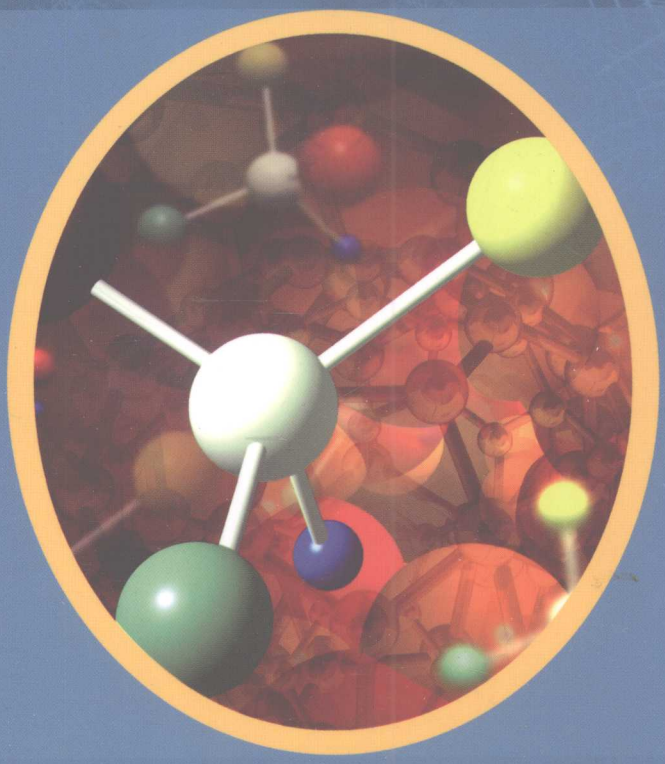
# 基因的分子生物学

# *Molecular Biology of the Gene*

杨业华　主编

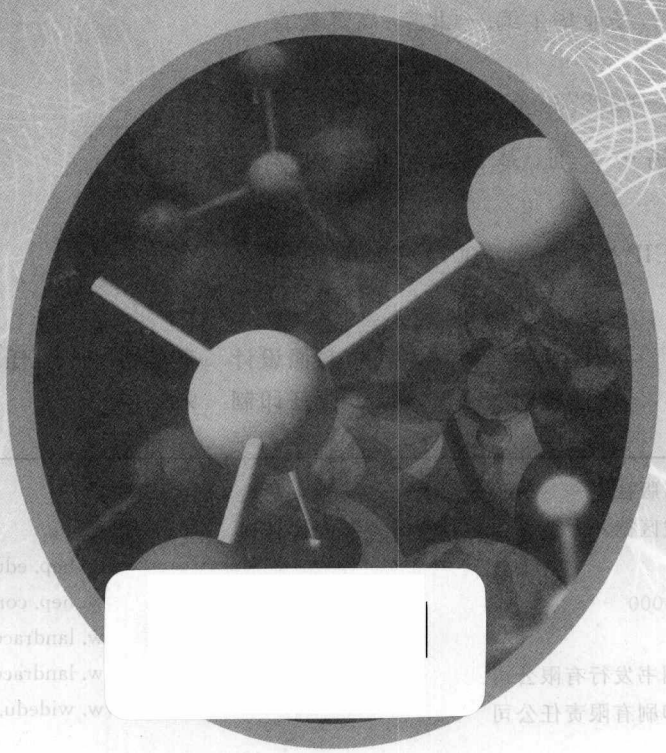


高等教育出版社
Higher Education Press

# 基因的分子生物学

# *Molecular Biology of the Gene*

杨业华　主编



高等教育出版社
Higher Education Press

## 内容简介

本书系统阐述基因的结构、功能、调节表达机理以及基因突变、重组的分子基础和基因操作的基本理论与方法。全书分7章，分别论述基因的结构特征与组织、原核生物和真核生物基因的调节表达、基因重组的分子基础及转座因子、基因操作与应用以及基因对发育、细胞分化、细胞增殖的控制等。为便于读者学习，每章之后都附有习题及参考答案，书末有中、英文索引和主要参考文献。

本书可作为高等农林院校生物类各专业本科生的基础课程教材，也可作为综合性大学和高等师范院校生物类有关专业本科生的专业基础课程教材或教学参考书，并可供有关专业的研究生和科技工作者参考。

**图书在版编目(CIP)数据**

基因的分子生物学/杨业华主编. —北京：高等教育出版社，2008.1

ISBN 978-7-04-022878-6

Ⅰ.基… Ⅱ.杨… Ⅲ.基因-分子生物学 Ⅳ.Q343.1

中国版本图书馆CIP数据核字(2007)第190429号

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **策划编辑** | 潘　超 | **责任编辑** | 田　军 | **封面设计** | 张　志 | **责任绘图** | 尹　莉 |
| **版式设计** | 王　莹 | **责任校对** | 金　辉 | **责任印制** | 宋克学 | | |

| | | | |
|---|---|---|---|
| **出版发行** | 高等教育出版社 | **购书热线** | 010-58581118 |
| **社　　址** | 北京市西城区德外大街4号 | **免费咨询** | 800-810-0598 |
| **邮政编码** | 100011 | **网　　址** | http://www.hep.edu.cn |
| **总　　机** | 010-58581000 | | http://www.hep.com.cn |
| | | **网上订购** | http://www.landraco.com |
| **经　　销** | 蓝色畅想图书发行有限公司 | | http://www.landraco.com.cn |
| **印　　刷** | 北京凌奇印刷有限责任公司 | **畅想教育** | http://www.widedu.com |
| **开　　本** | 787×1092　1/16 | **版　　次** | 2008年1月第1版 |
| **印　　张** | 19 | **印　　次** | 2008年1月第1次印刷 |
| **字　　数** | 460 000 | **定　　价** | 23.90元 |



