

组蛋白甲基化的阅读器识别机制研究进展

赵帅, 苏晓楠, 李元元, 李海涛

清华大学医学院基础医学系, 北京 100084

摘要 组蛋白甲基化修饰对遗传信息解读有着重要影响, 是表观遗传调控的主要机制之一。组蛋白甲基化可以被一类称作“阅读器”的结构域所特异识别并介导下游生物学事件。本文综述了目前已知的组蛋白甲基化阅读器(包括“皇室家族”成员、PHD 锌指及BAH等结构域)的结构特征及其对于甲基化修饰位点和程度特异性识别的分子基础。另外, 探讨了表观遗传修饰调控中的组合识别、修饰对话等概念与机制。

关键词 组蛋白阅读器; 组蛋白甲基化; 表观遗传调控; 组合识别; 修饰对话; 晶体结构

中图分类号 Q75

文献标志码 A

doi 10.3981/j.issn.1000-7857.2015.08.016

Research advances in the mechanism of histone methylation recognition by reader modules

ZHAO Shuai, SU Xiaonan, LI Yuanyuan, LI Haitao

Department of Basic Medical Sciences, School of Medicine, Tsinghua University, Beijing 100084, China

Abstract As a major mechanism for the epigenetic regulation, the histone methylation has a crucial impact on the decoding of the genetic information. The histone methylation can be recognized by a class of so-called “reader” modules to mediate the downstream functional outcomes. In this paper, we review the structural aspects of the reported histone methylation readers (e.g. “royal family” members, PHD finger, BAH), and illustrate the molecular basis underlying the site- and state-specific readouts of the histone methylation. This review also covers the concepts of the combinatorial readout and the modification crosstalk in the epigenetic regulation.

Keywords histone reader; histone methylation; epigenetic regulation; combinatorial readout; modification crosstalk; crystal structure

真核生物中基因组DNA以染色质形式存在。核小体是构成染色质的基本单位。核小体及高级染色质结构的形成一方面有效储存和保护了DNA序列所蕴含的遗传信息;另一方面作为基因组DNA的具体存在方式,染色质结构也成为各种需要接触DNA的细胞过程(如转录、复制、损伤修复等)的天然障碍,从而使染色质成为了一个重要的遗传信息表达调控平台。染色质凝聚、松弛结构的形成,以及开放、闭合等状态的转换提供了一种超越DNA序列本身的调控机制,即表观

遗传调控。

表观遗传调控机制涉及到组蛋白修饰、组蛋白变体、DNA甲基化、非编码RNA及染色质重塑等几个方面,是当前生命科学中的一个热点。其中组蛋白翻译后修饰有很多类型,包括甲基化(me)、乙酰化(ac)、丙酰化(pr)、丁酰化(bt)、巴豆酰化(cr)、甲酰化(fo)、磷酸化(ph)、泛素化(ub)、SUMO化、ADP核糖化和生物素化(bio)等。上述种种修饰可以发生在组蛋白不同氨基酸类型或组蛋白不同位点上,并可

收稿日期:2014-11-21;修回日期:2015-01-08

基金项目:国家自然科学基金面上项目(31270763);教育部新世纪优秀人才支持计划项目

作者简介:赵帅,博士研究生,研究方向为表观遗传调控的分子结构机理,电子信箱:zhao-s13@mails.tsinghua.edu.cn;李海涛(通信作者),教授,研究方向为表观遗传调控的分子结构机理,电子信箱:lht@tsinghua.edu.cn

引用格式:赵帅,苏晓楠,李元元,等.组蛋白甲基化的阅读器识别机制研究进展[J].科技导报,2015,33(8):94-100.

以成组或成簇地分布于染色质不同区段,被认为构成一类超越基因序列的“组蛋白密码”^[1,2],控制着细胞内遗传信息的组织层次,在基因表达以及细胞分化与发育调控中发挥着重要的作用。

在众多组蛋白修饰中,组蛋白的甲基化修饰有着重要的调控作用,并且其修饰方式更加复杂。甲基化可以发生在赖氨酸或精氨酸残基,赖氨酸残基可以发生单甲基化(Kme1)、二甲基化(Kme2)或三甲基化(Kme3)修饰,而精氨酸残基则可以发生单甲基化修饰(Rme1)、对称二甲基化修饰(Rme2s)或非对称二甲基化(Rme2a)修饰(图1)。对于组蛋白不同位点的甲基化修饰进一步增强了这一修饰的复杂性。

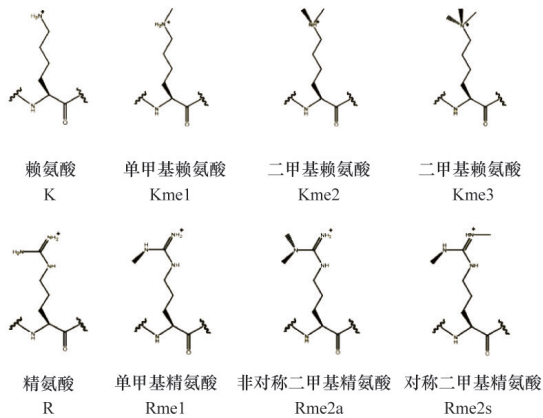


图1 组蛋白赖氨酸及精氨酸甲基化修饰
Fig. 1 Methylation types of histone lysine and arginine residues

为了完成精密的表观遗传调控,围绕“组蛋白密码”细胞进化出一系列修饰酶系统来完成特定组蛋白修饰的建立和消除。产生特定组蛋白修饰的酶称作“书写器”(writer),如组蛋白乙酰化酶(HAT),甲基化酶(HKMT)等;把功能相反并特异性去除特定组蛋白修饰的酶称作“擦除器”(eraser),如各种组蛋白去乙酰化酶(HDAC),去甲基化酶(HKDM)等;同样

