

**Isolation and Characterization of the
Genes Involved in the Biosynthesis of
Secologanin, an Intermediate in the
Formation of Vincristine, a Potent Anti-
Cancer Agent**

Thesis Submitted to **AcSIR**
For the Award of the Degree of
DOCTOR OF PHILOSOPHY
In **Biological Sciences**



By
Krithika Ramakrishnan
10BB13J26031

Under the guidance of
Dr. H. V. Thulasiram

CSIR-National Chemical Laboratory
Pune-411008, India
December 2016

You never fail until you stop trying.

-Albert Einstein

CERTIFICATE

This is to certify that the work incorporated in this Ph.D. thesis entitled “**Isolation and Characterization of the Genes Involved in the Biosynthesis of Secologanin, an Intermediate in the Formation of Vincristine, a Potent Anti-Cancer Agent**” submitted by **Miss. Krithika Ramakrishnan** to Academy of Scientific and Innovative Research (AcSIR) in fulfillment of the requirements for the award of the Degree of **Doctor of Philosophy** in **Biological Sciences**, embodies original research work under my supervision/guidance. I 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 obtained from other sources has been duly acknowledged in the thesis. Any text, illustration, table etc., used in the thesis from other sources, have been duly cited and acknowledged.

Krithika Ramakrishnan

Dr. H. V. Thulasiram
(Research Guide)
Division of Organic Chemistry
CSIR-National Chemical Laboratory
Pune-411008, India

DECLARATION

I, Krithika Ramakrishnan, hereby declare that the work incorporated in the thesis entitled “**Isolation and Characterization of the Genes Involved in the Biosynthesis of Secologanin, an Intermediate in the Formation of Vincristine, a Potent Anti-Cancer Agent**” submitted by me to to the Academy of Scientific and Industrial Research (AcSIR), for the degree of **Doctor of Philosophy in Biological Sciences** has been carried out by me at CSIR-National Chemical Laboratory under the guidance of Dr. H. V. Thulasiram. This work is original and has not been submitted to any other university or institution for the award of any degree or diploma. Such material, as has been obtained from other sources, has been duly acknowledged.

Krithika Ramakrishnan



Dedicated to my Parents.....

Acknowledgement

At the *itineris terminus* of my Ph.D., I would like to extend my sincere appreciation and gratefulness to all those who made this Ph.D. thesis possible.

First and foremost, I would like to extend my sincere gratitude to my research guide, Dr. H. V. Thulasiram for giving me the opportunity and introducing me to this exciting field of research. I sincerely thank him for his vital suggestions, invaluable advice, encouragement and continuous support, throughout my tenure.

I earnestly acknowledge the Lady Tata Memorial Trust, Mumbai, for providing with the Junior and Senior Research Scholarships, which enabled me to carry out my research at CSIR-NCL. I am grateful to The Director, CSIR-National Chemical Laboratory (CSIR-NCL), Pune and HOD, Division of Organic Chemistry, for infrastructure and support to realize my research work.

I would like to take this opportunity to express my obligations towards Dr. Mala Rao, Dr. Robert Raja and Dr. Paul Ratnasamy for pioneering me to CSIR-NCL. I am enormously obliged to Dr. Jomon Joseph, Dr. Ashok P. Giri, Dr. T. G. Ajithkumar, Dr. J. M. Khire, Dr. S. Dhanashekar and Dr. Kiran Kulkarni for their continuous evaluation of my Ph.D. work and their valued advice.

I express my heartfelt gratitude to my professors Dr. V. Ramamurthy, Dr. M. Ananthasubramaniam, Dr. P. Rani and Dr. Sivakumaran K. for sowing scientific seeds and inspiring me with their scientific attitude. I thank my teacher, Mrs. Lalitha Malviya whose teaching inspired me to never let go.

I thank Prof. K. N. Ganesh, Dr. Mahesh J. Kulkarni, Dr. C. V. Ramana, Dr. C. G. Suresh, Dr. Sanjay Nene, Dr. Alok Sen, Dr. Muthukrishnan, Dr. V. A. Kumar and Dr. Moneesha Fernandes for their support.

True friends stand by us during our best and worst times, all the while showering us with unconditional support and encouragement. I would like to acknowledge my friends, Soniya, Prabhakar, Shiva, Avinash, Vijayshree, Uttara, Nilofer, Rincy, Anurag and Sudhakar, for always being there for me. My friends from college, Thawfeek, Karthika, Vidhya, Sharanya, Padmaja, Manish and Muthu always encouraged me to move ahead.

I express my sincere gratitude to Soniya, Prabhakar, Avinash, Rani, Anirban, Manoj, Sunil, Chaya, Nivedita and Swati for their help in completing critical parts of my work, which would have been truly difficult without their help.

I would like to extend a huge thanks to my labmates, Pankaj, Swati, Atul, Prabhakar, Namdev, Sudhir, Rincy, Deepti, Nilofer, Trushna, Saikat, Devdutta, Dipesh, Harshal, Ashwini, Pratibha, Sreekant, Prasad, Sharanbasappa, Rohil, Soniya, Priya, Shiva, Avinash, Vijayshree, Uttara, Rani, Anurag, Jennifer, Montu, Aarthi, Supriya, Anirban, Manoj, Balaji, Fayaz, Pruthviraj, Ajay, Vijay, Yashpal, Nivedita, Chaya, Nagendra, Prabhu, Debjyoti, Shrikant, Sharvani, Govinda, Shruti, Aaditi, Singaravelan, Jeevitha, Bhagyashree, Sudha, Pooja, Sneha, Ashish, Ashwini, Nikita, Datta, Sonia Varkey, Sridhar, Joginder, Rahul and Mr. Khandekar for providing a healthy and pleasant environment for work.

My warm thanks to my room-mate, Poonam, who was indeed like an elder sister to me when I first came to Pune; Thank you, Poonam. I would like to give a special mention to some of my friends and colleagues at NCL, namely, Gyan, Yashwant, Pradeep, Shweta, Neena, Neha, Sandeep, Suresh, Arati, Reema, Rubina, Rashmi, Kedar, Tejal, Vishwa, Jyoti, Krishanu, Yugendra, Rupa, Prajakta, Arundhati, Sachin, Sarang, Richa, Tanaya, Dipanjan, Chaitanya Kiran, Rahul S., Gokul, Anand, Anuja, Deepanwita, Leena and Vadde Ramu.

“Families are God’s greatest gift to us.” I cannot thank my parents, Mr. P. R. Ramakrishnan and Mrs. Usha Ramakrishnan, enough for believing in me; if it wasn’t for their blessing and their faith in me, I would not have been able to come so far; Thank you Mom and Dad, for making my dreams yours. I thank my elders, especially my grandparents, Late Mr. P. S. Ramanathan, Late Mrs. Valli Ramanathan, Mr. K. N. Kalyanakrishnan and Mrs. Saroja Kalyanakrishnan for bestowing their affection and blessings on me and being my strength. I would also like to express my deepest regards for my Uncle, Dr. P. R. Subramanian for his constant support and encouragement during this journey and for truly inspiring me to walk on this path.

Last, but not least, I would like to thank the Almighty for giving me strength, health and courage in every phase of my life, to overcome all my hurdles and surge ahead with a new zest every time.

--Krithika Ramakrishnan

Contents

Acknowledgements	i
Contents	iii
List of tables	xiii
List of schemes	xiv
List of figures	xv-xxi
Abbreviations	xxii-xxiii
Abstract	xxiv-xxx
Chapter 1: Introduction	1-51
1.1 Plant biosynthetic pathways	3
1.2 <i>Catharanthus roseus</i>	4
1.3 Isoprenoids / Terpenoids	5
1.4 Classification of terpenes/isoprenes	5
1.4.1 Hemiterpenes	6
1.4.2 Monoterpenes	6
1.4.3 Sesquiterpenes	7
1.4.4 Diterpenes	8
1.4.5 Triterpenes	9
1.4.6 Tetraterpenes	10
1.4.7 Polyterpenes	12
1.5 Biosynthesis of Isoprenoids	12
1.6 Alkaloids	20
1.6.1 Types of Alkaloids	20
1.6.1.1 True Alkaloids	20
1.6.1.2 Protoalkaloids	21
1.6.1.3 Polyamine alkaloids	21
1.6.1.4 Peptide and cyclopeptide alkaloids	21
1.6.1.5 Pseudoalkaloids	22

1.6.1.6	Isoquinoline alkaloids	22
1.6.2	Terpene Indole Alkaloids / Isoprenoid Indole Alkaloids (TIAs)	23
1.6.2.1	Ergot alkaloids	23
1.6.2.2	Monoterpene Indole alkaloids (MIAs)	24
1.6.2.3	Bisindole alkaloids	24
1.6.3	Biosynthesis of alkaloids	25
1.6.3.1	Biosynthesis of Secologanin	25
1.6.3.2	Biosynthesis of Tryptamine	26
1.6.3.3	Biosynthesis of Vinblastine and Vincristine	26
1.7	Metabolic Engineering	29
1.7.1	Metabolic engineering for the production of isoprenoids	30
1.8	Scope of the work in monoterpene biosynthetic pathway	35
1.10	References	37

Chapter 2: Transcriptome sequencing, *de novo* assembly and functional annotation to gain insights for the secologanin biosynthetic pathway **52-75**

2.1	Introduction	54
2.2	Materials and Methods	55
2.2.1	Plant sample source	55
2.2.2	Reagents	55
2.2.3	Modified protocol for RNA isolation	56
2.2.4	Modified protocol for Isolation of total RNA using Spectrum TM Plant Total RNA Isolation Kit from Sigma	57
2.2.5	Quantification of total RNA	57
2.2.6	RT-PCR	57
2.2.7	Transcriptome sequencing	58
2.2.8	Raw data processing	58
2.2.9	Transcriptome assembly	58
2.2.10	Transcriptome annotation	59
2.3	Results and Discussion	60
2.3.1	Optimization of RNA isolation from various tissues of	60

	<i>Catharanthus roseus</i>	
2.3.2	Transcriptome sequencing and screening of genes involved in the biosynthesis of Secologanin in <i>Catharanthus roseus</i>	64
2.3.2.1	cDNA library preparation and sequencing	64
2.3.2.2	<i>De novo</i> transcriptome assembly	64
2.3.2.3	Functional annotation of unigenes	64
2.3.2.3.1	KASS analysis	65
2.3.2.3.2	Virtual Ribosome	67
2.3.2.3.3	Pfam analysis	68
2.3.3	Identification of transcripts involved in the MVA, MEP and Secologanin Biosynthetic Pathways	69
2.4	Conclusion	73
2.5	References	74

Chapter 3: Cloning, expression and characterization of the genes involved in iridoid biosynthesis **76-127**

3.1	Introduction	78
3.2	Materials and Methods	80
3.2.1	Materials used in the study	80
3.2.1.1	Strains and plasmids used in the study	80
3.2.1.2	Kits and reagents used in the study	80
3.2.1.3	Buffer compositions	80
3.2.1.3.1	CrGDS	80
3.2.1.3.2	CrGS	81
3.2.1.3.3	CrG10H	81
3.2.1.3.4	Cr10HGO	82
3.2.1.3.5	CrIDS	82
3.2.2	RNA isolation and cDNA preparation	82
3.2.3	Sequence analysis and ORF construction	83
3.2.4	Full-length gene isolation and cloning into expression vector	84
3.2.4.1	Isolation and cloning of ORF of <i>CrGDS</i> in pET32a expression	84

	vector	
3.2.4.2	Isolation and cloning of ORF of <i>CrGS</i> in pET28a expression vector	84
3.2.4.3	Isolation and cloning of ORF of <i>CrG10H</i> in pYES2 yeast expression vector	85
3.2.4.4	Isolation and cloning of ORF of <i>Cr10HGO</i> in pRSET B expression vector	85
3.2.4.5	Isolation and cloning of ORF of <i>CrIDS</i> in pET32a expression vector	86
3.2.5	Heterologous expressions and protein purifications	86
3.2.5.1	Heterologous expression and protein purification of <i>CrGDS</i>	86
3.2.5.2	Heterologous expression and protein purification of <i>CrGS</i>	87
3.2.5.3	Yeast expression and Microsome Preparation of <i>CrG10H</i>	87
3.2.5.4	Heterologous expression and protein purification of <i>Cr10HGO</i>	88
3.2.5.5	Heterologous expression and protein purification of <i>CrIDS</i>	88
3.2.6	Enzymatic characterization	89
3.2.6.1	Enzyme assay of geranyl diphosphate synthase (<i>CrGDS</i>)	89
3.2.6.2	Enzyme assay of geraniol synthase (<i>CrGS</i>)	89
3.2.6.3	Enzyme assay of geraniol 10-hydroxylase (<i>CrG10H</i>)	89
3.2.6.4	Enzyme assay of 10-hydroxygeraniol dehydrogenase (<i>Cr10HGO</i>)	90
3.2.6.5	Enzyme assay of iridoid synthase (<i>CrIDS</i>)	90
3.2.7	GC/GC-MS analyses	90
3.2.8	Determination of optimum temperature for Cr10HGO Enzyme assays	91
3.2.9	Determination of optimum pH	91
3.2.10	Determination of Kinetic Parameters of Cr10HGO	91
3.2.11	Combined assays	91
3.2.12	Phylogenetic analysis	92
3.3	Results and Discussion	93

3.3.1	Isolation, cloning, expression and characterization of geranyl diphosphate synthase (<i>CrGDS</i>)	93
3.3.1.1	Isolation and cloning of geranyl diphosphate synthase (<i>CrGDS</i>)	93
3.3.1.2	Bacterial expression and protein purification	94
3.3.1.3	Enzymatic characterization of <i>CrGDS</i>	95
3.3.2	Isolation, cloning, expression and characterization of Geraniol synthase (<i>CrGS</i>)	96
3.3.2.1	Isolation and cloning of geraniol synthase (<i>CrGS</i>)	96
3.3.2.2	Bacterial expression and protein purification	97
3.3.2.3	Enzymatic characterization of <i>CrGS</i>	98
3.3.3	Isolation, cloning, expression and characterization of Geraniol 10-hydroxylase (<i>CrG10H</i>)	99
3.3.3.1	Isolation and cloning of geraniol 10-hydroxylase (<i>CrG10H</i>)	99
3.3.3.2	Yeast expression and microsomes preparation	100
3.3.3.3	Enzymatic characterization of <i>CrG10H</i>	101
3.3.4	Isolation, cloning, expression and characterization of 10-hydroxygeraniol dehydrogenase (<i>Cr10HGO</i>)	102
3.3.4.1	Isolation and cloning of 10-hydroxygeraniol dehydrogenase (<i>Cr10HGO</i>)	102
3.3.4.2	Bacterial expression and protein purification	103
3.3.4.3	Enzymatic characterization of <i>Cr10HGO</i>	104
3.3.4.4	Kinetic characterization of <i>Cr10HGO</i>	107
3.3.5	Isolation, cloning, expression and characterization of Iridoid Synthase (<i>CrIDS</i>)	110
3.3.5.1	Isolation and cloning of Iridoid Synthase (<i>CrIDS</i>)	110
3.3.5.2	Bacterial expression and protein purification	111
3.3.5.3	Enzymatic characterization of <i>CrIDS</i>	112
3.3.6	Cascaded enzyme activity	113
3.3.7	Phylogenetic analysis of monoterpene synthases isolated from <i>Catharanthus roseus</i>	120
3.4	Conclusion	122

3.5	Appendix: Agarose gel electrophoresis for colony PCR screening for gene cloning	123
3.6	References	126
Chapter 4: Dissection of the activity of 10-hydroxygeraniol dehydrogenase (Cr10HGO) by site-directed mutagenesis studies		128-162
4.1	Introduction	130
4.2	Materials and Methods	132
4.2.1	Computational modeling and docking	132
4.2.2	Site-Directed mutagenesis	132
4.2.3	Heterologous expression and purification of Cr10HGO mutants	133
4.2.4	Enzymatic characterization and product analysis	134
4.2.5	Determination of kinetic parameters of Cr10HGO mutants	134
4.2.6	Substrate Specificity Studies	134
4.3	Results	135
4.3.1	Homology modeling and docking	135
4.3.2	Catalytic mechanism of Cr10HGO on 10-hydroxygeraniol	137
4.3.3	Site-directed mutagenesis and characterization	138
4.3.4	Heterologous Expression and Purification of Cr10HGO mutants	139
4.3.5	Residues not significantly affecting the activity	141
4.3.6	Residues causing significant changes in the product profile of Cr10HGO	142
4.3.6.1	Serine 301 (S301A)	142
4.3.6.2	Asparagine 115 (N115H)	143
4.3.7	Residues involved in the increased production of the di-aldehyde	145
4.3.7.1	Tyrosine 116 (Y116F)	145
4.3.7.2	Serine 256 and Alanine 302 (S256Y and A302F)	146
4.3.7.3	Cysteine 98 (C98A)	148
4.3.7.4	Histidine 51 and Tryptophan 61 (H51F and W61L)	149
4.3.7.5	Threonine 170 (T170A)	151
4.3.8	Double Mutants	152

4.3.9	Substrate Specificity Studies	154
4.3.10	Kinetic characterization of Cr10HGO mutants	155
4.4	Discussion	157
4.5	Conclusion	159
4.6	References	161

Chapter 5: Engineering of iridoid synthase (CrIDS) by site-directed mutagenesis to understand its activity profile **163-187**

5.1	Introduction	165
5.2	Materials and Methods	166
5.2.1	Computational modeling and docking	166
5.2.2	Site-Directed mutagenesis	166
5.2.3	Heterologous expression and purification of CrIDS mutants	167
5.2.4	Enzymatic characterization and product analysis	168
5.2.5	Determination of kinetic parameters of CrIDS mutants	168
5.3	Results	169
5.3.1	Homology modeling and docking	169
5.3.2	Mechanism of action of CrIDS on 10-oxogeranial	170
5.3.3	Site-directed mutagenesis and characterization	171
5.3.4	Heterologous Expression and Purification of CrIDS mutants	172
5.3.5	Residues showing a sharp decline in the activity of CrIDS	173
5.3.5.1	Tyrosine 176 (Y176F)	173
5.3.6	Residues not significantly affecting the activity	175
5.3.7	Residues involved in the enhanced formation of nepetalactol	176
5.3.7.1	Lysine 144 (K144L)	176
5.3.7.2	Tryptophan 107 (W107F)	178
5.3.7.3	Phenylalanine 175 (F175M)	179
5.3.7.4	Methionine 211 (M211A, M211F, M211S, M211K, M211H)	180
5.3.8	Kinetic characterization of CrIDS mutants	182
5.4	Discussion	184
5.5	Conclusion	186

5.6	References	187
Chapter 6: Screening and cloning of genes involved in the biosynthesis of secologanin from iridoid		188-217
6.1	Introduction	190
6.2	Materials and methods	192
6.2.1	Materials used in the study	192
6.2.1.1	Strains and plasmids used in the study	192
6.2.1.2	Kits and reagent used in the study	192
6.2.1.3	Buffer compositions	192
6.2.1.3.1	Cr7DLS, Cr7DLH and CrSLS	192
6.2.1.3.2	CrDLGT	193
6.2.1.3.3	CrLAMT	193
6.2.2	RNA isolation and cDNA preparation	194
6.2.3	Sequence analysis and ORF construction	194
6.2.4	Full-length gene isolation and cloning into expression vector	195
6.2.4.1	Isolation and cloning of ORF of <i>CrCPR</i> in MCS1 of pESC-URA yeast duet vector	195
6.2.4.2	Isolation and cloning of ORF of <i>Cr7DLS</i> in the MCS2 of pESC-URA expression vector	195
6.2.4.3	Isolation and cloning of ORF of CrDLGT in pET28a expression vector	196
6.2.4.4	Isolation and cloning of ORF of <i>Cr7DLH</i> in MCS2 of pESC-URA yeast expression vector	196
6.2.4.5	Isolation and cloning of ORF of <i>CrLAMT</i> in pRSET B expression vector	197
6.2.4.6	Isolation and cloning of ORF of <i>CrSLS</i> in the MCS2 of pESC-URA yeast expression vector	197
6.2.5	Expressions and protein purifications	198
6.2.5.1	Yeast expression of <i>Cr7DLS</i>	198
6.2.5.2	Heterologous expression and protein purification of <i>CrDLGT</i>	198

6.2.5.3	Yeast expression of <i>Cr7DLH</i>	198
6.2.5.4	Heterologous expression and protein purification of <i>CrLAMT</i>	199
6.2.5.5	Yeast expression of <i>CrSLS</i>	199
6.2.6	Phylogenetic analysis	200
6.3	Results	201
6.3.1	Isolation, and cloning of Cytochrome P450 reductase (<i>CrCPR</i>)	201
6.3.1.1	Isolation, and cloning of Cytochrome P450 reductase (<i>CrCPR</i>)	201
6.3.2	Isolation, cloning and expression characterization of 7-deoxyloganin synthase (<i>Cr7DLS</i>)	202
6.3.2.1	Isolation and cloning of 7-deoxyloganin synthase (<i>Cr7DLS</i>)	202
6.3.2.2	Yeast expression	203
6.3.3	Isolation, cloning and expression of 7-deoxyloganic acid glucosyltransferase (<i>CrDLGT</i>)	203
6.3.3.1	Isolation and cloning of 7-deoxyloganic acid glucosyltransferase (<i>CrDLGT</i>)	203
6.3.3.2	Bacterial expression and protein purification	204
6.3.4	Isolation, cloning and expression of 7-deoxyloganin hydroxylase (<i>Cr7DLH</i>)	205
6.3.4.1	Isolation and cloning of 7-deoxyloganin hydroxylase (<i>Cr7DLH</i>)	205
6.3.4.2	Yeast expression	206
6.3.5	Isolation, cloning, expression and characterization of loganic acid methyltransferase (<i>CrLAMT</i>)	206
6.3.5.1	Isolation and cloning of loganic acid methyltransferase (<i>CrLAMT</i>)	206
6.3.5.2	Bacterial expression and protein purification	207
6.3.6	Isolation, cloning and expression of secologanin synthase (<i>CrSLS</i>)	208
6.3.6.1	Isolation and cloning of secologanin synthase (<i>CrSLS</i>)	208
6.3.6.2	Yeast expression	209
6.3.7	Phylogenetic analysis of monoterpene synthases isolated from <i>Catharanthus roseus</i>	209
6.4	Discussion and Conclusion	211
6.5	Appendix: Agarose gel electrophoresis for colony PCR screening	213

	for gene cloning	
6.6	References	216
	Addendum	218-219
	Sequences of monoterpene synthases isolated from <i>Catharanthus roseus</i>	220-228
	Thesis Summary	229-230
	List of publications from thesis	231
	Curriculum Vitae	232-235

List of tables

1.1	Current knowledge of characterized plant MVA pathway enzymes	13
1.2	Current knowledge of characterized plant MEP pathway enzymes	17
2.1	Transcripts involved in the MVA, MEP and Secologanin biosynthetic pathways	70
3.2.1	Primer sequence for isolation of full length ORF of the genes involved in Iridoid biosynthesis	83
3.3.1	Kinetic Constants for Cr10HGO with various substrates	107
4.2.1	Primers for Cr10HGO site-directed mutagenesis	132
4.3.1	Product profiles of assays of Cr10HGO wild type and mutants with 10-hydroxygeraniol	152
4.3.2	Substrate specificity study of Cr10HGO wild type and mutants	154
4.3.3	Steady state kinetics values of Cr10HGO mutants	155
5.2.1	Primers for CrIDS site-directed mutagenesis	166
5.3.1	Product profiles of assays of CrIDS wild type and mutants with 10-oxogeranial	182
5.3.2	Steady state kinetics values of CrIDS mutants	182
6.2.1	Primer sequence for isolation of full length ORF of the genes involved in the biosynthesis of Secologanin from Iridoids	194

List of schemes

1.1	Biosynthesis of hemiterpenes	6
1.2	Biosynthesis of monoterpenes	7
1.3	Biosynthesis of sesquiterpenes	8
1.4	Biosynthesis of diterpenes	9
1.5	Biosynthesis of triterpenes	10
1.6	Biosynthesis of tetraterpenes	11
1.7	Mevalonate Pathway (MVA pathway)	13
1.8	Methylerythritol Pathway (MEP pathway)	16
1.9	Secologanin Biosynthetic Pathway	26
1.10	Biosynthesis of Tryptamine	27
1.11	Biosynthetic Pathway of Vincristine and Vinblastine	28
1.12	Chemical synthesis of Artemisinin	31
1.13	Chemical synthesis of Taxol	32
3.3.1	Schematic representation of <i>CrGDS</i> assay	95
3.3.2	Schematic representation of <i>CrGS</i> assay	98
3.3.3	Schematic representation of <i>CrG10H</i> assay	101
3.3.4	Schematic representation of <i>Cr10HGO</i> assay	104
3.3.5	Schematic representation of <i>CrIDS</i> assay	112
3.3.6	Concerted reaction of <i>Cr10HGO</i> and <i>CrIDS</i> on 10-hydroxygeraniol to form (1 <i>R</i> , 4 <i>aS</i> , 7 <i>S</i> , 7 <i>aR</i>)-nepetalactol	114
3.3.7	Concerted reaction of <i>CrGDS</i> , <i>CrGS</i> and <i>CrG10H</i>	116
3.3.8	Concerted reaction of <i>CrG10H</i> , <i>Cr10HGO</i> and <i>CrIDS</i>	118
3.3.9	Concerted reaction of <i>CrGDS</i> , <i>CrGS</i> , <i>CrG10H</i> , <i>Cr10HGO</i> and <i>CrIDS</i>	121

List of figures

1.1	Compartmentalization of the MVA and MEP Pathways	18
1.2	Isoprenoid carbon skeletons	19
1.3	Overview of terpene biosynthesis	20
1.4	Some examples of true alkaloids from plants and fungi	21
1.5	Some examples of Protoalkaloids from plants	21
1.6	Some examples of polyamine alkaloids from plants and animals	21
1.7	Some examples of Peptide and cyclopeptide alkaloids from plants	22
1.8	Some examples of pseudoalkaloids from plants	22
1.9	Some examples of isoquinoline alkaloids from plants	23
1.10	Some examples of ergot alkaloids from fungi	23
1.11	Some examples of Monoterpene Indole alkaloids (MIAs) from plants	24
1.12	Some examples of bisindole alkaloids from plants	25
1.13	Metabolic Engineering for the production of Artemisinic acid	32
1.14	Semi-synthesis of Artemisinin and Taxol in <i>S. cerevisiae</i>	33
1.15	Scope of the work in monoterpene biosynthesis	35
2.1	Agarose gel electrophoresis of total RNA isolated using Spectrum™ Plant Total RNA Isolation Kit from Sigma	60
2.2	Agarose gel electrophoresis of total RNA isolated using modified conventional protocol	61
2.3	Agarose gel electrophoresis of PCR of cDNAs with CrLAMT full-length primers for cDNA verification	62
2.4	Agarose gel electrophoresis of total RNA isolated using modified kit protocol	62
2.5	PCR of cDNA from leaves, stem and roots of white- and pink-flowering plants	63
2.6	KAAS analysis of unigenes for KEGG pathway mapping	65

2.7	KEGG Pathway analysis for unigenes present in the Terpenoid Biosynthetic Pathway	66
2.8	KEGG Pathway analysis for unigenes present in the Monoterpenoid Biosynthesis	67
2.9	ORF Prediction using Virtual Ribosome-V1.1	68
2.10	Pfam Analysis of transcripts with length ≥ 50 amino acids	69
3.1.1	Iridoid Biosynthetic Pathway cyclizes 10-oxogeraniol into equilibrium mixture of cis-trans-nepetalactol and iridodials	78
3.3.1	CrGDS full length ORF amplification	93
3.3.2	SDS-PAGE for CrGDS protein purification in pET 32a	94
3.3.3	SDS-PAGE for truncated CrGDS protein purification in pET 28a	95
3.3.4	Total Ion Chromatogram (TIC) of Assay of CrGDS with IPP and DMAPP as substrates	96
3.3.5	CrGS full length ORF amplification	97
3.3.6	SDS-PAGE for CrGS protein purification in pET 28a	98
3.3.7	Total Ion Chromatogram of Assay of CrGS with GPP as substrate	99
3.3.8	CrG10H full length ORF amplification	100
3.3.9	SDS-PAGE for CrG10H protein in pYES2	101
3.3.10	Ion Chromatogram of Assay of CrG10H with geraniol as substrate	102
3.3.11	Cr10HGO full length ORF amplification	103
3.3.12	SDS-PAGE for Cr10HGO protein purification in pRSET B	104
3.3.13	Total ion chromatograms (TICs) for 10-hydroxygeraniol dehydrogenase (Cr10HGO)	106
3.3.14	Product-ratio studies of Cr10HGO	107
3.3.15	Michaelis-Menten Plot for 10-hydroxygeraniol kinetics of Cr10HGO	108
3.3.16	Michaelis-Menten Plot for NADP ⁺ kinetics of Cr10HGO	108
3.3.17	Michaelis-Menten Plot for 10-oxogeraniol kinetics of	108

	Cr10HGO	
3.3.18	Michaelis-Menten Plot for NADPH kinetics of Cr10HGO	109
3.3.19	Michaelis-Menten Plot for 10-hydroxygeranial kinetics of Cr10HGO	109
3.3.20	Michaelis-Menten Plot for 10-oxogeranial kinetics of Cr10HGO	109
3.3.21	Michaelis-Menten Plot for 10-hydroxynerol kinetics of Cr10HGO	110
3.3.22	CrIDS full length ORF amplification	110
3.3.23	Box-shade analysis of CrIDS protein sequences	111
3.3.24	SDS-PAGE for CrIDS protein purification in pET 32a	112
3.3.25	Total Ion Chromatogram of Assay of CrIDS with 10-oxogeranial as substrate	113
3.3.26	Concerted reaction of Cr10HGO and CrIDS	114
3.3.27	Comparison of TICs of CrIDS, Cr10HGO+CrIDS and CrG10H+Cr10HGO+CrIDS assays with standard	115
3.3.28	Comparison of TICs of CrGDS, CrGDS+CrGS and CrGDS+CrGS+CrG10H enzyme assays	117
3.3.29	Comparison of TICs of CrG10H, CrG10H+Cr10HGO and CrG10H+Cr10HGO+CrIDS enzyme assays	119
3.3.30	Phylogenetic analysis of the monoterpene synthases isolated from <i>C. roseus</i>	120
3.5.1	Colony PCR screening for <i>CrGDS</i> truncated ORF cloned in pET 28a	122
3.5.2	Colony PCR screening of full-length ORF of <i>CrGS</i> cloned in pET 28a	122
3.5.3	Colony PCR Screening for <i>CrG10H</i> full length ORF cloned in pYES2	123
3.5.4	Colony PCR screening for full-length ORF cloning of <i>Cr10HGO</i> in pRSETB vector	123
3.5.5	Colony PCR screening for ORF of <i>CrIDS</i> cloned in pET 32a	124

	vector	
3.5.6	¹ H NMR spectrum of synthesized (1 <i>R</i> , 4 <i>aS</i> , 7 <i>S</i> , 7 <i>aR</i>)-nepetalactol in CDCl ₃	124
3.5.7	¹³ C NMR spectrum of synthesized (1 <i>R</i> , 4 <i>aS</i> , 7 <i>S</i> , 7 <i>aR</i>)-nepetalactol in CDCl ₃	125
3.5.8	DEPT NMR spectrum of synthesized (1 <i>R</i> , 4 <i>aS</i> , 7 <i>S</i> , 7 <i>aR</i>)-nepetalactol in CDCl ₃	125
4.3.1 A	Homology model of 10-hydroxygeraniol dehydrogenase built using co-ordinates of sinapyl alcohol dehydrogenase from <i>Populus tremuloides</i> as a template	135
4.3.1 B	10-hydroxygeraniol docked with Cr10HGO model	135
4.3.2	Ramachandran Plot of the built model of Cr10HGO	136
4.3.3 A	Cr10HGO mediated reaction in <i>C. roseus</i> for biosynthesis of iridoids	138
4.3.3 B	Product profile of Cr10HGO enzyme assay	138
4.3.4	Homology model of Cr10HGO showing amino acid interactions with the substrate	139
4.3.5	SDS gel images of Cr10HGO mutants' protein purification	140
4.3.6	Total Ion Chromatograms (TICs) of substrate control, Cr10HGO Wild Type, Cr10HGO_H55N and Cr10HGO_N115L	141
4.3.7	Total Ion Chromatograms (TICs) of substrate control, Cr10HGO Wild Type, Cr10HGO_S301A and Cr10HGO_N115H	144
4.3.8	Total Ion Chromatograms (TICs) of substrate control, Cr10HGO Wild Type and Cr10HGO_Y116F	146
4.3.9	Total Ion Chromatograms (TICs) of substrate control, Cr10HGO Wild Type, Cr10HGO_S256Y and Cr10HGO_A302F	147
4.3.10	Total Ion Chromatograms (TICs) of substrate control, Cr10HGO Wild Type and Cr10HGO_C98A	148

4.3.11	Total Ion Chromatograms (TICs) of substrate control, Cr10HGO Wild Type, Cr10HGO_H51F and Cr10HGO_W61L	150
4.3.12	Total Ion Chromatograms (TICs) of substrate control, Cr10HGO Wild Type and Cr10HGO_T170A	151
4.3.13	Total Ion Chromatograms (TICs) of substrate control, Cr10HGO Wild Type, Cr10HGO_DM1 and Cr10HGO_DM2	153
4.4.1	Total Ion Chromatograms (TICs) of Cr10HGO_WT, Cr10HGO_M5, Cr10HGO_M7, Cr10HGO_M9, Cr10HGO_M10, Cr10HGO_M11 and Cr10HGO_M13	157
4.4.2	Mutated amino acid positions around the substrate 10-hydroxygeraniol and NADP ⁺	158
5.3.1 A	Homology model of Iridoid synthase built using co-ordinates of Progesterone 5 β -Reductase from <i>Digitalis lanata</i> as a template	169
5.3.1 B	10-oxogeraniol docked with CrIDS model	169
5.3.2	Ramachandran Plot of the built model of CrIDS	170
5.3.3	CrIDS mediated reaction in <i>C. roseus</i> for biosynthesis of iridoids and Product profile of CrIDS enzyme assay	171
5.3.4	Homology model of CrIDS showing amino acid interactions with the substrate	172
5.3.5	SDS gel images of Cr10HGO mutants' protein purification	173
5.3.6	Total Ion Chromatograms (TICs) of substrate control, CrIDS Wild Type and CrIDS_Y176F assays	174
5.3.7	Total Ion Chromatograms (TICs) of substrate control, CrIDS Wild Type and C. CrIDS_I38A assays	175
5.3.8	Total Ion Chromatograms (TICs) of substrate control, CrIDS Wild Type and CrIDS_K144L assays	177
5.3.9	Total Ion Chromatograms (TICs) of A. Substrate Control, B. CrIDS Wild Type Assay, C. CrIDS_W107F Assay	178
5.3.10	Total Ion Chromatograms (TICs) of substrate control, CrIDS Wild Type and C. CrIDS_F175M assays	179

5.3.11	Total Ion Chromatograms (TICs) of substrate control, CrIDS Wild Type and CrIDS_M211A assays	180
5.3.12	Total Ion Chromatograms (TICs) of CrIDS Wild Type, CrIDS-M211F, CrIDS-M211S, CrIDS-M211K and CrIDS-M211H	181
5.5.1	Amino acids of CrIDS chosen for mutation, positioned around the substrate, 10-oxogeraniol and NADP ⁺	186
6.1.1	Biosynthesis of strictosidine via the key intermediate secologanin	191
6.3.1	CrCPR full length ORF amplification	201
6.3.2	Cr7DLS full length ORF amplification	202
6.3.3	SDS-PAGE for Cr7DLS protein in pESC-URA+CrCPR	203
6.3.4	CrDLGT full length ORF amplification	204
6.3.5	SDS-PAGE for CrDLGT protein purification in pET 28a	204
6.3.6	Cr7DLH full length ORF amplification	205
6.3.7	SDS-PAGE for Cr7DLH protein in pESC-URA+CrCPR	206
6.3.8	CrLAMT full length ORF amplification	207
6.3.9	SDS-PAGE for CrLAMT protein purification in pRSET B	207
6.3.10	CrSLS full length ORF amplification	208
6.3.11	SDS-PAGE for CrSLS protein expression in pESC-URA+CrCPR	209
6.3.12	Phylogenetic analysis of the monoterpene synthases isolated from <i>C. roseus</i>	210
6.4.1	Formation of Secologanin from Iridoids	211
6.5.1	Colony PCR screening for <i>CrCPR</i> ORF cloned in MCS1 of pESC-URA vector	213
6.5.2	Colony PCR screening of full-length ORF of <i>Cr7DLS</i> in MCS2 of pESC-URA+CrCPR	213
6.5.3	Colony PCR Screening for <i>CrDLGT</i> full length ORF cloned in pET28a	214
6.5.4	Colony PCR screening for full-length ORF <i>Cr7DLH</i> in MCS2 of pESC-URA+CrCPR	214

6.5.5	Colony PCR screening for ORF of <i>CrLAMT</i> cloned in pRSETB vector	215
6.5.6	Colony PCR screening for full-length ORF <i>CrSLS</i> in MCS2 of pESC-URA+CrCPR	215

Abbreviations

Å	Angstrom
Amp	Ampicillin
AA	Amino acid
BLAST	Basic Local Alignment Search Tool
bp	Base pair
Cam	Chloramphenicol
cDNA	Complementary DNA
CTAB	Cetyl Trimethyl Ammonium Bromide
C-terminal	Carboxy terminal
DEPC	Diethylpyrocarbonate
DTT	Dithiothritol
EDTA	Ethylene diamine tetra acetic acid disodium salt
FAD	Flavin adenine dinucleotide
FMN	Flavin mononucleotide
IPTG	Isopropyl β -D-1-thiogalactopyranoside
Kan	Kanamycin
kDa	Kilo dalton
KEGG	Kyoto Encyclopedia of Genes and Genomes
LA	Luria agar
LB	Luria broth
mg	Milligram
mL	Millilitre
μ g	Microgram
μ L	Microlitre
μ M	Micromolar
NADP	Nicotinamide adenine dinucleotide phosphate
NADPH	Nicotinamide adenine dinucleotide phosphate reduced
NCBI	National Center for Biotechnology Information
ng	Nanogram
N-terminal	Amino terminal

OD	Optical density
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
pI	Isoelectric point
PMSF	Phenyl methyl sulphonyl fluoride
Pfam	Protein families
PVPP	Polyvinylpyrrolidone
RACE	Rapid amplification of cDNA ends
RNA	Ribonucleic acid
Rt	Retention time
Sec	Second
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TB	Terrific broth

Thesis abstract

The *Catharanthus* alkaloids comprise a group of about 130 terpenoid indole alkaloids¹ of which Vinblastine and Vincristine have emerged as promising anti-cancer agents. These two compounds are synthesized via an intermediate, secologanin, the formation of which greatly depends on the formation of iridoids. This work describes the isolation and extensive characterization of the genes involved in the biosynthesis of iridoids in *Catharanthus roseus*, their scope for metabolic engineering and directed evolution of two important enzymes in the pathway. We have explored the identification and characterization of five genes involved in the Iridoid Biosynthetic Pathway, namely, Geranyl diphosphate synthase (CrGDS), Geraniol synthase (CrGS), Geraniol 10-hydroxylase (CrG10H), 10-hydroxygeraniol dehydrogenase (Cr10HGO) and Iridoid synthase (CrIDS) from the transcriptomic data of leaves, roots and stem tissues. Substrate specificity studies indicated that, Cr10HGO has good activity on substrates such as 10-hydroxygeraniol, 10-oxogeraniol or 10-hydroxygeranial over monohydroxy linear terpene derivatives. It was also observed that incubation of 10-hydroxygeraniol with Cr10HGO and CrIDS in the presence of NADP⁺ yielded a major metabolite, which was characterized as (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol by comparing its retention time, mass fragmentation pattern, and co-injection studies with that of the synthesized compound². These results indicate that there is concerted activity of Cr10HGO with iridoid synthase in the formation of (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol, an important intermediate in iridoid biosynthesis, and this paved the way for a combinatorial reaction between the first five enzymes of the pathway. Further, analysis of the homology-based models of Cr10HGO and CrIDS built using co-ordinates of highest similarity structures, led to the identification of residues that may affect their activity. Site-directed mutagenesis studies of Cr10HGO and CrIDS enabled us to better understand the activity profile of these two enzymes and also to pinpoint the amino acids to which the activity profiles are attributed. In an attempt to complete the biosynthetic pathway of secologanin, transcripts matching with 7- deoxyloganin synthase (Cr7DLS), 7-deoxyloganic acid glucosyltransferase (CrDLGT), 7-deoxyloganin hydroxylase (7DLH), Loganic acid methyltransferase (CrLAMT) and Secologanin synthase (CrSLS) were identified. Their cloning and characterization is in progress.

Chapter 1

Introduction

A wide array of secondary metabolites is produced by plant as a part of their defense system against herbivores³, pests and pathogens⁴. Terpenes are a large class of secondary metabolites produced in plants, which are known to possess numerous biological properties. They are the major building blocks of various biosynthetic pathways. Terpenes and their chemical modifications terpenoids (isoprenoids) possess various properties, especially insecticidal. They are the primary constituents of the essential oils of many plants and have strong aromatic properties. Owing to these properties, they are widely used in the fragrance industry⁵, as flavor additives in food, in medicines, etc. Terpene Indole Alkaloids (TIAs) are a class of secondary metabolites in plants, which are known to possess varied biological properties like antimicrobial, antioxidant, anthelmintic, antifeedant, antisterility, antidiarrheal, antidiabetic, etc⁶. *Catharanthus roseus*, an evergreen shrub, is a rich source of these (TIAs), which are known for their pharmacological and therapeutic importance. Serpentine and ajmalicine are employed for their relaxing properties, in the treatment of hypertension⁷; 3,4-anhydrovinblastine, for its anti-tumor activity⁸; and vincristine and vinblastine as anti-cancer drugs⁹, after they were initially used as anti-diabetic agents. These terpenes are produced in minute quantities in plants and take many years for their biosynthesis in most cases. Hence, in an attempt to produce these therapeutic compounds in sufficient amounts in heterologous systems poses a great challenge in the field of metabolic engineering.

Chapter 2

Transcriptome sequencing, *de novo* assembly and functional annotation to gain insights for the secologanin biosynthetic pathway

Study of the transcriptome at a global level can provide insights into the gene function, transcriptional pattern and molecular basis of various cellular processes in plants. Transcriptome analysis has been comprehended as an essential step for basic and applied research in any organism. This chapter deals with optimization of the conventional and kit-based protocols for the isolation of total RNA from the various tissues of *Catharanthus roseus*.

These RNA samples were further used for sequencing on Illumina GAI Analyzer and Fastq file of 4.88 GB was generated. The 15485359 high quality paired-end reads were assembled into 76075 contigs with an optimized hash length of 49, average contig length of 1024 bp and an N50 value of 2375. This transcriptome raw reads were submitted to NCBI database with the accession ID: SRR1693842. Functional annotation was carried out using KEGG database, KAAS server, NCBI Nr-database, SwissProt/ Uniprot database and Pfam database. These analyses finally led to the identification of the unigenes related to Secologanin biosynthesis. This data was used for cloning and functional characterization of the genes involved in secologanin biosynthesis in *Catharanthus roseus*.

Chapter 3

Cloning, expression and characterization of the genes involved in iridoid biosynthesis

Catharanthus roseus is a major source of the monoterpene indole alkaloids (MIAs), which are of significant interest due to their therapeutic value. These molecules are formed through an intermediate, *cis-trans*-nepetalactol, a cyclized product of 10-oxogeraniol. One of the key enzymes involved in the biosynthesis of MIAs is an NAD(P)⁺ dependent oxidoreductase system, 10-hydroxygeraniol dehydrogenase (Cr10HGO), which catalyses the formation of 10-oxogeraniol from 10-hydroxygeraniol via 10-oxogeraniol or 10-hydroxygeraniol.

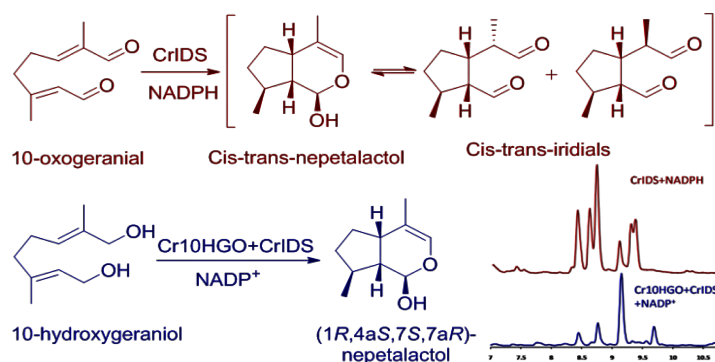


Figure 1: Concerted reaction of Cr10GO and CrIDS

This chapter describes the cloning and functional characterization of Cr10HGO from *C. roseus* and its role in the iridoid biosynthesis. We have shown that Cr10HGO possesses broad substrate specificity for 10-hydroxygeraniol, 10-oxogeraniol or 10-hydroxygeraniol over monohydroxy linear terpene derivatives. Concerted enzymatic function in the biosynthesis of *cis-trans*-nepetalactol has been demonstrated using 10-hydroxygeraniol and

NADP⁺ with Cr10HGO and CrIDS combined assay system. The stereochemistry of the enzymatic product was determined and is (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol, which is a key intermediate in the biosynthesis of iridoids and MIAs.

Further, the full-length unigenes, which showed high ranking with known Geranyl diphosphate synthase and Geraniol synthase from various sources, were used for the cloning and functional characterization of Geranyl diphosphate synthase (CrGDS) and Geraniol synthase (CrGS) from *C. roseus*. CrG10H was cloned using the primers designed from the reported gene¹⁰, which encodes geraniol hydroxylase in *C. roseus* as the unigene 742 had similar sequence to that of the reported one. Their concerted reaction was also carried out.

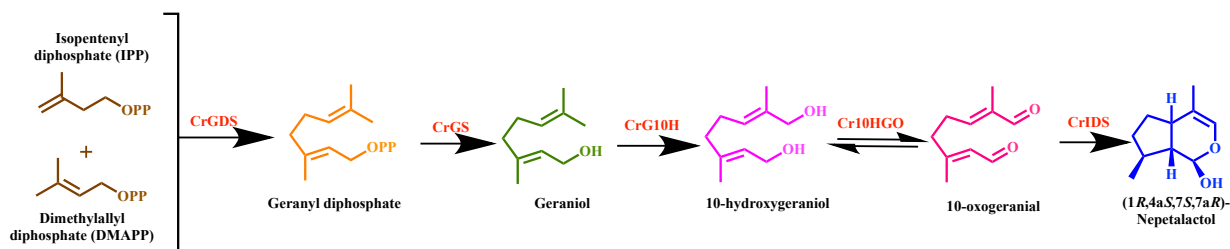


Figure 2: Concerted reaction of CrGDS, CrGS, CrG10H, Cr10HGO and CrIDS

Chapter 4

Dissection of the activity of 10-hydroxygeraniol dehydrogenase (Cr10HGO) by site-directed mutagenesis studies

10-hydroxygeraniol dehydrogenase (Cr10HGO) is an essential enzyme in the biosynthesis of Monoterpene Indole Alkaloids (MIAs). Cr10HGO catalyzes the formation of 10-oxogeraniol via two intermediates, 10-hydroxygeraniol and 10-oxogeraniol, in the presence of NADP⁺. Homology model was built using the co-ordinates of *Populus tremuloides* Sinapyl Alcohol Dehydrogenase¹¹ as a template with Schrodinger-Maestro 1.6. The model was validated using Ramachandran plot and used for docking.

Docking studies of the substrate with the homology-based model of Cr10HGO was studied in order to identify the substrate specificity of the enzyme. Mutation of the polar residue N115 to another polar amino acid residue, H resulted in notable decrease in product formation, which can be considered as an indication that the reaction is getting terminated prematurely. Another mutation at S301, replacing polar serine with hydrophobic alanine, also caused a substantial drop in product formation, such that 3/4th of the substrate was left unchanged and the remaining 1/4th constituted of only one of the intermediates.

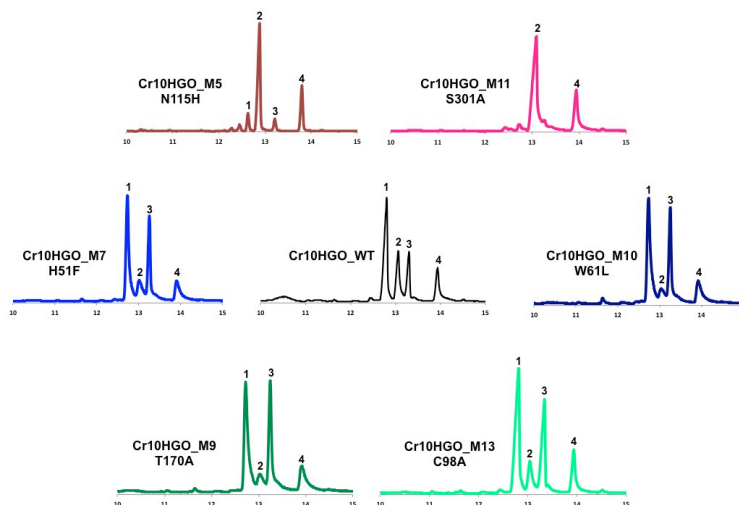


Figure 3. Total Ion Chromatograms (TICs) for reactions catalyzed by 10-hydroxygeraniol dehydrogenase wild type and mutants causing significant changes;1: 10- oxogeraniol, 2: 10-hydroxygeraniol, 3: 10-oxogeranial, 4: 10- hydroxygeranial.

The mutation W61 to Leucine resulted in lower levels of 10-hydroxygeraniol, indicating significant suppression of the reverse reaction, which would lead to the formation of substrate. This may be due to the removal of the aromatic ring of tryptophan. Mutation at the T170 position resulted in a 0.59 fold increase in the formation of 10-oxogeranial, which serves as a substrate for the next enzyme in the MIA biosynthetic pathway. These residues seem to be playing a major role in the stabilization of the carbocation intermediates. This study, thus, provides an insight into the residues that affect the catalytic activity of Cr10HGO and also provides a gateway for increasing the production of intermediates in the MIA pathway, which, in turn, would cause an increase in the biosynthesis of the final product of the pathway.

Chapter 5

Engineering of iridoid synthase (CrIDS) by site-directed mutagenesis to understand its activity profile

Iridoid synthase (CrIDS) is an NADPH-dependent enzyme, belonging to a class of monoterpene indole cyclases, which utilizes the linear 10-oxogeranial as a substrate and by a sequential reduction and cyclization, produces an equilibrium mixture of nepetalactols and iridoids, the final product in iridoid biosynthesis. Homology model was built using the coordinates of Progesterone 5 β -Reductase from *Digitalis lanata*¹² as template with

Schrodinger-Maestro 1.6. The model was validated using Ramachandran plot and used for docking.

Based on their affiliation with the substrate, 6 residues were identified to be interacting strongly with the substrates and 10 mutations were designed. The mutant I38A did not have much of an effect on the activity of the enzyme, whereas the mutation at Y156, replacing tyrosine with phenylalanine reduced the product (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol substantially. Although, this mutation only involved removal of the OH group from tyrosine, which was found to be involved in H-bonding with NADPH, it drastically reduced the activity of the enzyme. On the other hand, the mutants F175M and M211A increased the ratio of (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol in the product pool to a notable extent. Depending on these results, 4 more mutations were designed, M211F, M211S, M211K and M211H, which increased the desired stereospecific product formation by 23.3 %, 23.7 %, 21.3 % and 22.8 %, respectively.

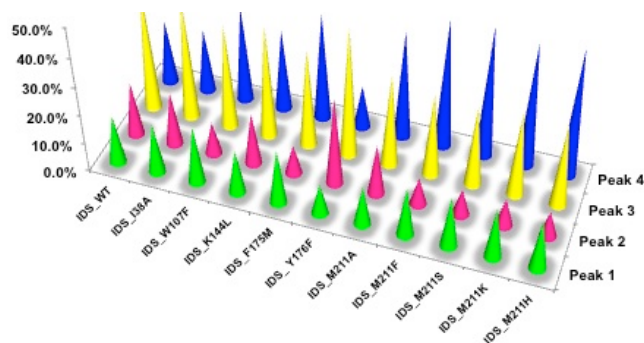


Figure 4. Product Profiles of CrIDS wild type and mutations; **Peaks 1, 2, 3:** *Cis-trans* Iridodials and *Cis-trans* Nepetalactols, **Peak 4:** (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol.

These product profiles indicated that the methionine at the 211th position is very essential for the product profile. Also, the OH group of tyrosine at the 156th position, forming a bond with NADPH is very essential for the overall activity of the enzyme. These mutants will aid in increasing the flux of Secologanin formation.

Chapter 6

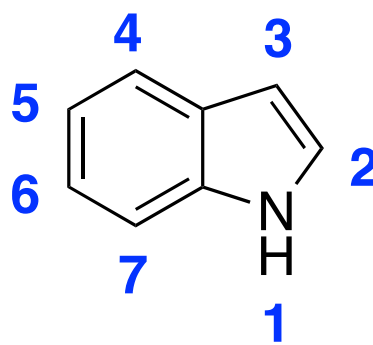
Screening and cloning of genes involved in the biosynthesis of secologanin from iridoid

Secologanin is a major intermediate in the formation of the potent anti-cancer agents, vincristine and vinblastine in *Catharanthus roseus*. Full-length unigenes showing matches with 7 deoxyloganin synthase (Cr7DLS), 7-deoxyloganetic acid glucosyltransferase

(CrDLGT), 7-deoxyloganin hydroxylase (7DLH), Loganic acid methyltransferase (CrLAMT) and Secologanin synthase (CrSLS) were identified, and their cloning in respective vectors is in progress. CrDLGT has been cloned in pET28a vector and CrLAMT in pRSETB vector. Their expressions were carried out.

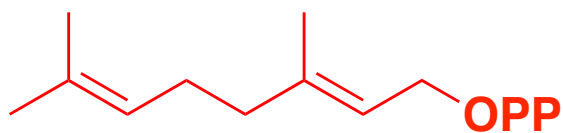
References:

1. Van der Heijden, R., Jacobs, D.I., Snoeijer, W., Hallared, D. & Verpoorte, R. The Catharanthus alkaloids: Pharmacognosy and biotechnology. *Curr. Med. Chem.* **11**, 607-628 (2004).
 2. Beckett, J.S., Beckett, J.D. & Hofferberth, J.E. A Divergent Approach to the Diastereoselective Synthesis of Several Ant-Associated Iridoids. *Org. Lett.* **12**, 1408-1411 (2010).
 3. Blatt, M.R. & Armstrong, F. K⁺ channels of stomatal guard cells: Abscisic-acid-evoked control of the outward rectifier mediated by cytoplasmic pH. *Planta* **191**, 330-341 (1993).
 4. Wyatt, T.D. Bombardier beetles and fever trees: A close-up look at chemical warfare and signals in animals and plants - Agosta, W. *Nature* **381**, 289-289 (1996).
 5. Monteiro, J.L.F. & Veloso, C.O. Catalytic conversion of terpenes into fine chemicals. *Top. Catal.* **27**, 169-180 (2004).
 6. Gajalakshmi, S., Vijayalakshmi & Devi, R.V. Pharmacological activities of *Catharanthus roseus*: a perspective review. *Int. J. Pharm. Biol. Sci.* **4**, 431-439 (2013).
 7. Feng, M., Zhao, M., Wang, Y., Xu, S., Wang, M. and Zhao, C. H-1-NMR metabonomics study of the therapeutic mechanism of total alkaloids and ajmalicine from *Rauvolfia verticillata* in spontaneously hypertensive rats. *Anal. Methods* **6**, 6473-6482.
 8. Kim, N. C. & Al., D.K. *Studies in Natural Products Chemistry: Bioactive Natural Products, Part H. 1 edn Vol. H* (ed. Rahman, A.U.) (Elsevier, 2002).
 9. Verpoorte, R. Pharmacognosy in the new millennium: Lead finding and biotechnology. *J. Pharm. Pharmacol.* **52**, 253-262 (2000).
 10. Collu, G., Unver, N., Peltenburg-Looman, A.M., van der Heijden, R., Verpoorte, R. & Memelink, J. Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Lett.* **508**, 215-220 (2001).
 11. Bomati, E.K. & Noel, J.P. Structural and kinetic basis for substrate selectivity in *Populus tremuloides* sinapyl alcohol dehydrogenase. *Plant Cell* **17**, 1598-1611 (2005).
 12. Thorn, A., Egerer-Sieber, C., Jäger, C.M., Herl, V., Müller-Uri, F., Kreis, W. & Müller, Y.A. The crystal structure of progesterone 5 beta-reductase from *Digitalis lanata* defines a novel class of short chain dehydrogenases/reductases. *J. Biol. Chem.* **283**, 17260-17269 (2008).
-



Chapter 1

Introduction



Chapter 1

Introduction

A wide array of secondary metabolites is produced by plants as a part of their defense system against herbivores, pests and pathogens. Terpenes are a large class of secondary metabolites in plants, which are known to possess numerous biological properties such as flavor and fragrance industries, in medicine, aromatherapy, as insecticides and pesticides, as antibacterial and antifungal agents. Terpenoids, also known as isoprenoids are modified terpenes, in which methyl groups have been moved or removed, or oxygen atoms added and also possess various insecticidal properties. They are the primary constituents of the essential oils of many plants and have strong aromatic properties. Owing to these properties, they are widely used in the fragrance industry, as flavor additives in food, in medicines, etc. Terpene Indole Alkaloids (TIAs) are a class of secondary metabolites produced in plants, which are known to possess varied biological properties. *Catharanthus roseus*, an evergreen shrub, is a rich source of these (TIAs), which are known for their pharmacological and therapeutic importance. For instance, serpentine and ajmalicine are employed for their relaxing properties, in the treatment of hypertension; 3,4-anhydrovinblastine, for its anti-tumor activity; and vincristine and vinblastine as anti-cancer drugs, after they were initially used as anti-diabetic agents. These terpenes are produced in minute quantities in plants and take many years for their biosynthesis. Hence, in an attempt to produce these therapeutic compounds in sufficient amounts in bacterial or yeast hosts poses a great challenge in the field of metabolic engineering.

1.1 Plant biosynthetic pathways

Plants have evolved specialized secondary biosynthetic pathways for the synthesis of structurally and functionally complex small molecules, which aid in the growth and development of the plants but are not required for the survival of the plants. In plants, the pattern of secondary metabolites is highly complex and well coordinated as it changes in a tissue- and/or organ-specific manner. Secondary metabolites play a key role in the protection of plants against microbial infections (fungal, bacterial), viral infections, herbivory, UV radiation, attraction of pollinators and frugivores, allelopathy and signalling¹. The abundance of secondary metabolites in plants is often less than 1 % of the total storage in a dedicated tissue. The unique ability to synthesize secondary metabolites is the result of a highly complex and sophisticated metabolic process, which has been selected throughout the course of evolution in different plant lineages. Plant genomes are estimated to contain 20,000-60,000 genes out of which 15-25 % of genes encode for the enzymes involved in secondary metabolism^{2,3}. Since ancient times, secondary metabolites have been used for several purposes, such as flavors, fragrances, dyes, stimulants, insecticides, hallucinogens as well as therapeutic agents⁴. Secondary metabolites have been classified into three major classes such as phenolic compounds, isoprenoids, and alkaloids depending on their structure and biosynthetic pathways, which are derived from primary metabolic pathways. These metabolites are unique to plants and are essential for adaptation to diverse and inconstant surroundings. Phenolic compounds are highly diverse and include phenylpropanoids, flavonoids, etc. which are derived from phenylalanine and acetyl-CoA and play important roles in defense, signaling, pathogenesis and symbiosis⁵. They are also known for providing pigments to flowers, fruits, leaves and seeds. Isoprenoids represent the largest class of secondary metabolites derived from two basic five carbon units, Isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). They play diverse functional roles in plants as hormones, photosynthetic pigments, electron carriers, structural components of membrane, as well as in communication and defense⁶. Alkaloids are nitrogen containing compounds found in about 20 % of plant species and play an important role in the defense of plants against herbivores and pathogens⁷. The genes responsible for generating great diversity of secondary metabolites arise from gene duplication, origin of new genes for secondary metabolism, gain and loss of function for specific

compounds, etc. The biosynthesis and accumulation of plant secondary metabolites is usually restricted to the specific organs/tissues at particular developmental stages. Owing to their wide industrial and pharmaceutical applications, the secondary metabolites can be produced in genetically tractable heterologous systems using synthetic biology tools, after the individual enzymes involved in the biosynthesis of the natural products have been characterized.

Metabolic engineering essentially deals with the genetic manipulation of microorganisms in order to increase productivity of a specific metabolite, like mutations, over-expressions, modifications in copy number of particular genes, etc. It is also applied for decreasing the production of some by-products, toxic substances, etc. In the last few years, metabolic engineering has been applied to various pathways for the large-scale production of metabolites of medical and commercial significance. Certain biosynthetic pathways in *Artemisia annua* and *Taxus brevifolia* have been subject to metabolic engineering to aid in semi-synthesis of commercially significant end-products. *Catharanthus roseus* is an excellent candidate for metabolic engineering due to the vast array of pharmaceutically significant metabolites that are produced in the plant.

1.2 *Catharanthus roseus*

Catharanthus roseus is an evergreen shrub, belonging to the *Apocynaceae* family. It is also commonly known as Madagascar periwinkle (due to its abundance in the Madagascar area) and *Vinca rosea*. *C. roseus* has been used extensively in Ayurvedic and ancient Chinese medicine for the treatment of diabetes, lymphoma, malaria, etc. in continuation, the plant extracts of *C. roseus* have been found to have antimicrobial, antioxidant, anthelmintic, antifeedant, antisterility, antidiarrheal, antidiabetic activities⁸. The main reason for these activities has been attributed to the presence of terpene indole alkaloids (TIAs), which are a type of secondary plant metabolites. Vincristine and vinblastine, which are potent anti-cancer agents, are TIAs, primarily produced through a key intermediate, secologanin in *C. roseus* (*Vinca rosea*). The main significance of this study is that these TIAs are produced at a very low level in *C. roseus* leaves (about 0.0002 % of dry weight)⁹. Hence, a better understanding of the pathway enzymes leading to the production of these anti-tumor compounds would aid in applying strategies for their large-scale production.

1.3 Isoprenoids / Terpenoids

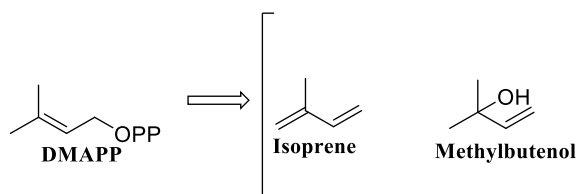
Isoprenoids, also known as terpenoids, are the most abundant and structurally diverse class of plant secondary metabolites. Over 70,000 individual structures, containing a truly incredible array of carbon skeletons and functional groups have been reported. All isoprenoid compounds are constructed from two simple five-carbon building blocks, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) which in turn, is synthesized through mevalonate (MVA) or methylerythritol phosphate (MEP) pathways. The MVA and MEP pathways converge at DMAPP. The reactions in isoprenoid biosynthesis and the structures of the enzymes in the pathway beyond DMAPP are similar in all organisms. After DMAPP, the pathway branches to give the multitude of metabolites found in nature. The basic building reactions in the isoprenoid biosynthetic pathway can be conveniently divided into four steps: the first is biosynthesis of the basic building blocks, isopentenyl diphosphate (IPP) and its isomer, dimethylallyl diphosphate (DMAPP), by the mevalonate (MVA) and non-mevalonate (MEP) pathways; the second step is the chain elongation by head to tail condensation of the basic C5 isoprene unit, to form chain elongated products such as geranyl diphosphate (GPP), farnesyl diphosphate (FPP) and geranylgeranyl diphosphate (GGPP)^{10,11}; the third step is where terpene synthases act upon respective prenyldiphosphates to produce a diverse array of terpene skeletons by cyclization or rearrangement and in the final step, modification on the parental backbone by tailored enzymes to form the final terpenoids or isoprenoids. Terpenes and their derivatives form the major constituents of the essential oils of various plants, and are known to have numerous applications in flavor and fragrance industries¹², in medicine^{13,14}, aromatherapy, as insecticides and pesticides¹⁵, as antibacterial and antifungal agents¹⁶, to name a few.

1.4 Classification of terpenes / isoprenes

The basic molecular formulae of terpenes are multiples of five carbon isoprene units (C₅H₈)_n, where n is the number of linked isoprene units (IPP or DMAPP). This is called the isoprene rule of C5. The isoprene units may be linked together in a head to tail fashion to form linear chains or they may be arranged to form rings, as well. Depending upon the number of isoprene units linked, terpenes are classified as hemiterpenes, monoterpenes, sesquiterpenes, diterpenes, sesteteterpenes, triterpenes, tetraterpenes and polyterpenes.

1.4.1 Hemiterpenes (C5)

Hemiterpenes are formed from a single isoprene unit, having five carbon atoms. Isoprenes are themselves considered the only hemiterpenes. Their oxygenated derivatives are termed as hemiterpenoids, for example, tiglig, angelic and isovaleric acids. Although several plant species such as mosses, fern, and trees emit isoprenes in large amounts, their occurrence is termed rare. Isoprenes are synthesized by chloroplastic enzymes directly by diphosphate elimination from DMAPP¹⁷⁻¹⁹ (Scheme 1.1).

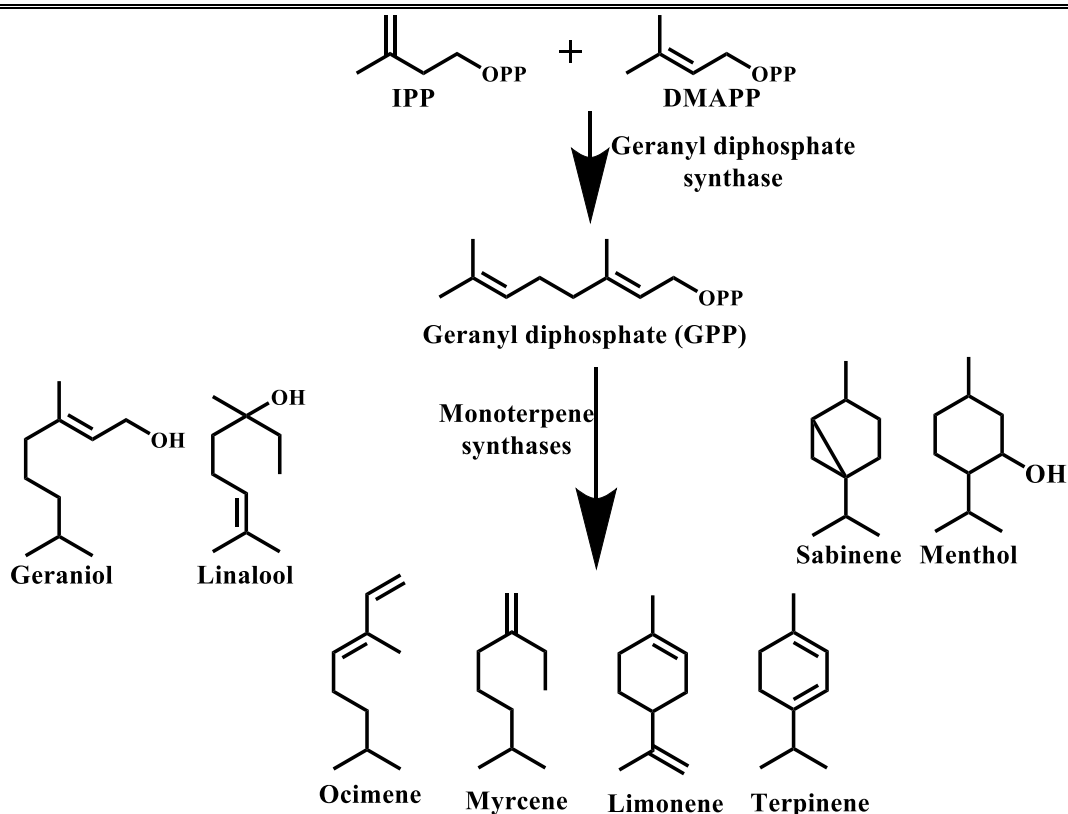


Scheme 1.1: Biosynthesis of hemiterpenes

Most hemiterpenes occur in oils and are hence, mostly, water insoluble. Two new hemiterpenes, namely utililactone and epiutililactone have been recently isolated from the plant *Prinsepia utilis*²⁰, which is known to have antioxidative activity. Also, two new hemiterpene glucosides named pubescenosides A and B have been isolated from the root of *Ilex pubescens*²¹. These were found to possess potent anti-platelet aggregation activity.

1.4.2 Monoterpenes (C10)

Monoterpenes derived by the combination of two isoprene units (one unit each of IPP and DMAPP) with the molecular formula C₁₀H₁₆, are extensively distributed in secretory tissues such as oil glands or chambers of higher plants, insects, fungi and marine organisms. Monoterpenes are the major components of plant essential oils and are highly volatile. This volatile nature is lost when these molecules are in their glycosidic forms. Monoterpenes are biosynthesized from geranyl diphosphate (GPP) catalyzed by monoterpene synthases (Scheme 1.2) and divided into three subgroups: acyclic (citronellol, geraniol, linalool), monocyclic (menthol, limonene) and bicyclic (thujane, fenchane, carane). Due to their fragrance, they are major ingredients in flavor and fragrance and cosmetic industries. For example, menthol has been used in toothpastes, mouth-fresheners, etc.

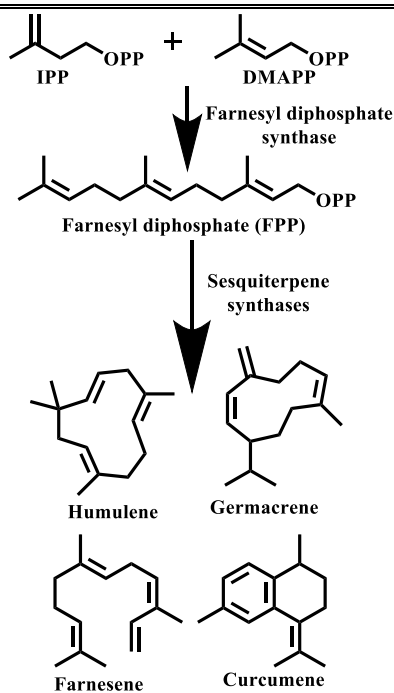


Scheme 1.2: Biosynthesis of monoterpenes

In addition to the flavor and fragrance properties, monoterpenes also possess various pharmacological activities including sedative, antitumor, cytotoxic, anti-inflammatory, insecticidal, antibacterial, antifungal and antioxidant²²⁻²⁴. For instance, linalool has been found to be able to moderately inhibit cell proliferation²⁵. In plants, monoterpenes are synthesized in plastids through the MEP pathway, whereas in other higher organisms and in yeast, they are synthesized through the MVA pathway²⁶.

1.4.3 Sesquiterpenes (C₁₅)

Sesquiterpenes are a diverse group of isoprenoids, derived by the combination of three isoprene units (2 units of IPP and 1 unit of DMAPP) with the molecular formula C₁₅H₂₄ (Scheme 1.3). This abundant group of isoprenoids consists of over 7300 molecules with over three hundred stereo-chemically distinct hydrocarbon skeletons. Sesquiterpenes are derived from the linear substrate, farnesyl diphosphate (C₁₅).



Scheme 1.3: Biosynthesis of sesquiterpenes

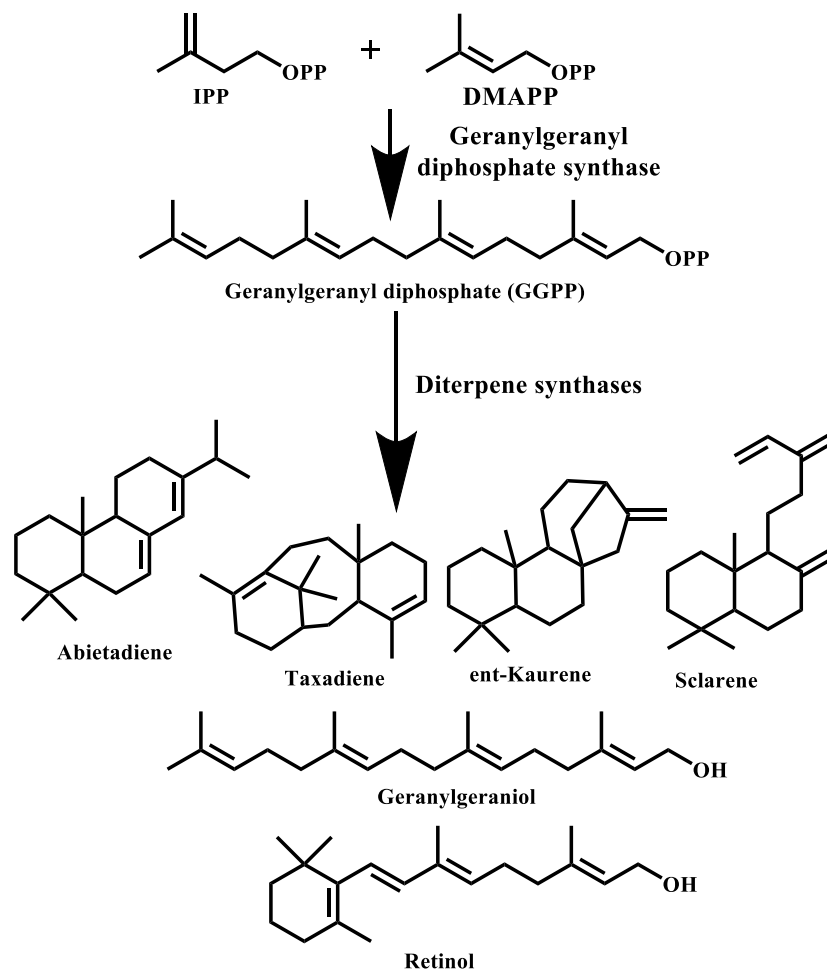
They represent one of the most important components of plant essential oils along with monoterpenes. The large carbon skeleton of Farnesyl diphosphate (FPP) and the presence of three double bonds greatly increase the structural diversity of the resulting products catalyzed by sesquiterpene synthases. Sesquiterpenes can be divided into subgroups depending on the number of cyclic rings as: acyclic (farnesene, farnesol), monocyclic (zingiberene, humulene), bicyclic (caryophyllene, vetivazulene) and tricyclic (copaene, longifolene)²⁷.

Sesquiterpenes have been used as flavors and fragrances but they also have the potential to serve as anti-cancer²⁸ and anti-malarial²⁹ agents. In recent years, sesquiterpenes have been recognized as replacements for petroleum-derived fuels³⁰. Modifications of sesquiterpenes, such as oxidation, produce sesquiterpenoids.

1.4.4 Diterpenes (C₂₀)

Diterpenes are composed of four isoprene units with the molecular formula C₂₀H₃₂. The diterpene compounds are derived from geranylgeranyl diphosphate (GGPP) and divided into several subgroups depending on the basis of cyclization such as: linear (geranylgeraniol, phytane), monocyclic (cembrene A), bicyclic (sclarene, labdane), tricyclic (abietane, taxadiene), tetracyclic (stemarene, stemodene, gibberellins) and macrocyclic (daphnanes, tigllianes) (Scheme 1.4). In addition to

these, taxanes are a class of diterpenoids featuring a taxadiene core and are produced by the plants of the genus *Taxus*. They are potent chemotherapy agents³¹.



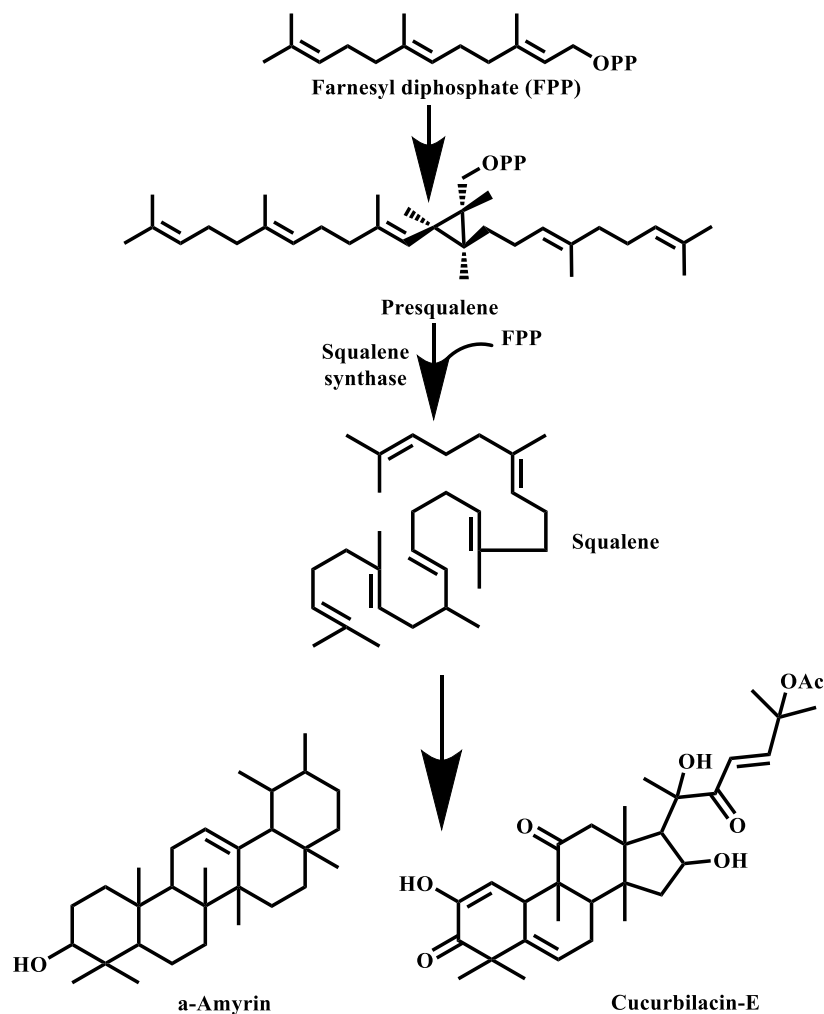
Scheme 1.4: Biosynthesis of diterpenes

Diterpenes have magnificent biological and pharmacological activities such as anti-bacterial, anti-fungal, anti-inflammatory and anti-leishmanial activity (for example, retinol, retinal and phytol).

1.4.5 Triterpenes (C₃₀)

Triterpenes are composed of six isoprene units and are derived from squalene, which is synthesized by head to head (1-1') condensation of two units of farnesyl diphosphate catalyzed by squalene synthase (Scheme 1.5). It has the molecular formula C₃₀H₄₈. Triterpenes are synthesized in plants, animals and fungi and exhibit a wide range of structural diversity³². Till date, at least 4000 triterpenes are known with great structural diversity. Sterols are precursors for a vast array of compounds (steroid hormone and bile acids in mammals, ecdysteroids, antheridiol and oogoniol in

fungi^{33,34}), involved in important cellular processes in animals whereas plant sterols are linked to brassinosteroids synthesis^{35,36}. Brassinosteroids are polyhydroxy plant steroids essential for normal growth and development and present in all parts of plants³⁷.



Scheme 1.5: Biosynthesis of triterpenes

1.4.6 Tetraterpenes (C₄₀)

Tetraterpenoids, which consist of eight isoprene units, having the molecular formula C₄₀H₅₆ instead of C₄₀H₆₄, and are found in all plants, bacteria and fungi. Carotenes and carotenoids, like lycopene and xanthophylls, are the biologically important tetraterpenoids^{38,39}. The first committed step in carotenoid biosynthesis is condensation of two molecules of geranylgeranyl diphosphate (GGPP) to form a 40C linear phytoene, with a central 15-15'-*cis*- double bond, catalyzed by phytoene synthase. Further 15-*cis*-phytoene undergoes several rearrangements including desaturation and isomerization of double bonds, via neurosporene and ζ-carotene as

well established that diversity in carotenoids structure and specific carotenoid accumulation in plant, highly depends on the geometric isomeric states of substrates utilized by carotene desaturases.

Some carotenoids such as β -carotene, lycopene and astaxanthin are commercially available and widely used as food colorants; food supplements, in cosmetics, etc. Carotenoids are known to possess antioxidant properties and an ability to assuage chronic diseases⁴¹.

1.4.7 Polyterpenes

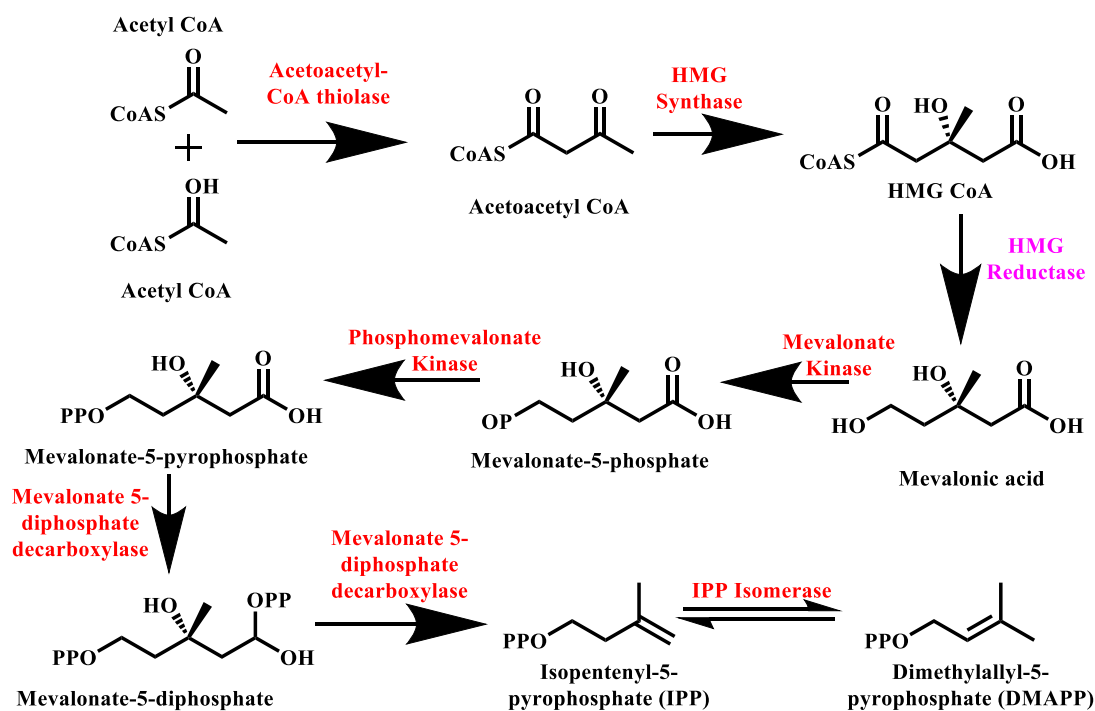
Polyterpenes are formed from many isoprene units joined head to tail, with the molecular formula $(C_5H_8)_n$. Polyprenyl diphosphate is converted to dolichol by dephosphorylation and saturation of α -isoprene unit⁴²⁻⁴⁴, which serve as glycosyl carriers in protein glycosylation. Rubber is the most well-known polyisoprenoid produced by more than 2000 plant species^{45,46} and is used in the preparation of tyres, surgical gloves, balloons, conveyer belts various other industrial applications. The source for natural rubber is *Hevea brasiliensis*⁴⁶, and it occurs on the surface of rubber particles suspended in latex⁴⁷. Rubber consists of polymers of isoprenes. FPP can bind to rubber transferase and elongate a new rubber polymer.

1.5 Biosynthesis of Isoprenoids / terpenoids

Isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) are two important units in the formation of terpenoid molecules. These two molecules can be synthesized by the mevalonate (MVA) pathway or the non-mevalonate/ 1-deoxy-D-xylulose-5-phosphate (DXP)/ methylerythritol phosphate (MEP) pathway. Non-mevalonate pathway occurs in plant plastids⁴⁸ as well as many bacteria such as *Escherichia coli*, *Mycobacterium tuberculosis*, *Neisseria meningitides*, etc. synthesize IPP and DMAPP through the non-mevalonate pathway⁴⁹.

The Mevalonate pathway (MVA) or HMG-CoA reductase pathway was first discovered in yeasts and animals in 1950, and it was thought to be responsible for the biosynthesis of IPP and DMAPP in all the organisms in six enzymatic steps starting from Acetyl-CoA^{50,51}, followed by IPP isomerase (IDI) which maintains the balance between IPP and DMAPP. This pathway occurs in cytosol and is present in animals, plants, fungi, archaea and some bacteria. The MVA pathway mainly leads to the formation of triterpenes and sesquiterpenes^{52,53} (Figure 1.1). Acetoacetyl-CoA

thiolase (EC 2.3.1.9; AACT; JF739870) catalyzes the condensation of two molecules of acetyl-CoA to acetoacetyl-CoA, the first step of the MVA pathway (Scheme 1.7). 3-hydroxy-3-methylglutaryl-CoA synthase (EC 4.1.3.5; HMGS), which catalyzes the condensation of acetoacetyl-CoA with acetyl-CoA to yield 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA), is the second enzyme in the pathway. 3-Hydroxy-3-methylglutaryl-CoA reductase (EC 1.1.1.34; HMGR) catalyzes the reduction of HMG-CoA to mevalonic acid. Mevalonate kinase (MVK) and mevalonate 5-phosphate (MVAP) kinase (PMK) catalyze the phosphorylation of mevalonic acid to form mevalonate 5-phosphate⁵⁴. Mevalonate kinase (MVK) and mevalonate 5-phosphate (MVAP) kinase (PMK) catalyze the phosphorylation of mevalonate to form mevalonate 5-diphosphate⁵⁵⁻⁵⁷. Finally, mevalonate 5-diphosphate is converted to IPP by the ATP-dependent mevalonate 5-diphosphate decarboxylase (MVD). Then IPP is isomerized to DMAPP by the catalysis of IPP isomerase (IDI)⁵⁸. Table 1.1 summarizes current knowledge regarding the molecular and functional characterization of the MVA pathway enzymes throughout the plant kingdom in various organisms.



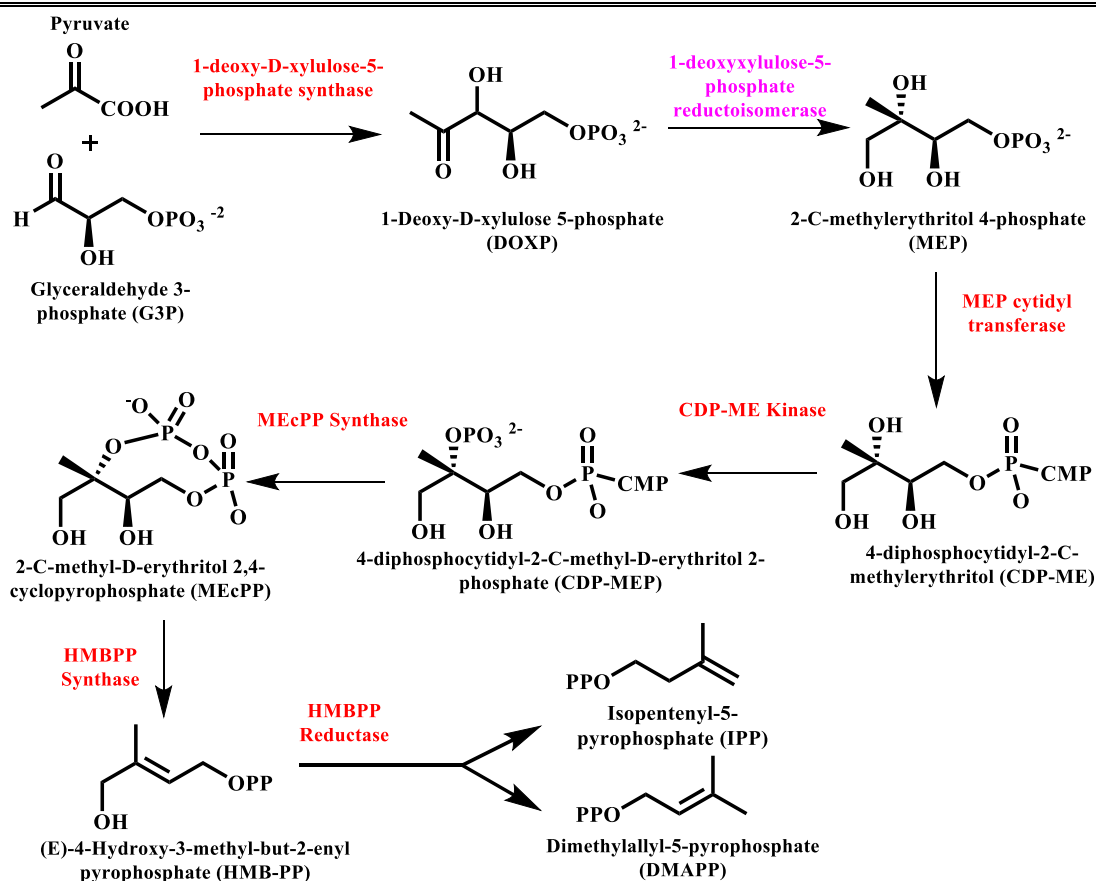
Scheme 1.7: Biosynthesis of IPP and DMAPP by the Mevalonate Pathway (MVA pathway).