**物料号　22878-00**

# 编 写 人 员


**主　　编**：杨业华（华中农业大学）

**副 主 编**：马正强（南京农业大学）

张天真（南京农业大学）

赵兴波（中国农业大学）

**参编人员**：王洪刚（山东农业大学）

余四斌（华中农业大学）

唐灿明（南京农业大学）

储存艮（南京农业大学）

刘　榜（华中农业大学）

赵书红（华中农业大学）

余　梅（华中农业大学）

# 前言

《基因的分子生物学》是以其作为《普通遗传学》第二版的配套教材为目的编写的。全书分七章，第一章基因的结构及组织特征是对基因特征的概述，也可看做是全书内容的导论，以后各章则是基因分子生物学的各论，包括原核生物基因表达及其调控、真核生物基因表达及其调控、基因重组的分子基础和转座因子、基因表达与发育、基因与癌及对细胞增殖的调控、基因操作及其应用等内容。

在编写过程中，我们尽可能引入了在基因研究方面的最新成果。名词、术语尽量采用标准名词和《英汉生物学词汇》(科学出版社，1997)中收录的词条，但对少数标准词汇中过时或错误的名词作了修正，如“operon”和“operator”在标准词汇中分别被译为“操纵子”和“操纵基因”，本书则采用了孙乃恩、孙东旭、朱德熙编著的《分子遗传学》中的译名“操纵元”和“操作子”。有些名词如“dimorphic gene”、“cryptomorphic gene”、“chaperonin”等在上述词典中均无相应译名，但在网络如“google”上均可查到英文名，它们在本书中则分别被译为“二型蛋白基因”、“隐形蛋白基因”和“伴侣蛋白宁”等。这些译法如果有误，待今后有了统一的名称后，再予以更正。

本书适合对遗传学基础知识要求较高的专业如生命科学、植物科学、动物科学、生物技术等专业的本科生在修完《普通遗传学》内容的基础上进一步学习，课堂讲授按60学时分配为宜。

本书与《普通遗传学》组成一套系列教材。参加此次遗传学系列教材修订工作的人员有山东农业大学王洪刚、南京农业大学马正强、中国农业大学赵兴波、华中农业大学杨业华、华中农业大学余四斌，张天真、唐灿明、刘榜、储存艮、赵书红、余梅仍是这套系列教材的编写成员。

虽然这套遗传学教材在结构和内容上都作了较大的调整和充实，但由于编者的水平有限，不足之处在所难免，还望有关专家和广大读者提出宝贵意见。

编　者

2007年8月28日

# 目　录

# 1. 基因的结构及组织特征

经典遗传学指出，每一种生物都具有许多控制性状发育的遗传单位，这些遗传单位称为基因(gene)。现在我们知道，基因是DNA分子中的一段核苷酸序列，个体的基因组成总括起来称为基因型(genotype)，一个细胞中所有的基因称为基因组(genome)。基因有一定的结构特征，基因能够发生重组、突变，决定某种性状发育或某种代谢途径的多个基因在基因组中有其一定的组织形式，有些基因还能从基因组的一个位置移动到另一个位置上。

## 1.1 基因的一般结构特征

### 1.1.1 转录单位

基因行使其基本功能的部分是编码某种特殊大分子化合物的核苷酸序列，这段序列称为转录单位(transcriptional unit)。RNA聚合酶可以将其转录成与之互补的RNA链，包括rRNA、tRNA和mRNA等。如果转录子，即以DNA为模板合成的各种RNA的统称，是mRNA，通过翻译，就能编码出蛋白质或者组成蛋白质的多肽亚基。DNA分子中这种决定蛋白质分子中氨基酸序列的区域称为可读框(open reading-frame，ORF)，它含有一系列编码氨基酸的三联体，不含任何终止密码子。由于基因的最初产物是能够与其编码区互补的RNA序列，所以习惯上就以双链DNA分子中与模板链互补的那条链即编码链来表示基因的组成。

### 1.1.2 基因间的间隔序列

在两个基因的编码区之间存在一些不编码的核苷酸序列，这类序列统称为基因间间隔序列(intergenic spacer)，其长度变化很大，从一对碱基到数千对碱基不等。间隔序列中与某个基因相连的一部分序列通常称为侧翼序列(flanking sequence)。由于间隔序列中某些部分对基因的功能具有某种特殊作用，所以有时基因5′端的这种序列称为前导序列(leader sequence)或5′非转录控制区，基因3′端的非编码序列称为拖尾序列(trail)或3′非转录控制区，亦可称为尾随序列(tailer sequence)。但是，在许多生物的基因组中，有些基因之间并不存在间隔序列，有些则是两个或更多的基因首尾连接在一起，甚至有些基因的首尾部分还相互重叠。

