

一个新的细胞粘菌肌动蛋白基因

周世宁

(中山大学生物系, 广州)

提 要

本文分析了一个细胞粘菌(*Dictyostelium discoideum*) 基因组克隆的核苷酸顺序。这段顺序编码 259 个氨基酸, 并含有一基因 5' 端非编码区结构。非编码区结构特点及氨基酸顺序表明这是一个新的肌动蛋白基因。基因组中约有 15 个内切酶消化所得的片段含有与此克隆同源的顺序。

导 言

细胞粘菌 (*Dictyostelium*) 是一种低等真核生物。由于它生命循环简单, 具有象细菌那样的单细胞生长特点和较高等生物那样的分化发育现象, 因而为我们提供了一种研究分化发育的理想系统。单细胞生长的细胞粘菌在一定条件下, 开始聚集成群, 约经一天的发育, 成群的细胞分化成孢子和柄细胞。在这一分化发育过程中, 有数千种基因表达^[1,2]。其中一部分基因已被克隆研究^[3-6]。这些研究丰富了对真核生物基因结构与表达的了解。

这里, 我们研究了一个编号为 79 的细胞粘菌基因组克隆片段。它含有一段与已知肌动蛋白基因相似的顺序, 它们之间有少数核苷酸不同, 所编码的氨基酸顺序却很接近, 它们的 5' 端非编码区结构也有很相似的特点。因此, 被认为是一个新的细胞粘菌肌动蛋白基因。

肌动蛋白存在于多种真核生物细胞中, 含量丰富, 在收缩运动, 细胞分裂, 以及受体分布等方面都有重要的作用和影响^[7-9]。据报道, 细胞粘菌可能有近 20 个肌动蛋白基因^[5]。其中 4 个基因的核苷酸顺序已知道, 另外几个的 5' 端及 3' 端的部分顺序也已清楚^[5,6,10,11]。现有资料表明这些基因是不同源的。不同的基因在发育

过程的不同阶段有不同的表达水平^[12, 13], 有的基因在细胞进入分化发育时才表达, 有的却关闭。这些观察暗示了肌动蛋白基因的多样性可能在细胞粘菌特殊的生命循环中, 对实现基因表达的调控是必要的。果蝇只有 6 个肌动蛋白基因, 酵母只有一个^[12], 而细胞粘菌的肌动蛋白基因数目相对大得多。研究这些基因的结构与表达对了解基因进化, 调控机理和深入了解发育分化过程都会得到启发。

材 料 和 方 法

一、材料 细胞粘菌菌株 *D. discoideum* Ax-3, 细胞粘菌 DNA 克隆 79 等实验材料由美国伊利诺大学生化系钟森文博士提供。内切酶, T4 DNA 连接酶, DNA 聚合酶 I 大片段, M13 顺序分析引物 (17 寡聚核苷酸) 为 New England Biolabs 公司产品; [α -³²P]-d ATP 为 Amersham 公司产品。

二、内切酶图谱和 DNA 亚克隆操作 按常规方法进行^[14]。

三、DNA 核苷酸顺序分析 基本原理为 Sanger 双脱氧链终止法^[15]。载体选用 M13_{mp} 18 和 M13_{mp} 19^[16]。现将经改进的步骤简述如下: (1) 小片段 DNA 的制备; (2) 克隆小片段 DNA 到 M13; (3) 制备双股 (RF) M13 和单

