STRUCTURAL STUDIES OF TRANSCRIPTIONAL REPRESSOR DOMAIN (TRD) OF MBD1 AND MIDA PROTEINS

UMAR FAROOK SHAHUL HAMEED

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOLOGICAL SCIENCES

NATIONAL UNIVERSITY OF SINGAPORE

2013

STRUCTURAL STUDIES OF TRANSCRIPTIONAL REPRESSOR DOMAIN (TRD) OF MBD1 AND MIDA PROTEINS

UMAR FAROOK SHAHUL HAMEED

(M. Tech Medical Biotechnology)

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOLOGICAL SCIENCES

NATIONAL UNIVERSITY OF SINGAPORE

2013

I Dedicate this Thesis

to

My Dear Parents, Family and Friends

ACKNOWLEDGEMENT

First of All, I would like to thank Almighty for giving me the strength and power, and I would like to thank my parents, family members, teachers and friends for their kind support and love, without them it could have been not possible for me to reach my goals.

This work would have been impossible without the relentless support and guidance of my supervisor, Prof. Kunchithapadam Swaminathan. I would like to thank him a lot for providing me excellent guidance and motivation for both the professional and personal aspects of this journey. I am grateful to him for patiently tolerating my mistakes and politely pointing out them, and teaching me to overcome difficult situations and, in the process, molding me into a better researcher. He is a real good teacher with plenty of enthusiasm always with his words “keep going”, which helped me to achieve my goals. I would also like to thank him for keeping his door always open, to listen to my problems and provide solutions.

I would like to thank our collaborator Prof. Yang Daiwen, for his precious time and suggestion, which really helped me shape my project and further proceed in my PhD. I am grateful to Dr. Lee Sang Hyun, for his financial support during my final year and for his kind support and valuable research inputs. I would also like to thank Prof. Luciano Di Croce, CRG, Barcelona, Spain, for allowing me to work for 4 months in his lab. He is really kind, supportive, helpful and nonetheless, a good teacher. I would like to thank Dr. Lim Jack Wee for his help in studying the TRD structure and it was really great to work with him.

It is my pleasure to specially thank Dr. Siva Kumar for his kind help and guidance to learn things, when I came as a naive person to the lab. I would also like to thank our SBL members Dr. Smarajit, Dr. Tan Kangwei, Nguyen Van Sang, Wentao,

Dr. B.C. Karthik, Lei Chang, Bhaskar, Bidhan, Dr. Kuntal Pal, Anupriya, Pavithra, Roopa, Divya, Madhuri, Kanmani, Dr. Feng xia, Anupama, Vindhya, Toan, Dr. Pankaj Kumar Giri, Priyanka, Nilofer, Dr. Jobichen Chacko, Lissa, Dr. Kumar, Deepthi, Dr. Suguna Badireddy, Karthiga, Rajesh Rattinam, Sarath, Digant, Dr. Sunil Tewary, Dr. Veerendra Kumar, Dr. Vinoth Roy, Dr. Jeremy Tan, other present and past members of Lab 4 and Lab5 and all friends from NUS during my PhD period for their help in my needy times.

I would like to specially thank my friends Abhilash, Manjeet and Thangavelu from SBL5 who are really supportive, encouraging and inspirational throughout my PhD period. We, as a group, go together for lunch, dinner and enjoy other wonderful times. I would like to thank my wife Thasneem, who really has been a boon to me throughout my hard times and for her patience. Finally, I would like to acknowledge the NUS research scholarship for pursuing my research for these 4 years.

TABLE OF CONTENTS

SUMMARY

MBD1 belongs to the family of proteins with the characteristic Methylated C_pG Binding Domain (MBD) and plays a very important role in neuronal development and division through transcriptional repression of genes involved in that pathway. MBD1 also reported to be involved in tumor progression, by repressing tumor suppressor genes when over expressed, especially in cancer cells. Until now, the MBD domain and CXXC domain, which has a cysteine rich zing finger binding domain and tends to bind unmethylated DNA, have been characterized but little is known about the potential Transcriptional Repressor Domain (TRD), which is involved in recruiting binding partners needed for repressor complex formation to suppress a gene through promoter methylation. We have solved the structure of TRD using the NMR technique and also characterized the mode of its binding with partners. To our surprise, such an effective repressor domain is unstructured with little defined secondary structures. TRD tends to bind to its interacting partners weakly, in a transient manner, but still with a lot of specificity. It can be concluded that TRD needs to be unstructured to bind several partners, rather than having a specific fold, to play a centered role in cell signaling pathways.

Mitochondrial Complex I is an important member of Oxidative Phosphorylation (OXPHOS) system involved in the production of ATP, which is the energy currency for cells. Up to 40% of various mitochondrial metabolic disorders that affect many vital organs of the body occur due to the deficiency in Complex I activity. At least six chaperones have been reported to be important for Complex I formation, where any mutations of those chaperones affect the function of the mitochondrion. MidA is a mitochondrial protein and has been classified as an uncharacterized protein with unknown function. Recent studies have revealed the

importance of MidA in the stability or assembly of mitochondrial Complex I. MidA was identified to be important for the survival of *Dictyostelium*, where MidA$^-$ knockout cells showed severe phenotypic defects due to the inefficiency of mitochondrial function and also knock down of the MidA homologue in humans resulted in reduced complex assembly which could result in disease conditions. MidA was also found to be interacting with NDUFS2, one of the core subunits of Complex I, an iron-sulfur protein and also possibly a chaperone. I have determined the crystal structure of the MidA protein of *Dictyostelium* with SAM and SAH at a resolution of 2.9 Å. Based on our sequence, structural and biophysical analyses, we have shown that MidA has a conserved methyltransferase motif I and and confirmed SAM binding property. We propose that MidA is a class I methyltransferase. Based on the earlier information that MidA interacts with the N-terminal 40 residues of NDUFS2, I have made a MidA: NDUFS2 complex model by docking a 3D model of NDUFS2 with our MidA structure. . The MidA: NDUFS2 model is quite convincing that MidA forms a specific fold *via* an interaction of its N-terminal helices with its C-terminal region and then is involved in binding to NDUFS2, which could be a possible substrate. However, this hypothesis should be further supported experimentally. The crystal structure of the MidA: NDUFS2 complex is a potential future goal to understand the role of MidA in mitochondrial function.

LIST OF FIGURES

Page

LIST OF TABLES

LIST OF ABBREVIATIONS

APL	Acute Promyelocytic Leukemia
ATP	Adenosine-5'-triphosphate
CD	Circular Dichroism
CSI	Chemical Shift Index
DLS	Dynamic Light Scattering
DNA	Deoxyribonucleic acid
DNMT	DNA methyl transferase
DTT	Dithiothreitol
EMSA	Electrophoretic Mobility Shift Assay
FGF2	Fibroblast Growth Factor 2
FID	Free Induction Decay
GFP	Green Fluorescent Protein
GST	Glutathione S-transferase
HDAC	Histone Deacetylase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMT	Histone Methyl Transferase
HP-1	Heterochromatin Protein-1
HSQC	Heteronuclear Single Quantum Coherence
IDP	Intrinsically Disordered Protein
IMM	Inner Mitochondrial Membrane
ITC	Isothermal Titration Calorimetry
kDa	kilo Daltons
MAD	Multi-wavelength Anomalous Dispersion
MBD	Methylated C_pG Binding Domain
MBP	Maltose Binding Protein
MCAF1	MBD1 Chromatin Associated Factor 1

MidA	Mitochondrial Dysfunction Protein A
MIR	Multiple Isomorphous Replacement
MMP	Mitochondrial Membrane Potential
MPG	3-methyl Purine DNA Glycosylase
MR	Molecular Replacement
MSA	Multiple Sequence Alignment
mt DNA	mitochondrial DNA
NADH	Nicotinamide Adenine Dinucleotide
NCS	Non Crystallographic Symmetry
NDUFS2	NADH dehydrogenase (ubiquinone) iron-sulfur protein 2
NLS	Nuclear Localization Signal
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
OXPHOS	Oxidative Phosphorylation
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEG	Poly Ethylene Glycol
pI	Isoelctric point
PKMT	Protein Lysine Methyl Transferase
PML	Promyelocytic Leukemia
PP2A	Protein Phosphatase 2A
ppm	parts per million
PRMT	Protein Arginine Methyl Transferase
RARα	Retinoic Acid Receptor α
RC	Respiratory Chain
RF	Radio Frequency
RMSD	Root Mean Square Deviation

RNA	Ribonucleic Acid
ROS	Reactive Oxygen Species
SAD	Single Wavelength Anomalous Dispersion
SAH	S-Adenosyl-L-Homocysteine
SAM	S-Adenosyl-L-Methionine
SEC	Size Exclusion Chromatography
SP-1	Specificity Protein-1
TFE	TriFluoroEthanol
TOCSY	Total Correlation Spectroscopy
TRD	Transcriptional Repressor Domain
TROSY	Transverse Relaxation Optimized Spectroscopy
WT	Wild Type

CHAPTER 1. STRUCTURE DETERMINATION METHODS

1.1 INTRODCUTION

Proteins decide the fate of cells, whether to survive or to die, and this decision will be executed by its 3-dimensional structure. The structure of proteins can be determined by techniques such as Nuclear Magnetic Resonance (NMR), X-ray crystallography and cryo-Electron Microscopy (cryo-EM). The ultimate goal is to decipher the role of these bio-machineries that are present in every cell of all living beings. However, each technique has its own limitations and more importantly it depends on several physical characteristics of proteins. Since the first NMR structure of protein inhibitor IIA from the bull seminal plasma (BUSI IIA, 6 kDa), NMR spectroscopy has emerged as one of the important techniques in solving protein structure. With the ongoing efforts to overcome the protein size barrier, new advances include higher multi-dimensional spectra, selective labeling for much larger proteins and pulse sequences. Since large proteins with more number of NMR-active nuclei, such as ^{1}H and ^{15}N nuclei, tend to overcrowd a 2D spectrum (e.g., ^{1}H-^{15}N HSQC), more dedicated attempts such as deuteration, near-automation and improved dimension spectroscopy fasten and ease peak assignments. Furthermore, NMR spectroscopy is not limited to the flexibility of a folded or unfolded protein.

On the other hand, X-ray crystallography has the advantage of obtaining structures of even bigger proteins. Although X-ray crystallography is not limited by protein size, still several bottlenecks including obtaining diffraction quality crystals impede the flow of structure determination. In recent time, the cryo-EM technique with at resolution of about 3.5 Å for a protein of more than 200 kDa is receiving much

attention. With proper combination of these three techniques, we will be able to study proteins of 50 residues to protein complexes of several kDa.

1.2 BASICS OF NMR SPECTROSCOPY

NMR stands for Nuclear Magnetic Resonance which can be employed for determining the structure of chemical compounds and proteins, and also it is used in other fields such as medicine for imaging. NMR depends on the magnetic nuclei present in the atom, so that it can align toward the magnetic field when applied to it and also it absorbs Radio Frequency (RF) when it is close to its resonance frequency. Thus absorbed Radio Frequency will give signal in NMR data collection and in particular those atoms with spin number other than 0 will absorb RF waves. For example, ^{1}H, ^{15}N, ^{13}C, etc have spin quantum number 1/2 and they are used for solving structures of proteins. For simple chemical compounds such as small molecules, the data obtained from ^{1}H chemical shift can be employed to interpret the structure. But in the meantime, proteins are macromolecules with multiple amino acids and thus the information obtained from 1D NMR, will give only basic idea about protein folding and also there will be a huge overlap of signal between different atoms. Thus advantage of using 3D NMR such as protein labeled with ^{15}N and ^{13}C, could provide us information about the three dimensional structure of the proteins from the peak shift of each amino acid present in the proteins and also the vicinity of each amino acid related to the other amino acid could be obtained, which would help us to study the structures of proteins. Thus NMR can be applied to various fields of chemistry and structural biology to determine the structures and to understand the importance of those structural arrangements.

1.2.1 NMR active nuclei

NMR spectroscopic data of proteins depend on the presence of nuclear spins. Nuclei with odd mass will have spin quantum numbers of half-integral; nuclei of even mass with an odd proton number will have an integral spin quantum number. While the above two are active for NMR, even mass nuclei with an even proton number (e.g., ^{12}C) will have a spin quantum number of zero and are inactive for NMR (Table 1.1).

Table 1.1 Properties of some selected NMR nuclei.

Isotope	**Spin *I***	**Natural abundance [%]**	**Gyromagnetic ratio, (*gamma*) [10^7*rad/(T*s)]**	**Relative sensitivity**	**Absolute sensitivity**
^{1}H	1/2	99.98	26.7519	1.00	1.00
^{2}H	1	0.016	4.1066	$9.65 \cdot 10^{-6}$	$1.45 \cdot 10^{-6}$
^{12}C	0	98.9	--	--	--
^{13}C	1/2	1.108	6.7283	$1.59 \cdot 10^{-2}$	$1.76 \cdot 10^{-4}$
^{14}N	1	99.63	1.9338	$1.01 \cdot 10^{-3}$	$1.01 \cdot 10^{-3}$
^{15}N	1/2	0.37	-2.712	$1.04 \cdot 10^{-3}$	$3.85 \cdot 10^{-6}$
^{16}O	0	98.9	--	--	--
^{17}O	5/2	0.037	-3.6279	$2.91 \cdot 10^{-2}$	$1.08 \cdot 10^{-5}$
^{31}P	1/2	100	10.841	$6.63 \cdot 10^{-2}$	$6.63 \cdot 10^{-2}$

For protein NMR, the nuclei with $I = ½$ (^{1}H, ^{13}C, ^{15}N, ^{19}F and ^{31}P) and the nuclei with $I = 1$ [^{2}H (deuteron)] give more important information. The quadrupolar nuclei with *I*

> ½, have more efficient nuclear spin relaxation and the magnetic state lifespan of it will be shorter and gives broader NMR signal, thus nuclei with I = ½ are usually more preferred and comparatively easier to study.

Nuclei having nonzero spin angular momentum also have nuclear magnetic moments. The Wigner-Eckart theorem is generally defined by equation 1.1:

$$\boldsymbol{\mu} = \gamma \mathbf{I} \qquad (1.1)$$

where $\boldsymbol{\mu}$, the nuclear magnetic moment is co-linear with **I**, the nuclear spin angular momentum vector, for a nucleus characterized by its gyromagnetic ratio γ. In the presence of an external magnetic field, the energies of the spin states of a nucleus are defined by equation 1.2:

$$E = -\boldsymbol{\mu}\mathbf{B} \qquad (1.2)$$

where **B** is the magnetic field vector.

For our study, we are only interested in spins of **I** = ½, e.g., ^{1}H, ^{13}C and ^{15}N with only two nuclear spin states, m = +½ and -½ and two energy levels separated by $\Delta E = \gamma \hbar B_0$. In addition, the sensitivity of NMR signal is proportional to γ^3 and among the three nuclei, ^{1}H has the highest magnetogyric ratio γ, which is four times that of ^{13}C and ten times that of ^{15}N. Since ^{1}H nuclei have the highest γ, the ^{1}H spins, due to dipole-dipole interactions, have the most efficient relaxation, providing Nuclear Overhauser Effects (NOEs), which are useful to determine the internuclear distances for a protein fold. Thus NMR solution structures can also be described as protonic structures.

1.2.2 Radiofrequency and free induction decay

Pulsed NMR are usually employed to disturb the thermal equilibrium state of the bulk magnetic moment of nuclei in a conditional static magnetic field (B_0) to

create an NMR free induction decay (FID) signal. Short radiofrequency (RF) bursts with pulse width in microseconds at first displace this equilibrium magnetization by a flip angle $\theta = \gamma B_1 t_p$ in radians, whereas B_1 is the RF field strength and t_p is the pulse width. In the absence of an RF field, the bulk magnetization of the non-equilibrium nuclei moves over the z-axis (the static field) with an angular Larmor frequency $\omega_0 = - \gamma B_0$ that is detected by a receiving coil. As the nuclear spins restore thermal or Boltzmann equilibrium, the precession releases energy in the form of sinusoidal RF waves or NMR signal. However this NMR signal will decay in the order of seconds, hence the term FID. The change of magnetization with time can be described by two first-order processes, spin-lattice and spin-spin relaxations. Spin-lattice relaxation shows the restoration of the longitudinal magnetization and the spin-spin relaxation shows the decay of the transverse magnetization in the x-y plane.

1.2.3 Phasing

The acquired NMR spectra often have both the absorptive and dispersive signals in the place of the Lorentzian absorptive signal, partly due to delays in pulse sequence between excitation and reception. Hence we have to correct the spectra by phasing two parameters, the zero-order phase, which is frequency independent and the first-order phasing, which is frequency dependent.

1.2.4 Chemical shift

Chemical shift is the result of the nucleus surrounded by anisotropic electronic environment and in general chemical shift is the electronic effect resulting from the static magnetic field and secondary magnetic shielding created by neighboring electrons, and used for residue assignments. This electronic effect is associated with

two phenomena, chemical exchange and chemical shift relaxation. Chemical exchange occurs when a nucleus is made to undergo conformational transition or a new chemical environment such as addition of a ligand and change in the environment parameters suc as pH, temperature, etc. Relaxation is due to the drop of an excited nucleus back to its ground state.

With the help of chemical shift differences, a chemical shift index (CSI) plot, based on deviations from known random coil values of H_α, C_α, H_β and C_β can predict secondary structure patterns in a protein. A combined chemical shift perturbation (CCSP) plot is based on the peak relocation upon addition of a ligand and it can be used to identify perturbed residues in a protein, which is binding to the ligand. Although the chemical shift effect for ^{1}H relaxation is small, until it is placed in a very large static magnetic field, an increase in magnetic field provides improved signal-to-noise ratio and will increase chemical shift relaxation and broaden the resonance of linewidths of nuclei, such as ^{13}C carbonyl and ^{31}P.

1.2.5 Chemical exchange

Chemical exchange occurs due to the exchange of a nucleus between two or more distinct chemical environments in a time-dependent manner. The simple model of such an exchange is interconversion between only two states such as an exchange between ligand-free ↔ ligand-bound states of a protein, monomer ↔ dimer, protonated ↔ deprotonated, or conformations A ↔ B (Fig. 1.1a). A simple two system state for NMR can be defined as the two states having resonance frequencies v_A and v_B and a chemical shift difference $\Delta v = |v_A - v_B|$. The appearance of the corresponding NMR spectrum relies on the populations of both the states, P_A and P_B, and the relative values of the exchange rate $k_{ex} = k_A + k_B$ and the chemical shift

difference Δv. k_{ex} quantifies the average number of stochastic exchange occurrences per unit time. k_{ex} and Δv can be compared directly since they will be having the same units. The effect of varying k_{ex} is shown in the NMR spectrum (Fig. 1.1b).

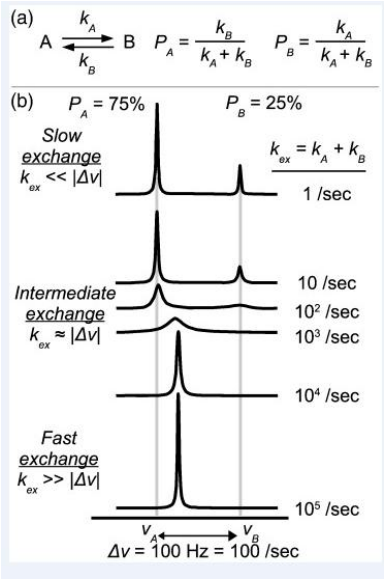


Figure 1.1. Chemical exchanges directly alter the three primary NMR observables. (**a**) Exchange between two states A and B can be described by their rates of departure k_A and k_B and chemical shifts ν_A and ν_B. k_{ex} is the total exchange rate and P_A and P_B are the population fractions of each state ($P_A + P_B = 1$). (**b**) The effect of varying k_{ex} on NMR spectra with P_A = 75% and Δv = 100 Hz. Adapted from: Yang, Zheng *et al*. 2004.

1.2.6 J-coupling

J-coupling is an indirect scalar interaction obtained from hyperfine interactions between nuclei and local electrons which occur between the two nuclear spins. The magnitude of J-coupling increases corresponding to the smaller number of bonds separating the two nuclei. Different amino acid residues can be distinguished based on J-coupling interactions obtained from NMR spectra (e.g., HSQC, HNCO and H(CCO)NH). In addition to that, the Karplus equation relating the 3J-coupling constant to dihedral angles is very useful for defining a protein structure. For

example, the main-chain torsion angle φ from 3JHNHα and side-chain torsion angle χ^1 from $3JH_\alpha H_\beta$ annotates the main chain and side-chain conformations of a peptide. Values of $3JHNH_\alpha$ are also used to index secondary structures in a protein, ranging 8-10 Hz (for β-sheets), 3-5 Hz (for α-helix) and 6-7.5 Hz (for random coil). Hence J-coupling is generally used to define a spin-system and the secondary structure of a protein.

1.2.7 Nuclear overhauser effect

The Nuclear Overhauser Effect (NOE) is the transfer of dipolar cross-relaxations between nuclear spins through space. In steady-state NOE for a two-spin, a low power but specific RF will equalize the average populations (α state = β state) of only I-spin by saturation. In the event of restoring Boltzmann population (α state > β state), I-spin relaxes via W0, W1, I and W2 processes. However, only W0 and W2 relaxations will modify the population difference across α and β energy states of S-spin via cross-relaxation within < 5 – 6 Å, thus creating NOE observed at S-spin (Fig. 1.2).

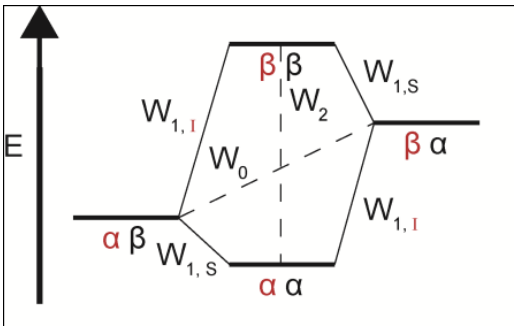


Figure 1.2. Transition processes for two-spin IS system. Energy α and β states for I-spin (red) and S-spin (black). Adapted from: (Johnson and Pinto 2004)

1.2.8 Relaxation

In NMR spectroscopy, the term "relaxation" can be defined as the return of nuclear magnetization to equilibrium. For an isotropic system (there is no chemical exchange), the relaxation process is typically represented by two independent and different rate constants, the longitudinal relaxation rate (R_1) and transverse relaxation rate (R_2) or their reciprocal relaxation time T_1 ($T_1 = R_1^{-1}$) and T_2 ($T_2 = R_2^{-1}$), respectively. For small molecules, the relaxation rates are measured using the 1D method but for proteins, they are determined on the same principle but using the 2D HSQC-based approach.

The T_1 (or spin-lattice relaxation time) is measured along the z-axis employing its decay with the inversion recovery (IR) method. For IR pulse sequence T_d-$180_x°$-τ-$90_x°$-FID, the time to come back to ground state of the excited magnetization along the z-axis back to equilibrium is defined by Equation 1.3:

$$M_z(t) = M_z(\infty)(1-2e^{-t/T_1}) \qquad (1.3)$$

T_1 is then measured by using a series of interval τ values, where $T_1 = \tau/\ln 2$ at node intensity, $M_z(\tau) = 0$ when $T_d > 5T_1$ interval.

The T_2 (or spin-spin relaxation time) is measured by the J-coupling of spins with various Larmor frequencies but can be precisely measured for heteronuclear spins (e.g., ^{15}N) using spin-echo variant experiments [e.g., Carr-Purcell-Meiboom-Gill (CPMG)]. The rate of recovery is similar to T_1 in an exponential form but the amplitude of magnetization reduces quickly in the x-y plane than the z-direction. T_2 is also the $T_{1\rho}$ (rotational relaxation time) in the absence of chemical exchange or anisotropy. Generally an unfolded protein has longer T_2 and it will be shorter for larger proteins. For a simple spin-echo pulse sequence T_d-$90_x°$-τ-$180_y°$-τ–FID, the decay of magnetization is defined in the x-y plane by Equation 1.4:

$$M_{xy}(t) = M_{xy}(0)e^{-t/T_2} \quad (1.4)$$

where t is the total evolution time in the x-y plane and M(0) is the initial magnetization.

For pulse sequence, $90_x°$-τ (spin-lock)-FID, $T_{1\rho}$ is obtained by fitting Equation 1.5:

$$M = M(\infty) + (M(0) - M(\infty))e^{\tau/T_{1\rho}} \quad (1.5)$$

where τ is the spin-lock duration by a continuous wave to irradiate the spins in the rotating frame and monitor its decay in a longitudinal-like manner.

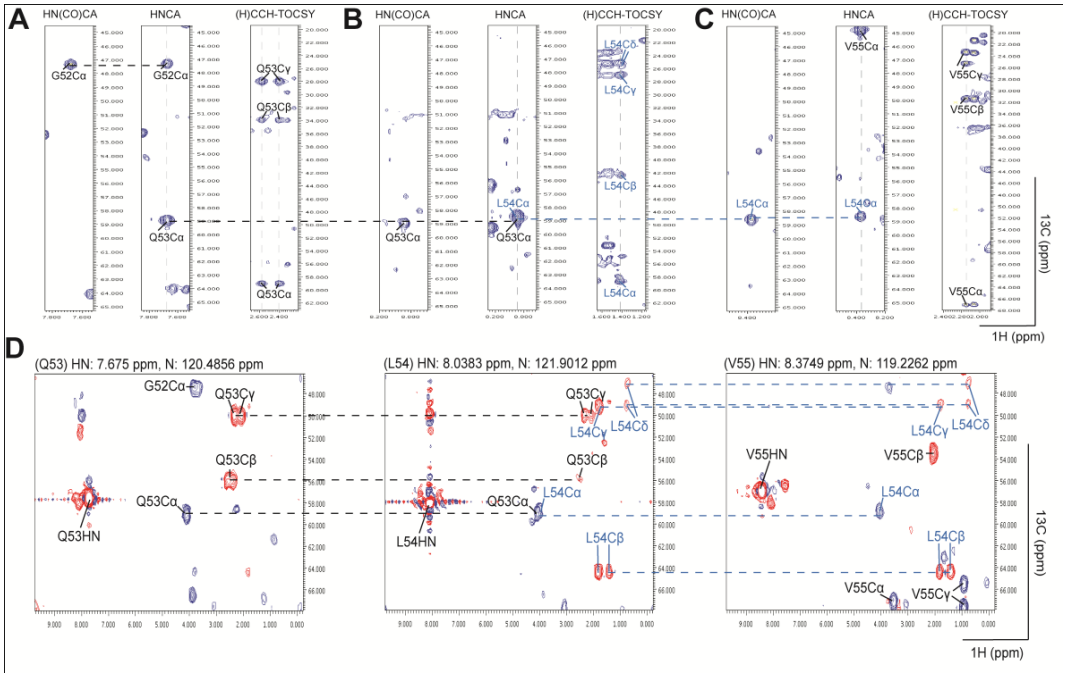
Heteronuclear NOE is another relaxation experiment via cross-relaxation where transfer of magnetization occurs from a proton to a heteronucleus. For example, ^{1}H-^{15}N heteronuclear NOE is calculated as the ratio of peak intensities in spectra recorded with and without proton saturation to identify flexible regions shown by decreased NOE intensities in backbone amides of a protein. Steady-state NOE value for each amide residue is calculated using Equation 1.6:

$$NOE = I_{saturated}/I_{equilibrium} \quad (1.6)$$

1.2.9 Structure determination

If we intend to solve a protein structure, a number of NMR spectra have to be acquired from a powerful NMR spectrometer using a typical ^{15}N, ^{13}C double-labeled protein sample. For our proteins, traditional 2D ^{1}H-^{15}N HSQC and 3D HNCA, 3D HNCOCA and 3D MQ-(H)CCH-TOCSY were collected for back-bone assignments. Instead of 3D ^{15}N or ^{13}C NOESY (Yang, Zheng et al. 2004), have taken an additional approach to better distinguish overcrowded peaks using ^{15}N, ^{13}C-time shared 4D NOESY without deuteration. Unlike 3D NOESY, 4D NOESY offers a fourth dimension to resolve ambiguous peaks due to overlapping in a better way. This is particularly helpful for proteins with poorly resolved peaks. Acquired NMR data are

then processed using NMRpipe (Xu, Zheng et al. 2006) and later analyzed using in-house NMRspy program (yangdw.science.nus.edu.sg/Software& Scripts/XYZ4D/index.htm). For a 3D (or higher dimensional) spectrum, it can be described as a 3D (or higher facet) cube composed of many 2D slices stacked together. In a conventional 2D ^{1}H-^{15}N HSQC, all peaks are from backbone amides NH (except proline, due to missing NH) and other peaks are from NH side-chains of Asn, Arg, Gln, Lys, Trp, etc. The first step is to match a peak to one amide residue. To attain the backbone assignment of a protein, 3D HNCA, 3D HN(CO)CA, 3D CCH-TOCSY and complemented with 4D ^{13}C, ^{15}N-edited sub-NOESY are used to identify intra-residue and sequential spin-systems which are related with each amide (Fig. 1.3). In 3D TOCSY spectrum, the resonances in all spins in a spin system of a residue will be seen (Fig. 1.3 a-c). For example in a 3D (H)CCH-TOCSY slice, one will see three ^{13}C peaks for Glu (C_α, C_β and C_γ) and five ^{13}C peaks for Leu (C_α, C_β, C_γ $C_{\delta1}$ and $C_{\delta2}$). After identifying the spin systems, the backbone is connected by matching sequential and intra-residue spin-systems based on the 4D ^{13}C, ^{15}N-edited NOESY sub-spectrum (Fig. 1.3d). On the basis of selective sequence assignments, NOEs arising from protons separated by 1.8 - 5.5 Å are also assigned from the ^{13}C, ^{15}N-edited, ^{13}C, ^{13}C-edited and ^{15}N, ^{15}N-edited sub-NOESY spectra. The NOEs are extracted as intensities or volumes and can be classified into three (or more) distance restraint classes: long (1.8 - 5.5 Å), medium (1.8 - 3.4 Å) and short range (1.8 - 2.4 Å). A lower bound of 1.8 Å, an approximate van der Waals contact for all restraint classes is set to account for nuclei, which are very close but having unobservable NOEs. The protonic structure is then calculated using Cyana 2.1 (Herrmann, Güntert et al. 2002) based on a series of dihedral and distance restraints and reassignment of ambiguous NOEs till minimal structural violations.



The 3D structure of a protein defines its function. We are interested in linking the structure of a protein to its function using several techniques including site-directed mutagenesis, chemical shift perturbation, NMR ligand titration and cell based study. They will be discussed in more details later for the protein of interest.

1.3 BASICS OF X-RAY CRYSTALLOGRAPHY

X-ray crystallography is one of the two most important techniques to determine 3D structures of proteins and chemical compounds. In crystallography, the most important requirement is to get a crystal that diffracts, so that we could gather atomic resolution information about the molecule we are analyzing. Chemical compounds, which are in general easy to crystallize, diffract to high resolution. However, obtaining crystals for macromolecules, such as proteins, DNA or protein-DNA complexes could be challenging, where we have to optimize crystallization conditions to obtain a good crystal that diffracts to a good resolution of at least better than 3 Å. X-ray crystallography has certain advantages over the other structure

determination method, nuclear magnetic resonance (NMR). NMR could be used for only small proteins of size up to 35 kDa, whereas X-ray crystallography has been used to solve even the structure of the whole ribosome (Weixlbaumer, Petry et al. 2007).

Protein crystals are formed when an aggregation of molecules forming a particularly ordered self arrangement is formed such that each molecule could be related to others following certain symmetry operations. In other words, a crystal is an anisotropic, homogeneous body possessing a three-dimensional periodic ordering of atoms, ions or molecules. The orderliness and diffraction of these crystals depend upon the intermolecular forces among the molecules. A well ordered crystal would need proper packing of molecules in a limited space (Helliwell 2005). When X-rays impinge on a well formed crystal, they are diffracted to give a diffraction pattern. In the place of lens in light microscope, we use computer to generate the arrangement of molecules from diffraction images.

1.3.1. Unit-cell

In a crystalline state, solute molecules will adopt one among a limited number of orientations, when they start to arrange in an ordered manner. An ordered arrangement will follow a fixed orientation of each molecule in relation to one another. Therefore, it is possible to divide a crystal into a minimum portion, which, when repeated multiple times, will generate the whole crystal. This minimal building block is known as the unit-cell. The unit-cell has three non coplanar lattice vectors a, b and c, with interaxial angles α, β and γ (Fig 1.4).

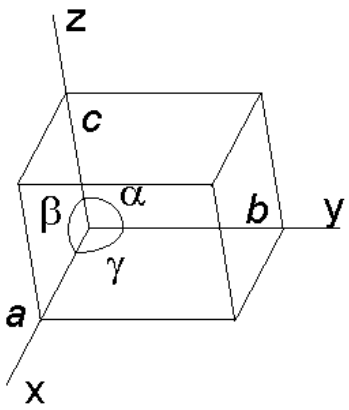


Figure 1.4. The unit-cell is the repeating building block which forms a crystal when repeated several times in accordance to the symmetry (Adapted from: http://xrayweb2.chem.ou.edu/notes/symmetry.html).

The magnitude of the lattice vectors a, b and c along with the angles α, β and γ are known as the unit-cell parameters. Based on the unit-cell parameters, every crystal can be divided into one among the following 7 crystal systems (Fig. 1.5). Based on this classification, one step further would allow the unit-cells to form interlaced arrangement (known as unit-cell centering) and the 7 crystal systems form 14 Bravais lattices (Fig. 1.6), named after the French physicist Auguste Bravais, who first proposed it in 1845.

