

植物分子生物学丛书 2

植物基因与分子操作

Plant Genes and Molecular Manipulations

顾红雅 瞿礼嘉 明小天 潘乃穰 陈章良 编著

北京大学出版社



植物分子生物学丛书-2

1924/07
植物基因与分子操作

Plant Genes and Molecular Manipulations

顾红雅 瞿礼嘉 明小天 编著
潘乃穧 陈章良

北京大学方正中青年学术骨干培养奖励基金部分资助



北林图 A00105722



北京大学出版社

北 京

436724

新登字(京)159号

图书在版编目(CIP)数据

植物基因与分子操作/顾红雅等编著. — 北京:北京大学出版社, 1995.4

(植物分子生物学丛书)

ISBN 7-301-02698-6

I. 植… II. 顾… III. 植物-遗传工程 IV. Q943

书 名: 植物基因与分子操作

著作责任者: 顾红雅 翟礼嘉 明小天 潘乃瑾 陈章良

责任编辑: 李宝屏

标准书号: ISBN 7-301-02698-6/Q·62

出版者: 北京大学出版社

地 址: 北京市海淀区中关村北京大学校内 100871

电 话: 出版部 2502015 发行部 2559712 编辑部 2502032

排 印 者: 北京大学印刷厂

发 行 者: 北京大学出版社

经 销 者: 新华书店

787×1092毫米 16开本 19.25印张 482千字

1995年4月第一版 1995年4月第一次印刷

定 价: 25元

前 言

1991年8月我们实验室举办了一个烟草基因工程培训班,一个月以后又与农业部热带作物生物工程重点实验室在海南举办了东南亚生物技术培训班,当时使用的是我们仓促写成的初步讲稿。在教学和实验过程中,我们迫切地感到需要这样一本教材,它既要讲透分子生物学基础理论,可以用作大学生和研究生的课本及其他研究人员的参考书;又要对基因操作的实验原理及操作过程有个明晰的阐述解释,在实验室里可以指导实验,并且应该以植物为研究对象的。目前符合以上几条要求的书国内不多,国外也少见。基于这样的考虑,我们着手编著了这本书。本书是植物分子生物学丛书的第二册,主要向读者介绍目前对于植物基因的一些基本研究方法和研究进展;对于基本原理,我们力求深入浅出,并大量使用图或表来辅助阐述,让初学者也能看懂;对于实验操作,我们则力图详细、清晰,使研究人员可以马上着手进行实验。

本书1992年开始筹划,1993年正式动笔。由于植物分子生物学各领域理论及技术研究的发展极为迅速,我们对1991年的初步讲稿作了彻底的更新和全面的补充,大大加重了基础理论的分量,并且增加了大量操作程序。春来秋去,历经一年的艰辛,本书终于成稿了。我们得到了来自实验室的精神和物质等各方面的支持,在此表示深深的谢意。

值得一提的是,由于我们实验室研究的面越来越广泛,几乎涵盖了植物分子生物学领域的各个部分,因此本书中的绝大部分实验操作程序都是我们实验室已经使用多年和正在使用的,由此我们要对所有在本实验室工作过的、对这些操作程序做出过贡献的朋友们表示衷心的感谢。刘春清老师为本书的制图付出了辛勤的劳动,在此再次感谢。

应该看到的是,本书仍是急就章,肯定会出现这样或那样的疏漏,我们诚恳地希望各位同仁朋友及时给我们指出来,以便进一步完善这本书,使它成为一本真正过硬的好教材。

编 著 者

于北京大学校园

1994年6月4日

《植物分子生物学丛书》是由北京大学生命科学学院蛋白质工程及植物基因工程国家重点实验室组织编写的一套专业性丛书。这套丛书总结归纳了他们自 1987 年建立以来在植物分子生物学领域所从事的多种有关教学和科研活动中逐渐积累起来的一些研究成果和教学资料,力求从植物分子生物学的基础知识、实验技术、理论发展和最新应用成果等多种角度和层次向读者介绍植物分子生物学的基础性和专业性知识,可供理、工、农、医各领域内相关专业的大学生、研究生和研究人员参考。

