

MOLECULAR CHARACTERIZATION OF
FRAGRANCE-RELATED TRANSCRIPTS IN
BENZENOID/PHENYLPROPANOID PATHWAY
OF *VANDA MIMI PALMER*

MOHD AIMAN BIN BARUDIN

MASTER OF SCIENCE (BIOTECHNOLOGY)

UNIVERSITI MALAYSIA PAHANG

**MOLECULAR CHARACTERIZATION OF FRAGRANCE-RELATED
TRANSCRIPTS IN BENZENOID/PHENYLPROPANOID PATHWAY OF VANDA
MIMI PALMER**



MOHD AIMAN BIN BARUDIN

**Thesis submitted in fulfilment of the requirements for the award of the degree of
Master of Science in Biotechnology (by Research)**

**Faculty of Industrial Sciences and Technology
UNIVERSITI MALAYSIA PAHANG**

MAY 2016

THESIS STATUS VALIDATION FORM

UNIVERSITI MALAYSIA PAHANG

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name : _____

Date of Birth : _____

Title : _____

Academic Session : _____

I declare that this thesis is classified as:

CONFIDENTIAL

(Contains confidential information under the Official Secret Act 1997)*

RESTRICTED

(Contains restricted information as specified by the organization where research was done)*

OPEN ACCESS

I agree that my thesis to be published as online open access (Full Text)

I acknowledge that Universiti Malaysia Pahang reserve the right as follows:

1. The Thesis is the Property of Universiti Malaysia Pahang
2. The Library of Universiti Malaysia Pahang has the right to make copies for the purpose of research only.
3. The Library has the right to make copies of the thesis for academic exchange.

Certified By: _____

(Student's Signature)

(Supervisor's Signature)

New IC/Passport Number
Date:

Name of Supervisor

Date

EXAMINERS APPROVAL DOCUMENT
UNIVERSITI MALAYSIA PAHANG
INSTITUTE OF POSTGRADUATE STUDIES

We certify that the thesis entitled Molecular Characterization of Fragrance Related-transcripts in Benzenoid/Phenylpropanoid Pathway of *Vanda* Mimi Palmer, is written by Mohd Aiman bin Barudin. We have examined the final copy of this thesis and in our opinion; it is fully adequate in terms of scope and quality for the award of the degree of Master of Science in Biotechnology. We herewith recommend that it be accepted in fulfilment of the requirements for the degree of Master of Science specializing in Biotechnology.

Name of External Examiner : Assoc. Prof. Dr. Roohaida Othman

Institution : Universiti Kebangsaan Malaysia

Name of Internal Examiner : Dr. Essam Abdellatif Makky Saleh

Institution : Universiti Malaysia Pahang

SUPERVISOR'S DECLARATION

We hereby declare that we have checked this thesis and in our opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Master of Science in Biotechnology.

(Supervisor's Signature)

Name of Supervisor : Dr Tan Suat Hian

Position : Senior Lecturer

Date :

(Co-supervisor's Signature)

Name of Co-supervisor : Mr. Mohd Hairul Ab Rahim

Position : Lecturer

Date :



STUDENT'S DECLARATION

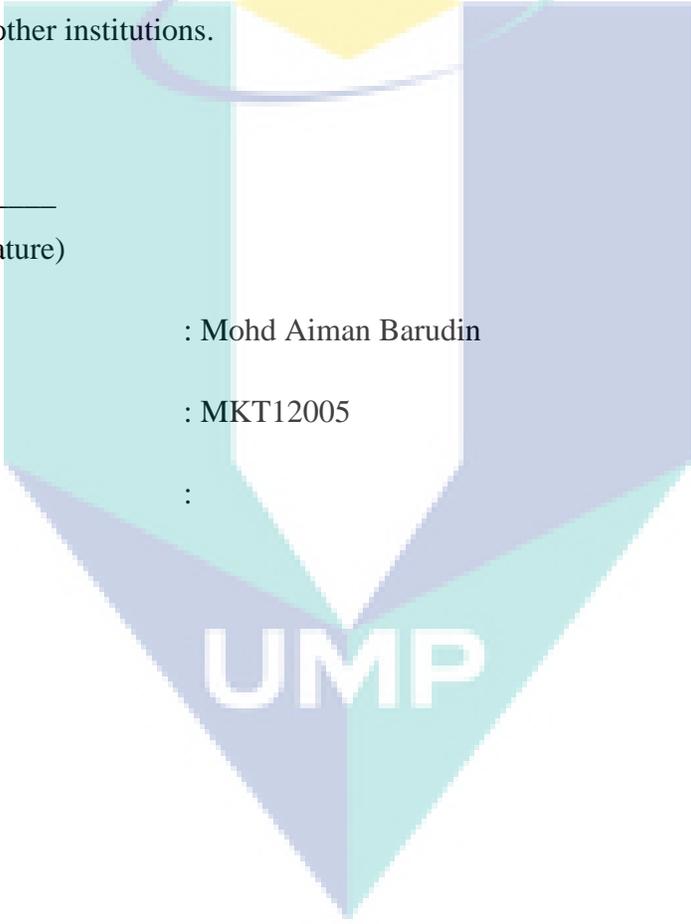
I hereby declare that the work in this thesis is based on my own original work except for quotations and citation which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

(Author's Signature)

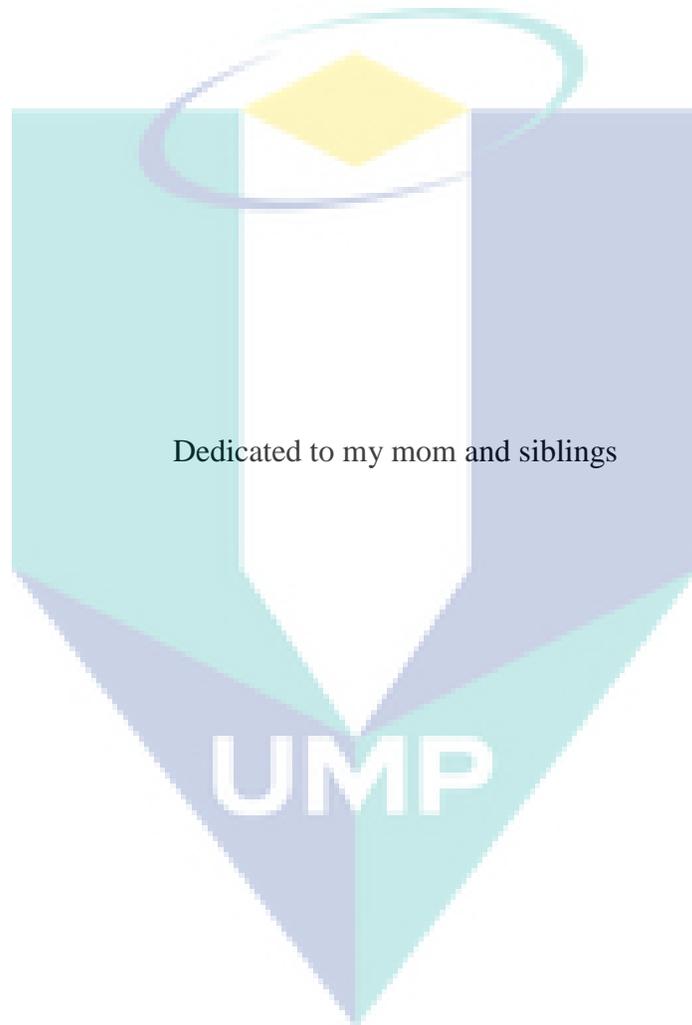
Full Name : Mohd Aiman Barudin

ID Number : MKT12005

Date :



UMP



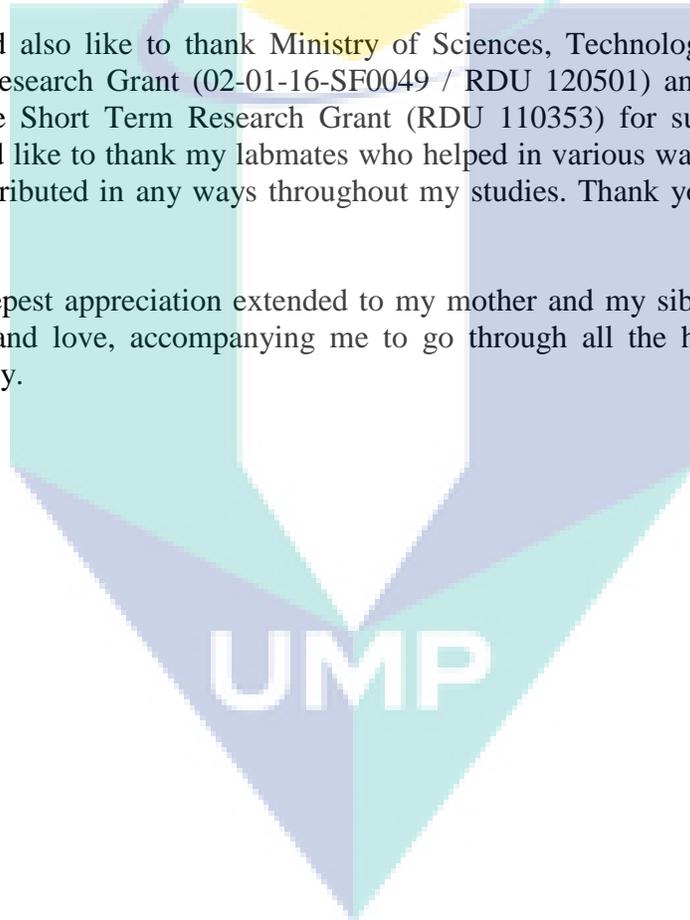
ACKNOWLEDGEMENTS

First and foremost, I would like to thank Allah for giving me strength and perseverance, for bringing me to this path and leading me through.

I would like to express my deep appreciation to my supervisor, Dr Tan Suat Hian and co-supervisor, Mr Mohd Hairul Ab. Rahim, for their invaluable guidance, advice, and mentorship. I am grateful for many opportunities that have been given to me to explore and persevere in the scientific research. Thank you for introducing me to the wonderful and exciting world of plant biotechnology.

I would also like to thank Ministry of Sciences, Technology and Innovation for Sciencefund Research Grant (02-01-16-SF0049 / RDU 120501) and Universiti Malaysia Pahang for the Short Term Research Grant (RDU 110353) for supporting this project. Lastly, I would like to thank my labmates who helped in various ways as well as to others who have contributed in any ways throughout my studies. Thank you from the bottom of my heart.

My deepest appreciation extended to my mother and my siblings for their endless support, care and love, accompanying me to go through all the happiness and sadness during my study.



ABSTRACT

Vanda Mimi Palmer (VMP) is a scented orchid that has received several international recognitions for its strong-sweet fragrance containing high percentage of benzenoids as well as phenylpropanoid compounds. Unfortunately, knowledge on the biosynthesis of benzenoids and phenylpropanoid compounds is still not fully-understood especially in Vandaceous orchid. Thus, this study was aimed to identify, isolate and molecular characterise potential transcripts that might be involved in the biosynthesis of benzenoids and phenylpropanoids in VMP. Screening of fragrance-related transcripts via cDNA-RDA approach has shown identification of partial sequences of phenylalanine ammonia lyase (VMPPAL) and S-adenosyl-L-methionine synthase (VMPSAMS) from VMP besides methionine synthases from *Vanda* Small Boy Leong (VSBLMS) and *Vandachostylis* Sri-Siam (VSSSMS). However, those transcripts were not further characterised due to their indirect involvement in final fragrance compounds biosynthesis. Thus, previously identified partial transcripts from floral cDNA library of VMP were isolated including 3-keto-acyl CoA thiolase (VMPKAT), benzoic acid/salicylic acid carboxyl methyltransferase (VMPBSMT), eugenol synthase (VMPEGS) and orcinol-O-methyltransferase (VMPOOMT), which resulted in the isolation of full ORF encoding 450, 378, 306 and 368 amino acids, respectively. All the transcripts were further characterised using molecular and bioinformatics approaches including domains identification, comparative modeling, molecular docking for substrates and ligands prediction as well as gene expression analysis using real-time RT-PCR. In this study, 3D-structural model was developed on those fragrance-related proteins for comparative modeling and molecular docking. Molecular docking on those proteins has shown their suitability towards specific substrates and ligands, reflecting their homologous putative functionality to fragrance-related proteins from other well-studied scented plants. However, gene expression analysis using real-time RT-PCR has shown that VMPKAT and VMPBSMT might be involved in benzenoids biosynthesis due to their up-regulated expression in floral tissues as well as developmentally-regulated expression pattern. Meanwhile for VMPEGS and VMPOOMT genes, their gene expression pattern might suggest their involvement as intermediates for biosynthesis of vanillin or other non-fragrance benzenoid and phenylpropanoid compounds, respectively. In conclusion, VMPKAT and VMPBSMT might be involved in fragrance benzenoid biosynthesis based on their functional catalytic prediction, up-regulated gene expression in floral compared to vegetative tissues and their developmentally-regulated expression pattern with the highest transcripts in fully-open flower.

ABSTRAK

Vanda Mimi Palmer (VMP) adalah orkid wangi yang menerima pengiktirafan antarabangsa kerana keharumannya yang merangkumi sebatian benzenoid dan fenilpropanoid. Namun begitu, pengetahuan mengenai biosintesis sebatian benzenoid dan phenylpropanoid masih tidak difahaminya sepenuhnya, terutamanya bagi orkid 'Vandaceous'. Maka, kajian ini adalah bertujuan bagi mengenalpasti, memencil dan mencirikan jujukan DNA yang terlibat dalam biosintesis sebatian benzenoid dan phenylpropanoid dalam *Vanda* Mimi Palmer. Melalui penyaringan jujukan DNA yang berkaitan dengan penghasilan wangian menggunakan kaedah cDNA-RDA, jujukan DNA separa bagi phenylalanine ammonia lyase (VMPPAL) dan S-adenosyl-L-methionine synthase (VMPSAMS) telah dipencilkan daripada *Vanda* Mimi Palmer selain methionine synthases daripada *Vanda* Small Boy Leong (VSBLMS) dan *Vandachostylis* Sri-Siam (VSSSMS). Walau bagaimanapun, jujukan-jujukan DNA tersebut tidak dipilih bagi pencirian selanjutnya disebabkan penglibatan yang secara tidak langsung didalam biosintesis sebatian wangian akhir. Justeru, jujukan separa yang sebelum ini telah dikenalpasti melalui perpustakaan-cDNA *Vanda* Mimi Palmer dipilih bagi pemencilan jujukan rangka terbuka (ORF) termasuk 3-keto-acyl CoA thiolase (VMPKAT), benzoic acid/salicylic acid carboxyl methyltransferase (VMPBSMT), eugenol synthase (VMPEGS) dan orcinol-O-methyltransferase (VMPOOMT) yang mengekod 450, 378, 306 dan 368 asid amino. Seterusnya, pencirian ORF jujukan DNA tersebut dijalankan melalui kaedah molekular dan bioinformatik merangkumi pengenalanpastian domain, perbandingan model protin, 'molecular docking' bagi pengenalanpastian substrat dan ligan, disamping analisis gen ekspresi menggunakan 'Real-time RT-PCR'. Dalam kajian ini, model 3-dimensi telah dihasilkan bagi kesemua protin yang dikodkan oleh jujukan-jujukan DNA tersebut bagi tujuan perbandingan model dan 'molecular docking'. Analisis 'molecular docking' bagi protin-protin tersebut menunjukkan kesesuaian terhadap substrat dan ligan yang spesifik. Ini menunjukkan protin-protin tersebut berpotensi bagi menjalankan fungsi yang sama dengan homolog protin daripada tumbuhan wangi yang lain. Walaubagaimanapun, analisis gen ekspresi menggunakan 'Real-time RT-PCR' menunjukkan hanya VMPKAT dan VMPBSMT berkemungkinan terlibat dalam biosintesis benzenoid disebabkan oleh ekspresi gen yang tinggi dalam tisu bunga disamping gen ekperesi yang meningkat bersama peningkatan peringkat pertumbuhan bunga. Manakala, VMPEGS dan VMPOOMT berkemungkinan terlibat sebagai perantara dalam biosintesis vanillin atau sebatian-sebatian lain berdasarkan kepada corak gen ekspresi kedua-duanya. Kesimpulannya, VMPKAT dan VMPBSMT berkemungkinan terlibat dalam biosintesis wangian benzenoid disebabkan fungsi pemangkinannya dan juga gen ekspresi yang tinggi dalam tisu bunga berbanding tisu vegetatif disamping peningkatan gen ekspresi dalam setiap peringkat pertumbuhan bunga.

TABLE OF CONTENTS

	Page
DECLARATION	
TITLE PAGE	
SUPERVISOR'S DECLARATION	iii
STUDENT'S DECLARATION	iv
DEDICATION	v
ACKNOWLEDGEMENTS	vii
ABSTRACT	viii
ABSTRAK	ix
TABLE OF CONTENTS	x-xiv
LIST OF TABLES	xv-xvi
LIST OF FIGURES	xvii-xix
LIST OF SYMBOLS	xx
LIST OF ABBREVIATIONS	xxi-xxx
CHAPTER 1 INTRODUCTION	
1.0 Introduction	1
1.1 Problem Statement	4
1.2 Research Objectives	5
1.3 Scope of Study	5

CHAPTER 2 LITERATURE REVIEW

2.1	Orchids: An Introduction	7
2.2	Vandaceous Orchids	9
2.2.1	<i>Vanda</i> Mimi Palmer	11
2.2.2	<i>Vanda</i> Tan Chay Yan	13
2.2.3	<i>Vanda</i> Small Boy Leong	13
2.2.4	<i>Vandachostylis</i> Sri-Siam	14
2.3	Biology of Floral Scent	14
2.4	Fragrance Biosynthesis Mechanisms	16
2.5	Benzenoid and Phenylpropanoid Biosynthetic Pathway	18
2.6	Scientific Studies on Fragrant-Orchids	22
2.7	Molecular and Biochemical Studies on the Scent of <i>Vanda</i> Mimi Palmer	24
2.8	The Economic Importance of Fragrant-orchids	25

CHAPTER 3 MATERIALS AND METHODS

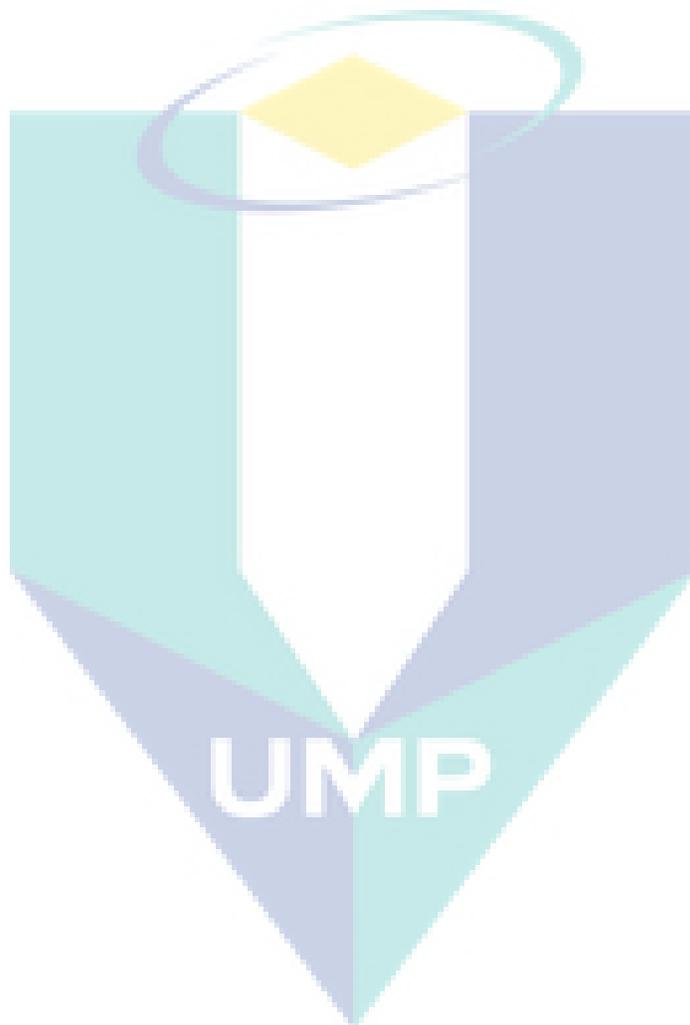
3.1	Plant Materials	27
3.2	Isolation of Total RNA	27
3.2.1	Formaldehyde Denaturing Agarose Gel Electrophoresis	28
3.2.2	Isolation of PolyA ⁺ mRNA	29
3.3	Synthesis of Double-stranded cDNA	30
3.4	cDNA Representational Difference Analysis (cDNA-RDA)	32
3.4.1	Digestion of Testers and Driver cDNA	32
3.4.2	cDNA-RDA, Oligonucleotide Adaptors and Primers	36
3.4.3	Ligation of Oligonucleotide Adaptors	38
3.4.4	Enrichment of Testers and Driver cDNA	39
3.4.5	Removal of Oligonucleotide Adaptors	40
3.4.6	First Round of cDNA-RDA	41
3.4.7	Subsequent Rounds of cDNA-RDA	42
3.5	Preparation of DH5 α <i>Escherichia coli</i> Competent Cells	42

3.6	Purification of Final cDNA-RDA Products	43
3.7	Cloning and Transformation of Purified cDNA-RDA Final Products	44
3.8	Plasmid Mini Preparation	46
3.9	Sequencing and Sequence Analysis of cDNA-RDA Products	47
3.10	Isolation of Full Length Fragrance-related Transcripts RACE-PCR	48
3.10.1	Designing Gene Specific Primers for Full length Isolation	51
3.10.2	Sequence analysis of full length sequences	51
3.11	Development of 3-Dimensional Structural Model of Fragrance-related Proteins	52
3.12	Functional Prediction on Fragrance-related Proteins	54
3.13	Expression Studies of Fragrance-related Genes by Real-time RT-PCR	55
CHAPTER 4	RESULTS AND DISCUSSION	
4.1	Identification of Fragrance-related Transcripts via cDNA-Representational Difference Analysis (cDNA-RDA)	59
4.2	Molecular characterisation of Full Length cDNA sequences of selected transcripts in benzenoid and phenylpropanoid pathway of <i>Vanda Mimi Palmer</i>	67
4.2.1	Sequence Analysis on Full length sequence of <i>Vanda Mimi Palmer</i> Keto-acyl CoA Thiolase (VMPKAT)	68
4.2.2	Sequence Analysis on Full length sequence of <i>Vanda Mimi Palmer</i> Benzoic Acid/Salicylic Acid Carboxyl methyltransferase (VMP BSMT)	75
4.2.3	Sequence Analysis on Full length sequence of <i>Vanda Mimi Palmer</i> Eugenol Synthase (VMPEGS)	79
4.2.4	Sequence Analysis on Full length sequence of <i>Vanda Mimi Palmer</i> Orcinol-O-methyltransferase (VMPOOMT)	87
4.3	Development of Three-Dimensional Structural Model of Fragrance-related Proteins for Their Structural and Functional Prediction	91
4.3.1	Structural and Functional Prediction of VMPKAT protein	94

4.3.2	Structural and Functional Prediction of VMPBSMT Protein	110
4.3.3	Structural and Functional Prediction of VMPEGS Protein	124
4.3.4	Structural and Functional Prediction of VMPOOMT Protein	140
4.4	Relative Expression Analysis of Fragrance-related Genes of <i>Vanda Mimi Palmer</i>	152
4.4.1	Relative Expression analysis of <i>Vanda Mimi Palmer</i> 3-Ketoacyl CoA Thiolase (VMPKAT)	154
4.4.2	Relative Expression analysis of <i>Vanda Mimi Palmer</i> Benzoic Acid/Salicylic Acid Carboxyl Methyltransferase (VMPBSMT)	156
4.4.3	Relative Expression analysis of <i>Vanda Mimi Palmer</i> Eugenol Synthase (VMPEGS)	158
4.4.4	Relative Expression analysis of <i>Vanda Mimi Palmer</i> Orcinol O-methyltransferase (VMPOOMT)	161
CHAPTER 5	CONCLUSIONS AND RECOMMENDATIONS	
	Conclusions and recommendations	164
	REFERENCES	167
	APPENDICES	184
A	General Solutions	184
B	List of Chemical Components and Its Compositions	185
C	pGEM®-T Easy Vector Map and Sequence Reference Points	186
D	List of Selected Full-length ORF Sequences of <i>Vanda Mimi Palmer</i>	187
E	List of Amino Acid sequences in <i>Vanda Mimi Palmer</i>	189
F	List of Primers for Fragrance-Related Transcripts used in RACE-PCR	190
G	List of Transcripts isolated via cDNA-RDA	192
H	Nucleotide and Deduced Amino Acid Sequences of full ORFs	195
I	Supplementary Data for Structural and Functional Prediction on Fragrance-related transcripts	203

J	Equations Used in Expression Analysis of the Putative Fragrance-related Transcripts	217
---	---	-----

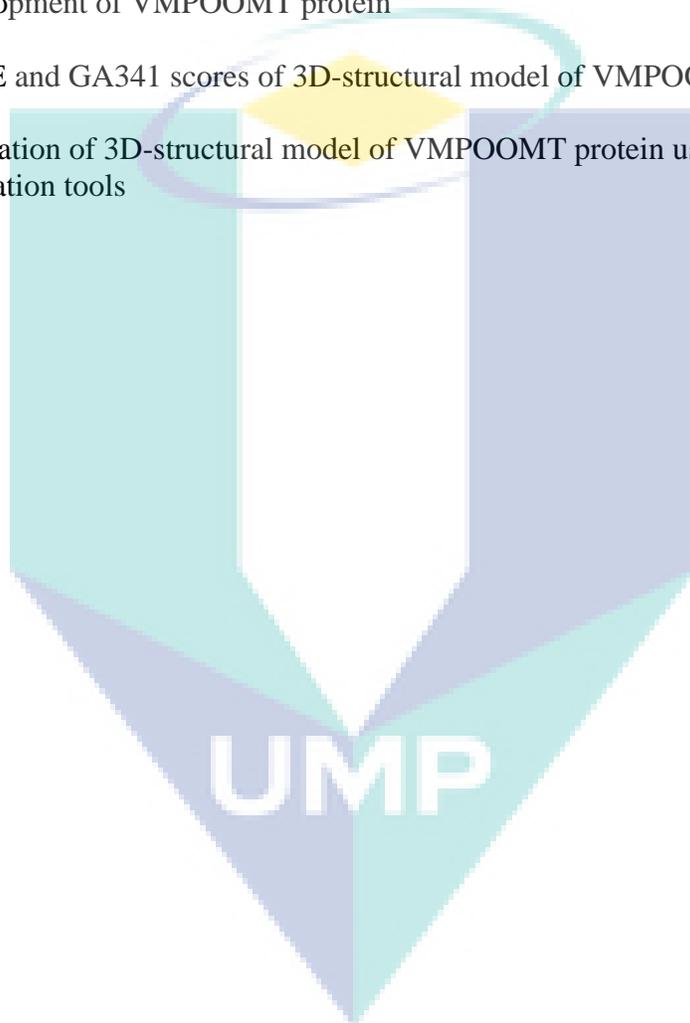
	LIST OF PUBLICATIONS	218
--	-----------------------------	-----



LIST OF TABLES

Table	Title	Page
3.1	Three sets of cDNA-RDA for identification of fragrance-related transcripts	34
3.2	Sequence of oligonucleotide adaptors for cDNA-RDA	37
3.3	Gene Specific Primers are used for 5'-regions, 3'-regions and Full Open Reading Frame (ORF) of Transcripts Isolation from <i>Vanda Mimi Palmer</i>	50
3.4	Primers Sequences for gene expression analysis using Real-time RT PCR	57
4.1	BLASTX analysis on final difference cDNA-RDA products	62
4.2	List of suitable protein templates to be used in 3D-structural model development of VMPKAT protein	95
4.3	DOPE and GA341 scores of 3D-structural model of VMPKAT protein	99
4.4	Evaluation on 3D-structural model of VMPKAT protein using multiple protein evaluation tools	99
4.5	Comparison of suitable model for development of 3D-structural model of VMPKAT protein using MODELLER version 9.14	101
4.6	Docking analysis of various putative substrates with fragrance-related proteins of <i>Vanda Mimi Palmer</i> using Autodock 4.2 software	107
4.7	Molecular docking of various putative ligands with fragrance-related proteins of <i>Vanda Mimi Palmer</i> using Autodock 4.2 software	107
4.8	List of suitable protein templates to be used in 3D-structural model development of VMPBSMT protein	111
4.9	DOPE and GA341 scores of 3D-structural model of VMPBSMT protein	114
4.10	Evaluation on 3D-structural model of VMPBSMT protein using protein evaluation tools	114
4.11	List of suitable protein templates to be used in development of 3D-structural model of VMPEGS protein	126

4.12	Comparison of suitable model for development of 3D-structural model of VMPEGS protein using MODELLER version 9.14	128
4.13	DOPE and GA341 scores of 3D-structural model of VMPEGS protein	128
4.14	Evaluation of 3D-structural model of VMPEGS protein using protein evaluation tools	133
4.15	List of suitable protein template to be used in 3D-structural model development of VMPOOMT protein	141
4.16	DOPE and GA341 scores of 3D-structural model of VMPOOMT protein	144
4.17	Evaluation of 3D-structural model of VMPOOMT protein using protein evaluation tools	144



LIST OF FIGURES

Table	Title	Page
2.1	Orchids used in this study	12
2.2	Benzenoid and Phenylpropanoid Biosynthetic Pathway in well-studied scented flowers	19
3.1	Flow chart of cDNA-RDA in this study	35
4.1	Difference products in three different rounds of cDNA-RDA	60
4.2	Postulation of benzenoid and phenylpropanoid pathway of Vandaceous orchid	66
4.3	Amplified PCR products of full ORF of selected fragrance-related transcripts	69
4.4	Phylogenetic tree of VMPKAT with homologous proteins	71
4.5	Alignment of VMPKAT amino acid sequence with closely related proteins from other plants	73
4.6	Alignment of VMPBSMT amino acid sequence with other proteins	76
4.7	Phylogenetic tree of VMPBSMT with homologous proteins	78
4.8	Alignment of amino acid sequences of VMPEGS with other closely related protein sequences from the NCBI GenBank database	81
4.9	Phylogenetic tree of VMPEGS with homologous proteins	83
4.10	Alignment of VMPEGS amino acid sequence with other eugenol synthases from <i>Gymnadenia</i> species	85
4.11	Phylogenetic tree of VMPEGS with eugenol synthases from <i>Gymnadenia</i> species	86
4.12	Alignment of VMPOOMT amino acid sequence with other closely related proteins that are retrieved from the NCBI GenBank database	88
4.13	Phylogenetic tree of VMPOOMT with homologous proteins	89
4.14	Secondary structure alignment between amino acid residue of	97

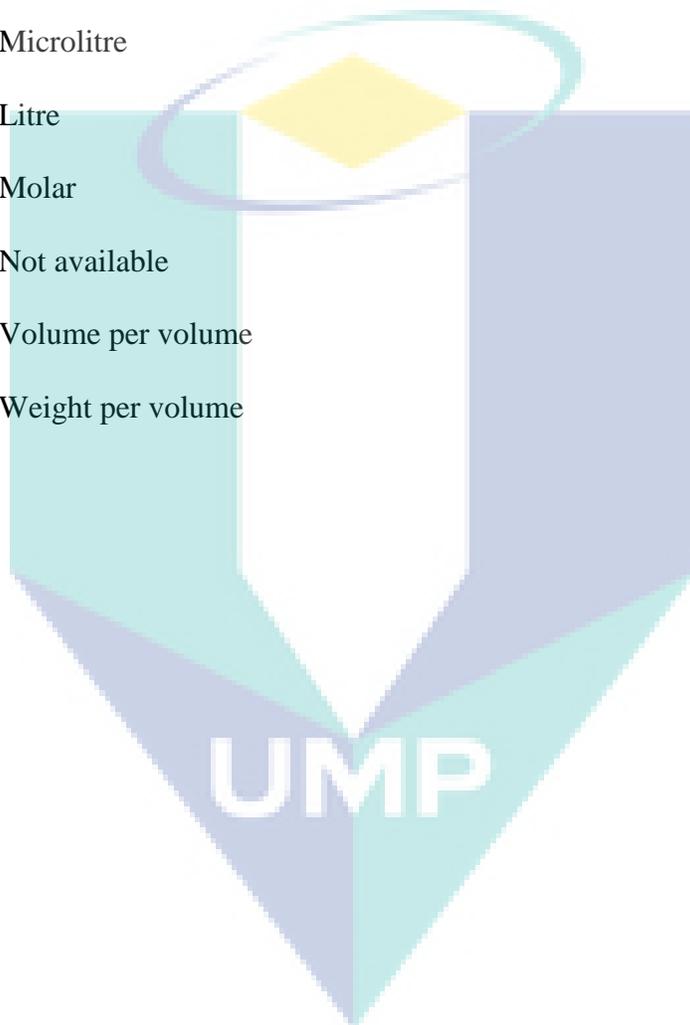
VMPKAT and 2wu9 template of 3-ketoacyl-CoA thiolase of
Arabidopsis thaliana

4.15	3D-structural model of VMPKAT protein	101
4.16	Comparative modeling of thiolase models	102
4.17	Predicted binding position (Cys 135, His 390, Cys 422) in thiolase active site domain of VMPKAT protein.	104
4.18	Superimposition of the 3D-structural model of VMPKAT protein	104
4.19	Involvement of 3-ketoacyl CoA thiolase (<i>PhKAT</i>) in benzenoid biosynthetic pathway of <i>Petunia hybrida</i>	105
4.20	Molecular docking of substrates and ligand with VMPKAT protein	108
4.21	Molecular docking of VMPKAT model with its substrate and ligand	109
4.22	Secondary structure alignment between amino acid residue of VMPBSMT and 1m6e template of <i>Clarkia breweri</i>	113
4.23	3D-structural model of VMPBSMT protein with labelled β strands (S1–S7) and α helices (H1–H10)	113
4.24	Comparative modeling of selected reference protein models to VMPBSMT	116
4.25	SAM binding sites of VMPBSMT model and other related models	117
4.26	Superimposition of VMPBSMT with chain X of 1m6e and other models	119
4.27	Molecular docking of VMPBSMT with its ligand and substrates	120
4.28	Specific amino acid residues of three different domains in VMPBSMT	122
4.29	VMPBSMT and its ligand and substrate-enzyme complex	123
4.30	Secondary structure alignment between amino acid residue of VMPEGS and 3c1o template of <i>Clarkia breweri</i> eugenol synthase	129
4.31	NAD(P) binding sites of VMPEGS protein	129
4.32	3D model of VMPEGS with the β strands, S1–S11 and the α helices, H1–H15, labelled.	130

4.33	NAD(P) binding sites of VMPEGS and other models	131
4.34	Comparative modeling of NmrA-like family enzymes in comparison with VMPEGS protein model	134
4.35	Superimposition of the refined VMPEGS model and its template, 3c1o	135
4.36	Eugenol biosynthesis	137
4.37	Molecular docking of VMPEGS with its ligand and substrate	138
4.38	Interactions between template model, 3c1o for VMPEGS and the NADP ⁺ cofactor	139
4.39	Molecular docking of VMPEGS protein with substrate and ligand using Autodock 4.2 software	139
4.40	Secondary structure alignment between amino acid residue of VMPOOMT and 1fp2 template of <i>Medicago sativa</i> isoflavone o-methyltransferase	143
4.41	Conserved domains and secondary structure of VMPOOMT	143
4.42	Comparative modeling of O-methyltransferase enzymes with VMPOOMT protein model	146
4.43	SAM binding sites of O-methyltransferase proteins	147
4.44	Enzymatic reaction of (a) orcinol and (b) guaiacol	148
4.45	VMPOOMT and its predicted substrates and ligands	150
4.46	Molecular docking of VMPOOMT protein with substrates and ligands using Autodock 4.2	151
4.47	Relative VMPKAT transcript levels in <i>Vanda Mimi Palmer</i> detected in (a) different tissues; (b) different flower developmental stages	155
4.48	Relative VMPBSMT transcript levels in <i>Vanda Mimi Palmer</i> detected in (a) different tissues; (b) different flower developmental stages	157
4.49	Relative VMPEGS transcript levels in <i>Vanda Mimi Palmer</i> detected in (a) different tissues; (b) different flower developmental stages	160
4.50	Relative VMPOOMT transcript levels in <i>Vanda Mimi Palmer</i> detected in (a) different tissues; (b) different flower developmental stages	162

LIST OF SYMBOLS

Å	Angstrom
°C	Degree Celcius
µg	Microgram
µl	Microlitre
L	Litre
M	Molar
n/a	Not available
v/v	Volume per volume
w/v	Weight per volume



LIST OF ABBREVIATIONS

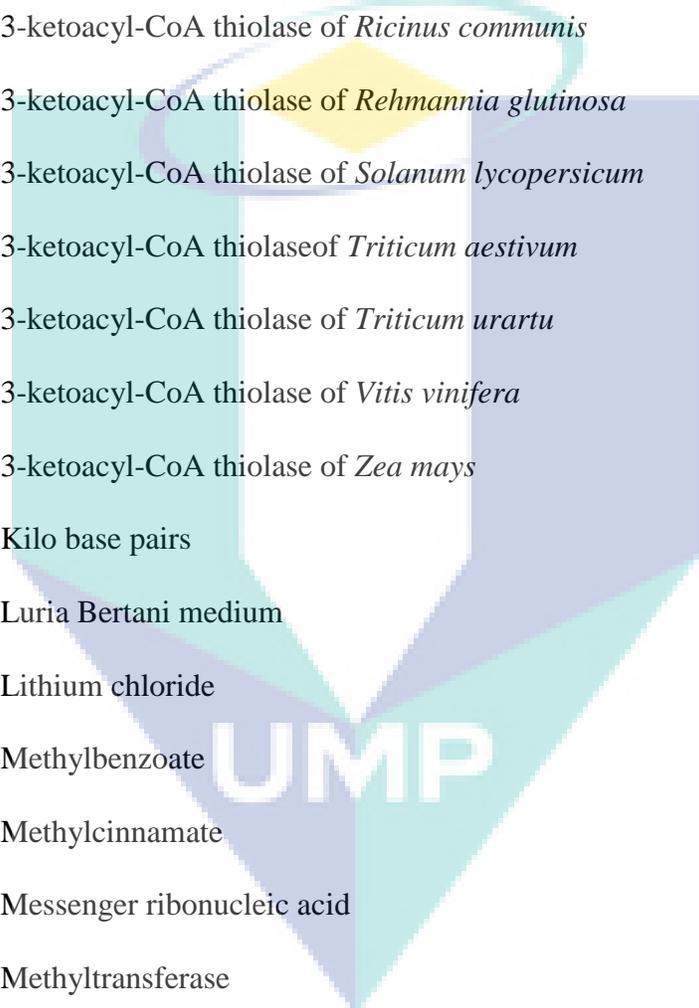


3H3PP	3-hydroxy-3-phenylpropionic acid
3H3PP-CoA	3-hydroxy-3-phenylpropionyl-CoA
3O3PP-CoA	3-oxo-3-phenylpropionyl-CoA
4CL	4-coumaroyl-CoA ligase
AADC	Aromatic amino acid decarboxylase
AMV	Avian Myeloblastosis Virus
ATP	Adenosine triphosphate
BA	Benzoic acid
BA-CoA	Benzoyl-CoA
BAHD	Benzyl alcoholacetyl-,anthocyanin-O-hydroxy-cinnamoyl-,anthranilate-N-hydrox y-cinnamoyl/benzoyl-, deacetylvind oline acetyltransferase
BAlc	Benzylalcohol
Bald	Benzaldehyde
BALDH	Benzaldehyde dehydrogenase
BAMT	Benzoic acid carboxyl methyltransferase
VvBAMT	Benzoic acid carboxyl methyltransferase of <i>Vitis vinifera</i>
BB	Benzylbenzoate
BLAST	Basic Local Alignment Search Tool
AtBMT	Benzoate carboxyl methyltransferase of <i>Aegilops tauschii</i>
RcBMT	Benzoate carboxyl methyltransferase of <i>Ricinus communis</i>
VvBMT	Benzoate carboxyl methyltransferase-like of <i>Vitis vinifera</i>
ZmBMT	Benzoate carboxyl methyltransferase of <i>Zea mays</i>

bp	Base pairs
BPBT	Benzoyl-CoA benzylalcohol/2-phenylethanol benzoyltransferase
BSA	Bovine Serum Albumin
BSMT	Benzoic acid/salicylic acid carboxyl methyltransferase
<i>LhBSMT</i>	SAM:benzoic acid/salicylic acid carboxyl methyltransferase of <i>Lilium hybrid cultivar</i>
<i>PhBSMT</i>	S-adenosyl-L-methionine: benzoic acid/salicylic acid carboxyl methyltransferase of <i>Petunia hybrida</i>
<i>PmBSMT</i>	Benzoic acid/salicylic acid methyltransferase of <i>Protoschwenkia mandonii</i>
C3H	p-coumarate-3-hydroxylase
C4H	Cinnamate-4-hydroxylase
CA	Cinnamic acid
CA-CoA	Cinnamoyl-CoA
CCoAOMT	Caffeoyl-CoA 3-O-methyltransferase
cDNA	Complementary deoxynucleotide acid
cDNA-RDA	cDNA-Representational difference analysis
CFA	Caffeic acid
CFA-CoA	Caffeoyl-CoA
CFAT	Coniferyl alcohol acyltransferase
CITES	Convention on International Trade in Endangered Species of Wild Fauna and Flora
CNL	Cinnamoyl-CoA ligase
COA	Coenzyme A
ConA	Coniferyl alcohol
ConAc	Coniferyl acetate

CTAB	Hexadecyltrimethyl-ammoniumbromide
DNA	Deoxynucleotide acid
DEPC	Diethylpyrocarbonate
DMT	3,5-dimethoxytoluene
dNTP	Deoxynucleoside triphosphate
EB	Elution buffer
EDTA	Ethylenediaminetetraacetic acid
EDO	1,2-ethanediol
EGS	Eugenol synthase
<i>CbEGS</i>	Eugenol synthase of <i>Clarkia breweri</i>
<i>GcEGS1</i>	<i>Gymnadenia conopsea</i> eugenol synthase 1
<i>GcEGS2</i>	<i>Gymnadenia conopsea</i> eugenol synthase 2
<i>GoEGS1</i>	<i>Gymnadenia odoratissima</i> eugenol synthase 1
<i>GoEGS2</i>	<i>Gymnadenia odoratissima</i> eugenol synthase 2
<i>PhEGS</i>	Eugenol synthase of <i>Petunia x hybrida</i>
EST	Expressed sequence tag
Eug	Eugenol
FA	Ferulic acid
FA-CoA	Feruloyl-CoA
FOMT	Flavonoid O-methyltransferase
<i>TcFOMT</i>	Flavonoid O-methyltransferase of <i>Theobroma cacao</i>
GC-MS	Gas Chromatography-Mass Spectrometry
GDPS	Geranyl Diphosphate Synthase
GSP	Gene specific primer

HMO	Isoformononetin
HPLC	High-performance liquid chromatography
IEMT	(iso)eugenol O-methyltransferase
IEug	Isoeugenol
IFR	Isoflavone reductase
<i>Bp</i> IFR	Isoflavone reductase of <i>Betula pendula</i>
<i>Dc</i> IFR	Isoflavone reductase of <i>Dendrobium catenatum</i>
<i>Fv</i> IFR	Isoflavone reductase of <i>Fragaria vesca</i>
<i>Mt</i> IFR	Isoflavone reductase of <i>Medicago truncatula</i>
<i>Pc</i> IFR	Isoflavone reductase of <i>Pyrus communis</i>
<i>Vv</i> IFR	Isoflavone reductase of <i>Vitis vinifera</i>
IGS	Isoeugenol synthase
IPTG	Isopropyl-beta -D-thiogalactopyranoside
<i>Gd</i> IGS1	<i>Gymnadenia densiflora</i> (iso)eugenol synthase 1
<i>Gd</i> IGS2	<i>Gymnadenia densiflora</i> (iso)eugenol synthase 2
KAT	3-Keto-acyl CoA Thiolase
<i>At</i> KAT	3-ketoacyl-CoA thiolase of <i>Aegilops tauschii</i>
<i>At</i> KAT	3-ketoacyl-CoA thiolase of <i>Arabidopsis thaliana</i>
<i>Bd</i> KAT	3-ketoacyl-CoA thiolase of <i>Brachypodium distachyon</i>
<i>Cp</i> KAT	3-ketoacyl-CoA thiolase of <i>Cavia porcellus</i>
<i>Cs</i> KAT	3-ketoacyl-CoA thiolase of <i>Cucumis sativus</i>
<i>Ec</i> KAT	3-ketoacyl-CoA thiolase of <i>Equus caballus</i>
<i>Fv</i> KAT	3-ketoacyl-CoA thiolase of <i>Fragaria vesca</i>
<i>Gm</i> KAT	3-ketoacyl-CoA thiolase of <i>Glycine max</i>



<i>MiKAT</i>	3-ketoacyl-CoA thiolase of <i>Mangifera indica</i>
<i>MmKAT</i>	3-ketoacyl-CoA thiolase of <i>Mus musculus</i>
<i>MtKAT</i>	3-ketoacyl-CoA thiolase of <i>Medicago truncatula</i>
<i>OsKAT</i>	3-ketoacyl-CoA thiolase of <i>Oryza sativa</i>
<i>PhKAT</i>	3-ketoacyl CoA thiolase of <i>Petunia hybrid</i>
<i>RcKAT</i>	3-ketoacyl-CoA thiolase of <i>Ricinus communis</i>
<i>RgKAT</i>	3-ketoacyl-CoA thiolase of <i>Rehmannia glutinosa</i>
<i>SlKAT</i>	3-ketoacyl-CoA thiolase of <i>Solanum lycopersicum</i>
<i>TaKAT</i>	3-ketoacyl-CoA thiolase of <i>Triticum aestivum</i>
<i>TuKAT</i>	3-ketoacyl-CoA thiolase of <i>Triticum urartu</i>
<i>VvKAT</i>	3-ketoacyl-CoA thiolase of <i>Vitis vinifera</i>
<i>ZmKAT</i>	3-ketoacyl-CoA thiolase of <i>Zea mays</i>
kb	Kilo base pairs
LB	Luria Bertani medium
LiCl	Lithium chloride
MeBA	Methylbenzoate
MeCA	Methylcinnamate
mRNA	Messenger ribonucleic acid
MT	Methyltransferase
<i>CsMT</i>	Carboxyl methyltransferase of <i>Crocus sativus</i>
<i>OsMT</i>	SAM dependent carboxyl methyltransferase of <i>Oryza sativa</i>
NADP	Nicotinamide adenosine diphosphate
NaOAc	Sodium acetate
NCBI	National Center for Biotechnology Information

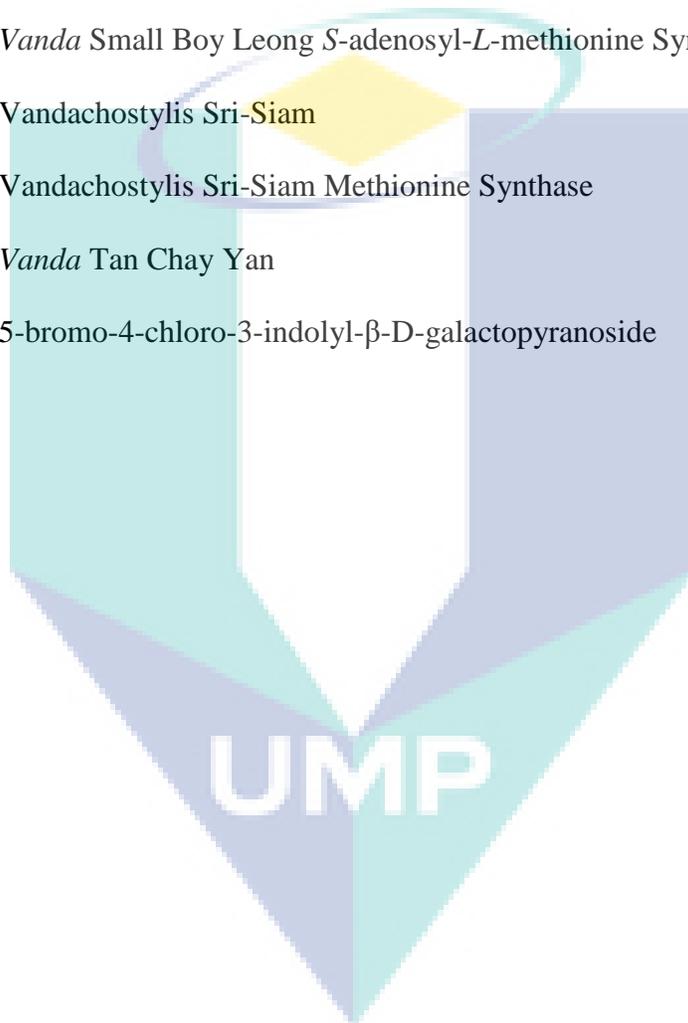
NmrA	Negative transcriptional regulator
OMT	O-methyltransferase
CcOMT	O-methyltransferase of <i>Citrus clementina</i>
RcOMT	O-methyltransferase of <i>Ricinus communis</i>
VpOMT	O-methyltransferase of <i>Vitis pseudoreticulata</i>
OMTs	O-methyltransferases
OOMT	Orcinol-O-methyltransferase
RcOOMT	Orcinol O-methyltransferase of <i>Rosa canina</i>
RcOOMT	Orcinol O-methyltransferase of <i>Rosa chinensis</i>
RgOOMT	Orcinol O-methyltransferase of <i>Rosa gallica</i>
RhOOMT	Orcinol O-methyltransferase of <i>Rosa hugonis</i>
RhOOMT	Orcinol O-methyltransferase of <i>Rosa hybrid cultivar</i>
ORF	Open reading frame
PAAS	Phenylacetaldehyde synthase
PAL	Phenylalanine ammonia lyase
PbGDPS	<i>Phalaenopsis bellina</i> geranyl diphosphate synthase
pCA	p-coumaric acid
pCA-CoA	p-coumaroyl-CoA
PCBER	Phenylcoumaran benzylic ether reductase
EgPCBER	Phenylcoumaran benzylic ether reductase of <i>Eucalyptus globulus</i>
EpPCBER	Phenylcoumaran benzylic ether reductase of <i>Eucalyptus pilularis</i>
EpPCBER	Phenylcoumaran benzylic ether reductase of <i>Eucalyptus pyrocarpa</i>
FiPCBER	Phenylcoumaran benzylic ether reductase of <i>Forsythia intermedia</i>
LcPCBER	Phenylcoumaran benzylic ether reductase of <i>Linum corymbulosum</i>

<i>Nt</i> PCBER	Phenylcoumaran benzylic ether reductase of <i>Nicotiana tabacum</i>
<i>Ps</i> PCBER	Phenylcoumaran benzylic ether reductase of <i>Pinus strobes</i>
<i>Pt</i> PCBER	Phenylcoumaran benzylic ether reductase of <i>Populus trichocarpa</i>
<i>Sa</i> PCBER	Phenylcoumaran benzylic ether reductase of <i>Striga asiatica</i>
<i>Sf</i> PCBER	Phenylcoumaran benzylic ether reductase of <i>Salvia fruticosa</i>
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEB	Phenylethylbenzoate
PhA	Phenylacetaldehyde
Phe	L-phenylalanine
PhEth	2-phenylethanol
PhEthA	Phenylethylamine.
pI	Isoelectric point
PIP	Pinoresinol-lariciresinol reductase, isoflavone reductase and phenylcoumaran benzylic ether reductase
PVP	Polyvinyl pyrrolidone
RACE	Rapid Amplification of cDNA Ends
RE	Restriction enzymes
RMSD	Root Mean Standard Deviation
RNA	Ribonucleic Acid
RNase	Ribonuclease
ROMT	Resveratrol O-methyltransferase
<i>Cs</i> ROMT	Trans-resveratrol di-O-methyltransferase of <i>Citrus sinensis</i>
<i>Eg</i> ROMT	Trans-resveratrol di-O-methyltransferase of <i>Elaeis guineensis</i>

<i>Ns</i> ROMT	Trans-resveratrol di-O-methyltransferase of <i>Nicotiana sylvestris</i>
<i>Nt</i> ROMT	Trans-resveratrol di-O-methyltransferase of <i>Nicotiana tomentosiformis</i>
<i>Pd</i> ROMT	Trans-resveratrol di-O-methyltransferase of <i>Phoenix dactylifera</i>
<i>Pm</i> ROMT	Trans-resveratrol di-O-methyltransferase of <i>Prunus mume</i>
<i>Si</i> ROMT	Trans-resveratrol di-O-methyltransferase of <i>Sesamum indicum</i>
<i>Si</i> ROMT	5-pentadecatrienyl resorcinol O-methyltransferase of <i>Setaria italica</i>
<i>Sl</i> ROMT	Trans-resveratrol di-O-methyltransferase of <i>Solanum lycopersicum</i>
<i>St</i> ROMT	Trans-resveratrol di-O-methyltransferase of <i>Solanum tuberosum</i>
<i>Vv</i> ROMT	Resveratrol O-methyltransferase of <i>Vitis vinifera</i>
RT-PCR	Reverse transcription polymerase chain reaction
SABATH	<i>S</i> -adenosyl- <i>L</i> -methionine: salicylic acid carboxyl methyltransferase (SAM:SAMT), <i>S</i> -adenosyl- <i>L</i> -methionine: benzoic acid carboxyl methyltransferase (SAM:BAMT), and theobromine synthase
SAM	<i>S</i> -adenosyl- <i>L</i> -methionine
SAMS	<i>S</i> -adenosyl- <i>L</i> -methionine synthase
SAMT	Salicylic acid carboxyl methyltransferase
<i>Am</i> SAMT	<i>S</i> -adenosyl- <i>L</i> -methionine: salicylic acid methyltransferase of <i>A. majus</i>
<i>Cp</i> SAMT	<i>S</i> -adenosyl- <i>L</i> -methionine: salicylic acid carboxyl methyltransferase of <i>Chimonanthus praecox</i>
<i>Hc</i> SAMT	<i>S</i> -adenosyl- <i>L</i> -methionine: salicylic acid carboxyl methyltransferase of <i>Hoya carnosa</i>
<i>Mt</i> SAMT	Salicylic acid carboxyl methyltransferase of <i>Medicago truncatula</i>
<i>Na</i> SAMT	<i>S</i> -adenosyl- <i>L</i> -methionine:salicylic acid carboxyl methyltransferase of <i>Nicotiana alata</i>
<i>Nb</i> SAMT	Salicylic acid carboxyl methyltransferase of <i>Nicotiana benthamiana</i>
<i>Ns</i> SAMT	<i>S</i> -adenosyl- <i>L</i> -methionine:salicylic acid carboxyl methyltransferase of <i>Nicotiana suaveolens</i>

<i>Nt</i> SAMT	Salicylic acid carboxyl methyltransferase of <i>Nicotiana tabacum</i>
<i>Sj</i> SAMT	Salicylic acid carboxyl methyltransferase of <i>Streptosolen jamesonii</i>
SA-PMPs	Streptavidin-paramagnetic particles
SDR	Short-chain dehydrogenase/reductase
SSC	Sodium Saline Citrate
SPME	Solid Phase Microextraction
TAE	Tris-acetate-EDTA
TE	Tris-EDTA
TOMT	Tabersonine 16-O-methyltransferase
<i>Mn</i> TOMT	Tabersonine 16-O-methyltransferase of <i>Morus notabilis</i>
Tris	Tris (hydroxymethyl)-aminomethane
UPM	Universal Primer Mix
UTR	Untranslated region
VMP	<i>Vanda Mimi Palmer</i>
VMPAAT	<i>Vanda Mimi Palmer</i> Alcohol acyl-transferase
VMPACA	<i>Vanda Mimi Palmer</i> Acetyl-CoA-acetyl transferase
VMPBSMT	<i>Vanda Mimi Palmer</i> Benzoic Acid/Salicylic Acid Carboxyl Methyltransferase
VMPCMEK	<i>Vanda Mimi Palmer</i> 4-(cytidine 5'-diphospho)-2-C-methyl-d-erythritol kinase
VMPCyP450	<i>Vanda Mimi Palmer</i> Cytochrome P450
VMPDXR	<i>Vanda Mimi Palmer</i> 1-deoxyd-xylulose 5-phosphate reductoisomerase
VMPEGS	<i>Vanda Mimi Palmer</i> Eugenol Synthase
VMPKAT	<i>Vanda Mimi Palmer</i> 3-ketoacyl-CoA thiolase
VMPOOMT	<i>Vanda Mimi Palmer</i> Orcinol O-methyltransferase

VMPPAAS	<i>Vanda</i> Mimi Palmer Phenylacetaldehyde Synthase
VMPPAL	<i>Vanda</i> Mimi Palmer Phenylalanine Ammonia Lyase
VMPSTS	<i>Vanda</i> Mimi Palmer Sesquiterpene Synthase
VSBL	<i>Vanda</i> Small Boy Leong
VSBLMS	<i>Vanda</i> Small Boy Leong Methionine Synthase
VSBLSAMS	<i>Vanda</i> Small Boy Leong <i>S</i> -adenosyl- <i>L</i> -methionine Synthase
VSSS	Vandachostylis Sri-Siam
VSSSMS	Vandachostylis Sri-Siam Methionine Synthase
VTCY	<i>Vanda</i> Tan Chay Yan
X-gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside



CHAPTER 1

INTRODUCTION

1.0 INTRODUCTION

Fragrances produced and emitted by floral tissues of flowering plants are very important for many crops whereby pollination of the flowers by pollinating agents, especially insects and birds will guarantee the yield as well as quality of their products for agricultural and horticultural industries (Andrews *et al.*, 2007). Besides, the presence of fragrance increases the aesthetic values of ornamental plants in cut-flower industry. Floral scent can be described as a concoction of low molecular weight compounds, including terpenoids, benzenoids and phenylpropanoids that are synthesised via terpenoid as well as benzenoid and phenylpropanoid biosynthetic pathways commonly found in various plants (Dudareva *et al.*, 2004). Investigation on fragrance biosynthesis via biochemistry and molecular approaches has been the focus of many researchers in the past decade in order to understand the role of fragrance-related genes for fragrance biosynthesis in scented flowers such as *Clarkia breweri*, *Anthirrhinum majus*, *Rosa hybrida* and *Petunia hybrida* (Dudareva *et al.*, 1996; Murfitt *et al.*, 2000; Lavid *et al.*, 2002; Verdonk *et al.*, 2005). Subsequently, the isolated fragrance-related genes from the scented flowers were further investigated for their characteristics via genetic engineering approach by transformation into several well-studied plants including *Arabidopsis thaliana* and *Nicotiana* spp. (Willmer *et al.*, 2009). Understanding floral scent biosynthesis in terpenoid as well as

benzenoid and phenylpropanoid pathway opens up a new opportunity to increase the commercial values of flowers and fruits that are used in fragrance and flavor as well as food industries (Colquhoun *et al.*, 2010).

Orchids belong to *Orchidaceae* family, the largest flowering plant family that consists of approximately 25,000 to 30,000 species belonging to more than 800 genera (Cozzolino and Widmer, 2005). Even though the actual number of orchid species is still far from being known clearly due to the continuous discovery of orchid species by orchid enthusiasts and botanists, their number of species are estimated to be four times higher than mammals and twice of birds. To date, orchids have evolved from their ancestral characteristics due to selection pressure as well as their adaptation. Orchid species have been hunted since mid-1700s all over the world due to their aesthetical values as well as exotic characteristics (Salzmann *et al.*, 2007). Due to the serious threat faced by wild orchid species, protection and conservation of the species have been seriously initiated by the Royal Botanic Garden in Kew, London. In addition, wild orchids covering approximately 90% of orchid species from the entire world are categorised as an endangered species in the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) (Eltz *et al.*, 2003). Thus in these days, a lot of orchid hybrids have been produced by crossing from the same genera (interspecific hybrids) as well as from different genera (inter-generic hybrids) for better characteristics including colors, shapes as well as fragrance for commercialisation purpose in floricultural industry (Scopece *et al.*, 2007).

Fragrant-orchids are commonly sold at higher price compared to orchids without any fragrance. In nature, volatiles from orchids and other plants are mainly used by the plants to attract specific pollinators to help in their pollination whereby different types of pollinators are attracted to different types of fragrance (Topik *et al.*, 2005). It has been reported that orchid species including *Vanda tessellata*, *Phalaenopsis bellina*, *Platanthera chlorantha*, *Polystachya cultriformis* and *Zygopetalum crinitum* are among orchid species that produce and emit strong fragrances (Kaiser, 1993). Unfortunately, extensive breeding work to produce attractive colours and shape of orchid flowers cause modern orchid hybrid to lose their fragrance characteristics because in previous traditional breeding, fragrance

was not included as the main objective of the new hybrids (Liu *et al.*, 2008). To date, some orchid hybrids with fragrant characteristic are extensively cultivated in Southeast-Asian countries including *Vanda* Mimi Palmer, *Vanda* Small Boy Leong, *Vanda* Johanna Ljunggren and *Vandachostylis* Sri-Siam (Chew, 2008). The orchid hybrids are propagated through tissue culture approach, mainly in Thailand to be commercialised not only in Thailand and Malaysia but also all around the world.

Research on fragrant-orchids has been started since early 1990s on determination of their volatile components via analysis using Gas Chromatography-Mass Spectrometry (GC-MS). Meanwhile molecular work on fragrance-related transcripts isolation and characterization was only reported in 2006 by identification of geranyl diphosphate synthase (GDPS) and some other putative fragrance-related transcripts from *Phalaenopsis bellina* through the establishment of Expressed Sequence Tags (ESTs) library (Hsiao *et al.*, 2006; 2008). Subsequently, some other fragrance-related transcripts, including alcohol acyltransferase (VMPPAAT), phenylacetaldehyde synthase (VMPPAAS) and sesquiterpene synthase (VMPSTS) have been isolated by a research group from Malaysia on another fragrant-orchid, *V. Mimi Palmer* that has won several international awards for its sweet-strong fragrance (Chan *et al.*, 2011; Mohd-Hairul, 2011). Lately, the sesquiterpene synthase (VMPSTS) of *V. Mimi Palmer* has been reported to be functionally expressed in *Lactococcus lactis*, a gram positive bacterium. *In vitro* analysis has shown that VMPSTS is involved in catalysing the biosynthesis of multiple sesquiterpene compounds including germacrene D, copaene and nerolidol (Song *et al.*, 2012).

In this study, screening of candidate for fragrance-related transcripts was carried out via cDNA Representational Difference Analysis (cDNA-RDA), a powerful PCR subtraction approach to identify fragrance-related transcripts of *V. Mimi Palmer* and other fragrant-orchids including *Vanda* Small Boy Leong and *Vandachostylis* Sri-Siam in comparison to a non-fragrant orchid, *Vanda* Tan Chay Yan. From the cDNA-RDA analysis, three putative fragrance-related transcripts that might be involved in benzenoid and phenylpropanoid pathway for fragrance biosynthesis have been identified including phenylalanine ammonia lyase (VMPPAL) and S-adenosyl-L-methionine synthase

(VMPSAMS) from *V. Mimi Palmer* as well as two methionine synthases from *V. Small boy Leong* (VSBLMS) and *Vandachostylis Sri-Siam* (VSSSMS), respectively. Unfortunately, all the three transcripts were not selected for further characterisation due to the fact that the enzymes encoded by the transcripts have less commercial values compared to the targeted transcripts encoding enzymes that are closely involved in the biosynthesis of final products of benzenoids as well as phenylpropanoids as fragrance compounds. Thus, in this study, other four fragrance-related transcripts that were previously identified from cDNA library of *V. Mimi Palmer* constructed by Chan *et al.* (2009) were selected of the full length clones namely eugenol synthase (VMPEGS), benzoic acid/salicylic acid carboxyl methyltransferase (VMPBSMT), keto-acyl CoA thiolase (VMPKAT) and orcinol-O-methyltransferase (VMPOOMT). Further characterisation of all the transcripts were carried out by sequence analysis and protein bioinformatics approaches as well as gene expression analysis using real-time RT-PCR.

1.1 PROBLEM STATEMENT

Floral fragrance is due to a complex mixture of compounds which remained yet to be fully discovered, although the most recent progress of floral fragrance research applying molecular as well as bioinformatics approaches. To date, advanced researches on floral fragrance comprise a few combined strategies related to biochemical, genomics and bioinformatics approaches for elucidation of fragrance biosynthetic pathways including their related genes, transcripts and enzymes. Thus, volatiles that are derived from dicotyledonous plants rather than monocotyledonous plants have been studied extensively to discover their fragrance biosynthesis mechanisms. In *Orchidaceae*, volatile compounds either in extracted essential oils or directly-captured volatiles using solid phase microextraction (SPME) have been analysed using Gas Chromatography-Mass Spectrometry (GC-MS). However, discovery of fragrance biosynthetic pathways in orchids is still incomplete especially in benzenoid and phenylpropanoid pathways whereby exact genes, transcripts as well enzymes that are involved in catalysing the biosynthesis of volatile products that are distributed among fragrant-orchids including methylbenzoate, benzyl acetate, phenylacetaldehyde, phenylethanol and phenylethyl acetate are still in

progress. Thus, full length cDNAs of fragrance-related transcripts in benzenoid and phenylpropanoid pathways as well as their characterisation using molecular and bioinformatics approaches will enhance understanding on these pathways in orchids especially on Vandaceous orchids.

1.2 RESEARCH OBJECTIVES

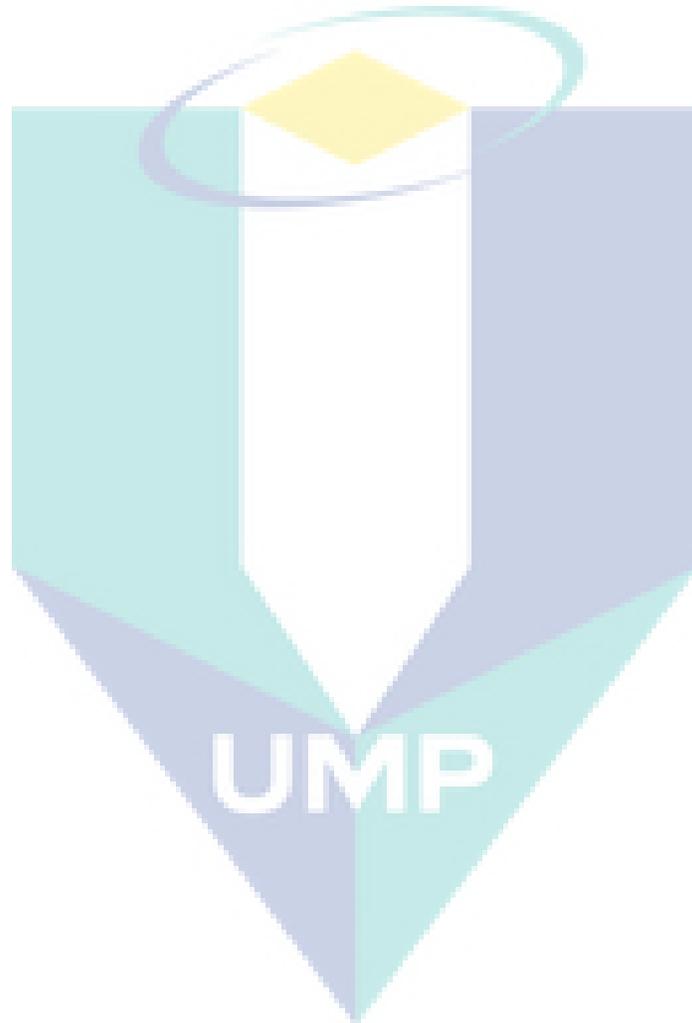
The specific objectives of this study were:

- 1) To screen candidates of fragrance-related transcripts in benzenoid and phenylpropanoid pathway of *Vanda Mimi Palmer* and other selected fragrant-orchids.
- 2) To isolate full cDNA sequence of four (4) selected fragrance-related transcripts in benzenoid and phenylpropanoid pathway of *Vanda Mimi Palmer*.
- 3) To characterise the full length fragrance-related cDNA sequences using molecular and bioinformatics approaches.

1.3 SCOPE OF STUDY

- 1) Identification of fragrance-related transcripts from selected fragrant-orchids including *Vanda Mimi Palmer*, *Vanda Small Boy Leong* and *Vandachostylis Sri-Siam* in comparison to *Vanda Tan Chay Yan*, a non-fragrant orchid via cDNA-RDA approach.
- 2) Amplification and isolation of full cDNA sequences of selected fragrance-related transcripts namely VMPKAT, VMPBSMT, VMPEGS and VMPOOMT that might be involved in benzenoid and phenylpropanoid pathways of *Vanda Mimi Palmer* via Rapid Amplification of cDNA-Ends PCR (RACE-PCR).

- 3) Characterisation of full length sequence of fragrance-related transcripts using molecular and bioinformatics approaches such as sequence analysis, identification of fragrance-related domains, development of 3D-structural model of fragrance-related proteins, molecular docking for substrate and ligand predictions as well as gene expression analysis using Real-time RT-PCR.



CHAPTER 2

LITERATURE REVIEW

2.1 ORCHIDS: INTRODUCTION

In the Plantae kingdom, seed plants can generally be divided into two groups which are flowering plants (angiosperms) and non-flowering plants (gymnosperms) (Audesirk *et al.*, 2002). Orchidaceae which has been reported as the largest family of flowering plants comprises 17,000 to 35,000 species that belong to 880 genera. Orchids are estimated to cover almost 30% of monocotyledonous plants as well as 10% of total flowering plants from all around the world (Dressler, 1993; Hossain, 2011). Besides that, Orchidaceae family has received a great interest among evolutionary biologist, botanist and orchid enthusiasts from all around the world after the establishment of a book on orchids 'Fertilization of Orchids' by Darwin in 1862. Orchids in Orchidaceae family are divided into five recognised subfamilies namely Apostasioideae, Cypripedioideae, Epidendroideae, Orchidoideae and Vanilloideae (Chase *et al.*, 2003). Among all the orchids in Orchidaceae family, more than 230 genera of orchids covering 898 to 4,000 species in 143 genera can be found in Peninsular Malaysia (Go *et al.*, 2010).

Orchid plants are naturally grown in a wide range of ecological habitats (Ramirez *et al.*, 2007) that are most suitable for their growth requirement based on their specialized morphological, structural and physiological characteristics (Dressler, 1990). Among all orchid species from the entire world, an estimated 70% of orchids are classified as epiphytic plants that grow on the trunks or branches of trees without consuming any nutrient from the host plants. In addition, epiphytic orchids have been reported to be nearly

two-thirds of total epiphytic flowering plants from all around the world (Gravendeel *et al.*, 2004; Hsu *et al.*, 2011). Besides epiphytic orchids, there are an estimated 25% of orchids from all around the world which have been reported to be terrestrial orchids whereby the orchids grow on soil and absorb nutrients directly from soil. Meanwhile the remaining 5% of orchids can be found to be grown naturally in various other supports (Atwood, 1986). It has been reported that almost all temperate orchids that grown in temperate regions are terrestrial orchids while tropical orchids grown in tropical regions are mostly epiphytic as well as lithophytic orchids (Dressler, 1990). Lithophytic orchids are another type of orchids that grow naturally on exposed rocks whereby their roots are directly attached to the extremely rocky terrain with slight humus materials (McDonald, 1999).

Among all types of orchids, lithophytic orchids are rarely selected to be cultivated by botanists and orchid enthusiasts due to their natural requirement either for high level of humidity or warm temperature throughout the year that can be quite difficult to be maintained in housing areas (Alikas, 2009; McDonald, 1999). Meanwhile for terrestrial orchids, they are mostly cultivated for cut flower industry due to their suitability to be grown on soil as well as their easier maintenance and requirement compared to epiphytic and lithophytic orchids (Ramirez *et al.*, 2007). Besides that, semi-terrestrial orchids that can be grown on the ground are also chosen as orchids to be cultivated whereby the root of the orchids do not directly penetrate the soil hence the factors including pH, salt levels and other materials in the soil will not directly inhibit the growth of the orchids. Instead of directly attached to soil, the root of this type of orchids prefer to find supports as well as nutrients from humus and leaves that are available on the ground surface (McDonald, 1999). In addition, this type of semi-terrestrial orchids require open space areas including coastal shrubs, wetlands as well as open shrubby forests where they can obtain adequate level of sunlight for their photosynthesis process without depending on any other plants for support (Alikas, 2009). Examples of terrestrial orchids that are favoured by botanists as well as orchid enthusiasts to be cultivated are *Cymbidium* orchids as well as slipper orchids (Dharmani, 2013).

In orchid industry, there are more than 100,000 orchid hybrids that have been produced by orchid breeders and the number keeps increasing every year in order to produce orchids with better characteristics including the brightness and exotic shape of inflorescent as well as their scent to increase their commercial values (Hands, 2006). Orchid hybrids are produced either by crossing species in the same genera (inter-specific hybrid) or by crossing orchids from different genera (inter-generic hybrid) (Hands, 2006). Examples of orchid hybrids used in this study are *V. Mimi Palmer* (*Vanda tessellata* x *Vanda* Tan Chay Yan), *V. Tan Chay Yan* (*Vanda dearea* x *Vanda* Josophine van Brero), *V. Small Boy Leong* (*Vanda tricolor* and *Vanda limbata*) and *Vandachostylis* Sri-Siam (*Vanda tessellata* x *Rhynchostylis gigantea*). From the list, *V. Mimi Palmer*, *V. Tan Chay Yan* and *V. Small Boy Leong* are inter-specific hybrids whereby all those orchids have been crossed between orchids in *Vanda* genera. Meanwhile, *Vandachostylis* Sri-Siam is an example of inter-generic hybrid whereby the orchid is produced by hybridisation between two orchid species from two different genera.

In floricultural industry, a large number of orchids have significant economic value where the products including cut-flowers and potted orchids from Malaysia, Singapore and Thailand have contributed an export value of RM200 millions annually (Raubeson *et al.*, 2005; Ooi, 2005). In Japan floral market, 23.4% of orchids have been imported from Malaysia (7,648 million yen), followed by Columbia (19.2%) and China (10.4%) as reported by the Japan Florists' Telegraph Delivery Association, 2010. In Malaysia and Singapore orchid industry, the development of the industry is enhanced by the cultivation of various orchid hybrids due to the easy cultivation of orchid hybrid compared to orchid species that are more difficult to be maintained, besides their free blooming habit as well as their various patterns of shapes, colours and a new array of flowers (Kishor *et al.*, 2006). The demand for orchid species and hybrids is still very high in the floricultural industry that is not limited to local market, but also from international market due to their aesthetic values (exotic and limited sources), including colour, scent and morphology (Hamdan, 2008).

2.2 VANDACEOUS ORCHIDS

Vandaceous orchids (subfamily Vandoideae) that consist of several genera including *Aranda*, *Arachnis*, *Ascocentrum*, *Mokara*, *Phalaenopsis* and *Vanda* are among the most popular orchids that have been cultivated in orchid industry where these types of orchids receive a high demand as cut-flowers as well as potted flowering plants for both local and international markets in this two decades (Lekawatana, 2010; Fadelah *et al.*, 2001). Among the Vandaceous orchids, orchids from *Vanda* genera are the second most popular orchids after *Phalaenopsis* (Cheam *et al.*, 2009). *Vanda* orchids are favoured by botanists and orchid enthusiasts due to their fast growth and frequent blooming compared to other orchids (Allikas, 2009). *Vanda* orchids are categorised as monopodial orchids that bloom frequently within three to five times in a year and the flowers in full-bloom stage have longer life-time (two to three weeks) compared to other orchids with shorter life-time (Dharmani, 2013). In addition, orchids from this genera have magnificent flowers with broad colour range including maroon, green, orange, red and white (Cheam *et al.*, 2009). Besides that, popularity of *Vanda* orchid keeps increasing due to a lot of successful orchid hybrids which have been produced by crossing between different species in this genera as well as different genera (Lekawatana, 2010). Exportation activity on Vandaceous orchid in cut flower industry has been initiated in Singapore in mid-1950 (Ng, 1984). To date, many different types of orchid cut-flowers have been exported to over 30 countries including Japan, Australia, USA and Western Europe (Goh and Kavaljian, 1989; Koay, 1993).