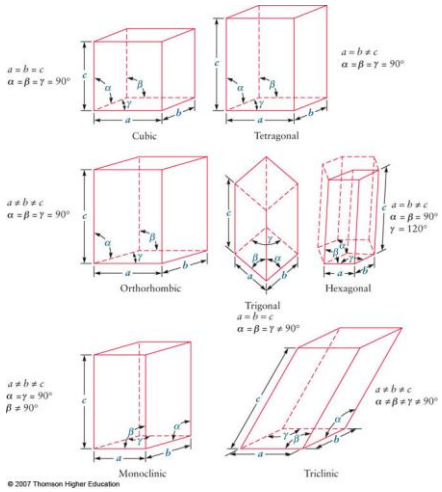


Figure 1.5. The seven crystal systems (http://nano-physics.pbworks.com/w/page/ 12296663/Group%202).

1.3.2 Point group and space group

The spatial relationship between the related neighbor motifs is called symmetry operations. The three basic symmetry operations followed by crystals are: inversion, reflection and rotation. The allowable combinations of these crystallographic symmetry operations in a primitive unit-cell of the 7 crystal systems are known as the crystallographic point groups. There are 32 point groups. When the basic symmetry elements rotation and reflection as well as the point groups are subject to translation and combined with the Bravais lattices, 230 possible types of crystallographic packing, known as space groups, are generated. Of these 230 space groups, many follow the inclusion of enantiomeric motifs. The reason that all proteins are made up of only L-amino acids and do not have the enantiomeric D-amino acids, will allow them to crystallize only in 65 chiral space groups (Hahn, 2006).

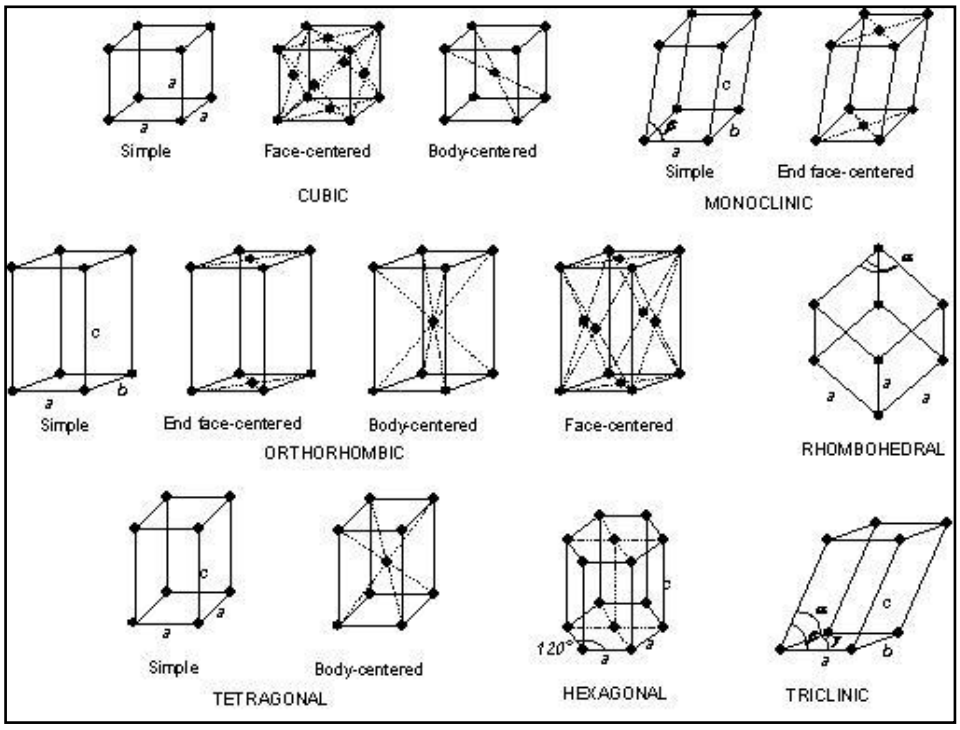


Figure 1.6. The 14 Bravais lattices of crystals (adapted from: http://www.askiitians.com/iit-jee-solid-state/bravais-lattices/).

1.3.3 Bragg's law

Bragg's law is a condition for a crystal to diffract, when electromagnetic radiation, of a wavelength that is comparable to atomic spacing, impinges upon a crystalline sample, scattered and undergoes constructive interference. For a crystalline solid, W.L. Bragg showed that a set of parallel planes with indices *hkl* and interplanar spacing d_{hkl} produces a diffracted beam when X-rays of wavelength λ impinge upon the planes at an angle θ and reflected, also at the same angle, only if θ meets the following condition

$$2d_{hkl}\, sin\theta = n\lambda \qquad (1.7)$$

where n is an integer (Rhodes, 2000).

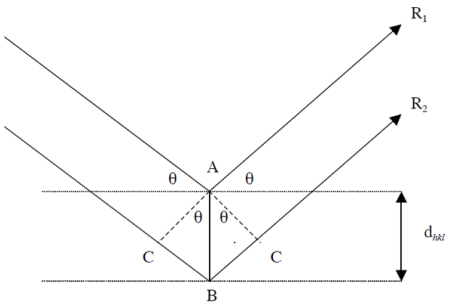


Figure 1.7. Bragg's law. Bragg's law defines the condition for diffraction (Rhodes, 2000).

The two planes (Fig. 1.7) represent two parallel rows of lattice points with interplanar spacing d_{hkl}. Two rays R_1 and R_2 are reflected from them at an angle θ. Lines AC are drawn from the point of reflection A of R_1, perpendicular to ray R_2. If ray R_2 is reflected at B, then the diagram shows that R_2 travels the same distance as wave R_1 plus an additional distance of 2BC. Since AB in triangle ABC, is perpendicular to the atomic plane and AC is perpendicular to the incident ray, angle CAB equals θ, the angle of incidence (two angles are equal if corresponding sides are perpendicular). Since ABC is a right-angled triangle, the sine of angle θ is BC/AB or BC/d_{hkl}. Thus

BC equals d_{hkl} sin θ, and the additional distance 2BC traveled by ray R_2 is $2d_{hkl}$ sinθ. If this difference in path length for rays reflected from successive planes is equal to an integral number of wavelengths of the incident X-rays (satisfying equation 1.7), then the rays are in phase with each other, resulting in constructive interference to produce a strong diffracted beam.

1.3.4 Reciprocal lattice and Ewald sphere

The concept of reciprocal lattice simplifies the mathematical calculation and illustration of diffraction which is obtained through Bragg's reflections and crystal geometry. A crystal, or the array of unit-cells, can be reduced to a set of repeating lattice points. The whole 3-dimentional lattice can be sliced using a set of parallel imaginary planes in a periodic fashion. The planes are then named as h, k, l, known as "Miller indices", which are the integral number of parts into which the plane divides each unit-cell parameter a, b and c, respectively. Therefore, an infinite set of planes can be made through the real-space lattice, with each set having a unique value for h, k and l. The perpendicular shortest distance between two adjacent parallel planes is then represented by the vector 'd_{hkl}' and is called the interplanar distance. To convert this real space lattice into a reciprocal lattice, a lattice point should be drawn from each point in the real-space at a distance of $1/d_{hkl}$ (represented as d^*_{hkl}) and normal to the *hkl* plane. The set of points drawn in this manner form the reciprocal lattice. The reciprocal lattice was first used by physicist Paul Ewald to give the general geometrical interpretation of diffraction, known as the Ewald sphere (Fig. 1.8) (Ewald, 1969).

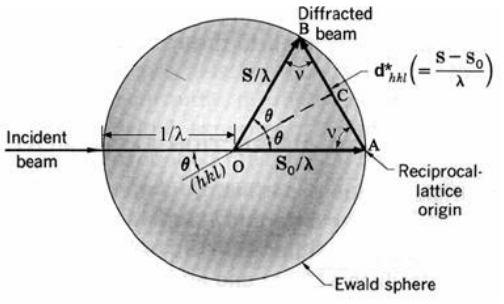


Figure 1.8. Ewald's Sphere. X-ray diffraction pattern due to constructive interference by Ewald (figure adapted from: Ewald,1969)

If a sphere of radius 'r' = 1/λ is drawn and the incident beam of X-rays hits the real space plane (hkl) at 'O' at an angle 'θ' and gets reflected making an equivalent angle COB then AC = sinθ/λ. By geometry, AC=BC, therefore, AB=2AC. Now, as shown above, 2AC = 2(sinθ/λ). According to Bragg's law, 2(sinθ/λ) = 1/d_{hkl}. As described earlier, 1/d_{hkl} is equal to d^*_{hkl} (Ewald, 1969). Therefore, point 'B' can be assumed to be a point in the reciprocal lattice and if it coincides with the Ewald sphere then it satisfies the Bragg's law.

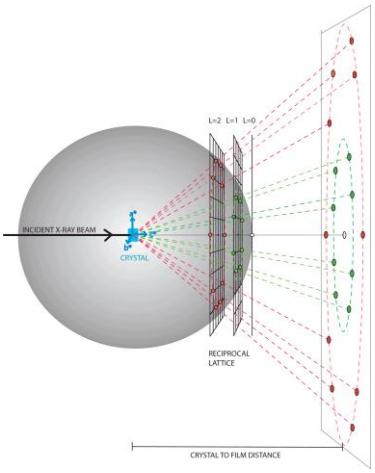


Figure 1.9. X-ray Diffraction by Ewald. In geometric interpretation, a crystal can be assumed to be at the centre of the Ewald sphere. The detector is placed behind the crystal to record diffraction spots (adapted from: people.mbi.ucla.edu/sawaya/m230d/Data/data.html).

According to the Ewald construction, diffraction from a hkl plane will only occur when the corresponding d^*_{hkl} lies on the Ewald sphere. Therefore, upon exposing a crystal to X-rays, only a few of the reciprocal lattice points will be sampled (Rupp, 2009). The crystal is, hence, rotated so that the reciprocal lattice also rotates and more points intersect with the Ewald sphere and are recorded in the detector (Fig. 1.9).

1.3.5 Structure factor

In a diffraction experiment, using Bragg's law and the Ewald construction, it is possible to determine where the spot from each hkl plane will get collected on a detector with known geometry. The real space lattice also could be obtained from the corresponding reciprocal lattice. However, the real space lattice obtained in this manner will give only a unit-cell with no information about its molecular contents. In order to learn about the molecular contents of a unit-cell, the contribution of each atom in the crystal towards each observed Bragg reflection needs to be calculated. In other words, the total contribution of all atoms in the direction of a particular hkl plane is known as the structure factor of that plane. To do this, each observed reflection is deconvoluted into the partial contribution of the individual atomic scattering factors (of all atoms in the unit-cell) in the direction of the observed reflection. The deconvolution into partial contribution of individual scattering factors is performed with the aid of the mathematical concept of complex numbers.

1.3.5.1 Wave as a complex number

Working with scattering of waves can be simplified with the use of complex numbers. A complex number can be represented as 'a + ib' where 'a' and 'b' are the 'real' component and 'i' is the imaginary component. In the complex number system,

the contribution of the imaginary component lies in an orthogonal direction to the real number. The complex number a+ib would then represent ‘a’ units in the x direction (real axis) and b units in the y direction (imaginary axis), Fig. 1.10

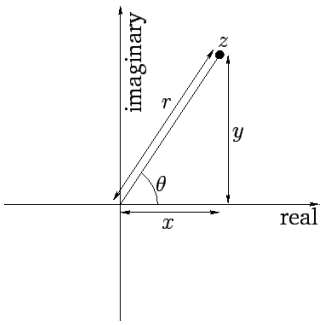


Figure 1.10. Wave as a complex number. Adapted from: http://farside.ph.utexas.edu/teaching/315/Waves/ node72.htm).

Waves can be conveniently represented using this system. The amplitude of a wave can be represented by ‘r’ and ‘θ’ can be used to represent the phase angle of the wave, (Fig. 1.10). This ‘a+ib’ representation is called the rectangular form of the wave and can be used to calculate the properties of waves.

For instance, the electrons present in the (protein) atoms scatters the incoming X-rays and the wavelength of the X-rays is about 1.5 to 0.5 Å, which is of the order of the atom diameter, and most of the scattering is in the forward direction. It is also known that the X-ray scattering power will correspond to the number of electrons present in the particular atom. The X-ray scattering power of an atom decreases with increasing scattering angle and is higher for heavier atoms.

1.3.5.2 Electron density

As defined earlier, each diffraction spot on the detector is the sum of all scattering factor of all atoms in the crystal in the direction hkl and given as F_{hkl} (Eq. 1.8).

$$\mathbf{F}_{hkl} = \sum_{j=1}^{n} f_j \, e^{2\pi i \, (hx_j + ky_j + lz_j)} \quad (1.8)$$

where 'f_j' represents the atomic scattering factor of atom 'j', hkl represent the Miller indices of the plane that is forming the Bragg reflection while x, y and z represent the fractional coordinates of atom j in the unit-cell. The atomic scattering factor of an atom decays with sinθ/λ. From the structure factors of all diffraction spots in an experiment, it is now desirable to compute the electron density at each point within the unit-cell which provides us a picture of the electron cloud of each atom, thereby locating the atom in the unit-cell. The relationship between the reciprocal space structure factors to the real space electron density is explained by the mathematical formulation of Fourier transformation, developed by the French mathematician Jean Baptiste Joseph Fourier. Using the principles of Fourier analysis, the calculation of electron density is a reverse Fourier transform of structure factors (Eq. 1.9), where V= volume of the entire unit cell.

$$\rho\,(x, y, z) = \frac{1}{\mathrm{V}} \sum_{h} \sum_{k} \sum_{l} \mathbf{F}_{hkl} \, e^{-2\pi i \, (h_x + k_y + l_z)} \quad (1.9)$$

To visualize the 3-dimentional structure of a protein, it is important to determine the location of each atom in the unit-cell of a protein crystal. As is evident from the above two equations 1.8 and 1.9, the problem of calculating the location of each atom in the unit-cell depends on the structure factor of as much reflections as possible, where the structure factor can be calculated by summing up the atomic scattering contribution, which in turn depends on the position of the atoms in the unit-cell. Thus the required solution of a problem depends on the problem itself. A technological inability limits the complete determination of experimental structure factors, due to the fact that structure factors are composed of experimentally measurable structure amplitude as well as immeasurable phase angles of the

diffraction waves. This phase problem of crystallography, in both small molecular and macromolecular, therefore, is reduced to the problem of accurately estimating the phase angles of diffracted reflections, with respect to the undiffracted ray at the origin.

1.3.6 Methods to solve the phase problem

As it is not possible to obtain the phase angles experimentally for all the spots recorded on the detector, it is possible to estimate and derive them indirectly or directly from the reflections. The method or strategy chosen for solving the phase problem depends on some considerations. In spite, all methods help us obtain enough initial phase angles to enable the development of interpretable electron density maps. Once we attain interpretable electron density map, a partial model of the protein can be built which then have to be refined to optimize selected parameters iteratively. Additional phases are then derived from the partial model to further improve the electron density maps and build more regions of the structure. Statistical parameters obtained from the model are analyzed at each step of model building and refinement to verify conformity with the experimentally obtained data. Iterative cycles of model building and refinement, while comparing the conformity with experimentally obtained data, therefore, gradually improves the electron density maps to yield the final structure. The final structure is, hence, a model whose conformity has adequately been refined using the experimental data.

1.3.6.1. Multi-wavelength anomalous dispersion (MAD)

Phases can be extracted by using the property of elements to absorb X-rays at a particular wavelength. The absorption of X-rays by atoms drops suddenly close to a characteristic wavelength(s), known as the absorption edge(s) of the element. When

data are collected at the absorption edge, Freidel's rule does not apply and $I_{h,k,l} \neq I_{-h,-k,-l}$. Certain atoms, such as selenium and sulphur, absorb at particular wavelengths that can be easily generated in today's synchrotron radiation facilities and can be effectively used in labeling a crystal for solving a structure with the MAD method. The scattering at the absorption edge is called anomalous scattering and can be characterized as,

$$F_{HP}^{\lambda 2} = F_{H}^{\lambda 1} + \Delta F_r + \Delta F_i \quad (1.10)$$

where, ΔF_r and ΔF_i are the real and imaginary components of the scattering contribution. $F_{HP}^{\lambda 2}$ and $F_{H}^{\lambda 1}$ are the structure factors of the heavy atom derivatized protein crystal at a wavelength where it does and does not show anomalous scattering respectively. Freidel pairs with superscripts + and – are equal in the absence of anomalous scattering $|F|^+ = |F|^-$ while $\alpha_{hkl}{}^+ = -\alpha_{hkl}{}^-$. It can also be shown that in Eq. 1.10, $\Delta F_r{}^+$ and $\Delta F_r{}^-$ are mirror images of each other at the anomalous wavelength. In addition, the imaginary contributions are inverted reflections of each other, i.e., $\Delta Fi+$ are a reflection of $\Delta F_i{}^-$ with the direction reversed. A Patterson function (detailed in the following part) can be used to locate the heavy atoms in the unit-cell and obtain the complete values of ΔF_r and ΔF_i. In order to break the phase ambiguity, data from another wavelength is needed, where phase ambiguity is better illustrated in the following Harker construction (Fig. 1.11). The two phases which are possibly present, where the circles intersect. The problem then arises as to which phase to choose. This requires a consideration of phase probabilities (Taylor 2003). To circumvent it, help is taken from the fact that the real and imaginary anomalous scattering factors, ΔF_r and ΔF_i, vary greatly at the absorption edge. Close to the absorption edge ΔF_r reaches a minimum value and as it begins to ascend, ΔF_i reaches a maximum value. These two wavelengths, therefore, can be used for exploiting the anomalous behavior while a

wavelength far from the absorption edge can be used for exploiting the non-anomalous behavior of atoms during diffraction.

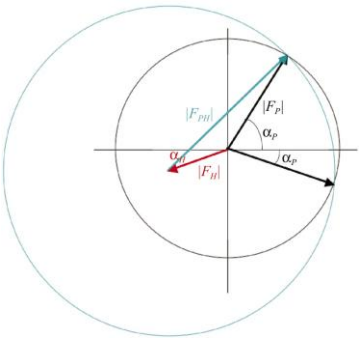


Figure 1.11. Harker Construction. Harker Construction for Single Isomorphous Replacement (SIR). Figure Adapted from (Taylor 2003)

1.3.6.2 Multiple isomorphous replacement (MIR)

In this method, a crystal is soaked in a heavy atom solution or a protein is co-crystallized with a heavy atom. The addition of the heavy atom must not affect the crystal formation, protein conformation or unit-cell dimensions, in comparison to its native form. In other words, the native and the derivative crystals should be truly isomorphic (Taylor 2003). Data sets from native and heavy-atom derivative crystals are collected. Interpretation of difference Patterson maps will reveal the locations of heavy atoms in the unit-cell, from which the phases of reflections can be computed. Since the structure factor of the heavy atom derivative ($\mathbf{F}_{ph}$) is the vector sum of the lone heavy atom ($\mathbf{F}_h$) and the native crystal ($\mathbf{F}_p$), the phase of the native $\mathbf{F}_p$ can be solved geometrically from the equation

$$F_{ph} = F_p + F_h \qquad (1.11)$$

or,

$$F_p = F_{ph} - F_h \qquad (1.12)$$

1.3.6.3 Patterson function

The Patterson function is crucial for determining the sub-structure of heavy atoms and is analogous to Eq. 1.9, which is used to calculate electron density from structure factors, albeit with a variation. It is a Fourier series that uses only the amplitude of the structure factors and does not use the phase (Eq. 1.12). As the structure factor amplitudes can be directly derived from spot intensities, no additional information is required for calculating a Patterson map.

$$\mathrm{P}\,(u,v,w) = \frac{1}{\mathrm{V}} \sum_{h} \sum_{k} \sum_{l} |\mathbf{F}_{hkl}^{2}| e^{-2\pi i\,(hu+kv+lw)} \quad (1.12)$$

A Patterson map constructed using Eq. 1.12 produces peaks at the heads of vectors drawn from every atom to all atoms, including itself, in the unit-cell. The peaks in the Patterson map are also referred to as Patterson atoms. Therefore, if there are n atoms in a unit-cell, then there would be n (n-1) Patterson atoms in the Patterson unit-cell. The enormity of the number of peaks makes interpretation of the Patterson map extremely difficult. However, interpretation can be simplified if the Patterson map has been calculated from only a few atoms. Hence, the Patterson map of only the heavy atoms is calculated in the MIR method. To calculate the Patterson map of the heavy atoms, a difference Patterson function is used. It can be shown that the vectors between atoms related by symmetry have one or two constant coordinates. Once the Patterson atoms have been obtained for the heavy atoms using the difference Patterson function, the real heavy atom positions in the unit-cell can be calculated largely by a procedure proposed by Harker, known as Harker sectioning, which uses the space group information.

1.3.6.4 Molecular replacement

The MIR and MAD methods are powerful techniques to obtain experimental phases for structure determination. However, both methods suffer from the major drawback of having the necessity to label the protein with heavy atoms and requiring more than one dataset. As more and more protein structures are being solved, another technique, called molecular replacement (MR), is gaining attention. MR makes use of the fact that if the structure of a target protein is likely to be similar to an already known structure, then the phases from the known structure will be a close approximation of the actual phases from the target structure. Proteins that share high degree of sequence similarity also tend to have significant structural similarity. Therefore, the calculated phases from a model structure can be combined with the measured intensities of the target protein to obtain interpretable electron density maps. After repeated cycles of iterative model building and refinement, the phases of the target structure will become more accurate.

In order to obtain meaningful phases from the phasing model, it is required to orient it similar to the target protein's orientation in the unit-cell. In the absence of the structure of the target protein, the correct orientation to choose is unknown. Space group considerations place some restrictions on the orientation of the phasing model inside the unit-cell of target protein. However, the number of possible orientations is still very large and extremely computer intensive for a brute force search. The criteria used for search is that the calculated structure factor intensities from the phasing model at a particular location should be similar to the observed structure factor intensities from a diffraction experiment. For a brute force approach to search for the best orientation, the phasing model needs to be rotated around and translated along the 3 unit-cell axes, producing 6 parameters for optimization. Exhaustively trying

different values for these 6 parameters at random is an extremely computer intensive task as the number of possible combinations will be extremely large. However, the task can be made less intensive if the rotation and translational searches are performed separately (Rossmann and Blow, 1962). In order to enable separate searches for the rotation and translation of the phasing model, the Patterson function is used. The unique property of the Patterson function is that it is the same for a phasing model with the same rotation axis values, irrespective of the translation axis values. This implies that two models will give the same Patterson function wherever they are located in the unit-cell as long as they have the same rotation values. Using this feature of the Patterson function, the correct values for the rotation axis can be estimated. Once these values are known, they can be applied along with different values of the translational axis to calculate structure factor intensities. These calculated values are then matched with experimental values to arrive at optimal translational axis values.

1.3.7 Refinement

The final model of a protein molecule is the result of appropriate refinement of an initial model. Model refinement refers to optimally fitting the atomic positions and temperature factors to the experimental data. In other words, the residues (or atoms) of the molecule are fit to electron density maps and their geometry is regularized. Since collecting data with Miller indices (hkl) $-\infty$ to $+\infty$ (where d_{hkl} approaches 0), which is a pre-requisite for an error free electron density map, is impractical, a model is fit to electron density maps at the available resolution of the experimental data. As the model keeps improving during the iterative cycles of model building and refinement, more accurate phases are available, which, in-turn, improve the electron

density map. As the electron density maps improve, more accurate phases are generated. The overall fit between the model and the experimental data is measured using the R-factor, Eq. 1.13, where F_{obs} = experimentally determined structure factors and F_{calc}= structure factors calculated from the target structure.

$$R = \frac{\sum | F_{obs} - F_{calc}|}{\sum F_{obs}} \qquad (1.13)$$

Traditionally, least-squares (LSQ) refinement has been used for optimizing model parameters. However, modern refinement programs attempt to minimize maximum likelihood (ML) functions or the energy of the crystal system, which tend to give better refinement as they are less susceptible to non-random errors and model bias.

CHAPTER 2. TRANSCRIPTIONAL REPRESSOR DOMAIN (TRD) OF MBD1

In this part of the thesis, I present the structure and function of the transcriptional repressor domain (TRD) of the methylated DNA binding protein MBD1. The structure was determined using NMR spectroscopy and based on that, functional studies were undertaken to elucidate the mechanism of target recognition in epigenetic gene silencing.

2.1 EPIGENETICS

Epigenetics is the study of hereditary inheritance of changes in gene function that happens without any change in the DNA sequence (Wolffe and Matzke 1999). Such modification allows the transfer of information between dividing cells, although such information is not present in the nucleotide sequence of the gene. For example when a bone cell divides, the daughter cells will express only those genes required for bone formation and other genes corresponding to other organ such as liver cells remains silent.

Epigenetics controls the regulation of gene expression through DNA methylation, chromatin modification and posttranslational changes of proteins, for example histones and transcriptional machineries, (Fig. 2.1) **(Nakao 2001)**. A transcriptionally active condition can be inferred from DNA hypomethylation, hyperacetylation of histones, and active chromatin. At the same time, repression of gene transcription occurs when there is DNA hypermethylation, deacetylation of histones and inactive chromatin. Furthermore, lysine 4 of histone H3 (H3K4) methylation is correlated with transcriptional activation while transcriptional

repression is marked by the methylation of lysine 9 of histone H3 (H3K9) (Fischle, Wang et al. 2003). Transcriptional activation is carried out by binding of transcription factors, which bind to specific DNA sequences and then recruit coactivators such as histone acetyltransferases and H3K4 methyltransferases. In the same way, methylated DNA-binding proteins and transcriptional repressors will recruit corepressors, such as histone deacetylases (HDACs) and H3K9 methyltransferases (Fig. 2.1). In addition to that, methylated and acetylated histone surfaces recruit additional factors, such as heterochromatin protein-1 (HP1), which specifically binds methyl-H3K9 (Nakao, Minami et al. 2004). This epigenetic mechanism is essential for normal cell division and differentiation, such as expression of a particular cell specific gene. In contrast to the normal physiological conditions, it is abnormally regulated in cancer cells (Chen, Pettersson et al. 1998).

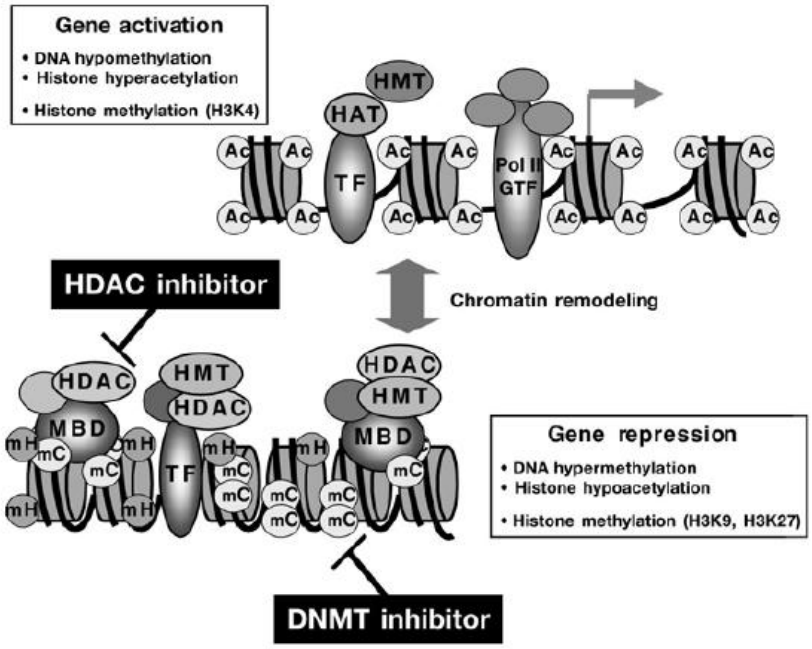


Figure 2.1. Epigenetics and chromatin. Gene activation marked by DNA hypomethylation, histone acetylation and binding of activators and Repression is marked by DNA methylation, histone deacetylation and binding of repressors (Nakao, Minami et al. 2004).

2.2 DNA METHYLATION

One of the major covalent modifications of eukaryotic DNA is the methylation of the palindromic C_pG sequence in a symmetrical manner (SINSHEIMER 1955). Scarano (Scarano 1971) first proposed DNA methylation Holliday and Pugh (Holliday and Pugh 1975) and Riggs (Riggs 1975) showed its involvement in cell differentiation and development. C_pG rich regions, found in mammalian gene promoters and known as "C_pG islands", are generally unmethylated and the remaining 70-80% of mammalian C_pG sites are usually methylated (Ehrlich and Wang 1981). C_pG islands are present in the promoter regions of all constitutively expressed genes (Zhu, He et al. 2008). However, not all C_pG islands are non-methylated and some of them that are involved in X-inactivation and genomic imprinting remain methylated depending on cell requirement (Edwards and Ferguson-Smith 2007). DNA methyl transferase (DNMT1) catalyses the methylation of the naked cytosine in a hemi-methylated C_pG (Fournier, Sasai et al. 2012). This process of DNA methylation, resulting in gene repression, will be inherited from mother to daughter in active cells (Bird and Wolffe 1999; Bird 2002; Edwards and Ferguson-Smith 2007).

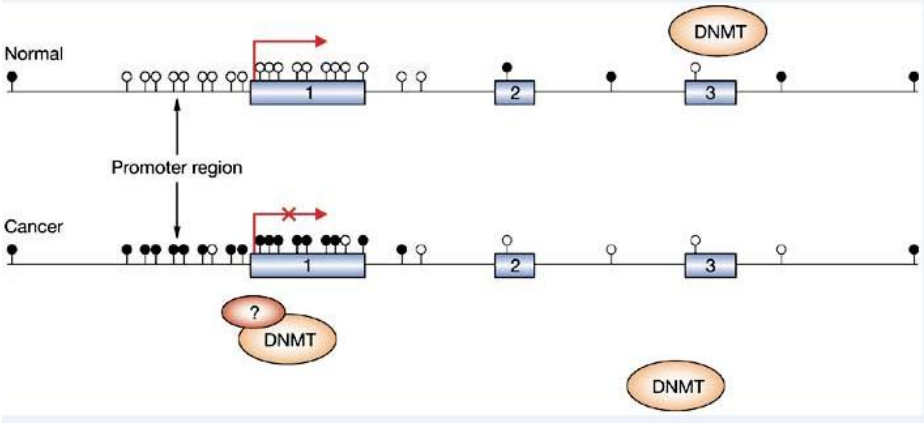


Figure 2.2. Gene silencing due to promoter methylation. The promoters of tumor suppressor genes get methylated in cancer cells and thus transcriptional activators are unable to bind for transcribing the gene (Baylin 2005).

Aberrant DNA methylation could result in severe disorders in the cell. Hypomethylation could lead to inappropriate expression of genes while hypermethylation could result in gene repression, (Fig. 2.2) (Costello, Frühwald et al. 2000; Esteller, Corn et al. 2001). Methylation of C_pG islands occurs in various cancers, thereby inducing silencing of tumor suppressor genes, like *Cdh1* in breast, bladder and prostate cancers (Graff, Herman et al. 1995), and the Rb gene in retinoblastomas (Stirzaker, Millar et al. 1997). This aberrant methylation generally occurs in cancer cells and there are ways to control this by administering DNA methyl transferase inhibitors such as 5-aza-2'-deoxycytidine or knock down by using antisense or siRNA. Results from the human colorectal cancer cell line HCT116 confirm the re-expression of the tumor suppressor gene Cdkn2a and provide a direct link between the methylation of its promoter and its transcription (Fuks, Burgers et al. 2000; Geiman, Sankpal et al. 2004). At the same time, DNMT1, DNMT3a or DNMT3b inhibition in a non-specific manner could result in aberrant expression of several genes, which are supposedly silenced in the normal cell, and could affect the physiological function of the cell.

DNA methylation plays an important role in genomic imprinting, the strategy by which a cell controls the expression of certain genes in the genome according to their parent of origin (Price, van der Leij et al. 2002). Thus the DNA methylation state plays a critical role in the transcriptional repression of genes, chromatin structure organization and embryo development while aberrant DNA methylation could result in diseases, including cancer and developmental disorders.

2.3 METHYLATED C_PG DNA BINDING PROTEINS

Identification and characterization of proteins that bind to methylated DNA in a specific manner is of major importance. With the help of electrophoretic mobility shift assay (EMSA), the binding of the MeCP1 protein, an MBD family member, to C_pG methylated DNA was observed in the nuclear extracts of a variety of mammalian tissues and cell lines (Meehan, Lewis et al. 1989). However, at that time it was difficult to differentiate the mediator of methylation dependent transcriptional repression (Boyes and Bird 1991). A second methyl-C_pG specific binding activity for MeCP2, which is distinct from that of MeCP1, was found in rat brain nuclear extracts (Lewis, Meehan et al. 1992). Followed by deletion analysis of recombinant MeCP2, a minimal region that is needed for specificity towards methylated DNA was identified and termed as methyl-C_pG binding domain or MBD (Nan, Meehan et al. 1993). Furthermore, searches of EST databases led to the identification of several MBD containing proteins.