主 编 陈章良

副主编 李宝屏*

瞿礼嘉

编 委 潘乃璁

顾红雅

朱玉贤

明小天

* 为北京大学出版社,其余编委均属于北京大学生命科学学院蛋白质工程及植物基因工程国家重点实验室。

内 容 简 介

本书是《植物分子生物学丛书》的第二分册,书中详细地阐述了当前植物分子生物学领域的发展状况及研究水平,深入浅出地介绍了研究植物基因的基本方法、原理及实验操作。全书包括四部分共十六章:第一部分介绍植物基因组研究,第二部分介绍植物基因操作基本技术,第三部分介绍植物基因的分析技术,第四部分介绍植物基因的转化。

本书可供理、工、农、医各领域内相关专业的大学生、研究生作教学课本,也可供相关领域的研究人员实验参考。

目 录

第一部分 植物基因组研究

第一章 植物核基因组	3
第一节 植物核基因组概况	3
第二节 核基因的表达和调控	6
第三节 已研究的植物基因	14
第四节 植物核基因基本技术	19
第二章 植物线粒体基因组	28
第一节 线粒体 DNA 的复杂性及多样性	28
第二节 植物线粒体基因结构及其表达调控	32
第三节 线粒体基因与细胞质雄性不育	35
第四节 植物线粒体的进化与起源	36
第五节 植物线粒体 DNA 的分离	36
第三章 植物叶绿体基因组	40
第一节 叶绿体 DNA 的性质	40
第二节 叶绿体基因图谱	41
第三节 叶绿体基因的表达调控	47
第四节 叶绿体 DNA 的分离	48

第二部分 植物基因操作基本技术

第四章 酶技术	53
第一节 核酸酶及应用	53
第二节 连接酶及应用	65
第三节 聚合酶与 DNA 修饰酶	70
第五章 植物 RNA 技术	74
第一节 植物细胞中的 RNA	74
第二节 植物 RNA 的分离	77
第三节 植物组织 mRNA 的分离	87
第四节 植物 RNA 杂交技术	89
第六章 基因克隆技术	94
第一节 基因克隆技术概述	94
第二节 质粒载体	96
第三节 DNA 导入宿主细胞——转化	100
第四节 重组子筛选	102
第五节 基因克隆中的一些基本技术	103

第七章 基因文库技术	108
第一节 噬菌体载体	108
第二节 新型载体 YAC	113
第三节 文库构建	114
第四节 cDNA 的合成	117
第八章 基因扩增技术	121
第一节 聚合酶链式反应 (PCR)	121
第二节 NASBA 技术	138
第九章 转位因子及基因标签法	142
第一节 转位因子导论	142
第二节 什么是转位因子基因标签实验?	143
第三节 植物中已成功的基因标签实验	145
第四节 转位因子的异源系统应用	147
第五节 基因标签法的应用前景	149

第三部分 植物基因的分析技术

第十章 植物基因的定位与结构分析	153
第一节 制定物理图谱	153
第二节 基因的初步定位	158
第三节 RFLP 分析与 DNA 序列测定	160
第四节 原位杂交操作与探针标记	163
第十一章 植物基因的表达研究	167
第一节 基因本身的研究	167
第二节 基因转录的调控研究	171
第三节 克隆基因表达产物的研究	175
第四节 基因调控的基本实验方法	177

第四部分 植物基因的转化

第十二章 土壤农杆菌的转化机理	191
第一节 Ti 质粒的 T-DNA 区	192
第二节 Ti 质粒的毒性区	199
第三节 细菌染色体上的基因	214
第四节 T-DNA 的整合过程	215
第十三章 植物土壤农杆菌转化技术	226
第一节 转化载体系统	226
第二节 中间质粒构建及三亲交配	229
第三节 土壤农杆菌转化方法	235
第十四章 植物原生质体的分离、转化和培养再生	239
第一节 原生质体分离纯化	239

