メチル化 DNA 結合タンパク質 MBD4 の可塑的な塩基認識機構

有吉眞理子,^{*,a,†} 大谷 淳二,^{b,††} 白川 昌宏^b

Structure Basis of Versatile Base Recognition of MBD4

Mariko Ariyoshi^{*,a,†}, Junji Otani^{b,††} and Masahiro Shirakawa^b^a*Institute for Integrated Cells-Material Sciences, Kyoto University; Yoshida-Honmachi, Sakyo-ku, 606-8501 Kyoto, Japan;* and ^b*Graduate School of Engineering, Kyoto University; Kyoto-daigaku-katsura, Nishikyo-ku, 615-8510 Kyoto, Japan.*

(Received July 26, 2014)

DNA methylation is one of the major epigenetic marks in the mammalian genome to define chromatin higher-order structure, and plays essential roles in various developmental processes. In the mammalian genome, DNA methylation mainly occurs at the 5th position of cytosine bases in a palindromic 5'-CG-3' dinucleotide sequence. Methyl CpG binding domain (MBD) proteins recognize symmetrically methylated CpG sites (^{5m}CG/^{5m}CG) through a conserved MBD, and recruit transcriptional repressors or chromatin modifiers. One of the MBD proteins, MBD4, uniquely contains a C-terminal glycosylation domain together with an N-terminal MBD, and functions as a mismatch DNA repair enzyme specific for T/G or U/G mismatch bases generated by spontaneous deamination of 5-methylcytosine. The base excision activity of MBD4 is also implicated in active DNA demethylation initiated by the conversion of 5-methylcytosine to thymine by deaminases. Unlike other MBD proteins, MBD4 recognizes not only ^{5m}CG/^{5m}CG but also T/G mismatched sites generated by spontaneous deamination of 5-methylcytosine (^{5m}CG/TG). In addition, our biochemical data demonstrate that MBD also binds to intermediates in DNA demethylation pathways, such as 5-hydroxymethyl-cytosine (^{hm}C), 5-carboxyl-cytosine and 5-hydroxy-uracil. The crystal structures of MBD_{MBD4} in complex with ^{5m}CG/TG, ^{5m}CG/^{5m}CG or ^{5m}CG/^{hm}CG provide new structural insights into the versatility of base recognition by MBD4. A DNA interface of MBD4 has flexible structural features, in which an extensive hydration water network supports the versatile base specificity of MBD4. The versatile base recognition by MBD_{MBD4} implies multi-functional roles of MBD4 in the regulation of dynamic DNA methylation patterns.

Key words—epigenetics; DNA methylation; crystal structure

1. シトシンのメチル化：ゲノム DNA のエピジェネティック修飾

真核生物ゲノムを格納しているクロマチンの高次構造はダイナミックに変化し、遺伝子発現や複製・修復などの核内現象を制御する。後生的なゲノムの化学修飾である DNA メチル化は、クロマチンの基本構造ユニットであるヌクレオソームを構成するヒストンタンパク質の翻訳後修飾・バリエーションとともにクロマチン高次構造を規定するエピジェネティック

クマーカーとして機能する。脊椎動物のゲノムでは、主として 3'-CG-5' の回文配列 (CpG 配列) 中のシトシンの 5 位の炭素がメチル化され、一般的には、クロマチンの凝集 (ヘテロクロマチン形成) を促し、転写を負に制御するエピジェネティックマーカーとして機能する。すなわち、ゲノム DNA のメチル化状態に応じて、転写抑制複合体やクロマチンリモデリング複合体がクロマチン上に会合し、ダイナミックな遺伝子発現制御が行われる。このような DNA メチル化を介したエピジェネティックな遺伝子発現制御は、発生・分化の過程と密接に係わっており、X 染色体の不活化、ゲノムインプリンティング、ゲノム安定性の維持など様々な高次の細胞機能において重要な役割を担っている。¹⁾

個々の細胞の機能や分化状態を規定する固有の DNA メチル化様式は、発生・分化の過程で新規メチル化酵素である DNA methyltransferase (DNMT)

^a京都大学物質-細胞統合システム拠点 (〒606-8501 京都市左京区吉田本町), ^b京都大学工学研究科 (〒615-8510 京都市西京区京都大学桂)