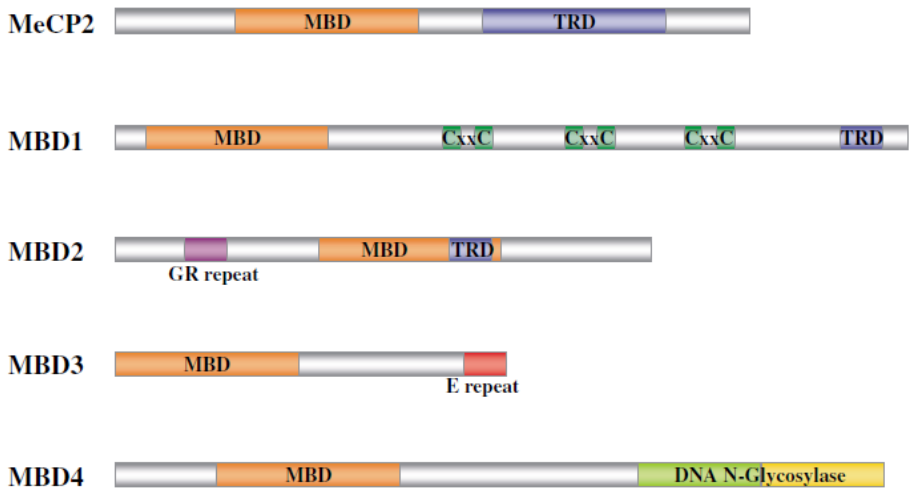


Figure 2.3. Domain Architecture of Methyl C_pG Binding Proteins: Methylated C_pG Binding Domain (MBD), Cysteine rich zing finger domain (CxxC), Transcriptional Repressor Domain (TRD) and DNA N-Glycosylase Domain are present among the family members of methylated C_pG binding proteins (Fatemi and Wade 2006).

The human MBD1 gene was the first to be cloned (Cross, Meehan et al. 1997), followed by the MBD2, MBD3 and MBD4 genes from mouse (Hendrich and Bird 1998). Like MeCP2, the newly identified MBD1, MBD2, and MBD4 proteins bind specifically to methylated C_pG DNA *in vitro*, while MBD3 being an exception. A schematic diagram of the architecture of MBD proteins is shown in (Fig. 2.3).

2.3.1 Structure of MBD

The structure of MBD of MeCP2 and MBD1 was solved using nuclear magnetic resonance (NMR), which reveals a wedge shaped α/β-sandwich structure comprising of a four stranded anti-parallel β-sheet, attached to a short α-helix (Ohki, Shimotake et al. 1999; Wakefield, Smith et al. 1999). Later the structure of MBD from MBD1 was solved in complex with methylated C_pG DNA using NMR (Ohki, Shimotake et al. 2001). Upon binding to DNA, a nine amino acid loop between the inner strands of the β-sheet gets structured and together with the inner β-strands and the loop connecting to the α-helix, this forms the majority of the interface with the major groove of the DNA. It was further noted from the structure that specific hydrophobic residues form the basis of interaction with the methyl groups on cytosine. However, a recent X-ray crystal structure of MBD from MeCP2, in complex with methylated DNA, (Fig. 2.4) shows that instead of hydrophobic contacts between methyl group, the water molecules are positioned around methyl groups make contact with the hydrophilic residues of MBD (Mayer-Jung, Moras et al. 1998; Ho, McNae et al. 2008).

2.3.2 MECP2

MeCP2 is a 53 kDa protein, which binds to methyl-C_pG pairs (Lewis, Meehan et al. 1992) and a flanking run of four or more A/T base pairs enhances the affinity

(Klose and Bird 2004). Expression of the MeCP2 protein level varies depending on cell type with the rat brain having the highest, followed by kidney, spleen, liver and the lowest in testes (Nan, Meehan et al. 1993). Inside the nucleus, once MeCP2 binds to methylated C_pG, it initiates transcriptional repression by recruiting the co-repressor molecules such as Sin3 and histone deacetylase (HDAC) to form a complex (Nan, Ng et al. 1998; Ng, Zhang et al. 1999). MeCP2 binds to HDAC through its transcriptional repression domain, residues 207-310 (Yu, Thiesen et al. 2000).

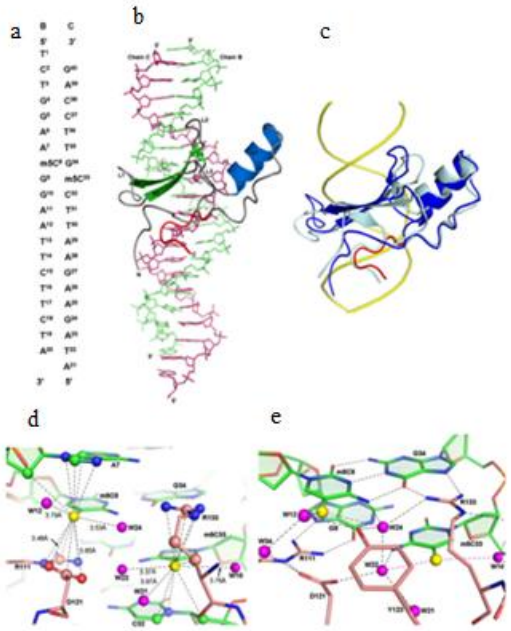


Figure 2.4. Crystal Structure of MBD of MeCP2 in complex with methylated DNA: From the structure, it can be inferred that methyl group in cytosine of DNA will recruit water and which in turn make contacts with hydrophilic residues such as Arg and Asp of MBD (**a**) Methylated C_pG DNA nucleotide sequence, (**b**) Complex structure of MeCp2-MBD with methylated DNA, (**c**) Overlay of native NMR structure with the complex structure of MeCp2-MBD, (**d**) Non bonded contacts with DNA, (**e**) Hydrogen bonding contacts with DNA (Ho, McNae et al. 2008).

When cells were treated with trichostatin A (an inhibitor of HDAC), they do not fully recover from transcriptional repression which confirms that MeCP2 does not totally depend on HDAC for repression and also it acts through other repressor complexes (Yu, Thiesen et al. 2000). MeCP2 was also found to be mutated at specific residues in Rett's syndrome and MeCP2 null mice showed abnormal behavior with loss of cognitive function, which confirms that MeCP2 has important role in neuronal function (Chahrour, Jung et al. 2008).

2.3.3 MBD2

MBD2 was first identified by Hendrich and Bird (Hendrich and Bird 1998) by searching for putative MBD domains in EST databases. It can be classified to MBD2a and MBD2b, where MBD2b lacks a glycine rich region of MBD2a. MBD2 might function as a histone deacetylase and chromatin remodeling transcriptional repressor based on reporter assays (Ng, Zhang et al. 1999; Feng and Zhang 2001). In addition, *in vitro* analysis proved that MBD2 hastens the binding of the NuRD complex to methylated DNA in nucleosomes, which are then remodeled and deacetylated. These findings also gave an idea that MBD2 and MBD3 might co-exist in the same NuRD complex, but another study showed the presence of distinct MBD2- and MBD3-containing NuRD complexes (Le Guezennec, Vermeulen et al. 2006). Mbd2 also seems to be needed in the colon to repress genes that are normally expressed in the exocrine pancreas (Berger, Sansom et al. 2007). Furthermore, the *Xist* gene promoter is also directly repressed by Mbd2 is (Barr, Hermann et al. 2007). In the above cases Mbd2 is found to represses transcription by recruiting histone deacetylase to promoters where DNA is methylated.

2.3.4 MBD3

MBD3 is a 265 amino acids (32 kDa) protein with an N-terminal MBD domain and a C-terminus region rich in acidic residues. MBD3 does not bind to methylated DNA, even though it has the MBD domain (Hendrich and Bird 1998). But unlike the human one, *Xenopus laevis* xMBD3 binds to methyl-C_pG (Wade, Gegonne et al. 1999). xMBD3 is expressed profusely in the early stages of development in eye, brain and bronchial arches. MBD3 was also found to co-purify with the Mi-2/NuRD complex (Le Guezennec, Vermeulen et al. 2006). The highly conserved MBD domain of MBD3 could also have a important role in protein-protein interactions.

2.3.5 MBD4

MBD4 is a 62.6 kDa protein and has a conserved MBD domain at the N-terminus and a typical glycosylase domain at the C-terminus involved in DNA repair (Hendrich and Bird 1998). The MBD domain of MBD4 binds to symmetrically methylated DNA and the C-terminal DNA glycosylase catalytic domain is a thymine and uracil glycosylase specific for G-T and G-U mismatches. MBD4 was found to suppress tumorigenesis by preventing C_pG mutability, i.e. C_pG to T_pG mutation, and thus MBD4 deficient mice are much prone to tumor (Millar, Guy et al. 2002).

2.3.6 MBD1

MBD1 is a 63 kDa protein and comprises of a highly conserved MBD domain, which binds to symmetrically methylated DNA at the N-terminus, a cysteine rich CXXC zing finger domain that binds unmethylated DNA and a number of CXXC domain, whose number varies upon isoforms and a transcriptional repressor domain (TRD), which is capable of forming an effective repressor complex by recruiting a co-

repressor complex (Ng, Jeppesen et al. 2000; Nakao, Matsui et al. 2001). The first human isoform, designated as PCM1, and an antibody raised against MBD1 was seen to 'supershift' the MeCP1 activity in EMSA and it was earlier thought that MBD1 might be a component of the MeCP1 complex (Cross, Meehan et al. 1997). However, later studies showed that there was cross reactivity of the MBD1 antibody with the MeCP1 complex (Ng, Zhang et al. 1999; Ng, Jeppesen et al. 2000). Full-length recombinant human MBD1 (PCM1) binds methylated DNA and suppresses transcription in a methylation dependent manner *in vitro* (Cross, Meehan et al. 1997). Recombinant mouse MBD1 was also able to bind to methylated DNA *in vitro* (Hendrich and Bird 1998). MBD1 was able to repress transcription on transfected reporter genes when transiently expressed in mammalian cells (Fujita, Takebayashi et al. 1999). Both isoforms MBD1v1 and MBD1v2 are able to repress transcription irrespective of whether it was a methylated or unmethylated promoter and repression of unmethylated reporters totally depends on the presence of the zinc finger domain CxxC3 in MBD1 (Fujita, Shimotake et al. 2000). In addition to this earlier study, an isolated CxxC3 domain from mouse Mbd1 binds to unmethylated C_pG containing DNA in a specific manner (Jørgensen, Ben-Porath et al. 2004). The localization of over expressed MBD1 to mouse pericentric heterochromatin depends upon the presence of wild type MBD and the presence of CXXC3 in the case of methylation deficient mice. The importance of MBD1 in neuronal cell differentiation has been reported. Mice deficient in MBD1 showed impairment in adult neurogenesis and differentiation of neural stem/progenitor cells (NSPCs), due to the over expression of fibroblast growth factor 2 (FGF2), which normally prevents neural cell differentiation (Zhao, Ueba et al. 2003; Li, Barkho et al. 2008). The importance of this epigenetic mechanism regulating the differentiation of NPSCs could be further understood from

another study, where MBD1 plays a major role in normal maintenance of neuronal cell differentiation. MBD1 was identified to be involved in the repression of miR-184 expression in neuronal cells, whereas miR-184 prevents the expression of Numbl, which is responsible for differentiation of NSPCs. It could be understood that it forms a signaling network between MBD1, miR-184 and Numbl (Liu, Teng et al. 2010).

From these above studies, we could understand the importance of MBD1 in neuronal cell regulation by regulating the epigenetic machinery inside cells. At the same time the level of MBD1 expression is found to be high in pancreatic cancer cell, which lowers the level of tumor suppressor gene expression. This effect can be reversed by inhibiting MBD1 expression (Xu, Liu et al. 2008). In all we could understand that MBD1 has high importance in physiological cell regulation and its over expression might results in cancer, where it silences several tumor suppressors. Thus it is clear that there should be a balance in the level of MBD1 in cells. The functional importance and roles of the MBD, CXXC and TRD domains of MBD1 in the regulation of gene expression are discussed in detail in the following sections.

2.3.6.1 DNA binding domain

MBD1 has two DNA binding domains, of which MBD binds specifically to methylated C_pG and the CXXC3 domain binds to unmethylated C_pG (Jørgensen, Ben-Porath et al. 2004). The structure of MBD of MBD1, in complex with methylated DNA, was solved using the NMR technique (Ohki, Shimotake et al. 2001) and there is a continuous patch of hydrophobic residues that makes contact with the methylated DNA, Fig. 2.5a. Also the structure of Myeloid/Lymphoid or Mixed-Lineage, Leukemia (MLL) and CXXC finger protein 1 (CFP1) CXXC domain in complex with unmethylated DNA which is highly homologous to the human MBD1 CXXC3

domain was solved using NMR and X-ray crystallography, respectively (Cierpicki, Risner et al. 2010; Xu, Bian et al. 2011). From the structures it is inferred that charged residues Arg and Lys are more important for recognizing unmethylated C_pGs (Figs. 2.5 b, c)**.**

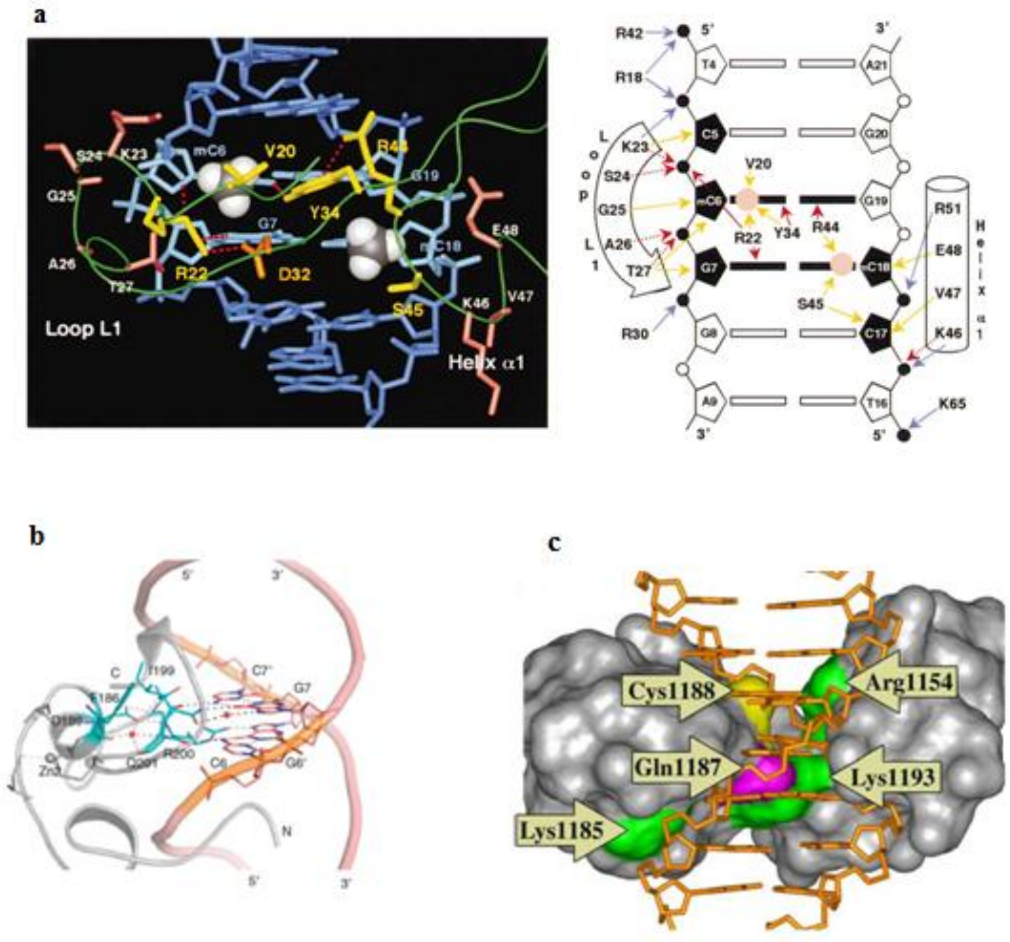


Figure 2.5. Known structures of MBD1.(**a**) NMR Structure of MBD of MBD1 in complex with methylated DNA (Ohki, Shimotake et al. 2001) (**b**) The CXXC domain in complex with unmethylated C_pG: The crystal structure of CXXC of CFP1 in complex with DNA (Xu, Bian et al. 2011) and (**c**) The NMR structure of CXXC of MLL in complex with DNA (Cierpicki, Risner et al. 2010).

The structure of the DNA binding domain of MBD1 gives us a clear idea on the specificity of the DNA binding domain towards methylated and umethylated C_pG. These domains are of high importance in cell signaling, such as heterochromatin formation and selective expression of certain genes on an organ specific manner. The presence of the unmethylated C_pG DNA binding domain CXXC3 of MBD1 could repress any gene expression once it binds to a promoter and recruits repressor complex proteins, but still the function of CXXC3 of MBD1 in cell signaling is unclear as promoters in the human genome contain unmethylated C_pG island.

2.3.6.2 Transcriptional Repressor Domain (TRD)

The human MBD1 variant PCM1 was found to repress transcription of transfected reporter genes when fused to a heterologous DNA binding domain and this allowed the mapping of a transcriptional repression domain (TRD) to the C-terminus of MBD1 (Ng, Jeppesen et al. 2000). TRD is about 64 amino acids long and acidic in nature with pI of 4.2. The ability of TRD to repress the expression of a transfected reporter gene has been reported (Ng, Jeppesen et al. 2000). When TRD is fused to a heterologous DNA binding domain, such as GAL4 DNA binding domain at the N-terminus, which binds to the reporter gene promoter, TRD recruits co-repressor complex proteins that are required for transcriptional repression. While TRD recruits HDAC proteins to repress gene expression, when treated with HDAC inhibitor, repressor activity is reduced but not abolished, (Fig, 2.6).

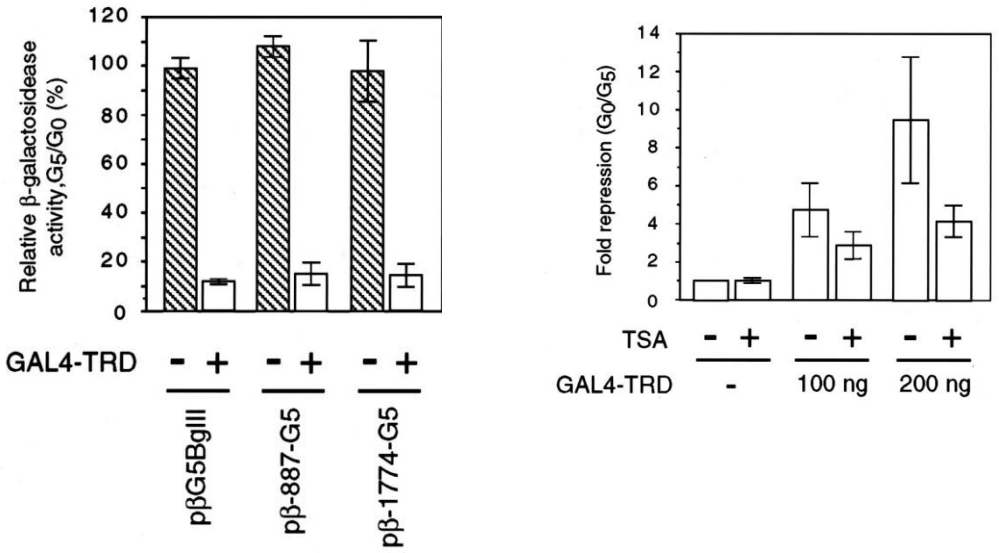


Figure 2.6. Repressor Activity of TRD. TRD was able to repress the reporter gene even at distance from transcription starter site when fused to GAL4 DNA binding domain and also the activity of TRD was reduced when treated with HDAC inhibitor (Ng, Jeppesen et al. 2000).

These finding confirms that TRD might be acting through other pathways that are involved in heterochromatin formation. It was later identified that TRD recruits binding partners such as the MCAF1 family of proteins and in turn recruits the SETDB1 complex to form an effective repressor system (Fujita, Watanabe et al. 2003; Ichimura, Watanabe et al. 2005).

2.3.6.3 MBD1 interacts with MCAF/AM family proteins through TRD to form heterochromatin

MBD1 containing chromatin association factor 1 (MCAF1) acts as an activator of transcription, when it binds to Specificity Protein 1 (Sp1) and then it recruits ATFa transcription factors involved in activation of various gene expression (Courey, Holtzman et al. 1989). MBD1, through its TRD region, is shown to interact with the C-terminus of MCAF1 (residues 1154 to 1270) (Fujita, Watanabe et al. 2003)

and in a similar manner, the homologous protein MCAF2 also interacts with MBD1 (Ichimura, Watanabe et al. 2005). Furthermore, selected hydrophobic residues of TRD are shown to be essential for interaction with MCAF/AM proteins, (Fig. 2.7).

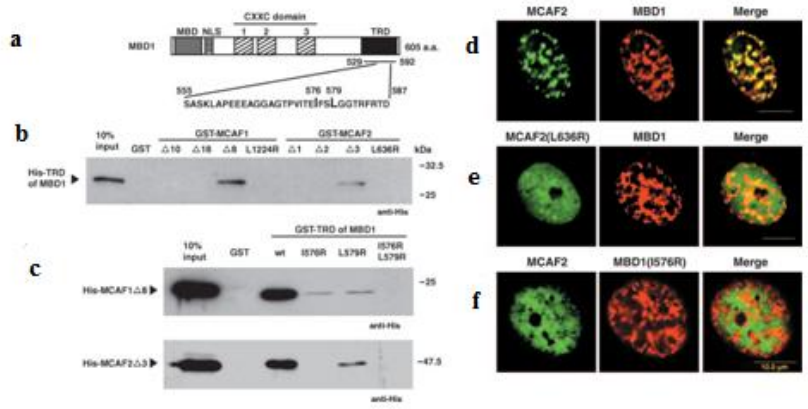


Figure 2.7. MBD1 Interaction with MCAF1 *via* TRD. Hydrophobic residues of MBD1 belongs to TRD region i.e. I576 and L579 are essential for binding to MCAF1 (Ichimura, Watanabe et al. 2005).

Thus MBD1 binds to MCAF1 and prevents its interaction with Sp1 and causes transcriptional repression. MBD1 interaction with MCAF1 will further recruit the SETDB1 complex to form a strong heterochromatin complex (Fujita, Watanabe et al. 2003). SETDB1 mediates the methylation of histone at lysine residues resulting in chromatin remodeling to form a compact mass and does not facilitate the binding of transcription factors to initiate gene expression. Thus it can be understood that MBD1 mediated transcriptional repression is majorly through MCAF1 binding rather than HDAC activity itself. Another related study showed that sumoylation of TRD will enhance the binding of MCAF1 to MBD1, which in turn helps in strong heterochromatin formation (Uchimura, Ichimura et al. 2006). Once it forms a heterochromatin complex, it could repress the gene expression of transfected reporter genes. In fact, the presence of TRD alone is enough to cause repression of a gene since it interacts with the repression complex machinery by itself, and the only

requirement is the presence of the DNA binding domain to repress any gene promoter, Fig. 2.8. Thus TRD is very important as a highly active repressor domain and its mode of partner recognition is also of high importance.

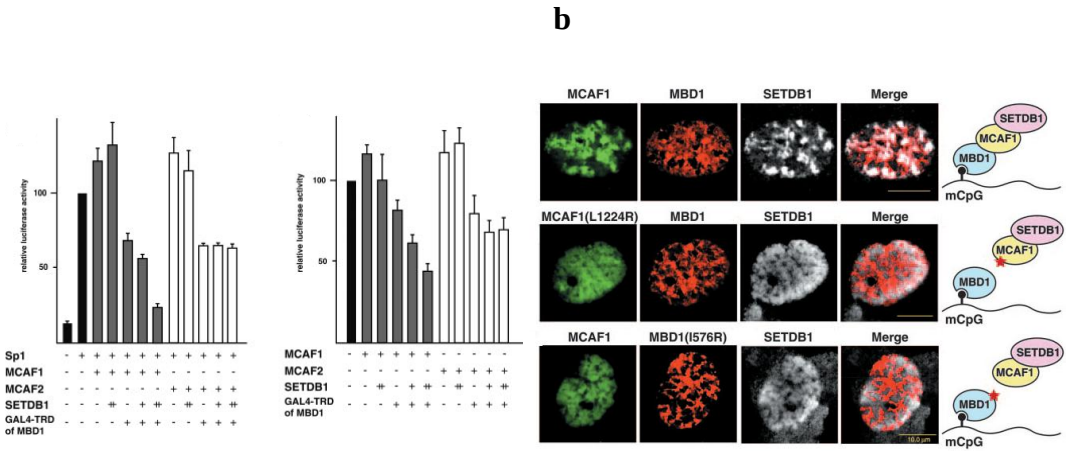


Figure 2.8. Heterochromatin formation by MBD1. (**a**) GAL4-TRD able to repress the reporter gene expression in the presence of MCAF1 and SETDB1, where clear repression occurs in the case of Sp1 mediated expression (**b**) MBD1 forms heterochromatin complex with MCAF and SETDB1 (Ichimura, Watanabe et al. 2005)

2.3.6.4 MBD1 is required for PML-RARα mediated gene repression where TRD is essential for co-repression

PML-RARα is an oncoprotein which blocks hematopoietic differentiation and results in acute promyelocytic leukemia (APL). It generally functions through the repression of proteins involved in differentiation of blood cells, such as RARβ2, by repressing RARE genes (Lin, Nagy et al. 1998). PML-RARα forms a strong repressor complex with DNA methyl transferases and HDAC3, and interestingly MBD1 is also part of the complex. PML-RARα recruits MBD1 through HDAC3, since MBD1 directly interacts with HDAC3 through the TRD. PML-RARα facilitates the methylation of DNA by recruiting Dnmt1 and then HDAC3, which is part of N-CoR complex, also will be present in contact with PML-RARα and in turn MBD1 will be

recruited to that methylated promoter region, thus further transcription of that particular gene will be affected, resulting in gene silencing. HDAC3 was found to interact with the TRD region of MBD1 through hydrophobic interaction, since mutants in that region fail to interact with HDAC3. Most importantly, TRD is needed to repress gene expression in acute myeloid leukemia (Villa, Morey et al. 2006). A schematic model of the PML-RARα repressor complex and the associated molecular events are explained in Fig. 2.9.

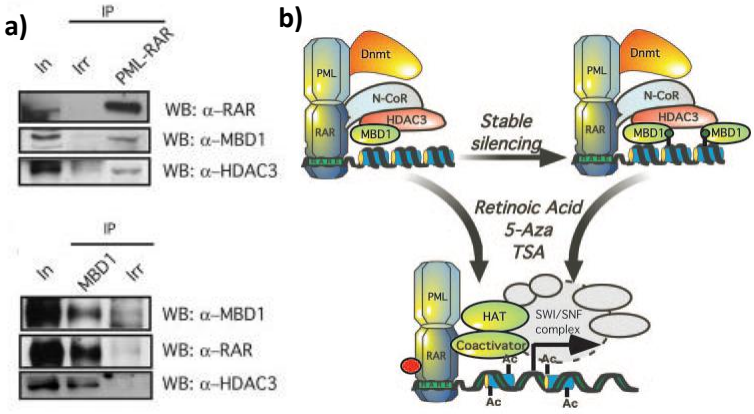


Figure 2.9. PML-RARα forms Co-Repressor with MBD1 and HDAC3. (a) Co-IP experiment confirms that the PML-RARα exists as complex with HDAC3 and MBD1 in APL cells (**b**) Schematic representation of PML-RARα co-repressor complex with N-CoR, HDAC3 and MBD1 (Villa, Morey et al. 2006).

2.3.6.5 MBD1 interaction with MPG occurs through TRD

3-methyl purine DNA glycosylase (MPG) is involved in a DNA repair mechanism. When there is any damage in DNA base pairs adenine or guanine, due to alkylation, deamination or oxidation, it binds to such a modified base and cleaves of the glycosidic linkage between that damaged base and the deoxy ribose sugar. This is

followed by endonuclease removal of that base, replacement with an intact base and ligation (Wyatt, Allan et al. 1999). MPG is an important enzyme in DNA repair, since MPG knock down mice showed high sensitivity to alkylation damage (Engelward, Weeda et al. 1997). MPG was found to be present in the MBD1 complex and yeast two hybrid screen performed with TRD as a bait showed that MPG interacts with it in a specific manner and fails to interact with CXXC or MBD of MBD1. Full length MPG was found to be essential for interaction with TRD of MBD1, since truncations at the N-terminus or C-terminus affected the interaction and co-localization inside mammalian cells. It was presumed that MPG could be a part of the heterochromatin complex with MBD1 and indeed the presence of MPG enhanced gene repression by MBD1. In other words, MBD1 is essential for effective DNA repair by MPG, since the cells that lack MBD1 are unable to carry out DNA repair effectively unlike in the presence of MBD1 (Watanabe, Ichimura et al. 2003).

The above studies confirm that the interaction between MBD1 and MPG is essential for effective DNA repair and heterochromatin formation, and such interaction is mediated through TRD of MBD1, (Fig. 2.10).

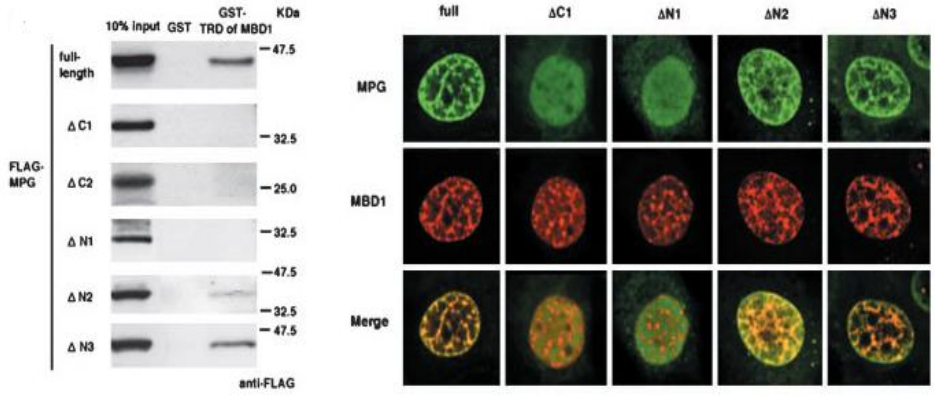


Figure 2.10. MBD1 interaction with MPG *via* TRD. GST pull-down and co-localization experiment confirms the presence of FL MPG is essential for interaction with TRD of MBD1 (Watanabe, Ichimura et al. 2003).

2.4 OBJECTIVES

1. To determine the structural features of TRD of MBD1 using NMR.
2. To determine the mode of interaction of TRD with its binding partners such as MPG and MCAF1Δ8, using NMR, ITC, etc.
3. To determine the importance of specific residues involved in interaction between TRD and its binding partners, by carrying out site-directed mutation based on the biophysical studies and verified using *in vitro* and *in vivo* studies.

CHAPTER 3. MATERIALS AND METHODS

3.1 SUBCLONING FOR PROTEIN EXPRESSION

The gene fragment encoding residues 507-605 of MBD1, including the transcription repressor domain (TRD, residues 529-592) was PCR amplified using the Pfu DNA polymerase (Fermentas) with the forward primer ATTGAATTCGACGAGTGGACACCAGand reverse primer TATCTCGAGCTACTGCTTTCTAGCTC (1^{st} Base), containing the EcoRI and the XhoI restriction sites (underlined), respectively. The gene fragment was subcloned into the pET-M (pET32a modified) vector with an N-terminal His tag and pGEX6P-1 vector with an N-terminus GST tag for protein expression and purification. In the case of MPG, the full length gene of MPG was amplified with a forward primer containing an EcoRI site (ATTGAATTCATGCCCGCGCGCAGC) and a reverse primer with an XhoI site (TTACTCGAGTCAGGCCTGTGTGTCCTGC)and then subcloned into the pETM vector. The gene encoding for the MCAF1 residues 1154-1270 that interact with TRD, hereafter named as MCAF1Δ8, was amplified with the forward primer ATTGAATTCGAAGCTGCCAGCACATCTCTGCCTC (with an EcoRI site) and the reverse primer ATTGCGGCCGCTTAACTGCTCTGGGTAGAAGAG (with a NotI site) and then subcloned into the multiple cloning site 1 (MCS1) of the pETduet-1 vector with an N-terminal MBP tag and the pET-M vector with an N-terminal His-tag for protein expression.

3.2 PROTEIN EXPRESSION AND PURIFICATION

Seed cultures of transformed *E. coli* BL21 (DE3) (Stratagene) competent cells were added to large-scaled LB medium containing 100 μg/ml ampicillin (Gold

Biotech). Cells were grown until $A_{600\ nm}$ was 0.6 to 0.7 before adding 0.15 mM isopropyl-1-thio-β-D (-) thiogalactopyranoside (IPTG) (Gold Biotech) for protein over-expression for another 14 to 18 h at 16 °C. Cells were harvested and the pellet was stored at -80 °C. The same expression protocol was used to prepare ^{13}C- or ^{15}N-labeled TRD protein by using the M9 medium containing ^{15}N-isotopic ammonium chloride and ^{13}C isotopic glucose (Cambridge Isotopes) or ^{12}C glucose.

The cell pellet was resuspended in binding buffer A: 50 mM Tris-HCl buffer, 150 mM NaCl, 2mM DTT, 5% Glycerol, 0.1% Triton X-100 with final pH adjusted to 7.0. The cell suspension was lysed by sonication on ice and the supernatant was loaded onto an Ni-NTA column and subsequently allowed to bind for 1 h at 4 °C. Weakly bound proteins were removed by extensive washing using binding buffer A and buffer B (buffer A with 10 mM imidazole). TRD was finally eluted with buffer B containing 300 mM imidazole. The eluted TRD was loaded to a Superdex 75 column (GE Healthcare) equilibrated with 20 mM sodium phosphate, 50 mM NaCl and 3 mM DTT, pH adjusted to 6.5. TRD eluted as a single homogenous peak, but at a different molecular weight (corresponding to 44 kDa instead of its 12 kDa), probably due to its intrinsic disorderliness. The same steps of purification was followed for unlabelled, ^{15}N labeled or ^{15}N and ^{13}C double labeled TRD. The purified protein was concentrated to 10 mg/ml using ultra 3 kDa Mol. Wt. cut off Centricon (Merck Millipore).

