

NMR Studies of the Methyl-CpG Binding Domain of Human Methylation-Dependent Repressor MBD1 and Its Interaction with Methylated DNA

(ヒトメチル化特異的リプレッサー-MBD1のメチル化CpG結合ドメインの
NMRによる研究)

The Doctoral Thesis
Submitted to the Graduate School of Biological Science of
Nara Institute of Science and Technology
February 2001

Izuru Ohki
(Dissertation Director: Professor Toshio Hakoshima)

大木 出
奈良先端科学技術大学院大学
バイオサイエンス研究科 生体高分子構造学講座
(箱嶋 敏雄 教授)

所属 (主指導教官)	バイオサイエンス研究科 生体高分子構造学講座 (箱嶋 敏雄 教授)		
氏名	大木 出	提出	平成13年 2月 2日
題目	NMR studies of the methyl-CpG binding domain of human methylation-dependent repressor MBD1 and its interaction with methylated DNA. (ヒトメチル化特異的リプレッサー-MBD1のメチル化CpG結合ドメインのNMRによる研究)		

要旨

DNA methylation in mammalian cells occurs at the 5-position of cytosine within the CpG sequence, and causes genome-wide phenomena such as gene silencing, gene imprinting, and X chromosome inactivation. A protein family that contains conserved methyl-CpG-binding domains (MBDs) recognizes this methylation signal and recruits another protein factors, such as corepressor complex including histone deacetylase and chromatin remodeling ATPase, at the methylation site.

One of such proteins, MBD1 is a mammalian methylation-dependent transcriptional repressor and binds specifically to a methylated CpG sequence via its MBD. This study reports the solution structure of MBD (residues 1-75) of human MBD1 protein in complexed with and in absence of a 12-mer methylated DNA by means of multi-dimensional heteronuclear NMR spectroscopy.

The three-dimensional structures of MBD1 MBD both in complexed with and absence of DNA have a compact α/β fold and consist of a four-stranded β -sheet and an α -helix packed approximately parallel to the β -sheet [1,2]. The residues formed hydrophobic core in MBD1 are highly conserved in the other MBDs, suggesting that the folding is essentially identical throughout the MBD family. Contacts to the DNA are made at the face of the β -sheet, which is positioned within the major groove. The methyl groups at the methylation site are recognized through hydrophobic contacts with aliphatic and aromatic portions of arginine, tyrosine, valine, and serine residues that are highly conserved among the MBD family. All interactions with DNA are observed in the major groove side of the six base pairs width and the specific base contacts are made only in methyl-CpG sequence.

The complex structure also provides insight into the mechanism of a disease. One of the MBD proteins, MeCP2 is associated with Rett syndrome, a childhood neurodevelopmental disorder; many mutations are found in its MBD sequence. The structures suggest that some residues of MeCP2 MBD that are mutated in Rett syndrome are located at the protein hydrophobic core or the protein-DNA interface, providing a structural basis to understand the consequence of these mutations.

[1] Ohki I., Shimotake N., Fujita N., Nakao M. and Shirakawa M. (1999). Solution structure of the methyl-CpG-binding domain of the methylation-dependent transcriptional repressor MBD. *EMBO J.* 18, 6653-61.

(This paper reports the structure of MBD1 in absence of DNA)

[2] Ohki I., Shimotake N., Fujita N., Jee J. -G., Ikegami T., Nakao M. and Shirakawa M. Structure of the methyl-CpG-binding domain of human MBD1 in a complex with methylated DNA. being submitted for publication.

CONTENTS

1. SUMMARY	4
2. INTRODUCTION	5
3. MATERIALS AND METHODS	
3.1 Structural analysis of DNA-free MBD1 MBD	12
3.1.1 Sample preparation of DNA-free MBD	12
3.1.2 NMR measurements of DNA-free MBD	14
Resonance assignments of MBD	15
Structure restraints of MBD	15
$\{^1\text{H}\}$ - ^{15}N steady-state heteronuclear NOE experiments	16
3.1.3 Structure calculation of MBD	16
3.2 Structural analysis of MBD-DNA complex	19
3.2.1 Sample preparation of MBD-DNA complex	19
Preparation of MBD	19
Preparation of 12-mer methylated DNA	21
Preparation of MBD- DNA complex	21
3.2.2 NMR measurements of MBD-DNA complex	22
Resonance assignments of MBD	22
Structure restraints of MBD	22
Resonance assignments of free-DNA	23
Resonance assignments and structure restraints of DNA in complex	23
Intermolecular restraints	24
$\{^1\text{H}\}$ - ^{15}N steady-state heteronuclear NOE experiments	25
3.2.3 Structure calculation of MBD-DNA complex	26
3.3 Mutational analyses	28
4. RESULTS	
4.1 Sample preparation of free MBD and MBD-DNA complex	30
4.2 Resonance assignments	31
DNA-free MBD	31
MBD in complex with DNA	31
DNA in complex with MBD, and free DNA	32

4.3 Structure determination and structural statistics	35
DNA-free MBD	35
MBD-DNA complex	35
4.4 Solution structure of DNA-MBD Complex	39
Structure of MBD	39
Structure of MBD-DNA complex	41
4.5 Molecular basis of methylated DNA recognition	45
Overview of recognition	45
Base recognition at the methyl-CpG site	45
DNA backbone contacts	47
5. DISCUSSION	
5.1 Comparison of the DNA-bound and DNA-free MBD	52
5.2 Comparison of MBD1 MBD and MeCP2 MBD	54
5.3 Rett mutations found in MeCP2	56
6. CONCLUSION	60
7. ACKNOWLEDGEMENTS	61
8. REFERENCES	62
9. LIST OF PUBLICATIONS	68

1 SUMMARY

DNA methylation in mammalian cells occurs at the 5-position of cytosine within the CpG sequence, and causes genome-wide phenomena such as gene silencing, gene imprinting, and X chromosome inactivation carcinogenesis. A protein family that contains conserved methyl-CpG-binding domains (MBDs) recognizes this methylation signal and recruits another protein factors, such as corepressor complex including histone deacetylase and chromatin remodeling factors, at the methylation site. One of such proteins, MBD1 is a mammalian methylation-dependent transcriptional repressor and binds specifically to a methylated CpG sequence via its MBD. This study reports the solution structure of MBD of human MBD1 protein both in complex with and absence of methylated DNA by means of multi-dimensional heteronuclear NMR spectroscopy. The structures of the MBD1 MBD consist of a four-stranded β -sheet and an α -helix packed approximately parallel to the β -sheet. The residues formed hydrophobic core in MBD1 are highly conserved in the other MBDs, which suggests that the folding is essentially identical throughout the MBD family. Contacts to the methylated DNA are made at the face of the β -sheet, which is positioned within the major groove. DNA binding induces the folding of a long loop that contributes a major DNA interface, establishing a novel mode of DNA binding. The methyl groups at the methylation site are recognized through extensive hydrophobic contacts with aliphatic and aromatic portions of arginine, tyrosine, valine, and serine residues that are highly conserved among the MBD family. Discrimination of the CG sequence is due to the same arginine and tyrosine residues. The structure also suggests that some residues of MeCP2 that are mutated in Rett syndrome, a childhood neurodevelopmental disorder, are located at the protein hydrophobic core or the protein-DNA interface, providing a structural basis to understand the consequence of these mutations.

2 INTRODUCTION

CpG methylation in vertebrates is important for control of gene activity either through effects on a single promoter region, or through global mechanisms that affect many genes (Fig. 2.1) (Razin, 1998; Tajima and Suetake, 1998; Siegfried and Cedar, 1997; Kass et al., 1997; Bird, 1992; Eden et al., 1998). In vertebrates, 60-90 % of all CpGs are methylated and differences in the CpG-methylation status correlate with a wide range of biological phenomena, including genomic imprinting, carcinogenesis, X inactivation and embryonic development (Surani, 1998; Chen et al., 1998; Tate et al., 1996; Li et al., 1993; Ballabio and Willard, 1992; Ng et al., 1999). The methylation of promoter-associated CpG islands is involved in the transcriptional repression of tumor suppressor genes in tumor cells and imprinted genes (Baylin et al., 1998; Feinberg et al., 1995; Sutcliffe et al., 1994; Lalande, 1996). On the other hand, genome-wide demethylation has been suggested to be an early step in carcinogenesis. DNA hypomethylation was shown to lead to elevated mutation rates, and is associated with the occurrence of chromosomal abnormalities (Fearon and Vogelstein, 1990; Chen et al., 1998).

The signals that DNA methylation represents are interpreted by protein factors that contain shared methyl-CpG-binding domains (MBDs) (Fig. 2.2A): there are at present five known family members, MeCP2, MBD1, MBD2, MBD3, and MBD4 (Fig. 2.2B) (Hendrich and Bird, 1998; Bhattacharya et al., 1999). Mammalian cells have the potential to encode both long and short forms of MBD2, and these have been termed MBD2a and 2b, respectively (Hendrich and Bird, 1998). One of them, MBD2b, was proposed to have methyl-CpG specific demethylase activity (Bhattacharya et al., 1999), but recent publications reported that MBD2b and its related proteins had failed to demethylate DNA, and instead implicate MBD2 in transcriptional repression (Wade, et al., 1999; Ng, et al., 1999). MBD4, another member of the MBD family, was shown to be an endonuclease that forms a complex with the DNA mismatch repair protein, MLH1, which

suggests that MBD4 may be involved in mismatch repair and be responsible for maintenance of genome stability (Bellacosa et al., 1999). Recently Hendrich and coworkers showed that MBD4 has DNA glycosylase activity and can remove thymine from a mismatch CpG site in vitro, and that its MBD binds preferentially to methyl-CpG·TpG mismatches (Hendrich et al., 1999). MBDs have been also found in *Xenopus laevis*. Wade et al. showed that *Xenopus laevis* Mi-2 complex contains MBD3 and MBD3 LF, whose sequences are substantially similar to mammalian MBD3 (Wade et al., 1999). MBD3 LF is a product of an alternative-splicing variant, and contains an insertion of 20 amino acid residues in MBD. Sequences of all the MBD family are highly conserved. All but mammalian MBD3 and *X. laevis* MBD3 LF have been shown to bind specifically to symmetric methyl-CpG, and the binding appears to be as a monomer and be independent of the sequence context outside of the CG sequence (Chandler et al., 1999; Hendrich and Bird, 1998; Nan et al., 1993).

Methyl-CpG binding proteins (MeCPs) 1 and 2 contribute to methylation-dependent gene silencing (Bergman and Mostoslavsky, 1998; Kass et al., 1997; Meehan et al., 1989; Boyes and Bird, 1991; Lewis et al., 1992; Nan et al., 1997). MeCP2 was shown to bind to a methylated DNA and to recruit the Sin3-histone deacetylase complex to promoters, resulting in deacetylation of core histones and subsequent transcriptional silencing (Fig. 2.3) (Nan et al., 1998; Jones et al., 1998). In like manner, MeCP1 is a histone deacetylase complex that includes MBD2, RbA p48/p46, histone deacetylases (HDACs) 1 and 2, Sin3A and its related proteins (Wade et al., 1999; Ng et al., 1999). This complex is capable of repressing transcription from closely and sparsely methylated promoters, and its activity is ubiquitous in somatic cells and tissues but is notably absent in ES cells (Boyes and Bird, 1991, 1992; Meehan et al., 1989). Besides these two complexes of MeCPs, other histone deacetylase core complexes were found to contain members of MBD family. NuRD, a multi-subunit complex having nucleosome remodeling and histone deacetylase activities, contains MBD3 and Mi-2 (Zhang et al., 1999; Wade et al., 1999). Mi-2 is known to be a member of the Swi2 / Snf2 helicase / ATPase family (Wade et al., 1999; Wade et

al., 1998). These observations illustrate a close link between CpG-methylation and histone deacetylation.

MBD1 was proposed to be a component of MeCP1, but this conclusion has been questioned (Cross et al., 1997; Ng et al., 1999). Fujita et al. showed that MBD1 can bind to methylated CpG islands from the tumor suppressors, p16, VHL and E-cadherin, and imprinted SNRPN genes, and inhibit promoter activities of these genes in a methylation dependent manner in vitro and in the cells (Fujita et al., 1999). Recently, MBD1 have effective transcriptional repression domains (TRD) in addition to the MBDs (Fig. 2.2B) (Nan et al., 1998; Ng et al., 2000; Fujita et al., 1999, 2000). The repression mechanism of MBD1 is also likely to involve deacetylation (Ng et al., 2000).

Relation to the disease has been found in these MBD proteins. Mutations in one such protein, MeCP2 can cause Rett syndrome, X-linked dominant disorder and is one of the most common causes of sporadic mental retardation in females (incidence of 1 in 10,000-15,000) (Amir et al., 1999). Most mutations are missense or truncating mutations that affect the integrity of the two main domains of MeCP2, the MBD and the TRD (Amir et al., 1999; Veyver and Zoghbi, 2000).

In this thesis, we present NMR studies of the solution structure of human methylation-dependent repressor MBD1 MBD (residues 1-75) complexed with a 12-mer methylated DNA. The purpose of this study is to understand the mechanisms of the methylated DNA recognition and the relation to the disease in terms of the three-dimensional structures and site-directed mutagenesis.

The determined structure of MBD1 MBD folds into a novel α/β -sheet sandwich structure with a long characteristic loop. Interaction between the MBD and DNA is made at the face of the β -sheet and the major groove. The structures of the MBD-DNA complex suggests that some residues of Rett mutations in MeCP2 MBD are located at the position of the protein hydrophobic core or the protein-DNA interface, and cause the loss of the function of MeCP2 protein. In this

study, the structure of MBD in absence of DNA also reported to enable better understanding of the structural requirements imposed on the protein structure for DNA recognition.

This study is roughly divided into two parts; study of MBD in absence of DNA (1), and study of MBD in complex with DNA (2). The paper published in 1999 (Ohki et al., 1999) describes a part 1, and is equivalent to Chapter 3.1 and a part of Chapter 4 in this thesis. Chapter 3.2 and the most part of Chapter 4 and 5 are the description about a part 2, and its data are submitted for publication now.

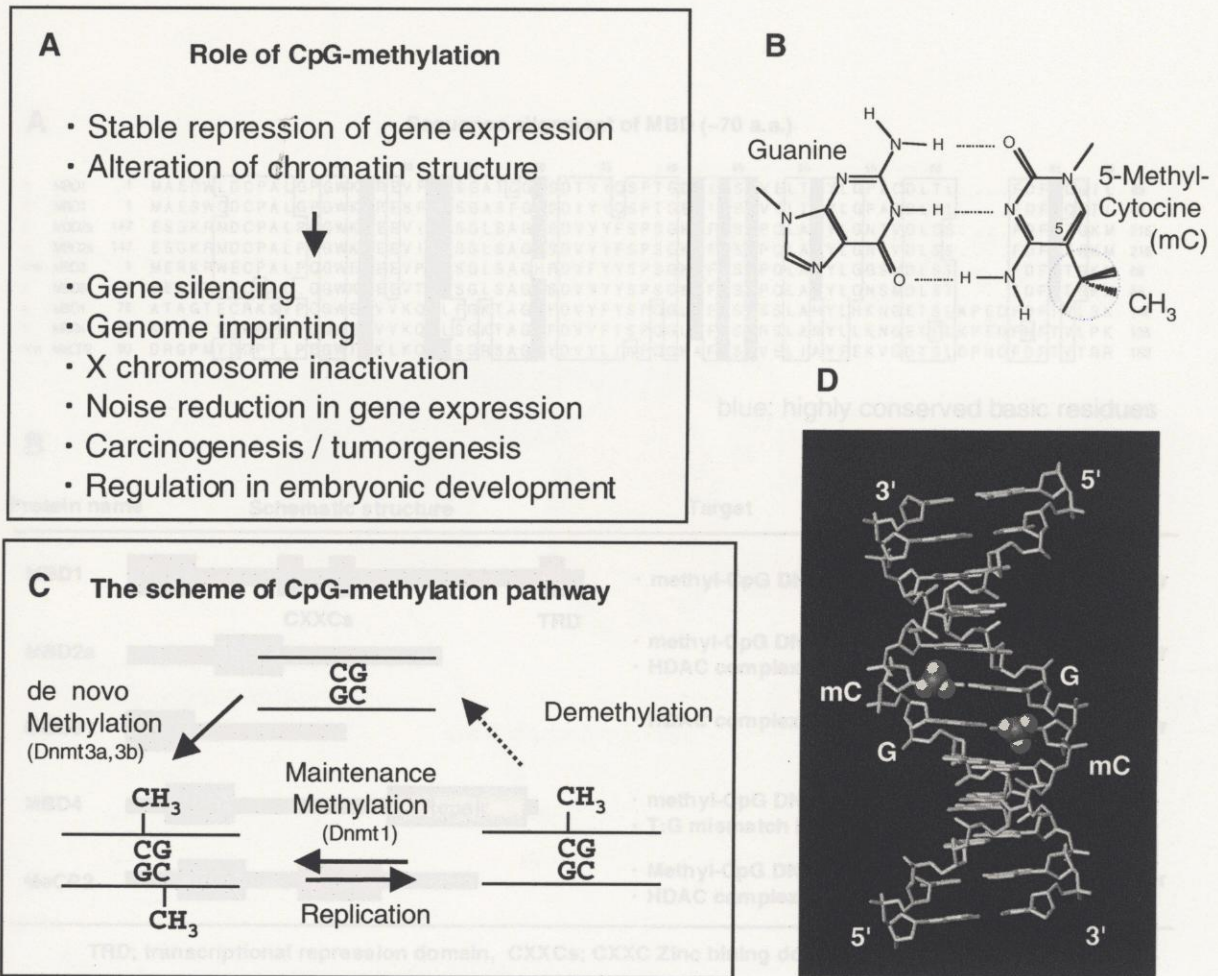
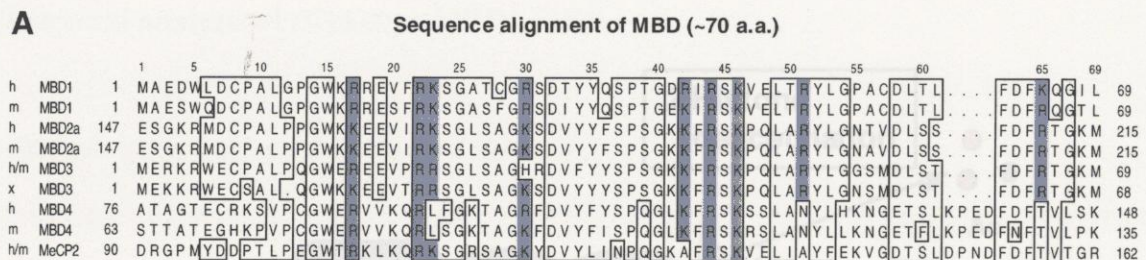


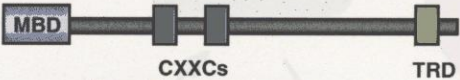




Figure 2.1 CpG methylation in vertebrates.

In vertebrates, cytosine methylation at the CpG site is important for the regulation of gene activity, genomic stability and chromatin structure; differences in the DNA-methylation status are associated with imprinting phenomena, development, and carcinogenesis (A) (Surani, 1998; Chen et al., 1998; Tate et al., 1996; Li et al., 1993; Ballabio and Willard, 1992; Ng et al., 1999). (B) The chemical structure of a guanine:5-methyl cytosine base pair. Hydrogen bonding arrangements are the same as guanine:cytosine base pair. (C) The scheme of CpG methylation pathway. In vertebrate, CpG methylation is carried out by methyltransferase 1 and 3 (Dnmt1, Dnmt3a and Dnmt3b). Dnmt1 is targeted to replication forks and introduces methyl groups on the newly synthesized DNA strand, thereby the methylation pattern is conserved throughout the DNA replications. (D) The methyl-CpG sequence in B-form DNA. The two methyl groups of methyl cytosine (mC) at the CpG site are exposed to the solvent from the side of DNA major groove.



blue; highly conserved basic residues

B

Protein name	Schematic structure	Target	Function
MBD1		<ul style="list-style-type: none"> • methyl-CpG DNA 	<ul style="list-style-type: none"> • Transcriptional repressor
MBD2a		<ul style="list-style-type: none"> • methyl-CpG DNA • HDAC complex (Mi-2) 	<ul style="list-style-type: none"> • Transcriptional repressor
MBD3		<ul style="list-style-type: none"> • HDAC complex (Mi-2) 	<ul style="list-style-type: none"> • Transcriptional repressor
MBD4		<ul style="list-style-type: none"> • methyl-CpG DNA • T:G mismatch DNA 	<ul style="list-style-type: none"> • Mismatch repair
MeCP2		<ul style="list-style-type: none"> • Methyl-CpG DNA • HDAC complex (Sin3) 	<ul style="list-style-type: none"> • Transcriptional repressor

TRD; transcriptional repression domain, CXXCs; CXXC Zinc binding domain

Figure 2.2 Methyl-CpG binding domain (MBD) and MBD-containing proteins.

