STRUCTURE AND FUNCTION OF METHYLTRANSFERASES FROM ANTIBIOTIC RESISTANCE *BACTEROIDES* OF HUMAN INTESTINE AND A STUDY ON Nm/Ng WITH CaM

Veerendra Kumar (M. Tech)

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY AT THE NATIONAL UNIVERSITY OF SINGAPORE

**DEPARTMENT OF BIOLOGICAL SCIENCES,
FACULTY OF SCIENCE,
NATIONAL UNIVERSITY OF SINGAPORE**

August, 2011

Acknowledgement

It would not have been possible to write this PhD thesis without the help and support of the kind people around me, to only some of whom it is possible to give particular mention here.

First of all, I wish to express my heartiest and sincere gratitude to my supervisor Prof J. Sivaraman for his invaluable guidance, suggestions and constant encouragement to do the research. Thanks for giving me the opportunity to work in your lab and excellent training in X-ray crystallography. Your patience, constant support, assistance and personal guidance have provided a good basis for the present thesis. I have learned a lot from you, not only the science and research, but also care and love that you share with student and others. Thank you for everything.

I wish to express my warm and sincere thanks to Professor Sheu Fwu-Shan who introduced me in the research and gave opportunity to work in his lab for two years which made the basis of this thesis. Thank you for offering me the scholarship in your lab.

It is also a great pleasure for me to thank Prof K. Swaminathan for the constructive and very informative discussion on crystallography. You offer me advice and suggestions whenever I need which overcome various stumbling blocks in my research. You always have been constant source of motivation and encouragement during my study.

I also extend my deep and sincere thanks to all the lab members of SBL4, SBL5 and other members of structure biology group for their kind help and creating friendly atmosphere. In particular, I wish to thank to thanks Dr Tang Xuhua for her help in teaching me to run different programmes. I thank Dr Jobi for all the scientific/ technical discussion. I thank Dr Mallika for her insightful comments, help and funny

jokes. I also wish to thank Lissa for her kind help in getting the materials on time and her kind help in conducting experiments. I extend my thanks to all my lab members Priyanka, Thangavelu, Abhilash, Manjeet, Nilofer, Vivek, Umar and Pankaj for their kind help through out my stay in lab. I also like to thank my wonderful new friend Magendran for being so supportive and helpful.

Finally, I would like to acknowledge the financial, academic and technical support of the Department of Biological Sciences and National University of Singapore for awarding me the Postgraduate Research Scholarship.

Table of contents

Summary

This thesis consists of two parts – Part I (Chapter 1-5): structural and functional analysis of two methyltransferases from two different species of the genus *Bacteroides*. Part II (Chapter 6-7) consist of the biophysical characterization of two neuron specific protein kinase C (PKC) substrate proteins Neuromodulin (Nm) and Neurogranin (Ng), and its structure of the IQ domain in complex with Calmodulin (CaM).

Methylation is important for various cellular activities. More often methyltransferases are involved in cellular and metabolic functions. To-date there is no report of any methyltransferase structure from human intestine antibiotic resistance pathogenic strains B. *thetaiotaomicron VPI-5482* and *B. vulgatus ATCC-8482.* Chapter 1 provides general introduction for this part. Chapter 2 report the expression, purification and crystallization of methyltransferases BT_2972 and BVU_3255 from the species *B. thetaiotaomicron VPI-5482* and *B. vulgatus ATCC-8482* respectively. These two MTases were cloned, over expressed and purified to yield approximately 120 mg of each protein from 1 L of the culture. *Apo* BT_2972 and BVU_3255 and their complexes with AdoMet/AdoHcy were crystallized in four different crystal forms using hanging drop vapour diffusion method. These crystals diffract to a resolution ranging from 2.9 to 2.2 Å.

Chapter 3, report the crystal structure of an AdoMet dependent methyltransferase BT_2972 and its complex with AdoMet and AdoHcy from *B. thetaiotaomicron VPI-5482* strain along with their isothermal titration calorimetric studies. The active site of *apo* BT_2972 and structures of its complex with AdoMet and AdoHcy revealed significant conformational changes which resulted in open and closed forms to regulate the movement of cofactors and to aid its interaction with the substrate.

Based on our analysis, supported with literature, we suggest that BT_2972 is a small molecule methyltransferase and might catalyze two O-methylation reaction steps in the ubiquinone biosynthesis pathway.

BVU_3255 belongs to an AdoMet- dependent methyltransferase. Chapter 4 report the crystal structure of *apo* BVU_3255, and its complexes with AdoMet and AdoHcy, which revealed a typical class I Rossmann Fold Methyltransferase. The isothermal titration calorimetric study showed that both AdoMet and AdoHcy bind with equal affinity. The structural and sequence analysis suggested that BVU_3255 is a small molecule methyltransferase and might involve in methylating the intermediates in ubiquinone biosynthesis pathway. The conclusions and future directions are provided in chapter 5.

In part II, the chapter 6 of this thesis discuss the biophysical characterization and structure of two neuron specific protein kinase C (PKC) substrate proteins Neuromodulin (Nm) and Neurogranin (Ng) fragments in complex with Calmodulin (CaM). The ubiquitous Ca^{2+} sensing protein CaM is also under go methylation at the Met residues as a part of post translational modification.

Biophysical studies clearly showed the unfolded state of Ng/Nm in the solutions. These classes of proteins are known as intrinsically unstructured protein and they are highly flexible and lack the globular fold. However they are functionally active proteins *in vivo* and *in vitro* conditions. Further we report the crystal structure of CaM binding motif (IQ motif) of Nm and Ng, in complex with *apo* and Ca^{2+}/CaM. In the presence of IQ peptides, Ca^{2+}/CaM adopt an unusual conformation, hither to not observed for any CaM structure. Moreover the crystal structure of *apo* CaM and IQ peptides showed that CaM adopts an extended conformation and the IQ peptides bind to C lobe of the CaM. In addition we have carried out interaction studies using

isothermal calorimetry. The ITC results and structural studies showed that only a small motif of full length Nm and Ng is sufficient to make interactions with CaM. Further the present studies clearly explains the reason why Nm and Ng shows 1) novel CaM binding properties; 2) low affinity for Ca^{2+}/CaM and 3) higher affinity for *apo* CaM. The present crystal structure is the first report of any neuron specific intrinsically unstructured proteins and explains how the unstructured proteins gain structure upon binding with its partners. Chapter 7 provides the conclusion and future direction for part II of this thesis.

List of Tables

List of Figures

List of abbreviations

AdoHcy/SAH	S-adenosyl-L-homocysteine
AdoMet/SAM	S-adenosyl-L-methionine
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
B. thetaiotaomicro	*Bacteroides thetaiotaomicron*
B. vulgatus	*Bacteroides vulgatus*
CaM	Calmodulin
CCD	charged coupled device
CNS	crystallography and NMR system
DEAE	diethyl aminoethyl
DLS	dynamic light scattering
DNA	deoxyribonucleic acid
DNAse	Deoxyribonuclease
DTT	Dithiothreitol
EC	Enzyme commission
E. coli	*Escherichia Coli*
EDTA	ethylenediamine tetraacetic acid
FPLC	fast performance liquid chromatography
GAP-43	Growth Associated Protein-43
Hepes	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IPTG	isopropyl thio-galactoside
ITC	isothermal titration calorimetry
LB	Luria-Bertani
MAD	multiwavelength anomalous dispersion

M. tuberculosis	*Mycobacterium tuberculosis*
MALDI-TOF	Matrix Assisted Laser Desorption Ionization –Time of Flight
MS	mass spectrometry
NCBI	National Center for Biotechnology Information
NCS	non crystallographic symmetry restraints
Ni-NTA	nickel-nitrilotriacetic acid
Ng	Neurogranin
Nm	Neuromodulin
NMR	nuclear magnetic resonance
OD	optical density
ORF	open reading frame
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol
PKC	protein kinase C
RFM	Rossmann fold methyltransferase
RMSD	root mean square deviation
RNA	ribonucleic acid
RNAse	Ribonuclease
rRNA	Ribosomal ribonucleic acid
SAD	single wavelength anomalous dispersion
SDS-PAGE	sodium dodecyl sulfate - polyacrylamide gel electrophoresis
SeMet	seleno-L-methionine
TLC	thin-layer chromatography
UbiA	4-hydroxybenzoate polyprenyltransferase,

UbiB	Ubiquinone biosynthesis monooxygenase UbiB
UbiC	Chorismate--pyruvate lyase
UbiD	3-polyprenyl-4-hydroxybenzoate carboxy-lyase
UbiE,	Ubiquinone/menaquinone biosynthesis methyltransferas UbiE
UbiF	Ubiquinone biosynthesis monooxgenase UbiF
UbiG	Ubiquinone biosynthesis SAM-dependent O-methyltransferase
UbiH,	Ubiquinone biosynthesis monooxgenase UbiH
WT	wild-type

Publications

1. **Veerendra Kumar** and Sivaraman J: A Conformational Switch in the Active Site of BT_2972, a Methyltransferase from an Antibiotic Resistant Pathogen *Bacteroides thetaiotaomicron* Revealed by its Crystal Structures. *PLoS One* 2011; 6(11): e27543 .

2. **Veerendra Kumar** and Sivaraman J: Structural and functional characterization a methyltransferase of ubiquinone biosynthesis pathway from antibiotic resistance pathogen *Bacteroides vulgatus ATCC 8482* of human intestine**.** *J Struct Biol.* 2011 Dec;176(3):409-13. Epub 2011 Aug 22.

3. **Veerendra Kumar**, Nagarajan Mallika and J.Sivaraman: Expression, Purification, Characterization and Crystallization of Methyltransferases BT_2972 and BVU_3255 of Antibiotic Resistant Pathogen Genus *Bacteroides* from the Human Intestine. *Acta Cryst.* (2011). F**67**, 1359-1362

4. Bokhari H, Smith C, **Veerendra Kumar**, Sivaraman J, Sikaroodi M, Gillevet P. Novel fluorescent protein from Hydnophora rigida possess cyano emission. *Biochem Biophys Res Commun*. **2010 Jun 4;396(3):631-6**.

5. **Veerendra Kumar**, Vishnu Priyanka Reddy Chichili, J. Seetharaman and J.Sivaraman. Structural basis for the role of intrinsically unstructured proteins neuromodulin and neurogranin in neurons. (***Manuscript under preparation***).

6. **Veerendra Kumar**, Vishnu Priyanka Reddy Chichili, and J.Sivaraman. A Novel Conformation of Calmodulin in the Presence of Ca^{2+}. (***Manuscript under preparation***).

7. Vishnu Priyanka Reddy Chichili, **Veerendra Kumar** and J.Sivaraman. A Method to Trap the Transiently Interacting Protein Complex for Structural Studies. (***Manuscript under preparation***).

Part I

Chapter 1

General Introduction

1.1 Introduction

1.1.1 Methyltransferase

Methylation of key biological molecules and proteins plays important roles in numerous biological systems, including signal transduction, biosynthesis, protein repair, gene silencing and chromatin regulation (Cheng, 1995). Methyltransferases (EC 2.1.1) catalyze the transfer of a methyl group from a donor molecule to variety of acceptor molecules. Methyltransferases use a reactive methyl group bound to sulfur in AdoMet as the methyl donor. In these reactions, methyl group is transferred from an AdoMet molecule to acceptor molecule yielding S-adenosylhomocysteine (AdoHCy or SAH) and a methylated target molecule. Methyltransferase are abundant, highly conserved across phylogeny, biologically important class of enzymes. Many of these reactions are very important for the proper functioning of life. The lack of the gene product that performs these methyl transfer reactions is sufficient to stop the normal functioning of organisms. 95% of methyltransferase use *S*-adenosyl-L-methionine (AdoMet or SAM) as the methyl donor. AdoMet is the second most commonly used enzymatic cofactor after ATP. The preference for AdoMet over other methyl donor (such as floate) is because of favourable energetic (Cantoni, 1975). Aberrant levels of SAM have been linked to many abnormalities, including Alzheimer's, depression, Parkinson's, multiple sclerosis, liver failure and cancer (Schubert *et al.*, 2003) . The comparison shows the importance of methyltransferase reactions in living organism. Thus it requires immediate need to characterize the methyltransferases structurally and functionally.

AdoMet is produced from methionine and ATP in the reactions catalyzed by S-adenosylmethionine synthetase (Figure 1.1). AdoMet serves as precursors for

numerous methyl transfer reactions. These methyltransferase enzymes are called as the AdoMet-dependent methyltransferases.

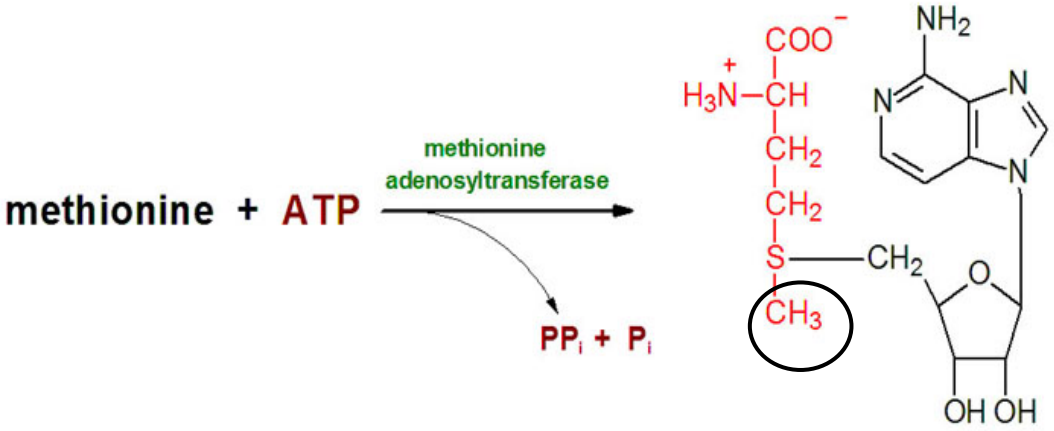


Figure 1.1: Biosynthesis reaction of AdoMet. A condensation of ATP and methionine is catalyzed by methionine adenosyltransferase yields AdoMet. The transferable methyl group is marked by circle. This figure is adapted from http://themedicalbiochemistrypage.org/amino-acid-metabolism.html.

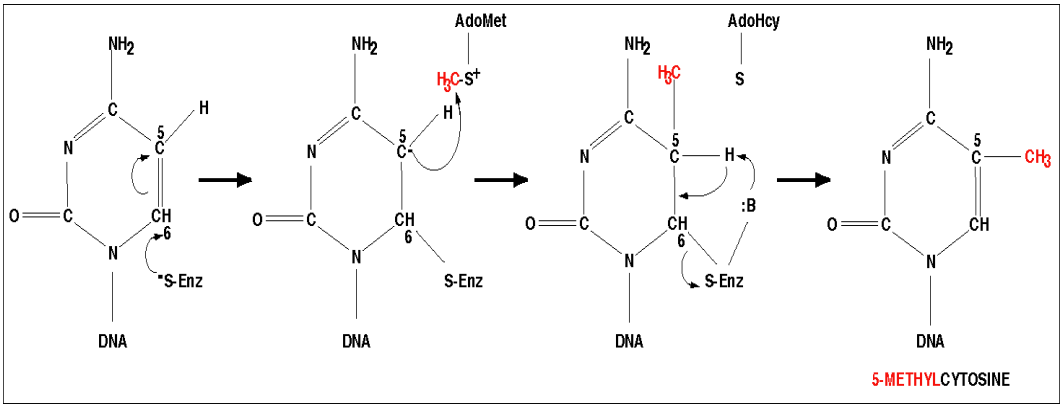


Figure 1.2: General methylation reaction mechanism catalyzed by methyltransferase. Adapted from http://edoc.hu-berlin.de/dissertationen/aguirre-arteta-ana-2000-06-28/HTML/aguirre-ch1.html

General methylation mechanism involves catalytic attack of a nucleophile (carbon, oxygen, nitrogen, sulfur or even halides) from the acceptor molecules on a methyl group of AdoMet. All methylation reaction takes place with direct transfer of methyl group to the acceptor molecule with inversion symmetry in an S_N2 like mechanism. It also requires that a proton be removed before or after methyl transfer (Woodard *et al.*, 1980). Final reaction products are methylated derivative of acceptor molecule and AdoHcy (SAH). Methylation reactions produce derivatives with less hydrophilicity than unmethylated counterpart (Figure 1.2).

AdoMet dependent methyltransferase methylate a wide variety of molecular targets (methyl group acceptor), including DNA, RNA, proteins, hormones, neurotransmitters and small molecules, thereby modulating important cellular and metabolic activities. These enzymes are classified according to the target they methylate like DNA methyltransferase, RNA methyltransferase, small molecule methyltransferase and so on.

The first structure of AdoMet dependent methyltransferase was solved for DNA C5-cytosine methyltransferase M.Hhai (Cheng *et al.*, 1993). Since then many methyltransferase structures have been solved and characterized. Structurally, there are five different class of AdoMet dependent methyltransferases (Schubert *et al.*, 2003). These methyltransferase exhibits enzyme analogy i.e. structurally distinct but catalyze the same reaction.

Class I:

This class of methyltransferase consist of a seven stranded β sheets (β3↓β2↓β1↓β4↓β5↓β7↑β6↓) and followed by α helixes on both side of sheets to make αβα sandwich. This fold is quite similar to the Rossmann-fold of NAD (P) binding domain of oxidoreducatases. The first β strands typically end in GXGXG motif (x-

any amino acids). The other conserved amino acid position is an acidic residue found at the end of β2 stand. This residue makes the hydrogen bond with ribose moiety of AdoMet. Class I methyltransferase are involved in gene regulation, in protein expression, in mutation repairing and in DNA protection from restriction enzymes and so on.

Classe II:

This class of methyltransferase has a long central antiparallel β strand which is flanked by groups α helices. AdoMet bound in shallow groove and makes hydrogen bond with conserved RXGY motif (Dixon *et al.*, 1996).

Class III:

The active site in this structural class of methyltransferase is hidden into a cleft between two αβα domains. Each domain contains 5 β strands and 4 α helices. Like class I methyltransferase, GXGXG motif occurs at C terminal of first β strands. But in this class of methyltransferase, this motif does not interact with AdoMet. AdoMet is buried between two domains in the active site.

Class IV:

This family consists of SPOUT family (**SpoU** and **T**rmD families) of RNA methyltransferase. They found to contain 1) six stranded parallel β sheet which is flanked by seven α helices, 2) active site is located close to the subunit interface of homodimer. Residues from both the monomers form the active site region.

Class V:

This family is known as SET (**S**u(var)3-9, **E**nhancer of Zeste, **T**rithorax) -domain proteins . Several proteins from this family shown to methylate lysine in flexible tails of histones or in Rubisco (Ribulose-1,5-bisphosphate carboxylase oxygenase). Structurally, this family consist of a series of 8 curved β strands forming 3 small

sheets.

The example and structure of each class of methyltransferases discussed above are shown in figure 1.3. Structural and sequence alignment data may enable the identification of similarities of structure and sequence conservation. Several structural features and sequence conservation separates small and macromolecule methyltransferases.

1.1.2 Macromolecule versus small molecule methyltransferase

Sequence analysis of small molecule methyltransferases suggests that they share a sequence identity of 15–20%. However the structural comparison showed that they share similarity more than the core Rossmann fold. Methyltransferases typically consist of well-conserved AdoMet-binding domains responsible for cofactor binding and methylation; and a second highly variable substrate-binding domain responsible for the substrate binding. The substrate binding domain varies in shape and size to accommodate the different substrate and is also indicative of the group to which these proteins belong to. Most of the macromolecular methyltransferase like DNA, RNA or proteins methyltransferase are found to contain substrate binding region as separate domain. These macromolecule methyltransferase are supplemented with additional substrate binding domain. In addition to the core Rossman fold, the macromolecular methyltransferase found to have major modification in the form of additional secondary structure element at C terminal. These two major structural features are found to be absent in small molecules methyltransferases. However, majority of small molecule methyltransferase has additional amino acids at N terminal region in the form of two α helices. Many small molecule methyltransferase have an active site cover formed by several core fold inserts. Small molecule methyltransferase also

found to have insertions between β5 and α7; and β6 and β7. These structural features can be used as an aid to distinguish the macromolecule and small molecule methyltransferase (Martin & McMillan, 2002).

1.1.3 Functional role of methyltransferases

Methylation reactions are very important for maintenance of life. Methyl transfer is an important biochemical reaction in the metabolism pathway of many drugs and xenobiotic compounds. For example, biosynthesis of endogenous compounds such as epinephrine which is a hormone and a neurotransmitter involves methylation of the primary distal amine of noradrenaline in a reaction catalyzed by phenylethanolamine N-methyltransferase (Goodall & Kirshner, 1958). N-, O-, and S-methylation have been reported to occur in many drugs. Some methylation reactions may terminate biological activity of the compounds.

DNA methylation plays an important role in gene transcription, gene expression, gene activation, and mutation repair in the cells. Methylations influence the gene expression by affecting the interactions between DNA and chromatin proteins as well as specific transcription factors. In addition, it has been noted that during development, tissue-specific genes undergo demethylation in their tissue of expression. Further, the X chromosome inactivation during development is accompanied by de novo methylation (Razin & Cedar, 1991). DNA methyltransferase binds to the DNA helix by base flipping and catalyzes the transfer of the methyl group to adenosine and cytosine (Cheng & Roberts, 2001). In addition, adenosine or cytosine methylation is part of the restriction modification system of many bacteria.

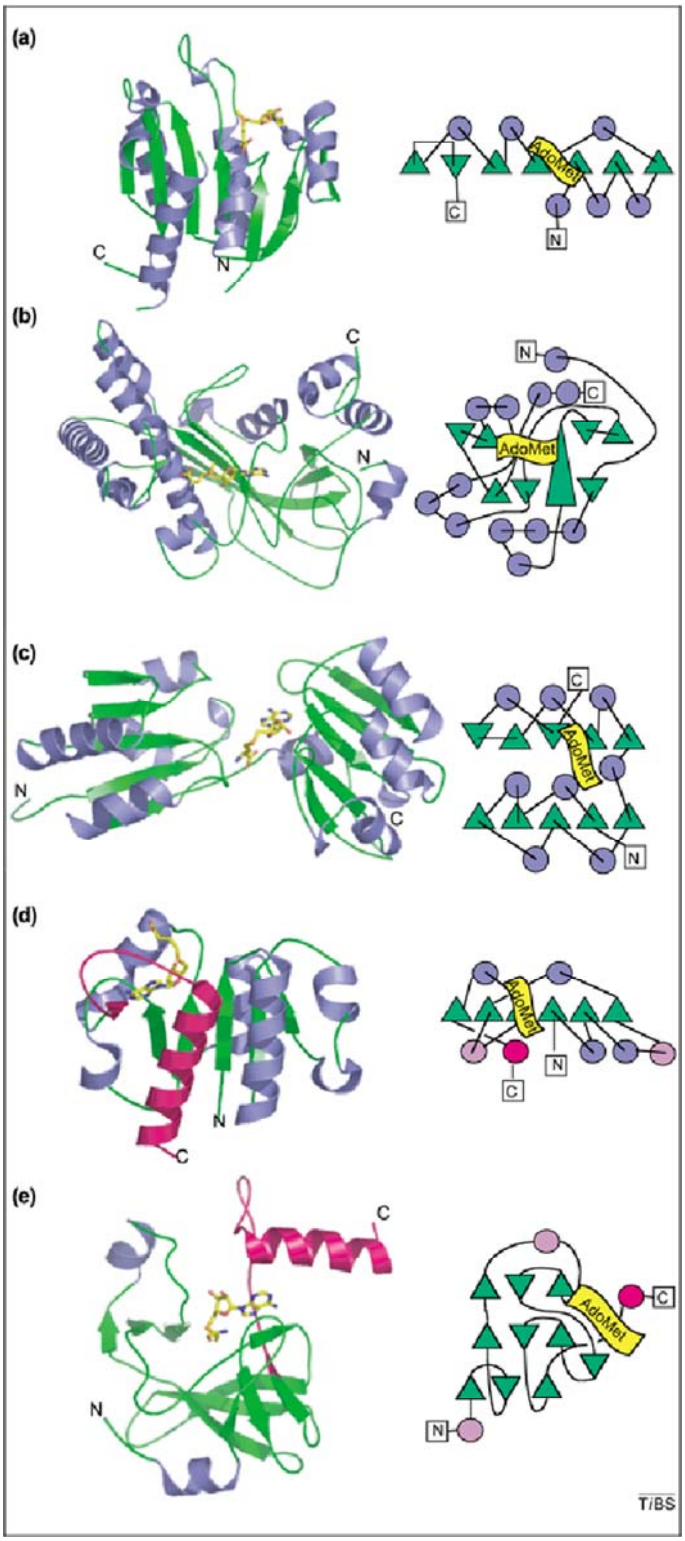


Figure 1.3: 3-Dimensional structure of five classes of AdoMet-dependent methyltransferase. In each class a representative structure and topology diagram is given. A) class I: enzyme M.HhaI (PDB code-6MHT), B) class II: reactive domain of methionine synthase (PDB code -1MSK), C) class III: the bilobal structure of CbiF (PDB code – 1CBF), D) class IV: enzyme Yibk (PDB code – 1MXI), and E) class V: histone lysine N-MTase family (PDB code – 1O9S). The figure adapted from (Schubert *et al.*, 2003).

Post translational modification of proteins in form of methylation is very common phenomenon in eukaryotes. Protein methylation typically takes place on arginine, lysine, histidine, glutamine, and asparagine amino acid residues in the protein sequence. Protein methylation has been most-studied in the histones catalyzed by enzymes known as histone methyltransferases. Histones that are methylated on certain residues can act epigenetically to repress or activate gene expression (Grewal & Rice, 2004). In protein O-methylation forming methyl esters on carboxyl groups are reversible and can modulate the activity of the target protein. N-methylations occurring on nitrogen atoms in N-terminal and side-chain positions are generally irreversible. These methylation reactions can modify the amino acid residues in protein and thus property as well as function of a protein (Clarke, 1993). The calcium-binding protein calmodulin (CaM) which regulates the activity of several enzymes by Ca^{2+} signal, undergoes N-methylation as one of the several post-translational modifications (Cobb and Roberts 2000). Methylation alters the function of CaM *in vivo*, and affects the regulation of nicotinamide adenine dinucleotide kinase and other enzymes. The part II of this thesis is discussing about the CaM and its interactions with neuron specific substrate proteins Ng/Nm, a different functional aspect of CaM

1.1.4 Antibiotic resistance and methyltransferase

Methyltransferases have been implicated to induce the antibiotic resistance of host cell. Antibiotic-producing bacteria protect themselves from the toxic effects of antibiotics by employing methyltransferase to methylate specific ribonucleotides in antibiotic-binding sites of the ribosome. Methylation of the 30S ribosomal subunit RNA (16S rRNA) is a significant mechanism of resistance to ribosome-targeting

antibiotics. Methylation of riboneucleotide shown to interfere the antibiotic binding without much interference with other functions of the ribosome (Poehlsgaard & Douthwaite, 2005). The 16S rRNA resistance methyltransferases act at riboneucleotide in close proximity to their respective antibiotic binding site and thus the sterically methyl group addition blocks antibiotic binding. For instance, two families of AdoMet-dependent aminoglycoside-resistance methyltransferases act upon 16S rRNA to produce either an N7-methyl G1405 (m7G1405) or N1-methyl A1408 (m1A1408) modification. The resistance spectrum conferred by m7G1405 is limited to the 4,6-DOS aminoglycosides (e.g. kanamycin and gentamicin). The m1A1408 modification confers a broad resistance spectrum that includes examples of the 4,6-DOS (e.g. kanamycin but not gentamicin) and 4,5-DOS aminoglycoside groups and also apramycin (Husain *et al.*, 2011, Husain *et al.*, 2010, Macmaster *et al.*, 2010).

The bacterial species from genera *Bacteroides* are shown to be resistant to a wide variety of antibiotics including β-lactams, aminoglycosides, erythromycin and tetracycline. Resistance to multiple antibiotics has been increasing in *Bacteroides* species for decades (Vedantam, 2009). This high level of antibiotic resistance has prompted concerns that *Bacteroides* species may become a reservoir for resistance in other, more highly-pathogenic bacterial strains. For this thesis work we have selected two methyltransferase from two different *Bacteroides* bacteria.

1.2 Bacteroides

The genus *Bacteroides* is consisting of Gram-negative, non-endospore-forming, anaerobes, and rod-shaped bacteria bacillus bacteria. Some of the species are motile and some are non-motile. Furthermore, *Bacteroides* species also stimulates the gut lining to produce fucosylated glycans, angiogenesis (formation of blood

vessels) in the newborn epithelium, enhancing human uptake of nutrients. So the *Bacteroides* bacteria live in symbiotic relationship with the human host (Figure 1.4).

Bacteroides were described in 1898 for the first time (Veillon, et al 1898). *Bacteroides* membranes contain sphingolipids which is not found in any other bacterial membrane. Some bacteroides also contain meso-diaminopimelic acid in their membrane.

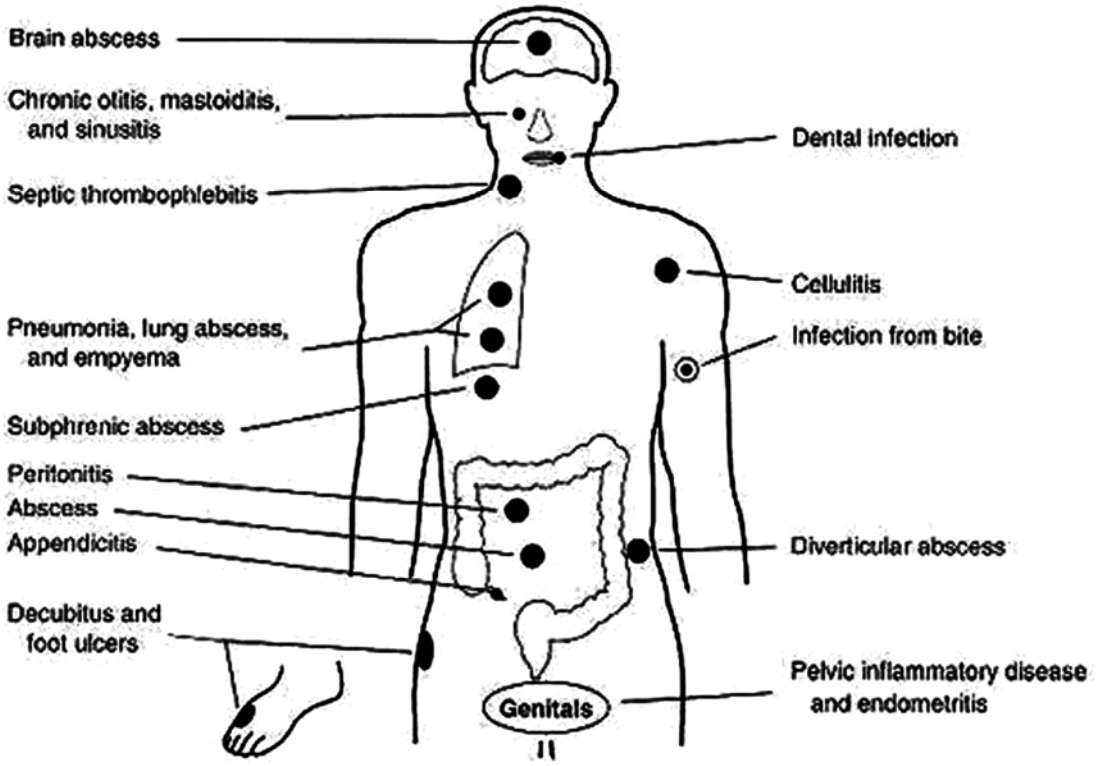


Figure 1.4: Common sites (●) of *Bacteroides* and other anaerobic bacterial infections human. Adapted from Anaerobic Infections in Humans by Sydney M. Finegold Academic Press Inc. 1989.

Bacteroides are commonly found in the human intestine where they have a symbiotic host-bacterial relationship with humans, but are also found to be opportunistic pathogens (McCarthy *et al.*, 1988). Roughly, 10^{10}-10^{11} cells per gram of human feces have been reported. The *Bacteroides* bacteria show resistance towards different antibiotics such as clindamycin, chloramphenicol, carbenicillin, lincomycin and

tetracycline (Bodner *et al.*, 1972,Rashtchian, 1982 #551). They are involved in digesting the otherwise non digestible food items and producing valuable nutrients and energy that the human body needs. Their main source of energy is polysaccharides from plant sources but whenever available they also can use simple sugars. They are also found to be involved in fermentation of carbohydrates, nitrogenous substances, and biotransformation of bile acids and other steroids. Hence *Bacteroides* are also classified as saccharolytic meaning that they obtain carbon and energy by hydrolysis of carbohydrate molecules(McCarthy *et al.*, 1988). They are mainly involved in degradation of polysaccharides from plant fibers, such as cellulose, xylan, arabinogalactan, and pectin, and vegetable starches such as amylose and amylopectin. *Bacteroides* are also known to produce several exoenzymes like DNAse, collagenase, neuraminidase, heparinase and some proteases. These enzymes assist the bacteria in the invasion of host tissues.

Since *Bacteroides* species colonize in the gut, they offer help to the host by excluding potential pathogens from colonizing in the gut. Although *Bacteroides* species are anaerobic, they are aerotolerant and thus can survive in the abdominal cavity (Xu *et al.*, 2007). However, if the other parts of the body (including the central nervous system, the head, the neck, the chest, the abdomen, the pelvis, the skin, and the soft tissues) get infected by *Bacteriodes*, they can cause or exacerbate abscesses and other infections (Xu *et al.*, 2003). Their entry into the host involves the following steps- disruptions of the intestinal wall, bacterial flora infiltrate the cavity, gram negative aerobes e.g. *E. coli* initiate the preliminary tissue destruction and reduces the oxidation-reduction potential of the oxygenated tissue which favors the anaerobe growth. Finally, anaerobic *Bacteroides* start growing, and dominate the infection. For example *B. fragilis* behaves like opportunistic human pathogens. It can cause

peritoneal cavity infection, gastrointestinal surgery, and appendicitis via abscess formation. *Bacteroides* have been known to be involved in cases of meningitis and shunt infections, especially in children. Hence they inhibit phagocytosis, and inactivate beta-lactam antibiotics. *Bacteroides* species are resistant to a wide variety of antibiotics like β-lactams, aminoglycosides, erythromycin and tetracycline (Salyers *et al.*, 2004).

Bacteroides species have an outer membrane, a peptidoglycan layer, and a cytoplasmic membrane. Outer membranes contain a mixture of long-chain fatty acids, mainly straight chain saturated, anteiso-methyl, and iso-methyl branched acids. They produce acetic acid, iso valeric acid, and succinic acid as their main by-products of their anaerobic respiration. Some of the *Bacteroides* species have been shown to bind to polysaccharides with their outer membrane receptor system before pulling the polysaccharides into the periplasm for monosaccharide degradation. Their DNA-base composition of *Bacteroides* is about 40-48% G-C.

Ability of *Bacteroides* bacteria to process the complex molecules into simpler compounds make them to play very important role in the ecosystem. The complex molecules degradation into simple molecule makes them to usable by the human host as well as the *Bacteroides*.

Physiological analysis of *Bacteroides* showed considerable heterogeneity in terms of to their biochemical properties. Based on the several phylogenetic analysis techniques like physiological characteristics, serotyping (Lambe, 1974), bacteriophage typing (Booth *et al.*, 1979), lipid analysis (Miyagawa *et al.*, 1978), oligonucleotide cataloging, and 5S - 16S rRNA sequence comparisons (Paster *et al.*, 1994, Weisburg *et al.*, 1985), the original *Bacteroides* members have been partitioned into three genera: *Bacteroides* (Shah & Collins, 1989), *Prevotella(Shah &*

Collins, 1990), and Porphyromonas (Shah & Collins, 1988) (Figure 1.5). The *Bacteroides* are found predominantly in the colon of mammals, while the *Prevotella* and *Porphyromonads* generally are associated with the oral cavity and rumen.

This classification restricts the *Bacteroides* to ten species: *B. fragilis, B. thetaiotaomicron, B. vulgatus, B. ovatus, B. distasonis, B. uniformis, B. stercoris, B. eggerthii, B. merdae, and B. caccae.*

For this study, we have selected the two species- *B. thetaiotaomicron VPI-5482* and *B. vulgatus ATCC 8482*. The AdoMet dependent methyltransferase BT_2972 (*B. thetaiotaomicron VPI-5482*) and BVU_3255 (*B. vulgatus ATCC 8482*) sequence analysis showed that they are involved in methylating the intermediates (2-polyprenyl-6-hydroxyphenol and 2-polyprenyl-3-methyl-5-hydroxy-6-methoxy-1, 4-benzoquinone) of Ubiquinone biosynthesis pathway. To the extent of our understanding there is no report as yet showing BT_2972 and BVU_3255 confer any antibiotic resistance. In the next section, we discuss about the ubiquinone/menaquinone biosynthesis pathway in bacteria.

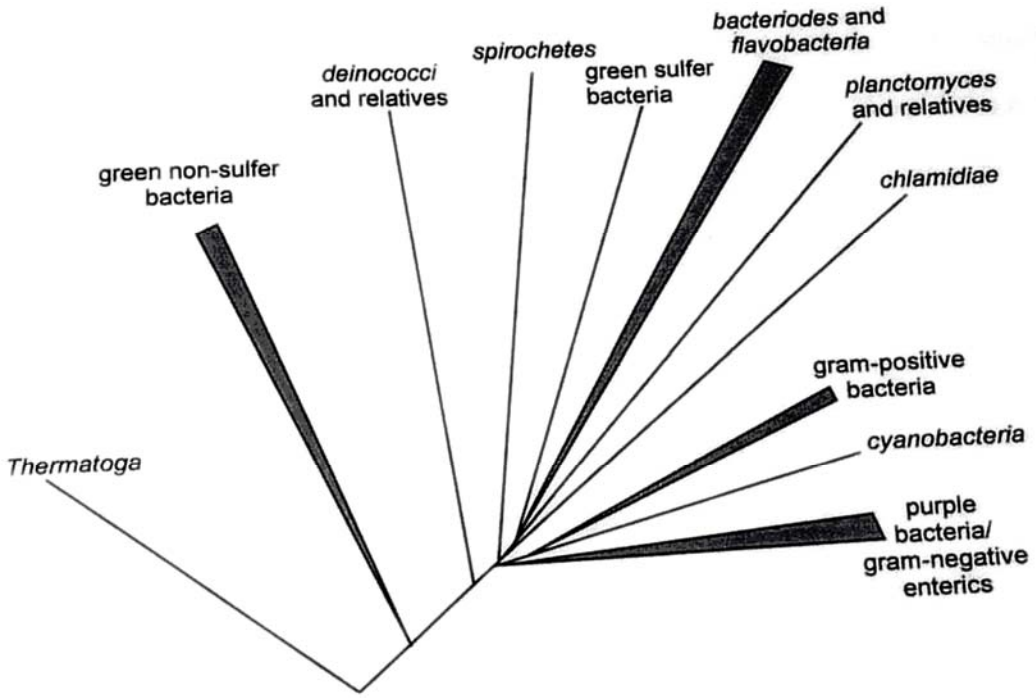


Figure 1.5: Phylogenetic tree of Eubacteria based on 16s rRNA sequence comparisons. The evolutionary relationships between prokaryotic phyla are shown. Branch lengths on the tree represent evolutionary distance. Adapted from (Weisburg *et al.*, 1985)

1.3 Ubiquinone biosynthesis pathway

Quinones are widely distributed in nature. They are best known as lipid-soluble components of membrane-bound electron transport chains, but in animal cells ubiquinone is found not only in the inner mitochondrial membrane but also in the endoplasmic reticulum, Golgi apparatus, lysosomes, peroxisomes, and in the plasma membrane. This distribution suggests that ubiquinones may be involved in a number of biological processes beyond the respiratory electron transport. Bacterial respiratory quinones can be divided into two groups. The first one comprises a benzoquinone termed ubiquinone or coenzyme ubiquinone. The second group contains

naphthoquinones – menaquinone and demethylmenaquinone (DMK) (Meganathan, 2001). In this study the ubiquinone is more relevant and we discuss this in detail.

The isoprenoid quinone ubiquinone (coenzyme Q, Figure 1.6) is an essential component in the respiratory electron transport chain of both eukaryotes and most prokaryotes, with the exception of the gram-positive bacteria and the blue-green algae (cyanobacteria). Chemically, ubiquinone is a 1,4-benzoquinone consist of quinone chemical group, and a variable isoprenyl chain (6-10) in its tail (Lee *et al.*, 1997).

