

高等植物核基因结构研究进展

朱 群 白永延

(中国科学院上海植物生理研究所)

与动物细胞核基因一样,高等植物的核基因在5'端、3'端及编码区都有精确的结构^[1,2],典型的植物细胞核基因结构具有以下一些特点:

一、5'端结构

这是基因表达顺式作用的主要区,结构极为复杂。由RNA聚合酶II负责转录的基因的5'端常包括翻译起始密码、转录起始位点、TATA盒、CCAAT盒、增强子等区域。

1. 翻译起始密码ATG 核基因编码的蛋白质翻译均在基因转录产物mRNA的5'端AUG(相当于DNA的ATG)处开始,编码甲硫氨酸。植物核基因对起始密码ATG附近的核苷酸碱基有所偏爱,如大麦B₁麦醇溶蛋白的基因ATG附近的顺序为CCACCATG^[3],玉米C₁基因ATG附近为GCCATGG^[4],这与动物翻译起始密码附近的DNA顺序(C₆CCATGG)同源。这段顺序是转录产生的mRNA链与核糖体结合的位点^[5-7]。在研究蛋白质翻译时将ATG的A处定为1,上游为负值,下游为正值。统计分析表明,95%真核生物的核基因翻译起始密码的ATG上游的-3位置均为嘌呤碱基,+4通常是G^[8]。1987年Taylor等利用赛氏粘菌几丁质酶的结构基因(Chi)和能在植物细胞中组成型表达的胭脂碱合成酶基因

(Nos)5'端和3'端构成融合基因,对ATG上游-3位和下游+4位进行了优化研究。结果表明,当-3位由A替换原来的C,细胞的几丁质酶含量提高2倍,如同时在+4位再由G替换C,则几丁质酶含量上升8倍,其mRNA含量上升4倍,证明蛋白质翻译和DNA的转录均受影响^[8]。

2. 转录起始位点(加帽位点) ATG密码上游-40—70bp左右为高等植物核基因转录起始点,大多数情况下其周围为嘧啶碱基的腺嘌呤残基,少数情况下为鸟嘌呤残基。转录起始位点发生变化或缺失时,影响转录效率甚至其发生。

编码鱼精蛋白的H₂A基因的TATA盒和转录起始位点间插入1—3bp片段DNA时,不影响基因体外转录,但影响体内外转录的正确性,为使转录正确而有效地进行TATA盒和转录起始位点在形成双螺旋结构时需在同一平面上,如发生缺失、插入或点突变,转录就从原起始处的A残基上开始^[9]。Jones等发现,为了不降低植物基因转录水平,在ATG和转录起始点之间应有合适的距离^[10]。

3. TATA盒 TATA盒通常位于转录起始点上游-25bp处的非编码链上。植物核基因的5'端也具有此盒,人们把它看作是原核

表 1 禾谷类植物基因的5'端区域的保守顺序

基因克隆	CATC 盒	TATA 盒	加帽位点
大麦醇溶蛋白基因 pBHR 184	- 139 ACATCCAAACA	- 80 CTATAAATA	- 51
麦醇溶蛋白基因 pW 8233	- 165 GCATCCAAGCA	- 105 CTATAAATA	- 67
Z 21 玉米醇溶蛋白基因 pML 1	- 177 CCATCTATAACC	- 112 GTATAAGCA	- 65, - 52
Z 19 玉米醇溶蛋白基因 ZG 99	- 168 TCATCTCTACC	- 90 GTATAAATA	- 57
玉米 Adh 1 基因 p 1S.1	- 167 CCATCTCTTCC	- 140 CTATATAAA	- 99
小麦 H ₃ 组蛋白基因 pTH012	- 201 CCATCTCGACC	- 97 CTATTTAAC	- 62
小麦 rbcS 基因 pWS 4.3	- 113 CCATCCCAACC	- 79 CTATATATA	- 42

