

苏云金芽孢杆菌杀虫晶体蛋白基因的启动子及其转录调控

程 萍* 王清锋 喻子牛

(华中农业大学微生物系, 农业微生物农业部重点实验室 武汉 430070)

关键词 苏云金芽孢杆菌, 杀虫晶体蛋白基因, 启动子, 转录调控

分类号 Q756 **文献标识码** C **文章编号** 0253-2654(1999)-02-0130-34

苏云金芽孢杆菌(*Bacillus thuringiensis*)是一种G⁺土壤杆菌,其之所以能成为世界上生产量最大的生物杀虫剂,是因为它能在细胞生长过程中,产生大量、高效的杀虫晶体蛋白(Insecticidal Crystal Proteins, ICPs),并在细胞中累积达细胞干重的25%以上。苏云金芽孢杆菌为什么能产生如此丰富的杀虫晶体蛋白?其调控的机制是什么?这正是近年来各国学者关注的课题。

细胞基因的表达,即由DNA转录成RNA再翻译成蛋白质的过程。在细胞的生长、发育和分化过程中,遗传信息的表达可按一定时间程序发生变化,并随着细胞内外环境条件的改变而加以调整,而转录水平的调控是关键环节^[1]。启动子是转录起始的控制部位,它的结构与功能直接影响到基因的表达。可见启动子的研究在苏云金芽孢杆菌杀虫晶体蛋白基因的有效表达上有着极为重要的意义。

1 杀虫晶体蛋白基因及其存在状态

1.1 杀虫晶体蛋白基因的分类

自1981年第一个ICP基因被克隆以来,到目前已有129种杀虫晶体蛋白基因被克隆和测序(Sasaki, 1996)。对于苏云金芽孢杆菌杀虫晶体蛋白基因的分类, Hofte 和 Whiteley (1989)曾按杀虫谱和分子特性将cry基因分为五大类,即cryI、cryII、cryIII、cryIV和细胞溶素cyr基因^[2]。1996年以Crickmore为首的cry基因命名委员会提出了按照晶体蛋白基因全长氨基酸序列的同源性将cry基因分类的新系统。本文的ICPs基因名称均按此新命名系统。

1.2 cry基因的存在状态

苏云金杆菌含有极其丰富的质粒(从1.5MD到大

于130MD),除了个别菌株的cry基因位于染色体上外,大多cry基因均位于较大的质粒上(大于30MD)。这些质粒的拷贝数往往很低。在苏云金杆菌的野生型菌株中,通常每个细胞具有多个带有cry基因的质粒,而有些质粒又带有1个到几个晶体蛋白基因,特别是cryI(即旧的cryI)、cry4和cry11(旧的cryIVD)基因。这些复合的cry基因可以直接合成相应的蛋白质,并形成异质(heterogeneous)的晶体内含物或合成特殊的蛋白质,进而形成明显形态不同的晶体。库斯塔克亚种的HDI菌株就是一明显的例子,它含有12个土族resident质粒,包含cry1Aa(旧的cry1Aa)、cry1Ac(旧的cry1Ac)、cry2Aa(旧的cryIIA)及一个沉默(silent)的cry2Ab(旧的cryIIB)基因在一个110MD的质粒上,而cry1Ab(旧的cry1Ab)基因在一个能自我转移的(self-transmissible)44MD的质粒上,当形成伴胞晶体时,cry1Aa、cry1Ab和cry1Ac蛋白组合形成HDI的菱形(bipyramid)晶体,而cry2Aa蛋白却单独组装成立方体(cuboidal-shaped)形的晶体。在以色列亚种中,cry4Aa(旧的cryIVA)、cry4Ba(旧的cryIVB)、cry11Aa(旧的cryIVD)及cyr1Aa(cyrA)基因却位于一个72MD的质粒上,并形成一个复合的伴胞晶体(Aronson, 1993)^[3]。由此可见,在苏云金芽孢杆菌菌株中,不同的cry基因转录合成不同量的相应蛋白质,并最终按一定比例和组成,精巧地组装成晶体。这说明,