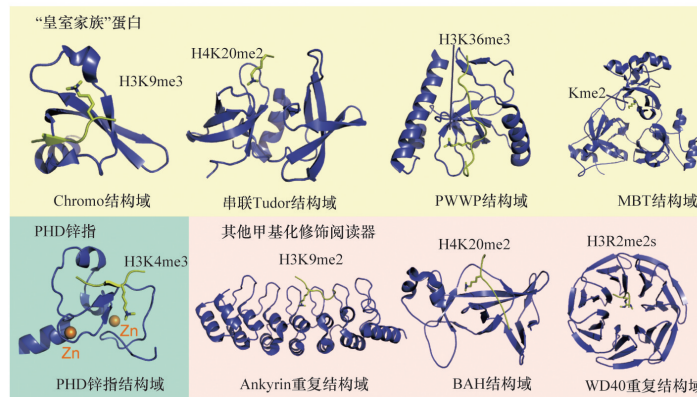
地,细胞内还存在一大类被称作“阅读器”(reader)的蛋白或结构域,可以特异地识别各种不同类型的组蛋白修饰,把特定组蛋白修饰与特定生物学功能后果偶联起来。以上组蛋白修饰调控因子的协同作用控制着诸如转录激活、转录沉默、DNA复制和修复、有丝分裂等重要细胞生理学过程,是表观遗传调控的重要生化和物质基础。

随着表观遗传学研究的深入,大量组蛋白阅读器被发现。目前鉴定出的参与特定组蛋白修饰识别的阅读器结构域类型已逾20多种,包括“皇室家族”(royal family)成员(如Chromo, Tudor, MBT, PWWP, Agenet, Spin/Ssty等)、PHD锌指、CW锌指、ADD锌指、Ankyrin重复、WD40重复、BAH、Bromo、双PHD、双PH、YEATS、BIR、BRCT和14-3-3结构域等^[3],这些结构域可以识别对应的甲基化、乙酰化和磷酸化等修饰类型。本文将以组蛋白甲基化修饰的阅读器识别为重点,从结构生物学角度总结已知组蛋白甲基化阅读器类型,及甲基化位点和程度特异性识别的分子机制,并探讨表观遗传调控中组合识别和修饰对话等概念与机制。希望通过对组蛋白甲基化阅读器的类型及识别机制介绍使读者进一步认识阅读器蛋白作为一类重要的衔接因子(adaptor)在甲基化修饰介导的表观遗传调控中的重要作用,以及基于结构的蛋白改造和药物发现。

1 组蛋白甲基化识别结构域

1.1 “皇室家族”成员

组蛋白甲基化修饰的阅读器主要包括“皇室家族”成员、PHD锌指、CW锌指、ADD锌指、Ankyrin重复、BAH和WD40重复等(图2)。其中,“皇室家族”成员是最早被鉴定也是研究最为深入的一类组蛋白甲基化阅读器,主要包括Chromo结构域、Tudor结构域、PWWP结构域、MBT结构域及植物中特有的Agenet结构域等。“皇室家族”成员拥有3~4个反平行 β 折叠片组成的 β 桶核心,被认为从同一个祖先蛋白结构域进化而来^[4]。



阅读器蛋白使用蓝色表示,组蛋白多肽使用黄色表示。蛋白质数据库(PDB)坐标依次为1KNE(HP1 Chromo结构域),2IG0(53BP1 串联Tudor结构域),2X4W(BRPF1 PWWP结构域),2RHX(L3MBTL1 MBT结构域),2F6J(BPTF PHD锌指),3B95(GLP Ankyrin重复结构域),4DOW(ORC1 BAH结构域),4A7J(WDR5 WD40结构域)

图2 组蛋白甲基化修饰阅读器
Fig. 2 Structure gallery of histone methylation reader modules

Chromo 结构域是最早被鉴定出的组蛋白甲基化修饰阅读器, HP1 的 Chromo 结构域可以特异地识别组蛋白 H3 第九位赖氨酸三甲基化(H3K9me3)修饰, 而不能识别 H3 第四位赖氨酸三甲基化(H3K4me3)^[5]。HP1 Chromo 结构域和 H3K9me3 多肽的复合物结构表明, H3 多肽形成了一条 β 折叠片填入 Chromo 结构域, 其中甲基化的 H3K9 残基插入了一个由三个芳香族残基组成的芳香笼(aromatic cage)中^[6]。组蛋白赖氨酸甲基化并不改变赖氨酸侧链的正电荷, 从而组蛋白 H3K9me3 与 Chromo 结构域的芳香族残基可以形成阳离子- π 相互作用(cation- π interaction), 是介导阅读器与赖氨酸甲基化修饰识别的主要作用力。除了单独的 Chromo 结构域可以识别组蛋白的甲基化修饰外, 双 Chromo 结构域(double chromodomain)也被发现可以识别甲基化赖氨酸修饰。在人 CHD1 双 Chromo 结构域同 H3K4me3 多肽的复合物中, 甲基化的赖氨酸残基被两个芳香族残基组成的芳香笼所识别^[7]。