表 2 高等植物中 rbcS 基因的 -140 区内“GT”顺序比较

基 因	顺 序	作者或文献
豌豆, rbcS-E 9	GTGTGGTTAATATG	Coruzzi et al. (1984)
3 A	GTGTGGTTAATATG	[19]
3 C	GTGTGGTTAATATG	[19]
3.6	GTGTGGTTAATATG	Herrera-Estrella et al. (1984)
大豆, SRS 1	GTGTGGCCAATATA	Berry-Lowe et al. (1982)
烟草, Ntss 23	GTGTGGATATTAAG	Mazur & Chui (1985)
N. plumbagini- folia rbcS-8 B	GTGTGGATATTAAG	C. Poulsen (未发表资料)
矮牵牛, SSU 8	GTGTGGATATTAATA	Tumer et al. (1986)
SSU 11	ATGTGGCCATTAAT	Tumer et al. (1986)

生物基因的 Pribnow 盒启动序列的同功部分。离体试验结果表明, TATA 盒影响转录的效率, 但它只是基因启动子的组分之一, 对 RNA 聚合酶 II 在转录起始位点的定向起重要作用, 从而影响基因转录效率。Kaulen 证明, 只有 TATA 盒时, 植物的查尔酮合成酶基因不能发生表达^[11]。

4. CCAAT 盒 在植物核基因转录起始点上游 -90 bp 处的典型顺序为 5'-GC-CCA-AT-3', 但禾谷类贮藏蛋白基因中无 CCAAT 盒, 被 CATC 盒取代(表 1)。CCAAT 或 CATC 盒也是植物启动子的组分之一^[3]。它们的效应可能是通过 RNA 聚合酶 II 或影响 DNA-蛋白质空间结构使 RNA 聚合酶 II 更易于迅速地转录, CCAAT 可能是 RNA 聚合酶 II 识别的位点, 使之容易与被转录基因的启动子结合, 形成转录起始复合物。有些植物基因, 如玉米醇溶蛋白基因, 在 TATA 盒和 CCAAT 盒之间存在有 AGGA 保守区, 该区是否起协调两盒的顺式作用还有待进一步研究^[12, 13]。

5. 增强子 是指基因的一些顺序无论是顺向还是反向, 在结构基因的 5'端或 3'端, 对基因的表达都有增强作用。增强子首先在 SV₄₀ 基因中发现^[14-17]。真核生物基因的增强子至少有两类, 一类与结构基因的大量转录有关, 另一类影响转录的水平和特异性表达。目前植物增加子的研究主要集中在后一类。

Simpson 等在研究叶绿素 a/b 结合蛋白基因(1hcp)5'端调控顺序时发现, AB 80 基因 5'端上游 -100 到 -347 区域的 DNA 片段能使植物中 Nos-NPT II 融合基因增强表达, 而且受光调控。如串联两个这样的区域, Nos 启动子转录效应更明显。这一区域还能使 Nos-NPT II 基因不表达^[18]。可认为, 在增强子的 DNA 片段中可能有两个功能不同的区域或同一区域在不同组织中起不同作用, 即增强子和沉默子。

Fluhr 等(1986)发现, rbc S 3A 5'端上游的 280 bp (-327 到 -48) 的顺向或是反向的都能使植物中 CaMV 35S RNA 启动子受光调控^[19]。rbc S 基因的 5'端至少有一区域与该基因的光

调控有关,在5'端约-140 bp处都有“GT”盒(表2),此盒与SV40增强子的核心顺序同源,因而有可能起相同的作用^[20,21]。

Voelker等(1987)研究了大豆细胞凝集素的两个等位基因,一个基因d1e 2来源于能合成白细胞凝集素的大豆,另一个基因pd1e 2来自不能合成凝集素的大豆,它们在转化烟草中表达情况不同,d1e 2基因比pd1e 2基因的转录水平高50倍。顺序分析结果表明,pd1e 2基因的5'端上游-114 bp处缺失,d1e 2基因的该区域含有4个特异结构:63 bp的“A”重复顺序,16 bp的倒转重复结构,种子蛋白基因都有“CATCAT”保守顺序和豆科植物贮藏蛋白基因具有豌豆球蛋白基因“GCCACCTCA”保守顺序^[22]。这些结构肯定与该基因的表达调控和特异表达有关。

