

植物 SET 蛋白

宋江华 曹家树*

(浙江大学蔬菜研究所, 杭州 310029)

摘要 SET 蛋白是一类包含保守的 SET 结构域、与组蛋白甲基化密切相关的蛋白质。组蛋白修饰作为调控基因表达的重要因素, 在植物体基因转录调控中发挥关键的作用。有关 SET 蛋白的研究为深入了解组蛋白修饰的机制提供了重要信息。植物 SET 蛋白具有保守的结构特征及进化机制, 参与众多细胞核内的反应过程, 如染色体的浓缩和分离, 基因的转录, 以及 DNA 的复制和修复等, 调控植物基因的表达, 影响植物体的发育。

关键词 SET 蛋白; 组蛋白密码; 甲基化; 基因表达

基因的转录调控是识别顺式作用元件的转录调控因子与基因在染色质环境中的时空表达协调作用的结果, 而影响基因在染色质环境中表达的因素包括 DNA 修饰、组蛋白修饰以及相关蛋白质的互作等^[1]。近年来, 组蛋白修饰已作为调控基因表达的重要因素来研究。组蛋白甲基化是由组蛋白甲基转移酶催化的, 尽管组蛋白作为甲基转移酶底物的最早记载始于 1964 年^[2], 但直到最近组蛋白甲基转移酶的分类及它们对功能调控的重要性才逐步引起重视。迄今为止已分离的赖氨酸甲基转移酶都包含有一个保守的 SET 结构域^[3]。包含 SET 结构域的蛋白质称为 SET 蛋白。

SET 蛋白广泛存在于真核生物中。目前, 有关 SET 蛋白的研究大多集中在酵母、动物及人类医学方面, 植物中 SET 蛋白的研究刚刚起步, 对 SET 蛋白结构、系统进化及其功能进行分析研究, 有助于人们深入理解 SET 蛋白对植物发育的调控机制。

1 SET 域的结构特征及家族分类

1.1 SET 域的结构特征

SET 结构域是一个包含约 130 个氨基酸残基的保守序列, 根据果蝇中位置效应花斑(position effect variegation, PEV)的抑制子 *Su (var) 3-9*、*Enhancer of zeste* 和 *Trithorax* 命名^[4]。果蝇的这 3 个基因包含 SET 结构域, 并且与基因遗传外修饰(epigenetic modification)有关。目前, 对 SET 域结构的相关研究为进一步深入了解 SET 蛋白催化及与底物结合的模式奠定了基础。

SET 结构域由 N 端和 C 端形成的两个不连续的



Fig. 1 Structure of a SET domain protein SET7/9^[5]

In the protein, SET-N and SET-C subdomains are shown on the left, the SET-I region is shown on the right, and the unusual knot structure is also shown.

区域组成, 称作 SET-N 与 SET-C。其 N 端和 C 端各自盘曲回旋成不相邻近的空间构象。3~4 个短的 β 折叠, 一个短螺旋和几个环将这些二级结构元件连接起来。有些含 SET 结构域的酶在 SET 结构域 N 端和 C 端之间还包含一个插入序列 SET-I, SET 域两侧富含的半胱氨酸区有前 -SET 域和后 -SET 域^[5](图 1)。蛋白质中的 SET 结构域与相邻的富含半胱氨酸区域共同影响酶活性。SET 结构域的 C 端大多形成一个拓扑学上不寻常的“假结”样结构, 即一个折叠穿过一个环, 形成“假结”的区域在 SET 蛋白中高度保守。SET 结构域两侧的前 -SET 和后 -SET 域, 造成蛋白质序列和结构的多样性。通过对多个 SET 蛋

收稿日期: 2006-09-26 接受日期: 2007-01-31

国家自然科学基金资助项目(No.30671426)

* 通讯作者。Tel/Fax: 0571-86971188, E-mail: jshcao@zju.edu.cn

白的比较发现, SET-I 区在序列长度上变异极大, 在人的组蛋白甲基转移酶 SUV39H1 中仅有 23 个氨基酸残基, 而在 SETDB1 中包含 361 个氨基酸残基^[6]。