Table 1.1: Current knowledge of characterized plant MVA pathway genes

Enzyme	Plant	Year(s)	Work done
Acetoacetyl-CoA thiolase (AACT1, AACT2)	<i>A. thaliana</i> ^{59,60}	2007, 2008	Histochemical, functional and GFP-fusion analysis
	<i>C. roseus</i> ⁶¹	1994	Purification and determination of enzyme activity
	<i>H. annuus</i> ^{62,63}	2002, 2009	Enzyme purification & activity
	<i>R. sativus</i> ⁶⁴	1996	cDNA characterization
Hydroxymethylglutaryl-CoA synthase (HMGS)	<i>A. thaliana</i> ^{65,66}	1995, 2010	cDNA isolation/ characterization
	<i>B. juncea</i> ⁶⁷⁻⁶⁹	2000, 2004, 2006	Kinetic analysis of WT & active-site mutants, crystal structure
	<i>C. roseus</i> ⁶¹	1994	Purification & determination of enzyme activity
	<i>H. brasiliensis</i> ⁷⁰ s	2004	Expression and activity
	<i>P. sylvestris</i> ⁷¹	1997	cDNA characterization in response to ozone treatment
	<i>Taxus</i> ⁷²	2006	Plant homolog expression in needles & stem
Hydroxymethylglutaryl-CoA reductase (HMGR)	<i>A. thaliana</i> ⁷³⁻⁸⁷	2005, 2009, 2004, 2006, 2007, 1995, 1997, 1998, 1994, 1996, 2003	Isozyme identification & characterization, subcellular localization, regulation of gene expression & enzyme activity
	<i>B. oleracea</i> ⁸⁸	1994	Enzyme activity assays
	<i>B. orellana</i> ⁸⁹	2001	Gene family identification, gene expression & enzyme activity
	<i>C. arabica</i> ⁹⁰	1997	Characterization of gene expression
	<i>C. melo L.</i> ⁹¹	2001	cDNA isolation/ characterization, heterologous expression for functional analysis
	<i>D. minor</i> ^{92,93}	2007	Characterization of transgenics
	<i>H. brasiliensis</i> ⁹³	1991	cDNA isolation/ characterization of two isoforms
	<i>H. vulgare</i> ⁹⁴	1996	Kinetic assay with partially purified enzyme
	<i>L. latifolia</i> ⁹⁵	2007	Characterization of transgenics
	<i>M. domestica</i> ⁹⁶	2002	cDNA isolation/ characterization
	<i>O. sativa L.</i> ⁹⁷	2001	Sequence & transcript analysis of HMG2
<i>P. sativum</i> ^{98,99}	1975	Activity assay, subcellular localization study	

	<i>R. sativus</i> ¹⁰⁰	1986	Enzyme purification, kinetic analysis
	<i>S. lycopersicum</i> ¹⁰¹	1989	cDNA characterization
	<i>S. tuberosum</i> ¹⁰²	1985	Effects of wounding and fungal infection on enzyme activity
	<i>T. aestivum</i> ¹⁰³	1993	cDNA characterization
Mevalonate Kinase (MK)	<i>A. americana</i> ^{104,105}	1972, 1973	MVA biosynthesis of cell-free extracts
	<i>A. thaliana</i> ^{106,107}	1994, 2000	cDNA characterization
	<i>C. roseus</i> ^{55,56}	1999, 2000	Enzyme kinetics
	<i>H. brasiliensis</i> ¹⁰⁸	1965	Reaction characteristics
	<i>P. strobus</i> ¹⁰⁹	2006	Enzyme activity assays
	<i>P. pinaster</i> ¹¹⁰	1974	Enzyme kinetics
	<i>P. vulgaris</i> ¹¹¹	1965	Enzyme activity assays, MK inhibition by prenyl diphosphates
Phosphomevalonate Kinase (PMK)	<i>C. roseus</i> ^{56,112}	1999	Enzyme kinetics
	<i>H. brasiliensis</i> ¹¹³	1966	Enzyme kinetics
	<i>S. tuberosum</i> ¹¹⁴	1996	Effects of wounding and pathogen-elicitation on enzyme activity
Diphosphomevalonate decarboxylase (MPDC1, MPDC2)	<i>A. thaliana</i> ¹¹⁵	1999	Complementation assay of an enzyme-deficient yeast strain
	<i>H. brasiliensis</i> ¹¹⁶	1968	Enzyme kinetics

Earlier it was believed that synthesis of terpenoids occur through MVA pathway, however, in the early 1990s Rohmer and co-workers discovered an alternative pathway for the synthesis of terpenoids when ¹³C-labelled studies of isoprenoid biosynthesis in bacteria differed from those expected from the classical isoprenoid biosynthetic route (MVA pathway). This newly discovered pathway is present in many eubacteria and plant plastids, but absent in humans, therefore, could be an excellent target for developing broad-spectrum antibiotics.



Scheme 1.8: Biosynthesis of IPP and DMAPP by Methylerythritol pathway (MEP pathway)/ 1-deoxy-D-xylulose 5-phosphate pathway (DXP pathway).

The MEP or DXP pathway is found to be highly responsible for the formation of monoterpenes, diterpenes, tetraterpenes and polyprenols. In the MEP pathway (Scheme 1.8), the first step involves a condensation of pyruvate with D-glyceraldehyde-3-phosphate to yield 1-deoxy-D-xylulose-5-phosphate (DXP), catalyzed by the enzyme 1-deoxy-D-xylulose-5-phosphate synthase (DXS)¹¹⁷. 1-Deoxyxylulose-5-phosphate reductoisomerase (DXR) catalyzes the rearrangement and subsequent reduction of DXP to 2-C-methylerythritol-4-phosphate (MEP)¹¹⁸. MEP is conjugated with 4-phosphocytidyl (CDP) by MEP cytidyltransferase (MCT) to form 4-(cytidine 5'-diphospho)-2-C-methylerythritol¹¹⁹.

4-(Cytidine 5'-diphospho)-2-C-methylerythritol kinase (CMK), which catalyzes the phosphorylation of 4-(cytidine 5'-diphospho)-2-C-methylerythritol¹²⁰ is, like MK and PMK of the MVA pathway, a member of the GHMP family of metabolite kinases. 4-(Cytidine 5'-diphospho)-2-C-methylerythritol 2-phosphate, the product of the reaction catalyzed by CMK, is then converted to 2-C-methylerythritol 2,4-cyclodiphosphate (MEcPP) by the action of 2-C-methylerythritol 2,4-

cyclodiphosphate synthase (MECPS)¹²¹. MEcPP is then converted to (E)-4-hydroxy-3-methyl-but-2-enyl diphosphate (HMBPP) by the HMBPP synthase enzyme. HMBPP is finally reduced by HMBPP reductase to form isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). IPP is the basic carbon building block of long chain prenyl phosphates. It has been reported that the IPP and DMAPP required for Secologanin synthesis in *C. roseus* cell culture is derived from the DXP Pathway¹²².

Table 1.2: Current knowledge of characterized plant MEP pathway genes

Enzyme	Plant	Year(s)	Work done
DOXP synthase (DXS)	<i>A. thaliana</i> ^{123,124}	2000	Putative plastidic localization
	<i>C. annuum</i> ¹²⁵	1998	Gene isolation & cloning
	<i>C. roseus</i> ¹¹⁷	2000	Gene isolation & cloning, Positive correlation between DXS transcript and terpene indole alkaloid production
	<i>L. esculentum</i> ¹²⁶	2000	Positive correlation between DXS transcript and lycopene production
	<i>M. piperita</i> ¹²⁷	1998	Gene isolation & cloning, Positive correlation between DXS transcript and carotenoid production
	<i>M. truncatula</i> ¹²⁸	2002	Gene isolation & heterologous expression
DOXP reductoisomerase (DXR)	<i>A. thaliana</i> ^{129,130}	2002, 1999	Gene isolation & heterologous expression, presence of an N-terminal transit peptide located to plastids
	<i>C. roseus</i> ¹³¹	2000	Positive correlation between DXR transcript and indole alkaloid production in suspension cultures
	<i>M. piperita</i> ^{52,132}	1999, 2001	Gene isolation & heterologous expression, Positive correlation between DXR overexpression and essential oil production in transgenics, fosmidomycin as an inhibitor
CDP-ME synthase	<i>A. thaliana</i> ¹³³	2000	Gene isolation & heterologous expression
	<i>C. annuum</i> ¹²⁰	2000	Radiolabelled CDP-ME is incorporated efficiently into carotenoids
CDP-ME kinase	<i>C. annuum</i> ¹³⁴	2000	C14-labelled CDP-ME was efficiently converted into

			carotenoids by isolated Chromoplast
	<i>L. esculentum</i> ¹³⁴	2000	Gene isolation & heterologous expression
	<i>M. piperita</i> ¹³⁵	1999	Partially purified enzyme assays
MECP synthase	<i>C. annuum</i> ^{121,136}	2000, 2001	Radiolabelled studies
	<i>C. roseus</i> ¹³¹	2000	cDNA isolation/ characterization
	<i>N. pseudonarcissus</i> ¹³⁷	1999	Cell free conversion of CDP-MEP into MECP
HMBPP synthase	<i>A. thaliana</i> ¹³⁸	2001	Gene identification by homology studies
HMBPP reductase	<i>C. annuum</i> ¹³⁹	2002	Catalytic activity from isolated chromoplasts of red pepper enhanced on addition of purified HMBPPR

The MVA pathway occurs in animals, plants, fungi and archea, whereas the MEP pathway is found to occur in bacteria, green algae and plants. In the case of plants, the MVA pathway occurs in the cytosol and the MEP pathway in the plastid (Figure 1.1).

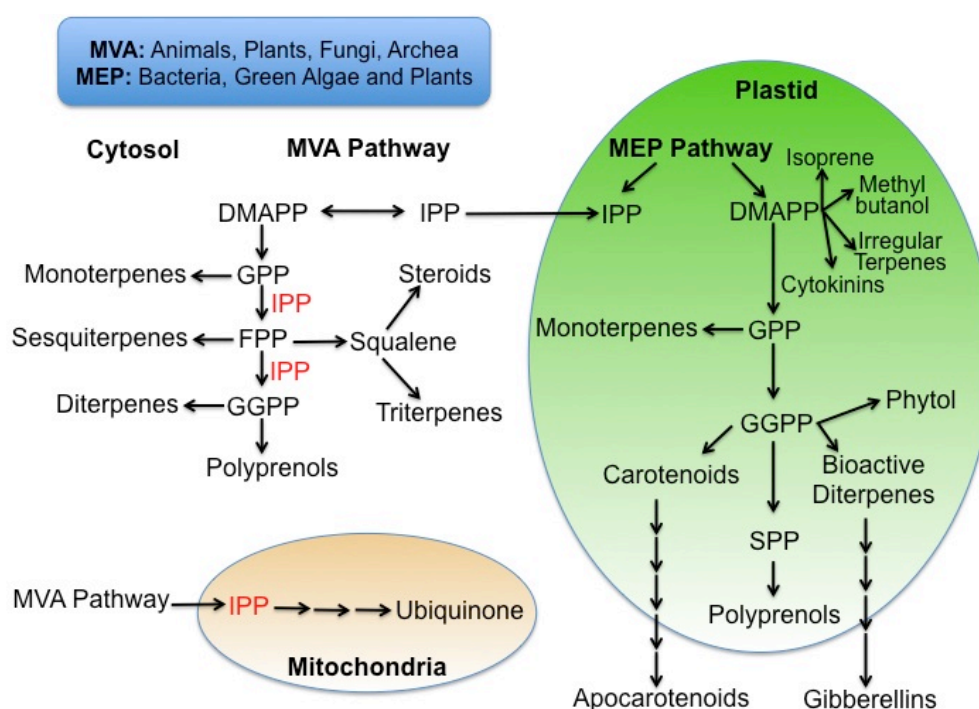


Figure 1.1: Compartmentalization of the MVA and MEP Pathways.

IPP and DMAPP are considered as the building blocks for the formation of isoprenoids across all living systems. The carbon skeletons of isoprenoid compounds

are constructed from 3-methyl-1-butyl units, which are joined in one of nine different patterns (Figure 1.2).

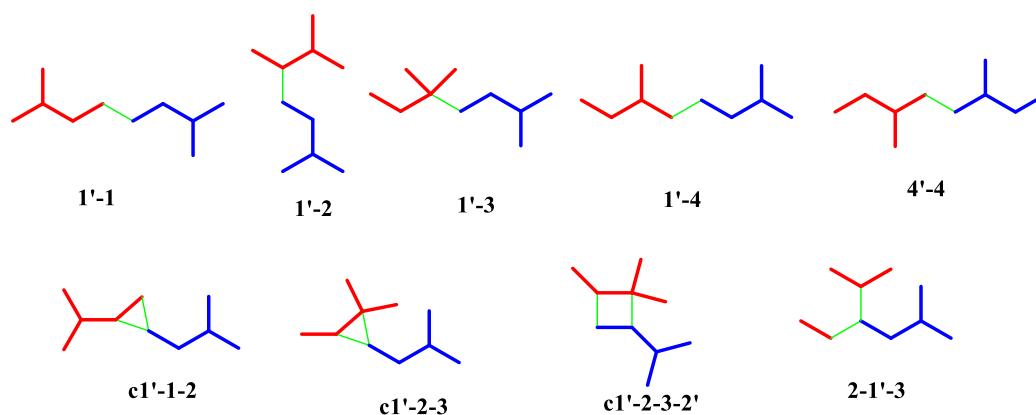


Figure 1.2: Isoprenoid carbon skeletons; 1'-carbon of one of the units (red) is joined to one of the four butyl carbons the other unit (black).

IPP and DMAPP are the elementary units of isoprenoid biosynthesis, which is found to occur by similar types of reactions in all living systems. DMAPP leads to the formation of hemiterpenes, whereas, IPP and DMAPP condense to form Geranyl diphosphate (GPP), which in turn, forms monoterpenes. GPP is further converted to farnesyl diphosphate (FPP) by the addition of an IPP molecule, which serves as the precursor for the biosynthesis of a number of classes of secondary metabolites (Figure 1.3) such as carotenoids, withanolides, ubiquinones, dolichols, sterols among others. The enzyme FPP synthase (FDS) is a chain elongation enzyme, which carries out head to tail condensation of two molecules of IPP with DMAPP to form FPP. FPP is directly involved in the biosynthesis of Heme A, Sesquiterpenes, farnesylated proteins, ubiquinones, while forms dolichol through dehydrololichol diphosphate and sterols and triterpenoids through squalene. FPP is further condensed to geranylgeranyl diphosphate (GGPP) by additions of another IPP molecule, which is the precursor for carotenoids, gibberellins, diterpenes, polyterpenes, etc. Rational approaches for exploitation of isoprenoid metabolism in drug development against human diseases depend on understanding how the pathway functions and how enzymes in the pathway catalyze reactions. Isoprenoid enzymes are not abundant in their host organisms and have only become readily available for study as the result of recombinant DNA studies. All organisms have an absolute requirement for isoprenoid compounds, and differences in the pathways and enzymes for biosynthesis of IPP and DMAPP offer an attractive opportunity for development on new drugs. Humans

synthesize isoprenoid compounds from acetyl-CoA by the MVA pathway, while most eubacteria, both gram negative and gram positive, and the plasmodium species responsible for malaria use the MEP pathway. Thus, the biosynthetic intermediates and the biosynthetic enzymes for humans and these microorganisms are orthogonal.

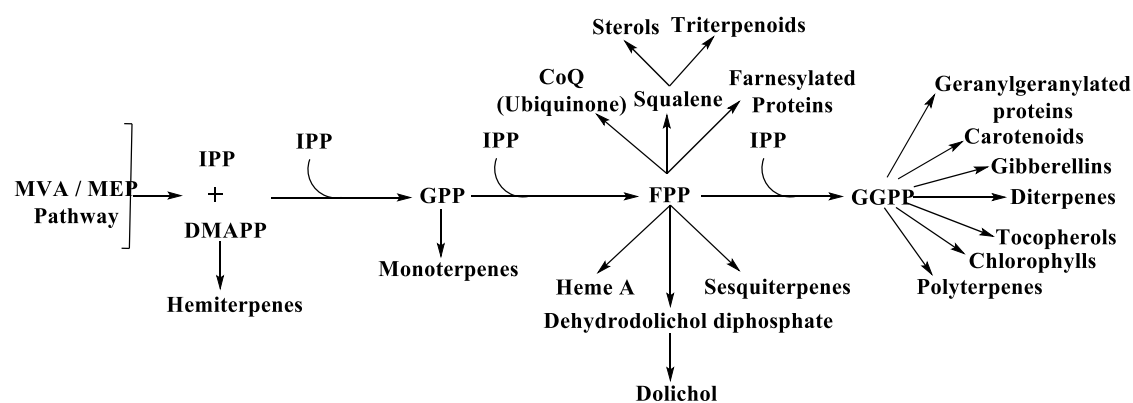


Figure 1.3: Overview of terpene biosynthesis.

1.6 Alkaloids

Alkaloids are a collection of naturally occurring nitrogen containing chemical compounds produced by a variety of organisms including bacteria, fungi, plants and animals. In addition to carbon, hydrogen and nitrogen, some alkaloids contain oxygen, sulfur, chlorine, bromine and phosphorous.

Alkaloids possess a wide array of pharmacological activities including anti-malarial (e.g. quinine)¹⁴⁰, anti-asthma (e.g. ephedrine)¹⁴¹, anti-cancer (e.g. homoharringtonine)¹⁴², cholinomimetic (e.g. galantamine)¹⁴³, vasodilatory (e.g. vincamine)¹⁴⁴, anti-arrhythmic (e.g. quinidine)¹⁴⁵, analgesic (e.g. morphine)¹⁴⁶, anti-bacterial (e.g. chelerythrine)¹⁴⁷ and anti-hyperglycemic activities (e.g. piperine)¹⁴⁸.

1.6.1 Types of Alkaloids

Based on the similarity of the carbon skeleton, alkaloids are classified into true alkaloids, protoalkaloids, polyamine alkaloids, peptide and cyclopeptide alkaloids and pseudoalkaloids.

1.6.1.1 True alkaloids

These are alkaloids containing nitrogen in the heterocyclic ring and are known to originate from amino acids¹⁴⁹. Their characteristic examples are atropine, nicotine, and morphine. This group also includes some alkaloids which in addition to nitrogen contain heterocyclic terpene (e.g., evonine¹⁴⁹) or peptide fragments (e.g.,

ergotamine¹⁴⁹). This group also includes piperidine alkaloids coniine and coniceine¹⁴⁹ although they do not originate from amino acids¹⁴⁹ (Figure 1.4).

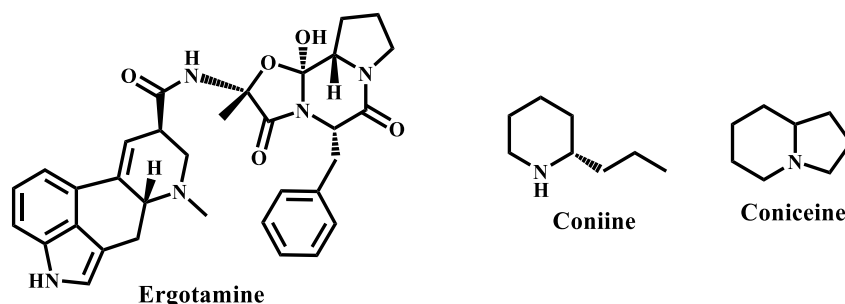


Figure 1.4: Some examples of true alkaloids from plants and fungi.

1.6.1.2 Protoalkaloids

Some examples of protoalkaloids include mescaline, adrenaline and ephedrine (Figure 1.5). These alkaloids contain nitrogen and also originate from amino acids¹⁴⁹. They are also known to possess adrenergic (stimulant) activity¹⁴⁹.

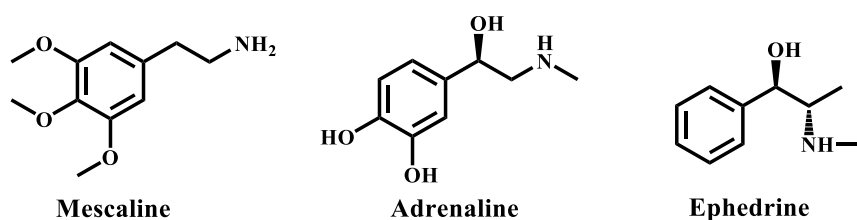


Figure 1.5: Some examples of Protoalkaloids from plants.

1.6.1.3 Polyamine alkaloids

Polyamine alkaloids are derivatives of putrescine (toxin), spermidine, and spermine (Figure 1.6). These alkaloids are found in ribosomes and living tissues, and have various metabolic functions within organisms like maintaining membrane potential and controlling intracellular pH and volume. They were primarily isolated from semen.

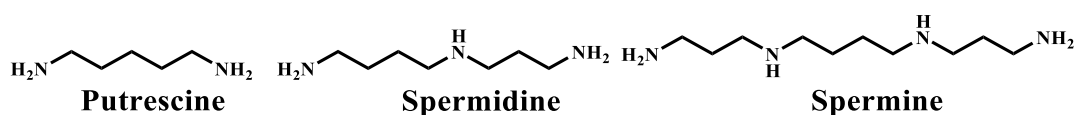


Figure 1.6: Some examples of polyamine alkaloids from plants and animals.

1.6.1.4 Peptide and cyclopeptide alkaloids

Cyclopeptide alkaloids are the compounds composed of simple, natural amino acids linked by amide bond (-CONH-) structure incorporated in them, in which a 10- or 12-membered peptide-type bridge spans the 1,3 or 1,4 positions of a benzene ring, which are used to heal ulcers and wounds¹⁵⁰. Some examples are sativanine, nummularine, rugosanine and sanjoinine (Figure 1.7).

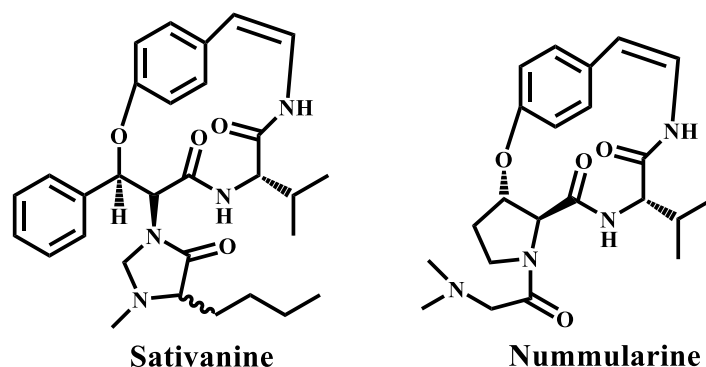


Figure 1.7: Some examples of Peptide and cyclopeptide alkaloids from plants.

1.6.1.5 Pseudoalkaloids

Compounds that do not originate from amino acids, but are alkaloid-like in nature, are classified as pseudoalkaloids¹⁴⁹. This group includes terpene-like and steroid-like alkaloids¹⁴⁹, as well as purine-like alkaloids such as caffeine, theobromine, theacrine and theophylline, used mainly as central nervous system stimulant, heart stimulant, as diuretic and for treatment of lung diseases¹⁴⁹ (Figure 1.8). Some authors classify compounds such as ephedrine and cathinone as pseudoalkaloids. These structures originate from the amino acid phenylalanine, but acquire their nitrogen atom not from the amino acid but through transamination¹⁴⁹.

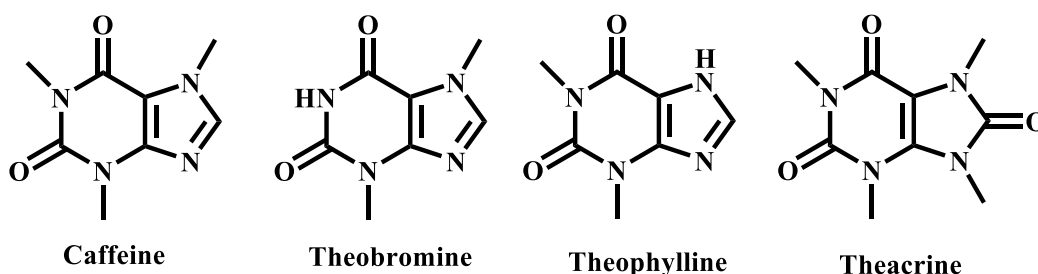


Figure 1.8: Some examples of pseudoalkaloids from plants.

1.6.1.6 Isoquinoline alkaloids

Some alkaloids do not have the carbon skeleton, which is characteristic of their group, but are still classified as isoquinoline alkaloids, like galantamine, homoaporphines, etc (Figure 1.9). These are mainly used for treatment of dementia.

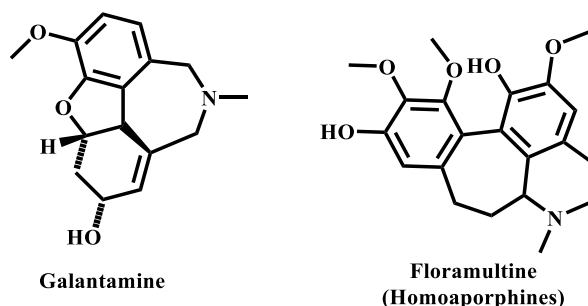


Figure 1.9: Some examples of isoquinoline alkaloids from plants.

1.6.2 Terpene Indole Alkaloids / Isoprenoid Indole Alkaloids (TIAs)

TIAs are made up of isoprenoid building blocks derived from IPP and DMAPP, along with residues of tryptophan or tryptamine. They are one of the major group of alkaloids with over 4000 known different compounds, mainly in plants. They are found mainly in eight plant families, of which the *Apocynaceae*, the *Loganiaceae*, and the *Rubiaceae* provide the best sources. Many TIAs are known to possess significant bioactivity. These are classified into ergot alkaloids, monoterpene indole alkaloids (MIAs) and bisindole alkaloids.

1.6.2.1 Ergot alkaloids

These are formed by multi-stage reactions involving tryptophan and DMAPP, related to lysergic acid and are hemiterpenoids (Figure 1.10).

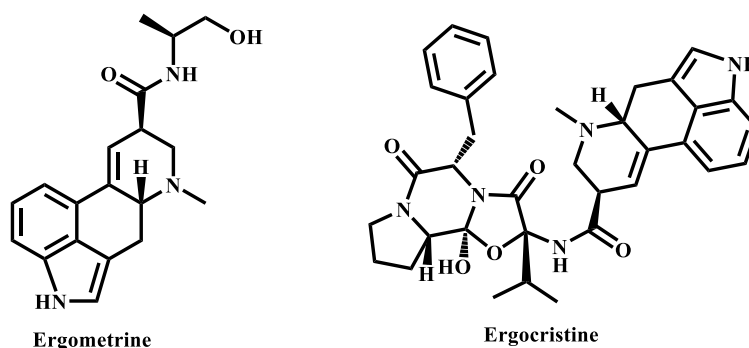


Figure 1.10: Some examples of ergot alkaloids from fungi.

Some examples are ergometrine (water soluble amino alcohol derivatives used to cause contraction of the uterus during childbirth), ergotamine, ergostine and ergocristine (water-insoluble polypeptide derivatives). They are also used to treat headaches due to migraine.

1.6.2.2 Monoterpene Indole alkaloids (MIAs)

The MIAs mainly include a C₉ or C₁₀ fragment (Figure 1.11) originating from secologanin, which is formed from IPP and DMAPP through 10 enzymatic steps. Secologanin combines with tryptamine to form strictosidine, which further is converted to various MIAs. Depending on the structure of this fragment, they are divided into Corynanthe (ajmaline, sarpagin, reserpine), Iboga (ibogaine, voacangine) and Aspidosperma (eburnamin, tabersonine, vindoline). They are mainly used as anti-hypertensive drugs in the treatment of high blood pressure.

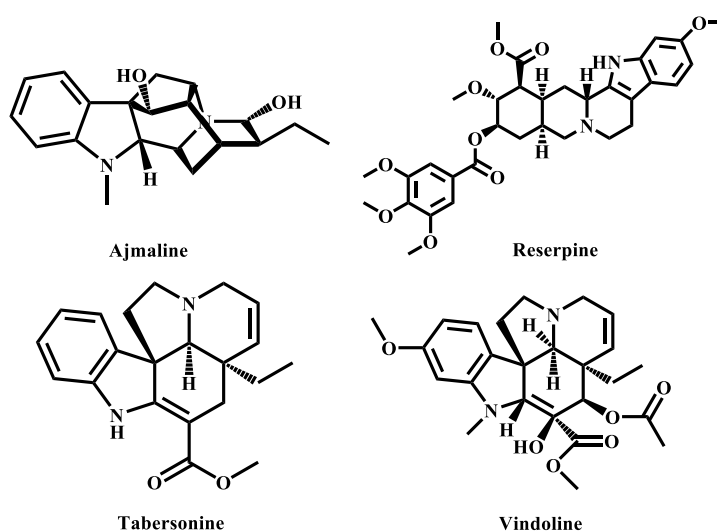


Figure 1.11: Some examples of monoterpene indole alkaloids (MIAs) from plants.

1.6.2.3 Bisindole alkaloids

These are also known as dimeric indole alkaloids and have more than 200 alkaloids classified under this category, for which MIAs serve as the precursor. Some examples are voacamine, villalstonine, toxiferine, vinblastine and vincristine (Figure 1.12). Most of these are used in combination therapy for the treatment of various types of cancer⁸.

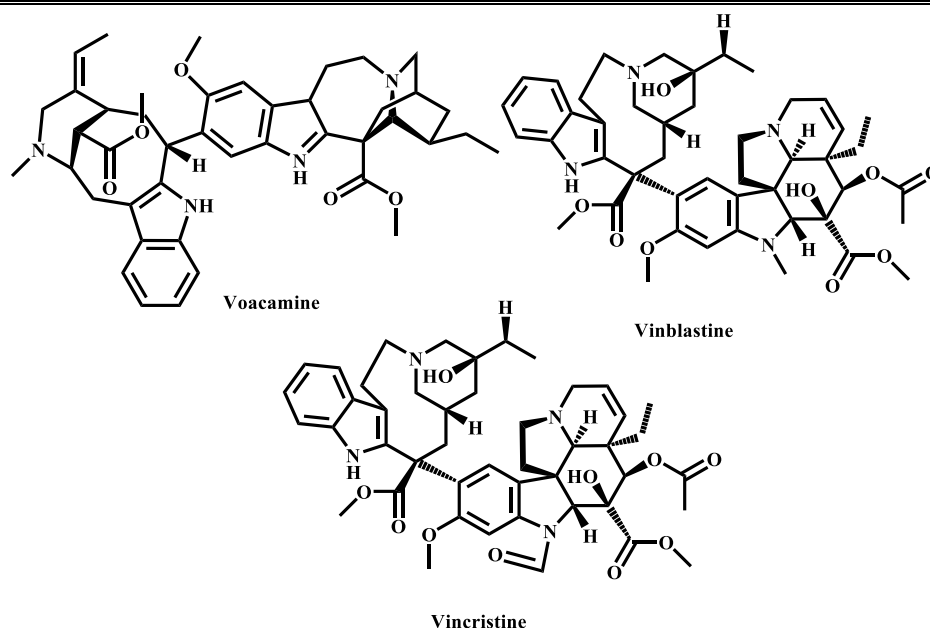


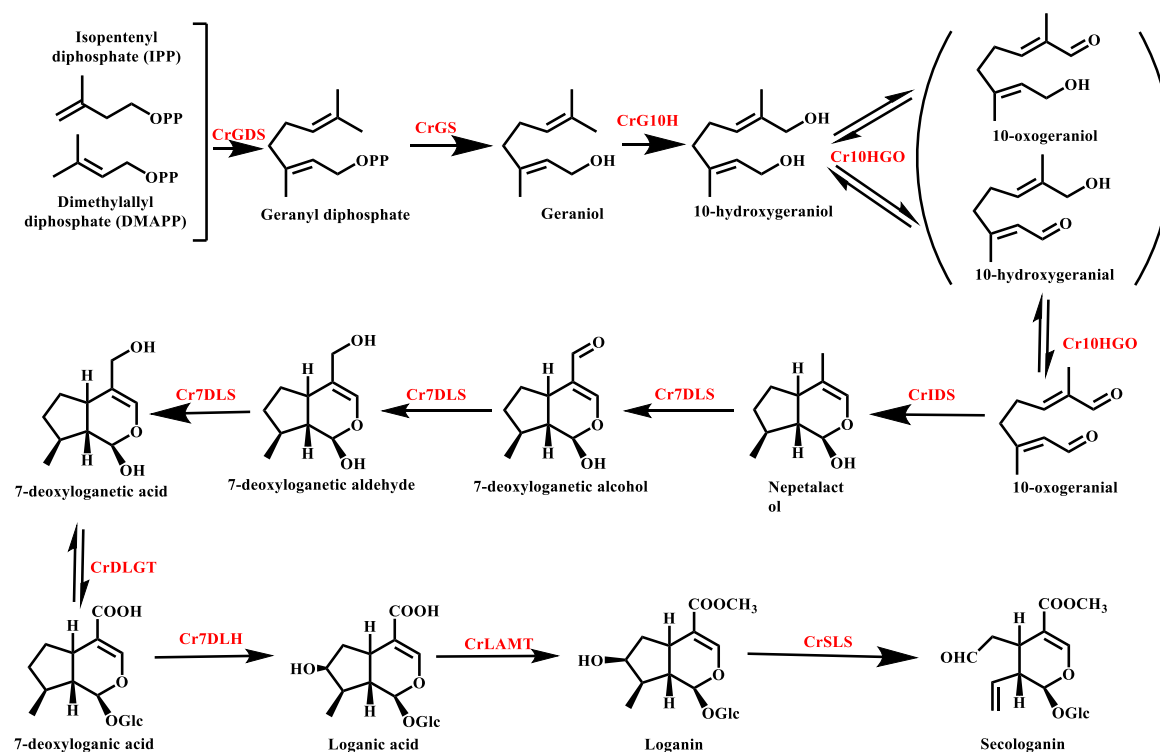
Figure 1.12: Some examples of bisindole alkaloids from plants.

1.6.3 Biosynthesis of Alkaloids

1.6.3.1 Biosynthesis of Secologanin

The monoterpene indole alkaloids (MIAs) biosynthesis diverges from the isoprenoid biosynthetic pathway at 1'-4 chain elongation intermediate geranyl diphosphate (GPP) formed through head-to-tail condensation of isopentenyl diphosphate (IPP) with dimethylallyl diphosphate (DMAPP) catalyzed by Geranyl diphosphate synthase (GDS)¹⁵¹. Further, geraniol synthase (GS)¹⁵² hydrolyses GPP into geraniol, which undergoes hydroxylation at C10 to form 10-hydroxygeraniol catalysed by the cytochrome P450 system, Geraniol 10-hydroxylase (G10H)¹⁵³ (Scheme 1.18). Feeding experiments with labelled 10-hydroxygeraniol, 10-hydroxynerol, iridodiol in *Catharanthus roseus* and *Lonicera morrowii* suspension cultures clearly indicated that 10-hydroxygeraniol is oxidized to 10-oxogeraniol by the oxidoreductase system, 10-hydroxygeraniol dehydrogenase (10HGO)¹⁵⁴⁻¹⁵⁷ (Scheme 1.18). Recently, a short chain reductive cyclase, iridoid synthase (IDS)¹⁵⁸, which cyclises 10-oxogeraniol into an equilibrium mixture of *cis-trans*-nepetalalcol and iridodials, has been characterized. This mixture of nepetalactols is converted into 7-deoxyloganetic alcohol, 7-deoxyloganetic aldehyde and then 7-deoxyloganetic acid by a single enzyme, 7-deoxyloganetic acid synthase (7DLS) in a series of reactions. Glycosylation of 7-deoxyloganetic acid by the 7-deoxyloganetic acid glucosyltransferase (DLGT) enzyme results in the formation of 7-deoxyloganic acid,

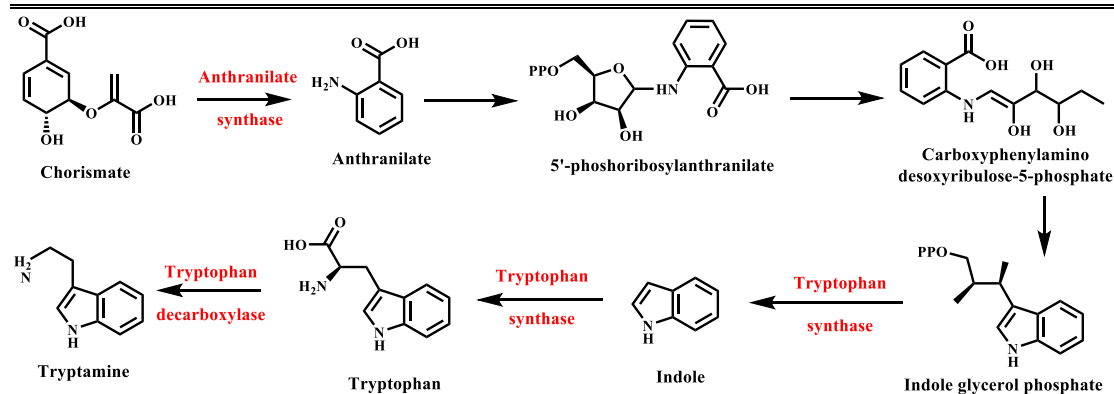
which is then hydroxylated by 7-deoxyloganic acid hydroxylase (7DLH) to form loganic acid. Loganic acid is acted upon by a methyltransferase enzyme, Loganic acid methyltransferase (LAMT) to form loganin, which is then converted to secologanin by Secologanin Synthase (SLS)¹⁵⁹ (Scheme 1.9).



Scheme 1.9: Secologanin Biosynthetic Pathway

1.6.3.2 Biosynthesis of Tryptamine

Tryptamine is produced by the action of tryptophan decarboxylase (TDC) on tryptophan. TDC is encoded by a single gene in *C. roseus*¹⁶⁰. The first committed step of tryptophan synthesis is the conversion of chorismate into anthranilate by Anthranilate synthase (AS). AS and TDC are thought to be the key regulated enzymes in the production of tryptamine (Scheme 1.10).



Scheme 1.10: Biosynthesis of Tryptamine

1.6.3.3 Biosynthesis of Vinblastine and Vincristine

Strictosidine is the vital intermediate in the biosynthesis of many indole alkaloids, not only in *C. roseus*, but a host of other organisms too. It is derived from the condensation of secologanin and tryptamine by strictosidine synthase (STR). The complete genomic sequence of STR gene was cloned by Pasquali *et al.*¹⁶¹. The expression of the STR gene is down regulated by auxin and upregulated by methyl jasmonic acid (MeJA)¹⁶¹.

The first committed step in the vincristine and vinblastine biosynthesis is the removal of glucose moiety of stricosidine by the enzyme strictosidine β -D-glucosidase (SGD) to form strictosidine aglycoside and then it undergoes a series of reactions to produce 4,21-dehydrogeissoschizine^{162,163}. SGD shares ~ 60 % homology with other plant glucosidases.

vindoline by deacetylvindoline-4-O-acetyltransferase (DAT)¹⁷⁰. The enzymes, T16H, D4H and DAT are thought to be the main regulatory targets of transcription factors (Scheme 1.11).

Catharanthine and vindoline condense together to form α -3',4'-anhydrovinblastine (AVLB), the reaction mediated by AVLB synthase¹⁷¹. The mechanism of formation of vinblastine and vincristine from AVLB remains unclear till date.

1.7 Metabolic Engineering

Catharanthus roseus is an evergreen herbaceous plant belonging to the family *Apocynaceae*, with flowers ranging from white to dark pink. It has been long raised for herbal medicine and also as an ornamental plant. It is native to the Indian Ocean island of Madagascar, and hence the name Madagascar periwinkle. It can be found commonly in tropical and sub-tropical regions of the world^{172,173}.

The leaf extracts of *C. roseus* have been used in ancient medicine against various disorders and deficiencies¹⁷⁴. Leaf and flower extracts are found to be used as a treatment for wasp sting, as an anti-hemorrhagic, to treat laryngitis, as an eyewash for infants, for menorrhagia, rheumatism, asthma, tuberculosis, etc^{173,175}. Studies have shown that *C. roseus* leaf extracts reduce blood glucose level in rabbits¹⁷⁶ and mice¹⁷⁷; act as *in vitro* anti-angiogenesis agents¹⁷⁸, cause larval mortality of the malarial vector, *Anopheles stephensi*¹⁷⁹, exhibit anti-proliferative activity¹⁸⁰. Extracts of all parts of *C. roseus* have been used to treat malaria, diarrhea, diabetes, cancer and skin diseases¹⁸¹.

The activities of *C. roseus* tissues are attributed to the wide range of terpene indole alkaloids (TIAs) present in the plant. *C. roseus* is known to produce more than 200 alkaloids in various tissues. These TIAs are a multifarious class of natural products with significant structural diversity. They possess various pharmaceutical properties such as the anti-cancer activity of vincristine and vinblastine^{182,183}, the anti-hypertensive activity of reserpine and ajmalicine¹⁸⁴ and the anti-arrhythmic activity of ajmaline¹⁸⁵. Certain TIAs even have potent activities against pathogens¹⁶³.

Only very small quantities of these alkaloids can be obtained from plant extracts as the alkaloid content in *C. roseus* is only 0.005 % dry weight¹⁸⁶. Owing to this, it is very difficult to isolate these alkaloids in sufficient amounts from plant

tissue samples to meet the requirement. Hence, an insight into their biosynthesis is very valuable and can be used for *in vivo* production of the alkaloids in bacterial or yeast hosts, by means of metabolic engineering.

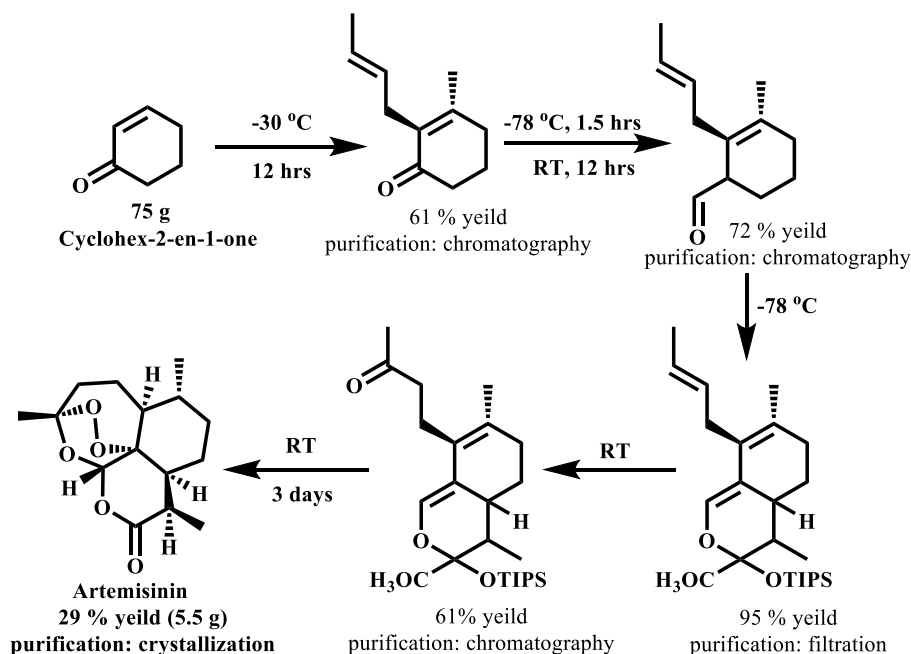
Metabolic Engineering essentially deals with the genetic manipulation of microorganisms in order to increase productivity of a specific metabolite, like mutations, over expressions, modifications in copy number of particular genes, etc. It is also applied for decreasing the production of some by-products, toxic substances, etc. In the last few years, metabolic engineering has been applied to various pathways for the large-scale production of metabolites of medical and commercial significance.

1.7.1 Metabolic engineering for the production of isoprenoids

Isoprenoids are the most structurally diverse natural products and have found various applications in medicine (artemisinin as antimalarial, taxol as anticancer), nutrition (carotenoids), flavours and fragrances (essential oils). Plants are one of the major sources for the production of these isoprenoids. However the supply of these products suffers from low yields, impurities and consumption of large amounts of natural resources. Chemical synthesis of these natural isoprenoids is difficult and costly because of their structural complexity and low yields. Engineering of metabolic pathway in microbial systems for the production of terpenes in large scale presents a cost-effective alternative tool for isoprenoid biosynthesis. Recent advances in molecular techniques have enabled the engineering of microorganisms for the production of isoprenoids. Metabolic engineering for the production of isoprenoids requires two main steps: (i) the production of common intermediate prenyldiphosphate (farnesyldiphosphate) by FDS from basic building blocks (IPP and DMAPP) synthesized by either MVA or MEP pathway, and (ii) conversion of prenyldiphosphate into products of interest by incorporating heterologous genes into *Escherichia coli* or yeast (like *Saccharomyces cerevisiae*) through synthetic biology tools, examples are microbial engineering of artemisinin and taxol.

Artemisinin, a sesquiterpene lactone extracted from *Artemisia annua* is an antimalarial drug and highly effective against multidrug resistant strains of *P. falciparum*¹⁸⁷. Artemisinin is accumulated in glandular trichome to the level of 0.01-1 % of dry weight¹⁸⁸. Supply from natural resource is limiting because of low yield. Although, chemical synthesis is possible, structural complexity, multiple time-

consuming steps and low yield makes it economically nonviable for drug production¹⁸⁹ (Scheme 1.12).



Scheme 1.12: Chemical synthesis of Artemisinin

The cost of one ton leaves was 1200-1500 USD. The global demand of malarial treatments is 230 – 260 million per year and subsequently, the global requirement of artemisinin is 130-150 tons per year. The lower content in the leaves leads to an increase in the price of artemisinin. The market price of artemisinin was 750 USD/Kg in 2004 and 1100 USD/Kg in 2005. Therefore, a semi-synthetic route for the production of artemisinin could be utilized by starting from artemisinic acid¹⁹⁰. 25 g/litre of artemisinic acid was produced in an engineered yeast strain (*S. cerevisiae*) starting from a simple inexpensive carbon source by fermentation¹⁹¹ (Figure 1.13). This result would constitute a superior route for the development of economically viable processes for the production of semi synthetic artemisinin.

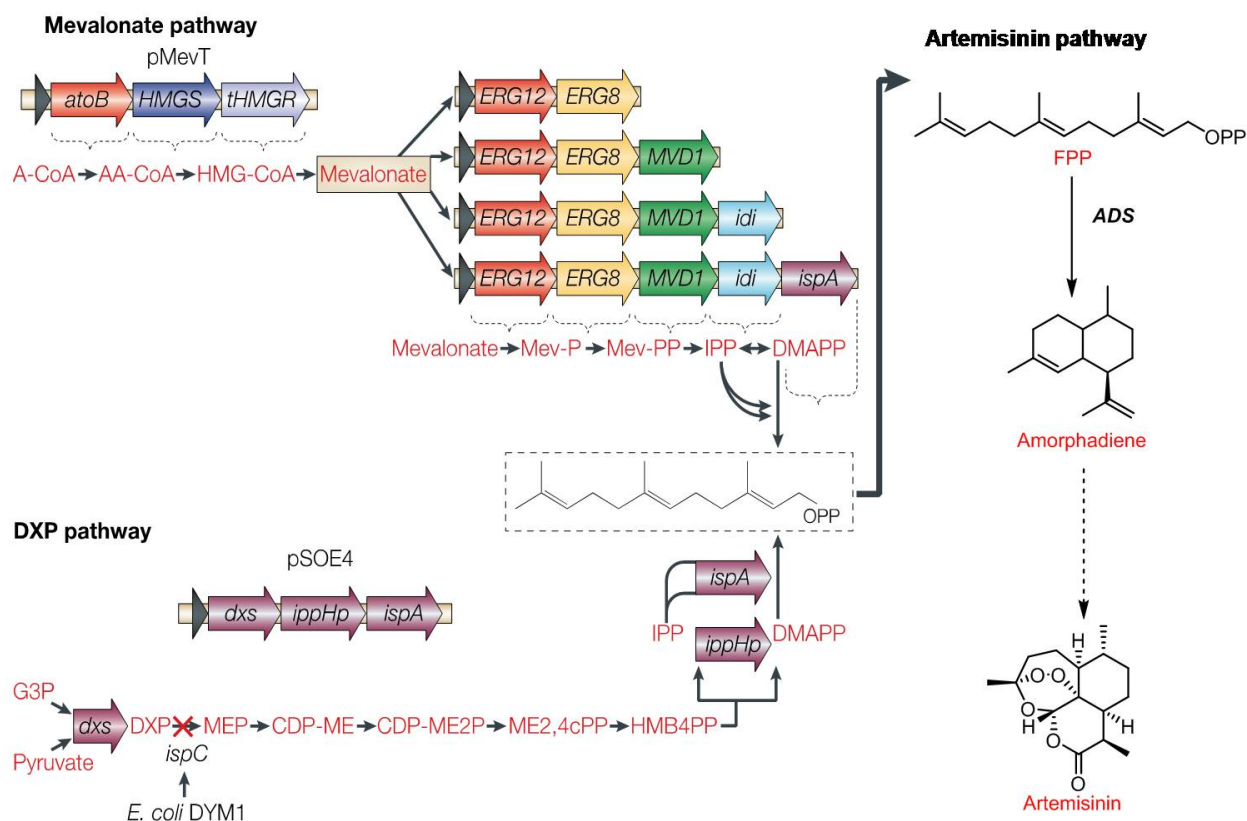


Figure 1.13: Metabolic Engineering for the production of Artemisinic acid¹⁹¹.

Over expression of tHMGR resulted in a 5-fold increase in amorphadiene production; where downregulation of ERG9 (Squalene Synthase) resulted in a 2-fold increase in amorphadiene production. Over expression of the gene encoding FPP Synthase had little effect on amorphadiene production. Combining all these modifications resulted in a 500-fold increase in amorphadiene production than normal ADS gene expression.

This construct was further expressed in yeast (*S. cerevisiae*), wherein by increasing FPP production, expressing the ADS gene and expressing a novel CYP450 and its redox partner, 25 g/litre of artemisinic acid was produced (Figure 1.14). This reduced the total cost of artemisinin to 230-215 USD/Kg.

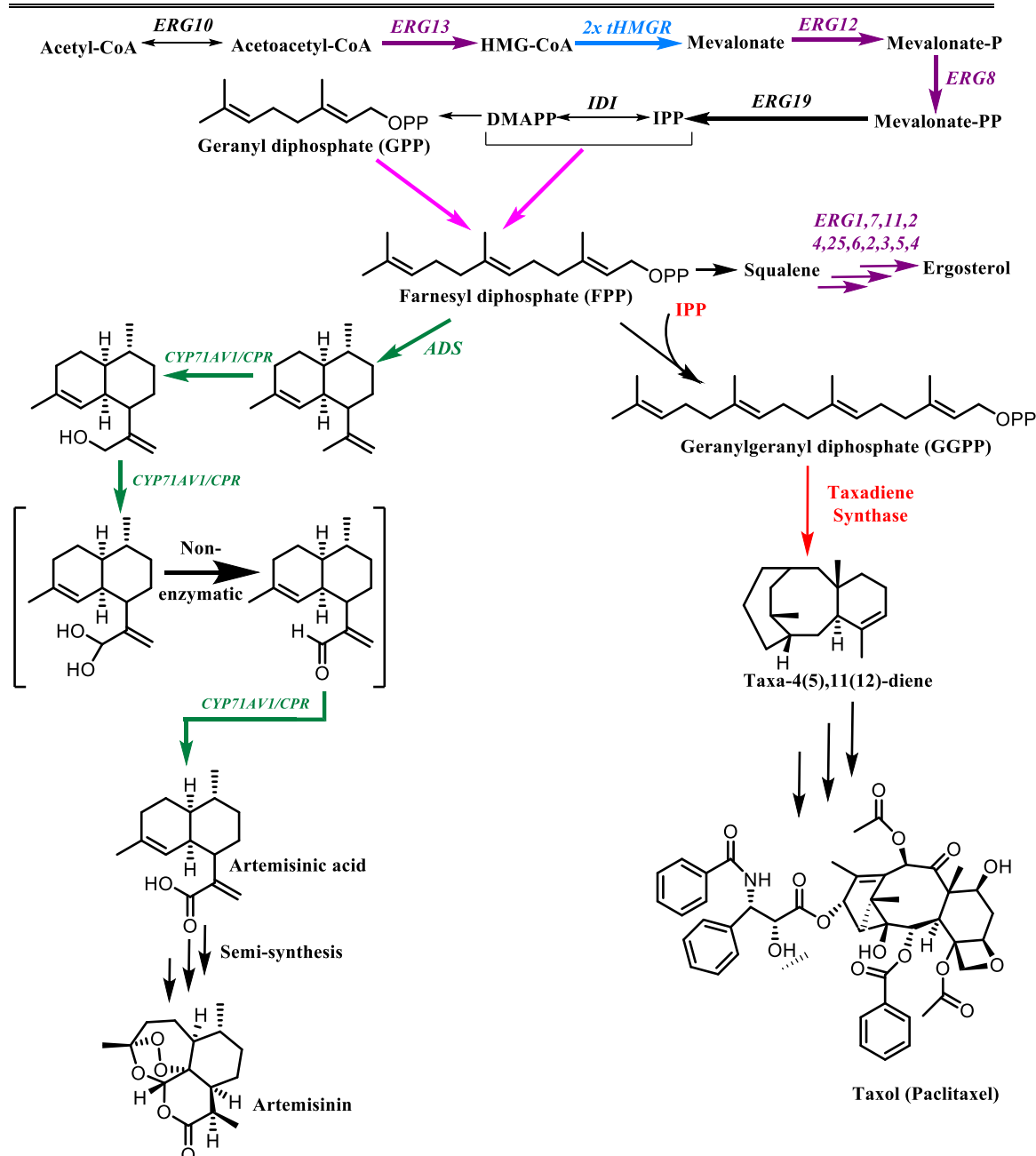
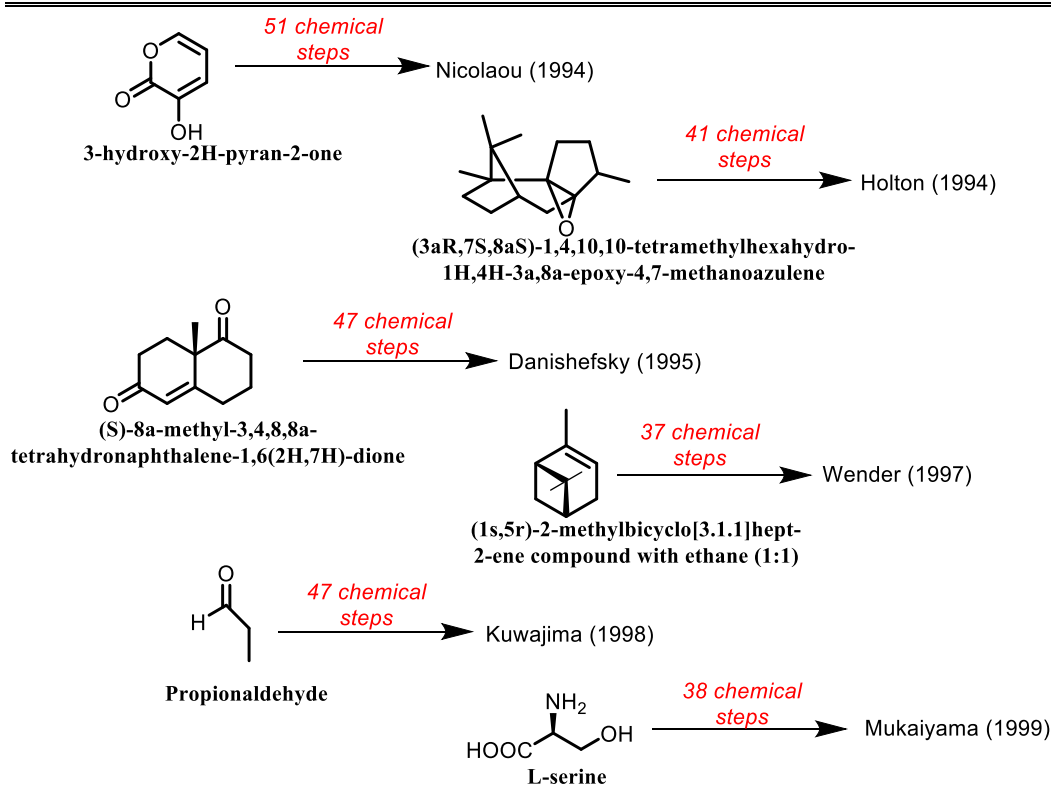


Figure 1.14: Semi-synthesis of Artemisinin and Taxol in *S. cerevisiae*.

Taxol is a polyoxygenated cyclic diterpene and is well-established as an anti cancer molecule¹⁹² and as a potent antineoplastic drug against a wide range of cancer¹⁹³. Naturally, taxol is isolated from the bark of *taxus sp.* with a yield of 0.001-0.05 % of dry weight¹⁹⁴. Total synthesis of taxol is reported but the approach is commercially not viable¹⁹⁵, therefore increased demand for drug shifted from large scale destruction of low yielding *Taxus sp.* to semi-synthetic approach¹⁹⁶ (Scheme 1.13).



Scheme 1.13: Chemical synthesis of Taxol.

Taxol can be produced semi-synthetically from more abundant taxoid like baccatin III and 10-deacetyl baccatin III isolated from needles or cell cultures of various *Taxus* species as renewable resources¹⁹⁷⁻²⁰⁰. However, the extraction of semi-synthetic precursors from *Taxus* is also very difficult and expensive. It takes 3 trees to give 10 Kg bark, from which a meager 1 g Taxol is obtained²⁰¹ and its cost is 200,000 USD/ Kg. Therefore, the production of taxol or its precursor in a genetically engineered microorganism (Figure 1.14) is a desirable alternative for economic viability and increase in yield.

The use of *T. chinensis* geranylgeranyl diphosphate synthase and *T. chinensis* Taxadiene Synthase in the construct resulted in a 67 % increase in taxadiene production. The implementation of the truncated HMGR gene increased the production of IPP for terpenoid production and resulted in a further 50 % increase in taxadiene production. The UPC2-1 aided in an exogenous steroid uptake, which only marginally increased taxadiene production. The heterologous expression of *S. acidocaldarius* geranylgeranyl diphosphate synthase in addition to *T. chinensis* taxadiene synthase, tHMG-CoA reductase and upc2-1 resulted in marginal increase in taxadiene production, but a 100-fold increase in GGPP production. A further codon-

optimized *T. chinensis* taxadiene synthase, truncated HMG-CoA reductase, the UPC2-1 transcription factor gene and *S. acidocaldarius* geranylgeranyl diphosphate synthase resulted in a 40-fold increase in taxadiene production.

1.8 Scope of the work in Monoterpenoid Biosynthetic Pathway

Monoterpene indole alkaloids (MIAs) are a multifarious class of natural products with distinct chemical and biological properties²⁰²⁻²⁰⁴ found in numerous species of plants belonging to the Apocynaceae, Loganiaceae and Rubiaceae families. To date, over 3000 MIAs are known with diverse structures and biological activities.

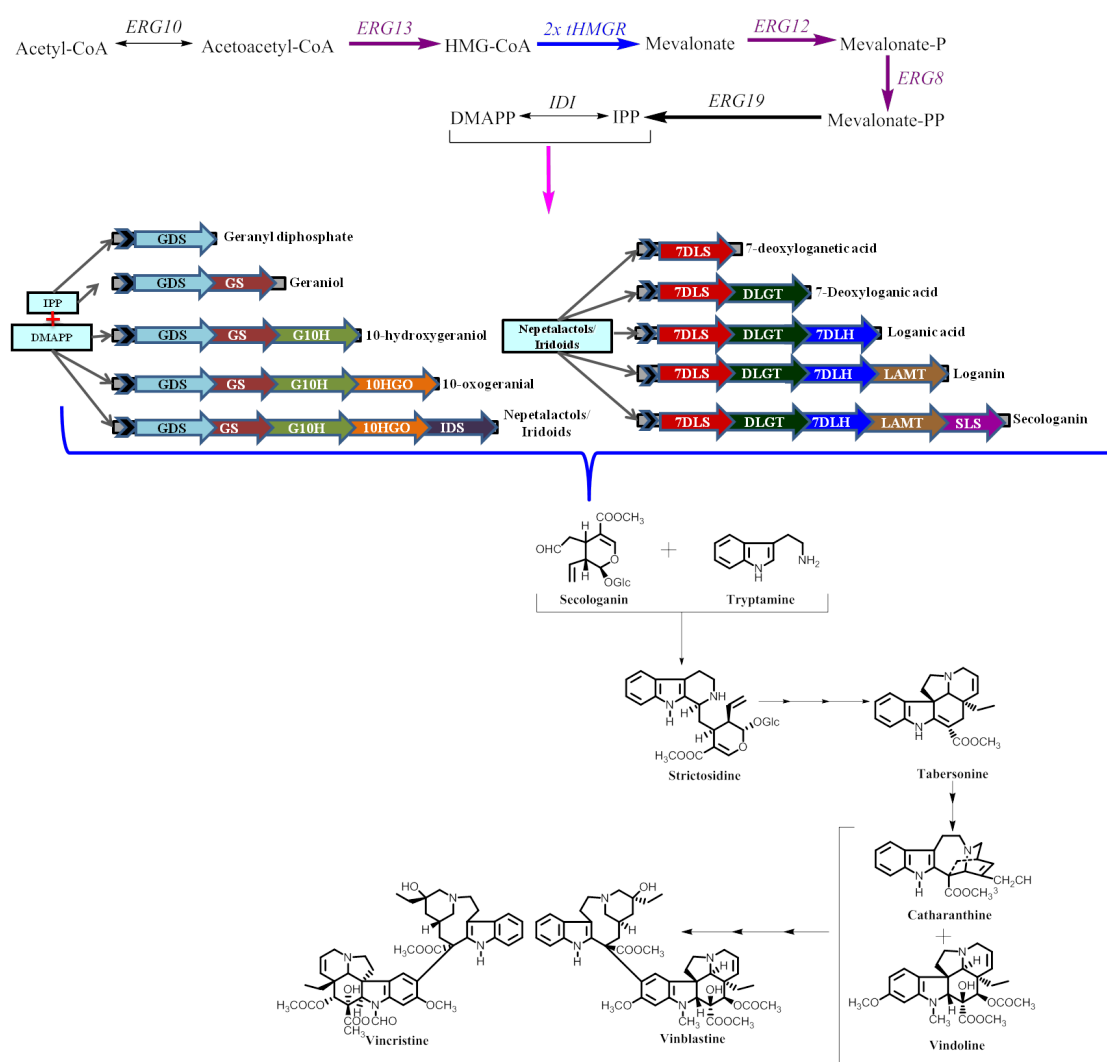


Figure 1.15: Scope of the work in monoterpene biosynthesis.

These MIAs are inclusive of the anticancer alkaloids vincristine/vinblastine^{182,183} and camptothecin²⁰⁵, the anti-malarial quinine²⁰⁶, the antihypertensive drug ajmalicine²⁰⁷, the anti-arrhythmic drug, sarpagine²⁰⁸ and rat poison/ pesticide strychnine-n-oxide²⁰⁹. These MIAs are synthesized by the

condensation of tryptamine and the iridoid monoterpene, secologanin to give strictosidine, which is the precursor for all the afore-mentioned therapeutic compounds.

Over 200 indole alkaloids are known from *C. roseus*, with catharanthine and vindoline present in greatest abundance. Like in the case of metabolic engineering of taxol and artemisinin, the pathway enzymes leading to the formation of IPP and DMAPP can be modified to boost the biosynthesis of terpene indole alkaloids. This can further aid in large-scale production of the anti-cancerous compounds, vincristine and vinblastine (Figure 1.15), by suitable scaling-up experiments. Owing to these prospective, the objectives of this thesis were designed.

OBJECTIVES OF THE THESIS

- Identification of the genes involved in Secologanin biosynthesis in *Catharanthus roseus* by total RNA isolation and transcriptome sequencing.
- Cloning and characterization of the genes involved in the Secologanin biosynthetic pathway.
- Employing strategies to gain insights into the factors involved in the formation of iridoids.

1.10 References

1. Dixon, R.A. Plant natural products: the molecular genetic basis of biosynthetic diversity. *Curr. Opin. Biotech.* **10**, 192-197 (1999).
2. Bevan, M., Bancroft, I., Bent, E., Love, K., Goodman, H., Dean, C., Bergkamp, R., Dirkse, W., Van Staveren, M., Stiekema, W. & Drost, L. Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. *Nature* **391**, 485-488 (1998).
3. Somerville, C. & Somerville, S. Plant Functional Genomics. *Science* **285**, 380-383 (1999).
4. Uarrotta, V.G., Severino, R.B. & Maraschin, M. Maize Landraces (*Zea mays L.*): A New Prospective Source for Secondary Metabolite Production. *Int. J. Agric. Res.* **6**, 218-226 (2011).
5. Koes, R.E., Quattrocchio, F. & Mol, J.N.M. The flavonoid biosynthetic pathway in plants: Function and evolution. *BioEssays* **16**, 123-132 (1994).
6. Singh, B. & Sharma, R.A. Plant terpenes: defense responses, phylogenetic analysis, regulation and clinical applications. *Biotech.* **5**, 129-151 (2015).
7. Caporale, L.H. Chemical ecology: a view from the pharmaceutical industry. *Proc. Natl. Acad. Sci. USA* **92**, 75-82 (1995).
8. Gajalakshmi, S., Vijayalakshmi, S. & Rajeswari, V.D. Pharmacological activities of *Catharanthus roseus*: A perspective review. *Int. J. Pharm. Biol. Sci.* **4**, 431-439 (2013).
9. Miettinen, K., Dong, L., Navrot, N., Schneider, T., Burlat, V., Pollier, J., Woittiez, L., van der Krol, S., Lugin, R., Ilc, T. & Verpoorte, R. The seco-iridoid pathway from *Catharanthus roseus*. *Nat. Commun.* **5**, 3606-3617 (2014).
10. Thulasiram, H.V., Erickson, H.K. & Poulter, C.D. Chimeras of two isoprenoid synthases catalyze all four coupling reactions in isoprenoid biosynthesis. *Science* **316**, 73-76 (2007).
11. Thulasiram, H.V. & Poulter, C.D. Farnesyl Diphosphate Synthase: The Art of Compromise between Substrate Selectivity and Stereoselectivity. *J. Am. Chem. Soc.* **128**, 15819-15823 (2006).
12. Caputi, L. & Aprea, E. Use of terpenoids as natural flavouring compounds in food industry. *Recent Pat. Food Nutr. Agric.* **3**, 9-16 (2011).
13. Belanger, J.T. Perillyl alcohol: applications in oncology. *Altern. Med. Rev.* **3**, 448-57 (1998).
14. Funhoff, E.G., Bauer, U., Garcia-Rubio, I., Witholt, B. & van Beilen, J.B. CYP153A6, a soluble P450 oxygenase catalyzing terminal-alkane hydroxylation. *J. Bacteriol.* **188**, 5220-5227 (2006).
15. Isman, M.B. Plant essential oils for pest and disease management. *Crop Prot.* **19**, 603-608 (2000).
16. Knobloch, K., Pauli, A., Iberl, B., Weigand, H. & Weis, N. Antibacterial and Antifungal Properties of Essential Oil Components. *J. Essent. Oil Res.* **1**, 813-829 (1989).

17. Sharkey, T.D. & Yeh, S. Isoprene Emission from Plants. *Annu. Rev. Plant Phys.* **52**, 407-436 (2001).
18. Silver, G.M. & Fall, R. Enzymatic Synthesis of Isoprene from Dimethylallyl Diphosphate in Aspen Leaf Extracts. *Plant Physiol.* **97**, 1588-1591 (1991).
19. Kuzma, J. & Fall, R. Leaf Isoprene Emission Rate Is Dependent on Leaf Development and the Level of Isoprene Synthase. *Plant Physiol.* **101**, 435-440 (1993).
20. Xu, Y.Q., Yao, Z., Hu, J.Y., Teng, J.I.E., Takaishi, Y.O. & Duan, H.Q.. Immunosuppressive terpenes from *Prinsepia utilis*. *J. Asian Nat. Prod. Res.* **9**, 637-642 (2007).
21. Jiang, Z.H., Wang, J.R., Li, M., Liu, Z.Q., Chau, K.Y., Zhao, C. & Liu, L. Hemiterpene glucosides with anti-platelet aggregation activities from *Ilex pubescens*. *J. Nat. Prod.* **68**, 397-399 (2005).
22. Garcia, R., Alves, E.S., Santos, M.P., Aquije, G.M., Fernandes, A.A.R., Santos, R.B.D., Ventura, J.A. & Fernandes, P. Antimicrobial activity and potential use of monoterpenes as tropical fruits preservatives. *Braz. J. Microbiol.* **39**, 163-168 (2008).
23. Singh, P., Shukla, R., Prakash, B., Kumar, A., Singh, S., Mishra, P.K. & Dubey, N.K. Chemical profile, antifungal, antiaflatoxic and antioxidant activity of *Citrus maxima* Burm. and *Citrus sinensis* (L.) Osbeck essential oils and their cyclic monoterpene, DL-limonene. *Food Chem. Toxicol.* **48**, 1734-1740 (2010).
24. Karkabounas, S., Kostoula, O.K., Daskalou, T., Veltsistas, P., Karamouzis, M., Zelovitis, I., Metsios, A., Lekkas, P., Evangelou, A.M., Kotsis, N. & Skoufos, I. Anticarcinogenic and antiplatelet effects of carvacrol. *Exp. Oncol.* **28**, 121-125 (2006).
25. Kladniew, B.R., Polo, M., Villegas, S.M., Galle, M., Crespo, R. & de Bravo, M.G. Synergistic antiproliferative and anticholesterogenic effects of linalool, 1,8-cineole, and simvastatin on human cell lines. *Chem. Biol. Interact.* **214**, 57-68 (2014).
26. Hampel, D., Mosandl, A. & Wust, M. Biosynthesis of mono- and sesquiterpenes in carrot roots and leaves (*Daucus carota* L.): metabolic cross talk of cytosolic mevalonate and plastidial methylerythritol phosphate pathways. *Phytochemistry* **66**, 305-311 (2005).
27. Yu, F. & Utsumi, R. Diversity, regulation, and genetic manipulation of plant mono- and sesquiterpenoid biosynthesis. *Cell. Mol. Life Sci.* **66**, 3043-3052 (2009).
28. Bommareddy, A., Rule, B., VanWert, A.L., Santha, S. & Dwivedi, C. \pm -Santalol, a derivative of sandalwood oil, induces apoptosis in human prostate cancer cells by causing caspase-3 activation. *Phytomedicine* **19**, 804-811 (2012).
29. Klayman, D.L. Qinghaosu (artemisinin): an antimalarial drug from China. *Science* **228**, 1049-1055 (1985).
30. Tippmann, S., Chen, Y., Siewers, V. & Nielsen, J. From flavors and pharmaceuticals to advanced biofuels: Production of isoprenoids in *Saccharomyces cerevisiae*. *Biotechnol. J.* **8**, 1435-1444 (2013).
31. Rowinsky, E.K. The development and clinical utility of the taxane class of antimicrotubule chemotherapy agents. *Annu. Rev. Med.* **48**, 353-374 (1997).

32. Ohvo-Rekila, H., Ramstedt, B., Leppimäki, P. & Slotte, J.P. Cholesterol interactions with phospholipids in membranes. *Prog. Lipid Res.* **41**, 66-97 (2002).
33. Hartmann, M.A. Plant sterols and the membrane environment. *Trends Plant Sci.* **3**, 170-175 (1998).
34. Lutjohann, D. Sterol autoxidation: from phytosterols to oxyphytosterols. *Br. J. Nutr.* **91**, 3-4 (2004).
35. Sakurai, A. Brassinosteroid biosynthesis. *Plant Physiol. Bioch.* **37**, 351-361 (1999).
36. Bishop, G.J. & Yokota, T. Plants Steroid Hormones, Brassinosteroids: Current Highlights of Molecular Aspects on their Synthesis/Metabolism, Transport, Perception and Response. *Plant Cell Physiol.* **42**, 114-120 (2001).
37. Clouse, S.D., Langford, M. & McMorris, T.C. A brassinosteroid-insensitive mutant in *Arabidopsis thaliana* exhibits multiple defects in growth and development. *Plant Physiol.* **111**, 671-678 (1996).
38. Lichtenthaler, H.K., Rohmer, M. & Schwender, J. Two independent biochemical pathways for isopentenyl diphosphate and isoprenoid biosynthesis in higher plants. *Physiol. Plantarum* **101**, 643-652 (1997).
39. Hirschberg, J. Carotenoid biosynthesis in flowering plants. *Curr. Opin. Plant Biol.* **4**, 210-218 (2001).
40. Demmig-Adams, B., Gilmore, A.M. & Adams, W.W. Carotenoids 3: *in vivo* function of carotenoids in higher plants. *FASEB J.* **10**, 403-412 (1996).
41. Rao, A.V. & Rao, L.G. Carotenoids and human health. *Pharmacol. Res.* **55**, 207-216 (2007).
42. Sagami, H., Igarashi, Y., Tateyama, S., Ogura, K., Roos, J. & Lennarz, W.J. Enzymatic Formation of Dehydrodolichal and Dolichal, New Products Related to Yeast Dolichol Biosynthesis. *J. Biol. Chem.* **271**, 9560-9566 (1996).
43. Sagami, H., Kurisaki, A. & Ogura, K. Formation of dolichol from dehydrodolichol is catalyzed by NADPH-dependent reductase localized in microsomes of rat liver. *J. Biol. Chem.* **268**, 10109-10113 (1993).
44. Chojnacki, T. & Dallner, G. The biological role of dolichol. *Biochem. J.* **251**, 1-9 (1988).
45. Pootakham, W., Chanprasert, J., Jomchai, N., Sangsrakru, D., Yoocha, T., Therawattanasuk, K. & Tangphatsornruang, S. Single nucleotide polymorphism marker development in the rubber tree, *Hevea brasiliensis* (*Euphorbiaceae*). *Am. J. Bot.* **98**, e337-e338 (2011).
46. Backhaus, R.A. Rubber formation in plants: a mini-review. *Israel J. Bot.* **34**, 283-293 (1985).
47. Oh, S.K., Kang, H., Shin, D.H., Yang, J., Chow, K.S., Yeang, H.Y., Wagner, B., Breiteneder, H. & Han, K.H. Isolation, characterization, and functional analysis of a novel cDNA clone encoding a small rubber particle protein from *Hevea brasiliensis*. *J. Biol. Chem.* **274**, 17132-17138 (1999).
48. Lichtenthaler, H.K. The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 47-65 (1999).

49. Lange, B.M., Rujan, T., Martin, W. & Croteau, R. Isoprenoid biosynthesis: The evolution of two ancient and distinct pathways across genomes. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13172-13177 (2000).
50. McGarvey, D.J. & Croteau, R. Terpenoid metabolism. *Plant Cell* **7**, 1015-1026 (1995).
51. Bach, T.J. Some new aspects of isoprenoid biosynthesis in plants-a review. *Lipids* **30**, 191-202 (1995).
52. Lange, B.M. & Croteau, R. Isoprenoid biosynthesis via a mevalonate-independent pathway in plants: Cloning and heterologous expression of 1-deoxy-D-xylulose-5-phosphate reductoisomerase from peppermint. *Arch. Biochem. Biophys.* **365**, 170-174 (1999).
53. Newman, J.D. & Chappell, J. Isoprenoid biosynthesis in plants: Carbon partitioning within the cytoplasmic pathway. *Crit. Rev. Biochem. Mol. Biol.* **34**, 95-106 (1999).
54. Maldonadomendoza, I.E., Burnett, R.J. & Nessler, C.L. Nucleotide sequence of a cDNA encoding 3-Hydroxy-3-Methylglutaryl Coenzyme A reductase from *Catharanthus roseus*. *Plant Physiol.* **100**, 1613-1614 (1992).
55. Schulte, A.E., Duran, E.M.L., van der Heijden, R. & Verpoorte, R. Mevalonate kinase activity in *Catharanthus roseus* plants and suspension cultured cells. *Plant Sci.* **150**, 59-69 (2000).
56. Schulte, A.E., van der Heijden, R. & Verpoorte, R. Purification and characterization of phosphomevalonate kinase from *Catharanthus roseus*. *Phytochemistry* **52**, 975-983 (1999).
57. Schulte, A.E., van der Heijden, R. & Verpoorte, R. Purification and characterization of mevalonate kinase from suspension-cultured cells of *Catharanthus roseus* (L.) G. Don. *Arch. Biochem. Biophys.* **378**, 287-298 (2000).
58. Ramos-Valdivia, A.C., van der Heijden, R. & Verpoorte, R. Isopentenyl diphosphate isomerase and prenyltransferase activities in rubiaceous and apocynaceous cultures. *Phytochemistry* **48**, 961-969 (1998).
59. Ahumada, I., Cairó, A., Hemmerlin, A., González, V., Pateraki, I., Bach, T.J., Rodríguez-Concepción, M., Campos, N. and Boronat, A. Characterisation of the gene family encoding acetoacetyl-CoA thiolase in *Arabidopsis*. *Funct. Plant Biol.* **35**, 1100-1111 (2008).
60. Carrie, C., Murcha, M.W., Millar, A.H., Smith, S.M. & Whelan, J. Nine 3-ketoacyl-CoA thiolases (KATs) and acetoacetyl-CoA thiolases (ACATs) encoded by five genes in *Arabidopsis thaliana* are targeted either to peroxisomes or cytosol but not to mitochondria. *Plant Mol. Biol.* **63**, 97-108 (2007).
61. Vanderheijden, R., Verpoorte, R. & Duine, J.A. Biosynthesis of 3S-hydroxy-3-methylglutaryl-coenzyme A in *Catharanthus roseus*: acetoacetyl-CoA thiolase and HMG-CoA synthase show similar chromatographic behaviour. *Plant Physiol. Biochem.* **32**, 807-812 (1994).
62. Dyer, J.H., Maina, A., Gomez, I.D., Cadet, M., Oeljeklaus, S. & Schiedel, A.C. Cloning, expression and purification of an acetoacetyl CoA thiolase from sunflower cotyledon. *Int. J. Biol. Sci.* **5**, 736-744 (2009).

63. Schiedl, A.C., Oeljeklaus, S., Minihan, P. & Dyer, J.H. Cloning, expression, and purification of glyoxysomal 3-oxoacyl-CoA thiolase from sunflower cotyledons. *Protein Expr. Purif.* **33**, 25-33 (2004).
64. Vollack, K.U. & Bach, T.J. Cloning of a cDNA encoding cytosolic acetoacetyl-coenzyme A thiolase from radish by functional expression in *Saccharomyces cerevisiae*. *Plant Physiol.* **111**, 1097-1107 (1996).
65. Ishiguro, S., Nishimori, Y., Yamada, M., Saito, H., Suzuki, T., Nakagawa, T., Miyake, H., Okada, K. & Nakamura, K. The Arabidopsis FLAKY POLLEN1 gene encodes a 3-Hydroxy-3-Methylglutaryl-Coenzyme A synthase required for development of tapetum-specific organelles and fertility of pollen grains. *Plant Cell Physiol.* **51**, 896-911 (2010).
66. Montamat, F., Guilloton, M., Karst, F. & Delrot, S. Isolation and characterization of a cDNA encoding *Arabidopsis thaliana* 3-hydroxy-3-methylglutaryl-coenzyme A synthase. *Gene* **167**, 197-201 (1995).
67. Alex, D., Bach, T.J. & Chye, M.L. Expression of *Brassica juncea* 3-hydroxy-3-methylglutaryl CoA synthase is developmentally regulated and stress-responsive. *Plant J.* **22**, 415-426 (2000).
68. Nagegowda, D.A., Bach, T.J. & Chye, M.L. *Brassica juncea* 3-hydroxy-3-methylglutaryl (HMG)-CoA synthase 1: expression and characterization of recombinant wild-type and mutant enzymes. *Biochem. J.* **383**, 517-527 (2004).
69. Pojer, F., Ferrer, J.L., Richard, S.B., Nagegowda, D.A., Chye, M.L., Bach, T.J. & Noel, J.P. Structural basis for the design of potent and species-specific inhibitors of 3-hydroxy-3-methylglutaryl CoA synthases. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11491-11496 (2006).
70. Suwanmanee, P., Sirinupong, N. & Suvachittanont, W. Regulation of the expression of 3-hydroxy-3-methylglutaryl-CoA synthase gene in *Hevea brasiliensis* (BHK) Mull. Arg. *Plant Sci.* **166**, 531-537 (2004).
71. Wegener, A., Gimbel, W., Werner, T., Hani, J., Ernst, D. & Sandermann, H. Molecular cloning of ozone-inducible protein from *Pinus sylvestris* L with high sequence similarity to vertebrate 3-hydroxy-3-methylglutaryl-CoA-synthase. *Biochim. Biophys. Acta* **1350**, 247-252 (1997).
72. Kai, G., Miao, Z., Zhang, L., Zhao, D., Liao, Z., Sun, X., Zhao, L. & Tang, K. Molecular cloning and expression analyses of a new gene encoding 3-hydroxy-3-methylglutaryl-CoA synthase from *Taxus x media*. *Biologia Plantarum* **50**, 359-366 (2006).
73. Enjuto, M., Balcells, L., Campos, N., Caelles, C., Arro, M. & Boronat, A. *Arabidopsis thaliana* contains two differentially expressed 3-hydroxy-3-methylglutaryl-CoA reductase genes, which encode microsomal forms of the enzyme. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 927-931 (1994).
74. Enjuto, M., Lumbreras, V., Marin, C. & Boronat, A. Expression of the Arabidopsis HMG2 gene, encoding 3-hydroxy-3-methylglutaryl coenzyme A reductase, is restricted to meristematic and floral tissues. *Plant Cell* **7**, 517-527 (1995).
75. Godoy-Hernández, G.C., Chappell, J., Devarenne, T.P., García-Pineda, E., Guevara-García, A.A. & Lozoya-Gloria, E. Antisense expression of *hmg1* from *Arabidopsis*

-
- thaliana* encoding 3-hydroxy-3-methylglutaryl coenzyme A reductase, reduces isoprenoid production in transgenic tobacco plants. *J. Plant Physiol.* **153**, 415-424 (1998).
76. Hey, S.J., Powers, S.J., Beale, M.H., Hawkins, N.D., Ward, J.L. & Halford, N.G. Enhanced seed phytosterol accumulation through expression of a modified HMG-CoA reductase. *Plant Biotechnol. J.* **4**, 219-229 (2006).
77. Learned, R.M. Light suppresses 3-hydroxy-3-methylglutaryl coenzyme a reductase gene expression in *Arabidopsis thaliana*. *Plant Physiol.* **110**, 645-655 (1996).
78. Learned, R.M. & Connolly, E.L. Light modulates the spatial patterns of 3-hydroxy-3-methylglutaryl coenzyme A reductase gene expression in *Arabidopsis thaliana*. *Plant J.* **11**, 499-511 (1997).
79. Leivar, P., González, V.M., Castel, S., Trelease, R.N., López-Iglesias, C., Arró, M., Boronat, A., Campos, N., Ferrer, A. & Fernandez-Busquets, X. Subcellular localization of Arabidopsis 3-hydroxy-3-methylglutaryl-coenzyme A reductase. *Plant Physiology* **137**, 57-69 (2005).
80. Lumberras, V., Campos, N. & Boronat, A. The use of an alternative promoter in the Arabidopsis thaliana HMG1 gene generates an mRNA that encodes a novel 3-hydroxy-3-methylglutaryl coenzyme A reductase isoform with an extended N-terminal region. *Plant J.* **8**, 541-549 (1995).
81. Nieto, B., Fores, O., Arro, M. & Ferrer, A. Arabidopsis 3-hydroxy-3-methylglutaryl-CoA reductase is regulated at the post-translational level in response to alterations of the sphingolipid and the sterol biosynthetic pathways. *Phytochemistry* **70**, 53-59 (2009).
82. Ohyama, K., Suzuki, M., Masuda, K., Yoshida, S. & Muranaka, T. Chemical phenotypes of the hmg1 and hmg2 mutants of Arabidopsis demonstrate the in-plant role of HMG-CoA reductase in triterpene biosynthesis. *Chem. Pharm. Bull.* **55**, 1518-1521 (2007).
83. Re, E.B., Brugger, S. & Learned, M. Genetic and biochemical analysis of the transmembrane domain of Arabidopsis 3-hydroxy-3-methylglutaryl coenzyme A reductase. *J. Cell. Biochem.* **65**, 443-459 (1997).
84. Re, E.B., Jones, D. & Learned, R.M. Co-expression of native and introduced genes reveals cryptic regulation of HMG CoA reductase expression in Arabidopsis. *Plant J.* **7**, 771-784 (1995).
85. Rodriguez-Concepcion, M., Fores O., Martinez-Garcia J. F., González V, Phillips MA, Ferrer A, Boronat A. Distinct light-mediated pathways regulate the biosynthesis and exchange of isoprenoid precursors during Arabidopsis seedling development. *Plant Cell* **16**, 144-156 (2004).
86. Suzuki, M., Kamide, Y., Nagata, N., Seki, H., Ohyama, K., Kato, H., Masuda, K., Sato, S., Kato, T., Tabata, S. & Yoshida, S. Loss of function of 3-hydroxy-3-methylglutaryl coenzyme A reductase 1 (HMG1) in Arabidopsis leads to dwarfing, early senescence and male sterility, and reduced sterol levels. *Plant J.* **37**, 750-761 (2004).
87. Suzuki, M., Nakagawa, S., Kamide, Y., Kobayashi, K., Ohyama, K., Hashinokuchi, H., Kiuchi, R., Saito, K., Muranaka, T. & Nagata, N. Complete blockage of the
-

-
- mevalonate pathway results in male gametophyte lethality. *J. Exp. Bot.* **60**, 2055-2064 (2009).
88. Ball, K.L., Dale, S., Weekes, J. & Hardie, D.G. B Biochemical characterization of two forms of 3-hydroxy-3-methylglutaryl-CoA reductase kinase from cauliflower (*Brassica oleracea*). *Eur. J. Biochem.* **219**, 743-750 (1994).
89. Narvaez, J.A., Canche, B.B.C., Perez, P.F. & Madrid, R.R. Differential expression of 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR) during flower and fruit development of *Bixa orellana*. *J. Plant Physiol.* **158**, 1471-1477 (2001).
90. RojasHerrera, R., MonforteGonzalez, M., MendezZeel, M. & LoyolaVargas, V.M. Analysis of the 3-hydroxy 3-methylglutaryl-CoA reductase (HMGR) gen expression in *Coffea arabica* L tissue cultures. *Plant Physiol.* **114**, 1245-1245 (1997).
91. Kato-Emori, S., Higashi, K., Hosoya, K., Kobayashi, T. & Ezura, H. Cloning and characterization of the gene encoding 3-hydroxy-3-methylglutaryl coenzyme A reductase in melon (*Cucumis melo* L. *reticulatus*). *Mol. Genet. Genomics* **265**, 135-142 (2001).
92. Sales, E., Munoz-Bertomeu, J., Arrillaga, I. & Segura, J. Enhancement of cardenolide and phytosterol levels by expression of an N-terminally truncated 3-hydroxy-3-methylglutaryl CoA reductase in transgenic *Digitalis minor*. *Planta Medica* **73**, 605-610 (2007).
93. Chye, M.L., Kush, A., Tan, C.T. & Chua, N.H. Characterization of cDNA and genomic clones encoding 3-hydroxy-3-methylglutaryl-coenzyme A reductase from *Hevea brasiliensis*. *Plant Mol. Biol.* **16**, 567-577 (1991).
94. Barker, J.H., Slocombe, S.P., Ball, K.L., Hardie, D.G., Shewry, P.R. & Halford, N.G. Evidence that barley 3-hydroxy-3-methylglutaryl-coenzyme A reductase kinase is a member of the sucrose nonfermenting-1-related protein kinase family. *Plant Physiol.* **112**, 1141-1149 (1996).
95. Munoz-Bertomeu, J., Sales, E., Ros, R., Arrillaga, I. & Segura, J. Up-regulation of an N-terminal truncated 3-hydroxy-3-methylglutaryl CoA reductase enhances production of essential oils and sterols in transgenic *Lavandula latifolia*. *Plant Biotechnol. J.* **5**, 746-758 (2007).
96. Pechous, S.W. & Whitaker, B.D. Cloning and bacterial expression of a 3-hydroxy-3-methylglutaryl-CoA reductase cDNA (HMG 1) from peel tissue of apple fruit. *J. Plant Physiol.* **159**, 907-916 (2002).
97. Ha, S.H., Lee, S.W., Kim, Y.M. & Hwang, Y.S. Molecular characterization of Hmg2 gene encoding a 3-hydroxy-methylglutaryl-CoA reductase in rice. *Mol. Cells* **11**, 295-302 (2001).
98. Brooker, J.D. & Russell, D.W. Subcellular localization of 3-hydroxy-3-methylglutaryl coenzyme A reductase in *Pisum sativum* seedlings. *Arch. Biochem. Biophys.* **167**, 730-737 (1975).
99. Brooker, J.D. & Russell, D.W. Properties of microsomal 3-hydroxy-3-methylglutaryl coenzyme A reductase from *Pisum sativum* seedlings. *Arch. Biochem. Biophys.* **167**, 723-729 (1975).
-

-
100. Bach, T.J., Rogers, D.H. & Rudney, H. Detergent-solubilization, purification, and characterization of membrane-bound 3-hydroxy-3-methylglutaryl-coenzyme A reductase from radish seedlings. *Eur. J. Biochem.* **154**, 103-111 (1986).
 101. Narita, J.O. & Gruissem, W. Tomato hydroxymethylglutaryl-CoA reductase is required early in fruit development but not during ripening. *Plant Cell* **1**, 181-190 (1989).
 102. Oba, K., Kondo, K., Doke, N. & Uritani, I. Induction of 3-Hydroxy-3-methylglutaryl CoA Reductase in Potato Tubers after Slicing, Fungal Infection or Chemical Treatment, and Some Properties of the Enzyme E. *Plant and Cell Physiol.* **26**, 873-880 (1985).
 103. Aoyagi, K., Beyou, A., Moon, K., Fang, L. & Ulrich, T. Isolation and characterization of cDNAs encoding wheat 3-hydroxy-3-methylglutaryl coenzyme A reductase. *Plant Physiol.* **102**, 623-628 (1993).
 104. Garciapae.E, Aragon, M.C., Suarez, M.D. & Mayor, F. Mevalonate phosphorylation in *Agave americana*. *Phytochemistry* **11**, 2495-2498 (1972).
 105. Garciapae.E, Suarez, M.D. & Mayor, F. Isolation of two fractions with mevalonate kinase activity from *Pinus pinaster* and *Agave americana*. *FEBS Lett.* **30**, 15-17 (1973).
 106. Lluch, M.A., Masferrer, A., Arro, M., Boronat, A. & Ferrer, A. Molecular cloning and expression analysis of the mevalonate kinase gene from *Arabidopsis thaliana*. *Plant Mol. Biol.* **42**, 365-376 (2000).
 107. Riou, C., Tourte, Y., Lacroute, F. & Karst, F. Isolation and characterization of a cDNA encoding *Arabidopsis thaliana* mevalonate kinase by genetic complementation in yeast. *Gene* **148**, 293-297 (1994).
 108. Williamson, I. P., & Kekwick, R. G. O. The formation of 5-phosphomevalonate by mevalonate kinase in *Hevea brasiliensis* latex. *Biochem J.* **96**, 862-871 (1965).
 109. Tang, W. & Newton, R.J. Mevalonate kinase activity during different stages of plant regeneration from nodular callus cultures in white pine (*Pinus strobus*). *Tree Physiol.* **26**, 195-200 (2006).
 110. Suarez, D., Garciapae.E & Mayor, F. Mevalonate kinase from *Pinus pinaster* seedlings. *Phytochemistry* **13**, 1059-1063 (1974).
 111. Gray, J.C. & Kekwick, R.G.O. Mevalonate kinase from etiolated cotyledons of French beans. *Biochem. J.* **113**, 113-142 (1969).
 112. Schulte, A.E., van der Heijden, R. & Verpoorte, R. Microplate enzyme-coupled assays of mevalonate and phosphomevalonate kinase from *Catharanthus roseus* suspension cultured cells. *Anal. Biochem.* **269**, 245-254 (1999).
 113. Skilleter, D. N., I. P. Williamson, and G. O. Kekwick. Phosphomevalonate Kinase from *Hevea brasiliensis* latex. *Biochem. J.* **98**, P27 (1966).
 114. Bianchini, G.M., Stermer, B.A. & Paiva, N.L. Induction of early mevalonate pathway enzymes and biosynthesis of end products in potato (*Solanum tuberosum*) tubers by wounding and elicitation. *Phytochemistry* **42**, 1563-1571 (1996).
 115. Cordier, H., Karst, F. & Berges, T. Heterologous expression in *Saccharomyces cerevisiae* of an *Arabidopsis thaliana* cDNA encoding mevalonate diphosphate decarboxylase. *Plant Mol. Biol.* **39**, 953-967 (1999).
-

-
116. Skilleter, D. N., & Kekwick, R. G. O. Some Characteristics of Pyrophosphomevalonate Decarboxylase from *Hevea brasiliensis* Latex. *Biochem. J.* **108**, P11 (1968).
117. Chahed, K., Oudin, A., Guivarc'h, N., Hamdi, S., Chénieux, J.C., Rideau, M. & Clastre, M. L-Deoxy-D-xylulose 5-phosphate synthase from periwinkle: cDNA identification and induced gene expression in terpenoid indole alkaloid-producing cells. *Plant Physiol. Biochem.* **38**, 559-566 (2000).
118. Takahashi, S., Kuzuyama, T., Watanabe, H. & Seto, H. A 1-deoxy-D-xylulose 5-phosphate reductoisomerase catalyzing the formation of 2-C-methyl-D-erythritol 4-phosphate in an alternative nonmevalonate pathway for terpenoid biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9879-9884 (1998).
119. Rohdich, F. Cytidine 5'-triphosphate-dependent biosynthesis of isoprenoids: YgbP protein of *Escherichia coli* catalyzes the formation of 4-diphosphocytidyl-2-C-methylerythritol. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11758-11763 (1999).
120. Lüttgen, H., Rohdich, F., Herz, S., Wungsintaweekul, J., Hecht, S., Schuhr, C.A., Fellermeier, M., Sagner, S., Zenk, M.H., Bacher, A. & Eisenreich, W. Biosynthesis of terpenoids: YchB protein of *Escherichia coli* phosphorylates the 2-hydroxy group of 4-diphosphocytidyl-2C-methyl-D-erythritol. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1062-1067 (2000).
121. Herz, S., Wungsintaweekul, J., Schuhr, C.A., Hecht, S., Lüttgen, H., Sagner, S., Fellermeier, M., Eisenreich, W., Zenk, M.H., Bacher, A. & Rohdich, F. Biosynthesis of terpenoids: YgbB protein converts 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate to 2C-methyl-D-erythritol 2,4-cyclodiphosphate. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2486-2490 (2000).
122. Contin, A., van der Heijden, R., Lefeber, A.W.M. & Verpoorte, R. The iridoid glucoside secologanin is derived from the novel triose phosphate/pyruvate pathway in a *Catharanthus roseus* cell culture. *FEBS Lett.* **434**, 413-416 (1998).
123. Araki, N., Kusumi, K., Masamoto, K., Niwa, Y. & Iba, K. Temperature-sensitive *Arabidopsis* mutant defective in 1-deoxy-D-xylulose 5-phosphate synthase within the plastid non-mevalonate pathway of isoprenoid biosynthesis. *Physiol. Plant.* **108**, 19-24 (2000).
124. Estévez, J.M., Cantero, A., Romero, C., Kawaide, H., Jiménez, L.F., Kuzuyama, T., Seto, H., Kamiya, Y. & León, P. Analysis of the expression of CLA1, a gene that encodes the 1-deoxyxylulose 5-phosphate synthase of the 2-C-methyl-D-erythritol-4-phosphate pathway in *Arabidopsis*. *Plant Physiol.* **124**, 95-103 (2000).
125. Bouvier, F., d'Harlingue, A., Suire, C., Backhaus, R.A. & Camara, B. Dedicated roles of plastid transketolases during the early onset of isoprenoid biogenesis in pepper fruits. *Plant Physiol.* **117**, 1423-1431 (1998).
126. Lois, L.M., Rodriguez-Concepcion, M., Gallego, F., Campos, N. & Boronat, A. Carotenoid biosynthesis during tomato fruit development: regulatory role of 1-deoxy-D-xylulose 5-phosphate synthase. *Plant J.* **22**, 503-513 (2000).
127. Lange, B.M., Wildung, M.R., McCaskill, D. & Croteau, R. A family of transketolases that directs isoprenoid biosynthesis via a mevalonate-independent pathway. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2100-2104 (1998).
-

128. Walter, M.H., Hans, J. & Strack, D. Two distantly related genes encoding 1-deoxy-D-xylulose 5-phosphate synthases: differential regulation in shoots and apocarotenoid-accumulating mycorrhizal roots. *Plant J.* **31**, 243-254 (2002).
129. Carretero-Paulet, L., Ahumada, I., Cunillera, N., Rodríguez-Concepción, M., Ferrer, A., Boronat, A. & Campos, N. Expression and molecular analysis of the Arabidopsis DXR gene encoding 1-deoxy-D-xylulose 5-phosphate reductoisomerase, the first committed enzyme of the 2-C-methyl-D-erythritol 4-phosphate pathway. *Plant Physiol.* **129**, 1581-1591 (2002).
130. Schwender, J., Muller, C., Zeidler, J. & Lichtenthaler, H.K. Cloning and heterologous expression of a cDNA encoding 1-deoxy-D-xylulose-5-phosphate reductoisomerase of Arabidopsis thaliana. *FEBS Lett.* **455**, 140-144 (1999).
131. Veau, B., Courtois, M., Oudin, A., Chénieux, J.C., Rideau, M. & Clastre, M. Cloning and expression of cDNAs encoding two enzymes of the MEP pathway in Catharanthus roseus. *Biochim. Biophys. Acta* **1517**, 159-163 (2000).
132. Mahmoud, S.S. & Croteau, R.B. Metabolic engineering of essential oil yield and composition in mint by altering expression of deoxyxylulose phosphate reductoisomerase and menthofuran synthase. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8915-8920 (2001).
133. Rohdich, F., Wungsintaweekul, J., Eisenreich, W., Richter, G., Schuhr, C.A., Hecht, S., Zenk, M.H. & Bacher, A.. Biosynthesis of terpenoids: 4-Diphosphocytidyl-2C-methyl-D-erythritol synthase of Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6451-6456 (2000).
134. Lange, B.M. & Croteau, R. Isopentenyl diphosphate biosynthesis via a mevalonate-independent pathway: Isopentenyl monophosphate kinase catalyzes the terminal enzymatic step. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13714-13719 (1999).
135. Rohdich, F. et al. Biosynthesis of terpenoids: 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase from tomato. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8251-8256 (2000).
136. Fellermeier, M., Raschke, M., Sagner, S., Wungsintaweekul, J., Schuhr, C.A., Hecht, S., Kis, K., Radykewicz, T., Adam, P., Rohdich, F. & Eisenreich, W. Studies on the nonmevalonate pathway of terpene biosynthesis - The role of 2C-methyl-D-erythritol 2,4-cyclodiphosphate in plants. *Eur. J. Biochem.* **268**, 6302-6310 (2001).
137. Fellermeier, M., Kis, K., Sagner, S., Maier, U., Bacher, A. & Zenk, M.H. Cell-free conversion of 1-deoxy-D-xylulose 5-phosphate and 2-C-methyl-D-erythritol 4-phosphate into beta-carotene in higher plants and its inhibition by fosmidomycin. *Tetrahedron Lett.* **40**, 2743-2746 (1999).
138. Hecht, S., Eisenreich, W., Adam, P., Amslinger, S., Kis, K., Bacher, A., Arigoni, D. & Rohdich, F. Studies on the nonmevalonate pathway to terpenes: The role of the GcpE (IspG) protein. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14837-14842 (2001).
139. Adam, P., Hecht, S., Eisenreich, W., Kaiser, J., Gräwert, T., Arigoni, D., Bacher, A. & Rohdich, F. Biosynthesis of terpenes: Studies on 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12108-12113 (2002).
140. Warhurst, D.C. The quinine-haemin interaction and its relationship to antimalarial activity. *Biochem. Pharmacol.* **30**, 3323-3327 (1981).