Tudor 或 Tudor-like 结构域不仅可以识别甲基化赖氨酸, 也可以被用于识别甲基化精氨酸残基, 这一能力是其他“皇室家族”成员所不具有的^[8]。两个 Tudor 结构域可以同时存在于一个蛋白, 并且共同配合识别甲基化的赖氨酸残基。例如, JMJD2A 的双 Tudor 结构域由两个交错的 Tudor 结构域所组成, 该双 Tudor 结构域可以识别 H3K4me3^[9]。其识别口袋由位于第 2 个 Tudor 结构域上的 3 个芳香族残基组成。相比于 JMJD2A 中相互交错的两个 Tudor 结构域, 53BP1 的串联 Tudor 结构域中两个 Tudor 结构域则彼此相对独立。这两个 Tudor 结构域共同组成了一个界面来识别 H4K20 二甲基化修饰多肽, 这一识别过程涉及了 4 个芳香族残基组成的芳香笼和一个天冬氨酸残基^[10]。

PWWP 结构域因为含有一个保守的“脯氨酸-色氨酸-色氨酸-脯氨酸”(残基单字母缩写为“P-W-W-P”)序列而得名。同其他“皇室家族”的成员一样, PWWP 结构域也通过芳香笼来识别甲基化的赖氨酸修饰。同时, PWWP 结构域还发现有同 DNA 相互作用的能力, 暗示其可能还参与了 DNA 介导的表观调控现象^[11]。

MBT 结构域对于赖氨酸甲基化识别的一个特点是低甲基化状态(单甲基化或双甲基化)特异性。MBT 结构域对于低甲基化状态的赖氨酸残基的识别不明显依赖于其周围的序列, 但不能识别三甲基化修饰的赖氨酸残基。对于 L3MBTL1 的结构研究表明, MBT 结构域对于低甲基化赖氨酸残基的识别是通过口袋插入(cavity insertion)的方式来实现的, 由于三甲基化的基团存在较大的空间位阻(steric clash), 因此不能被 MBT 结构域所识别^[12]。

1.2 PHD 锌指蛋白

PHD 锌指结构域是不同于“皇室家族”蛋白的一类全新的组蛋白甲基化阅读器, 并且在人体内广泛存在, 人类基因组中编码了 200 余个 PHD 锌指结构域, 隶属于 90 多种蛋白, 其中多数与基因转录及染色质状态的调控有关。PHD 锌指蛋白有“Cys₄HisCys₃”的锌指序列特征, 其对于组蛋白修饰的

识别有很广的范围, 可以识别非修饰的赖氨酸或精氨酸残基, 二甲基化或三甲基化的赖氨酸残基以及乙酰化的赖氨酸残基^[13]。

PHD 锌指蛋白在一级序列上相似性不高, 但所有的 PHD 锌指蛋白均含有一个由两条反平行 β 折叠片组成的核心。虽然 PHD 锌指蛋白有着相同的核心区域, 但组蛋白或非组蛋白多肽可以结合到不同的 PHD 锌指表面, 如 β 1 表面、 β 2 表面、 α 1 表面和 N 表面^[13]。另外, PHD 锌指蛋白还可以与 GATA-like 锌指(GATA-like finger)融合形成 ADD 结构域, ADD 结构域可以识别组蛋白 H3 非修饰的 K4(通过 PHD 锌指结构域)和三甲基化修饰的 K9(通过 GATA-like 锌指)^[14]。

PHD 锌指蛋白识别甲基化赖氨酸残基的口袋同“皇室家族”成员类似, 也通过芳香笼与甲基化赖氨酸残基间的阳离子- π 相互作用实现。但是有意思的是, 在 ATRX 蛋白的 ADD 结构域中, 三甲基化的赖氨酸残基是通过一个非典型的“极性口袋”被识别的。这一口袋仅有一个酪氨酸残基参与了阳离子- π 相互作用, 口袋的其他组成氨基酸同三甲基化的赖氨酸残基形成了一个非常规的“C-H...O”氢键网络^[14]。

1.3 其他组蛋白甲基化修饰识别蛋白

除了“皇室家族”成员与 PHD 锌指蛋白外, 还有一些结构域也相继被发现可以识别甲基化修饰的赖氨酸残基。G9a 和 GLP 的 ankyrin 结构域被发现可以识别单甲基化和二甲基化的赖氨酸残基^[15]。ORC1 的 BAH 结构域对 H4K20 的二甲基化修饰有特异性的识别^[16]。WDR5 的 WD40 结构域则可以识别 H3R2 的对称二甲基化修饰^[17]。

GLP 含有 6 个 ankyrin 重复(Ankyrin repeats), 其中组蛋白 H3 多肽结合在了 4 个和第 5 个 ankyrin 重复之间。Ankyrin 结构域也是通过一个芳香笼识别甲基化的赖氨酸残基, 识别口袋由 3 个色氨酸残基和一个酸性残基组成。由于三甲基化的赖氨酸残基存在空间位阻以至于不能插入识别口袋, 因此 GLP 的 ankyrin 结构域只能识别单甲基化和二甲基化的赖氨酸残基。GLP 和 G9a 蛋白本身是 H3K9 甲基化的书写器, 这表明同一蛋白可以同时产生和识别某一特定的组蛋白修饰^[15]。ORC1 是参与到 DNA 复制中的一个蛋白, 其 BAH 结构域被发现可以识别 H4K20me2。BAH 结构域识别甲基化的赖氨酸残基也是通过一个芳香族的口袋来实现^[16]。WDR5 的 WD40 结构域中心口袋由 3 个苯丙氨酸残基和 1 个丝氨酸残基组成, 在对于非修饰的精氨酸残基的识别过程中, 精氨酸残基的疏水部分由两个苯丙氨酸残基稳定, 两个结构水则参与了与胍基的识别。而对于对称二甲基化修饰的精氨酸残基来说, 一个水分子被二甲基化的侧链替换掉, 并且整个侧链更靠近口袋底部的苯丙氨酸残基从而获得了更强的相互作用^[17]。