現所属: [†]京都大学工学研究科分子工学専攻 (〒615-8510 京都市西京区京都大学桂), ^{††}京都大学大学院生命科学研究科統合生命科学専攻 (〒606-8501 京都市左京区吉田近衛町)

*e-mail: ariyoshi.mariko.2x@kyoto-u.ac.jp

本総説は、日本薬学会第 134 年会シンポジウム S40 で発表した内容を中心に記述したものである。

Fig. 1. Regulation Pathways of DNA Methylation

A) Regulation of DNA methylation pattern; *de novo* methylation by DNMT3, maintenance DNA methylation by UHRF1 and DNMT1 and demethylation by TET proteins and DNA excision repair enzymes. B) DNA deamination pathways *via* oxidation and/or deamination of 5-methylcytosine couples with base excision repair enzymes, such as MBD4 and TDG. Comparison of affinity of MBD_{MBD4} for XG/5mCG containing DNA is indicated; X represents 5mC, T or oxidation or deamination products. K_d values for 5mCG/5mCG, TG/5mCG and hmCG/5mCG are indicated.

3 ファミリータンパク質によって形成され、維持メチル化酵素である DNMT1 とメチル化 DNA 結合タンパク質である ubiquitin-like PHD and RING finger domain containing protein 1 (UHRF1) によって、DNA 複製・細胞分裂サイクルを通して忠実に継承される [Fig. 1 (A)].²⁻⁴⁾ DNA メチル化様式の形成・継承は正常な個体形成や生命機能の維持に不可欠であり、DNMT 欠損マウスは、胎生致死となることが報告されている。また、がん細胞では、慢性炎症や喫煙、加齢等に伴う様々なレベルのゲノム DNA のメチル化異常の蓄積が報告されている。一

方、生殖細胞の分化や胚発生の過程では、ゲノムワイドな DNA 脱メチル化による細胞のリプログラミングがおこる。⁵⁾ 分化した細胞においても特異的なゲノム領域における脱メチル化が観察されている。^{6,7)} すなわち、ゲノムの DNA メチル化様式は、個体の生命活動のサイクルに応じて、精緻かつダイナミックに制御されている。さらに、最近、5-メチルシトシン (5mC) がさらに酸化修飾を受け、5-ヒドロキシメチルシトシン (5hmC)、5-ホルミルシトシン (5foC)、5-カルボキシルシトシン (5caC) へと変換され、よりダイナミックな DNA メチル化制御

Fig. 2. Structures of Methylated DNA Binding Proteins

A) Overall structure of MeCP2 MBD bound to methyl CpG site (upper panel) and a close-up view of its DNA binding interface (lower panel). B) Overall structure of the SRA domain of UHRF1 bound to hemi-methyl CpG site (upper panel) and a close-up view of its DNA binding interface (lower panel). C) Overall structure of the three zinc fingers of Kaiso, which is a representative of ZF-MDBP, bound to methyl CpG site (upper panel) and a close-up view of its DNA binding interface (lower panel).

を可能にしていることがわかってきた。^{5mC}の酸化酵素である ten-eleven translocation (TET) タンパク質が同定され、^{5mC}の酸化を介した脱メチル化の新規の分子機構が明らかになってきた [Fig. 1(B)].⁸⁾

現在までに、筆者らの研究グループは、このようなダイナミックな DNA メチル化制御に係わる様々なタンパク質因子の構造機能解析を行い、その分子機構の解明に取り組んできた。本稿では、メチル化 DNA 領域の DNA 修復、及び DNA 脱メチル化への関与も示唆される methyl CpG binding domain 4 (MBD4) タンパク質のメチル化シトシン及びその酸化誘導体認識のユニークな構造基盤について、既知のメチル化 DNA 結合タンパク質の塩基認識機構との比較も含めて解説する。⁹⁾

2. メチル化 DNA 結合タンパク質ファミリーの構造基盤

CpG 配列中のシトシンのメチル化は、DNA2 重鎖の両方の鎖で対称的に起こる [Fig. 1(A)]. ゲノム DNA においては、このようなメチル化修飾は比較的小さな官能基の付加であり、典型的な DNA2