股 M13 样板: 用 50 μ l 生长过夜的大肠杆菌 (JM101) 培养液与一个重组 M13 噬菌斑接种 2ml 肉汤培养基, 在转动的试管中培养 6 小时 (37 $^{\circ}$ C), 离心, 1.2ml 和 0.3ml 上清液分别留作单股 M13 制备和互补试验。沉淀细胞用于 RF M13 的制备, 方法与质粒制备相似^[14]。单股 M13 样板制备: 1.2ml 上清液加到 300 μ l 2.5 mol/L NaCl 和 20% PEG (聚乙二醇) 溶液中, 于冰浴放置 30 分钟, 离心去上清, 沉淀溶于 TE 缓冲液, 然后经酚抽提, 乙醇沉淀纯化; (4) 用 RF DNA 酶切检查插入片段; (5) 为了确保选作顺序分析用的二个重组 M13 含有方向相反的外源片段, 通常需对这对重组 M13 单股 DNA 作互补试验。如两个 M13 含有方向相反的插入片段, 其单股会因为互补而连系在一起, 在电泳时移动较慢。当外源 DNA 是用单一内切酶制备并插入 M13 的单一位点时, 此步不可缺少。互补试验方法: 从上述保留的 2 种 M13 上清液各抽取 20 μ l 与 2 μ l SDS (2%) 和 6 μ l 染料混合, 于 67 $^{\circ}$ C 保温 1 小时, 冷却后电泳分析; (6) 顺序分析反应: 5 μ l 样板 DNA, 0.1 μ l M13 顺序分析用引物与 1.5 μ l 缓冲液于 63 $^{\circ}$ C 保温 2 小时, 加入 [α - 32 P]-dATP 及 DNA 聚合酶 I 的 Klenow 片段, 混合物分配到含适量 dNTP 和 dd NTP 的四个管中, 一般在室温下反应, 有时在稍高的温度下进行反应以克服由于某些特殊结构造成的问题; (7) 聚丙烯酰胺高压电泳; (8) 放射自显影和核苷酸顺序释读。

四、细胞粘菌基因组 DNA 制备 细胞培养法按照已报道的步骤进行^[17]。在 MES-HL5 培养基中生长至 10^6 细胞/ml 密度的粘菌被离心收集, 重悬于缓冲液 (50mmol/L Tris, 10 mmol/L NaCl, 2mmol/L MgCl₂, 3mmol/L CaCl₂, 0.5mol/L 山梨糖醇, 0.3% Cemulsol, 0.3% Triton x-100) 中, 停留 1 至 2 分钟, 离心收集细胞核, 沉淀重悬于水中并加入 1/4 体积的 Saewsil (20%), 添加 CsCl₂ (使浓度达 0.96 克/ml) 和溴化乙锭, 于 45,000 r/min 离心 24 小时, 乙醇沉淀法纯化 DNA。

五、Southern 杂交 基本方法如前所

述^[14]。程序简述如下: 基因组 DNA 用适当的内切酶处理, 接着电泳分离 DNA 片段, 含 DNA 的琼脂糖凝胶经 HCl 和 NaOH 先后处理, 转移核酸到硝酸纤维滤膜上, 干燥, 于 40 $^{\circ}$ C 预杂交 4 小时, 加入 32 P 标记的 DNA 探针, 40 $^{\circ}$ C 保温 2 天, 用 0.1% SDS 和 0.1 $\times$ SSC 溶液于 55 $^{\circ}$ C 漂洗, 然后置于 -70 $^{\circ}$ C 放射自显影。

32 P 同位素标记的 DNA 探针制备通过缺口翻译法进行^[14]。

结 果

克隆 79 的内切酶图谱、亚克隆和顺序分析策略

细胞粘菌基因组克隆 79 最初从以 λ gt WES λ B 为载体的基因文库中选出^[17], 随后, 其外源片段被转移到质粒载体 pMB9 的 EcoRI 位点, 全长约 8Kb。经杂交分析 (结果未发表), 相应的 mRNA 被定位于 Hind III 到 Bgl II 的区域, 这样, 研究的注意力将集中在 XbaI 到 BglII 之间的名为 79A 的片段 (图 1)。这一片段含有 1 个 HindIII 切点, 2 个 HpaII 切点, 1 个 TaqI 切点。DNA 顺序分析应用了下列亚克隆片段: XbaI-HindIII, HindIII-TaqI, TaqI-BglII, XbaI-HpaII, HpaII-HpaII, HpaII-BglII。

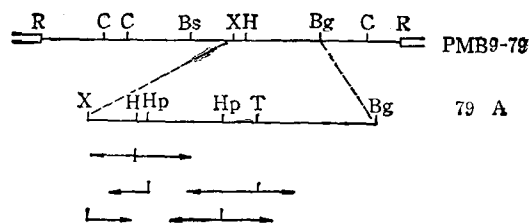


图 1 79A 的内切酶图及顺序分析策略

箭头表示顺序分析的片段和方向
内切酶代号: R, EcoRI; C, ClaI; Bs, BstEII; X, XbaI; H, HindIII; Bg, BglII; Hp, HpaII; T, TaqI

79A 编码区核苷酸及氨基酸顺序

图 2 列出了已经分析的 79A DNA 顺序。从转译起始密码子 ATG (mRNA 上为 AUG) 起, 编码区的 780 对核苷酸顺序已被读出。在密码子 ATG 下游约 20 个核苷酸位置处, 有一

