SEQUENCING AND ANALYSIS OF A cDNA ENCODING A PUTATIVE COCLAURINE N-METHYLTRANSFERASE (CNMT) FROM *Aristolochia fimbriata*, A BASAL ANGIOSPERM

DISSERTATION

Submitted To

THE FACULTY OF LIFE & ENVIRONMENTAL SCIENCES
UNIVERSITY OF PESHAWAR

In fulfillment for the degree

Of

DOCTOR OF PHILOSOPHY

by

ROSHAN ALI

THE CENTER OF BIOTECHNOLOGY AND MICROBIOLOGY
UNIVERSITY OF PESHAWAR
2010

CENTER OF BIOTECHNOLOGY AND MICROBIOLOGY
UNIVERSITY OF PESHAWAR
SEMESTER 2010

This Dissertation is Submitted
By
ROSHAN ALI

Titled

SEQUENCING AND ANALYSIS OF A cDNA ENCODING A PUTATIVE COCLAURINE N-METHYLTRANSFERASE (CNMT) FROM *Aristolochia fimbriata,* A BASAL ANGIOSPERM

In fulfillment of the requirements for the degree

Of

DOCTOR OF PHILOSOPHY

In

BIOTECHNOLOGY

APPROVAL

The dissertation of Mr. Roshan Ali is approved

SUPERVISOR:

Dr. Ghosia Lutfullah — Signature:_____

Professor, — Date:_____

Center of Biotechnology & Microbiology,

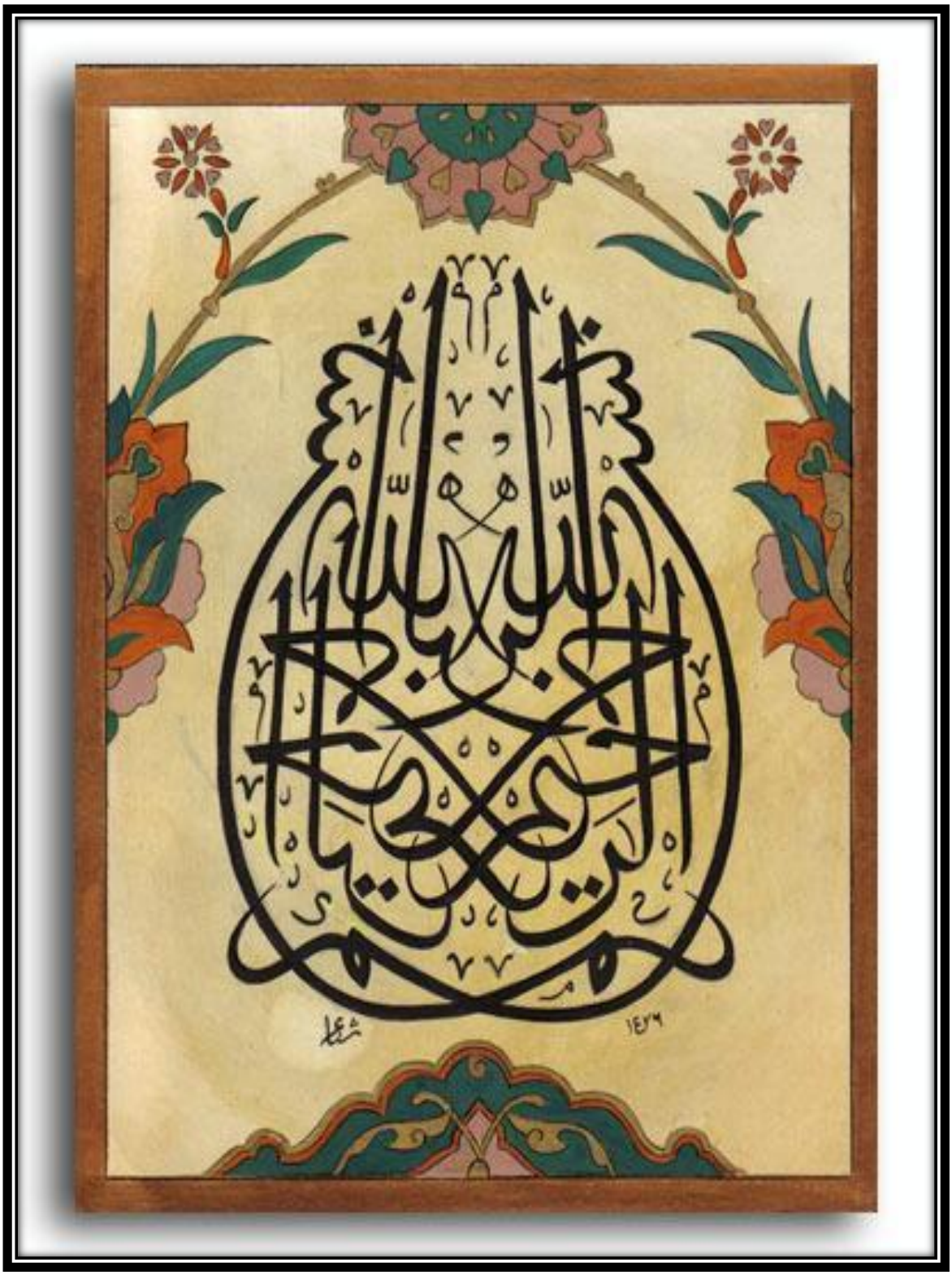
University of Peshawar.

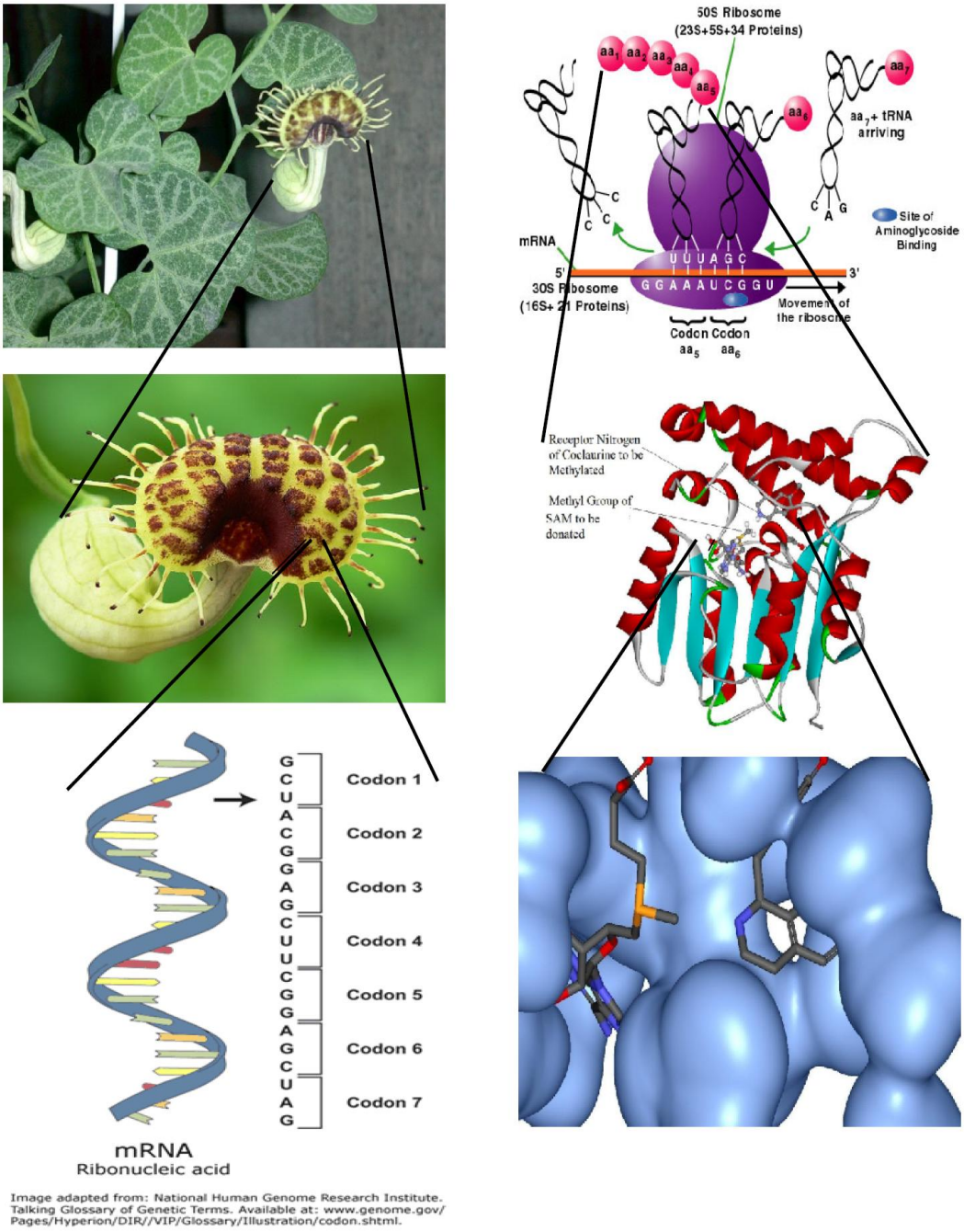
EXAMINER: — Signature:_____

Date:_____


Dedicated to
Our Beloved Prophet Muhammad (SAWWS), His Ahlul Bayt (AS)
and
Their Great Companions

In the name of Allah, the Beneficent, the Merciful





Graphical Abstract of the dissertation:

Image sources are: (http://1.bp.blogspot.com/_vKxDQBwCntA/S-DMy0e3gVI/AAAAAAAAAJo/e2ickYGXEas/s1600/2516243964_5e41ba8d14.jpg), (http://images19.fotki.com/v27/photos/6/642761/4094024/Aristolochiafimbriata37-vi.jpg?1182176542), (http://www.scripps.edu/chem/wong/PIX/ribosome.jpg) and this dissertation.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

To begin with, all praise and thanks to Allah, God Almighty, most Beneficent and Merciful for giving me tremendous courage and power to complete this research and dissertation work. May Allah shower his blessings on Muhammad (SAWWS) and his Ahlulbait (AS), who always have been the sources of knowledge and guidance for humanity.

I am greatly thankful to my supervisor Prof. Dr. Ghosia Lutfullah, Professor at the Center of Biotechnology and Microbiology, University of Peshawar, Pakistan for her great help, nice advices, guidance, and encouragement.

I am happy to acknowledge my deep sense of gratitude to the Director, Center of Biotechnology and Microbiology Prof. Dr. Bashir Ahmad for his great help, inspiring encouragement and overwhelming support.

I wish to express my deep gratitude to my ever smiling Professor Dr. Claude dePamphilis for his excellent guidance, constant encouragement and affectionate patronage throughout the experimental work conducted at dePamphilis laboratory, Department of Biology, The Huck Institute of Life Sciences, University Park, The Pennsylvania State University, USA under his kind supervision. The experimental work embodied in this dissertation was carried out at his laboratory. I am also thankful to him for providing me with all the facilities in the laboratory and financial support.

This work would not have been possible without the fellowship award from the Higher Education Commission, Pakistan for which I am thankful to the Government of

Pakistan and particularly grateful to Dr. Atta-u-Rehman who devised this fellowship program.

I wish to express my special thanks to Dr. Sayed Awlad Hussain and Dr. Sayed Mussarat Hussain for their sincere guidance, valuable suggestions, inspiring encouragement, fruitful discussion, nice advices and kind prayers. I would like to pay my special regards to their family, Baba buzargawar and their forefathers.

I wish my gratitude to the research fellows Kerr Wall, Paula Ralph, Lena Landherr and Laura Warg of USA, Nadira Naznin and Bindhu Abinah of India, Yan Zhang, Yuannian Jiao of China, Norman Wickett of Canada in the dePamphilis laboratory for their great help, cooperation, sincere wishes, invaluable assistance and for their nice company during my stay in USA. I am also thankful to Dr. Ahmad, Zaid, Najeem, Khalid, Wael, Jouke Postma, Maria Postma and their daughter Leizbeth for their beautiful company during my stay in USA.

My expressions of gratitude and sincere thanks are due to Mr. Zahid Khan, Mr. Abid Ali Khan, Aziz-ur-Rehman, Ashfaq Ahmad, Sadiq Azam, Ibrar Khan, Sajid Ali, Johar Jameel, Riaz Khan and other research fellows who helped me immensely with their fruitful discussions, valuable suggestions and continuous cooperation.

I am thankful to Faculty members Dr. Nafees Bacha, Mr. Akhtar Hussain, Mr. Noor Muhammad, Mr. Jamshid Ahmad, Mr. Abu Nasar Sidiqui, Mr. Momin Khan and Mr. Khalid Javed lecturers at the Center of Biotechnology and Microbiology, University of Peshawar, for their sincere wishes and help during my course work.

Great contribution has been made in polishing my dissertation by Dr. Claude, Paula Ralph, Lena Landherr, Yuannian Jiao. I would convey my thanks to Prof. Dr. S. Mustafa and Dr. Rasool Khan for their guidance, encouragement and cooperation.

I am also thankful to all my friends Shuja-u-Din, Aftab Khan, Ashiq Hussain, Nabi Hassan, Muzamil Hussain, Mudasir Hussain, Wahab Ali, S. Mahboob Ali Shah, Izhar Ali, Anwar Ali, Iqbal Hussain and Javed Hussain for their nice company, prayers and encouragement.

I wish to express my special thanks to Mr. Javid Hussain and Mr. Shafaat Hussain for their help, cooperation, encouragement, nice company and financial support.

I am thankful to all my relatives Noor Alam Jan, Jabir Hussain, Muhammad Ayub, Aftab Hussain, Kamal Hussain, Javid Hussain, Amjad Hussain, Muhammad Ibrahim, Khalid Hussain, Wajid Hussain, Kiramat Hussain, Amir Hussain, Abdullah Jan, Faras Ali, Asad Ali, Tanvir Hussain, Mehdi Hussain, Muhammad Zulqarnain, Hadi Hussain, Muhammad Saki, Qamar Abbas, Ansar Hussain, Ihtisham, Basit Ali, Hassanain, Awais, Nishat Hassan, Waqar Hussain, Anis, Mahmood Ali, Maroof Ali, Ashfaq Hussain, Banaras Hussain, Qaim Hussain, Khatib Hussain and Komail Hussain for their kind prayers.

Finally I would like to forward a special word of gratitude to my very supportive mother, my father, Muslim Jan, and my sisters for their enormous cooperation throughout my life, especially my mother whose love, prayers and moral support enabled me to continue with perseverance.

Roshan Ali

ABSTRACT

ABSTRACT

Alkaloids are produced in plants through various pathways involving several enzymes that lead to diverse alkaloids. One of the most important alkaloid biosynthetic enzymes is coclaurine N-methyltransferase (CNMT) which is an S-adenosyl-L-methionine-dependent methyltransferase (SAM-MTase). SAM-MTases utilize S-adenosyl-L-methionine (SAM) as a cofactor to methylate other molecules. CNMT catalyzes the methylation of coclaurine. Crystal structures of more than hundred SAM-MTases have been investigated. Several O-methyltransferases have been characterized at the molecular as well as structural levels, but there have been very few molecular studies of N-methyltransferases especially about CNMTs.

In this study, the amino acids sequence of *Aristolochia fimbriata* putative CNMT has been determined by isolating and translating the full-length cDNA. In order to investigate the mechanism of methylation by this putative CNMT, three-dimensional homology model has been built and the ligand (SAM) as well as the substrate (S-Coclaurine) has been docked into its active site. Phylogenetic analyses were performed using the MEGA 4.0 software. The phylogenetic relationship of *A. fimbriata* putative CNMT with their homologs has also been analyzed. In order to identify the putative CNMT gene and determine its function, online similarity searches were performed by BLAST program using the cDNA sequence as well as the putative protein that could be encoded by the gene. All the methods, applied, predicted that the gene identified might be involved in the production of CNMT.

The predicted homology model consists of two domains: the N-terminal catalytic core domain and the C-terminal domain. The catalytic core domain has a central sheet of β-strands surrounded by α helices. The catalytic core domain contains binding site for

SAM. The C-terminal domain consists of alpha helices and a few beta sheets creating a pocket for the substrate in between them. The SAM-binding pocket is located next to substrate binding pocket and there is an opening in between these two cavities through which the methyl group of SAM projects towards the substrate. The most important residues involved in the methyl transfer reaction seem to be Tyr-79 and Glu-96.

LIST OF ABBREVIATIONS

ABBREVIATION	FULL WORD
3D	Three-dimensional
4OMT	4-O-methyltransferase
6OMT	6-O-methyltransferase
7OMT	7-O-methyltransferase
ADT	Autodocktools
APSSP	Advanced Protein Secondary Structure Prediction server
BBE	Berberine bridge enzyme
BIA	Benzylisoquinoline alkaloid
CASTp	Compound atlas of surface topology of proteins
CFAPS	Cyclopropane fatty acyl phospholipid synthase
CFS	Cheilanthifoline synthase
CJNCS1	2-Oxoglutarate-dependent dioxygenase
CJPR10A	PR10-like protein
CNMT	Coclaurine N-methyltransferase
COR	Codeinone reductase
CYP719	P450-dependent monoxygenase
CYP80B3	P450-dependent monoxygenase, (S)-N-methylcoclaurine-3-hydroxylase
DBOX	Dihydrobenzophenanthridine oxidase
EA	Ergot alkaloid
EST	Expressed sequence tag
HNN	Hierarchical neural network
MACPS	Mycolic acid cyclopropane synthase

MP	Maximum parsimony
MSH	N-methylstylopine 14-hydroxylase
MTase	Methyltransferase
NCS	Norcoclaurine synthase
NJ	Neighbor joining
NMCH	(S)-N-methylcoclaurine-3-hydroxylase
NMR	Nuclear magnetic resonance spectroscopy
NMT	N-methyltransferase
OMT	O-methyltransferase
P6H	Protopine-6-hydroxylase
PDB	Protein Data Bank
Pfam	Protein families
RMSD	Root mean square deviation
SalAT	Salutaridinol 7-O-acetyltransferase
SalR	Salutaridine reductase
SalSyn	Salutaridine synthase
SAM/Adomet	(S)-Adenocyl-L-methionine
SAM-MTase	SAM-dependent methyltransferase
THS	Thebaine synthase
TIA	Monoterpenoid indole alkaloid
TNMT	Tetrahydroprotoberberine cis-N-methyltransferase
TYDC	Tyrosine decarboxylase
STS	Stylopine synthase

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

INTRODUCTION

INTRODUCTION

Coclaurine N-methyltransferase (CNMT) [1-3], involved in the biosynthesis of benzylisoquinoline alkaloids (BIAs), is a unique S-adenosyl-L-methionine (SAM)-dependent N-methyltransferase (NMT) that catalyzes the transfer of a methyl group from SAM to the amino group of the tetrahydrobenzylisoquinoline (a sub-type of BIA) alkaloid coclaurine. This enzyme is considered to be very important because N-methylation of coclaurine enhances the 4-O-methylation of 3-hydroxy-N-methylcoclaurine-4-O-methyltransferase and enables the sequential metabolic conversion [4, 5]. Studies have shown that CNMT is non-stereospecific and has wide substrate specificity. It can also catalyze the transfer of methyl group to several alkaloids structurally similar to coclaurine such as (R)-coclaurine, (S)-norcoclaurine, (R, S)-6-O-methylcoclaurine, (R, S)-norlaudanosoline and (R, S)-norreticuline etc [1, 6].

Whereas several O-methyltransferases (OMTs) have been characterized at the molecular [7-9] and structural levels [10-12], little is known about N-methyltransferases (NMTs) [13, 14]. CNMT has been purified [1] and the corresponding cDNAs have been isolated from *Coptis japonica* [15], *Papaver somniferum* [16] and *Thalictrum flavum* [17]. Tetrahydroprotoberberine-NMT (TNMT) has been isolated from *Corydalis vaginans*, *Eschscholzia californica* and *Sanguinaria canadensis* [18, 19]. However CNMT has not yet been reported in *Aristolochia* species, which is a very useful genus to study evolutionary biology.

Aristolochia has been selected as a model for our project because it is an important medicinal plant and occupies an important position in the phylogeny of angiosperms. As a basal angiosperm from the magnoliid clade [20], it offers opportunities for the

study of developmental evolution in flowering plants. Developmental pathways that are conserved evolutionarily, such as pathways that control the floral organ identities in angiosperms, have been identified through functional studies in plants, such as *Arabidopsis, Oryza*, *Petunia* and *Zea* [21-24]. Comparative studies of these plant species have led to the hypotheses about the evolution of gene function in flowering plants [25, 26]. Studies have shown that monocot and eudicot lineages diverged only about 113–133 million years ago, some 28–48 million years after the angiosperm divergence [27, 28]. It has been proposed that if biochemical pathways found in core eudicots and/or monocots are also found in a basal angiosperm that would imply that the conserved regions were present in angiosperms before the origin of the two major classes of angiosperms [29]. It has been suggested that further establishment of *Aristolochia* as a basal angiosperm experimental model would be of great benefit for evolutionary biology [29].

Genome sequencing provides us with a large database that can be used in developing amino acid sequences of proteins of living organisms. To get full benefits out of these genome sequences, we need to understand the function of the proteins encoded by these genomes. This task can be achieved by the knowledge of the three-dimensional (3D) structure of the proteins [30, 31]. But unfortunately, the structures of only a small number of proteins [32] have been solved by X-ray crystallography or nuclear magnetic resonance spectroscopy (NMR) as compared to the large number of sequences [33-35]. Therefore, we need to predict the structure of most protein sequences by computation [36]. Among the two approaches of protein structure prediction, the ab-initio method predicts the structure from sequence alone [37], but is not a successful method due to large number of errors in the structures created through this method. Another method of protein structure prediction, including

threading and comparative modeling, make use of similarity between the modeled sequence and at least one known structure [38]. In this study, a cDNA from *A. fimbriata,* a medicinally important basal angiosperm, has been isolated and identified to be a putative CNMT. Crystal structure and mechanism of reaction have been explained for some methyltransferases (MTases) [39-42]. However three-dimensional crystal structure and the mechanism of reaction have not yet been explained for CNMTs. In this study, a 3D homology model of the *A. fimbriata* putative CNMT has been built by comparative modeling, as well as a putative mechanism has been proposed of methyl transfer in the reaction on the basis of previously proposed mechanisms for other enzymes. As no crystal structure was available for any known CNMT, a 3D homology model of this putative CNMT was predicted by using other MTases as templates. The present study also includes the phylogenetic analysis of the *A. fimbriata* CNMT with homologous sequences of other plants.

1.1. S-Adenosyl-L-Methionine-Dependent Methyltransferases

CNMT is a subclass of SAM-dependent methyltransferases (SAM-MTases) [E.C.2.1.1.-], which are a diverse and important class of enzymes that methylate proteins, lipids, nucleic acids and other small molecules [43, 44]. SAM-MTases are key enzymes in cellular biochemistry because they are involved in important biological processes, including protein trafficking and sorting, metabolism, biosynthesis, signal transduction and gene expression. SAM-MTases belong to the class II methyltransferases [45]. Some members of this class show strict substrate specificities, however there are others that are quite promiscuous [46]. All SAM-MTases possess a conserved SAM-binding domain consisting of a seven-stranded β-sheet flanked by three α-helices per side of the sheet. A substrate-recognition domain

usually happens near the SAM-binding domain [43]. However, this domain is decidedly variable, which is steady with the fact that these enzymes methylate a variety of substrates of different kinds.

MTases exhibit a great variety in their structures, binding sites and the conformations of SAM/S-adenosylhomocysteine (SAH). MTases have been divided into five classes on a structural basis. The similarities of the sequence among the MTases are very low even within the same class [47].

1.2. S-Adenosyl-L-Methionine (SAM or AdoMet)

All CNMTs use SAM to methylate their substrates. SAM, also known as AdoMet, is a conjugate of the nucleotide (adenosine), and the amino acid (methionine) [48]. SAM is an important metabolic intermediate, and all cellular organisms have many SAM-binding enzymes. One of the several roles of SAM is to provide methyl group for covalent alteration of different substrates, such as oxidized arsenic, chloride and iodine ions [49-51], tRNA, rRNA and vital proteins. The methylation status of SAM can function as a regulatory signal for the maturation and control of various types of interactions with other molecules [52-54]. Methyl transfer is just one of the several biochemical reactions requiring SAM. There are several other enzymatic, as well as non-enzymatic, reactions where SAM is used to donate a methyl group to a long list of potential acceptors in transmethylation reactions.

1.3. Genes of Bezylisoquinoline Alkaloid Pathway

Modern innovations in genomic technologies have improved the pace of discovery of new alkaloid biosynthetic genes. A number of methyltransferase cDNAs of BIAs of biosynthetic pathway have been isolated [55, 56], including the berberine bridge

enzyme (BBE) [57] a methylenedioxy bridge-forming P450 monooxygenase (CYP719) [58], the salutaridinol 7-O-acetyltransferase (SalAT) [59] and codeinone reductase (COR) [60]. Two cDNAs encoding *P. somniferum* norcoclaurinesynthase (NCS) isoforms have been isolated from an elicitor-treated cell culture expressed sequence tags (EST) database on the basis of resemblance to *T. flavum* NCS [61]. In spite of only 39% matching identity between *P. somniferum* and *T. flavum* NCS, both are concerned with the catalyses of the formation of norcoclaurine. Lately, cDNAs encoding a 2-oxoglutarate-dependent dioxygenase (CjNCS1) and a PR10-like protein (CjPR10A) that is 62% identical to *T. flavum* NCS have been isolated and characterized from *C. japonica* [62]. All the identified BIA methyltransferase orthologs are well conserved between diverse species. Comparative transcript and alkaloid profiling of morphine-producing and morphine-free *Papaver* species led to the isolation of corresponding cDNAs for salutaridine-synthase (SalSyn) and salutaridine reductase (SalR) [63]. The exposition of morphine biosynthesis in opium poppy has recently included the partial purification and functional characterization of thebaine-synthase (THS), which converts salutaridinol-7-O-acetate to thebaine, which was previously thought to happen non-enzymatically [64]. Cognate cDNAs for two enzymes involved in the conversion of (S)-scoulerine to (S)-cis-N-methylstylopine have been isolated from *E. californica* corresponding to CYP719A2 and CYP719A3, both of which exhibit steroid-sulphatase activity [65]. A full-length cDNA, encoding Tetrahydroprotoberberine cis-N-methyltransferase (TNMT), has also been isolated in an opium poppy elicitor treated cell culture EST collection [66], representing how the conservation of BIA methyltransferases can be used to discover new BIA biosynthetic genes. Another enzyme sanguinarine reductase (SanR) has also been isolated from *E. californica* cell cultures [67]. A molecular clone that encodes a P450-dependent C–C

coupling enzyme corytuberine synthase (CTS) has been isolated from *C. japonica* [68].

It is very difficult to isolate molecular CNMT based on the structural similarity of MTases because of the very minute differences found in the primary structures, including the SAM binding site of NMTs and OMTs, reported to date. Therefore, cDNA of the CNMT was amplified and isolated based on homology to the mRNA sequence of a purified enzyme. Based on this observation, the amino acids sequence of the *A. fimbriata* putative CNMT was determined by isolating the full-length cDNA. The nucleotide sequence of cDNA suggested that the deduced amino acid sequence showed similarity to known NMTs especially CNMTs of other plants.

1.4. Alkaloids

CNMT is involved in the catalyses of transmethylation reaction in the BIA synthesis pathway. Here, it is important to briefly describe about alkaloids and the BIA synthesis pathway. Alkaloids are low-molecular-weight, nitrogen-containing, physiologically active compounds that have been found in approximately 20% of plant species and which are synthesized mainly in higher plants [69]. It has been investigated that alkaloids function as defense compounds as it is apparent from their biological activity [70].

Approximately 200,000 compounds have been identified that are believed to be produced by a variety of biosynthetic pathways in plants [71]. These compounds include more than 25,000 terpenoids, about 8,000 phenolic compounds, roughly 12,000 alkaloids and many other types of chemicals such as flavonoids, lignans, quinines and steroids [70], which are supposed to play key roles in adaptation of

plants to the environment. Such secondary metabolites have also been used as drugs, dyes, flavors, food additives, fragrances and insecticides due to their chemical and pharmaceutical characteristics [70, 72]. Thus, plant alkaloids are among the major groups of plant compounds synthesized in nature. Most of the alkaloids are derived from amino acids, such as lysine, ornithine, tryptophan and tyrosine.

Many higher plants synthesize numerous diverse types of alkaloids, and like uric acid and urea in animals, it was proposed that alkaloids were produced as nitrogenous wastes and had little or no significance for host plants [73]. However, there is some data available about alkaloids that propose them to be involved in protecting plants against herbivores and microbial pathogens [74], whereas others, such as nicotinic acid [75] may act as precursors of vitamins.

The discoverer of the word "alkaloid" (German: *Alkaloide)*, German chemist Carl F. Wilhelm Meissner (in 1818), realized that active chemicals of lethal plants are alkaline in character, and thus can be isolated from plants by using techniques extraction. Individual alkaloids such as atropine from *Atropa belladonna* L., and hyoscyamine from *Hyoscyamus niger* L., and later scopolamine were isolated from plants in 1830 [76-78]. Ethno-pharmacology traditionally included various plants whose extracts were used throughout the human history as magic potions and poisons of a variety of types that today are known to contain tropane alkaloids, including *Atropa mandragora* (deadly mandrake), *A. belladonna* (nightshade), *Scopolia carniolica* (scopola), *Datura stramonium* (thornapple), and others [79]. In today's market, above 40% of all pharmaceuticals are natural products. From 1941 to 2002, over 50% of all the drugs, accessible for cancer treatment were derivative of natural sources [80].

1.4.1. Types of Alkaloids

The categorization of the alkaloids is complex and is generally guided by a set of rules that take into account the structure and other chemical characteristics of the alkaloid molecule, its biogenetic basis, as well as biological origin [81, 82]. There are three main types of alkaloids according to Britanica online (http://www.britannica.com/EBchecked/topic/15672/alkaloid):

i. True alkaloids: These have nitrogen-containing heterocyclic ring, derived from amino acids.
ii. Proto alkaloids: These do not have a heterocyclic ring containing nitrogen. These alkaloids are also derived from amino acids.
iii. Pseudo alkaloids: These have a nitrogen-containing heterocyclic ring. However, these alkaloids are derived from Purines or Terpenoids.

As it has been stated that CNMT catalyzes an important step in the BIA biosynthetic pathway; therefore, it is necessary to have a brief overview of BIAs and the pathway that lead to the synthesis of different BIAs.

1.4.2. Benzilisoquinoline Alkaloids (BIAs)

BIAs, a group of about 2,500 compounds, is a large class of plant alkaloids [83, 84]. These alkaloids are derivative of tyrosine and are further modified by an intricate scheme of biosynthetic enzymes. These are not compulsory for normal growth and development of plants, but seem to function in the protection of plants against herbivores and pathogens [85]. Primarily BIAs are found in basal angiosperms (i.e. Ranunculales) and in the eumagnoliids [61], including the Berberidaceae, Fumariaceae, Papaveraceae, Ranunculaceae [86] especially *Papaver* species that

accumulate different types of BIAs [87] and infrequently throughout the order Piperales [61]. The most well-known and important species of genus *Papaver* is the opium poppy. Many BIAs have pharmacological activity including the antimicrobial, berberine; the vasodilator, papaverine; the analgesic, morphine; the antitussive, codeine and the muscle relaxants, tubocurarine and papaverine [83, 84].

Opium poppy (*P. somniferum*), an extensively cultivated medicinal plant, produces more than 80 benzylisoquinoline alkaloids, including the antimicrobial, sanguinarine, and the narcotic analgesics, morphine and codeine. Sanguinarine is an effective inhibitor of fungal and bacterial growth, and hence, protects the plant against pathogens [88]. This property of sanguinarine has prompted its use as an antiplaque agent in oral hygiene products [89].

In the BIAs pathway, protoberberine-type and aporphine-type alkaloids are produced through (S)-reticuline from tyrosine. (S)-reticuline is the branch-point intermediate alkaloid in the biosynthesis of several types of BIAs, and is also a non-narcotic alkaloid of pharmaceutical importance used in the development of anticancer and antimalarial drugs [90]. New studies have also recommended these alkaloids as useful medicines. For example, magnoflorine has been revealed to save from harm of high-density-lipoprotein (HDL) during oxidant stress to stop the development of atherosclerotic disease and to inhibit human lymphoblastic cell-killing by human immuno deficiency virus [91-93]. The antimicrobial agent, berberine, has been reported to have cholesterol-lowering properties [94].

The BIAs can be classified into a variety of groups: aporphines, benzophenanthridines, true benzylisoquinolines, morphinans, pavines, phthalideisoquinolines, protoberberines and rhoeadines.

Due to the pharmaceutical value of BIAs, their biosynthetic pathways have been studied extensively, and many of them have been understood completely at the enzyme level [70, 95].

1.4.3. BIA Producing Pathway

Alkaloid synthesis and accumulation is a constitutive, cell type-specific process in plants. For example, aerial organs mostly contain noscapine, morphine and papaverine, whereas roots generally accumulate sanguinarine [96]. Alkaloid biosynthetic enzymes are localized to sieve elements of the phloem, and the cognate transcripts have been associated with companion cells of sieve elements [97, 98].

It is known that several BIAs share a common biosynthetic pathway [Fig. 1.1] from L-tyrosine to (S)-reticuline. (S)-reticuline is an important transitional chemical that functions as a central precursor for different types of BIAs [70, 87].

1.4.4. Common Pathway

Studies have shown that a variety of oxidative steps in this pathway are catalyzed by cytochrome P450 enzymes [5, 58, 65, 99, 100] whose members are found in a number of species, particularly in the plant kingdom [101, 102], and lots of them are involved in the secondary metabolism in plants [103].

The basic pathway of BIAs biosynthesis up to (S)-reticuline has already been investigated at the enzymatic level and cDNAs have been isolated for all the enzymes involved in various plant species [4, 8, 84, 104-106]. The steps at the start of this biosynthetic pathway are general to all the members of these classes that lead to the central intermediate (S)-reticuline. The first step in BIAs biosynthesis is the

condensation of dopamine and 4-hydroxyphenylacetaldehyde by NCS [62, 105, 107] to yield tetrahydrobenzylisoquinoline alkaloid; (S)-norcoclaurine [83, 84].

Formation of dopamine involves the decarboxylation of tyrosine and/or dihydroxyphenylalanine by tyrosine/dopa decarboxylase (TYDC) [Fig. 1.1] [104]. A series of enzymes is involved in the common pathway that leads to the formation of (S)-reticuline from (S)-norcoclaurine. The enzymes involved are norcoclaurine-6-O-methyltransferase (6OMT), CNMT, (S)-N-methylcoclaurine-3-hydroxylase (NMCH), and 3-hydroxy-N-methylcoclaurine-4′-O-methyltransferase (4OMT) [108]. 6OMT methylates the 6-hydroxy position of (S)-norcoclaurine to (S)-coclaurine, which is then N-methylated by CNMT, yielding (S)-N-methylcoclaurine [15, 56]. The cytocrome P450-dependent monoxygenase, (S)-N-methylcoclaurine-3-hydroxylase (CYP80B3), catalyzes the 3-hydroxylation of (S)-N-methylcoclaurine [99] prior to the formation of (*S*)-reticuline by 4OMT [4, 16, 84]. (S)-Reticuline represents the last common intermediate in the biosynthetic pathway that leads to the production of both morphine and sanguinarine [86].

From this step onward the pathway divides into diverse benzylisoquinoline classes and a number of modifications and rearrangements of the benzylisoquinoline backbone takes place leading to a large number of structurally different alkaloids. For example, (S)-scoulerine is formed by the formation of an oxidative C–C bond between the N-methyl group and carbon-2 of the benzyl unit, giving rise to protoberberines, phthalideisoquinolines and protopines, which are mainly found in species of the Berberidaceae, Menispermaceae, Papaveraceae and Ranunculaceae. These alkaloids can further be modified to benzophenanthridines and papaverrubines [83, 109].

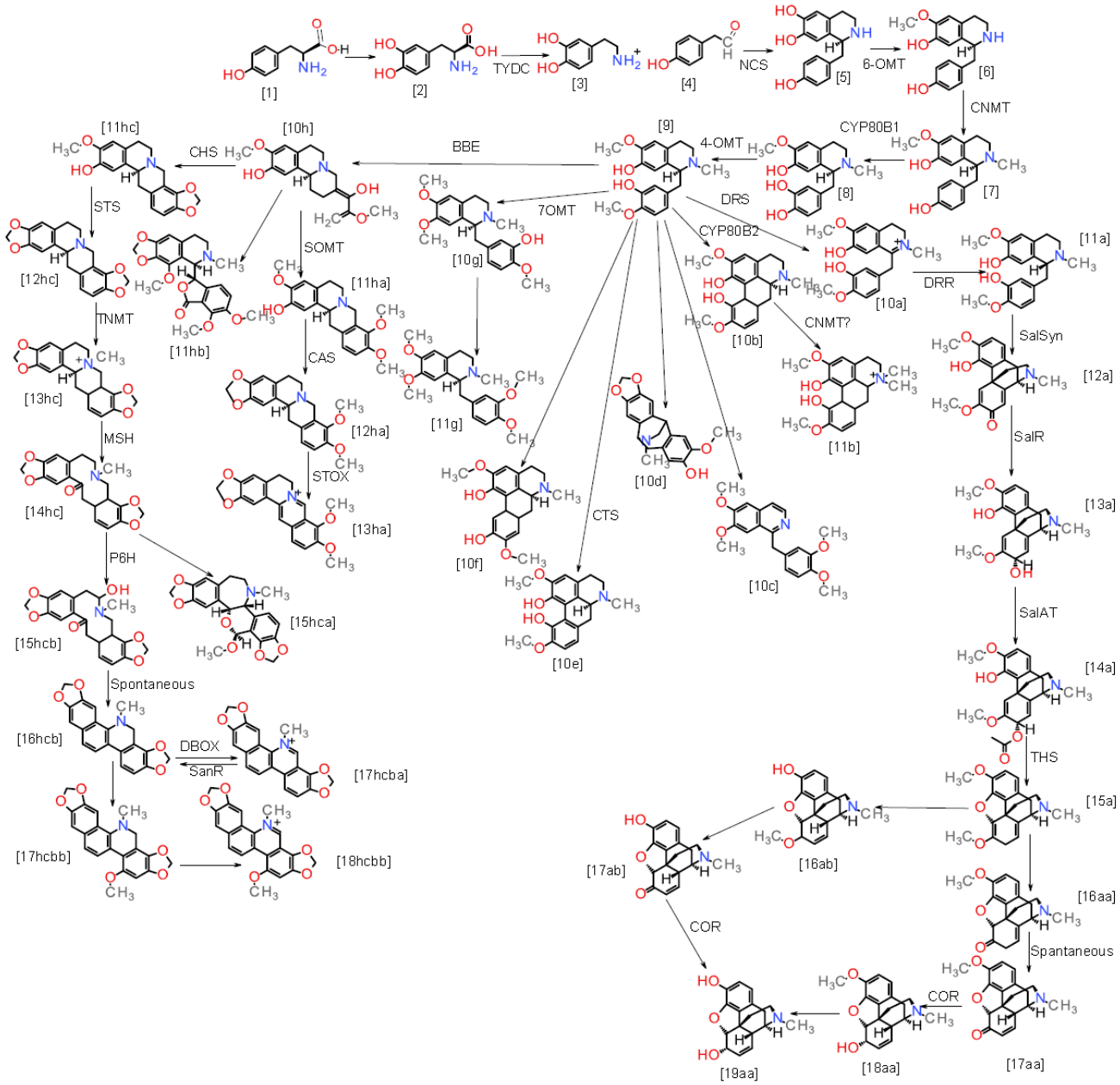


Fig. 1.1: The benzylisoquinoline alkaloids pathway: The pathway and all the structural formulae were created using Symyx-Draw v3.3 (www.symyx.com) combining data from several pathways [5, 58, 83, 99, 107, 108]. The abbreviations used have been given in the list of abbreviations as well as in the text. The alkaloids have been given number codes in brackets. These codes have been explained in APPENDIX.

Very little information is available about the pathways downstream from (S)-reticuline. For the biosynthesis of benzophenanthridines and protoberberines, the metabolic flow has been studied and enzymes have been isolated in partially purified form [83]. Several cDNAs of this pathway have been isolated, such as (S)-scoulerine-9-O-methyltransferase, BBE, (S)-canadine-Synthase, columbamine-O-methyltransferase and (R,S)-reticuline 7-O-methyltransferase (7OMT) [8, 55-58]. (S)-reticuline can also produce derivatives, such as laudanine, by a methylation process with the help of 7OMT [108]. There are other pathways leading to a vast variety of alkaloids but we will limit our discussion to sanguinarine and morphine. Sanguinarine and morphine biosynthesis has been clarified at the enzyme level [83], as explained below.

1.4.5. Sanguinarine Pathway

The first step in the pathways that leads to the production of several kinds of BIAs is catalyzed by BBE, which is concerned with the methylene bridge formation in (S)-reticuline to yield (S)-scoulerine [108]. The next step involves the formation of two methylenedioxy bridges in (S)-scoulerine by the stylopine synthase (STS) and cheilanthifoline synthase (CFS), to form (S)-stylopine [65, 110, 111]. (S)-Stylopine is converted to (S)-cis-N-methylstylopine by TNMT [66], which is subsequently hydroxylated by N-methylstylopine 14-hydroxylase (MSH) [108, 112]. The product of MSH tautomerizes to protopine, which is then hydroxylated to produce dihydrosanguinarine by protopine 6-hydroxylase (P6H) [113], later oxidation by dihydrobenzophenanthridine oxidase (DBOX) give rise to sanguinarine [Fig. 1.1] [114].

1.4.6. Morphine Pathway

Due to the high economic significance of the morphinan branch of the BIAs pathway [Fig. 1.1], it has been investigated intensively for several past years. Several enzymes involved in the pathway have been purified partially or completely, and the encoding cDNAs have been cloned [59, 115-117]. The pathway leading to the morphinans is started by the conversion of (S)- to (R)-reticuline, which is catalyzed in a reaction with two steps that includes the oxidation of (S)-reticuline by 1, 2-dehydroreticuline synthase (DRS) to the 1, 2-dehydroreticulinium ion and following reduction of the ion to (R)-reticuline by 1, 2-dehydroreticulinium ion reductase (DRR) [108]. In the next step, salutaridine is produced by the carbon–carbon phenol coupling between C-12 and C-13 of (R)-reticuline by the enzyme SalSyn. 7-(S)-salutaridinol is formed by the reduction of the keto group at C-7 by stereospecific, NADPH-dependent enzyme SalR [118]. Salutaridinol is converted to salutaridinol-7-O-acetate by the acetyl coenzyme-A-dependent enzyme SalAT [115]. The first pentacyclic alkaloid of the morphinan group, thebaine, is formed by the elimination of the acetyl residue with the formation of an oxide bridge [64, 83]. The last steps of morphine biosynthesis can take place through two different routes: either thebaine is oxidized to codeinone via neopinone, and is later reduced by codeinone reductase (COR), yielding codeine, or thebaine is first demethylated to oripavine and then to morphinone, which is reduced to morphine by COR [84].

1.4.7. Medicinal Value of Alkaloids

The use of plants as medicinal agents can be traced back thousands of years, and certainly, today's modern pharmaceutical industry has its roots in the physiologically active secondary metabolites of these plants. This section reviews some of the varied

biologically active and medicinally significant alkaloids that have been isolated from diverse flowering plants, animals, fungi, and bacteria.

The alkaloids have key pharmacological uses. For example, the muscle relaxant, bisbenzylisoquinoline tubocurarine alkaloid, is employed in modern medicine during abdominal surgery; the anti-tussive, codeine; the analgesic, morphine; the antimicrobials, berberine and sanguinarine [119]. Strychnine is usually used as an important component in Chinese herbal-medicines to treat arthritic and traumatic, pains and diseases related to central nervous system [120]. Colchicine, found in *Colchicum autumnale*, has been used as a treatment for arthritis and gout [121, 122]. Ephedrine has been mostly utilized as central nervous system stimulant, an antitussive, an antipyretic and an anti-inflammatory in various Chinese traditional medicines [123]. Aconitine has excellent efficacy against rheumatosis and some other inflammative diseases. Central to the biosynthesis of all of these alkaloids is the tri-oxygenated intermediate (S)-N-methylcoclaurine [124] that on hydroxylation produces (S)-reticuline, which then leads to structurally diverse classes of alkaloids such as sanguinarine, berberine and morphine by subsequent regio- and stereo-specific oxidation of (S)-reticuline. Plants produce various types of alkaloids; some of them have been discussed below.

The Rutaceae family chiefly produces acridone alkaloids [125, 126]. For example, various parts of *Atalantia monophylla* have been used as folk medicine for several purposes, such as treatment of paralysis [127], as a stimulant and for hemiplegia [128]. The essential oil from the leaves has shown antimicrobial and strong inhibitory activities against some pathogenic fungi [129], whereas a decoction of the leaves is often applied for itching and other skin complaints [128]. Acridone alkaloids have

also been isolated from the root bark of *A. monophylla* [130, 131]. Acridone alkaloids have many biological activities such as induction of differentiation in human promyelocytic leukemia cells (HL-60) [132], induction of the inhibition of Epstein-Barr virus (EBV)-EA [133], and anti-proliferative activity [134].

The Piperaceae family has great potential for medicinal uses. Several *Piper* species have been reported to contain a large number of natural compounds including amides and alkaloids [135]. *P. nigrum* has been reported to be an analgesic, central nervous system stimulant and an antipyretic [136]. The leaves of another species, *P. sarmentosum* have been used to treat coughs, malaria and toothache. This species has also been shown to have cytotoxic [137], antituberculosis [138] and antifungal activities [139].

Papaver (Papaveraceae), a genus with global distribution, is rich in morphinone and isoquinoline alkaloids [87]. For example, *P. nudicaule* L., a widely distributed species in the Mongolian–Siberian area of Central Asia [140], has commonly been used in Mongolian folk medicine as a sedative, antitussive and respiratory regulative remedy, as well as for headaches caused by nervous disorders, and for acute and chronic inflammation of stomach [141]. *Papaver* species have been investigated to contain an abundance of alkaloids [87]. The “sleep-inducing” property and the medicinal value of its latex have also been known to human beings for ages [142]. The plant has been reported to contain about eighty different tetrahydrobenzylisoquinoline-derived alkaloids, including the analgesic and narcotic drug, morphine [143], broadly used for the control of pain [144]; the muscle relaxant, papaverine; the cough suppressant, codeine; and the antitumoric agent, noscapine [145]. Morphine and sanguinarine have been reported to function as part of the chemical defense system for *P. somniferum*

[146]. Benzophenanthridines have attracted interest for their anti-microbial and anti-oxidant properties [147]. For example, sanguinarine has been recorded to treat leukemia [148].

Argemone mexicana L. (Papaveraceae), a bitter, purgative and diuretic spiny herbaceous plant distributed throughout India, has been used to kill intestinal worms, and as cures for leprosy, diarrhea, inflammation, dysentery and various skin diseases [149]. A number of berberine, benzylisoquinoline and benzophenanthridine alkaloids have been isolated from this plant [150-153].

Doryphora sassafras (Atherospermataceae), commonly known as Sassafras, have been reported to produce 1-(4-hydroxybenzyl)-6,7-methylenedioxy-2-methylisoquinolinium trifluoroacetate, which has been reported to possess antimalarial activities without any cytotoxicity for a human embryonic kidney cell line [154].

Erythrina crista-galli L. (Fabaceae), a popular ornamental plant in subtropical areas, has been verified to have diuretic, hypertensive and sedative activities [155]. The seed extracts of some species possess a curare-like effect on the central nervous system [90]. These properties of the plant are due to the Erythrina-type alkaloids that they contain.

Roots and rhizomes of *Psychotria ipecacuanha* (Rubiaceae), a medicinal plant native to Central and South America, have been used as an anti-amebic and an emetic [156]. These medicinal effects derive from its principal alkaloid, emetine, occurring together with a number of related alkaloids, such as ipecoside and cephaeline, collectively called Ipecac alkaloids [157].

Madagascar periwinkle and *Catharanthus roseus* (Apocynacrar) produce variety of monoterpenoid indole alkaloids (TIAs) in different organs. The leaves and stems are the sources of vincristine and vinblastine, which are used in a number of anticancer chemotherapy drugs [158]. Their roots gather antihypertensive alkaloids, such as serpentine [159].

The azaoxoaporphine alkaloid sampangine (aporphine family of alkaloids) have been reported to show antitumor, antifungal, antibacterial and antiviral activities [160-165]. Sampangine is mainly produced by the plant family Annonaceae [166-168]. Sampangine exhibits a very strong inhibitory activity against the human fungal pathogens [166, 168].

Poisonous Solanaceae family have been reported to contain many alkaloid-producing species [79] that were in use in ancient times as folk medicine in a variety of ethnic groups [169]. These plants synthesize tropane alkaloids, which constitute a group of medicinally useful, potentially toxic, and very addictive substances [170].

Another important group of alkaloids, the steroidal glycoalkaloids (SGAs), are also found in a number of species of the genus *Solanum* (Solanaceae), both edible and poisonous [171]. Several toxic effects, such as fruit poisoning due to unripe ackee, and Jamaican vomiting sickness, have been attributed to these alkaloids [172].

The tobacco alkaloid, nicotine, has been shown to slow down apoptosis of neoplastic and normal cell lines, resulting in tumor formation [173], and also causes DNA damage in epithelial and non-epithelial cells [174].

Plants in the genus *Daphniphyllum* contain daphniphyllum alkaloids [175, 176]. The species of this genus in the family Daphniphyllaceae have long been used as

medicines. For example, the leaves and seeds of *D. calycinum* have been used in traditional medicine in China to treat influenza, fever and inflammation [177]. Daphniphyllum alkaloids also show cytotoxic activity against many tumor cell lines [178-180].

Alkaloids are not only produced by plants, but also by animals. Alkaloids found in skins of anurans were first discovered in a Colombian poison-dart frog (*Phyllobates aurotaenia*) of the neotropical family Dendrobatidae [181]. After this discovery, several other alkaloids were identified in various species of dendrobatid frogs, collectively known as ‘‘dendrobatid alkaloids’’ [182]. In 1984, such alkaloids were also found in a bufonid toad (*Melanophryniscus moreirae*), a myobatrachid toadlet and two mantellid frogs [183]. Later, various alkaloids were found in the skin extracts of several other species of toads [184, 185]. All the skin alkaloids of the mantellid, dendrobatid and bufonid anurans are sequestered from dietary alkaloid-containing arthropods [186-189], but the skin pseudophrynamines of the myobatrachid frogs (Pseudophryne) appear to be synthesized by the frogs [190]. The over 800 alkaloids detected in amphibian skin extracts have recently been summarized, along with their occurrence in anuran families, and putative dietary arthropod sources [191].

Several phytopathogenic fungi that belong to the genus *Claviceps* synthesize ergot alkaloids (EAs), which are derivatives of prenylated tryptophan. Due to the structural similarity of some EAs to the neurotransmitters dopamine and serotonin, affinity to the cognate receptors of the central nervous system [192, 193] have long been used for the treatment of various disorders of the peripheral and central nervous systems [194].

A large family of isoprenoid quinoline alkaloids, the aurachins, has been found in myxobacteria [195, 196].

A new derivative of the natural alkaloid ellipticine, 8-methyl-4-(3-diethylaminopropylamino)-pyrimido-[4′,5′:4,5]-selenolo(2,3-b)quinoline, has been shown to posses cytotoxicity to leukemic cell lines by inducing necrosis in them [197].

1.5. *Aristolochia*

Aristolochia belong to order Piperales, which is one of the most species-rich clades among basal angiosperms, comprising about 3,300 species [198]. Piperales includes the families Aristolochiaceae, Hydnoraceae, Lactoridaceae, Piperaceae, and Saururaceae [199, 200]. This order contains approximately all forms of plant life, including lianas, geophytes, succulents, herbs, shrubs, trees, epiphytes and parasites. Members of Piperales demonstrate a diverse spectrum of specializations in the morphology of their flowers, and pollination. The flowers of Aristolochiaceae are highly specialized which are helpful in attracting insects. Saururaceae and Piperaceae have perianthless flowers which are pollinated mostly by bees and flies [201-203]. Secondary metabolites of Aristolochiaceae (e.g. aristolochic acids) are chief phamacological compounds [204].

Most of the species of Aristolochiaceae [205] are tropical or subtropical. Although generic circumscription within the family has been in dispute for about two centuries [206, 207], recent authors have documented four genera divided into two subfamilies: Aristolochioideae and Asaroideae, each with two genera [208-210]. The subfamily Asaroideae is characterized by an actinomorphic perianth, and has two genera: the

monotypic *Saruma,* and *Asarum*. The second subfamily, Aristolochioideae includes *Thottea*, with an actinomorphic perianth, and *Aristolochia*, the largest genus of the family.

The genus *Aristolochia* consists of approximately 500 species, mostly found in tropical, subtropical, and Mediterranean regions [206]. Flowers of *Aristolochia* have a monosymmetric perianth and are highly derived to function as a trap for insects in order to ensure pollination [211]. The flowers produce a stinky odor, which attracts predominantly fly pollinators [212, 213]. *Aristolochia* s. str. has been divided into two subclasses based on their morphology: subclass *Podanthemum* (unilabiate flowers with a striped utricle) and *Aristolochia* (unilabiate or bilabiate flowers with a sessile utricle). Molecular based phylogenies suggest that *Podanthemum* and *Aristolochia* are monophyletic [214] sister groups [215]. The old world distribution of the two subclasses is similar in that both are found in Africa and Asia. However the African species of the two groups are in distinctly different areas. *Podanthemum* is found south of the Sahara, while *Aristolochia* is found in North Africa, specifically, Algeria, Morocco and Tunisia. In the Mediterranean and the adjacent Near East, only members of *Aristolochia* are found. The distribution of *Aristolochia* starting from Asia towards Europe can be assumed.

The extensive use of *Aristolochia* species in herbal medicine makes its study essential. Fresh investigations have shown that some of the secondary compounds (e.g. aristolochic acids) are nephrotoxic and carcinogenic. Therefore many countries have banned the supply and use of herbal medicines that use extracts from *Aristolochia*.

Aristolochia fimbriata, a flowering and perennial plant, has been cultivated for its interesting flowers that attract butterflies, and for conventional medicinal uses. The flower of *A. fimbriata* has a monosymmetric, unipartite perianth adapted for pollination by insects, while other genera in Aristolochiaceae have radially symmetric flowers [216].

1.5.1. Medicinal Value of *Aristolochia*

Aristolochia has considerable medicinal significance and has long been used in medicines. Secondary metabolites of *Aristolochia* are critical for the survival of butterflies during their larval feeding stage [217-219].

Earlier studies of *Aristolochia pubescens* have led to the isolation of five neolignans, seven lignans, two diterpenes, two sesquiterpenes, eight aristolochic acids and three aristolactams [220-222]**.** Various lignans, aristolochic acids and diterpenes of the plant play a very important role in chemical defense against insects [223-226]. Several biflavonoids and a tetraflavonoid have also been isolated from the stem of *Aristolochia ridicula* [227]. Reports on the natural distribution of tetraflavonoids are very rare, and are limited to *Lophira alata, A. ridicula* (Ochnaceae) [228, 229] and *Cephalotaxus wilsoniana* (Cephalotaxaceae) [230].

A. fimbriata has been reported to possess aristolochic acid and the structurally related alkaloid, magnoflorine [231-233]. They have been reported to be the major toxic components found in *Aristolochia* plants. Aristolochic acid has been found to be carcinogenic and nephrotoxic [234-236].

Aristolochia species have widely been used as abortifacients, anodynes, antiasthmatics, antiophidians, expectorants and stomachics in Brazilian traditional

medicine and, recently, in slimming therapies as an alternative for traditional Chinese herbs [237, 238].

Some species are used to stimulate uterine contractions [239]; others have antimitotic [240] and antiviral activity [241].

The aerial parts of *Aristolochia constricta* (Saragosa), have been used in folk medicine as an anticancer agent, an antimalarial, an antispasmodic, an emmenagogue, and in the treatment of snake bites [242-244]. Some parts of *A. fimbriata* have been used in the treatment of asthma, cold and fever [245].

Renal failure, due to an excessive use of *A. manshuriensi*, has been reported in Japan and China [246, 247]**.** Aristolochic acids of *A. manshuriensis* have been reported to cause the nephrotoxicity called Chinese herb nephropathy or aristolochic acid nephropathy [248, 249]**.** Luckily, healthcare practitioners have begun to be cautioned against the use of *Aristolochia* species (therapeutic goods administration, (Tga). Health and aged care. Published on August, 2000. Available online from: http://www.quackwatch.org/01Quackery RelatedTopics /DSH/aristalochia.pdf).

The anti-inflammatory activity of aristolochic acid led to the development of pharmaceutical preparations containing aristolochic acid in Germany in the 1970s, until it was shown that aristolochic acid was a strong carcinogen [250]. In 1982 all pharmaceutical preparations that contained aristolochic acid were removed from the market in Germany and a number of other countries.

Aristolochic acids are now known for their nephrotoxic effects as well as for their carcinogenic properties [251]. Herbal remedies containing species of the genus *Aristolochia* have now been classified as carcinogenic to humans by the International

Agency for Research on Cancer (IARC) [252]. However, it has been reported that *Aristolochia* plants are still in use as traditional medicine in some parts of the world [234].

1.6. Aim of the Present Study

A. fimbriata is a medicinally as well as an evolutionary important basal angiosperm. It has long been used in medicines. *Aristolochia* species have widely been used as abortifacients, anodynes, antiasthmatics, antiophidians, expectorants and stomachics in Brazilian traditional medicine and, recently, in slimming therapies as an alternative for traditional Chinese herbs. But due to the presence of aristolochic acids it has been found to be carcinogenic and nephropathic. Aristolochic acid is one of the toxic compounds produced by this plant. Aristolochic acid is synthesized through a special pathway in *Aristolochia* species that involve several intermediate compounds. CNMT is an important enzyme involved in catalyzing the methylation of coclaurine in BIA biosynthetic pathway. CNMT not only catalyzes coclaurine methylation, but it can also catalyze the methylation of (R, S)-norlaudanosoline, an important intermediate of aristolochic acid biosynthesis. No data is available about the presence of CNMT in *A. fimbriata.* Keeping in view the above points, the present research work aims at the identification, sequencing and analysis of a cDNA from *A. fimbriata* that encodes for a putative CNMT, its homology modeling and analysis so that the mechanism of reaction catalyzed by this enzyme could be explored. CNMT can be used in the methylation of several organic compounds once its three-dimensional structure and mechanism of methylation is understood. As no crystal structure is available for any CNMT known, therefore a theoretical model is necessary to be built, using the homology modeling so that its mechanism of reaction could be analyzed. Another

important investigation includes the phylogeny of CNMTs to better understand the evolutionary relationship of *A. fimbriata* CNMT with CNMTs from different groups of plants and an Alga. As has been stated that establishment of *Aristolochia* as a basal angiosperm experimental model would be of great benefit for evolutionary biology.

MATERIALS AND METHODS

MATERIALS AND METHODS

2.1. Tissue Source

Flower buds, young fruits, roots and young leaves from *A. fimbriata*, grown under standard greenhouse conditions [253] at the Pennsylvania State University, USA, were collected in different sizes [254]. The tissues were then quick frozen immediately, using liquid nitrogen, and stored at -80°C.

2.2. RNA Extraction

Total RNA was extracted from each of the tissues collected according to the manufacturer's protocol for the RNAqueous®-Midi Kit (Ambion, Inc., Catalog # 1911; Available on: www.ambion.com/jp/techlib/prot/bp_1911.pdf) [253] or a modified CTAB method [255].

2.2.1. Ambion Method

For the Ambion kit RNA isolation method, approximately 200 mg of each tissue was ground separately in RNase-free pestles and mortars pre-chilled with liquid nitrogen. The ground tissue was then mixed in lysis buffer that was already prepared by mixing with "Plant Isolation Aid®" (Ambion, Inc., Catalog # 9690) in a ratio of 8:1 to improve yields. The solution was then vortexed. The solution was centrifuged by spinning at 12,000g at 22°C for 10 minutes. The clarified lysate was mixed with equal volume of ethanol (64 %). This mixture was filtered by passing it through a glass fiber filter and washed with an equal volume of Wash solution # 1 and 70% volume of Wash solution # 2/3. To elute the RNA from the filter, a 0.5 ml preheated Elution solution was passed through the filter three times. A total of 1.5 ml eluate was

collected in a collection tube. The RNA was precipitated overnight using 3 ml ethanol (100 %) and 0.1 ml sodium acetate. After centrifugation, the RNA pellet was washed with cold 70% ethanol and resuspended in RNase-free diethylpyrocarbonate (DEPC)-treated water and stored in a freezer at -80°C [253].

2.2.2. Modified CTAB Method

Total RNA was extracted using a modified version of the cetyl trimethyl ammonium bromide (CTAB) protocol developed by Chang et al. (1993) [255], except that two to three grams of frozen tissue, was ground in an RNase-free, chilled mortar and pestle under liquid nitrogen and suspended in warm (65ºC) CTAB buffer which was made fresh same day using RNase-free stock solutions.

2.3. Quality and Purity of the Total RNA

Determination of quality and purity of total RNA was done by micro-capillary electrophoresis on the Agilent-2100 Bioanalyzer according to the manufacturer's suggested protocol (RNA 6000 Nano Kit and Assay Protocol; Available online: http/www.chem.agilent.com/Library/usermanuals/Public/G2938-90034_KitRNA6000 Nano_ebook.pdf). The 18S and 28S ribosomal peaks were used to determine the quality and quantity of RNA.

2.4. Messenger RNA (mRNA) Isolation

Total RNA extracted from the each tissue was treated with RNase free DNase 1 (Invitrogen DNase I, Amplification Grade, Catalog # 18068-015). Messenger RNA (mRNA) was isolated from the total RNA using the manufacturer's protocol for the poly(A)purist^{TM} mRNA purification Kit (Ambion, Inc, Catalog # 1916; Available

online: http://www.ambion.com/jp/techlib/prot/bp_1916.pdf). Two poly-(A) purist columns were used for the purification of the total RNA. Each column was loaded with 800 μg of total RNA to be purified. The purified RNA was precipitated with ethanol, resuspended in the RNA storage solution (Ambion, Inc, Catalog # 1916), and stored in the freezer at -80°C.

2.5. Unigene Selection from *A. fimbriata* EST Database

2.5.1. *Papaver somniferum* CNMT Gene Sequence Information

Complete CDS of *P. somniferum* S-adenosyl-L-methionine: coclaurine N-methyltransferase mRNA (GenBank Accession Number: AY217336) [16] was obtained from Genbank database through the NCBI database (http://www.ncbi.nlm.nih.gov/genbank/).