(A) Sequence alignment of the MBD family members of human (h), mouse (m) and *X. laevis* (x).. The numbering is shown for human MBD1 (above). Conserved residues are boxed and conserved basic residues are colored in blue. MBD2 of human and mouse has a potential to encode both long (MBD2a) and short (MBD2b) forms. MBD2b starts at methionine 152 of MBD2a.

(B) MBD-containing proteins. MBD proteins share a well-conserved MBD. CXXC domains occur in DNA methyltransferase 1 (DNMT1), HRX protein and MBD1. TRD in both MBD1 and MeCP2 is referred to the transcriptional repression domain. The repair domain of MBD4 is a TG mismatch glycosylase.

3 MATERIALS AND METHODS

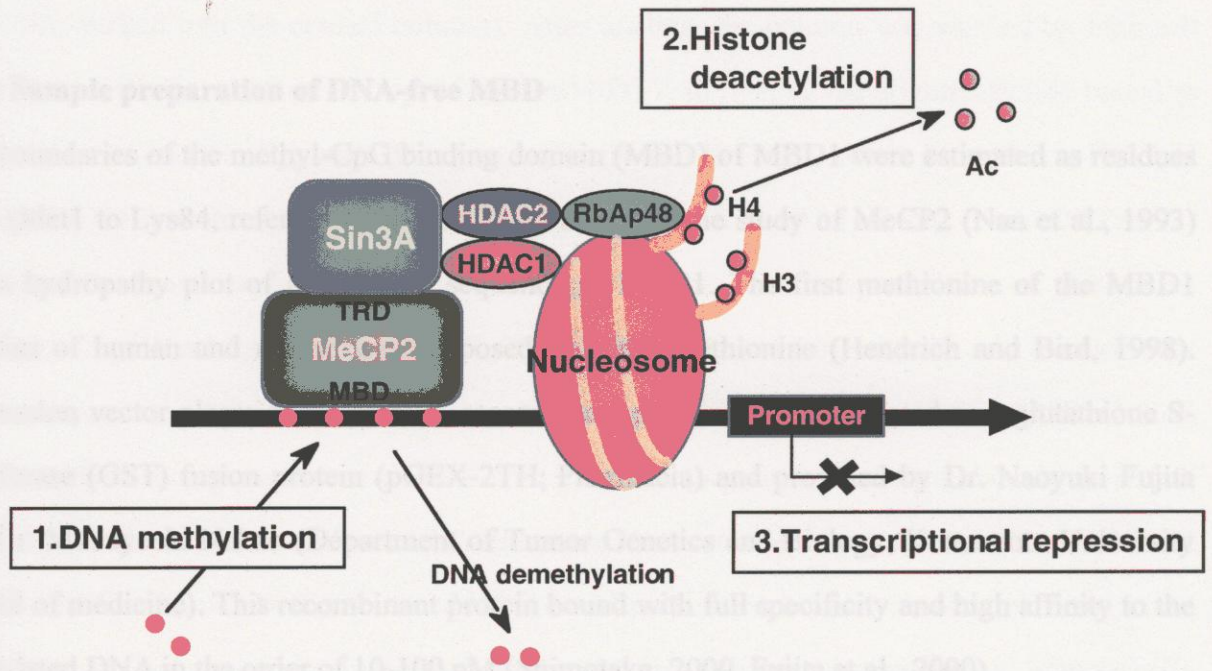


Figure 2.3 MeCP2-induced transcriptional repression by CpG methylation.

Transcriptional repression complexes. The repression complex is formed initially by the interaction of MeCP2 with the DNA at methyl-CpG site, which then recruits and interacts with Sin3A to which the two deacetylases HDAC1 and HDAC2 are associated. The deacetylases remove acetyl moieties from the ϵ -amino acetyl groups of specific lysine residues on histone H3 and H4. Finally, the interactions between the positively charged lysines with the DNA presumably restricts nucleosome mobility on the DNA, rendering the promoter inaccessible to the transcriptional machinery (Bazin, 1998).

3 MATERIALS AND METHODS

3.1 Structural analysis of DNA-free MBD1 MBD

3.1.1 Sample preparation of DNA-free MBD

The boundaries of the methyl-CpG binding domain (MBD) of MBD1 were estimated as residues 1-84 (Met1 to Lys84, referred to as MBD1₁₋₈₄) based on the study of MeCP2 (Nan et al., 1993) and a hydropathy plot of amino acid sequence of MBD1. The first methionine of the MBD1 proteins of human and mouse was proposed in-frame methionine (Hendrich and Bird, 1998). Expression vector plasmid of MBD1₁₋₈₄ recombinant protein was constructed as a glutathione S-transferase (GST) fusion protein (pGEX-2TH; Pharmacia) and provided by Dr. Naoyuki Fujita and Dr. Mitsuyoshi Nakao (Department of Tumor Genetics and Biology, Kumamoto University School of medicine). This recombinant protein bound with full specificity and high affinity to the methylated DNA in the order of 10-100 nM (Shimotake, 2000, Fujita et al., 2000).

Isotopically ¹⁵N-labelled and ¹⁵N/¹³C-labelled MBD1₁₋₈₄ proteins were expressed in *E. Coli* (DH5 α) as a GST fusion protein in M9 synthetic media (7 g/L Na₂HPO₄, 3 g/L KH₂PO₄, 0.5 g/L NaCl, 20 mg/L thymidine, 20 mg/L adenosine, 20 mg/L guanosine, 20 mg/L cytidine, 20 mg/L biotin, 20 mg/L thiamine, 1 mM MgSO₄, 3.3 μ M FeCl₃, 50 μ M MnCl₂, and 100 μ M CaCl₂) containing ¹⁵N-ammonium chloride (¹⁵NH₄Cl, 0.5 g/L) and ¹⁵NH₄Cl (0.5 g/L) / ¹³C-glucose (1 g/L), respectively. When ¹⁵N-labelled protein was expressed, 0.1 % glycerol and 4 g/L glucose were added in M9 medium. All mediums used for expression of MBD1₁₋₈₄ contained 50 mg/ml ampicillin. The culture was started at 37 °C. When the OD₆₆₀ reached to 0.5, the culture was cooled down from 37 °C to 25 °C, and then 1- β -D thiogalactopyranoside (IPTG) was added to a final concentration of 0.1mM. After 6 hr induction at 25 °C, bacteria were harvested by centrifugation (6,000 rpm, 30 min, 4 °C, in a Beckman JA-10 rotor) and stored at -80 °C.

The stored cells at -80 °C were resuspended in a solution containing 400 mM KCl, 20 mM

Tris-HCl (pH 8.0), 1 mM EDTA, and 5 mM dithiothreitol (DTT) (10ml per 1g wet cells), and sonicated immediately on ice after addition of 0.1mM phenylmethyl sulphonyl fluoride (PMSF). The lysate was centrifuged (19,000rpm, 30min, 4 ° C, in a Beckman JA-20 rotor), and the supernatant was loaded onto a glutathione column (3ml glutathione-Sepharose 4B resin per 1g wet cells, packed into the opened column). After loading, the column was washed by high salt buffer (1M KCl, 20 mM Tris-HCl (pH 8.0), 5 mM DTT, to remove the oligonucleotide bound to the protein non-specifically), mild detergent buffer (Triton X-100 0.1% v/v, 20 mM Tris-HCl (pH 8.0), 5 mM DTT, without salt), and salt-free buffer (20 mM Tris-HCl (pH 8.0), 5 mM DTT) at 4 ° C. Then, the fusion protein was eluted from glutathione column by addition of glutathione elution buffer (10 mM glutathione (GSH), 20 mM Tris-HCl (pH 8.0)) after 30 min incubation at room temperature. After concentration up to 1 mg/ml (Centriplus 10, 10,000 M.W. cutoff, Amicon), the fusion protein was cleaved with thrombin (Sigma, human thrombin; 20 NIH unit per 1 mg fusion protein) for 6 hr at 4 ° C. Thrombin cleavage was stopped by addition of 0.1mM PMSF, and then MonoS buffer A was added (20 mM potassium phosphate buffer (pH 6.5), 5 mM DTT) until pH value was decreased to 7.0. The cleavage mixture was purified to homogeneity using a Mono S FPLC column (Pharmacia, 1 ml column) with a linear salt gradient from buffer A (20 mM potassium phosphate buffer (pH 6.5), 5 mM DTT) to buffer B (20 mM potassium phosphate buffer (pH 6.5), 1 M KCl, 5 mM DTT). The C-terminal 9 residues (AHPVAVASK) were degraded in the bacterial cells and during steps of purification. The MonoS fractions contained MBD1 which consisted of the residues 1-75 were collected and diluted by NMR measurement buffer (20 mM potassium phosphate buffer (pH 6.5), 50 mM KCl, 5mM DTT in 90% H₂O/10% D₂O) and concentrated to 0.7-1.5 mM using a Centricon-3 concentrator (Amicon, 3,000 M.W. cutoff). The purified protein was checked by N-terminal sequencing and TOF-mass spectroscopy. Its final purity was checked by 17.5% SDS- polyacrylamide gel electrophoresis (SDS-PAGE).

The molecular mass of recombinant MBD1 (residues 1-75) is 8,639 Dalton. The calculated isoelectric point is 8.69. The molar absorption coefficient at 280 nm is calculated to be 16,460

according to the equations $(5,550 \times 2(\text{Trp}) + 1,340 \times 4(\text{Tyr}))$, where $N(X)$ represents the number of amino acid X in the protein.

3.1.2 NMR measurements of DNA-free MBD

The protein used for NMR measurement consisted of residues 1-75 of MBD1 with two additional residues derived from the vector on N-terminus (GS), referred to as MBD. NMR spectra were acquired at 303 K with Bruker DMX500, DRX500 or DRX800 NMR spectrometer equipped with a triple resonance probe ($^1\text{H}/^{15}\text{N}/^{13}\text{C}$) and triple axis gradient (xyz axis). Samples for NMR measurements typically comprised 1.3 mM MBD in 20 mM potassium phosphate buffer (pH 6.5), 50 mM KCl, 5 mM DTT. The sample tubes for NMR measurements were a 5 mm silicone coated micro glass tubes (Shigemi) with a flat bottom and a cylindrical adapter inserted from upward. The proton base frequency (0 ppm) was set to the absolute resonance frequency of DSS dissolved in 90% $\text{H}_2\text{O}/10\%$ D_2O at 30 °C. The base frequencies of ^{15}N and ^{13}C nuclei were set to the values of 0.10132905 ($^{15}\text{N}/^1\text{H}$ frequency ratio) and 0.25144952 ($^{13}\text{C}/^1\text{H}$ frequency ratio), respectively. Of all NMR experiments, quadrature detection in the indirect dimensions was achieved by States-TPPI methods. In experiments with gradient pulses applied for coherence selection (gradient echo), quadrature detection was achieved by linear combinations of P- and N-type data. All NMR data were processed using program NMRPipe (Delaglio et al., 1995). In general, time-domain data were apodized with sine-bell or Lorentzian-to-Gaussian window functions and zero-filled once, and then transformed to frequency-domain data by Fourier transformation. Linear prediction (LP) was applied in only one dimension where appropriate. Mirror-image LP for indirect dimension acquired with constant-time evolution. Processed spectra were analyzed using PIPP (Garrett et al., 1991) and NMRVIEW (Johnson et al., 1994) softwares.

Resonance assignments of MBD

For assignments of the ^1H , ^{15}N and ^{13}C resonances, a series of two- and three-dimensional experiments (2D ^{15}N - ^1H HSQC, ^{13}C - ^1H HSQC, and 3D CBCA(CO)NH, CBCANH, HNCO, HN(CA)CO, C(CO)NH, H(CCO)NH, HCCH-TOCSY, ^{15}N -TOCSY-HSQC) were performed with the ^{15}N - or ^{15}N , ^{13}C -labelled protein dissolved in 90% H_2O /10% D_2O or 100% D_2O solution. (Cavanagh et al., 1996). The Table 3.1 summarized the information of these experimental parameters and combinations of the experiment and the sample.

Of these experiments for assignments, the 2D ^{15}N - ^1H -HSQC experiment included WATER-GATE technique with 3-9-19 composite pulse and WATER-FLIP-BACK technique to suppress the water signal. Other 3D NMR experiments in which amide protons were directly detected included the sensitivity-enhancement and gradient-echo methods on the indirect ^{15}N dimensions. The 2D- ^{13}C - ^1H HSQC experiments also included the sensitivity-enhancement method in the indirect ^{13}C dimension. The 3D H(CCO)NH, C(CO)NH, and HCCH-TOCSY experiments were acquired with 3 cycles of DIPSI-3 ^{13}C -homonuclear mixing (20.6 ms of total mixing time) at the rf field strength of 8.1 kHz. The 3D ^{15}N -TOCSY-HSQC experiment was acquired with 21 cycles of DIPSI-2 ^1H -homonuclear mixing (70 ms of mixing time) at the rf field strength of 8.6 kHz.

The stereospecific assignment of the methyl groups of Leu and Val residues was achieved using a 15 % fractionally ^{13}C -labelled protein (Hu and Zuiderweg, 1996).

Chemical shifts of the amino groups of Arg, Gln side-chains were assigned with 2D ^{15}N - ^1H -HSQC and 3D ^{15}N -NOESY-HSQC spectra. The assignment of the aromatic proton of Tyr, Phe, Trp residues was obtained by ^1H - ^1H 2D NOESY and TOCSY spectra recorded in D_2O .

Structure restraints of MBD

Distance information (Nuclear Overhauser Effects; NOEs) was obtained by means of 3D ^{15}N - or ^{13}C - , or 4D ^{15}N , ^{13}C - or ^{13}C , ^{13}C - NOESY experiments with a mixing time of 100 ms or 150ms (Cavanagh et al., 1996). In the 4D NOESY experiments, WATER-GATE and WATER-FLIP-BACK techniques were used.

For torsion angle restraints (ϕ), the backbone vicinal coupling constants (${}^3J_{\text{HNH}\alpha}$) were determined by means of HMQC-J (Cavanagh et al., 1996). The torsion angles, χ^1 , of Tyr 35, Tyr 52, Phe 62 and Phe 64 were estimated from ${}^3J_{\text{C}\alpha\text{C}\gamma}$ and ${}^3J_{\text{N}\alpha\text{C}\gamma}$ coupling constants obtained from ${}^{15}\text{N}$ - or ${}^{13}\text{C}'$ - ${}^{13}\text{C}\gamma$ spin echo difference HSQC spectra (Hu et al., 1997).

$\{^1\text{H}\}$ - ${}^{15}\text{N}$ steady-state heteronuclear NOE experiments

To estimate the flexibility of unliganded protein backbone roughly, $\{^1\text{H}\}$ - ${}^{15}\text{N}$ steady-state heteronuclear NOE experiments were performed. The experiments were acquired at 303 K with 500 MHz spectrometer. In the experiments, values of relaxation delay and ${}^1\text{H}$ saturation were set to 3.6 sec and 3.0 sec, respectively. The ${}^1\text{H}$ saturation was achieved with 120 degree pulses applied every 5 ms. The spectra width were 6510.417 Hz (${}^1\text{H}$) and 742.942 Hz (${}^{15}\text{N}$) with 512 (${}^1\text{H}$) x 128 (${}^{15}\text{N}$) complex points. The ${}^1\text{H}$ carrier was set to frequency of the water resonances (4.7 ppm) and the ${}^{15}\text{N}$ carrier was set to 122.471 ppm. Values of $\{^1\text{H}\}$ - ${}^{15}\text{N}$ heteronuclear NOE were obtained from the ratio of intensities of two spectra (measured with and without saturation, respectively). Two sets of spectra with and without saturation were collected for the noise estimation.

3.1.3 Structure calculation of MBD

The NOE connectivities from strong, medium and weak cross peaks were categorized, and assumed to correspond to the upper limits for proton-proton distances of 3.0, 4.0, and 5.0 Å, respectively. In the final steps of structure calculation, restraints were included for 15 slowly exchanging backbone amides obtained at 283 K, pH 5.0 (2.8 - 3.4 Å (N-O), 1.8 - 2.4 Å (H-O)). Structure calculation of MBD1 was performed by DYANA (Guntert, 1998) and XPLOR (Bruenger, 1993), which calculate dynamics with simulated annealing protocols in torsion angle space and Cartesian space, respectively. Final structure was calculated in Cartesian space using X-PLOR. The X-PLOR energy function included a quadratic harmonic potential term for covalent geometry, a quartic harmonic potential term for Van der Waals repulsion, and the

pseudo potential term of NMR restraints. The force constants for NMR pseudo potential was set to 50 kcal mol⁻¹ Å⁻² for distance restraints and 200 kcal mol⁻¹ rad⁻² for torsion angle restraints. Since the NMR data provided no indication of any cis-peptide bonds, the omega dihedral angle was restrained to 180 ± 10 with force constant 500 kcal mol⁻¹ rad⁻². Chirality restraints were also applied. Center averaging of distance involving methylene, methyl, and aromatic ring protons was used with appropriate corrections to the distance. In total, 1270 NOE constraints (491 intra-residual, 302 sequential, 165 medium range and 312 long range) and 44 dihedral angle constraints were used in final calculation. A total of 125 structures were calculated. Of these, the 25 structures, which showed the lowest energy, were selected, and analyzed using MOLMOL (Koradi et al., 1996), AQUA and PROCHECK-NMR (Laskowski et al., 1996).

Experiment	dim	Field (MHz)	Time domain sizes (points)	Dwell time (μ s)	Carrier frequency (ppm)	NS	Tm (ms)
^{15}N-MBD1 sample in 90% $\text{H}_2\text{O}/10\%$ D_2O							
^{15}N -HSQC	2D	500	128(N)x512(HN)	673x76.8	122.47(N)-4.7(HN)	16	
^{15}N -NOESY-HSQC	3D	800	200(H)x32(N)x1024(HN)	50x420.8x41.6	4.7(H)-122.47(N)-4.7(HN)	8	150
^{15}N -TOCSY-HSQC	3D	500	24(N)x70(H)x512(HN)	673x76.8x76.8	122.47(N)-4.7(H)-4.7(NH)	32	70
CN-NOESY	4D	500	15(C)x35(H)x12(N)x512(HN)	165x142x673x76.8	43(C)-2.5(H)-122.47(N)-4.7(HN)	4	100
HMQC-J	2D	800	420(N)x1024(HN)	420.8x41.6	122.47(N)-4.7(HN)	16	
$^{15}\text{N}/^{13}\text{C}$-MBD1 sample in 90% $\text{H}_2\text{O}/10\%$ D_2O							
CBCA(CO)NH	3D	500	23(N)x51(Cab)x512(HN)	673x62.5x76.8	122.47(N)-46(Cab)-4.7(HN)	8	
CBCANH	3D	500	23(N)x29(Cab)x512(HN)	673x62.5x76.8	122.47(N)-46(Cab)-4.7(HN)	32	
HNCO	3D	500	64(CO)x23(N)x512(HN)	350x673x76.8	177.31(CO)-122.47(N)-4.7(HN)	8	
HN(CA)CO	3D	500	23(N)x70(CO)x512(HN)	673x350x76.8	122.47(N)-177.31(CO)-4.7(HN)	16	
C(CO)NH	3D	500	23(N)x64(C)x512(HN)	673x56x76.8	122.47(N)-43(C)-4.7(HN)	16	20.6
H(CCO)NH	3D	500	23(N)x64(H)x512(HN)	673x142x76.8	122.47(N)-4.7(H)-4.7(NH)	16	20.6
$^{15}\text{N}/^{13}\text{C}$-MBD1 sample in 100% D_2O							
^{13}C -HSQC	2D	500	128(C)x512(H)	50x76.8	45.3(C)-4.7(H)	32	
HCCH-TOCSY	3D	500	110(H)x30(C)x512(H)	142x165x76.8	4.7(H)-43(C)-4.7(H)	8	20.6
^{13}C -NOESY-HSQC	3D	800	32(C)x114(H)x1024(H)	77.7x89.2x46.4	39.1(C)-4.7(H)-4.7(H)	8	150
CC-NOESY	4D	800	15(C)x52(H)x15(C)x512(H)	108x72x108x50	44.31(C)-3.6(H)-44.31(C)-4.7(H)	2	100
^{15}N-MBD1 sample in 100% D_2O							
NOESY	2D	800	800(H)x4096(H)	104x41.6	4.7(H)-4.7(H)	32	150
TOCSY	2D	500	512(H)x4096(H)	166x62.4	4.7(H)-4.7(H)	16	50

Table 3.1 Summary of NMR experiments of the DNA-free MBD

NMR experiments and parameters used in the collection of the NMR data of unliganded MBD1. Tm, mixing time; NS, number of scan; dim, dimension of experiment; time domain size are in complex points.