第二节	原生质体的培养方法	244
第三节	影响原生质体培养再生的因素	247
第四节	通过原生质体的基因转化方法	248
第五节	水稻原生质体转化程序	254
第十五章	基因枪技术在植物基因转化中的应用	262
第一节	基因枪转化的原理	262
第二节	基因枪转化的一般步骤	264
第十六章	转基因植物的检测	266
第一节	报告基因检测法	266
第二节	杂交检测法	279

第五部分 附 录

附录 I	简要名词解释	291
附录 II	主要缩写名称解释	295
附录 III	密码子与氨基酸对应表	296
附录 IV	氨基酸名称缩写及分子量	296
附录 V	一个 A_{260} 单位相对应的核酸含量	297
附录 VI	各 λ 分子量标记的大小	297
附录 VII	一些 β 射线放射性同位素半衰期及强度范围	298
附录 VIII	离心机转速与离心力关系表	298

第一部分

植物基因组研究

第一章

生物体因基因

第一章 植物核基因组

一棵植物的发生、发育、生长及开花结果,这些复杂的生理过程每一步都涉及到特异基因的表达,即基因表达生成蛋白质(酶等),蛋白质再直接决定植物每一个细胞的结构与功能。整个基因表达的过程是极其复杂的,因为基因表达不仅存在一个要受到光照、营养、病菌侵染等外界刺激影响的问题,而且在植物细胞中还有一个三个独立的基因组——核基因组、叶绿体基因组和线粒体基因组——如何协同作用和表达的问题。在这三个基因组中,叶绿体基因组和线粒体基因组都无法编码所有其细胞器本身发育所需的蛋白,都需要核基因组编码的酶。因此我们不难看出,作为细胞的“最高司令部”,核基因组在细胞发育过程中扮演着举足轻重的角色。本章主要介绍植物的核基因组,其他两个基因组另列章节介绍。值得一提的是,目前我们对于植物基因组的结构和表达方面的知识大多都是通过基因操作技术获得的,有关这些技术操作我们后面都将予以详细介绍。

第一节 植物核基因组概况

在植物细胞中,核基因组无论是从其DNA的含量,还是从其编码的基因数目来看都是最大的。核DNA、组蛋白和其他一些非组蛋白聚在一起从而形成染色质,核DNA不仅要编码植物细胞中其他各个发育过程所需的蛋白,而且也要编码核DNA自身的转录翻译等过程所需要的蛋白。我们先来看看其自身的发育过程。

所有的发育过程中最基本的一个环节就是核内遗传物质的复制及其在子代细胞中的分配,子代细胞中的DNA应与母代细胞中的DNA完全一致。如果没有有丝分裂,就不会有什么发育过程。如图1-1所示,我们粗略地把细胞的生活周期划分为四个期: S期、G₂期、M期和G₁期。在S期时,细胞合成组蛋白,这时DNA即开始复制。学过生物化学的人都会知道,DNA的复制是多位点起始,而且是双向的;从一个复制起始点合成的DNA称为复制子。合成完毕后,合成的DNA与组蛋白组装成染色质,这时细胞进入G₂期。在G₂期细胞主要是为下一步有丝分裂作准备,合成一些有丝分裂过程中所需要的蛋白质。M期为有丝分裂时期。紧接着有丝分裂期后边有一个G₁期,在G₁期中细胞合成一些复制过程必需的蛋白和其他组分。对于某一植物品种来说,细胞的S期、G₂期和M期的长短相对来说都是较为一定、不变的,而G₁期则可以有很大的变化,主要可能与细胞当时所处的环境有关;一般来说,DNA复制一旦开始,那么细胞周期就要进行到有丝分裂完成。当然也有例外,在高度分化的组织中,细胞周期可以停在不同的阶段,而这时往往是DNA本身发生了某些变化,这样就使细胞丧失了全能性(totipotency)。

