

MBD 结构域和 SRA 结构域识别 甲基化 DNA 的结构机理 *

杨 娜 许瑞明 **

(中国科学院生物物理研究所, 生物大分子国家重点实验室, 北京 100101)

摘要 DNA 胞嘧啶 5- 甲基化修饰是表观遗传重要的修饰之一, 其对基因表达的调控依赖下游的识别蛋白识别和传递甲基化信号. 本文围绕两种主要的甲基化 DNA 识别结构域——MBD 结构域和 SRA 结构域, 综述了它们识别不同修饰形式 DNA 的结构基础以及发挥功能的分子机理.

关键词 DNA 修饰识别, 5- 甲基胞嘧啶, MBD 结构域, SRA 结构域, 转录调控, 结构机理

学科分类号 Q6

DOI: 10.3724/SP.J.1206.2014.00284

DNA 甲基化修饰是最为重要的表观遗传修饰之一. 分子遗传学研究表明, DNA 胞嘧啶甲基化修饰通常与基因表达沉默密切相关, 在 X 染色体失活、基因印记、转座子失活等重要发育进程中发挥作用^[1]. 相对于其他表观遗传修饰(如组蛋白修饰), DNA 甲基化修饰更具稳定性, 可在细胞分裂过程中稳定遗传, 并且已有证据表明父本 DNA 的甲基化模式可遗传给子代^[2-4].

脊椎动物中 DNA 甲基化修饰主要发生在 CpG 二核苷酸位点的胞嘧啶第 5 位碳原子上(5mC)(图 1). 哺乳动物中约有 80% 的 CpG 位点均发生了甲基化修饰. 基因启动子位置的甲基化修饰会造成基因表达沉默^[5], 而通常看家基因(housekeeping genes)的启动子位置富含 CpG 孤岛, 是不发生甲基化修饰的 CpG 位点. 近年来, 在鼠和人的胚胎干细胞和卵母细胞中发现大量存在非 CpG 位点(CpH, H 代表 A、C 或 T)的甲基化修饰, 并且该甲基化修饰程度与基因表达水平成正相关, 但这些非 CpG 甲基化修饰的生物学意义还有待进一步研究^[2].

1 DNA 甲基化修饰酶

DNA 甲基化修饰是由 DNA 甲基化酶催化完成的, 目前在哺乳动物中共发现了两类催化 5mC 甲

基化修饰的 DNA 甲基化酶, 根据催化的底物不同, 分为从头甲基化酶和维持型甲基化酶两类. 从头甲基化酶以双链都未发生甲基化的 DNA 为底物, 可完成两条链上的甲基化修饰, 包括 DNMT3a、DNMT3b 和没有催化活性的 DNMT3L^[6-7]. 而维持型甲基化酶只能以一条链上已经发生甲基化修饰的 DNA(半甲基化 DNA)为底物, 完成配对链上的甲基化反应, 主要在 DNA 半保留复制过程中维持新生链上的 DNA 甲基化水平, 目前只发现了一种, 即最早发现的 DNA 甲基化酶 DNMT1^[8].

类似组蛋白上的修饰可写可擦, DNA 的甲基化修饰也是可以去掉的, 这种去甲基化又可分为被动过程和主动过程. 其中被动去甲基化过程主要是依赖 DNA 复制的稀释过程, 通过抑制 DNMT1 的活性, 如阻止 DNMT1 入核, 使新生链上不能发生甲基化修饰, 因此 DNA 甲基化程度随着复制逐渐降低, 例如受精卵发育过程中来自母本 DNA 的去

* 国家自然科学基金(31370734, 31430018)和国家重点基础研究发展计划(973)(2015CB856202)资助项目.

** 通讯联系人.

Tel: 010-64888797, E-mail: rmxu@sun5.ibp.ac.cn

收稿日期: 2014-10-08, 接受日期: 2014-10-11

甲基化过程^[9-10]；主动去甲基化现象在胚胎发育早期以及体细胞中特异位点的基因上都有发现，但这种主动去甲基化过程的分子机制还不是很清楚。多年来，科学家们都在寻找哺乳动物中能够直接切除 DNA 5mC 甲基的去甲基化酶，试图找到不依赖于碱基错配而发生碱基切除修复的主动去甲基化过程。先后有报道称 ribozyme 和甲基化 DNA 结合蛋白 MBD2 具有直接切除 5mC 碱基的活性，但被证明都是错的^[11-12]。直到 2009 年 Rao 和 Zhang 实验室先后发现 TET 蛋白(ten-eleven translocation)可以氧化 5mC DNA 生成 5-羟甲基化胞嘧啶(5hmC) DNA^[13-15]。随后 Xu 和 Zhang 实验室分别报道了 5hmC 可被 TET 蛋白进一步氧化生成 5-甲酰基胞嘧啶(5fC)和 5-羧基胞嘧啶(5caC)^[16-17]。5hmU(5hmC 脱氨后生成)、5fC 以及 5caC 均可在 DNA 糖基化酶 TDG 的作用下去掉整个碱基，产生无碱基位点，进而通过碱基切除修复(base excision repair, BER)加上一个新的胞嘧啶，实现去甲基化过程。然而是否存在一个直接的脱羧酶，可以将 5caC 上的羧基直接去除，或是存在可以直接切除羟甲基的酶，进而简化 DNA 主动去甲基化过程，仍然是科学家们正在探索和寻找的问题(图 1)。

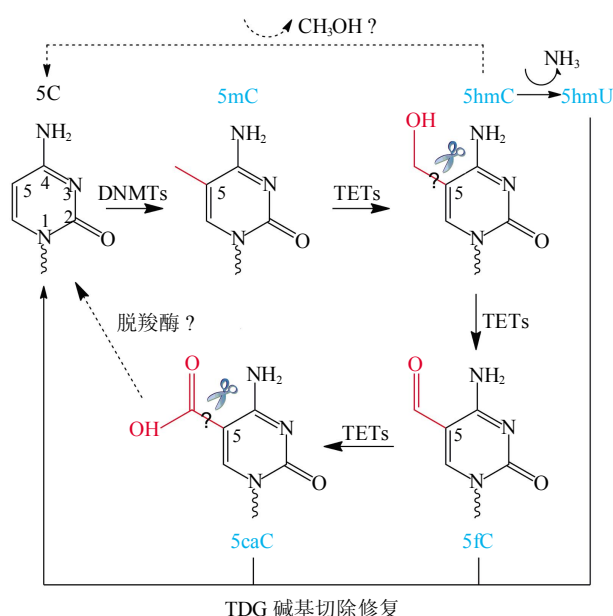


Fig. 1 Possible pathways of TET-mediated active DNA demethylation

图 1 TET 蛋白介导的 DNA 主动去甲基化过程的分子机制示意图

2 DNA 甲基化修饰的识别

DNA 甲基化修饰如何实现调控基因表达功能是表观遗传领域的一个核心问题。一种可能性是修饰的甲基基团妨碍转录因子与 DNA 的结合，特别是一些偏好结合富含 GC 序列的转录因子^[18]，如 CTCF 蛋白^[19]。另外一种可能性刚好和前者相反，甲基化的 DNA 促进了蛋白质因子的结合，一些具有特殊结构域的甲基化 DNA 结合蛋白可以识别甲基化修饰的 DNA^[20]，进而招募下游的共抑制因子，完成对目的基因表达的调控。

2.1 MBD 家族蛋白

MBD(methyl-CpG-binding domain)家族蛋白是最早发现的甲基化 DNA 结合蛋白。Bird 课题组早在 1989 年就从多种哺乳动物细胞中检测到 MeCP1 蛋白复合体和 MeCP2 蛋白可以特异地结合甲基化 DNA，其中 MeCP2 是其中第一个被纯化和克隆基因的蛋白^[21]。之后通过 MeCP2 DNA 结合结构域的同源比对又发现了该家族的另外几个成员 MBD1~4，其中 MBD2 是最早发现的 MeCP1 复合体中的重要组分(图 2a)。MBD 家族蛋白均包含 MBD 结构域和其他功能域，通过招募不同的共抑制复合体到目标基因的甲基化位点，抑制目标基因的表达(图 2b)^[5]。