2 组蛋白甲基化位点及程度的特异性识别

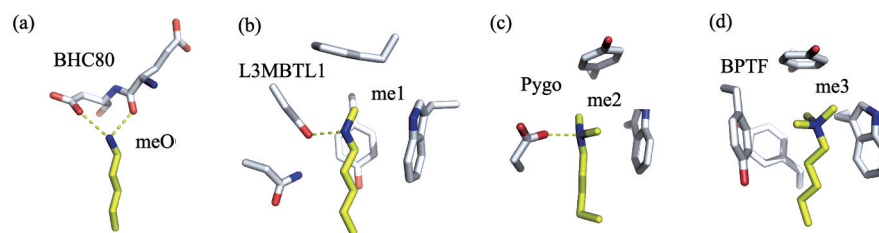
甲基化修饰是最为复杂的组蛋白翻译后修饰之一, 不同的赖氨酸残基位点可以发生甲基化修饰(如 H3K4, H3K9 和 H3K36 等), 同一位点的赖氨酸残基也可以发生不同程度的

甲基化(如 H3K4 可以发生单甲基化、二甲基化或三甲基化)。令人惊奇的是,对于不同位点甚至不同程度的甲基化状态,组蛋白的阅读器可以实现特异性的识别。对于阅读器和组蛋白多肽复合物的结构生物学研究揭示了这些特异性识别的分子机制。

对于 BPTF 的 PHD 锌指和 H3K4me3 多肽的复合物结构研究表明, BPTF 对于 H3K4 这一位点有序列特异性的识别^[18]。当复合物形成时,组蛋白 H3 多肽被诱导形成 β 折叠片并占据了 BPTF 的 PHD 锌指的表面沟(surface groove),这建立了主链层次的相互作用。在此基础上,三组主要的相互作用介导了 H3K4me3 的特异性识别。首先, H3 多肽的 N 末端通过 Loop2 的 A45-D49 的氢键作用被束缚。另外, H3R2 和 H3K4me3 的两个长侧链分别插入了 PHD 锌指中两个临近的口袋。H3R2 的识别口袋是一个酸性的极性口袋,可以通过盐桥和氢键介导同精氨酸残基的识别; H3K4me3 的识别口袋则是经典的芳香笼,通过阳离子- π 和疏水相互作用介导同 H3K4me3 的识别(图 3(d))。这两个识别口袋由一个色氨酸残基隔开,因此要求 R2 和 K4me3 之间需要有一个氨基酸残基的间隔。而“RXK”这一基序只有 H3K4 才具备,对于 H3K9

和 H3K27 均为“RK”基序(组蛋白 H3 N 端前 30 个残基的序列为 ARTK₁QTARK₂STGGKAPRKQLATKAARK₂₇SAP),这也保证了 BPTF 的 PHD 锌指对于 H3K4me3 识别的序列特异性。

虽然绝大多数的阅读器均采用了芳香笼口袋来识别甲基化的赖氨酸残基,但是阅读器可以通过配基识别位点处的关键残基的改变来精确地区分甲基化的不同状态。最显著的改变便是通过酸性残基区域来识别非甲基化修饰的赖氨酸残基。例如,在 BHC 80 的 PHD 锌指中,第 488 位的谷氨酸残基和第 489 位的天冬氨酸残基对于非修饰的 H3K4 识别起到了重要作用(图 3(a))^[9]。对于甲基化的 H3K4 来说,空间位阻和不相容的极性-疏水性相互作用大大降低了结合亲和力。另一种相对温和的改变方式是在芳香笼中引入酸性残基来精细地区分二甲基化和三甲基化赖氨酸残基。例如, Pygo 的 PHD 锌指含有一个由“YDW”残基组成的芳香笼^[20]。在复合物结构中,第 352 位的天冬氨酸残基与二甲基化的赖氨酸残基之间形成了一对直接的氢键,使得 H3K4 二甲基化的亲和力相对于三甲基化强了 1~2 倍(图 3(b))。这种温和的改变在“皇室家族”成员也被广泛使用,例如 53BP1 和 L3MBTL1 的芳香笼中均有酸性残基的替换(图 3(c))^[10,12,21]。



阅读器蛋白使用蓝色表示,不同甲基化状态的赖氨酸残基使用黄色表示。(a) BHC80 的 PHD 锌指识别非修饰的赖氨酸残基(PDB 2PUY); (b) L3MBTL1 的 MBT 结构域识别单甲基化修饰的赖氨酸残基(PDB 2RHY); (c) Pygo 的 PHD 锌指识别二甲基化修饰的赖氨酸残基(PDB 2VPG); (d) BPTF 的 PHD 锌指识别三甲基化修饰的赖氨酸残基(PDB 2F6J)

图 3 阅读器对不同甲基化赖氨酸状态的识别

Fig. 3 Recognition of different methylation states by reader modules

3 组蛋白修饰的组合识别

愈来愈多的证据表明,染色质“打开”或“关闭”的状态往往不仅由单一的组蛋白修饰所决定。例如,虽然 H3K9me3 在异染色质区富集并且同转录抑制高度关联,这一修饰在基因组水平的研究中也被发现同转录活跃区域共存^[22]。事实上,组蛋白的修饰经常成对或成组存在并介导下游的生物学事件。这些修饰可以共存于同一组蛋白尾巴或不同组蛋白尾巴(包括同一核小体的不同组蛋白尾巴或不同核小体的不同组蛋白尾巴)^[9]。相对应地,组蛋白阅读器元件(histone reader module)也可以识别这些组蛋白修饰的组合。这些阅读器元件可以在同一蛋白或同一蛋白复合物中共存,奠定了组蛋白修饰“多价态识别”(multivalent engagement)的分子基础。同一类型的组蛋白阅读器元件可以串联识别同一修饰,