对 SET 结构域晶体结构的研究表明, 有些 SET 结构域的 N 端侧链包括前 -SET, 距离酶的活性中心较远, 因此 N 端可能仅仅起到增强 SET 域结构稳定性的作用, 与活性中心的组成无关; 而 SET 结构域保守的 C 端侧链则与酶的催化活性密切相关, 该序列的任何突变都会造成酶的催化活性显著下降^[7]。例如, 切除组蛋白甲基转移酶 SET7/9 位于 SET 域 C 端的氨基酸残基 K344, 该酶的催化活性发生了剧烈的变化, 说明 C 端与酶的活性中心密切相关^[5]。同时, 组蛋白甲基转移酶家族中 SET 结构域 C 端的毗邻序列各不相同, 这种结构可能是造成各种酶的特异性表现出极大差异的原因。

关于 SET 蛋白具体作用机制的研究才刚刚起步, 许多问题尚有待于进一步探索。例如, 不同的 SET 蛋白如何确定赖氨酸残基甲基化转移甲基的数量? 甲基化的不同方式与转录信号之间的关系如何? 不同 SET 蛋白的功能保守程度有何区别? SET 蛋白是在改变甲基化活性还是本身具有甲基化活性? 等等。因此, 要了解 SET 蛋白的作用机制, 还需要对其序列及结构的保守性及生物学功能进行深入研究。

1.2 SET 蛋白家族分类

Baumbusch 等^[8]采用生物信息学方法研究包含 SET 结构域的基因, 通过比较模式植物拟南芥中 30 多个 SET 基因的 SET 结构域, 半胱氨酸区, 以及其他相关结构域, 将这些基因划分为 4 类。RT-PCR 表达分析表明这些基因在植物发育过程中的时空表达具有差异, 也反映出它们对基因活性调控的高度复杂性。

Springer 等^[9]将果蝇、大鼠、酵母中的 SET 蛋白与玉米和拟南芥中的 SET 蛋白进行了系统进化分析, 表明真核生物中的 SET 蛋白具有一定的保守性; 并且根据各种结构域的保守性及其分布, 将 SET 蛋白划分为 5 类。第一类 SET 蛋白包括 CLF 及 MEA 蛋白。它们都包含 5 个保守的结构域, 即 EZD1、EZD2、SANT 域、富含半胱氨酸域和 SET 域^[10]。根据 SET 域外的结构域分类, SET 蛋白的第二类是一个结构多样的集合, 尽管在不同类植物之间序列聚类较低, 但这类蛋白质仍然具有一些共同的特征, 即包含 AWS 域, 且 SET 域的位置相同, 表明这类蛋白质

在起源和功能上可能具有相关性。第三类 SET 蛋白包括拟南芥的 ATX1 和 ATX2 蛋白^[11]。根据蛋白质包含结构域的不同, 又将其分为 4 个分支, 它们分别包含 PWWP 域、FYR 域、PHD 域及 GYF 域等结构。第四类 SET 蛋白仅在植物和酵母中存在。它们均包含一个 SET 域和一个 PHD 域, 没有前 -SET 和后 -SET, 在进化上高度保守, 对基因表达的遗传外修饰起重要作用。第五类 SET 蛋白在植物中最多, 也是唯一一类包含前 -SET 和后 -SET 的 SET 蛋白。

有关 SET 结构域基因的进化分析表明在植物中发生了大量的基因复制事件, 造成植物中 SET 基因的数量较其他物种中偏多。这可能与植物中发生的染色体异源多倍化, 非常规重组及逆转录转座事件相关。研究发现, 保守的 SET 蛋白在植物各组织中均有表达, 这表明 SET 蛋白可能在植物器官分化和发育过程中发挥作用。因此, 对 SET 蛋白进行时空表达模式的进一步分析十分必要。总之, 对 SET 蛋白保守性及系统进化的分析, 有助于推测一些未知 SET 蛋白的功能, 为进一步揭示 SET 蛋白对植物发育的调控机制创造条件。