3.2 Structural analysis of MBD1 MBD-DNA complex

3.2.1 Sample preparation of MBD-DNA complex

Preparation of MBD

To avoid a loss of samples caused by degradation in purification steps, we reconstructed GST-MBD1 expression vector using by GeneEditor in vitro site-directed mutagenesis system (Promega) (Shimotake, 2000). The reconstructed vector had a sequence of the residues 1-75 of MBD1 (MBD1₁₋₇₅) instead of 1-84, and showed a good expression of MBD1 in comparison with previous one. This recombinant MBD1₁₋₇₅ bound with same specificity and affinity as MBD1₁₋₈₄ to the methylated DNA (Shimotake, 2000). The pattern of ¹⁵N-¹H HSQC spectrum of MBD1₁₋₇₅ was essentially identical of MBD1₁₋₈₄, which suggested the two proteins had a same tertiary structure. All the samples of MBD1 for NMR measurements of the complex were prepared with this expression vector. This recombinant protein consisted of residues 1-75 of MBD1 with two additional residue derived from the vector on N-terminus (GS). The molecular mass of recombinant MBD1₁₋₇₅ is 8,639 Dalton. The calculated isoelectric point is 8.69. The molar absorption coefficient at 280 nm is calculated to be 16,460.

MBD1₁₋₇₅ recombinant proteins, referred to as MBD were expressed in *E. coli* (DH5α) as the GST fusion protein. The expression procedures of MBD1₁₋₇₅-GST protein were slightly changed from MBD1₁₋₈₄-GST described in chapter 3.1.1. The values changed from previous one were underlined in the following lines. ¹⁵N-labelled and ¹⁵N/¹³C-labelled MBD1₁₋₇₅, referred to as MBD, were expressed in M9 synthetic media containing ¹⁵NH₄Cl (0.5 g/L) and ¹⁵NH₄Cl (0.5 g/L) / ¹³C-glucose (1 g/L), respectively. When ¹⁵N-labelled MBD was expressed, 0.1 % glycerol and 4 g/L glucose were added in M9 medium. The culture was performed at 37 °C. When the OD₆₆₀ reached 0.6, IPTG was added to final concentration of 1 mM. After 3 hr induction at 37 °C, bacteria were harvested by centrifugation (6,000 rpm, 30 min, 4 °C) and stored at -80 °C.

Not isotopically labeled MBD was expressed in LB medium (10 g/L Tryptone, 5 g/L Yeast extract, 10 g/L NaCl, 1 g/L glucose). The culture started at 37 °C and when the OD₆₆₀ reached

0.6, IPTG was added to a final concentration of 1mM. After 3 hrs induction at 37 ° C, bacteria were pelleted by centrifugation (6,000 rpm, 30 min, 4°C) and stored at -80 ° C.

Fractionally deuterated ¹⁵N-labelled MBD containing only Arginine residues labelled with ¹³C, ¹⁵N and ¹H was expressed in M9 media (in this case, M9 medium contained 90 % D₂O, 0.5 g/L ¹⁵NH₄Cl, 1 g/L ¹³C-D-glucose and 100 mg/L ¹³C, ¹⁵N-labelled Arginine). The culture started at 37 ° C and when the OD₆₆₀ reached 0.25, IPTG was added to a final concentration of 1mM. After 6 hr induction at 37 ° C (OD₆₆₀ reached 0.4), bacteria were pelleted by centrifugation (6,000 rpm, 30 min, 4 ° C) and stored at -80 ° C. In this medium contained 90 % D₂O, the bacteria growth became very slow (doubling time 2 hr) because of isotope effects.

A purification procedure of MBD was the same as MBD₁₋₈₄ (as described previously, in chapter 3.1.1). The cells were sonicated and centrifuged, and then fusion protein was purified from its supernatant using by glutathione column. After the elution from glutathione column, the fusion protein was cleaved by thrombin. Finally, MBD was purified from the cleavage solution by Mono S FPLC column and centrifuged. Sample integrity was confirmed by N-terminal sequencing, TOF-mass spectroscopy, and SDS-PAGE.

Preparation of 12-mer methylated DNA

The palindromic 12-mer deoxy-oligonucleotide containing a single methyl-CpG in the center of the sequence, 5'-GTATCMGGATAC-3' (M; 5-methyl-cytosine) was chemically synthesized at the NIPPON SEIFUN Corporation (npdna@nippon.co.jp) at 500 nmol scale with HPLC grade. The synthesized oligonucleotide was dissolved in distilled water and loaded onto a Resource-Q anion-exchange column (Pharmacia, 1ml). After washing the column with 100 mM KCl solution, the oligonucleotide was eluted with 1M KCl solution. The solution was then applied onto HiTrap desalting column (Pharmacia, 5ml x 3 column) for desalting and eluted with distilled water. Dried oligonucleotide was dissolved in an annealing solution containing 50 mM KCl, and annealed by decreasing the temperature from 95 ° C to 4 ° C for 3hr. The annealed oligonucleotide was checked by 10% native-PAGE, performed at 4 ° C. The oligonucleotide was stored at 4 ° C.

The concentration of DNA solution was obtained from the value of the absorption (260 nm) at 80 °C. The molar absorption coefficient at 260 nm is calculated to be 1.2×10^5 . The melting temperature of this oligonucleotide was calculated as 36 °C.

Preparation of MBD-DNA complex

The MBD-DNA complex was formed by titration of the 0.2 mM oligonucleotide into the 0.2 mM protein, tracing resonances both of the protein amide groups in the ^{15}N - ^1H HSQC spectrum and the DNA imino protons in the ^1H 1D jump-and-return spectrum (Cavanagh et al., 1996) to indicate 1:1 stoichiometry. The formed MBD-DNA complex was concentrated using a Centricon-3 (Amicon, 3,000 M.W. cutoff) for NMR measurements. The resonances of the complex were in slow exchange on the chemical shift time scale.

3.2.2 NMR measurements of MBD-DNA complex

NMR spectra were acquired at 303 K with a Bruker DMX500, DRX500 or DRX800 NMR spectrometer equipped with a triple resonance probe and triple axis gradient. Samples for NMR measurements typically comprised 1.3 mM MBD-DNA complex in 20 mM potassium phosphate buffer (pH 6.5), 5 mM deuterated DTT (d_{10} -DTT). Other NMR conditions were same as DNA-free MBD1. The Table 3.2 summarized these experimental parameters and combinations of the experiment and the sample.

Resonance assignments of MBD

For assignments of the ^1H , ^{15}N and ^{13}C resonances of protein in complex, a series of two- and three-dimensional experiments were performed (2D ^{15}N - ^1H HSQC, ^{13}C - ^1H HSQC, 3D CBCA(CO)NH, CBCANH, HNC(O), HN(CA)CO, C(CO)NH, H(CCO)NH, HCCH-TOCSY and ^{15}N -TOCSY-HSQC) with the ^{15}N - or ^{15}N , ^{13}C -labelled protein dissolved in 90% H_2O /10% D_2O or 100% D_2O solution. (Cavanagh et al., 1996).

The stereospecific assignment of the methyl groups of the Leu and Val residues was achieved using a 15% fractionally ^{13}C -labelled protein. To avoid ambiguity, several side-chain chemical shifts were assigned by the experiments which use one-bond scalar coupling for coherence transfer. 2D (HB)CB(CGCD)HD and 2D (HB)CB(CGCDCE)HE experiments were used for assignment of $\text{H}\delta$ and $\text{H}\epsilon$ chemical shifts of aromatic residues (Yamazaki et al., 1993). Assignment of guanidine ^1H and ^{15}N chemical shifts of Arg residues was obtained, using a fractionally deuterated, ^{15}N -labelled sample containing only Arg labelled with ^{13}C , ^{15}N , and ^1H , as described (Yamazaki et al. 1995). The side-chain CG chemical shifts of Asp were obtained from 2D CO(CA)HA experiment (Dijkstra et al., 1994).

Structure restraints of MBD

Distance information of protein was obtained by means of 3D ^{15}N - or 4D ^{15}N , ^{13}C - or ^{13}C , ^{13}C -NOESY experiments with a mixing time of 120 ms (Cavanagh et al., 1996).

Backbone ϕ and ψ torsion angle restraints were derived from $^3J_{\text{HNH}\alpha}$ coupling constant obtained from 3D HNHA spectrum, in combination with a database search based on ^{15}N , H_N , $\text{C}\alpha$, $\text{C}\beta$, C' and $\text{H}\alpha$ secondary chemical shifts using the program TALOS (Cornilescu et al. 1999). The torsion angles χ^1 of Trp 15, Tyr 35, Tyr 52, Phe 62 and Phe 64 were estimated from $^3J_{\text{C}'\text{C}\gamma}$ and $^3J_{\text{N}\text{C}\gamma}$ coupling constants obtained from ^{15}N - or $^{13}\text{C}'$ - $^{13}\text{C}\gamma$ spin echo difference HSQC spectra (Hu et al., 1997). Those of Asp 7, Arg 22, Asp 32 and Tyr 34 were estimated from $^3J_{\text{NH}\beta}$, $^3J_{\text{C}'\text{C}\gamma}$ and/or $^3J_{\text{H}\alpha\text{H}\beta}$ couplings, obtained from 3D- HNHB, HN(CO)CO, and HACAHB-COSY spectra (Vuister et al., 1998, Grzesiek et al., 1995).

Resonance assignments of free-DNA

Free DNA samples for NMR measurements comprised 0.7 mM DNA in 20 mM potassium phosphate buffer (pH 6.5), 50 mM KCl, 5 mM deuterated DTT (d_{10} -DTT) at 303K in D_2O solution. Assignments of proton chemical shifts of the free-DNA were obtained from analysis of

^1H - ^1H 2D NOESY and TOCSY recorded in D_2O , employing the sequential assignment procedure for DNA (Wuthrich, 1986) (Table 3.2).

Resonance assignments and structure restraints of DNA in complex

Assignment of ^1H chemical shifts of the DNA was obtained from analysis of 2D (F1, F2)- ^{13}C -filtered TOCSY (30 ms mixing at rf field strength of 8.6 kHz) experiment and four NOESY experiments at 20, 40, 60, 120 ms mixing time, respectively. In these experiments, adiabatic WURST pulse was used for wide range inversion (80 kHz width at 500 MHz spectrometer) of ^{13}C magnetization. Assignments for DNA in complex were obtained in a similar manner as free-DNA. Imino proton chemical shifts were obtained from 2D Jump-and-Return NOESY experiments with 120 ms mixing time (Sklenar and Bax, 1987).

Interproton distances within the DNA were derived from 2D ^1H - ^1H NOE spectra and (F1, F2)- ^{13}C -filtered NOE spectra at a mixing time of 40 ms. Restraints of δ angles for all sugar rings were estimated based on (H1', H2') or (H3', H4') cross peak patterns in a DQF-COSY spectrum.

When the asymmetric MBD monomer is bound to the symmetric double-stranded oligonucleotide, each strand of the DNA interacts with the MBD in a distinctly different way, and thus exhibits chemical shifts that are different from those of the opposite strand. For the MBD-DNA complex, we observed NOESY cross peaks that arose from chemical exchange of the two DNA strands, as a consequence of a reversal in the orientation of the bound MBD on time scale of the NOESY mixing time, as seen for example, in the case of HIV Protease/inhibitor KNI-529 (Kato et al., 1999). Several cross peaks arose from a combined effect of the chemical exchange and NOE at a mixing time of 120 ms. We distinguished cross peaks that were due to the chemical exchange process by measuring NOE spectra at a variety of mixing times, 40, 60, and 120 ms at temperatures varying from 15 to 35 °C, as described Kato et al.. The build-up rates of these exchange-originating cross peaks were found to be strongly dependent on temperature, increasing as the temperature increased (Kato et al., 1999). We also measured ROESY spectrum (40ms mixing at rf field strength of 2.5 kHz) in order to identify cross peaks that arose solely

from the chemical exchange. Intensities of cross peaks that arose from a combined effect of the chemical exchange and NOE are sufficiently small in NOE spectra at a mixing time of 40 ms; thus, all interproton distances within the DNA. For the protein, cross peaks that were due to the chemical exchange process were not observed.

Intermolecular restraints

Assignments of intermolecular interproton NOEs were identified using 3D (F1)-¹³C-filtered (F2)-¹³C-edited NOESY-HSQC and series of two dimensional NOESY spectra, ((F1,F2)-¹³C-filtered NOESY, (F1)-¹³C-decoupled (F2)-¹³C-filtered NOESY, and (F2)-¹³C-filtered NOESY), of ¹⁵N, ¹³C-labelled MBD1-DNA complex in D₂O solution. In these isotope-filtered experiments, adiabatic WURST pulse was used for wide range inversion (80 kHz width at 500 MHz spectrometer) of ¹³C magnetization. Intermolecular interproton distance restraints including protein main chain or Arg epsilon protons were identified using 3D ¹⁵N-separated NOESY and 3D (F1)-¹³C/¹⁵N-filtered (F2)-¹⁵N-edited NOESY spectra of fractionally deuterated, ¹⁵N-labelled MBD1 sample containing only Arg residues labelled with ¹³C, ¹⁵N, and ¹H.

Intermolecular interproton distances were calculated from the intensity of assigned NOE peaks of 2D (F2)-¹³C-filtered NOESY and 2D ¹H-¹H NOESY spectra at a short mixing time of 40 ms to avoid the effects of chemical exchange as described previously.

{¹H}-¹⁵N steady-state heteronuclear NOE experiments

To roughly estimate the flexibility of protein backbone in complex with DNA, {¹H}-¹⁵N steady-state heteronuclear NOE experiments were also performed in MBD-DNA complex to estimate the flexibility change induced upon DNA binding. The parameters for NMR measurements of MBD in complex were essentially the same as those of unliganded MBD. The values of relaxation delay and ¹H saturation were set to 3.6 sec and 3.0 sec respectively, and the saturation was achieved with 120 degree pulses applied every 5 ms. The spectra width were 6510.417 Hz (¹H) and 886.997 Hz (¹⁵N) with 512 (¹H) x 110 (¹⁵N) complex points. The ¹H carrier was set to

frequency of the water resonances (4.7 ppm) and the ^{15}N carrier was set to 122.250 ppm. The two sets of spectra with and without saturation were collected for the noise estimation.

3.2.3 Structure calculation of MBD-DNA complex

The structures of the MBD1-DNA complex were calculated with simulated annealing protocols using X-PLOR (Bruenger 1993). The NOE connectivities within the protein, and between the protein and DNA, from strong, medium, weak, and very weak cross peaks were categorized, and assumed to correspond to the upper limits for proton-proton distances of 3.0, 4.0, 5.0 and 6.0 Å, respectively. Interproton distance restraints for the DNA were classified into five ranges, 1.8-2.5, 1.8-3.0, 1.8-3.5, 2.3-5.0, and 3.5-6.0 Å. Distances involving non-stereospecifically assigned methylene protons and aromatic ring protons in the protein were represented as a $(\sum r^{-6})^{-1/6}$ sum (Omichinski et al., 1997). Watson-Crick base pairing was maintained in the DNA by the following hydrogen bond restraints: for GC base pairs $r_{\text{G(N1)-C(N3)}}=2.95 \pm 0.2$ Å, $r_{\text{G(N2)-C(O2)}}=2.86 \pm 0.2$ Å, and $r_{\text{G(O6)-C(N4)}}=2.91 \pm 0.2$ Å; for AT base pairs $r_{\text{A(N1)-T(N3)}}=2.82 \pm 0.2$ Å, and $r_{\text{A(N6)-T(O4)}}=2.86 \pm 0.2$ Å (Wojciak et al., 1999). Loose torsion angle restraints for the DNA were used to alleviate problems associated with mirror images, covering both A- and B-form DNA conformers ($\alpha = -65 \pm 50^\circ$, $\beta = 180 \pm 50^\circ$, $\gamma = 60 \pm 50^\circ$, $\epsilon = 180 \pm 50^\circ$ and $\zeta = -85 \pm 50^\circ$) (Wojciak et al., 1999; Omichinski et al., 1997). In addition, δ angles for all sugar rings were estimated to be $145 \pm 50^\circ$, based on (H1', H2') or (H3', H4') cross peak patterns in a DQF-COSY spectrum. These hydrogen bond and torsion angle restraints for DNA are justified, because the pattern of NOEs for the DNA is typical of B-DNA. Weak planarity restraints for base pairs were included during the simulated-annealing protocol (a force constant of $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$). The X-PLOR target function also included a quadratic harmonic potential term for covalent geometry, a quartic harmonic potential term for van der Waals repulsion, and the pseudo potential term of NMR restraints. In the final steps of the structure determination, hydrogen bond restraints within the protein were included for 15 slowly exchanging protein backbone amides obtained HD-exchange experiments (2.8-3.4

Å (N-O), 1.8-2.4 Å (H-O)). No protein-DNA interfacial hydrogen bonds were assumed during the calculation of the complex.

The procedure for calculation of the complex structure was essentially same as described by Omichinski and Wright (Omichinski et al., 1997; Wright et al., 1997). First, the structure of the protein was calculated by simulated annealing using only restraints within the protein. Then, the simulated annealing calculation of the complex was performed using all the experimental restraints, starting from B-form DNA and the calculated structure of the protein, using a variety of protein-DNA orientations with a separation of about 50 Å. In total, 200 structures of the complex were calculated. Of these, 81 structures exhibited neither distance violations greater than 0.3 Å nor dihedral violations greater than 5 degree. These structures were further refined by a simulated annealing procedure. Finally, the 20 structures that showed the lowest energy were selected and analyzed with MOLMOL (Koradi et al., 1996), AQUA, PROCHECK-NMR (Laskowski et al., 1996) and CURVES (Dickerson et al., 1989).

Experiment	dim	Field (MHz)	Time domain sizes (points)	Dwell time (ns)	Carrier frequency (ppm)	NS	Tm (ms)	Information
[¹⁵N-MBD-DNA, 90% H₂O/10% D₂O]								
¹ H-HSQC	2D	500	128(Nx512)(IN)	563.7x76.8	122.25(N)-4.7(HN)	32		MBD Assign
¹ N-NOESY-HSQC	3D	500	128(1N)x128(Nx512)(IN)	84x563.7x76.8	4.7(H)-122.25(N)-4.7(HN)	16	100	MBD/MBD-DNA NOEs
¹ N-TOCSY-HSQC	3D	500	32(N)x80(1N)x512(IN)	563.7x84x76.8	122.25(N)-4.7(H)-4.7(HN)	16	72	MBD Assign
CN-NOESY	4D	500	16(CN)x512(N)x512(Nx512)(IN)	165x142x563.7x76.8	4.7(C)-2.5(D)-122.25(N)-4.7(HN)	4	100	MBD NOEs
HN1FA	3D	500	45(D)x25(N)x512(IN)	84x563.7x76.8	4.7(H)-122.25(N)-4.7(HN)	32		MBD 3J
HN1FB	3D	500	100(HN)x30(N)x512(IN)	86x563.7x76.8	4.7(H)-122.25(N)-4.7(HN)	16		MBD 3J
IR-NOESY	2D	800	512(Hx8192(H))	56x8x25	4.7(H)-4.7(H)	96	60/20	DNA(amino) Assign/NOEs
[¹⁵N¹³C-MBD-DNA, 90% H₂O/10% D₂O]								
CBCA(CO)NH	3D	500	23(N)x56(Cx)N512(IN)	563.7x62.5x76.8	122.25(N)-46(Cx)-4.7(HN)	8		MBD Assign
CBCAMH	3D	500	23(N)x56(Cx)N512(IN)	563.7x62.5x76.8	122.25(N)-46(Cx)-4.7(HN)	32		MBD Assign
HNCO	3D	500	64(CO)x23(N)x512(IN)	35x563.7x76.8	177.31(CO)-122.25(N)-4.7(HN)	8		MBD Assign
HN(CA)CO	3D	500	23(N)x64(CO)x512(IN)	563.7x35x76.8	122.25(N)-177.31(CO)-4.7(HN)	16		MBD Assign
C(CO)NH	3D	500	23(N)x64(CO)x512(IN)	563.7x35x76.8	122.25(N)-4.7(HN)	16	20.6	MBD Assign
H(CCO)NH	3D	500	23(N)x96(HN)x512(IN)	563.7x142x76.8	122.25(N)-4.7(H)-4.7(HN)	16	20.6	MBD Assign
[¹⁵N¹³C-MBD-DNA, 100% D₂O]								
¹³ C-HSQC	2D	500	30(N)x512(D)	50x62.4	45.3(C)-4.7(H)	32		MBD Assign
¹³ C-HSQC	2D	500	128(CN)x512(D)	100x83.2	127.3(C)-4.7(H)	64		MBD(assign) Assign
H(C)H-TOCSY	3D	500	110(HN)x30(CN)x512(D)	142x105x76.8	4.7(H)-4.7(C)-4.7(H)	16	20.6	MBD Assign
CC-NOESY	4D	800	15(CN)x512(HN)x64(CN)x512(D)	108x98x108x52	44.3(C)-2.5(D)-44.31(C)-4.7(H)	2	100	MBD NOEs
HACAHB-COSY	3D	500	54(Hx62(CN)x512(D))	111.1x120x55.6	4.7(H)x56.3(CN)-4.7(H)	8		MBD 3J
HNCOCCO	3D	500	20(N)x32(CO)x512(D)	563.7x35x87x14	122.25(N)-176.78(CO)-4.7(H)	64		MBD 3J
CBHD	2D	500	18(CN)x512(D)	22x62.4	35.78(CN)-4.7(H)	16		MBD(assign) Assign
CRHE	2D	500	18(CN)x512(D)	22x62.4	35.78(CN)-4.7(H)	16		MBD(assign) Assign
F1- ¹³ C-filtered F3- ¹³ C-edited NOESY	3D	800	110(HN)x24(CN)x512(D)	113.6x77.6x41.6	4.7(H)x39.1(CN)-4.7(H)	32	100	MBD-DNA NOEs
F1- ¹³ C-filtered F3- ¹³ C-edited NOESY	3D	500	128(HN)x36(CN)x512(D)	176x62.5x62.4	4.7(H)x42.31(CN)-4.7(H)	16	120	MBD-DNA NOEs
F1,F2- ¹³ C-filtered NOESY	2D	800	400(HN)x4096(H)	113.6x41.6	4.7(H)x4.7(H)	160	20/40/60/20	DNA Assign/NOEs
F1,F2- ¹³ C-filtered ROESY	2D	800	400(HN)x4096(H)	113.6x41.6	4.7(H)x4.7(H)	80	31.1	DNA Assign
COSY	2D	500	256(HN)x2048(H)	176x62.4	4.7(H)x4.7(H)	128	40	DNA Assign
F1- ¹³ C-filtered NOESY(F1- ¹³ C-decoupled)	2D	800	1024(HN)x4096(H)	113.6x41.6	4.7(H)x4.7(H)	64		DNA 3J
F1- ¹³ C-filtered NOESY	2D	800	310(HN)x4096(H)	113.6x41.6	4.7(H)x4.7(H)	64	40/60/20	MBD-DNA NOEs
F1- ¹³ C-filtered NOESY	2D	800	400(HN)x4096(H)	113.6x41.6	4.7(H)x4.7(H)	176	120	MBD-DNA NOEs
[¹⁵ N ¹³ C/ ¹⁵ N-filtered F3- ¹³ N-edited NOESY	3D	800	128(HN)x21(NN)x2048(H)	83.3x352.4x38.4	4.7(HN)-122.25(N)x4.7(H)	16	120	MBD-DNA NOEs
[¹⁵N-MBD-DNA, 100% D₂O]								
NOESY	2D	800	512(HN)x192(H)	100x41.6	4.7(H)-4.7(H)	80	40/20	MBD(assign)/MBD(assign)-DNA NOEs
TOCSY	2D	500	200(HN)x2048(H)	174x62.4	4.7(H)-4.7(H)	96	68.4	MBD(assign) Assign
[Free DNA, 100% D₂O]								
NOESY	2D	500	400(HN)x4096(H)	176x50	4.7(H)x4.7(H)	48	100	DNA Assign/NOEs
TOCSY	2D	500	400(HN)x4096(H)	176x50	4.7(H)x4.7(H)	32	50	DNA Assign