重鎖構造を大きく破壊することはない。生体内には、このようなメチル化修飾を精確に読み取り、様々なクロマチン制御因子や転写因子を適切にリクルートするメチル化 DNA 結合タンパク質が存在する。現在までに 3 つのメチル化 DNA 結合タンパク質ファミリーが同定されている；① methyl CpG binding domain (MBD) タンパク質、② SET and RING finger associated domain (SRA) タンパク質、③ zinc finger type methylated DNA binding protein (ZF-MDBP)。Figure 2 に示すように、代表的なファミリータンパク質の機能ドメインの立体構造が決定され、メチル化 DNA 認識の構造基盤が明らかにされている。

MBD タンパク質は、進化的に保存された MBD を介して、メチル化 CpG 配列を認識する [Fig. 2 (A)]. 典型的な MBD タンパク質である methyl CpG binding protein 2 (MeCP2), MBD1, MBD2 に関して、MBD とメチル化 DNA の複合体の立体構造が決定されており、共通したメチル化 CpG 認識機構が明らかになっている。MBD は DNA2 重鎖の構造を歪ませることなく、主溝側に結合する。2 つ

の保存されたアルギニン側鎖を合理的に配置することによって、DNA鎖中に対称的に配置されたメチル化 CpG 配列中のグアニン骨格と^{5m}Cの5位のメチル基を特異的に認識している。¹⁰⁻¹²⁾メチル化 CpG 配列に隣接する塩基配列の認識はみられない。これら MBD タンパク質は、MBD 以外にそれぞれに特徴的な転写抑制ドメインやタンパク質-タンパク質相互作用ドメインを持ち、メチル化 DNA 上への転写抑制複合体やクロマチンリモデリング複合体の会合の要として機能している。

SRA タンパク質である UHRF1 は、SRA ドメインを介して DNA 複製後に生じる非対称な片鎖メチル化状態を特異的に認識し、DNA 維持メチル化に必須の因子である。¹³⁾ UHRF1 はメチル化されたシトシンを DNA2 重鎖からフリップアウトさせ、SRA ドメイン表面に形成された疎水性のポケットに取り込むことによって、シトシンのメチル化状態を厳密に認識している [Fig. 2(B)]。

第3の ZF-MDBP ファミリーに属する Kaiso 及び ZFP57 タンパク質は、複数の Cys₂His₂ タイプの亜鉛フィンガーモチーフを介してメチル化 CpG 配列を認識する [Fig. 2(C)]。^{14,15)} MBD が2つの^{5m}CG 塩基対を等価に認識しているのとは対照的に、ZF-MDBP のメチル化 DNA 認識では、対称的に配置されている^{5m}Cの一方は特異的に認識されているが、もう一方の^{5m}Cの認識はあいまいである。また、複数の Zn フィンガーモチーフを用いることによって、^{5m}CpG 配列周辺の塩基配列も認識している。このような塩基認識の特徴は、これら ZF-MDBP が特定のプロモーター領域のメチル化依存的な転写抑制に関与することを考慮すると合理的である。細胞内には、同様の亜鉛フィンガーモチーフを持つ DNA 結合タンパク質が多数同定されており、その一部は DNA のメチル化、酸化修飾を認識する因子であると推測される。多様な ZF-MDBP が、組織特異的若しくは経時的な遺伝子発現制御に係わっている可能性があるだろう。

3. MBD4 の塩基認識特異性

^{5m}C 塩基は生理条件下での自発的な脱アミノ化などを介して容易にチミン塩基に変換され (1細胞あたり 2-300 変換/日)、がんやその他の疾患に関連した変異の原因となる T/G ミスマッチを生じることが知られている。MBD ファミリーに属する MBD4

は、N 末端領域の MBD とともに C 末端領域に塩基除去活性ドメインを持つ T/G ミスマッチ塩基除去修復酵素である。通常、転写制御因子として同定されている他の MBD タンパク質とは異なり、メチル化 CpG 配列に生じたミスマッチ塩基除去活性を有し、ゲノム安定性などに重要な役割を果たす。MBD4 による DNA 修復機能は、ゲノム DNA のメチル化状態の保持に貢献する。MBD4 の MBD (MBD_{MBD4}) は、他の MBD タンパク質の MBD と類似のアミノ酸配列を保持しているにもかかわらず、両鎖メチル化 CpG (^{5m}CG/^{5m}CG) と、脱アミノ化により生じる T/G ミスマッチを含む配列 (^{5m}CG/TG) の両方に結合するというユニークな DNA 結合能を有している。