2 植物 SET 蛋白的生物学功能

SET 蛋白作为调控蛋白或存在于蛋白质复合体中, 参与一些下游的细胞核内反应过程, 如染色体的浓缩和分离, 基因的转录, 以及 DNA 的复制和修复等。大量包含 SET 结构域蛋白的发现, 及其对于组蛋白赖氨酸甲基化的作用, 为人们深入理解组蛋白密码的调节机制提供了重要的信息^[12]。

2.1 与组蛋白甲基化的相关性

植物细胞中染色质状态对于基因在发育中的表达调控及遗传外修饰极其重要。关于植物染色质蛋白的突变对植物发育产生影响的研究早有报道^[13-15], 但调控植物染色质状态的详细的分子机制尚未明确。近年来, 有关“组蛋白密码”的研究为揭示植物染色质结构改变的分子机制提供了必要的信息。组蛋白翻译后修饰可能影响染色质结构和染色质相关蛋白的合成^[16]。

众多研究表明 SET 蛋白对组蛋白甲基化起作用。组蛋白甲基化位点多位于 H3、H4 的赖氨酸和精氨酸残基上, 组蛋白 H3 氨基尾端上的 K4 和 K9 便是其中的两个甲基化的常发位点。不同位点上的甲基化, 由不同的酶负责。例如, 果蝇中的 Su(var)3-9 是最早发现的催化赖氨酸甲基化的组蛋白甲基转移

酶,专一作用于组蛋白 H3 中的第 9 位赖氨酸,引起基因沉默^[17]。哺乳动物和裂殖酵母中 *dim-5* 基因编码的同源蛋白 SU(VAR)3-9 和 CLR4,可以催化组蛋白 H3 第 9 位赖氨酸残基甲基化,与染色质沉默密切相关^[18]。酵母 SET1 甲基化组蛋白 H3 的第 4 位赖氨酸,同时对 rDNA 沉默是必须的^[19]。在哺乳动物中,SET7 甲基化组蛋白 H3 的第 4 位赖氨酸^[20]。虽然不同真核生物中催化组蛋白发生甲基化的酶有所不同,但大部分组蛋白甲基转移酶都含有一个保守的 SET 结构域,表明 SET 蛋白能使组蛋白发生甲基化。

许多植物 SET 蛋白也具有组蛋白甲基转移酶活性,对于特定的底物和基因的表达起调节作用^[21]。在拟南芥和玉米已发现的大量 SET 蛋白中,拟南芥 ATX-1 能对组蛋白 H3K4 产生离体甲基化,维持花同源异型基因的表达^[22]。拟南芥 *KYP* 基因编码一个含有 SET 结构域蛋白甲基转移酶,此蛋白质可以特异性地使 H3K9 发生甲基化。试验表明,剔除 *KYP* 基因会使组蛋白 H3K9 和基因组 DNA 的甲基化水平大大降低,但对拟南芥表型性状没有产生影响。最近分离了拟南芥 *suvh4* 基因突变体,发现 *SUVH4* 基因抑制植物开花相关基因 *SUPERMAN* 的甲基化和沉默,同时对该基因进行离体表达,显示其具有 H3K9 组蛋白甲基转移酶活性^[23]。然而,另有研究表明拟南芥中的 SET 蛋白 *SUVH4*,对维持基因组 DNA 甲基化起重要作用,但与组蛋白甲基化无关^[24]。这一发现不仅为全面了解 SET 蛋白的甲基化功能提供了必要信息,而且对于研究组蛋白甲基化和基因组 DNA 甲基化之间的关系具有重要意义。

烟草中的 NtSET1 被证实与染色质结合,具有 H3K9 组蛋白甲基转移酶活性,其异位表达加强了 H3K9 的甲基化,影响烟草悬浮培养细胞的染色质分离,抑制植株生长。从细胞水平分析发现可能是由于不正常的细胞分裂和分化阻碍了细胞的伸长,这也说明组蛋白 H3K9 发生甲基化对植物发育的调控起重要作用^[25]。Yu 等^[26]通过悬浮培养烟草 BY2 细胞,进一步研究了 NtSET1 的酶活性及对细胞的影响,选用 PER8 诱导型表达载体异位表达 NtSET1,促进了 H3K9 的二甲基化,影响烟草 BY2 悬浮细胞染色体的分离。因此,进一步分析组蛋白 H3K9 甲基化对植物生长调控的作用,对于阐明细胞伸长和分化的下游调控机制十分必要。