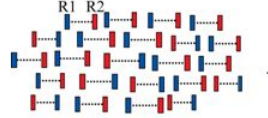
141. Barlow, O.W. & Frye, E.F. The antiasthmatic efficiency of epinephrine ephedrine and atropine - Their comparative effects on a series of experimental attacks in a subject with a complex type of asthma. *Arch. Int. Med.* **45**, 538-545 (1930).
142. Kittakoop, P., Mahidol, C. & Ruchirawat, S. Alkaloids as Important Scaffolds in Therapeutic Drugs for the Treatments of Cancer, Tuberculosis, and Smoking Cessation. *Curr. Top. Med. Chem.* **14**, 239-252 (2014).
143. Russo, P., Frustaci, A., Del Bufalo, A., Fini, M. & Cesario, A. Multitarget Drugs of Plants Origin Acting on Alzheimer's Disease. *Curr. Med. Chem.* **20**, 1686-1693 (2013).
144. Caravaggi, A.M., Sardi, A., Baldoli, E., Difrancesco, G.F. & Luca, C. Hemodynamic profile of a new cerebral vasodilator, vincamine and of one of its derivatives, apovincaminic acid ethylester (RGH-4405). *Arch. Int. Pharmacodyn. Ther.* **226**, 139-148 (1977).
145. Hollman, A. Plants in Cardiology: Quinine and Quinidine. *Br. Heart J.* **66**, 301-301 (1991).
146. Sinatra, R.S., Jahr, J.S. & Watkins-Pitchford, M.J. *The Essence of Analgesia and Analgesics*, (Cambridge University Press, 2010).
147. Cushnie, T.P.T., Cushnie, B. & Lamb, A.J. Alkaloids: An overview of their antibacterial, antibiotic-enhancing and antivirulence activities. *Int. J. Antimicrob. Agents* **44**, 377-386 (2014).
148. Shi, Q.I.U., Hui, S.U.N., Zhang, A.H., Hong-Ying, X.U., Guang-Li, Y.A.N., Ying, H.A.N. & Xi-Jun, W.A.N.G. Natural alkaloids: basic aspects, biological roles, and future perspectives. *Chin. J. Nat. Med.* **12**, 401-406 (2014).
149. Plemenkov, V. V. "Introduction to the chemistry of natural compounds." *Kazan* 242 (2001).
150. Gournelis, D.C., Laskaris, G.G. & Verpoorte, R. Cyclopeptide alkaloids. *Nat. Prod. Rep.* **14.1**, 75-82 (1997).
151. Rai, A., Smita, S.S., Singh, A.K., Shanker, K. & Nagegowda, D.A. Heteromeric and Homomeric Geranyl Diphosphate Synthases from *Catharanthus roseus* and Their Role in Monoterpene Indole Alkaloid Biosynthesis. *Mol. Plant* **6**, 1531-1549 (2013).
152. Simkin, A.J., Miettinen, K., Claudel, P., Burlat, V., Guirimand, G., Courdavault, V., Papon, N., Meyer, S., Godet, S., St-Pierre, B. & Giglioli-Guivarc'h, N. Characterization of the plastidial geraniol synthase from Madagascar periwinkle which initiates the monoterpene branch of the alkaloid pathway in internal phloem associated parenchyma. *Phytochemistry* **85**, 36-43 (2013).
153. Collu, G., Unver, N., Peltenburg-Looman, A.M., van der Heijden, R., Verpoorte, R. & Memelink, J. Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Lett.* **508**, 215-220 (2001).
154. Herbert, R.B. The biosynthesis of plant alkaloids and nitrogenous microbial metabolites. *Nat. Prod. Rep.* **2**, 163-179 (1985).
155. Uesato, S., Matsuda, S., Iida, A., Inouye, H. & Zenk, M.H. Intermediacy of 10-hydroxygeraniol, 10-hydroxynerol and iridodial in the biosynthesis of ajmaline and vomilenine in *Rauwolfia serpentina* suspension cultures. *Chem. Pharm. Bull.* **32**, 3764-3767 (1984).

-
156. Uesato, S., Matsuda, S. & Inouye, H. Mechanism for iridane skeleton formation from acyclic monoterpenes in the biosynthesis of secologanin and vindoline in *Catharanthus roseus* and *Lonicera morrowii*. *Chem. Pharm. Bull.* **32**, 1671-1674 (1984).
 157. Uesato, S., Ueda, S., Kobayashi, K. & Inouye, H. Mechanism of iridane skeleton formation in the biosynthesis of iridoid glucosides in *Gardenia jasminoides* cell cultures. *Chem. Pharm. Bull.* **31**, 4185-4188 (1983).
 158. Geu-Flores, F., Sherden, N.H., Courdavault, V., Burlat, V., Glenn, W.S., Wu, C., Nims, E., Cui, Y. & O'Connor, S.E. An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature* **492**, 138-142 (2012).
 159. Yamamoto, H., Katano, N., Ooi, A. & Inoue, K. Secologanin synthase which catalyzes the oxidative cleavage of loganin into secologanin is a cytochrome P450. *Phytochemistry* **53**, 7-12 (2000).
 160. Guillet, G., Poupart, J., Basurco, J. & De Luca, V. Expression of tryptophan decarboxylase and tyrosine decarboxylase genes in tobacco results in altered biochemical and physiological phenotypes. *Plant Physiol.* **122**, 933-943 (2000).
 161. Pasquali, G. et al. Coordinated regulation of two indole alkaloid biosynthetic genes from *Catharanthus roseus* by auxin and elicitors. *Plant Mol. Biol.* **18**, 1121-1131 (1992).
 162. Facchini, P.J. Alkaloid biosynthesis in plants: Biochemistry, cell biology, molecular regulation, and metabolic engineering applications. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**, 29-66 (2001).
 163. Luijendijk, T.J.C., Nowak, A. & Verpoorte, R. Strictosidine glucosidase from suspension cultured cells of *Tabernaemontana divaricata*. *Phytochemistry* **41**, 1451-1456 (1996).
 164. esseau, S., Kellner, F., Lanoue, A., Thamm, A.M., Salim, V., Schneider, B., Geu-Flores, F., Höfer, R., Guirimand, G., Guihur, A. and Oudin, A. A Pair of Tabersonine 16-Hydroxylases Initiates the Synthesis of Vindoline in an Organ-Dependent Manner in *Catharanthus roseus*. *Plant Physiol.* **163**, 1792-1803 (2013).
 165. Guirimand, G., Guihur, A., Poutrain, P., Héricourt, F., Mahroug, S., St-Pierre, B., Burlat, V. & Courdavault, V. Spatial organization of the vindoline biosynthetic pathway in *Catharanthus roseus*. *J. Plant Physiol.* **168**, 549-557 (2011).
 166. Schröder, G., Unterbusch, E., Kaltenbach, M., Schmidt, J., Strack, D., De Luca, V. & Schröder, J. Light-induced cytochrome P450-dependent enzyme in indole alkaloid biosynthesis: tabersonine 16-hydroxylase. *FEBS Lett.* **458**, 97-102 (1999).
 167. Stpierre, B. & Deluca, V. A A Cytochrome P-450 Monooxygenase Catalyzes the First Step in the Conversion of Tabersonine to Vindoline in *Catharanthus roseus*. *Plant Physiol.* **109**, 131-139 (1995).
 168. Deluca, V., Balsevich, J., Tyler, R.T. & Kurz, W.G.W. Characterization of a novel N-methyltransferase (NMT) from *Catharanthus roseus* plants : Detection of NMT and other enzymes of the indole alkaloid biosynthetic pathway in different cell suspension culture systems. *Plant Cell Rep.* **6**, 458-461 (1987).
-

-
169. Decarolis, E. & Deluca, V. Purification, characterization, and kinetic analysis of a 2-oxoglutarate-dependent dioxygenase involved in vindoline biosynthesis from *Catharanthus roseus*. *J. Biol.Chem.* **268**, 5504-5511 (1993).
170. Magnotta, M., Murata, J., Chen, J. & De Luca, V. Expression of deacetylvindoline-4-O-acetyltransferase in *Catharanthus roseus* hairy roots. *Phytochemistry* **68**, 1922-1931 (2007).
171. Sottomayor, M., Lopez-Serrano, M., DiCosmo, F. & Barcelo, A.R. Purification and characterization of alpha-3',4'-anhydrovinblastine synthase (peroxidase-like) from *Catharanthus roseus* (L.) G. Don. *FEBS Lett.* **428**, 299-303 (1998).
172. Chevallier, A. *The encyclopedia of medicinal plants*, (London: Dorling Kindersley, 1996).
173. Dwyer, J. & Rattray, D. *Magic and medicine of plants*, (Reader's Digest Association, Inc. Pleasantville, New York, 1986).
174. Singh, S.N., Vats, P., Suri, S., Shyam, R., Kumria, M.M.L., Ranganathan, S. & Sridharan, K. Effect of an antidiabetic extract of *Catharanthus roseus* on enzymic activities in streptozotocin induced diabetic rats. *J. Ethnopharmacol.* **76**, 269-277 (2001).
175. Duke, J.A. Medicinal Plants. *Science* **229**, 1036-1036 (1985).
176. Nammi, S., Boini, M.K., Lodagala, S.D. & Behara, R.B.S. The juice of fresh leaves of *Catharanthus roseus* Linn. reduces blood glucose in normal and alloxan diabetic rabbits. *BMC Complement. Altern. Med.* **3**, 4-4 (2003).
177. Vega-Ávila, E., Cano-Velasco, J.L., Alarcón-Aguilar, F.J., Fajardo Ortíz, M.D.C., Almanza-Pérez, J.C. & Román-Ramos, R. Hypoglycemic Activity of Aqueous Extracts from *Catharanthus roseus*. *Evid. Based Complement. Alternat. Med.: eCAM* **2012**, 934258-934258.
178. Wang, S., Zheng, Z., Weng, Y., Yu, Y., Zhang, D., Fan, W., Dai, R. & Hu, Z. Angiogenesis and anti-angiogenesis activity of Chinese medicinal herbal extracts. *Life Sci.* **74**, 2467-2478 (2004).
179. Panneerselvam, C., Murugan, K., Kovendan, K., Kumar, P.M., Ponarulselvam, S., Amerasan, D., Subramaniam, J. & Hwang, J.S. Larvicidal efficacy of *Catharanthus roseus* Linn. (Family: *Apocynaceae*) leaf extract and bacterial insecticide *Bacillus thuringiensis* against *Anopheles stephensi* Liston. *Asian Pac. J. Trop. Dis.* **6**, 847-853 (2013).
180. Wong, S.K., Lim, Y.Y., Abdullah, N.R. & Nordin, F.J. Antiproliferative and phytochemical analyses of leaf extracts of ten *Apocynaceae* species. *Pharmacognosy Res.* **3**, 100-106 (2011).
181. Sutarno, H. & Rudjiman. *Catharanthus roseus* (L.) G. Don. In. *Plant Resources of South-East Asia* **12**, 185-190 (1999).
182. van der Heijden, R., Jacobs, D.I., Snoeijer, W., Hallared, D. & Verpoorte, R. The *Catharanthus* alkaloids: Pharmacognosy and biotechnology. *Curr. Med. Chem.* **11**, 607-628 (2004).
183. Shurin, G.V., Tourkova, I.L., Kaneno, R. & Shurin, M.R. Chemotherapeutic Agents in Noncytotoxic Concentrations Increase Antigen Presentation by Dendritic Cells via an IL-12-Dependent Mechanism. *J. Immunol.* **183**, 137-144 (2009).
-

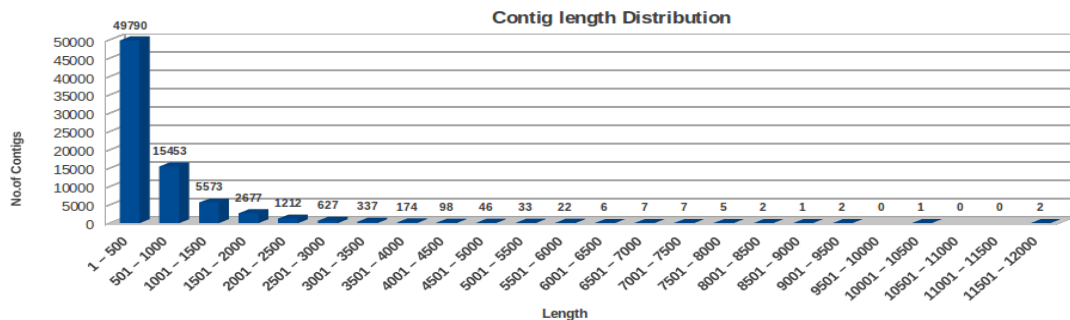
-
184. Jordan, M.A., Thrower, D. & Wilson, L. Mechanism of inhibition of cell-proliferation by vinca alkaloids. *Cancer Res.* **51**, 2212-2222 (1991).
 185. Gary A. Thompson & Bel, A.J.E.v. *Phloem: Molecular Cell Biology, Systemic Communication, Biotic Interactions*, 368 (John Wiley & Sons, 2012).
 186. Barthe, L., Ribet, J.P., Péliou, M., Degude, M.J., Fahy, J. & Duflos, A. Optimization of the separation of Vinca alkaloids by nonaqueous capillary electrophoresis. *J. Chromatogr. A* **968**, 241-250 (2002).
 187. Lin, A.J., Klayman, D.L. & Milhous, W.K. Antimalarial activity of new water-soluble dihydroartemisinin derivatives. 3. Aromatic amine analogues. *J. Med. Chem.***30**, 2147-2150 (1987).
 188. Duke, M.V., Paul, R.N., Elshohly, H.N., Sturtz, G. & Duke, S.O. Localization of Artemisinin and Artemisitene in Foliar Tissues of Glanded and Glandless Biotypes of *Artemisia annua* L. *Int. J. Plant Sci.* **155**, 365-372 (1994).
 189. Avery, M.A., Chong, W.K.M. & Jenningswhite, C. Stereoselective total synthesis of (+)-artemisinin, the antimalarial constituent of *Artemisia annua* L. *J. Am. Chem. Soc.* **114**, 974-979 (1992).
 190. Roth, R.J. & Acton, N. A Simple Conversion of Artemisinic Acid into Artemisinin. *J. Nat. Prod.* **52**, 1183-1185 (1989).
 191. Paddon, C.J., Westfall, P.J., Pitera, D.J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D. & Polichuk, D.R. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496**, 528-532 (2013).
 192. Wall, M.E. & Wani, M.C. Paclitaxel - from discovery to clinic. *Taxane Anticancer Agents: Basic Science and Current Status*, **583**, 18-30 (1995).
 193. Goldspiel, B.R. Clinical overview of the taxanes. *Pharmacotherapy* **17**, S110-S125 (1997).
 194. Malik, S., Cusidó, R.M., Mirjalili, M.H., Moyano, E., Palazón, J. & Bonfill, M. Production of the anticancer drug taxol in *Taxus baccata* suspension cultures: A review. *Process Biochem.* **46**, 23-34 (2011).
 195. Borman, S.T.U. Total Synthesis of Anticancer Agent Taxol Achieved by Two Different Routes. *Chem. Eng. News* **72**, 32-34 (1994).
 196. Wuts, P.G. Semisynthesis of Taxol. *Curr. Opin. Drug Disc.* **1**, 329-37 (1998).
 197. Kingston, D.G.I. Recent Advances in the Chemistry of Taxol. *J. Nat. Prod.* **63**, 726-734 (2000).
 198. Cragg, G.M., Schepartz, S.A., Suffness, M. & Grever, M.R. The Taxol supply crisis. New NCI policies for handling the large-scale production of novel natural product anticancer and anti-HIV agents. *J. Nat. Prod.* **56**, 1657-1668 (1993).
 199. Ketchum, R.E.B. & Gibson, D.M. Paclitaxel production in suspension cell cultures of *Taxus*. *Plant Cell Tiss. Org.* **46**, 9-16 (1996).
 200. Ketchum, R.E.B., Gibson, D.M., Croteau, R.B. & Shuler, M.L. The kinetics of taxoid accumulation in cell suspension cultures of *Taxus* following elicitation with methyl jasmonate. *Biotechnol. Bioeng.* **62**, 97-105 (1999).
 201. Nicolaou, K.C., Yang, Z., Liu, J.J., Ueno, H., Nantermet, P.G., Guy, R.K., Claiborne, C.F., Renaud, J., Couladouros, E.A., Paulvannan, K. & Sorensen, E.J. Total synthesis of taxol. *Nature* **367**, 630-634 (1994).
-

-
202. Hays, S.B. Some effects of reserpine, a tranquilizer, on House Fly. *J. Econ. Entomol.* **58**, 782-& (1965).
 203. Yui, T. & Takeo, Y. Neuropharmacological studies on a new series of ergot alkaloids; elymoclavine as a potent analeptic on reserpine-sedation. *Jpn. J. Pharmacol.* **7**, 157-161 (1958).
 204. Yui, T. & Takeo, Y. Neuropharmacological studies on a new series of ergot alkaloids; the effects on electrocorticogram of rabbits. *Jpn. J. Pharmacol.* **7**, 162-168 (1958).
 205. Govindachari, T. R., & Viswanathan, N. Alkaloids of *Mappia foetida*. *Phytochemistry* **2**, 3529-3531 (1972).
 206. Staines, H.M. & Krishna, S. Treatment and Prevention of Malaria : Antimalarial Drug Chemistry, Action and Use, *Springer Science & Business Media* (2012).
 207. Heckendorf, A.H. & Hutchinson, C.R. Biosynthesis of camptothecin. II. Confirmation that isovincoside, not vincoside, is the penultimate biosynthetic precursor of indole alkaloids. *Tetrahedron Lett.* 4153-4154 (1977).
 208. Ma, X., Koepke, J., Bayer, A., Linhard, V., Fritzsche, G., Zhang, B., Michel, H. & Stöckigt, J. Vinorine synthase from *Rauvolfia*: the first example of crystallization and preliminary X-ray diffraction analysis of an enzyme of the BAHD superfamily. *Biochim. Biophys. Acta* **1701**, 129-132 (2004).
 209. Sharma, R.K. Concise textbook of forensic medicine and toxicology, *J Punjab Acad. Forensic Med. Toxicol.* **5**, 54-54 (2005).



Chapter 2

Transcriptome sequencing, *de novo* assembly and functional annotation to gain insights for the secologanin biosynthetic pathway



Chapter 2

Transcriptome sequencing, *de novo* assembly and functional annotation to gain insights for the secologanin biosynthetic pathway

RNA-sequencing (RNA-seq) or Transcriptome sequencing is a technique that employs next-generation DNA-sequencing technology to simultaneously sequence all of the RNA transcripts in a cell. It is a quick and cost effective method to profile complete gene space of any organism due to its high throughput accuracy and reproducibility. Isolation of total RNA from stem, leaves, roots and flowers of *Catharanthus roseus* posed certain degradation problems, which were overcome by adopting a method for RNA isolation, which is an extensively modified version of an earlier reported procedure. The total RNA, thus obtained from each tissue was mixed in equal ratio and used for transcriptome sequencing using Illumina GAI Analyzer. Following *de novo* assembly, further screening for identification of transcripts involved in secologanin biosynthesis was carried out by annotation of the assembled transcriptome data.

2.1 Introduction

Catharanthus roseus is an evergreen shrub belonging to the family *Apocynaceae*. It is a rich source of the monoterpene indole alkaloids (MIAs), many of which possess magnificent pharmacological and therapeutic interest such as the tranquillizer reserpine¹ or the ergot alkaloids² employed for their relaxing and migraine relieving properties, the biggest breakthrough being the discovery of vincristine and vinblastine as anti-cancer drugs^{3,4}. These MIAs were synthesized from the condensation of tryptamine and the iridoid monoterpene, secologanin. The monoterpene alcohols themselves possess chemopreventive activities against various forms of cancer. For example, perillyl alcohol, formed by the hydroxylation of limonene, has been proved to cause apoptosis of liver tumors in rats⁵. Also, the acyclic monoterpene alcohol, geraniol, possesses anti-tumor activity against rat and mice hepatoma, leukemia and melanoma cells⁶.

Isolation of total RNA from the stem, leaf and root tissues of *C. roseus* white-flowering and pink-flowering plants was carried out primarily using various kits, but posed several problems in the form of degradation and genomic DNA contamination. The quality of RNA is very important for downstream processing so as to avoid interference in reverse transcription, as well as in PCR reactions. In this study, a protocol for the isolation of high quality total RNA from the leaf, stem and root tissues of *C. roseus* was used, with extensive modifications of the protocol reported for isolation of high quality of RNA from gymnosperm and angiosperm tissue⁷. This protocol was developed earlier in our lab for the isolation of high quality total RNA from the interface of heartwood and sapwood of Indian Sandalwood, *Santalum album*⁸. Later on, high quality total RNA could be obtained using Spectrum™ Plant Total RNA Isolation Kit from Sigma-Aldrich, USA, also, by including a few modifications.

Transcriptome sequencing is a powerful technique for profiling transcriptome because of its high-throughput accuracy and reproducibility. In plants, high-throughput RNA sequencing has accelerated the discovery of novel genes, transcription pattern, and functional analysis. In the present study, we have performed RNA sequencing on Illumina GAI Analyzer, and used for screening of unigenes involved in the biosynthesis

of secologanin in *Catharanthus roseus*. The raw RNA-seq paired reads from stem, root and leaf RNA have been assembled using various assembly tools such as Velvet_1.1.05⁹ and Oases¹⁰. Various approaches for functional annotation of the assembled transcripts have been used to identify the unigenes, which are involved in MIA biosynthesis in *C. roseus*.

2.2 Materials and Methods

2.2.1 Plant Sample Source

The various plant tissues such as leaves, stem and roots were collected from *C. roseus* white-flowering and pink-flowering plants of height 26 cm above the ground, grown in a greenhouse (temperature: 25 °C, humidity: 40 % relative humidity). The plants were grown in shade with a 12-hour light and dark period. The tissues were flash frozen in liquid nitrogen and stored at -80 °C till further use.

2.2.2 Reagents

All the reagents and chemicals were procured from Sigma-Aldrich or Invitrogen (Life Technologies/ Thermo-Scientific), USA. All the solutions were prepared in 0.1 % diethylpyrocarbonate (DEPC) and sterilized in an autoclave, except Tris buffer, which was prepared in sterile DEPC treated water. All the plastic and glasswares were soaked in freshly prepared DEPC water, dried, and sterilized in the autoclave before use.

Extraction buffer

- ◆ Lithium dodecyl sulphate (1.5 %), Lithium Chloride (300 mM), EDTA disodium salt (20 mM), Sodium Deoxycholate (1 % w/v), Tergitol nonidet NP-40 (1 % v/v). The buffer containing these constituents was autoclaved and the Tris-HCl (200 mM) pH 8.5 was added thereafter and following constituents were added just before use: Thiourea (5 mM), Aurintricarboxylic acid (1 mM), and DTT (10 mM).
- ◆ PVPP (Polyvinylpolypyrrolidone)
- ◆ Chloroform/Isoamylalcohol (24:1; v/v)
- ◆ 3.3 M Sodium Acetate pH 6.1 (prepared in DEPC treated water and autoclaved)
- ◆ 5 M NaCl (prepared in DEPC treated water and autoclaved)

-
- ◆ 10 % CTAB (prepared in DEPC treated water and autoclaved)
 - ◆ Isopropanol
 - ◆ TE: 10 mM Tris-HCl; 1 mM EDTA; pH 8.0
 - ◆ 10 M LiCl (prepared in DEPC treated water and autoclaved)
 - ◆ 70 % Ethanol (prepared with DEPC treated autoclaved water)
 - ◆ DEPC treated water (DEPC from Sigma-Aldrich)
 - ◆ Centrifuges
 - ◆ Water bath (wet and dry)

2.2.3 Modified protocol for RNA isolation

1. 1 g of each plant tissue was ground in liquid nitrogen in presence of 0.3 g of PVPP per gram of tissue in bead-beater at a frequency of 30 vibrations/second for 5 minutes.
2. 20 mL extraction buffer/g of tissue was added.
3. This was vortexed at 25-30 °C for 10 minutes.
4. This was centrifuged at $3000 \times g$ for 20 minutes at 4 °C.
5. The supernatant was transferred into a fresh tube and kept on ice.
6. To this, 2 mL 10 % CTAB solution was added at 25-30 °C and incubated for 5 minutes at 60 °C, to remove the residual polysaccharides.
7. This mixture was extracted with Chloroform: Isoamyl alcohol (24:1) till the interface became clear and the supernatant was retrieved each time.
8. The supernatant was transferred into another tube and to this, 1/9th volume of 3.3 M Sodium acetate and 0.6 volume of ice-cold isopropanol were added.
9. This was incubated at -20 °C for 3 h and then centrifuged at $14000 \times g$ for 20 minutes at 4 °C and the supernatant was discarded.
10. To the pellet, 1 mL of TE Buffer (10 mM) and 1 mL of 5M NaCl were added and incubated on ice for 30 minutes, with periodic vortexing.
11. This mixture was extracted with chloroform: isoamyl alcohol (24:1) and the supernatant was retrieved each time. This was repeated till the interface became clear.

12. To the supernatant, 2.5 M final concentration of LiCl was added and incubated overnight at -20 °C.
13. The RNA was pelleted down by centrifugation at $14000 \times g$ for 30 minutes at 4 °C.
14. The resultant pellet was washed with 70 % ethanol by centrifugation at $14,000 \times g$ for 10 minutes at 4 °C.
15. The pellet was air dried at room temperature and re-suspended in 50 μ l DEPC treated water.

2.2.4 Modified protocol for Isolation of total RNA using Spectrum™ Plant Total RNA Isolation Kit from Sigma-Aldrich, USA

Initially, 0.1 g of each plant tissue was ground in liquid nitrogen with 0.3 g of PVPP per gram of tissue in bead-beater at a frequency of 30 vibrations /second for 5 minutes. Then the sample was processed as per the manufacturer's instructions. Before binding the RNA to the column, extractions were carried out using Chloroform: Isoamyl alcohol mixture, in the ratio of 24:1, to remove protein/polysaccharide/lipid contamination. Elution of RNA at the final step was carried out in DEPC-treated water.

2.2.5 Quantification of total RNA

Total RNA was quantified using a spectrophotometer (NanoDrop, Thermo Scientific) by measuring optical density of isolated RNA (re-suspended in DEPC-treated water), at wavelengths 230, 260, and 280 nm. As RNA absorbs strongly at 260 nm, its purity can be assessed by the ratios A_{260}/A_{280} and A_{260}/A_{230} , which should be over 1.8 to guarantee the absence of proteins and other organic compounds. The integrity of total RNA was assessed by sharpness of rRNA (28S and 18S rRNA) bands on a 1 % agarose gel and visualized by GelRed™ (Biotium Inc., USA).

2.2.6 RT-PCR

To test RNA quality, 5 μ g of total RNA was used for cDNA synthesis using SuperScript® III First-Strand Synthesis System for RT-PCR from Invitrogen (Life Technologies), USA, according to manufacturer's instructions. This cDNA was used in a

PCR reaction with the full length primers of the Loganic acid methyltransferase (CrLAMT) gene¹¹ as follows: Initial denaturation at 95 °C for 5 min, followed by 35 cycles of denaturation at 95 °C for 30 sec, annealing at 60 °C for 30 sec and extension at 72 °C for 1 min 20 sec and a final extension at 72 °C for 10 min. The samples were run on a 1 % agarose gel and visualized by GelRedTM (Biotium).

2.2.7 Transcriptome sequencing

Transcriptome library was constructed according to the IlluminaTruSeq RNA library protocol outlined in “TruSeq RNA Sample Preparation Guide” (Part # 15008136; Rev. A; Nov 2010) by Genotypic Pvt. Ltd. In brief, mRNA was purified from 1 µg of high quality intact total RNA (pooled as an equivalent mixture of total RNA isolated from leaves, stem and roots) using oligo dT beads (TruSeq RNASample Preparation Kit, Illumina). The purified mRNA was fragmented for 2 minutes at an elevated temperature (94 °C) in the presence of divalent cations and reverse transcribed with Superscript II Reverse transcriptase by priming with random hexamers. Second strand cDNA was synthesized in the presence of DNA polymerase I and RnaseH. The cDNA was cleaned up using Agencourt Ampure XP SPRI beads (Beckman Coulter). Illumina adapters were ligated to the cDNA molecules after end repair and addition of an ‘A’ base, and SPRI cleanup was performed after ligation. The library was amplified using 8 cycles of PCR for enrichment of adapter-ligated fragments. The prepared library was quantified using Nanodrop and validated for quality by running an aliquot on High Sensitivity Bioanalyzer Chip (Agilent).

2.2.8 Raw data processing

The raw reads generated were 150 bp paired-ends. Primary QC check of the raw data was done using the in-built tool SeqQC-V2.1. In order to obtain better quality of reads, the raw reads were filtered using adapter trimming, B trimming and Low quality trimming throughout the read sequences (Phred score <20).

2.2.9 Transcriptome assembly

The clean reads were first assembled into contigs using the Velvet_1.1.05⁹ with an optimized hash length of 49. Assembled contigs were submitted as inputs for Oasis_0.2.01¹² to generate transcripts. These transcripts were further subjected to cluster and assembly analysis using CD-HIT to remove the redundancy and generate unique unigenes (transcripts).

2.2.10 Transcriptome annotation

In order to assign molecular function, biological processes and cellular components of unigenes, functional annotation of the assembled transcriptomic data was carried out. The transcript sequences (DNA sequences) were translated into their corresponding peptide sequences (all 6 reading frames) using the online tool “Virtual Ribosome – version 1.1”¹³, which has an integrated ORF finder and a full support for the IUPAC alphabet of nucleotides, including degenerate nucleotides. The longest ORFs were selected for each of the unigenes and submitted to Pfam-A database¹⁴, to identify protein domains and architecture. The FASTA formats of all the unigenes were submitted to KEGG (Kyoto Encyclopedia of Genes and Genomes)¹⁵ database to assign KO (KEGG Orthology) numbers and to KAAS (KEGG Automatic Annotation Server) to generate KEGG pathways.

2.3 Results and Discussion

2.3.1 Optimization of protocol for isolation of total RNA from various tissues of *Catharanthus roseus*

Several standard protocols have been reported for isolation of total RNA from various plant tissues. Initially, total RNA isolation from the root, stem and leaf tissues (of white and pink flowering plants) was carried out using Spectrum™ Plant Total RNA Isolation Kit from Sigma-Aldrich, USA. Although the 28S rRNA and 18S rRNA bands were slightly visible, degradation was observed each time using this protocol (Figure 2.1).

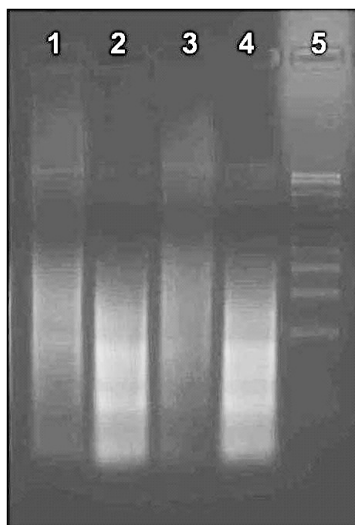


Figure 2.1: Agarose gel electrophoresis of total RNA isolated using Spectrum™ Plant Total RNA Isolation Kit from Sigma; **Lanes 1 – 2:** Total RNA isolated from leaves and stem of white-flowering plant, **Lanes 3 – 4:** Total RNA isolated from leaves and stem of pink-flowering plants, **Lane 5:** 1 Kb Plus DNA Ladder (Appendix Figure A1).

Hence, a conventional protocol standardized in our lab for the isolation of total RNA from the interface of heartwood and sapwood from Sandalwood⁸, was used with slight modifications. The initial solvent washes were eliminated, as the phenolics released by the tissues of *C. roseus* were not as high as in the case of the bark of wood. The initial crushing with PVPP was sufficient to remove the sparse phenolic content in the tissues of *C. roseus* and preventing them from degrading the RNA at a later stage. Treatment with

CTAB and heating is known to remove polysaccharide contamination, which might co-precipitate with nucleic acid and lead to degradation of RNA.

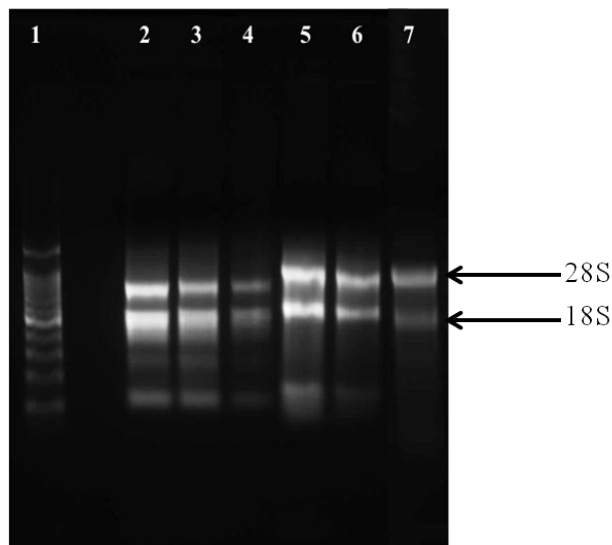


Figure 2.2: Agarose gel electrophoresis of total RNA isolated using modified conventional protocol; **Lane 1:** 1 KB Plus DNA Ladder (Appendix Figure A1), **Lanes 2 – 4:** Total RNA isolated from leaves, stem and roots of white-flowering plants, **Lanes 5 – 7:** Total RNA isolated from leaves, stem and roots of pink-flowering plants.

Total RNA extracted from the root, stem and leaf tissues of *Vinca* plant showed two distinct rRNA bands (28S rRNA and 18S rRNA) on a 1 % agarose gel electrophoresis in the ratio of 2:1, without degradation and also showed no genomic DNA contamination (Figure 2.2).

The $A_{260/280}$ ratios were found to be 1.95, 1.85, 1.96, 1.95, 2.04 and 1.97 and the $A_{260/230}$ ratios were found to be 2.17, 2.18, 2.13, 2.19, 2.13 and 2.10 respectively, for RNAs from root, stem and leaf tissues of white and pink flowering plants, indicating no contamination of protein, polysaccharides, polyphenolics, salts and solvents. Total RNA in the range of 40-130 $\mu\text{g/g}$ of root tissues, 250-400 $\mu\text{g/g}$ of stem tissues and 500-700 $\mu\text{g/g}$ of leaf tissues were obtained. To check the integrity of isolated total RNA, it was reverse transcribed as discussed earlier, and PCR amplification with primer of CrLAMT resulted in amplification of the expected ≈ 1150 bp amplicon (Figure 2.3).

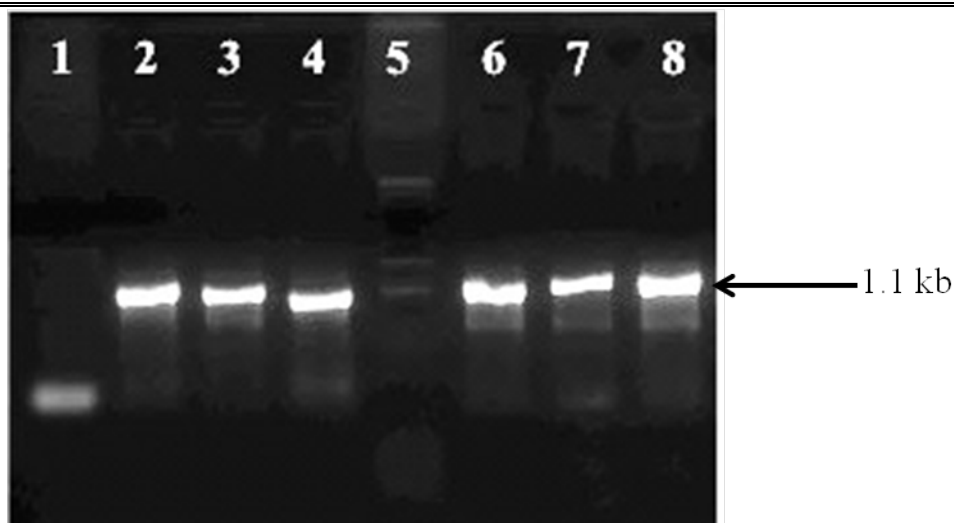


Figure 2.3: Agarose gel electrophoresis of PCR of cDNAs with CrLAMT full-length primers for cDNA verification; **Lane 1:** Negative control, **Lanes 2, 3 and 4:** PCR of cDNA from leaves, stem and roots of white-flowering plants, **Lane 5:** 1Kb DNA Ladder (Appendix Figure A1), **Lanes 6, 7 and 8:** PCR of cDNA from leaves, stem and roots of pink-flowering plants

On employing certain modifications (initial crushing with PVPP, extraction with chloroform:isoamyl alcohol::24:1 mixture and final elution in DEPC-treated water), high quality RNA could also be obtained using the SpectrumTM Plant Total RNA Isolation Kit from Sigma-Aldrich, USA (Figure 2.4).

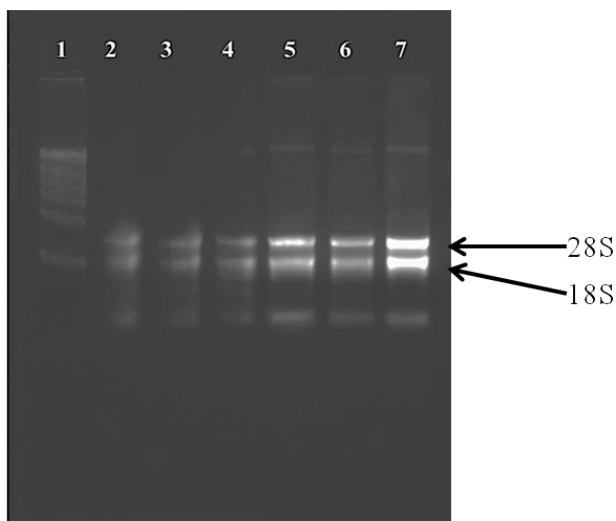


Figure 2.4: Agarose gel electrophoresis of total RNA isolated using modified kit protocol; **Lane 1:** 1 Kb DNA Ladder (Appendix Figure A1), **Lanes 2 – 4:** Total RNA isolated from root, stem and leaves of pink-flowering plants, **Lanes 5 – 7:** Total RNA isolated from roots, stem and leaves of white-flowering plants.

Similarly, the RNA samples isolated by this protocol were also checked for their integrity by construction of cDNA and PCR amplification using full-length CrLAMT primers (Figure 2.5 (a) and Figure 2.5 (b)).

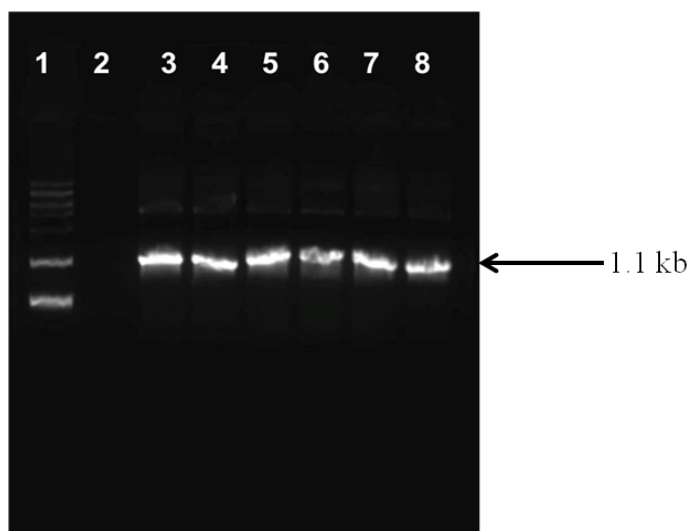


Figure 2.5: Agarose gel electrophoresis of PCR of cDNAs with CrLAMT full-length primers for cDNA verification; **Lane 1:** 1 Kb DNA Ladder (Appendix Figure A1), **Lane 2:** Negative Control, **Lane 3-8:** PCR of cDNA from leaves, stem and roots of pink-flowering plants and white-flowering plants, respectively.

In summary, the modified protocol used by us is simple and highly effective for extraction of high quality total RNA from the various tissues rich in polysaccharides and polyphenolics. The use of PVPP along with tissue crushing aided in avoiding contaminations that are usually caused during the breakage of cells. The optimized protocol enabled us to isolate high quality total RNA without degradation and contamination, which was used for downstream processing such as reverse transcription, EST library preparation and PCR amplifications.

2.3.2 Transcriptome sequencing and screening of genes involved in the biosynthesis of Secologanin in *Catharanthus roseus*

2.3.2.1 cDNA library preparation and sequencing

cDNA library was constructed from the mRNA purified from total RNA isolated from the root, leaf and stem tissues of *Catharanthus roseus* and was subjected to RNA

sequencing using Illumina GAII Analyzer. cDNA library was constructed as discussed earlier and amplified by PCR to enrich the adaptor ligated fragments. The cDNA library was sequenced on one lane of the flow cell using paired end sequencing. A total of 15485359 raw reads were generated with a length of 150 bp corresponding to 4.88 GB. Primary QC check for raw reads was performed using the inbuilt tool SeqQC-V2.1. From the observation of QC report, adapter trimming and low quality reads (Phred score <20) trimming was performed throughout the sequence to get better quality of reads.

2.3.2.2 *De novo* transcriptome assembly

A total 15485359 high quality paired end reads were generated with read length of 150 bp using deep sequencing. These high quality reads were preliminarily assembled into contigs using Velvet_1.1.05⁹ with varying hash lengths (k-mer) from 49 to 121. A Total 76075 contigs were generated with an optimized hash length of 49 with largest average contig length (548.5 bp) and an N50 length (881 bp). These contigs were given as input for Oasis_0.2.01¹² to generate 70779 transcripts having N50 length of 2485 bp and average transcript length 1594.83 bp. The total transcripts were further subjected to cluster and assembly analysis using CD-HIT¹⁶ to remove the redundancy. Finally, it resulted in a total of 62352 unique transcripts with an average size of 1024 bp and N50 length 2375 bp, which contains 28429 transcripts (45.59 %) with length greater than 1 kb and 37088 transcripts (59.48 %) with length greater than 500 bp. These results showed that the high throughput sequencing and assembly quality is good, which could be used for functional annotation.

2.3.2.3 Functional annotation of unigenes

Functional annotation mainly provides an insight into the biochemical functions, biological processing and information relating to regulation, interactions and expression of the unigenes represented in an organism. Various approaches were employed to functionally annotate the assembled transcripts based on the sequence similarity with proteins/transcripts in the publically available databases.

2.3.2.3.1 KAAS analysis

In order to identify the various terpenoid and alkaloid biosynthetic pathways present in *Catharanthus roseus*, all the 62352 unigenes were submitted to KAAS database, a web-based server, which carries out bi-directional BLAST comparisons against the manually curated KEGG GENES database and provides functional annotation of genes.

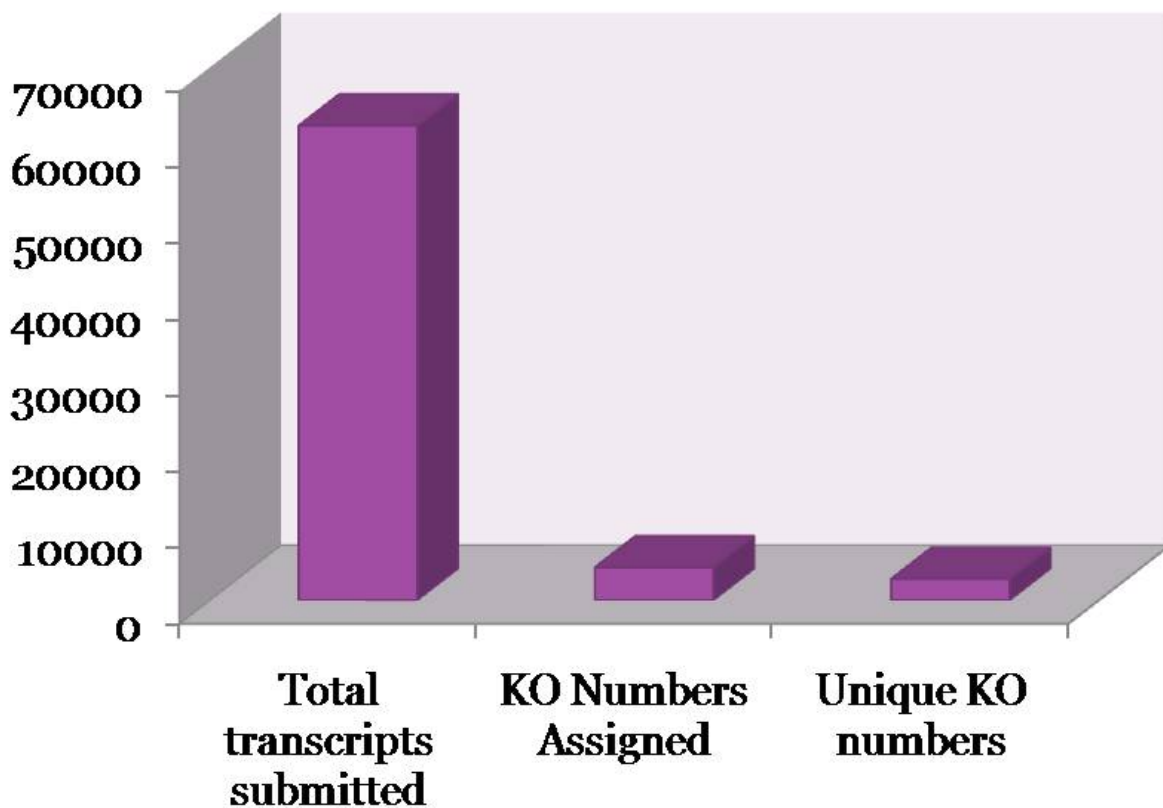


Figure 2.6: KAAS analysis of unigenes for KEGG pathway mapping

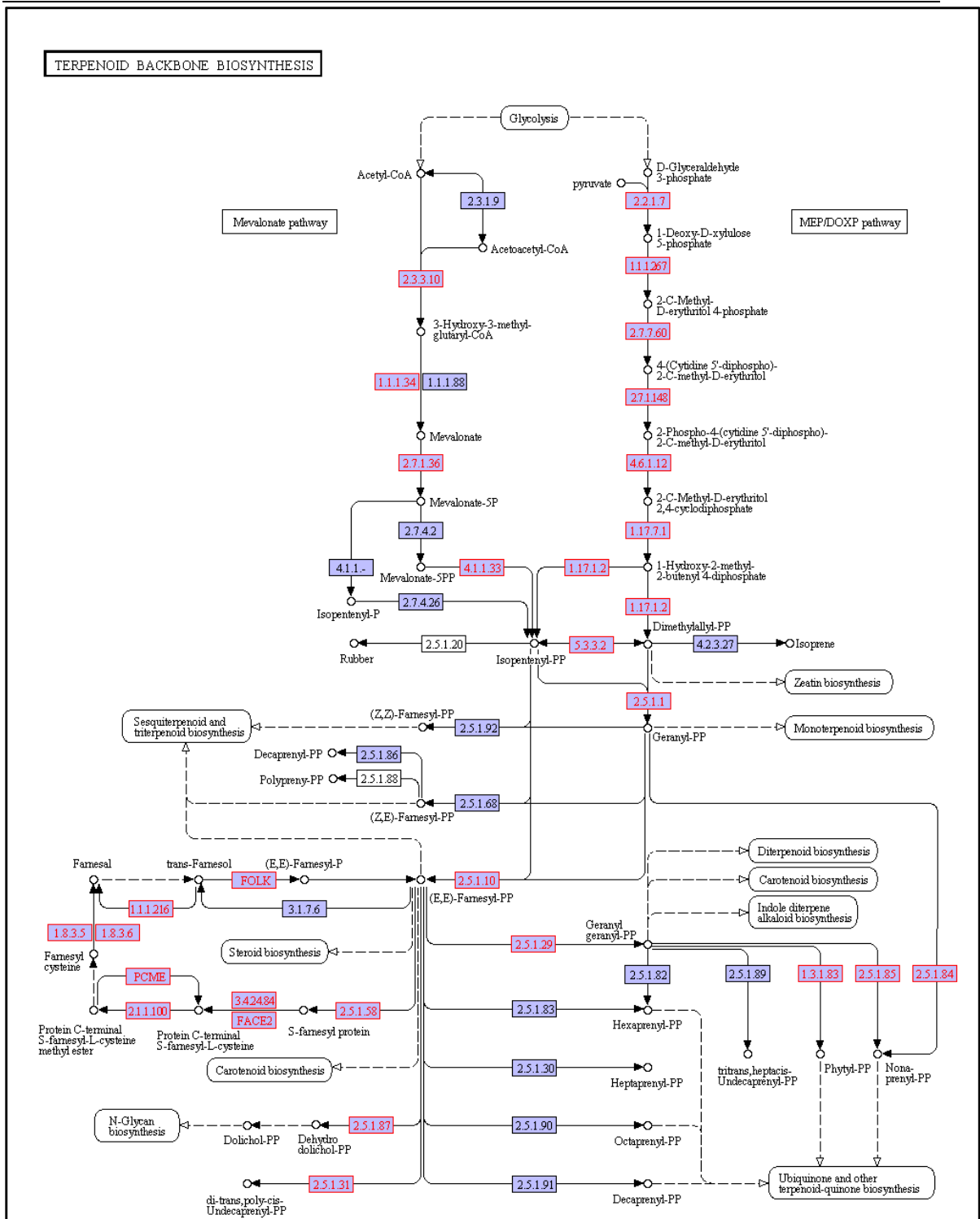


Figure 2.7: KEGG Pathway analysis for unigenes present in the Terpenoid Biosynthetic Pathway (highlighted in colour)

A total of 4335 unigenes were assigned with KO number representing 327 KEGG¹⁵ pathways, from which 2810 unigenes represent unique KO number. The predicted pathways represented the majority of plant biochemical pathways including metabolism, cellular processes and genetic information processing. A KEGG pathway representing the Indole alkaloid biosynthetic pathway is represented in figure 2.7.

According to KEGG pathway mapping of our sequence dataset, forty-five (45) unigenes were found to be involved in terpenoid backbone biosynthesis, ten (10) in Indole Alkaloid Biosynthesis and twenty-eight (28) in Monoterpenoid Biosynthesis (Figure 2.8).

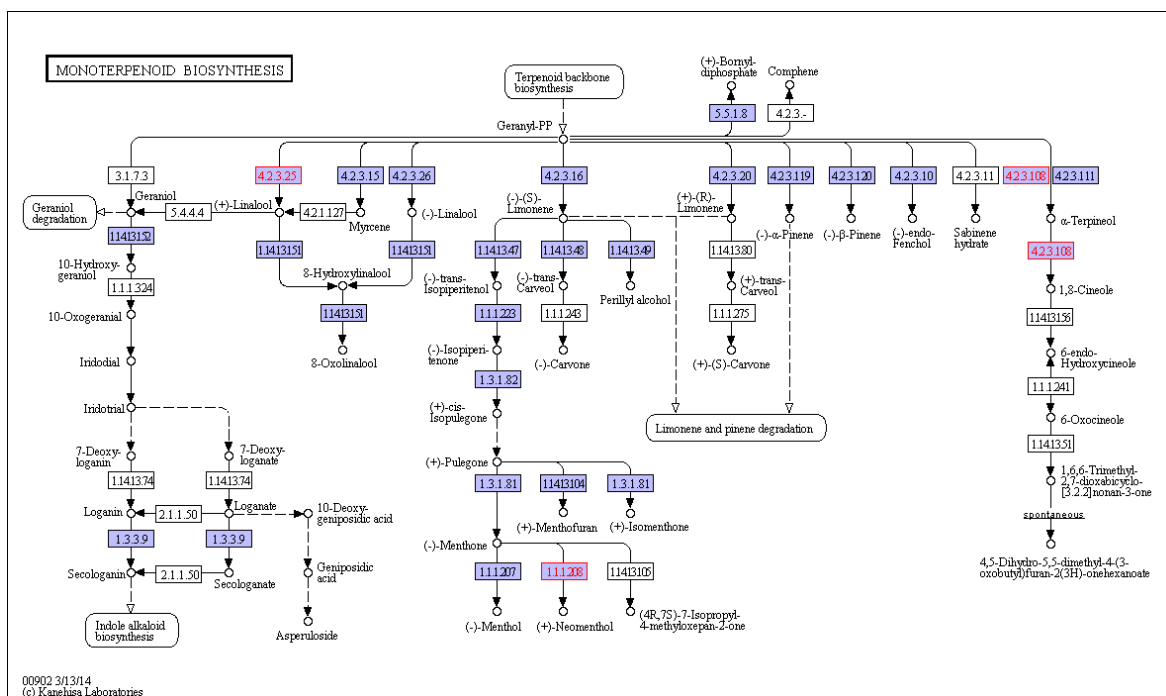


Figure 2.8: KEGG Pathway analysis for unigenes present in the Monoterpenoid Biosynthesis (highlighted in colour)

2.3.2.3.2 Virtual Ribosome

Virtual Ribosome¹³ is a DNA translation tool with two features: 1) providing a tool for translation using an integrated ORF finder and 2) a feature for providing additional sequence annotation such as genes and associated coding regions, variation information, especially exon-intron information, etc. This tool (Virtual Ribosome-V1.1)

was employed for predicting the largest ORF of the transcripts, which can be made use of for further analysis.

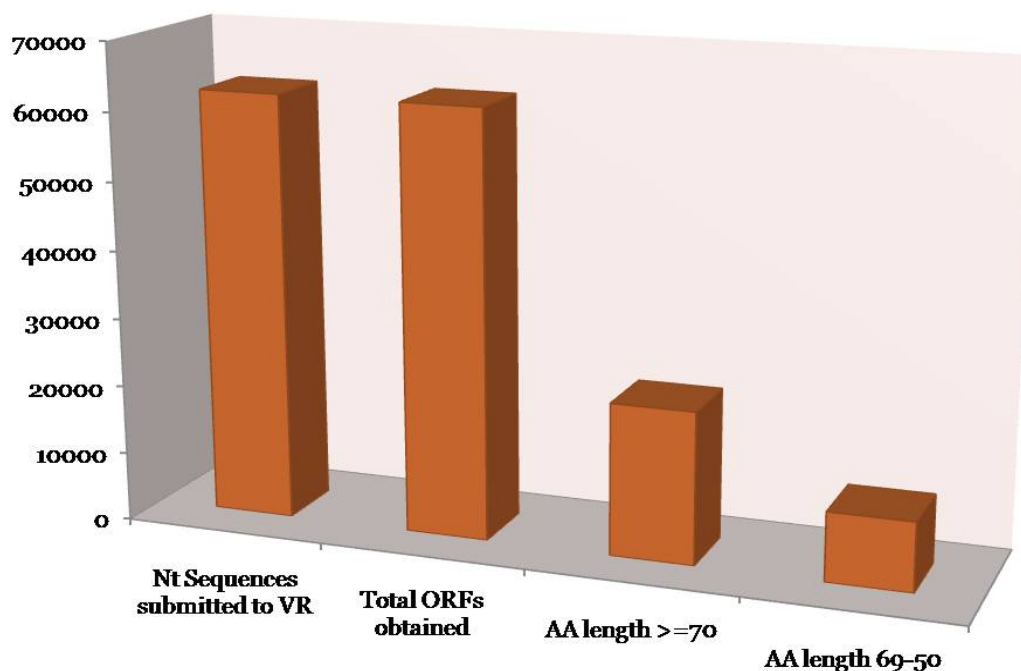


Figure 2.9: ORF Prediction using Virtual Ribosome-V1.1

All the unique unigenes (62352) were submitted to Virtual Ribosome to predict the largest ORF for each of the unigenes, in all six reading frames. Total 62290 unigenes (99.90%) were identified as having ORFs starting at an ATG codon, out of which 22224 unigenes (35.64 %) were found to have the ORFs with length ≥ 70 amino acids without redundancy, 10120 (16.23 %) unigenes contained ORFs with length $69 \geq 50$ amino acids and 29946 (48.03 %) unigenes had ORFs with length < 50 amino acids, whereas only 62 (0.01 %) unigenes were found without having any ORFs in any of the frames (Figure 2.9).

2.3.2.3.3 Pfam analysis

Pfam is a database of protein families, which contains functional descriptions based on protein domain and a profile hidden Markov models¹⁴. The hidden Markov model is searched on sequences based on UniProt Knowledge base (UniProtKB) for information pertaining to the protein family¹⁴.

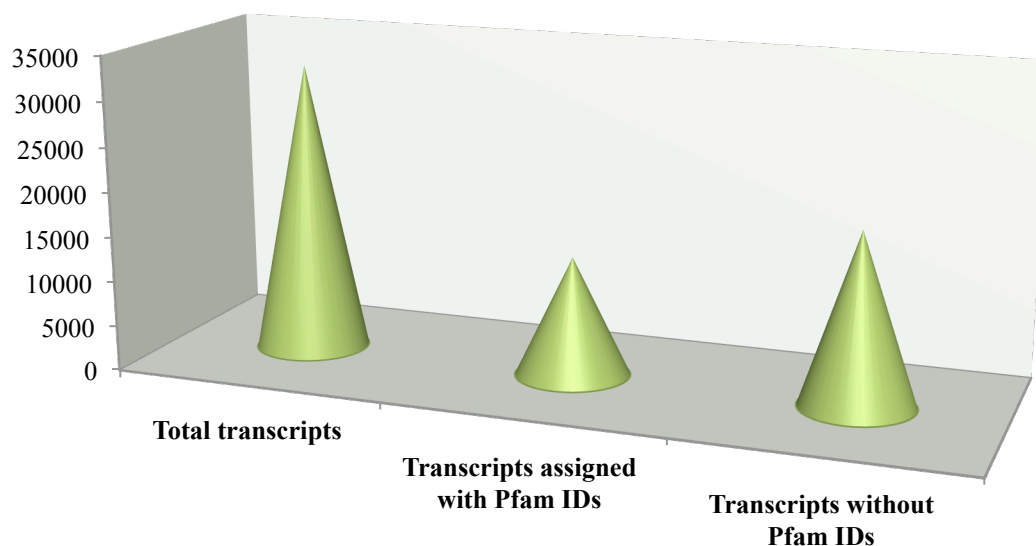


Figure 2.10: Pfam Analysis of transcripts with length ≥ 50 amino acids

The 32344 unigenes having ORFs ≥ 50 amino acids were submitted to Pfam-A database, which contains entries that are curated to eliminate false positives (Figure 2.10). From the Pfam analysis only 13251 unigenes were assigned with Pfam IDs. Further analysis of the 13251 unigenes having Pfam IDs revealed that 8 unigenes belonged to terpenoid biosynthesis, represented by the Pfam IDs: PF01397 (terpene synthase N-terminal domain) and PF03936 (terpene synthase family metal binding domain) and 6 unigenes belonged to Indole Alkaloid Biosynthesis with Pfam ID: PF03088 (strictosidine synthase).

13251 unigenes assigned with Pfam ID were subjected to Blast2Go analysis against NCBI Nr-database, Swissprot/ Uniprot database to screen for the genes involved in monoterpenoid indole alkaloid biosynthesis in *Catharanthus roseus*.

2.3.3 Identification of transcripts involved in the MVA, MEP and Secologanin Biosynthetic Pathways

A compilation of the results of KAAS analysis, Virtual Ribosome and Pfam analysis resulted in the identification of transcripts matching with the various genes

involved in the MVA, MEP and Secologanin Biosynthetic Pathways. These results are summarized in Table 2.1.

Table 2.1: Transcripts involved in the MVA, MEP and Secologanin biosynthetic pathway

Transcript ID	Length of ORF	BLAST Results	% Similarity
MEP Pathway Transcripts			
Vinca_1853	2151	<i>Catharanthus roseus</i> mRNA for 1-deoxyxylulose 5-phosphate synthase	99 %
Vinca_1617	1425	<i>Catharanthus roseus</i> 1-deoxy-D-xylulose-5-phosphate reductoisomerase (dxr) mRNA, complete cds	99 %
Vinca_9009	936	<i>Catharanthus roseus</i> 4-diphosphocytidyl-methylerythritol 2-phosphate synthase mRNA, complete cds	99 %
Vinca_2073	1227	<i>Catharanthus roseus</i> 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK) mRNA, complete cds	99 %
Vinca_2177	711	<i>Catharanthus roseus</i> 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MECS) mRNA, complete cds	98 %
Vinca_481	2223	<i>Catharanthus roseus</i> GCPE protein mRNA, complete cds	99 %
Vinca_465	1389	<i>Catharanthus roseus</i> 1-hydroxy-2-methyl-butenyl 4-diphosphate reductase mRNA, complete cds	99 %
Vinca_775	939	<i>Catharanthus roseus</i> plastid isopentenyl pyrophosphate:dimethylallyl	99 %

		pyrophosphate isomerase isoform 1 mRNA, complete cds; nuclear gene for plastid product	
MVA Pathway Transcripts			
Vinca_1158	1218	<i>Catharanthus roseus</i> acetoacetyl-CoA thiolase mRNA, complete cds	98 %
Vinca_2418	1392	<i>Catharanthus roseus</i> hydroxymethylglutaryl-CoA synthase mRNA, complete cds	99 %
Vinca_1892	1614 (3' and 5' ends missing)	<i>Camellia sinensis</i> cultivar Longjing 43 HMG-CoA reductase (HMGR1) mRNA, complete cds	79 %
Vinca_4366	1163	<i>Catharanthus roseus</i> mevalonate kinase mRNA, complete cds	99 %
Vinca_9276	1497	<i>Catharanthus roseus</i> 5- phosphomevalonate kinase mRNA, complete cds	99 %
Vinca_8419	1265	<i>Catharanthus roseus</i> mevalonate 5- diphosphate decarboxylase mRNA, complete cds	99 %
Vinca_775	939	<i>Catharanthus roseus</i> plastid isopentenyl pyrophosphate:dimethylallyl pyrophosphate isomerase isoform 1 mRNA, complete cds; nuclear gene for plastid product	99 %
Secologanin Biosynthetic Pathway Transcripts			
Vinca_5406, Vinca_48434	1263	Geranyl diphosphate synthase (CrGDS)	99 %
Vinca_2171,	1770	Geraniol synthase (CrGS)	99 %

Vinca_3238			
Vinca_742, Vinca_20416	1482	Geraniol 10-hydroxylase (CrG10H)	100 %
Vinca_536, Vinca_18357	1083	10-hydroxygeraniol dehydrogenase (Cr10HGO)	99 %
Vinca_2280, Vinca_1143	1161	Iridoid synthase (CrIDS)	96 %
Vinca_734, Vinca_6889	1548	7-deoxyloganin synthase (Cr7DLS)	99 %
Vinca_1690, Vinca_15559	1671	7-deoxyloganic acid glucosyltransferase (CrDLGT)	100 %
Vinca_2445, Vinca_891	1566	7-deoxyloganic acid hydroxylase (Cr7DLH)	98 %
Vinca_917, Vinca_34458	1116	Loganic acid methyltransferase (CrLAMT)	99 %
Vinca_263, Vinca_2131	1575	Secologanin synthase (CrSLS)	100 %

2.4 Conclusion

Study of the transcriptome at a global level can provide insights into the gene function, transcriptional pattern and molecular basis of various cellular processes in plants. Transcriptome analysis has been comprehended as an essential step for basic and applied research in any organism.

The present work deals with the optimization of conventional and kit-based protocols for the isolation of high quality total RNA from the various tissues of *Catharanthus roseus*. These total RNA samples were further used for RNA sequencing on Illumina GAI Analyzer and Fastq file of 4.88 GB was generated. The 15485359 high quality paired end raw reads were assembled into 76075 contigs with an optimized hash length of 49, largest average contig length of 548.5 bp and an N50 value of 881 bp.

Functional annotation was carried out using KEGG database, KAAS server, NCBI Nr-database, SwissProt/ Uniprot database and Pfam database. These analyses finally led to the identification of the unigenes related to Secologanin biosynthesis. This data was used for cloning and functional characterization of the genes involved in Secologanin biosynthesis in *Catharanthus roseus*.

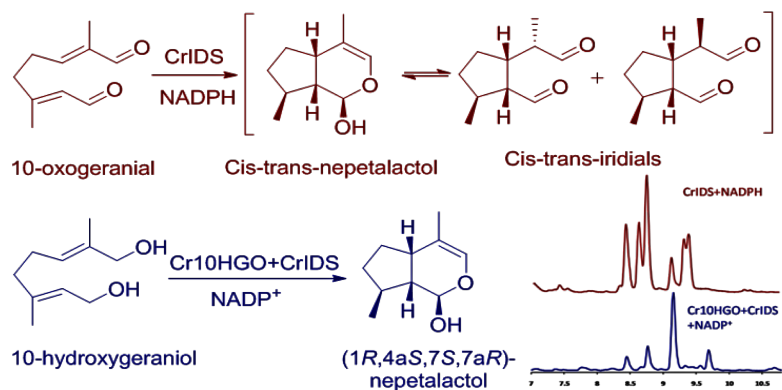
2.5 References

1. Hays, S.B. Some effects of reserpine, a tranquilizer, on the house fly. *J. Econ. Entomol.* **58**, 782-783 (1965).
2. Yui, T. & Takeo, Y. Neuropharmacological studies on a new series of ergot alkaloids; elymoclavine as a potent analeptic on reserpine-sedation. *Jpn. J. Pharmacol.* **7**, 157-161 (1958).
3. Shurin, G.V., Tourkova, I.L., Kaneno, R. & Shurin, M.R. Chemotherapeutic Agents in Noncytotoxic Concentrations Increase Antigen Presentation by Dendritic Cells via an IL-12-Dependent Mechanism. *J. Immunol.* **183**, 137-144 (2009).
4. van der Heijden, R., Jacobs, D.I., Snoeijer, W., Hallared, D. & Verpoorte, R. The *Catharanthus* alkaloids: Pharmacognosy and biotechnology. *Curr. Med. Chem.* **11**, 607-628 (2004).
5. Mills, J.J., Chari, R.S., Boyer, I.J., Gould, M.N. & Jirtle, R.L. Induction of apoptosis in liver tumors by the monoterpene perillyl alcohol. *Cancer Res.* **55**, 979-983 (1995).
6. Shoff, S.M., Grummer, M., Yatvin, M.B. & Elson, C.E. Concentration-dependent increase of murine P388 and B16 population doubling time by the acyclic monoterpene geraniol. *Cancer Res.* **51**, 37-42 (1991).
7. Kolosova, N., Miller, B., Ralph, S., Ellis, B.E., Douglas, C., Ritland, K. & Bohlmann, J. Isolation of high-quality RNA from gymnosperm and angiosperm trees. *Biotechniques* **36**, 821-824 (2004).
8. Srivastava, P.L., Daramwar, P.P., Krithika, R., Pandreka, A., Shankar, S.S. & Thulasiram, H.V. Functional Characterization of Novel Sesquiterpene Synthases from Indian Sandalwood, *Santalum album*. *Sci. Rep.* **5**, 10095-10107 (2015).
9. Zerbino, D.R. & Birney, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821-829 (2008).
10. Schulz, M.H., Zerbino, D.R., Vingron, M. & Birney, E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**, 1086-1092 (2015).
11. Murata, J., Roepke, J., Gordon, H. & De Luca, V. The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization. *Plant Cell* **20**, 524-542 (2008).
12. Schulz, M.H., Zerbino, D.R., Vingron, M. & Birney, E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**, 1086-1092 (2015).
13. Wernersson, R. Virtual Ribosome - a comprehensive DNA translation tool with support for integration of sequence feature annotation. *Nucleic Acids Res.* **34**, W385-W388 (2006).
14. Finn, R.D. et al. Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222-D230 (2014).

-
15. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27-30 (2000).
 16. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659 (2006).

Chapter 3

Cloning, expression and characterization of the genes involved in iridoid biosynthesis



Chapter 3

Cloning, expression and characterization of the genes involved in iridoid biosynthesis

Catharanthus roseus [L.] is a major source of the monoterpene indole alkaloids (MIAs), which are of significant interest due to their therapeutic value. These molecules are formed through an intermediate, *cis-trans*-nepetalactol, a cyclized product of 10-oxogeraniol. One of the key enzymes involved in the biosynthesis of MIAs is an NAD(P)⁺ dependent oxidoreductase system, 10-hydroxygeraniol dehydrogenase (Cr10HGO), which catalyzes the formation of 10-oxogeraniol from 10-hydroxygeraniol via 10-oxogeraniol or 10-hydroxygeraniol. This chapter describes the cloning and functional characterization of Cr10HGO from *C. roseus* and its role in the iridoid biosynthesis. Substrate specificity studies indicated that, Cr10HGO has good activity on substrates such as 10-hydroxygeraniol, 10-oxogeraniol or 10-hydroxygeraniol over monohydroxy linear terpene derivatives. Further it was observed that incubation of 10-hydroxygeraniol with Cr10HGO and iridoid synthase (CrIDS) in the presence of NADP⁺ yielded a major metabolite, which was characterized as (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol by comparing its retention time, mass fragmentation pattern, and co-injection studies with that of the synthesized compound. These results indicate that there is concerted activity of Cr10HGO with iridoid synthase in the formation of (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol, an important intermediate in iridoid biosynthesis. Further, the full-length unigenes, which showed high-ranking with known Geranyl diphosphate synthase and Geraniol synthase from various sources, were used for the cloning and functional characterization of Geranyl diphosphate synthase (CrGDS) and Geraniol synthase (CrGS) from *C. roseus*. CrG10H was cloned using the primers designed from the reported gene, which encodes Geraniol hydroxylase in *C. roseus* as the unigenes 742 had similar sequence to that of the reported one.

3.1 Introduction

Monoterpene indole alkaloids (MIAs) are a multifarious class of natural products with distinct chemical and biological properties¹⁻³. To date, over 3000 MIAs are known with diverse structures and biological activities. The Apocynaceae family plant, *C. roseus* is a rich source of the iridoid derived MIAs and is known to contain over 200 alkaloids in various tissues. Two MIAs from this plant, vincristine and vinblastine, are widely prescribed as potent anti-cancer agents^{4,5}. These MIAs were synthesized from the condensation of tryptamine and the iridoid monoterpene, secologanin. The MIAs' biosynthesis diverges from the isoprenoid biosynthetic pathway at the 19-4 chain elongation intermediate geranyl diphosphate (GPP) formed through head-to-tail condensation of isopentenyl diphosphate (IPP) with dimethylallyl diphosphate (DMAPP) catalyzed by Geranyl diphosphate synthase (CrGDS)⁶. Geraniol synthase (CrGS)⁷ hydrolyses GPP into geraniol, which undergoes hydroxylation at C10 to form 10-hydroxygeraniol by the cytochrome P450 system, Geraniol 10-hydroxylase (CrG10H)⁸. Feeding experiments with labelled 10-hydroxygeraniol, 10-hydroxynerol and iridodiol in *C. roseus* and *Lonicera morrowii* suspension cultures clearly indicated that 10-hydroxygeraniol is oxidized to 10-oxogeraniol by the oxidoreductase system, 10-hydroxygeraniol dehydrogenase (10HGO)⁹⁻¹² (Figure 3.1.1).

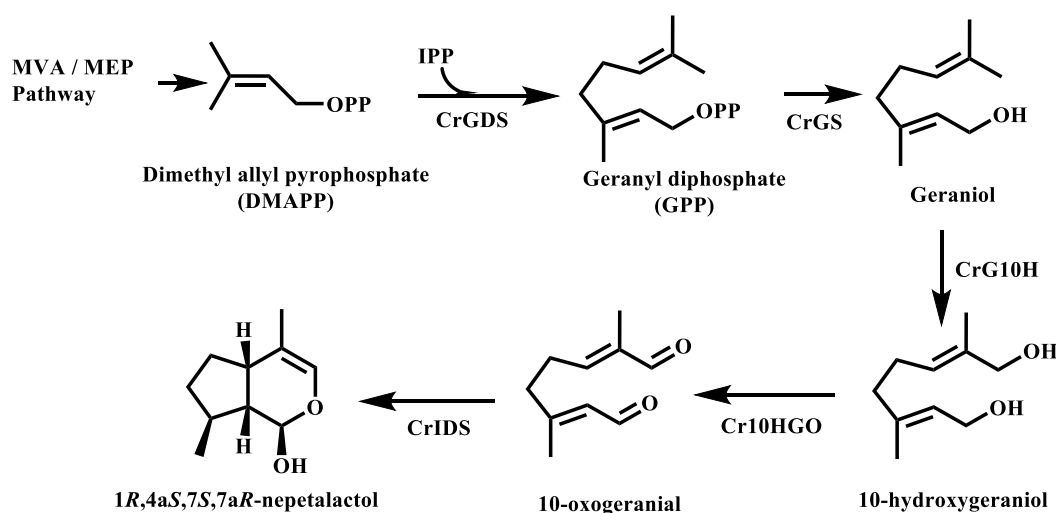


Figure 3.1.1: Iridoid Biosynthetic Pathway cyclizes 10-oxogeraniol into equilibrium mixture of *cis-trans*-nepetalactol and iridodials. The bicyclic compound, *cis-trans*-nepetalactol is the key intermediate.

Recently, a short chain reductive cyclase, iridoid synthase (CrIDS)¹³, which

cyclises 10-oxogeraniol into an equilibrium mixture of *cis-trans*-nepetalactol and iridodials, has been characterized. The bicyclic compound, *cis-trans*-nepetalactol, is the key intermediate involved in the biosynthesis of a structurally diverse array of MIAs. Experiments using labelled intermediates indicated that one of the committed steps during the biosynthesis of iridoids is the oxidation of 10-hydroxygeraniol to its dialdehyde cognate, 10-oxogeraniol¹¹. Ikeda *et al.*¹⁴ had purified the NADP⁺ dependent oxidoreductase protein from *Rauwolfia serpentina* cells, which could convert 10-hydroxygeraniol into 10-oxogeraniol, 10-hydroxygeraniol and 10-oxogeraniol. However, it was found to have better activity on nerol and geraniol. The present work describes the cloning and functional characterization of Cr10HGO and the study on the orchestration of enzyme activity of Cr10HGO with CrIDS in the biosynthesis of desired (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol (Figure 3.1.1). Also, substrate specificity studies of Cr10HGO indicated that, it has broad substrate specificity with 10-hydroxygeraniol, 10-oxogeraniol or 10-hydroxygeraniol over monohydroxy linear terpene derivatives.

3.2 Materials and Methods

3.2.1 Materials used in the study

3.2.1.1 Strains and plasmids used in the study

Escherichia coli TOP10 (Invitrogen/ Life Technologies, USA)

Escherichia coli BL21DE3 (Invitrogen/ Life Technologies, USA)

Escherichia coli Rosetta2DE3 (Novagen, Addgene, USA)

pRSET expression vector (Invitrogen/ Life Technologies, USA)

pET32 expression vector (Novagen, Addgene, USA)

pET28 expression vector (Novagen, Addgene, USA)

pYES2 Yeast Vector (Invitrogen/ Life Technologies, USA)

Saccharomyces cerevisiae Invsc1 (Invitrogen/ Life Technologies, USA)

Restriction enzymes (New England Biolabs, USA)

3.2.1.2 Kits and reagents used in the study

SuperScript® III Reverse Transcriptase (ThermoFisher Scientific, USA)

JumpStart™ Taq DNA Polymerase (Sigma-Aldrich, USA)

AccuTaq™ LA DNA Polymerase (Sigma-Aldrich, USA)

AccuPrime™ Pfx DNA Polymerase (ThermoFisher Scientific, USA)

GenElute™ PCR Clean-Up Kit (Sigma-Aldrich, USA)

GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich, USA)

GelRed™ (Biotium Inc., USA)

3.2.1.3 Buffer compositions

3.2.1.3.1 CrGDS

Lysis buffer

25 mM MOPSO, 300 mM NaCl, pH 8.0, 10 % glycerol, containing 1 mg/mL lysozyme, 1 mM PMSF and 0.5 % CHAPS.

Wash buffer

25 mM MOPSO, 300 mM NaCl, 30 mM Imidazole, pH 8.0, 10 % glycerol.

Elution buffer

25 mM MOPSO, 300 mM NaCl, 250 mM Imidazole, pH 8.0, 10 % glycerol.

Desalting buffer

25 mM MOPSO, 10 % glycerol, pH 7.0.

Enzyme assay buffer

50 mM MOPSO, pH 7.0, 10 % v/v glycerol, containing 2 mM DTT and 10 mM MgCl₂.

3.2.1.3.2 CrGS

Lysis buffer

50 mM Tris-HCl, 300 mM NaCl, pH 8.0, 10 % glycerol, containing 1 mg/mL lysozyme, 1 mM PMSF and 0.5 % CHAPS.

Wash buffer

50 mM Tris-HCl, 300 mM NaCl, 30 mM Imidazole, pH 8.0, 10 % glycerol.

Elution buffer

50 mM Tris-HCl, 300 mM NaCl, 250 mM Imidazole, pH 8.0, 10 % glycerol.

Desalting buffer

100 mM HEPES-KOH, 10 % glycerol, pH 7.0.

Enzyme assay buffer

100 mM HEPES-KOH, pH 7.0, 10 % v/v glycerol, containing 1 mM MgCl₂, 100 μM MnCl₂.

3.2.1.3.3 CrG10H

Lysis buffer

50 mM Tris-HCl, 1 mM EDTA, pH 7.4, containing 600 mM Sorbitol, 5 mM DTT, 0.25 mM PMSF.

Wash buffer – TEK buffer

50 mM Tris-HCl, 100 mM KCl, 1 mM EDTA, pH 7.4.

Suspension buffer – TEG buffer

50 mM Tris-HCl, 1 mM EDTA, 30 % glycerol, pH 7.5.

Enzyme Assay Buffer

100 mM K₂HPO₄, pH 7.6, 1 mM EDTA, 1 mM DTT.

3.2.1.3.4 Cr10HGO**Lysis buffer**

50 mM NaH₂PO₄, 300 mM NaCl, pH 8.0, 10 % glycerol, containing 1 mg/mL lysozyme, 1 mM PMSF and 0.5 % CHAPS.

Wash buffer

50 mM NaH₂PO₄, 300 mM NaCl, 30 mM Imidazole, pH 8.0, 10 % glycerol.

Elution buffer

50 mM NaH₂PO₄, 300 mM NaCl, 250 mM Imidazole, pH 8.0, 10 % glycerol.

Desalting buffer

20 mM MOPS, 10 % glycerol, pH 8.0.

Enzyme assay buffer

20 mM Sodium bicarbonate, 10 % v/v glycerol, pH 10.0.

3.2.1.3.5 CrIDS**Lysis buffer**

50 mM NaH₂PO₄, 300 mM NaCl, pH 8.0, 10 % glycerol, containing 1 mg/mL lysozyme, 1 mM PMSF and 0.5 % CHAPS.

Wash buffer

50 mM NaH₂PO₄, 300 mM NaCl, 30 mM Imidazole, pH 8.0, 10 % glycerol.

Elution buffer

50 mM NaH₂PO₄, 300 mM NaCl, 250 mM Imidazole, pH 8.0, 10 % glycerol.

Desalting buffer

20 mM MOPS, 10 % glycerol, pH 8.0.

Enzyme assay buffer

20 mM MOPS, 10 % glycerol, pH 7.0

3.2.2 RNA isolation and cDNA preparation

As discussed in Chapter 2, total RNA was isolated from the leaf, stem and root tissues of white and pink-flowering plants. 5 µg of total RNA was used for first strand cDNA synthesis using SuperScript® III Reverse Transcription Kit (ThermoFisher Scientific, USA). The cDNA thus constructed, was stored at -20 °C till further use.