例如 SGF29 的串联 Tudor 结构域可以识别 H3K4me3^[23];不同类型的组蛋白阅读器元件也可以组合在一起识别不同组蛋白修饰的组合,例如 TRIM33 的 PHD-Bromo 组合结构域可以识别“H3K4me0-H3K9me3-H3K18ac”这一组合修饰^[24];3 种不同的组蛋白阅读器元件也可以在同一蛋白中共存,例如 ZYMND8 和 ZYMND11 的 PHD-Bromo-PWWP 组合结构域^[25]。值得注意的是,在组合结构域中并不是每个结构域都有组蛋白阅读器活性,有些结构域起到了结构支架或调节作用。对于组蛋白甲基化的组合修饰目前已经有了较为深入的研究,发现了“赖氨酸甲基化-赖氨酸乙酰化修饰组合识别”、“赖氨酸甲基化-精氨酸甲基化修饰组合识别”和“赖氨酸甲基化-组蛋白变体组合识别”等多种组合识别方式。

组蛋白赖氨酸残基的甲基化与乙酰化修饰经常共存,对

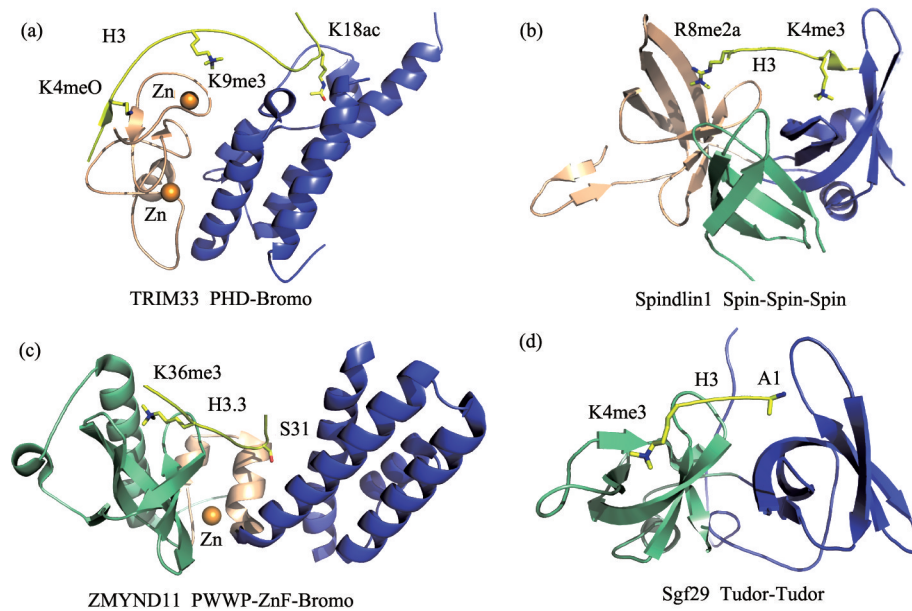
于这一修饰的组合识别有着重要的生物学意义。TRIM33的PHD-Bromo组合结构域对于组蛋白H3多肽的“K4me0-K9me3-K18ac”修饰组合有高达60nM的结合能力^[24]。“K4me0-K9me3-K18ac”修饰组合在细胞内表征了一种“待命”(poised)的染色质状态,基因表达可由细胞外信号激活。“K18ac”或“K9me3-K18ac”的丧失可以导致8~50倍的结合能力的减弱。其中,K4me0被PHD锌指识别,K18ac被Bromo结构域识别,K9me3则由两结构域间的结合口袋识别(图4(a))。这3个残基之间适当间距对于有效识别是必需的。

赖氨酸甲基化-精氨酸甲基化修饰组合的识别机制由对于Spindlin1蛋白的结构研究所揭示^[26]。Spindlin1是Spin/Ssty家族成员之一,参与细胞周期的调控,在多种肿瘤中高表达。结构研究表明Spindlin1分别通过串联Spin/Ssty结构域2和1特异性识别组蛋白H3K4me3和H3R8me2a甲基化修饰(图4(b));利用等温量热滴定法测定该识别的结合常数高达45nM,是目前已报导的结合力最强的组蛋白修饰识别事件,充分显示了组蛋白修饰多价态识别的潜力。由于组蛋白H3K4me3和H3R8me2a均为活跃转录的标志,对于这两种修饰的组合识别也吻合了Spindlin1的转录激活功能。

对于肿瘤抑制因子ZMYND11利用其串联“Bromo-ZnF-

PWWP”结构域识别组蛋白变体H3.3K36me3修饰的研究首次揭示了组蛋白甲基化修饰和组蛋白变体的双重识别机制^[25]。ZMYND11的Bromo-PWWP组合结构域可以区分H3.1(31位为丙氨酸残基)和H3.3(31位为丝氨酸残基)之间仅一个氨基酸残基的差异,ZMYND11对于H3.3K36me3修饰的识别相比于H3K36me3修饰提高了近7倍。由复合物的结构解析发现,H3K36me3插入了PWWP结构域中的芳香笼而H3.3的第31位丝氨酸残基则由Bromo-PWWP交界处的氢键相互作用所稳定(图4(c))。这种对组蛋白变体和甲基化修饰类型的双重识别,是组蛋白修饰识别机制上的新亮点,体现了真核生物表观遗传调控的复杂性和重要性。