2.2 蛋白质复合体互作及对转录调控的影响

组蛋白对于转录调控等过程至关重要,通过对其

末端的化学修饰作用,参与细胞核中的生命活动。组蛋白赖氨酸和精氨酸的甲基化同转录调节和异染色质的形成相关。组蛋白甲基化修饰的结果相对复杂,可以使转录增强或转录抑制^[27]。

有研究已表明 SET 蛋白对于转录调控发挥着重要的作用,在染色质相关复合体中可能促进或抑制基因表达。果蝇中的 Pc-G 和 trx-G 蛋白与染色质结构密切相关,通过形成稳定的染色质复合体,在发育过程中起着稳定转录状态的作用。PcG 蛋白维持了转录的沉默状态,而 TrxG 维持转录的激活状态^[28]。酵母的 SET1 属于 8 个蛋白质组成的复合体,大多参与催化 H3K4 甲基化,且甲基化状态与特定的生物反应相关。例如, H3K4 的三甲基化与转录激活相关^[29]。SET1 与 Paf1 复合体互作,影响 RNA 聚合酶 II 延伸^[30],而 RNA 聚合酶 II 对于转录的顺利完成具有决定性作用。SET1 也调控 rDNA 和端粒的沉默,端粒长度,以及 DNA 修复等^[31]。这些过程均与转录调控密切相关。

植物减数分裂的正常进行需要一些基因紧凑协调的转录过程来保证。减数分裂维持有性生殖中细胞倍性的稳定性,因此是有性植物繁殖的一个非常重要的过程。尽管目前在植物中有关 SET 蛋白对基因转录调控影响的研究较少,但根据 SET 蛋白在动物与植物中的结构及进化关系可以推测植物中 SET 蛋白也对发育过程中的转录调控起作用,具体的作用机制还有待于深入研究^[32]。

许多 SET 蛋白存在于大的蛋白质复合体中,也具有其他结构域,像 PHO、PWPP 和 YDG 域等,这些结构域都参与蛋白质间的互作调控。而且 SET 蛋白在 N 端半胱氨酸富含区高度保守,N 端的结构域可能决定着 SET 蛋白和其他蛋白质间的互作^[33]。推测,SET 蛋白可能决定着蛋白质复合体的特异性,一个蛋白质复合体可能和多个 SET 蛋白产生互作。拟南芥中的 E(z) 亚家族中的 3 个蛋白质,CLF、MEA 和 EZA1,都与同一个蛋白 FIE 互作^[34]。如果其他 SET 蛋白复合体也是这样,那么可以认为 SET 蛋白对于决定蛋白质复合体的特异性是非常重要的。

2.3 对植物发育的影响

植物中发现的许多 SET 蛋白对植物发育产生了重要影响。为揭示 SET 蛋白在植物中的功能,在烟草、玉米、菠菜、豌豆、水稻和矮牵牛等多种植物中,相继报道了有关 SET 基因的研究^[35-37]。

植物中发现的第一个包含 SET 结构域的基因是

CURLY LEAF (CLF)。研究表明, 该基因在拟南芥的叶、茎、花序和花等组织中均有表达, 可以抑制花的同源异型基因 *AGAMOUS* 的转录, 影响花的形态和开花时间^[13]。另一个植物 SET 基因是 *MEDEA (MEA)*, 它也被发现在植物的各个器官均有不同程度的表达, 与拟南芥胚芽分化有关^[35]。CLF 和 MEA 这两个 SET 蛋白与果蝇的 Pc-G 蛋白同属于 SET 蛋白第一亚家族。果蝇中的 Pc-G 蛋白和染色质结构密切相关, 通过形成稳定的染色质复合体, 对基因表达造成长期的抑制。但植物中 CLF 和 MEA 与染色质结构的关系尚未报道。Mayama 等^[36]在矮牵牛中也克隆到两个 SET 基因 (*PhCLF1* 和 *PhCLF2*), 与拟南芥 *CLF* 的同源性较高。两个基因功能虽然略有差异, 但在花器官和叶片中均有表达, 对 C 级开花基因 *AGAMOUS (AG)* 的转录起抑制作用。