2.1.1 MeCP2 蛋白

MeCP2 蛋白是 MBD 家族最早被鉴定的成员。MeCP2 蛋白广泛存在于哺乳动物神经细胞中，是神经元发育分化与功能维持的重要调节因子^[22-24]。已知 MeCP2 基因突变会引发一类严重的女童神经运动发育疾病，Rett 综合征^[25]。发病者主要是女性儿童，患者表现为头部萎缩、智力低下、呼吸异常、丧失四肢和语言功能等，并伴有孤独症行为。Rett 综合征的发病率约为 1/10 000，其中约有 87% 的患儿体内发生了 MeCP2 基因突变。

关于 MeCP2 致病的分子机制一个合理的解释是基因突变解除了 MeCP2 对目的基因的抑制，造成沉默基因的异常表达。MeCP2 蛋白主要包含两个结构域，甲基化 DNA 结合结构域(MBD)和转录抑制结构域(transcriptional repression domain, TRD)(图 2a)。研究表明，MBD 结构域偏好结合在周边富含 AT 碱基的甲基化 CpG 位点，一半以上的 MeCP2 致病突变都发生在 MBD 区域；MBD 和 TRD 结构域可招募共抑制复合体 Sin3A Complex^[26-27]。Sin3A 复合体包含组蛋白去乙酰化酶 HDAC 以及

组蛋白赖氨酸甲基化酶 HKMT, 通过组蛋白赖氨酸上的去乙酰化反应以及特定位点上的甲基化修饰, 影响染色质的空间结构, 形成相对紧密的染色质结构, 抑制基因表达. *Bdnf* (Brain-derived neurotrophic factor) 基因是第一个被鉴定的 MeCP2 目标基因^[28-29], 其编码的蛋白是一种脑源性神经营养因子, 与神经细胞活性相关, 对大脑接受的短期刺激转变为长程变化至关重要, 是学习和记忆过程的基础. 因此 MeCP2 蛋白对 *Bdnf* 基因表达的异常调控是 Rett 综合症的潜在致病原因之一. 后续又在老鼠中鉴定出其他几个 MeCP2 目标基因, 都可能与 Rett 综合症的发生密切相关.

2.1.2 MBD 蛋白

MBD1 蛋白是一个转录调节因子, 除了包含一个 MBD 结构域外, 还包含 2~3 个 CXXC 结构域, 以及 1 个 C 端的 TRD 结构域(图 2a). 其中 MBD 和 CXXC 结构域分别结合 mCpG 和 CpG DNA, CXXC 和 TRD 结构域可招募共抑制因子组蛋白

H3K9 甲基化酶 SETDB1, 并可在细胞有丝分裂的 S 期特异地形成 CAF1-MBD1-SETDB1 复合体, 在 CAF1 介导的复制依赖型染色质组装过程中促进 H3K9 位点甲基化反应, 沉默基因表达^[30].

MBD2 蛋白包含 MBD 结构域和相邻的 TRD 结构域(图 2a), 是最早发现的甲基化结合蛋白复合体 MeCP1 中的组分, 可特异地结合 mCpG DNA^[21]. 已知 MeCP1 复合体是由 NuRD 复合体(也称为 Mi-2 共抑制复合体)和 MBD2 蛋白组成的大复合体, 其中 NuRD 复合体包含染色质重塑因子 Mi-2、组蛋白去乙酰化酶 HDAC 以及 MBD3 等组分^[31-32]. MBD2 通过与 HDAC1 的相互作用, 促进组蛋白去乙酰化反应, 抑制特定位点的基因表达. MBD3 蛋白是 NuRD 复合体中的一个组分, 也包含 1 个 N 端的 MBD 结构域, 但对 mCpG DNA 的结合力并不强^[33]. MBD3 蛋白可介导 MTA2 蛋白(metastasis-associated protein 2)与组蛋白去乙酰化酶复合体间的相互作用.

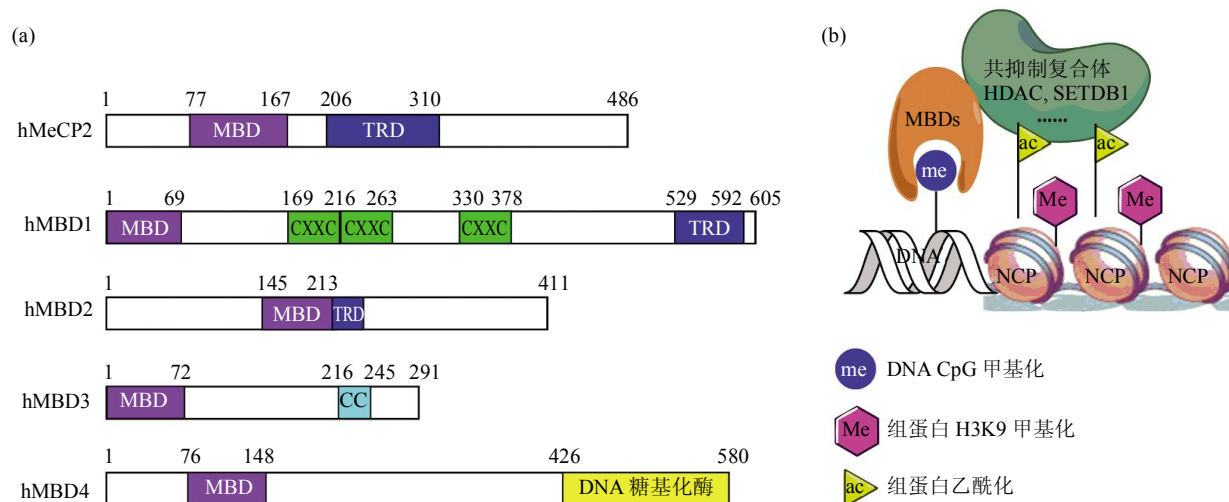


Fig. 2 Domain architecture of MBD proteins and model of their functions on gene expression

图 2 MBD 家族蛋白结构域组成以及调控基因表达的模式图

(a) 人源 MBD 结构域蛋白的结构域组成和分布: MBD 结构域紫色标注, TRD 结构域蓝色标注, CXXC DNA 结合结构域绿色标注, Coiled-Coil(CC)结构域青色标注, DNA 水解酶催化结构域黄色标注. 数字代表氨基酸残基数. (b) MBD 结构域蛋白抑制基因表达的模式图: 灰色的 DNA 缠绕在粉色的组蛋白八聚体上组成核小体(NCP). MBD 结构域橙色表示, 可识别蓝色圆圈代表的 DNA CpG 甲基化修饰. 共抑制复合体主要包括组蛋白去乙酰化酶 HDAC 和组蛋白甲基化酶 HKMT/SETDB1, 用墨绿色标识. 亮粉色六边形代表组蛋白 H3K9 上的甲基化修饰, 黄色三角形代表组蛋白上的乙酰化修饰. MBD 结构域介导了 DNA 甲基化修饰和组蛋白修饰之间的交互作用(Crosstalk).

MBD4 蛋白与其他 MBD 蛋白差别较大, 是一个核酸水解酶, 主要与 DNA 错配修复密切相关. MBD4 包含 1 个 MBD 结合结构域和 1 个 C 端的糖基化酶结构域, 可识别发生 T:G 错配的 mCpG 位

点, 将错配的胸腺嘧啶 T 碱基切除^[34]. 近期有报道称 MBD4 的糖基化酶结构域具有切除 5hmU 碱基的活性, 完成类似 TDG 的 DNA 主动去甲基化过程^[35-36].

2.1.3 MBD 结构域识别甲基化 DNA 的结构机理

MBD 结构域蛋白具有较高的序列同源性, 三维结构十分保守. MBD 结构域是一种典型的 α/β 夹层结构 (α/β sandwich), 其中一层由 2~4 条 β -strands 组成, 中间的 2 条 β -strands $\beta 2$ 和 $\beta 3$ 较长, 在所有的 MBD 结构域中均保守; 两侧的 2 条 β -strands $\beta 1$ 和 $\beta 4$ 较短, 在有些 MBD 结构域中以 loop 形式存在. 另一层由一个 α 螺旋 $\alpha 1$ 和一个 C 端的发夹 loop (hairpin) 构成, 只有 MBD4 的 MBD 结构域没有折成类似的发夹 loop, 而是伸向了相反的一侧 (图 3a).