* 现在珠海市农业科学研究中心工作。联系地址:广东省珠海市香洲区前山梅溪珠海市农业科学研究中心,519070

1998-02-09收稿,1998-07-31修回

在苏云金杆菌的体内必然存在着精密的调控机制,以使各种杀虫晶体蛋白基因的表达在产量、比例和时间等诸方面满足细胞生存的最佳需要。

2 杀虫晶体蛋白基因启动子的类型与功能

2.1 启动子的基本特点

在苏云金芽孢杆菌的杀虫晶体蛋白基因中,除了 *cry2Aa*(旧的 *cryIIA*), *cry2Ac*(旧的 *cryIIC*) 和 *cry11Aa*(*cryIVD*) 是以操纵子形式存在外,其他所有的 *cry* 基因都是单顺反子,即每个 *cry* 基因转录成 mRNA 都通过自身的启动子起始转录^[3]。因此,启动子是影响 *cry* 基因表达的一个重要因子。

所谓启动子,即 RNA 聚合酶结合到基因转录起始的一段 DNA 序列。参与启动子功能的 DNA 序列,首先是在大肠杆菌 (*E. coli*) 中通过一系列不同基因的核苷酸序列分析比较而获得的。经比较发现,转录起始位点的上游区含有两段序列,它们在各种基因中都十分相近,这些共有序列,每个由 6 个核苷酸组成,并位于转录起始位点上游大约 10 和 35 个碱基对处,它们被称为 -10 和 -35 因子 (elements),这只是表明它们与转录起始点的相对位置,因转录起始位点被定义为 +1。不同启动子在 -10 和 -35 位置上的序列是不同的,但却非常相似,以致足以建立共有序列——即在每个启动区的相应位置都能发现的、最频繁出现的碱基对。-10 区的保存序列为 TATAAT,称为 Pribnow 框。或 -10 序列,-35 区的保存序列为 TTGACA,称为识别区或 -35 序列^[5]。

经突变分析表明,-35 序列的突变将降低 RNA 聚合酶与启动子结合的速度,但不影响转录起点附近 DNA 双链的解开;而 -10 序列的突变效应相反。这表明,-35 序列提供 RNA 聚合酶识别的信号,-10 序列则有助于 DNA 局部双链解开。从序列的结构也可看出,-10 序列含较多的 A-T 碱基对,因而双链分开所需的能量也较低。

2.2 杀虫晶体蛋白基因启动子类型

当第一个杀虫晶体蛋白基因被克隆后,就对其启动子进行了研究,发现 *cry1Aa*[旧的 *cryIA(a)*] 启动子是两个重叠的启动子^[6],从而引起研究者的极大关注,并对各种 *cry* 基因的启动子结构、特点及作用机制进行了大量的研究,结果表明,不同 *cry* 基因,其启动子的结构和作用不同。因此,*cry* 基因启动子可根据其结构特

征不同而划分为不同的类型。

2.2.1 双重重叠启动子 (Dual overlapping promoters):所有的 *cry1*(旧的 *cryI*) 基因和 *cry11Aa*(旧的 *cryIVD*) 操纵子都含有两个重叠的启动子^[7~9]分别称为 BtI(下游启动子, downstream promoter) 和 BtII(上游启动子, upstream promoter)。它们的结构特征是:上游 BtII 启动子的 -10 区位于 BtI 的 -10 与 -35 区之间(Hofte 等, 1989; Dervyn, 1995)。BtI 启动子可被含 σ^E (σ^{35}) 的 RNA 聚合酶识别而启动转录,作用从芽孢形成 II 到 IV 期,而 BtII 则被含 σ^K (σ^{28}) 的 RNA 聚合酶识别而启动转录,并从芽孢形成 IV 到 VI 期发生作用。因此,这两个启动子的协同作用就能够使得这些 *cry* 基因的转录在整个芽孢形成期不断进行。