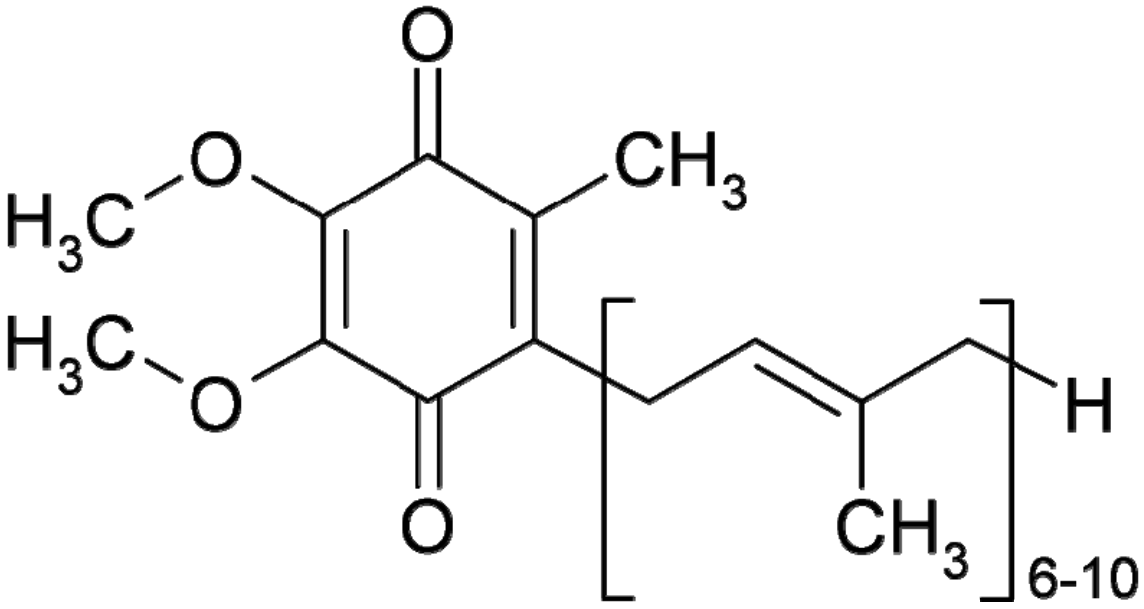


Figure 1.6: Chemical structure of Ubiquinone (Coenzyme Q).

Animal cells synthesize only ubiquinone, but MK is obtained from the diet. Most Gram-positive bacteria and anaerobic Gram-negative bacteria contain only MK, whereas the majority of strictly aerobic Gram-negative bacteria contain exclusively ubiquinone. Both types of isoprenoid quinones are found only in facultative anaerobic Gram-negative bacteria. Archaea bacteria lack ubiquinone. A nearly complete set of

orthologs of the ubiquinone biosynthesis genes is apparent in all cyanobacterial genomes, in spite of the apparent absence of ubiquinone in these organisms. The possibility for these genes to be involved in plastoquinone (rather than ubiquinone) biosynthesis in cyanobacteria has been proposed. The biosynthesis of ubiquinone is studied in great detail in bacteria and yeast but only to a limited extent in animal tissues (Meganathan, 2001).

The biosynthesis of ubiquinone and menaquinone begins in shikimate pathway via chorismate in bacteria or tyrosine in higher eukaryotes. Here we focus the ubiquinone biosynthesis in the bacteria (Figure 1.7). The biosynthesis of ubiquinone includes at least nine reactions. The benzene ring derived from chorsimate intermediate of shikimate pathway. The prenyl side chain of ubiquinone is derived from prenyl diphosphate and methyl groups are derived from AdoMet. The first committed step is the formation of 4-hydroxybenzoate from chorismate in a reaction catalyzed by ubiA enzyme. Three hydroxylation reactions introduce hydroxyl groups at positions C-6, C-4, and C-5 of the benzene ring of ubiquinone. They are catalyzed by *ubiB, ubiH,* and *ubiF* enzymes respectively. Ubiquinone synthesis requires two O methylation steps and one C methylation. The C-methylation reaction is catalyzed by UbiE methyltransferase and O-methylation reaction is catalyzed by UbiG methyltransferase. The same O-methyltransferase with dual specificity (UbiG) catalyzes both O-methylation steps in ubiquinone biosynthesis(Poon *et al.*, 1999). Thus ubiquinone biosynthesis pathway involved two methyltransferase encoded by *UbiE* and *UbiG* gene (Figure 1.7).

In summary, the adult human intestine is dominated by the Bacteria from the genus *Bacteroides*. In general *Bacteroides* habitats in symbiotic relationship, but are also found be opportunistic pathogens. The *Bacteroides* bacteria show the antibiotic

resistance for different antibiotics such as clindamycin, chloramphenicol, carbenicillin, lincomycin and tetracycline. However, the protein under study BT_2972 and BVU_3255 have not been shown yet to confer any antibiotic resistance.

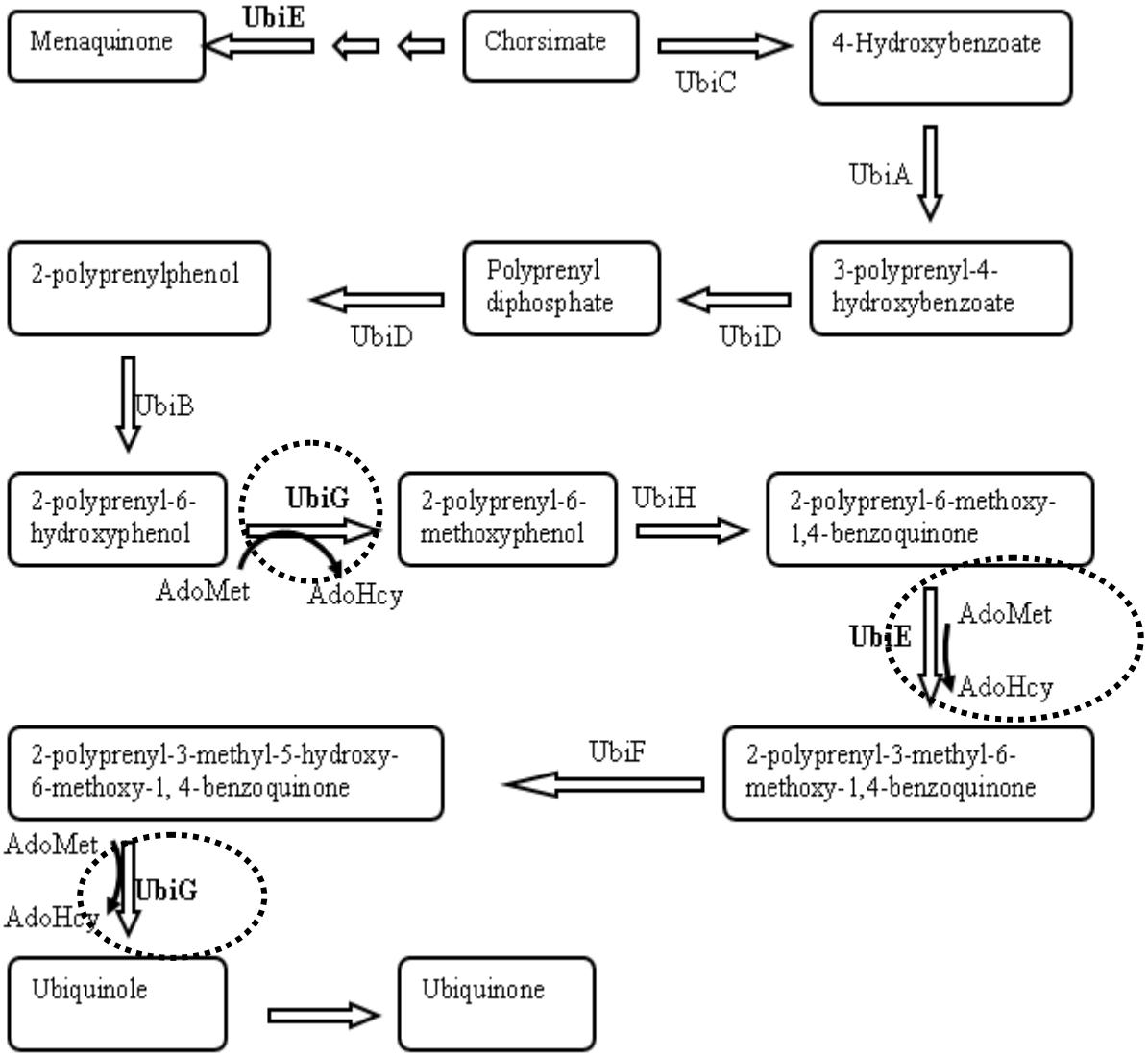


Figure 1.7: The schematic representation of the proposed biosythesis of ubiquinone in bacteria. There are two methylation reactions in the pathway catalyzed by UbiE and UbiG (shown by dotted ellipsoid). This figure is prepared based on the literature with possible intermediates and possible enzymes involved in this biosynthesis (Meganathan, 2001, Poon *et al.*, 1999).

1.4 Aims and Objective

In this study we have selected two methyltransferase from *Bacteroides* bacteria *B. thetaiotaomicron VPI-5482*, and *B. vulgatus ATCC 8482*. Complete genome sequence of these two species are available (Xu *et al.*, 2003, Xu *et al.*, 2007). *B. thetaiotaomicron VPI-5482* genome is 6.26-Mb long while that for *B. vulgatus ATCC 8482* is *5.3* MB. The target methyltransferase BT_2972 (*B. thetaiotaomicron VPI-5482*) and BVU_3255 (*B. vulgatus ATCC 8482*) are class 1 AdoMet dependent methyltransferase. Sequence identity between these two enzymes is 59% which indicates that these two enzymes are evolutionary related. Based on the sequence analysis we proposed that these two enzymes catalyze the two O-methylation reactions in the ubiquinone biosynthesis pathway. The objective of this study is to understand the structure and function of these two MTases through the following experimental approaches-

1. To clone the gene *BT_2972* and BVU_3255 into the expression vector followed by protein expression and purification.
2. Biophysical characterization of the recombinant proteins using different techniques like mass spectrometry analysis, peptide mass finger printing, size exclusion chromatography and Dynamic light scattering (to access the oligomeric state of the protein).
3. Isothermal calorimetric analysis to characterize the thermodynamics of ligand (SAM and SAH) binding.
4. To determine the *Apo* and complex structures of these proteins with ligands (AdoMet/SAM and AdoHcy/SAH) using X-ray crystallography.
5. Structure function analysis to understand the role of these enzymes in methylation.

Chapter 2

Purification, Crystallization and Diffraction studies of Methyltransferases BT_2972 and BVU_3255 of Antibiotic Resistant Pathogen Genus *Bacteroides* from the Human Intestine

2.1 Introduction

The adult human intestine is dominated by bacteria from the genus *Bacteroides*. In general *Bacteroides* habitats in symbiotic relationship, but are also found to be opportunistic pathogens (McCarthy *et al.*, 1988). The *Bacteroides* bacteria show resistance towards different antibiotics such as clindamycin, chloramphenicol, carbenicillin, lincomycin and tetracycline (Bodner *et al.*, 1972,Rashtchian, 1982 #551). Recently, the complete genome of two species *B. thetaiotaomicron VPI-5482 and B. vulgatus ATCC 8482* of the genus *Bacteroides* has been sequenced (Xu *et al.*, 2007, Xu *et al.*, 2003). An open reading frame (ORF) in the chromosomal genome of *B. thetaiotaomicron VPI-5482* encodes a putative methyltransferase BT_2972 (accession no NP_811884.1). Similarly BVU_3255 (accession no. YP_001300506.1) is a putative methyltransferase from *B. vulgatus ATCC 8482.* Previously the *apo* structure of BT_2972 (pdb code 3F4K) and BVU_3255 (pdb code 3E7P) and that of a closely related protein Q5LES9_BACFN-SAM complex (pdb code 3KKZ) from *B.fragilis NCTC 9343* had been determined by Northeast Structural Genomics consortium, but not yet reported in the literature. Towards understanding the structure and function of these two homologous (sequence identity 59%) methyltransferases, in this chapter we report the cloning, protein expression and biophysical characterization of BT_2972 and BVU_3255. Besides, both the proteins were crystallized in the *apo* form and as complex with ligands such as SAM and SAH. The sequence analysis suggests that both BT_2972 and BVU_3255 are small molecule methyltransferase and are involved in the methylation of intermediates of ubiquinone biosynthesis pathway.

2.2 Materials and Methods

2.2.1 Cloning

Gene encoding the BT_2972 (gene ID 1075985) and BVU_3255 (gene ID 5304216) were chemically synthesized (GeneScript, USA) and inserted into pUC57 cloning vector. Subsequently, to clone them into the expression vector, these genes were further PCR amplified from the respective pUC57 vector using the following primer-

BT_2972:
Forward:
5'CTTTCATATGCATCATCATCATCATCATAGTAACAATAATACAT'3
Reverse:
5' CTTTCTCGAGTCATCTTTTTTGTCCTATATAGAATACGTA'3

BVU_3255:
Forward: 5' CTTTCATATGCATCATCATCATCATCATAATAATGAC '3
Reverse: 5' CTTTCTCGAGTCACCTTCTCAGTGAGAATCC'3

This PCR amplification introduced Nde1 and Xho1 restriction sites at 5' and 3' end of each gene respectively. This also resulted in the introduction of Met start codon followed by 6 His codon at 5' end of each gene. Both the PCR products were gel purified and then digested with Nde1 and Xho1 (NEB, England) along with vector pGS21a (GeneScript, USA). The digested gene products were ligated with pGS21a vector and transformed into *E.coli* DH5α cells. Positive colonies were selected based on colony PCR and sequence of insert was verified by DNA sequencing.

2.2.2 Expression and purification

For large scale protein expression, recombinant plasmid pGS21a-BT_2972 and pGS21a-BVU_3255 were transformed into chemically competent *E. coli BL21 (DE3)* cells. A single colony was chosen to inoculate 100 ml of the LB media containing 100µg/ml ampicillin and this was incubated overnight at 37 ^{0}C with continuous shaking. 50 ml of this overnight grown culture was used to inoculate 1 L of the LB

media which was allowed to grow at 37 ^{0}C till the OD_{600} attained 0.6-0.8 AU. The culture was then induced with 0.15 mM IPTG and further allowed to grow for 16 hrs at 16 ^{0}C. The cells were harvested by centrifugation at 9800 g for 10 min.

Cell pellets were re-suspended in 100 ml of lysis buffer composed of 50 mM Tris–HCl (pH 7.4), 500 mM NaCl, 10% v/v glycerol, 20 mM imidazole, and 20 mm BME. Cell suspension was sonicated for 5 min at 28% amplitude and an ON/OFF 1 second pulse in Vibra-Cell VCX750 ultrasonic processor using 13 mm probe (Sonics). The cell lysate was centrifuged at 39000 g for 30 min at 4 ^{0}C to remove the cell debris and the insoluble proteins. The clear supernatant was mixed with 5 ml of Ni–NTA (Qiagen) resin pre-equilibrated in lysis buffer and kept for binding at 4 ^{0}C for 2 hrs. After binding, the supernatant was removed and the Ni–NTA resin was washed thrice, using 10 ml of lysis buffer each time, with a 15 min. interval. After washing, the bound proteins were released from Ni-NTA beads using lysis buffer supplemented with 300 mM imidazole. As a next step of purification, the BT_2972 and BVU_3255 fractions collected from Ni-NTA beads were passed through the gel filtration column- HiLoad™ 16/60 Super- dex™ 200 prep grade (Amersham Biosciences, Sweden). The column was pre-equilibrated with a buffer consisting of 20 mM Tris–HCl (pH 7.4) and 200 mM NaCl. Gel filtration chromatography was performed at 4 ^{0}C using the AKTA FPLC. Protein quantification was done by Bradford method (Zor & Selinger, 1996) using reagent from BioRad (USA).

2.2.3 Dynamic light scattering

Fractions of individual proteins from the gel filtration chromatography were pooled together and concentrated to 10 mg/ml. Homogeneity of each of the concentrated proteins was checked by Dynamic Light Scattering (DLS) using a DynaPro

instrument (Protein Solutions, USA). All samples were centrifuged at 10000 g for 20 min before the experiment. The measurements were carried out at 25 ^{0}C. The software provided by the manufacturer was used to calculate the hydrodynamic properties of BT_2972 and BVU_3255.

2.2.4 Crystallization

BT_2972 and BVU_3255 at a concentration of 10 mg/ml were mixed with SAM/SAH (ratio 1:5 protein: ligand) which is well above the micro molar binding affinity indicated by preliminary calorimetry experiments (results not shown) for making their respective complexes. Proteins and SAM/SAH were kept in a buffer consisting of 20 mM Tris–HCl (pH 7.4) and 200 mM NaCl. Crystallization trials were performed for *apo*, SAM and SAH complex of each of the proteins. Crystallization experiments were performed manually by hanging drop vapour diffusion method at room temperature. Typically 1µl of protein and 1µl of reservoir solutions were mixed together and equilibrated against 500 µl of reservoir solutions. Initial crystallization conditions were identified from the Hampton Research screens (PEG/ION and PEGRx2). Further, these conditions were optimized to obtain diffraction quality crystals.

2.2.5 Data collection

Crystals of both the proteins and their SAM/SAH complexes were soaked briefly (30 sec) in a cryo-protectant solution consisting of mother liquor supplemented with 10 % (v/v) glycerol, picked up in a nylon loop, and flash cooled at 100K in a N_2 cold stream (McFerrin & Snell, 2002). The complete diffraction data sets for the *apo* protein crystals were collected at the Bruker AXS X8 Proteum X-ray system (Bruker

AXS Inc., Madison, USA) while that for the SAM/SAH: protein complex crystals were collected in the beamline X-12B1 (Quantum 4-CCD detector -Area Detector Systems Corp Poway, CA, USA) at the National Synchrotron Radiation Research Centre (NSRRC), Taiwan. A complete data set (360°) was collected for some crystals in the in house X-ray machine (wavelength – 1.5418 Å). Some data was collected in synchrotron ((wavelength – 1.000 Å). In order to utilize the beam time, minimum no of frames were collected. No of frames were decided based completeness (greater than 90 %), redundancy (more than 3) in the higher resolution shell. The data collection strategy such as number of images and degrees per image are provided in Table 2.1. The collected data sets were processed using HKL2000 (Otwinowski & Minor, 1997).

2.3. Results

The optimized expression of BT_2972 and BVU_3255 from the pGS21a vector gave high yields and resulted in approximately 200 mg of each protein in the supernatant from 1L cell culture (Figure 2.1A and 2.1B). BT_2972 and BVU_3255 proteins were purified by two chromatographic steps: - the Ni-NTA affinity chromatography followed by gel filtration. Both proteins eluted at a volume which corresponded to the monomer size of these proteins, as compared with molecular weight standards (Figure 2.2). The final yield of each of the purified proteins was approximately 120 mg/litre culture and showed 99% purity in SDS-PAGE.

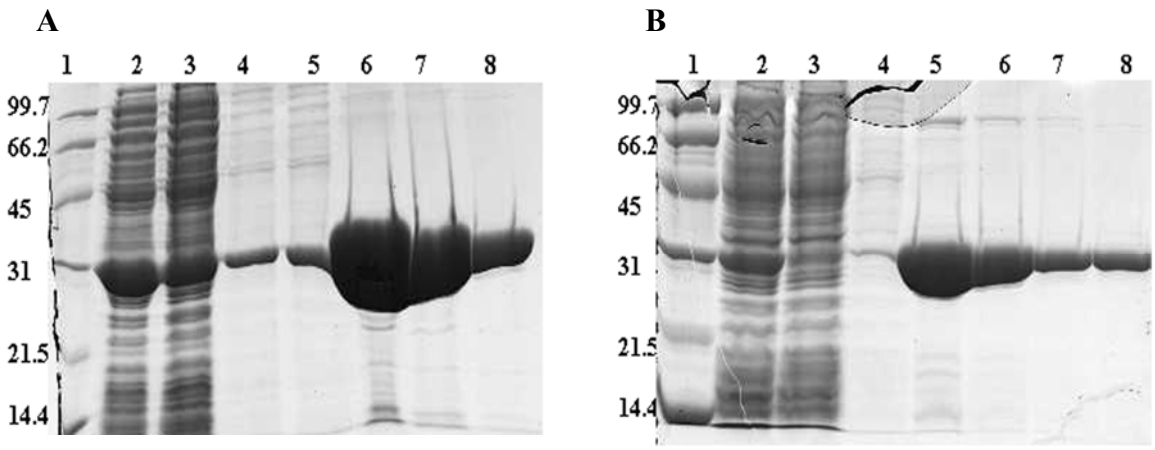


Figure 2.1: SDS-PAGE showing the expression and purification profile of BT_2972 (**A)** and BVU_3255 (B). Lane 1- marker, lane 2- cleared lysate, lane 3- flow through after affinity binding, lane 4 and 5- washes of Ni-NTA beads, lane 6 and 7- elution from Ni-NTA beads, and lane 8 - gel filtration fraction. Affinity purified protein was overloaded in order to check the traces of impurities (lane 6).

Table 2.1: Crystallization conditions and data collection statistics

	BT_2972			BVU_3255		
	Native	**AdoMet (SAM)**	**AdoHcy (SAH)**	**Native**	**AdoMet (SAM)**	**AdoHcy (SAH)**
Crystallization conditions	0.12 M Magnesium acetate and 16% polyethylene glycol (PEG) 3350	25% (v/v) 2-propanol; 0.1 M MES monohydrate (pH 6.0) and 18% (w/v) PEG monomethyl ether 2000 (PEG MME 2000)	25% (v/v) 2-propanol; 0.1 M MES monohydrate (pH 6.0) and 18% (w/v) PEG MME 2000	170 mM $Mg(CH_3COO)_2$ and 15% Polyethylene Glycol (PEG)3350	150 mM NH_4Cl, 17% w/v PEG3,350 and 3% v/v 2-Propanol	120 mM $(NH_4)(NO_3)$, 16% w/v PEG3350 and 200 mM non-detergent sulfobetaine 221 (NDSB-221)
Cell parameters (Å °)	a= 60.55 b= 74.38 c= 117.11	a = 120.79 b = 59.68 c = 77.62 β = 104.61	a=60.64 b=74.52 c=117.22	a= 132.06 b= 72.14 c= 55.90	a = 73.42 b = 106.65 c = 133.0	a=73.98, b=107.05, c=133.10
Space group	$P\ 2_1\ 2_1\ 2_1$	C 2	$P\ 2_1\ 2_1\ 2_1$	$P2_12_12$	$P2_12_12_1$	$P2_12_12_1$
No. of molecules in the A.U	2	2	2	2	4	4
Matthews coefficient (V_M, A^3/Da)	2.20	2.26	2.21	2.22	2.17	2.20
Solvent content (%)	44.1	45.5	44.3	44.6	43.4	44.0
Resolution range (Å) *	50.0 – 2.8 (2.9-2.8)	50.0 - 2.4 (2.49-2.4)	50.0 - 2.3 (2.38-2.3)	50.0 - 2.9 (3.0-2.9)	50.0 - 2.2 (2.28 2.2)	50.0 - 2.49(2.53-2.49)
Wavelength (Å)	1.5418	1.0000	1.0000	1.5418	1.000	1.5418
Oscillation angle (°)	1	1	1	1	1	1
No. of images	360	180	120	360	150	360
Observed reflections	108383	63495	82597	64399	258707	276109
Unique reflections	13019	20044	22841	11578	50893	36997
Redundancy	8.3(3.4)	3.2(2.4)	3.6(2.6)	5.6 (2.7)	5.1(4.4)	7.5 (5.3)
Completeness (%)	95.3 (74.7)	94.7 (73.1)	93.5(86.0)	92.8 (74.5)	94.7 (68.0)	95.6(91.4)
Overall (I/ σ (I)	4.1	9.9	10.2	23.3	15.0	15.4
R_{sym} [a]	0.110 (0.185)	0.059 (0.200)	0.074 (0.215)	0.105 (0.18)	0.055 (0.24)	0.091 (0.52)

[a] $R_{sym} = \sum|I_i - \langle I \rangle|/|I_i|$ where I_i is the intensity of the i^{th} measurement, and $\langle I \rangle$ is the mean intensity for that reflection.

* The high resolution bin details are in the parenthesis.

The purified proteins were concentrated upto 10 mg/ml. Size homogeneity of the proteins was monitored by dynamic light scattering experiments (Figure 2.3) during concentration and prior to crystallization. This showed the molecular weights to correspond to the monomers in solution and consistent with the gel filtration observations.

Additionally, the identity of the purified proteins, BT_2972 and BVU_3255, was verified by the peptide mass fingerprinting (PMF) experiments (data not shown). In both the proteins PMF detected peptide from C-terminal and Ni-NTA binding indicated the presence of His tag at N-terminus. This suggests that there was no proteolytic cleavage of N and C terminus.

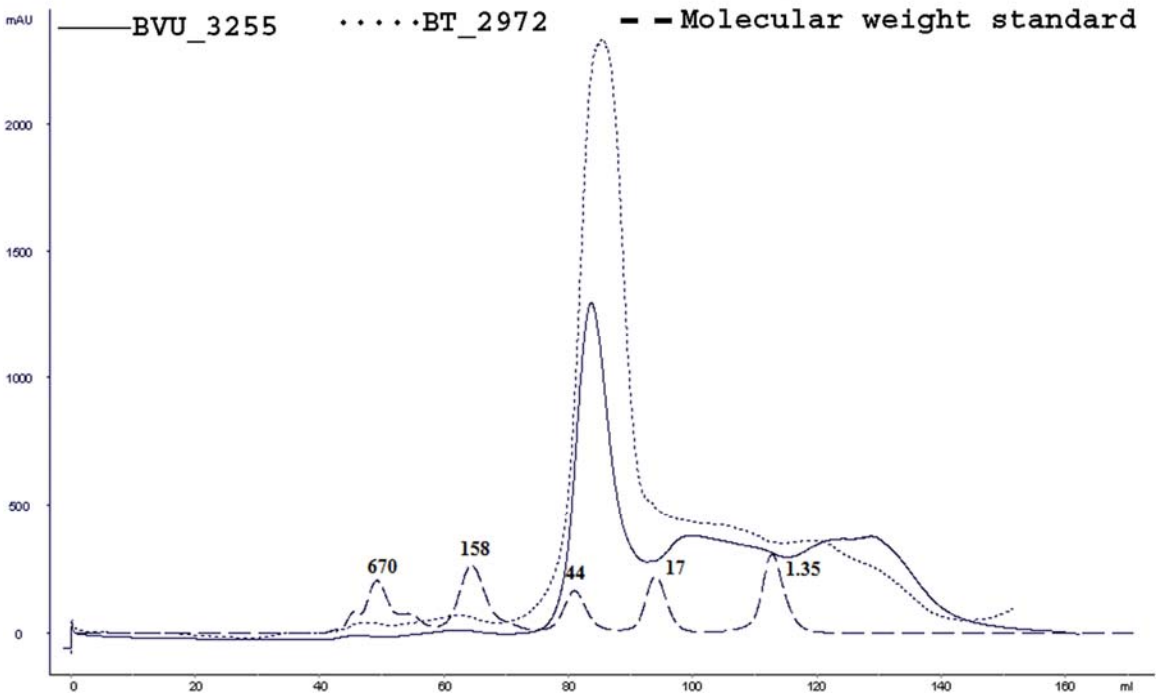


Figure 2.2: Gel filtration chromatography elution profile of BT_2972 (dotted line) and BVU_3255 (solid line). Elution profile of molecular weight standard is shown as dashed line and marked with different size of proteins (kDa). Gel filtration profile shows both proteins elute as monomer.

The noncleavable His tagged proteins were used for crystallization. The crystals appeared after 2 days and grown upto 7 days and have crystallized in different space groups. The collected data sets were processed to a maximum possible resolution with acceptable redundancy; completeness and R_{sym} in highest resolution bin (Figure 2.4). The Vm calculation indicated the presence of 2 molecules in the asymmetric unit of *apo* BT_2972, its SAM/SAH complexes and in *apo* BVU_3255, whereas 4 molecules in the asymmetric unit of BVU_3255 complexed with SAM/SAH (Table 2.1). Further this is confirmed by the self-rotation function calculations. Notably, while we were in the data collection stage, the *apo* structures of both proteins BT_2972 (3F4K) and BVU_3255 (3E7P) were available in the pdb data base and these structures had one molecule in asymmetric unit.

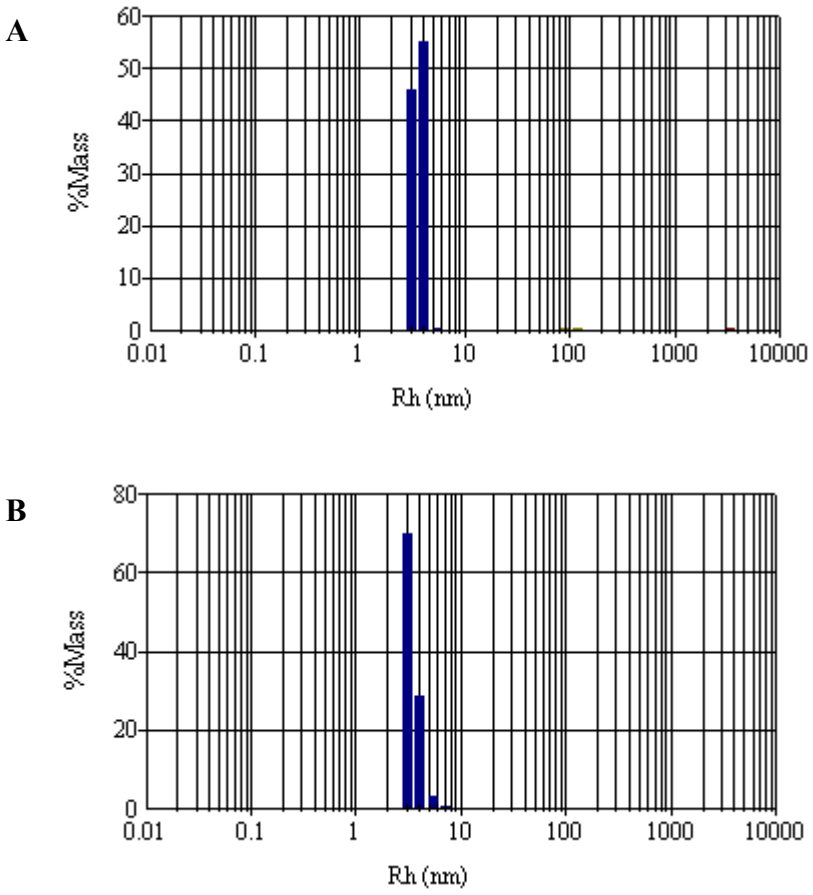


Figure 2.3: Dynamic light scattering histogram of BT_2972 (A) and BVU_3255 (B). Both proteins exist as monomer in solution.

The BT_2972 structures (*apo*, SAM and SAH complexes) and 3F4K all crystallized in three different space groups. In the case of BVU_3255 all of our structures and 3E7P were crystallized in different orthorhombic space groups. Our structures were crystallized in $P2_12_12$ (Andreasen *et al.*) and $P2_12_12_1$ (SAM/SAH), whereas 3E7P was in I222. Similarly a closely related SAM-dependent methyltransferase from *B.fragilis* was crystallized (in complex with SAM) in space group C2. The structure of both proteins and their complexes was determined by molecular replacement method using the program Molrep-auto MR in CCP4 suite (Vagin & Teplyakov, 2010). The structure solution clearly indicated the expected number of molecules in the asymmetric unit of BT_2972 and BVU_3255 with the initial R-factors in the range of 0.39 to 0.42. The models were refined using CNS (Brunger *et al.*, 1998). The non crystallographic symmetry (NCS) was used during the refinement. When the R-factors were close to 30s the difference maps were calculated and clearly showed the presence of the ligands.

The *PISA* server (Krissinel & Henrick, 2007) showed that the symmetry related molecule of 3F4K and 3E7P can form the dimer with complexation score 1. All BT_2972 and *apo* BVU_3255 crystals showed two molecules in the asymmetric unit with a similar dimer interface region as predicted for 3F4K and 3E7P. Similarly SAM/SAH complexes of BVU_3255 with 4 molecules in the asymmetric unit have a similar dimer interface region. The observed common dimer interface suggests that the dimerization might be biologically relevant.

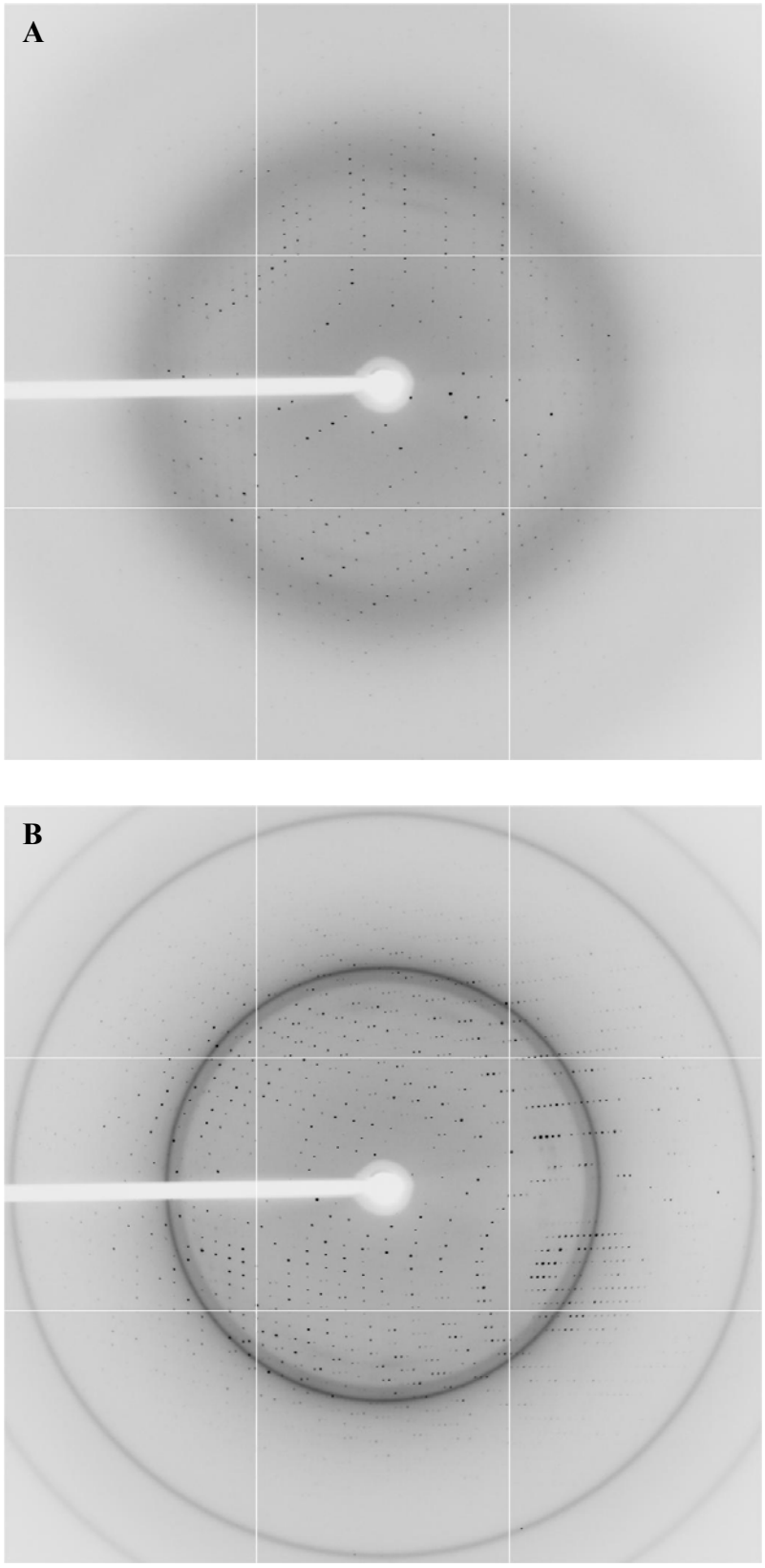


Figure 2.4: Representative diffraction pattern of both proteins crystals BT_2972-SAM (A) and BVU_3255-SAH (B).

In summary, this chapter reports the expression, purification, crystallization (of *apo* as well as SAM/SAH complexes), data collection and diffraction analysis of the methyltransferase BT_2972 and BVU_3255. These studies lead to the determination of the first representative methyltransferase structure in complex with SAM/SAH from the respective *Bacteroides* bacteria. Sequence analysis of these two proteins suggest that both are homologous with similar SAM/SAH binding regions and are small molecule methyltransferase involved in methylating the intermediates of ubiquinone biosynthesis pathway.

Chapter 3

A Conformational Switch in the Active Site of BT_2972, a MTase from an Antibiotic Resistant Pathogen *B. thetaiotaomicron* Revealed by its Structures

3.1 Introduction

Methyltransferases (EC 2.1.1) comprise a group of approximately 140 transferase enzymes that catalyze the transfer of a methyl group from a universal donor molecule, such as *S*-adenosyl-L-methionine (AdoMet), to a target substrate acceptor molecule, such as DNA, RNA, protein, lipid, or other small molecules. These enzymes play a crucial role in numerous biochemical processes, including signal transduction, biosynthesis, metabolism, protein modification, gene silencing, and chromatin regulation (Martin & McMillan, 2002, Miller *et al.*, 2003). Methyltransferases are structurally categorized into 5 classes (I-V). Whilst class I methyltransferases differ in their overall structure and substrate binding domain, all of the AdoMet-dependent methyltransferases in class I share a common Rossmann-fold at their catalytic site (Bujnicki, 1999).

Bacteroides thetaiotaomicron is a gram-negative anaerobic bacterial pathogen with extreme disease-causing potential and antibiotic resistance. Predominantly found in the human intestinal tract (Xu *et al.*, 2003), these bacteria dominate over other bacterial species and are involved in the uptake and degradation of otherwise non-digestible polysaccharides (e.g. amylose, amylopectin and pullulan), as well as in capsular polysaccharide biosynthesis, environmental sensing, signal transduction and DNA mobilization (Bjursell *et al.*, 2006, Xu *et al.*, 2003). The complete genome of the strain *B. thetaiotaomicron VPI-5482* has been sequenced (Xu *et al.*, 2003) and an open reading frame (ORF) encodes a protein BT_2972 (accession no NP_811884.1) that is predicted to have a conserved AdoMet binding domain, which is a characteristic of most methyltransferases (Loenen, 2006).

As a continuation of our studies toward understanding the structure and function of methyltransferases, in this chapter we report the crystal structures of

BT_2972, and the thermodynamics of AdoMet/AdoHcy binding with BT_2972. This study revealed significant conformational changes to a loop in the region of the active site (Glu121–Ile127), resulting in open and closed forms of the active site. In addition, our analysis combined with literature suggests that BT_2972 is a small molecule methyltransferase, and may be involved in catalyzing the O-methylation reaction in the ubiquinone biosynthesis pathway.

3.2 Material and Methods

3.2.1 Cloning and protein purification

The BT_2972 gene was cloned into expression vector pGS21a (GeneScript, USA) and the recombinant plasmid was transformed into *E. coli BL21 (DE3)* competent cells. The protein was purified to homogeneity using a two-step procedure involving Ni^{2+}-NTA affinity (Hengen, 1995) and gel filtration chromatography in a buffer consisting of Tris-HCl (pH 8.0) and 200 mM NaCl. Prior to crystallization, the homogeneity of BT_2972 was verified by dynamic light scattering (DLS) experiments. A detailed cloning, expression and purification procedure is described in chapter 1.

3.2.2 Crystallization and structure determination

Diffraction quality crystals of *apo* BT_2972 were obtained from a reservoir solution consisting of 0.12 M magnesium acetate and 16% (w/v) PEG3350, while the crystals of AdoMet and AdoHcy complexes were each grown from 25% (v/v) 2-propanol, 0.1 M MES monohydrate (pH 6.0) and 18% (w/v) polyethylene glycol monomethyl ether 2,000, respectively. Crystals were cryo-protected with 10% glycerol supplemented with reservoir condition and flash cooled in N_2 cold stream at

100 K (McFerrin & Snell, 2002). A detailed crystallization screening, crystal optimization and data collection strategy is given in chapter 2.

The molecular replacement method was used to solve the structure of BT_2972 using the program Molrep-auto MR in CCP4 suite (Vagin & Teplyakov, 2010) using 3F4K as the search model. The molecular replacement solution clearly indicated the expected number of molecules in the asymmetric unit of BT_2972, predicted based on the Matthew's constant. The initial R-factors of the unrefined models were in the range of 0.39–0.42 with a correlation coefficient of ~0.6. When required, protein models were manually built using the program COOT (Emsley & Cowtan, 2004) and refinement performed using the program CNS (Brunger *et al.*, 1998). Difference maps were calculated to position the ligands. At the final stage of the refinement, well-ordered water molecules were included. The models have good stereochemistry, with all residues within the allowed region of Ramachandran plot as analyzed by PROCHECK (Laskowski *et al.*, 1993). All structure-related figures reported were generated using PyMol (DeLano & Lam, 2005).

3.2.3 Isothermal titration calorimetry

The binding of AdoMet and AdoHcy to BT_2972 was studied using isothermal titration calorimetry (Pierce *et al.*, 1999). Protein and stock solutions of AdoMet and AdoHcy were kept in a buffer consisting of 20 mM Tris-HCl (pH 8.0) and 200 Mm NaCl. The ITC experiments were performed using a VP-ITC calorimeter (Microcal, LLC) at 24^{o}C with 0.4 ml of AdoMet or AdoHcy in the injector cell and 1.8 ml of protein in the sample cell. All samples were thoroughly degassed and centrifuged to remove precipitates. 10 µl injection volumes were used for all experiments. Two consecutive injections were separated by 5 min to reset the

baseline. The control experiment, consisting of titration of AdoMet/AdoHcy against buffer, was performed and subtracted from each experiment to adjust for the heat of dilution of ligands. ITC data were analyzed with a single site fitting model using Origin 7.0 software (OriginLab Corp.).

3.2.4 PDB deposition

The coordinates and structure factors of *apo*, AdoMet and AdoHcy with BT_2972 complexes were deposited in the RCBS database (Bernstein *et al.*, 1977) with the access code 3SVZ, 3SXJ, and 3T0I, respectively.