在组蛋白甲基化的组合识别中,一个最为极端的例子仅需要4个残基的参与(H3“A1-K4me3”)^[23]。SGF29的串联Tudor结构域形成了紧凑的“面对面”结构来识别H3K4me3,其中组蛋白H3的前4个残基ARTK参与了识别过程。N端的残基H3A1被第1个Tudor结构域所识别,H3K4me3则由第2个Tudor结构域中的芳香笼所识别(图4(d))。SGF29是组蛋白乙酰化催化复合物SAGA的组分之一,这种对H3短多肽的识别被认为可以更加有效地呈递H3K9进行进一步的乙酰化。



阅读器的不同结构域分别使用图中所标注颜色表示,组蛋白多肽使用黄色表示。(a) TRIM33的PHD-Bromo结构域识别H3“K4me0-K9me3-K18ac”修饰(PDB 3U50);(b) Spindlin1的串联Spin结构域识别H3“K4me3-R8me2a”修饰(PDB 4MZF);(c) ZMYND11的PWWP-ZnF-Bromo结构域识别H3.3“S31-K36me3”修饰(PDB 4N41);(d) Sgf29的串联Tudor结构域识别H3“A1-K4me3”修饰(PDB 3MEA)

图4 阅读器对组蛋白甲基化修饰的组合识别

Fig. 4 Combinatorial readout of histone methylation patterns

4 新型组蛋白修饰的鉴定及“修饰对话”

对于新型组蛋白修饰组合的鉴定将进一步扩宽人们对于组蛋白修饰所介导的调控的认识,目前基于抗体技术和质

谱技术的方法是鉴定组蛋白翻译后修饰的主要手段。随着高通量表观基因组和定量蛋白组学的发展,多种更加综合的技术已经被用于组蛋白修饰组合及其识别的鉴定。例如,在

拥有高质量抗体的情况下,ChIP-Seq技术可以提供组蛋白翻译后修饰在基因组水平的分布。ChIP-reChIP是一种连续的染色质免疫沉淀技术,可以直接鉴定多种同时共存的组蛋白翻译后修饰。目前,组蛋白翻译后修饰的ChIP-reChIP偶联液相色谱-质谱技术已经成功被应用于在单核小体水平分析对称或非对称的组蛋白修饰共存^[27]。这项工作也成功发现了在胚胎干细胞“双价启动子”(bivalent promoter)的核小体上存在H3K27me3和H3K4me3的非对称共存。Middle-down和top-down质谱技术是在单组蛋白多肽水平鉴定翻译后修饰的重要技术^[28]。近年来,基于质谱技术的方法已经鉴定出百余种组蛋白翻译后修饰。例如,通过middle-down质谱偶联在线反相高效色谱(RP-HPLC)和电子转移解离(ETD)的方法,超过200余种组蛋白H3.2和70余种组蛋白H4的修饰被鉴定出来,其中包括H3.2中高度修饰的组合“K4me3-K9ac-K14ac-K18ac-K23ac-K27ac-K36me3”^[29]。另外,通过二维液相色谱和top-down的质谱技术,仅用7.5 μg纯化后的Hela细胞核心组蛋白(core histones)鉴定出了708种不同的组蛋白修饰的组合^[30]。定量质谱偶联非标记定量(label-free quantitation)或多重表位标记(multiple isotope labeling)的方法可以在蛋白质组水平进行组蛋白翻译后修饰的比较和动态分析^[31,32]。目前对于组蛋白甲基化修饰识别的研究主要集中在多肽水平,在“设计核小体”(designer nucleosome)水平进行识别机制尤其是组合识别机制的研究将成为进一步了解组蛋白阅读器机理的关键。原则上说,“设计核小体”可以将感兴趣的组蛋白或DNA修饰进行组装,因此这项技术将有望提供更加深入和详尽的表观相互作用谱^[33]。近期,“设计核小体”技术已经可以文库化从而在核小体或染色质水平上分析组蛋白翻译后修饰,例如DNA条形码文库(DNA-barcoded designer nucleosome libraries)和基质辅助染色质阅读器捕获(matrix-assisted reader chromatin capture)技术。

除了多种组蛋白翻译后修饰的组合识别外,组蛋白修饰间的“修饰对话”(modification crosstalk)对于了解组蛋白修饰介导的表观遗传调控也有着重要的意义。组蛋白“阅读器”元件可以同“擦除器”或“书写器”的催化结构域共存,这奠定了“修饰对话”的生化基础。例如,组蛋白赖氨酸去甲基化酶PHF8和KIAA1718均含有一个PHD锌指和JmjC结构域。这两个蛋白的PHD锌指均识别H3K4me3这一修饰且该修饰可以增强去甲基化酶的活性。PHF8同H3“K4me3-K9me2”多肽的复合物结构表明,K4me3插入了PHD锌指的结合口袋并且帮助K9me2定向插入到PHF8的催化口袋^[34]。PHD锌指与H3K4me3的结合或许起到了提高JmjC结构域周围H3K9me2局部浓度的作用,从而促进了催化作用。对于KIAA1718来讲,其PHD锌指和JmjC结构域间的距离远远长于H3K4和H3K9间的距离,因此H3K4me3促进了对H3K27me2的去甲基化活性^[34]。多种翻译后修饰的对话可以以顺序相互作用的形式发生^[35]。比如在MOZ复合物(MOZ, BPRF1, hEAF6和ING5)的4个核心亚基中包含了一个“书写器”元件(MOZ的