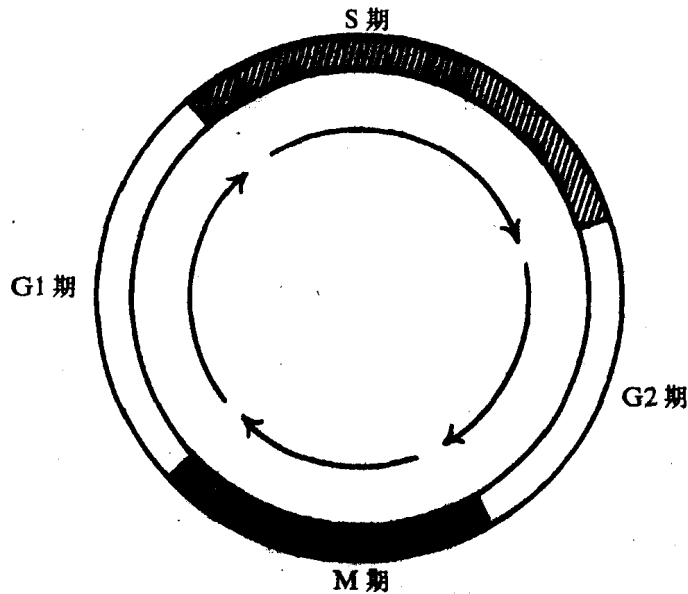


图 1-1 植物细胞周期划分示意图

植物细胞中含有大量的核 DNA, 但 DNA 含量因植物种的不同而变化极大。如表 1-1 所示, 植物中含核 DNA 最少的是拟南芥, 单倍体中只含有 0.07pg DNA; 多的如槲寄生 (*Viscum album*), 单倍体中含有 100pg DNA 以上。即便是拟南芥, 它含有的 DNA 量也是果蝇 (*Drosophila melanogaster*) 核基因组的 5 倍。显然, 植物细胞中的核 DNA 太多了, 大大超过了编码各种蛋白质所需的 DNA 量。那么那些多余的 DNA 是干什么用的呢? 人们采用了多种方法来研究这些 DNA 的功能, 包括 DNA 的变性动力学分析和单个基因的序列分析等。

DNA 变性动力学分析的原理是, 一条双链 DNA 可以经变性变成两条单链 DNA, 在合适的温度和离子强度下, 这两条单链 DNA 又可以重新复性, 粘在一起形成双链。由于这一过程是由随机碰撞决定的, 因此复性的速度是由各个不同的 DNA 序列的起始浓度决定的, 这就是说, 如果一个 DNA 序列有许多拷贝 (即在某一序列群体中有大量该序列), 那么它就会比只有一个拷贝的序列复性速度要快得多。具体的操作为, 先将植物的核基因 DNA 切成小片段 (约 200—400bp), 然后变性, 再复性。按时间取出部分混合物, 测定其中单链 DNA 的含量, 然后将这个含量与 C_0t 一起作图。 C_0t 是指复性的时间与 DNA 起始浓度的乘积, 因此人们又用 $C_0t_{1/2}$ 来表示一半 DNA 完成复性时的状况, 只需对不同样品的 $C_0t_{1/2}$ 值进行比较即可判断出它们的复杂度有什么不同。含有重复多次序列的 DNA 复性更快, $C_0t_{1/2}$ 值更低; 等量的含有单拷贝序列的 DNA 复性速度就要慢得多, $C_0t_{1/2}$ 值就高。自 70 年代中期以来, 人们一直成功地采用这种动力学分析方法分析不同基因组 DNA 的复杂度。研究表明, 植物的核基因组 DNA 中含有大量的重复序列, 因为许多植物核 DNA 的 $C_0t_{1/2}$ 值比等量的纯单拷贝 DNA 的 $C_0t_{1/2}$ 值要低得多; 例如, 在豌豆中, 只有 15% 的 DNA 的复性行为表现得与单拷贝或低拷贝序列类似, 85% 的 DNA 都是中度甚至高度重复序列, 甚至我们还能分辨出中度重复序列和高度重复序列。通常高度重复序列都存在于异染色质 (heterochromatin) 区。尽管目前已鉴定了许多重复序列 DNA 家族, 但对于它们的具体功能仍旧不甚了解。有一个例外, 那就是编码核蛋白体 RNA 的基因家族, 这些基因在植物核