双重重叠启动子是很多 *cry* 基因的一个结构特征,它与单一启动子和非重叠双启动子在功能上有何区别?通过突变重叠区,使 BtI 的转录失活,结果却导致从 BtI 启动的转录率提高 5~6 倍。相反,突变 BtII,则依赖 BtII 的转录率很低(Aronson 和 T. Walter, 1996; 个人交流资料)。可见 BtI 是一强启动子,BtII 为弱启动子,BtI 通过 BtII 而得到调节。

值得注意的是,尽管 *cry1* 和 *cry11Aa*(旧的 *cryIVD*) 具有类似的双重重叠启动子,但对 *cry* 基因的转录效率却仍有差异。Wu 和 Federici (1995) 利用 *cry1Ac*[*cryIA(c)*] 的启动子和 *cry11Aa*(*cryIVD*) 的启动子分别检测 P20 对 *cry11Aa*(*cryIVD*) 表达的影响,发现利用 *cry1Ac* 启动子控制 P20 的表达,而 *cry11Aa* 用其自身的启动子结果形成一个很大的 *cry11Aa*(*cryIVD*) 的菱形晶体,其体积为野生型 *cry11Aa*(*cryIVD*) 晶体的 5 倍,而为 *cryIVD* 操纵子形成晶体体积的 1.7 倍,可见 *cry1* 和 *cry11Aa* 的启动子类型相同,但在对 *cry* 基因的表达上却有差异^[10]。这也说明,利用 *cry1Ac* 启动子增加 P20 的表达,能导致 *cry11Aa* 蛋白的增加。

事实上,双重重叠启动子并非仅仅发生在苏云金芽孢杆菌中,在枯草芽孢杆菌的 *spoVG* 基因中也发现这类启动子(Johnson 等, 1983)。

2.2.2 非重叠的双启动子 (Dual unoverlapping promoters):Yoshisue 等 (1993; 1995) 对以色列亚种的 *cry4Aa*(旧的 *cryIVA*) 和 *cry4Ba*(旧的 *cryIVB*) 的启动子区进行了研究,发现这两个基因的转录都是受 σ^{35} (σ^E) 的 RNA 聚合酶控制,即它们都具有 BtI 启动子,并指

出, *cry4Aa* 也能被含 σ^H 的 RNA 聚合酶转录^[11~12]。最近, Yoshisue 等 (1997) 对 *cry4Aa* 的上游非翻译区进行了核苷酸序列分析, 并鉴定出 *cry4Aa* 的第二个转录起点, 这个转录是依赖于 σ^{38} (σ^K) 因子的, 并在芽孢形成后期发生作用。序列分析的结果显示, *cry4Aa* 基因具有两个启动子, 下游启动子 P1 位于 -349 区段, 上游启动子 P2 位于 -649 区段, 即这两个启动子的 -10 区和 -35 区并未重叠, 在两个启动子之间相距近 300bp。进一步证明, 从 *cry4Aa* 的 P2 启动子的转录是在芽孢发育的后期 (T20), 而且 P2 启动子的 -10 和 -35 的核苷酸顺序与 σ^{38} (σ^K) 因子特异识别的启动子相近^[13]。可见, *cry4A* 的 P2 启动子是被含 σ^K 的 RNA 聚合酶识别, 而 P1 早已证实是依赖 σ^E 因子 (Yoshisue, 1993)。通过缺失分析发现, P2 启动子也是一弱启动子, 它的缺失对 *cry4Aa* 的转录无明显的影响, 这一特点与 *cryI* 基因的 *BtII* 启动子有相似之处^[13]。另外, *cryIIAa* 的启动子也属于非重叠的双启动子 (Dervyn 等, 1995)。

2.2.3 单启动子 (mono-promoter): 除 *cryI*、*cryIIAa*、*cry4Aa* 和 *crtA* 外, 其余的 *cry* 基因都检测到含一个启动子, 且为 *BtI* 类, 这目前还没有发现只含 *BtII* 启动子的 *cry* 基因。

