



第七讲

真核基因表达调控

真核生物（除酵母、藻类和原生动物等单细胞类之外）主要由多细胞组成，每个细胞基因组中蕴藏的遗传信息量及基因数量都大大高于原核生物。人类细胞单倍体基因组有 3×10^9 bp，为大肠杆菌总DNA的800倍，噬菌体的10万左右！



自然选择倾向于保留高效率的生命过程。在每30分钟增殖一次的 10^9 细菌群体中，若一个细菌变成29.5分钟增殖，经过80天的连续生长后，这个群体中的99.9%都将具有29.5分钟增殖一倍的生长速度。



每个大肠杆菌细胞有约**15 000**个核糖体，**50**种核糖体蛋白、糖酵解体系的酶、**DNA**聚合酶、**RNA**聚合酶等都是代谢过程中必需的，其合成速率不受环境变化或代谢状态的影响，这一类蛋白质被称为永久型（**constitutive**）合成的蛋白质。



另一类则被称为适应型或调节型 (**adaptive or regulated**)，因为这类蛋白质的合成速率明显地受环境的影响而改变。如大肠杆菌细胞中一般只有15个 β -半乳糖苷酶，但若将细胞培养在只含乳糖的培养基中，每细胞中这个酶的量可高达几万个分子。



真核生物染色质被包裹在细胞核内，基因的转录（核内）和翻译（细胞质内）被核膜所隔开，核内RNA的合成与转运，细胞质中RNA的剪接和加工等都属于真核生物基因调控的范围。



随着生物个体的发育，**DNA**分子能有序地将其所承载的遗传信息，通过密码子-反密码子系统转变成蛋白质，执行各种生理生化功能。

从**DNA**到蛋白质的过程称为基因表达（**gene expression**），对这个过程的调节就称为基因表达调控（**gene regulation**或**gene control**）。

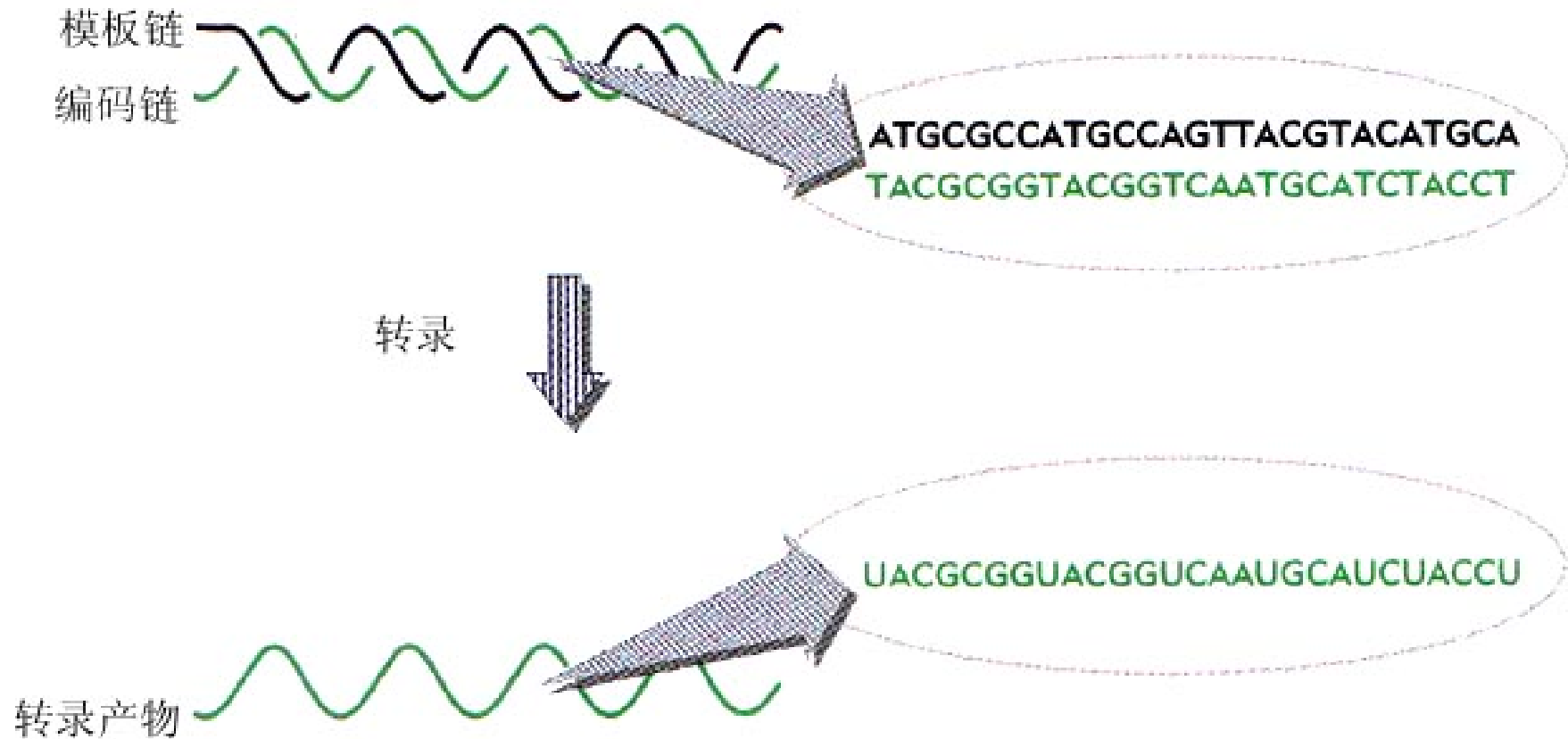


图7-1基因表达的第一步是由**RNA**聚合酶拷贝**DNA**双链中的模板链，生成与该序列完全互补的**RNA**链。



基因表达调控主要表现在以下二方面：

转录水平上的调控（**transcriptional regulation**）；

转录后水平上的调控（**post-transcriptional regulation**），包括

（1）**mRNA**加工成熟水平调控（**differential processing of RNA transcript**）；

（2）翻译水平调控（**differential translation of mRNA**）。



真核生物基因调控主要包括：

瞬时调控或称可逆性调控；

发育调控又称不可逆调控，是真核基因调控的精髓部分，它决定了真核细胞生长、分化、发育的全部进程。

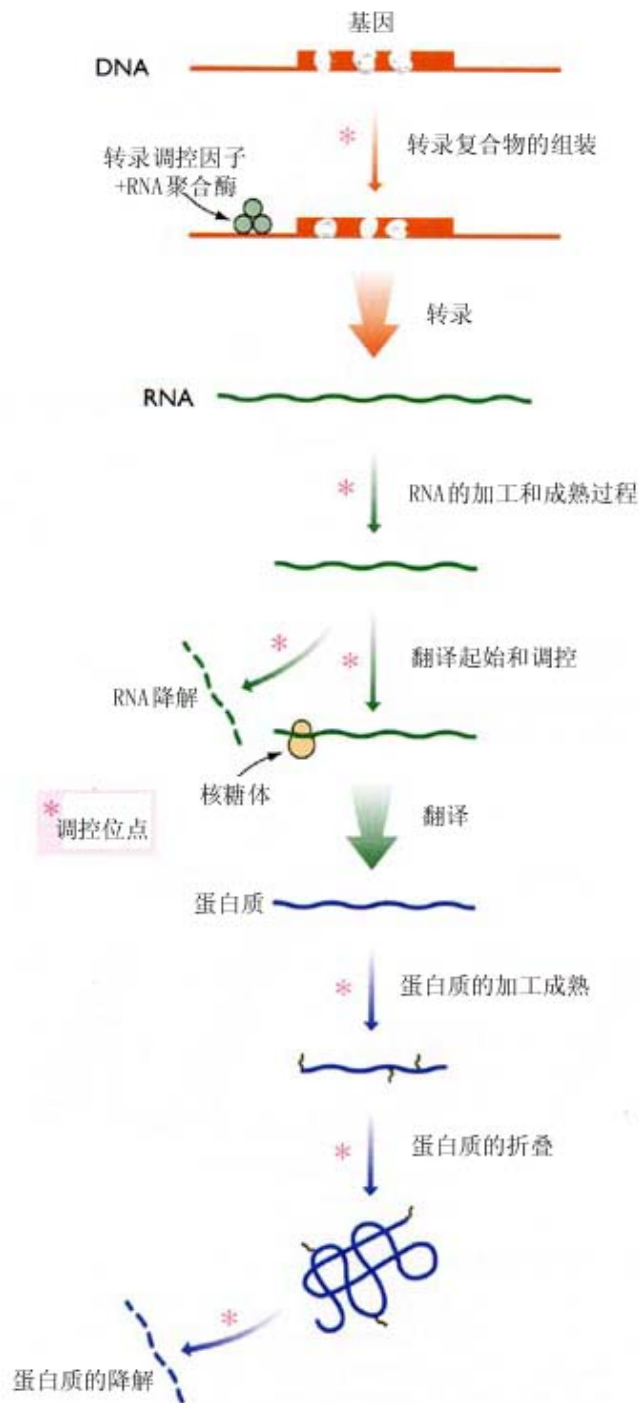


图7-2 真核基因表达调控的主要步骤示意图



7.1 真核生物的基因结构与转录活性

- ①在真核细胞中，一条成熟的mRNA链只能翻译出一条多肽链，很少原核生物中常见的多基因操纵子形式。
- ②真核生物能够有序地根据生长发育阶段的需要进行DNA片段重排，增加细胞内某些基因的拷贝数。



③基因转录的调节区很大，可能远离转录起始位点达几百个甚至上千个碱基对，主要通过改变整个所控制基因5'上游区DNA构型来影响它与RNA聚合酶的结合能力。



④真核生物的RNA在细胞核中合成，只有经转运穿过核膜，到达细胞质后，才能被翻译成蛋白质。

⑤许多真核生物基因只有经过复杂的成熟和剪接过程，才能被顺利地翻译成蛋白质。



7. 1. 1 基因家族 (gene family)

真核细胞中许多相关的基因常按功能成套组合，被称为基因家族。

同一家族中的成员有时紧密地排列在一起，成为一个基因簇，也可能分散在同一染色体的不同部位，甚至位于不同的染色体上，具有各自不同的表达调控模式。



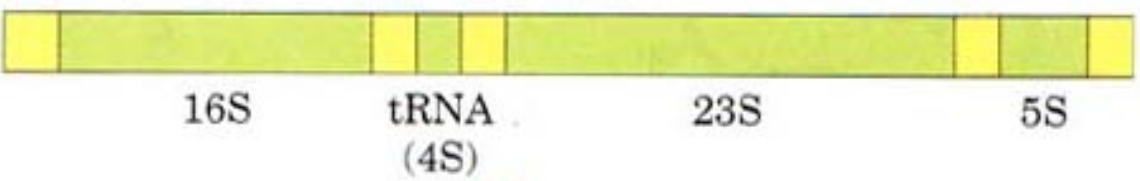
1、简单多基因家族

简单多基因家族中的基因一般以串联方式前后相连。细菌中所有rRNA和部分tRNA都来自这个分子量为30S（约6500个核苷酸）的前rRNA。

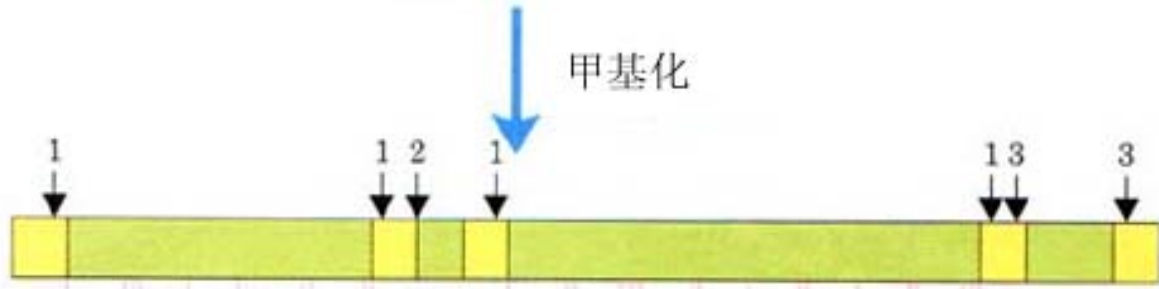
在真核生物中，前rRNA转录产物的分子量为45S，（约有14 000个核苷酸），包括18S，28S和5.8S三个主要rRNA分子。



前 rRNA 转录物



(a)



(b)



中间产物



(c)