基因组中高度重复。编码 18S rRNA 和 25S rRNA 的基因串接成一个单元,由单元再形成重复。整个单元通过转录先形成一大的前体 RNA 分子,然后该 RNA 前体分子再经过一系列剪切等步骤变成成熟的分子(见图 1-2)。这些基因间的间隔长短因植物种类的不同而有很大差异,有时甚至在同一种的植物间也会存在一定程度的差异。

表 1-1 不同植物种的核基因组大小

植 物 种		单倍体核基因组 DNA/pg
拟南芥	(<i>Arabidopsis thaliana</i>)	0.07
萝卜	(<i>Raphanus sativus</i>)	0.4
绿豆	(<i>Vigna radiata</i>)	0.53
番茄	(<i>Lycopersicon esculentum</i>)	0.7
鹅掌楸	(<i>Liriodendron tulipifera</i>)	0.8
大豆	(<i>Glycine max</i>)	0.9
菠菜	(<i>Spinacea oleracea</i>)	1.0
西瓜	(<i>Cucumis melo</i>)	1.0
水稻	(<i>Oryza sativa</i>)	1.0
甜瓜	(<i>Cucumis sativus</i>)	1.1
甜菜	(<i>Beta vulgaris</i>)	1.25
茄子	(<i>Solanum tuberosum</i>)	1.4
油菜	(<i>Brassica napus</i>)	1.63
南瓜	(<i>Cucurbita pepo</i>)	2.8
棉花	(<i>Gossypium hirsutum</i>)	3.0
菜豆	(<i>Phaseolus coccineus</i>)	3.5
烟草	(<i>Nicotiana tabacum</i>)	3.9
玉米	(<i>Zea mays</i>)	3.9
豌豆	(<i>Pisum sativum</i>)	4.9
大麦	(<i>Hordeum vulgare</i>)	5.5
辣椒	(<i>Capsicum frutescens</i>)	6.0
单小麦	(<i>Triticum monococcum</i>)	6.2
黑麦	(<i>Secale cereale</i>)	9.5
蚕豆	(<i>Vicia faba</i>)	13.3
燕麦	(<i>Avena sativa</i>)	13.7
洋葱	(<i>Allium cepa</i>)	16.8
小麦	(<i>Triticum aestivum</i>)	17.3
百合	(<i>Lilium longiflorum</i>)	43.2
松	(<i>Pinus sylvestris</i>)	47.9
云杉	(<i>Picea abies</i>)	50.0
槲寄生	(<i>Viscum album</i>)	107.0

植物的核蛋白体基因中没有内含子(intron)序列,但在许多其他的植物基因中却含有内含子,这一结论是通过基因操作的方法证明的。内含子插在编码序列中,同编码序列一起转录,但经RNA后加工即被切去,因而也就不参与翻译过程。另外,人们还在植物基因的5'上游区域和3'下游区域中发现一些可调控基因表达的序列,这些我们在下一节中再详细讨论。