3.2.3 Sequence analysis and ORF construction

After identification of the genes involved in Iridoid biosynthesis, sequences were analyzed by NCBI GenBank database and overlapped with the EST fragments to generate full-length sequences. ORFs were selected using online ORF finder software for all the genes for their cloning and functional characterization. Full length ORF primers were designed, both blunt and those containing restriction enzyme sites at both the ends, for all the genes for cloning them in expression vector(s).

Table 3.2.1 Primer sequence for isolation of full length ORF of the genes involved in Iridoid biosynthesis.

Primer name	5'-3' Primer sequence
CrGDS_F	ATGTTGTTTTCCAGAGGATTGTATA
CrGDS_R	TCACTTTCTTCTTGTAATAACGCGT
CrGDS_RE_F	GATATCATGTTGTTTTCCAGAGGATTGTATA
CrGDS_RE_R	CTCGAGTCACTTTCTTCTTGTAATAACGCGT
CrGDS_trnc_RE_F	GCATATGGTAGTTGCAGAGGTCC
CrGS_F	ATGGCAGCCACAATTAGTAACCTTT
CrGS_R	TTAAAAACAAGGTGTAAAAACAAAGC
CrGS_RE_F	GGATCCATGGCAGCCACAATTAGTAACCTTT
CrGS_RE_R	GCGGCCGCTTAAAAACAAGGTGTAAAAACAAA
CrG10H_F	ATGGATTACCTTACCATAATATTAA
CrG10H_R	TTAAAGGGTGCTTGGTACAGCACGC
CrG10H_RE_F	GTCGACATGGATTACCTTACCATAATATTAA
CrG10H_RE_R	CTCGAGTTAAAGGGTGCTTGGTACAGCACGC
Cr10HGO_F	ATGGCGAAATCACCGGAAGTCGAGC
Cr10HGO_R	TTATGCAGATTTCAAGTGTGTTGGCT
Cr10HGO_RE_F	GATATCATGGCGAAATCACCGGAAGTCGAGC
Cr10HGO_RE_R	CTCGAGTTATGCAGATTTCAAGTGTGTTGGCT
CrIDS_F	ATGAGTTGGTGGTGGGAAGAGGTCCA
CrIDS_R	CTAAGGAATAAACCTATAATCCCTC
CrIDS_RE_F	GGATCCATGAGTTGGTGGTGGGAAGAGGTC
CrIDS_RE_R	GCGGCCGCCTAAGGAATAAACCTATAATCCCTC

Bold sequences are the Restriction site sequences used in cloning strategy

3.2.4 Full-length gene isolation and cloning into expression vector

3.2.4.1 Isolation and cloning of ORF of *CrGDS* in pET 32a expression vector

Full-length primers for Geranyl diphosphate Synthase ORF were designed using the transcript of *GDS* as a template. Synthesized cDNA was used for PCR reaction using Accuprime *Pfx* Supermix (Invitrogen) using the PCR program: 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 60 °C for 30 sec, 68 °C for 1 min 20 sec followed by final extension at 68 °C for 10 min. PCR product of 1263 bp was cloned in Zero blunt vector and sequenced. *CrGDS* was sub-cloned in pET32a expression vector after digesting both with *EcorV* and *XhoI* restriction sites, using T₄ DNA ligase (Invitrogen) by incubating overnight at 14 °C. The ligation mixture was transformed in TOP10 chemically competent cells, plated on LA containing 100 µg/mL of ampicillin and incubated overnight at 37 °C. Positive clones were screened by colony PCR with T7 promoter and T7 terminator primers. Plasmids were isolated from 3 positive clones of colony PCR and sequenced with T7 promoter and T7 reverse primers to check for the sequence in the correct vector frame. The gene was expressed in Rosetta 2DE3 cells, but when the pellet was lysed, it was found that the protein was completely insoluble. Hence, according to previous reports of GDS from other plant systems, the *CrGDS* gene was truncated from the 5' end till the second Methionine residue (1-99 amino acids) and sub-cloned in pET 28a vector using the same protocol as before with the restriction enzymes *BamHI* and *XhoI*.

3.2.4.2 Isolation and cloning of ORF of *CrGS* in pET 28a expression vector

Full-length primers for Geraniol Synthase ORF were designed using the transcript of *CrGS* as a template. Synthesized cDNA was used for PCR reaction using Accuprime *Pfx* Supermix (Invitrogen) using the PCR program: 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 58 °C for 30 sec, 68 °C for 2 min followed by final extension at 68 °C for 10 min. PCR product of 1770 bp was cloned in Zero blunt vector and sequenced. *CrGS* was sub-cloned in pET 28a expression vector by digesting both with *BamHI* and *NotI* restriction enzymes and ligating them using T₄ DNA ligase (Invitrogen) by incubating overnight at 14 °C. Ligation mixture was transformed in TOP10 chemically competent cells, plated on LA containing 100 µg/mL of ampicillin and incubated overnight at 37 °C. Positive clones were screened by colony PCR with T7 promoter and T7 terminator primers. Plasmids were isolated

from 2 positive clones of colony PCR and sequenced with T7 promoter and T7 reverse primers to check for the sequence in the correct vector frame.

3.2.4.3 Isolation and cloning of ORF of *CrG10H* in pYES2 expression vector

Full-length primers for Geraniol 10-hydroxylase ORF were designed from the reported gene, which encodes geraniol hydroxylase in *C. roseus*. Synthesized cDNA was used for PCR reaction using Accuprime *Pfx* Supermix (Invitrogen) using the PCR program: 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 60 °C for 30 sec, 68 °C for 1 min 35 sec followed by final extension at 68 °C for 10 min. PCR product of 1482 bp was cloned in Zero blunt vector and sequenced. *CrG10H* was sub-cloned in pYES2 expression vector (Yeast Expression Vector from Invitrogen) after digesting both with *KpnI* and *XhoI* restriction sites, using T₄ DNA ligase (Invitrogen) by incubating overnight at 14 °C. Ligation mixture was transformed in TOP10 chemically competent cells, plated on LA containing 100 µg/mL of ampicillin and incubated overnight at 37 °C. Positive clones were screened by colony PCR with T7 promoter and gene-specific reverse primers. Plasmids were isolated from 4 positive clones of colony PCR and sequenced with T7 promoter primer to check for the sequence in the correct vector frame.

3.2.4.4 Isolation and cloning of ORF of *Cr10HGO* in pRSET B expression vector

Full-length primers for 10-hydroxygeraniol dehydrogenase ORF were designed using the transcript of *Cr10HGO* as a template. Synthesized cDNA was used for PCR reaction using LA DNA Polymerase (Sigma-Aldrich) using the PCR program: 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 58 °C for 30 sec, 68 °C for 1 min 20 sec followed by final extension at 68 °C for 10 min. PCR product of 1083 bp was cloned in Zero blunt vector and sequenced. *Cr10HGO* was sub-cloned in pRSET B expression vector after digesting both with *EcorV* and *XhoI* restriction sites, using T₄ DNA ligase (Invitrogen) by incubating overnight at 14 °C. Ligation mixture was transformed in TOP10 chemically competent cells, plated on LA containing 100 µg/mL of ampicillin and incubated overnight at 37 °C. Positive clones were screened by colony PCR with T7 promoter and gene-specific reverse primers. Plasmids were isolated from 5 positive clones of colony PCR and sequenced with T7 promoter and T7 terminator primers primer to check for the sequence in the correct vector frame.

3.2.4.5 Isolation and cloning of ORF of *CrIDS* in pET 32a expression vector

Full-length primers for Iridoid synthase ORF were designed using the transcript of *CrIDS* as a template. Synthesized cDNA was used for PCR reaction using Accuprime *Pfx* Supermix (Invitrogen) using the PCR program: 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 52 °C for 30 sec, 68 °C for 1 min 20 sec followed by final extension at 68 °C for 10 min. PCR product of 1161 bp was cloned in pET 32a expression vector after digesting both with *BamHI* and *NotI* restriction sites, using T₄ DNA ligase (Invitrogen) by incubating overnight at 14 °C. Ligation mixture was transformed in TOP10 chemically competent cells, plated on LA containing 100 µg/mL of ampicillin and incubated overnight at 37 °C. Positive clones were screened by colony PCR with T7 promoter and gene-specific reverse primers. Plasmids were isolated from 4 out of the 10 positive clones of colony PCR and sequenced with T7 promoter and T7 terminator primers to check for the sequence in the correct vector frame.

3.2.5 Heterologous expressions and protein purifications

3.2.5.1 Heterologous expression and protein purification of *CrGDS*

Expression was carried out in Rosetta2 DE3, cells were grown in Terrific Broth at 37 °C and induced with IPTG at a final concentration of 1 mM and incubated for 18 hours at 16 °C. After the induction, the culture was harvested by centrifugation at 4500 × g for 20 minutes. The cell pellet (5 g/L) was re-suspended in 10 mL/g of cell pellet of Lysis buffer (50 mM MOPSO, 10 % Glycerol, pH 8.0) with 1 mg/mL concentration of lysozyme and 100 µL/5 mL of protease inhibitor cocktail from Sigma and incubated on ice for 30 minutes. After the incubation period, the cells were sonicated with a pulse of 30 sec ON and 30 sec OFF for 10 cycles. The sample was then centrifuged at 3,000 × g for 10 minutes at 4 °C. The crude lysate was purified using Ni-NTA column (2 mL resin / g cell pellet). The protein was finally eluted out in the elution buffer (50 mM MOPSO, 10 % Glycerol, pH 7.0, 250 mM Imidazole) each. Purified protein fractions (after checking on 12 % SDS gel) were pooled together and desalted on Hi-PrepTM 26/10 Desalting Columns with desalting buffer (20 mM MOPS, pH 7.0, 10 % v/v Glycerol) using AKTA (GE Healthcare). The desalted proteins were estimated using Bradford reagent (Bio-Rad) and a Bovine Serum Albumin standard.

3.2.5.2 Heterologous expression and protein purification of *CrGS*

Expression was carried out in Rosetta2 DE3, cells were grown in Terrific Broth at 37 °C and induced with IPTG at a final concentration of 1mM and incubated for 12 hours at 16 °C. After the induction, the culture was harvested by centrifugation at 4500 × g for 20 minutes. The cell pellet (5 g/L) was re-suspended in 10 mL/g of cell pellet of Lysis buffer (20 mM Tris, 250 mM NaCl, 1 % TritonX100, pH 8.0) with 1 mg/mL concentration of lysozyme and 100 µL/5 mL of protease inhibitor cocktail from Sigma and incubated on ice for 30 minutes. After the incubation period, the cells were sonicated with a pulse of 30 sec ON and 30 sec OFF for 10 cycles. The sample was then centrifuged at 10,000 × g for 10 minutes at 4 °C. The crude lysate was purified using Ni-NTA column (2 mL resin / g cell pellet). The protein was finally eluted out in the elution buffer (20 mM Tris, 250 mM NaCl, 1 % Triton X100, pH 8.0, 250 mM Imidazole) each. Purified protein fractions (after checking on 12 % SDS gel) were pooled together and desalted on Hi-Prep™ 26/10 Desalting Columns with desalting buffer (25 mM HEPES, 100 mM KCl, pH 8.0, 10 % v/v Glycerol) using AKTA (GE Healthcare). The desalted proteins were estimated using Bradford reagent (Bio-Rad) and a Bovine Serum Albumin standard.

3.2.5.3 Yeast Expression and Microsome Preparation of *CrG10H*

Expression of active protein was carried out in INVSc1 yeast competent cells. Cells were grown overnight in synthetic complete medium without Uracil (SC-U), containing 2 % glucose at 30 °C, then transferred to induction medium (SC-U, containing 2 % galactose) and further incubated at 30 °C for 12 hours. The cells were centrifuged at 3000 × g for 10 minutes at 4 °C. The cell pellet obtained was washed with TEK buffer (50 mM Tris-HCl, 1 mM EDTA, pH 7.4, 100 mM KCl) (1 mL/g of cell pellet) and centrifuged. The cell pellet (1 g/5 mL) was re-suspended in TES buffer (50 mM Tris-HCl, 1 mM EDTA, 600 mM Sorbitol, 5 mM DTT, 0.25 mM PMSF and pH 7.4) and cells were lysed using a bead-beater (with acid washed glass beads, 425–600 µm) for 6 cycles (pulse on 30 sec, pulse off 30 sec, manual rocking for 3 × 30 sec). The lysed cells were centrifuged at 1000 × g for 5 min at 4 °C to remove the glass beads. Further, the supernatant was subjected to centrifugation at 10,000 × g for 30 min at 4 °C. The 10,000 × g supernatant was centrifuged at 1,00,000 × g for 1 hr 30 min at 4 °C. The microsomal pellet, thus obtained, was suspended in TEG buffer (50 mM Tris-HCl, 1 mM EDTA, 30 % glycerol, pH 7.5)

and homogenized. The homogenized microsomal fraction was aliquoted (0.2 mL), flash-frozen in liquid nitrogen and stored at -80 °C.

3.2.5.4 Heterologous expression and protein purification of *Cr10HGO*

Expression was carried out in BL21 DE3, cells were grown in Luria Bertani Broth at 37 °C and induced with IPTG at a final concentration of 1mM and incubated for 6 hours at 30 °C. After the induction, the culture was harvested by centrifugation at 4500 × g for 20 minutes. The cell pellet (5 g/L) was re-suspended in 10 mL/g of cell pellet of Lysis buffer (50 mM NaH₂PO₄, 300 mM NaCl, 0.5 % CHAPS, pH 8.0, 10 % v/v Glycerol) with 1 mg/mL concentration of lysozyme and 100 µL/5 mL of protease inhibitor cocktail from Sigma and incubated on ice for 30 minutes. After the incubation period, the cells were sonicated with a pulse of 30 sec ON and 30 sec OFF for 10 cycles. The sample was then centrifuged at 5,000 × g for 10 minutes at 4 °C. The crude lysate was purified using Ni-NTA column (2 mL resin / g cell pellet). The protein was finally eluted out in the elution buffer (50 mM NaH₂PO₄, 300 mM NaCl, pH 8.0, 250 mM Imidazole, 10 % v/v Glycerol) each. Purified protein fractions (after checking on 12 % SDS gel) were pooled together and desalted on Hi-Prep™ 26/10 Desalting Columns with desalting buffer (20 mM MOPS, pH 8.0, 10 % v/v Glycerol) using AKTA (GE Healthcare). The desalted proteins were estimated using Bradford reagent (Bio-Rad) and a Bovine Serum Albumin standard.

3.2.5.5 Heterologous expression and protein purification of *CrIDS*

Expression was carried out in Rosetta2 DE3, cells were grown in Terrific Broth at 37 °C and induced with IPTG at a final concentration of 1mM and incubated for 12 hours at 16 °C. After the induction, the culture was harvested by centrifugation at 4500 × g for 20 minutes. The cell pellet (5 g/L) was re-suspended in 10 mL/g of cell pellet of Lysis buffer (50 mM NaH₂PO₄, 300 mM NaCl, 0.5% CHAPS, pH 8.0, 10 % v/v Glycerol) with 1 mg/mL concentration of lysozyme and 100 µL/5 mL of protease inhibitor cocktail from Sigma and incubated on ice for 30 minutes. After the incubation period, the cells were sonicated with a pulse of 30 sec ON and 30 sec OFF for 10 cycles. The sample was then centrifuged at 5,000 × g for 10 minutes at 4 °C. The crude lysate was purified using Ni-NTA column (2 mL resin / g cell pellet). The protein was finally eluted out in the elution buffer (50 mM NaH₂PO₄, 300 mM NaCl, pH 8.0, 250 mM Imidazole, 10 % v/v Glycerol) each. Purified protein fractions (after

checking on 12 % SDS gel) were pooled together and desalted on Hi-PrepTM 26/10 Desalting Columns with desalting buffer (20 mM MOPS, pH 8.0, 10 % v/v Glycerol) using AKTA (GE Healthcare). The desalted proteins were estimated using Bradford reagent (Bio-Rad) and a Bovine Serum Albumin standard.

3.2.6 Enzymatic characterization

3.2.6.1 Enzyme assay of geranyl diphosphate synthase (*CrGDS*)

CrGDS was assayed for its activity using 100 µg desalted protein and 1mM each of Isopentenyl pyrophosphate (IPP) and Dimethylallyl pyrophosphate (DMAPP) as the substrates MOPSO buffer (50 mM MOPSO, pH 7.0, 10 % v/v Glycerol, 2 mM DTT, 10 mM MgCl₂), incubated at 30 °C for 6 hours with 0.2 mM each of IPP and DMAPP. Further incubation at 30 °C, overnight, with 5 U Alkaline phosphatase to hydrolyze the GPP formed. The reaction mixture was extracted twice with 500 µL Dichloromethane. The combined organic phase was dried over sodium sulphate, reduced to ~50 µL with a stream of dry nitrogen and subjected to GC and GC-MS analyses for product characterization by comparing with authentic standards.

3.2.6.2 Enzyme assay of Geraniol synthase (*CrGS*)

CrGS was assayed for its activity using 100 µg desalted protein and 200 µM Geranyl pyrophosphate (GPP) as the substrate in 500 µl HEPES buffer (100 mM HEPES-KOH, pH 7.0, 1 mM MgCl₂, 100 µM MnCl₂, 10 % v/v Glycerol) at 30 °C for 1 hour. Assay mixtures were extracted with Tertiary Butyl Methyl Ether (TBME). The combined organic phase was dried over sodium sulphate, reduced to ~50 µL with a stream of dry nitrogen and subjected to GC and GC-MS analyses for product characterization by comparing with authentic standards.

3.2.6.3 Enzyme assay of Geraniol 10-hydroxylase (*CrG10H*)

CrG10H was assayed for its activity using 1 mg of microsomal fraction in 1 ml potassium phosphate buffer (0.1 M K₂HPO₄, pH 7.6, 1 mM EDTA, 1 mM DTT), in the presence of 1 mM NADPH, 2.5 µM Glucose 6-phosphate, 1U Glucose 6-phosphate dehydrogenase, 10 µM FAD, 10 µM FMN, with the addition of 0.5 mM Geraniol and incubation at 30 °C for 1 hour. Assay mixtures were extracted with Tertiary Butyl Methyl Ether (TBME). The combined organic phase was dried over sodium sulphate, reduced to ~50 µL with a stream of dry nitrogen and subjected to

GC and GC-MS analyses for product characterization by comparing with authentic standards.

3.2.6.4 Enzyme assay of 10-hydroxygeraniol dehydrogenase (*Cr10HGO*)

Cr10HGO was assayed for its activity using 100 µg desalted protein and 1 mM 10-hydroxygeraniol as the substrate in 500 µl Sodium bi-carbonate buffer (20 mM Sodium bi-carbonate, 10 % v/v Glycerol, pH 10.0) at 30 °C for 30 minutes. Assay mixtures were extracted with dichloromethane (DCM). The combined organic phase was dried over sodium sulphate, reduced to ~50 µL with a stream of dry nitrogen and subjected to GC and GC-MS analyses for product characterization by comparing with authentic standards. Product-ratio studies of Cr10HGO were carried out with various substrates using the same procedure.

3.2.6.5 Enzyme assay of Iridoid synthase (*CrIDS*)

CrIDS was assayed for its activity using 100 µg desalted protein and 1 mM 10-Oxogeraniol as the substrate in 500 µl MOPS Buffer (20 mM MOPS, pH 7.0, 10 % v/v Glycerol), in the presence of 200 µM NADPH at 30 °C for 1 hour. Assay mixtures were extracted with Tertiary Butyl Methyl Ether (TBME). The combined organic phase was dried over sodium sulphate, reduced to ~50 µL with a stream of dry nitrogen and subjected to GC and GC-MS analyses for product characterization by comparing with authentic standards.

3.2.7 GC/GC-MS analyses

1 µl of the extract was injected onto a i) 30 m × 0.25 mm × 0.25 µm HP-5 capillary GC column with a temperature gradient from 60 to 120 °C at 20 °C per min, followed by a temperature gradient from 120 to 170 °C at 2.5 °C per min and a final temperature gradient from 170 to 190 °C at 20 °C per min (program 1) or ii) 30 m × 0.25 mm × 0.12 µm Astec CHIRAL DEXTM B-DA capillary column (Sigma-Aldrich, USA) with a temperature gradient from 60 to 100 °C at 4 °C per min, followed by a temperature gradient from 60 to 160 °C at 1 °C per min, followed by a temperature gradient from 160 to 215 °C at 10 °C per min (program 2) or iii) 30 m × 0.25 mm × 0.12 µm Astec CHIRAL DEXTM B-DA Capillary Column (Sigma-Aldrich, USA) with a temperature gradient from 60 to 160 °C at 1 °C per min, followed by a temperature gradient from 160 to 215 °C at 10 °C per min (program 3). Nitrogen was

used as a carrier gas at a flow rate of 1 mL/min. Analyses by GC-MS were carried out under similar conditions at a helium flow rate of 1 mL/min.

3.2.8 Determination of optimum temperature for Cr10HGO Enzyme assays

The enzyme assay was carried out at 22 °C, 25 °C and 30 °C for the same duration. The samples were extracted with DCM, injected in GC / GC-MS and analyzed. It was observed that the percentage of product formation was higher at 30 °C than at the lower temperatures.

3.2.9 Determination of optimum pH

For determining the optimum pH, the enzyme assay was carried out in 20 mM of each of the buffers of different pH: sodium citrate (pH 3.0 – 4.0), Bis-tris (pH 5.0 – 7.0), Trizma (pH 8.0 – 9.0) and sodium bi-carbonate (pH 10.0). The extent of product formation was analyzed by injecting the extracted samples in GC / GC-MS. The enzyme showed minimal activity at pH 3.0, the product formation increasing with increase in pH, and maximum activity at pH 10.0. Also, the buffers with glycerol showed very high activity when compared to those without glycerol.

3.2.10 Determination of Kinetic Parameters of Cr10HGO

Steady-state kinetics was performed in 20 mM Sodium bi-carbonate, 10 % v/v Glycerol, pH 10.0 at 30 °C with varying substrate concentrations, ranging from 0.25 to 500.0 mM with saturation concentration of cofactor, NADP⁺ (500.0 mM) and vice versa. The reactions were followed by measuring changes in NADPH concentration at 340 nm. The kinetic data were fitted with the Graph Pad Prism software and the parameters calculated using Michaelis-Menten plots. Similarly, kinetic parameters for 10-hydroxygeraniol (with NAD⁺), 10-oxogeraniol (with NADP⁺ or NADPH), 10-oxogeraniol (with NADPH), 10-hydroxygeraniol (with NADP⁺ or NADPH) and 10-hydroxynerol (with NADP⁺) were determined.

3.2.11 Combined Assays

The combined assay of CrGDS, CrGS and CrG10H was carried out by adding 0.1 mg of each purified protein (microsomal pellet in case of CrG10H) to an assay mixture containing IPP (0.1 mM), DMAPP (0.1 mM), NADPH (1 mM), Glucose 6-phosphate (2.5 μM), Glucose 6-phosphate dehydrogenase (1U), FAD (10 μM), FMN (10 μM) in buffer (100 mM K₂HPO₄, 50 mM MOPSO, pH 7.6, 1 mM EDTA, 1 mM

DTT, 10 mM MgCl₂, 0.1 mM MnCl₂) and incubated on a metabolic shaker for 3 hours at 50 rpm. After this incubation period, the aqueous phase was extracted three times with 0.5 mL of dichloromethane. The combined organic phase was dried over sodium sulphate, concentrated and subjected to GC and GC-MS analyses.

The Cr10HGO and CrIDS combined assay was carried out by adding 0.1 mg of each purified protein to an assay mixture containing 0.2 mM 10-hydroxygeraniol and 0.2 mM of NADP⁺ in assay buffer (20 mM MOPS, pH 7.0, 10% v/v Glycerol) and incubated at 30 °C on a rotary shaker. Incubations were carried out for different time intervals (30 min to 6 hours). After this incubation period, the assay samples were extracted thrice with 0.5 mL dichloromethane. Similarly, the combined assay for CrG10H, Cr10HGO and CrIDS was carried out in 2 mL assay buffer (100 mM K₂HPO₄, 50 mM MOPS, pH 7.6, 1 mM EDTA, 1 mM DTT, 10 mM MgCl₂, 0.1 mM MnCl₂) containing the required cofactors and geraniol as the substrate.

3.2.12 Phylogenetic analysis

A phylogenetic tree was constructed for all the five genes from *Vinca*, in comparison with terpene synthases from other species. Multiple sequence alignments were performed using software Clustal Omega and phylogenetic tree was generated using the nearest neighbour joining method^{15,16} through MEGA6¹⁷. The evolutionary distances were computed using the Poisson correction method¹⁸.

3.3 Results and Discussion

3.3.1 Isolation, cloning, expression and characterization of Geranyl diphosphate synthase (*CrGDS*)

3.3.1.1 Isolation and cloning of Geranyl diphosphate synthase (*CrGDS*)

Analysis of the transcriptomic data revealed that one transcript, Locus_5406, with an Open reading frame (ORF) of 1263 bp (Figure 3.3.1), encoding a polypeptide of 420 amino acids, displaying 79 % sequence identity with geranyl pyrophosphate synthase from *Nicotiana tabacum* (Genbank ID: AHL84161.1) and 77 % sequence identity with geranyl pyrophosphate synthase from *Solanum lycopersicum* (GenBank ID: NP_001234089.1) at the amino acid level, was identified as geranyl diphosphate synthase (*CrGDS*: KF561462). Deduced amino acid sequence of *CrGDS* was found to have a calculated molecular weight of 46.4 kDa, comprising a polyprenyl synthase 1 domain (positions 176–190) with the PROSITE ID: PS00723, and another polyprenyl synthase 2 domain (positions 298-310) with the PROSITE ID: PS00444 containing the consensus sequences LLhDDvl..DdaetRRG and LGlaFQLiDDVID, respectively.

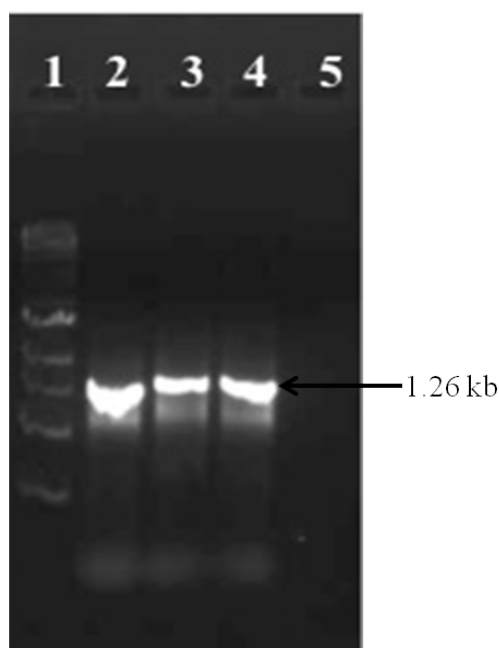


Figure 3.3.1: *CrGDS* full length ORF amplification, **Lane 1:** 1 Kb DNA ladder (Addendum Figure A1), **Lane 2:** PCR of *CrGDS* full-length primers with Root cDNA, **Lane 3:** PCR of *CrGDS* full-length primers with Leaf cDNA, **Lane 4:** PCR of *CrGDS* full-length primers with Stem cDNA, **Lane 5:** Negative Control

The ORF of *CrGDS* was cloned in pET32a vector frame with N terminal His₆ tag for affinity purification under the control of T7-RNA polymerase promoter for

expression of active protein in *E.coli* Rosetta2 (DE3) cells. Positive clones were screened by colony PCR, followed by sequencing with T7 promoter and T7 reverse primer.

3.3.1.2 Bacterial expression and protein purification

Expression was carried out in Rosetta2 DE3. The recombinant protein was expressed and purified as discussed earlier. But, solubilization of the CrGDS protein proved to be very difficult even after administering several buffers like HEPES, Tris, sodium phosphate and MOPSO (all at pH 8.0), with a combination of detergents like CHAPS, Triton X100, etc., the highest solubilization obtained with the combination of MOPSO and CHAPS, even after which purification could not be achieved (Figure 3.3.2). Hence, the CrGDS was truncated of 99 amino acids (1-99) and sub-cloned in pET 28a vector and once again expression and purification was carried out.

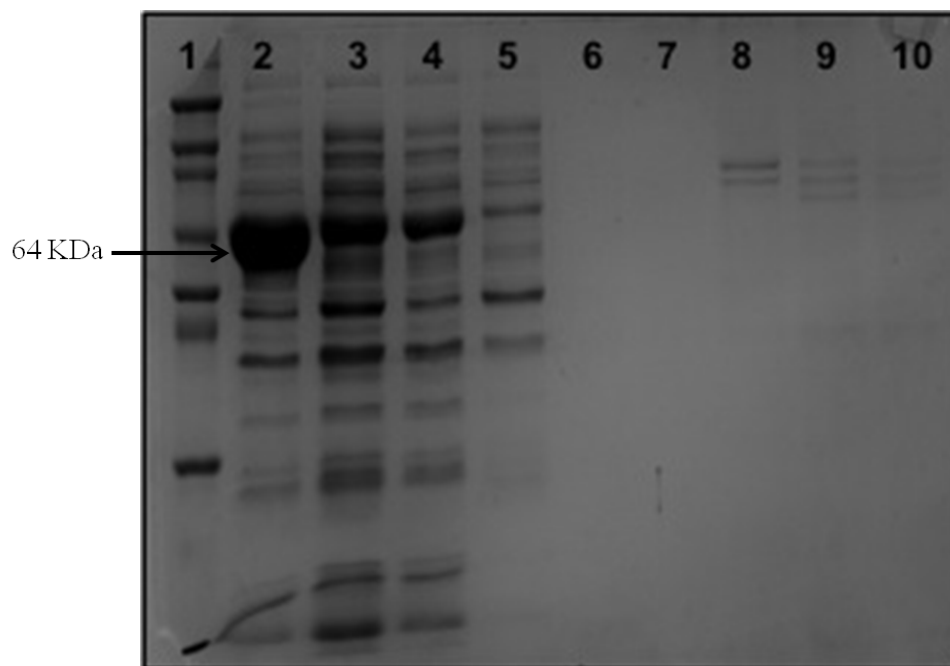


Figure 3.3.2: SDS-PAGE for CrGDS protein purification in pET 32a, **Lane 1:** BenchMark™ Protein Ladder (Addendum Figure A2), **Lane 2:** Pellet fraction, **Lane 3:** Crude lysate (supernatant fraction), **Lane 4:** Unbound fraction, **Lanes 5-7:** Wash fractions, **Lanes 8-10:** Elution fractions.

The purified protein fractions gave a band at 40 kDa on a 12 % SDS-PAGE (Figure 3.3.3). The purified fractions were desalted, flash-frozen and store at -80 °C till used.

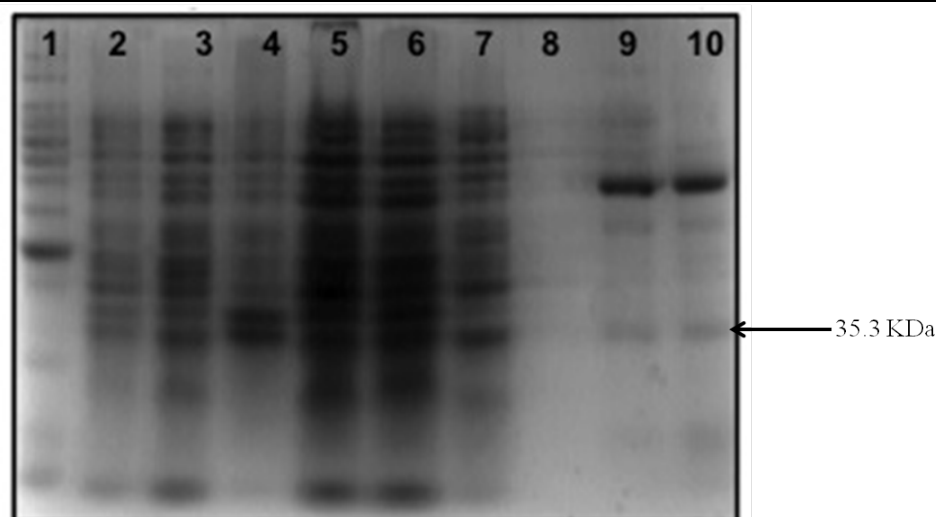
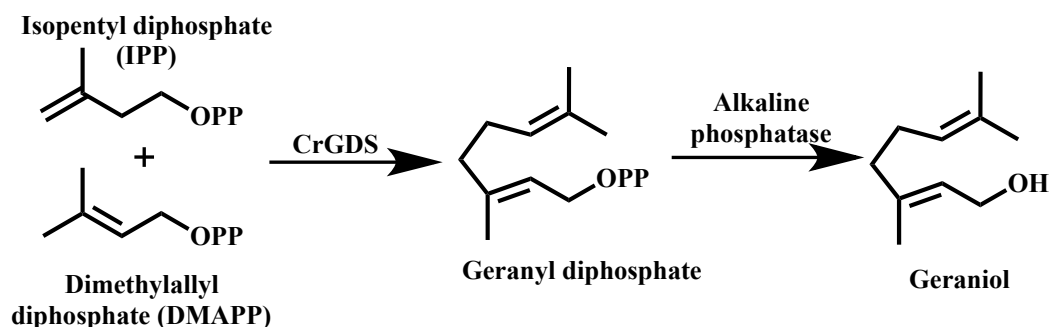


Figure 3.3.3: SDS-PAGE for truncated CrGDS protein purification in pET 28a, **Lane 1:** BenchMark™ Protein Ladder (Addendum Figure A2), **Lane 2:** Uninduced Pellet fraction, **Lane 3:** Uninduced Crude lysate (supernatant fraction), **Lane 4:** Induced Pellet fraction, **Lane 5:** Induced Crude lysate (supernatant fraction), **Lane 6:** Unbound fraction, **Lanes 7-8:** Wash fractions, **Lanes 9-10:** Elution fractions.

3.3.1.3 Enzymatic characterization of *CrGDS*



Scheme 3.3.1: Schematic representation of *CrGDS* assay

Geranyl diphosphate, also known as geranyl pyrophosphate (GPP), is an important intermediate in the Terpene biosynthetic pathways. *CrGDS* was characterized by incubating it with IPP and DMAPP, which resulted in the formation of GPP, which on subsequent treatment with alkaline phosphatase yielded the hydrolyzed product, geraniol (Scheme 3.3.1). This product was characterized by mass fragmentation pattern and co-injection with standard geraniol (Figure 3.3.4).

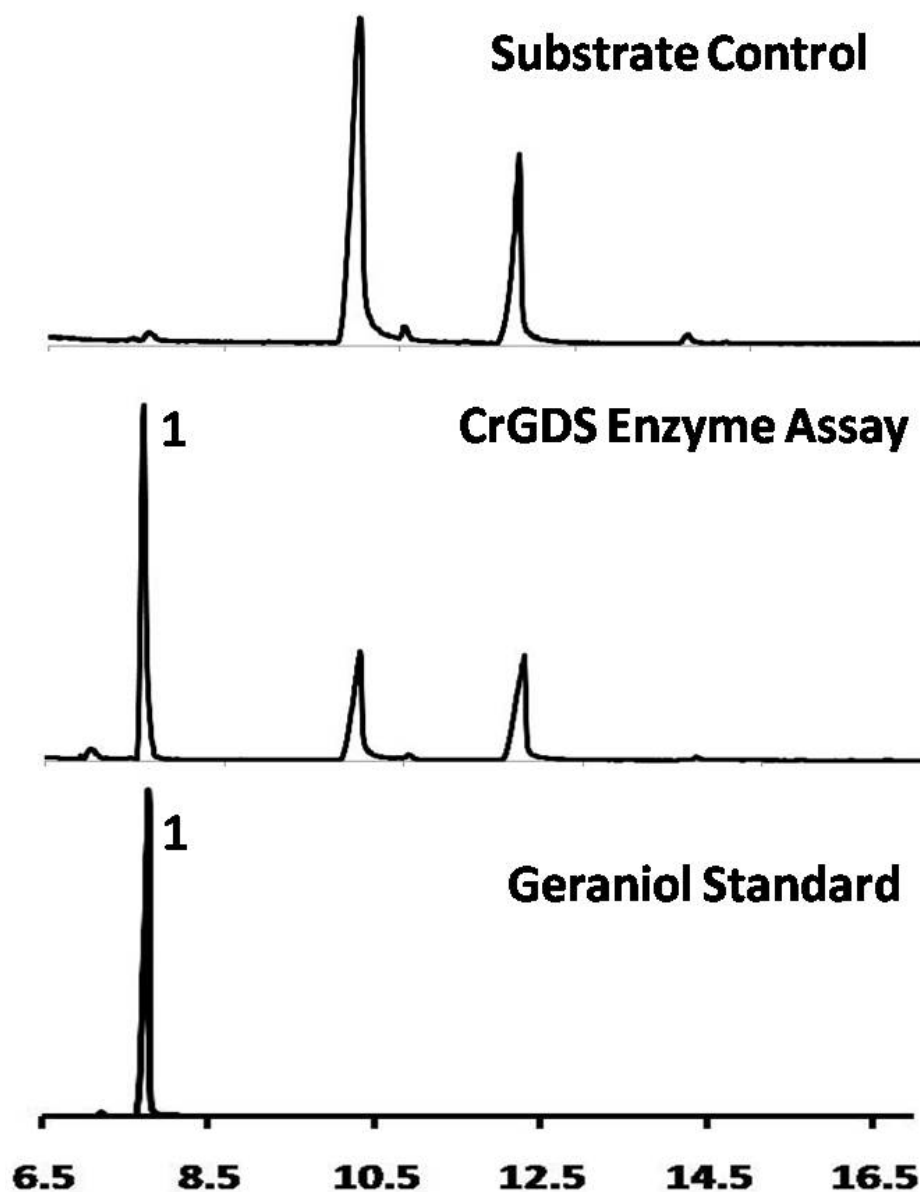


Figure 3.3.4: Total Ion Chromatogram (TIC) of Assay of CrGDS with IPP and DMAPP as substrates; 1-geraniol.

3.3.2 Isolation, cloning, expression and characterization of Geraniol synthase (*CrGS*)

3.3.2.1 Isolation and cloning of Geraniol synthase (*CrGS*)

Analysis of the transcriptomic data revealed that one transcript, Locus_2171, with an Open reading frame (ORF) of 1770 bp (Figure 3.3.5), encoding a polypeptide of 589 amino acids, displaying 67 % sequence identity with geraniol synthase from *Phyla dulcis* (Genbank ID: ADK62524.1) and 66 % sequence identity with geraniol synthase from *Valeriana officinalis* (GenBank ID: AHE41084.1), at the amino acid

level, was identified as geraniol synthase (CrGS: KF561459). Deduced amino acid sequence of CrGS was found to have a calculated molecular weight of 67.8 kDa, comprising a Serine-rich motif (positions 6-54) with the PROSITE ID: PS50324, containing the consensus sequences Snlsflak...sssssssssmS and another Plant Terpene Cyclase domain, which includes a diverse group of monomeric plant terpene cyclases (Tspa-Tspf) that convert the acyclic isoprenoid diphosphates into cyclic monoterpenes, diterpenes, or sesquiterpenes; a few form acyclic species.

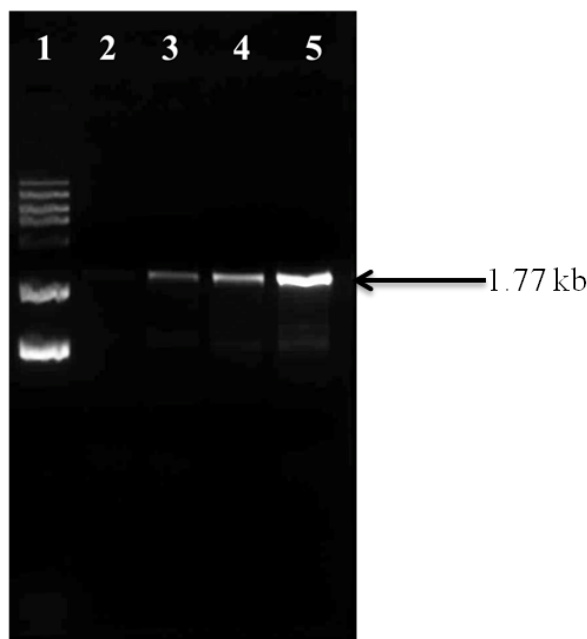


Figure 3.3.5: CrGS full length ORF amplification, **Lane 1:** 1 Kb DNA ladder (Addendum Figure A1), **Lane 2:** Negative Control, **Lane 3:** PCR of CrGS full-length primers with Root cDNA, **Lane 4:** PCR of CrGS full-length primers with Leaf cDNA, **Lane 5:** PCR of CrGS full-length primers with Stem cDNA.

The ORF of *CrGS* was cloned in pET 28a vector frame with N terminal His₆ tag for affinity purification under the control of T7-RNA polymerase promoter for expression of active protein in *E.coli* Rosetta2 (DE3) cells. Positive clones were screened by colony PCR, followed by sequencing with T7 promoter and T7 reverse primer.

3.3.2.2 Bacterial expression and protein purification

Expression was carried out in Rosetta2 DE3. The recombinant protein of CrGS in pET28a was expressed and purified as discussed earlier. The purified protein fractions gave a band at 71 kDa on a 12 % SDS-PAGE (Figure 3.3.6). The purified fractions were desalted, flash-frozen and store at -80 °C till used.

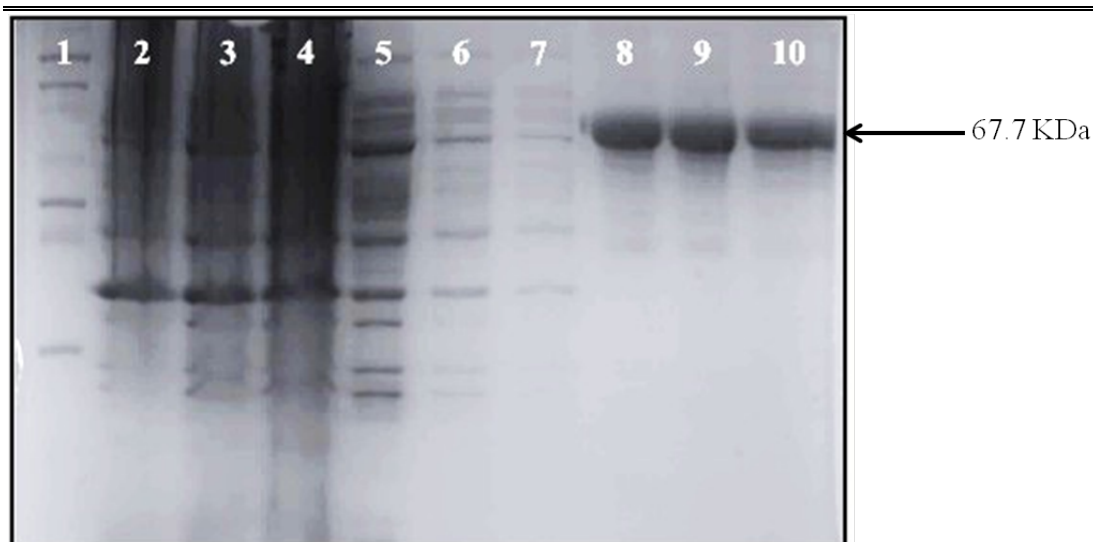
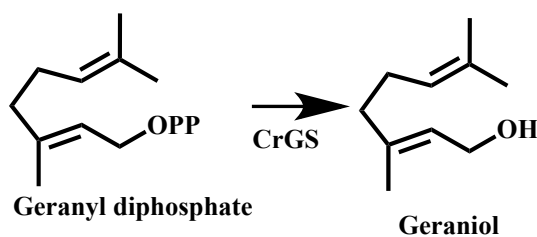


Figure 3.3.6: SDS-PAGE for CrGS protein purification in pET 28a, **Lane 1:** BenchMark™ Protein Ladder (Addendum Figure A2), **Lane 2:** Pellet fraction, **Lane 3:** Crude lysate (supernatant fraction), **Lane 4:** Unbound fraction, **Lanes 5-7:** Wash fractions, **Lanes 8-10:** Elution fractions.

3.3.2.3 Enzymatic characterization of *CrGS*



Scheme 3.3.2: Schematic representation of *CrGS* assay

CrGS carries out the hydroxylation of Geranyl diphosphate, to geraniol. Geraniol is present as a major component of rose oil, palm oil and also present in minute quantities in lemon and other essential oils. CrGS was characterized by incubating it GPP, which resulted in the formation of geraniol (Scheme 3.3.2). This product was characterized by mass fragmentation pattern and co-injection with standard geraniol (Figure 3.3.7).

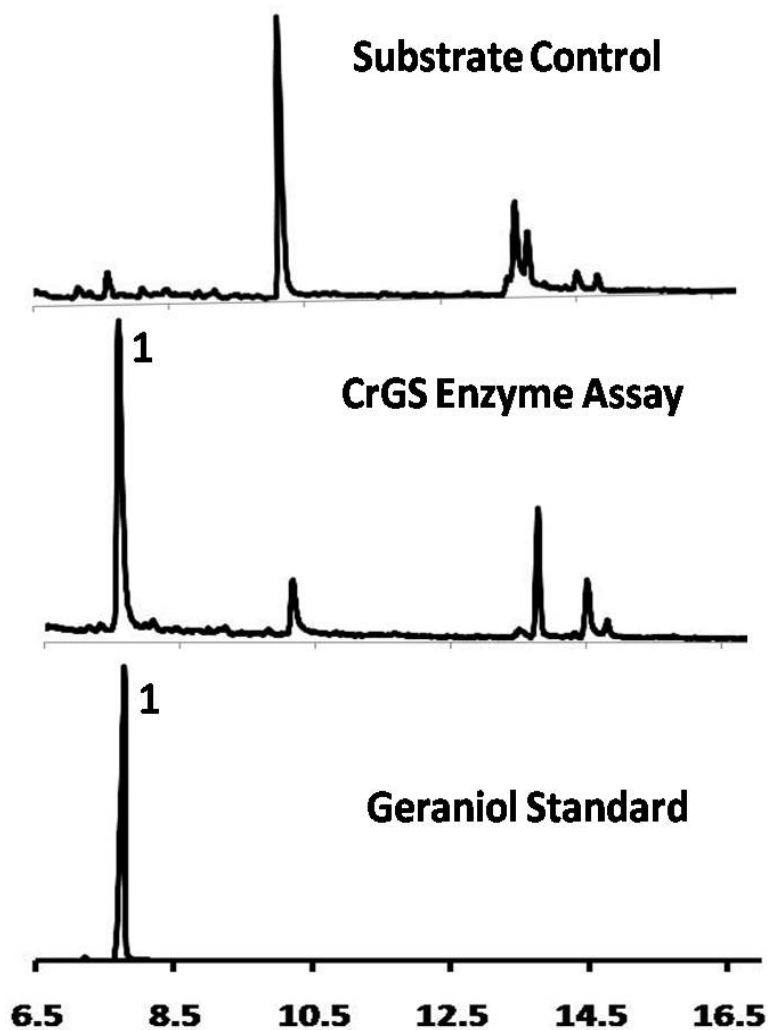


Figure 3.3.7: Total Ion Chromatogram of Assay of CrGS with GPP as substrate; 1-geraniol.

3.3.3 Isolation, cloning, expression and characterization of Geraniol 10-hydroxylase (*CrG10H*)

3.3.2.1 Isolation and cloning of Geraniol 10-hydroxylase (*CrG10H*)

Analysis of the transcriptomic data revealed that one transcript, Locus_742, with an open reading frame (ORF) of 1482 bp (Figure 3.3.8), encoding a polypeptide of 493 amino acids, displaying 83 % sequence identity with geraniol 10-hydroxylase from *Ophiorrhiza pumila* (Genbank ID: BAP90522.1) and 66 % sequence identity with geraniol 10-hydroxylase from *Swertia asarifolia* (GenBank ID: ALJ75585.1), was identified as CrG10H (KF561461). Deduced amino acid sequence of CrG10H was found to have a calculated molecular weight of 55.7 kDa, comprising a

CYTOCHROME_P450 domain (positions 429-438) with the PROSITE ID: PS00086, containing the consensus sequences FGaGRRICPG.

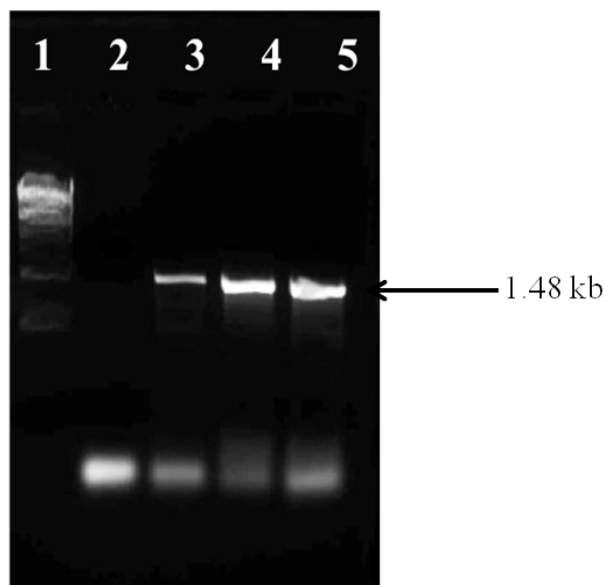


Figure 3.3.8: CrG10H full length ORF amplification, **Lane 1:** 1 Kb DNA ladder (Addendum Figure A1), **Lane 2:** Negative Control, **Lane 3:** PCR of CrG10H full-length primers with Root cDNA, **Lane 4:** PCR of CrG10H full-length primers with Leaf cDNA, **Lane 5:** PCR of CrG10H full-length primers with Stem cDNA.

The ORF of *CrG10H* was cloned in pYES2 vector frame under the control of T7-RNA polymerase promoter for expression of active protein in *Saccharomyces cerevisiae* INVSc1 cells. Positive clones were screened by colony PCR, followed by sequencing with T7 promoter primer.

3.3.2.2 Yeast expression and microsomes preparation

Expression was carried out in INVSc1 cells. The recombinant protein of CrG10H in pYES2 was expressed and microsomes prepared, as discussed earlier. The protein fractions at each stage were loaded on a 12 % SDS-PAGE (Figure 3.3.9). A band at 55.7 KDa confirmed expression of CrG10H. The microsomal fractions were flash-frozen and store at -80 °C till used.

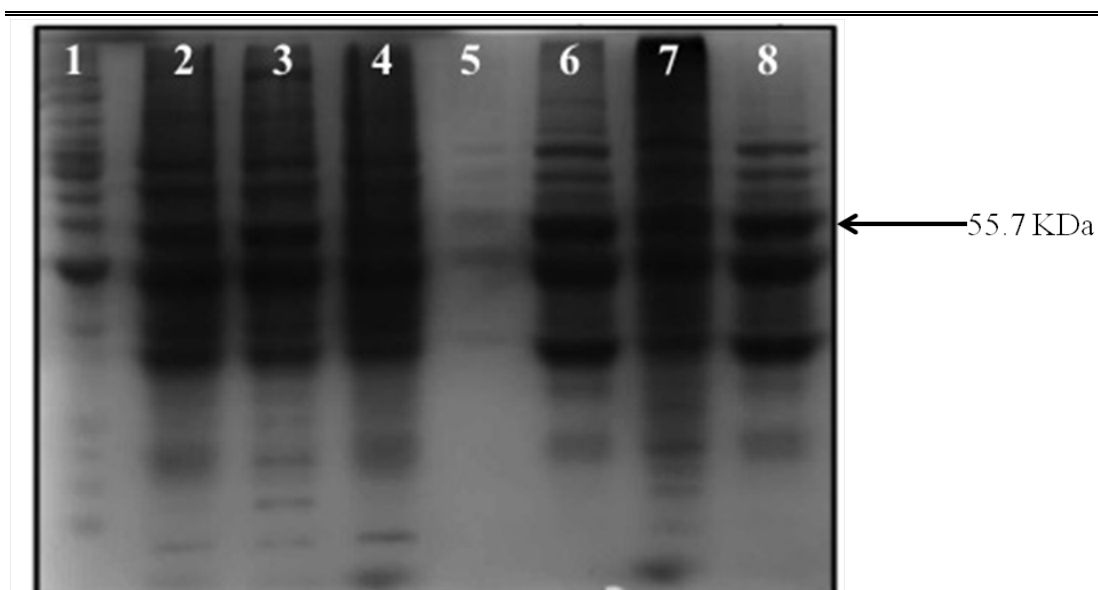
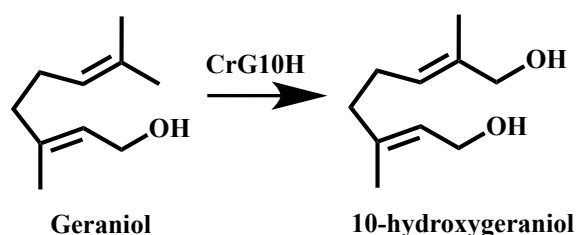


Figure 3.3.9: SDS-PAGE for CrG10H protein in pYES2, **Lane 1:** BenchMark™ Protein Ladder (Addendum Figure A2), **Lane 2:** Uninduced Pellet fraction, **Lane 3:** Uninduced Crude lysate (supernatant fraction), **Lane 4:** Induced Pellet fraction, **Lane 5:** Induced Supernatant fraction, **Lane 6:** Supernatant 1 (10000 × g centrifugation), **Lanes 7:** Microsomal fraction, **Lane 8:** Supernatant 2 (100000 × g centrifugation).

3.3.2.3 Enzymatic characterization of *CrG10H*



Scheme 3.3.3: Schematic representation of *CrG10H* assay

CrG10H carries out the hydroxylation of geraniol at the 10th position to give 10-hydroxygeraniol. 10-hydroxygeraniol is an important intermediate in the formation of Terpene indole alkaloids. CrG10H was characterized by incubating it with geraniol, which resulted in the formation of 10-hydroxygeraniol (Scheme 3.3.3). This product was characterized by mass fragmentation pattern and co-injection with standard 10-hydroxygeraniol (Figure 3.3.10). CYP450 mono-oxygenase enzymes normally require a reductase partner for activity, which in the case of CrG10H has been played by the yeast reductase. Due to this complementation, only very little of the substrate (geraniol) has been utilized by CrG10H to produce the small peak of 10-hydroxygeraniol (Figure 3.3.10).

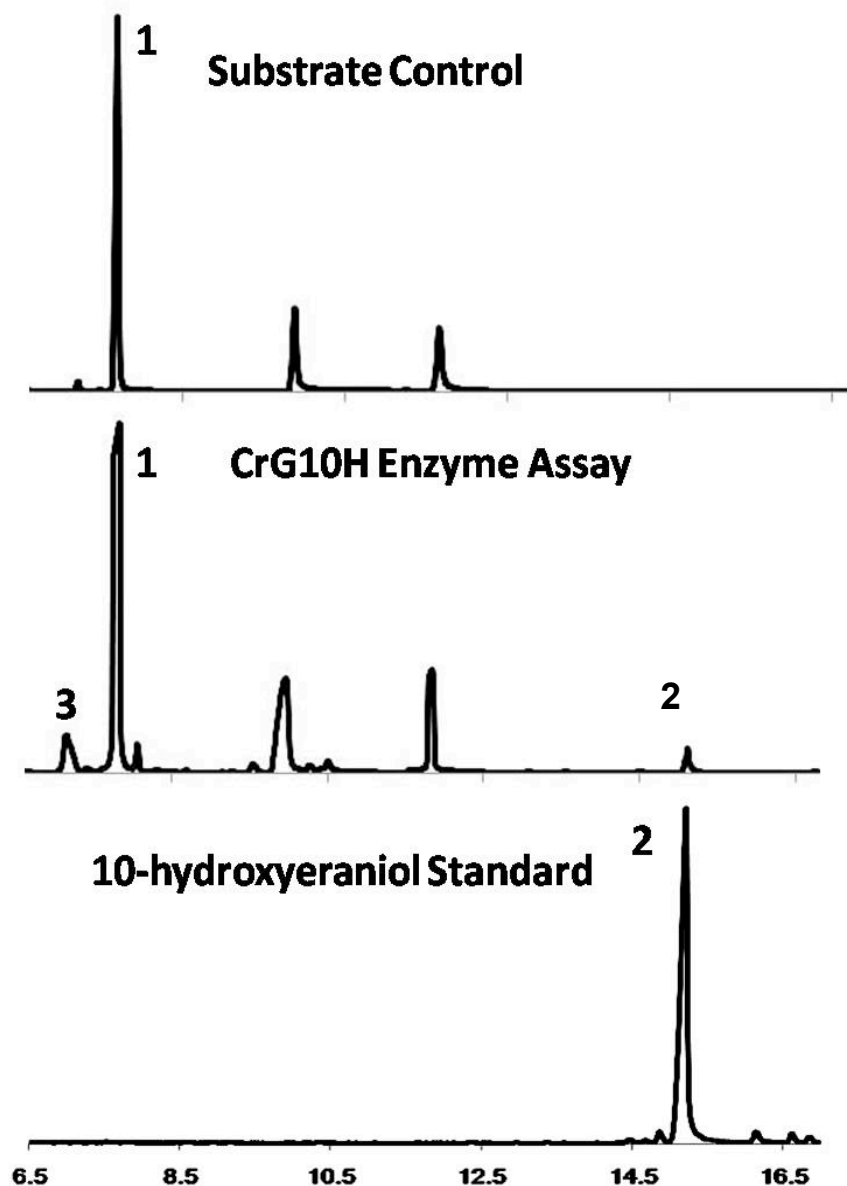


Figure 3.3.10: Total Ion Chromatogram of Assay of CrG10H with geraniol as substrate; 1-geraniol, 2-10 hydroxygeraniol, 3-nerol.

3.3.4 Isolation, cloning, expression and characterization of 10-hydroxygeraniol dehydrogenase (*Cr10HGO*)

3.3.4.1 Isolation and cloning of 10-hydroxygeraniol dehydrogenase (*Cr10HGO*)

An Open reading frame (ORF) of 1083 bp (Figure 3.3.11), encoding a polypeptide of 360 amino acids, displaying 78 % sequence identity with both cinnamyl alcohol dehydrogenase from *Populus trichocarpa* (Genbank ID: ACC6387415) and geraniol dehydrogenase from *Ocimum basilicum* (GenBank ID: Q2KNL616), was identified as Cr10HGO (KF561458). Deduced amino acid sequence

of Cr10HGO was found to have a calculated molecular weight of 38.9 kDa, comprising zinc binding alcohol dehydrogenase signature sequence (positions 71–85) with the PROSITE ID: ADH_ZINC (PS00059), and the consensus sequence G-H-E-x-EL-G-7-x(4)-[GA]-x(2)-[IVSAC]. Although the sequence of this gene was shown to be similar to the gene AY352047¹⁹ found in the data bank, its functional characterization has not been published.

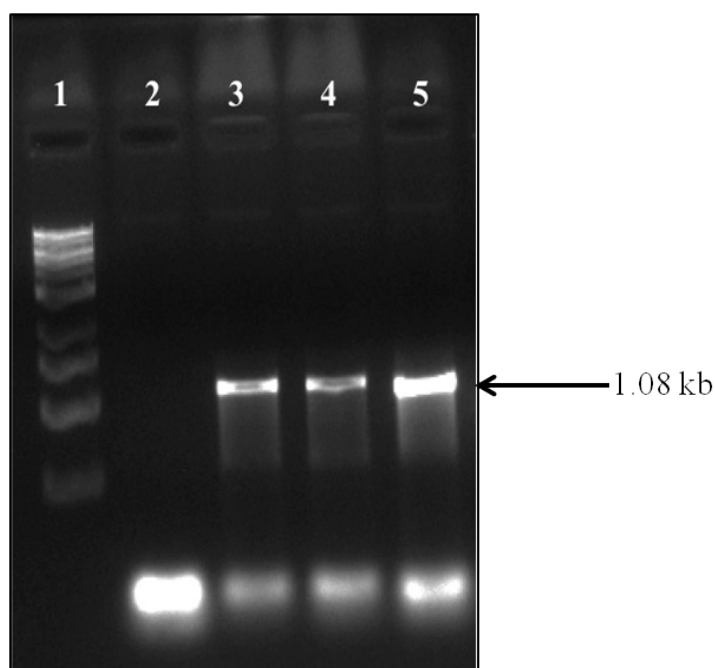


Figure 3.3.11: Cr10HGO full length ORF amplification, **Lane 1:** 1 Kb DNA ladder (Addendum Figure A1), **Lane 2:** Negative Control, **Lane 3:** PCR of Cr10HGO full-length primers with Stem cDNA, **Lane 4:** PCR of Cr10HGO full-length primers with Leaf cDNA, **Lane 5:** PCR of Cr10HGO full-length primers with Root cDNA.

This unigene was cloned in pRSET B expression vector with N-terminal His₆-tag under the control of T7-RNA polymerase promoter for expression of soluble active proteins in *E. coli* BL21 (DE3) cells. Recombinant His₆-tagged proteins were purified to homogeneity by Ni²⁺-affinity chromatography with the yield of 12 mg/L.

3.3.4.2 Bacterial Expression and protein purification

Expression was carried out in Rosetta2 DE3. The recombinant protein of Cr10HGO in pRSETB was expressed and purified as discussed earlier. The purified protein fractions gave a band at 40 kDa on a 12 % SDS-PAGE (Figure 3.3.12). The purified fractions were desalted, flash-frozen and store at -80 °C till used.

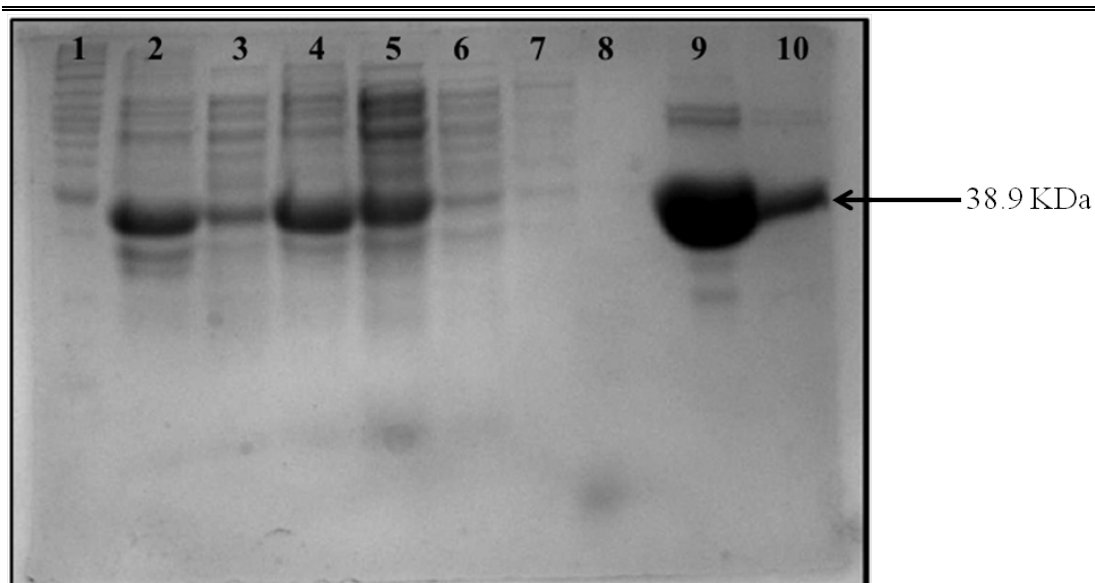
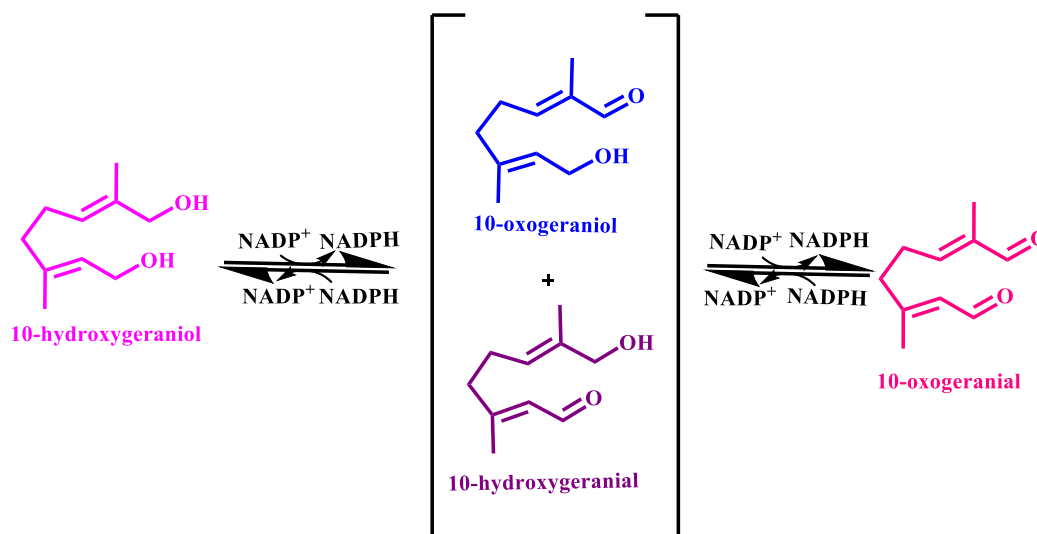


Figure 3.3.12: SDS-PAGE for Cr10HGO protein purification in pRSET B, **Lane 1:** BenchMark™ Protein Ladder (Addendum Figure A2), **Lane 2:** Uninduced Pellet fraction, **Lane 3:** Uninduced Crude lysate (supernatant fraction), **Lane 4:** Induced Pellet fraction, **Lane 5:** Induced Crude lysate (supernatant fraction), **Lane 6:** Unbound fraction, **Lanes 7-8:** Wash fractions, **Lanes 9-10:** Elution fractions.

3.3.4.3 Enzymatic characterization of *Cr10HGO*



Scheme 3.3.4: Schematic representation of *Cr10HGO* assay. NADP⁺ is utilized in the forward reaction, whereas NADPH is utilized in the reverse reactions, generating vice-versa in the process.

Gas chromatography and mass spectrometric analyses of the reaction products after incubation of purified Cr10HGO protein with 10-hydroxygeraniol in the presence of NADP⁺ resulted in the formation of 10-oxogeraniol along with 10-oxogeraniol and 10-hydroxygeranial as minor products (Figure 3.3.13). The formation of these products was further confirmed by comparing the fragmentation pattern as

well as co-injection studies using corresponding synthesized compounds, 10-oxogeraniol, 10-oxogeraniol and 10-hydroxygeraniol. Further, Cr10HGO efficiently converted 10-oxogeraniol and 10-hydroxygeraniol into 10-oxogeraniol in the presence of NADP⁺. However, when NADPH was used as cofactor, 10-hydroxygeraniol was found to be the major enzymatic product with substrates, 10-oxogeraniol, 10-hydroxygeraniol and 10-oxogeraniol, indicating that the Cr10HGO mediated reaction is reversible. The NADP⁺ dependent oxidoreductase protein purified from *R. serpentina* catalyzes dehydrogenation of nerol and geraniol in an efficient manner compared to 10-hydroxygeraniol¹⁴. Similarly, the oxidoreductase purified from *Nepeta racemosa*²⁰ also showed better activity towards geraniol, nerol and 10-hydroxynerol than towards 10-hydroxygeraniol. The recently reported 8-HGO, which encodes the NADP⁺ dependent oxidoreductase from *C. roseus* carries out the dehydrogenation of 10-hydroxygeraniol and also other acyclic monoterpenes²¹, but does not possess much sequence similarity with Cr10HGO. In contrast to these observations, monohydroxy terpene derivatives such as geraniol, nerol, and farnesol were found to be poor substrates for Cr10HGO as compared to the reported 8HGO²¹ (Figure 3.3.14). Studies on the effects of temperature on Cr10HGO mediated reaction revealed that the Cr10HGO activity was found to be optimum at 30 °C.

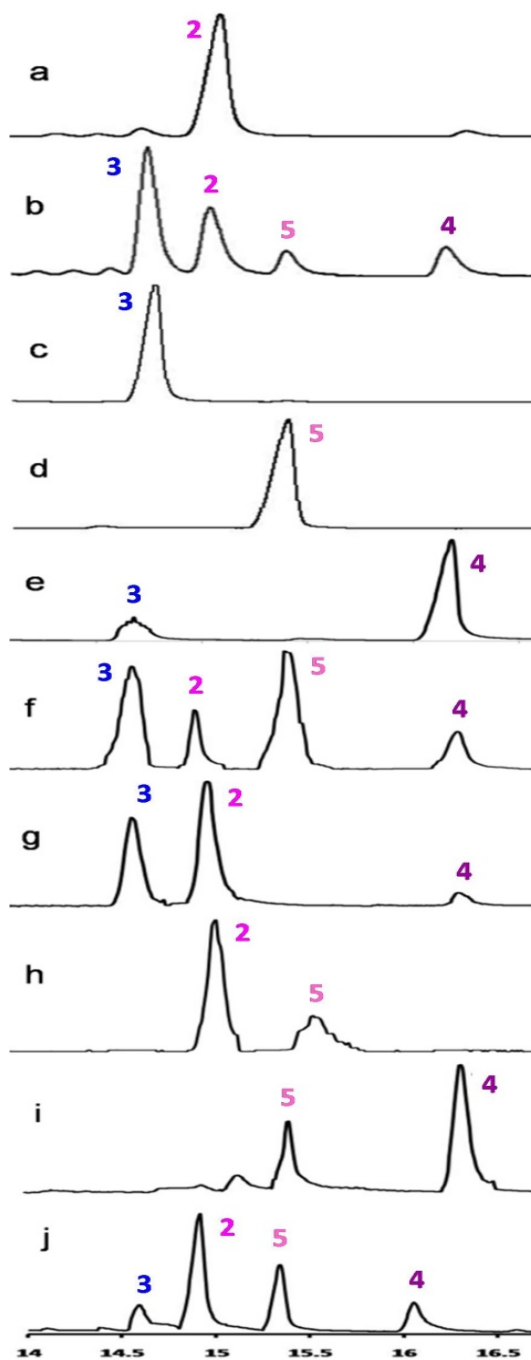


Figure 3.3.13: Total ion chromatograms (TICs) for 10-hydroxygeraniol dehydrogenase (Cr10HGO) catalyzed reactions with

- 10-hydroxygeraniol
- 10-hydroxygeraniol and NADP^+
- 10-oxogeraniol standard
- 10-oxogeraniol standard
- 10-hydroxygeraniol standard
- 10-oxogeraniol and NADP^+
- 10-oxogeraniol and NADPH
- 10-oxogeraniol and NADPH
- 10-oxogeraniol and NADPH
- 10-hydroxygeraniol and NADP^+

j) 10-hydroxygeraniol and NADPH.
The peaks represent (2) 10-hydroxygeraniol, (3) 10-oxogeraniol, (4) 10-hydroxygeraniol and (5) 10-oxogeraniol.

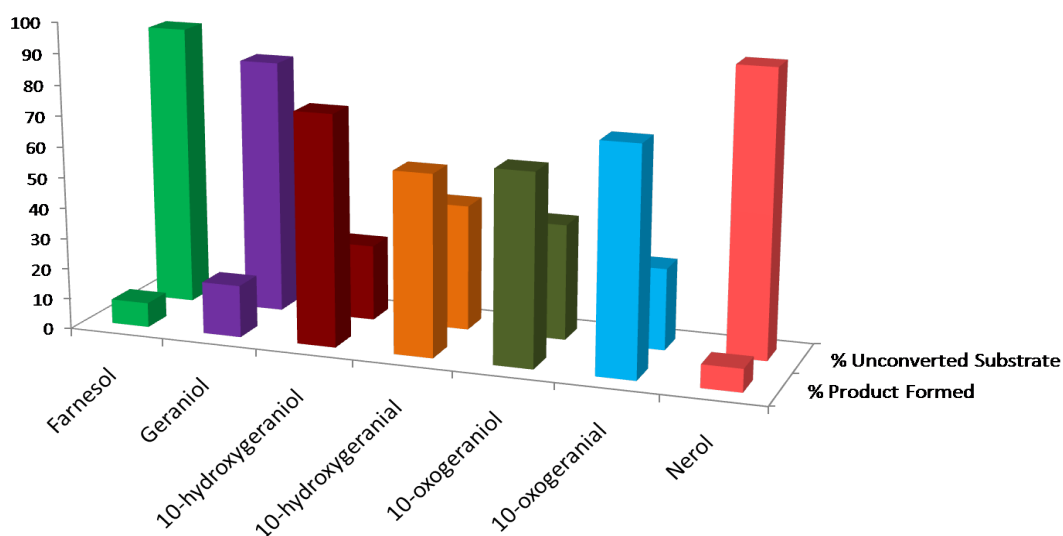


Figure 3.3.14: Product-ratio studies of Cr10HGO

3.3.4.4 Kinetic characterization of *Cr10HGO*

Table 3.3.1: Kinetic Constants for Cr10HGO with various substrates

Substrate	K_m (μM)	V_{max} ($\mu\text{M}/\text{sec}^{-1}$)	V_{max}/E_t (sec^{-1})
10-hydroxygeraniol	1.50	0.019	7.25
NADP ⁺	1.34	0.020	7.75
NAD ⁺	0.39	0.008	3.29
10-hydroxygeraniol	1.10	0.016	6.10
10-oxogeraniol	1.40	0.013	5.05
NADPH	1.07	0.014	5.65
10-oxogeraniol	1.32	0.014	5.44
10-hydroxynerol	1.35	0.013	4.90

The apparent K_m values were found to be 1.5 mM for 10-hydroxygeraniol, 1.0 mM for 10-oxogeraniol and 10-hydroxygeraniol at saturated concentrations of NADP⁺ (Figures 3.3.15–3.3.21). The kinetic studies also indicated that between NAD⁺/(H) and NADP⁺/(H), the latter was found to be the preferred coenzyme for Cr10HGO (Table 3.3.1).

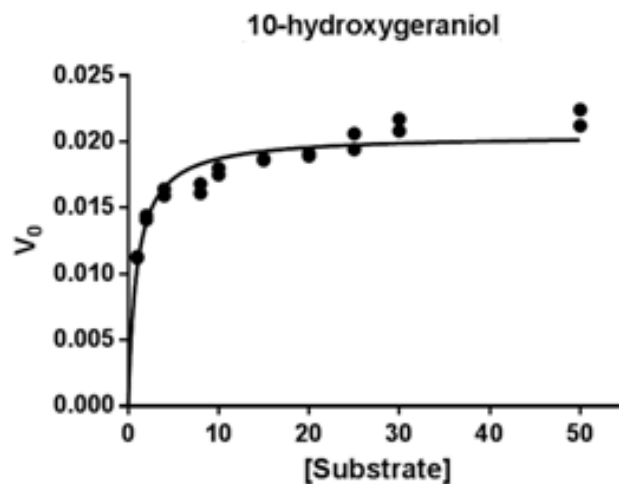


Figure 3.3.15: Michaelis-Menten Plot for 10-hydroxygeraniol kinetics of Cr10HGO

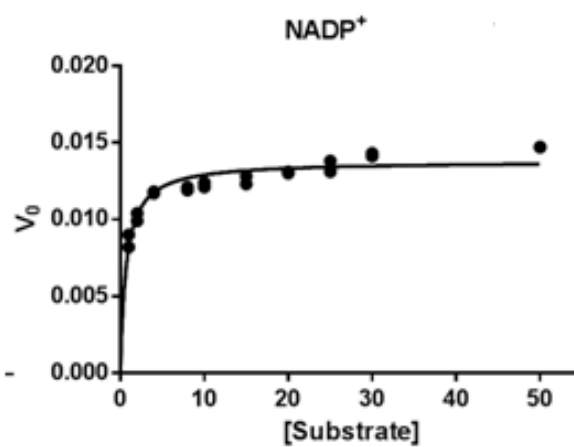


Figure 3.3.16: Michaelis-Menten Plot for NADP⁺ kinetics of Cr10HGO

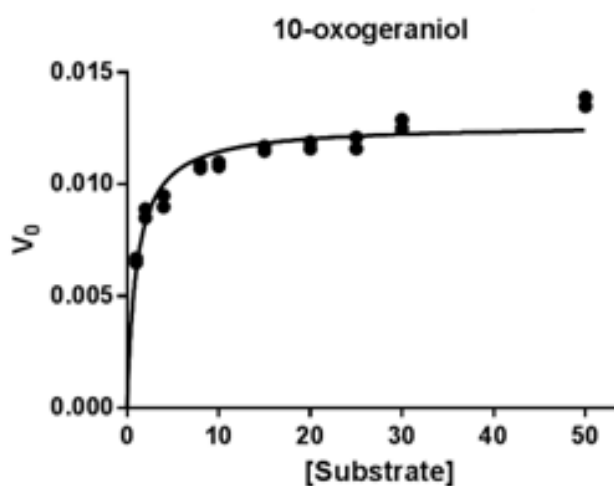


Figure 3.3.17: Michaelis-Menten Plot for 10-oxogeraniol kinetics of Cr10HGO

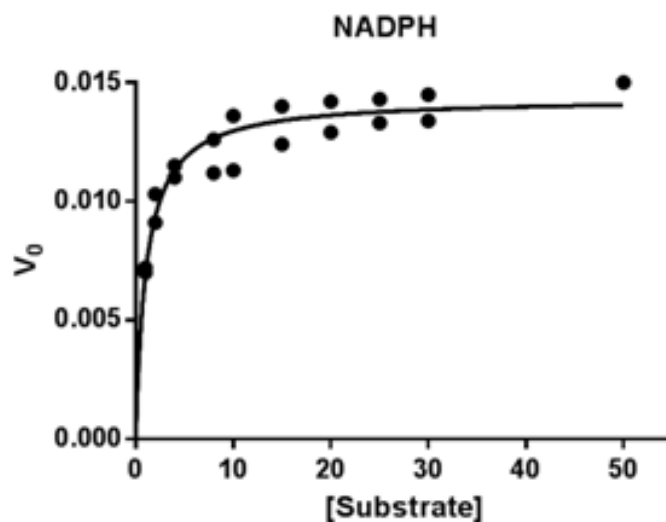


Figure 3.3.18: Michaelis-Menten Plot for NADPH kinetics of Cr10HGO

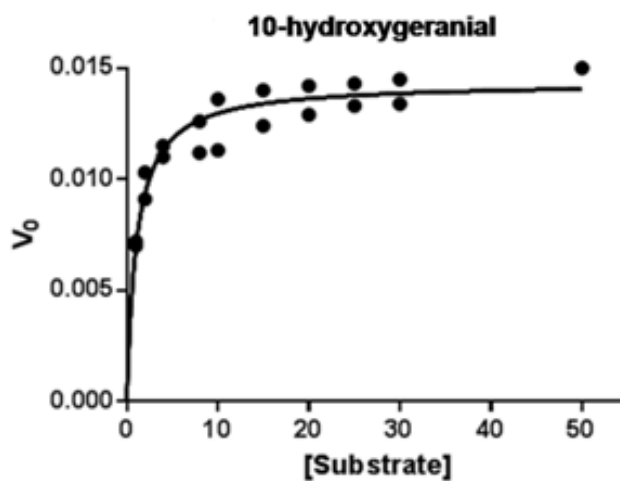


Figure 3.3.19: Michaelis-Menten Plot for 10-hydroxygeranial kinetics of Cr10HGO

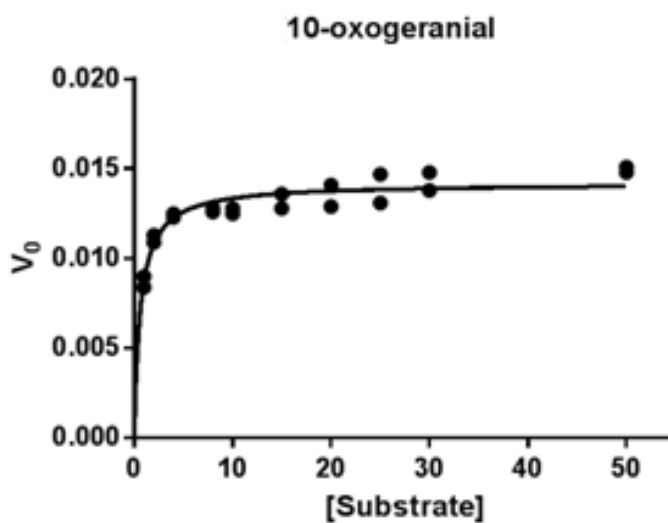


Figure 3.3.20: Michaelis-Menten Plot for 10-oxogeranial kinetics of Cr10HGO

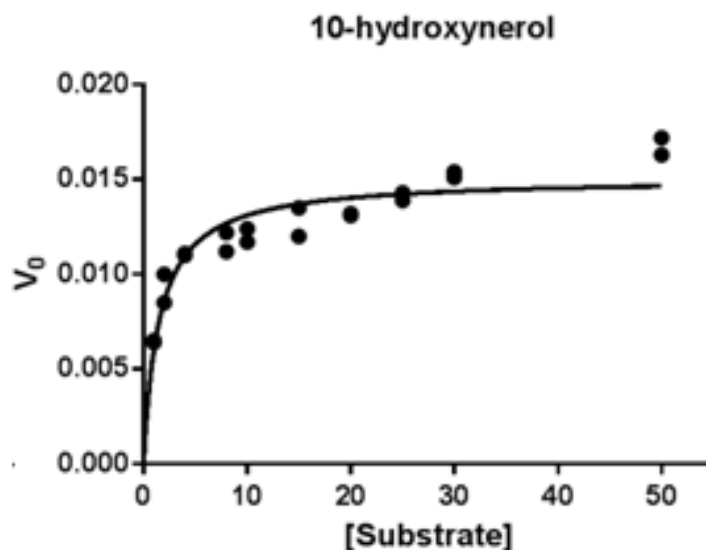


Figure 3.3.21: Michaelis-Menten Plot for 10-hydroxynerol kinetics of Cr10HGO

3.3.5 Isolation, cloning, expression and characterization of Iridoid synthase (*CrIDS*)

3.3.5.1 Isolation and cloning of Iridoid synthase (*CrIDS*)

Full length cDNA sequence of *CrIDS* unigene was obtained by using the primers designed from the transcript number 2280. The ORF of *CrIDS* (KF56146) composed of 1161 bp (Figure 3.3.22) encoding a polypeptide of 386 amino acids, with a calculated molecular weight of 43.69 KDa.

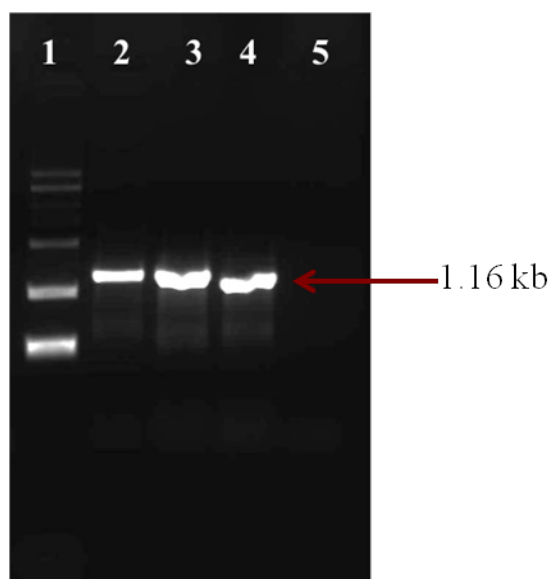


Figure 3.3.22: *CrIDS* full length ORF amplification, **Lane 1:** 1 Kb DNA ladder (Addendum Figure A1), **Lane 2:** PCR of *CrIDS* full-length primers with Root cDNA, **Lane 3:** PCR of *CrIDS* full-length primers with Leaf cDNA, **Lane 4:** PCR of *CrIDS* full-length primers with Stem cDNA, **Lane 5:** Negative Control.

While cloning of this gene was in progress, a report on the same gene from the same source appeared¹³. The polypeptide sequence of CrIDS showed high sequence similarity with the reported IDS¹³ but was shorter by six nucleotides (nucleotides from 47–52 missing), thereby causing a deletion of two amino acids (at positions 16 and 17), Pro and Asn (Figure 3.3.23). In addition to this major change, the amino acids at a few other positions were changed (N27S, V56G, A71V, N87D, I198D). Nevertheless, this shorter CrIDS efficiently carried out the reductive cyclization of 10- oxogeranial into four compounds. The reason these changes did not effect the substrate specificity of the two enzymes may be that these particular amino acids do not have a close interaction with the substrate / co-factor, and hence, no change in activity.

CrIDS	1	MSWWWKRSIGAGKNL--QNKENGVC	KNYKSVALVVGVTGIVGSSSLAEVLK
IDS	1	MSWWWKRSIGAGKLNLPNQNKENGVC	KSYSKSVALVVGVTGIVGSSSLAEVLK
CrIDS	49	LPDTPVGPWKVYGVARRPCPAW	LAKKPVEYIQCDVSNNOETISKLSPLKD
IDS	51	LPDTPGGPWKVYGVARRPCPV	WLAKKPVEYIQCDVSDNNOETISKLSPLKD
CrIDS	99	ITHIFYVSWIGSEDCQTNATMFKNILNSVIPNASNLQHVCLQTGIKHYFG	
IDS	101	ITHIFYVSWIGSEDCQTNATMFKNILNSVIPNASNLQHVCLQTGIKHYFG	
CrIDS	149	IFEEGSKVVPHDSPFTEDLPRLNVPNFYHDLEDILYEETGKNNLTWSVHR	
IDS	151	IFEEGSKVVPHDSPFTEDLPRLNVPNFYHDLEDILYEETGKNNLTWSVHR	
CrIDS	199	PALVFGFSPCSMMNIVSTLCVYATICKHENKALVYPGSKNSWNCYADA	VAVD
IDS	201	PALVFGFSPCSMMNIVSTLCVYATICKHENKALVYPGSKNSWNCYADA	VAVD
CrIDS	249	ADLVAEHEIWA AVDPKAKNQVLNCNNGDVFKWKHIWKKLAEFGIEMVGY	
IDS	251	ADLVAEHEIWA AVDPKAKNQVLNCNNGDVFKWKHIWKKLAEFGIEMVGY	
CrIDS	299	VEGKEQVSLAELMKDKDQVWDEIVKKNLVP TKLKEIAAFWFADIAFCSE	
IDS	301	VEGKEQVSLAELMKDKDQVWDEIVKKNLVP TKLKEIAAFWFADIAFCSE	
CrIDS	349	NLISSMNSKELGFLGFRNSMKS FVSCIDKMRDYRFIP	
IDS	351	NLISSMNSKELGFLGFRNSMKS FVSCIDKMRDYRFIP	

Figure 3.3.23: Box-shade analysis of CrIDS protein sequences

The ORF of *CrIDS* was cloned in pET 32a vector frame with N terminal His₆ tag for affinity purification under the control of T7-RNA polymerase promoter for expression of active protein in *E.coli* Rosetta2 (DE3) cells. Positive clones were screened by colony PCR, followed by sequencing with T7 promoter and T7 reverse primer.

3.3.5.2 Bacterial expression and protein purification

Expression was carried out in Rosetta2 DE3. The recombinant protein of CrIDS in pET32a was expressed and purified as discussed earlier. The purified

protein fractions gave a band at 61.5 kDa on a 12 % SDS-PAGE (Figure 3.3.24). The purified fractions were desalted, flash-frozen and store at -80 °C till used.

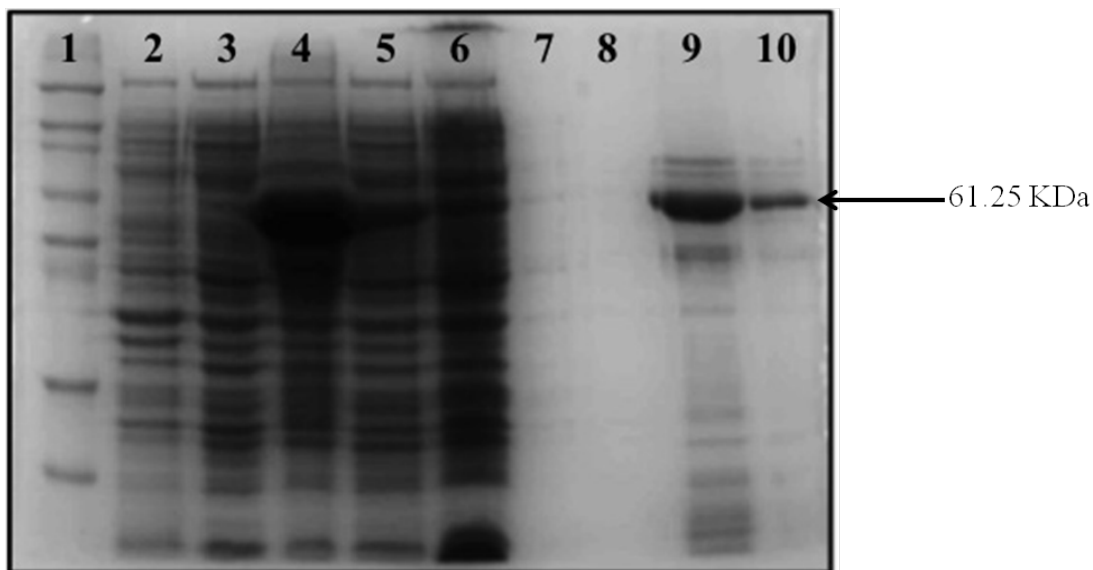
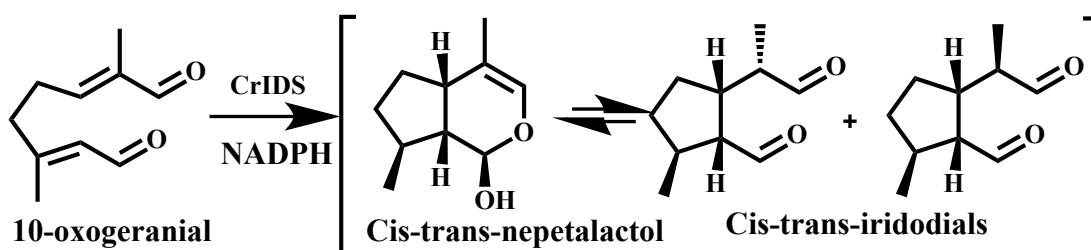


Figure 3.3.24: SDS-PAGE for CrIDS protein purification in pET 32a, **Lane 1:** BenchMark™ Protein Ladder (Addendum Figure A2), **Lane 2:** Uninduced Pellet fraction, **Lane 3:** Uninduced Crude lysate (supernatant fraction), **Lane 4:** Induced Pellet fraction, **Lane 5:** Induced Crude lysate (supernatant fraction), **Lane 6:** Unbound fraction, **Lanes 7-8:** Wash fractions, **Lanes 9-10:** Elution fractions.