Stougarrd等(1987)发现,大豆血红蛋白lbc 3基因5'端上游-100到-950区域存在有强的增强子,能使CAT基因在植物中表达增强20倍,-139到-102区域决定该基因组织特异性的表达^[23]。所以lbc 3基因表达受多层次的调控,起组织特异性调控区域可能是协调不同增强子作用的环节,也可能不同增强子通过途径起作用。

豌豆热休克蛋白基因hs 6871的5'端有两类顺序调控基因的表达:一类为热休克调控顺序hsE,在hs 6871中有6个这种顺序,其中位于-181到-154的两个交错的hsE决定该基因热诱导;另一类与增强基因表达有关,位于-439到-298间的顺序具有增强子的效应,这段序列正、反向都能极显著地增强hs 6871基因的表达。结构分析表明,该序列中有与酵母增强子相似的14个A连续的顺序,另有一以-407区为中心的34 bp不完全轴对称结构^[13,24]

查尔酮合成酶基因、玉米醇溶蛋白基因、蔗糖合成酶基因、燕麦和南瓜的光敏色素基因中都有与动物基因增强子同源顺序的存在^[11,25,26]。已知植物基因增强子与动物、病毒基因5'端增强子一样是位于TATA盒和

CCAAT盒的上游,有的距TATA盒近,也有相距数kb以上,结构复杂,但均具有以下几个特点:

(1) 起顺式作用。

(2) 对基因转录均起增强作用,但有的增强子在不同组织中的功能完全不同。

(3) 位于基因5'端上游的增强子,无论是顺向或是反向都能起增强基因转录的作用,但与动物基因增强子不同,位于结构基因3'末端的增强子通常不起作用,这特点与酵母的增强子相似。

(4) 有些增强子有相似的保守序列,可能是在基因活化前或活化过程中受扭力或结合蛋白的相互作用形成发夹式结构。有的增强子可能形成“Z-DNA”(左旋DNA),可能影响染色质细微结构的变化,从而增强基因的转录活性。

综上所述,高等植物基因5'端结构复杂,执行不同功能的DNA区域被隔开,可人为地分为3个主要区域:基因表达的起始区(即转录和翻译的起始点),决定转录起始点的选择区(TATA盒和CCAAT盒)以及转录水平的调控区(增强子及特异性表达元件)。不同区域不仅结构复杂,而且多态性,但它们之间作用的协调均通过RNA聚合酶II进行。

二、外显子和内含子结构

与其他高等生物基因相似,植物基因通常也有内含子,但有些基因,如热休克蛋白基因和凝集素基因则无内含子^[24-27]。不同植物中编码相同蛋白的基因复杂程度不同,但也有保守性,如拟南芥菜的酒精脱氢酶基因中的6个内含子与玉米的该酶基因的9个内含子中6个的位置完全相同^[28]。此外植物内含子还含有类似哺乳动物和果蝇等基因内含子中的切割信号。经统计分析证明,内含子5'端为一保守区域(GT₀AGT),而且其中的GT两核苷酸碱基的保守性最强。改变这区域结构,不仅影响切割效率,而且导致隐蔽切割位点的产生,如5'端第1,2位的G和T改变时,mRNA前体的

切割加工就不能进行。

内含子3'端也有切割信号[(T/C)_nNCC/TAG/G]，基因的特定内含子3'端切割信号的嘧啶(T/C)_n数一定，当发生缺失突变时，切割效率下降，有时则产生新的切割位点。