Table 3.2 Summary of NMR experiments of the MBD-DNA complex

NMR experiments and parameters used in the collection of the NMR data of MBD-DNA complex. Tm, mixing time; NS, number of scan; dim, dimension of experiment; time domain size are in complex points.

3.3 Mutational analyses

We also performed a mutational analysis to obtain the effects of the substitutions to DNA binding activity, which could not be obtained from structural data directly. For the mutational analysis, the residues expected to play an important role for DNA binding were selected and its site-directed mutants were prepared in a GST-fusion form, and each was examined as to the potential to bind to a double-stranded oligonucleotide carrying a methyl-CpG pair, by means of gel-shift assays.

Based on a large chemical shift change caused by DNA binding, the sequence homology in the MBD family, and the NMR structure, the following 11 residues were selected; Arg 22, Thr 27, Arg 30, Asp 32, Tyr 34, Arg 44, Ser 45, Lys 46, Tyr 52, Phe 64, and Lys 65. All of these residues were changed to alanine. In addition to these substitutions to alanine, the residue suggested the possibility of the base recognition was changed to another residue of similar property; Arg 22, Arg 30, and Arg 44 to lysine, and Tyr 34 to phenylalanine. Consequently, we prepared following 15 mutants of MBD1 for mutational analysis; R22A, R22K, T27A, R30A, R30K D32A, Y34A, Y34F, R44A, R44K, S45A, K46A, Y52A, F64A, and K65A mutants.

Mutant constructs were prepared from GST-MBD1₁₋₈₄ expression vector (pGEX-2TH; Pharmacia) with GeneEditor in vitro site-directed mutagenesis system (Promega), according to the manufacturer's instructions. The coding regions of all mutants were sequenced. GST-MBD1 fusion proteins were purified and used for DNA binding assay. The cold 12-mer methylated and unmethylated DNA (5'-GTATCMGGATAC-3', M; 5-methyl-cytosine or cytosine) were prepared as described in chapter 3.2.1.

To examine DNA binding affinity, reactions were carried out in 10 μ l of 10 mM Tris-HCl, pH 8.0, 5 mM MgCl₂, 5 mM DTT, 5 % glycerol. The mixtures contained 1.7 ng of each purified protein and 74 ng of the methylated or unmethylated DNA. The reactions were carried for 30 min at room temperature. The binding reaction mixtures were loaded on a 10% acrylamide gel and run in TBE buffer. The gels were stained with SYBR green I (Molecular Probes).

The tertiary structures of each mutant were checked by the ^{15}N - ^1H HSQC spectrum. Of all mutants, only F64A MBD1 was showed a mutation-induced disruption.

4 RESULTS

4.1 Sample preparation of free MBD and MBD-DNA complex

The methyl-CpG binding domain (MBD) of MBD1, comprising residues 1-75, was expressed in *E. coli* with the addition of ^{13}C -glucose and $^{15}\text{NH}_4\text{Cl}$ to prepare 99% isotopically enriched protein. The recombinant MBD₁₋₇₅, referred to as MBD, was purified chromatographically as described in “Materials and Methods” (Chapter 3), and final purity was more than 95% judged by 17.5% SDS-PAGE. The MBD was in the soluble fraction through the purification steps. Finally yields of 2.2 mg protein per liter of minimal M9 culture, which contained 1g/l ^{13}C -glucose and 0.5 g/l $^{15}\text{NH}_4\text{Cl}$, could be obtained. Yields of 1.1 mg/l protein were obtained from the M9 culture containing 90 % D₂O.

The condition of NMR measurements of the free MBD and MBD-DNA complex was decided previous to starting the structural analysis. ^{15}N - ^1H HSQC NMR experiment provides a spectrum, which correlates the resonances of an amide proton ($^1\text{H}_\text{N}$), and the directly attached amide nitrogen (^{15}N). In the HSQC spectrum, each peak represents the backbone amide group of each amino acid or side-chain amide/amino groups. Therefore, a condition of a protein can be judged from its dispersion, a number, and intensities of peaks. Comparison of the HSQC spectra recorded in various conditions show that free MBD is soluble and stable in following conditions;

pH: 5.0-6.5

Ionic strength (KCl): 0 - 150 mM

Temperature: 293K - 310K

Protein concentration: < 2.0 mM.

For NMR measurements, a lower pH value, lower salt, higher protein concentration, and higher temperature are generally preferable. Consequently, the final NMR condition for free MBD was decided to be 20mM potassium phosphate buffer (pH 6.5), 50 mM KCl, 5 mM DTT at 303K. In this condition, MBD gave well-dispersed ^{15}N - ^1H HSQC spectrum and was stable for 5 months.

MBD in complex with DNA was also stable and showed good dispersions in the same condition of free MBD. Generally, most of the interactions between protein and DNA were held by electrostatic forces, so a lower ionic strength was preferable to increase the electrostatic interaction. As the spectrum pattern was nearly the same in the range of the ionic condition from 0 to 50 mM of KCl, the final NMR condition for MBD-DNA complex was decided to 20mM potassium phosphate buffer (pH 6.5), 5mM DTT without salt at 303K. In this condition, MBD in complex gave well-dispersed ^1H - ^{15}N HSQC spectrum and MBD-DNA was stable for 3 months.

4.2 Resonance assignments

DNA-free MBD

The resonance assignments for the backbone $^1\text{H}/^{15}\text{N}/^{13}\text{C}$ chemical shifts of free MBD were completed except for the vector-derived residues, Gly-2 and Ser-1, on N-terminus. The resonance of these two residues could not be observed in ^{15}N - ^1H HSQC spectrum, probably due to conformational averaging or hydrogen exchange with solvent. Most of the side-chain chemical shifts were assigned except for Gly-2 and Ser-1. All the methyl groups of Leu and Val were assigned stereospecifically with a 15% fractionally ^{13}C labeled protein.

MBD in complex with DNA

The resonance assignments for the backbone $^1\text{H}/^{15}\text{N}/^{13}\text{C}$ chemical shifts of MBD in complex with DNA were completed except for the vector-derived residues, Gly-2 and Ser-1. These resonances could not be observed in spectra of the complex. Most of the side-chain chemical shifts were assigned except for Gly-2 and Ser-1. All the methyl groups of Leu and Val were assigned stereospecifically. In complex, guanidine ^1H and ^{15}N chemical shifts of all the arginine residues were further assigned. Fig. 4.1 shows a ^{15}N - ^1H correlation HSQC spectrum of the DNA-bound and DNA-free MBD with the assignment of backbone and side-chain amide/amino groups.

DNA in complex with MBD, and free DNA

The resonance assignments for non-exchangeable protons of 12 base pairs DNA in complex were complete for H1', H2', H2'', H3', H4', H2, H5, H6, H8 and methyl groups of thymine and methyl cytosine. Generally, when the symmetric double-stranded oligonucleotide binds to asymmetric protein, each strands of oligonucleotide give different chemical shifts. For the MBD-DNA complex, each strand of the base pairs 3-10 gave different chemical shifts. Only two base pairs in the both ends of DNA gave same chemical shifts, because the both ends of DNA were little influenced by MBD-DNA interaction. The resonances in the base pairs 3-10 were also affected by chemical exchange caused by MBD-DNA interaction. Assignments of the imino proton of DNA bases were complete except for two base pairs in the both ends of DNA probably due to signal overlapping. At 283 K, 11 imino signals out of 12 base pairs were observed in the chemical shift range from 12 ppm to 14 ppm, indicating that these imino protons formed Watson-Crick type hydrogen bonds.

The resonances of free DNA were also assigned to compare with those of DNA in complex with MBD. The resonance assignments for non-exchangeable protons of free DNA were also complete for H1', H2', H2'', H3', H4', H2, H5, H6, H8 and methyl groups of thymine and methyl cytosine. In the case of a symmetric double-stranded oligonucleotide, each strand gives same chemical shifts. Actually, in this free DNA, each strand had same chemical shifts.

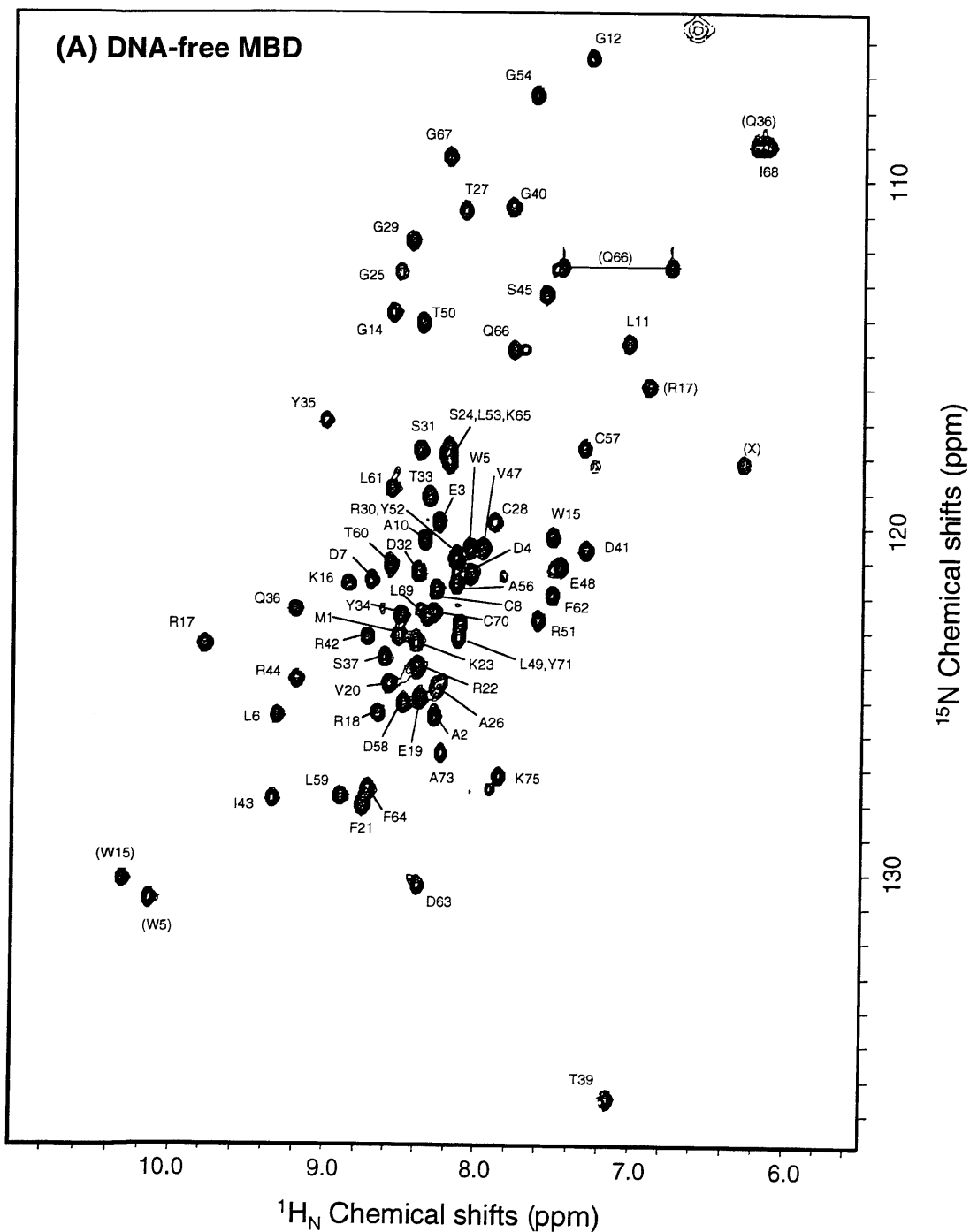


Figure 4.1 Resonance assignments of DNA-free and DNA-bound MBD.

(A) ^{15}N - ^1H HSQC spectrum of DNA-bound MBD. The assignments of the backbone amide groups of DNA-free MBD are shown. The assignments in parentheses indicate those of side-chain amide groups. Figure 4.1(B) is in the next page.

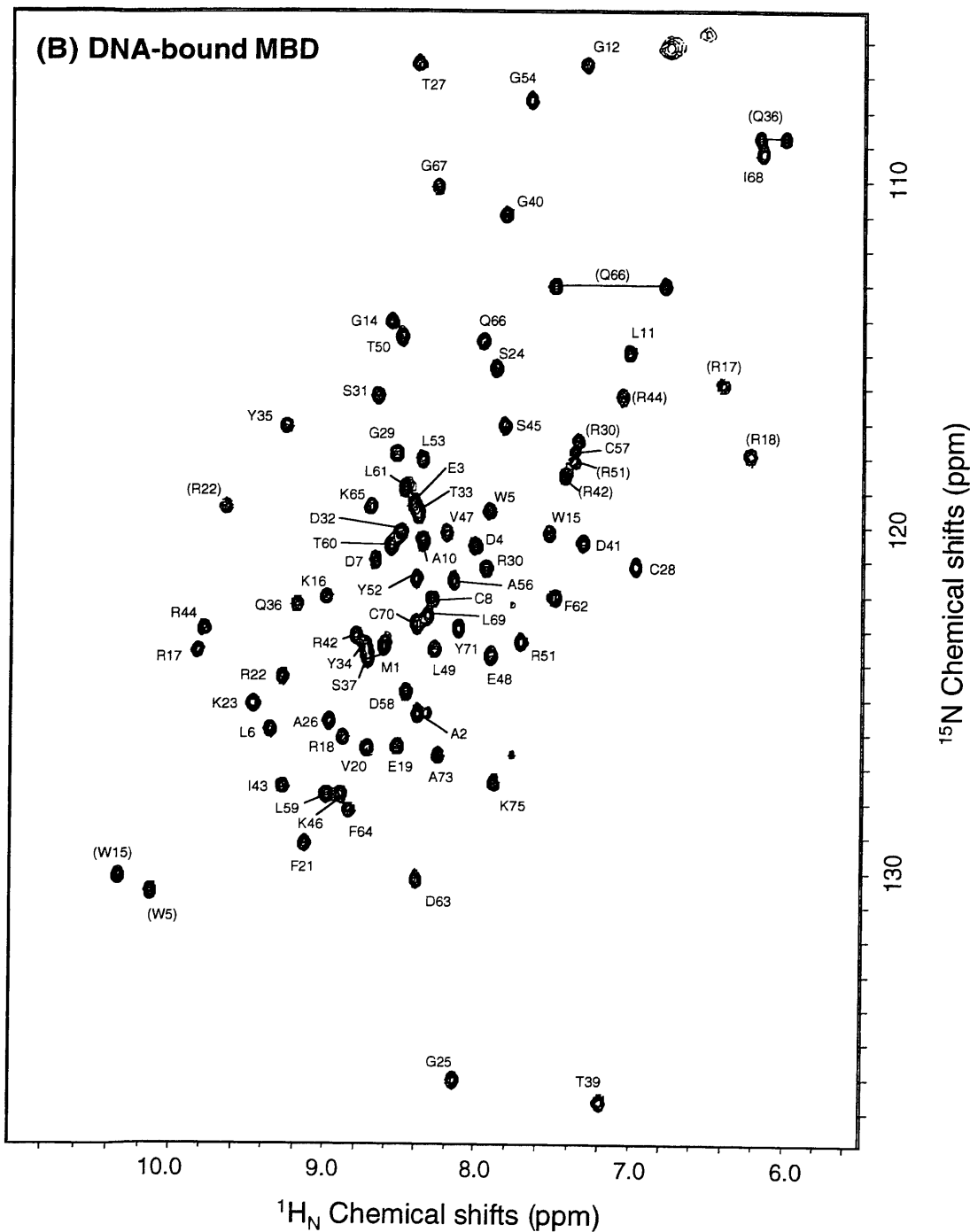


Figure 4.1 Resonance assignments of DNA-free and DNA-bound MBD.

(B) ^{15}N - ^1H HSQC spectrum of DNA-bound MBD. The assignments of the backbone amide groups of DNA-bound MBD are shown. The assignments in parentheses indicate those of side-chain amide groups.

4.3 Structure determination and structural statistics

DNA-free MBD

Structure of MBD was determined from a total of 1329 NMR-derived distance and torsion angle restraints. A total of 125 structures was calculated, and 98 of them showed no violation greater than 0.5 Å or 5 degree. Of these, the 25 structures, which showed the lowest energy, were selected, and analyzed. The statistics for the 25 final structures are shown in Table 4.1. The backbone and hydrophobic side chains have been well defined, except for a loop region (L2; residues 22-30), and N-terminal and C-terminal residues. For loop L2, few long-range NOEs were observed. The local sequential NOE patterns and ^1H - ^{15}N heteronuclear NOE values of this region are characteristic of an unstructured and flexible loop. In addition to this region, the N-terminal residues, including Met1 and Ala2, and the C-terminal six residues have not been defined regarding structure.

The root-mean-square deviation (rmsd) of residues 3-69 as to the mean coordinate positions, excluding loop L1, is 0.45 Å for the backbone heavy atoms, and 0.98 Å for all heavy atoms.

MBD-DNA complex

Structure of MBD-DNA complex was determined from a total of 2022 NMR-derived distance and torsion angle restraints (1452 for protein, 479 for DNA, and 91 for protein-DNA). The statistics for the 20 final structures are shown in Table 4.2. The MBD-DNA complex was well defined except for N-terminal and C-terminal residues of protein and the both end of DNA. The N-terminal two and C-terminal six residues of the protein, and two base pairs in the both ends of DNA have not been defined regarding structure. These residues were far from the MBD-DNA interface.

The rmsd of residues 3-69 of the protein and base pairs 3-10 of the DNA is 0.57 Å for the protein backbone and DNA heavy atoms, and 0.82 Å for all heavy atoms.