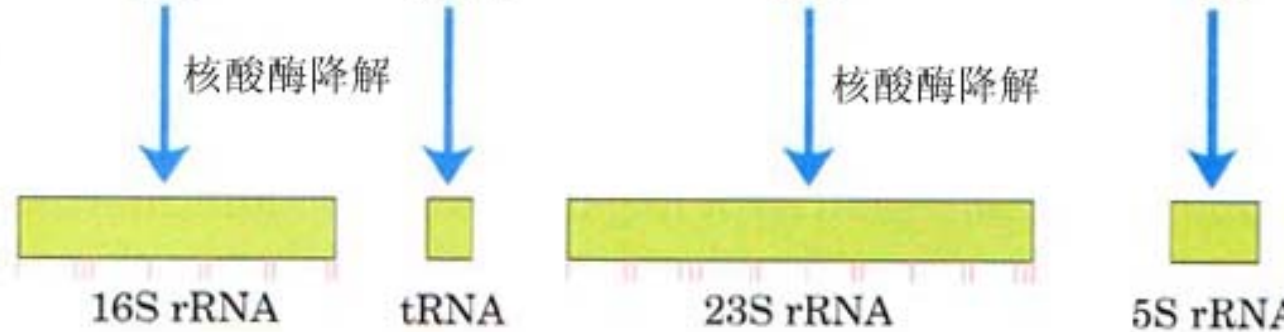


图7-3 细菌中rRNA基因家族各成员的分布与成熟过程分析

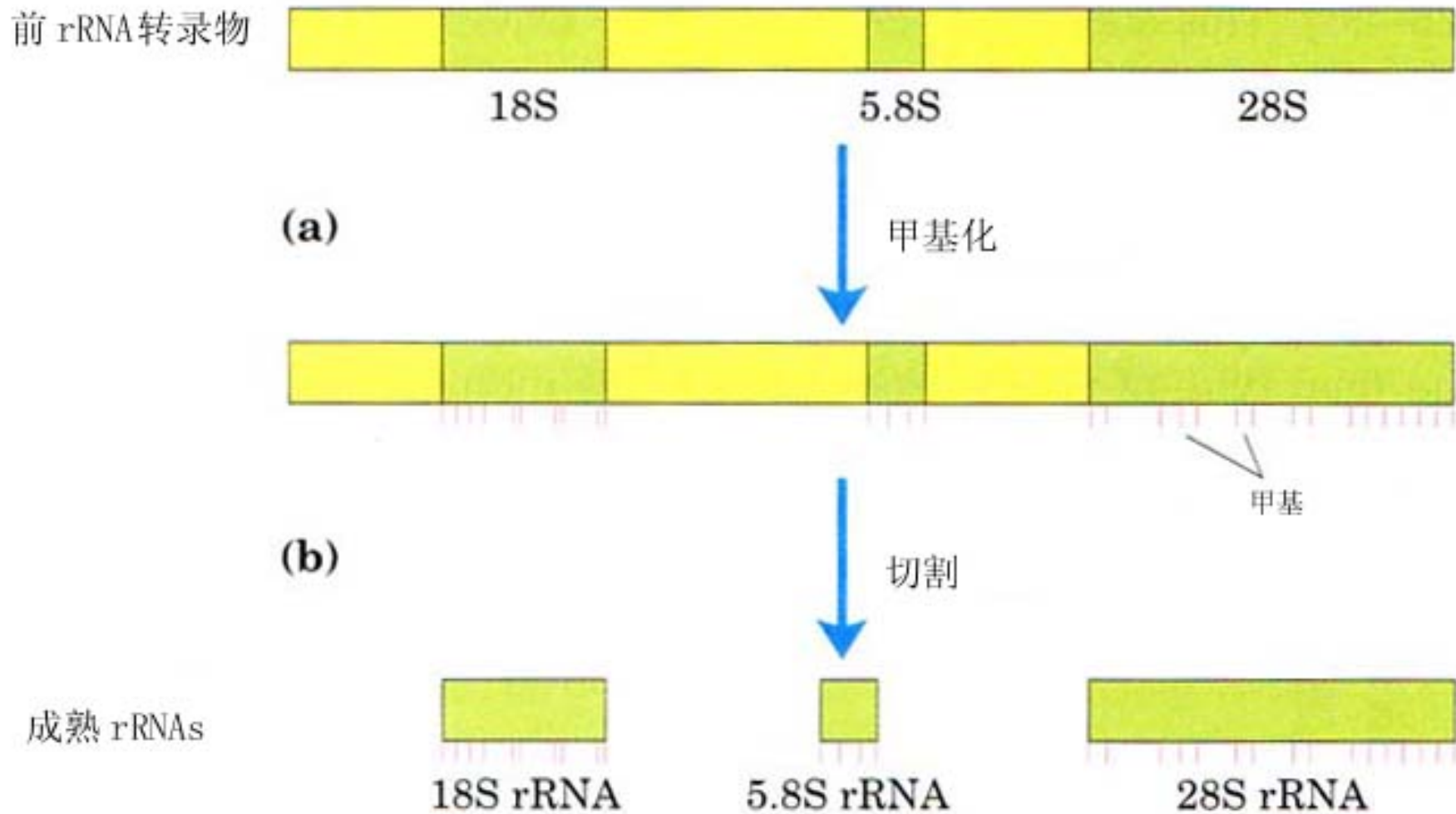


图7-3 脊椎动物中rRNA基因家族主要成员的分布与成熟过程分析。



2、复杂多基因家族

复杂多基因家族一般由几个相关基因家族构成，基因家族之间由间隔序列隔开，并作为独立的转录单位。

海胆组蛋白基因家族（图7-4）：编码不同组蛋白的基因处于一个约为6 000bp的片段中，分别被间隔序列所隔开。这5个基因组成的串联单位在整个海胆基因组中可能重复多达1 000次。

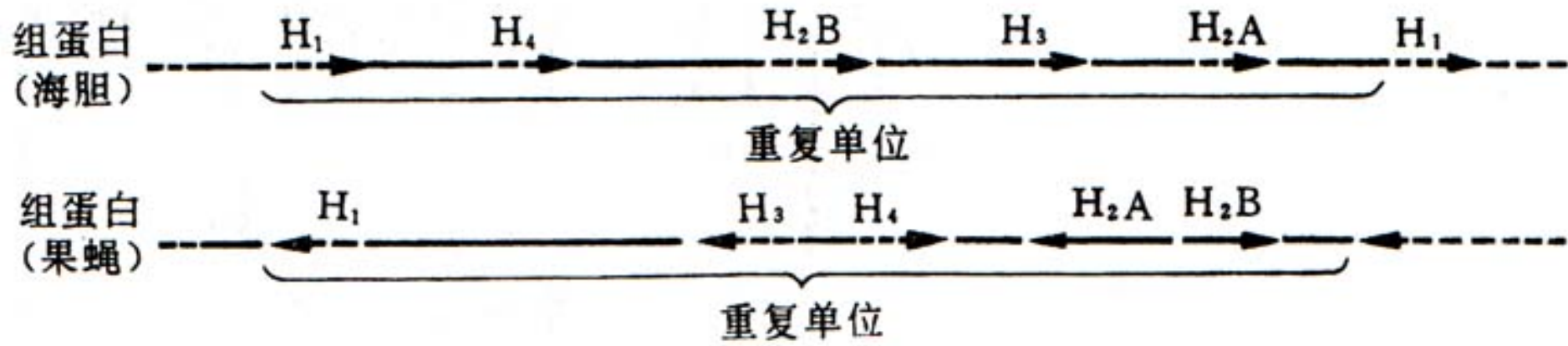


图 7-5 复杂的多基因家族



3、发育调控的复杂多基因家族

血红蛋白是所有动物体内输送分子氧的主要载体，由 $2\alpha 2\beta$ 组成的四聚体加上一个血红素辅基（结合铁原子）后形成功能性血红蛋白。在生物个体发育的不同阶段出现几种不同形式的 α 和 β 亚基。



人 α 珠蛋白基因簇位于16号染色体短臂上，约占30kb左右，其中 ζ 为胚胎期基因（表7-1）。 β 珠蛋白基因簇位于11号染色体短臂上，约占50-60kb，其中 ϵ 为胚胎期基因， $G\gamma$ 和 $A\gamma$ 为胎儿型基因， δ 和 β 为成人期基因。



表7-1 不同发育阶段血红蛋白亚型

发育阶段	组成
胚胎期（8周以前）	$\zeta_2 \epsilon_2$ 、 $\zeta_2 \gamma_2$ 和 $\alpha_2 \epsilon_2$
胎儿期（8~41周）	$\alpha_2 \gamma_2$
成人期（出生以前）	$\alpha_2 \delta_2$ 和 $\alpha_2 \beta_2$

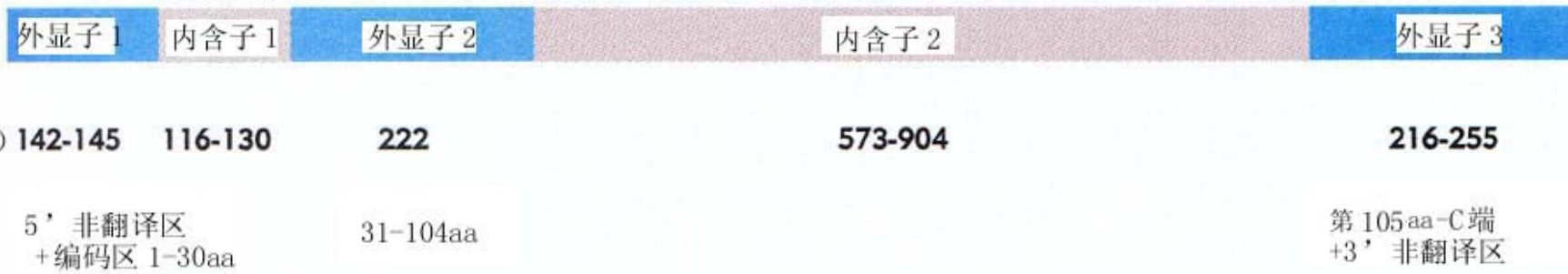


图7-5 人 β -珠蛋白基因的基本结构。

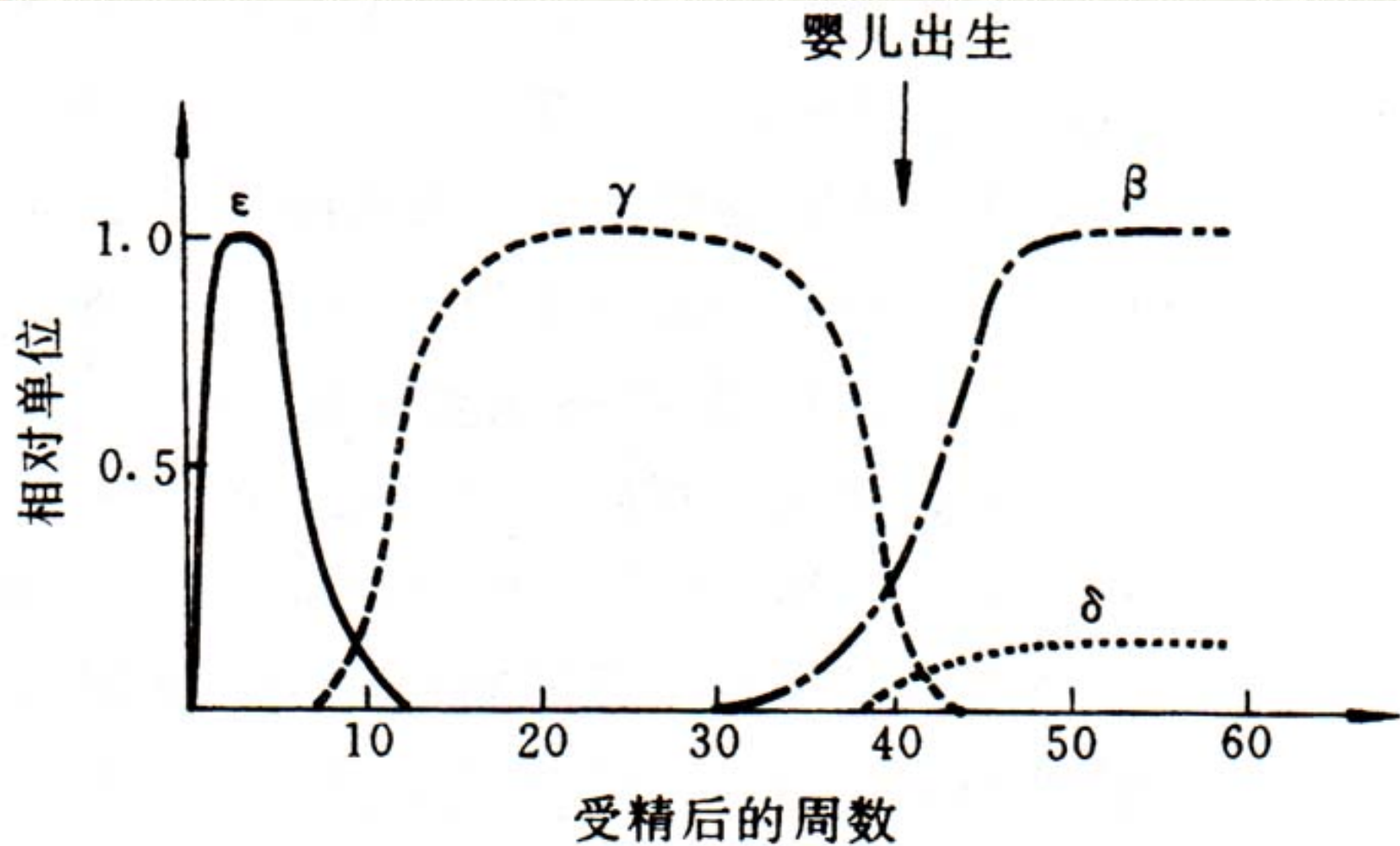


图7-7 人体发育过程中不同类型β-珠蛋白的含量变化

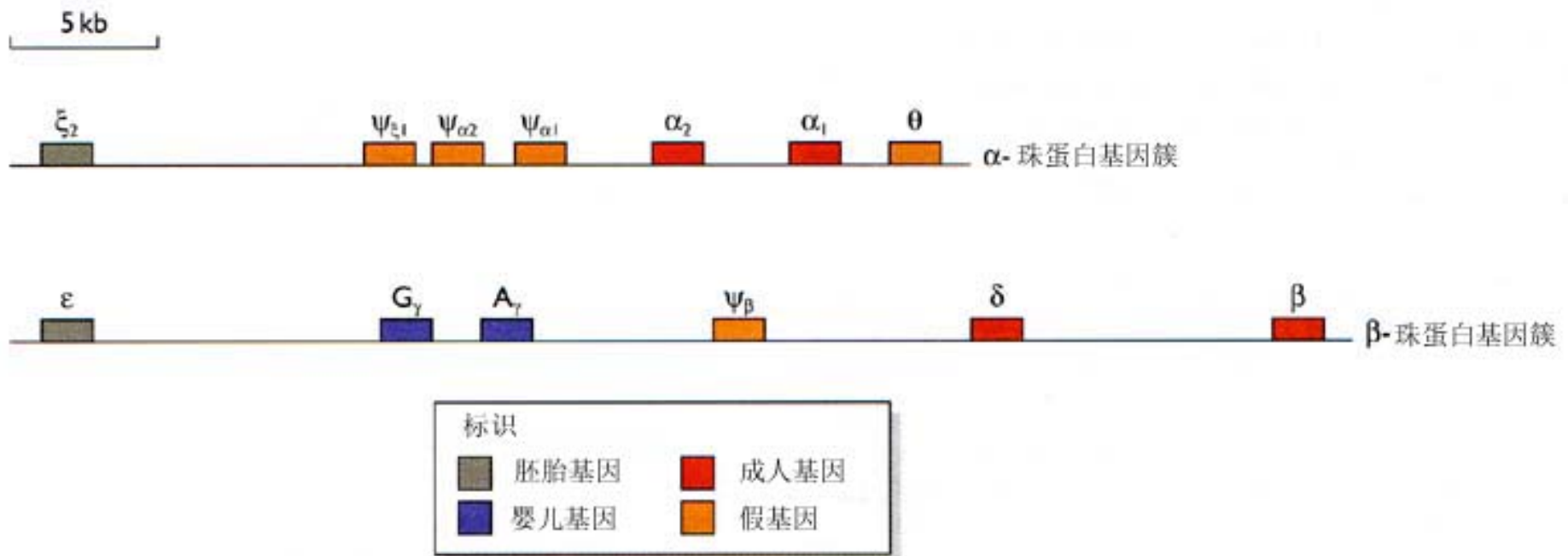


图7-7 人细胞中 α 和 β -珠蛋白基因簇结构示意图



7. 1. 2 真核基因的断裂结构

1、外显子与内含子

“intron”是指存在于原始转录物或基因组DNA中，但不存在于成熟mRNA、rRNA或tRNA中的那部分核苷酸序列。大多数真核基因都是由蛋白质编码序列和非蛋白质编码序列两部分组成的。