一方、近年、メチル化シトシンの酸化・脱アミノ化と塩基除去修復を介した脱メチル化経路が提唱されている [Fig. 1(B)].⁸⁾ その詳細な分子機構に関しては、不明な点も残されているが、^{5m}C の酸化、脱アミノ化に係わる酵素として、それぞれ TET タンパク質、及び activation-induced DNA-cytosine deaminase (AID) が同定されている。さらに、T/G ミスマッチ塩基除去修復酵素である MBD4 や thymine DNA glycosylase (TDG) の DNA 脱メチル化反応への関与が示唆されている。このような研究背景をふまえ、メチル化シトシンの酸化誘導体である^{hm}C、^{fo}C、^{ca}C、及び^{hm}Cの脱アミノ化誘導体、^{hm}U も含めて、MBD_{MBD4} とエピジェネティック修飾された CpG 配列の結合を定量的に解析した。等温滴定カロリーメトリー (isothermal titration calorimetry; ITC) 法を用いて得られたそれぞれの解離定数 (K_d) を Fig. 1(B) に示す。筆者らの解析の結果から、MBD_{MBD4} は両鎖メチル化 (^{5m}CG/^{5m}CG)、若しくは、その脱アミノ化産物である T/G ミスマッチ (^{5m}CG/TG) を含む DNA 断片と同等の親和性で結合することが示された。典型的な MBD タンパク質である MBD1 の MBD (MBD_{MBD1}) の場合、T/G ミスマッチに対する結合親和性は、両鎖メチル化 DNA との親和性の 1/10 程度であった。さらに、MBD_{MBD1} がメチル化 DNA にのみ厳密な結合特異性を示すのとは対照的に、MBD_{MBD4} は、^{hm}C、^{fo}C、^{hm}U を含む DNA に対しても有意に結合することが明らかになった。筆者らは、このような MBD4 にユニークな基質認識を実現する構造要因を明らかに

Fig. 3. Structural Comparison of DNA Binding Interface between MBD_{MBD4} (upper panels) and MBD_{MeCP2} (lower panels)

Structures of DNA binding interface in the MBD_{MBD4} - $5mC/G/5mC/G$ and MBD_{MeCP2} - $5mC/G/5mC/G$ complexes are shown. A) Symmetrical arrangement of Arg-fingers are involved in $5mC/G$ recognition. B) Recognition of $5mC/G$ via Arg-finger 1. C) Recognition of $5mC/G$ via Arg-finger 2. Orientation of Tyr96 of $MBD4$ is opposite to that of Tyr123 of $MeCP2$. Glu 137 makes hydrogenbonds with Arg-finger 2 in MBD_{MeCP2} . The conserved acidic residue (Glu137 in $MeCP2$) is replaced by Ser in $MBD4$ (Ser110).

するために、 MBD_{MBD4} と $5mC/G/TG$, $5mC/G/5mC/G$ 若しくは $5mC/G/hmC/G$ を含む DNA 断片との複合体の結晶構造を決定した。さらに、結晶中では、 MBD_{MBD4} の非特異的な DNA 結合状態の構造も観察された。

4. $MBD4$ の可塑的な DNA 結合面が可能にする寛容な基質特異性

まず、 MBD_{MBD4} - $5mC/G/5mC/G$ 複合体の結晶構造を代表的な MBD タンパク質である $MeCP2$ の MBD (MBD_{MeCP2}) と DNA の複合体構造と比較する。¹¹⁾ 骨格構造、及び対称的な $5mC/G/5mC/G$ 配列を認識する 2 つのアルギニン側鎖 (Arg-finger 1, 2), を含む基本的な DNA 認識モチーフは $MBD4$ においても保存されている。Figure 3(A) に示すように、Arg-finger 1 と 2 の配置は 2 つの結晶構造でよく一致していた。しかしながら、 $MBD4$ と $MeCP2$ の DNA 相互作用表面の構造を詳細に比較すると、メチル化 CpG の認識に重要だとされるチロシン残基側鎖の配向の違いと、片方の Arg-finger (Arg-finger 2) の側鎖の柔軟性という 2 つの構造要因が $MBD4$ のユニークな基質特異性を説明し得ることが判明した [Fig. 3(C)]。