在所有基因中，双链DNA分子中只有一条链含有编码序列。但是在许多生物的基因组中，有些基因位于双链DNA分子的一条链上，而另一些基因则位于另一条互补链上。由于两条互补链具有相反的极性，而转录机器对基因的阅读又始终是按3′到5′方向进行，所以在两条互补链上的基因就具有不同的极性。

### 1.1.3　基因的调控序列

基因在表达过程中，受到一系列具有特殊功能的、短的 DNA 序列的调节。其中有些控制基因的转录起始和终止过程，有些确保翻译过程中核糖体与 mRNA 的结合，而另一些则与基因接受某些特殊信号有关，所以分子遗传学中就把这些控制和调节基因表达的特殊序列统称为调控序列(controlling sequence)(图 1-1)。在这些序列中，有些位于编码序列上游并同基因的编码序列位于同一条 DNA 分子上，所以它们又称为顺式作用元件(*cis*-acting element)。

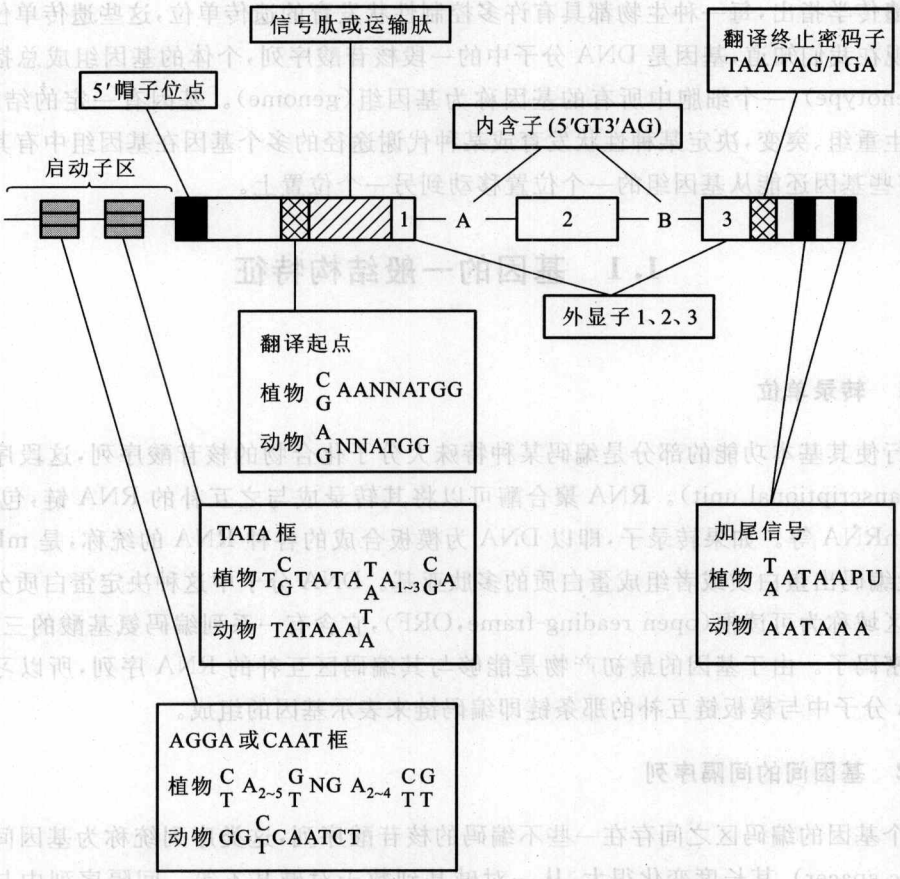



图 1-1　真核生物蛋白质基因的一般结构模型

#### 1.1.3.1　启动子

在基因的 5′端即 5′非转录调控区中有一段能被 RNA 聚合酶识别，并可与 RNA 聚合酶结合的序列，这段序列称为启动子(promotor)，它控制基因的转录起始过程。在大肠杆菌中，启动子含有两段结构保守的序列。一段位于转录起点上游第 10 个碱基以上，其共有序列(consesus sequence)为 TATAAAT，称为普里布诺框(Pribnow box)。另一段位于这一段序列上游约 25 个碱基的位置，其共有序列为 TTGACA，这段序列有时也叫 RNA 聚合酶识别位点。在真核生物基因中，启动子通常是指准确而有效地起始基因转录所需的任何 DNA 序列。对编码蛋白质的

基因来说，启动子还包括位于转录起点(+1 位)周围的帽子位点(cap site)，−30 位的 TATA 框(TATA box，又称 Goldberg-Hogness box)，−40 至−110 位之间的其他上游因子(upstream element 又称上游启动元件)等。