基因中的内含子数量和大小都不同。胶原蛋白基因长约40kb，至少有40个内含子，其中短的只有50bp，长的可达到2000bp。



表7-2 不同真核生物基因的平均长度及单个基因平均含有外显子数量比较

物 种	外显子数/基 因	核基因平均长度(kb)	mRNA平均长 度
酵 母	1	1.6	1.6
真 菌	3	1.5	1.5
线 虫	4	4.0	3.0
果 蝇	4	11.3	2.7
鸡	9	13.9	2.4
哺乳类	7	16.6	2.2



图7-8是哺乳动物二氢叶酸还原酶基因，全长25-31kb左右，但其6个外显子总长只有2kb。少数基因，如组蛋白及 α 型、 β 型干扰素基因，根本不带内含子。

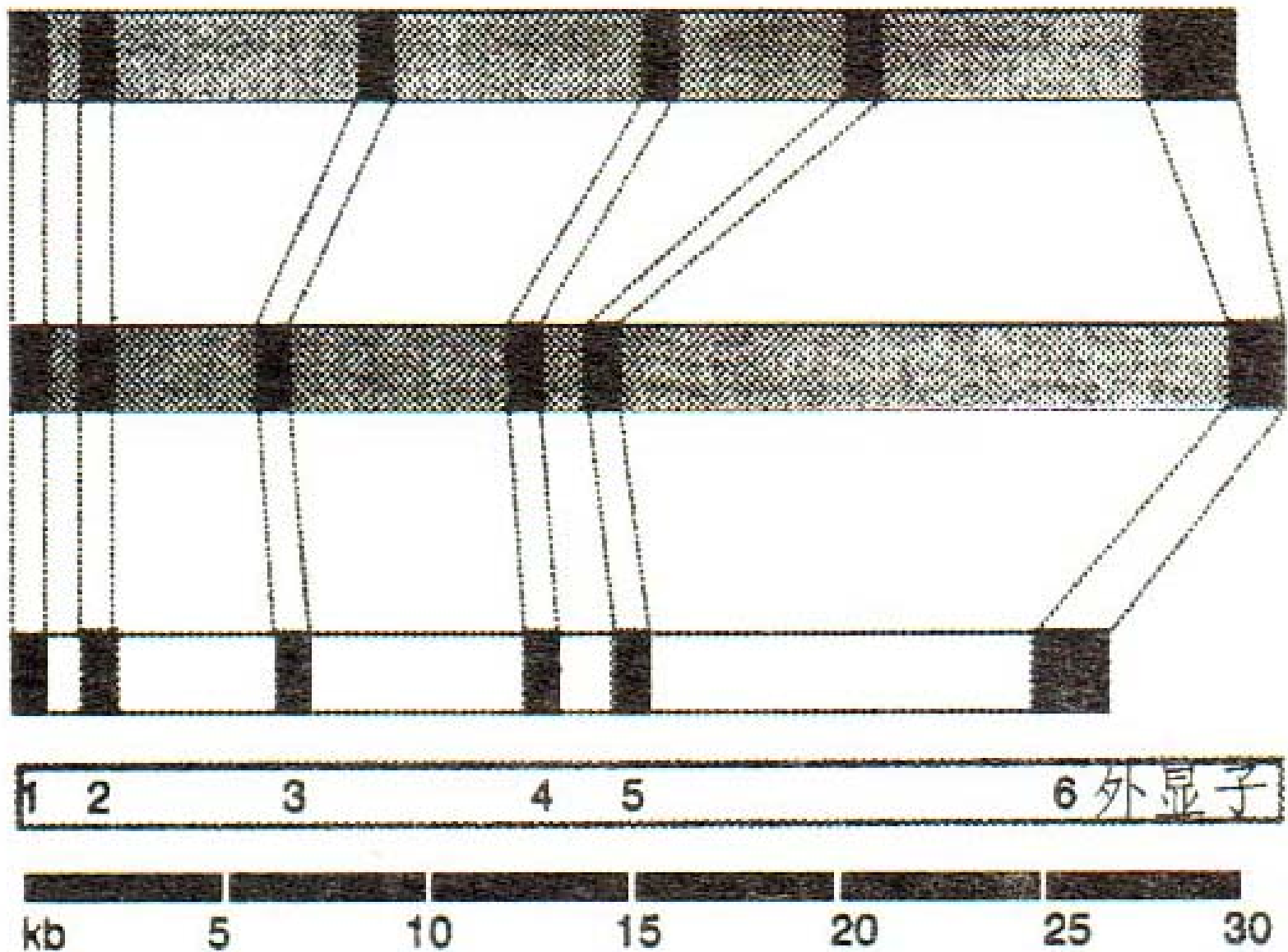


图7-8 哺乳动物二氢叶酸还原酶的基因结构示意图



2、外显子与内含子的连接区。

断裂结构的一个重要特点是外显子-内含子连接区（exon—intron junction）的高度保守性和特异性碱基序列。内含子的两端序列之间没有广泛的同源性，因此内含子两端序列不能互补，说明在剪接加工之前，内含子上游序列和下游序列不可能通过碱基配对形成发卡式二级结构。



序列分析表明，几乎每个内含子5' 端起始的两个碱基都是GT，而3' 端最后两个碱基总是AG，由于这两个碱基的高度保守性和广泛性，有人把它称为GT-AG法则，即：5' GT.....AG 3' 。



3、外显子与内含子的可变调控

真核基因的原始转录产物可通过不同的剪接产生不同的mRNA，翻译成不同的蛋白质。有些真核基因，如肌红蛋白重链基因虽有41个外显子，却能精确地剪接成一个成熟的mRNA，我们称这种方式为**组成型剪接**。一个基因的转录产物通过组成型剪接只能产生一种成熟的mRNA，编码一个多肽。



有些基因选择了不同的启动子，或者选择了不同的多聚（A）位点而使原始转录物具有不同的二级结构，产生不同的mRNA分子。同一基因的转录产物由于不同的剪接方式形成不同mRNA的过程称为**选择性剪接**。

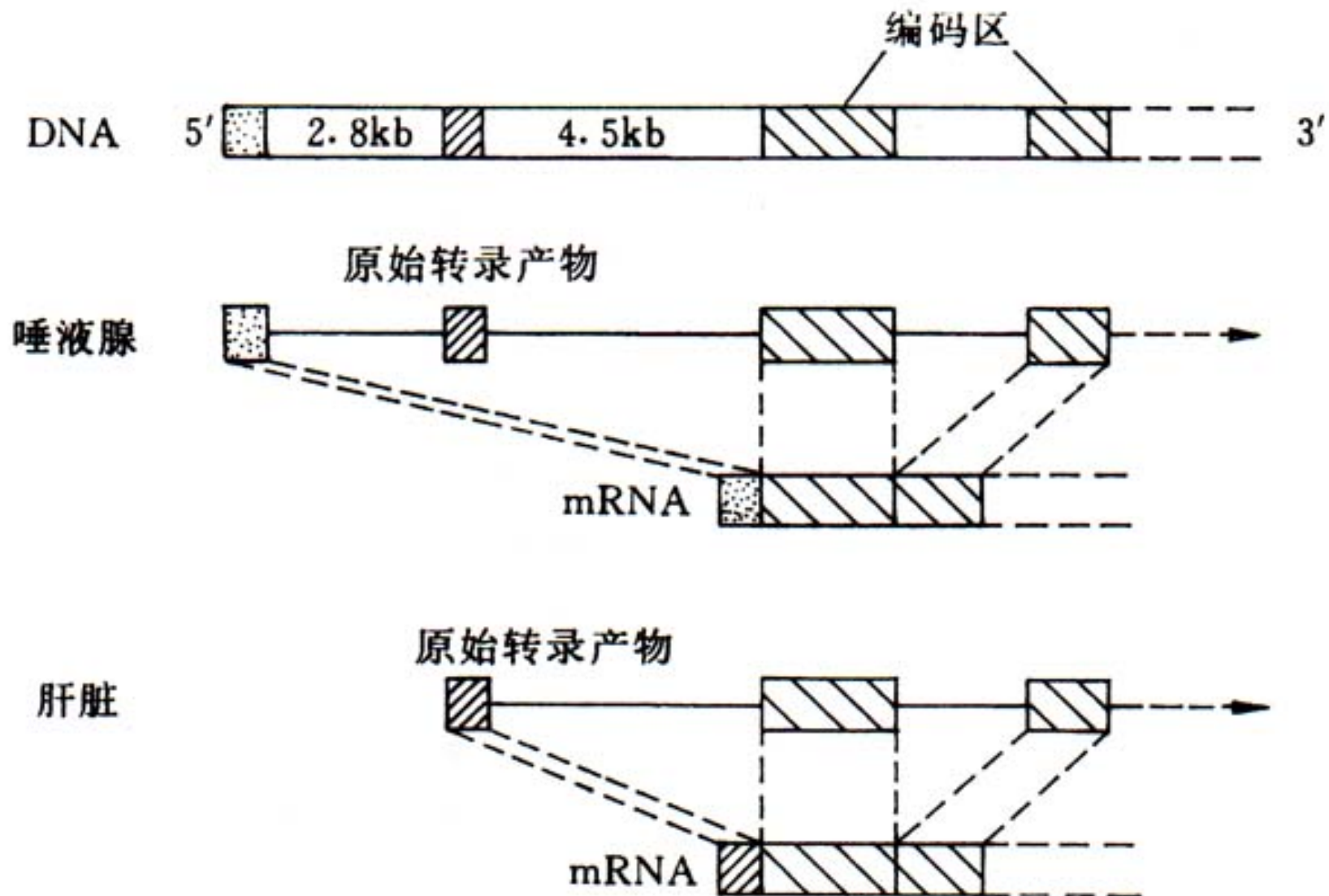


图7-10 不同外显子的使用导致 α -淀粉酶基因在不同组织中表达水平很不相同



在肝脏中，mRNA 5' 端的161个碱基是由位于第2号外显子转录起始点上游4 500bp处的L外显子编码的。

在唾液腺中，mRNA 5' 端的50个碱基是由位于转录起始点上游7 300bp处的S外显子编码的。



L外显子和S外显子分别为淀粉酶mRNA提供了不同的起始序列。其实，L外显子只是唾液腺淀粉酶基因中内含子序列的一部分，将在mRNA成熟过程中被切除。

一个基因的内含子成为另一个基因的外显子，形成基因的差别表达，这是真核基因的一个重要特点。



7. 1. 3 真核生物DNA水平上的 基因表达调控

DNA水平的调控是真核生物发育调控的一种形式，包括**基因丢失、扩增、重排和移位**等。



1、基因扩增

基因扩增是指某些基因的拷贝数专一性大量增加的现象，它使细胞在短期内产生大量的基因产物以满足生长发育的需要，是基因活性调控的一种方式。

非洲爪蟾的卵母细胞中原有rRNA基因（rDNA）约500个拷贝，在减数分裂粗线期，基因开始迅速复制，到双线期拷贝数约为200万个，扩增近4000倍，可用于合成 10^{12} 个核糖体。



2、基因重排与变换

将一个基因从远离启动子的地方移到较近的位点从而启动转录，被称为基因重排。

免疫球蛋白的肽链主要由可变区（V区）、恒定区（C区）以及两者之间的连接区（J区）组成，V、C和J基因片段在胚胎细胞中相隔较远。编码产生免疫球蛋白的细胞发育分化时，通过染色体内DNA重组把4个相隔较远的基因片段连接在一起，产生具有表达活性的免疫球蛋白基因。



表7-3 人类基因组中免疫球蛋白基因主要片段的数量比较

成份	基因位点	染色体	基因片段数量			
			V	D	J	C
重链	IGH	14	86	30	9	11
轻链(K链)	IGK	2	76	0	5	1
轻链(λ 链)	IGL	22	52	0	7	7