$MeCP2$ の保存されたチロシン残基、Tyr123 は、DNA 側を向いて水を介した CH-O タイプの水素結合により $5mC$ の 5 位のメチル基を認識している。それに対して、 $MBD4$ では対応する Tyr96 の側鎖が DNA 結合表面とは反対側を向いており、DNA 結合表面に広い空間が形成されている。この $MBD4$ の DNA 結合表面に生じた空間には、タンパク質分子と DNA 分子の間を埋めるように水と水のネットワークが形成されている。一方、 MBD_{MeCP2} の構造にみられるように、典型的な MBD では、メチル化 DNA 認識に重要な 2 つの Arg-finger の側鎖は、それぞれの近傍の酸性アミノ酸残基との水素結合によって、メチル化 CpG 配列の認識に有利なコンフォメーションに固定されている。 $MBD4$ においても、Arg-finger 1 の側鎖は同様に酸性残基 Asp94 との水素結合によって固定されており、その $5mC/G$ 塩基対の認識は、DNA 結合に関する水分子の配置も含めて $MBD4$ と $MeCP2$ で保存されている [Fig. 3(B)]。しかし、 $MBD4$ では、Arg-finger 2 を固定する酸性残基を欠いている。そのために、この Arg 側鎖のコンフォメーションは柔軟に変化することが可能である。

Fig. 4 Structural Flexibility of DNA Binding Interface of MBD_{MBD4}

A) Recognition of the TG mismatch base pair by Arg-finger 2. B) Recognition of the ^{5m}CG base pair by Arg-finger 2. C) Structural comparison of Arg-finger 2 in the specific ^{5m}CG/^{5m}CG and non-specific DNA bound forms. Arg-finger 2 is involved in base recognition in the specific complex (upper panel) while it makes a contact with phosphate backbone in the non-specific complex (lower panel).

水和水のネットワークと Arg-finger 2 の側鎖の自由度から生じる MBD_{MBD4} の DNA 結合面の可塑性は、MBD_{MBD4}-TG/^{5m}CG 複合体における塩基認識に寄与している [Fig. 4(A)]. MBD4 の 2 つの Arg-finger 2 による ^{5m}CG/TG の認識様式は、^{5m}CG/^{5m}CG の場合と大局的には類似している。一見すると、チミン塩基とメチル化シトシン塩基の化学構造の違いは顕著ではないが、シトシンがメチル化されても C/G 塩基対の水素結合パターンは維持されるのに対して、T/G ミスマッチ塩基対では Watson-Click 型の水素結合が形成できなくなる。それゆえ、^{5m}C/G 塩基対のメチル化シトシンと比較すると、T/G 塩基対中のチミン塩基は DNA 二重鎖の主溝側へ 1 Å 程度突出している。MBD4 に特徴的な DNA 結合面の空間は、立体障害を引き起こすことなく突出したチミン塩基を受け入れること可能である。メチル化 DNA とミスマッチ DNA の違いに応じて、水分子間の水素結合のネットワークパターンと Arg-finger 2 の側鎖のコンフォメーションが変化してい

るのが見て取れる。MBD_{MBD4}-^{hm}C/^{5m}CG 複合体においても、DNA 結合表面の水和水のネットワークが変化して、^{hm}C 塩基のヒドロキシ基と水素結合を形成している [Fig. 4(B)]. さらに、この DNA 結合面の可塑性は、非特異な DNA 複合体における DNA 結合面の構造を比較してみるとさらに顕著である [Fig. 4(C)]. 通常、MBD4 で観察された非特異な DNA 結合は、他の MBD タンパク質では検出されない。

5. まとめ

筆者らの MBD_{MBD4} の結晶構造解析から、MBD_{MBD4} の寛容な塩基認識を支える 2 つの特徴、①DNA 結合面の可塑的な水和水のネットワーク形成、②Arg-finger 2 の側鎖の柔軟性、が示された。MBD4 の DNA 結合表面の柔軟性は、厳密にメチル化シトシンを認識する他の MBD や UHRF1 の SRA ドメインにはみられないユニークなものである。MBD_{MBD4} の寛容な DNA 認識特異性は、MBD4 が転写抑制と T/G ミスマッチ修復という複