3.3 Results

3.3.1 BT_2972 Sequence Analysis

A PSI-BLAST (Altschul *et al.*, 1997) search on BT_2972 sequence indicated that it belongs to the AdoMet-dependent methyltransferase family, with 80% and 51% sequence identity to two methyltransferases involved in the ubiquinone/menaquinone biosynthesis pathways of *Bacteroides xylanisolvens XB1A* (score = 259) and *Gordonibacter pamelaeae 7-10-1-b* (score = 266), respectively. The PSI-Blast e-score indicates that these two sequences are evolutionarily related to BT_2972. Sequence alignment of the cluster of orthologous groups (COG) of similar ubiquinone/menaquinone methyltransferase (COG2226H) revealed several conserved residues (Tatusov *et al.*, 2000) (Figure 3.1). These analyses suggest that BT_2972 may be involved in the biosynthesis of ubiquinone/menaquinone; these pathways share a common intermediate and common methyltransferases.

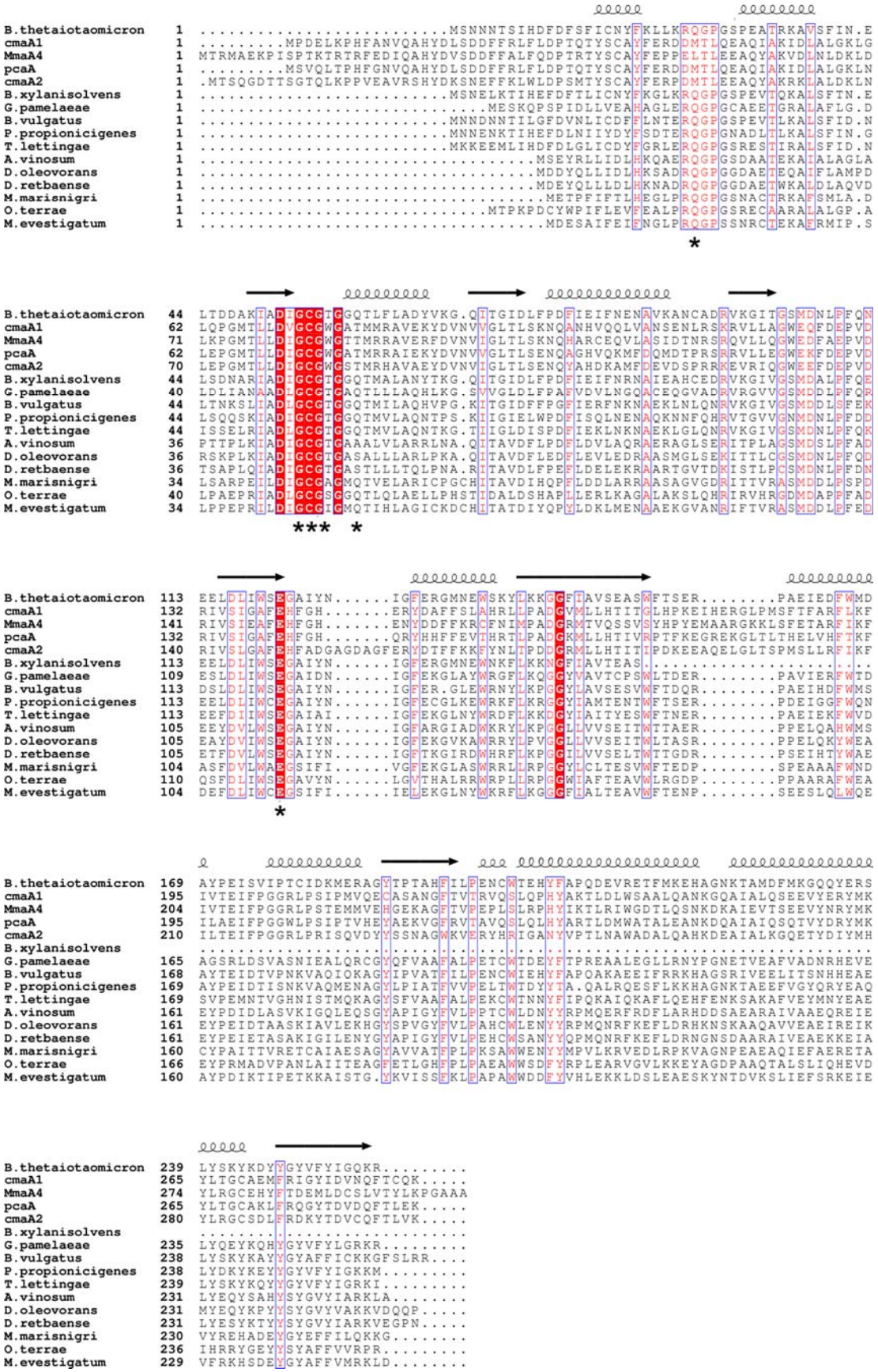

B.thetaiotaomicron
cmaA1
MmaA4
pcaA
cmaA2
B.xylanisolvens
G.pamelaeae
B.vulgatus
P.propionicigenes
T.lettingae
A.vinosum
D.oleovorans
D.retbaense
M.marisnigri
O.terrae
M.evestigatum

Figure 3.1: Multiple sequence alignment of BT_2972 with selected sequences of methyltransferase of ubiquinone/menaquinone biosynthesis pathway and mycolic acid modifying methyltransferases. The top five sequences (1-5) show the structure based sequence alignment of BT_2972 with mycolic acid methyltransferases (CmaA1, PDB code 1KPG; MmaA4, PDB code 2FK8; PcaA, PDB code 1L1E and CmaA2, PDB code 1KPI). Sequences 6-16 represent the sequence based alignment of BT_2972 with the ubiquinone/menaquinone methyltransferases (*B. xylanisolvens* - CBK67164.1, *G. pamelaeae*- CBL04788.1, *B. vulgatus* -YP_001300506.1, *P. propionicigenes*- YP_004041425.1, *T. lettingae* -YP_001471184.1 , *A. vinosum*- YP_003442314.1, *D. oleovorans* -YP_001528472.1 , *D. retbaense* - YP_003197230.1, *M. marisnigri*- YP_001046804.1, *O. terrae*- YP_001818708.1, *M. evestigatum*- YP_003726621.1). Sequence similarities are highlighted in red, whereas sequence identities are shown as white letters on a red background. The residues that interact with AdoMet/AdoHcy are marked with asterisks. The structure based alignment was obtained using DALI(Holm & Sander, 1995). Alignment was carried out using ClustalW(Chenna *et al.*, 2003). The secondary structure for BT_2972 is shown on the top. This diagram was generated using the program ESPript(Gouet *et al.*, 1999). Abbreviation- B: *Bacteroides*, G: *Gordonibacter*, P: *Paludibacter*,T: *Thermotoga*, A: *Allochromatium*, D: *Desulfococcus*, M: *Methanoculleus*, O: *Opitutus.*

3.3.2 Structure of BT_2972 and its AdoMet/AdoHcy complexes

Crystal structure of the *apo* BT_2972 and its complex with AdoMet and AdoHcy were solved and refined up to a resolution of 2.9, 2.5 and 2.4 Å, respectively (Table 3.1, Figure 3.2). The BT_2972 structures (*apo*, AdoMet and AdoHcy complexes) and the search model (PDB code 3F4K) all crystallized in three different space groups. BT_2972 structures crystallized with two molecules in the asymmetric unit (Table 3.1), while 3F4K crystallized with one molecule in the asymmetric unit. Notably, each structure crystallized in different conditions.

The pair wise comparison of these structures did not reveal any major conformational changes other than to the active site loop region, residues Glu121-Ile127. 3F4K is similar to BT_2972 *apo* form (RMSD 0.7Å for all Cα atoms). Despite the presence of two molecules of BT_2972 in the asymmetric unit, gel filtration and Dynamic Light Scattering (DLS) experiments showed that they exist as monomers in solution.

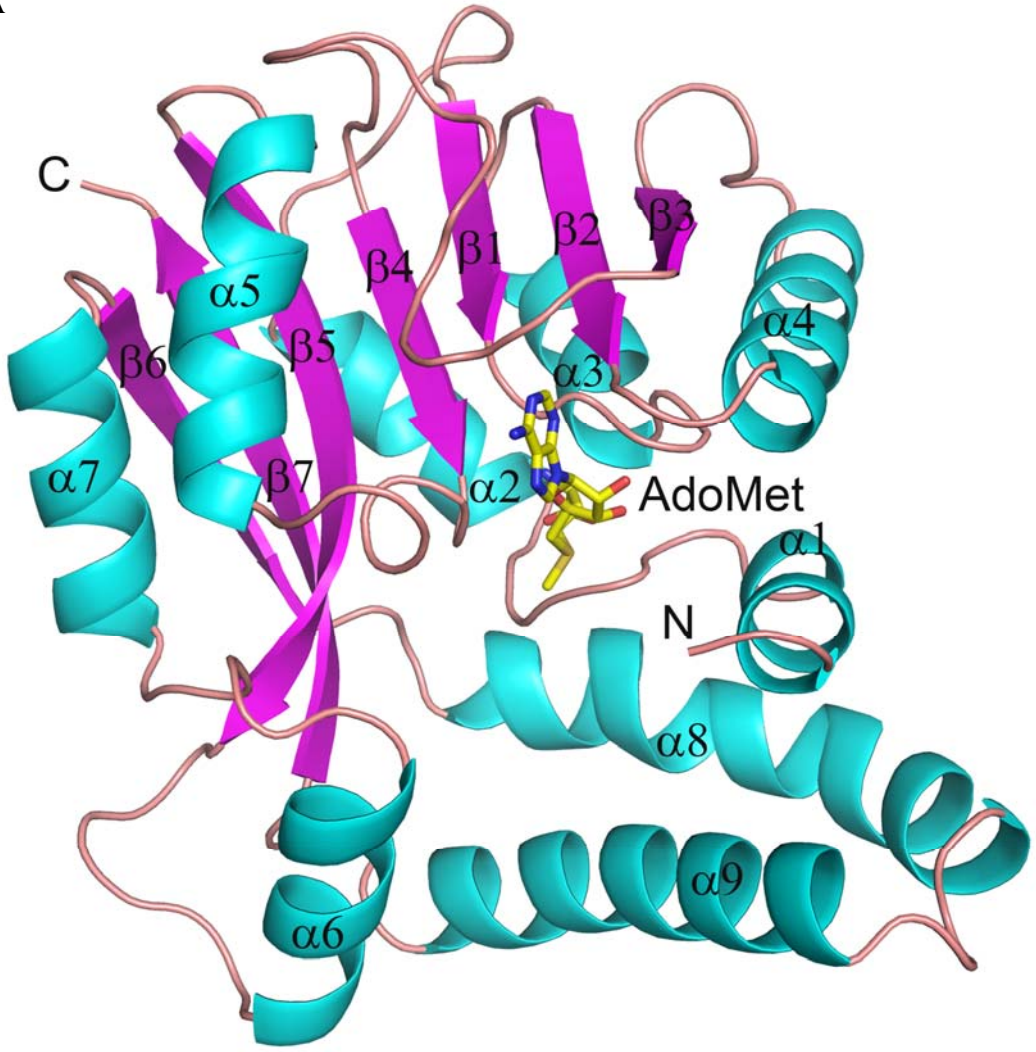
The BT_2972 construct has a total of 257 amino acids and a $(His)_6$-tag at the N-terminus. In all three models, the electron density map for the first 12 amino acids and His-tag was not well-defined and, as such, these residues were not included in the model.

Table 3.1: Crystallographic refinement statistics

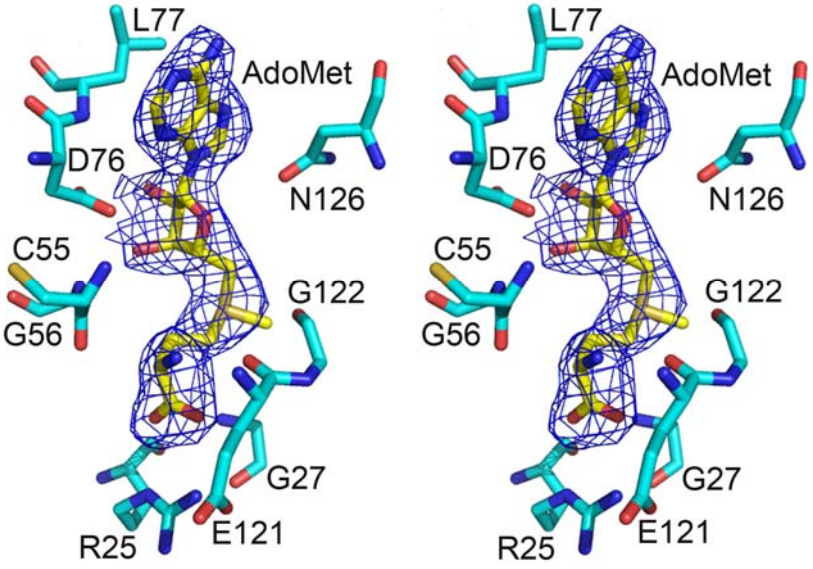
	Native	**AdoMet**	**AdoHcy**
Refinement and quality [a]			
Resolution range (Å)	50.0 – 2.9	50.0 - 2.5	50.0 - 2.4
R_{work} [b] (no. of reflections)	0.20(8952)	0.23 (13792)	0.23(16142)
R_{free} [c] (no. of reflections)	0.26 (674)	0.27 (1013)	0.26(1189)
RMSD bond lengths (Å)	0.008	0.008	0.007
RMSD bond angles(°)	1.17	1.12	1.00
Average B-factors[d] (Å^2)			
Main chain	22.1	47.1	32.0
Side chain	25.2	48.2	34.4
Ramachandran plot			
Most favored regions (%)	91.0	90.1	92.1
Additional allowed regions (%	8.5	9.6	7.9
Generously allowed regions (%)	0.5	0.4	0.0
Disallowed regions (%)	0.0	0.0	0.0

[a]Reflections with $I>\sigma$ was used in the refinement.
[b]$R_{work} = |F_{obs} - F_{calc}|/|F_{obs}|$ where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively.
[c]R_{free} = as for R_{work}, but for 5% of the total reflections chosen at random and omitted from refinement.
[d]Individual B-factor refinements were calculated.
* The high resolution bin details are in the parenthesis.

A


C
β4
β1
β2
β3
α5
α4
β6
β5
α3
α7
β7
α2
AdoMet
α1
N
α8
α9
α6

B


L77
AdoMet
D76
N126
C55
G122
G56
G27
E121
R25
L77
AdoMet
D76
N126
C55
G122
G56
G27
E121
R25

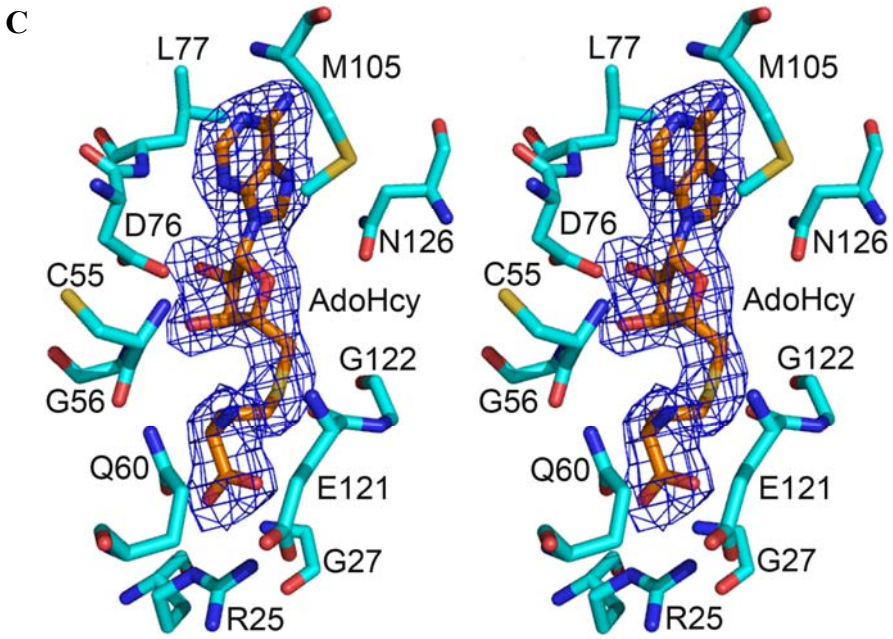


Figure 3.2: Structure of BT_2972. ***(A)*** Ribbon representation of the crystal structure of BT_2972-AdoMet complex. The α helices and β sheets are shown in cyan and magenta, respectively. The N- and C- termini and the secondary structural elements are labelled. The AdoMet is shown in stick representation (yellow). Stereo view of *2Fo-Fc* map of **(B)** AdoMet and **(C)** AdoHcy in BT_2972 complexes. Map is contoured at a level of 1.0σ. AdoMet is shown in yellow **(B)** and AdoHcy is shown in orange **(C)**. Selected interacting residues from BT_2972 are shown in cyan colour.

The BT_2972 molecule is a single domain, globular protein comprising both the cofactor and substrate binding site within the same domain (Figure 3.2A). It consists of a total of nine α-helices and seven β-strands of different lengths that are distributed throughout the protein sequence. The core domain adopts a typical class I Rossmann-like AdoMet-binding fold that is common to all class I AdoMet-dependent methyltransferases, comprising a seven-stranded β-sheet (β3↓β2↓β1↓β4↓β5↓β7↑β6↓) flanked by three α helices on both sides (Schubert *et al.*, 2003). In this fold, all β-strands in the active site are planner and parallel to each other, with the exception of the anti-parallel strand β7. The AdoMet/AdoHcy binding region is primarily located

within the N-terminus (Figure 3.2A), while the substrate is proposed to bind to C-terminal residues. Depending on the target for methylation, substrate binding sites of different methyltransferases vary in structure and size to accommodate the different targets.

3.3.3 Thermodynamics of AdoMet/AdoHcy binding

The interactions between BT_2972 with AdoMet and AdoHcy were studied using ITC (Figure 3.3). As predicted from the crystal structure, ITC experiments showed a single site binding model for both AdoMet and AdoHcy. The thermodynamic binding parameters (where K_a is the association constant, ΔH is the change in enthalpy, and N is the number of binding sites) were calculated from ITC data fitting, as follows: for AdoMet: $K_a=0.395\times10^5$ M^{-1} ($\pm$ 0.0435×10^5), $\Delta H=-3.775$ kcal/mol (±0.42), N=0.78.$\pm$0.07; and for AdoHcy: $K_a=3.498\times10^5$ M^{-1} ($\pm0.66\times10^5$), $\Delta H=-18.63$ kcal/mol (±5.6), N=0.82$\pm$0.04. Although both AdoMet and AdoHcy have similar interactions with BT_2972, the affinity between the ligands and BT_2972 is different.

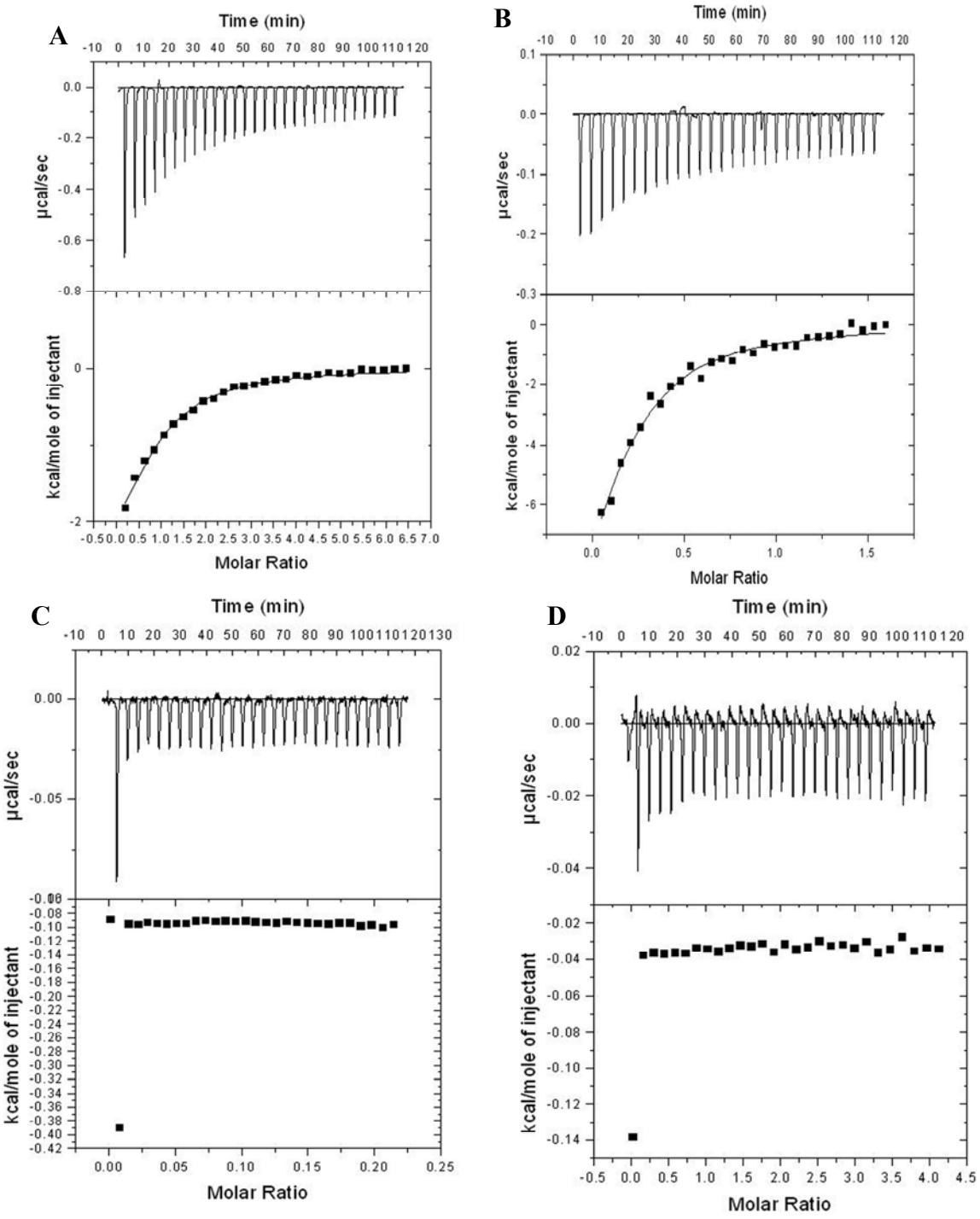


Figure 3.3: ITC profiles for BT_2972 titrated against the cofactor **(A)** AdoMet and **(B)** AdoHcy. The ITC control experiments: **C)** Titration profile for AdoMet against buffer. A similar figure was obtained for AdoHcy titration against buffer. **D)** Titration of buffer against BT_2972 protein solution. The raw ITC data for injections of ligands into the sample cell containing the native protein are shown in the upper panels of ITC profiles. The peaks were normalized to the ligand:protein molar ratio and were integrated as shown in the bottom panels. Solid dots indicate the experimental data, and their best fit was obtained from a nonlinear least squares method, using a one-site binding model (depicted by a continuous line).

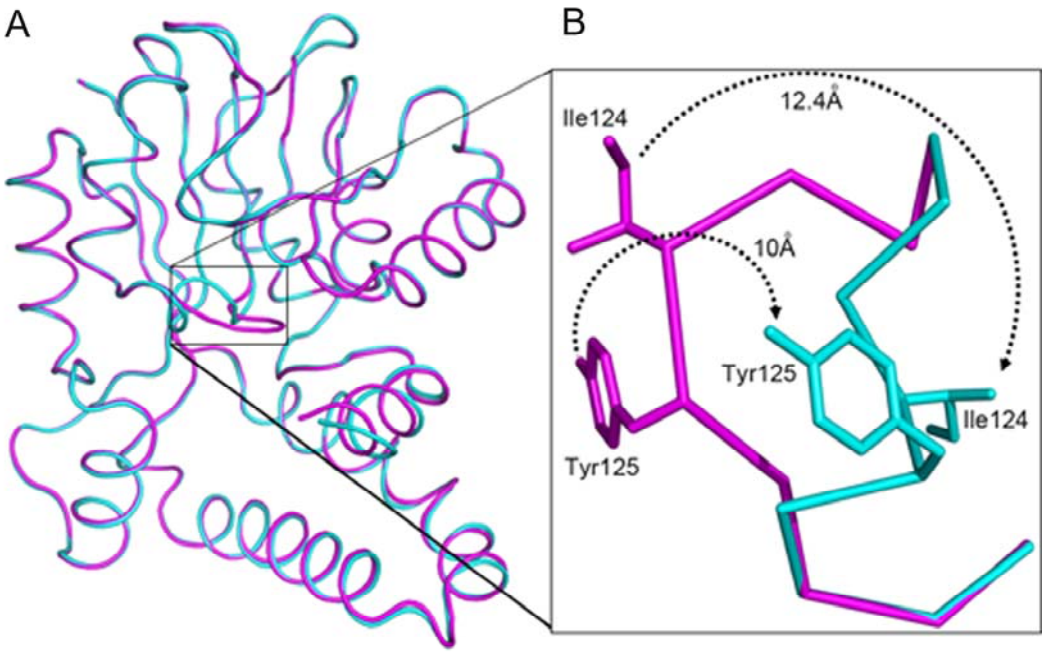
3.3.4 AdoMet/AdoHcy Binding Pocket of BT_2972

BT_2972 was crystallized with and without AdoMet and AdoHcy (Figure 3.2). In the active site, both ligands are buried and are oriented in a similar way, with their adenosyl base moiety facing outside and the carboxypropyl moiety facing inside the molecule. Residues located in the active site region, such as Arg25, Gln26, Gly27, Cys55, Gly56, Gln60 and Gly54; interact with the carboxypropyl moiety of AdoMet and AdoHcy in both complexes. Similarly, residues within the region of the active site, such as Leu77, Met105, Asn126, Gly103, Glu121 and Gly122, interact with the adenosyl moiety of the ligands (Figure 3.2B and 3.2C). These residues line up in the active site, and all interact with the ligands either via hydrophobic or hydrogen bonding contacts (Figure 3.2B and 3.2C). There are ten hydrogen bonding contacts (<3.3Å) between the ligand and BT_2972 in both complexes. The superposition of these two complex structures gave an rmsd of 0.4Å. The carboxyl group of Asp76 in β2 strand was found to interact with the oxygen atom of the AdoMet/AdoHcy ribose ring. This is a conserved position with Asp or Glu in a number of methyltransferases and interacts with ribose ring of the bound AdoMet/AdoHcy through hydrogen bonding contacts (Lim *et al.*, 2001).

3.3.5 Conformational switch acts as a gate to the active site

Superposition of the *apo* BT_2972 with AdoMet- and AdoHcy-bound structures gave a root mean square deviation (RMSD) of 1.0 and 0.8 Å, respectively, for 245 Cα atoms (Figure 3.4A). A large conformational change was observed between the residues Glu121 and Ile127 upon the binding of AdoMet/AdoHcy in the active site region. This region is located in the loop between β4 and α5, with Glu121, Gly122 and Asn126 making direct contact with AdoMet/AdoHcy. Compared with the *apo*

BT_2972, the backbone atoms of the AdoMet/AdoHcy complexes of this region are moved approximately 9Å (with a maximum side chain movement of Ile124 by 12.4Å and Tyr125 by 10Å), opening up the active site (Figure 3.4B). The *apo* BT_2972 thus represents the closed form of the active site. In the *apo* structure, the loop Glu121-Ile127 protrudes into the cofactor binding region, with the side chains of Ile124 and Tyr125 occluding the active site from engaging with cofactors (Figure 3.4C). However, on the other side of the active site cleft, there was no conformational change observed in the wall of the cleft, containing residues Leu77 and Phe81. This suggests that the loop region (Glu121-Ile127) acts as a flexible gate that enables the cofactors to enter or leave the active site region. In addition to opening the gate, this loop region may also enhance the hydrophobic interactions with substrate (Figure 3.5D) (see possible substrate binding site and substrate section below).



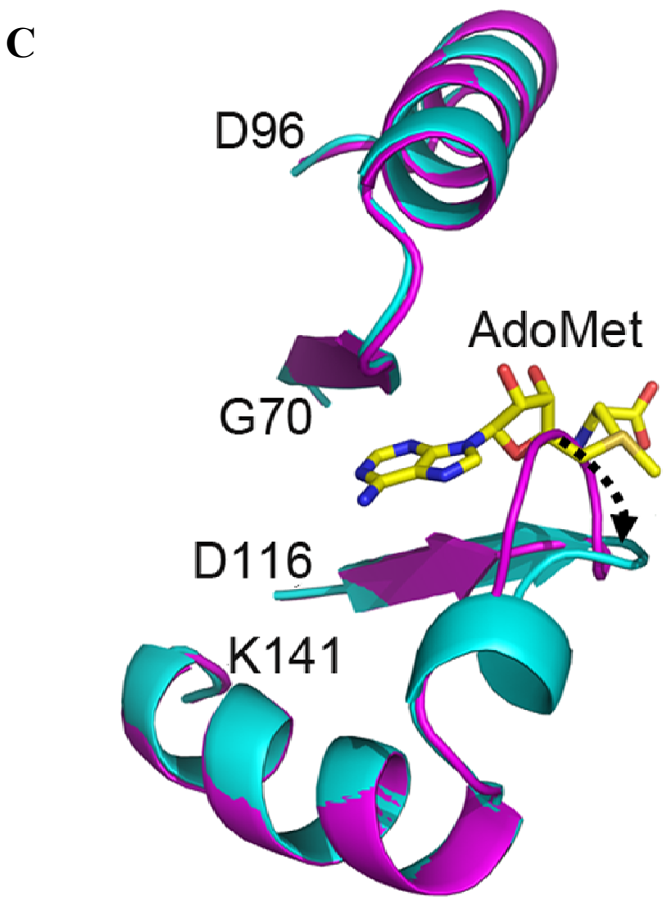
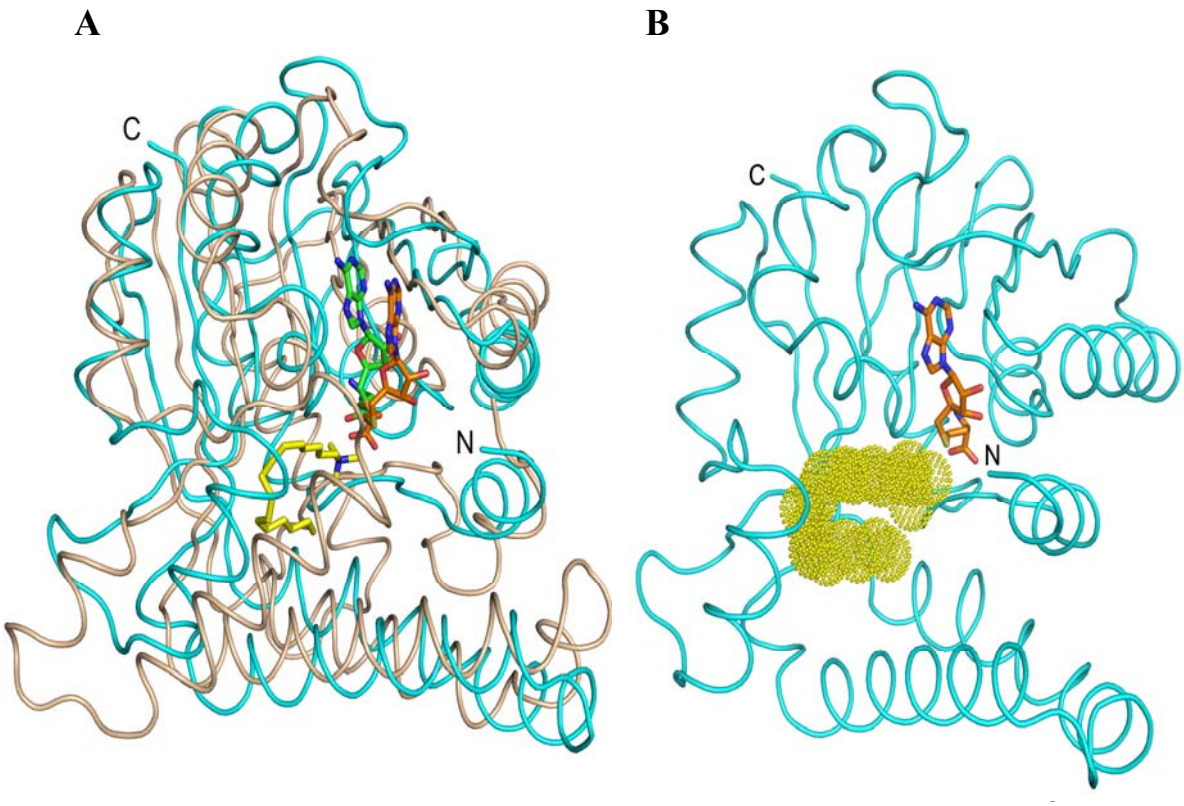


Figure 3.4: A) The superposition of *apo* BT_2972 (magenta) and BT_2972-AdoMet (cyan) complex. Conformational change in the fragment Glu121-Ile127 is marked by a rectangular box. **B)**. A close up view of the conformational change **C)** Conformational change in the fragment Glu121-Ile127 with bound ligand (AdoMet is shown here).

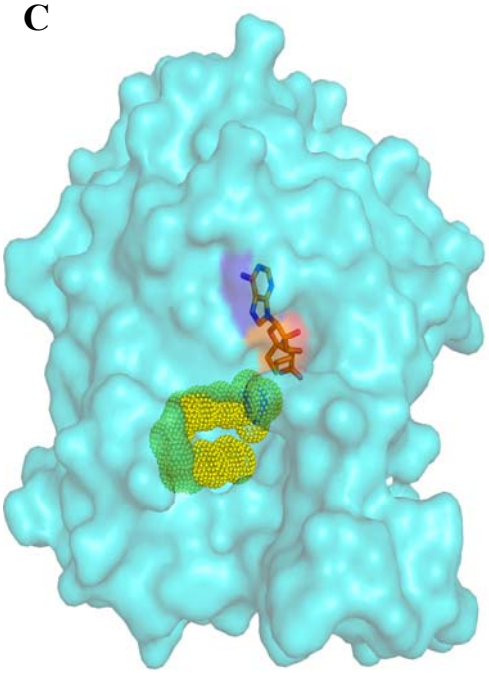
3.3.6 Structural Comparison with other Homologs

A search for structurally similar proteins in PDB database using DALI server (Holm & Sander, 1995) revealed that BT_2972 shared the highest homology with several class I Rossmann folded methyltransferases (RFM). The most structurally similar protein we identified (with known function) was Hma enzyme (PDB code 2FK8: Score 20.4 and rmsd 3.3 Å for 237 Cα atoms; sequence identity 12%), an AdoMet-dependent methyltransferase from *Mycobacteriam tuberculosis* involved in the production of branched chain mycolic acid (Boissier *et al.*, 2006). Next, was mycolic acid cyclopropane synthase CmaA1 from *M. Tuberculosis* (PDB code 1KPG,

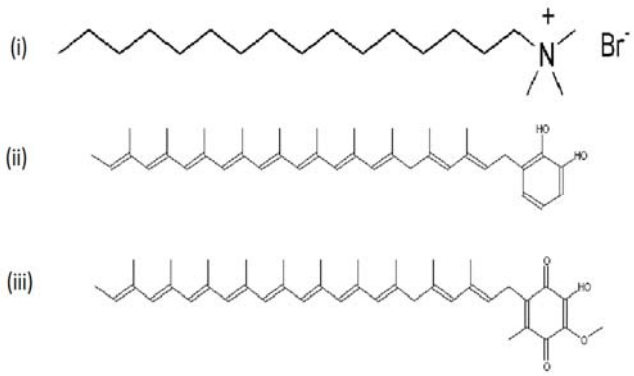
Score 19.9 and rmsd of 3.3 Å for 232 Cα atoms; sequence identity 12%) (Huang *et al.*, 2002). CmaA1 is also an AdoMet-dependent protein with a similar α/β Rossmann fold that is involved in site-specific methylation of the mycolic acid. Figure 3.5A shows the superposition of the structure of BT_2972- AdoHcy and CmaA1-CTAB-AdoHcy(Huang *et al.*, 2002). We observed close alignment of the active site regions of these two structures with the cofactors (Figure 3.5A). The carboxypropyl portions of AdoMet in both proteins are in the same plane, but the ribose and adenosyl moieties show some deviation. This similarity in the active site suggests a similar substrate. To date, only one structure of an uncharacterized ubiquinone/menaquinone methyltransferase is available in PDB: TM1389 from *Thermotoga maritima MSB8* (Code 2AVN, Score 14.4 and RMSD 3.3 Å for 179 Cα atoms; sequence identity 14%); this structure is not reported in the literature. Based on these analyses, we propose that BT_2972 belongs to the class of small molecule methyltransferases involved in the ubiquinone/menaquinone pathway.



C



D



E

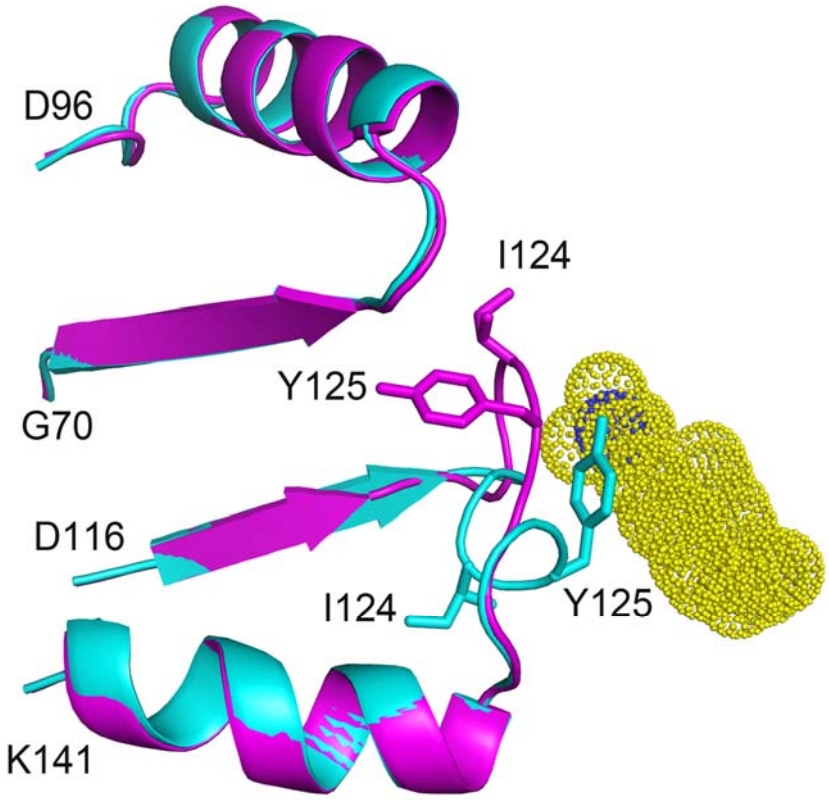


Figure 3.5: A) Cα trace for the superposition of BT_2972-AdoHcy (cyan) and mycolic acid cyclopropane synthase CmaA1-AdoHcy-CTAB (light orange) from *M. tuberculosis* (PDB code 1KPG). AdoHcy of the two proteins occupies the same region in the active site. The carboxypropyl moiety is lying in same plane but ribose and adenosyl moieties are separated by 3.5Å. Methylation substrate CTAB of the CmaA1 is shown as stick representation in yellow. The alignment was carried out in PyMol (DeLano & Lam, 2005). **B)** The proposed substrate binding region of BT_2972-AdoHcy is shown as yellow dotted surface. **C)** The inferred substrate binding site also is shown in surface diagram **D)** Similarity between CTAB and the proposed substrate for BT_2972:- (i) CTAB contain a 16-carbon-long alkyl chain with a positively charged quaternary ammonium group at one end. (ii) 2-polyprenyl-6-hydroxyphenol and (iii) 2-polyprenyl-3-methyl-5-hydroxy-6-methoxy-1,4-benzoquinone; both substrate have similar 10- carbon long alkyl chain. **E)** Conformation change in the fragment Glu121-Ile127 with respect to the substrate binding site. Side chains of Ile124 and Tyr125 in AdoMet/AdoHcy complex structure project towards the substrate site and will enhance the hydrophobic interactions with substrate.

3.3.7 Possible Substrate Binding Site and Substrate

The superposition of CmaA1 substrate (CTAB) complex with our BT_2972 structure allowed us to propose the substrate binding site for BT_2972. The substrate CTAB of CmaA1 occupies a space that presents as a cavity in the BT_2972 complex adjacent to the AdoMet/AdoHcy binding site, and it is reasonable to suggest that this proposed substrate binding site in BT_2972 is where methylation occurs (Figure 3.5B). The site primarily consists of C-terminal residues and is surrounded by helices α1, α6 and α9 (yellow dotted region in Figure 3.5B and 3.5C) of the active site region. Most of the amino acids are bulky and hydrophobic in nature and therefore suitable to interact with a substrate of a hydrophobic nature. The substrate of CmaA1, i.e. CTAB, and the intermediates which undergo methylation in the ubiquinone pathway are similar, with long aliphatic chains (Figure 3.5D). This suggests that substrates for BT_2972 are 2-polyprenyl-6-hydroxyphenol and 2-polyprenyl-3-methyl-5-hydroxy-6-methoxy-1, 4-benzoquinone; intermediates of the ubiquinone pathway. In addition to the aliphatic chains, BT_2972 substrate has a benzene ring (Figure 3.5D). The hydrophobic environment and shape of the proposed substrate binding site of BT_2972 could

potentially form a hydrophobic cluster with the incoming substrate. Moreover, upon an open conformation, the loop (Glu121-Ile127) that is located between the AdoMet/AdoHcy binding site and the proposed substrate binding site will enhance the hydrophobic interactions with the bound substrate by bringing its own hydrophobic residues (Ile124 and Tyr125) closer to the substrate (Figure 3.5E).