**1.1.3.2 增强子(enhancer)**

增强子首先见于病毒基因中，后来在许多真核生物基因中都有发现。在某些病毒基因的启动子附近，有一种特殊的结构保守序列 GGTGTG－G $\begin{matrix}AAA\\TTT\end{matrix}$ G(－横线表示任意碱基)，它可以显著地提高启动子起始基因转录的效率。增强子有以下几个特征：① 其增效作用具有双方向性，即它既可以作用于上游基因，也可以作用于下游基因，其作用范围可以远达与其相距 3 kb 的基因；② 无论增强子与某个基因的极性相同或相反，对该基因都具有活性；③ 增强子具有明确的组织和物种特异性。在猴肾病毒 SV40 中，增强子是一个 72 bp 的重复序列的组成部分，但几个增强子中只有一个具有增效功能。

**1.1.3.3 核糖体结合位点**

在基因编码区的上游，除了存在上述非转录区中的调控因子以外，在基因的转录区中，即在 mRNA 的翻译起始位点周围有一组特殊的序列，也可以控制 mRNA 的翻译过程，其中主要为 ATG(mRNA 中为 AUG)起始密码子及其前后的若干碱基。在细菌基因中，这段短的核苷酸序列称为 SD 序列(shine-dalgarno sequence，SD)，共有序列为 AGGAGG，它是核糖体的结合位点，与 16S rRNA3′端 CCUCCU 结合。真核生物基因中没有 SD 序列，核糖体是通过其他机制与 mRNA 结合的(见 3.5)。

密码子 ATG 的位置和可读框的方向并不是决定翻译起始过程的唯一因素。例如将 AUG 插入到 mRNA 的前导序列中，这一密码子对翻译起始的影响依赖于其侧翼序列。如果 AUG 前有 A(或 G)CC，其后有 G，这种序列组合对起始转录过程具有最高的效率。在某些基因中，翻译起始密码子为 GUG 和 UUG，虽然通常前者编码缬氨酸，后者编码亮氨酸，但当作为起始密码子时，二者都编码甲硫氨酸。在大肠杆菌的氨甲酰磷酸合成酶基因中，翻译起始密码子则是 AUU(通常编码异亮氨酸)。

**1.1.3.4 基因拖尾区中的信号序列**

基因 3′端非编码区中的信号主要是与终止转录过程有关的序列，称为终止子(terminator)。根据对许多转录子的核苷酸序列分析，RNA 聚合酶终止转录的位点正好位于转录子本身的末端。终止子主要有两种类型，第一种终止子由一对富含碱基 GC 的反向重复(inverted repeat)序列以及寡聚 T 组成，这段序列可以形成发夹环结构(图 1－2)。可能当转录到达这一区域时，RNA 通过自身碱基配对，形成发夹环结构而终止转录过程。第二类终止子的结构目前还不是很清楚，只知道它们不富含 GC 碱基和寡聚 T 序列。现在

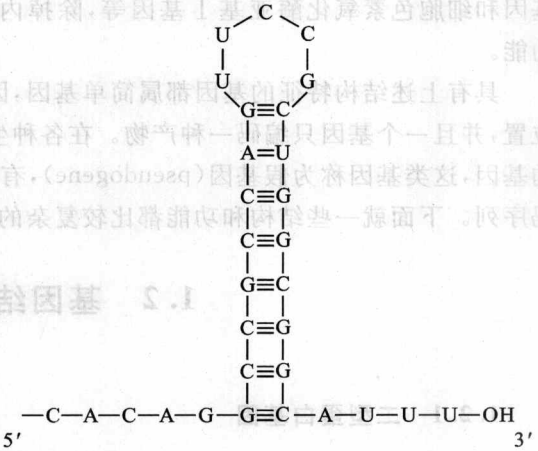



图 1－2 转录终止位点的结构

知道，细菌中这类终止子行使功能时需要 ρ 因子(Rho factor)和一种称为 nusA 的蛋白质。需指出的是，并非所有基因的终止子都相同。例如在酵母中，mRNA 有一段由 8 个碱基 TTTTTATA 组成的序列，当这段序列活化后，转录过程就终止在另一组序列 CAATATTTG 的最后一个 T 碱基上。

#### 1.1.3.5　加尾信号序列

真核生物 mRNA 的 3′端都有一段多聚 A 尾巴[poly(A) tail]，这种尾巴不由基因编码，而是转录之后加到 mRNA 上的。加尾过程受位于终止密码子 3′端的加尾信号序列控制。现在知道，动、植物基因都有一个以上的加尾信号序列，动物基因的加尾信号共有序列为 AATAAA，植物基因的加尾信号序列变化较大，为 $\begin{matrix}\mathrm{T}\\\mathrm{A}\end{matrix}$ATAApu。原核生物的 mRNA 没有多聚 A 尾巴，故基因中没有加尾信号序列。

#### 1.1.3.6　内部调控信号

某些较小的 RNA 分子如 tRNA、5S rRNA 以及病毒的某些 RNA，在它们的基因中，启动子序列位于基因的编码序列之内。关于这种位置与启动转录之间的关系目前还不是很清楚。

### 1.1.4　割裂基因

在真核生物中，大多数基因编码蛋白质的序列都被若干不编码的序列隔开，编码区中这种不编码的部分称为内含子(intron)，编码的部分称为外显子(exon)。基因在转录以后通过加工，将内含子去掉，然后再将外显子连接起来，形成一个连续的编码区，像这样的基因称为割裂基因(split gene)。