-20

核苷酸顺序为转录股的互补链，图中 T 改为 U 即为 mRNA 顺序。核苷酸编号时把转录起始密码 ATG 的 A 定为 1。氨基酸编号时是把紧跟转录起始密码后面的 Asp 定为 1。

核苷酸顺序为转录股的互补链，图中 T 改为 U 即为 mRNA 顺序。核苷酸编号时把转录起始密码 ATG 的 A 定为 1。氨基酸编号时是把紧跟转录起始密码后面的 Asp 定为 1。

内切酶 HindIII 切点,这是已知肌动蛋白基因特征性保守顺序之一。对比已知的肌动蛋白基因 5'端编码顺序^[5,11],79A 中有些核苷酸顺序是与之不同的。如表 1 所示,这些不同的核苷酸多数是在密码子的第三位上。这种情况,如基因 B1; A8 和 A13 等^[5,11],并没有改变其所编码的氨基酸顺序的一致性。然而,79A 与基因 A3-Sub1, A3-Sub2, AM6 和 A12 相比,有较多的核苷酸不同,并且这些不同出现在密码子的第一或第二位上,这样 79A 和这些基因所编码的氨基酸也就发生了较多的变化。

表 1 79A 编码区与已知基因比较

肌动蛋白基因	与 79A 不同的核苷酸			氨基酸变化	
	密码子 1 位 2 位 3 位	总数	%	总数	%
B1	7	7	4.6	0	0
A8	19	19	2.4	0	0
A3-Sub 1	8 3 76	87	11.1	9	3.4
A3-Sub 2	23 13 90	126	16.1	35	13.4
A12	1 2 47	50	6.4	1	0.3
AM 6	2 8	10	6.4	2	3.9
A13	8	8	2.8	0	0

注: 基因名称 B1 据文献^[11],其它据^[5]。各基因已知的核苷酸顺序并用于此处比较的长度是: B1, 150bp; AM6, 155bp; A13, 285bp; 其余为 780 bp。

和其它肌动蛋白基因一样,在 79A 编码区里没有发现内含子(intron),而在其它细胞粘菌基因已发现有内含子结构。

把从 79A 推算出来的 259 个氨基酸顺序直接与一个取自细胞粘菌肌动蛋白的氨基酸顺序^[18]作比较,其中 257 个氨基酸是相同的,只有 2 个不同,它们是 225 位的 Ala 和 228 位的 Gln。在上述蛋白中,这两个位置分别是 Gln 和 Ala。在已知顺序的基因中,就这二个位置的氨基酸而言,79A 与 A8, A12, A3-Sub1 都是相同的, A3-Sub2 的 225 位相同,而 228 位的 Gln 变成了 Leu。由此可见,这个已知的肌动蛋白并不是上述基因的产物。

79A 的非编码区结构

大部分真核细胞基因的启动子都含有一段普遍相似的顺序,称 TATA 盒。对细胞粘菌

基因初步研究知道,他们的 TATA 盒有自己的特征,其结构是 TATAAA^TA^[9]。活性基因中此段顺序的变化很少超过 2 个碱基。如图 2 所示,79A 的 5'端转译起点 ATG 的上游—66 个碱基处有一段 TATA 盒结构 CATAAAAA,在这一顺序的下游相隔 2 个嘌呤碱基处,可以找到一段由 19 个 T 组成的多聚 T 顺序——这也是细胞粘菌基因的特征顺序。在紧靠转译启动密码子 ATG 的 5'端,有连在一起的 4 个碱基 A。整个非编码区是极端 A-T 富集的,它含有 90%以上的 A-T 成份。所有这些特点均符合细胞粘菌基因的结构特点,也与已知的肌动蛋白基因 5'端结构相似^[6,11]。但它们也有不同之处,79A 的 TATA 盒以 C 碱基开头,其多聚 T 顺序较其它基因长些,而 TATA 盒与聚 T 结构之间只隔 2 个碱基。从 TATA 盒到起始密码子 ATG 的距离也与其它基因不同,有 65 个碱基。在 -140 处有一顺序 AATGGGA-TT,与此相似的结构可在基因 A9, A13, A2-Sub1, A6, A7 和 A12 的 5'端上游找到^[6]。