Interactions between MBD and DNA were identified based on the calculated ensemble structures. MBD-DNA contacts illustrated in Figure 4.6 (Chapter 4.5) were observed in most members of the NMR ensemble, except for the hydrogen bond from the Arg 44 guanidinium to either N7 or O6 of G19, or to the both of them. This Arg 44-G19 hydrogen bond was assumed based on the following observations. First, single substitution of Arg 44 to either alanine or lysine totally abolished DNA binding, indicating that Arg 44 is important for DNA binding (Fig. 4.6). Then, in members of the NMR ensemble, only N7 or O6 of G19 and/or G20 can accept hydrogen bond from the N ϵ atom of Arg 44, while no other group in the DNA is found to accept the bond. Of these two Arg 44-base contacts, the significance or existence of the Arg 44-G20 hydrogen bond can be excluded by the observation that MBD1 MBD can efficiently bind to DNA of the sequence 5'-GATCGAm⁵CGACGTAC-3' (data not shown). This observation indicates that the base following the methyl CpG sequence (positions 8 and 20 in Figs. 4.5) can be adenine or thymine for the specific binding. Taken together, we assumed that the Arg 44-G19 hydrogen bond found in the NMR ensemble is significant to DNA recognition.

NMR restraints for structure calculation	
Distance restraint	1,315
Intra-residue	491
Sequential i-j =1	302
Short-range 1< i-j _4	206
Long-range i-j >4	301
Hydrogen bond	15
Dihedral angle restraints (ϕ)	44
Statistics for structure calculation	
R.m.s.d. from experimental restraints	
Distances (Å)	0.007±0.0007
Angles (°)	0.23±0.08
Deviations from idealized geometry	
Bonds (Å)	0.0024±0.00001
Angles (°)	0.67±0.002
Impropers (°)	0.36±0.005
PROCHECK analysis (residues 3-20,32-69)	
Most favored regions	75.8 %
Additionally allowed regions	20.7 %
Generously allowed regions	3.1 %
Disallowed regions	0.4 %
Coordinate precision (residues 3-20,32-69)	
Protein backbone atoms	0.45 Å
All protein heavy atoms	0.98 Å

Table 4.1 Structural statistics for DNA-free MBD.

The reported values refer to the final 25 simulated annealing structures. None of the structures exhibited distance restraint violations > 0.3 Å or dihedral angle restraint violations > 5°. Coordinate precision is defined as the average r.m.s. difference between the final 25 structures and the mean coordinates. The backbone atoms include the N, C α and C' atoms.

NMR restraints for structure calculation	
Protein	
Distance restraint	1,389
Intra-residue	520
Sequential i-j =1	324
Short-range 1< i-j _4	178
Long-range i-j >4	341
Hydrogen bond	26
Dihedral angle restraints (ϕ, ψ, χ_1)	34,20,9
DNA	
Distance restraints (a)	343
Dihedral angle restraints	136
Protein-DNA distance restraints	91
Statistics for structure calculation	
R.m.s.d. from experimental restraints	
Distances (Å)	0.019 ± 0.0003
Angles (°)	0.12 ± 0.03
Deviations from idealized geometry	
Bonds (Å)	0.0044 ± 0.00002
Angles (°)	0.91 ± 0.002
Impropers (°)	0.35 ± 0.01
PROCHECK analysis (residues 3-69 of the protein)	
Most favored regions	64.7 %
Additionally allowed regions	28.1 %
Generously allowed regions	5.7 %
Disallowed regions	1.4 %
Coordinate precision	
Protein backbone plus DNA heavy atoms	0.57 Å
All protein heavy plus DNA heavy atoms	0.82 Å
Protein backbone atoms	0.42 Å
All protein heavy atoms	0.87 Å
DNA heavy atoms	0.46 Å

(a) There were a total of 140 intra-residue, 148 sequential intra-strand, 25 inter-strand and 30 hydrogen bond distance restraints.

Table 4.2 Structural statistics for MBD-DNA complex .

The reported values refer to the final 20 simulated annealing structures. None of the structures exhibited distance restraint violations > 0.3 Å or dihedral angle restraint violations > 5°. Coordinate precision is defined as the average r.m.s. difference between the final 25 structures and the mean coordinates. The values refer to residues 3-69 of the protein and base pairs 3-10 of the DNA. The backbone atoms include the N, C α and C' atoms.

4.4 Solution structure of MBD-DNA Complex

Structure of MBD

The three-dimensional structure of the MBD in complex with DNA has a compact fold, the N- and C- termini being on opposite faces of the molecule. Figure 4.2 shows the protein backbone of the 20 final structures derived from NMR data. It assumes an α/β -sandwich fold: one layer consists of a four-stranded twisted β -sheet (strand β A, Leu 6 - Asp 7; β B, Lys 16 - Phe21; β C, Ser 31 - Gln 36; β D, Arg 42 - Ile 43), and the other layer comprises α -helix (α A, Lys 46-Leu 53) with a characteristic hairpin loop at the C-terminus. Both helix α A and the β -sheet have an amphipathic character, and their hydrophobic faces are tightly packed against each other, so that helix α A and strand β D are arranged in a roughly antiparallel manner. The overall fold of DNA-bound MBD is similar to that of the unliganded MBD with the exception of a loop L2 that emanates between β B and β C.

The MBD contains four characteristic long loops, L1, L2, L3 and the C-terminal hairpin loop (Fig. 4.2). Loop L1 (Cys 8 - Trp 15), which connects β A and β B, is largely bent and projects beyond the strand β B and β C. Of this loop, Leu 11 comprises the hydrophobic core and forms many interactions with the rest of the protein. Loop L3 (Gly 54 - Leu 61), that connects helix α A and the hairpin loop, is well defined. It starts at a characteristic short turn at Gly 54 - Pro 55, followed by a helical coil arranged antiparallel to α 1. Loop L2 (Arg 22 - Arg30) is also well defined in complexed with DNA. However, in contrast to L1 and L3, loop L2 is highly flexible in the unliganded protein judging from the small ^{15}N heteronuclear NOE value. As described below, this L2 flexibility in unliganded form plays an important role for DNA binding. The amino acid sequences of L2 are highly conserved, which suggests an existence of structural and dynamical similarity in L2 region throughout the MBD proteins. The C-terminal hairpin loop (Phe 62 - Ile 68) resembles a β -hairpin, with a mainchain hydrogen bond between Gly 67 HN and Asp 63 C'=O in its stem region. This loop has structurally vital hydrophobic residues, Phe 62 and Phe 64, which stabilize orientation of the hairpin relative to the rest of the protein, and are identical in all MBD family members (Fig. 4.3C). Substitution of Phe 64 by alanine caused disruption of the

native tertiary structure and a consequent loss of the DNA binding (see "Materials and methods"). Deletion of residues 157-162 from MeCP2, which corresponds to the most of the hairpin loop of MeCP2, resulted in a total loss of methyl-CpG binding activity, further indicating the structural significance of the hairpin loop (Nan et al., 1993). With knowledge of the tertiary structure of the MBD1 and structurally vital residues, we could align sequences of the MBD family, as shown in Figure 4.3, finding the insertion of four residues at the junction between L3 and the hairpin loop in both MeCP2 and MBD4. Considering the insertion position in the tertiary structure, these residues would be expected not to have a large impact on the overall folding of these proteins.

This MBD fold generates a well defined hydrophobic core that includes the entire polypeptide domain, and is formed by Leu 11, Trp 15, Thr 33, Tyr 35, Ile 43, Leu 49, Tyr 52, Leu 53, Leu 59, Phe 62 and Phe 64 (Figs. 4.3A and B). The high degree of conservation of these residues in mammalian and *X. laevis* MBDs, which are responsible for structural integrity of the domain, suggests that the folding is essentially identical throughout the MBD family (Fig. 4.3C). The geometry of the α -helix relative to the β -sheet was determined by the hydrophobic interaction of Leu 49 and Leu 53 in the α -helix and Tyr 35 and Ile 43 in the β -sheet. Extensive hydrophobic contacts of residues in loops, Leu 11 and Trp 15 in loop L1, Leu 59 in loop L3, and Phe 62 and Phe 64 in hairpin-loop, further stabilized this interaction. Side-chains of most of the conserved hydrophobic residues point to the interior of the protein. Exceptions are Tyr 34 and Tyr 52. These residues are highly conserved and their aromatic rings are largely exposed to the solvent (Fig. 4.3A).

In addition to hydrophobic residues that form the core, several basic residues are well conserved in the mammalian and *X. laevis* MBD family. Side-chains of these basic residues are exposed to the solvent, are largely confined to one side of the molecule and forms a binding surface to DNA (Figs. 4.4 B and C). Well conserved Arg 30, Arg 42, Arg 44, Lys 46 and Lys 65 are lined up in the middle of this basic surface, and Arg 22 and Lys 23 are clustered in loop L2. Of these, only Arg 17, which highly conserved in MBD family, is located on the other side, however it supports one side of protein hydrophobic core by long aliphatic polypeptide of the

side-chain. In addition to these basic ones, three conserved residues, Asp 32, Tyr 34 and Ser 45, are located in the middle of this surface and these side-chains are exposed to DNA bases (Figs. 4.4 and 4.6).

Structure of MBD-DNA complex

The final 20 structures of the MBD-DNA complex are shown superimposed in stereo in Figure 4.5D. The location of MBD on the DNA has been well defined supported by a large number of intermolecular NOEs. Contacts to the DNA are made through the face of the β -sheet, which is positioned within the major groove. The beta strands β B and β C are across the major groove, and two different structural components at the both end of β -strands, loop L2 and helix α A stabilize the MBD-DNA interaction from each side of DNA. The MBD contacts to the six base pairs and the specific base contacts are made only in methyl-CpG sequence.

One characteristic feature of the MBD-DNA interaction is its small interfacial contact area. Calculation of the change in solvent-accessible surface area of MBD upon DNA binding gives a value of only 810 \AA^2 , which is much smaller than the average value of 1340 \AA^2 for 9 monomeric proteins (Jones et al., 1999). This small value reflects the fact that MBD contacts DNA exclusively within a narrow area of the major groove of the duplex: no part of the protein invades the minor groove side (Fig. 4.6). This feature is probably necessary so that MBD can access exposed CpG sites on nucleosome cores without steric clash (Luger et al., 1997). Indeed, the MBD of MeCP2 can bind methyl-CpG sites exposed in the major groove of nucleosomal DNA without major impedance (Chandler et al., 1999). Probably owing to this small contact surface, the DNA in the complex is, in turn, unbent and otherwise displays typical features of a B form duplex, except for a small widening of the major groove at the MBD contact site.

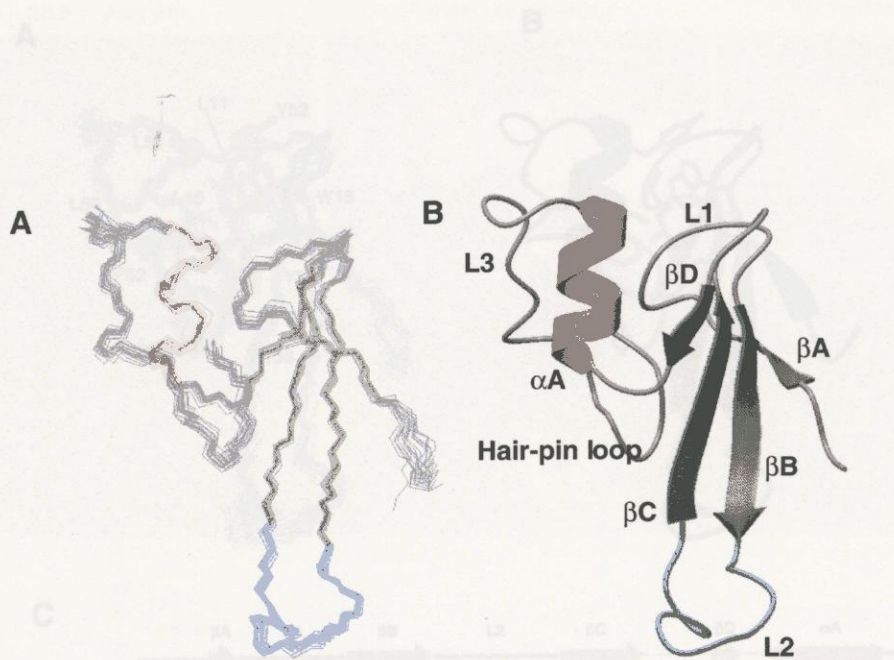


Figure 4.2 Solution structure of the MBD1 MBD in complex with DNA.

(A) Best-fit superposition of the backbone (N,C α , and C') atoms of the 20 final structures. The disordered two N-terminal and six C-terminal residues are omitted from figures for clarity. The residues in β -strand and α -helix are colored in green and red, respectively. Loop L2 which is important for DNA binding is colored in blue. (B) Schematic ribbon drawing of the energy minimized average structure. Secondary structure elements are indicated.

Figure 4.3 The hydrophobic core residues of the MBD.

The best-fit superposition (A) and schematic ribbon drawing (B) of the side-chain heavy atoms (red) of the residues which formed hydrophobic core in MBD1 MBD. The backbone atoms are also superimposed in blue. (C) The sequence alignment of MBDs. The residues which formed hydrophobic core in MBD1 are colored in yellow. These residues are highly conserved in the other MBDs, which suggests that the folding is essentially identical throughout the MBD family. The secondary structure of MBD1 is indicated at the top.

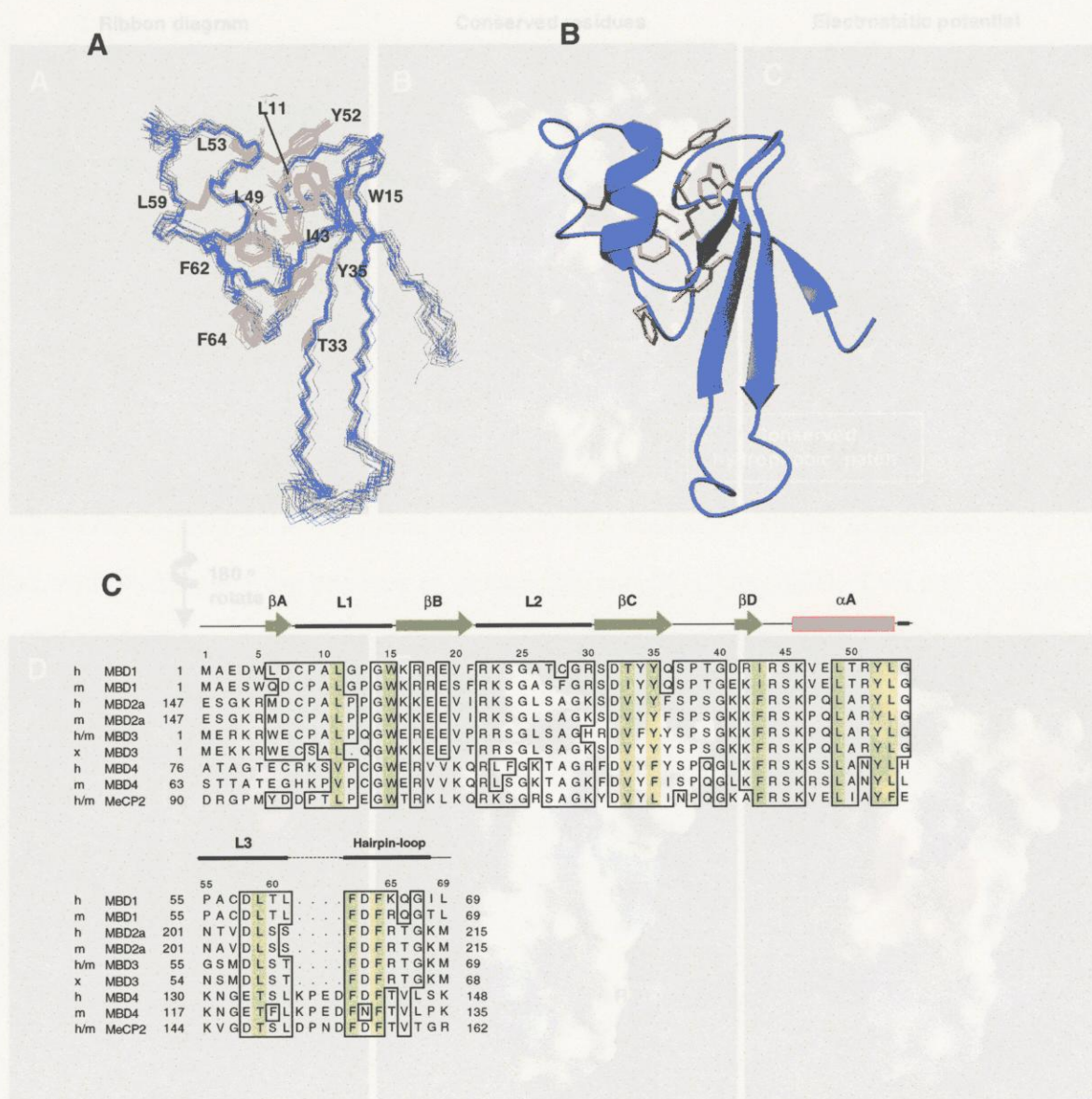


Figure 4.3 The hydrophobic core residues of the MBD.

The best-fit superposition (A) and schematic ribbon drawing (B) of the side-chain heavy atoms (red) of the residues which formed hydrophobic core in MBD1 MBD. The backbone atoms are also superimposed in blue. (C) The sequence alignment of MBDs. The residues which formed hydrophobic core in MBD1 are colored in yellow. These residues are highly conserved in the other MBDs, which suggests that the folding is essentially identical throughout the MBD family. The secondary structure of MBD1 is indicated at the top.

4.5 Molecular basis of methylated DNA recognition

Overview of recognition

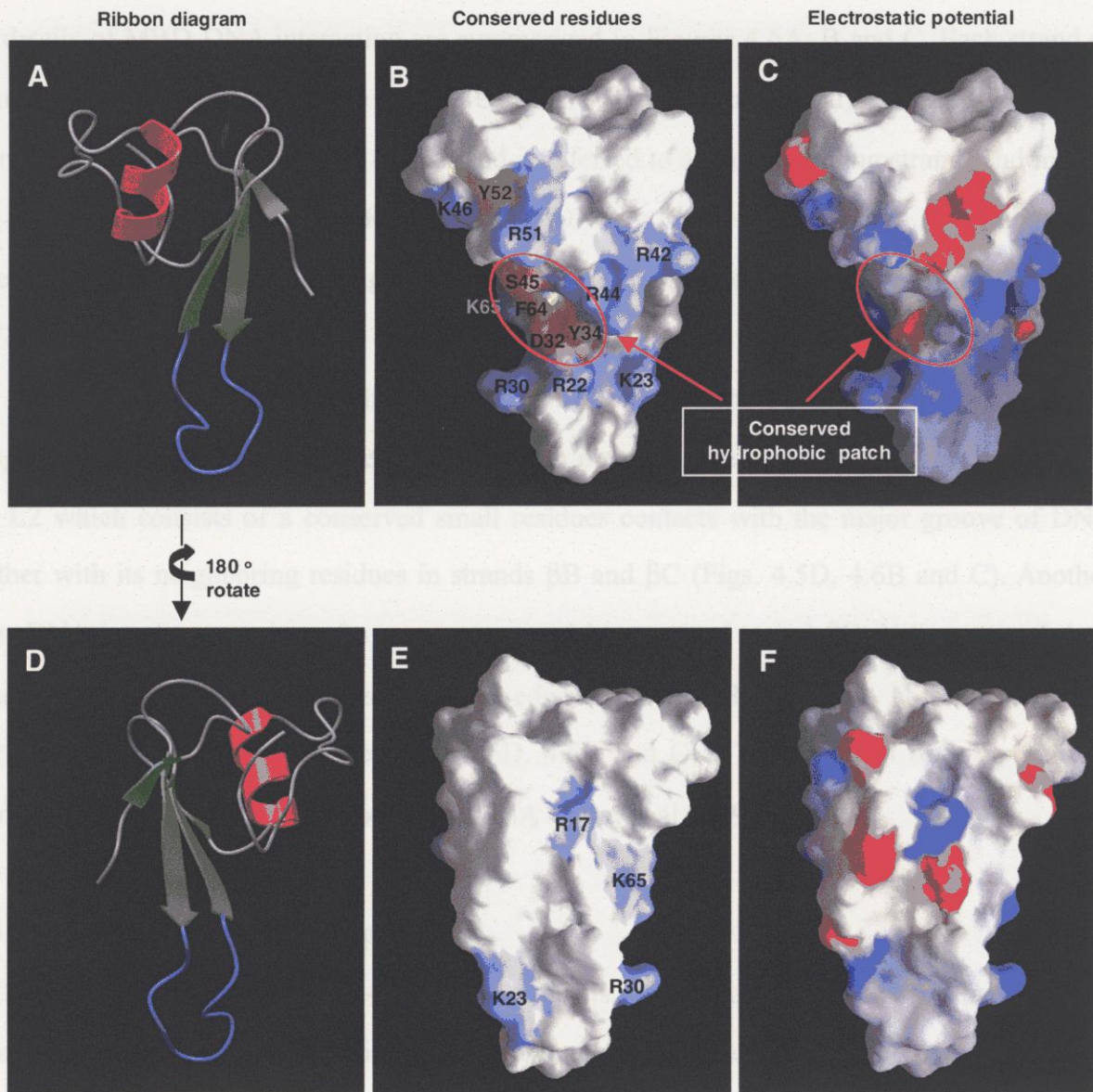


Figure 4.4 Surface diagrams of MBD.