2001 年, Shirakawa 课题组报道了人源 MBD1 蛋白的 MBD 结构域与 mCpG DNA 复合物的核磁结构, 揭示了 MBD 结构域识别甲基化 DNA 的结构基础^[37-38]. 如图 3b 所示, MBD 结构域结合在双链 DNA 的大沟处, 非对称地与 DNA 的两条链相互作用. 其中 $\beta 2$ 和 $\beta 3$ 以及它们之间连接的 loop L1 与 DNA 的一条链相互作用, $\beta 4$ 与 $\alpha 1$ 之间的连接 loop L2 以及 $\alpha 1$ 的 N 端区域与 DNA 的另一条链相互作用 (图 3c). MBD 结构域中特异识别 mCpG 位点的残基包括 Val20、Arg22、Tyr34 以及 Arg44 和 Ser45, 主要通过疏水和极性作用力与两侧的 5mC 碱基相互作用. 除了 Val20 外, 其余 4 个残基在所有的 MBD 结构域中均保守, 针对它们

的突变均会影响 MBD 结构域的甲基化 DNA 结合能力. 其中 Tyr34 的侧链苯环上的羟基对特异识别甲基化的 mCpG 碱基对起到至关重要的作用 (图 3c), 可以同两条链上的 mC 碱基第 4 位氮原子形成两对氢键. 将酪氨酸突变为苯丙氨酸去掉侧链的羟基, MBD 结构域不再结合 mCpG DNA.

MeCP2 蛋白的 MBD 结构域结构最早被解析^[39], 但直到 2008 年, 其 MBD 结构域与调控基因 *Bdnf* 启动子 DNA 复合体的晶体结构才被报道^[40]. MeCP2 的 MBD 结构域结合 DNA 的方式与 MBD1 蛋白非常类似, 但它识别 mCpG 位点的残基并不与 5mC 发生直接的相互作用, 而是通过 DNA 周围结合的水分子介导了 MeCP2 对甲基化 DNA 的识别.

对 MBD2、MBD3 和 MBD4 蛋白的 MBD 结构域与 DNA 复合体的结构也都有报道^[41-43], 它们结合 DNA 的方式与其他 MBD 蛋白十分类似, 水分子在二者的相互作用中发挥了重要的作用. 不同的是, MBD2 与 MBD1 对甲基化的 CpG 位点的亲和力要比未甲基化的位点高很多; 而 MBD3 对 5mC 的亲和力仅略高于 5C 和 5hmC; MBD4 对 CpG 位点的识别则更加宽泛, 可结合 5C、5mC、5hmC 和 5hmU 等多种修饰的嘧啶环^[44].

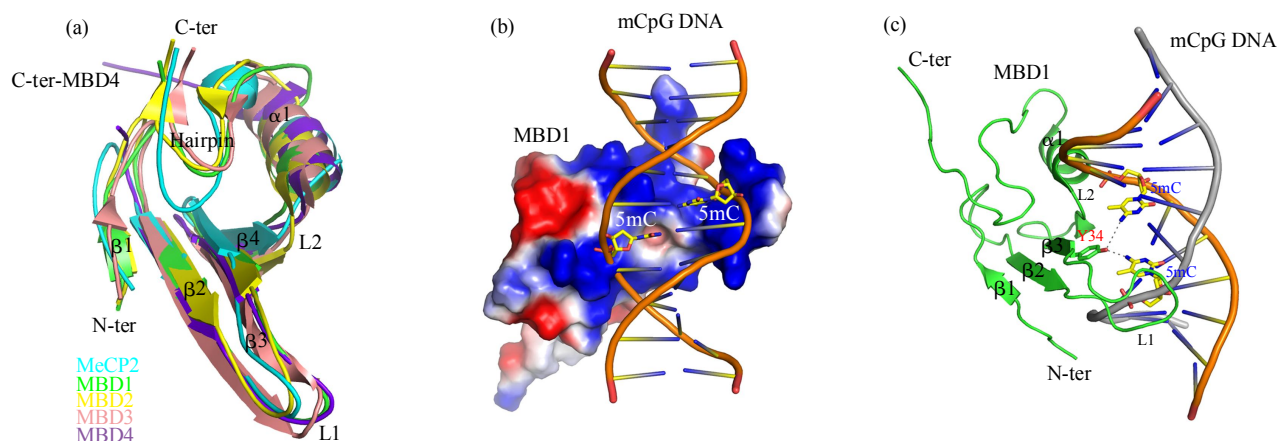


Fig. 3 Structural bases for methyl DNA recognition by MBD proteins

图 3 MBD 结构域蛋白结合甲基化 DNA 的结构机理

(a) MBD 结构域蛋白结构比对图. MeCP2、MBD1~4 的 MBD 结构域 PDB 坐标分别使用 3C2I、1IG4、2KY8、2MB7 和 3VYB 绘图, 颜色分别用青色、绿色、黄色、粉色和紫色表示, 图中显示它们具有十分相似的三维结构. (b) MBD1-MBD 结合甲基化 DNA 的表面电势图. 蓝色代表正电荷碱性区域, 红色代表负电荷酸性区域, 白色为疏水区域. 甲基化 DNA 结合在 MBD1-MBD 的一个富含正电荷的表面, 两侧的 5mC (黄色标注) 均未发生翻转, 结合在 MBD 一个相对疏水的区域. (c) MBD1-MBD 结合甲基化 DNA 的结构基础. MBD 结构域结合在双链 DNA 的大沟处, 非对称地与 DNA 的 2 条链 (分别用橙色和灰色表示) 相互作用. 相互作用主要集中在连接 loop L1 和 L2 位置. Y34 的侧链苯环上的羟基可同 2 条链上的 5mC 碱基第 4 位氮原子形成 2 对氢键.

2.2 SRA 结构域蛋白

SRA 结构域 (SET and RING finger-associated domain) 是另一类可以结合甲基化 DNA 的结构域, 因其相邻位置通常包含组蛋白甲基化酶 SET 结构域或泛素连接酶 RING 结构域而得名 (图 4a)。相对于 MBD 结构域特异识别 CpG 位点的全甲基化修饰, SRA 结构域对甲基化 DNA 的识别更具全能性, 除了 CpG 位点, 也可识别 CHH (其中 H 代表 A, T 或 C 碱基) 位点的甲基化修饰, 以及半甲基化修饰、非对称修饰的 DNA。

2.2.1 植物中的 SRA 结构域蛋白

最早在拟南芥中发现组蛋白 H3K9 甲基化酶

KRYPTONITE (也称为 KYP 或 SUVH4) 可以调控 DNA 甲基化酶 CMT3 的活性^[45]。KYP 蛋白除了包含 C 端的组蛋白甲基化酶催化结构域 SET 外, 还包含一个 N 端相邻的 SRA 结构域。进一步研究发现 KYP 的 SRA 结构域可识别 CHH 位点甲基化的 DNA, 其 H3K9 甲基化酶活性与其甲基化 DNA 结合能力密切相关^[46]。KYP 通过其 SRA 结构域的甲基化 DNA 识别能力将 DNA 甲基化修饰与组蛋白 H3K9 的甲基化修饰联系到一起, 介导了这两种不同的表观遗传修饰在基因沉默和异染色质生成中的交互作用 (crosstalk)。另一类在植物中发现的 SRA 结构域蛋白是拟南芥 VIM 蛋白 (variant in