Vanda orchids have originated from tropical countries including Malaysia, Philippines and Thailand as their natural habitats (Chew, 2007). There are about 50 species from *Vanda* genera that have been identified to be native to tropical Asian countries and distribution of *Vanda* species has been reported from Sri Lanka and southern India to Papua New Guinea and, northern Australia and Solomon Islands as well as to the north region of China, Taiwan and the Philippines (Chew, 2008). *Vanda* orchids are also known as sympodial epiphytic orchids that can grow easily in full sun exposure, bright light, warm temperatures and in high humidity areas. Other than that, *Vanda* orchids have long and trailing roots to absorb nutrients and moisture from the atmosphere (Kishor *et al.*, 2006). Most orchids in *Vanda* genera are epiphytic but some of them have been reported to be either lithophytic or terrestrial that can survive very well in warm and full-sun habitats

(Dharmani, 2013). However, there are some orchid species in *Vanda* genera that have been identified to grow naturally in cool and low temperature highland regions including the north of India, Himalaya, Southeast Asia, Papua New Guinea, south China as well as northern Australia (Cheam *et al.*, 2009).

All *Vanda* orchids possess thick stems and roots and the orchids normally require at least for three and a half to 10 years to reach their maturity stage as flowering plants (Kishor *et al.*, 2008). Leaves of *Vanda* orchids are flat, strap-shaped and close together but might be different in size and morphology which depend on their habitats (Cheam *et al.*, 2009). Basically, *Vanda* orchids can be categorised into three main groups according to their leaf morphological structures which are strap-leaves, semi-terete and terete *Vanda*. Strap-leaf *Vanda* is characterised by the presence of flat leaves while terete *Vanda* orchids is described by their round and pencil-shape leaves. Meanwhile, semi-terete *Vanda* is characterised by the hybrid of both leaf characters, with intermediate size of leaves (American Orchid Society, 2013).

2.2.1 *Vanda* Mimi Palmer

Vanda Mimi Palmer is an orchid hybrid that emits strong sweet fragrance from its floral tissues (Figure 2.1(a)). This orchid is produced from the crossing between a well-known fragrant-orchid *Vanda tessellata* with an orchid hybrid *Vanda* Tan Chay Yan. *Vanda* Mimi Palmer received its sweet fragrance as well as tri-colour (purplish-green-brown) flower from *Vanda tessellata* while the shape of petal, sepal and leaves might be inherited from *Vanda* Tan Chay Yan (Mohd-Hairul, 2011). The fragrance emitted by floral parts of this orchid has been reported to be developmentally regulated, whereby none of fragrance emission has been detected in bud stage and starts to be detected in half-open flower stage and reaches the highest peak during fully-open flower stage (Mohd-Hairul *et al.*, 2010a). Interestingly, this fragrant-orchid has not only received high demand among orchid enthusiasts but also received several international awards for its strong sweet-smelling fragrance including The Champion Award of Fragrant-Orchids by the Royal Horticultural Society of Thailand in 1993 and the Best Fragrant Orchid in the 17th World



Figure 2.1: Orchids used in this study; (a) *Vanda Mimi Palmer*, (b) *Vanda Tan Chay Yan*, (c) *Vanda Small Boy Leong* and (d) *Vandachostylis Sri-Siam*.

Orchid Conference in 2002 (Nair and Arditti, 2002). Besides that, another reason why this orchid hybrid receives high demand among orchid lovers is due to its frequent blooming within a year with maximum emission of floral scent during fully-open flower stage (Chan *et al.*, 2011). Scientific studies on this fragrant-orchid by researchers from Universiti Putra Malaysia have reported the identification, isolation and characterisation of fragrance-related transcripts that are involved in fragrance biosynthesis in this fragrant-orchid via molecular and genetic engineering works (Mohd-Hairul *et al.*, 2010b; Teh *et al.*, 2011; Chan *et al.*, 2011; Song *et al.*, 2012). Interestingly, a sesquiterpene synthase transcript from this orchid has been successfully expressed in *L. lactis*, a gram positive bacterial system whereby the recombinant enzyme has been reported to be involved in catalysing the biosynthesis of multiple sesquiterpene compounds including germacrene D, copaene and nerolidol (Song *et al.*, 2012).

2.2.2 Vanda Tan Chay Yan

Vanda Tan Chay Yan (Figure 2.1(b)) that was used in this study is an orchid hybrid that had been produced by the crossing between *Vanda Josephine* and *Vanda dearei* (Yeoh, 1978) with golden colour of round and flat petals and sepals (Mohd-Hairul *et al.*, 2010a). This orchid hybrid has been reported to be popular in 1960s and had won several international awards including the First Class certificate in 1954 by the Royal Horticultural Society and the Trophy for The Best *Vanda* at the Second World Conference in Hawaii. This orchid hybrid was first produced by Robert Tan Hoon Siang, son of a Malaysian rubber plantation merchant, Tan Chay Yan (<http://eresources.nlb.gov.sg/>). However, this golden colour beautiful orchid hybrid has been reported to lose its popularity as cut flower due to its infrequent blooming of twice a year (Yeoh, 1978). Studies on this orchid by Mohd-Hairul *et al.* (2010a) has shown that this orchid hybrid emits β -ocimene compound that is normally produced by plants for pollination and plant defense. In molecular aspects, none of the β -ocimene synthase gene, transcript and enzyme that are responsible for the biosynthesis of β -ocimene has been reported so far from any orchid.

2.2.3 Vanda Small Boy Leong

Vanda Small Boy Leong is a delightful primary orchid hybrid with brown yellowish sepals and petals containing white spots (Figure 2.1(c)). This orchid hybrid was produced by the crossing of two orchid species, *Vanda tricolor* (seed parent) and *Vanda limbata* (pollen parent). This orchid hybrid was registered by International Orchid Registrar under the Royal Horticultural Society, England on 1st January 1955 (<https://www.rhs.org.uk/>). This orchid is classified as fragrant-orchid that emits fragrance by its fully-open flower and might receive its strong fragrance characteristics from its pollen parent *Vanda tricolor*. Unfortunately, no study on the biochemical analysis of its fragrance as well as fragrance-related genes, transcripts and enzymes have been reported so far.

2.2.4 *Vandachostylis* Sri-Siam

Vandachostylis Sri-Siam (Figure 2.1(d)) is an orchid hybrid of *Vanda tessellata* and *Rhynchostylis gigantea*. Seed parent for this orchid hybrid is *Vanda tessellata* while the pollen parent is *Rhynchostylis gigantea*. The purplish-grape colour of *Vandachostylis* Sri-Siam as well as the shape and size of this orchid hybrid might be inherited from *Rhynchostylis gigantea* while its fragrant characteristic might be inherited from the other parent, *Vanda tessellata*. This orchid hybrid was registered with the Royal Horticultural Society in England on 1st January 1978 (<https://www.rhs.org.uk/>). Unfortunately, until this moment, no scientific finding has been reported on any biochemical and molecular biology aspects of this fragrant orchid.

2.3 BIOLOGY OF FLORAL SCENT

Floral scent or floral fragrance is due to the production of a complex mixture of low molecular mass molecules including monoterpenes, sesquiterpenes, benzenoids, phenylpropanoids and fatty acid derived compounds that play a very important role in pollination in floral tissues and plant defense in vegetative tissues (Raguso, 2008; Knudsen *et al.*, 1993). In general, emission of floral scent by flowering plants appears during anthesis stage when the petals are already opened and ready to be pollinated by specific or non-specific pollinators (Schade *et al.*, 2001). Specific odours are commonly produced by

flowering plants in order to attract specific pollinators such as bees, moths and butterflies to help in their pollination (Dobson, 1994). Besides that, specific emission time of floral scent by floral organ of flowering plants have been reported to be different due to the different active time of specific pollinators as well as the control by circadian clock and photoperiod (Verdonk *et al.*, 2003). This plant-insect interaction is very important to determine the accomplishment of pollination that leads to fruit development in many crop species (Shuttleworth and Johnson, 2009). In addition, biochemical compounds that are present in floral scent play a very important role for plant protection against their natural enemies including herbivores, ants and other pathogens (Schiestl, 2010).

Floral scents of different plant species have been identified to be different due to the combination and compositional level of each compound (Dudareva *et al.*, 2000; Knudsen *et al.*, 1993). Among flowering plants that have been reported to synthesise and emit floral scent, petals of the flowers play an important role for scent emission rather than the entire floral organ (Pichersky *et al.*, 1994) whereby synthesised essential oils are temporarily stored in special oil glands that exist in petals of the flowers before being released as volatiles to the environment (Effmert *et al.*, 2006). However, this concept is different in orchids whereby sepals and lips of orchid flowers also contribute to the fragrance biosynthesis and emission. This is due to the fact that both petals and sepals share almost the same morphological structure in matured flowers where petal becomes the first whorl while sepals become the second whorl of the flowers (Schiestl, 2010). Meanwhile, a special structure of lip or labellum which is a modified petal in orchid flowers is totally different compared to petals and sepals where expression of fragrance-related genes have been detected in the organ but in much lower level compared to petals and sepals (Raguso, 2008). The component of floral scent or floral volatiles that are emitted by flowers can be detected using Gas Chromatography-Mass Spectrometry (GC-MS) which has specific database containing mass spectra of isolated and well-studied compounds (Knudsen and Gershenzon, 2006).

In 1990s, research on floral scent was focused on isolation, structural elucidation of fragrance compounds as well as their chemical synthesis mechanisms in order to fulfil high

demand from perfumery and food industries (Knudsen, 1993). Significant research on natural fragrance compounds are very important where the information on their precursors and compound derivatives is very important in production of synthetic perfume that mimic natural floral scent such as rose, jasmine, lavender and petunia for commercialization in perfumery industry (Verdonk *et al.*, 2003; Guterman *et al.*, 2002; Zuker *et al.*, 1998; Kaiser, 1993). Unfortunately, commercialisation of plants in different regions of the world has faced problem on the subsequent progeny production due to the lack of natural local pollinators while traditional breeding of flowering plants to produce better hybrid features has also faced problems to attract their natural pollinators due to the drastic changes of their morphology and biochemistry (Pichersky and Dudareva, 2007). In addition, effort to bring their natural pollinators into the new areas or territory was also unsuccessful the failure of their natural pollinators to adapt to the new environment (Buchmann and Nabhan, 1996).

Thus, introduction of scent engineering approach by transforming fragrance-related genes into plant of interest has opened a new opportunity for pollination where local pollinators that are attracted to the scent emitted by the plants, may visit the plants for honey hunting and indirectly become their pollinating agents (Pichersky and Dudareva, 2007). Besides that, transformation of fragrance-related genes into hybrid plants might restore the original scent produced by their mother plants where original scent from their mother plants had been suppressed due to previous selective breeding that focused on traits other than their fragrant characteristic (Spitcer *et al.*, 2007). Moreover, introduction of fragrance biosynthesis and emission in scentless flowering plants by genetic engineering has been reported to increase their commercial values in cut flower industry (Verdonk *et al.*, 2003).

2.4 FRAGRANCE BIOSYNTHESIS MECHANISMS

Studies on fragrance biosynthetic pathways via biochemical and molecular biology approaches have revealed that the complex mixture of floral volatile from various flowering plants consist of compounds derived from terpenoid, benzenoid and phenylpropanoid pathways (Pichersky and Dudareva, 2007). Terpenoid pathway that has been reported to

occur in both plastid and cytosol of plants is responsible for the presence of a wide number of terpenoid compounds including monoterpenoids (C₁₀ compounds), sesquiterpenoids (C₁₅ compounds) and diterpenoids (C₂₀ compounds) (Knudsen and Gershenzon, 2006). Examples of monoterpenoid compounds that have been identified as the constituents of floral scent are linalool, ocimene, β -pinene, citral and geraniol while germacrene D, nerolidol, farnesene, copaene and caryophyllene are some of the natural sesquiterpenoids that have been identified to be the constituents of various scents produced by different species of flowering plants (Knudsen and Gershenzon, 2006). Besides terpenoid pathway, benzenoid and phenylpropanoid biosynthetic pathways comprise more than 300 compounds including methylbenzoate, methylsalicylate, phenylacetaldehyde, phenylethyl acetate, benzyl acetate, phenylethanol, eugenol and isoeugenol which are considered as the second largest contributor for floral scent (Pott *et al.*, 2002; Raguso *et al.*, 2003; Verdonk *et al.*, 2003). In addition, there is another important class of floral scent compounds known as fatty acid derived compounds such as hexanol, hexanal, nonanal, pentadecane, decanal and dodecanal that are synthesised via lipoxygenase pathway and by fatty acid degradation that have been reported to contribute extensively for the constituent of floral scent (Knudsen and Gershenzon, 2006). Volatile of fatty acid derivatives are mostly detected in vegetative tissues in high levels, specifically playing a very important role in plant defense (Schaller, 2001). However this class of compound has also been reported as minor constituent of floral scent that might contribute to their odour that specifically participate in attracting pollinators for pollination purpose (D'Auria *et al.*, 2007).

Volatile compound identification that has been focused by researchers in early 1990s has led to fragrance-related enzyme isolation and characterization starting from the middle of 1990s whereby *in vitro* characterisation has been carried out by purification of specific enzymes (Guterman *et al.*, 2002; Vainstein *et al.*, 2001). Many fragrance-related enzymes have been identified and characterised including linalool synthase, eugenol synthase, salicylic acid carboxyl methyltransferase and geranyl diphosphate synthase from *C. breweri*, phenylacetaldehyde synthase and eugenol synthase from *P. hybrida* as well as ocimene synthase and mycrene synthase from *A. majus* (Chen *et al.*, 2003; Dudareva *et al.*, 2003; Nagegowda *et al.*, 2008; Pichersky *et al.*, 1995). After that, with the development of

molecular biology approach, a lot of work has been carried out employing Expressed Sequence Tags (ESTs) library construction for identification of fragrance-related transcripts from scented flowers including *R. hybrida*, *P. hybrida* and fragrant orchids such as *Phalaenopsis bellina* and *V. Mimi Palmer* (Hsiao *et al.*, 2006; 2008; Teh *et al.*, 2011). In recent years, development of Next Generation Sequencing (NGS) approach that can directly sequence total cDNA that has been transcribed from polyA⁺ mRNA has enhanced the identification of fragrance-related transcripts involved in fragrance biosynthetic pathway in scented flower with much cheaper cost, faster speed and larger output of data compared to conventional EST libraries that are extensively studied in the previous decade (Graciet *et al.*, 2014).

2.5 BENZENOID AND PHENYLPROPANOID BIOSYNTHETIC PATHWAYS

Benzenoid and phenylpropanoid biosynthetic pathway (Figure 2.2) that has shown a high contribution in fragrance compound biosynthesis produces the second largest class of volatile organic compounds including benzenoids, phenylpropanoids and their derivatives including alcohols and esters (Gang *et al.*, 2001). This pathway involves the main important precursor which is phenylalanine, an aromatic amino acid that is synthesised via shikimate pathway, one of the most important pathways for plant secondary metabolite biosynthesis (Boatright *et al.*, 2004). Biochemical compounds synthesised via this pathway are categorised into two major groups based on their carbon skeleton structure whereby benzenoids are referred to C6-C1 backbone while C6-C2 backbone is categorised as phenylpropanoid related compounds (Knudsen and Gershenzen, 2006). Among available benzenoid and phenylpropanoid compounds, about 24% of the compounds have been found to be derived from phenylalanine compound (Pichersky and Dudareva, 2007). Meanwhile the remaining compounds (76%) are derived from non-phenylalanine compounds including benzylbenzoate, benzylacetate, methylbenzoate and phenylethyl acetate (Boatright *et al.*, 2004; Pichersky and Dudareva, 2007).

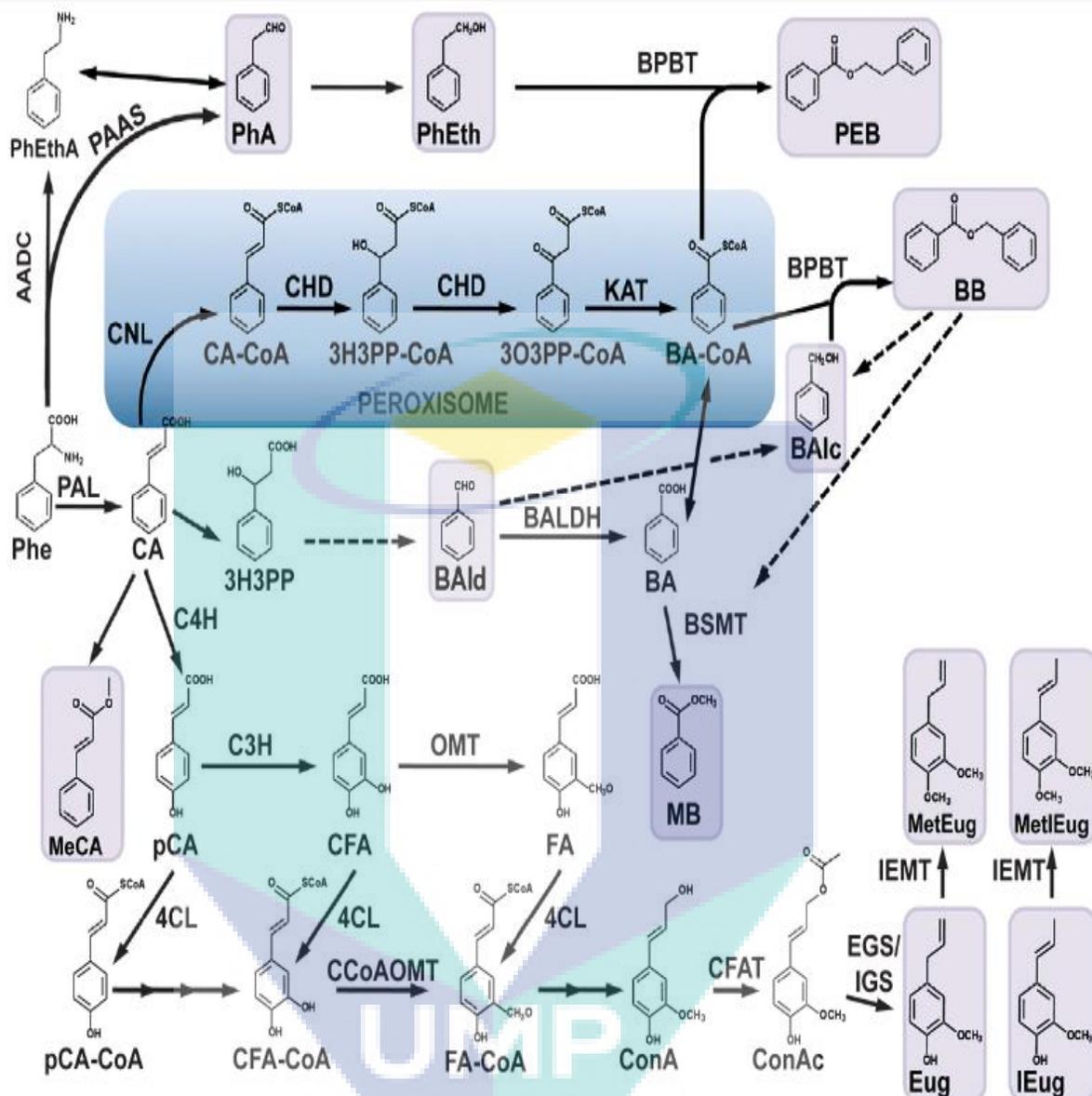


Figure 2.2: Benzenoid and Phenylpropanoid Biosynthetic Pathway in well-studied scented flowers. This figure was adopted from Dudareva *et al.*, 2013. Solid arrows indicate established biochemical steps, while hypothetical steps are represented by broken arrows. Stacked arrows illustrate the involvement of multiple enzymatic reactions. Volatile benzenoid/phenylpropanoid compounds are highlighted in boxes.

In benzenoid and phenylpropanoid pathway, biosynthesis of fragrance compounds is initiated by a well characterised enzyme which is *L*-phenylalanine ammonia lyase that is very important in catalysing the deamination of *L*-phenylalanine to trans-cinnamic acid. The step is followed by benzaldehyde formation via C2 shortening of trans-cinnamic acid. Formation of benzenoids (C6-C1) from cinnamic acid involves shortening of the propyl side chain by two carbons and was shown to proceed either via β -oxidative pathway, non- β -oxidative pathway or a combination of both routes (Boatright *et al.* 2004; Orlova *et al.* 2006). The β -oxidative pathway has only recently been fully elucidated in the flower of *P. hybrida* and appears to resemble to fatty acid catabolism and certain branched chain amino acids that are localized in peroxisomes. The pathway begins with an activation of cinnamic acid to cinnamoyl-CoA (CoA thioester), which undergoes hydration, oxidation and cleavage of the β -keto thioester, resulting in subsequent formation of benzoyl-CoA (Van Moerkercke *et al.*, 2009; Klempien *et al.*, 2012; Qualley *et al.*, 2012). Benzoyl-CoA which is localised in peroxisomes for β -oxidative pathway might be transported to cytosol using a special mechanism for benzyl benzoate and phenylethyl benzoate biosynthesis (Kaminaga *et al.*, 2006).

In contrast to benzenoids and phenylpropanoids, biosynthesis of volatile phenylpropanoid related compounds (C6-C2), such as phenylacetaldehyde and 2-phenylethanol has been reported to be synthesised via another route whereby phenylalanine is utilized as the main substrate to phenylacetaldehyde synthase as a competitor to cinnamic acid route (Kaminaga *et al.*, 2006; Tieman *et al.*, 2007). Interestingly, the genes involved in both phenylacetaldehyde and 2-phenylethanol biosynthesis have been isolated and characterised (Hirata *et al.*, 2012). In addition, phenylacetaldehyde biosynthesis has been reported to be different in both petunia (*P. hybrida*) and rose (*R. hybrida*) petals in the presence of two main routes which are decarboxylation-amine oxidation reaction and decarboxylation of formed phenylpyruvate intermediate (Kaminaga *et al.*, 2006; Sakai *et al.*, 2007). Specifically, conversion of phenylacetaldehyde to 2-phenylethanol is catalysed by a phenylacetaldehyde reductase that has been reported to be present in rose flowers while decarboxylation reaction catalysed by phenylacetaldehyde synthase (PAAS) has been detected to occur in petals of both *P. hybrida* and *R. hybrida* (Kaminaga *et al.*, 2006; Farhi

et al., 2010). Besides that, another alternative route has been reported whereby deamination process of phenylalanine by aromatic amino acid aminotransferase first occurs to form phenylpyruvate as intermediate compound and is subsequently followed by decarboxylation of phenylpyruvate to phenylacetaldehyde, representing second alternative route that only occurs in *R. hybrida* (Farhi *et al.*, 2010). Other than that, the deamination of phenylalanine in tomato also occurs via two separate steps, where it is first converted to phenylethylamine by an aromatic amino acid decarboxylase and then, the action of a hypothesised amine oxidase, dehydrogenase, or transaminase for the formation of phenylacetaldehyde (Tieman *et al.*, 2006). Meanwhile, Gonda *et al.* (2010) discovered that melon fruit (*Cucumis melo* L.) has a third enzymatic route whereby transaminated phenylalanine is converted to its corresponding α -keto acid, phenylpyruvate and subsequently followed by decarboxylation to phenylacetaldehyde.

Meanwhile, formation of floral volatile phenylpropanoids (C6-C3) such as eugenol, isoeugenol, methyleugenol, methyl-isoeugenol, chavicol and methylchavicol has been identified to share the same initial biosynthetic steps with lignin biosynthetic pathway up to the step of coniferyl alcohol biosynthesis (Koeduka *et al.*, 2006). Subsequent esterification step of coniferyl alcohol to coniferyl acetate that has been identified to occur in petals of *P. hybrida* is catalysed by a coniferyl alcohol acetyltransferase (Dexter *et al.*, 2007). Coniferyl acetate is subsequently converted to phenylpropanoid compounds which are eugenol and isoeugenol by eugenol synthase (EGS) and isoeugenol synthase (IGS), respectively. EGS and IGS have been reported to belong to pinoresinol-lariciresinol reductase, isoflavone reductase and phenylcoumaran benzylic ether reductase (PIP) family of NADPH dependent reductases (Koeduka *et al.*, 2006; 2008).

Furthermore, the diversification of phenylpropanoid and benzenoid compounds is further increased by modification process on the compounds by methylation, hydroxylation and acetylation reaction that enhance their volatility or olfactory properties of scent compounds (Muhlemann *et al.*, 2014). In addition, further methylation reactions on the modified benzenoids and phenylpropanoids are catalysed by either O-methyltransferases (OMTs) or carboxyl methyltransferases to eugenol, isoeugenol and chavicol for the

downstream production of methyleugenol, isomethyleugenol and methylchavicol (Gang *et al.*, 2002). Meanwhile, O-methyltransferases (OMTs) have been reported to be responsible for the biosynthesis of a diverse array of benzenoids and phenylpropanoids including veratrole in white campion flower (*Silene latifolia*) (Gupta *et al.*, 2012; Akhtar and Pichersky, 2013), 3,5- dimethoxytoluene and 1,3,5-trimethoxybenzene in roses (Lavid *et al.*, 2002; Scalliet *et al.* 2002) as well as methyleugenol and isomethyleugenol in *C. breweri* (Wang and Pichersky, 1998).

Besides that, in benzenoid and phenylpropanoid pathway, *S*-adenosyl-*L*-methionine: salicylic acid carboxyl methyltransferase (SAM:SAMT), *S*-adenosyl-*L*-methionine: benzoic acid carboxyl methyltransferase (SAM:BAMT) and theobromine synthase (SABATH) are well-known as carboxyl methyltransferases (D'Auria *et al.*, 2003) responsible for volatile ester biosynthesis such as methylbenzoate in *A. majus* and *P. hybrida* flowers (Murfitt *et al.*, 2000; Negre *et al.*, 2003), as well as methylsalicylate in *C. breweri* and *P. hybrida* (Ross *et al.*, 1999; Negre *et al.*, 2003). In addition, enzymes from the benzyl alcoholacetyl-,anthocyanin-O-hydroxy-cinnamoyl-,anthranilate-N-hydroxy-cinnamoyl/benzoyl deacetyl vindoline acetyl trans ferase (BAHD) superfamily of acyltransferases including acetyl CoA:(*Z*)-3-hexen-1-ol acetyltransferase, benzylalcohol O-acetyltransferase, benzoyl-CoA benzyl alcohol phenylethanol benzoyltransferase, and benzoyl-CoAbenzylalcohol O-benzoyltransferase (D'Auria, 2006) has been shown to be responsible for acetylation of scent compounds such as benzyl acetate in *C. breweri* (Dudareva *et al.*, 1998), benzoyl benzoate in *C. breweri* and *P. hybrida* (D'Auria *et al.*, 2002; Boatright *et al.*, 2004; Orlova *et al.*, 2006), and phenylethyl benzoate in *P. hybrida* (Boatright *et al.*, 2004; Orlova *et al.*, 2006).

2.6 SCIENTIFIC STUDIES ON FRAGRANT-ORCHIDS

Identification of biochemical compounds in complex mixture of fragrant-orchid volatiles has been initiated in late 1980s using Gas Chromatography-Mass Spectrometry (GC-MS) (Kaiser, 1993). A scientific book on orchid scent by Kaiser (1993) has shown GC-MS analysis data of more than 180 orchid species and hybrids including *Cattleya*

araguaiensis, *Cymbidium formosanum*, *Dendrobium carniferum*, *Dendrobium superbum*, *Oncidium curcutum*, *Phalaenopsis violacea* and *Vanda tessellata*. While in 2006, scent constituent of *Phalaenopsis bellina* was reported by Hsiao *et al.* (2006) and followed by V. Mimi Palmer (Mohd-Hairul *et al.*, 2010a) whereby most of their scents are dominated by terpenoids as well as benzenoids and phenylpropanoids. Monoterpenoids and sesquiterpenoids are the highly distributed compounds among orchid species based on the GC-MS analysis on scented orchids including linalool, myrcene, ocimene, germacrene D and nerolidol. Besides terpenoids, benzenoids and phenylpropanoid compounds are also found to be described among in scented orchids including methylbenzoate, benzyl benzoate, benzyl acetate, phenylethyl acetate, eugenol and isoeugenol. In addition, there were traces of other compounds detected in scented orchids such as fatty acid derivatives, indole and formanilide (Kaiser, 1993; Hsiao *et al.*, 2006). More recently, Gallego *et al.* (2012) reported that scent constituents of three orchid species namely *Himantoglossum robertianum*, *Ophrys apifera* and *Gymnadenia conopsea* containing 106 different volatile compounds have been identified in their scent. Among the volatile compounds, 54% of the compounds have been confirmed including α -pinene, β -pinene and limonene in *Himantoglossum robertianum*, 1-butanol, butyl ether and caryophyllene in *Ophrys apifera* and phenethyl acetate, eugenol and benzaldehyde in *Gymnadenia conopsea*.

Meanwhile, knowledge on molecular biology aspects of fragrance biosynthesis in orchid plants was firstly reported on *Phalaenopsis bellina* by Taiwanese researchers by identifying fragrance-related transcripts from constructed cDNA library (Hsiao *et al.*, 2006). Subsequently, characterisation of a monoterpene synthase geranyldiphosphate synthase (PbGDPS) that lack DXXD motif has been reported for the same orchid (Hsiao *et al.*, 2008). In addition, expression analysis using real-time PCR and Northern-blot has shown that PbGDPS gene expression is developmentally regulated whereby the transcript level has been detected to increase gradually once the bud opened and reached the highest peak on the fifth-day after bud-opening, and followed by a gradual decrease of its expression towards the end of the flower life-cycle. Subsequently, several fragrance-related transcripts have been identified from V. Mimi Palmer, a well-known fragrant-orchid that received several international awards for its strong sweet smelling fragrant (Chan *et al.*,

2009; 2011; Mohd-Hairul *et al.*, 2010b; Mohd-Hairul, 2011; Teh *et al.*, 2011; Song *et al.*, 2012). More recently, another work has been reported on isolation and characterisation of a eugenol synthase transcript from an orchid species *Gymnadenia odoratissima* where functional expression of the transcript in *Escherichia coli* system has shown that the recombinant enzyme was capable of catalysing the biosynthesis of eugenol from its substrate coniferyl acetate (Gupta *et al.*, 2014).

2.7 MOLECULAR AND BIOCHEMICAL STUDIES ON THE SCENT OF VANDA MIMI PALMER

Biochemical analysis on the scent of *V. Mimi Palmer* has been reported by Mohd-Hairul *et al.* (2010a) whereby the volatile comprises high level of benzenoids as well as phenylpropanoid compounds including methylbenzoate, benzylacetate, phenylethyl acetate and phenylethyl alcohol besides terpenoid compounds including linalool, ocimene and nerolidol. Besides that, emission of volatile from the orchid hybrid has been reported to be developmentally regulated, in the same pattern of *Phalaenopsis bellina* that has been reported by Hsiao *et al.* (2006). In addition, emission pattern of volatiles has been reported to be time regulated whereby the volatile emission has been detected at very low level at 6.00am, and increased gradually until it reached the highest peak in the afternoon and decreased gradually in the evening. The volatile emission could not be detected at night (Mohd-Hairul *et al.*, 2010a).

In molecular biology aspects, a floral cDNA library has been constructed by Chan *et al.* (2009) and several fragrance-related transcripts have been identified from the library covering both terpenoid as well as benzenoid and phenylpropanoid biosynthetic pathway (Chan *et al.*, 2009; Mohd-Hairul, 2011; Teh *et al.*, 2011). From their findings, the expression of most fragrance-related transcripts showed the same expression pattern whereby the transcripts level in the floral tissues are developmentally regulated. The expression of fragrance-related transcripts including *V. Mimi Palmer* 1-deoxyd-xylulose 5-phosphate reductoisomerase (VMPDXR), *V. Mimi Palmer* phenylacetaldehyde synthase (VMPPAAS), *V. Mimi Palmer* alcohol acyltransferase (VMPAAT), *V. Mimi Palmer*

sesquiterpene synthase (VMPSTS), *V. Mimi Palmer* 4-(cytidine 5 -diphospho)-2-C-methyl-d-erythritol kinase (VMPCMEK), *V. Mimi Palmer* cytochrome P450 (VMPCyP450) and *V. Mimi Palmer* Acetyl-CoA-C-acetyltransferase (VMPACA) started to be detected during the opening of the bud, increased gradually until reached the highest peak during fully-open flower stage and decreased gradually until their senescence. Interestingly, expression analysis on the fragrant-related transcripts in different tissues has shown that the transcripts have been detected at very high level in petal and sepal compared to bud and vegetative tissues. The results suggested that sepals of the orchid are playing the same role as petals for fragrance biosynthesis in which they share the same morphological structures and functions. Recently, genetic engineering work on *V. Mimi Palmer* Sesquiterpene synthase (VMPSTS) was successfully carried out by expressing the transcript in *L. lactis*, a gram positive bacterium (Song *et al.*, 2012) where the sesquiterpene synthase has been found to be responsible in catalysing the biosynthesis of Germacrene D as the major compound besides some other sesquiterpene compounds as its by-products including nerolidol and copaene.

2.8 ECONOMIC IMPORTANCE OF FRAGRANT-ORCHIDS

For orchid, trait such as floral scent is a primary novel marker because it is a key determinant of consumer choice (Hsiao *et al.*, 2006). However, traditional breeding programmes have caused the loss of the scent of orchid varieties. Many orchids have been focused on other non-fragrance purpose like improving the visual aesthetic and vase life, and for producing either cut-flower or ornamental-type orchids (Hsu *et al.*, 2011). Orchid industry thrives on novelty for producing high quality varieties which also focused on colour and shape of orchid flower as well as their floral scent (Hsiao *et al.*, 2011). Studies on orchid scent biology is difficult due to different odours, genome size, period of life cycle, regeneration time and inefficient transformation systems (Chang *et al.*, 2011). For production of orchid hybrids, several scented and scentless orchids which are cross-incompatible probably lead to restriction of scented progeny, either with diluted scent or complete loss of the scent characteristics (Hsiao *et al.*, 2008). Other than that, fragrance biosynthetic pathway of orchids remains yet to be understood because limited studies have

been carried out so far, especially in discovery of scent-related enzymes and genes in monocotyledonous plants such as orchids generally, and vandaceous orchids specifically (Teh *et al.*, 2011). In addition, new scented orchid varieties may have been developed successfully for cut-flower, but, the rational design of suitable choice of species is also important for genetic manipulation in order to understand the composition of scent components, their specific fragrance biosynthetic pathways, key scented enzymes and other few related features (Hsiao *et al.*, 2011).

More recently, orchids have become the centre of attention to new areas of plant research, including genetic engineering, functional genomics, proteomics and metabolomics (Hsu *et al.*, 2011; Chang *et al.*, 2011), which require advanced transgenic strategies (da Silva *et al.*, 2011) to achieve their different goals. The successful application of these new approaches will help to further improve orchids and orchid products. However, different situation happened in cut-flower industry when breeding of cultivated flowers has been extensively studied in order to improve their vase life, shipping characteristics and visual aesthetic values such as shape and colour, leading to the lost of floral scent originality eventually (Vainstein *et al.*, 2001). Moreover, due to the rising demand and popularity of orchids, it is nowadays a multimillion dollar business, primarily as pot plants and cut flower stalks (Winkelmann *et al.*, 2006). To meet the demand on orchids in the future, as well as to conserve their biodiversity, the development and deployment of new technologies are very important for improving flower quality, resistance to biotic and abiotic stresses, and to rapidly produce mass propagation members of this phylum (Raubeson *et al.*, 2005).

CHAPTER 3

MATERIALS AND METHODS

3.1 PLANT MATERIALS

Mother plants with fully-opened flowers of selected fragrant-orchids namely *V. Mimi Palmer*, *V. Small Boy Leong* and *Vandachostylis Sri-Siam* together with a non-fragrant orchid *V. Tan Chay Yan* used in this study were bought from the United Orchid Plants, an orchid nursery located in Rawang, Selangor. Samples for RNA extraction which are fully-open flower from each selected fragrant and non-fragrant orchids were detached from their mother plants. Each sample of 2 g was wrapped in aluminium foil and followed by quickly frozen in liquid nitrogen. All the samples were then stored at -80 °C freezer prior to total RNA extraction.

3.2 ISOLATION OF TOTAL RNA

Total RNA was extracted from fully-open flowers of *V. Mimi Palmer*, *Small Boy Leong*, *Vandachostylis Sri-siam* and *V. Tan Chay Yan* separately following RNA extraction procedure established by Yu and Goh method (2000) with slight modifications (Chan *et al.*, 2009). 2 g of fully-open flowers for each sample that have been stored at -80 °C freezer for less than a week were quickly transferred into a special container containing liquid nitrogen. Each sample was ground separately using mortar and pestle in the presence of liquid nitrogen until the flowers became fine powder form. All the ground samples were then transferred into a 50 ml Nalgene centrifuge tube and mixed thoroughly in 20 ml of pre-warmed (65 °C) CTAB RNA extraction buffer [100 mM Tris-Cl (pH 7.5) containing

2M NaCl, 2 % (w/v) hexadecyl (cetyl) trimethyl ammonium bromide (CTAB), 20 mM EDTA (pH 8), 2 % (w/v) polyvinylpyrrolidone (PVP), and 4 % (v/v) β -mercaptoethanol]. All the samples were then incubated at 65 °C for 20 minutes and shaken vigorously every 5 minutes. Next, a centrifugation step was carried out at 12,857 x g for 15 minutes at 4 °C. After the centrifugation step, recovered supernatant for each sample was mixed vigorously with an equal volume of chloroform:isoamylalcohol (24:1) and followed by another centrifugation at 12,857 x g, 4 °C for 15 minutes. After the centrifugation step, the top aqueous phase of each sample was carefully transferred into fresh 50 ml Nalgene centrifuge tubes and followed by chloroform:isoamylalcohol (24:1) extraction twice. Final recovered aqueous phase of each sample was then subjected to overnight precipitation in 2 M lithium chloride (LiCl) at 4 °C overnight. After the overnight precipitation, total RNA for each sample was recovered by a centrifugation step at 12,857 x g, 4 °C for 30 minutes. The recovered total RNA pellet at the bottom of the tube was then rinsed in cold 80 % (v/v) ethanol and air-dried at room temperature to remove excess ethanol. Subsequently, each total RNA pellet was dissolved in 200 μ l of diethyl pyrocarbonate (DEPC)-treated water. The isolated RNA samples were then stored at -20 °C for further use. The yield and purity of each total RNA sample were quantified using a NanoVue Spectrophotometer (GE, Healthcare, UK, 2010). While the integrity of the total RNA samples were verified using formaldehyde denaturing agarose gel electrophoresis.

3.2.1 Formaldehyde Denaturing Agarose Gel Electrophoresis

Formaldehyde agarose gel solution containing 1.2 % (w/v) agarose gel was prepared by boiling 0.6 g agarose powder in 50 ml of 1 X TAE buffer (Appendix A) containing 1 X F buffer [20 mM MOPS buffer (pH 7) containing 1 mM EDTA and 5 mM NaOAc] and 0.2 X of Gel-red solution (Biothium, USA). The boiled agarose gel solution was allowed to cool to 50 °C under running tap water before adding 37 % (v/v) formaldehyde solution for a final concentration of 6 % (v/v). The formaldehyde agarose gel solution was mixed by swirling and poured into a gel cast to be solidified. While waiting for the gel to be solidified, 1 μ g of total RNA for each sample was added into RNA sample buffer [1X F buffer containing 50 % (v/v) formamide and 6 % formaldehyde (v/v)] and

followed by addition of a final concentration of 0.24 X loading dye [0.25 % (w/v) bromophenol blue, 0.25 % xylene cyanol, 30 % (v/v) glycerol]. All the samples were then incubated at 65 °C for 10 minutes to denature any secondary structure of RNA in all the samples. All the prepared RNA samples were then quickly chilled on ice and loaded into the solidified formaldehyde agarose gel. Gel electrophoresis was carried out in a formaldehyde running buffer [1X F buffer containing 6 % (v/v) formaldehyde] at 60 Volts for 45 minutes. The electrophoresed RNA samples were then visualized under a UV gel documentation system (Alpha Imager, USA, 2009).

3.2.2 Isolation of PolyA⁺ mRNA

All the total RNA samples were subjected to polyA⁺ mRNA isolation using the PolyAtract[®] mRNA Isolation Systems III (Promega, USA). 1 mg of each total RNA sample was adjusted to a final volume of 500 µl by adding nuclease-free water provided in the kit and followed by incubation at 65 °C for 10 minutes. After that, annealing reaction was carried out by adding 150 pmol of biotinylated oligo(dT) probe and Sodium Saline Citrate (SSC) buffer (pH 7) to a final concentration of 0.5 X (diluted from 20 X SSC provided in the kit) to the RNA solution to allow the oligomers to anneal with polyA⁺ mRNA. The mixture of each sample was then mixed gently and allowed to completely cool at room temperature.

While waiting for the RNA sample to completely cool, streptavidin-paramagnetic particles (SA-PMPs) (provided in the kit) were subjected to a washing step by gently flicking the bottom of the tube until the particles were completely dispersed. After that, a step of particles capturing were carried out by placing the respective tubes on a magnetic stand for 30 seconds to accumulate all the particles at the wall of each tube. Next, recovered supernatant was carefully discarded and followed by another three times washing steps with 300 µl of 0.5 X SSC buffer (diluted from 20 X SSC provided in the kit). For each washing step, the particles were flicked four to five times and captured using the magnetic stand for 30 seconds as described above. The washed SA-PMPs particles were subsequently resuspended in 100 µl of 0.5 X SSC buffer.

Total RNA samples from the annealing reaction was then added into tubes containing washed SA-PMPs particles, followed by incubation at room temperature for 10 minutes with gentle mixing every two minutes. Each mixture containing SA-PMPs and total RNA was then captured on the magnetic stand for 30 seconds. After the capturing step, supernatant of each sample was discarded and the remaining pellet of SA-PMPs particles bound with polyA⁺ mRNA was subjected to four times washing steps with 300 µl of 0.1 X SSC buffer (diluted from 20 X SSC provided in the kit). For each washing step, the bottom of the each tube was gently flicked until the SA-PMPs particles were completely dispersed. Subsequently, the final supernatant was removed as much as possible without disturbing the SA-PMPs particles.

After that, elution step of polyA⁺ mRNA from the recovered SA-PMPs particles was carried out for each sample. In this elution step, the SA-PMPs particles were resuspended by gently flicking in 100 µl of RNase free water and followed by a magnetically capturing step for 30 seconds. Next, the eluted polyA⁺ mRNA solution was transferred carefully into sterile RNase free microcentrifuge tubes without disturbing the SM-PMPs pellet at the bottom of each tube. The recovered polyA⁺ mRNA was then subjected to a precipitation step by adding a final concentration of 0.3 M sodium acetate (pH 5.2) and 80 % (v/v) isopropanol. The precipitation was carried out overnight at -20 °C. After the overnight incubation, the mixture was centrifuged at 12,857 x g at 4 °C for 20 minutes. Subsequently, pellet of mRNA for each sample at the bottom of microcentrifuge tubes was then washed with 75 % (v/v) ethanol and allowed to air-dry at room temperature. The mRNA pellet was then dissolved in 20 µl RNase-free water and stored at -20 °C for subsequent use. The concentration and purity of the eluted polyA⁺ mRNA for each sample was measured using Nanovue spectrophotometer (GE Healthcare, UK, 2010).

3.3 SYNTHESIS OF DOUBLE-STRANDED cDNA

The isolated polyA⁺ mRNA samples were subjected to double-stranded cDNA synthesis using the Universal Riboclone[®] cDNA synthesis system (Promega, USA). The cDNA synthesis procedure is divided into two major parts which were first-strand cDNA

synthesis and second-strand cDNA synthesis. In first-strand cDNA synthesis, 1 µg of oligo (dT) was added to 2 µg of mRNA followed by a final volume adjustment to 15 µl by adding nuclease-free water (provided in the kit). The mixture was then incubated at 70 °C for 10 minutes and quickly chilled on ice. Then, 40 units of RNAsin® and 5 µl of 5 X first-strand buffer (provided in the kit) was added into the mixture and followed by heating at 42 °C for 5 minutes. After the incubation period, 30 units of AMV reverse transcriptase and 2 mM of sodium pyrophosphate were added into the mixture followed by incubation at 42 °C for 60 minutes. The synthesized first-strand cDNA was then used as template for the synthesis of second-strand cDNA.

In second strand cDNA synthesis, 40 µl of 2.5 X second-strand buffer (provided in the kit) was added into the synthesized first-strand cDNA template together with 25 units of DNA polymerase, 5 µg of acetylated bovine serum albumin (BSA) and 0.8 units of RNase H. After that, the volume of second-strand cDNA synthesis reaction was adjusted to a volume of 100 µl by adding nuclease-free water. The synthesis of second strand cDNA from the synthesized first-strand cDNA was carried out by incubation at 14 °C in a mastercycler (Biorad, USA) for 2 hours followed by heating at 70 °C for 10 minutes to stop the reaction. Then, 0.2 unit of T4 DNA polymerase was pipetted into the reaction mixture and followed by incubation at 37 °C for 10 minutes to produce blunt end of 3'- and 5'-end of the overhanged synthesized double-stranded cDNA. The blunting reaction was then terminated by adding 20 mM EDTA into the reaction mixture.

The blunt end double-stranded cDNA of each sample was then subjected to purification by adding an equal volume of phenol: chloroform: isoamyl-alcohol (25:24:1) into the double-stranded cDNA of each sample. Subsequently, the mixture was mixed by vortexing briefly followed by centrifugation at room temperature at 15,490 x g for 60 seconds. The top aqueous phase of each cDNA sample was then transferred carefully into a fresh 1.5 ml microcentrifuge tube followed by adding 0.3 M sodium acetate (pH 5.2) and 2 volumes of ice-cold absolute ethanol for precipitation purpose. Each sample was then mixed gently by inverting four times and incubated at -80 °C for 30 minutes for precipitation of purified double stranded cDNA. After the incubation period, another

centrifugation step was carried out at 15,490 x *g* for 20 minutes at 4 °C. The supernatant of each sample was then discarded and the recovered pellet of purified double stranded cDNA of each sample was then subjected to a washing step with ice-cold 70 % (v/v) ethanol. The washing step was carried out by adding 300 µl of ice-cold 70 % (v/v) ethanol into each tube and followed by a centrifugation step at 4 °C at 15,490 x *g* for 10 minutes. The supernatant was then discarded and another washing step with ice-cold absolute ethanol was carried out in the same manner. After that, all the tubes containing recovered purified double-stranded cDNA pellet of each sample was air dried at room temperature for 30 minutes to remove any excess ethanol. After totally dried, all of the pellets of double-stranded cDNA for each sample were diluted in 20 µl of Tris-EDTA (TE) buffer, pH 8 by pipetting. Concentration and purity of purified double-stranded cDNA samples were then determined using the NanoVue™ Plus Spectrophotometer (GE Healthcare, UK, 2010). All the purified double stranded cDNA samples were then stored at -20 °C freezer for further use.

3.4 cDNA-REPRESENTATIONAL DIFFERENCE ANALYSIS (cDNA-RDA)

In this study, cDNA representational difference analysis (cDNA-RDA) was carried out in forward way. The cDNA of *V. Mimi Palmer*, *Vandachostylis Sri-Siam* and *V. Small Boy Leong* (fragrant-orchids) were designated as testers cDNA (the source of the sequences of interest to be isolated) while *V. Tan Chay Yan* (non-fragrant orchid) was chosen as a driver cDNA (the source of the sequences to be eliminated) in this cDNA-RDA experiment (Table 3.1). Figure 3.1 showed the tester and driver cDNA were subjected to downstream steps such as digestion, ligation and enrichment in three rounds of cDNA-RDA.

3.4.1 Digestion of Testers and Driver cDNA

Each tester and driver cDNA listed in Table 3.1 was digested using three restriction enzymes (*Bgl*II, *Bam*HI and *Hind*III), separately. The digestion reaction of each sample was carried out in a total volume of 50 µl containing 2 µg of double-stranded cDNA, 1X NEB Buffer 3 (New England Biolabs, UK) and 10 U of respective enzymes. For digestion reactions containing *Bam*HI, bovine serum albumin (BSA) was added in a final

concentration of 0.1 mg/ml. All the mixtures were mixed by pipetting followed by incubation at 37 °C for 2 hours. After 1 hour incubation, an additional of 10 U of respective enzyme was added into each reaction mixture to enhance the effectiveness of digestion reaction.

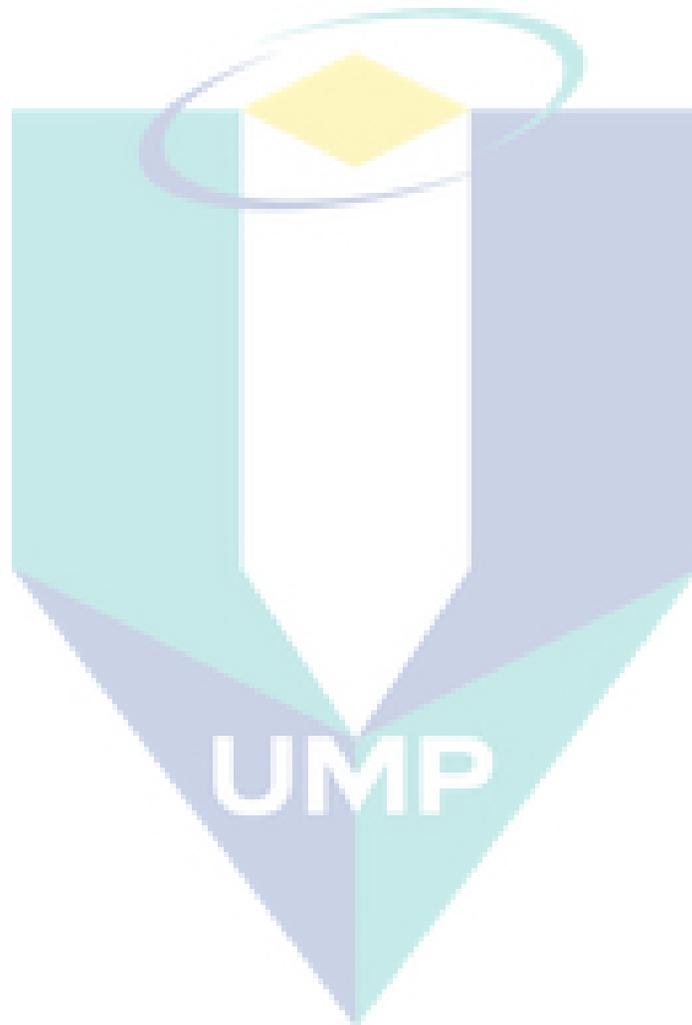
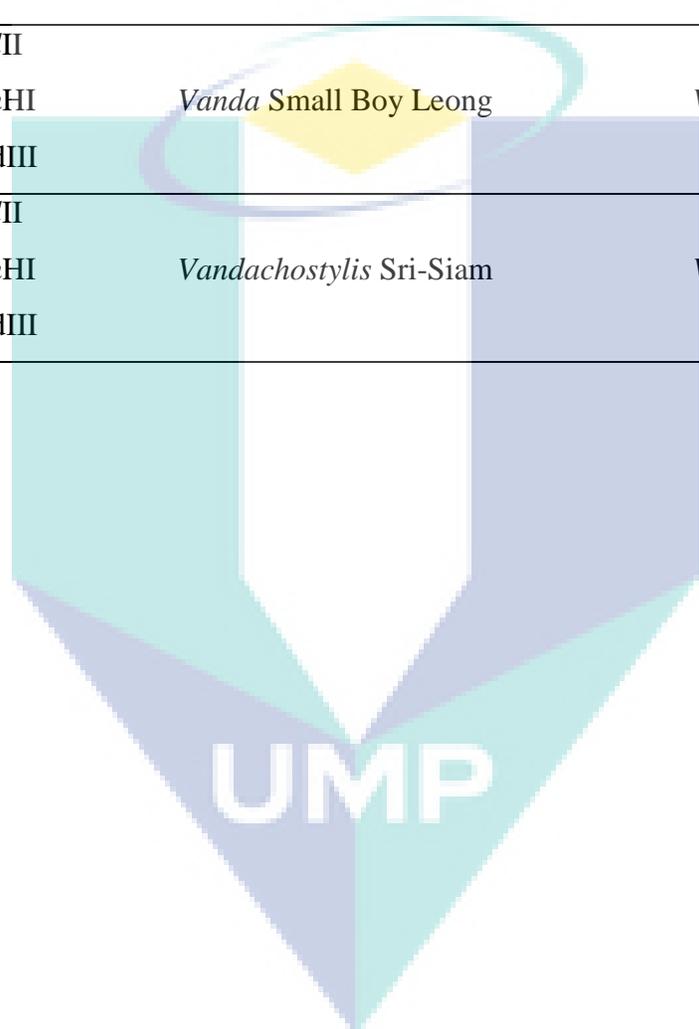


Table 3.1: Three sets of RDA for identification of fragrance-related transcripts

RDA set	RE digestion	Tester cDNA	Driver cDNA
A	<i>Bgl</i> II <i>Bam</i> HI <i>Hind</i> III	<i>Vanda</i> Mimi Palmer	<i>Vanda</i> Tan Chay Yan
B	<i>Bgl</i> II <i>Bam</i> HI <i>Hind</i> III	<i>Vanda</i> Small Boy Leong	<i>Vanda</i> Tan Chay Yan
C	<i>Bgl</i> II <i>Bam</i> HI <i>Hind</i> III	<i>Vandachostylis</i> Sri-Siam	<i>Vanda</i> Tan Chay Yan



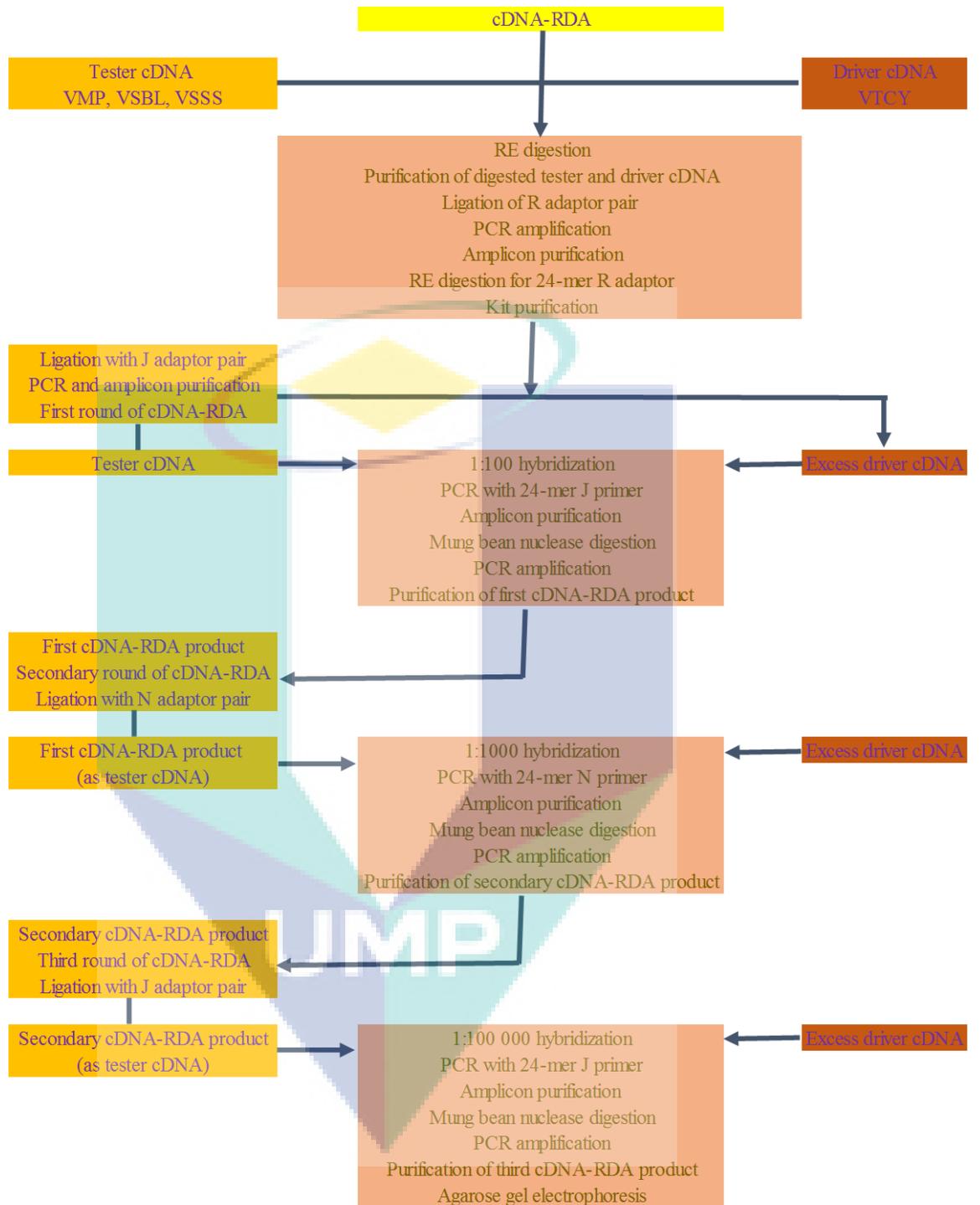


Figure 3.1: Flow chart of cDNA-RDA in this study.

After the incubation period, digestion mixture was then mixed thoroughly by pipetting. Next, the digested cDNA was subjected to purification using an equal volume of phenol: chloroform: isoamyl alcohol (25:24:1). The mixtures were briefly vortexed and followed by a centrifugation step at 15,490 x g for 15 minutes at room temperature. Top aqueous phase of each sample was then transferred into a fresh 1.5 ml microcentrifuge tube separately and subjected to another purification step with equal volume of chloroform: isoamylalcohol (24:1) in the same manner. The final aqueous phase of each sample was then transferred into a fresh 1.5 ml microcentrifuge tube and followed by addition of 0.1 volume of 3.0 M sodium acetate (pH 5.2) and 2.5 volume of ice-cold absolute ethanol. The mixture of each sample was then mixed gently by inverting 4-5 times and followed by incubation at -20 °C for 30 minutes. Subsequently, all the samples were subjected to another centrifugation at 15,490 x g at 4 °C for 20 minutes. The pellet of each sample was subjected to a washing step in ice-cold 70 % (v/v) ethanol. Another washing step was then applied using an ice-cold absolute ethanol in the same manner. Finally, the pellet of each tester and driver was resuspended in 30 µl of TE buffer, pH 8.0 containing 10 mM Tris and 1 mM EDTA. All the purified digested testers and driver cDNA were further subjected to concentration and purity determination using the NanoVue™ Plus Spectrophotometer (GE Healthcare, UK, 2010). All the purified digested testers and driver cDNA were then stored at -20 °C freezer for further use.

3.4.2 cDNA-RDA, Oligonucleotide Adaptors and Primers

Throughout the entire cDNA-RDA experiment for identification of fragrant-related transcripts based on a subtraction approach, three adaptor sets for each restriction enzyme digestion were used (R oligonucleotide pair, J oligonucleotide pair and N oligonucleotide pair) as listed in Table 3.2. The sequences of all of the adaptors were adopted from Lisitsyn and Wigler (1999) where each adaptor pair comprises both 24-mer adaptor (long oligonucleotide) and 12-mer adaptor (oligonucleotide). All the oligonucleotide adaptors used were synthesised and purified using HPLC (Bioneer, Korea). Meanwhile the same sequence of 24-mer adaptors for each set of cDNA-RDA was synthesised by Bioneer,

Table 3.2: Sequence of oligonucleotide adaptors for cDNA-RDA experiment

RE	Adaptors	Name	Sequences 5'-3'
<i>Bgl</i> III	R oligonucleotide pair	R Bgl24	AGCACTCTCCAGCCTCTCACCGAG
		R Bgl12	GATCCTCGGTGA
	J oligonucleotide pair	J Bgl24	ACCGACGTCGACTATCCATGAACG
		J Bgl12	GATCCTCCCTCG
	N oligonucleotide pair	N Bgl24	AGGCAACTGTGCTATCCGAGGGAG
		N Bgl12	GATCCTCCCTCG
<i>Bam</i> HI	R oligonucleotide pair	R Bam24	ACCGACGTCGACTATCCATGAACG
		R Bam12	GATCCGTTTCATG
	J oligonucleotide pair	J Bam24	AGGCAACTGTGCTATCCGAGGGAG
		J Bam12	GATCCTCCCTCG
	N oligonucleotide pair	NBam24	AGCACTCTCCAGCCTCTCACCGAG
		NBam12	GATCCTCGGTGA
<i>Hind</i> III	R oligonucleotide pair	RHind24	AGCACTCTCTCCAGCCTCTCACCGCA
		RHind12	AGCTTGCGGTGA
	J oligonucleotide pair	J Hind24	ACCGACGTCGACTATCCATGAACA
		JHind12	AGCTTGTTTCATG
	N oligonucleotide pair	NHind24	AGGCAGCTGTGGTATCGAGGGAGA
		NHind12	AGCTTCTCCCTC

Korea without HPLC-purification step, to be used as a specific primer during PCR amplification in each round of cDNA-RDA. In this cDNA-RDA experiment, R pair adaptors of each restriction enzyme were used for preparation of driver and testers cDNA in representation stage. On the other hand, J pair adaptors used in the first round of cDNA-RDA was applied to the tester DNA for the first round of RDA. Meanwhile, N pair adaptors were applied for second round of cDNA-RDA that used the first round-difference products as cDNA template. After that, next rounds of cDNA-RDA were carried out by changing between J and N pair adaptors alternately in order to isolate different transcripts that were up-regulated expressed in fragrant-orchid and down-regulated expressed in non-fragrant orchid.

3.4.3 Ligation of Oligonucleotide Adaptors

Preparation of testers cDNA (cDNA of fragrant-orchids) and driver cDNA (cDNA of non-fragrant orchid) was carried out by a ligation reaction of the digested cDNA (refer to section 3.4.1) with their respective pair of R adaptors (refer to Table 3.2). Ligation reaction for each tester and driver was carried in a volume of 30 μ l containing ~50 ng of digested cDNA, 25 μ M of R 12-mer adaptor (short oligonucleotide), 25 μ M of R 24-mer adaptor (long oligonucleotide) and 1 X T4 DNA ligase buffer (New England Biolabs, UK). The mixture of ligation reaction was then incubated at 55 °C in a mastercycler (BioRad, USA) for 3 minutes followed by an incubation step at 16 °C for 60 minutes in the same mastercycler. The mixture were then quickly placed on ice and followed by an addition of 0.7 U T4 DNA ligase (New England Biolabs, UK). The mixture of each tester and driver cDNA was then incubated at 16 °C in the same mastercycler. After the incubation period, the mixture of each tester and driver cDNA was diluted two times in nuclease-free water.

After 16 hours ligation reaction, determination on the success of ligation reaction was carried out using polymerase chain reaction (PCR) approach. The PCR reaction was carried out in a volume of 50 μ l reaction containing 5 μ l of diluted ligation product, 1 X PCR buffer (GeneAll, Korea) and 0.2 mM dNTP mix (GeneAll, Korea). The PCR reaction mixture was initially incubated at 72 °C in a mastercycler for 5 minutes to melt off the R

12-mer oligonucleotide adaptor. The mixture was then quickly chilled on ice followed by addition of 2.5 U of *Taq* DNA polymerase (GeneAll, Korea). After that, each sample mixture was incubated at 72 °C in the same mastercycler for 5 minutes. Finally, 0.5 µM of R 24-mer primer was added into the mixture prior to the actual amplification with the following parameters: 94 °C for 3 minutes; 30 cycles at (94 °C for 1 minute; 72 °C for 3 minutes) and 72 °C for 10 minutes.

3.4.4 Enrichment of Testers and Driver cDNA

Enrichment of testers and driver cDNA was performed using PCR approach in the same manner as described in section 3.4.3. However, total volume of the reaction was doubled (10 tubes x 100 µl). A volume of 5 µl of each PCR product was subjected to agarose gel electrophoresis, while the remaining PCR products were pooled together for each respective driver and tester. All the PCR-amplified products of driver and testers were further subjected to a purification step using phenol: chloroform: isoamylalcohol (25:24:1). An equal volume of phenol: chloroform: isoamylalcohol (25:24:1) was transferred into each PCR product of respective driver and tester and mixed vigorously by vortexing. After that, all the samples were centrifuged at 15,940 x g for 15 minutes. After the centrifugation step, aqueous phase of each sample was transferred into a fresh 1.5 ml microcentrifuge tube. The purification step was then repeated in the same manner using chloroform: isoamylalcohol (24:1). Next, aqueous phase of each purified testers and driver cDNA was subjected to a precipitation step by adding 0.1 volume of 3 M sodium acetate (pH 5.2) and 2.5 volume of absolute ethanol. All the testers and driver cDNA were then incubated at -20 °C for 30 minutes. After the precipitation step, a centrifugation step was carried out at 15,490 x g for 30 minutes to recover the purified testers and driver cDNA at the bottom of the tubes in pellet form. After the centrifugation step, supernatants were discarded while recovered pellet of each tester and driver cDNA was subjected to a washing step with 70 % (v/v) ethanol. The pellet of each tester and driver cDNA was then air dried for 20 minutes and each pellet was then dissolved in 100 µl of nuclease-free water. Subsequently, the enriched testers and driver cDNA were subjected to concentration and purity determination

using the NanoVue™ Plus Spectrophotometer (GE Healthcare, UK, 2010). Remaining testers and driver cDNA were then stored in -20 °C freezer for further use.

3.4.5 Removal of Oligonucleotide Adaptors

It was essential to remove the R 24-mer oligonucleotide adaptors from the testers and driver cDNA before proceed to the subtraction steps. Thus, digestion of the enriched testers and driver in section 3.4.4 was carried out in the same manner as described in section 3.4.2 using their respective restriction enzyme. Subsequently, the digestion products were purified using a PCR purification kit (GeneAll, Korea) by following the manufacturer's instruction with minor modifications. The volume of each adaptor-digested tester and driver was adjusted to 100 µl volume by adding nuclease-free water. After that, 5 volumes of PB buffer (provided by the kit) was transferred into each adaptor-digested tester and driver and followed by a mixing step by pipetting. The mixture of each adaptor-digested tester and driver was then transferred into a SV spin column (supplied together in the kit) separately and followed by a centrifugation step at 15,940 x g at room temperature for 30 seconds. Then, pass-through solution of each sample was discarded. All SV spin columns were re-inserted back in their respective receiver tube. Subsequently, a washing step was carried out for each adaptor-digested tester and driver sample by adding a volume of 700 µl of NW washing buffer (provided by the kit) and followed by another centrifugation step at 15,940 x g at room temperature for 30 seconds. The pass-through washing buffer for each tester and driver sample was discarded while all the SV columns were re-inserted back into their respective receiver tube. Next, another centrifugation step was carried out for 2 minutes in the same manner in order to remove excess washing buffer containing ethanol from the membrane of each spin column. All the receiver tubes were discarded and replaced with fresh 1.5 ml microcentrifuge tubes. A volume of 50 µl of EB buffer (provided in the kit) was then pipetted to the centre of the membrane of each SV spin column and followed by incubation at room temperature for 60 seconds. After the incubation period, a final centrifugation step was carried out in the same manner for 60 seconds. The recovered solution at the bottom of each receiver tube was the subjected to concentration and purity measurement as described in section 3.4.4. All the recovered

purified adaptor-digested testers and driver were then subjected to agarose gel electrophoresis on 1.2 % (w/v) agarose gel containing 0.2 X gelRed solution (Biothium, USA) in 1 X TAE buffer. The agarose gel was then viewed under a UV gel documentation system (Alpha Imager, USA, 2009) for adaptor-digested verification of each tester and driver. All adaptor-digested testers were then subjected to a ligation reaction with a new set of J pair oligonucleotide adaptor (refer Table 3.2) to be used in the first round of cDNA-RDA. The ligation as well as enrichment and purification of ligated testers and driver cDNA were carried out in the same manner as described in sections 3.4.3 to 3.4.5. Meanwhile, adaptor-digested driver was maintained throughout cDNA-RDA experiment without the presence of any adaptor.

3.4.6 First Round of cDNA-RDA

In the first round of cDNA-RDA, the digested PCR products were ligated with their respective J adaptors. An amount of 30 ng of each tester was mixed with 3,000 ng of each driver, resulting for a ratio of 1:100 (tester: driver). The mixtures were denatured at 95 °C for 5 minutes and hybridised at 67 °C for 20 hours in a thermocycler (Biorad, USA). The hybridised products were adjusted to a final volume of 200 µl by adding TE buffer. Then, 10 µl of the diluted hybridised products were selectively PCR-amplified in a 100 µl reaction containing 1 X PCR buffer, 0.2 mM dNTPs, 2.5 U of Taq DNA polymerase (GeneAll, Korea) and 24-mer J primers (J Bgl24, J Hind24 and JBam24). The PCR mixtures were then incubated at 72 °C for 5 minutes to fill in the ends of the re-annealed fragments. After that, a PCR cycle was performed using the following parameter: 94 °C for 3 minutes; 35 cycles at (94 °C for 1 minute; 72 °C for 2 minutes) and 72 °C for 5 minutes. The PCR products were then purified with phenol: chloroform: isoamyl alcohol (25:24:1) extraction and precipitated with 0.1 volume of 3 M sodium acetate (pH 5.2) and 2.5 volume of absolute ethanol as described in section 3.4.4. The precipitated pellets were dissolved in 40 µl of TE buffer.

After that, a volume of 20 µl of the purified amplified hybridization products were then subjected to mung bean nuclease digestion for removal of linear amplified tester-driver

hybrids as well as single stranded driver. The Mung Bean Nuclease digestion was carried out in 40 µl digestion reaction mixture containing 1 X Mung Bean Nuclease buffer and 2 U of Mung Bean Nuclease enzymes (New England Biolabs, UK). The digestion reaction was carried out at 30 °C for 30 minutes followed by addition of 160 µl of 50 mM Tris-Cl, (pH 8.9). The Mung Bean Nuclease activity of the mixture was then terminated at 99 °C for 5 minutes. Next, four replicates of 100 µl PCR mixture containing 250 ng of Mung Bean Nuclease digested product, 1 X PCR buffer (GeneAll, Korea), 2.5 U of *Taq* DNA polymerase (GeneAll, Korea), 2.5 mM dNTPs mix (GeneAll, Korea) were prepared and mixed by pipetting. PCR amplification of the Mung Bean Nuclease digested products was then carried out using the following protocol: 94 °C for 3 minutes; 35 cycles at (94 °C for 1 minute; 72 °C for 2 minutes) and 72 °C for 5 minutes. Subsequently, 5 µl of each PCR-amplified product was subjected to agarose gel electrophoresis for verification of the accomplishment of PCR amplification while the remaining PCR products for each sample were combined together and subjected to phenol: chloroform: isoamylalcohol (25:24:1) purification and precipitated with 0.1 volume of 3 M sodium acetate (pH 5.2) and 2.5 volume of absolute ethanol as described in section 3.4.4. The precipitated pellets were dissolved in 50 µl of TE buffer.

3.4.7 Subsequent Rounds of cDNA-RDA

The first round RDA products were used as drivers for second round of cDNA-RDA. Second round of cDNA-RDA was carried out by digesting the cDNA-RDA products with their respective restriction enzymes (*Bgl*II, *Bam*HI and *Hind*III). Then, the digested products were ligated with N adaptors at the ratio of tester: driver (1: 1,000). The next steps were carried out in the same manner until a mung bean nuclease digestion step. Subsequently, third round of cDNA-RDA was carried out using J adaptors with the ratio of tester: driver (1: 100,000). Products of each round of cDNA-RDA were electrophoresed on 1.2 % (w/v) agarose gel containing 1 X TAE buffer and 0.2 X gel-red solution (Biotium, USA) and viewed under a UV gel documentation system.

3.5 PREPARATION OF DH5α *ESCHERICHIA COLI* COMPETENT CELLS

DH5 α *Escherichia coli* competent cells were prepared by calcium chloride treatment following the method established by Sambrook *et al.* (1989). The preparation of competent cells was initiated by inoculating a colony of DH5 α *E. coli* cells grown on LB plate without the presence of any antibiotic into a volume of 5 ml of LB broth (Appendix A) in a 15 ml Falcon tube. The bacterial cells were grown for 16 hours in a 37 °C incubator shaker with shaking at 250 rpm. After the 16 hours incubation period, 2 ml of the bacterial culture was pipetted into 20 ml of fresh LB broth and followed by incubation in the same 37 °C incubator shaker with shaking at 250 rpm for two hours. After that, the two-hour bacterial culture was centrifuged at 1,000 x g at 4 °C for 5 minutes. The supernatant was discarded from the tube while recovered bacterial pellet at the bottom of the tube was resuspended in 10 ml of 75 mM cold-CaCl₂. The resuspended bacterial cells were then placed on ice for 20 minutes and followed by another centrifugation at 1,000 x g at 4 °C for 5 minutes. Supernatant was then discarded and recovered CaCl₂-treated bacterial pellet was resuspended in 2 ml of fresh 75 mM CaCl₂. Subsequently, a final volume of 15 % (v/v) glycerol was added into the resuspended competent bacterial cells and aliquoted in 100 μ l volume into several 1.5 ml microcentrifuge tubes. All the bacterial competent cells were stored in -80 °C freezer for further use.

3.6 PURIFICATION OF FINAL cDNA-RDA PRODUCTS

Each different product in third round of cDNA-RDA was excised from 1.2 % (w/v) agarose gel and purified using GeneAll Expin Combo GP kit (GeneAll, Korea) following the manufacturer's instructions. The product was excised from the agarose gel using ethanol-cleaned razor blade on a UV transilluminator that is available in UV Gel Documentation System (Alpha Imager, USA, 2009). The weight of each excised agarose gel containing specific cDNA-RDA final product was measured using an analytical balance before transferred into a 1.5 ml microcentrifuge tube. After that, 3 volumes of GB buffer (provided in the kit) was transferred into each 100 mg of agarose-excised cDNA-RDA final product. The mixture was incubated at 50 °C incubator oven for 10 minutes with gradual vigorous mixing by vortexing for every 2-3 minutes to enhance the melting of the agarose gel slice. After the incubation period, a volume of 700 μ l of each sample was transferred

into a SV spin column (provided in the kit) and followed by a centrifugation step at 15,940 x g at room temperature for 30 seconds. The pass-through solution from each SV spin column was then discarded and all the SV spin columns were reinserted back onto their respective receiver tube. The step was then repeated until all the remaining mixture of melted agarose gel slices containing respective cDNA-RDA final product were finished. After that, 500 µl of fresh GB buffer was transferred into each SV spin column and centrifuged in the same manner in order to remove the traces of agarose gel. The flow-through buffer for each SV spin column was discarded and all the SV spin columns were reinserted back to their respective receiver tube in the same manner. Next, a washing step was carried out by applying 700 µl of NW buffer (provided in the kit) into each SV spin column. The centrifugation step was then performed in the same manner for 30 seconds. Flow-through solution for each sample was then discarded and each SV spin column was re-inserted back into its respective receiver tube. All the SV spin columns were then centrifuged in the same manner for 2 minutes to remove any trace of washing solution containing ethanol from the membranes that might be present in the SV spin columns. The receiver tubes were then discarded and replaced with fresh 1.5 ml microcentrifuge tubes as the new receiver tubes. A volume of 50 µl of elution buffer (EB buffer) that is provided in the kit was then pipetted directly into the membrane located in the centre of each SV spin column. All the samples were then incubated at room temperature for 60 seconds followed by a final centrifugation step in the same manner for 2 minutes. The recovered gel-purified cDNA-RDA final product in each tube was further subjected to concentration and purity measurement as well as agarose gel electrophoresis as described in section 3.4.5. The remaining purified cDNA-RDA final products were kept in -20 °C freezer for subsequent use.