割裂基因中内含子的数目因物种和基因不同而有所不同，有些基因只有一个内含子，有些则有多个内含子。例如小鼠肌肉 α-肌动蛋白基因(muscleo-actin gene)有 6 个内含子，酵母线粒体的细胞色素氧化酶亚基Ⅰ基因则有 7～9 个内含子。

在转录过程中，割裂基因的内含子和外显子一起被转录。对某些基因来说，除掉内含子并不影响基因的正常功能和基因产物的加工；而对另一些基因来说，如 SV40 的基因、酵母 $tRNA^{Tyr}$ 基因和细胞色素氧化酶亚基Ⅰ基因等，除掉内含子往往导致转录子不能进行加工和失去正常功能。

具有上述结构特征的基因都属简单基因，因为这些基因在基因组中都有相对固定的结构和位置，并且一个基因只编码一种产物。在各种生物的基因组中还存在一些结构和功能都不完整的基因，这类基因称为假基因(pseudogene)，有的缺少转录起始和终止信号，有的缺少完整的编码序列。下面就一些结构和功能都比较复杂的基因分别加以说明。

## 1.2　基因结构的变异类型

### 1.2.1　二型蛋白基因

二型蛋白基因(dimorphic gene)是指含有两个相互连接的编码区的基因，它编码两种不同功能的多肽，其特征是 5′编码区的产物为信号肽(signal peptide)或运输肽(transit peptide，亦称转

运肽),它们负责其下游功能部分的跨膜运输。例如小鼠的激肽释放酶(kallikrein)基因的前体mRNA长867个碱基,其中前导序列为24个碱基,3′端拖尾区为48个碱基,编码序列为795个碱基,编码长265个氨基酸的蛋白质(图1-3)。翻译的初级产物称为前酶原(preproenzyme),其前面长17个氨基酸的亚基行使一种运输蛋白的功能,称为信号肽。在内质网膜上合成蛋白质的过程中,信号肽与一种称为运输识别颗粒(transit recognition particle)的特殊受体互作,识别内质网的表面位点并且将与之相连的激肽释放酶原运输过内质网膜。然后,在运输识别颗粒的帮助下,整个前酶原进入内质网泡腔中,最后通过酶切,将信号肽除掉,形成激肽释放酶原。激肽释放酶原分泌到细胞外之后,还需要经过进一步加工,将前面的11个氨基酸除掉,才形成具有活性功能的激肽释放酶。因此,激肽释放酶基因编码两个功能部分,即信号肽和激肽释放酶,二者被一段不具活性的区域隔开。

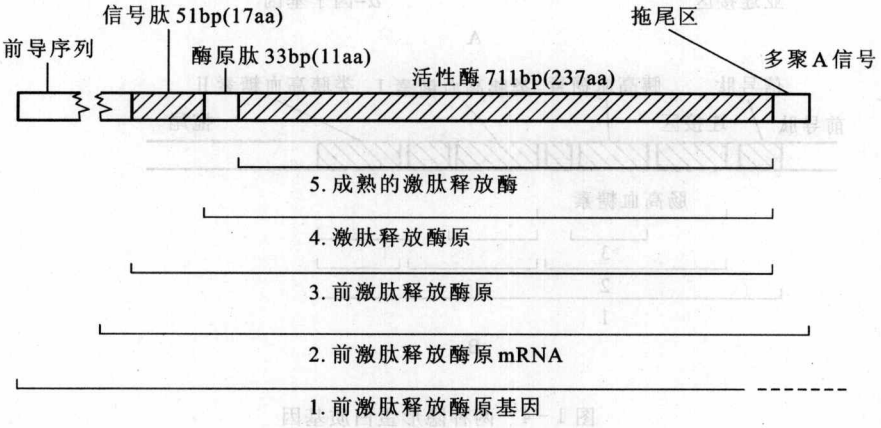



图1-3 二型蛋白基因的结构

图中"aa"表示氨基酸。小鼠的激肽释放酶基因具有一段编码信号肽的前导序列,其后为一段编码酶原肽的序列,当这两段序列被除掉以后,才能形成具有活性的激肽释放酶

### 1.2.2 隐形蛋白基因

在真核细胞中,有些基因的具活性的最终编码产物隐蔽在前体蛋白质之中,只有当前体蛋白质通过酶切或加工之后,才能形成具有功能的蛋白质,编码这类蛋白质的基因称为隐形蛋白基因(cryptomorphic gene)。例如啤酒酵母(*Saccharomyces cerevisiae*)的交配外激素(mating pheromone,亦可称为性外激素或交配信息素)基因,也叫α-因子基因(*MFα-1*)就属于这类基因。交配外激素基因的转录子长495个核苷酸,翻译产物长165个氨基酸。这段肽链包括4段首尾连接的α-因子,每段长13个氨基酸。这4个α-因子均被一段短的前导肽隔开,除第一个前导肽长6个氨基酸以外,其余均长8个氨基酸(图1-4A)。前体蛋白质在翻译后以完整的形式贮存于细胞之中,当酵母细胞受到某种未知的内、外因素影响时,前体蛋白质通过水解,释放出α-因子和前导肽。总的来说,隐形蛋白基因表达时,首先编码出一条大蛋白质分子,然后其近羧基的半部分经过切割和加工,最终形成具有活性的交配外激素。前体交配外激素的近氨基的半部分相当于二型蛋白基因的前半部分,其作用是使其实际功能部分停留在不活动状态,当后半部分被