Expression of pGEX-6P-1:TRD followed almost the same above conditions, but during purification, the lysate was allowed to bind to glutathione sepharose beads (GE Healthcare), equilibrated with buffer A for 3 h at 4 °C, then washed with Buffer A for 3 times. The GST tag was removed from TRD using 1 unit of Prescission protease per mg of GST-TRD at 4 °C. The cleaved TRD protein was loaded to a

Superdex 75 column (GE Healthcare), equilibrated with the same elution buffer as before and concentrated.

MPG was over-expressed with an N-terminal 6xHis tag in the pET-M vector. In short, the construct was transformed into *E. coli* BL21 (DE3) competent cells and the protein was over expressed and first affinity purified using Ni-NTA resin. The protein was further purified using a cation exchange Hi-Trap SP-HP (GE Healthcare) column (MPG is highly basic with pI of about 10), with buffer A: 20 mM sodium phosphate and buffer B: 20 mM sodium phosphate, 1 M NaCl pH adjusted to 8.0. MPG was again purified using a Superdex 75 column (GE Healthcare) equilibrated with 20 mM sodium phosphate, 50 mM NaCl and 3 mM DTT, pH adjusted to 6.5. The protein eluted as a single homogenous peak, at its theoretical Mol. Wt. of 35 kDa. The purified MPG protein was concentrated using ultra filtration membrane (Merck Millipore) with 10 kDa Mol. Wt. cut off.

MCAF1Δ8 was over-expressed using the pETduet-1 vector as an N-terminal MBP-fusion protein. Seed cultures of transformed *E. coli* BL21 (DE3) competent cells were added to large scale LB medium containing 100 µg/ml ampicillin. The cell density was monitored until A_{600nm} was 0.6 to 0.7, before adding 0.1 mM IPTG for protein over-expression for another 16 h at 18 °C. Cells were harvested and cell pellet was stored at -80 °C until needed. The cell pellet was resuspended in binding buffer A: 50 mM Tris-HCl buffer, 150 mM NaCl, 2 mM DTT, 5% Glycerol, 0.1% Triton X-100 and final pH adjusted to 8.0. The cell suspension was lysed by sonication on ice and the supernatant was loaded onto an amylose (NEB) column and subsequently allowed to bind for 2 h at 4 °C. The weakly bound proteins were removed by extensive washing using binding buffer A for 3 to 4 times. The MBP: MCAF1Δ8 fusion protein was eluted with 20 mM maltose in buffer A. Thus purified MBP-

MCAF1Δ8 was passed through a Superdex 200 column (GE Healthcare) equilibrated with 20 mM sodium phosphate, 50 mM NaCl and 3 mM DTT, pH adjusted to 6.5. MBP: MCAF1Δ8 eluted as a single homogenous peak, at its theoretical Mol. Wt. of 55 kDa. The purified protein was concentrated using ultra filtration membrane (Merck Millipore) with 10 kDa Mol. Wt. cut off. The MBP tag was not cleaved, as MCAF1Δ8 by itself is very unstable and precipitates immediately after purification.

Table 3.1 List of Constructs used in this study of TRD

S. No		Fwd Primer	Rev Primer	Vector
1	MBD1 Δ	5'ATTGAATTCGACGA GTGGACACCAG3'	5'TATCTCGAGCTAC TGCTTTCTAGCTC3'	pET-M
2	MBD1 Δ	5'ATTGAATTCGACGA GTGGACACCAG3'	5'TATCTCGAGCTAC TGCTTTCTAGCTC3'	pGEX 6P-1
3	MCAF 1Δ8	5'ATTGAATTCGAAGC TGCCAGCACATCTCT GCCTC3'	5'ATTGCGGCCGCTT AACTGCTCTGGGTA GAAGAG3'	pET-M
4	MCAF 1Δ8	5'ATTGAATTCGAAGC TGCCAGCACATCTCT GCCTC3'	5'ATTGCGGCCGCTT AACTGCTCTGGGTA GAAGAG3'	pET duet-1
6	MPG	5'ATTGAATTCATGCC CGCGCGCAGC3'	5'TTACTCGAGTCAG GCCTGTGTGTCCTG C3'	pET-M

3.3 CIRCULAR DICHROISM SPECTROSCOPY

Circular dichroism (CD) spectroscopy experiments were performed on a JASCO J-810 spectropolarimeter. A quartz cuvette with 1.0 mm optical path was filled with 200 µl proteins (~ 25 µM) in 10 mM NaH2PO4/ Na2HPO4, pH 6.5 and CD spectra were recorded from 190 to 260 nm at 25 °C.

3.4 ISOTHERMAL TITRATION CALORIMETRY (ITC)

The interaction of TRD with MPG and MBP: MCAF1Δ8 was verified using a VP-ITC microcalorimeter (MicroCal) at 25 °C in 20 mM NaH2PO4/ Na2HPO4 buffer, pH 6.5. The cell contained 50 μM MPG or 20μM MBP: MCAF1Δ8. The syringe contained 500 μM of TRD against MPG and 200μM of TRD against MCAF1Δ8, was prepared in the same buffer. Samples were first degassed for 15 min and titration was performed using a stirring speed of 390 rpm. The initial injection for TRD was 2 μl for 10 s and subsequent injections were 10 μl for 10 s with 240 s spacing. Data points were fitted using the integrated program ORIGIN (MicroCal) for one-site binding.

3.5 DYNAMIC LIGHT SCATTERING (DLS)

Dynamic light scattering measurements were performed at room temperature on a DynaPro (Protein Solutions) DLS instrument. The homogeneity of the protein samples was monitored at various stages of concentration in order to avoid aggregation. The percentage of polydispersity was below 16% and the SOS error was less than 50 for all protein samples at various concentrations.

3.6 GST PULL-DOWN ASSAY

For GST pull down assay, about 1 μg of GST alone and GST-fused TRD wild type and TRD mutants was allowed to bind to glutathione sepharose beads (GE Healthcare) in buffer containing 50 mM Tris-HCl buffer, 150 mM NaCl, 2 mM DTT, 5% Glycerol, 0.1% Triton X-100, pH 7.0. MBP:MCAF1Δ8 and 6xHis:MPG were also prepared with the same buffer and equal amount of proteins was allowed to bind to GST:TRD and mutants. The beads were washed with buffer alone for three times.

Finally complexes were eluted with 20 mM reduced glutathione and western blotted using anti-His antibody.

3.7 NMR DATA ACQUISITION AND ANALYSIS

All NMR experiments were performed on a 500 or 800 MHz NMR spectrometer (Bruker) at 25 °C. We collected the following NMR spectra for each ^{13}C, ^{15}N-labeled protein sample: 2D ^{1}H-^{15}N HSQC, 2D ^{1}H-^{13}C HSQC, 3D HN(CO)CA, 3D HNCA, 3D MQ-(H)CCH-TOCSY (Yang, Zheng et al. 2004) and 4D timeshared ^{13}C/ ^{15}N edited-NOESY (Xu, Long et al. 2007).

All data were processed with NMRpipe and analyzed with NMRview or in-house NMRspy and an extension XYZ4D (http://yangdw.science.nus.edu.sg/Software&Scripts/XYZ4D/index.htm). NMRspy was recently developed to facilitate resonance assignment with the 4D NOESY-based strategy (Xu, Zheng et al. 2006).

3.7.1 $\Delta(C_\alpha$-$C_\beta)$ combined chemical shift plot

$\Delta(C_\alpha$-$C_\beta)$ combined chemical shift plot based on (C_α, TRD - C_α, random coil) - (C_β, TRD - C_β, random coil) was used to identify well-defined secondary structure, where continuous positive values (> 3 ppm) represent the α-helix region and continuous negative values (< -2 ppm) represent β-sheet region. Random coil values are obtained from the averaged shift values of each amino acid available at the Biological Magnetic Resonance Data Bank (BMRB).

3.7.2 Corrected residual chemical shift plot

To identify possible residual secondary structure, a corrected residual chemical shift plot was used. Unlike the $\Delta(C_\alpha$ - $C_\beta)$ combined chemical shift plot, the

corrected residue chemical shift plot takes into account the nearest neighbor effect up to two residues away. ΔH_α is calculated based on (H_α, TRD - H_α, random coil), where the random coil values take the random coil shifts reported for the Ac-GGXPGG-NH_2 peptide if the preceding residue (X+1) is a proline and Ac-GGXAGG-NH_2 peptide if the preceding residue is not a proline. The correction factors A, B C, D are then added, where A and D account for long range effect due to X-2 and X+2 residues particularly if they are aromatic residues, respectively. The same is applied for ΔC_α using equation 3.1

$$\Delta R(\text{corrected}) = (R - R_{\text{random coil}}) + A + B + C + D \qquad (3.1)$$

where R is H_α or C_α of a residue and four correction factors (A, B, C, D) are obtained from the study by (Schwarzinger, Kroon et al. 2001).

3.8 STRUCTURE CALCULATION

NOE restraints were obtained from the timeshared 4D $^{13}C/^{15}N$-edited NOESY (containing ^{13}C, ^{15}N-edited, ^{13}C, ^{13}C-edited and ^{15}N, ^{15}N-edited subspectra) using NMRview or NMRspy. Backbone torsional angle restraints, Φ and Ψ were obtained from chemical shift data using TALOS (Cornilescu, Delaglio et al. 1999). Ambiguous NOEs were assigned with the iterative structure calculation method using Cyana 2.1 (Herrmann, Güntert et al. 2002). The final 10 lowest energy structures out of 100 calculated were selected. The quality of the structure was assessed using PROCHECK (Morris, MacArthur et al. 1992). All data were processed using NMRPipe program and analyzed with NMRspy (Xu, Zheng et al. 2006).

3.9 NMR TITRATION

For titration experiments using ^{1}H-^{15}N HSQC-TROSY, an initial concentrated unlabeled MPG/ MBP-MCAF/ MBP was minimally added to 0.2 mM ^{15}N-labeled TRD with the His-tag removed until a final molar ratio [MBP-MCAF/ MCAF/ MBP]:[TRD] of 3:1. The ^{1}H and ^{15}N chemical shift difference, before and after titration was calculated with the equation

$$\Delta\delta = ((\Delta\delta_{HN})^2+(\Delta\delta_N/7)^2)^{0.5} \quad (3.2)$$

where $\Delta\delta_{HN}$ and $\Delta\delta_N$ are the respective differences of ^{1}H and ^{15}N chemical shifts of an amide in the free and bound TRD. For intensity analysis, the ratio of the bound to the free TRD was derived from the ratio of the peak heights extracted using NMRspy and corrected with a scaling factor to account for dilution.

CHAPTER 4. RESULTS

4.1 EXPRESSION, PURIFICATION AND CHARACTERIZATION OF TRD

The transcriptional repressor domain (TRD) of MBD1, present at the C-terminus, residues (507 to 605), was cloned into the pET-M vector and over expressed in BL21 (DE3) cells. The over-expressed protein was purified by two step purification, first with Ni-NTA purification and then by size exclusion chromatography, where it eluted as a single peak, corresponding to 45 kDa (Ziemer, Mason et al. 1982), instead of 12 kDa (Fig. 4.1). The fractions were verified by SDS-PAGE, (Fig. 4.2).

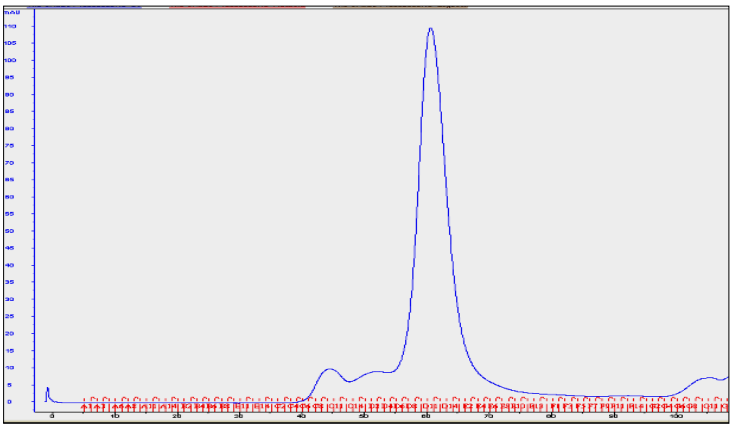


Figure 4.1. Size exclusion chromatography of TRD.

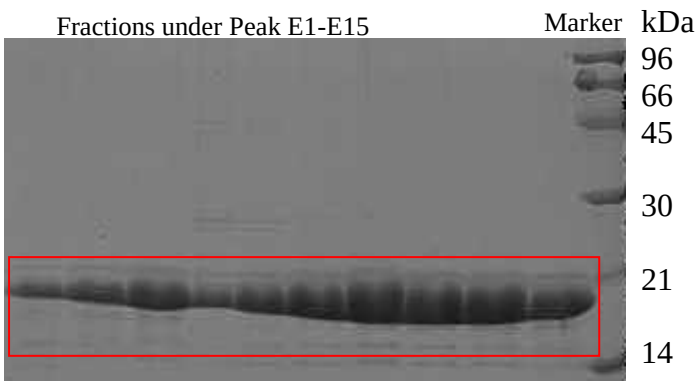


Figure 4.2. SDS-PAGE of the TRD fractions under peak.

The molecular weight obtained from SDS PAGE seems to be higher than 12 kDa and then mass spectrometry was performed to confirm the molecular weight is 12 kDa for TRD (Fig. 4.3). And the higher molecular weight obtained in size exclusion could be due to non-globular nature of the protein.

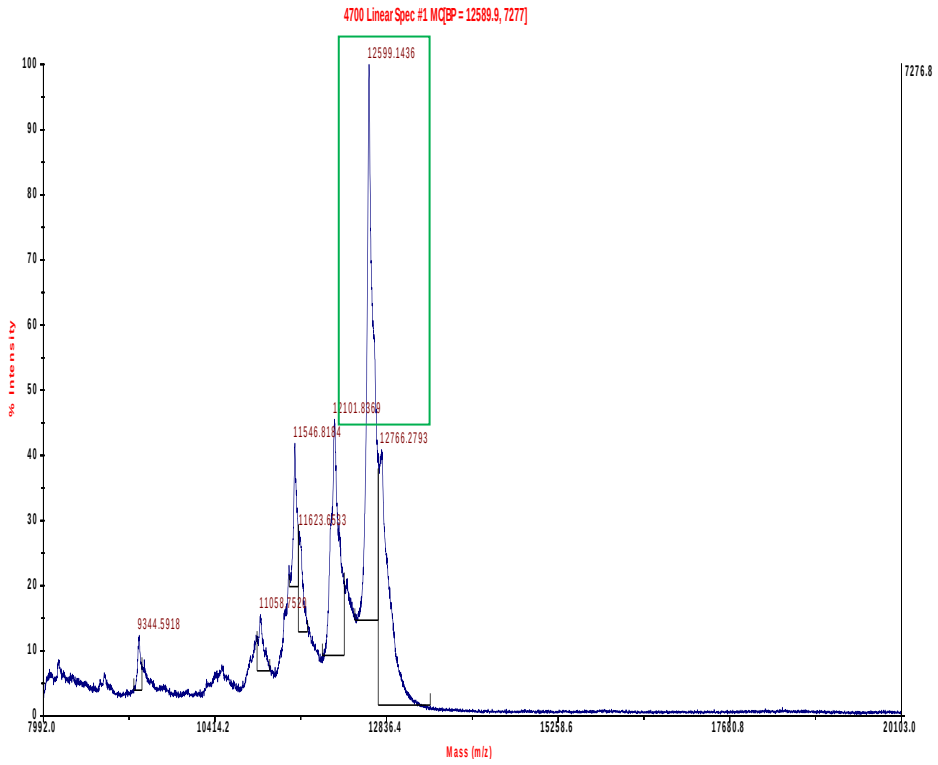
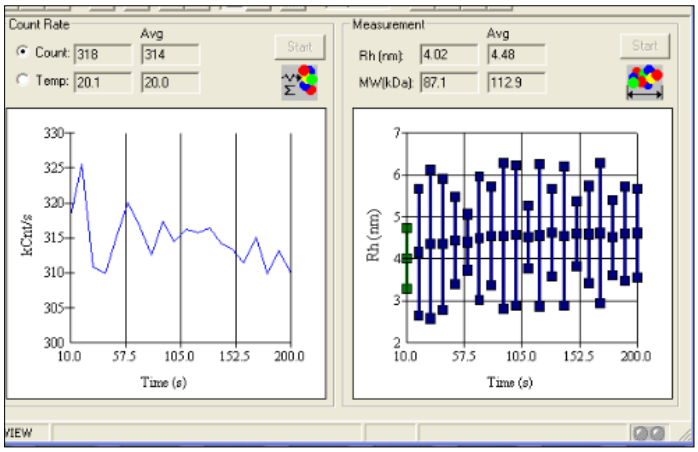


Figure 4.3. Molecular weight determination of TRD using mass spectrometry. The molecular weight was determined as 12 kDa

The TRD protein was concentrated to 10 mg/ml and its homogeneity was analyzed using dynamic light scattering (DLS). The results confirm that TRD is homogenous with a polydispersity index of 9% (limit is 20%), radius 4.48 nm and SOS error 12.8 (limit is 50), (Fig. 4.4).



Mst#	Time(s)	Temp(C)	Count Rate	Ampl	Diff Coeff	Radius(nm)	Polyd(nm)	PolydIndx	MW(KDa)	%Mass	Baseline	Sos Error
1	10.0	20.1	318187	0.768	510.	4.02	0.730	0.03	87.1	100.0	1.000	12.0
2	20.0	20.1	325516	0.745	492.	4.16	1.51	0.13	94.6	100.0	1.001	9.18
3	30.0	20.1	310851	0.667	470.	4.36	1.78	0.17	106.	100.0	1.000	12.2
4	40.0	20.0	309921	0.721	469.	4.35	1.57	0.13	105.	100.0	1.000	9.16
5	50.0	19.9	315301	0.791	458.	4.44	1.04	0.05	110.	100.0	1.000	12.4
6	60.0	20.0	320153	0.807	464.	4.40	0.668	0.02	108.	100.0	1.000	12.4
7	70.0	20.0	316386	0.821	454.	4.50	1.47	0.11	114.	100.0	0.999	13.3
8	80.0	20.0	312573	0.821	448.	4.56	1.18	0.07	117.	100.0	1.000	15.3
9	90.0	20.0	317432	0.816	450.	4.54	1.74	0.15	116.	100.0	1.000	14.2
10	100.0	20.0	314459	0.824	447.	4.57	1.67	0.13	118.	100.0	1.000	13.4
11	110.0	20.0	316175	0.820	452.	4.52	0.748	0.03	115.	100.0	1.000	12.0
12	120.0	20.0	315824	0.815	448.	4.56	1.70	0.14	117.	100.0	1.000	12.7
13	130.0	20.0	316451	0.813	441.	4.63	1.05	0.05	121.	100.0	1.000	21.2
14	140.0	20.0	314185	0.810	449.	4.55	1.65	0.13	116.	100.0	1.000	11.0
15	150.0	20.0	313366	0.807	444.	4.60	0.770	0.03	120.	100.0	1.000	12.0
16	160.0	20.0	311397	0.811	445.	4.59	1.15	0.06	119.	100.0	1.000	8.05
17	170.0	20.0	315113	0.799	443.	4.61	1.68	0.13	121.	100.0	1.000	12.6
18	180.0	20.0	309784	0.794	452.	4.51	0.896	0.04	114.	100.0	1.000	15.3
19	190.0	20.0	313138	0.790	444.	4.60	1.13	0.06	120.	100.0	1.001	15.4
20	200.0	20.0	310033	0.785	443.	4.61	1.05	0.05	121.	100.0	1.000	12.0
Aves:												
Mono		20.0	314812	0.791	456.	4.48	1.26	0.09	113.	100.0	1.000	12.8
Bi-1		0.0	0	0.000	0.000	0.000	---	---	0.000	0.0	0.000	0.000
Bi-2				0.000	0.000	0.000			0.000	0.0		

Figure 4.4. DLS of TRD at 10mg/ml.

4.3 STRUCTURAL CHARACTERIZATION OF TRD

4.3.1 Back bone assignment using 2D ^{15}N HSQC NMR

TRD was concentrated to 10 mg/ml for structure determination using NMR. The TRD protein was labeled with ^{15}N for detection in Heteronuclear Single Quantum Coherence (HSQC), where amide peaks can be visualized, so that the number of peaks corresponds to the number of amino acids of the protein excluding proline, which has an imide instead of an amide bond and cannot be seen in HSQC. HSQC will give us preliminary information about the structural contents of a protein, and if a protein is structured with defined secondary structures then we would be able to see well dispersed peaks and it will be easy to differentiate them. If there is no defined

structure or minimal secondary structure, it will lead to narrow dispersion of peaks and some peaks could remain overlapped.

In the 2D ^{15}N-HSQC, the narrow dispersion of the peaks between ^{1}H: 7.6 - 8.7 ppm hinted an overall degree of unfolded state. Although TRD is highly unstructured, the buffer condition was optimized to obtain well resolved ^{15}N HSQC peaks within the narrow ^{1}H and ^{15}N ppm window, (Fig. 4.5). This is atypical of an unstructured protein, which is often characterized by peak overcrowding in ^{15}N-HSQC spectrum. For performing backbone assignment, I have double-labeled the protein with ^{13}C and ^{15}N, and data were collected. Based on the fact that the chemical shift of each amino acid is unique, each peak was assigned, based on chemical shift and also from the information of neighboring amino acids in the structure in the Fig 4.5.

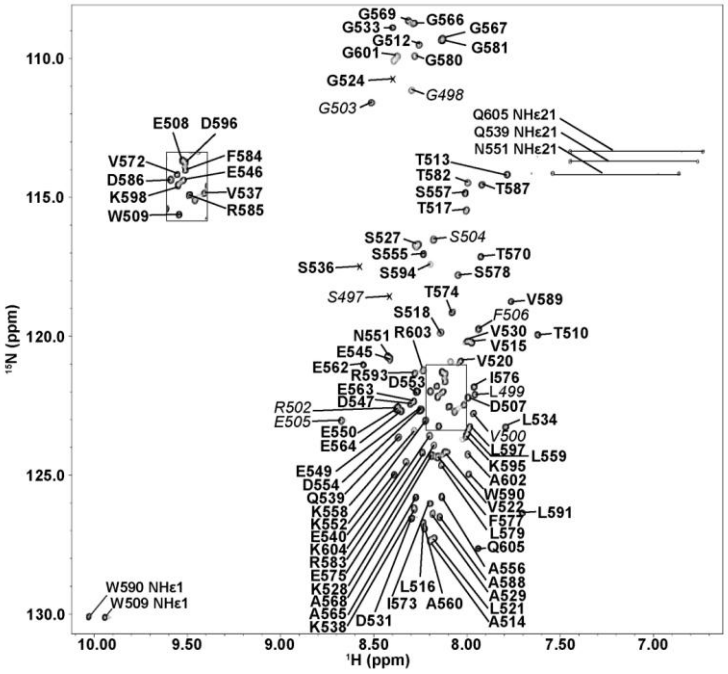


Figure 4.5. Backbone Assignment of TRD of MBD1. The HSQC spectra of TRD showing the assigned residues in the region between ^{1}H: 8.0-8.2 ppm and ^{15}N: 121.0-123.3 ppm. Residues belonging to the 6xHis-tag are italicized and those of MBD1 are in bold. The residues are numbered according to MBD1, isoform 1.

To check further on whether TRD has any residual secondary structure, circular dichroism (CD) experiments were performed. From the CD data, (Fig. 4.6), it is clear that TRD by itself may not have any defined secondary structure, since there are no peaks between 190 and 260 nm, the specific wavelengths at which absorption peaks are produced by α-helices and β-strands. However, significant secondary structure formation is observed in the presence of 40% trifluoroethanol (TFE), an alpha helical inducer.

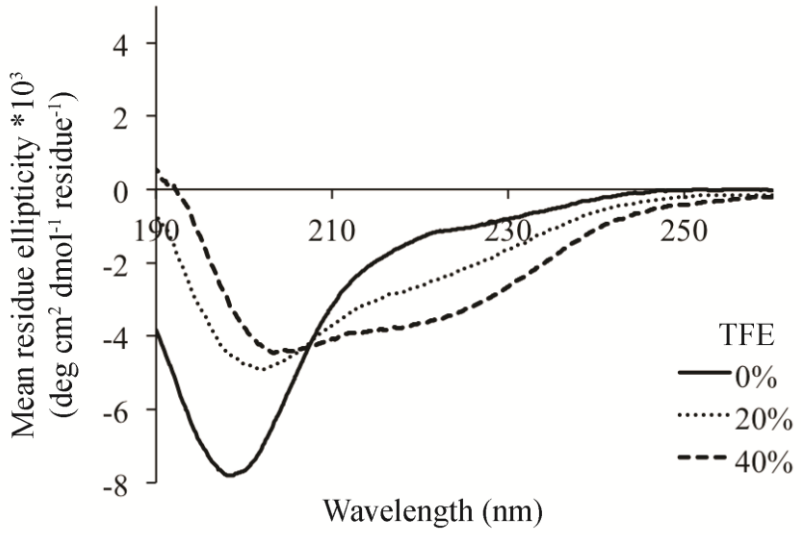


Figure 4.6. CD Spectra for TRD. TRD by itself does not have secondary structures and 40% TFE induces secondary structures.

4.3.2 NMR characterization of TRD

TRD of MBD1 is conserved in closely related organisms, where sequence identity is more than 50%, (Fig 4.7). TRD was known for its importance in transcriptional gene silencing through recruiting binding partners to the promoter of the gene need to be silenced (Ichimura, Watanabe et al. 2005).

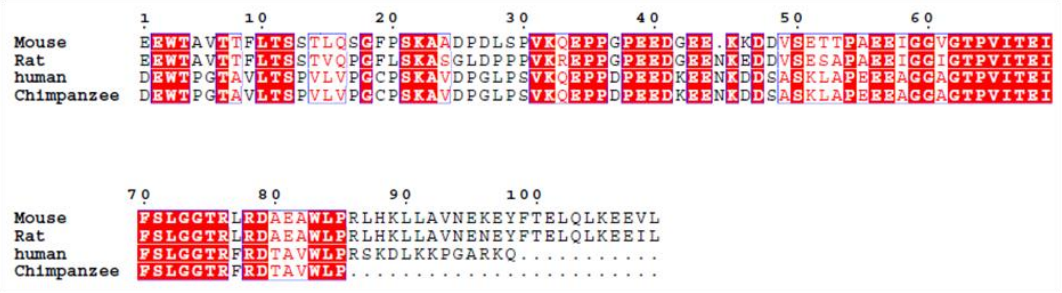


Figure 4.7. Multiple Sequence Alignment of TRD of MBD1 from closely related organisms. The charged residues and hydrophobic residues are highly conserved.

The collected 4D data showed only a few medium and long range NOEs, suggesting the lack of a global fold, (Fig. 4.8a). Even though the standard chemical shift index for individual amino acids (based on the BioMagResBank) also suggested the absence of a well defined secondary structure, some residual helical structure may exist in the protein sequence, (Fig. 4.8 b-d). The initial Δ(Cα-Cβ) suggests a possible short helical structure between K552 and A556 despite a random coil circular dichroism profile. To further validate this, both Cα and Hα resonance assignments of TRD were used to probe for residual secondary structure as previously developed (Wishart, Bigam et al. 1995; Schwarzinger, Kroon et al. 2001). In the region between K552 and A556, a contiguous stretch of residual ΔCα'/Hα' values are in the range of -0.25 to +1, (Fig. 4.8 c and d), suggesting a transient helical structure. The value for properly formed helical or extended structures is -1 and +1. This residual helical propensity in unstructured TRD is further shown by the presence of dips at 205 nm and 222 nm, upon adding 20 to 40% TFE, a known helical inducer, (Fig 4.6). Nonetheless, the protein is overall highly unstructured, implying that the vital transcriptional repressor role of TRD does not require a well-defined secondary structural motif.

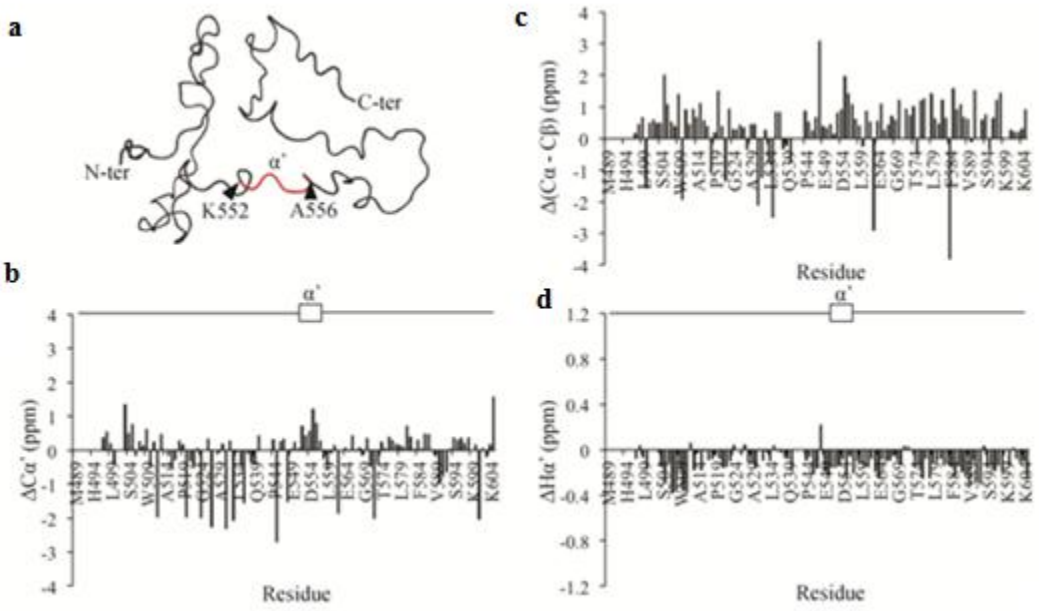
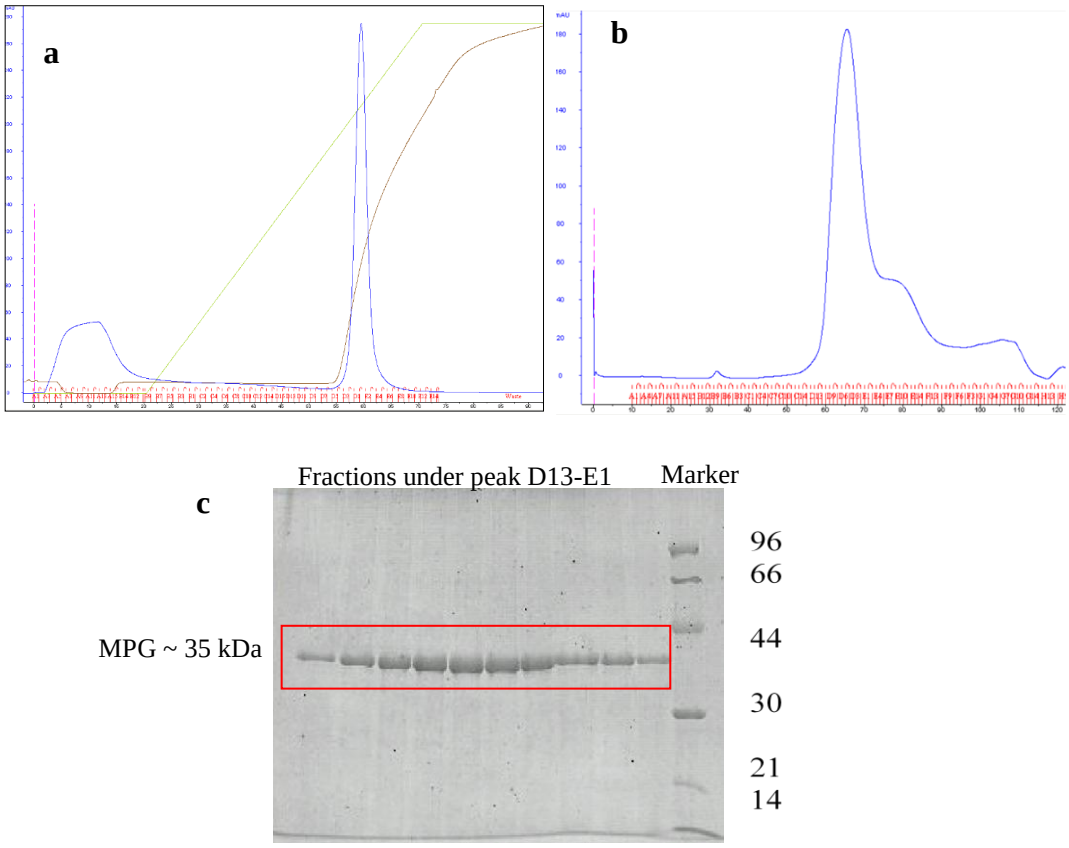


Figure 4.8. NMR analysis of TRD. (**a**) Lowest energy structure of MBD1Δ showing putative residual helix α' in red. **(b)** Traditional chemical shift index (CSI) plot based on ($C_{\alpha,\ TRD}$ - $C_{\alpha,\ random\ coil}$ − $C_{\beta,\ TRD}$ - $C_{\beta,\ random\ coil}$). (**c**) Residual CSI plot of Cα' based on corrected random coil values of Cα (**d**) Residual CSI plot of Hα' based on corrected random coil values of Hα. The putative residual helix α' is shown above in (**c**) and (**d**).