2.5.2. Tribe Identification of the *P. somniferum* CNMT Gene and Unigene Selection

In order to identify all potential CNMT genes in *A. fimbriata*, the known CNMT gene in *P. somniferum* was blasted against the PlantTribes database [http://fgp.huck.psu.edu/] for homologous genes (or tribe) in the published genomes [256]. All genes in the identified potential CNMT gene cluster (tribe) were then used as queries to blast against *A. fimbriata* EST database (http://ancangio.uga.edu/ancangiodb.html) to identify the other CNMT paralogous genes. Unigenes of *A. fimbriata* with stringency value of five were downloaded from the database. All the Unigenes were searched for a start and stop codon to find out the complete and longest ORF. The one with complete and longest ORF i.e. CL957Contig5, was selected for further analysis.

2.6. Analysis and Amplification of Gene Expression by Reverse Transcriptase RT-PCR

2.6.1. Polymerase Chain Reaction

cDNA was amplified from mRNA using the manufacturer's protocol for PROMEGA Kit (PROMEGA Access RT-PCR System, Catalog # A1250; available online: http://www.promega.com/tbs/tb220/tb220.pdf). A master mixture of 50 µl was prepared by mixing, 2 µl of total RNA solution (25 ng/µl), 10 µl of AMV/Tfl reaction buffer, 1 µl of dNTPs (10 µM each), 1 µl each forward and reverse primers (50 µM each), 2 µl $MgSO_4$, 1 µl AMV reverse transcriptase, 1 µl Tfl DNA polymerase (5 µ/µl) and 31 µl of nuclease-free water. PCR was run on a thermal cycler (Eppendorf mastercycler gradient) using the following program: For reverse transcription; 48°C for 45 min (once only); 94°C for 2 min (once only), the amplification was done for 45 cycles using the program: 94°C for 30 s, 48.2°C or 49.8°C (gradient) for 1 min and 70°C for 2 min. The mixture was left in thermal cycler for 10 min at 68°C to complete the reaction. Primers for PCR amplification of the complete CDS sequence were designed from CL957Contig5 using MacVector (Cambridge, U.K.). The forward primer used was 5′-CAT GGC GTC TGA AAA GCT C-3′ and the reverse primer was 5′-CCT ATT TTT TCT TGA AAA GCA GAT G-3′. The amplicon was expected to be 1076 bp.

In order to analyze and visualize the PCR product, samples were run on a 2 % agarose gel, stained with ethidium bromide, and viewed using a Bio-Rad Gel Doc XR using the image software (Quantity One version 4.6.6; Available online: http://legacy.lclark.edu/~lycan/Bio312/Quantity%20onerev%27d08.pdf). The expected bands obtained on the agarose gel were extracted from the gel with the help

of QIAquick® Gel Extraction Kit (50), (Cat. # 28704) using the protocol provided with the kit (Available online: http://www1.qiagen.com/literature/ render.aspx?id=103715). We tried to sequence the PCR product, but it was not pure enough and resulted in a bad sequence probably due to multiple copies were present in the band extracted. Therefore the PCR product was further purified as explained below.

2.6.2. Gene Cloning

Amplified PCR product was cloned into pCR®II-TOPO® TA vector system (Invitrogen) and One Shot® TOPO10 chemically competent *E. coli,* (Cat. No. C4040-10), using the manufacturer's protocol (Available online: http://tools.invitrogen.com/ content/sfs/manuals/topota_qrc.pdf) with minor modifications. A mixture was prepared by mixing 2 µl fresh PCR product, 1 µl salt solution, 1 µl TOPO vector and 1 µl water (sterile), to make the volume 5 µl, was prepared and incubated for the night at room temperature. The *E. coli* was thawed on ice at -80°C for One Shot® chemical transformation and 2 µl of the TOPO® cloning reaction was supplemented and incubated for 30 min on ice. The cells were then heat-shocked for 30 sec at temperature of 42°C and 250 µl SOC medium was added to the cells. The cells were incubated at 37°C for 1 hour. Prewarmed LB plates (100 mm) containing 50 µg/ml ampicillin were spread with X-gal (40 mg/ml) and 30 µl of the transformation mixture and incubated all night at 37°C. The next day, 10 white colonies were chosen and grown them during the night in a 5ml liquid culture with ampicilin (50µg/ml). The plasmid was isolated using Qiagen QIAprep Spin, {Miniprep Kit (250) Cat. #27106} according to the manufacturer's protocol (Available online: http://www1.qiagen.com/literature/render. aspx?id=370).

The product was sequenced by The Penn State Genomics Core Facility at University Park (http://www.huck.psu.edu/facilities/genomics-core-up/) using the same primers used in the RT-PCR. The results were analyzed by Sequencher software (ver 4.9). FinchTV (ver 1.4.0: http://www.geospiza.com/Products/finchtv.shtml), a free software, was used to create the chromatogram for reverse complement of reverse chromatogram. Geospiza's FinchTV is the well-liked way to view DNA sequence traces on all the operating systems. FinchTV launched as the only chromatogram viewer that can show a complete trace in a scalable multi-pane view and it leads the way with raw data views, BLAST probing and the facility to reverse complement sequences and traces. The low quality regions were trimmed from both the chromatograms and a manual alignment was made between them. The consensus sequence was then aligned with *A. fimbriata* Unigene CL957Contig5 to check for mismatches. Both the sequences were 100% identical.

2.7. Gene Identification and Prediction of the Function of its Putative Protein

Genes can be identified through several tools and methods [257]. In this study, following methods were used for the identification of gene and its putative function.

2.7.1. Similarity Searches

To determine the function of the gene, identified in the Unigene i.e. CL957Contig5, similarity searches were performed by non-redundant (database nucleotide collection nr/nt) BLASTn ([Table.3.3], tBLASTx [Table.3.4] and BLASTx [Table.3.5] program of the NCBI-BLAST package [256] using the default parameters. BLASTn programs explore nucleotide databases using a nucleotide sequence. In BLASTx the nucleotide sequence is translated in all possible six frames to protein and then the exploration is

performed against the protein sequences. tBLASTx scans translated nucleotide databases using a translated nucleotide sequence.

2.7.2. Sequence Translation

The unigene CL957Contig5 was translated by the "Translate" Tool available on SwissProt database [258]. This tool performs a conceptual translation of a nucleotide (DNA/RNA) query to a protein sequence. It allows the translation of the nucleotide sequences to proteins in all possible six frames.

Similarity search was also performed by non-redundant BLASTp [Table.3.7] of the NCBI-BLAST package, with default parameters, using the translated protein sequence as a query. BLASTp is a package that performs similarity searches against protein sequences of the database using the sequences of protein as query.

ORF (Open Reading Frame Finder: http://www.ncbi.nlm.nih.gov/gorf/gorf.html) was also used to identify the longest ORF of CL957Contig5.

2.7.3. Protein Parameters

ProtParam Tool (http://www.expasy.ch/tools/protparam.html) [259] of the SWISSProt ExPASy database was used to calculate the molecular weight, amino acid composition [Table. 3.8], number of negative and positive charged residues, half-life and instability index of the translated protein sequence (putative).

2.7.4. ModBase Search

The putative protein sequence was BLASTed [Table.3.13] against the ModBase database [260] (http://ModBase.compbio.ucsf.edu/ModBase-cgi/index.cgi) to search

for the structures theoretically modeled. ModBase is a database storage-area for hypothetically predicted models on the basis of homology modeling. This database is run by Ursula Pieper in the group of Andrej Sali, Department of Bioengineering and Therapeutic Sciences and California Institute for Quantitative Biomedical Research, Mission Bay Campus, Byers Hall, University of California San Francisco, San Francisco, CA 94158-2330.

2.7.5. Function Predictions

Function of the putative protein was predicted from the BLAST results and InterproScan [261, 262] analysis of EMBL-EBI database (http://www.ebi.ac.uk/interpro/). InterPro is a database that contains protein domains, families, regions, repeats and sites in which particular features found in previously known proteins can be useful to newly sequenced proteins. Signal peptides were analyzed with Signal-CF (predicting signal peptide and cleavage site) [263] and Signal-3L (a 3-layer approach for predicting signal peptides) [264]. Mem Type-2L, predicts the membrane protein types [265], was used to analyze whether it is a membrane protein or not. Enzyme functional class and subclass analysis was performed with EnzyPred (A top-down approach for the prediction of enzyme functional classes and sub-classes) [266]. Whether the protein is a protease or not was analyzed by ProtIdent [267] that identifies the proteases and their types by fusing the functional domain and sequential evolution information. Subcellular localization of the protein was predicted by Proteome Analyst Specialized Subcellular Localization Server v2.5 [268], WoLF PSORT [269, 270], LOCtree [271] and CELLO v.2.5 (subCELlular LOcalization predictor) [272].

For the identification of the protein AACompIdent [273] and TagIdent [259] were also used by searching the proteins with similar amino acid composition or with pI values and molecular weights in the specified range [Table.3.11]. For AACompIdent the parameters used were IP: 5.21±0.5, Constellation: 0 and Database: Swissprot/TREMBL, and for TagIdent the parameters used were pI Range: 5.0-5.4 and mw: 41325.2 (theoretically calculated by ProtParam) within 20 % range. The pI value of 5.21 was theoretically calculated by ProtParam. AACompIdent is a tool which allows the identification of a protein from its amino acid composition. This tool is used to search the SWISS-Prot and/or TrEMBL databases for proteins, whose amino acid compositions are closest to the given amino acid composition.

TagIdent, a tool that generates a listing of proteins nearer to a given value of pI and mw, identifies the proteins by matching a short sequence tag of upto six amino acids in length against proteins in the UniProt knowledgebase (Swiss-Prot and TrEMBL) databases near to a given pI and mw, and also identifies the proteins by their mass, if this mass has been determined by mass spectrometric techniques for one or more species and with an optional keyword.

Phylogenetic tree was also used in identifying the function of the protein. For this purpose BLASTp was performed against the non-redundant protein database using *A. fimbriata*'s putative CNMT as query. The first top one hundred best hit sequences were downloaded from NCBI. Additional seven sequences were obtained by BLASTing against PDB database including sequences of proteins used as templates for homology modeling. Proteins annotated as unknown, putative, and predicted were removed from these hundred sequences except for the putative CNMTs of *Arabidopsis* and *Oryza*. The multiple sequence alignment of the remaining thirty-

eight sequences was performed using CLUSTALx [274, 275]. The remaining sequences were split into six groups: coclaurine N-methyltransferases (CNMT: 6 sequences), cyclopropane fatty acyl phospholipid synthases (CFAPS: 15 sequences), mycolic acid cyclopropane synthases (MACPS: 7 sequences), methyltransferase type 11 (4 sequences) sequences, SAM-dependent methyltransferases (SAM-MTases: 2 sequences) and tetrahydroprotoberberine N-methyltransferases (TNMT: 3 sequences). First, each group was aligned separately, and then the six groups were aligned gradually using the profile alignment mode of CLUSTALx. In the end, the sequence obtained from *A. fimbriata* was also aligned with the aligned profiles. Gap columns were removed from the final alignment. The final alignment was used in generating the phylogenetic tree. Phylogenetic tree was constructed using MEGA 4.0 [276, 277] with "complete deletion" option and "poisson correction" [278] model by the Neighbor Joining (NJ) method [279]. The tree was rooted with MACPS (2fk7 from *Mycobacterium*).

For the protein family and domain identification Pfam [280] database was searched by using the sequence search option.

2.7.6. Secondary Structure Prediction

Secondary structure predictions were performed using different methods; Viz. HNN (Hierarchical Neural Network) at NPS@ (Network Protein Sequence Analysis) [281], APSSP (Advanced Protein Secondary Structure Prediction Server) [282] and JPred-3 (http://www.compbio.dundee.ac.uk/www-jpred), a Secondary Structure Prediction Server [283, 284] of SwissProt. @TOM 2.1 (http://atome.cbs.cnrs.fr/AT2/meta.html), an online server for homology modeling [285], was used to align the *A. fimbriata* putative CNMT protein sequence with the templates. Secondary structure of the

putative CNMT protein sequence of *A. fimbriata* was compared with the secondary structures of templates (PDB, IDs: 1kpiA [286], 1kph [286] and 2fk8A [287]). After the idenfication, the cDNA sequence of the *A. fimbriata* putative CNMT was submitted to the Genebank (Accession number: HQ343195).

2.8. Homology Model Building

Comparative modeling techniques were employed to gain insight into the structural characteristics of the putative CNMT protein of *A. fimbriata.* Molecular modeling is a technique which is used to predict homology models in case where experimentally solved crystal structure does not exist. It can compute good and best quality structural model for a wide range of applications. The method, used to develop the putative CNMT model, has been divided into four steps: template identification and selection, sequence alignment, model generation and evaluation.

2.8.1. Template Identification

To perform similarity searches for template selection, the putative CNMT sequence of *A. fimbriata* was BLASTed [Table.3.14] against the PDB database using program PSI-BLASTp of NCBI-BLAST package with default parameters. Three crystal structure coordinates of mycolic acid fatty acyl cyclopropane synthase from *M. tuberculosis* (PDB id: 1kpi, PDB id: 1kph both crystallized at 2.65 A° and 2fk8 crystallized at 2.00 A°) were selected as the templates because of their highest homology with the target sequence and similarity of their secondary structures as predicted by @TOME v2.1 homology modeling server. The putative protein sequence of *A. fimbriata* shows 30 %, 27% and 25 % identity with the template sequences of

1kph, 1kpi and 2fk8 respectively. The 3D structure coordinates of 1kph, 1kpi and 2fk8 were obtained from Brookhaven Protein Databank (PDB) [288, 289].

For the identification of a best template the *A. fimbriata* putative CNMT sequence was aligned with the selected templates sequences. A tree was generated using Maximum Parsimony (MP) [290] to identify *A. fimbriata* closer relative among all the best hit templates searched by BLAST search against PDB. The tree identified that 1kpi was a closer relative.

2.8.2. Target-Template Alignment

Alignment of the *A. fimbriata* sequence with the template sequences was created with the help of @TOME 2.1. This tool aligns the sequences on the basis of their sequence similarities as well as secondary structure similarities for homology modeling. The meta-server, named @TOME v2.1, allows us to submit an amino acid sequence to the embedded software devoted to similarity searches, structural predictions and fold recognition. @TOME v2.1 facilitates us to recognize a suitable 3D template and the best automatic alignment according to an important set of evaluation tools results. This tool also has the ability to create 3D model by homology modeling or a 3D complex by comparative docking of a ligand between the experimental protein and model. The alignment created by this tool was then manually edited and the SAM ligand was included in the alignment file from the template 2fk8.

2.8.3. Model Generation

MODELLER 9v7 [291, 292] was used to build the three-dimensional homology models of the *A. fimbriata* putative CNMT using 1kph, 1kpi and 2fk8 as templates. MODELLER is a comparative modeling program that calculates and predicts a

refined 3D homology model of a protein based on a given sequence alignment and template/templates. MODELLER employs probability density functions that are derived logically using the statistical mechanics and empirically using a database of known protein structures as the spatial restraints rather than energy [293]. MODELLER infers distance and angle constraints from the crystal coordinates, the template structure and coalesce them with the energy terms for sufficient stereochemistry in an objective function which is optimized afterwards in Cartesian space with conjugate gradients and the methods molecular dynamic.

2.8.4. Best Structure Selection and its Evaluation

Reliabilities of the predicted homology models were assessed by PROCHECK [294] and ProSA [295]. Evaluation of the predicted models included analysis of the geometry, stereochemistry and energy distribution. The stereochemistry of the models was evaluated with PROCHECK. The energy graphs were calculated with the help of ProSA. The best model was selected on the basis of PROCHECK and ProSA results [Table.3.16 and 3.17]. Z-score values of the putative CNMT model and templates were calculated with ProSA-web [296] for comparison purposes.

In order to examine any alteration in the Cα backbone of the model, it was superimposed onto the templates 1kpi, 1kph and 2fk8 crystal structure coordinates separately with the help of SUPPERPOSE command of DS-visualizer. Superimposition allowed us to calculate the root mean square deviation (RMSD) values for positional differences between equivalent atoms of the model and templates. All protein structures and models were visualized and analyzed with the help of DS Visualizer® (v. 2, Accelrys Software Inc). The topology of the best model

was then compared with the topology of available theoretically modeled CNMTs (downloaded from ModBase database).

2.9. Active Site Identification

The binding site residues of the protein were identified and assessed using CASTp (Computed Atlas of Surface Topology of Proteins) [297]. CASTp is a program that helps in detection and characterization of the active sites, binding sites of proteins, functional residues situated on the protein surface and voids covered under the interior of proteins, by measuring concave surface regions of 3D structures of proteins. It also calculates the area and volume of the active site by solvent accessible surface model.

Similarly the active site residues of the modeled protein and the active site residues of the top five identical theoretically modeled CNMTs were also analyzed on ModBase, a database of annotated comparative protein structure models and associated resources. The conserved residues of the model were also analyzed by the Conserved Domain Feature [298] of NCBI after BLASTp.

2.10. Docking

On the bases of similarity of the putative protein of *A. fimbriata* with SAM-dependent CNMTs, it was also predicted to be a SAM-dependent CNMT.

The three-dimensional structures of (S)-coclaurine (the substrate) was downloaded from Chemspider database (Chemspider Building Community for Chemists. Available From: http://www.chemspider.com). ChemSpider is a free access service that gives a structure centric population for the chemists. ChemSpider is the richest single source of structure-based chemistry information that provides access to many

of the chemical structures of compounds and integration to a number of other online services.

The SAM molecule was already docked into the protein model from the template structure of 2fk8 with the help of MODELLER. Blind docking of (S)-coclaurine into the protein was performed with the help of online web server PatchDock [299, 300]. 3D homology model (PDB format) of the *A. fimbriata* protein sequences and 3D structure of (S)-coclaurine (PDB format) were uploaded and submitted on the server. The results were obtained through email. Top 20 docking models were downloaded. Some of the poorly docked models were rejected. These 20 models were analyzed for the active site for (S)-coclaurine. A rough estimation of potential active sites for the (S)-coclaurine was drawn from models. In autodocking the grid-box around these potential active sites was created according to these estimations.

Docking of the substrate (S)-coclaurine in the active site of the predicted protein model of *A. fimbriata* putative CNMT was carried out using the AutoDock4 and Autogrid4 [301] with the help of graphical user interface; AutoDockTools (ADT) [302]. The molecules were prepared for AutoDock4 and AutoGrid4 with the help of ADT using the default parameters, Lamarkian genetic algorithm with local search, 25 million energy evaluations (Medium) per run at population size of 150, the number of GA runs was changed to 100 instead of 10 as a default. The protein was held rigid during the docking process while the ligand was allowed to be flexible. Number of rotatable bonds in the (S)-coclaurine was set to five. The grid-box size was 104 points × 122 points × 108 points in the *x*, *y* and *z* dimensions. The grid-box was centered with the values of 92.713, 38.928 and 9.126 for X_center, Y_center and Z_center. The grid-box was set with AutoGrid4 and then the docking was performed

with AutoDock4. AutoGrid4 and AutoDock4 were run from the graphical user interface of ADT. The results were analyzed with the help of analyze option of ADT.

AutoDockTools is one of the three applications of MGLTools (Molecular Graphics Laboratory) of The Scripps Research Institute for visualization and analysis of the molecular structures. ADT is a graphical interface for setting up and running AutoDock. It's a free software.

AutoDock is a suite of computerized docking tools. It is intended to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure. AutoDock actually consists of two main programs AutoDock and AutoGrid. AutoDock performs the docking of the ligand to a set of grids describing the target protein while AutoGrid pre-calculates these grids. In addition to using them for docking, it can also be used for the atomic affinity grids visualization. This can help organic synthetic chemists to design better binders, for example.

AutoGrid is a program that pre-calculates grid maps of interaction energies for various atom types, with a macromolecule such as a DNA/RNA or a protein. AutoDock then uses these grid maps for docking calculations from where it can conclude the total interaction energy for a ligand with the given macromolecule.

2.11. Analysis of Ligand and Substrate Binding Interactions with the Active Site Residues

LigPlot [303] was used to analyze ligand and substrate binding residues of the protein and the interactions of ligand and substrate with the protein. Ligplot, written by Andrew Wallace and Roman Laskowski, is software that generates automatically the schematic diagrams of protein-ligand interactions for a given PDB file.

2.12. Phylogenetic Analysis

To find out the closer relatives among CNMT producing plants, the *A. fimbriata* putative CNMT was aligned with the pre-aligned CNMTs (six sequences).

Trees were constructed using MEGA 4.0 [276] with “complete deletion” option also applied in the case of Maximum Parsimony (MP) analysis. MP tree was obtained using the Close-Neighbor-Interchange algorithm [304]. To evaluate the reliability of internal branches, a bootstrap test of 1000 replicates [305] was carried out. Branches corresponding to partitions reproduced in less than 50% of the bootstrap replicates were collapsed. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches [305].

RESULTS

AND

DISCUSSION

RESULTS AND DISCUSSION

3.1. Qualitative Assessment of Extracted Total RNA

Total RNA was extracted from different parts of the plant. Several samples were prepared from each tissue. The quality of the RNA was evaluated by microcapillary electrophoresis using the Agilent Bioanalyzer. The RNA from all the tissues yielded fairly sharp peaks corresponding to the expected sizes of the 18S and 28S RNAs. The baseline was almost flat for all the tissues. The ratio of 18S and 28S was very close to the normal ratio i.e. 1:2, which means a very good quality RNA was obtained. Good concentrations of RNA were obtained from each tissue. The highest concentration (257 ng/µl) was obtained from young fruit [Fig. 3.1a] and lowest (33 ng/µl) from the roots [Fig. 3.1b] and 35 ng/µl from leaves [Fig. 3.1c]. Flower buds yielded 149 ng/µl [Fig. 3.1d] total RNA. It was hard to get good quality and quantity of RNA from roots [Fig. 3.1b] and leaves [Fig. 3.1c] with the help of Ambion Kit; however a fairly good result was obtained by using CTAB method [Fig. 3.1e]. This is because a larger quantity of tissue is used and hence higher concentration of RNA is obtained. Each of the samples had a perfect RNA integrity number (RIN) of 10 except for samples from roots and leaves. The RIN for samples from roots and leaves was somewhat lower, but exceeded 9 for all samples. As is clear from the results that the samples from flower bud and young fruit were of good quality and quantity therefore these samples were selected for further analysis.

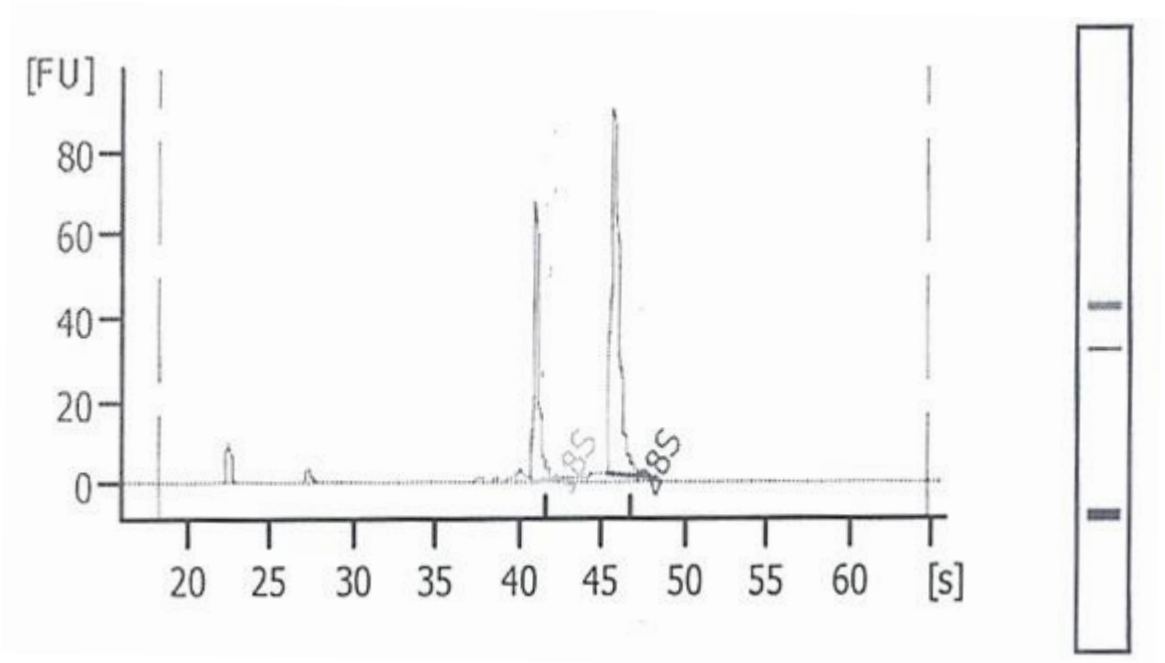


Fig. 3.1a: Gel-like image (right) and electropherogram (left) of intact total RNA obtained from young fruits of *A. fimbriata* using the Ambion method

The analyses were performed using the Agilent Bioanalyser. The results obtained are

RNA area:	213.9
RNA concentration:	257 ng/µl
rRNA ratio (28s/18s):	1.9
RNA integrity number (RIN):	10

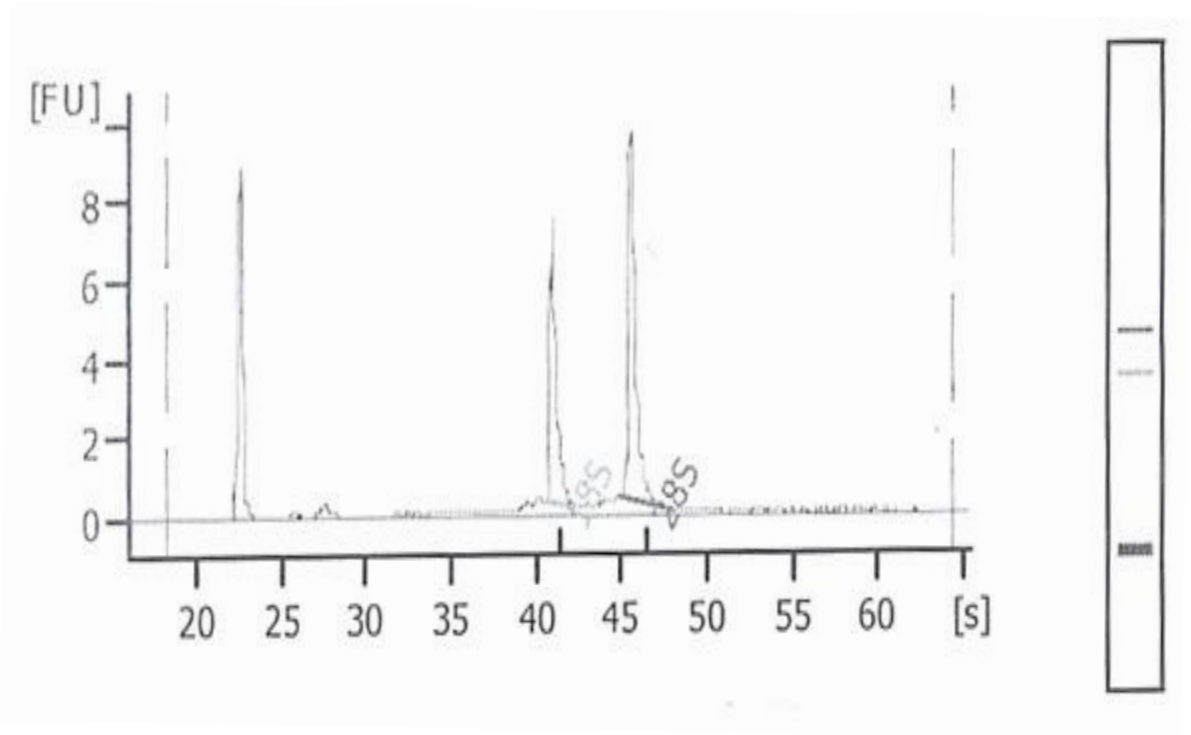


Fig. 3.1b: Gel-like image (right) and electropherogram (left) of intact total RNA obtained from young roots of *A. fimbriata* using the Ambion method

The analyses were performed using the Agilent Bioanalyser. The results obtained are

RNA area: 27.9

RNA concentration: 33 ng/µl

rRNA ratio (28s/18s): 1.3

RNA integrity number (RIN): 9.8

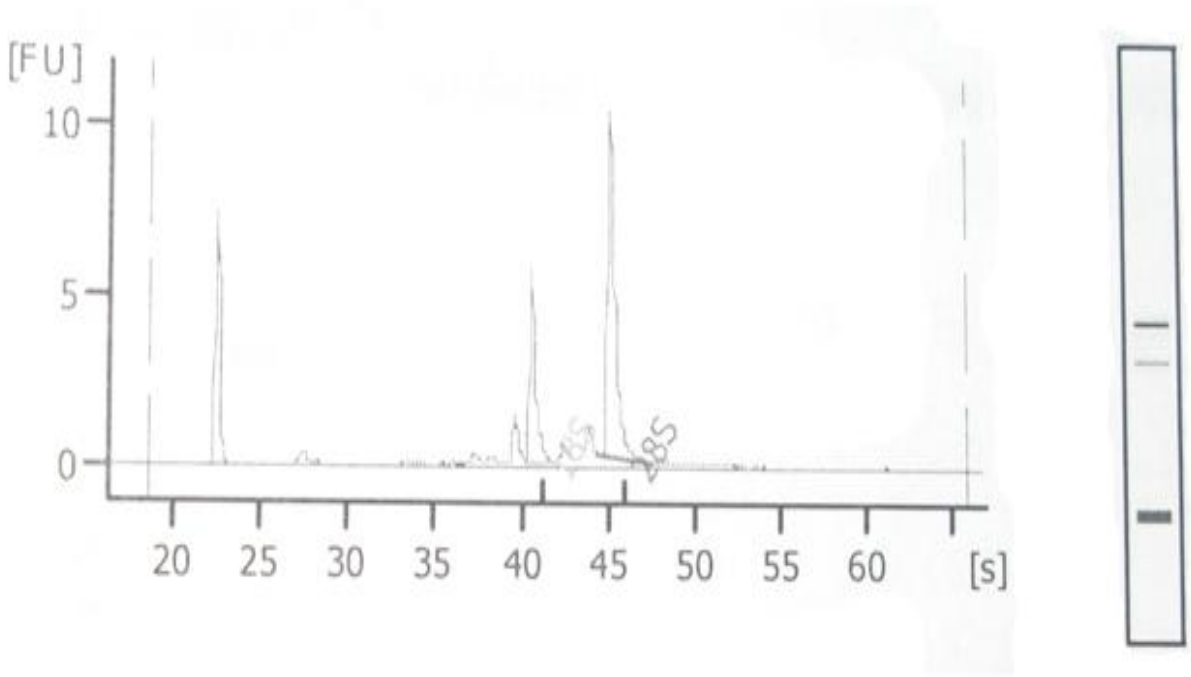


Fig. 3.1c: Gel-like image (right) and electropherogram (left) of intact total RNA obtained from young leaves of *A. fimbriata* using the Ambion method

The analyses were performed using the Agilent Bioanalyser. The results obtained are

RNA area: 30.8

RNA concentration: 35 ng/µl

rRNA ratio (28s/18s): 2.2

RNA integrity number (RIN): 9.3

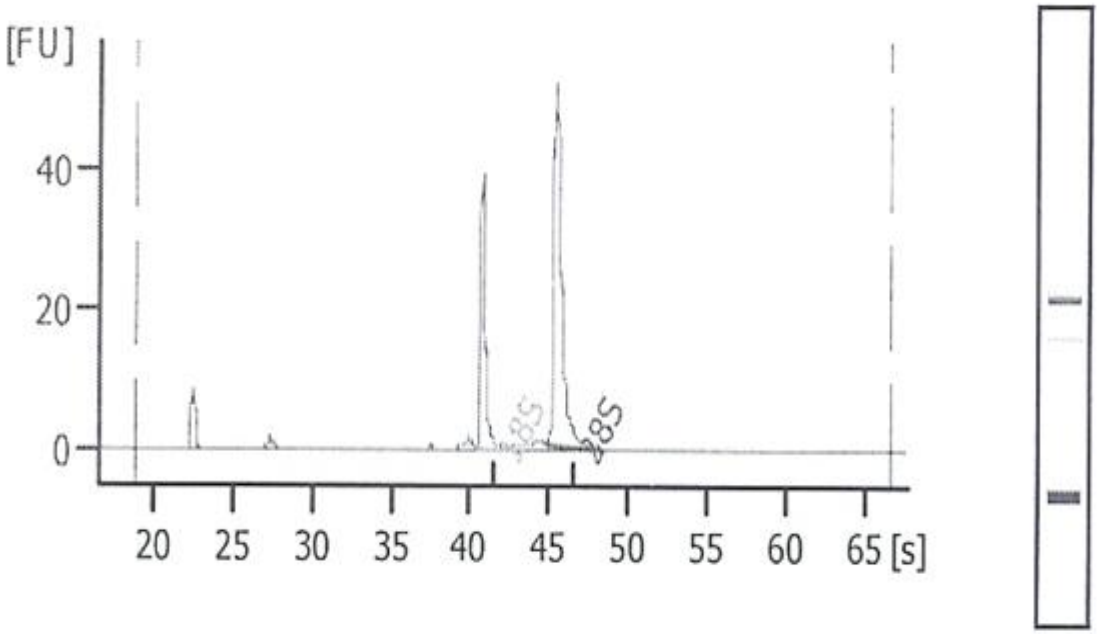


Fig. 3.1d: Gel-like image (right) and electropherogram (left) of intact total RNA obtained from flower buds of *A. fimbriata* using the Ambion method

The analyses were performed using the Agilent Bioanalyser. The results obtained are

RNA area: 124.0

RNA concentration: 149 ng/µl

rRNA ratio (28s/18s): 2.0

RNA integrity number (RIN): 10

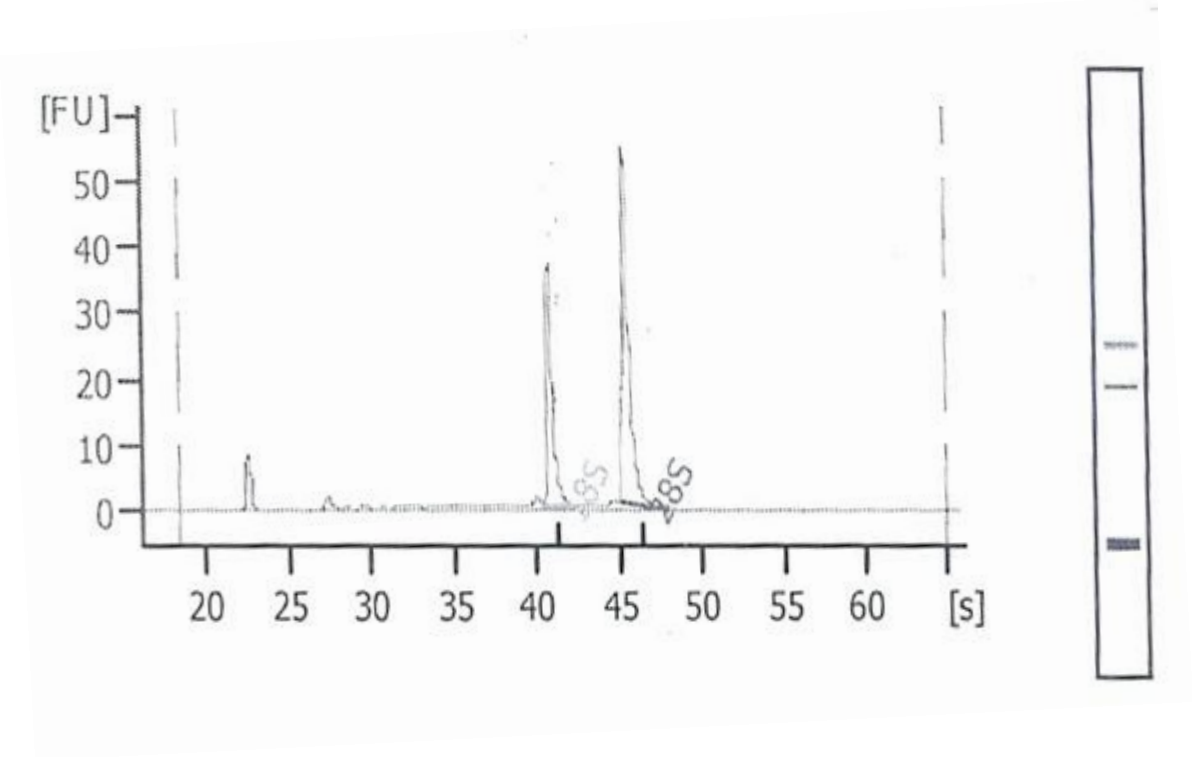


Fig. 3.1e: Gel-like image (right) and electropherogram (left) of intact total RNA obtained from young roots of *A. fimbriata* using the CTAB method

The analyses were performed using the Agilent Bioanalyser. The results obtained are

RNA area: 135.9

RNA concentration: 162 ng/µl

rRNA ratio (28s/18s): 1.7

RNA integrity number (RIN): 9.8

3.2. PCR and Gel Extraction

Oligonucleotide primers were designed for PCR amplification of the expected gene. Gradient PCR was run for four samples of two different tissues (fruit and flower buds) at two different annealing temperatures (Ta: 48.2°C and 49.8°C) for each of the tissue's RNA sample. The PCR products were loaded on agarose gel to look for the possibility of non-specific amplicons. Several bands of non-specific amplicons were obtained. Each tissue also gave bands between 1000bp and 1200bp [Fig. 3.2a]. The bands were closer to 1000bp than to 1200bp as is clear from the figure. These bands are very close to the expected size of 1076bp. The bands of the samples of both the tissues whose annealing temperature was at 49.8°C are brighter than the bands of 48.2°C, but unclear. The bands of 48.2°C are clearer for both the tissues. The bands of young fruit at both annealing temperatures are clearer than those of flower buds at respective temperature. The brighter, broader, bands in the gel show that many of the amplicons have sizes that are so close to each other that they cannot be distinguished [306]. For this reason, only the clearer bands 1 and 2 were excised from the gel for further analysis. Even these bands contained multiple amplicons; hence for further purification amplification was done for bands 1 and 2 [Fig. 3.2b] using the same protocol except the Ta which was kept constant at 48.2°C for both young fruit and flower bud samples (this time with negative control and the concentration of ladder was also increased). Multiple bands at the same place show that there might be other potential CNMT like genes in *A. fimbriata* and should be examined further. However, our study is limited to this single gene. The bands were then extracted, cloned in *E. coli*, plasmid isolated and sent to the Penn State Genomics Core Facility at University Park for sequencing. The sequence obtained was exactly identical to the expected 1076bp part of the *A. fimbriata* Unigene (CL957Contig5).

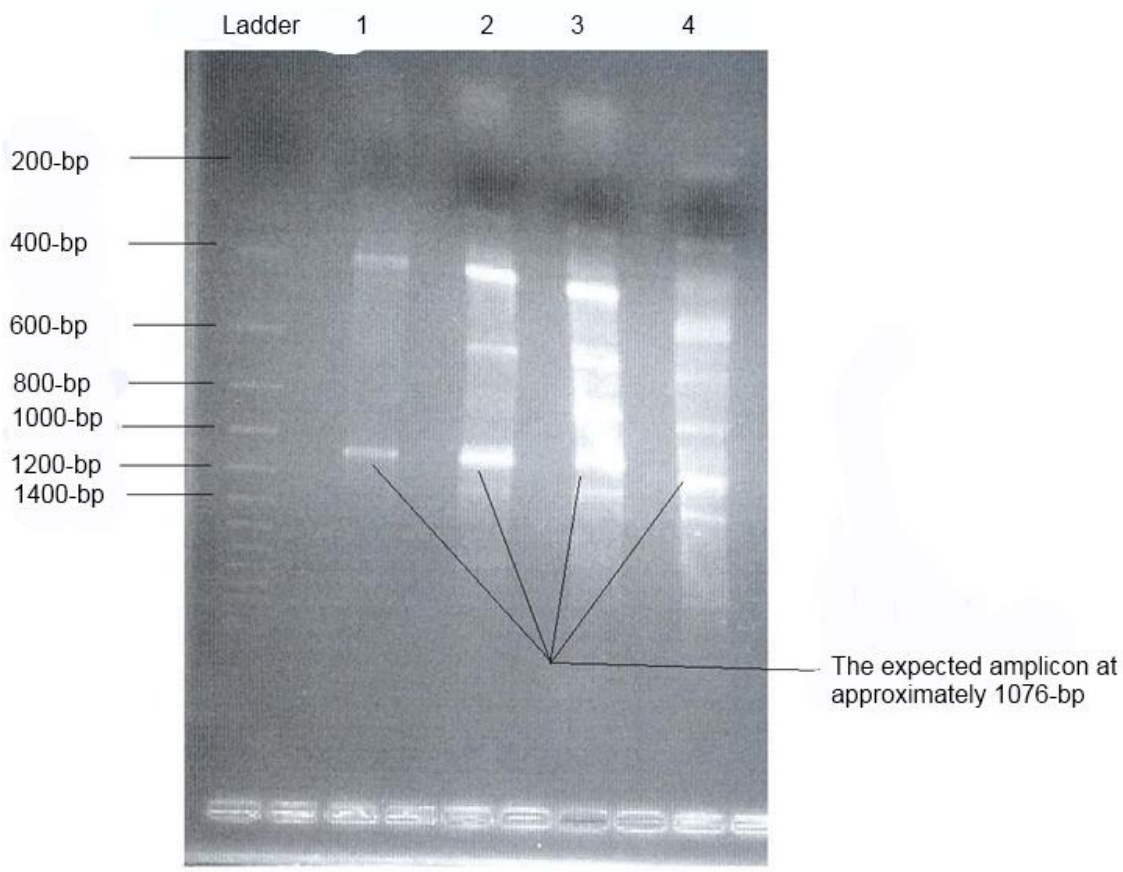


Fig. 3.2a: Agarose gel electrophoresis of PCR amplified products: Two different samples at two different annealing temperatures have been run on the gel i.e. 1: Young fruit RNA at Ta 48.2°C, 2: Flower bud RNA at 48.2°C, 3: Flower bud RNA at 49.8°C, 4: Young fruit RNA at 49.8°C

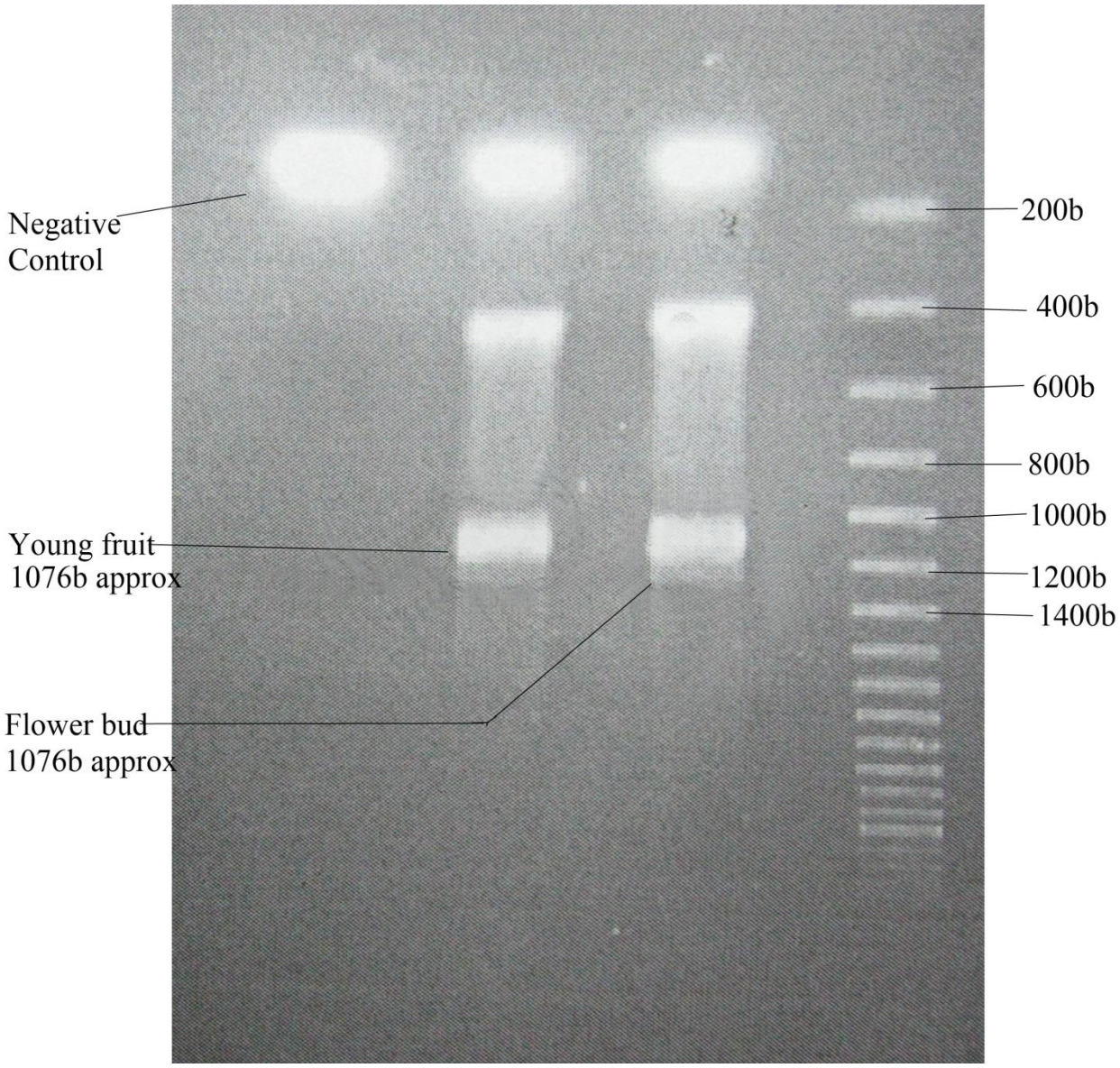


Fig. 3.2b: Agarose gel electrophoresis of PCR amplified products: Two different samples have been run on the gel i.e. 1: Young fruit RNA at Ta 48.2°C, 2: Flower bud RNA at 48.2°C.

3.3. Nucleotide Sequence Chromatogram Analysis

The gene was sequenced by the Penn State Core facility and the chromatograms were received through email. A total of eight chromatograms were received, four forward and four reverse. Only two of the chromatograms have been given in the figures [Fig. 3.3a, 3.3b]. Peaks of a good chromatogram must be sharp, evenly spaced and baseline should be flat. The peaks of forward chromatogram [Fig. 3.3a] uptill 12 are quite noisy which is normal and happens in every chromatogram. Peaks from 12 to 26 are broader, but it is normal with all chromatograms. Soon after 26 the peaks become sharper and are evenly spaced. The baseline is absolutely flat. After 730 the peaks starts broadening again and after 840 the peaks become much broader and also noisy but it is still ok and reliable. After 950 the peaks are quite broader, noisy and unreliable. This is normal with all the sequencers. At the end of a chromatogram the noise is due to the problem with the capillary electrophoresis of these days technologies. If salt is present in the sample the noise become worse. The forward chromatogram is quite reliable from 30 to 720. This can be used undoubtedly as the first part of the sequence. The last part of the sequence is inferred from reverse chromatogram [Fig. 3.3b]. The peaks in reverse chromatogram are also very nice from 30 to 730 and are reliable. The peaks are evenly spaced, sharp and the baseline is quite flat which shows its authenticity. So it can be used quite confidently, but we need its reverse complement to get the complete sequence of the gene. FinchTV 1.4.0 was used to get the reverse complement [Fig. 3.3c] of the reverse chromatogram for alignment purposes. In reverse complement chromatogram, the peaks from 480 to 1180 (Corresponding complement peaks are 744 to 34 of reverse chromatogram) are sharp, evenly spaced and clear.

Fig. 3.3a: Forward chromatogram obtained with the forward primer

Fig. 3.3b: Reverse chromatogram obtained with the reverse primer

Fig. 3.3c: Reverse complement of reverse chromatogram obtained by using the FinchTV software

The first and last noisy and unreliable parts of the sequence of the forward chromatogram were trimmed off and the sequence was taken from only the most reliable portion [Fig. 3.4a: peaks from 35 to 726]. Similarly the unreliable portions of the sequence from reverse complement of the reverse chromatogram were also trimmed and the most reliable portion [Fig. 3.4b: peaks from 494 to 1183] was taken. Both these sequences of reliable portions were aligned manually [Fig. 3.4c]. The last portion (500-726: forward chromatogram numbering) of the trimmed sequence of forward chromatogram overlapped with the first portion (494-720: reverse complement of reverse chromatogram numbering) of the reverse complement of reverse chromatogram. The alignment has been shown in figure 3.4c. Consensus sequence [Fig. 3.4c] was derived and the longest ORF was identified in the consensus sequence. The longest ORF of this consensus sequence was aligned with *A. fimbriata* Unigene CL957Contig5 [Fig. 3.4d]. The longest ORF of the consensus sequence was 100 % identical with the longest ORF of CL957Contig5. This shows that the exact same gene has been sequenced.

3.4. Tribe Identification

To search for a putative CNMT encoding gene in *A. fimbriata*, the database containing the ESTs of *A. fimbriata* was examined for sequences similar to *P. somniferum* CNMT, indirectly (because the ESTs results were not published at that time). In this indirect search, the tribe containing *A. thaliana* coding sequence (CDS), sequences similar to *P. somniferum* CNMT were found. Using the *Arabidopsis* CNMT sequence the *Aristolochia* EST database was probed with BLAST for similar sequences.

```
  1 CATGCTCGAG CGGCCGCCAG TGTGATGGAT ATCTGCAGAA TTCGCCCTTC ATGGCGTCTG
 61 AAAAGCTCAA CAAAACGGAG ATGCTGAGGA GGCTAGAAGA AGGGTCTGTT CCAGACGAGG
121 AGTTCCGGAG GCTGATAAGA ATCGAGCTCG GAAGGCGTCT CCGATGGTAC TGTCAGAAGA
181 AACCCACATA CGAGGAACAA ACGGCCGAGA TCGTCGCCTT GGTTAAGGCT CTGCGGCAGA
241 TGGGTATTAC CGGTGATCAG TCTGATCAGT TGAACTCCGA CTTGTACGAT ATGCCCATGT
301 CCTTCCTCAA GATCACCTTC GGCAAATTAT TGAAAGAAAG TGGCAGCTAC TTCAAGGATG
361 ACTCCATGAC ATTAGATGAG GCAGAGGAAG CAATGCTGGA TTTGTACTGT GAAAGAGCTC
421 AATTGAAGGA TGGACAAAAG ATTCTGGACC TGGGTTGTGG CCAAGGAGCC TTCACTCTTC
481 ATGCTGCTCA GAAATATAAG AAGAGTCATG TCACTGCAGT CACAAACTCA GCCACACAGA
541 AGAAGTACAT CGAAGATCAG TGCCAGATAC TGGAGCTTTC AAATGTGGAG GTCTTACTAG
601 AAGATATTAC CCAACTTACA ATGGAGGCCA CATTCGATAG AATAATAGTG ATTGGATTAC
661 TTGAGCATAT GAAGAACTAT GGCCTTCTTC TT
```

Fig. 3.4a: Trimmed sequence of forward chromatogram sequence (from 35 to 726: numbering on forward chromatogram)

```
  1 GAGCCTTCAC TCTTCATGCT GCTCAGAAAT ATAAGAAGAG TCATGTCACT GCAGTCACAA
 61 ACTCAGCCAC ACAGAAGAAG TACATCGAAG ATCAGTGCCA GATACTGGAG CTTTCAAATG
121 TGGAGGTCTT ACTAGAAGAT ATTACCCAAC TTACAATGGA GGCCACATTC GATAGAATAA
181 TAGTGATTGG ATTACTTGAG CATATGAAGA ACTATGGCCT TCTTCTTCAA AATATTTCTC
241 AATGGATGGC TCCAGCTGAT AGCCTTCTCT TTATAGACCA TGTTTGTCAC AAAAGCTATT
301 TTTACCAGTG TGAGCCTCTT GATGAAGATG ACTGGTTCGC TGAATATTTT TTCCCCCCTG
361 GCTCATTTGC TATGCCTTCT GCTTCTTTTC TTCTCTACTT CCAAGATGAT GTTTCCATCG
421 TGGATCATTG GATTCTTAGC GGGAAACATT TTCATCGCAC TGCGGAGGAG TGGGTGAAGC
481 AGTTGGATAC TAATTTGGAG AAGGGGAAGG AAATACTGGA GTCAAAATAT GGTAGTAAAG
541 AAGCAGCCCT GAAAGCGTTT AACCACTGGA GGGGATTATG CATATTTAGC AGTGAGATAT
601 TTGGCTACAA TGGAGGAGAA GAATGGATGA CTTCTCATCT GCTTTTCAAG AAAAAATAGG
661 AAGGGCGAAT TCCAGCACAC TGGCGGCCGT
```

Fig. 3.4b: Trimmed Sequence of reverse complement (from 494 to 1183: numbering on reverse complement of reverse chromatogram) of reverse chromatogram sequence (from 720 to 31: numbering on reverse chromatogram)

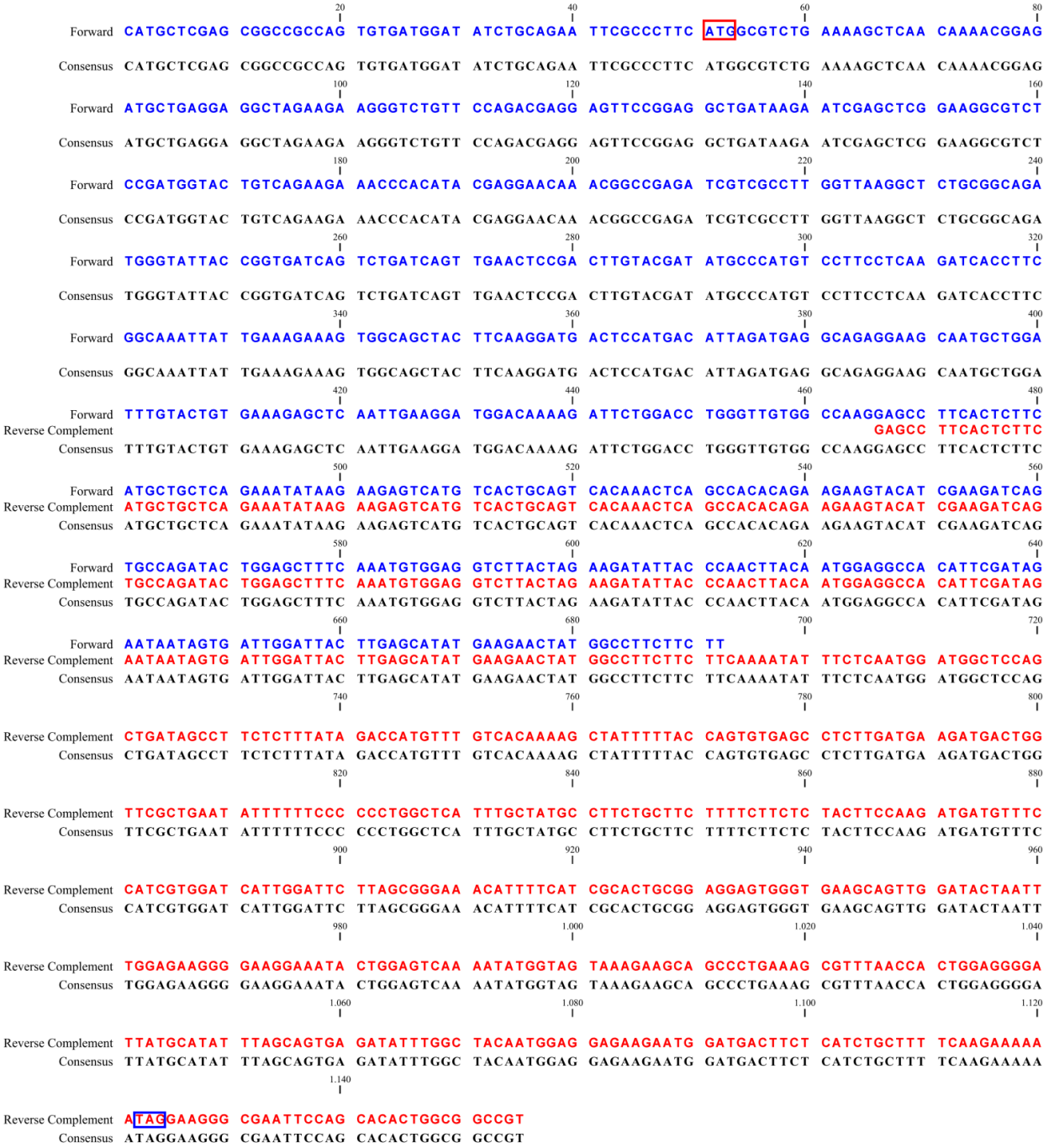


Fig. 3.4c: Alignment of trimmed forward chromatogram sequence (blue) with trimmed reverse complement (red) of reverse chromatogram sequence: Start and stop codons have been enclosed in red and blue rectangles respectively. The consensus sequence has been shown in black font. The last portion of the trimmed sequence of forward chromatogram overlapped with the first portion of the reverse complement of reverse chromatogram (from 466 to 692 in this diagram).

```
                                 ....|....| ....|....| ....|....| ....|....| ....|....|
                                        10         20         30         40         50
lcl|CL957Contig5                 AAATTTCCGT CCAGAAGCAC TCTGTCCGGC AGAGCAGAAA CAGAGTCCCC
Consensus_Sequence               ---------- ---------- ---------- ---------- ----------
Clustal Consensus


                                 ....|....| ....|....| ....|....| ....|....| ....|....|
                                        60         70         80         90        100
lcl|CL957Contig5                 AGTCTCCACT AAGTTGTAGC CGATAGATAG AGAGCGAGAG AGAGATTCTC
Consensus_Sequence               ---------- ---------- ---------- ---------- ----------
Clustal Consensus


                                 ....|....| ....|....| ....|....| ....|....| ....|....|
                                       110        120        130        140        150
lcl|CL957Contig5                 AGTCTCAGCT CCTAATTAAG CTGGTGCCTT GCCTCACGCC ATGGCGTCTG
Consensus_Sequence               ---------- ---------- ---------- ---------- ATGGCGTCTG
Clustal Consensus                                                           **********


                                 ....|....| ....|....| ....|....| ....|....| ....|....|
                                       160        170        180        190        200
lcl|CL957Contig5                 AAAAGCTCAA CAAAACGGAG ATGCTGAGGA GGCTAGAAGA AGGGTCTGTT
Consensus_Sequence               AAAAGCTCAA CAAAACGGAG ATGCTGAGGA GGCTAGAAGA AGGGTCTGTT
Clustal Consensus                ********** ********** ********** ********** **********


                                 ....|....| ....|....| ....|....| ....|....| ....|....|
                                       210        220        230        240        250
lcl|CL957Contig5                 CCAGACGAGG AGTTCCGGAG GCTGATAAGA ATCGAGCTCG GAAGGCGTCT
Consensus_Sequence               CCAGACGAGG AGTTCCGGAG GCTGATAAGA ATCGAGCTCG GAAGGCGTCT
Clustal Consensus                ********** ********** ********** ********** **********


                                 ....|....| ....|....| ....|....| ....|....| ....|....|
                                       260        270        280        290        300
lcl|CL957Contig5                 CCGATGGTAC TGTCAGAAGA AACCCACATA CGAGGAACAA ACGGCCGAGA
Consensus_Sequence               CCGATGGTAC TGTCAGAAGA AACCCACATA CGAGGAACAA ACGGCCGAGA
Clustal Consensus                ********** ********** ********** ********** **********


                                 ....|....| ....|....| ....|....| ....|....| ....|....|
                                       310        320        330        340        350
lcl|CL957Contig5                 TCGTCGCCTT GGTTAAGGCT CTGCGGCAGA TGGGTATTAC CGGTGATCAG
Consensus_Sequence               TCGTCGCCTT GGTTAAGGCT CTGCGGCAGA TGGGTATTAC CGGTGATCAG
Clustal Consensus                ********** ********** ********** ********** **********


                                 ....|....| ....|....| ....|....| ....|....| ....|....|
                                       360        370        380        390        400
lcl|CL957Contig5                 TCTGATCAGT TGAACTCCGA CTTGTACGAT ATGCCCATGT CCTTCCTCAA
Consensus_Sequence               TCTGATCAGT TGAACTCCGA CTTGTACGAT ATGCCCATGT CCTTCCTCAA
Clustal Consensus                ********** ********** ********** ********** **********


                                 ....|....| ....|....| ....|....| ....|....| ....|....|
                                       410        420        430        440        450
lcl|CL957Contig5                 GATCACCTTC GGCAAATTAT TGAAAGAAAG TGGCAGCTAC TTCAAGGATG
Consensus_Sequence               GATCACCTTC GGCAAATTAT TGAAAGAAAG TGGCAGCTAC TTCAAGGATG
Clustal Consensus                ********** ********** ********** ********** **********
```

```
                          ....|....| ....|....| ....|....| ....|....| ....|....|
                                  460        470        480        490        500
lcl|CL957Contig5          ACTCCATGAC ATTAGATGAG GCAGAGGAAG CAATGCTGGA TTTGTACTGT
Consensus_Sequence        ACTCCATGAC ATTAGATGAG GCAGAGGAAG CAATGCTGGA TTTGTACTGT
Clustal Consensus         ********** ********** ********** ********** **********

                          ....|....| ....|....| ....|....| ....|....| ....|....|
                                  510        520        530        540        550
lcl|CL957Contig5          GAAAGAGCTC AATTGAAGGA TGGACAAAAG ATTCTGGACC TGGGTTGTGG
Consensus_Sequence        GAAAGAGCTC AATTGAAGGA TGGACAAAAG ATTCTGGACC TGGGTTGTGG
Clustal Consensus         ********** ********** ********** ********** **********

                          ....|....| ....|....| ....|....| ....|....| ....|....|
                                  560        570        580        590        600
lcl|CL957Contig5          CCAAGGAGCC TTCACTCTTC ATGCTGCTCA GAAATATAAG AAGAGTCATG
Consensus_Sequence        CCAAGGAGCC TTCACTCTTC ATGCTGCTCA GAAATATAAG AAGAGTCATG
Clustal Consensus         ********** ********** ********** ********** **********

                          ....|....| ....|....| ....|....| ....|....| ....|....|
                                  610        620        630        640        650
lcl|CL957Contig5          TCACTGCAGT CACAAACTCA GCCACACAGA AGAAGTACAT CGAAGATCAG
Consensus_Sequence        TCACTGCAGT CACAAACTCA GCCACACAGA AGAAGTACAT CGAAGATCAG
Clustal Consensus         ********** ********** ********** ********** **********

                          ....|....| ....|....| ....|....| ....|....| ....|....|
                                  660        670        680        690        700
lcl|CL957Contig5          TGCCAGATAC TGGAGCTTTC AAATGTGGAG GTCTTACTAG AAGATATTAC
Consensus_Sequence        TGCCAGATAC TGGAGCTTTC AAATGTGGAG GTCTTACTAG AAGATATTAC
Clustal Consensus         ********** ********** ********** ********** **********

                          ....|....| ....|....| ....|....| ....|....| ....|....|
                                  710        720        730        740        750
lcl|CL957Contig5          CCAACTTACA ATGGAGGCCA CATTCGATAG AATAATAGTG ATTGGATTAC
Consensus_Sequence        CCAACTTACA ATGGAGGCCA CATTCGATAG AATAATAGTG ATTGGATTAC
Clustal Consensus         ********** ********** ********** ********** **********

                          ....|....| ....|....| ....|....| ....|....| ....|....|
                                  760        770        780        790        800
lcl|CL957Contig5          TTGAGCATAT GAAGAACTAT GGCCTTCTTC TTCAAAATAT TTCTCAATGG
Consensus_Sequence        TTGAGCATAT GAAGAACTAT GGCCTTCTTC TTCAAAATAT TTCTCAATGG
Clustal Consensus         ********** ********** ********** ********** **********

                          ....|....| ....|....| ....|....| ....|....| ....|....|
                                  810        820        830        840        850
lcl|CL957Contig5          ATGGCTCCAG CTGATAGCCT TCTCTTTATA GACCATGTTT GTCACAAAAG
Consensus_Sequence        ATGGCTCCAG CTGATAGCCT TCTCTTTATA GACCATGTTT GTCACAAAAG
Clustal Consensus         ********** ********** ********** ********** **********

                          ....|....| ....|....| ....|....| ....|....| ....|....|
                                  860        870        880        890        900
lcl|CL957Contig5          CTATTTTTAC CAGTGTGAGC CTCTTGATGA AGATGACTGG TTCGCTGAAT
Consensus_Sequence        CTATTTTTAC CAGTGTGAGC CTCTTGATGA AGATGACTGG TTCGCTGAAT
Clustal Consensus         ********** ********** ********** ********** **********

                          ....|....| ....|....| ....|....| ....|....| ....|....|
```

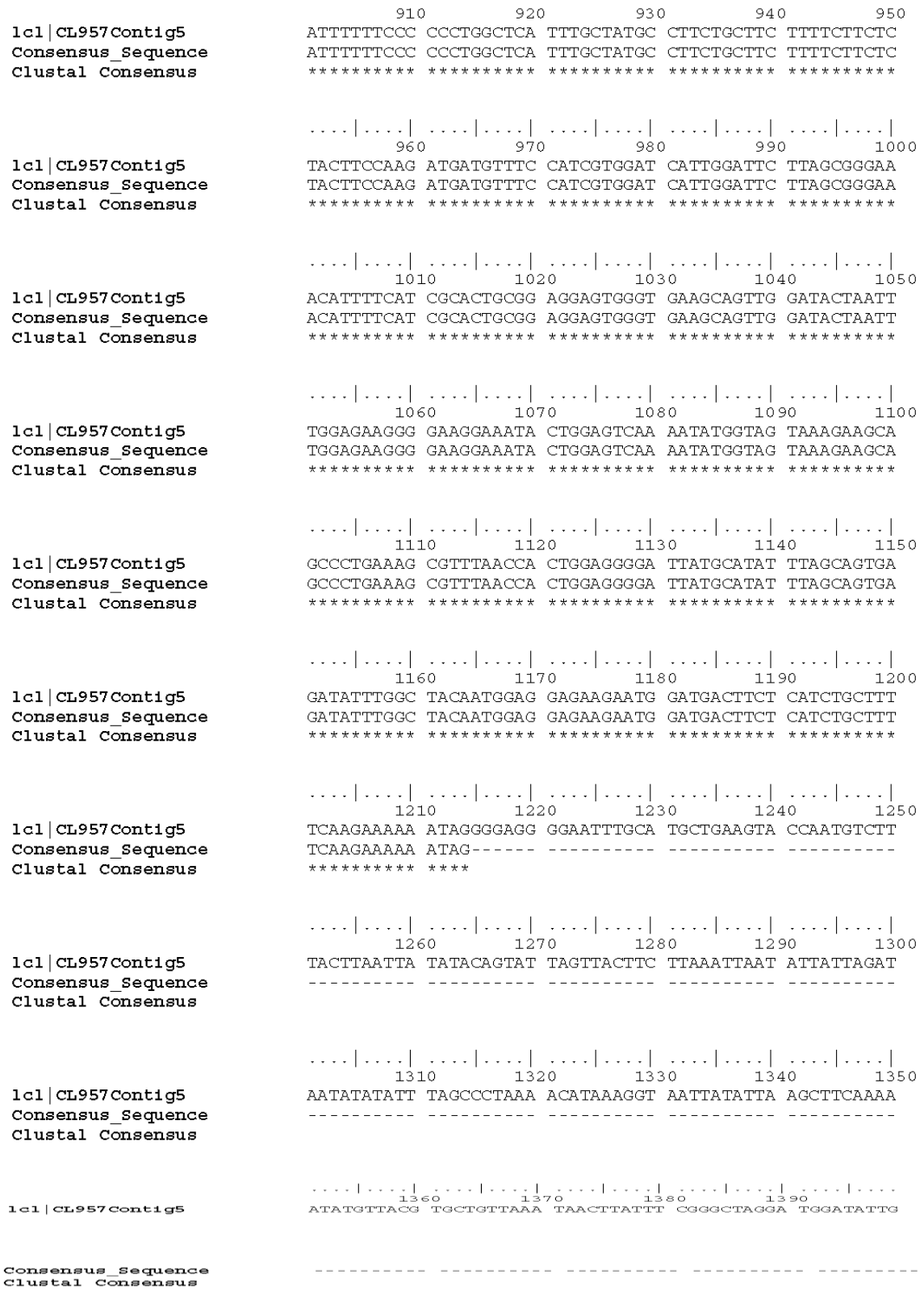


Fig. 3.4d: Pairwise sequence alignment of consensus sequence with *A. fimbriata* Unigene CL957Contig5

The 1054bp long CDS sequence [Fig. 3.5] of *P. somniferum* CNMT (GenBank accession number: AY217336) was downloaded from the GenBank and was submitted to the PlantTribe-BLASTx for sequence comparison of the CDS (coding sequence) with the CDS of the PlantTribe database (*Arabidopsis* + *Oryza* + *Populus* protein). All the sequences obtained belong to the same cluster (tribe) at stringency 1.2, 3.0 and 5.0 with tribe IDs 835, 4714 and 4676 respectively (Table. 3.1). The result contained a total of seven hits. There were two sequences from *A. thaliana* (putative CNMTs), one from *Carica papaya* (with no description), one from *Medicago truncatula* (SAM dependent methyltransferase), one from *Oryza sativa* (cyclopropane-fatty-acyl-phospholipid-synthase) and two from *Populus trichocarpa* (no description) in the tribe at stringency 5 (Tribe ID: 4676). All these seven proteins are SAM-MTases, however only two of the sequences from *A. thaliana* have been annotated as putative CNMTs.