3.7 CLONING AND TRANSFORMATION OF PURIFIED cDNA-RDA FINAL PRODUCTS

Each purified cDNA-RDA final product in section 3.6 was subjected to cloning into pGEM-T Easy vector (Appendix C) (Promega, USA) following the instructions provided in its manual that has been downloaded from the manufacturer's website. All the purified

cDNA-RDA final products was subjected to A-tailing reaction in a volume of 10 μ l reaction mixture containing 10-25 ng of purified cDNA-RDA final product, 1 X PCR buffer (GeneAll, Korea) and 2.5 U of *Taq* DNA polymerase (GeneAll, Korea) in the presence of 0.2 mM dATP (Promega, USA). The A-tailing reaction was carried out at 70 °C for 30 minutes in a mastercycler (Biorad, USA). After the 30 minutes A-tailing reaction, a volume of 2 μ l of each A-overhanged cDNA-RDA final products were ligated with pGEM-T easy Vector in a 10 μ l ligation mixture comprising 50 ng of pGEM-T Easy Vector (provided in the kit), 1 X Rapid Ligation buffer (Appendix B) (provided in the kit), and 3 U of T4 DNA ligase (provided in the kit). The ligation was then carried out by incubating the ligation mixture in 4 °C chiller for 16 hours.

After the 16 hours ligation in 4 °C chiller, 2.5 μ l of ligation product of each cDNA-RDA final product was subjected to transformation into *E. coli* competent cells (refer to section 3.5). The transformation procedure was initiated by thawing the previously prepared *E. coli* competent cells in respective microcentrifuge tubes on ice followed by addition of 2.5 μ l of each ligation product into the respective tubes containing *E. coli* competent cells. The mixture of each sample was mixed by pipetting and incubated on ice for 20 minutes. A heat shock step was then carried out by placing the respective mixture of each sample on a thermal block that had been set at 42 °C for 90 seconds. All the samples were then quickly transferred onto ice and allowed to stand for about 60 seconds. Subsequently, a volume of 1,000 μ l of fresh LB broth was added into each sample and followed by incubation at 37 °C incubator shaker with shaking at 250 rpm for 60 minutes.

While waiting for the incubation period to be finished, 40 μ g/ml of 5-bromo-4-chloro-3-indolyl- β -D galactoside (X-gal) and 0.1 mM Isopropyl β -D-1-thiogalactopyranoside (IPTG) were spread onto pre-warmed LB agar plates containing 100 μ M ampicillin antibiotics. All the plates were allowed to be dried in a sterile condition for about 30 minutes. After the 60 minutes incubation period, all the samples were centrifuged at 15,940 x g for 60 seconds. A volume of 800 μ l of each supernatant was then discarded, while tiny bacterial pellets were resuspended with the remaining LB broth by vortexing. Next, 100 μ l of each bacterial sample was spread on the LB agar plates containing 40

$\mu\text{g/ml}$ of 5-bromo-4-chloro-3-indolyl- β -D galactoside (X-gal) and 0.1 mM Isopropyl β -D-1-thiogalactopyranoside (IPTG) as well as 100 μM of ampicillin antibiotic that have been prepared earlier. All the plates were then incubated at 37 °C for 16 hours in an inverted position.

After the 16 hours incubation period, 8 white colonies from each LB agar plate were re-streaked on fresh LB agar plates containing 100 μM of ampicillin. All the plates were then incubated at 37 °C incubator oven for 16 hours. After the incubation period, a single colony of each re-streaked white colonies for each sample was subjected to colony PCR for verification of successful transformants. The colony PCR was carried out by transferring a single colony of bacterial cells from their master plate by a sterile toothpick into 20 μl PCR mixture containing 1 X PCR buffer (GeneAll, Korea), 1 U of *Taq* DNA polymerase (GeneAll, Korea), 0.25 μM Sp6 primer, 0.25 μM T7 primer, 0.2 mM dNTPs mix (GeneAll, Korea). The PCR amplification was carried out using the following parameter: pre-denaturation at 95 °C for 5 minutes; 35 cycles at (denaturation at 94 °C for 30 seconds; annealing at 50 °C for 30 seconds; extension at 72 °C for 45 seconds) and final extension at 72 °C for 5 minutes. After finished the PCR amplification, a volume of 5 μl of each PCR product was electrophoresed on 1.2 % agarose gel and viewed under a UV Gel Documentation System (Alpha Imager, USA, 2009).

3.8 PLASMID MINI PREPARATION

A positive transformant for each sample that has been confirmed using colony PCR approach in section 3.7 was further selected for plasmid mini preparation. Initially, a single bacterial colony of positive transformant for each sample was inoculated into 5 ml LB broth in a 15 ml Falcon tube. All the samples were then incubated at 37 °C incubator shaker with shaking at 250 rpm for 16 hours. A volume of 4 ml of each overnight cultured transformant was subjected to plasmid DNA extraction using GeneAll Hybrid-Q™ Plasmid Rapid prep kit (GeneAll, Korea). All the overnight cultures of bacterial transformants were centrifuged for 2 minutes at room temperature with a centrifugation speed of 15,490 x *g*. The supernatants were discarded while the remaining bacterial pellet for each sample were

dissolved by resuspending thoroughly in 250 μ l of S1 buffer (a lysis buffer provided in the kit) containing 133.25 μ g of RNase A. Subsequently, 250 μ l of S2 buffer (provided in the kit) was added into the samples followed by gentle mixing step by inverting 4 to 5 times. Next, 350 μ l of G3 Buffer (provided in the kit) was added into the mixture and followed by immediately inverting the mixture for 4 to 5 times. Then, all the samples were centrifuged at room temperature for 10 minutes at 15,490 x g. Supernatant of each sample containing plasmid DNA was then transferred into a special spin column (provided in the kit) and centrifuged at 15,490 x g for 30 seconds. Flow-through solution was then discarded and the spin column was re-inserted back into its respective receiver tube. After that, two washing steps were carried out by applying 500 μ l of AW Washing Buffer (provided in the kit) followed by 700 μ l of PW Buffer (provided in the kit) into each spin column. Both washing steps were carried out by transferring the respective buffer into the spin column followed by a centrifugation step at 15,490 x g for 30 seconds. In each washing step, the flow-through solution for each sample was discarded and the spin-columns were reinserted backed into their respective receiver tubes. After the two washing steps using AW as well as PW washing buffer were done, all the spin columns were subjected to another centrifugation in the same manner for 2 minutes to remove the trace of ethanol that available in PW washing buffer. Subsequently, the dried spin-columns were placed onto fresh 1.5 ml microcentrifuge tubes as their new receiver tubes. Then, 35 μ l of elution buffer (EB buffer) (provided in the kit) was pipetted directly to the membrane in the centre of the spin columns and followed by incubation step at room temperature for 60 seconds. Finally, a final centrifugation was carried out at 15,490 x g in the same manner for 60 seconds to recover the recombinant plasmid containing respective cDNA-RDA final product. The isolated plasmid of each clone was then electrophoresed on 1.0 % (w/v) agarose gel containing 0.2 X Gelred solution and viewed under a UV Gel Documentation System (Alpha Imager, USA, 2009).

3.9 SEQUENCING AND SEQUENCE ANALYSIS OF cDNA-RDA PRODUCTS

All the plasmid samples extracted in section 3.8 were then sent for sequencing service provider (Macrogen, Korea) for single pass sequencing using SP6 and T7 universal

primers (Appendix G). Sequencing result of each clone was subjected to removal of vector sequences using Vecscreen tool at the NCBI GenBank database (<http://www.ncbi.nlm.nih.gov/>). Then, all the nucleotide sequences were compared to their closely related sequences using BLASTX tool that is available at the NCBI GenBank database (<http://www.ncbi.nlm.nih.gov/>) in order to identify their putative functions based on their hit to homologous protein that are available in the database.

3.10 ISOLATION OF FULL LENGTH FRAGRANCE-RELATED TRANSCRIPTS USING RACE-PCR

There were several candidates of fragrance-related transcripts which have been identified from cDNA-RDA work but the enzymes encoded by the transcripts have less commercial values compared to fragrance-related transcripts that encoded enzymes involved in the biosynthesis of final products of benzenoids and phenylpropanoids. Thus, another four fragrance-related transcripts that might be involved directly in biosynthesis of benzenoids and phenylpropanoids were selected to be characterised in this study namely eugenol synthase (VMPEGS), benzoic acid carboxymethyltransferase (VMPBSMT), 3-keto-acyl CoA thiolase (VMPKAT) and orcinol-O-methyltransferase (VMPOOMT). Partial sequence of all the samples were previously identified from floral cDNA library of *V. Mimi Palmer* constructed by Chan *et al.* (2009) and sequenced by Teh *et al.* (2012).

Initially, 5'- and 3'-RACE Ready cDNA template was synthesized from 500 ng of total RNA isolated from petal and sepal of *V. Mimi Palmer* using SMARTer™ RACE cDNA Amplification Kit (Clontech, USA) by following the manufacturer's instructions. In 5'-RACE Ready cDNA preparation, 500 ng of total RNA was mixed with 2 µl of 5 X First-Strand Buffer (provided by the kit) 1 µl of 20 mM DTT (provided by the kit) and 1 µl of 10 mM dNTP Mix, 1 µl of 5'-CDS primer A (provided in the kit). The mixture was adjusted to a total volume of 3.75 µl with nuclease-free water (provided in the kit) and mixed by pipetting. Meanwhile for 3'-RACE Ready cDNA preparation, 3'-CDS primer A was used instead of the 5'-CDS primer A (provided in the kit) and the final volume was adjusted to 4.75 µl using the provided nuclease-free water. Both 5'- and 3'-samples were

then subjected to incubation at 72 °C for 3 minutes and followed by 42 °C for 2 minutes in a mastercycler (Biorad, USA). Both reaction mixtures of 5'- and 3'- samples were then quickly chilled on ice. Then, 1 µl of SMARTer IIA oligo (provided in the kit) was added into only the 5'- reaction mixture. Next, 0.25 µl of RNase A inhibitor and 1.0 µl of SMARTScribe™ Reverse Transcriptase (100 U) were added into both 5'- and 3'-reaction mixtures and mixed by pipetting. Next, the reaction mixtures were incubated at 42 °C for 90 minutes in the same mastercycler (Biorad, USA) and followed by a heating step at 70 °C to deactivate the reverse transcriptase enzyme. Subsequently, the ready 5'- and 3'-RACE cDNA samples were diluted in 100 µl of tricine-EDTA buffer (provided in the kit) and stored in -20 °C freezer for further use. Isolation of targeted 5'- and 3'- regions of interest was then carried out by mixing each 5'- and 3'-Ready cDNA separately into PCR mixture containing 1 X Advantage2 PCR buffer (Clontech, USA), 1 X Universal Primer Mix (Clontech, USA), 0.2 µM gene specific primer (GSP) (Table 3.3), 0.2 mM dNTP mix (GeneAll, Korea) and 1 X Advantage 2 polymerase mix enzymes (Clontech, USA). After that, the PCR mixtures for both 5'- and 3'-regions of respective gene of interest were subjected to PCR-amplified in the same mastercycler (BioRad, USA) using the following parameter: pre-denaturation at 94 °C for 3 minutes, 35 cycles of (denaturation at 94 °C for 30 seconds; annealing at 65 °C for 30 seconds; extension at 72 °C for 3 minutes) followed by a final extension at 72 °C for 5 minutes.

Table 3.3: Gene Specific Primers are used for 5'-regions, 3'-regions and Full Open Reading Frame (ORF) of Transcripts Isolation from *Vanda Mimi Palmer*.

PCR products	Primer Sequences	Annealing Temperature (°C)
VMPBSMT ORF	VMPBSMT ORF Forward: 5'-GGAGATTATAACCATGATTGCTGAACAGG-3'	65
	VMPBSMT ORF Reverse: 5'-GACAATCAAGACATTGCTTGGCTCGGGAG-3'	
VMPBSMT 3'-region	VMPBSMT 3'-GSP: 5'-CCACTCTTCATACAGCCTCATGTGGCT-3'	65
VMPKAT ORF	VMPKAT ORF Forward: 5'-GAGAGATGGAGAAAGCCATCGACAGGC-3'	65
	VMPKAT ORF Reverse: 5'-TCCACAGCGCAATTAAAGGAGCATAAGC-3'	
VMPKAT 5'-region	VMPKAT 5'-GSP: 5'-CACAGCAGGATCAACCCCAACTGCAGC-3'	65
VMPKAT 3'-region	VMPKAT 3' GSP: 5'-CCATGTGCATTGGTTCTGGGATGGGAGC-3'	65
VMPEGS ORF	VMPEGS ORF Forward: 5'-CTTACATTCAATGGCGTCCGAGAAGAGC-3'	65
	VMPEGS ORF Reverse: 5'-CAGACTTCAAACAAGGATGGCAAAGATGAGC-3'	
VMPEGS 5'-region	VMPEGS 5'-GSP 5'-GAGGGCGAAGGTGGGATGTCCAGTGCG-3'	65
VMPOOMT ORF	VMPOOMT ORF Forward: 5'-CATGACGACTGCCGTGGAACAAGAAACCC-3'	65
	VMPOOMT ORF Reverse: 5'-GCTACATTCACAATGCTCAGCCATAGC-3'	
VMPOOMT 5'-region	5'-GAGACAGGACGCGCATGATGCTGCGAAGG-3'	65

All the amplified RACE-PCR products were electrophoresed on 1.2 % agarose gel and the successfully amplified PCR products were excised from the gel and subjected to gel purification using GeneAll Exspin Combo GP (GeneAll, Korea) as described in section 3.6. The purified PCR products were then transformed into DH5 α *E. coli* competent cells as described in section 3.7 and further subjected to plasmid mini preparation in the same manner as described in section 3.8. All the isolated recombinant plasmids containing the respective 5'- and 3'-region of interest were then sent for single pass sequencing to a sequencing service provider (Macrogen, Korea). The PCR amplification of full ORF for each transcript of interest was carried out in the same manner as carried out for 5'- and 3'-regions isolation with a minor modifications whereby a single gene specific primer (GSP) and Universal Primer Mix (UPM) were replaced with 0.2 μ M ORF forward primer as well as 0.2 μ M ORF reverse primer (refer Table 3.3). The amplified ORF sequences were then subjected to gel purification, cloning into pGEM-T Easy vector (Appendix C), transformation into DH5 α *E. coli* competent cells, plasmid mini preparation as carried out for cDNA-RDA final products (refer section 3.6) as well as 5'- and 3'-regions of interest transcripts. The recombinant plasmids of all the cloned ORF sequences were then subjected to sequencing reactions using SP6 and T7 universal primers. All the sequences were further submitted to the NCBI GenBank database.

3.10.1 Designing Gene Specific Primers for Full Length cDNA Isolation

Gene specific primers as listed in Table 3.3, Appendix F were designed for isolation of their missing regions, either at 3'- or 5'-end regions. At the same time, gene specific primers (forward and reverse primers) of each identified ORF were designed at specific location that is slightly earlier than start codon for forward primer and slightly after stop codon to ensure further amplified PCR products consist of full ORF sequence.

3.10.2 Sequence Analysis of Full Length cDNA Sequences

All the sequences of 5'- and 3'-regions of interest transcripts were further subjected to contig analysis with the previously identified partial sequence of the transcripts from the

floral cDNA library of *V. Mimi Palmer* using the CAP3 Assembly Programme (<http://doua.prabi.fr/software/cap3>) that is free to be accessed via online. The presence of Full Open Reading Frame (ORF) of each transcript of interest was carried out by translating the assembled sequences from contig analysis into six frame translation region of the respective transcripts using BioEdit software (Appendix D). The longest translated amino acid sequences were then subjected to BLASTP analysis at the NCBI GenBank database (<http://www.ncbi.nlm.nih.gov>) in order to predict their putative function by their sequence homology to amino acid sequences of proteins that are available at the NCBI GenBank database.

The translated amino acid sequences of all the fragrance-related transcripts were then subjected to ClustalW multiple alignment analysis with other homologous amino acid sequences of homologous proteins that have been retrieved from the database in order to identify their conserved regions. In addition, the theoretical isoelectric point (pI) and molecular weight of the protein sequences were predicted using ExPasy (<http://br.expasy.org/tools/>), a free online tool. Besides that, phylogenetic tree was constructed for each protein of interest by comparing with their homologous protein sequences retrieved from the NCBI GenBank database. A Neighbour Joining Method was used to develop phylogenetic trees using the MEGA software, version 4.0 (Tamura *et al.*, 2007). Phylogenetic construction was developed using bootstrapping of neighbouring-joining method as recommended by Tamura *et al.* (2007). Furthermore, Motifs search and signal peptide prediction was carried out using ExPasy tool (<http://br.expasy.org/tools/>) as well as by using Localizome software (<http://localodom.kobic.re.kr/LocaloDom/index.htm>) (Lee *et al.*, 2006). Besides that, motif analyses were carried out using multiple online bioinformatics servers including ScanProsite (de Castro *et al.*, 2006), ProP 1.0 server (Duckert *et al.*, 2004), TMHMM server 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>), SignalP 4.1 server (Petersen *et al.*, 2011), ChloroP 1.1 server (Emanuelsson *et al.*, 1999) and TargetP 1.1 server (Emanuelsson *et al.*, 2000).

3.11 DEVELOPMENT OF 3-DIMENSIONAL STRUCTURAL MODEL OF FRAGRANCE-RELATED PROTEINS

All the deduced amino acid sequence of fragrance-related transcripts (Appendix E) including VMPKAT, VMPBSMT, VMPEGS and VMPOOMT were subjected to various protein sequence analyses using several bioinformatics servers including BLAST-PDB (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>), PSI-BLAST (<http://blast.ncbi.nlm.nih.gov/Blastp/PSI-BLAST>) as well as SUPERFAMILY HMM (Gough and Chothia, 2002) (Appendix Ib and Ic). PSI-BLAST and BLAST-PDB were used to perform threading method on the amino acid sequence of the isolated transcript while SUPERFAMILY HMM was used to identify the presence of any conserved domain and possible families of the protein.

A threading method was performed to screen for potential protein structural templates by submitting the amino acid sequence of all the fragrance-related proteins to the available library of known protein folds using multiple bioinformatics servers including PSI-BLAST (<http://blast.ncbi.nlm.nih.gov/Blastp/PSIBLAST>), Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/>), HHpred (<http://toolkit.tuebingen.mpg.de/hhpred>), pGenThreader (Jones 1999) and LOMETS (<http://zhanglab.ccmb.med.umich.edu/LOMETS/>). The most common protein template that appeared in all the threading approaches with the highest identity was selected as the best protein template to be used as a reference in 3D-structural development of each fragrance-related protein.

Three-dimensional structure of each fragrance-related protein was developed using the MODELLER version 9.14 and the chosen protein template was used as reference protein structure. Development of 3D-structure of each fragrance-related protein was performed using the protocol described by Eswar *et al.* (2007) as well as the recommended basic procedure that was available at the MODELLER website (Appendix Ih). A number of five models of 3D-protein structure were generated using the MODELLER version 9.14 software for each fragrance-related protein. After that, the best 3D-protein structure was determined based on two main factors. The first factor in choosing the best 3D-protein structure was based on the lowest dope-score with the smallest value to minimize the spatial restraints of the structure. Meanwhile the second factor was based on the GA341 value that is equal to 1.0. The best developed model of 3D-protein structure for each fragrance-related protein was saved in PDB format and subjected to TM-score and root

mean square deviation (RMSD) evaluation by submitting the model to the TM-align program (<http://zhanglab.ccmb.med.umich.edu/TM-align>). Based on the evaluation, the best developed 3D-protein model for each fragrance-related protein is chosen based on its significant TM-score of 0.4 -1.0 and RMSD value of approximately 1.0 (Zhang and Skolnick, 2005) (Appendix Ie).

Then, the chosen 3D-protein model for each fragrance-related protein was further assessed using different verification tools including VERIFY-3D (http://services.mbi.ucla.edu/Verify_3D/), ERRAT (<http://services.mbi.ucla.edu/ERRAT/>), PRO-CHECK (Laskowski *et al.*, 1993) as well as ProSA-web (Wiederstein and Sippl, 2007). After that, all the chosen 3D-protein models of fragrance-related proteins were subjected to a structural refinement process including structural optimisation and energy minimisation by submitting the 3D-protein models to the Modrefiner server (<http://zhanglab.ccmb.med.umich.edu/ModRefiner/>) and run using a default setting (Xu and Zhang, 2011). After the refinement process, the developed 3D-protein model for each fragrance-related protein was visualized in 3D-graphical image by using the UCSF CHIMERA software (Pettersen *et al.*, 2004) (Appendix Id).

3.12 FUNCTIONAL PREDICTION ON FRAGRANCE-RELATED PROTEINS

The binding pockets and active sites (Appendix I(f)) of selected four fragrance-related protein models of *V. Mimi Palmer* were recognized by Computed Atlas of Surface Topography of Proteins (CASTp) server (UIC, Chicago, IL, USA) (<http://sts.bioengr.uic.edu/castp/>). The active sites present on enzymes were selected for docking the ligands. The SDF files of various ligands were retrieved from the Pubchem databases of NCBI (<http://pubchem.ncbi.nlm.nih.gov>), Zinc docking database (<http://zinc.docking.org>) and ligand explorer in Protein Data Bank database (http://www.rcsb.org/pdb/ligandExplorer_viewer.html). These SDF files were converted into PDB files with the help of an open babel converter (<http://openbabel.org>). Molecules were docked at the binding sites by Autodock 4.2 (<http://autodock.scripps.edu>).

Structural data of protein models are subjected to Autodock Tools for docking the putative ligands and substrates with the selected protein models of interest, namely as VMPKAT, VMPBSMT, VMPEGS and VMPOOMT (Appendix Ia and Ig). The requirements for molecular docking of enzymes and substrates using Autodock 4.2 are to install MGL tools, Cygwin, binary files of autodock 4.0 and autogrid 4.0 as well as Discovery studio visualizer. All selected programs were installed properly for 32-bits version on operating system of Windows 7 before using Autodock 4.2 software. The first step is to retrieve required ligand and target.pdb files from major databases. The second step is preparing PDBQT format files for target and ligand (Target.pdbqt, Ligand.pdbqt) and grid and docking parameter file (a.gpf and a.dpf) using AutoDock 4.2. The third step is to perform molecular docking using Cygwin and finally the results are analyzed. Finally, ligand-enzyme interaction complex.pdb was retrieved and complex.pdb was opened in Discovery Studio Visualizer software in order to view the docking between ligand and substrates with its receptor of enzyme in catalytic cavity, surrounded by neighbouring amino acid residues. The image of complex.pdb was saved as image file format.

3.13 EXPRESSION STUDIES OF FRAGRANCE-RELATED GENES BY REAL TIME-PCR

Expression analysis of the selected fragrance-related transcripts (VMPKAT, VMPBSMT, VMPEGS and VMPOOMT) was carried out using real-time RT-PCR in different tissues and developmental stages. For expression studies in different tissues, total RNA was isolated from both floral (bud, petal, sepal and lip) and vegetative tissues (leaf, shoot and root). Meanwhile, for expression studies at different flower developmental stages, total RNA was isolated from bud, half-open flower and fully-open flower. All the samples for expression analysis on different tissues and developmental stages were collected at 8.00 am in the morning. The procedure used for total RNA extraction from all the above mentioned samples was as described in section 3.2.

The extracted total RNA samples were used as templates for first strand cDNA synthesis. In the total RNA, poly-A tail mRNAs were reverse-transcribed into first strand

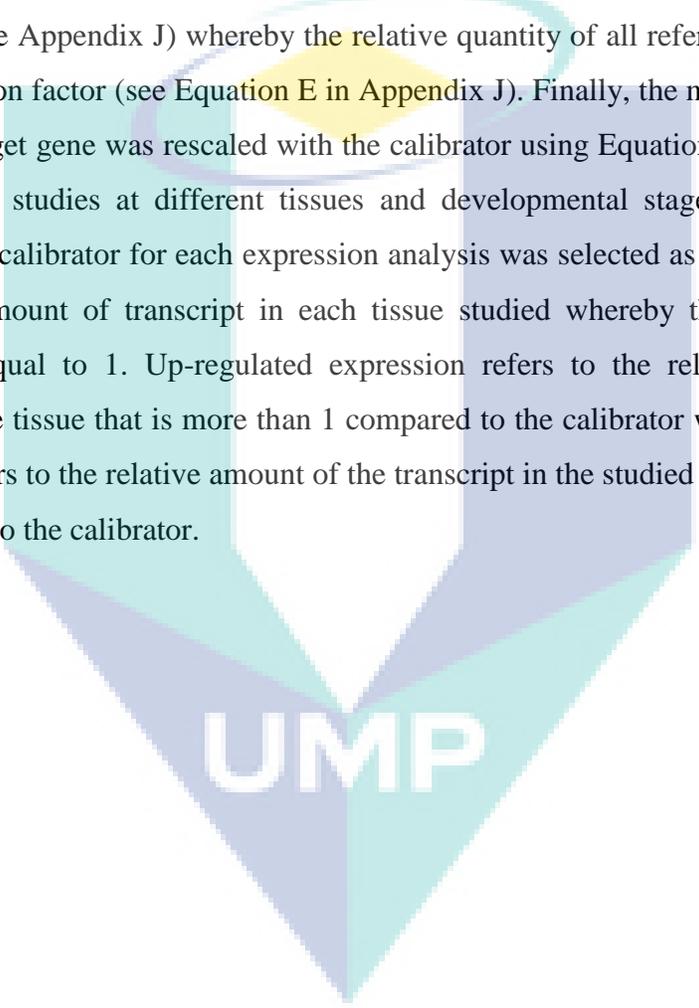
cDNA by using Quantitect reverse transcription kit (Qiagen, Germany) according to the manufacturer's instruction. Firstly, 500 ng of total RNA template and 2 μ l of 7 X of gDNA wipe buffer (provided in the kit) were transferred into a clean PCR tube, followed by addition of RNase-free water to a total volume of 14 μ l. The mixture was then incubated at 42 °C for 2 minutes and chilled on ice quickly. The mixture consisting of reverse transcriptase enzyme (1 μ l), Quantiscript RT buffer (4 μ l), and RT primer mix (1 μ l) were mixed together with the remaining components of the kit. The mixture was then incubated at 42 °C for 30 minutes, followed by incubation at 95 °C for 3 minutes to inactivate the reverse transcriptase enzyme. The first strand cDNA for each sample was aliquoted into a few microcentrifuge tubes and kept at -20 °C for further use.

Four fragrance-related transcripts selected in this study (VMPKAT, VMPBSMT, VMPEGS and VMPOOMT) together with three housekeeping genes namely actin (GenBank accession no: AF246716.1), alpha tubulin (GenBank accession no: GW687608.1) and cyclophilin (GenBank accession no: GU942927.1) as internal controls were selected for expression analysis in different tissues as well as at different developmental stages using real-time RT-PCR. The real-time RT-PCR reactions were performed in four replicates for each samples by using Mastercycler[®] ep Realplex Real-time PCR (Eppendorf, Germany). The real-time RT-PCR reactions were carried out in a total volume of 20 μ l containing 1 X SYBR[®] Green I Hot-start Real-time PCR mix (GeneCraft, Germany), 1 μ l of 10 X dilution of cDNA template and 150 nM of gene-specific primers (Table 3.4) using the following protocol: 95 °C for 2 minutes; 40 cycles of 95 °C for 30 s, annealing for 30 s (see Table 3.4) and extension at 72 °C for 30 s; 81 cycles for melting curve analysis; and 10 s for each 0.5 °C (55-95 °C).

Table 3.4: Primer sequences for gene expression analysis using Real-time RT-PCR

Target/ amplicon length (bp)	Primers	Primer sequences	Annealing temperature (°C)
VMPKAT (175)	Forward	5'-TGTAGCTGATCGGTTGAGTAATGT3'	58
	Reverse	5'-AGCGCAATTAAGGAGCATAAG-3'	
VMPBSMT (236)	Forward	5'-ATAATGGGAAGCCAAAGCACAC-3'	56
	Reverse	5'- TGGGAATAGTAAGGGTTGTATGTTC- 3'	
VMPEGS (151)	Forward	5'-TCCCGAAGTCAAATACACTACAG-3'	58
	Reverse	5'-AAACAAGGATGGCAAAGATGAG-3'	
Endogenous control			
Actin (236)	Forward	5'-CAGTGTTTGGATTGGAGGTTC-3'	56
	Reverse	5'-CCAGCAGCAGTCAGGAAA-3'	
Tubulin (227)	Forward	5'-CTCCCGCATTGACCATAAAT-3'	56
	Reverse	5'-GGAACCACACCCAAACTCTC-3'	
Cyclophilin (200)	Forward	5'-TTGGATGTCGTGAAGGCAAT-3'	58
	Reverse	5'-CAACACAAGAAGATAGCACAGCA- 3'	

Relative quantification was carried by applying $\Delta\Delta C_T$ comparative method (Livak and Schmittgen, 2001) and normalized with three housekeeping gene transcripts; alpha tubulin (GenBank accession no: GW687608.1), actin (GenBank accession no: AF246716.1), and cyclophilin (GenBank accession no: GU942927.1). Relative quantity (Q) of cDNA transcript of each gene was calculated using Equation C (see Appendix J). After that, normalization expression was carried out by normalizing the relative cDNA quantity of each gene of interest with cDNA of reference genes (endogenous control) using Equation D (see Appendix J) whereby the relative quantity of all reference genes was used for normalization factor (see Equation E in Appendix J). Finally, the normalized expression level of the target gene was rescaled with the calibrator using Equation F (see Appendix J). For expression studies at different tissues and developmental stages, bud was used as calibrator. The calibrator for each expression analysis was selected as a reference tissue for comparative amount of transcript in each tissue studied whereby the expression of the calibrator is equal to 1. Up-regulated expression refers to the relative amount of the transcript in the tissue that is more than 1 compared to the calibrator while down-regulated expression refers to the relative amount of the transcript in the studied tissue that less than 1 in comparison to the calibrator.

The logo for UMP (University of Malaya Press) is a large, downward-pointing arrow shape. It is composed of several overlapping triangles in shades of teal, light blue, and yellow. The letters 'UMP' are written in a bold, white, sans-serif font across the center of the arrow's shaft.

UMP

CHAPTER 4

RESULTS AND DISCUSSION

4.1 IDENTIFICATION OF FRAGRANCE-RELATED TRANSCRIPTS VIA cDNA- REPRESENTATIONAL DIFFERENCE ANALYSIS (cDNA-RDA)

cDNA Representational Difference Analysis (cDNA-RDA) of selected fragrant-orchids including *V. Mimi Palmer*, *V. Small Boy Leong* and *Vandachostylis Sri-Siam* in comparison to *V. Tan Chay Yan*, a non-fragrant orchid was carried out in three rounds of cDNA-RDA comprising hybridization as well as selective amplification in each round. Based on the cDNA-RDA experiment, a number of amplified cDNA bands have been detected in first, second as well as third round of cDNA-RDA (Figure 4.1) in the size range of ~250 to 600 bp. From the cDNA-RDA analysis, 21 cDNA-RDA products were successfully identified. All the final cDNA-RDA products in the third round of cDNA RDA (Figure 4.1c) have been successfully purified using agarose gel purification approach. Among the 21 of cDNA-RDA final products, six cDNA-RDA products have been isolated from cDNA-RDA of *V. Mimi Palmer* (tester cDNA) versus *V. Tan Chay Yan* (driver cDNA) while another six products have been subtracted from cDNA-RDA of *V. Small Boy Leong* (tester cDNA) versus *V. Tan Chay Yan* (driver cDNA). Meanwhile, the remaining nine cDNA-RDA products were successfully subtracted from cDNA-RDA of *Vandachotylis Sri-Siam* (tester cDNA) versus *V. Tan Chay Yan* (driver cDNA). In this study, *V. Tan Chay Yan* was chosen as driver cDNA in all the three sets of cDNA-RDA due to the natural characteristic of *V. Tan Chay Yan* as a non-fragrant orchid, which does not emit any scent that can be detected by human nose. While, *V. Mimi Palmer*, *V. Small*

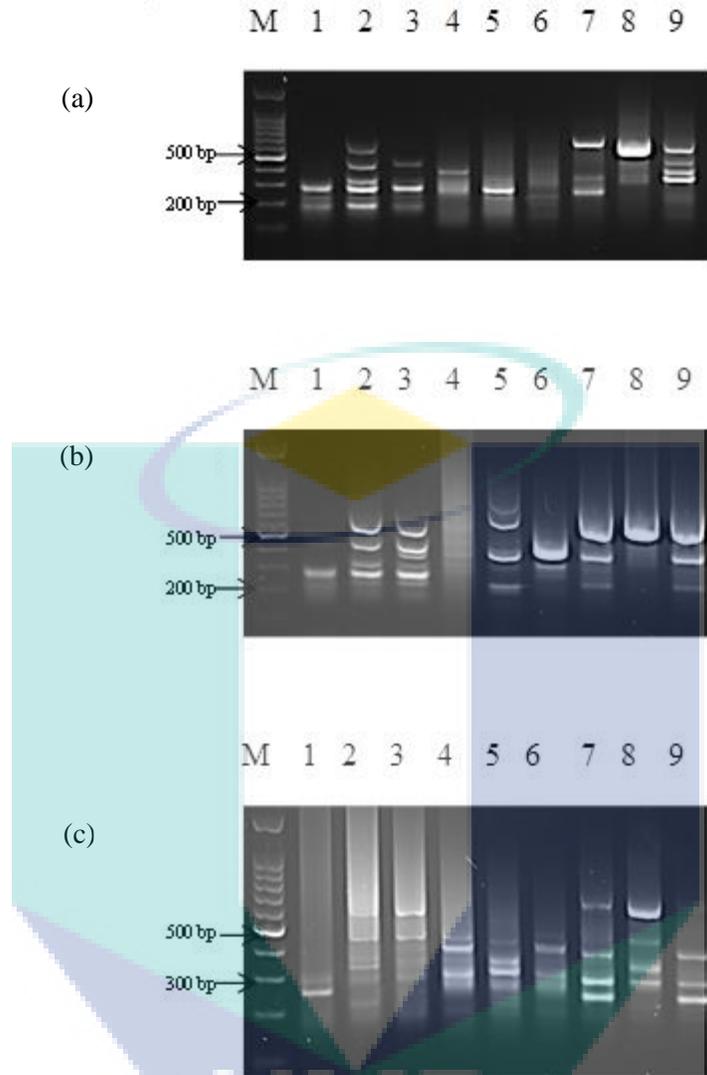


Figure 4.1: Difference products in three different rounds of cDNA-RDA; (a) first round cDNA-RDA; (b) second round cDNA-RDA; (c) third round cDNA-RDA. Each sample was electrophoresed on 1.2% (w/v) agarose gel. Lane M: 100 bp DNA marker (Vivantis, Malaysia); lane 1: Set A *Bgl*III; lane 2: Set A *Bam*HI; Lane 3: Set A *Hind*III; Lane 4: Set B *Bgl*III; Lane 5: Set B *Bam*HI; Lane 6: Set B *Hind*III; Lane 7: Set C *Bgl*III; Lane 8: Set C *Bam*HI; Lane 9: Set C *Hind*III.

Boy Leong and *Vandachostylis Sri-siam* were selected separately as testers cDNA in the three sets of cDNA-RDA due to their potential to be the sources of fragrance-related transcripts screening and isolation.

BLASTX analysis of nucleotide sequences of all the 21 cDNA-RDA products that have been cloned separately into pGEM-Easy plasmid (Promega, USA) and sequenced using a single pass sequencing approach has shown the presence of three fragrance-related transcripts including phenylalanine ammonia-lyase (VMP/BamHI_1), S-adenosyl-L-methionine synthase (VSBL/Bgl II_1) and methionine synthase (VSBL/BgIII_3 and VSSS/BgIII_3) (refer Table 4.1). All the three transcripts have been reported to be involved in benzenoid and phenylpropanoid pathway, as one of the fragrance biosynthetic pathways that are very important in biosynthesis of fragrance components as reported previously in other scented flowers including *P. hybrida*, *A. majus* and *C. breweri* (Verdonk *et al.*, 2003; Boatright *et al.*, 2004; Dudareva *et al.*, 1996; 1998). Interestingly, in *V. Mimi Palmer*, the main orchid hybrid plant used in this study, it has been previously reported that there are four fragrance compounds from the class of benzenoid and phenylpropanoid compounds including methylbenzoate, benzyl acetate, phenylethyl acetate and phenylethyl alcohol emitted by the orchid (Mohd-Hairul *et al.*, 2010a). Thus, all the three isolated transcripts might contribute in the biosynthesis of the benzenoid and phenylpropanoid compounds as the component of orchid scent.

The first candidate of fragrance-related transcript (VMP/BamHI_1 clone) has shown the highest homology to phenylalanine ammonia lyase (PAL) of another orchid *Dendrobium candidum*, an orchid from other genera with the highest score of 167 with E-value of $5e^{-46}$. Phenylalanine ammonia lyase (PAL; EC 4.3.1.5) has been reported to be involved in benzenoid and phenylpropanoid pathway in various plants including *Cistanche deserticola* (Gao *et al.*, 2012), *Musa acuminata* (Alvarez *et al.*, 2013) *Rhus chinensis* (Ma *et al.*, 2013) and *Dendrobium candidum* (Jin *et al.*, 2013). PAL is the first and the key enzyme in the regulation of overall carbon flux in the benzenoid and phenylpropanoid pathway (Camm and Towers, 1973; Xiang and Moore, 2002). This enzyme catalyzes the deamination of L-phenylalanine to produce cinnamic acid (trans-cinnamate) and ammonia

Table 4.1: BLASTX analysis on final difference cDNA-RDA products.

Set A [*Vanda Mimi Palmer* (Tester) vs *Vanda Tan Chay Yan* (Driver)]

No	Clone name	Putative identity	E value	% identity	Genbank accession number
1	VMP/ <i>Bgl</i> III_1	Cobalamin biosynthesis protein	0.92	39	JZ480886
2	VMP/ <i>Bam</i> HI_1	Phenylalanine ammonia- lyase*	5e ⁻⁴⁶	79	JZ480887
3	VMP/ <i>Bam</i> HI_2	Uncharacterized protein	8e ⁻¹⁹	57	JZ480888
4	VMP/ <i>Hind</i> III_1	RNA-dependent RNA polymerase	1e ⁻²²	98	-
5	VMP/ <i>Hind</i> III_2	Stearoyl-acyl carrier protein desaturase	4e ⁻³²	76	JZ480889
6	VMP/ <i>Hind</i> III_3	RNA-dependent RNA polymerase	3e ⁻²⁹	96	-

Set B [*Vanda Small Boy Leong* (Tester) vs *Vanda Tan Chay Yan* (Driver)]

No	Clone name	Putative identity	E value	% identity	Genbank accession number
1	VSBL/ <i>Bgl</i> III_1	s-adenosyl methionine synthase*	2e ⁻²²	96	JZ480893
2	VSBL/ <i>Bgl</i> III_2	Gag-pol polyprotein	2e ⁻¹²	61	JZ480894
3	VSBL/ <i>Bgl</i> III_3	Methionine synthase*	3e ⁻⁸⁴	85	JZ480895
4	VSBL/ <i>Bam</i> HI_1	Clathrin assembly protein	2e ⁻¹⁵	64	JZ480896
5	VSBL/ <i>Hind</i> III_2	RNA-dependent RNA polymerase	7e ⁻⁵¹	97	-
6	VSBL/ <i>Hind</i> III_3	RNA-dependent RNA polymerase	9e ⁻²⁸	95	-

Set C [*Vandachostylis* Sri-Siam (Tester) vs *Vanda* Tan Chay Yan (Driver)]

No	Clone name	Putative identity	E value	% identity	Genbank accession number
1	VSSS/ <i>Bg</i> III_1	Auxin-repressed protein	2e ⁻⁰⁶	57	JZ480890
2	VSSS/ <i>Bg</i> III_2	No significant similarity found	-	-	JZ480891
3	VSSS/ <i>Bg</i> III_3	Methionine synthase*	1e ⁻⁸³	92	JZ480892
4	VSSS/ <i>Bam</i> HI_1	RNA-dependent RNA polymerase	7e ⁻³⁸	91	-
5	VSSS/ <i>Bam</i> HI_2	RNA-dependent RNA polymerase	2e ⁻²⁴	98	-
6	VSSS/ <i>Hind</i> III_1	RNA-dependent RNA polymerase	3e ⁻²⁹	97	-
7	VSSS/ <i>Hind</i> III_2	RNA-dependent RNA polymerase	5e ⁻⁵⁰	90	-
8	VSSS/ <i>Hind</i> III_3	RNA-dependent RNA polymerase	7e ⁻⁵¹	96	-
9	VSSS/ <i>Hind</i> III_4	RNA-dependent RNA polymerase	8e ⁻⁵⁰	94	-

Note: Putative fragrance-related sequences were denoted in asterisk form (*).

(Xiang and Moore, 2002). Besides that, PAL enzyme has been reported to be very important in biosynthesis of flower pigments including anthocyanins, carotenoids and chlorophylls as the key regulator enzyme in benzenoid and phenylpropanoid pathway (Xiang and Moore, 2002). In addition, the presence of PAL enzyme as the key regulator for the pathway plays a very important role in the biosynthesis of various secondary metabolites including benzenoids, phenylpropanoids and their derivatives that are very important as antibiotics, insect repellent, protection agent against ultraviolet radiation as well as in molecular signals involving plant and microbes interactions (Calabrese *et al.*, 2004). Benzenoid and phenylpropanoid pathways have also been reported to be important for production of plant hormones such as methyl salicylate, methyl benzoate and auxin (Orlova *et al.*, 2006) as well as in the biosynthesis of cell wall components of plant systems including suberin, lignin and other related compounds (Liang *et al.*, 1989; Calabrese *et al.*, 2004). In *V. Mimi Palmer*, the presence of phenylalanine ammonia lyase (VMPPAL) might contribute as the main precursor for biosynthesis of benzyl acetate and methylbenzoate via cinnamic acid route. Meanwhile, for biosynthesis of volatile phenylpropanoids including phenylethyl acetate and phenylethyl alcohol that have been previously identified as the components of *V. Mimi Palmer*'s scent might be initiated by carboxylation of transaminated phenylacetaldehyde synthase (VMPPAAS) via the other route.

BLASTX analysis on the second candidate of fragrance-related transcripts which is represented by VSBL/BglII_1 clone has shown the highest homology to S-adenosyl-L-methionine synthase (SAMS: EC 2.5.1.6) of *Hosta ventricosa* with the highest score of 97.1 and *E*-value of $2e^{-22}$ as shown in Table 4.1. In molecular biology aspect, S-adenosyl-L-methionine synthase genes and transcripts have been studied in several plants including *Oryza sativa* (rice) and *Brassica rapa* (chinese cabbage) (Jae-Gyeong *et al.*, 2012; Saet-Byul *et al.*, 2012). Previous studies have shown that S-adenosyl-L-methionine synthase is involved as the key enzyme in S-adenosyl-L-methionine (SAM) biosynthesis (Roeder *et al.*, 2009). The enzyme is one of the crucial enzymes in many enzyme-catalytic reactions in many organisms whereby SAM is the final product synthesized via adenylation of methionine (Met) and ATP (Jae-Gyeong *et al.*, 2012). In benzenoid and phenylpropanoid pathway, SAM plays a very important role as biological cofactor of several

methyltransferase enzymes including salicylic acid carboxymethyltransferase (SAMT), benzoic acid carboxymethyltransferase (BAMT) and orcinol-O-methyltransferase (OOMT) that have been reported to be involved in the biosynthesis of fragrance compounds in scented flowers including *C. breweri*, *A. majus*, *R. hybrida* and *P. hybrida* (Dudareva *et al.*, 1996; 1998; Boatright *et al.*, 2004; Guterman *et al.*, 2002; Verdonk *et al.*, 2003).

Besides that, SAM has also been reported to be important as methyl group donor in most of transmethylation reactions in various reactions including biosynthesis of nucleic acids, lipids, polysaccharides, proteins and various plant secondary metabolites including polyamines, nicotianamine, biotin and ethylene (Roeder *et al.*, 2009). In *V. Mimi Palmer*, partial sequence of benzoic acid carboxymethyltransferase (VMPBSMT) that has been previously identified in the floral cDNA library constructed from the orchid (unpublished data) might be responsible for biosynthesis of methylbenzoate compound. Thus, methylbenzoate compound detected in the scent of *V. Mimi Palmer* might be converted from benzoic acid by catalytic reaction involving BSMT enzyme encoded by VMPBSMT in the presence of S-adenosyl-L-methionine compound as its biological cofactor. Meanwhile, the third candidate of fragrance-related transcripts identified from the cDNA-RDA analysis is two methionine synthase transcripts of *V. Small Boy Leong* (clone VSBL/BgIII_3) and *Vandachostylis Sri-Siam* (clone VSSS/BgIII_3), respectively.

In addition, methionine synthase (MS; EC 2.1.1.14) is an enzyme that has been reported to be responsible in catalytic-reaction of homocysteine methylation into methionine (Roeder *et al.*, 2009) as shown in Figure 4.2. Benzenoids production in fragrant orchids might be regulated by the transcriptional control of shikimate pathway with the production of chorismate which is then converted to phenylalanine and further metabolized to benzenoids and related products. The role of PAL becomes important when it also converts phenylalanine to trans-cinnamic acid which is a precursor for downstream production of benzoic acid. The indirect relationship between PAL and s-adenosyl-L-methionine (SAM) cycle genes (SAMS and MS) might happen when SAM cycle contributes to volatile benzenoid biosynthesis by providing methyl group from the methyl

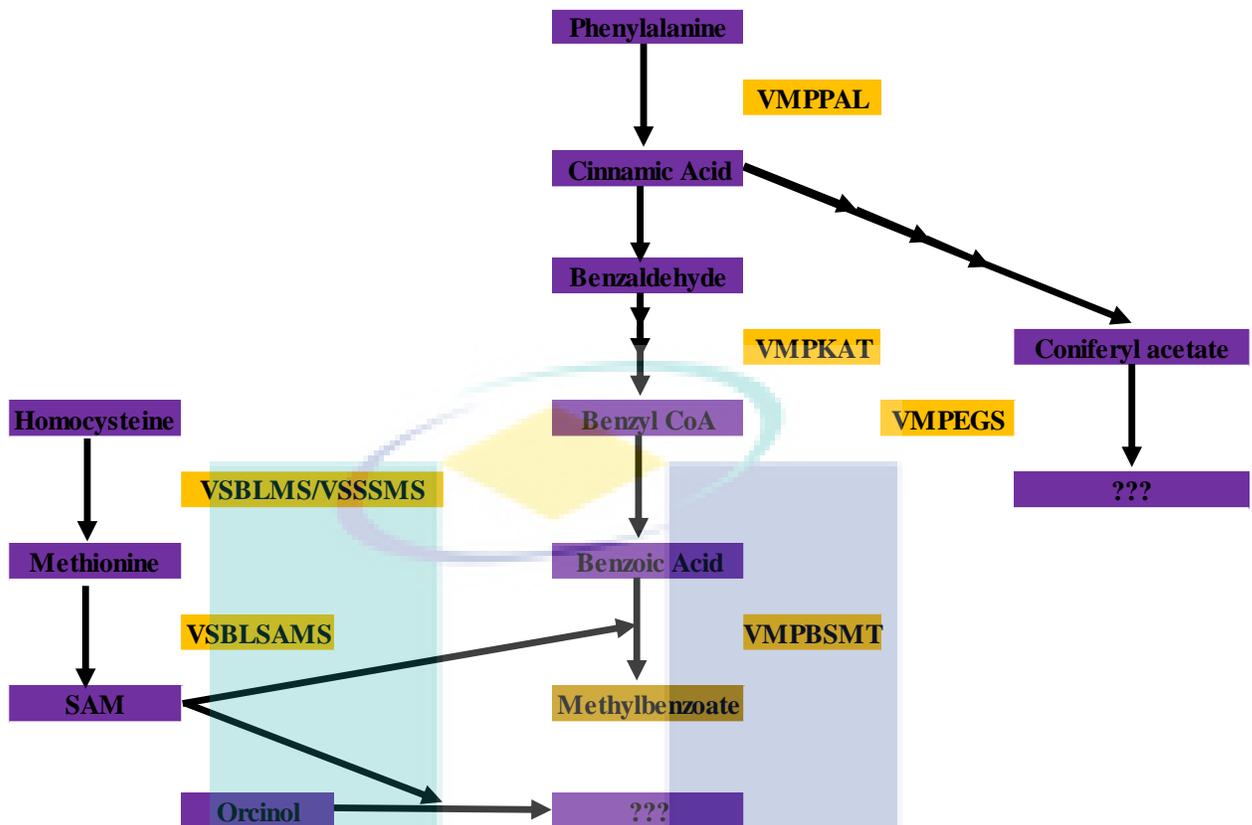


Figure 4.2: Postulation of benzenoid and phenylpropanoid pathway of Vandaceous orchid. This figure shows the self-designed pathway based on the postulation regarding to the regulation of benzenoid phenylpropanoid pathways in selected fragrant flowers (Dudareva *et al.*, 2013; Schuurink *et al.*, 2006; Scalliet *et al.*, 2008). The involvement of fragrance-related enzymes encoded by VMPKAT, VMPBSMT, VMPEGS and VMPOOMT transcripts are shown in the figure together with VMPPAL, VSBLMS, VSBLMS and VSSMS that have been identified via cDNA-RDA experiment.

donor SAM to benzoic acid for biosynthesis of methylbenzoate. After the methylation of benzoic acid with the catalytic activity of benzoic acid salicylic acid methyltransferase (BSMT) enzyme, methylbenzoate compound is produced. Methylbenzoate is also a benzenoid product besides phenylethyl alcohol and phenylacetaldehyde that are also synthesized from phenylalanine. This suggested theory about the benzenoid production in fragrant orchids might be related to the previous study on *P. hybrida* 'Mitchell' (Robert *et al.*, 2006). From the incomplete result of our study, we prefer to suggest further studies need to be done in order to enable further elucidation of the regulation of benzenoid biosynthesis such as full-length isolation and molecular characterization of SAM cycle-regulated biosynthetic genes and PAL gene in vandaceous orchids.

4.2 MOLECULAR CHARACTERIZATION OF FULL LENGTH SEQUENCES OF SELECTED TRANSCRIPTS IN BENZENOID AND PHENYLPROPANOID PATHWAY OF VANDA MIMI PALMER

In this study, partial sequence of four selected transcripts that might be involved in benzenoid and phenylpropanoids pathway of *V. Mimi Palmer* were selected for molecular characterisation by isolation of full length sequences followed by sequence analyses using molecular and bioinformatics approaches. Through cDNA-RDA analysis (section 4.1), there are four candidates of fragrance-related transcripts possibly involved in benzenoid and phenylpropanoid biosynthesis including phenylalanine ammonia lyase (VMPPAL) and S-adenosyl-L-methionine synthase (VMPSAM) from *V. Mimi Palmer* as well as two methionine synthase (VSBLMS and VSSSMS) transcripts that have been identified from *V. Small Boy Leong* and *Vandachostylis Sri-Siam*, respectively. Unfortunately, the involvement of all the transcripts and their respective enzymes in benzenoid and phenylpropanoid pathway of Vandaceous orchids can be considered as indirect and not specific enough for final fragrance product biosynthesis in the orchids. Besides that, all the enzymes might also contribute for biosynthesis of non-fragrance compounds either in the benzenoid and phenylpropanoid pathway or other metabolisms. Thus, another four fragrance-related transcripts encoding enzymes that are directly involved in the biosynthesis of final fragrance products that have been previously identified from the floral

cDNA library of *V. Mimi Palmer* namely benzoic acid carboxymethyltransferase (VMPBSMT), 3-keto-acyl-CoA thiolase (VMPKAT), eugenol synthase (VMPEGS) and orcinol-O-methyltransferase (VMPOOMT) were selected for further sequence characterisation using molecular and bioinformatics approaches. The full length sequences of isolated fragrance-related transcripts subjected to sequence analysis which included Clustal W multiple alignment, motif identification and phylogenetic tree construction. Furthermore, motifs search and signal peptide prediction was carried out using ExPASy tool (<http://br.expasy.org/tools/>) and Localizome software (<http://localodom.kobic.re.kr/LocaloDom/index.htm>) (Lee *et al.*, 2006). Besides that, the deduced amino acid sequence of the transcripts were analysed using multiple online bioinformatics servers including ScanProsite (de Castro *et al.*, 2006), ProP 1.0 server (Duckert *et al.*, 2004), TMHMM server 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>), SignalP 4.1 server (Petersen *et al.*, 2011), ChloroP 1.1 server (Emanuelsson *et al.*, 1999) and TargetP 1.1 server (Emanuelsson *et al.*, 2000).

4.2.1 Sequence Analysis on Full Length Sequence of *Vanda Mimi Palmer* 3-Keto-acyl CoA Thiolase (VMPKAT)

The full-length sequence of VMPKAT transcript (GenBank accession no: KF278718) comprises 1,577 bp (Figure 4.3), encoding a polypeptide of 450 amino acid residues consisting 1,350 bp open reading frame (ORF) flanked by 41 bp of 5'-untranslated region (UTR) and 185 bp of 3'-UTR including a poly-A tail (Appendix H(a)). The predicted molecular weight of this protein is 47.1 kD with an isoelectric point (pI) of 7.04. BLASTP analysis (NCBI) shows that the deduced amino acid sequence of VMPKAT is 73% to 76% homologous to 3-keto-acyl-CoA thiolase from other plants such as *Cucumis sativus* (NP_001267690.1), *Brachypodium distachyon* (XP_003570403.1), *Petunia x hybrida* (ACV70032.1), *Arabidopsis thaliana* (NP_180873.1), *Triticum aestivum* (BAI66423.1) and *Oryza sativa* (AIV98504.1). To our knowledge, there is no reported work on 3-ketoacyl coA thiolase (KAT) transcript from vandaceous orchids. In *A. thaliana*, a well-studied plant model, ketoacyl-CoA thiolase (KAT) has been reported to play an important role in the β -oxidative pathway leading to the production of benzoyl-CoA from 3-oxo-3-

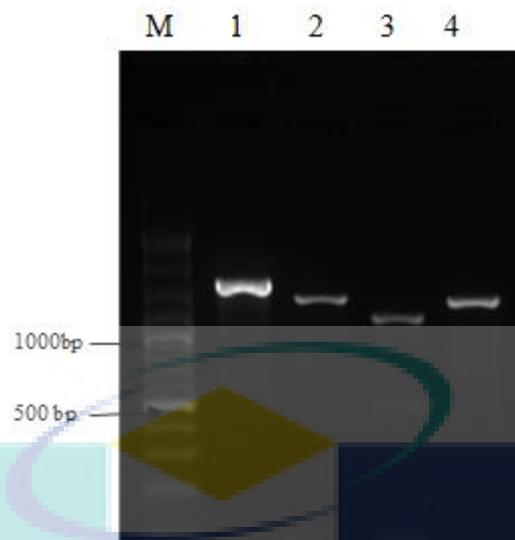


Figure 4.3: Amplified PCR products of full ORF of selected fragrance-related transcripts. The PCR products were electrophoresed on 1.2 % (w/v) agarose gel. Lane M: 100 bp DNA marker (Vivantis, Malaysia); Lane 1: *Vanda Mimi Palmer* of 3-ketoacyl coA thiolase (VMPKAT); Lane 2: *Vanda Mimi Palmer* benzoic acid/salicylic acid carboxyl-methyltransferase (VMPBSMT); Lane 3: *Vanda Mimi Palmer* eugenol synthase (VMPEGS); Lane 4: *Vanda Mimi Palmer* orcinol-O-methyltransferase (VMPOOMT).

phenylpropionyl-CoA (benzoylacetyl-CoA) (Carrie *et al.*, 2007). Besides that, KAT transcript has been reported to be involved in benzenoid pathway of *P. hybrida*, a well studied scented flowers in genomics, transcriptomics and metabolomics. In floral tissue of *P. hybrida*, a 3-ketoacyl-CoA thiolase enzyme (*PhKAT*) was detected to be present in the peroxisomes of its petals suggesting the involvement of this organelle in volatile benzenoid biosynthesis and the production of benzoic acid (Van Moerkercke *et al.*, 2009). Meanwhile in *V. Mimi Palmer*, VMPKAT might be involved in the upstream step of benzoic acid biosynthesis that becomes the main substrate for methylbenzoate compound biosynthesis whereby the methylbenzoate compound has been detected to be emitted by fully-open flower of *V. Mimi Palmer* as reported by Mohd-Hairul *et al.* (2010a).

A constructed phylogenetic tree using the MEGA version 4.0 software (Tamura *et al.*, 2007) based on neighbouring-joining method on the deduced amino acid sequence of VMPKAT and other homologous proteins that were retrieved from the NCBI GenBank database has shown that VMPKAT is clustered together in the same clade with a KAT protein from *Oryza sativa* Japonica Group (NP001048523.1) with the bootstrap value of 100 (Figure 4.4). The bootstrap value of 100 reflects 100% confidence level on the clade and can be considered as statistically significant whereby the statistical data is considered as significant for the bootstrap value of more than 95% as proposed by Felsenstein (1985). This result was as expected since both *V. Mimi Palmer* and *Oryza sativa* are classified as monocotyledonous angiosperm plants and might have stronger genetic relationship compared to other dicotyledonous plants that have been selected to be used in the phylogenetic tree construction including *A. thaliana*, *Glycine max* and *Solanum lycopersicum*. This result agrees with the previous phylogenetic tree constructed on phenylacetaldehyde synthase of *V. Mimi Palmer* (VMPPAAS) with homologous protein sequence that are available in the NCBI GenBank database whereby the VMPPAAS is clustered together with tryptophan decarboxylase of *Oryza sativa* suggesting the close genetic relatedness between vandaceous orchid and *Oryza sativa* as monocotyledonous angiosperm plants (Mohd-Hairul *et al.* 2010b). Unfortunately, no 3-keto-acyl coA thiolase amino acid sequence from any other orchid species or hybrids is currently available at the NCBI GenBank database to be used in this phylogenetic tree construction and analysis. If

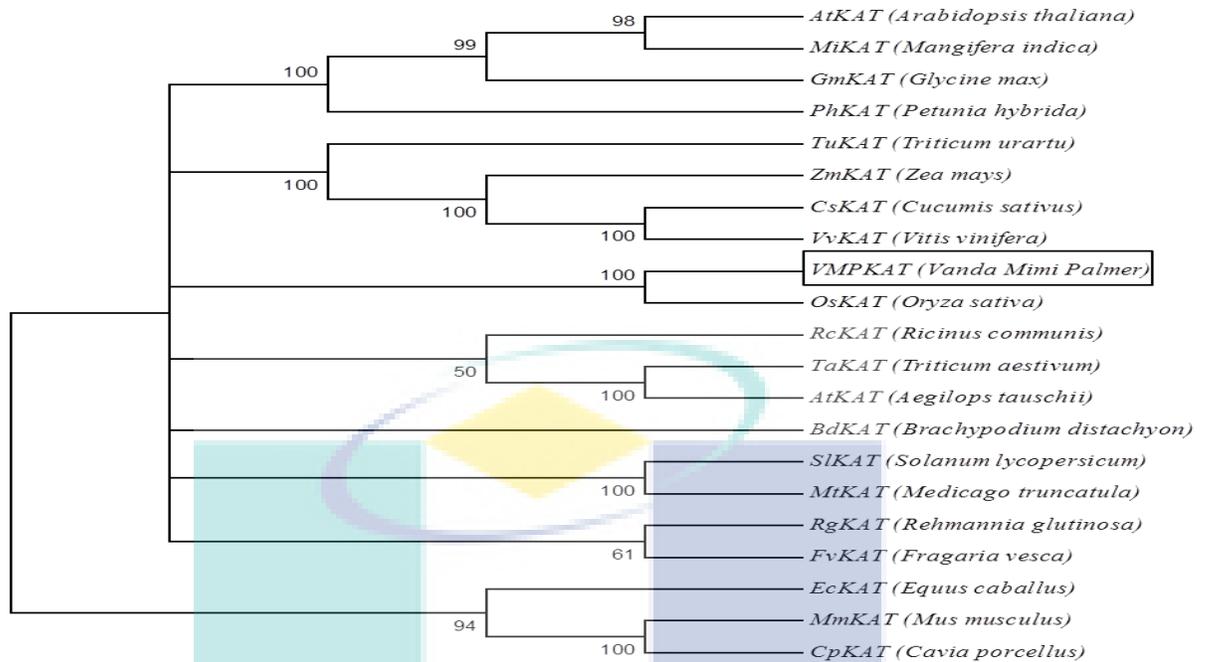


Figure 4.4: Phylogenetic tree of VMPKAT with homologous proteins. The phylogenetic tree is constructed by MEGA software version 4.0 using amino acid sequence of VMPKAT (highlighted in box) with homologous proteins that are available at the NCBI GenBank database.

Note: AtKAT (Accession no. NP180873.1), MiKAT (Accession no. CAA53078.1), GmKAT (Accession no. XP003555712.1), PhKAT (Accession no. ACV70032.1), TuKAT (Accession no. EMS60447.1), ZmKAT (Accession no. NP001241698.1), CsKAT (Accession no. XP004138952.1), VvKAT (Accession no. XP002285653.1), OsKAT (Accession no. NP001048523.1), RcKAT (Accession no. XP002518136.1), TaKAT (Accession no. BAI66423.1), AtKAT (Accession no. EMT31542.1), BdKAT (Accession no. XP003570403.1), SlKAT (Accession no. XP004247103.1), MtKAT (Accession no. XP003604976.1), RgKAT (Accession no. AGC59769.1), FvKAT (Accession no. XP004290230.1), EcKAT (Accession no. XP001488609.1), MmKAT (Accession no. NP570934.1) and CpKAT (Accession no. XP003464147.1).

homologous amino acid sequences from orchids are available, the sequences might show a closer genetic relatedness with the VMPKAT of *V. Mimi Palmer* due to the fact that orchid species or hybrids as well as *V. Mimi Palmer* are genetically related and grouped in the same Orchidaceae family. In contrast, homologous KAT protein from *Equus caballus* (XP001488609.1), *Mus musculus* (NP570934.1) and *Cavia porcellus* (XP003464147.1) are branched out from the main cluster with a bootstrap value of 94, representing 94% of the branch confidence level. The outside cluster of those KAT proteins might reflect their far genetic -relatedness to *V. Mimi Palmer* compared to other plants that are used for the phylogenetic tree construction and analysis. The confidence level of the far genetic relatedness between this orchid hybrid with those three plants is quite high since it is supported by the bootstrap value of 94. Eventhough the the bootstrap value is less than 95, the branch is still reliable since the bootstrap value for the branch is more than 70 and thus it is recommended for phylogenetic tree reliability in estimating evolutionary lineages as suggested by Efron *et al.* (1996).

Besides that, sequence analysis on the deduced amino acid sequence of VMPKAT using PROSITE (<http://prosite.expasy.org/>) has shown that the protein is hit to three distinct patterns of sequence at high probability of motifs, predicted as Thiolase I (Thiolase acyl-enzyme intermediate)(aa₁₃₁₋₁₄₉), Thiolase II (aa₃₈₀₋₃₉₆) and Thiolase active site (aa₄₁₇₋₄₃₀) (Figure 4.5). In *P. hybrida*, a well studied scented flower for benzenoid and phenylpropanoid pathway, the presence of Thiolase I domain (3-ketoacyl-CoA thiolase, E.C. 2.3.1.16) has been identified in its 3-ketoacyl-CoA thiolase (PhKAT). The Thiolase I domain has been reported to be involved in catalyzing the removal of an acetyl group from an acyl-CoA in fatty acid β -oxidation. Meanwhile, thiolase II domain (acetoacetyl-CoA synthase, E.C. 2.3.1.9) has been identified to be involved in catalysing the reverse reaction whereby two molecules of acetyl-CoA is condensed to acetoacetyl-CoA in *P. hybrida* (Kursula *et al.*, 2005). Furthermore, Thiolase II domain has also been reported to play another role as a regulatory enzyme in isoprenoid biosynthesis via mevalonate (MVA) pathway as previously reported in *Medicago sativa*, *A. thaliana*, *Populous trichocarpa* and *Hevea brasiliensis* (Kirby and Keasling, 2009). Besides that, sequence analysis on the deduced VMPKAT amino acid sequence using ScanProsite (de Castro *et al.*, 2006), a more

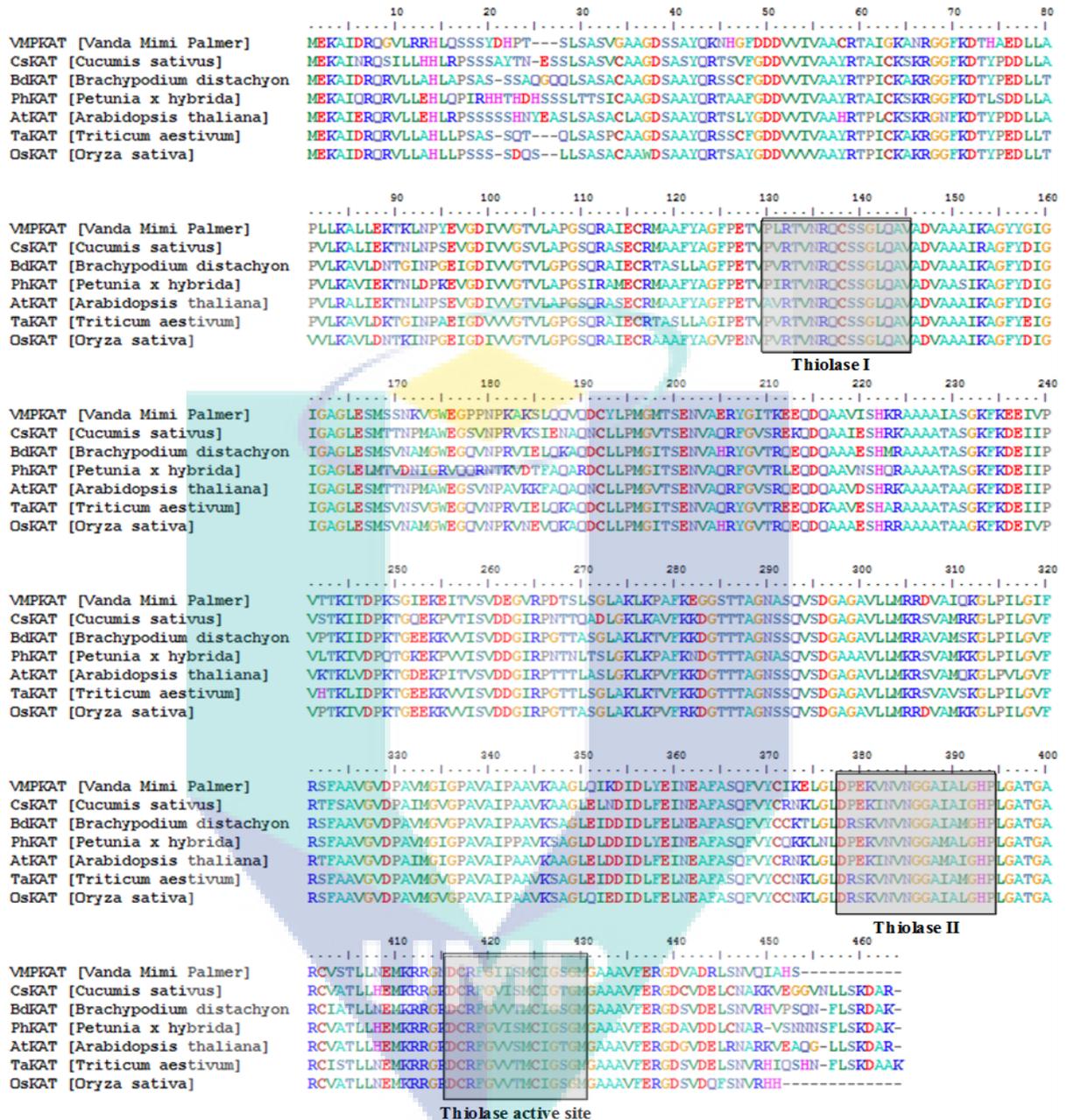


Figure 4.5: Alignment of VMPKAT amino acid sequence with closely related proteins from other plants. Thiolase I (Thiolase acyl-enzyme intermediate)(aa₁₃₁₋₁₄₉), Thiolase II (aa₃₈₀₋₃₉₆) and Thiolase active site (aa₄₁₇₋₄₃₀) are highlighted in box.

Note: CsKAT (Accession no. NP_001267690.1), BdKAT (Accession no. XP_003570403.1), PhKAT (Accession no. ACV70032.1), AtKAT (Accession no. NP_180873.1), TaKAT (Accession no. BAI66423.1) and OsKAT (Accession no. AIV98504.1)].

detail analysis program has shown the presence of all the three thiolase domains including Thiolase I, Thiolase II and Thiolase active site at the same amino acid positions as predicted in the Prosite analysis. Besides that, the ScanProsite program has predicted the presence of another active site for proton acceptor at amino acid position of 422 in the VMPKAT protein. In addition, the ScanProsite program has also predicted the presence of twelve N-myristoylation sites (GsqrAI, GigiGA, GigaGL, GgstTA, GsttAG, GifrSF, GvdpAV, GigpAV, GgaiAL, GatgAR, GsgmGA and GMgaAA), five casein kinase II phosphorylation sites (SsyD, ThaE, Tke E, SgiE and SvdE), five protein kinase C phosphorylation sites (SqR, SnK, ShK, SgK and TtK), a N-glycosylation site (NASQ), and a site for cell attachment sequence (RGD). Unfortunately, the involvement of the N-myristoylation sites, casein kinase II phosphorylation sites, protein kinase C phosphorylation site, N-glycosylation site and cell attachment site in fragrance biosynthesis especially in KAT enzymes has not been yet discovered. Further investigation on the function role of those sites might be determined via direct mutagenesis studies.

Meanwhile, sequence analysis of the deduced VMPKAT amino acid sequence using Localizome software (Lee *et al.*, 2006) has shown the presence of another two specific domains which are thiolase N-terminal and thiolase C-terminal, at the position of amino acid residue 46 to 304 and 311 to 435, respectively. In addition, the Localizome software has also predicted that no signal peptide is present in the amino acid sequence of VMPKAT suggesting that the VMPKAT might be localised in cytosol instead of plastid. This analysis result is supported by sequence analysis using multiple subcellular localization analysis servers including TMHMM 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>), SignalP 4.1 (Petersen *et al.*, 2011), ChloroP 1.1 (Emanuelsson *et al.*, 1999) and TargetP 1.1 (Emanuelsson *et al.*, 2000) servers. From the analyses using multiple servers, it has been predicted that the deduced amino acid sequences of VMPKAT do not have any signal peptide cleavage site, propeptide cleavage site (Arginine/Lysine), transmembrane helices, secretory pathway signal peptide, chloroplast transit peptide, N-terminal presequences and mitochondrial targeting peptide due to the score of less than 0.5. Thus, it indicates that VMPKAT might be localised in cytosol since there is no chloroplast signal peptide and mitochondrial targeting peptide which are involved in transportation of VMPKAT protein

from chloroplast to cytosol or mitochondria to cytosol vice versa. This analyses results agree with the benzenoid and phenylpropanoid pathway in well studied scented flowers including *P. hybrida*, *R. hybrida* and *A. majus* whereby all their fragrance-related transcripts in the pathway have not been reported to be localised in plastid (Colquhoun *et al.*, 2010; 2011; 2012). Meanwhile, for the other class of fragrance compounds which are terpenoid compounds, the responsible enzymes in catalysing the biosynthesis of terpenoid compounds have been reported to be localised in cytosol for sesquiterpene synthases while monoterpene synthases have been reported to be localised in plastid (Chen *et al.*, 2011; Gutensohn *et al.*, 2013).

4.2.2 Sequence Analysis on Full Length Sequence of *Vanda Mimi* Palmer Benzoic Acid/Salicylic acid Carboxyl Methyltransferase (VMPBSMT)

Analysis on full length cDNA of VMPBSMT (GenBank accession no: KF278719) of 1,456 bp (refer Figure 4.3) showed that it contains 1,134 bp of open reading frame (ORF), 90 bp of 5'-untranslated region (UTR) and 231 bp of 3'-UTR including a poly-A tail (Appendix H(b)). The predicted molecular weight of this protein is 43.2 kD with an isoelectric point (pI) of 4.97. BLASTP analysis showed that the deduced amino acid sequence of VMPBSMT is 43 to 47 % homologous to carboxyl methyltransferase from other plants such as *Lilium hybrid cultivar* (AIG92833.1), *Vitis vinifera* (XP_002265700.3), *Medicago truncatula* (XP_003612068.1), *Nicotiana suaveolens* (ACZ55216.1) and *Nicotiana alata* (ACZ55219.1). From the BLASTP analysis, VMPBSMT has been detected to have a conserved motif for SAM dependent carboxyl methyltransferase group as found in BSMT sequences from other plants (Figure 4.6). Benzoic acid/salicylic acid carboxyl methyltransferase (BSMT) has been reported to catalyse the formation of methylbenzoate (MeBA) from benzoic acid in the presence *S*-adenosyl-L methionine (SAM) as cofactor for the enzyme (Chen *et al.*, 2003). This type of methyltransferase uses benzoic acid as substrate and included in SABATH family of methyltransferase (MT), a class of small molecule MTs that are found in plants. The name of 'SABATH' are given for the first three enzymes to be characterised in the family including salicylic acid MT, benzoic acid MT and theobromine synthase (D'Auria *et al.*, 2003). Thus, benzoic acid carboxyl



Figure 4.6: Alignment of amino acid sequence of VMPBSMT with other closely related proteins that are available at the NCBI GenBank database. SAM dependent carboxyl methyltransferase attachment site is highlighted in box (HSSYSLMWLSQVP).

Note: LhBSMT (Accession no. AIG92833.1), VvBAMT (Accession no. XP_002265700.3), MtSAMT (Accession no. Accession no. XP_003612068.1), NsSAMT (Accession no. ACZ55216.1)] and NaSAMT (Accession no. ACZ55219.1).

methyltransferase of *V. Mimi Palmer* (VMPBSMT) might be involved in catalysing the biosynthesis of methylbenzoate from benzoic acid as its main substrate in the presence of SAM as cofactor for VMPBSMT enzyme in its catalytic reaction. This postulation is strongly supported by the presence of methylbenzoate compound as detected previously in the scent of fully-opened flower of the *V. Mimi Palmer* as reported by Mohd-Hairul *et al.* (2010a).

Phylogenetic tree of VMPBSMT and its homologous proteins from other plants (Figure 4.7) that has been constructed using MEGA version 4.0 based on neighbour-joining method (Tamura *et al.*, 2007) has shown that VMPBSMT is clustered together with BSMT proteins from *Crocus sativus* (CAD70566.1) in the same clade and supported by the bootstrap value of 69. The confidence level of 69 % on the clade reflecting less reliability due to the fact that the bootstrap value of less than 70 as recommended by Efron *et al.* (1996). This result could be due to the less genetic relatedness between clustered VMPBSMT and the protein sequences from *Crocus sativus*. However, both *V. Mimi Palmer* and *C. sativus* are classified as monocotyledonous angiosperm plants and interestingly both of the plants are classified in the same order of Asparagales. Meanwhile, homologous protein of *P. hybrida* and other well-studied dicotyledonous plants including *Nicotiana benthamiana* and *Ricinus communis* have been identified to be not in the same cluster with VMPBSMT. This result is as expected due to the fact that orchids are monocotyledonous angiosperm plants and their morphological characteristics are totally different compared to all the selected dicotyledonous plants. Besides that, a homologous protein from another well-studied scented plant, *A. majus* or well-known as snapdragon, is branched out from the cluster, reflecting the far genetic relatedness between snapdragon and vandaceous orchid. This result is exactly as expected since snapdragon is a dicotyledonous plant and has different morphological structure and grown in totally different areas and conditions compared to orchids.