MYST结构域)和5个“阅读器”元件(MOZ的双锌指, BPRF1的PHD-Zinc knuckle-PHD、Bromo、PWWP、ING5的PHD)^[36]。MOZ的双锌指识别H3K14乙酰化修饰, BPRF1的PHD-Zinc knuckle-PHD、Bromo、PWWP分别识别非修饰的H3、H3/H4非位点特异的赖氨酸乙酰化修饰、H3K36me3、ING5的PHD锌指特异性识别H3K4me3。被识别的组蛋白非修饰H3、H3K14ac、H3/H4ac、H3K4me3和H3K36me3可以以顺序的形式发生。首先,ING5的PHD锌指结合在H3K4me3富集的启动子来招募MOZ从而产生新的H3乙酰化修饰,被乙酰化的H3则作为双锌指和Bromo结构域的识别锚点进一步稳定了MOZ复合物。BPRF1对于非修饰H3和H3K36me3的识别则促进了MOZ复合物到周边未被H3K4me3标记的基因区域的扩展。

5 展望

表观遗传调控从本质上讲是在染色质层面的遗传信息解读。作为研究最为深入的组蛋白修饰之一,甲基化修饰及其组合修饰和“修饰对话”对于表观遗传调控起到了至关重要的作用。结构生物学在对组蛋白甲基化修饰及其与“阅读器”蛋白相互作用的研究中起到了不可替代的作用。对于结构生物学的一大挑战是在体内功能背景的基础上进行体外的结构生物学研究。另一个重要的挑战则是在“设计核小体”水平进行组蛋白组合修饰的研究。一些新的组蛋白甲基化修饰类型也陆续被报道,如CENP-A的氨基端甲基化修饰^[37]和H2A的104位谷氨酰胺的甲基化修饰^[38],暗示了组蛋白甲基化修饰调控的潜能。对于这些新型甲基化修饰的鉴定、功能上的核实及其与阅读器的识别的结构机制都将为组蛋白修饰介导的表观遗传调控提供新内容和新视角。

除了对于新型组蛋白阅读器机制的研究外,对于组蛋白阅读器在人类疾病及药物研发中的作用也有着越发深入的研究。组蛋白甲基化修饰被发现同癌症、衰老、代谢及神经相关疾病有着密切的联系^[39,40]。例如,ING1-5和PHF23的PHD锌指结构域同肿瘤发生有关^[41]而ATRAX的阅读器结构域则影响到人类智力发育迟滞综合征^[39]。基于蛋白结构的重要信息,以组蛋白甲基化阅读器蛋白作为靶标的药物研发也有了长足的进展。例如,数个靶向“皇室家族”中的MBT结构域的小分子被发现是潜在的药物分子^[42]。值得期待的是,对于组蛋白新型翻译后修饰的鉴定及其组合的结构与功能研究将有助于更加深入地了解和破译“组蛋白密码”,并指导表观因子靶向的创新药物发现。

参考文献(References)

- [1] Strahl B D, Allis C D. The language of covalent histone modifications [J]. *Nature*, 2000, 403(6765): 41-45.
- [2] Jenuwein T, Allis C D. Translating the histone code [J]. *Science*, 2001, 293(5532): 1074-1080.
- [3] Taverna S D, Li H, Ruthenburg A J, et al. How chromatin-binding modules interpret histone modifications: Lessons from professional