3.3.5.3 Enzymatic characterization of *CrIDS*



Scheme 3.3.5: Schematic representation of *CrIDS* assay

CrIDS, in the presence of NADPH, carries out the cyclization of 10-oxogeranial to an equilibrium mixture of *cis-trans*-nepetalactol and iridodials (Scheme 3.3.5, Figure 3.3.25). Iridoids belong to a class of plant secondary metabolites. They normally form an integral part of plant defense mechanisms. They are also produced in some animals/ insects as part of their defense mechanisms and are characterized by a bitter taste. Iridoids are known to exhibit wide range of pharmacological activities like anti-tumor, anti viral, hypoglycemic, analgesic, etc. This product was characterized by mass fragmentation pattern.

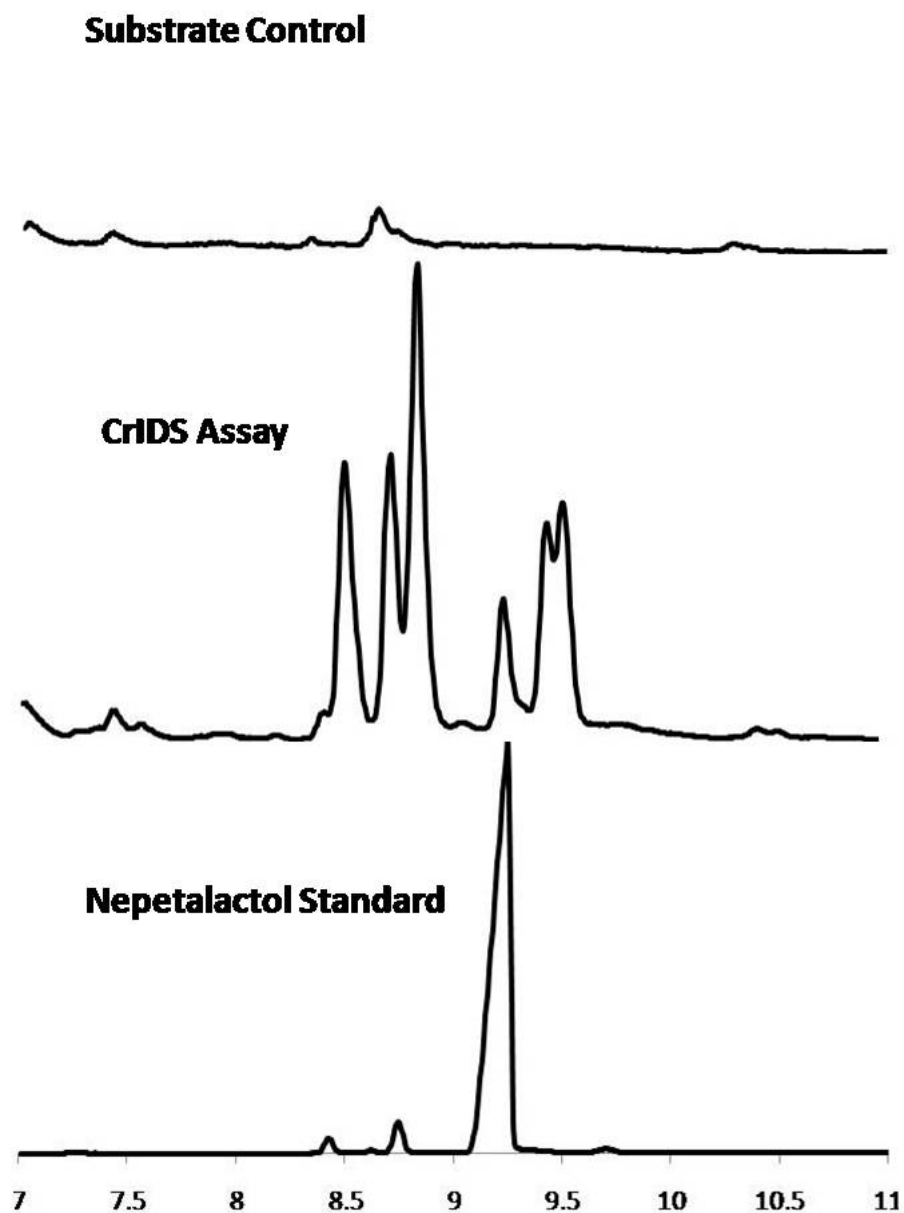
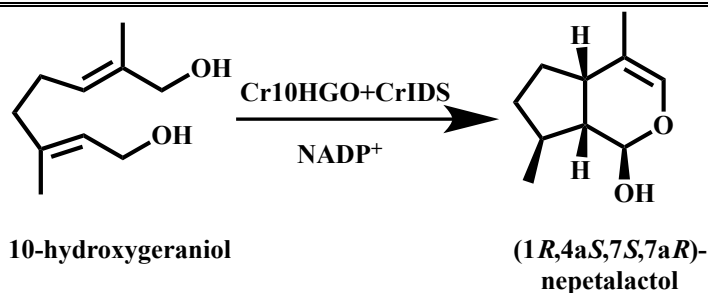


Figure 3.3.25: Total Ion Chromatogram of Assay of CrIDS with 10-oxogeraniol as substrate.

3.3.6 Cascaded enzyme activity

Surprisingly, when 10-hydroxygeraniol was incubated with Cr10HGO and CrIDS for 30 min, in the presence of NADP^+ , it yielded a major metabolite (Scheme 3.3.6).



Scheme 3.3.6: Concerted reaction of Cr10HGO and CrIDS on 10-hydroxygeraniol to form (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol.

The major metabolite (Rt: 9.2 min) (Figure 3.3.26) formed was identified as *cis-trans*-nepetalactol with stereochemistry 4*aS*, 7*S*, 7*aR* by comparing the retention time and co-injection studies in GC and GC-MS analyses with the synthesized diastereomeric mixture of *cis-trans*-nepetalactols [containing (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol as a major diastereomer]²²⁻²⁴ arising due to the asymmetry at carbon 1.

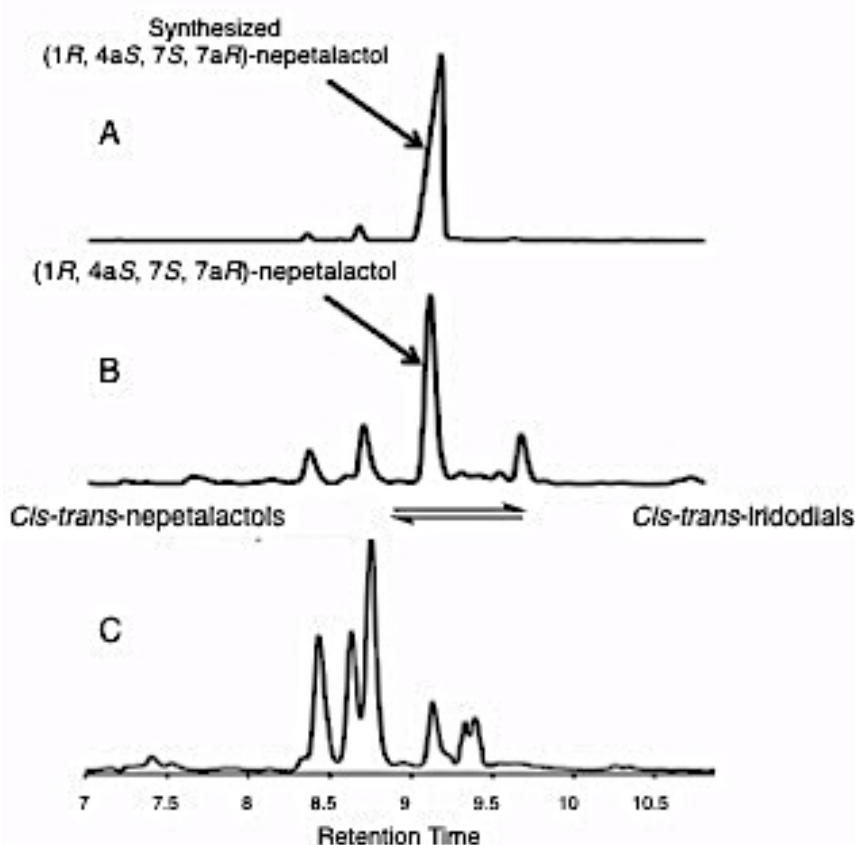


Figure 3.3.26: Concerted reaction of Cr10HGO and CrIDS. Comparison of TICs of A) Synthesized (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol, B) CH₂Cl₂ extract of combined assay of Cr10HGO and CrIDS with 10-hydroxygeraniol and NADP⁺ as substrates, and C) CH₂Cl₂ extract of CrIDS assay with 10-oxogeraniol and NADPH as substrates.

These two diastereomers were not resolved under the GC conditions even with chiral capillary column. However, the acetylated diastereomers were separated under similar GC conditions (Figure 3.3.27). The stereochemistry of the major enzymatic product formed was determined as (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol acetylation of assay mixture, followed by GC and GC-MS analyses and comparing the retention time and co-injection studies with that of acetylated derivatives of synthesized nepetalactols' mixture containing (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol as a major diastereomer²³.

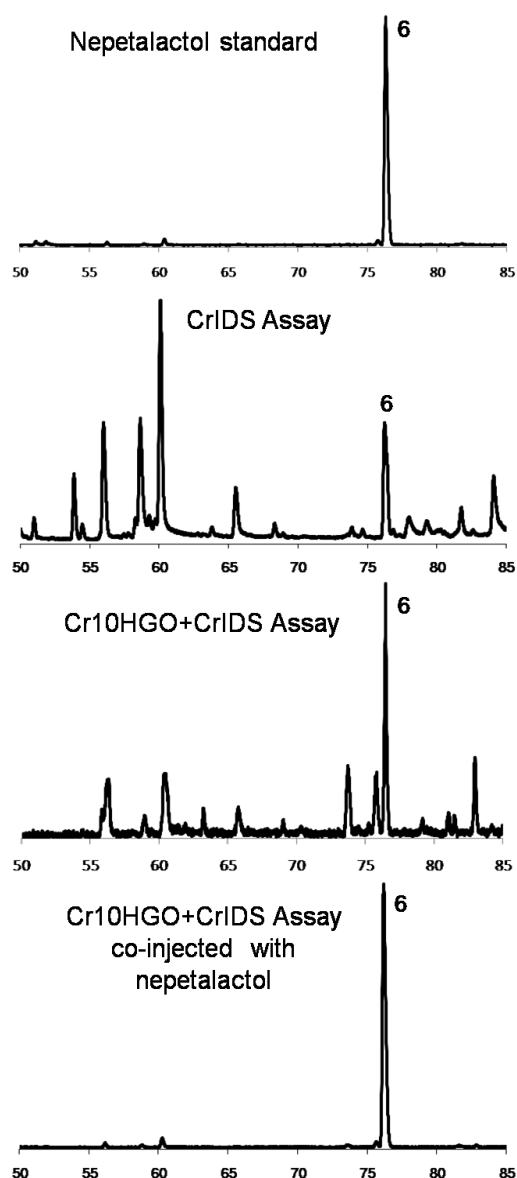
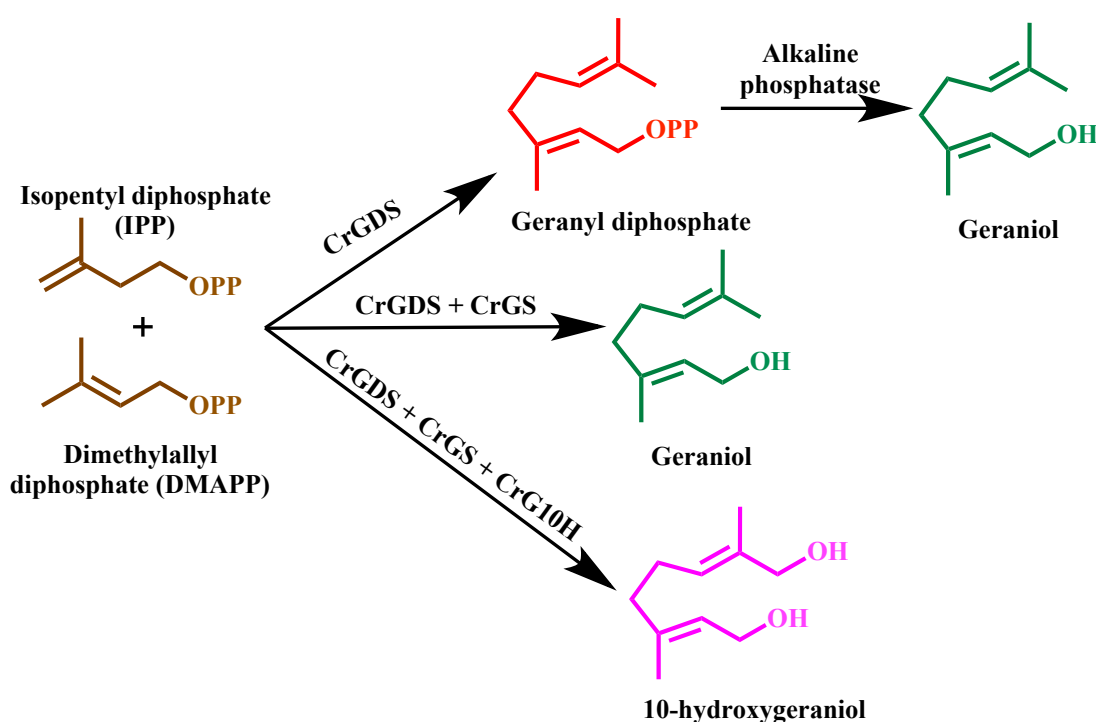


Figure 3.3.27: Comparison of TICs of CrIDS, Cr10HGO+CrIDS and Cr10HGO+CrIDS assays with standard (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol on Astec CHIRALDEX™ B-DA column; (6): (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol.

Further, when 10-hydroxygeraniol was incubated with Cr10HGO and CrIDS in the presence of NADP^+ for prolonged incubation beyond 30 min led to the formation of open structures of iridodials in equilibrium with *cis-trans*-nepetalactol (Figure 3.3.26). The formation of a major metabolite by the combined action of two enzymes clearly indicates the concerted enzymatic action of Cr10HGO and CrIDS in the formation of desired (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol, an important intermediate in iridoids and MIAs biosynthesis. As both Cr10HGO and CrIDS are cytoplasmic enzymes, presumably, the products of Cr10HGO [10-oxogeraniol and NAD(P)H] will be used by CrIDS to synthesize (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol, indicating a physiological enzyme cascade.



Scheme 3.3.7: Concerted reaction of CrGDS, CrGS and CrG10H

To understand the combinatorial action of these enzymes on their cognate substrates, we incubated dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP) with purified CrGDS and CrGS proteins and yeast expressed microsomal pellet containing CrG10H, in the presence of coenzymes, which yielded 10-hydroxygeraniol as the enzymatic product (Scheme 3.3.7, Figure 3.3.28).

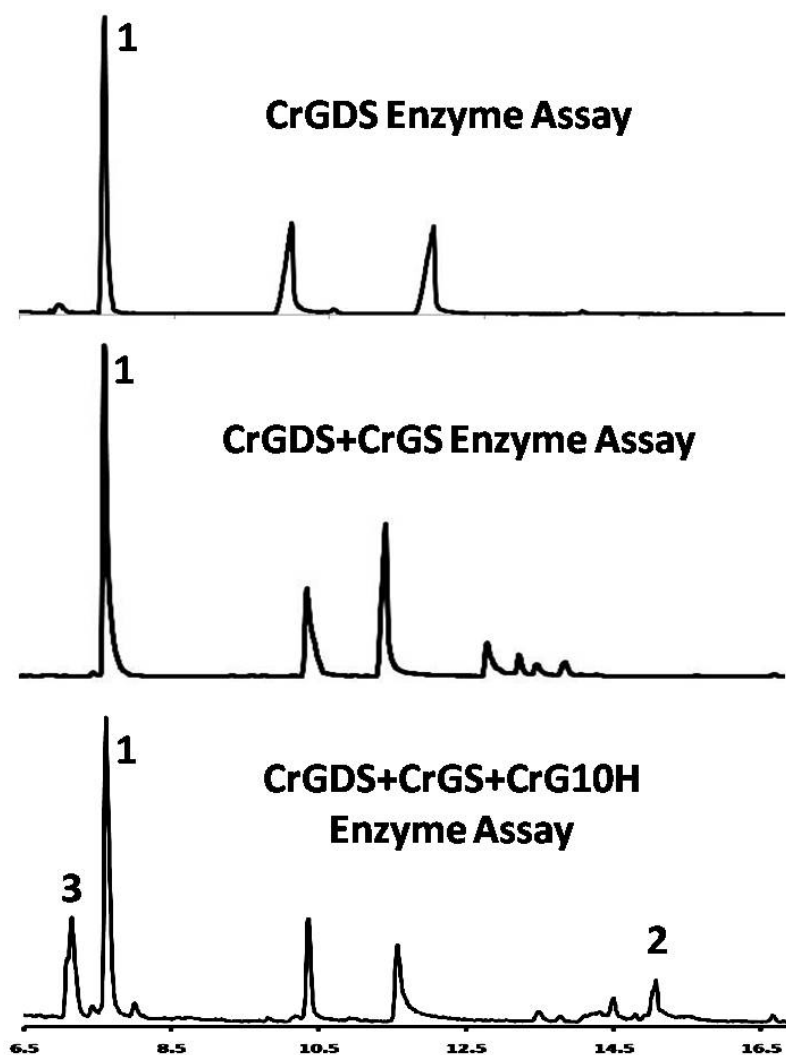
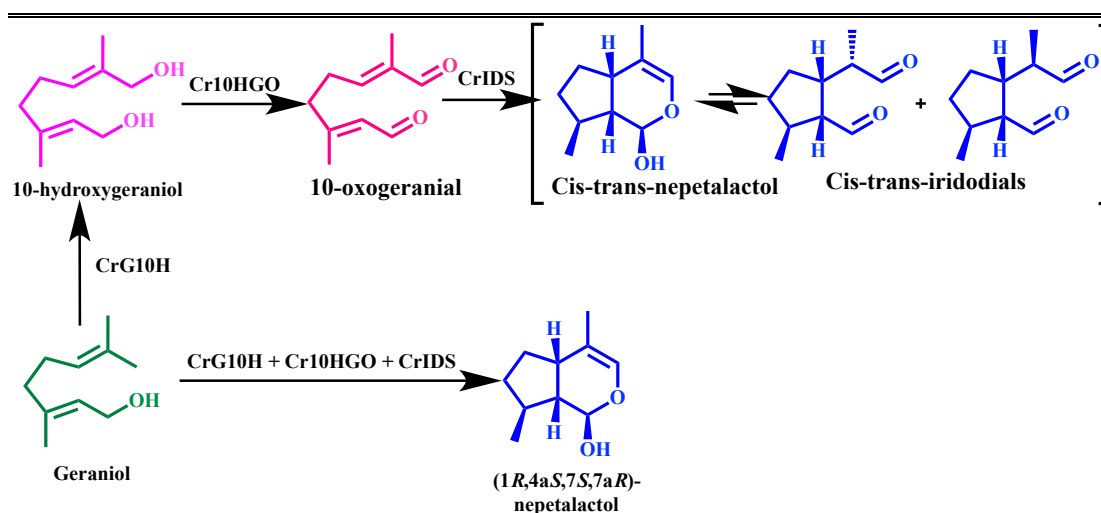


Figure 3.3.28: Comparison of TICs of CrGDS, CrGDS+CrGS and CrGDS+CrGS+CrG10H enzyme assays; (1):geraniol, (2):10-hydroxygeraniol, (3):nerol.

Furthermore, incubation of geraniol in combination with CrG10H, Cr10HGO and CrIDS revealed the formation of (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol by GC and GC-MS analyses (Scheme 3.3.8, Figure 3.3.28).



Scheme 3.3.8: Concerted reaction of CrG10H, Cr10HGO and CrIDS

These experiments suggest *in vivo* proximity and “cross talk” of the enzymes involved in the biosynthetic cascade leading to the formation of the desired product.

As CrG10H is localized on endoplasmic reticulum, the product of this enzyme, 10-hydroxygeraniol might be sequestered out to the subsequent enzymes of MIAs pathway. These observations are further supported by the studies on cellular localization of MIAs biosynthetic pathway enzymes in a particular cell type in *C. roseus*^{24–27}. It also appears that the formation of 10-hydroxygeraniol through CrG10H catalyzed reaction is the rate-limiting step for iridoid biosynthesis.

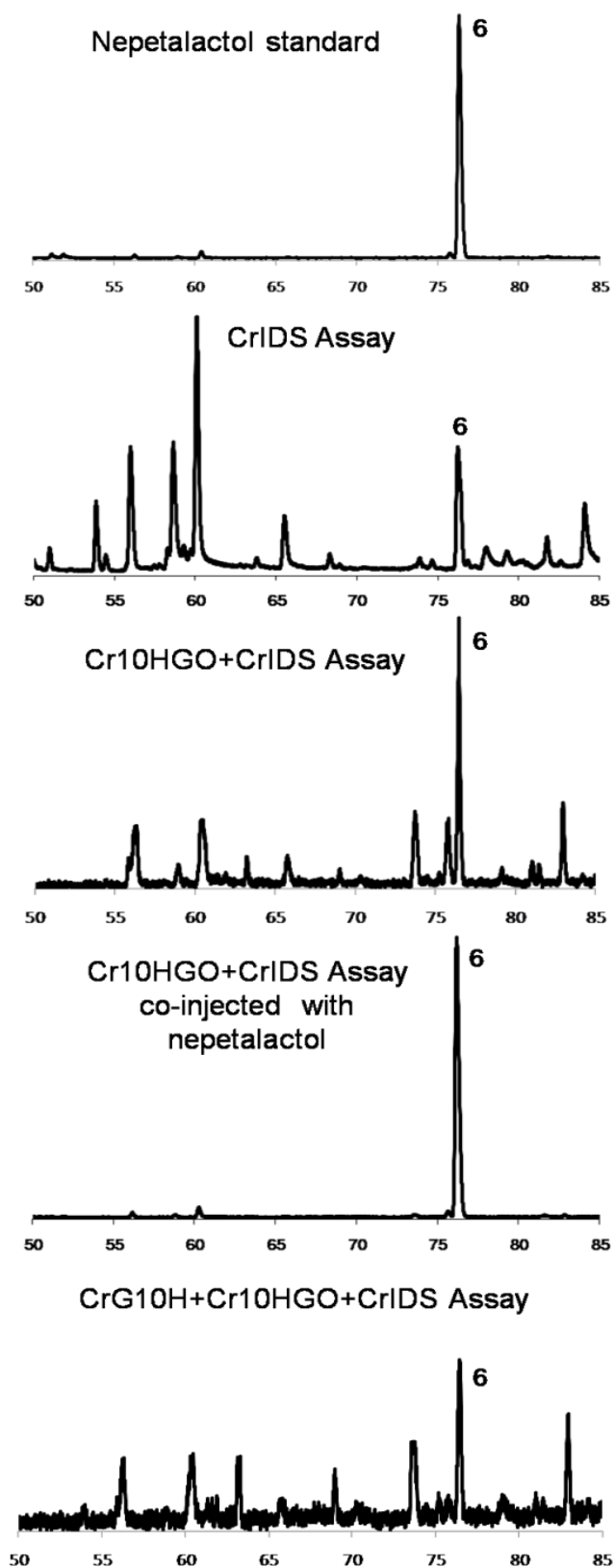


Figure 3.3.29: Comparison of TICs of CrG10H, CrG10H+Cr10HGO and CrG10H+Cr10HGO+CrIDS enzyme assays; (6): (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol.

3.3.7 Phylogenetic analysis of monoterpene synthases isolated from *Catharanthus roseus*

A neighbor joining phylogenetic tree placed all the full-length sequence of monoterpene synthase isolated from *C. roseus* in separate clades. CrIDS seems to have evolved from a family of Progesterone reductases, having a link with monoterpene alcohol synthases, from which CrGS showed lineage. CrGDS was placed in a separate clade for Terpene diphosphate synthases. While Cr10HGO was clustered with a family of alcohol dehydrogenases, CrG10H showed similarity with other geraniol 10-hydroxylases.

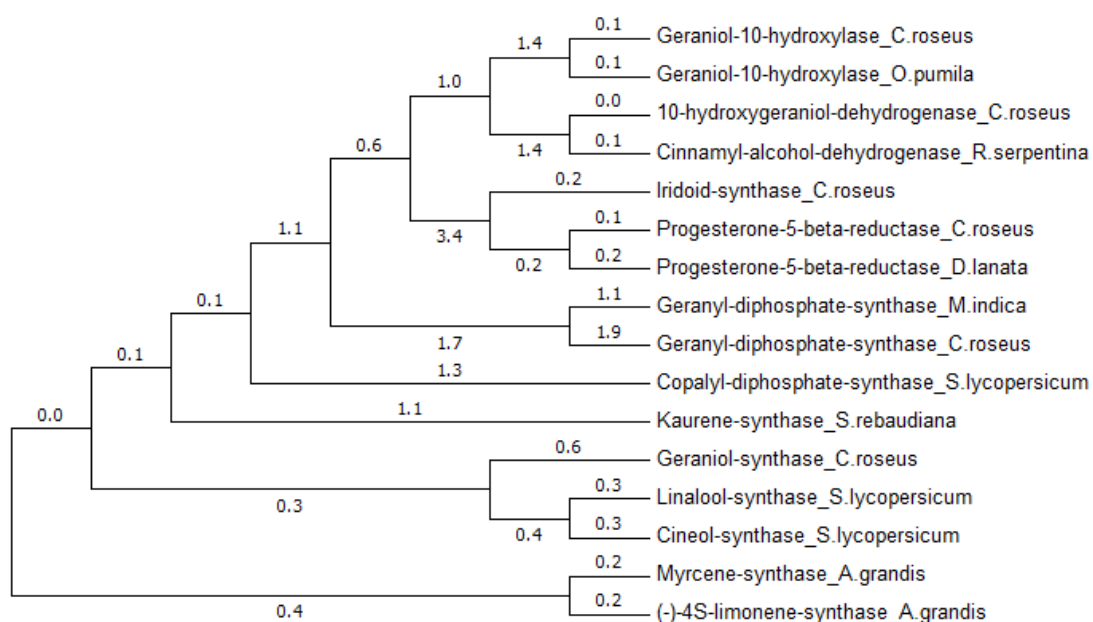
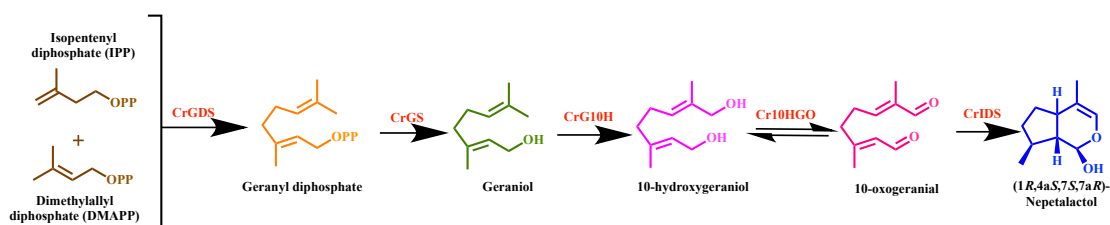


Figure 3.3.30: Phylogenetic analysis of the monoterpene synthases isolated from *C. roseus*; Sequences used for phylogenetic tree construction are: Iridoid synthase from *Catharanthus roseus* (AHA82033), Progesterone 5- β reductase from *Catharanthus roseus* (AIW09147), Progesterone 5- β reductase from *Digitalis purpurea* (ACZ66261), Geranyl diphosphate synthase from *Mangifera indica* (AFJ52721), Geraniol synthase from *Catharanthus roseus* (AHA82032), Copalyl diphosphate synthase from *Solanum lycopersicum* (BAA84918), Geranyl diphosphate synthase from *Catharanthus roseus* (AHA82035), Kaurene synthase from *Stevia rebaudiana* (AAD34295), Cinnamyl alcohol dehydrogenase from *Rauvolfia serpentina* (ALW82980), 10-hydroxygeraniol dehydrogenase from *Catharanthus roseus* (AHA82031), Linalool synthase from *Solanum lycopersicum* (NP_001233805), Geraniol 10-hydroxylase from *Catharanthus roseus* (AHA82034), Geraniol 10-hydroxylase from *Ophiorrhiza pumila* (BAP90522), Cineol synthase from *Solanum lycopersicum* (AEM05857), Myrcene synthase from *Abies grandis* (AAB71084), (-)-4S-limonene synthase from *Solanum lycopersicum* (AAB70907).

Although the terpene synthases from *Abies grandis* fall under the class of monoterpene synthases of secondary metabolite origin, they are quite distinct from the other monoterpene synthases represented in the phylogenetic tree due to the divergence between angiosperms and gymnosperms²⁵. It is due to this bifurcation that the gymnosperm monoterpene synthases are more closely related to each other than to their counterpart angiosperms. Copalyl diphosphate synthase and kaurene synthase are genes of the primary metabolism seem to be expressed as single copy genes²⁶ and hence are placed in separate clades. They have very low similarity to the five genes of *C. roseus*.

3.4 Conclusion

Cloning and functional characterization of 10-hydroxygeraniol dehydrogenase (Cr10HGO) system from *C. roseus* indicated that Cr10HGO showed broad substrate specificity for 10-hydroxygeraniol, 10-oxogeraniol or 10-hydroxygeranial over monohydroxy linear terpene derivatives. Concerted enzymatic function in the biosynthesis of *cis-trans*-nepetalactol has been demonstrated using 10-hydroxygeraniol and NADP⁺ with Cr10HGO and CrIDS combined assay system. The stereochemistry of the enzymatic product was determined and is (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol, which is a key intermediate in the biosynthesis of iridoids and MIAs. Further, we have demonstrated the *in vitro* formation of (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol when geraniol was incubated with CrG10H, Cr10HGO and CrIDS.



Scheme 3.3.9: Concerted reaction of CrGDS, CrGS, CrG10H, Cr10HGO and CrIDS

3.5 Appendix: Agarose gel electrophoresis for colony PCR screening for gene cloning

3.5.1 Cloning of *CrGDS* in pET 28a vector frame

Colony PCR screening of geranyl diphosphate synthase cloned in pET 28a vector frame with T7 promoter primer and T7 reverse primer.

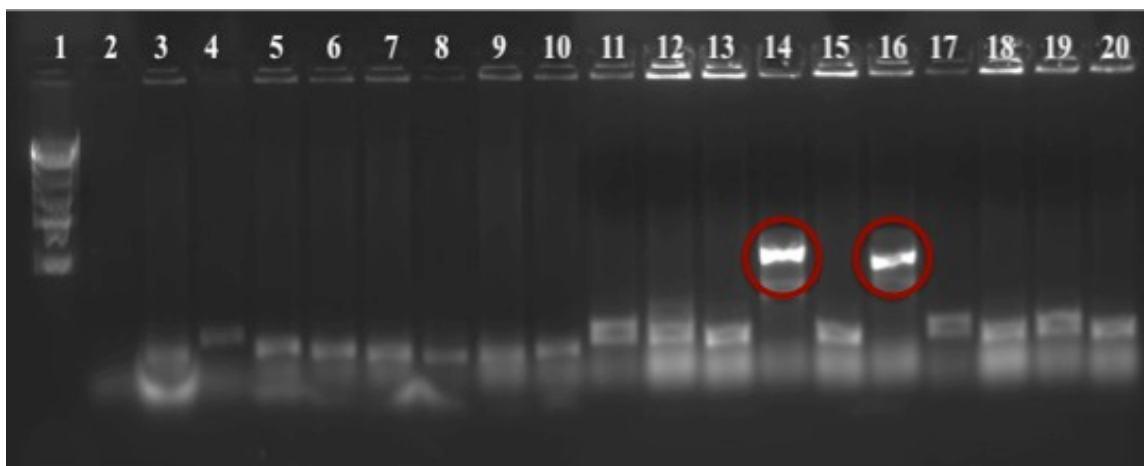


Figure 3.5.1: Colony PCR screening for *CrGDS* truncated ORF cloned in pET 28a on 1 % agarose gel, **Lane 1:** 1 Kb DNA ladder (Addendum Figure A1), **Lane 2:** negative control, **Lanes 3-20:** PCR with T7 promoter and T7 reverse primer (encircled is band for 966 bp).

3.5.2 Cloning of *CrGS* in pET 28a vector frame

Colony PCR screening of geraniol synthase cloned in pET 28a vector frame with T7 promoter primer and T7 reverse primer.



Figure 3.5.2: Colony PCR screening of full-length ORF of *CrGS* cloned in pET 28a on 1 % agarose gel, **Lane 1:** 1 Kb DNA ladder (Addendum Figure A1), **Lane 2:** negative control, **Lanes 3-6:** PCR with T7 promoter and T7 reverse primer

3.5.3 Cloning of *CrG10H* in pYES2 vector

Colony PCR screening of geraniol 10-hydroxylase cloned in pYES2 vector frame with T7 promoter primer and CrG10H reverse primer.

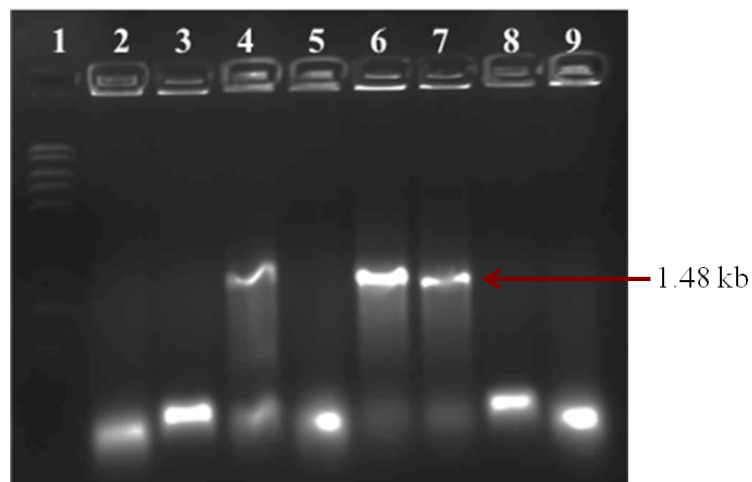


Figure 3.5.3: Colony PCR Screening for *CrG10H* full length ORF cloned in pYES2 on 1 % agarose gel, **Lane 1:** 1 Kb DNA ladder (Addendum Figure A1), **Lane 2:** negative control, **Lanes 3-9:** PCR with T7 promoter primer and CrG10H reverse primer

3.5.4 Cloning of ORF of *Cr10HGO* in pRSET B vector frame

Colony PCR screening of 10-hydroxygeraniol dehydrogenase, cloned in pRSET B vector frame with T7 promoter and T7 reverse primers.

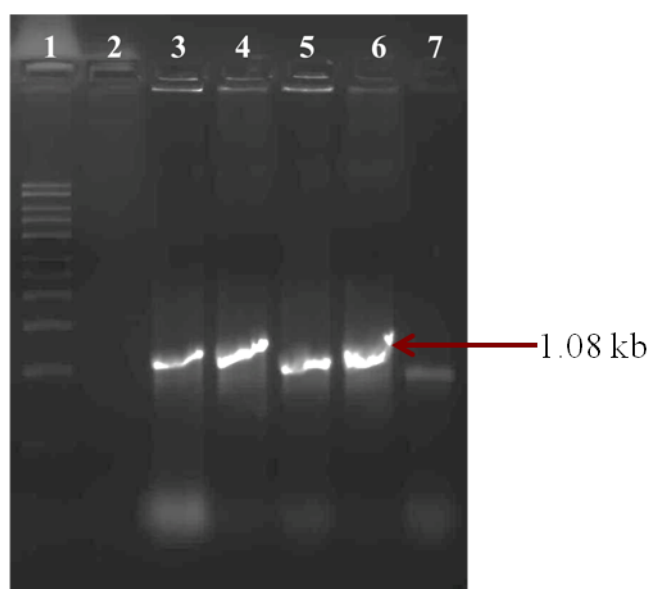


Figure 3.5.4: Colony PCR screening for full-length ORF cloning of *Cr10HGO* in pRSETB vector on 1 % agarose gel, **Lane 1:** 1 Kb DNA ladder (Addendum Figure A1), **Lane 2:** no insert control **Lanes 3-7:** PCR with T7 promoter and T7 reverse primer

3.5.5 Cloning of ORF of *CrIDS* in pET 32a

Colony PCR screening of ORF of *CrIDS* cloned in pET 32a vector frame with T7 promoter and T7 reverse primers.

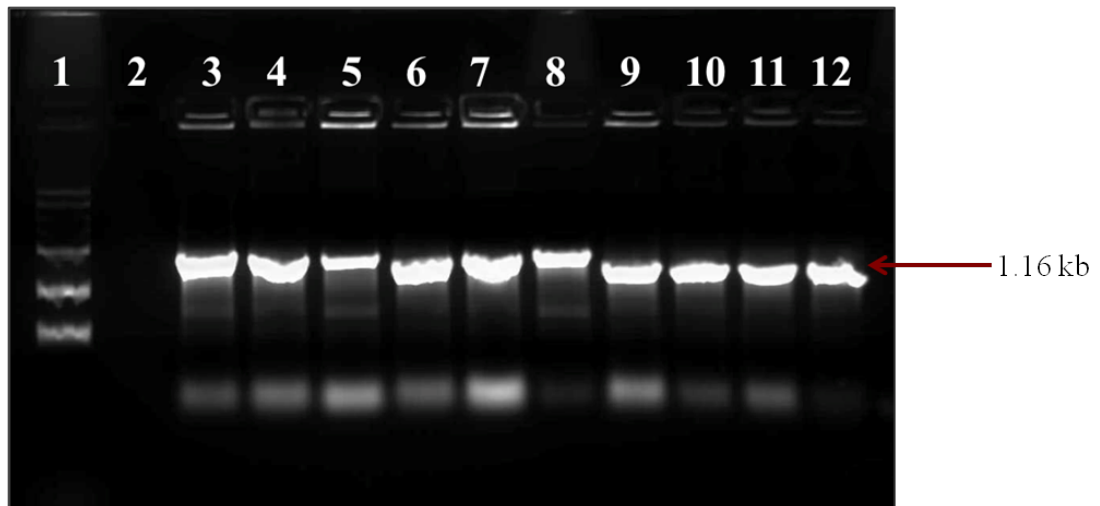


Figure 3.5.5: Colony PCR screening for ORF of *CrIDS* cloned in pET 32a vector on 1 % agarose gel, **Lane 1:** 1 Kb DNA ladder (Addendum Figure A1), **Lane 2:** no insert control, **Lanes 3-12:** PCR amplicon with T7 promoter and T7 reverse primer

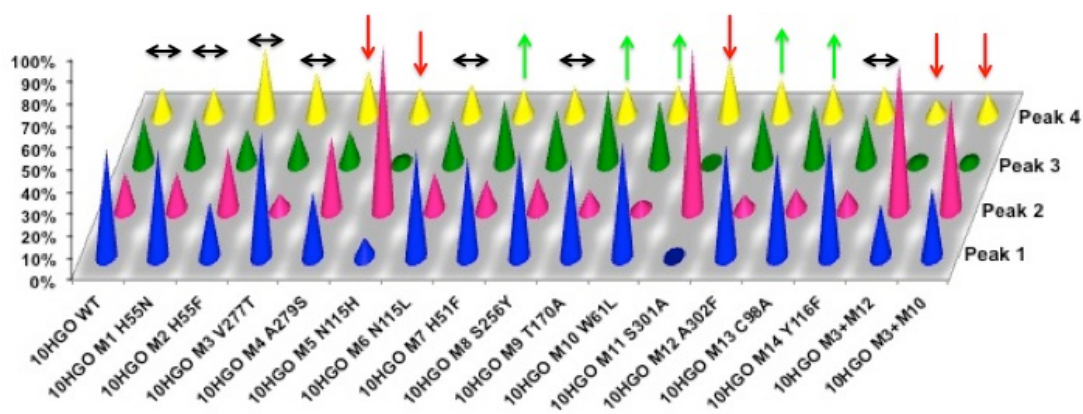
3.6 References

1. Hays, S.B. Some effects of reserpine, a tranquilizer, on House Fly. *J. Econ. Entomol.* **58**, 782-783 (1965).
2. Yui, T. & Takeo, Y. Neuropharmacological studies on a new series of ergot alkaloids; elymoclavine as a potent analeptic on reserpine-sedation. *Jpn. J. Pharmacol.* **7**, 157-161 (1958).
3. Yui, T. & Takeo, Y. Neuropharmacological studies on a new series of ergot alkaloids; the effects on electrocorticogram of rabbits. *Jpn. J. Pharmacol.* **7**, 162-168 (1958).
4. Shurin, G.V., Tourkova, I.L., Kaneno, R. & Shurin, M.R. Chemotherapeutic Agents in Noncytotoxic Concentrations Increase Antigen Presentation by Dendritic Cells via an IL-12-Dependent Mechanism. *J. Immunol.* **183**, 137-144 (2009).
5. van der Heijden, R., Jacobs, D.I., Snoeijer, W., Hallared, D. & Verpoorte, R. The Catharanthus alkaloids: Pharmacognosy and biotechnology. *Curr. Med. Chem.* **11**, 607-628 (2004).
6. Rai, A., Smita, S.S., Singh, A.K., Shanker, K. & Nagegowda, D.A. Heteromeric and Homomeric Geranyl Diphosphate Synthases from *Catharanthus roseus* and Their Role in Monoterpene Indole Alkaloid Biosynthesis. *Mol. Plant* **6**, 1531-1549 (2013).
7. Simkin, A.J., Miettinen, K., Claudel, P., Burlat, V., Guirimand, G., Courdavault, V., Papon, N., Meyer, S., Godet, S., St-Pierre, B. & Giglioli-Guivarc'h, N. Characterization of the plastidial geraniol synthase from Madagascar periwinkle which initiates the monoterpene branch of the alkaloid pathway in internal phloem associated parenchyma. *Phytochemistry* **85**, 36-43 (2013).
8. Madyastha, K.M., Meehan, T.D. & Coscia, C.J. Characterization of a cytochrome P-450 dependent monoterpene hydroxylase from the higher plant *Vinca rosea*. *Biochemistry* **15**, 1097-1102 (1976).
9. Herbert, R.B. The biosynthesis of plant alkaloids and nitrogenous microbial metabolites. *Nat. Prod. Rep.* **2**, 163-179 (1985).
10. Uesato, S., Matsuda, S., Iida, A., Inouye, H. & Zenk, M.H. Intermediacy of 10-hydroxygeraniol, 10-hydroxynerol and iridodial in the biosynthesis of ajmaline and vomilenine in *Rauwolfia serpentina* suspension cultures. *Chem. Pharm. Bull.* **32**, 3764-3767 (1984).
11. Uesato, S., Matsuda, S. & Inouye, H. Mechanism for iridane skeleton formation from acyclic monoterpenes in the biosynthesis of secologanin and vindoline in *Catharanthus roseus* and *Lonicera morrowii*. *Chem. Pharm. Bull.* **32**, 1671-1674 (1984).
12. Uesato, S., Ueda, S., Kobayashi, K. & Inouye, H. Mechanism of iridane skeleton formation in the biosynthesis of iridoid glucosides in *Gardenia jasminoides* cell cultures. *Chem. Pharm. Bull.* **31**, 4185-4188 (1983).
13. Geu-Flores, F., Sherden, N. H., Courdavault, V., Burlat, V., Glenn, W. S., Wu, C., Nims, E., Cui, Y., & O'Connor, S. E. An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature* **492**, 138-142 (2013).
14. Ikeda, H., Esaki, N., Nakai, S., Hashimoto, K., Uesato, S., Soda, K., & Fujita, T. Acyclic monoterpene primary alcohol - NADP⁺ oxidoreductase of *Rauwolfia*

-
- serpentina* cells - the key enzyme in biosynthesis of monoterpene alcohols. *J. Biochem.* **109**, 341-347 (1991).
15. Saitou, N. & Nei, M. The neighbor-joining method - a new method for reconstructing phylogenetic trees. *Jpn. J Genet.* **61**, 611-611 (1986).
 16. Saitou, N. & Nei, M. The neighbor-joining method - a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406-425 (1987).
 17. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **30**, 2725-2729 (2013).
 18. Zuckerkandl, E. & Pauling, L. Evolutionary divergence and convergence in proteins. *New York: Academic Press* **97**, 97-166 (1965).
 19. Teoh, K.H., Gorman, E.B. & McKnight, T.D. Characterization and cloning of 10-hydroxygeraniol oxidoreductase. Vol. 2012 (2000).
 20. Hallahan, D.L., West, J.M., Wallsgrove, R.M., Smiley, D.W., Dawson, G.W., Pickett, J.A. & Hamilton, J.G. Purification and Characterization of an Acyclic Monoterpene Primary Alcohol-NADP(1) Oxidoreductase from Catmint (*Nepeta racemosa*). *Arch. Biochem. Biophys.* **318**, 105-112 (1995).
 21. Miettinen, K., Dong, L., Navrot, N., Schneider, T., Burlat, V., Pollier, J., Woittiez, L., Van Der Krol, S., Lugan, R., Ilc, T. & Verpoorte, R. The seco-iridoid pathway from *Catharanthus roseus*. *Nature Comm.* **5**, 3606-3617 (2014).
 22. Beckett, J.S., Beckett, J.D. & Hofferberth, J.E. A Divergent Approach to the Diastereoselective Synthesis of Several Ant-Associated Iridoids. *Org. Lett.* **12**, 1408-1411 (2010).
 23. Dawson, G.W., Pickett, J.A. & Smiley, D.W.M. The aphid sex pheromone cyclopentanoids: Synthesis in the elucidation of structure and biosynthetic pathways. *Bioorg. Med. Chem.* **4**, 351-361 (1996).
 24. Fairlamb, I.J.S., Dickinson, J.M. & Pegg, M. Selenium dioxide E-methyl oxidation of suitably protected geranyl derivatives - synthesis of farnesyl mimics. *Tetrahedron Lett.* **42**, 2205-2208 (2001).
 25. Bohlmann, J., Meyer-Gauen, G. & Croteau, R. Plant terpenoid synthases: Molecular biology and phylogenetic analysis. *Proc. Natl. Acad. Sci. USA* **95**, 4126-4133 (1998).
 26. Bohlmann, J., Phillips, M., Ramachandiran, V., Katoh, S. & Croteau, R. cDNA cloning, characterization, and functional expression of four new monoterpene synthase members of the Tpsd gene family from grand fir (*Abies grandis*). *Arch. Biochem. Biophys.* **368**, 232-243 (1999).
-

Chapter 4

Dissection of the activity of 10-hydroxygeraniol dehydrogenase (Cr10HGO) by site-directed mutagenesis studies



Chapter 4

Dissection of the activity of 10-hydroxygeraniol dehydrogenase (Cr10HGO) by site-directed mutagenesis studies

Catharanthus roseus 10-hydroxygeraniol dehydrogenase (Cr10HGO) catalyzes a crucial step in the biosynthesis of Monoterpene Indole Alkaloids (MIAs), resulting in the formation of 10-oxogeranial via two intermediates, 10-hydroxygeranial and 10-oxogerniol, in the presence of NADP⁺ in a reversible reaction. In order to understand the structural basis for the catalysis of Cr10HGO, homology model was built and molecular docking studies were performed using substrate and intermediates. We selected the amino acids residues lining the first tier of active site pocket, which might play a crucial role in the catalysis of Cr10HGO. Mutation of the residue N115 to H resulted in a notable decrease in substrate conversion, which could be due to premature termination of the enzymatic reaction. Another mutation at S301, where serine was replaced with alanine, also caused a substantial drop in substrate conversion such that 3/4th of the substrate was left unutilized and the remaining 1/4th constituted of only one of the intermediates. Mutation at W61 resulted in the lower levels of the substrate, 10-hydroxygeraniol, indicating suppression of the reverse reaction. Mutation at the T170 position resulted in an increase in the formation of the dialdehyde, 10-oxogeranial, which serves as a substrate for the next enzyme in the MIA biosynthetic pathway. These residues seem to be playing a major role in the stabilization of the carbocation intermediates. This study, thus, provides an insight into the residues that affect the catalytic activity of Cr10HGO and also provides a gateway for increasing the production of intermediates in the MIA pathway, which, in turn, would cause an increase in the biosynthesis of the final product of the pathway.

4.1 Introduction

Catharanthus roseus is an evergreen shrub belonging to the family Apocynaceae. It is a rich source of the monoterpene indole alkaloids, many of which possess considerable pharmacological and therapeutic interests such as the tranquillizer reserpine¹ or the ergot alkaloids¹⁻³ employed for their relaxing and migraine relieving properties, the biggest breakthrough being the discovery of vincristine and vinblastine as anti-cancer drugs^{4,5}. These anti-cancer compounds are produced via the Monoterpene Indole alkaloid pathway, where secologanin is a major intermediate. In order to establish the pathway from IPP and DMAPP to secologanin, the genes involved in the pathway were isolated and characterized. Of this, one of the major enzymes is 10-hydroxygeraniol dehydrogenase (Cr10HGO), belonging to the alcohol dehydrogenase family and catalyzing the reaction from an alcohol to di-aldehyde.

Monoterpene alcohols and aldehydes have several therapeutic values. The monoterpene alcohols themselves possess chemopreventive activities against various forms of cancer. For example, perillyl alcohol, formed by the cyclization of limonene, has been proved to cause apoptosis of liver tumors in rats⁶. Also, the acyclic monoterpene alcohol, geraniol, possesses anti-tumor activity against rat and mice hepatoma, leukemia and melanoma cells^{7,8}. Monoterpene aldehydes are also known to possess various therapeutic qualities, like citral (a racemic mixture of geranial and neral). Citral has been proved to be an antimicrobial, anti-inflammatory, antiparasitic, allelopathic and mosquito repellent⁹. These monoterpene alcohols and aldehydes are produced in various plants. In most plant species, the leaf epidermis plays a very crucial role in the biosynthesis of a wide range of these secondary metabolites like flavonoids, terpenes and alkaloids, and their accumulation. Similarly, in *C. roseus*, the leaf epidermal cells are specialized in the biosynthesis of monoterpene indole alkaloids¹⁰.

All Terpene indole alkaloids are derivatives of the tryptamine moiety, tryptophan and the iridoid terpene, Secologanin¹¹, which are condensed into strictosidine, the precursor for the indole alkaloids. The initial committed step for the synthesis of secologanin is the hydroxylation of geraniol to produce 10-hydroxygeraniol^{12,13}. The diol thus formed, is subsequently oxidized to the dialdehyde

cognate, 10-oxogeranial, which is further cyclized to form the basic iridoid backbone¹⁴.

In order to assess the substrate specificity and catalytic domain of Cr10HGO, homology modeling and site-directed mutagenesis studies were adopted. In this chapter, a homology-based model of Cr10HGO has been created and used for identifying the residues, which may affect the catalytic activity of the enzyme.

4.2 Materials and Methods

4.2.1 Computational modeling and docking

10-hydroxygeraniol dehydrogenase (Cr10HGO) homology model was built using the co-ordinates of *Populus tremuloides* Sinapyl Alcohol Dehydrogenase¹⁵ as the template with Schrodinger-Maestro 1.6. The lowest energy structures were validated using Ramachandran plot. 3D structure for reaction intermediates were constructed using ChemBioDrawUltra 13.0 for docking studies. To determine the amino acid interactions involved in the product formation, docking was performed using the substrate, 10-hydroxygeraniol and images were generated by PYMOL (v.1.6).

4.2.2 Site-Directed mutagenesis

Table 4.2.1 Primers for Cr10HGO site-directed mutagenesis. The mutagenesis bases are shown in red. Double mutants were generated using plasmid of Cr10HGO_M3 as template and primers of Cr10HGO_M10 and Cr10HGO_M12, respectively.

Cr10HGO_M1	His55Asn_F His55Asn_R	CACTCGGATCTT AAC ATGATCAAGAAC GTTCTTGATCAT GTT AAGATCCGAGTG
Cr10HGO_M2	His55Phe_F His55Phe_R	CACTCGGATCTT TTC ATGATCAAGAAC GTTCTTGATCAT GAA AAGATCCGAGTG
Cr10HGO_M3	Val277Thr_F Val277Thr_R	AAGCTAATTCTT ACT GGTGCACCAGAA TTCTGGTGCACC AGT AAGAATTAGCTT
Cr10HGO_M4	Ala279Ser_F Ala279Ser_R	ATTCTTGTTGGT TCA CCAGAAAAGCCA TGGCTTTTCTGG TGA ACCAACAAGAAT
Cr10HGO_M5	Asn115His_F Asn115His_R	AAAGATCTTGAA CAT TATTGTCCGGGC GCCCCGACAATA ATG TTCAAGATCTTT
Cr10HGO_M6	Asn115Leu_F Asn115Leu_R	AAAGATCTTGAA CTT TATTGTCCGGGC GCCCCGACAATA AAG TTCAAGATCTTT
Cr10HGO_M7	His51Phe_Fwd His51Phe_Rev	TGCGGGATCTGT TTCT CGGATCTTCAC GTGAAGATCCGA GAA ACAGATCCCGCA
Cr10HGO_M8	Ser256Tyr_Fwd Ser256Tyr_Rev	ATTGACACCGTT TAT GCAATTCATCCA TGGATGAATTGC ATA AACGGTGTCAAT
Cr10HGO_M9	Thr170Ala_Fwd Thr170Ala_Rev	TGTGCTGGGATT GCT ACATACAGTCCA TGGACTGTATGT AGC AATCCCAGCACA

Cr10HGO_M10	Trp61Leu_Fwd	ATCAAGAACGAA <u>T</u> TGGGCTTCACGAAA
	Trp61Leu_Rev	TTTCGTGAAGCC <u>CA</u> AATTCGTTCTTGAT
Cr10HGO_M11	Ser301Ala_Fwd	ATAATAGCTGGAG <u>GCT</u> GCAATTGGAGGG
	Ser301Ala_Rev	CCCTCCAATTGC <u>AGCT</u> TCCAGCTATTAT
Cr10HGO_M12	Ala302Phe_Fwd	ATAGCTGGAAGT <u>TTC</u> ATTGGAGGGTTG
	Ala302Phe_Rev	CAACCCTCCAAT <u>GAA</u> ACTTCCAGCTAT
Cr10HGO_M13	Cys98Ala_Fwd	GTAGGAGTGGGT <u>GCC</u> CTTGTGGGATCA
	Cys98Ala_Rev	TGATCCCACAAG <u>GGC</u> ACCCACTCCTAC
Cr10HGO_M14	Tyr116Phe_Fwd	GATCTTGAAAAT <u>TTT</u> TGTCCGGGCCAA
	Tyr116Phe_Rev	TTGGCCCGGACA <u>AAA</u> ATTTTCAAGATC

All the mutation reactions were carried out with a pRSET B expression vector (Invitrogen) harboring the coding sequence for 10-hydroxygeraniol dehydrogenase as a template. The mutants were generated using the QuikChange Lightning site directed mutagenesis kit from Stratagene according to the manufacturer's instructions. The mutagenesis reaction mixtures were transformed into *Escherichia coli* XL10 Gold Cells and plated onto LB- agar plates containing ampicillin. Individual colonies were picked and grown in selection media. Plasmid DNA was then isolated and the mutant clones were identified by sequencing. The mutagenic primers (with mutagenic bases shown in boldface and underlined), which were designed on the basis of homology-based model of *Cr10HGO* are listed in the table 4.2.1.

4.2.3 Heterologous Expression and Purification of Cr10HGO mutants

pRSET B vector harboring Cr10HGO mutants were expressed and purified as done earlier¹⁶. Protein expression was carried out in BL21 DE3 cells in Luria Bertani Broth at 37 °C and induced with IPTG at a final concentration of 1 mM and incubated for 6 hours at 30 °C. After the induction, the culture was harvested by centrifugation at 4500 × g for 20 minutes. The cell pellet was re-suspended in 10 mL/g of Lysis buffer (50 mM NaH₂PO₄, 300 mM NaCl, 0.5 % CHAPS, pH 8.0, 10 % v/v Glycerol) with 1 mg/mL concentration of lysozyme and 100 µL/5 mL of protease inhibitor cocktail from Sigma and incubated on ice for 30 minutes. After the incubation period, the cells were sonicated with a pulse of 30 sec ON and 30 sec OFF for 10 cycles. The sample was then centrifuged at 5,000 × g for 10 minutes at 4 °C. The crude lysate was purified using Ni-NTA column (2 mL resin / g cell pellet lysate). The protein was

finally eluted out in the elution buffer (50 mM NaH₂PO₄, 300 mM NaCl, pH 8.0, 250 mM Imidazole, 10 % v/v Glycerol) each. Purified protein fractions (after checking on 12 % SDS gel) were pooled together and desalted on Hi-PrepTM 26/10 Desalting Columns with desalting buffer (20 mM MOPS, pH 8.0, 10 % v/v Glycerol) using AKTA (GE Healthcare). The desalted proteins were estimated using Bradford reagent (Bio-Rad) and a Bovine Serum Albumin standard.

4.2.4 Enzymatic Characterization and Product Analysis

Cr10HGO was assayed for its activity using 100 µg desalted protein and 1 mM 10-hydroxygeraniol as the substrate in 500 µL Sodium bi-carbonate Buffer (20 mM Sodium bi-carbonate, 10 % v/v Glycerol, pH 10.0) at 30 °C for 30 minutes. Assay mixtures were extracted with dichloromethane (DCM). The combined organic phase was dried over sodium sulphate, reduced to ~50 µL with a stream of dry nitrogen and subjected to GC and GC-MS analyses for product characterization by comparing with authentic standards. 1 µL of the extract was injected onto a 30 m × 0.25 mm × 0.25 µm HP-5 capillary GC column with a temperature gradient from 60 to 120 °C at 20 °C per min, followed by a temperature gradient from 120 to 170 °C at 2.5 °C per min and a final temperature gradient from 170 to 190 °C at 20 °C per min.

4.2.5 Determination of Kinetic Parameters of Cr10HGO mutants

Steady-state kinetic constants were determined for all the mutants in 20 mM Sodium bi-carbonate, 10 % v/v Glycerol, pH 10.0 at 30 °C with varying substrate concentrations, ranging from 0.25 to 500.0 mM with saturation concentration of cofactor, NADP⁺ (500.0 mM) and vice versa. The reactions were followed by measuring changes in NADPH concentration at 340 nm. The kinetic data were fitted with the Graph Pad Prism software and the parameters calculated using Michaelis-Menten plots.

4.2.6 Substrate Specificity Studies

Various alcohols namely, geraniol, farnesol, nerol, α -bisabolol, α -santalol, β -citronellol, chrysanthemol, eugenol, linalool and menthol were assayed with the prominent Cr10HGO mutants and checked for their activity on different substrates. The enzyme assay, characterization and product analysis were the same as in Section 4.2.4.

4.3 Results

4.3.1 Homology Modeling and Docking

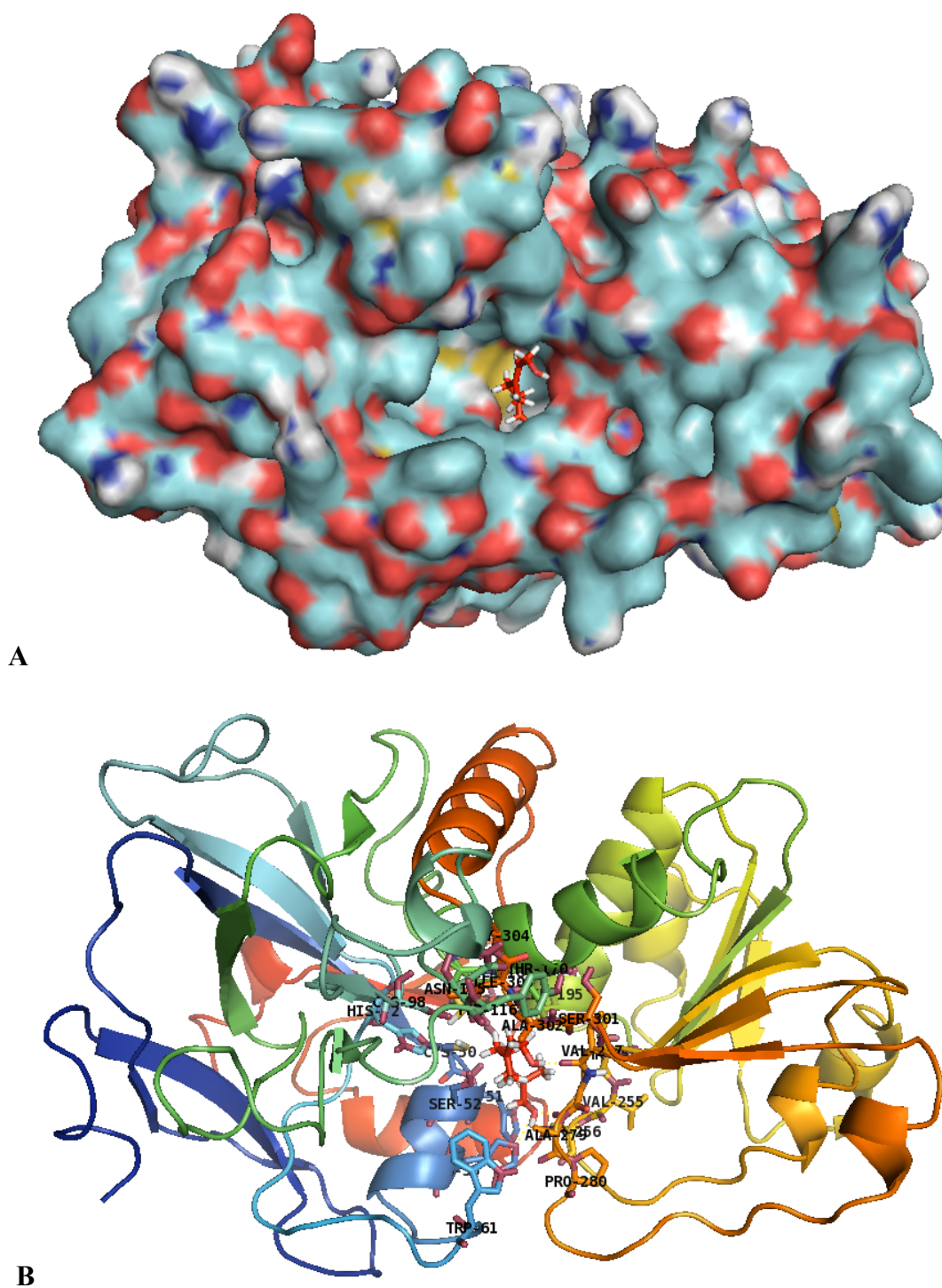


Figure 4.3.1 A. Homology model of 10-hydroxygeraniol dehydrogenase built using coordinates of sinapyl alcohol dehydrogenase from *Populus tremuloides* as a template. B. 10-hydroxygeraniol docked with Cr10HGO model.

Alcohol dehydrogenases are a group of enzymes belonging to the oxidoreductase family, found in various living organisms. They are known to catalyze the reversible oxidation of primary or secondary alcohols to aldehydes or ketones, respectively. The Cr10HGO amino acid sequence displays 78 % sequence identity with Sinapyl Alcohol Dehydrogenase from *Populus tremuloides* (PDB ID: 1YQD¹⁵) and 52 % with Cinnamyl Alcohol Dehydrogenases from *Arabidopsis thaliana* (PDB ID: 2CF5¹⁷).

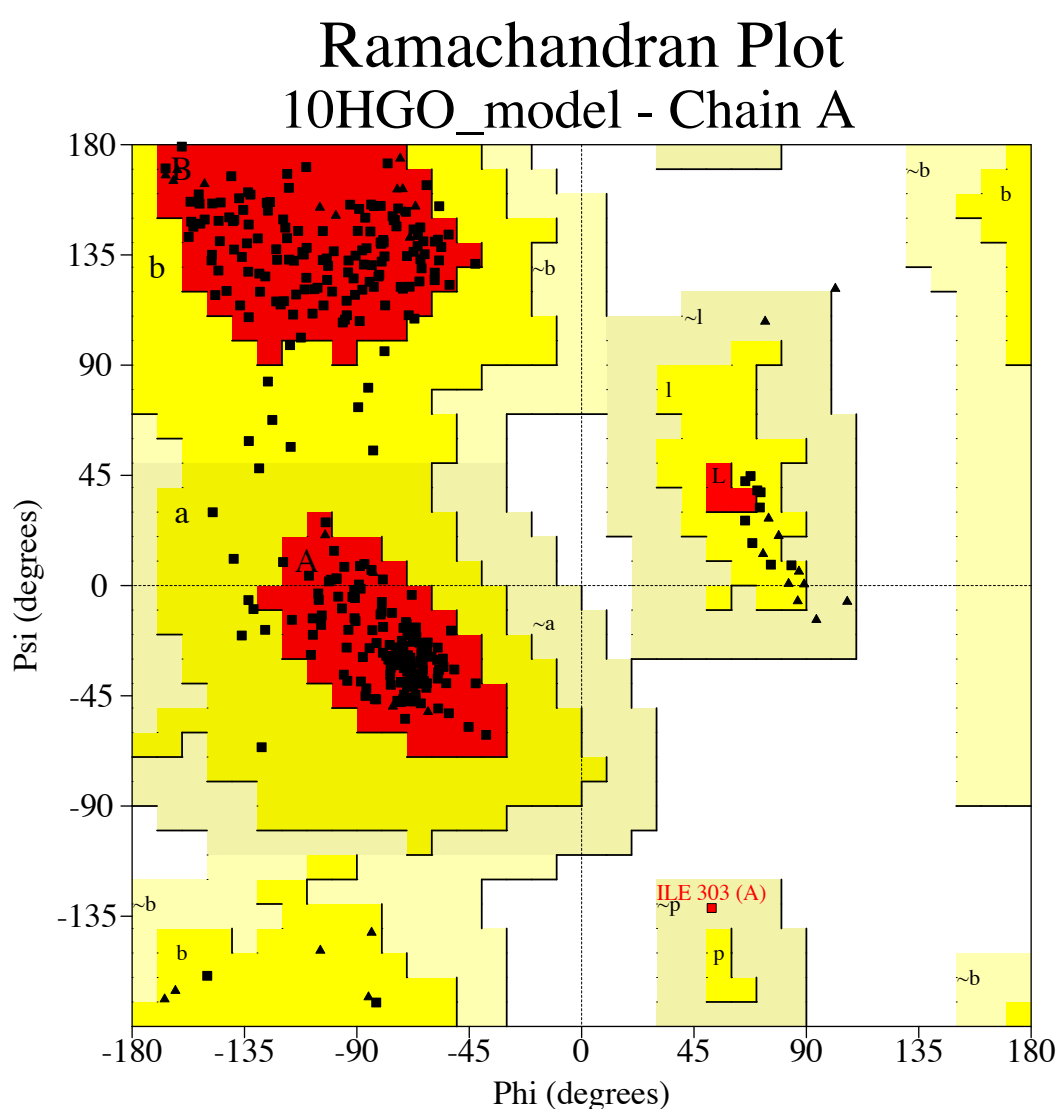


Figure 4.3.2: Ramachandran Plot of the built model of Cr10HGO, showing maximum residues (90.1 %) in the most favored region.

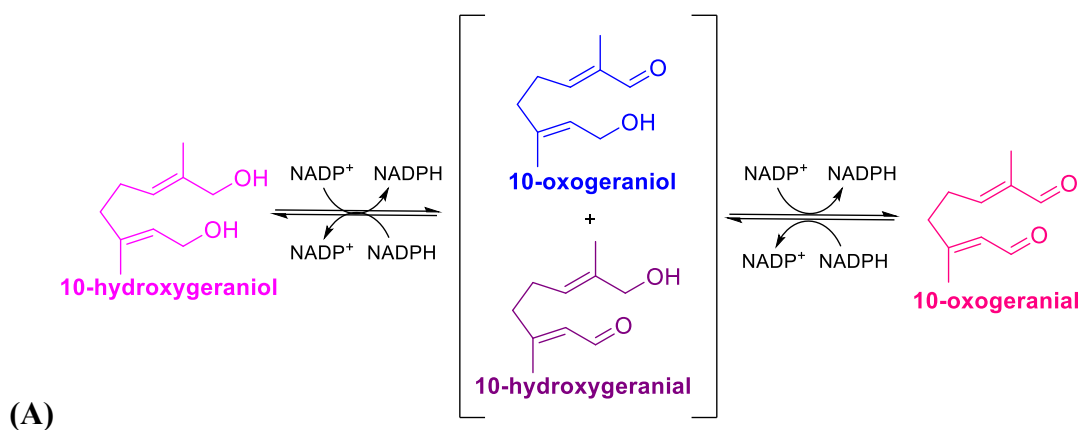
Deduced amino acid sequence of Cr10HGO was found to contain zinc binding alcohol dehydrogenase signature sequence (positions 71–85) with the PROSITE ID: ADH_ZINC (PS00059), and the consensus sequence G-H-E-x-EL-G-7-x(4)- [GA]-

x(2)-[IVSAC]. In order to understand the basis for the product profile of Cr10HGO, homology based model was built using the co-ordinates of sinapyl alcohol dehydrogenase from *Populus tremuloides* (PDB ID: 1YQX¹⁵) as a template, based on an overall sequence similarity of 77 % at amino acid level, using Schrodinger (Figure 4.3.1 A & B).

The homology model of Cr10HGO built using Schrodinger was validated using Ramachandran plot¹⁸ (Figure 4.3.2) with default parameters. Ramachandran PROCHEK validation tools suggested that 272 residues (90.1 %) of the modeled Cr10HGO were present in the most favored region, whereas 29 residues (9.6 %) were found to be present in the additional allowed regions. There were no residues (0.0 %) present in the disallowed region. This result clearly suggests that the quality of model built is appropriate and can be used for further docking studies.

4.3.2 Catalytic mechanism of Cr10HGO on 10-hydroxygeraniol

Cr10HGO acts on the alcohol, 10-hydroxygeraniol, converting the alcohols at the positions 1 and 10 into aldehydes, forming 10-hydroxygeranial and 10-oxogeraniol, respectively. Through these intermediates, the di-aldehyde, 10-oxogeraniol is formed. Cr10HGO is an NADP⁺ and NADPH dependent enzyme¹⁹, and is active only in their presence. As it is a reversible enzyme, it utilizes NADP⁺ in the forward reaction and NADPH in the reverse reaction. In the usual reaction of wild type Cr10HGO with 10-hydroxygeraniol and NADP⁺, 83.2 % of the substrate was found to be converted into two intermediates and 1 product, namely 10-oxogeraniol (49.5 %), 10-hydroxygeranial (12.8 %) and 10-oxogeraniol (20.9 %).



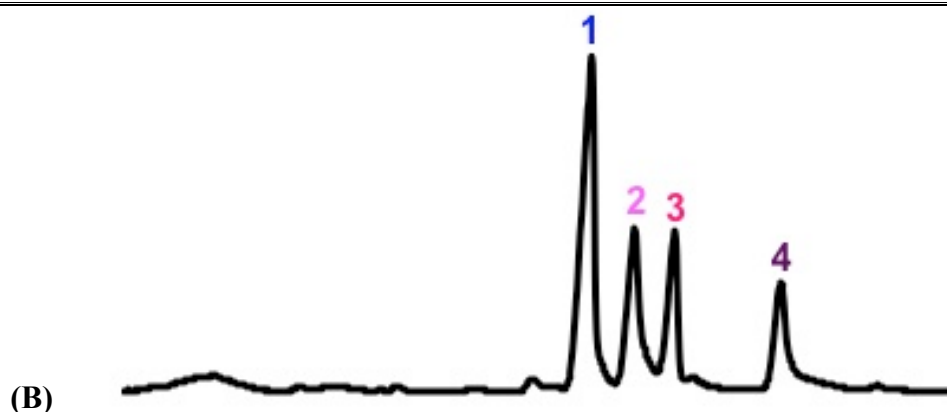


Figure 4.3.3 (A) Cr10HGO mediated reaction in *C. roseus* for biosynthesis of iridoids (B) Product profile of Cr10HGO enzyme assay, (1): 10-oxogeraniol; (2): 10-hydroxygeraniol (substrate); (3): 10-oxogeranial; (4): 10-hydroxygeranial.

4.3.3 Site-directed mutagenesis and characterization

In order to understand **the exact role of various residues present in close vicinity of substrate**, 12 amino acid residues were identified based on their interactions with the docked substrate, which may affect the product profile of Cr10HGO significantly, and 14 mutations were designed. These mutants were expressed in *E. coli* and the purified proteins were assayed for their activity with 10-hydroxygeraniol as substrate, in the presence of NADP⁺ as the co-factor. The products of the reaction were analyzed on a HP-5 capillary GC column. In addition, two double mutants were also designed based on the preliminary results.

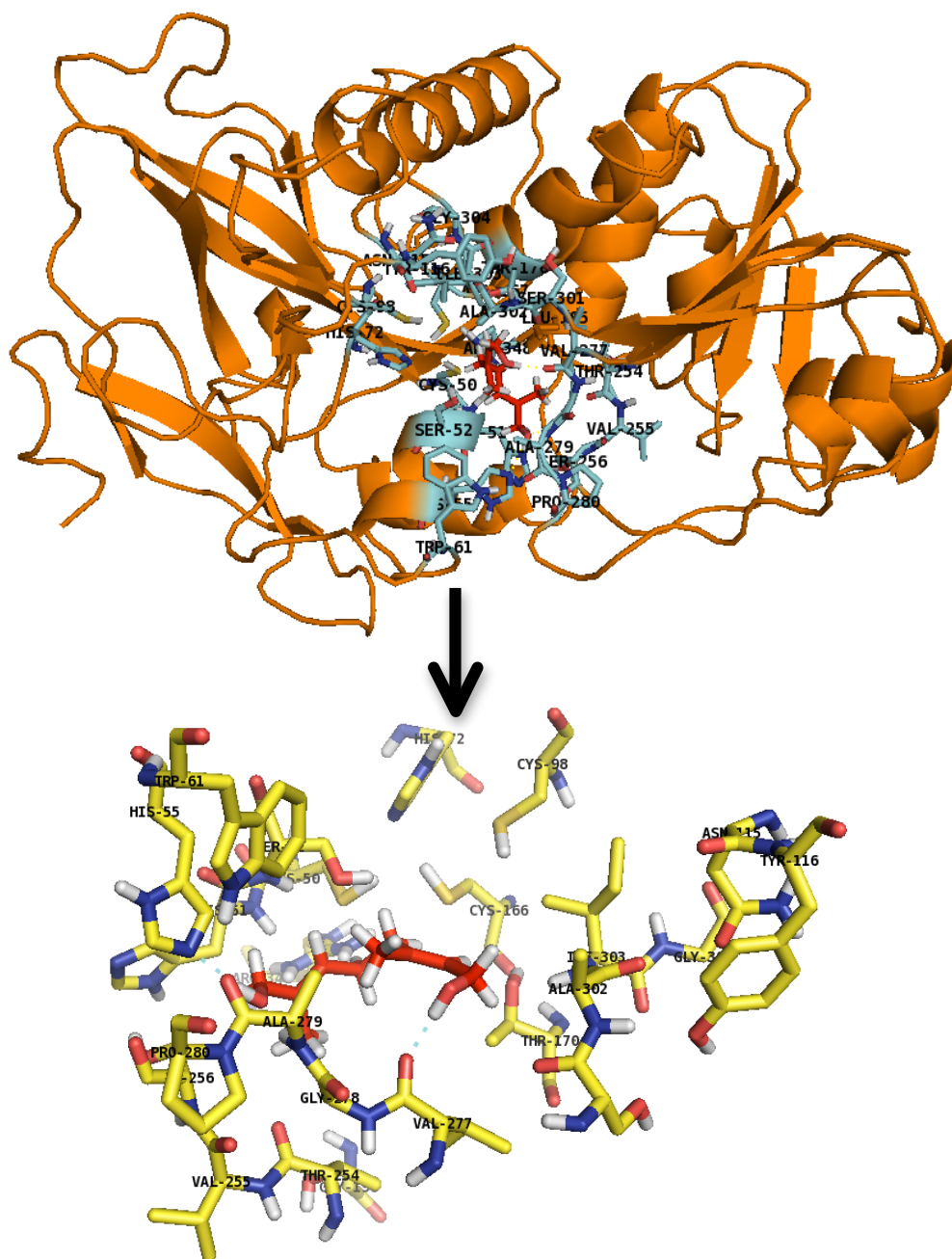


Figure: 4.3.4 Homology model of Cr10HGO showing amino acid interactions with the substrate.

4.3.4 Heterologous Expression and Purification of Cr10HGO mutants

Expression was carried out in Rosetta2 DE3 cells. The recombinant protein of Cr10HGO in pRSETB was expressed and purified to the homogeneity as discussed earlier. The purified protein fractions showed distinct bands of 40 kDa on a 12 % SDS-PAGE (Figure 4.3.5). The purified fractions were desalted, flash-frozen and store at -80 °C till used.

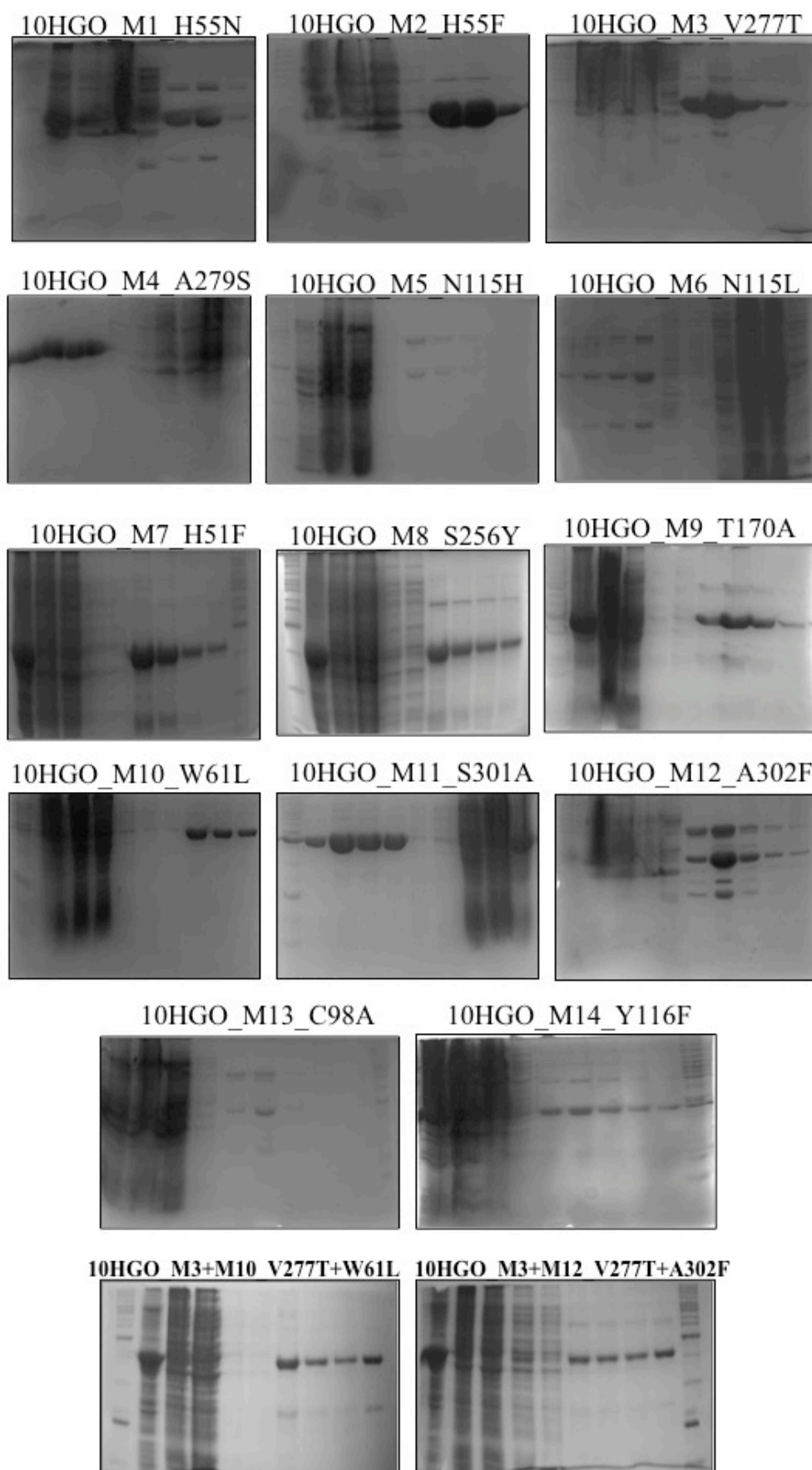


Figure 4.3.5 SDS gel images of Cr10HGO mutants' protein purification (M. Wt: 40 KDa)

4.3.5 Residues not significantly affecting the activity

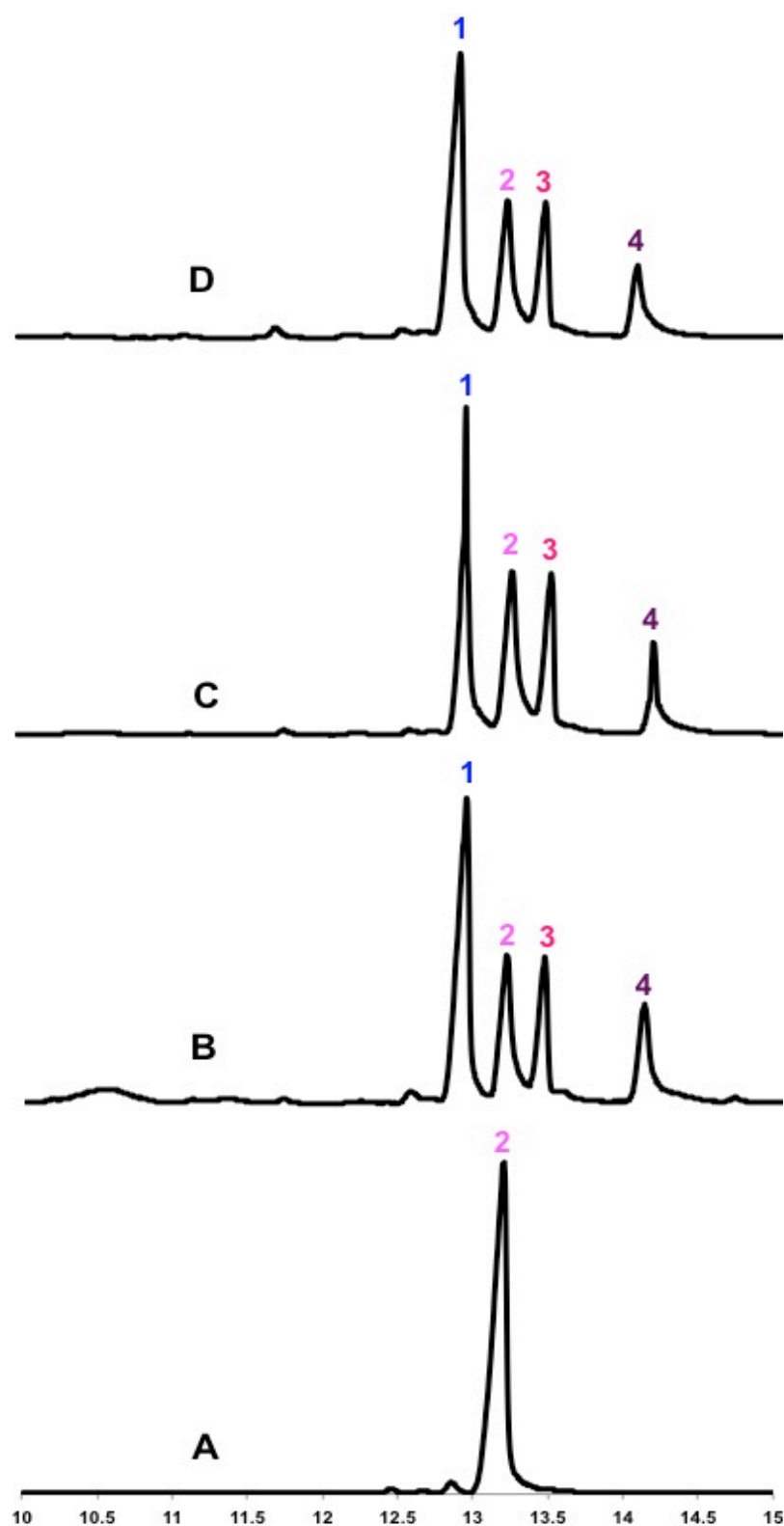


Figure 4.3.6 Total Ion Chromatograms (TICs) of A. Substrate Control, B. Cr10HGO Wild Type Assay, C. Cr10HGO_H55N Assay, D. Cr10HGO_N115L Assay; 10-oxogeraniol (1), 10-hydroxygeraniol (2), 10-oxogeraniol (3), 10-hydroxygeraniol (4).

On replacing asparagine 115 (N115) with the positively charged leucine (N115L), there was not much change observed in the activity of the protein or in the product profile. The substitution of leucine for asparagine does not affect the ability of the protein to bind to the substrate and produced similar product profile as the wild type protein. Whereas, when this asparagine (N115) was replaced with histidine (N115H), the activity of the protein was drastically reduced. When Histidine was replaced with asparagine at H55 position, we would expect a similar result. But, surprisingly, this mutation did not cause any change in the activity of the protein which seemed to be interacting well with the substrate, as can be observed from the model. However, when this histidine was replaced with the bulkier aromatic residue, phenylalanine, the intermediate 10-oxogeraniol reduced by half, while the other intermediate, 10-hydroxygeraniol was increased by 18.6 %, reducing the formation of 10-oxogeraniol by 5.3 %. In another case, valine 277, present near the 1-hydroxy position of the substrate, was mutated to threonine, where introduction of a hydroxyl group increased the polarity near the 1-hydroxy position. This change in polarity caused a slight decrease in the formation of the final di-aldehyde product, i.e., caused a stunted forward reaction. Although, the substrate conversion was increased by almost 10 %, this was compensated in the intermediate, 10-oxogeraniol and not the final di-aldehyde product.

4.3.6 Residues causing significant changes in the product profile of Cr10HGO

Among the 14 mutants, two mutants were identified as being able to reduce the substrate conversion. The mutants S301A and N115H reduced the substrate conversion from 83.2 % in the wild type to 25.7 % and 24.4 %, respectively.

4.3.6.1 Serine 301 (S301A)

The residue S301, which is found to be conserved among alcohol dehydrogenases across species, plays an important role in the activity profile of Cr10HGO. The corresponding serine-301 residue in the sinapyl alcohol dehydrogenase was shown to be involved in H- bond formation with the substrate near the 1-hydroxy position of the substrate, 10-hydroxygeraniol. When Serine was replaced with alanine (S301A), causing a dehydroxylation, only 25.7 % of the substrate was converted into product and only one product, 10-hydroxygeraniol at

that. The other intermediate, 10-oxogeraniol and the final di-aldehyde product, 10-oxogeranial were not formed at all. The change in polarity caused due to the replacement of serine with alanine may have affected the H- bond interaction. It may also be affecting the stabilization of the intermediate for further activity. By these observations, we may predict that the first reaction might be the oxidation of the –OH at 1C position.

4.3.6.2 Asparagine 115 (N115H)

When the residue N115, which is also conserved across species, was mutated, varied results were observed. This residue is present in the active site pocket and on replacing asparagine with the positively charged leucine (N115L), there was not much change observed in the activity of the protein or in the product profile, as described earlier. On the other hand, when asparagine was replaced with another positively charged amino acid, histidine (N115H), there was a drastic change in the product profile of Cr10HGO. The mutant N115H caused 75.6 % retention of the substrate. The intermediates and the final product formed comprised the remaining 24.4 % of the total, with the di-aldehyde being formed only to the effect of 2.6 %. In the case where N115 was mutated to leucine, there is only a replacement of one linear chain with another, whereas, in the second case, the linear chain in asparagine is replaced by the heterocyclic chain in histidine, which seems to be affecting the oxidation of the substrate, and hence the conversion of alcohol to aldehydes. It can also be predicted that an increase in the polarity of the active site pocket due to the change from asparagine to histidine is causing the product formation to decrease, i.e, the reaction to decrease.

In another case, when a change was incorporated in A279 position in the form of serine, there was once again a reduction in the enzyme activity. In this case, 33.7 % of the substrate remained unchanged as opposed to 16.8 % in the case of wild type. Here again, this decrease in activity could be attributed to the increase in polarity. Interestingly, all three amino acids at the different positions (H55, N115 and S301) were conserved and corresponded to the same amino acid in the model sequence, as well.

The product profiles of these three mutations leads us to an observation that serine plays a very essential role in governing the activity of the enzyme.