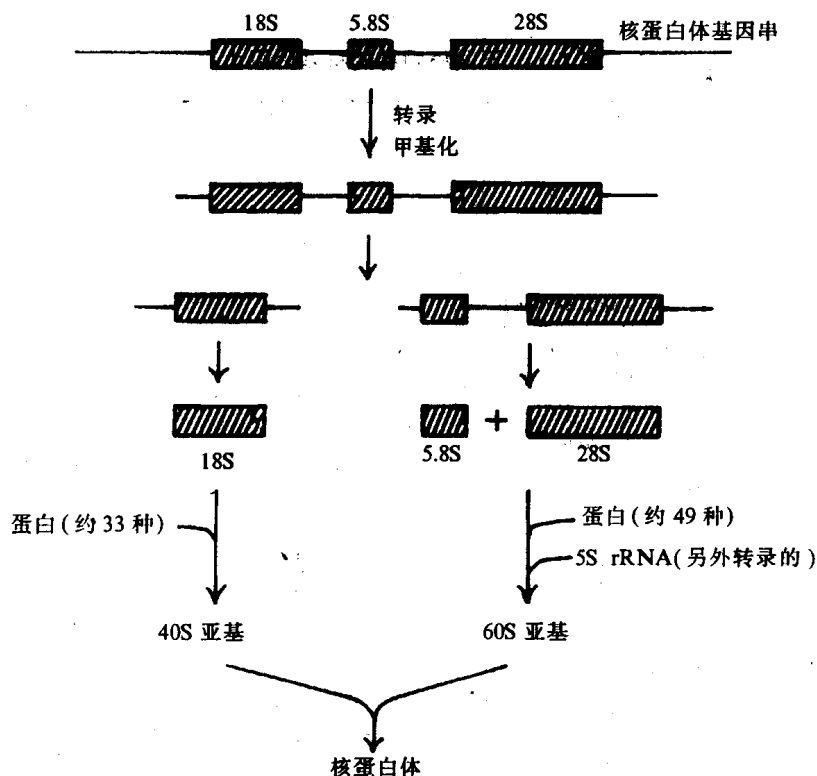


图 1-2 植物核蛋白体基因转录及核蛋白体形成示意图

第二节 核基因的表达和调控

谈到基因调控,我们一定会想到原核生物细胞中的操纵子(operon),但植物基因的调控与原核操纵子有许多不同之处。首先在结构上就有很大的差异,例如植物基因是每个编码序列前都带有它自己的调控序列,而操纵子则是几个基因共用某个调控序列。又如上一节所述,植物基因的编码序列往往被内含子截断,需要经过转录后的剪接过程才能形成功能性的成熟mRNA。同时植物基因的转录是由RNA聚合酶II来完成的(RNA聚合酶I和III分别合成rRNA和tRNA),而原核基因都是由一个单一的RNA聚合酶来完成。这些差异同时也告诉我们,植物基因的调控过程与原核基因调控不同;是一个极其复杂的过程。

一、编码蛋白的植物基因的结构

将植物编码蛋白的基因序列与其他真核基因序列进行比较,我们通常都能发现,在基因的5'端、3'端甚至编码区内存在着一些保守的结构序列和调控序列,这些保守的序列对于基因准确、协同的表达具有决定性作用。在转录起始位点上游5'端的序列我们称之为启动子(promoter)区域,这个启动子区域主要负责转录的起始和调控。植物基因的启动子包括三

个较重要的区域：一是转录起始位点（常称为+1处），二是在转录起始点上游25—40bp处，三是在-75处或更远些。在转录起始位点上游约25bp（即-25）处，我们通常能找到一个所谓的TATA盒子（TATA box，亦称Hogness box），植物基因的TATA盒子中心序列为TCACTATATATAG。TATA盒子的功能是将RNA聚合酶II引导至正确的转录起始位点。实验证明，转录的准确起始必须要有完整的TATA盒子存在；在TATA盒子两边都是富含GC的序列（GC-rich），这些序列可能也对TATA盒子的功能起一定的作用。熟悉细菌的人可能会发现，植物基因的TATA盒子与细菌的Pribnow盒子十分近似，只是在位置上有些不同而已。