The sequence data available at EST databases is quite helpful in gene discovery and functional prediction [307, 308]. As this study is about looking for genes involved in the encoding of putative CNMT in *A. fimbriata*, CDS sequence of *A. thaliana* (ID: AT4G33110.1) was selected as a query, because this is the only CNMT (as annotated) in this cluster that is most similar with *P. somniferum* CNMT and it has also been shown previously that the predicted translation product of the *A. thaliana* gene At4g33120 and At4g33110 are close homologs of CNMT [15]. *A. thaliana* CNMT was BLASTed against *A. fimbriata* EST database. It was found to be a potential ortholog of *Aristolochia* Contig's tribe 3389 (Tribe ID). Tribe 3389 [Table. 3.2] contains twelve *A. fimbriata* unigene sequences. Although CL2801contig1 was the longest contig, it contained shorter ORFs as compared to some other contigs.

```
   1 ATGCAGCTAA AGGCAAAGGA AGAGCTGTTG AGAAACATGG AGCTTGGGTT GATACCGGAC
  61 CAAGAGATTA GACAACTGAT TAGAGTCGAA TTAGAAAAGC GTCTCCAATG GGGTTACAAA
 121 GAAACCCATG AAGAACAGCT TTCTCAGCTT CTTGACTTGG TTCACTCTCT GAAAGGTATG
 181 AAAATGGCAA CTGAGATGGA GAATTTGGAT TTAAAACTAA CGAAGCACCT ATGGAATTCT
 241 TAAAGATCCA ACATGGAAGC AACATGAAAC AAAGTGCGGG TTACTACACT GATGAATCAA
 301 CAACGTTGGA CGAAGCTGAG ATAGCGATGC TGGATTTGTA TATGGAGAGG GCACAGATCA
 361 AAGATGGTCA GAGTGTACTT GATTTAGGGT GTGGATTGGG TGCTGTAGCT CTTTTCGGGG
 421 CAAACAAATT TAAGAAATGT CAATTTACTG GAGTTACAAG TTCTGTGGAA CAAAAAGATT
 481 ACATTGAAGG AAAATGCAAG GAGCTTAAGT TGACCAACGT CAAGGTTTTA TTAGCTGATA
 541 TAACAACTTA CGAAACCGAA GAAAGATTTG ATCGGATTTT CGCTGTTGAA TTGATTGAGC
 601 ATATGAAGAA CTATCAATTG CTTCTTAAGA AAATATCAGA GTGGATGAAA GATGATGGAC
 661 TTCTTTTCGT TGAACATGTT TGCCATAAGA CTCTAGCTTA CCATTATGAG CCCGTTGATG
 721 CAGAAGATTG GTACACTAAT TACATATTCC CTGCTGGAAC TCTGACTCTT CATCTGCTTC
 781 AATGCTTCTT TACTTTCAAG ATGATGTTTC AGTTGTGAAT CAATGGACTC TGAGTGGAAA
 841 GCATTACTCT AGATCCCATG AAGAATGGTT GAAAAACATG GATAAAAACA TTGTTGAATT
 901 CAAAGAAATA ATGAGGTCCA TAACTAAAAC TGAGAAAGAA GCAATCAAAT TGCTCAATTT
 961 TTGGAGGATT TTTTGCATGT GTGGAGCAGA GTTGTTTGGG TATAAGAATG GTGAAGAATG
1021 GATGCTCACC CATCTTCTCT TCAAGAAAAA ATGA
```

Fig. 3.5: The 1054 bases long (from 76-1131 base) complete CDS of *P. somniferum* S-adenosyl-L-methionine: CNMT mRNA, GenBank accession number: AY217336

Table 3.1: BLASTx results of *P. somniferum* CDS against *Arabidopsis+Oryza+Populus*-Protein sequences at PlantTribe database

ID	Species	Description	Score	E.Value	Stringency		
					1.2	**3.0**	**5.0**
					Cluster (Tribe) ID		
Os06g37610.1	Orysa5	CFAPS	219	9e-57	835	4714	4676
AT4G33110.1	Arath7	CNMT Putative	209	9e-54			
CT967316_8.3	Medtr1	SAM-MTase	208	2e-53			
estExt_fgenesh4_pg.C_LG_XVIII0095	Poptr1	No description	207	3e-53			
evm.model.supercontig_91.83	Carpa1	No description	207	6e-53			
AT4G33120.1	Arath7	CNMT Putative	203	6e-52			
gw1.VI.341.1	Poptr1	No description	143	8e-34			

Abbreviations:

CFAPS: Cyclopropane-fatty-acyl-phospholipid-synthase

CNMT: Coclaurine N-methyltransferase

SAM: S-adenocyl-L-methionine

Orysa: *Oryza sativa*

Arath: *Arabidopsis thaliana*

Medtr: *Medicago truncatula*

Poptr: *Populus trichocarpa*

Carpa: *Carica papaya*

Table 3.2: List of unigenes included in tribe 3389

Unigene	Length	Number of ESTs	Longest ORF	Super Tribe	Tribe
CL957Contig5	1399	78	1074b: 357aa	438	3389
CL1Contig6640	1272	61	150b: 49aa		
CL1Contig6030	1220	58	987b: 328aa		
CL2801Contig1	1410	40	696b: 231aa		
CL957Contig3	612	14	598b: 199aa		
CL957Contig4	256	6	237b: 78aa		
CL16496Contig1	289	6			
CL20924Contig1	392	5	354b: 117aa		
CL1Contig10736	343	4			
CL53903Contig1	246	2	165b: 54aa		
CL2801Contig2	208	2	150b: 49		
371184_0645_1136	235	1	No ORF		

Therefore this contig was not selected because shorter ORFs cannot be taken as genes. CL957Contig5 was the second longest contig with a length of 1399 bases and with the largest number of ESTs among the contigs. This contig contained the longest ORF with 1074 bases starting from start codon (ATG) at base number 141 and ending at stop codon (TAG) at base number 1214 [Fig. 3.6], which means that it might contain full CDS sequence for a putative CNMT. The second longest contig (987 bases: 328 amino acids) was found in CL1contig6030, but it did not show similarities to CNMTs by BLAST search. The remaining sequences either contained shorter ORFs or the contigs were too small to be considered as complete coding sequence for a CNMT. However there might be other potential CNMT like genes in the *A. fimbriata* EST database that should be examined extensively.

3.5. Identification of the Gene and Prediction of the Putative Function of its Protein

Several methods and tools were used for the identification of the gene and to predict the function of its translated putative protein. The results are discussed below.

3.5.1. Pairwise Sequence Alignment

In bioinformatics, a sequence alignment is a way of arranging the sequences of DNA, RNA, or protein to identify regions of similarity that may be a consequence of functional, structural, or evolutionary relationships between the sequences [309]. Pairwise sequence alignment can give us a clue about the function of the sequence under consideration.

Fig. 3.6: Unigene CL957Contig5 sequence of *A. fimbriata*: Primers were designed for the segments enclosed in blue color boxes. The longest ORF has been colored red.

The pairwise alignment of the longest ORF of *A. fimbriata* unigene CL957Contig5 shows that it is 61% identical to the *P. somniferum* CDS [Fig. 3.7a]; and the translated ORF was 52% identical to the sequence of *P. somniferum* CNMT [Fig. 3.7b]. This is a clue that CL957Contig5 might contain a full length CDS encoding a putative CNMT. As is clear from the alignment, there are continuous identities at several regions throughout the sequences; this is likely to indicate either close relatedness or identity due to functional constraints on the sequences.

3.5.2. BLASTn, BLASTx and tBLASTx Similarity Searches

BLAST finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families [310]. Several types of BLAST were used in this study to identify the gene and predict its function.

The BLASTn results [Table.3.3] show that unigene CL957Contig5 is 69 % identical (with 72 % coverage) to a putative CNMT-like mRNA [311]. The second, third, sixth and seventeenth most similar CDS sequences are CNMT mRNAs from *T. flavum* [17], *C. japonica* (experimentally expressed) [15], *P. somniferum* [98] and *A. thaliana* with 68 %, 67 %, 66 % and 63 % identities respectively. The eighth [311] and fourteenth [311] most similar matches are also CNMT-like mRNAs. Most of the similar matches are annotated as mRNAs of SAM-MTases. Only the top twenty most similar matches have been shown in the table [66, 312-314].

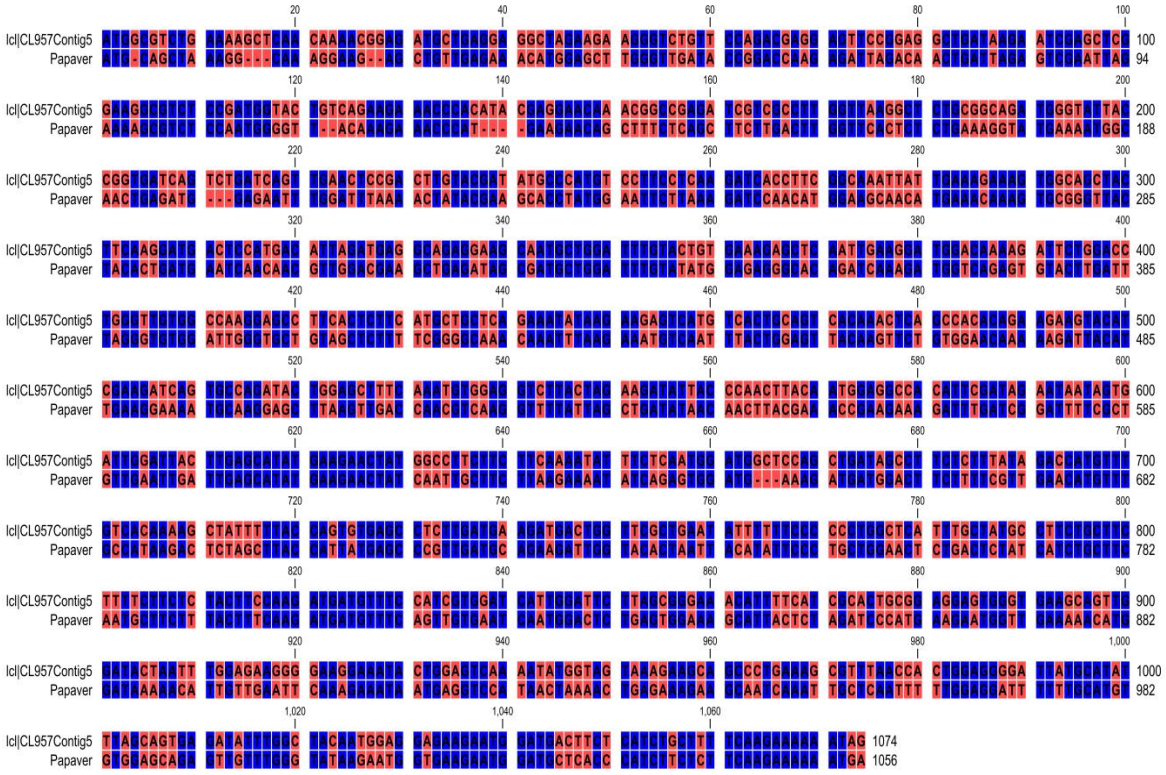


Fig. 3.7a: Pairwise sequence alignment of the longest ORF of *A. fimbriata* (the longest ORF of the unigene CL957Contig5) with the complete CDS of *P. somniferum* CNMT: There is 61 % (653/1071; excluding the stop codon) identity (blue) between the two sequences. The conserved residues have been shown in blue color.

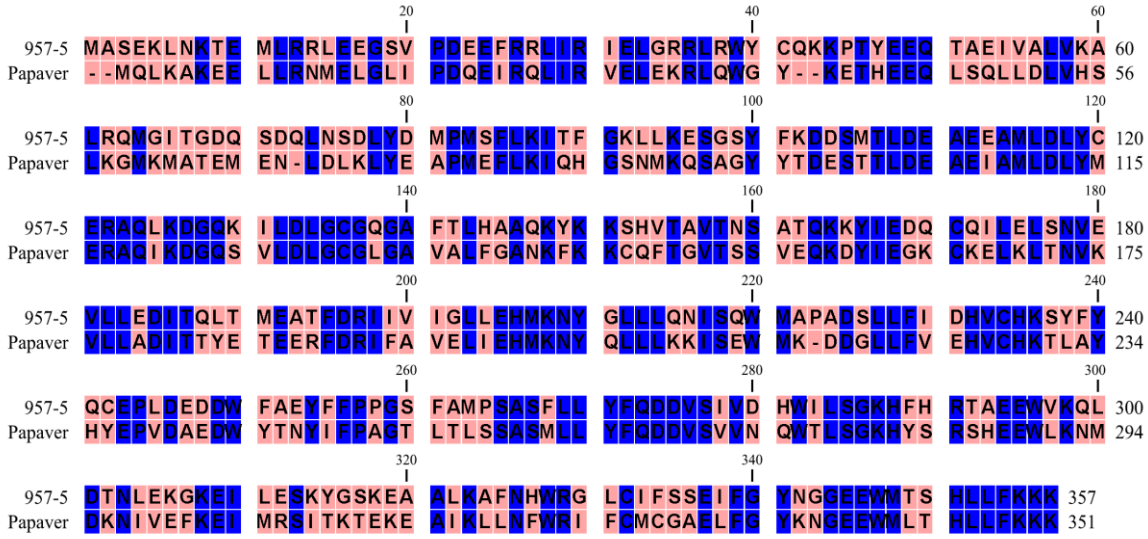


Fig. 3.7b: Pairwise sequence alignment of the longest translated ORF of *A. fimbriata* unigene CL957Contig5 with *P. somniferum* CNMT: There is 52% (186/357) identity (blue) between the two sequences. The conserved residues have been shown in blue color.

Table 3.3: Top 20 best hits for cDNA sequence (longest ORF) of putative CNMT from *A. fimbriata* obtained through BLASTn search of NCBI

S.No.	Accession	Organism	Description	Score	Coverage	E.V.	Identity
1	EU883009.1	*T. flavum*	CNMT like NMT mRNA (putative), complete cds	286	72%	2e-73	69%
2	AY610508.1	*T. flavum*	CNMT mRNA, complete cds	279	73%	3e-71	68%
3	AB061863.1	*C. japonica*	CNMT mRNA, complete cds	262	73%	2e-66	67%
4	EU883010.1	*T. flavum*	PNMT (Pavine-NMT) mRNA, complete cds	190	72%	1e-44	65%
5	EU882993.1	*Papaver bracteatum*	TNMT like NMT mRNA (putative); clone PBRSC1NG_027_E03, complete cds	159	60%	2e-35	66%
6	AY217336.1	*P. somniferum*	CNMT mRNA, complete cds	158	54%	7e-35	66%
7	XM_002324706	*P. trichocarpa*	predicted protein, mRNA	147	51%	1e-31	66%
8	EU882973.1	*E. californica*	CNMT like mRNA;clone ECASC1NG_058_A10, partial sequence	141	25%	6e-30	71%
9	AK323570.1	*S. lycopersicum*	cDNA, clone: LEFL1059DA09, HTC in leaf	136	53%	2e-28	66%
10	AK321345.1	*S. lycopersicum*	cDNA, clone: LEFL1023CA04, HTC in leaf	136	53%	2e-28	66%
11	EU882994.1	*P. bracteatum*	TNMT (tetrahydroprotoberberine NMT) mRNA; clone PBRSC1NG_051_D09, complete cds	129	52%	3e-26	65%
12	DQ028579.1	*P. somniferum*	TNMT mRNA, complete cds	118	51%	6e-23	65%
13	EU882977.1	*E. californica*	TNMT mRNA, complete cds	116	52%	2e-22	65%
14	EU883002.1	*T. flavum*	CNMT-like protein mRNA; clone TFLSC1NG_041_D11, partial cds	96.9	30%	2e-16	67%
15	XM_002280256	*Vitis vinifera*	hypothetical protein LOC100259368 (LOC100259368), mRNA	93.3	39%	3e-15	65%
16	AK227357.1	*A. thaliana*	hypothetical protein mRNA, complete cds, clone: RAFL14-03-H17	80.6	50%	2e-11	63%
17	NM_119466.2	*A. thaliana*	CNMT (putative), (AT4G33120) mRNA, complete cds	80.6	50%	2e-11	63%
18	AY136408.1	*A. thaliana*	putative protein (At4g33110) mRNA, complete cds	80.6	50%	2e-11	63%
19	BT005878.1	*A. thaliana*	At4g33120 gene, complete cds	80.6	50%	2e-11	63%
20	AY088221.1	*A. thaliana*	clone 4369 mRNA, complete sequence, Conceptual translation	80.6	50%	2e-11	63%

BLASTx results have been given in table.3.4. The top 20 most similar matches have been shown in the table. The greatest identity match was an NMT from *T. flavum* [311] with an E-Value of 1e-111. Second, third, fifth, sixteenth and nineteenth most similar sequences were also annotaed as CNMTs from *C. japonica* (E.V: 3e-110, experimentally expressed) [15], *T. flavum* (E.V: 5e-109) [17], *P. somniferum* (E.V: 1e-95) [98] and *A. thaliana* (2 sequences with E.V: 1e-82 and 4e-82) [314] respectively. Most of the matches of *A. fimbriata* sequence were to SAM-MTases. Below the top 20 most similar matches (results not shown in the table) were mostly CFAPS.

From tBLASTx (Table. 3.5) only the top five closest matches have been given in the table. The most identical match was CNMT mRNA sequence from *C. japonica* (experimentally expressed) with an E.V. of 9e-123. Third and fifth best matches were also annotated as CNMT mRNAs from *T. flavum* (2e-115) and *P. somniferum* (E.V: 6e-103), however second and fourth best matches were annotated as putative CNMT-like N-methyltransferase [311] and pavine N-methyltransferase [311] respectively both from *T. flavum*.

From all the above identity matches, it can be hypothesized that the cDNA isolated from *A. fimbriata* might be involved in the production of a methyltransferase, and possibly a putative CNMT. However for further confirmation, the sequence was translated into protein sequence and used to BLAST for similar protein sequences, because protein search is superior to nucleotide-nucleotide search and gives less false positive results [257]. These results also predict that *Aristolochia* sequence could be most closely related to that of *Thalictrum* and *Coptis*.

Table 3.4: Top 20 best hits for cDNA sequence (longest ORF) of putative CNMT from *A. fimbriata* obtained through BLASTx search of NCBI

S. No	Accession	Organism	Description	Score	E.V.
1	ACO90253.1	*T. flavum*	NMT	407	1e-111
2	BAB71802.1	*C. japonica*	CNMT	402	3e-110
3	AAU20766.1	*T. flavum*	CNMT	398	5e-109
4	ACO90251.1	*T. flavum*	PNMT	373	2e-101
5	AAP45316.1	*P. somniferum*	CNMT	353	1e-95
6	ACO90222.1	*E. californica*	TNMT	338	8e-91
7	XP_002324742	*P. trichocarpa*	SAM-MTase: predicted	322	6e-86
8	ACO90236.1	*P. bracteatum*	TNMT like NMT (putative)	320	2e-85
9	EEE65917.1	*O. sativa*	SAM-MTase (hypothetical protein)	317	1e-84
10	XP_002280292	*V. vinifera*	SAM-MTase (hypothetical protein)	317	1e-84
11	BAD61850.1	*O. sativa*	CNMT (putative)	317	1e-84
12	EEC80845.1	*O. sativa*	SAM-MTase (hypothetical protein)	317	2e-84
13	ACO90237.1	*P. bracteatum*	TNMT	313	2e-83
14	XP_002437192	*Sorghum bicolor*	CNMT like hypothetical protein	311	6e-83
15	ACF85951.1	*Z. mays*	SAM-MTase like (unknown function)	311	6e-83
16	AAM65762.1	*A. thaliana*	CNMT	310	1e-82
17	AAY79177.1	*P. somniferum*	TNMT	309	3e-82
18	BAC42939.1	*A. thaliana*	SAM-MTase like (unknown function)	309	3e-82
19	NP_567912.1	*A. thaliana*	CNMT (putative)	309	4e-82
20	ACJ84709.1	*M. truncatula*	SAM-MTase like Unknown function	305	4e-81

Table 3.5: Top 5 best hits for cDNA sequence (ORF) of *A. fimbriata* putative CNMT obtained through tBLASTx search

S. No	Accession	Organism	Description	Score	E.V.
1	AB061863.1	*C. japonica*	CNMT mRNA	228	9e-123
2	EU883009.1	*T. flavum*	CNMT like NMT (putative)	240	5e-121
3	AY610508.1	*T. flavum*	CNMT mRNA	222	2e-115
4	EU883010.1	*T. flavum*	PNMT	229	1e-110
5	AY217336.1	*P. somniferum*	CNMT mRNA	206	6e-103

Although the homologous sequences found in these searches were genes from a wide diversity of eudicot (*Thalictrum*, *Coptis*, *Papaver*, *Eschscholzia*, *Populus*, *Solanum*, *Vitis*, *Arabidopsis*, *Medicago*) and monocot (*Oryza*, *Zea*) species, none of the top hits in the search was identified as coming from a basal angiosperm. This suggests that this could be the first CNMT homolog that has been cloned and characterized from a basal angiosperm lineage.

3.5.3. Sequence Translation and BLASTp Search

For most common purposes, gene is usually defined as the longest ORF for a given region of DNA [309]. Translation of a DNA sequence in all six reading frames is a straightforward task, which can be performed on line using, for example, the Translate tool on the ExPASy server (http://www.expasy.org/tools/dna.html).

Translate tool translated the nucleotide sequence into six possible frames [Fig. 3.8a]. All the six frames were analyzed for the ORF lengths. There are 14 possible ORFs in 5′ 3′ Frame-1, the longest one being 65 amino acid (aa) residues long. There are 5 possible ORFs in 5′ 3′ Frame-2 with the longest ORF being 31 aa residues long. Fourteen ORFs were found in 5′3′Frame-3. The longest ORF of this frame was 357 aa residues in length. There are 8 ORFs in 3′5′Frame-1 with the longest one being 43 aa residues long. There are a total of 4 ORFs in 3′5′Frame-2 among which the longest one was 40 aa residues long and the last Frame i.e. 3′5′Frame-3 contains 8 ORFs with the longest one being 45 aa residues long. The longest ORF found among all the six frames was 357 residues in length.

Open Reading Frame (ORF) finder was also used to identify the longest ORF of CL957Contig5.

5'3' Frame 1

K F P S R S T L S G R A E T E S P V S T K L Stop P I D R E R E R D S Q S Q L L I K L V P C L T P W R L K S S T K R R C Stop G G Stop K K G L F Q T R S S G
G Stop Stop E S S S E G V S D G T V R R N P H T R N K R P R S S P W L R L C G R W V L P V I S L I S Stop T P T C T I C P C P S S R S P S A N Y Stop K K V
A A T S R M T P Stop H Stop M R Q R K Q C W I C T V K E L N Stop R M D K R F W T W V V A K E P S L F M L L R N I R R V M S L Q S Q T Q P H R R S T
S K I S A R Y W S F Q M W R S Y Stop K I L P N L Q W R P H S I E Stop Stop Stop L D Y L S I Stop R T M A F F F K I F L N G W L Q L I A F S L Stop T M F
V T K A I F T S V S L L M K M T G S L N I F S P L A H L L C L L L L F F S T S K M M F P S W I I G F L A G N I F I A L R R S G Stop S S W I L I W R R G R K Y
W S Q N M V V K K Q P Stop K R L T T G G D Y A Y L A V R Y L A T M E E K N G Stop L L I C F S R K N R G G E F A C Stop S T N V F T Stop L Y T V L V T
S Stop I N I I R Stop Y I F S P K T Stop R Stop L Y Stop A S K I C Y V L L N N L F R A R M D I

5'3' Frame 2

N F R P E A L C P A E Q K Q S P Q S P L S C S R Stop I E S E R E I L S L S S Stop L S W C L A S R H G V Stop K A Q Q N G D A E E A R R R V C S R R G V P
E A D K N R A R K A S P M V L S E E T H I R G T N G R D R R L G Stop G S A A D G Y Y R Stop S V Stop S V E L R L V R Y A H V L P Q D H L R Q I I E R K
W Q L L Q G Stop L H D I R Stop G R G S N A G F V L Stop K S S I E G W T K D S G P G L W P R S L H S S C C S E I Stop E E S C H C S H K L S H T E E V H
R R S V P D T G A F K C G G L T R R Y Y P T Y N G G H I R Stop N N S D W I T Stop A Y E E L W P S S S K Y F S M D G S S Stop Stop P S L Y R P C L S Q K
L F L P V Stop A S Stop Stop R Stop L V R Stop I F F P P W L I C Y A F C F F S S L L P R Stop C F H R G S L D S Stop R E T F S S H C G G V G E A V G Y
Stop F G E G E G N T G V K I W Stop Stop R S S P E S V Stop P L E G I M H I Stop Q Stop D I W L Q W R R R M D D F S S A F Q E K I G E G N L H A E V
P M S L L N Y I Q Y Stop L L L K L I L L D N I Y L A L K H K G N Y I K L Q K Y V T C C Stop I T Y F G L G W I L

5'3' Frame 3

I S V Q K H S V R Q S R N R V P S L H Stop V V A D R Stop R A R E R F S V S A P N Stop A G A L P H A M A S E K L N K T E M L R R L E E G S V P D E E F
R R L I R I E L G R R L R W Y C Q K K P T Y E E Q T A E I V A L V K A L R Q M G I T G D Q S D Q L N S D L Y D M P M S F L K I T F G K L L K E S G S Y F K
D D S M T L D E A E E A M L D L Y C E R A Q L K D G Q K I L D L G C G Q G A F T L H A A Q K Y K K S H V T A V T N S A T Q K K Y I E D Q C Q I L E L S N
V E V L L E D I T Q L T M E A T F D R I I V I G L L E H M K N Y G L L L Q N I S Q W M A P A D S L L F I D H V C H K S Y F Y Q C E P L D E D D W F A E Y F
F P P G S F A M P S A S F L L Y F Q D D V S I V D H W I L S G K H F H R T A E E W V K Q L D T N L E K G K E I L E S K Y G S K E A A L K A F N H W R G L C
I F S S E I F G Y N G G E E W M T S H L L F K K K Stop G R G I C M L K Y Q C L Y L I I Y S I S Y F L N Stop Y Y Stop I I Y I Stop P Stop N I K V I I L S F K N
M L R A V K Stop L I S G Stop D G Y

3'5' Frame 1

Q Y P S Stop P E I S Y L T A R N I F L K L N I I T F M F Stop G Stop I Y I I Stop Stop Y Stop F K K Stop L I L Y I I K Stop R H W Y F S M Q I P L P Y F F L K
S R Stop E V I H S S P P L Stop P N I S L L N M H N P L Q W L N A F R A A S L L P Y F D S S I S F P F S K L V S N C F T H S S A V R Stop K C F P L R I Q
Stop S T M E T S S W K Stop R R K E A E G I A N E P G G K K Y S A N Q S S S S R G S H W Stop K Stop L L Stop Q T W S I K R R L S A G A I H Stop E I F
Stop R R R P Stop F F I C S S N P I T I I L S N V A S I V S W V I S S S K T S T F E S S S I W H Stop S S M Y F F C V A E F V T A V T Stop L F L Y F Stop A A
Stop R V K A P W P Q P R S R I F C P S F N Stop A L S Q Y K S S I A S S A S S N V M E S S L K Stop L P L S F N N L P K V I L R K D M G I S Y K S E F N
Stop S D Stop S P V I P I C R R A L T K A T I S A V C S S Y V G F F Stop Q Y H R R R L P S S I L I S L R N S S S G T D P S S S L L S I S V L L S F S D A M A
Stop G K A P A Stop L G A E T E N L S L A L Y L S A T T Stop W R L G T L F L L C R T E C F W T E I

3'5' Frame 2

N I H P S P K Stop V I Stop Q H V T Y F Stop S L I Stop L P L C F R A K Y I L S N N I N L R S N Stop Y C I Stop L S K D I G T S A C K F P S P I F S Stop K A
D E K S S I L L L H C S Q I S H C Stop I C I I P S S G Stop T L S G L L L Y Y H I L T P V F P S P S P N Stop Y P T A S P T P P Q C D E N V S R Stop E S N D P
R W K H H L G S R E E K K Q K A Stop Q M S Q G G K N I Q R T S H L H Q E A H T G K N S F C D K H G L Stop R E G Y Q L E P S I E K Y F E E E G H S S S
Y A Q V I Q S L L F Y R M W P P L Stop V G Stop Y L L V R P P H L K A P V S G T D L R C T S S V W L S L Stop L Q Stop H D S S Y I S E Q H E E Stop R L L
G H N P G P E S F V H P S I E L F H S T N P A L L P L P H L M S W S H P Stop S S C H F L S I I C R R Stop S Stop G R T W A Y R T S R S S T D Q T D H R
Stop Y P S A A E P Stop P R R R S R P F V P R M W V S S D S T I G D A F R A R F L S A S G T P R L E Q T L L L A S S A S P F C Stop A F Q T P W R E A R H
Q L N Stop E L R L R I S L S L S I Y R L Q L S G D W G L C F C S A G Q S A S G R K F

3'5' Frame 3

I S I L A R N K L F N S T Stop H I F E A Stop Y N Y L Y V L G L N I Y Y L I I L I Stop E V T N T V Y N Stop V K T L V L Q H A N S P P L F F L E K Q M R S H P
F F S S I V A K Y L T A K Y A Stop S P P V V K R F Q G C F F T T I F Stop L Q Y F L P L L Q I S I Q L L H P L L R S A M K M F P A K N P M I H D G N I I L E V
E K K R S R R H S K Stop A R G E K I F S E P V I F I K R L T L V K I A F V T N M V Y K E K A I S W S H P L R N I L K K K A I V L H M L K Stop S N H Y Y S I E
C G L H C K L G N I F Stop Stop D L H I Stop K L Q Y L A L I F D V L L L C G Stop V C D C S D M T L L I F L S S M K S E G S L A T T Q V Q N L L S I L Q L S
S F T V Q I Q H C F L C L I Stop C H G V I L E V A A T F F Q Stop F A E G D L E E G H G H I V Q V G V Q L I R L I T G N T H L P Q S L N Q G D D L G R L F
L V C G F L L T V P S E T P S E L D S Y Q P P E L L V W N R P F F Stop P P Q H L R F V E L F R R H G V R Q G T S L I R S Stop D Stop E S L S R S L S I G Y
N L V E T G D S V S A L P D R V L L D G N

Fig. 3.8a: Translation of CL957Contig5 into six frames by Translate tool: The longest ORF has been colored in red. All potential “M (Start)” and “Stop” codons have been colored blue and brown respectively.

ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) is a graphical analysis tool which finds all open reading frames of a selectable minimum size in a user's sequence or in a sequence already in the database. This tool identifies all open reading frames using the standard or alternative genetic codes. This tool gave the same results [Fig. 3.8b] with the longest ORF being 1074b (357 aa) long [Table.3.6]. Other ORFs were too short to be considered as protein coding, and no other ORFs overlap the 5' or 3' end of this longest ORF. According to this tool no ORF was found in frame +2. The length (357 aa) of the longest ORF [Fig. 3.8a] is in the range of average length of other known CNMTs (*T. flavum*: 361, *O. sativa*: 359, *C. japonica*: 358, *Chlamydomonas reinhardtii*: 357, *A. thaliana*: 355 and *P. somniferum*: 351). The length of the longest ORF shows that it could represent a full length putative CNMT. However this technique is not very reliable for the identification of the gene and its function. Therefore, the translated sequence of the longest ORF was also searched in the protein database for homologs [309] that further helped in identification.

After translating the unigene sequence, the longest ORF was selected and similarity searches were performed with BLASTp against protein sequences to see whether the ORF of the translated sequence was correctly selected or not, to what proteins it is similar, and what the function of the putative protein would be.

The results of BLASTp show that the most identical protein was a putative NMT of *T. flavum* [311] with an E.V of 1e-121 [Table. 3.7]. Among the rest top 20 results second, third, fifth, twelfth, fourteenth and eighteenth were all CNMTs from *C. japonica* [15], *T. flavum* [17], *P. somniferum* [98] and *A. thaliana* (three sequences) respectively [314]. Similar to BLASTx results, most of the matches after the top 20 results were CFAPS.

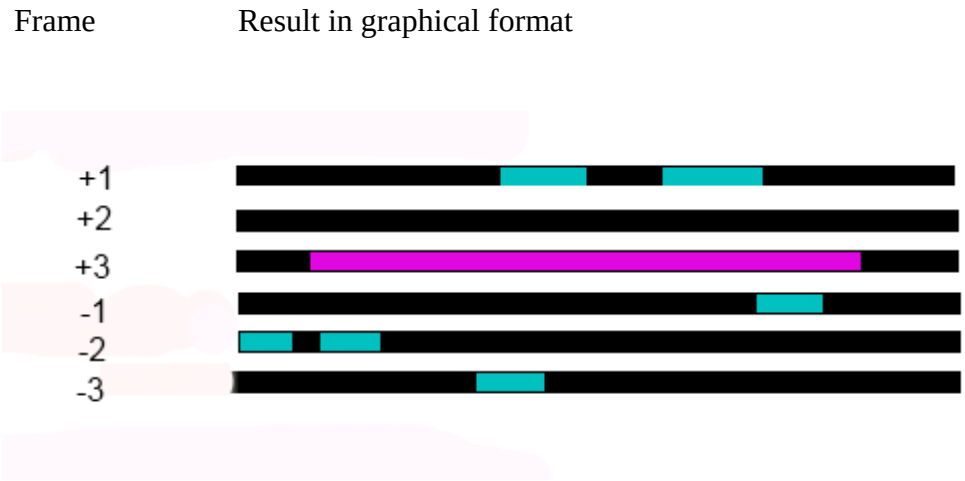


Fig. 3.8b: Graphical representation of the ORFs for all the six frames of CL957Contig5 obtained with the help of ORF finder tool of NCBI: The ORFs have been highlighted in green; however the longest ORF, identified in frame +3, is given in purple color.

Table 3.6: Lengthwise ranking of the ORFs, identified in the CL957Contig5, by ORF finder

Rank	Frame	From	To	Length	Protein Length
1	+3	141	1214	1074	357
2	+1	835	1032	198	65
3	+1	520	690	171	56
4	-3	462	599	138	45
5	-1	1007	1138	132	43
6	-2	157	279	123	40
7	-2	1	108	108	35

Table 3.7: Top 20 best hits for *A. fimbriata* putative CNMT sequence obtained through BLASTp search of NCBI

S. No	Accession	Organism	Description	Score	E.V.
1	ACO90253.1	*T. flavum*	NMT (putative)	440	1e-121
2	BAB71802.1	*C. japonica*	CNMT	431	5e-119
3	AAU20766.1	*T. flavum*	CNMT	424	8e-117
4	ACO90251.1	*T. flavum*	PNMT	399	4e-109
5	AAP45316.1	*P. somniferum*	CNMT	382	3e-104
6	ACO90222.1	*E. californica*	TNMT	362	3e-98
7	XP_002324742	*P. trichocarpa*	SAM-MTase: predicted	345	4e-93
8	ACO90236.1	*P. bracteatum*	TNMT like NMT (putative)	345	53-93
9	ACO90237.1	*P. bracteatum*	TNMT	335	6e-90
10	AAY79177.1	*P. somniferum*	TNMT	332	5e-89
11	ACF85951.1	*Z. mays*	SAM-MTase like (unknown function)	329	3e-88
12	AAM65762.1	*A. thaliana*	CNMT	329	3e-88
13	BAC42939.1	*A. thaliana*	SAM-MTase like (unknown function)	328	6e-88
14	NP_567912.1	*A. thaliana*	CNMT (putative)	327	8e-88
15	EEE65917.1	*O. sativa*	SAM-MTase (hypothetical protein)	321	1e-85
16	BAD61850.1	*O. sativa*	CNMT (putative)	320	1e-85
17	EEC80845.1	*O. sativa*	SAM-MTase (hypothetical protein)	320	1e-85
18	NP_195038.2	*A. thaliana*	CNMT (putative)	320	1e-85
19	XP_002280292	*V. vinifera*	SAM-MTase (hypothetical protein)	320	2e-85
20	XP_002437192	*S. bicolor*	CNMT like hypothetical protein	316	2e-84

Almost all the sequences were MTases and many of them were SAM-dependent. Most of the similarity matches of *A. fimbriata* CNMT were annotated based on their homology to other experimentally expressed proteins. *C. japonica* CNMT however is the only CNMT that has been experimentally proved to be a CNMT. On the basis of similarity of *A. fimbriata* CNMT to *C. japonica* CNMT (both nucleotide and protein sequences), it can be confidently predicted that the cDNA isolated from *A. fimbriata* encodes a putative CNMT. In addition to coclaurine N-methyltransferases, some TNMTs also showed a very good similarity e.g. sixth, ninth and tenth most similar sequences were TNMTs from *E. californica*, *P. bracteatum* and *P. somniferum* (experimentally confirmed) [66]. This is because CNMTs and TNMTs are very closely related proteins.

3.5.4. Protein Parameters

Several tools are available online for the identification and analysis of a protein sequence [315]. Some of them have been used in this study for the identification of the predicted protein. To find out the composition of the translated protein ProtParam tool of ExPASy was used. The results show that the protein contains all the amino acids with Leu in the highest percentage (12 %), and Cys and Trp in the lowest (2 % each) [Table. 3.8]. The amino acid composition is very similar to the amino acid compositions of other plant CNMTs [Table.3.9]. The number of negatively charged residues (Asp + Glu) was greater than the number of positively charged residues (Arg + Lys) in *A. fimbriata* putative CNMT. It means the protein is intracellular. It has been shown that intracellular and extracellular proteins possess different amino acid compositions. Intracellular proteins tend to have a higher fraction of negatively charged residues than extracellular proteins [316, 317].

Table 3.8: Amino acid composition of *A.fimbriata* putative CNMT, calculated using the ProParam tool of ExPASy

Amino Acid	Number of Residues	Percentage
Leucine	43	12.0
Glutamic acid	33	9.2
Lysine	28	7.8
Serine	24	6.7
Aspartic acid	23	6.4
Alanine	22	6.2
Glycine	20	5.6
Phenylalanine	19	5.3
Isoleucine	19	5.3
Glutamine	18	5.0
Threonine	16	4.5
Tyrosine	14	3.9
Arginine	13	3.6
Methionine	12	3.4
Valine	12	3.4
Histidine	10	2.8
Asparagine	9	2.5
Proline	8	2.2
Cystine	7	2.0
Triptopane	7	2.0

Table 3.9: Comparison of amino acid compositions of known CNMT sequences from different species of the plants and an alga

Amino acid	Percentage of amino acids in CNMTs of different plants and alga					
	A. fimbriata	*A. thaliana*	*C. japonica*	*T. flavum*	*O. sativa*	*C. reinhardtii*
Ala	6.2	5.9	5.9	5.3	9.5	7.0
Arg	3.6	3.9	4.2	3.9	5.6	7.3
Asn	2.5	2.8	4.2	4.2	3.6	5.0
Asp	6.4	5.4	5.3	5.3	4.5	3.9
Cys	2.0	2.0	2.0	2.2	1.9	1.1
Gln	5.0	1.7	4.2	3.9	2.2	3.4
Glu	9.2	8.7	10.6	9.1	8.1	6.4
Gly	5.6	5.6	4.5	5.0	3.6	6.4
His	2.8	3.1	2.8	2.8	3.3	3.1
Ile	5.3	6.8	6.1	6.1	5.6	3.6
Leu	12.0	10.1	11.7	10.8	11.4	10.4
Lys	7.8	8.7	7.0	8.3	6.1	3.4
Met	3.4	3.7	3.4	3.9	3.1	2.0
Phe	5.3	5.6	4.2	4.4	5.8	6.7
Pro	2.2	1.7	2.0	2.2	2.2	3.6
Ser	6.7	6.5	5.0	5.0	7.5	7.0
Thr	4.5	4.8	5.3	5.3	5.8	5.6
Trp	2.0	2.3	2.0	1.9	2.2	2.2
Tyr	3.9	4.8	4.5	3.9	4.7	5.3
Val	3.4	5.9	5.3	6.6	3.1	6.4

ProtParam results show that 56 (15.7 %) residues are negatively charged and 41(11.5 %) residues are positively charged. The Instability Index (A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable) [318] of the protein was calculated to be 43.05. This classifies the protein as unstable. The molecular weight of the predicted protein is 41325.2. The isoelectric point was calculated to be 5.21 which indicates that the protein have slight acidic characters. This lower value is due to the more negatively charged (acidic) residues as compared to the positively charged residues.

Different parameters calculated by ProtParam of five species (*C. reinhardtii, A. thaliana, T. flavum, C. japonica* and *O. sativa*) were compared with *A. fimbriata* putative CNMT sequence [Table. 3.9 and Table. 3.10]. *A. fimriata* values for different parameters {length of the CNMT, molecular weight (mw), isoelectric point (pI), total number of negatively charged and positively charged residues, instability index, aliphatic index [319] (higher value shows higher thermostability) and grand average of hydropathicity [320]} are very similar and are in the range of average values for other known CNMTs. These similarities confirmed further identity of *A. fimbriata* putative CNMT.

3.5.5. Protein Properties and Functional Analysis

No signal peptide was detected in the protein sequence according to the computational results of Signal-CF (A subsite-coupled and window-fusing approach for predicting signal peptides) and Signal-3L (a 3-layer approach for predicting signal peptides). The results of ProtIdent (A web server for identifying proteases and their types by fusing functional domain and sequential evolution information) show that the protein will not show any protease activity.

Table 3.10: Comparison of different parameters for known CNMTs sequences

Species	Length	mw	pI	Asp+Glu	Arg+Lys	Instability Index	Aliphatic index	Grand average hydropathicity
A. fimbriata	357	41325.2	5.21	56	41	43.05	83.64	-0.374
C. reinhardtii	357	41239.8	7.84	37	38	45.16	80.31	-0.275
A. thaliana	355	41229.7	6.16	50	45	37.23	88.99	-0.164
T. flavum	361	41903.4	5.77	52	44	32.67	90.44	-0.261
C. japonica	358	41780.8	5.12	57	40	42.00	90.98	-0.322
O.sativa	359	41623.7	6.47	45	42	45.40	84.62	-0.218

The parameters include length of the CNMT, molecular weight (mw), isoelectric point (pI), total number of negatively charged (Asp+Glu) and positively charged (Arg+Lys) residues, instability index, aliphatic index and grand average of hydropathicity.

EnzyPred predicted the protein to be a transferase enzyme that transfers one carbon group. The protein is not a membrane protein according to the Mem-Type-2L approach. All the methods applied for sub-cellular localization predictions showed that this is a cytoplasmic protein.

To identify the function of the protein two other approaches were also used. One was to search for the proteins that have nearly the same pI and mw and the other was to identify proteins with nearly similar amino acid compositions. For this purpose, TagIdent and AACompIdent were used. TagIdent searched out five (top three shown in the table) proteins from SwissProt and two (only first hit is shown in the table) proteins from TrEMBL. The closest protein to our protein in terms of pI and mw was CNMT from *C. japonica* [Table. 3.11].

The closest SwissProt entries (in terms of aa composition) and having pI and mw values in the specified range (mw: 33060-49590, pI: 4.71-5.71) for the "COPTIS" species, was (S)-norcoclaurine synthase 1 (mw: 39964, pI: 5.27) with a score of 44. The closest TrEMBL entries (in terms of aa composition) and having pI and mw values in the specified range for the species "COPTIS" was CNMT (mw: 41781, pI: 5.12) with a score of 19. The pI and mw values of the *A. fimbriata* putative CNMT have already been compared with other available CNMTs [Table. 3.10].

The superfamily of the protein was identified to be S-adenosyl-L-methionine-dependent methyltransferases by InterProScan.

Pfam identified four domains in the putative CNMT sequence. Three of them i.e. NodS, Methyltransf_11 and CMAS, have been given in the table 3.12.

Table 3.11: TagIdent results showing the proteins with specified ranges of isoelectric point and molecular weight

Database	ID	Name	pI	mw	Species
UniProtKB/ Swiss-Prot	A2A1A0	S-norcoclaurine synthase 1	5.28	39964	*C. japonica*
	Q39522	(S)-scoulerine 9-O-methyltransferase	5.37	41665	
	Q9LEL5	3'-hydroxy-N-methyl-(S)-coclaurine 4'-O-methyltransferase	5.33	38776	
UniProtKB/TrEMBL	Q948P7	CNMT	5.12	41781	

The specified range was according to the calculated values of pI and mw for the putative CNMT of *A. fimbriata*

Table 3.12: Pfam domains identified in the predicted CNMT sequence

Domain	Accession	Go Process	Go Function	Start	End
NodS	PF05401	Oligosaccharide biosynthetic	SAM-MTase	85	261
Methyltransf_11	PF08241	Metabolic	Methyltransferase	132	229
CMAS	PF02353	Lipid biosynthetic	CFAPS	69	323

Three domains have been identified by the Pfam. The start residue number and the end residue number for the given domains have been given in the table.

These results show that the protein is included in CMAS (cyclopropane-fatty-acyl-phospholipid-synthase) family. NodS domain shows that the superfamily of the protein is S-adenosyl-L-methionine-dependent methyltransferases, which shows that the protein is dependent upon a SAM ligand for its function. It might be using SAM to transfer its methyl group to the substrate. Methyltransf_11 domain shows that the protein has the function of transferring of a methyl group. For confirmation of the results, Pfam database was also scanned for the CNMT sequence of *P. somniferum*. The same domains were identified in the CNMT sequence of *P. somniferum*. This shows that the Pfam results were correctly identified for the putative CNMT of *A. fimbriata* and that these two proteins belong to the same family, hence it can be concluded that they might have similar functions. All the above mentioned predictions are in the favor of its CNMT nature.

3.5.6. ModBase Search

The ModBase database was scanned for theoretical models of similar sequences. In the option window the sequence identity was changed to 40% while the other options were left default. Only 10 models were found, out of which five were CNMTs from *C. japonica*, *T. flavum*, *P. somniferum*, *O. sativa* and *A. thaliana* with sequence identities of 57 %, 55 %, 50 %, 50 % and 46 % respectively [Table. 3.13]. Other five models sequences were either less similar or the proteins were uncharacterized. These results show that the sequence isolated from *A. fimbriata* might also be a CNMT encoding gene.

Table 3.13: Top 10 best hits of *A. fimbriata* putative CNMT sequence obtained through BLAST search against ModBase database that contains theoretical models of proteins

S.No	ID	Annotation	Organism	Size	Coverage	Coverage (%)	Identity (%)	E.V.
1	Q9SMZ7	AT4G3 3120 (Putative protein)	*A. thaliana*	296	79-357	78	47	3e-74
2	Q8L9U0	CNMT	*A. thaliana*	355	44-357	86	46	3e-83
3	Q9SMZ8	AT4G3 3110 (Putative protein)	*A. thaliana*	174	44-172	35	40	3e-25
4	Q5C9L6	CNMT	*T. flavum*	361	1-357	99	55	e-109
5	Q84TE2	AT4G33120 (Putative protein)	*A. thaliana*	355	44-357	86	46	1e-79
6	Q8L788	AT4G33110 (Putative protein)	*A. thaliana*	355	44-357	86	46	8e-83
7	Q8GXB6	AT4G33110 (Putative protein)	*A. thaliana*	355	44-357	86	46	6e-83
8	Q948P7	CNMT	*C. japonica*	358	3-357	98	57	e-111
9	Q7XB08	CNMT	*P. somniferum*	351	8-357	96	50	3e-96
10	Q5Z7K6	CNMT (Putative)	*O. sativa*	359	54-357	83	50	2e-85

3.5.7. Conserved Domain Database Predictions

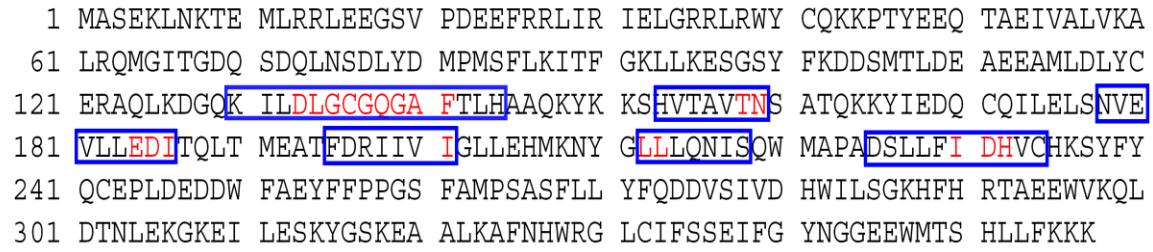
The Conserved Domain Database (CDD) is the protein classification component of NCBI's Entrez query and retrieval system. This tool can help us identify the functional class of a protein [321]. The *A. fimbriata* putative CNMT sequence was scanned for conserved residues by CDD. The CDD results show that the most conserved residues are in between residues 130-234 [Fig. 3.9] of the putative CNMT protein sequence of *A. fimbriata*. Thirteen residues (134-140 158-159, 184-186 and 201) of the segment were found to be the most conserved residues. These residues are 134-LGCGQGA-140, 158-TN-159, 184-EDI-186 and 201-I. These residues might be the ligand binding residues. All the SAM-MTase enzymes contain almost the same residues and are the binding sites for SAM ligand. From this it was predicted that the *A. fimbriata* putative protein might also be a SAM-MTase.

3.6. Homology Model Building

3.6.1. Template Identification

Top 10 best hits of BLAST results of the putative CNMT sequence against PDB have been shown in table. 3.14. The most similar hit was that of 1kp9 and 1kph with E.V. of 1e-11 and a 30 % sequence identity. Other most similar matches were 1l1e, 1kpg, 1kpi, 1tpy, 2fk8/2fk7, 3bus, 3A27, 2cww and 1wxw with E.values; 1e-11, 8e-11, 2e-10, 9e-09, 2e-08, 2e-07, 4e-05, 0.016, 0.044 and 0.051 respectively.

Three crystal-structure coordinates i.e. 1kph, 1kpi and 2fk8 were selected as the templates because of their highest identity with the target sequence and similarity of their secondary structures (next topic).



1: 130-KILDLGCGQGAFTLH-144

2: 153-HVTAVTN-159

3: 178-NVEVLLEDI-186

4: 195-FDRIIVI-201

5: 212-LLLQNIS-218

6: 225-DSLLFIDHVC-234

Fig. 3.9: Conserved residues of the *A. fimbriata* putative CNMT as identified by conserved domain database tool of NCBI: The conserved residues have been enclosed in blue boxes and the residues that were identified as the most conserved have been colored red. Six positions have been identified as the conserved regions as shown with numbers indicating the starting residue number and the ending residue number.

Table 3.14: Top 10 best templates for *A. fimbriata* putative CNMT sequence obtained through PSI-BLAST against PDB database

S.No	PDB ID	Chain ID	Name	Bits	E.V.
1	1kp9, 1kph	A/B A/B/C/D	Crystal structure of MACPS	67.0	1e-11
2	1l1e	A/B	Crystal structure of MACPS	63.9	8e-11
3	1kpg	A/B/C/D	Crystal structure of MACPS	62.8	2e-10
4	1kpi	A	Crystal structure of MACPS	57.4	9e-09
5	1tpy	A	Crystal structure of Cyclopropane Synthase Mmaa2 from *M. tuberculosis*	56.2	2e-08
6	2fk7/2fk8	A	Crystal structure of Hma (Mmaa4) from *M. tuberculosis*	53.1	2e-07
7	3bus	A	Crystal structure of Rebm	45.1	4e-05
8	3A27	A	Crystal structure of Tyw2 from *M. Jannaschii*	36.6	0.017
9	2cww	A/B	Crystal structure of Ttha1280 from *Thermus thermophilus*	35.0	0.044
10	1wxw	A/B/C/D	Crystal structure of Tt1595, a putative SAM-MTase from *T. thermophillus*	34.7	0.052

The putative CNMT sequence of *A. fimbriata* shows 30 %, 27% and 25 % identity with the template sequences of 1kph, 1kpi and 2fk8 respectively. The comparison of the putative CNMT's secondary structure with that of the selected templates is shown in figure 3.10. 1kph was selected on the basis of highest identity, 1kpi was selected because the online homology modeling server SWISS-MODEL [322] selected it as the most suitable template on the bases of "HHSearch template library search". Also the phylogenetic tree calculated using all the top templates showed that 1kpi was the best template [Fig. 3.11]. 2fK8 was selected on two bases, first it aligns to the greater part of our sequence and second it contains the ligand SAM (to include SAM in the model). Over and above, all the matches are reliable when the corresponding E-value is less than 10^{-4} [323]. Most of the templates fall in this category, but due to less sequence identity, they were not selected as templates because the accuracy of comparative modeling is related to the percentage sequence identity on which the model is based [324, 325].

Highly reliable models are usually based on >50% sequence identity, medium-accuracy models are based on 30–50% sequence identity and low accuracy models are based on <30% sequence match [323]. As we see the identity is low in case of putative CNMT with its templates, therefore to improve the accuracy of models, three templates were selected, because using multiple templates give better results [326, 327]. To improve the model accuracy the alignment will be based on both sequence as well as secondary structure identity with the templates.

An MP tree rooted with 2fk7 [Fig. 3.11] based on multiple alignment [Fig. 3.12] was also calculated to search for the best template for homology modeling of the *A. fimbriata* putative CNMT.

Fig. 3.10: Comparison of the secondary structure of *A. fimbriata* putative CNMT sequence with that of the templates

Rows: 1st=*A. fimbriata* CNMT, 2nd=HNN prediction, 3rd=JPred-3 prediction, 4th=APSSP prediction, 5th= our own predictions based on the above three predictions, 6th=1kph, 7th=1kpi, 8th=2fk8

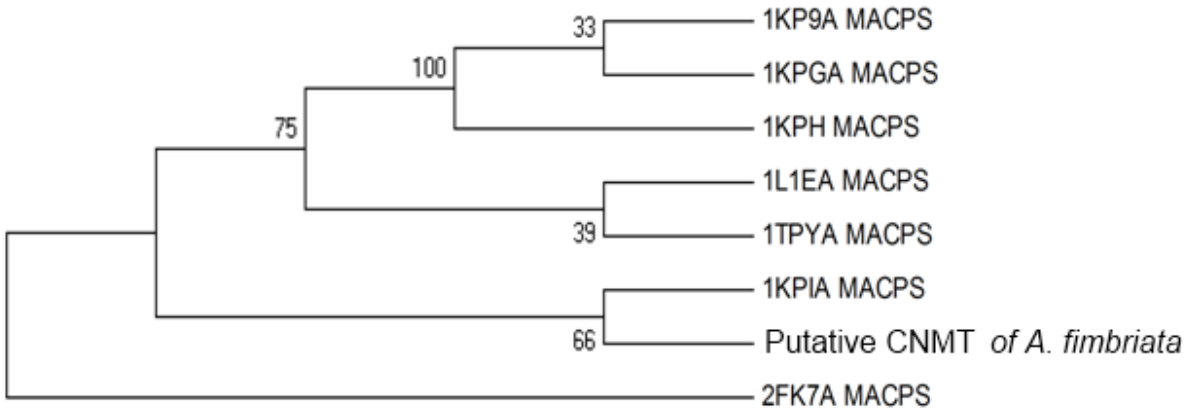


Fig. 3.11: Evolutionary relationships of seven homologous template protein sequences with the putative CNMT sequence from *A. fimbriata*

The evolutionary history was inferred using the Maximum Parsimony method. The unrooted bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the proteins analyzed. Rooting is arbitrary and could potentially be placed anywhere in the tree until further information becomes available. In this tree 2fk7A sequence has been selected as root. 1kpi sequence clustered with the putative CNMT indicating that 1kpi is the closest homologue among the seven selected template sequences and was selected as one of the templates for homology model building. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated sequences clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The MP tree was obtained using the Close-Neighbor-Interchange algorithm with search level 3 in which the initial trees were obtained with the random addition of sequences (10 replicates). All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 266 positions in the final dataset, out of which 74 were parsimony informative. Phylogenetic analyses were conducted in MEGA4.