The sequence alignment shows that the AdoMet/AdoHcy binding residues are strictly conserved in all of the compared sequences, and suggests that the AdoMet/AdoHcy binding site will be the same in all the remaining 11 proteins whose structures are not known (5 structure-based and 11 sequence-based alignments, outlined in Figure 3.1). Notably, the confirmed substrate binding site of CmaA1 and the proposed substrate binding site for BT_2972 are conserved. Moreover, this proposed substrate binding site in BT_2972 is close to the transferable methyl group of AdoMet, indicating that it fulfils the geometric requirements for methylation SN2 reactions: a linear and planar arrangement of the acceptor substrate atom, the transferred methyl group and the sulphur atom of AdoMet to form the appropriate transition state. However, the proposed substrate warrants further experimental verification.

3.4 Discussion

Class I AdoMet-dependent methyltransferases consist of a well-conserved AdoMet-binding region that is responsible for cofactor binding and methylation and a highly variable substrate-binding region (Kozbial & Mushegian, 2005). Macromolecule and small molecule methyltransferases have several distinguishing features in addition to their core Rossmann fold. Macromolecular methyltransferases have additional secondary structural elements at their C terminus (Martin & McMillan, 2002) and in most cases, their substrate binding region is a separate domain to engage with the

substrate such as DNA, RNA or proteins. In contrast, small molecule methyltransferases lack these two major structural features, with the majority having two additional α helices at the N-terminus and others showing insertion sites between β5 and α7; and β6 and β7(Martin & McMillan, 2002).

Here, we report the crystal structure of the AdoMet-dependent methyltransferase, BT_2972, from *B. thetaiotaomicron VPI-5482* strain which shows various structural features: 1) there is no separate substrate binding domain; 2) there are no additional secondary structure at the C-terminus; 3) there are two small α helices (α1 and α2) at the N-terminus; and 4) there are helices between β5 and α7 (α6) and between β6 and β7 (α8 and α9), respectively (Figure 3.2A). Furthermore, *E. coli Bl21 (DE3)* harbouring the BT_2972 gene was tested for different antibiotic resistance (kanamycin, tetracyclins, erythromycin and chloroamphenicol). Each of these antibiotics retained their activity, suggesting that BT_2972 might not have a macromolecule substrate for methylation that confers antibiotic resistance. Based on these structural features, sequence analysis and resistance tests; we suggest that BT_2972 is a small molecule methyltransferase of the ubiquinone/menaquinone biosynthesis pathway.

The ubiquinone pathway involves three methylation steps while menaquinone pathway has one, with chorismate as the common intermediate for both pathways (Meganathan, 2001). The three steps of ubiquinone pathway are catalyzed by two methyltransferases encoded by the *UbiG* and *UbiE* genes. The *UbiE*-encoded methyltransferase is common amongst both pathways and is highly conserved amongst the bacteria. In the case of *B. thetaiotaomicron VPI-5482*, this methyltransferase is annotated as BT_4216 (Xu *et al.*, 2003). In contrast, the sequences for *UbiG*-encoded methyltransferases (ubiquinone biosynthesis SAM-

dependent O-methyltransferase) are highly variable amongst the bacteria. This leads us to speculate that BT_2972 is a methyltransferase of the ubiquinone biosynthesis pathway and may represent the UbiG-encoded methyltransferase in *B. thetaiotaomicron VPI-5482* that catalyzes the two O-methyltransformation reactions in the ubiquinone biosynthetic pathway in bacteria.

A key observation from our crystal structure in BT_2972 is the conformational change in the active site region upon cofactor binding which shows the opening and closing of the gate access to the active site. Conformational changes near the AdoMet-binding site have been observed in other methyltransferases, such as in rat catechol-O-methyltransferase (Tsuji *et al.*, 2009), L-isoaspartyl (D-aspartyl) methyltransferases (Griffith *et al.*, 2001), Betaine homocysteine S-methyltransferase (Gonzalez *et al.*, 2004). In the case of rat catechol-O-methyltransferasee (PDB code 1VID and 2ZLB), the conformational change occurs in the backbone atoms of the loop Lys36-Val42 (~12.0Å) upon AdoMet binding close to the carboxylpropyl moiety of AdoMet. However, this loop does not occupy the AdoMet binding site, and only upon cofactor binding the backbone of this loop move towards the caroboxyl moiety of AdoMet. Further, the side chains that occupied the AdoMet binding site in the *apo* structure (His142 and Trp143) moved away (Figure 3.6A) to accommodate the incoming ligand (Tsuji *et al.*, 2009). In the case of L-isoaspartyl (D-aspartyl) methyltransferases (PDB code 1JG1 and 1JG4) (Griffith *et al.*, 2001), only the side chains of Tyr192 and His193 flip outward between the AdoMet and AdoHcy complexes. In AdoMet-bound structure, these side chains project toward the cofactor; but in AdoHcy complex, these side chains flip outward (Figure 3.6B). For the betaine homocysteine S-methyltransferase (Gonzalez *et al.*, 2004), the backbone and the side

chain atoms of Phe76 and Tyr77 move upon substrate binding in comparison with its *apo* structure without interacting with the substrate (Figure 3.6C).

Unlike with these previous examples, the conformational change in BT_2792 is important for the ligand binding. The observed conformational change in the loop (Glu121-Ile127) in the active site is near to the ribose and adenosyl base binding region of the ligands, and is accomplished by the movement of the backbone atoms up to 9Å along with the side chain movement up to 12Å. Thus, to our knowledge, the conformational change observed in BT_2972 is unique.

In summary, we report the crystal structures of an AdoMet-dependent methyltransferase BT_2972 from an antibiotic resistant bacterium *B. Thetaiotaomicron VPI-5482*. BT_2972 is the first representative structure of an AdoMet-dependent methyltransferase of the ubiquinone biosynthesis pathway from *Bacteroides genera*. The comparison between the structures of the *apo* and AdoMet/AdoHcy complexes revealed the closed and open forms of the active site that may regulate cofactor movement and substrate interactions. The isothermal titration calorimetric studies showed a different binding affinity for BT_2972 with AdoHcy and AdoMet. The structural and sequence analyses, combined with literature, suggest that BT_2972 is a small molecule methyltransferase that may catalyze two O-methylation reaction steps in the ubiquinone biosynthesis pathway.

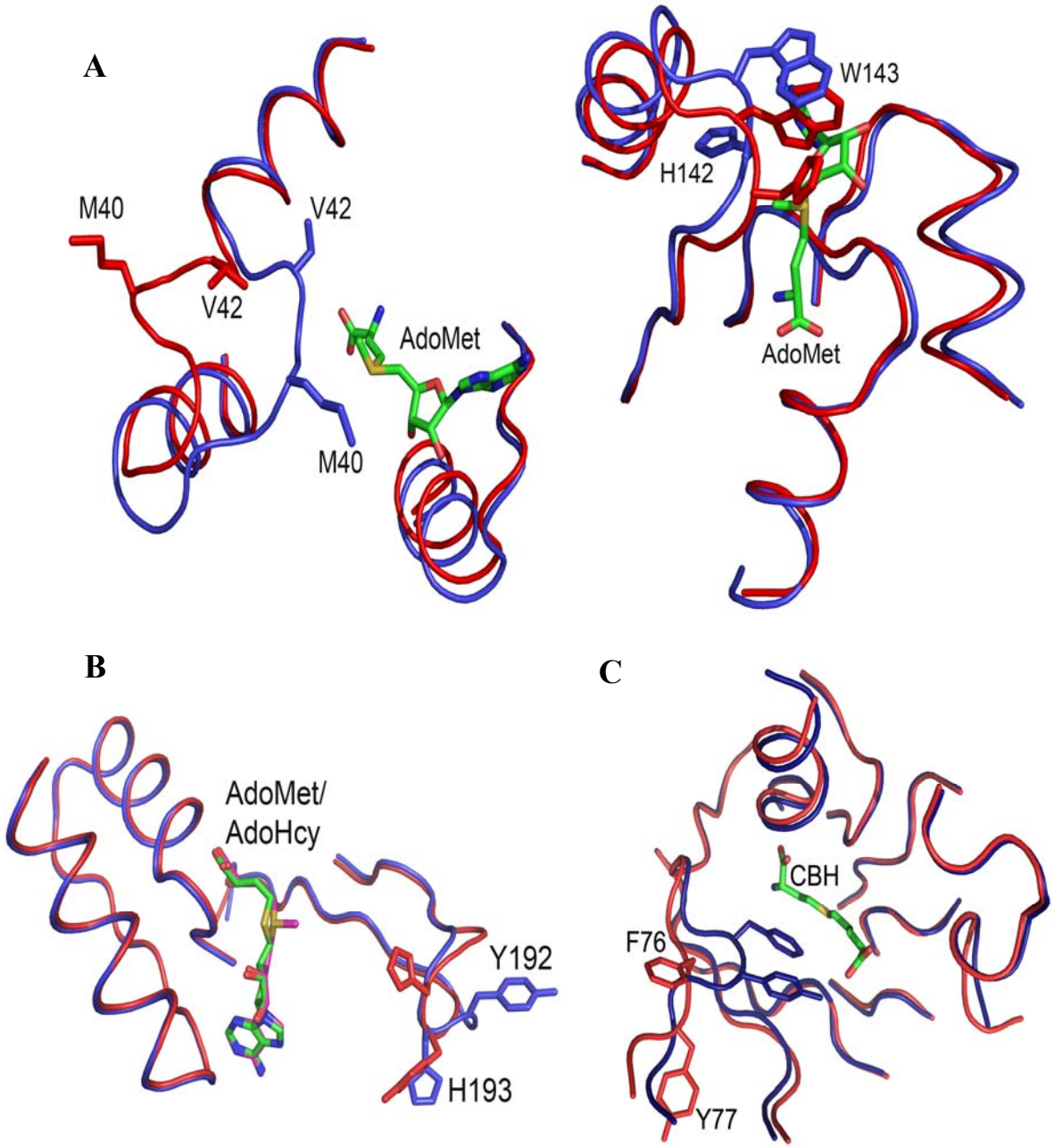


Figure 3.6: A) Comparison of *apo* (red) and AdoMet bound (blue) rat catechol-O-methyltransferase (PDB code- 2ZLB and 1VIB, respectively). A large conformational change takes place in the loop Lys36-Val42 upon AdoMet binding that moves this loop towards the AdoMet (left panel). Also in the same structure a side chain movement was observed for the residue His 142 and Trp 143 (right panel). AdoMet is shown as stick representation in green **B)** In L-isoaspartyl (D-aspartyl) methyltransferases (PDB code 1JG1 and 1JG4), the side chain was flipped out in the residues Tyr192 and His193 between AdoMet (red) and AdoHcy (blue) complexes. **C)** Figure shows the conformational change in betaine homocysteine S-methyltransferase (PDB code: 1UMY and 1LT8) upon substrate binding. The backbone and the side chain atoms of Phe76 and Tyr77 were moved in the substrate (CBH) complex in comparison with its *apo* structure. *Apo* protein is shown in red and CBH complex is in blue.

Chapter 4

Structural Characterization of BVU_3255, a Methyltransferase from Human Intestine Antibiotic Resistant Pathogen *Bacteroides vulgatus*

4.1 Introduction

The adult human intestine contains trillions of bacteria dominated by the genus *Bacteroides(McCarthy et al., 1988)*, a group of gram-negative, rod-shaped, anaerobic bacteria that do not form spores and are non-motile (Bakir *et al.*, 2006). Typically, *Bacteroides* are symbionts, but they can also be opportunistic pathogens. In the human intestine, specific *Bacteroides* are responsible for digesting different polysaccharides, such as amylose, amylopectin and pullulan, which support rapid and high growth yields; however, *Bacteroides* are unable to degrade plant cell wall polysaccharides. One of the most common species in the human intestine, *Bacteroides vulgatus*, accounts for approximately 12% of this bacterial population (Bakir *et al.*, 2006).

Previously, the complete genome sequence of the *B. vulgatus ATCC 8482* was reported (Xu *et al.*, 2007). BVU_3255 is a 29.8 kDa protein (accession no YP_001300506.1) of this proteome and is a putative methyltransferase, with an S-adenosyl-L-methionine (SAM)-binding domain. Methyltransferases (EC 2.1.1) catalyze the transfer of a methyl group from a donor molecule (SAM) to an acceptor molecule (substrate), such as DNA, RNA, proteins and other small molecules and leaving the product SAH (S-Adenosyl-L-Homocysteine). Methyltransferases regulate important cellular and metabolic functions, including antibiotic resistance (Miller *et al.*, 2003, Husain *et al.*, 2010); yet, to date, there has been no report of any methyltransferase structure from *B. vulgatus.*

In this chapter we report the crystal structure of *apo* BVU_3255 and its complex with SAM and SAH, along with the isothermal titration calorimetric studies. These studies revealed that BVU_3255 adopts a classical SAM-dependent methyltransferase architecture with a characteristic α/β fold and thus belongs to the class I Rossmann-

Fold Methyltransferases (RFM) (Bujnicki, 1999). In addition, the structural and sequence analysis provide evidence that BVU_3255 is a small molecule methyltransferase of ubiquinone biosynthesis pathway.

4.2 Materials and Methods

4.2.1 Cloning, expression and protein purification

BVU_3255 was cloned into expression vector pGS21a (GeneScript, USA) and was purified to homogeneity using a two-step procedure involving Ni^{2+}-NTA affinity and gel filtration chromatography in a buffer consisting of Tris-HCl (pH 8.0) and 200 mM NaCl. The homogeneity of BVU_3255 was verified by dynamic light scattering (DLS) experiments. Please refer chapter 2 for detailed cloning, expression and purification.

4.2.2 Crystallization and structure determination

A detailed crystal optimization and data collection strategy are described in chapter two of this thesis. Initial phases were obtained by molecular replacement method using the program Molrep-auto MR (Vagin & Teplyakov, 2010) in CCP4 suite. Where required, manual model building was done using COOT (Emsley & Cowtan, 2004) and alternatively refined by CNS (Brunger *et al.*, 1998). The final models have good stereochemistry, as analysed by PROCHECK (Laskowski *et al.*, 1993). All structure-related figures reported here were generated using PyMol (DeLano & Lam, 2005).

4.2.3 Isothermal titration calorimetry

The ITC binding studies of SAM and SAH with *apo* BVU_3255 were performed using a VP-ITC calorimeter (Microcal, LLC) at room temperature with 0.2 ml of SAM/SAH in the injector cell and 1.8 ml of BVU_3255 in the sample cell. The protein and ligands were kept in a buffer consisting of 20 mM Tris-HCl (pH 8.0) and 200 mM NaCl. 10 µl injection volumes were used for all experiments. Two consecutive injections were separated by 5 min to reset the baseline. The control experiment, consisting of titration of SAM/SAH against buffer, was performed and subtracted from each experiment to adjust for the heat of dilution of ligands. ITC data were analyzed with a single-site fitting model, using Origin 7.0 software (OriginLab Corp.).

4.2.4 PDB Code

The coordinates and structure factors of *apo*, SAM and SAH complex of BVU_3255 were deposited in the RCBS (Bernstein *et al.*, 1977) database with the access code 3T7R, 3T7S, and 3T7T respectively.

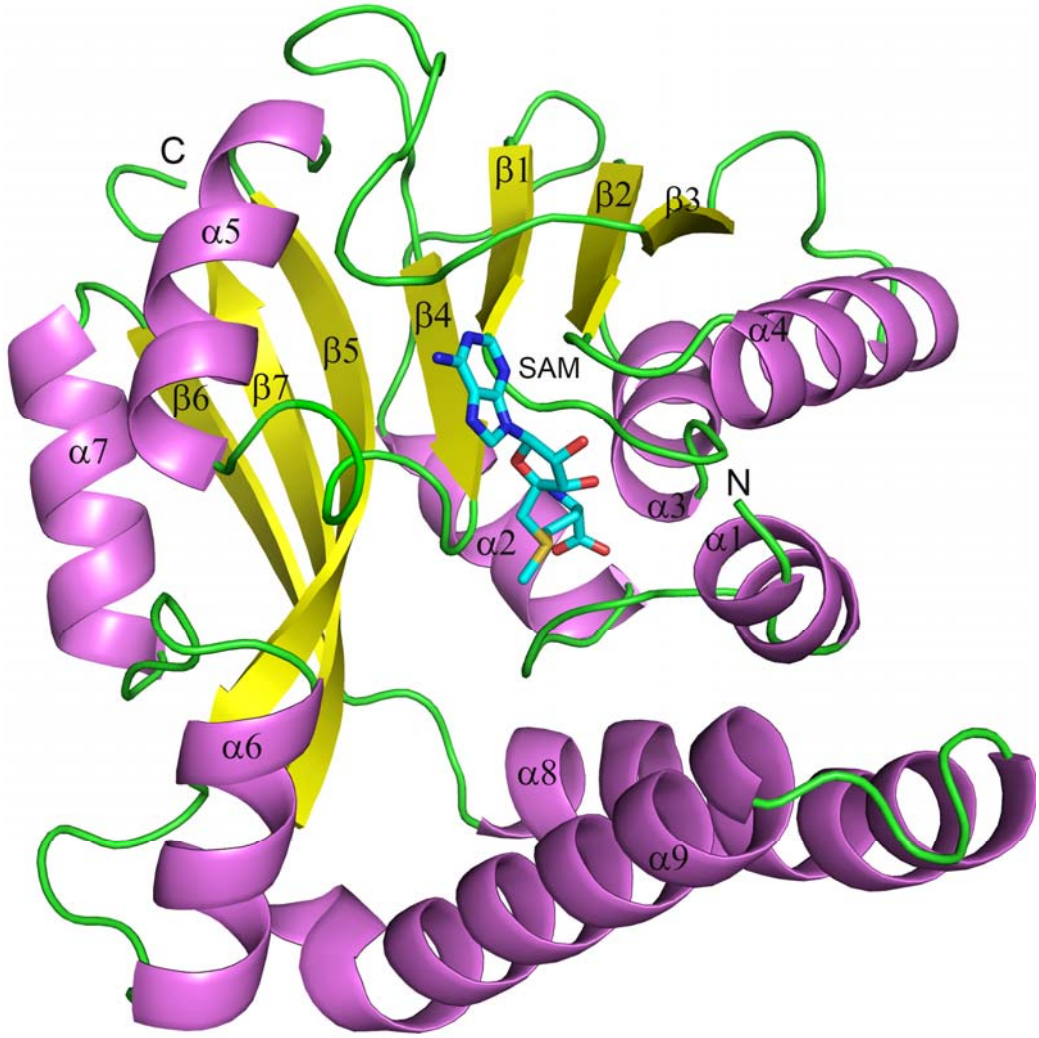
4.3. Results and discussion

4.3.1 Overall structure

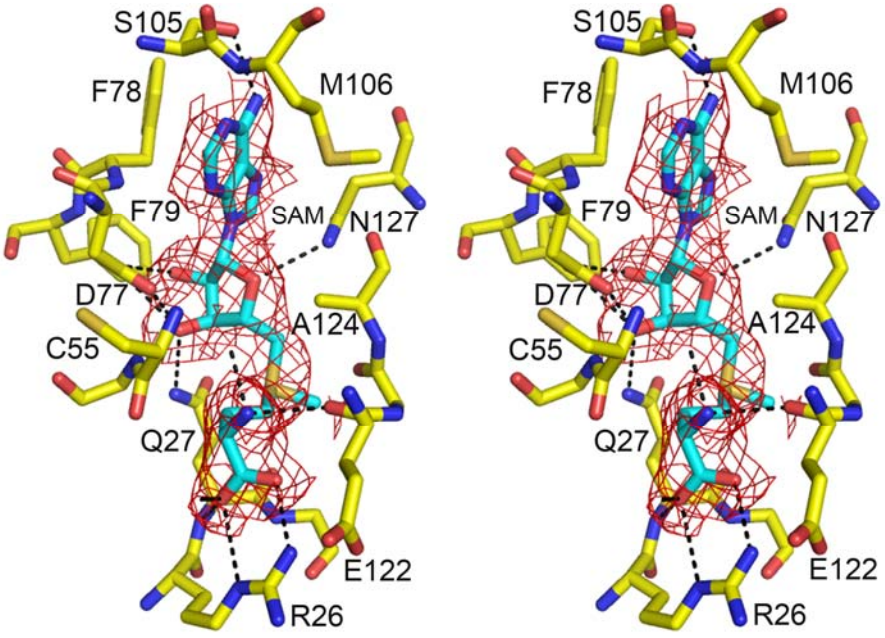
The crystal structure of the *apo* BVU_3255, as well as the co-crystal structures of BVU_3255 in complex with SAM and SAH, were solved by molecular replacement method and refined to a resolution 2.9, 2.2 and 2.5 Å, respectively (Figure 4.1, Table 4.1). There were two molecules in the asymmetric unit of the *apo* BVU_3255 and four molecules in the SAM/SAH complexes. However, the gel filtration and dynamic

light scattering experiments showed that BVU_3255 exists as a monomer in solution. The BVU_3255 construct had a total of 262 amino acids and an N-terminal $(His)_6$-tag. This tag and the first 12 residues at the N-terminus were not well-defined in the electron density map and were not included in the model. The BVU_3255 molecule exists as a single-domain globular protein, with both the cofactor and substrate binding site in the same domain, a common feature observed for the small molecules' methyltransferases (Martin & McMillan, 2002). The active site of BVU_3255 comprised seven-stranded β sheets (β3↓-β2↓-β1↓-β4↓-β5↓-β7↑-β6↓), flanked by three α-helices at both the N- (α2,α3,α4) and C-termini (α5, α6, α7) (Figure 4.1), characteristic of a typical SAM-dependent class I RFM (Bujnicki, 1999). The helices α1, α8 and α9 are clustered together and covered the active site region of BVU_3255.

A


C
α5
β1
β2
β3
β4
β5
β6
β7
α7
SAM
α4
N
α3
α2
α1
α6
α8
α9

B


S105
F78
M106
F79
SAM
N127
D77
A124
C55
Q27
E122
R26
S105
F78
M106
F79
SAM
N127
D77
A124
C55
Q27
E122
R26

C

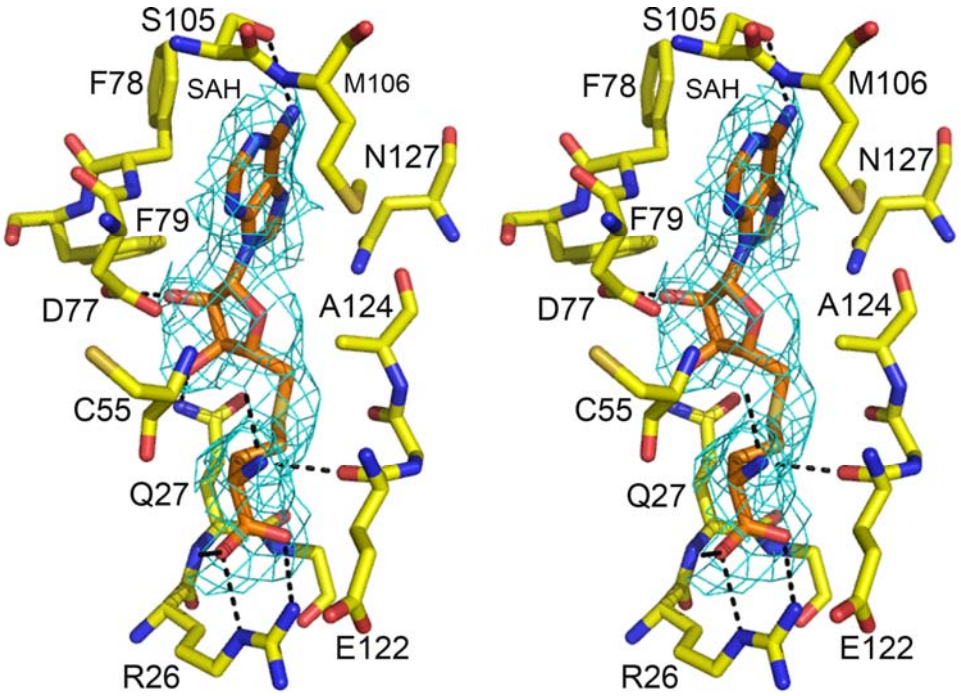


Figure 4.1: A) The crystal structure of BVU_3255-SAM complex. The N- and C-termini and the secondary structural elements are labelled (α helices in violet; β strands in yellow; loops in green). Stereo view of SAM **(B)** and SAH **(C)** complexes at the active site region, with SAM in cyan, SAH in orange and interacting residues from BVU_3255 in yellow. Simulated annealed *Fo-Fc* omit maps (ligands and 3.5 Å of its surroundings were omitted) were contoured at a level of 3σ. For clarity, only selected side chains are shown.

Several structural features separate small molecule and macromolecular methyltransferases. By examining the structures of known methyltransferases, the SAM-binding region is conserved but the substrate binding site is highly variable depending upon the methylation substrate (Martin & McMillan, 2002). In the case of macromolecular methyltransferases, where the substrate is relatively large in size, the substrate binding site has been shown to be located in a different portable domain, with an additional secondary structural element at the C-terminus. In contrast, in

small molecule methyltransferases, the substrate binding site is generally located in the same domain adjacent to the cofactor binding site, and these methyltransferases lack any additional C-terminal secondary structures (Martin & McMillan, 2002). Further, the majority of small molecules methyltransferases have two additional helices at the N-terminal region, and have insertions in the C-terminal part of the core SAM-binding domain (between β5 and α7; β6 and β7). Overall, the current structure points to BVU_3255 as a small molecule methyltransferase.

4.3.2 SAM and SAH binding site

Both SAM and SAH were shown to bind within the active site, with their adenosyl base moiety facing outside and their carboxypropyl moiety buried inside. Both ligands made equal interactions with BVU_3255. Residues, such as Arg26, Gln27, Gly55, Gly57, Gln61, Glu122, and Gly123 were involved in hydrogen bonding contacts (< 3.2 Å) with the carboxypropyl moiety of the ligands, while the ribose ring contacted with residues Asp77, Phe78, Phe79 and Asn127 (Figure 4.1B and 4.1C). The Asp77 position is conserved in all known SAM-dependent methyltransferases, with either an Asp or a Glu residue (Lim *et al.*, 2001). Gly104, Ser105, and Met106 made interactions with the adenosyl moiety of the ligands. The superposition of *apo* BVU_3255 onto both SAM and SAH complexes yielded an RMSD less than 0.6 Å for 245 Cα atoms.

Table 4.1: Crystallographic refinement statistics

	Native	SAM	SAH
Refinement and quality [a]			
Resolution range (Å)	30.0 – 2.9	30.0 - 2.2	30.0 - 2.5
R_{work} [b] (no. of reflections)	0.21(9971)	0.23(41510)	0.22(28599)
R_{free} [c] (no. of reflections)	0.28(746)	0.26(2176)	0.27(2163)
RMSD bond lengths (Å)	0.009	0.008	0.009
RMSD bond angles(°)	1.226	1.244	1.305
Average B-factorse[d] (Å^2)			
Main chain	41.7	40.5	31.3
Side chain	42.7	42.5	32.4
Ligand	51.8	33.7	22.4
Ramachandran plot			
Most favored regions (%)	91.0	90.1	92.1
Additional allowed regions (%	8.5	9.6	7.9
Generously allowed regions (%)	0.5	0.4	0.0
Disallowed regions (%)	0.0	0.0	0.0

[a]Reflections with I>σ was used in the refinement.
[b]R_{work} = $|F_{obs} - F_{calc}|/|F_{obs}|$ where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively.
[c]R_{free} = as for R_{work}, but for 5% of the total reflections chosen at random and omitted from refinement.
[d]Individual B-factor refinements were calculated.
*The high resolution bin details are in the parenthesis.

4.3.3 Isothermal titration calorimetry

The interactions between BVU_3255 and SAM/SAH were studied by ITC experiments (Figure 4.2). Consistent with the crystal structure, the ITC experiments showed a single-site binding model for both ligands. The binding parameters for SAH were: K_a=2.955×10^4 M^{-1} (± 0.2404 × 10^4), ΔH=−5.74 kcal/mol (±0.219), and N=1.23.±0.20; and for SAM were: K_a=4.45×10^4 M^{-1} (±0.63×10^4), ΔH=−5.533 kcal/mol (±0.359), and N=0.85±0.01 (where K_a is the association constant, ΔH is the change in enthalpy, and N is the number of binding sites). These ITC experiments showed that both SAM and SAH bind to BVU_3255 with a similar affinity, consistent with the observation from the complex structures that both ligands had similar interactions with BVU_3255.

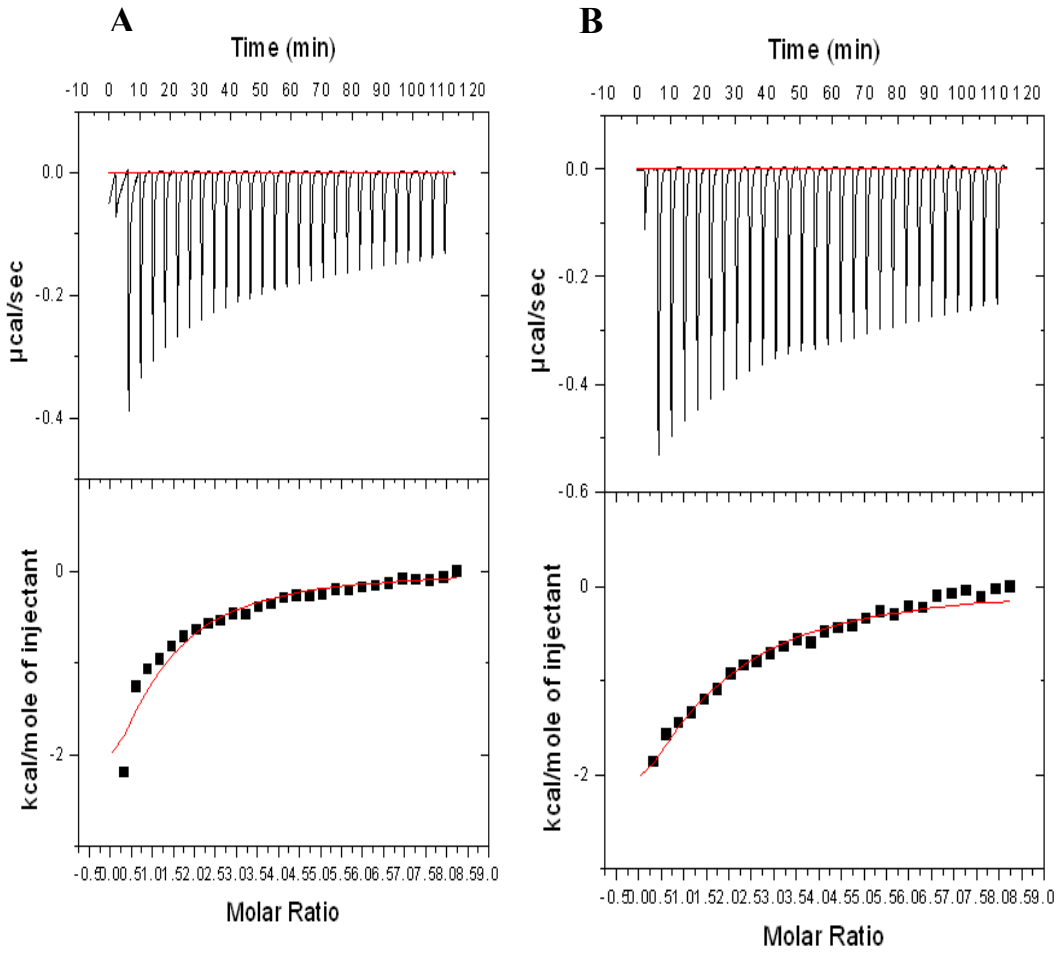


Figure 4.2: ITC profile of BVU_3255 titrated against (A) SAM and (B) SAH. The upper panels show the raw ITC data for injection of ligands into the sample cell containing the native protein. The peaks were normalized to the ligand:protein molar ratio and were integrated as shown in the bottom panels. Solid dots indicate the experimental data, and their best fit was obtained from a nonlinear least squares method, using a one-site binding model depicted by a continuous line.

4.3.4 Sequence and structural homology

Class I methyltransferase share low overall sequence homology. Depending on the methylation target, their substrate binding site varies. However they share structure and sequence similarity at the active site. Previous bioinformatics studies on Class I methyltransferases have shown that the structure of the AdoMet binding site is conserved in the primary sequences of four short signature motifs designated Motifs I, Post I, II, and III. A PSI-BLAST search revealed that BVU_3255 shared sequence

homology with several SAM-dependent class I RFMs. It exhibited a maximum identity of 62% with a methyltransferase of known function involved in ubiquinone biosynthesis from *B. Xylanisolvens*. This was followed by 48% identity with a similar ubiquinone methyltransferase from *Gordonibacter pamelaeae* (Fig 4.3). BT_2972 and BVU_3255 share ~59% sequence identity. This sequence analysis indicated that BVU_3255 is a small molecule methyltransferase and might be involved in methylating the intermediates in the ubiquinone/menaquinone biosynthesis pathway.

In the ubiquinone/menaquinone biosynthesis pathway there are three methylation reactions catalyzed by UbiG and UbiE methyltransferases (Meganathan, 2001). Previously Xu *et al*. annotated BVU_2542 as the UbiE in *B. vulgatus ATCC 8482* (Xu *et al.*, 2007). BVU_2542 is common for ubiquinone and menaquinone biosynthesis pathways and catalyzes a methylation reaction in both the pathways. Thus, BVU_3255 may be the UbiG SAM-dependent O-methyltransferase of the ubiquinone biosynthesis pathway; UbiG enzymes act at two intermediates in the ubiquinone biosynthetic pathway.

A search for structurally similar proteins in DALI server (Holm & Sander, 1995) resulted in similarity with several proteins belonging to class 1 RFMs. The highest homology with known function was found to be mycolic acid cyclopropane synthases.

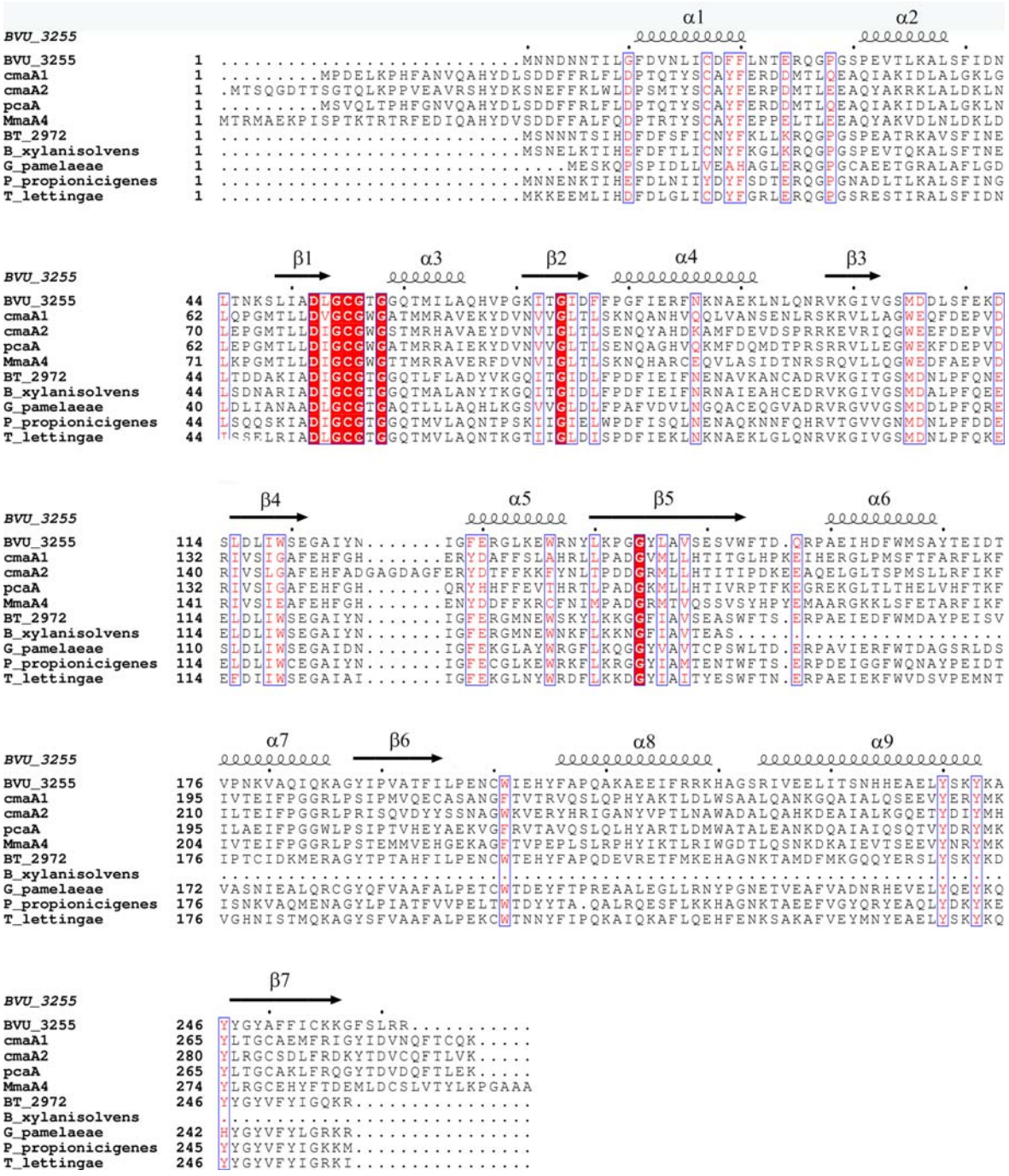


Figure 4.3: Structure-based sequence alignment of BVU_3255 with different mycolic acid methyltransferases from *M. tuberculosis*. Sequence similarities are highlighted in red, whereas sequence identities are shown as white letters on a red background. The secondary structural elements are shown on the top. Structural alignment was performed in DALI server while sequence alignment was carried out using ClustalW (Chenna *et al.*, 2003) with default parameters. The diagram shown here was generated using ESPript (Gouet *et al.*, 1999).

Crystal structures of three members of mycolic acid cyclopropane synthases have been reported, namely CmaA1, CmaA2, and PcaA (Huang *et al.*, 2002). All three family members are highly similar to each other and to BVU_3255 (DALI score 22.1, RMSD 3.0 Å for Cα 235 residue pdb code 1KPG for CmaA1). Mycolic acid cyclopropane synthases are also SAM-dependent proteins, with a similar α/β Rossmann Fold, and are involved in the site-specific methylation of mycolic acid, a major component of the *Mycobacterium* cell wall (Huang *et al.*, 2002). All of these structures are similar, and the structure-based sequence alignment shows several conserved key residues in their active site region (Figure 4.3).

4.3.5 Inferred substrate binding site

A one-to-one structural comparison between BVU_3255-SAH/SAM and CmaA1-SAH-CTAB (substrate, Cetyltrimethylammonium Bromide) (PDB: 1KPG) showed the existence of a highly conserved active site and substrate-binding site. The orientation of the cofactor for both proteins is the same (Figure 4.4A). On closer examination of the active site, we observed that the carboxypropyl portions of SAH from both proteins are in the same plane, but the ribose and adenosyl moieties are not. The N6 atoms of the adenosyl bases are approximately 3.5 Å apart. Nonetheless, the cofactors in both proteins occupy the same region in the active site. In addition, an extra electron density map was observed in the *apo* BVU_3255 in the same place where the CTAB (substrate) binds CmaA1. This electron density lies adjacent to the transferable methyl group of SAM (Figure 4.4B) and is surrounded by bulky amino acids that make the environment highly hydrophobic for interaction with the hydrophobic ligand. These observations s suggests that this additional electron density might be a naturally bound substrate of BVU_3255 (Figure 4.4B).