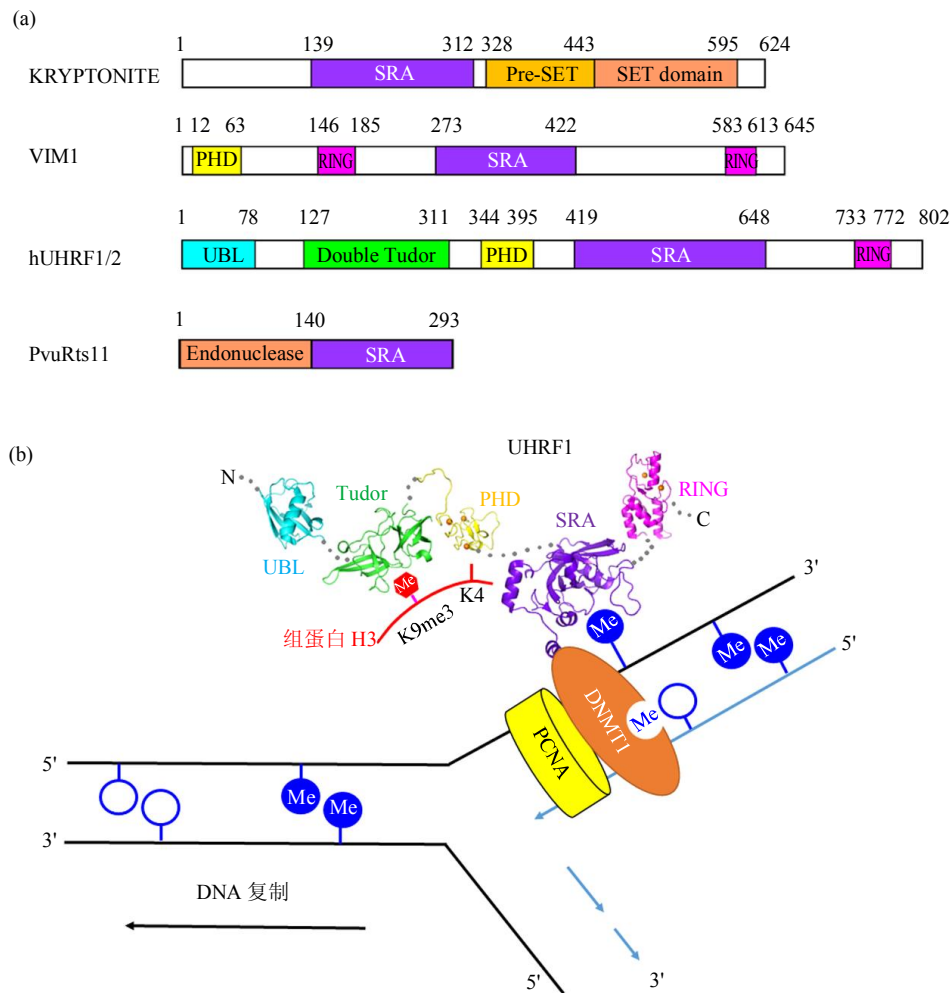


Fig. 4 Domain architecture of SRA proteins and a model of UHRF1-mediated DNA methylation during DNA replication
图 4 SRA 结构域蛋白结构域组成以及维持 DNA 甲基化水平的示意图

(a) SRA 结构域蛋白的结构域组成和分布。紫色代表 SRA 结构域, 橙色代表催化结构域, 黄色代表 PHD finger, 亮粉色代表泛素连接酶 RING 结构域, 青色代表类泛素结构域 UBL, 绿色代表双 Tudor 结构域。(b) UHRF1 蛋白在 DNA 复制过程中维持新生链 DNA 甲基化水平的示意图。蓝色圆圈代表 5CpG 碱基, 空心表示未发生甲基化, 蓝色实心为发生甲基化修饰。UHRF1 由已解析结构的各个结构域串联而成, 颜色同(a)。组蛋白 H3 红色表示, 其甲基化修饰以红色六边形表示。UHRF1 的 SRA 结构域识别 DNA 半甲基化的位点, 招募维持型甲基化酶 DNMT1 完成新生链上的甲基化反应, 并通过其双 Tudor 结构域和 PHD finger 识别组蛋白 H3 上的甲基化修饰, 介导 DNA 甲基化修饰和组蛋白修饰间的交互作用(crosstalk)。

methylation)^[47], 包括高度同源的 VIM1~3 三个成员, 它们均包含 1 个中间的 SRA 结构域、两侧相邻的 2 个 RING 泛素连接酶结构域和 1 个 N 端的 PHD 结构域(图 4a). VIM 蛋白的 SRA 结构域也可识别甲基化的 DNA, 并且对 CpG 位点具有特异性, 在拟南芥基因组 CpG DNA 的甲基化维持以及基因沉默中发挥重要作用.

2.2.2 哺乳动物中的 SRA 结构域蛋白

哺乳动物中同样存在可识别甲基化 DNA 的 SRA 结构域蛋白. UHRF1(ubiquitin-like, containing PHD and RING finger domains 1)是 VIM1 蛋白的同源蛋白, 也称为 Np95 和 ICBP90. 在 DNA 半保留复制中, UHRF1 蛋白可招募维持型 DNA 甲基化酶 DNMT1 至半甲基化 DNA 位点, 完成新合成 DNA 链上的甲基化修饰, 对维持基因组 DNA 的甲基化图谱和水平起到至关重要的作用^[48-49]. UHRF1 蛋白也是一个包含多个结构域的多功能蛋白, 其 N 端的 UBL 结构域是一个类泛素结构域, 双 Tudor 结构域可以识别组蛋白 H3K9 上的三甲基化修饰, 中间的 PHD finger 可识别组蛋白 H3K4 位点, SRA 结构域则可识别半甲基化的 CpG DNA, C 端的 RING 结构域是组蛋白泛素连接酶. UHRF1 蛋白的 SRA 结构域通过识别半甲基化 DNA, 帮助 DNMT1 定位在半甲基化 CpG 位点区域, 完成新生链上的甲基化修饰. 同时, UHRF1 蛋白通过其 Tudor 结构域和 PHD finger 识别组蛋白上的基因沉默信号, 介导 DNA 甲基化修饰和组蛋白修饰间的交互作用(crosstalk), 共同维持异染色质的形成.

2.2.3 SRA 结构域蛋白识别甲基化 DNA 的结构机理

UHRF1 蛋白的 SRA 结构域与半甲基化 CpG DNA 复合体的晶体结构展示了 SRA 结构域蛋白识别 mCpG DNA 的结构特点^[50-52]. SRA 结构域折叠成一个类似马鞍形状, 其主体部分是由 5 个 β -strands 构成的 β 桶, 外周包含 4 段 α 螺旋作为结构组分(图 5a). 与 MBP 结构域不同, SRA 结构域同时结合在 DNA 的大沟和小沟处(图 5b). 有意思的是, 不配对的 5mC 碱基翻转出来结合在 SRA 结构域的一个疏水口袋中, 这种碱基翻转机制(base flipping)通常在 DNA 修饰酶与 DNA 复合物中出现, 可帮助酶定位在待修饰的碱基处^[53]. 两段关键的 loop、Finger loop 和 Thumb loop, 分别结合在 DNA 的大沟和小沟处, 与 mCG:GC 二核苷酸对中的其他三个碱基发生相互作用, 确保了 CpG 位

点的特异性. 其中 Thumb loop 的 V446 残基侧链深入到 DNA 内部, 占据了原本 mC 碱基的位置, 而 Finger loop 中的 NKR 残基主要与互补链上未甲基化的 C 碱基发生极性相互作用, 如果 mC 取代 C, 则会阻碍 Finger Loop 与 DNA 的结合, 因此也就确保了 UHRF1 的 SRA 结构域识别半甲基化 DNA 的特异性.