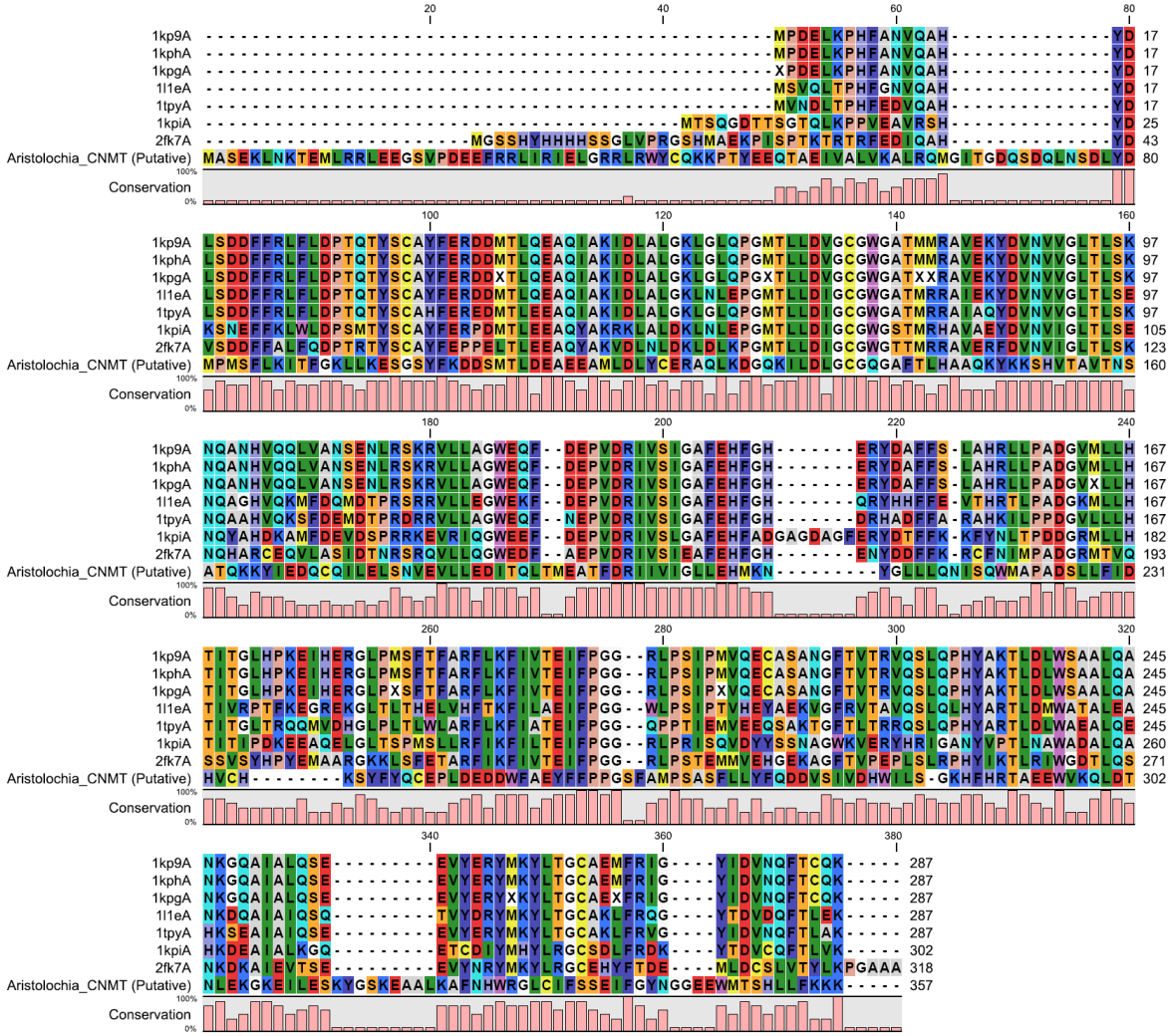


Fig. 3.12: Multiple alignment among the templates sequences with the *A. fimbriata* sequence: Only the top seven templates were selected for this multiple alignment, i.e. 1kp9A, 1kph, 1kpgA, 1l1eA, 1tpyA, 1kpiA and 2fk7A. The image has been created using CLC-sequence viewer version-6.4 (ww.clcbio.com) with rasmol background color option. The graph shows the percent conservation of the residues.

The tree shows that 1kpi is the best template for modeling the *A. fimbriata* putative CNMT. Both the sequences are found in a single clade with a good support of 66; however in this investigation multiple templates have been used as it calculates a better model as compared to models calculated using single templates. Two other templates 2fk8 and 1kph were also selected on the bases already explained.

3.6.2. Secondary Structure Analysis

Secondary structure predictions are important for calculating 3D models. Errors that result due to incorrect alignment, and fold assignment are the most important limitations in comparative and homology modeling [328]. Protein sequence gives us important information about the secondary structure of the protein, as history tells Linus Pauling correctly guessed the formation of helices and strands [329, 330] from the sequence of the protein, which were later on confirmed by the X-ray crystallographic studies [331, 332]. Therefore we analyzed the sequence for secondary structure of the putative CNMT sequence of *A. fimbriata*. The secondary structure of the sequence was analyzed by three ab initio protein secondary structure prediction programs i.e. HNN, JPred-3 and APSSP [Fig. 3.13, Table. 3.15]. Predictions were performed on full length of the putative CNMT of *A. fimbriata* and compared with that of the templates. In general, the various secondary structure prediction methods give similar results. The HNN results show that 54.62 % of the residues of the sequence are related to alpha helix, 9.8 % to extended beta strand and 35.57 % to random coil.

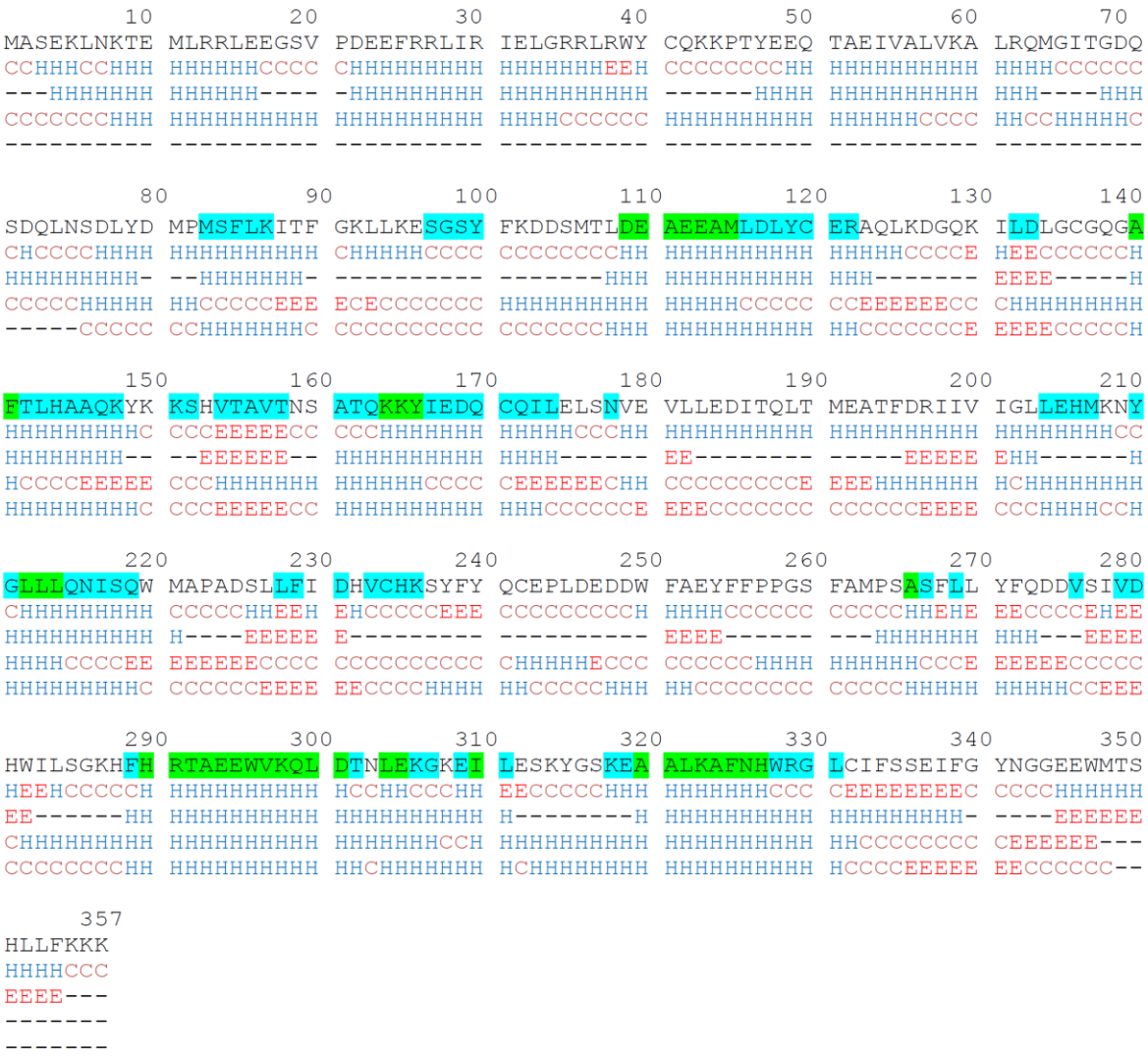


Fig. 3.13: Secondary structure comparison of the putative CNMT model of *A. fimbriata* with three different ab initio prediction methods

The first row represents the query sequence, the second row shows the HNN predictions, the third row shows the JPred-3 predictions, the fourth row shows the APSSP predictions and the fifth row shows the secondary structure derived from the predicted model. Alpha helix, beta strands and coil regions have been represented with H (blue), E (red) and C (brown) respectively. The highlighted regions show that the model prediction is more reliable in these regions because these regions show that either all methods have predicted the same secondary structure (green) or three of the methods have predicted the same secondary structure (blue).

Table 3.15: Secondary structure analysis of the putative CNMT of *A. fimbriata* with three different methods

Method	Alpha helix	Beta strands	Random coil
HNN	3-5,8-16, 22-37, 40, 59-64, 72, 77-90, 92-96, 109-125, 131, 140-149, 164-175, 179-208, 212-220, 226-227, 230, 232, 250-254, 266-267, 269, 278, 281, 284, 290-301, 304-305, 309-310, 318-327, 345-354	38-39, 130, 132-133, 154-158, 228-229, 231, 238-240, 268, 270-272, 277, 279-280, 282-283, 311-312, 323-339	1-2, 6-7, 17-21, 41-58, 65-71, 73-76, 91, 97-108, 126-129, 134-139, 150-153, 159-163, 176-178, 209-211, 212-225, 233-237, 241-249, 255-265, 273-276, 285-289, 302-303, 306-308, 313-317, 328-331, 340-344, 355-357
JPred-3	4-16, 22-40, 47-63, 68-79, 83-89, 108-123, 140-148, 161-174, 202-203, 210-221, 264-273, 289-311, 320-339	131-134, 153-158, 181-182, 196-201, 226-231, 251-254, 277-282, 345-354	1-3, 17-21, 41-46, 64-67, 80-82, 90-107, 124-130, 135-139, 149-152, 159-160, 175-180, 183-195, 204-209, 222-225, 232-250, 255-263, 274-276, 283-288, 312-319, 340-344, 355-357.
APSSP	8-34, 41-56, 61-62, 65-69, 76-82, 101-115, 132-141, 154-166, 179-180, 194-201, 203-214, 242-246, 257-266, 282-307, 310-332	88-91, 93, 123-128, 146-150, 172-177, 190-193, 219-226, 247, 270-275, 242-247	1-7, 35-40, 57-60, 63-64, 70-75, 83-87, 92, 94-100, 116-122, 129-131, 142-145, 151-153, 167-171, 178, 181-189, 202, 215-218, 227-241, 248-256, 267-269, 276-281, 308-309, 233-241
Homology Model	83-89, 108-122, 140-149, 161-173, 204-207, 210-219, 237-242, 248-252, 266-275, 289-302, 304-311, 313-331	130-134, 154-158, 180-183, 197-200, 227-232, 278-280, 336-342	76-82, 90-107, 123-129, 135-139, 150-153, 159-160, 174-179, 184-196, 201-203, 208-209, 220-226, 233-236, 243-247, 253-265, 276-277, 281-288, 332-335, 343-348

First row contains the HNN predictions, the second row shows the JPred-3 predictions the third row shows the APSSP predictions and the fourth row shows the secondary structure observed in the homology model. The numbers represent the number of amino acid of query sequence. The secondary structure predictions that agree with the homology model have been highlighted in blue. The red colored residue numbers in the homology model shows that they didn't agree with any of the three predictions.

As is clear from the results, segments 8-16, 21-34, 49-56, 77-82, 109-115, 164-166, 212-214, 290-301 and 320-327 have been predicted to be alpha helices by all three methods applied. In addition, sequence segments 35-37, 47-48, 57-60, 83-89, 106-123, 142-149, 167-174, 194-211, 215-220 and 328-332 have been predicted to be alpha helices by two methods (any two of the three) while the third (any one of the three) method has given different predictions. The sequence segments 125-132, 145-158, 171-177, 190-201, 222-231, 251-254, 270-282 and 341-355 were predicted to be extended beta sheets by either one or two methods. The segments 6-7, 69-75, 96-100 and 231-237 were predicted to be random coils by two methods (any two) and the third (any one) method has given different predictions. The predictions of the APSSP were most different from the other two methods' predictions. Some parts predicted to be coils by the APSSP, were predicted to be helices by the other two methods and vice versa. The secondary structure of the putative CNMT, as evident from homology model, was very much similar to the predictions made by JPred-3 and HNN; however it was more identical to the JPred-3 predictions. It showed lesser identity with the APSSP predictions. As a whole, the model was very much similar to the secondary structure predicted by the above mentioned methods. Only two segments 237-242 and 343-348 of the model were not according to the predicted secondary structures. The segment 237-242 was an alpha helix in the model, however it was predicted to be either a beta strand (HNN) or a random coil (APSSP and JPred-3). The segment 343-348 was built as a random coil; however it was predicted to be alpha helix (HNN) or beta strand (JPred-3). Both these segments are not the part of the active site.

The comparison of secondary structure was made by aligning the putative CNMT sequence with the selected templates sequences. The secondary structure of 1kphA showed the highest identity in its secondary structure to the secondary structure of the

putative CNMT. Its secondary structure was mostly similar to the predictions of JPred-3 [Fig. 3.10]. The secondary structure of 1kpi was mostly similar to the HNN and APSSP predictions and 2fk8 showed highest similarity with the APSSP predictions; however it also showed a very good similarity with HNN predictions. These two templates showed a very low similarity to JPred3 predictions. As a whole, all the three templates were very similar in their secondary structures to that of the putative CNMT secondary structure. These three structures were used as templates on the basis of their similarity both in their sequences and secondary structures.

3.6.3. Model Generation

It has already been shown that BLAST searching of the non-redundant database of all known protein sequences showed that the *A. fimbriata* unigene CL957Contig5 sequence as well as its translated putative CNMT sequence is similar to many other proteins defined as SAM-MTases especially CNMTs from different plants. Although the sequence shows similarity with CNMTs, however no CNMT with experimentally determined three-dimensional structure on PDB database was found. The most identical sequences with experimentally determined three-dimensional structures were MACPS from *M. tuberculosis* as discussed above. As shown above the templates showed not more than 30 % identities. The similarity at this very low level needs a very careful analysis to create a 3D model of a protein by comparative homology modeling. The main challenge here is related to the finding of the right sequence alignment, because sequences with low similarity can be aligned differently giving rise to similar scores, and gaps, required to align the sequences in a better way, can happen inside the secondary structure elements, which can lead to obtaining a wrong model [333, 334]. Therefore @TOME 2.1 was used for the alignment [Fig. 3.14].

```
>P1;CNMT
sequence:CNMT:     : :     : :CNMT:Aristolochia fimbriata: 0.00: 0.00
SDLYDMPMSFLKITFGKLLKESGSYFKDDSMTLDEAEEAMLDLYCERAQLKDGQKILDLGCGQGAFTLHAAQKYKKSHVTAVTNSATQKKYIEDQCQ
ILE-LS-NV---EVLL---EDITQLTMEATFDRIIVIGLLEHM--KNYGLLLQNISQWMAPADSLLFI-DHVCHKSYFYQCEPLD---EDDWFAEY-
---FFPPGSFAMPSASFLLYFQDDVSIVDHWILSGK-HFHRTAEEWVKQLDTNLEKGKEIL-ESKYGSKEAALKAFNHWRGLCIFSSEIFGYNGGEE
WM,*
>P1;1KPH
structureX:1KPH.pdb:16 :A:169 :A:CFAPS:Mycobacterium tuberculosis: 2.00:-1.00
---YDLSDDFFRLFLDPTQTYSCAYFERDDMTLQEAQIAKIDLALGKLGLQPGMTLLDVGCGWGATMMRAVEKY-DVNVVGLTLSKNQANHVQQLVA
NSENLR-SK---RVLL---AGWEQF--DEPVDRIVSIGAFEHFGHERYDAFF-SLAHRLLPADGVMLL-HTI-------------------------
------------------------------------------------------------------------------------------------
---*
>P1;1KPI
structureX:1KPI.pdb:24 :A:152 :A:CFAPS:Mycobacterium tuberculosis: 2.00:-1.00
---YDKSNEFFKLWLDPSMTYSCAYFERPDMTLEEAQYAKRKLALDKLNLEPGMTLLDIGCGWGSTMRHAVAEYD---VNVIGLTLSENQYAHDKAM
FDE-VD-SPRRKEVRI---QGWEEF--DEPVDRIVSLGAFEHF--AD----------------------------------------------------
------------------------------------------------------------------------------------------------
---*
>P1;2FK8
structureX:2FK8.pdb:25 :A:302 :A:CFAPS:Mycobacterium tuberculosis: 2.00:-1.00
---YDVSDDFFALFQDPTRTYSCAYFEPPELTLEEAQYAKVDLNLDKLDLKPGMTLLDIGCGWGTTMRRAVERF-DVNVIGLTLSKNQHARCEQVLA
SID-TNRSR---QVLL---QGWEDFAE--PVDRIVSIEAFEHFGHENYDDFFKRCFNIM-PADGRMTVQSSVSYHPYEMAARGKKLSFETARFIKFI
VTEIFPGGR--LPSTEMMVEHGEKAGFTVPEPLSLRPHYIKTLRIWGDTLQSNKDKAIEVTSEEVYNRYMKYLRGCEHYFTDEMLDCSLVTYLKPGA
AA,*
```

Fig. 3.14: Alignment of *A. fimbriata* putative CNMT sequence with the three selected templates i.e. 1kph, 1kpi and 2fk8. The sequences were aligned by @tome2.1 for homology modeling: The conserved residues 59-65, 83-84, 125 of CNMT have been aligned with the conserved residues of the templates as highlighted in red.

@TOME 2.1 is a homology modeling server, already explained in methodology chapter. It uses both sequence identity as well as secondary structure similarity for making an alignment between the target and templates. The server aligned the conserved parts of the sequence with the conserved parts of the templates. The conserved part of the sequence i.e. 76-348 was aligned with the conserved parts of the templates i.e. 16-169, 24-152 and 25-302 of 1kph, 1kpi and 2fk8 respectively.

Starting from this alignment, one hundred ten (110) models were created for the *A. fimbriata* putative CNMT protein. The models were built for the segment starting from residue No. 76 to 348 which is also the most conserved region of the protein family as has been clear from multiple alignment, and Pfam results.

3.6.4. Models Evaluation and Selection of the Best Model

All the models created were evaluated and the best model was selected on the basis of results of PROCHECK and ProSA as described below. ProSA is a tool that is used to evaluate 3D models of protein structures for potential errors. This tool is used for error recognition in experimentally determined structures [335, 336], theoretical models [337, 338] and protein engineering [339].

3.6.5. PROCHECK and ProSA Results

All models were ranked on the basis of the values obtained for the core, disallowed, number of bad contacts, maximum deviation and G-factor (overall) with PROCHECK. According to core region values models no. 108, 109 and 27 (model no. 77 is identical to model no. 27) were ranked first, second and third respectively. According to disallowed region values models no. 41 (91 identical to 41), 102 and 110 were ranked first because they had 0.0 % of residues in disallowed region.

Models no. 1, 3, 8, 20, 26, 28, 29, 35, 37, 43, 45, 51, 53, 58, 70, 76, 78, 79, 85, 87, 93, 95, 103 and 104 were ranked second with disallowed region of 0.4 %. On the basis of number of bad contacts model No. 104 was ranked first with the lowest number of bad contacts (only 10), models no. 44, 94, 106 were ranked second and models No.13, 46, 63, 96 and 105 were ranked third. According to maximum deviation values models no. 2, 16, 45, 52, 66 and 95 were ranked first, models no. 3, 21, 53 and 71 as second, 19 and 69 as third, and 17, 25, 37, 67, 75 and 87 as fourth. On the basis of G-factor value (for overall model quality: the closer it is to zero the better the structure) [340, 341] models no. 2, 16, 45, 52, 66 and 95 were ranked first, 3, 21, 53 and 71 as second, 19 & 69 as third while 17, 25, 37, 67, 75 and 87 as fourth. On the basis of these results, we selected top first and second ranked models from each group for further evaluation with ProSA. Model no. 27 (77) was also included because of its high value (88.4) of core region [Tables.3.16 and 3.17].

All the selected models were evaluated further with ProSA. From the energy diagram [Fig. 3.15a] it is clear that models no. 3, 19, 41 (91), 35 (85), 39, 104 and 107 have the lowest energy values as compared to the rest of the selected models. A separate graph [Fig. 3.15b] was calculated for these selected models to see which may be the best. All these models showed very similar energy values for the first 1-110 residues and 140-160. Model no. 41 has the lowest overall value. Model no. 104 has the lowest value in between residues 170-195, but for the rest of the part it has higher value as compared to model no. 41. Although among other models some shows lower energy as compared to model no. 41, but in this study model no. 41 was preferred over the other due to two reasons.

Table 3.16: PROCHECK results for 110 homology models created with MODELLER

S.No	Core	Disallowed	Bad contacts	Max deviation	G-factor: overall
1	84.7	0.4	26	19.0	-0.34
2	86.7	1.2	14	4.0	-0.28
3	87.6	0.4	23	4.1	-0.25
4	84.3	0.8	15	12.4	-0.27
5	86.7	0.8	15	18.6	-0.25
6	85.9	2.4	18	18.7	-0.20
7	82.3	1.2	25	18.3	-0.27
8	85.5	0.4	25	18.3	-0.28
9	86.7	1.6	22	4.5	-0.27
10	86.7	1.2	23	4.8	-0.23
11	85.9	1.6	17	18.7	-0.27
12	85.9	1.6	24	10.4	-0.28
13	87.1	1.2	13	5.3	-0.23
14	86.7	1.2	26	6.4	-0.25
15	85.9	0.8	19	18.3	-0.27
16	85.5	1.6	18	4.0	-0.29
17	86.3	0.8	20	4.3	-0.28
18	83.5	0.8	16	5.4	-0.25
19	87.6	1.6	17	4.2	-0.20
20	85.9	0.4	16	18.3	-0.30
21	85.1	1.6	22	4.1	-0.22
22	88.0	2.0	19	18.5	-0.26
23	87.1	1.6	20	17.8	-0.28
24	83.9	1.2	20	18.4	-0.27
25	85.5	1.6	14	4.3	-0.23
26	85.1	0.4	25	6.7	-0.28
27	88.4	1.6	15	4.9	-0.23
28	86.7	0.4	22	18.5	-0.22
29	87.6	0.4	20	17.9	-0.25
30	85.1	0.8	21	12.2	-0.24
31	85.1	0.8	23	4.4	-0.22
32	85.1	0.8	22	18.4	-0.30
33	86.7	0.8	19	5.9	-0.23
34	85.5	1.6	22	18.6	-0.30
35	86.7	0.4	18	4.4	-0.19
36	84.7	0.8	23	18.7	-0.24
37	83.5	0.4	17	4.3	-0.30
38	83.5	1.2	21	18.7	-0.27
39	86.7	1.2	15	18.5	-0.22
40	84.7	1.2	22	18.3	-0.32
41	88.0	0.0	24	5.8	-0.20
42	86.3	1.6	18	18.7	-0.27
43	85.5	0.4	22	6.7	-0.27
44	84.3	0.8	12	15.2	-0.25
45	86.7	0.4	27	4.0	-0.25
46	86.3	0.8	13	4.5	-0.25
47	86.7	2.0	23	12.2	-0.26
48	85.9	0.8	15	4.7	-0.24
49	83.9	1.6	22	18.7	-0.34
50	84.3	2.0	25	6.3	-0.27
51	84.7	0.4	26	19.0	-0.34
52	86.7	1.2	14	4.0	-0.28
53	87.6	0.4	23	4.1	-0.25
54	84.3	0.8	15	12.4	-0.27
55	86.7	0.8	15	18.6	-0.25

56	85.9	2.4	18	18.7	-0.20
57	82.3	1.2	25	18.3	-0.27
58	85.5	0.4	25	18.3	-0.28
59	86.7	1.6	22	4.5	-0.27
60	86.7	1.2	23	4.8	-0.23
61	85.9	1.6	17	18.7	-0.27
62	85.9	1.6	24	10.4	-0.28
63	87.1	1.2	13	5.3	-0.23
64	86.7	1.2	26	6.4	-0.25
65	85.9	0.8	19	18.3	-0.27
66	85.5	1.6	18	4.0	-0.29
67	86.3	0.8	20	4.3	-0.28
68	83.5	0.8	16	5.4	-0.25
69	87.6	1.6	17	4.2	-0.20
70	85.9	0.4	16	18.3	-0.30
71	85.1	1.6	22	4.1	-0.22
72	88.0	2.0	19	18.5	-0.26
73	87.1	1.6	20	17.8	-0.28
74	83.9	1.2	20	18.4	-0.27
75	85.5	1.6	14	4.3	-0.23
76	85.1	0.4	25	6.7	-0.28
77	88.4	1.6	15	4.9	-0.23
78	86.7	0.4	22	18.5	-0.22
79	87.6	0.4	20	17.9	-0.25
80	85.1	0.8	21	12.2	-0.24
81	85.1	0.8	23	4.4	-0.22
82	85.1	0.8	22	18.4	-0.30
83	86.7	0.8	19	5.9	-0.23
84	85.5	1.6	22	18.6	-0.30
85	86.7	0.4	18	4.4	-0.19
86	84.7	0.8	23	18.7	-0.24
87	83.5	0.4	17	4.3	-0.30
88	83.5	1.2	21	18.7	-0.27
89	86.7	1.2	15	18.5	-0.22
90	84.7	1.2	22	18.3	-0.32
91	88.0	0.0	24	5.8	-0.20
92	86.3	1.6	18	18.7	-0.27
93	85.5	0.4	22	6.7	-0.27
94	84.3	0.8	12	15.2	-0.25
95	86.7	0.4	27	4.0	-0.25
96	86.3	0.8	13	4.5	-0.25
97	86.7	2.0	23	12.2	-0.26
98	85.9	0.8	15	4.7	-0.24
99	83.9	1.6	22	18.7	-0.34
100	84.3	2.0	25	6.3	-0.27
101	88.0	0.8	15	18.1	-0.24
102	87.1	0.0	18	18.6	-0.26
103	88.0	0.4	19	6.6	-0.27
104	87.1	0.4	10	4.5	-0.22
105	87.6	1.2	13	4.8	-0.19
106	88.0	0.8	12	4.8	-0.25
107	88.0	0.8	16	5.4	-0.26
108	89.6	1.2	14	4.5	-0.23
109	88.8	1.2	18	5.1	-0.22
110	85.5	0.0	20	5.0	-0.30

The top best values have been highlighted red in each column for each model.

Table 3.17: Models ranked as first, second, third and fourth on the basis of best values for core region, disallowed region, bad contacts, max deviation and the G-factor (overall) value as noted from the PROCHECK results

Rank	Models numbers				
	Core	**Disallowed**	**Bad contacts**	**Max dev**	**G-factor: overall**
1	108	41,91,102,110	104	2,16,45,52,66,95	35,85,105
2	109	1,3,8,20,26,28,29,35,37,43,45,51,53,58,70,76,78,79,85,87,93,95,103,104	44,94,106	3,21,53,71	6,19,41,56,69,91
3	27, 77	4,5,15,17,18,30-33,36,44,46,48,54,55,65,67,68,80-83,86,94,96,98,101,106,107	13,46,63,96,105	19,69	21,28,31,39,71,78,81,89,104,109
4	22,41,72,91,101,103,106,107	2,7,10,13,14,24,38-40,52,57,60,63,64,74,88-90,105,108,109	2,25,52,75,108	17,25,37,67,75,87	10,13,25,27,33,60,63,75,77,83,108

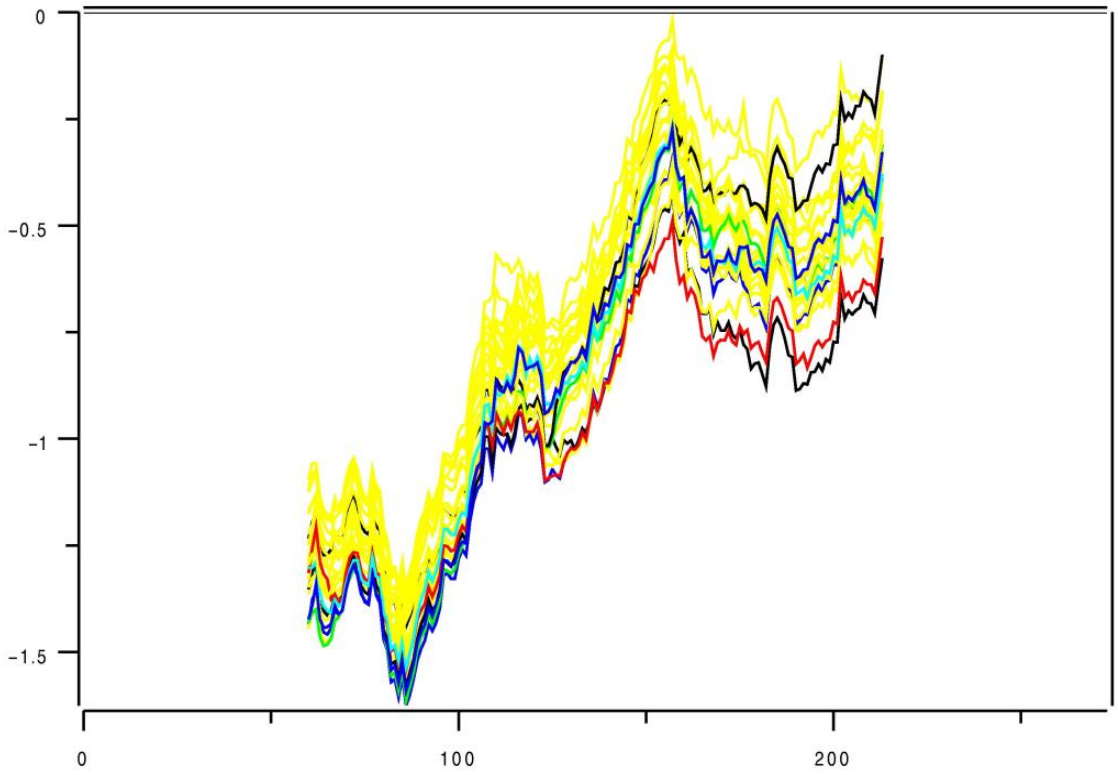


Fig. 3.15a: ProSA energy graphs for the models selected as best models on the basis of PROCHECK results

The models having lowest energy values have been colored 41=red, 107=blue, 35=85=cyan, 104=blue, 19=green, 3=black, 39=magenta, and the models with higher energy values have all been colored yellow. The winsize value was 120.

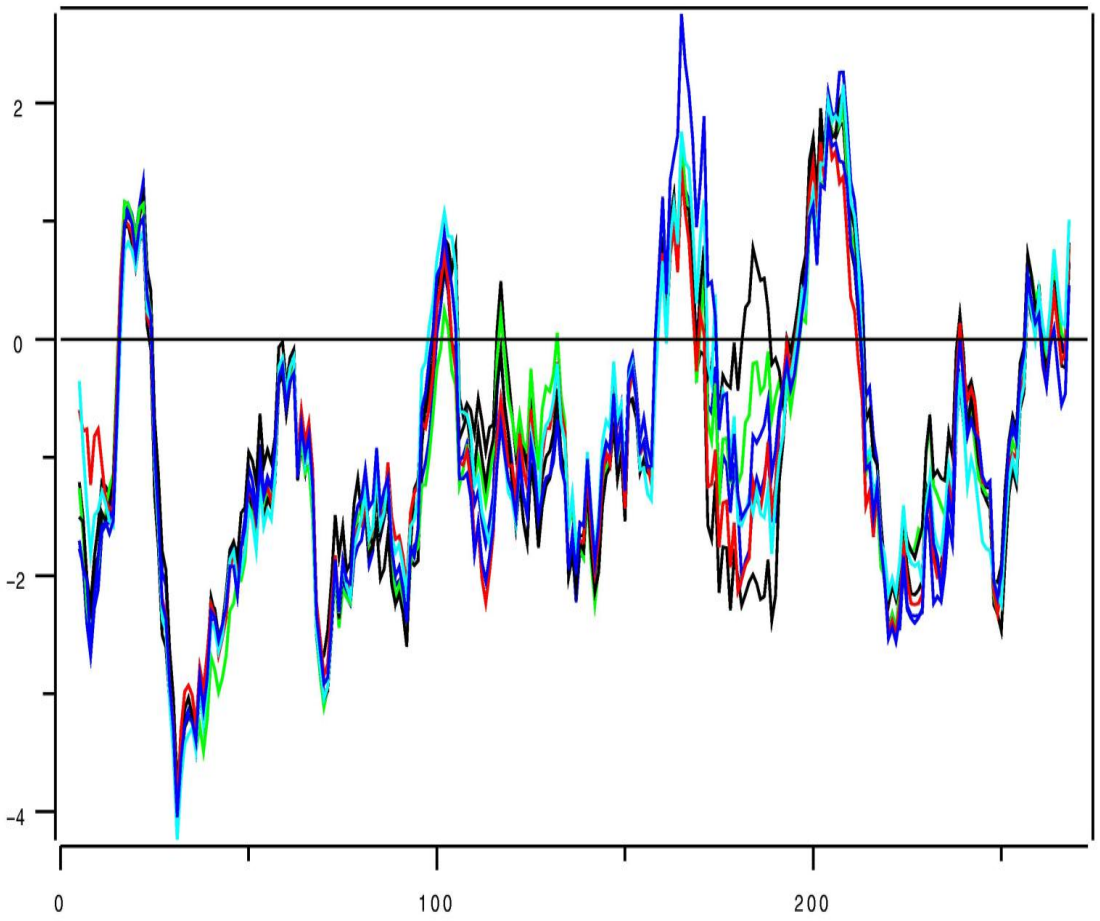


Fig. 3.15b: ProSA energy graphs for the models selected as best models on the basis of PROCHECK as well as ProSA results

The models have been colored as 41=red, 107=blue, 35=85=cyan, 104=blue, 19=green, 3=black, and 39=magenta. The winsize was selected as 10 to view minor differences in the energy values. There is no big difference among the energy values of these selected models therefore any of these models can be considered as a good model. However with a careful observation it is evident from the graph that model number 41 has lower overall energy value as compared to other models.

First, it has low energy value in the active site region. Second, it has the highest number of residues in the core region and there is no residue in the disallowed region. On the basis of these results model no. 41 was selected as the best model. This model has the lowest energy value in the conserved region of the sequence. The energy of the selected model was compared with all the three templates [Fig. 3.15c]. The energy of the model is higher at residues 15-25, 50-60, 80-88, 95-105, 160-170, 200-210 and 240-last as compared to the templates, however it is the lowest at residues 26-40, 66-75, 108-112, 182-193 and 215-235. The overall energy is comparable with the templates especially with 2fk8. The overall energy is in the range of a good enough for a good quality model.

3.6.6. Z-Score Values and Their Comparison

The Z-score of a protein is defined as the energy separation between the native fold and the average of an ensemble of misfolds in the units of the standard deviation of the ensemble [342, 343]. Z-scores outside a range characteristic for native proteins indicate erroneous structures. However, it is not known what range of values of Z-scores should be expected for real proteins with exact potentials. The Z-scores vary tremendously from protein to protein [344]. The Z-score of the selected model #41 was -6.16 [Fig. 3.16a] and is within the acceptable region of the graph. The Z-scores for the templates 1kph [Fig. 3.16b], 1kpi [Fig. 3.16c] and 2fk8 [Fig. 3.16d] were -9.01, -8.72 and -9.86 respectively. These values lie near the model's Z-score value. The Z-score values for templates are a bit better than the model's Z-score. This is because templates are experimentally determined structures.

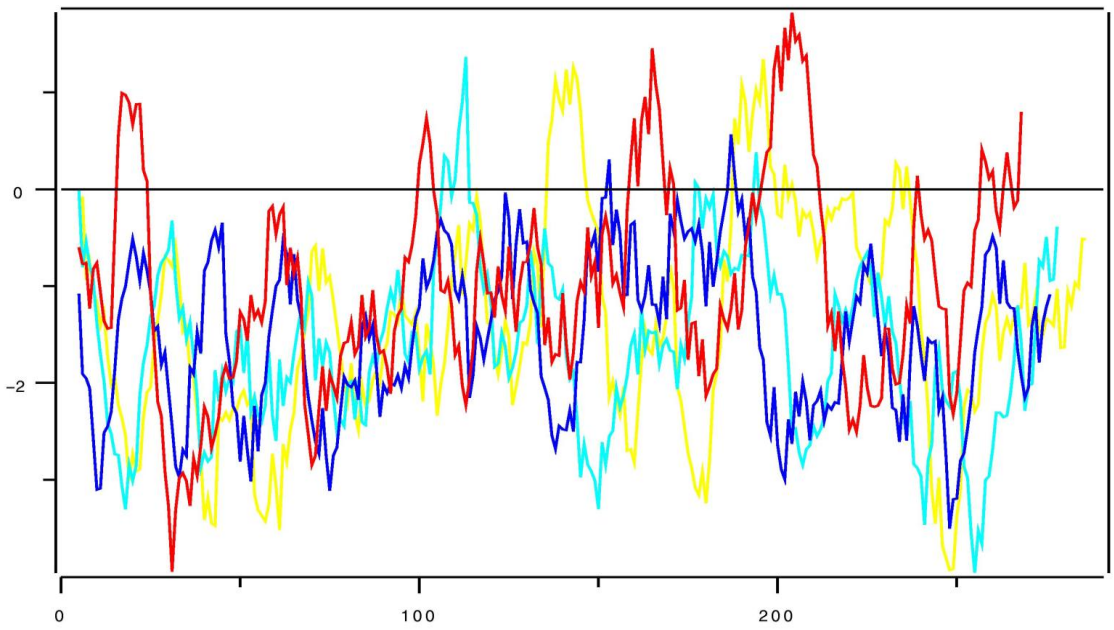


Fig. 3.15c: Comparison of energy graphs of the selected model number 41=red, with three selected templates 1kpi=yellow, 1kph=cyan, 2fk8=blue for modeling: The winsize was selected as 10.

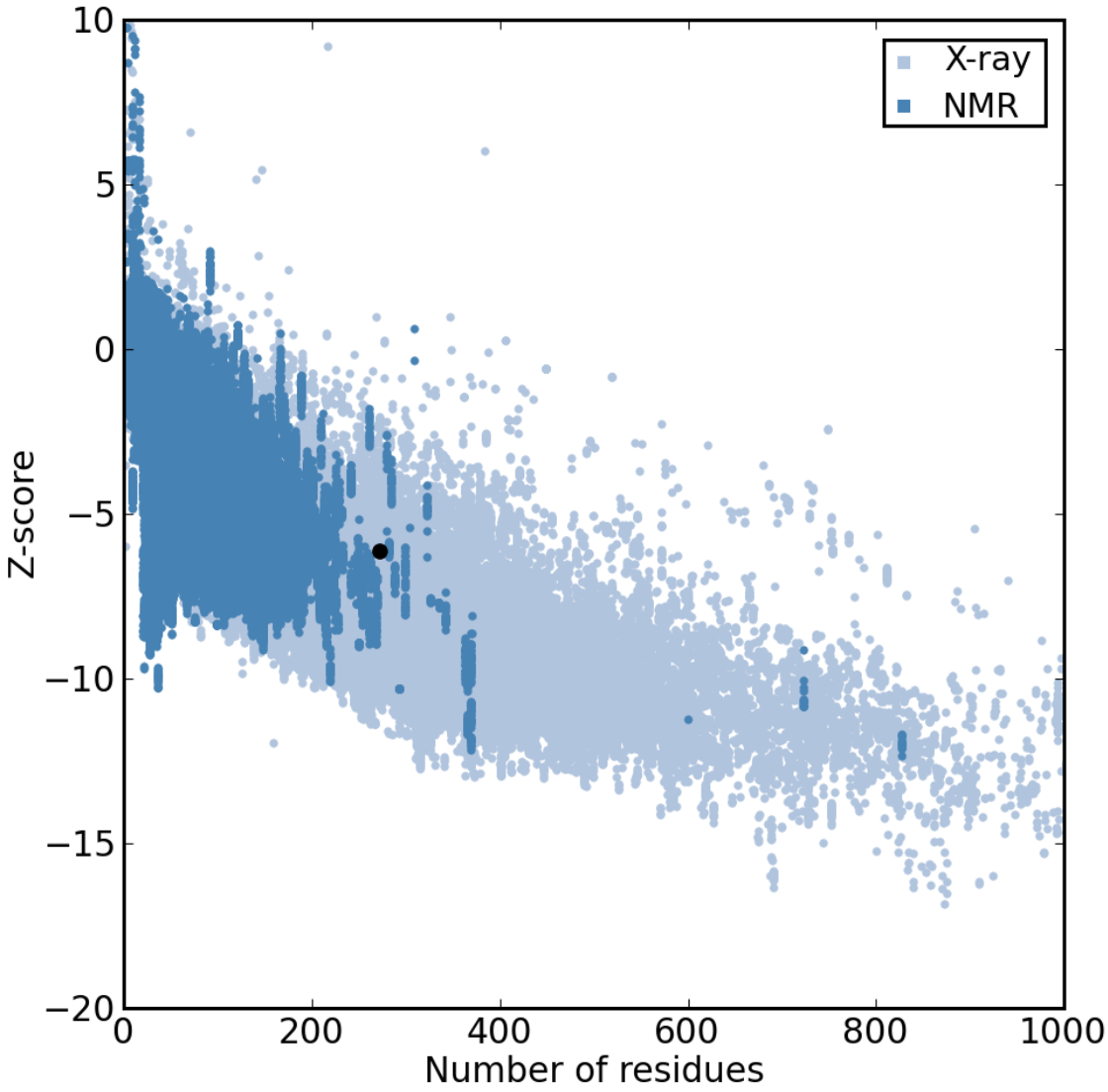


Fig. 3.16a: Graph representing the Z-score value for selected model #41

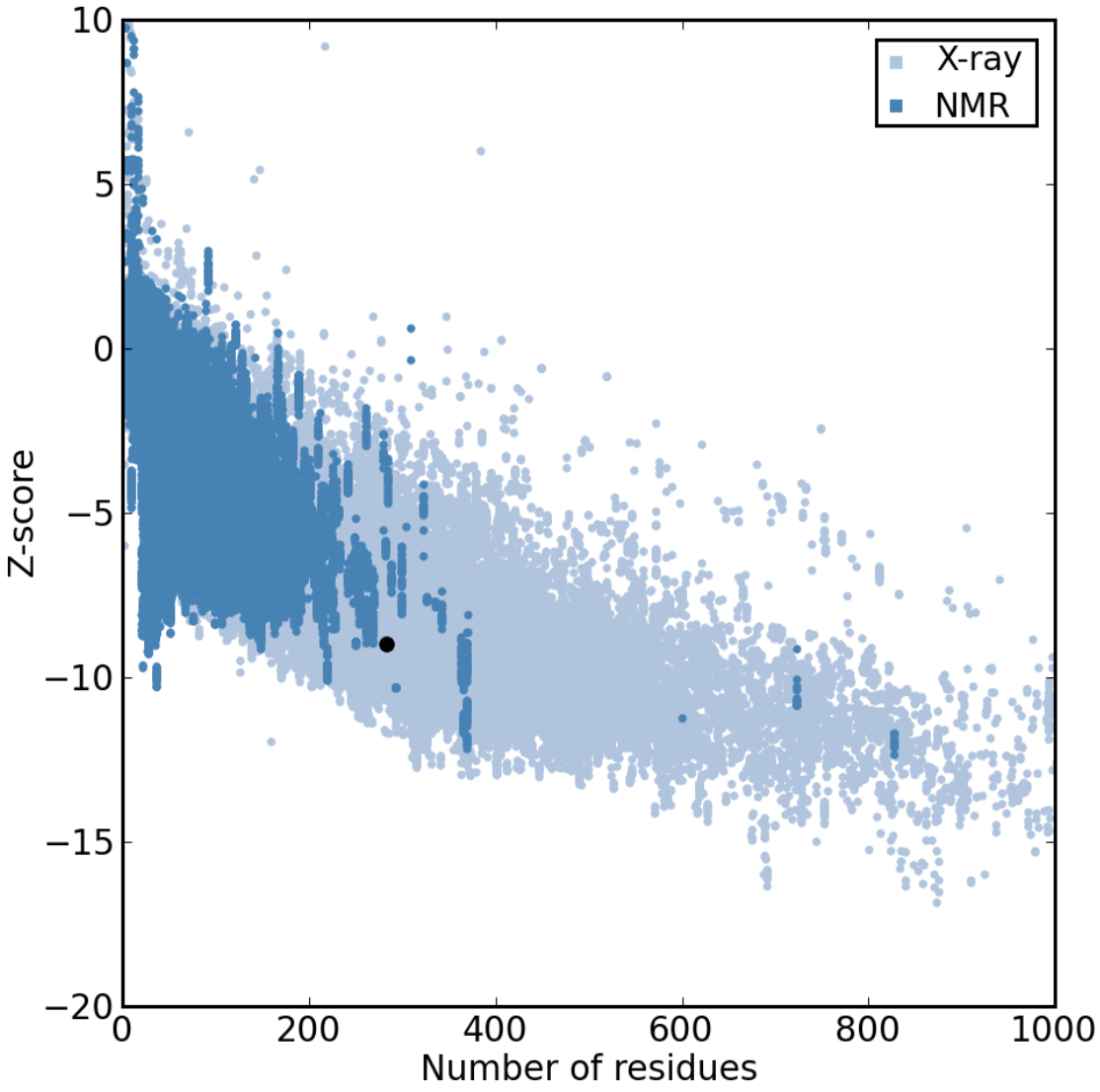


Fig. 3.16b: Graph representing the Z-score value for the template 1kph

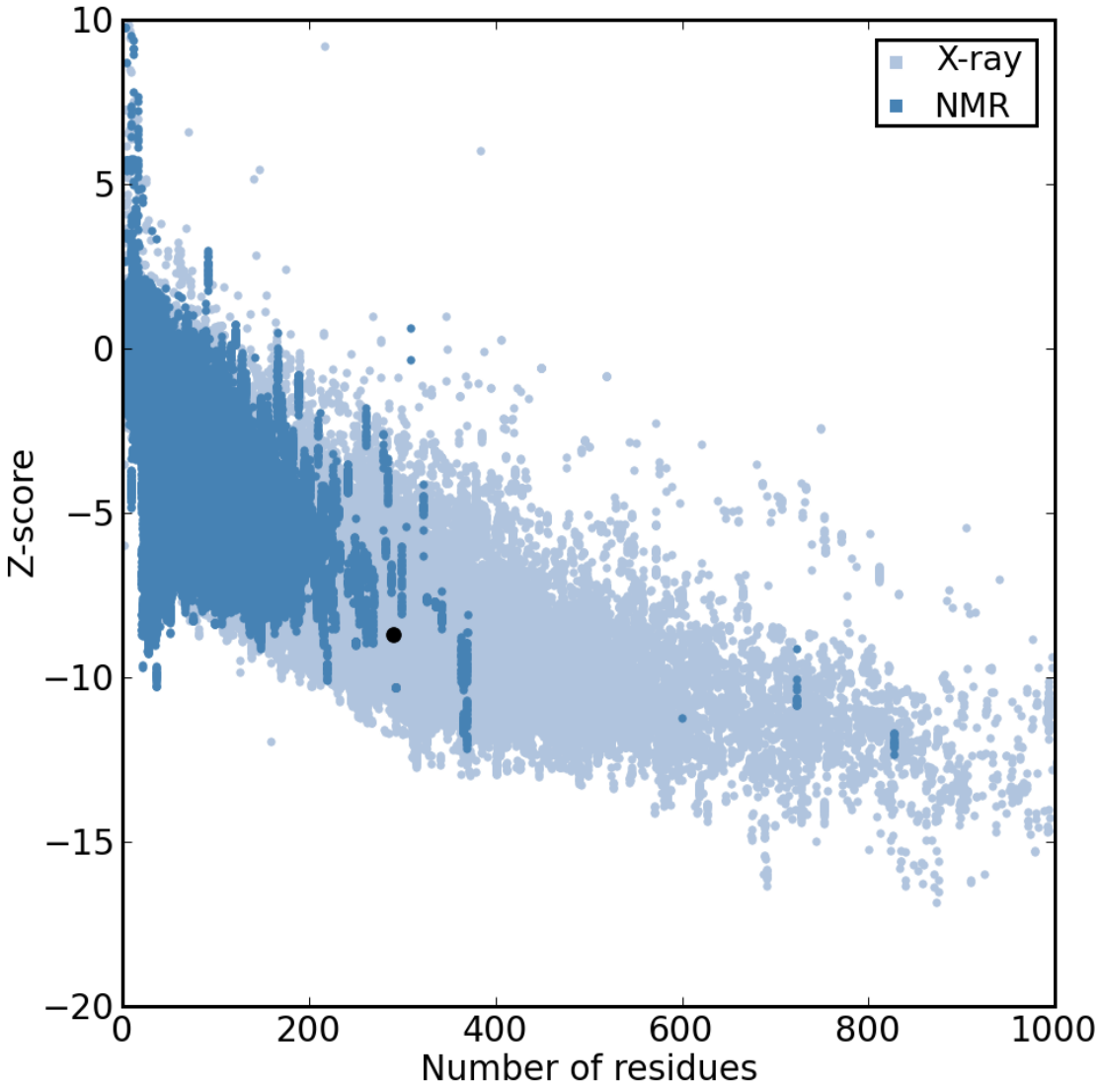


Fig. 3.16c: Graph representing the Z-score value for the second template 1kpi

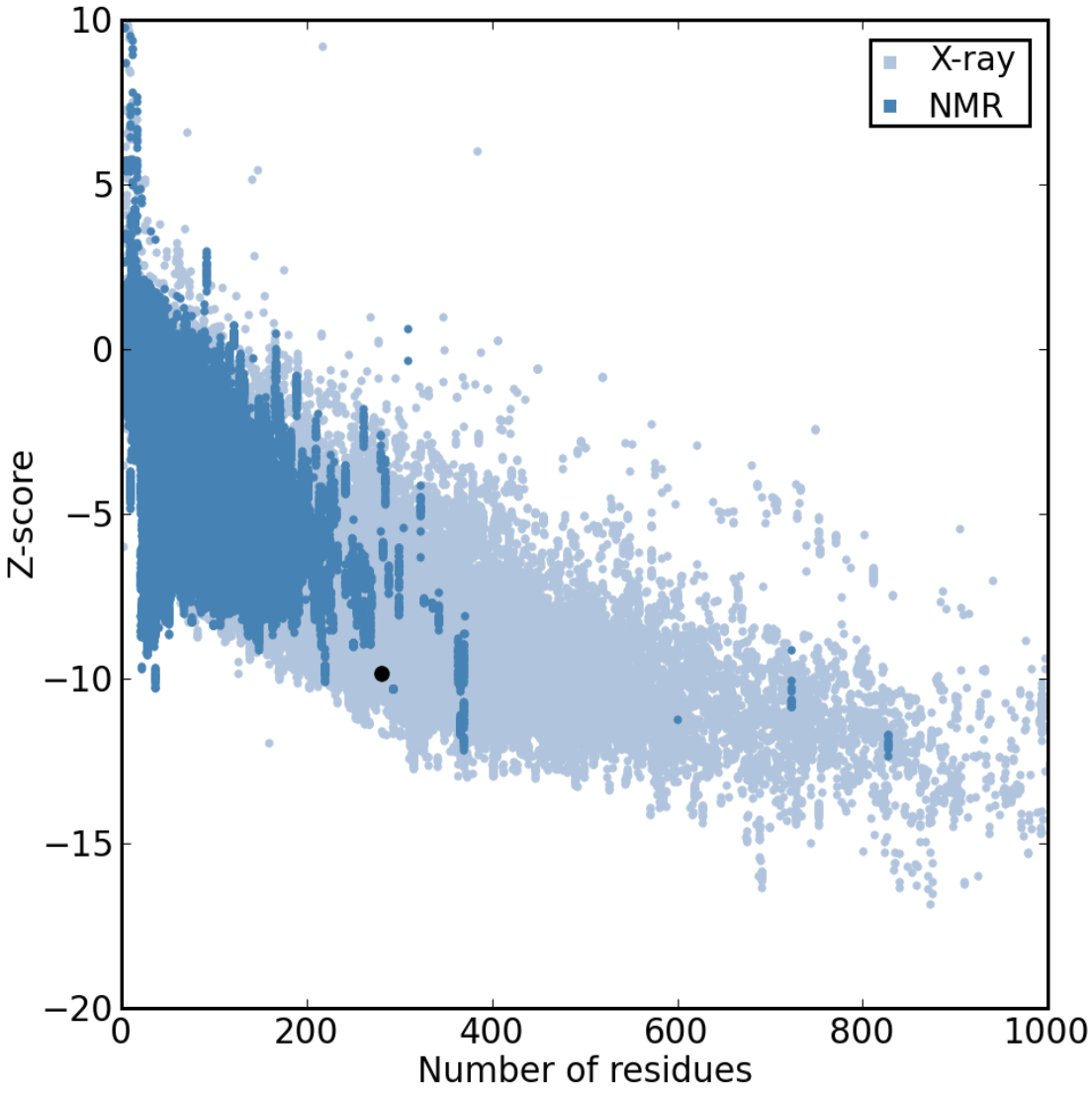


Fig. 3.16d: Graph representing the Z-score value for the third template 2fk8

Another reason is that there is less than 30 % sequence identity between the model and templates sequences. Still the value for the model is within the range of a good quality model.

3.6.7. Ramachandran Plot

Ramachandran plot [Fig. 3.17a] shows that model no. 41 has 88.0 % residues in the most favoured region, 9.6 % residues in additional allowed region, 2.4 % residues in generously allowed region, and 0.0 % residues in disallowed region. Models having more than 80 % residues in the most favoured region are considered as quite satisfactory [345].

For comparison, an analogous analysis was performed on the experimental templates [345]. Ramachandran plots for the templates 1kpi [Fig. 3.17b], 1kph [Fig. 3.17c] and 2fk8 [Fig. 3.17d] shows that they have 89.5 %, 93.2 %, 94.4 % residues in the most favoured region, 10.1 %, 5.2 % and 5.2 % residues in additional allowed region, 0.4 %, 0.8 % and 0.4 % residues in generously allowed region, and 0.0 %, 0.8 % and 0.0 % residues in disallowed region respectively. These values are either identical or very close to that of model's values. The comparison of the model's values with the templates' values shows that the model has been made accurately and can be considered to be a good quality model. Also important is that model #41 does not have any residues in the disallowed region. For further evaluation the model was superimposed over all the three templates separately.