Localizome analysis has shown the presence of a domain for SAM dependent carboxyl methyltransferase in the amino acid sequence of VMPBSMT at amino acid residue 36 to 368 thus supporting the prediction of functional roles to VMPBSMT in

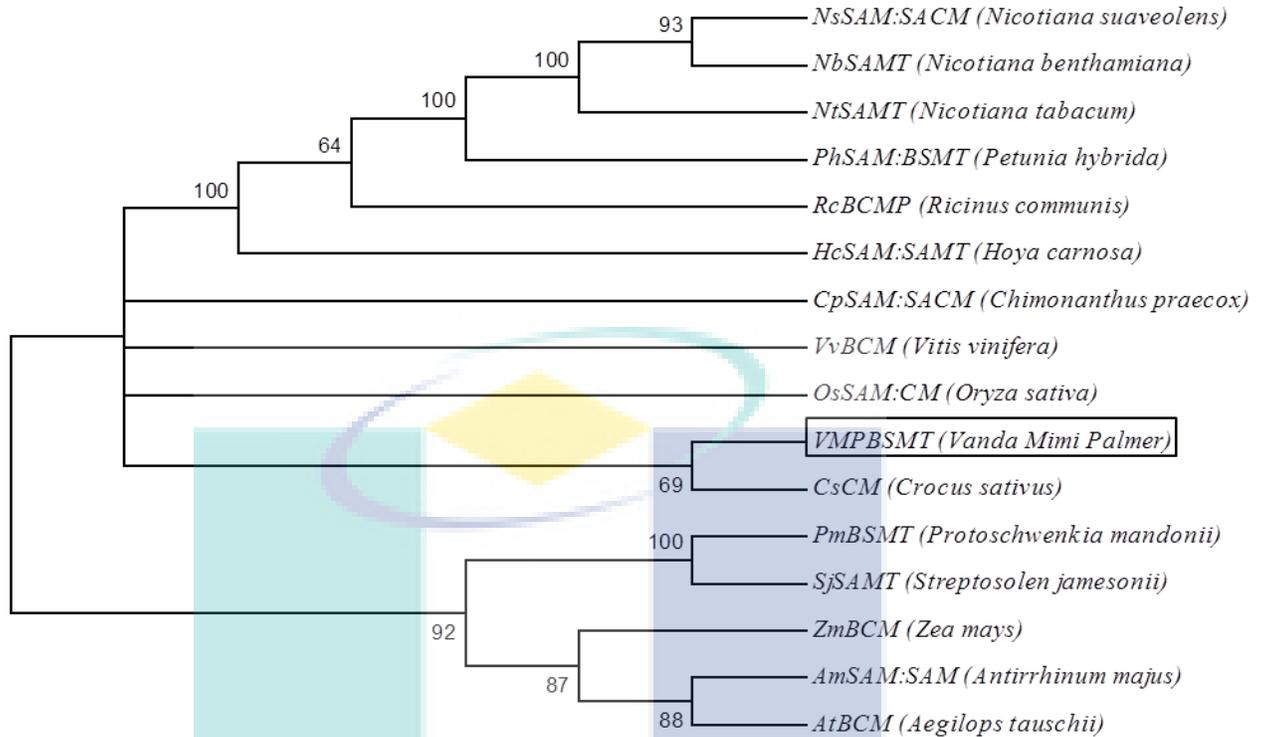


Figure 4.7: Phylogenetic tree of VMPBSMT with homologous proteins. The phylogenetic tree was constructed using MEGA software version 4.0 using amino acid sequence of VMPBSMT (as boxed) with homologous proteins available at the NCBI GenBank database.

Note: NbSAMT (Accession no. AFS35577.1), NtSAMT (Accession no. AAW66850.1), PhBSMT (Accession no. ABF50941.1), RcBMT (Accession no. XP002521542.1), HcSAM:SAMT (Accession no. CA105934.1), CpSAMT (Accession no. ABU88887.2), CpSAM:SAMT (Accession no. ABU88887.2), VvBMT (Accession no. XP002265700.2), OsSAM:MT (Accession no. OS06G20920.1), CsMT (Accession no. CAD70566.1), PmBSMT (Accession no. ABO71012.1), SjSAMT (Accession no. AAW66842.1), ZmBMT (Accession no. NP001147683.1), AmSAMT (Accession no. AF515284) and AtBMT (Accession no. EMT10135.1).

catalysing the methylation of benzoic acid in the biosynthesis of methylbenzoate as one of the final product in the scent of *V. Mimi Palmer*. Besides that, the software predicts that VMPBSMT protein does not have any signal peptide as well as transmembrane domain. In addition, sequence analysis on VMPBSMT using multiple subcellular localisation analysis servers including TMHMM 2.0, SignalP 4.1 (Petersen *et al.*, 2011), ChloroP 1.1 (Emanuelsson *et al.*, 1999), and TargetP 1.1 (Emanuelsson *et al.*, 2000) servers has shown that VMPBSMT protein is predicted to have no signal peptide cleavage site, propeptide cleavage site (Arginine/Lysine), transmembrane helices, secretory pathway signal peptide, chloroplast transit peptide, N-terminal presequences as well as mitochondrial targeting peptide due to the score of less than 0.5. Thus, VMPBSMT protein is predicted to be localised in cytosol instead of plastid or mitochondria. This postulation agrees with the characteristic of fragrance-related transcripts in benzenoid and phenylpropanoid pathway of other scented plants including *P. hybrida*, *A. majus* as well as *R. hybrida* whereby the respective enzymes in benzenoid and phenylpropanoid pathway of the plants are localised in cytosol instead of plastid (Boatright *et al.*, 2004; Effmert *et al.*, 2005; Negre *et al.*, 2003; Pott *et al.*, 2002). In addition, sequence analysis that was carried out using ScanProsite has predicted the presence of two N-myristoylation sites (GsgNS and GSpgSF), six casein kinase II phosphorylation sites (SyaE, TlpE, SlcD, SigE, SnwD and TlrE), two N-glycosylation sites (NRTA and NESY) and two protein kinase C phosphorylation sites (SsR and TIR). Unfortunately, the involvement of those motifs that have been identified using the ScanProsite software in fragrance aspect is still unclear and far from being understood. Thus in future, site-directed mutagenesis approach might be the best suitable option in determining the contribution and involvement of each motif in fragrance biosynthesis especially in benzenoid and phenylpropanoid pathway of Vandaceous orchids.

4.2.3 Sequence Analysis on Full Length Sequence of *Vanda Mimi Palmer* Eugenol Synthase (VMPEGS)

The full transcript of VMPEGS (NCBI GenBank accession number: KF278720) consists of 1,314 bp (refer Figure 4.3), encoding a polypeptide of 306 amino acid residues (Appendix Hc) with a molecular mass of 33.3 kD and an isoelectric point (pI) of 5.69.

BLASTP analysis on the deduced amino acid sequence of VMPEGS has shown that VMPEGS is homologous to eugenol synthase (EGS) sequences from other plants such as *Vitis vinifera* (XP_002283921.2), *Petunia x hybrida* (ABR24115.1), *Clarkia breweri* (ABR24114.1), *Dendrobium catenatum* (AID53186.1) and phenylcoumaran benzylic ether reductase of *Linum corymbulosum* (ACA60731.1) with 59 % to 64 % identity. Based on BLASTP analysis, VMPEGS has been detected to have a conserved motif for phenylcoumaran benzylic ether reductase / NmrA-like family group as found in EGS sequences from other plants (Figure 4.8). Furthermore, EGS has been reported to be involved in catalysing a reductive displacement of acetate group from propenyl side chain of its substrate coniferyl acetate to produce the allyl-phenylpropene eugenol by converting coniferyl acetate to eugenol in the presence of NADPH (Koeduka *et al.*, 2006; Dexter *et al.*, 2007). This phenylpropene-forming enzyme belongs to a structural family of NADPH-dependent reductases including pinosresinol–lariciresinol reductase, isoflavone reductase and phenylcoumaran benzylic ether reductase (Gang *et al.*, 1999). Thus in *V. Mimi Palmer*, VMPEGS is possibly involved in catalysing the biosynthesis of eugenol compound from coniferyl acetate in the presence of NADPH as cofactor in the enzymatic reaction. In future, the functional role of VMPEGS should be carried out by cloning and expression in any bacterial system and followed by recombinant enzyme production and determination of its activity using coniferyl acetate as the main substrate and NADPH as cofactor in the *in vitro* enzymatic assay.

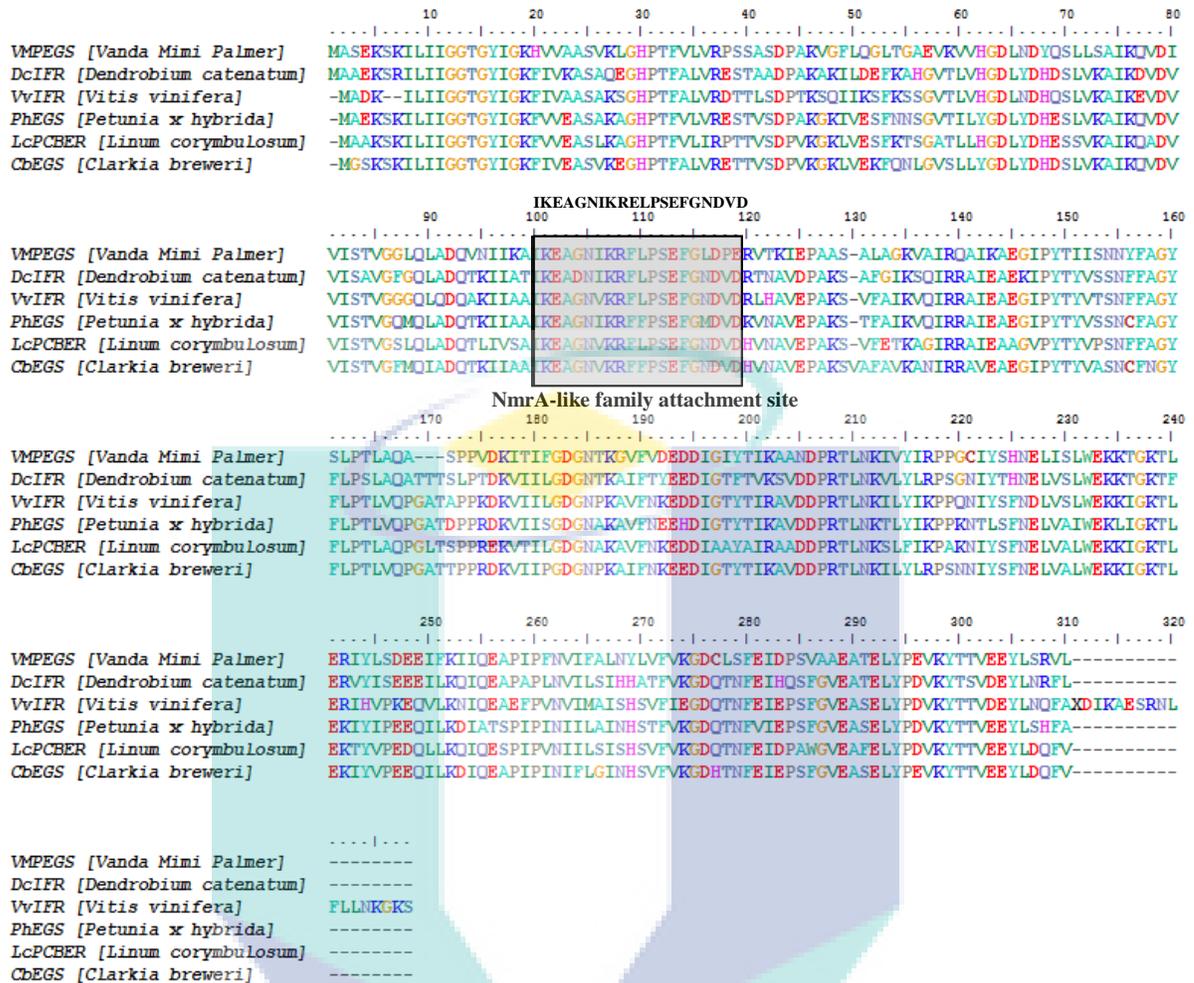
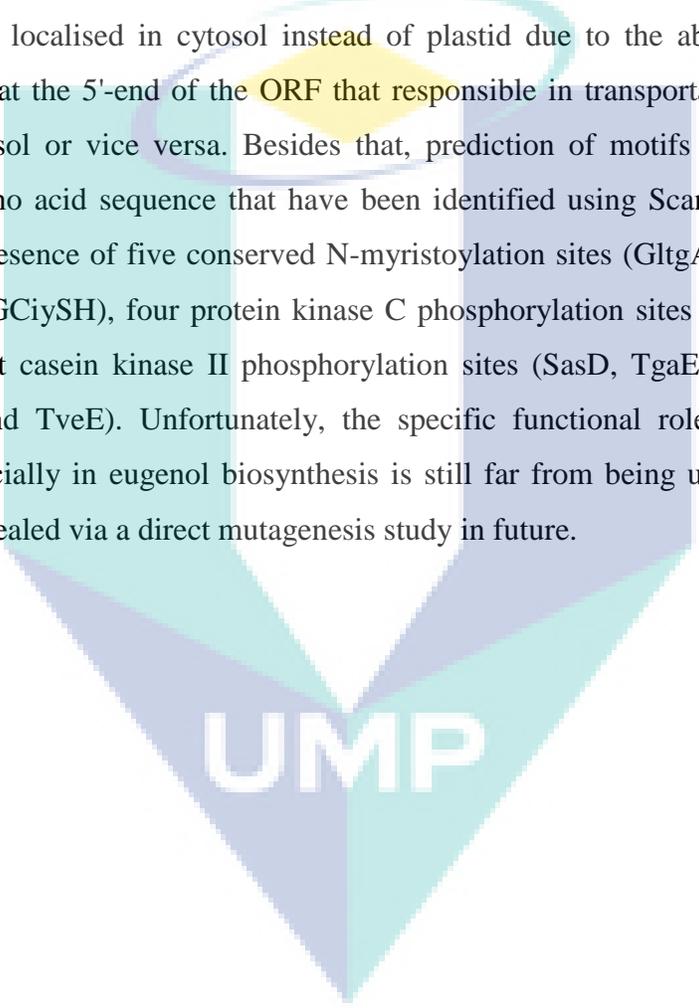


Figure 4.8: Alignment of amino acid sequences of VMPEGS with other closely related protein sequences from the NCBI GenBank database. NmrA-like family attachment site is highlighted in box (IKEAGNIKRELPSEFGNDVD).

Note: VvIFR (Accession no. XP_002283921.2), PhEGS (Accession no. ABR24115.1), CbEGS (Accession no. ABR24114.1), DcIFR (Accession no. AID53186.1) and LcPCBER (Accession no. ACA60731.1).

A phylogenetic tree (Figure 4.9) constructed using MEGA version 4.0 based on neighbour-joining method (Tamura *et al.*, 2007) to estimate the genetic relatedness between amino acid sequence of VMPEGS with other plant eugenol synthases or phenylpropanoid-like enzyme family has shown that VMPEGS is clustered together with EGS proteins from *Salvia fruticosa* (ADD14078.1) and *Vitis vinifera* (CAI56334.1) in the same clade and supported by a bootstrap value of 90, reflecting 90 % confidence level on the clade. Based on the constructed phylogenetic tree, only phenylcoumaran benzylic ether reductase of *Pinus strobus* (ABF39004.1) is located outside of the cluster, suggesting that the plant might have different ancestral line with the other eugenol synthase from other plants including *V. Mimi Palmer*. Meanwhile, a well characterised eugenol synthase from *C. breweri* (CbEGS) are located in different clade of the same cluster and supported by a bootstrap value of 63. The bootstrap value of 63 on the cluster that is lower than 70% reflecting the low confidence level on the cluster. This is as expected since *C. breweri* might have very far genetic relationship compared to orchids due to its morphological structure as well as the different order and family in taxonomical classification. However, VMPEGS might not be closely related to several eugenol synthases that have been recently isolated from *Gymnadenia* orchids (*Gymnadenia conopsea*, *Gymnadenia densiflora* and *Gymnadenia odoratissima*) by Gupta *et al.* (2014) due to the fact that VMPEGS is excluded from eugenol synthases of *Gymnadenia* species as shown in phylogenetic tree. Besides that, VMPEGS is also detected to be not well-aligned with amino acid sequence of those eugenol synthases from *Gymnadenia* species, even though they classified in the same Orchidaceae family (Figure 4.10 and Figure 4.11). The difference could be due to their different genera which are *Vanda* and *Gymnadenia*, respectively. However, eugenol synthase of *V. Mimi Palmer* might share similar functionality in catalysing biosynthesis of eugenol compound from coniferyl acetate due the presence of NmrA-family domain in the sequence of VMPEGS that are also identified to be presence in eugenol synthases from well studied scented flowers including *C. breweri* and *P. hybrida* as discussed earlier.

Other than that, subcellular localisation analysis using multiple servers including TMHMM 2.0, SignalP 4.1 (Petersen *et al.*, 2011), ChloroP 1.1 (Emanuelsson *et al.*, 1999) and TargetP 1.1 (Emanuelsson *et al.*, 2000) has shown that the deduced amino acid sequence of VMPEGS protein is predicted to have no signal peptide cleavage site, propeptide cleavage site (Arginine/Lysine), transmembrane helices, secretory pathway signal peptide, chloroplast transit peptide, N-terminal presequences as well as mitochondrial targeting peptide due to the score of less than 0.5. Thus, VMPEGS is expected to be localised in cytosol instead of plastid due to the absence of chloroplast transit peptide at the 5'-end of the ORF that responsible in transportation of protein from plastid to cytosol or vice versa. Besides that, prediction of motifs that are available in VMPEGS amino acid sequence that have been identified using ScanProsite program has revealed the presence of five conserved N-myristoylation sites (GltgAE, GlqlAD, GipyTI, GNtkGV and GCiySH), four protein kinase C phosphorylation sites (SeK, SvK, TiK and TgK) and eight casein kinase II phosphorylation sites (SasD, TgaE, TkiE, ShnE, SlwE, SdeE, TtvE and TveE). Unfortunately, the specific functional role of those motifs in fragrance especially in eugenol biosynthesis is still far from being understood and might possibly be revealed via a direct mutagenesis study in future.

The logo for UMP (Universiti Malaysia Perlis) is a large, downward-pointing triangle composed of four smaller triangles meeting at the center. The top-left triangle is light blue, the top-right is light green, the bottom-left is light purple, and the bottom-right is light teal. The letters 'UMP' are written in a bold, white, sans-serif font across the center of the triangle.

UMP

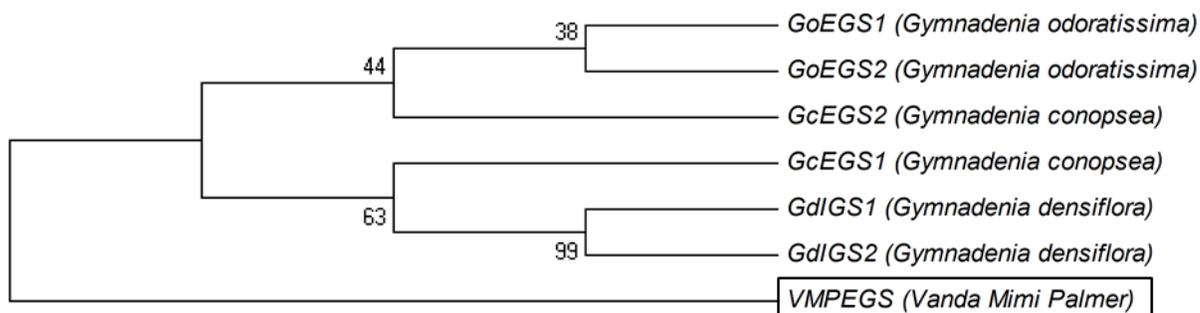
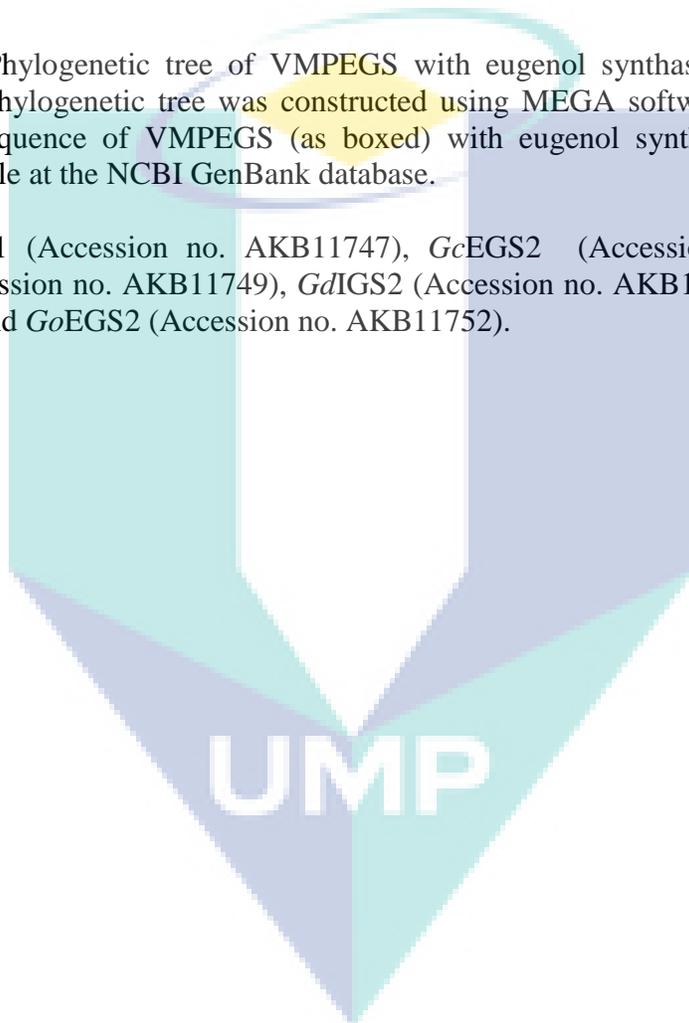


Figure 4.11: Phylogenetic tree of VMPEGS with eugenol synthases from *Gymnadenia* species. The phylogenetic tree was constructed using MEGA software version 4.0 using amino acid sequence of VMPEGS (as boxed) with eugenol synthases of *Gymnadenia* species available at the NCBI GenBank database.

Note: GcEGS1 (Accession no. AKB11747), GcEGS2 (Accession no. AKB11748), GdIGS1 (Accession no. AKB11749), GdIGS2 (Accession no. AKB11750), (Accession no. AKB11751) and GoEGS2 (Accession no. AKB11752).



4.2.4 Sequence Analysis on Full Length Sequence of *Vanda Mimi Palmer* Orcinol-O-methyltransferase (VMPOOMT)

The full-length VMPOOMT transcript (Genbank Accession no: KF278721) comprises 1,289 bp (refer Figure 4.3), encoding a polypeptide of 368 amino acid residues consisting of 1,104 bp open reading frame (ORF) and flanked by 2 bp of 5'-untranslated region (UTR) and 182 bp of 3'-UTR including a poly-A tail (Appendix H(d)). The predicted molecular weight of this protein is 41.6 kD with an isoelectric point (pI) of 5.74. The BLASTP analysis (NCBI) has shown that the deduced amino acid sequence of VMPOOMT is 45 to 47 % homologous to O-methyltransferase sequences from other plants such as *Vitis vinifera* (CAQ76879.1), *Rosa hybrid cultivar* (CAH05083.1), *Rosa chinensis* (AEC13057.1), *Rosa hugonis* (CAJ65638.1) and *Rosa gallica* (Accession no. CAJ65624.1). VMPOOMT has a conserved motif for O-methyltransferase attachment site as found in OOMT sequences from other plants (Figure 4.12). Orcinol O-methyltransferase (OOMT) is the enzyme which catalyses the transfer of a methyl group from *s*-adenosyl-*l*-methionine (SAM) to a hydroxyl functionality. It is closely related to other plant methyltransferases in which the substrates range from isoflavones to phenylpropenes with activity toward several phenolic substrates including 3,5-dihydroxytoluene (orcinol), 3-methoxy,5-hydroxytoluene (orcinol monomethyl ether), 1- methoxy, 2-hydroxy benzene (guaiacol), and eugenol. In *R. hybrida*, *Rh*OOMT has been reported to catalyse the last two steps of the biosynthetic pathway leading to the phenolic methyl ether 3,5-dimethoxytoluene (DMT) (Scalliet *et al.*, 2006; 2008). Thus in *V. Mimi Palmer*, VMPOOMT might be involved in methylation of either single or multiple phenolic substrate depending on its catalytic specificity. However, further investigation should be carried by *in vitro* enzymatic assay with different substrate in the presence of SAM compound as cofactor to confirm the exact functional role of VMPOOMT.

A phylogenetic tree constructed using MEGA version 4.0 (Tamura *et al.*, 2007) is to estimate the genetic-relatedness between VMPOOMT amino acid sequence with other plant O-methyltransferases (Figure 4.13). The phylogenetic result has shown that VMPOOMT is clustered together with OOMT proteins from *Rosa chinensis* (AEC13057.1)

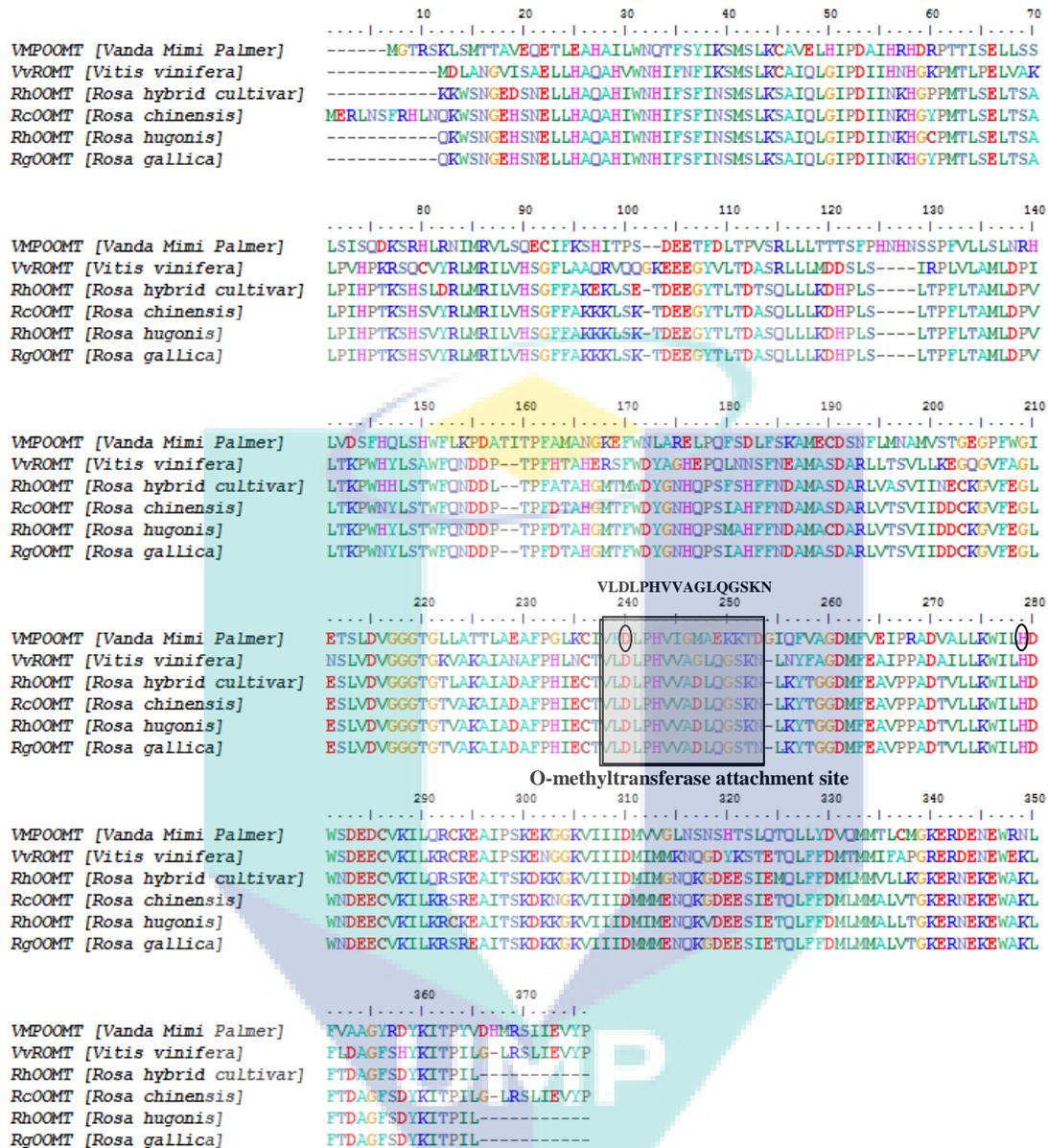


Figure 4.12: Alignment of VMPOOMT amino acid sequence with other closely related proteins that are retrieved from the NCBI GenBank database. O-methyltransferase attachment site is highlighted in box (VLDLPHVVAGLQGSKN).

Note: VvROMT (Accession no. CAQ76879.1), RhOOMT (Accession no. CAH05083.1), RcOOMT (Accession no. AEC13057.1), RhOOMT (Accession no. CAJ65638.1) and RgOOMT (Accession no. CAJ65624.1).

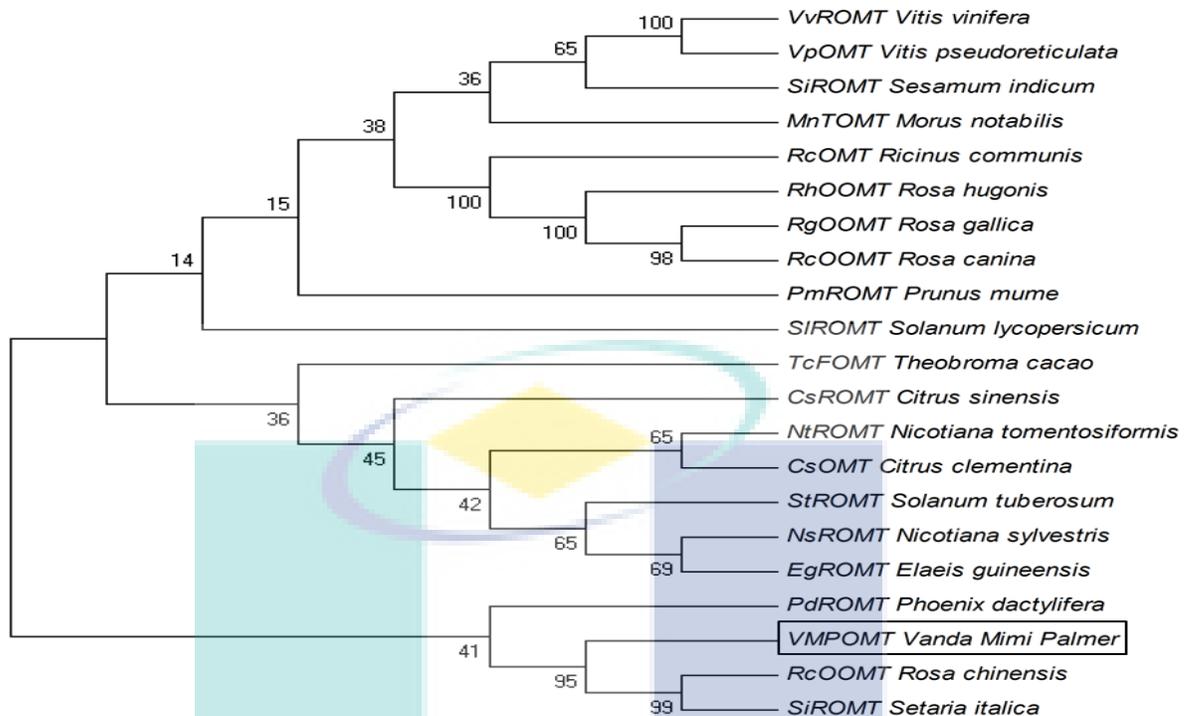


Figure 4.13: Phylogenetic tree of VMPOOMT with homologous proteins. The phylogenetic tree is constructed by MEGA software version 4.0 using amino acid sequence of VMPOOMT (as boxed) with the homologous proteins available at NCBI GenBank database.

Note: PdROMT (Accession no. XP_008783119.1), VvROMT (Accession no. XP_002278294.1), VpOMT (Accession no. AEF97347.1), NtROMT (Accession no. XP_009602136.1), NsROMT (Accession no. XP_009780941.1), EgROMT (Accession no. XP_010916033.1), SIROMT (Accession no. XP_004235335.1), StROMT (Accession no. XP_006366298.1), PmROMT (Accession no. XP_008229109.1), CsROMT (Accession no. XP_006469069.1), CcOMT (Accession no. XP_006446728.1), RcOMT (Accession no. XP_002525159.1), RhOOMT (Accession no. AAM23004.1), RcOOMT (Accession no. AEC13057.1), TcFOMT (Accession no. XP_007043718.1), RhOOMT (Accession no. CAJ65638.1), SiROMT (Accession no. XP_011091580.1), MnTOMT (Accession no. XP_010093978.1), SiROMT (Accession no. XP_004975403.1), RgOOMT (Accession no. CAJ65624.1) and RcOOMT (Accession no. CAJ65610.1).

and *Setaria italica* (XP_004975403.1), and supported by a bootstrap value of 95 for the clade that really can reflect homologous catalytic functionality based on reliability and confidence of the clade with bootstrap value more than 70 as recommended by Efron *et al.* (1996). Based on previous functional study carried out on RcOoMT, the enzyme has been identified to be involved in O-methylation process of orcinol to 3,5-dimethoxytoluene (DMT) in the presence of SAM as cofactor (Scalliet *et al.*, 2006). Interestingly, based on ClustalW multiple alignment, the homologous domain has been identified to be present in SiOoMT as well as VMPOoMT (Figure 4.13). However, all the VMPOoMT, RcOoMT and SiOoMT are not grouped in the main cluster with O-methyltransferase enzymes from other plants. This result could be due to the far genetic relatedness to the plants as well as their functional role in O-methylation of orcinol compounds instead of other substrates for O-methylation. In addition, there are a wide range of plant O-methyltransferase proteins that prefer different substrates including resorcinol, revestrol, eugenol as well as isoeugenol in the presence of SAM as cofactor (Ibrahim *et al.* 1998).

Meanwhile, localisation analysis using Localizome software (<http://localodom.kobic.re.kr/LocaloDom>; Lee *et al.*, 2006) has shown that VMPOoMT does not have any signal peptide and thus, reflecting its localisation in cytosol. Based on Localizome analysis, the presence of two functional domains which are dimerisation domain and O-methyltransferase domain have been detected in amino acid sequence of VMPOoMT at amino acid residue (aa₃₅₋₈₈) and (aa₉₃₋₃₄₄), respectively. The presence of those functional domains might reflect the catalytic functional role of VMPOoMT in O-methylation reaction. However, confirmation functional role of VMPOoMT could be confirmed in future by *in vitro* enzymatic assay. In addition to Localizome analysis result, subcellular localisation analysis using multiple online servers including TMHMM 2.0, SignalP 4.1 (Petersen *et al.*, 2011), ChloroP 1.1 (Emanuelsson *et al.*, 1999) and TargetP 1.1 (Emanuelsson *et al.*, 2000) has supported the result whereby no signal peptide cleavage site, propeptide cleavage site (Arginine/Lysine), transmembrane helices, secretory pathway signal peptide, chloroplast transit peptide, N-terminal presequences and mitochondrial targeting peptide have been detected to be present in amino acid sequence of VMPOoMT based on the score of less than 0.5. Thus, VMPOoMT is expected to be localised in cytosol

instead of plastid due to the absence of chloroplast transit peptide at the 5'-end of the ORF that responsible in transportation of protein from plastid to cytosol or vice versa.

Besides that, motifs prediction on VMPOOMT amino acid sequence by ScanProsite (de Castro *et al.*, 2006) has revealed the presence of nine casein kinase II phosphorylation sites (TavE, TisE, TpsD, SdeE, StgE, TslD, TlaE, SdeD and SiiE), a N-glycosylation site (NQTF), a protein kinase C phosphorylation site (SIK), four N-myristoylation sites (GietSL, GGgtGL, GllaTT and GLnsNS) and a conserved domain for SAM-dependent O-methyltransferase class II-type profile, starting from amino acid residue 25 to 368, yielding motif score of 75.677 generated by ScanProsite. Interestingly, other specific sites in the amino acid sequence of VMPOOMT are also detected including *S*-adenosyl-*L*-methionine binding site (D) and proton acceptor (active site) (H) at amino acid residue of 243 and 282, respectively as circled in Figure 4.12. Unfortunately, the specific functional role of those motifs in fragrance biosynthesis especially in *Vanda Mimi Palmer* is still far from clear.

4.3 DEVELOPMENT OF 3D-STRUCTURAL MODEL OF FRAGRANCE-RELATED PROTEINS FOR THEIR STRUCTURAL AND FUNCTIONAL PREDICTION

The deduced amino acid sequence of fragrance-related transcripts of *V. Mimi Palmer* including VMPKAT, VMPBSMT, VMPEGS and VMPOOMT were subjected to various threading analysis using multiple online servers by comparing the sequences with available protein models in Protein Data Bank (PDB) (<http://www.rcsb.org/pdb>) for selection of the best protein template to be used in development of their 3D-structural models. According to Vyas *et al.* (2012), selection of reference protein template is very important in development of 3D-structural model of protein of interest especially in predicting their structural and functional role based on their homologous catalytic regions, ligands and substrate specificity. Secondary structures of all fragrance-related proteins were aligned with secondary structure of their respective protein template for prediction of their homologous position of helix structures and β -strands. Secondary structural alignment is

very important in determining the reliability of 3D-structure of proteins of interest in comparison to their respective protein templates (Vyas *et al.*, 2012).

In this study, five 3D-structural models of each fragrance-related protein were developed for each protein of interest by using the MODELLER software, version 9.14 (Eswar *et al.* 2007). Among the five developed 3D-structural models, the most significant model was chosen as the best model to describe their tertiary structure. The most significant 3D-structural model of fragrance-related proteins was determined by the lowest DOPE-score as well as GA341 value. The lowest DOPE-score value is very important to minimise spatial restraint of the structure (Shen and Sali, 2006) while GA341 value that is mostly equal to 1.0 is another criteria in choosing the best 3D-structural model of protein of interest. This is due to the fact that GA341 is a score for reliability of a model which derives from statistical potentials and it is also considered to be reliable since the protein model is predicted to have GA341 score that is more than 0.7 as pre-specified cutoff value (Melo *et al.*, 2002). Besides that, the quality of the chosen 3D-structural model of all those fragrance-related proteins were measured by investigating their TM-score and RMSD values in comparison to their respective template proteins whereby Tm-score more than 0.4 is considered to be significant and meaningful while RMSD value that is lower than 1.0 is considered as accepted RMSD value (Xu and Zhang, 2011).

Besides measuring the quality of developed 3D-structural protein models with DOPE-score, GA341 value, TM score as well as RMSD value, the developed 3D-structural model of fragrance-related proteins were also assessed using multiple online protein assessment tools including VERIFY-3D (http://services.mbi.ucla.edu/Verify_3D/), ERRAT (<http://services.mbi.ucla.edu/ERRAT/>), PRO-CHECK (Laskowski *et al.*, 1993) as well as ProSA-web (Wiederstein and Sippl, 2007). Multiple assessments using various assessment protein tools were performed to further evaluate the quality of developed 3D-structural model of proteins of interest in different criteria (Dalton and Jackson, 2007). For example, quality assessment of the developed 3D-structural model of all the fragrance-related proteins that has been carried out using PROCHECK, was determined by distribution of accepted favorable protein regions based on the Ramachandran plot whereby the

disallowed region is much lower than favourable region of amino acid residues in 3D-structural protein model. In addition, stereochemical quality of a protein structure was also evaluated using PROCHECK programme by analysing residue-by-residue geometry and overall structure geometry of the developed 3D-structural model of the proteins (Laskowski *et al.*, 2002). Meanwhile, quality assessment on the developed 3D-structural model of each fragrance-related protein using VERIFY-3D programme was determined by compatibility of their atomic 3D models with their own amino acid sequence (1D), with minimum accepted quality of 80 percent (Bowie *et al.*, 1991; Luthy *et al.*, 1992).

In addition, a structural refinement process were carried out on all the developed 3D-structural model of those fragrance-related proteins by using the Modrefiner tool (<http://zhanglab.ccmb.med.umich.edu/ModRefiner/>) in order to avoid high violations of the restrains in preventing instability of the 3D-structural model of the proteins (Xu and Zhang, 2011). Finally, all the refined 3D-structural model of those fragrance-related proteins were subjected to a viewing process using the UCSF CHIMERA software (Pettersen *et al.*, 2004) to visualise the model in 3D-view as well as to view the structure of their coils, sheets and helixes. Besides that, superimposition between the developed 3D-structural models of those proteins of interest with their respective protein template from PDB database was also visualised using the software. In addition, the UCSF CHIMERA software is also used to visualise catalytic residues that are available on those of those fragrance-related in a close-up view.

Furthermore, the 3D-structural models of all four fragrance-related proteins of *V. Mimi Palmer* were docked with selected ligands and substrates using Autodock 4.2 software (<http://autodock.scripps.edu>) for prediction of their catalytic functionality. All selected ligands and substrates were chosen based on previous substrates and ligands that have been previously reported to be suitable for reference proteins besides available information at the Protein Data Bank (PDB) website (http://www.rcsb.org/ligand_Explorer_viewer.html) as well as BRENDA database (<http://www.brenda-enzymes.org>). Subsequently, prediction on position of active sites that are available in each protein of interest was performed using multiple online servers including CASTp (<http://sts>.

bioe.uic.edu/castp/), Galaxy site (<http://galaxy.seoklab.org/site>) and BSpred (<http://zhanglab.ccmb.med.umich.edu/BSpred/>) for prediction of their binding activity as well as substrate and ligand specificity analysis. Molecular docking analysis using Autodock 4.2 software (<http://autodock.scripps.edu>) was performed to investigate suitability of their potential substrates and ligands binding to bind at catalytic region of protein of interest for their catalytic functionality. The binding interaction between protein of interest with suitable substrates and ligands were visualised graphically using Discovery Studio Visualizer (<http://accelrys.com/products/discovery-studio/>).

4.3.1 Structural and Functional Prediction of VMPKAT Protein

Threading analysis on the deduced amino acid sequence of VMPKAT using multiple online tools including PSI-BLAST (<http://blast.ncbi.nlm.nih.gov/Blastp/PSIBLAST>), Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/>), HHpred (<http://toolkit.tuebingen.mpg.de/hhpred>), pGenThreader (Jones 1999) and LOMETS (<http://zhanglab.ccmb.med.umich.edu/LOMETS/>) has shown that VMPKAT protein hit to 2wu9 protein template of 3-ketoacyl-CoA thiolase 2 of *A. thaliana*, the well-studied plant model, with 75% identity (Table 4.2). Interestingly, in all the threading methods used, the 2wu9 protein template has been identified by almost all the servers to be homologous to VMPKAT protein with identity score of 75. Thus, the 2wu9 protein model of ketoacyl-CoA thiolase 2 of *A. thaliana* was selected as the best protein template to be used as reference in development of 3D-structural model of VMPKAT protein due to the appearance of the template in most of the threading methods with higher identity score compared to other homologous protein templates with much lower identity score. High identity score is an important criterion in choosing the best template for 3D-structural model development of any protein of interest as recommended by Saxena *et al.* (2013). Besides that, consensus secondary structure that aligned between amino acid sequence of VMPKAT protein and the 2wu9 template based on location of α -helices and β -strands as well as the presence of any gap in the amino acid sequences alignment (Figure 4.14) are very important in determination of reliability of the template protein to be used in development of 3D-structural model of each protein of interest as recommended by Drozdetskiy *et al.* (2015).

Table 4.2: List of suitable protein templates is used in 3D-structural model development of VMPKAT protein.

Server name	Template	Protein	Organism	Identity	E-value
PSI-BLAST	2c7y	3-ketoacyl-coa thiolase	<i>Arabidopsis thaliana</i>	76	0
	2wu9	3-ketoacyl-coa thiolase	<i>Arabdiopsis thaliana</i>	75	0
	2wua	3-ketoacyl-Coa acyl transferase	<i>Helianthus annuus</i>	73	0
	2iik	acetyl-coa acyl transferase	<i>Homo sapiens</i>	59	1e ⁻¹⁵³
	1pxt	3-ketoacyl-coa thiolase	<i>Saccharomyces cerevisiae</i>	46	6e ⁻¹⁰⁸
HHPred	2wu9	3-ketoacyl-CoA thiolase	<i>Arabdiopsis thaliana</i>	75	1.2e ⁻⁷¹
	2vu1	Acetyl-CoA acetyltransferase	<i>Zoogloea ramigera</i>	44	3.3e ⁻⁶⁸
	3goa	3-ketoacyl-CoA thiolase	<i>Salmonella enterica</i>	44	4.5e ⁻⁶⁶
	1afw	3-ketoacetyl-CoA thiolase	<i>Saccharomyces cerevisiae</i>	47	4.3e ⁻⁶³
	1ulq	acetyl-CoA acetyltransferase	<i>Thermus thermophilus</i>	47	2.5e ⁻⁶²
Phyre2	2wua	3-ketoacyl-Coa acyl transferase	<i>Helianthus annuus</i>	75	100
	1ulq	acetyl-CoA acetyltransferase	<i>Thermus thermophiles</i>	47	100
	2c7y	Plant enzyme	<i>Arabdiopsis thaliana</i>	76	100
	3ss6	Acetyl-CoA acetyltransferase	<i>Bacillus anthracis</i>	42	100

	2vu2	Thiolase	<i>Zoogloea ramigera</i>	45	100
pGenThreader	4n46	Thiolase	<i>Clostridium acetobutylicum</i> <i>Arabidopsis thaliana</i>	n/a	n/a
	2wu9	3-ketoacetyl-CoA thiolase	<i>Helianthus annuus</i>	n/a	n/a
	2wua	3-ketoacyl-Coa acyl transferase	<i>Homo sapiens</i>	n/a	n/a
	2iik	acetyl-coa acyl transferase	<i>Peptoclostridium difficile</i>	n/a	n/a
	4e11	Acetoacetyl-CoA thiolase		n/a	n/a
LOMETS	2wua	3-ketoacyl-Coa acyl transferase	<i>Helianthus annuus</i>	n/a	n/a
	2c7y	Plant enzyme	<i>Arabidopsis thaliana</i>	n/a	n/a
	4c2j	3-ketoacyl CoA thiolase	<i>Homo sapiens</i>	n/a	n/a
	2wu9	3-ketoacyl CoA thiolase	<i>Arabidopsis thaliana</i>	n/a	n/a

Note: Based on threading analysis using various protein threading programmes, protein template of 2wu9 that appeared in most threading analysis with the highest identity (75%) was chosen as the protein template for 3D-structural model development of VMPKAT.

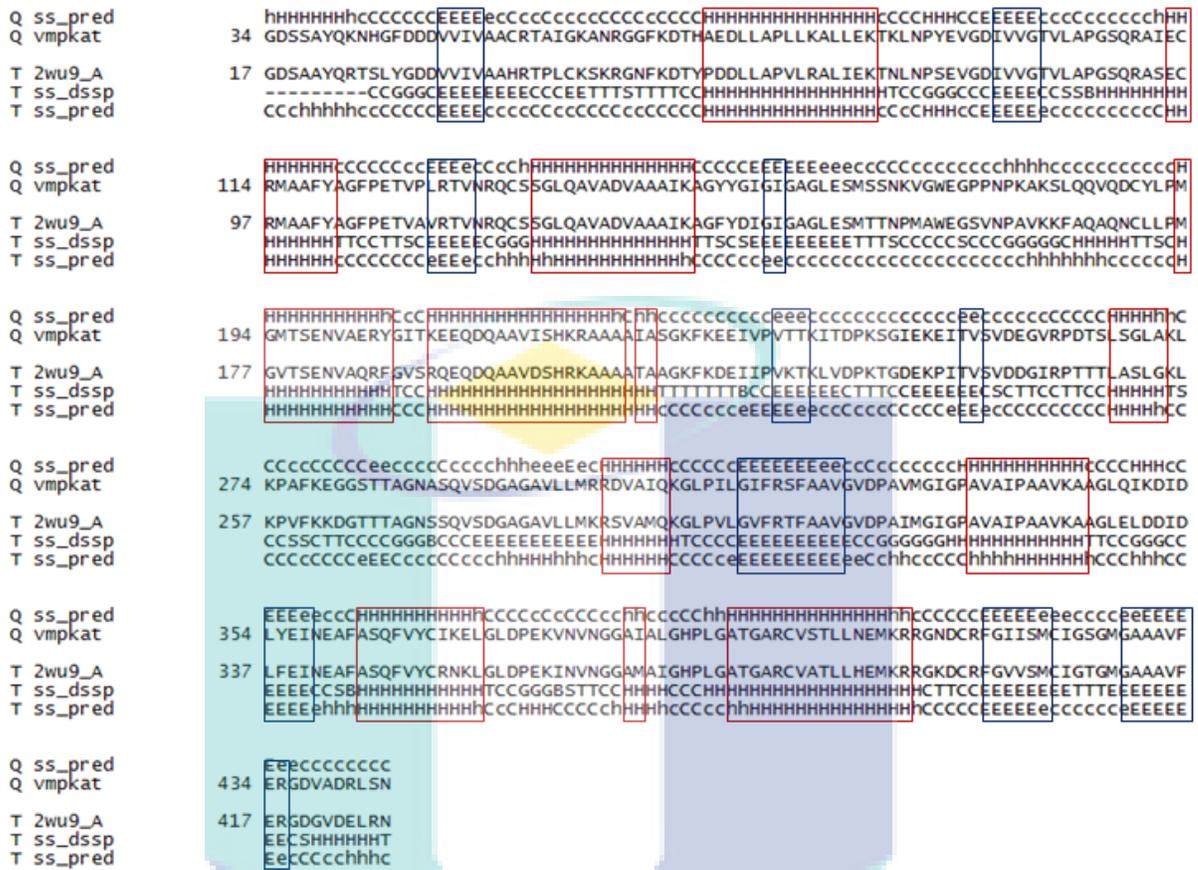


Figure 4.14: Secondary structure alignment between amino acid residue of VMPKAT and 2wu9 template of 3-ketoacyl-CoA thiolase of *Arabidopsis thaliana*. Positions of β -strands and α -helices are highlighted in blue and red boxes, respectively.

In this study, a number of five 3D-structural models of VMPKAT protein were developed using MODELLER version 9.14 based on alignment between amino acid sequences of VMPKAT protein with 3-ketoacyl-CoA thiolase from *A. thaliana* (2wu9) as its reference protein template. The best 3D-structural model of VMPKAT protein was chosen based on the lowest DOPE-score and GA341 value of 1.0 as shown in Table 4.3. Based on DOPE-score and GA341 value for all the developed models, the best model which is vmpkat.B99990001.pdb was selected as final model to be used in this study. Evaluation of the developed 3D-structural model of VMPKAT protein using TM-align Program has shown that the developed 3D-structural model of VMPKAT protein has a Tm Score of 0.887 and RMSD value of 0.79 Å. Thus, the developed 3D-structural model of VMPKAT protein is considered to be significant and reliable due to its Tm-Score of more than 0.4 and RMSD value of lower than 1.0 as recommended by Zhang and Skolnick (2005).

In addition, multiple quality evaluations on the developed 3D-structural model of VMPKAT protein were performed using multiple online protein assessment tools including VERIFY-3D (http://services.mbi.ucla.edu/Verify_3D/), ERRAT (<http://services.mbi.ucla.edu/ERRAT/>), PRO-CHECK (Laskowski *et al.*, 1993) and ProSA-web (Wiederstein and Sippl, 2007). Based on the evaluation results that are shown in Table 4.4, the developed 3D-structural model of VMPKAT protein is considered to be significant and reliable since the model has shown an overall quality of 84.434 based on the evaluation performed using ERRAT program whereby the accepted overall quality is more than 50 percent as recommended by Colovos and Yeates (1993). Besides that, assessment on the developed 3D-structural model of VMPKAT protein using VERIFY-3D programme has shown that the model has at least 80% amino acid that scored ≥ 0.2 . In addition, assessment on the 3D-structural model of VMPKAT protein using PROCHECK online protein assessment server has shown a score of 99.7%, and thus, it is considered to be acceptable since the score is more than 90% as minimum accepted value. Therefore, the developed 3D-structural model of VMPKAT is considered to be acceptable to be used for further structural and functional prediction and analyses. Besides that, the developed 3D-structural model of VMPKAT

Table 4.3: DOPE and GA341 scores of developed 3D-structural models of VMPKAT protein

Filename	Molpdf	DOPE score	GA341 score
vmpkat.B99990001.pdb	2118.94409	-45752.74609	1.00000
vmpkat.B99990002.pdb	2413.57080	-44974.19922	1.00000
vmpkat.B99990003.pdb	2413.90527	-45363.93359	1.00000
vmpkat.B99990004.pdb	3028.49561	-45134.46094	1.00000
vmpkat.B99990005.pdb	2448.99194	-44938.90625	1.00000

Table 4.4: Evaluation on developed 3D-structural model of VMPKAT protein using multiple protein evaluation tools

Evaluation tools	Criteria	Value/ Score/ Percentage
ERRAT	Overall Quality Factor	84.434
VERIFY 3D	At least 80% of amino acid have score ≥ 0.2	Yes
Prosa-Web	Z-score	-9.68
PROCHECK	Percentage	99.7%
RMSD	Less than 1.0	0.79 Å
TM-score	Approximately 1.0	0.887

protein was subjected to a refinement process using Modrefiner online programme (<http://zhanglab.ccmb.med.umich.edu/ModRefiner/>) to minimize their energy level as well as avoiding high violation of spatial restrains in order to stabilise the developed 3D-structural model of VMPKAT protein as recommended by Xu and Zhang (2011).

VMPKAT has been detected to share three homologous domains that are available in 3-ketoacyl-CoA thiolase of *A. thaliana* (2wu9) including Thiolase I, Thiolase II and Thiolase active site (Figure 4.15). In this study, ketoacyl-CoA thiolase of *A. thaliana* (2wu9) was chosen as reference protein template due to several reasons including its 75% identity with VMPKAT as well as the presence of 2wu9 protein template in most of threading analysis results (refer Table 4.2). Besides that, selection of the best protein template to be used as reference protein model of VMPKAT that has been performed using MODELLER software version 9.14 has shown that 2wu9 protein (3-ketoacyl-CoA thiolase of *A. thaliana*) as the best template to be selected due to the lowest resolution of 1.5 Å compared to other protein models including 2wua, 2c7y, 2iik and 1pxt (Table 4.5). The lowest resolution is recommended for protein template selection in *in silico* protein modeling due to the stability of the protein structure (Sousa and Grigorieff, 2007). Unfortunately, in fragrance aspect, none of 3-ketoacyl-CoA thiolase that involves in biosynthesis of fragrance-related compounds and its precursors including benzenoids and phenylpropanoids in scented plant has been reported so far. However, *in vitro* enzymatic reaction on keto-acyl-CoA thiolase of *P. hybrida* (PhKAT) has shown that 3-oxo-3-phenylpropionyl-CoA as suitable substrate for biosynthesis of benzoyl-CoA, a precursor for fragrance compounds including benzoic acid, methylbenzoate, benzyl acetate and phenylethyl acetate (Van Moerkercke *et al.*, 2009). Interestingly, threading analysis of PhKAT using multiple online servers including HHpred and PSI-BLAST has shown that 3-ketoacyl-CoA thiolase of *A. thaliana* (2wu9) that is used as the template protein of VMPKAT appeared in all the threading results with 81 % identity (Appendix I(k)). In addition, BLASTP analysis of VMPKAT using the BLASTP tool at the NCBI GenBank database has shown that VMPKAT shares 73% identity with PhKAT (Accession number: ACV70032.1). Moreover, the same three thiolase domains including thiolase I, thiolase II and thiolase active site have been detected to be present in PhKAT at the same position

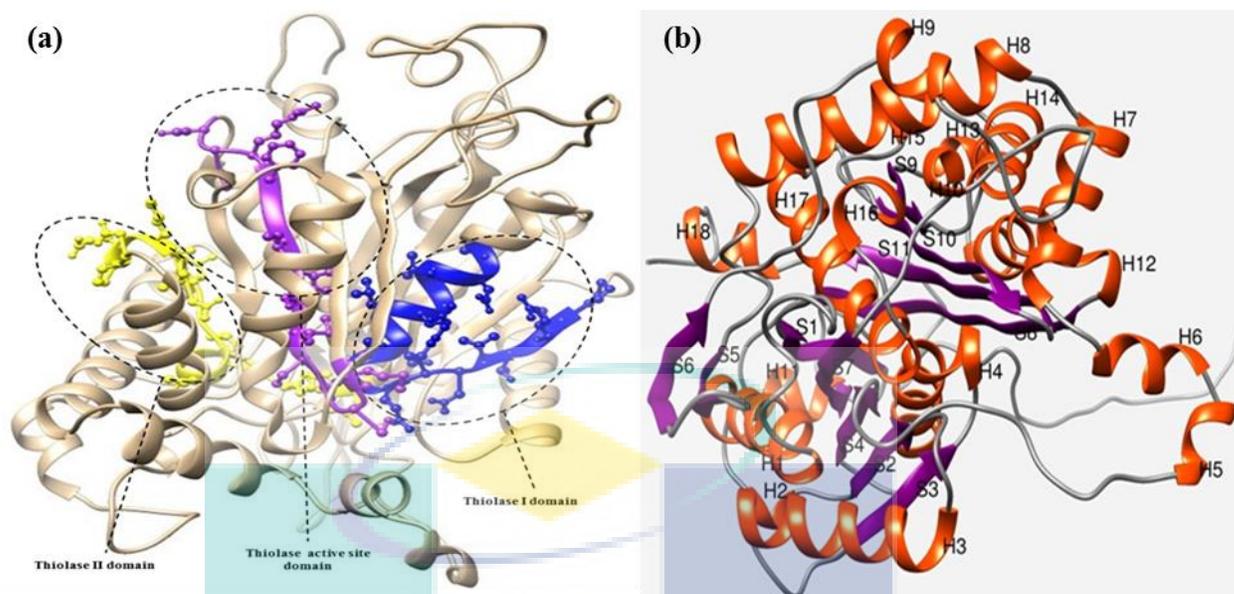


Figure 4.15: 3D-structural model of VMPKAT protein; (a) Position of Thiolase I, Thiolase II and Thiolase active site domains (b) Position of labeled β strands (S1–S11) and α helices (H1–H18).

Table 4.5: Comparison of suitable model for development of 3D-structural model of VMPKAT protein using MODELLER version 9.14

Template	Protein	Organism	Resolution
2wu9	3-ketoacyl-coA thiolase	<i>Arabidopsis thaliana</i>	1.5 Å
2wua	3-ketoacyl-coa thiolase	<i>Helianthus annuus</i>	1.8 Å
2c7y	3-ketoacyl-coa thiolase	<i>Arabidopsis thaliana</i>	2.1 Å
1pxt	3-ketoacyl-coa thiolase	<i>Saccharomyces cerevisiae</i>	2.8 Å
2iik	acetyl-coaacyl transferase	<i>Homo sapiens</i>	2.5 Å

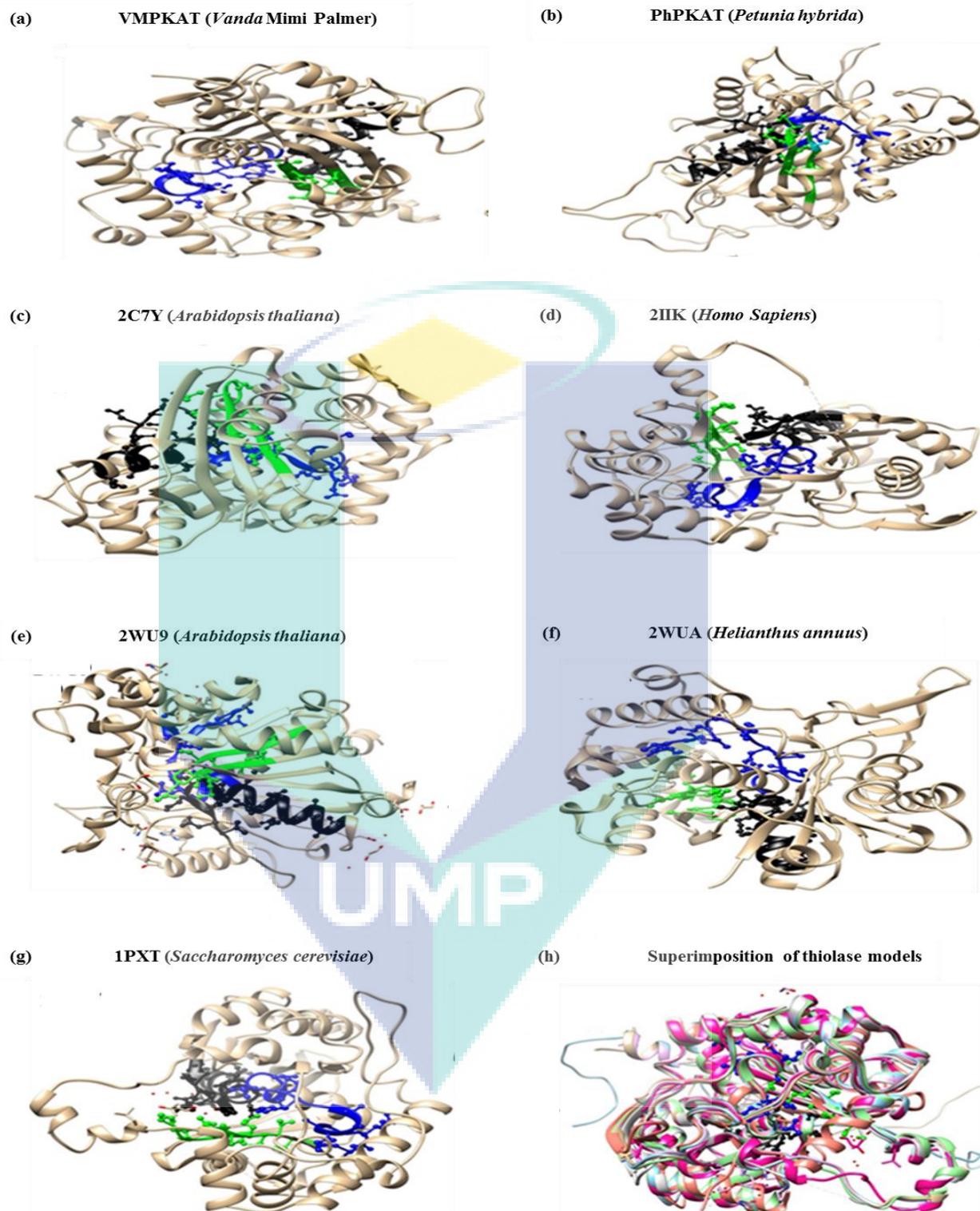


Figure 4.16: Comparative modeling of thiolase models. Thiolase I, Thiolase II and Thiolase active site in (a-g) were shown in blue, green and black color, respectively. Meanwhile, superimposition of thiolase models is shown in (h).

with VMPKAT as shown in Figure 4.5 in section 4.2.1 and Figure 4.16 as well as three corresponding active sites of VMPKAT as shown in Figure 4.16. Thus, in this study, PhKAT is selected for a comparison study to VMPKAT due to their homologous structural and functional characteristics. Interestingly, superimposition between VMPKAT and PhKAT also showed high structural similarity as shown in Figure 4.17, together with other thiolase protein models. The high structural similarity in superimposition is reflecting that superimposed structures of VMPKAT protein with other thiolase protein models have slightly similar structure due to the fact that all the protein share the same homologous thiolase domain even though the location of starting amino acid residues in thiolase domains are slightly different (Figure 4.18). Thus, VMPKAT protein has a reliable structural protein model based on the structure of chain A in 3-ketoacyl-CoA thiolase of *A. thaliana* (2wu9).

In this study, prediction on functional catalytic reaction of VMPKAT was determined by *in silico* docking approach using Autodock 4.2 software (<http://autodock.scripps.edu>). In docking analysis on 3D-structural model of VMPKAT protein, 3-oxo-3-phenylpropionyl-CoA was selected as substrate candidate due to the reported *in vitro* enzymatic work on recombinant 3-ketoacyl-CoA thiolase of *P. hybrida* (PhKAT) that catalyses the biosynthesis of benzoyl-CoA from 3-oxo-3-phenylpropionyl-CoA as its substrate (Van Moerkercke *et al.*, 2009; Figure 4.19). Besides that, another two suitable substrates were also selected for docking analysis of 3D-structural model VMPKAT protein including acetoacetyl-CoA (ACO) and coenzyme A (COA) that have been previously reported to be successfully utilised by 3-ketoacyl-CoA thiolase of *A. thaliana* (2wu9) (Pye *et al.*, 2010). Unfortunately, molecular docking analysis on 3D-structural model of VMPKAT cannot be performed using 3-oxo-3-phenylpropionyl-CoA as substrate due to the instability of 3-oxo-3-phenylpropionyl-CoA 3D-conformer whereby acyl-CoA chain of this substrate is longer compared to other CoA-derivatives that are usually used as substrate in *in vitro* enzymatic test of thiolase enzymes (Appendix I(i)). Theoretically, long ester bonds of CoA-derivatives will disrupt molecular docking of thiolase protein resulting with torsional bonds error of substrates besides yielding very high free binding energy as well as dissociation constant leading to poor molecular docking

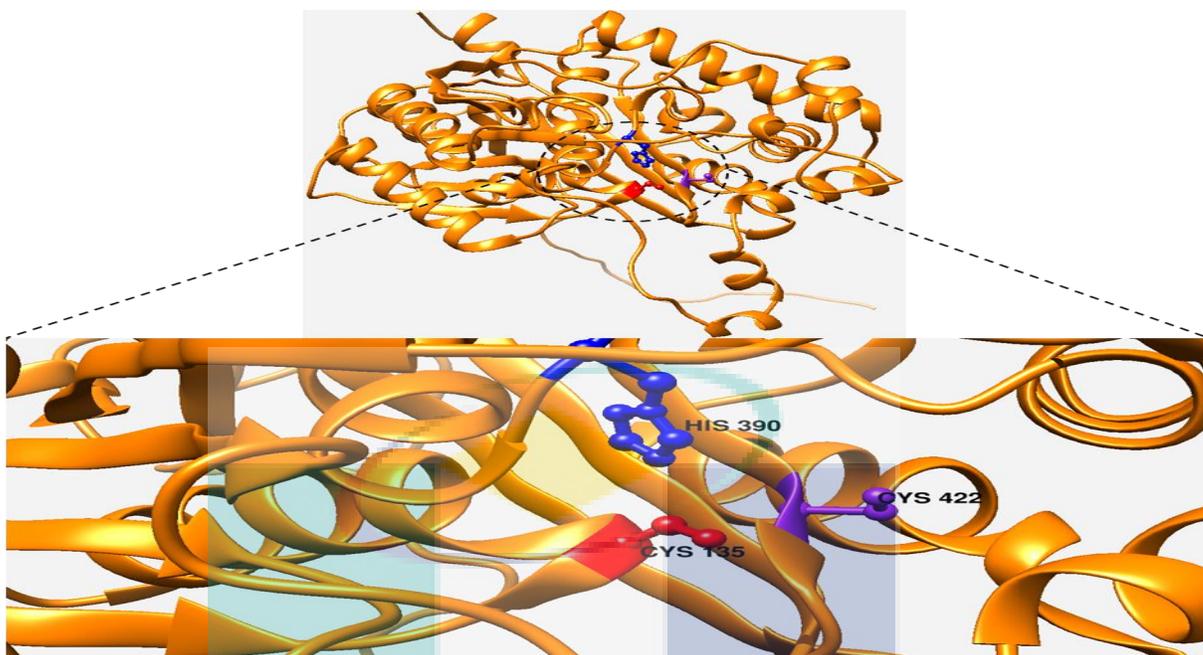


Figure 4.17: Predicted binding position (Cys 135, His 390, Cys 422) in thiolase active site domain of VMPKAT protein.

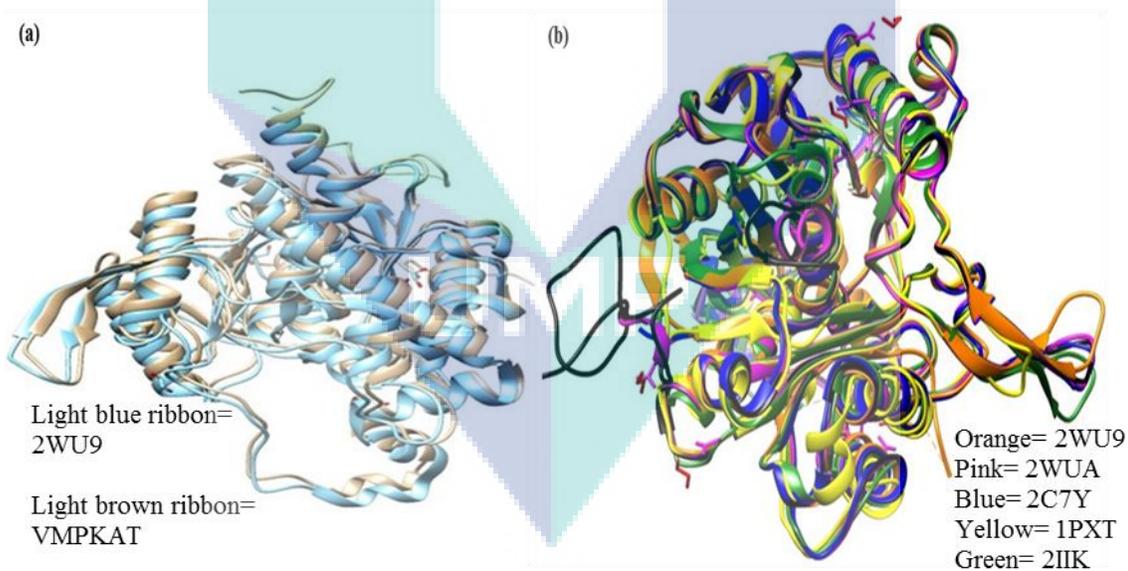


Figure 4.18: Superimposition of 3D-structural model of VMPKAT protein with (a) 2wu9 protein template of *Arabidopsis thaliana* and its RMSD value of 1.23 Å (b) five thiolase protein models, including 2c7y, 2wu9, 2wua, 2iik and 1pxt, representing RMSD value of superimposition is 0.733 Å.

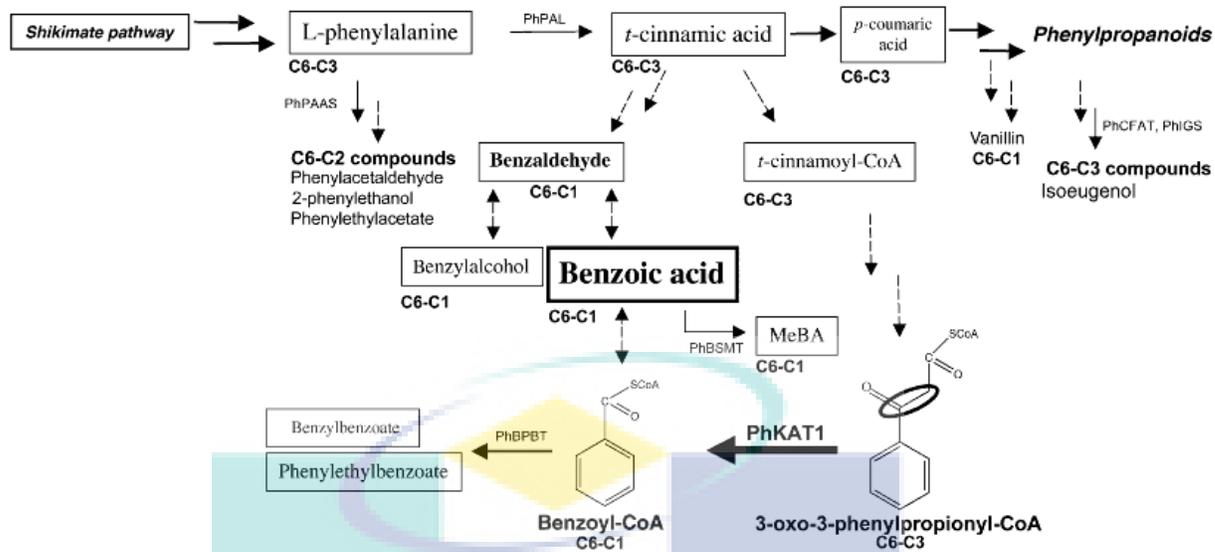


Figure 4.19: Involvement of 3-ketoacyl CoA thiolase (*PhKAT*) in benzenoid biosynthetic pathway of *Petunia hybrida*. This schematic figure was adopted from Van Moerkercke *et al.*, (2009).

UMP

between protein of interest and substrate (Pye *et al.*, 2010). Moreover, none of *in silico* docking analysis work has been reported on 3D-structural model of 3-ketoacyl-CoA thiolase from any organisms that utilises 3-oxo-3-phenylpropionyl-CoA as its substrate. Therefore, only two substrates were used in this *in silico* docking study of 3D-structural model of VMPKAT protein which is acetoacetyl-CoA (ACO) and coenzyme A (COA). Besides the two substrates, *in silico* docking analysis of 3D-structural model of VMPKAT was also performed using a specific ligand which is 1,2-ethanediol whereby the ligand is the only ligand that is listed as suitable ligand for 3-ketoacyl-CoA thiolase of *A. thaliana* (2wu9) in ligand explorer database that available at the Protein Data Bank (PDB) (<http://www.rcsb.org/pdb>).

Based on docking analysis of 3D-structural model of VMPKAT protein, both acetoacetyl-CoA and coenzyme A have shown suitable values of binding energy with -2.21 kcal/mol and -3.79 kcal/mol, respectively (Table 4.6; Figure 4.20; Figure 4.21). However, based on the result, coenzyme A might be more suitable to be used as the substrate to VMPKAT protein due to its lower binding energy compared to the other substrates due to its higher specificity towards the VMPKAT protein. Meanwhile, ligand specificity test by the docking analysis using Autodock 4.2 software has predicted that 1,2-ethanediol (EDO) is suitable to be used as ligand for VMPKAT protein. This result is reasonable since 2-ethanediol (EDO) has been reported to be used in enzymatic test of 3-ketoacyl CoA thiolase of *A. thaliana* (2wu9) (Pye *et al.*, 2010) that has been selected in this study as reference protein template for 3D-structural development of VMPKAT protein due to its 75% homologous identity to VMPKAT. In addition, molecular docking analysis of 3-ketoacyl-CoA thiolase of *P. hybrida* (PhKAT), a well-studied scented flower has shown a similar result whereby Coenzyme A and 1,2-ethanediol (EDO) are also suitable to be used as substrate and ligand, respectively (Tables 10 and 11). However, validation on suitability of the predicted substrate and ligand could be tested in future by *in vitro* enzymatic test to confirm the specific functional catalytic role of VMPKAT.

Table 4.6: Docking analysis of various putative substrates with fragrance-related proteins of *Vanda Mimi Palmer* using Autodock 4.2 software

Name of protein	Putative Substrate	Binding energy (kcal/mol)	Dissociation constant (mM)
VMPKAT	Coenzyme A	-3.79	3.20
	Aceto-acetyl-CoA	-2.21	1.16
VMPBSMT	Benzoic acid	-2.76	2.45
	Salicylic acid	-1.89	1.43
VMPEGS	Coniferyl acetate	-6.78	0.042
	Coniferyl alcohol	-5.62	0.075
VMPOOMT	Orcinol	-5.54	0.986
	Resorcinol	-2.41	2.44
	Isoformononetin (HMO)	-1.68	1.24

Table 4.7: Molecular docking analysis of various putative ligands with fragrance-related proteins of *Vanda Mimi Palmer* using Autodock 4.2 software.