Figures B and C are the surface diagrams of MBD, viewed in the same orientation as in Figure A. Figures D, E and F are viewed in the opposite direction as A. The conserved basic, acidic, and hydrophobic residues are colored in blue, pink, and yellow, respectively (Figs. B and E). The hydrophobic superficial patches made up of the conserved basic residues (Fig. B). Distribution of the electrostatic potential (± 10 kT) on the solvent accessible surface displayed with GRASP (Figs. C and F). Blue corresponds to positive potential and red to negative potential.

4.5 Molecular basis of methylated DNA recognition

Overview of recognition

The details of MBD-DNA interaction are summarized in Figures 4.6A, B and C. Each strand of 12-mer double-stranded DNA are numbered to G1-C12 (coding strand) and G13-C24 (non-coding strand). Methyl cytosine in each strand is referred to as m⁵C6 (coding strand) and m⁵C18 (non-coding strand), respectively. Recognition of DNA by MBD results from a large number of the electrostatic, hydrogen bonding, and hydrophobic interactions, which are formed between MBD and DNA.

Major binding interface in the protein consists of two parts, which each part recognizes phosphate backbone and bases of each strand from opposite side of DNA. One major interface, loop L2 which consists of a conserved small residues contacts with the major groove of DNA together with its neighboring residues in strands β B and β C (Figs. 4.5D, 4.6B and C). Another major DNA interface consists of a segment comprising parts of strand β D, helix α A and their connecting loop. These interactions anchor an edge of the beta-sheet and the N-terminal part of helix α A within the major groove (Figs. 4.5D, 4.6B and C). The alignment of α A permits a favorable interaction of its helix dipole with DNA (Zhou et al., 1998).

Base recognition at the methyl-CpG site

MBD binds specifically to methyl-CpG DNA, and its recognition is achieved by hydrophobic interactions between the MBD residues and two methyl groups of two methyl-cytosine (m⁵C6, m⁵C18) at the methyl-CpG site. This two methyl groups are distinguished by major groove contacts made by five residues, Val 20, Arg 22, Tyr 34, Arg 44 and Ser 45 (Fig. 4.6B). These residues form a continuous hydrophobic patch that mediates the interfacial contacts, and are invariant in all functional MBD family members, except for Val 20 (Figs. 4.4B and C). The side chains of Val 20, Arg 22 and Tyr 34 create a hydrophobic pocket, which packs against the methyl group of m⁵C6, while the aliphatic portions of Arg 44 and Ser 45 contacts m⁵C18. Mutation of Ser 45 to alanine causes a moderate loss of DNA binding affinity (Fig. 4.7A). Fujita et al.

reported that localization patterns of MBD1 protein carrying a single mutation of Ser45 to alanine in HeLa nucleus were similar to those of wild-type protein (Fujita et al., 2000). These suggest that the side-chain of Ser 45 mainly contributes the hydrophobic interaction rather than the electrostatic. The interactions between the five residues of MBD and the methyl groups of the methyl-cytosine are evidenced by 20 intermolecular NOE observations.

In addition to the two methyl groups of methyl-cytosine bases, two guanine bases at the methyl-CpG site are recognized by two conserved arginine residues through electrostatic interactions. The side chain guanidiniums of Arg 22 and Arg 44 are in positions to donate hydrogen bonds to G7 and G19, respectively (Fig. 4.6). Arg 22 also donates a hydrogen bond to the phosphate located between C5 and m⁵C6 through its ε-NH. These contacts are significant to recognition, as substitution of Arg 22 or Arg 44 with either alanine or lysine abolishes DNA binding (Fig. 4.7). Specificity for the m⁵C:G base pair is further conferred by Tyr 34; its side chain hydroxyl group is poised to accept a hydrogen bond from the 4-amino group of m⁵C6 (Fig. 4.6). The existence of this bond is supported by mutagenesis of Tyr 34 to phenylalanine or alanine, either of which markedly decreases the affinity for methylated DNA, suggesting that the Tyr 34 hydroxyl group participates in DNA binding (Figs. 4.6 and 4.7A). In mammalian MBD3, this tyrosine substitution at position 34 may cause a loss of the binding to methyl-CpG (Fig. 4.5A) (Hendrich and Bird, 1998). Taken together, recognition of the methyl-CpG site appears to be the result exclusively of hydrophobic and polar contacts mediated mainly by the five residues. Base contacts by MBD are limited to the m⁵CG sequence within the major groove.

Side chain-side chain interactions involving these five residues are crucial for proper generation of the interfacial contact surface; in particular, those involving Asp 32 and Tyr 34 are of special importance. The carboxylate of conserved Asp 32 is in a position to form a hydrogen bond salt bridge interaction with the guanidinium of Arg 22, facilitating contacts of Arg 22 with G7 and m⁵C6 (Fig. 4.6B). Although Asp 32 appears not to mediate a direct contact with DNA, its substitution to alanine significantly impairs DNA binding, suggesting the existence and importance of this Asp 32-Arg 22 interaction (Figs. 4.6B and 4.7A). This Arg-Asp interaction in

the guanine recognition is also observed in Zif268 Zinc fingers (Reber et al., 1994). The aromatic ring of Tyr 34 on one face mediates an interfacial contact with m⁵C6 and on the other side buttresses the side chain of Arg 44 through hydrophobic interactions, supporting the Arg 44 side chain contact with G19 and m⁵C18.

DNA backbone contacts

DNA sugar-phosphate backbone contacts are mediated mainly by residues in L2 and β D- α A loop, and helix α A. Loop L2, which becomes ordered upon DNA binding, spreads out from the body of the protein, its tip latching onto the DNA backbone of one strand (Figs. 4.5D and 4.6). Invariant Gly 25 forms a characteristic sharp turn at the extreme of the loop, and the main chain amide groups of its neighboring residues, Lys 23, Ser 24, Ala 26 and Thr 27, are favorably positioned to donate hydrogen bonds to the two phosphates located between C5 and G7 (Fig. 4.6C). Furthermore, the hydrogen-bond between Arg 30 HN and Thr 27 C'=O stabilizes the loop structure. Some of these residues on the loop concurrently make hydrophobic contacts with the backbone sugars.

The side-chains of Arg 18, Lys 23, Arg 30, Arg 42, Lys 46, Arg 51 and Lys 65 are favorably positioned to donate hydrogen bonds to DNA phosphates (Fig. 4.6). Alanine substitutions of Arg 30, Lys 46, or Lys 65 indicate that their importance for DNA binding, each of which significantly decreased DNA binding (Fig. 4.7A). On the other hand, substitution of Arg 30 to lysine has little effect on DNA binding, probably because the positively charged side chain of lysine can mediate a similar contact with the DNA backbone. In contrast, the lysine substitution of either Arg 22 or Arg 44 totally abolishes DNA binding, because guanidiniums of these residues are crucial for base discrimination (Fig. 4.7A). The backbone interaction is further stabilized by two α A residues, Val 47 and Glu 48, that make hydrophobic contacts with sugar rings of C17 and m⁵C18, respectively. The strong conservation of residues that are important for DNA binding and structural integrity of the domain suggests that the folding and mode of DNA interaction is similar throughout the MBD family.

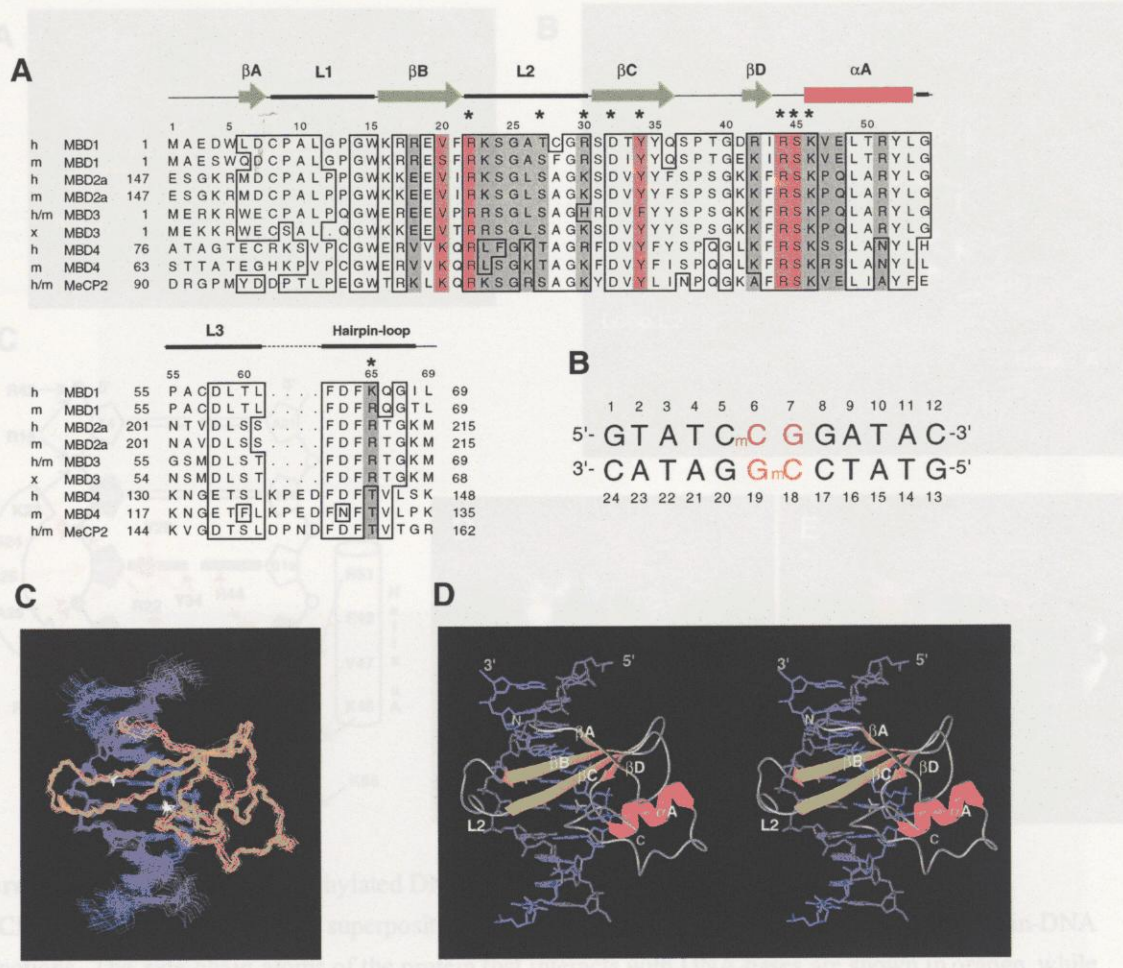


Figure 4.5 Solution structure of MBD1 MBD-DNA complex.

(A) Sequence alignment of the MBD family. Conserved residues are boxed. Red denotes MBD1 residues interacting with DNA bases and gray identifies residues interacting with DNA sugar-phosphate backbone. The secondary structure of MBD1 in complex with DNA is indicated at the top. Residues that have been shown to be important for DNA binding through mutational analysis are marked with asterisk above the sequences. (B) Sequence of oligonucleotide used in this study. "mC" represents 5-methyl cytosine. The methyl-CpG sequence is highlighted in red. (C) Superposition of 20 NMR structures of MBD-DNA complex. The protein backbone atoms (residues 3-69) and DNA heavy atoms (base pairs 2-11) are shown. The two methyl groups at the methyl-CpG site are colored in white. (D) Schematic ribbon diagrams of the energy minimized average structure in stereo. Secondary structure elements of the protein are shown.

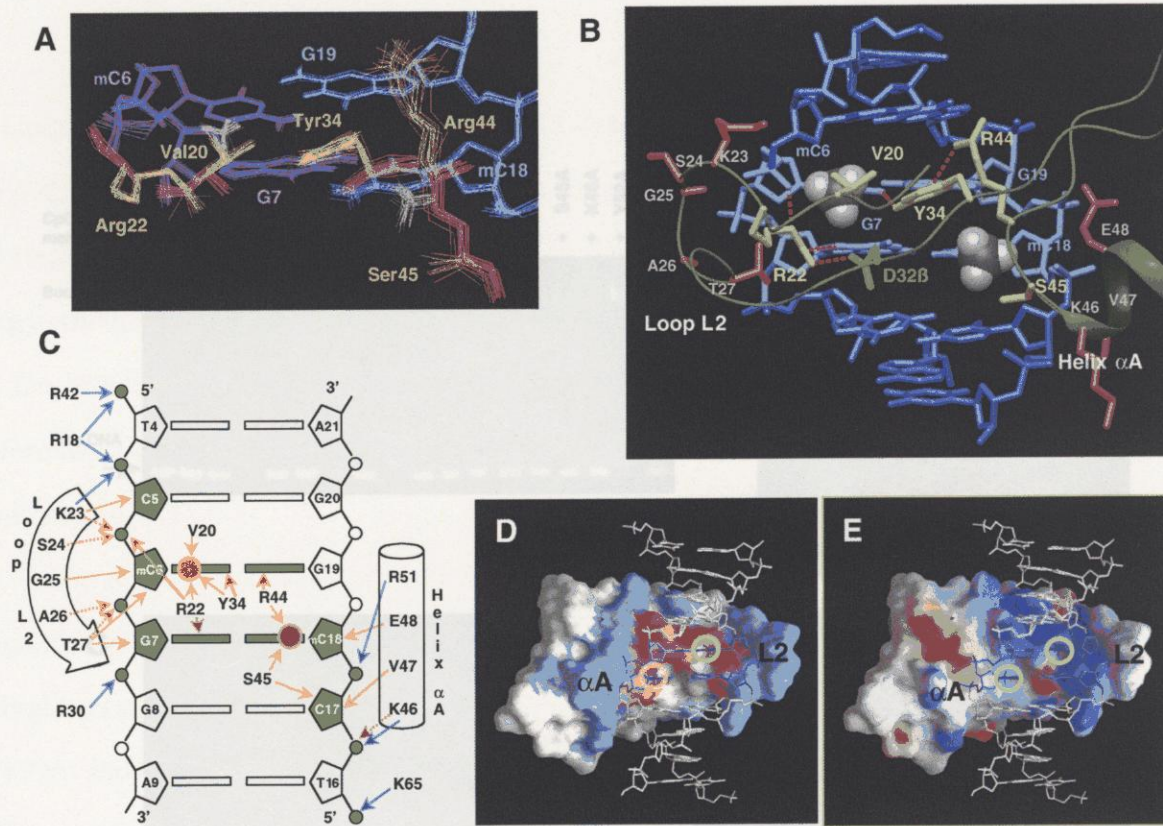


Figure 4.6 Molecular basis of methylated DNA recognition.

(A) Close-up view of the best fit superposition of the 20 final structures at the position of protein-DNA interactions. The side-chain atoms of the protein that interacts with DNA bases are shown in orange, while protein backbone atoms are in red. (B) A detail view of the protein-DNA contact site. Protein side-chains interacting with DNA bases and backbone are shown in yellow and red, respectively. The side-chain of D32 are colored in green. Red dashed lines are indicates proposed inter-molecular hydrogen bonds, which are observed in the most members of NMR structures. R22-D32 hydrogen bond is also displayed. (C) Schematic summary of protein-DNA interactions. DNA bases, sugar rings, and phosphates are indicated as boxes, pentagons, and small circles, respectively. Two methyl groups at the methyl-CpG site are represented as red filled circles. Arrows are colored differently depending on the interaction mode of the protein-DNA: hydrophobic (yellow), electrostatic (blue), and hydrogen bonds (red). Red dashed arrows indicate the hydrogen bond interactions involving backbone amide groups of the protein. DNA bases, sugar rings, and phosphates which contact with protein are colored green. (D) Protein surface diagrams at the DNA contact site. The protein surfaces which interact with DNA bases and backbones are shown in red and blue, respectively. Two methyl groups are marked with yellow circles. (E) Electrostatic surfaces of the protein at the DNA binding site. Two methyl groups are also marked with yellow circles. The basic residues are positioned in one side (side of loop L2) of the protein surface.

5 DISCUSSION

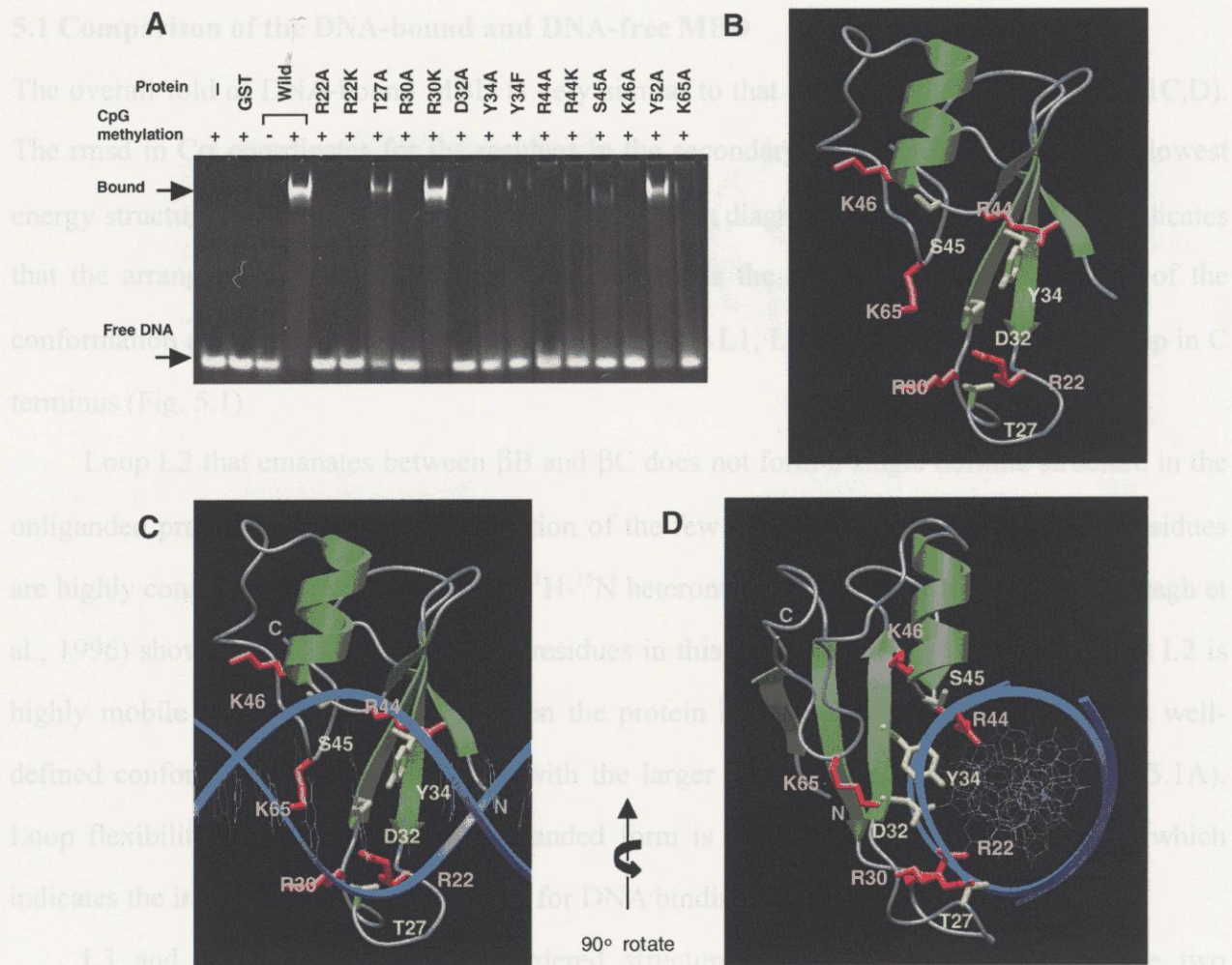


Figure 4.7 Mutational analysis

(A) Gel-shift assay was performed to obtain the effects of the substitutions to DNA binding activity, which could not be obtained from structural data directly. The wild-type and mutated residues are indicated above. (B) Only the residues which substitutions cause a loss of DNA binding activity are indicated on the MBD structure. Basic residues are colored in red, while others are in yellow. (C) The structure of the MBD-DNA complex, viewed in the same orientation as in Figure B. The residues which recognize DNA base and sugar-phosphate backbone are colored in red and yellow, respectively. (D) Side view of the Figure C.

5 DISCUSSION

5.1 Comparison of the DNA-bound and DNA-free MBD

The overall fold of DNA-bound MBD is very similar to that of unliganded MBD (Fig. 5.1C,D). The rmsd in C α coordinates for the residues in the secondary structure between the two lowest energy structures is 1.2 Å. A superposition of the ribbon diagrams of the two molecules indicates that the arrangement of the β -sheet and the α -helix is the same. A significant change of the conformation are observed in several loop regions; loop L1, L2, and L3 and the hairpin loop in C terminus (Fig. 5.1).