(A) IGH基因座的分布情况



(B) 通过基因重排 (rearrangement) 获得功能型免疫球蛋白重链基因

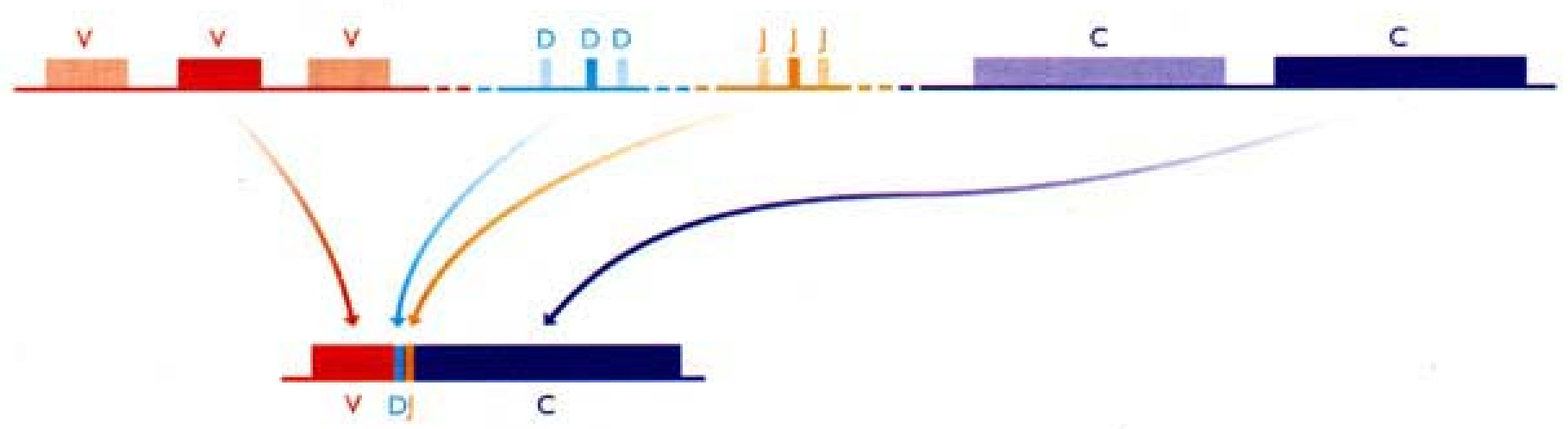


图7-11 免疫球蛋白重链基因片段重排与组织特异性表达



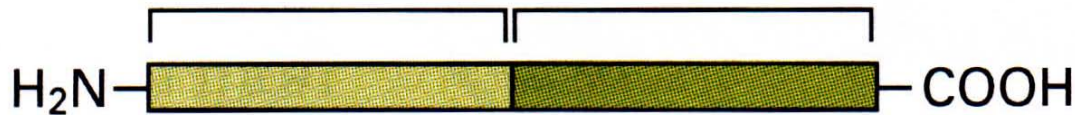
抗体分子由4条（两对）多肽链组成，包括两条相同的轻链（L-chain）和两条相同的重链（H-chain）。轻链和重链在相对分子质量上有较大差别，前者约 2.3×10^4 ，后者则介于 5.3×10^4 – 7.0×10^4 之间。



LIGHT CHAIN

variable
region

constant
region (κ type or λ type)



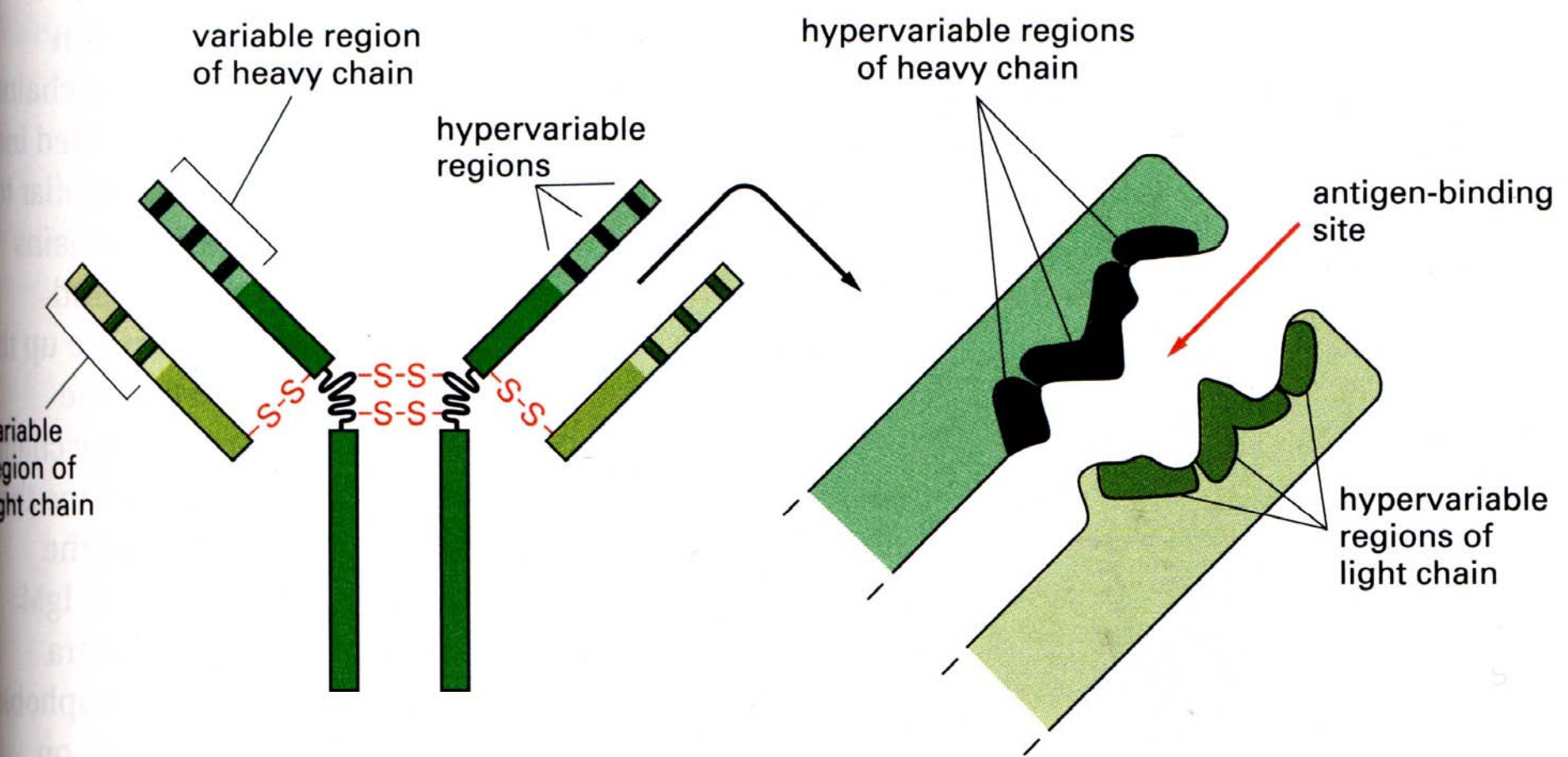
HEAVY CHAIN



variable region

constant region
(α , δ , ϵ , γ , or μ type)

23 - 30



23-31



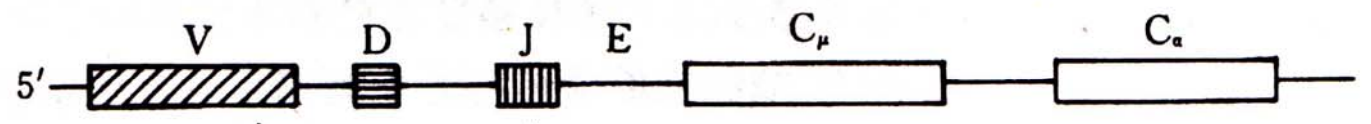
所有Ig分子都含有两类轻链中的一类，即 κ 型或 λ 型。

κ 型和 λ 型轻链的恒定区和可变区的氨基酸序列是不同的。

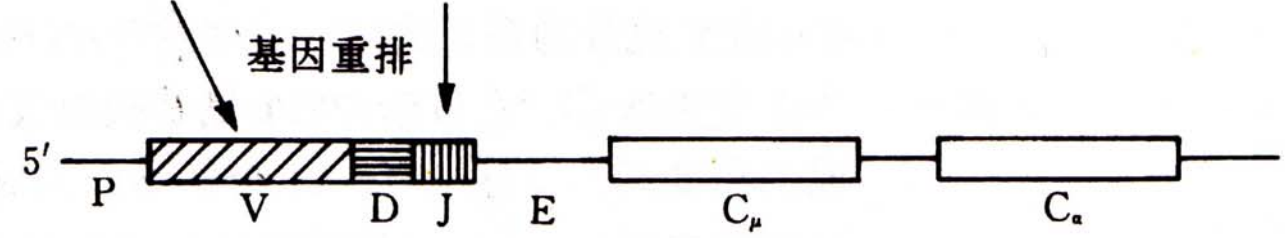
在小鼠中，95%的抗体轻链是 κ 型，而人类抗体轻链中， κ 型和 λ 型各占50%左右。



不表达 Ig 基因的
种系 DNA



表达 C_μ 基因的
DNA 结构



表达 C_α 基因的
DNA 结构

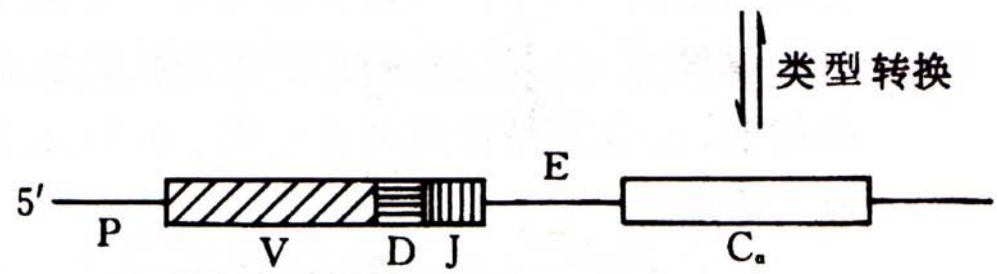


图 7-20 增强子序列 (E) 对免疫球蛋白重链基因表达的影响



免疫球蛋白重链基因DNA重排以后，大量间隔序列被切除，使位于J-C μ 之间的增强子序列得以发挥作用，增强基因转录。



7. 1. 4 DNA甲基化与基因调控

DNA甲基化能引起染色质结构、DNA构象、DNA稳定性及DNA与蛋白质相互作用方式的改变，从而控制基因表达。研究证实，CpG二核苷酸中胞嘧啶的甲基化导致了人体1/3以上由于碱基转换而引起的遗传病。



DNA 甲基化主要形成5-甲基胞嘧啶（5-mC）和少量的N⁶-甲基腺嘌呤（N⁶-mA）及7-甲基鸟嘌呤（7-mG）。真核生物中，5-甲基胞嘧啶主要出现在CpG、CpXpG、CCA/TGG和GATC中。



高等生物CpG二核苷酸中的C通常被甲基化，极易自发脱氨，生成胸腺嘧啶，所以CpG二核苷酸序列出现的频率远远低于按核苷酸组成计算出的频率。

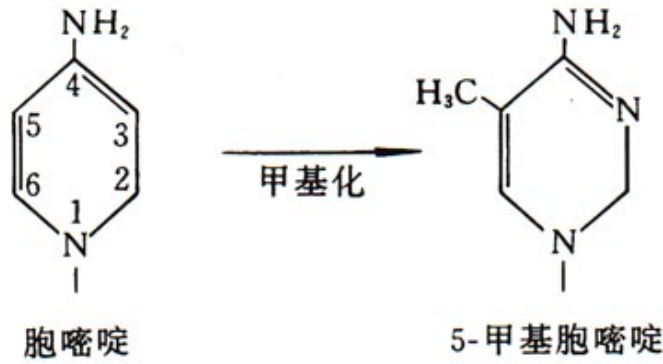


图 7-14 5-甲基胞嘧啶、N⁶-甲基腺嘌呤和 7-甲基鸟嘌呤结构图



对于弱启动子来说，稀少的甲基化就能使其完全失去转录活性。当这一类启动子被增强时（带有增强子），即使不去甲基化也可以恢复其转录活性。若进一步提高甲基化密度，即使增强后的启动子仍无转录活性。



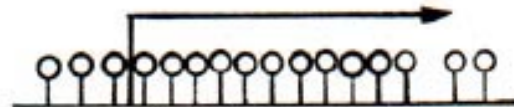
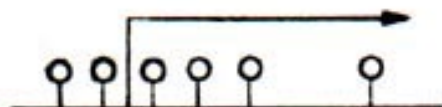
启动子强度 甲基化

低密度 CpG

高密度 CpG

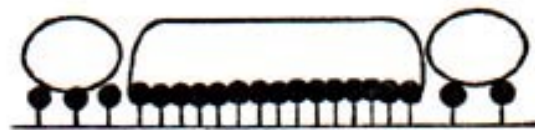
弱

-



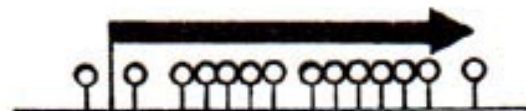
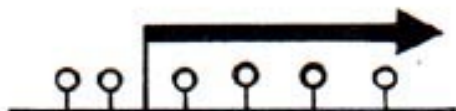
弱

+



强

-



强

+

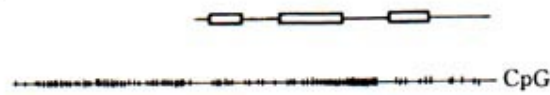


CG 岛

图 7-15 DNA 甲基化对基因转录的抑制作用 箭头粗细表示转录活性强弱，空心圆表示 DNA 未甲基化，小黑点表示 DNA 甲基化，椭圆形表示 MeCP1 松散结合，长方形表示 MeCP1 紧密结合



A. 人 α -珠蛋白基因



小鼠 α -珠蛋白基因



人 γ -珠蛋白基因



100bp

B.