在含单启动子的 *cry* 基因中, *cry3Aa* (旧的 *cryIIIA*) 基因的启动子却明显不同^[14]。*cry3Aa* 在营养生长期表达, 并在稳定生长初期被激活 (Sekar, 1988; Malvar 等, 1994)。然而, 在枯草芽孢杆菌中的研究发现, *cry3Aa* 的表达并不是依赖于芽孢形成的特异 σ 因子, 而且能在 *spoOA* 的遗传背景下增加表达^[14]。这在苏云金芽孢杆菌中也得到了证实, 即将 *cry3Aa* 基因转入 *spoOA* 突变体中, 这种苏云金杆菌的菌株仍能产生大量的 *cry3Aa* 蛋白^[15]。这说明, *cry3Aa* 的表达是不依赖于芽孢形成因子。

对 *cry3Aa* (*cryIIIA*) 上游核苷酸序列分析发现, *cry3Aa* 的启动区远离该基因翻译起始密码, 在 -635 和 -553 区域并证实主要的转录起始位点为 -558。其 -35 和 -10 区的序列与 A 因子识别序列相似, 且在营养生长期表达, 因此认为该启动子是由含 σ^A 因子的 RNA 聚合酶启动转录。同时也发现, 相应的细胞中有大量的以 -129 为 5' 末端的 mRNA 的存在, 开始认为这种 mRNA 是来自从 *cry3Aa* 的 -129 位点开始的转录, 即在此位点上游有相应的启动子。但缺失分析表

明, -129 区域缺失对 T-129 的形成并无改变, 这表明, -129 位点并非真正的转录起始位点。这种 T-129 mRNA 的形成只是在 -558 位点转录后, 加工形成的一种稳定的 mRNA 形式 (Agaisse 和 Lerechus, 1994), 它是 *cry3Aa* 基因 mRNA 存在的主要形式。可见, *cry3Aa* 启动子的结构与功能十分复杂, 它们转录主要依赖于 -558 区的启动了, 这一启动子又依赖于 σ^A , 而下游区 -129 区域的作用可能为转录后的修饰, 具体作用机制尚不清楚。

另外, *cry3Ba* (旧的 *cryIIIB*) 和 *cry3Bb1* (旧的 *cryIIIB2*) 也具有 *cry3Aa* 类似的启动子结构。

2.2.4 存在于操纵子中的启动子

(1) *cry2Aa* 启动子: 苏云金杆菌的汤普逊亚种中的 *cry2Aa* (旧的 *cryIIA*) 基因编码 65kD 蛋白质, 在胞内形成立方体的晶体, 且对鳞翅目和双翅目昆虫均有毒力。该基因却不像大多数 *cry* 基因为单顺反子, 而是位于一含三个基因的操纵子中最远的结构基因位置 (Widner 和 Whiteley, 1989)。因此, *cry2Aa* 基因的转录受该操纵子的控制。经序列分析发现, 单一的、依赖芽孢形成的启动子位于 *cry2Aa* 操纵子的第一个结构基因 *orf1* 上游 36bp 的位置。第二个结构基因 *orf2* 含有一段 15 个氨基酸重复 11 次的氨基酸序列。在大肠杆菌中, 只表达 *orf1* 和 *orf2* 的细胞裂解物对测试昆虫无毒 (Widner 和 Whiteley, 1989)。这表明, *orf1* 和 *orf2* 不属于杀虫基因 (图 1B)。

Crickmore 和 Ellar (1992) 的研究表明, 单独的 *cry2Ac* 基因不能直接在测试的苏云金杆菌菌株中形成晶体, 而 *cry2Aa* 晶体的形成需要 *orf2* 的存在。*orf2* 的基因产物为 -252 个氨基酸的多肽, 在 *cry2Aa* 的晶体中也检测到含量很低的 ORF2 蛋白, 可见 *orf2* 的多肽产物对 *cry2Aa* 的表达是必需的, 而 *orf1* 对 *cry2Aa* 的表达无明显的作用 (Crickmore and Ellar, 1992)。但 Wu 等 (1991) 指出, *orf1* 和 *orf2* 对库斯塔克亚种中的 *cry2Ac* 晶体的形成并非必要。这说明 *cry2Aa* 和 *cry2Ac* 基因结构不同, 对 *orf2* 的需要也不同。