79A 在基因组中的同源片段

图 3 (见图版 III)是 Southern 杂交的结果。探针是 79A 上的片段 HindIII-BglII。经 TaqI 和 BglII, HpaII 和 BglII 或 HindIII 和 BglII 双酶消化的细胞粘菌基因组 DNA 与 ³²P 标记的探针杂交,可以看到约有 15 条杂交带,暗示着基因组约有 15 个左右与 79A 片段相似的区域。这些片段大小不相同,表明它们是不完全同源的,或者它们分布在基因组不同的环境里,因而使它们有着不一致的邻接的 DNA 顺序。事实上,比较 79A 和已知基因(图 1,文献^[5]),它们的内切酶图谱确是不相同的。

讨 论

79A 的核苷酸及氨基酸顺序表明它是一个肌动蛋白基因 5'端结构。这一 DNA 顺序包括了该基因的 5'端非转译区及 260 个氨基酸的编码区(包括了起点 MET)。据初步测定,79A 的 mRNA 约有 1200 至 1400bp 长(未发表结果)。

因此,其所编码的氨基酸应约为 400 个。与已知肌动蛋白的氨基酸数 375^[10,18]接近。这样,已被分析的 79A 的氨基酸约占整个基因产物的 $\frac{2}{3}$ 。虽然,79A 与其它肌动蛋白基因的已知顺序比较,核苷酸有多处不同源,但其氨基酸顺序却是较少变化的(表 1)。特别是出现频率高的密码子第 3 位上的变化,很少引起氨基酸的改变。这种通过密码子简并原理保持基因产物稳定的现象,可能使肌动蛋白经受了强烈的突变选择。

从表 1 知道,79A 与基因 A3-Sub1, A3-Sub 2, M6 和 A12^[9] 所编码的氨基酸顺序存在差别;由于缺少更多的数据,也不能确定 79A 与基因 B1, A8, 和 A13^[5,11] 的产物是否完全一样,尽管他们已比较的氨基酸顺序是一致的。从编码区核苷酸顺序看,79A 与 A8 较为相似,两者间只有 2.4% 的变化。其次是 A13。然而,79A 的 5' 端非编码区结构与其它基因均有明显不同。

杂交分析得知细胞粘菌基因组约有 15 个左右与 79A 有相似顺序的片段。这些片段数可能代表了肌动蛋白基因的大约数目。前报道估计这种基因数目为 17^[10] 或 20^[6], 差异可能由于所用的分析条件不同,或应用了不同的内切酶所致。细胞粘菌含有如此多的肌动蛋白基因是有趣的。在较高等的生物果蝇中,有 6 个肌动蛋白基因,因它们有组织特异性,因而基因有相应的不同^[19]是容易理解的。粘菌是一个相对简单得多的生物,还不清楚为什么需要更多的基因。一种解释是,这些基因在粘菌分化发育的不同阶段,各基因有不同的表达水平^[10,13,18], 基因表达的调节要通过基因非编码区控制。因此 5' 及 3' 端非编码区的不同^[10], 可能是表达水平调控所必需的。另一解释是,细胞粘菌在发育

的某些阶段,肌动蛋白及其 mRNA 在短时间内急剧升高^[12],要使基因产物这样快速的增加,单个基因可能难以实现。另外,在已知的细胞粘菌肌动蛋白基因中,还发现有象 2-Sub2 那样的假基因^[10,18]。肌动蛋白多基因的表达和调控问题仍然不很清楚,需要更多的研究和调查更多的基因。