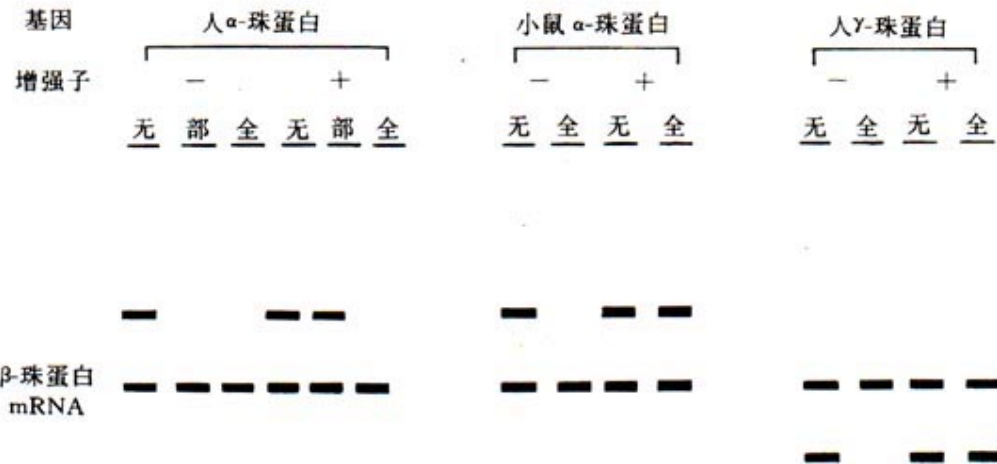


图 7-16 3 个珠蛋白的基因编码区及启动区的 CpG 密度 (A) 和甲基化对这些基因表达强度的影响 (B) 无: 没有甲基化; 部: 部分甲基化; 全: 全部甲基化; β -珠蛋白 mRNA 作为对照



7.2 真核基因的转录

真核基因调控主要在转录水平上进行，受大量特定的**顺式作用元件**（*cis-acting element*）和**反式作用因子**（*transacting factor*，又称跨域作用因子）调控。



一个完整的基因，不但包括编码区（**coding region**），还包括5'和3'端长度不等的特异性序列，它们虽然不编码氨基酸，却在基因表达的过程中起着重要作用。所以，“基因”的分子生物学定义是：**产生一条多肽链或功能RNA所必需的全部核苷酸序列。**

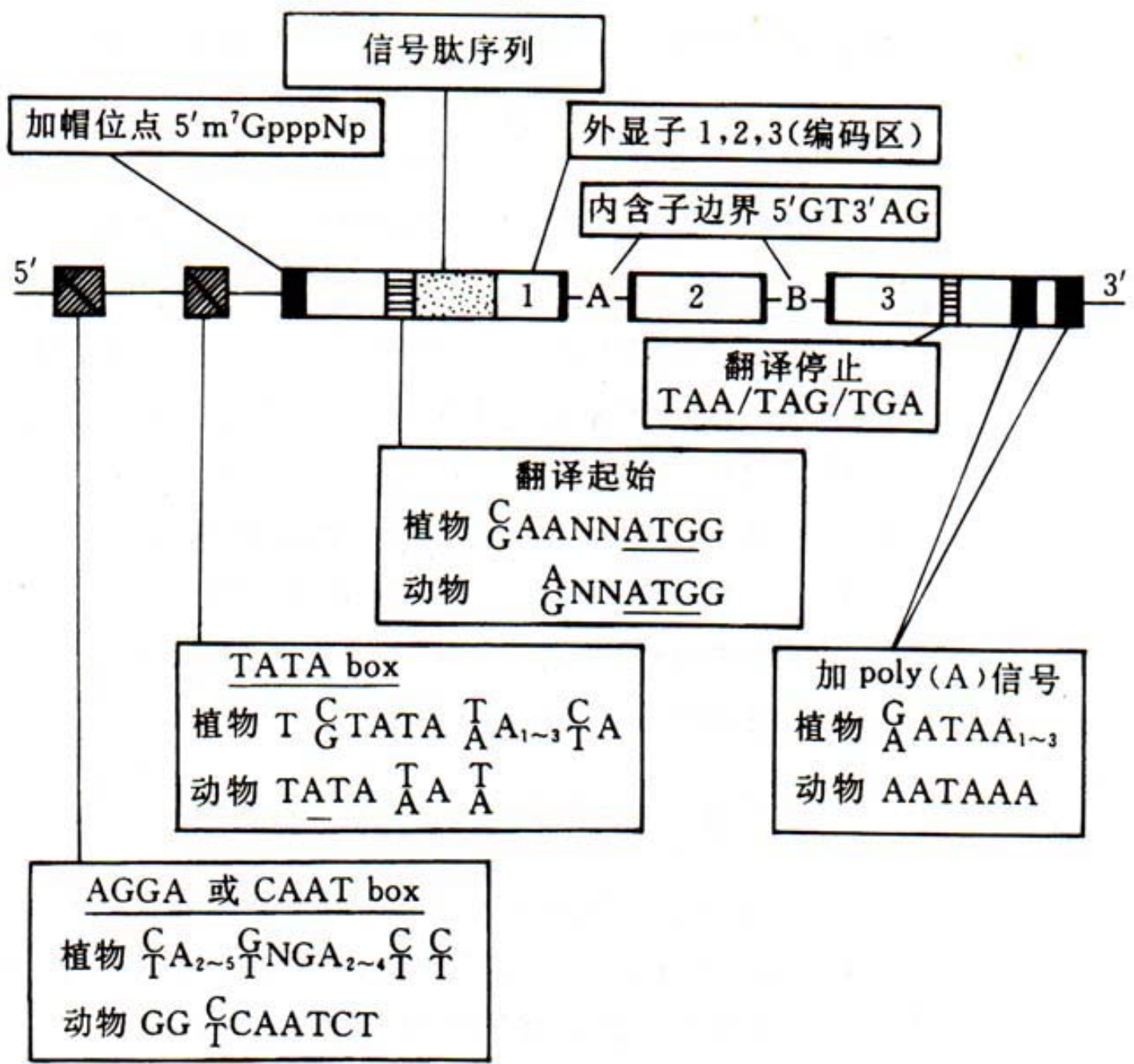


图7-18 真核基因的一般构造示意图



1、启动子（promoter）。

真核基因启动子由核心启动子和上游启动子两个部分组成，是在基因转录起始位点（+1）及其5'上游大约100~200bp以内的一组具有独立功能的DNA序列，每个元件长度约为7~20bp，是决定RNA聚合酶II转录起始点和转录频率的关键元件。



核心启动子 (core promoter) :
是指保证RNA聚合酶II转录正常起始所必需的、最少的DNA序列，包括转录起始位点及转录起始位点上游-25~-30bp处的TATA盒。核心启动子确定转录起始位点并产生基础水平的转录。



上游启动子元件（**upstream promoter element, UPE**）包括通常位于**-70bp**附近的**CAAT盒**（**CCAAT**）和**GC盒**（**GGGCGG**）等，能通过**TFIID**复合物调节转录起始的频率，提高转录效率。



2、增强子及其对转录的影响。
增强子是指能使与它连锁的基因转录频率明显增加的**DNA**序列，最早发现于**SV40**早期基因的上游，有两个长**72bp**的正向重复序列。



增强子通常具有下列特性：

A、增强效应十分明显。

B、增强效应与其位置和取向无关。

C、大多为重复序列（50bp）。

D、其增强效应有严密的组织和细胞特异性。

E、没有基因专一性，可以在不同的基因组合上表现增强效应；

F、许多增强子还受外部信号的调控。



7.3 反式作用因子

根据各个蛋白质成分在转录中的作用，
能将整个复合物分为3部分：

参与所有或某些转录阶段的RNA聚合酶亚基，不具有基因特异性。

与转录的起始或终止有关的辅助因子，不具有基因特异性。

与特异调控序列结合的转录因子。

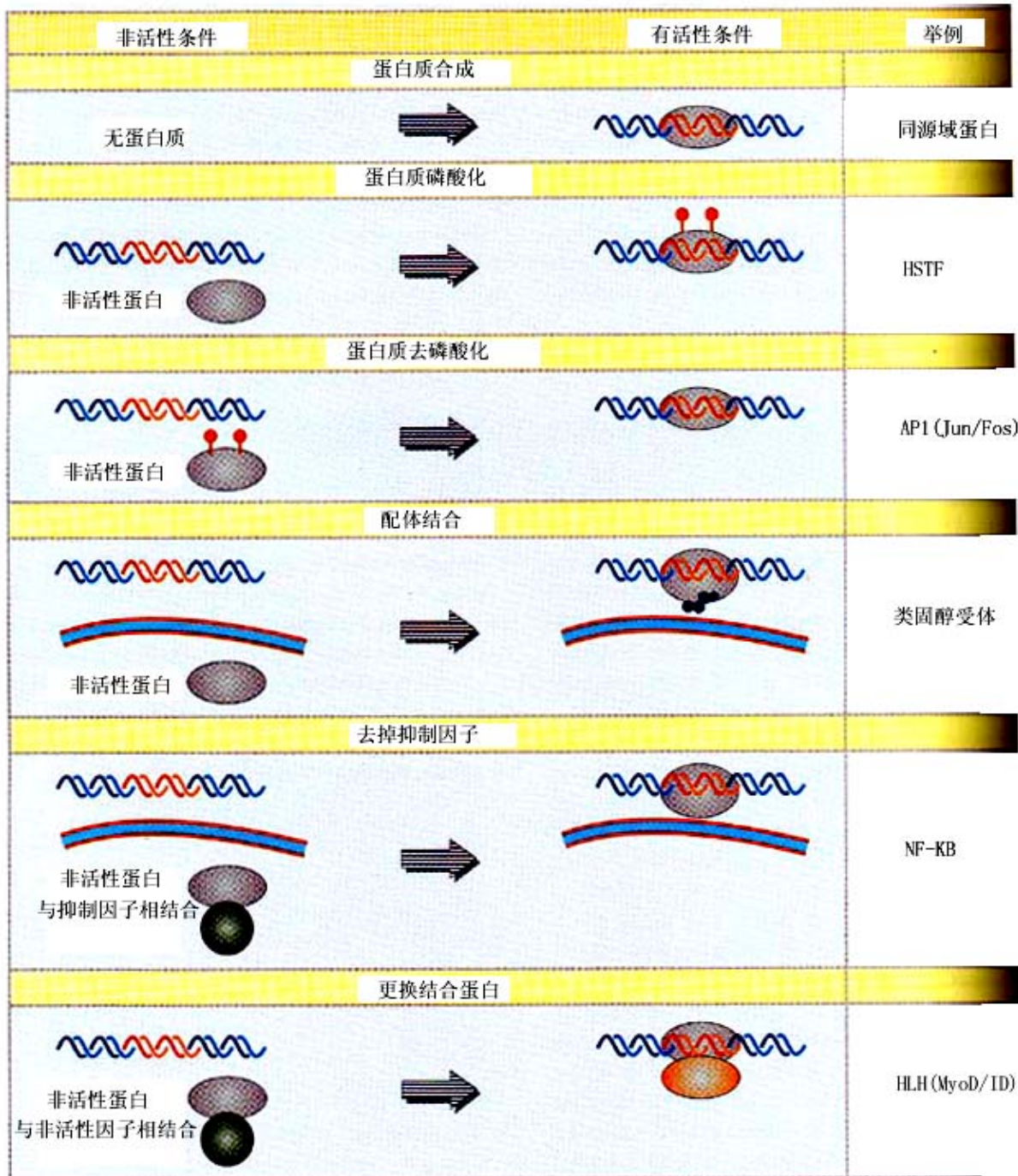


图7-17 真核生物中转录因子活性调节的主要方式



7.3.1 反式作用因子中的DNA识别或结合域

反式作用因子是能直接或间接地识别或结合在各类顺式作用元件核心序列上（表7-5），参与调控靶基因转录效率的蛋白质。



表7-5 部分球蛋白基因5'调控区序列的保守性比较。斜体表示高度保守的序列

基 因	-80区	-30区
小鼠 β^{min}	<i>GGCCAATCTGCTC</i>	<i>GGTATATAAA</i>
小鼠 β^{maj}	<i>GGCCAATCTGCTC</i>	<i>AGCATATAAG</i>
兔 β	<i>GGCCAATCTAC--</i>	<i>GGCATAAAAG</i>
山羊 β^{A}	<i>AGCCAATCTGCTC</i>	<i>GGCATAAAAG</i>
山羊 β^{C}	<i>AGCCAATCTGCTC</i>	<i>GGCATAAAAG</i>
人 β	<i>GGCCAATCTACTC</i>	<i>GGCATAAAAG</i>
人 δ	<i>AACCAATCTGCTC</i>	<i>-GCATAAAAG</i>
人 A^{γ}	<i>GACCAATAGCCTT</i>	<i>GGCATAAAAG</i>
人 G^{γ}	<i>GACCAATAGCCTT</i>	<i>GGCATAAAAG</i>
人 α	<i>GACCAATGACTTT</i>	<i>GGCATAAAAG</i>



A、螺旋-转折-螺旋（**helix-turn-helix, H-T-H**）结构。这一类蛋白质分子中有至少两个 α 螺旋，中间由短侧连氨基酸残基形成“转折”，近羧基端的 α 螺旋中氨基酸残基的替换会影响该蛋白质在**DNA**双螺旋大沟中的结合。