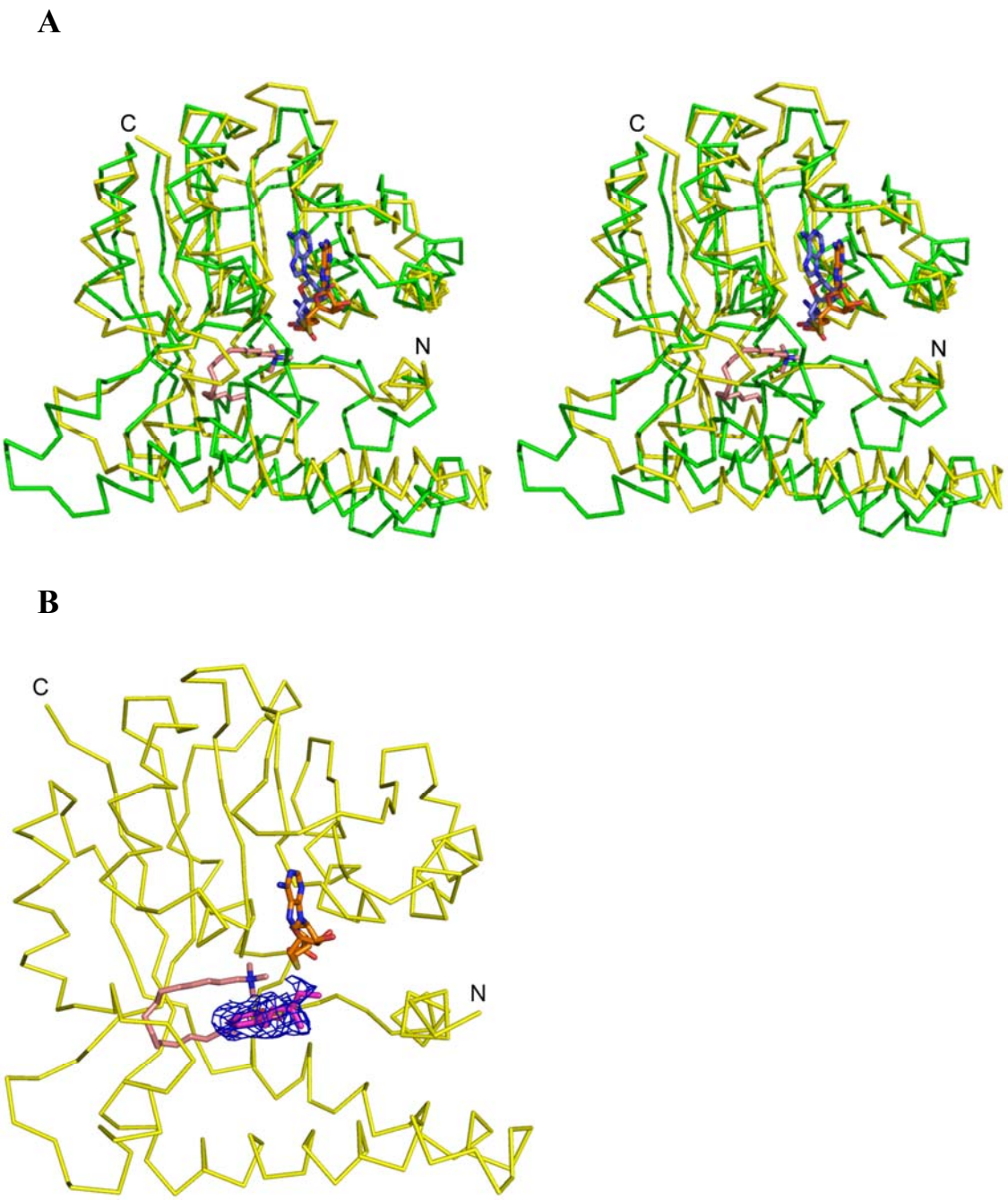


Figure 4.4: **A)** Stereo view of superposition of Cα trace of BVU_3255-SAH (yellow) and mycolic acid cyclopropane synthase CmaA1-CTAB-SAH complex (green) from *M. tuberculosis* (PDB 1KPG). SAH of BVU_3255 is shown in orange and SAH of CmaA1 in light blue. The active site of the two proteins superimposed well, and SAH occupies the same region in the active site. Carboxypropyl moieties of SAH from both the proteins are lying in same plane but ribose and adenosyl moieties are separated by 3.5Å. The substrate CTAB of CmaA1 is shown in pink. **B)** Based on the superposition the substrate binding site on BVU_3255-SAH complex was inferred. A part of the proposed substrate (5 carbon atoms of the aliphatic chain attached to benzene ring) of BVU_3255 was modelled into the observed additional electron density at the region where CTAB is mapped onto BVU_3255. Simulated annealed *Fo-Fc* omit maps (ligand and 3.5 Å of its surroundings were omitted) were contoured at a level of 3σ.

Sequence analysis suggested that BVU_3255 is involved in the ubiquinone biosynthesis pathway and may function to methylate two intermediates of the pathway *i.e.* 2-polyprenyl-6-hydroxyphenol and 2-polyprenyl-3-methyl-5-hydroxy-6-methoxy-1, 4-benzoquinone. The substrate CTAB (for CmaA1) and the proposed BVU_3255 substrates are similar, except BVU_3255 substrates have a benzene ring that contains the methylation site (Figure 4.4B). The observed electron density allowed us to build a benzene ring and 5 carbon atoms of the aliphatic chain attached to benzene ring of the substrate. However, no well-defined density was observed for the rest of the alkyl side chain. Overall, our analysis and observations support the view that substrates for BVU_3255 are similar to mycolic acid, with a long aliphatic chain attached to the benzene ring, although further experimental verification is warranted.

In summary, in this chapter we report the first representative crystal structures of a SAM-dependent methyltransferase, BVU_3255, from an antibiotic-resistant bacterium, *B. vulgatus*. The isothermal titration calorimetric studies showed that both SAM and SAH bind to BVU_3255 with a similar affinity. The structural and sequence analysis indicated that BVU_3255 is a small molecule methyltransferase that may be involved in catalyzing (methylates) the two O-methylation reaction steps for intermediates in the ubiquinone biosynthesis pathway.

Chapter 5

Conclusions and Future Directions

5.1 Conclusion

In this study we have met our objectives of studying the two O-methyltransferase BT_2972 and BVU_3255. These studies lead to the first complete structural characterization of representative methyltransferase structures from the respective *Bacteroides* bacteria. Structure of BT_2972 adopts a typical class I Rossmann-like AdoMet-binding fold at the catalytic site that is common to all class I AdoMet-dependent methyltransferases, comprising a seven-stranded β-sheet (β3↓β2↓β1↓β4↓β5↓β7↑β6↓) flanked by three α helices on both sides. The *apo* and ligand bound BT_2972 structure are similar other than the active sites. The comparison between the structures of *apo* and AdoMet/AdoHcy revealed the closed and open forms of the active site which could potentially regulate the movement of the cofactors and the interactions with the substrate. The isothermal titration calorimetric studies showed that BT_2972 binding affinity with AdoHcy and AdoMet are slightly different.

Similarly, the all BVU_3255 crystal structures also adopt class I Rossmann fold at the active site. The three structures are very similar and no major conformational change was observed between *apo* and ligand bound structures. An additional electron density was observed in the native structure at the active site, next to the AdoMet binding site. This density was modelled with proposed substrate. The isothermal titration calorimetric studies showed that both AdoMet and AdoHcy bind with BVU_3255 in a similar affinity.

The structural and sequence analysis suggested that both proteins BT_2972 and BVU_3255 are a small molecule methyltransferase and probably catalyzes (methylates) two O-methylation reaction step intermediates in the ubiquinone

biosynthesis pathway. ITC study of ligand binding suggested that BT_2972 bind to the ligand (AdoMet/AdoHcy) with 10 fold higher affinity than BVU_3255.

In both the structures, the substrate binding site is located within the same domain where AdoMet/AdoHcy binds. This is the characteristic feature of small molecule methyltransferase. Structure and sequence similarity allowed us to predict the target substrate for the two methyltransferases. Structural comparison suggests that these two methyltransferases are very similar to the mycolic acid methyltransferase from *M. tuberculosis*; the AdoMet/AdoHcy binding sites are conserved and the substrate binding site is next to AdoMet/AdoHcy in order to facilitate the methyl transfer.

5.2 Future Directions

Despite the fact that we have presented the crystal structural and biophysical characterization of two O-methyltransferases from ubiquinone pathway, some of experiments need to be performed to validate the proposed substrate. The predicted substrate for these two methyltransferases, i.e. the intermediates of the ubiquinone biosynthesis pathway -1) 2-polyprenyl-6-hydroxyphenol and 2) 2-polyprenyl-3-methyl-5-hydroxy-6-methoxy-1, 4-benzoquinone needs to be verified experimentally both *in vitro* and *in vivo* study. *In vitro*, this can be performed by incubating the substrate with enzymes and AdoMet followed monitoring the profile using Thin Layer chromatography (TLC). For *in vivo* based study, it might be verified by knocking out the genes and then monitoring quantity of the intermediates of ubiquinone pathway.

Our future work will also focus on co-crystallize the ternary complex of enzymes-substrate-AdoMet/AdoHcy to understand the exact mechanism of methylation.

Part II

Chapter 6

Interactions of the Intrinsically Disordered Neuron Specific Substrate Proteins Neuromodulin (Nm) and Neurogranin (Ng) with Calmodulin (CaM) Revealed by Biophysical and Structural Studies

6.1 Introduction

6.1.1 Neuron

The nervous system is an organ system made of a network of specialized cells known as neurons (Figure 6.1). The nervous system controls the actions and behaviour of the animals by transmitting the signals between different parts of the body. It consists of two parts called the central nervous system and the peripheral nervous system. The central nervous system (CNS) consists of the brain, the spinal cord, (in some cases) retina and the cranial nerves. CNS integrates the information that it receives from, and coordinates the activity of, all parts of the bodies of animals. The peripheral nervous system (PNS) consists of the nerves and ganglia and together with CNS, it controls the behaviour of animal (Mineta *et al.*, 2003).

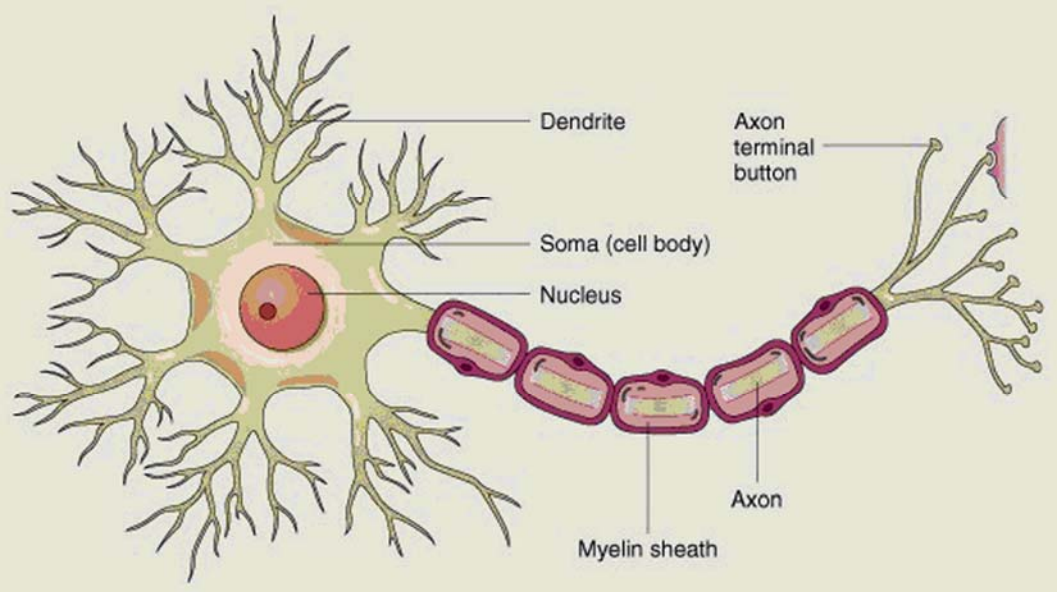


Figure 6.1: Structure of a typical neuron cell. (This figure is adapted from http://scienceblogs.com/purepedantry/2006/07/background_to_the_20_year_coma.php)

The neuron transmits information in the form of chemical and electrical signal. Neurons have three distinct structures known as the dendrites, the cell body and the axon. The axon and dendrites are responsible for transmitting and receiving the information (Figure 6.1 and 6.2). The connections between two neurons are known as synapses. Neurons release chemicals known as neurotransmitters into these synapses to communicate with other neurons (Figure 6.2). Unlike other body cells, neurons do not reproduce shortly after birth. Other body cells keep on reproducing new cells throughout the life. But neurons stop reproducing new neuron after birth. Because of this, some parts of the brain have more neurons at birth than later in life because neurons die but are not replaced. There are several types of neurons responsible for different tasks in the human body. Sensory neurons are involved in carrying information from the sensory receptor cells throughout the body to the brain. Motor neurons are responsible for transmitting information from the brain to the muscles of the body. Inter neurons are involved in communicating information between different neurons in the body.

6.1.2 Action Potentials

Neurons communicate by transmitting information within the neuron and from one neuron to the next. The dendrites of neurons receive information from sensory receptors or from other neurons. This information is then passed down to the cell body and on to the axon. Information in the neuron travels in the form of electrical signals and chemical messengers. The information (after arriving at the axon) travels down the length of the axon in the form of an electrical signal known as an action potential (Barnett & Larkman, 2007).

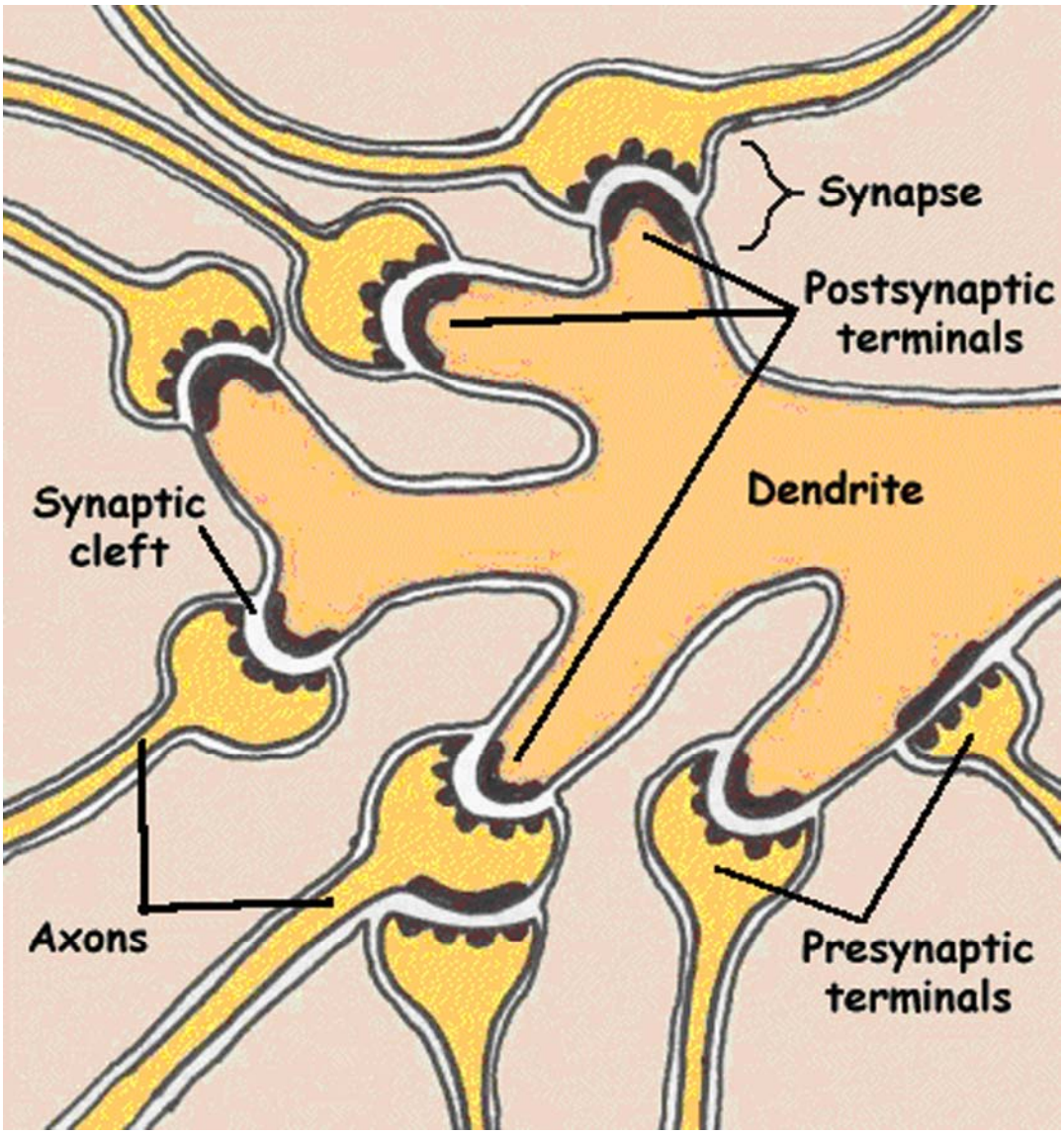


Figure 6.2: The synapse is the connection between nerve cells (neurons) in animals including humans. The synapse joins the axons in one neuron to the dendrites in another. (Adapted from http://www.mult-sclerosis.org/synapse.html)

6.1.3 Communication between Synapses

Once an action potential has reached the end of an axon, the information from axon transmits across the synaptic gap to the dendrites of the adjoining neuron by one of the two processes. In some cases, the action potential can almost instantaneously bridge the gap between the neurons and continue along its path. In other cases,

neurotransmitters are released that carry the information from one neuron to the next. Neurotransmitters are chemical messengers that are released from the axon terminals, travels cross the synaptic gap and reach the receptor sites of other neurons. These neurotransmitters attach to the receptor site and are reabsorbed by the neuron to be reused. Some examples of neurotransmitter are acetylcholine, endorphins, and dopamine.

6.1.4 Long term potentiation (LTP)

LTP is a long-lasting enhancement in signal transmission between two neurons that results from stimulating them synchronously. It is one of the several phenomena underlying synaptic plasticity and is considered as one of the major cellular mechanisms that underlie learning and memory. LTP is thus associated with long-term memory formation. LTP and long-term memory are triggered rapidly, both depends upon the synthesis of new proteins, each has properties of associativity, and each can last for many months (Malenka & Bear, 2004). LTP enhances synaptic transmission by improving the ability of two neurons (presynaptic and postsynaptic) to communicate with one another across a synapse.

6.1.5 Long term depression (LTD)

LTD is a reduction in the efficacy of neuronal synapses lasting hours or longer. LTD has been found to occur in different types of neurons at the synapses. LTD is the opposing process to LTP. LTD mainly takes place because of decrease in postsynaptic receptor density or a decrease in presynaptic neurotransmitter release. It is important for motor learning, clearing of old memory traces. LTD is also found to

selectively weaken specific synapses in order to make constructive use of synaptic strengthening caused by LTP.

In this part of the thesis we focus on two neuron specific proteins Neuromodulin(Nm) and neurogranin (Ng) and their interacting partner in neuron Calmodulin (CaM).

6.2 Growth Associated Protein-43 (GAP-43, Nm)

The neuron specific substrate protein Neuromodulin (23.6kda, aka Nm, B50, GAP-43, F1, or P-57) is acid stable and hydrophilic in nature. It was discovered and isolated by virtue of its novel CaM binding properties (Benowitz & Routtenberg, 1987, Andreasen *et al.*, 1981). Neuromodulin (here after referred as Nm) was purified by application of detergent-solubilized bovine brain membranes to CaM-sepaharose in the presence of excess Ca^{2+} chelator and then eluted from CaM-sepharose with a Ca^{2+} containing buffer. Nm binds CaM through a short conserved IQ motif [I,L,V]QxxxR[G,x]xxx[R,K] under a low Ca^{2+} micro-environment while reduces that affinity when Ca^{2+} level increases. IQ motif contains mainly basic and hydrophobic residues, in particular lysine and arginine (Slemmon *et al.*, 1996), with a high propensity to form an amphipathic α helix and is thought to be the main structural and functional domain. A proposed biochemical function for Nm is to sequester CaM in the vicinity of CaM – activated enzymes under low Ca^{2+} condition. Elevations of intracellular free Ca^{2+} would promote dissociation of CaM from Nm, allowing rapid CaM activation of enzyme activity. The Nm- CaM interaction is also regulated by protein kinase C (PKC) mediated phosphorylation. PKC phosphorylates Ser41 residue within IQ motif of Nm and reduces its affinity towards CaM (Alexander *et al.*, 1987). Phosphorylation of Ser41 residue of Nm by PKC regulates neurite

formation, regeneration, and synaptic plasticity (Benowitz & Routtenberg, 1997a) (Figure 6.3).

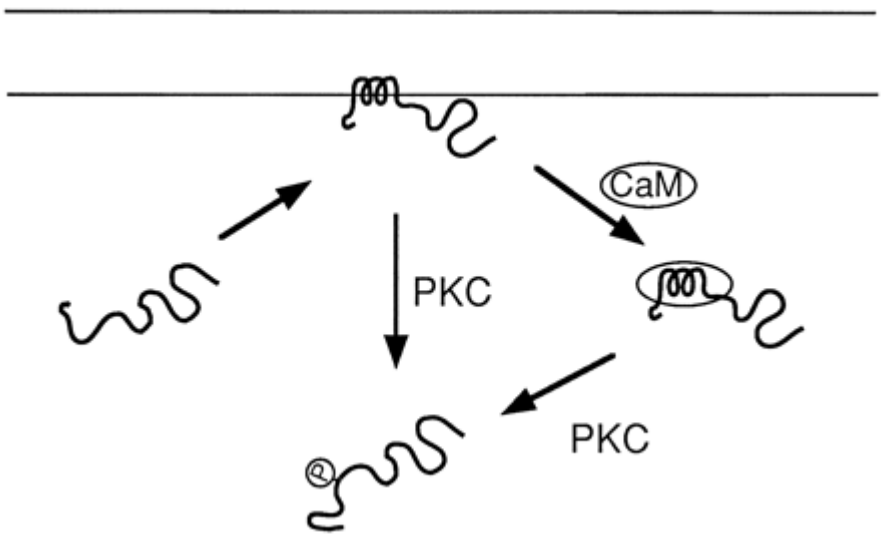


Figure 6.3: Schematic model of Nm membrane interaction and its regulation by PKC and calmodulin. Figure adapted from (Hayashi *et al.*, 1997).

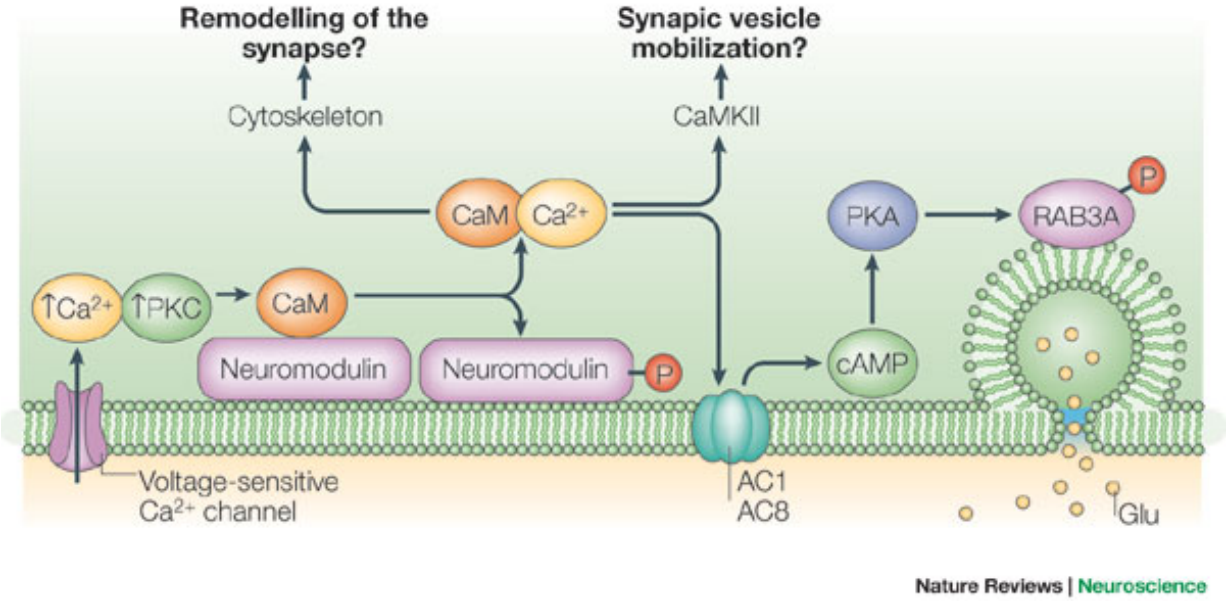


Figure 6.4: Role of CaM and Nm at presynaptic loci in LTP. This figure is adapted from(Xia & Storm, 2005).

The amino acid composition of the Nm is very unusual. The protein contains a single phenylalanine with no other aromatic residue, while alanine, glutamic acid, lysine and proline are abundant. Nm can attach to the membrane via a dual palmitoylation sequence on Cys3 and Cys4. It is also found as non bound form in cytoplasm and centrosome. This sequence (Cys3 and Cys4) targets Nm to lipid rafts (Figure 6.3). Centrosome and mitotic spindles are found to be mislocalized in the neuron cells that unable to express Nm during cell differentiation. Nm also plays a critical role in the development of the central nervous systems (Strittmatter *et al.*, 1995). Nm is an intrinsic determinant of structural change at the synapse. Nm binds to actin (He *et al.*, 1997) and spectrin (Riederer & Routtenberg, 1999), and by such protein–protein interactions affects morphological change at the synapse. This synaptic plasticity has been implicated in memory storage processes (Cammarota *et al.*, 1997).

In the neuron cell, Nm is localized at the cytoplasmic face of the plasma membrane of axonal and presynaptic terminals of synapses. During the neural development, Nm accumulates in axonal growth cones and helps them to navigate exactly to their appropriate targets (Benowitz & Routtenberg, 1997b, Shen *et al.*, 2002). Nm is also involved in neurite extension and neuronal plasticity, neuroregeneration, regulation of neurotransmitter release at presynaptic terminal and long term potentiation (LTP) (Denny, 2006, Mosevitsky *et al.*, 2001, Mosevitsky, 2005, Doster *et al.*, 1991, Fitzgerald *et al.*, 1991, Gianotti *et al.*, 1992) Figure 6.4). Nm is not reported to found in tissue other than brain, retina and spinal cord (Cimler *et al.*, 1985). In the nervous system, Nm overexpression is accompanied by enhanced learning and regenerative capabilities (Routtenberg *et al.*, 2000). Recent study suggests that the protein kinase

C phosphorylation site on Nm differentially regulates performance of three different memory-associated tasks (Holahan & Routtenberg, 2008).

Genetically over expression of Nm is able to enhance the learning and LTP in transgenic mice. A substitution mutation in Nm at Ser41 (which is the phosphorylation by protein kinase C) to Ala has been shown to block phosphorylation. The over expression of mutated Nm does not show any enhancement in learning. These findings suggest that a growth-related gene (GAP-43/Nm) regulates learning and memory and also suggest that the Nm phosphorylation site is the main target for enhancing cognitive ability. Since learning increases Nm phosphorylation, one might expect that a transgenic mouse that over expresses phosphorylable Nm would demonstrate enhanced learning. A contradiction to this prediction is that such genetically enhanced learning would not occur if the PKC site of the over expressed Nm were mutated to prevent its phosphorylation. Figure 6.5 explains the proposed mechanisms of enhanced learning in transgenic animals over expressing Nm. At the presynaptic terminal PKC is activated by an NMDA-dependent postsynaptic retrograde signal. Phosphorylated endogenous Nm (black circle) interacts only with calcium-sensor proteins of the exocytotic protein machine (EPM) to enhance release when intraterminal calcium is raised sufficiently.

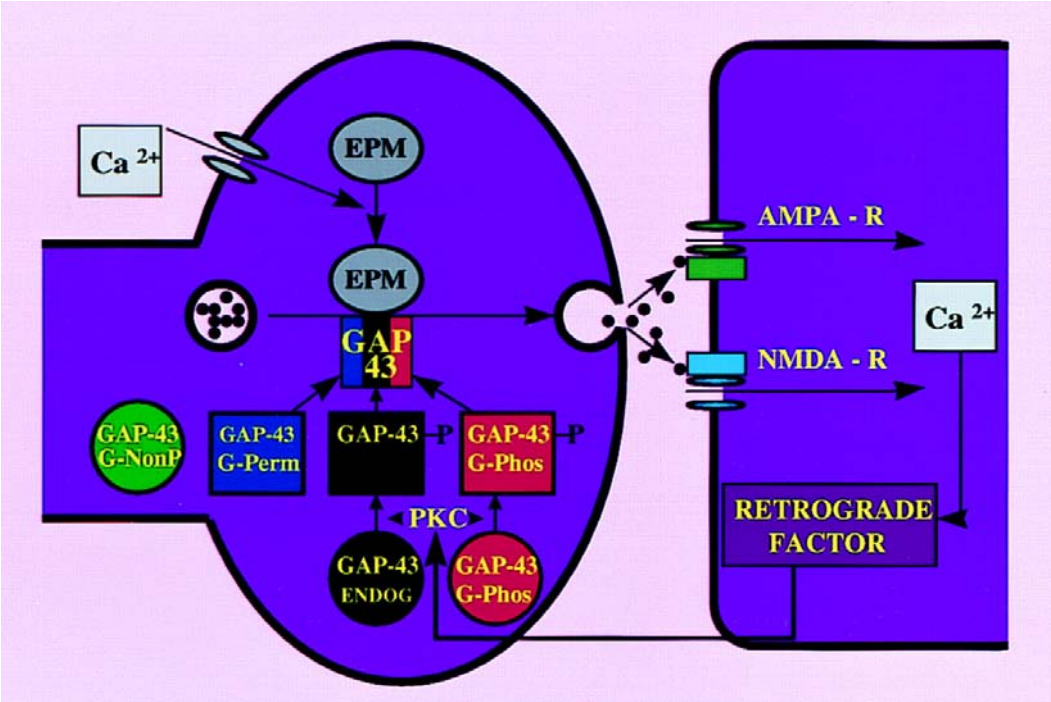


Figure 6.5: A Proposed mechanistic model to elucidate the enhanced LTP in transgenic animals over expressing Nm. Figure adapted from (Routtenberg *et al.*, 2000)

But this is not sufficient to induce LTP. So, PKC also phosphorylates both endogenous and transgenic G-Phos (red circle) and now the terminal is "primed" to release more transmitters upon subsequent depolarization of the presynaptic terminal. Because the G-Perm *v*ariant of Nm (blue square) can bind to activated EPM without PKC phosphorylation, this mutated form of Nm does not require the influence of NMDA receptor activation.

Nm also has been shown to interact with a number of different molecules, like PKC, PIP2, actin, CaM, spectrin, palmitate, synaptophysin, amyloid and tau protein. Its important role in memory and information storage is executed through its cell

biological mechanisms of phosphorylation, palmitoylation, protein-protein interaction and structural remodeling via actin polymerization.

6.3 Neurogranin (Ng)

Another neuron specific PKC substrate protein called neurogranin (7.6 kDa, Ng) is also acid stable but expressed at postsynaptic cytosol loci in several telencephalic areas such as cerebral cortex, hippocampus, amygdale, striatum and olfactory bulb in mammalians (Neuner-Jehle *et al.*, 1996). Its expression controlled by thyroid hormone (Bernal *et al.*, 1992), retinoic acid, and vitamin A(Husson *et al.*, 2003), is developmentally regulated, showing a relatively low expression in the embryonic and neonatal stages and a dramatic increase from 2 to 3 weeks of age (Alvarez-Bolado *et al.*, 1996). In neurons, it is found in higher concentration at dendritic spines where it participates in synaptic signalling events.

Ng is the member of calpactin protein family which includes other protein such as Nm, peptide protein 19 (PEP-19), Igloo and sperm protein 17 (SP-17) (Ran *et al.*, 2003). Similar to other proteins of this family Ng binds to CaM in low Ca^{2+} environment and releases CaM under high Ca^{2+} condition. Its CaM-binding affinity is modulated by phosphorylation (Huang *et al.*, 1993), oxidation (Ran *et al.*, 2003) or S-glutathiolation(Li *et al.*, 2001). Ng is involved in the regulation of CaM and CaM – activated protein enzymes, such as nitric oxide synthase, Ca^{2+}/CaM-dependent protein kinase II and CaM-dependent adenylate cyclise (Huang *et al.*, 2004). Since most CaM-activated proteins are involved in long term potentiation (LTP) and long term depression (LTD), the timing pattern of Ng gene expression and protein synthesis appears to coincide with CaM-related synaptogenesis and development. The

Ng protein may play an important role in the neuroplasticity mechanism of learning and memory (Huang *et al.*, 2004). Ng features an IQ motif that mediates its interaction with CaM and also with phosphatidic acid (PA) and IQ motif becomes phosphorylated by protein kinase C (PKC) at Ser36. Ser36-phosphorylated Ng is unable to bind either CaM or PA (Diez-Guerra, 2010). In addition, changes in Ng expression levels may be associated with some cerebral pathophysiologic conditions, such as hypothyroidism, sleep deprivation, and brain aging(Li *et al.*, 2003). Transgenic knock-out mice lacking expression of Ng shows deficiency in learning and memory, activation of Ca^{2+}/CaM-dependent protein kinase II, attenuation of PKC and cAMP-dependent protein kinase signal transduction (Miyakawa *et al.*, 2001, Wu *et al.*, 2002)

Synaptic plasticity in CA1 hippocampal neurons depends on Ca^{2+} elevation and the resulting activation of CaM-dependent enzymes. Induction of long-term depression (LTD) depends on calcineurin, whereas long-term potentiation (LTP) depends on Ca^{2+}/ CaM-dependent protein kinase II (CaMKII). The concentration of CaM in neurons is considerably less than the total concentration of the *apo* CaM-binding proteins Ng and NM, resulting in a low level of free CaM in the resting state. Ng is highly concentrated in dendritic spines.

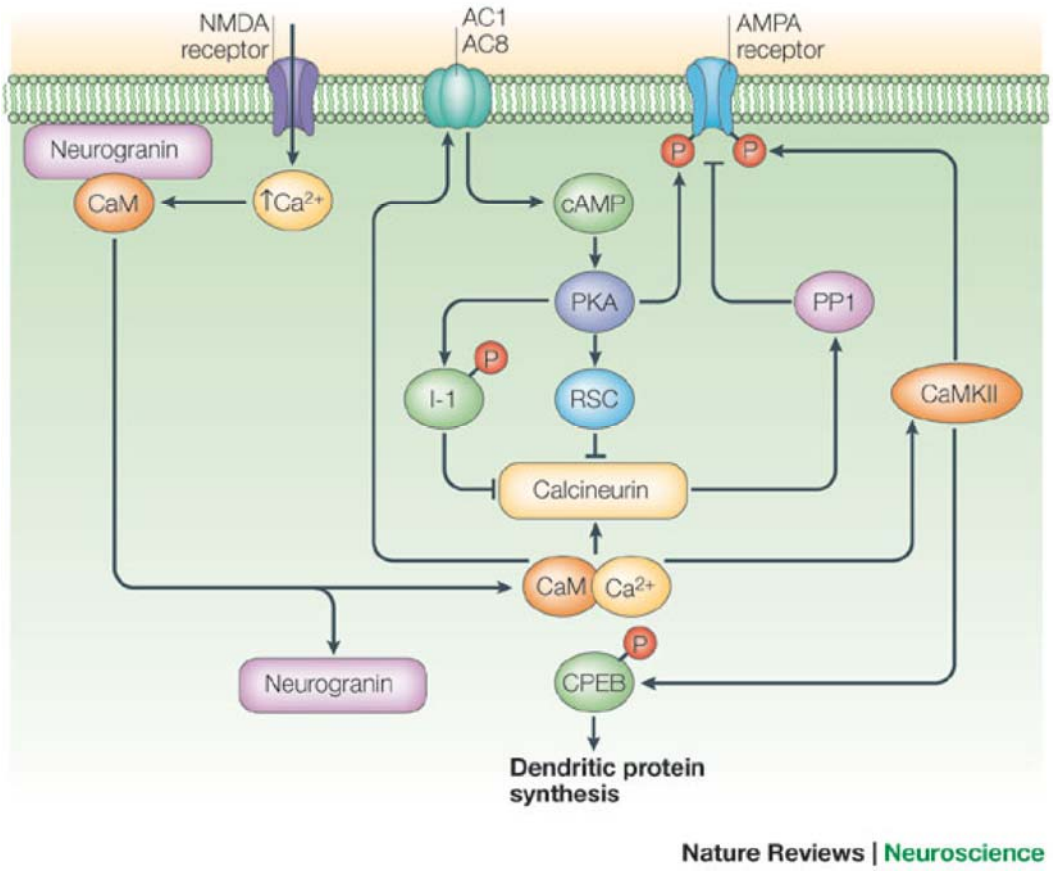


Figure 6.6: Role of Ng and CaM at postsynaptic loci in LTP and LDP. Figure adapted from (Xia & Storm, 2005).

Ng also acts as a pro-apoptotic factor and regulates the cell death by apoptosis. It has been shown that the overexpression of Ng in a T-cell line (dependent on IL2) can induce apoptosis even in the presence of IL2. This apoptosis inducement by Ng coincides with a rise in intracellular Ca^{2+} levels. The effect can be prevented by treatment with a chelator of Ca^{2+}. Mutants of Ng that are unable to bind CaM fail to raise intracellular Ca^{2+} levels and are unable to cause apoptosis (Devireddy & Green, 2003).

The role of Ng in synaptic plasticity has been elucidated with emphasis on the interaction of CaM with Ng, calcineurin, and CaMKII (Figure 6.6). The Ca^{2+} transients that occur during LTD or LTP induction affect CaM which results in activation of calcineurin and CaMKII affects AMPA receptor-mediated transmission. Knocking out of Ng protein strongly diminishes the LTP and slightly enhances LTD. Thus CaM should be stored in spines in the form of rapidly dissociating CaM-Ng complexes to induce LTP (Zhabotinsky *et al.*, 2006).

Ng enhances synaptic strength through its interaction with CaM. Learning-correlated plasticity at CA1 hippocampal excitatory synapses is dependent on neuronal activity and (NMDAR) activation (Figure 6.7). Ng enhances postsynaptic sensitivity and increases synaptic strength in an activity- and NMDA receptor-dependent manner. In addition, Ng-mediated potentiation of synaptic transmission mimics and occludes LTP. Expression of Ng mutants that lack the ability to bind, or dissociate from, CaM fails to potentiate synaptic transmission, strongly suggesting that regulated Ng–CaM binding is necessary for Ng-mediated potentiation. Thus, Ng–CaM interaction can provide a mechanistic link between induction and expression of postsynaptic potentiation (Gerges *et al.*, 2009).

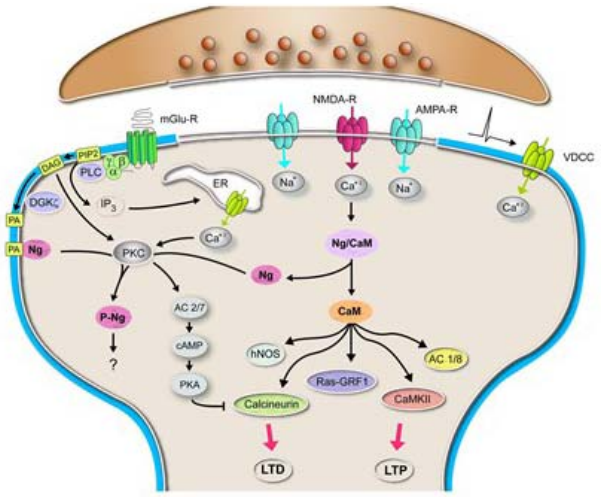


Figure 6.7: Schematic diagram illustrating Neurogranin involvement in postsynaptic signaling. This figure is adpated from http://www.cbm.uam.es/mkfactory.esdomain/webs/cbmso/plt_LineasInvestigacion.aspx?IdObjeto=33&ChangeLanguage=2

Learning-related modifications of synaptic transmission at CA1 hippocampal excitatory synapses are activity- and NMDA receptor-dependent. As mentioned above a postsynaptic increase in Ca^{2+} is absolutely required for synaptic plasticity induction, the molecular mechanisms underlying the transduction of synaptic signals to postsynaptic changes are not clearly understood. Zhong and Gerges have reported that the postsynaptic CaM-binding protein Ng enhances synaptic strength in an activity- and NMDAR-dependent manner (Zhong & Gerges, 2010).

6.4 Calmodulin (CaM)

CaM, a 148 residue protein (16.7 kDa) found in all eukaryotic cells, has been extensively studied as a primary Ca^{2+}-binding protein. CaM is expressed in almost all cell types and assumes different subcellular locations like the cytoplasm, endoplasmic reticulum, sarcoplasmic reticulum, within organelles, or associated with the plasma or organelle membranes (Fok *et al.*, 2008). CaM is thought to mediate processes such as inflammation, metabolism, apoptosis, smooth muscle contraction, intracellular movement, short-term and long-term memory, nerve

growth and the immune response (Geiser *et al.*, 1991). Post-translational modifications, such as phosphorylation, acetylation, methylation and proteolytic cleavage take place in CaM and affect the function of CaM (Murtaugh *et al.*, 1983).

CaM has four EF-hand motifs (two in each domain) that are site for calcium binding and change conformation upon binding calcium ions (Yap *et al.*, 1999). EF motifs provide suitable electronegative environment for Ca^{2+} ion coordination. EF-hand motif consists of two alpha helices that are connected by a 12-residue loop. In *apo* CaM (Ca^{2+} free), the alpha helices in the EF-hand motifs are positioned almost parallel to each other. This is referred as the closed conformation of CaM. When the intracellular calcium level rises to 10^{-5} M, four Ca^{2+} ions bind to CaM, one in each EF motif. Ca^{2+} induces a large conformational change in CaM molecule. In Ca^{2+}-CaM, the alpha helices of the EF-hand motifs change their position relative to each other, forming an almost perpendicular conformation (Figure 6.8). This form is refereed as open conformation of CaM. The crystal structure of Ca^{2+}-CaM exhibits a dumbbell shape (Babu *et al.*, 1988, Babu *et al.*, 1985). This conformational change exposes hydrophobic methyl groups from methionine residue in CaM that facilitates Ca^{2+}-CaM to increase its binding affinity for a number of Basic Amphiphilic Helices (BAA helices) on target proteins (Dash *et al.*, 1997). It has been shown that upon binding to its target protein or peptides, CaM "wrap around" its target peptide and form a compact globular conformation by bending its central helix. The central helix of CaM is highly flexible and is key to its ability to bind a wide range of targets (Chou *et al.*, 2001).