Shen 等^[25]将烟草中的 *NtSET1* 基因与绿色荧光蛋白融合进行细胞定位, 发现它在烟草悬浮培养细胞内细胞间期的分布不均匀。有丝分裂期规则分布于细胞核内, 与浓缩的染色体紧密结合。在过量表达 *NtSET1* 基因的转基因烟草中, 植物表现生长缓慢, 叶片窄小, 茎间变短, 花蕾易脱落等表型特征。推测 *NtSET1* 可能通过抑制基因的转录活性, 造成植物生长延缓。

Liang 等^[37]通过结合 cDNA 文库筛选和 5' RACE 技术, 克隆到水稻的一个全长为 2 957 bp 的 SET 基因 *OsSET1*。OsSET1 的表达模式与在拟南芥和玉米中的 SET 基因相似, 在植物各个器官中均有表达, 没有器官特异性。在拟南芥中过量表达 *OsSET1*, 发现 53.8% 植株幼苗期茎的生长受到抑制, 子叶较大, 没有真叶, 茎尖不具有原套-原体 (tunic-carpus) 结构。

几项研究表明, 拟南芥中 3 个 Su(z)12 同源蛋白——Fis2、Emf2 和 Vrn2 在植物体中调控不同的发育阶段。Fis2 调节胚乳发育, Emf2 调节花的发育, Vrn2 对不同温度处理下花的发育起调控作用^[14,38,39]。今后, 进一步研究植物多个 SET 蛋白之间的相互关系及其作用机制, 将有助于人们揭示植物发育及基因表达的调控机制, 从而对调节植物生长发育产生实际的意义。

综合上述研究, 可以发现 SET 基因在植物的不同器官和发育的不同时期都可能表达。植物在发育的不同阶段需要基因表达系统的遗传外修饰来维持分生组织和器官的特定特征, 从而积极应对遗传和环境产生的影响, 顺利地完发育过渡, 而 SET 蛋白在

此过程中发挥了重要作用^[9,40,41]。

我们在白菜核雄性不育 *mmc* 突变体与其野生型植株花蕾差异表达的基因中, 发现了一个与 SET 蛋白家族众多成员相似的基因 (*BcMF-SET*), 尽管该基因在花粉发育中的确切功能还未知, 但根据其在花粉发育不同阶段的表达特征, 推测该基因可能在花粉发育中具有一定的作用 (未发表工作)。在植物花粉发育过程中孢子的形成必需经过减数分裂, 我们分离鉴定的 SET 基因是否通过转录调控来调节一些减数分裂中间基因的表达, 或者是通过与其他基因发生互作, 从而对白菜的花粉发育产生影响, 还需要进一步的实验证明。

3 小结与展望

组蛋白甲基化在植物细胞染色质的遗传外修饰中占有中心地位, 对细胞繁殖与分化、基因表达、转录、基因组稳定性等均有深远的影响, 而 SET 蛋白的大量研究表明其在组蛋白甲基化中扮演着十分重要的角色。近年来, 虽然关于 SET 蛋白的研究进展迅速, 但仍存在许多问题需要深入。例如, SET 蛋白在转录调控中起作用的具体分子机制, 组蛋白甲基化与其他蛋白质翻译后修饰的关系, 以及相关的生物学功能等等。了解 SET 蛋白结构特征及系统进化机制不仅有利于深入研究它们在植物体中发挥的功能, 而且可以揭示其在发育调控过程中的分子机制, 深化人们对蛋白质翻译后修饰及基因表达调控的认识, 从而为解释困扰人们的一些生物学现象提供新的思路。

参考文献 (References)