利用时，近氨基的半部分也随之被切割。至于其中的连接区是否具有运输蛋白质的功能，目前尚不清楚。

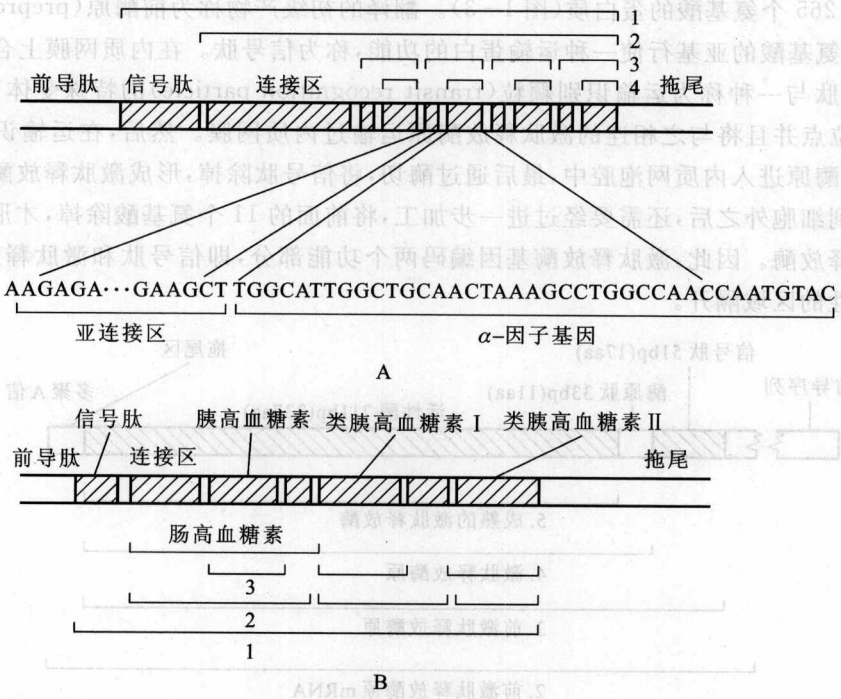



图 1-4　两种隐形蛋白质基因

图中数字表示加工或合成过程中各步骤的产物，阴影区示初级转录子，但其中只有一部分编码最终的功能产物。A. 啤酒酵母交配外激素基因　B. 哺乳动物胰高血糖素基因

高等真核细胞的隐形蛋白质基因比酵母交配外激素基因复杂得多，例如哺乳动物胰的胰高血糖素基因。胰高血糖素基因编码的初级翻译产物称为前高血糖素原，其前面的 20 个氨基酸为信号肽。前体蛋白质的主要部分为高血糖素原，长 160 个氨基酸(图 1-4B)，包括高血糖素的结构序列和两条不同的称为类胰高血糖素的多肽。前体蛋白质在加工过程中，首先去掉类胰高血糖素多肽，形成长 69 个氨基酸的、称为肠高血糖素的激素，然后再通过第二次加工才形成具有功能的胰高血糖素。

### 1.2.3　隐密基因

在原生动物的线粒体以及某些真核生物的线粒体基因组中，有些基因以一种节略的形式存在，它们缺少完整编码蛋白质的序列和 mRNA 分子中与翻译有关的其他一些功能结构。这类基因在转录后通过一种特殊的加工方式即 RNA 编辑(RNA editing)，对 RNA 进行加工，将转录子变成可翻译的 mRNA，这类基因称为隐密基因(cryptogene)。根据对锥虫(*Trypanosomatide brucei*)以及其他几种原生动物线粒体基因组的分析，细胞色素 b 基因(*cyb*)、细胞色素氧化酶亚基Ⅱ(coⅡ)和亚基Ⅲ(coⅢ)基因都是隐密基因。例如，在锥虫中，*cyb* 基因在转录后，mRNA 5′端加进了 34 个 U，而在另外两种锥虫 *Leishmania tarentolae* 和 *Crithidia fasiculate* 的线粒体

中，*cyb* 基因的 mRNA5′端则加进了 39 个 U。又如锥虫的 *co*Ⅲ基因，其初级转录子在 158 个位点上共加进了 398 个 U，在 9 个位点上缺失了 19 个 U。隐密基因在转录后，通过 RNA 编辑，或者是将一种不具功能的转录子加工成可翻译的 mRNA，或者是通过修改已具功能的 mRNA 分子而改变其编码产物的氨基酸序列。

### 1.2.4 分段基因

高等真核生物的某些基因可以分割成几个独立转录的部分，各部分转录以后，不同转录子再通过剪接连接在一起，形成具有翻译功能的 mRNA（见 3.4.5）。像这样的一类基因称为分段基因（divided gene），其转录子的这种剪接方式称为反式剪接（trans-splicing）。烟草的编码一种叶绿体核糖体蛋白 rpS12 的基因、莱茵衣藻的光合系统Ⅰ的 *psaA* 基因、线虫（*Caenorhabditis elegans*）的肌动蛋白（actin）基因等都是分段基因，虽然它们在基因组中为数不多，但在细胞或生物体的整个生命活动过程中扮演着十分重要的角色。