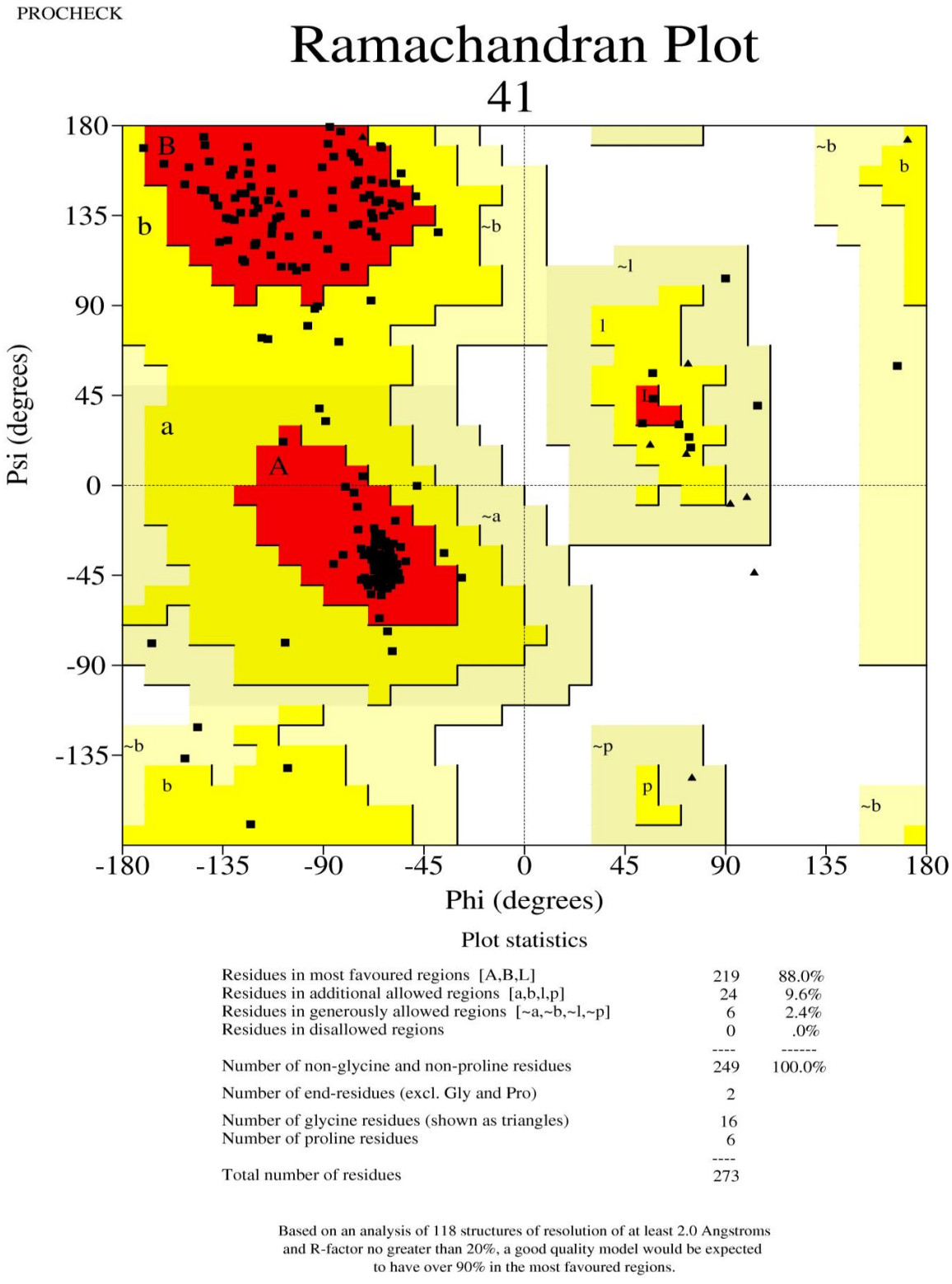


Fig. 3.17a: Ramachandran plot for model number 41

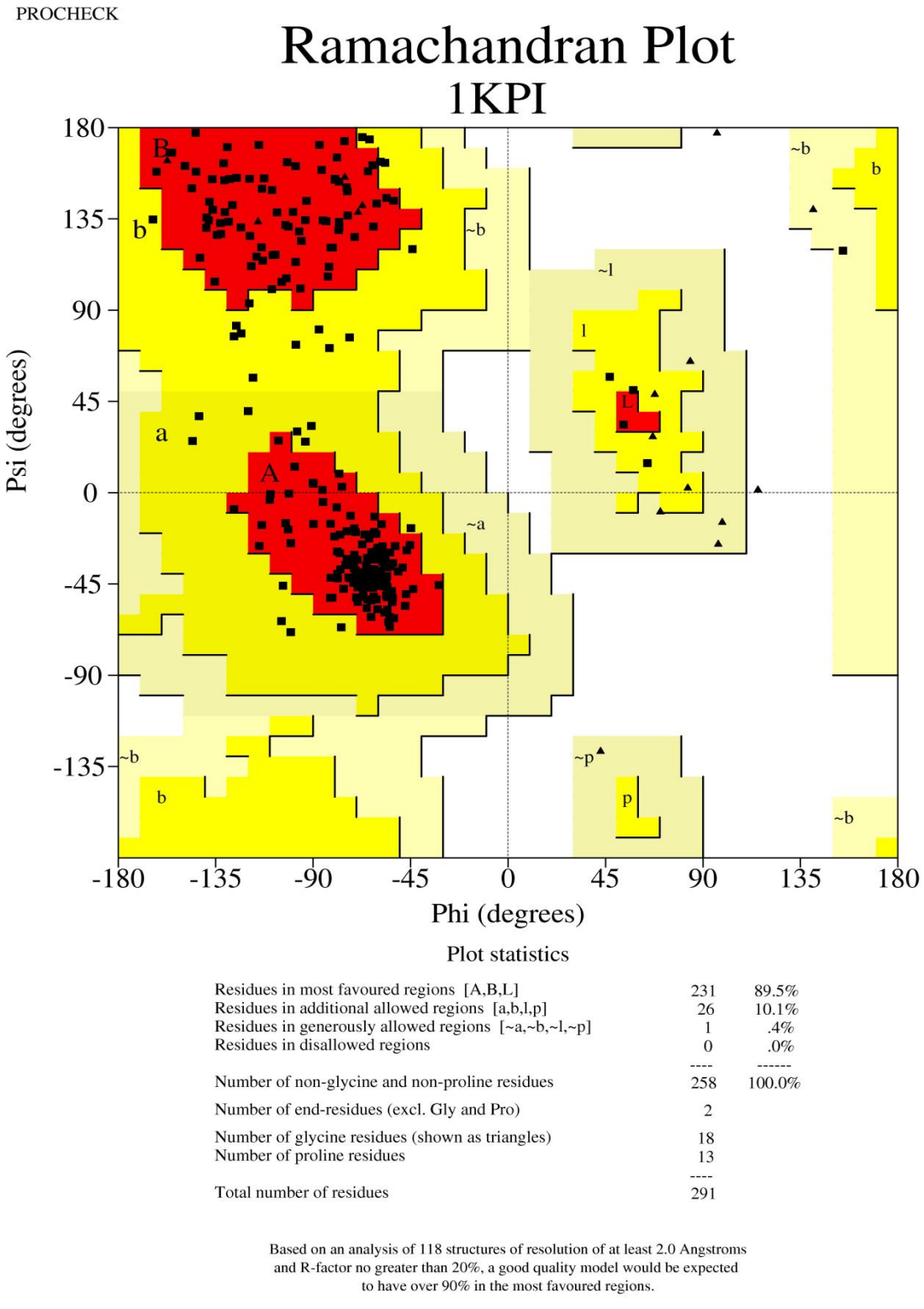


Fig. 3.17b: Ramachandran plot for the first template 1kpi

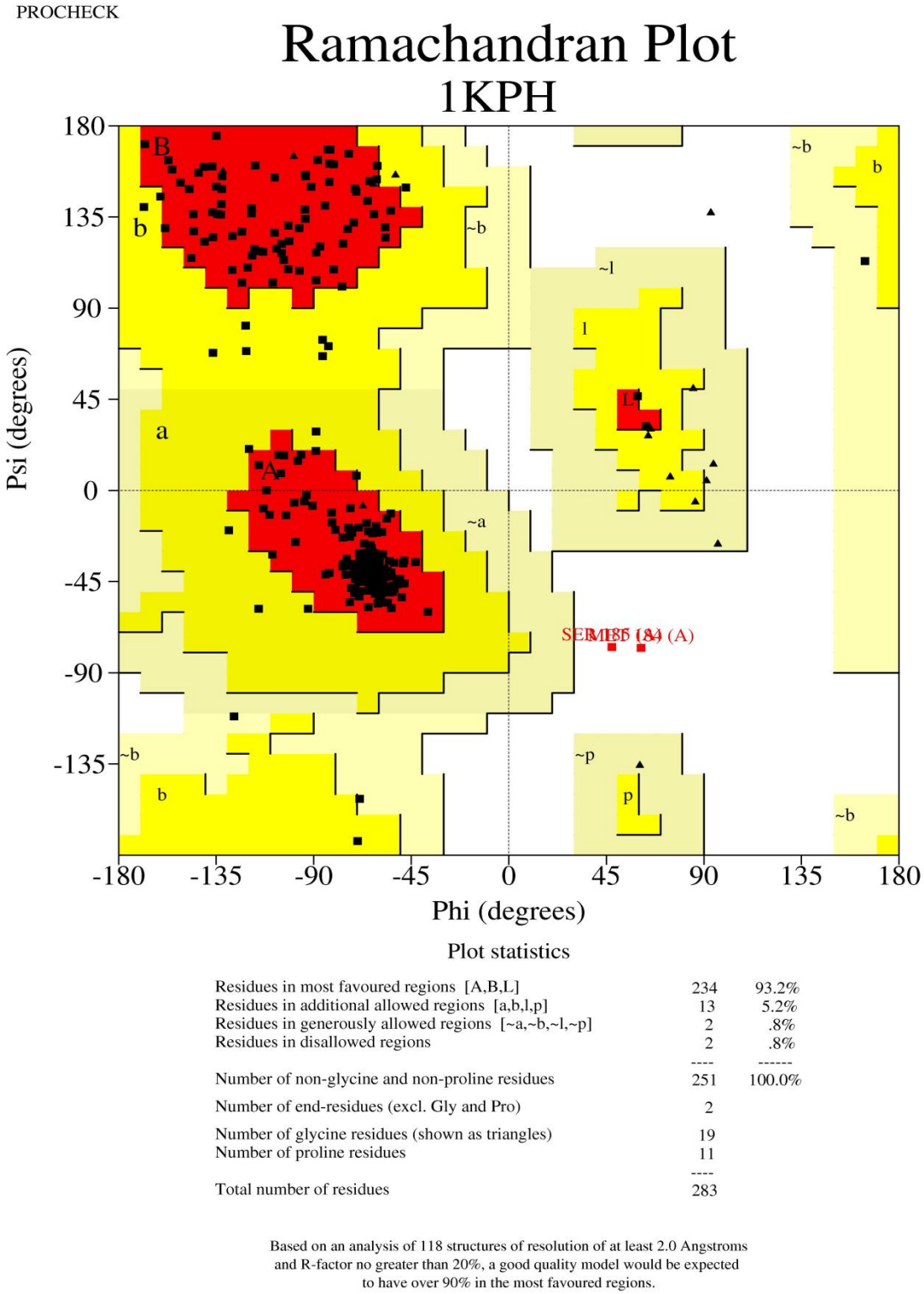
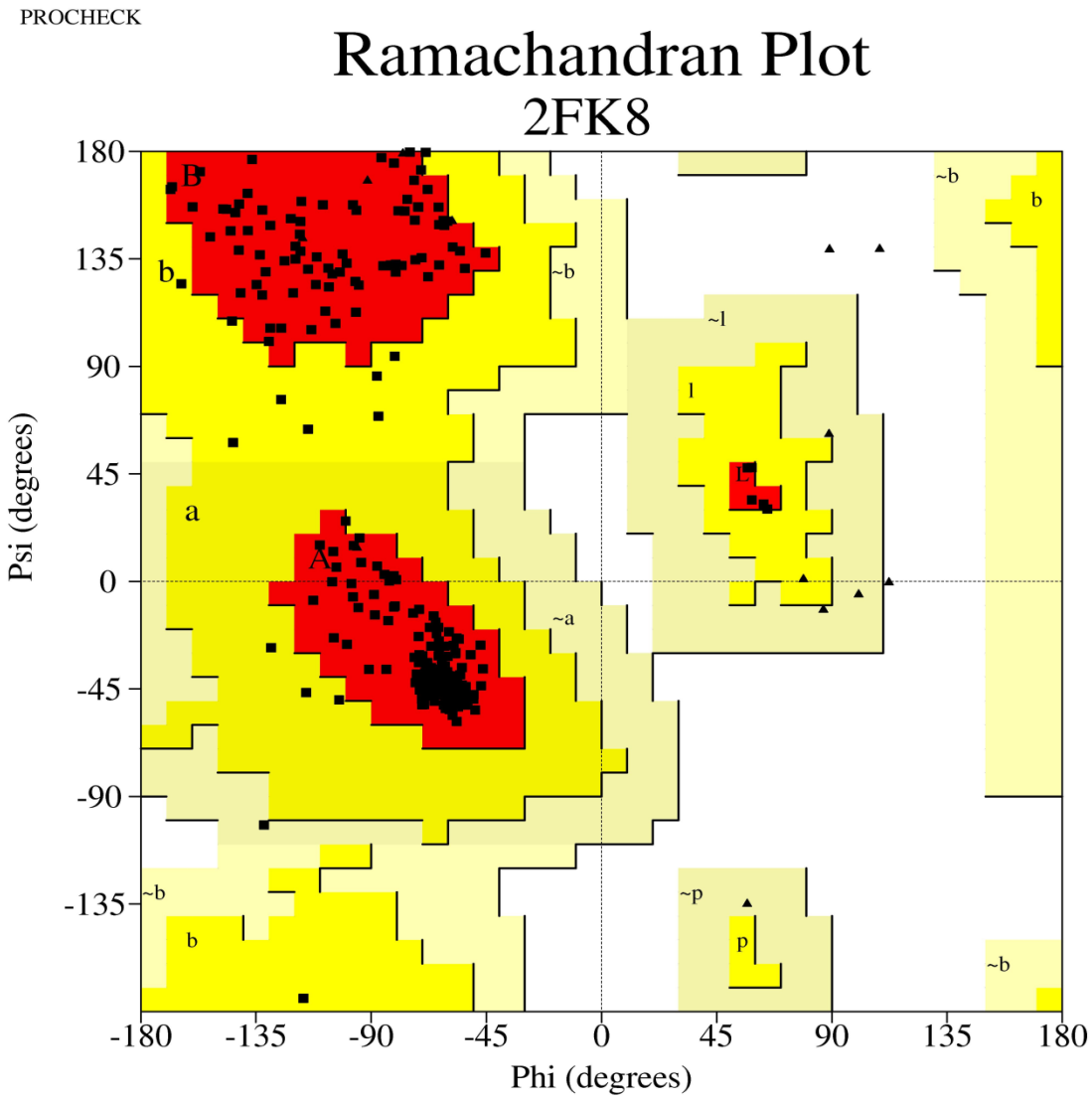


Fig. 3.17c: Ramachandran plot for the second template 1kph



Plot statistics

Residues in most favoured regions [A,B,L]	236	94.4%
Residues in additional allowed regions [a,b,l,p]	13	5.2%
Residues in generously allowed regions [~a,~b,~l,~p]	1	.4%
Residues in disallowed regions	0	.0%
Number of non-glycine and non-proline residues	250	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	16	
Number of proline residues	13	
Total number of residues	281	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Fig. 3.17d: Ramachandran plot for the template 2fk8

3.6.8. Superimposition

The Cα backbones of all the three templates were individually superimposed over that of the homology model of the putative CNMT. The superimposition was according to alignment file created for homology modeling. Only those parts of the templates and CNMT (the conserved parts) were superimposed over each other and aligned in the alignment file. Only 131 residues of the model were superimposed to the templates 1kph, 1kpi and 2fk8 with root-mean-square-deviation (rmsd) of 2.71, 2.90 and 2.94 respectively.

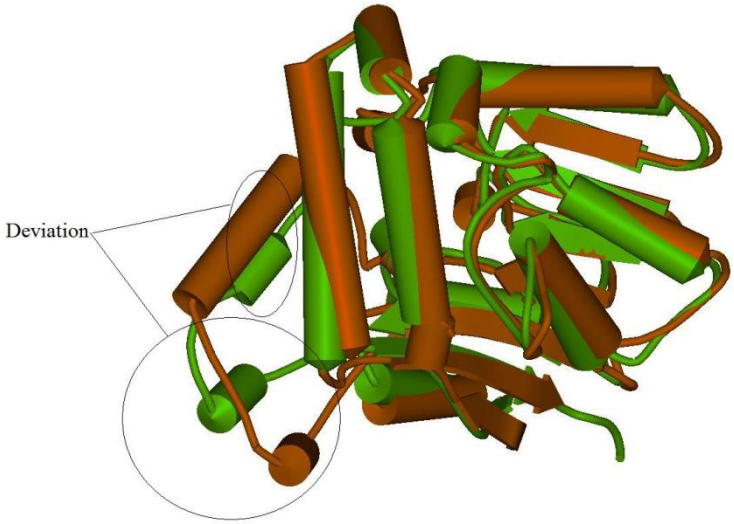
Two proteins, even with a high level of sequence similarity and very similar secondary and tertiary structure, cannot have precisely matching backbone conformations, even when determined under analogous conditions. A homology model can be estimated to differ from the real crystal structure to at least this degree. Overall differences in protein Cα backbone structures can be quantitated with the root mean square deviation of the positions of alpha carbons, or rmsd. A model can be considered 'accurate enough' when its rmsd is within the range of deviations observed for experimental structures displaying a similar sequence identity level as the target and template sequences [322].

To provide a frame of reference for rmsd values, up to 0.5 Å rmsd of Cα carbons occurs in independent determinations of the same protein [346]. On average, 1 Å rmsd have been observed for proteins with 50% sequence identity [322]. These values are for X-ray crystallographic determinations; NMR determinations have higher rmsd's.

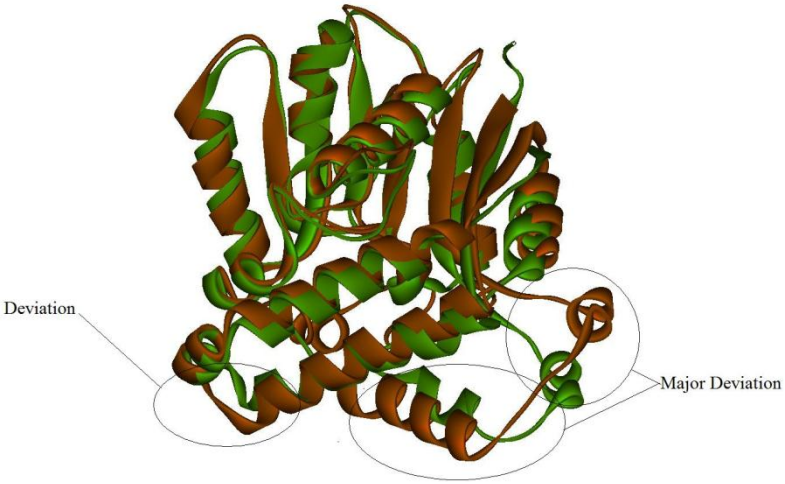
If a successful model is defined as one having <=2 Å rmsd from the experimental structure, then the template must have >=60% identity with the target sequence for a success rate >70%. Even at high sequence identities (60%-95%), as many as one in ten models have an rmsd >5 Å vs. the experimental structure. With sequence identity below 40%, serious errors can appear (Martz, E. 2001; online article: http://www.umass.edu/molvis/workshop/ homol mod.htm).

As our sequence showed only 30 % or less identity with the templates, serious errors were expected in our model, however the rmsd values show that it is fairly accurate and is identical to a greater extent to the templates. Our results show that the conserved active part of the model is more identical to 1kph as compared to other two templates. In fact, the overall structure might be greatly identical to 1kph crystal structure.

The overall close similarity of predicted model for our protein and related known proteins, even when sequence identities are only 30 %, suggests that many of the amino acid substitutions separating these proteins have small impact on the backbone configuration. The main deviation of the model from the template 1kpi was at three regions [Fig. 3.18a]. The first region in the template was 185-200, 152-160 and 236-244, while the corresponding segments on the model were 233-248, 207-210 and 279-287. The first one was an alpha helix in both the template and model.



A



B

Fig. 3.18a: A: Schematic and B: Solid ribbon representation of the superimposition of the model #41 colored in green over the template 1kpi colored as brown with rmsd =2.90: The regions deviated have been encircled. The model has been shown from two different angles.

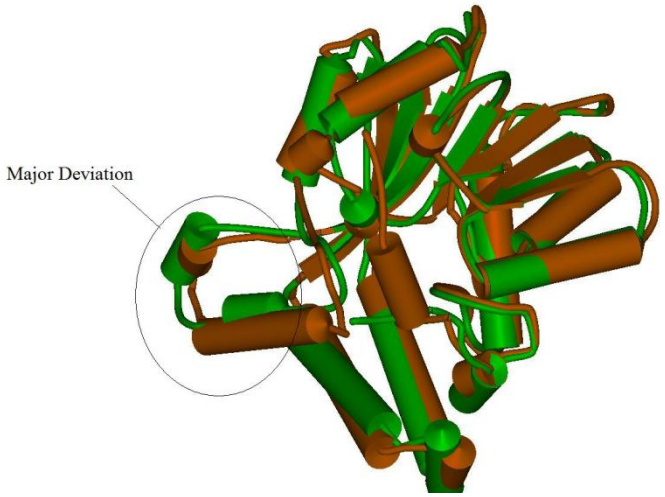
The second segment is a loop region. The loop in this region is longer in the model than that of the template. This shows a gap in the model sequence in the alignment. The @TOME has not aligned this segment of the predicted CNMT with the template. The third segment is a beta sheet in the template, while it is a coil in the model. All these three regions are not in the active site.

The main deviation of the model from the template 1kph was at only one region i.e. 173-206 of the template with 236-265 of the model [Fig. 3.18b]. This region is divided into four parts two helices and two coils in the sequence helix-coil-helix-coil. Here also the region is not in the active site and is not the part where the sequences were aligned.

The deviations from 2fk8 were mainly in the loop regions [Fig. 3.18c and 3.18d]. The whole model superimposed well over the template. This was because the alignment made was to a longer region as compared to the other two templates, although its rmsd was higher than the other two templates.

The ligand is nearly at the same position and in the same orientation [Fig. 3.18e] as it is in the template 2fk8. There is a very minute difference in the ligand's position in the model; however for such a low identity this difference is expected.

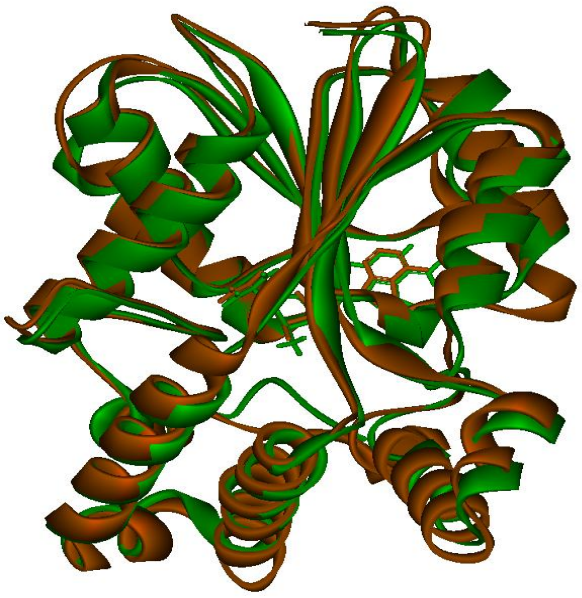
Except for the deviated parts of the model the overall model superimposed well over all the templates. This supports the correctness of the model.



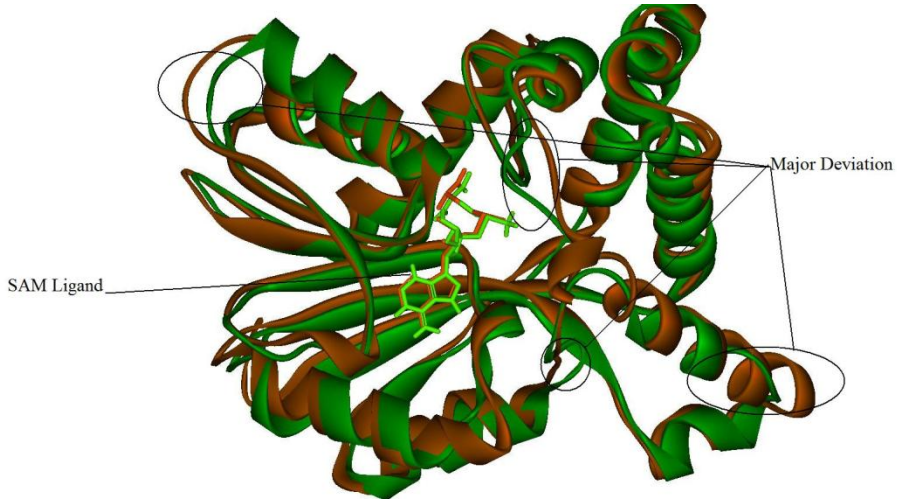
A

B

Fig. 3.18b: A: Schematic and B: Solid ribbon representation of the superimposition of the model #41 colored in green over the template 1Kph colored as brown with rmsd =2.71: The regions deviated have been encircled. The model has been shown from two different angles.



A



B

Fig. 3.18c: Solid ribbon representation of the superimposition of the model colored in green over the template 2fk8 colored as brown with rmsd =2.94: This model also includes the SAM residues of both the templates and the model. The model has been shown from two different angles as shown in figure A and B.

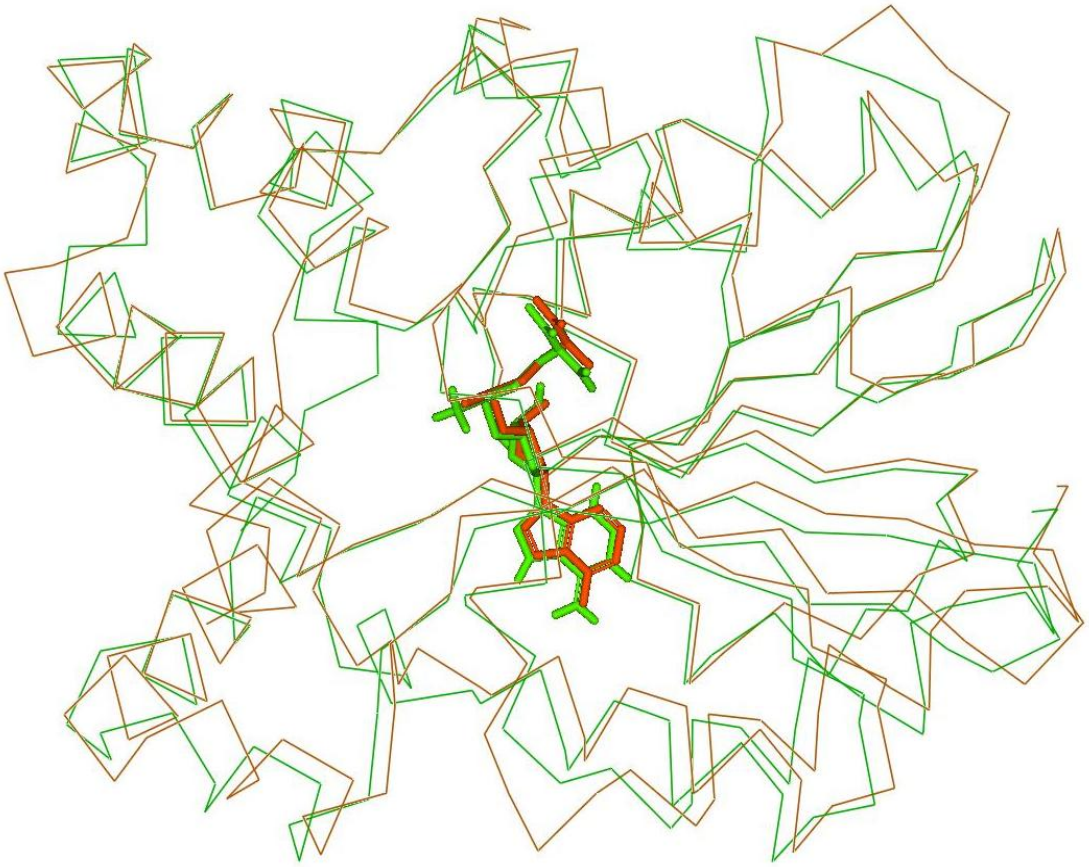


Fig. 3.18d: Carbon alpha wire representation of the superimposition of the model colored in green over the template 2fk8 colored as brown with rmsd =2.94: This model also includes the SAM residues of both the templates and the model.

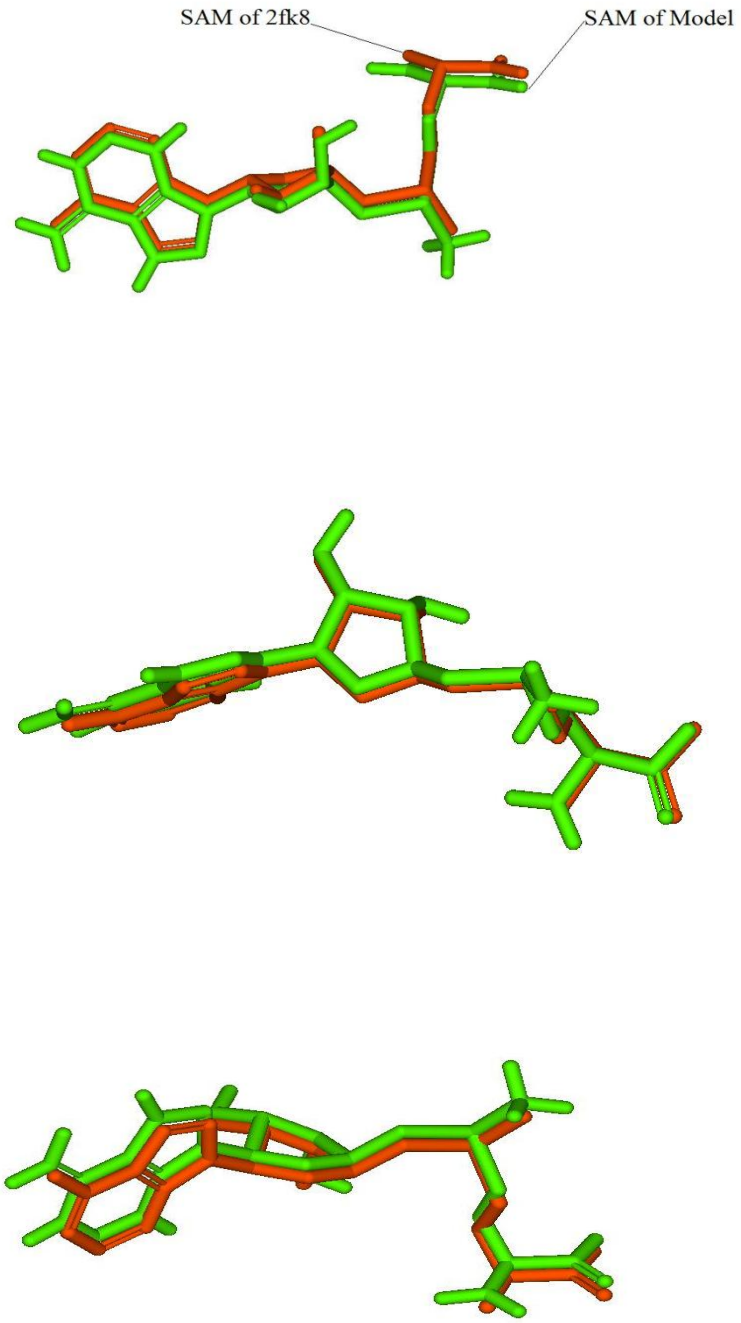


Fig. 3.18e: Stick representation of the superimposition of ligands (SAM) of the template colored as brown over that of model colored as green: The view has been shown from three different angles. The model's SAM has been shown with hydrogen atoms attached.

3.7. 3D Structure of the Model

The structure of the model can be divided into two main domains: the N-terminal catalytic core domain and the C-terminal domain. It has been shown that SAM-MTases show a similar folding pattern with central parallel b-strands surrounded by alpha-helices [347-351]. All SAM-MTases possess a conserved SAM-binding domain consisting of a seven-stranded β-sheet flanked by three α-helices per side of the sheet. Depending on the size of the substrate, a substrate-recognition domain may be appended to the SAM-binding domain [43]. However, this domain is highly variable, which is consistent with the fact that these enzymes methylate different kinds of substrates.

The same pattern is observed in the putative CNMT model that the catalytic domain of the putative CNMT comprises a central beta sheet of twisted β-strands surrounded by alpha helices. The topology of the model is a mixed α/β structure, almost identical to the templates used for building the model. The model mainly composed of seven beta-strands (β1-β7; strand order is β 3-2-1-4-5-7-6) and twelve α-helices [Fig. 3.19a]. The β-strands are surrounded from three sides by α-helices. The first five β-strands (3-2-1-4-5) and the sixth beta strand are all parallel to each other, while the last beta strand (in between beta 5 and 6) is antiparallel to all the other beta strands [Fig. 3.19b]. The catalytic core domain contains the binding sites for SAM. The cavity for SAM and substrate [Fig. 3.19c] is surrounded by three α-helices from one side and the C-terminal residues of β1, β4 and β5 from the other side. The β-strands are perpendicular to the SAM and substrate cavity. The amino acid residues making these α-helices and β-sheets are α1: 83-89, α2: 108-123, β1: 130-135, α3: 140-150, β2: 153-159, α4: 161-173, β3: 179-183, β4: 197-201, α5: 204-207, α6: 210-220, β5: 227-233, α7: 237-242, α8: 248-252, α9: 256-275, β6: 279-281, α10: 289-302, α11: 304-311, α12: 313-330 and β7: 335-342.

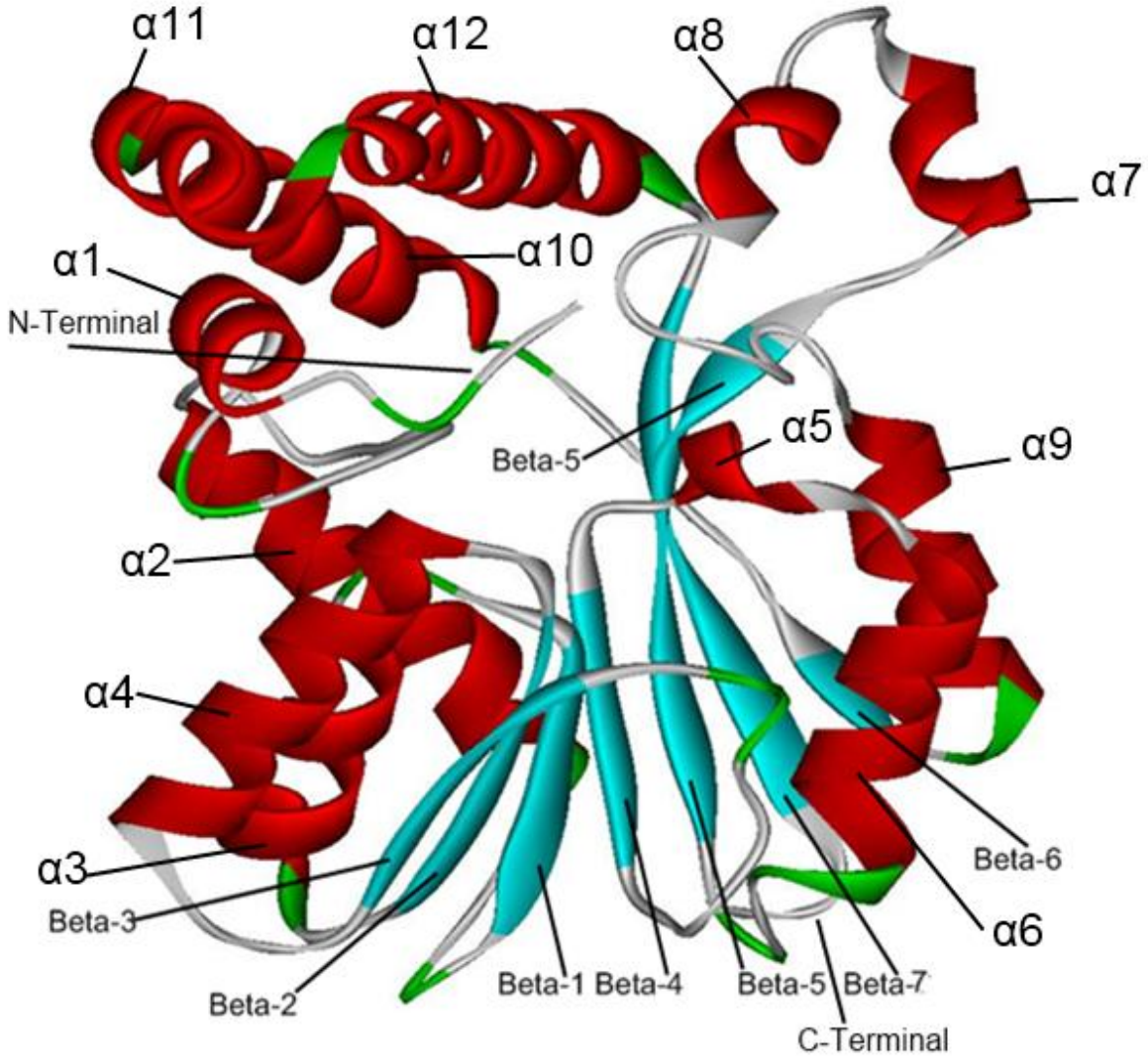


Fig. 3.19a: Solid ribbon representation of the putative CNMT model of *A. fimbriata*: The model consists of 7 beta strands (blue) surrounded by 12 alpha helices (red).

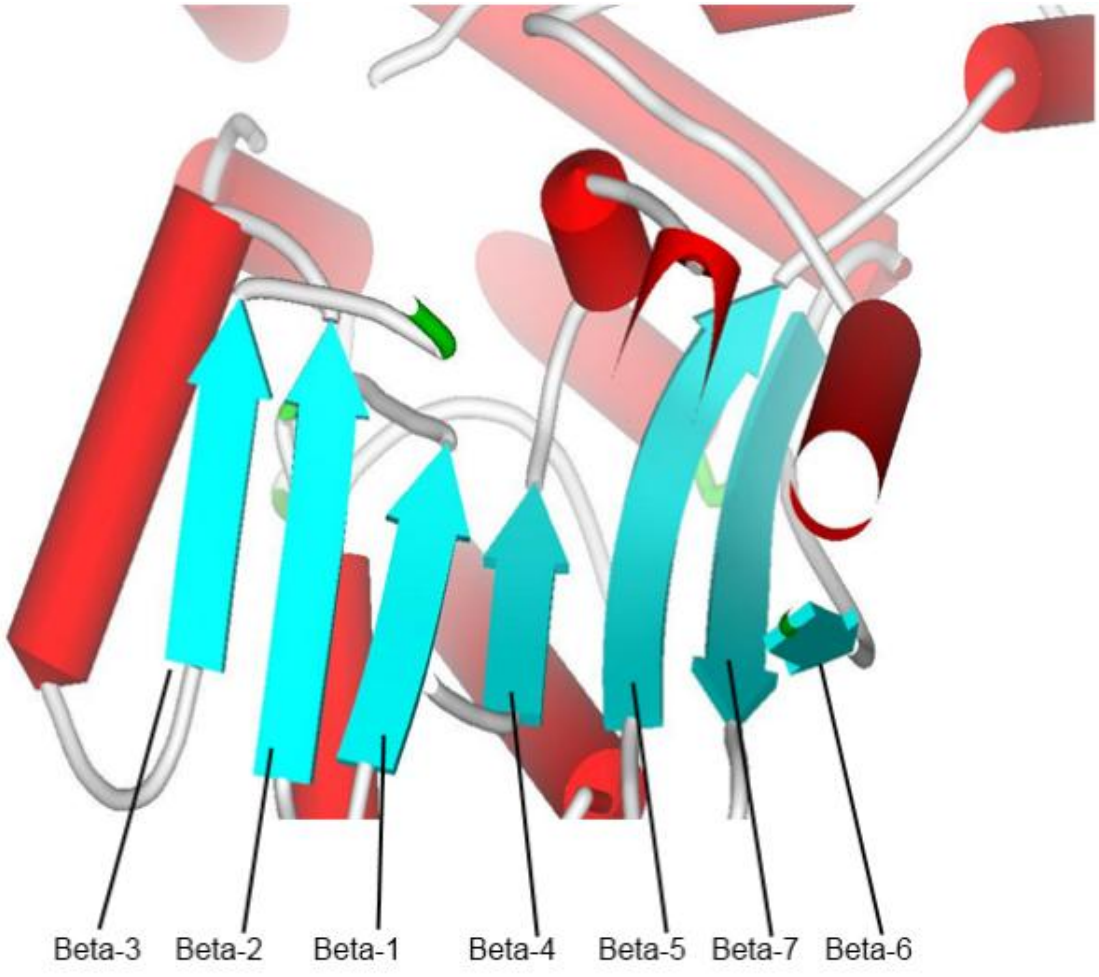


Fig. 3.19b: Schematic representation of a part of the putative CNMT model of *A. fimbriata* showing the arrangement of beta-strands: The alpha helices in front of the beta sheet have been hidden to make the beta strands visible. The first five beta strands are parallel and the last two are antiparrallel.

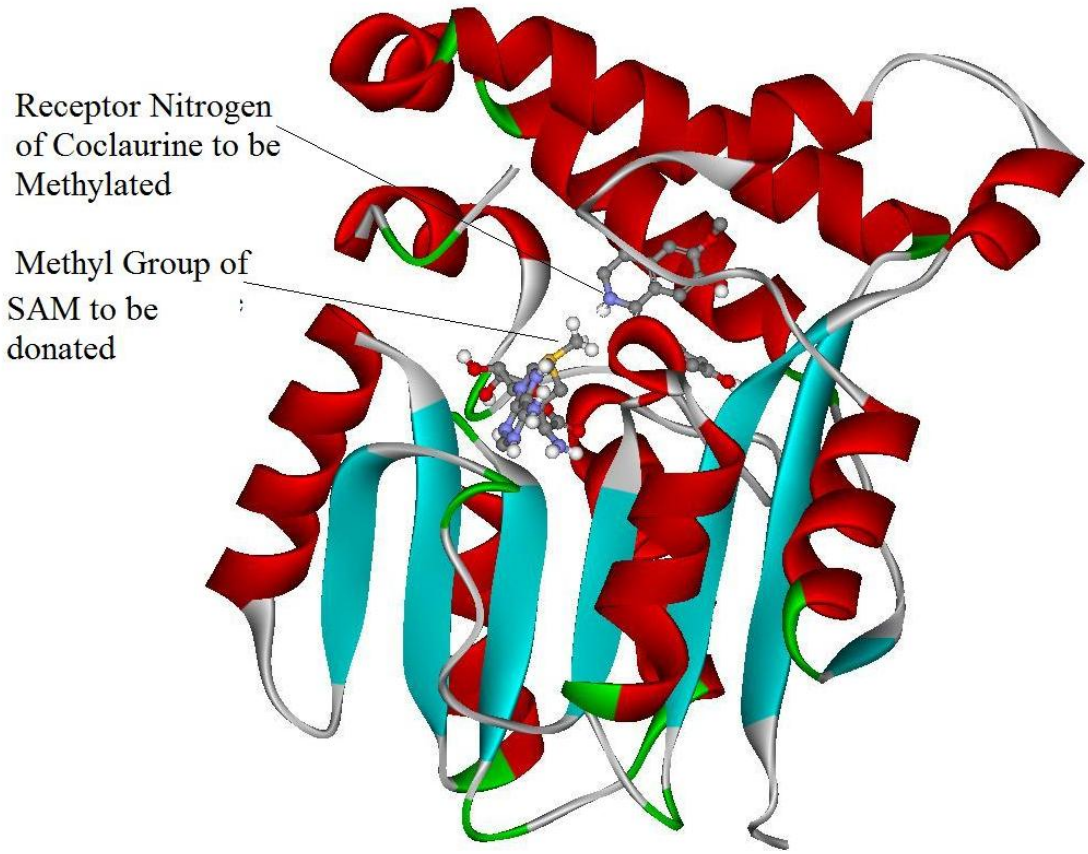


Fig. 3.19c: Homology model of *A. fimbriata* putative CNMT in complex with the ligand SAM and the substrate (S)-coclaurine showing the positions of the substrate and SAM in the model: The model has been represented in solid ribbon while the substrate and the ligand have been represented in ball and stick representation with atoms colored differently. Beta strands= blue, alpha helices= red, carbon atoms = grey, oxygen atoms=red, nitrogen atoms=purple, hydrogen atoms=white and sulfur atoms=yellow. Methyl group of SAM to be donated and the N4 of (S)-coclaurine which will receive the methyl group has been pointed.

The SAM binding residues that make hydrogen bonding interactions with SAM are residues Glu96 and Ser97 are located in the loop between α1 and α2, residue Gly135 is the top end residue of β1, residue Gly137 of the loop between β1 and α3, Thr158 and Asn159 come from β2, Gln163 from α4, Ile186 from the loop between β3 and β4, and Ile201 the top end residue of β4. The residues that make hydrophobic interactions with SAM are Leu134, Val157, Asp185, Leu203, His206 and Phe256 lying on β1, β2, loop between β3 and β4, loop between β4 and α5, α5, and α9 respectively.

The substrate is surrounded by one Asp (no. 231), one His (no. 232), one Glu (no. 96), one Gly (no. 202) and four Phe (no. 251, 256, 325 and 349) groups. All the surrounding residues are arranged in such ways that form a pocket for coclaurine. The four Phe residues and the Gly cannot make hydrogen bonds; Asp is oriented in such a way that is unable to make a hydrogen bond. However the N_{D1} of His232 forms a hydrogen bond with O2 of coclaurine. Similarly O_{E1} of Glu96 forms a Hydrogen bond with N4 (the nitrogen to be methylated) of the coclaurine. Glu96 seems to play important role in the transfer of methyl group.

3.8. Comparison with Theoretical Models of Other CNMTs

ModBase data was searched for theoretically modeled CNMTs. Five models were found. These models were downloaded and there images were saved with the help of DS-visualizer. Alpha helices and beta strands of all the six models has been given numbering in a series, starting from N-terminal towards C-terminal. The comparative order of helices (H) and strands (E) of all the six models has been given in table 3.18.

Table 3.18: Comparison of alpha helices (H: red) and beta strands (E: black) of *A. fimbriata* with *A. thaliana, C. japonica, P. somniferum, T. flavum* and *O. sativa*. Some helices or strands are split into two or three or may have a part of the segment in the model (as compared to the segment of the model with the longest segment of that part). Due to these reasons, those cells of the table have been split up accordingly. There are three types of splitting of a cell in the following table.

1: If the segment has split up into two (one happening towards each end as compared to the longest segment amongst the models) or having only half of the segment then the cell has been split up into two.

N-Terminal Part	C-Terminal Part

2: If the segment has either split into three parts or has one third of the longest corresponding segment amongst the models then the cell has been split into three.

N-Terminal	Middle	C-Terminal

3: If the segment has two third as compared to its corresponding longest segment amongst the models then the cell has been split into two unequal segments.

N-Terminal	Middle + C-erminal Part

The table consists of eighteen main columns. Due to the length of the table, it has been split into two parts. The first table is from column one to thirteen and the second table is from column fourteen to column twenty three.

First part of the table;

	1	2	3	4	5	6	7	8	9	10	11	12	13
Species		*		*	*	*	*	*	*		*	*	*
A. fimbriata		H		H	E	H	E	H	E		E	H	H
A. thaliana		H	H	H	E	H	E	H	E	H	E	H	H
C. japonica	H	H	H	H	E	H	E	H	E		E	H	H
P. somniferum	H	H		H	E	H	E	H	E		E	H	H
T. flavum	H	H		H	E	H	E	H	E		E	H	H
O. sativa	H	H		H H	E	H H	E	H	E		E	H	H

Second part of the table

	14	15	16	17	18	19	20	21	22	23
Species	+		+		+	+	*	*	+	+
A. fimbriata	E	H	H		H	E	H	H	H	E
A. thaliana	E		H		H	E	H	H	H H	E E
C. japonica	E	H	H H			E	H	H	H	E
P. somniferum	E		H		H	E	H	H	H	E
T. flavum		H	H H		H		H	H	H	E
O. sativa	E E		H	E	H	E				E

* sign means; the structure is conserved in all the CNMTs models given in the table.

+ sign means; the structure is conserved in most of the CNMTs models in the table.

As is clear from the results the general topology is almost the same for all the CNMTs' models of the species. Most of the secondary structures are conserved in all the CNMTs given.

The number of helices in *A. fimbriata* [Fig. 3.20a], *A. thaliana* [Fig. 3.20b], *C. japonica* [Fig. 3.20c], *P. somniferum* [Fig. 3.20d], *T. flavum* [Fig. 3.20e] and *O. sativa* [Fig. 3.20f] for the modeled segments are 12, 14, 14, 12, 14 and 11 respectively, while the number of beta strands are 7, 8, 7, 7, 5 and 8 respectively. The structural conservation of the putative CNMT model of *A. fimbriata* further confirms its identity. The most conserved segments of the homology model of the putative CNMT of *A. fimbriata* are α1, α2, β1, α3, β2, α4, β3, β4, α5, α6, β5, α8, α9, β 6, α10, α11, α12 and β7. Model of *O. sativa* has a bit different topology. Helix number 2 and 3 (as numbered in *A. fimbriata* model) have each been split into two in *O. sativa* (3, 4 and 5, 6 respectively, as numbered in *O. sativa* model). Fifth beta strand of *A. fimbriata* has been split into two beta strands i.e. number 5 and 6 as numbered in *O. sativa* model. Three conserved helices (10, 11 and 12: *A. fimbriata* numbering) between beta strands number 6 and 7 (*A. fimbriata* numbering; corresponding beta strands are number 7 and 8 in *O. sativa*) are missing in *O. sativa*. Two conserved beta strands (5 and 6: *A. fimbriata* model numbering) between helices 6, 7 and 9, 10 (*A. fimbriata* numbering; corresponding helices in *T. flavum* are 7, 8 and 11, 12 respectively) respectively are missing in *T. flavum* model. There are two main differences with *A. thaliana* model. One extra helix (helix number 6; *A. thaliana* numbering) has appeared between beta strands number 3 and 4 (numbering is the same for all the models). No other model has such a helix in between these two beta strands except *A. thaliana*.

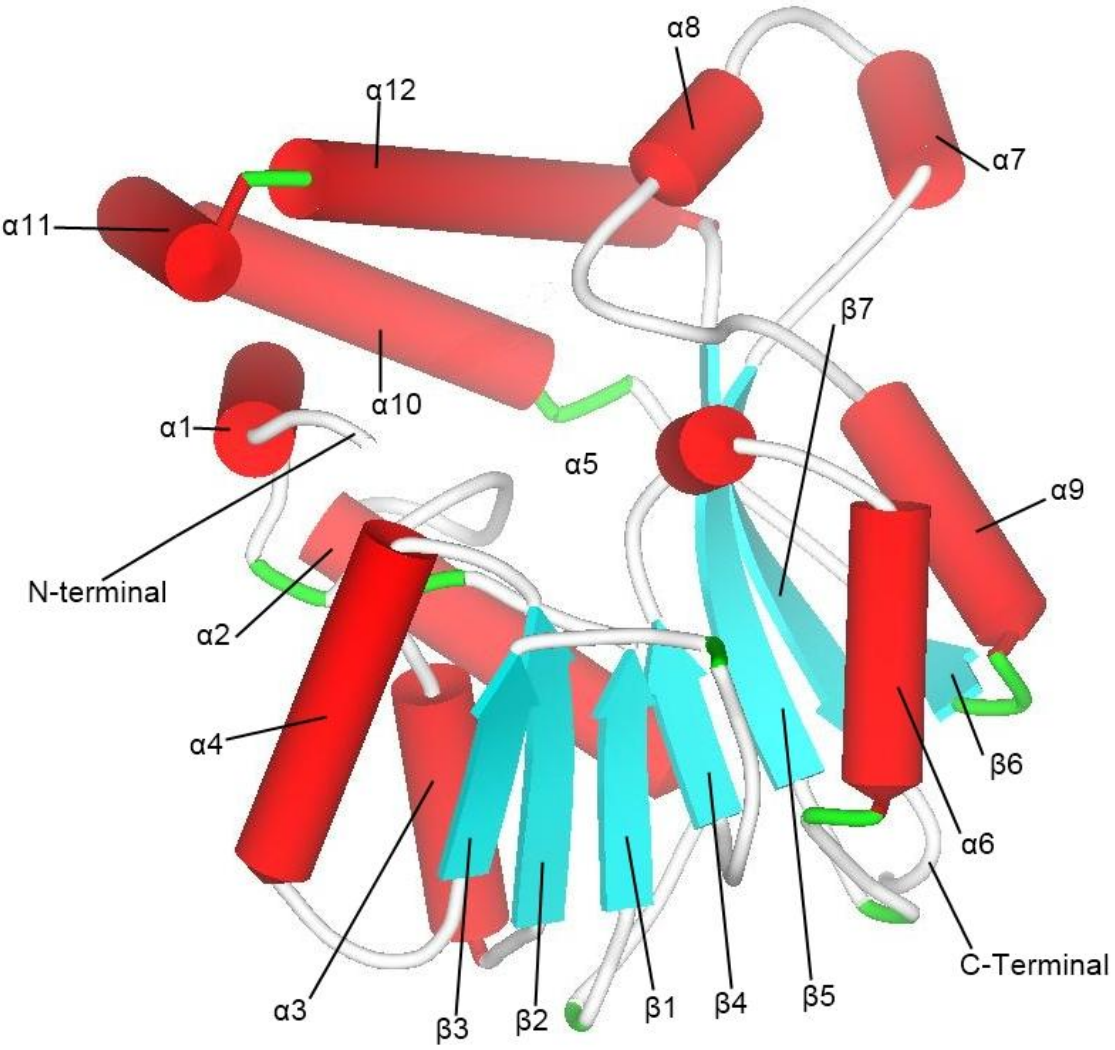


Fig. 3.20a: Schematic representation of the homology model of *A. fimbriata* putative CNMT: The number of helices and sheets only represents the model core. Numbering starts from N-terminal and ends at C-terminal.

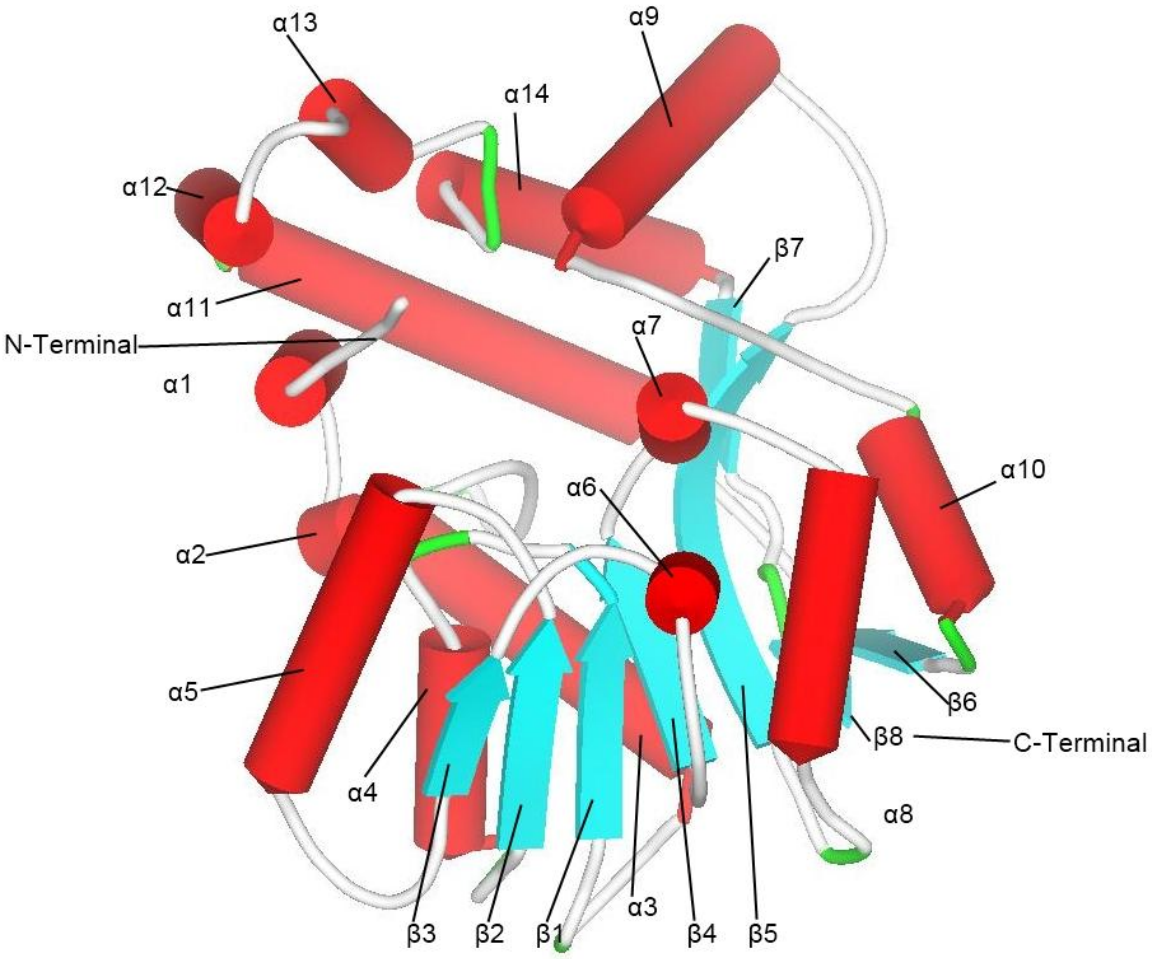


Fig. 3.20b: Schematic representation of the homology model of *A. thaliana* CNMT: The model was downloaded from ModBase database and was pictured with DS-visualizer.

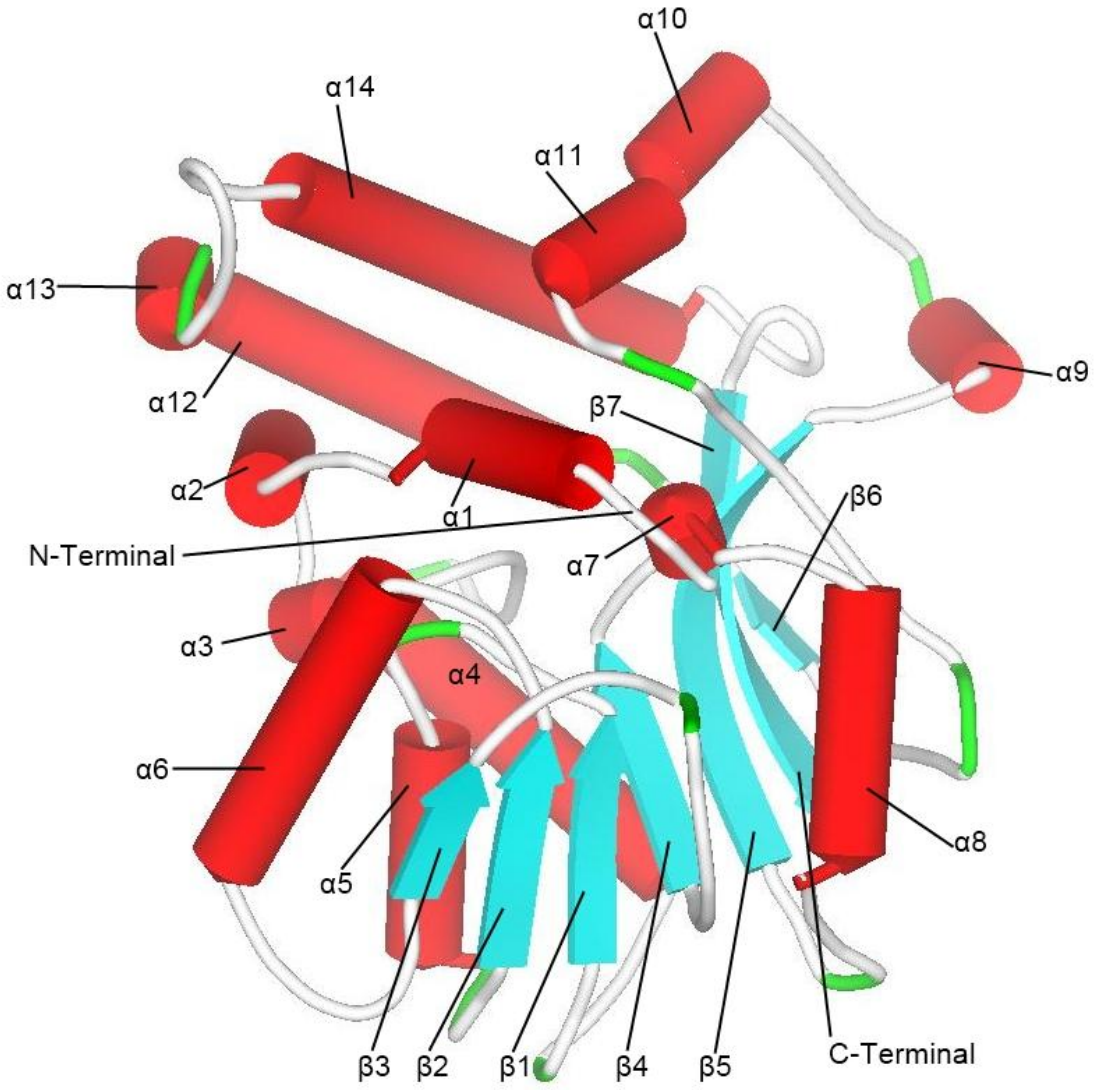


Fig. 3.20c: Schematic representation of the homology model of *C. japonica* CNMT: The model was downloaded from ModBase database and was pictured with DS-visualizer.

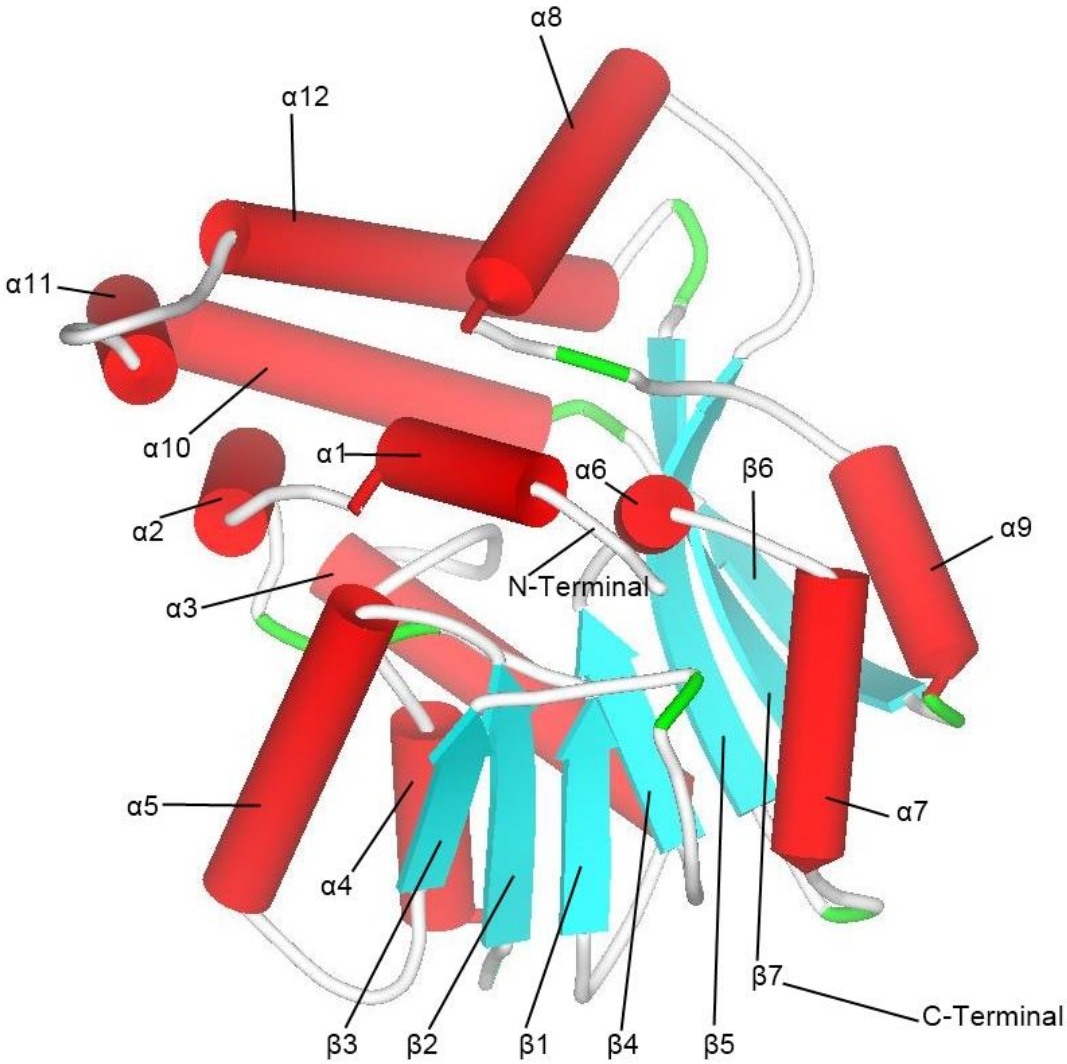


Fig. 3.20d: Schematic representation of the homology model of *P. somniferum* CNMT: The model was downloaded from ModBase database and was pictured with DS-visualizer.

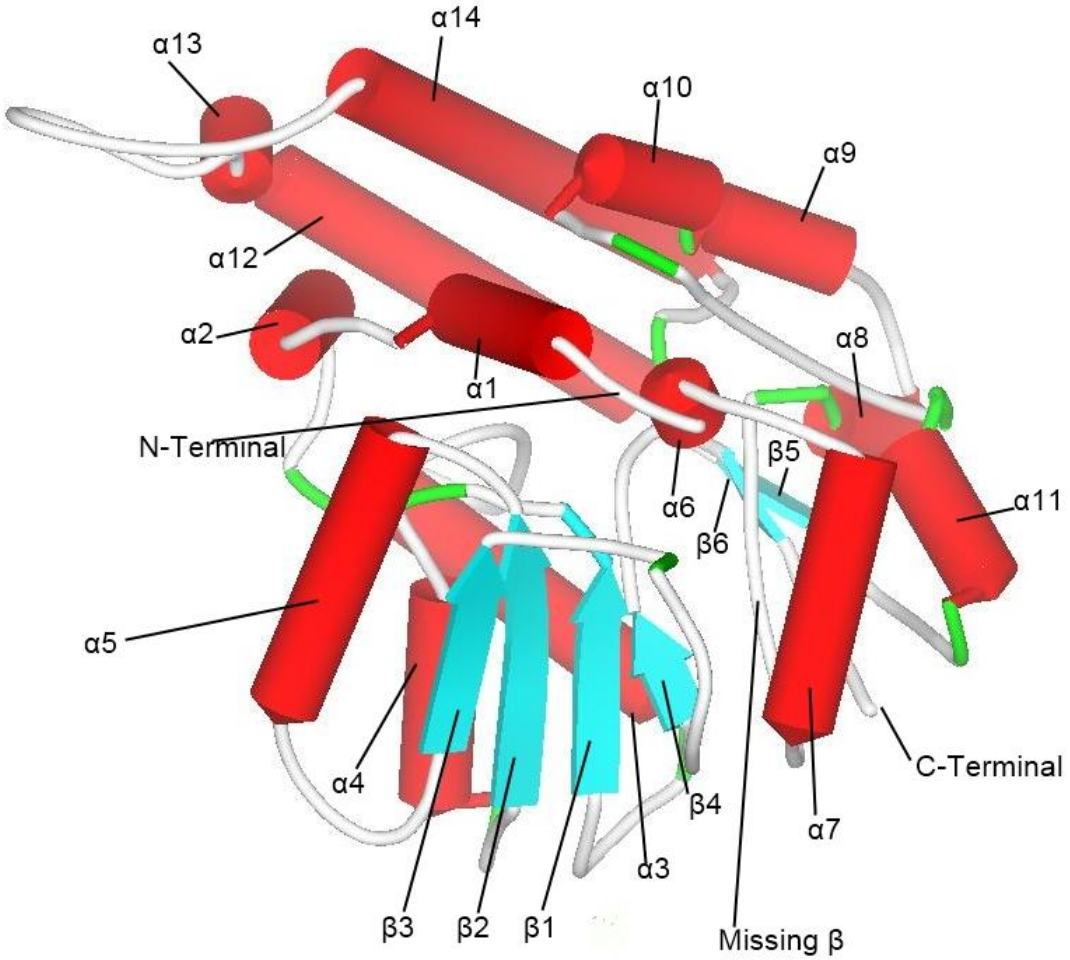


Fig. 3.20e: Schematic representation of the homology model of *T. flavum* CNMT: The model was downloaded from ModBase database and was pictured with DS-visualizer.