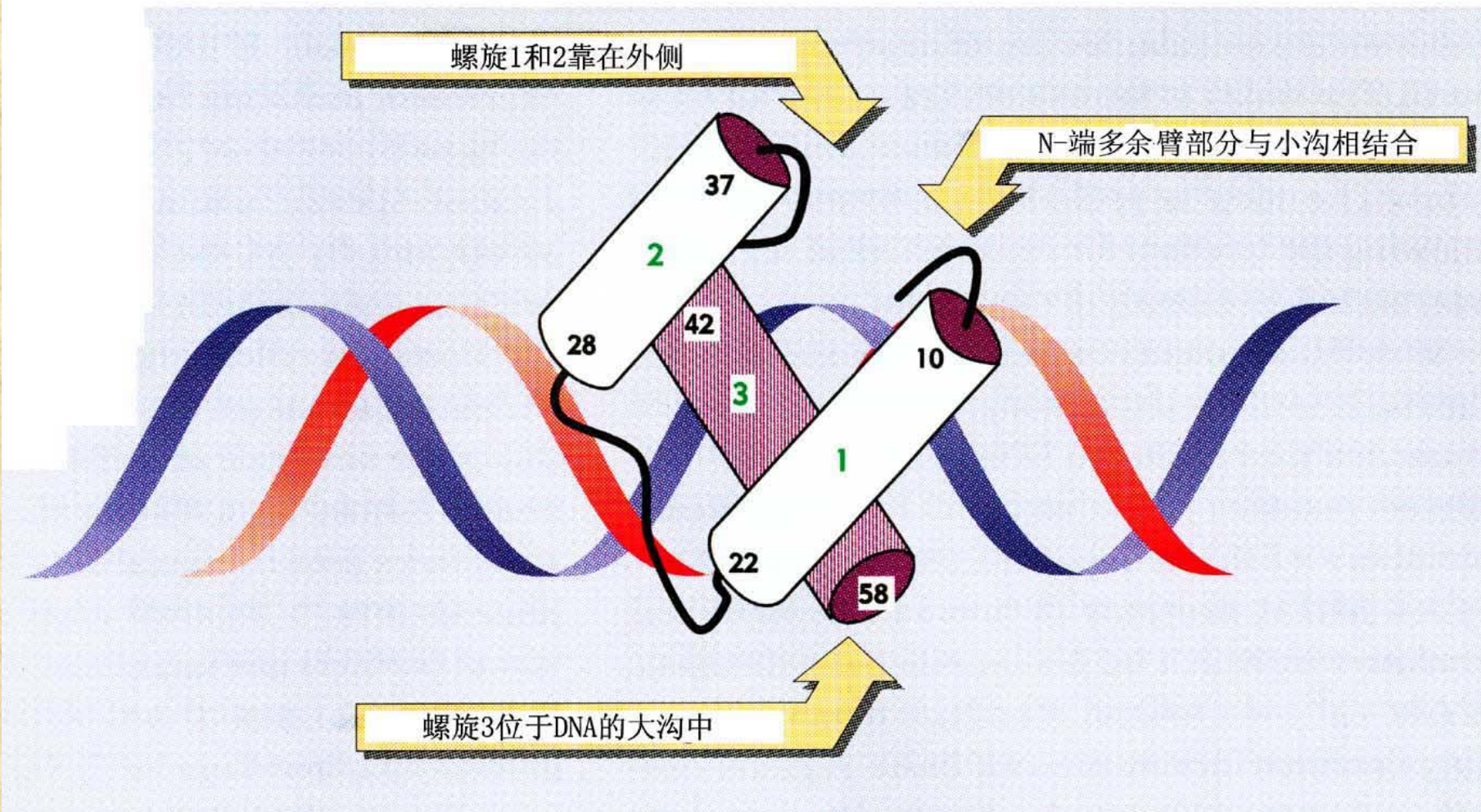


图7-18 同源域蛋白通过其第三个螺旋与双链DNA的大沟相结合，其N端的多余臂部分则与DNA的小沟相结合，提高了稳定性。



B、锌指（Zinc finger）蛋白：包括锌指、锌钮(**twist**)和锌簇(**cluster**)结构，有锌参与时才具备转录调控活性。与**DNA**的结合较为牢固，特异性也很高。

类固醇激素受体家族含有连续的两个锌指结构，以同源或异源性二聚体的方式将两个 α 螺旋结合在相邻的两个大沟中。

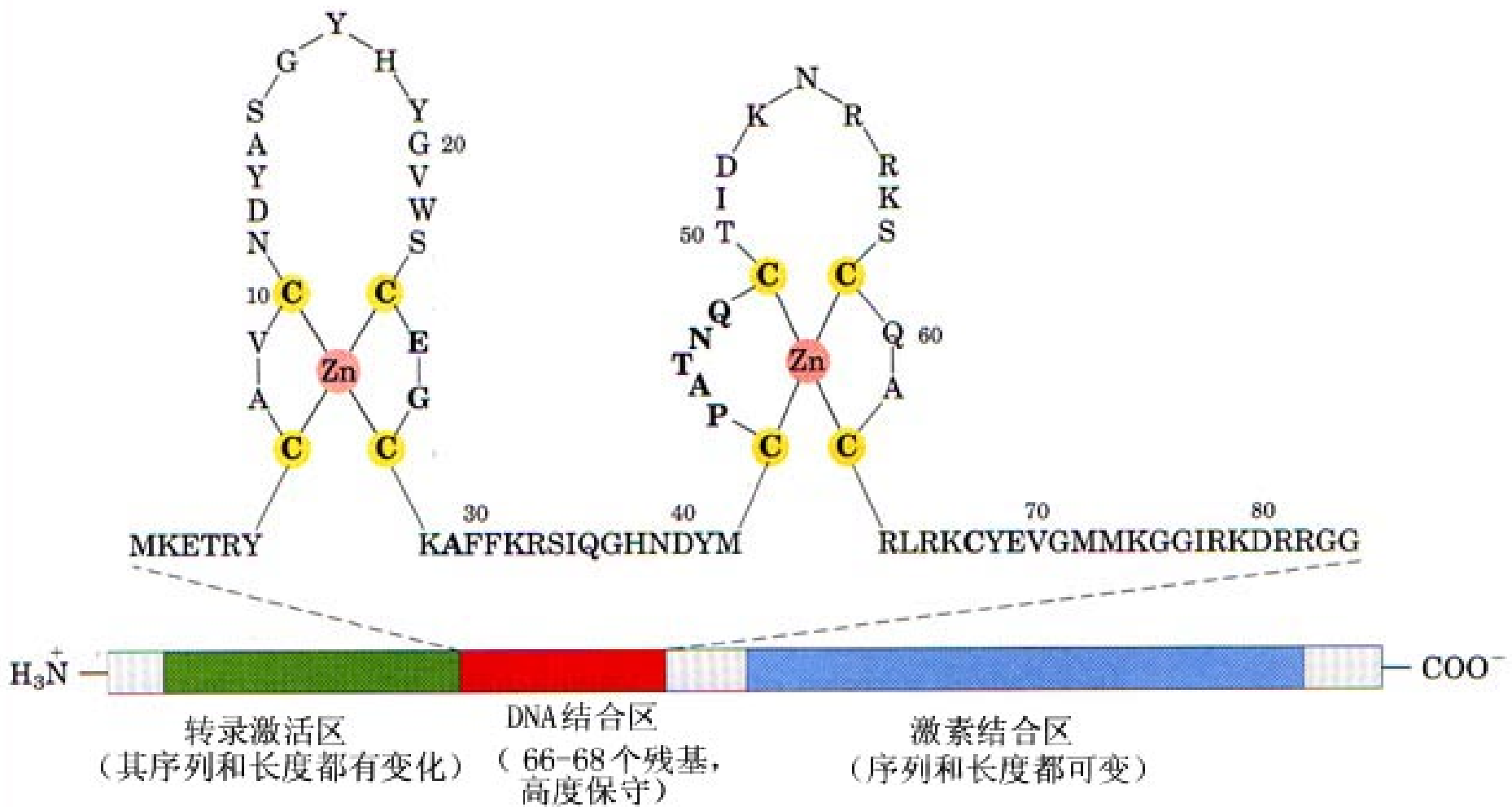


图7-19 典型的类固醇激素受体结构示意图



Cys2/Cys2 锌指序列:

Cys—X₂—Cys—X₁₃—

Cys—X₂—Cys



表7-6 一些转录因子和蛋白质通过Cys2/His2与DNA相结合

蛋白质	来源	相对分子质量	锌指数	DNA结合区/bp	靶序列	功能
TFIIIA	哺乳类	7.3×10^4	9	50	5S基因	与RNA聚合酶III作用，促进5S基因转录。
SP1	哺乳类	1.1×10^4	3	10	GC区	与RNA聚合酶II作用。广泛性调控因子。
ADR1	果蝇	1.5×10^4	2	22	ADH2基因	与RNA聚合酶II作用，激活ADH2基因。
TDF	哺乳类	?	13	?	?	决定雄性性别。
Kruppel	果蝇	6.0×10^4	5	?	?	胚胎分化。
Hunch-Back	果蝇	8.0×10^4	4+2	?	?	胚胎分化。



表7-7 具有Cys2/Cys2锌指区的转录因子

蛋 白	相对分子质量	锌指数	靶序列
皮质糖受体	9.4×10^4	2	GRE中20bp
雌激素受体	6.6×10^4	2	ERE中20bp
GAL4	9.9×10^4	1	UAS _G 中 17bp
腺病毒E1A	约 3.0×10^4	1	?



C、碱性-亮氨酸拉链（basic-leucine zipper），即bZIP结构。蛋白中每隔6个氨基酸就有一个亮氨酸残基，导致第7个亮氨酸残基都在螺旋的同一方向出现。

以二聚体形式与**DNA**结合，亮氨酸拉链区并不直接结合**DNA**，肽链氨基端**20~30**个富含碱性氨基酸结构域与**DNA**结合，但以碱性区和亮氨酸拉链结构域整体作为基础。

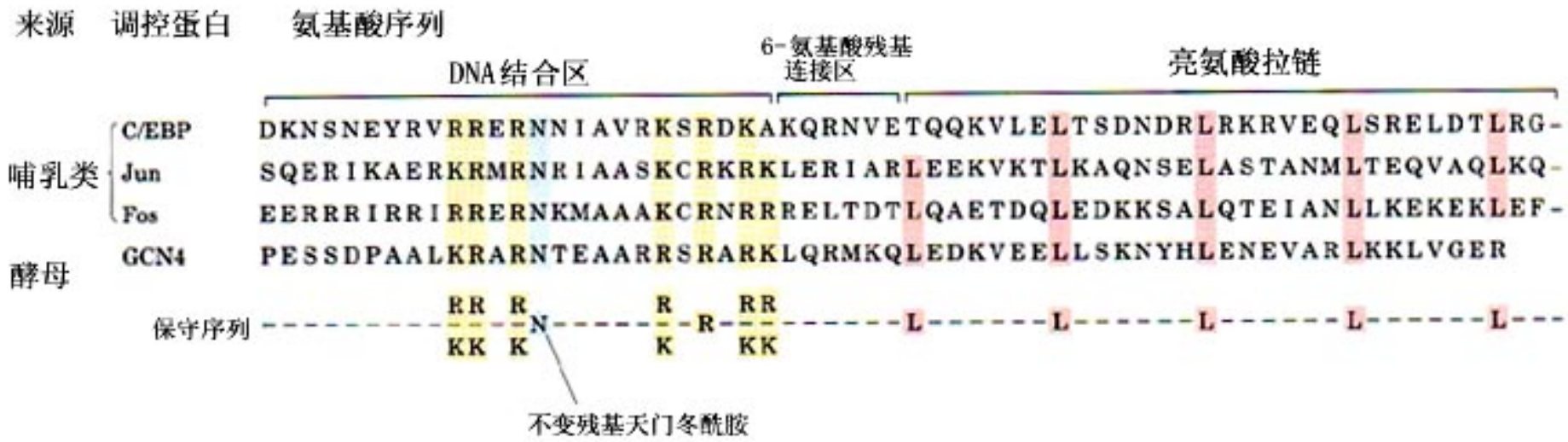


图7-20 碱性亮氨酸拉链结构域转录激活因子。

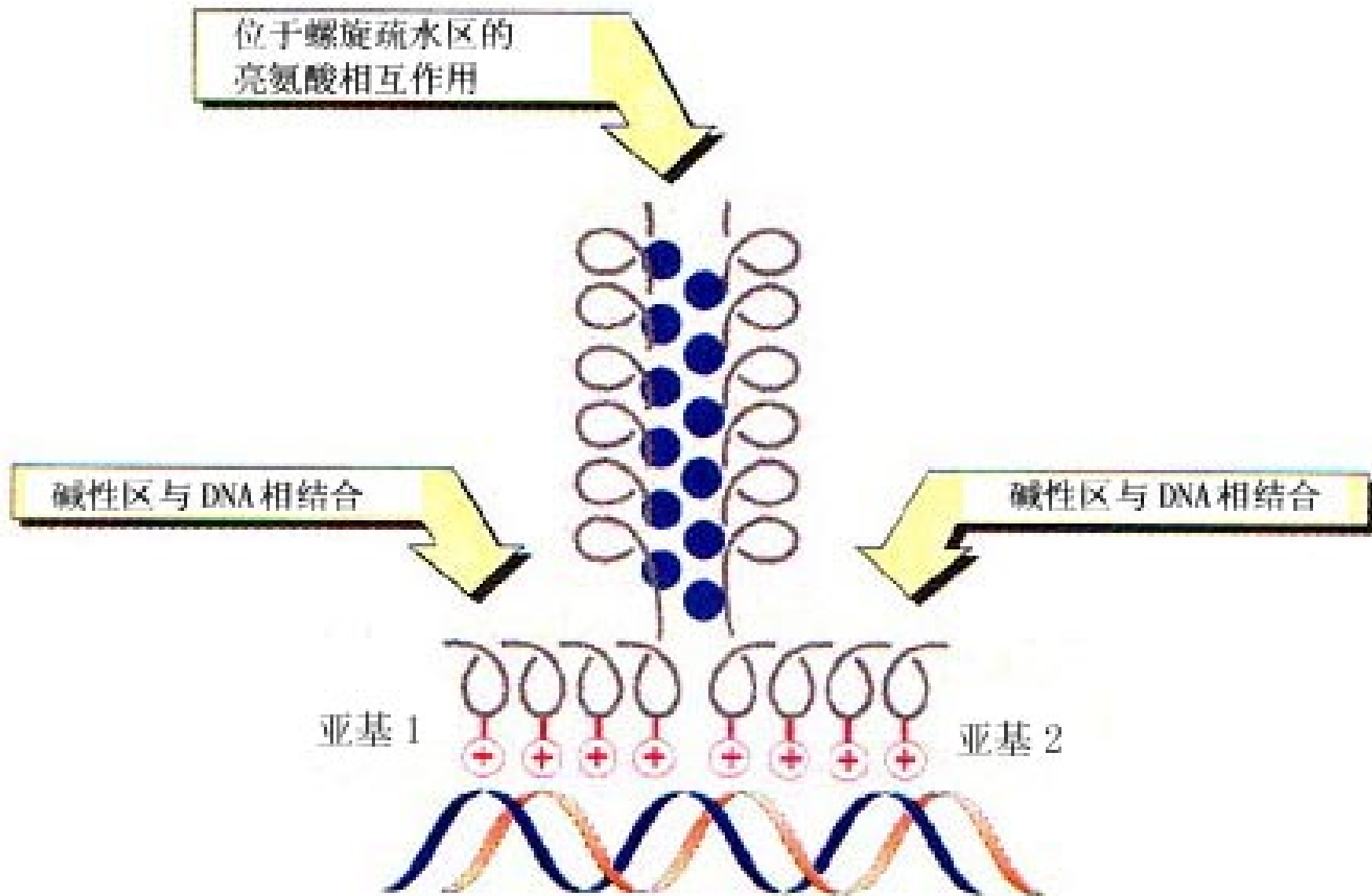


图7-21 碱性亮氨酸拉链转录激活蛋白与调控区DNA相结合的示意图



D、碱性-螺旋-环-螺旋（basic-helix/loop/helix，即bHLH结构）

在免疫球蛋白 κ 轻链基因的增强子结合蛋白 E12 与 E47 中，羧基端 100 ~ 200 个氨基酸残基可形成两个两性 α 螺旋，被非螺旋的环状结构所隔开，蛋白质的氨基端则是碱性区，其 DNA 结合特性与亮氨酸拉链类蛋白相似。