(2) *cry15Aa* 基因的启动子: Brown 和 Whiteley (1992) 在苏云金杆菌汤普逊亚种 (*subsp. thompsoni*) 发现了两种新的杀虫晶体蛋白基因, 当时定名为 *cry34* 和 *cry40*。1993 年 Brown 经序列分析, 证实 *cry40* 位于 *cry34* 上游 64bp, 并各具有自己的核糖体结合位点^[15]。

在两个基因间未发现启动子与终止子,而在 *cry34* 基因后发现一终止子结构。Western 分析检测了启动子的活性在 *cry40* 基因上游,而并非两基因之间。由此确定这两个基因属于一个操纵子,该操纵子的转录是从 *cry40* 翻译起始密码上游 84 和 85bp 之间的连接点开始,该启动子的 -10 和 -35 区的序列与苏云金杆菌 RNA 聚合酶 σ^{35} (σ^E) 识别的共有序列相一致 (图 1c)。

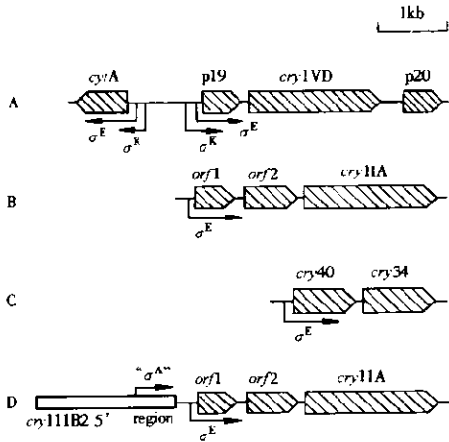


图 1 编码杀虫晶体蛋白的 Bt 操纵子

A. 以色列亚种的 *cry11Aa* (*cryIVD*) 操纵子; B. 库斯塔克亚种的 *cry2A* (*cryIIA*) 操纵子; C. 汤普逊亚种的 *cry40-cry34* (*cry15A*) 操纵子; D. 启动子的位置和起始用箭头表示。控制每个启动子的 σ 因子也指明。

(引自 Brown J. of Bacteriol, 1993)

进一步的毒力测定表明, *cry40* 蛋白对测试昆虫无毒 (Brown 和 Whiteley, 1992)。因此, *cry34* 现定名为 *cry15A*, 而 *cry40* 的功能尚不十分清楚, 它的存在可能类似于 *orf2* 对 *cry2A* 的作用。由此可见, *cry15A* 的启动子是位于操纵子内, 并受 RNA 聚合酶 σ^E 的识别而启动转录, *cry15A* 的表达只在芽孢形成期, 且开始于芽孢形成初期, 说明该启动子属于 BtI 类。

(3) 杀蚊蛋白基因的启动子: 苏云金杆菌以色列亚种 (*subsp. israelensis*) 在芽孢形成期能产生一种对苍蝇和蚊幼虫有极高毒力的伴胞晶体, 该晶体主要由四种多肽组成: *cry4Aa* (*cryIVA*, 128kD), *cry4Ba* (*cryIVB*, 134kD), *cry11Aa* (*cryIVD*, 72kD) 和 *cyl1Aa* (*cylA*, 27.3kD)。编码这些蛋白的基因却在同一质粒 (72MD) 上, 并在芽孢形成期高度表达, 从而导致形成一个直径约 $1\mu\text{m}$ 的大球状伴胞晶体 (Ibarm 和 Federici, 1996)。这

些单一蛋白的毒力比复合蛋白晶体的毒力低 10 到 100 倍。这说明, 这些蛋白质之间具有协同作用, 而 *Cyl1Aa* 蛋白, 由于它的疏水特性和作为伴胞晶体中的优势蛋白而起着特别重要的协同作用 (Wu and Chang, 1985; Wu and Federici, 1993)。经进一步深入的分析这些基因的核苷酸序列, 发现它们都有各自的启动子, 并在结构和功能上有很大的差异。