高等植物核编码的叶绿体蛋白(RuBP羧化酶)的结构基因5'端有一编码转运肽区，转运肽通常由30个左右的氨基酸残基组成的疏水性的一段小肽，将成熟蛋白穿越叶绿体膜运进功能部位。Schreier等(1985)曾利用rbc S基因启动子和编码转运肽的DNA区域与NPT II构成融合基因后转化烟草，结果NPT II蛋白能被转运肽导入叶绿体^[29]。

Lycett等(1983)分析了多种植物基因密码子利用情况，发现植物与动物基因密码子的偏爱有所不同，在植物中带CG密码子利用频率较低，编码Arg 6个密码中AGA和AGG利用频率较高，不像动物基因对编码Leu的CVC，Ala的GCC，Lys的AAG和Glu的CAG密码那样偏爱^[30]。

三、3'末端结构

植物基因3'末端是终止基因转录的功能区，基因转录的终止是3'末端不同功能区协调作用的结果。

1. 终止密码 植物基因翻译的终止密码为TAA、TAG和TGA，如玉米花色素苷调控基因C1的3'端有TAG、TAA两个终止密码^[4]。

2. AATAAA加尾信号 植物基因的3'下游、poly A加尾位点上游也有AATAAA保守序列，数量为1个或多个不等，有的两个加尾信号交错形成AATAAATAAA序列。有的加尾信号附近还有GATAAA或AATAAGAAA顺序出现。有的与动物基因一样在AATAAA下游有一个TTTTCCTGTC保守顺序出现，其作用尚不清楚^[30]。AATAAA序列发生点突变或缺失，均导致mRNA加尾不能正常而有效地进行。加尾信号的上游和下游的其他顺序也起协调作用。

基因转录的终止处位于poly A加尾的开始点上，终止后加尾，产生的成熟mRNA运入细胞质发挥作用。

在动物红细胞β-珠蛋白基因及组蛋白H₂基因的3'端发现有增强子存在，增强子与基因的特异性表达有关^[31,32]。虽然迄今在植物基因3'端还未发现有类似的调控顺序，但启发我们今后应注意3'端在基因表达调控中的作用。

摘 要

本文较系统地介绍了高等植物核基因的5'末端和3'末端的结构研究新进展，并着重介绍了增强子结构及其对基因表达的调控作用。

参 考 文 献

- [1] Dynen W. S. et al., 1985, *Nature*, 316: 774—778.
- [2] Reudelhuber T., 1984, *Nature*, 312: 700—701.
- [3] Forde B. G. et al., 1985, *Nucleic Acid Res.*, 13: 7327—7339.
- [4] Paz-Ares J. et al., 1987, *The EMBO J.*, 6: 3553—3558.
- [5] Kozak M., 1986, *Cell*, 44: 283—293.
- [6] Lutcker H., et al., 1987, *The EMBO J.*, 6: 43—48.
- [7] Green P. J., et al., 1987, *The EMBO J.*, 6: 2543—2549.
- [8] Taylor J. L., et al., 1987, *Mol. Gen. Genet.*, 210: 572—577.
- [9] Kovacs B. J., et al., 1986, *Nucleic Acids Res.*, 14: 2429—2442.
- [10] Jones J. D. G., et al., 1985, *The EMBO J.*, 4: 2411—2418.
- [11] Kaulen H., et al., 1986, *The EMBO J.*, 5: 1—8.
- [12] Kreis B. M., et al., 1986, *Phil. Trans. R. Soc. Lond. B* 314: 355—365.
- [13] Heidecker G. and Messing J., 1986, *Ann. Rev. Plant Physiol.*, 37: 439—466.
- [14] Voss S. D., et al., 1986, *TIBS*, 11: 287—289.
- [15] Struhl K., 1987, *Cell*, 49: 293—297.
- [16] Mitchell P. J., et al., 1987, *Cell*, 50: 847—861.
- [17] Dynan W. S., 1987, *TTG*, 3: 121—122.