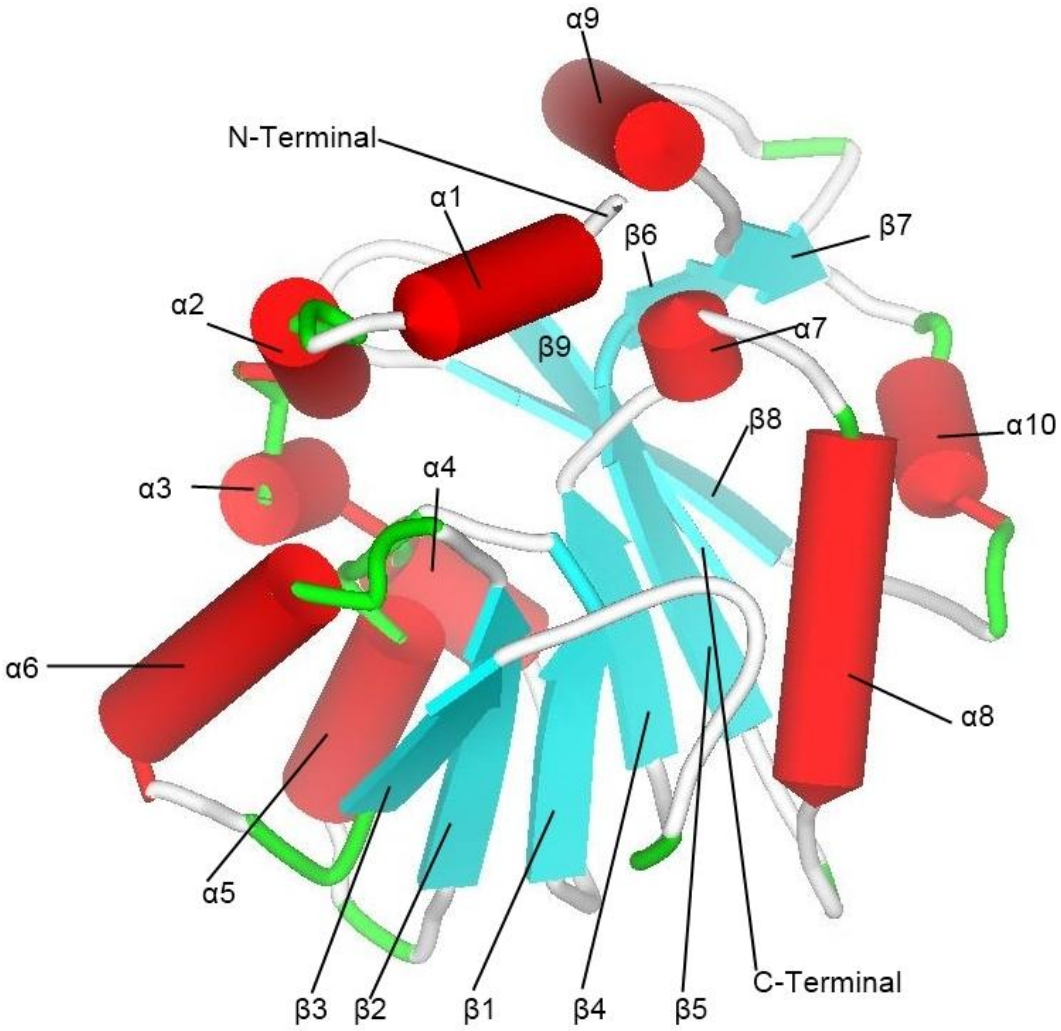


Fig. 3.20f: Schematic representation of the homology model of *O. sativa* putative CNMT: The model was downloaded from ModBase database and was pictured with DS-visualizer.

However models show some signs, that there might be an alpha helix in other models too as there appears a green region at the segment of the coil where *A. thaliana* has a helix. Another difference with *A. thaliana* is that the last beta strand (number 7; *A. fimbriata* model numbering) has been split up into two shorter beta strands i.e. 7 and 8 (as numbered in *A. thaliana*). These differences in the topology are due to the templates and the alignment used to create these models and of course due to the differences in their sequences. There are other minor differences in the models topology which can be seen in the table 3.18 as well as in figures 3.20a-3.20f. The arrangement of beta strands in all of them is the same except their numbers. All except the last beta strands i.e. number 7 as numbered in *A. fimbriata* model (numbering is different for other models: split into two parts in *A. thaliana*) are parallel.

3.9. Binding Site

3.9.1 Active Site Identification

A large class of SAM-MTases has been found to share a conserved catalytic domain structure due to the interaction of the enzymes with a common cofactor, SAM [352]. The common 3D structure of these enzymes is reflected in sequence motifs that are conserved among a large number of SAM-MTases [353-355]. To identify SAM-binding residues of the protein three different methods were used; CDD search, CASTp search and ModBase search. All the three searches gave almost similar results. According to the CDD results [Fig. 3.9] the conserved region consisted of seven motifs 130-144, 151, 153-159, 179-186, 195-201, 212-218 and 225-234. The SAM-binding residues were predicted to be 133-DLGCGQGAF-141, 158-TN-159, 184-EDI-186, 201-I-201, 212-LL-213, and 231-DH-232. CASTp predicted a similar

result with few extra motifs [Fig. 3.21a]. The predictions consisted of 26 motifs which might also include substrate binding sites. The residues which were indicated as ligand binding included 94-LKES-97, 115-M-115, 133-DLGCG-137, 141-F-141, 157-VTNS-160, 163-Q-163, 186-IT-187, 201-IGL-203, 206-HM-207 and 209-N-209.

ModBase identified several CNMTs from different plants with theoretically predicted models. It searched out five CNMTs of high sequence identity. The SAM-binding residues were predicted for all the five models. Those residues were compared with the CDD and CASTp results. The main binding site residues for SAM were almost identical [Fig. 3.21b]. Active site residues 95-K-S-97, 133-D-GCG-137, 159-N, 163-Q, 186-I, 206-H (*A. fimbriata* sequence numbering, different numbering for other CNMTs) were identical in all five ModBase results; however residues 94-L-E-96, 115-M, 134-L, 141-F, 157-V, 185-D, 201-IGL-203 and 209-N of *A. fimbriata* putative CNMT were similar but not identical to all other CNMTs. In two of the CNMTs “L” was replaced with “M”, “E” was replaced with either “G”, “Y” or “Q”, “M” was replaced with “I” in only one, “L” was replaced with either “V” or “I”, “F” was replaced with either “L” or “V”, “V” was replaced with either “L” or “V”, “V” was replaced with “I” in only one, “I” with “V”, “G” with “E”, and “L” with “M”.

Multiple alignment of all the CNMTs selected [Fig. 3.21c] shows that residues E-16, P-21, R-30, L-33, R-36, Y-79, P-82, G-91, K-95, S-97, L-108, A-111, E-112, L-116,118, Y-119, 121-ERA-123, G-128, L-132, 135-GCG-137, G-139, L-143, A-146, S-160, Q-163, K-164, I-167, N-178, 196-DR-197, 205-EHMKNY-210, 213-LL-214, W-220, 228-LF-229, H-235, Y-240, E-243, 249-DW-250, F-256, G-259, S-265, 270-LYFQ-273, W-282, G-286, H-288, E-294, W-296, D-301, A-321, 328-WR-329, 339-F, Y-341,G-344, 346-EW-347, H-351, and 353-LF-354, are conserved in the CNMTs.

```
  1 MASEKLNKTE MLRRLEEGSV PDEEFRRLIR IELGRRLRWY CQKKPTYEEQ TAEIVALVKA
 61 LRQMGITGDQ SDQLNSDLYD MPMSFLKITF GKLLKESGSY FKDDSMTLDE AEEAMLDLYC
121 ERAQLKDGQK ILDLGCGQGA FTLHAAQKYK KSHVTAVTNS ATQKKYIEDQ CQILELSNVE
181 VLLEDITQLT MEATFDRIIV IGLLEHMKNY GLLLQNISQW MAPADSLLFI DHVCHKSYFY
241 QCEPLDEDDW FAEYFFPPGS FAMPSASFLL YFQDDVSIVD HWILSGKHFH RTAEEWVKQL
301 DTNLEKGKEI LESKYGSKEA ALKAFNHWRG LCIFSSEIFG YNGGEEWMTS HLLFKKK
```

Fig. 3.21a: Conserved residues of *A.fimbriata* putative CNMT sequence as identified by CASTp: The conserved residues have been highlighted in red color.

```
                                *    *    *       * *  *   **  *
Coptis      : MAVEAKQT-- KKAAIVELLK QLELGLVPYD DIKQLIRREL ARRLQWGY-- --KPTYEEQI AEIQNLTHSL
Papaver     : ---------M QLKAKEELLR NMELGLIPDQ EIRQLIRVEL EKRLQWGY-- --KETHEEQL SQLLDLVHSL
Thalictrum  : MAVEGKQVAP KKAIIVELLK KLELGLVPDD EIKKLIRIQL GRRLQWGC-- --KSTYEEQI AQLVNLTHSL
Oryza       : --MAMAARAA YLAATRAALA ALERNALPDA VTRRLTRLLL AQRLRLGYLP SSSSSAPLHL HHLLLFAHAL
Arabidopsis: --MEKIIDVA YGASVKVGLT LLEMNLLPDI VIRRLTRLLL AGRLRSGYKP ----TAEMQL SDLLRFVDSI

              *  *     *      **  *  *     *  * * *  *     *  *  * * * * *  * **   *
Coptis      : RQMKIAT-EV ETLDSQLYEI PIEFLKIMNG SNLKGSCCYF KEDSTTLDEA EIAMLDLYCE RAQIQDGQSV
Papaver     : KGMKMAT-EM ENLDLKLYEA PMEFLKIQHG SNMKQSAGYY TDESTTLDEA EIAMLDLYME RAQIKDGQSV
Thalictrum  : RQMKIAT-EV ETLDDQMYEV PIDFLKIMNG SNLKGSCCYF KNDSTTLDEA EIAMLELYCE RAQIKDGHSV
Oryza       : EEMPIAI-ET EKAKDQHYEL PTTFFKLVLG RNLKYSSCYF PDESSTLEDA EVAMLELYCE RAQLQDGQTI
Arabidopsis: KKMPIAI-HT EKPKTQHYEL PTAFFELVLG RNMKYSSCYF SNDSSSLEDA EEAILALYCE RAKVEDGQSV

              ** ***      *  * *       *   **   *   *          * *    ***          ***
Coptis      : LDLGCGQGAL TLHVAQKYKN CRVTAVTNSV SQKEYIEEES RRRNLLNVEV KLADITTHEM AETYDRILVI
Papaver     : LDLGCGLGAV ALFGANKFKK CQFTGVTSSV EQKDYIEGKC KELKLTNVKV LLADITTYET EERFDRIFAV
Thalictrum  : LDLGCGQGAL TLYVAQKYKN SRVTAVTNSV SQKEFIEEES RKRNLSNVEV LLADITTHKM PDTYDRILVV
Oryza       : LDVGCGWGSL SLYIAKKYSK CSITGICNST TQKAFIEEQC RENELSNVEI IVADISKFEM ERSFDRIISI
Arabidopsis: LDIGCGWGSL SLYIARKYSK CKLTGICNSK TQKAFIDEKC RKLGLQNIEI IVADISTFEH EGTYDRIFSI

              *  ******   ** *   *     * ***     **** ***  *     **  *   *         **   **
Coptis      : ELFEHMKNYE LLLRKISEWI S-KDGLLFLE HICHKTFAYH YEPLDDDDWF TEYVFPAGTM IIPSASFFLY
Papaver     : ELIEHMKNYQ LLLKKISEWM K-DDGLLFVE HVCHKTLAYH YEPVDAEDWY TNYIFPAGTL TLSSASMLLY
Thalictrum  : ELFEHMKNYE LLLRKIKEWM A-KDGLLFVE HICHKTFAYH YEPIDEDDWF TEYVFPAGTM IIPSASFFLY
Oryza       : EMFEHMKNYK ALLKKLSRWM K-EDSLLFVH YFCHKTFAYH FEDNNEDDWI TRYFFTGG-- TMPSANLLLY
Arabidopsis: EMFEHMKNYG ELLKKIGKWM K-EDSLLFVH YFCHKTFAYH FEDVNDDDWI TRHFFSGG-- TMPSADLLLY

              **  **     *  ** *      *****  *                    *      **  *     * **
Coptis      : FQDDVSVVNH WTLSGKHFSR TNEEWLKRLD ANLDVIKPMF ETLMGN-EEE AVKLINYWRG FCLSGMEMFG
Papaver     : FQDDVSVVNQ WTLSGKHYSR SHEEWLKNMD KNIVEFKEIM RSITKT-EKE AIKLLNFWRI FCMCGAELFG
Thalictrum  : FQDDVSVVNH WTLSGKHFSR TNEEWLKRLD ANVELIKPMF VTITGQCRQE AMKLINYWRG FCLSGMEMFG
Oryza       : FQDDVSIANH WLVSGTHYAR TSEEWLKRMD KNITSIRPIF EKTYGK--ES ATKWIAYWRT FFISVAELFG
Arabidopsis: FQENVSIVDH WLVSGTHYAK TSEEWLKRMD KEIVAVKEIM EVTYGK--EE AVKWMVYWRT FFIAVAELFG

              * ** ***   * ** **
Coptis      : YNNGEEWMAS HVLFKKK
Papaver     : YKNGEEWMLT HLLFKKK
Thalictrum  : YNNGEEWMAS HVLFKKK
Oryza       : YNNGDEWMVA HFLFRKK
Arabidopsis: YNNGEEWMVA HFLFKKK
```

Fig. 3.21b: Active site residues of different species CNMTs, as predicted by ModBase, have been highlighted in red color. The sequences of 5 species i.e. *C. japonica, P. somniferum, T. flavum, O. sativa* and *A. thaliana*, have been aligned to each other. Most of the active site residues are conserved in all five sequences. The conserved residues have been starred.

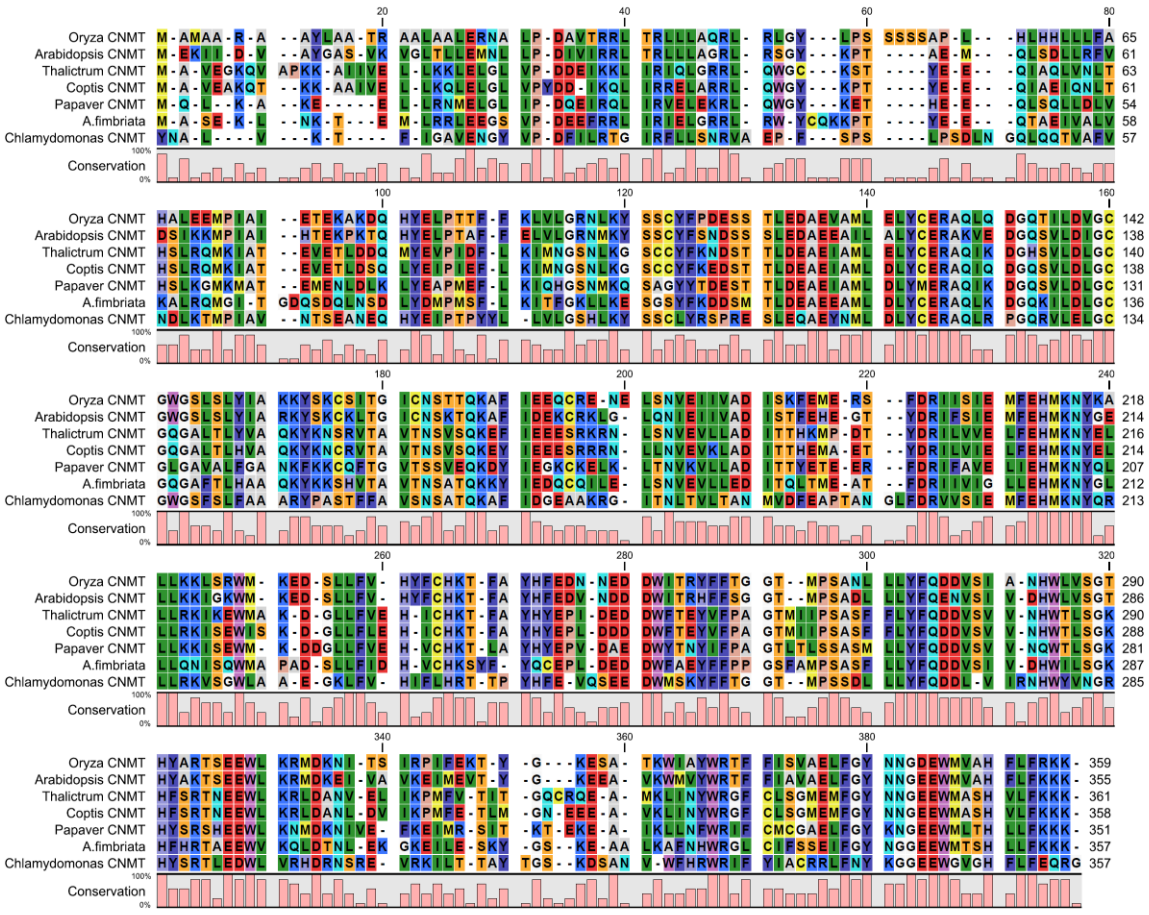


Fig. 3.21c: Multiple Alignment of CNMTs of different plants including *A. thaliana, C. japonica, T. flavum, P. somniferum,* and *A. fimbriata*, and the green alga; *C. reinhardtii* CNMT sequences: The image has been created using CLC-sequence viewer version-6.4 (ww.clcbio.com) with rasmol background color option. The graph shows the percent conservation of the residue.

The last three residues of Lys are conserved in all except *Chlamydomonas*. In the *O. sativa* putative CNMT, only one of the three Lys residues is different. When compared SAM binding residues, it was found that most of the SAM-binding residues are among these conserved residues.

Only one residue i.e. F-256 (among coclaurine binding residues) is found among the conserved residues. Other substrate and SAM-binding residues are also conserved, although the residues are different in some of the CNMTs, but those different residues are characteristically similar to one another, therefore we can say that they are also conserved.

3.9.2 *A. fimbriata* Putative CNMT Model Active Site Residues as Calculated by Ligplot

SAM is bound in a similar position and conformation in *A. fimbriata* putative CNMT model to other SAM-MTases, which is evident from the superimposition of the model over SAM-dependent 2fk8 crystal structure [Fig. 3.18c and 3.18d]. SAH/SAM binding is mediated through extensive hydrogen bonding and other van der Waals interactions. The number of hydrogen bonds and the residues involved in hydrogen bonding with SAM were highly conserved. There are a total of ten hydrogen-bonding interactions to the Ligand (SAM) [Fig. 3.22a], however there are eleven in 2fk8 [Fig. 3.22b]. There is only one hydrogen bond with the adenine ring of the SAM that comes from 186-Ile and is between N1 of the adenine ring and the backbone amide of Ile in *A. fimbriata* CNMT, however there are two hydrogen bonds with Adenine ring in 2fk8.

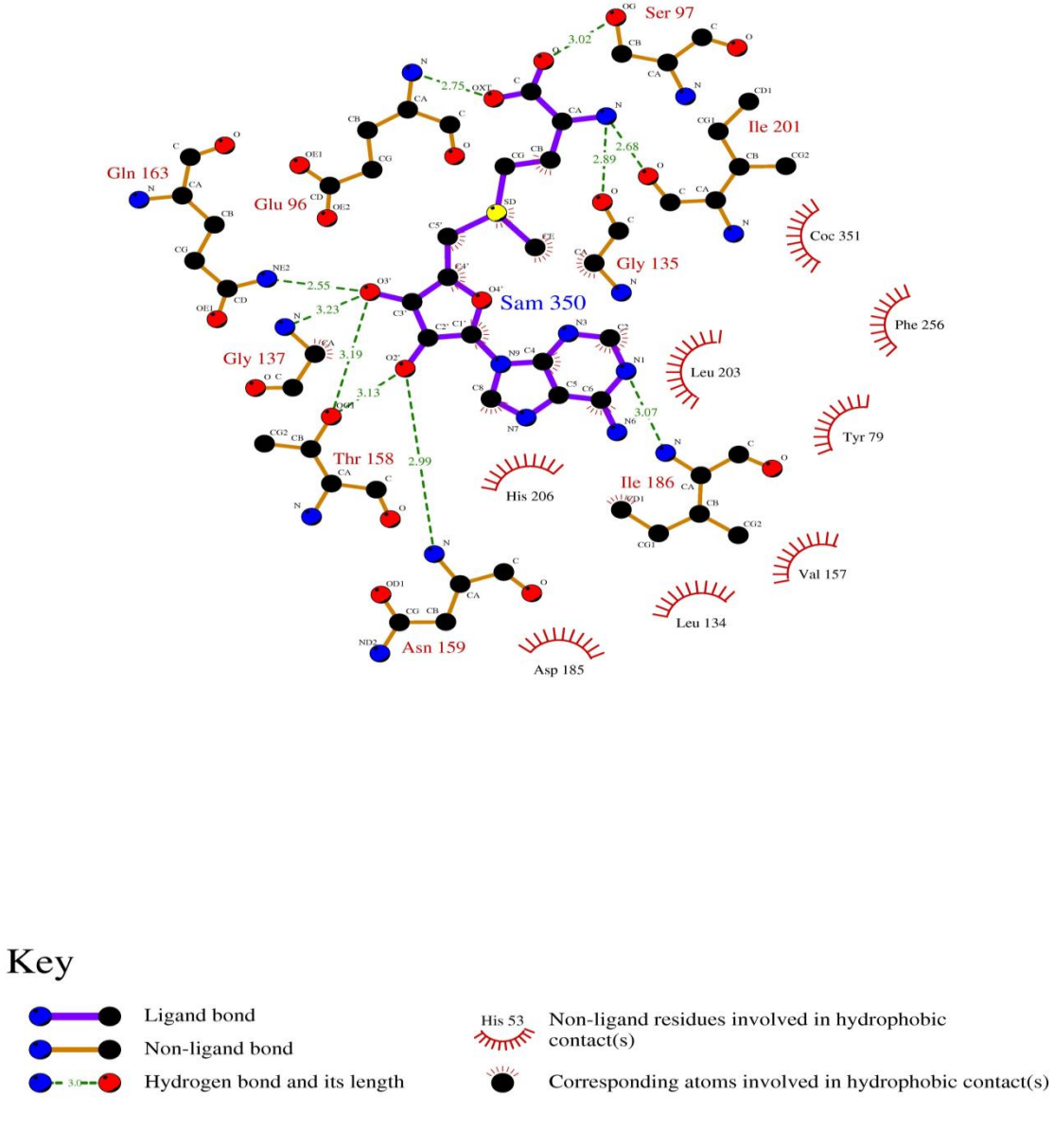


Fig. 3.22a: Ligplot results showing the SAM binding residues in the *A. fimbriata* putative CNMT model: Atoms have been colored as follows: carbon=black, nitrogen=blue, oxygen=red and sulfur=yellow.

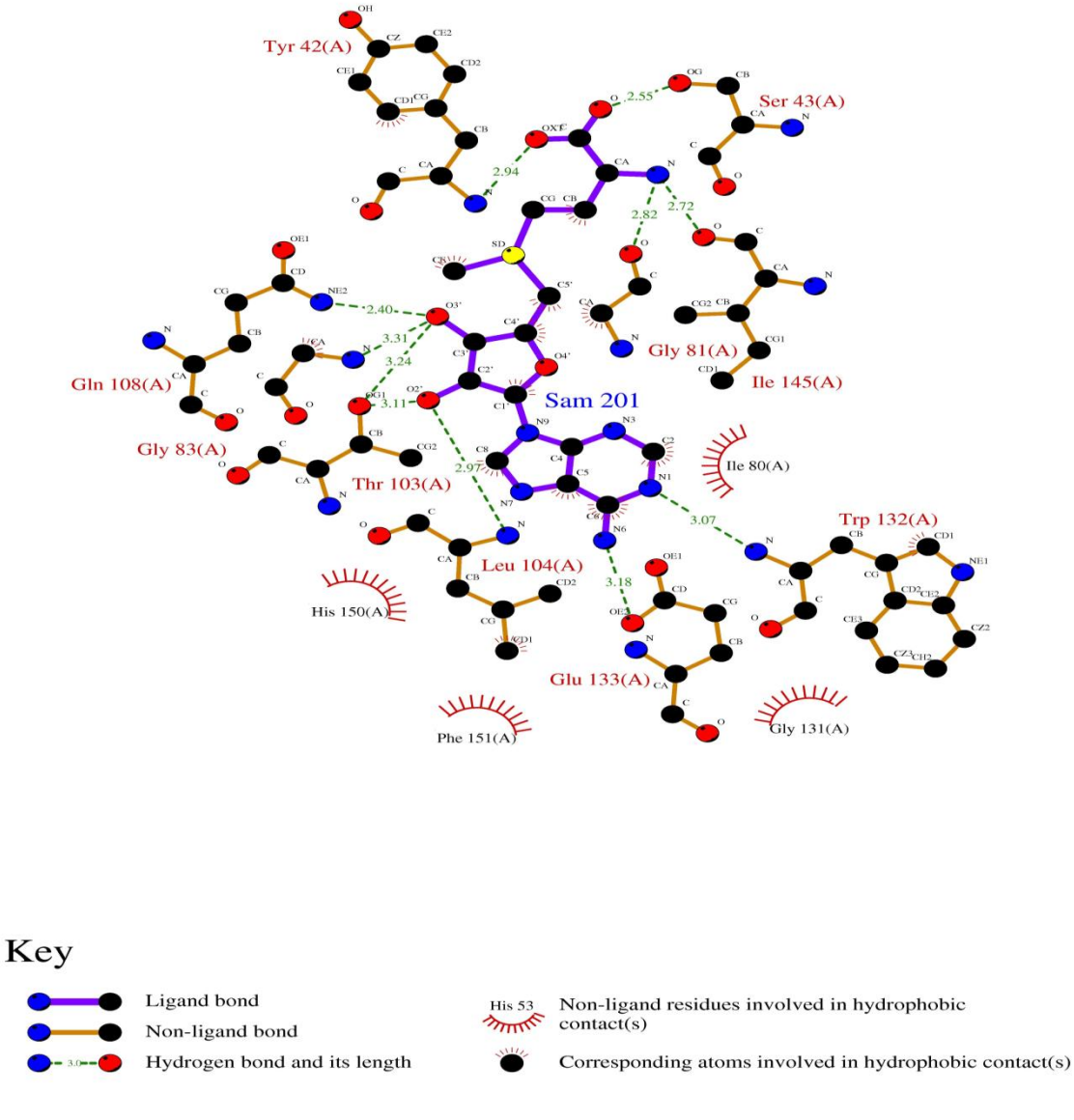


Fig. 3.22b: Ligplot results showing the SAM binding residues in the template 2fk8 crystal structure: Atoms have been colored as follows carbon=black, nitrogen=blue, oxygen=red and sulfur=yellow.

The first one is between the exocyclic amino group N6 and the side chain carboxyl group of Glu-133 and the second one occurs between N1 and the backbone amide of Trp-132. In CNMTs Glu is replaced with a shorter residue Thr which cannot make hydrogen bond with the N6 of SAM. The distance of the side chain hydroxyl group of Thr is 5.243 Å in CNMT which is too far for a hydrogen bond to be formed. The rib hydroxyls of SAM form four hydrogen bonding interactions with strictly conserved residues Gln-163, Gly-137 and Thr-158 both in *A. fimbriata* CNMT and 2fk8 crystal structure (Gln-108, Gly-83 and Thr-103 in 2fk8). Rib O3′ of SAM forms hydrogen bonds with the side chain amide group of Gln, amide group of Gly and hydroxyl group of Thr. Hydroxyl group of Thr also forms another hydrogen bond with O2′ of the rib of SAM. The O2′ of the rib forms another hydrogen bond with the backbone amide group of Asn-159 in CNMT and backbone amide group of Leu-104 in 2fk8. The tail amino group of SAM forms hydrogen-bonding interactions with the backbone carboxyl groups of Gly-135 and Ile-201 in CNMT (Gly-81 and Ile-145 in 2fk8). The carboxyl tail of SAM forms two hydrogen-bonding interactions, one with the hydroxyl group of Ser-97 and the other with backbone amide of Glu-96 in CNMT; however Glu is replaced with Tyr in 2fk8. The bond is between the tail carboxyl of SAM and the backbone amide of Tyr-42. Other than hydrogen-bonding interactions there are hydrophobic interactions between SAM and Tyr-79, Leu-134, Val-157, Asp-185, Leu-203, His-206 and Phe-256. There are fewer hydrophobic interactions between SAM and other amino acid residues in 2fk8. These interactions are with Ile-80, Gly-131, His-150 and Phe-151.

3.9.3. SAM Binding Site

SAM and SAH are metabolites that are involved in the conversion of methionine to homocysteine in the proximal end of the methionine cycle. SAM is synthesized from methionine and ATP in a reaction catalyzed by methionine adenosyltransferase. The methyl group is provided by SAM which is also involved in several dozens of transmethylation reactions of crucial biological importance [356].

In general, MTases posses three conserved binding motifs: motif 1 containing the GXGXGG sequence, is involved in binding to the amino acid portion of SAM; motif 2 having an acidic and a hydrophobic residue, the acidic residue is responsible for making bonds to the ribose moiety and the hydrophobic residue is involved in forming the hydrophobic pocket to stabilize the adenine ring while motif 3 make contacts to the adenine ring via ionic or hydrogen bonds with D/E/N/Q residues [357].

The conserved regions; motifs 1, post-1, 2, and 3 are always found in the same order on the polypeptide chain and are separated by comparable intervals [354]. The three-dimensional structures of SAM-MTases have shown that motif 1 and post-1 interact directly with the SAM residue, while motifs 2 and 3 make contacts with each other and with a portion of motif 1 to form the main central portion of the b-sheet.

The glycine-rich sequence containing Gly-72 and Gly-74 ((E/ D)XGXGXG), often referred to as motif 1, is highly conserved in many SAM-MTases and is also present in TNMTs, CNMTs, and MtPcaA [15, 43, 286]. MtPcaA exhibits a structural fold which is most similar to the small molecule subclass of methyltransferases [43]. The conservation of critical residues involved in SAM-binding suggests that TNMT and CNMTs also maintain this core small molecule methyltransferase fold. A

methyltransferase fold model also predicts that SAM-binding motifs happen in the N-terminal domains of TNMT and CNMTs. As such, amino acid residues that confer substrate specificity are most likely situated in the C-terminal regions of these proteins [43]. In the absence of crystal structures for CNMTs and TNMT, residues that confer substrate and reaction specificity with respect to nitrogen methylation might be identified among C-terminal residues that are conserved in CNMTs.

SAM/SAH binds to the pocket that is long and wide enough to hold the ligand. This pocket is located near the substrate binding pocket where the donor-methyl group protrudes outward from the substrate pocket toward the ligand binding pocket [Fig. 3.23a]. This type of situation is almost similar to the binding of SAM in *Streptococcus pneumoniae* Sp1610 (a putative tRNA methyltransferase) [39]. The bound SAM is located on the inner surface of a large cavity [Fig. 3.23a]. SAM binds to Glu-96, Ser-97, Gly-135, Gly-137, Thr-158, Asn-159, Gln-163, Ile-186 and Ile-201, as well as interacting with Tyr-79, Leu-134, Val-157, Asp-185, Leu-203, His-206 and Phe-256 through weak Van der Waals interaction [Fig. 3.22a]. However in case of Sp1610 the SAM is stabilized by Glu46 and Arg5 by forming hydrogen and ionic contacts with SAM and is stabilized further by several residues in the binding pocket. The adenine moiety of SAM makes hydrophobic contacts with Val22, Val47, Asn74, Leu97, Met93, and Ile101 while the N6 atom of adenine and Asn74 forms a hydrogen bond and the hydroxyl groups in the ribose ring form hydrogen bonds with Glu46. The carboxyl groups of SAM in Sp1610 are recognized by Arg5 with the charge enhanced hydrogen bond and the amino groups of SAM are in contact with the backbone carbonyl groups of Gly23 and Ala91 through hydrogen bonding [39].

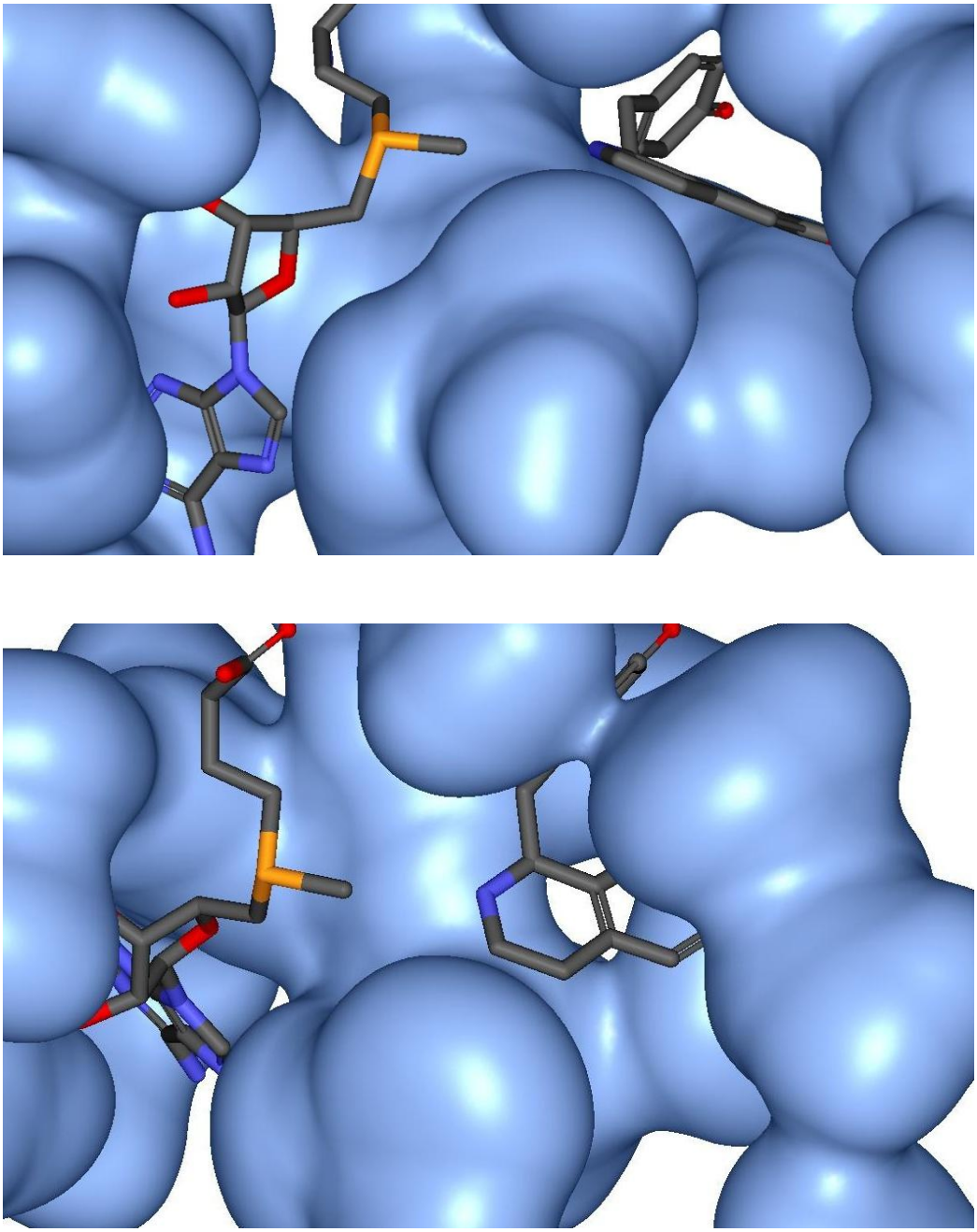


Fig. 3.23a: Stick representation of the substrate (S)-coclaurine and SAM from two different angles within their pockets showing the predicted positions of both the residues with respect to each other in the molecule: Each of these residues is in their own cavity situating next to each other. The two cavities are connected through a narrow opening in between the two cavities through which the methyl group transfer takes place.

The *A. fimbriata's* putative CNMT shows a total of 10 hydrogen bonds between CNMT and SAM [Fig. 3.22a], which seems to fix exactly the position and orientation of SAM. Notably, the OH of Tyr-79 and the sulfur atom of SAM are positioned in close vicinity [Fig. 3.23b]. Most of the MTases have been observed to possess the interactions between the sulfur atom of SAM/SAH and an oxygen atom of an amino acid of the enzyme [358] for example in rRNA N6-MTase, the interaction between the sulfur atom of SAM and an oxygen atom of a main chain Asn residue [359], and in histone-lysine N-MTase, the interaction between the sulfur atom of SAH and the Oδ atom of an Asn residue [360]. These interactions are required to be studied further for the elucidation of their precise role.

The active sites of some of the methyltransferases have been studied in detail e.g. Glycine-N-methyltransferase (GNMT), which catalyzes the SAM-dependent methylation of glycine producing sarcosine [361]. It has been shown that the carboxylate and amino groups of the Met moiety of SAM in GNMT are involved in hydrogen bonding with Trp30, Arg40, Ala64, and Leu136, however in case of *A. fimbriata* putative CNMT model, the carboxylate and amino groups of Met moiety of SAM are involved in hydrogen bonding with Glu-96, Ser-97, Gly-135 and Ile-201 [Fig. 3.22a]. The number of hydrogen-bonded residues is four in both the GNMT and the putative CNMT. The residues of CNMT are almost identical in properties with the residues of GNMT involved in hydrogen bonding with SAM.

Another important residue that helps in methyl transfer reaction is Tyr.

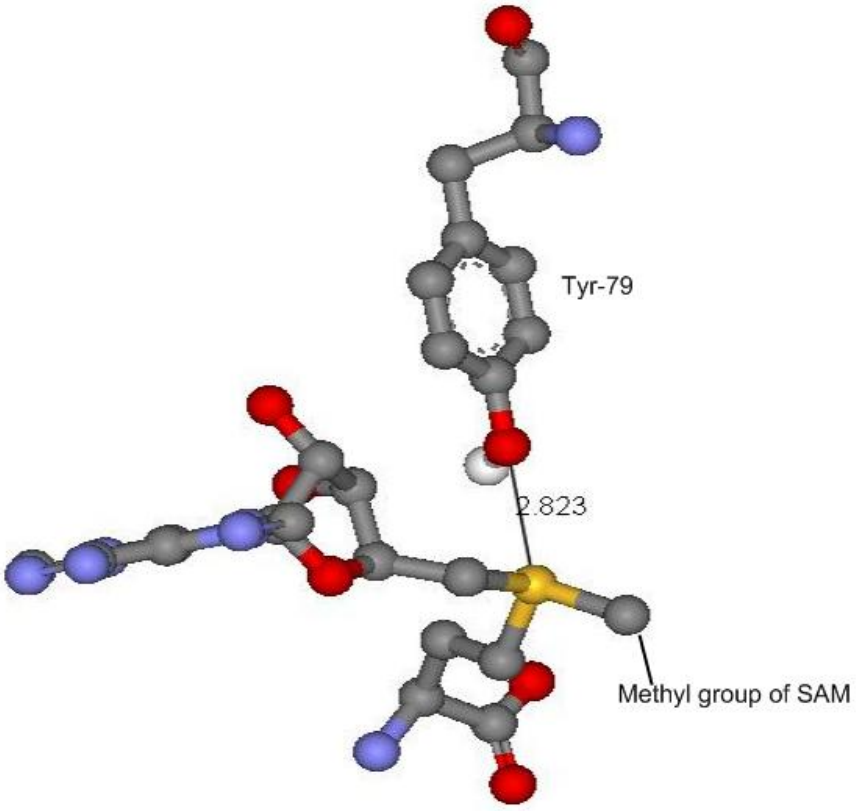


Fig. 3.23b: Ball and stick representation of SAM and Tyr-79: The OH of Tyr79 is positioned to have a charge-dipole interaction with the positively charged S_D of SAM, which might help in the loosening of the methyl group of SAM.

In GNMTs Tyr-21 has been implicated in its methyl transfer mechanism [41] however in CNMTs Tyr-79 (As numbered in *A fimbriata* putative CNMT) is conserved [Fig. 3.21c] and its hydroxyl group is in close proximity to the positively charged S_D atom of SAM and is positioned and oriented in such a way to make interactions with the positively charged S_D atom of SAM which facilitates the transfer of methyl group [40, 41]. From this it can be inferred that Tyr-79 (numbering is different for different CNMTs) might be involved in transfer of methyl group in CNMTs of plants [Fig. 3.23b].

3.9.4. Substrate Binding Pocket

The substrate is located just next to the SAM-binding cavity because the SAM-dependant methylation reaction involves the direct transfer of the methyl group from SAM to the substrates, and thus the cavities for both the ligand and substrate should be positioned close to each other and oriented in such a way that they can transfer the methyl group easily [39]. Coclaurine is properly positioned in the active site for methylation. The residues surrounding the substrate-binding cavity are different in different MTases and are less conserved. They change according to the substrate nature and its interactions with the SAM. These differences are important for both substrate recognition and catalysis [40]. The substrate is surrounded by four phenyl rings and a guanidinium ring of His in such a way that it makes a pocket that just suits the (S)-coclaurine. One of the four Phe (Phe-256) surrounding the coclaurine is conserved in all CNMTs as shown by the multiple alignment [Fig. 3.21c]. There is a large cavity between the two domains of CNMT, which is large enough to hold the substrate, coclaurine [Fig. 3.24a].

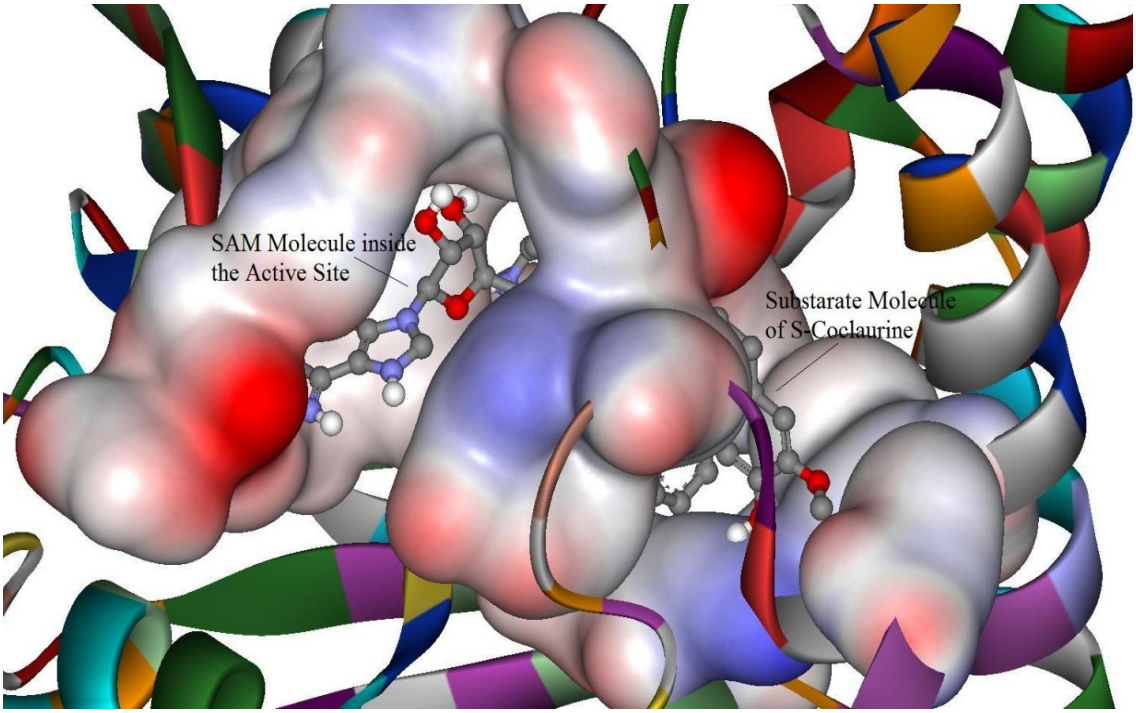


Fig. 3.24a: Two cavities for SAM and (S)-coclaurine: SAM-binding cavity is on the left and the substrate-binding cavity has been shown on the right. The ligand and substrate have been shown in ball and stick form inside their active sites. The two cavities are situated next to each other. The cavities have been made apparent by adding the closed surface around the ligand and substrate binding residues, leaving the remaining chain in the solid ribbon conformation.

Moreover, the SAM-binding site is located next to this cavity and there is an opening in between these two cavities through which the methyl group of SAM projects toward this cavity, suggesting strongly that coclaurine binds to this cavity. The substrate is oriented in such a way that the nitrogen atom of coclaurine to be methylated comes just in front of the opening that connects the two cavities, which is just suitable for the transfer of the methyl group to the coclaurine [Fig. 3.23a and 3.24b]. The N4 of coclaurine receives the methyl group from SAM through this opening [Fig. 3.24b and 3.24c]. Coclaurine binds to Glu-96 and His-232 through hydrogen bonds. Coclaurine makes a total of two hydrogen bonds. One hydrogen bond is between N_4 of tetrahydroisoquinoline moiety of the substrate and O_{E1} group of Glu-96. Second hydrogen bond is between exocyclic O2 of tetrahydroisoquinoline of the substrate and N_{D1} of the ring of His-232. There are a total of six residues that make hydrophobic interactions with coclaurine. These residues are Gly-202, Asp-231, Phe-251, Phe-256, Phe-325, and Phe-349 [Fig. 3.25a]. The two hydrogen bonds between CNMT and coclaurine fix exactly the position and orientation of coclaurine. Notably, the O_{E1} atom of Glu-96 and the N_4 atom of coclaurine are positioned in close vicinity forming a hydrogen bond [Fig. 3.25b] that might facilitate the transfer of methyl to coclaurine.

3.10. Putative Reaction Mechanism

The model of the putative CNMT presented here has a loop at the N-terminal [Fig. 3.19a] part that might be flexible enough to allow the substrate and SAM to enter into the active site as well as the reaction products to exit [40]. After the ligand and the substrate get entered, the ligand (SAM) can get the potential to transfer its methyl group to the substrate.

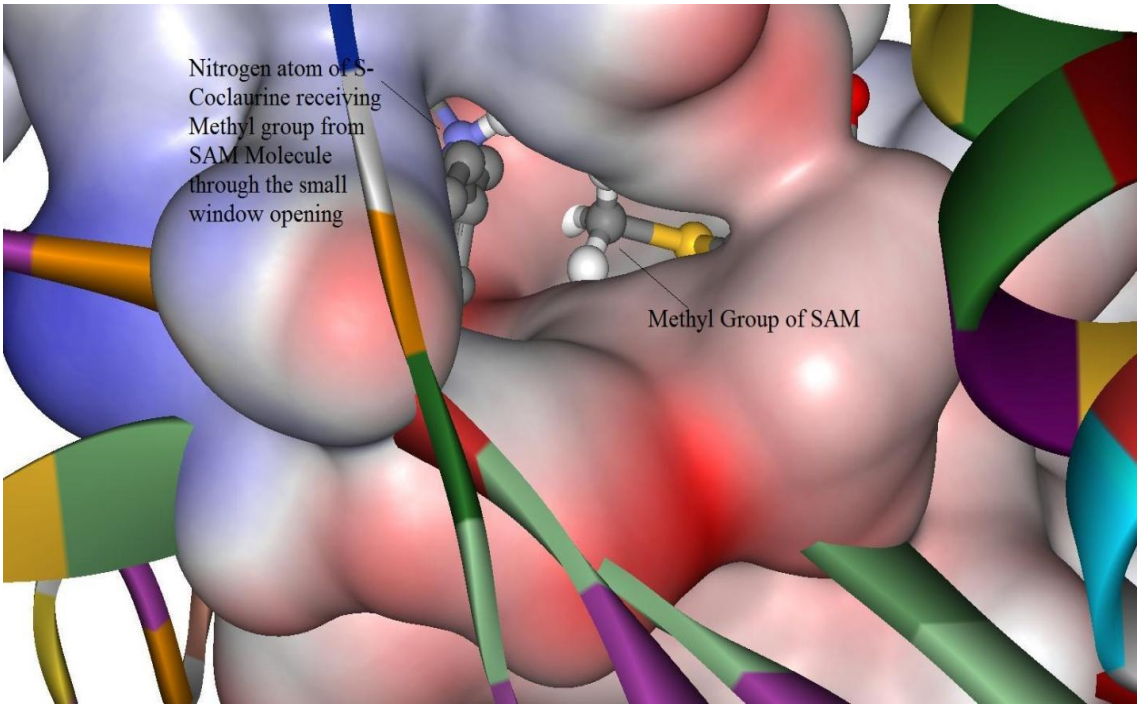


Fig. 3.24b: The small opening between the two cavities, through which the methyl group has been shown, projected towards the nitrogen of coclaurine: The methyl group of the SAM and the receiver N4 of coclaurine has been shown in the diagram with the ball and stick representation. The binding residues are illustrated with a solid surface around them to make the channel opening between the cavities clear.

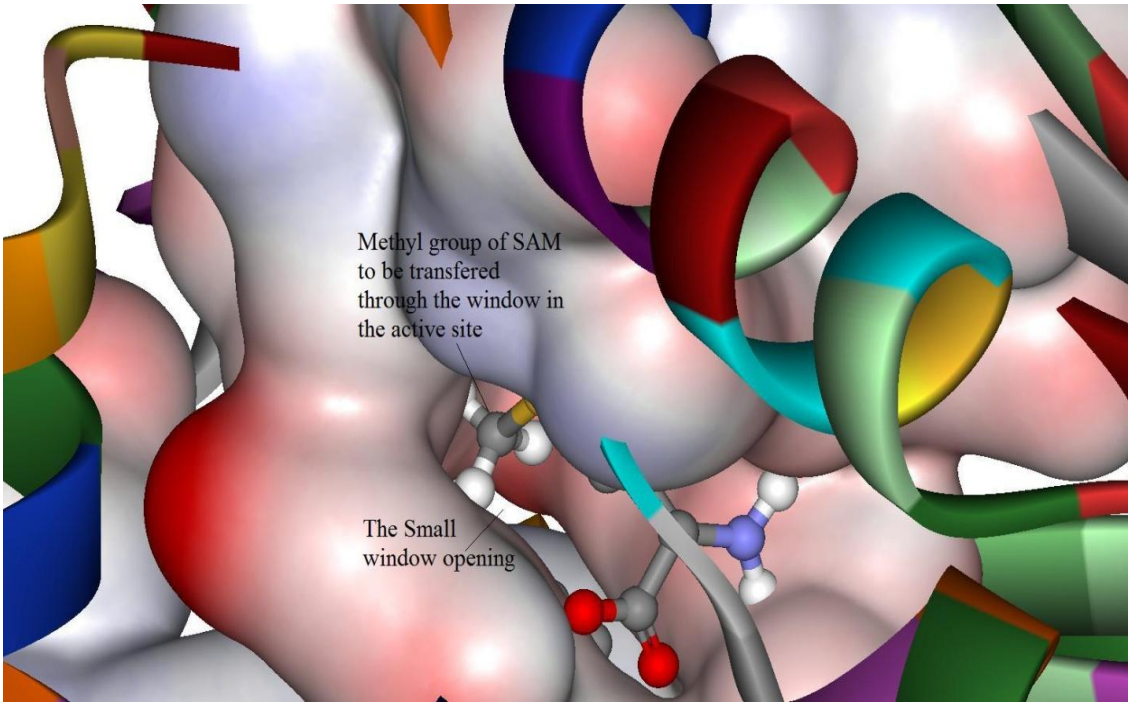


Fig. 3.24c: The Small opening through which the methyl group of SAM is projecting toward the substrate cavity, where the substrate nitrogen atom can receive the methyl group: The cavities have been made apparent by adding the close surface around the ligand and substrate-binding residues leaving the remaining chain in the solid ribbon conformation.

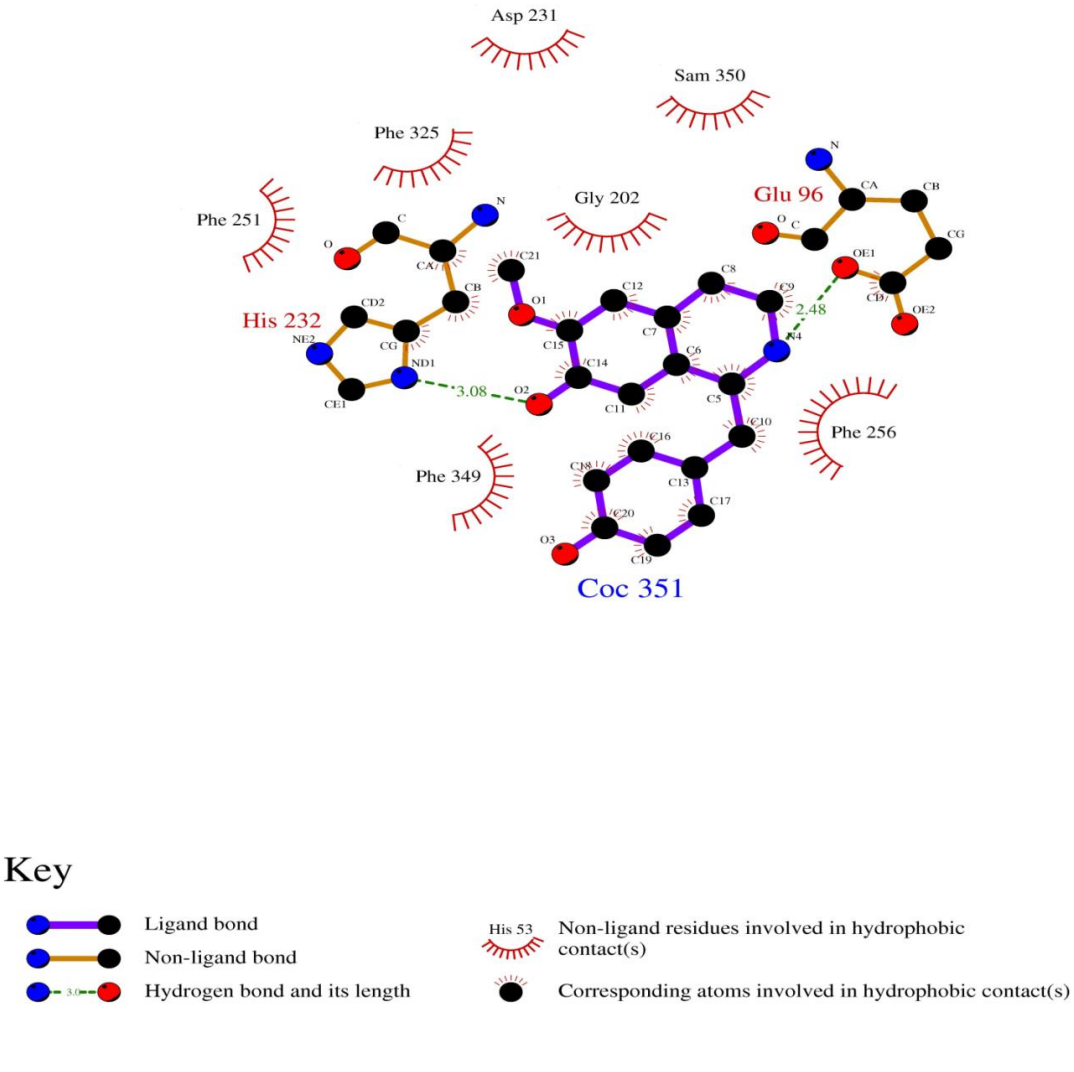


Fig. 3.25a: Ligplot results showing the coclaurine binding residues in the *A. fimbriata* putative CNMT homology model: Atoms have been colored as follows: carbon=black, nitrogen=blue, oxygen=red and sulfur=yellow.

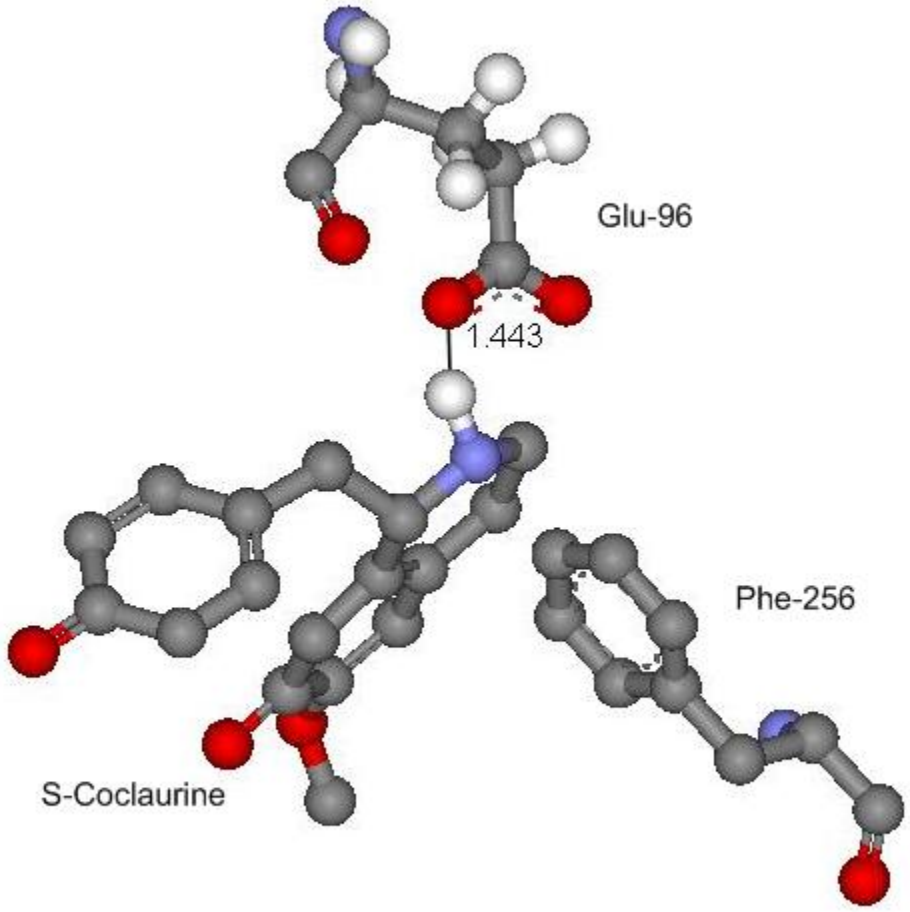


Fig. 3.25b: Ball and stick representation of coclaurine, Glu-96 and Phe-256: The carboxyl O of Glu-96 is positioned to have hydrogen bond interactions with the hydrogen attached to N_4 of coclaurine and might help in deprotonating the N_4 of coclaurine.

Residues, surrounding the ligand and substrate, play key role in this type of transfer reaction. Among these residues Tyr79 [Fig. 3.23b], Phe256 and Glu96 [Fig. 3.25b] might be the most important residues involved in the transfer of the methyl group to the substrate, because these groups are at the contact opening between the pockets of the substrate and SAM. These residues might be involved in inducing charges on the sulfur of SAM and nitrogen of coclaurine through various types of interactions such as hydrogen bonding and Van der Waals interactions.

SAM is highly reactive in donating methyl group. When electrons are donated to the S_D atom of SAM, it looses the methyl group which makes it ready for methylation reactions. On the basis of the arrangement of the conserved residues in the CNMT active site cavity, the location and orientation of the bound SAM inside its cavity, the proposed binding mode of coclaurine, and proposed mechanism for GNMT [41]; the following putative mechanism has been proposed for transmethylation by the putative CNMT.

It has been proposed that SAM gets entered into its active site first producing another cavity nearby for the substrate [41]. As is evident from the predicted model of CNMT, SAM and the substrate are properly oriented in their active sites in CNMT. From their orientation it can be predicted that CNMT catalyzes the methyltransfer reaction by "proximity and orientation effects" just like GNMT does. The coclaurine binds in its active site in a way that the lone pair of electrons on the N_4 of coclaurine is in front of the methyl carbon of SAM. Similarly SAM is oriented in such a way that its methyl carbon projects toward the N_4 of coclaurine. After binding of ligand and substrate, the transfer might happens in a single S_N2 step [42] as was proposed previously for GNMT. Tyr-79 noted earlier could form a charge dipole interaction with the

positively charged S_D atom of SAM, facilitating the methyl transfer reaction in this family of methyltransferases [41] [Fig. 3.23b]. Similarly Glu96 is positioned in such a way that it can take part in the removal of the hydrogen atom from N_4 of coclaurine [Fig. 3.25b].

The OH of Tyr79 has the potential to induce positive charge on S_D atom of SAM. First, the hydroxyl group of Tyr-79 is possibly deprotonated. His-206 may act as a base and has the potential to accept the proton from Tyr-79 [358]. The resulting oxyanion of Tyr-79 may interact with S_D atom of SAM to form Van der Waals interactions. The electrons are attracted from the methyl group toward the S_D atom of SAM due to which the methyl group becomes electrophilic for the lone pair available at N_4 of coclaurine. The bond between S_D atom of SAM and the methyl group is weakened. Interactions between the sulfur atom of SAM and an oxygen atom of an enzyme have been observed in other MTases as well [360], however in some cases the situation is different for example BchU where the Nδ atom of His150 and the sulfur atom of SAM are positioned in close vicinity (3.1 Å) which was proposed to be the first example of the case [358]. In the case of putative CNMT model His-206 is in close proximity to the S_D atom of SAM, but the distance is 5.28Å which is greater than that of BchU. The positive charge induced on S_D atom of SAM also induces positive charge on the carbon atom of the methyl group attached to the S_D atom of SAM. The negatively charged unshared pair of electrons on N_4 atom of coclaurine is attracted by the positively charged methyl group of SAM. In other Mtases e.g. OMTs the substrates have been proposed to be fully or predominantly deprotonated at cellular pH values and should only require their correct positioning for methylation to occur [362]. In case of CNMT, the carboxyl oxygen of the Glu-96 seems to deprotonate the N_4 atom of coclaurine. In the mean time the bond between methyl

group and S_D atom of SAM may break and a new bond might be formed between the methyl group and N_4 atom of coclaurine and the hydrogen seems to be taken by Glu96. The deprotonation of N4 atom with the concomitant cleavage of the C–S bond in SAM and a new bond formation between the methyl group and the N4 atom of coclaurine seems to happen simultaneously and hence follows the S_N2 mechanism. The NMTs have been thought to require some additional help for methylation because nitrogen is not as electronegative as oxygen [40]. The reaction mechanism seems very much identical to that of GNMT. However, this is just a putative mechanism based on other experimentally determined mechanisms; therefore, this mechanism needs to be confirmed experimentally.