Loop L2 that emanates between β B and β C does not form a single definite structure in the unliganded protein judging from observation of the few long-range NOEs, although its residues are highly conserved in the MBD family. ^1H - ^{15}N heteronuclear NOE measurements (Cavanagh et al., 1996) showed small NOE values for residues in this loop (0.34 ± 0.17) indicating that L2 is highly mobile in solution. However, when the protein binds to DNA, this loop adopts a well-defined conformation, which supported with the larger NOE values (0.70 ± 0.05) (Fig. 5.1A). Loop flexibility at this position in unliganded form is also observed in MeCP2 MBD, which indicates the importance of this flexibility for DNA binding (Wakefield et al., 1999).

L3 and the hairpin loop have ordered structure and same conformation in the two molecules. However, the position of these loops in MBD structure is different by about 2.5 Å. Fitting the coordinate of the unliganded protein to those of the protein-DNA complex, the hairpin loop of unliganded protein causes a steric clash with major groove backbone atoms. The small values of ^1H - ^{15}N heteronuclear NOE of Ala 56 - Asp 58 in L3 (smaller than 0.6) suggests that the loop flexibility reduces steric crash with DNA. The position of the loop L1 that connects β A and β B is also changed upon DNA binding. As L1 interacts directly with the hairpin loop and is in opposite side of DNA binding surface, this change appears to be caused by the movement of the hairpin loop upon DNA binding.

5.2 Comparison of MBD1 MBD and MeCP2 MBD

Recently, the structure of the MBD of another MBD family protein MeCP2 was reported (Wakefield et al., 1999). Although the sequence homology of MBDs is not so high (39%) between the MBD1 and MeCP2 proteins, the tertiary structures of the two MBDs are very similar

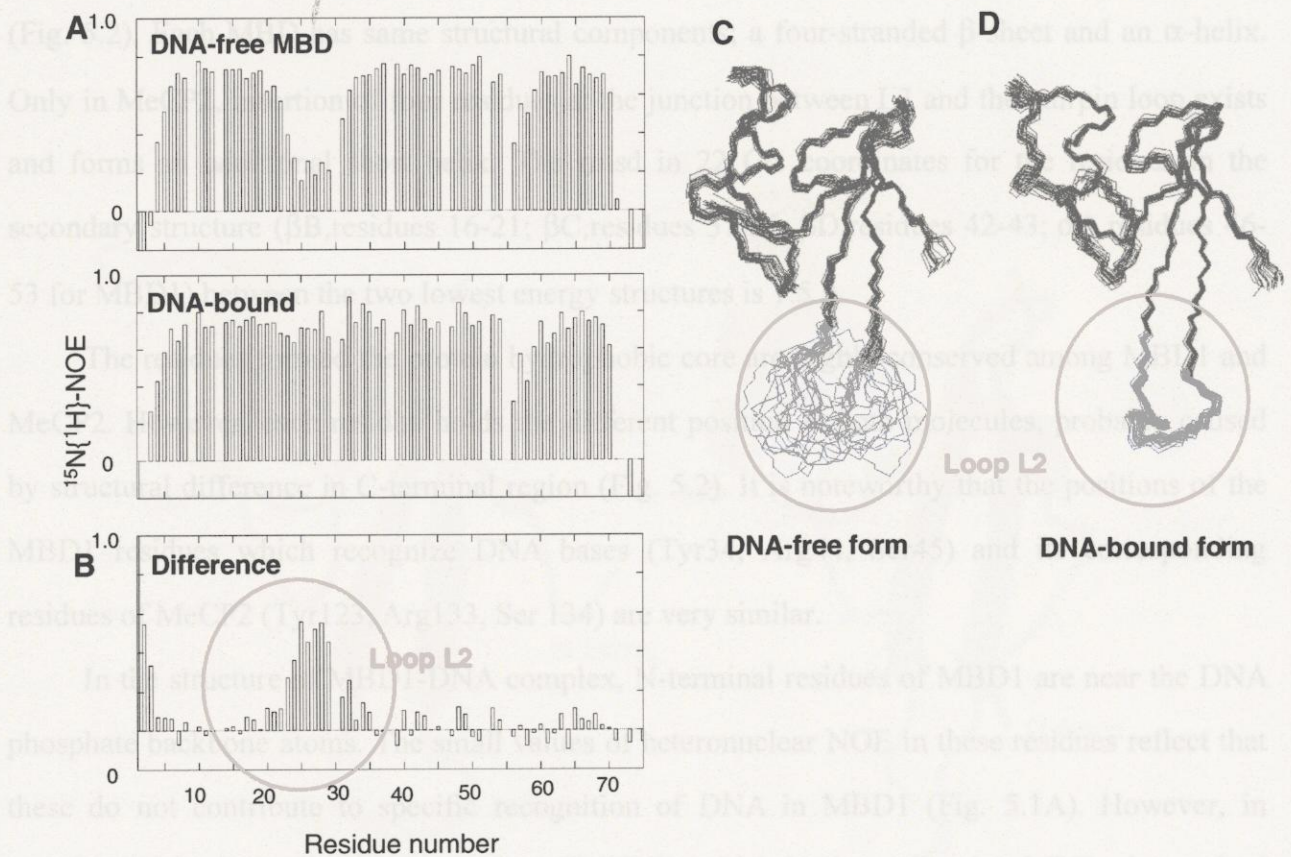


Figure 5.1 $^{15}\text{N}\text{-}^1\text{H}$ heteronuclear NOE of the DNA-free and DNA-bound MBD.

(A). Plots of the values of $^{15}\text{N}\text{-}^1\text{H}$ heteronuclear NOE of the DNA-free (above) and DNA-bound (below) MBD against the residue number. Figure (B) shows a difference of these two NOE values. Largest difference is observed in loop L2. (C) and (D). The backbone superposition of 20 final structures of MBD in DNA-free (C) and DNA-bound forms (D) are shown. Loop L2 is indicated in blue. The L2 becomes ordered in the presence of DNA.

5.2 Comparison of MBD1 MBD and MeCP2 MBD

Recently, the structure of the MBD of another MBD family protein MeCP2 was reported (Wakefield et al., 1999). Although the sequence homology of MBDs is not so high (39%) between the MBD1 and MeCP2 proteins, the tertiary structures of the two MBDs are very similar (Fig. 5.2). Each MBD has same structural components; a four-stranded β -sheet and an α -helix. Only in MeCP2, insertion of four residues at the junction between L3 and the hairpin loop exists and forms an additional short helix. The rmsd in 22 C α coordinates for the residues in the secondary structure (β B, residues 16-21; β C, residues 31-36; β D, residues 42-43; α A, residues 46-53 for MBD1) between the two lowest energy structures is 1.5 Å.

The residues formed the protein hydrophobic core are highly conserved among MBD1 and MeCP2. However, each residue holds the different position in both molecules, probably caused by structural difference in C-terminal region (Fig. 5.2). It is noteworthy that the positions of the MBD1 residues which recognize DNA bases (Tyr34, Arg44, Ser45) and its corresponding residues of MeCP2 (Tyr123, Arg133, Ser 134) are very similar.

In the structure of MBD1-DNA complex, N-terminal residues of MBD1 are near the DNA phosphate backbone atoms. The small values of heteronuclear NOE in these residues reflect that these do not contribute to specific recognition of DNA in MBD1 (Fig. 5.1A). However, in MeCP2, N-terminal polypeptide segment (IRDRG), which is absent in other MBDs, is required for DNA binding (Wakefield et al., 1999; Nan et al., 1993). Variety of N-terminal sequence among MBD family probably gives an additional function to MBD.

5.3 Rett mutations found in MeCP2

The structure of the MBD-DNA complex thus allows us to interpret rapidly accumulating mutational data for MeCP2 in patients suffering from Rett syndrome, a leading cause of mental retardation and autistic behavior in females. In total, substitutions of 15 different residues have been reported in MeCP2, 10 of which are in the MBD, as shown in Figure 5.3B (Amar et al., 1999; Veyver and Zoghbi, 2000).

Of the residues in the MBD, Arg 133 and Ser 134 of MeCP2 are equivalent to Arg 44 and Ser 45 of MBD1. Interestingly, these are two of the most conserved residues that form the methyl-binding pocket in the MBD1-DNA complex (Fig. 4.6B). In MeCP2, it is rather straightforward to identify the location of the Rett mutations. The largest number of contacts with DNA is made by the MBD1 MBD. The MBD1 MBD is a β -sheet structure with a β -strand (green) and an α -helix (red). The overall topology is essentially the same. Additional α -helix (α B) is observed in C-terminal of MeCP2.

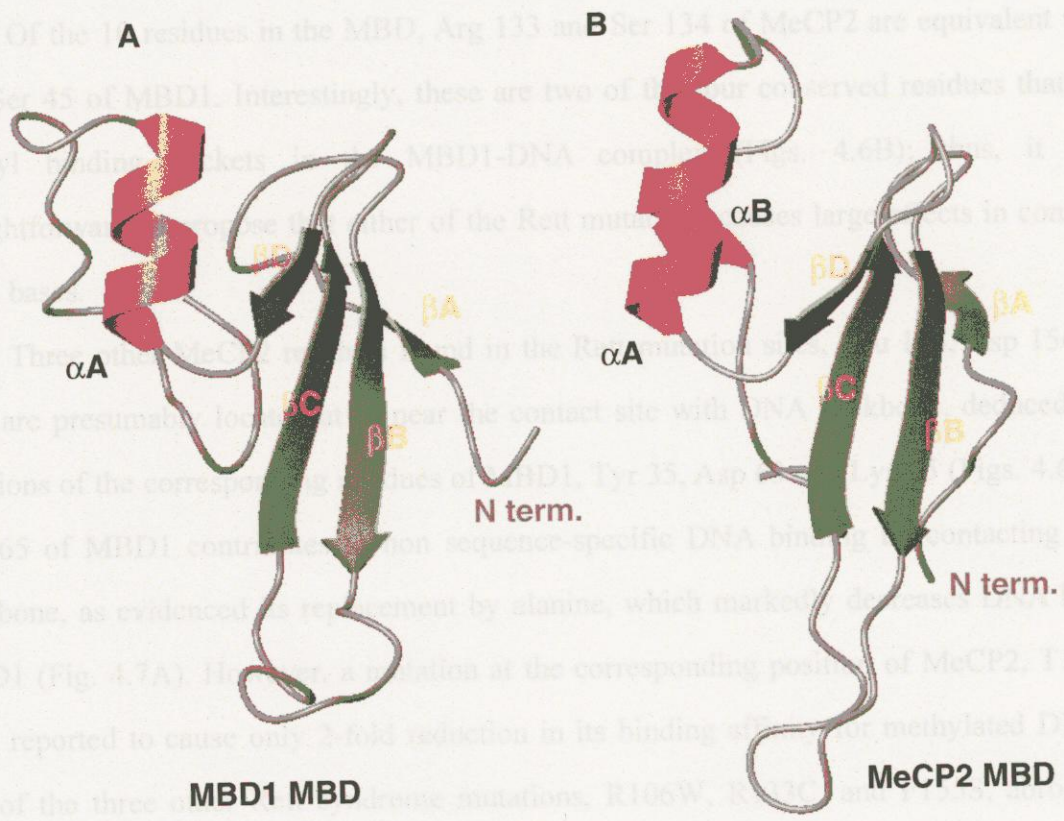


Figure 5.2 Comparison of the overall topologies of MBD1 MBD and MeCP2 MBD.

Ribbon diagrams of MBD1 MBD (A) and MeCP2 MBD (B) (Wajefield et al., 1999) are shown with the secondary structural elements. The secondary structures in both structures are indicated in green (β -strand) and red (α -helix). The overall topology is essentially the same. Additional α -helix (α B) is observed in C-terminal of MeCP2.

...the MBD family, but are located remote from the DNA interface. Rett mutations affecting these MeCP2 residues are likely to have global structural consequences, because they are presumably important for proper folding or maintenance of the protein's

5.3 Rett mutations found in MeCP2

The structure of the MBD-DNA complex thus allows us to interpret rapidly accumulating mutational data for MeCP2 in-patients suffering from Rett syndrome, a leading cause of mental retardation and autistic behavior in females. In total, substitutions of 15 different residues have been reported in MeCP2, 10 of which are in the MBD, as shown in Figure 5.3B (Amir et al., 1999; Veyver and Zoghbi, 2000).

Of the 10 residues in the MBD, Arg 133 and Ser 134 of MeCP2 are equivalent to Arg 44 and Ser 45 of MBD1. Interestingly, these are two of the four conserved residues that form the methyl binding pockets in the MBD1-DNA complex (Figs. 4.6B); thus, it is rather straightforward to propose that either of the Rett mutations causes large effects in contacts with DNA bases.

Three other MeCP2 residues found in the Rett mutation sites, Leu 124, Asp 156 and Thr 158, are presumably located at or near the contact site with DNA backbone, deduced from the positions of the corresponding residues of MBD1, Tyr 35, Asp 63 and Lys 65 (Figs. 4.6 and 5.4). Lys 65 of MBD1 contributes to non sequence-specific DNA binding by contacting the DNA backbone, as evidenced its replacement by alanine, which markedly decreases DNA binding of MBD1 (Fig. 4.7A). However, a mutation at the corresponding position of MeCP2, T158M, has been reported to cause only 2-fold reduction in its binding affinity for methylated DNA, while any of the three other Rett syndrome mutations, R106W, R133C, and F155S, abrogate DNA binding significantly (Ballestar et al., 2000). Only MeCP2 and MBD4 possess a threonine at position 65 in the MBD family (Fig. 4.5A); thus, this residue might be involved in other roles related to the specific function of MeCP2 and MBD4, as suggested by Ballestar et al.. In contrast, Gly 12 (Pro 101 in MeCP2), Arg 17 (Arg 106) and Phe 62 (Phe 155) of MBD1 are well conserved among the MBD family, but are located remote from the DNA interface. Rett mutations affecting these MeCP2 residues are likely to have global structural consequences, because they are presumably important for proper folding or maintenance of the protein's

structure in both MBD1 and MeCP2 (Ohki et al. 1999; Wakefield et al., 1999; Ballestar et al., 2000).

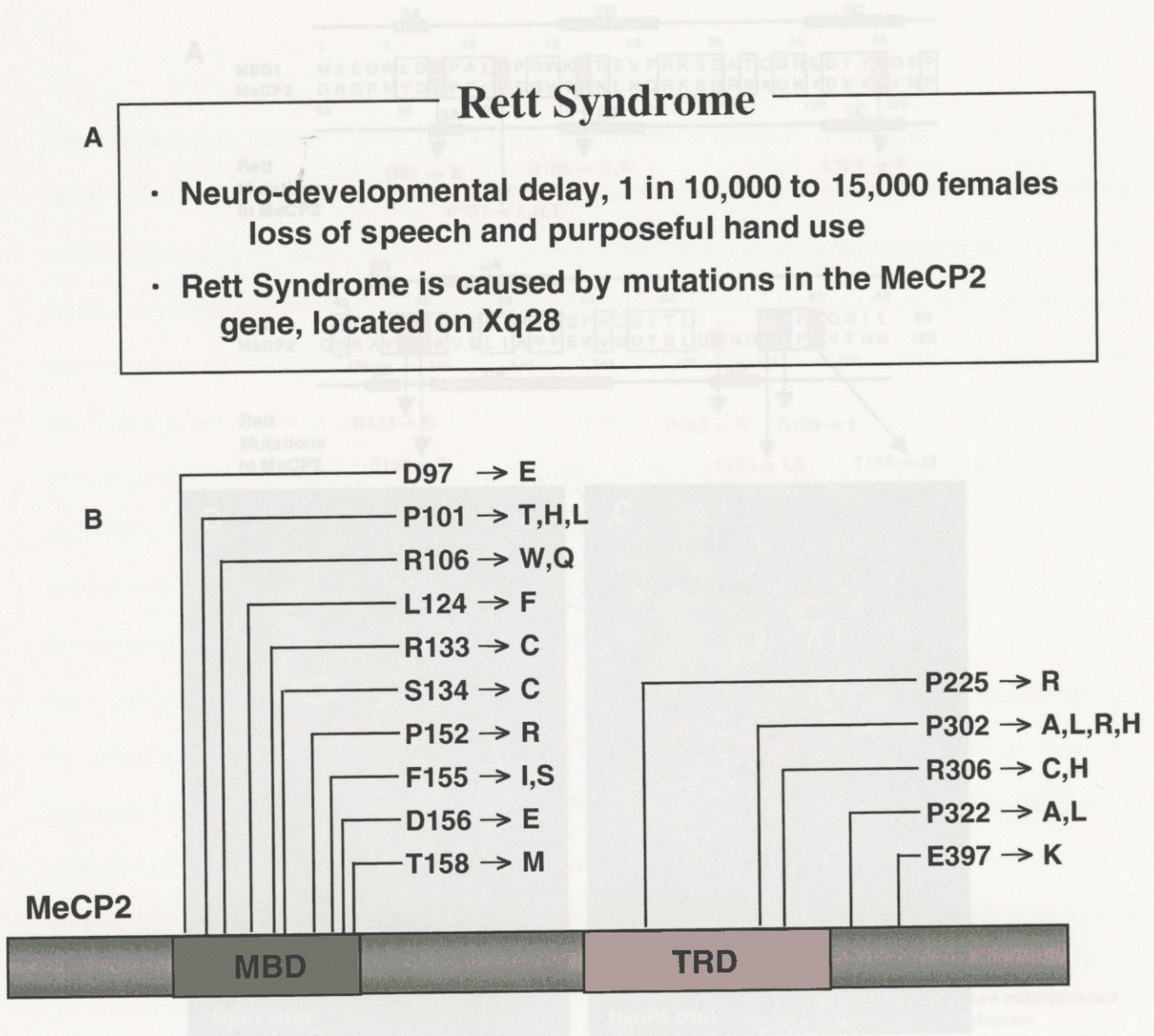


Figure 5.1 Positions of Rett syndrome mutations.

(A) Sequential alignment of MBD1 and MeCP2. Conserved residues are boxed and secondary structures are indicated. (B) MeCP2 structure with mutations indicated. The missense mutations of MeCP2 found in Rett syndrome are shown in orange. The missense mutations of MeCP2 found in Rett syndrome are shown in orange. (C) MeCP2 MBD structure with mutations indicated. The missense mutations of MeCP2 found in Rett syndrome are shown in orange. These substitutions may cause a loss of the function of MeCP2. Blue residues are slightly away from DNA interaction site in MBD1, while they are exposed to the solvent. T158 mutation of MeCP2 has been reported to cause only 2-fold reduction in DNA binding affinity (Ballenger et al., 2000), which suggests the other function of T158. The positions of these mutations are showed on the structures of MBD1 MBD (B) and MeCP2 MBD (C).

6. CONCLUSION

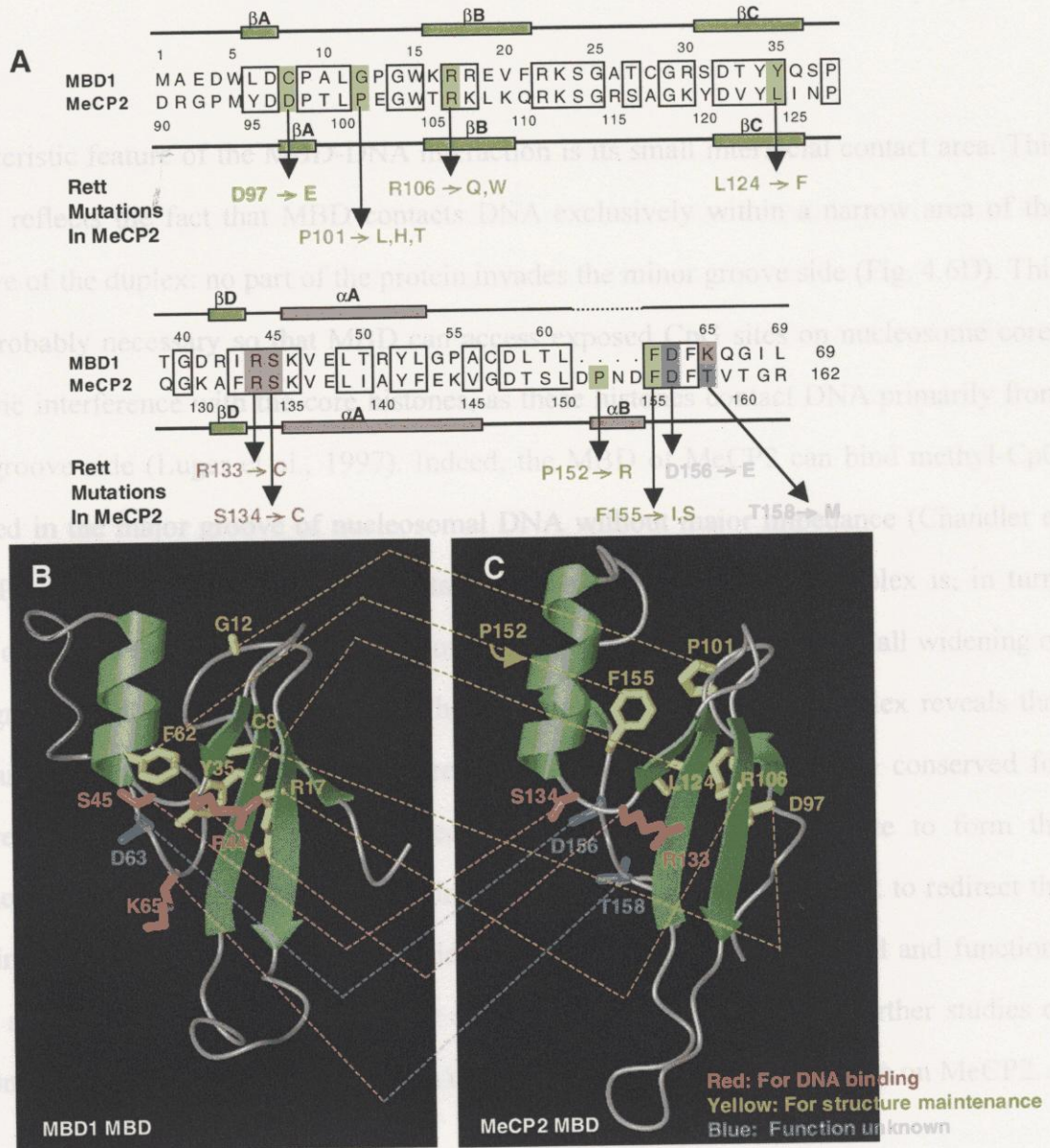


Figure 5.4 Positions of Rett syndrome mutations.