本文得到美国伊利诺大学生化系钟森文博士的指导和支持,谨致谢意。

参 考 文 献

- [1] Loomis, W. F.: *Dictyostelium discoideum, A Developmental System*, Acad. Press, New York, 1975.
- [2] Blumberg, D. et al.: *Devel. Biol.*, 1980, **78**, 285.
- [3] Kimmel, A. R. et al.: *Nuc. Acids Res.*, 1983, **11**, 541.
- [4] Poole, S. J. et al.: *J. Mol. Biol.*, 1984, **172**, 203.
- [5] Romans, P. et al.: *J. Mol. Biol.*, 1985, **186**, 321.
- [6] Romans, P. et al.: *J. Mol. Biol.*, 1985, **183**, 311.
- [7] Clarke, M. et al.: *Annu. Rev. Biochem.*, 1977, **46**, 797.
- [8] Korn, E. D.: *Proc Natl. Acad. Sci. USA* 1978, **75**, 588.
- [9] Loomis, W. F.: *The Development of Dictyostelium*, Acad. Press, New York, 1982, 233—324.
- [10] Mckeown, M. et al.: *J. Mol. Biol.*, 1981, **131**, 593.
- [11] Firtel, R. A. et al.: *Proc. Natl. Acad. Sci., USA*, 1979, **76**, 6206.
- [12] Mckeown, M. et al.: *Cell*, 1981, **24**, 799.
- [13] Romans, P. et al.: *J. Mol. Biol.*, 1985, **186**, 337.
- [14] Maniatis, T. et al.: *Molecular Cloning, a Laboratory Manual*, Cold Spring Harbor Laboratory, New York, 1982.
- [15] Sanger, F. et al.: *Proc. Natl. Acad. Sci. USA*, 1977, **74**, 5463.
- [16] Messing, J.: *Methods in Enzymology* 101 (part C), *Recombinant DNA* (R. Wu, et al eds), Academic Press, New York, 1983, 20—78.
- [17] Chung, S. et al.: *Cell*, 1981, **24**, 785.
- [18] Vandekerckhove, J. et al.: *Nature*, 1980, **284**, 475.
- [19] Fyrberg, E. A. et al.: *Cell*, 1983, **33**, 115.

[本文于 1987 年 9 月 7 日收到]

王理开等：“DNA “拓扑学与拓扑异构酶”一文的附图 12

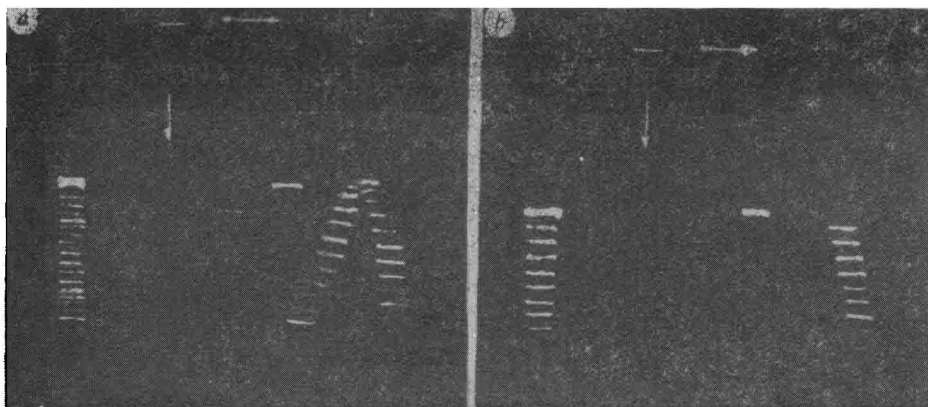


图 12 双向电泳鉴别正、负超螺旋 DNA

- a. 对照,含正和负超螺旋 DNA,左侧为单向电泳,右侧为双向电泳,DNA 呈拱形分形,拱形左翼为负超螺旋 DNA,右翼为正超螺旋 DNA
- b. 实验,左侧为单向电泳,右侧为双向电泳,只见存在拱形右翼,说明样品只含正超螺旋 DNA

周世宁：“一个新的细胞粘菌肌动蛋白基因”一文中的图 3

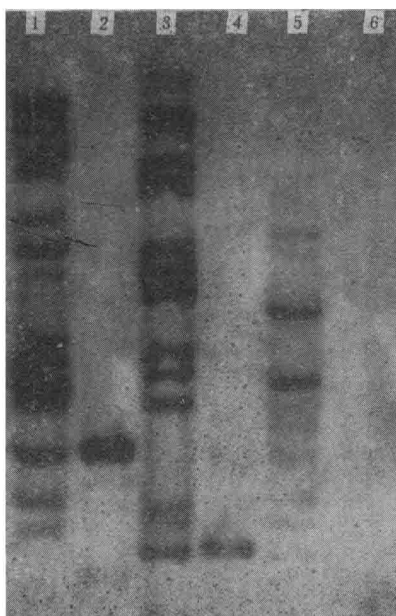


图 3 79A 片段与细胞粘菌基因组 DNA 的 Southern 杂交

样品 1, 3, 5 均为基因组 DNA, 分别被下列各组酶消化: HindIII/Bgl II, HpaII/Bgl II, TagI/BglII, 样品 2, 4, 6 为 pMB9—79 DNA, 依次分别被上述各组酶消化, 作为对照。探针为 79A 的 Hind III-Bgl II 片段