Name of protein	Putative ligand	Binding energy (kcal/mol)	Dissociation constant (mM)
VMPKAT PhKAT	1,2-ethanediol (EDO)	-3.42	1.25
VMPBSMT	S-adenosyl-L-methionine (SAM)	-4.93	0.243
	S-adenosyl-L-homocysteine (SAH)	-7.30	0.0044
VMPEGS	Nicotinamide-adenine phosphate (NAP)	-5.45	2.11
VMPOOMT	S-adenosyl-L-methionine (SAM)	-2.81	2.276
	S-adenosyl-L-homocysteine (SAH)	-4.74	1.145

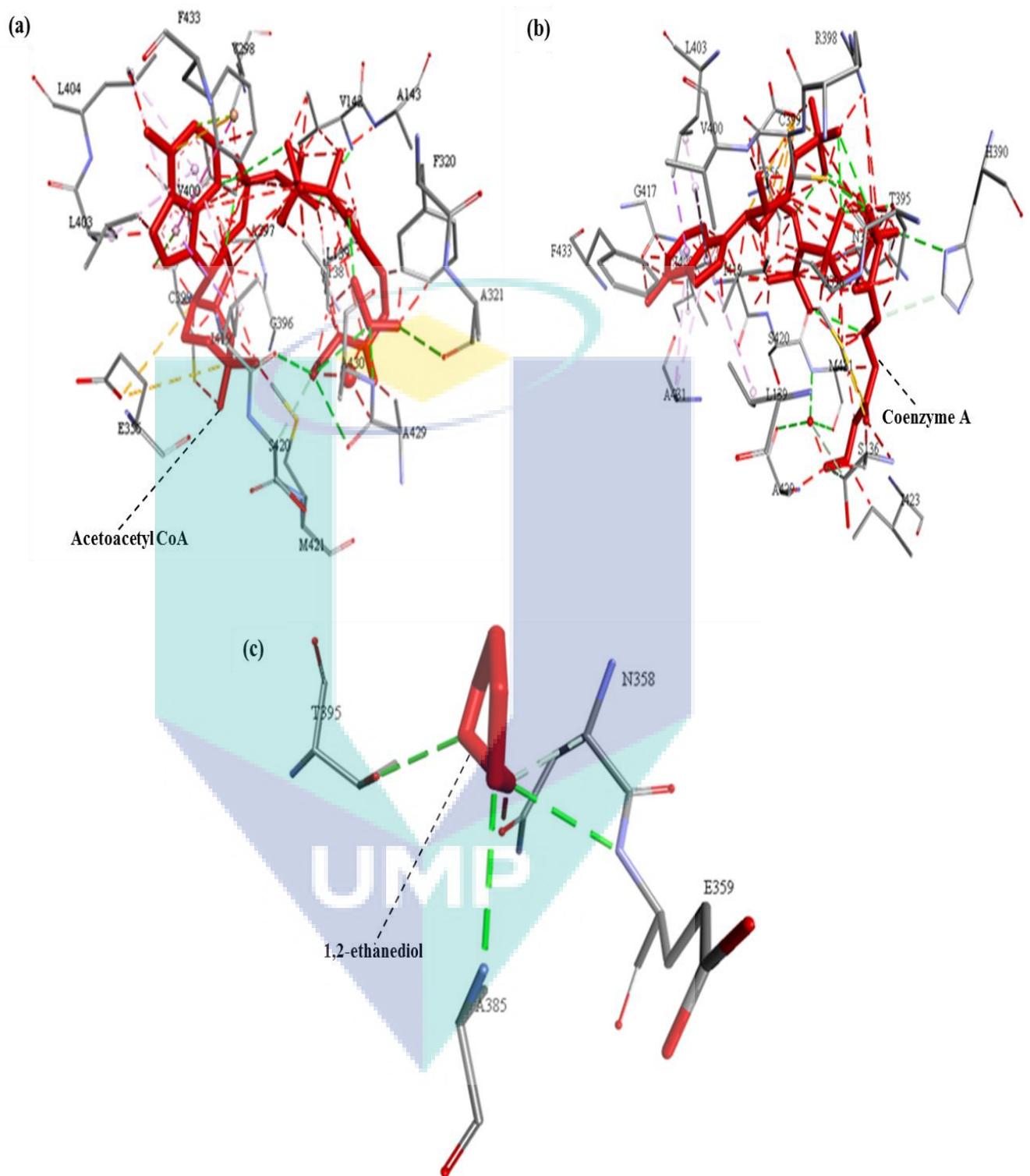


Figure 4.20: Molecular docking of substrates and ligand with VMPKAT protein using Autodock 4.2; (a) VMPKAT-ACO complex, (b) VMPKAT-COA complex, (c) VMPKAT-EDO complex

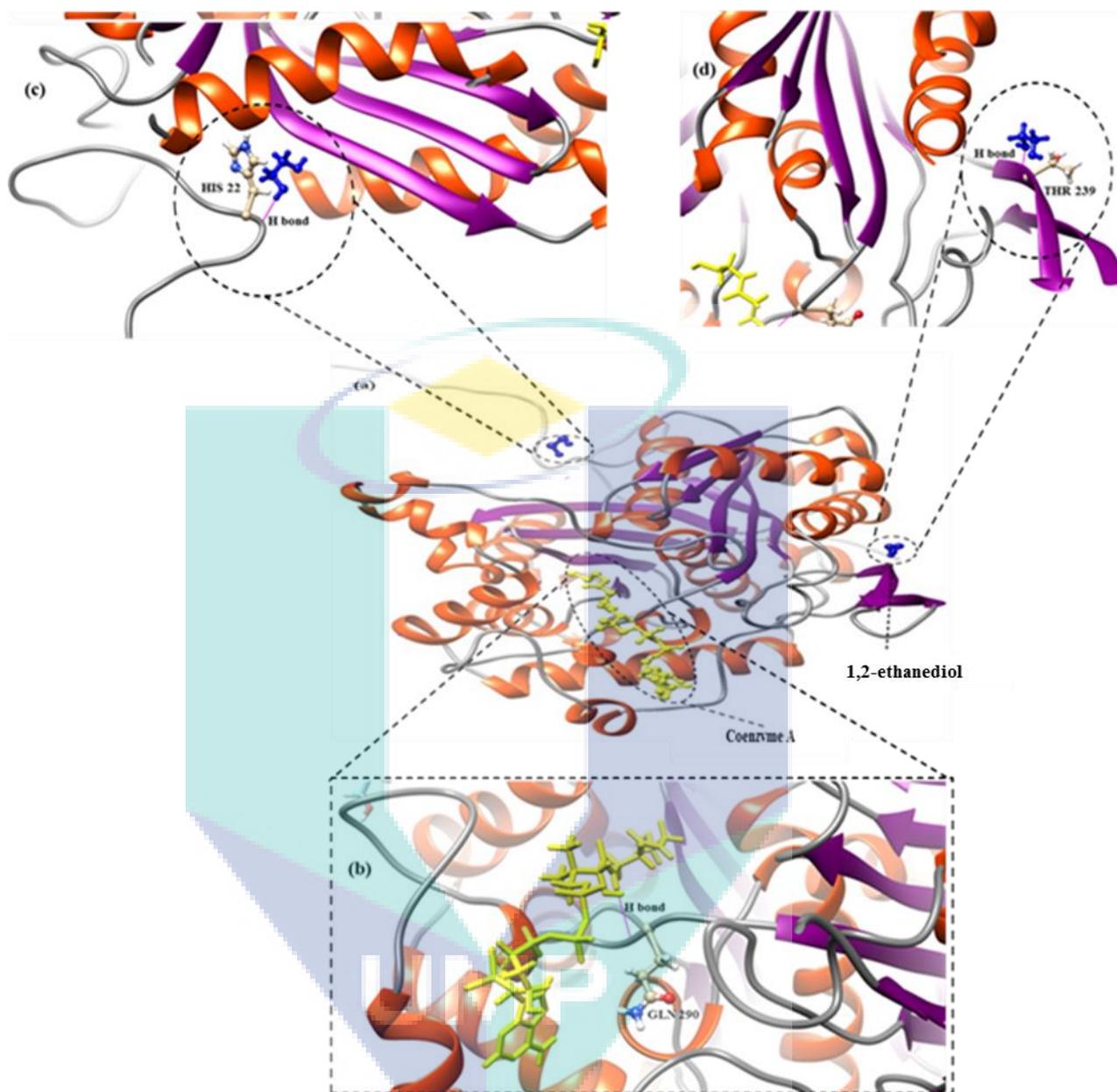


Figure 4.21: Molecular docking of VMPKAT model with its substrate and ligand, using Swissdock and viewed with CHIMERA; (a) general picture; (b) 3-oxo-3- phenylpropionyl-CoA; (c) and (d) 1,2-ethanediol.

4.3.2 Structural and Functional Prediction of VMPBSMT Protein

Based on threading analysis of VMPBSMT protein using multiple online protein threading servers, several homologous protein templates that hit with VMPBSMT protein are listed in Table 4.8. Among the templates, 1m6e protein template of salicylic acid carboxyl methyltransferase from *C. breweri* has been identified to be homologous to VMPBSMT protein, with 41 to 42% identity. Thus, 1m6e protein of *C. breweri* was chosen as protein template to be used as reference in the development of 3D-structural model of VMPBSMT protein due to the appearance of the template in the results of all threading methods. Besides that, consensus secondary structure of α -helices and β -strands that aligned between amino acid sequence of VMPBSMT and 1m6e protein *C. breweri* is shown in Figure 4.22. In addition, β -strands and α -helices positions are visualised in 3D-structure as shown in Figure 4.23. Based on secondary structure alignment, both VMPBSMT and 1m6e might share similar structural as well as functional role in biosynthesis of either methyl benzoate or methyl salicylate, depending on substrate specificity towards its catalytic regions (Zubieta *et al.*, 2003). In V. Mimi Palmer, VMPBSMT is predicted to utilise benzoic acid instead of salicylic acid since methylbenzoate, the subsequent product to be catalysed by the enzyme was detected in the scent of V. Mimi Palmer (Mohd-Hairul *et al.*, 2010).

Five 3D-structural models of VMPBSMT protein have been constructed using the MODELLER software, version 9.14 based on alignment between amino acid sequences of VMPBSMT with salicylic acid carboxyl methyltransferase of from *C. breweri* (1m6e). In this study, salicylic acid carboxyl methyltransferase (SAMT) was selected as reference protein template instead of benzoic acid carboxyl methyltransferase (BSMT) due the no crystallised structure of BSMT that has been deposited at the Protein Data Bank (PDB). Among the five constructed models, vmpbsmt.B99990005.pdb was selected as the best model to be used in this study based on the lowest DOPE-score and GA341 value (Table 4.9). The selected model was further evaluated for its Tm-score and RMSD value using TM-align Program (<http://zhanglab.ccmb.med.umich.edu/TM-score/>). Based on the

Table 4.8: List of suitable protein templates to be used in 3D-structural model development of VMPBSMT protein

Server name	Template	Protein	Organism	Identity	E-value
PSI-	1m6e	Salicylic acid carboxyl methyltransferase	<i>Clarkia breweri</i>	41	2e-81
BLAST	2eg5	Xanthosine methyltransferase	<i>Coffea canephora</i>	38	5e-75
	2efj	1,7-dimethylxanthine methyltransferase	<i>Coffea canephora</i>	37	6e-73
	3b5i	Indole-3-acetic acid methyltransferase	<i>Arabidopsis thaliana</i>	33	4e-59
HHPred	3b5i	salicylic acid carboxyl methyltransferase	<i>Arabidopsis thaliana</i>	32	1.9e-72
			<i>Coffea canephora</i>		
	2efj	3,7-dimethylxanthine methyltransferase	<i>Clarkia breweri</i>	39	7e-71
	1m6e	salicylic acid carboxyl methyltransferase	<i>Anabaena variabilis</i>	42	3.4e-64
	3ccf	Cyclopropane-fatty-acyl-phospholipid synthase	<i>Agrobacterium</i>	16	3.1e-25
		Trans-aconitate 2-methyltransferase	<i>fabrum</i>		
2p35			13	1.3e-22	

Phyre2	1m6e	Salicylic acid carboxyl methyltransferase	<i>Clarkia breweri</i>	42	100
	2eg5	Xanthosine methyltransferase	<i>Coffea canephora</i>	39	100
	3b5i	salicylic acid carboxyl methyltransferase	<i>Arabidopsis thaliana</i>	31	100
			<i>Galdieria</i>		
	2o57	Sarcosine dimethylglycine methyl transferase	<i>sulphuraria</i>	14	100
		Phosphoethanolamine	<i>Haemonchus</i>		
	4krh	N-methyltransferase	<i>contortus</i>	18	100
pGen	1m6e	Salicylic acid carboxyl methyltransferase	<i>Clarkia breweri</i>	n/a	n/a
Threader	2efj	1,7-dimethylxanthine methyltransferase	<i>Coffea canephora</i>	n/a	n/a
	2eg5	Xanthosine methyltransferase	<i>Coffea canephora</i>	n/a	n/a
	3b5i	salicylic acid carboxyl methyltransferase	<i>Arabidopsis thaliana</i>	n/a	n/a
			<i>Anabaena variabilis</i>		
	3ccf	Cyclopropane-fatty-acyl-phospholipid synthase		n/a	n/a

Note: The 1m6e template appeared in the all of the results given by the different methods and was chosen as the model template ofVMPBSMT.

Table 4.9: DOPE and GA341 scores of constructed 3D-structural models of VMPBSMT protein

Filename	Molpdf	DOPE score	GA341 score
vmpbsmt.B99990001.pdb	2185.14941	-43568.03516	1.00000
vmpbsmt.B99990002.pdb	2430.09766	-44028.48047	1.00000
vmpbsmt.B99990003.pdb	2212.34326	-44148.22656	1.00000
vmpbsmt.B99990004.pdb	2509.18848	-43342.49219	1.00000
vmpbsmt.B99990005.pdb	2300.20508	-44376.35156	1.00000

Table 4.10: Evaluation on developed 3D-structural model of VMPBSMT protein using protein evaluation tools

Evaluation tools	Criteria	Value/ Score/ Percentage
ERRAT	Overall Quality Factor	53.951
VERIFY 3D	At least 80% of amino acid have score ≥ 0.2	Yes
Prosa-Web	Z-score	-8.17
PROCHECK	Percentage	99.8%
RMSD	Less than 1.0	0.79 Å
TM-score	Approximately 1.0	0.9

evaluation, the selected model has shown an acceptable protein model quality with Tm Score of 0.9 and RMSD value of 0.79 Å as recommended by Zhang and Skolnick (2005) with minimum Tm-score of 0.4 and RMSD value of lower than 1.0. Further quality assessment on the developed 3D-structural model of VMPBSMT protein by using multiple online protein assessment servers has shown the significant of the constructed model as summarised in Table 4.10. Based on the evaluation result, the model is considered to be significant and reliable due to its overall quality of 53.951% that is higher than the minimum acceptable score of 50% as recommended by Colovos and Yeates (1993) for model evaluation using ERRAT Program. Besides that, assessment using VERIFY-3D and PROCHECK program (Table 4.10) has also shown acceptable score for the developed 3D-structural model of VMPBSMT protein. Thus, the developed 3D-structural model of VMPBSMT protein is considered to be significant and reliable to be used for further structural and functional prediction and analysis.

Comparison between the developed 3D-structural model of VMPBSMT protein with homologous *S*-adenosyl-*L*-methionine methyltransferase-family protein from other plants including *C. breweri*, *Coffea canephora* and *A. thaliana* has shown that the structural features of the VMPBSMT is almost similar to the salicylic acid carboxyl methyltransferase of *C. breweri* (1m6e) based on the location of *S*-adenosyl-*L*-methyltransferase domain (Figure 4.24). This result is as expected since secondary structural alignment between VMPBSMT protein with the 1m6e protein template (refer Figure 4.22) has shown well-alignment between consensus α -helices as well as β -strands. Meanwhile, superimposition of VMPBSMT with other homologous proteins including *S*-adenosyl-*L*-methionine methyltransferase-family protein from other plants including *Coffea canephora* (2efj) and *A. thaliana* (3b5i) has shown that both proteins are not well-aligned with the developed 3D-structural protein of VMPBSMT (Figure 4.25). This results is as expected since threading analysis using multiple protein threading servers (Table 4.8) has shown that the homologous protein of *Coffea canephora* (2efj) and *A. thaliana* (3b5i) share less than 40% identity with VMPBSMT of *V. Mimi Palmer*. In addition, the far genetic-relatedness between Vandaceous orchid with the *Coffea canephora* and *A. thaliana* might contribute to the lower identity.

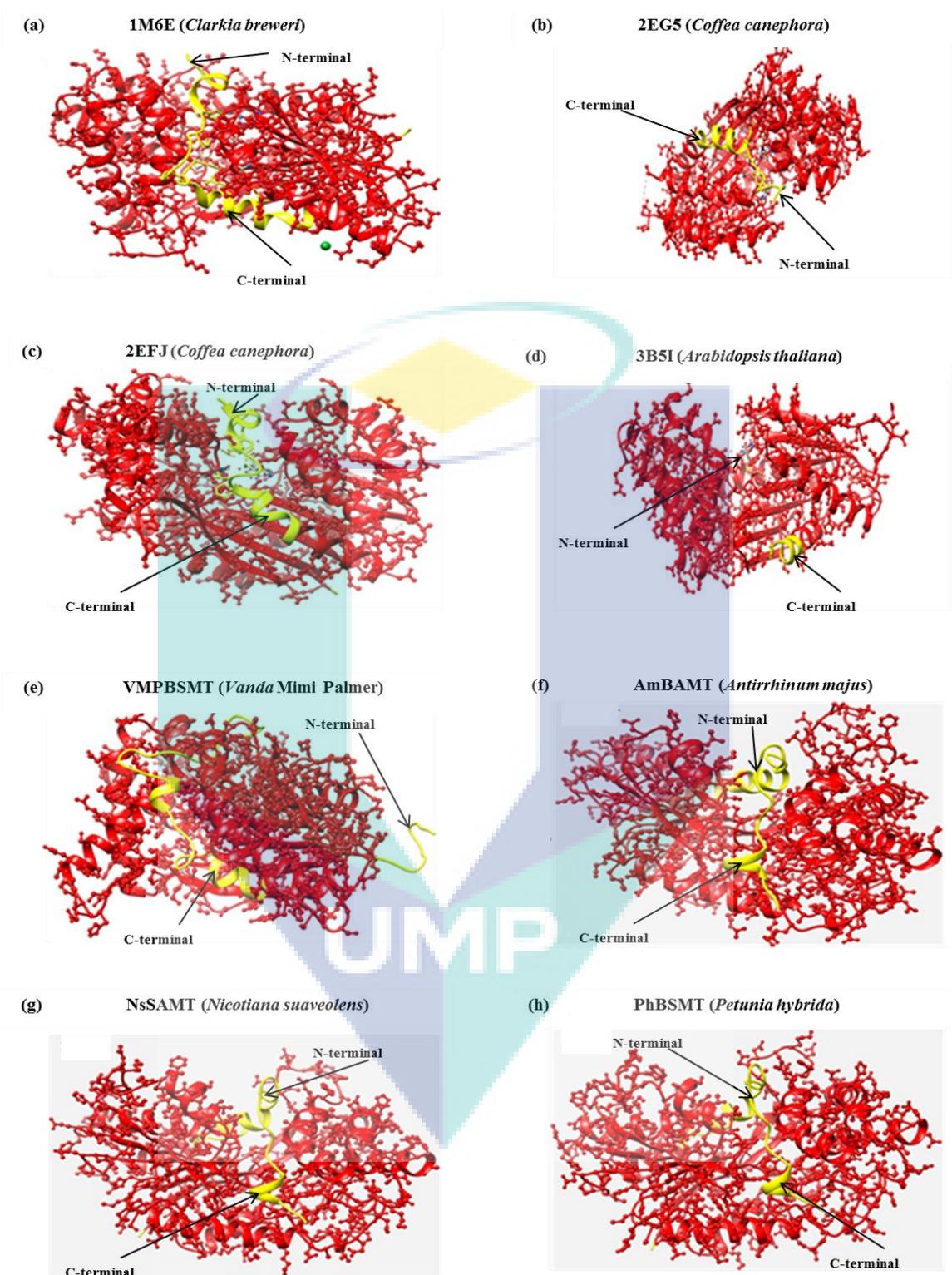


Figure 4.24: Comparative modeling of selected reference protein models to VMPBSMT. SAM-dependent methyltransferase domain was colored in red meanwhile yellow structures are the terminal N and C domain.

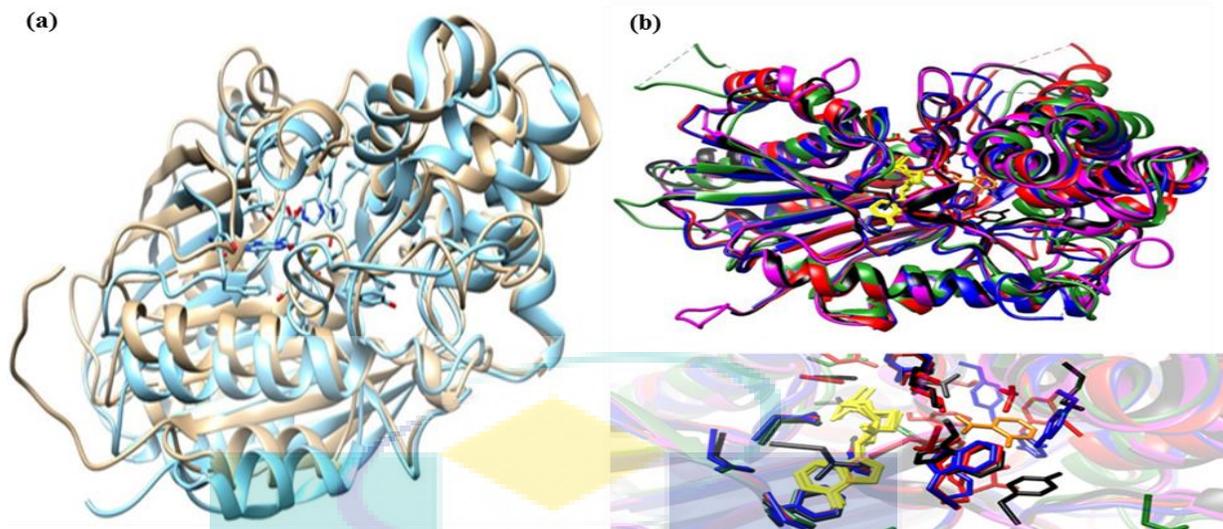


Figure 4.25: (a) Superimposition of VMPBSMT with chain X of 1m6e. Light brown chain indicates VMPBSMT meanwhile light blue chain indicates chain X of 1m6e. (b) Superimposition of VMPBSMT with other protein models, retrieved from PDB database (1m6e, 2eg5, 2efj and 3b5i). Yellow ligand= SAH, orange ligand= SAL. For 1m6e, chain X was used to superpose with VMPBSMT model, others (2eg5, 2efj and 3b5i) used chain A for superimposition in open conformation.

In this study, another three well-studied fragrance-related proteins including SAMT of *A. majus*, BAMT of *Nicotiana suaveolens* and BSMT of *P. hybrida* that have been reported to catalyse biosynthesis of methylbenzoate from benzoic acid (Negre *et al.* 2003; Pott *et al.* 2004; Boatright *et al.* 2004) were selected for comparative modeling (refer Figure 4.24). Unfortunately, none of the crystallised structure of those proteins is available in the Protein Data Bank. Thus, all of those proteins were subjected to threading analysis using multiple online servers and followed by development of 3D-structural protein models. Interestingly, based on the threading analysis, 1m6e protein template of *C. breweri* has been identified to be suitable to be used as reference protein template for development of 3D-structural protein models due to the presence of the template in all threading results with more than 40% identity (Table 4.8). The suitability of 1m6e template of *C. breweri* to be used in the development of 3D-structural protein of VMPBSMT as well as well studied fragrance-related proteins including SAMT of *A. majus*, BAMT of *Nicotiana suaveolens* and BSMT of *P. hybrida* might reflect their similar role of VMPBSMT and all of those proteins in utilising benzoic acid as the main substrate for biosynthesis of methylbenzoate as one of the constituents of their fragrance. However, in previous *in vitro* enzymatic studies of SAMT of *A. majus*, both salicylic acid and benzoic acid have been reported to be suitable as its substrate but salicylic acid is more preferable due to its catalytic specificity. Meanwhile in *Nicotiana suaveolens*, benzoic and cinnamic acids have been reported to be accepted as suitable substrates (Pott *et al.*, 2004). Thus in *V. Mimi Palmer*, the VMPBSMT protein might be involved in catalysing the biosynthesis of methylbenzoate compound from benzoic acid as its precursor since the methyl benzoate compound has been reported to be emitted by the fully- open flower during day time as reported by Mohd-Hairul *et al.* (2010a).

Functional domain analysis of VMPBSMT using multiple online tools including ScanProsite, PSI-BLAST and Localizome has shown the presence of S-adenosyl-L-methionine methyltransferase domain in the amino acid sequence of the VMPBSMT protein (Figure 4.26). Interestingly, similar functional domain has been identified in other homologous protein including SAMT from *C. breweri*, SAMT from *A. majus* as well as BSMT from *P. hybrida* (Zubieta *et al.*, 2003). In addition, S-adenosyl-L-methionine

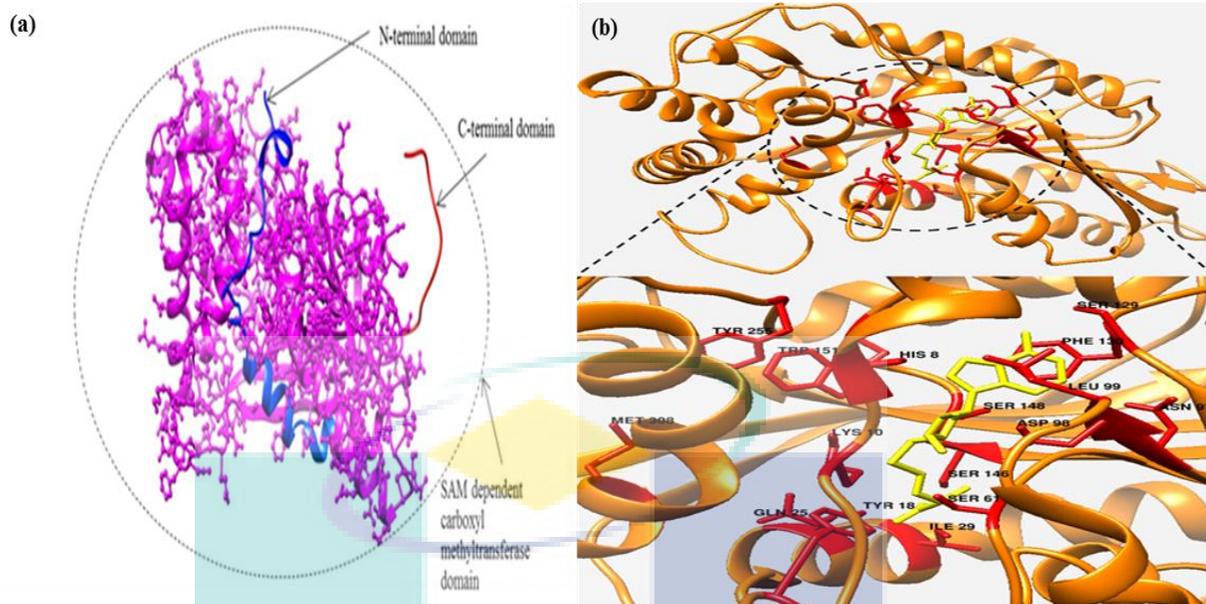


Figure 4.26: SAM binding sites of VMPBSMT; (a) VMPBSMT model show the SAM domain binding site (magenta color). (b) SAM binding sites of salicylic acid carboxyl methyltransferase of *Clarkia breweri*, chain X (1m6e_X) (reference protein model to VMPBSMT)

UMP

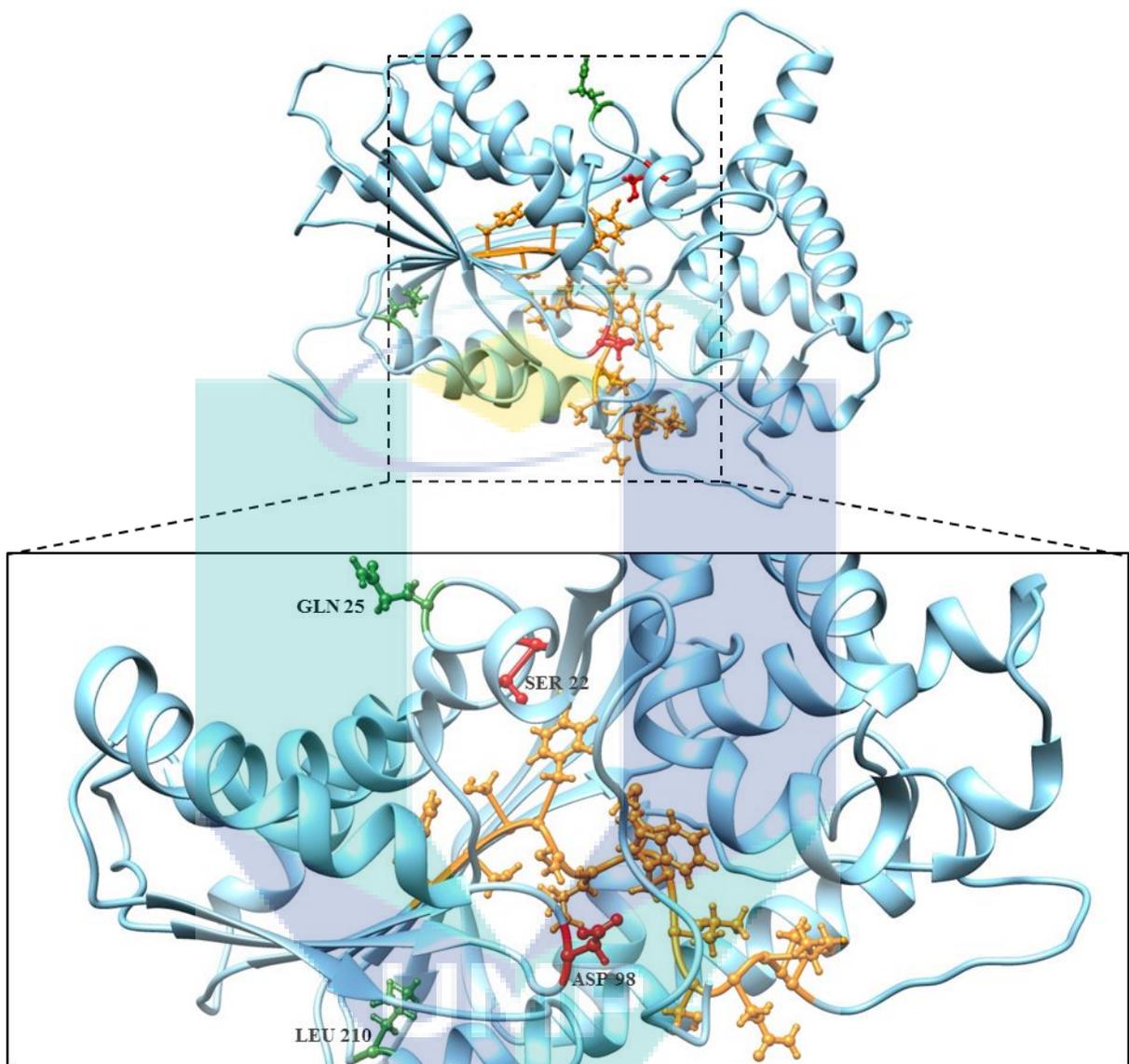


Figure 4.27: Specific amino acid residues involved in three different domains in VMPBSMT. The colored parts are domains for VMPBSMT. Orange: SAM dependent carboxyl methyltransferase attachment site; green: salicylate binding residues; red: SAM/SAH binding residues.

methyltransferase domain has been found in all members of methyltransferase family proteins whereby the catalytic reaction has been reported to be occurred in the presence of S-adenosyl-L-methionine (SAM) as important cofactor for catalytic reaction of the enzymes (Negre *et al.*, 2002; Figure 4.27). Even though a lot of *in vitro* enzymatic assays have proven the functional role of SAM-dependent proteins including BSMT from *P. hybrida*, SAMT from *C. breweri* and SAMT from *A. majus* in methylation of salicylic acid and benzoic acid to respective methyl salicylate and methyl benzoate, S-Adenosyl-L-Homocysteine or well known as SAH has been reported to be suitable ligand for methyltransferase enzymes including SAMT of *C. breweri* (1m6e), SAMT of *A. thaliana* (3b5i), 3,7-dimethylxanthine methyltransferase of *Coffea canephora* (2efj) and xanthosine methyltransferase of *Coffea canephora* (2eg5) (Zubieta *et al.*, 2003; McCarthy and McCarthy, 2007) as shown in Protein Data Bank. In the catalytic reaction, SAM has been reported as initial cofactor for methyltransferase enzymes and subsequently converted to SAH by alternating of methionine to homocysteine (Effmert *et al.*, 2005). The preference to SAH as ligand might be due to the minor differences of homocysteine amino acid as functional group in SAH instead of methionine in SAM (Effmert *et al.*, 2005).

Interestingly, molecular docking study on VMPBSMT using Autodock software, version 4.2 has shown that both benzoic acids as well as salicylic acid are suitable substrates for its catalytic reaction based on their negative value of free binding energy (Table 4.6; Figure 4.28; Figure 4.29). Even though the development of 3D-structural model of VMPBSMT protein was based on the 3D-protein model of 1m6e (salicylic acid carboxyl methyltransferase of *C. breweri*) that is involved in catalysing biosynthesis of methyl salicylate from salicylic acid, there is a possibility of VMPBSMT either as monofunctional enzyme to utilise benzoic acid as the only substrate or bifunctional enzyme that utilises both benzoic acid as well as salicylic acid as preferable substrate for biosynthesis of methylbenzoate and methylsalicylate, respectively, depending on the availability of the substrates. This is due to the fact that its homologous protein PhBSMT from *P. hybrida* has been reported previously as a bifunctional enzyme that utilise both benzoic acid as well as salicylic acid as suitable substrates in its methylation reaction (Verdonk *et al.* 2003). However, 1m6e has been selected in this study as reference protein template for the

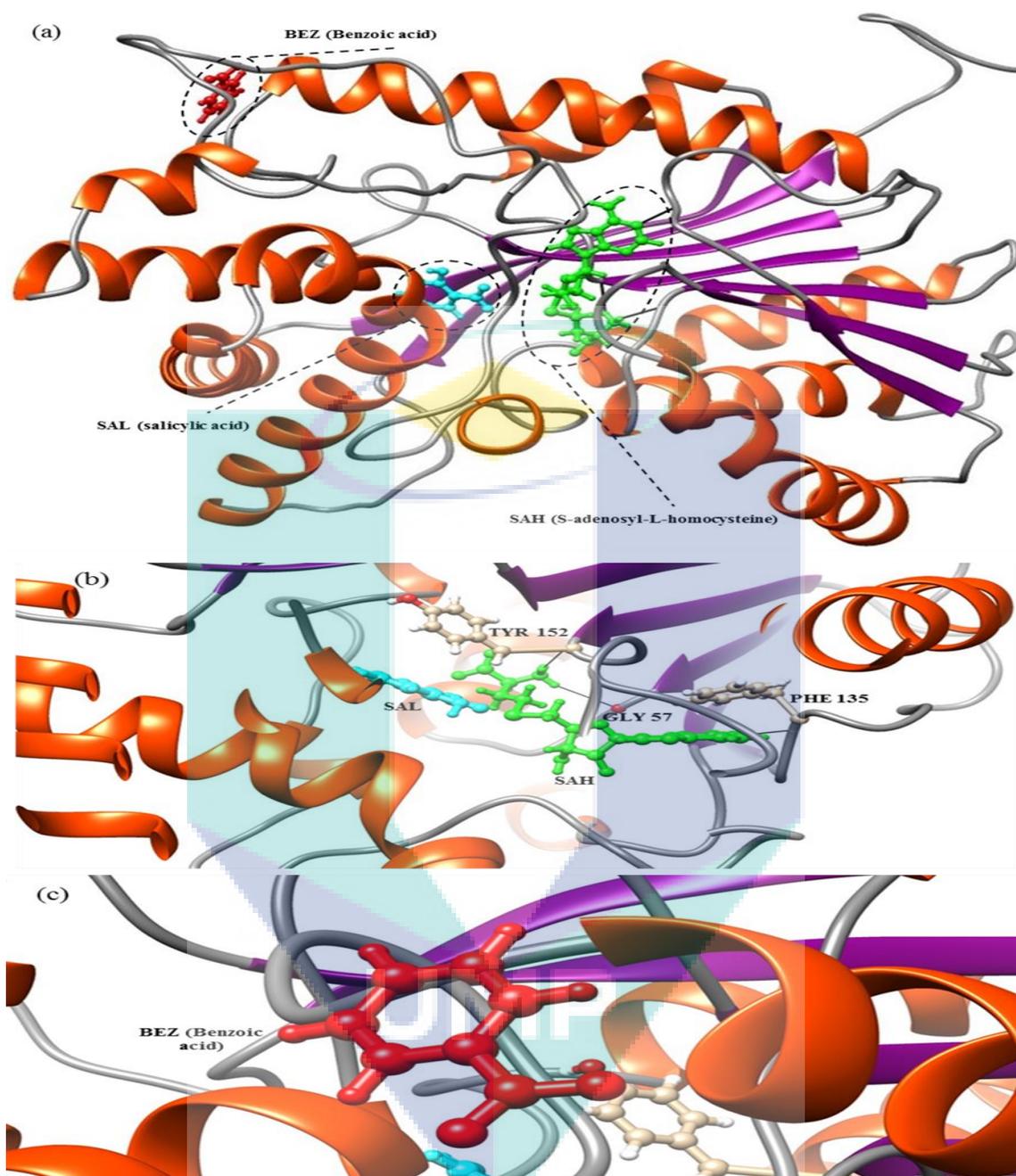


Figure 4.28: Molecular docking of VMPBSMT with its ligand and substrates. (a) general picture; (b) salicylic acid was colored in blue as a substrate and S-adenosyl homocysteine was colored in green as a ligand; (c) benzoic acid as a substrate was colored in red.

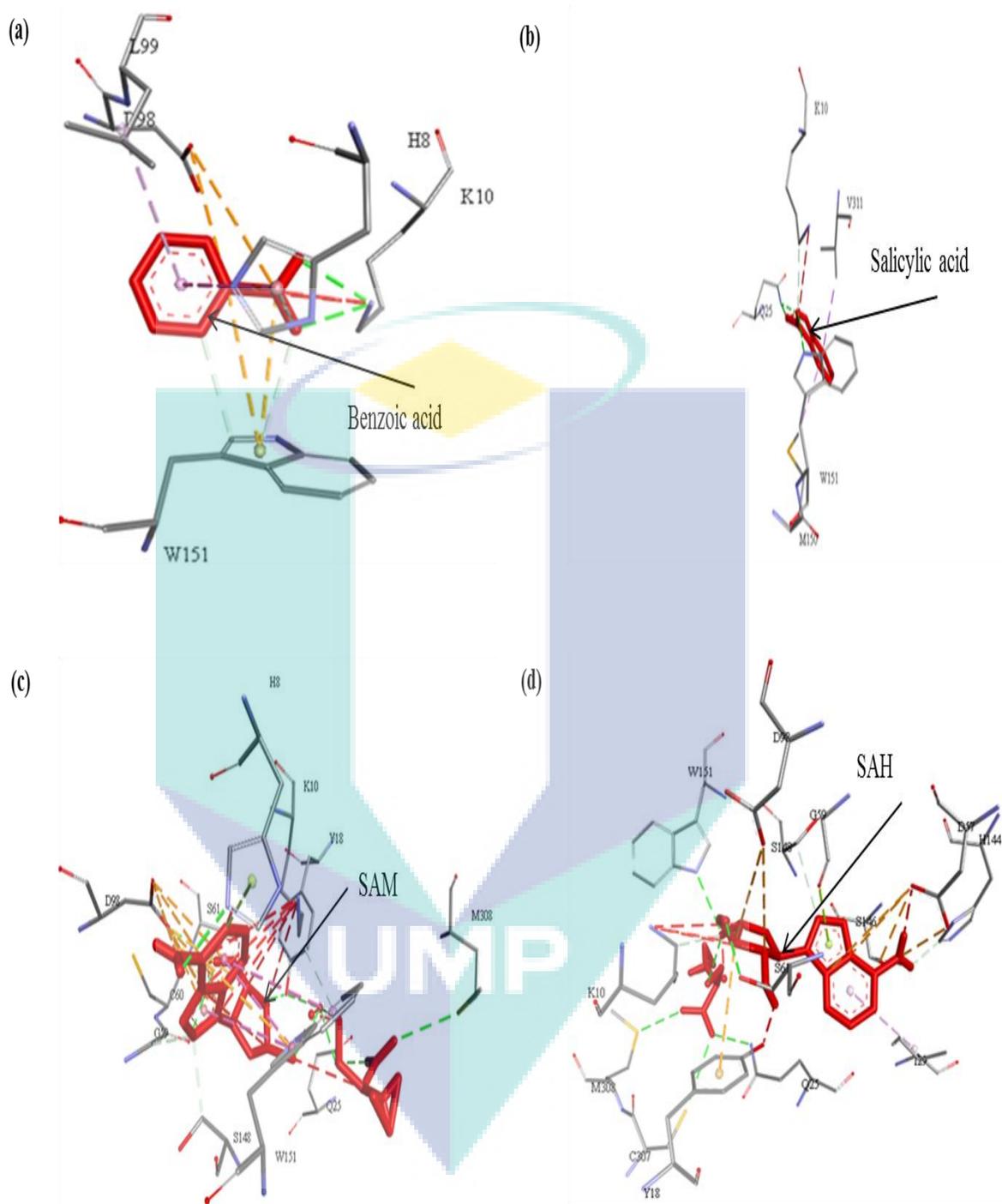


Figure 4.29: VMPBSMT and its ligand and substrate-enzyme complex; (a) VMPBSMT-Benzoic acid complex; (b) VMPBSMT-salicylic acid complex; (c) VMPBSMT-SAM complex; (d) VMPBSMT-SAH complex

development of 3D-structural model of BMPBSMT protein due to no crystalised protein model of benzoic acid carboxyl methyltransferase (BAMT) or benzoic acid/salicylic acid carboxyl methyltransferase (BSMT) is unavailable at the Protein Data Bank.

Meanwhile, testing on suitable ligands including SAM and SAH has shown that VMPBSMT prefers SAH more than SAM based on minimum free binding energy value (refer Table 4.7; Figure 4.28; Figure 4.29). Even though SAH is predicted to be the suitable ligand for VMPBSMT, SAM is also considered as important co-factor for enhancing the methylation reaction in order to produce a methylated-benzenoid volatile product, together with any ester which is preferable with this enzyme. However, in floral tissues of *V. Mimi Palmer*, VMPBSMT might have high possibility to utilise benzoic acid as main substrate in the presence of SAM as cofactor due to the fact that methylbenzoate compound has been identified previously in the fragrance of the orchid instead of methylsalicylate (Mohd-Hairul *et al.*, 2010a). In future, confirmation on the bifunctional characteristics of the protein in substrate specificity to either benzoic acid or salicylic acid can be determined by *in vitro* enzymatic test using recombinant protein of VMPBSMT that has been produced in any microbial system.

4.3.3 Structural and Functional Prediction of VMPEGS Protein

Threading analysis on the deduced amino acid sequence of VMPEGS (Table 4.11) has shown that 3c1o protein template which is eugenol synthase of *C. breweri* as the best suitable protein template to be used as reference protein template in development of 3D-structural model of VMPEGS protein. The 3c10 protein template was chosen as reference protein due to the presence of the template in all the results given by various threading analysis servers, with 43% identity. Even though amino acid sequence of VMPEGS hit only 43% identity to eugenol synthase of *C. breweri* (3c1o) as compared with other non-eugenol synthase subgroup (1qyc, 2gas, 1qyd) with higher identity percentage, 3c1o was selected as a reference protein to VMPEGS due to the fact that the crystal structure of eugenol synthase from *C. breweri* (3c1o) was solved to have lower resolution which is 1.8 Å (Table 4.12) by using x-ray diffraction as reported by Koeduka *et al.*, (2008).

Furthermore, in the MODELLER analysis, the selection of model comparison between 3c1o and other reductases protein models (1qyc, 2gas, 1qyd) was confirmed that the reductases protein models have higher resolution as compared to 3c1o. For this reason, 3c1o is suitable to be used as a protein template for modeling 3D structural model of VMPEGS protein due to lower resolution of crystallised template protein will result in more stable protein structure as recommended by MODELLER procedure for basic modeling (Eswar *et al.* 2007).

Secondary structure analysis on the alignment between VMPEGS and the eugenol synthase of *C. breweri* (3c10) has shown a similar pattern of alternating positions of α -helices and β -strands and thus suggesting their similar structural and functional characteristics (Figure 4.30) as well as having similar domain of NAD(P) binding sites (Figure 4.31 and Figure 4.32). Besides that, the positions of α -helices and β -strands for VMPEGS were also visualised in 3D-structure as shown in Figure 4.33. In this study, a number of five 3D-structural models have been developed for VMPEGS protein by using the MODELLER software, version 9.14 as listed in Table 4.13. Among the developed models, vmpegs.B99990002.pdb has been identified to be the most suitable model of VMPEGS protein based on the lowest DOPE-score as well as GA341 value of 1.0. In addition, further TM-alignment analysis on the chosen 3D-structural model of VMPEGS has shown a Tm-score of 0.987 and RMSD of 0.35 Å. Meanwhile, multiple quality assessment of the developed 3D-model using various protein assessment tools including ERRAT, PROCHECK, VERIFY-3D and Prosa-Web (Table 4.14) has shown that the developed 3D-structural model of VMPEGS protein is considered to be significant and reliable based on their measurement criteria.

Table 4.11: List of suitable protein templates to be used in development of 3D-structural model of VMPEGS protein

Server name	Template	Protein	Organism	Identity	E-value
PSI-BLAST	1qyc	Phenylcoumaran benzylic ether reductase	<i>Pinus taeda</i>	56	7e-118
	2gas	Isoflavone reductase	<i>Medicago sativa</i>	51	5e-102
	1qyd	Pinoresinol-lariciresinol reductase	<i>Thuja plicata</i>	47	3e-98
	3c1o	Eugenol synthase	<i>Clarkia breweri</i>	43	6e-82
	2qw8	Eugenol synthase	<i>Ocimum basilicum</i>	44	1e-77
HHPred	1qyc	Phenylcoumaran benzylic ether reductase	<i>Pinus taeda</i>	56	3.8e-49
		Isoflavone reductase			
	2gas	Pinoresinol-lariciresinol reductase	<i>Medicago sativa</i>	51	4.1e-48
	1qyd	leucoanthocyanidin reductase	<i>Thuja plicata</i>	48	4.5e-48
	3i6i	Eugenol synthase	<i>Vitis vinifera</i>	39	3.5e-41
	3c1o		<i>Clarkia breweri</i>	43	9.2e-42
Phyre2	2qx7	Eugenol synthase	<i>Ocimum basilicum</i>	45	100
	3c1o	Eugenol synthase	<i>Clarkia breweri</i>	43	100
	2gas	Isoflavone reductase	<i>Medicago sativa</i>	50	100
	3i5m	Leucoanthocyanidin reductase	<i>Vitis vinifera</i>	40	100
	1qyc	Phenylcoumaran benzylic ether reductase	<i>Pinus taeda</i>	56	100

pGenThreader	2gas	Isoflavone reductase	<i>Medicago sativa</i>	n/a	3e-16
	1qyc	Phenylcoumaran benzylic ether reductase	<i>Pinus taeda</i>	n/a	4e-15
		Eugenol synthase			
	2r6j	Putative leucoanthocyanidin reductase	<i>Ocimum basilicum</i>	n/a	4e-15
	3i6i	Eugenol synthase	<i>Vitis vinifera</i>	n/a	7e-15
	3c1o		<i>Clarkia breweri</i>	n/a	1e-14
LOMETS	1qyc	Phenylcoumaran benzylic ether reductase	<i>Pinus taeda</i>	n/a	n/a
		Pinoresinol-lariciresinol reductase			
	1qyd	Isoflavone reductase	<i>Thuja plicata</i>	n/a	n/a
	2gas	Eugenol synthase	<i>Medicago sativa</i>	n/a	n/a
	3clo	Leucoanthocyanidin reductase	<i>Clarkia breweri</i>	n/a	n/a
	3i52		<i>Vitis vinifera</i>	n/a	n/a

Table 4.12: Comparison of suitable model for development of 3D-structural model of VMPEGS protein using MODELLER version 9.14

Template	Protein	Organism	Resolution
1qyc	Phenylcoumaran benzylic ether reductase	<i>Pinus taeda</i>	1.6 Å
2gas	Isoflavone reductase	<i>Medicago sativa</i>	1.8 Å
1qyd	Pinoresinol-lariciresinol reductase	<i>Thuja plicata</i>	1.8 Å
3c1o	Eugenol synthase	<i>Clarkia breweri</i>	1.6 Å
2qw8	Eugenol synthase	<i>Ocimum basilicum</i>	1.8 Å

Table 4.13: DOPE and GA341 scores of developed 3D-structural model of VMPEGS protein

Filename	molpdf	DOPE score	GA341 score
vmpegs.B99990001.pdb	1588.41821	-35216.58203	1.00000
vmpegs.B99990002.pdb	1456.62109	-35778.46484	1.00000
vmpegs.B99990003.pdb	1432.38660	-35229.31250	1.00000
vmpegs.B99990004.pdb	1502.86084	-35543.42188	1.00000
vmpegs.B99990005.pdb	1550.18066	-35662.64453	1.00000

```

Q ss_pred      CCCcEEEEEECCCcHHHHHHHHHHHHCCcCEEEEECCc-cchhhHHHHHHhhcCCCEEEEECCCCcHHHHHHHHhhCCCEE
Q VMPEGS      3 SEKSKILIIIGGTGYIGKHVVAASVKLGHPTFVLVVRPSS-ASDPAKVGFLQGLTGAEVKVHGDLDNDYQSLLSAIKQVDIV

T 3c1o_A      2 SHMEKIIYGGTGYIGKFMVRASLSFSHPTFIYARPLTPDSTPSSVQLREEFRSMGVITIEGEMEEHEKMSVSKQVDIV
T ss_dssp     --CCCEEEEETTTTTHHHHHHHHHHHHTCCCEEEEECCcCTTCHHHHHHHHHHHHTTCEEEEECTTCHHHHHHHHTTCSSE
T ss_pred     CCCcEEEEEECCCcHHHHHHHHHHHHCCcCEEEEECCcCCcchhhhhHHHHhhcCCCEEEEECCCCcHHHHHHHHhhCCCEE

Q ss_pred      EECcCCcchhhHHHHHHHHHHCCcCEEEeCCcCCcCCcCCcCCcchHHHHHHHHHHHHHHCCCEEEEECCcchhhc
Q VMPEGS      82 ISTVGGQLADQVNIKAKEAGNKRFLPSEFGLDPERVTKIEPAASALAGKVAIRQAIAEGIPYITIIISNNYFAGYSL

T 3c1o_A      82 ISALPFPMISSQIHIINAIKAAGNKRFLPSEFGLDPERVTKIEPAASALAGKVAIRQAIAEGIPYITIIISNNYFAGYFV
T ss_dssp     EECcCGGSGGGHHHHHHHHHHHHCCcCEEEcCCSScGGGCCCCcHHHHHHHHHHHHHHHHHHHTTCCBEEEECCEHHHHH
T ss_pred     EECcCCcchHHHHHHHHHHHHHHCCcCEEEeCCcCCcCCcCCcCCcchHHHHHHHHHHHHHHCCCEEEEECCcchhhc

Q ss_pred      ccccccc---CCcEEeCCcCEEEEEHHHHHHHHHHHHcChhhcCCEEEeCCcCCcCCcHHHHHHHHHHHHCCcCE
Q VMPEGS      162 PTLAQASP---PVDKITIFGDGNTKGVFVDEDDIGIYTIKAANDPRTLNKIIVYIRPPGCIYSHNELISLWEKKTGKTLER

T 3c1o_A      162 NYLLHPSPHPNRNDIVYGTGETKFLVNYEEDIKAYTIKIVACDPRCCNRIVYIRPPKNIISQNELISLWEAKSGLSFKK
T ss_dssp     HHHHCcSScCTTSCEEEETTSCCEEEEEcHHHHHHHHHHHHcGGGTTEEEEECCcGGGEEEEHHHHHHHHHHHTTSCCE
T ss_pred     ccccccccccEEeCCcCEEEeHHHHHHHHHHHHcChhhcCCEEEeCCcCCcCCcHHHHHHHHHHHHCCcCE

Q ss_pred      EeCHHHHHHHHHcCCcchHHHHHHHHHHcCCcCCcCCcCCcChhhcCCcCCcCCcHHHHHHHHc
Q VMPEGS      239 IYLSDEEIFKIIQEAIPFNVIIFALNYLVFVKGDLSFEIDPSVAEAATELYPEVKYTTVEEYLSRVL

T 3c1o_A      242 VHMPDEQLVRLSQELPQFNIPVSIILHSIFVKGDLMSYEMRKDD-IEASNLYPELEFTSIDGLLDLFI
T ss_dssp     EEECHHHHHHHHHHSGTITHHHHHHHHHHCCcCTTSSCCSSc-EEGGTCTTCCcCCcHHHHHHHHH
T ss_pred     eeCHHHHHHHHHcCCcchhhHHHHHHHHcCCcCCcCCcCCc-cchhhcCCcCCcCCcHHHHHHHHH

```

Figure 4.30: Secondary structure alignment between amino acid residue of VMPEGS and 3c1o template of *Clarkia breweri* eugenol synthase. Positions of β -strands and α -helices are highlighted in blue and red boxes, respectively.

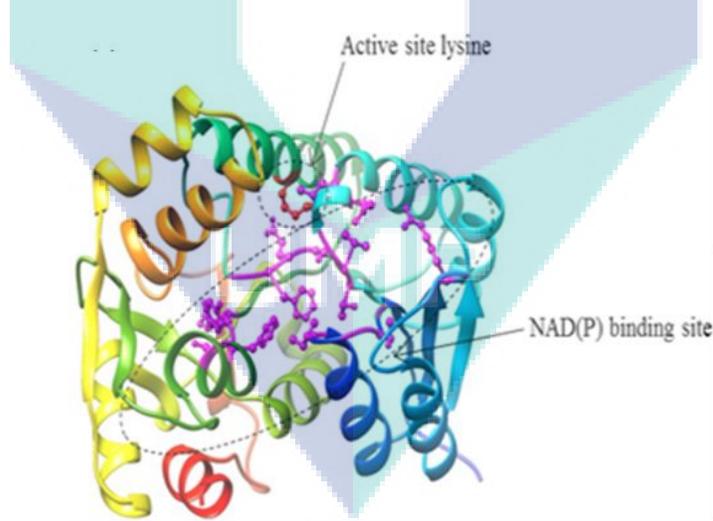


Figure 4.31: VMPEGS model represents NAD(P) binding site and active site of lysine residue. Active site of lysine is shown in red color and NAD(P) binding site in magenta

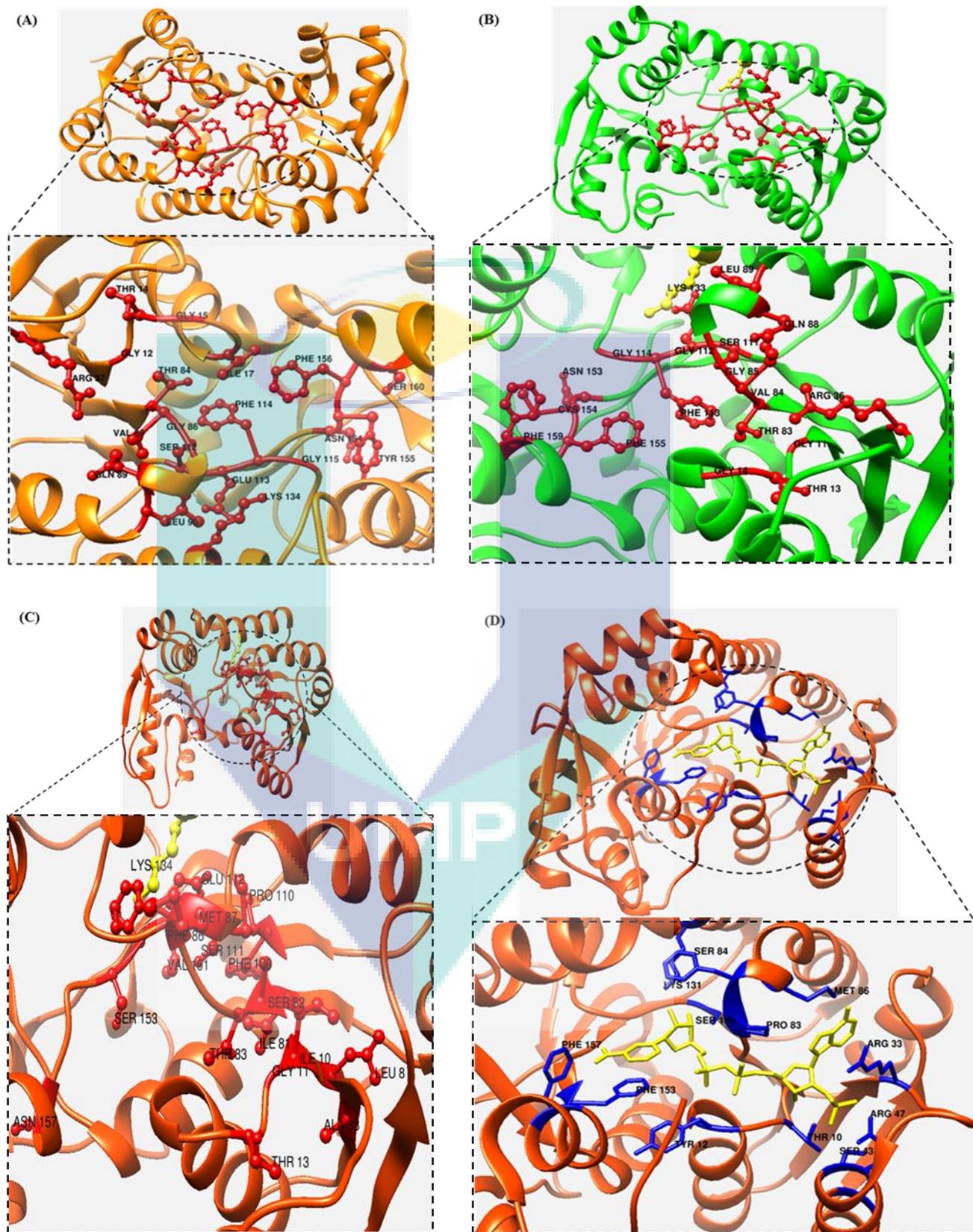


Figure 4.32: NAD(P) binding sites of; (A) VMPEGS; (B) PhEGS; (C) CbEGS; (D) 3c1o

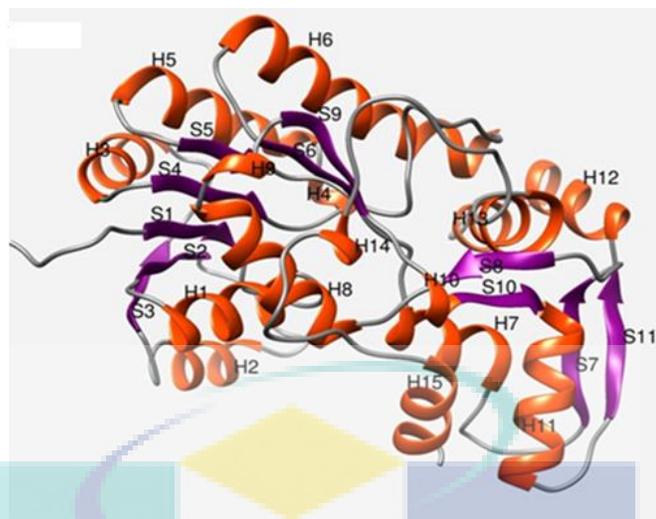


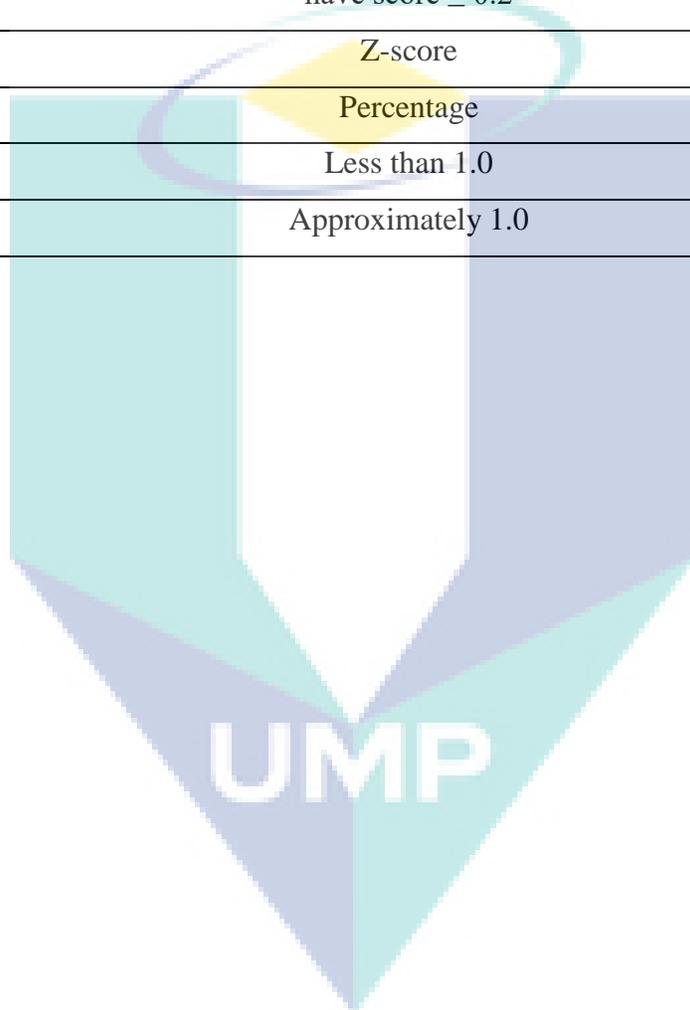
Figure 4.33: 3D model of VMPEGS with the β strands, S1–S11 and the α helices, H1–H15, labelled. The color of α -helices strands and β -strands are orange and purple, respectively. The coil is grey in color.

UMP

Prediction of functional domain that is available in VMPEGS amino acid sequence by using multiple online protein servers including Localizome, ScanProsite and PSI-BLAST has shown the presence of Negative Transcriptional Regulator domain or well known as NmrA domain. Interestingly, the same functional domain is also detected in homologous NmrA-family protein from *C. breweri* (3c1o) that has been chosen as reference protein template for development of 3D-structural model of VMPEGS proteins. Interestingly, the same domain is also identified in homologous proteins from other plants including *Pinus taeda* (1qyc), *Medicago sativa* (2gas), *Thuja plicata* (1qyd), and *Ocimum basilicum* (2qw8). Based on BLASTP analysis VMPEGS amino acid sequence by comparing the sequence with other homologous proteins that are available at the NCBI GenBank database, a NADP binding site consisting of 19 amino acid residues has been identified to be presence in VMPEGS protein and visualised in 3D-structure (refer Figure 4.31 and 36). Interestingly, the same NADP binding site has been identified to be presence in all the homologous proteins including eugenol synthase of *C. breweri* (Figure 4.33). The presence of NADP binding site in eugenol synthase of *C. breweri* and other plants has been reported to be important of specific binding of NADP molecule as ligand for catalytic reaction of eugenol synthase for reduction of coniferyl acetate to eugenol compound (Koeduka *et al.* 2006; 2008; 2009). Besides that, comparative modelling between VMPEGS and its homologous proteins from scented plants including *P. hybrida* (PhEGS) and *Gymnadenia conopsea* (GcEGS) as well as *Vitis venifera* (VvEGS) (Figure 4.34), the same functional domain has been identified. This result is as expected since all the 3D-structural model of VMPEGS, PhEGS, GcEGS and VvEGS were developed using 3c1o protein template of *C. breweri* eugenol synthase as their reference protein models based on their homologous identity of more than 40% and superimposition of models are shown in similar structure based on the same reference protein template (Figure 4.35).

Table 4.14: Evaluation of developed 3D-structural model of VMPEGS protein using protein evaluation tools

Evaluation tools	Criteria	Value/ Score/ Percentage
ERRAT	Overall Quality Factor	85.570
VERIFY 3D	At least 80% of amino acid have score ≥ 0.2	Yes
Prosa-Web	Z-score	-11.05
PROCHECK	Percentage	100%
RMSD	Less than 1.0	0.35 Å
TM-score	Approximately 1.0	0.987



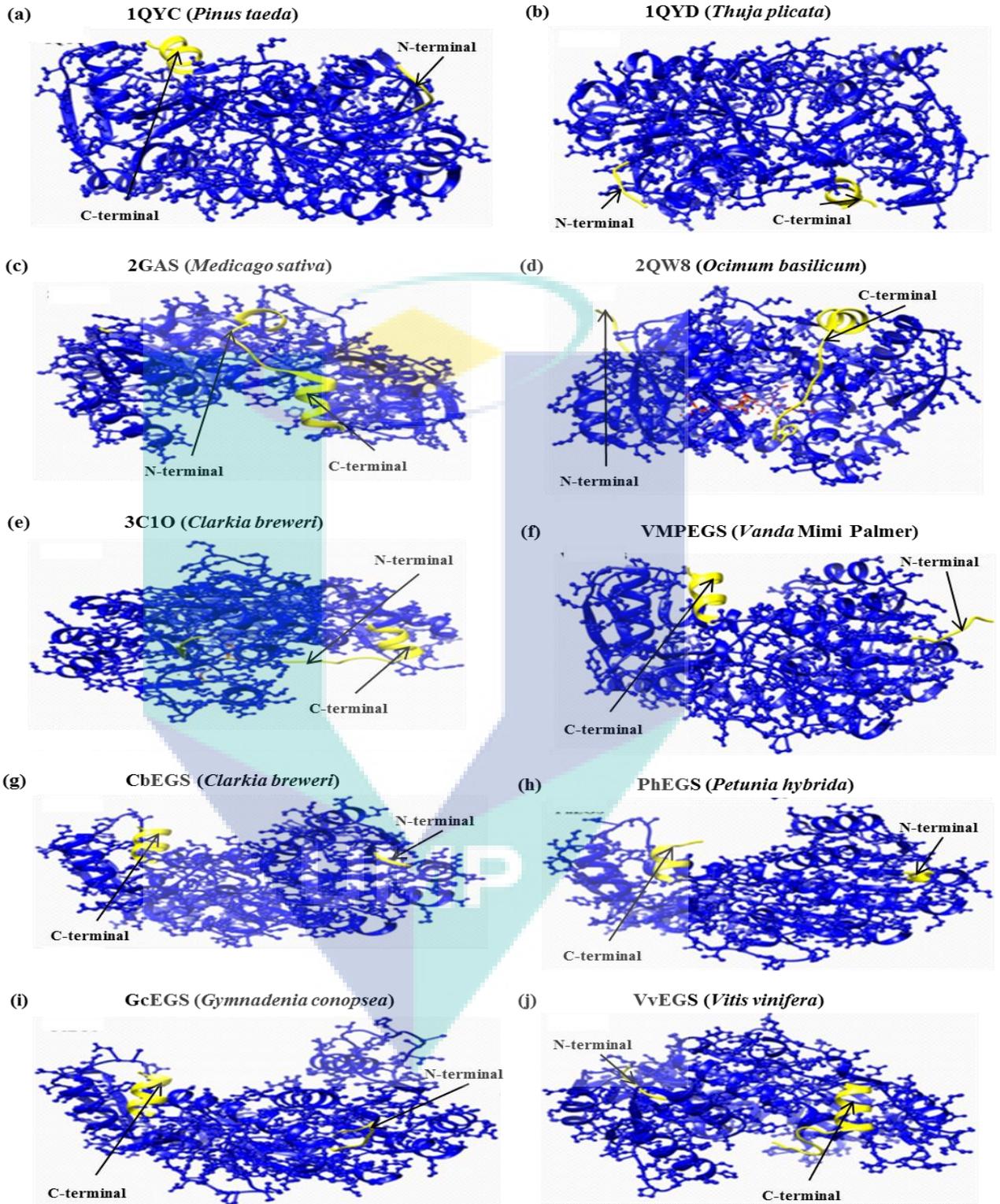


Figure 4.34: Comparative modeling of NmrA-like family enzymes in comparison with VMPEGS protein model. Blue domain represents a NmrA-like family domain and yellow parts are terminal domain at N and C domain.

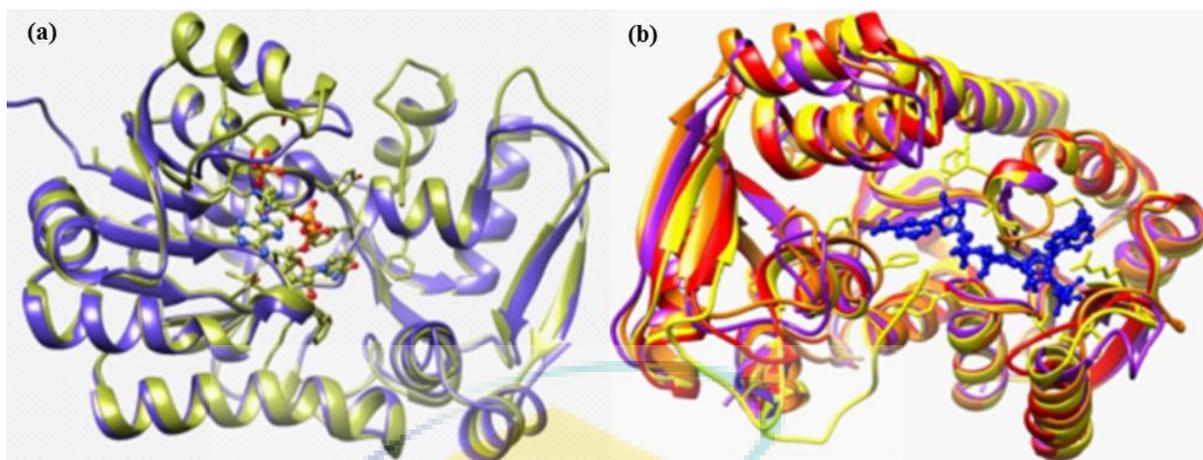


Figure 4.35: (a) Superimposition of the refined VMPEGS model and its template, 3c1o. Both structures are shown in a cartoon representation. The blue ribbon represents the VMPEGS structure, while the light green ribbon represents the 3c1o model. The catalytic regions of both models are represented as balls and sticks. (b) Superimposition of the polypeptide-chain backbones of VMPEGS 3D model and other PIP-family enzymes. The NADP⁺ cofactor of EGS is also colored as blue. The structures of the other PIP-family enzymes were determined in the absence of cofactor. Note: Red ribbon= VMPEGS; purple ribbon= 1qyc; orange ribbon= 2gas; yellow ribbon= 3c1o.

UMP

In fragrance aspect, the involvement of VMPEGS in fragrance compound biosynthesis is still far from understood since no eugenol compound has been detected to be present in as scent constituent of *V. Mimi Palmer* based on scent analysis using Gas Chromatography-Mass Spectrometry (GC-MS) that carried out by Mohd-Hairul *et al.* (2010). In contrast, eugenol compound that is produced by a reduction process of coniferyl acetate, catalysed by eugenol synthase has been detected as one of the scent constituent of *P. hybrida*, *C. breweri* as well as *Gymnadenia conopsea* as volatile fragrance that might be primarily used for their pollination purpose as well as defense against herbivores and pathogens (Koeduka *et al.* 2008; Gupta *et al.* 2014). However, in *V. Mimi Palmer*, the VMPEGS might be involved as biological catalyst in reduction of coniferyl acetate but might not be accumulated in the cells in the form of eugenol compound. The eugenol compound might be subjected to vanillin biosynthesis via a few subsequent reaction steps as trace vanillin compound has been reported to be detected in essential oil extracted from floral tissues of *V. Mimi Palmer* (Mohd-Hairul, 2010; Figure 4.36).

In this study, molecular docking of VMPEGS was carried out by docking between the developed 3D-structural protein of VMPEGS with nicotinamide adenosine phosphate (NAP) as specific ligand besides coniferyl acetate and coniferyl alcohol as its potential substrate (refer Tables 10 and 11; Figure 4.37; Figure 4.38; Figure 4.39). From the molecular docking analysis, coniferyl acetate and nicotinamide adenosine phosphate (NAP) have been identified to be suitable substrate and ligand for catalytic activity of VMPEGS based on their low score of free binding energy and dissociation constant as recommended by Cozzini *et al.* (2004). However, based on docking analysis that has been carried out in this study, coniferyl alcohol might also be suitable to be used as potential substrate with slightly higher free binding energy compared to coniferyl acetate (Table 4.6). Thus, coniferyl alcohol is predicted to bind slightly less specific to VMPEGS compared to coniferyl acetate. Based on the docking analysis, coniferyl alcohol is also detected to be used as suitable substrate for VMPEGS and might be due to closely-related structure between coniferyl alcohol and coniferyl acetate. However, based on previous study on well-studied scented flowers including *C. breweri* and *P. hybrida*, coniferyl alcohol is esterified to coniferyl acetate in the presence of acetyl-CoA in a catalytic reaction by coniferyl

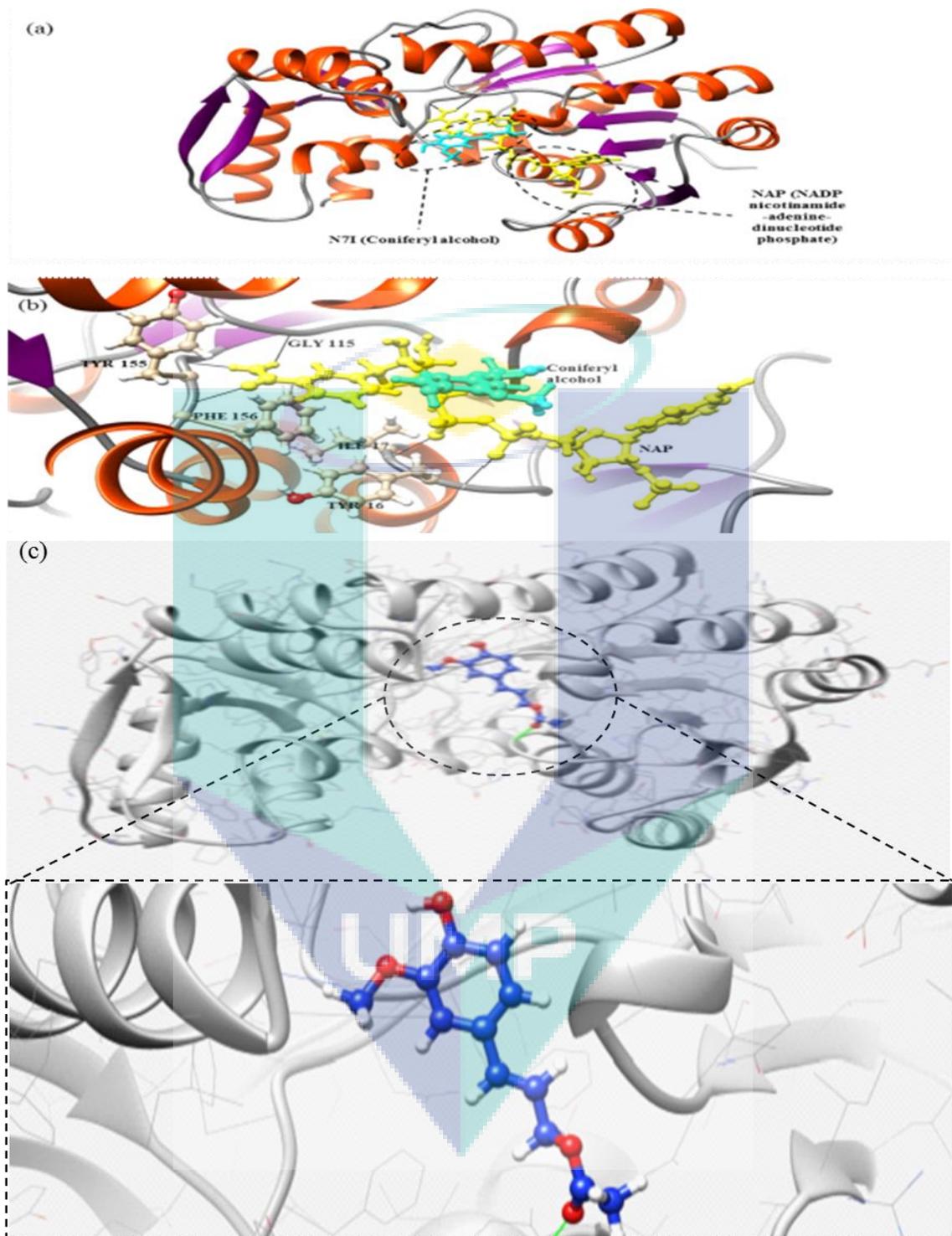


Figure 4.37: Molecular docking of VMPEGS with its ligand and substrate. (a) and (b) represent NAP and coniferyl alcohol as a ligand and substrate, respectively. (c) The grey picture represent coniferyl acetate (blue,red,white) as a substrate for VMPEGS and green line is hydrogen bond.