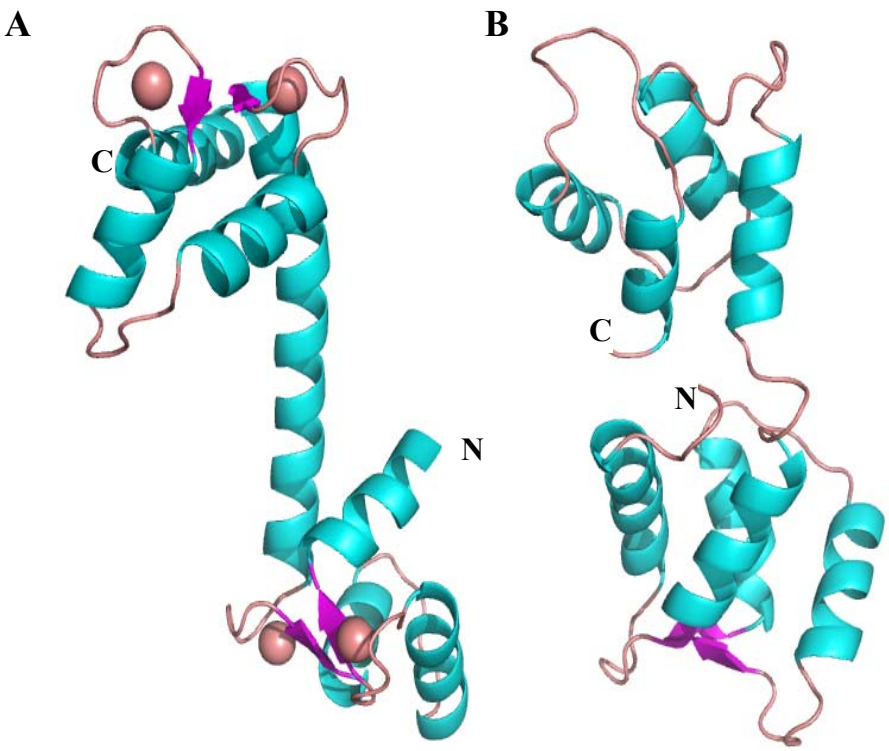


Figure 6.8: (A) Ca^{2+}-CaM structure (PDB code 3CLN) and **(B)** *apo* CaM structure (PDB code 1CFC). Figure generated using PyMol (DeLano & Lam, 2005).

6.4.1 CaM-binding proteins

CaM-binding proteins can be divided into two classes-1) calcium-dependent and 2) calcium-independent modes of binding. Ca^{2+}-CaM can bind to target proteins to alter their function, acting as part of a calcium signal transduction pathway. Ca^{2+}-CaM binding proteins include kinases, phosphatases, second messenger signalling proteins, cytoskeletal and muscle proteins. CaM binding proteins can have one or more CaM-binding domains. Key interaction occurs through bulky hydrophobic residues such as Phe, Trp, Ile, Leu, or Val. C-terminal domain of CaM is more negatively charged than the N-terminal domain. Peptides with +ve charged (H, K, and R) interacts with – ve charged residue on CaM. C-terminal domain of CaM was able to accommodate larger volume than the N-terminal domain (Osawa *et al.*, 1999). Figure 6.9 shows some CaM binding protein and their sub cellular localization.

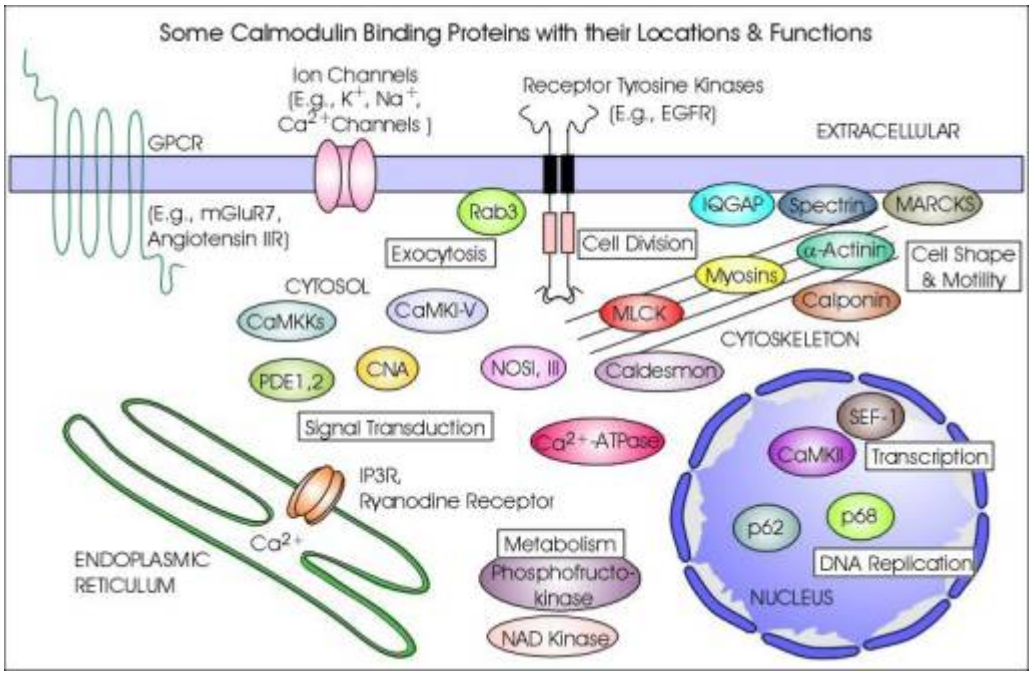


Figure 6.9: Example of CaM binding proteins. This figure is adapted from (O'Day, 2003)

CaM binding motifs have been divided into following category based on the distance of key bulky hydrophobic residue-

1-14 motif:

In this class, key bulky residues are separated by12 residues. Examples are -smooth muscle and skeletal muscle myosin light chain kinase (smMLCK, skMLCK).

1-10 motif:

The key bulky residues are separated by 8 residues. For example- Ca^{2+}/CaM dependent kinase II (CaMKII).

1-16 motif:

Here the key bulky residues are separated by 14 residues. For Example- Ca^{2+}/CaM dependent protein kinase kinase (CaMKK).

IQ motif:

IQ motif is the group of sequences on the target peptide through which it interacts with CaM mostly in Ca^{2+} dependent manner. A general sequence of IQ motif is given below

IQ (I/L/V)QXXXRXXXX(R/K)

IQ motifs have been first identified in Nm (Alexander *et al.*, 1988a) but was first characterized in myosins (Zuhlke *et al.*, 1999, Cheney & Mooseker, 1992) and calcium channels, such as the L-type calcium channel.

6.4.2 *apo* CaM-binding proteins

apo CaM utilizes unique binding motifs, which show greater homology than Ca^{2+}-CaM binding motifs. *apo* CaM binding proteins are neuroproteins, receptors, second messenger signalling proteins, cytoskeletal and muscle proteins. *apo* CaM and target protein interaction are shown to affect neurotransmitter production, nerve growth, synaptic development, intracellular movement and smooth muscle relaxation.

In summary, Neuromodulin (Nm) and Neurogranin (Ng) are the two neuron specific proteins expressed at pre and post synaptic terminal, respectively for the synaptic dialogue of efficacy changes underlying in learning and memory. Both proteins bind to CaM through IQ motif (I/L/V)QXXXRXXXX(R/K) under a low Ca^{2+} micro-environment and dissociates when Ca^{2+} level increases (Andreasen *et al.*, 1983). In resting neurons, much of the CaM is associated with these two proteins and also with

regulator of CaM signalling (RCS). Ca^{2+}-dependent phosphorylation of Nm (Ser41) and Ng (Ser36) by PKC also prevents its binding to *apo* CaM (Caroni, 2001) (Huang *et al*, 1993). Ser36-phosphorylated Ng is unable to bind either CaM or phospahtidic acid (Diez-Guerra, 2010). Thus Nm and Ng play the similar role at the pre and post synaptic termini respectively by controlling the concentration of *apo* CaM in neurons. Nm and Ng share high sequence similarity within a 20-amino-acid region that contains an IQ motif. Both are acid stable, hydrophilic proteins and belongs to the calpactin protein family (Ran *et al.*, 2003).

Our study shows that Nm and Ng are intrinsically unstructured proteins *in vitro* under physiological conditions. The unstructured nature of the two proteins probably better suited for its interaction with multiple proteins and to perform the various function as mentioned above. In order to gain the further insight into the molecular mechanism of learning and memory formation, in this chapter we report the structural and functional studies of these two proteins involved in important cognitive functions. This is the first time the structure of Nm/Ng and CaM complexes have been reported.

6.5 Objective of this study

The goal of the proposed study is to gain insight into the molecular mechanism of learning and memory formation through the structural and functional studies of these two (Nm and Ng) proteins involved in important cognitive functions through its interactions with CaM in neuron cell. Even though these proteins have been known for many years, so far there is no representative structure available for this family of neuron specific substrate proteins. Structure-function relationships that will be established from the studies of these neuron specific substrate proteins will help to

understand their implications in the functional regulation of the nervous system. The objective of this study is to understand the structure and function of these two proteins through the following experimental approaches:-

1. Cloning of various genes constructs to over express the recombinant proteins.
2. Biophysical characterization of Nm and Ng by using different techniques to understand the *in vitro* behaviour of these proteins.
3. Thermodynamic characterization of interaction with full length Nm and Ng and their IQ peptides with *apo* and Ca^{2+}/CaM using isothermal calorimetry.
4. Determination of the complex crystal structures' of IQ motif peptides derived from Nm and Ng with *apo* CaM and Ca^{2+}/CaM.
5. Structure function analysis to understand the role of these proteins in the molecular mechanism of learning and memory formation.

6.6 Methods

6.6.1 Expression, purification and characterization of Nm and Ng

Full-length Nm (accession no NP_032109) and Ng (accession no NP_071312) proteins were expressed from *E. coli BL21 (DE3)* cells harboring respective genes in pQE30 (Qiagen, USA) plasmids. *E. coli* cells were cultured in 1L LB media (supplemented with 100 µg/mL ampicillin) at 37°C until the OD_{600} reached between 0.6-0.8 AU. The culture was then maintained at 16°C before protein expression was induced with 0.15 mM IPTG. Cells were grown for 16 h at 16°C, and harvested by centrifugation at 9800 g for 10 min. Cell pellets from 2 L cell culture were resuspended in 100 ml lysis buffer (50 mM Tris-HCl (pH-8.0), 500 mM NaCl, 10% v/v glycerol, 20 mM Imidazole, 20 mM BME and 0.1 mM PMSF). After sonication,

the cell lysate was centrifuged at 39,000 g for 30 min. The supernatant was mixed with 5 ml of Ni–NTA (Qiagen, USA), and washed thrice using 30 ml lysis buffer. The protein was released from the Ni-NTA resin using 10 ml lysis buffer supplemented with 300 mM imidazole. The protein was then passed through a HiLoad™ 16/60 Superdex™ 200 prep grade gel filtration column and eluted in a buffer (50 mM $NaH_2PO_4.2H_2O$ (pH 7.4), 300 mM NaCl, 20 mM BME, and 0.1% Triton X100). The mean stokes radius and apparent molecular weight of Nm and Ng were calculated using standard curves (Figure 10B and 10C) generated based on the elution volumes of proteins of known molecular weight and stokes radius(le Maire *et al.*, 1987). Dynamic Light Scattering (Protein solutions, USA) and Circular Dichroism (JASCO J-175, Japan) experiments were performed in the same buffer. ^1H NMR experiments were performed in 50 mM sodium phosphate (pH 7.4), 100 mM NaCl and 95% H_2O plus 5% D_2O at 25^0 C with 16 K data point. Suppression of the water resonance was achieved through the WATERGATE technique.

Based on the complex structures, an arginine to alanine point mutation of full-length Nm/Ng (Arg43Ala/Arg38Ala) was generated using inverse PCR and purified as described above.

6.6.2 Cloning, expression and purification of CaM constructs

NmIQ2 (34-57) and NgIQ2 (27-50) motifs were linked to the C-terminus of CaM *via* a 5-glycine flexible linker (CaM-$(Gly)_5$-NmIQ2 and CaM-$(Gly)_5$-NgIQ2) using a three-step phusion PCR procedure, as described by Ye *et al* (Ye *et al.*, 2006). The final PCR product was digested with Nde1 and Xho1 restriction enzymes (New England Biolabs, UK) along with the pGS21a vector (GeneScript, USA). Predigested

CaM-$(Gly)_5$-NmIQ2 and CaM-$(Gly)_5$-NgIQ2 genes and pGS21a vector were ligated, transformed into chemically competent *E. coli DH5α* cells and screened for positive colony formation. The CaM-$(Gly)_5$-NmIQ2 and CaM-$(Gly)_5$-NgIQ2 gene sequences were verified by DNA sequencing.

For protein expression, the recombinant plasmids (pGS21a- CaM-$(Gly)_5$-NmIQ2 and pGS21a- CaM-$(Gly)_5$-NgIQ2) were transformed into *E. coli BL21 (DE3)* chemically competent cell and plated onto agar plates. A single colony was used to inoculate 100 ml of the LB media containing 100 µg/ml ampicillin. The expression and purification strategy was similar to that of full-length Nm/Ng described above, with the exception that the proteins were passed through HiLoad™ 16/60 Superdex™ 75 prep grade columns (Amersham Biosciences, Sweden) and eluted with a buffer consisting of 20 mM Imidazole pH 8.0, 100 mM NaCl, 2 mM EGTA (buffer A).

A similar expression and purification procedure was adopted for $(His)_6$-tagged CaM (accession no NP_033920) from pETDuet-1 vector (Novagen, USA). CaM protein was eluted from gel filtration column in either buffer A or a buffer containing 20 mM HEPES (pH 8.0), 100 mM NaCl and 10 mM $CaCl_2$ (buffer B) to elute *apo* CaM or Ca^{2+}/CaM, respectively. For SeMet SAD phasing, selenomethionine-labeled proteins were produced using LeMaster media(Hendrickson *et al.*, 1990) following a similar procedure as described above. All protein purification steps were carried out at 4°C unless otherwise indicated.

6.6.3 Isothermal titration calorimetry

Isothermal titration calorimetry was used to study the binding of full-length wild type Nm/Ng proteins, full-length mutant Nm/Ng proteins and their synthetic wild-type IQ motif peptides (Nm (NmIQ1: aa32-50 and NmIQ2: aa34-57) and Ng

(NgIQ1: aa27-45 and NgIQ2: aa27-50)) (Figure 13-15 and Figure 18) with Ca^{2+}/CaM (in a buffer B) and *apo* CaM (in a buffer A). ITC experiments were performed using either VP-ITC or iTC200 calorimeter (indicated in brackets) (Microcal, LLC) at room temperature (24°C) with 0.2 (or 0.04) ml of Ca^{2+}/CaM or *apo* CaM in the injector cell, and 1.8 (or 0.2) ml of Nm/NmIQ or Ng/NgIQ in the sample cell, respectively. Samples were thoroughly degassed and centrifuged to remove precipitates. All experiments used 10 (or 2) µl volumes per injection. Two consecutive injections were separated by 5 (or 2) min to allow the peak to return to baseline. Data from control experiments, i.e., titration of protein/peptides into buffer, were subtracted from each experiment to compensate for the heat of dilution. ITC data were analyzed with a single-site fitting model using the following equation

$$Q_i^{tot} = V_o E_{tot}\left(\left(K_1 P\right)\Delta H_1 / \left(1 + K_1 P\right)\right)$$

Where Q_i^{tot} is the total heat after the *i*th injection, V_0 is the volume of calorimetric cell, K_1 is the observed equilibrium constants for each site, P is the concentration of free ligand, and ΔH_1 is enthalpy changes. The corresponding of K_{obs} and ΔH were obtained by fitting the experimental data to model using Origin 7.0 software (OriginLab Corp. MA, USA).

6.6.4 Crystallization and structure determination

Crystallization trails for *apo* CaM-$(Gly)_5$-NmIQ2, *apo* CaM-$(Gly)_5$-NgIQ2 and complexes of Ca^{2+}/CaM with NmIQ2 and NgIQ2 synthetic peptides (1:3 molar ratios) were performed with a protein concentration of 10-12 mg/ml using the hanging drop vapor diffusion method at room temperature. The initially identified conditions from Hampton Research (Aliso Viejo, CA, USA) and Qiagen (Valencia, CA, USA) were further optimized. Best crystals of Ca^{2+}/CaM-NmIQ2 and Ca^{2+}/CaM-

NgIQ2 complexes were obtained using a reservoir solution consisting of 100 mM Tris-HCl (pH 8.0), 8-10% v/v PEG 6 K, 5-10 mM $ZnCl_2$ and 10% v/v glycerol. *apo* CaM-$(Gly)_5$-NmIQ2 complex crystals were obtained from a condition consisting of 0.12 mM MgAcO, 8% PEG 3350 and 10% EtOH. Similarly, the *apo* CaM-$(Gly)_5$-NgIQ2 complex crystals were obtained using a condition consisting of 0.1 M imidazole (pH 8.0) and 1.2 M sodium citrate tribasic dihydrate. Where necessary, crystals were cryo-protected with reservoir condition supplemented with 10% glycerol and flash-cooled in N_2 cold stream at 100 K.

The molecular replacement method did not yield any structure solution, which led us to collect the Single wavelength Anomalous Dispersion (SAD) data sets in the synchrotron beam line X8C (NSLS, Brookhaven National Laboratory) and 13B1 SW6 (National Synchrotron Radiation Research Center (NSRRC), Taiwan) using a Quantum 4-CCD detector (Area Detector Systems Corp Poway, CA, USA). All data sets were processed using HKL2000(Otwinowski & Minor, 1997). Heavy atom (Se) location, phasing and density modification were performed using the program ShelxC/D/E (Sheldrick, 2008), and model building was carried out with the program Buccaneer (Cowtan, 2006) in CCP4 . Where necessary, the model was manually built in COOT(Emsley & Cowtan, 2004) and refinement was carried out in Refmac5(Vagin *et al.*, 2004). At the final stage of refinement, well-ordered water molecules were included. The models have good stereochemistry, as analyzed by PROCHECK(Laskowski *et al.*, 1993) (Table 2). All structure-related figures reported in this manuscript were generated using PyMol(DeLano & Lam, 2005).

6.6.5 Protein Data Bank accession code

Coordinates and structure factors of all the *apo* CaM-$(Gly)_5$-NmIQ2 and *apo* CaM-$(Gly)_5$-NgIQ2 has been deposited with RCSB Protein Data Bank with codes 4E53 and 4E50.

6.7 Results

6.7.1 Nm and Ng are intrinsically unstructured proteins

6.7.1.1 *Sequence analysis predicts Nm and Ng are intrinsically unstructured proteins*

Compared with the average folded protein (Romero *et al.*, 2001), the signature for proteins with a probable lack of folding is a low sequence complexity and an amino acid composition with a low proportion of order-promoting amino acids (W, Y, F, C, I, L and N) and a high proportion of disorder-promoting amino acids (P, E, K, S and Q). Consequently, intrinsically unstructured proteins have high net charge at neutral pH and a low overall hydrophobicity. The amino acid compositions of Nm and Ng are unique. Nm consists of 4.4% order-promoting amino acids, and 45.8% disorder-promoting amino acids. Similarly, Ng consists of 15.4% order-promoting and 27% disorder-promoting amino acids. Thus, the amino acid composition of both proteins suggests that their structures are intrinsically unstructured (Table 6.1). Further the acidic pI and low hydrophobicity of these two proteins indicates that they are unfolded or unstable. Moreover Nm/Ng migrates slowly in SDS-PAGE, consistent with previous findings (Baudier *et al.*, 1989).

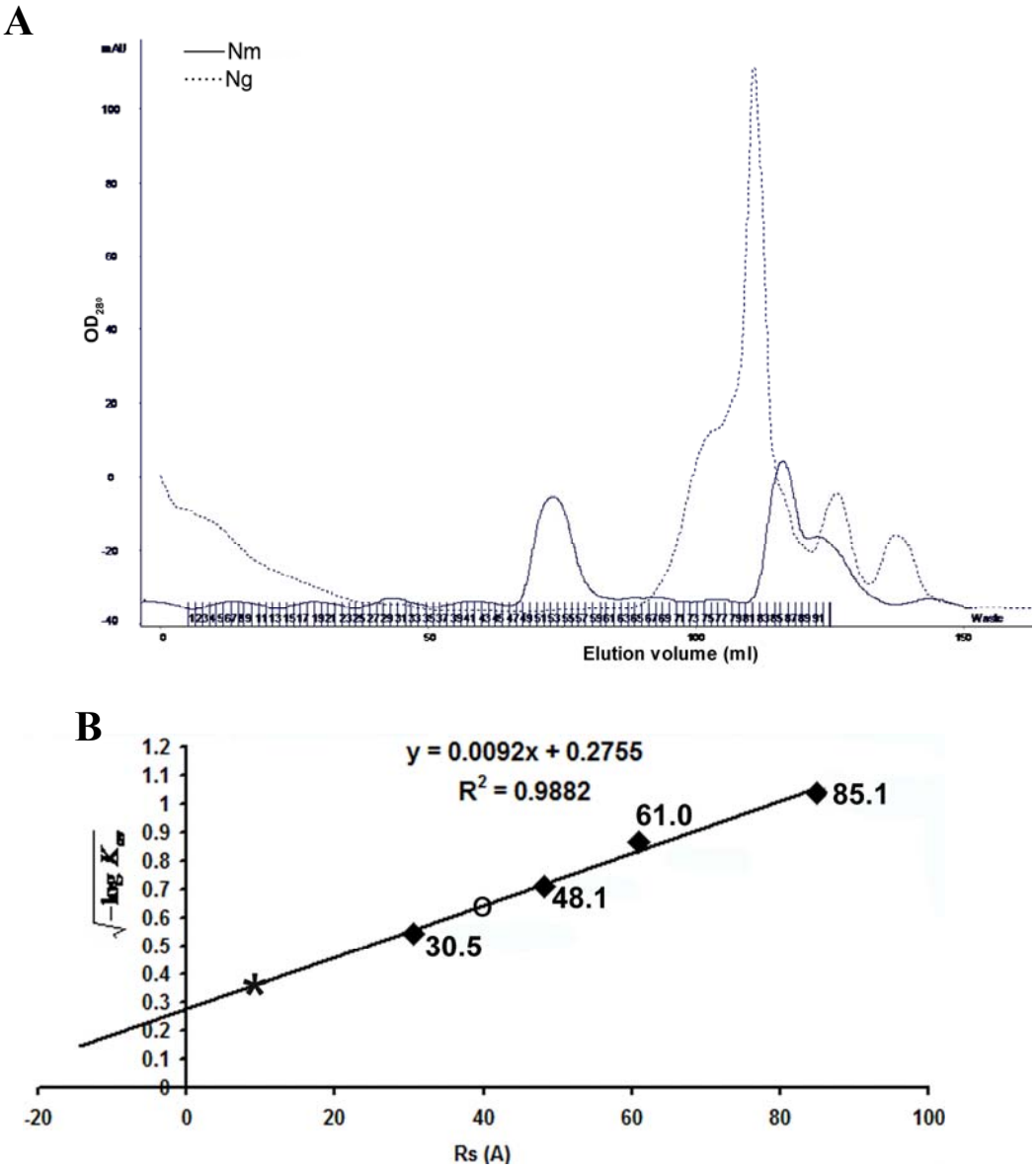
Table 6.1: The amino acid composition of the Nm and Ng. The order promoting amino acids (*) (W, Y, F, C, I, L and N) and disorder promoting amino acids ($) (P, E, K, S and Q) are indicated.

	Nm		Ng	
	No of amino acids	**Composition (%)**	**No of amino acids**	**Composition (%)**
Ala (A)	46	20.3	11	14.1
Arg (R)	5	2.2	4	5.1
Asn (N)*	2	0.9	1	1.3
Asp (D)	22	9.7	8	10.3
Cys (C)*	2	0.9	4	5.1
Gln (Q)$	9	4.0	1	1.3
Glu (E)$	34	15.0	2	2.6
Gly (G)	13	5.7	18	23.1
His (H)	3	1.3	1	1.3
Ile (I)*	3	1.3	4	5.1
Leu (L)*	2	0.9	2	2.6
Lys (K)$	30	13.2	6	7.7
Met (M)	2	0.9	2	2.6
Phe (F)*	1	0.4	1	1.3
Pro (P)$	17	7.5	7	9.0
Ser (S)$	14	6.2	5	6.4
Thr (T)	15	6.6	1	1.3
Trp (W)*	0	0.0	0	0.0
Tyr (Y)*	0	0.0	0	0.0
Val (V)	7	3.1	0	0.0

6.7.1.2 *Gel Filtration shows Nm and Ng exist in unfolded globular state*

The elution profiles of Nm and Ng from Superdex 200 gel filtration columns revealed an elution time shorter than that which would be expected for a monomer of a similar sized globular protein. The symmetry of the elution peaks indicates that Nm and Ng both eluted as a homogeneous species (Figure 6.10). The calculated mean stokes radius (R_S) of 39 Å for Nm and 10 Å for Ng corresponds to a molecular weight of ~100 kDa and ~22 kDa, respectively. Thus, the estimated molecular weight from gel filtration chromatography is around 3-4 times higher than that for the expected monomeric MW of 23.6 kDa (Nm) and 7.6 kDa (Ng). The elution volume for the

denatured (6 M Guanidinium HCl-treated) Nm and Ng corresponds to an R_S of 46 Å and 14 Å, respectively. No significant differences were observed in the gel filtration profile upon treating these proteins with the denaturing agent. Further, the Dynamic Light Scattering (DLS) experiments for these two proteins showed higher molecular weights and were highly monodispersed in solution. Together, these results suggest that Nm and Ng exist in a largely unfolded premolten globular state.



C

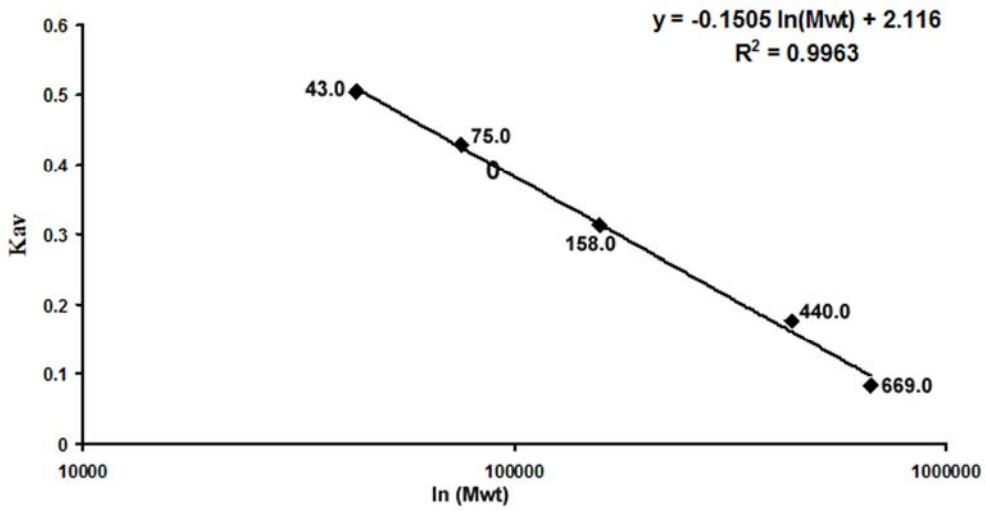
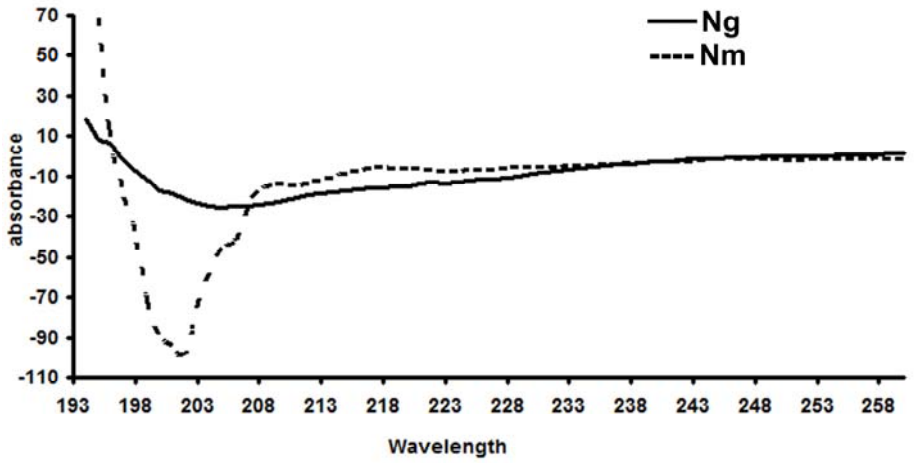


Figure 6.10: (A) Hydrodynamic analyses of Nm (solid line, 73 ml) and Ng (dotted line, 103 ml) monitored at 280 nm in Superdex 200 Gel filtration column. The partition coefficient, K_{av}, was calculated using the expression $K_{av} = \frac{V_e - V_0}{V_c - V_0}$, where, V_e = elution volume, V_0 = void volume (45 ml as calculated using Blue dextran 2000) and V_c = total column volume (120 ml). Analytical gel filtration chromatography standard curves for (B) K_{av} versus Stokes radius (Rs); and (C) K_{av} versus molecular weight (MW) of standard monomeric MW markers (High Molecular Weight Gel Filtration Calibration Kit, Amersham Biosciences, Sweden). The Stokes radius and hydrodynamic molecular weight was calculated using the equations shown. Open ellipsoid and asterisk corresponds to Nm and Ng, respectively. Because of the very high K_{av} value (0.773), Ng is not shown in graph C.

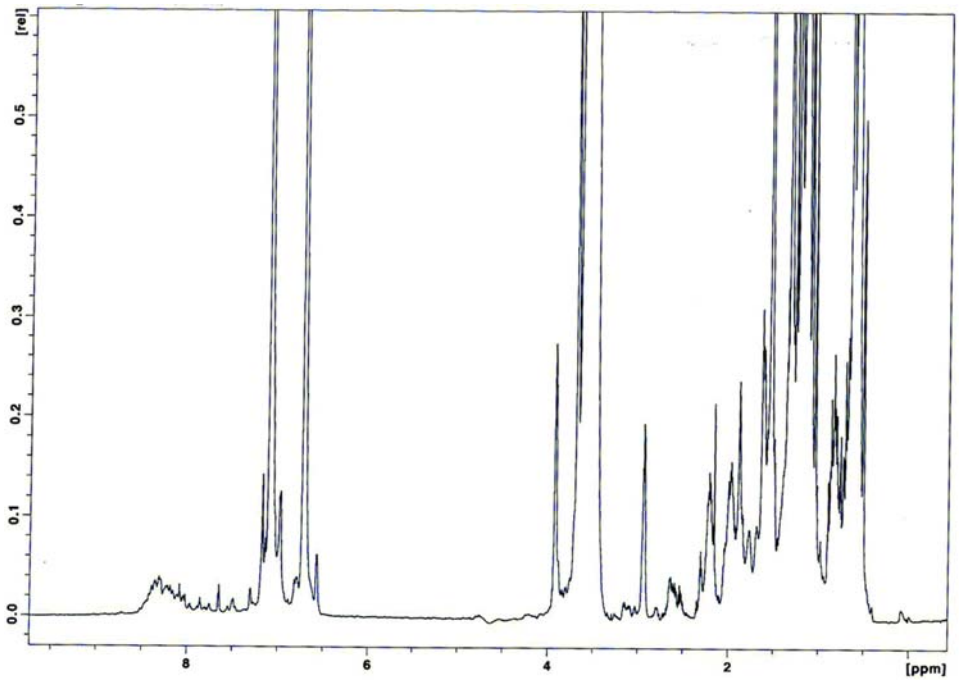
6.7.1.3 *Residual Secondary structure from Far UV-CD*

CD spectra of Nm and Ng showed an intense minimum at 201 and 204 nm respectively, indicating the presence of disordered regions in these proteins. The CD analysis further supports that Nm and Ng both have residual secondary structural elements (mainly α helix) and suggests that these two proteins exist in premolten globular conformational state of protein quartet model (Uversky, 2002) under physiological conditions (Figure 6.11A) , supporting the results from the gel filtration experiments.

A


Ng
Nm
absorbance
70
50
30
10
-10
-30
-50
-70
-90
-110
193
198
203
208
213
218
223
228
233
238
243
248
253
258
Wavelength

B


[rel]
0.5
0.4
0.3
0.2
0.1
0.0
8
6
4
2
[ppm]

C

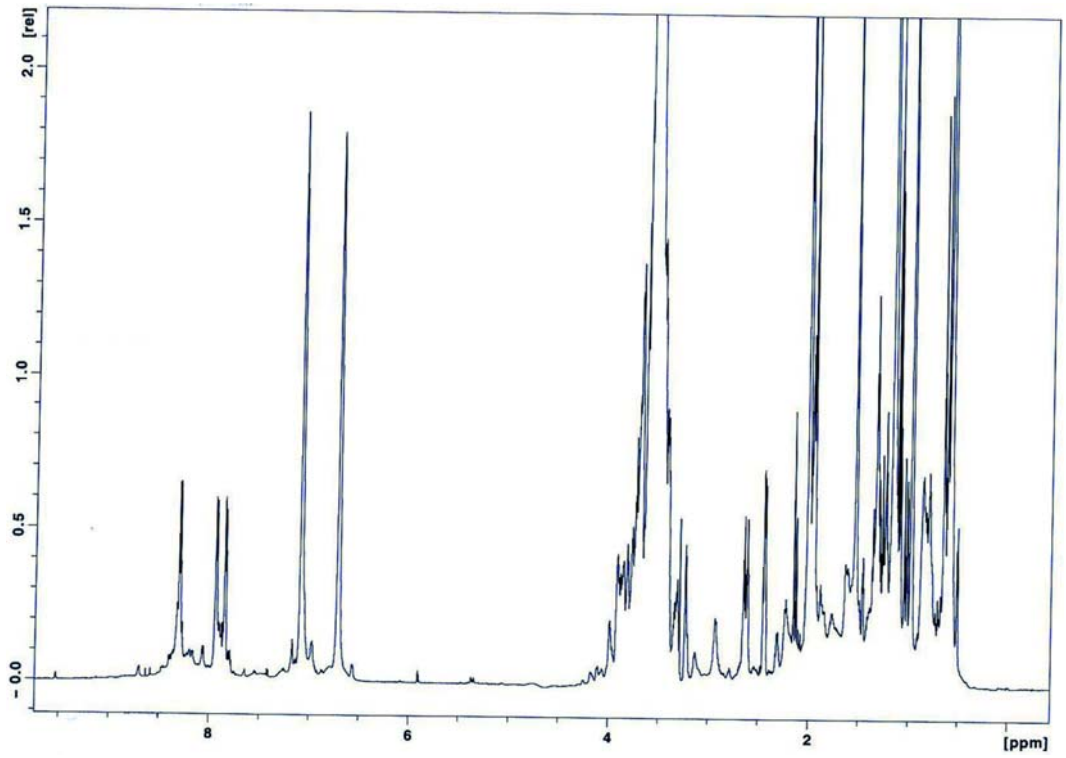


Figure 6.11: (A) Far-UV CD spectra of the purified native Nm (dotted line) and Ng (solid line). Data from three independent scans were averaged and the background spectrum of the buffer was subtracted. ^{1}H NMR spectrum of **(B)** Nm and **(C)** Ng. Far-UV CD and ^{1}H NMR spectrum suggest unfolded nature of these proteins in solution.

6.7.1.4 *NMR spectroscopy suggests Nm and Ng are natively unfolded proteins*

^{1}H NMR is sensitive method capable of distinguishing between folded and unfolded proteins through a small dispersion of the amide backbone chemical shift(Rehm *et al.*, 2002). Particularly, the appearance of a large and broad signal at approximately 8.3 ppm is an indicator for a disordered protein. On the other hand, signal dispersion beyond 8.5 ppm (8.5-11 ppm) indicates a folded protein. Further, in the aliphatic region of the ^{1}H NMR spectrum between +1.0 and -1.0 ppm, a large-signal dispersion versus a steep flank of the dominant peak at approximately 1.0 ppm will separate a structured protein from an unfolded protein. The NMR spectra of Nm and Ng were very similar to those previously observed for other unstructured proteins (Figure 6.11B and C). The resonance of both protein's methyl group protons (at 0.9

and 0.95 ppm) and amide groups (around 8.4 ppm); along with very limited spectral dispersion of these signals, indicate the lack of a stable tertiary structure in both Nm and Ng (Figure 6.11B and C).

Thus, the biophysical studies on Nm and Ng support the classification of these proteins as intrinsically unstructured, lacking a compact globular fold, and having very little or no secondary structure (Permyakov *et al.*, 2003, Kleerekoper & Putkey, 2009, De Guzman *et al.*, 2004).

6.7.2 Isothermal calorimetry

The sequence alignment of known CaM binding peptides showed that each has one or two key aromatic residues, a few basic amino acids (located at N or C terminal) and other hydrophobic amino acids (Figure 6.12). Nm and Ng protein sequences do not have any aromatic residues at the N-terminal region of the IQ motif. However, both IQ motifs are enriched with basic amino acids in the C-terminal region. Based on the sequence alignment, we selected two different lengths of peptides surrounding the IQ motif for Nm (NmIQ1 and NmIQ2) and Ng (NgIQ1 and NgIQ2) in order to determine which length better mimics the interactions of their full-length counterparts with CaM. The ITC experiments were performed with full length and IQ peptides of Nm/Ng with CaM (Table 6.2 and Figure 6.13-6.15). Table 6.2 report the preliminary results of the ITC experiments. Currently we are in the process of analyzing all ITC data that we have obtained. Based on the analysis, if needed additional ITC experiments will be performed.

ITC experiments revealed that the full-length Nm and Ng bind to *apo* CaM (in the absence of Ca^{2+}) with a higher affinity than CaM in the presence of Ca^{2+}. These observations are consistent with the previous fluorescence experiments (Andreasen *et*

al., 1983) (Table 6.2). Further, the ITC results show that NmIQ2 and NgIQ2 peptides have a similar affinity toward *apo* CaM as their full-length counterparts; thus, the IQ2 peptides almost mimic their full-length proteins. Nm and Ng interact with CaM mainly through the IQ motif region with a stoichiometric ratio of 1:1.

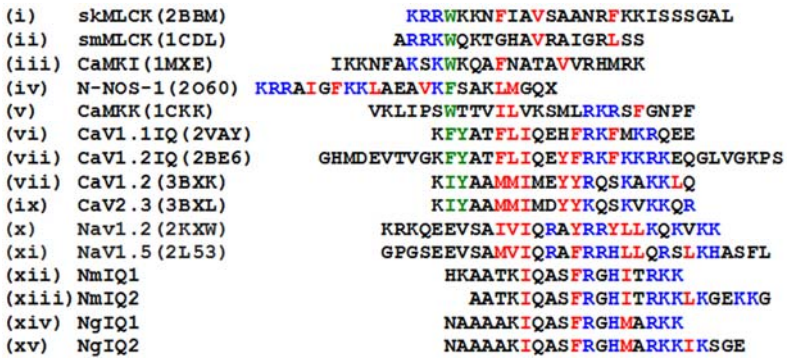


Figure 6.12: Structure-based sequence alignment of CaM binding peptides from different proteins. Sequence analysis shows a critical aromatic amino acid that anchors the peptide into CaM; a positively charged amino acid (blue) that determines the orientation of the peptide between two lobes of CaM; and hydrophobic amino acids (red) that are involved in binding. The pdb codes are given in parenthesis.