4.4 TRD INTERACTS WITH MPG AND MCAF1Δ8

4.4.1 Expression and purification of MPG

Full length MPG (1-298 amino acids) was overexpressed in BL21 (DE3) cells under the T7 promoter in the pET-M vector. The overexpressed protein was purified in three steps: affinity purification using Ni-NTA resin, cation exchange chromatography (the pI of MPG is 10) and size exclusion chromatography, (Fig. 4.9). The MPG protein eluted as a monomer in the size exclusion chromatography, corresponding to its molecular weight of 35 kDa. The eluted fractions of gel filtration show the protein is about 95% pure, (Fig. 4.9c).



4.9. Purification of MPG. (**a**) Cation exchange chromatography of MPG (**b**) Size exclusion chromatography of MPG (**c**) 12% SDS-PAGE gel for the eluted fractions of size exclusion.

4.4.2 Expression and purification of MCAF1Δ8.

MCAF1Δ8 (1154-1270) , fused with MBP at the N-terminus in a pET Duet-1 modified vector and with a 6xHis tag at the N-terminus in the pET-M vector, was expressed in BL21 (DE3) cells. The MBP tagged protein was purified in two steps, affinity purification using amylose and size exclusion, (Fig. 4.10). The protein eluted as a monomer at its molecular weight of 55 kDa in an S200 column and was concentrated.

a **b**

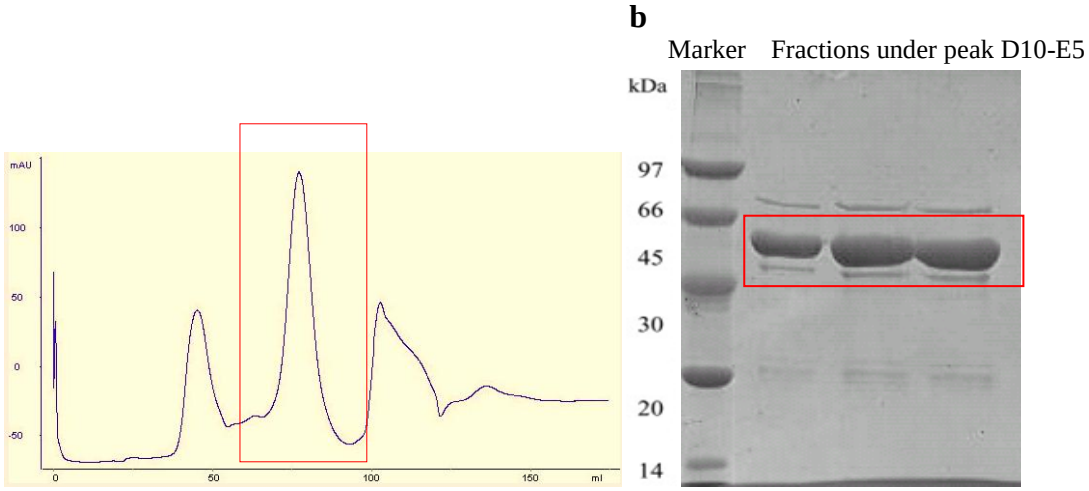
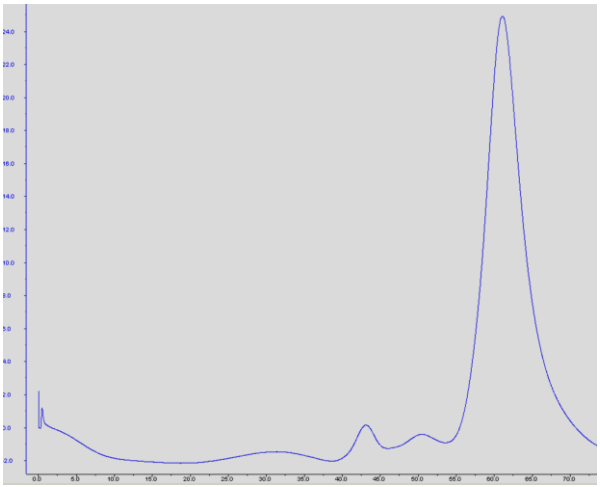


Figure 4.10. Purification of MCAF1Δ8 with n-terminal MBP. (**a**) Size exclusion chromatography (**b**) 12% SDS-PAGE of elution fractions.

MCAF1Δ8 with an N-terminal His tag was purified in two steps: Ni-NTA affinity purification and size exclusion chromatography, (Fig. 4.11). The protein eluted as a monomer, corresponding to its molecular weight 15 kDa, in gel filtration and the fraction confirm the protein is 95% pure.

a



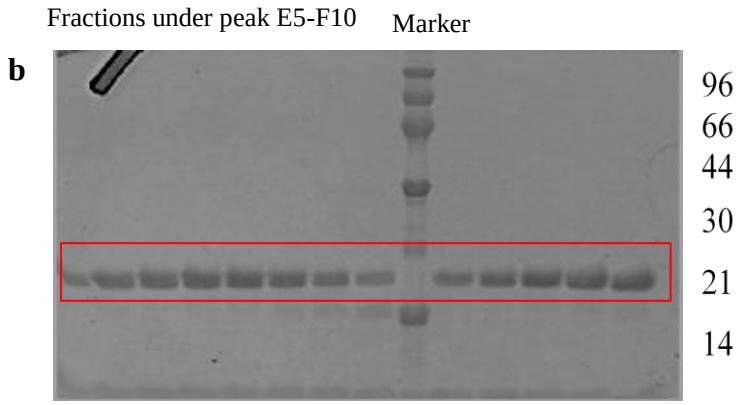


Figure 4.11. Purification of MCAF1Δ8 with N-terminal 6x His tag. (**a**) MCAF1Δ8 eluted as monomer in size exclusion (**b**) SDS-PAGE shows the protein is highly pure.

4.4.3 HSQC titration for TRD interaction with MCAF1Δ8 and MPG

MBD1 is known to interact with MCAF and MPG through its TRD. We added unlabeled MBP:MCAF, MBP:MCAF1Δ8 or MBP alone (as a control) to ^{15}N labeled TRD at molar ratios of 1:0 to 1:3. The combined chemical shift perturbation and peak intensities of the residues were plotted. As there was negligible peak shift (< 0.03 ppm), even at excess 1:3 molar ratio, (Fig. 4.12), the intensity ratio per residue for bound to free TRD was analyzed. Now, apart from an overall loss in peak intensities across the protein sequence, a continuous region of residues between Thr574 and Ser597 disappeared which coincides within the TRD domain. This disappearance of residues is due to line-width broadening by interaction, suggesting a slow tumbling of TRD and MBP-MCAF1Δ8 complex (Fig 4.12). Moreover, the disappearance is oblivious in the presence of MBP alone. This is in accordance with the earlier study of TRD interaction with MCAF1 through Ile576 and Leu579 (Ichimura, Watanabe et al. 2005). Interestingly, two previously uncharacterized but highly conserved charged residues, R583 and R585 of TRD were also found to be interacting with MBP: MCAF1Δ8-.

Unlabeled full length MPG was titrated against ^{15}N labeled TRD in the molar ratio of 1:0 to 1:3. A general attenuation in peak intensities per residue was observed but to a lesser extent compared to MCAF1Δ8 **(**Fig. 4.13**)**. Significant intensity loss was observed for Glu545, Asp554, Trp590 and Leu591 of TRD. Thus we could infer that TRD interacts with both MCAF1 and MPG but in a different manner. More importantly, TRD remains intrinsically unstructured upon binding, indicating the interactions are relatively very weak.

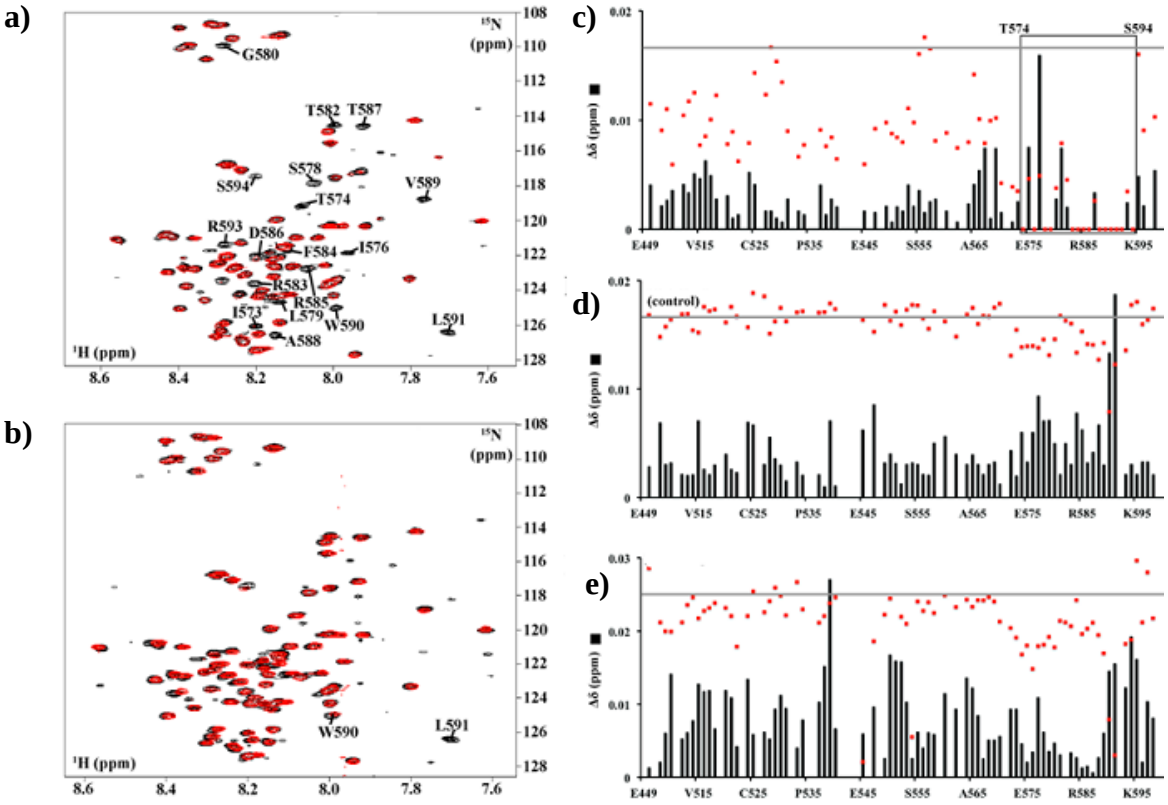
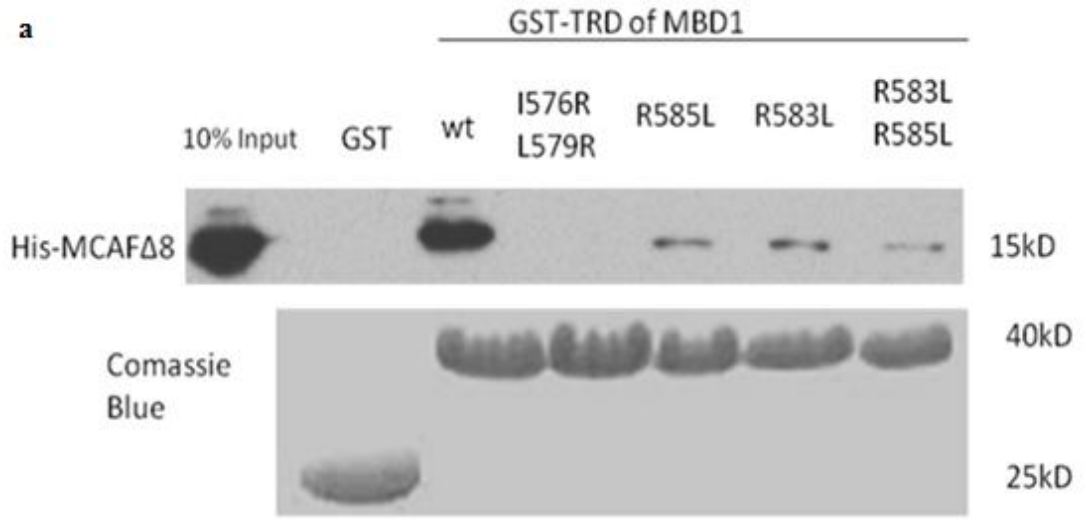


Figure 4.12. ^{15}N HSQC-TROSY of TRD titrated with protein. Titration was performed with the ^{15}N TRD at molar ratios of 1:0 (black) and 1:3 (red). (**a**) with MBP:MCAFΔ8 and (**b**) with MPG. Chemical shift perturbation (Δδ) and intensity ratios of TRD (**c**) with MBP-MCAFΔ8 and (**d**) MBP alone, as a control (**e**) with MPG. The peak intensity ratio after and before titration is based on ^{15}N-HSQC-TROSY. Gain/loss of interaction will lead to decrease/increase of peak intensity ratio, respectively. For any absence or similar interaction in free and bound TRD, the peak intensity ratio would be 1 (gray line). Selected disappeared residues in TRD domain upon adding MBP-

MCAFΔ8/ MPG are labeled in (**a**) and (**b**) and boxed between Thr574 and Ser594 in (**c**). The residues are numbered according to MBD1 isoform 1.

4.4.4 GST pull down for TRD with MCAF1Δ8 and MPG

Following the interaction study results, GST pull down was performed, (Fig. 4.13a). The GST pull-down results support the NMR interaction studies. Binding of MPG to TRD is weaker than that of MCAF1Δ8. The binding between TRD to MCAF1Δ8 and MPG is mostly due to the stretch of residues 571 to 592 of TRD where peak weakening occurs. The interaction between TRD and MCAF1Δ8 is not only due to the hydrophobic interaction involving Ile576 and Leu579 but also due to the charged residues Arg583 and Arg585, which are highly conserved, (Fig. 4.7). From the pull-down results, it could be seen that mutation of the charged residues to hydrophobic residues leads to loss of binding. This clearly shows that the binding is highly specific and in the case of MPG, mutation of even one residue causes loss of interaction, (Fig 4.14 b)**.**



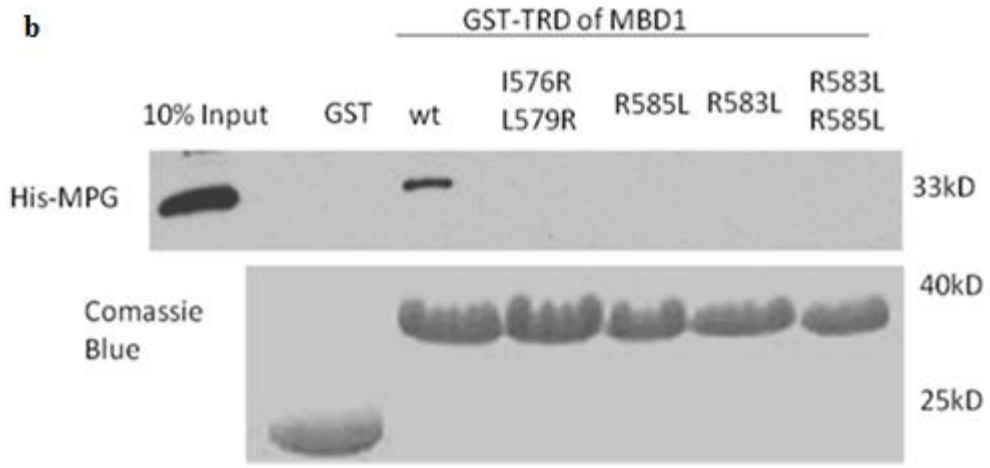
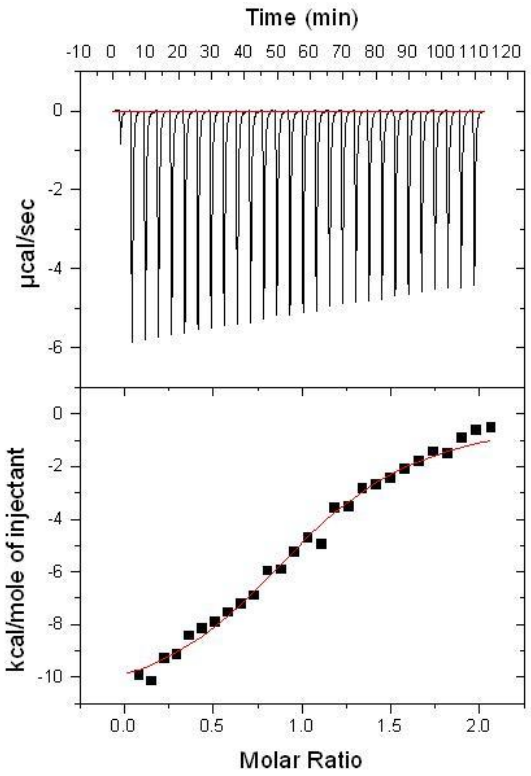


Figure 4.13. GST pull down of MCAF1Δ8 and MPG with TRD. (**a**) GST and GST- TRD proteins were immobilized on glutathione-agarose beads and incubated with the 6xHis: MCAF1Δ8. Wild type (wt) and mutant (I576R, L579R, R583L and R585L) TRD proteins were tested. The input indicates 10% of the indicated proteins in the reaction mixture. (**b**) As in panel (**a**) but with 6xHis:MPG.

4.4.5 Isothermal titration calorimetry (ITC)

The interaction of TRD with MPG and MCAF1Δ8 was analyzed using ITC to determine affinity and the stoichiometry ratio. Earlier studies have indicated that the interaction is relatively weak and could be only transient in nature.. The K_d of TRD-MPG interaction was 10 μM, with a stoichiometry close to 1 (Fig. 4.14a). The K_d of TRD-MCAF1Δ8 interaction was 10 μM with a stoichiometry of 0.9 (Fig. 4.14b). Other relevant interaction parameters are given in Table 4.1.

a



b

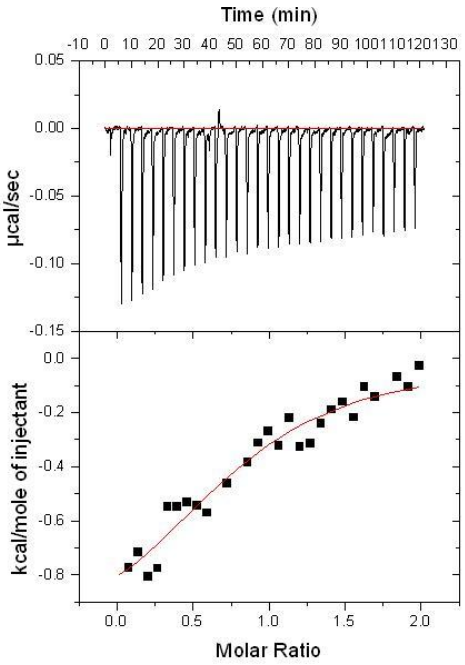


Figure 4.14. Interaction of TRD with MPG and MCAF1Δ8 using ITC. (a) TRD titrated against MPG. MPG did not saturate, but the heat exchange for each injection was recorded **(b)** TRD titrated against 6xHis:MCAF1Δ8-His. Similar to the previous case, the protein did not saturate.

Table 4.1. Summary of the Values recorded in the ITC between TRD and MPG, also with MCAF1Δ8

No	Protein	N	K_a (10^5 M^{-1})	K_d μM	ΔH (kcal/mol)	TΔS (kcal/mol)	ΔG (kcal/mol)
1	MPG	1.044 ± 0.0220	1.277 ± 0.1612	7.8	-11.132 ± 0.0355	-4.353	-6.779
2	MCAF1Δ8	0.8386 ± 0.1004	1.127 ± 0.4244	8.8	-0.1139 ± 0.02024	5.751	-5.8649

CHAPTER 5. DISCUSSION

MBD1 is an essential cell signaling protein required to regulate neuronal cell differentiation and division (Liu, Teng et al. 2010). Its over expression is often linked to tumorigenesis by recruiting transcriptional repressor machinery, which unfortunately suppresses the tumor suppressor genes that are required to control invariably proliferating cells (Nakao, Matsui et al. 2001; Sakamoto, Watanabe et al. 2007). Presently it is known that MBD1, as a transcriptional repressor, requires its transcription repressor domain (TRD) for interaction with different transcriptional repressor proteins in either a histone deacetylase dependent or independent manner. MBD1 could repress transcription in HDAC dependent manner in case of Acute Promyelocytic Leukemia through HDAC3 (Villa, Morey et al. 2006) and also in HDAC independent manner through MCAF1/AM family proteins to form heterochromatin (Ichimura, Watanabe et al. 2005). Other studies reveal that MPG is also part of such heterochromatin formation through its interaction with MBD1 via TRD (Watanabe, Ichimura et al. 2003). It has been shown earlier that cell signaling regulated by MBD1 involves the TRD domain. Earlier results have also clearly shown that MBD1ΔTRD (lacking TRD) is not efficient in transcriptional silencing albeit the other domains such as the DNA binding MBD domain and the CXXC domain are intact (Ng, Jeppesen et al. 2000). Thus TRD is a key part of MBD1 to act as a repressor and thus it is necessary to study its structure and mode of recognizing different binding partners.

MCAF1 consists of two conserved domains, allowing the protein to interact with transactivator Sp1 and MBD1. MCAF1 can also act as an activator when it binds to transactivator Sp1 but at the same time, it acts as a transcriptional repressor when it

binds to MBD1 by further recruiting SETDB1 to form a strong heterochromatin complex (Fujita, Watanabe et al. 2003). The C-terminus of MCAF1 (MCAF1Δ8) is involved in interaction with TRD and a TRD deletion mutant of MBD1 is not able to bind MCAF.

Here, NMR spectroscopy is used to elucidate both the structure of TRD and probe its interaction with MPG and MCAF1Δ8. We have found the TRD domain to be intrinsically unstructured. Besides the TRD domain, which is highly conserved among the seven isoforms (except isoform v6), we used constructs with 22 residues before the TRD domain. Notably, these 22 residues (Asp507 to Lys528) are only conserved among the four isoforms: v1, v5, v7 and v9. Next, an additional 13-residue sequence after the TRD domain is also highly conserved but not in isoforms v2 and v6, (Fig. 5.1a). Upon addition to MPG and MCAF1Δ8, an overall intensity loss across the TRD 99-residue sequence was observed after correcting the dilution effect. Interestingly, a previously uncharacterized stretch of residues (Thr574 to Ser594) disappeared in the presence of MPG. The same region also showed intensity loss in the presence of MCAF1Δ8 but to a lesser extent. Coincidentally, this 21-residue region lies within the 64-residue TRD domain and is rich in hydrophobic residues. Compared to the adjacent highly acidic region (Val537 to Ile573), both MPG and MCAF1Δ8 reduce the peak intensities more within this 21-residue hydrophobic region, bearing significantly fewer but conserved acidic (Glu575 and Asp586) and basic (Arg583 and Arg585) Residues. Moreover, there was no great loss in intensity in another hydrophobic region (Asp507 to Ala536). Ichimura and co-workers have previously shown that mutation of two residues (Ile576 and Leu579) of TRD to basic Arg residues clearly abolished MBD1 binding to MCAF (Ichimura, Watanabe et al. 2005). We further investigated the need for more favorable hydrophobicity.

Surprisingly mutating two basic residues to hydrophobic residues, R583L and R585L, did not enhance TRD binding to both MPG and MCAF1Δ8 but there was a significant loss in binding. To check whether the double mutant of charged residues has the same structure as the wild type, a ^{15}N HSQC spectrum of the R583L/R585L mutant protein was collected and it showed similar well-resolved peak pattern as the native TRD, suggesting no protein aggregation, (Fig. 5.1b). The key differences are only the appearance of two new peaks in the place of the mutated R583L and R585L residues, and peak shifts of residues Ile576 and Leu579, which were only minimally perturbed due to local chemical changes near the point mutations (Fig. 5.1a). Additionally, both Arg583 and Arg585 are not conserved in isoform v6 only, suggesting a unique recognition motif within the primary binding site (TRD domain) for the remaining six isoforms. Hence, the newly identified charge-charge interactions of TRD with MPG and MCAF1Δ8, besides the previously reported Ile576 and Leu579, are crucial. Furthermore, the formation of the TRD-MCAFΔ8 or TRD-MPG complexes was further confirmed by GST pull-down assays.

The need for this charge-charge interaction within the primary TRD domain can be attributed to compensate the lack of conformational fit. Perhaps, this lack of globular fold may enable the unstructured TRD domain to act as a recognition domain to sample a wider gallery of binding proteins and fueled by electrostatic forces. Notably, the TRD-protein interaction is weak and does not induce visible conformational changes in TRD. The weak interaction is seen by the negligible chemical shifts in ^{15}N HSQC and line-width broadening, suggesting a slow tumbling complex. The unique mode of partner proteins recognition through weak but specific interaction may enable MBD1 to shuffle between at least two different functional forms for regulating a cobweb of cell functions, primarily through its TRD domain.

While MBD1 is an important cell signaling protein to maintain normal cell function, it plays a significant role at the same time during tumor progression in a promoter methylation dependant manner. It is obviously essential to control its activity and this may be achieved by targeting its interaction with other repressor complex proteins, particularly through its TRD domain.

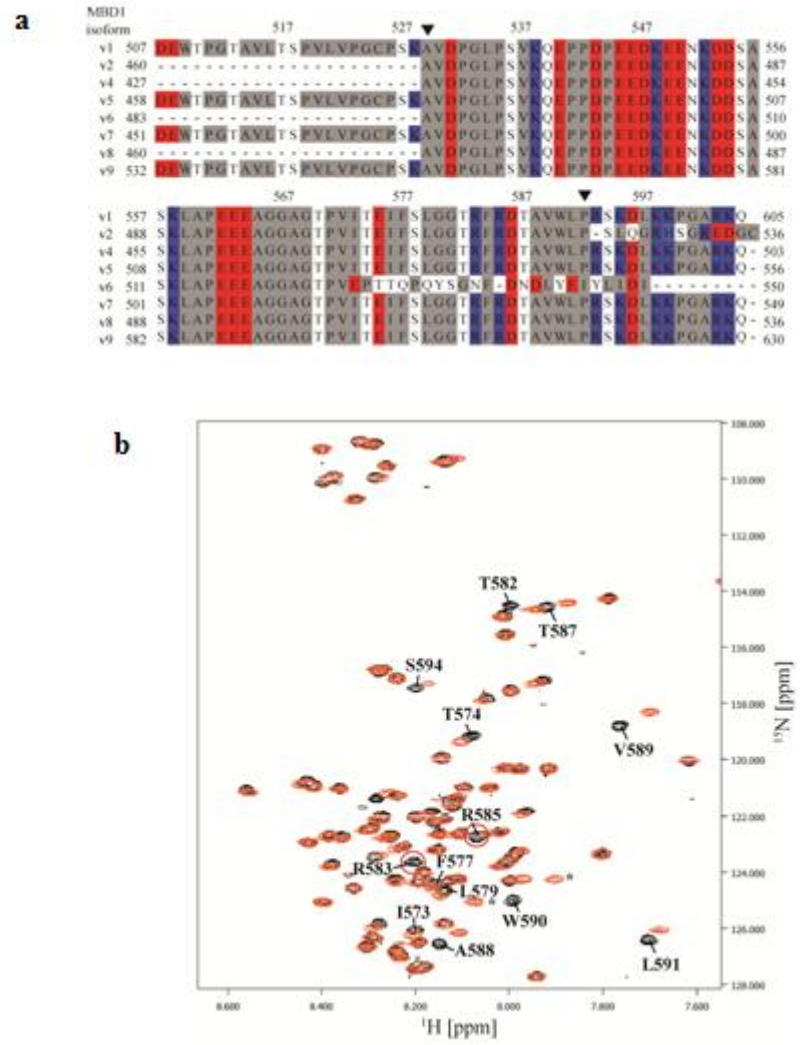


Figure 5.1. Conservation of TRD between MBD1 isoforms and the HSQC Spectra for the Double mutant. (**a**) Multiple sequence alignment of TRD of MBD1 isoforms (**b**) ^{15}N HSQC spectra for the arginine double mutant (R583L/R585L) to confirm there is no structural change.

We could design peptide inhibitors corresponding to the core residue stretch (Thr574 to Ser594) that interacts with MCAF1, MPG and HDAC3. Targeted silencing of these 21 residues to prevent the interaction between TRD and its binding partners is an interesting hypothesis and will be of high therapeutic interest.

CHAPTER 6. MITOCHONDRIAL DYSFUNCTION PROTEIN A (MIDA)

In this part of the thesis, I will discuss about the structure of the MidA protein, determined by X-ray crystallography. MidA has a methyltransferase domain, which binds to co-factors and a substrate binding domain. From our structure we could come up with the important residues of MidA that are involved in binding co-factors required for methylating substrate. MidA is an important member of the mitochondrion, since it helps in proper Mitochondrial Complex I formation and is also presumed to methylate one of the sub-units in Complex I.

6.1 THE MITOCHONDRION

The mitochondrion is an important sub-cellular organelle and plays a major role in energy metabolism. Its outer membrane contains high concentrations of porins, which facilitate the permeability of solutes, whereas the inner membrane is impermeable to metabolites and ions (Jonckheere, Smeitink et al. 2012). A family of about 50 mitochondrial carriers has been discovered in the human genome which enable exchange between the inter-membrane space and matrix (Kunji 2004). The inner membrane contains protein complexes and redox cofactors that are involved in electron transfer and ATP synthesis (Boyer 1997). The mitochondrial matrix, encapsulated by the inner membrane, contains a mixture of enzymes that are involved in all aspects of metabolism, in addition to the mitochondrial genome, with 39 genes in mammals, known to be important for mitochondrial function.

The mitochondrion is known to form highly dynamic and interconnected tubular networks throughout the cell, fusing and dividing, and undergoing regulated

turnover, enabling it to readily adapt to changes in the cellular environment (Egner, Jakobs et al. 2002). Disruption of mitochondrial dynamics results in cellular dysfunction and several major neurodegenerative pathologies, including Parkinson's, Alzheimer's and Huntington's diseases. Remarkably, in several disease models, the manipulation of mitochondrial fusion and fission can partially rescue disease phenotypes (Chen and Chan 2009).

In contrast to the traditional 'baffle model' of mitochondrial structure, Fig 6.1 shows how invagination of cristae from the mitochondrial inner membrane and throughout the matrix produces compartments that are likely to impede the lateral diffusion of metabolites (Mannella 2006). These cristae junctions have great importance in the regulation of metabolic processes within the mitochondrion, and also in other processes that are key to cellular biochemistry, such as apoptosis. The model reveals the point of contact between the inner and outer mitochondrial membranes, postulated to be the sites of large macromolecular assemblies involved in metabolite transport into and out of the mitochondrion (Frey and Mannella 2000).

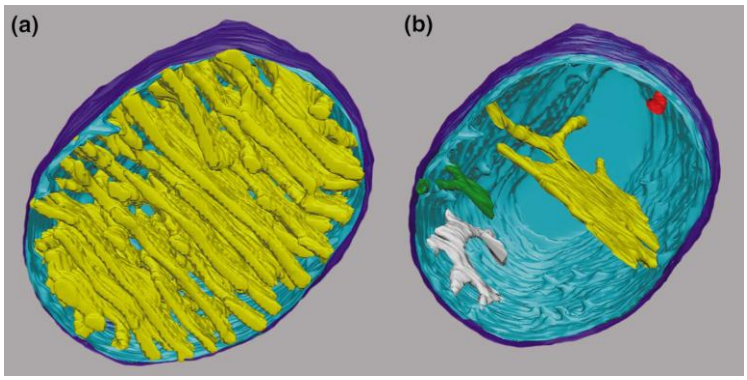


Figure 6.1 The Mitochondrion. Computer models generated from segmented 3D tomograms of a mitochondrion in chick cerebellum. (**a**) The entire model showing all cristae in yellow, the inner boundary membrane in light blue, and the outer membrane in dark blue. (**b**) Outer membrane, inner boundary membrane and four representative

cristae in different colors. Figure adapted from (Frey and Mannella 2000).

6.2 MITOCHONDRIAL FUNCTIONS

6.2.1 Energy transduction

Energy transduction in the mitochondrion is carried out by five protein complexes located in the inner membrane. A flavin mononucleotide, which is non-covalently bound to NADH:ubiquinone oxidoreductase (complex I), is reduced by the low potential electrons of NADH. Subsequently the electrons are passed down to a series of redox centres, approximately arranged in the order of increasing potential, until they are used up to reduce molecular oxygen. Thus released energy is used to pump protons across the impermeable mitochondrial inner membrane into the inter-membrane space. This proton motive force is used by ATP synthase to synthesize ATP from ADP (Jonckheere, Smeitink et al. 2012). In addition to oxidative phosphorylation, the mitochondrion also plays an important role in upstream of the electron transfer chain reactions, including the citric acid cycle. Recent studies clearly show that, impairment in the components of the citric acid cycle could also result in tumor formation (Raimundo, Baysal et al. 2011).

6.2.2 Apoptosis

It is well known that the mitochondrion plays a central role in apoptosis, which is important in many physiological and pathological processes. The Bcl-2 family of proteins controls the cell death by regulating the permeability of the mitochondrial outer membrane and the release of pro-apoptotic factors. The formation of mitochondrial apoptosis induced channels is important for the regulation of apoptosis by the Bcl-2 family of proteins. Pro-apoptotic members, such as Bax and

Bak, promote the formation of channels, whereas anti-apoptotic members inhibit the formation of channels (Rodriguez, Rojas-Rivera et al. 2011).