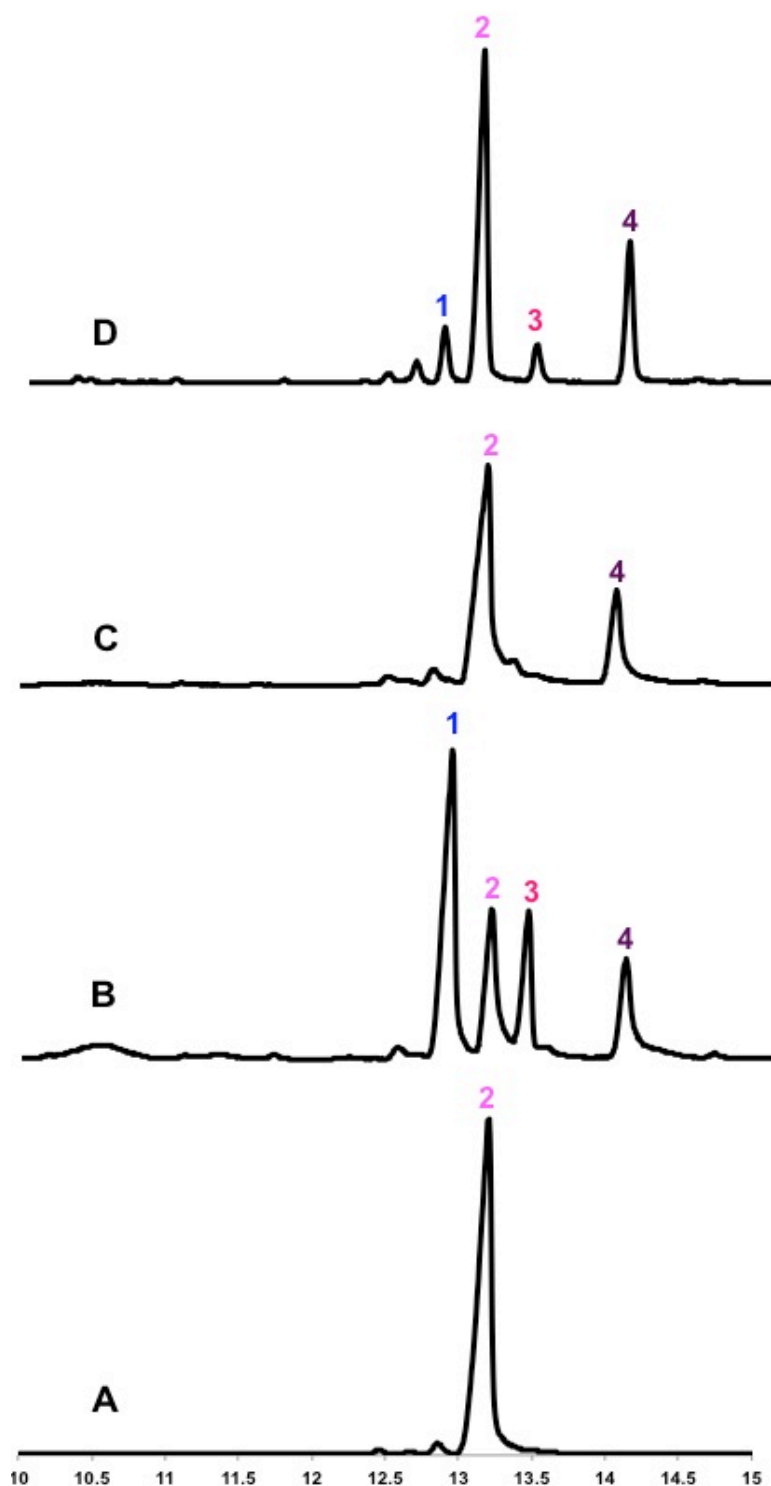


Figure 4.3.7 Total Ion Chromatograms (TICs) of A. Substrate Control, B. Cr10HGO Wild Type Assay, C. Cr10HGO_S301A Assay, D. Cr10HGO_N115H Assay; 10-oxogeraniol (1), 10-hydroxygeraniol (2), 10-oxogeraniol (3), 10-hydroxygeraniol (4).

4.3.7 Residues involved in the increased production of the di-aldehyde

The di-aldehyde, 10-oxogeranial, which is the final product of the reaction carried out by Cr10HGO, is the substrate for the next enzyme in the iridoid biosynthetic pathway. It is very essential that this product should form in a significant quantity so as to facilitate the increased production of iridoids. Some of the residues mutated in our study resulted in increase in the formation of this product and hence were identified as essential to the retention of the di-aldehyde.

4.3.7.1 Tyrosine 116 (Y116F)

Tyrosine at 116th position is a conserved one among alcohol dehydrogenases of related species and hence corresponds to the same in the model enzyme too. The mutant Y116F increases the formation of 10-oxogeranial only marginally, however reduces the substrate retention by 1.8 times, wherein only 9.3 % of the substrate remained unchanged as opposed to 16.8 % in the case of wild type. In the case of the intermediates, both showed only a marginal increase. As was the case before, an increase in hydrophobicity supports the forward reaction to a higher extent.

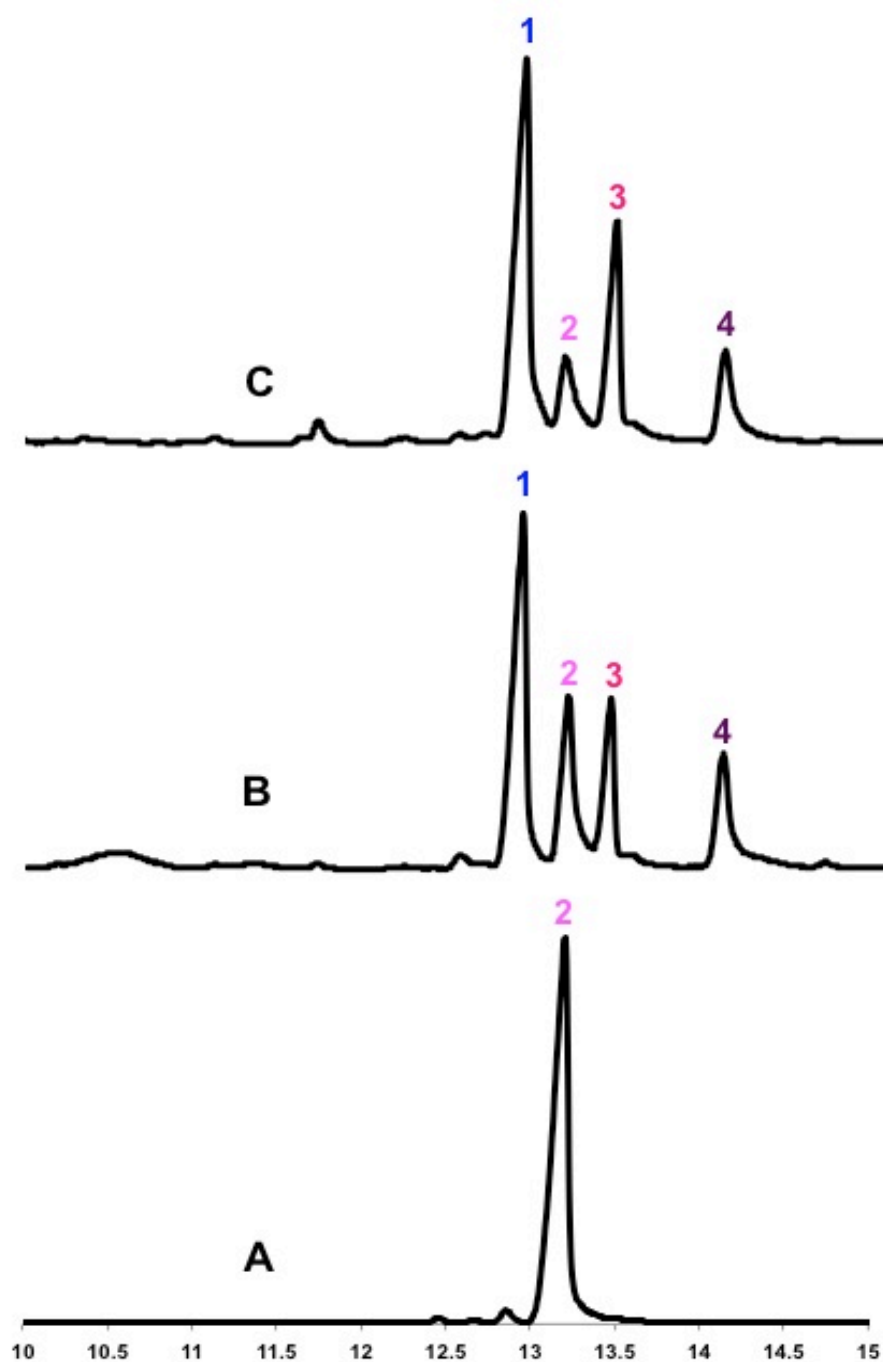


Figure 4.3.8 Total Ion Chromatograms (TICs) of **A.** Substrate Control, **B.** Cr10HGO Wild Type Assay, **C.** Cr10HGO_Y116F Assay; 10-oxogeraniol (1), 10-hydroxygeraniol (2), 10-oxogeraniol (3), 10-hydroxygeraniol (4).

4.3.7.2 Serine 256 and Alanine 302 (S256Y and A302F)

The mutation to S256Y and A302F resulted in the increase level of the 10-oxogeraniol production by 3 – 4 %, the former only marginally decreased the

retention of the substrate while the latter was successfully able to convert 93.1 % of the substrate to products.

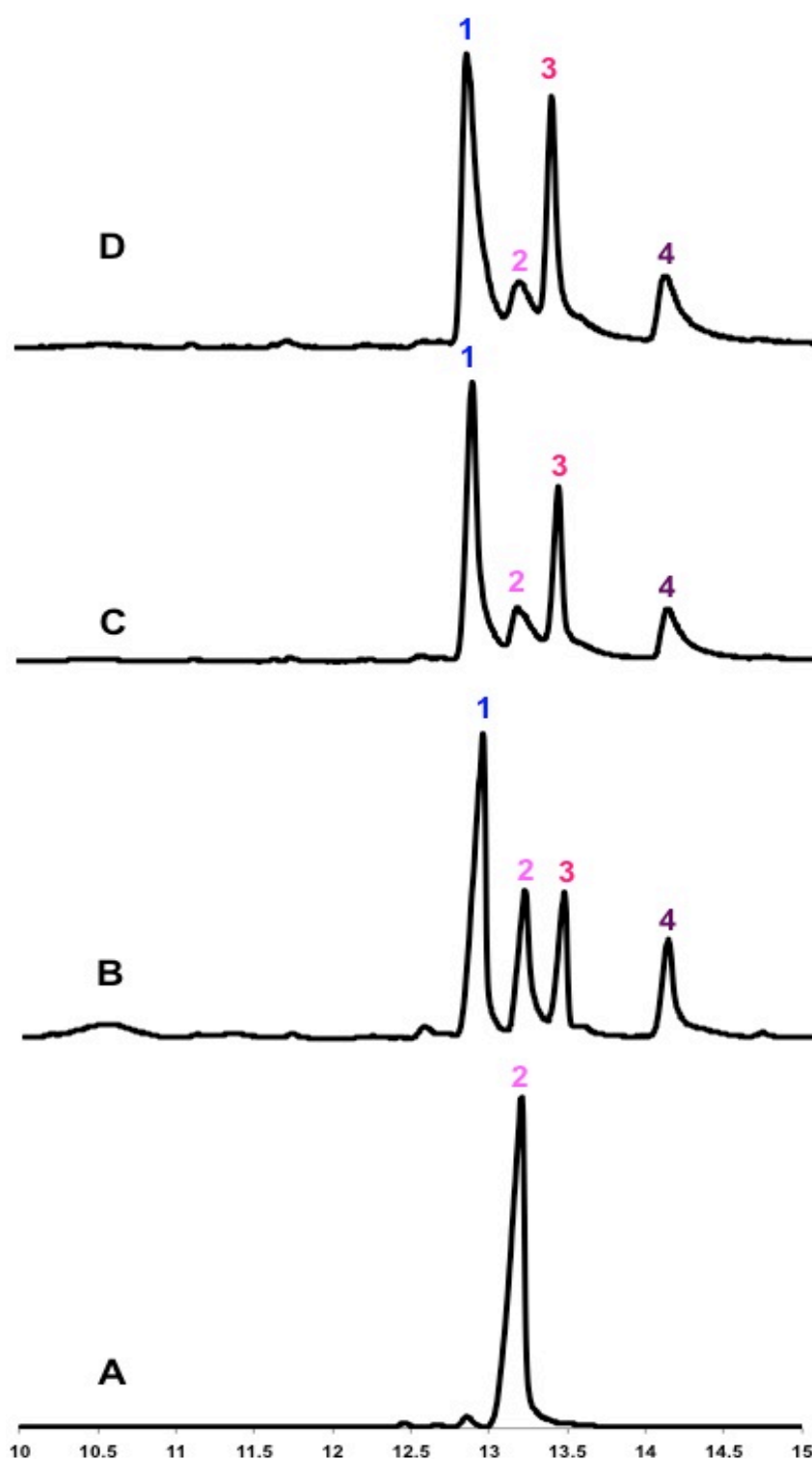


Figure 4.3.9 Total Ion Chromatograms (TICs) of **A.** Substrate Control, **B.** Cr10HGO Wild Type Assay, **C.** Cr10HGO_S256Y Assay, **D.** Cr10HGO_A302F Assay; 10-oxogeraniol (**1**), 10-hydroxygeraniol (**2**), 10-oxogeraniol (**3**), 10-hydroxygeraniol (**4**).

As before, in this case also these two residues, S256 and A302 were highly conserved among similar species. However, where S256 in *C. roseus* corresponded to S256 in Sinapyl alcohol dehydrogenase, A302 in *C. roseus* corresponded with G302 in Sinapyl alcohol dehydrogenase. Here both the residues in Cr10HGO were replaced by amino acids with aromatic side chains, and both caused an almost similar effect. Once again, we can observe here that both S256Y and A302F mark an increase in hydrophobicity, which subsequently causes an increase in the forward reaction.

4.3.7.3 Cysteine 98 (C98A)

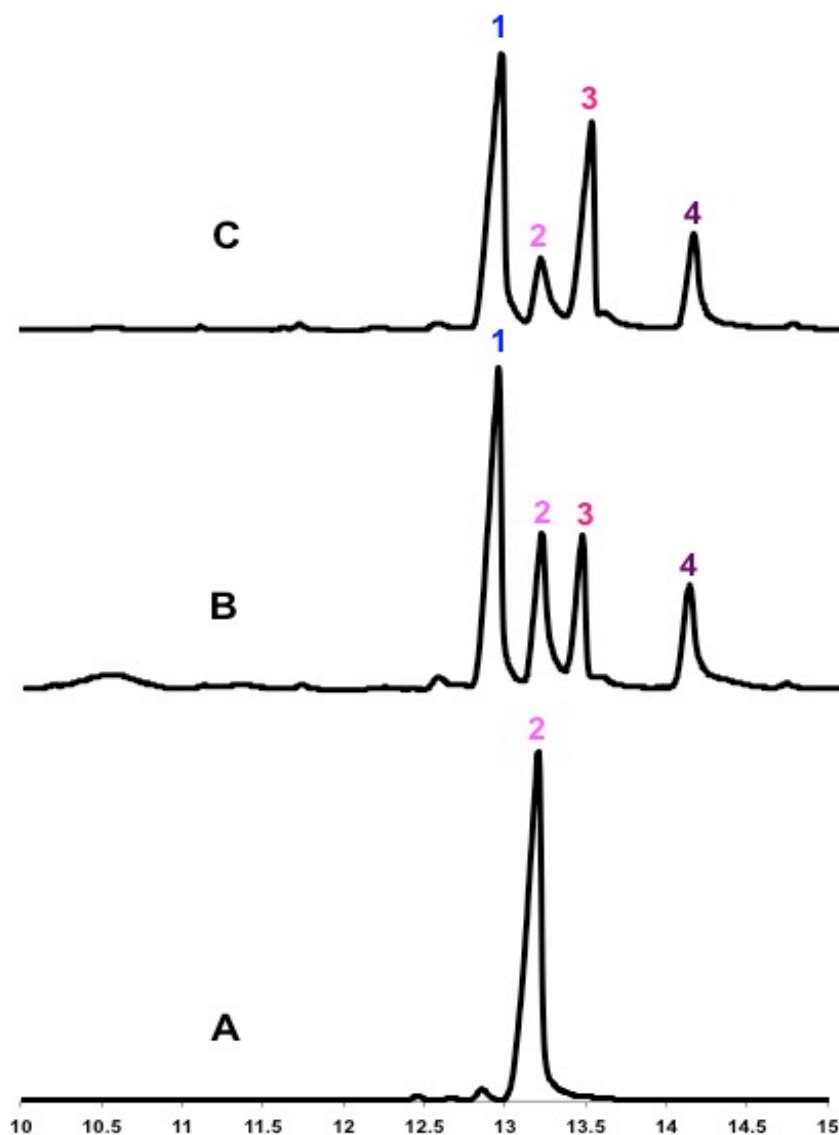


Figure 4.3.10 Total Ion Chromatograms (TICs) of A. Substrate Control, B. Cr10HGO Wild Type Assay, C. Cr10HGO_C98A Assay; 10-oxogeraniol (1), 10-hydroxygeraniol (2), 10-oxogeraniol (3), 10-hydroxygeraniol (4).

The cysteine at the 98th position is also a highly conserved residue, and upon replacement with the much smaller alanine, resulted in a 29-fold increase in the formation of 10-oxogeranial making its contribution in the final product pool to 26.9 % from 20.9 %. This mutation also played a very vital role in the substrate conversion factor, where it resulted in 73-fold decrease in the retention of the substrate. Hence, where normally in the case of wild type, 16.8 % of the substrate is retained, in this mutant, only 9.7 % of the substrate is retained. An increase in the hydrophobicity can be attributed the reason for the increase in the forward reaction.

4.3.7.4 Histidine 51 and Tryptophan 61 (H51F and W61L)

The Histidine at the 51st position and Tryptophan at the 61st position are quite conserved across various species. While H51 is found to be involved in H-bond formation with the substrate at the C10 position, W61 and A302 envelop the binding site pocket. Mutations at H51 to phenylalanine, replacing the heterocyclic ring of histidine with the aromatic ring in phenylalanine and W61 to leucine, removing the aromatic ring of tryptophan, caused almost an 8 % increase in the formation of the di-aldehyde. Whilst these two mutants had the exact same effect on the di-aldehyde formation, their influence on substrate retention was quite divergent. The former caused only a marginal decrease in substrate retention, whereas, the latter decreased the substrate percentage by 3.7 times (4.6 %), the lowest among all mutants. In the case of the W61L mutant, 95.4 % of the substrate was converted to products. These results indicate that when a bulkier group like tryptophan is replaced by the lighter leucine, the interaction between the substrate and the protein increases substantially to push higher product formation.

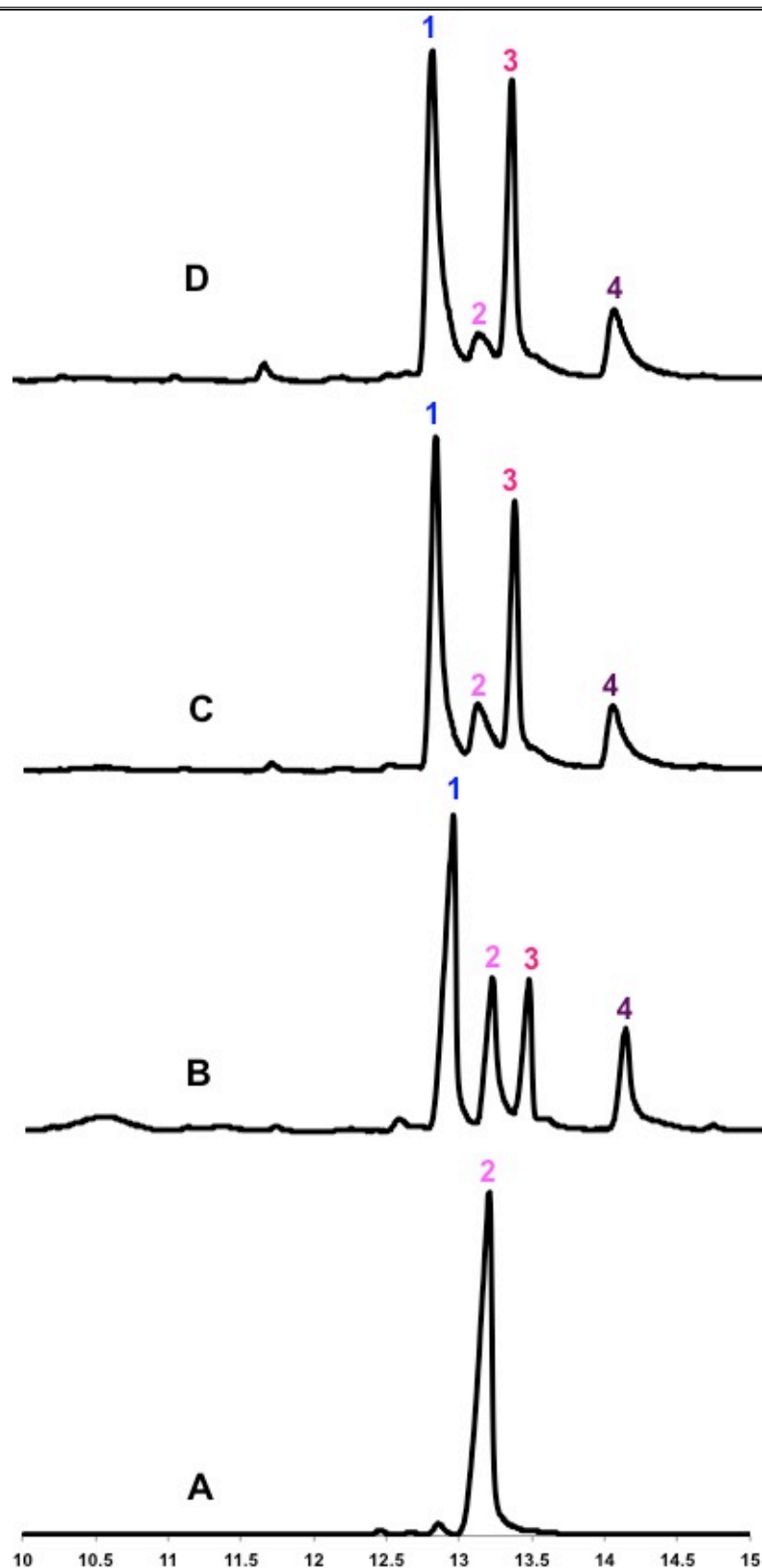


Figure 4.3.11 Total Ion Chromatograms (TICs) of A. Substrate Control, B. Cr10HGO Wild Type Assay, C. Cr10HGO_H51F Assay, D. Cr10HGO_W61L; 10-oxogeraniol (1), 10-hydroxygeraniol (2), 10-oxogeraniol (3), 10-hydroxygeraniol (4).

4.3.7.5 Threonine 170 (T170A)

In the case of the mutant T170A, where the replacement of threonine with alanine caused demethylation and dehydration of a 2° alcohol, the formation of 10-

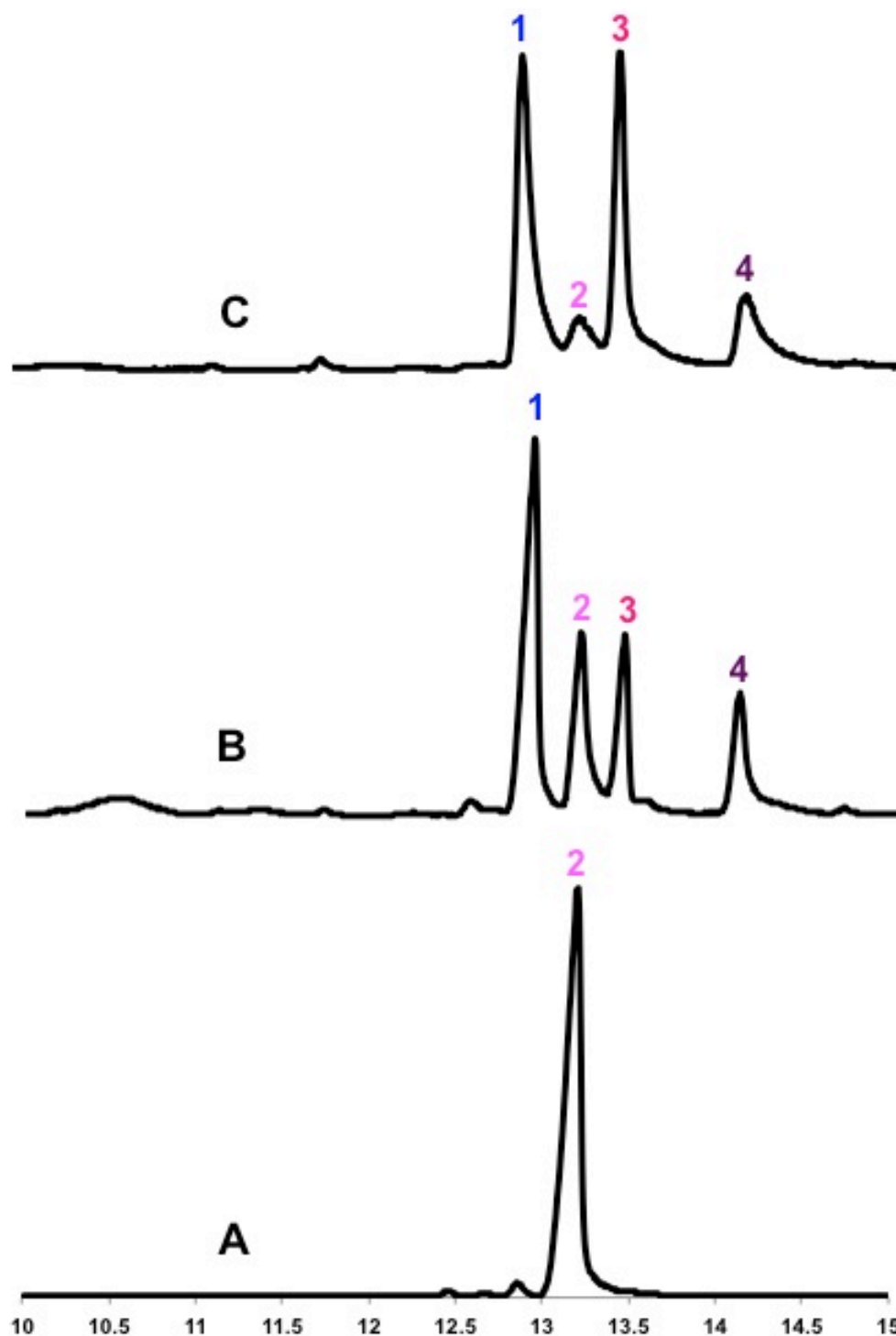


Figure 4.3.12 Total Ion Chromatograms (TICs) of **A.** Substrate Control, **B.** Cr10HGO Wild Type Assay, **C.** Cr10HGO_T170A Assay; 10-oxogeraniol (1), 10-hydroxygeraniol (2), 10-oxogeraniol (3), 10-hydroxygeraniol (4).

oxogeraniol increases from 20.9 % to 33.7 %. The threonine residue at this position belongs to the dinucleotide binding domain, interacting with the NADP⁺ molecule. The mutation causes a drastic change in polarity, which affects the interaction of the protein with the substrate and co-factor. At the same time, the amount of substrate contributing to the product profile dropped from 16.8 % to 9 %. Of all the mutations carried out in this study, T170A resulted in diverting the highest flux towards 10-oxogeraniol (33.7 %).

Table 4.3.1 Product profiles of assays of Cr10HGO wild type and mutants with 10-hydroxygeraniol.

	Peak 1 : 10-oxogeraniol Rt: 12.802	Peak 2 : 10-hydroxygeraniol Rt: 13.033	Peak 3 : 10-oxogeraniol Rt: 13.308	Peak 4 : 10-hydroxygeraniol Rt: 13.921
Cr10HGO WT	49.5 %	16.8 %	20.9 %	12.8 %
Cr10HGO M1 H55N	49.2 %	17.1 %	20.7 %	13.0 %
Cr10HGO M2 H55F	25.0 %	28.1 %	15.6 %	31.4 %
Cr10HGO M3 V277T	57.2 %	7.2 %	15.5 %	20.1 %
Cr10HGO M4 A279S	29.8 %	33.7 %	14.9 %	21.6 %
Cr10HGO M5 N115H	8.8 %	75.6 %	2.6 %	13.0 %
Cr10HGO M6 N115L	48.4 %	17.0 %	19.7 %	14.9 %
Cr10HGO M7 H51F	45.2 %	13.6 %	28.5 %	12.8 %
Cr10HGO M8 S256Y	47.8 %	14.7 %	24.1 %	13.4 %
Cr10HGO M9 T170A	43.6 %	9.0 %	33.7 %	13.7 %
Cr10HGO M10 W61L	52.3 %	4.6 %	28.5 %	14.6 %
Cr10HGO M11 S301A	0.0 %	74.3 %	0.0 %	25.7 %
Cr10HGO M12 A302F	51.5 %	6.9 %	24.5 %	17.1 %
Cr10HGO M13 C98A	48.4 %	9.7 %	26.9 %	15.0 %
Cr10HGO M14 Y116F	54.8 %	9.3 %	22.0 %	13.9 %
Cr10HGO M3+M12	24.2 %	66.7 %	1.8 %	7.4 %
Cr10HGO M3+M10	31.7 %	51.2 %	2.5 %	10.7 %

4.3.8 Double Mutants

Based on the low substrate retention, two double mutants were designed by combining **a)** V277T and W61L and **b)** V277T and A302F, to check if these double mutations would help in completely eliminating the substrate from the profile.

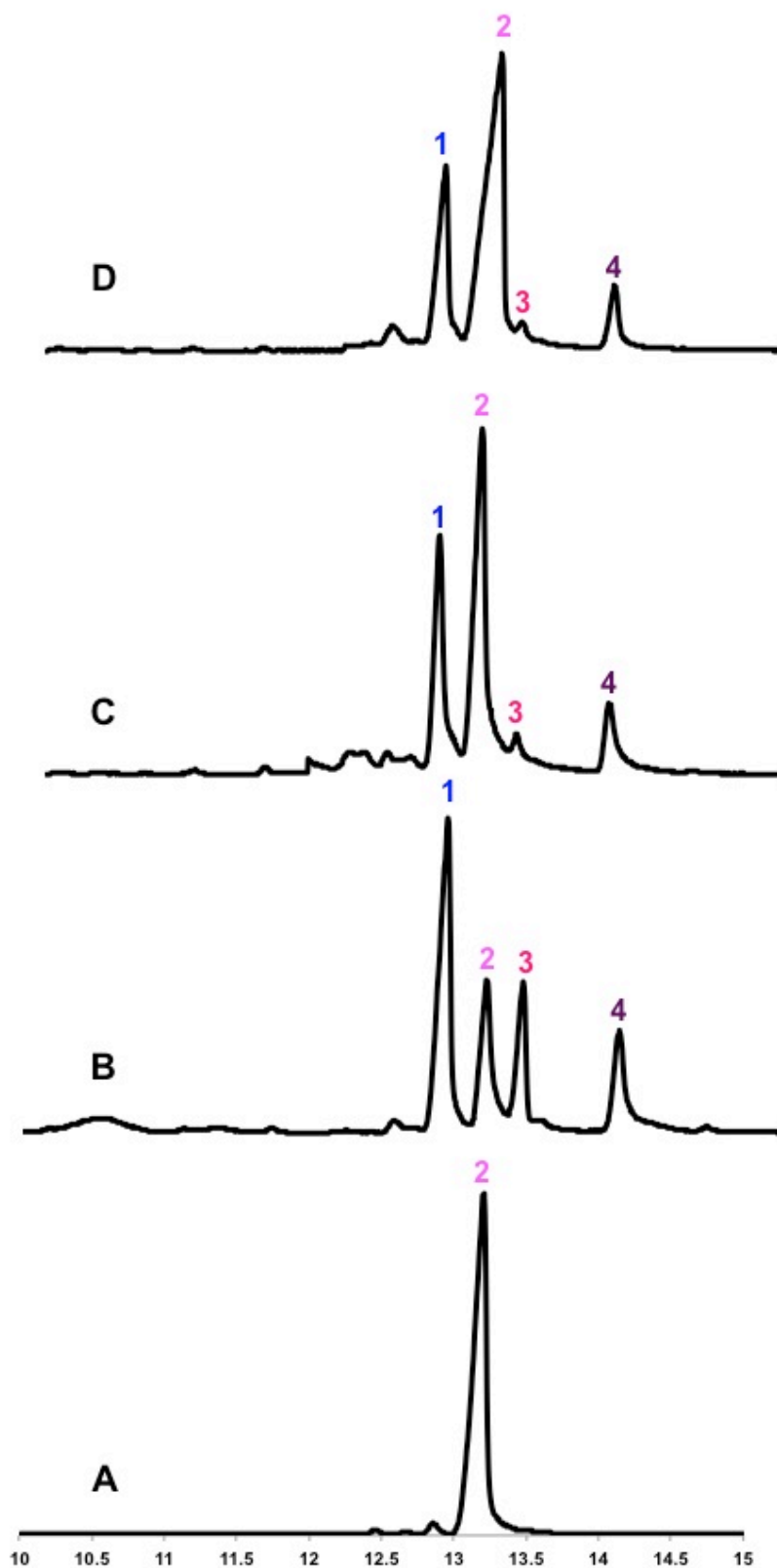


Figure 4.3.13 Total Ion Chromatograms (TICs) of A. Substrate Control, B. Cr10HGO Wild Type Assay, C. Cr10HGO_DM1 Assay, D. Cr10HGO_DM2 Assay; 10-oxogeraniol (1), 10-hydroxygeraniol (2), 10-oxogeraniol (3), 10-hydroxygeraniol (4).

As is evident from the product profiles that three mutants showed the lowest substrate presence in the ratio, namely V277T (7.2 %), W61L (4.6 %) and A302F (6.9 %). In an attempt to completely nullify the presence of the substrate completely, which may contribute to an increase in the final product formation, we thought of combining these changes to obtain a combined activity. Hence two double mutants, namely V277T+W61L (DM1) and V277T+A302F (DM2) were prepared. We expected to see complete elimination of the substrate in the final product profile, but, as a consequence, these double mutants increased the substrate retention to 51.2 % and 66.7 %, respectively (Figure 4.3.13), these being the highest substrate retaining enzymes of all the mutants. Although the individual mutant V277T showed high substrate conversion, when combined with W61, the pushing of the substrate due to increase in the size from valine to threonine is enhanced due to the increase in the cavity size due to W61L, causing the substrate to get further displaced and hence causing a decrease in the activity. In the second case, since V277 and A302 are present towards the same end of the substrate (10C position), the decrease in cavity size triggered due to replacement of both residues by bulkier groups causes to push the substrate in the opposite direction, which once again displaces the substrate and causes a disruption in its interactions with the enzyme.

4.3.9 Substrate Specificity Studies

Cr10HGO wild type and some of the mutants were studied for their substrate specificity with a host of substrates, namely, geraniol, farnesol, nerol, α -bisabolol, α -santalol, β -citronellol, chrysanthemol, eugenol, linalool and menthol.

Table 4.3.2 Substrate specificity study of Cr10HGO wild type and mutants

	WT	W61L	C98A	N115H	H51F	T170A
Geraniol	88.7 %	83.5 %	86.9 %	100.0 %	80.9 %	68.7 %
Geranial	16.3 %	16.5 %	13.1 %	0.0 %	19.1 %	31.3 %
Farnesol	96.6 %	94.7 %	96.4 %	100.0 %	95.7 %	91.7 %
Farnesal	8.4 %	5.3 %	3.6 %	0.0 %	4.3 %	8.3 %
Nerol	89.4 %	66.0 %	66.6 %	100.0 %	53.0 %	59.4 %

Neral	10.6 %	30.8 %	30.1 %	0.0 %	42.4 %	37.1 %
		Geraniol: 2.1 %	Geraniol: 1.9 %		Geraniol: 0.5 %	Geraniol: 2.0 %
β-Citronellol	No activity	95.4 %	93.5 %	100.0 %	92.4 %	94.8 %
		4.6 %	6.5 %	0.0 %	7.6 %	5.2 %
Chrysanthemol	No activity	95.4 %	95.4 %	100.0 %	94.6 %	95.1 %
		4.6 %	4.6 %	0.0 %	5.4 %	4.9 %

While Cr10HGO wild type only reacted with the linear alcohols, geraniol, farnesol and nerol, converting them to their respective aldehydes, the case with the mutants was quite different (Table 4.3.2).

The mutant N115H did not show significant activity with any of the substrates, which agrees with the fact that it had very low activity with the native substrate as well. Whereas, the other mutants, such as W61L, C98A, H51F and T170A, all showed activity with geraniol, farnesol, nerol, β -citronellol and chrysanthemol. In the case of nerol as substrate, all the mutants produced geraniol and geranial in addition to neral which is not observed in case of the wild type protein. The mutant T170A instigated maximum conversion of geraniol (31.3 %) and farnesol (8.3 %) to their respective aldehydes, while the mutant H51F converted nerol (47 %), β -citronellol (7.6 %) and chrysanthemol (5.4 %) to the maximum extent. All these results point towards the fact that an increase in hydrophobicity favors the effective conversion of alcohol group to aldehyde.

4.3.10 Kinetic characteristics of the Cr10HGO mutants

Table 4.3.3 Steady state kinetics values of Cr10HGO mutants

		K_m (μ M)	V_{max} (μ M/sec)	K_{cat} (sec ⁻¹)
10HGO	NADP ⁺	1.50	0.019	3.63
Wild type	Substrate	1.34	0.020	3.88
10HGO_M5	NADP ⁺	6.05	0.018	3.51
N115H	Substrate	6.67	0.019	3.70

10HGO_M7	NADP ⁺	1.65	0.017	3.31
H51F	Substrate	2.03	0.019	3.70
10HGO_M9	NADP ⁺	1.88	0.052	10.12
T170A	Substrate	2.06	0.044	8.56
10HGO_M10	NADP ⁺	2.02	0.083	16.15
W61L	Substrate	2.03	0.074	14.40
10HGO_M11	NADP ⁺	4.03	0.052	10.12
S301A	Substrate	4.18	0.054	10.51
10HGO_M13	NADP ⁺	2.32	0.060	11.67
C98A	Substrate	2.12	0.065	12.65

The mutant N115H showed very less activity with the natural substrate, 10-hydroxygeraniol. This observation finds support in the kinetic data of the enzyme, where the mutant is shown to possess a high K_m of 6.67 μM (substrate) and 6.05 μM (NADP⁺). The mutant S301A, which also shows reduced activity, has higher K_m values as compared to the wild type protein. However, the V_{max} and turnover numbers of these two mutants are the same or slightly lower than those of the wild type. The kinetic constants for the other mutants, namely, N115H, H51F, T170A, W61L and C98A are all higher than those of the wild type, with the mutant W61L having highest V_{max} values (0.083 $\mu\text{M}/\text{sec}$ for NADP⁺ and 0.074 $\mu\text{M}/\text{sec}$ for substrate), supplementing its prominent activity (Table 4.3.3). These kinetic parameters also support the higher activity of these mutants with the other substrates such as geraniol, nerol, farnesol, etc, when compared to the wild type.

4.4 Discussion

Alcohol dehydrogenases are NADP⁺- or NAD⁺- dependent oxidoreductase enzymes, which carry out reversible reduction of ketones and aldehydes. In this study also, Cr10HGO is an oxidoreductase, which catalyzes the reversible reaction(s) between four substrates, 1-hydroxygeraniol, 10-oxogeraniol, 10-hydroxygeranial and 10-oxogeranial. We have performed the site directed mutagenesis of residues present in active site cleft so that the flux of this reaction can be directed more in forward direction for the formation of the di-aldehyde, this would cause a potential increase of flux in the pathway. Similar studies have been performed in the case of the isozymes of Horse Liver Alcohol Dehydrogenases (HLADH)²⁰, and observed that small variation in the amino acid sequences of the two isozymes played a crucial role in differing their substrate specificities²¹. This attribution of difference in substrate specificity with differences in the active site pocket residues was also explicitly described while studying the crystal structure of *Arabidopsis* cinnamyl alcohol dehydrogenases²². In order to alter the substrate specificity, site-directed mutagenesis has been widely used on alcohol dehydrogenases from a range of sources²³⁻²⁶.

Ben-David, Lynn Kamerlin and co-workers have described in their study²⁷ how manipulating the active site hydrophobicity of enzymes, their substrate specificities can be altered. This deduction was based on the saturation mutagenesis carried out on a tryrosine residue in the active site pocket. Studies on altering substrate specificity concerned with branched chain alcohols was conducted by Benner and co-workers²⁸ to identify residues affecting the activity of yeast alcohol dehydrogenases and demonstrated that, W54L showed better activity with straight chain than branched alcohols. However, mutation to corresponding residue in Cr10HGO to W61L resulted in the highest substrate conversion, with a substantial increase in the formation of the desired di-aldehyde product, 10-oxogeranial.

A study on the proton/ hydride transfer pathways in liver alcohol dehydrogenase shed light that H51 and S48 were the residues most likely to affect the proton/ hydride transfers²⁹. The residues R47, D49, D50, E267, R363 and R369, all interact with H51²⁹. H51 is also one of the residues found in the active site of Yeast alcohol dehydrogenase, along with T45, H48 and D53³⁰. Any change at H51 position would influence coenzyme binding and further trigger changes in the catalytic

activity. As expected, when this residue was mutated to phenylalanine, there was a notable change in the product profile, which illustrates the importance of this residue in Cr10HGO as well.

In the alcohol dehydrogenase family of enzymes, the residues W54, L116, M270, I290 were specified by Branden *et al.*³¹, as side chains that could be altered in order to improve activity because of their location at the entrance of the active site. In Cr10HGO, N115 was mutated to L115 (N115L), but this mutant did not show much change in activity of the protein. In the cases of Horse Liver alcohol dehydrogenase and Yeast alcohol dehydrogenase, the corresponding residue at 116 is already a Leucine residue. This mutant of Cr10HGO (N115L) emphasizes that the change from Asparagine to Leucine does not cause a notable change in the activity of the protein. However, when this residue (N115) was mutated to Histidine (N115H), the activity of the protein was substantially decreased (only 24.4 % of the substrate was converted to products). Following this mutation, the active site cavity was increased due to replacement of a bigger residue by a smaller one, increasing the polarity and altering the hydrophobicity of the active site. However, either a saturation mutagenesis at this position and comparison with the role of this residue in other alcohol dehydrogenases would put us in a better position to ascertain its effect on the activity of the protein.

The mutations carried out at the identified residues of Cr10HGO played with the hydrophobicity of the active site pocket and we were able to perceive how this change in polarity directly affected the activity of the protein.

4.5 Conclusion

In order to establish the structure-function relationship, we have identified 12 amino acids present in the active site pocket. Functional characterization of the 14 mutants generated from these amino acids suggests the crucial role in the catalysis of Cr10HGO (Figure 4.4.1). The mutants S301A and reduce the activity of the enzyme to a substantial amount. Serine at 301st position and asparagine at the 115th are highly conserved residues among alcohol dehydrogenases across the species. When these residues were replaced with smaller residues, the activity showed a sudden decline, probably not allowing for enough interactions between the substrate and the mutated residues in the enzyme. Furthermore when T170, C98 and W61 were replaced with residues smaller than present, alanine and lysine, respectively, the activity was once again affected. In this case, the formation of the di-aldehyde metabolite, 10-oxogeranial, shows an increase. The residues W61 and A302 envelop the binding site pocket and the mutations W61L and A302F, causing a change in the hydrophobicity greatly affect the forward reaction.

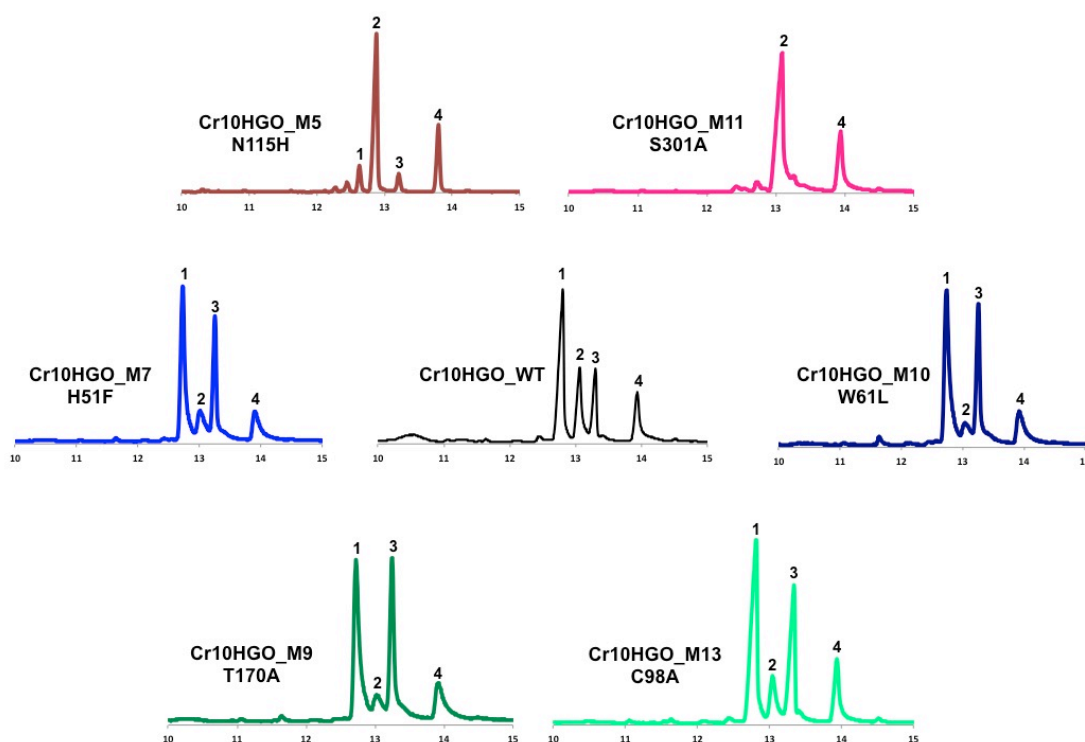


Figure 4.5.1 Total Ion Chromatograms (TICs) of Cr10HGO_WT, Cr10HGO_M5, Cr10HGO_M7, Cr10HGO_M9, Cr10HGO_M10, Cr10HGO_M11 and Cr10HGO_M13; 10-oxogeranial (1), 10-hydroxygeranial (2), 10-oxogeranial (3), 10-hydroxygeranial (4).

As can be evident from both the different scenarios, replacement of the highly conserved residues with smaller amino acids affects the interaction of the substrate and the protein, which in turn affects the activity of the protein. The results observed indicate that an increase in the hydrophobicity surrounding the active site pocket greatly favors the reaction in the forward direction, and hence enhances the percentage formation of the di-aldehyde, 10-oxogeranial. The kinetic parameters of these significant mutants are in well agreement with the characteristics of their product profile. Conclusively, mutation of Cr10HGO so as to increase the production of 10-oxogeranial facilitates higher substrate availability for the next enzyme in the Iridoid biosynthetic pathway, which would substantially increase the flux of iridoid formation.

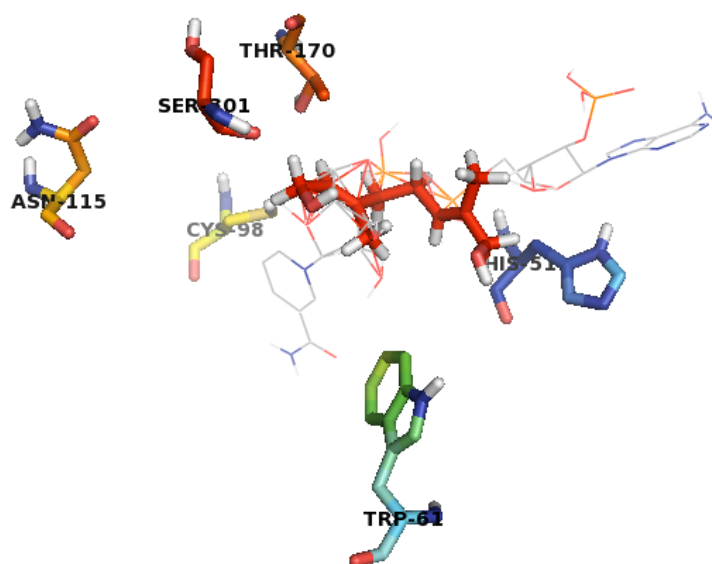


Figure: 4.5.2 Mutated amino acid positions around the substrate 10-hydroxygeraniol and NADP⁺.

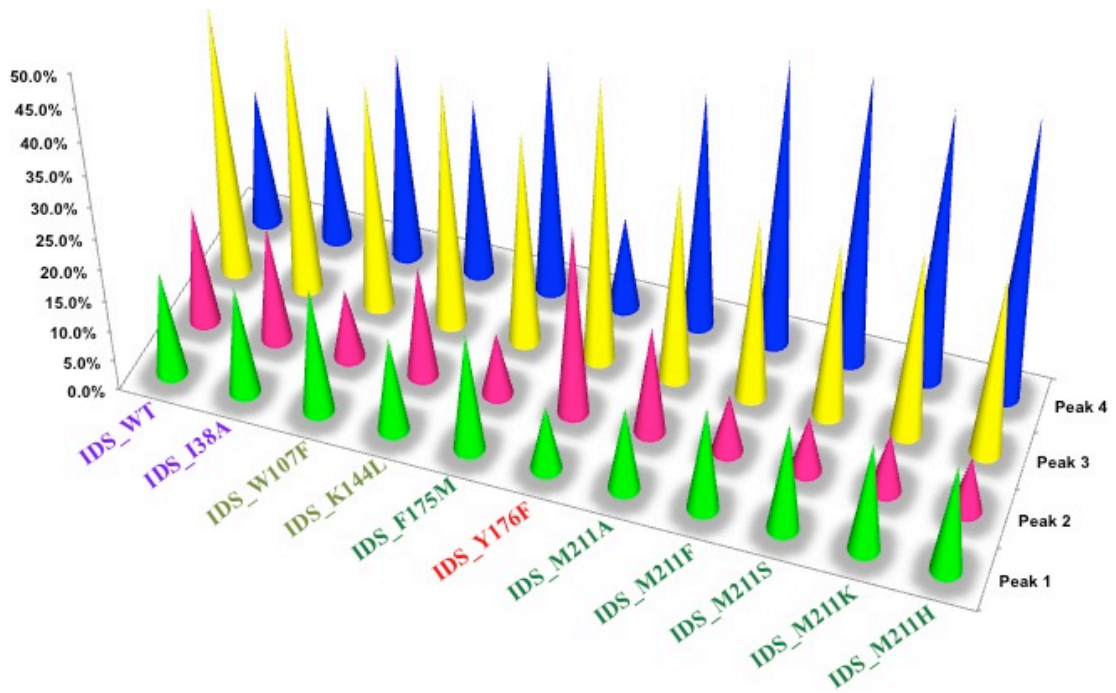
4.6 References

1. Hays, S.B. Some effects of reserpine, a tranquilizer, on House Fly. *J. Econ. Entomol.* **58**, 782-784 (1965).
2. Yui, T. & Takeo, Y. Neuropharmacological studies on a new series of ergot alkaloids; elymoclavine as a potent analeptic on reserpine-sedation. *Jpn. J. Pharmacol.* **7**, 157-161 (1958).
3. Yui, T. & Takeo, Y. Neuropharmacological studies on a new series of ergot alkaloids; the effects on electrocorticogram of rabbits. *Jpn. J. Pharmacol.* **7**, 162-168 (1958).
4. Shurin, G.V., Tourkova, I.L., Kaneno, R. & Shurin, M.R. Chemotherapeutic Agents in Noncytotoxic Concentrations Increase Antigen Presentation by Dendritic Cells via an IL-12-Dependent Mechanism. *J. Immunol.* **183**, 137-144 (2009).
5. van der Heijden, R., Jacobs, D.I., Snoeijer, W., Hallared, D. & Verpoorte, R. The *Catharanthus* alkaloids: Pharmacognosy and biotechnology. *Curr. Med. Chem.* **11**, 607-628 (2004).
6. Rai, A., Smita, S.S., Singh, A.K., Shanker, K. & Nagegowda, D.A. Heteromeric and Homomeric Geranyl Diphosphate Synthases from *Catharanthus roseus* and Their Role in Monoterpene Indole Alkaloid Biosynthesis. *Mol. Plant* **6**, 1531-1549 (2013).
7. Madyastha, K.M., Meehan, T.D. & Coscia, C.J. Characterization of a cytochrome P-450 dependent monoterpene hydroxylase from the higher plant *Vinca rosea*. *Biochem.* **15**, 1097-1102 (1976).
8. Simkin, A. J., Miettinen, K., Claudel, P., Burlat, V., Guirimand, G., Courdavault, V., Papon, N., Meyer, S., Godet, S., St-Pierre, B., Giglioli-Guivarc'h, N., Fischer, M. J. C., Memelink, J., & Clastre, M. Characterization of the plastidial geraniol synthase from Madagascar periwinkle which initiates the monoterpene branch of the alkaloid pathway in internal phloem associated parenchyma. *Phytochem.* **85**, 36-43 (2013).
9. Herbert, R.B. The biosynthesis of plant alkaloids and nitrogenous microbial metabolites. *Nat. Prod. Rep.* **2**, 163-179 (1985).
10. Uesato, S., Matsuda, S., Iida, A., Inouye, H. & Zenk, M.H. Intermediacy of 10-hydroxygeraniol, 10-hydroxynerol and iridodial in the biosynthesis of Ajmaline and Vomilenine in *Rauwolfia serpentina* suspension cultures. *Chem. Pharm. Bull.* **32**, 3764-3767 (1984).
11. Uesato, S., Matsuda, S. & Inouye, H. Mechanism for iridane skeleton formation from acyclic monoterpenes in the biosynthesis of secologanin and vindoline in *Catharanthus roseus* and *Lonicera morrowii*. *Chem. Pharm. Bull.* **32**, 1671-1674 (1984).
12. Geu-Flores, F., Sherden, N. H., Courdavault, V., Burlat, V., Glenn, W. S., Wu, C., Nims, E., Cui, Y. H., & O'Connor, S. E. An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature* **492**, 138-142 (2013).
13. Uesato, S., Ueda, S., Kobayashi, K. & Inouye, H. Mechanism of iridane skeleton formation in the biosynthesis of iridoid glucosides in *Gardenia jasminoides* cell cultures. *Chem. Pharm. Bull.* **31**, 4185-4188 (1983).
14. Ikeda, H., Esaki, N., Nakai, S., Hashimoto, K., Uesato, S., Soda, K., & Fujita, T. Acyclic monoterpene primary alcohol:NADP⁺ oxidoreductase of *Rauwolfia serpentina* cells: the key enzyme in biosynthesis of monoterpene alcohols. *J. Biochem.* **109**, 341-347 (1991).
15. Bomati, E.K. & Noel, J.P. Structural and kinetic basis for substrate selectivity in *Populus tremuloides* sinapyl alcohol dehydrogenase. *Plant Cell* **17**, 1598-1611 (2005).
16. Krithika, R., Srivastava, P. L., Rani, B., Kolet, S. P., Chopade, M., Soniya, M., & Thulasiram, H. V. Characterization of 10-Hydroxygeraniol Dehydrogenase from

-
- Catharanthus roseus* Reveals Cascaded Enzymatic Activity in Iridoid Biosynthesis. *Sci. Rep.* **5**, 8258-8264 (2015).
17. Youn, B., Camacho, R., Moinuddin, S. G. A., Lee, C., Davin, L. B., Lewis, N. G., & Kang, C. H. Crystal structures and catalytic mechanism of the Arabidopsis cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4. *Org. Biomol. Chem.* **4**, 1687-1697 (2006).
 18. Ramachandran, G.N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95-99 (1963).
 19. Jornvall, H., Persson, B. & Jeffery, J. Characteristics of alcohol polyol dehydrogenases - the zinc-containing long-chain alcohol dehydrogenases. *Eur. J. Biochem.* **167**, 195-201 (1987).
 20. Theorell, H., Åkeson, Å., Liszka-kopeć, B. & Zalenski, C.d. Equilibrium and rate constants for the reaction between NADH and horse liver alcohol dehydrogenases "EE", "ES", and "SS". *Arch. Biochem. Biophys.* **139**, 241-247 (1970).
 21. Park, D.H. & Plapp, B.V. Isoenzymes of horse liver alcohol dehydrogenase active on ethanol and steroids. cDNA cloning, expression, and comparison of active sites. *J. Biol. Chem.* **266**, 13296-13302 (1991).
 22. Youn, B., Camacho, R., Moinuddin, S. G. A., Lee, C., Davin, L. B., Lewis, N. G., & Kang, C. H. Crystal structures and catalytic mechanism of the Arabidopsis cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4. *Org. Biomol. Chem.* **4**, 1687-1697 (2006).
 23. Green, D.W., Sun, H.W. & Plapp, B.V. Inversion of the substrate specificity of yeast alcohol dehydrogenase. *J. Biol. Chem.* **268**, 7792-7798 (1993).
 24. Light, D. R., Dennis, M. S., Forsythe, I. J., Liu, C. C., Green, D. W., Kratzer, D. A., & Plapp, B. V. Alpha-isoenzyme of alcohol dehydrogenase from monkey liver. Cloning, expression, mechanism, coenzyme, and substrate specificity. *J. Biol. Chem.* **267**, 12592-12599 (1992).
 25. Murali, C. & Creaser, E.H. Protein engineering of alcohol dehydrogenase 1. Effects of two amino acid changes in the active site of yeast ADH-1. *Protein Eng.* **1**, 55-57 (1986).
 26. Park, D.H. & Plapp, B.V. Interconversion of E and S isoenzymes of horse liver alcohol dehydrogenase. Several residues contribute indirectly to catalysis. *J. Biol. Chem.* **267**, 5527-5533 (1992).
 27. Blaha-Nelson, D., Kruger, D.M., Szeler, K., Ben-David, M. & Kamerlin, S.C.L. Active Site Hydrophobicity and the Convergent Evolution of Paraoxonase Activity in Structurally Divergent Enzymes: The Case of Serum Paraoxonase 1. *J. Am. Chem. Soc.* **139**, 1155-1167 (2017).
 28. Weinhold, E.G. & Benner, S.A. Engineering yeast alcohol dehydrogenase. Replacing Trp54 by Leu broadens substrate specificity. *Protein Eng.* **8**, 457-461 (1995).
 29. Cui, Q., Elstner, M. & Karplus, M. A theoretical analysis of the proton and hydride transfer in liver alcohol dehydrogenase (LADH). *J. Phys. Chem. B* **106**, 2721-2740 (2002).
 30. Raj, S.B., Ramaswamy, S. & Plapp, B.V. Yeast Alcohol Dehydrogenase Structure and Catalysis. *Biochem.* **53**, 5791-5803 (2014).
 31. Eklund, H., Nordström, B., Zeppezauer, E., Söderlund, G., Ohlsson, I., Boiwe, T., Söderberg, B.-O., Tapia, O., Brändén, C.-I., & Åkeson, Å. Three-dimensional structure of horse liver alcohol dehydrogenase at 2.4 Å resolution. *J. Mol. Biol.* **102**, 27-59 (1976).
-

Chapter 5

Engineering of iridoid synthase (CrIDS) by site-directed mutagenesis to understand its activity profile



Chapter 5

Engineering of iridoid synthase (CrIDS) by site-directed mutagenesis to understand its activity profile

Iridoid synthase (CrIDS) is an NADPH-dependent enzyme, belonging to a class of monoterpene indole cyclases, which utilizes the linear 10-oxogeranial as a substrate and through a sequential reduction and cyclization, produces an equilibrium mixture of nepetalactols and iridoids, the final product in iridoid biosynthesis. In an attempt to establish the structure-function relationship, a homology model of CrIDS was built and used for docking studies. Based on amino acid interaction with the substrate (10-oxogeranial) in active site pocket, 6 residues were identified and 10 mutations were designed. Analysis of these mutants revealed that I38A did not have much effect on the activity of the enzyme, whereas the mutation at Y156 (replacing tyrosine with phenylalanine) reduced the product (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol to a notable extent. On the other hand, the mutants F175M and M211A increased the ratio of (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol in the product pool to 17 folds. Depending on these results, another 4 mutations were designed, M211F, M211S, M211K and M211H, which increased the desired stereospecific product formation substantially. These site-directed mutagenesis studies established the importance of the methionine at the 211th position for the determination of product profile of CrIDS. These studies would greatly benefit in increasing the flux towards secologanin formation.

5.1 Introduction

Iridoids are a class of monoterpenoids that are distinctly bicyclic in nature and are found in a wide family of plants like *Apocynaceae*, *Loganiaceae*, *Rubiaceae*, *Lamiaceae*, etc. and some animal species. They are mostly known to occur as glycosides in plants, i.e., bound to glucose. Iridoids possess a range of significant biological and pharmacological activities like Geniposidic acid from *Alibertia myrcifolia* serves as an antifungal agent, Phlomiol from *Eremostachys laciniata* as an anti-bacterial agent, the anti-cancer agent, Acevaltrate from *Valeriana glechomifolia* and Catalpol from *Rahmannia glutinosa*, an anti-hyperglycemic agent^{1,2}. *Paederia scandens*, a *Rubiaceae* plant, produces paederoside, an iridoid glycoside that acts as a potent restraint against the ladybird beetle^{3,4}. The anti-inflammatory activities of several iridoid glycosides reptoside⁵, boschnaloside⁶, 8- epideoxyloganic acid⁷, etc have been well studied.

Iridoids are made up of a six-membered ring containing an oxygen, bound to a cyclopentane ring. The iridoid skeleton is formed by the cyclisation of 10-oxogeraniol, which is biosynthesized from geraniol through 10-hydroxygeraniol to yield iridodials. Further, iridodial is subsequently oxidised into iridotrial. *Catharanthus roseus* Iridoid synthase (CrIDS)⁸ is an NADPH-dependent enzyme, belonging to a class of monoterpene indole cyclases. The deduced amino acid sequence of CrIDS was found to have a calculated molecular weight of 43.69 kDa and pI 7.08, comprising 5- β progesterone reductase-like active site domain (positions 29–326) and the NAD dependent epimerase/dehydratase family (positions 30–255). CrIDS utilizes the linear 10-oxogeraniol as a substrate and by a sequential reduction and cyclization, produces an equilibrium mixture of nepetalactols and iridoids, the ultimate product in iridoid biosynthesis. This cyclization leading to the formation of the basic iridoid skeleton is a very significant step in the biosynthesis of strictosidine, which is the precursor for the formation of a number of pharmaceutically valuable potent drugs like quinine, camptothecin, ajmalicine, serpentine, vinblastine and vincristine⁹.

In order to carry out the structure-function studies of *Catharanthus roseus* Iridoid synthase (CrIDS), site-directed mutagenesis of the residues that seem to be affecting the catalytic site pocket were carried out and studied.

5.2 Materials and Methods

5.2.1 Computational modeling and docking

Homology model of Iridoid synthase (CrIDS) was built using the co-ordinates of Progesterone 5 β -Reductase from *Digitalis lanata*¹⁰ as template using Schrodinger-Maestro 1.6. The lowest energy structures were validated using Ramachandran plot. Reaction substrates were constructed using ChemBioDraw Ultra 13.0 for docking studies. To determine the amino acid interactions involved in the product formation, docking was performed using the substrate, 10-oxogeranial and images were generated by PYMOL (v.1.6).

5.2.2 Site-Directed mutagenesis

All the mutation reactions were carried out with a pET32a expression vector (Invitrogen) harboring the coding sequence for Iridoid synthase as a template⁸. The mutants were generated using the QuikChange Lightning site directed mutagenesis kit from Agilent according to the manufacturer's instructions. The mutagenesis reaction mixtures were transformed into *Escherichia coli* XL10 Gold Cells and plated onto Luria Bertani agar plates containing ampicillin. Individual colonies were picked and grown under selection. Plasmid DNA was then isolated and the mutant clones were identified by sequencing. The mutagenic primers (with mutagenic bases shown in boldface and underlined), which were designed on the basis of homology-based model of *CrIDS*, are listed in the table 5.2.1.

Table 5.2.1. Primers for CrIDS site-directed mutagenesis. The mutagenesis bases are shown in red.

CrIDS_M1	IDS_I38A_FWD	GGAGTCACCGGC GCT GTTGGCAGCAGT
	IDS_I38A_REV	ACTGCTGCCAAC AGC GCCGGTGACTCC
CrIDS_M2	IDS_W107F_FWD	TTCTATGTTTCT TTT ATTGGATCAGAA
	IDS_W107F_REV	TTCTGATCCAAT AAA AGAAACATAGAA
CrIDS_M3	IDS_K144L_FWD	CAAACAGGAATC CTA CATTATTTTGGT
	IDS_K144L_REV	ACCAAATAATG TAG GATTTCCTGTTG

CrIDS_M4	IDS_F175M_FWD	AATGTCCAAAAT ATG TACCATGATCTT
	IDS_F175M_REV	AAGATCATGGTAC CAT ATTTTGGACATT
CrIDS_M5	IDS_Y176F_FWD	GTCCAAAATTT TTC CATGATCTTGAA
	IDS_Y176F_REV	TTCAAGATCATG GAA GAAATTTGGAC
CrIDS_M6	IDS_M211A_FWD	CCATGTAGTATG GCG AACATTGTCAGT
	IDS_M211A_REV	ACTGACAATGTT CGC CATACTACATGG
CrIDS_M7	IDS_M211F_FWD	CCATGTAGTATG TTC AACATTGTCAGT
	IDS_M211F_REV	ACTGACAATGTT GAA CATACTACATGG
CrIDS_M8	IDS_M211S_FWD	CCATGTAGTATG AGC AACATTGTCAGT
	IDS_M211S_REV	ACTGACAATGTT GCT CATACTACATGG
CrIDS_M9	IDS_M211L_FWD	CCATGTAGTATG AAG AACATTGTCAGT
	IDS_M211L_REV	ACTGACAATGTT CTT CATACTACATGG
CrIDS_M10	IDS_M211H_FWD	CCATGTAGTATG CAT AACATTGTCAGT
	IDS_M211H_REV	ACTGACAATGTT ATG CATACTACATGG

5.2.3 Heterologous expression and purification of CrIDS mutants

pET32a vector harboring CrIDS mutants were expressed and purified as done earlier⁸. Expression was carried out in Rosetta2 (DE3) cells, cells were grown in Terrific Broth at 37 °C and induced with IPTG at a final concentration of 1mM and incubated for 12 hours at 16 °C. After the induction, the culture was harvested by centrifugation at 4500 × g for 20 minutes. The cell pellet (5 g/L) was re-suspended in 10 mL/g of cell pellet of Lysis buffer (50 mM NaH₂PO₄, 300 mM NaCl, 0.5 % CHAPS, pH 8.0, 10 % v/v Glycerol) with 1 mg/mL concentration of lysozyme and 100 µL/5 mL of protease inhibitor cocktail from Sigma and incubated on ice for 30 minutes. After the incubation period, the cells were sonicated with a pulse of 30 sec ON and 30 sec OFF for 10 cycles. The sample was then centrifuged at 5,000 × g for 10 minutes at 4 °C. The crude lysate was purified using Ni-NTA column (2 mL resin /

g cell pellet). The protein was finally eluted out in the elution buffer (50 mM NaH_2PO_4 , 300 mM NaCl, pH 8.0, 250 mM Imidazole, 10 % v/v Glycerol) each. Purified protein fractions (after checking on 12 % SDS gel) were pooled together and desalted on Hi-PrepTM 26/10 Desalting Columns with desalting buffer (20 mM MOPS, pH 8.0, 10 % v/v Glycerol) using AKTA (GE Healthcare). The desalted proteins were estimated using Bradford reagent (Bio-Rad) and a Bovine Serum Albumin standard.

5.2.4 Enzymatic characterization and product analysis

CrIDS was assayed for its activity using 100 μg purified protein and 1 mM 10-oxogeranial as the substrate in 500 μl MOPS Buffer (20 mM MOPS, 10 % v/v Glycerol, pH 7.0) at 30 °C for 2 hours. Assay mixtures were extracted with dichloromethane (DCM). The combined organic phase was dried over sodium sulphate, reduced to ~ 50 μL with a stream of dry nitrogen and subjected to GC and GC-MS analyses for product characterization by comparing with authentic standards. 1 μL of the extract was injected onto a 30 m \times 0.25 mm \times 0.25 μm HP-5 capillary GC column with a temperature gradient from 60 to 120 °C at 20 °C per min, followed by a temperature gradient from 120 to 170 °C at 2.5 °C per min and a final temperature gradient from 170 to 190 °C at 20 °C per min, with a final hold temperature for 2.5 min. For better resolution, the M211 mutants were analyzed on the following temperature program: 60 to 120 °C at 5 °C per min, followed by a temperature gradient from 120 to 135 °C at 1 °C per min and a final temperature gradient from 135 to 190 °C at 20 °C per min, with a final hold temperature for 2.5 min.

5.2.5 Determination of kinetic parameters of CrIDS mutants

Steady-state kinetics was performed in 20 mM MOPS, 10 % v/v Glycerol, pH 7.0 at 30 °C with varying substrate concentrations, ranging from 1 to 100 μM with saturation concentration of cofactor, NADPH (200.0 μM), using 5 nmoles of the enzyme, and vice versa. The reactions were followed by measuring changes in NADPH concentrations at 340 nm. The kinetic data were fitted with the Graph Pad Prism software and the kinetic parameters calculated using Michaelis-Menten plots.

5.3 Results

5.3.1 Homology modeling and docking

The CrIDS amino acid sequence displays 59 % sequence identity with Progesterone 5- β -reductase from *Digitalis lanata* (PDB ID: 2V6F¹⁰). As mentioned, the deduced amino acid sequence of CrIDS was found to have a calculated molecular weight of 43.69 kDa and pI 7.08, comprising 5-beta progesterone reductase-like active site domain (positions 29–326) and the NAD dependent epimerase/dehydratase family (positions 30–255). To understand the basis of the product profile of *CrIDS*, homology based model was built using co-ordinates of Progesterone 5- β -reductase from *Digitalis lanata* (PDB ID: 2V6F) as a template, based on an overall sequence similarity of 59 %, using Schrodinger (Figure 5.3.1).

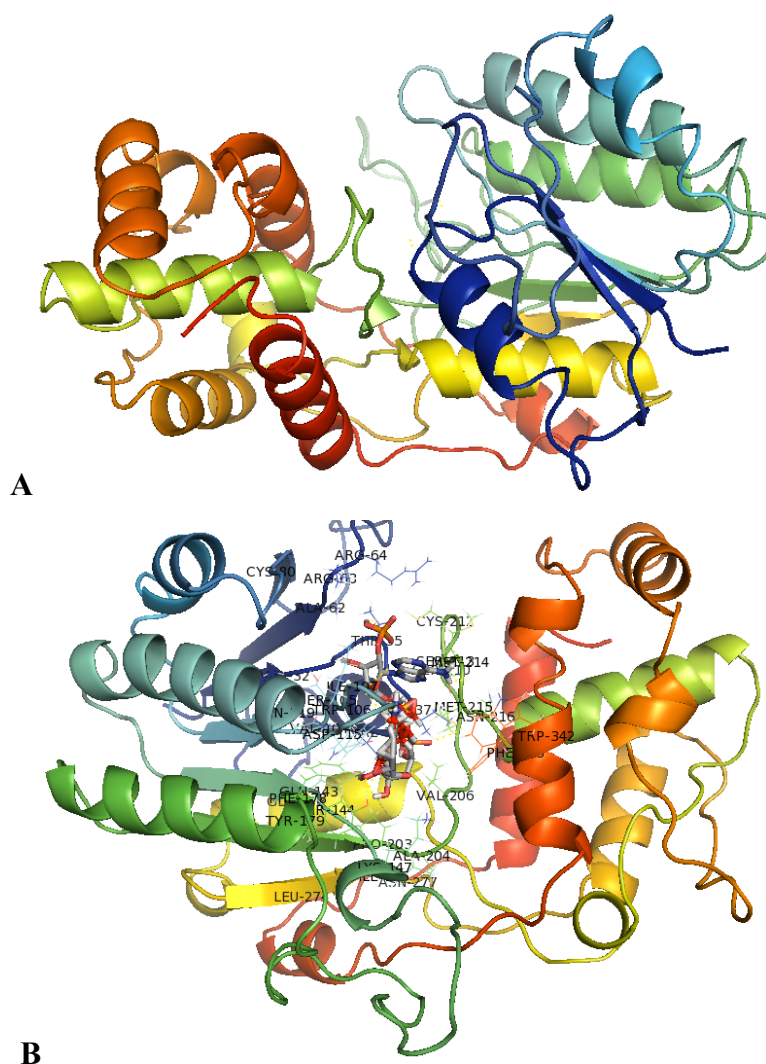


Figure 5.3.1 A. Homology model of Iridoid synthase (CrIDS) built using co-ordinates of Progesterone 5 β -Reductase from *Digitalis lanata* as a template. B. 10-oxogeranial docked with CrIDS model.

The homology model of CrIDS was validated using Ramachandran plot with default parameters. Ramachandran PROCHEK validation tools suggested that 282 residues (87.3 %) of the modeled *CrIDS* were present in the most favored region, whereas 32 residues (9.9 %) were present in the additional allowed regions. There were only 2 residues (0.6 %) present in the disallowed region (Figure 5.3.2). This result clearly suggests that the model built is satisfactory and can be used for further docking studies.

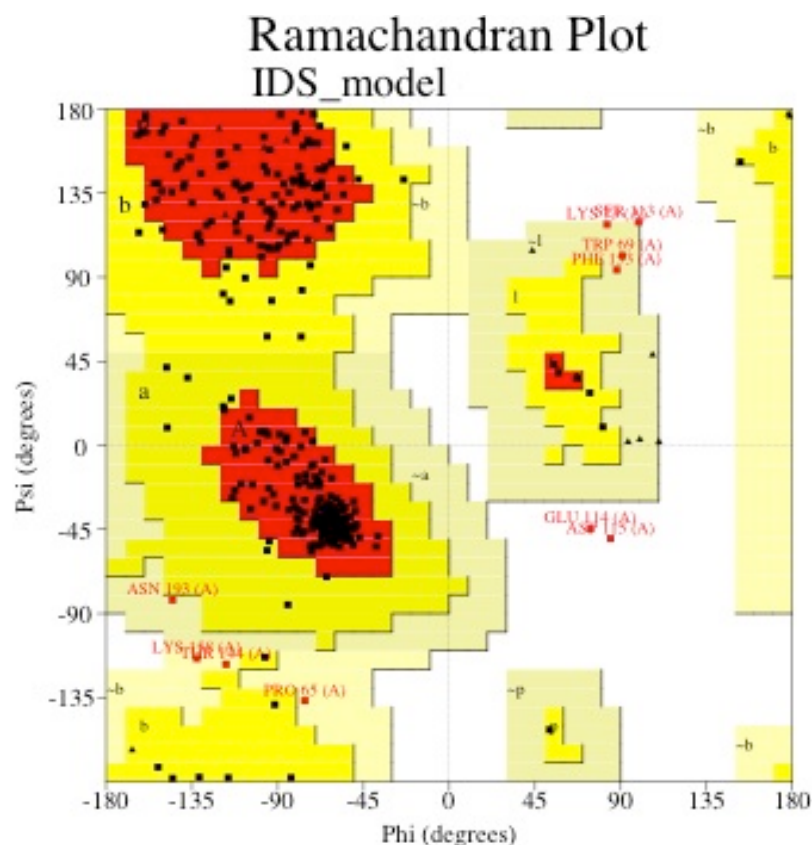


Figure 5.3.2: Ramachandran Plot of the built model of CrIDS, showing maximum residues (90.1 %) in the most favored region

5.3.2 Mechanism of action of CrIDS on 10-oxogeranial

CrIDS acts on the di-aldehyde, 10-oxogeranial, converting it into an equilibrium mixture of *cis-trans*-iridodials and *cis-trans*-nepetalactols. CrIDS is an NADPH dependent enzyme, and is active only in its presence. Iridoids are basically a type of cyclopentanopyrans, i.e., they are bicyclic *cis*-fused cyclopentane-pyrans found in numerous plants and a few animals. In the wild-type assay of CrIDS, 4 distinct peaks can be observed (Figure 5.3.3), of which the first 3 are the mixtures of *cis-trans*-iridodials and *cis-trans*-nepetalactols, and the fourth peak represents

nepetalactol, particularly, (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol (21.7 %), which is utilized in further reactions in the pathway.

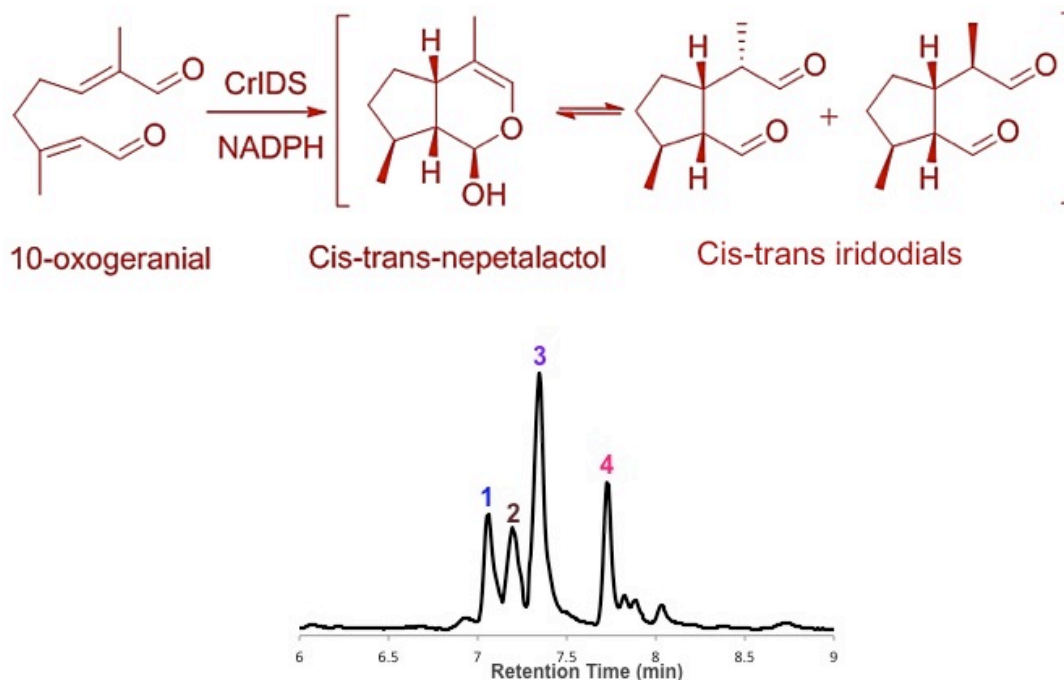


Figure 5.3.3: CrIDS mediated reaction in *C. roseus* for biosynthesis of iridoids (top) and product profile of CrIDS enzyme assay (bottom); Peaks 1,2 and 3: *Cis-trans*-iridoids/ *cis-trans*-nepetalactols, Peak 4: (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol.

5.3.3 Site-directed mutagenesis and characterization

In order to understand catalytic determinants of CrIDS, 6 residues were identified based on their interactions with the docked substrate as well as co-factor (NADPH) and mutations were designed. After incorporating the site-specific mutations, the CrIDS mutants were expressed in *E. coli* Rosetta 2(DE3) cells and the purified proteins were assayed for their activity with 10-oxogeranial as substrate, in the presence of NADPH as the co-factor (Figure 5.3.4). The products of the reaction were analyzed on a GC fitted with HP-5 capillary column. Taking into account the initial results, 4 additional mutations were designed and studied as before.

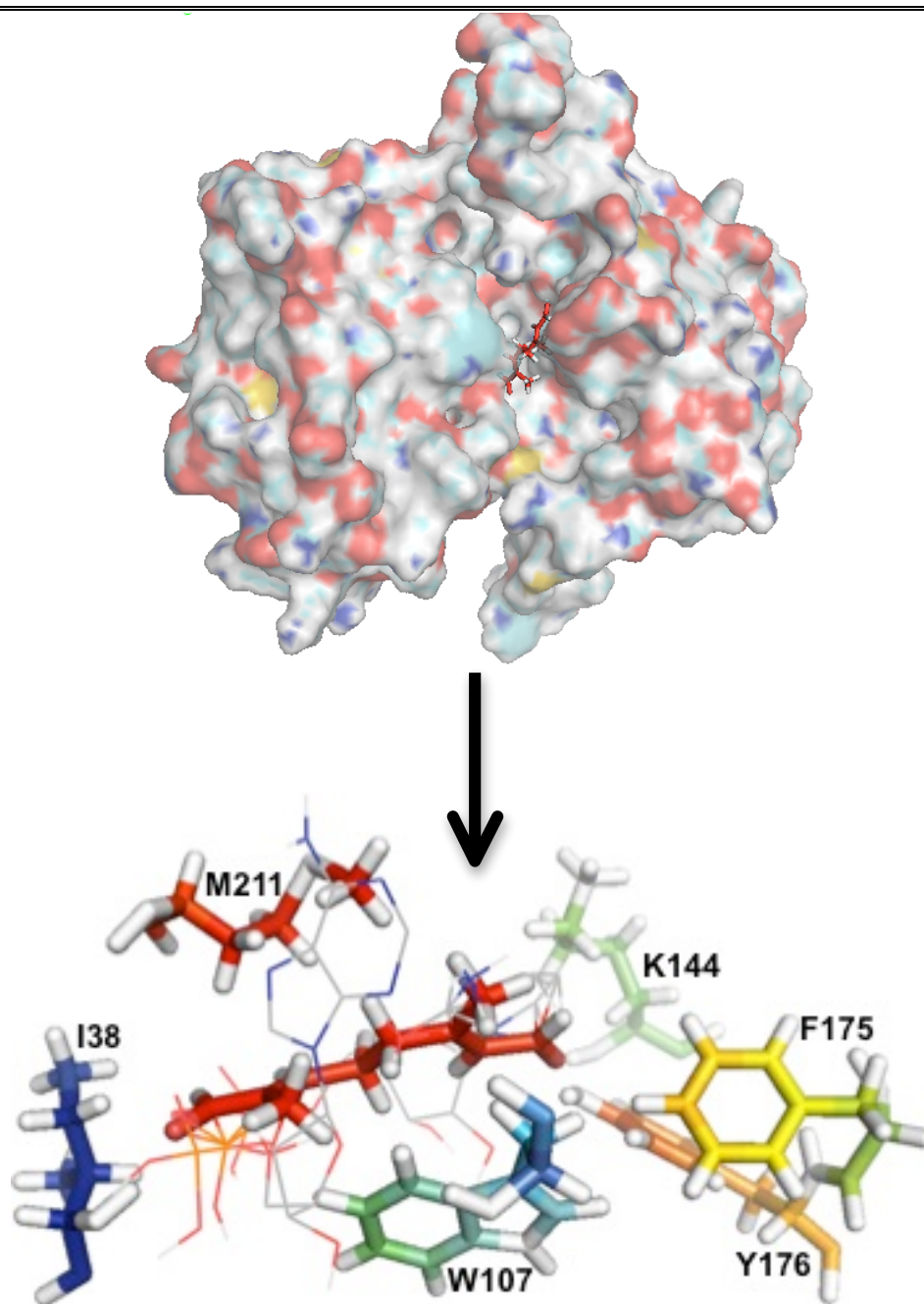


Figure 5.3.4: Homology model of CrIDS showing amino acid interactions with the substrate.

5.3.4 Heterologous expression and purification of CrIDS mutants

Protein expression was carried out in Rosetta2 DE3 cells. The recombinant protein of CrIDS in pET32a was expressed and purified as discussed earlier. The purified protein fractions resulted in bands at 61.5 kDa on a 12 % SDS-PAGE (Figure 4.3.5). The purified fractions were desalted, flash-frozen and stored at -80 °C till used. The product profiles of all the mutants are summarized in Table 5.3.1.

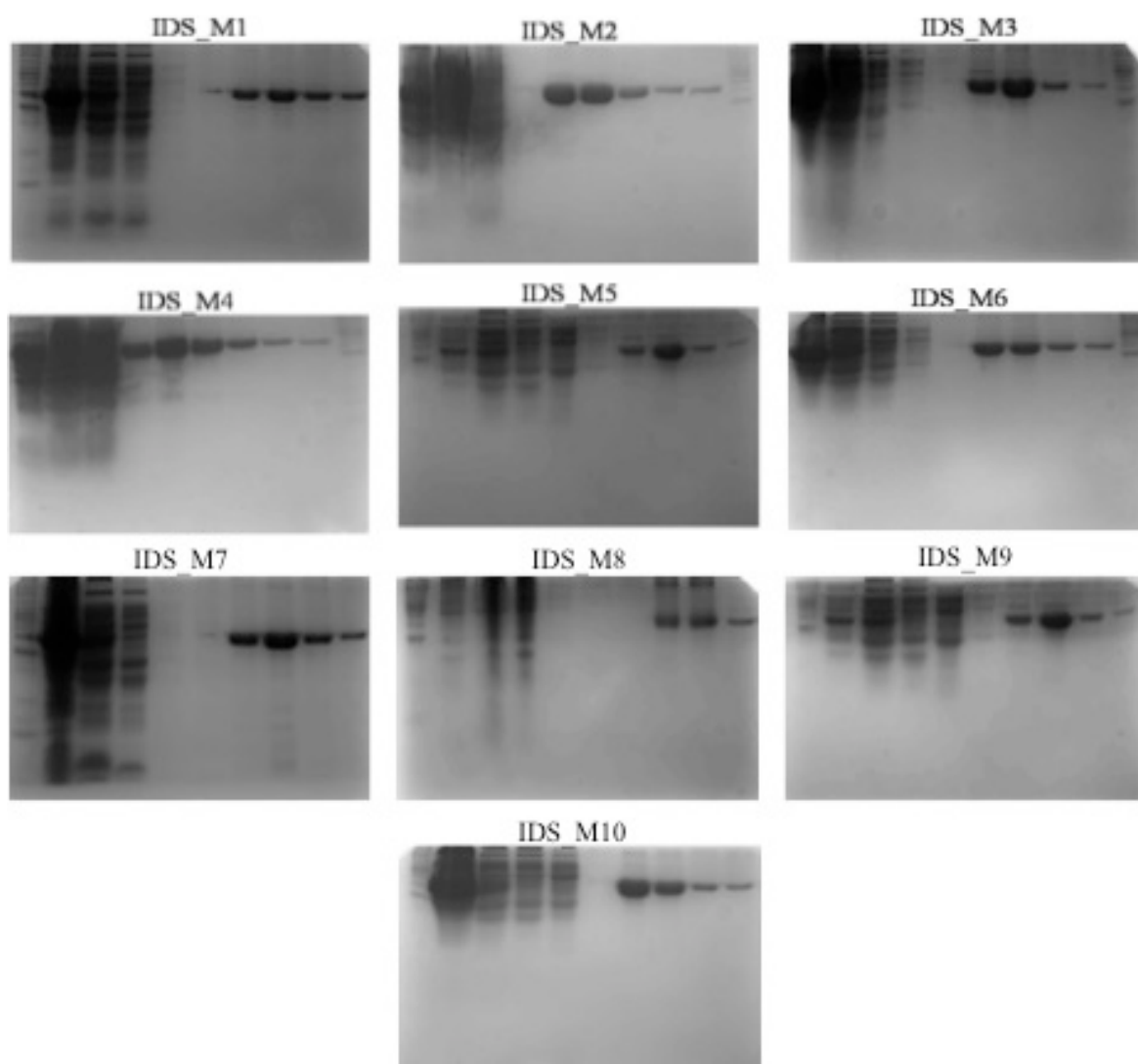


Figure 5.3.5: SDS gel images of CrIDS mutants' protein purification (M. Wt: 61.5 KDa).

5.3.5 Residues showing a sharp decline in the activity of CrIDS

Of the 10 mutations carried out, almost all of them seemed to significantly affect the activity of the protein.

5.3.5.1 Tyrosine 176 (Y176F)

As is characteristic in short-chain dehydrogenase/reductase (SDRs) superfamily, progesterone reductases possess a catalytic triad comprising Tyr/ Lys/ Ser¹¹, which seemed to be lacking in the case of CrIDS. However, the residue Tyr176 was found to be present in CrIDS, where it played a crucial role in the stability of the enzyme as the –OH of tyrosine was found to be involved in a hydrogen-bond formation with NADPH. Also, this tyrosine was near the 10-CHO of the substrate.