近期, 关于植物中的 SRA 结构域与甲基化 DNA 识别的结构机理研究也有不少报道. 拟南芥中组蛋白 H3K9 甲基化酶 SUVH5 的 SRA 结构域识别甲基化 DNA 的特异性较差, 可以识别双甲基化、半甲基化以及非 CpG 甲基化的 DNA, 这点与其同源蛋白 KRYPTONITE(也称为 SUVH4)很类似. 在 SUVH5 的 SRA 结构域与多种甲基化修饰 DNA 复合体的晶体结构中(包括全甲基化 CpG、半甲基化 CpG 以及 CHH DNA)^[54], 其 SRA 结构域与 UHRF1 的 SRA 结构域结构十分相似(Dali 比对结果: Z-score 18.8, RMSD 1.8). 不同的是 DNA 分子与 SRA 的结合比例为 1:2, SUVH5 两个 SRA 结构域分别结合在 DNA 双螺旋的两侧, 同时将 mC 和互补链上的 mC/C/G 碱基翻转出来(图 5c). SUVH5-SRA 结合 DNA 的方式与 UHRF1 蛋白十分类似, 翻转的碱基均结合在一个疏水口袋中, 不同的是, SUVH5-SRA 的 Finger loop 不包含 NKR 残基, 该 loop 在结构中呈现一种无规则的状态, 因此和 DNA 相互作用的部位主要集中在 Thumb loop 区域. 类似 UHRF1 的 V446 残基, SUVH5 的一个 Q392 残基的侧链深入到 DNA 中, 占据了翻转碱基原本的位置, 对 DNA 的结合至关重要. 由于缺乏 Finger loop 区域对 CpG 二核苷酸对中的其他三个碱基的识别, 因此 SUVH5 识别 CpG 位点的特异性较低, 对 CHH 甲基化的 DNA 均可结合. 关于为何 SUVH5 蛋白能采用这种双翻转机制的一个合理解释是, 其 Finger loop 的可塑性允许一个 DNA 分子同时结合两个 SRA 分子, 而在 UHRF1 中, 其 Finger loop 折叠成相对稳定的结构结合在 DNA 的大沟处, 如将两个 UHRF1-SRA 分子对接到 SUVH5 结合的 DNA 分子的两侧, 则在 Finger loop 区域会产生空间位置上的冲撞, 无法形成一个稳定的结构(图 5d).

KRYPTONITE/SUVH4 与甲基化 CHH DNA 复合体的晶体结构近期也获得了解析, 其 SRA 结构域结合 mCHH DNA 的方式与 UHRF1 十分类似, 也是单碱基翻转, 包含 NKR 残基、相对稳定的

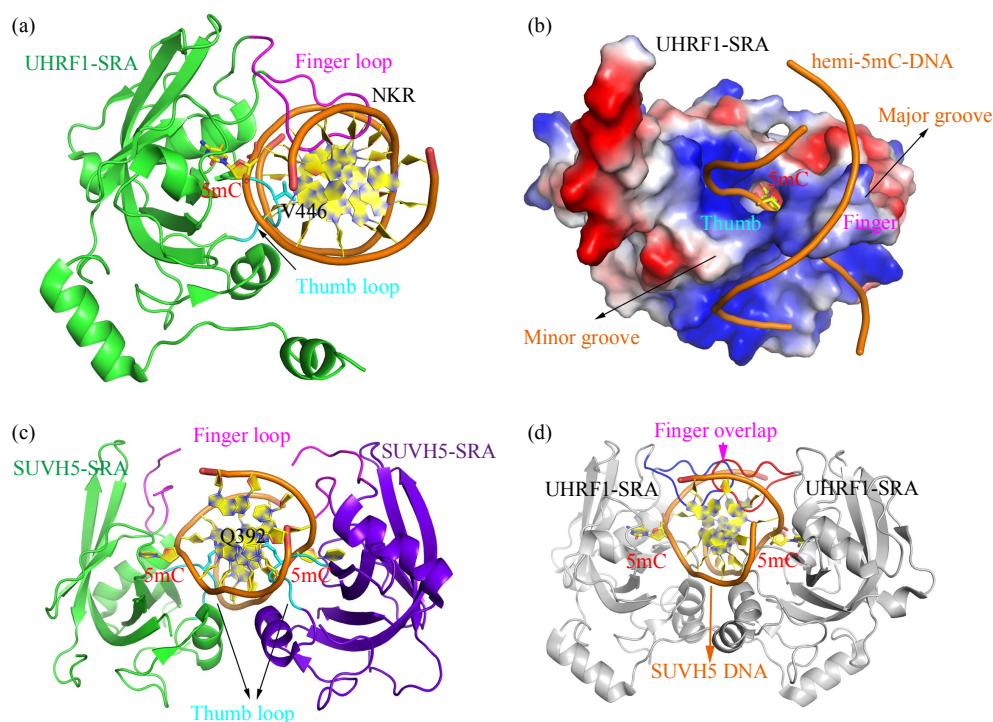


Fig. 5 Structural bases for methyl DNA recognition by SRA domains

图 5 SRA 结构域蛋白识别甲基化 DNA 的结构机理

(a)UHRF1-SRA(绿色)与半甲基化 DNA(橙色)复合物的结构图。UHRF1-SRA 通过 2 个重要的 loop, Finger loop(亮紫色)和 Thumb loop(青色)与 DNA 相互作用。一侧的 5mC 从 DNA 中翻转出来,深入 UHRF1-SRA 中,而 Thumb loop 的 V446 残基的侧链则深入 DNA 中,填补了翻转碱基原本的位置。Finger loop 主要通过 NKR 残基与 DNA 发生直接的相互作用。(b)UHRF1-SRA 与半甲基化 DNA 结合的表面电势图。DNA 结合在 SRA 结构域富含正电荷的表面,SRA 的 Finger loop 和 Thumb loop 分别结合在 DNA 的大沟和小沟处。翻转的 5mC 结合在 SRA 的一个疏水口袋中。(c)植物 SUVH5-SRA 与全甲基化 DNA 的结构图。一个 DNA 分子同时结合了 2 个 SUVH5-SRA,两侧的 5mC 均翻转结合到 SRA 疏水口袋中。两侧的 Q392 残基的侧链占据了翻转碱基的位置。由于 Finger loop 十分柔性,中间密度不可见,SUVH5-SRA 与 DNA 的相互作用主要发生在 Thumb loop 区域。(d)将 UHRF1-SRA 替代 SUVH5-SRA 放入双侧 DNA 结合位置,2 个 UHRF1-SRA 的 Finger loop 相互重叠,空间位置上不允许一个 DNA 分子结合 2 个 UHRF1-SRA 结构域。

Finger loop 与 Thumb loop 同时结合在 DNA 的大沟和小沟处。但 SUVH4 的 Finger loop 中的 NKR 残基与 CpH 二核苷酸对中的其他三个碱基未发生直接的相互作用,因此对 CpG 的选择性也较差^[55]。

细菌中也存在 SRA 结构域蛋白,主要是一大类限制性核酸内切酶,如不动杆菌(*Acinetobacter bau.*)来源的 AbaSI^[56],变形杆菌(*Proteus vulgaris*)来源的 PvuRts1I^[57],分枝杆菌(*Mycobacterium sp.*)来源的 MspJI^[58]以及固氮弧菌(*Azoarcus sp.*)来源的 AspBI^[59]等。它们均包含一个 N 端的核酸内切酶结构域和一个 C 端的 SRA 结构域(图 4a),可识别甲基化或羟甲基化的 DNA(5mC 和 5hmC)位点,完成临近固定位点上的双链 DNA 切割反应。这类核酸内切酶作为工具酶已广泛应用到 5mC 和 5hmC 的单碱基分辨率的 DNA 测序中。关于这类酶的 SRA 结构域识别 DNA 的结构机理研究目前还非常有限,唯

一个 AbaSI 与 DNA 复合体的晶体结构中^[56],SRA 结构域并没有与 DNA 发生紧密的相互作用,Abasi 与 DNA 的相互作用全部发生在 DNA 的磷酸骨架,类似非特异的电荷吸附作用。因此关于此类 SRA 结构域如何特异识别 5mC 和 5hmC 的分子机制还有待进一步的研究和发现。

3 DNA 羟甲基化修饰的识别

如前所述,5hmC 是 TET 蛋白介导的 DNA 主动去甲基化过程中一个主要的中间产物。同时 5hmC 作为一种稳定的表观遗传标记,大量存在于基因组中,特别是在神经系统和胚胎干细胞中的分布明显高于其他组织和细胞^[60]。新一代的测序技术实现了全基因组范围内的 5hmC 定位,发现在人及鼠的 ESC 细胞中 5hmC 多定位在基因的增强子位置和基因本身^[61-62],与基因表达的激活密切相关。

5mC 与 5hmC 只相差一个羟基, 但二者对基因表达的调控呈相反的趋势. 因此关于 5hmC 又是如何被特异识别, 如何上调基因表达的机理引起了很多研究兴趣.