CNMT can also catalyze the transfer of methyl group to several other alkaloids structurally similar to (S)-coclaurine such as (R)-coclaurine, (S)-norcoclaurine, (R,S)-6-O-methylcoclaurine, (R, S)-norlaudanosoline and (R, S)-norreticuline etc [1, 6]. It is already known that norlaudanosoline is a precursor in the biosynthesis of aristolochic acid, and N-methylation of norlaudanosoline is an important step [233] in this catalysis. So it can be predicted that the putative CNMT of *A. fimbriata* might be involved in the N-methylation of norlaudanosoline which ultimately leads to the production of aristolochic acid. The mechanism needs to be elucidated experimentally.

3.11. Phylogenetic Analysis

To confirm the identity of the *A. fimbriata* putative CNMT, an NJ consensus tree [Fig. 3.26] based on 1000 bootstrap was generated using a protein multiple sequence alignment [Fig. 3.27], which helped us in finding out the closer relatives of the *A. fimbriata* putative CNMT.

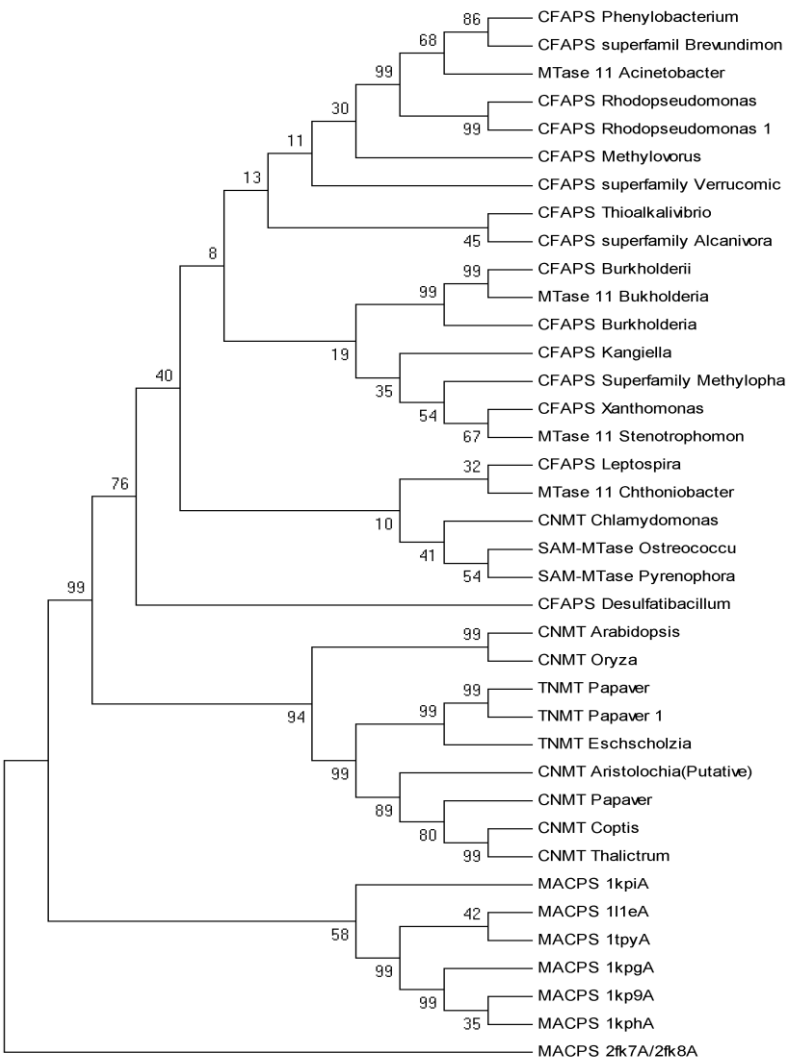
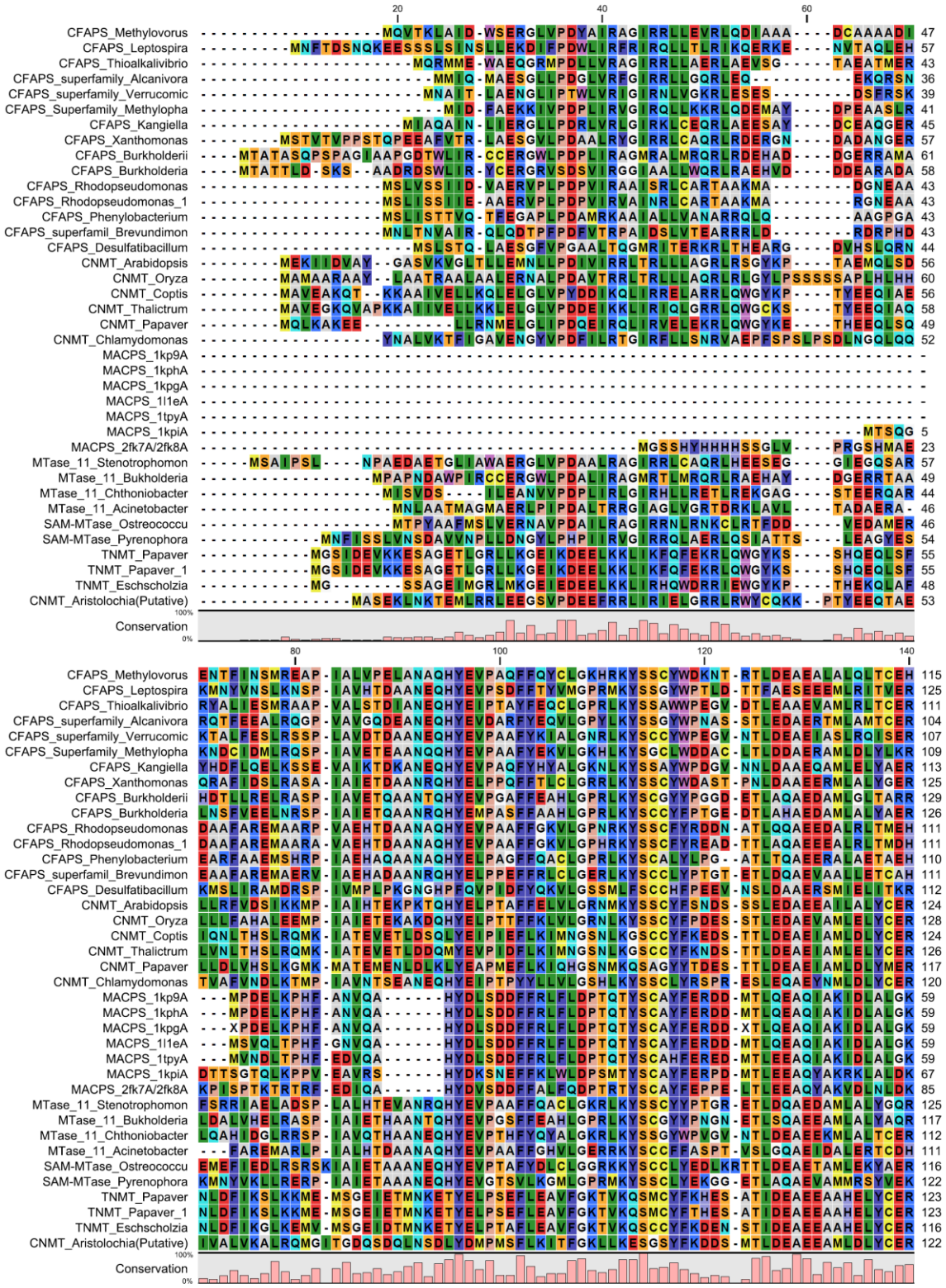
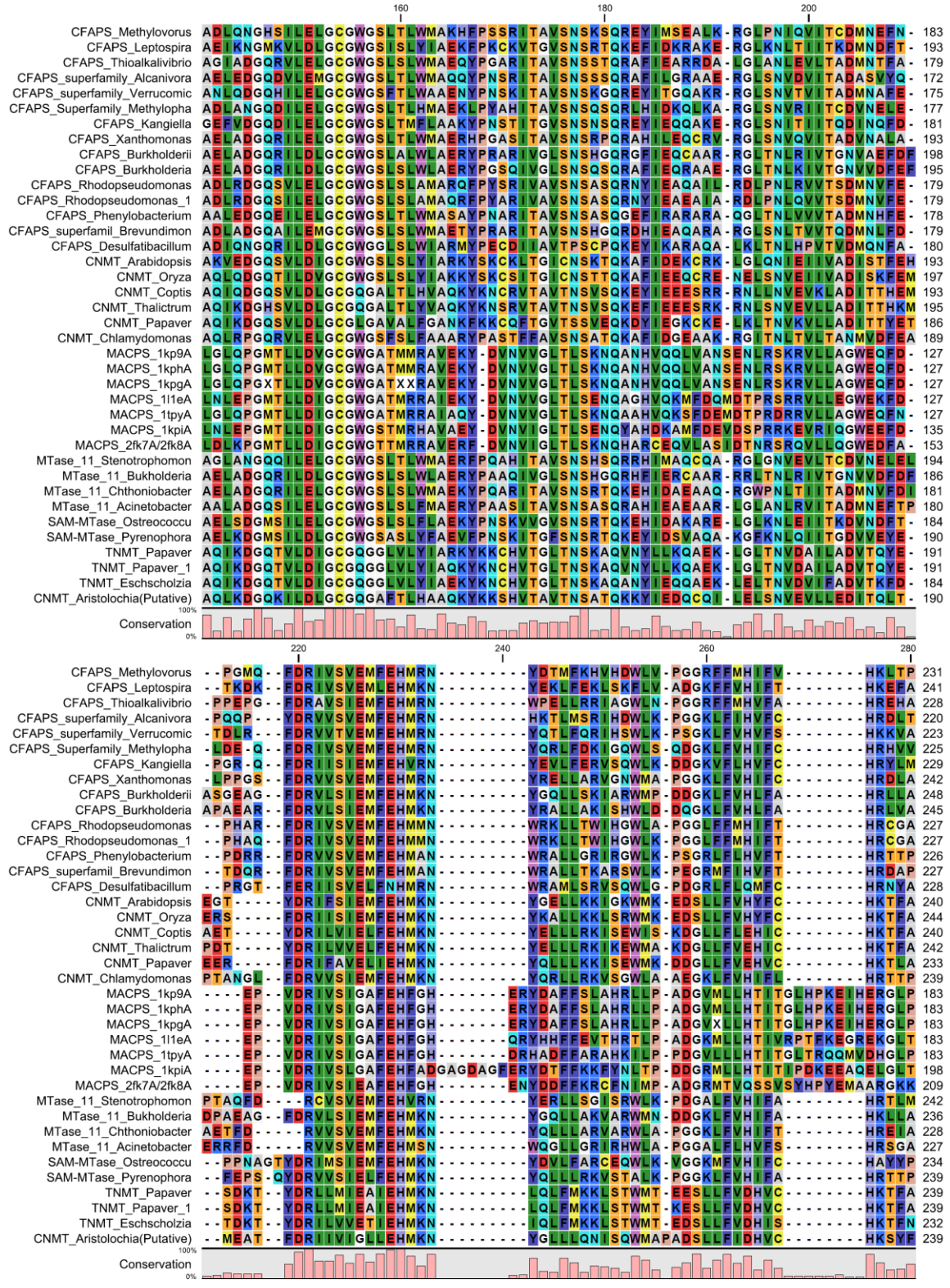


Fig. 3.26: Evolutionary relationships of 38 methyl transferase enzymes: The evolutionary history was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent an estimate of the evolutionary history of the sequences analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated sequences clustered together in the bootstrap test (1000 replicates) are shown next to the branches. All positions containing gaps and missing data were eliminated from the dataset (complete deletion and p-distance option). There were a total of 256 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4.


CFAPS_Methylovorus
CFAPS_Leptospira
CFAPS_Thioalkalivibrio
CFAPS_superfamily_Alcanivora
CFAPS_superfamily_Verrucomic
CFAPS_Superfamily_Methylopha
CFAPS_Kangiella
CFAPS_Xanthomonas
CFAPS_Burkholderii
CFAPS_Burkholderia
CFAPS_Rhodopseudomonas
CFAPS_Rhodopseudomonas_1
CFAPS_Phenylobacterium
CFAPS_superfamil_Brevundimon
CFAPS_Desulfatibacillum
CNMT_Arabidopsis
CNMT_Oryza
CNMT_Coptis
CNMT_Thalictrum
CNMT_Papaver
CNMT_Chlamydomonas
MACPS_1kp9A
MACPS_1kphA
MACPS_1kpgA
MACPS_1l1eA
MACPS_1tpyA
MACPS_1kpiA
MACPS_2fk7A/2fk8A
MTase_11_Stenotrophomon
MTase_11_Bukholderia
MTase_11_Chthoniobacter
MTase_11_Acinetobacter
SAM-MTase_Ostreococcu
SAM-MTase_Pyrenophora
TNMT_Papaver
TNMT_Papaver_1
TNMT_Eschscholzia
CNMT_Aristolochia(Putative)
Conservation
100%
0%


CFAPS_Methylovorus
CFAPS_Leptospira
CFAPS_Thioalkalivibrio
CFAPS_superfamily_Alcanivora
CFAPS_superfamily_Verrucomic
CFAPS_Superfamily_Methylopha
CFAPS_Kangiella
CFAPS_Xanthomonas
CFAPS_Burkholderii
CFAPS_Burkholderia
CFAPS_Rhodopseudomonas
CFAPS_Rhodopseudomonas_1
CFAPS_Phenylobacterium
CFAPS_superfamil_Brevundimon
CFAPS_Desulfatibacillum
CNMT_Arabidopsis
CNMT_Oryza
CNMT_Coptis
CNMT_Thalictrum
CNMT_Papaver
CNMT_Chlamydomonas
MACPS_1kp9A
MACPS_1kphA
MACPS_1kpgA
MACPS_1l1eA
MACPS_1tpyA
MACPS_1kpiA
MACPS_2fk7A/2fk8A
MTase_11_Stenotrophomon
MTase_11_Bukholderia
MTase_11_Chthoniobacter
MTase_11_Acinetobacter
SAM-MTase_Ostreococcu
SAM-MTase_Pyrenophora
TNMT_Papaver
TNMT_Papaver_1
TNMT_Eschscholzia
CNMT_Aristolochia(Putative)
Conservation
100%
0%

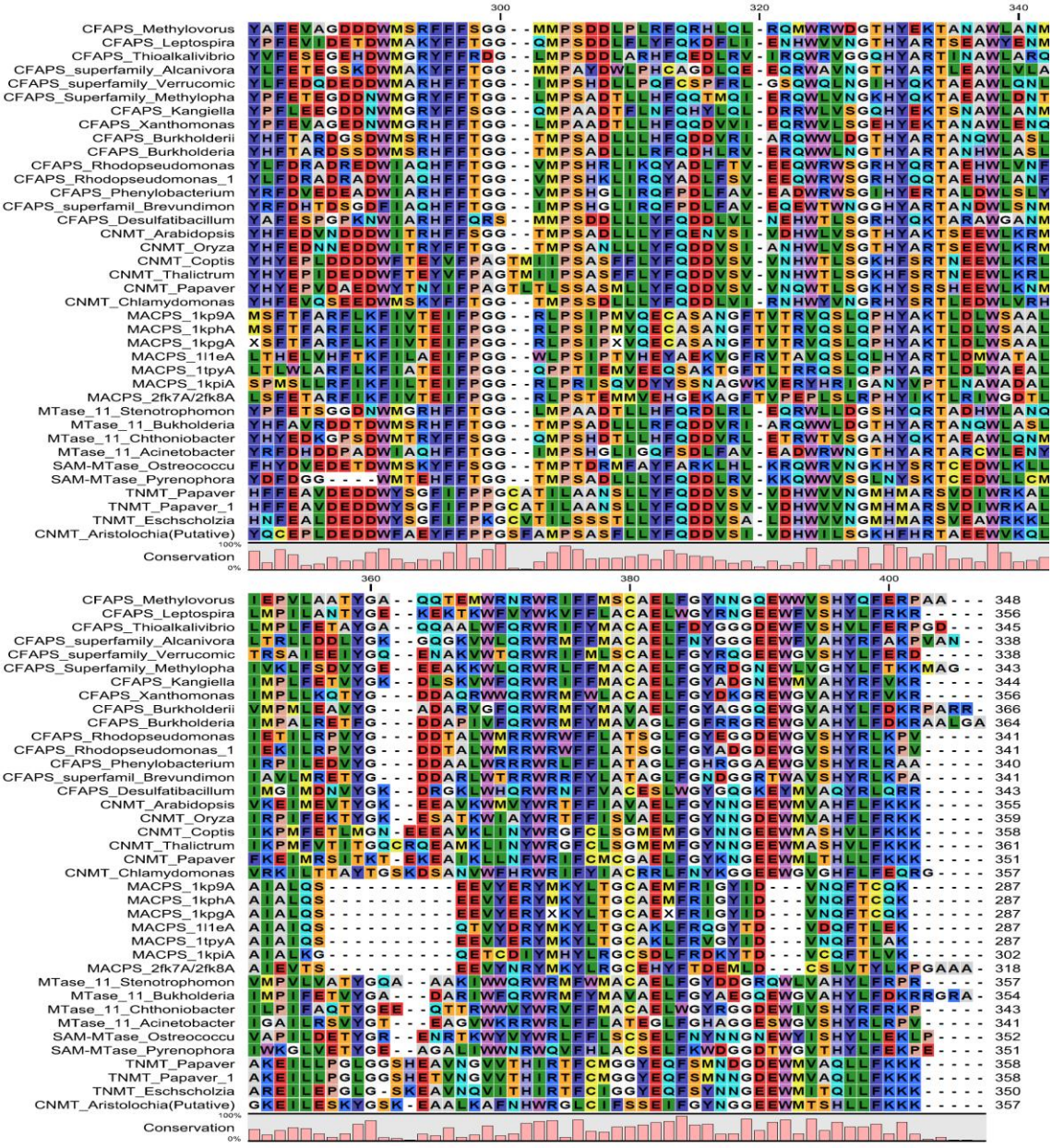


Fig. 3.27: Multiple Alignment of CFAPS (*Methylovorus, Leptospira, Thioalkalivibrio, Alcanivora, Verrucomic, Methylopha, Kangiella, Xanthomonas, Burkholderii, Burkholderia, Rhodopseudomonas, Phenylobacterium, Brevondioun, Desulfatibacillum*), CNMTS (*Arabidopsis, Orysa, Coptis, Thalictrum, Papaver, Chlamydomonas*), MACPS (1kp9A, 1kph, 1kpgA, 1l1eA, 1tpyA, 1kpiA, 2fk7A), Methyltransferase_type-11 (*Stenotrophomon, Burkholderia, Chthoniobacter, Acinetobacter*), SAM-MTases (*Ostreococcu, Pyrenophora*), TNMTs (*Papaver, Eschscholzia*) and *Aristolochia* putative CNMT sequence. The image has been created using CLC-sequence viewer version-6.4 (ww.clcbio.com) with rasmol background color option. The graph shows the percent conservation of the residue.

It also helped us identifying the potential function of the putative CNMT. The tree was rooted with MACPS (2fk7 from *Mycobacterium*). This phylogenetic tree showed that the *A. fimbriata* sequence was similar to the CNMTs from other plants because it has clustered with them, but the cluster was not monophyletic because TNMTs were also clustered with CNMTs which can also mean that TNMTs might be very closely related to CNMTs and they might have evolved from a single ancestral gene. Previous phylogenetic analysis also shows a monophyletic origin of the two N-methyltransferases i.e. CNMT and TNMT [66]. Although TNMTs and CNMTs share limited sequence similarity with cyclopropane synthases, several residues known to be involved in SAM binding in MtPcaA (bacterial SAM-dependent cyclopropane fatty acid synthases) are strictly conserved [66] as explained above. This is probably the reason that they are making the same cluster. The metabolic role also suggests that TNMT has appeared by the duplication of a gene after the more ancient recruitment of CNMT, which is supported by the widespread occurrence of TNMT activity in the Papaveraceae but not in other members of plant families that are involved in the accumulation of benzylisoquinoline alkaloids [66]. The glycine-rich sequence containing Gly-72 and Gly-74 ((E/D)XGXGXG), often referred to as motif I is highly conserved in many SAM-MTases and is also present in TNMT, CNMTs, and MtPcaA [15, 43, 286]. Structure and sequence conservation shows that MtPcaA exhibits a structural fold most similar to the small molecule subclass of methyltransferases [43].

The *A. fimbriata* putative sequence is sister to an orthologous clade of CNMTs of *Papaver, Coptis* and *Thalictrum* and supported with a strong bootstrap value of 89. This reveals that the *A. fimbriata* sequence might be the CNMT sequence and may be involved in transferring of a methyl group from SAM to coclaurine.

All CFAPS, methyltransferase_Type_11 (MTase 11) and SAM-Mtases have clustered together with a moderate bootstrap value of 76. Similarly the MACPS have been clustered into a single monophyletic group with multiple duplication events.

To show the evolutionary relationship of the *A. fimbriata* putative CNMT with CNMTs of other species of known CNMT sequences, the CNMT sequences were aligned [Fig. 3.21c] and an MP tree [Fig. 3.28] was generated. The tree was rooted with *Chlamydomonas* CNMT. The tree gave quite strange results which were different from the results of the proposed species tree generated for basal Angiosperms [363] and APG (Angiosperm Phylogeny Group) III classifications [364]. The tree shows that *A. fimbriata* might be a closer relative of *C. japonica* and *T. flavum* as compared to *A. thaliana* and *O. sativa*. This is clear from the high bootstrap value of *A. fimbriata* CNMT with *C. japonica* and *T. flavum* CNMTs. Although *A. thaliana* is a eudicot but it has clustered with *O. sativa*, a monocot and, in turn *A. fimbriata* which is a magnoliid has been clustered with eudicots. The *A. thaliana* protein seems to be closely related to *O. sativa*, because they have a strong support of 91, but this situation seems quite odd, because they have very far relationship taxonomically (*A. thaliana* is a eudicot while *O. sativa* is a monocot). *C. japonica* and *T. flavum* make a single cluster with high bootstrap support of 100. It seems that *C. japonica* and *T. flavum* might have diverged recently from one another, which is according to their accepted classification both being Ranunculales (eudicot).

If the genes are all considered to be potential orthologs, the gene tree is clearly in conflict with the expected organismal phylogeny. One possible explanation is that the phylogeny is reflecting complex processes of gene duplication and gene loss that are common within gene families.

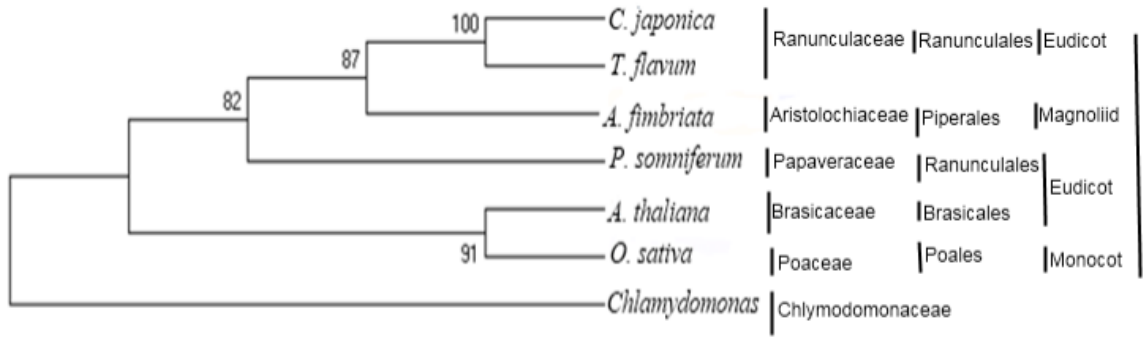


Fig. 3.28: Evolutionary relationships of seven CNMT proteins: The evolutionary history was inferred using the Maximum Parsimony method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the proteins analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated proteins clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The MP tree was obtained using the Close-Neighbor-Interchange algorithm with search level 3 in which the initial trees were obtained with the random addition of sequences (10 replicates). All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 212 positions in the final dataset, out of which 92 were parsimony informative. Phylogenetic analyses were conducted in MEGA4.

In this case, a deeper sampling of related CNMTs may help to clarify the pattern of gene duplications that would be required to obtain this specific tree. Another possibility is horizontal gene transfer, which can cause gene exchanges to occur across sometimes distant phylogenetic boundaries. Horizontal gene transfer has been shown to primarily occur in bacteria [365, 366] however, there are evidences for eukaryotic gene transfer [367] even in higher plants [368], but is not prevalent in multicellular organism [369]. Horizontal gene transfer makes it very difficult to determine evolutionary relationships by analyzing only a single gene or a protein (http://opbs.okstate.edu/~melcher/MG/MGW3/MG334.html). A much deeper sampling of putative CNMTs from many more plants is necessary to distinguish gene family dynamics from possible horizontal gene transfer events.

CONCLUSIONS

CONCLUSIONS

Following conclusion has been made based on the investigation carried out in this study.

1. A cDNA has been cloned and sequenced from the basal angiosperm *A. fimbriata*. From all the BLAST searches performed on the gene sequence as well as on the translated sequence, secondary structure identification results, function prediction results, ModBase searches, CDD search, the phylogenetic analysis and from the templates identified, it was predicted that the gene is involved in the production of CNMT. This might be an indication that such pathways are operating in *A. fimbriata* and hence in basal angiosperm. Basal angiosperms should be investigated for BIA biosynthetic pathways.

2. The predicted CNMT model can be divided into two domains: the N-terminal catalytic core domain and the C-terminal domain. The catalytic domain comprises a central sheet of twisted β-strands surrounded by alpha helices. The catalytic core domain contains the binding sites for SAM and coclaurine.

3. The mechanism of the transfer of methyl has been predicted to be done by "proximity and orientation effects" and is dependent on SAM for its methyl group.

4. CNMT and TNMT might have evolved from a single ancestor as has been previously hypothesized.

5. The *Aristolochia* sequence appears to be more closely related to the basal eudicots *Coptis* and *Thalictrum*, than it is to homologs known from *Oryza* and *Arabidopsis*.

6. *Aristolochia* might have acquired the putative CNMT gene through horizontal transfer from dicots; however *Arabidopsis* might have acquired it from monocots through horizontal transfer. Alternatively, complex gene family dynamics, including gene duplication and loss of members of the CNMT group, might explain the phylogenetic relationships that are obtained.

7. Glu-96 and Tyr-79 might be the most important residues involved in the methylation of coclaurine by *Aristolochia* putative CNMT.

Suggestion: The presence of CNMT in *A. fimbriata* should be confirmed experimentally. The molecule of CNMT should be crystallized and its mechanism of action should be determined through experimental work.

APPENDICES

APPENDIX

Abbreviations used in the alkaloid biosynthesis pathway

Code	Alkaloid name	Code	Alkaloid name
[1]	L-Tyrosine	**[11b]**	Magnoflorine
[2]	L-Dopa	**[10c]**	Papaverine
[3]	Dopamine	**[10d]**	Amurensine
[4]	4-Hydroxyphenalacetaldehyde	**[10e]**	(S)-Corytuberine
[5]	(S)-Norcoclaurine	**[10f]**	Bractuline
[6]	(S)-Coclaurine	**[10g]**	Laudanine
[7]	(S)-N-Methylcoclaurine	**[11g]**	Laudanosine
[8]	3'-Hydroxy-N-Methylcoclaurine	**[10h]**	(S)-Scoulerine
[9]	(S)-Reticuline	**[11ha]**	(S)-Tetrahydrocolumbamine
[10a]	1, 2-Dehydroreticuline	**[12ha]**	(S)-Canadine
[11a]	(R)-Reticuline	**[13ha]**	Berberine
[12a]	Salutaridine	**[11hb]**	Noscapine
[13a]	Salutaridinol	**[11hc]**	(S)-Cheilanthifoline
[14a]	Salutaridinol-7-O-acetate	**[12hc]**	(S)-Stylopine
[15a]	Thebaine	**[13hc]**	(S)-cis-N-Methylstylopine
[16aa]	Neopinone	**[14hc]**	Protopine
[17aa]	Codeinone	**[15hca]**	Rhoeadine
[18aa]	Codeine	**[15hcb]**	6-Hydroxyprotopine
[19aa]	Morphine	**[16hcb]**	Dihydrosanguinarine
[16ab]	Oripavine	**[17hcba]**	Sanguinarine
[17ab]	Morphinone	**[17hcbb]**	Dihydrochelirubine
[10b]	(S)-Coritubrine	**[18hcbb]**	Chelirubine

REFERENCES

REFERENCES

1. Choi, K., Morishige, T. and Sato, F.: **Purification and Characterization of Coclaurine N-Methyltransferase from Cultured *Coptis japonica* Cells**. *Phytochemistry*, 56, 649-655, (2001).

2. Frenzel, T. and Zenk, M.: **Purification and Characterization of Three Isoforms of S-Adenosyl-L-Methionine: (R, S)-Tetrahydrobenzylisoquinoline-N-Methyltransferase from *Berberis koetineana* Cell Cultures**. *Phytochemistry*, 29, 3491-3497, (1990).

3. Wat, C. K., Steffens, P. and Zenk, M. H.: **Partial-Purification and Characterization of S-Adenosyl-L-Methionine - Norreticuline N-Methyltransferases from *Berberis* Cell-Suspension Cultures**. *Zeitschrift Fur Naturforschung C-a Journal of Biosciences*, 41, 126-134, (1986).

4. Morishige, T., Tsujita, T., Yamada, Y. and Sato, F.: **Molecular Characterization of the S-Adenosyl-L-Methionine: 3'-Hydroxy-N-Methylcoclaurine 4'-O-Methyltransferase Involved in Isoquinoline Alkaloid Biosynthesis in *Coptis japonica***. *Journal of Biological Chemistry*, 275, 23398-23405, (2000).

5. Stadler, R. and Zenk, M. H.: **The Purification and Characterization of a Unique Cytochrome-P-450 Enzyme from *Berberis stolonifera* Plant-Cell Cultures**. *Journal of Biological Chemistry*, 268, 823-831, (1993).

6. Loeffler, S., Deusneumann, B. and Zenk, M. H.: **S-Adenosyl-L-Methionine-(S)-Coclaurine-N-Methyltransferase from *Tinospora cordifolia***. *Phytochemistry*, 38, 1387-1395, (1995).

7. Ibrahim, R. K., Bruneau, A. and Bantignies, B.: **Plant O-Methyltransferases: Molecular Analysis, Common Signature and Classification**. *Plant Molecular Biology*, 36, 1-10, (1998).

8. Takeshita, N., Fujiwara, H., Mimura, H., Fitchen, J. H., Yamada, Y. and Sato, F.: **Molecular-Cloning and Characterization of S-Adenosyl-L-Methionine-Scoulerine-9-O-Methyltransferase from Cultured-Cells of *Coptis-japonica***. *Plant and Cell Physiology*, 36, 29-36, (1995).

9. Ibrahim, R. K.: **Plant O-methyltransferase Signatures**. *Trends in Plant Science*, 2, 249-250, (1997).

10. Zubieta, C., He, X. Z., Dixon, R. A. and Noel, J. P.: **Structures of Two Natural Product Methyltransferases Reveal the Basis for Substrate Specificity in Plant O-Methyltransferases**. *Nature Structural Biology*, 8, 271-279, (2001).

11. Zubieta, C., Kota, P., Ferrer, J. L., Dixon, R. A. and Noel, J. P.: **Structural Basis for the Modulation of Lignin Monomer Methylation by Caffeic Acid/5-Hydroxyferulic Acid 3/5-O-Methyltransferase**. *Plant Cell*, 14, 1265-1277, (2002).

12. Ferrer, J. L., Zubieta, H., Dixon, R. A. and Noel, J. P.: **Crystal Structures of *Alfalfa* Caffeoyl Coenzyme A 3-O-Methyltransferase**. *Plant Physiology*, 137, 1009-1017, (2005).

13. Anaya, A. L., Cruz-Ortega, R. and Waller, G. R.: **Metabolism and Ecology of Purine Alkaloids**. *Frontiers in Bioscience*, 11, 2354-2370, (2006).

14. Suzuki, K., Yamada, Y. and Hashimoto, T.: **Expression of *Atropa belladonna* Putrescine N-Methyltransferase Gene in Root Pericycle**. *Plant and Cell Physiology*, 40, 289-297, (1999).

15. Choi, K. B., Morishige, T., Shitan, N., Yazaki, K. and Sato, F.: **Molecular Cloning and Characterization of Coclaurine N-Methyltransferase from Cultured Cells of *Coptis japonica***. *Journal of Biological Chemistry*, 277, 830-835, (2002).

16. Facchini, P. J. and Park, S. U.: **Developmental and Inducible Accumulation of Gene Transcripts Involved in Alkaloid Biosynthesis in Opium Poppy**. *Phytochemistry*, 64, 177-186, (2003).

17. Samanani, N., Park, S. U. and Facchini, P. J.: **Cell Type-Specific Localization of Transcripts Encoding Nine Consecutive Enzymes Involved in Protoberberine Alkaloid Biosynthesis**. *Plant Cell*, 17, 915-926, (2005).

18. Rueffer, M., Zumstein, G. and Zenk, M. H.: **Partial-Purification and Properties of S-Adenosyl-L-Methionine - (S)-Tetrahydroprotoberberine-Cis-N-Methyltransferase from Suspension-Cultured Cells of *Eschscholtzia* and *Corydalis***. *Phytochemistry*, 29, 3727-3733, (1990).

19. Okeefe, B. R. and Beecher, C. W. W.: **Isolation and Characterization of S-Adenosyl-L-Methionine-Tetrahydroberberine-Cis-N-Methyltransferase from Suspension-Cultures of *Sanguinaria-canadensis* L**. *Plant Physiology*, 105, 395-403, (1994).

20. Jansen, R. K., Cai, Z., Raubeson, L. A., Daniell, H., Depamphilis, C. W., Leebens-Mack, J., Muller, K. F., Guisinger-Bellian, M., Haberle, R. C., Hansen, A. K., Chumley, T. W., Lee, S. B., Peery, R., McNeal, J. R., Kuehl, J. V. and Boore, J. L.: **Analysis of 81 Genes from 64 Plastid Genomes Resolves Relationships in Angiosperms and Identifies Genome-Scale**

Evolutionary Patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 19369-19374, (2007).

21. Agrawal, G., Abe, K., Yamazaki, M., Miyao, A. and Hirochika, H.: **Conservation of the E-Function for Floral Organ Identity in Rice Revealed by the Analysis of Tissue Culture-Induced Loss-of-Function Mutants of the OsMADS1 Gene**. *Plant Molecular Biology*, 59, 125-135, (2005).
22. Kramer, E. M., Dorit, R. L. and Irish, V. F.: **Molecular Evolution of Genes Controlling Petal and Stamen Development: Duplication and Divergence within the APETALA3 and PISTILLATA MADS-Box Gene Lineages**. *Genetics*, 149, 765-783, (1998).
23. Ma, H. and dePamphilis, C.: **The ABCs of Floral Evolution**. *Cell*, 101, 5-8, (2000).
24. Zahn, L. M., King, H. Z., Leebens-Mack, J. H., Kim, S., Soltis, P. S., Landherr, L. L., Soltis, D. E., dePamphilis, C. W. and Ma, H.: **The Evolution of the SEPALLATA Subfamily of MADS-Box Genes: A Preangiosperm Origin with Multiple Duplications Throughout Angiosperm History**. *Genetics*, 169, 2209-2223, (2005).
25. Jaramillo, M. A. and Kramer, E. M.: **APETALA3 and PISTILLATA Homologs Exhibit Novel Expression Patterns in the Unique Perianth of *Aristolochia* (Aristolochiaceae)**. *Evolution & Development*, 6, 449-458, (2004).
26. Kim, S., Soltis, P. S., Wall, K. and Soltis, D. E.: **Phylogeny and Domain Evolution in the APETALA2-like Gene Family**. *Molecular Biology and Evolution*, 23, 107-120, (2006).

27. Bell, C. D., Soltis, D. E. and Soltis, P. S.: **The Age of the Angiosperms: A Molecular Timescale Without a Clock**. *Evolution*, 59, 1245-1258, (2005).

28. Leebens-Mack, J., Raubeson, L. A., Cui, L. Y., Kuehl, J. V., Fourcade, M. H., Chumley, T. W., Boore, J. L., Jansen, R. K. and dePamphilis, C. W.: **Identifying the Basal Angiosperm Node in Chloroplast Genome Phylogenies: Sampling One's Way out of the Felsenstein Zone**. *Molecular Biology and Evolution*, 22, 1948-1963, (2005).

29. Bliss, B. J., Landherr, L., dePamphilis, C. W., Ma, H., Hu, Y. and Maximova, S. N.: **Regeneration and Plantlet Development from Somatic Tissues of *Aristolochia fimbriata***. *Plant Cell Tissue and Organ Culture*, 98, 105-114, (2009).

30. Domingues, F. S., Koppensteiner, W. A. and Sippl, M. J.: **The Role of Protein Structure in Genomics**. *Febs Letters*, 476, 98-102, (2000).

31. Skolnick, J., Fetrow, J. S. and Kolinski, A.: **Structural Genomics and its Importance for Gene Function Analysis**. *Nature Biotechnology*, 18, 283-287, (2000).

32. Deshpande, N., Addess, K. J., Bluhm, W. F., Merino-Ott, J. C., Townsend-Merino, W., Zhang, Q., Knezevich, C., Xie, L., Chen, L., Feng, Z. K., Green, R. K., Flippen-Anderson, J. L., Westbrook, J., Berman, H. M. and Bourne, P. E.: **The RCSB Protein Data Bank: A Redesigned Query System and Relational Database Based on the mmCIF Schema**. *Nucleic Acids Research*, 33, D233-D237, (2005).

33. Bairoch, A., Consortium, U., Bougueleret, L., Altairac, S., Amendolia, V., Auchincloss, A., Argoud-Puy, G., Axelsen, K., Baratin, D., Blatter, M. C., Boeckmann, B., Bolleman, J., Bollondi, L., Boutet, E., Quintaje, S. B., Breuza,

L., Bridge, A., Decastro, E., Ciapina, L., Coral, D., Coudert, E., Cusin, I., Delbard, G., Dornevil, D., Roggli, P. D., Duvaud, S., Estreicher, A., Famiglietti, L., Feuermann, M., Gehant, S., Farriol-Mathis, N., Ferro, S., Gasteiger, E., Gateau, A., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Hulo, N., James, J., Jimenez, S., Jungo, F., Junker, V., Kappler, T., Keller, G., Lachaize, C., Lane-Guermonprez, L., Langendijk-Genevaux, P., Lara, V., Lemercier, P., Le Saux, V., Lieberherr, D., Lima, T. D., Mangold, V., Martin, X., Masson, P., Michoud, K., Moinat, M., Morgat, A., Mottaz, A., Paesano, S., Pedruzzi, I., Phan, I., Pilbout, S., Pillet, V., Poux, S., Pozzato, M., Redaschi, N., Reynaud, S., Rivoire, C., Roechert, B., Schneider, M., Sigrist, C., Sonesson, K., Staehli, S., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Veuthey, A. L., Yip, L., Zuletta, L., Apweiler, R., Alam-Faruque, Y., Antunes, R., Barrell, D., Binns, D., Bower, L., Browne, P., Chan, W. M., Dimmer, E., Eberhardt, R., Fedotov, A., Foulger, R., Garavelli, J., Golin, R., Horne, A., Huntley, R., Jacobsen, J., Kleen, M., Kersey, P., Laiho, K., Leinonen, R., Legge, D., Lin, Q., Magrane, M., Martin, M. J., O'Donovan, C., Orchard, S., O'Rourke, J., Patient, S., Pruess, M., Sitnov, A., Stanley, E., Corbett, M., di Martino, G., Donnelly, M., Luo, J., van Rensburg, P., Wu, C., Arighi, C., Arminski, L., Barker, W., Chen, Y. X., Hu, Z. Z., Hua, H. K., Huang, H. Z., Mazumder, R., McGarvey, P., Natale, D. A., Nikolskaya, A., Petrova, N., Suzek, B. E., Vasudevan, S., Vinayaka, C. R., Yeh, L. S. and Zhang, J.: **The Universal Protein Resource (UniProt) 2009**. *Nucleic Acids Research*, 37, D169-D174, (2009).

34. Kiefer, F., Arnold, K., Kunzli, M., Bordoli, L. and Schwede, T.: **The SWISS-MODEL Repository and Associated Resources**. *Nucleic Acids Research*, 37, D387-D392, (2009).

35. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. and Sayers, E. W.: **GenBank**. *Nucleic Acids Research*, 37, D26-D31, (2009).

36. Baker, D. and Sali, A.: **Protein Structure Prediction and Structural Genomics**. *Science*, 294, 93-96, (2001).

37. Bonneau, R. and Baker, D.: **Ab Initio Protein Structure Prediction: Progress and Prospects**. *Annual Review of Biophysics and Biomolecular Structure*, 30, 173-189, (2001).

38. Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F. and Sali, A.: **Comparative Protein Structure Modeling of Genes and Genomes**. *Annual Review of Biophysics and Biomolecular Structure*, 29, 291-325, (2000).

39. Ta, H. M. and Kim, K. K.: **Crystal Structure of *Streptococcus pneumoniae* Sp1610, a Putative tRNA Methyltransferase, in Complex with S-Adenosyl-L-Methionine**. *Protein Science*, 19, 617-624, (2010).

40. McCarthy, A. A. and McCarthy, J. G.: **The Structure of Two N-Methyltransferases from the Caffeine Biosynthetic Pathway**. *Plant Physiology*, 144, 879-889, (2007).

41. Takata, Y., Huang, Y. F., Komoto, J., Yamada, T., Konishi, K., Ogawa, H., Gomi, T., Fujioka, M. and Takusagawa, F.: **Catalytic Mechanism of Glycine N-Methyltransferase**. *Biochemistry*, 42, 8394-8402, (2003).

42. Coward, J. K.: **Chemical Mechanisms of Methyl Transfer Reactions: Comparison of Methylases with Nonenzymic 'Model Reactions'**. in *The*

Biochemistry of Adenosylmethionine : [Proceedings of an International Symposium on the Biochemistry of Adenosylmethionine, Sponsored by the Accademia Nazionale Dei Lincei, Held in Rome, Italy May 21-26, 1974] (ed Salvatore, F.) 127-144 (Columbia University Press, New York, 1977).

43. Martin, J. L. and McMillan, F. M.: **SAM (Dependent) I AM: the S-Adenosylmethionine-Dependent Methyltransferase Fold**. *Current Opinion in Structural Biology*, 12, 783-793, (2002).

44. Miller, D. J., Ouellette, N., Evdokimova, E., Savchenko, A., Edwards, A. and Anderson, W. F.: **Crystal Complexes of a Predicted S-Adenosylmethionine-Dependent Methyltransferase Reveal a Typical AdoMet Binding Domain and a Substrate Recognition Domain**. *Protein Science*, 12, 1432-1442, (2003).

45. Joshi, C. P. and Chiang, V. L.: **Conserved Sequence Motifs in Plant S-Adenosyl-L-Methionine-Dependent Methyltransferases**. *Plant Molecular Biology*, 37, 663-674, (1998).

46. Frick, S. and Kutchan, T. M.: **Molecular Cloning and Functional Expression of O-Methyltransferases Common to Isoquinoline Alkaloid and Phenylpropanoid Biosynthesis**. *Plant Journal*, 17, 329-339, (1999).

47. Schubert, H. L., Blumenthal, R. M. and Cheng, X. D.: **Many Paths to Methyltransfer: a Chronicle of Convergence**. *Trends in Biochemical Sciences*, 28, 329-335, (2003).

48. Waddell, T. G., Eilders, L. L., Patel, B. P. and Sims, M.: **Prebiotic Methylation and the Evolution of Methyl Transfer Reactions in Living Cells**. *Origins of Life and Evolution of the Biosphere*, 30, 539-548, (2000).

49. Thomas, D. J., Waters, S. B. and Styblo, M.: **Elucidating the Pathway for Arsenic Methylation**. *Toxicology and Applied Pharmacology*, 198, 319-326, (2004).

50. Wuosmaa, A. M. and Hager, L. P.: **Methyl-Chloride Transferase - a Carbocation Route for Biosynthesis of Halometabolites**. *Science*, 249, 160-162, (1990).

51. Saxena, D., Aouad, S., Attieh, J. and Saini, H. S.: **Biochemical Characterization of Chloromethane Emission from the Wood-Rotting Fungus *Phellinus pomaceus***. *Applied and Environmental Microbiology*, 64, 2831-2835, (1998).

52. Anantharaman, V., Koonin, E. V. and Aravind, L.: **Comparative Genomics and Evolution of Proteins Involved in RNA Metabolism**. *Nucleic Acids Research*, 30, 1427-1464, (2002).

53. Hopper, A. K. and Phizicky, E. M.: **tRNA Transfers to the Limelight**. *Genes & Development*, 17, 162-180, (2003).

54. Kouzarides, T.: **Histone Methylation in Transcriptional Control**. *Current Opinion in Genetics & Development*, 12, 198-209, (2002).

55. Morishige, T., Dubouzet, E., Choi, K. B., Yazaki, K. and Sato, F.: **Molecular Cloning of Columbamine O-Methyltransferase from Cultured *Coptis japonica* Cells**. *European Journal of Biochemistry*, 269, 5659-5667, (2002).

56. Ounaroon, A., Decker, G., Schmidt, J., Lottspeich, F. and Kutchan, T. M.: **(R,S)-Reticuline 7-O-Methyltransferase and (R,S)-Norcoclaurine 6-O-Methyltransferase of *Papaver somniferum* - cDNA Cloning and Characterization of Methyl Transfer Enzymes of Alkaloid Biosynthesis in Opium Poppy**. *Plant Journal*, 36, 808-819, (2003).

57. Dittrich, H. and Kutchan, T. M.: **Molecular-Cloning, Expression, and Induction of Berberine Bridge Enzyme, an Enzyme Essential to the Formation of Benzophenanthridine Alkaloids in the Response of Plants to Pathogenic Attack**. *Proceedings of the National Academy of Sciences of the United States of America*, 88, 9969-9973, (1991).

58. Ikezawa, N., Tanaka, M., Nagayoshi, M., Shinkyo, R., Sakaki, T., Inouye, K. and Sato, F.: **Molecular Cloning and Characterization of CYP719, a Methylenedioxy Bridge-forming Enzyme that Belongs to a Novel P450 Family, from Cultured *Coptis japonica* Cells**. *Journal of Biological Chemistry*, 278, 38557-38565, (2003).

59. Grothe, T., Lenz, R. and Kutchan, T. M.: **Molecular Characterization of the Salutaridinol 7-O-acetyltransferase Involved in Morphine Biosynthesis in Opium Poppy *Papaver somniferum***. *Journal of Biological Chemistry*, 276, 30717-30723, (2001).

60. Unterlinner, B., Lenz, R. and Kutchan, T. M.: **Molecular Cloning and Functional Expression of Codeinone Reductase: The Penultimate Enzyme in Morphine Biosynthesis in the Opium Poppy *Papaver somniferum***. *Plant Journal*, 18, 465-475, (1999).

61. Liscombe, D. K., MacLeod, B. P., Loukanina, N., Nandi, O. I. and Facchini, P. J.: **Evidence for the Ponophyletic Evolution of Benzylisoquinoline Alkaloid Biosynthesis in Angiosperms (vol 66, pg 1374, 2005)**. *Phytochemistry*, 66, 2500-2520, (2005).

62. Minami, H., Dubouzet, E., Iwasa, K. and Sato, F.: **Functional Analysis of Norcoclaurine Synthase in *Coptis japonica***. *Journal of Biological Chemistry*, 282, 6274-6282, (2007).

63. Ziegler, J., Voigtlander, S., Schmidt, J., Kramell, R., Miersch, O., Ammer, C., Gesell, A. and Kutchan, T. M.: **Comparative Transcript and Alkaloid Profiling in *Papaver* Species Identifies a Short Chain Dehydrogenase/Reductase Involved in Morphine Biosynthesis**. *Plant Journal*, 48, 177-192, (2006).

64. Fisinger, U., Grobe, N. and Zenk, M. H.: **Thebaine Synthase: A New Enzyme in the Morphine Pathway in *Papaver somniferum***. *Natural Product Communications*, 2, 249-253, (2007).

65. Ikezawa, N., Iwasa, K. and Sato, F.: **Molecular Cloning and Characterization of Methylenedioxy Bridge-forming Enzymes Involved in Stylopine Biosynthesis in *Eschscholzia californica***. *FEBS Journal*, 274, 1019-1035, (2007).

66. Liscombe, D. K. and Facchini, P. J.: **Molecular Cloning and Characterization of Tetrahydroprotoberberine cis-N- methyltransferase, an Enzyme Involved in Alkaloid Biosynthesis in Opium Poppy**. *Journal of Biological Chemistry*, 282, 14741-14751, (2007).

67. Weiss, D., Baumert, A., Vogel, M. and Roos, W.: **Sanguinarine Reductase, a Key Enzyme of Benzophenanthridine Detoxification**. *Plant Cell and Environment*, 29, 291-302, (2006).

68. Ikezawa, N., Iwasa, K. and Sato, F.: **Molecular Cloning and Characterization of CYP80G2, a Cytochrome p450 that Catalyzes an Intramolecular C - C Phenol Coupling of (S)-reticuline in Magnoflorine Biosynthesis, from Cultured *Coptis japonica* Cells**. *Journal of Biological Chemistry*, 283, 8810-8821, (2008).

69. Kutchan, T. M.: **Alkaloid Biosynthesis - the Basis for Metabolic Engineering of Medicinal-Plants**. *Plant Cell*, 7, 1059-1070, (1995).

70. Corteau, R., Kutchan, T. M. and Lewis, N. G.: **Natural Products (Secondary Metabolites). In: Biochemistry & Molecular Biology of Plants**. in *American Society of Plant Physiologists* (eds. Buchanan, B., Gruissem, W. and Jones, R.) (Rockville, Md., 2000).

71. Ober, D. and Hartmann, T.: **Homospermidine Synthase, the First Pathway-Specific Enzyme of Pyrrolizidine Alkaloid Biosynthesis, Evolved from Deoxyhypusine Synthase**. *Proceedings of the National Academy of Sciences of the United States of America*, 96, 14777-14782, (1999).

72. Wildung, M. and Croteau, R. B.: **Genetic Engineering of Peppermint for Improved Essential Oil Composition and Yield**. *Transgenic Research*, 14, 365-372, (2005).

73. Robinson, T.: **The Biochemistry of Alkaloids**. in *Molecular Biology, Biochemistry, and Biophysics* (Springer-Verlag, Berlin, Germany, 1981).

74. Robinson, T.: in *Herbivores, Their Interaction with Secondary Plant Metabolites* (eds. Rosenthal, G. and AJanzen, D. H.) 413-448 (Academic Press, New York, 1979).

75. Neuhann, H., Leienbach, K. W. and Barz, W.: **Metabolism of Nicotinic-Acid in Plant-Cell Suspension Cultures .5. Degradation of Nicotinamide Adenine-Dinucleotide in Cell-Suspension Cultures**. *Phytochemistry*, 18, 61-64, (1979).

76. Fodor, G.: **The Tropane Alkaloids**. in *The Alkaloids: Chemistry and Physiology*. (ed Manske, R. H. F.) 351-396, (Academic Press. New York, New York, 1971).

77. Henry, T. A.: **The Tropane Alkaloids**. in *The Plant Alkaloids.* (ed Henry, T. A.) (JA Churchill Ltd, London, 1949).

78. Sneader, W.: **Plant Products Analogues and Compounds Derived from Them**. in *Drug Discovery : A History* (ed Sneader, W.) 115-150 (Wiley & Sons, Chichester, 2005).

79. Griffin, W. J. and Lin, G. D.: **Chemotaxonomy and Geographical Distribution of Tropane Alkaloids**. *Phytochemistry*, 53, 623-637, (2000).

80. Newman, D. J., Cragg, G. M. and Snader, K. M.: **Natural Products as Sources of New Drugs Over the Period 1981-2002**. *Journal of Natural Products*, 66, 1022-1037, (2003).

81. Roberts, M. F. and Wink, M.: in *Alkaloids : Biochemistry, Ecology, and Medicinal Applications* 2-3 (Plenum Press, New York, 1998).

82. Hesse, M.: in *Alkaloids: Nature's Curse or Blessing?* 11-114 (Plenum Press, ZèurichWeinheim ; New York, 2002).

83. Kutchan, T. M.: **Molecular Genetics of Plant Alkaloid Biosynthesis**. in *The Alkaloids-Chemistry and Biology* (ed Cordell, G. A.) 257-316 (Academic Press, San Diego, 1998).

84. Kutchan, T. M., Frick, S. and Weid, M.: **Engineering Plant Alkaloid Biosynthetic Pathways – Progress and Prospects.** in *Advances in Plant Biochemistry and Molecular Biology* (eds. Bohnert, H. J. and Nguyen, H. T.) 281-308 (Elsevier Science Ltd, Oxford, 2007).

85. Wink, M.: **Plant-Breeding - Importance of Plant Secondary Metabolites for Protection Against Pathogens and Herbivores**. *Theoretical and Applied Genetics*, 75, 225-233, (1988).

86. Facchini, P. J.: **Alkaloid Biosynthesis in Plants: Biochemistry, Cell Biology, Molecular Regulation, and Metabolic Engineering Applications**. *Annual Review of Plant Physiology and Plant Molecular Biology*, 52, 29-66, (2001).

87. Preininger, V.: **Chemotaxonomy of Papaveraceae and Fumariaceae**. in *The Alkaloids (Chemistry and Pharmacology)* (ed Brossi, A.) 1-98 (Academic Press, Orlando, San Diego, New York, Austin, Boston, London, Sydney, Tokyo and Toronto, 1986).

88. Schmeller, T., LatzBruning, B. and Wink, M.: **Biochemical Activities of Berberine, Palmatine and Sanguinarine Mediating Chemical Defence Against Microorganisms and Herbivores**. *Phytochemistry*, 44, 257-266, (1997).

89. Paraskevas, S.: **Randomized Controlled Clinical Trials on Agents Used for Chemical Plaque Control**. *International Journal of Dental Hygiene*, 3, 162-78, (2005).

90. Hider, R. C., Walkinshaw, M. D. and Saenger, W.: **Erythrina Alkaloid Nicotinic Antagonists - Structure-Activity-Relationships**. *European Journal of Medicinal Chemistry*, 21, 231-234, (1986).

91. Hung, T. M., Na, M. K., Min, B. S., Zhang, X. F., Lee, I. S., Ngoc, T. M., Thuong, P. T., Sok, D. E. and Bae, K. H.: **Protective Effect of Magnoflorine Isolated from *Coptidis rhizoma* on Cu2+-Induced Oxidation of Human Low Density Lipoprotein**. *Planta Medica*, 73, 1281-1284, (2007).

92. Hung, T. M., Lee, J. P., Min, B. S., Choi, J. S., Na, M. K., Zhang, X. F., Ngoc, T. M., Lee, I. and Bae, K. H.: **Magnoflorine from *Coptidis rhizoma* Protects High Density Lipoprotein During Oxidant Stress**. *Biological & Pharmaceutical Bulletin*, 30, 1157-1160, (2007).

93. Rashid, M., Gustafson, K., Kashman, Y., Cardellina, J., McMahon, J. and Boyd, M.: **Anti-HIV Alkaloids from *Toddalia asiatica***. *Natural Product Research*, 6, 153-156, (1995).

94. Kong, W. J., Wei, J., Abidi, P., Lin, M. H., Inaba, S., Li, C., Wang, Y. L., Wang, Z. Z., Si, S. Y., Pan, H. N., Wang, S. K., Wu, J. D., Wang, Y., Li, Z. R., Liu, J. W. and Jiang, J. D.: **Berberine is a Novel Cholesterol-lowering Drug Working Through a Unique Mechanism Distinct from Statins**. *Nature Medicine*, 10, 1344-1351, (2004).

95. Zenk, M. H.: **The Formation of Benzophenanthridine Alkaloids**. *Pure and Applied Chemistry*, 66, 2023-2028, (1994).

96. Facchini, P. J. and Deluca, V.: **Phloem-Specific Expression of Tyrosine Dopa Decarboxylase Genes and the Biosynthesis of Isoquinoline Alkaloids in Opium Poppy**. *Plant Cell*, 7, 1811-1821, (1995).

97. Bird, D. A., Franceschi, V. R. and Facchini, P. J.: **A Tale of Three Cell Types: Alkaloid Biosynthesis is Localized to Sieve Elements in Opium Poppy**. *Plant Cell*, 15, 2626-2635, (2003).

98. Samanani, N., Alcantara, J., Bourgault, R., Zulak, K. G. and Facchini, P. J.: **The Role of Phloem Sieve Elements and Laticifers in the Biosynthesis and Accumulation of Alkaloids in Opium Poppy**. *Plant Journal*, 47, 547-563, (2006).

99. Pauli, H. H. and Kutchan, T. M.: **Molecular Cloning and Functional Heterologous Expression of Two Alleles Encoding (S)-N-methylcoclaurine 3'-hydroxylase (CYP80B1), a New Methyl Jasmonate-inducible Cytochrome P-450-dependent Mono-oxygenase of Benzylisoquinoline Alkaloid Biosynthesis**. *Plant Journal*, 13, 793-801, (1998).

100. Kraus, P. F. X. and Kutchan, T. M.: **Molecular-Cloning and Heterologous Expression of a cDNA-Encoding Berbamunine Synthase, a C-O Phenol-Coupling Cytochrome-P450 from the Higher-Plant *Berberis-stolonifera***. *Proceedings of the National Academy of Sciences of the United States of America*, 92, 2071-2075, (1995).

101. Nelson, D. R., Schuler, M. A., Paquette, S. M., Werck-Reichhart, D. and Bak, S.: **Comparative Genomics of Rice and *Arabidopsis*. Analysis of 727 Cytochrome P450 Genes and Pseudogenes from a Monocot and a Dicot**. *Plant Physiology*, 135, 756-772, (2004).

102. Nelson, D. R., Zeldin, D. C., Hoffman, S. M. G., Maltais, L. J., Wain, H. M. and Nebert, D. W.: **Comparison of Cytochrome P450 (CYP) Genes from the Mouse and Human Genomes, Including Nomenclature Recommendations for Genes, Pseudogenes and Alternative-splice Variants**. *Pharmacogenetics*, 14, 1-18, (2004).

103. Chapple, C.: **Molecular-Genetic Analysis of Plant Cytochrome P450-Dependent Monooxygenases**. *Annual Review of Plant Physiology and Plant Molecular Biology*, 49, 311-343, (1998).

104. Facchini, P. J. and Deluca, V.: **Differential and Tissue-Specific Expression of a Gene Family for Tyrosine Dopa Decarboxylase in Opium Poppy**. *Journal of Biological Chemistry*, 269, 26684-26690, (1994).

105. Samanani, N., Liscombe, D. K. and Facchini, P. J.: **Molecular Cloning and Characterization of Norcoclaurine Synthase, an Enzyme Catalyzing the First Committed Step in Benzylisoquinoline Alkaloid Biosynthesis**. *Plant Journal*, 40, 302-313, (2004).

106. Ziegler, J., Diaz-Chavez, M., Kramell, R., Ammer, C. and Kutchan, T. M.: **Comparative Macroarray Analysis of Morphine Containing *Papaver somniferum* and Eight Morphine Free *Papaver* Species Identifies an O-Methyltransferase Involved in Benzylisoquinoline Biosynthesis**. *Planta*, 222, 458-471, (2005).

107. Luk, L. Y. P., Bunn, S., Liscombe, D. K., Facchini, P. J. and Tanner, M. E.: **Mechanistic Studies on Norcoclaurine Synthase of Benzylisoquinoline Alkaloid Biosynthesis: An Enzymatic Pictet-Spengler Reaction**. *Biochemistry*, 46, 10153-10161, (2007).

108. Facchini, P., Hagel, J., Liscombe, D., Loukanina, N., MacLeod, B., Samanani, N. and Zulak, K.: **Opium Poppy: Blueprint for an Alkaloid Factory**. *Phytochemistry Reviews*, 6, 97-124, (2007).

109. Ronsch, H.: **Rhoeadine Alkaloids**. in *The Alkaloids* (ed Brossi, A.) 1-93 (Academic Press, San Diego, 1986).

110. Bauer, W. and Zenk, M. H.: **2 Methylenedioxy Bridge Forming Cytochrome-P-450 Dependent Enzymes Are Involved in (S)-Stylopine Biosynthesis**. *Phytochemistry*, 30, 2953-2961, (1991).

111. Bauer, W. and Zenk, M. H.: **Formation of Both Methylenedioxy Groups in the Alkaloid (S)-Stylopine Is Catalyzed by Cytochrome-P-450 Enzymes**. *Tetrahedron Letters*, 30, 5257-5260, (1989).

112. Rueffer, M. and Zenk, M. H.: **Enzymatic Formation of Protopines by a Microsomal Cytochrome-P-450 System of *Corydalis-vaginans***. *Tetrahedron Letters*, 28, 5307-5310, (1987).

113. Tanahashi, T. and Zenk, M. H.: **Elicitor Induction and Characterization of Microsomal Protopine-6-Hydroxylase, the Central Enzyme in**

Benzophenanthridine Alkaloid Biosynthesis. *Phytochemistry*, 29, 1113-1122, (1990).

114. Schumacher, H. M. and Zenk, M. H.: **Partial-Purification and Characterization of Dihydrobenzophenanthridine Oxidase from *Escholtzia-californica* Cell-Suspension Cultures**. *Plant Cell Reports*, 7, 43-46, (1988).

115. Lenz, R. and Zenk, M. H.: **Acetyl Coenzyme A: Salutaridinol-7-O-Acetyltransferase from *Papaver somniferum* Plant Cell Cultures - The Enzyme Catalyzing the Formation of Thebaine in Morphine Biosynthesis**. *Journal of Biological Chemistry*, 270, 31091-31096, (1995).

116. Deeknamkul, W. and Zenk, M. H.: **Purification and Properties of 1,2-Dehydroreticuline Reductase from *Papaver-somniferum* Seedlings**. *Phytochemistry*, 31, 813-821, (1992).