Table 6.2: Thermodynamic parameters for Nm and Ng with CaM interactions

	Nm-*apo* CaM				Nm-Ca^{2+}/CaM			
	Nm	**NmIQ1**	**NmIQ2**	**Nm(R43A)**	**Nm**	**NmIQ1**	**NmIQ2**	**Nm(R43A)**
$K_A X\ 10^6\ (M^{-1})$	3.2 (± 0.4)	0.02(± 0.09)	0.56(± 0.042)	0.077(± 0.018)	1.95(± 0.68)	0.0096(± 0.0024)	0.039(± 0.0039)	0.077(± 0.009)
N	0.97 (±0.02)	1.12(±0.19)	0.95(±0.00)	1.04(±0.2)	0.99.(±0.04)	0.84(±0.10)	0.93(±0.02)	1.06(±0.17)
ΔH (kcal/mol)	-163.9(±4.9)	-0.87(±0.15)	-230.2(±8.5)	2.49(±0.7)	-61 (±3.9)	-2.08(±0.35)	-14.6(±0.38)	-1.514 (±0.3)
ΔS (cal/mol/deg)	-520	15.36	-745.8	30.6	-176	-11.4	-27	17.28
	Ng-*apo* CaM				**Ng-Ca^{2+}/CaM**			
	Ng	**NgIQ1**	**NgIQ2**	**Ng(R38A)**	**Ng**	**NgIQ1**	**NgIQ2**	**Ng(R38A)**
$K_A X\ 10^6\ (M^{-1})$*	13.0(± 2.1)	0.043(± 0.0040)	0.17(± 0.04)	0.2(± 0.04)	1.01(± 0.24)	0.0093(± 0.003)	0.061(± 0.007)	0.063(± 0.03)
N	0.96(±0.01)	1.11(±0.013)	0.94 (±0.03)	1.00(±0.07)	0.96(±0.03)	1.24(±0.16)	1.07(±0.018)	1.05(±0.07)
ΔH (kcal/mol)	-28.3(±0.5)	-6.5(±0.10)	-5.8 (±0.34)	1.07(±0.12)	-70.1(±3.9)	0.96(±0.24)	-15.4(±0.3)	-8.814(±0.80)
ΔS (cal/mol/deg)	-62.5	-0.17	4.77	27.86	-207.4	16.8	-28.8	-7.57

Ka- Association constant
N- Number of binding sites
ΔH- Change in enthalpy
ΔS- Change in entropy

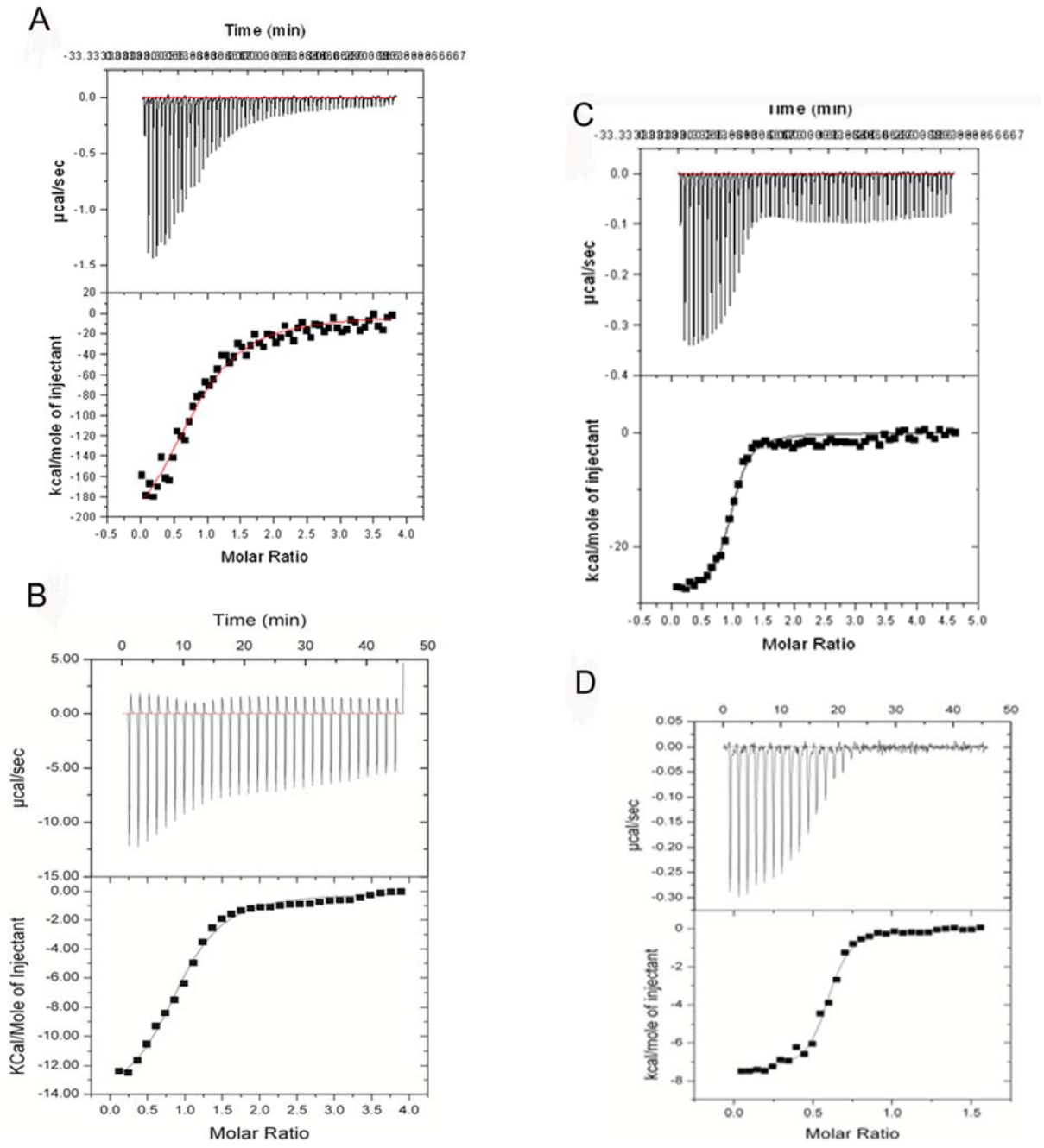


Figure 6.13: ITC profiles of **(A)** Nm **(B)** NmIQ2 **(C)** Ng and **(D)** NgIQ2 peptides titrated against *apo* CaM. Upper panels show the raw data; lower panels show the heat release for each injection. The data is fitted using a single-site binding model.

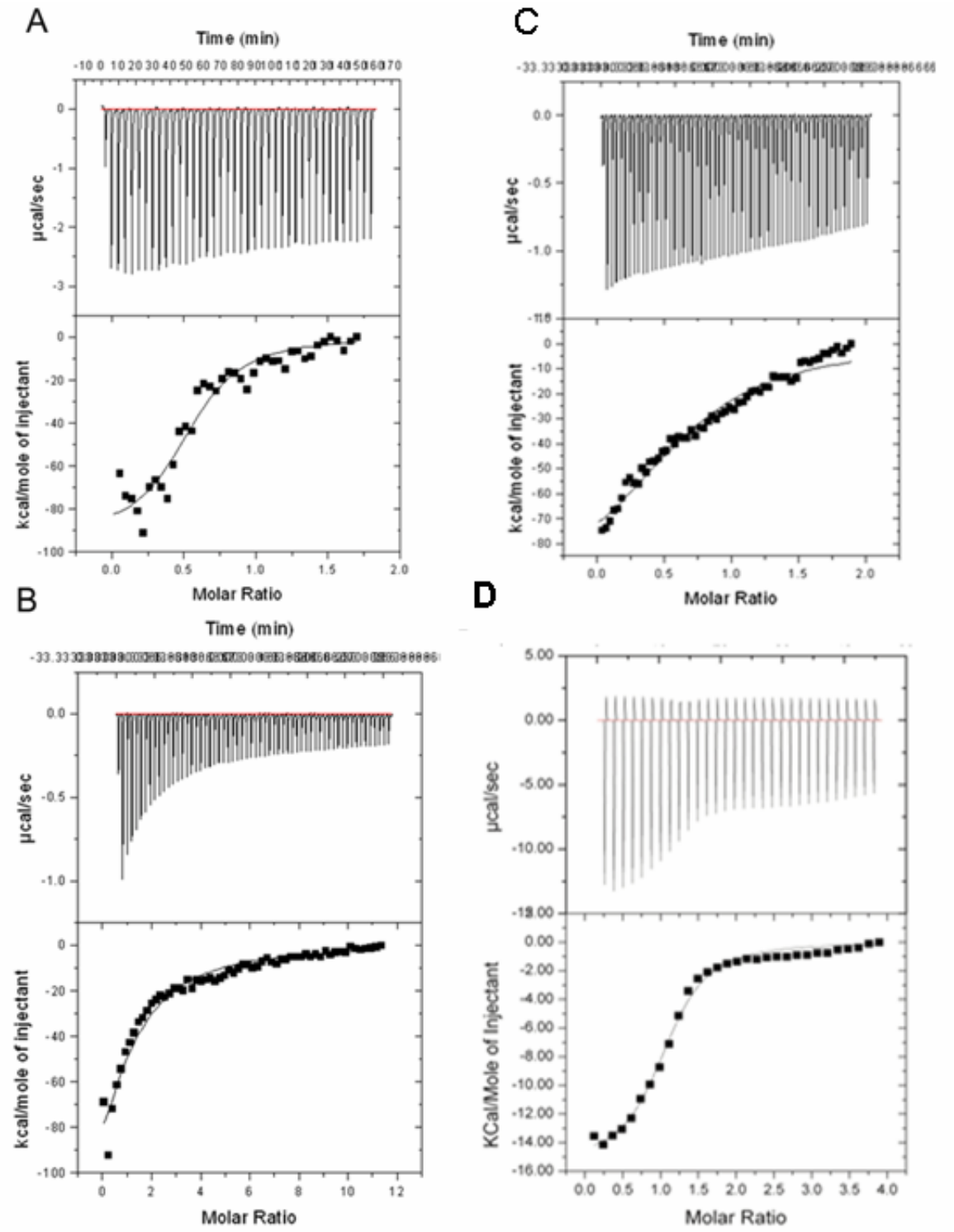


Figure 6.14: ITC profiles of Nm/Ng and NmIQ2/NgIQ2 peptides titrated against Ca^{2+}/CaM. **(A)** Nm titrated against Ca^{2+}/CaM; **(B)** NmIQ2 peptide titrated against Ca^{2+}/CaM; **(C)** Ng titrated against Ca^{2+}/CaM; and **(D)** NgIQ2 peptide titrated against Ca^{2+}/CaM. Upper panels show the raw data; lower panels show the heat release for each injection. The data is fitted using a single-site binding model.

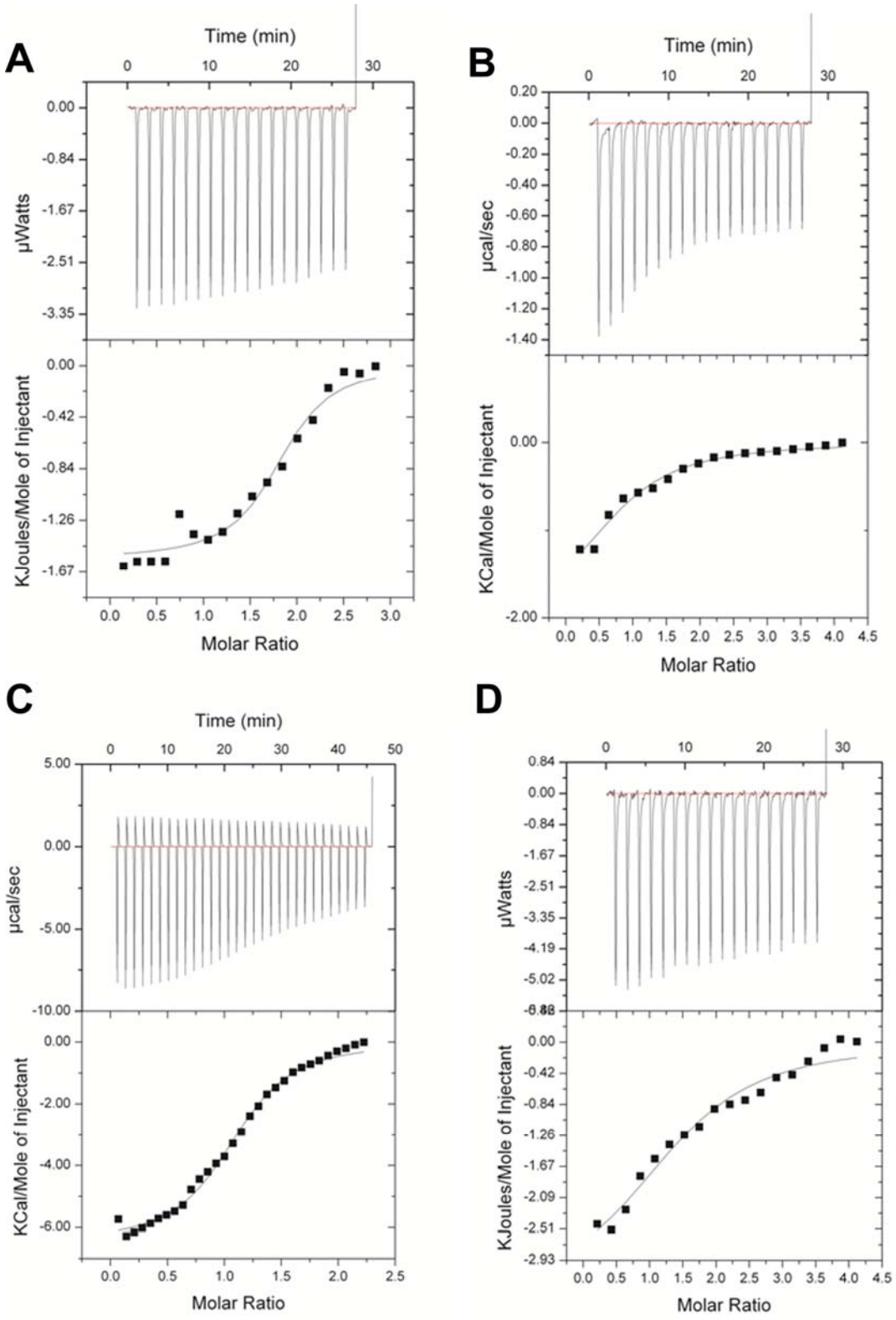


Figure 6.15: ITC profiles of NmIQ1 and NgIQ1 peptides titrated against CaM. **(A)** NmIQ1 titrated against Ca^{2+}/CaM; **(B)** NmIQ1 titration against *apo* CaM; **(C)** NgIQ1 titrated against Ca^{2+}/CaM; **(D)** NgIQ1 titrated against *apo* CaM. Upper panels show the raw data; lower panels show the heat release for each injection. The data is fitted using a single-site binding model.

6.7.3 Nm/Ng CaM complex structural studies

The present studies combined with literature clearly showed that Nm and Ng are intrinsically unstructured proteins. Brief attempts to crystallize the full-length Nm and Ng alone and in complex with *apo* and Ca^{2+}/CaM did not yield crystals. This led us to crystallize the IQ motif peptides in complex with CaM instead. The best peptides to better mimic the full-length proteins were selected. Initially, we co-crystallized the synthetic IQ motif peptides with *apo* CaM and Ca^{2+}/CaM. Crystals were obtained only for Ca^{2+}/CaM-NmIQ2 and Ca^{2+}/CaM-NgIQ2 complexes, with similar crystallization conditions. Although the presence of these peptides in their complex crystals was verified, there was no electron density observed for the peptides.

6.7.3.1 *Ca^{2+}/CaM-NmIQ2 and Ca^{2+}/CaM-NgIQ2 structures*

In both structures, the His-tag and the N-terminal 5 residues were disordered and not included in the model. The two molecules of the asymmetric unit were tightly packed. Ca^{2+} ions were well-defined in the electron density map, located at each EF-hand (helix-loop-helix) motif, while no electron density was observed for the IQ peptides. The structure of Ca^{2+}/CaM in the presence of Nm/Ng-IQ peptides was similar, with an rmsd of 0.7 Å for 139 Cα atoms. The following section discusses the high resolution structure of Ca^{2+}/CaM crystallized in the presence of NmIQ peptide (Table 6.3).

A search for structural homologs in the pdb database using DALI program (Holm & Sander, 1995) did not provide any similar structures. In the present structure, the Ca^{2+}/CaM adopt a unique conformation, with the relative disposition of its two lobes being completely different from any of the previously reported CaM

structures. However, a one-to-one comparison of individual domains showed no significant structural differences (rmsd less than 0.8Å for all Cα atoms).

The close examination of current Ca^{2+}/CaM structure with previously reported *apo* and Ca^{2+}-bound CaM complexes (including IQ motif complexes) showed that residues Ala74-Asp79 of the central helix (aa 65-92) were unwound, and bent by ~90° near Arg75, which reoriented the C-lobe in a perpendicular direction to the central helix (Figure 6.16). While a transformation of α-helix-to-loops has been previously reported (Shen *et al.*, 2004, Aoyagi *et al.*, 2003), the kink observed at Arg75 is unique, the current structure of CaM likely to be one among its many possible conformations.

It must be noted that the Ca^{2+}/CaM-NmIQ2 crystals were grown in conditions containing 5-10 mM $ZnCl_2$ and 10 mM $CaCl_2$. Zn^{2+} has been reported to bind CaM with higher affinity than Ca^{2+}. Based on the B-factor and structural comparison, we assigned the electron density in our structures as Ca^{2+} ions. Nonetheless, the previously reported Zn^{2+}-bound, N-lobe CaM structure resembles *apo* CaM structures, with no conformational changes observed in the central helix, despite the Zn^{2+} binding to both EF-hand motifs in N-lobe(Warren *et al.*, 2007). Thus, the present conformational change observed in the central helix of the CaM is independent of the bound metal ion.

Table 6.3: Crystallographic data and refinement statistics

	Ca^{2+}/CaM	*apo* CaM-$(Gly)_5$-NmIQ2		*apo* CaM-$(Gly)_5$-NgIQ2	
	Peak	**Peak**	**Native**	**Inflection**	**Peak**
Cell parameters (Å, °)	a=38.83, b= 116.40, c= 38.80, β= 94.83	a=56.30, b=56.40, c=135.80, β=90.76	a=79.38 b=79.28 c=136.06, β=90.19	a=76.57, b=76.57, c=46.99	a=76.46, b=76.46, c=46.95
Space group	$P2_1$	P2	C2	$P4_1$	$P4_1$
Data collection					
Resolution range (Å) *	50.00-2.0(2.05-2.00)	50.00-3.00 (3.11- 3.00)	50.00-2.65(2.74-2.65)	30-2.70 (2.80-2.70)	30.00-2.60 (2.69-2.60)
Wavelength (Å)	0.978	0.978	1.541	0.964	0.979
Observed reflections >1σ	165195	77387	105895	38257	157832
Unique reflections	45014(2341)	31242(2393)	23563(2086)	7588(742)	8503(779)
Completeness (%)	99.3 (98.9)	93.4 (71.6)	95.2(85.5)	100(99.6)	98.9(92.8)
Overall (I/ σ (I))	34.6(5.2)	15.8(5.5)	24.7(3.7)	44.6(2.7)	18.8(2.0)
R_{sym}[a] (%)	5.5(38.1)	9.5(18.3)	7.3(22.5)	7.2(42.8)	12.7(33.3)
Refinement and quality [b]					
Resolution range (Å)	30.0-2.0		30.0-2.69		30.00-2.70
R_{work}[c] (no. of reflections)	0.23(21619)		0.265 (21094)		0.24(7200)
R_{free}[d] (no. of reflections)	0.25(1178)		0.30(1278)		0.27(359)
RMSD bond lengths (Å)	0.01		0.007		0.012
RMSD bond angles(°)	1.045		1.12		1.43
Average B-factorse ($Å^2$)					
Main chain	41.7		56.8		61.6
Side chain	43.4		57.4		62.6
Ramachandran plot					
Most favored regions (%)	91.6		84.6		89.4
Additional allowed regions (%	7.6		14.8		9.2
Generously allowed regions (%)	0.8		0.7		0.7
Disallowed regions (%)	0.0		0.0		0.7

[a]$R_{sym} = \sum|I_i - \langle I \rangle|/|I_i|$ where I_i is the intensity of the ith measurement, and $\langle I \rangle$ is the mean intensity for that reflection.
[b]Reflections with I>σ was used in the refinement.
[c]$R_{work} = |F_{obs} - F_{calc}|/|F_{obs}|$ where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively.
[d]R_{free} = as for R_{work}, but for 5-7% of the total reflections chosen at random and omitted from refinement.
[e]Individual B-factor refinements were calculated.
*The high resolution bin details are in the parenthesis.

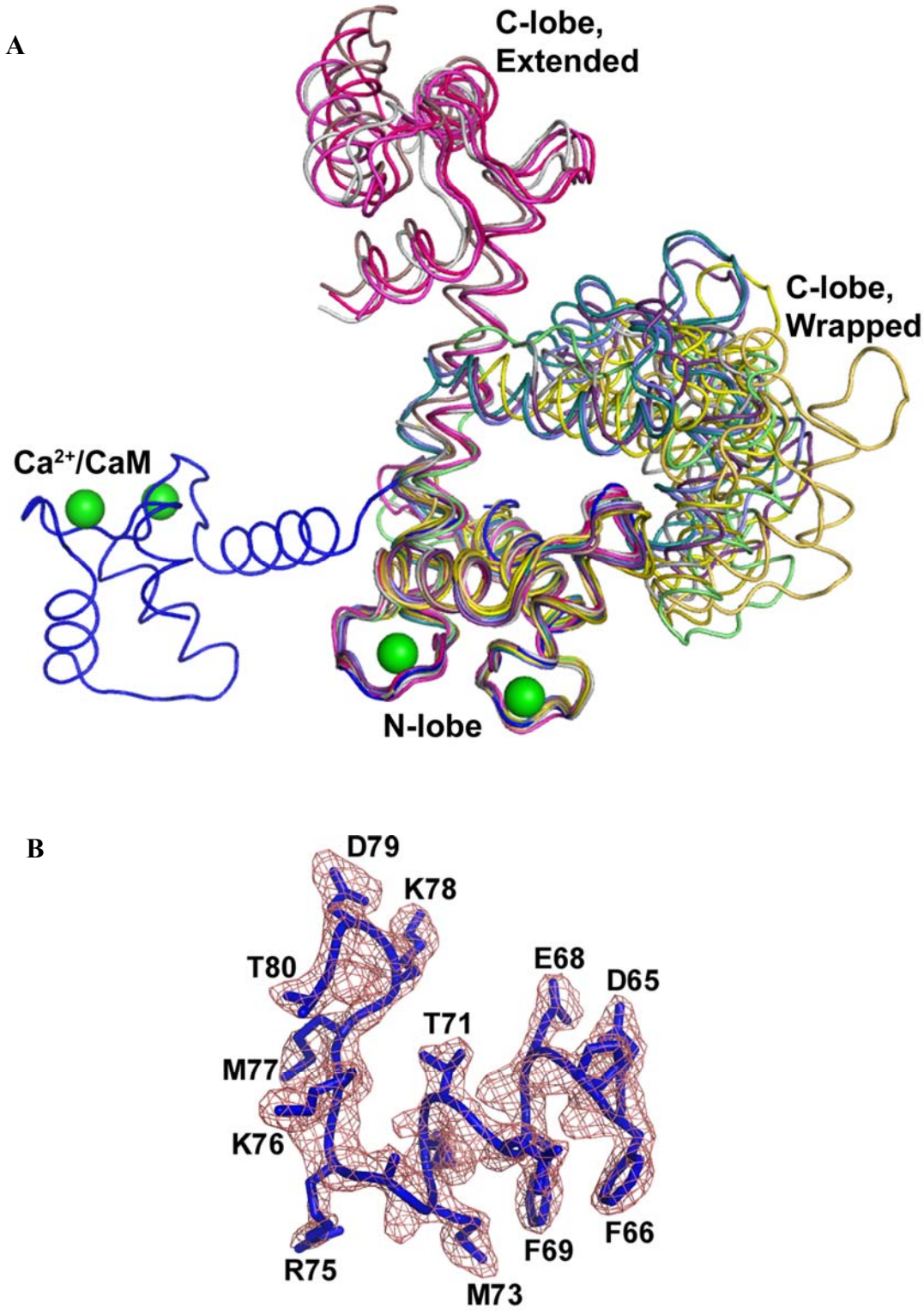


Fig. 6.16: (A) Cα superimposition of Ca^{2+}/CaM (blue) with *apo* CaM-$(Gly)_5$-NmIQ2 (magenta) and other structures of CaM from the pdb database 2BE6 (yellow), 2F2P (white), 2F3Y (light blue), 2O60 (pale green), 2VAY (teal), 2W73 (red), 2X0G (orange), 3BXQ (deep purple), 3DVE (gray), 1CDM (olive) and 3CLN (dark salmon). The results show that Ca^{2+}/CaM-NmIQ2 adopts a unique conformation. **(B)** *2Fo-Fc* electron density map for the fragment aa65-80 of Ca^{2+}/CaM-NmIQ2. This map is contoured at a level of 1σ.

6.7.3.2 *apo CaM-(Gly)$_5$-NmIQ2 and apo CaM-(Gly)$_5$-NgIQ2 structures*

NmIQ2 and NgIQ2 both 24aa peptides were linked to the C-terminal of CaM through 5 aa Gly linker (referred to as *apo* CaM-(Gly)$_5$-NmIQ2 and *apo* CaM-(Gly)$_5$-NgIQ2), crystallized, and their structure independently solved by Single wavelength Anomalous Dispersion (SAD) method. Both models were refined up to 2.6 Å resolutions, with good stereo chemical parameters (Table 6.3). The *apo* CaM-(Gly)$_5$-NmIQ2 has 3 molecules in the asymmetric unit, whereas *apo* CaM-(Gly)$_5$-NgIQ2 has 1 molecule in the asymmetric unit. CaM in both structures existed in an extended conformation, with peptide binding to the C-lobe. The structures of both complexes were different (rmsd 4.1Å for 122 Cα atoms). However, the peptide bound to the C-lobes of both CaM structures superimposed with an rmsd of 2.1 Å (Figure 6.19). The bound peptides are well defined in the electron density map (Figure 6.17) and the models have good geometry. No density was observed for any metal ion in any of the divalent metal ion binding sites.

The amino acid distribution of CaM makes the C-lobe more negatively charged than the N-lobe(Osawa *et al.*, 1999). Notably, NmIQ2/NgIQ2 peptides lacked key, bulky hydrophobic residues, but were rich in positively charged amino acids (Figure 6.12). The crystal structure revealed that the bound NmIQ2/NgIQ2 peptide adopted an α-helical structure that was almost perpendicular to the central α-helix of CaM, making several contacts with the C-lobe and a few contacts with central helix of CaM (Figure 6.17 and 6.19 and Table 6.4).

6.7.3.2.1 *apo CaM-(Gly)$_5$-NmIQ2*

Residues Lys55-Gly57 of the NmIQ2 peptide were not well-defined in the electron density map and were not included in the model. PKC phosphorylation site

Ser41 of NmIQ2 was observed to be in contact with Glu115 of CaM, residing in a negatively charged cavity consisting of Asp119, Glu120 and Glu121 (Figure 6.19). Previous studies have shown that a Ser41-to-Asn (S41N) mutation in full-length Nm allows Nm to retain its affinity for *apo* CaM-Sepharose, whereas a Ser41-to-Asp (S41D) mutation causes the affinity to be lost (Chapman *et al.*, 1991). Another report shows that the phosphorylation of Ser41 by PKC blocks the Nm and CaM association and interrupts several learning- and memory-associated functions (Holahan & Routtenberg, 2008, Hulo *et al.*, 2002). Therefore, we suggest that upon phosphorylation by PKC or substitution with a negatively charged amino acid at position 41 (e.g. Asp), may interrupt the interactions between Nm and CaM due to electrostatic repulsion and steric hindrances. By comparison, substitution with a neutral amino acid (e.g. Asn) will not affect the interaction (Figure 6.19). A separate study shows that genetically over expressed wild type Nm can enhance the learning and LTP in transgenic mice; but that over expressed mutant Nm (S41A), does not(Routtenberg *et al.*, 2000). Others have shown that interactions between a S41A Nm mutant and CaM in PC12B cells become insensitive to Ca^{2+} concentrations *in vitro*, with the mutation inhibiting the association of Nm with the membrane skeleton of PC12B cells(Meiri *et al.*, 1996). These findings suggest that the Ser41 phosphorylation site is the main target for enhancing cognitive ability.

The fragment Arg43-Leu51 of Nm has been reported to bind CaM-Sepharose both in the presence of Ca^{2+} or in an excess of EGTA, and can be eluted with 150 mM KCl(Alexander *et al.*, 1988b). By comparison, the Gln39-Lys55 fragment (Trp substituted for Phe42) binding to CaM-Sepharose becomes Ca^{2+}-sensitive (Alexander *et al.*, 1988b). Phe42 of Nm interacts with CaM hydrophobically, and an increase in hydrophobicity at this position (Phe-to-Trp substitution) increases the Nm-CaM

affinity by 10-fold (Chapman *et al.*, 1991). The present crystal structure showed that the fragment Ile38-Phe42 of Nm is in the vicinity of the EF-motifs (i.e., a Ca^{2+}-binding site) of CaM (Figure 6.19). Thus, the binding of fragment Ile38-Phe42 (hence, Gln39-Lys55) is affected by binding of Ca^{2+} to the EF-motif. However, the Arg43-Leu51 fragment (mainly basic amino acids) binds through electrostatic interactions near the central helix of CaM, away from the EF-motifs, and thus remains insensitive to Ca^{2+}. Residues, such as Ile38, Gln39, Phe42 and Arg43, anchor the NmIQ2 peptide in a groove formed by the central helix and EF-motifs of CaM. Further, Arg43 of the IQ peptide makes several contacts with the residues of the central helix of CaM, such as Glu85, Asp81 and Ser82. Our site-directed point mutation of Arg43Ala on the full-length Nm drastically reduced its affinity with CaM (Table 6.2 and Figure 6.18).

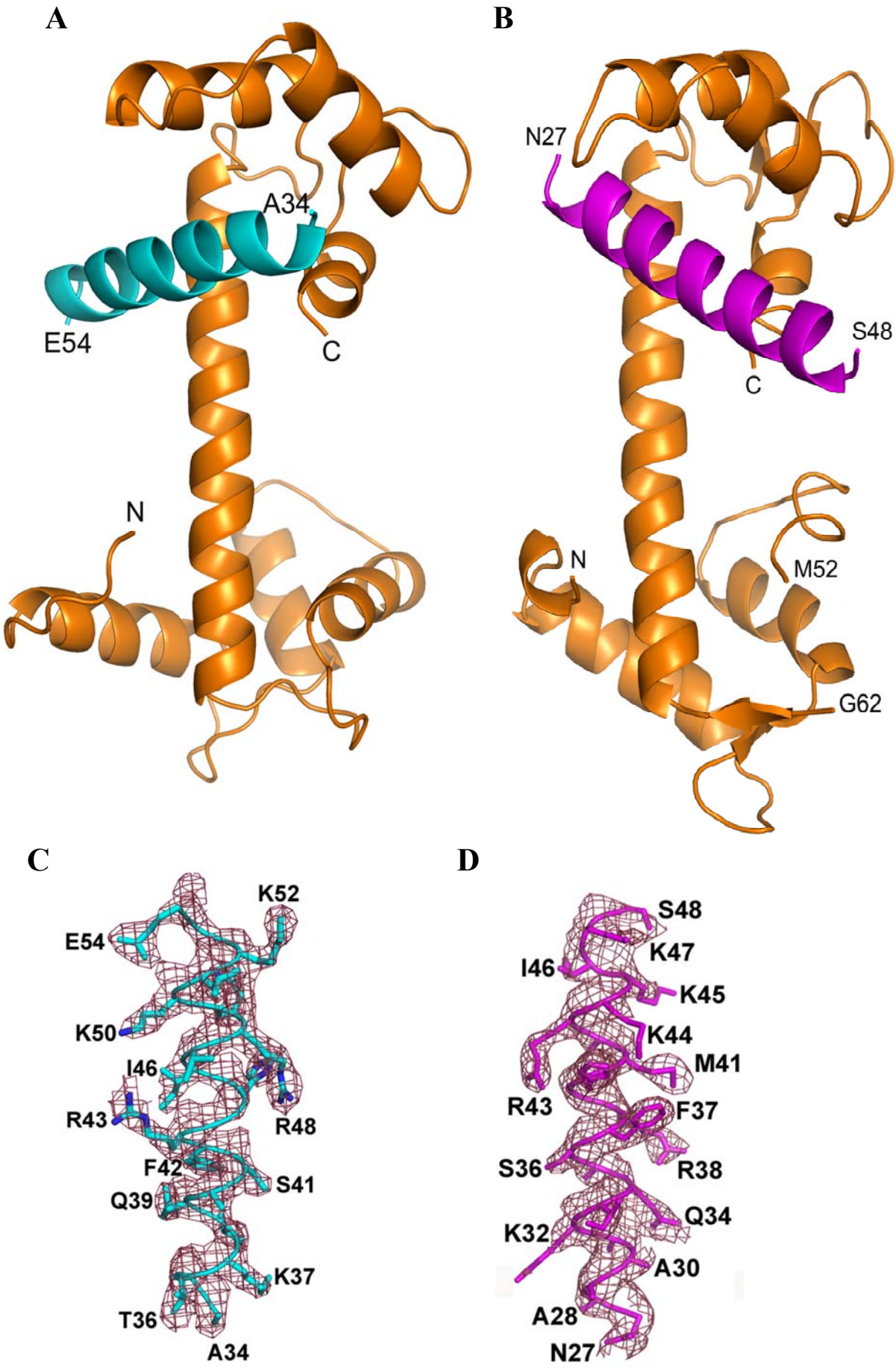

A
A34
E54
C
N
B
N27
S48
C
N
M52
G62
C
K52
E54
K50
I46
R43
R48
F42
S41
Q39
K37
T36
A34
D
S48
K47
I46
K45
K44
M41
R43
F37
S36
R38
Q34
K32
A30
A28
N27

Figure 6.17: Cartoon representations of the structure of **(A)** *apo* CaM-$(Gly)_5$-NmIQ2 and **(B)** *apo* CaM-$(Gly)_5$-NgIQ2 complexes, with CaM (orange), NmIQ2 (cyan) and NgIQ2 (magenta). N and C-termini of the IQ peptides are labeled. *2Fo-Fc* electron density maps of **(C)** NmIQ2 and **(D)** NgIQ2 peptides. Maps are contoured at a level of 1σ.

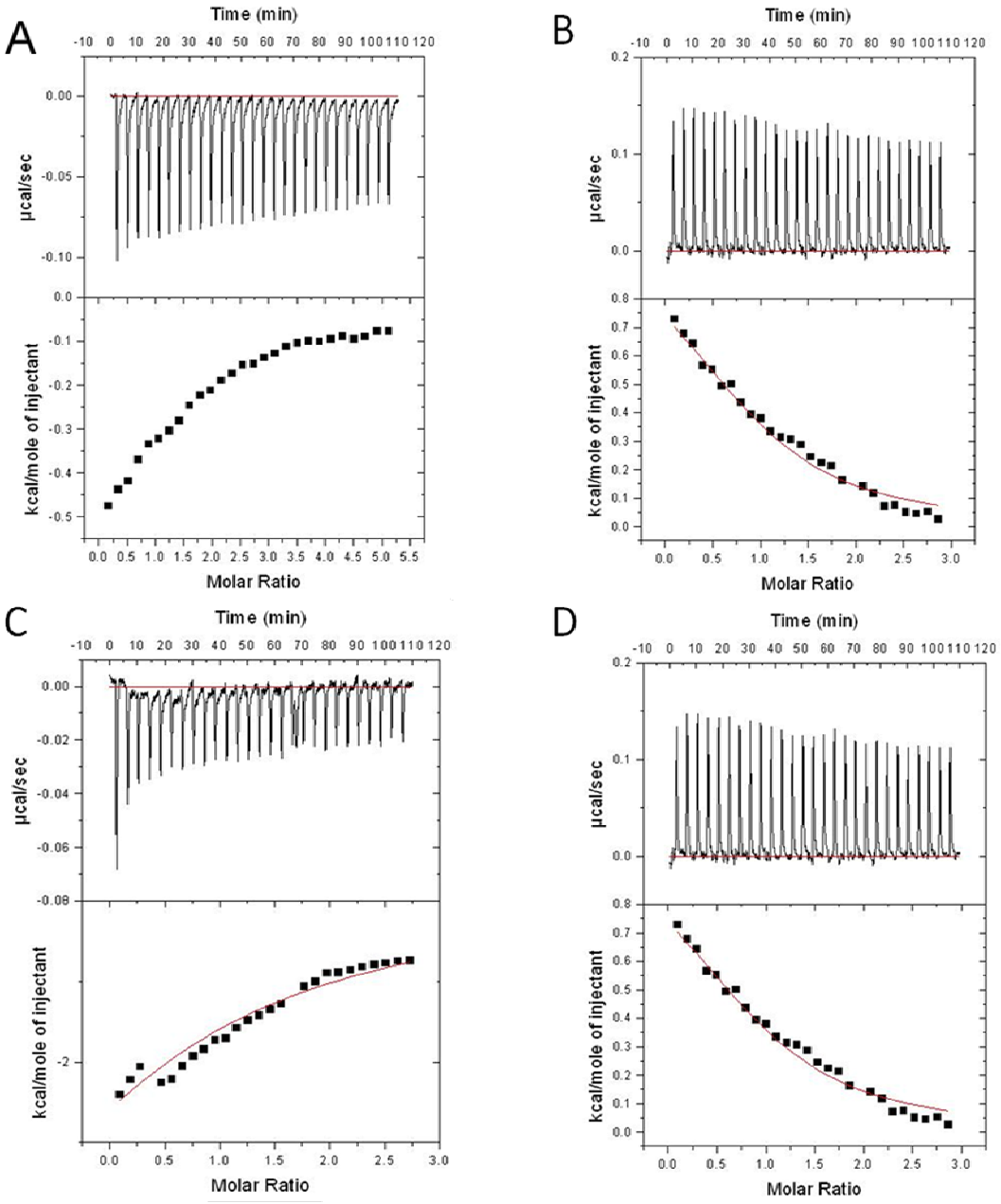


Figure 6.18: ITC profile of R43A Nm and R38A Ng titrated against CaM- **(A)** R43A Nm titrated against Ca^{2+}/CaM; **(B)** R43A Nm titrated against *apo* CaM; **(C)** R38A Ng titrated against Ca^{2+}/CaM; **(D)** R38A Ng titrated against *apo* CaM. Upper panels show the raw data; lower panels show the heat release for each injection. The data is fitted using a single-site binding model.

6.7.3.2.2 *apo CaM-$(Gly)_5$-NgIQ2*

Similar to the *apo* CaM-$(Gly)_5$-NmIQ2 complex, the CaM in *apo* CaM-$(Gly)_5$-NgIQ2 complex existed in extended conformation (Figure 6.17B). No electron density was observed for the loop between Met52 and Gly62 of CaM or for the last two residues (Gly49 and Glu50) of NgIQ2; these residues were not included in the model. Interestingly, the CaM interacting peptide of Ng is from the nearest symmetry-related molecule and the bound peptide is in the opposite direction to that observed in the CaM-$(Gly)_5$-NmIQ2 complex (Figure 6.20). The N-terminal part of the peptide is deeply buried in a cleft formed by the helices from the EF-motif of CaM. The conserved residues of the IQ motif [(I/L/V)QXXXRXXXX(R/K)] interact with CaM and anchor the Ng molecule (Figure 6.19B). Studies have previously shown that the Ser36-phosphorylated Ng is unable to bind CaM and phosphatidic acid (Dominguez-Gonzalez *et al.*, 2007, Diez-Guerra, 2010). In the *apo* CaM-$(Gly)_5$-NgIQ2 structure, Ser36 is buried in a pocket surrounded by negatively charged residues, such as Glu83, Glu84, Glu85, Glu88 and Asp96 (Figure 6.19B). We suggest that upon phosphorylation by PKC, Ser36 may have electrostatic repulsion by these negative charged residues of CaM and also may have steric effects; hence, preventing the interaction between Ng and CaM, and similarly, its interactions with phosphatidic acid.

Table 6.4: Interactions between NmIQ2 and NgIQ2 with CaM.

NmIQ2 residue (atom)	CaM residue (atom)	Interatomic distances in Å
H-bond		
Lys37 (NZ)	Glu121 (OE)	2.98
Ser41 (OG)	Glu115 (OE2)	2.81
Arg43 (NE)	Glu85 (OE1)	2.42
Arg43 (NH2)	Ser82 (OG)	3.49
Lys49 (O)	Glu88 (OE2)	3.27
Lys50 (NZ)	Asp81 (OD2)	2.92
Hydrophobic interactions		
Ile38 (CD1)	Met125 (SD)	3.54
Phe42(CD2)	Phe142 (CZ)	3.23
Ile46 (CG1)	Ala89 (CB)	2.96
NgIQ2 residue (atom)		
H-bond		
Gln34 (OE1)	Glu115 (N)	2.78
Gln34 (NE2)	Leu113 (O)	2.55
Gln34 (NE2)	Met110 (O)	2.64
Arg38 (NH1)	Glu115 (OE1)	2.45
Arg38 (NH2)	Glu121 (OE2)	2.91
Arg43 (NH2)	Met146 (O)	3.32
Hydrophobic interactions		
Ile33 (CD)	Phe90 (CD2)	3.19
Ile33 (CD)	Ala89 (CB)	3.23
Ile33 (CG2)	Met125 (CE)	3.40
Phe37 (CZ)	Met125 (CA)	3.31

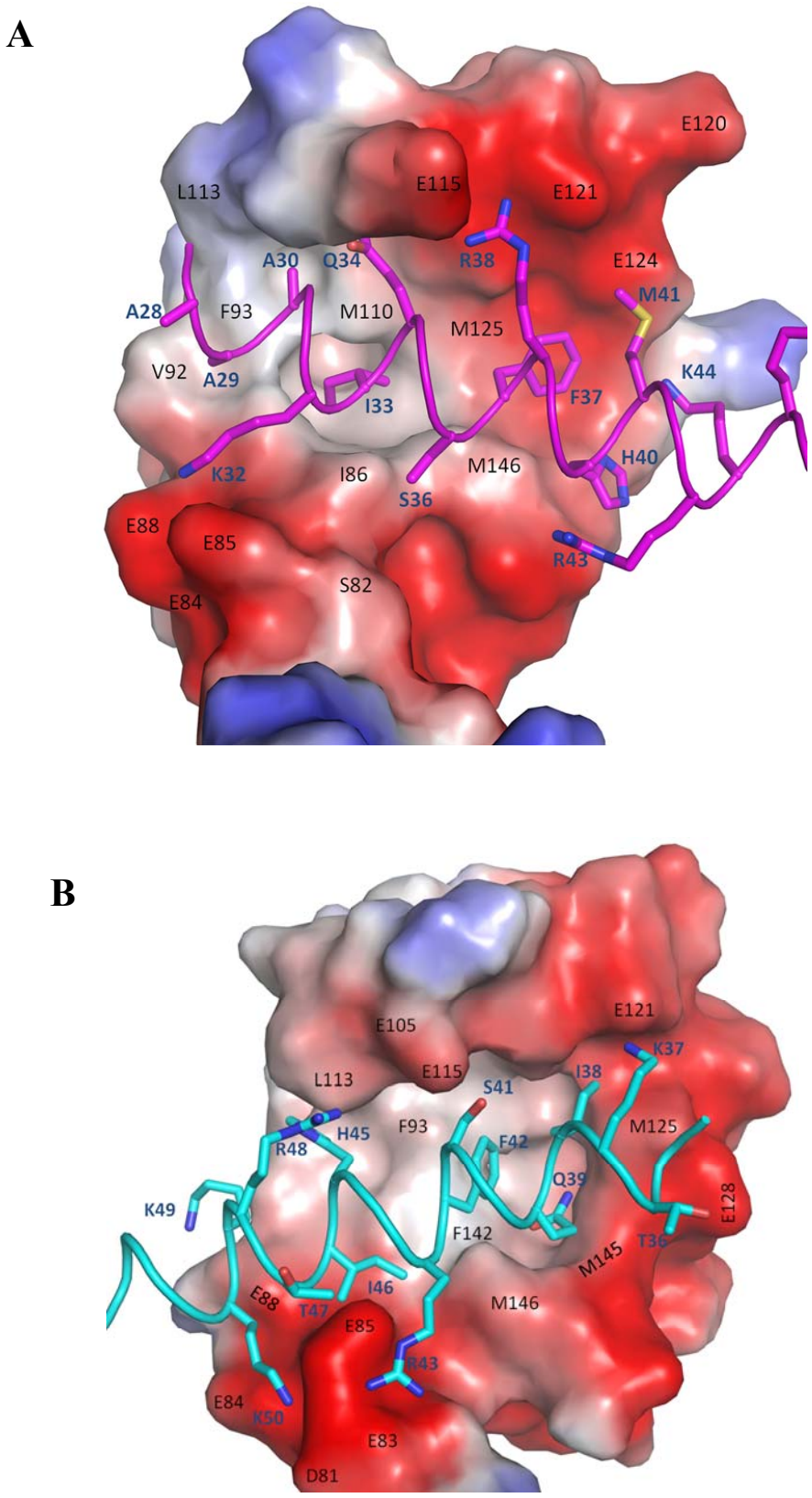


Figure 6.19: Interactions of (**A**) NmIQ2 and (**B**) NgIQ2 peptides with the C-lobe of CaM. IQ peptides are shown in surface representation and key side chains involved in interactions are shown as sticks. The electrostatic surface potential of C-lobes of CaM are shown (negative charge, red; positive charge, blue).