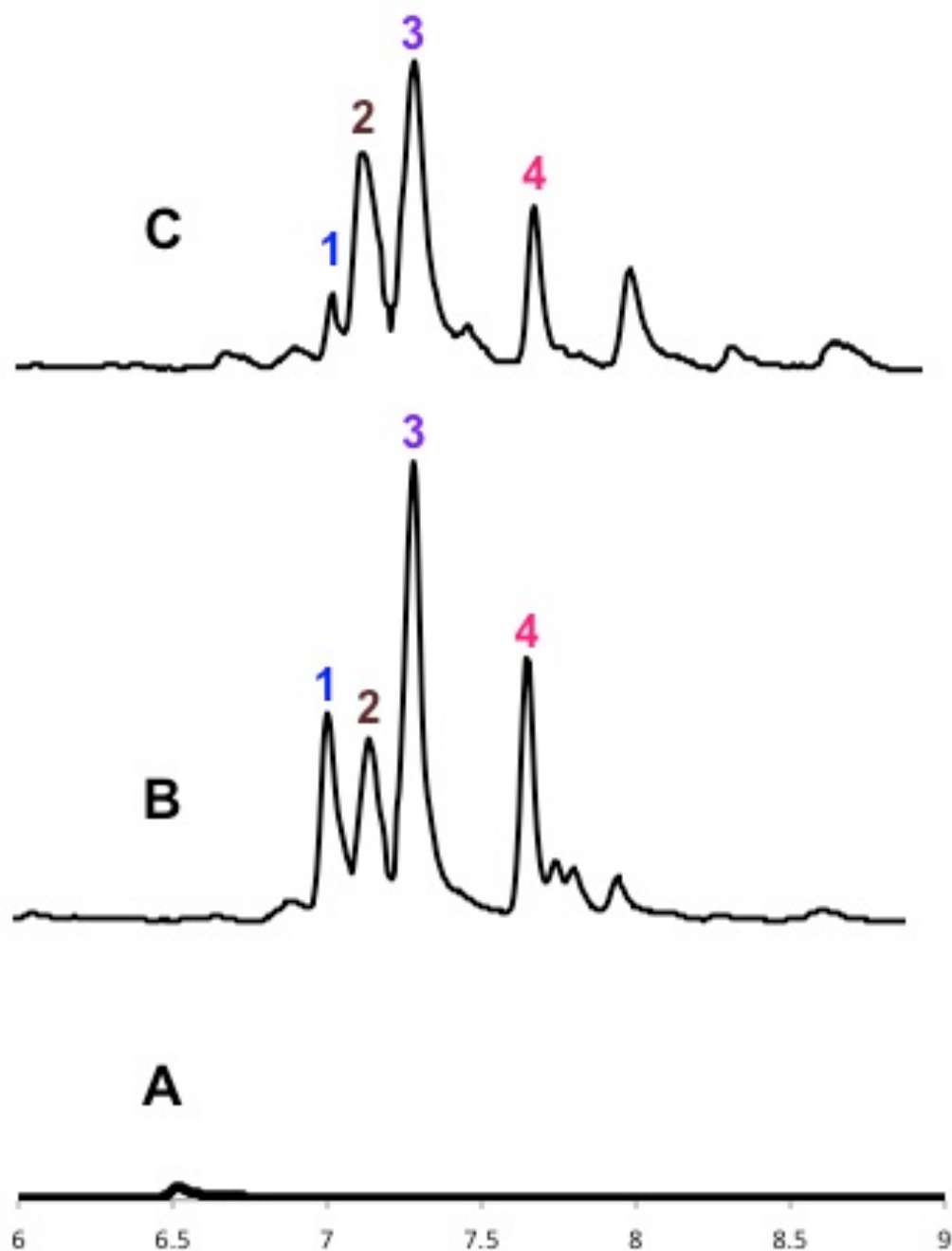


Figure 5.3.6: Total Ion Chromatograms (TICs) of **A.** Substrate Control, **B.** CrIDS Wild Type Assay, **C.** CrIDS_Y176F Assay, 1-3: *Cis-trans*-iridodials and *cis-trans*-nepetalactols, 4: *Cis-trans*-nepetalactol.

It is also predicted to be involved in rotation about C4-C5¹², so as to enable cyclization and lactol production, leading to the formation of the final product, nepetalactol. The significance of Tyr176 was emphasized when tyrosine was replaced

with phenylalanine, which has a similar backbone, except for an –OH group, thus causing an increase in hydrophobicity and eliminating the group that was involved in the interactions. Upon enzymatic characterization, it was observed that the activity of CrIDS was drastically reduced, also affecting the formation of nepetalactol.

5.3.6 Residues not significantly affecting the activity

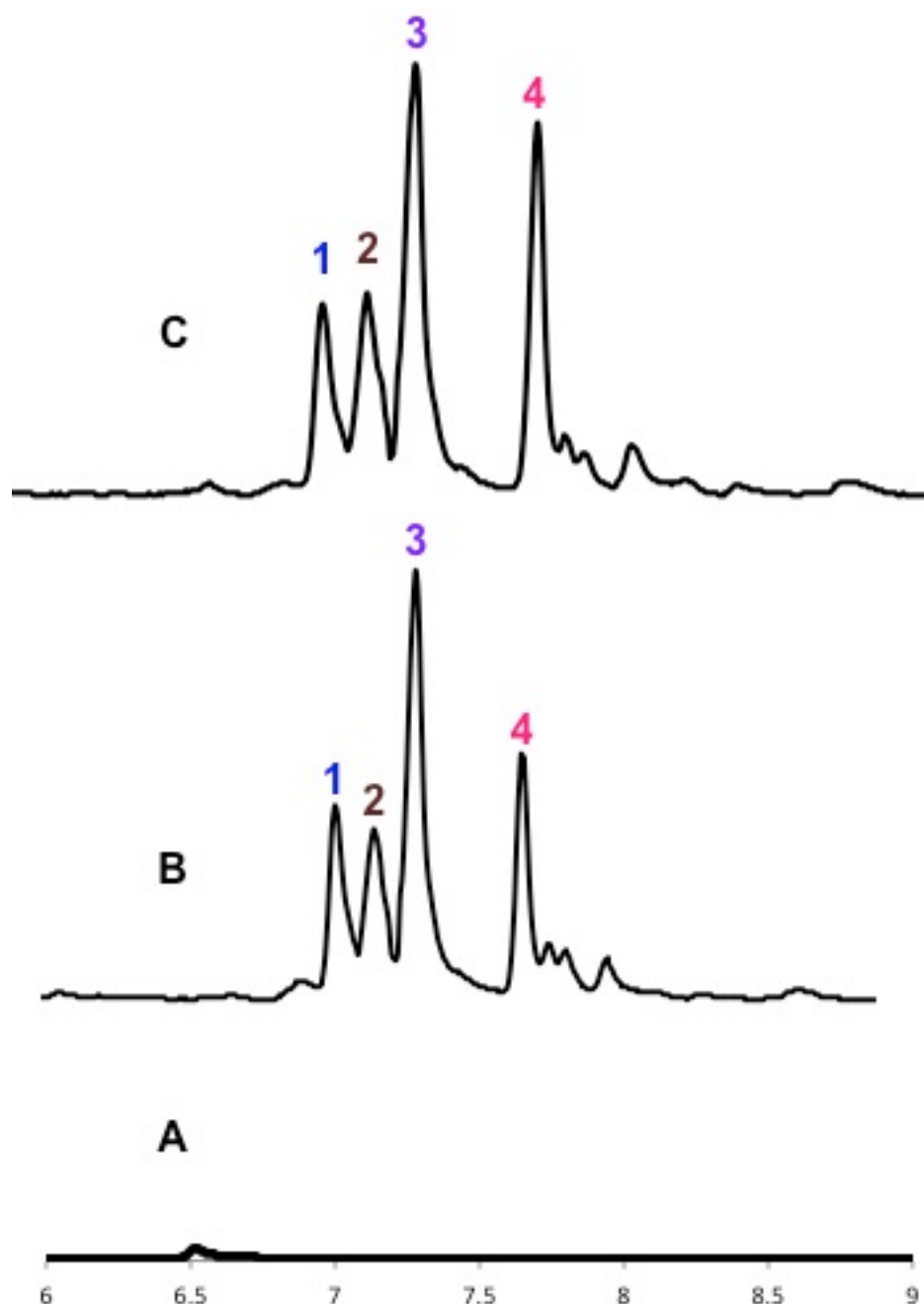


Figure 5.3.7: Total Ion Chromatograms (TICs) of **A.** Substrate Control, **B.** CrIDS Wild Type Assay, **C.** CrIDS_I38A Assay; 1-3: *Cis-trans*-iridodials and *cis-trans*-nepetalactols, 4: *Cis-trans*-nepetalactol.

Isoleucine at the 38th position (I38) is found to be positioning NADPH for CrIDS catalysis near the 1-CHO end of the substrate. One would expect any change in this residue to affect the activity profile of CrIDS, but surprisingly, the mutant I38A seemed to cause but a little change in the activity profile. This difference can be attributed to the incorporation of a smaller residue and a minute increase in polarity.

5.3.7 Residues involved in the enhanced formation of nepetalactol

Cis-trans-nepetalactol / (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol is formed by the cyclization of 10-oxogeraniol by the enzyme Iridoid synthase (CrIDS). The downstream enzyme to complete the secologanin biosynthetic pathway, and in turn the strictosidine and vinblastine/ vincristine pathways further utilizes this nepetalactol. It, however, forms only a portion of the product profile of CrIDS. If the formation of this metabolite could be enhanced, it would greatly aid in affecting the flux of the pathway. Some residues were identified in CrIDS, mutations of which led to an enhancement in the ratio of this metabolite, which would in turn, would enhance the further reactions of the pathway.

5.3.7.1 Lysine 144 (K144L)

The crystal structure of Progesterone 5- β -reductase (P5 β R-*Dl*) from *D. lanata* exemplifies that the protein belongs to short-chain dehydrogenases/ reductases, which can be said in the case of Iridoid synthase also. In the case of P5 β R-*Dl*, the residue K147 is present in its active site and is the part of novel sequence motifs. This is analogous to K144 in CrIDS, which is present in the hydrophobic active site pocket surrounding NADPH, near the 10-CHO end of the substrate. When this residue was mutated to the linear hydrophobic leucine (K144L) residue, an increase in hydrophobicity was caused and there was a 25 % increase *cis-trans*-nepetalactol / (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol (Figure 5.3.8). This may be due to the removal of the -NH₂, which causes for a better interaction between the substrate and the now -CH₂ group in the enzyme at the 144th position.

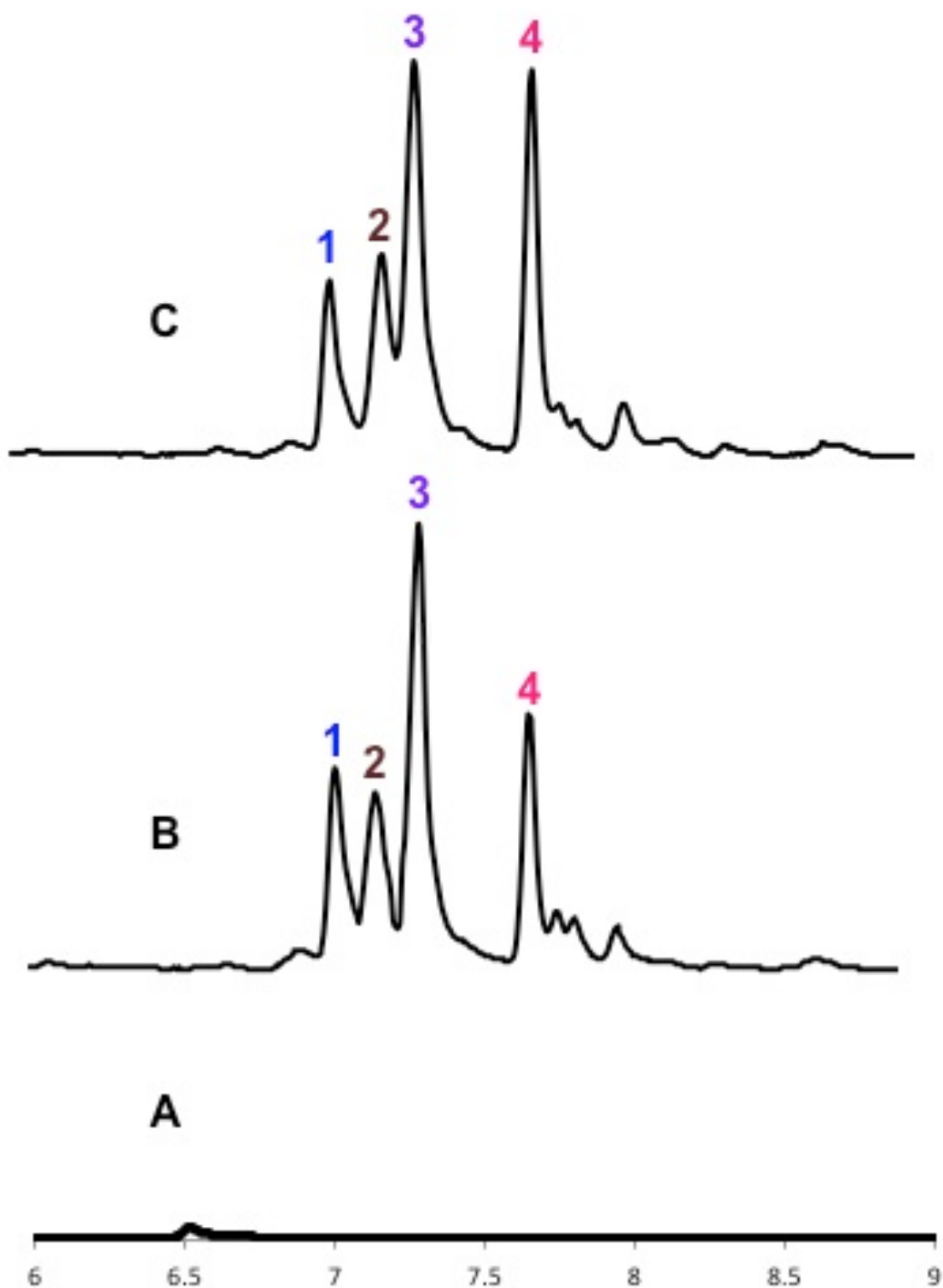


Figure 5.3.8: Total Ion Chromatograms (TICs) of **A.** Substrate Control, **B.** CrIDS Wild Type Assay, **C.** CrIDS_K144L Assay; 1-3: *Cis-trans*-iridodials and *cis-trans*-nepetalactols, 4: *Cis-trans*-nepetalactol.

5.3.7.2 Tryptophan 107 (W107F)

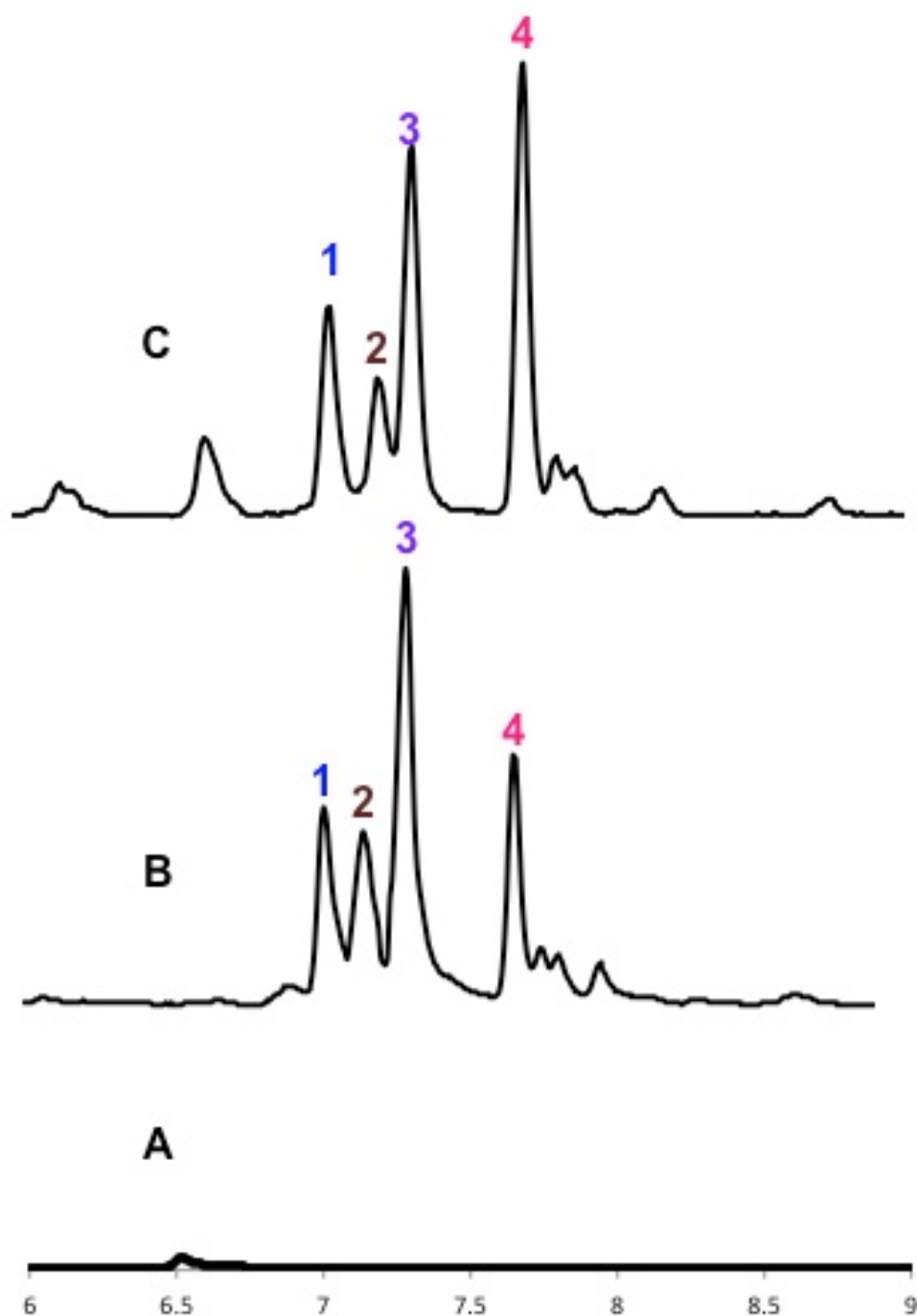


Figure 5.3.9: Total Ion Chromatograms (TICs) of **A.** Substrate Control, **B.** CrIDS Wild Type Assay, **C.** CrIDS_W107F Assay; 1-3: *Cis-trans*-iridodials and *cis-trans*-nepetalactols, 4: *Cis-trans*-nepetalactol.

In addition to the Y176 and K144, the nicotinamide ring of NADP^+ is found to be buried in a hydrophobic pocket with the residues W107, F175, M211 and F340¹³. These residues are highly conserved in the binding pocket. Naturally a change in the highly conserved W107 would cause a change in the activity of the protein. When this

residue was mutated to phenylalanine (W107F), there was a 51 % increase in *cis-trans*-nepetalactol / (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol, resulting in 32.8 % of this product. The smaller size of the phenylalanine introduced (in comparison with tryptophan) would also increase the cavity size.

5.3.7.3 Phenylalanine 175 (F175M)

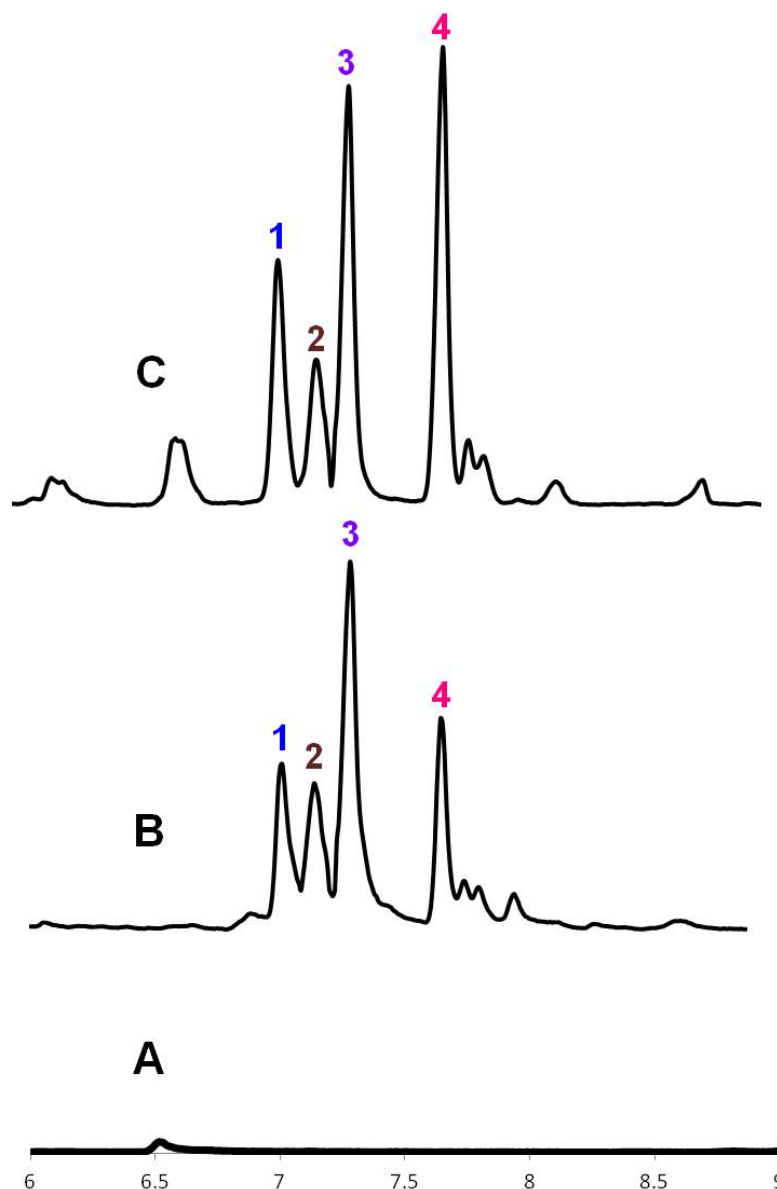


Figure 5.3.10: Total Ion Chromatograms (TICs) of **A.** Substrate Control, **B.** CrIDS Wild Type Assay, **C.** CrIDS_F175M Assay; 1-3: *Cis-trans*-iridodials and *cis-trans*-nepetalactols, 4: *Cis-trans*-nepetalactol.

As mentioned earlier, F175 is another residue present in the hydrophobic pocket surrounding NADPH. This residue, present near the 10-CHO end of the substrate, was mutated to methionine (F175M) and resulted in a 70 % increase in the

formation of *cis-trans*-nepetalactol / (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol. This mutation suggests that the aromatic ring of phenylalanine seems to have a great impact on the interaction between the substrates and protein and hence the change in product ratio. Also, the thiol group introduced because of methionine may be responsible for this increase.

5.3.7.4 Methionine 211 (M211A, M211F, M211S, M211K, M211H)

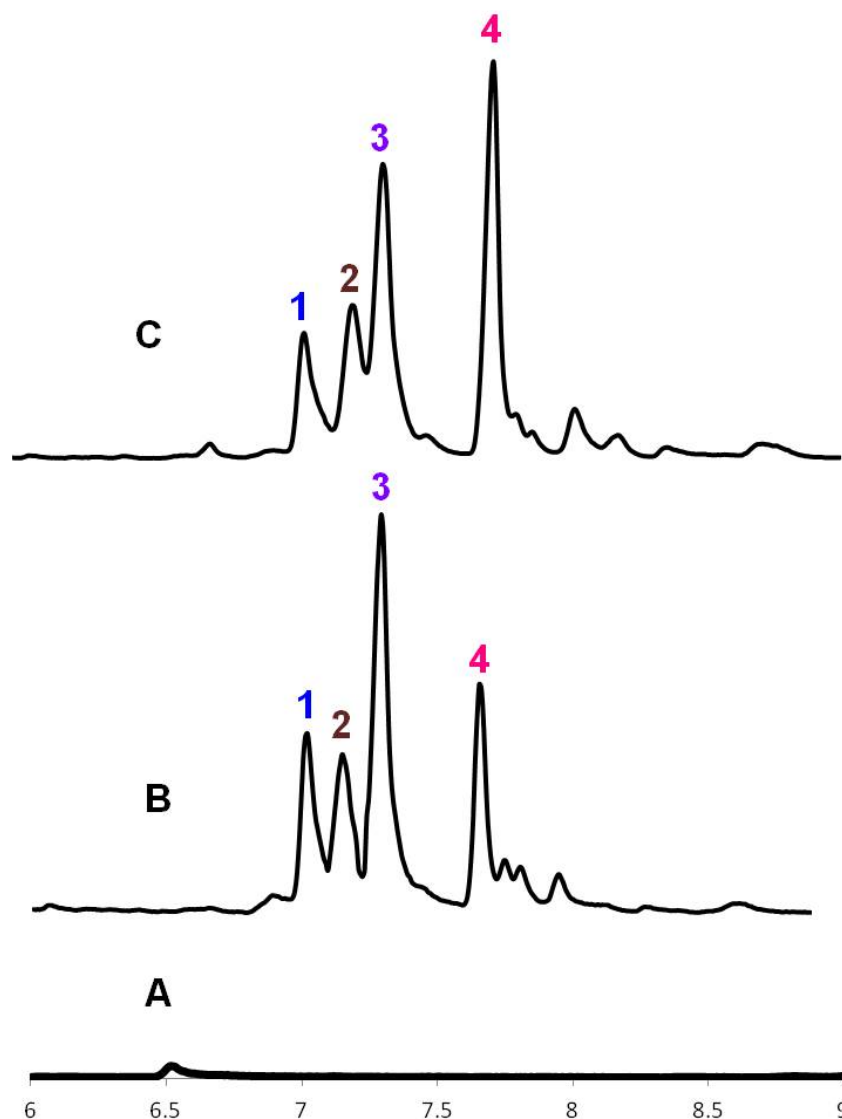


Figure 5.3.11: Total Ion Chromatograms (TICs) of **A.** Substrate Control, **B.** CrIDS Wild Type Assay, **C.** CrIDS_M211A Assay; 1-3: *Cis-trans*-iridodials and *cis-trans*-nepetalactols, 4: *Cis-trans*-nepetalactol.

Methionine 211 is another one of the residues which is highly conserved and present in the hydrophobic pocket surrounding NADPH. A change at this position from methionine to the smaller residue alanine (M211A) resulted in a 72 % increase in the formation of *cis-trans*-nepetalactol / (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol. Once

again, replacement by a smaller residue seems to have enhanced our desired product formation. As the change at this residue caused highest formation of our desired product, which would in turn increase the flux towards the subsequent products of the pathway, four more mutations were designed at this position.

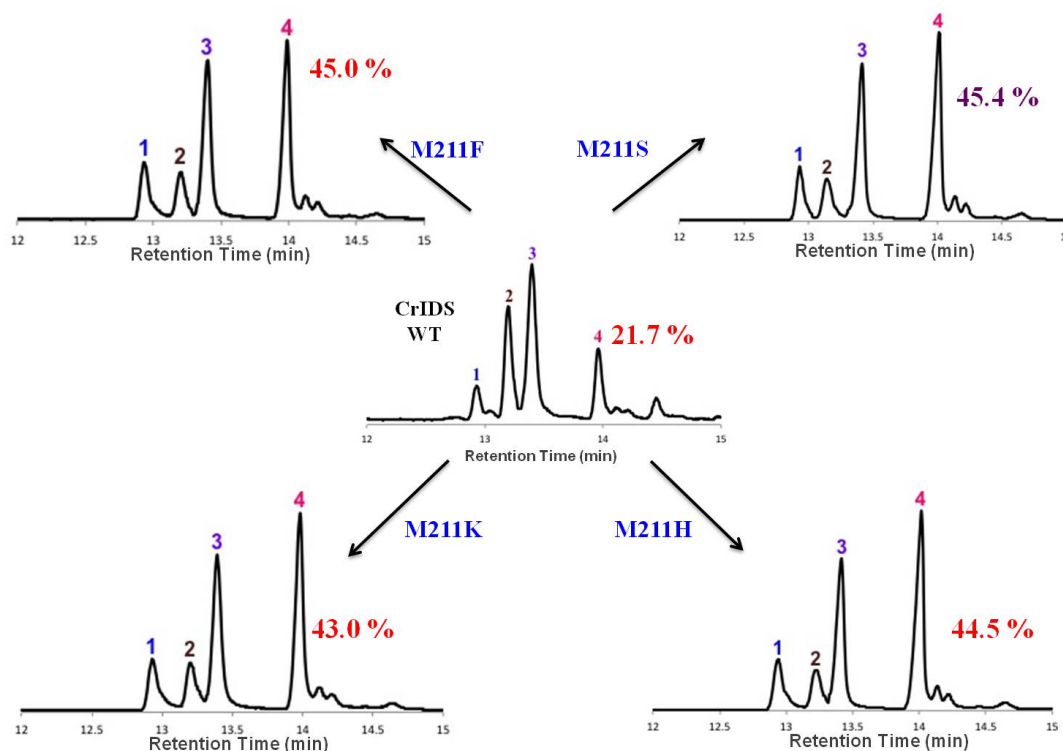


Figure 5.3.12 Total Ion Chromatograms (TICs) of CrIDS Wild Type, CrIDS-M211F, CrIDS-M211S, CrIDS-M211K and CrIDS-M211H.

The non-polar linear chained Methionine-211 was mutated to the hydrophobic phenylalanine (M211F), which has an aromatic ring and this yielded a 107.4 % increase in the formation of *cis-trans*-nepetalactol / (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol. In this change, the aromatic ring of phenylalanine may enhance the activity of the enzyme towards the formation of *cis-trans*-nepetalactol / (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol. In the case where the non-polar linear residue was replaced by the polar serine (M211S), this change was to the effect of 109.2 %. This residue was then mutated to the basic linear residue lysine (M211K) and histidine (M211H), which has a cyclopentane side chain. This yielded a 98 % and 105 % respective increase in the formation of our desired product. Serine is one of the amino acids normally forming a catalytic triad in most enzymes, mutation of which normally results in a drastic change in the activity of the enzyme. Although, serine was missing in the catalytic triad in this case, an introduction of serine in place of methionine, which is also an

essential residue present in the hydrophobic pocket enveloping NADP⁺, also caused a change in the activity of the enzyme. This fact emphasises on the importance of serine for the activity of the protein.

Table 5.3.1 Product profiles of assays of CrIDS wild type and mutants with 10-oxogeranial.

	Peak 1	Peak 2	Peak 3	Peak 4 : <i>Cis-trans-</i> nepetalactol
	RT: 7.061	RT: 7.237	RT: 7.355	RT: 7.740
CrIDS_WT	16.90 %	18.60 %	42.70 %	21.70 %
CrIDS_M1_I38A	17.50 %	18.50 %	42.20 %	21.90 %
CrIDS_M2_W107F	20.20 %	11.20 %	35.90 %	32.80 %
CrIDS_M3_K144L	15.20 %	18.00 %	38.80 %	28.00 %
CrIDS_M4_F175M	19.20 %	10.10 %	33.90 %	36.90 %
CrIDS_M5_Y176F	10.40 %	30.20 %	44.60 %	14.80 %
CrIDS_M6_M211A	13.40 %	17.40 %	31.80 %	37.40 %
CrIDS_M7_M211F	16.80 %	9.30 %	28.90 %	45.00 %
CrIDS_M8_M211S	17.50 %	9.00 %	28.10 %	45.40 %
CrIDS_M9_M211K	17.80 %	9.50 %	29.80 %	43.00 %
CrIDS_M10_M211H	17.60 %	8.80 %	29.10 %	44.50 %

5.3.8 Kinetic characterization of the CrIDS mutants

The mutant Y176F showed very less activity with the natural substrate, 10-oxogeranial. This observation finds support in the kinetic data of the enzyme, where the mutant is shown to possess a high K_m of 5.77 μM (substrate) and 5.26 μM (NADPH). In addition to this, their turnover number(s), K_{cat} , was very low, i.e., 0.3 sec^{-1} (NADPH) and 0.61 sec^{-1} (substrate), in comparison with that of the wild type (9.70 sec^{-1} for NADPH and 10 sec^{-1} for substrate). The K_m for all other mutants, especially the ones at Methionine-211, was lower than the wild type (Table 5.3.2).

Table 5.3.2 Steady state kinetics values of CrIDS mutants.

		K_m	V_{max}	K_{cat}
		(μM)	($\mu\text{M}/\text{sec}$)	(sec^{-1})
CrIDS	NADPH	4.33	0.032	9.70
Wild type	10-oxogeranial	4.49	0.033	10.00

CrIDS_M4	NADPH	2.38	0.005	1.52
F175M	10-oxogeranial	2.01	0.025	7.58
CrIDS_M5	NADPH	5.26	0.001	0.30
Y176F	10-oxogeranial	5.77	0.002	0.61
CrIDS_M6	NADPH	2.91	0.003	0.91
M211A	10-oxogeranial	3.12	0.007	2.12
CrIDS_M7	NADPH	1.53	0.007	2.12
M211F	10-oxogeranial	1.16	0.009	2.73
CrIDS_M8	NADPH	2.63	0.007	2.12
M211S	10-oxogeranial	2.74	0.009	2.73
CrIDS_M9	NADPH	0.77	0.011	3.33
M211K	10-oxogeranial	0.83	0.111	3.64
CrIDS_M10	NADPH	6.93	0.088	6.67
M211H	10-oxogeranial	6.82	0.031	9.39

5.4 Discussion

CrIDS belongs to a family of progesterone 5 β -reductase (P5 β R) family, which has been shown by way of transcriptome mining¹³. Although it might be confusing as it belongs to short chain dehydrogenases / reductases family, sequence alignments and structural studies of various P5 β Rs suggests its diversification from SDRs. This P5 β R family contains six other progesterone 5 β -reductase genes from *C. roseus*. Due to this similarity of CrIDS with P5 β R, homology model of CrIDS was built using the co-ordinates of Progesterone 5 β -Reductase from *Digitalis lanata* to understand the catalytic mechanism.

CrIDS, in the presence of NADPH, carries out the cyclization of 10-oxogeranial to an equilibrium mixture of *cis-trans*-nepetalactol and iridodials^{8,14}. In order to understand the structural determinants responsible for the formation of this mixture of products, we have manipulated the CrIDS active site pocket to cause a considerable change in this product ratio profile such that it leans more towards the formation of *cis-trans*-nepetalactol / (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol using site-directed mutagenesis of certain identified residues.

A close observance into the SDR-specific catalytic triad residues¹⁵, Tyr/ Lys/ Ser, shows that the conserved motifs of this triad are missing in CrIDS. Although, in comparison with the crystal structure of P5 β R from *D. lanata*, it can be predicted that the residue Tyr-176 might serve the purpose of the tyrosine in the triad. This Tyr-176 is found to interact with the nicotinamide ring in both 17 β -hydroxysteroid dehydrogenase type 1¹⁶ and P5 β R. Due to this interaction, we can say is the reason that the mutation at this position (Y176F) causes CrIDS to lose its activity to such an extent.

Characteristically in SDRs there is a lysine residue, which is considered to be catalytically important in order to provide binding affinity for the co-factor^{10,17}. In P5 β R, the corresponding residue was identified as K147 and in the case of CrIDS as K144. When this residue was mutated to leucine (K144L) there was a substantial increase in the formation of *cis-trans*-nepetalactol / (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol, which is quite surprising as previous reports suggested a loss in the activity of the protein when this lysine residue was mutated^{18,19}. This variance cannot be explained without further detailed studies from various other related proteins.

Mutations at F175 and M211 further enabled us to obtain *cis-trans*-nepetalactol / (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol in higher amounts than the wild type. In addition to the Y176 and K144 residues, these 2 residues (F175 and M211) have also previously been shown to affect the activity of the protein, upon modifications, as has been observed in the case of CrIDS as well^{10,19}.

The pharmaceutical significance of CrIDS poses its structural and mutational studies as a broad topic of interest. Our mutagenesis study has led us to the identification of essential residues governing the substrate specificity of CrIDS, which can be further manipulated for subsequent modification of the downstream metabolites of CrIDS.

5.5 Conclusion

The main aim of carrying out directed mutagenesis of CrIDS was to identify the amino acids controlling the substrate specificity of the enzyme. Through these studies, it could be observed that Y176 is most essential for the activity of the enzyme, which distorted the activity of the enzyme greatly upon mutation. The residues F175 and M211 were responsible for the substrate specificity and product profile, selective mutations of which led to higher formation of *cis-trans*-nepetalactol. In the case of CrIDS, it was observed that an increase in polarity near the active site pocket favored the formation of *cis-trans*-nepetalactol / (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol more than normally. The lactol ring formation of iridoids/ nepetalactol is a very essential step in secologanin biosynthesis as it forms the backbone of secologanin-derived alkaloids in *Catharanthus roseus*. Hence, these mutants would be of great value for the metabolic engineering of the vindoline pathway in *C. roseus*.

Considering the therapeutic significance of iridoids and their metabolites, these studies can also be effectively put to use in the medical field for the selective production of desirable molecules. Multiple site mutations can also be further designed for enhancing the activity of CrIDS.

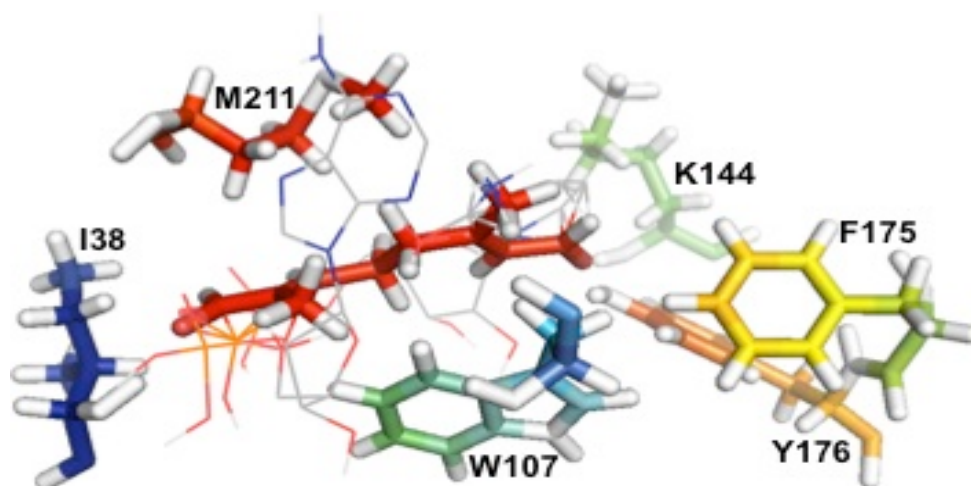


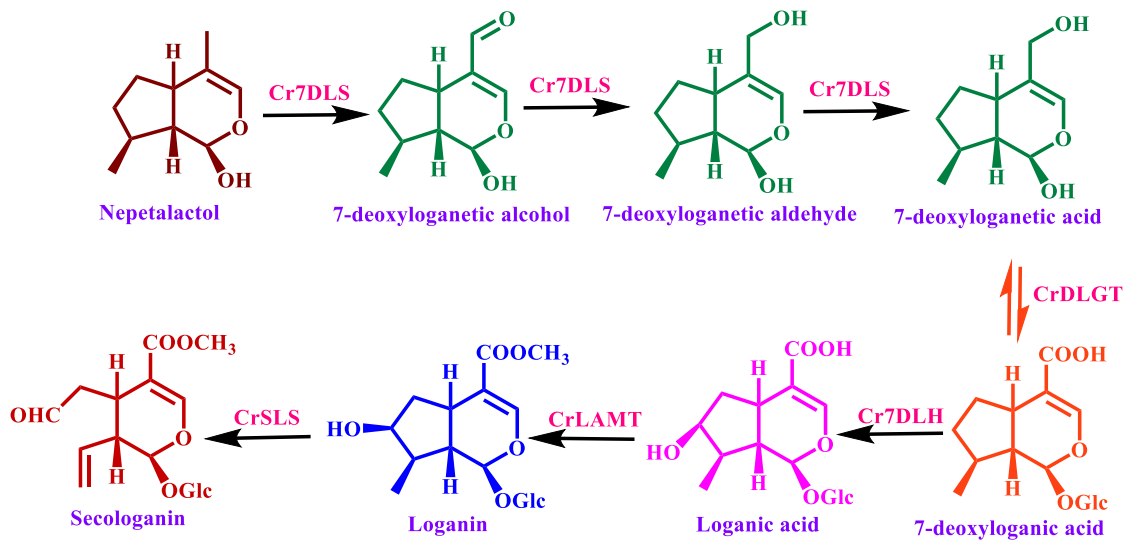
Figure 5.5.1: Amino acids of CrIDS chosen for mutation, positioned around the substrate, 10-oxogeranial and NADP⁺.

5.6 References

1. Dinda, B., Debnath, S. & Harigaya, Y. Naturally occurring secoiridoids and bioactivity of naturally occurring iridoids and secoiridoids. A review, part 2. *Chem. Pharm. Bull.* **55**, 689-728 (2007).
2. Dinda, B., Debnath, S. & Harigaya, Y. Naturally occurring iridoids. A review, part 1. *Chem. Pharm. Bull.* **55**, 159-222 (2007).
3. Nishida, R. Chemical ecology of insect-plant interactions: ecological significance of plant secondary metabolites. *Biosci. Biotechnol. Biochem.* **78**, 1-13 (2014).
4. Nishida, R. & Fukami, H. Host plant iridoid-based chemical defense of an aphid, *Acyrtosiphon nipponicus*, against ladybird beetles. *J. Chem. Ecol.* **15**, 1837-1845 (1989).
5. Gautam, R., Jachak, S.M. & Saklani, A. Anti-inflammatory effect of *Ajuga bracteosa* Wall Ex Benth. mediated through cyclooxygenase (COX) inhibition. *J. Ethnopharmacol.* **133**, 928-930 (2011).
6. Viljoen, A., Mncwangi, N. & Vermaak, I. Anti-Inflammatory Iridoids of Botanical Origin. *Curr. Med. Chem.* **19**, 2104-2127 (2012).
7. Lin, L. C., Wang, Y. H., Hou, Y. C., Chang, S., Liou, K. T., Chou, Y. C., Wang, W. Y., & Shen, Y. C. The inhibitory effect of phenylpropanoid glycosides and iridoid glucosides on free radical production and beta 2 integrin expression in human leucocytes. *J. Pharm. Pharmacol.* **58**, 129-135 (2006).
8. Krithika, R., Srivastava, P. L., Rani, B., Kolet, S. P., Chopade, M., Soniya, M., & Thulasiram, H. V. Characterization of 10-Hydroxygeraniol Dehydrogenase from *Catharanthus roseus* Reveals Cascaded Enzymatic Activity in Iridoid Biosynthesis. *Sci. Rep.* **5**, 8258-8264 (2015).
9. Treimer, J.F. & Zenk, M.H. Purification and properties of strictosidine synthase, the key enzyme in indole alkaloid formation. *Eur. J. Biochem.* **101**, 225-233 (1979).
10. Thorn, A., Egerer-Sieber, C., Jager, C. M., Herl, V., Muller-Uri, F., Kreis, W., & Muller, Y. A The crystal structure of progesterone 5 beta-reductase from *Digitalis lanata* defines a novel class of short chain dehydrogenases/reductases. *J. Biol. Chem.* **283**, 17260-17269 (2008).
11. Qin, L. L., Zhu, Y., Ding, Z. Q., Zhang, X. J., Ye, S., & Zhang, R. G. Structure of iridoid synthase in complex with NADP(+)/8-oxogeraniol reveals the structural basis of its substrate specificity. *J. Struct. Biol.* **194**, 224-230 (2016).
12. Hu, Y. M., Liu, W. D., Malwal, S. R., Zheng, Y. Y., Feng, X. X., Ko, T. P., Chen, C. C., Xu, Z. X., Liu, M. X., Han, X., Gao, J., Oldfield, E., & Guo, R. T. Structures of Iridoid Synthase from *Cantharanthus roseus* with Bound NAD(+), NADPH, or NAD(+)/10-Oxogeraniol: Reaction Mechanisms. *Angew. Chem. Int. Ed.* **54**, 15478-15482 (2015).
13. Munkert, J., Pollier, J., Miettinen, K., Van Moerkercke, A., Payne, R., Muller-Uri, F., Burlat, V., O'Connor, S. E., Memelink, J., Kreis, W., & Goossens, A. Iridoid Synthase Activity Is Common among the Plant Progesterone 5 β -Reductase Family. *Mol. Plant* **8**, 136-152 (2015).
14. Geu-Flores, F., Sherden, N. H., Courdavault, V., Burlat, V., Glenn, W. S., Wu, C., Nims, E., Cui, Y. H., & O'Connor, S. E. An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature* **492**, 138-142 (2012).
15. Filling, C., Berndt, K. D., Benach, J., Knapp, S., Prozorovski, T., Nordling, E., Ladenstein, R., Jornvall, H., & Oppermann, U. Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. *J. Biol. Chem.* **277**, 25677-25684 (2002).
16. Breton, R., Housset, D., Mazza, C. & Fontecilla Camps, J.C. The structure of a complex of human 17 beta-hydroxysteroid dehydrogenase with estradiol and NADP(+) identifies two principal targets for the design of inhibitors. *Structure* **4**, 905-915 (1996).
17. Parikh, S., Moynihan, D.P., Xiao, G.P. & Tonge, P.J. Roles of tyrosine 158 and lysine 165 in the catalytic mechanism of InhA, the enoyl-ACP reductase from *Mycobacterium tuberculosis*. *Biochem.* **38**, 13623-13634 (1999).
18. Nakajin, S., Takase, N., Ohno, S., Toyoshima, S. & Baker, M.E. Mutation of tyrosine-194 and lysine-198 in the catalytic site of pig 3 alpha/beta,20 beta-hydroxysteroid dehydrogenase. *Biochem. J.* **334**, 553-557 (1998).
19. Petersen, J., Lanig, H., Munkert, J., Bauer, P., Muller-Uri, F., & Kreis, W. Progesterone 5-reductases/iridoid synthases (PRISE): gatekeeper role of highly conserved phenylalanines in substrate preference and trapping is supported by molecular dynamics simulations. *J. Biomol. Struct. Dyn.* **34**, 1667-1680 (2016).

Chapter 6

Screening and cloning of genes involved in the biosynthesis of secologanin from iridoid



Chapter 6

Screening and cloning of genes involved in the biosynthesis of secologanin from iridoid

Secologanin is a major intermediate in the formation of the potent anti-cancer agents, vincristine and vinblastine in *Catharanthus roseus*. In order to complete the secologanin biosynthetic pathway, full-length unigenes showing matches with 7 deoxyloganin synthase (Cr7DLS), 7-deoxyloganetic acid glucosyltransferase (CrDLGT), 7-deoxyloganin hydroxylase (Cr7DLH), loganic acid methyltransferase (CrLAMT) and secologanin synthase (CrSLS) were identified, and their cloning have been carried out. CrDLGT has been cloned in pET28a and CrLAMT in pRSETB bacterial vectors, whereas, the cytochrome 450 hydroxylase enzymes Cr7DLS, Cr7DLH and CrSLS were cloned in the MCS2 of the yeast duet vector pESC-URA containing cytochrome P450 reductase (CrCPR) from *C. roseus* in its MCS1.

6.1 Introduction

Monoterpene indole alkaloids (MIAs) are a multifarious class of natural products with distinct chemical and biological properties¹⁻³ found in numerous species of plants belonging to the *Apocynaceae*, *Loganiaceae* and *Rubiaceae* families. To date, over 3000 MIAs are known with diverse structures and biological activities. These MIAs are inclusive of the anticancer alkaloids vincristine/ vinblastine from *Catharanthus roseus*^{4,5} and camptothecin from *Camptotheca acuminata*⁶/*Ophiorhiza pumila*⁷, the anti-malarial quinine from *Cinchona ledgeriana*/*C. succirubra*⁸, the antihypertensive drug ajmalicine from *Catharanthus roseus*/*Rauvolfia serpentina*⁹, the anti-arrhythmic drug, sarpagine from *Alstonia macrophylla*¹⁰ and rat poison/pesticide strychnine-n-oxide from *Strychnos nux-vomica*¹¹. These MIAs are synthesized by the condensation of tryptamine and the iridoid monoterpene, secologanin to give strictosidine, which is the precursor for all the afore-mentioned therapeutic compounds.

Over 200 MIAs are known from *C. roseus*, with catharanthine and vindoline present in greatest abundance. Chapter 3 dealt with the cloning and characterization of the genes from geranyl diphosphate synthase to iridoid synthase, the final product of which was *cis-trans* iridoids and *cis-trans* nepetalactols. This chapter deals with the screening and cloning of genes for the production of secologanin from iridoids, in an attempt to complete the secologanin biosynthetic pathway.

The gene 7-deoxyloganetic acid synthase (7DLS)¹² or iridoid oxidase (IO)¹³ converts nepetalactol to 7-deoxyloganetic acid via a three-step oxidation mechanism where nepetalactol is converted to an alcohol, then an aldehyde, and finally to a carboxylic acid, 7-deoxyloganetic acid. This acid is converted to 7-deoxyloganic acid by a glucosylation reaction catalyzed by the 7-deoxyloganetic acid glucosyltransferase (7-DLGT)^{13,14} enzyme, using UDP-glucose as the sugar donor. 7-deoxyloganic acid hydroxylase (7DLH)^{13,15} catalyzes the conversion of 7-deoxyloganic acid into loganic acid by a hydroxylation reaction. Loganic acid is converted to loganin by the oanic acid methyltransferase (LAMT)^{16,17} enzyme. The cytochrome P450 enzyme, secologanin synthase¹⁸, catalyzes the final step in the secologanin biosynthesis, which involves the conversion of loganin to secologanin.

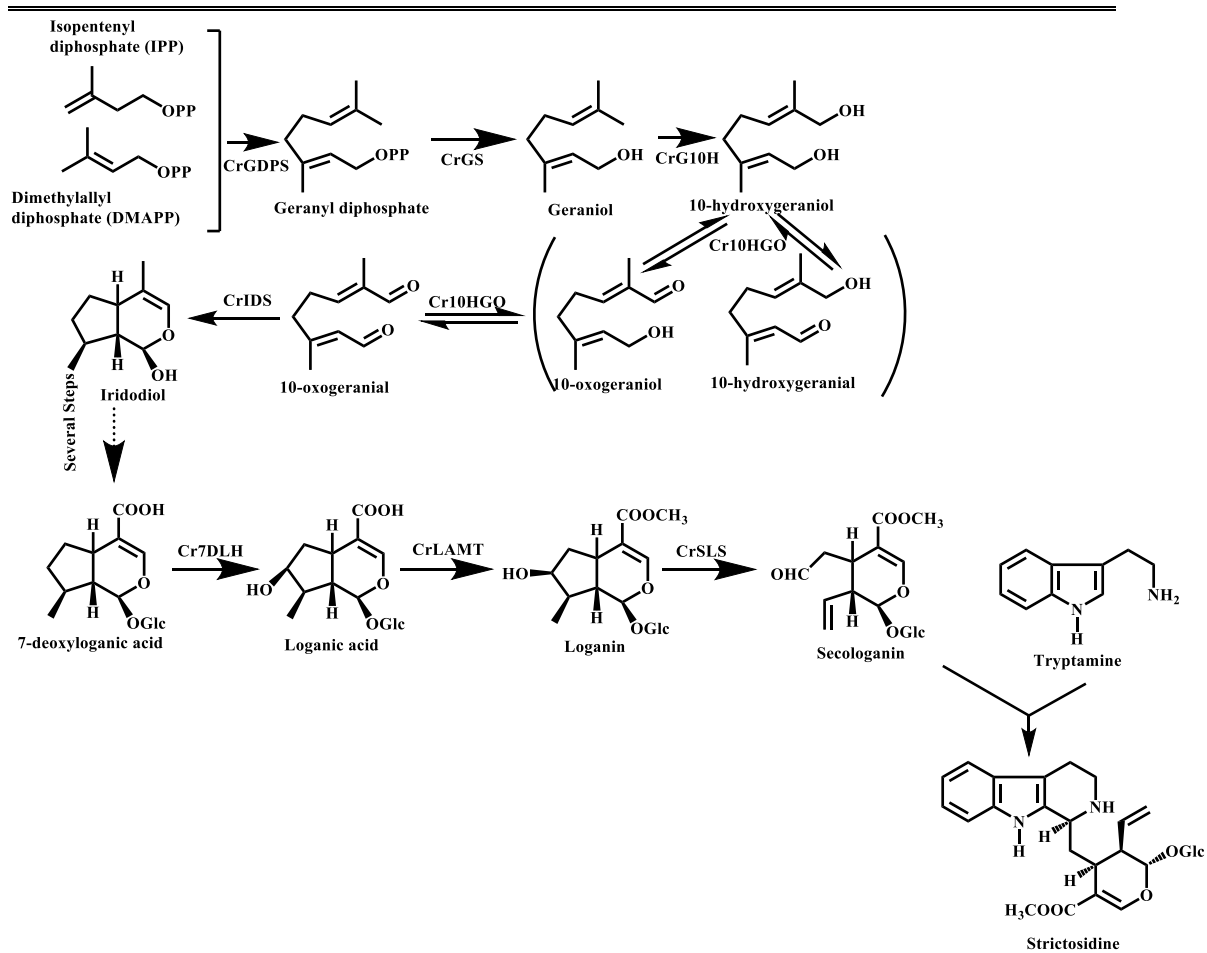


Figure 6.1.1: Biosynthesis of strictosidine via the key intermediate secologanin.

6.2 Materials and Methods

6.2.1 Materials used in the study

6.2.1.1 Strains and plasmids used in the study

Escherichia coli DH5 α (Invitrogen)

Escherichia coli BL21DE3 (Invitrogen)

Escherichia coli Rosetta2DE3 (Novagen)

Saccharomyces cerevisiae Invsc1 (Invitrogen)

pRSET expression vector (Invitrogen)

pET28 expression vector (Novagen)

pESC-URA Yeast Duet Expression Vector (Agilent)

Restriction enzymes (New England Biolabs)

6.2.1.2 Kits and reagent used in the study

SuperScript® III Reverse Transcriptase (ThermoFisher Scientific)

JumpStart™ Taq DNA Polymerase (Sigma-Aldrich)

AccuPrime™ Pfx DNA Polymerase (ThermoFisher Scientific)

GenElute™ PCR Clean-Up Kit (Sigma-Aldrich)

GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich)

GelRed™ (Biotium)

6.2.1.3 Buffer compositions

6.2.1.3.1 Cr7DLS, Cr7DLH and CrSLS

Lysis buffer – TES buffer

50 mM Tris-HCl, 1 mM EDTA, pH 7.4, containing 600 mM Sorbitol, 5 mM DTT, 0.25 mM PMSF.

Wash buffer – TEK buffer

50 mM Tris-HCl, 100 mM KCl, 1 mM EDTA, pH 7.4.

Suspension buffer – TEG buffer

50 mM Tris-HCl, 1 mM EDTA, 30 % glycerol, pH 7.5.

6.2.1.3.2 CrDLGT**Lysis buffer**

50 mM Tris-HCl, 300 mM NaCl, pH 8.0, 10 % glycerol, containing 1 mg/mL lysozyme, 1 mM PMSF and 0.5 % CHAPS.

Wash buffer

50 mM Tris-HCl, 300 mM NaCl, 30 mM Imidazole, pH 8.0, 10 % glycerol.

Elution buffer

50 mM Tris-HCl, 300 mM NaCl, 250 mM Imidazole, pH 8.0, 10 % glycerol.

Desalting buffer

100 mM HEPES-KOH, 10 % glycerol, pH 7.0.

6.2.1.3.3 CrLAMT**Lysis buffer**

50 mM NaH₂PO₄, 300 mM NaCl, pH 7.4, 10 % glycerol, containing 1 mg/mL lysozyme, 1 mM PMSF and 0.5 % CHAPS.

Wash buffer

50 mM NaH₂PO₄, 300 mM NaCl, 30 mM Imidazole, pH 7.4, 10 % glycerol.

Elution buffer

50 mM NaH₂PO₄, 300 mM NaCl, 250 mM Imidazole, pH 7.4, 10 % glycerol.

Desalting buffer

100 mM HEPES-KOH, 10 % glycerol, pH 7.0.

6.2.2 RNA isolation and cDNA preparation

As discussed in Chapter 2, total RNA was isolated from the leaf, stem and root tissues of white and pink-flowering plants. 5 µg of total RNA was used for first strand cDNA synthesis using SuperScript® III Reverse Transcription Kit (ThermoFisher Scientific). The cDNA thus constructed, was stored at -20 °C till further use.

6.2.3 Sequence analysis and ORF construction

After identification of the genes involved in secologanin biosynthesis, sequences were analyzed by NCBI GenBank database and overlapped with the EST fragments to generate full-length sequences. ORFs were selected using online ORF finder software for all the genes for their cloning and functional characterization. Full length ORF primers were designed, both blunt and those containing RE sites at both the ends, for all the genes for cloning them in expression vector(s).

Table 6.2.1 Primer sequence for isolation of full length ORF of the genes involved in the biosynthesis of Secologanin from Iridoids

Primer name	5'-3' Primer sequence
CrCPR_F	ATGGATTCTAGCTCGGAGAAGTTGT
CrCPR_R	TCACCAGACATCTCGGAGATACCTT
CrCPR_RE_F	GACGACGCGGCCGCATGGATTCTAGCTCGGAGAAG
CrCPR_RE_R	GACGACTTAATTAATCACCAGACATCTCGGAGATA
Cr7DLS_F	ATGGCGACCATCACTTTCGATTCACT
Cr7DLS_R	GATATGAACTCTCTTCTTAGGGATG
Cr7DLS_RE_F	GACGACGTCGACATGGCGACCATCACTTTC
Cr7DLS_RE_R	GACGACGGTACCGATATGAACTCTCTTCTTAG
CrDLGT_F	ATGGGTTCTCAAGAAACAAAT
CrDLGT_R	TCAAATAATCAGTGATTTTATGTA
CrDLGT_RE_F	GACGACGAGCTCATGGGTTCTC
CrDLGT_RE_R	GACGACCTCGAGTCAAATAATCAG
Cr7DLH_F	ATGGAATTGAACTTCAAATCAATT
Cr7DLH_R	TTAGAGTTTGTGCAGAATCAAATGA
Cr7DLH_RE_F	GACGACGTCGACATGGAATTGAACTTC
Cr7DLH_RE_R	GACGACGGTACCTTAGAGTTTGTGCAG
CrLAMT_F	ATGGTTGCCACAATTGATTCCATT

CrLAMT_R	TTAATTTCCCTTGCGTTTCAAGAC
CrLAMT_RE_F	GATATCATGGTTGCCACAATTGATTCCATTG
CrLAMT_RE_R	CTCGAGTTAATTTCCCTTGCGTTTCAAGACA
CrSLS_F	ATGGAGATGGATATGGATAACCATT
CrSLS_R	TTAAAAATTCTGTCTCTCAAGCTT
CrSLS_RE_F	GACGACGTCGACGAATGGAGATGGATATG
CrSLS_RE_R	GACGACGGTACCTTAAAAATTCTGTCT

Bold sequences are the Restriction site sequences used in cloning strategy

6.2.4 Full-length gene isolation and cloning into expression vector

6.2.4.1 Isolation and cloning of ORF of *CrCPR* in MCS1 of pESC-URA yeast duet vector

Full-length primers for Cytochrome P450 reductase ORF were designed using the transcript of *CPR* as a template. Synthesized cDNA was used for PCR reaction using Accuprime Pfx Supermix (Invitrogen) using the PCR program: 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 65 °C for 30 sec, 68 °C for 2 min 20 sec followed by final extension at 68 °C for 10 min. PCR product of 2145 bp was cloned in the MCS1 of pESC-URA yeast expression vector by digesting both with NotI and PacI restriction sites, using T₄ DNA ligase (Invitrogen) by incubating overnight at 14 °C. Ligation mixture was transformed into DH5 α chemically competent cells, plated on LA containing 100 μ g/mL of ampicillin and incubated overnight at 37 °C. Positive clones were screened by colony PCR with Gal10 forward and reverse primers. Plasmids were isolated from 4 positive clones of colony PCR and sequenced with Gal10 forward and reverse primers to check for the sequence in the correct vector frame.

6.2.4.2 Isolation and cloning of ORF of *Cr7DLS* in the MCS2 of pESC-URA expression vector

Full-length primers for 7-deoxyloganin synthase ORF were designed using the transcript of *Cr7DLS* as a template. Synthesized cDNA was used for PCR reaction using Accuprime Pfx Supermix (Invitrogen) using the PCR program: 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 62 °C for 30 sec, 68 °C for 1 min 50 sec followed by final extension at 68 °C for 10 min. PCR product of 1548 bp was cloned in the MCS2 of pESC-URA yeast expression vector by digesting both with Sall and

KpnI restriction sites, using T₄ DNA ligase (Invitrogen) by incubating overnight at 14 °C. Ligation mixture was transformed in DH5 α chemically competent cells, plated on LA containing 100 μ g/mL of ampicillin and incubated overnight at 37 °C. Positive clones were screened by colony PCR with Gal1 forward and reverse primers. Plasmids were isolated from 4 positive clones of colony PCR and sequenced with Gal1 forward and reverse primers to check for the sequence in the correct vector frame.

6.2.4.3 Isolation and cloning of ORF of *CrDLGT* in pET28a expression vector

Full-length primers for 7-deoxyloganic acid glucosyltransferase ORF were designed using the transcript of *CrDLGT* as a template. Synthesized cDNA was used for PCR reaction using Accuprime Pfx Supermix (Invitrogen) using the PCR program: 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 57 °C for 30 sec, 68 °C for 1 min 35 sec followed by final extension at 68 °C for 10 min. PCR product of 1449 bp was cloned in pET28a expression vector by digesting both with SacI and XhoI restriction sites, using T₄ DNA ligase (Invitrogen) by incubating overnight at 14 °C. Ligation mixture was transformed in DH5 α chemically competent cells, plated on LA containing 50 μ g/mL of kanamycin and incubated overnight at 37 °C. Positive clones were screened by colony PCR with T7 promoter and T7 reverse primers. Plasmids were isolated from 4 positive clones of colony PCR and sequenced with T7 promoter and reverse primer to check for the sequence in the correct vector frame.

6.2.4.4 Isolation and cloning of ORF of *Cr7DLH* in MCS2 of pESC-URA yeast expression vector

Full-length primers for 7-deoxyloganin hydroxylase ORF were designed using the transcript of *Cr7DLH* as a template. Synthesized cDNA was used for PCR reaction using Accuprime Pfx Supermix (Invitrogen) using the PCR program: 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 60 °C for 30 sec, 68 °C for 1 min 50 sec followed by final extension at 68 °C for 10 min. PCR product of 1566 bp was cloned in the MCS2 of pESC-URA expression vector by digesting both with Sall and KpnI restriction sites, using T₄ DNA ligase (Invitrogen) by incubating overnight at 14 °C. Ligation mixture was transformed in DH5 α chemically competent cells, plated on LA containing 100 μ g/mL of ampicillin and incubated overnight at 37 °C. Positive clones were screened by colony PCR with Gal1 forward and reverse primers.

Plasmids were isolated from 4 positive clones of colony PCR and sequenced with Gal1 forward and reverse primers to check for the sequence in the correct vector frame.

6.2.4.5 Isolation and cloning of ORF of *CrLAMT* in pRSET B expression vector

Full-length primers for loganic acid methyltransferase ORF were designed using the transcript of *CrLAMT* as a template. Synthesized cDNA was used for PCR reaction using Accuprime Pfx Supermix (Invitrogen) using the PCR program: 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 60 °C for 30 sec, 68 °C for 1 min 20 sec followed by final extension at 68 °C for 10 min. PCR product of 1116 bp was cloned in Zero Blunt vector and sub-cloned in pRSET B expression vector by digesting both with EcorV and XhoI restriction sites, using T₄ DNA ligase (Invitrogen) by incubating overnight at 14 °C. The ligation mixture was transformed in TOP10 chemically competent cells, plated on LA containing 100 µg/mL of ampicillin and incubated overnight at 37 °C. Positive clones were screened by colony PCR with T7 promoter and T7 reverse primers. Plasmids were isolated from 6 positive clones of and sequenced with T7 promoter and T7 terminator primers to check for the sequence in the correct vector frame.

6.2.4.6 Isolation and cloning of ORF of *CrSLS* in the MCS2 of pESC-URA yeast expression vector

Full-length primers for Loganic acid methyltransferase ORF were designed using the transcript of *CrSLS* as a template. Synthesized cDNA was used for PCR reaction using Accuprime Pfx Supermix (Invitrogen) using the PCR program: 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 58 °C for 30 sec, 68 °C for 1 min 40 sec followed by final extension at 68 °C for 10 min. PCR product of 1575 bp was cloned in Zero Blunt vector and sub-cloned in pRSET B expression vector by digesting both with SalI and KpnI restriction sites, using T₄ DNA ligase (Invitrogen) by incubating overnight at 14 °C. The ligation mixture was transformed in DH5α chemically competent cells, plated on LA containing 100 µg/mL of ampicillin and incubated overnight at 37 °C. Positive clones were screened by colony PCR with Gal1 forward and reverse primers. Plasmids were isolated from 4 positive clones of and sequenced with Gal1 forward and reverse primers to check for the sequence in the correct vector frame.

6.2.5 Expression and protein purification

6.2.5.1 Yeast expression of *Cr7DLS*

Expression of active protein was carried out in INVSc1 yeast competent cells. Cells were grown overnight in synthetic complete medium without uracil (SC-U), containing 2 % glucose at 30 °C, then transferred to induction medium (SC-U, containing 20 % galactose) and further incubated at 30 °C for 12 hours. The cells were centrifuged at $3000 \times g$ for 10 minutes at 4 °C. The cell pellet obtained was washed with TEK buffer (1 mL/g of cell pellet) and centrifuged. The cell pellet (1 g/5 mL) was re-suspended in TES buffer and cells were lysed using a bead-beater (with acid washed glass beads, 425–600 mm) for 6 cycles (pulse on 30 sec, pulse off 30 sec, manual rocking for 3×30 sec). The lysed cells were centrifuged at $1000 \times g$ for 5 min at 4 °C to remove the glass beads. Further, the supernatant was subjected to centrifugation at $10,000 \times g$ for 30 min at 4 °C. All the fractions were run on a 12 % SDS gel.

6.2.5.2 Heterologous expression and protein purification of *CrDLGT*

Expression was carried out in Rosetta2 DE3, cells were grown in Terrific Broth at 37 °C and induced with IPTG at a final concentration of 1 mM and incubated for 12 hours at 16 °C. After the induction, the culture was harvested by centrifugation at $4500 \times g$ for 20 minutes. The cell pellet (5 g/L) was re-suspended in 10 mL/g of cell pellet of lysis buffer (20 mM Tris, 300 mM NaCl, 10 % glycerol, 0.5 % CHAPS, pH 8.0) with 1 mg/mL concentration of lysozyme and 100 μ L/5 mL of protease inhibitor cocktail from Sigma and incubated on ice for 30 minutes. After the incubation period, the cells were sonicated with a pulse of 30 sec ON and 30 sec OFF for 10 cycles. The sample was then centrifuged at $10,000 \times g$ for 10 minutes at 4 °C. The crude lysate was purified using Ni-NTA column (2 mL resin / g cell pellet). The protein was finally eluted out in the elution buffer (20 mM Tris, 300 mM NaCl, 10 % glycerol, pH 8.0, 250 mM Imidazole) each. Purified protein fractions were checked on 12 % SDS gel.

6.2.5.3 Yeast Expression of *Cr7DLH*

Expression of active protein was carried out in INVSc1 yeast competent cells. Cells were grown overnight in synthetic complete medium without uracil (SC-U), containing 2 % glucose at 30 °C, then transferred to induction medium (SC-U,

containing 20 % galactose) and further incubated at 30 °C for 12 hours. The cells were centrifuged at $3000 \times g$ for 10 minutes at 4 °C. The cell pellet obtained was washed with TEK buffer and centrifuged. The cell pellet (1 g/5 mL) was re-suspended in TES buffer and cells were lysed using a bead-beater (with acid washed glass beads, 425–600 μ m) for 6 cycles (pulse on 30 sec, pulse off 30 sec, manual rocking for 3×30 sec). The lysed cells were centrifuged at $1000 \times g$ for 5 min at 4 °C to remove the glass beads. Further, the supernatant was subjected to centrifugation at $10,000 \times g$ for 30 min at 4 °C. All the fractions were run on a 12 % SDS gel.

6.2.5.4 Heterologous expression and protein purification of *CrLAMT*

Expression was carried out in C41 DE3, cells were grown in Luria Bertani Broth at 37 °C and induced with IPTG at a final concentration of 4 mM and incubated for 12 hours at 30 °C. After the induction, the culture was harvested by centrifugation at $4500 \times g$ for 20 minutes. The cell pellet (10 g/L) was re-suspended in 5 mL/g of cell pellet of Lysis buffer (50 mM NaH_2PO_4 , 300 mM NaCl, 0.5 % CHAPS, pH 7.4, 10 % v/v Glycerol) with 1 mg/mL concentration of lysozyme and 1 mM PMSF and incubated on ice for 30 minutes. After the incubation period, the cells were sonicated with a pulse of 30 sec ON and 30 sec OFF for 10 cycles. The sample was then centrifuged at $5,000 \times g$ for 10 minutes at 4 °C. The crude lysate was purified using Ni-NTA column (2 mL resin / g cell pellet). The protein was finally eluted out in the elution buffer (50 mM NaH_2PO_4 , 300 mM NaCl, pH 7.4, 250 mM Imidazole, 10 % v/v Glycerol) each. Purified protein fractions were checked on 12 % SDS gel.

6.2.5.5 Yeast expression of *CrSLS*

Expression of active protein was carried out in INVSc1 yeast competent cells. Cells were grown overnight in synthetic complete medium without uracil (SC-U), containing 2 % glucose at 30 °C, then transferred to induction medium (SC-U, containing 20 % galactose) and further incubated at 30 °C for 12 hours. The cells were centrifuged at $3000 \times g$ for 10 minutes at 4 °C. The cell pellet obtained was washed with TEK buffer (1 mL/g of cell pellet) and centrifuged. The cell pellet (1 g/5 mL) was re-suspended in TES buffer and cells were lysed using a bead-beater (with acid washed glass beads, 425–600 μ m) for 6 cycles (pulse on 30 sec, pulse off 30 sec, manual rocking for 3×30 sec). The lysed cells were centrifuged at $1000 \times g$ for 5 min at 4 °C to remove the glass beads. Further, the supernatant was subjected to

centrifugation at $10,000 \times g$ for 30 min at 4 °C. All the fractions were run on a 12 % SDS gel.

6.2.6 Phylogenetic analysis

The evolutionary history of the six genes from *Vinca* described in this chapter was inferred using the Neighbor-Joining method¹⁹. The evolutionary distances were computed using the Poisson correction method²⁰. The analysis involved 16 amino acid sequences. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA6²¹.

6.3 Results

6.3.1 Isolation, cloning and expression of Cytochrome P450 Reductase (*CrCPR*)

6.3.1.1 Isolation and cloning of Cytochrome P450 Reductase (*CrCPR*)

Analysis of the transcriptomic data revealed that one transcript, Locus_375, with an Open reading frame (ORF) of 2145 bp (Figure 6.3.1), encoding a polypeptide of 714 amino acids, displaying 100 % sequence identity with Full=NADPH--cytochrome P450 reductase from *Catharanthus roseus* (Genbank ID: Q05001.1²²) at the amino acid level, was identified as Cytochrome P450 reductase (*CrCPR*). Deduced amino acid sequence of Cr7DLS was found to have a calculated molecular weight of 78.96 kDa, comprising the NADPH cytochrome P450 reductase (CYPOR) domain (positions 310-714) and FAD binding domain (positions 310-531).

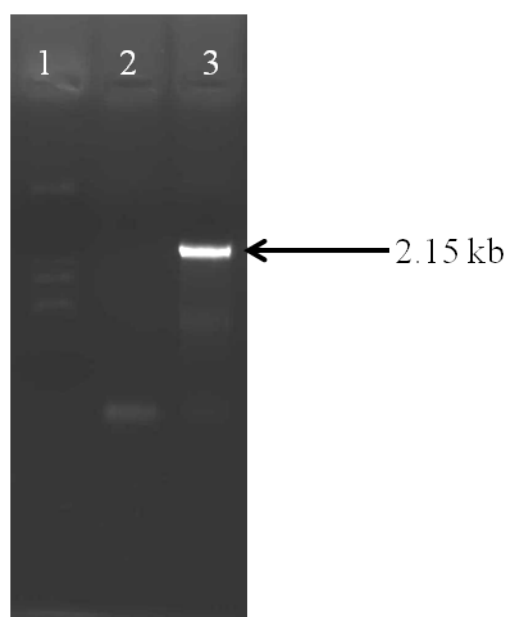


Figure 6.3.1: CrCPR full length ORF amplification, **Lane 1:** 1 kb DNA Ladder (Addendum Figure A1), **Lane 2:** Negative control, **Lane 3:** PCR product of CrCPR from cDNA

It functions in electron transfer to cytochrome P450 mono-oxygenases and that functional domains involved in binding of the cofactors FMN, FAD and NADPH are highly conserved between all kingdoms²².

6.3.2 Isolation, cloning and expression of 7-deoxyloganin synthase (*Cr7DLS*)

6.3.2.1 Isolation and cloning of 7-deoxyloganin synthase (*Cr7DLS*)

Analysis of the transcriptomic data revealed that one transcript, Locus_734, with an Open reading frame (ORF) of 1548 bp (Figure 6.3.2), encoding a polypeptide of 516 amino acids, displaying 88 % sequence identity with CYP1 from *Gentiana rigescens* (Genbank ID: AKJ26120.1²³) at the amino acid level, was identified as 7-deoxyloganin synthase (*Cr7DLS*). Deduced amino acid sequence of *Cr7DLS* was found to have a calculated molecular weight of 58.73 kDa, comprising the P450 superfamily domain (positions 39-505) and cytochrome P450 [Secondary metabolites biosynthesis, transport and catabolism, defense mechanisms] (positions 37-512).

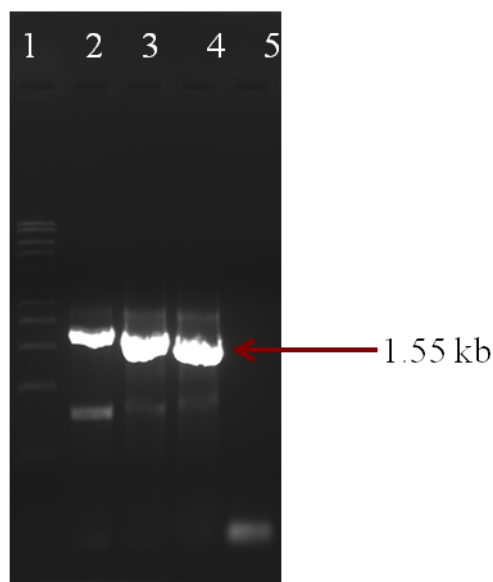


Figure 6.3.2: *Cr7DLS* full length ORF amplification, **Lane 1:** 1 kb DNA Ladder (Addendum Figure A1), **Lane 2:** PCR product of *Cr7DLS* from stem cDNA, **Lane 3:** PCR product of *Cr7DLS* from leaf cDNA, **Lane 4:** PCR product of *Cr7DLS* from root cDNA, **Lane 5:** Negative control

The ORF of *Cr7DLS* was cloned in the MCS2 of pESC-URA vector, containing the cytochrome P450 reductase (*CrCPR*) in the MCS1, so that the hydroxylase activity of *Cr7DLS* can be complemented with the activity of the reductase for expression of active protein in *S. cerevisiae* InvscI cells. Positive clones were screened by colony PCR, followed by sequencing with Gal1 forward and reverse primers. The sequence analysis revealed that the ORF of the transcript, Locus_734,

was successfully cloned in the MCS2 of the pESC-URA vector containing the cytochrome P450 reductase (CrCPR) in the MCS1.

6.3.2.2 Yeast expression

Expression was carried out in INVSc1. The recombinant protein of Cr7DLS in pESC-URA was expressed and processed, as discussed earlier. The protein fractions were loaded on a 12 % SDS-PAGE (Figure 6.3.3). A band at ~58 KDa confirmed expression of Cr7DLS. The microsomal fractions were flash-frozen and stored at -80 °C till used.

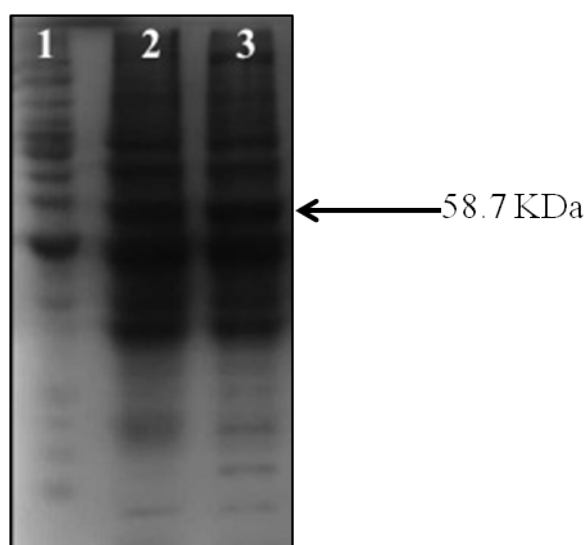


Figure 6.3.3: SDS-PAGE for Cr7DLS protein in pESC-URA+CrCPR, **Lane 1:** BenchMark™ Protein Ladder (Addendum Figure A2), **Lane 2:** Pellet fraction, **Lane 3:** Crude lysate (supernatant fraction).

6.3.3 Isolation, cloning and expression of 7-Deoxyloganetic Acid Glucosyltransferase (*CrDLGT*)

6.3.3.1 Isolation and cloning of 7-Deoxyloganetic Acid Glucosyltransferase (*CrDLGT*)

Analysis of the transcriptomic data revealed that one transcript, Locus_1690, with an Open reading frame (ORF) of 1449 bp (Figure 6.3.4), encoding a polypeptide of 482 amino acids, displaying 97 % sequence identity with UGT4 from *Panax ginseng* (Genbank ID: AIE12477.1²⁴) at the amino acid level, was identified as 7-deoxyloganetic acid glucosyltransferase (*CrDLGT*). Deduced amino acid sequence of CrDLGT was found to have a calculated molecular weight of 54.3 kDa, comprising a UDP-glucuronosyl/UDP-glucosyl transferase family protein (positions 12-480).

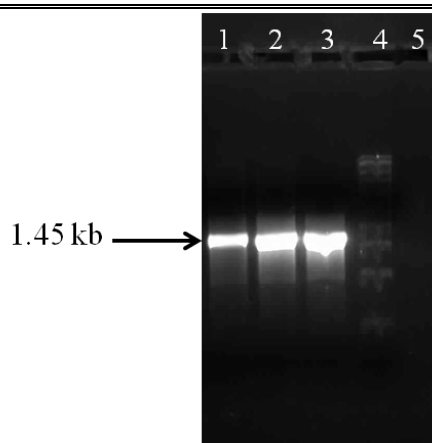


Figure 6.3.4: CrDLGT full length ORF amplification, **Lane 1:** PCR product of CrDLGT from root cDNA, **Lane 2:** PCR product of CrDLGT from stem cDNA, **Lane 3:** PCR product of CrDLGT from leaf cDNA, **Lane 4:** 1 kb DNA Ladder (Addendum Figure A1), **Lane 5:** Negative control

The ORF of *CrDLGT* was cloned in pET 28a vector frame with N terminal His₆ tag for affinity purification under the control of T7-RNA polymerase promoter for expression of active protein in *E.coli* Rosetta2 (DE3) cells. Positive clones were screened by colony PCR, followed by sequencing with T7 promoter and T7 reverse primer.

6.3.3.2 Bacterial expression and protein purification

Expression was carried out in Rosetta2 DE3 cells. The recombinant protein of CrDLGT in pET28a was expressed and purified as discussed earlier. The purified protein fractions gave a band at ~55 kDa on a 12 % SDS-PAGE (Figure 6.3.5). The purified fractions were desalted, flash-frozen and stored at -80 °C till used.

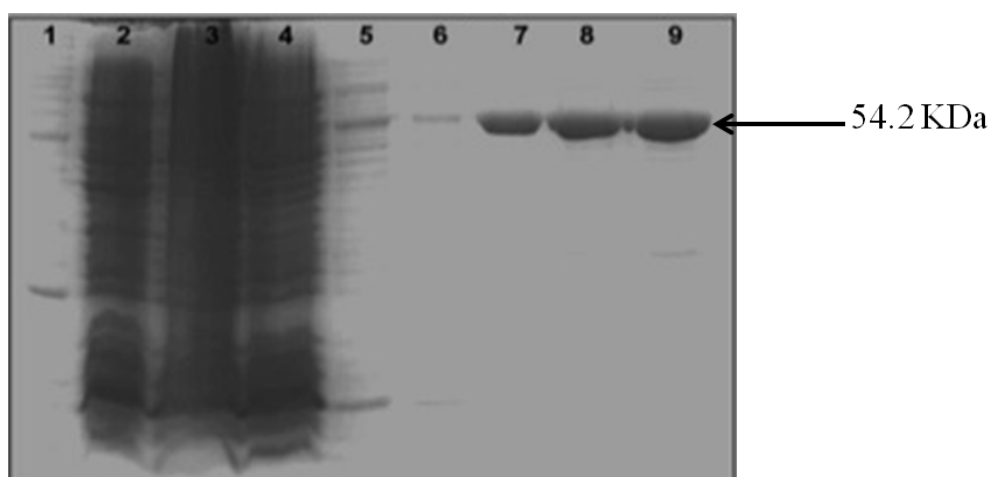


Figure 6.3.5: SDS-PAGE for CrDLGT protein purification in pET 28a, **Lane 1:** BenchMark™ Protein Ladder (Addendum Figure A2), **Lane 2:** Pellet fraction, **Lane 3:** Crude lysate (supernatant fraction), **Lane 4:** Unbound fraction, **Lanes 5-6:** Wash fractions, **Lanes 7-9:** Elution fractions.

6.3.4 Isolation, cloning and expression of 7-deoxyloganin hydroxylase (*Cr7DLH*)

6.3.4.1 Isolation and cloning of 7-deoxyloganin hydroxylase (*Cr7DLH*)

Analysis of the transcriptomic data revealed that one transcript, Locus_2445, with an open reading frame (ORF) of 1566 bp (Figure 6.3.6), encoding a polypeptide of 521 amino acids, displaying 100 % sequence identity with CYP72A224 7-deoxyloganic acid 7-hydroxylase from *Catharanthus roseus* (Genbank ID: AGX93062.1²⁵) was identified as 7-deoxyloganin hydroxylase (*Cr7DLH*). Deduced amino acid sequence of *Cr7DLH* was found to have a calculated molecular weight of 60 kDa, comprising a CYTOCHROME_P450 domain (positions 99-496) with the PROSITE ID: PS00086, containing the consensus sequences FGaGRRICPG.

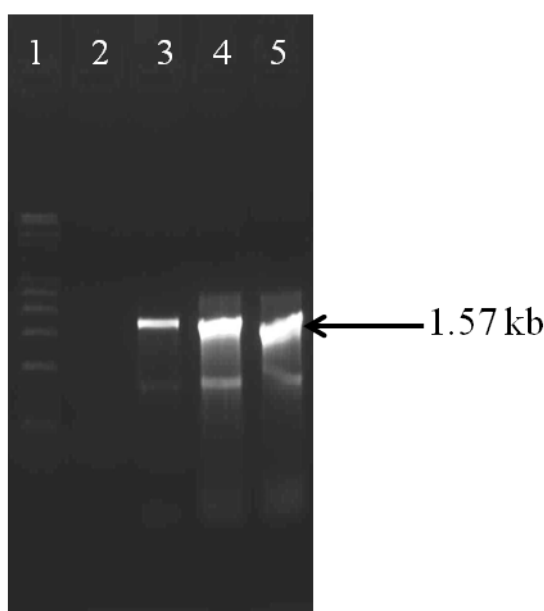


Figure 6.3.6: *Cr7DLH* full length ORF amplification, **Lane 1:** 1 kb DNA Ladder (Addendum Figure A1), **Lane 2:** Negative control, **Lane 3:** PCR product of *Cr7DLH* from leaf cDNA, **Lane 4:** PCR product of *Cr7DLH* from stem cDNA, **Lane 5:** PCR product of *Cr7DLH* from root cDNA

The ORF of *Cr7DLH* was cloned in the MCS2 of pESC-URA vector, containing the cytochrome P450 reductase (*CrCPR*) in the MCS1, so that the hydroxylase activity of *Cr7DLH* can be complemented with the activity of the reductase for expression of active protein in *S. cerevisiae* InvscI cells. Positive clones were screened by colony PCR, followed by sequencing with Gal1 forward and reverse primers.

6.3.4.2 Yeast expression

Expression was carried out in INVSc1. The recombinant protein of Cr7DLH in pESC-URA was expressed and processed, as discussed earlier. The protein fractions were loaded on a 12 % SDS-PAGE (Figure 6.3.7). A band at ~60 KDa confirmed expression of Cr7DLH. The protein fractions were flash-frozen and stored at -80 °C till used.

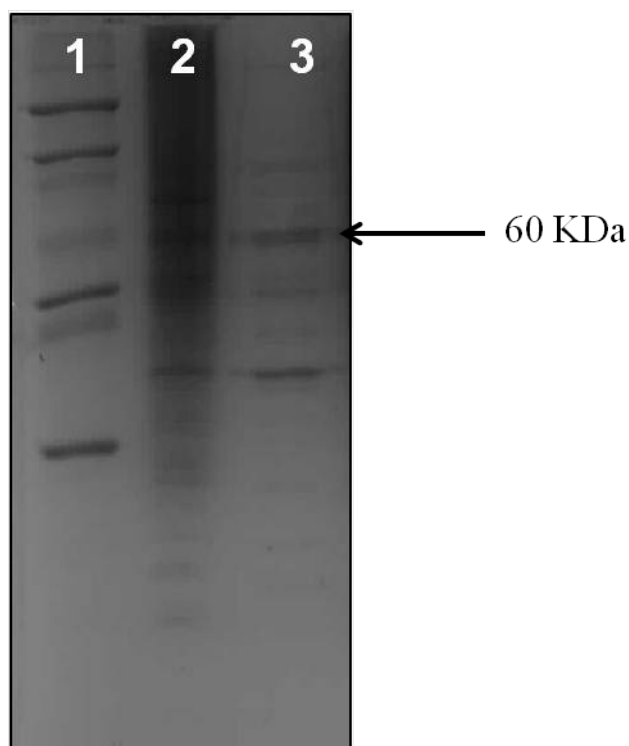


Figure 6.3.7: SDS-PAGE for Cr7DLH protein in pESC-URA+CrCPR, **Lane 1:** BenchMark™ Protein Ladder (Addendum Figure A2), **Lane 2:** Pellet fraction, **Lane 3:** Crude lysate (supernatant fraction).

6.3.5 Isolation, cloning, expression and characterization of Loganic acid Methyltransferase (*CrLAMT*)

6.3.5.1 Isolation and cloning of Loganic acid Methyltransferase (*CrLAMT*)

An Open reading frame (ORF) of 1116 bp (Figure 6.3.8), encoding a polypeptide of 371 amino acids, displaying 100 % sequence identity with loganic acid methyltransferase from *Catharanthus roseus* (GenBank ID: ABW38009.1¹⁷), was identified as loganic acid methyltransferase (*CrLAMT*). Deduced amino acid sequence of *CrLAMT* was found to have a calculated molecular weight of 42 kDa,

comprising SAM dependent carboxyl methyltransferase domain (positions 52–367) and the indole-3-acetate carboxyl methyltransferase domain (positions 22-352).

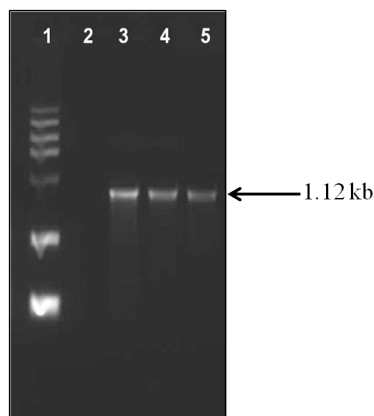


Figure 6.3.8: CrLAMT full length ORF amplification, **Lane 1:** 1 Kb DNA ladder (Addendum Figure A1), **Lane 2:** Negative Control, **Lanes 3-5:** PCR of CrLAMT full-length primers with Stem, Leaf and Root cDNAs, respectively.

This unigene was cloned in pRSET B expression vector with N-terminal His₆-tag under the control of T7-RNA polymerase promoter for expression of soluble active proteins in *E. coli* C41 (DE3) cells. Recombinant His₆-tagged proteins were purified to homogeneity by Ni²⁺-affinity chromatography.

6.3.5.2 Bacterial Expression and protein purification

Expression was carried out in C41 DE3. The recombinant protein of CrLAMT in pRSETB was expressed and purified as discussed earlier. The purified protein fractions gave a band at 42 kDa on a 12 % SDS-PAGE (Figure 6.3.9). The purified fractions were desalted, flash-frozen and store at -80 °C till used.

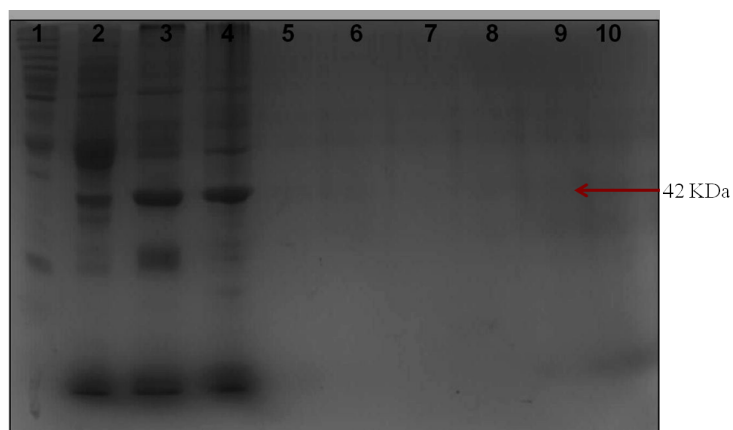


Figure 6.3.9: SDS-PAGE for CrLAMT protein purification in pRSET B, **Lane 1:** BenchMark™ Protein Ladder (Addendum Figure A2), **Lane 2:** Pellet fraction, **Lane 3:** Crude lysate (supernatant fraction), **Lane 4:** Unbound fraction, **Lanes 5-6:** Wash fractions, **Lanes 7-10:** Elution fractions.

6.3.6 Isolation, cloning and expression of secologanin synthase (*CrSLS*)

6.3.6.1 Isolation and cloning of secologanin synthase (*CrSLS*)

Full length cDNA sequence of *CrSLS* unigene was obtained by using the primers designed from the transcript number 263. The ORF of *CrSLS* composed of 1575 bp (Figure 6.3.10) encoding a polypeptide of 524 amino acids, with a calculated molecular weight of 60.6 KDa. This showed a 98 % identity with cytochrome P450 from *Catharanthus roseus* (GenBank Accession: AEX07771.1²⁶) comprising the cytochrome P450 domain (positions 102-497).