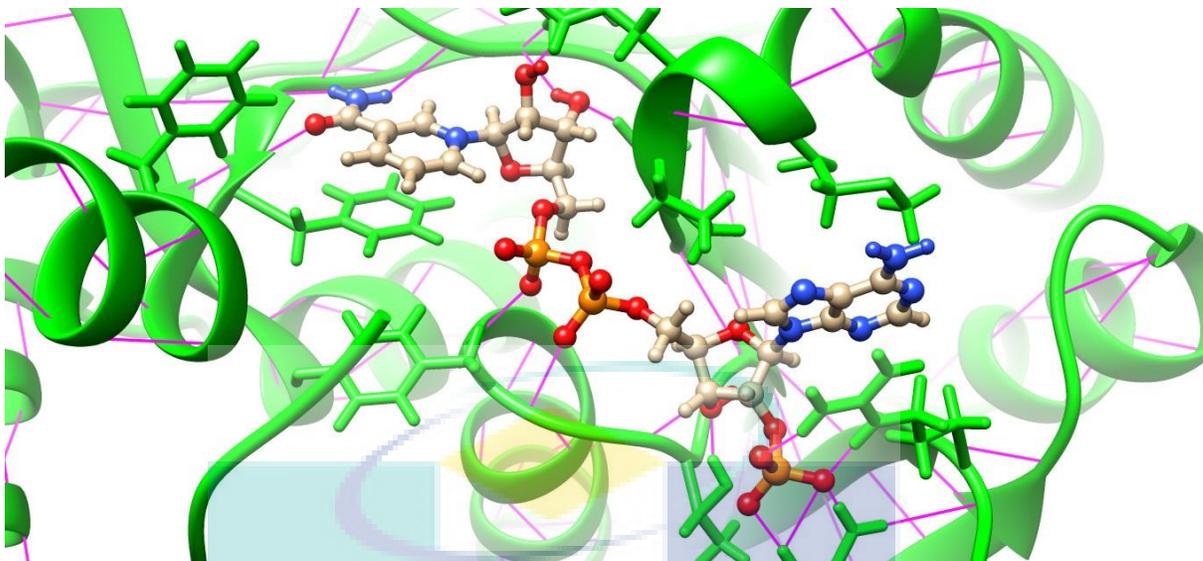


Figure 4.38: Interactions between template model, 3c1o for VMPEGS and the NADP⁺ cofactor. The polypeptide chain segments of 3c1o is designated with green color without changing the original structure from PDB database. Note: The atoms of the NADP⁺ cofactor are drawn as balls and sticks, and are colored coded according to element (carbon: gray; nitrogen: blue; oxygen: red; phosphorus: orange).

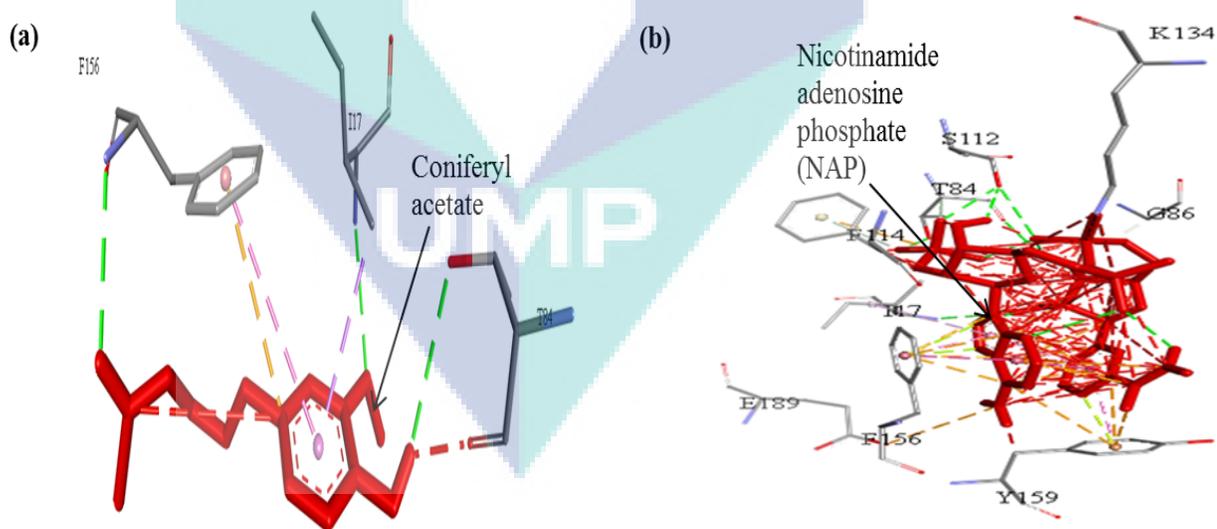


Figure 4.39: Molecular docking of VMPEGS protein with substrate and ligand using Autodock 4.2 software. (a) VMPEGS-coniferyl acetate complex; (b) VMPEGS-NAP complex

acetyltransferase (CFAT) and subsequently reduced to eugenol compound by eugenol synthase (refer Figure 4.36) as described by Koeduka *et al.* (2006). In future, confirmation on suitability of both coniferyl alcohol and coniferyl acetate as suitable substrate for VMPEGS protein can be further tested via *in vitro* enzymatic assay.

4.3.4 Structural and Functional Prediction of VMPOOMT Protein

Threading analysis using multiple protein threading servers on the deduced amino acid sequence of VMPOOMT (Table 4.15) has shown that VMPOOMT is predicted to be well-aligned with chain A of isoflavone o-methyltransferase from *Medicago sativa* (1fp2) with 40-41% identity. Thus, the 1fp2 protein template was selected as reference protein for development of 3D-structural model of VMPOOMT. Secondary structure analysis on the alignment between VMPOOMT and isoflavone o-methyltransferase of *Medicago sativa* (1fp2) has shown a similar pattern of alternate positions of α -helices and β -strands, and thus suggesting their similar structural and functional characteristics (Figure 4.40). Besides that, the positions of 19 α -helices and 7 β -strands in the structure of VMPOOMT protein was visualised in 3D-structure using Chimera software as shown in Figure 4.41(b). In this study, five potential 3D-structural models of VMPOOMT have been developed using MODELLER software, version 9.14 and further subjected to model evaluation based on DOPE-score and GA-341 value. Based on the evaluation (Table 4.16), vmpoomt.B99990005.pdb has to be identified as the most suitable 3D-structural model of VMPOOMT protein. Besides that, the chosen 3D-structural model of VMPOOMT protein has shown a Tm-score of 0.924 and RMSD value of 0.88 Å, reflecting significance and reliability of the protein model based on its Tm-Score of more than 0.4 as well as RMSD value that is lower than 1.0 as recommended by Zhang and Skolnick (2005). In addition, multiple quality assessment the developed 3D-structural model of VMPOOMT protein using various protein assessment on tools including ERRAT, PROCHECK, VERIFY-3D and ProSA-Web (Table 4.17) has also shown the significant and reliability of the developed 3D-structural model of VMPOOMT protein to be used for further molecular docking analysis. In this study, functional domain analysis of amino acid sequence of VMPOOMT protein has shown the presence of two important domains which are dimerisation domain

Table 4.15: List of suitable protein template to be used in 3D-structural model development of VMPOOMT protein

Server name	Template	Protein	Organism	Identity	E-value
PSI-	1fp2	Isoflavone o-methyltransferase	<i>Medicago sativa</i>	40%	3e-85
BLAST	2qyo	Isoflavone o-methyltransferase	<i>Medicago truncatula</i>	38%	1e-83
	4e70	Coniferyl alcohol 9-o-methyltransferase	<i>Linum nodiflorum</i>	39%	3e-81
	1fpx	Selenomethionine isoflavone o-methyltransferase	<i>Medicago sativa</i>	39%	9e-79
	1zga	Isoflavone 4'-o-methyltransferase	<i>Medicago truncatula</i>	39%	1e-67
HHPred	3gwz	Methyltransferase	<i>Streptomyces lavendulae</i>	27%	2.2e-49
	3lst	CALO1 methyltransferase	<i>Micromonospora echinospora</i>	25%	1.3e-48
	3p9c	Caffeic acid O-methyltransferase	<i>Lolium perenne</i>	27%	2.9e-46
	1fp2	Isoflavone O-methyltransferase	<i>Medicago sativa</i>	41%	4.3e-46
	4e70	Coniferyl alcohol 9-O-methyltransferase	<i>Linum nodiflorum</i>	39%	3.7e-46

Phyre2	4e70	Coniferyl alcohol 9-O-methyl transferase	<i>Linum nodiflorum</i>	39%	100
	1zga	Isoflavone 4'-o-methyltransferase	<i>Medicago truncatula</i>	41%	100
	1fp2	Isoflavone o-methyltransferase	<i>Medicago sativa</i>	41%	100
	1kyz	Caffeic acid/5-hydroxyferulic acid 3/5-o-methyltransferase	<i>Medicago sativa</i>	31%	100
	3p9k	Caffeic acid o-methyltransferase	<i>Lolium perenne</i>	28%	100
pGen	2qyo	Isoflavone o-methyltransferase	<i>Medicago truncatula</i>	n/a	n/a
Threader	4e70	Coniferyl alcohol 9-O-methyltransferase	<i>Linum nodiflorum</i>	n/a	n/a
		Isoflavone o-methyltransferase			
	1fp2	Monolignol o-methyltransferase	<i>Medicago sativa</i>	n/a	n/a
	3reo	Isoflavone 4'-o-methyltransferase	<i>Clarkia breweri</i>	n/a	n/a
	1zga		<i>Medicago truncatula</i>	n/a	n/a
LOMETS	4e70	Coniferyl alcohol 9-o-methyl transferase	<i>Linum nodiflorum</i>	n/a	n/a
	1zg3	Isoflavone 4'-o-methyltransferase	<i>Medicago truncatula</i>	n/a	n/a
	4evi	Coniferyl alcohol 9-o-methyl transferase	<i>Linum nodiflorum</i>	n/a	n/a
	1fp2	Isoflavone O-methyltransferase	<i>Medicago sativa</i>	n/a	n/a

Note: Based on threading analysis using various protein threading program, protein template of isoflavone O-methyltransferase of *Medicago sativa* with 40% identity was chosen as the protein template for 3D-structural model development of VMPOOMT protein. n/a= not available.

Table 4.16: DOPE and GA341 scores of developed 3D-structural model of VMPOOMT protein

Filename	molpdf	DOPE score	GA341 score
vmpoomt.B99990001.pdb	1850.55359	-39770.98438	1.00000
vmpoomt.B99990002.pdb	1872.11182	-39549.80859	1.00000
vmpoomt.B99990003.pdb	2006.90283	-39962.87109	1.00000
vmpoomt.B99990004.pdb	1839.79309	-39534.85938	1.00000
vmpoomt.B99990005.pdb	1903.21228	-39970.28516	1.00000

Table 4.17: Evaluation of developed 3D-structural model of VMPOOMT protein using protein evaluation tools

Evaluation tools	Criteria	Value/ Score/ Percentage
ERRAT	Overall Quality Factor	76.190
VERIFY 3D	At least 80% of amino acid have score ≥ 0.2	Yes
Prosa-Web	Z-score	-7.92
PROCHECK	Percentage	99.7%
RMSD	Less than 1.0	0.88 Å
TM-score	Approximately 1.0	0.924

as and *S*-adenosyl-*L*-methionine binding domain as shown in Figure 4.41(a). Dimerisation domain has been reported in most of well-studied plant O-methyltransferase proteins to mediate dimerization process of O-methyltransferase protein at their N-terminal (Zubieta *et al.* 2005). Meanwhile, the other domain which is SAM binding domain has also been reported to be present in most of plant O-methyltransferase proteins as binding site for SAM cofactor in O-methylation reaction.

Interestingly, comparative modeling between VMPOOMT with other plant O-methyltransferases including isoflavone o-methyltransferase (2qyo) and isoflavone 4'-o-methyltransferase (1zga) of *Medicago truncatula*, coniferyl alcohol 9-o-methyltransferase of *Linum nodiflorum* (4e70) and selenomethionine isoflavone o-methyltransferase of *Medicago sativa*, O-methyltransferase of *Vitis vinifera* (VvOMT) and orcinol O-methyltransferase from *R. hybrida* cultivar (RhOOMT) have shown the presence of dimerisation as well as SAM binding domains in the proteins (Figure 4.42; Figure 4.43). In addition, for both VvOMT and RhOOMT, the 3D structure models were developed based on the crystal structure of isoflavone o-methyltransferase of *Medicago sativa* (1fp2) which are similar to VMPOOMT (refer Figure 4.42). Both have similar functional domain of dimerization domain within their structural protein model that are designed using MODELLER software (Figure 4.42). Thus, VMPOOMT resembles the common characteristics of most O-methyltransferase family proteins that contain both dimerization domain and *S*-adenosyl-*L*-methionine (SAM) binding domain as with other o-methyltransferase protein models retrieved from Protein Data Bank (PDB) database including isoflavone o-methyltransferase of *Medicago sativa* (1fp2). Therefore, in *V. Mimi Palmer*, VMPOOMT is predicted to be involved in O-methylation of compound with benzene ring based on the presence of those dimerisation and SAM binding domains.

In fragrance aspect, comparative modeling between VMPOOMT with well-studied OOMT transferase (*RhOOMT*) from *R. hybrida* has also shown the presence of dimerisation as well as SAM binding domains in the proteins (Figure 4.42; Figure 4.43). In *R. hybrida* the RhOOMT was detected to be involved in catalysing methylation of orcinol to 3-hydroxy 5-methoxytoluene and subsequent methylation of 3-hydroxy 5-

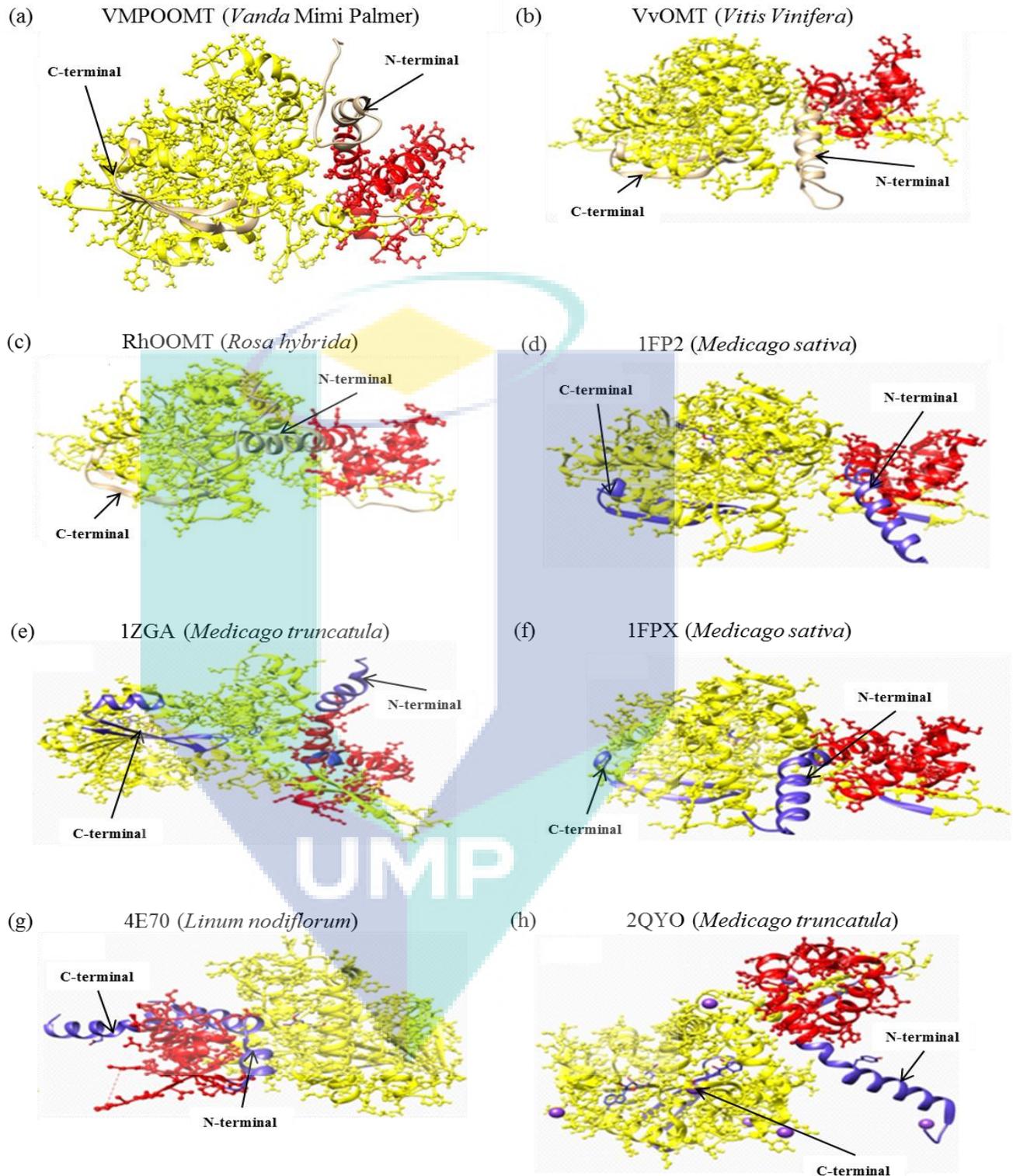


Figure 4.42: Comparative modeling of O-methyltransferase enzymes with VMPOOMT protein model. Red domain is dimerization domain and yellow domain is O-methyltransferase domain.

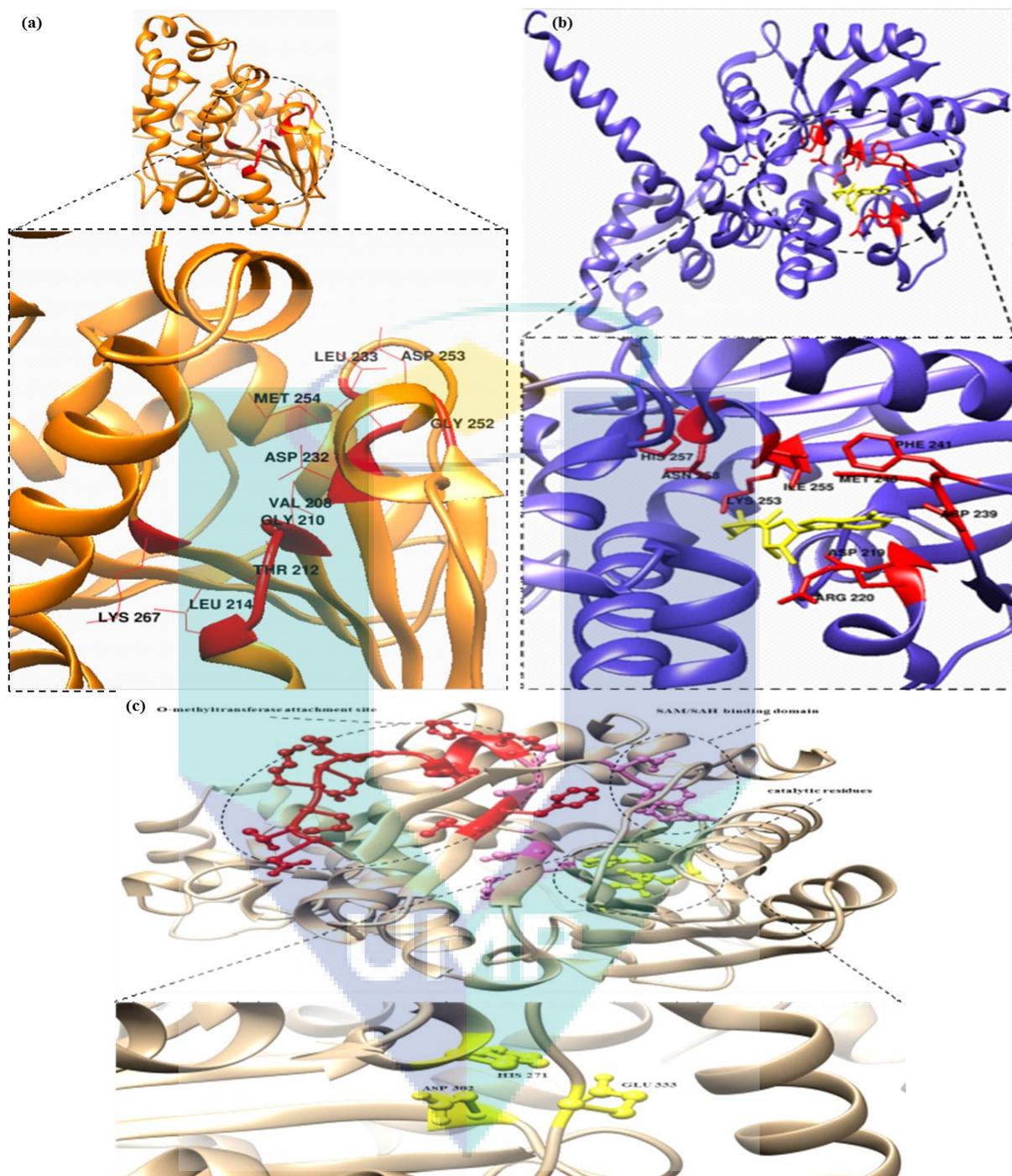


Figure 4.43: SAM binding sites of O-methyltransferase proteins. (a) orcinol o-methyltransferase of *Vanda Mimi Palmer*; (b) Isoflavone o-methyltransferase of *Medicago sativa*, 1fp2. (c) The predicted amino acid residues involved in important domain of VMPOOMT protein model. Red: O-methyltransferase attachment site; Pink: SAM/SAH binding domain; Yellow: catalytic residues represent His 271, Asp 302 and Glu 333 act as catalytic residues in catalytic region of VMPOOMT, docking with possible interacting ligand and substrates.

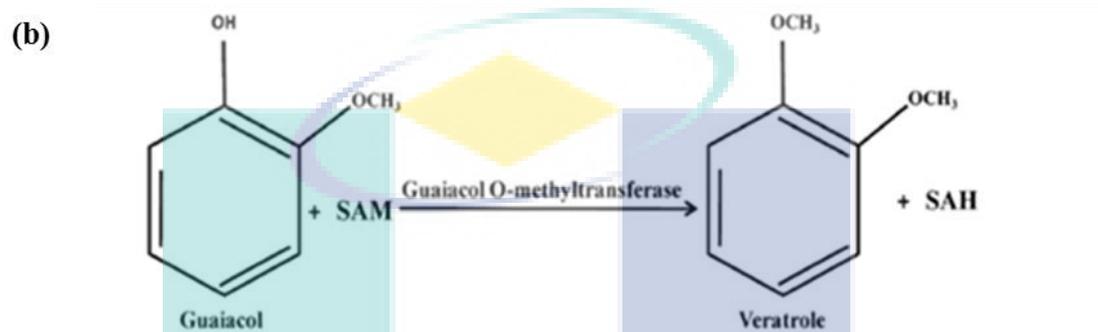
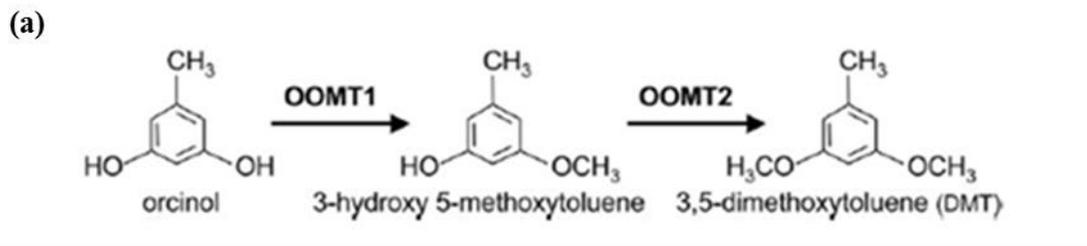


Figure 4.44: Enzymatic reaction of (a) orcinol and (b) guaiacol to DMT and veratrole using *ReOOMT* and *SIGOMT*, respectively. This pathway has been adopted from Scalliet *et al.* 2008 and Gupta *et al.*, 2012.

UMP

methoxytoluene to 3,5-dimethoxytoluene (DMT) (Scalliet *et al.* 2008; refer Figure 4.44a). In fragrance studies, no other orcinol methyltransferase has been reported from other well studied scented plants including *A. majus*, *C. breweri* as well as *P. hybrida*. However, there are several other O-methyltransferases has been studied in other scented plants including guaiacol O-methyltransferase of *Silene latifolia* that catalyse methylation of guaiacol for veratrole biosynthesis (Gupta *et al.*, 2012; refer Figure 4.44(b)), phloroglucinol O-methyltransferase from *Rosa chinensis* that catalyse biosynthesis of estragole to volatile 1,3,5-trimethoxybenzene (Wu *et al.*, 2004), allylphenol O-methyltransferase in sweet basil (*Ocimum basilicum*) that is involved in methyl-eugenol biosynthesis (Lewinsohn *et al.*, 2000), O-methyltransferases in *Vitis vinifera* (grape) that is involved in the biosynthesis of methoxypyrazines as one of the constituents in its aroma (Dunlevy *et al.*, 2010), catechol-O-methyltransferase of *Solanum lycopersicum* that involved as catalyst in guaiacol biosynthesis (Mageroy *et al.*, 2012) as well as eugenol O-methyltransferase (*CbEOMT*) of *C. breweri* that involved in biosynthesis of methyleugenol (Wang *et al.*, 1997). In this study, isoflavone o-methyltransferase of *Medicago sativa* (1fp2) was selected as reference protein template instead of OOMT from *R. hybrida* as well as OOMT due to the unavailability of crystallised structure of OOMT protein at the PDB databank. However, 3D-structural model of OOMT from *R. hybrida* (RhOOMT) has been developed in this study to investigate and predict its similar putative structural and functionality to VMPOOMT. Interestingly, threading analysis on RhOOMT using multiple online servers (Appendix I(k)) has shown that the same isoflavone o-methyltransferase of *Medicago sativa* (1fp2) is suitable to be used for the development of 3D-structural protein of RhOOMT protein. Comparative modeling between VMPOOMT and RhOOMT has also shown the presence of similar dimerisation as well as *S*-adenosyl-*L*-methionine binding domain (refer Figure 4.43) that might reflect their similar structural and functionality in catalysing biosynthesis of DMT from orcinol compound.

Furthermore, molecular docking has been carried out on the developed 3D-structural model of VMPOOMT protein with potential substrates and ligands by using Autodock software, version 4.2. Based on the molecular docking analysis, three substrates have been identified to be suitable for catalytic binding to VMPOOMT with low free

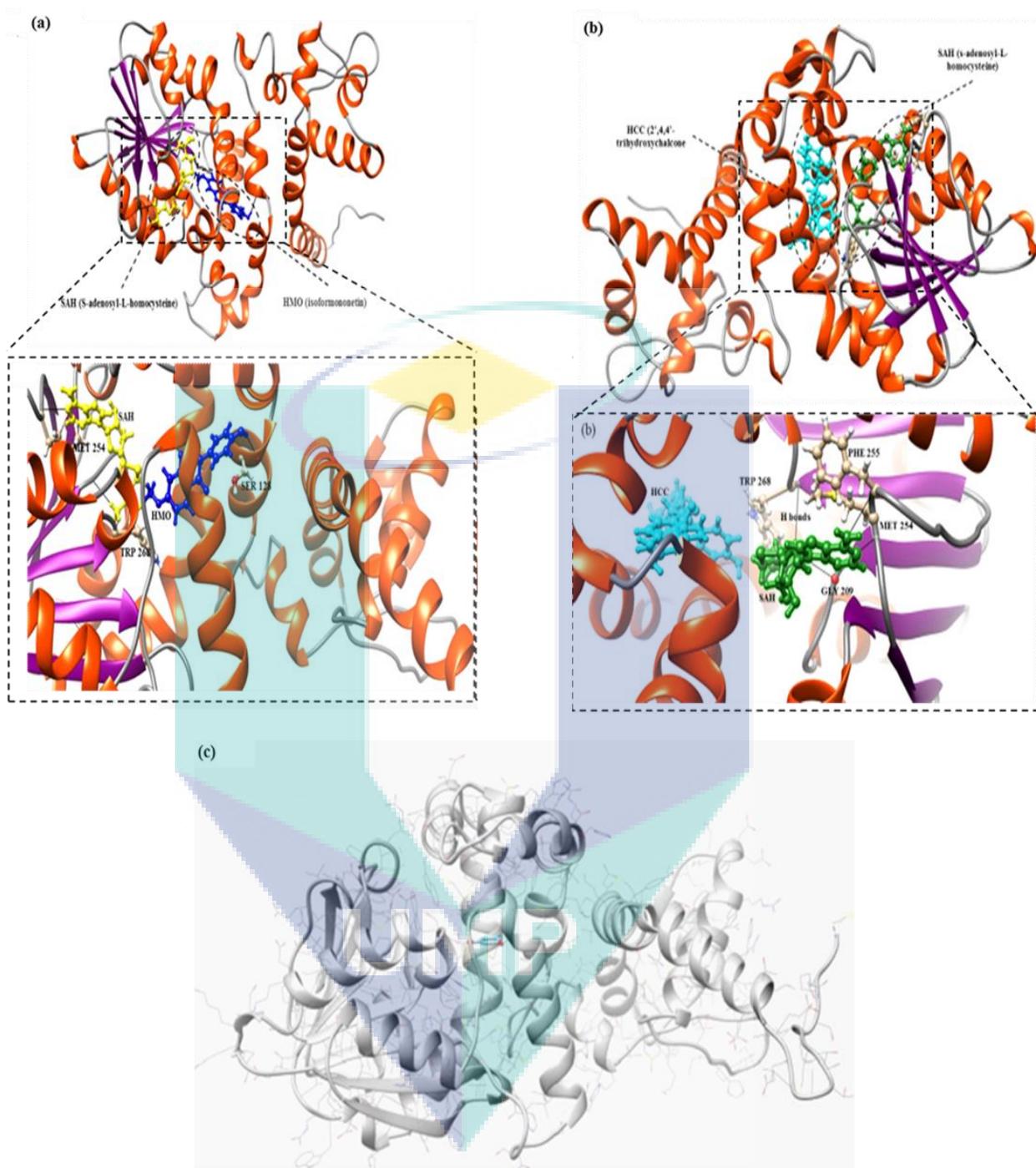


Figure 4.45: VMPOOMT and its predicted substrates and ligands; (a) represent SAH as a ligand and HMO as a substrate for VMPOOMT protein model; (b) represent HCC as a substrate and SAH as a ligand; (c) Blue-red ligand is denoted as orcinol in catalytic region of VMPOOMT protein model.

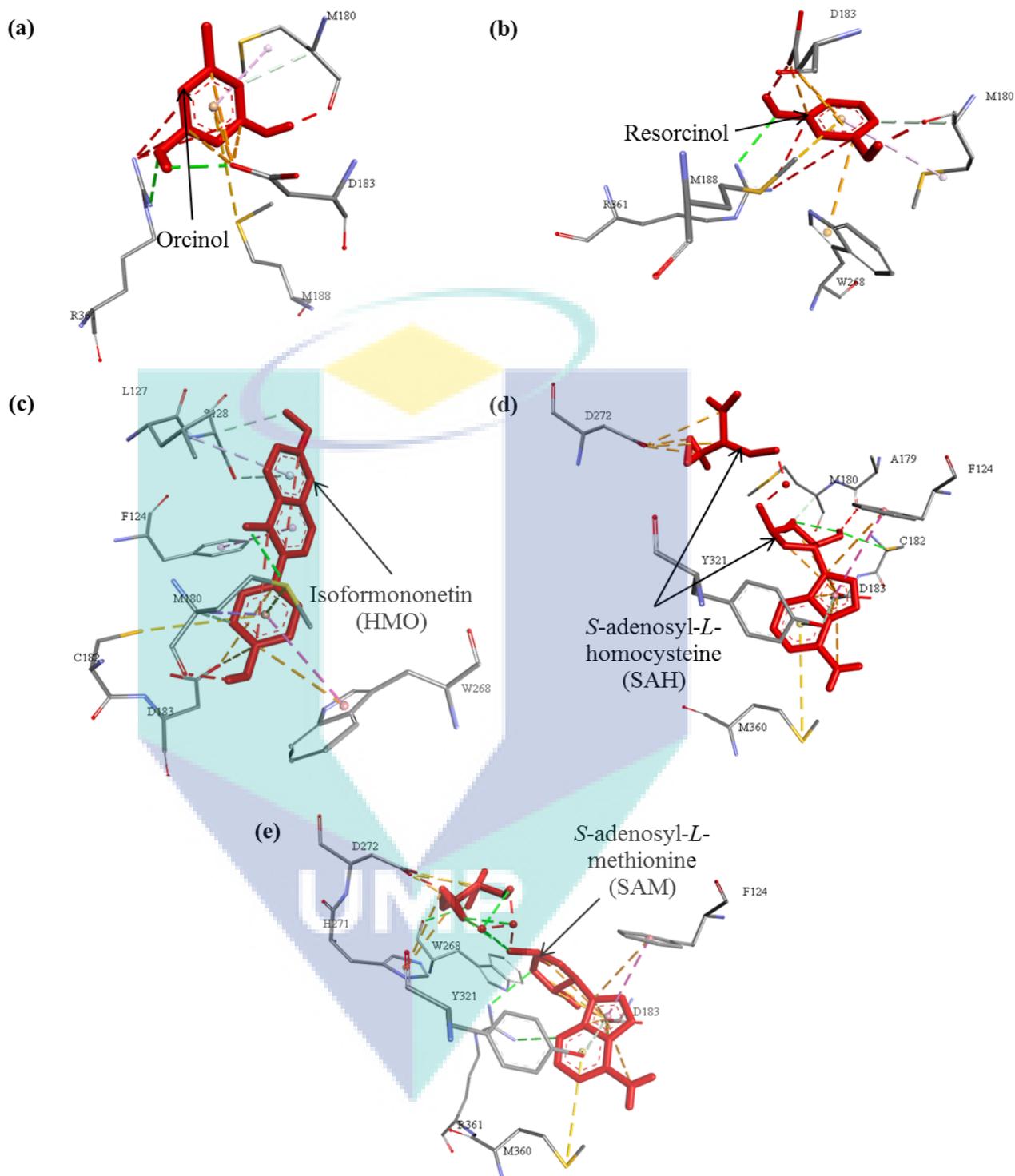


Figure 4.46: Molecular docking of VMPOOMT protein with substrates and ligands using Autodock 4.2 (a) VMPOOMT-Orcinol complex; (b) VMPOOMT-resorcinol complex; (c) VMPOOMT-HMO complex; (d) VMPOOMT-SAH complex; (e) VMPOOMT-SAM complex

binding energy including orcinol, resorcinol and isoformononetin (HMO) (Table 4.6; Figure 4.45). Meanwhile, S-adenosyl-L-homocystein (SAH) and S-adenosyl-L-methionine (SAM) have been identified to be used as ligand in catalytic reaction of VMPOOMT (Table 4.7; Figure 4.46). In this study, orcinol is selected as potential substrate due to the success of previously reported *in vitro* enzymatic assay of RhOOMT in catalysing methylation of orcinol for the biosynthesis of 3,5-dimethoxytoluene in *R. hybrida*. Interestingly, other compounds which are resorcinol and isoformononetin (HMO) have also shown their suitability as potential substrate candidates for VMPOOMT protein. The suitability of resorcinol could be due to its similar main structure with the lacking of a methyl group at carbon-5 of benzene group of the structure (Lavid *et al.* 2002). Meanwhile another potential substrate candidate which is isoformononetin (HMO) is detected to be suitable for catalytic activity of VMPOOMT could be due to the successful *in vitro* enzymatic assay on isoflavone o-methyltransferase of *Medicago sativa* (1fp2) that has been used as reference protein template in development of 3D-structural model of VMPOOMT protein. Meanwhile, testing on ligand including SAM and SAH has shown that SAH is more preferable for catalytic binding of VMPOOMT instead of SAM based on its lower free binding energy (refer Table 4.7). However, further *in vitro* enzymatic tests should be carried out in future to confirm the suitability of those substrate and ligand candidates in catalytic function of VMPOOMT protein.

4.4 RELATIVE EXPRESSION ANALYSIS OF FRAGRANCE-RELATED GENES OF VANDA MIMI PALMER

The four fragrance-related genes of *V. Mimi Palmer* including VMPKAT, VMPBSMT, VMPEGS and VMPOOMT were selected for gene expression analysis using real-time RT-PCR to determine their relative expression levels in different tissues and developmental stages. Studies on gene expression analysis in different tissue binding energy including orcinol, resorcinol and isoformononetin (HMO) (Table 4.6; Figure 4.45). Meanwhile, S-adenosyl-L-homocystein (SAH) and S-adenosyl-L-methionine (SAM) have been identified to be used as ligand in catalytic reaction of VMPOOMT (Table 4.7; Figure 4.46). In this study, orcinol is binding energy including orcinol, resorcinol and

isoformononetin (HMO) (Table 4.6; Figure 4.45). Meanwhile, S-adenosyl-L-homocystein (SAH) and S-adenosyl-L-methionine (SAM) have been identified to be used as ligand in catalytic reaction of VMPOOMT (Table 4.7; Figure 4.46). In this study, orcinol is binding energy including orcinol, resorcinol and isoformononetin (HMO) (Table 4.6; Figure 4.45). Meanwhile, S-adenosyl-L-homocystein (SAH) and S-adenosyl-L-methionine (SAM) have been identified to be used as ligand in catalytic reaction of VMPOOMT (Table 4.7; Figure 4.46). In this study, orcinol is and at different developmental stages might reflect the mechanism of final fragrance product biosynthesis. Previously, expression analyses by real-time RT-PCR have shown that all the fragrance-related genes of *V. Mimi Palmer* including phenylacetaldehyde synthase (VMPPAAS), sesquiterpene synthase (VMPSTS) and alcohol acyltransferase (VMPPAAT) are up-regulated in floral tissues including petal and sepal but down-regulated in vegetative tissues including leaf, root and shoot. The results were in accordance to the presence of trichome glands in petal and sepal of the flower as storage and release site for fragrance that are important to attract pollinators including insects and birds for pollination purpose (Mohd-Hairul *et al.*, 2010b; Chan *et al.*, 2011). In addition, all those fragrance-related genes have been detected to be developmentally-regulated whereby no expression was detected in bud tissues and started to be detected during half-open flower stage and reached the highest expression level in fully-open flower. Expression of those fragrance-related genes agrees with emission of fragrance compounds that was detected to be developmentally regulated as reported by Mohd-Hairul *et al.*, (2010a), reflecting fragrance biosynthesis in *V. Mimi Palmer* is developmentally regulated. Thus, in this study, expression of four selected fragrance-related genes including VMPKAT, VMPBSMT, VMPEGS and VMPOOMT is expected to be up-regulated in floral tissues especially in petal and sepal as the main location of fragrance biosynthesis. Besides that, expression of those fragrance-related genes is expected to be developmentally regulated, reflecting the developmentally-regulated pattern of fragrance compounds biosynthesis including benzenoids phenylpropanoids that contribute to the sweet fragrance of *V. Mimi Palmer*.

4.4.1 Relative Expression Analysis of *Vanda Mimi Palmer* 3-Keto-acyl CoA Thiolase (VMPKAT)

Relative expression studies of VMPKAT gene in different floral and vegetative tissues of *V. Mimi Palmer* (Figure 4.47(a)) showed up-regulated expression in floral tissues especially in petal and sepal compared to vegetative tissues. The level of VMPKAT transcript in petal and sepal was detected to be more than twice compared to bud as selected calibrator tissue. In *P. hybrida*, another well-studied scented flower, a homologous 3-ketoacyl-CoA thiolase (PhKAT1) gene expression was detected to be up-regulated in petal and down-regulated in sepal and all vegetative tissues (Van Moerkercke *et al.*, 2009). This contrasting result on the expression of the VMPKAT transcripts in sepal of *V. Mimi Palmer* might be due to the nature of this orchid whereby its sepals and petals have similar structure and function (Mohd-Hairul *et al.*, 2011). This condition can be proven by previous histological studies on *V. Mimi Palmer* that showed the availability of a high number of trichomes in petal and sepal that represent the presence of potential sites for accumulation of fragrance compounds that are derived from fragrance biosynthetic pathway of *V. Mimi Palmer* (Janna Ong Abdullah, unpublished data).

Meanwhile, in different floral developmental stages, the expression of VMPKAT gene has been detected to increase gradually from bud to half-open flower and reaches the highest level in fully-open flower stage (Figure 4.47(b)). This result is in accordance to the emission increment of volatile benzenoids including benzyl acetate, methylbenzoate, phenylethyl acetate as well as phenylethylalcohol by fully-open flower of *V. Mimi Palmer* compared to half-open flower, and none of the compounds has been detected to be emitted by bud of the flower (Mohd-Hairul *et al.*, 2010a). This expression profile result was as expected since in *P. hybrida*, PhKAT1 and PhKAT2 were also detected to be developmentally regulated in petal as the main floral tissue for fragrance biosynthesis. Interestingly, similar expression pattern was also detected for fragrance-related transcripts in other well-studied scented flowers including *Phalaenopsis bellina*, *A. majus* and *R. hybrida* (Hsiao *et al.*, 2008; Kaminaga *et al.*, 2006; Farhi *et al.*, 2010). In *V. Mimi Palmer*, VMPKAT might be involved indirectly in catalysing the biosynthesis of precursors for

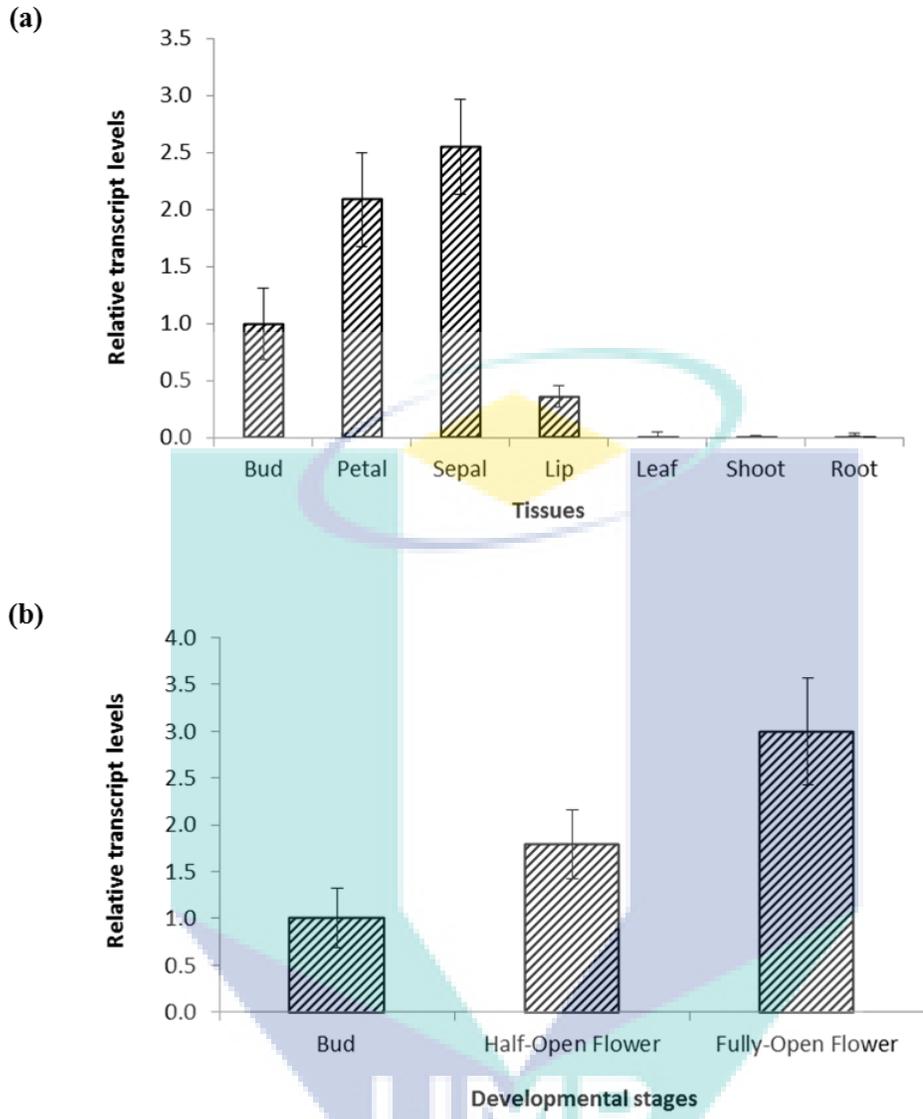


Figure 4.47: Relative VMPKAT transcript levels in *Vanda Mimi Palmer* detected in (a) different tissues; (b) different flower developmental stages

volatile benzenoid and phenylpropanoid compounds via either β -oxidative or non- β -oxidative pathways. In this study, VMPKAT was expected to catalyse the biosynthesis of precursor of benzoic acid as intermediate volatile benzenoid that will be further subjected to a methylation step for methylbenzoate (MeBA) production which is one of the abundant volatile compounds emitted by *V. Mimi Palmer*. Based on expression analysis of VMPKAT gene in different developmental stages, the result was in accordance to previous volatile analysis on *V. Mimi Palmer* whereby methylbenzoate emission profile was found to be developmentally-regulated.

4.4.2 Relative Expression Analysis of *Vanda Mimi Palmer* Benzoic Acid/Salicylic Acid Carboxyl Methyltransferase (VMPBSMT)

Relative expression analysis of VMPBSMT gene in different tissues (Figure 4.48(a)) of *V. Mimi Palmer* has shown a similar expression pattern like VMPKAT whereby the level of the transcript is detected to be up-regulated in floral tissues especially in petal and sepal and down-regulated in vegetative tissues including leaf, shoot and root. In this study, expression level of VMPBSMT transcript in petal and sepal has been detected to be more than 3,000 times higher compared to bud as selected calibrator tissue. This result was as expected due to the fact that VMPBSMT is expected to be involved in catalysing biosynthesis of methylbenzoate from benzoic acid as its precursor as reported on homologous benzoic acid carboxylmethyltransferase from other plants including *A. majus* and *Stephanotis floribunda* (Pott *et al.*, 2004; Negre *et al.*, 2003). In addition, methylbenzoate compound that might be directly derived from benzoic acid in the presence of VMPBSMT as its biological catalyst has been detected to be emitted in abundant level between 8.00am to 12.00pm as reported by Mohd-Hairul *et al.* (2010). While in lip, expression level of VMPBSMT has been detected to be much lower (~1,500 times lower) compared to both petal and sepal of *V. Mimi Palmer* but much higher (~200 times higher) compared to bud as calibrator tissue. This result agreed with histological analysis on *V. Mimi Palmer* whereby much lower number of trichome presence in lip compared to petal and sepal of this orchid (Janna Ong Abdullah, unpublished data).

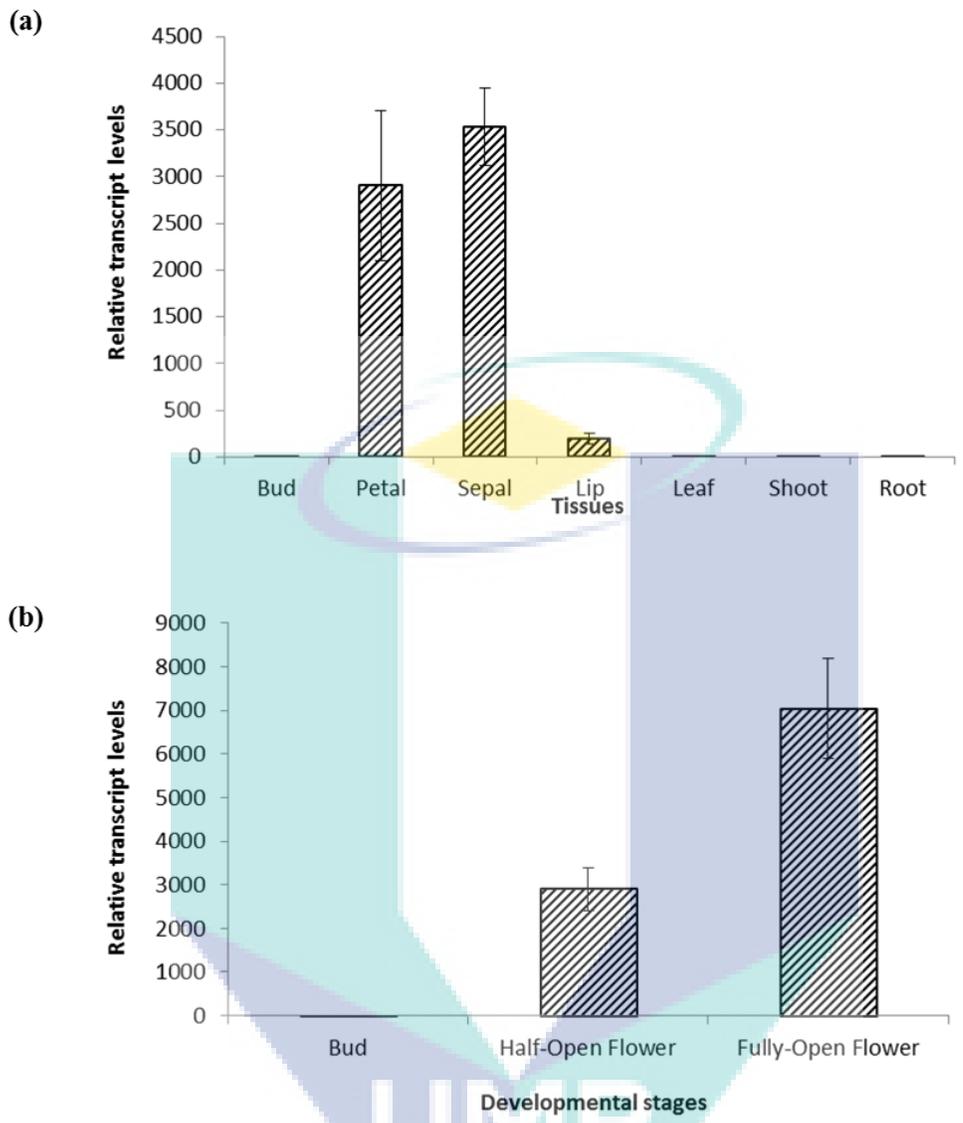


Figure 4.48: Relative VMPBSMT transcript levels in *Vanda Mimi Palmer* detected in (a) different tissues; (b) different flower developmental stages

Meanwhile, in different floral developmental stages (refer Figure 4.48(b)), expression of VMPBSMT gene has been detected to be developmentally regulated whereby the transcript level increases gradually from bud to half-open flower and reaches the highest level in fully-open flower stage. This result was in accordance with the increase of methylbenzoate emission levels by the fully-open flower of *V. Mimi Palmer* in fully-open flower as compared to half-open flower and none of the volatile compound was detected in bud stage as reported by Mohd-Hairul *et al.* (2010a). Interestingly, a similar expression pattern of fragrance-related transcripts has also been detected in other well-studied scented flowers including *Phalaenopsis bellina*, *A. majus*, *P. hybrida* and *R. hybrida* (Farhi *et al.*, 2010; Hsiao *et al.*, 2008; Nagegowda *et al.*, 2008; Kaminaga *et al.*, 2006). In *V. Mimi Palmer*, VMPBSMT is predicted to catalyse a methylation reaction of benzoic acid for production of methylbenzoate as discussed previously in Section 4.4.1.

4.4.3 Relative Expression Analysis of *Vanda Mimi Palmer* Eugenol Synthase (VMPEGS)

Relative expression studies on VMPEGS gene in different floral and vegetative tissues of *V. Mimi Palmer* (Figure 4.49(a)) has shown a contradict result compared to VMPKAT and VMPBSMT. In petal and sepal of *V. Mimi Palmer*, the expression level of VMPEGS transcript has been detected to be almost half lower than bud as selected calibrator tissue. This unexpected result was possibly due to the involvement of VMPEGS in catalysing biosynthesis of eugenol as intermediate for other phenylpropanoid derivatives. For example, eugenol compound might be further utilised as the main precursor for subsequent reactions in vanillin biosynthesis whereby traces of vanillin compound has been detected previously in essential oil extracted from the flower of *V. Mimi Palmer* (Mohd-Hairul, 2011). However, no investigation has been carried out on phenylpropanoid compounds in bud tissues whether eugenol compound is accumulated in the tissues or the compound is converted to some other phenylpropanoid derivatives.

Interestingly, the presence of VMPEGS transcripts in lip and root was detected to be slightly higher than petal and sepal, and slightly lower compared to bud. This result might

possibly be due to the involvement of eugenol compound as the main precursor in the biosynthesis of other phenylpropanoids that might have some medicinal properties. Previously, extracts from flowers, leaves and roots of *V. tessellata*, one of the parents of *V. Mimi Palmer*, has been reported to be traditionally used by local folks in India to treat various medicinal conditions including inflammatory treatment, otitis remedy, and rheumatic treatment besides other treatment for fever, dyspepsia, bronchitis and neurophatological problems (Chowdhury *et al.*, 2014). In addition, the anti-inflammatory effectiveness of *V. tessellata* extract has been proven scientifically on mice besides the potential of the extract to enhance male sexuality in normal mice (Kumar *et al.*, 2000). Thus, *V. Mimi Palmer* might inherit some of the medicinal properties from *V. tessellata* as one of its parents due to the fact that half of the gene pool in *V. Mimi Palmer* is derived from *V. tessellata*. Unfortunately, the expression of VMPEGS in leaf and shoot of *V. Mimi Palmer* is detected to be much lower compared to lip, root and bud, possibly due to the focus of leaf and shoot in photosynthesis as its primary metabolism instead of phenylpropanoid biosynthesis as its secondary metabolism.

Meanwhile, in different developmental stages, the expression pattern of VMPEGS is totally different compared to other fragrance-related transcripts including VMPBSMT and VMPKAT as well as phenylacetaldehyde synthase (VMPPAAS), alcohol acyltransferase (VMPAAT) and sesquiterpene synthase (VMPSTS) that lowly expressed in bud and followed by gradual expression increment in half-open flower and reaches the highest level during fully-open flower stage (Mohd-Hairul *et al.*, 2010b; Chan *et al.*, 2011). In this study, the expression of VMPEGS has been detected to be almost five times higher in bud compared to half-open and fully open flower stages (refer Figure 4.49(b)). This contradicting result was reasonable since during half-open and fully-open flower stage, floral tissues of *V. Mimi Palmer* are focusing on fragrance-compounds biosynthesis including volatile monoterpenes, sesquiterpene as well as benzenoids and phenylpropanoids as reported by Mohd-Hairul *et al.* (2010) instead of production of non-fragrance phenylpropanoid compounds that might have some medicinal or other properties.

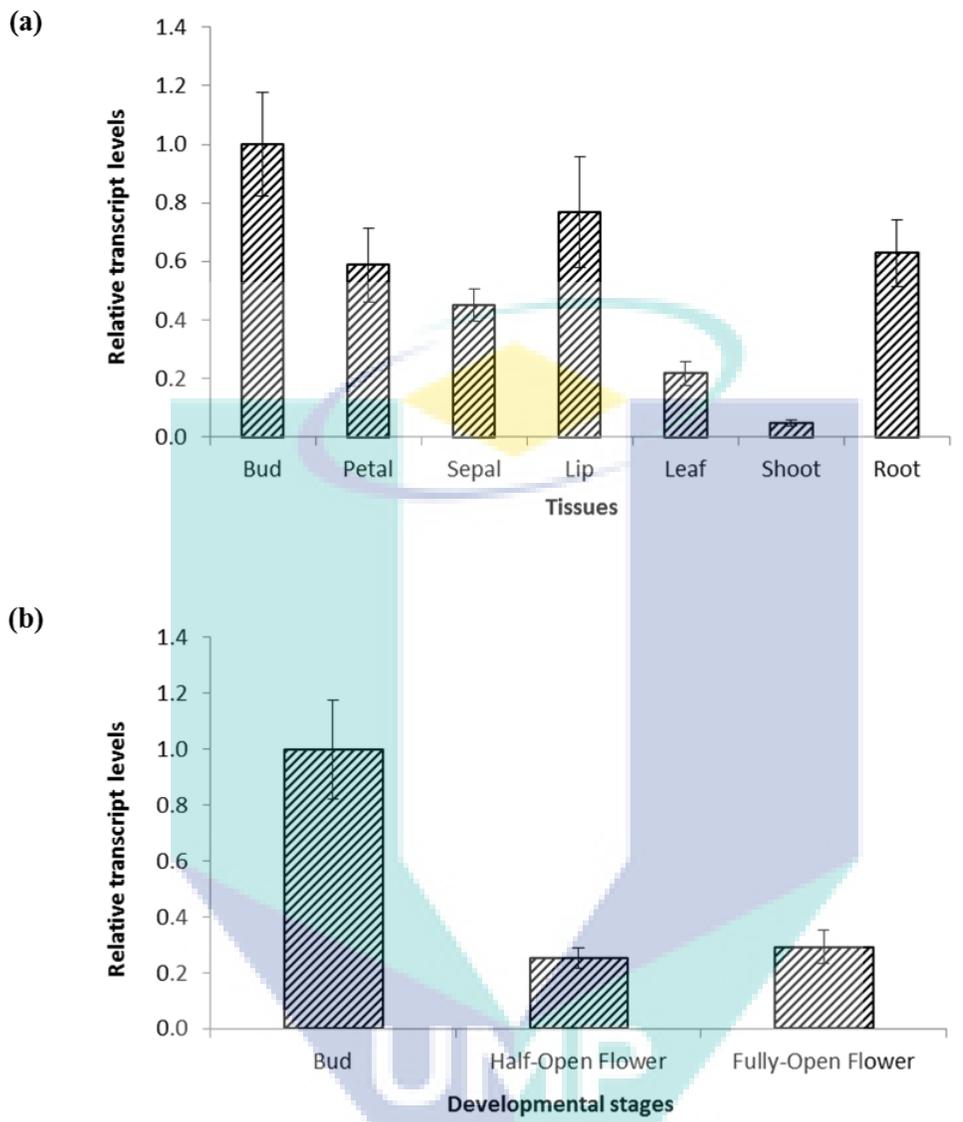


Figure 4.49: Relative VMPEGS transcript levels in *Vanda Mimi Palmer* detected in (a) different tissues; (b) different flower developmental stages

On the contrary, recent studies on three *Gymnadenia* species including *Gymnadenia conopsea*, *Gymnadenia densiflora* and *Gymnadenia odoratissima* has shown a high expression pattern of their eugenol synthases in floral tissues (Gupta *et al.*, 2014). Interestingly, expression pattern of their eugenol synthases agrees to the emission of eugenol compound by the flowers. Besides that, preliminary work on volatile analysis on more than 100 orchids including species and hybrids by Kaiser (1993) has shown the presence of eugenol as one of the scent constituents that are emitted by their floral organs in order to attract pollinators. However, in *V. Mimi Palmer*, no eugenol compound has been detected to be emitted by the flower (Mohd-Hairul *et al.*, 2010a). Thus, it is expected that the expression pattern of VMPEGS transcript is totally different compared to those *Gymnadenia* orchids, reflecting that VMPEGS has more tendency to be involved in biosynthesis of non-fragrant properties due to the fact that VMPEGS is highly expressed in non-fragrant tissues including bud, lip and root compared to petal and sepals of *V. Mimi Palmer* that have been identified as specific tissues for fragrance biosynthesis, accumulation and emission.

4.4.4 Relative Expression Analysis of *Vanda Mimi Palmer* Orcinol O-methyltransferase (VMPOOMT)

Expression analysis of VMPOOMT gene in different tissues of *V. Mimi Palmer* (Figure 4.50(a)) has shown that the gene was expressed at the highest level in shoot followed by leaf. In contrast, gene expression level of VMPOOMT has been detected to be much lower in floral tissues including petal, sepal and lip in comparison to leaf and shoot. Meanwhile, among the floral tissues, expression of VMPOOMT has been identified to be three to five times higher in lip, petal and sepal in comparison to bud as calibrator. Thus, the comparative gene expression level of VMPOOMT in different tissues might indicate that the gene might be not totally involved in fragrance biosynthesis in floral tissues of this orchid but also contribute to other metabolisms in vegetative tissues. In addition, expression pattern of the gene contradicted with the other fragrance-related genes that were previously reported on *V. Mimi Palmer* including phenylacetaldehyde synthase

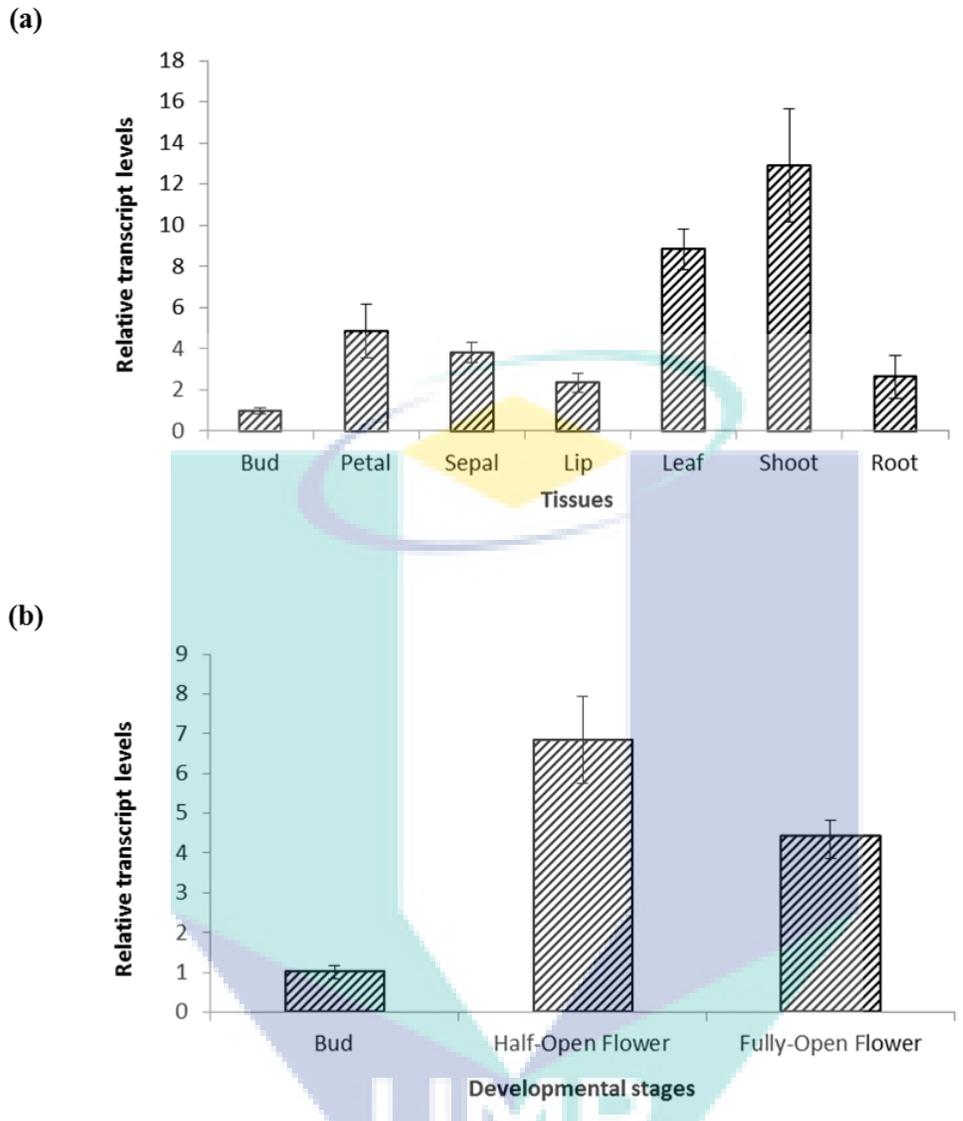


Figure 4.50: Relative VMPOOMT transcript levels in *Vanda Mimi Palmer* detected in (a) different tissues; (b) different flower developmental stages

(VMPPAAS), sesquiterpene synthase (VMPSTS) and alcohol acyltransferase (VMPPAAT) whereby all the fragrance-related genes have been detected to be expressed at much higher level in floral tissues especially in petal and sepal compared to vegetative tissues (Mohd-Hairul *et al.*, 2010b; Chan *et al.*, 2011).

In *R. hybrida*, one of the well-studied scented flower, its orcinol O-methyltransferase (RhOOMET) has been reported to be involved in methylation of orcinol to DMT (3,5-dimethoxytoluene) (Lavid *et al.*, 2002; Scalliet *et al.*, 2002). However, in *V. Mimi Palmer*, the involvement of VMPOOMET in methylation of orcinol to DMT (3,5-dimethoxytoluene) still remains unclear due to the fact that none of DMT compound or its derivatives has been identified to be emitted by floral tissues of the orchid (Mohd-Hairul *et al.*, 2010a). Thus, the expression pattern of VMPOOMET was expected to be totally different compared to fragrance-related transcripts that have been previously reported on *V. Mimi Palmer* including sesquiterpene synthase (VMPSTS), alcohol acyltransferase (VMPPAAT) and phenylacetaldehyde synthase (VMPPAAS) (Chan *et al.*, 2011; Mohd-Hairul *et al.*, 2010b) since the involvement of VMPOOMET in fragrance or non-fragrance biosynthesis is still remaining unclear. However, docking analysis on the developed 3D-structural model of VMPOOMET protein with several potential substrates (Appendix Ij) that have been previously reported to plant O-methyltransferases showed that orcinol might be the best substrate, with higher specificity compared to other substrate candidates. Interestingly, relative expression analysis of VMPOOMET in different developmental stages of *V. Mimi Palmer* (Figure 4.50(b)) has shown very low level in bud stage and increased to almost 7-fold expression during half-open flower stage and decreased a little bit in fully-open flower stage. Thus, VMPOOMET might be involved in catalysing compound that is involved in progressive growth of the plants such as growth hormone or other signalling mechanisms for the plant growth. However, the exact functional catalytic role of VMPOOMET could be further investigated in future by *in vitro* enzymatic assay of its recombinant protein.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

In conclusion, all the three objectives of this study have been successfully achieved including screening and identification of fragrance-related transcripts of selected fragrant-orchids, isolation of four full length sequences of fragrance-related transcripts of *V. Mimi Palmer* and characterisation of the full length sequences using molecular and bioinformatics approaches. In the first objective, partial sequences of three fragrance-related transcripts that might be involved in benzenoid and phenylpropanoid pathway in fragrant orchids have been successfully isolated including phenylalanine ammonia lyase (VMPPAL) from *V. Mimi Palmer*, *S*-adenosyl-*L*-methionine synthase (VSBSAMS) from *V. Small Boy Leong* as well as methionine synthases from *V. Small Boy Leong* (VSBLMS) and *Vandachostylis Sri-Siam* (VSSSMS), respectively. However, in the second objective, all the three transcripts were not further selected for full length cDNA isolation due to their indirect involvement in the biosynthesis of final fragrance products of volatile benzenoid and phenylpropanoids that have previously been reported to be emitted by the fully-open flower of *V. Mimi Palmer* including methylbenzoate, benzyl acetate, phenylethyl acetate and phenylethyl alcohol. Thus, partial sequences of another four fragrance-related transcripts of *V. Mimi Palmer* that have been previously identified from cDNA library of *V. Mimi Palmer* including 3-ketoacyl-CoA thiolase (VMPKAT), benzoic acid carboxyl methyltransferase (VMPBSMT), eugenol synthase (VMPEGS) and orcinol *O*-methyltransferase (VMPOOMT) have been selected for full length cDNA isolation in the second objective and further characterised using molecular and bioinformatics approaches in the third objective of this study.

In this study, full length sequences of the four selected candidates of fragrance-related transcripts have been successfully isolated via RACE-PCR approach, they have shown the presence of fragrance-related domains including thiolase domain for VMPKAT, S-adenosyl-L-methionine binding domain for VMPBSMT, Nmr-A like family attachment site for VMPEGS and O-methyltransferase attachment site for VMPOOMT. Besides that, 3D-structural reference template models of fragrance-related proteins that are retrieved from Protein Data Bank (PDB) which are 3-keto acyl-CoA thiolase of *A. thaliana* (2wu9), salicylic acid carboxyl methyltransferase of *C. breweri* (1m6e), Eugenol synthase of *C. breweri* (3c1o) and isoflavone O-methyltransferase of *Medicago sativa* (1fp2) selected for the development of 3D protein structure of VMPKAT, VMPBSMT, VMPEGS and VMPOOMT, respectively. Subsequent molecular docking on the developed 3D-structural model of those fragrance-related proteins using Autodock software version 4.2 has shown suitability of substrate and ligand candidates for prediction of their catalytic functionality. Docking analysis of VMPKAT has shown that Coenzyme A as suitable substrate while 1,2-ethanediol is a suitable ligand for its catalytic reaction. Meanwhile, docking analysis of VMPBSMT has shown that both benzoic and salicylic acids are the potential substrates to be used in its catalytic reaction in the presence of S-adenosyl-L-methionine as its specific ligand. Docking analysis on both VMPEGS and VMPOOMT has shown that coniferyl acetate and orcinol are suitable substrate, respectively, to be used in their catalytic reaction as reported by *in vitro* enzymatic test that have been carried out on their homologous proteins from other well-studied scented flowers. Thus, eugenol as well as DMT might be possibly produced by *V. Mimi Palmer* in floral as well as vegetative tissues but might not be accumulated in the cells for fragrance purpose.

Besides that, expression analysis of VMPKAT and VMPBSMT genes by real-time RT-PCR has shown a similar expression pattern of the genes in different tissues whereby the expression of both genes have been detected to be up-regulated in floral tissues especially in petal and sepal, but down-regulated in vegetative tissues including leaf, shoot and root. In addition, expression of both genes have been identified to be developmentally-regulated whereby the level of the transcripts have been detected at very low level in bud stage, increase gradually to half-open flower stage and reaches the highest level during

fully-open flower stage. In contrast, the involvement of VMPEGS and VMPOOMT in benzenoid and phenylpropanoid pathway of *V. Mimi Palmer* is still far from understood since expression analysis results on those genes has shown totally different patterns compared to VMPKAT and VMPBSMT as well as other reported fragrance-related transcripts of *V. Mimi Palmer*. The expression analysis result on both VMPEGS and VMPOOMT is expected to be dissimilar compared to VMPKAT, VMPBSMT as well as other reported fragrance-related transcripts of *V. Mimi Palmer* due to the fact that eugenol as well as 3,5-dimethoxytoluene (DMT) were not detected in the scent of *V. Mimi Palmer*.

In fragrance aspect, VMPKAT and VMPBSMT might have a bright future prospect to be used in fragrance and flavour industry through genetic engineering by expressing the ORF of the transcripts into microbial systems including bacterial and yeasts for biosynthesis of respective fragrance benzenoid compounds. Successful expression of recombinants that produce the respective compounds *in vivo* might have potential to be grown in larger scale using fermentor or bioreactor to fulfil the high demand from fragrance, flavour and perfumery industries. Besides that, the knowledge on benzenoid and phenylpropanoid pathway that lead to biosynthesis of fragrance compounds in orchids could be utilised for orchid breeding in horticultural industry. Selection of the correct parents for orchids with strong fragrance leads to production of new orchid hybrids with scent characteristic that will have potential to be commercialised as potted plants and cut-flowers in floricultural industry. Orchids with additional fragrance are normally received high demand from orchid enthusiasts with higher price that can increase the profit of Malaysia's floriculture industry.

REFERENCES

- Akhtar, T.A. and Pichersky, E. 2013. Veratrole biosynthesis in *Silene latifolia* (white campion). *Plant Physiol.* 162: 52-62.
- Allikas, G. 2009. Epiphyte or Terrestrial? Sympodial or Monopodial? (Online). Available at: http://www.aos.org/images/img_content/newsletter_issues/aug09.html
- Alvarez-Pérez, S. and Herrera, C.M. 2013. Composition, richness and non-random assembly of culturable bacterial–microfungal communities in floral nectar of Mediterranean plants. *FEMS Microbiol. Ecol.* 83: 685–699.
- Alvarez, J.C., Rodriguez, H.A., Rodriguez-Arango, E., Monsalve, Z.I., Morales, J.G. and Arango, R.E. 2013. Characterization of a differentially expressed phenylalanine ammonia-lyase gene from banana induced during *Mycosphaerella fijiensis* infection. *J. Plant Stud.* 2: 35-46.
- American Orchid Society, 2013. (Online). <http://www.aos.org/Default.aspx?id=207>
- Andrews, E., Theis, N. and Adler, L. 2007. Pollinator and herbivore attraction to *Cucurbita* floral volatiles. *J. Chem. Ecol.* 33: 1682–1691.
- Attwood, J.T. 1986. The size of the Orchidaceae and the systematic distribution of epiphytic orchids. *Selbyana.* 9: 171–186.
- Audesirk, T., Audesirk, G. and Byers, B.E. 2002. *Biology Life On Earth*. New Jersey: Prentice-Hall Incorporation.
- Blom, N., Gammeltoft, S. and Brunak, S. 1999. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* 294(5): 1351-1362.
- Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S. and Brunak, S. 2004. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics Review.* 4(6): 1633-1649.
- Boatright, J., Negre, F., Chen, X., Kish, C.M, Wood, B., Peel, G., Orlova, I., Gang, D., Rhodes, D. and Dudareva, N. 2004. Understanding *in vivo* benzenoid metabolism in petunia petal tissue. *Plant Physiology.* 135: 1993-2011.
- Bowie, J.U., Lüthy, R. and Eisenberg, D. 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science.* 253(5016): 164-170.
- Buchmann, S. and Nabhan, G. 1996. *The Forgotten Pollinators*. Washington: Island Press.