- [1] Jenuwein T et al. *Science*, 2001, **293**: 1074
- [2] Murray K. *Biochemistry*, 1964, **3**: 10
- [3] Yeates TO. *Cell*, 2002, **111**: 5
- [4] Alvarez-Venegas R et al. *Gene*, 2002, **285**: 25
- [5] Wilson JR et al. *Cell*, 2002, **111**: 105
- [6] Marmorstein R et al. *Trends Biochem Sci*, 2003, **28**: 59
- [7] Trievel RC et al. *Cell*, 2002, **111**: 91
- [8] Baumbusch LO et al. *Nucleic Acids Res*, 2001, **29**: 4319
- [9] Springer NM et al. *Plant Physiol*, 2003, **132**: 907
- [10] Springer NM et al. *Plant Physiol*, 2002, **128**: 1332
- [11] Alvarez-Venegas R et al. *Gene*, 2001, **271**: 215
- [12] Dillon SC et al. *Genome Biol*, 2005, **6**: 227
- [13] Goodrich J et al. *Nature*, 1997, **386**: 44
- [14] Yoshida N et al. *Plant Cell*, 2001, **13**: 2471
- [15] Wagner D et al. *Curr Biol*, 2002, **12**: 85
- [16] Zhang Y et al. *Genes Dev*, 2001, **15**: 2343
- [17] Rea S et al. *Nature*, 2000, **406**: 593

- [18] Fishcher A *et al.* *J Plant Physiol*, 2006, **163**: 358
[19] Briggs SD *et al.* *Genes Dev*, 2001, **15**: 3286
[20] Wang H *et al.* *Mol Cell*, 2001, **8**: 1207
[21] Ebbs ML *et al.* *Plant Cell*, 2006, **18**: 1166
[22] Alvarez-Venegas R *et al.* *Curr Biol*, 2003, **13**: 627
[23] Jackson JP *et al.* *Nature*, 2002, **416**: 556
[24] Malagnac F *et al.* *EMBO J*, 2002, **21**: 6842
[25] Shen WH. *Plant J*, 2001, **28**: 371
[26] Yu Y *et al.* *Plant J*, 2004, **40**: 699
[27] 胡 笳等. *科学通报*, 2005, **50**: 1061
[28] Kouzarides T. *Curr Opin Genet Dev*, 2002, **12**: 198
[29] Santos-Rosa H *et al.* *Nature*, 2002, **419**: 407
[30] Krogan NJ *et al.* *Mol Cell*, 2003, **11**: 721
[31] Bryk M *et al.* *Curr Biol*, 2002, **12**: 165
[32] Hayashi K *et al.* *Nature*, 2005, **438**: 374
[33] Rozovskaia T *et al.* *Oncogene*, 2000, **19**: 351
[34] Luo M *et al.* *Proc Natl Acad Sci USA*, 2000, **97**: 10637
[35] Kiyosue T *et al.* *Proc Natl Acad Sci USA*, 1999, **96**: 4186
[36] Mayama T *et al.* *Plant Cell Physiol*, 2003, **44**: 811
[37] Liang YK *et al.* *J Exp Bot*, 2003, **54**: 1995
[38] Luo M *et al.* *Proc Natl Acad Sci USA*, 1999, **96**: 296
[39] Gendall AR *et al.* *Cell*, 2001, **11**: 525
[40] He Y *et al.* *Trends Plant Sci*, 2005, **10**: 30
[41] Raynaud C *et al.* *Plant J*, 2006, **47**: 395

SET Domain Proteins in Plants

Jiang-Hua Song, Jia-Shu Cao*

(Institute of Vegetable Science, Zhejiang University, Hangzhou 310029, China)

Abstract SET domain proteins contain conserved SET domains, which intensively involved in histone methylation. Histone modification plays an important part in transcriptional regulation in plants emerging as a central theme in the control of gene expression. SET domain proteins have been extensively studied in recent years, which are required to elucidate the mechanism of histone modification. In this review, the conserved structure and evolutionary mechanism of SET domain proteins in plants was first described. Furthermore, their complex biological functions were discussed, involved in the condensation and isolation of chromosome, the transcription of gene, and the replication and repair of DNA, and so on. It is revealed that SET domain proteins in plants are important for determining epigenetic regulation of gene expression and plant development.

Key words SET domain protein; histone code; methylation; gene expression

Received: September 26, 2006 Accepted: January 31, 2007

This work was supported by the National Natural Science Foundation of China (No.30671426)

*Corresponding author. Tel/Fax: 86-571-86971188, E-mail: jshcao@zju.edu.cn