### 1.2.5 复合基因

复合基因（complex gene）主要为脊椎动物免疫系统的基因，其结构相当复杂，现以免疫球蛋白（immunoglobulin）基因为例加以说明。

免疫球蛋白又称抗体（antibody），现在已知有 5 种类型的免疫球蛋白，即 A、D、E、G 和 M，其中有些还可以分成几个亚种。一般说来，每种免疫球蛋白都由 4 个亚单位组成（图 1-5），即两条重链（H 链）和两条轻链（L 链）。轻链又可分成 κ（kappa）链和 λ（lambda）链，所以免疫球蛋白的一般结构可用 κ2H2 或 λ2H2 表示。

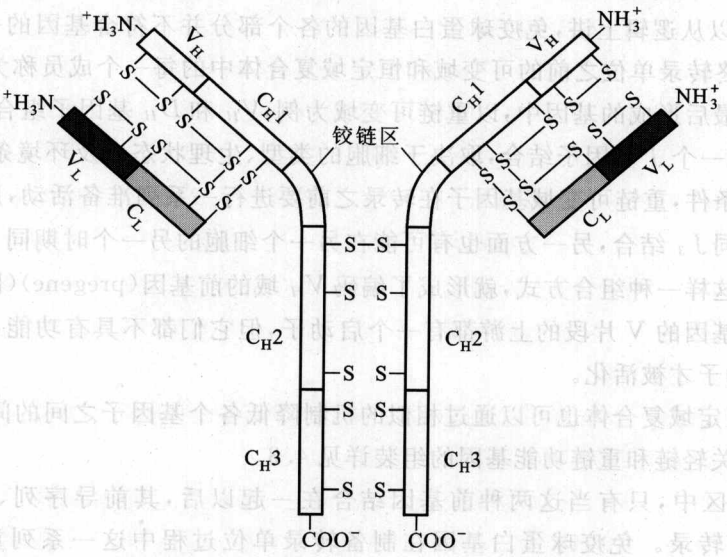



图 1-5 免疫球蛋白的结构模型

两条轻链（阴影区）和两条重链（无阴影区）通过二硫键结合在一起，每一条重链和轻链的氨基末端部分都有一个可变域（$V_H$ 和 $V_L$），轻链只有一个恒定域（$C_L$），而重链则有三个恒定域（$C_H1$、$C_H2$、$C_H3$）和一个铰链区

免疫球蛋白的两条轻链分别由 212～214 个氨基酸组成，每一条链又大致均等地分成一对相对应的区域。近羧基末端的半部分通常变异较少，所以称为恒定域(constant domain)，近氨基末端的半部分称为可变域(variable domain)，其变化范围很大，主要与抗体的特异性有关。

上述 5 种免疫球蛋白重链的变化都比轻链的变化更大。从结构上讲，重链与轻链相似，都由恒定域和可变域组成，但是重链有 3～4 个恒定域，而轻链 $\kappa$ 和 $\lambda$ 变异体都只有 1 个恒定域。在重链的恒定域中，前面的两个由一段短的铰链区(hinge)隔开。除了 L 链和 H 链以外，有些成熟的免疫球蛋白分子(如 IgA 和 IgM)还具有一条 J 链，其功能是将某些亚基连接在一起。在 IgA 中还有一个 S 链，这条链只在腺体中形成。

上述 5 种免疫球蛋白的主要差别在于恒定域不同，因此每一类型的恒定域的编码序列各不相同。恒定域的编码序列常用希腊字母表示，如 IgM 的恒定域写为 $C_\mu$，IgA 的恒定域写为 $C_\alpha$，IgD 为 $C_\delta$ 等加以区别。又由于同一类型的免疫球蛋白有多个恒定域，所以各个恒定域的编码序列又可进一步用数字加以区别，如 IgM 恒定域的各个编码区分别为 $C_{\mu 1}$、$C_{\mu 2}$、$C_{\mu 3}$ 和 $C_{\mu 4}$。此外，这 5 种免疫球蛋白的每一条链在基因组中都有多个拷贝的编码序列，所以其基因结构组成就非常复杂。

编码重链可变域的基因的结构也分为几个不同部分。第一个区域为 $V_H$，编码重链可变部分的大部分肽链，其后为一条短的、称为 $D_H$ 的片段和几条称为 $J$ 的片段，$J$ 片段最后能将可变域和恒定域连接在一起。免疫球蛋白基因的前导区也是一种复合结构，称为前导复合体。原始的前导区很长，其中只有一部分最后拼接在一起用作转录的控制信号。