bHLH类蛋白只有形成同源或异源二聚体时，才具有足够的**DNA**结合能力。当这类异源二聚体中的一方不含有碱性区（如**Id**或**E12**蛋白）时，该二聚体明显缺乏对靶**DNA**的亲合力。



bHLH蛋白形成二聚体，能结合 DNA

无碱性区的 HLH 蛋白不能结合 DNA

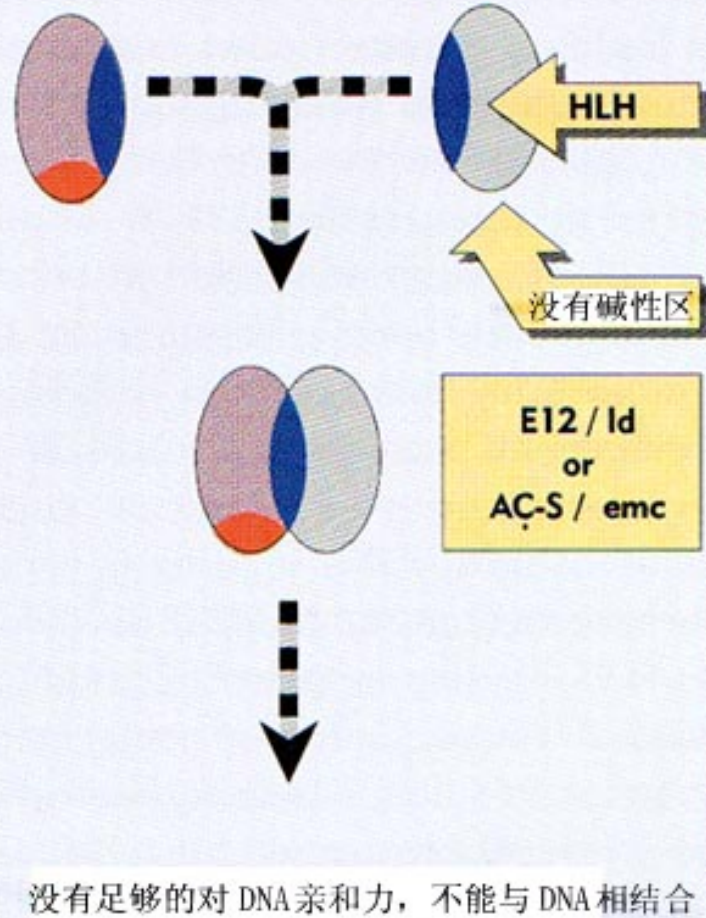
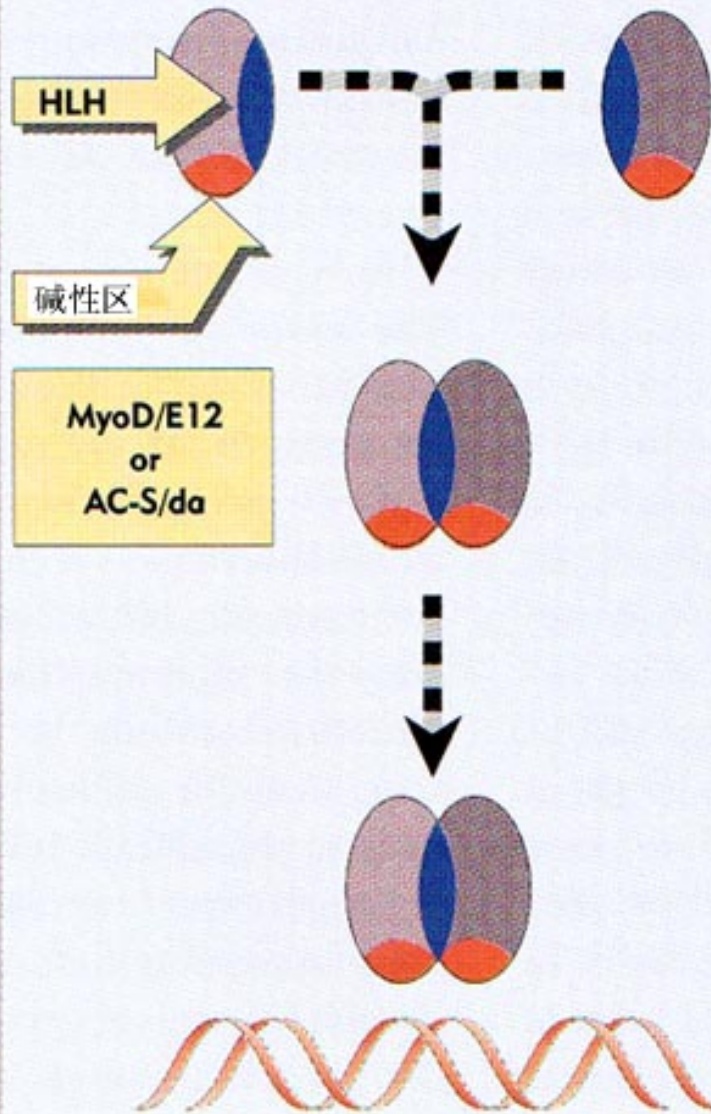


图7-22 bHLH蛋白的DNA结合能力分析



E、同源域 (homeo domains) 是指编码60个保守氨基酸序列的DNA片段，它广泛存在于真核生物基因组内，由于最早从果蝇homeotic loci (该遗传位点的基因产物决定了躯体发育) 中克隆得到而命名，同源转换基因与生物有机体的生长、发育和分化密切相关。



表7-8 含有同源转换区的转录因子及其DNA靶序列

蛋白因子	靶 DNA 序列
Oct-1、 Oct-2	ATTTGCAT
Pit-1/GHF-1	TATGCATAA
en、 ftz、 ubx	TCAATTAAAT
Eve	TCAATTAAAT TCAGCACCG
Bicoid	TCAATCCC



7.3.2 转录活化结构域

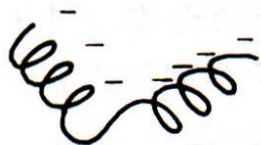
在真核生物中，反式作用因子的功能由于受蛋白质-蛋白质之间相互作用的调节变得精密、复杂，完整的转录调控功能通常以复合物的方式来完成，因此，是否具有转录活化域就成为反式作用因子中唯一的结构基础。



不同的转录活化域有下列特征性结构：

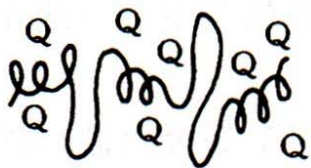
1、带负电荷的螺旋结构。哺乳动物细胞中糖皮质激素受体的两个转录活化域，**AP1**家族的**Jun**及**GAL4**都有酸性的螺旋结构，它们可能与**TFIID**复合物中某个通用因子或**RNA**聚合酶**II**本身结合，具有稳定转录起始复合物的作用。

A. GAL4酸性结构域



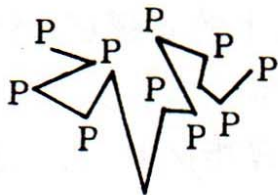
DSAAAHHDNSTIPLDFMPRDALH
GFDWSEEDDMSDGLPFLKTDPNNGF

B. SP1谷氨酰胺结构域



QGQTPQRVSGLQGSDALNIQQNQTSGGSLQAGQQKEGEQNQQQTQQQQQLQPQLVQ
GGQALQALQAAPLSGQTFTTQAISQETLQNLQAVPNSGPIIIRTPTVGPNGQVSW
QTLQLQNLQVQNPQAQTITLAPMQGVSLGQ

C. 富含脯氨酸结构域



PPHLNPQDPLKDLVSLACDPASQQPGRLNGSGQLKMPSHCLSAQMLPPPPGLPR
LPPATKPATTSEGGATSPSYSPDTSF

图7-24 几种常见的转录活化结构域



2、富含谷氨酰胺的结构。**SP1**是启动子**GC**盒的结合蛋白，共有**4**个参与转录活化的区域，其中最强的转录活化域含**25%**左右的谷氨酰胺。

Oct1/2、**Jun**、**AP2**、血清应答因子（**SRF**）等都有相同的富含谷氨酰胺的结构域。



3、富含脯氨酸的结构。CTF-NF1因子的羧基端富含脯氨酸（达20%~30%），很难形成 α 螺旋。在Oct2、Jun、AP2、SRF等哺乳动物因子中也有富含脯氨酸的结构域。



7.4 真核基因转录调控的主要模式

细胞是生命活动的基本单位。细胞通过**DNA**的复制和细胞分裂将本身所固有的遗传信息由亲代传至子代，实现增殖繁衍。它们还不断地“感知”环境变化，并对其作出特定的应答。



细胞应答可以分为3个阶段：

①外界信息的“感知”，即由细胞膜到细胞核内的信息传递；

②染色质水平上的基因活性调控；

③特定基因的表达，即从
DNA→**RNA**→蛋白质的遗传信息传递过程。



7.4.1 蛋白质磷酸化、信号转导及基因表达

蛋白质的磷酸化与去磷酸化过程是生物体内普遍存在的信息传导调节方式，几乎涉及所有的生理及病理过程，如糖代谢、光合作用、细胞的生长发育、神经递质的合成与释放甚至癌变等等。

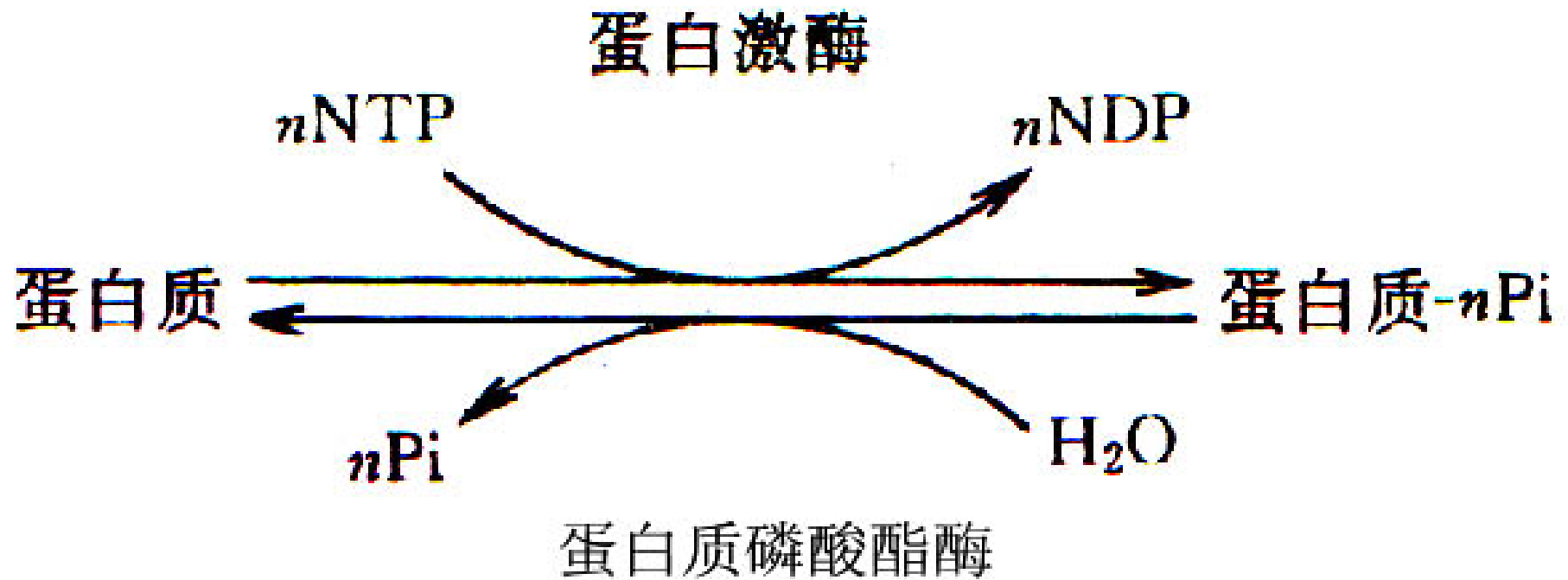


图7-25 蛋白质的磷酸化及去磷酸化



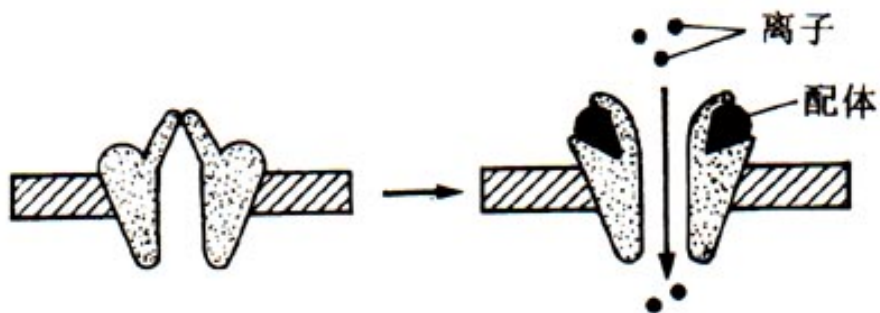
细胞表面受体与配体分子的高亲和力特异性结合，能诱导受体蛋白构象变化，使胞外信号顺利通过质膜进入细胞内，或使受体发生寡聚化而被激活。