117. Hirata, K., Poeaknapo, C., Schmidt, J. and Zenk, M. H.: **1,2-Dehydroreticuline Synthase, the Branch Point Enzyme Opening the Morphinan Biosynthetic Pathway**. *Phytochemistry*, 65, 1039-1046, (2004).

118. Gerardy, R. and Zenk, M. H.: **Purification and Characterization of Salutaridine - NADPH 7-Oxidoreductase from *Papaver-somniferum***. *Phytochemistry*, 34, 125-132, (1993).

119. Lajide, L., Escoubas, P. and Mizutani, J.: **Antifeedant Activity of Metabolites of Aristolochia-Albida Against the Tobacco Cutworm, *Spodoptera-litura***. *Journal of Agricultural and Food Chemistry*, 41, 669-673, (1993).

120. hui, Z. R. G. w. s. b. y. d. w. y.: **The Committee of the Pharmacopoeia of the Ministry of Heath of the People's Republic of China, Pharmacopoeia**

of the People's Republic of China. (Chemical Industry Press, Beijing, China, 2000).

121. Dehon, B., Chagnon, J. L., Vinner, E., Pommery, J., Mathieu, D. and Lhermitte, M.: **Colchicine Poisoning: Report of a Fatal Case with Body Fluid and Post-Mortem Tissue Analysis by High-Performance Liquid Chromatography**. *Biomedical Chromatography*, 13, 235-238, (1999).

122. Brvar, M., Ploj, T., Kozelj, G., Mozina, M., Noc, M. and Bunc, M.: **Case Report: Fatal Poisoning with *Colchicum autumnale***. *Critical Care*, 8, R56-R59, (2004).

123. Konno, C., Taguchi, T., Tamada, M. and Hikino, H.: **Studies on the Constituents of Ephedra .3. Ephedroxane, Anti-Inflammatory Principle of Ephedra Herbs**. *Phytochemistry*, 18, 697-698, (1979).

124. Stadler, R. and Zenk, M. H.: **A Revision of the Generally Accepted Pathway for the Biosynthesis of the Benzyltetrahydroisoquinoline Alkaloid Reticuline**. *Liebigs Annalen Der Chemie*, 555-562, (1990).

125. Gibbons, S., Craven, L., Dunlop, C., Gray, A. I., Hartley, T. G. and Waterman, P. G.: **The Secondary Metabolites of *Aff samadera* SAC-2825: An Australian Simaroubaceae with Unusual Chemistry**. *Phytochemistry*, 44, 1109-1114, (1997).

126. Skaltsounis, A. L.: **Acridone Alkaloids**. in *The Alkaloids* (ed Cordell, G. A.) 259-377 (Elsevier, The Netherlands, 2000).

127. Basa, S. C.: **Atalaphyllinine, a New Acridone Base from *Atalantia-monophylla***. *Phytochemistry*, 14, 835-836, (1975).

128. Panda, H.: in *Handbook on Medicinal Herbs with Uses* 166-167 (Asia Pacific Business Press, 2004).

129. Prasad, Y. R.: **Chemical Investigation and Antimicrobial Efficacy of the Volatile Leaf Oil of *Atalantia monophylla* Corr**. *Prafuemeria and Kosmetic*, 69, 418-419, (1988).

130. Govindachari, T. and Viswanathan, B.: **Alkaloids of *Atalantia monophylla correa* 1***. *Tetrahedron*, 26, 2905-2910, (1970).

131. Kulkarni, G. H. and Sabata, B. K.: **An Acridone Alkaloid from the Root Bark of *Atalantia-monophylla***. *Phytochemistry*, 20, 867-868, (1981).

132. Kawaii, S., Tomono, Y., Katase, E., Ogawa, K., Yano, M., Takemura, Y., Ju-Ichi, M., Ito, C. and Furukawa, H.: **Acridones as Inducers of HL-60 Cell Differentiation**. *Leukemia Research*, 23, 263-269, (1999).

133. Itoigawa, M., Ito, C., Wu, T. S., Enjo, F., Tokuda, H., Nishino, H. and Furukawa, H.: **Cancer Chemopreventive Activity of Acridone Alkaloids on Epstein-Barr Virus Activation and Two-Stage Mouse Skin Carcinogenesis**. *Cancer Letters*, 193, 133-138, (2003).

134. Kawaii, S., Tomono, Y., Katase, E., Ogawa, K., Yano, M., Takemura, Y., Motoharu, J., Ito, C. and Furukawa, H.: **The Antiproliferative Effect of Acridone Alkaloids on Several Cancer Cell Lines**. *Journal of Natural Products*, 62, 587-589, (1999).

135. Scott, I. M., Puniani, E., Jensen, H., Livesey, J. F., Poveda, L., Sanchez-Vindas, P., Durst, T. and Arnason, J. T.: **Analysis of Piperaceae Germplasm by HPLC and LCMS: A Method for Isolating and Identifying Unsaturated Amides from *Piper spp* Extracts**. *Journal of Agricultural and Food Chemistry*, 53, 1907-1913, (2005).

136. MIYAKADO, M., NAKAYAMA, I. and YOSHIOKA, H.: **Insecticidal Joint Action of Pipercide and Co-occurring Compounds Isolated from *Piper nigrum L***. *Agricultural and Biological Chemistry*, 44, 1701-1703, (1980).

137. Ee, G. C. L., Lim, C. M., Lim, C. K., Rahmani, M., Shaari, K. and Bong, C. F. J.: **Alkaloids from *Piper sarmentosum* and *Piper nigrum***. *Natural Product Research*, 23, 1416-1423, (2009).

138. Rukachaisirikul, T., Siriwattanakit, P., Sukcharoenphol, K., Wongvein, C., Ruttanaweang, P., Wongwattanavuch, P. and Suksamrarn, A.: **Chemical Constituents and Bioactivity of *Piper sarmentosum***. *Journal of Ethnopharmacology*, 93, 173-176, (2004).

139. Tuntiwachwuttikul, P., Phansa, P., Pootaeng-On, Y. and Taylor, W. C.: **Chemical Constituents of the Roots of *Piper sarmentosum***. *Chemical & Pharmaceutical Bulletin*, 54, 149-151, (2006).

140. Veznik, F., Taborska, E., Bochorakova, H., Turecek, F., Hanus, V. and Slavik, J.: **Alkaloids of the Papaveraceae .86. Alkaloids of *Papaver-nudicaule* Subsp *Xanthopetalum* (Trautv) Fedde and *Papaver-nudicaule* Subsp *Album* (Regel) Fedde from the Section *Scapiflora reichb***. *Collection of Czechoslovak Chemical Communications*, 52, 1634-1640, (1987).

141. Ligaa, U.: **Medicinal Plants of Mongolia Used in Mongolian Traditional Medicine**. (KCA Press, Seoul, Korea, 1996).

142. Brownstein, M. J.: **A Brief-History of Opiates, Opioid-Peptides, and Opioid Receptors**. *Proceedings of the National Academy of Sciences of the United States of America*, 90, 5391-5393, (1993).

143. Sertuerner, F. W. A. F.: **Darstellung Der Reinen Mohnsaè Ure (Opiumsaè Ure) Nebst Einer Chemischen Untersuchung Des Opiums Mit Vorzuè**

Glicher Hinsicht Auf Einen Darin Neu Entdeckten Stoff Und Die Dahin Gehoe Rigen Bemerckungen. *J Pharm Aerzte Apoth Chem*, 14, 47-93, (1806).

144. Mcquay, H. J.: **Opioids in Chronic Pain**. *British Journal of Anaesthesia*, 63, 213-226, (1989).

145. Ye, K. Q., Ke, Y., Keshava, N., Shanks, J., Kapp, J. A., Tekmal, R. R., Petros, J. and Joshi, H. C.: **Opium Alkaloid Noscapine Is an Antitumor Agent that Arrests Metaphase and Induces Apoptosis in Dividing Cells**. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 1601-1606, (1998).

146. Cline, S. D. and Coscia, C. J.: **Stimulation of Sanguinarine Production by Combined Fungal Elicitation and Hormonal Deprivation in Cell-Suspension Cultures of *Papaver-bracteatum***. *Plant Physiology*, 86, 161-165, (1988).

147. Lenfeld, J., Kroutil, M., Marsalek, E., Slavik, J., Preininger, V. and Simanek, V.: **Antiinflammatory Activity of Quaternary Benzophenanthridine Alkaloids from *Chelidonoum majus**,****. *Planta Medica*, 43, 161-5, (1981).

148. Ahmad, N., Gupta, S., Husain, M. M., Heiskanen, K. M. and Mukhtar, H.: **Differential Antiproliferative and Apoptotic Response of Sanguinarine for Cancer Cells versus Normal Cells**. *Clinical Cancer Research*, 6, 1524-8, (2000).

149. Asolkar, L. V.: **Glossary of Indian Medicinal Plants, Supplement**. in *CSIR, Publication and Information Directorate* 86-88 (New Delhi, 1992).

150. Dopke, W., Hess, U. and Jimenez, V.: **Structure of New Alkaloid from *Argemone-mexicana***. *Zeitschrift Fur Chemie*, 16, 54-55, (1976).

151. Nakkady, S. and Sharma, M.: **Studies on the Chemical Constitution of *Argemone mexicana***. *Egyptian Journal of Pharmaceutical Sciences*, 29, 53-61, (1988).

152. Chang, Y. C., Chang, F. R., Khalil, A. T., Hsieh, P. W. and Wu, Y. C.: **Cytotoxic Benzophenanthridine and Benzylisoquinoline Alkaloids from *Argemone mexicana***. *Zeitschrift Fur Naturforschung C-a Journal of Biosciences*, 58, 521-526, (2003).

153. Singh, S., Singh, T. D., Singh, V. P. and Pandey, V. B.: **A New Benzylisoquinoline Alkaloid from *Argemone mexicana***. *Natural Product Research*, 24, 63-67, (2010).

154. Buchanan, M. S., Davis, R. A., Duffy, S., Avery, V. M. and Quinn, R. J.: **Antimalarial Benzylisoquinoline Alkaloid from the Rainforest Tree *Doryphora sassafras***. *Journal of Natural Products*, 72, 1541-1543, (2009).

155. Greshoff, M.: **Chemische Berichte**. 23, 3537, (1890).

156. Janot, M. M.: in *The Alkaloids* (eds. Manske, R. H. F. and Holmes, H. L.) 363-394 (Academic Press, New York, 1953).

157. Fujii, T. and Ohba, M.: in *The Alkaloids* (ed Cordel, G. A.) 271-321 (Academic Press, New York, 1998).

158. Jordan, M. A., Thrower, D. and Wilson, L.: **Mechanism of Inhibition of Cell-Proliferation by Vinca Alkaloids**. *Cancer Research*, 51, 2212-2222, (1991).

159. Kulkarni, R. N., Baskaran, K., Chandrashekara, R. S. and Kumar, S.: **Inheritance of Morphological Traits of Periwinkle Mutants with Modified Contents and Yields of Leaf and Root Alkaloids**. *Plant Breeding*, 118, 71-74, (1999).

160. Clark, A. M., Watson, E. S., Ashfaq, M. K. and Hufford, C. D.: **In Vivo Efficacy of Antifungal Oxoaporphine Alkaloids in Experimental Disseminated Candidiasis**. *Pharmaceutical Research*, 4, 495-8, (1987).

161. Guinaudeau, H., Leboeuf, M. and Cave, A.: **Aporphinoid Alkaloids .5.** *Journal of Natural Products*, 57, 1033-1135, (1994).

162. Hufford, C. D., Sharma, A. S. and Oguntimein, B. O.: **Antibacterial and Antifungal Activity of Liriodenine and Related Oxoaporphine Alkaloids**. *Journal of Pharmaceutical Sciences*, 69, 1180-3, (1980).

163. Hufford, C. D., Liu, S., Clark, A. M. and Oguntimein, B. O.: **Anticandidal Activity of Eupolauridine and Onychine, Alkaloids from *Cleistopholis patens***. *Journal of Natural Products*, 50, 961-4, (1987).

164. Montanha, J. A., Amoros, M., Boustie, J. and Girre, L.: **Anti-Herpes Virus Activity of Aporphine Alkaloids**. *Planta Medica*, 61, 419-24, (1995).

165. Stevigny, C., Bailly, C. and Quetin-Leclercq, J.: **Cytotoxic and Antitumor Potentialities of Aporphinoid Alkaloids**. *Current Medicinal Chemistry Anticancer Agents*, 5, 173-82, (2005).

166. Muhammad, I., Dunbar, D. C., Takamatsu, S., Walker, L. A. and Clark, A. M.: **Antimalarial, Cytotoxic, and Antifungal Alkaloids from *Duguetia hadrantha***. *Journal of Natural Products*, 64, 559-62, (2001).

167. Rao, J. U. M., Giri, G. S., Hanumaiah, T. and Rao, K. V. J.: **Sampangine, a New Alkaloid from *Cananga-odorata***. *Journal of Natural Products*, 49, 346-347, (1986).

168. Peterson, J. R., Zjawiony, J. K., Liu, S., Hufford, C. D., Clark, A. M. and Rogers, R. D.: **Copyrine Alkaloids: Synthesis, Spectroscopic Characterization, and Antimycotic/Antimycobacterial Activity of A- and**

B-Ring-Functionalized Sampangines. *Journal of Medicinal Chemistry*, 35, 4069-77, (1992).

169. Leake, C. D. and Pelikan, E. W.: **An Historical Account of Pharmacology to the 20th Century (Book Review)**. *The Journal of Clinical Pharmacology*, 16, 669-671, (1976).

170. Lounasmaa, M. and Tamminen, T.: **The Tropane Alkaloids**. in *The Alkaloids* (ed Brossi, A.) 1-114 (Academic Press, New York, 1993).

171. Morris, T.: **Case of Poisoning by Potatoe-Berries**. *British Medical Journal*, 1, 719-720, (1859).

172. Joskow, R., Belson, M., Vesper, H., Backer, L. and Rubin, C.: **Ackee Fruit Poisoning: An Outbreak Investigation in Haiti 2000-2001, and Review of the Literature**. *Clinical Toxicology*, 44, 267-273, (2006).

173. Arredondo, J., Chernyavsky, A. I., Jolkovsky, D. L., Pinkerton, K. E. and Grando, S. A.: **Receptor-Mediated Tobacco Toxicity: Cooperation of the Ras/Raf-1/MEK1/ERK and JAK-2/STAT-3 Pathways Downstream of Alpha 7 Nicotinic Receptor in Oral Keratinocytes**. *FASEB Journal*, 20, 2093-2101, (2006).

174. Sassen, A. W., Richter, E., Semmler, M. P., Harreus, U. A., Gamarra, F. and Kleinsasser, N. H.: **Genotoxicity of Nicotine in Mini-Organ Cultures of Human Upper Aerodigestive Tract Epithelia**. *Toxicological Sciences*, 88, 134-141, (2005).

175. Kobayashi, J. and Morita, H.: in *The Alkaloids* (ed Cordell, G. A.) 165-205 (Academic Press, New York, 2003).

176. Heathcock, C. H.: **Nature Knows Best: An Amazing Reaction Cascade Is Uncovered by Design and Discovery**. *Proceedings of the National Academy of Sciences of the Uninted States of America*, 93, 14323-7, (1996).

177. Zhen, M. and Min, T. L.: in *Flora Republicae Popularis Sinicae (Zhongguo Zhi Wu Zhi)* 1-11 (Science Press, Beijing, 1980).

178. Morita, H., Ishioka, N., Takatsu, H., Shinzato, T., Obara, Y., Nakahata, N. and Kobayashi, J.: **Daphmanidins C and D, Novel Pentacyclic Alkaloids from *Daphniphyllum teijsmanii***. *Organic Letters*, 7, 459-462, (2005).

179. Kobayashi, J., Takatsu, H., Shen, Y. C. and Morita, H.: **Daphniglaucins A and B, Novel Polycyclic Quaternary Alkaloids from *Daphniphyllum glaucescens***. *Organic Letters*, 5, 1733-1736, (2003).

180. Yang, S. P. and Yue, J. M.: **Two Novel Alkaloids with a Unique Fused Hexacyclic Skeleton from *Daphniphyllum subverticillatum***. *Journal of Organic Chemistry*, 68, 7961-7966, (2003).

181. Tokuyama, T., Daly, J. and Witkop, B.: **The Structure of Batrachotoxin, a Steroidal Alkaloid from the Colombian Arrow Poison Frog, *Phyllobates aurotaenia*, and Partial Synthesis of Batrachotoxin and its Analogs and Homologs**. *Journal of the American Chemical Society*, 91, 3931-8, (1969).

182. Daly, J. W., Brown, G. B., Mensahdwumah, M. and Myers, C. W.: **Classification of Skin Alkaloids from Neotropical Poison-Dart Frogs (Dendrobatidae)**. *Toxicon*, 16, 163-188, (1978).

183. Daly, J. W., Highet, R. J. and Myers, C. W.: **Occurrence of Skin Alkaloids in Non-Dendrobatid Frogs from Brazil (Bufonidae), Australia (Myobatrachidae) and Madagascar (Mantellinae)**. *Toxicon*, 22, 905-919, (1984).

184. Garraffo, H. M., Spande, T. F., Daly, J. W., Baldessari, A. and Gros, E. G.: **Alkaloids from Bufonid Toads (Melanophryniscus) - Decahydroquinolines, Pumiliotoxins and Homopumiliotoxins, Indolizidines, Pyrrolizidines, and Quinolizidines**. *Journal of Natural Products*, 56, 357-373, (1993).

185. Daly, J. W., Wilham, J. M., Spande, T. F., Garraffo, H. M., Gil, R. R., Silva, G. L. and Vaira, M.: **Alkaloids in Bufonid Toads (Melanophryniscus): Temporal and Geographic Determinants for Two Argentinian species**. *Journal of Chemical Ecology*, 33, 871-887, (2007).

186. Daly, J. W., Kaneko, T., Wilham, J., Garraffo, H. M., Spande, T. F., Espinosa, A. and Donnelly, M. A.: **Bioactive Alkaloids of Frog Skin: Combinatorial Bioprospecting Reveals that Pumiliotoxins have an Arthropod Source**. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 13996-14001, (2002).

187. Saporito, R. A., Garraffo, H. M., Donnelly, M. A., Edwards, A. L., Longino, J. T. and Daly, J. W.: **Formicine Ants: An Arthropod Source for the Pumiliotoxin Alkaloids of Dendrobatid Poison Frogs**. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 8045-8050, (2004).

188. Saporito, R. A., Donnelly, M. A., Norton, R. A., Garraffo, H. M., Spande, T. F. and Daly, J. W.: **Oribatid Mites as a Major Dietary Source for Alkaloids in Poison Frogs**. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 8885-8890, (2007).

189. Takada, W., Sakata, T., Shimano, S., Enami, Y., Mori, N., Nishida, R. and Kuwahara, Y.: **Scheloribatid Mites as the Source of Pumiliotoxins in Dendrobatid Frogs**. *Journal of Chemical Ecology*, 31, 2403-2415, (2005).

190. Smith, B. P., Tyler, M. J., Kaneko, T., Garraffo, H. M., Spande, T. F. and Daly, J. W.: **Evidence for Biosynthesis of Pseudophrynamine Alkaloids by an Australian Myobatrachid Frog (Pseudophryne) and for Sequestration of Dietary Pumiliotoxins**. *Journal of Natural Products*, 65, 439-447, (2002).

191. Daly, J. W., Spande, T. F. and Garraffo, H. M.: **Alkaloids from Amphibian Skin: A Tabulation of Over Eight-Hundred Compounds**. *Journal of Natural Products*, 68, 1556-1575, (2005).

192. Berde, B. and Sturmer, E.: **Introduction to the Pharmacology of Ergot Alkaloids and Related Compounds**. in *Ergot Alkaloids and Related Compounds* (ed Aellig, W. H.) 1-28 (Springer, Berlin, Germany, 1978).

193. Mantegani, S., Brambilla, E. and Varasi, M.: **Ergoline Derivatives: Receptor Affinity and Selectivity**. *Farmaco*, 54, 288-296, (1999).

194. Tudzynski, P., Correia, T. and Keller, U.: **Biotechnology and Genetics of Ergot Alkaloids**. *Applied Microbiology and Biotechnology*, 57, 593-605, (2001).

195. Hofle, G. and Irschik, H.: **Isolation and Biosynthesis of Aurachin P and 5-Nitroresorcinol from *Stigmatella erecta***. *Journal of Natural Products*, 71, 1946-1948, (2008).

196. Hoefle, G. and Kunze, B.: **Biosynthesis of Aurachins A-L in *Stigmatella aurantiaca*: A Feeding Study**. *Journal of Natural Products*, 71, 1843-1849, (2008).

197. Shahabuddin, M. S., Nambiar, M., Moorthy, B. T., Naik, P. L., Choudhary, B., Advirao, G. M. and Raghavan, S. C.: **A Novel Structural Derivative of Natural Alkaloid Ellipticine, MDPSQ, Induces Necrosis in Leukemic Cells**. *Investigational New Drugs*, (2010).

198. Frodin, D. G.: **History and Concepts of Big Plant Genera**. *Taxon*, 53, 753-776, (2004).

199. Nickrent, D. L., Blarer, A., Qiu, Y. L., Soltis, D. E., Soltis, P. S. and Zanis, M.: **Molecular Data Place Hydnoraceae with Aristolochiaceae**. *American Journal of Botany*, 89, 1809-1817, (2002).

200. Bremer, B., Bremer, K., Chase, M. W., Reveal, J. L., Soltis, D. E., Soltis, P. S., Stevens, P. F., Anderberg, A. A., Fay, M. F., Goldblatt, P., Judd, W. S., Kallersjo, M., Karehed, J., Kron, K. A., Lundberg, J., Nickrent, D. L., Olmstead, R. G., Oxelman, B., Pires, J. C., Rodman, J. E., Rudall, P. J., Savolainen, V., Sytsma, K. J., van der Bank, M., Wurdack, K., Xiang, J. Q. Y., Zmarzty, S. and Grp, A. P.: **An Update of the Angiosperm Phylogeny Group Classification for the Orders and Families of Flowering Plants: APG II**. *Botanical Journal of the Linnean Society*, 141, 399-436, (2003).

201. Semple, K. S.: **Pollination of Piperaceae**. *Annals of Missouri Botanical Garden*, 61, 868-871, (1974).

202. De Figueiredo, R. A. and Sazima, M.: **Pollination Biology of Piperaceae Species in Southeastern Brazil**. *Annals of Botany*, 85, 455-460, (2000).

203. Bornstein, A. J.: **Taxonomic Studies in the Piperaceae .1. The Pedicellate Pipers of Mexico and Central America (*Piper subg arctottonia*)**. *Journal of the Arnold Arboretum*, 70, 1-55, (1989).

204. Nortier, J. L., Martinez, M. M., Schmeiser, H. H., Arlt, V. M., Bieler, C. A., Petein, M., Depierreux, M. F., De Pauw, L., Abramowicz, D., Vereerstraeten, P. and Vanherweghem, J. L.: **Urothelial Carcinoma Associated with the Use of a Chinese Herb (*Aristolochia fangchi*).** *New England Journal of Medicine*, 342, 1686-1692, (2000).

205. Borsch, T., Löhne, C., Müller, K., Hilu, K., Wanke, S., Worberg, A., Barthlott, W., Neinhuis, C. and Quandt, D.: **Towards Understanding Basal Angiosperm Diversification: Recent Insights Using Rapidly Evolving Genomic Regions**. *Nova Acta Leopoldina NF*, 92, 85–110, (2005).

206. Neinhuis, C., Wanke, S., Hilu, K. W., Muller, K. and Borsch, T.: **Phylogeny of Aristolochiaceae Based on Parsimony, Likelihood, and Bayesian Analyses of trnL-trnF Sequences**. *Plant Systematics and Evolution*, 250, 7-26, (2005).

207. González, F. and Stevenson, D.: **A Phylogenetic Analysis of the Subfamily Aristolochioideae (Aristolochiaceae)**. *Rev. Acad. Colomb. Cienc. Exact. Fis. Nat*, 26, 25-57, (2002).

208. Wanke, S., Gonzalez, F. and Neinhuis, C.: **Systematics of Pipevines: Combining Morphological and Fast-Evolving Molecular Characters to Investigate the Relationships within Subfamily Aristolochioideae (Aristolochiaceae)**. *International Journal of Plant Sciences*, 167, 1215-1227, (2006).

209. Wanke, S., Jaramillo, M. A., Borsch, T., Samain, M. S., Quandt, D. and Neinhuis, C.: **Evolution of Piperales - matK Gene and trnK Intron Sequence Data Reveal Lineage Specific Resolution Contrast**. *Molecular Phylogenetics and Evolution*, 42, 477-497, (2007).

210. Huber, H.: **Aristolochiaceae**. in *The Families and Genera of Vascular Plants* (eds. Kubitzki, K., Rohwer, J. G. and Bittrich, V.) 129-137 (Springer-Vlg, Berlin ; New York, 1993).

211. Knoll, F.: **Die Gleitfalle *Als Blumentypus***. *Verh Zool-Bot Ges Wien*, 79, 9-12, (1929).

212. Sprengel, C. K.: **Das entdeckte Geheimnis der Natur im Bau und in der Befruchtung der Blumen**. in *Ostwald's Klassiker Der Exakten Naturwissenschaften* (ed Knuth, P. E.) 48-51 (Engelmann. Leipzig, Leipzig, 1793).

213. Proctor, M., Yeo, P. and Lack, A.: **The Natural History of Pollination**. in *The New Naturalist, 83* (ed Collins, H.) 479 (London, (New), 1996).

214. Ohi-Toma, T., Sugawara, T., Murata, H., Wanke, S., Neinhuis, C. and Murata, J.: **Molecular Phylogeny of *Aristolochia sensu lato* (Aristolochiaceae) Based on Sequences of rbcL, matK, and phyA Genes, with Special Reference to Differentiation of Chromosome Numbers**. *Systematic Botany*, 31, 481-492, (2006).

215. De Groot, H., Wanke, S. and Neinhuis, C.: **Revision of the Genus *Aristolochia* (Aristolochiaceae) in Africa, Madagascar and Adjacent Islands**. *Botanical Journal of the Linnean Society*, 151, 219-238, (2006).

216. Gonzalez, F. and Stevenson, D. W.: **Perianth Development and Systematics of *Aristolochia***. *Flora*, 195, 370-391, (2000).

217. Klitzke, C. F. and Brown, K. S.: **The Occurrence of Aristolochic Acids in Neotropical Troidine Swallowtails (Lepidoptera: Papilionidae)**. *Chemoecology*, 10, 99-102, (2000).

218. Rausher, M. D.: **Host Plant-Selection by *Battus-philenor* Butterflies - the Roles of Predation, Nutrition, and Plant Chemistry**. *Ecological Monographs*, 51, 1-20, (1981).

219. Vieira, L., De Pascoli, I., De Bortoli, S. and Lopes, L.: **Efeito De Extratos De *Aristolochia lagesiana* (Aristolochiaceae) Sobre A Lagarta-Da-Soja, *Anticarsia gemmatalis* (Lepidoptera: Noctuidae)**. *Arquivos Instituto Biol., São Paulo*, 76, 245-250, (2009).

220. Nascimento, I. R. and Lopes, L. M. X.: **2,3-Dihydrobenzofuran Neolignans from *Aristolochia pubescens* (vol 52, pg 345, 1999)**. *Phytochemistry*, 53, 621-621, (2000).

221. Nascimento, I. R. and Lopes, L. M. X.: **Diterpene Esters of Aristolochic Acids from *Aristolochia pubescens***. *Phytochemistry*, 63, 953-957, (2003).

222. Nascimento, I. R., Lopes, L. M. X., Davin, L. B. and Lewis, N. G.: **Stereoselective Synthesis of 8,9-Licarinediols**. *Tetrahedron*, 56, 9181-9193, (2000).

223. Nascimento, I. R., Murata, A. T., Bortoli, S. A. and Lopes, L. M. X.: **Insecticidal Activity of Chemical Constituents from *Aristolochia pubescens* Against *Anticarsia gemmatalis* Larvae**. *Pest Management Science*, 60, 413-416, (2004).

224. Harmatha, J. and Dinan, L.: **Biological Activities of Lignans and Stilbenoids Associated with Plant-Insect Chemical Interactions**. *Phytochemistry Reviews*, 2, 321-330, (2003).

225. Klein Gebbinck, E., Jansen, B. and de Groot, A.: **Insect Antifeedant Activity of Clerodane Diterpenes and Related Model Compounds**. *Phytochemistry*, 61, 737-770, (2002).

226. Saxena, B., Koul, O., Tikku, K., Atal, C., Suri, O. and Suri, K.: **Aristolochic Acid—an Insect Chemosterilant from *Aristolochia bracteata* Retz**. *Indian Journal of Experimental Biology*, 17, 354-360, (1979).

227. Carneiro, F., Boralle, N., Silva, D. and Lopes, L.: **Bi-and Tetraflavonoids from *Aristolochia ridicula***. *Phytochemistry*, 55, 823-832, (2000).

228. Murakami, A., Tanaka, S., Ohigashi, H., Hirota, M., Irie, R., Takeda, N., Tatematsu, A. and Koshimizu, K.: **Possible Antitumor Promoters - Biflavonoids and Tetraflavonoids from *Lophira-alata***. *Phytochemistry*, 31, 2689-2693, (1992).

229. Tih, A., Martin, M. T., Tih, R. G., Vuidepot, I., Sondengam, B. L. and Bodo, B.: **Lophiroflavans-B and Lophiroflavans-C Tetraflavonoids of *Lophira-alata***. *Phytochemistry*, 31, 3595-3599, (1992).

230. Wang, L., Su, H., Yang, S., Won, S. and Lin, C.: **New Alkaloids and a Tetraflavonoid from *Cephalotaxus wilsoniana***. *Journal of Natural Products*, 67, 1182-1185, (2004).

231. Pailer, M. and Pruckmayr, G.: **Über die Basischen Inhaltsstoffe der *Aristolochia clematitis L*.** *Monatshefte für Chemie/Chemical Monthly*, 90, 145-147, (1959).

232. Pailer, M.: **Nat u rally Oc curring Ni tro gen Com pounds**. *Fortschritte der Chemie Organischer Naturstoffe*, 18, 55-82, (1960).

233. Schütte, H., Orban, U. and Mothes, K.: **Biosynthesis of Aristolochic Acid**. *European Journal of Biochemistry*, 1, 70-72, (1967).

234. Arlt, V., Stiborova, M. and Schmeiser, H.: **Aristolochic Acid as a Probable Human Cancer Hazard in Herbal Remedies: A Review**. *Mutagenesis*, 17, 265-277, (2002).

235. Cosyns, J.: **Aristolochic Acid and Chinese Herbs Nephropathy': A Review of the Evidence to Date**. *Drug Safety*, 26, 33-48, (2003).

236. Wojcikowski, K., Johnson, D. and Gobe, G.: **Medicinal Herbal Extracts–Renal Friend or Foe? Part One: The Toxicities of Medicinal Herbs**. *Nephrology*, 9, 313-318, (2004).

237. Lopes, L., Nascimento, I. and Silva, T.: **Phytochemistry of the Aristolochiaceae family**. *Research Advances in Phytochemistry*, 2, 19–108, (2001).

238. Ioset, J., Raoelison, G. and Hostettmann, K.: **Detection of Aristolochic Acid in Chinese Phytomedicines and Dietary Supplements Used as Slimming Regimens**. *Food and Chemical Toxicology*, 41, 29-36, (2003).

239. Shaw, F.: **The Pharmacological Testing of Alkaloids from Australian Flora.** *The Australasian Journal of Pharmacy*, 28, 857, (1947).

240. Barnard, C.: *Australian Journal of Science,*, 12, 30, (1949).

241. Berghe, D., Ieven, M., Mertens, F., Vlietinck, A. and Lammens, E.: **Screening of Higher Plants for Biological Activities 11 Antiviral Activity**. *Journal of Natural Products*, 11, 463-467, (1978).

242. Branch, L. and Silva, M.: **Folk Medicine of Alter do Chão, Pará, Brazil**. *Acta Amazonica*, 13, 737-797, (1983).

243. Montes, M. and Wilkomirsky, T.: **Medicina Traditional Chile**. *Editorial de la Universidad de Concepcion. Chile*, 15, 52, (1985).

244. Li, H., Youji, S., Shingo, M. and Chen, X.: **Eleven Aristolochic Acid Derivatives from *Aristolochia cinnabarina***. *Phytochemistry*, 37, 237-239, (1994).

245. Lewis, W. and Elvin-Lewis, M.: *Medical botany: plants affecting human health*, (Wiley, Hoboken, N. J., 2003).

246. Yu, Y., Zheng, F. and Li, H.: **Chinese Herbs-Induced Renal Failure with Fanconi Syndrome: A Report of 6 Cases**. *Zhonghua Nei Ke Za Zhi [Chinese journal of internal medicine]*, 42, 110-112, (2003).

247. Tanaka, A., Nishida, R., Maeda, K., Sugawara, A. and Kuwahara, T.: **Chinese Herb Nephropathy in Japan Presents Adult-Onset Fanconi Syndrome: Could Different Components of Aristolochic Acids Cause a Different Type of Chinese Herb Nephropathy?** *Clinical Nephrology*, 53, 301-306, (2000).

248. Qiu, Q., Liu, Z., Chen, H., Yin, H., Li, L., Drugs, C., Neoplasms, K., Plants, M. and Allocation, R.: **Long-Term Outcome of Acute Renal Injury Induced by *Aristolochia manshuriensis Kom* in Rats**. *Acta pharmacologica Sinica*, 21, 1129-1135, (2000).

249. Hu, S., Zhang, H., Chan, K. and Mei, Q.: **Studies on the Toxicity of *Aristolochia manshuriensis* (Guanmuton)**. *Toxicology*, 198, 195-201, (2004).

250. Mengs, U., Lang, W. and Poch, J.: **The Carcinogenic Action of Aristolochic Acid in Rats**. *Archives of Toxicology*, 51, 107-119, (1982).

251. Mengs, U.: **Acute Toxicity of Aristolochic Acid in Rodents**. *Archives of Toxicology*, 59, 328-331, (1987).

252. *International Agency for Research on Cancer (IARC). Some traditional herbal medicines, some mycotoxins, nafhtalene and styrene*, (IARC, Lyon, France, 2002).

253. Carlson, J., Leebens-Mack, J., Wall, P., Zahn, L., Mueller, L., Landherr, L., Hu, Y., Ilut, D., Arrington, J. and Choirean, S.: **EST Database for Early Flower Development in California Poppy (*Eschscholzia californica Cham*., Papaveraceae) Tags Over 6000 Genes from a Basal Eudicot**. *Plant Molecular Biology*, 62, 351-369, (2006).

254. Becker, A., Gleissberg, S. and Smyth, D.: **Floral and Vegetative Morphogenesis in California Poppy (*Eschscholzia californica Cham*.)**. *International Journal of Plant Sciences*, 166, 537-555, (2005).

255. Chang, S., Puryear, J. and Cairney, J.: **A Simple and Efficient Method for Isolating RNA from Pine Trees**. *Plant Molecular Biology Reporter*, 11, 113-116, (1993).

256. Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.: **Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs**. *Nucleic Acids Research*, 25, 3389-3402, (1997).

257. Koonin, E. and Galperin, M.: **Identification of Genes in a Genomic DNA Sequence**. in *Sequence-evolution-function: computational approaches in comparative genomics* 461 (Springer, Netherlands, 2003).

258. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. and Bairoch, A.: **ExPASy: The Proteomics Server for in-Depth Protein Knowledge and Analysis**. *Nucleic Acids Research*, 31, 3784-3788, (2003).

259. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M., Appel, R., Bairoch, A. and Walker, J.: **Protein Identification and Analysis Tools on the ExPASy Server**. in *The proteomics protocols handbook* (ed Walker, J. M.) 571–607 (Humana Press, Totowa, N. J., 2005).

260. Pieper, U., Eswar, N., Webb, B., Eramian, D., Kelly, L., Barkan, D., Carter, H., Mankoo, P., Karchin, R. and Marti-Renom, M.: **MODBASE, a Database of Annotated Comparative Protein Structure Models and Associated Resources**. *Nucleic Acids Research*, 37, D347-D354, (2009).

261. Zdobnov, E. M. and Apweiler, R.: **InterProScan--an Integration Platform for the Signature-Recognition Methods in InterPro**. *Bioinformatics*, 17, 847-848, (2001).

262. Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Laugraud, A., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., Mistry, J., Mitchell, A., Mulder, N., Natale, D., Orengo, C., Quinn, A. F., Selengut, J. D., Sigrist, C. J., Thimma, M., Thomas, P. D., Valentin, F., Wilson, D., Wu, C. H. and Yeats, C.: **InterPro: The Integrative Protein Signature Database**. *Nucleic Acids Research*, 37, D211-D215, (2009).

263. Chou, K. C. and Shen, H. B.: **Signal-CF: A Subsite-Coupled and Window-Fusing Approach for Predicting Signal Peptides**. *Biochemical and Biophysical Research Communications*, 357, 633-640, (2007).

264. Shen, H. B. and Chou, K. C.: **Signal-3L: A 3-Layer Approach for Predicting Signal Peptides**. *Biochemical and Biophysical Research Communications*, 363, 297-303, (2007).

265. Chou, K. C. and Shen, H. B.: **MemType-2L: A Web Server for Predicting Membrane Proteins and Their Types by Incorporating Evolution Information Through Pse-PSSM**. *Biochemical and Biophysical Research Communications*, 360, 339-345, (2007).

266. Shen, H. B. and Chou, K. C.: **EzyPred: A Top-Down Approach for Predicting Enzyme Functional Classes and Subclasses**. *Biochemical and Biophysical Research Communications*, 364, 53-59, (2007).

267. Chou, K. C. and Shen, H. B.: **ProtIdent: A Web Server for Identifying Proteases and Their Types by Fusing Functional Domain and Sequential Evolution Information**. *Biochemical and Biophysical Research Communications*, 376, 321-325, (2008).

268. Lu, Z., Szafron, D., Greiner, R., Lu, P., Wishart, D. S., Poulin, B., Anvik, J., Macdonell, C. and Eisner, R.: **Predicting Subcellular Localization of Proteins Using Machine-Learned Classifiers**. *Bioinformatics*, 20, 547-556, (2004).

269. Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J. and Nakai, K.: **Protein Subcellular Localization Prediction with Wolf Psort. Proceedings of the 4th Annual Asia Pacific Bioinformatics Conference APBC06**. in 39-48 (Taipei, Taiwan, 2006).

270. Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J. and Nakai, K.: **WoLF PSORT: Protein Localization Predictor**. *Nucleic Acids Research*, 35, W585-W587, (2007).

271. Nair, R. and Rost, B.: **Mimicking Cellular Sorting Improves Prediction of Subcellular Localization**. *Journal of Molecular Biology*, 348, 85-100, (2005).

272. Yu, C. S., Chen, Y. C., Lu, C. H. and Hwang, J. K.: **Prediction of Protein Subcellular Localization**. *Proteins*, 64, 643-651, (2006).

273. Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J. C., Yan, J. X., Gooley, A. A., Hughes, G., Humphery-Smith, I., Williams, K. L. and Hochstrasser, D. F.: **From Proteins to Proteomes: Large Scale Protein**

Identification by Two-Dimensional Electrophoresis and Amino Acid Analysis. *Biotechnology (N Y)*, 14, 61-65, (1996).

274. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G.: **The CLUSTAL_X Windows Interface: Flexible Strategies for Multiple Sequence Alignment Aided by Quality Analysis Tools**. *Nucleic Acids Research*, 25, 4876-4882, (1997).

275. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G.: **Clustal W and Clustal X Version 2.0**. *Bioinformatics*, 23, 2947-2948, (2007).

276. Tamura, K., Dudley, J., Nei, M. and Kumar, S.: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0**. *Molecular Biology and Evolution*, 24, 1596-1599, (2007).

277. Kumar, S., Nei, M., Dudley, J. and Tamura, K.: **MEGA: A Biologist-Centric Software for Evolutionary Analysis of DNA and Protein Sequences**. *Briefings in Bioinformatics*, 9, 299-306, (2008).

278. Zuckerkandl, E. and Pauling, L.: **Evolutionary Divergence and Convergence in Proteins. In: Evolving Genes and Proteins : A Symposium Held at the Institute of Microbiology of Rutgers, the State University with Support from the National Science Foundation**. (eds. Bryson, V. and Vogel, H. J.) 97-166 (Academic Press, New York, 1965).

279. Saitou, N. and Nei, M.: **The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees**. *Molecular Biology and Evolution*, 4, 406-425, (1987).

280. Finn, R. D., Tate, J., Mistry, J., Coggill, P. C., Sammut, S. J., Hotz, H. R., Ceric, G., Forslund, K., Eddy, S. R., Sonnhammer, E. L. and Bateman, A.: **The Pfam Protein Families Database**. *Nucleic Acids Research*, 36, D281-D288, (2008).

281. Combet, C., Blanchet, C., Geourjon, C. and Deleage, G.: **NPS@: Network Protein Sequence Analysis**. *Trends in Biochemical Sciences*, 25, 147-150, (2000).

282. Raghava, G.: **Protein Secondary Structure Prediction Using Nearest Neighbor and Neural Network Approach**. *CASP*, 4, 75–76, (2000).

283. Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M. and Barton, G. J.: **JPred: A Consensus Secondary Structure Prediction Server**. *Bioinformatics*, 14, 892-893, (1998).

284. Cole, C., Barber, J. D. and Barton, G. J.: **The Jpred 3 Secondary Structure Prediction Server**. *Nucleic Acids Research*, 36, W197-W201, (2008).

285. Pons, J. L. and Labesse, G.: **@TOME-2: A New Pipeline for Comparative Modeling of Protein-Ligand Complexes**. *Nucleic Acids Research*, 37, W485-W491, (2009).

286. Huang, C. C., Smith, C. V., Glickman, M. S., Jacobs, W. R., Jr. and Sacchettini, J. C.: **Crystal Structures of Mycolic Acid Cyclopropane Synthases from *Mycobacterium tuberculosis***. *Journal of Biological Chemistry*, 277, 11559-11569, (2002).

287. Boissier, F., Bardou, F., Guillet, V. R., Uttenweiler-Joseph, S., Daffe, M., Quemard, A. and Mourey, L.: **Further Insight into S-Adenosylmethionine-Dependent Methyltransferases - Structural Characterization of Hma, an Enzyme Essential for the Biosynthesis of Oxygenated Mycolic Acids in**

Mycobacterium tuberculosis. *Journal of Biological Chemistry*, 281, 4434-4445, (2006).

288. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E.: **The Protein Data Bank**. *Nucleic Acids Research*, 28, 235-242, (2000).

289. Berman, H., Henrick, K., Nakamura, H. and Markley, J. L.: **The Worldwide Protein Data Bank (wwPDB): Ensuring a Single, Uniform Archive of PDB Data**. *Nucleic Acids Research*, 35, D301-D303, (2007).

290. Eck, R. and Dayhoff, M.: **Atlas of Protein Sequence and Structure**. *National Biomedical Research Foundation, Silver Spring, MD*, 161–202, (1966).

291. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U. and Sali, A.: **Comparative Protein Structure Modeling Using Modeller**. *Current Protocols in Bioinformatics*, Chapter 5, Unit 5 6, (2006).

292. Eswar, N., Eramian, D., Webb, B., Shen, M. Y. and Sali, A.: **Protein Structure Modeling with MODELLER**. *Methods in Molecular Biology*, 426, 145-159, (2008).

293. Shen, M. Y. and Sali, A.: **Statistical Potential for Assessment and Prediction of Protein Structures**. *Protein Science*, 15, 2507-2524, (2006).

294. Laskowski, R. A., Macarthur, M. W., Moss, D. S. and Thornton, J. M.: **Procheck - a Program to Check the Stereochemical Quality of Protein Structures**. *Journal of Applied Crystallography*, 26, 283-291, (1993).

295. Sippl, M. J.: **Recognition of Errors in Three-Dimensional Structures of Proteins**. *Proteins*, 17, 355-362, (1993).

296. Wiederstein, M. and Sippl, M. J.: **ProSA-web: Interactive Web Service for the Recognition of Errors in Three-Dimensional Structures of Proteins**. *Nucleic Acids Research*, 35, W407-W410, (2007).

297. Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y. and Liang, J.: **CASTp: Computed Atlas of Surface Topography of Proteins with Structural and Topographical Mapping of Functionally Annotated residues**. *Nucleic Acids Research*, 34, W116-W118, (2006).

298. Marchler-Bauer, A., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., He, S., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Liebert, C. A., Liu, C., Lu, F., Lu, S., Marchler, G. H., Mullokandov, M., Song, J. S., Tasneem, A., Thanki, N., Yamashita, R. A., Zhang, D., Zhang, N. and Bryant, S. H.: **CDD: Specific Functional Annotation with the Conserved Domain Database**. *Nucleic Acids Research*, 37, D205-D210, (2009).

299. Duhovny, D., Nussinov, R. and Wolfson, H. J.: **Efficient Unbound Docking of Rigid Molecules**. *Algorithms in Bioinformatics, Proceedings*, 2452, 185-200, (2002).

300. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. and Wolfson, H. J.: **PatchDock and SymmDock: Servers for Rigid and Symmetric Docking**. *Nucleic Acids Research*, 33, W363-W367, (2005).

301. Huey, R., Morris, G. M., Olson, A. J. and Goodsell, D. S.: **A Semiempirical Free Energy Force Field with Charge-Based Desolvation**. *Journal of Computational Chemistry*, 28, 1145-1152, (2007).

302. Sanner, M. F.: **Python: A Programming Language for Software Integration and Development**. *Journal of Molecular Graphics and Modelling*, 17, 57-61, (1999).

303. Wallace, A. C., Laskowski, R. A. and Thornton, J. M.: **LIGPLOT: A Program to Generate Schematic Diagrams of Protein-Ligand Interactions**. *Protein Engineering*, 8, 127-134, (1995).

304. Nei, M. and Kumar, S.: *Molecular Evolution and Phylogenetics*, (Oxford University Press, USA, 2000).

305. Felsenstein, J.: **Confidence-Limits on Phylogenies - an Approach Using the Bootstrap**. *Evolution*, 39, 783-791, (1985).

306. Fuhrman, J. A., Griffith, J. F. and Schwalbach, M. S.: **Prokaryotic and Viral Diversity Patterns in Marine Plankton**. *Ecological Research*, 17, 183-194, (2002).

307. Lindqvist, C., Scheen, A. C., Yoo, M. J., Grey, P., Oppenheimer, D. G., Leebens-Mack, J. H., Soltis, D. E., Soltis, P. S. and Albert, V. A.: **An Expressed Sequence Tag (EST) Library from Developing Fruits of an Hawaiian Endemic Mint (*Stenogyne rugosa*, Lamiaceae): Characterization and Microsatellite Markers**. *BMC Plant Biology*, 6, 16, (2006).

308. Varshney, R. K., Graner, A. and Sorrells, M. E.: **Genic Microsatellite Markers in Plants: Features and Applications**. *Trends Biotechnol*, 23, 48-55, (2005).

309. Mount, D.: *Bioinformatics: sequence and genome analysis*, (CSHL press, Cold Spring Harbor, N. Y., 2004).

310. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J.: **Basic Local Alignment Search Tool**. *Journal of Molecular Biology*, 215, 403-410, (1990).

311. Liscombe, D. K., Ziegler, J., Schmidt, J., Ammer, C. and Facchini, P. J.: **Targeted Metabolite and Transcript Profiling for Elucidating Enzyme Function: Isolation of Novel N-Methyltransferases from Three Benzylisoquinoline Alkaloid-Producing Species**. *Plant Journal*, 60, 729-43, (2009).

312. Tuskan, G. A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R. R., Bhalerao, R. P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G. L., Cooper, D., Coutinho, P. M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., Dejardin, A., Depamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjarvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leple, J. C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D. R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouze, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C. J., Uberbacher, E.,

Unneberg, P., Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G., Van de Peer, Y. and Rokhsar, D.: **The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray)**. *Science*, 313, 1596-604, (2006).

313. Aoki, K., Yano, K., Suzuki, A., Kawamura, S., Sakurai, N., Suda, K., Kurabayashi, A., Suzuki, T., Tsugane, T., Watanabe, M., Ooga, K., Torii, M., Narita, T., Shin, I. T., Kohara, Y., Yamamoto, N., Takahashi, H., Watanabe, Y., Egusa, M., Kodama, M., Ichinose, Y., Kikuchi, M., Fukushima, S., Okabe, A., Arie, T., Sato, Y., Yazawa, K., Satoh, S., Omura, T., Ezura, H. and Shibata, D.: **Large-Scale Analysis of Full-Length cDNAs from the Tomato (*Solanum lycopersicum*) Cultivar Micro-Tom, a Reference System for the Solanaceae genomics**. *BMC Genomics*, 11, 210.

314. Haas, B. J., Volfovsky, N., Town, C. D., Troukhan, M., Alexandrov, N., Feldmann, K. A., Flavell, R. B., White, O. and Salzberg, S. L.: **Full-Length Messenger RNA Sequences Greatly Improve Genome Annotation**. *Genome Biology*, 3, RESEARCH0029, (2002).

315. Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D. and Hochstrasser, D. F.: **Protein Identification and Analysis Tools in the ExPASy Server**. *Methods in Molecular Biology*, 112, 531-52, (1999).

316. Nishikawa, K., Kubota, Y. and Ooi, T.: **Classification of Proteins into Groups Based on Amino-Acid-Composition and Other Characters .1. Angular-Distribution**. *Journal of Biochemistry*, 94, 981-995, (1983).

317. Nishikawa, K., Kubota, Y. and Ooi, T.: **Classification of Proteins into Groups Based on Amino-Acid-Composition and Other Characters .2. Grouping into 4 Types**. *Journal of Biochemistry*, 94, 997-1007, (1983).

318. Guruprasad, K., Reddy, B. V. B. and Pandit, M. W.: **Correlation between Stability of a Protein and Its Dipeptide Composition - a Novel-Approach for Predicting Invivo Stability of a Protein from Its Primary Sequence**. *Protein Engineering*, 4, 155-161, (1990).

319. Ikai, A.: **Thermostability and Aliphatic Index of Globular Proteins**. *J Biochem*, 88, 1895-1898, (1980).

320. Kyte, J. and Doolittle, R. F.: **A Simple Method for Displaying the Hydropathic Character of a Protein**. *Journal of Molecular Biology*, 157, 105-132, (1982).

321. Marchler-Bauer, A., Anderson, J. B., Derbyshire, M. K., DeWeese-Scott, C., Gonzales, N. R., Gwadz, M., Hao, L., He, S., Hurwitz, D. I., Jackson, J. D., Ke, Z., Krylov, D., Lanczycki, C. J., Liebert, C. A., Liu, C., Lu, F., Lu, S., Marchler, G. H., Mullokandov, M., Song, J. S., Thanki, N., Yamashita, R. A., Yin, J. J., Zhang, D. and Bryant, S. H.: **CDD: A Conserved Domain Database for Interactive Domain Family Analysis**. *Nucleic Acids Research*, 35, D237-D240, (2007).

322. Arnold, K., Bordoli, L., Kopp, J. and Schwede, T.: **The SWISS-MODEL Workspace: A Web-Based Environment for Protein Structure Homology Modelling**. *Bioinformatics*, 22, 195-201, (2006).

323. Pieper, U., Eswar, N., Stuart, A. C., Ilyin, V. A. and Sali, A.: **MODBASE, a Database of Annotated Comparative Protein Structure Models**. *Nucleic Acids Research*, 30, 255-259, (2002).

324. Sanchez, R. and Sali, A.: **Large-Scale Protein Structure Modeling of the *Saccharomyces cerevisiae* Genome**. *Proceedings of the National Academy of Sciences of the Uninted States of America*, 95, 13597-13602, (1998).

325. Koehl, P. and Levitt, M.: **A Brighter Future for Protein Structure Prediction**. *Nature Structural Biology*, 6, 108-111, (1999).

326. Bates, P. A., Kelley, L. A., MacCallum, R. M. and Sternberg, M. J.: **Enhancement of Protein Modeling by Human Intervention in Applying the Automatic Programs 3D-JIGSAW and 3D-PSSM**. *Proteins*, Suppl 5, 39-46, (2001).

327. Moult, J.: **A Decade of CASP: Progress, Bottlenecks and Prognosis in Protein Structure Prediction**. *Current Opinion in Structural Biology*, 15, 285-289, (2005).

328. John, B. and Sali, A.: **Comparative Protein Structure Modeling by Iterative Alignment, Model Building and Model Assessment**. *Nucleic Acids Research*, 31, 3982-3992, (2003).

329. Pauling, L. and Corey, R. B.: **Configurations of Polypeptide Chains with Favored Orientations around Single Bonds: Two New Pleated Sheets**. *Proceedings of the National Academy of Sciences of the Uninted States of America*, 37, 729-740, (1951).

330. Pauling, L., Corey, R. B. and Branson, H. R.: **The Structure of Proteins; Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain**. *Proceedings of the National Academy of Sciences of the Uninted States of America*, 37, 205-211, (1951).

331. Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. and Shore, V. C.: **Structure of Myoglobin: A Three-**

Dimensional Fourier Synthesis at 2 A. Resolution. *Nature*, 185, 422-427, (1960).

332. Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G. and North, A. C.: **Structure of Haemoglobin: A Three-Dimensional Fourier Synthesis at 5.5-A. Resolution, Obtained by X-ray Analysis**. *Nature*, 185, 416-422, (1960).

333. Tramontano, A.: **Homology Modeling with Low Sequence Identity**. *Methods*, 14, 293-300, (1998).

334. Kopp, J. and Schwede, T.: **Automated Protein Structure Homology Modeling: A Progress Report**. *Pharmacogenomics*, 5, 405-416, (2004).

335. Llorca, O., Betti, M., Gonzalez, J. M., Valencia, A., Marquez, A. J. and Valpuesta, J. M.: **The Three-Dimensional Structure of an Eukaryotic Glutamine Synthetase: Functional Implications of its Oligomeric Structure**. *Journal of Structural Biology*, 156, 469-479, (2006).

336. Teilum, K., Hoch, J. C., Goffin, V., Kinet, S., Martial, J. A. and Kragelund, B. B.: **Solution Structure of Human Prolactin**. *Journal of Molecular Biology*, 351, 810-823, (2005).

337. Petrey, D. and Honig, B.: **Protein Structure Prediction: Inroads to Biology**. *Molecular Cell*, 20, 811-819, (2005).

338. Ginalski, K.: **Comparative Modeling for Protein Structure Prediction**. *Current Opinion in Structural Biology*, 16, 172-177, (2006).

339. Wiederstein, M. and Sippl, M. J.: **Protein Sequence Randomization: Efficient Estimation of Protein Stability Using Knowledge-Based Potentials**. *Journal of Molecular Biology*, 345, 1199-1212, (2005).

340. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R. and Thornton, J. M.: **AQUA and PROCHECK-NMR: Programs for Checking the Quality of Protein Structures Solved by NMR**. *Journal of Biomolecular NMR*, 8, 477-486, (1996).

341. Laskowaski, M. J.: **Procheck**. *Annual Review of Biochemistry*, 49, 593-626, (1980).

342. Bowie, J. U., Luthy, R. and Eisenberg, D.: **A Method to Identify Protein Sequences That Fold into a Known 3-Dimensional Structure**. *Science*, 253, 164-170, (1991).

343. Sippl, M. J.: **Boltzmann's Principle, Knowledge-Based Mean Fields and Protein Folding. An Approach to the Computational Determination of Protein Structures**. *Journal of Computer-Aided Molecular Design*, 7, 473-501, (1993).

344. Zhang, L. and Skolnick, J.: **What Should the Z-Score of Native Protein Structures be?** *Protein Science*, 7, 1201-1207, (1998).

345. Andreini, C., Banci, L., Bertini, I., Luchinat, C. and Rosato, A.: **Bioinformatic Comparison of Structures and Homology-Models of Matrix Metalloproteinases**. *Journal of Proteome Research*, 3, 21-31, (2004).

346. Chothia, C. and Lesk, A. M.: **The Relation between the Divergence of Sequence and Structure in Proteins**. *EMBO Journal*, 5, 823-826, (1986).

347. Cheng, X., Kumar, S., Posfai, J., Pflugrath, J. W. and Roberts, R. J.: **Crystal Structure of the HhaI DNA Methyltransferase Complexed with S-Adenosyl-L-Methionine**. *Cell*, 74, 299-307, (1993).

348. Labahn, J., Granzin, J., Schluckebier, G., Robinson, D. P., Jack, W. E., Schildkraut, I. and Saenger, W.: **Three-Dimensional Structure of the**

Adenine-Specific DNA Methyltransferase M.Taq I in Complex with the Cofactor S-Adenosylmethionine. *Proceedings of the National Academy of Sciences of the Uninted States of America*, 91, 10957-10961, (1994).

349. Vidgren, J., Svensson, L. A. and Liljas, A.: **Crystal Structure of Catechol O-Methyltransferase**. *Nature*, 368, 354-358, (1994).

350. Fu, Z., Hu, Y., Konishi, K., Takata, Y., Ogawa, H., Gomi, T., Fujioka, M. and Takusagawa, F.: **Crystal Structure of Glycine N-Methyltransferase from Rat Liver**. *Biochemistry*, 35, 11985-11993, (1996).

351. Djordjevic, S. and Stock, A. M.: **Chemotaxis Receptor Recognition by Protein Methyltransferase CheR**. *Nature Structural Biology*, 5, 446-450, (1998).

352. Schluckebier, G., O'Gara, M., Saenger, W. and Cheng, X.: **Universal Catalytic Domain Structure of AdoMet-Dependent Methyltransferases**. *Journal of Molecular Biology*, 247, 16-20, (1995).

353. Lauster, R., Trautner, T. A. and Noyer-Weidner, M.: **Cytosine-Specific Type II DNA Methyltransferases. A Conserved Enzyme Core with Variable Target-Recognizing Domains**. *Journal of Molecular Biology*, 206, 305-312, (1989).

354. Kagan, R. M. and Clarke, S.: **Widespread Occurrence of Three Sequence Motifs in Diverse S-Adenosylmethionine-Dependent Methyltransferases Suggests a Common Structure for These Enzymes**. *Archives of Biochemistry and Biophysics*, 310, 417-427, (1994).

355. Malone, T., Blumenthal, R. M. and Cheng, X.: **Structure-Guided Analysis Reveals Nine Sequence Motifs Conserved Among DNA Amino-**

Methyltransferases, and Suggests a Catalytic Mechanism for These Enzymes. *Journal of Molecular Biology*, 253, 618-632, (1995).

356. Clarke, S. and Banfield, K.: in *Homocysteine in health and disease* (eds. Carmel, R. and Jacobsen, D.) 63 (Cambridge Univ Pr, Cambridge, UK ; New York, 2001).

357. Bheemanaik, S., Reddy, Y. V. and Rao, D. N.: **Structure, Function and Mechanism of Exocyclic DNA Methyltransferases**. *Biochemical Journal*, 399, 177-190, (2006).

358. Wada, K., Yamaguchi, H., Harada, J., Niimi, K., Osumi, S., Saga, Y., Oh-oka, H., Tamiaki, H. and Fukuyama, K.: **Crystal Structures of BchU, a Methyltransferase Involved in Bacteriochlorophyll c Biosynthesis, and its Complex with S-Adenosylhomocysteine: Implications for Reaction Mechanism**. *Journal of Molecular Biology*, 360, 839-849, (2006).

359. Bussiere, D. E., Muchmore, S. W., Dealwis, C. G., Schluckebier, G., Nienaber, V. L., Edalji, R. P., Walter, K. A., Ladror, U. S., Holzman, T. F. and Abad-Zapatero, C.: **Crystal Structure of ErmC', an rRNA Methyltransferase which Mediates Antibiotic Resistance in Bacteria**. *Biochemistry*, 37, 7103-7112, (1998).

360. Jacobs, S. A., Harp, J. M., Devarakonda, S., Kim, Y., Rastinejad, F. and Khorasanizadeh, S.: **The Active Site of the SET Domain is Constructed on a Knot**. *Nature Structural Biology*, 9, 833-838, (2002).

361. Huang, Y., Komoto, J., Konishi, K., Takata, Y., Ogawa, H., Gomi, T., Fujioka, M. and Takusagawa, F.: **Mechanisms for Auto-Inhibition and Forced Product Release in Glycine N-Methyltransferase: Crystal Structures of Wild-Type, Mutant R175K and S-Adenosylhomocysteine-**

Bound R175K Enzymes. *Journal of Molecular Biology*, 298, 149-162, (2000).

362. Zubieta, C., Ross, J. R., Koscheski, P., Yang, Y., Pichersky, E. and Noel, J. P.: **Structural Basis for Substrate Recognition in the Salicylic Acid Carboxyl Methyltransferase Family**. *Plant Cell*, 15, 1704-1716, (2003).

363. Qiu, Y., Li, L., Hendry, T., Li, R., Taylor, D., Issa, M., Ronen, A., Vekaria, M. and White, A.: **Reconstructing the Basal Angiosperm Phylogeny: Evaluating Information Content of Mitochondrial Genes**. *Taxon*, 55, 837-856, (2006).

364. Bremer, B., Bremer, K., Chase, M. W., Fay, M. F., Reveal, J. L., Soltis, D. E., Soltis, P. S., Stevens, P. F., Anderberg, A. A., Moore, M. J., Olmstead, R. G., Rudall, P. J., Sytsma, K. J., Tank, D. C., Wurdack, K., Xiang, J. Q. Y., Zmarzty, S. and Grp, A. P.: **An Update of the Angiosperm Phylogeny Group Classification for the Orders and Families of Flowering Plants: APG III**. *Botanical Journal of the Linnean Society*, 161, 105-121, (2009).

365. Ochiai, K., Yamanaka, T., Kimura, K. and Sawada, O.: **Inheritance of Drug Resistance (and its Transfer) between *Shigella* Strains and between *Shigella* and *E. coli* Strains**. *Nihon Iji Shimpo*, 1861, 34, (1959).

366. Akiba, T., Koyama, K., Ishiki, Y., Kimura, S. and Fukushima, T.: **On the Mechanism of the Development of Multiple-Drug-Resistant Clones of *Shigella***. *Japanese Journal of Microbiology*, 4, 219-227, (1960).

367. Blanchard, J. L. and Lynch, M.: **Organellar Genes - Why do They End Up in the Nucleus?** *Trends in Genetics*, 16, 315-320, (2000).

368. Ho, M., Ryan, A. and Cummins, J.: **Cauliflower Mosaic Viral Promoter-A Recipe for Disaster?** *Microbial Ecology in Health and Disease*, 11, 194-197, (1999).

369. Richardson, A. O. and Palmer, J. D.: **Horizontal Gene Transfer in Plants**. *Journal of Experimental Botany*, 58, 1-9, (2007).