最先围绕可结合 5mC 的 MBD 结构域、UHRF1 的 SRA 结构域以及 TET 蛋白的 CXXC DNA 结合结构域开展了不少特异性筛选工作, 体外结合实验表明, 这些结构域虽然都可结合 5hmC, 但也同时结合 5mC, 甚至结合 5mC 的能力超过 5hmC^[44, 63-66]. 2013 年, 荷兰的 Vermeulen 课题组通过体外 DNA pull-down 的方法, 在 NPC 神经前体细胞中找到了可结合 5hmC DNA 的 UHRF2 蛋白, 后又发现重组的 UHRF2-SRA 结构域可以体外特异地结合 5hmC DNA, 与 5caC 也有一定的结合, 但对 5fC、5mC 和 5C 的结合能力较弱. 同时 UHRF2-SRA 结构域可以提高 TET1 蛋白的催化活性, 特别是以 5hmC 为底物的催化活性^[67].

UHRF2 蛋白是 UHRF1 蛋白的近亲, 二者序列相似度很高, 特别是 SRA 结构域的序列相似性高达 76%, 但二者的功能却并不完全相同. 虽然它们体外均可结合半甲基化 DNA, 并且可与 DNMT1 以及组蛋白甲基化酶 G9a 相互作用, 但 UHRF1 蛋白主要在 ESC 细胞中表达而 UHRF2 蛋白多在已分化的细胞中存在; 同时 UHRF2 不能像 UHRF1 一样, 在 DNA 复制过程中招募 DNMT1 到半甲基化 CpG 位点^[68]. Vermeulen 课题组的发现证明了

UHRF2-SRA 是一个潜在的 5hmC DNA 特异识别蛋白, 但其识别的分子机制需要进一步的结构生物学证据. 2014 年, Xu 课题组报道了 UHRF2-SRA 与包含 5hmC、5mC 和 5C 多种 DNA 复合物的晶体结构. 发现 UHRF2-SRA 与不同修饰 DNA 的结合方式与 UHRF1-SRA 十分类似, 无论是 5hmC、5mC 和 5C 碱基均翻转出来, 结合在类似的疏水口袋中, Val475 残基的侧链深入 DNA 中填补了翻转碱基原本的位置^[69]. 进一步比较 UHRF2-SRA 与 UHRF1-SRA 结合翻转碱基的口袋发现(图 6a), UHRF1-SRA 中一个 Tyr471 残基的侧链与翻转碱基的位置很近, 只能允许较小侧链的 5mC 结合, 而 5hmC 伸出的羟基会在位置上与 Tyr471 的侧链发生冲突, 因此 UHRF1-SRA 不适合结合羟甲基化的 DNA. 而在 UHRF2 中, 相应位置的残基是 Phe495, 其侧链的位置给翻转碱基结合口袋留出了更大的空间, 可以容纳 5C、5mC 乃至 5hmC、5caC 碱基. 这也部分解释了 UHRF2 的 SRA 结构域对 DNA 的选择性比 UHRF1 要差的原因.

分析不同修饰的胞嘧啶在 UHRF2-SRA 结合口袋中的差异, 发现 5hmC 的羟基与 Thr508 的 C 羰基形成的一个额外氢键, 是 UHRF2-SRA 对 5hmC 具有更强亲和力的原因(图 6b). 同时也解释了为何 UHRF2-SRA 对 5caC 也有较强的结合力, 因为 5caC 羧基的羟基集团也可形成类似的额外氢键, 而 5fC、5mC 和 5C 则没有相应的 H 原子与 Thr508

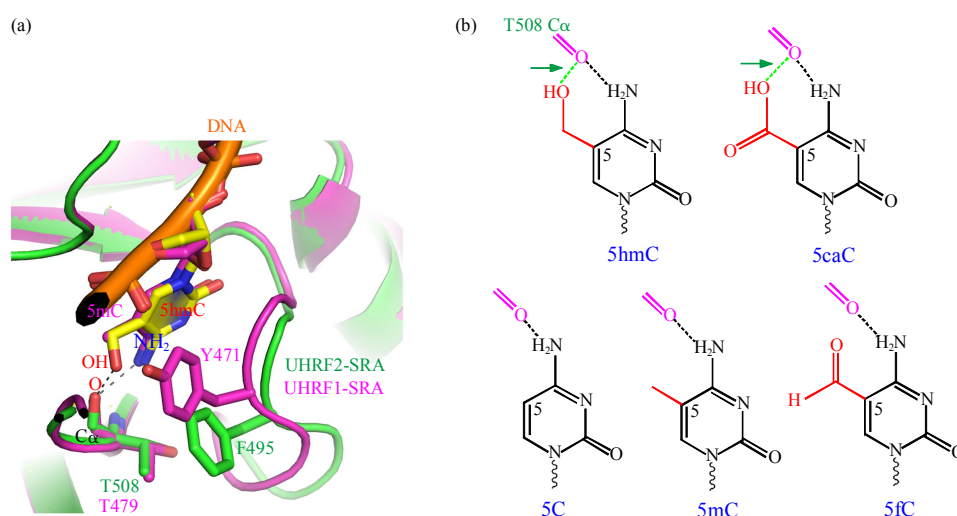


Fig. 6 Structural basis for 5hmC DNA recognition by the SRA domain of UHRF2

图 6 UHRF2-SRA 识别羟甲基化 DNA 的结构机理

(A)UHRF2-SRA(绿色)与 UHRF1-SRA(亮紫色)结合翻转碱基的结构对比图. 5hmC 和 5mC 碱基分别用黄色和紫色表示, UHRF1-SRA 的 Y471 残基与 5hmC 的羟基在位置上有冲突. (B)UHRF2-SRA T508 残基的 C 羰基与 5hmC、5caC 的羟基形成的一对多余氢键(绿色表示), 决定了其对 5hmC 和 5caC 的识别特异性.

的羧基氧原子形成氢键. 同时, 体外结合实验显示了 UHRF2-SRA 更偏好结合双边都修饰的 5hmC DNA, 这点从结构上也可获得解释, 其类似 SUVH5 的柔性 NKR loop 允许两个 SRA 结构域结合在一个 DNA 分子的两侧, 将 5hmC 和互补链上的碱基同时翻转出来(图 5c).

4 展 望

可与 DNA 结合的蛋白质结构域很多, 如 Helix-Turn-Helix、CXXC 结构域、Winged-helix 结构域、Zinc-finger 结构域等, 但可特异结合 5mCpG 甲基化 DNA 的结构域迄今只发现了 MBD 结构域和 SRA 结构域蛋白. 从图 7 可以看出, MBD 结构域蛋白对全甲基化 DNA 的结合能力高于其他修饰形式的 DNA, 特别是 MBD2(MeCP1 复合体中的组分)和 MeCP2 对全甲基化 DNA 的亲合力是非甲基化 DNA 的上百倍. MBD1 的 MBD 结

构域本身对甲基化 DNA 的偏好性也非常高, 但由于 MBD1 还包含 3 个 CXXC DNA 结合结构域, 因此其全长蛋白对甲基化 DNA 的偏好性受到一定影响. MBD3 和 MBD4 的 MBD 结构域对甲基化 DNA 的结合特异性较低, 相对于非甲基化 DNA 只有 4~5 倍的差别, 特别是 MBD4 的 C 端包含一个糖基化酶结构域, 可识别发生 T:G 错配的 mCpG 位点, 因此 MBD4 蛋白对 DNA 的识别特异性很大程度上是通过其糖基化酶结构域实现的; SRA 结构域蛋白对 DNA 修饰的识别具有多样性. 植物中的 VIM 蛋白和哺乳动物中同源的 UHRF1 蛋白均可特异识别半甲基化的 CpG DNA, 而与 UHRF1 高度同源的 UHRF2 蛋白却对半甲基化 DNA 缺乏特异性, 而对全羟甲基化 DNA 有一定的偏好性. 而植物中的 KRYPTONITE、SUVH5 的 SRA 结构域不但对 DNA 修饰形式(甲基化, 半甲基化)没有很强的选择性, 对 CpG 位点的特异性也不高, 可识别 CHH 多种位点处的甲基化修饰.