- pocket pickers[J]. *Nature Structural & Molecular Biology*, 2007, 14(11): 1025–1040.
- [4] Maurer–Stroh S, Dickens N J, Hughes–Davies L, et al. The Tudor domain ‘Royal Family’: Tudor, plant Agenet, Chromo, PWWP and MBT domains[J]. *Trends in Biochemical Sciences*, 2003, 28(2): 69–74.
- [5] Bannister A J, Zegerman P, Partridge J F, et al. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain[J]. *Nature*, 2001, 410(6824): 120–124.
- [6] Nielsen P R, Nietlispach D, Mott H R, et al. Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9[J]. *Nature*, 2002, 416(6876): 103–107.
- [7] Flanagan J F, Mi L Z, Chruszcz M, et al. Double chromodomains cooperate to recognize the methylated histone H3 tail[J]. *Nature*, 2005, 438(7071): 1181–1185.
- [8] Chen C, Nott T J, Jin J, et al. Deciphering arginine methylation: Tudor tells the tale[J]. *Nature Reviews Molecular Cell Biology*, 2011, 12(10): 629–642.
- [9] Huang Y, Fang J, Bedford M T, et al. Recognition of histone H3 lysine–4 methylation by the double tudor domain of JMJD2A[J]. *Science*, 2006; 312(5774): 748–751.
- [10] Botuyan M V, Lee J, Ward I M, et al. Structural basis for the methylation state–specific recognition of histone H4–K20 by 53BP1 and Crb2 in DNA repair[J]. *Cell*, 2006, 127(7): 1361–1373.
- [11] Wu H, Zeng H, Lam R, et al. Structural and histone binding ability characterizations of human PWWP domains[J]. *Plos One*, 2011, 6(6): e18919.
- [12] Li H, Fischle W, Wang W, et al. Structural basis for lower lysine methylation state–specific readout by MBT repeats of L3MBTL1 and an engineered PHD finger[J]. *Molecular Cell*, 2007, 28(4): 677–691.
- [13] Li Y Y, Li H T. Many keys to push: Diversifying the ‘readership’ of plant homeodomain fingers[J]. *Acta Biochimica et Biophysica Sinica*, 2012, 44(1): 28–39.
- [14] Iwase S, Xiang B, Ghosh S, et al. ATRX ADD domain links an atypical histone methylation recognition mechanism to human mental–retardation syndrome[J]. *Nature Structural & Molecular Biology*, 2011, 18(7): 769–776.
- [15] Collins R E, Northrop J P, Horton J R, et al. The ankyrin repeats of G9a and GLP histone methyltransferases are mono– and dimethyllysine binding modules[J]. *Nature Structural & Molecular Biology*, 2008, 15(3): 245–250.
- [16] Kuo A J, Song J K, Cheung P, et al. The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier–Gorlin syndrome [J]. *Nature*, 2012, 484(7392): 115–119.
- [17] Migliori V, Muller J, Phalke S, et al. Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance[J]. *Nature Structural & Molecular Biology*, 2012, 19(2): 136–144.
- [18] Li H T, Ilin S, Wang W K, et al. Molecular basis for site–specific read–out of histone H3K4me3 by the BPTF PHD finger of NURF[J]. *Nature*, 2006, 442(7098): 91–95.
- [19] Lan F, Collins R E, De Cegli R, et al. Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1–mediated gene repression [J]. *Nature*, 2007, 448(7154): 718–722.
- [20] Fiedler M, Sanchez–Barrena M J, Nekrasov M, et al. Decoding of methylated histone H3 tail by the Pygo–BCL9 Wnt signaling complex [J]. *Molecular Cell*, 2008; 30(4): 507–518.
- [21] Min J R, Allali–Hassani A, Nady N, et al. L3MBTL1 recognition of mono– and dimethylated histones[J]. *Nature Structural & Molecular Biology*, 2007, 14(12): 1229–1230.
- [22] Vakoc C R, Sachdeva M M, Wang H X, et al. Profile of histone lysine methylation across transcribed mammalian chromatin[J]. *Nature Structural & Molecular Biology*, 2006, 26(24): 9185–9195.
- [23] Bian C B, Xu C, Ruan J B, et al. Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation[J]. *The EMBO Journal*, 2011, 30(14): 2829–2842.
- [24] Xi Q R, Wang Z X, Zaromytidou A I, et al. A poised chromatin platform for TGF–beta access to master regulators[J]. *Cell*, 2011, 147(7): 1511–1524.
- [25] Wen H, Li Y Y, Xi Y X, et al. ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression[J]. *Nature*, 2014, 508(7495): 263–268.
- [26] Su X N, Zhu G X, Ding X Z, et al. Molecular basis underlying histone H3 lysine–arginine methylation pattern readout by Spin/Ssty repeats of Spindlin1 [J]. *Genes & Development*, 2014, 28(6): 622–636.
- [27] Voigt P, LeRoy G, Drury W J, et al. Asymmetrically modified nucleosomes [J]. *Cell*, 2012, 151(1): 181–193.
- [28] Karch K R, Denizio J E, Black B E, et al. Identification and interrogation of combinatorial histone modifications[J]. *Frontiers in Genetics*, 2013, 4: 264.
- [29] Young N L, DiMaggio P A, Plazas–Mayorca M D, et al. High throughput characterization of combinatorial histone codes[J]. *Molecular & Cellular Proteomics*, 2009, 8(10): 2266–2284.
- [30] Tian Z X, Tolic N, Zhao R, et al. Enhanced top–down characterization of histone post–translational modifications[J]. *Genome Biology*, 2012, 13(10): R86.
- [31] Britton L M P, Gonzales–Cope M, Zee B M, et al. Breaking the histone code with quantitative mass spectrometry[J]. *Expert Review of Proteomics*, 2011, 8(5): 631–643.
- [32] Eberl H C, Mann M, Vermeulen M. Quantitative proteomics for epigenetics [J]. *Chemical & Pharmaceutical Bulletin*, 2011, 12(2): 224–234.
- [33] Allis C D, Muir T W. Spreading chromatin into chemical biology[J]. *Chemical & Pharmaceutical Bulletin*, 2011, 12(2): 264–279.
- [34] Horton J R, Upadhyay A K, Qi H H, et al. Enzymatic and structural insights for substrate specificity of a family of jumoni histone lysine demethylases[J]. *Nature Structural & Molecular Biology*, 2010, 17(1): 38–43.
- [35] Seet B T, Dikic I, Zhou M M, et al. Reading protein modifications with interaction domains[J]. *Nature Reviews Molecular Cell Biology*, 2006, 7(7): 473–483.
- [36] Klein B J, Lalonde M E, Cote J, et al. Crosstalk between epigenetic readers regulates the MOZ/MORF HAT complexes[J]. *Epigenetics*, 2014, 9(2): 186–193.
- [37] Bailey A O, Panchenko T, Sathyan K M, et al. Posttranslational modification of CENP–A influences the conformation of centromeric chromatin[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2013, 110(29): 11827–11832.
- [38] Tessarz P, Santos–Rosa H, Robson S, et al. Glutamine methylation in histone H2A is an RNA– polymerase– I– dedicated modification[J]. *Nature*, 2014, 505(7484): 564–568.
- [39] Greer E L, Shi Y. Histone methylation: A dynamic mark in health, disease and inheritance[J]. *Nature Reviews Genetics*, 2012, 13(5): 345–357.
- [40] Jakovcevski M, Akbarian S. Epigenetic mechanisms in neurological disease[J]. *Journal of Natural Medicines*, 2012, 18(8): 1194–1204.
- [41] Chi P, Allis C D, Wang G G. Covalent histone modifications: Miswritten, misinterpreted and mis–erased in human cancers[J]. *Nature Reviews Cancer*, 2010, 10(7): 457–469.
- [42] James L I, Barsyte–Lovejoy D, Zhong N, et al. Discovery of a chemical probe for the L3MBTL3 methyllysine reader domain[J]. *Physical Chemistry*, 2013, 9(3): 184–191.

(责任编辑 刘志远)