6.3 MITOCHONDRIAL COMPLEXES

The mitochondrial respiratory chain (RC) assembly consists of, as shown in Fig. 6.2, four membrane-bound (Complex I, II, III and IV) multimeric RC complexes (RCCs) and function as a catalyst for the oxidation of reducing equivalents, importantly of nicotinamide adenine dinucleotide (NADH), using oxygen (O_2) as terminal electron acceptor in the inner mitochondrial membrane (IMM) (Lenaz and Genova 2010). The electron transfer mediated by the RC is connected to ATP synthase (complex V), the final complex that generates ATP. The coupling between ATP synthesis and RC is called oxidative phosphorylation (OxPhos) (Lemarie and Grimm 2011).

Reactive oxygen species (ROS) are generated if electron leakage occurs in RC (Lemarie and Grimm 2011). In many cancer cells, deficiencies in RC are frequently observed mainly due to numerous mutations in the mitochondrial genome (mtDNA), which directly affect the subunits in RCC I, III and IV (Polyak, Li et al. 1998), and in the nuclear-coded components of complex II (Hensen and Bayley 2011), whose assembly factors are also found to be mutated (Ghezzi, Goffrini et al. 2009). These alterations lead to specific defects of the respiratory chain.

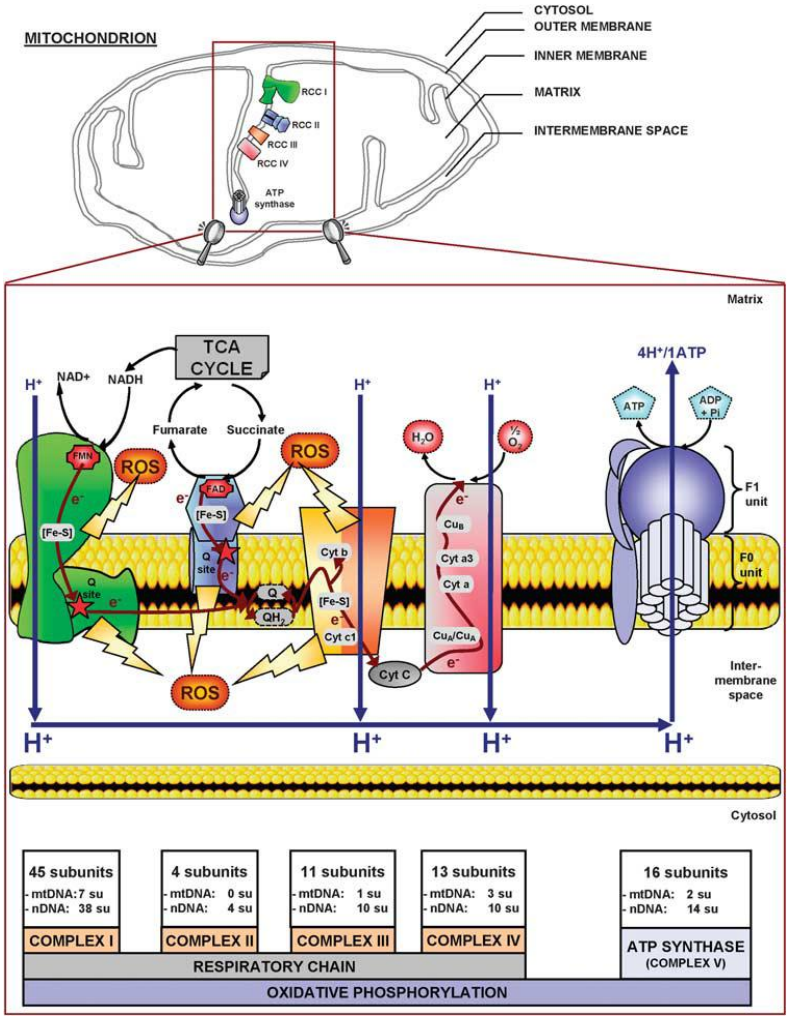


Figure 6.2. Respiratory Chain Complex. Representation of the mitochondrial respiratory chain complexes and the oxidative phosphorylation system. Figure adapted from (Lemarie and Grimm 2011).

6.3.1 Mitochondrial complex I

Complex I (CI), comprising of NADH:ubiquinone oxidoreductase or NADH dehydrogenase, is the largest of the mitochondrial respiratory complexes, with 45 subunits. Seven subunits are encoded by the mtDNA while others are encoded by the nuclear DNA. CI is involved in the catalysis of the transfer of electrons from NADH, which is produced during the tricarboxylic acid cycle (TCA) to coenzyme Q. A recent effort to study the three-dimensional structure using electron microscopy and

crystallographic studies reveal that the structure of the complex resembles an ‘L’ with the presence of a peripheral hydrophilic arm (matrix arm) and a highly hydrophobic membrane arm, (Fig 6.3) (Tocilescu, Fendel et al. 2007; Clason, Ruiz et al. 2010).

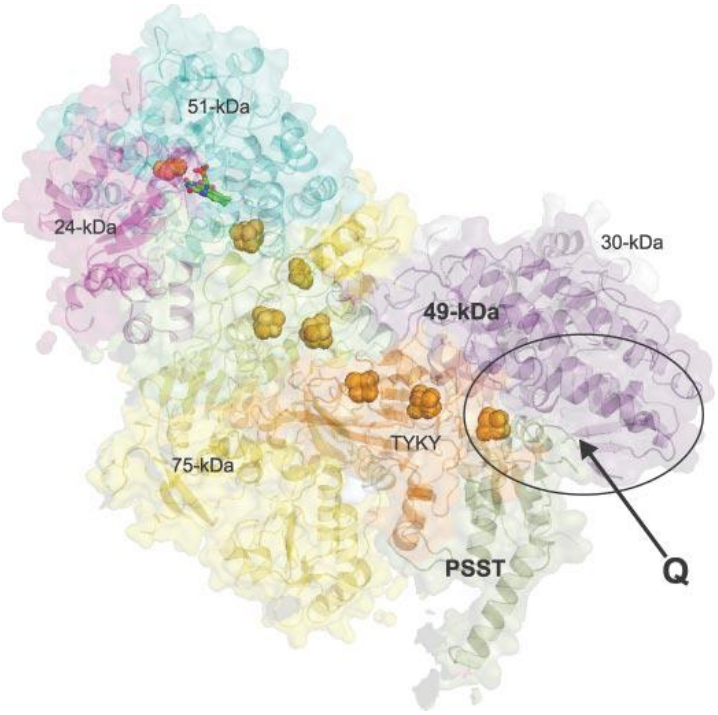


Figure 6.3. **Overall structure of the peripheral arm of *T. thermophilus* complex I.** It shows the wire of iron-sulfur clusters connecting the NADH-binding site near FMN with a broad cavity formed by the PSST and the 49-kDa subunit that should comprise the active site for ubiquinone reduction and the binding region for the large number of inhibitors that have been found for complex I. Figure adapted from (Tocilescu, Fendel et al. 2007).

The assembly of this enzyme is quite elaborate, and even though many steps have been identified in mutants of *Neurospora crassa* and by studies of patients with CI defects, the complete sequence of events need to be determined. Various models, which vary in details but with several common features, have been proposed for the assembly of the holoenzyme. The complex is assembled in several modules that go on to form the peripheral and membrane arms (Mckenzie and Ryan 2010; Perales-

Clemente, Fernández-Vizarra et al. 2010). Patients will have severe manifestations during birth and early childhood whenever there are mutations of nuclear origin leading to CI defects, which very often lead to premature death (Kirby, Crawford et al. 1999). Various mutations in CI nuclear and mtDNA-encoded subunits have been reported and part of the CI deficiency cases are due to mutations in the structural subunits (Distelmaier, Koopman et al. 2009). CI defects are also due to mutations in auxiliary proteins, which do not form part of the final holoenzyme but assist in the assembly process (Küffner, Rohr et al. 1998). In addition to the Complex I, other complexes such as Complex II, III, IV and V plays important roles in the oxidative phosphorylation which controls various steps of the electron transport chain (Lemarie and Grimm 2011).

6.2 METHYL TRANSFERASES

A protein methyl transferase enzyme employs a common bimolecular nucleophilic substitution (S_N2) methyl transfer mechanism (Zhang and Bruice 2008; Smith and Denu 2009). The lone pair electrons of a nitrogen atom (either from lysine or arginine) 'attacks' the electrophilic methylsulphonium cation of SAM (S-Adenosyl-L-Methionine) at an angle of 180° to the leaving group, to form a penta-coordinate carbon transition state, which then collapses. The methyl group gets relocated to the nitrogen atom of the lysine or arginine side chain and forms *S*-adenosyl-l-homocysteine (SAH; also known as AdoHcy) as a product, (Fig 6.4).

The two highly studied methyl transferases, owing to their importance in cell signaling are protein lysine methyl transferase (PKMT) and protein arginine methyl transferase (PRMT). The common structural feature of these two types, which differentiates them from other proteins that use SAM, is confined to the overall

architecture of their extended catalytic active sites. They generally consist of a SAM-binding pocket that can be accessed from one face of the protein and a narrow and hydrophobic acceptor (lysine or arginine) channel, which extends to the opposite face of the protein surface, in a way the two substrates enter the active site from opposite sides of the enzyme surface (Cheng, Collins et al. 2005).

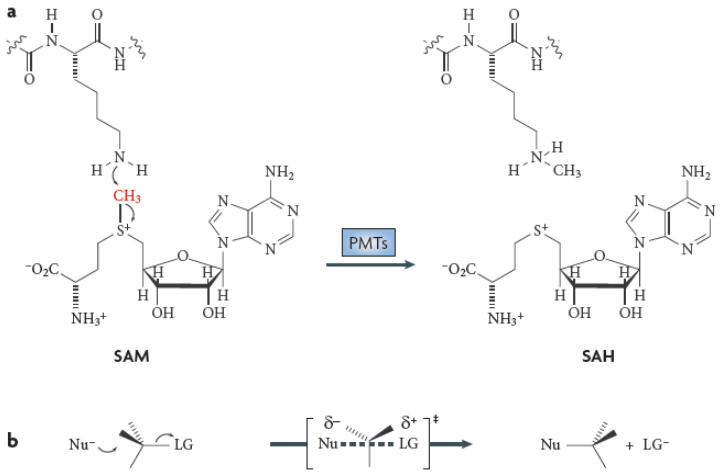


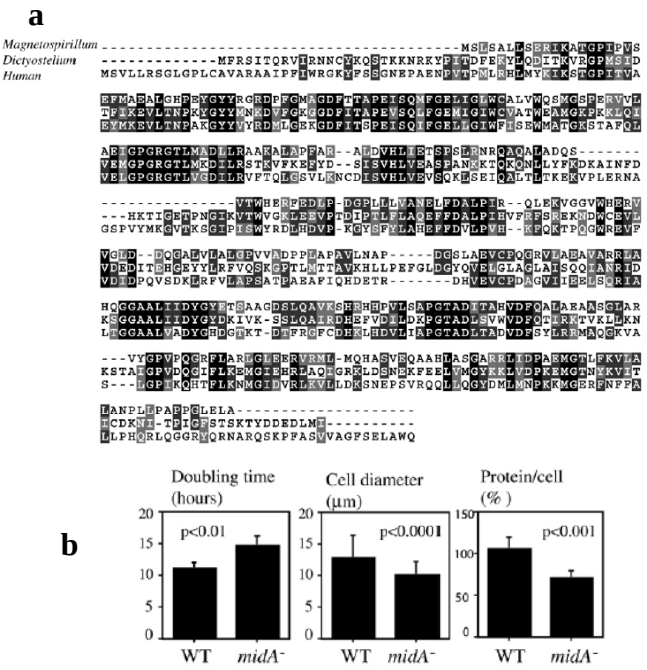
Figure 6.4. Protein methyl transferase. PMT- catalyzed methylation of proteins by an S_N2 reaction with SAM as the methyl donor**.** (**a**) The methyl group of the SAM is 'attacked' by the lone pair electrons of a lysine side-chain nitrogen atom and it results in transfer of the methyl group to the attacking nitrogen atom and the production of SAH from the reaction cofactor. (**b**) A more generalized chemical scheme of a bimolecular nucleophilic substitution (S_N2) group transfer. Figure adapted from (Copeland, Solomon et al. 2009).

6.3 MITOCHONDRIAL DYSFUNCTION PROTEIN A (MIDA)

6.3.1 MidA belongs to an uncharacterized protein family

MidA was first identified in *Dictyostelium* from a group of 30 uncharacterized proteins with highly conserved sequences, (Fig. 6.5a). MidA is important for the normal cell functioning of *Dictyostelium*, since the knockdown of MidA resulted in pleiotropic effects. A large part of mutant cells had a diameter of 8-12 μm, but wild-

type cells displayed a wider distribution and also a significant proportion of cells were more than 12 μm in diameter (Fig. 6.5b). In supportive to the data, the amount of total protein per cell was also found to be to 70% less than that of wild type (Torija, Vicente et al. 2006). Macropinocytosis responsible for most of fluid-phase uptake (Hacker, Albrecht et al. 1997), and phagocytosis accounts for the uptake of particles in bacteria (Maniak 2001). In the mutant cells, both the processes were down regulated as shown in (Fig 6.5c). The level of uptake of fluorescent particles and fluorescent soluble material was found to be less in the mutant, after 30 minutes of incubation (Torija, Vicente et al. 2006). Flow-cytometry also confirmed that the number of cells that were able to ingest multiple fluorescent microspheres was reduced in *midA*⁻ cells, (Fig. 6.5d). Even though MidA was initially thought to be an uncharacterized and non-functional protein, the above results clearly show the conservation from bacteria to human and knock out of MidA results in defects of physiological functions in *Dictyostelium*. Thus it can be inferred that MidA plays an important role in a particular cell organelle (Torija, Vicente et al. 2006).



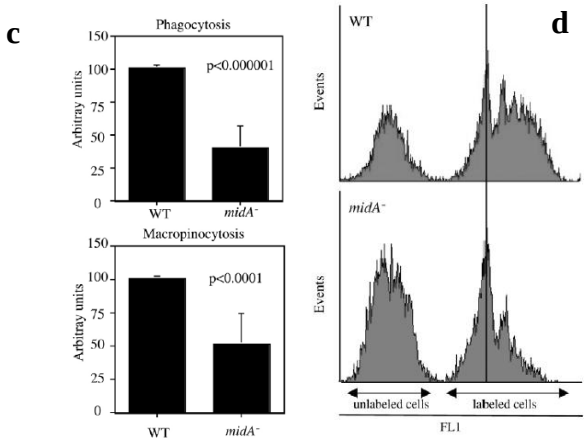


Figure 6.5. Knock out of MidA in *Dictysotelium*. (**a**) Multiple Sequence Alignment of MidA from bacteria to humans (**b**) Cell diameter of mutants enlarged compared to wild type (**c**) Phagocytosis and Macropinocytosis is reduced in mutant cells (**d**) The uptake of fluorescent labeled materials is reduced in mutant cells. Figure adapted from (Torija, Vicente et al. 2006).

6.3.2 Loss of MidA results in mitochondrial dysfunction

The localization of MidA in cells was determined by the expression of green fluorescent protein (GFP) fused to the C-terminal end of the *Dictyostelium* full length MidA protein. The expression of this construct was supported by the actin15 promoter. Furthermore, transfection into *midA*$^{-}$ cells has shown recovery in growth rate in bacteria (Fig. 6.6). The recovery strains were monitored with confocal microscopy for GFP fluorescence and the GFP distribution showed a punctuate pattern, which co-localized with the mitochondrial specific marker MitoTracker Red (Fig 6.6).

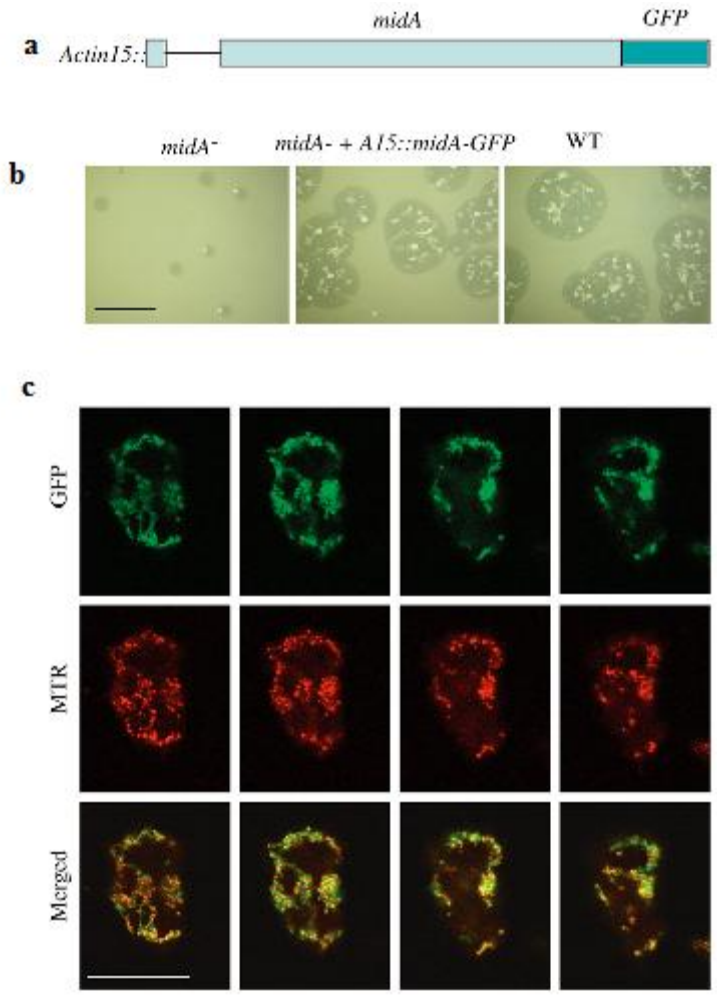


Figure 6.6. Localization of MidA in mitochondria. (**a**) The construct used for expression of GFP fused to MidA. (**b**) Effect of deletion of MidA in mutant and rescue with MidA expression (**c**) The localization of GFP fused MidA and mitochondrial specific marker (MTR), where Series of 2μm confocal images from the same cell show GFP location in green and the mitochondrial marker in red. Figure adapted from (Torija, Vicente et al. 2006).

The reason for such defects in mutants in mitochondrial function was analyzed using MitoTracker Red (CMXRos), which get incorporated to active mitochondria and in a way that depends on the mitochondrial membrane potential (MMP) (Pendergrass, Wolf et al. 2004). The cells that were stained with these markers can be

observed by fluorescent microscopy and quantitatively analyzed fluorimetry. There was no change in the level of intensity in staining of these markers, suggesting that the amount of mitochondria per cell and the MMP was not affected in the mutant. The cellular oxygen consumption was also found to be unaffected in *midA*$^{-}$ cells (Fig. 6.7). Interestingly, the amount of ATP per cell was about 70 % less than that of the wild type which could be the probable reason for the mitochondrial defect.

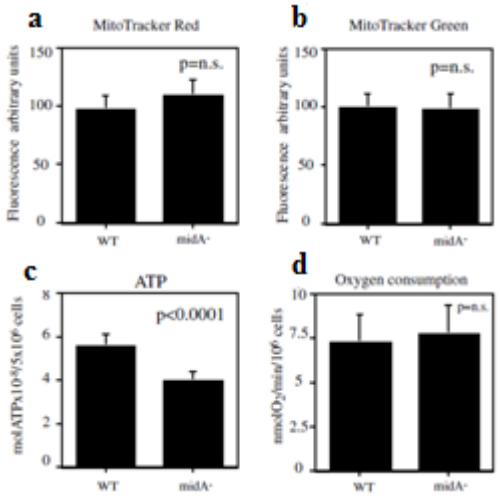


Figure 6.7. Effects of MidA knock out in mitochondria. (**a**) Mitrotracker red level marks active mitochondria (**b**) the level of mitochondrial membrane potential in mutant and WT (**c**) the level of ATP production in mutant and WT (**d**) the level of oxygen consumption in mutant and WT. Figure adapted from (Torija, Vicente et al. 2006).

6.3.3 MidA is important for complex I formation in mitochondria

The importance of MidA in complex I formation was analyzed by measuring the level of activity in the OXPHOS complexes (I, II, III and IV) in *midA*$^{-}$ null *Dictyostelium*. About 50% decrease in the activity of CI in *Dictyostelium* was observed (Fig. 6.8). It was quite interesting to see that the activity of other complexes remain unaffected and or significantly higher in the mutant (Carilla-Latorre, Gallardo

et al. 2010). This could be due to the compensatory response to the loss of CI activity. In support, there was an increase in the level of mtDNA and most mtRNA transcripts in the mutant (Fig 6.8). Also an insignificant increase in the level of citrate synthase was observed in the mutant cells.

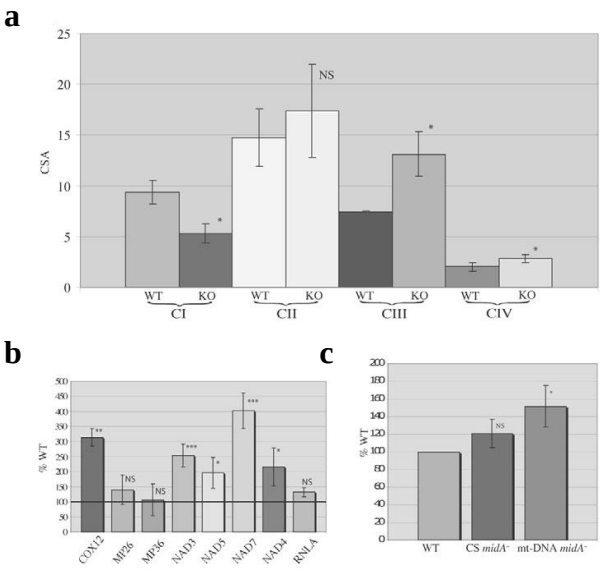


Figure 6.8. The activity of complex I is reduced in cells lacking MidA. (**a**) The level of activity of the OXPHOS complexes in the WT and Knockout cells. (**b**) The level of mtRNA transcripts level in mutant cells (**c**) The expression level of citrate synthase in mutant and WT, and also mtDNA in mutant. Figure adapted from (Carilla-Latorre, Gallardo et al. 2010).

The level of complex I assembly was analyzed using blue native (BN) Poly Acrylamide Gel Electrophoresis (PAGE) in a stable transfection cell line of HEK293T in which the human MidA protein was down regulated by the shRNAmir (micro RNA adapted shRNA) technology. HEK293T cells transfected with the scramble vector was used as a control. Quantification by densitometry of three independent experiments is shown, (Fig. 6.9), and it was seen only the clones S1 and S19 produced a significant reduction in the level of MidA and because of that there was only a modest (60 - 80%) of CI assembly was observed. The activity of CI,

determined by spectrophotometric analysis, is also reduced to 75%. These results confirm the involvement of MidA in complex I assembly or stability (Carilla-Latorre, Gallardo et al. 2010).

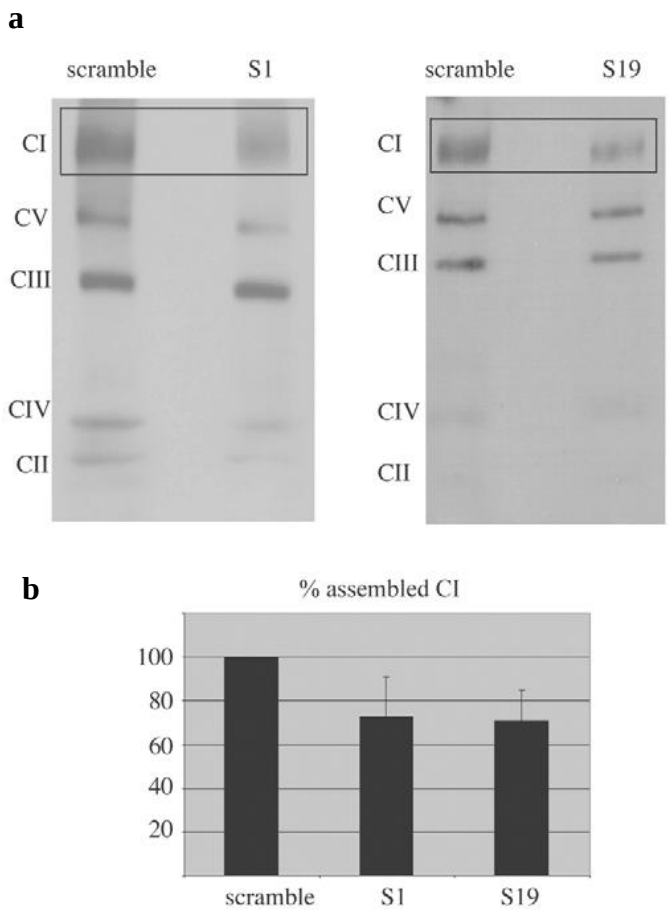


Figure 6.9. Down regulation of MidA results in reduction of fully assembled complex I. (**a**) Blue-native gel shows the level of complex assembly in the cell, where MidA is knocked down. S1 and S9 mark different shRNA clones. (**b**) Spectrophotometric analysis of Complex I assembly in the knock down cells (Figure adapted from : (Carilla-Latorre, Gallardo et al. 2010).

6.3.4 MidA interacts with ndufs2 subunit of mitochondrial complex I

To identify the proteins that interact with MidA, a yeast two hybrid screen was carried out using the whole *Dictyostelium* MidA protein, except for the first 21 amino acids, as a bait. About ten independent positive clones were identified which encode

Dictyostelium NADH-ubiquinone oxidoreductase chain 49, a homologue of the human complex I subunit NDUFS2. Fig. 6.10 shows the common region that is present in different clones which helped in restraining the minimal interaction region to 40 amino acids. To validate this result, pull down assays were performed and the N-terminal fragment of *Dictyostelium* NDUFS2, which contains the minimal region of interaction, was fused to GST, incubated with cell extracts from *Dictyostelium* expressing MidA, which was fused to GFP. MidA-GFP was pulled down by GST-NDUFS2 and even the human MidA was pulled down. This again confirms the interaction between MidA and NDUFS2 as well as the functional conservation between *Dictyostelium* and human MidA proteins (Carilla-Latorre, Gallardo et al. 2010).

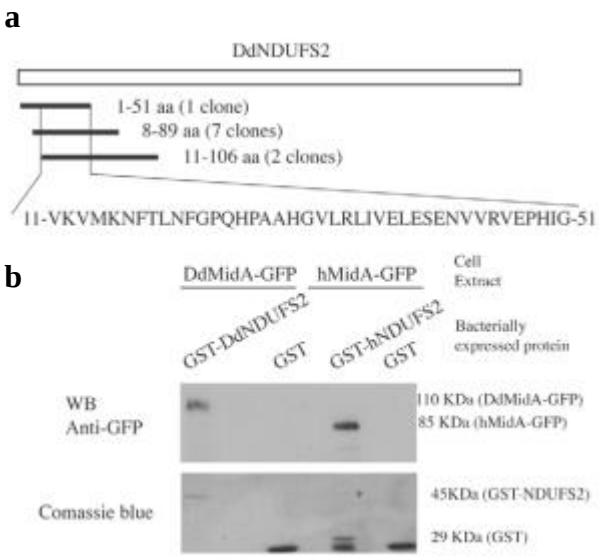


Figure 6.10. MidA interacts with NDUFS2. (**a**) The clones that were found to be positive in a yeast-two hybrid screen. (**b**) GST Pull down of *Dictyostelium* and human MidA by NDUFS2. Figure adapted from (Carilla-Latorre, Gallardo et al. 2010).

6.3.5 MidA has a methyl transferase domain

The MidA sequence was compared to all known structures from the Protein Data Bank. Even though the MidA protein is of unknown function, the structural similarity showed that it has a methyl transferase domain similar to the human dimethyl adenosine transferase, which belongs to the family of S-adenosyl-L-methionine (SAM) dependent transferases. Based on homology modeling, the important residues involved in SAM binding were speculated to be Gly170, Glu200 and Asn257. Furthermore, G170V and G172V mutants were prepared and the effect of these mutations in *Dictyostelium* was studied. Both residues have previously been suggested to be important for SAM binding, and are well conserved among methyltransferases (Niewmierzycka and Clarke 1999).

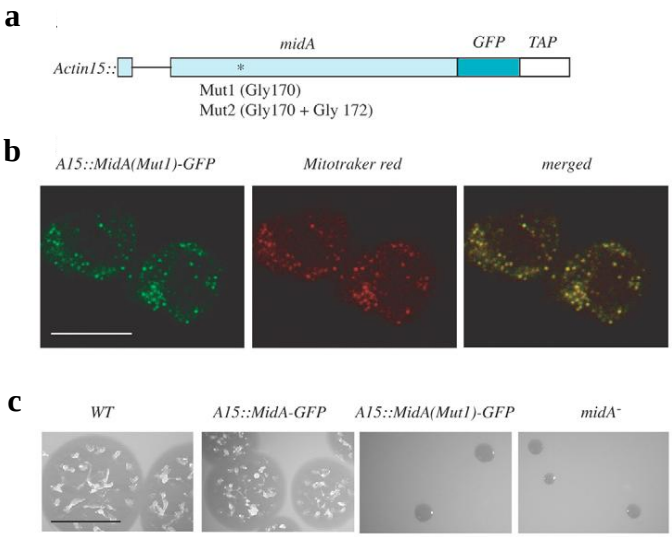


Figure 6.11. Importance of SAM binding residues of MidA. (**a**) Schematic diagram of the mutation and the lone (**b**) Localization of mutant in the mitochondrion. (**c**) WT MidA protein is able to complement in the MidA$^-$ null cells but the G170V mutant protein failed (Carilla-Latorre, Gallardo et al. 2010).

The WT and mutated forms of MidA were fused to GFP to visualize their expression and localization and transformed into the *Dictyostelium* MidA$^-$ null mutant (Fig.

6.11). As known earlier, MidA$^-$ cells showed defects in growth in both axenic medium and in association with bacteria, as a consequence of mitochondrial dysfunction (Torija, Vicente et al. 2006). The WT protein was able to complement the phenotype completely, but the G170V mutant protein was unable to complement the phenotype. The lack of reversing the phenotype confirms that the single mutation in Gly170 from the methyltransferase domain is enough to make the protein ineffective. These results suggest that the putative methyltransferase domain is required for MidA function in *Dictyostelium*.

6.4 OBJECTIVES

1. To determine the crystal structure of MidA and its complex with SAM and SAH using X-ray crystallography
2. To validate the structure by performing site-directed mutation of important residues and verified by biophysical studies such as ITC
3. To model MidA complex with NDUFS2 and validate it by experiments
4. To determine the importance of MidA as a methyl transferase in the *Dictyostelium* cells

CHAPTER 7. MATERIALS AND METHODS

7.1 SUBCLONING

The gene encoding residues 76 to 484 amino acids was PCR amplified using Pfu DNA polymerase (Fermentas) with the forward primer ATTGGATCCAAA CAAGTTTTCAAAAATAG containing a BamHI site (underlined) and the reverse primer ATTCTCGAGTTAAATCATTAAATCTT CATCATCATAAG with an XhoI site and sub-cloned into the pGEX6P-1 vector with N-terminus GST tag.

7.2 PROTEIN EXPRESSION AND PURIFICATION

The MidA protein was over-expressed using the pGEX 6P-1 vector as an N-terminal GST-fusion protein. Seed cultures of transformed *E. coli* BL21 (DE3) competent cells were added to large scale LB medium containing 100 μg/ml ampicillin. The cell density was monitored until A_{600nm} was 0.6 to 0.7, before adding 0.15 mM IPTG for protein over-expression for another 18h at 16 °C. Cells were harvested and cell pellet was stored at -80 °C. The cell pellet was resuspended in binding buffer A: 50 mM Tris-HCl buffer, 150 mM NaCl, 2mM DTT, 5% Glycerol, 0.1% Triton X-100 and its final pH adjusted to 8.5. The cell suspension was lysed by sonication on ice and the supernatant was loaded onto a GST column (GE Healthcare) and subsequently allowed to bind for 4 h at 4 °C. Weakly bound proteins were removed by extensive washing using buffer A for 3 to 4 times. The GST tag was cleaved off with Prescission Protease overnight using 1unit/mg of GST:MidA at 4 °C. Thus purified protein was passed through a Superdex 200 column (GE Healthcare) equilibrated with 20 mM HEPES, 200 mM NaCl and 3 mM DTT, pH adjusted to 8.0. MidA eluted as a single homogenous peak, at its theoretical Mol. Wt. of 45 kDa. The

purified protein was concentrated to 5 mg/ml using ultra filtration membrane (Merck Millipore) with 10 kDa Mol. Wt. cut off. The same protocol was followed for the seleno-methionine derivative protein.

7.3 CRYSTALLIZATION, DATA COLLECTION AND STRUCTURE DETERMINATION

The purified MidA protein was mixed with cofactors S-adenosyl-L-methionine (SAM) and S-adenosyl-homocysteine (SAH) in a ratio of MidA:SAM (or SAH) 1:20 and setup for crystallization using commercial screens using the hanging drop method in VDX plates (Hampton Research) where 1 µl of MidA, MidA:SAM or MidA:SAH was mixed with 1 µl of the corresponding reservoir solution and allowed to equilibrate against 1 ml of the reservoir solution at room temperature. Crystal appeared in a few hours and grew to maximum size of 0.1-0.2mm in one day. Initial crystals were obtained for MidA alone and MidA:SAH complex in Wizard Screen III (Emerald Biosystems) contain 10% PEG8K, 0.1 M HEPES pH 7.5 and 8% ethylene glycol and for MidA:SAM complex in PEGRx 2 (Hampton Research) containing 10% PEG10K, 0.1 M HEPES pH 7.5 and 0.2 M sodium potassium tartrate. The seleno-methionine derivative protein was also crystallized using the same condition.