在转录起始位点上游约-75处有一个或几个调控基因转录活性的元件，不同基因间的这些序列通常都有一定的同源性，其中较普遍存在的是两个盒子：GC盒子（GC box）和CAAT盒子（CAAT box），前者的中心序列为GGGCGG，后者的中心序列为GG^T_CCAATCT。这些元件可以以任一方向同时存在于某一基因中，拷贝数也不定。有趣的是，CAAT盒子还有极少数的例外，即在极少数植物基因上没有CAAT盒子，例如Rubisco小亚基基因等。尽管如此，通过对许多植物基因的5'端区域进行缺失分析，人们已经证明这些元件对于基因的表达具有重要的作用。在有些已经分析过的植物基因中没有CAAT盒子，而是由另一个类似的AGGA盒子代替CAAT盒子行使调控功能。除了以上提到的几种保守的盒子（属结构序列）以外，在5'上游区域还有一些序列是负责调控植物基因发育阶段性表达和组织特异性表达的，属于调控序列，在后边的章节中我们将作简略介绍。

有些玉米醇溶蛋白基因的结构很复杂，它们带有两个启动子（P₁和P₂），每一启动子都带有不止一个的转录起始位点。在P₂启动子上有两个转录起始位点，其中一个离编码区只有50bp远，这个转录起始位点并不起始转录，却是内含子的剪切位点，因此人们推测可能有一个内含子插入到非翻译区中了。对于内含子（intron），我们已有了一些了解，它可能是细胞内原本分开的基因区域间发生重组的产物，通常内含子都是插在基因的编码序列中间的。与内含子相对，人们还常用到外显子（exon）一词，外显子指的是在前体hnRNA分子和mRNA分子中共有的片段，外显子通常是被内含子分开的，内含子与外显子之间有明显的界限（见图1-3）。

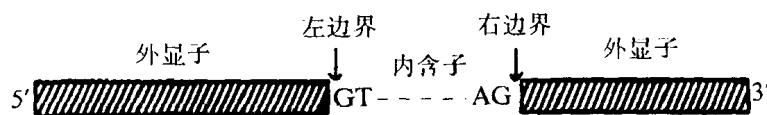


图1-3 内含子-外显子边界示意图

植物基因转录首先是形成核内不均一RNA(hnRNA)，它是mRNA的前体，通常是mRNA分子的4—5倍长。hnRNA形成后很快即在5'端加上“帽子”，3'端加上poly(A)尾巴（详见“植物RNA技术”一章）。带“帽子”和尾巴的hnRNA要变成mRNA，就必须将插在编码序列中内含子剪切掉。切去内含子需要ATP参与，主要分两步进行：最先是内含子左边界断开；然后其断开的游离5'端与距右边约18—40个核苷酸的TACTAAC盒子中的第六位A形成一个5'—2'磷酸二酯键，从而变成一个套（lariat）的形状。这时内含子的右边界才被切开，套再变成线形，最终完成切去内含子的工作。

绝大多数的植物基因都含有内含子，但也有些例外，如玉米的贮藏蛋白基因和小麦的叶

绿素 a/b 结合蛋白基因就没有内含子，它们的转录后加工就省去了剪切内含子这一道“工序”。

基因的编码区域是以翻译起始密码 ATG 开始的，到终止密码子 (TAG、TAA 或 TGA) 结束，中间都称为开放读码框架。通常我们能在一个保守的序列 (AACAATGGCT) 中找到起始密码子 ATG。当 40S 核蛋白体亚基从 mRNA 5' 端开始寻找其结合位点时，这个保守序列就能被它识别。

在基因的 3' 端也有一个保守的序列，即 AATAAA，绝大多数人都认为它是一个加 A 序列，即有它存在，基因的 mRNA 3' 端的多聚腺苷化过程才能得以完成。