(A) Sequential alignment of MBD1 and MeCP2. Conserved residues are boxed and secondary structures are indicated at the top (MBD1) and bottom (MeCP2) of the sequence. The missense mutations of MeCP2 found in Rett syndrome are indicated below the sequences. Rett mutations in MeCP2 are colored in proposed function based on an analysis of the structure of MBD-DNA complex. The residues colored in orange are supposed to possess a function of structure maintenance and reds need for interaction with DNA. These substitutions may cause a loss of the function of MeCP2. Blue residues are slightly away from DNA interaction site in MBD1, while they are exposed to the solvent. T158 mutation of MeCP2 has been reported to cause only 2-fold reduction in DNA binding affinity (Ballestar et al., 2000), which suggests the other function of T158. The positions of these mutations are shown on the structures of MBD1 MBD (B) and MeCP2 MBD (C).

6 CONCLUSION

One characteristic feature of the MBD-DNA interaction is its small interfacial contact area. This small value reflects the fact that MBD contacts DNA exclusively within a narrow area of the major groove of the duplex: no part of the protein invades the minor groove side (Fig. 4.6D). This feature is probably necessary so that MBD can access exposed CpG sites on nucleosome cores without steric interference with the core histones, as these histones contact DNA primarily from the minor groove side (Luger et al., 1997). Indeed, the MBD of MeCP2 can bind methyl-CpG sites exposed in the major groove of nucleosomal DNA without major impedance (Chandler et al., 1999). Probably owing to this small contact surface, the DNA in the complex is, in turn, unbent and otherwise exhibits typical features of a B-form duplex, except for a small widening of the major groove at the MBD contact site. The structures of MBD-DNA complex reveals that many residues are conserved for functional reasons. Other residues clearly are conserved for structural reasons. The majority of conserved hydrophobic residues contribute to form the hydrophobic core, and several conserved glycine and proline residues are present to redirect the protein main chain. Conservation of these residues indicated that the overall fold and functions will be the same across the MBD. This study also provides a framework for further studies of Rett syndrome by facilitating the interpretation of the accumulating mutational data on MeCP2.

Coordinates.

The coordinates of DNA-free MBD1 MBD have been deposited in the Protein Data Bank (accession code 1D9N).

7 ACKNOWLEDGEMENTS

The present studies have been performed under the direction of Dr. Masahiro Shirakawa (Graduate school of Biological Science, Nara Institute of Science and Technology). I would like to express my gratitude to him for his guidance and discussions throughout my studies.

I wish to thank Dr. Mitsuyoshi Nakao and Dr. Naoyuki Fujita (Department of Tumor Genetics and Biology, Kumamoto University School of Medicine) for gifts of the genes of the methyl-CpG binding domain of MBD1, and for the precious discussion.

I wish to thank Dr. Haruki Nakamura (Research Center for Structural Biology, Institute of Protein Research) for advices in the structure calculation and Dr. Masato Katahira (Department of Chemistry and Biotechnology, Faculty of Engineering, Yokohama National University) for discussions about NMR technique in the analysis of nucleic acid. I also thanks to Dr. Markus Waelchli (Bruker Japan) for NMR measurements of the isotope-filtered spectra with Bruker spectrometer equipped with a newest cryo-probe.

I wish to thank Mr. Nobuya Shimotake, Mr. Jun-Goo Jee (Graduate school of Biological Science, Nara Institute of Science and Technology) and Dr. Takahisa Ikegami (Institute for Organic Chemistry, University Frankfurt) for their great help and discussions as collaborators.

Thanks to Dr. Toshio Hakoshima and all the members of his group (Graduate school of Biological Science, Nara Institute of Science and Technology), I could carry out my experiments comfortably. I would like to appreciate their kinness.

Finarlly, I would like to thank my familiy for their supports and kindness during my doctoral course.

Izuru Ohki

木 木 出

January 9, 2001.

8 REFERENCES

- Allen, M.D., Yamasaki, K., Ohme-Takagi, M., Tateno, M., and Suzuki, M. (1998). A novel mode of DNA recognition by a b-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *EMBO J.* 15, 5484-5496.
- Amir, R.E., Van den Veyver I.B., Wan M., Tran C.Q., Francke U. and Zoghbi H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl- CpG-binding protein 2. *Nat. Genet.* 23, 185-8.
- Ballestar, E., Yusufzai, T.M. & Wolffe, A.P. (2000). Effects of rett syndrome mutations of the methyl-CpG binding domain of the transcriptional repressor MeCP2 on selectivity for association with methylated DNA. *Biochemistry* 39, 7100-6.
- Ballobio, A., and Willard, H.F. (1992). Mammalian X-chromosome inactivation and the XIST gene. *Curr. Opin. Genet. Dev.* 2, 439-447.
- Baylin, S.B., Herman, J.G., Graff, J.R., Vertino, P.M., and Issa, J.P. (1998). Alterations in DNA methylation: a fundamental aspect of neoplasia. *Adv. Cancer. Res.* 72, 141-196.
- Bellacosa, A., Cicchillitti, L., Schepis, F., Riccio, A., Yeung, A.T., Matsumoto, Y., Golemis, E.A., Genuardi, M., and Neri, G. (1999). MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. *Proc. Natl. Acad. Sci. USA* 96, 3969-3974.
- Bergman, Y., and Mostoslavsky, R. (1998). DNA demethylation: turning genes on. *Biol. Chem.* 379, 401-407.
- Bhattacharya, S.K., Ramchandani, S., Cervoni, N., and Szyf, M. (1999). A mammalian protein with specific demethylase activity for mCpG DNA. *Nature* 397, 579-583.
- Bird, A. & Wolffe, A. P. (1999). Methylation-induced repression - belts, braces, and chromatin. *Cell* 99, 451-454.
- Bird, A. (1992). The essentials of DNA methylation. *Cell* 70, 5-8.
- Boyes, J., and Bird, A. (1991). DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* 64, 1123-1134.
- Boyes, J., and Bird, A. (1992). Repression of genes by DNA methylation depends upon CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. *EMBO J.* 11, 327-333.

- Bruenger, A.T. (1993). X-PLOR 3.1 : a system for X-ray crystallography and NMR. Yale Univ. Press, New Haven, USA.
- Cavanagh, J., Fairbrother, W.J., Palmer III, A.G., and Skelton, N.J. (1996). Protein NMR Spectroscopy. Academic Press, San Diego, USA.
- Chandler, S. P., Guschin, D., Landsberger, N. and Wolffe, A. P. (1999). The methyl-CpG binding transcriptional repressor MeCP2 stably associates with nucleosomal DNA. *Biochemistry* 38, 7008-7018.
- Chen, R.Z., Pettersson, U., Beard, C., Jackson-Grusby, L. and Jaenisch, R. (1998). DNA hypomethylation leads to elevated mutation rates. *Nature* 395, 89-93.
- Chuprina, V.P., Rullmann, J.A.C., Lamerichs, R.M.J.N., Boom, J.H.v., Boelens, R., and Kaptein, R. (1993). Structure of the complex of lac repressor headpiece and an 11 base-pair half-operator determined by nuclear magnetic resonance spectroscopy and restrained molecular dynamics. *J. Mol. Biol.* 234, 446-462.
- Clark, K.L., Halay, E.D., Lai, E., and Burley, S.K. (1993). Co-crystal structure of the HNF-3/fork head DNA recognition motif resembles histone H5. *Nature* 364, 412-420.
- Cornilescu, G., Delaglio, F. and Bax, A. (1999). Protein backbone angle restraints from searching a database for protein chemical shift and sequence homology. *J. Biomol. NMR* 13, 289-302.
- Costello, J. F. et al. (2000). Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat. Genet.* 24, 132-8.
- Cross, S.H., Meehan, R.R., Nan, S., and Bird, A. (1997). A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. *Nature Genet.* 16, 256-259.
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRpipe: a multi-dimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6, 277-293
- Dijkstra, K., Kroon, G.J.A., Vannuland, N.A.J., and Scheek, R.M. (1994). The COCAH Experiment to Correlate Intraresidue Carbonyl, C, and H Resonances in Proteins. *J. Magn. Reson. A* 107, 102-105.
- Eden, S., Hashimshony, T., Keshet, I., and Cedar, H. (1998). DNA methylation models histone acetylation. *Nature* 394, 842.
- Fearon, E.R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell* 61, 759-767.
- Feinberg, A.P., Rainier, S., and DeBaun, M.R. (1995). Genomic imprinting, DNA methylation, and cancer. *J. Natl. Cancer Inst. Monogr.* 17, 21-26.

- Foster, M.P., Wuttke, D.S., Clemens, K.R., Jahnke, W., Radhakrishnan, I., Tennant, L., Reymond, M., Chung, J., and Wright, P.E. (1998). Chemical shift as a probe of molecular interfaces: NMR studies of DNA binding by the three amino-terminal zinc finger domains from transcription factor IIIA. *J. Biomol. NMR* 12, 51-71.
- Free, A., Wakefield, R., Smith, O.B., Dryden, D., Barlow, P.N., and Bird, A. (2000). DNA Recognition by the methyl-CpG binding domain of MeCP2, *J. Biol. Chem.* in press, the manuscript is available at <http://www.jbc.org/>.
- Fujita, N., Takebayashi, S., Okumura, K., Kudo, S., Chiba, T., Saya, H., and Nakao, M. (1999). Methylation mediated transcriptional silencing in euchromatin by methyl-CpG binding protein MBD1 isoforms. *Mol. Cell. Biol.* 19, 6415-26.
- Fujita, N., Shimotake, N., Ohki, I., Chiba, T., Saya, H., Shirakawa, M. and Nakao, M. (2000). Mechanism of transcriptional regulation by methyl-CpG binding protein MBD1. *Mol. Cell. Biol.* 20, 5107-18.
- Garrett, D.S., Powers, R., Gronenborn, A.M., and Clore, G.M.. (1991). A common sense approach to peak picking in two-, three, and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson.* 95, 214-220
- Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Molecular and Cellular Biology* 18, 6538-6547.
- Hendrich, B., Hardeland, U., Ng, H., Jiricny, J. and Bird, A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* 401, 301-304.
- Hu, W., and Zuiderweg, E.R.P. (1996). Stereospecific assignments of Val and Leu methyl groups in a selectively ¹³C-labeled 18 kDa polypeptide using 3D CT-(H)CCH-COSY and 2D 1JC-C edited heteronuclear correlation experiments. *J. Magn. Reson. B* 113, 70-75.
- Hu, J.-S., Grzesiek, S., and Bax, A. (1997). Two-dimensional NMR methods for determining χ_1 angles of aromatic residues in proteins from three-bond JC'Cg and JNCg couplings. *J. Am. Chem. Soc.* 119, 1803-1804.
- Johnson, B.A., and Blevins, J. (1994). NMRView: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* 4, 603-614.
- Jones, P.L., Veenstra, G.J.C., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, H., and Wolffe, A.P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genet.* 19, 187-191.
- Jones, S., Heyningen, P.V., Berman, H.M. and Thornton, J.M. (1999). Protein-DNA Interactions: A Structural Analysis. *J. Mol. Biol.* 287, 877-896.

- Kass, S.U., Landsberger, N. and Wolffe, A.P. (1997) DNA methylation directs a time-dependent repression of transcription initiation. *Curr. Biol.* 7, 157-165.
- Kass, S.U., Pruss, D., and Wolffe, A.P. (1997). How does DNA methylation repress transcription? *Trends Genet.* 13, 444-449.
- Katoh, E., Yamazaki, T., Kiso, Y., Wingfield, P.T. and Stahl, S.J. (1999). Determination of the Rate of Monomer Interchange in a Ligand-Bound Homodimeric Protein from NOESY Cross peaks- application to the HIV Protease/KNI-529 Complex. *J. Am. Chem. Soc.* 121, 2607-2608.
- Koradi, R., Billeter, M., and Wuthrich, K. (1996). MOLMOL: a program for the display and analysis of macromolecular structures. *J. Mol. Graph.* 14, 51-55.
- Lalande, M. (1996). Parental imprinting and human disease. *Annu. Rev. Genet.* 30, 173-195.
- Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8, 477-486.
- Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* 69, 905-914.
- Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. *Nature* 366, 362-365.
- Luger, K., Mader, A.W., Richmond, R. K., Sargent, D.F. and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251-60.
- Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L., and Bird, A.P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* 69, 905-914.
- Nan, X., Meehan, R.R., and Bird, A. (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Research* 21, 4886-4892.
- Nan, X., Campoy, F.J., and Bird, A. (1997). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* 88, 471-481.
- Nan, X., Ng, H.-H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393, 386-389.
- Ng, H. -H., Zhang, Y., Hendrich, B., Johnson, C. A., Turner, B. M., Erdjument-Bromage, H., Tempst, P., Reinberg, D. and Bird, A. (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nature Genet.* 23, 58-61.

- Ng, H.H., Jeppesen, P. and Bird, A. (2000). Active repression of methylated genes by the chromosomal protein MBD1. *Mol. Cell. Biol.* 20, 1394-406.
- Nicholls, A., Sharp, K.A., and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* 11, 281-296.
- Ohki, I., Shimotake, N., Fujita, N., Nakao, M. and Shirakawa, M. (1999). Solution structure of the methyl-CpG-binding domain of the methylation-dependent transcriptional repressor MBD1. *EMBO J.* 18, 6653-61.
- Ohki, I., Shimotake, N., Fujita, N., Jee, J.-G., Ikegami, T., Nakao, M., and Shirakawa, M., Structure of the methyl-CpG-binding domain of human MBD1 in complex with methylated DNA, *submitted for publication*
- Omichinski, J.G., Pedone, P.V., Felsenfeld, G., Gronenborn, A.M. and Clore, G.M. (1997). The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode. *Nat. Struct. Biol.* 4, 122-132.
- Qian, X., Gozani, S. N., Yoon, H., Jeon, C. J., Agarwal, K. and Weiss, M. A. (1993). Novel zinc finger motif in the basal transcriptional machinery: three-dimensional NMR studies of the nucleic acid binding domain of transcriptional elongation factor TFIIS. *Biochemistry* 32, 9944-9959.
- Razin, A. (1998). CpG methylation, chromatin structure and gene silencing - a three-way connection. *The EMBO J.* 17, 4905-4908.
- Shimotake, N. (2000). thesis for master degree. Nara Inst. Sci. Tech., Nara, JAPAN
- Siegfried, Z., and Cedar, H. (1997). DNA methylation: a molecular lock. *Curr. Biol.* 7, 305-307.
- Sklenar, V., and Bax, A. (1987). Spin echo water suppression for the generation of pure phase two-dimensional NMR spectra. *J. Magn. Reson.* 74, 469-479.
- Surani, M.A. (1998). Imprinting and the initiation of gene silencing in the germ line. *Cell* 93, 309-312.
- Sutcliffe, J.S., Nakao, M., Christian, S., Orstavik, K.H., Tommerup, N., Ledbetter, D.H., and Beaudet, A.L. (1994). Deletions of a differentially methylated CpG island at the SNRPN gene define a putative imprinting control region. *Nature Genet.* 8, 52-58.
- Suzuki, M., Yagi, N., and Gerstein, M. (1995). DNA recognition and superstructure formation by helix-turn helix proteins. *Protein Engineering* 8, 329-338.
- Tajima, S., and Suetake, I. (1998). Regulation and function of DNA methylation in vertebrates. *J. Biochem.* 123, 993-999.

- Tate, P., Skarnes, W., and Bird, A. (1996). The methyl-CpG binding protein MeCP2 is essential for embryonic development in the mouse. *Nature Genet.* 12, 205-208.
- Van den Veyver, I.B. and Zoghbi, H.Y. (2000). Methyl-CpG-binding protein 2 mutations in rett syndrome. *Curr. Opin. Genet. Dev.* 10, 275-9.
- Vuister, G.W., Tessari, M., Krimi-Nejad, Y. and Whitehead, B. (1998). in *Biological Magnetic Resonance*, Krishna, N. R. and Berliner, L. J., eds. (Kluwer Academic / Plenum Publishers, New York), 195-257
- Wade, P. A., Jones, P. L., Vermaak, D. and Wolffe, A. P. (1998). A multiple subunit Mi-2 histone deacetylase from *Xenopus laevis* cofractionates with an associated Snf2 superfamily ATPase. *Currnet Biology* 8, 843-846.
- Wade, P. A., Geggone, A., Jones, P. L., Ballestar, E., Aubry, F., and Wolffe, A. P. (1999). Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. *Nature Genet.* 23, 62-66.
- Wakefield, R.I., Smith, B.O., Nan, X., Free, A., Soteriou, A., Uhrin, D., Bird, A.P. and Barlow, P.N. (1999). The solution structure of the domain from MeCP2 that binds to methylated DNA. *J. Mol. Biol.* 291, 1055-65.
- Wishart, D.S., and Sykes, B.D. (1994). Chemical shifts as a tool for structure determination. *Methods Enzymol.* 239, 363-392.
- Wojciak, J.M., Connolly, K.M., and Clubb, R.T. (1999). NMR structure of the Tn916 integrase-DNA complex. *Nature Structure Biology* 6, 366-373.
- Wuthrich, K. (1986). *NMR of proteins and nucleic acid.* (Wiley, New York). 292 pages.
- Xu, G. L., Bestor, T.H., Bourc'his, D., Hsieh, C.L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J. and Viegas-Pequignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* 402 187-91.
- Yamazaki, T., Forman-Kay, J.D., and Kay, L.E. (1993). Two dimensional NMR experiments for correlating ^{13}C and ^1H d/e chemical shifts of aromatic residues in ^{13}C -labelled proteins via scalar coupling. *J. Am. Chem. Soc.* 115, 11054-11055.
- Yamazaki, T., Pascal, S.M., Singer, A.U., Forman-Kay, J.D. and Kay, L.E. (1995). NMR pulse schemes for the sequence-specific assignment of arginine guanidino ^{15}N and ^1H chemical shifts in proteins. *J. Am. Chem. Soc.* 117, 3556-3564.
- Yusufzai, T.M., and Wolffe, A.P. (2000). Functional consequences of rett syndrome mutations on human MeCP2. *Nucleic Acids Res.* 28, 4172-4179.
- Zhou, P., Sun, L.G., Dotsch, V., Wagner, G., Verdine, G.L. (1998). Solution structure of the core NFATC1/DNA complex. *Cell* 92, 687-696.

9 LIST OF PUBLICATIONS

1. Ohki, I., Shimotake, N., Fujita, N., Nakao, M., and Shirakawa, M. (1999). Solution structure of the methyl-CpG-binding domain of the methylation-dependent transcriptional repressor MBD1. EMBO J. 18, 6653-6661.
2. Fujita, N., Shimotake, N., Ohki, I., Chiba, T., Saya, H., Shirakawa, M., and Nakao, M. (2000). Mechanism of transcriptional regulation by methyl-CpG binding protein MBD1. Mol. Cell. Biol. 20, 5107-5118.
3. Ikegami, T., Okada, T., Ohki, I., Hirayama, J., Mizuno, T., and Shirakawa, M. (2001). Solution Structure and Dynamic Character of the Histidine-Containing Phosphotransfer Domain of Anaerobic Sensor Kinase ArcB from Escherichia coli. Biochemistry 40, 375-386.

In preparation:

Ohki, I., Shimotake, N., Fujita, N., Jee, J.-G., Ikegami, T., Nakao, M., and Shirakawa, M., Structure of the methyl-CpG-binding domain of human MBD1 in complex with methylated DNA, *being submitted for publication*