- Calabrese, J.C., Jordan, D.B., Boodhoo, A., Sariaslani, S. and Vannelli, T. 2004. Crystal structure of phenylalanine ammonia-lyase: multiple helix dipoles implicated in catalysis. *Biochem.* 43: 11403–11416.
- Camm, El. and Towers, Ghn. 1973. Phenylalanine ammonia-lyase. *Photochemistry.* 12: 961- 973.
- Carrie, C., Murcha, M.W., Millar, A.H., Smith, S.M. and Whelan, J. 2007. 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(1): 97-108.
- Chaitanya, M. 2010. Exploring the molecular basis for selective binding of Mycobacterium tuberculosis Asp kinase toward its natural substrates and feedback inhibitors: a docking and molecular dynamics study. *J Mol Model.* 16(8): 1357–1367.
- Chan, W.S., Abdullah, J.O., Namasivayam, P. and Mahmood, M. 2009. Molecular characterization of a new 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR) transcript from *Vanda Mimi Palmer*. *Scientia Horticulturae.* 121: 378-382.
- Chan, W.S. 2010. *Identification, Cloning and Characterization of Selected Full-Length Fragrance-Related Transcripts from Orchid (Vanda Mimi Palmer)*. MSc. Thesis. Universiti Putra Malaysia, Serdang, Selangor, Malaysia.
- Chan, W.S., Janna, O.A. and Parameswari, N. 2011. Isolation, cloning and characterization of fragrance-related transcripts from *Vanda Mimi Palmer*. *Scientia Horticulturae.* 127: 388-397.
- Chang, Y.Y., Chu, Y.W., Chen, C.W., Leu, W.M. and Yang, C.H. 2011. Characterization of *Oncidium* ‘Gower Ramsey’ transcriptomes using 454 GS-FLX pyrosequencing and their application to the identification of genes associated with flowering time. *Plant Cell Physiol.* 52(9): 1532-1545.
- Chase, M.W., Cameron, K.M., Barrett, R.L. and Freudenstein, J.V. 2003. DNA data and *Orchidaceae* systematics: a new phylogenetic classification. In: Dixon, K.W., Kell, S.P., Barrett, R.L. and Cribb, P.J. eds. *Orchid conservation*. Kota Kinabalu: Natural History Publications (Borneo): 69–89.
- Cheam, M.C., Antony, E., Abdullah, E. and Perumal, B. 2009. *Wild Orchids of Cameron Highlands*. Malaysia: Regional Environmental Awareness Cameron Highlands.
- Chen, F., D’Auria, J.C., Tholl, D., Ross, J.R., Gershenzon, J., Noel, J.P. and Pichersky, E. 2003. An *Arabidopsis thaliana* gene for methylsalicylate biosynthesis, identified by biochemical genomic approach, has a role in defense. *Plant J.* 36: 577-588.

- Chen, F., Tholl, D., Bohlmann, J. and Pichersky, E. 2011. The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.* 66(1): 212-229.
- Cheng, X. and Blumenthal, R.M. 1999. *S-Adenosylmethionine-dependent methyltransferases: structures and functions*. World Scientific Publication Co.
- Chew, E. 2007. A celebration of heritage orchids. *Malayan Orchid Review.* 41: 45-47.
- Chew, E. 2008. *Vanda* Tan Chay Yan: golden heritage, glorious controversy. *Malayan Orchid Review.* 42: 54-57, 110-113.
- Chowdhury, M.A., Rahman, M.M., Chowdhury, M.R., Uddin, M.J., Sayeed, M.A. and Hossain, M.A. 2014. Antinociceptive and cytotoxic activities of an epiphytic medicinal orchid: *Vanda tessellata* Roxb. *BMC Complement Altern Med.* 3(14): 464-471.
- Claros, M.G. and Vincens, P. 1996. Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur. J. Biochem.* 241: 779-786.
- Colovos, C. and Yeates, T.O. 1993. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* 2(9): 1511-1519.
- Colquhoun, T.A., Verdonk, J.C., Schimmel, B.C., Tieman, D.M., Underwood, B.A. and Clark, D.G. 2010. *Petunia* floral volatile benzenoid/phenylpropanoid genes are regulated in a similar manner. *Phytochemistry.* 71: 158-167.
- Colquhoun, T.A. and Clark, D.G. 2011. Unraveling the regulation of floral fragrance biosynthesis. *Plant Signaling and Behavior.* 6: 378-381.
- Colquhoun, T.A., Marciniak, D.M., Wedde, A.E., Kim, J.Y., Schwieterman, M.L., Levin, L.A., Moerkercke, A.V., Schuurink, R.C. and Clark, D.G. 2012. A peroxisomally localized acyl-activating enzyme is required for volatile benzenoid formation in a *Petunia*×*hybrida* cv. 'Mitchell Diploid' flower. *J Exp Bot.* 63(13): 4821-4833.
- Cour, T.L., Kiemer, L., Mølgaard, A., Gupta, R., Skriver, K. and Brunak, S. 2004. Analysis and prediction of leucine rich nuclear export signals. *Protein Eng. Des. Sel.* 17(6): 527-36.
- Cozzolino, S. and Widmer A. 2005. Orchid diversity: an evolutionary consequence of deception? *Trends in Ecology and Evolution.* 20(9): 487-494.
- Cozzini, P., Fornabaio, M., Marabotti, A., Abraham, D.J., Kellogg, G.E. and Mozzarelli, A. 2004. Free energy of ligand binding to protein: evaluation of the contribution of water molecules by computational methods. *Curr Med Chem.* 11(23): 3093-3118.

- D'Auria, J.C., Chen, F. and Pichersky, E. 2002. Characterization of an acyltransferase capable for synthesizing benzyl benzoate and other volatile esters in flowers and damaged leaves of *Clarkia breweri*. *Plant Physiology*. 130: 466-476.
- D'Auria, J.C., Chen, F., Pichersky, E. and John, T.R. 2003. The SABATH family of MTS in *Arabidopsis thaliana* and other plant species. *In Recent Advances in Phytochemistry* (ed. J.T. Romeo): 253–283.
- D'Auria, J.C. 2006. Acyltransferases in plants: a good time to be BAHD. *Current Opinion in Plant Biology*. 9: 331–340.
- D'Auria, J.C., Pichersky, E., Schaub, A., Hansel, A. and Gershenzon, J. 2007. Characterization of a BAHD acyltransferase responsible for producing the green leaf volatile (Z)-3-hexen-1-yl acetate in *Arabidopsis thaliana*. *Plant J.* 49: 194–207.
- Da Silva, J.A., Chin, D.P., Van, P.T. and Mii, M. 2011. Transgenic orchids. *Scientia Horticulturae*. 130: 673- 680.
- Dalton, J.A. and Jackson, R.M. 2007. An evaluation of automated homology modelling methods at low target template sequence similarity. *Bioinformatics*. 23(15): 1901-1908.
- de Castro, E., Sigrist, C.J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., Bairoch, A. and Hulo, N. 2006. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* 34 (Web Server issue): W362-365.
- Dexter, R., Qualley, A., Kish, C.M., Ma, C.J., Koeduka, T., Nagegowda, D.A. and Clark, D. 2007. Characterization of a petunia acetyltransferase involved in the biosynthesis of the floral volatile isoeugenol. *The Plant Journal*. 49: 265–275.
- Dharmani, A.D. 2013. Orchid Growth Types: Epiphytes, Litophytes, Terrestrial. (Online). Available at: <http://www.bellaonline.com/articles/art33604.asp>
- Dobson, H.E.M. 1994. Floral volatiles in insect biology. In E.A. Bernays. *Insect Plant Interactions*. 5: 47-81. Boca Raton: CRC Press.
- Dobson, H.E.M., Groth, I. and Bergstrom, G. 1996. Pollen advertisement: chemical contrasts between whole flower and pollen odors. *Am J Bot.* 83: 877–885.
- Dressler, R.L. 1990. *The orchids—natural history and classification*. Harvard University Press, London.
- Dressler, R.L. 1993. *Phylogeny and classification of the orchid family*. Oregon: Dioscorides Press.

- Drozdetskiy, A., Cole, C., Procter, J. and Barton, G.J. 2015. JPred4: a protein secondary structure prediction server. *Nucl. Acids Res.* 43(W1): W389-W394.
- Duckert, P., Brunak, S. and Blom, N. 2004. Prediction of proprotein convertase cleavage sites. *Protein Engineering, Design and Selection.* 17: 107- 112.
- Dudareva, N., Cseke, L., Blanc, V. and Pichersky, E. 1996. Evolution of floral scent in *Clarkia*: novel patterns of S-linalool synthase gene expression in the *C. breweri* flower. *Plant Cell.* 8: 1137-1148.
- Dudareva, N., D'Auria, J.C., Nam, K.H., Raguso, R.A. and Pichersky, E. 1998. Acetyl-CoA: benzylalcohol acetyltransferase – an enzyme involved in floral scent production in *Clarkia breweri*. *Plant Journal.* 14: 297-304.
- Dudareva, N., Murfitt, L.M., Mann, C.J., Gorestein, N., Kolosova, N., Kish, C.M., Bonham, C. and Wood, K. 2000. Developmental regulation of methyl benzoate biosynthesis and emission in snapdragon flowers. *Plant Cell.* 12: 949-961.
- Dudareva, N., Martin, D., Kish, C.M., Kolosova, N., Gorestein, N., Faldt, J., Miller, B. and Bohlman, J. 2003. (*E*)- β -ocimene and myrcene synthase genes of floral scent biosynthesis in snapdragon: Functional and expression of three terpene synthase genes of a new terpene synthase subfamily. *Plant Cell.* 15: 1227-1241.
- Dudareva, N., Pichersky, E. and Gershenzon, J. 2004. Biochemistry of plant volatiles. *Plant Physiol.* 135: 1893–1902.
- Dudareva, N., Klempien, A., Muhlemann, J.K. and Kaplan, I. 2013. Biosynthesis, function and metabolic engineering of plant volatile organic compounds. *New Phytologist.* 198: 16–32.
- Effmert, U., Buss, D., Rohrberk, D. and Piechulla, B. 2006. Localization of the synthesis and emission of scent compounds within the flower. In N. Dudareva and E. Pichersky. *Biology of Floral Scent*: 105-124. Florida: CRC Press.
- Efron, B., Halloran, E. and Holmes, S. 1996. Bootstrap confidence levels for phylogenetic trees. *Proc Natl Acad Sci U S A.* 93(23): 13429-13434.
- Eltz, T., Roubik, D.W. and Whitten, M.W. 2003. Fragrances, male display and mating behaviour of *Euglossa hemichlora*: a flight cage experiment. *Physiological Entomology.* 28: 251–260.
- Emanuelsson, O., Nielsen, H. and von Heijne, G. 1999. ChloroP, A neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* 8(5): 978-984.

- Emanuelsson, O., Nielsen, H., Brunak, S. and Heijne, G.V. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300: 1005-1016.
- Erich, B.B., Eric, R. and Martin, V. 1998. Computational Approaches to Identify Leucine Zippers. *Nucleic Acids Res.* 26(11): 2740-2746.
- Eswar, N., Webb, B., Marti-Renom, M., Madhusudhan, M.S., Eramian, D., Shen, M.Y., Pieper, U. and Sali, A. 2007. Comparative protein structure modeling using MODELLER. *Curr protoc protein sci.* 2(1): 1–30.
- Fadelah, A.A., Zaharah, H., Rozaily, Z., Nuraini, I., Tan, S.L. and Hamidah, S. 2001. *Orchids The Living Jewels of Malaysia*. Kuala Lumpur: Malaysia Agricultural Research and Development Institute.
- Farhi, M., Lavie, O., Masci, T., Hendel-Rahmanim, K., Weiss, D., Abeliovich, H. and Vainstein, A. 2010. Identification of rose phenylacetaldehyde synthase by functional complementation in yeast. *Plant Molecular Biology.* 72: 235-245.
- Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution.* 39:783-791.
- Ferrer, J.L., Zubieta, C., Dixon, R. A. and Noel, J. P. 2005. Crystal structures of alfalfa caffeoyl coenzyme A 3-O-methyltransferase. *Plant Physiol.* 137(3): 1009–1017.
- Flament, I., Debonneville, C. and Furrer, A. 1993. Volatile constituents of roses: characterization of cultivars based on the headspace analysis of living flower emissions. *Bioactive Volatile Compounds from Plants*. American Chemical Society, Washington, DC. 269–281.
- Gallego, S.M., Pena, L.B., Barcia, R.A., Azpilicueta, C.E., Lannone, M.F., Rosales, E.P., Zawoznik, M.S., Groppa, M.D. and Benavides, M.P. 2012. Unravelling cadmium toxicity and tolerance in plants: Insight into regulatory mechanisms. *Environmental and Experimental Botany.* 83: 33–46.
- Gang, D.R., Kasahara, H. and Xia, Z.Q. 1999. Evolution of plant defense mechanisms. Relationships of phenylcoumaran benzylic ether reductases to pinoresinol-lariciresinol and isoflavone reductases. *J Biol Chem.* 274: 7516–7527.
- Gang, D.R., Wang, J., Dudareva, N., Kyoung, H.N., Simon, J.E., Lewinsohn, E. and Pichersky, E. 2001. An investigation of the storage and biosynthesis of phenylpropenes in sweet basil. *Plant Physiology.* 125: 539-555.
- Gang, D.R., Lavid, N., Zubieta, C., Chen, F., Beuerle, T., Lewinsohn, E., Noel, J.P. and Pichersky, E. 2002. Characterization of phenylpropene O-methyltransferases from sweet basil: facile change of substrate specificity and convergent evolution within a plant O-methyltransferase family. *Plant Cell.* 14: 505–519.

- Gao, Z.M., Wang, X.C., Peng, Z.H., Zheng, B. and Liu, Q. 2012. Characterization and primary functional analysis of phenylalanine ammonia-lyase gene from *Phyllostachys edulis*. *Plant Cell Reports*. 31: 1345–1356.
- Go, R., Yong, W.S.Y., Unggang, J. and Salleh, R. 2010. *Orchids of Perlis, Jewels in the Forests, Revised Edition*. Perlis Forestry Department and Universiti Putra Malaysia Press, Kuala Lumpur, Malaysia: 152.
- Goh, C.J. and Kavaljian, L.G. 1989. Orchid industry of Singapore. *Economic Botany*. 43: 241-254.
- Gonda, I., Bar, E., Portnoy, V., Lev, S., Burger, J., Schaffer, A.A., Tadmor, Y., Gepstein, S., Giovannoni, J.J. and Katzir, N. 2010. Branched-chain and aromatic amino acid catabolism into aroma volatiles in *Cucumis melo* L. Fruit. *Journal of Experimental Botany*. 61: 1111–1123.
- Gough, J. and Chothia, C. 2002. SUPERFAMILY: HMMs representing all proteins of known structure. SCOP sequence searches, alignments and genome assignments. *Nucleic Acids Res*. 30(1): 268-272.
- Graciet, E., O Maoileidigh, D.S. and Wellmer, F. 2014. Next-generation sequencing applied to flower development (ChIP-Seq). *Methods in Molecular Biology*. 1110: 413-429.
- Gravendeel, B., Smithson, A., Slik, F.J.W. and Schuiteman, A. 2004. Epiphytism and pollinator specialization: drivers for orchid diversity? *Philos Trans R Soc Lond B Biol Sci*. 359: 1523–1535.
- Gupta, A.K., Akhtar, T.A., Widmer, A., Pichersky, E. and Schiestl, F.P. 2012. Identification of white campion (*Silene latifolia*) guaiacol O-methyltransferase involved in the Biosynthesis of veratrole, a key volatile for pollinator attraction. *BMC Plant Biol*. 12: 158.
- Gutensohn, M., Orlova, I., Nguyen, T.T., Davidovich-Rikanati, R., Ferruzzi, M.G., Sitrit, Y., Lewinsohn, E., Pichersky, E. and Dudareva, N. 2013. Cytosolic monoterpene biosynthesis is supported by plastid-generated geranyl diphosphate substrate in transgenic tomato fruits. *Plant J*. 75(3): 351-363.
- Guterman, I., Monshe, S., Naama, M., Dan, P., Mery, D.Y., Gil, S., Einat, B., Olga, D., Mariana, O., Michal, E., Jihong, W., Zach, A., Pichersky, E., Efraim, L., Dani, Z., Vainstein, A. and David, W. 2002. Rose Scent: Genomic approach to discovering novel floral fragrance-related genes. *The Plant Cell*. 14: 2325-2338.
- Hamdan, O. 2008. *Variasi Dunia Orkid : Panduan Penanaman dan Penjagaan*. Shah Alam: Alaf 21 Sdn. Bhd.

- Hands, G. 2006. *Pocket Guide to Orchids*. England: D&S Books Ltd
- Hirata, H., Ohnishi, T., Ishida, H., Tomida, K., Sakai, M., Hara, M. and Watanabe, N. 2012. Functional characterization of aromatic amino acid aminotransferase involved in 2-phenylethanol biosynthesis in isolated rose petal protoplasts. *J Plant Physiol.* 169(5): 444–451.
- Hossain, M.M. 2011. Therapeutic orchids: traditional uses and recent advances-An overview. *Fitoterapia.* 82: 102-140.
- Hsiao, Y.Y., Tsai, W.C., Huang, T.H., Wang, H.C., Wu, T.S., Leu, Y.L., Chen, W. H. and Chen, H.H. 2006. Comparison of transcripts in *Phalaenopsis bellina* and *Phalaenopsis equestris* (Orchidaceae) flowers to deduce monoterpene biosynthesis pathway. *BMC Plant Biology.* 6: 14.
- Hsiao, Y.Y., Jeng, M.F., Tsai, W.C., Chuang, Y.C., Li, C.Y., Kuoh, C.S., Chen, W.H. and Chen, H.H. 2008. A novel homodimeric geranyl diphosphate synthase from the orchid *Phalaenopsis bellina* lacking a DD (X)2-4D motif. *The Plant Journal.* 55: 719-733.
- Hsiao, Y.Y., Pan, Z.J., Hsu, C.C., Yang, Y.P., Hsu, Y.C., Chuang, Y.C., Shih, H.H., Chen, W.H., Tsai, W.C. and Chen, H.H. 2011. Research on Orchid Biology and Biotechnology. *Plant and Cell Physiology.* 52(9): 1467-1486.
- Hsu, C.C., Chung, Y.L., Chen, T.C., Lee, Y.L., Kuo, Y.T., Tsai, W.C., Hsiao, Y.Y., Chen, Y.W., Wu, W.L. and Chen, H.H. 2011. An overview of the *Phalaenopsis* orchid genome through BAC end sequence analysis. *BMC Plant Biology.* 11: 3-13.
- Ibrahim, R.K., Bruneau, A., and Bantignies, B. 1998. Plant O-methyltransferases: Molecular analysis, common signature and classification. *Plant Mol. Biol.* 36: 1–10.
- iPSORT server. Available from <http://ipsort.hgc.jp/>
- Jae-Gyeong, Yu., Gi-Ho, Lee. and Young-Doo, Park. 2012. Physiological role of endogenous S-adenosyl-L-methionine synthetase in Chinese cabbage. *Horticulture, Environment, and Biotechnology.* 53(3): 247-255.
- Jin, Q., Yao, Y., Cai, Y. and Lin, Y. 2013. Molecular cloning and sequence analysis of a phenylalanine ammonia-lyase gene from *Dendrobium*. *PLoS One.* 8: e62352.
- Jones, D.T. 1999. pGenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J Mol Biol.* 287(4): 797–815.
- Kaiser, R. 1993. *The scent of orchids: Olfactory and chemical investigations*. Amsterdam: Elsevier Science Publishers B.V.

- Kaminaga, Y., Schnepf, J., Peel, G., Kish, C.M., Gili, B.N., Weiss, D., Orlova, I., Lavie, O., Rhodes, D., Wood, K., Porterfield, D.M., Cooper, A.J.L., Schloss, J.V., Pichersky, E., Veinstein, A. and Dudareva, N. 2006. Plant phenylacetaldehyde synthesis is a dysfunctional homotetrameric enzyme that catalyses phenylalanine decarboxylation and oxidation. *Journal of Biological Chemistry*. 281: 23357-23366.
- Kato, M., Mizuno, K., Crozier, A., Fujimura, T. and Ashihara, H. 2000. Caffeine synthase gene from tea leaves. *Nature*. 406: 956-957.
- Kirby, J., Romanini, D.W., Paradise, E.M. and Keasling, J.D. 2008. Engineering triterpene production in *Saccharomyces cerevisiae*-beta-amyrin-synthase from *Artemisia annua*. *FEBS J*. 275(8): 1852-1859.
- Kirby, J. and Keasling, J.D. 2009. Biosynthesis of plant isoprenoids: perspectives for microbial engineering. *Annu Rev Plant Biol*. 60: 335-355.
- Kishor, R., ShaValli Khan, P.S. and Sharma, G.J. 2006. Hybridization and in vitro culture of an orchid hybrid *Ascocenda 'Kangla'*. *Scientia Horticulturae*. 108: 66-73.
- Kishor, R., Devi, H., Jeyaram, K. and Singh, M. 2008. Molecular characterization of reciprocal crosses of *Aerides vandarum* and *Vanda stangeana* (Orchidaceae) at the protocorm stage. *Plant Biotechnology Reports*. 2: 145-152.
- Klempien, A., Kaminaga, Y., Qualley, A., Nagegowda, D.A., Widhalm, J.R., Orlova, I., Shasany, A.K., Taguchi, G., Kish, C.M. and Cooper, B.R. 2012. Contribution of CoA ligases to benzenoid biosynthesis in petunia flowers. *Plant Cell*. 24: 2015-2030.
- Knudsen, J.T., Tollesten, L. and Bergstrom, G.L. 1993. Floral scents-A checklist of volatile compounds isolated by headspace techniques. *Phytochemistry*. 33: 253-280.
- Knudsen, J.T. and Gershenzon, J. 2006. The chemical diversity of floral scent. In N. Dudareva and E. Pichersky. In N. Dudareva and E. Pichersky. *Biology of Floral Scent* (pp. 105-124). Florida: CRC Press.
- Koay, S.H. 1993. Overview of the Singapore orchid industry. *Malayan Orchid Rev*. 27: 73-75.
- Koeduka, T., Fridman, E., Gang, D.R., Vassão, D.G., Jackson, B.L., Kish, C.M., Orlova, I., Spassova, S.M., Lewis, N.G. and Noel, J.P. 2006. Eugenol and isoeugenol, characteristic aromatic constituents of spices, are biosynthesized via reduction of a coniferyl alcohol ester. *Proc Natl Acad Sci USA*. 103: 10128-10133.
- Koeduka, T., Louie, G.V., Orlova, I., Kish, C.M., Ibdah, M., Wilkerson, C.G., Bowman, M.E., Baiga, T.J., Noel, J.P. and Dudareva, N. 2008. The multiple phenylpropene synthases in both *Clarkia breweri* and *Petunia hybrida* represent two distinct protein lineages. *Plant J*. 54: 362-374.

- Koeduka, T., Suzuki, S., Iijima, Y., Ohnishi, T., Suzuki, H., Watanabe, B., Shibata, D., Umezawa, T., Pichersky, E. and Hiratake, J. 2013. Enhancement of production of eugenol and its glycosides in transgenic aspen plants via genetic engineering. *Biochem Biophys Res Commun.* 436(1): 73-78.
- Kolosova, N., Gorenstein, N., Kish, C.M. and Dudareva, N. 2001. Regulation of circadian methyl benzoate emission in diurnally and nocturnally emitting plants. *Plant Cell.* 13: 2333–2347.
- Kumar, P.K.S., Subramoniam, A. and Pushpangadan, P. 2000. Aphrodisiac activity of *Vanda tessellata* (Roxb.) Hook. ex Don extract in male mice. *Ind J Pharmacol.* 32: 300–304.
- Kursula, P., Sikkilä, H., Fukao, T., Kondo, N. and Wierenga, R.K. 2005. High resolution crystal structures of human cytosolic thiolase (CT): a comparison of the active sites of human CT, bacterial thiolase, and bacterial KAS I. *J. Mol. Biol.* 347: 189–201.
- Laskowski, R.A. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr.* 26(2): 283–291.
- Lavid, N., Wang, J., Shalit, M., Guterman, I., Bar, E., Beuerle, T., Menda, N., Shafir, S., Zamir, D., Adam, Z., Vainstein, A., Weiss, D., Pichersky, E. and Lewinsohn, E. 2002. O-methyltransferase involved in biosynthesis of volatile phenolic derivatives in rose petals. *Plant Physiology.* 129: 1899-1907.
- Lee, S., Lee, B., Jang, I., Kim, S. and Bhak, J. 2006. Localizome: a server for identifying transmembrane topologies and TM of eukaryotic proteins utilizing domain information. *Nucleic Acids Research.* 34: 99-103.
- Lekawatana, S. 2010. *Thai Orchid: Current Situation. In: Taiwan International Orchid Symposium.* International Conference Hall, Taiwan Orchid Plantation, Tainan, Taiwan.
- Liang, X., Dron, M., Cramer, C.L., Dixon, R.A. and Lamb, C.J. 1989. Differential regulation of phenylalanine ammonia-lyase genes during plant development and environmental cues. *J. Bid. Chem.* 264: 14486-14492.
- Lisitsyn, N. and Wigler, M. 1999. Representational difference analysis in detection of genetic lesions in cancer. *Methods Enzymol.* 254: 291–304.
- Liu, Z.J., Chen, L.J., Liu, K.W., Li, L.Q. and Ma, X.Y. 2008. *Chenorchis*: A new orchid genus and its eco-strategy of ant pollination. *Acta Ecologica Sin.* 28(6): 2433–2444.
- Luthy, R., Bowie, J.U. and Eisenberg, D. 1992. Assessment of protein models with three-dimensional profiles. *Nature.* 356(6364): 83–85.

- Ma, Z., Yang, L., Yan, H., Kennedy, J.F. and Meng, X. 2013. Chitosan and oligochitosan enhance the resistance of peach fruit to brown rot. *Carbohydrate Polymers*. 94: 272-277.
- Mathieu, M., Modis, Y., Zeelen, J.P., Engel, C.K., Abagyan, R.A., Ahlberg, A., Rasmussen, B., Lamzin, V.S., Kunau, W.H. and Wierenga, R.K. 1997. The 1.8 Å crystal structure of the dimeric peroxisomal 3-ketoacyl-CoA thiolase of *Saccharomyces cerevisiae*: implications for substrate binding and reaction mechanism. *J Mol Biol*. 273(3): 714-728.
- McCarthy, A.A. and McCarthy, J.G. 2007. The Structure of Two N-Methyltransferases from the Caffeine Biosynthetic Pathway. *Plant Physiol*. 144(2): 879–889.
- McDonald, E. 1999. *100 Orchids for the American Gardener*. Workingman Publishing, New York.
- Melo, F., Sanchez, R. and Sali, A. 2002. Statistical potentials for fold assessment. *Protein Sci*. 11: 430–448.
- Mohd-Hairul, A.R., Parameswari, N., Gwendoline Ee, C.L. and Janna O.A. 2010(a). Terpenoid, benzenoid and phenylpropanoid compounds in the floral scent of *Vanda Mimi Palmer*. *Journal of Plant Biology*. 53: 358-366.
- Mohd-Hairul, A.R., Parameswari, N., Janna O.A. and Gwendoline Ee, C.L. 2010(b). Screening, isolation, and molecular characterization of putative fragrance-related transcripts from *Vanda Mimi Palmer*. *Acta Physiologiae Plantarum*. 33(5): 1651-1660.
- Mohd-Hairul, A.R. 2011. Chemical composition of floral volatiles and expression of scent related genes in *Vanda Mimi Palmer*. Masters dissertation. Universiti Putra Malaysia, Selangor, Malaysia.
- Muhlemann, J.K., Klempien, A. and Dudareva, N. 2014. Floral volatiles: from biosynthesis to function. *Plant, Cell and Environment*. 37: 1936–1949.
- Murfitt, L.M., Kolosova, N., Mann, C.J. and Dudareva, N. 2000. Purification and characterization of S-adenosyl-L-methionine:benzoic acid carboxyl methyltransferase, the enzyme responsible for biosynthesis of the volatile ester methyl benzoate in flowers of *Anthirrhinum majus*. *Arch Biochem Biophys*. 382: 145–151.
- Nagegowda, D.A., Gutensohn, M., Wilkerson, C.G. and Dudareva N. 2008. Two nearly identical terpene synthases catalyze the formation of nerolidol and linalool in snapdragon flowers. *The Plant Journal*. 55: 224-239.

- Nair, H. and Arditti, J. 2002. Award and trophies. *Proceedings of the 17th World Orchid Conference "Sustaining Orchids for the Future"*. Sabah: Natural History Publications (Borneo) Sdn. Bhd.
- Negre, F., Kish, C.M., Boatright, J., Underwood, B., Shibuya, K., Wagner, C., Clark, D.G. and Dudareva, N. 2003. Regulation of methylbenzoate emission after pollination in snapdragon and petunia flowers. *Plant Cell*. 15: 2992–3006.
- Ng, H.G. 1984. Commercial orchid growing in Singapore . *Malayan Orchid Review*. 18: 41-45.
- Noel, J.P., Dixon, R.A., Pichersky, E., Zubieta, C. and Ferrer, J.L. 2003. Structural, functional, and evolutionary basis for methylation of plant small molecules. *Recent Adv Phytochem*. 37: 37–58.
- Obeng-Ofori, D. and Reichmuth, C. 1997. Bioactivity of eugenol, a major component of essential oil of *Ocimum suave* (Wild.) against four species of stored-product *Coleoptera*. *Int J Pest Manage*. 43: 89–94.
- Ooi, J.M. 2005. Screening of a random peptide library with CymMV for potential development of diagnostic kits. Master dissertation. Malaysia University of Science and Technology, Selangor Darul Ehsan, Malaysia.
- Orlova, I., Marshall-Colon, A. and Schnepf, J. 2006. Reduction of benzenoid synthesis in petunia flowers reveals multiple pathways to benzoic acid and enhancement in auxin transport. *The Plant Cell*. 18: 3458–3475.
- Petersen, T.N., Brunak, S., Heijne, G.V. and Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*. 8: 785-786.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*. 25(13):1605-1612.
- Pichersky, E., Raguso, R.A., Lewinsohn, E. and Croteau, R. 1994. Floral scent production in *Clarkia* (Onagraceae) I. Localization and developmental modulation of monoterpenes emission and linalool synthase activity. *Plant Physiology*. 126: 1533-1540.
- Pichersky, E., Lewinsohn, E. and Croteau, R. 1995. Purification and characterization of S-linalool synthase, and enzyme involved in the production of floral scent in *Clarkia breweri*. *Archives of Biochemistry and Biophysics*. 306: 803-807.
- Pichersky, E. and Dudareva, N. 2007. Scent engineering: toward the goal of controlling how flowers smell. *Trends in Biotechnology*. 25: 105-110.

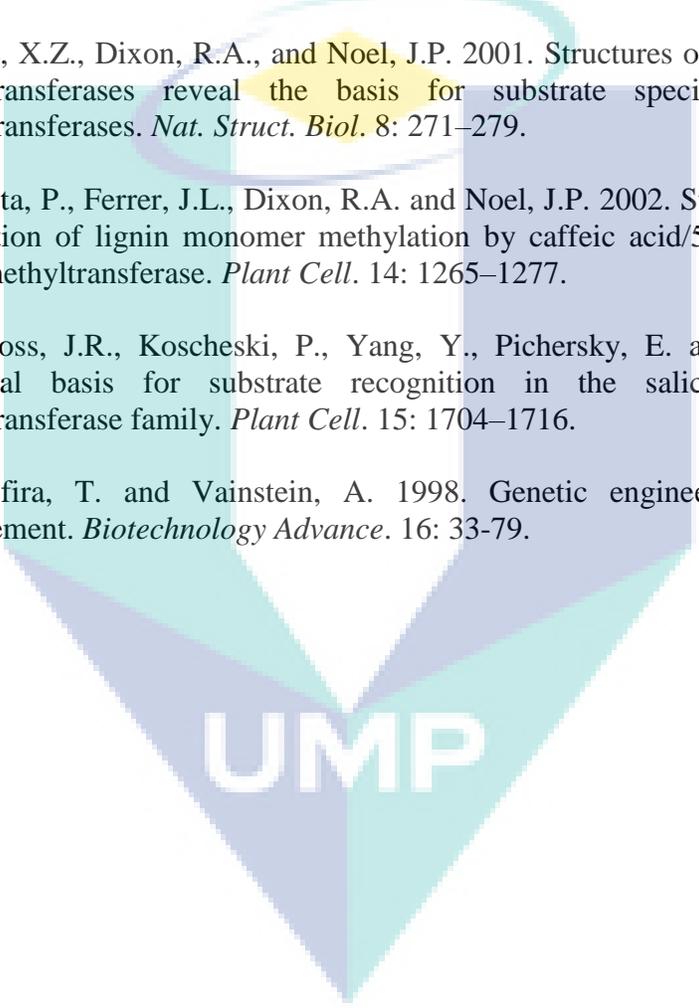
- Pott, M.B., Pichersky, E. and Piechulla, B. 2002. Evening-specific oscillation of scent emission, SAMT enzyme activity, and mRNA in flowers of *Stephanotis floribunda*. *Journal of Plant Physiology*. 159: 925-934.
- Pott, M., Hippauf, F., Saschenbrecker, S., Chen, F., Ross, J., Kiefer, I., Slusarenko, A., Noel, J.P., Pichersky, E., Effmert, U. and Piechulla, B. 2004. Biochemical and structural characterization of benzenoid carboxyl methyltransferases involved in floral scent production in *Stephanotis floribunda* and *Nicotiana suaveolens*. *Plant Physiol*. 135: 1946–1955.
- Pye, V.E., Christensen, C.E., Dyer, J.H., Arent, S. and Henriksen, A. 2010. Peroxisomal Plant 3-Ketoacyl-CoA Thiolase Structure and Activity Are Regulated by a Sensitive Redox Switch. *J Biol Chem*. 285(31): 24078 –24088.
- Qualley, A.V., Widhalm, J.R., Adebessin, F., Kish, C.M. and Dudareva, N. 2012. Completion of the core β -oxidative pathway of benzoic acid biosynthesis in plants. *Proceedings of the National Academy of Sciences of the United States of America*. 109: 16383–16388.
- Raguso, R.A., Levin, R.A., Foose, S.E., Holmberg, M.W. and McDade, L.A. 2003. Fragrance chemistry, nocturnal rhythms and pollination “syndromes” in *Nicotiana*. *Phytochemistry*. 63: 265-284.
- Raguso, R. 2008. Wake up and smell the roses: The ecology and evolution of floral scent. *Annual Review of Ecology, Evolution and Systematics*. 39: 549-569.
- Ramirez, S.R., Gravendeel, B., Singer, R.B., Marshall, C.R. and Pierce, N.E. 2007. Dating the origin of the Orchidaceae from a fossil orchid with its pollinator. *Nature*. 448: 1042–1045.
- Raubeson, L.A. and Jansen, R.K. 2005. Chloroplast genomes of plants. In: Henry, R.J., editor. *Plant diversity and evolution: genotypic and phenotypic variation in higher plants*. Cambridge: CAB International: 45–68.
- Roeder, S., Dreschler, K., Wirtz, M., Cristescu, S.M., Harren, F.J.V., Hell, R. and Piechulla, B. 2009. SAM levels, gene expression of SAM synthetase, methionine synthase and ACC oxidase, and ethylene emission from *N. suaveolens* flowers. *Plant Mol Biol*. 70: 535-546.
- Ross, J.R., Nam, K.H., D’Auria J.C. and Pichersky, E. 1999. S-Adenosyl-Lmethionine: salicylic acid carboxyl methyltransferase, an enzyme involved in floral scent production and plant defense, represents a new class of plant methyltransferases. *Archives of Biochemistry and Biophysics*. 367: 9–16.
- Saet-Byul, Kim., Jae-Gyeong, Yu., Gi-Ho, Lee. and Young-Doo, Park. 2012. Characterization of *Brassica rapa* S-adenosyl-L-methionine synthetase gene

- including its roles in biosynthesis pathway. *Horticulture, Environment, and Biotechnology*. 53(1): 57-65.
- Sakai, M., Hirata, H., Sayama, H., Sekiguchi, K., Itano, H., Asai, T. and Watanabe, N. 2007. Production of 2-phenylethanol in roses as the dominant floral scent compound from L-phenylalanine by two key enzymes, a PLP-dependent decarboxylase and a phenylacetaldehyde reductase. *Bioscience, Biotechnology, and Biochemistry*. 71: 2408–2419.
- Salzmann, C., Brown, A. and Schiestl, F. 2007. Floral scent in food-deceptive orchids: Species specificity and sources of variability. *Plant Biol*. 9: 720–729.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. *Molecular Cloning: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory Press.
- Scalliet, G., Journot, N., Jullien, F., Baudino, S., Magnard, J.L., Channeliere, S., Vergne, P., Dumas, C., Bendahmane, M., Cock, J.M. and Huguene, P. 2002. Biosynthesis of the major scent components 3,5-dimethoxytoluene and 1,3,5-trimethoxybenzene by novel rose O-methyltransferases. *FEBS Lett*. 523(1-3):113-118.
- Scalliet, G., Piola, F., Douady, C.J., Réty, S., Raymond, O., Baudino, S., Bordji, K., Bendahmane, M., Dumas, C., Cock, J.M. and Huguene, P. 2008. Scent evolution in Chinese roses. *Proc Natl Acad Sci*. 105: 5927–5932.
- Schade, F., Legge, R.L. and Thompson, J.E. 2001. Fragrance volatiles of developing and senescing carnation flowers. *Phytochemistry*. 56: 703-710.
- Schaller, F. 2001. Enzymes of the biosynthesis of octadecanoid-derived signaling molecules. *J Exp Bot*. 52: 11–23.
- Schiestl, F.P. 2010. The evolution of floral scent and insect chemical communication. *Ecology Letters*. 13: 643-656.
- Scopece, G., Musacchio, A., Widmer, A. and Cozzolino, S. 2007. Patterns of reproductive isolation in Mediterranean deceptive orchids. *Evolution*. 11: 2623–2642.
- Shen, M.Y. and Sali, A. 2006. Statistical potential for assessment and prediction of protein structures. *Protein Sci*. 15(11): 2507-2524.
- Shuttleworth, A. and Johnson, S. 2009. A key role for floral scent in a wasp-pollination system in *Eucomis* (Hyacinthaceae). *Annals of Botany*. 103: 715-725.
- Schuurink R.C., Haring, M.A. and Clark, D.G. 2006. Regulation of volatile benzenoid biosynthesis in petunia flowers. *Trends Plant Sci*. 11: 20–25.

- Song, A.A.L., Janna, O.A., Mohd, P., Norazizah, S., Raha, A.R. and Puad, A. 2012. Functional expression of an orchid fragrance gene in *Lactococcus lactis*. *International Journal of Molecular Sciences*. 13: 1582-1597.
- Sousa, D. and Grigorieff, N. 2007. Ab initio resolution measurement for single particle structures. *J Struct Biol*. 157(1): 201-210.
- Spitcer, B., Ben Zvi, M., Ovadis, M., Marhevla, E., Barkai, O., Edelbaum, O., Marton, I., Masci, T., Alon, M., Morin, S., Rogachev, I., Aharoni, A. and Vainstein, A. 2007. Reverse Genetics of floral scent: application of tobacco rattle virus-based gene silencing in petunia. *Plant Physiology*. 145: 1241-1250.
- Sundaramoorthy, R., Micossi, E., Alphey, M.S., Germain, V., Bryce, J.H., Smith, S.M., Leonard, G.A. and Hunter, W.N. 2006. The crystal structure of a plant 3-ketoacyl-CoA thiolase reveals the potential for redox control of peroxisomal fatty acid beta-oxidation. *J Mol Biol*. 359(2): 347-357.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30: 2725–2729.
- Teh, S.L., Chan, W.S., Janna, O.A. and Parameswari, N. 2011. Development of expressed sequence tag resources for *Vanda Mimi Palmer* and data mining for EST-SSR. *Molecular Biology Reports*. 38: 3903-3909.
- Teh, S.L., Janna, O.A., Parameswari, N. and Go, R. 2012. Discovering Fragrance Biosynthesis Genes from *Vanda Mimi Palmer* Using the Expressed Sequence Tag (EST) Approach. *Intech*. 9: 205-224.
- Tieman, D., Taylor, M., Schauer, N., Fernie, A.R., Hanson, A.D. and Klee, H.J. 2006. Tomato aromatic acid decarboxylases participate in synthesis of the flavor volatiles 2-phenylethanol and 2-phenylacetaldehyde. *Proc. Natl Acad. Sci. U S A*. 103: 8287–8292.
- Tieman, D.M., Lucas, H.M., Kim, J.Y., Clark, D.G. and Klee, H.J. 2007. Tomato phenylacetaldehyde reductase catalyze the last step in the synthesis of the aroma volatile 2-phenylethanol. *Phytochemistry*. 68: 2660–2669.
- TMHMM server v.2.0. Available from <http://www.cbs.dtu.dk/services/TMHMM/>
- Topik, H., Yukawa, T. and Ito, M. 2005. Molecular phylogenetics of subtribe *Aeridinae* (*Orchidaceae*): insights from plastid matK and nuclear ribosomal ITS sequences. *J Plant Res*. 118(4): 271–284.
- Vainstein, A., Lewinsohn, E., Pichersky, E. and Weiss, D. 2001. Floral fragrance. New Inroads into an Old Community. *Plant Physiol*. 127: 1383-1389.

- Van Moerkercke, A., Schauvinhold, I., Pichersky, E., Haring, M.A. and Schuurink, R.C. 2009. A plant thiolase involved in benzoic acid biosynthesis and volatile benzenoid production. *The Plant Journal*. 60: 292-302.
- Vassão, D.G., Kim, S.J. and Milhollan, J.K. 2007. A pinoresinol-lariciresinol reductase homologue from the creosote bush (*Larrea tridentata*) catalyzes the efficient *in vitro* conversion of p-coumaryl/coniferyl alcohol esters into the allylphenols chavicol/eugenol, but not the propenylphenols p-anol/isoeugenol. *Arch Biochem Biophys*. 465: 209–218.
- Verdonk, J.C., de Vos, C.H.R., Verhoeven, H.A., Haring, M.A., Van Tunen, A.J. and Schuurink, R.C. 2003. Regulation of floral scent production in petunia revealed by targeted metabolomics. *Phytochemistry*. 62: 997-1008.
- Verdonk, J.C., Haring, M.A., van Tunen, A.J. and Schuurink, R.C. 2005. ODORANT1 regulates fragrance biosynthesis in petunia flowers. *The Plant Cell*. 17: 1612-1624.
- Vyas, V.K., Ukawala, R.D., Ghate, M. and Chintia, C. 2012. Homology modeling a fast tool for drug discovery: current perspectives. *Indian J. Pharm. Sci.* 74: 1–17.
- Wang, J. and Pichersky, E. 1998. Characterization of S-adenosyl-L-methionine: (iso)eugenol O-methyltransferase involved in floral scent production in *Clarkia breweri*. *Arch. Biochem. Biophys*. 349: 153-160.
- Wiederstein, M. and Sippl, M.J. 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acid Res.* 35(2):W407–W410
- Willmer, P.G., Nuttman, C.V., Raine, N.E., Stone, G.N., Patrick, J.G. and Henson, K. 2009. Floral volatiles controlling ant behaviour. *Funct. Ecol.* 23: 888–900.
- Winkelmann, T., Heintz, D., Van Dorsselaer, A., Serek, M. and Braun, H.P. 2006. Proteomic analyses of somatic and zygotic embryos of *Cyclamen persicum* Mill. reveal new insights into seed and germination physiology. *Planta*. 224: 508–519.
- Wu, S., Ueda, Y., He, H., Nishihara, S. and Matsumoto, S. 2000. Phylogenetic analysis of Japanese Rosa species using matK sequences. *Breed Sci.* 50: 275–281.
- Xiang, L. and Moore, B.S. 2002. Inactivation, complementation, and heterologous expression of encP, a novel bacterial phenylalanine ammonio-lyase gene. *Journal of Biology and Chemistry*. 277: 32505-32509.
- Xu, D. and Zhang, Y. 2011. Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-step Atomic-level Energy Minimization. *Biophys J.* 101: 2525-2534.

- Yeoh, B.C. 1978. In *Orchids: A Publication Commemorating The Golden Anniversary of The Orchid Society of South East Asia*, ed. Teoh, E.S. pp. 18-23. Singapore: Times Periodicals.
- Yu, H. and Goh, C.J. 2000. Identification and characterization of three orchid MADS-Box genes of the AP1/AGL9 subfamily during floral transition. *Plant Physiology*. 123: 1325-1336.
- Zhang, Y. and Skolnick, J. 2005. TM-align: A protein structure alignment algorithm based on TM-score. *Nucleic Acids Research*. 33: 2302-2309.
- Zubieta, C., He, X.Z., Dixon, R.A., and Noel, J.P. 2001. Structures of two natural product methyltransferases reveal the basis for substrate specificity in plant O-methyltransferases. *Nat. Struct. Biol.* 8: 271–279.
- Zubieta, C., Kota, P., Ferrer, J.L., Dixon, R.A. and Noel, J.P. 2002. Structural basis for the modulation of lignin monomer methylation by caffeic acid/5-hydroxyferulic acid 3/5-O methyltransferase. *Plant Cell*. 14: 1265–1277.
- Zubieta, C., Ross, J.R., Koscheski, P., Yang, Y., Pichersky, E. and Noel, J.P. 2003. Structural basis for substrate recognition in the salicylic acid carboxyl methyltransferase family. *Plant Cell*. 15: 1704–1716.
- Zuker, A., Tzfira, T. and Vainstein, A. 1998. Genetic engineering for cut-flower improvement. *Biotechnology Advance*. 16: 33-79.

The logo for UMP (Universiti Malaysia Perlis) is a large, stylized letter 'U' composed of four overlapping triangles in shades of teal and light blue. The letters 'UMP' are printed in white, bold, sans-serif font across the center of the 'U' shape.

UMP

APPENDIX A

GENERAL SOLUTIONS

TAE buffer

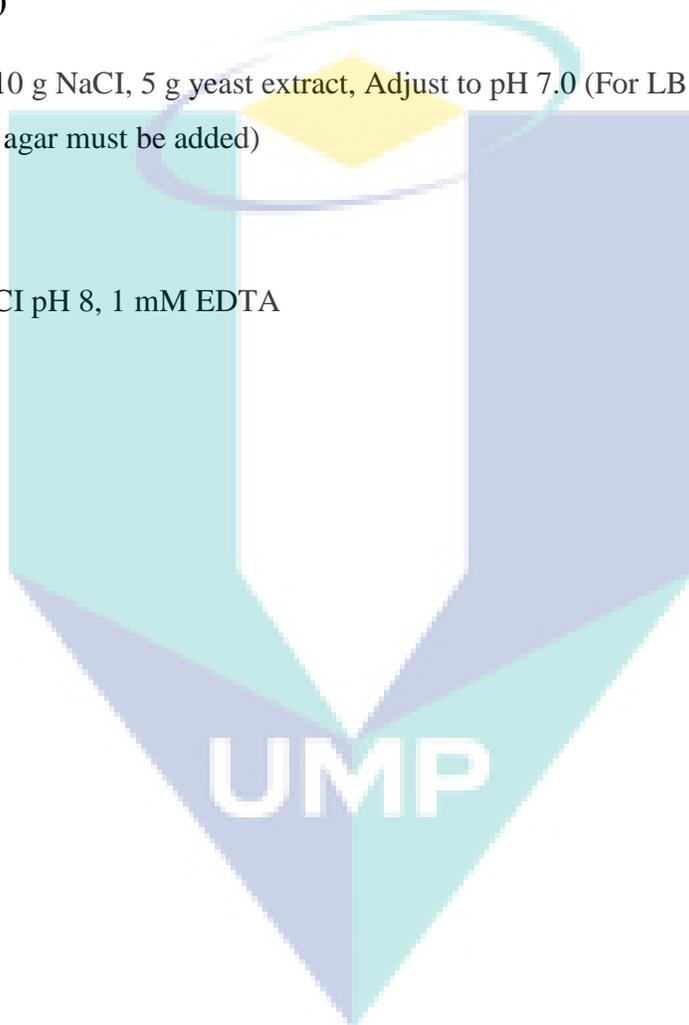
242 g Tris base, 57.1 ml glacial acetic acid, 100 ml 0.5 M EDTA pH 8

LB broth (1 L)

10 g tryptone, 10 g NaCl, 5 g yeast extract, Adjust to pH 7.0 (For LB agar, 20 g of bacteriological agar must be added)

TE buffer

10 mM Tris-HCl pH 8, 1 mM EDTA



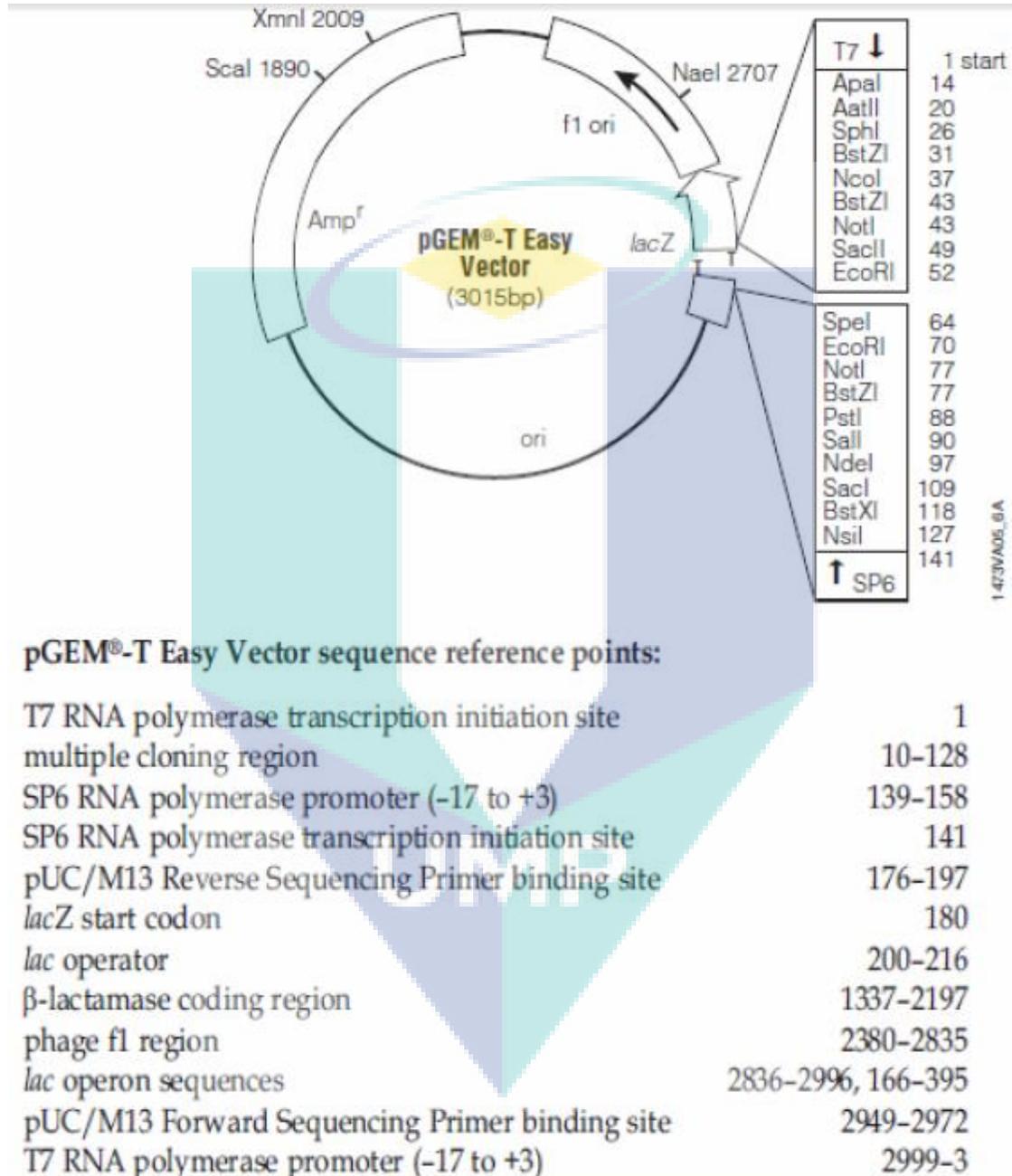
APPENDIX B

List of chemical components and its compositions:

- 1) 6X Loading Dye (Fermentas, Canada) - 10mM Tris-HCl (pH 7.6), 0.03% bromophenol blue, 0.03% xylene cyanol FF, 60% glycerol, 60mM EDTA Fermentas
- 2) 10X PCR reaction buffer (Fermentas, Canada) - 100mM Tris-HCl (pH 7.5 at 25 °C), 25mM MgCl₂, 1mM CaCl₂
- 3) Wizard™ SV Gel and PCR Clean Up System (Promega, USA)
- 4) pGEM®-T Easy Vector (Promega, USA)
- 2X Rapid Ligation Buffer (Promega, USA) – 60 mM Tris HCl (pH 7.8), 20 mM MgCl₂, 20 mM DTT, 2 mM ATP, 10% polyethylene glycol
- 5) 10X T4 DNA Ligase Buffer (Fermentas, Canada) - 400mM Tris-HCl, 100mM MgCl₂, 100mM DTT, 5mM ATP (pH 7.8 at 25 °C)

APPENDIX C

pGEM[®]-T Easy vector map and sequence reference points



APPENDIX D

List of selected full-length ORF sequences of *Vanda Mimi Palmer*

>VMPKAT ORF

ATGGGGGAGGGAAAGAGGGGAGGAAGAGAGAACGGGAGAGAGATGGAGAAAAGCCATCGACAGGCAGGGAGTACTTCGCC
GGCATCTTCAATCCTCTTCTTACGACCATCTACATCTACTATCGGCGTCTGTAGGTGCTGCGGGAGACAGTTCCGGCTACC
AGAAGAACCATGGCTTTGACGATGATGTTGTTATCGTGGCTGCTTGCCGGACTGCAATTGGCAAGGCCAATAGAGGAGGT
TTCAAGGATACACACGCAGAGGACCTTCTAGCTCCTTCTGTAAGGCTCTGTTGGAGAAGACTAAACTGAACCCATATGA
AGTTGGTGATATAGTTGTTGGTACTGTCTGCGCCTGGTTCCTCAAAGAGCTATTGAATGCAGGATGGCAGCCTTTATGC
TGGTTTCCCAGAGACCGTTCCTCTTAGAACTGTGAACCGACAATGCTCATCTGGTCTTCAGGCTGTTGCAGATGTTGCTGCT
GCTATCAAAGCAGGATACTATGGCATTGGTATTGGTCTGGTGGTGGAGTCTATGAGTTCAAACAAAAGTTGGATGGGAGGG
ACCACCTAATCAAAGGCCAAATCTCTTCAACAAGTCCAAAGATTGTTATCTACCCATGGGAATGACATCTGAAAAATGTTGC
AGAGCGGTATGGAATAACAAAGGAAGAACAAGACCAAGCTGCTGTGATATCTCATAAGCGTGCAGGCTGCTGCAATTGCTT
CTGGTAAATTTAAAGAAGAAATAGTACCCGTGACAACAAGATTACGGATCCAAAAAGTGAATTGAAAAAGGAGATTAC
GGTCTCTGTGGATGAGGGAGTTCCGCCGATACGTCCTTGTCCGGCCTTGCAAACTCAAACAGCATTAAAGGAGGGTG
GAAAGCACTACGGCAGGCAATGCTAGTCAAGTTAGTATGGTGGTGGGGCGTCTACTAATGAGGAGGGATGTTGCAATA
CAAAAAGGGCTTCCATTCTTGAATTTTCAGGAGCTTTGCTGCAGTTGGGGTTGATCCTGCTGTGATGGGCATTGGTCTT
GCTGTTGCAATACCAGCTGCAGTAAAAGCAGCTGGTCTTCAAATGAAAGATATTGACTTGTATGAAATAAATGAGGCTTTT
CGCTCGCAGTTGTATACTGCATCAAAGAGTTGGGGCTGTATCCAGAAAAAGTCAATGTTAATGGCGGTGCTATTGCTCTT
GGCCATCCATTGGGTGCAACGGGTGCACGCTGTGTACTTATTAATGAAATGAAACCGAGGGGCAATGACTGCGG
ATTTGGAATATTTCATGTGCATTGGTCTGGGATGGGAGCTGCGGCTGTTTTGAGAGAGGGCGATGTAGCTGATCGGTT
GAGTAATGTGCAGATTGCCATTCTAGAAATCTGCTATTGCGGATTTCTACTGTTAAGTAGAGCTCAGTTTTAAAGCACTT
AATTAATAATTAATATAGACTACCTTTGTTTTCCTTAATCAACAATAAATCAAAGCTTATGCTCCTTAATTTGCGCTGTGG
ATTGGTGATAATTTGGTTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>VMPBSMT ORF

TCTCTCTCTCTCTCTTCTACAACAGGGAAAGGTTACTGACCAATATAAATATTGTGCAGACGAGTCTTTCTAGAGGAG
ATTATAACCATGATTGCTGAACAGGTCTTGAATGACGGGAGGTAGCGGAGAGAAACAGCTATGCTGAAAACCTCAAAGCT
TCAAGAGAAGGCAGGATATGACGAAACATTAGGGGAAGAAGCTTAAAGAGAAATATGTAAAACTCTCCAGAAAAG
CTTGTGTGGCTGACTTGGGATGCTCCTTGGTCCCAACACATTACTTGTGCTCTCAGATCATCAATGCAGTGGATGAGT
ATTGCTCTAAATCGAGCCAACAAGGCTGAAATCCAATATTCTTTGTGATCTTCTGACAATGACTTTAACACTTTGTT
CCAATTATCGGAGCAATATCAGAAGAAAATTAGAGAAGAGAAAAGGGGATAGATATGAACCACTTATATTGTTGGATCGC
CTGGTTCGTTTTATAGAAGGTTATTTCTTCAAAAAGTGTGCAGCTTTTCCACTTTCATACAGCCTCATGTGGCTATCGCA
GGTACCTAAAGGGCTAAAAGATGAAGCCTTTGATCACTTCAATAGGACAGCCATATATATCAATAATGAAAGCTATTCTA
TCGTGTCGCAATTTGATGTTGAACAATTTAAAACAGATTTCTCGGATTTCTCAAATCCCCTTAAAGGAGCTTATTTCTGG
AGGGAGCATGATTTATCATTTTGGGAAGAAGTGACTTTCAATCTATTGGTGAAATGTTATATTTATGGGAGCTTCTTGCA
GAAGCGTAACTGCAATGGTTTCTGAGGGACTAGTGGAGGAAGATAAACTAAAAGACTTTAATTTACCATTTTATACGCC
ATTGATGGAAGAAGTGAGATCTATTATCAGTATGGAAGGATCAATTTTCACTTTAAACAAGCACAATCTTTGAGTCAAAATG
GGATCCTTTTGTGATGAATCACATGAAAATTTTGTGGAAGACAAAAGCTCGAGTAGTAGAAAATGTTGCAAAATACATGAGAG
CAGTGCTAGAACCCTAGTTACATCTTATTTGGAAGTATTTTGAAGATTTATCTCGGTTATGCAAAATAATTTGC
AAGGCACATGTATAATGGGAAGCCAAAGCACACTGTTTTCATTTCTTGAATTTAAAGGTGAAGAGCAAAACACTTC
GAGAATGCCAATAAATGCAATATATAAGTGAAGATGGTGCAGTGAAGCAATCTTAACACTTCAGAATAAGACAAAATAA
ATTTACTGTTCTCCGAGCAAGCAATGTCTTGATTGCTCAACTTTAAGAGAAGAATAAGTGGGAACATACAACCCTTACT
ATCCCATGTATCCATCATCTCAACTTGAATAAAAATTAATGTTTTCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAA

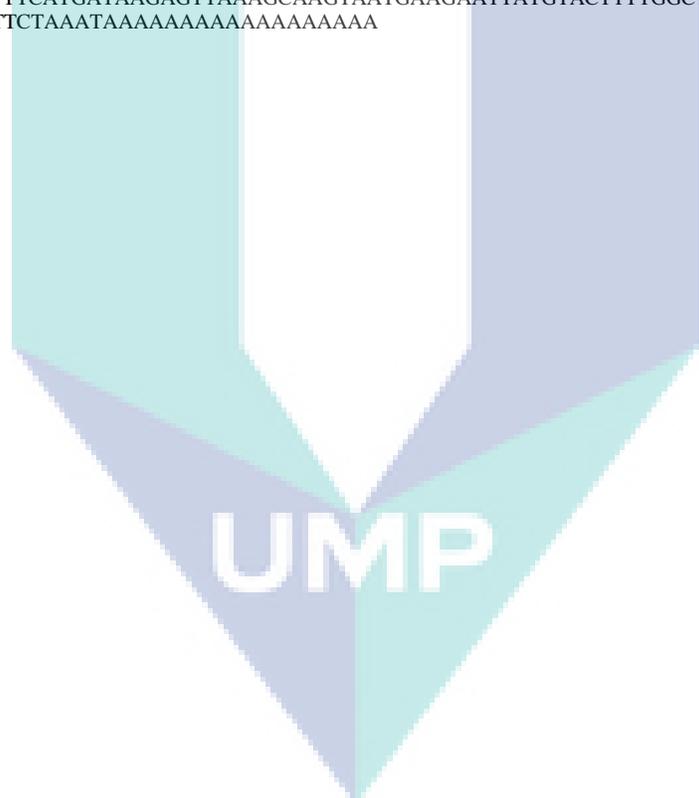
>VMPEGS ORF

ATGGGGAACTGACAGGAAGTGAACGGACCACAAACGGCGACCACAAACGGCGACCGGATGTTCTATAAAAAGGGGGG
GTTAGCGTATGCGTGCTTCGCAATATACGATCCATTGTCTCGGCGGATAAGCTTAAACACGCATTTTATCGTGCTCCATCTTC
TTACATTC AATGGCGTCCGAGAAGAGCAAGATCCATCATCGGAGGCACGGGATACATCGGGAAGCATGTCGTGGCTGC
GAGCGTCAAGTTGGGCCACCCTACCTTCGTTCTCGTAGGCCAAGCTCAGCATCCGACCCTCGGAAGGTTGGATTCTTCCA
GGCTTAAACGGGGCGGAAGTCAAAGTTGTTTATGGGATTTGATGACTATCAGAGCTTGTGAGTGAATCAAAGCAGG
TGGATATCGTCATCTCCACCGTGGAGGGCTACAGTTGGCAGATCAGGTTAACATCATCAAAGCCATTAAGAAGCAGGA
AACATCAAGAGGTTTCTGCCATCAGAGTTTGGGTTGGATCCGGAACCGGTGACTAAGATAGAACCAGGCAGCATCCGCTCT
CGCTGGCAAGTTGCAATCCGGCAAGCCATCAAGGCAAGGCAGAACCCCTACACGATCATCTTAATAACTACTTCGCGG
GTTACTCCCTCCCACTGGCTCAAGCCAGCCCTCCGTTGACAAAATCACCATCTTTGGTGAATGGAACACAAAAGGTG
TCTTTGTTGATGAAGATGATATTGGCATAATACCATCAAGGCGGCTAATGATCCAAGAACCCTAAAACAAGATTGTTTACA
TAAGGCCCCCTGGTTGCATCTACTCTCACAACGAGCTCATCTCTTTGGGAGAAGAAGACCGGCAAAACCCTAGAGAGG
ATTTATCTCTCGGATGAAGAAATTTTAAAGATTATCCAAGAAACCAATCCCTTTCAACGTAATATTGCTTTAAACTACT
TGGTGTGTTGAAAGGAGACTGTCTCAGCTTTGAGATTGACCATCCGTTGCTGCTGAAGCCACTGAGCTCTATCCCCAAG

TCAAATACACTACAGTTGAGGAATATCTCAGTCGTGTTCTGTAAATAAAGTTTACAACATCTCCTTCTCAAGTTTTTCAGA
AAAGTCCTATGGAAATTTCTATTATGTGTTGCTGTTTTGCTCATCTTTGCCATCCTTGTGTTGAAGTCTGTTTCTGTGATGTCT
TGACCTTCTGTCTGTTTTGTTGGACTGTTGATCTATGAGTATTGTGTATTTTGGGATCTTTCATGAGTTCTGATTTTTTATG
TGAAAAAAAAAAAAAAAAAAAA

>VMPOOMT ORF

CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACATGGGAACCAGAAGTAAATTAAGCATGACGACT
GCCGTGGAACAAGAAACCTAGAAAGCTCATGCCATTCTATGGAACCAAACCTTTAGCTACATCAAATCCATGTCCCTAAA
ATGCGCCGTTGAACTTACATCCCCGACGCCATCCACCGCCATGACCGGCCAACCACATTTCCGAACTCCTTTCCTCTCTC
TCCATCTCTCAGGACAAATCTCGCCACCTTCGCAACATCATGCGCGTCTGTCTCAAGAATGCATCTTTAAATCCCACATC
ACCCCTCTGATGAAGAAACCTTCGACCTCACTCTGTCTCCCGCCTCTTCTACCACCACCTCCTTCCCCACAACCATA
ACAGCTCTCCCTTTGTCTCCTAAGCCTCAATCGCCACCTCGTCGACTCGTTTCATCAACTCTCCCACTGGTTCCTTAAACC
GGATGCCACCATAACCCCTTTGCGATGGCAAACGGGAAGGAGTTCTGGAATCTGGCAAGGGAGCTACCACAATTTAGTG
ATTTGTTTAGTAAGGCAATGGAATGTGATTCCAACCTTTCTTATGAATGCTATGGTGAGTACTGGAGAGGGGCCATTTGGG
GGATTGAGACATCACTGGATGTTGGAGGTGGCACTGGCTTGCTGGCAACAACATTGGCGGAGGCATTTCTGGGTTAAAA
TGCATCGTGTGATCTACCACATGTCATPGGTATGGCAGAGAAGAAGACGGACGGGATTCAATTCGTCGCCGGTGATATG
TTTGTGAGATTCCACGTGCCGATGTTGCTCTGCTTAAGTGGATATTGCATGATTGGAGTGATGAAGATTGCGTCAAAAATA
CTACAACGTTGCAAGGAAGCAATTCCTTCTAAAGAGAAAAGGTGAAAAGTGATTATAATCGACATGGTGGTTGGTTTGA
CTCCAACCTCATACTTCATTGCAGACACAACCTTCTGTATGACGTGCAAAATGATGACTTTGTGCATGGGAAAAGAAAGGGA
TGAAAATGAATGGCGTAATCTATTTGTCGCGGCAGGTTATAGAGACTACAAGATAACACCATATGTTGATCACATGAGAT
CAATTATTGAAGTCTATCCTTGATTAATTATTATTGATTTCAAGTCTTATTATTAGTACATGTTATTGTGCTTATTATGATT
AGTTAACTTTTCTTTCATGATAAGAGTTAAAGCAAGTAATGAAGAATTATGTACTTTTGGCTATGGCTGAGCATTGTGAA
TGTAGCTATTATTTCTAAATAAAAAAAAAAAAAAAAAAAAA



APPENDIX E

Lists of protein sequences in *Vanda Mimi Palmer*

>VMPKAT

MEKAIDRQGVLRRLQSSSYDHPTSLSASVGAAGDSSAYQKNHGFDDDDVVIVAAC
RTAIGKANRGGFKDTHAEDLLAPLLKALLEKTKLNPYEVGDIVVGTVLAPGSQRAI
ECRMAAFYAGFPETVPLRTVNRQCSSGLQAVADVAAAIKAGYYGIGIGAGLESMS
SNKVGWEGPPNPKAKSLQQVQDCYLPMGMTSENAERYGITKEEQDQAAVISHK
RAAAAIASGKFKEEIVPVTTKITDPKSGIEKEITVSVDEGVRPDTSLSLAKLKPFAK
EGGSTTAGNASQVSDGAGAVLLMRRDVAIQKGLPILGIFRSFAAVGVDPAVMGIGP
AVAIPA AVKAAGLQIKDIDLYEINEAFASQFVYCIKELGLDPEK VNVNGGAIALGHP
LGATGARC VSTLLNEMKRRGND CRFGIISMCI GSGMGAAAVFERGDVADRLSNVQ
IAHS

>VMPBSMT

MIAEQVLAMTGGSGENSYAENSKLQEKAGIMTKPLGEEALREICKTLPEKLVVAD
LGCSSGPNTLLVLSQIINAVDEYCSKSSQQRPEIQYSLCDLPDNDNFNTLFLQSEYQ
KKIREEKGDREYFPYIVGSPGSFYRRLFPSKSVQLFHSSYSLMWLSQVPKGLKDEAF
DHFNRTAIYINNESYSIVSHLYVEQFKTDFSDFLKSRSELISGGSMILSFLGRSDFQS
IGEMLYLWELLAEALTAMVSEGLVEEDKCLKDFNLPFYTPLMEEVRSIISMEGSFHL
KQAQSFESNWDPFDESHENFVEDKSSSRNVANYMRAVLEPLVTSYFGTDILEDLF
SRYANNIARHMYNGKPKHTVFILFLKLGEEQTLRECE

>VMPEGS

MASEKSKILIIGGTGYIGKHVVAASVKLGHPTFVLVVRPSSASDPAKVGFLQGLTGAE
VKVVHGDLDNDYQSLLSAIKQVDIVISTVGGLQLADQVNIKAIKEAGNIKRFLPSEF
GLDPERVTKIEPAASALAGKVAIRQAIAEGIPYTIISNNYFAGYSLPTLAQASPPVD
KITIFGDGNTKGVFVDEDDIGIYTIKAANDPRTLKIVYIRPPGCIYSHNELISLWEK
KTGKTLERIYLSDEEIFKIIQEAPIPFNVIFALNYLVFVKGDCLSFEIDPSVAAEATEL
YPEVKYTTVEEYLSRVL

>VMPOOMT

MGTRSKLSMTTAVEQETLEAHAILWNQTFSYIKSMSLKCAVELHIPDAIHRHDRPT
TISELLSSLSISQDKSRHLRNIMRVLSQECIFKSHITPSDEETFDLTPVSRLLLTTTSFP
HNHNSSPFVLLSLNRHLVDSFHQLSHWFLKPDATITPFAMANGKEFWNLARELPQF
SDLFSKAMECDSNFLMNAMVSTGEGPFWGIETSLDVGGGTGLLATTLAFAFPGLK
CIVFDLPHVIGMAEKKTDGIQFVAGDMFVEIPRADVALLKWILHDWSEDCVKILQ
RCKEAIPSKEKGGKVIIDMVVGLNSNSHTSLQTQLLYDVQMMTLCMGKERDENE
WRNLFVAAGYRDYKITPYVDHMRSIIEVYP

APPENDIX F

List of VMP primers for fragrance-related transcripts used in Rapid Amplification of cDNA ends (RACE).

a) 5'-VMP KAT

5'-CACAGCAGGATCAACCCCAACTGCAGC-3'

b) 3'-VMP KAT

5'-CCATGTGCATTGGTTCTGGGATGGGAGC-3'

c) 3'-VMP BSMT

5'-CCACTCTTCATACAGCCTCATGTGGCT-3'

d) 5'-VMP EGS1

5'-GAGGGCGAAGGTGGGATGTCCAGTGCG-3'

e) 5'-VMP EGS2

5'-CCAACCCAAACTCTGATGGCAGAAACCTC-3'

f) 5'-VMP OOMT

5'-GAGACAGGACGCGCATGATGCTGCGAAGG-3'

g) Oligo dT primer

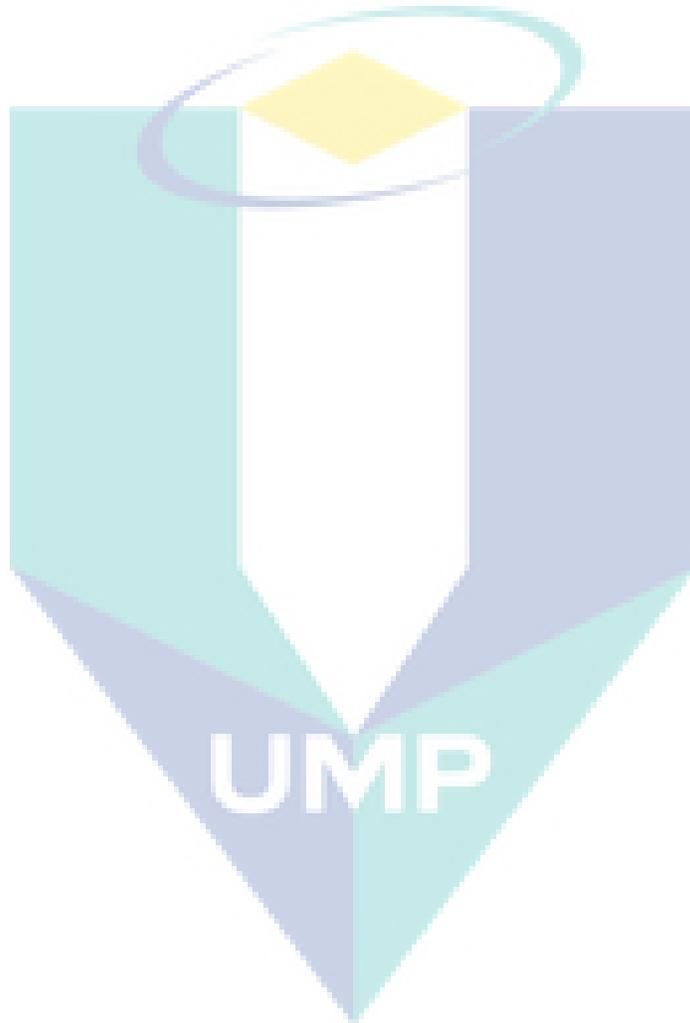
5'-TTTTTTTTTTTTTTT-3'

h) UPM long primer

5'-CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT-3'

i) UPM short primer

5'-CTAATACGACTCACTATAGGGC-3'



APPENDIX G

List of ESTs in cDNA-Representational difference analysis

>gi|537538021|gb|JZ480886.1|JZ480886 RDA001h Floral cDNA library of *Vanda Mimi Palmer Vanda* hybrid cultivar cDNA clone RDA001 similar to putative transposase, mRNA sequence

```
AGATCTCGTGCACCCAGGAGAGATTCCAACAGCTGAAGGAGAGTCGATCCCACGCGTCAGGAGGTAATT
ACGATACACGAGCGTGATCCTTCCGAGTACGACCTTTGGAGGGAGGCTTCTGGAGGGAAGAATAACGGTC
GAGTCTATGGGCCAGGTCTCAGAGGTTTGCTTACTACGCTTACCCCTCCACCAGTGCACCTGTAGCTTC
TATCTCTACCGTACAGGATGAGGAGATCTTCCCTCGGATAGCACAGTTGCCTTG
```

>gi|537538025|gb|JZ480890.1|JZ480890 RDA002h Floral cDNA library of *Vandachostylis Sri-Siam x Vandachostylis* hybrid cultivar cDNA clone RDA002 similar to auxin-repressed protein, mRNA sequence

```
AAGCACTCCCCAGCCTCTCACCGCAGATCTGATCCGAGCGCCTTCGTGTTAATATTGCTTCCAGGGTGA
AGACGCTTCGCCAAACATTAGCTTTCTTCGCCGTAGAAGTAGCCGGAGACACGGGCGTCACCGGCGTGAT
CGGCGTCGTCGGCGTCGCTGGCAACGACAACGACCGCTGGTAAACCTTCCCGCTGCTCTCCCCTCACCA
CCTGGGGGGTTGATGACTAAGGGTTTGGAGGAGAT
```

>gi|537538026|gb|JZ480891.1|JZ480891 RDA003h Floral cDNA library of *Vandachostylis Sri-Siam x Vandachostylis* hybrid cultivar cDNA clone RDA003, mRNA sequence

```
ACACAATCATTTACCTTGAACCTCAACTTTCGGCCAACAACCAATTTTCCACAACCTTGATGTCCTTCACTC
CAACAACCTAACACAAGGTGAATTCACCATCAATAATCATCATACATGGACTAATTTCCATGAATTCCAA
GAAGAACATCAAGAACATTCATTCCTCCATTTTCGGTCAAGCTAGGGTTTTCCACTTGAACATAATCAATG
CTCAATTTACTCAAATTAACAACATTTGAGCACTCATAATCCTTGTTTCATGGAAGCAAATACCGTGTATT
CAAGGAAATCAACAAGTTTTACCTTAGAACCTAGCTCAACGAACTCGGACATCCTTTGAATTTGAAGAA
CAAACCTTGAATCCAAGCCAACAATGGTCTAGGTTTGTCCGCCAAG
```