由于恒定域和可变域复合体的各个成员既不编码某个具体蛋白质的亚基，也不编码某种功能 RNA 分子，所以从逻辑上讲，免疫球蛋白基因的各个部分并不符合基因的一般定义，但目前通常把组装成最终转录单位之前的可变域和恒定域复合体中的每一个成员称为基因节段或基因子(genelet)。在最后形成的基因中，以重链可变域为例，$V_H$ 和 $D_H$ 基因子组合只与一个 $J$ 基因子结合，究竟同哪一个 $J$ 基因子结合，取决于细胞的类型、生理状态以及环境条件等一系列复杂因素。根据环境条件，重链可变域基因子在转录之前要进行一系列准备活动，所以 $V_H$ 和 $D_H$ 部分一方面有可能同 $J_3$ 结合，另一方面也有可能在另一个细胞的另一个时期同 $J_1$ 或 $J$ 组的其他成员结合。通过这样一种组合方式，就形成了编码 $V_H$ 域的前基因(pregene)(图 1－6)。

重链和轻链基因的 V 片段的上游都有一个启动子，但它们都不具有功能，只有在 V 区和 C 区结合后，该启动子才被活化。

同样，重链恒定域复合体也可以通过相似的机制降低各个基因子之间的间隔片段的长度而形成前基因。有关轻链和重链功能基因的组装详见 4.3。

在重链编码区中，只有当这两种前基因结合在一起以后，其前导序列、肽链编码区和间隔序列才能用于转录。免疫球蛋白基因在制备转录单位过程中这一系列重要的、具有高度选择性的步骤可能是在各种内、外因素的影响下，基因组中其他各种有关基因共同作用的结果。在上述 5 种免疫球蛋白中，细胞究竟合成哪一种，完全由细胞内的活动决定。例如 IgM 首先出现在许多幼期细胞中，在成熟细胞中 IgM 停止合成，而合成第二类免疫球蛋白(通常为 IgG)。

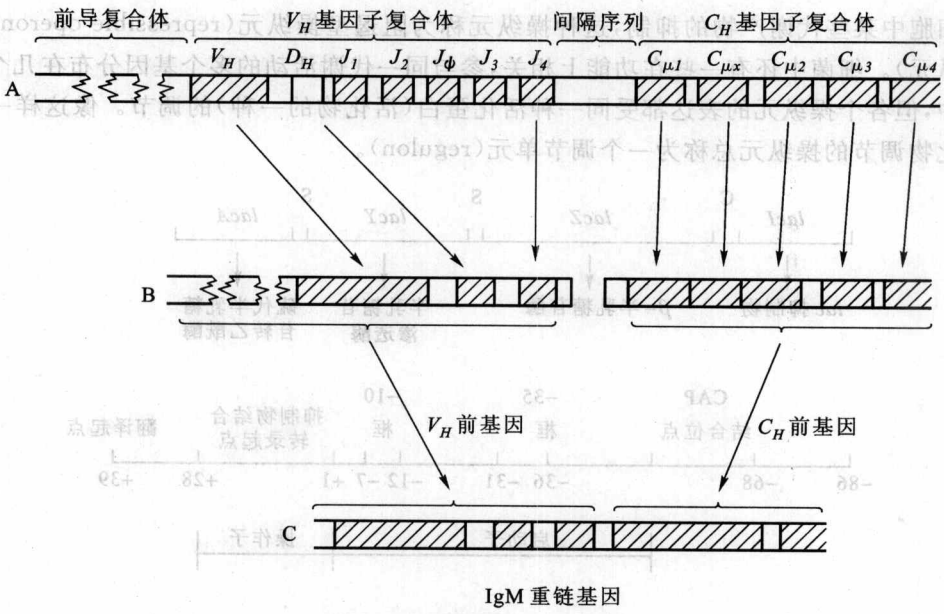



图 1-6 免疫球蛋白重链基因结构及其组合过程
A. 基因子复合体 B. 降解成前基因 C. 组装成 IgM 重链基因

## 1.3 基因的组织

### 1.3.1 操纵元

操纵元(operon)一词首先来自对细菌突变体的分析。在细菌中,编码若干种功能上相关的酶类的基因,在转录时受一个开关单位控制,这些基因作为一个单位转录,形成一条多顺反子mRNA(polycistronic mRNA),各个顺反子之间被一段短的间隔序列隔开(图 1-7)。像这样一种"超级基因"就称为操纵元或操纵子。由于转录子在转录后迅速翻译,所以就有可能等量合成几种在功能上相关的酶,这种蛋白质合成过程就称为协同控制(coordinate control)。当然,并非所有多顺反子 mRNA 的翻译都是协同控制的,在有些操纵元中,上游顺反子的翻译效率高于下游顺反子,而在另一些操纵元中,下游顺反子的翻译效率则可能高于上游顺反子。

图 1-7 的上图表示整个操纵元的结构,其中 *lacI*、*lacZ*、*lacY*、*lacA* 分别为调节基因及三个结构基因的符号,操纵元之下为这些基因编码的相应的产物,其中 C 为控制区,S 为内部转录间隔序列。下图为控制区的详细结构,数字表示某个核苷酸距转录起点核苷酸的距离,转录起点上游的核苷酸以负数表示。

有些操纵元的表达是相当稳定的,也就是说由其结构基因编码的若干种酶能够稳定地合成,较少受到细胞内、外其他因素的影响,这类操纵元称为组成型操纵元(constitutive operon)。还有一些操纵元的表达则受到严格控制,它们受细菌细胞中某些代谢产物的调节。如果某种代谢产物能够促使操纵元表达,这种操纵元称为诱导型操纵元(inducible operon)。如果操纵元的表