- [18] Simpson J., et al., 1986, *Nature*, 323: 551-553.
- [19] Najy B. F., et al., 1986, *Phil. Trans. R. Soc. Lond. B* 314: 493-500.
- [20] Fluhr R., et al., 1986, *Science*, 232: 1106-1112.
- [21] Kuhlemeier C., et al., *Ann. Rev. Plant Physiol.* 38: 221-257.
- [22] Voelker T., et al., 1987, *The EMBO J.*, 6: 3571-3577.
- [23] Slougaard J., et al., 1987, *The EMBO J.*, 6: 3565-3570.
- [24] Schoffl B. F., et al., 1986, *Phil. Trans. R. Soc. Lond. B* 314: 353-368.
- [25] Hershey H. P., et al., 1987, *Gene*, 61: 339-348.
- [26] Chen Z. L., et al., 1988, *The EMBO J.*, 7: 297-302.
- [27] Goldberg B. R. B. 1986, *Phil. Trans. R. Soc. Lond. B* 314: 343-353.
- [28] Chang, et al 1986, *Proc. Natl Acad. Sci. USA.*, 83: 1411-1412.
- [29] Schreier P. H. et al., 1985, *The EMBO J.*, 4: 25-32.
- [30] Lycett D. W. et al., 1983, *FEBS Letters*, 153: 43-46.

真核基因反义 RNA 研究进展

金 健 健

(中国科学院上海细胞生物学研究所)

反义 RNA (antisense RNA) 是指能与特定的 mRNA 互补的 RNA 片段, 1981 年 Tomizawa 首先报道^[1]。目前已发现在原核生物中存在不少天然的反义 RNA 系统, 它们对基因的复制、转录和翻译起着重要的调节控制作用^{[2][3]}。在真核生物中, 至今尚未观察到有天然的反义 RNA 存在。但通过不同的途径, 把体外人工构建的反义 RNA (或基因) 引入到细胞内进行基因调控的研究, 已有不少报道。本文介绍近几年来这方面工作的一些进展。

一、反义 RNA 的作用

1. 表达反义 RNA 的基因质粒

Izant 和 Weintraub^[4]最先提出了真核细胞反义 RNA 研究的重要性并建立了一个模型系统。他们将单纯疱疹病毒 (HSV) 的胸腺嘧啶核苷激酶 (TK) DNA 反向插入在 HSV-TK 启动子或小鼠肉瘤病毒 (MSV) LTR 与 SV 40 polyA 信号之间, 获得 HSV-TK 的反义 RNA 基因重

组质粒。将上述反义 HSV-TK 基因质粒和 TK 基因质粒以 100:1 引入到小鼠 LTK⁻细胞, 结果发现, 与单独引入 TK 基因质粒的对照组相比, 细胞内 TK 的瞬时表达大大下降。这个实验表明: 一个外源反义 RNA 基因可以抑制真核细胞内特定基因的表达。在上述工作的基础上, 其它一些基因的反义 RNA 基因质粒也进行了构建并研究, 如鸡的 TK 基因^[5], 小鼠 β -肌动蛋白基因^[5], 劳氏肉瘤病毒 (RSV) env 基因^[6], 果蝇的 hsp 26 基因^[7], 氯霉素乙酰转移酶 (CAT) 基因^[5], c-fos 基因^[8]和 c-myc 基因^[9]等等。

考虑到某些基因的抑制对细胞可能产生致死效应, 或实验要求反义 RNA 只在特定的阶段表达, 不少反义 RNA 重组质粒采用了可诱导的启动子^{[5][9]}。较为常用的是小鼠乳腺肿瘤病毒 (MMTV) 启动子, 在地塞米松的诱导下, 带有上述启动子的反义 RNA 基因可在细胞内表达, 其表达量在一定范围内随药