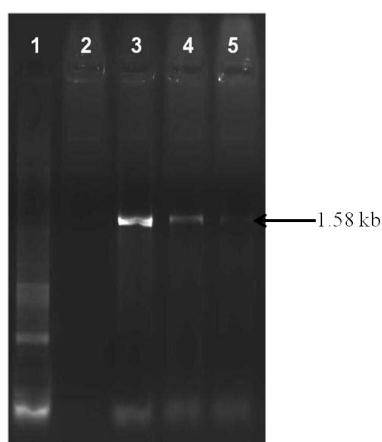


Figure 6.3.10: *CrSLS* full length ORF amplification, **Lane 1:** 1 Kb DNA ladder (Addendum Figure A1), **Lane 2:** Negative Control, **Lane 3:** PCR of *CrSLS* full-length primers with Stem cDNA, **Lane 4:** PCR of *CrSLS* full-length primers with Leaf cDNA, **Lane 5:** PCR of *CrSLS* full-length primers with Root cDNA.

The ORF of *CrSLS* was cloned in the MCS2 of pESC-URA vector, containing the cytochrome P450 reductase (*CrCPR*) in the MCS1, so that the hydroxylase activity of *CrSLS* can be complemented with the activity of the reductase for expression of active protein in *S. cerevisiae* InvscI cells. Positive clones were screened by colony PCR, followed by sequencing with Gal1 forward and reverse primers.

6.3.6.2 Yeast expression

Expression was carried out in INVSc1. The recombinant protein of *CrSLS* in pESC-URA was expressed and processed, as discussed earlier. The protein fractions were loaded on a 12 % SDS-PAGE (Figure 6.3.11). A band at ~61 KDa confirmed expression of *CrSLS*. The protein fractions were flash-frozen and store at -80 °C till used.

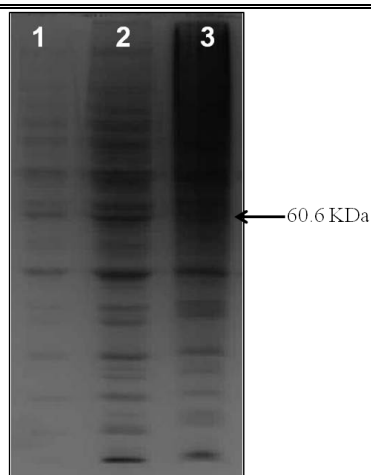


Figure 6.3.11: SDS-PAGE for CrSLS protein expression in pESC-URA+CrCPR, **Lane 1:** BenchMark™ Protein Ladder (Addendum Figure A2), **Lane 2:** Crude lysate (supernatant fraction), **Lane 3:** Pellet fraction.

6.3.7 Phylogenetic analysis of monoterpene synthases isolated from *Catharanthus roseus*

A neighbor joining phylogenetic tree placed all the full-length sequence of monoterpene synthase isolated from *C. roseus* in separate clades. All the cytochrome P450 enzymes, Cr7DLS, CrSLS and Cr7DLH were grouped together as branches of one clade, with CrSLS and Cr7DLH having the closest proximity. CrDLGT was grouped in a separate clade with other glucosyltransferase, while CrLAMT and CrCPR were placed in a separate clade. Kaurene synthase is a terpene synthase of primary metabolism and are only distantly related to those of secondary metabolism. Although the terpene synthases from *Abies grandis* (Abietadiene, Camphene and Pinene synthases) belong to that of secondary metabolism, they are branched separately due to the evolutionary separation between angiosperms and gymnosperms. It can be deciphered from the tree that monoterpene synthases from the angiosperm family are quite distinct from the monoterpene synthases from the gymnosperm family. Hence, the terpene synthases from gymnosperms are seen to form a separate branch from those of angiosperms. This indicates that the functional specialization of these terpene synthases has occurred after the divergence of gymnosperms and angiosperms from the common ancestor.

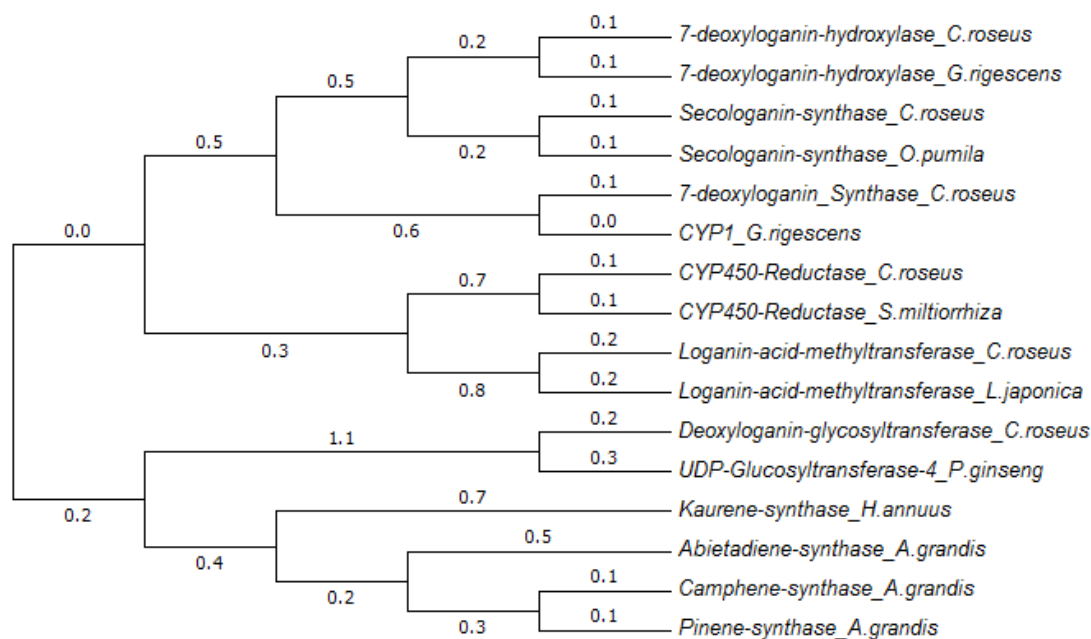
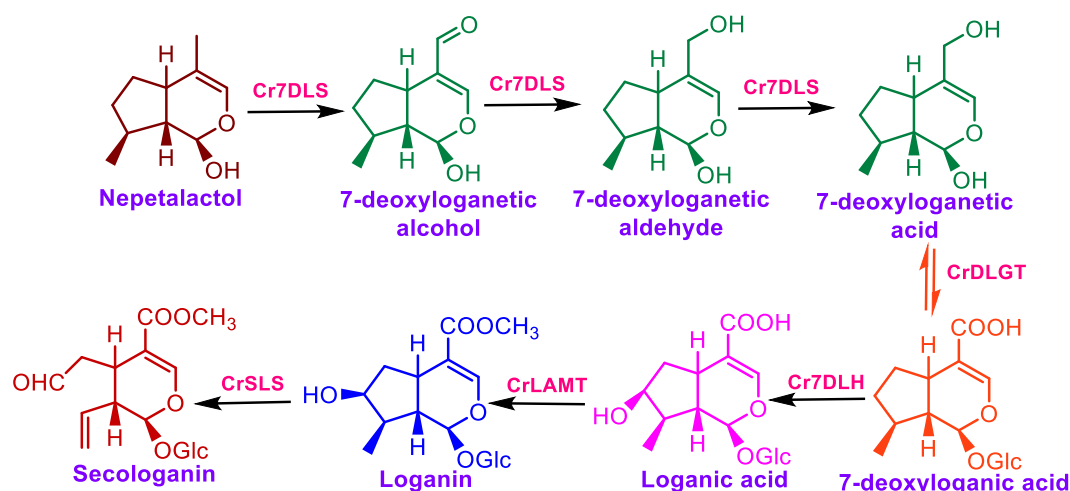


Figure 6.3.12: Phylogenetic analysis of the monoterpene synthases isolated from *C. roseus*; Sequences used for phylogenetic tree construction are: Cytochrome P450 reductase from *Catharanthus roseus*, Cytochrome P450 reductase from *Salvia miltiorrhiza* (**AGL46979.1**), 7-Deoxyloganin Synthase from *Catharanthus roseus*, CYP1 from *Gentiana rigescens* (**AKJ26120.1**), Secologanin synthase from *Catharanthus roseus*, Secologanin synthase from *Ophiorrhiza pumila* (**BAP90521.1**), 7-deoxyloganin hydroxylase from *Catharanthus roseus*, 7-deoxyloganin hydroxylase from *Gentiana rigescens* (**ALJ10970.1**), 7-deoxyloganic acid glucosyltransferase from *Catharanthus roseus*, UDP-Glucosyltransferase-4 from *Panax ginseng* (**AIE12477.1**), Loganic acid methyltransferase from *Catharanthus roseus*, Loganic acid O-methyltransferase from *Lonicera japonica* (**AMB61019.1**), Kaurene synthase from *Helianthus annuus* (**CBL42917**), Abietadiene synthase from *Abies grandis* (**AAK83563**), Camphene synthase from *Abies grandis* (**AAB70707**), Pinene synthase from *Abies grandis* (**AAB71085**).

6.4 Discussion and Conclusion

The transcriptome was screened for the remaining genes of the Secologanin biosynthetic pathway resulting in the identification of 6 transcripts, cytochrome P450 reductase (**CrCPR**), 7-deoxyloganin synthase (**Cr7DLS**), 7-deoxyloganetic acid glucosyltransferase (**CrDLGT**), 7-deoxyloganin hydroxylase (**Cr7DLH**), loganic acid methyltransferase (**CrLAMT**) and secologanin synthase (**CrSLS**). Each of the cytochrome P450 genes, namely, Cr7DLS, Cr7DLH and CrSLS were coned in the MCS2 of the pESC-URA yeast duet vector, containing CrCPR in its MCS1, and their expressions were carried out. CrDLGT was cloned in pET28a, whereas, CrLAMT was cloned in pRSETB bacterial vectors and their expressions were carried out under T7 promoter. These bacterial proteins were then purified.



Scheme 6.4.1: Formation of Secologanin from Iridoids

While we were carrying out this work, the cloning and characterization studies of these genes were published by different groups working on this plant^{12,14,15,17,22,26}. However, their cloning and characterization were carried out in an attempt to carry out the seven steps of the pathway (from nepetalactol to secologanin) in a single step either by way of co-expression of these six genes or by constructing an operon model of the same, which is under progress. In comparison with the reported data on these genes^{12,14,15,17,22,26}, the sequences show a 100 % similarity. CrDLGT and CrLAMT were expressed in *E. coli* and purified. Although CrLAMT enzyme has been partially purified from *Vinca rosea*²⁷, this is the first report for its purification after heterologous expression in a host.

Phylogenetic analyses of these monoterpene synthases was carried out with various other terpene synthases, which showed the distinct bifurcation between gymnosperm and angiosperm terpene synthases as well as the diverse evolutionary branching of primary terpene synthases from secondary terpene synthases.

6.5 Appendix: Agarose gel electrophoresis for colony PCR screening for gene cloning

6.5.1 Cloning of *CrCPR* in MCS1 of pESC-URA vector frame

Colony PCR screening of Cytochrome P450 reductase cloned in MCS1 of pESC-URA vector frame with Gal10 forward and reverse primers.

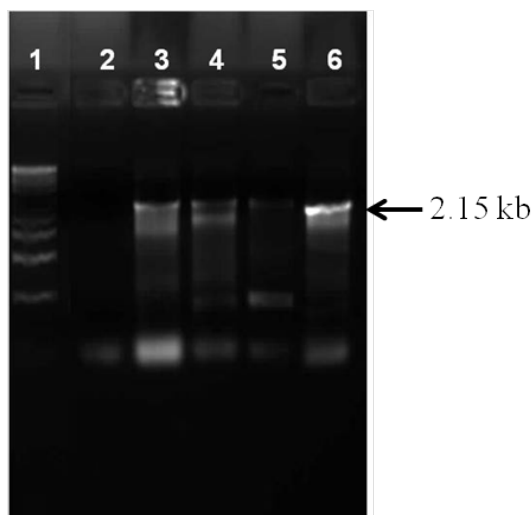


Figure 6.5.1: Colony PCR screening for *CrCPR* ORF cloned in MCS1 of pESC-URA vector, **Lane 1:** 1 Kb DNA ladder (Addendum Figure A1), **Lane 2:** Negative control, **Lanes 3-6:** Colony PCR with Gal10 forward and reverse primers.

6.5.2 Cloning of *Cr7DLS* in MCS2 of pESC-URA+CrCPR vector frame

Colony PCR screening of 7-deoxyloganin synthase cloned in MCS2 of pESC-URA+CrCPR vector frame with Gal1 forward and reverse primers.

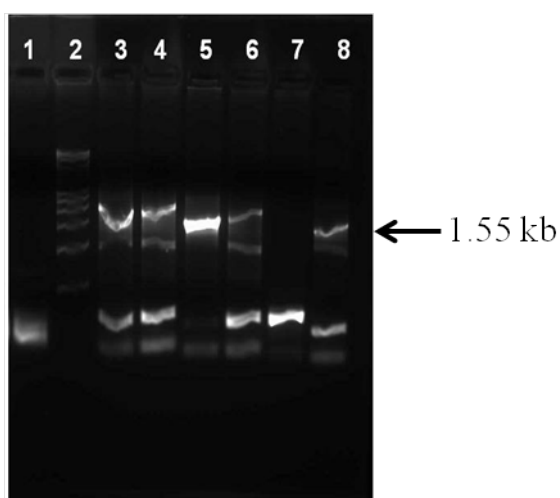


Figure 6.5.2: Colony PCR screening of full-length ORF of *Cr7DLS* in MCS2 of pESC-URA+CrCPR, **Lane 1:** Negative control, **Lane 2:** 1 Kb DNA ladder (Addendum Figure A1), **Lanes 3-8:** Colony PCR with Gal1 forward and reverse primers.

6.5.3 Cloning of *CrDLGT* in pET28a vector

Colony PCR screening of 7-deoxyloganic acid glucosyltransferase cloned in pET28a vector frame with T7 promoter and reverse primers.

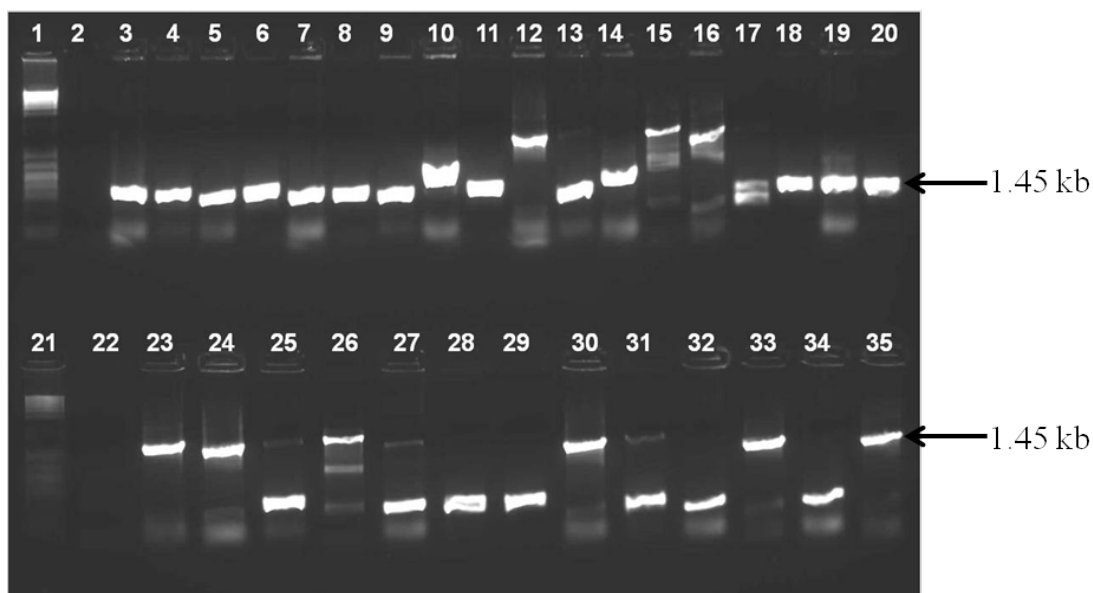


Figure 6.5.3: Colony PCR Screening for *CrDLGT* full length ORF cloned in pET28a, **Lanes 1 & 21:** 1 Kb DNA ladder (Addendum Figure A1), **Lane 2:** Negative control, **Lanes 3-35:** Colony PCR with T7 promoter and reverse primers.

6.5.4 Cloning of ORF of *Cr7DLH* in MCS2 of pESC-URA+CrCPR vector frame

Colony PCR screening of 7-deoxyloganin hydroxylase cloned in MCS2 of pESC-URA+CrCPR vector frame with Gal1 forward and reverse primers.

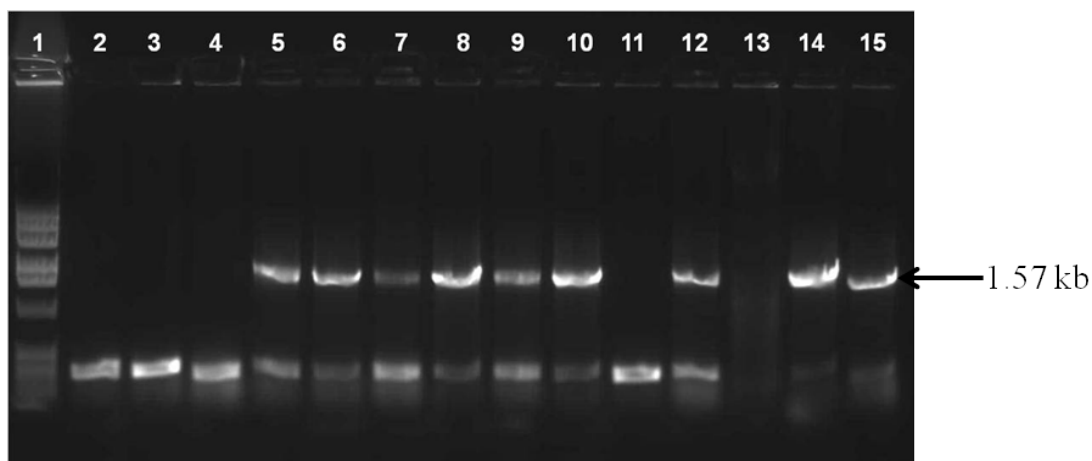


Figure 6.5.4: Colony PCR screening for full-length ORF *Cr7DLH* in MCS2 of pESC-URA+CrCPR, **Lane 1:** 1 Kb DNA ladder (Addendum Figure A1), **Lane 2:** Negative control, **Lanes 3-15:** Colony PCR with Gal1 forward and reverse primers.

6.5.5 Cloning of ORF of *CrLAMT* in pRSETB

Colony PCR screening of ORF of *CrLAMT* cloned in pRSETB vector frame with T7 promoter and reverse primers.

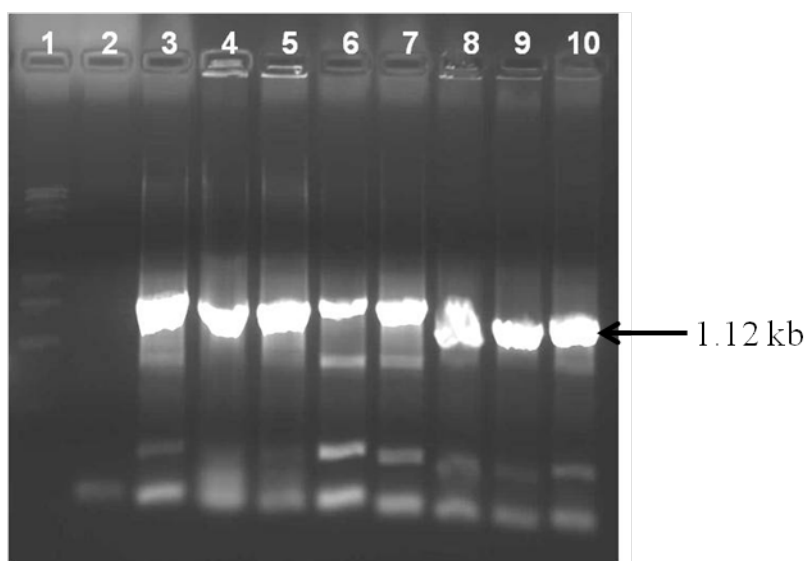


Figure 6.5.5: Colony PCR screening for ORF of *CrLAMT* cloned in pRSETB vector, **Lane 1:** 1 Kb DNA ladder (Addendum Figure A1), **Lane 2:** Negative control, **Lanes 3-10:** Colony PCR with T7 promoter and reverse primers.

6.5.6 Cloning of ORF of *CrSLS* in MCS2 of pESC-URA+CrCPR vector frame

Colony PCR screening of Secologanin synthase cloned in MCS2 of pESC-URA+CrCPR vector frame with Gal1 forward and reverse primers.

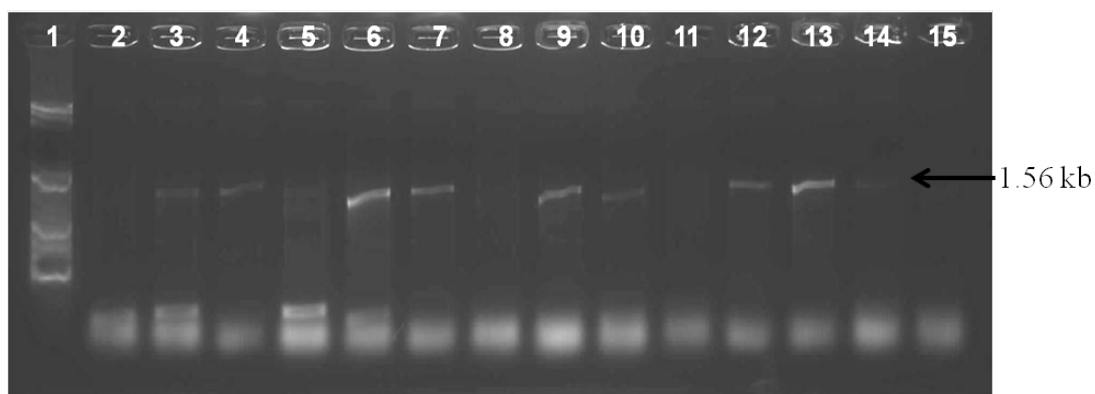


Figure 6.5.6: Colony PCR screening for full-length ORF *CrSLS* in MCS2 of pESC-URA+CrCPR, **Lane 1:** 1 Kb DNA ladder (Addendum Figure A1), **Lane 2:** Negative control, **Lanes 3-15:** Colony PCR with Gal1 forward and reverse primers.

6.6 References

1. Hays, S.B. Some effects of reserpine, a tranquilizer, on House Fly. *J. Econ. Entomol.* **58**, 782-784 (1965).
2. Yui, T. & Takeo, Y. Neuropharmacological studies on a new series of ergot alkaloids; elymoclavine as a potent analeptic on reserpine-sedation. *Jpn. J. Pharmacol.* **7**, 157-161 (1958).
3. Yui, T. & Takeo, Y. Neuropharmacological studies on a new series of ergot alkaloids; the effects on electrocorticogram of rabbits. *Jpn. J. Pharmacol.* **7**, 162-168 (1958).
4. Shurin, G.V., Tourkova, I.L., Kaneno, R. & Shurin, M.R. Chemotherapeutic Agents in Noncytotoxic Concentrations Increase Antigen Presentation by Dendritic Cells via an IL-12-Dependent Mechanism. *J. Immunol.* **183**, 137-144 (2009).
5. van der Heijden, R., Jacobs, D.I., Snoeijer, W., Hallared, D. & Verpoorte, R. The *Catharanthus* alkaloids: Pharmacognosy and biotechnology. *Curr. Med. Chem.* **11**, 607-628 (2004).
6. Govindachari, T.R. & Viswanat.N. Alkaloids of *Mappia foetida*. *Phytochemistry* **2**, 3529-3531 (1972).
7. Yamazaki, M., Mochida, K., Asano, T., Nakabayashi, R., Chiba, M., Udomson, N., Yamazaki, Y., Goodenowe, D. B., Sankawa, U., Yoshida, T., Toyoda, A., Totoki, Y., Sakaki, Y., Gongora-Castillo, E., Buell, C. R., Sakurai, T., & Saito, K. Coupling Deep Transcriptome Analysis with Untargeted Metabolic Profiling in *Ophiorrhiza pumila* to Further the Understanding of the Biosynthesis of the Anti-Cancer Alkaloid Camptothecin and Anthraquinones. *Plant Cell Physiol.* **54**, 686-696 (2013).
8. Staines, H.M. & Krishna, S. *Treatment and Prevention of Malaria : Antimalarial Drug Chemistry, Action and Use*, (Springer Science & Business Media, 2012).
9. Heckendorf, A.H. & Hutchinson, C.R. Biosynthesis of camptothecin. II. Confirmation that isovincoside, not vincoside, is the penultimate biosynthetic precursor of indole alkaloids. *Tetrahedron Lett.*, 4153-4154 (1977).
10. Ma, X. Y., Koepke, J., Bayer, A., Linhard, V., Fritzsche, G., Zhang, B., Michel, H., & Stockigt, J. Vinorine synthase from *Rauvolfia*: the first example of crystallization and preliminary X-ray diffraction analysis of an enzyme of the BAHD superfamily. *Biochim. Biophys. Acta, Proteins Proteomics* **1701**, 129-132 (2004).
11. Sharma, R.K. *Concise textbook of forensic medicine and toxicology*, (Journal of Punjab Academy of Forensic Medicine & Toxicology, 2005).
12. Salim, V., Wiens, B., Masada-Atsumi, S., Yu, F. & De Luca, V. 7-Deoxyloganetic acid synthase catalyzes a key 3 step oxidation to form 7-deoxyloganetic acid in *Catharanthus roseus* iridoid biosynthesis. *Phytochemistry* **101**, 23-31 (2015).
13. Miettinen, K., Dong, L. M., Navrot, N., Schneider, T., Burlat, V., Pollier, J., Woittiez, L., van der Krol, S., Lugan, R., Ilc, T., Verpoorte, R., Oksman-Caldentey, K. M., Martinoia, E., Bouwmeester, H., Goossens, A., Memelink, J., & Werck-Reichhart, D. The seco-iridoid pathway from *Catharanthus roseus*. *Nature Comm.* **5**, 3606-3617(2014).
14. Asada, K., Salim, V., Masada-Atsumi, S., Edmunds, E., Nagatoshi, M., Terasaka, K., Mizukami, H., & De Luca, V. A 7-Deoxyloganetic Acid Glucosyltransferase Contributes a Key Step in Secologanin Biosynthesis in Madagascar Periwinkle. *Plant Cell* **25**, 4123-4134 (2013).

-
15. Salim, V., Yu, F., Altarejos, J. & De Luca, V. Virus-induced gene silencing identifies *Catharanthus roseus* 7-deoxyloganin acid-7-hydroxylase, a step in iridoid and monoterpene indole alkaloid biosynthesis. *Plant J.* **76**, 754-765 (2013).
 16. Madyastha, K.M., Guarnaccia, R., Baxter, C. & Coscia, C.J. S-Adenosyl-L-methionine: loganic acid methyltransferase. A carboxyl-alkylating enzyme from *Vinca rosea*. *J. Biol. Chem.* **248**, 2497-2501 (1973).
 17. Murata, J., Roepke, J., Gordon, H. & De Luca, V. The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization. *Plant Cell* **20**, 524-542 (2008).
 18. Vetter, H. P., Mangold, U., Schröder, G., Marner, F. J., Werck-Reichhart, D., & Schröder, J. Molecular Analysis and Heterologous Expression of an Inducible Cytochrome P-450 protein from Periwinkle (*Catharanthus roseus* L. *Plant Physiol.* **100**, 998-1007 (1992).
 19. Saitou, N. & Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406-425 (1987).
 20. Zuckerkandl, E. & Pauling, L. (eds.). *Evolutionary divergence and convergence in proteins. Evolving Genes and Proteins*, **69** (Academic Press, New York, New York, 1965).
 21. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **30**, 2725-2729 (2013).
 22. Meijer, A.H. et al. Isolation and characterization of a cDNA clone from *Catharanthus roseus* encoding NADPH:cytochrome P-450 reductase, an enzyme essential for reactions catalysed by cytochrome P-450 mono-oxygenases in plants. *Plant J.* **4**, 47-60 (1993).
 23. Zhang, X.D., Allan, A.C., Li, C.X., Wang, Y.Z. & Yao, Q.Y. De Novo Assembly and Characterization of the Transcriptome of the Chinese Medicinal Herb, *Gentiana rigescens*. *Int. J. Mol. Sci.* **16**, 11550-11573 (2015).
 24. Yan, X., Fan, Y., Wei, W., Wang, P., Liu, Q., Wei, Y., Zhang, L., Zhao, G., Yue, J. & Zhou, Z. Production of bioactive ginsenoside compound K in metabolically engineered yeast. *Cell Res.* **24**, 770-773 (2014).
 25. Salim, V., Yu, F., Altarejos, J. & De Luca, V. Virus-induced gene silencing identifies *Catharanthus roseus* 7-deoxyloganin acid-7-hydroxylase, a step in iridoid and monoterpene indole alkaloid biosynthesis. *Plant Journal* **76**, 754-765 (2013).
 26. Huang, L., Li, J., Ye, H., Li, C., Wang, H., Liu, B., & Zhang, Y. Molecular characterization of the pentacyclic triterpenoid biosynthetic pathway in *Catharanthus roseus*. *Planta* **236**, 1571-1581 (2012).
 27. Madyastha, K.M., Guarnaccia, R., Baxter, C. & Coscia, C.J. S-Adenosyl-L-Methionine - Loganic Acid Methyltransferase - Carboxyl-Alkylating Enzyme From *Vinca-Rosea*. *J. Biol. Chem.* **248**, 2497-2501 (1973).

Addendum

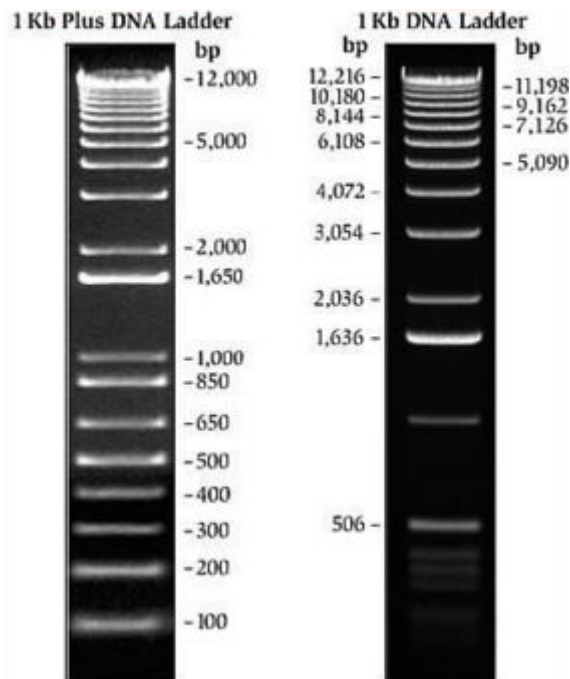
DNA and Protein Ladders

Figure A1: Comparison between 1 Kb Plus DNA Ladder and 1 Kb DNA Ladder from Thermo Fischer Scientific, USA on a 0.9 % agarose gel stained with Ethidium bromide.

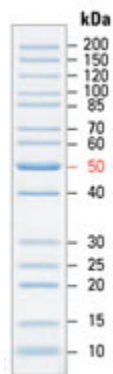


Figure A2: Band distribution of the BenchMark™ Protein Ladder from Thermo Fischer Scientific, USA on a 12.5% Tris-Glycine Gel stained with Coomassie® Brilliant Blue R-250.

**Sequences of monoterpene synthases
isolated from *Catharanthus roseus***

1) Geranyl diphosphate synthase (*CrGDS*; Accession Number: KF561462)

>Vinca_5406_CrGDS_1263bp

ATGTTGTTTTCCAGAGGATTGTATAGGATCGCAAGGACGAGTTTGAACAGAAGTCGATTGCT
TTACCCGTTACAAAGTCAGTCGCCGGAGCTGCTGCAGTCTTTTCAGTTTCGCTCTCCTATTG
GTTCTTCTCAAAAGTTTTAGTTTTAGAGTAATCTATTTCATGGGTCTCAAGTGCCCTGGCC
AATGTTGGACAGCAGGTACAGCGCCAGAGCAACTCTGTTGCCGAGGAGCCACTAGATCCATT
TTCACCTGTTGCTGATGAATTGTCCATTCTTGCTAATAGACTGAGGTCAATGGTAGTTGCAG
AGGTCCCGAAGCTTGCTTCAGCTGCCGAATATTTTTTCAAGTTAGGGGTGGAAGGAAAGAGG
TTTCGACCAACAGTTTTGCTATTGATGGCGACAGCTATAGATGCACCAATATCTAGAACC
TCCTGATACATCACTTGATACTTTATCCACAGAACTACGCCTAAGGCAGCAGTCGATTGCTG
AGATCACTGAGATGATCCATGTTGCTAGTCTTCTTCATGACGATGTATTAGATGATGCTGAA
ACAAGCGAGGGATTGGTTCTCTAAATTTGTGATGGGAAATAAGTTAGCAGTGTGGCTGG
TGATTTCTGCTATCAAGAGCCTGTGTTGCACTTGCCTCTTTGAAAAATACAGAGGTCGTGT
CCCTCTTGGCAACAGTTGTGGAGCATCTTGTACGGGTGAAACGATGCAAATGACCACCACA
TCTGATCAACGTTGTAGCATGGAGTACTATATGCAAAAGACATACTATAAGACGGCATCCTT
GATCTCAAACAGTTGCAAAGCAATTGCCCTTCTTGCTGGGCAAACATCAGAAGTTGCAATGT
TGGCTTATGAGTATGGAAAAATCTGGGATTGGCGTTTTAGTTAATAGATGATGTTCTTGAT
TTCACCTGGCACATCAGCTTCCCTTGGCAAGGGCTCTCTGTCTGACATTCGCCACGGAAATGT
TACTGCTCCAATATTATTTGCCATAGAAGAGTTCCCTGAACTACGTGCTGTTGTTGACGAGG
GATTTGAAAATCCATATAATGTAGATCTTGCTCTACATTACCTTGGAAAAGAGTAGAGGAATA
CAACGAACGAGGGAAGTGGCAATAAAGCATGCTAACCTTGCCTCTGATGCAATCGACTCTCT
TCCGGTGACCGATGATGAACATGTTTTAAGGTCAAGAAGAGCTCTTGTGGAACCTTACTCAAC
GCGTTATTACAAGAAGAAAGTGA

>CrGDS_Protein Sequence

MLFSRGLYRIARTSLNRSRLLYPLQSQSPELLQSFQFRSPIGSSQKVSGFRVIYSWVSSALA
NVGQQVQRQSNVAEEPLDPFSLVADELSILANRLRSMVVAEVPKLAASAEYFFKLGVEGKR
FRPTVLLLMATAIDAPISRTPPDTSLDLSTELRLRQQSIAEITEMIHVASLLHDDVLDLDAE
TRRGIGSLNFVMGNKLAVLAGDFLLSRACVALASLKNTEVVSLLATVVEHLVLTGETMQMTT
SDQRCSMEYYMQKTYKASLISNSCKAIALLAGQTSEVAMLAYEYKGNLGLAFQLIDDVLD
FTGTSASLGKGSLSDIRHGIVTAPILFAIEEFPELRAVVDEGFENPYNVDLALHYLKGSRGI
QRTRELAIKHANLASDAIDSLPVTDDHVLRSRRALVELTQRVITRRK-

2) Geraniol Synthase (*CrGS*; Accession Number: KF561459)

>Vinca_2171_CrGS_1770bp

ATGGCAGCCACAATTAGTAACCTTTCTTTCTTAGCAAAATCTAGGGCACTTTCAAGGCCTTC
TTCTTCTTCACTTTCATGGCTAGAAAGGCCTAAACTTTCATCGACTATTTGCATGTCTATGC
CATCATCTTTCATCATCATCATCATCTTTCATCCATGTCTCTGCCTTTGGCAACTCCATTGATC
AAAGACAATGAATCTCTCATCAAGTTCTTGCGCAACCCATGGTGCTTCTCATGAGGTTGA
TGACAACACCAAAAGGAGGGAAATTGCTGGAAAGAACAAGAAAAGAACTAGAATTAATGCAG
AAAACCATTGGAGGCCTTGAAGATGATAGATATAAATTCAAAGATTGGGATTATCATATCAT
TTTGAAGATGATATTAATTCAATTCTCACAGGATTTTCAAATATTAGCAGCCAAACTCATGA
AGATCTCCTCACTGCTTCACTTCGTTTTTCGATTGCTTTCGACACAATGGGCATAAGATCAGTC
CTGATATATTCCAAAATTCATGGACAACAATGGAAAGTTTAAAGATTCATTAAAGGATGAC
ACATTAGGCATGTTAAGCTTATATGAAGCTTCATATTTGGGAGCCAATGGAGAAGAAAATATT
GATGGAAGCCCAAGAATTCACCAAAACTCACCTGAAAACCTCATTGCCAGCCATGGTACCAT

CTCTTTCTAAGAAGGTTTCTCAAGCTTTAGAGCAACCAAGACATAGAAGAATGTTGAGGTTA
GAAGCTAGAAGATTTATTGAAGAATATGGTGCTGAAAATGACCATAATCCAGACCTTCTTGA
GCTTGCAAAATTGGATTATAACAAAGTCCAATCTCTACACCAATGGAATTGTCTGAGATAA
CAAGGTGGTGGAAACAATTAGGGCTTGTGGATAAACTCACCTTTGCTCGAGATCGACCCCTT
GAATGCTTTCTTTGGACAGTGGGATTATTACCAGAGCCTAAGTATTCAGGTTGCAGAAATGA
GCTTGCAAAAACCATAGCCATTTTGCTTGTCAATTGATGATATCTTTGATACTCATGGTACCC
TAGATGAGCTTCTTCTATTCACTAATGCCATTTAAAGATGGGATCTTGAGGCCATGGAAGAT
TTACCAGAATATATGAGAATTTGTTACATGGCATTGTACAATACTACTAATGAAATTTGCTA
TAAAGTCTTAAAGGAAAATGGTTGGAGTGTCTTCCCTTACCTAAAGGCAACGTGGATTGATA
TGATTGAAGGATTCATGGTTGAAGCAGAATGGTTCAATTCTGATTATGTACCAAACATGGAA
GAATATGTAGAAAATGGAGTTAGAACAGCAGGATCATATATGGCCTTAGTCCATTTGTTCTT
TCTAATAGGGCAAGGTGTCACACAAGATAATGTGAAATTACTGATTAAACCCTATCCAAAGC
TCTTTTCCCTCCTCAGGAAGAATCCTTCGACTTTGGGATGATTTGGGAACCTCAAAGGAGGAA
CAAGAAAGAGGAGACTTGGCATCAAGCATAACAATTGTTTATGAGAGAGAAAGAGATAAAATC
AGAAGAAGAAGGAAGGAAAGGAATATTGGAAATTATAGAGAATTTATGGAAAGAATTGAATG
GAGAATTAGTTTATAGAGAAGAAATGCCTCTTGAATAATCAAGACAGCATTCAACATGGCA
AGAGCTTCCCAAGTTGTGTATCAACATGAAGAAGACACCTATTTTTCAAGTGTAGATAATTA
TGTAAGGCTTTGTTTTTTTACACCTTGTTTTTAA

>CrGS_Protein_Sequence

MAATISNLSFLAKSRALSRPSSSSLSWLERPKTSSTICMSMPSSSSSSSSSSMSLPLATPLI
KDNESLIKFLRQPMVLPHEVDDNTRKRELLERTRKELELNAEKPLEALKMIDI IQRLGLSYH
FEDDINSILTGFSNISSQTHEDLLTASLRFRLLRHNGHKISPDIFQKFMNNGKFKDSLKDD
TLGMLSLEYEASYLEGANGEEILMEAEQFTKTHLKNLSPAMVPSLSKKVVSQALEQPRHRRLRL
EARRFIEEYGAENDHNPDLLELAKLDYKNVQSLHQMELSEITRWWKQLGLVDKLTFFARDRPL
ECFLWTVGLLPEPKYSGRIELAKTIAILLVIDDIFDTHGTLDELLFTNAIKRWDLEAMED
LPEYMRICYMALYNTTNEICYKVLKENGWSVLPYLKATWIDMIEGFMVEAEWFNSDYVPNME
EYVENGVRTAGSYMALVHLFFLIGQGVTDQNVKLLIKPYPKLFSSSGRILRLWDDLGTSKEE
QERGLASSIQLFMREKEIKSEEEGRKGILEI IENLWKELNDELVYREEMPLAI IKTAFNMA
RASQVVYQHEEDTYFSSVDNYVKALFFTPCF

3) Geraniol 10-hydroxylase (*CrG10H*; Accession Number: KF561461)

>Vinca_742_CrG10H_1482bp

ATGGATTACCTTACCATAATATTAACCTTACTATTTGCCTTGACTCTCTATGAAGCCTTCAG
CTACCTATCCAGAAGAACCAAAAACCTTCCCTCCAGGACCATCGCCATTGCCGTTTCATCGGAA
GCCTCCATTTATTAGGCGACCAACCACACAAATCCTTAGCAAAAACCTTCCAAAAACACGGT
CCAATTATGAGTCTCAAATTAGGCCAGATCACTACAATCGTCATATCTTCATCAACAATGGC
GAAAGAAGTTCTTCAAAAACAGGATTTAGCATTTTCAAGCAGATCAGTTCCAAACGCACTCC
ACGCTCACAATCAATTCAAATTCCTCCGTTGTATGGCTTCCGGTAGCCTCACGATGGAGAAGT
CTTCGAAAAGTTTTGAATTCTAATATATTTTCCGGCAATCGGCTCGACGCTAATCAACATTT
GAGAAGTAGAAAAGTACAGGAACATAATTGCGTATTGCCGAAAAATAGCCAGAGCGGAGAAG
CGGTTGACGTCGGCCGAGCTGCTTTTAGAACTTCGTTGAATTTGTTGTCGAATTTGATTTTT
TCAAAGGATTTGACGGATCCTTATTCGGATTCTGCCAAGGAATTCAGGATTTGGTTTGAA
TATAATGGTTGAGGCGGGGAAACCTAATTTGGTTCGATTTTTTTCCCTGCTTGAAAAAGTTG
ATCCTCAAGGTATACGACATCGTATGACGATTCACCTTTGGGGAAGTCTTAAGCTTTTTGGT
GGACTTGTTAATGAAAGATTGGAGCAAAGAAGATCAAAAAGGGGAAAAAATGATGTGTGGA

TGTACTTCTAACTACCAGCCAAGAAAGCCCTGAGGAAATCGATAGAACTCACATTGAGCGAA
TGTGCTTGGACCTGTTTGTAGCAGGGACGGACACAACATCAAGCACATTAGAATGGGCAATG
TCAGAAATGCTTAAAAACCCAGACAAAATGAAGAAAACCCAAGATGAACTTGCACAAGTAAT
CGGCAGAGGAAAAACAATAGAAGAATCCGATATTAACCGCTTACCTTACTTAAGATGCGTTA
TGAAAGAAACCTTAAGGATACATCCACCAGTTCCTTCTTAATTCCTCGCAAAGTGGAAACAA
AGTGTGAGGTTTGTGGATACAATGTCCCTAAAGGATCACAAGTTCTTGTGAATGCTTGGGC
AATGGACGTGATGAACTGTTGGGATGATGCTTGGCATTCAAACCCGAGAGATTTATGG
AATCTGAATTGGATATCCGTGGAAGAGATTTGAGCTGATTCGGTTCGGTCTGGCCGAAGA
ATTTGCCAGGGTTGCCATTGGCACTAAGGACTGTGCCTTTGATGCTTGGTCTTTTGTGAA
CTCTTTTAATTGGAAGCTTGAAGGTGGGATGGCTCCAAAAGATTTGGATATGGAGGAGAAGT
TTGGTATTACACTGCAGAAGGCTCATCCTTTGCGTGCTGTACCAAGCACCTTTAA

>CrG10H_Protein Sequence

MDYLTIIILTLFALTYEAFSYLSRRTKNLPPGPSPLPFIGSLHLLGDQPHKSLAKLSKKHG
PIMSLKLGQITTIIVISSSTMAKEVQLQKDLAFSSRSVFNALHAHNQFKFSVWVLPVSRWRS
LRKVLNSNIFSGNRLDANQHLRTRKVVQELIAYCRKNSQSGEAVDVGRAAFRTSLNLLSNLIF
SKDLTDPYSDSAKEFKDLVWNIMVEAGKPNLVDFPPLLEKVDPPQIRHRMTIHFGEVLKLF
GLVNERLEQRRSKGEKNDVLDVLLTTSQESPEEIDRTHIERMCLDLFVAGTDTTSSTLEWAM
SEMLKNPDKMKTQDELAQVIGRGTIEESDINRLPYLRVCMKETLRIHPPVPLIPRKVEQ
SVEVCGYNVPKGSQVLVNAWAIGRDETVDLALAFKPERFMESELDIRGRDFELIPFGAGRR
ICPGLPLALRTVPLMLGSLNLSFNWKLEGGMAPKDLDMEEKFGITLQKAHPLRAVPSTL-

4) 10-hydroxygeraniol dehydrogenase (*Cr10HGO*; Accession Number: KF561458)

>Vinca_536_Cr10HGO_1083bp

ATGGCGAAATCACCGGAAGTCGAGCATCCAGTTAAGGCCTTTGGATGGGCTGCTAGAGACACTTCTGGG
CATCTCTCTCCTTTTCATTTCTCCCGAAGGGCTACCGGAGAACATGATGTGCAATTCAAAGTTTTATAC
TGCGGGATCTGTCACTCGGATCTTACATGATCAAGAACGAATGGGGCTTACGAAATATCCCATTGTA
CCTGGGCATGAAATTGTGGGTATAGTTACAGAGGTTGGTAGTAAGGTTGAGAAATTCAGGTTGGGGAT
AAAGTAGGAGTGGGTTGCCTTGTGGGATCATGCCGAAATGTGATATGTGCACTAAAGATCTTAAAAAT
TATTGTCCGGGCCAAATACTCACATACAGCGCTACTTATACTGATGGAACAACCTACCTATGGAGGCTAT
TCTGATCTAATGGTTGCTGATGAACACTTTGTGATTCGTTGGCCTGAGAAATTTGCCTATGGATATTGGT
GCTCCTTTGCTTTGTGCTGGGATTACTACATACAGTCCATTGAGATATTTTGGACTTGATAAACAGGA
ACCCATGTTGGCGTCGTTGGTCTTGGTGGTCTTGGCCATGTAGCTGTGAAATTTGCCAAGGCTTTTGGT
GCCAAAGTTACTGTCAATTAGCACATCTGAGAGTAAAAAACAGGAAGCCTTAGAGAACTTGGTGTGAT
TCTTTCTTGGTTAGTCGTGACCCTGAACAGATGAAGGCAGCAGCGGCTAGTTTGGATGGAATTATTGAC
ACCGTTTCTGCAATTCATCCAATCATGCCCTGCTTAGCATATTGAAGTCTCATGGGAAGCTAATTCCTT
GTTGGTGCACCAGAAAAGCCACTTGAGTTGCCATCGTTCCTCTTATTGCTGGGAGGAAAATAATAGCT
GGAAGTGAATTTGGAGGGTTGAAAGAACTCAAGAAATGATAGATTTTGCAGCAAAGCACAAATGTTTTG
CCTGATGTGGAACCTTGTCTCGATGGATTATGTAACACTGCAATGGAGAGGCTACTGAAAGCTGATGTT
AAATATCGTTTTGTCAATTGATGTAGCCAACACACTGAAATCTGCATAA

>Cr10HGO_Protein Sequence

MAKSPEVEHPVKAFGWAARDTSGHLSPFHFSRRATGEHDVQFKVLYCGICHSDLHMIKNEWG
FTKYPIVPGHEIVGIVTEVGSKVEKFKVGDVGVGCLVGSCKMCTKDLENYCPGQILTY

SATYTDGTTTTYGGYS DLMVADEHFVIRWPENLPM DIGAPLLCAGITTY SPLRYFGLDKPGTH
VGVVGLGGLGHVAVKFAKAFGAKVTVISTSESKQEALEKLGADSFVLSRDPEQMKAASL
DGIIDTVSAIHPIMPLLSILKSHGKLLILVGAPEKPLELPSFPLIAGRKIIAGSAIGGLKETQ
EMIDFAAKHNVLDPVELVSM DYVNTAMERLLKADV KYRFVIDVANTLKSA-

5) Iridoid synthase (*CrIDS*; Accession Number: KF561460)

>Vinca_2280_CrIDS_1161bp

ATGAGTTGGTGGTGGAAAGAGGTCCATTGGTGCTGGAAAGAACTTACAGAACAAAGAAAATGG
AGTTTGTA AAAACTACAAGAGTGTGGCACTAGTAGTAGGAGTCACCGGCATTGTTGGCAGCA
GTTTAGCCGAAGTCTCTAAAGCTACCAGACTCCTGGCGGTCCCTGGAAGGTATACGGCGTG
GCAAGGCGCCCATGCCCTGCATGGCTAGCCAAAAACCAGTTGAGTACATCCAATGTGATGT
CTCAAATAATCAAGAAACAATTTCTAAATTATCTCCACTTAAGGACATAACTCATATATTCT
ATGTTTCTTGGATTGGATCAGAAGATTGCCAAACAAATGCTACAATGTTCAAAAACATCCTT
AATTCAGTGATTCCTAATGCTTCAAATCTCCAACATGTGTGTCTACAAACAGGAATCAAACA
TTATTTTGGTATTTTGAAGAAGGTTCTAAAGTTGTACCACATGATTCGCCTTTTACCGAGG
ATTTACCTAGGCTAAATGTCCAAAATTTCTACCATGATCTTGAAGATATACTGTATGAAGAA
ACTGGCAAGAACAATTTAACATGGTCTATCCATAGACCAGCTCTGGTTTTCGGGTTTTCCCC
ATGTAGTATGATGAACATTGTCAGTACATTATGCGTTTTACGCCACAATTTGTAAACATGAAA
ATAAGGCCTTGGTTTTATCCAGGTAGTAAAAATTCATGGAATTGTTATGCTGATGCTGTTGAT
GCAGATTTAGTAGCCGAGCACGAAATTTGGGCAGCAGTTGATCCTAAGGCTAAAAATCAAGT
ATTGAATTGTAACAATGGGGATGTTTTTAAATGGAACATATTTGGAAGAAATTAGCAGAGG
AATTTGGGATTGAAATGGTGGGATATGTTGAAGGTAAAGAACAAGTTAGTTTGGCTGAGTTA
ATGAAAGATAAAGATCAAGTTTGGGATGAAATGTTAAGAAGAACAATTTAGTACCTACTAA
ATTGAAAGAAATGCTGCCTTTTGGTTTTGCTGATATTGCTTTTTTGTCTGAGAATTTGATTA
GTAGTATGAACAAAAGTAAGGAATTAGGGTCTTGGGGTTTAGAAATTCGATGAAATCGTTT
GTTTCTTGTATTGATAAGATGAGGGATTATAGGTTTTATTCCTTAG

>CrIDS_Protein Sequence

MSWWWKRSIGAGKNLQNKENGVCNKYKSVALLVVGVTGIVGSSLAEVLKLPDTPGGPWKVYGV
ARRPCPAWLAKKPVEYIQCDVSNQETISKLSPLKDITHIFYVSWIGSEDCQTNATMFKNIL
NSVIPNASNLQHVCLQTGIKHYFGIFEEGSKVVP HDSPTEDLPRLNVQNFYHDLEDILYEE
TGKNNLTWSIHRPALVFGFSPCSMMNIVSTLCVYATICKHENKALVYPGSKNSWNCYADAVD
ADLVAEHEIWA AVDPKAKNQVLNCNNGDVFKWKHIWKKLAE EFGIEMVGYVEGKEQVSLAEL
MKDKDQVWDEIVKKNLVP TKLKEIAAFWFADIAFCSENLISSMNKSKELGFLGFRNSMKSF
VSCIDKMRDYRFIP-

6) 7-deoxyloganetic acid synthase (*Cr7DLS*)

>Vinca_734_Cr7DLS_1548bp

ATGGCGACCATCACTTTTCGATTCACTAAACCCGGTGACCGTCGCCATCTCAGCCGGTTTTCT
TCTCCTCCTTATCATCTTTGTGAAATCAAGAACTGGTTCATCCAAAAGAAAACCGCCGGGTC
CTCCAGGATGGCCGATTTTCGGCAATATGTTGACCTCGGAGATTTACCTCACCAAACGCTG
TACAAGCTCAAGTCAAATACGGGCCGATCGTATGGCTCCAGCTCGGCTCAATTAATACTAT
GGTTGTTCAAAATGCAGTTTCCGCCGCCGAGCTTTTTAAGAAACACGATGTCCCCTTTTGTG
ACCGAAAAGTCCCTGATACTCTCACTGCCTTCAATTTCAACCAGGGCTCCCTGGGTATGAAT
ACTTACGGCGGACATTGGCGAGTGCTTAGACGCCTCTGTTCCATGGAGTTTTTGGTAAATAA
ACGTATGAACGAAACTACAGATCTAAGAAGGAGAATTGAGGACAATATGGTCCGTTGGATTG

AAGAAGATTCATTGGCATCTAAAGCACAAAGGAGGAACCGGAGCGGTTCAATTGAGCAGGTTT
CTTTTTCTAATGGCGTTTAATTTGGTGGGTAACCTTATGTTATCGAGAGATTTAATGGATAA
TAAGGATCCAGAAGGAAGGGAATTTTTTGGATTGTATGAATGAGATATTGGAATTGGCTGGTA
CTCCTAATATTGCTGATTTTTTGCCTTTGTTGAAGAAATGGATCCACTTGAATGAAGAAA
AGGATGGTTGATAATATGTCTAGAACTATGAAAATTTCTTCAAAGTTTGTTCAGAAAAGACT
TGATAATAGAAAAGCTGGAAAGATTAATGAAAAGAAAGATTTCTTGGATGTTATGCTTGAAT
ACCAAGGTGATGGTAAAGATGGTCCCGATAAATTCACTGAGCAACATGTCAACATTGTCATC
ATGGAAATGTTTTTGCAGGATCAGAGACAACAAGTATCAGCATTGAATGGGGATTACACAGA
ACTTCTCCGAAACCCACACGCATTCAAGAAAGTAAGGGAAGAGATCGACAGAGTAGTGGGAG
TAAACAGAATGGTGAAGAGAGCGACATGGAGAATTTGCCGATTTGCAAGCAGTTGTAAAG
GAGACGCTTAGATTGCATCCGGCACTTCCAATGCTTCTACCAAGAAACACAATGGAAGACAC
TGAATACATGGGATACCTCATTCCTAAAGGCACACAAGTTTTCGTGAATGCATGGGCAATCG
GAAGGGATCCTGAGTACTGGCAGGACCCTTTGAGTTTCAAACCAGAAAGGTTTATCAATTCC
AGTGTGAATATAAAGGACAACATTTTGAATTGATTCTTTTGGTTCTGGAAGAAGAATTTG
TGTGGGATTTCTTTGGCTCATAGAGTTGTTTCATCTCACTTTGGCTACATTAGTCCAGGCTT
TTGATTGGGATCTTGGTGCTGGCGTTAAACCACAAGATATTGATTGGAAGAGAGATTGGGA
TTGACTCTGAGGAAGAAGAATCCATTGAATGTCATCCCTAAGAAGAGAGTTCATATCTAA

>Cr7DLS_Protein Sequence

MATITFDLSLNPVTVVAISAGFLLLLLIIFVKSRGTSSKRKPPGPPGWPIFGNMFDLGDLPHQTL
YKLKSKYGPVWLQLGSINTMVVQNAVSAEELFKKHDVPFCDRKVPDTLTAFFNFNQSLGMN
TYGGHWRVLRRLCSMEFLVNKRNETTDLRRRIEDNMVRWIEEDSLASKAQGGTGAVQLSRF
LFLMAFNLVGNLMLSRDLMDNKPDEGREFFDCMNEILELAGTPNIADFLPLLKLDPLGMKK
RMVDNMSRTMKISSKFVQERLDNRKAGKINEKKDFLDVMLEYQGDGKDPDKFTEQHVNI
MEMFFAGSETTSISIEWGFTELLRNPHAFKKVREEIDRVVGVNRMVEESDMENLPYLQAVVK
ETLRLHPALPMLLPRNTMEDTEYMGYLI PKGTQVFNAWAIGRDPEYWQDPLSFKPERFINS
SVEYKQHFELIPFGSGRRICVGFPLAHRVHLTLATLVQAFDWDLDGAGVKPQDIDLEERLG
LTLRKNPLNVI PKRVHI -

7) 7-deoxyloganetic acid glucosyltransferase (*CrDLGT*)

>Vinca_1690_CrDLGT_1449bp

ATGGGTTCTCAAGAAACAAATTTGCCTCCTCATGTTCTCATATTCCCTTTGCCAATTCAAGG
CCATGTGAACTCCATGCTCAGACTAGCAGAACTTCTCTGTCTAGCTGAATTAGATATCACTT
TTATTGTTTTAGAGTTCAGCCACAGTCGCTTAATTAAGCATACAAATGTTGCCTCCCGTTTT
GCCCCGTTACCCAGGATTTAGTTCAGCCCATATCCGATGGCCTTCCCGATGATCACCCAG
AGCAGGCGAAAGAGTCATGGATATTTTGCCTTCAACTAAAAATGTAACGGGACCCCTGTTTA
AGCAGATGATGGTTGAGAATAAGTGCTTTTCTTCTGCTACTCGAAGGCCGATTACTTGATA
ATCGCTGATGGGGTTTTGAGCTTTGCTGGGGATTTTCGCTCAAGAAAAGGAATCCCTCTTAT
TTACTTCAGAACTGTTTCTGCTTGTCTTTCTGGGCTTGTTTTTGTATGCCTGAACTAATTG
AATCTGGTGACATTCGGATTAAAGGAAATGGTATGGATCTGATAGTGAAAAGCGTGCCGGGG
ATGGAACTTTTCTCAGGCGACGAGATCTGCCGGGGTTTTGCCGAGTCAATGACATTAATGA
ACCGAACTCCAGATCCTGAAAACAGAGACAAGACAGACAACGAGAGCCCAAGCGGCTATTT
TGAACACTTTTGGAGATTTGGAAGGACCAATTCATCTCAAATTCGCAAACACATGCCAAGA
CTTTTACAATTGGACCAAGTCATTCACCTAACATCTAGACTTGAAACCAAGAATATCAA
GACTTTAATTTCTCCGGTAGTTTCTGGGAAGAAGATCGAAGCTGTGTGGATTGGCTTGATG
CACAGCCACCAAGATCTGTTTTATATGTAAGTTTTGGAAGTATTACAGTTGTAACAAGAGAC

CAACTCCTGGAATTCTGGTATGGTCTTGTAAATAGTGGACAACGGTCTTGTGGGTAATGAG
GCCAGATTCAATTATGGGCAAAGATGGGCAAAGCCAGATTCCTGCAGATCTTGAAGAGGGCA
CAAAGCCAGAGGTTATATGGTGGGATGGGCACCTCAAGAAGAGGTCTTGAATCACCCCTGCC
ATTGGAGGATTTCTGACTCACAGTGGGTGGAATTCACCTTAGAGAGTATTGTGGCTGGTGT
CCCAATGATATGTTGGCCATACTTTGCGGACCAAATGATAAACAGCAGATTTGTGAGTGAAA
TTTGAAGATTGGATTGGATATGAAAGATACATGTGATAGAGAGACAATTGTGAAGATGGTT
AGGGAGTTGATGGAGATTAGGAAAGATGAGTCTTGCAAAGGGCTGATCATATGGCTAAATT
GGCAAAGAGGGCTGTTAGTGAAGGCGGATCCTCATATTCTAACTTGGATGGCCTCGTTGATT
ACATAAAATCACTGATTATTTGA

>CrDLGT_Protein Sequence

MGSQETNLPPHVLIFPLPIQGHVNSMLRLAELLCLAELDITFIVSEFSHSRLIKHTNVASRF
ARYPGFQFQPI SDGLPDDHPRAGERVMDILPSTKNVTGPLFKQMMVENKCFSSATTRRPITCI
IADGVLSFAGDFAQEKGIPLIYFRTVSACSFWACFCMPELIESGDIPIKGNMDLIVKSVPG
METFLRRDLPGFCRVNDINEPKLQILKTETRQTTRAQAAILNTFEDLEGPILSQIRKHMPR
LFTIGPSHSHLTSRLETKNIKTLISSGSFWEEDRSCVDWLDAQPPRSVLYVSFGSITVVTRD
QLLEFWYGLVNSGQRFLWVMRPDSIMGKDGQSQIPADLEEGTKARGYMWGAPQEEVLNHPA
IGGFLTHSGWNSTLESIVAGVPMICWPYFADQMINSRFVSEIWKIGLDMKDTCDRETIVKMV
RELMEIRKDEFLQRADHMAKLAKEAVSEGGSSYSNLDGLVDYIKSLII-

8) 7-deoxyloganic acid hydroxylase (*Cr7DLH*)

>Vinca_2445_Cr7DLH_1556bp

ATGGAATTGAACTTCAAATCAATTATTTTCTTGGTTTTTGTAGTCTAACCCCTTTACTGGGT
CTATAGAATCTTGGATTGGGTATGGTTTTAAGCCCAAGAAGCTGGAAAAATGCTTAAGAGAAC
AAGGGTTTTAAGGGAAATCCTTACAGATTGTTCTTGGGAGATCAGTATGATAGTGGAAAAATTG
ATCAGACAAGCCTTGACTAAACCTATTGGTGTGAAGAAGATGTTAAGAAACGAATTGTACC
TCATATACTCAAACTGTTGGGACTCATGGCAAGAAGTCTTTTATGTGGGTGGAAGAATAC
CAAGGGTGAATATCACAGATCCAGAGCTGATTAAGAGGTCTTAACAAAATACTATAAGTTC
CAGAAGAATCATCATGATCTTGACCCAATTACCAAATTGCTTTTACTGGAATTGGAAGCTT
AGAAGGTGACCCTGGGCTAAACGCCGAAAAATCATCAATGCTGCTTTTACTTTGAAAAGT
TAAAGCTTATGCTTCCTGCATTCTATTTGAGTTGCCGTGATATGGTGACAAAATGGGATAAT
AAGGTCCTGAAGGAGGATCAGCAGAGGTGGATGTGTGGCATGATATTGAAACCTTAACTGG
AGATGTAATTTCAAGAACATTATTTGGAAGTAACTTCGAGGAAGGGAGAAGAATATTTGAAC
TCATGAAAGAACTTACTGCTCTTACTATTGATGTCATTCGCTCGGTTTATATCCCTGGACAG
AGGTTTCTTCCAATAAGAGGAACAATAGGATGAGAGCAATTGACAAAGAAGTTAGAGTAAG
AATAACAGAAATATCAATAAGAAAATGAAGGTAATGAAATCAGGAGAAGCAGCAAGTGCTG
CTGATGATTTTTTGGGAATATGTTGGAATGCAATCTGAATGAAATAAAGGAGCAAGGAAAT
AACAAAAGTGCTGGAATGACTATTGGAGAGATTATTGGTGAGTGCAAACCTTCTATTTTGC
TGGCCAGGACACTACCTCGACTTTACTCGTGTGGACAATGGTCCTCTTGTCTAGATTCCCCG
AATGGCAAACCTCGTGCCAGAGAAGAAGTTTTTCAAGTCTTTGGTAATAAAACCCCGGACTAT
GATGGTATCAGCCATCTTAAAGTTATAACGATGATCTTGTACGAGGTTTTGAGGTTATATAC
TCCAGTGGCTGAACTGACAAAAGTGGCTCATGAAGCTACACAACCTGGGAAAATACTTCATTC
CAGCGGGCGTGACGCTAATGATGCCACAAATCCTACTTCATCATGACCCTGAAATATGGGGT
GAAGATGTAATGGAGTTCAAACCAGAAAGATTTGCAGAGGGAGTTCTCAAGGCAACTAAGAG
CCAAGGCTCTTTTTTCCATTTAGCTTGGGACCAAGAATGTGCATTGGCCAAAACCTTTGCTC
TATTGGAAGCCAAAATGGCGATGTCTCTAATCTTGAGACGTTTTTCTTTCGAGCTTTCTCCC

TCCTATGTTTCATGCTCCTTTCACTCTCATCACAATGCAACCCCAATATGGTGCTCATTGAT
TCTGCACAAACTCTAA

>Cr7DLH_Protein Sequence

MELNFKSIIIFLVFVSLTLYWVYRILDWVWFKPKKLEKCLREQGFKGNPYRFLFLGDQYDSGKL
IRQALTKPIGVEEDVKKRIVPHILKTVGTHGKKSFMWVGRI PRVNITDPELIKEVLTKYYKF
QKNHHDLDPI TKLLLTGIGSLEGPWAKRRKI INAAFHFEEKLKLMLPAFYLS CRDMVTKWDN
KVPEGGSAEVDVWHDIETLTGDVISRTLFGSNFEEGRRI FELMKELTALTIDVIRSVYIPGQ
RFLPTKRNNRMRAIDKEVRVRITEI INKMKVMKSGEAAASAADDFLGILLEC NLNEIKEQGN
NKSAGMTIGEI IGECKLFYFAGQD TTSTLLVW TMVLLSRFPEWQTRAREEVFQVFGNKTPDY
DGI SHLKVITMILYEVLRLYTPVAELTKVAHEATQLGKYFI PAGVQLMMPQILLHHDPEIWG
EDVMEFKPERFAEGVLKATKSQGSFFPFS LGPRMCIGQNFALLEAKMAMSLILRRFSFELSP
SYVHAPFTLITMQPQYGAHLILHKL-

9) Loganic acid methyltransferase (*CrLAMT*)

>Vinca_917_CrLAMT_1116bp

ATGGTTGCCACAATTGATTCCATTGAAATGCCTGCTCTTCCAACAGCTGTGGAAGCCCACCC
AATGAAAGGTGGTGTGACTCCCACAGTTACTCCAAAATTC TTGCTACCAGAAAGGAGTAA
TTGATGCAGCCAAGGCAGTGATTGTTGAAGCAGTGAATGAAAAATTGGATCTTGAAAAATAAT
CCAATCTTTGATCCAATCAAACCTTTCCGCATTGCAGATTTTGGATGTTCAACAGGACCAAA
TACTTTCCATGCTATGCAAAACATAGTAGAATCAGTTGAAACAAAATACAAATCACTTCAA
AAACCCCTGAATCCATGTTTTCTTCAATGATCATGTTAACAATGATTTCAATGTTCTCTTT
AGATCCCTTCCACCAAACAGGGAATTTTTTGGCTGCTGGTGTTCCTGGATCTTTCTACTAG
AGTTTTTCCAAAAATAGCATTCATTTTTGCTCATTGTTCTTATGCACTTCATTGGTTATCTA
AAGTGCCCAAGGAAATCAAGATAAAAATTC TTGGCTTATAATAAGGGAAGAATTCATTAC
ACTGGTACTGAAAAACATGTTGTTAAGGCTTATTTTTGGTCAATTCCAAGAGATTTTGAAGG
GTTTTTAAAAGCAAGAGCTCAAGAAATTGTTGTTGGAGGGTTGATGGTGATTCAAATACCTG
GGCTTCCTAGTGGTGAAGTCTTTTCTCAAGGACTGGTGTGCTGGTTTGCTTCACTTCCCTTTG
GGAACCTCCTTGATGGAATTGGTTAACAAGGGAATCATCAATGAAGAAAGTGTAGATTCCTT
CAATTTGCCTCAATATCATCCCTCAGTTGAAGACTTGGAATGGTGATAGAAATGAACGATT
GTTTCACAATTGAAAGGGTTGGAAC TTTACCTCATCCCATGAAGAATTTACCATTTGATGTT
CAAAGGACTTCTTTACAAGTAAGAGCAATCATGGAATGCATTCTCACTGAACATTTTGGGGA
AAATATTTTGGATCCATTATTTGAAATCTACCCAAAAATTTGCAAGAGAATTTCCATGTTT
TTGATAAGGAAATTAGAAAGGATGCAGATTTGTACCTTGTCTTGAAACGCAAGGGAAATTA

>CrLAMT_Protein Sequence

MVATIDSIEMPALPTAVEAHPMKGGDDSHSYSQNSCYQKGVIDA AKAVIVEAVNEKLDLENN
PIFDPIKPFRIADFGCSTGPNTFHAMQNI VESVETKYKSLQKTP EFHVFFNDHVNND FNVLF
RSLPPNREFFAAGVPGSFYTRVFPKNSIHF AHCSYALHWLSKVPKEIQDKNSLAYNKGR IHY
TGTEKHVVKAYFGQFQRDFEGFLKARAQEI VVGGLMVIQIPGLPSGEVLF SRTGAGLLHFL L
GTSLMELVNKGI INEESVDSFNLPQYHPSVEDLEMVIEMNDCFTIERVGTLPHPMKNLPFDV
QRTSLQVRAIMECILTEHFGENILDPLFEIYTKNLQENFHVFDKEIRKDADLYLV LKRKGN

10) Secologanin synthase (*CrSLS*)

>Vinca_917_CrSLS_1116bp

ATGGAGATGGATATGGATACCATTAGAAAGGCAATTGCTGCCACTATTTTTGCATTGGTAAT
GGCTTGGGCATGGAGAGTGTGGATTGGGCATGGTTTACTCCTAAGAGGATCGAGAAACGTC
TAAGGCAGCAAGGTTTTAGAGGAAATCCTTATAGATTCTTGGTTGGAGATGTTAAGGAGAGT
GGAAAAATGCATCAAGAAGCCTTGTCTAAACCCATGGAGTTCACAATGATATTGTTCCCTCG
CCTCATGCCACATATTAACCACACTATCAATACTTACGGTAGGAATTCCCTTACATGGATGG
GAAGGATTCCAAGAATTCATGTTATGGAACCTGAACTTATTAAGGAAGTATTGACCCACTCA
AGCAAATACAAAAGAAGCTTTGATGTTTACAATCCCCTTGTTAAGTTCCCTTCTACCCGGAGT
TGGAAGCTTTGAGGGTGCAAAATGGTCAAAACACAGAAGAATTATTTCCCTGCCTTCACTC
TTGAGAAACTAAAGTCAATGCTGCCAGCTTTTGCCATATGCTACCATGACATGTTGACCAAA
TGGGAGAAAATAGCTGAAAAACAAGGATCCCATGAAGTTGATATCTTTCCACGTTTGATGT
TTTAACAAGTGATGTGATTTCAAAGGTTGCATTTGGTAGCACATATGAAGAAGGAGGCAAAA
TCTTCAGACTATTGAAAGAAGCTCATGGATCTCACAATTGACTGCATGAGAGATGTCTACATT
CCAGGATGGAGCTACTTGCCAACCAAGAGGAACAAGAGGATGAAAGAAAATTAACAAAGAGAT
CACAGATATGCTAAGGTTTATCATCAACAAGAGAATGAAGGCTTTGAAGGCTGGAGAGCCAG
GTGAGGATGACTTGCTGGGAGTATTGTTGGAATCAAACATTCAGAAAATTCAAAAGCAAGGA
AACAAGAAGGATGGTGAATGTCAATCAATGATGTAATTGAGGAGTGCAAATTGTTCTACTT
TGCTGGTCAAGAACTACTGGAGTTTTACTGACATGGACCACCATCTTATTGAGCAAGCACC
CTGAGTGGCAAGAACGAGCCAGAGAAGAAGTTCTCCAAGCCTTTGGCAAGAATAAACCTGAG
TTTGAACGCTTAAATCACCTCAAATATGTGTCTATGATCTTGTACGAGGTTCTAAGGTTGTA
CCCACCAGTGATTGATCTAACAAAGATTGTCCACAAGGACACAAAGTTAGGGTCGTACACAA
TCCCTGCAGGAACACAAGTGATGTTGCCGACAGTAATGCTTCACAGAGAGAAGAGCATTG
GGAGAAGATGCAATGGAATTCAACCAATGAGATTTGTTGATGGAGTTGCCAATGCAACCAA
GAACAATGTAACATATTTGCCATTGAGCTGGGGACCTAGGGTTTGTCTTGCCAAAACCTTG
CACTTCTGCAAGCAAATTAGGATTGGCAATGATTTTACAACGCTTCAAGTTTGTGTTGCT
CCATCCTATGTTTATGCTCCTTTTACTATTCTCACAGTTCAACCCAGTTTGTTTCTCATGT
CATCTACAAGAAGCTTGAGAGCTAG

>CrSLS_Protein Sequence

MEMDMDTIRKAI AATIFALVMAWAWRVLDWAWFTP KRIEKRLRQQGFRGNPYRFLVGDVKES
GKMHQEALSKPMEFNNDIVPRLMPHINHTINTYGRNSFTWMGRI PRIHVMEPELIKEVLTHS
SKYQKNFDVHNPLVKFLLTG VGSFEGAKWSKHRRII SPAFTLEKLSMLPAFAICYHDM LTK
WEKIAEKQGSHEVDIFPTFDVLTSDVISKVAFGSTYEEGGKIFRLKELMDLTI DCMRDVYI
PGWSYLPTRKRNKRMKEINKEITDMLRFI INKRMKALKAGEPGEDDLLGV LLESNIQEI QKQG
NKKDGGMSINDVIEECKLFYFAGQETTGVLLTWTTI LLSKHP EWQERAREEVLQAFGKNKPE
FERLNHLKYVSMILYEVLRLYPPVIDLTKIVHKDTKLGSYTI PAGTQVMLPTVMLHREKSIW
GEDAMEFNPMRFVDGVANATKNNVTYLPFSWGRVCLGQNFALLQAKLGLAMILQRFKFDVA
PSYVHAPFTILT VQPQFGSHVIYKKLES-

Thesis Summary

A wide array of secondary metabolites is produced by plant as a part of their defense system. Terpenes are a large class of secondary metabolites produced in plants, which are known to possess numerous biological properties. *Catharanthus roseus*, an evergreen shrub, is a rich source of a class of compounds called terpene indole alkaloids (TIAs), which are known for their pharmacological and therapeutic importance. The *Catharanthus* alkaloids comprise a group of numerous terpenoid indole alkaloids, of which vinblastine and vincristine have emerged as promising anti-cancer agents. These two compounds are synthesized via an intermediate, secologanin, the formation of which greatly depends on the formation of iridoids. This thesis describes the isolation and extensive characterization of the genes involved in the biosynthesis of iridoids in *Catharanthus roseus*, their scope for metabolic engineering and directed evolution of two important enzymes in the pathway.

Isolation of high quality total RNA from the various tissues of *Catharanthus roseus* was carried out after optimization of conventional and kit-based protocols. These total RNA samples were further used for RNA sequencing on Illumina GAI Analyzer and Fastq file of 4.88 GB was generated. The raw RNA-seq paired reads were deposited with NCBI (Accession ID: SRR1693842). The 15485359 high quality paired end raw reads were assembled into 76075 contigs with an optimized hash length of 49, largest average contig length of 548.5 bp and an N50 value of 881 bp. Functional annotation was carried out using KEGG database, KAAS server, NCBI Nr-database, SwissProt/ Uniprot database and Pfam database. These analyses finally led to the identification of the unigenes related to secologanin biosynthesis.

Using the data from transcriptome analysis, five genes involved in the biosynthesis of iridoids were identified as geranyl diphosphate synthase (CrGDS), geraniol synthase (CrGS) and geraniol 10-hydroxylase (CrG10H), 10-hydroxygeraniol dehydrogenase (Cr10HGO) and iridoid synthase (CrIDS) and their cloning and functional characterizations were carried out. Cr10HGO showed broad substrate specificity for 10-hydroxygeraniol, 10-oxogeraniol or 10-hydroxygeraniol over monohydroxy linear terpene derivatives. Concerted enzymatic function in the biosynthesis of *cis-trans*-nepetalactol has been demonstrated using 10-hydroxygeraniol and NADP⁺ with Cr10HGO and CrIDS combined assay system. The stereochemistry of the enzymatic product was determined as (1*R*, 4*aS*, 7*S*, 7*aR*)-

nepetalactol, which is a key intermediate in the biosynthesis of iridoids and MIAs. Further, we demonstrated the *in vitro* formation of (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol when geraniol was incubated with CrG10H, Cr10HGO and CrIDS.

Cr10HGO catalyzes the formation of 10-oxogeranial via two intermediates, 10-hydroxygeranial and 10-oxogeraniol, in the presence of NADP⁺. In order to establish the structure-function relationship, we identified 12 amino acids present in the active site pocket. Functional characterization of the 14 mutants generated from these amino acids suggests the crucial role in the catalysis of Cr10HGO. The product profiles of the 14 mutants suggested that, replacement of the highly conserved residues with smaller amino acids affects the interaction of the substrate and the protein, which in turn affects the activity of the protein. The results observed indicate that an increase in the hydrophobicity surrounding the active site pocket greatly favors the reaction in the forward direction, and hence enhances the percentage formation of the di-aldehyde, 10-oxogeranial. The kinetic parameters of these significant mutants were well in agreement with the characteristics of their product profile.

Iridoid synthase (CrIDS) is an NADPH-dependent enzyme, belonging to a class of monoterpene indole cyclases, which utilizes the linear 10-oxogeranial as a substrate and by a sequential reduction and cyclization, produces an equilibrium mixture of nepetalactols and iridoids, the final product in iridoid biosynthesis. The main aim of carrying out directed mutagenesis of CrIDS was to identify the amino acids controlling the substrate specificity of the enzyme. Studies of the activity profiles of the 10 CrIDS mutants constructed, led to the observation that an increase in polarity near the active site pocket favored the formation of *cis-trans*-nepetalactol / (1*R*, 4*aS*, 7*S*, 7*aR*)-nepetalactol more than normally.

Full-length unigenes showing matches with 7 deoxyloganin synthase (Cr7DLS), 7-deoxyloganic acid glucosyltransferase (CrDLGT), 7-deoxyloganin hydroxylase (7DLH), loganic acid methyltransferase (CrLAMT) and Secologanin synthase (CrSLS) were identified. Their cloning and expressions were carried out and this study gives scope for the extension for metabolic engineering of the ten genes of the secologanin biosynthetic pathway for expression in heterologous systems.

List of publications

1. **Ramakrishnan Krithika**, Prabhakar Lal Srivastava, Bajaj Rani, Swati P. Kolet, Manojkumar Chopade, Mantri Soniya, and Hirekodathakallu V. Thulasiram. "Characterization of 10-hydroxygeraniol dehydrogenase from *Catharanthus roseus* reveals cascaded enzymatic activity in iridoid biosynthesis." *Scientific reports* 5 (2014): 8258-8258.
2. Srivastava, Prabhakar Lal, Pankaj P. Daramwar, **Ramakrishnan Krithika**, Avinash Pandreka, S. Shiva Shankar, and Hirekodathakallu V. Thulasiram. "Functional characterization of novel sesquiterpene synthases from Indian sandalwood, *Santalum album*." *Scientific reports* 5 (2015).
3. Daramwar, Pankaj P., Raju Rincy, Siddiqui Niloferjahan, **Ramakrishnan Krithika**, Arvind Gulati, Amit Yadav, Rakesh Sharma, and Hirekodathakallu V. Thulasiram. "Transformation of (\pm)-lavandulol and (\pm)-tetrahydrolavandulol by a fungal strain *Rhizopus oryzae*." *Bioresource technology* 115 (2012): 70-74.
4. **Ramakrishnan Krithika** and Hirekodathakallu V. Thulasiram. "Dissection of the activity of *Catharanthus roseus* 10-hydroxygeraniol dehydrogenase (Cr10HGO) by site-directed mutagenesis studies." (**Manuscript under preparation**)
5. **Ramakrishnan Krithika** and Hirekodathakallu V. Thulasiram. "Structure-function studies of *Catharanthus roseus* Iridoid synthase through site-directed mutagenesis." (**Manuscript under preparation**)

Permanent Address

“Bharadwaj Apartments”, Flat No:1,
15, Hanumantha Road, Ground Floor,
Balaji Nagar, Royapettah,
Chennai-600 014. India.

Phone no: +919764207018, +919840352775

Email: krithikarhythm@gmail.com, krithika.ramakrishnan10@gmail.com

Education

Doctor of Philosophy (Ph.D.), Academic CSIR, National Chemical Laboratory, Pune

Title of Thesis: **Isolation and Characterization of the Genes Involved in the Biosynthesis of Secologanin, an Intermediate in the Formation of Vincristine, a Potent Anti-Cancer Agent**

Academic Record

Course	Institution	Board/ University	Year of Completion	Marks % / CGPA
Ph. D.	CSIR-National Chemical Laboratory, Pune	AcSIR, NEW DELHI	May 2017	
B.Tech. Biotechnology	PSG College of Technology, Coimbatore	ANNA UNIVERSITY, CHENNAI	2009	7.14
XII (Bio/Chem/Phy/Maths)	P.N.D. Adarsh Vidyalaya M.H.S. School, Chennai	STATE BOARD, TAMIL NADU	2005	91.6
X	P.N.D. Adarsh Vidyalaya M.H.S. School, Chennai	MATRICULATION	2003	88.55

Fields of Interest

- *Genetics*
- *Molecular Biology*
- *Genetic and Protein Engineering*

Research Experience

- **“Biodiversity of *aac6'-aph2*” Gene in different strains of *Enterococcus faecalis*,** June - October 2007, PSG College of Technology, Coimbatore, India.
- **“Protein-protein Interactions using Phylogenetic Analysis,”** June -October 2008, PSG College of Technology, Coimbatore, India,
- **“Modeling in *Enterococcus faecalis*,** June – October 2008, PSG College of Technology, Coimbatore, India.
- **“Characterisation of the Genes Involved in the Santalane Biosynthetic Pathway in the Indian Sandalwood *Santalum album*,** December 2008 - April 2009, CSIR-National Chemical Laboratory, Pune, India (B.Tech Dissertation)
- **“Biocatalysis of Artemisinin by Soil Microbe to give A Crystallized Non-Peroxide Derivative”,** June 2009-June 2011, CSIR-National Chemical Laboratory, Pune, India.

- **“Utilization of Oxidoreductases System of a Fungal Strain for Reductive Biotransformation of Aceto- & Benzo- Phenones to Chiral Alcohols”**, June 2009-June 2011, CSIR-National Chemical Laboratory, Pune, India.
- **“Santalene synthase: A comprehensive site directed mutagenesis study to understand the dynamic nature of sesquiterpene synthase active site pocket”**, July 2011-July 2012, CSIR-National Chemical Laboratory, Pune, India.
- **“Isolation and Characterization of the Genes Involved in the Biosynthesis of Secologanin, an Intermediate in the Formation of Vincristine, a Potent Anti-Cancer Agent”**, August 2012-December 2016, CSIR-National Chemical Laboratory, Pune, India (Ph. D.) (Supervisor: Dr. H. V. Thulasiram)
 - RNA Isolation, transcriptome sequencing and functional annotation
 - Cloning and functional characterization of 5 genes involved in Iridoid biosynthesis in *Catharanthus roseus*: 1 prenyl transferase-geranyl diphosphate synthase (CrGDS), two terpene cyclases – geraniol synthase (CrGS) and iridoid synthase (CrIDS), a cytochrome P450 hydroxylase- geraniol 10-hydroxylase (CrG10H) and an alcohol dehydrogenase – 10-hydroxygeraniol dehydrogenase (Cr10HGO)
 - Site-directed mutagenesis of Cr10HGO and CrIDS in order to understand the structure basis for the product profile of each
 - Cloning of three cytochrome P450 hydroxylases – 7-deoxyloganetic acid synthase (Cr7DLS), 7-deoxyloganin hydroxylase (Cr7DLH) and Secologanin synthase (CrSLS), a cytochrome P450 reductase (CrCPR), a glycosyltransferase – 7-deoxyloganetic glucosyltransferase (CrDLGT) and a methyltransferase – loganic acid methyltransferase (CrLAMT) involved in the biosynthesis of secologanin from iridoids

Research Expertise

Molecular Biology

- Isolation of RNA/DNA from various plant tissues such wood tissue of Sandalwood, leaves, roots and stems, fungal systems, etc.
- Primer designing, cDNA preparation and PCR amplification
- Plasmid isolation, gel elution, PCR purification and restriction digestion reactions
- Primer designing for site directed mutagenesis, mutation studies and screening of mutants using sequencing
- cDNA library preparation and screening
- Gene cloning in bacterial and yeast vectors for protein expression
- Cloning of multiple genes in dual vectors and *in vivo* production of metabolites in bacteria.

Downstream Processing

- Protein expression, homogenization, microsome preparation, high speed and ultra-centrifugation
- Protein purification using chromatographic techniques such as affinity purification, gel filtration, ion exchange chromatography using AKTA Avant and manual procedure
- Southern, Northern and Western blotting for the detection of DNA, RNA and protein
- Enzymatic characterization of purified enzymes and kinetic characterization

Analytical Techniques and Instrumentation

- PCR, Real-time PCR
- Gas Chromatography/GC-MS analysis, GC Qtof analysis

- UV-spectrophotometer – Cary 300 for Kinetic characterizations
- Chemiluminescence and phosphorescence imaging system
- High speed and Ultra centrifuges
- AKTA Avant
- Fermentation

Organic Chemistry Techniques

- Biotransformation of various organic compounds using microbial systems
- Extraction of metabolites from various systems such as plants, microbes, etc.
- Thin layer chromatography (TLC)
- Silica gel column chromatography

Software Skills

- Bioinformatics tools: Sequence analysis
- Transcriptome analysis using Virtual ribosome translation, Pfam domain analysis, KAAS-KEGG analysis, BLAST2GO analysis, BLAST analysis
- Multiple sequence alignment using ClustalW, Boxshade analysis, phylogenetic analysis
- Schrodinger for homology modeling of protein, PyMol for visualization
- MS Office, Chemdraw, Bioedit, Graphpad prism, Sigmaplot, Endnote

Conferences and Symposia

- International symposium on “**Proteomics beyond ID’s**”, CSIR-NCL, Pune, November, 2012.
- “**Indo-Mexico workshop on Biotechnology: Beyond borders**” organized by CSIR-NCL, Pune and CINVESTAV, Mexico, October, 2013.
- “**ISBOC-2015**, IISER, Pune, January 2015.

Awards and Honors

- “**Best Poster Award**”, “**Screening and Utilization of Oxidoreductase Systems for Organic Synthesis**”, National Science Day, CSIR-NCL, Pune, February 2011.
- “**Best Poster Award**”, “**Isolation and Characterization of genes involved in Biosynthesis of Secologanin in Vinca rosea**”, “Indo-Mexico workshop on Biotechnology: Beyond borders” CSIR-NCL, Pune, October 2013.
- “**Best Poster Award**”, “**Studies on genes involved in Iridoid Biosynthesis in Catharanthus roseus**”, ISBOC - International Symposium on Bioorganic Chemistry-2015, IISER, Pune, January 2015.
- “**Best Poster Award**”, “**Studies on genes involved in Iridoid Biosynthesis in Catharanthus roseus**”, National Science Day, CSIR-NCL, Pune, February 2015.
- **Lady Tata Memorial Trust Fellowship for Junior and Senior Research Scholar**, 2012 and 2014, respectively.
- **Dr. Prema Krishnaswami Award for Highest Scorer in Science** in school, Matriculation Examination, 2003.
- “**Best Student**” Award, LION’S CLUB, 2003.

Research Publications

- **Ramakrishnan Krithika**, Prabhakar Lal Srivastava, Bajaj Rani, Swati P. Kolet, Manojkumar Chopade, Mantri Soniya, and Hirekodathakallu V. Thulasiram. "Characterization of 10-hydroxygeraniol dehydrogenase from *Catharanthus roseus* reveals cascaded enzymatic activity in iridoid biosynthesis." *Scientific reports* 5 (2015): 8258-8258. DOI:10.1038/srep08258.
- Srivastava, Prabhakar Lal, Pankaj P. Daramwar, **Ramakrishnan Krithika**, Avinash Pandreka, S. Shiva Shankar, and Hirekodathakallu V. Thulasiram. "Functional characterization of novel sesquiterpene synthases from Indian sandalwood, *Santalum album*." *Scientific reports* 5 (2015): 10095-10107. DOI: 10.1038/srep10095.
- Daramwar, Pankaj P., Raju Rincy, Siddiqui Niloferjahan, **Ramakrishnan Krithika**, Arvind Gulati, Amit Yadav, Rakesh Sharma, and Hirekodathakallu V. Thulasiram. "Transformation of (±)-lavandulol and (±)-tetrahydrolavandulol by a fungal strain *Rhizopus oryzae*." *Bioresource technology* 115 (2012): 70-74. DOI: 10.1016/j.biortech.2011.11.038.
- **Ramakrishnan Krithika** and Hirekodathakallu V. Thulasiram. "Dissection of the activity of *Catharanthus roseus* 10-hydroxygeraniol dehydrogenase (Cr10HGO) by site-directed mutagenesis studies." (**Manuscript under preparation**)
- **Ramakrishnan Krithika** and Hirekodathakallu V. Thulasiram. "Structure-function studies of *Catharanthus roseus* iridoid synthase through site-directed mutagenesis." (**Manuscript under preparation**)

Languages

Read: English, Hindi and Tamil

Write: English, Hindi and Tamil

Speak: English, Hindi, Tamil and Bengali

Declaration

I hereby declare that the above mentioned details are true to the best of my knowledge and belief.

Krithika Ramakrishnan