由以上分析不难看出, 包含 MBD 及 SRA 结构域的蛋白质对不同修饰形式 DNA 的识别特异性均与其生物学功能密切相关, 如 UHRF1 在 DNA 半保留复制过程中, 对新生链上甲基化修饰的维持需要其能够特异识别半甲基化的 DNA, 而高度同源的 UHRF2 蛋白不具备这样的功能, 因此也就不需要对半甲基化 DNA 产生偏好性. UHRF2 是目前发现的对羟甲基化 DNA 有识别特异性的蛋白, 结构生物学研究也揭示了其识别特异性的分子机制, 当然是否存在对羟甲基化 DNA 有很高特异性的识别蛋白仍是未来值得探索的方向.

另外值得一提的是, 无论是 MBD 结构域还是 SRA 结构域蛋白, 它们发挥生物学功能的方式是类似的, 都是介导了 DNA 甲基化修饰与组蛋白修饰间的交互作用(crosstalk). a. 首先它们通过甲基化 DNA 结合结构域识别 DNA 甲基化修饰, 然后通过不同的方式与组蛋白修饰发生联系, 如 MeCP2 和 MBD2 蛋白均可招募组蛋白去乙酰化酶 HDAC, 介导 DNA 甲基化与组蛋白去乙酰化间的协同作用; b. MBD1 蛋白则通过与组蛋白甲基化酶 SETDB1 的作用, 完成与 H3K9 甲基化修饰间的协同作用; c. 植物中的 KRYPTONITE 和 SUVH5 本身就是组蛋白甲基化催化酶, 可以自身完成 DNA 的识别和组蛋白 H3K9 上的甲基化修饰; d. UHRF1 蛋白除可与 DNA 甲基化酶 DNMT1、组蛋白 H3K9 甲基化酶 G9a 直接相互作用外, 还

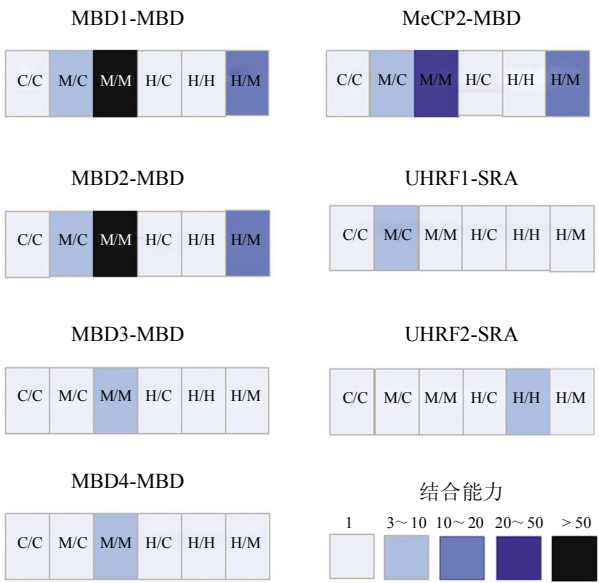


Fig. 7 Degree of specific methyl DNA binding abilities of various proteins toward different forms of modified CpG DNA

图 7 甲基化 DNA 结合蛋白识别不同修饰形式 5CpG DNA 的特异性

其中 C/C 表示未甲基化 DNA, M/C 表示半甲基化 DNA, M/M 表示全甲基化 DNA, H/C 表示半羟甲基化 DNA, H/H 表示全羟甲基化 DNA, H/M 表示一条链甲基化、另一条链羟甲基化 DNA. 图中方框不同颜色表示不同的 DNA 能力, 以未甲基化 DNA 的结合能力为基数 1, 其他修饰 DNA 的结合能力以倍数计算. 绘图的数据依据来自参考文献中的荧光偏振实验结果, 实验中 MBD 蛋白均使用其 MBD 结构域^[44], SRA 结构域使用的是 Tudor-PHD-SRA 串联结构域^[69].

包含多个组蛋白修饰识别结构域 Tudor 和 PHD 以及 DNA 识别结构域 SRA, 更紧密地将 DNA 甲基化修饰和组蛋白上的修饰联系在一起。我们知道组蛋白去乙酰化和 H3K9 上的三甲基化修饰都是组成型异染色质(constitutive heterochromatin)生成的标志, DNA 甲基化修饰通过与这些异染色质相关的组蛋白修饰联系在一起, 共同维持异染色质的生成与基因表达的沉默。这也为 DNA 甲基化修饰如何调控基因表达给出了一种很好的阐释。

参 考 文 献

- [1] Bird A. Perceptions of epigenetics. *Nature*, 2007, **447** (7143): 396–398
- [2] Guo H, Zhu P, Yan L, *et al.* The DNA methylation landscape of human early embryos. *Nature*, 2014, **511**(7511): 606–610
- [3] Jiang L, Zhang J, Wang J J, *et al.* Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Cell*, 2013, **153**(4): 773–784
- [4] Wang L, Zhang J, Duan J, *et al.* Programming and inheritance of parental DNA methylomes in mammals. *Cell*, 2014, **157**(7): 1735–1735
- [5] Bird A P, Wolffe A P. Methylation-induced repression—belts, braces, and chromatin. *Cell*, 1999, **99**(5): 451–454
- [6] Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet*, 1998, **19**(3): 219–220
- [7] Okano M, Bell D W, Haber D A, *et al.* DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, 1999, **99**(3): 247–257
- [8] Bestor T H, Ingram V M. Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proc Natl Acad Sci USA*, 1983, **80**(18): 5559–5563
- [9] Monk M, Boubelik M, Lehnert S. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development*, 1987, **99**(3): 371–382
- [10] Carlson L L, Page A W, Bestor T H. Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting. *Genes Dev*, 1992, **6**(12B): 2536–2541
- [11] Weiss A, Keshet I, Razin A, *et al.* DNA demethylation *in vitro*: involvement of RNA. *Cell*, 1996, **86**(5): 709–718
- [12] Bhattacharya S K, Ramchandani S, Cervoni N, *et al.* A mammalian protein with specific demethylase activity for mCpG DNA. *Nature*, 1999, **397**(6720): 579–583
- [13] Iyer L M, Tahiliani M, Rao A *et al.* Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle*, 2009, **8**(11): 1698–1710
- [14] Tahiliani M, Koh K P, Shen Y, *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, 2009, **324**(5929): 930–935
- [15] Ito S, D'Alessio A C, Taranova O V, *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, 2010, **466**(7310): 1129–1133
- [16] Ito S, Shen L, Dai Q, *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, 2011, **333**(6047): 1300–1303
- [17] He Y F, Li B Z, Li Z, *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, 2011, **333**(6047): 1303–1307
- [18] Watt F, Molloy P L. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev*, 1988, **2**(9): 1136–43.
- [19] Bell A C, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, 2000, **405**(6785): 482–485
- [20] Boyes J, Bird A. DNA methylation inhibits transcription indirectly *via* a methyl-CpG binding protein. *Cell*, 1991, **64**(6): 1123–1134
- [21] Meehan R R, Lewis J D, McKay S, *et al.* Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell*, 1989, **58**(3): 499–507
- [22] Cohen S, Gabel H W, Hemberg M, *et al.* Genome-wide activity-dependent MeCP2 phosphorylation regulates nervous system development and function. *Neuron*, 2011, **72**(1): 72–85
- [23] Chen R Z, Akbarian S, Tudor M, *et al.* Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nat Genet*, 2001, **27**(3): 327–331
- [24] Guy J, Hendrich B, Holmes M, *et al.* A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat Genet*, 2001, **27**(3): 322–326
- [25] Amir R E, Van den Veyver I B, Wan M, *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet*, 1999, **23**(2): 185–188
- [26] Jones P L, Veenstra G J, Wade P A, *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet*, 1998, **19**(2): 187–191
- [27] Nan X, Ng H H, Johnson C A, *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, 1998, **393**(6683): 386–389
- [28] Chen W G, Chang Q, Lin Y, *et al.* Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. *Science*, 2003, **302**(5646): 885–889
- [29] Martinowich K, Hattori D, Wu H, *et al.* DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science*, 2003, **302**(5646): 890–893
- [30] Sarraf S A, Stancheva I. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol Cell*, 2004, **15**(4): 595–605
- [31] Wade P A, Geggion A, Jones P L, *et al.* Mi-2 complex couples