Crystals were cryoprotected in the reservoir solution supplemented with 20% glycerol and flash cooled at 100 K. Since MidA has 11 methionines, the structure was solved by the multi wavelength anomalous dispersion (MAD) method (Terwilliger and Berendzen 1997). X-ray diffraction data were collected at beamline 13B, National Synchrotron Radiation Research Centre (NSRRC), Taiwan using a Quantum-315r CCD area detector (ADSC) and processed with HKL2000 (Otwinowski and Minor 1997). All Se sites in the asymmetric unit were located for the MidA:SAH complex

using the program Phenix-Autosol (Terwilliger, Adams et al. 2009). Phases were further improved by density modification using RESOLVE (Terwilliger 2003). Over 50% of the backbone atoms of the model were built by RESOLVE. The remaining residues were manually built using COOT (Emsley and Cowtan 2004), followed by refinement using phenix-Refine (Adams, Afonine et al. 2010). Refinement was continued until the R-value converged to 0.20 (R_{free} = 0.255) for reflections I>σ (I) to 2.8 Å. For determination of MidA and MidA-SAM structures, molecular replacement was performed using the program Phaser-MR (Simple Interface) from phenix (McCoy, Grosse-Kunstleve et al. 2007) with the MidA:SAH structure used as the template model and refined using Phenix-Refine until the R value converged to 0.214 (R_{free} = 0.274) for MidA with reflections I>σ (I) to 2.8 Å resolution and for MidA:SAM complex until R was 0.206 (R_{free} 0.251) for reflections I>σ (I) to 3.0 Å. About 4.97% of reflections were used for R_{free} calculation.

7.4 PROTEIN-PROTEIN DOCKING FOR MIDA WITH NDUFS2

MidA is known to interact with the NDUFS2 subunit of mitochondrial Complex I (Carilla-Latorre, Gallardo et al. 2010), primarily at the N-terminus of NDUFS2. As no experimental structure is available for NDUFS2, its 3D model was built using the CPH model 3.0 server (Nielsen, Lundegaard et al. 2010). This model was then docked against the MidA structure using the Cluspro 2.0 protein-protein docking server (Comeau, Kozakov et al. 2007; Kozakov, Hall et al. 2010). The output of Cluspro was analyzed using the CONTACT program of CCP4. The residues within 3 Å contact distance are considered to be probably involved in interaction.

CHAPTER 8. RESULTS

8.1 PROTEIN EXPRESSION AND PURIFICATION

Residues 76-484 (of 484 aa) of the MidA protein, also known as NADH dehydrogenase (ubiquinone) complex I assembly factor 7 was cloned into the pGEX 6P-1 vector and overexpressed in BL21 (DE3) cells. The overexpressed protein was purified by two step purification, first with GST, following which the N-terminal GST tag was cleaved overnight at 4 °C using Prescission protease, and then by size exclusion chromatography. The protein was eluted as a monomer, corresponding to 45 kDa (Fig. 8.1). The fractions were verified by SDS-PAGE, Fig. 8.2.

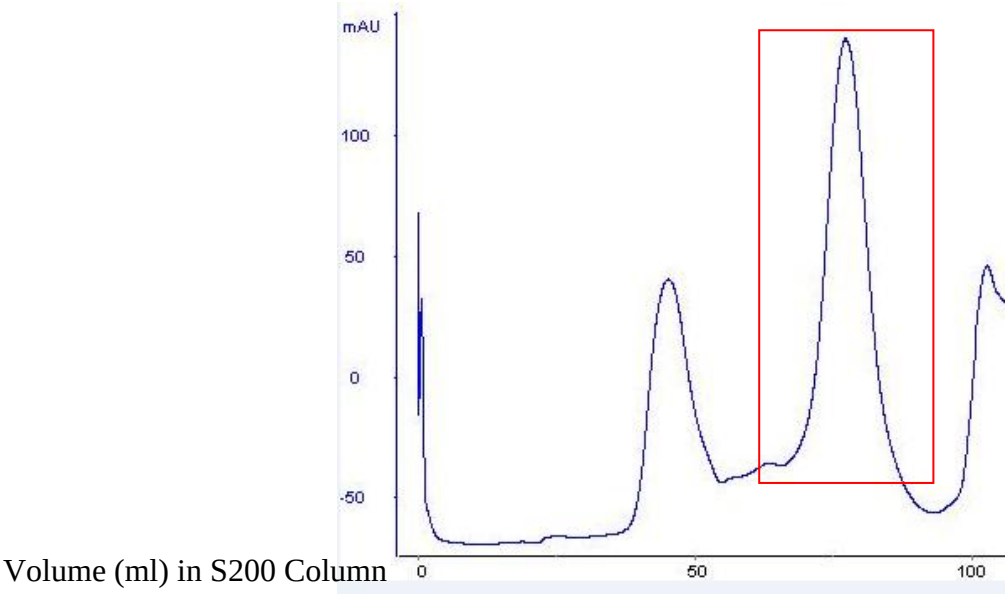


Figure 8.1. Size exclusion chromatography of MidA.

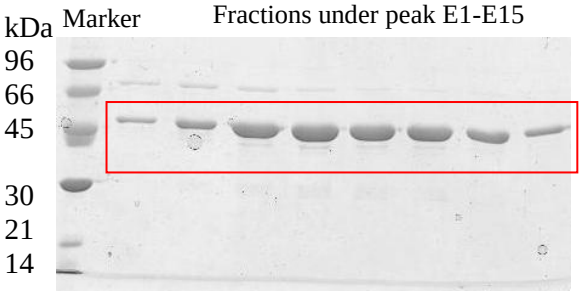
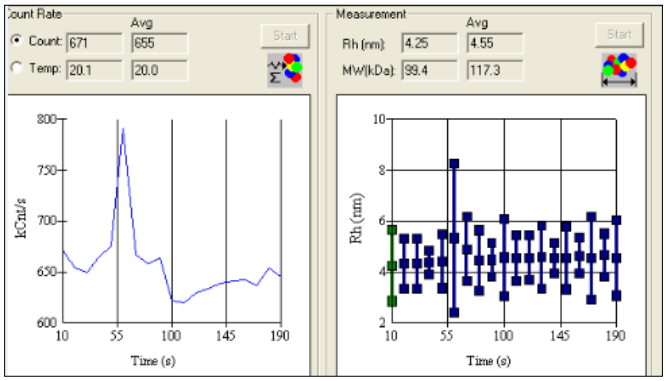


Figure 8.2. SDS-PAGE of the fractions of Fig. 8.1.

The MidA protein was concentrated to 5 mg/ml and its homogeneity was analyzed using dynamic light scattering (DLS). The results confirm that MidA is homogenous with a polydispersity index of 9% (limit is 20%), radius 4.25 nm and SOS error 5.51 (limit is 50), Fig. 8.3.



Msr#	Time[s]	Temp[C]	Count Rate	Ampl	Diff Coeff	Radius[nm]	Polyd[nm]	PolydIndx	MW[KDa]	%Mass	Baseline	Sos Error
1	10.0	20.1	671115	0.706	505.	4.25	1.41	0.11	99.4	100.0	1.000	4.54
2	20.0	20.1	653398	0.732	496.	4.32	0.980	0.05	103.	100.0	1.000	3.00
3	30.0	20.0	649682	0.765	495.	4.32	0.971	0.05	103.	100.0	1.000	2.19
4	40.0	19.9	664082	0.773	489.	4.36	0.469	0.01	106.	100.0	1.001	3.29
5	50.0	20.0	675408	0.790	485.	4.42	1.06	0.06	109.	100.0	1.000	2.97
6	60.0	20.0	790668	0.760	401.	5.33	2.95	0.31	169.	100.0	1.004	38.7
7	70.0	20.0	666517	0.754	437.	4.89	1.27	0.07	138.	100.0	1.004	14.9
8	80.0	20.0	657696	0.743	482.	4.44	1.19	0.07	110.	100.0	0.999	3.40
9	90.0	20.0	662941	0.718	478.	4.47	0.653	0.02	112.	100.0	0.999	1.87
10	100.0	20.0	622062	0.623	470.	4.55	1.51	0.11	117.	100.0	1.001	2.68
11	110.0	20.0	619907	0.628	471.	4.54	0.888	0.04	116.	100.0	1.000	2.55
12	120.0	20.0	628987	0.633	470.	4.55	0.876	0.04	117.	100.0	1.000	2.56
13	130.0	20.0	633676	0.648	467.	4.58	1.23	0.07	118.	100.0	1.000	2.43
14	140.0	20.0	638298	0.653	471.	4.54	0.608	0.02	116.	100.0	1.000	3.00
15	150.0	20.0	640694	0.664	470.	4.55	1.23	0.07	117.	100.0	1.000	2.44
16	160.0	20.0	641923	0.678	462.	4.64	0.704	0.02	122.	100.0	1.000	2.89
17	170.0	20.0	636619	0.681	472.	4.54	1.63	0.13	116.	100.0	0.999	5.40
18	180.0	20.0	654063	0.694	459.	4.66	0.839	0.03	123.	100.0	1.000	3.04
19	190.0	20.0	645431	0.697	470.	4.55	1.49	0.11	117.	100.0	1.000	2.79
Aves:												
Mono		20.0	655429	0.702	471.	4.55	1.15	0.07	117.	100.0	1.000	5.51
Bi-1		20.0	655429	0.214	948.	2.65	----	----	40.5	67.2	1.000	1.47
Bi-2				0.493	397.	6.08			339.	32.8		

Figure 8.3. DLS of MidA at 5mg/ml.

8.2 CRYSTALLIZATION AND STRUCTURE DETERMINATION

8.2.1 Crystallization

MidA was concentrated to 5 mg/ml and then screened for crystallization using various commercial screens. Several conditions gave crystals but with poor diffraction. The seleno-methionine derivative protein gave a crystal in a condition containing 8% ethylene glycol, 0.1 M HEPES pH 7.5 and 10% PEG 8000 at room

temperature in about 16 to 24 hours (Fig. 8.4). Cryo-protected (with 25% glycerol) crystals diffracted to 3.0 Å resolution. Also, MidA was mixed with 20 times excess S-adenosyl-homocysteine (SAH), which is a cofactor, and setup for crystallization. MidA:SAH crystal, produced at the same condition, diffracted to 2.8 Å.

a) **b)**

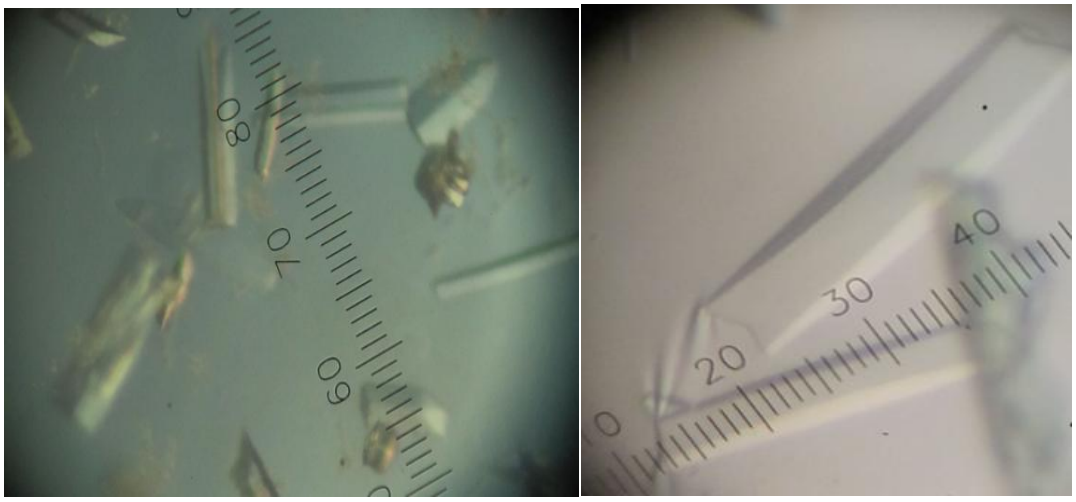


Figure 8.4. Crystal of MidA. (**a**) Crystals of apo MidA and (**b**) MidA:SAH.

8.2.2 Structure determination

MidA has 11 methionines and the structures of Se-Met apo MidA and MidA:SAH complex were solved by the single wavelength anomalous dispersion (SAD) method (Hendrickson et al., 1990). The MidA:S-Adenosyl-L-Methionine (SAM) complex structure was solved by molecular replacement using the MidA:SAH structure as the starting model (Table 8.1 and Fig. 8.5). Electron density for the N-terminal residues 124 to 133 was not well defined and the residues were not modeled. The asymmetric unit consists of three MidA molecules, without any local non-crystallographic symmetry (NCS). Refinement details for apo MidA and the MidA:SAM complex structures are given in Table 8.2.

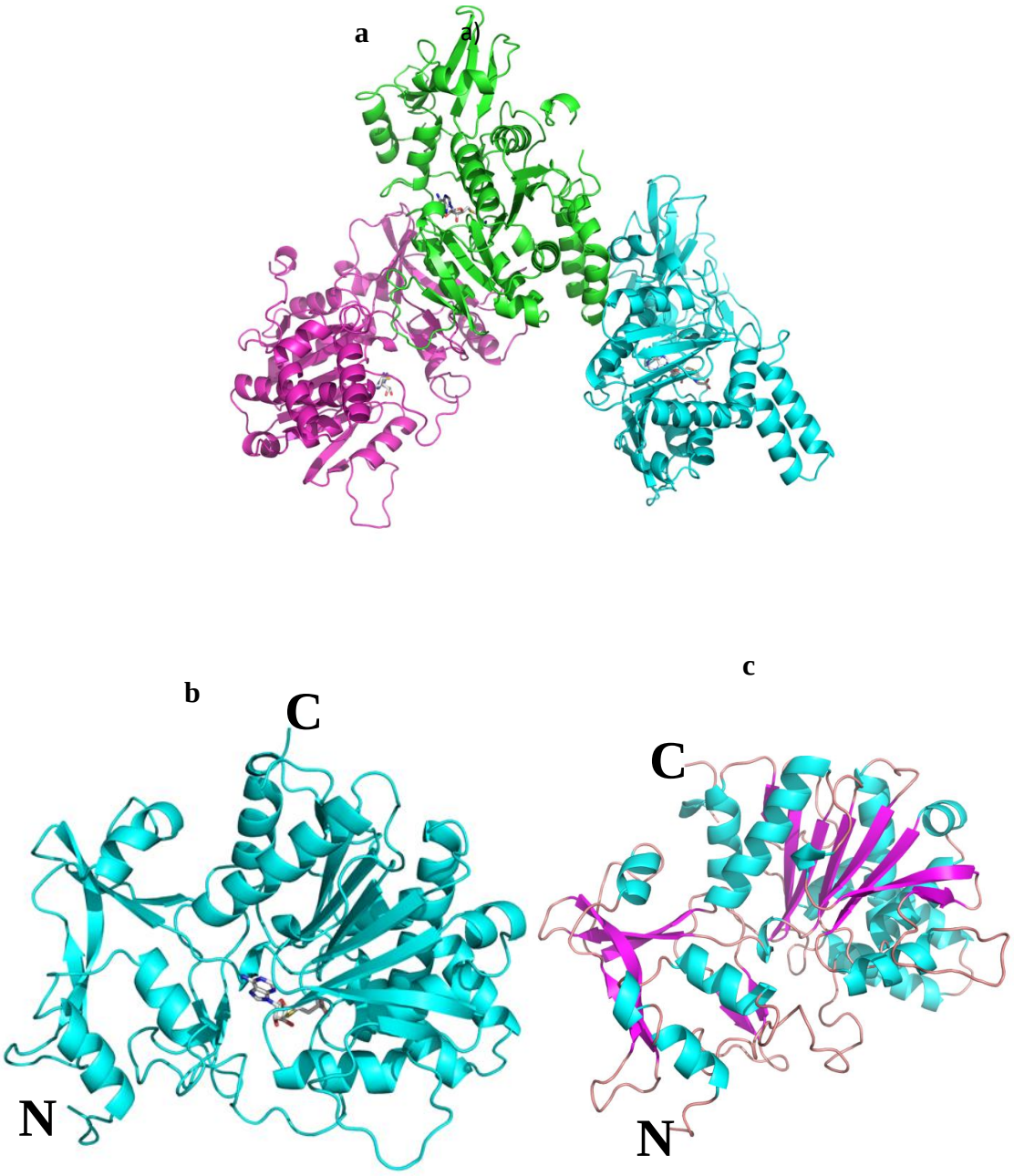


Figure 8.5. Crystal structure of MidA. (**a**) The three MidA:SAH molecules in the asymmetric unit (**b)** MidA:SAH monomer and (**c**) apo MidA molecule.

MidA adopts a Class I methyltransferase fold consisting of a parallel seven-stranded β sheet and a pair of α-helices (Fig. 8.6). A search for structurally similar proteins within the PDB database was performed with DALI (Holm, Kääriäinen et al.

2008), (Table 8.1). MidA shows the highest structural similarity to certain proteins, whose functions are still unknown. However, they are likely to be methyl transferases. Furthermore, MidA matches with the carboxy methyl transferase-1, which methylates the terminal carboxyl group of leucine of protein phosphatase 2A (PP2A) in humans. Based on sequence and structural homology, MidA is proposed to be a macromolecular methyltransferase, which could methylate either proteins or DNA or RNA molecules in cell signaling.

The central region, residues Thr135 to Leu263, binds the co-factors SAM or SAH. In addition, residues Ile265 to Gly344, which also interact with the N-terminal helices, could be the region that binds substrate. According to DALI, MidA shares structural similarities with other Rossmann Fold Methyltransferase (RFM) proteins with predicted or determined methyltransferase activity. There are evidences that MidA is as an important assembly factor in mitochondrial complexI NADH dehyrogenase (ubiquinone) complex I assembly factor 7 (NDUF7), in addition to its possible methyltransferase activity on selected components of complex I.

8.2.3 MIDA interaction with SAM and SAH

The ligands in MidA:SAM and MidA:SAH complexes display a similar mode of binding (Fig. 8.7). The highly conserved consensus motif hh(D/E)hGxGxG, where h represents a hydrophobic residue and x can be any residue, mostly involved in SAM binding (Niewmierzycka and Clarke 1999) is present in the loop between residues Ile166 to Gly174. This loop connecting a β-strand to its successive α-helix forms the bottom of the cofactor-binding site, while residues Thr135, Gly170, Pro171, Gly172, Glu200, Asn205, Gln257 and Phe259 delineate the cofactor-binding site. Residues Gly170 and Glu200 are highly conserved in class I SAM dependent MTases and are

involved in cofactor binding. The hydroxyl group of Thr135 interacts with the carboxylic group of SAH or SAM and Glu200 coordinates both hydroxyls of the ribose moiety. In particular, the side chain of Gln257 (like its equivalent in other related enzymes) coordinates the exocyclic amino group of the SAM adenine moiety, while Gly170, Gly172 and Asn205 are engaged in hydrogen bonds with the cofactor via their backbone. The adenine moiety is stacked towards Phe259, where the hydrophobic benzene ring forms stacking interaction with the hydrophobic indole group of the adenine ring and Pro171 helps in formation of a pocket to hold the co-factors. In total, Thr135, Gly170, Gly172, Glu200, Asn205 and Gln257 will form six hydrogen bonds with SAM/SAH. The cofactor binding site in apo-MidA is highly similar to that of the MidA–SAM/SAH complex.

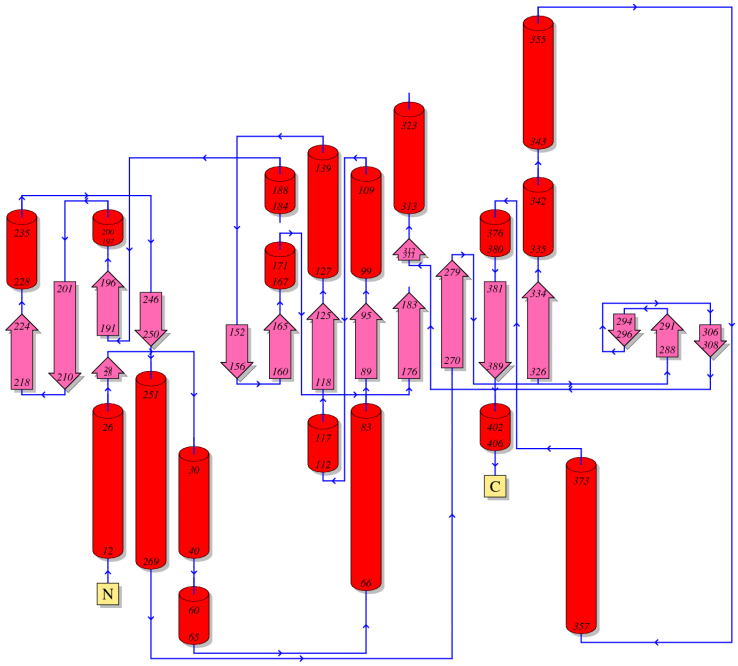


Figure 8.6. Topology Diagram of MidA. α-helices are shown as red cylinders while β-strands are shown as pink arrows. The figure was drawn with the PDBsum online server.

Table 8.1. Structural alignment of the MidA protein.

DALI match	PDB ID	Z-Score	RMSD (Å)	Length (aa)	Identity (%)	Organism	Protein
1	1zkd	42.0	2.0	349	35	*Rhodopseudomonas palustris*	Q6N1P6
2	4q67	37.1	2.7	348	24	*Burkholderia thailandensis*	Likely methyltransferase
3	3mnt	13.8	3.6	204	11	*Homo sapiens*	Leucine carboxyl methyltransferase-1
4	2zzk	13.6	3.3	228	10	*Saccharomyces cerevisae*	Leucine carboxyl methyltransferase-2
5	3bgv	13.6	3.6	209	9	*Homo sapiens*	mRNA cap guanine-N^7 methyltransferase
6	1ri5	13.4	3.1	198	12	*Encephalitozoon cuniculi*	mRNA capping enzyme
7	2zwa	13.2	3.4	227	10	*Saccharomyces cerevisae*	Leucine carboxyl methyltransferase-2
8	3ha5	13.2	3.5	201	9	*Mycobacterium Tuberculosis*	Methoxy mycolic acid synthase 4
9	2vdw	13.0	3.7	228	10	*Vaccinia virus*	Capping enzyme

Table 8.2. Data collection and refinement statistics of MidA structure.

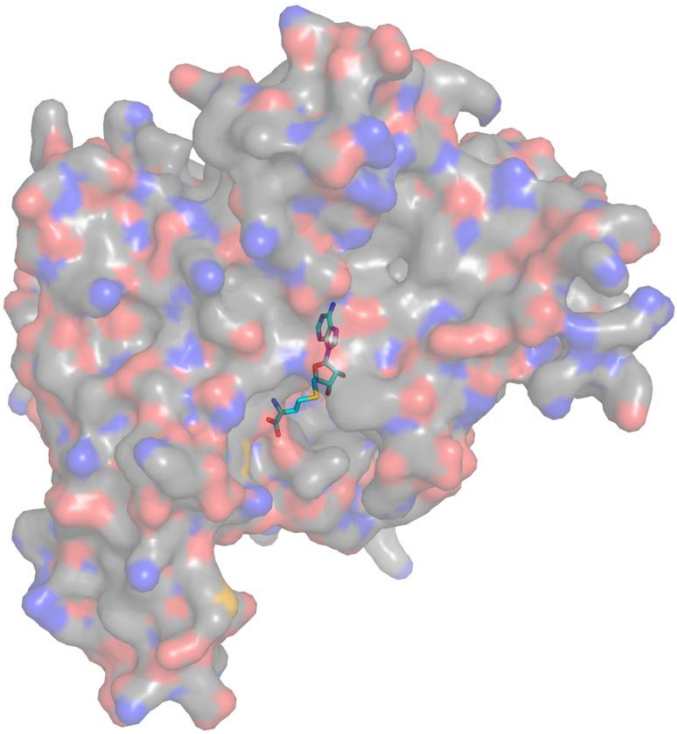
Data collection	Native	SAM	SAH
Unit cell (A):	a=72.82 b=105.39, c= 201.57	a=72.85 b=106.04, c= 200.77	a=73.04 b=105.64, c= 201.12
Space group:	P2 2 2 1 1 1	P2 2 2 1 1 1	P2 2 2 1 1 1
Resolution range (A):	50.0-2.8 (2.85-2.80)	50.0-2.9 (2.95-2.90)	50.0-2.80 (2.90-2.80)
Wavelength(A):	0.97879	0.97864	0.96357
Observed reflections:	421912	256356	541083
Unique reflections:	38951	35124	38959
Completeness (%):	99.7(98.9)	100 (100)	100(99.9)
Overall (I/σI):	12.5	11.3	11.0
Redundancy:	10.8(9.5)	7.3(7.0)	13.9(11.8)
R Sym:	0.17	0.18	0.18
Refinement and quality			
Resolution range:	32.0-2.8	32.0-2.90	32.0-2.80
R work:	0.248	0.244	0.190
R free:	0.275	0.278	0.242
RMSD bond length (A)	0.009	0.009	0.010
RMSD bond angles (Degree)	1.097	1.144	1.365
Average B factors Main Chain: Side Chain and waters:	38.24 39.38	35.03 36.34	45.68 48.42
Ramachandran Plot Most Favoured regions (%) Additionally allowed regions (%) Generously allowed regions (%) Disallowed regions (%)	93.6 4.01 2.34 0	92.4 6.18 1.42 0	92.19 6.05 1.76 0

[a]$R_{sym}=\sum|Ii - <I>I|/\sum|Ii|$, where Ii is the intensity of the *i*-th measurement, and <*I*> is the mean intensity for that reflection.

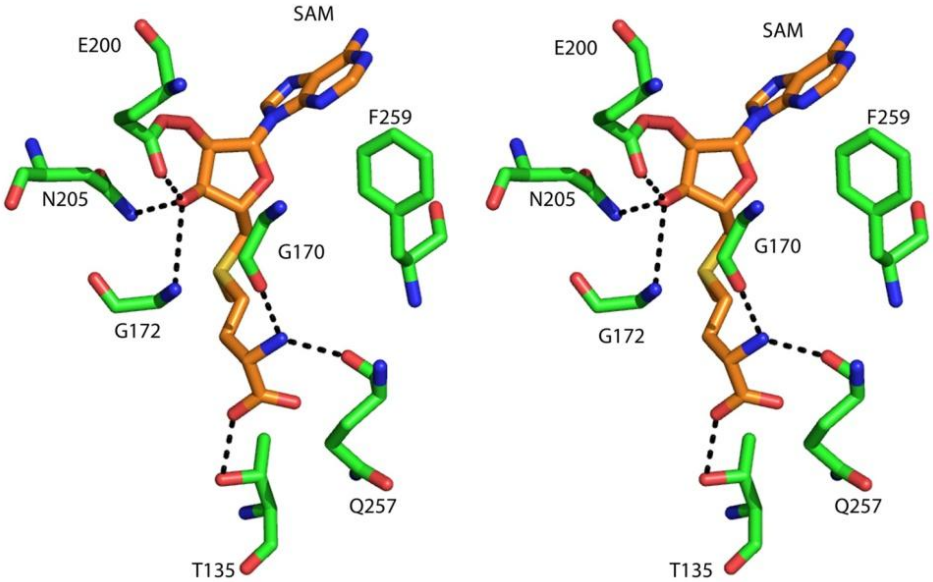
[b]$R_{work}=\sum||Fobs| - |Fcalc||/\sum|Fobs|$, where *F*calc and *F*obs are the calculated and observed structure factor amplitudes, respectively.

[c]R_{free} = as for Rwork, 5% of the total reflections chosen at random and omitted from refinement.

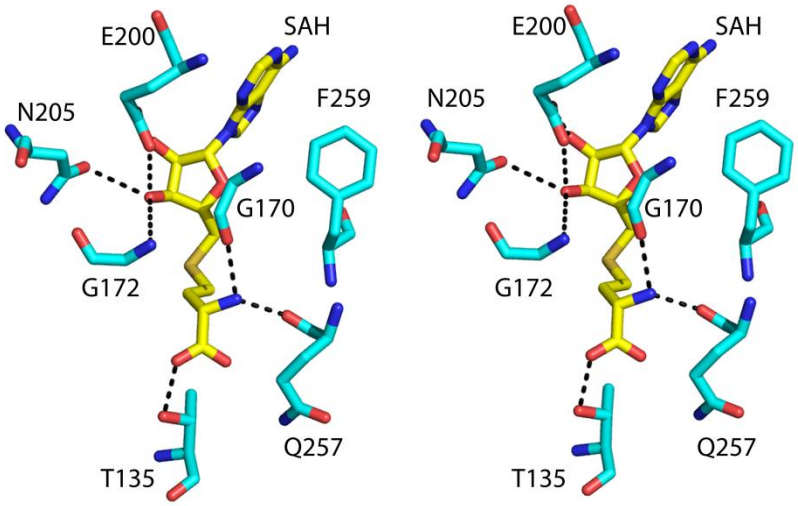
a



b


SAM
E200
F259
N205
G170
G172
Q257
T135
SAM
E200
F259
N205
G170
G172
Q257
T135

c



d

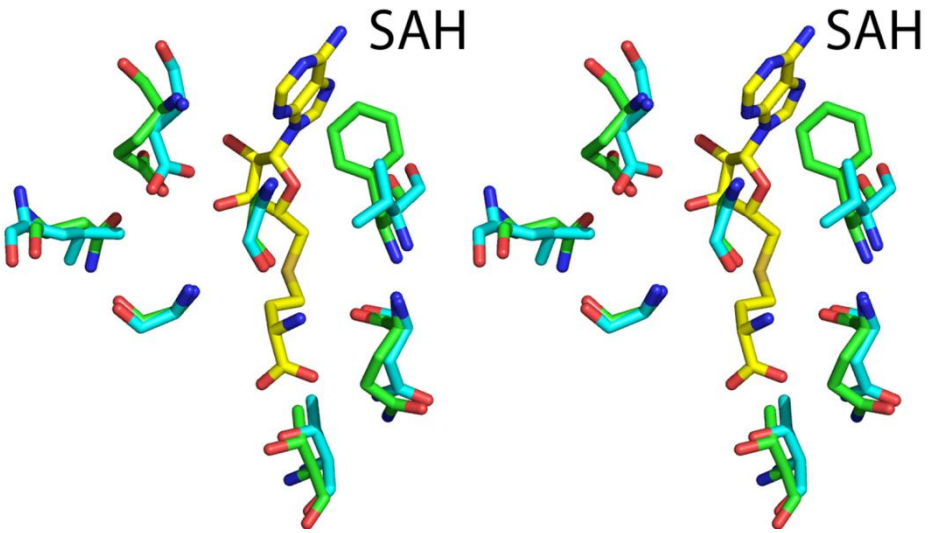


Figure 8.7. MidA Interaction with SAM/SAH. (**a**) Electrostatic surface of MidA, where S-Ado Hcy is bound at the pocket (**b**) MidA residues involved in interaction with SAM (**c**) MidA residues involved in interaction with SAH (**d**) Superimposition of COG1565 super-family member (PDB 4q67, colored in cyan blue) and MidA residues (colored in green) involved in interaction with SAH. The figure was drawn with the Pymol program.

8.3. DOCKING MODEL OF MIDA-NDUFS2 INTERACTION

NDUFS2 is an important subunit of mitochondrial Complex I and is part of the catalytic core (Loeffen, Elpeleg et al. 2001). It interacts with MidA through its N-terminal 40 residues, Val11 to Gly51 (Carilla-Latorre, Gallardo et al. 2010). Earlier reports show that human NDUFS2 is methylated at Arg323 (Loeffen, Elpeleg et al. 2001) and hence NDUFS2 might be a substrate for MidA (Carilla-Latorre, Gallardo et al. 2010). It is important to know the structure of the MidA: NDUFS2 complex which will provide more insight about the mode of interaction and the signal cascade between the two proteins. The knowledge of this interaction will enable us understand about mitochondrial disorders.

a **b**

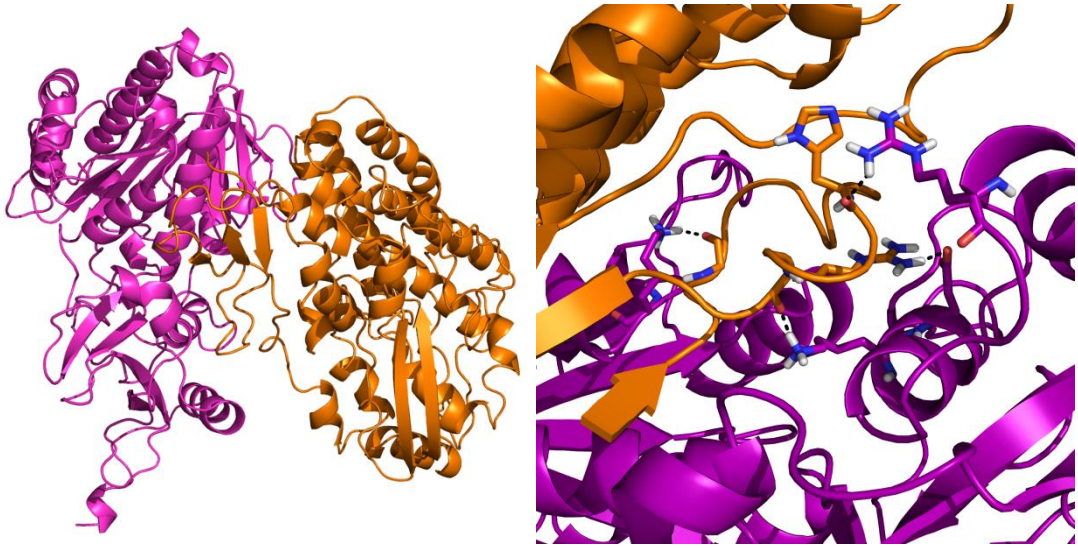


Figure 8.8. MidA-NDUFS2 interaction model. (**a**) A docking model for MidA: NDUFS2 interaction (MidA is shown in purple and NDUFS2 shown in orange color respectively) (**b**) A close-up view showing electrostatic and hydrogen bonding. The figure was drawn with the Pymol program.