图 1-4 所示为高等植物编码蛋白的基因结构图，从图中可以清楚地看到各种基因元件的不同位置。

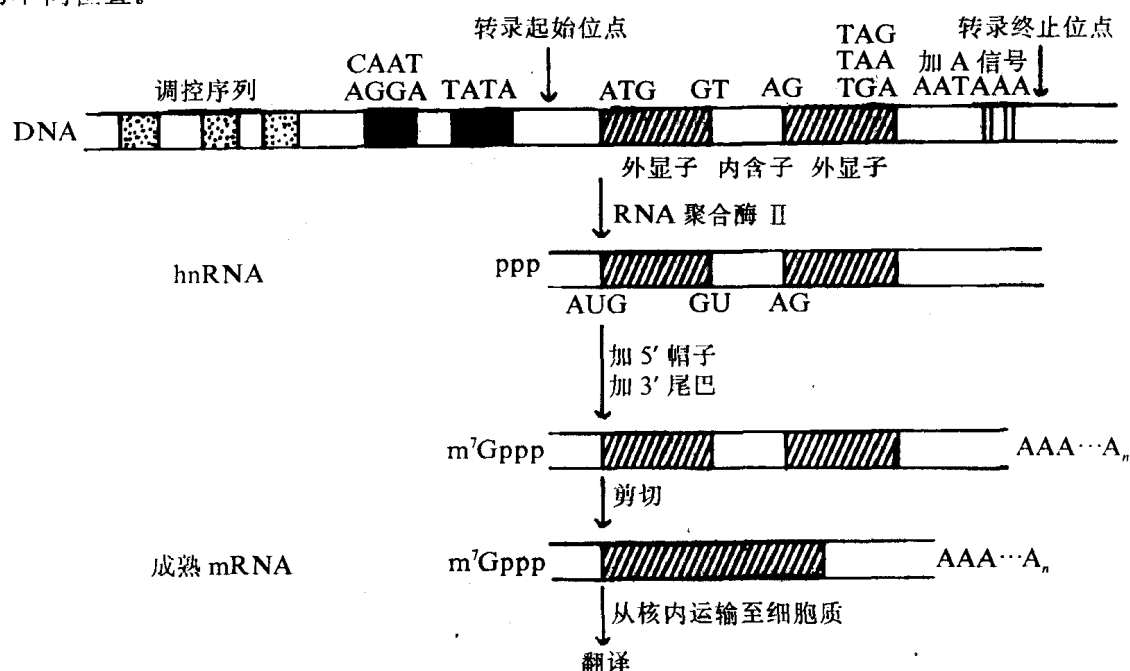


图 1-4 植物编码蛋白基因的结构与表达

二、植物基因的调控

植物基因表达的调控是由调控基因转录、mRNA 的加工形成、mRNA 从核中运出、mRNA 在核质胞质中降解以及翻译等众多过程的速率入手的。一般来说，植物基因表达的调控都是转录调控，即调控基因转录的速率，这种调控可以通过基因转录出的 mRNA 量的多少来判断，mRNA 增加了，即可说明该基因的转录被调控加快了，反之则说明转录速度放慢了。除了转录调控外，各种转录后的加工过程的调控也很重要，因为这些过程直接关系到 mRNA 的稳定性。在植物中，基因的转录调控过程是很复杂的，它包括基因的某些区域与结构蛋白或是其他某些特异性蛋白的相互作用。正如我们在图 1-4 中所看到的，植物基因的启动子区域有一些调控序列 (元件) 或盒子，这些元件和盒子上都能结合上某种 (些) 蛋白，这些 DNA 与蛋白质之间的相互作用将直接影响基因在何时开始转录，转录速率多大。研究植物基因表达的调控目前主要包括研究转录调控、DNA 甲基化及转录后调控等，其中转录调控的研究就包括鉴定植物基因上的保守的调控序列、将基因修饰改造后在转基因植物中表