- DNA methylation to chromatin remodelling and histone deacetylation. *Nat Genet*, 1999, **23**(1): 62–66
- [32] Zhang Y, Ng H H, Erdjument-Bromage H, *et al.* Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev*, 1999, **13** (15): 1924–1935
- [33] Saito M, Ishikawa F. The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. *J Biol Chem*, 2002, **277**(38): 35434–35439
- [34] Wu P, Qiu C, Sohail A, *et al.* Mismatch repair in methylated DNA. Structure and activity of the mismatch-specific thymine glycosylase domain of methyl-CpG-binding protein MBD4. *J Biol Chem*, 2003, **278**(7): 5285–5291
- [35] Hashimoto H, Zhang X, Cheng X. Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: structural basis and implications for active DNA demethylation. *Nucleic Acids Res*, 2012, **40**(17): 8276–8284
- [36] Moréra S, Grin I, Vigouroux A, *et al.* Biochemical and structural characterization of the glycosylase domain of MBD4 bound to thymine and 5-hydroxymethyluracil-containing DNA. *Nucleic Acids Res*, 2012, **40**(19): 9917–9926
- [37] Ohki I, Shimotake N, Fujita N, *et al.* Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell*, 2001, **105**(4): 487–497
- [38] Ohki I, Shimotake N, Fujita N, *et al.* Solution structure of the methyl-CpG-binding domain of the methylation-dependent transcriptional repressor MBD1. *EMBO J*, 1999, **18** (23): 6653–6661
- [39] Wakefield R I, Smith B O, Nan X, *et al.* The solution structure of the domain from MeCP2 that binds to methylated DNA. *J Mol Biol*, 1999, **291**(5): 1055–1065
- [40] Ho K L, McNae I W, Schmiedeberg L, *et al.* MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol Cell*, 2008, **29**(4): 525–531
- [41] Scarsdale J N, Webb H D, Ginder G D, *et al.* Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic Acids Res*, 2011, **39**(15): 6741–6752
- [42] Cramer J M, Scarsdale J N, Walavalkar N M, *et al.* Probing the dynamic distribution of bound states for methylcytosine-binding domains on DNA. *J Biol Chem*, 2014, **289**(3): 1294–1302
- [43] Manvilla B A, Maiti A, Begley M C, *et al.* Crystal structure of human methyl-binding domain IV glycosylase bound to abasic DNA. *J Mol Biol*, 2012, **420**(3): 164–175
- [44] Hashimoto H, Liu Y, Upadhyay A K, *et al.* Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res*, 2012, **40**(11): 4841–4849
- [45] Jackson J P, Lindroth A M, Cao X, *et al.* Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature*, 2002, **416**(6880): 556–560
- [46] Johnson L M, Bostick M, Zhang X, *et al.* The SRA methylcytosine-binding domain links DNA and histone methylation. *Curr Biol*, 2007, **17**(4): 379–384
- [47] Woo H R, Dittmer T A, Richards E J. Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in *Arabidopsis*. *PLoS Genet*, 2008, **4**(8): e1000156
- [48] Bostick M, Kim J K, Estève P O, *et al.* UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science*, 2007, **317**(5845): 1760–1764
- [49] Sharif J, Muto M, Takebayashi S, *et al.* The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*, 2007, **450**(7171): 908–912
- [50] Arita K, Ariyoshi M, Tochio H, *et al.* Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature*, 2008, **455**(7214): 818–821
- [51] Avvakumov G V, Walker J R, Xue S, *et al.* Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature*, 2008, **455**(7214): 822–825
- [52] Hashimoto H, Horton J R, Zhang X, *et al.* The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature*, 2008, **455**(7214): 826–829
- [53] Roberts R J. On base flipping. *Cell*, 1995, **82**(1): 9–12
- [54] Rajakumara E, Law J A, Simanshu D K, *et al.* A dual flip-out mechanism for 5mC recognition by the *Arabidopsis* SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation *in vivo*. *Genes Dev*, 2011, **25**(2): 137–152
- [55] Du J, Johnson L M, Groth M, *et al.* Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. *Mol Cell*, 2014, **55**(3): 495–504
- [56] Horton J R, Borgaro J G, Griggs R M, *et al.* Structure of 5-hydroxymethylcytosine-specific restriction enzyme, AbaSI, in complex with DNA. *Nucleic Acids Res*, 2014, **42**(12): 7947–7959
- [57] Kazrani A A, Kowalska M, Czapinska H, *et al.* Crystal structure of the 5hmC specific endonuclease PvuRts1I. *Nucleic Acids Res*, 2014, **42**(9): 5929–5936
- [58] Horton J R, Mabuchi M Y, Cohen-Karni D, *et al.* Structure and cleavage activity of the tetrameric MspJI DNA modification-dependent restriction endonuclease. *Nucleic Acids Res*, 2012, **40**(19): 9763–9773
- [59] Horton J R, Nugent R L, Li A, *et al.* Structure and mutagenesis of the DNA modification-dependent restriction endonuclease AspBHI. *Sci Rep*, 2014, **4**: 4246
- [60] Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, 2009, **324**(5929): 929–930
- [61] Yu M, Hon G C, Szulwach K E, *et al.* Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell*, 2012, **149**(6): 1368–1380
- [62] Stroud H, Feng S, Morey Kinney S, *et al.* 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol*, 2011, **12**(6): R54
- [63] Xu Y, Xu C, Kato A, *et al.* Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for *Xenopus* eye and

- neural development. *Cell*, 2012, **151**(6): 1200–1213
- [64] Mellén M, Ayata P, Dewell S, *et al.* MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell*, 2012, **151**(7): 1417–1430
- [65] Frauer C, Hoffmann T, Bultmann S, *et al.* Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS One*, 2011, **6**(6): e21306
- [66] Yildirim O, Li R, Hung J H, *et al.* Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell*, 2011, **147**(7): 1498–1510
- [67] Spruijt C G, Gnerlich F, Smits A H, *et al.* Dynamic readers for 5-(hydroxy) methylcytosine and its oxidized derivatives. *Cell*, 2013, **152**(5): 1146–1159
- [68] Zhang J, Gao Q, Li P, *et al.* S phase-dependent interaction with DNMT1 dictates the role of UHRF1 but not UHRF2 in DNA methylation maintenance. *Cell Res*, 2011, **21**(12): 1723–1739
- [69] Zhou T, Xiong J, Wang M, *et al.* Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. *Mol Cell*, 2014, **54**(5): 879–886

Structural Basis for Methyl DNA Readout by MBD and SRA Domains*

YANG Na, XU Rui-Ming**

(National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China)

Abstract DNA methylation at the 5 position of cytosine is one of the most important epigenetic modifications in vertebrate, which regulate gene expression in conjunction with specific protein readers to readout the methyl marks. This review focuses on two major classes of methyl DNReaders, MBD and SRA domains, and discusses their structural bases and molecular mechanisms of recognizing different types of DNA modification marks.

Key words readout of DNA modification, 5-methylcytosine, MBD domain, SRA domain, transcriptional regulation, structural mechanism

DOI: 10.3724/SP.J.1206.2014.00284

* This work was supported by grants from The National Natural Science Foundation of China (31370734, 31430018) and National Basic Research Program of China (2015CB856202).

**Corresponding author.

Tel: 86-10-64888797, E-mail: rmxu@sun5.ibp.ac.cn

Received: October 8, 2014

Accepted: October 11, 2014