前面已经提到 *cry4Aa* 和 *cyl1Aa* 具有非重叠的双启动子, *cry4Ba* 具有 BtI 单一启动子, *cry11Aa* (*cryIVD*) 也位于一操纵子中, 其启动子属于重叠的双启动子。在 *E. coli* 的实验中证明, *CytA* 的有效产生需要一个位于 *cry11Aa* (*cryIVD*) 基因下游的、产物为 20kD 辅助蛋白的存在。该 p20 基因的表达由 *cry11Aa* (*cryIVD*) 的启动子控制, 即 *cry11Aa* 和 p20 基因的表达是作为一个转录单位 (Adams 等, 1989) (图 1A)。从杀蚊晶体蛋白基因的种类、存在状态及启动子类型可以看到, *cry* 基因即使在同一质粒上, 其产物的杀虫特异性也相同, 但它们却在启动子的类型及转录率上存在着很大的差异。

综上所述, 苏云金芽孢杆菌杀虫晶体蛋白基因的启动子类型不同, 转录表达的时间、方式、产量等各不相同。

(1) 大多数晶体蛋白能在苏云金杆菌细胞内大量积累, 与其含有 BtI 类强启动子有关, 这类启动子能高效转录 *cry* 基因。

(2) 双重重叠的启动子及非重叠双启动子的存在, 是在转录水平上对同一 *cry* 基因表达在时间和产量调控, 以使晶体蛋白的产生不致于过多或过快而打破细胞生长的平衡。

(3) 不同的 *cry* 基因具有不同类型的启动子, 同一启动子对不同 *cry* 基因的转录效率不同。如 Wu 等 (1995) 利用 *cry1Ac* 的启动子启动 P20 的转录并促进表达 *cry11Aa* (*cryIVD*), 结果 *cry11Aa* 的表达量大大高于 *cry11Aa* 本身启动子的表达 (P20 位于该操纵子内)。这可能与基因所处的位置和启动子结构差异等因素有关。

因此, 深入研究苏云金芽孢杆菌的启动子及有关 *cry* 基因表达的调节因子, 对于了解杀虫晶体蛋白的调控机制, 构建高效的工程菌具有十分重要的理论和实践意义。

参 考 文 献

- [1] 沈 同, 王镜岩. 生物化学. (第二版). 北京: 高等教育出版社, 1996.
- [2] Hofte H, Whiteley H R. Microbiol Rev, 1989, **53**: 242~255.
- [3] Aronson A I. Mol Microbiol, 1993, **7**: 489~498.
- [4] Walter T. Regulation of *cry* gene transcription in *Bacillus thuringiensis* by a DNA binding protein which recognizes DNA elements upstream of the *cry1Ab*, *cry1C* and *cry1D* gene promoters. Thesis for the degree of philosophy in Purdue University, 1995.
- [5] Snyder L, Champness W. Molecular Genetics. ASM Press. Washington, D. C. 1997.
- [6] Wong H C, Schnepf H E, Whiteley H R. J. Biol. Chem.. 1983, **258**: 1960~1967.
- [7] Brown R L, Whiteley H R. Proc Natl Acad Sci USA, 1988, **85**: 4166~4170.
- [8] Brown K L, Whitely H R. J of Bacteriol, 1990, **172**(12): 6682~6688.
- [9] Dervyn E, Poncet S, Klier A *et al.* J Bacteriol, 1995, **177**: 2283~2291.
- [10] Wu D, Federici B A. Appl Microbiol Biotechnol, 1995, **42**: 697~702.
- [11] Yoshisue H, Fukada T, Yoshida K *et al.* J Bacteriol, 1993, **175**: 2750~2753.
- [12] Yoshisue H, Ihara K, Nishimoto T *et al.* FEMS Microbiol Lett, 1995, **127**: 65~72.
- [13] Yoshisue H, Sakai H, Sen K, Yamagiwa M *et al.* Gene, 1997, **185**: 251~255.
- [14] Agaisse H, Lereclus D. J of Bacteriol, 1994, **176**: 474~4741.
- [15] Brown K L. J of Bacteriol, 1993, **175**(24): 7951~7957.