It has been shown that Ser36-phosphorylated Ng does not bind to CaM and is unable to potentiate synaptic transmission and synaptic strength (Han *et al.*, 2007, Zhong *et al.*, 2009, Gerges *et al.*, 2009, Zhong & Gerges, 2010). The identified key CaM interacting residues of NgIQ2 (Ile33, Ser36 and Arg38) are consistent with the residues identified in the full-length Ng for *in vivo* CaM binding (Prichard *et al.*, 1999). The Ile33Gln (I33Q) point mutation completely inhibits the binding, while Ser36Asp (S36D) and Arg38Gln (R38Q) mutations only reduce Ng-CaM binding (Prichard *et al.*, 1999). In the crystal structure, Ile33 makes strong interactions with the hydrophobic cluster consisting of Ala89, Phe90 and Met125 (Table 6.4). A point mutation at this position with a non-hydrophobic amino acid (e.g. I33Q) might disrupt the hydrophobic interaction, also create steric hindrances and abolishes the interaction between Ng and CaM (Prichard *et al.*, 1999). Moreover, a point mutation of S36D in Ng would add a negative charge to this position, but cause less of an effect than if it were to be phosphorylated on Ser36 (Prichard *et al.*, 1999); i.e., while a S36D mutation will reduce Ng-CaM binding, a phosphorylation event will completely abolish it. Similarly, Arg38 of Ng makes hydrogen bonding contacts with Glu115 and Glu121 of CaM (Table 6.4). Our site-directed point mutation of Arg38Ala on the full-length Ng resulted in a reduced binding with CaM in the presence or absence of Ca^{2+} (Table 6.2 and Figure 6.18).

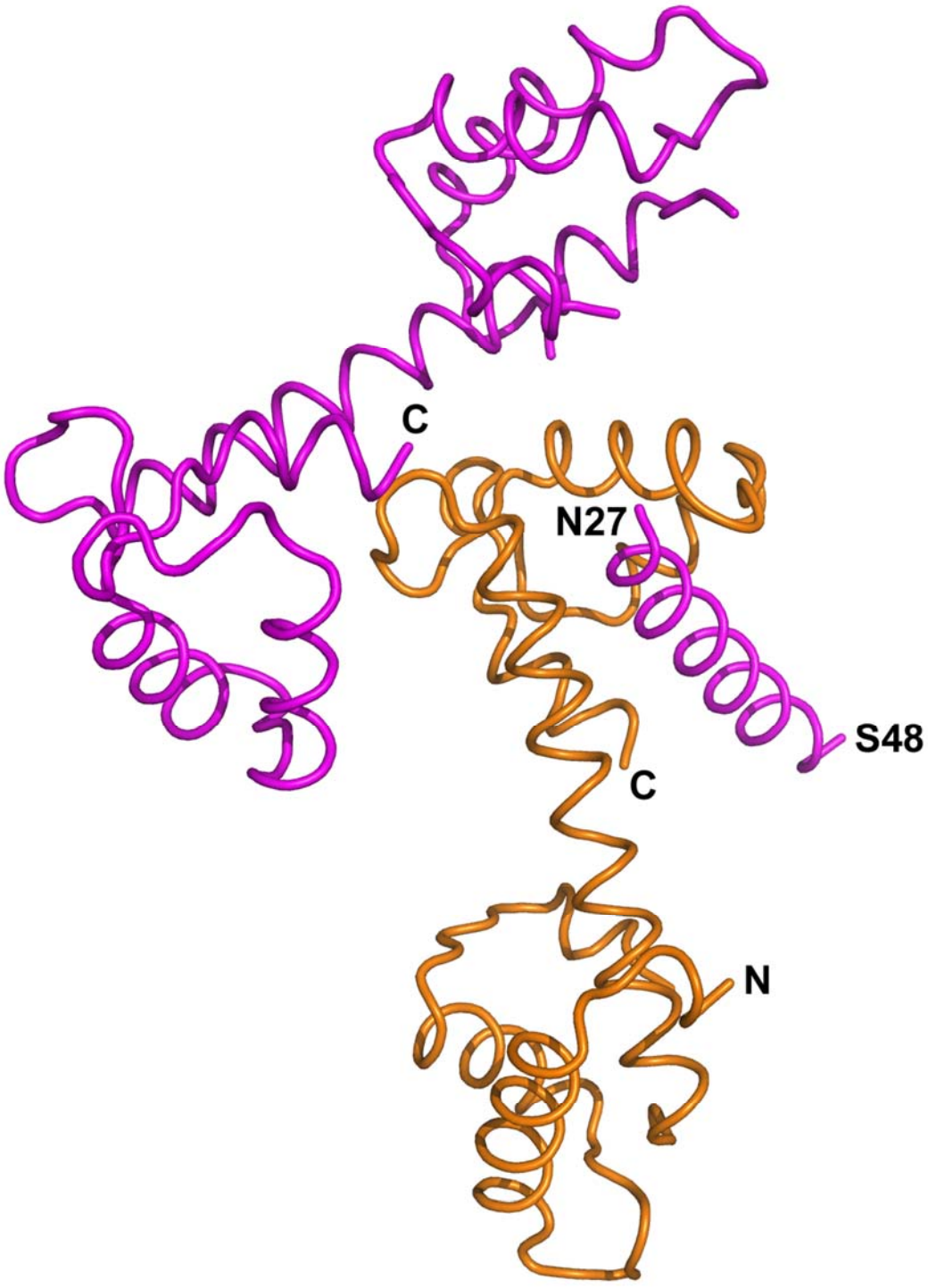


Figure 6.20.: In the crystal structure of *apo* CaM-$(Gly)_5$-Ng, the CaM interacting peptide of Ng is from the nearest symmetry-related molecule. Each molecule of CaM is shown in a different color.

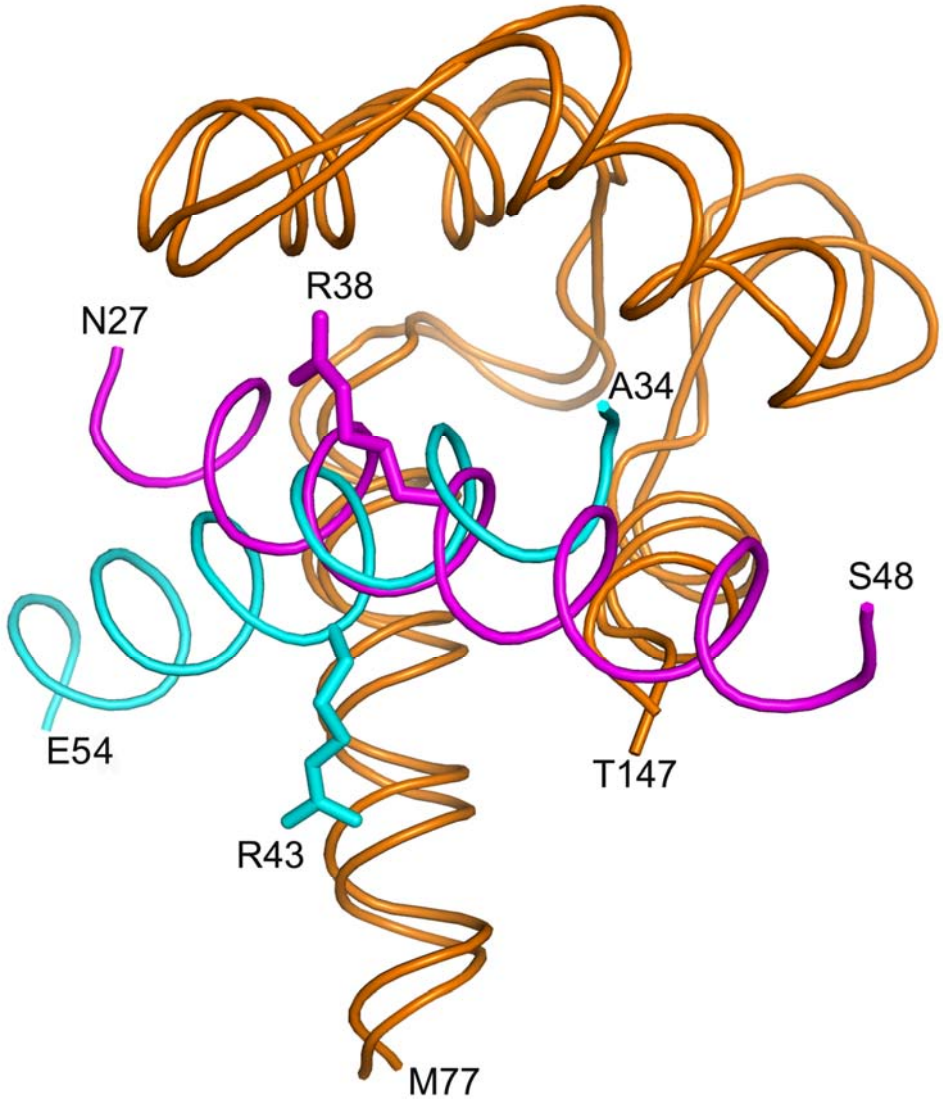


Figure 6.21: Superposition of the structures of NmIQ2 (cyan) and NgIQ2 (magenta) bound to C-lobe of *apo* CaM (orange). Side chains of Arg43 of Nm and Arg38 of Ng are shown as sticks.

6.8 Discussion

Nm and Ng have been the subjects of intense study for their potential roles in brain development and neural plasticity (Zhong *et al.*, 2009, Gerges *et al.*, 2009, Routtenberg *et al.*, 2000). Both proteins are members of the calpacitin family and share a conserved IQ domain that mediates interactions between Ca^{2+}, CaM, and PKC signalling pathways (Gerendasy, 1999, Chapman *et al.*, 1991, Huang *et al.*, 1993).

PKC phosphorylates Ser41 (Nm) and Ser36 (Ng) within the IQ motif and phosphorylated proteins are unable to bind CaM. One proposed biochemical function for Nm and Ng is that it sequesters CaM at the membrane in the vicinity of 'CaM-activated enzymes' under low Ca^{2+} conditions at the pre- and post-synaptic terminals, respectively (Andreasen *et al.*, 1983, Alexander *et al.*, 1988b, Huang *et al.*, 1993). Therefore, Nm and Ng might serve as a Ca^{2+}-regulated modulators of CaM activity in neurons. Moreover, the strict conservation of the region containing the IQ motif in all vertebrates suggests that both the PKC phosphorylation site and the CaM binding feature are essential to the functions of Nm and Ng (Clayton *et al.*, 2009). Although the functions of Nm and Ng have been well-established, the intrinsically unstructured nature of these two proteins, and the consequential lack of structural information, has seriously hampered a complete understanding of the interaction between Nm/Ng and CaM.

From our study, we show that once the Nm/Ng binds with CaM, the interacting IQ motif regions of Nm/Ng adopt an α-helical conformation. Similar conformational changes of intrinsically unstructured binding partners of CaM have been shown for chicken gizzard caldesmon via physico-chemical experiments (Permyakov *et al.*, 2003), and for PEP-19 (Kleerekoper & Putkey, 2009) and calponin (Pfuhl *et al.*, 2011) via NMR. Further, a previous NMR study reported a helical structure for the activation domain of CITED2, which is unstructured in its free form, in complex with its partner TAZ1 (De Guzman *et al.*, 2004). The present report on the Nm/Ng with CaM is the first crystal structure of any neuron specific intrinsically disordered proteins complexed with their binding partner. Besides the IQ motif, Nm and Ng share no sequence homology. However, the N-terminal domain of both

proteins is highly conserved among vertebrates, and the structural information provided here might be extended to other homologs.

Our initial aim was to crystallize the synthetic IQ peptides in complex with Ca^{2+}/CaM and *apo* CaM; however, crystals were only obtained for the Ca^{2+}-bound form. Further, no electron density was observed for the peptides in Ca^{2+}/CaM, and the CaM adopted a unique conformation so far not observed. CaM has been shown to adopt a wide variety of conformations to interact with different targets (Figure 6.16). The N- and C-terminal lobes move in to wrap around the hydrophobic residues of a target molecule (Meador *et al.*, 1993, Gifford *et al.*, 2011). Besides this classical mode of binding, CaM bound to Oedema factor adopts an extended conformation(Drum *et al.*, 2002), where part of the central α-helix transforms into loops for peptide interactions(Shen *et al.*, 2004, Aoyagi *et al.*, 2003), and binds to the target peptide only via the C-terminal lobe(Elshorst *et al.*, 1999, Pagnozzi *et al.*, 2010). In addition, NMR and other spectroscopic studies have shown that the central α-helix is flexible in solution, and that the backbone atoms between residues Lys77 and Asp80 undergo conformational changes(van der Spoel *et al.*, 1996).

Our previous study has shown that the complex between Ng and CaM is not stable for the structure determination by NMR (Ran *et al.*, 2003). In order to study the interactions of Nm/Ng with CaM, the full-length mimic IQ peptide fusion proteins were generated, wherein the IQ peptides were fused at the gene level to the C-terminus of CaM via a flexible linker. The linking of peptides to CaM has been previously reported(Ye *et al.*, 2006) and, based on the analysis on CaM structure, we generated various length linker constructs and finally optimized with a linker of $(Gly)_5$ to provide an appropriate degree of flexibility for the peptide to bind at its naturally preferred site on CaM. For the linked peptide complexes, crystals were

obtained only for *apo* CaM (Ca^{2+} free). It is worth mentioning here that previously the crystal structure of a chimeric protein containing CaM linked via $(Gly)_5$ to the CaM-binding domain (CBD) peptide of calcineurin was determined (Ye *et al.*, 2006). Subsequently CaM structure was solved in complexed with synthetic peptide without any linker [$(Gly)_5$]. The structure was essentially similar to the others, with the same interacting residues and binding regions (Ye *et al.*, 2006, Ye *et al.*, 2008). Although the CBD peptide and the present study IQ peptides of Nm/Ng were linked to CaM with $(Gly)_5$, the interactions of these peptides with CaM were all different. Interestingly, the bound peptides in the NmIQ2-CaM and NgIQ2-CaM complexes were orientated in the opposite direction (Figure 6.21). Besides, in the NmIQ2-CaM complex, Arg43 of Nm makes several contacts with the central helix; whereas, in the NgIQ2-CaM complex, Arg38 of Ng mainly interact the C-lobe of CaM. Yet, the interacting residues of Nm/Ng identified through the IQ motif crystal complexes are consistent with our site-directed mutagenesis on the full-length proteins, as well as those identified through *in vivo* mutagenesis approaches (Prichard *et al.*, 1999, Chapman *et al.*, 1991, Ran *et al.*, 2003). Together all these observations confirm that linker does not restrict the range of possible orientations for intermolecular interactions. The observed interactions of Nm/Nm peptides with CaM represent their natural interactions. The linked complex approach provides a useful mechanism to study protein-protein interactions of weakly interacting proteins, as well as for the interactions of intrinsically unstructured proteins.

Located in the IQ motif, the PKC phosphorylation site is the functionally important site of Nm and Ng. In the complex crystal structure, the PKC phosphorylation sites, Ser41 and Ser36 of Nm and Ng, respectively, are completely surrounded by negatively charged amino acids. Consequently, the phosphorylation of

this Ser residue will repel Nm/Ng from CaM due to electrostatic repulsion and steric hindrance. This clearly explains why the phosphorylation of Ser by PKC blocks the Nm/Ng association with CaM and interrupts several learning- and memory-associated functions (Routtenberg *et al.*, 2000, Hulo *et al.*, 2002, Holahan & Routtenberg, 2008, Zhong & Gerges, 2010, Chapman *et al.*, 1991, Meiri *et al.*, 1996, Han *et al.*, 2007).

In summary, this is the first report of the crystal structures of the intrinsically unstructured, neuron-specific substrate proteins Nm/Ng as IQ peptides in complex with CaM. The unstructured IQ peptides (24 aa) interact with the C-lobe of CaM and gain an α-helical conformation. Biophysical studies with full-length Nm/Ng confirmed their unstructured properties in solution. Further, ITC studies revealed that full-length Nm/Ng, their mutants, and their IQ peptides bind strongly to *apo* CaM than to Ca^{2+}/CaM. This study provides the structural basis for the association of Nm/Ng with CaM, a crucial interaction for several learning- and memory-associated functions in neuronal cells.

Chapter 7

Conclusions and future directions

7.1 Conclusion

We have met our objectives of studying the two neuron specific proteins Nm and Ng and their peptide complexes with *apo* CaM and Ca^{2+}/CaM. Biophysical studies indicate that these two proteins, Nm and Ng are "intrinsically unstructured proteins" belonging to a class that exhibit little secondary structure, high flexibility, and low compactness. Gel filtration chromatography of the purified proteins clearly indicates their respective Stoke radius is larger than what would be expected from crystal structure of the other globular proteins of similar molecular weight. Far-UV circular dichroism (CD), nuclear magnetic resonance (NMR), sequence analysis and aberrant mobility in SDS-PAGE indicate the intrinsically unstructured nature of these two proteins.

Crystallographic study showed that the Ca^{2+}/CaM adopts an unique conformation in the presence of IQ peptides and explains the reason why these two proteins show unusual CaM binding features; low affinity for Ca^{2+}/CaM and higher affinity for *apo* CaM. Crystal structure of *apo* CaM and IQ peptides showed that CaM adopts an extended conformation. In both structures, *apo* CaM-$(Gly)_5$-NmIQ2 and *apo* CaM-$(Gly)_5$-NgIQ2, the IQ peptides interact with mainly C-domain of the CaM. The interacting residues are consistent with the earlier predictions arrived through mutagenesis and NMR studies. Linking of the peptides to CaM has been reported previously for at least two cases and each of them has different interactions with CaM (which depends on the interacting partner) was shown(Ye *et al.*, 2006), and is independent of whether the peptide is linked or not. Length of the linker (5XGly) is large enough to give flexibility to peptide for docking into its specific binding site on CaM. This method might be applicable for other low affinity protein-protein or protein-peptides interactions. However, the length of the linkers needs to be

optimized. Besides we have carried out interaction studies using ITC. The ITC results and structural studies showed that only a small motif of full length Nm and Ng is sufficient to make interactions with CaM. Consistent with the previous studies, ITC results showed that both proteins (and IQ peptides) interact with higher affinity to *apo* CaM than with Ca^{2+}/CaM. In this study we show for the first time the crystal structure of the neuron specific intrinsically unstructured proteins and this study explains how the unstructured proteins gain structure upon binding with its partners.

7.2 Future studies

With the knowledge gained through this study, our future objective is to crystallize the full length Nm/Ng complexed with CaM. Following this, structure based functional studies will be conducted to understand the role of these substrate proteins. We shall be verifying the structural finding of IQ peptides interaction with full length proteins.

Reference:

Alexander, K. A., Cimler, B. M., Meier, K. E. & Storm, D. R. (1987). *J Biol Chem* **262**, 6108-6113.

Alexander, K. A., Wakim, B. T., Doyle, G. S., Walsh, K. A. & Storm, D. R. (1988a). *J Biol Chem* **263**, 7544-7549.

Alexander, K. A., Wakim, B. T., Doyle, G. S., Walsh, K. A. & Storm, D. R. (1988b). *The Journal of biological chemistry* **263**, 7544-7549.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res* **25**, 3389-3402.

Alvarez-Bolado, G., Rodriguez-Sanchez, P., Tejero-Diez, P., Fairen, A. & Diez-Guerra, F. J. (1996). *Neuroscience* **73**, 565-580.

Andreasen, T. J., Keller, C. H., LaPorte, D. C., Edelman, A. M. & Storm, D. R. (1981). *Proc Natl Acad Sci U S A* **78**, 2782-2785.

Andreasen, T. J., Luetje, C. W., Heideman, W. & Storm, D. R. (1983). *Biochemistry* **22**, 4615-4618.

Aoyagi, M., Arvai, A. S., Tainer, J. A. & Getzoff, E. D. (2003). *Embo J* **22**, 766-775.

Babu, Y. S., Bugg, C. E. & Cook, W. J. (1988). *J Mol Biol* **204**, 191-204.

Babu, Y. S., Sack, J. S., Greenhough, T. J., Bugg, C. E., Means, A. R. & Cook, W. J. (1985). *Nature* **315**, 37-40.

Bakir, M. A., Sakamoto, M., Kitahara, M., Matsumoto, M. & Benno, Y. (2006). *Int J Syst Evol Microbiol* **56**, 1639-1643.

Barnett, M. W. & Larkman, P. M. (2007). *Pract Neurol* **7**, 192-197.

Baudier, J., Bronner, C., Kligman, D. & Cole, R. D. (1989). *J Biol Chem* **264**, 1824-1828.

Benowitz, L. I. & Routtenberg, A. (1987). *Trends Neurosci* **10**, 527-532.

Benowitz, L. I. & Routtenberg, A. (1997a). *Trends Neurosci* **20**, 84-91.

Benowitz, L. I. & Routtenberg, A. (1997b). *Trends Neurosci* **20**, 84-91.

Bernal, J., Rodriguez-Pena, A., Iniguez, M. A., Ibarrola, N. & Munoz, A. (1992). *Acta Med Austriaca* **19 Suppl 1**, 32-35.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J Mol Biol* **112**, 535-542.

Bjursell, M. K., Martens, E. C. & Gordon, J. I. (2006). *J Biol Chem* **281**, 36269-36279.

Bodner, S. J., Koenig, M. G., Treanor, L. L. & Goodman, J. S. (1972). *Antimicrob Agents Chemother* **2**, 57-60.

Boissier, F., Bardou, F., Guillet, V., Uttenweiler-Joseph, S., Daffe, M., Quemard, A. & Mourey, L. (2006). *J Biol Chem* **281**, 4434-4445.

Booth, S. J., Van Tassell, R. L., Johnson, J. L. & Wilkins, T. D. (1979). *Rev Infect Dis* **1**, 325-336.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Crystallogr D Biol Crystallogr* **54**, 905-921.

Bujnicki, J. M. (1999). *In Silico Biol* **1**, 175-182.

Cammarota, M., Paratcha, G., Levi de Stein, M., Bernabeu, R., Izquierdo, I. & Medina, J. H. (1997). *Neurochem Res* **22**, 499-505.

Cantoni, G. L. (1975). *Annu Rev Biochem* **44**, 435-451.

Caroni, P. (2001). *Embo J* **20**, 4332-4336.
Chapman, E. R., Au, D., Alexander, K. A., Nicolson, T. A. & Storm, D. R. (1991). *J Biol Chem* **266**, 207-213.
Cheney, R. E. & Mooseker, M. S. (1992). *Curr Opin Cell Biol* **4**, 27-35.
Cheng, X. (1995). *Annu Rev Biophys Biomol Struct* **24**, 293-318.
Cheng, X., Kumar, S., Klimasauskas, S. & Roberts, R. J. (1993). *Cold Spring Harb Symp Quant Biol* **58**, 331-338.
Cheng, X. & Roberts, R. J. (2001). *Nucleic Acids Res* **29**, 3784-3795.
Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. & Thompson, J. D. (2003). *Nucleic Acids Res* **31**, 3497-3500.
Chou, J. J., Li, S., Klee, C. B. & Bax, A. (2001). *Nat Struct Biol* **8**, 990-997.
Cimler, B. M., Andreasen, T. J., Andreasen, K. I. & Storm, D. R. (1985). *J Biol Chem* **260**, 10784-10788.
Clarke, S. (1993). *Curr Opin Cell Biol* **5**, 977-983.
Clayton, D. F., George, J. M., Mello, C. V. & Siepka, S. M. (2009). *Dev Neurobiol* **69**, 124-140.
Cowtan, K. (2006). *Acta Crystallogr D Biol Crystallogr* **62**, 1002-1011.
Dash, S., Niemaczura, W. & Harrington, H. M. (1997). *Biochemistry* **36**, 2025-2029.
De Guzman, R. N., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. (2004). *The Journal of biological chemistry* **279**, 3042-3049.
DeLano, W. L. & Lam, J. W. (2005). *Abstracts of Papers of the American Chemical Society* **230**, U1371-U1372.
Denny, J. B. (2006). *Curr Neuropharmacol* **4**, 293-304.
Devireddy, L. R. & Green, M. R. (2003). *Mol Cell Biol* **23**, 4532-4541.
Diez-Guerra, F. J. (2010). *IUBMB Life* **62**, 597-606.
Dixon, M. M., Huang, S., Matthews, R. G. & Ludwig, M. (1996). *Structure* **4**, 1263-1275.
Dominguez-Gonzalez, I., Vazquez-Cuesta, S. N., Algaba, A. & Diez-Guerra, F. J. (2007). *Biochem J* **404**, 31-43.
Doster, S. K., Lozano, A. M., Aguayo, A. J. & Willard, M. B. (1991). *Neuron* **6**, 635-647.
Drum, C. L., Yan, S. Z., Bard, J., Shen, Y. Q., Lu, D., Soelaiman, S., Grabarek, Z., Bohm, A. & Tang, W. J. (2002). *Nature* **415**, 396-402.
Elshorst, B., Hennig, M., Forsterling, H., Diener, A., Maurer, M., Schulte, P., Schwalbe, H., Griesinger, C., Krebs, J., Schmid, H., Vorherr, T. & Carafoli, E. (1999). *Biochemistry* **38**, 12320-12332.
Emsley, P. & Cowtan, K. (2004). *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132.
Fitzgerald, M., Reynolds, M. L. & Benowitz, L. I. (1991). *Neuroscience* **41**, 187-199.
Fok, A. K., Aihara, M. S., Ishida, M. & Allen, R. D. (2008). *J Eukaryot Microbiol* **55**, 481-491.
Geiser, J. R., van Tuinen, D., Brockerhoff, S. E., Neff, M. M. & Davis, T. N. (1991). *Cell* **65**, 949-959.
Gerendasy, D. (1999). *J Neurosci Res* **58**, 107-119.
Gerges, N. Z., Zhong, L., Cherry, T., Bies, C. E. & Florence, M. A. (2009). *Embo J* **28**, 3027-3039.
Gianotti, C., Nunzi, M. G., Gispen, W. H. & Corradetti, R. (1992). *Neuron* **8**, 843-848.
Gifford, J. L., Ishida, H. & Vogel, H. J. (2011). *J Biomol NMR* **50**, 71-81.
Gonzalez, B., Pajares, M. A., Martinez-Ripoll, M., Blundell, T. L. & Sanz-Aparicio, J. (2004). *J Mol Biol* **338**, 771-782.

Goodall, M. & Kirshner, N. (1958). *Circulation* **17**, 366-371.
Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). *Bioinformatics* **15**, 305-308.
Grewal, S. I. & Rice, J. C. (2004). *Curr Opin Cell Biol* **16**, 230-238.
Griffith, S. C., Sawaya, M. R., Boutz, D. R., Thapar, N., Katz, J. E., Clarke, S. & Yeates, T. O. (2001). *J Mol Biol* **313**, 1103-1116.
Han, N. L., Wen, J., Lin, Q., Tan, P. L., Liou, Y. C. & Sheu, F. S. (2007). *Int J Biol Sci* **3**, 263-273.
Hayashi, N., Matsubara, M., Titani, K. & Taniguchi, H. (1997). *J Biol Chem* **272**, 7639-7645.
He, Q., Dent, E. W. & Meiri, K. F. (1997). *J Neurosci* **17**, 3515-3524.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *Embo J* **9**, 1665-1672.
Hengen, P. (1995). *Trends Biochem Sci* **20**, 285-286.
Holahan, M. & Routtenberg, A. (2008). *Hippocampus* **18**, 1099-1102.
Holm, L. & Sander, C. (1995). *Trends Biochem Sci* **20**, 478-480.
Huang, C. C., Smith, C. V., Glickman, M. S., Jacobs, W. R., Jr. & Sacchettini, J. C. (2002). *J Biol Chem* **277**, 11559-11569.
Huang, K. P., Huang, F. L. & Chen, H. C. (1993). *Arch Biochem Biophys* **305**, 570-580.
Huang, K. P., Huang, F. L., Jager, T., Li, J., Reymann, K. G. & Balschun, D. (2004). *The Journal of neuroscience : the official journal of the Society for Neuroscience* **24**, 10660-10669.
Hulo, S., Alberi, S., Laux, T., Muller, D. & Caroni, P. (2002). *Eur J Neurosci* **15**, 1976-1982.
Husain, N., Obranic, S., Koscinski, L., Seetharaman, J., Babic, F., Bujnicki, J. M., Maravic-Vlahovicek, G. & Sivaraman, J. (2011). *Nucleic Acids Res* **39**, 1903-1918.
Husain, N., Tkaczuk, K. L., Tulsidas, S. R., Kaminska, K. H., Cubrilo, S., Maravic-Vlahovicek, G., Bujnicki, J. M. & Sivaraman, J. (2010). *Nucleic Acids Res* **38**, 4120-4132.
Husson, M., Enderlin, V., Alfos, S., Feart, C., Higueret, P. & Pallet, V. (2003). *Br J Nutr* **90**, 191-198.
Kleerekoper, Q. K. & Putkey, J. A. (2009). *The Journal of biological chemistry* **284**, 7455-7464.
Kozbial, P. Z. & Mushegian, A. R. (2005). *BMC Struct Biol* **5**, 19.
Krissinel, E. & Henrick, K. (2007). *J Mol Biol* **372**, 774-797.
Lambe, D. W., Jr. (1974). *Appl Microbiol* **28**, 561-567.
Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *Journal of Applied Crystallography* **26**, 283-291.
le Maire, M., Ghazi, A., Moller, J. V. & Aggerbeck, L. P. (1987). *Biochem J* **243**, 399-404.
Lee, P. T., Hsu, A. Y., Ha, H. T. & Clarke, C. F. (1997). *J Bacteriol* **179**, 1748-1754.
Li, H. Y., Li, J. F. & Lu, G. W. (2003). *Sheng Li Ke Xue Jin Zhan* **34**, 111-115.
Li, J., Huang, F. L. & Huang, K. P. (2001). *The Journal of biological chemistry* **276**, 3098-3105.
Lim, K., Zhang, H., Tempczyk, A., Bonander, N., Toedt, J., Howard, A., Eisenstein, E. & Herzberg, O. (2001). *Proteins* **45**, 397-407.
Loenen, W. A. (2006). *Biochem Soc Trans* **34**, 330-333.
Macmaster, R., Zelinskaya, N., Savic, M., Rankin, C. R. & Conn, G. L. (2010). *Nucleic Acids Res* **38**, 7791-7799.
Malenka, R. C. & Bear, M. F. (2004). *Neuron* **44**, 5-21.

Martin, J. L. & McMillan, F. M. (2002). *Curr Opin Struct Biol* **12**, 783-793.
McCarthy, R. E., Pajeau, M. & Salyers, A. A. (1988). *Appl Environ Microbiol* **54**, 1911-1916.
McFerrin, M. B. & Snell, E. H. (2002). *Journal of Applied Crystallography* **35**, 538-545.
Meador, W. E., Means, A. R. & Quiocho, F. A. (1993). *Science* **262**, 1718-1721.
Meganathan, R. (2001). *FEMS Microbiol Lett* **203**, 131-139.
Meiri, K. F., Hammang, J. P., Dent, E. W. & Baetge, E. E. (1996). *J Neurobiol* **29**, 213-232.
Miller, D. J., Ouellette, N., Evdokimova, E., Savchenko, A., Edwards, A. & Anderson, W. F. (2003). *Protein Sci* **12**, 1432-1442.
Mineta, K., Nakazawa, M., Cebria, F., Ikeo, K., Agata, K. & Gojobori, T. (2003). *Proc Natl Acad Sci U S A* **100**, 7666-7671.
Miyagawa, E., Azuma, R. & Suto, T. (1978). *Journal of General and Applied Microbiology* **24**, 341-348.
Miyakawa, T., Yared, E., Pak, J. H., Huang, F. L., Huang, K. P. & Crawley, J. N. (2001). *Hippocampus* **11**, 763-775.
Mosevitsky, M. I. (2005). *Int Rev Cytol* **245**, 245-325.
Mosevitsky, M. I., Konovalova, E. S., Bitchevaya, N. K. & Klementiev, B. I. (2001). *Neurosci Lett* **297**, 49-52.
Murtaugh, T. J., Rowe, P. M., Vincent, P. L., Wright, L. S. & Siegel, F. L. (1983). *Methods Enzymol* **102**, 158-170.
Neuner-Jehle, M., Denizot, J. P. & Mallet, J. (1996). *Brain Res* **733**, 149-154.
O'Day, D. H. (2003). *Cell Signal* **15**, 347-354.
Osawa, M., Tokumitsu, H., Swindells, M. B., Kurihara, H., Orita, M., Shibanuma, T., Furuya, T. & Ikura, M. (1999). *Nat Struct Biol* **6**, 819-824.
Otwinowski, Z. & Minor, W. (1997). *Macromolecular Crystallography, Pt A* **276**, 307-326.
Pagnozzi, D., Birolo, L., Leo, G., Contessi, S., Lippe, G., Pucci, P. & Mavelli, I. (2010). *Biochemistry* **49**, 7542-7552.
Paster, B. J., Dewhirst, F. E., Olsen, I. & Fraser, G. J. (1994). *J Bacteriol* **176**, 725-732.
Permyakov, S. E., Millett, I. S., Doniach, S., Permyakov, E. A. & Uversky, V. N. (2003). *Proteins* **53**, 855-862.
Pfuhl, M., Al-Sarayreh, S. & El-Mezgueldi, M. (2011). *Biophys J* **100**, 1718-1728.
Pierce, M. M., Raman, C. S. & Nall, B. T. (1999). *Methods* **19**, 213-221.
Poehlsgaard, J. & Douthwaite, S. (2005). *Nat Rev Microbiol* **3**, 870-881.
Poon, W. W., Barkovich, R. J., Hsu, A. Y., Frankel, A., Lee, P. T., Shepherd, J. N., Myles, D. C. & Clarke, C. F. (1999). *J Biol Chem* **274**, 21665-21672.
Prichard, L., Deloulme, J. C. & Storm, D. R. (1999). *J Biol Chem* **274**, 7689-7694.
Ran, X., Miao, H. H., Sheu, F. S. & Yang, D. (2003). *Biochemistry* **42**, 5143-5150.
Razin, A. & Cedar, H. (1991). *Microbiol Rev* **55**, 451-458.
Rehm, T., Huber, R. & Holak, T. A. (2002). *Structure* **10**, 1613-1618.
Riederer, B. M. & Routtenberg, A. (1999). *Brain Res Mol Brain Res* **71**, 345-348.
Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J. & Dunker, A. K. (2001). *Proteins* **42**, 38-48.
Routtenberg, A., Cantallops, I., Zaffuto, S., Serrano, P. & Namgung, U. (2000). *Proc Natl Acad Sci U S A* **97**, 7657-7662.
Salyers, A. A., Gupta, A. & Wang, Y. (2004). *Trends Microbiol* **12**, 412-416.

Schubert, H. L., Blumenthal, R. M. & Cheng, X. (2003). *Trends Biochem Sci* **28**, 329-335.
Shah, H. N. & Collins, D. M. (1990). *International Journal of Systematic Bacteriology* **40**, 205-208.
Shah, H. N. & Collins, M. D. (1988). *International Journal of Systematic Bacteriology* **38**, 128-131.
Shah, H. N. & Collins, M. D. (1989). *International Journal of Systematic Bacteriology* **39**, 85-87.
Sheldrick, G. M. (2008). *Acta Crystallogr A* **64**, 112-122.
Shen, Y., Guo, Q., Zhukovskaya, N. L., Drum, C. L., Bohm, A. & Tang, W. J. (2004). *Biochem Biophys Res Commun* **317**, 309-314.
Shen, Y., Mani, S., Donovan, S. L., Schwob, J. E. & Meiri, K. F. (2002). *J Neurosci* **22**, 239-247.
Slemmon, J. R., Morgan, J. I., Fullerton, S. M., Danho, W., Hilbush, B. S. & Wengenack, T. M. (1996). *J Biol Chem* **271**, 15911-15917.
Strittmatter, S. M., Fankhauser, C., Huang, P. L., Mashimo, H. & Fishman, M. C. (1995). *Cell* **80**, 445-452.
Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. (2000). *Nucleic Acids Res* **28**, 33-36.
Tsuji, E., Okazaki, K., Isaji, M. & Takeda, K. (2009). *J Struct Biol* **165**, 133-139.
Uversky, V. N. (2002). *Protein science : a publication of the Protein Society* **11**, 739-756.
Vagin, A. & Teplyakov, A. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 22-25.
Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. & Murshudov, G. N. (2004). *Acta Crystallogr D Biol Crystallogr* **60**, 2184-2195.
van der Spoel, D., de Groot, B. L., Hayward, S., Berendsen, H. J. & Vogel, H. J. (1996). *Protein Sci* **5**, 2044-2053.
Vedantam, G. (2009). *Future Microbiol* **4**, 413-423.
Warren, J. T., Guo, Q. & Tang, W. J. (2007). *J Mol Biol* **374**, 517-527.
Weisburg, W. G., Oyaizu, Y., Oyaizu, H. & Woese, C. R. (1985). *J Bacteriol* **164**, 230-236.
Woodard, R. W., Tsai, M. D., Floss, H. G., Crooks, P. A. & Coward, J. K. (1980). *J Biol Chem* **255**, 9124-9127.
Wu, J., Li, J., Huang, K. P. & Huang, F. L. (2002). *J Biol Chem* **277**, 19498-19505.
Xia, Z. & Storm, D. R. (2005). *Nat Rev Neurosci* **6**, 267-276.
Xu, J., Bjursell, M. K., Himrod, J., Deng, S., Carmichael, L. K., Chiang, H. C., Hooper, L. V. & Gordon, J. I. (2003). *Science* **299**, 2074-2076.
Xu, J., Mahowald, M. A., Ley, R. E., Lozupone, C. A., Hamady, M., Martens, E. C., Henrissat, B., Coutinho, P. M., Minx, P., Latreille, P., Cordum, H., Van Brunt, A., Kim, K., Fulton, R. S., Fulton, L. A., Clifton, S. W., Wilson, R. K., Knight, R. D. & Gordon, J. I. (2007). *PLoS Biol* **5**, e156.
Yap, K. L., Ames, J. B., Swindells, M. B. & Ikura, M. (1999). *Proteins* **37**, 499-507.
Ye, Q., Li, X., Wong, A., Wei, Q. & Jia, Z. (2006). *Biochemistry* **45**, 738-745.
Ye, Q., Wang, H., Zheng, J., Wei, Q. & Jia, Z. (2008). *Proteins* **73**, 19-27.
Zhabotinsky, A. M., Camp, R. N., Epstein, I. R. & Lisman, J. E. (2006). *J Neurosci* **26**, 7337-7347.
Zhong, L., Cherry, T., Bies, C. E., Florence, M. A. & Gerges, N. Z. (2009). *The EMBO journal* **28**, 3027-3039.
Zhong, L. & Gerges, N. Z. (2010). *Commun Integr Biol* **3**, 340-342.

Zor, T. & Selinger, Z. (1996). *Anal Biochem* **236**, 302-308.
Zuhlke, R. D., Pitt, G. S., Deisseroth, K., Tsien, R. W. & Reuter, H. (1999). *Nature* **399**, 159-162.