>gi|537538027|gb|JZ480892.1|JZ480892 RDA004h Floral cDNA library of *Vandachostylis Sri-Siam x Vandachostylis* hybrid cultivar cDNA clone RDA004 similar to methionine synthase, mRNA sequence

```
AGATCTTCAACTTCCAGCTTGATGGCGAGGGCAATCTGGTAGCATGTTTCAAACCTTGGTTGGTCATTGC
GGGCAAAGGACCAGTTAAGGATTGTAACAGGGCCGGTAAGCATTCCTCCATCGGGCGGGAGGTCATGCT
CTGGGCCAAGGTGGACCAGAAGACAGTCATGGCCTTGGGGCGACTCACATCCCCATATATGATCGGGGGC
TTTACACATCGTGAGCCATAAGACTGGACCAACCATTAACAGTGAAAGCAAAACCAGAGAGCTGCTCCC
CGAAATACTCCACCATATCATTCTCTCAGGTTACCATGGACAAGGACATCGATGTCAAACCTCCTCCTG
GATCTTGACAACCTTGTGATTTCCTCCTTGATTGCATGGATATAAGCTTCTCCGAGAT
```

>gi|537538028|gb|JZ480893.1|JZ480893 RDA005h Floral cDNA library of *Vanda Small Boy Leong Vanda* hybrid cultivar cDNA clone RDA005 similar to S-adenosyl methionine synthetase, mRNA sequence

AAGGCAACTGTGCTATCCGAGGGAAGATCGTCCGGACAATCTTCTCATAGTCAACGGTGGCCTTAGTAGT
GATCTCACCAAAGACCATGACCATATTGGTCTTGCTGCAGGTCTCACAGGCAACCTTGCTGTCGGGATCC
TGAGCGAGGCAAGCATCGAGAATGGCATCTGA

>gi|537538029|gb|JZ480894.1|JZ480894 RDA006h Floral cDNA library of *Vanda Small Boy Leong Vanda* hybrid cultivar cDNA clone RDA006 similar to putative gag-pol polyprotein, mRNA sequence

AAGCTTCAACTTCCCCACCATGGTTGAACTTGCTACATTGGCGCAACTTCCTCCATCAACAATGAGGCTG
CACACTTGATCTTAACTCGGCATCTAGTGTGAAAGATGCACTCCCTTTGATTGGCCTCCTCCTTGGCTT
GCACACTTAGAGCTCTCCTAGTAACAAAAGCAAAGTCTCCTTCCCTCGGGCTCAAGTGTGTCTTGCTCCTC
CTCCAAAGGTGGCGTCTCCTCCATGGCATCCTCCTC

>gi|537538030|gb|JZ480895.1|JZ480895 RDA007h Floral cDNA library of *Vanda Small Boy Leong Vanda* hybrid cultivar cDNA clone RDA007 similar to methionine synthase, mRNA sequence

AGGCAACTGTGCTATCCGAGGGAAGATCTTCGACTTCTGCTTGATGGCGAGGGCAATCTGGTAGCATGT
TTCAAACCTTGGTTGGTCATTGCGGACGAAGGACCAGTTAAGGATCGTAACAGGGCCGGTAAGCATTCCC
TTCATCGGGCGGGAGGTCATGCGCTGGGCCAAGGTGGACCAGAAGACAGTCATGGCCTTGGGGCGACTCA
CATCCCCATATATGATCGGGGGCTTTACACATCGTGAGCCATAAGACTGGACCCAACCATTGACAGTGAA
AGCAAAACCAGAGAGCTGCTCCCCGAAATACTCCACCATATCATTCTCTCAGGTTACCATGGACAAGA
ACATCGATGTCAAACCTCCTCGGATCTTAACAACCTTGTTGATTCTCCTTGGATTGCATGGATATAAG
CTTCTCCGAGAT

>gi|537538022|gb|JZ480887.1|JZ480887 RDA008h Floral cDNA library of *Vanda Mimi Palmer Vanda* hybrid cultivar cDNA clone RDA008 similar to phenylalanine ammonia-lyase, mRNA sequence

GGATCCGAAAATTCCCGCATTAAAGGAATCTGATGAGCTCTTCTGAAGGGCGCCGCCCGCTTAGTGCGA
CGGTGAGAGGTGGCTCCGAAGCCAGTGGTAACGCCGTAGTAATCGCCGCCATGATTCGTGTTATCCAAAA
CCCAGTCGCTGCTGGCCTTCACGCCGGCACGCGCCGACTCCGCAAGCTCAACGGCGGTAAGTGCAGCCGGA
GGCCACTGCAGCGACCTGAGAAATACTTAGTCCGCTCCCCGAGCAATACCAGGGGTTTCTGAAACCCC
TCCACCATTCTTACCTCCTCGAGGTGGCTCCCTCCGACTTCTTCGCCGCCGCGACCCATCCCAGTG
GATTC

>gi|537538023|gb|JZ480888.1|JZ480888 RDA009h Floral cDNA library of *Vanda Mimi Palmer Vanda* hybrid cultivar cDNA clone RDA009, mRNA sequence

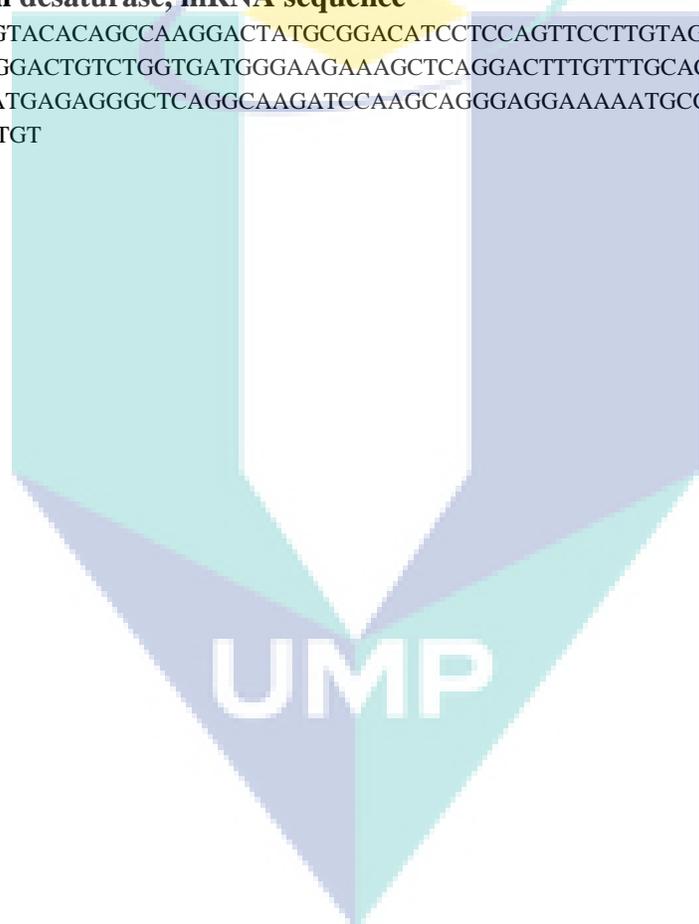
GATATCCGTTGGGAAGATAGCAAATACTATGATCCGCATATTCAGGAAGCACAGGAGTTACAGAACGCTC
CGATAGTGGCATCTGAGCCTTAGGTGCTTCAACTCCATTTGAAGATCCATACTGAATTCCAGAAGTCTTC
TTGGAAACCTCATGACCTTCATGCGTTTTGTTCTTGGATTCCATTGACAACCTTCTGAAGCAATTCTGTTG
CTTGATCTGCAGCCGAAGCAACAGCCGCCATGGTTGGATCC

>gi|537538031|gb|JZ480896.1|JZ480896 RDA010h Floral cDNA library of *Vanda* Small Boy Leong *Vanda* hybrid cultivar cDNA clone RDA010 similar to clathrin assembly protein, mRNA sequence

```
GGATCCGCCGCATAAAGGGATTTCGGAGCTGCAACATCGTGATACTGCTGCTGCTGTTGTTGTTGTCGAT
ATGCAGCTTCATCATAACAAGCTATCCAAAGTGAGTTTGTGCGAACCCGCCGCCAACTGGCTTTCAATGGC
AGATGAATGAGCAGCATTAACGAGGGCCAACTCCCATCCTGATAGATCCCACGATGGCTCATCTTGAGCA
CGAACGACACCGGTGAGAAAGGCAACTTCGCCACCGCCAAGGGGGCGCCTCGGGGGCTGGGCTTTTCCCGA
TGCCGGATCTCACCGAGATCCTCGGTGAGAGGCTGGAGAGTGCTC
```

>gi|537538024|gb|JZ480889.1|JZ480889 RDA011h Floral cDNA library of *Vanda* Mimi Palmer *Vanda* hybrid cultivar cDNA clone RDA011 similar to stearyl-acyl carrier protein desaturase, mRNA sequence

```
AAGCTTGGGGTGTACACAGCCAAGGACTATGCGGACATCCTCCAGTTCCTTGTAGCTAGGTGGAAGGTGG
AAGAACCTACGGGACTGTCTGGTGATGGGAAGAAAGCTCAGGACTTTGTTTGC ACTCTGGTGCCTAGAAT
TCGTAGGCTGGATGAGAGGGCTCAGGCAAGATCCAAGCAGGGAGGAAAAATGCGTTTCAGTTGGATCTAT
GATAGGGAAGCTGT
```



APPENDIX H

The Nucleotide and Deduced Amino Acid Sequences of ORFs

(a) VMPKAT

3	GGG	GGA	GGG	AAA	GAG	GGA	GGA	AGA	GAG	AAC	GGG	AGA	GAG	ATG	GAG	47
0	G	G	G	K	E	G	G	R	E	N	G	R	E	M	E	14
48	AAA	GCC	ATC	GAC	AGG	CAG	GGA	GTA	CTT	CGC	CGG	CAT	CTT	CAA	TCC	92
15	K	A	I	D	R	Q	G	V	L	R	R	H	L	Q	S	29
93	TCT	TCT	TAC	GAC	CAT	CCT	ACA	TCA	CTA	TCG	GCG	TCT	GTA	GGT	GCT	137
30	S	S	Y	D	H	P	T	S	L	S	A	S	V	G	A	44
138	GCG	GGA	GAC	AGT	TCG	GCC	TAC	CAG	AAG	AAC	CAT	GGC	TTT	GAC	GAT	182
45	A	G	D	S	S	A	Y	Q	K	N	H	G	F	D	D	59
183	GAT	GTT	GTT	ATC	GTG	GCT	GCT	TGC	CGG	ACT	GCA	ATT	GGC	AAG	GCC	227
60	D	V	V	I	V	A	A	C	R	T	A	I	G	K	A	74
228	AAT	AGA	GGA	GGT	TTC	AAG	GAT	ACA	CAC	GCA	GAG	GAC	CTT	CTA	GCT	272
75	N	R	G	G	F	K	D	T	H	A	E	D	L	L	A	89
273	CCT	CTT	CTG	AAG	GCT	CTG	TTG	GAG	AAG	ACT	AAA	CTG	AAC	CCA	TAT	317
90	P	L	L	K	A	L	L	E	K	T	K	L	N	P	Y	104
318	GAA	GTT	GGT	GAT	ATA	GTT	GTT	GGT	ACT	GTC	CTG	GCG	CCT	GGT	TCT	362
105	E	V	G	D	I	V	V	G	T	V	L	A	P	G	S	119
363	CAA	AGA	GCT	ATT	GAA	TGC	AGG	ATG	GCA	GCC	TTT	TAT	GCT	GGT	TTC	407
120	Q	R	A	I	E	C	R	M	A	A	F	Y	A	G	F	134
408	CCA	GAG	ACC	GTT	CCT	CTT	AGA	ACT	GTG	AAC	CGA	CAA	TGC	TCA	TCT	452
135	P	E	T	V	P	L	R	T	V	N	R	Q	C	S	S	149
453	GGT	CTT	CAG	GCT	GTT	GCA	GAT	GTT	GCT	GCT	GCT	ATC	AAA	GCA	GGA	497
150	G	L	Q	A	V	A	D	V	A	A	A	I	K	A	G	164
498	TAC	TAT	GGC	ATT	GGT	ATT	GGT	GCT	GGT	TTG	GAG	TCT	ATG	AGT	TCA	542
165	Y	Y	G	I	G	I	G	A	G	L	E	S	M	S	S	179
543	AAC	AAA	GTT	GGA	TGG	GAG	GGA	CCA	CCT	AAT	CCA	AAG	GCC	AAA	TCT	587
180	N	K	V	G	W	E	G	P	P	N	P	K	A	K	S	194
588	CTT	CAA	CAA	GTC	CAA	GAT	TGT	TAT	CTA	CCC	ATG	GGA	ATG	ACA	TCT	632
195	L	Q	Q	V	Q	D	C	Y	L	P	M	G	M	T	S	209
633	GAA	AAT	GTT	GCA	GAG	CGG	TAT	GGA	ATA	ACA	AAG	GAA	GAA	CAA	GAC	677
210	E	N	V	A	E	R	Y	G	I	T	K	E	E	Q	D	224
678	CAA	GCT	GCT	GTG	ATA	TCT	CAT	AAG	CGT	GCG	GCT	GCT	GCA	ATT	GCT	722
225	Q	A	A	V	I	S	H	K	R	A	A	A	A	I	A	239
723	TCT	GGT	AAA	TTT	AAA	GAA	GAA	ATA	GTA	CCC	GTG	ACA	ACA	AAG	ATT	767
240	S	G	K	F	K	E	E	I	V	P	V	T	T	K	I	254
768	ACG	GAT	CCA	AAA	AGT	GGA	ATT	GAA	AAG	GAG	ATT	ACG	GTC	TCT	GTG	812

255	T	D	P	K	S	G	I	E	K	E	I	T	V	S	V	269
813	GAT	GAG	GGA	GTT	CGC	CCC	GAT	ACG	TCC	TTG	TCC	GGC	CTT	GCA	AAA	857
270	D	E	G	V	R	P	D	T	S	L	S	G	L	A	K	284
858	CTC	AAA	CCA	GCA	TTT	AAG	GAG	GGT	GGA	AGC	ACT	ACG	GCA	GGC	AAT	902
285	L	K	P	A	F	K	E	G	G	S	T	T	A	G	N	299
903	GCT	AGT	CAA	GTT	AGT	GAT	GGT	GCT	GGG	GCG	GTC	CTA	CTA	ATG	AGG	947
300	A	S	Q	V	S	D	G	A	G	A	V	L	L	M	R	314
948	AGG	GAT	GTT	GCA	ATA	CAA	AAA	GGG	CTT	CCC	ATT	CTT	GGA	ATT	TTC	992
315	R	D	V	A	I	Q	K	G	L	P	I	L	G	I	F	329
993	AGG	AGC	TTT	GCT	GCA	GTT	GGG	GTT	GAT	CCT	GCT	GTG	ATG	GGC	ATT	1037
330	R	S	F	A	A	V	G	V	D	P	A	V	M	G	I	344
1038	GGT	CCT	GCT	GTT	GCA	ATA	CCA	GCT	GCA	GTA	AAA	GCA	GCT	GGT	CTT	1082
345	G	P	A	V	A	I	P	A	A	V	K	A	A	G	L	359
1083	CAA	ATT	AAA	GAT	ATT	GAC	TTG	TAT	GAA	ATA	AAT	GAG	GCT	TTT	GCG	1127
360	Q	I	K	D	I	D	L	Y	E	I	N	E	A	F	A	374
1128	TCG	CAG	TTT	GTA	TAC	TGC	ATC	AAA	GAG	TTG	GGG	CTT	GAT	CCA	GAA	1172
375	S	Q	F	V	Y	C	I	K	E	L	G	L	D	P	E	389
1173	AAA	GTC	AAT	GTT	AAT	GGC	GGT	GCT	ATT	GCT	CTT	GGC	CAT	CCA	TTG	1217
390	K	V	N	V	N	G	G	A	I	A	L	G	H	P	L	404
1218	GGT	GCA	ACG	GGT	GCA	CGC	TGT	GTT	AGT	ACC	TTA	TTA	AAT	GAA	ATG	1262
405	G	A	T	G	A	R	C	V	S	T	L	L	N	E	M	419
1263	AAA	CGC	AGG	GGC	AAT	GAC	TGC	CGA	TTT	GGA	ATT	ATT	TCC	ATG	TGC	1307
420	K	R	R	G	N	D	C	R	F	G	I	I	S	M	C	434
1308	ATT	GGT	TCT	GGG	ATG	GGA	GCT	GCG	GCT	GTT	TTT	GAG	AGA	GGC	GAT	1352
435	I	G	S	G	M	G	A	A	A	V	F	E	R	G	D	449
1353	GTA	GCT	GAT	CGG	TTG	AGT	AAT	GTG	CAG	ATT	GCC	CAT	TCC	TAG	AAA	1397
450	V	A	D	R	L	S	N	V	Q	I	A	H	S	*	K	464
1398	TCT	GCT	ATT	CGG	ATT	TCT	ACT	GTT	AAG	TAG	AGC	TCA	GTT	TTT	AAG	1442
465	S	A	I	R	I	S	T	V	K	*	S	S	V	F	K	479
1443	CAC	TTT	AAT	TAA	TAA	TTT	AAT	ATA	GAC	TAC	CTT	TGT	TTT	CCT	TAA	1487
480	H	F	N	*	*	F	N	I	D	Y	L	C	F	P	*	494
1488	TCA	ACA	ATA	AAT	CAA	AGC	TTA	TGC	TCC	TTT	AAT	TGC	GCT	GTG	GAT	1532
495	S	T	I	N	Q	S	L	C	S	F	N	C	A	V	D	509
1533	TGG	TGA	TAA	TTT	GGT	TAT	AAA	AAA	1577							
510	W	*	*	F	G	Y	K	K	K	K	K	K	K	K	K	

The Nucleotide and Deduced Amino Acid Sequences of VMPKAT. The open reading frame (ORF) of VMPKAT starts at nucleotide 42. The asterisk (*) indicates stop codon.

(b) VMPBSMT

1	TCT	CTC	TCT	CTC	TCT	CTT	CTA	CAA	CAG	GGA	AAG	GTT	ACT	GAC	CAA	45
1	S	L	S	L	S	L	L	Q	Q	G	K	V	T	D	Q	15
46	TAT	AAA	TAT	TGT	GCA	GAC	GAG	TTC	TTT	CTA	GAG	GAG	ATT	ATA	ACC	90
16	Y	K	Y	C	A	D	E	F	F	L	E	E	I	I	T	30
91	ATG	ATT	GCT	GAA	CAG	GTC	CTT	GCA	ATG	ACG	GGA	GGT	AGC	GGA	GAG	135
31	M	I	A	E	Q	V	L	A	M	T	G	G	S	G	E	45
136	AAC	AGC	TAT	GCT	GAA	AAC	TCC	AAG	CTT	CAA	GAG	AAG	GCA	GGG	ATT	180
46	N	S	Y	A	E	N	S	K	L	Q	E	K	A	G	I	60
181	ATG	ACG	AAA	CCA	TTA	GGG	GAA	GAA	GCC	TTA	AGA	GAA	ATA	TGT	AAA	225
61	M	T	K	P	L	G	E	E	A	L	R	E	I	C	K	75
226	ACT	CTC	CCA	GAA	AAG	CTT	GTT	GTG	GCT	GAC	TTG	GGA	TGC	TCC	TCT	270
76	T	L	P	E	K	L	V	V	A	D	L	G	C	S	S	90
271	GGT	CCC	AAC	ACA	TTA	CTT	GTG	CTC	TCT	CAG	ATC	ATC	AAT	GCA	GTG	315
91	G	P	N	T	L	L	V	L	S	Q	I	I	N	A	V	105
316	GAT	GAG	TAT	TGC	TCT	AAA	TCG	AGC	CAA	CAA	AGG	CCT	GAA	ATC	CAA	360
106	D	E	Y	C	S	K	S	S	Q	Q	R	P	E	I	Q	120
361	TAT	TCT	CTT	TGT	GAT	CTT	CCT	GAC	AAT	GAC	TTT	AAC	ACT	TTG	TTC	405
121	Y	S	L	C	D	L	P	D	N	D	F	N	T	L	F	135
406	CAA	TTA	TCG	GAG	CAA	TAT	CAG	AAG	AAA	ATT	AGA	GAA	GAG	AAA	GGG	450
136	Q	L	S	E	Q	Y	Q	K	K	I	R	E	E	K	G	150
451	GAT	AGA	TAT	GAA	CCA	TTC	TAT	ATT	GTT	GGA	TCG	CCT	GGT	TCG	TTT	495
151	D	R	Y	E	P	F	Y	I	V	G	S	P	G	S	F	165
496	TAT	AGA	AGG	TTA	TTT	CCT	TCA	AAA	AGT	GTG	CAG	CTT	TTC	CAC	TCT	540
166	Y	R	R	L	F	P	S	K	S	V	Q	L	F	H	S	180
541	TCA	TAC	AGC	CTC	ATG	TGG	CTA	TCG	CAG	GTA	CCT	AAA	GGG	CTA	AAA	585
181	S	Y	S	L	M	W	L	S	Q	V	P	K	G	L	K	195
586	GAT	GAA	GCC	TTT	GAT	CAC	TTC	AAT	AGG	ACA	GCC	ATA	TAT	ATC	AAT	630
196	D	E	A	F	D	H	F	N	R	T	A	I	Y	I	N	210
631	AAT	GAA	AGC	TAT	TCT	ATC	GTG	TCG	CAT	TTG	TAT	GTT	GAA	CAA	TTT	675
211	N	E	S	Y	S	I	V	S	H	L	Y	V	E	Q	F	225
676	AAA	ACA	GAT	TTC	TCG	GAT	TTC	CTC	AAA	TCC	CGT	TCT	AAG	GAG	CTT	720
226	K	T	D	F	S	D	F	L	K	S	R	S	K	E	L	240
721	ATT	TCT	GGA	GGG	AGC	ATG	ATT	TTA	TCA	TTT	TTG	GGA	AGA	AGT	GAC	765
241	I	S	G	G	S	M	I	L	S	F	L	G	R	S	D	255
766	TTT	CAA	TCT	ATT	GGT	GAA	ATG	TTA	TAT	TTA	TGG	GAG	CTT	CTT	GCA	810
256	F	Q	S	I	G	E	M	L	Y	L	W	E	L	L	A	270
811	GAA	GCG	CTA	ACT	GCA	ATG	GTT	TCT	GAG	GGA	CTA	GTG	GAG	GAA	GAT	855
271	E	A	L	T	A	M	V	S	E	G	L	V	E	E	D	285
856	AAA	CTA	AAA	GAC	TTT	AAT	TTA	CCA	TTT	TAT	ACG	CCA	TTG	ATG	GAA	900
286	K	L	K	D	F	N	L	P	F	Y	T	P	L	M	E	300

901	GAA	GTG	AGA	TCT	ATT	ATC	AGT	ATG	GAA	GGA	TCA	TTT	CAT	CTT	AAA	945
301	E	V	R	S	I	I	S	M	E	G	S	F	H	L	K	315
946	CAA	GCA	CAA	TCT	TTT	GAG	TCA	AAT	TGG	GAT	CCT	TTT	GAT	GAA	TCA	990
316	Q	A	Q	S	F	E	S	N	W	D	P	F	D	E	S	330
991	CAT	GAA	AAT	TTT	GTG	GAA	GAC	AAA	AGC	TCG	AGT	AGT	AGA	AAT	GTT	1035
331	H	E	N	F	V	E	D	K	S	S	S	S	R	N	V	345
1036	GCA	AAT	TAC	ATG	AGA	GCA	GTG	CTA	GAA	CCT	CTA	GTT	ACA	TCT	TAT	1080
346	A	N	Y	M	R	A	V	L	E	P	L	V	T	S	Y	360
1081	TTT	GGA	ACT	GAT	ATT	TTA	GAA	GAT	TTA	TTC	TCG	CGT	TAT	GCA	AAT	1125
361	F	G	T	D	I	L	E	D	L	F	S	R	Y	A	N	375
1126	AAT	ATT	GCA	AGG	CAC	ATG	TAT	AAT	GGG	AAG	CCA	AAG	CAC	ACT	GTT	1170
376	N	I	A	R	H	M	Y	N	G	K	P	K	H	T	V	390
1171	TTC	ATT	CTT	TTC	TTG	AAA	TTA	AAA	GGT	GAA	GAG	CAA	ACA	CTT	CGA	1215
391	F	I	L	F	L	K	L	K	G	E	E	Q	T	L	R	405
1216	GAA	TGC	GAA	TAA	ATG	CAA	TAT	ATA	AGT	GAA	GAT	GGT	GCA	GTG	AGC	1260
406	E	C	E	*	M	Q	Y	I	S	E	D	G	A	V	S	420
1261	AAT	CTT	AAC	ACT	TCA	GAA	TAA	GAC	AAA	ATA	AAA	TTT	ACT	GTT	CTC	1305
421	N	L	N	T	S	E	*	D	K	I	K	F	T	V	L	435
1306	CCG	AGC	AAG	CAA	TGT	CTT	GAT	TGT	CTC	AAC	TTT	AAG	AGA	AGA	ATA	1350
436	P	S	K	Q	C	L	D	C	L	N	F	K	R	R	I	450
1351	AGT	GGG	AAC	ATA	CAA	CCC	TTA	CTA	TTC	CCA	TGT	ATC	CAT	CAT	TCT	1395
451	S	G	N	I	Q	P	L	L	F	P	C	I	H	H	S	465
1396	CAA	CTT	GGA	ATA	AAA	TTA	ATG	TTT	TCT	AAA	AAA	AAA	AAA	AAA	AAA	1440
466	Q	L	G	I	K	L	M	F	S	K	K	K	K	K	K	480
1441	AAA	AAA	AAA	AAA	AAA											1455
481	K	K	K	K	K											

The Nucleotide and Deduced Amino Acid Sequences of VMPBSMT. The open reading frame (ORF) of VMPBSMT starts at nucleotide 91. The asterisk (*) indicates stop codon.

(c) **VMPEGS**

3	GGG	GGA	ACT	GAC	AGG	AAG	TGA	ACG	GAC	CAC	CAA	ACG	GCG	ACC	ACC	47
0	G	G	T	D	R	K	*	T	D	H	Q	T	A	T	T	14
48	AAA	CGG	CGA	CCG	GAT	GTT	CTA	TAA	AAG	GGG	GGG	TTA	GCG	TAT	GCG	92
15	K	R	R	P	D	V	L	*	K	G	G	L	A	Y	A	29
93	TGC	TTC	GCA	ATA	TAC	GAT	CCA	TTG	TCT	CGG	CGG	ATA	AGC	TTA	ACA	137
30	C	F	A	I	Y	D	P	L	S	R	R	I	S	L	T	44
138	CGC	ATT	TTA	TCG	TGC	TCC	ATC	TTC	TTA	CAT	TCA	ATG	GCG	TCG	GAG	182
45	R	I	L	S	C	S	I	F	L	H	S	M	A	S	E	59
183	AAG	AGC	AAG	ATC	CTC	ATC	ATC	GGA	GGC	ACG	GGA	TAC	ATC	GGG	AAG	227
60	K	S	K	I	L	I	I	G	G	T	G	Y	I	G	K	74
228	CAT	GTC	GTG	GCT	GCG	AGC	GTC	AAG	TTG	GGC	CAC	CCT	ACC	TTC	GTT	272
75	H	V	V	A	A	S	V	K	L	G	H	P	T	F	V	89
273	CTC	GTC	AGG	CCA	AGC	TCA	GCA	TCC	GAC	CCT	GCG	AAG	GTT	GGA	TTC	317
90	L	V	R	P	S	S	A	S	D	P	A	K	V	G	F	104
318	CTC	CAG	GGC	TTA	ACC	GGG	GCG	GAA	GTC	AAA	GTT	GTT	CAT	GGG	GAT	362
105	L	Q	G	L	T	G	A	E	V	K	V	V	H	G	D	119
363	TTG	AAT	GAC	TAT	CAG	AGC	TTG	CTG	AGT	GCA	ATC	AAG	CAG	GTG	GAT	407
120	L	N	D	Y	Q	S	L	L	S	A	I	K	Q	V	D	134
408	ATC	GTC	ATC	TCC	ACC	GTT	GGA	GGG	CTA	CAG	TTG	GCA	GAT	CAG	GTT	452
135	I	V	I	S	T	V	G	G	L	Q	L	A	D	Q	V	149
453	AAC	ATC	ATC	AAA	GCC	ATT	AAA	GAA	GCA	GGA	AAC	ATC	AAG	AGG	TTT	497
150	N	I	I	K	A	I	K	E	A	G	N	I	K	R	F	164
498	CTG	CCA	TCA	GAG	TTT	GGG	TTG	GAT	CCG	GAA	CGC	GTG	ACT	AAG	ATA	542
165	L	P	S	E	F	G	L	D	P	E	R	V	T	K	I	179
543	GAA	CCG	GCA	GCA	TCC	GCT	CTC	GCT	GGC	AAG	GTT	GCA	ATC	CGG	CAA	587
180	E	P	A	A	S	A	L	A	G	K	V	A	I	R	Q	194
588	GCC	ATC	AAG	GCA	GAA	GGC	ATC	CCC	TAC	ACG	ATC	ATC	TCT	AAT	AAC	632
195	A	I	K	A	E	G	I	P	Y	T	I	I	S	N	N	209
633	TAC	TTC	GCC	GGT	TAC	TCC	CTC	CCC	ACC	TTG	GCT	CAA	GCC	AGC	CCT	677
210	Y	F	A	G	Y	S	L	P	T	L	A	Q	A	S	P	224
678	CCC	GTT	GAC	AAA	ATC	ACC	ATC	TTT	GGT	GAT	GGA	AAC	ACC	AAA	GGT	722
225	P	V	D	K	I	T	I	F	G	D	G	N	T	K	G	239
723	GTC	TTT	GTT	GAT	GAA	GAT	GAT	ATT	GGC	ATA	TAT	ACC	ATC	AAG	GCG	767
240	V	F	V	D	E	D	D	I	G	I	Y	T	I	K	A	254
768	GCT	AAT	GAT	CCA	AGA	ACC	CTA	AAC	AAG	ATT	GTT	TAC	ATA	AGG	CCC	812
255	A	N	D	P	R	T	L	N	K	I	V	Y	I	R	P	269
813	CCT	GGT	TGC	ATC	TAC	TCT	CAC	AAC	GAG	CTC	ATC	TCT	CTT	TGG	GAG	857
270	P	G	C	I	Y	S	H	N	E	L	I	S	L	W	E	284
858	AAG	AAG	ACC	GGC	AAA	ACC	CTA	GAG	AGG	ATT	TAT	CTC	TCG	GAT	GAA	902
285	K	K	T	G	K	T	L	E	R	I	Y	L	S	D	E	299

903	GAA	ATT	TTT	AAG	ATT	ATC	CAA	GAA	GCA	CCA	ATC	CCT	TTC	AAC	GTA	947
300	E	I	F	K	I	I	Q	E	A	P	I	P	F	N	V	314
948	ATA	TTT	GCT	TTA	AAC	TAC	TTG	GTG	TTT	GTG	AAA	GGA	GAC	TGT	CTC	992
315	I	F	A	L	N	Y	L	V	F	V	K	G	D	C	L	329
993	AGC	TTT	GAG	ATT	GAC	CCA	TCC	GTT	GCT	GCT	GAA	GCC	ACT	GAG	CTC	1037
330	S	F	E	I	D	P	S	V	A	A	E	A	T	E	L	344
1038	TAT	CCC	GAA	GTC	AAA	TAC	ACT	ACA	GTT	GAG	GAA	TAT	CTC	AGT	CGT	1082
345	Y	P	E	V	K	Y	T	T	V	E	E	Y	L	S	R	359
1083	GTT	CTG	TAA	ATA	AAG	TTT	ACA	ACT	ATC	TCC	TTC	TCA	AGT	TTT	TCA	1127
360	V	L	*	I	K	F	T	T	I	S	F	S	S	F	S	374
1128	GAA	AAG	TCC	TAT	GGA	AAT	TTC	TAT	TAT	GTG	TTG	CTG	TTT	TGC	TCA	1172
375	E	K	S	Y	G	N	F	Y	Y	V	L	L	F	C	S	389
1173	TCT	TTG	CCA	TCC	TTG	TTT	GAA	GTC	TGT	TTC	TGT	GAT	GTC	TTG	ACC	1217
390	S	L	P	S	L	F	E	V	C	F	C	D	V	L	T	404
1218	TTC	TGT	CTG	TTT	TGT	TGG	ACT	GTT	GAT	CTA	TGA	GTA	TTG	TGT	ATT	1262
405	F	C	L	F	C	W	T	V	D	L	*	V	L	C	I	419
1263	TTG	GGA	TCT	TTG	CAT	GAG	TTC	TGA	TTT	TTT	ATG	TGA	AAA	AAA	AAA	1307
420	L	G	S	L	H	E	F	*	F	F	M	*	K	K	K	434
1308	AAA	AAA	1313													
435	K	K	436													

The Nucleotide and Deduced Amino Acid Sequences of VMPEGS. The open reading frame (ORF) of VMPEGS starts at nucleotide 71. The asterisk (*) indicates stop codon.

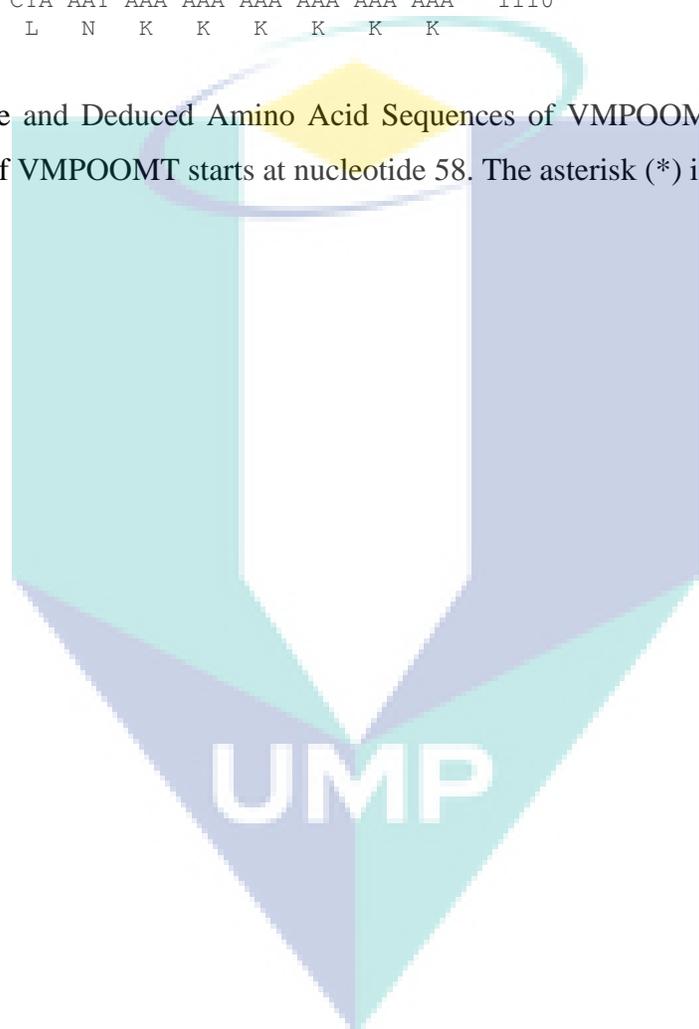
UMP

(d) VMPOOMT

1	GAA	CTC	CTT	TCC	TCT	CTC	TCC	ATC	TCT	CAG	GAC	AAA	TCT	CGC	TAC	45
1	E	L	L	S	S	L	S	I	S	Q	D	K	S	R	Y	15
46	CTT	CGC	AGC	ATC	ATG	CGC	GTC	CTG	TCT	CAA	GAA	TGC	ATC	TTT	AAA	90
16	L	R	S	I	M	R	V	L	S	Q	E	C	I	F	K	30
91	TCC	CAC	ATC	ACC	CCC	TCT	GAT	GAA	GAA	ACC	TTC	GAC	CTC	ACT	CCT	135
31	S	H	I	T	P	S	D	E	E	T	F	D	L	T	P	45
136	GTC	TCC	CGC	CTC	CTT	CTC	ACC	ACC	ACC	TCC	TTT	CCC	CAC	AAC	CAT	180
46	V	S	R	L	L	L	T	T	T	S	F	P	H	N	H	60
181	AAC	AGC	TCT	CCC	TTT	GTC	CTC	CTA	AGC	CTC	AAT	CGC	CAC	CTC	GTC	225
61	N	S	S	P	F	V	L	L	S	L	N	R	H	L	V	75
226	GAC	TCG	TTT	CAT	CAA	CTC	TCC	CAC	TGG	TTC	CTT	AAA	CCG	GAT	GCC	270
76	D	S	F	H	Q	L	S	H	W	F	L	K	P	D	A	90
271	ACC	ATA	ACC	CCT	TTT	GCG	ATG	GCA	AAC	GGG	AAG	GAG	TTC	TGG	AAT	315
91	T	I	T	P	F	A	M	A	N	G	K	E	F	W	N	105
316	CTG	GCA	AGG	GAG	CTA	CCA	CAA	TTT	AGT	GAT	TTG	TTT	AGT	AAG	GCA	360
106	L	A	R	E	L	P	Q	F	S	D	L	F	S	K	A	120
361	ATG	GAA	TGT	GAT	TCC	AAC	TTT	CTT	ATG	AAT	GCT	ATG	GTG	AGT	ACT	405
121	M	E	C	D	S	N	F	L	M	N	A	M	V	S	T	135
406	GGA	GAG	GGG	CCA	TTT	TGG	GGG	ATT	GAG	ACA	TCA	CTG	GAT	GTT	GGA	450
136	G	E	G	P	F	W	G	I	E	T	S	L	D	V	G	150
451	GGT	GGC	ACT	GGC	TTG	CTG	GCA	ACA	ACA	TTG	GCG	GAG	GCA	TTT	CCT	495
151	G	G	T	G	L	L	A	T	T	L	A	E	A	F	P	165
496	GGG	TTA	AAA	TGC	ATC	GTG	TTT	GAT	CTA	CCA	CAT	GTC	ATT	GGT	ATG	540
166	G	L	K	C	I	V	F	D	L	P	H	V	I	G	M	180
541	GCA	GAG	AAG	AAG	ACG	GAC	GGG	ATT	CAA	TTC	GTC	GCC	GGT	GAT	ATG	585
181	A	E	K	K	T	D	G	I	Q	F	V	A	G	D	M	195
586	TTT	GTT	GAG	ATT	CCA	CGT	GCC	GAT	GTT	GCT	CTG	CTT	AAG	TGG	ATA	630
196	F	V	E	I	P	R	A	D	V	A	L	L	K	W	I	210
631	TTG	CAT	GAT	TGG	AGT	GAT	GAA	GAT	TGC	GTC	AAA	ATA	CTA	CAA	CGT	675
211	L	H	D	W	S	D	E	D	C	V	K	I	L	Q	R	225
676	TGC	AAG	GAA	GCA	ATT	CCT	TCT	AAA	GAG	AAA	GGT	GGA	AAA	GTG	ATT	720
226	C	K	E	A	I	P	S	K	E	K	G	G	K	V	I	240
721	ATA	ATC	GAC	ATG	GTG	GTT	GGT	TTG	AAC	TCC	AAC	TCT	CAT	ACT	TCA	765
241	I	I	D	M	V	V	G	L	N	S	N	S	H	T	S	255
766	TTG	CAG	ACA	CAA	CTT	CTG	TAT	GAC	GTG	CAA	ATG	ATG	ACT	TTG	TGC	810
256	L	Q	T	Q	L	L	Y	D	V	Q	M	M	T	L	C	270
811	ATG	GGA	AAA	GAA	AGG	GAT	GAA	AAT	GAA	TGG	CGT	AAT	CTA	TTT	GTC	855
271	M	G	K	E	R	D	E	N	E	W	R	N	L	F	V	285
856	GCG	GCA	GGT	TAT	AGA	GAC	TAC	AAG	ATA	ACA	CCA	TAT	GTT	GAT	CAC	900
286	A	A	G	Y	R	D	Y	K	I	T	P	Y	V	D	H	300

901	ATG	AGA	TCA	ATT	ATT	GAA	GTC	TAT	CCT	TGA	TTA	ATT	ATT	ATT	TGA	945
301	M	R	S	I	I	E	V	Y	P	*	L	I	I	I	*	315
946	TTT	CAA	GTC	TTA	TTA	TTA	GTA	CAT	GTT	ATT	GTG	CTT	ATT	ATG	ATT	990
316	F	Q	V	L	L	L	V	H	V	I	V	L	I	M	I	330
991	AGT	TTA	ACT	TTT	CTT	TCA	TGA	TAA	GAG	TTA	AAG	CAA	GTA	ATG	AAG	1035
331	S	L	T	F	L	S	*	*	E	L	K	Q	V	M	K	345
1036	AAT	TAT	GTA	CTT	TTG	GCT	ATG	GCT	GAG	CAT	TGT	GAA	TGT	AGC	TAT	1080
346	N	Y	V	L	L	A	M	A	E	H	C	E	C	S	Y	360
1081	TAT	TTT	CTA	AAT	AAA	1110										
361	Y	F	L	N	K	K	K	K	K	K	K	K	K	K		

The Nucleotide and Deduced Amino Acid Sequences of VMPOOMT. The open reading frame (ORF) of VMPOOMT starts at nucleotide 58. The asterisk (*) indicates stop codon.



APPENDIX I

Supplementary Data for Structural and Functional Prediction on Fragrance-related Proteins

(a) Predicted ligand for VMP protein sequences

Protein name		Ligand		
VMPKAT		EDO (1,2-ethanediol) Eg. Of PDB code: 2wu9		
VMPBSMT	SAH (s-adenosyl-L-homocysteine) Example of PDB code: 1m6e, 2eg5, 2efj, 3b5i	37T (theobromine) Example of PDB code: 2efj	SAL (2-hydroxybenzoic acid) Example of PDB code: 1m6e	Magnesium ion Example of PDB code: 3b5i
VMPEGS	NAP = NADP nicotinamide-adenine-dinucleotide phosphate Example of PDB code: 3c1o, 2qw8	Nitrate ion Example of PDB code: 2qw8	PEG di(hydroxyethyl) ether Example of PDB code: 2qw8	
VMPOOMT	SAH Example of PDB code: 1fp2, 2qyo, 1zga	SAM Example of PDB code: 1fpx	Glycerol Example of PDB code: 4e70	Potassium ion Example of PDB code: 2qyo

(b) Predicted homologous family on VMP protein sequences

PROTEIN NAME	PSI-BLAST	SUPERFAMILY HMM LIBRARY
VMPKAT	Thiolase-like family	-Thiolase-like family -region in input sequence (40-306, 314-435) -strong classification (low E-value)
VMPBSMT	Methyltransferases	-S-adenosyl-L-

	superfamily	methionine-dependent methyltransferases superfamily -Salicylic acid carboxyl methyltransferase (SAMT) family -region in input sequence (3-366) -strong classification(low E-value)
VMPEGS	NADP binding site Active site lysine Short-chain dehydrogenase/reductase superfamily	-NAD(P)-binding Rossmann-fold domains superfamily -Tyrosine-dependent oxidoreductases family -Region in input sequence (5- 273)
VMPOOMT	AdoMet-dependent methyltransferases superfamily Dimerization superfamily S-adenosylmethionine binding site	-Region (14-112): “winged helix” DNA-binding domain superfamily, plant O-methyltransferase family at N-terminal domain (weak HMM library due to E-value > 0.0001) -Region (118-366): S-adenosyl-L-methionine-dependent methyltransferases superfamily, plant O-methyltransferase family at C-terminal domain (strong HMM library and strong classification due to low E-value)

(c) **Predicted conserved domains on VMP protein sequences**

Protein	Conserved domains on proteins
VMPKAT	dimer interface (20 residues): D69, D96, A110, R114, F118, R129, T130, V131, N132, Q140, A147, A148, A151, S165, F320, A322, G324, A344, S425, G426 -active site (RTVN)- 3 residues: C135, H390, C422 -thiolase: I51- R435 (specific hits) -Condensing enzymes superfamily: I51-R435 -Thiolase, C-terminal domain: P312-R435 -beta-ketoacyl synthase,C-terminal domain: P312- R435 -multi domain= 3-ketoacyl-CoA thiolase: E2- Q445 -thiolase,N-terminal domain:V50- R435 acetyl-CoA acetyltransferases -putative acyltransferases
VMPBSMT	SAM dependent carboxyl methyltransferase: E37-K368 indole-3-acetate carboxyl methyltransferase: E37-K368
VMPEGS	NAD(P) binding site on conserved domain PCBER_SDR_a. 19 of 19 of the residues that compose this conserved feature have been mapped to the query sequence: G12, T14, G15, I17, R37, T84, V85, G86, Q89, L90, S112, E113, F114, G115, K134, N154, Y155, F156, S160 active site lysine on conserved domain PCBER_SDR_a. 1 of 1 of the residues that compose this conserved feature have been mapped to the query sequence: K134
VMPOOMT	Specific hits and superfamily: Dimerization domain ;This domain is found at the N-terminus of a variety of plant O-methyltransferases. It has been shown to mediate dimerization of these proteins: M35-S83 S-adenosylmethionine binding site on conserved domain AdoMet_MTases. 13 of 13 of the residues that compose this conserved feature have been mapped to the query sequence.: V208, G209, G210, G211, T212, G213, L214, D232, L233, G252, D253, M254, K267

(d) Summary of PDB codes and related organisms on VMP protein models

Name of studied query protein	PDB ID of template protein	Name of selected protein	Organism	Classification of protein enzymes
VMPKAT	2c7y	3-ketoacyl coa thiolase	<i>Arabidopsis thaliana</i>	Transferase
	2wu9	3-ketoacyl coa thiolase	<i>Arabidopsis thaliana</i>	Transferase
	2wua	Acetoacteyl coa thiolase	<i>Helianthus annuus</i>	Transferase
	2iik	3-ketoacyl coa thiolase	<i>Homo sapiens</i>	Transferase
	1pxt	3-ketoacyl coa thiolase	<i>Saccharomyces cerevisiae</i>	Thiolase
VMPBSMT	1m6e	S-adenosyl-L-methionine:salicylic acid carboxyl methyltransferase	<i>Clarkia breweri</i>	Transferase
	2eg5	Xanthosine methyltransferase	<i>Coffea canephora</i>	Transferase
	2efj	3,7-dimethylxanthine methyltransferase	<i>Coffea canephora</i>	Transferase
	3b5i	S-adenosyl-L-methionine:salicylic acid carboxyl methyltransferase	<i>Arabidopsis thaliana</i>	Transferase
VMPEGS	1qyc	Phenylcoumaran benzylic ether reductase	<i>Pinus taeda</i>	Plant protein
	2gas	Isoflavone reductase	<i>Medicago sativa</i>	Oxidoreductase
	1qyd	Pinoresinol-lariciresinol reductase	<i>Thuja plicata</i>	Plant protein
	3c1o	Eugenol synthase	<i>Clarkia breweri</i>	Oxidoreductase
	2qw8	Eugenol synthase	<i>Ocimum basilicum</i>	Oxidoreductase
VMPOOMT	1fp2	Isoflavone o-methyltransferase	<i>Medicago sativa</i>	Transferase
	2qyo	O-methyltransferase	<i>Medicago truncatula</i>	Transferase
	4e70	Coniferyl alcohol 9-O-methyltransferase	<i>Linum nodiflorum</i>	Transferase
	1fpx	Isoflavone o-methyltransferase	<i>Medicago sativa</i>	Transferase
	1zga	Isoflavanone 4'-O-methyltransferase	<i>Medicago truncatula</i>	Plant protein, transferase

(e) **Predicted 3D Model of VMP protein models**

Quality criteria	VMPKAT	VMPBSMT	VMPEGS	VMPOOMT
Target region	35-437	1-370	4-306	9-368
Protein length	450	378	306	368
Template PDB code	2wu9_A	1m6e_X	3c1o_A	1fp2_A
Template region	38-440	1-359	3-308	1-368
Sequence identity	77%	43%	55%	42%
E-value	0	0	0	0
GA341	1.00	1.00	1.00	1.00
MPQS	1.83896	1.50314	1.7321	1.48046
z-DOPE	-1.12	-0.65	-1.3	-0.29
TSVMod	5.015	3.024	1.522	3.568
RMSD				
TSV NO35	0.865	0.836	0.959	0.823

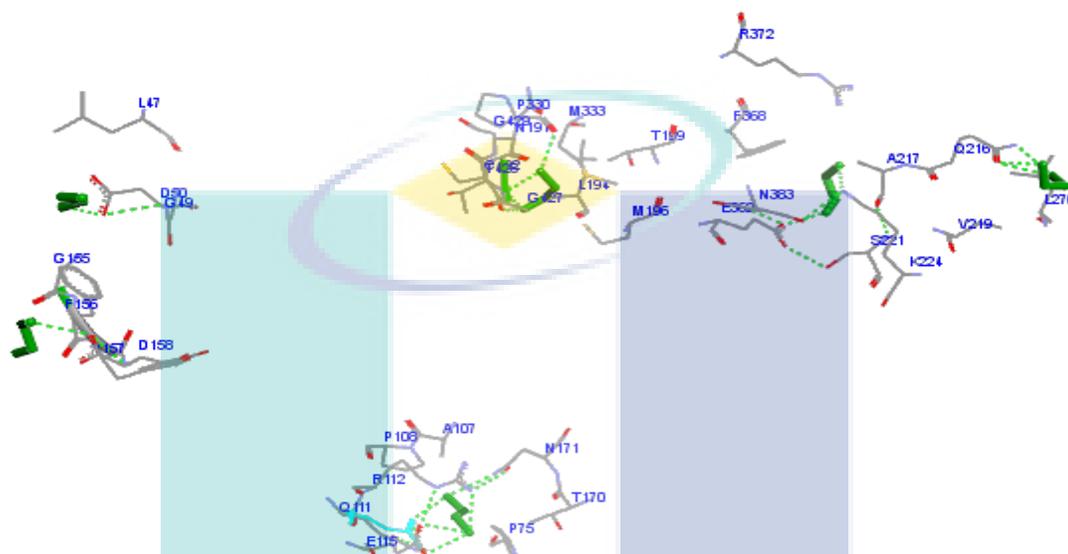
UMP

(f) **Regio-specificity of binding of various putative substrates (ligands) to fragrance-related proteins of *Vanda Mimi Palmer* (targets) indicating residues involved in interaction.**

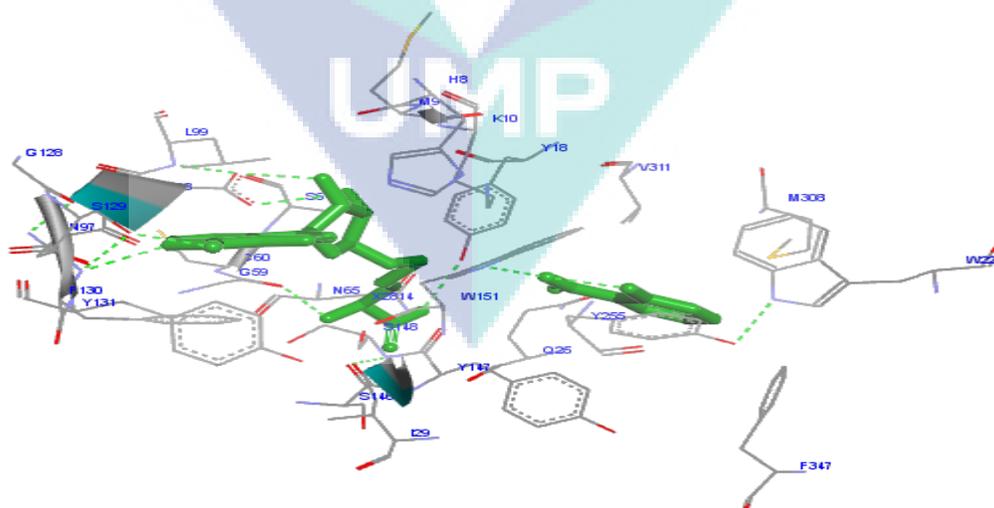
Name of proteins	Substrates	interacting residues and their number in parentheses
VMPKAT	COA	dimer interface (20 residues): D69, D96, A110, R114, F118, R129, T130, V131, N132, Q140, A147, A148, A151, S165, F320, A322, G324, A344, S425, G426 -active site (RTVN)- 3 residues: C135, H390, C422
VMPBSMT	Benzoic acid	SAM dependent carboxyl methyltransferase: E37-K368
VMPEGS	Coniferyl acetate	G12, T14, G15, I17, R37, T84, V85, G86, Q89, L90, S112, E113, F114, G115, K134, N154, Y155, F156, S160
VMPOOMT	Orcinol	V208, G209, G210, G211, T212, G213, L214, D232, L233, G252, D253, M254, K267

(g) Interactions between ligand and interacting amino acid residues of catalytic region in 3D structural model of VMPKAT, VMPBSMT, VMPEGS and VMPOOMT protein.

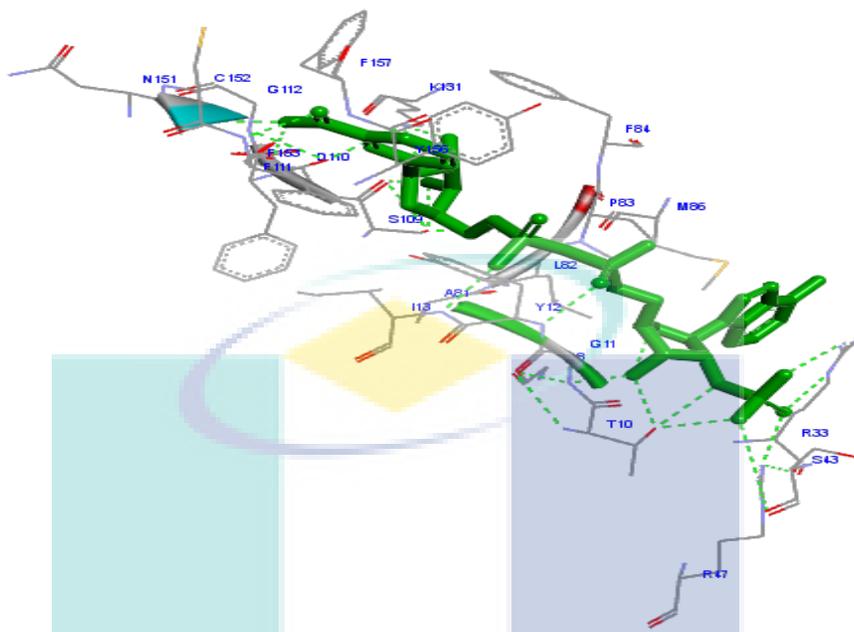
(i) VMPKAT



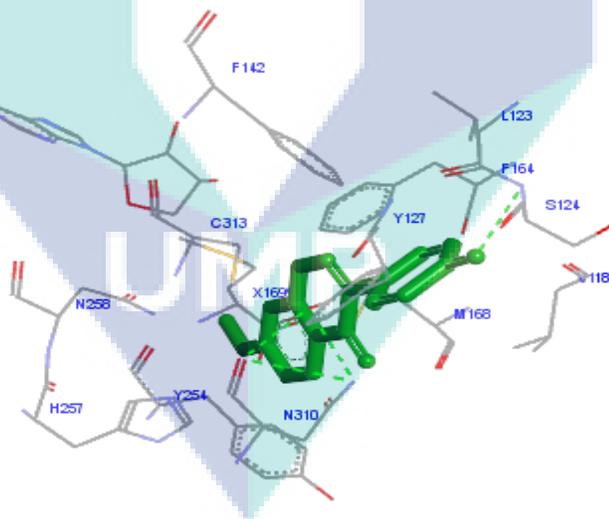
(ii) VMPBSMT



(iii) VMPEGS

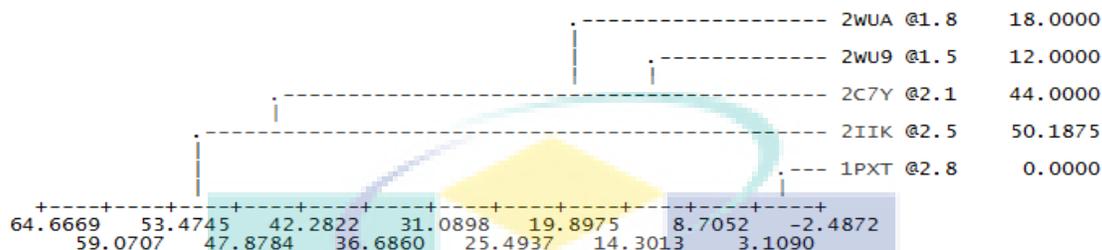


iv) VMPOOMT

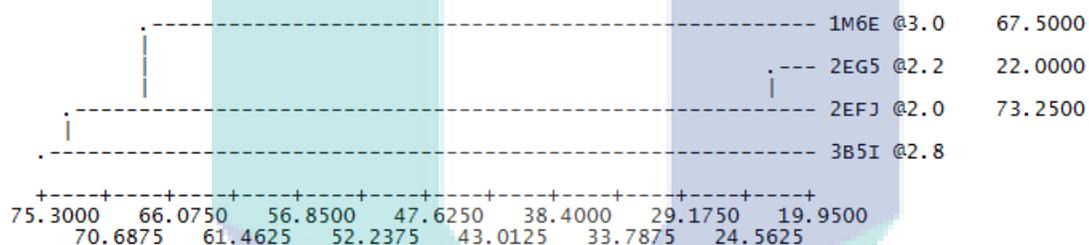


**(h) Weighted-pair group average clustering based on a distance matrix in
MODELLER software, version 4.2.**

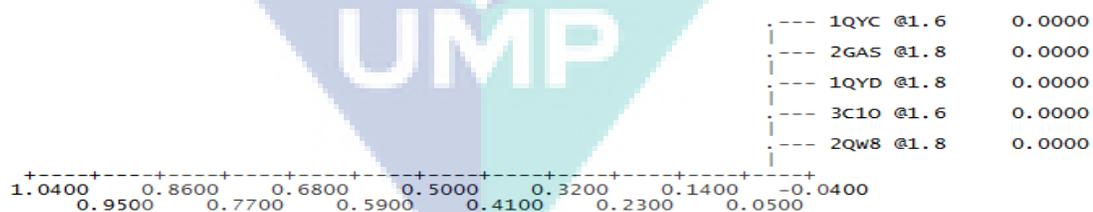
(i) VMPKAT



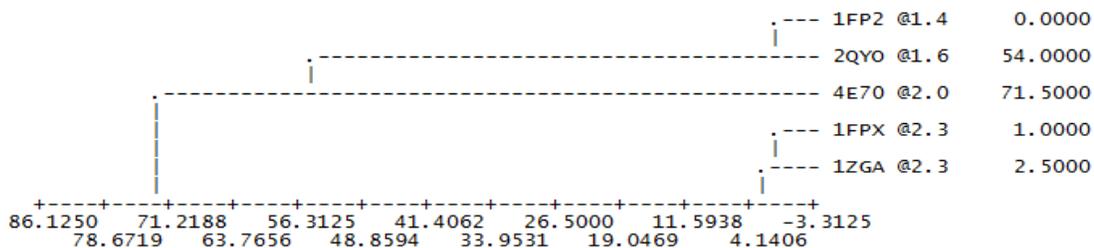
(ii) VMPBSMT



(iii) VMPEGS



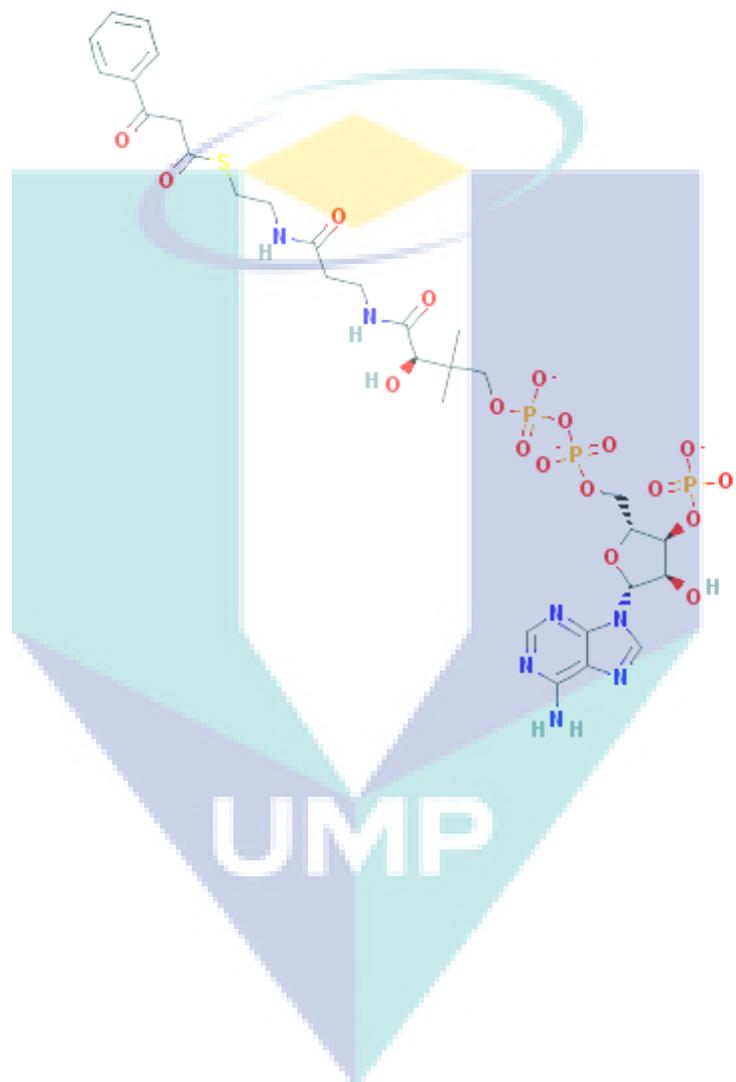
(iv) VMPOOMT



(i) **Putative substrate of VMPKAT protein: 3-oxo-3-phenylpropionyl-CoA**

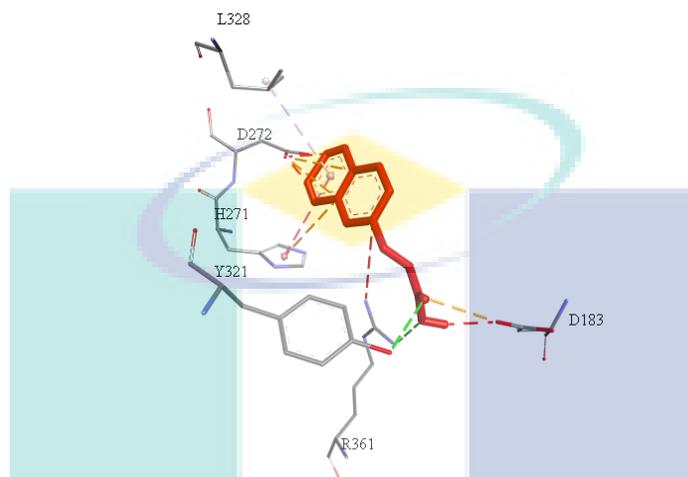
Compound: CID 25203310 (PubChem database)

Conformer generation is disallowed since too many atoms, too flexible

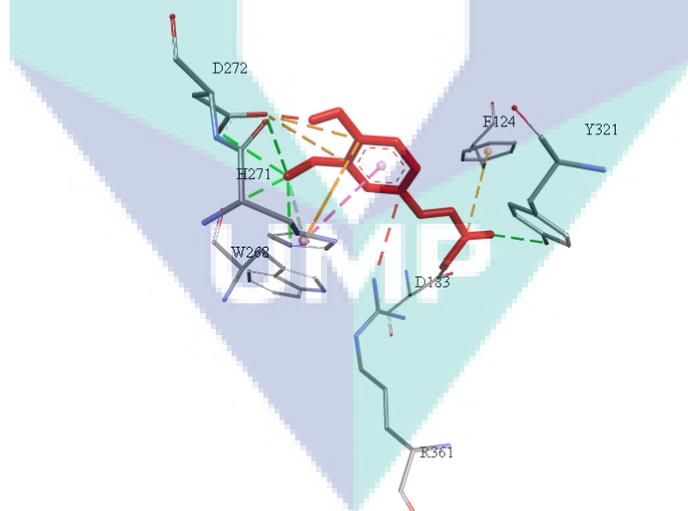


(j) Molecular docking of VMPOOMT protein with several substrates performed using Autodock software, version 4.2 and visualized by Discovery Studio Visualizer.

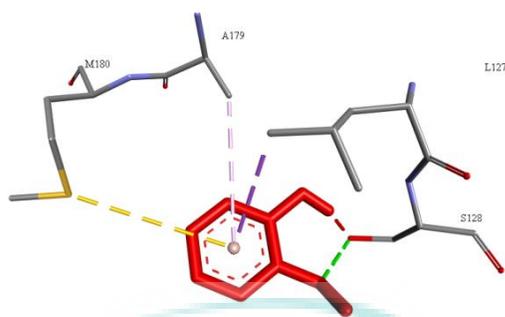
(i) VMPOOMT-caffeic acid complex



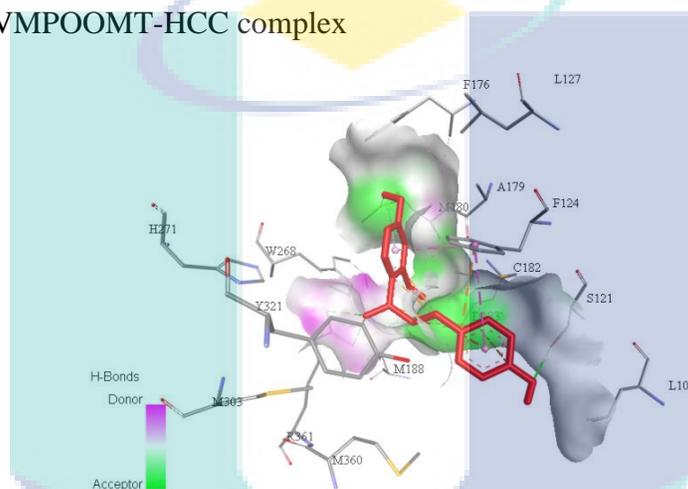
(ii) VMPOOMT-ferulic acid complex



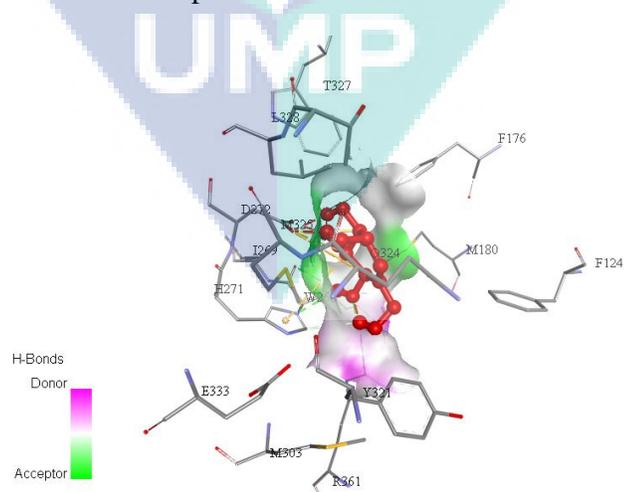
(iii) VMPOOMT-guaiacol complex



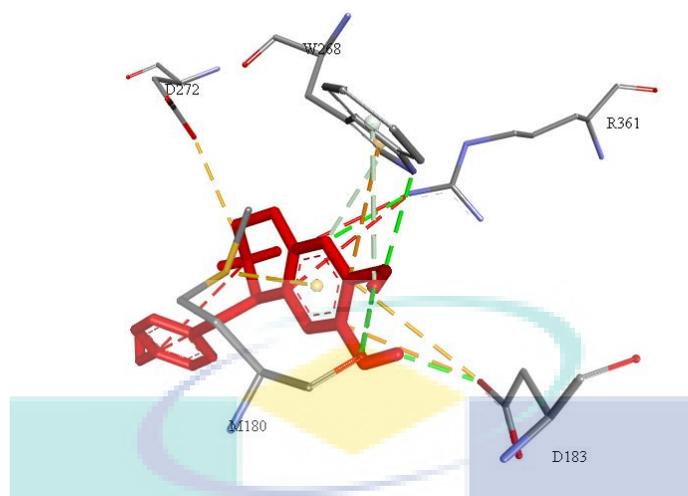
(iv) VMPOOMT-HCC complex



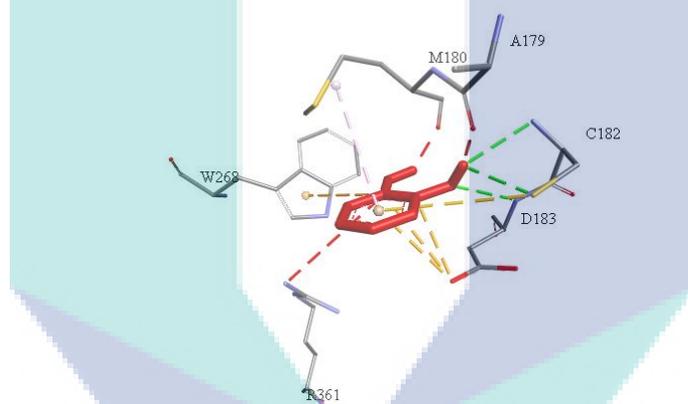
(v) VMPOOMT-HMC complex



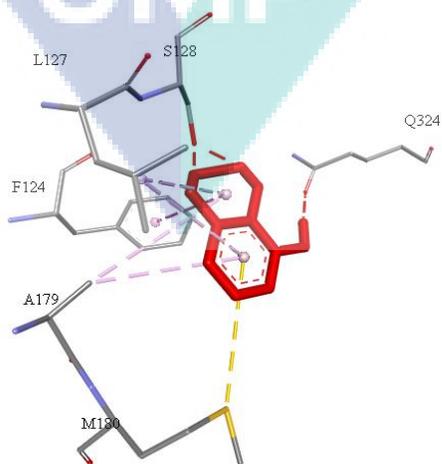
(vi) VMPOOMT-N-methyl-coclaurine complex



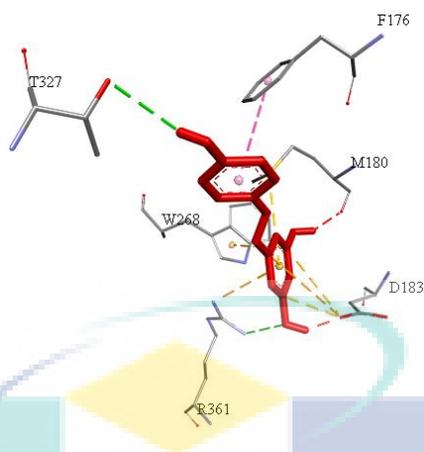
(vii) VMPOOMT-pyrocatechol complex



(viii) VMPOOMT-pyrogallol complex



(ix) VMPOOMT-resveratrol complex



(k) Threading analysis result of related comparative protein models with VMPKAT, VMPBSMT, VMPEGS and VMPOOMT using PSI-BLAST program.

Protein Name	Organism	E-value	Percentage identity	Reference protein template from Protein Data Bank (PDB)
PhKAT	<i>Petunia hybrida</i>	0.0	81%	2wu9
AmBAMT	<i>Antirrhinum majus</i>	1e-98	44%	1m6e
NsSAMT	<i>Nicotiana suaveolens</i>	5e-151	58%	1m6e
PhBSMT	<i>Petunia hybrida</i>	7e-145	57%	1m6e
CbEGS	<i>Clarkia breweri</i>	0.0	100%	3c1o
PhEGS	<i>Petunia hybrida</i>	5e-87	45%	3c1o
GcEGS	<i>Gymnadenia conopsea</i>	4e-63	35%	3c1o
VvEGS	<i>Vitis vinifera</i>	5e-152	63%	3c1o
VvOMT	<i>Vitis vinifera</i>	3e-111	47%	1fp2
RhOOMT	<i>Rosa hybrida</i>	2e-117	48%	1fp2

APPENDIX J

Equations Used in Expression Analysis of the Putative Fragrance-related Transcripts

<p>Equation A (Standard curve)</p> $y = mx + c$ <p>where y = CT values, m = slope of the standard curve, x = log input of template amount, c = y-intercept of the standard curve line.</p>
<p>Equation B (PCR amplification efficiency)</p> $E = (10^{-1/\text{slope}} - 1) \times 100$ <p>where E = PCR amplification efficiency, slope = m as mentioned in Equation A</p>
<p>Equation C (Relative quantity)</p> $Q = E^{(\text{lowest } C_T - \text{sample } C_T)}$ <p>where Q = Relative sample quantity, E = PCR efficiency</p>
<p>Equation D (Normalization expression)</p> $\text{Normalized expression level of target n} = Q_{\text{target n}} / \text{NF}_{\text{target n}}$ <p>where $Q_{\text{target n}}$ = relative sample quantity of target n, $\text{NF}_{\text{target n}}$ = normalization factor of target n</p>
<p>Equation E (Normalization Factor)</p> $\text{NF} = (Q_{\text{sample (ref 1)}} * Q_{\text{sample (ref 2)}} * \dots * Q_{\text{sample (ref n)}})^{1/n}$ <p>Where NF = normalization factor, $Q_{\text{sample (ref 1)}}$ = relative quantity of reference gene 1 (endogenous control), $Q_{\text{sample (ref 2)}}$ = relative quantity of reference gene 2 (endogenous control), n = number of endogenous control genes used</p>
<p>Equation F (Scaled Normalized Expression)</p> $\text{Rescaled normalized expression target} = \frac{\text{normalized expression level}_{\text{target n}}}{\text{normalized expression level}_{\text{calibrator}}}$

LIST OF PUBLICATIONS

Journal publications

1. **Mohd-Aiman, B.**, Mohd-Hairul, A. R. and Tan, S.H. 2015. Identification of fragrance-related transcripts from selected orchids using cDNA Representational Difference Analysis (cDNA-RDA) approach. *IJIMS*. **2(4)**: 198-207.
2. **Mohd-Aiman, B.**, Mohd-Hairul, A. R. and Tan, S.H. 2015. Sequence Analysis of New Isolated O-methyltransferase Transcript (VMPOMT) from *Vanda Mimi Palmer*. *Sch. Acad. J. Biosci.* **3(3)**: 239-247.

Poster presentation

1. **Mohd-Aiman, B.**, Mohd-Hairul, A.R., Tan, S.H., Abdullah, J.O. and Namasivayam, P. (2013) Molecular studies of selected fragrance-related transcripts in benzenoid and phenylpropanoid pathway of *Vanda Mimi Palmer*. Poster Presentation at the National Conference on Industry-academia Initiatives in Biotechnology (CIA-BIOTECH13) on 5-7th December 2013 at Equatorial Hotel, Cameron Highlands

Sequence submission to NCBI GenBank database

Complete mRNA sequences;

1. Keto-acyl-CoA thiolase (VMPKAT) [Accession number: KF278718]
2. Benzoic acid carboxyl methyltransferase (VMPBSMT) [Accession number: KF278719]
3. Eugenol synthase (VMPEGS) [Accession number: KF278720]
4. Orcinol O-methyltransferase (VMPOOMT) [Accession number: KF278721]

cDNA-RDA sequence (EST submission)

1. RDA001h [Accession number: JZ480886]= putative transposase
2. RDA002h [Accession number: JZ480890]= auxin-repressed protein
3. RDA003h [Accession number: JZ480891]= mRNA sequence
4. RDA004h [Accession number: JZ480892]= methionine synthase
5. RDA005h [Accession number: JZ480893]= S-adenosyl methionine synthetase
6. RDA006h [Accession number: JZ480894]= putative gag-pol polyprotein
7. RDA007h [Accession number: JZ480895]= methionine synthase
8. RDA008h [Accession number: JZ480887]= phenylalanine ammonia-lyase
9. RDA009h [Accession number: JZ480888]= mRNA sequence
10. RDA0010h [Accession number: JZ480896]= clathrin assembly protein

11. RDA0011h [Accession number: JZ480889]= stearyl-acyl carrier protein desaturase