数の機能を果たすことを反映していると考えられることができる。また、MBD4は、^{hm}C, ^{fo}C, ^{hm}Uを含むメチル化 CpG 配列も認識できることが明らかになった。しかし、筆者らが行った MBD4 の塩基除去実験では、T/G, U/G, ^{hm}U/G といった、ミスマッチ塩基対のみが基質として切除された。^{hm}C, ^{fo}C, ^{ca}C 塩基は、現在 DNA 脱メチル化の中間体として非常に注目されているが、細胞内でこれらの塩基は長寿命であり、エピジェネティックマークとして機能する可能性も唆されている。^{16,17)} MBD4 がこのような酸化シトシンのシグナルを読み取るリーダーとして働いている可能性もあるであろう。DNA 脱メチル化や酸化シトシン塩基の機能における MBD4 の役割については今後の研究が待たれる。

謝辞 本研究は、大阪大学蛋白質研究所の田嶋正二教授、末武 勲准教授、木村博信助教との共同研究として行われました。本研究の遂行にあたり、MBD4 の結晶構造解析にご協力頂いた有田恭平博士(現・横浜市立大学准教授)に深く感謝致します。

利益相反 開示すべき利益相反はない。

REFERENCES

- 1) Bird A., *Genes Dev.*, **16**, 6–21 (2002).
- 2) Okano M., Bell D. W., Haber D. A., Li E., *Cell*, **99**, 247–257 (1999).
- 3) Bostick M., Kim J. K., Estève P. O., Clark A., Pradhan S., Jacobsen S. E., *Science*, **317**, 1760–1764 (2007).
- 4) Sharif J., Muto M., Takebayashi S., Suetake I., Iwamatsu A., Endo T. A., Shinga J., Mizutani-Koseki Y., Toyoda T., Okamura K., Tajima S., Mitsuya K., Okano M., Koseki H., *Nature (London)*, **450**, 908–912 (2007).
- 5) Shearstone J. R., Pop R., Bock C., Boyle P., Meissner A., Socolovsky M., *Science*, **334**, 799–802 (2011).
- 6) Kangaspeska S., Stride B., Métivier R., Polycarpou-Schwarz M., Ibberson D., Carmouche R. P., Benes V., Gannon F., Reid G., *Nature (London)*, **452**, 112–115 (2008).
- 7) Métivier R., Gallais R., Tiffocche C., Le Péron C., Jurkowska R. Z., Carmouche R. P., Ibberson D., Barath P., Demay F., Reid G., Benes V., Jeltsch A., Gannon F., Salbert G., *Nature (London)*, **452**, 45–50 (2008).
- 8) Kohli R. M., Zhang Y., *Nature (London)*, **502**, 472–479 (2013).
- 9) Otani J., Arita K., Kato T., Kinoshita M., Kimura H., Suetake I., Tajima S., Ariyoshi M., Shirakawa M., *J. Biol. Chem.*, **288**, 6351–6362 (2013).
- 10) Ohki I., Shimotake N., Fujita N., Jee J., Ikegami T., Nakao M., Shirakawa M., *Cell*, **105**, 487–497 (2001).
- 11) Ho K. L., McNae I. W., Schmiedeberg L., Klose R. J., Bird A., Walkinshaw M. D., *Mol. Cell*, **29**, 525–531 (2008).
- 12) Scarsdale J. N., Webb H. D., Ginder G. D., Williams D. C. Jr., *Nucleic Acids Res.*, **39**, 6741–6752 (2011).
- 13) Arita K., Ariyoshi M., Tochio H., Nakamura Y., Shirakawa M., *Nature (London)*, **455**, 818–821 (2008).
- 14) Buck-Koehntop B. A., Stanfield R. L., Ekiert D. C., Martinez-Yamout M. A., Dyson H. J., Wilson I. A., Wright P. E., *Proc. Natl. Acad. Sci. USA*, **109**, 15229–15234 (2012).
- 15) Liu Y., Toh H., Sasaki H., Zhang X., Cheng X., *Genes Dev.*, **26**, 2374–2379 (2012).
- 16) Inoue A., Shen L., Dai Q., He C., Zhang Y., *Cell Res.*, **21**, 1670–1676 (2011).
- 17) Inoue A., Zhang Y., *Science*, **334**, 194 (2011).