The MidA:NDUFS2 complex structure was prepared using the Bioinfromatic server Cluspro 2.1 (Fig. 8.8). NDUFS2 structure was modeled using the 3D CPH model

server (Nielsen, Lundegaard et al. 2010). The model that showed a pose close to the previously predicted orientation was chosen. NDUFS2 interacts with MidA through electrostatic interactions, as verified by the CONTACTS program of CCP4 (Potterton, Briggs et al. 2003; Winn, Ballard et al. 2011). MidA residues Lys120, Glu121, Lys242, Gln334, Asn338, Lys342 and Lys399 form contacts with Asn20, Arg33, Gly22, Lys15, Ser40, Glu41 and Glu9 of NDUFS2 within 3.0 Å. However, an actual experimental structure of the complex is needed to further validate this mode of interaction.

CHAPTER 9. DISCUSSION

MidA is conserved from bacteria to humans and it has a characteristic DUF185 motif of unknown function, as discussed in the results part 8.2.2. MidA from *Dictyostelium* and the one from human share significant sequence identity. The proposed methyltransferase property of this protein is likely to be important for mitochondrial Complex I function. Another putative methyltransferase, C20orf7 from human, has also been proposed to be inevitable for Complex I assembly (Sugiana, Pagliarini et al. 2008; Gerards, Sluiter et al. 2010). The presence of these methyltransferases certainly highlights the importance, but yet poorly defined role, of methylation in Complex I function. Overall 40% of inherited disorders related to the OXPHOS system are due to Complex I deficiencies.

The direct role of MidA in human mitochondrial disorders is still unknown but it could be a strong candidate. The loss of function of MidA in *Dictyostelium* resulted in a complex phenotypic outcome, such as growth and developmental defects (Carilla-Latorre, Gallardo et al. 2010). *Dictyostelium midA*$^{-}$ null cells were found to have decreased activity of Complex I and Blue Native (BN) PAGE studies in human cells ascertain that the level of complex assembly goes down, when MidA level is down regulated, suggesting that MidA could have a role as an assembly factor (Carilla-Latorre, Gallardo et al. 2010). Complex I (CI) is formed by about 45 subunits in humans. Chaperones have an important role in the assembly and stability of such a large multi protein complex and presently about six such assembly factors, including NDUFAF1-NDUFAF4 (Hoefs, Dieteren et al. 2009; Saada, Vogel et al. 2009), C8ORF38 (Pagliarini, Calvo et al. 2008) and C20ORF7 (Gerards, Sluiter et al. 2010), have been identified and studied for their role in human CI deficiency. Despite its

large size, the basic catalytic core of CI is formed by only 14 proteins, which are evolutionarily conserved from prokaryotes to humans. All the 14 proteins, including NDUFS2 (Loeffen, Elpeleg et al. 2001; Ugalde, Janssen et al. 2004), are implicated in human CI disorders. NDUFS2 is an iron-sulfur protein and has been identified to interact with MidA. Even though NDUFS2 is highly homologous in human and *Dictyostelium,* it is encoded by the nuclear genome in humans whereas it is encoded by the mitochondrial genome in *Dictyostelium*, correlating its ancient origin as an endosymbiont protein. Surprisingly, proteobacteria are found to have MidA homologues, giving us a clear message that the MidA protein and the catalytic CI are of ancient origin and conserved from bacteria to humans.

Despite little sequence identity, SAM-dependent methyltransferase enzymes have a conserved structural fold.. Interestingly, the residues that bind the co-factor SAM are poorly conserved (Schubert, Blumenthal et al. 2003; Loenen 2006). Our crystal structure and several structural features support that that MidA could act as a methyltransferase. We suggest that MidA is a methyltransferase involved in SAM binding and could be involved in methylation of its substrate, probably the subunits of mitochondrial Complex I. The proteins that bind to SAM and catalyze a methyl transfer are classified into five groups and our structure fits into class I, the most abundant (Martin and McMillan 2002; Schubert, Blumenthal et al. 2003). Further biochemical characterization identification of targets are needed.

Methyltransferases are involved in the methylation of various substrates, such as DNA, RNA, proteins and small molecules. The atomic targets are carbon, oxygen, nitrogen and sulfur. In most cases, no specific signature in the sequence or structure of methyltransferase proteins can be employed to predict substrates. For example, motif I is found in DNA, RNA, protein and small molecule methyltransferases (Kagan and

Clarke 1994). Methylation has an important role in the mitochondrion, since the targets of specific methylation events are mitochondrial DNA, tRNA and rRNA (Helm, Brulé et al. 1998) and there are specific carriers to transport SAM into the mitochondrion (Palmieri, Arrigoni et al. 2006). Protein methylation is also of importance in cell signaling. Only two methylated subunits have been found in mitochondrial complex I subunits. Bovine NDUFB3 (B12), which is methylated at conserved His residues and human NDUFS2, which is methylated at Arg323 (Carilla-Latorre, Gallardo et al. 2010). Indeed, the possible hypothesis would be that MidA interacts with NDUFS2 and methylates it. However this hypothesis needs adequate experimental validation. The importance of such mitochondrial methylation is not known yet, but such modification of important core member of Complex I might carry a lot of relevance in mitochondrial function.

The role of MidA in Complex I assembly and its interaction with the NDUFS2 subunit of complex I has been reported earlier (Carilla-Latorre, Gallardo et al. 2010), and based on our structure, it could be understood that it binds to SAM with its conserved motif I and might function as a methyltransferase. It is also possible to project that MidA can possibly have dual functions as a chaperone and a methyltransferase. Based on our docking model, it can be speculated that the N-terminus of MidA makes contact with its C-terminus, in addition to binding the co-factor SAM, to interact with NDUFS2. This hypothesis needs to be supported by experimental evidences. It confers the possibility of the substrate binding region existing as a separate domain and the co-factor binding regions existing as a separate unit which would help MidA binding and methylating its substrate it. In case of certain methyltransferases reported earlier, where substrates are known such as Ribosome binding methyltransferase, which tend to have separate catalytic domain

other than the co-factor binding region (Husain, Tkaczuk et al. 2010; Husain, Obranic et al. 2011). In case of MidA, it possibly follows the same mechanism, but the specificity towards substrate will be known only after confirming the substrate and interaction with it. Our model is in accordance with the interaction study results (Carilla-Latorre, Gallardo et al. 2010). Importantly, MidA is not stably bound to CI as shown by BN-PAGE (Carilla-Latorre, Gallardo et al. 2010). Thus it may be possible that MidA, as a protein methyltransferase, interacts with and methylates NDUFS2 during the assembly process of the complex as a transitory protein that is required for correct Complex I formation and then leaves. Crystal structure determination of the MidA:NDUFS2 complex, additional mutational studies of relevant MidA and NDUFS2 residues and identification of specific targets of MidA will give better more details about the mode of interaction of MidA and NDUFS2 and their roles in mitochondrial cell signaling, phenotypic outcome and mitochondrial disorders.

CHAPTER 10. CONCLUSION AND FUTURE PERSPECTIVES

In this thesis, I have reported the details of my studies on two proteins that are involved in gene regulation and cell signaling. Here I am adding the conclusion and future direction, separately for both the proteins.

10.1 NMR STUDY OF TRD

The TRD domain of MBD1 has been known to be highly important for gene repression by recruiting several repressor proteins to a gene promoter. TRD by itself, even in the absence of the other domains of MBD1, is sufficient enough to suppress a gene, as demonstrated when it was fused to the GAL4 DNA binding domain to facilitate localizing to any promoter. Little was known about its structural features and how it recruits multiple proteins with high specificity. The NMR technique was employed to study the structural feature of TRD of MBD1 and it was surprisingly found to be an intrinsically unstructured protein with negligible secondary structure. Interestingly, it was found to be interacting with its binding partners in a specific manner, since NMR titration against the binding partners showed that the mode of binding is specific for particular proteins. GST pull-down, based on the NMR titration, showed that particular residues of TRD favor binding to certain proteins. MCAF1Δ8 is found to be interacting with TRD in a stronger manner than MPG and even a single mutation of TRD could prevent MPG interaction with TRD, whereas MCAF1Δ8 could still interact with TRD. Similarly, the binding affinities, based on ITC, showed that MCAF1Δ8 interacts stronger than MPG. Thus it can be implied that TRD interacts in a specific, but transient, manner with its binding partners. After complex formation and its function are exhibited, TRD might present as an individual protein.

The future direction is to study the importance of the stretch of residues and identify the individual amino acids that are involved in interaction with MCAF1Δ8 and MPG using cell based studies. Site-directed mutations of important residues of TRD of MBD1, based on the current study and check the level of co-localization and suppressor activity, in the presence of MCAF1 and MPG, is also a potential future extension. This study helps us expand the knowledge of the functional importance of intrinsically unstructured proteins in cell signaling.

10.2 CRYSTAL STRUCTURE OF MIDA

MidA was first identified to be important for mitochondrial complex I assembly and interacts with the NDUFS2 subunit of it. It was also proposed to be a methyltransferase, since it has the conserved SAM binding region. Its mutants were unable to bind SAM and *Dictyosteium* MidA$^-$ cells could be rescued with those mutants. However, the above study did not report a structural basis. Apo MidA and in complex with co-factors SAM and SAH were crystallized and the structure was determined using X-ray crystallography. The important residues that are involved in recognizing SAM were identified. While it could be proposed with confidence, based on the structure, that MidA is a methyltransferase, its substrate is still not known. Based on the earlier reports that the N-terminus of NDUFS2 interacts with MidA, we have prepared a docking model of MidA with NDUSF2 and shown that the interaction is mostly through charged residues. Thus the MidA structure justifies it to be a methyltransferase; it recognizes SAM with its conserved SAM binding motif and interacts with NDUFS2. The future direction for this study would be to prove that NDUFS2 is the substrate of MidA and also to determine its other substrates. Once identified, the importance of MidA, as a methyltransferase, can be established by

performing several *in vivo* studies and showing how important is the methylation of such substrates by MidA and what are the subsequent changes endured when the methylation is affected. Also, it is of high interest to know the complex structure of MidA-NDUFS2 by X-ray crystallography which will gives us an idea on how MidA helps in complex I assembly and stability.

REFERENCES

Adams, P. D., P. V. Afonine, et al. (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 213-221.

Barr, H., A. Hermann, et al. (2007). "Mbd2 contributes to DNA methylation-directed repression of the Xist gene." Mol Cell Biol **27**(10): 3750-3757.

Baylin, S. B. (2005). "DNA methylation and gene silencing in cancer." Nat Clin Pract Oncol **2 Suppl 1**: S4-11.

Berger, J., O. Sansom, et al. (2007). "MBD2 is required for correct spatial gene expression in the gut." Mol Cell Biol **27**(11): 4049-4057.

Bird, A. (2002). "DNA methylation patterns and epigenetic memory." Genes Dev **16**(1): 6-21.

Bird, A. P. and A. P. Wolffe (1999). "Methylation-induced repression--belts, braces, and chromatin." Cell **99**(5): 451-454.

Boyer, P. D. (1997). "The ATP synthase--a splendid molecular machine." Annu Rev Biochem **66**: 717-749.

Boyes, J. and A. Bird (1991). "DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein." Cell **64**(6): 1123-1134.

Carilla-Latorre, S., M. E. Gallardo, et al. (2010). "MidA is a putative methyltransferase that is required for mitochondrial complex I function." J Cell Sci **123**(Pt 10): 1674-1683.

Chahrour, M., S. Y. Jung, et al. (2008). "MeCP2, a key contributor to neurological disease, activates and represses transcription." Science **320**(5880): 1224-1229.

Chen, H. and D. C. Chan (2009). "Mitochondrial dynamics--fusion, fission, movement, and mitophagy--in neurodegenerative diseases." Hum Mol Genet **18**(R2): R169-176.

Chen, R. Z., U. Pettersson, et al. (1998). "DNA hypomethylation leads to elevated mutation rates." Nature **395**(6697): 89-93.

Cheng, X., R. E. Collins, et al. (2005). "Structural and sequence motifs of protein (histone) methylation enzymes." Annu Rev Biophys Biomol Struct **34**: 267-294.

Cierpicki, T., L. E. Risner, et al. (2010). "Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia." Nat Struct Mol Biol **17**(1): 62-68.

Clason, T., T. Ruiz, et al. (2010). "The structure of eukaryotic and prokaryotic complex I." J Struct Biol **169**(1): 81-88.

Comeau, S. R., D. Kozakov, et al. (2007). "ClusPro: performance in CAPRI rounds 6-11 and the new server." Proteins **69**(4): 781-785.

Copeland, R. A., M. E. Solomon, et al. (2009). "Protein methyltransferases as a target class for drug discovery." Nat Rev Drug Discov **8**(9): 724-732.

Cornilescu, G., F. Delaglio, et al. (1999). "Protein backbone angle restraints from searching a database for chemical shift and sequence homology." J Biomol NMR **13**(3): 289-302.

Costello, J. F., M. C. Frühwald, et al. (2000). "Aberrant CpG-island methylation has non-random and tumour-type-specific patterns." Nat Genet **24**(2): 132-138.

Courey, A. J., D. A. Holtzman, et al. (1989). "Synergistic activation by the glutamine-rich domains of human transcription factor Sp1." Cell **59**(5): 827-836.

Cross, S. H., R. R. Meehan, et al. (1997). "A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins." Nat Genet **16**(3): 256-259.
Distelmaier, F., W. J. Koopman, et al. (2009). "Mitochondrial complex I deficiency: from organelle dysfunction to clinical disease." Brain **132**(Pt 4): 833-842.
Edwards, C. A. and A. C. Ferguson-Smith (2007). "Mechanisms regulating imprinted genes in clusters." Curr Opin Cell Biol **19**(3): 281-289.
Egner, A., S. Jakobs, et al. (2002). "Fast 100-nm resolution three-dimensional microscope reveals structural plasticity of mitochondria in live yeast." Proc Natl Acad Sci U S A **99**(6): 3370-3375.
Ehrlich, M. and R. Y. Wang (1981). "5-Methylcytosine in eukaryotic DNA." Science **212**(4501): 1350-1357.
Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-2132.
Engelward, B. P., G. Weeda, et al. (1997). "Base excision repair deficient mice lacking the Aag alkyladenine DNA glycosylase." Proc Natl Acad Sci U S A **94**(24): 13087-13092.
Esteller, M., P. G. Corn, et al. (2001). "A gene hypermethylation profile of human cancer." Cancer Res **61**(8): 3225-3229.
Fatemi, M. and P. A. Wade (2006). "MBD family proteins: reading the epigenetic code." J Cell Sci **119**(Pt 15): 3033-3037.
Feng, Q. and Y. Zhang (2001). "The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes." Genes Dev **15**(7): 827-832.
Fischle, W., Y. Wang, et al. (2003). "Binary switches and modification cassettes in histone biology and beyond." Nature **425**(6957): 475-479.
Fournier, A., N. Sasai, et al. (2012). "The role of methyl-binding proteins in chromatin organization and epigenome maintenance." Brief Funct Genomics **11**(3): 251-264.
Frey, T. G. and C. A. Mannella (2000). "The internal structure of mitochondria." Trends Biochem Sci **25**(7): 319-324.
Fujita, N., N. Shimotake, et al. (2000). "Mechanism of transcriptional regulation by methyl-CpG binding protein MBD1." Mol Cell Biol **20**(14): 5107-5118.
Fujita, N., S. Takebayashi, et al. (1999). "Methylation-mediated transcriptional silencing in euchromatin by methyl-CpG binding protein MBD1 isoforms." Mol Cell Biol **19**(9): 6415-6426.
Fujita, N., S. Watanabe, et al. (2003). "MCAF mediates MBD1-dependent transcriptional repression." Mol Cell Biol **23**(8): 2834-2843.
Fuks, F., W. A. Burgers, et al. (2000). "DNA methyltransferase Dnmt1 associates with histone deacetylase activity." Nat Genet **24**(1): 88-91.
Geiman, T. M., U. T. Sankpal, et al. (2004). "DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone methylation system." Biochem Biophys Res Commun **318**(2): 544-555.
Gerards, M., W. Sluiter, et al. (2010). "Defective complex I assembly due to C20orf7 mutations as a new cause of Leigh syndrome." J Med Genet **47**(8): 507-512.
Ghezzi, D., P. Goffrini, et al. (2009). "SDHAF1, encoding a LYR complex-II specific assembly factor, is mutated in SDH-defective infantile leukoencephalopathy." Nat Genet **41**(6): 654-656.

Graff, J. R., J. G. Herman, et al. (1995). "E-cadherin expression is silenced by DNA hypermethylation in human breast and prostate carcinomas." Cancer Res **55**(22): 5195-5199.
Hacker, U., R. Albrecht, et al. (1997). "Fluid-phase uptake by macropinocytosis in Dictyostelium." J Cell Sci **110 (Pt 2)**: 105-112.
Helliwell, J. R. (2005). "Protein crystal perfection and its application." Acta Crystallogr D Biol Crystallogr **61**(Pt 6): 793-798.
Helm, M., H. Brulé, et al. (1998). "The presence of modified nucleotides is required for cloverleaf folding of a human mitochondrial tRNA." Nucleic Acids Res **26**(7): 1636-1643.
Hendrich, B. and A. Bird (1998). "Identification and characterization of a family of mammalian methyl-CpG binding proteins." Mol Cell Biol **18**(11): 6538-6547.
Hensen, E. F. and J. P. Bayley (2011). "Recent advances in the genetics of SDH-related paraganglioma and pheochromocytoma." Fam Cancer **10**(2): 355-363.
Herrmann, T., P. Güntert, et al. (2002). "Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS." J Biomol NMR **24**(3): 171-189.
Ho, K. L., I. W. McNae, et al. (2008). "MeCP2 binding to DNA depends upon hydration at methyl-CpG." Mol Cell **29**(4): 525-531.
Hoefs, S. J., C. E. Dieteren, et al. (2009). "Baculovirus complementation restores a novel NDUFAF2 mutation causing complex I deficiency." Hum Mutat **30**(7): E728-736.
Holliday, R. and J. E. Pugh (1975). "DNA modification mechanisms and gene activity during development." Science **187**(4173): 226-232.
Holm, L., S. Kääriäinen, et al. (2008). "Searching protein structure databases with DaliLite v.3." Bioinformatics **24**(23): 2780-2781.
Husain, N., S. Obranic, et al. (2011). "Structural basis for the methylation of A1408 in 16S rRNA by a panaminoglycoside resistance methyltransferase NpmA from a clinical isolate and analysis of the NpmA interactions with the 30S ribosomal subunit." Nucleic Acids Res **39**(5): 1903-1918.
Husain, N., K. L. Tkaczuk, et al. (2010). "Structural basis for the methylation of G1405 in 16S rRNA by aminoglycoside resistance methyltransferase Sgm from an antibiotic producer: a diversity of active sites in m7G methyltransferases." Nucleic Acids Res **38**(12): 4120-4132.
Ichimura, T., S. Watanabe, et al. (2005). "Transcriptional repression and heterochromatin formation by MBD1 and MCAF/AM family proteins." J Biol Chem **280**(14): 13928-13935.
Johnson, M. A. and B. M. Pinto (2004). "NMR spectroscopic and molecular modeling studies of protein-carbohydrate and protein-peptide interactions." Carbohydr Res **339**(5): 907-928.
Jonckheere, A. I., J. A. Smeitink, et al. (2012). "Mitochondrial ATP synthase: architecture, function and pathology." J Inherit Metab Dis **35**(2): 211-225.
Jørgensen, H. F., I. Ben-Porath, et al. (2004). "Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains." Mol Cell Biol **24**(8): 3387-3395.
Kagan, R. M. and S. Clarke (1994). "Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes." Arch Biochem Biophys **310**(2): 417-427.

Kirby, D. M., M. Crawford, et al. (1999). "Respiratory chain complex I deficiency: an underdiagnosed energy generation disorder." Neurology **52**(6): 1255-1264.
Klose, R. J. and A. P. Bird (2004). "MeCP2 behaves as an elongated monomer that does not stably associate with the Sin3a chromatin remodeling complex." J Biol Chem **279**(45): 46490-46496.
Kozakov, D., D. R. Hall, et al. (2010). "Achieving reliability and high accuracy in automated protein docking: ClusPro, PIPER, SDU, and stability analysis in CAPRI rounds 13-19." Proteins **78**(15): 3124-3130.
Küffner, R., A. Rohr, et al. (1998). "Involvement of two novel chaperones in the assembly of mitochondrial NADH:Ubiquinone oxidoreductase (complex I)." J Mol Biol **283**(2): 409-417.
Kunji, E. R. (2004). "The role and structure of mitochondrial carriers." FEBS Lett **564**(3): 239-244.
Le Guezennec, X., M. Vermeulen, et al. (2006). "MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties." Mol Cell Biol **26**(3): 843-851.
Lemarie, A. and S. Grimm (2011). "Mitochondrial respiratory chain complexes: apoptosis sensors mutated in cancer?" Oncogene **30**(38): 3985-4003.
Lenaz, G. and M. L. Genova (2010). "Structure and organization of mitochondrial respiratory complexes: a new understanding of an old subject." Antioxid Redox Signal **12**(8): 961-1008.
Lewis, J. D., R. R. Meehan, et al. (1992). "Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA." Cell **69**(6): 905-914.
Li, X., B. Z. Barkho, et al. (2008). "Epigenetic regulation of the stem cell mitogen Fgf-2 by Mbd1 in adult neural stem/progenitor cells." J Biol Chem **283**(41): 27644-27652.
Lin, R. J., L. Nagy, et al. (1998). "Role of the histone deacetylase complex in acute promyelocytic leukaemia." Nature **391**(6669): 811-814.
Liu, C., Z. Q. Teng, et al. (2010). "Epigenetic regulation of miR-184 by MBD1 governs neural stem cell proliferation and differentiation." Cell Stem Cell **6**(5): 433-444.
Loeffen, J., O. Elpeleg, et al. (2001). "Mutations in the complex I NDUFS2 gene of patients with cardiomyopathy and encephalomyopathy." Ann Neurol **49**(2): 195-201.
Loenen, W. A. (2006). "S-adenosylmethionine: jack of all trades and master of everything?" Biochem Soc Trans **34**(Pt 2): 330-333.
Maniak, M. (2001). "Fluid-phase uptake and transit in axenic Dictyostelium cells." Biochim Biophys Acta **1525**(3): 197-204.
Mannella, C. A. (2006). "Structure and dynamics of the mitochondrial inner membrane cristae." Biochim Biophys Acta **1763**(5-6): 542-548.
Martin, J. L. and F. M. McMillan (2002). "SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold." Curr Opin Struct Biol **12**(6): 783-793.
Mayer-Jung, C., D. Moras, et al. (1998). "Hydration and recognition of methylated CpG steps in DNA." EMBO J **17**(9): 2709-2718.
McCoy, A. J., R. W. Grosse-Kunstleve, et al. (2007). "Phaser crystallographic software." J Appl Crystallogr **40**(Pt 4): 658-674.
Mckenzie, M. and M. T. Ryan (2010). "Assembly factors of human mitochondrial complex I and their defects in disease." IUBMB Life **62**(7): 497-502.

Meehan, R. R., J. D. Lewis, et al. (1989). "Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs." Cell **58**(3): 499-507.
Millar, C. B., J. Guy, et al. (2002). "Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice." Science **297**(5580): 403-405.
Morris, A. L., M. W. MacArthur, et al. (1992). "Stereochemical quality of protein structure coordinates." Proteins **12**(4): 345-364.
Nakao, M. (2001). "Epigenetics: interaction of DNA methylation and chromatin." Gene **278**(1-2): 25-31.
Nakao, M., S. Matsui, et al. (2001). "Regulation of transcription and chromatin by methyl-CpG binding protein MBD1." Brain Dev **23 Suppl 1**: S174-176.
Nakao, M., T. Minami, et al. (2004). "Epigenetic system: a pathway to malignancies and a therapeutic target." Int J Hematol **80**(2): 103-107.
Nan, X., R. R. Meehan, et al. (1993). "Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2." Nucleic Acids Res **21**(21): 4886-4892.
Nan, X., H. H. Ng, et al. (1998). "Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex." Nature **393**(6683): 386-389.
Ng, H. H., P. Jeppesen, et al. (2000). "Active repression of methylated genes by the chromosomal protein MBD1." Mol Cell Biol **20**(4): 1394-1406.
Ng, H. H., Y. Zhang, et al. (1999). "MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex." Nat Genet **23**(1): 58-61.
Nielsen, M., C. Lundegaard, et al. (2010). "CPHmodels-3.0--remote homology modeling using structure-guided sequence profiles." Nucleic Acids Res **38**(Web Server issue): W576-581.
Niewmierzycka, A. and S. Clarke (1999). "S-Adenosylmethionine-dependent methylation in Saccharomyces cerevisiae. Identification of a novel protein arginine methyltransferase." J Biol Chem **274**(2): 814-824.
Ohki, I., N. Shimotake, et al. (2001). "Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA." Cell **105**(4): 487-497.
Ohki, I., N. Shimotake, et al. (1999). "Solution structure of the methyl-CpG-binding domain of the methylation-dependent transcriptional repressor MBD1." EMBO J **18**(23): 6653-6661.
Otwinowski, Z. and W. Minor (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods in enzymology, Academic press. **276:** 307-326.
Pagliarini, D. J., S. E. Calvo, et al. (2008). "A mitochondrial protein compendium elucidates complex I disease biology." Cell **134**(1): 112-123.
Palmieri, L., R. Arrigoni, et al. (2006). "Molecular identification of an Arabidopsis S-adenosylmethionine transporter. Analysis of organ distribution, bacterial expression, reconstitution into liposomes, and functional characterization." Plant Physiol **142**(3): 855-865.
Pendergrass, W., N. Wolf, et al. (2004). "Efficacy of MitoTracker Green and CMXrosamine to measure changes in mitochondrial membrane potentials in living cells and tissues." Cytometry A **61**(2): 162-169.
Perales-Clemente, E., E. Fernández-Vizarra, et al. (2010). "Five entry points of the mitochondrially encoded subunits in mammalian complex I assembly." Mol Cell Biol **30**(12): 3038-3047.
Polyak, K., Y. Li, et al. (1998). "Somatic mutations of the mitochondrial genome in human colorectal tumours." Nat Genet **20**(3): 291-293.

Potterton, E., P. Briggs, et al. (2003). "A graphical user interface to the CCP4 program suite." Acta Crystallogr D Biol Crystallogr **59**(Pt 7): 1131-1137.
Price, N., F. van der Leij, et al. (2002). "A novel brain-expressed protein related to carnitine palmitoyltransferase I." Genomics **80**(4): 433-442.
Raimundo, N., B. E. Baysal, et al. (2011). "Revisiting the TCA cycle: signaling to tumor formation." Trends Mol Med **17**(11): 641-649.
Riggs, A. D. (1975). "X inactivation, differentiation, and DNA methylation." Cytogenet Cell Genet **14**(1): 9-25.
Rodriguez, D., D. Rojas-Rivera, et al. (2011). "Integrating stress signals at the endoplasmic reticulum: The BCL-2 protein family rheostat." Biochim Biophys Acta **1813**(4): 564-574.
Saada, A., R. O. Vogel, et al. (2009). "Mutations in NDUFAF3 (C3ORF60), encoding an NDUFAF4 (C6ORF66)-interacting complex I assembly protein, cause fatal neonatal mitochondrial disease." Am J Hum Genet **84**(6): 718-727.
Sakamoto, Y., S. Watanabe, et al. (2007). "Overlapping roles of the methylated DNA-binding protein MBD1 and polycomb group proteins in transcriptional repression of HOXA genes and heterochromatin foci formation." J Biol Chem **282**(22): 16391-16400.
Scarano, E. (1971). "The control of gene function in cell differentiation and in embryogenesis." Adv Cytopharmacol **1**: 13-24.
Schubert, H. L., R. M. Blumenthal, et al. (2003). "Many paths to methyltransfer: a chronicle of convergence." Trends Biochem Sci **28**(6): 329-335.
Schwarzinger, S., G. J. Kroon, et al. (2001). "Sequence-dependent correction of random coil NMR chemical shifts." J Am Chem Soc **123**(13): 2970-2978.
SINSHEIMER, R. L. (1955). "The action of pancreatic deoxyribonuclease. II. Isomeric dinucleotides." J Biol Chem **215**(2): 579-583.
Smith, B. C. and J. M. Denu (2009). "Chemical mechanisms of histone lysine and arginine modifications." Biochim Biophys Acta **1789**(1): 45-57.
Stirzaker, C., D. S. Millar, et al. (1997). "Extensive DNA methylation spanning the Rb promoter in retinoblastoma tumors." Cancer Res **57**(11): 2229-2237.
Sugiana, C., D. J. Pagliarini, et al. (2008). "Mutation of C20orf7 disrupts complex I assembly and causes lethal neonatal mitochondrial disease." Am J Hum Genet **83**(4): 468-478.
Taylor, G. (2003). "The phase problem." Acta Crystallogr D Biol Crystallogr **59**(Pt 11): 1881-1890.
Terwilliger, T. (2003). "SOLVE and RESOLVE: automated structure solution and density modification." Methods Enzymol **374**: 22-37.
Terwilliger, T. C., P. D. Adams, et al. (2009). "Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard." Acta Crystallogr D Biol Crystallogr **65**(Pt 6): 582-601.
Terwilliger, T. C. and J. Berendzen (1997). "Bayesian correlated MAD phasing." Acta Crystallogr D Biol Crystallogr **53**(Pt 5): 571-579.
Tocilescu, M. A., U. Fendel, et al. (2007). "Exploring the ubiquinone binding cavity of respiratory complex I." J Biol Chem **282**(40): 29514-29520.
Torija, P., J. J. Vicente, et al. (2006). "Functional genomics in Dictyostelium: MidA, a new conserved protein, is required for mitochondrial function and development." J Cell Sci **119**(Pt 6): 1154-1164.
Uchimura, Y., T. Ichimura, et al. (2006). "Involvement of SUMO modification in MBD1- and MCAF1-mediated heterochromatin formation." J Biol Chem **281**(32): 23180-23190.

Ugalde, C., R. J. Janssen, et al. (2004). "Differences in assembly or stability of complex I and other mitochondrial OXPHOS complexes in inherited complex I deficiency." Hum Mol Genet **13**(6): 659-667.
Villa, R., L. Morey, et al. (2006). "The methyl-CpG binding protein MBD1 is required for PML-RARalpha function." Proc Natl Acad Sci U S A **103**(5): 1400-1405.
Wade, P. A., A. Gegonne, et al. (1999). "Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation." Nat Genet **23**(1): 62-66.
Wakefield, R. I., B. O. Smith, et al. (1999). "The solution structure of the domain from MeCP2 that binds to methylated DNA." J Mol Biol **291**(5): 1055-1065.
Watanabe, S., T. Ichimura, et al. (2003). "Methylated DNA-binding domain 1 and methylpurine-DNA glycosylase link transcriptional repression and DNA repair in chromatin." Proc Natl Acad Sci U S A **100**(22): 12859-12864.
Weixlbaumer, A., S. Petry, et al. (2007). "Crystal structure of the ribosome recycling factor bound to the ribosome." Nat Struct Mol Biol **14**(8): 733-737.
Winn, M. D., C. C. Ballard, et al. (2011). "Overview of the CCP4 suite and current developments." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 235-242.
Wishart, D. S., C. G. Bigam, et al. (1995). "1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects." J Biomol NMR **5**(1): 67-81.
Wolffe, A. P. and M. A. Matzke (1999). "Epigenetics: regulation through repression." Science **286**(5439): 481-486.
Wyatt, M. D., J. M. Allan, et al. (1999). "3-methyladenine DNA glycosylases: structure, function, and biological importance." Bioessays **21**(8): 668-676.
Xu, C., C. Bian, et al. (2011). "The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain." Nat Commun **2**: 227.
Xu, J., C. Liu, et al. (2008). "[Activation of multiple tumor suppressor genes by MBD1 siRNA in pancreatic cancer cell line BxPC-3]." Zhonghua Yi Xue Za Zhi **88**(28): 1948-1951.
Xu, Y., D. Long, et al. (2007). "Rapid data collection for protein structure determination by NMR spectroscopy." J Am Chem Soc **129**(25): 7722-7723.
Xu, Y., Y. Zheng, et al. (2006). "A new strategy for structure determination of large proteins in solution without deuteration." Nat Methods **3**(11): 931-937.
Yang, D., Y. Zheng, et al. (2004). "Sequence-specific assignments of methyl groups in high-molecular weight proteins." J Am Chem Soc **126**(12): 3710-3711.
Yu, F., J. Thiesen, et al. (2000). "Histone deacetylase-independent transcriptional repression by methyl-CpG-binding protein 2." Nucleic Acids Res **28**(10): 2201-2206.
Zhang, X. and T. C. Bruice (2008). "Enzymatic mechanism and product specificity of SET-domain protein lysine methyltransferases." Proc Natl Acad Sci U S A **105**(15): 5728-5732.
Zhao, X., T. Ueba, et al. (2003). "Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function." Proc Natl Acad Sci U S A **100**(11): 6777-6782.
Zhu, J., F. He, et al. (2008). "On the nature of human housekeeping genes." Trends Genet **24**(10): 481-484.
Ziemer, M. A., A. Mason, et al. (1982). "Cell-free translations of proline-rich protein mRNAs." J Biol Chem **257**(18): 11176-11180.