表7-9 真核细胞主要跨膜信号传导途径

体系	作用方式
蛋白激酶受体	配体-受体结合激活受体细胞质内蛋白酪氨酸激酶或丝氨酸/苏氨酸激酶
受体耦联的G蛋白	配体结合后激活特定的G蛋白，活化其 α 亚单位，并调节特定靶分子，如腺苷酸环化酶、磷脂酶、cGMP磷酸二酯酶和离子通道的活性
离子通道	配体结合特定可关闭或开启的离子通道
心钠素（ANF）受体	配体结合活化鸟苷酸环化酶
细胞因子受体	细胞因子与受体结合，抑制或激活核基因表达
类固醇激素基因家族的核受体	与配体结合诱导或抑制靶基因表达

A. 受配体结合调控的离子通道



B. 与 G 蛋白相耦联的受体系统



C. 跨膜蛋白激酶-受体系统

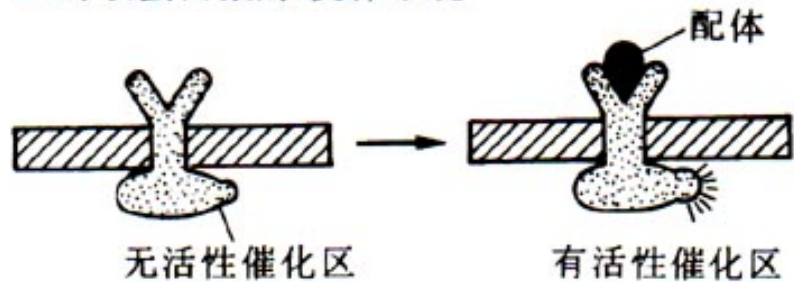


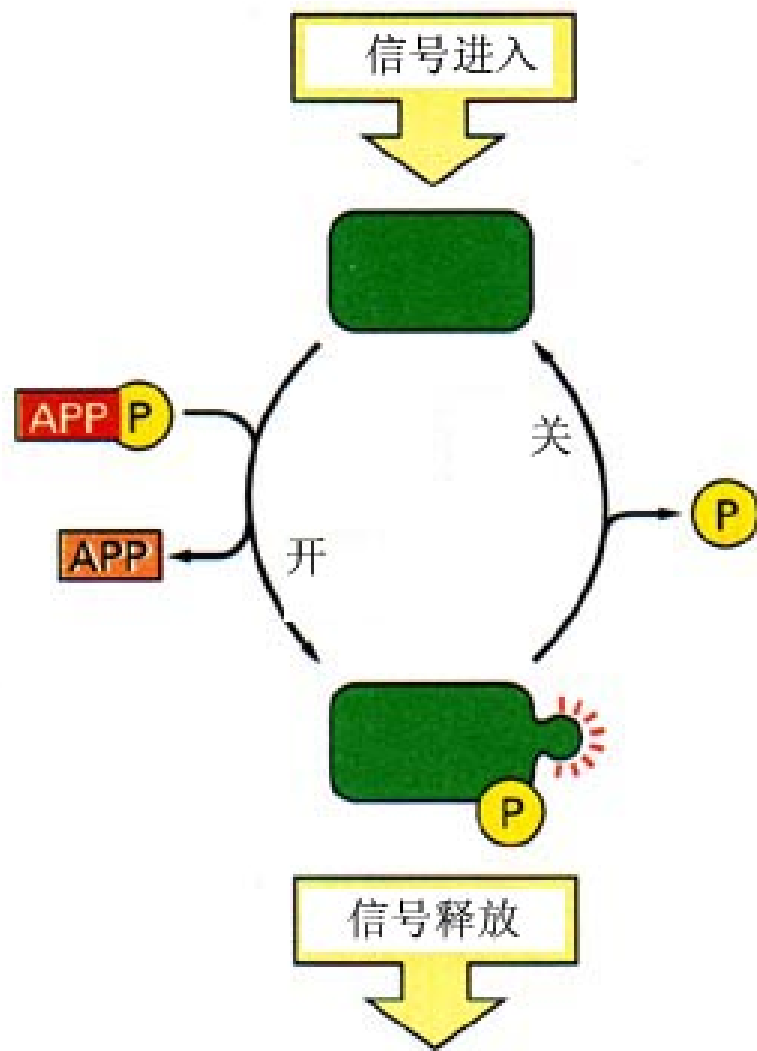
图7-27 细胞表面的三类受体示意图。



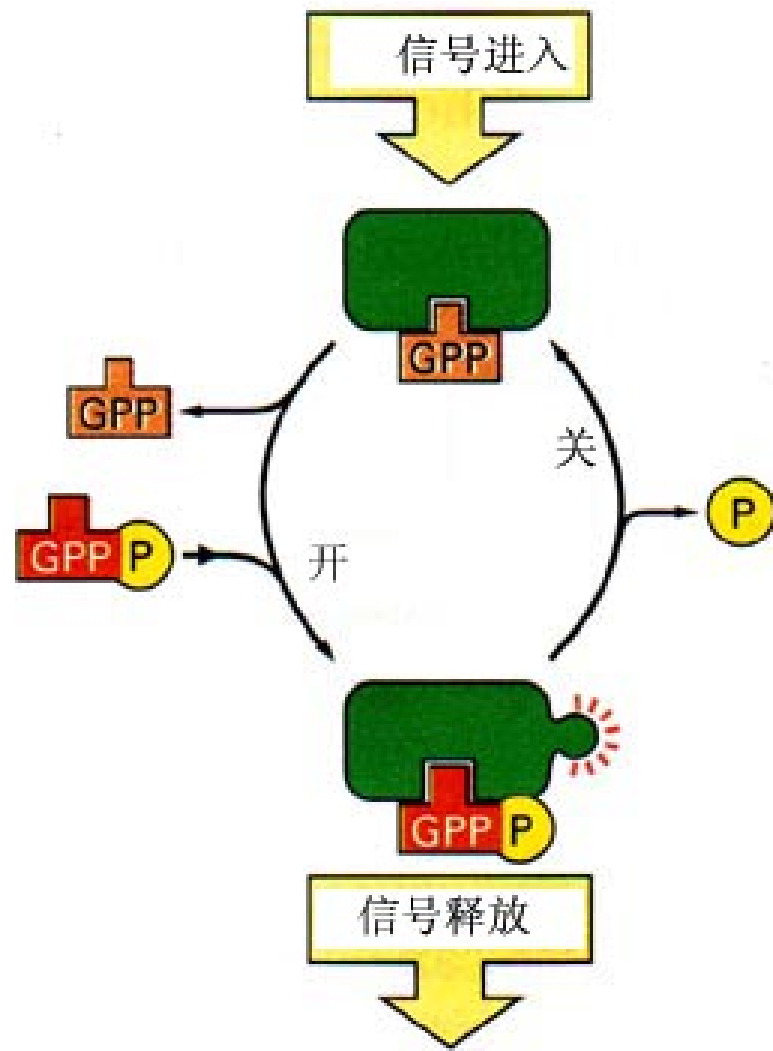
受体分子活化细胞功能的途径主要有两条：

一是受体本身或受体结合蛋白具有内源酪氨酸激酶活性，胞内信号通过酪氨酸激酶途径得到传递；

二是配体与细胞表面受体结合，通过G蛋白介异的效应系统产生介质，活化丝氨酸/苏氨酸或酪氨酸激酶，从而传递信号。



(A) 蛋白质磷酸化介导的信号转导



(B) GTP 结合蛋白介导的信号转导

图7-28 蛋白质磷酸化和GTP结合蛋白参与的信号转导过程

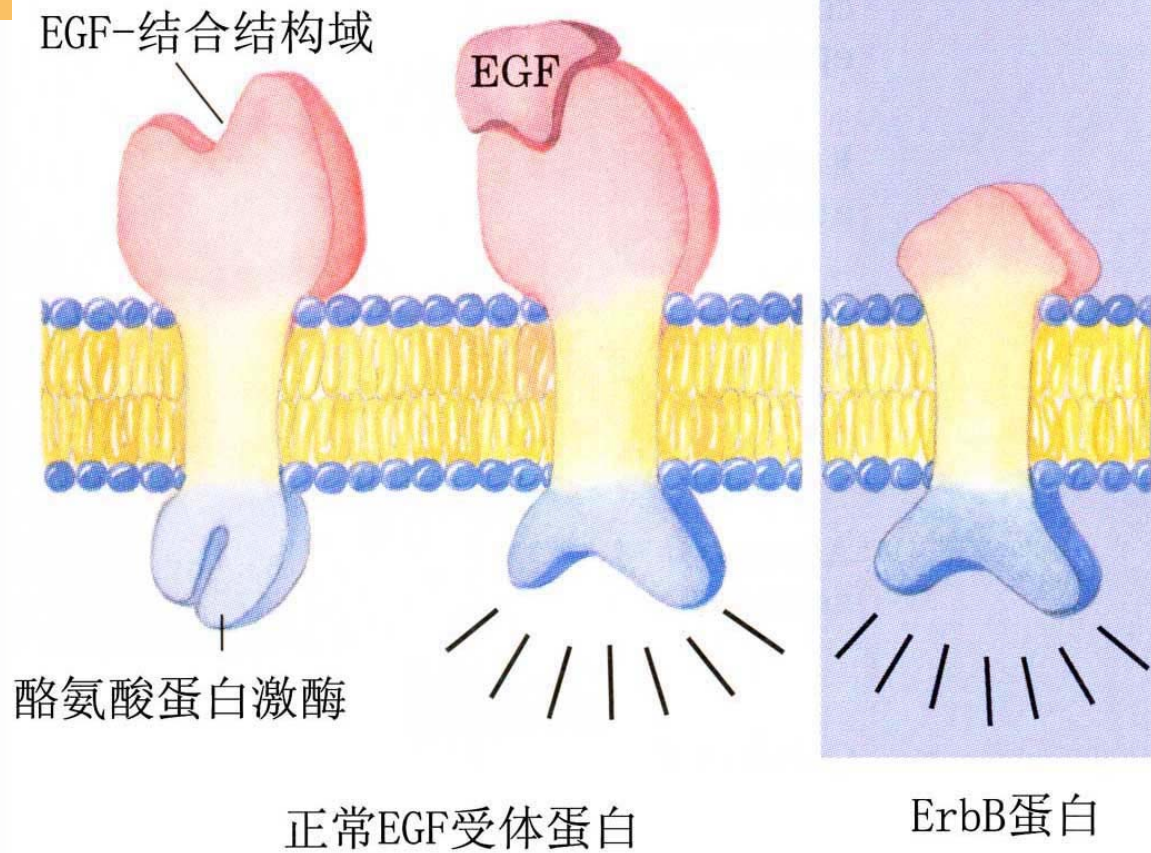


图7-26 酪氨酸受体蛋白激酶与表皮生长因子(EGF)相结合，刺激该受体蛋白的激酶活性，引发生理反应。原癌蛋白ErbB虽然没有正常酪氨酸受体蛋白激酶的胞外结构域，其胞内结构域却具有永久性蛋白激酶活性，永久刺激细胞分裂，诱发癌变。



已发现蛋白激酶（**protein kinase, PK**）基因**2000**余个和**1000**个左右蛋白质去磷酸化酶基因。根据底物蛋白质被磷酸化的氨基酸残基可分为：

- ①丝氨酸/苏氨酸型；
- ②酪氨酸型；
- ③组氨酸型。



根据是否有调节物参与蛋白激酶活性可分为两大类：信使依赖型（又分胞内信使、调节因子依赖型和激素或生长因子依赖型，表7-10）和非信使依赖型（表7-11）。



表7-10 信使依赖型蛋白激酶

胞内信使或调节因子依赖型PK				激素或生长因子依赖型PK			
名称	调节物	动物	植物	名称	调节物	动物	植物
cAMP依赖型PK (PKA)	cAMP	+	?	表皮生长因子(EGF) 依赖型PK	EGF	+	-
cAMP依赖型PK(PKG)	cGMP	+	?	胰岛素依赖型PK (胰岛素样生长因子)	胰岛素	+	-
Ca²⁺、CaM 依赖型PK	Ca²⁺、CaM	+	+	血小板衍生生长因子(PDGF) 依赖型PK	PDGF	+	-
Ca²⁺、磷脂依赖型PK(PKC)	Ca²⁺、磷脂	+	+	类胰岛素生长因子-1 (IGF-1) 依赖型PK	IGF-1	+	-
只依赖于Ca²⁺的PK	Ca²⁺	+	+	集落刺激因子-1 (CSF-1) 依赖型PK	CSF-1	+	-



表7-11 非信使依赖型蛋白激酶

名 称	动物	植物
酪蛋白激酶 (I, II 型)	+	+
糖原合成酶激酶	+	?
组蛋白激酶II	+	+
视紫红质蛋白激酶	+	-
β -肾上腺素受体蛋白激酶	+	-



1、受cAMP水平调控的A激酶

依赖于cAMP的蛋白激酶称为A激酶（PKA），它能把ATP分子上的末端磷酸基团加到某个特定蛋白质的丝氨酸或苏氨酸残基上。

被A激酶磷酸化的氨基酸N端上游往往存在两个或两个以上碱性氨基酸，特定氨基酸的磷酸化（X-Arg-Arg-X-Ser-X）改变了这一蛋白的酶活性。



在不同的细胞中，**A**激酶的反应底物不一样，所以，**cAMP**能在不同靶细胞中诱发不同的反应。

非活性状态的**PKA**全酶由4个亚基**R2C2**所组成，分子量约为**150-170**，调节亚基与**cAMP**相结合，引起构象变化并释放催化亚基，后者随即成为有催化活性的单体（图7-30）。

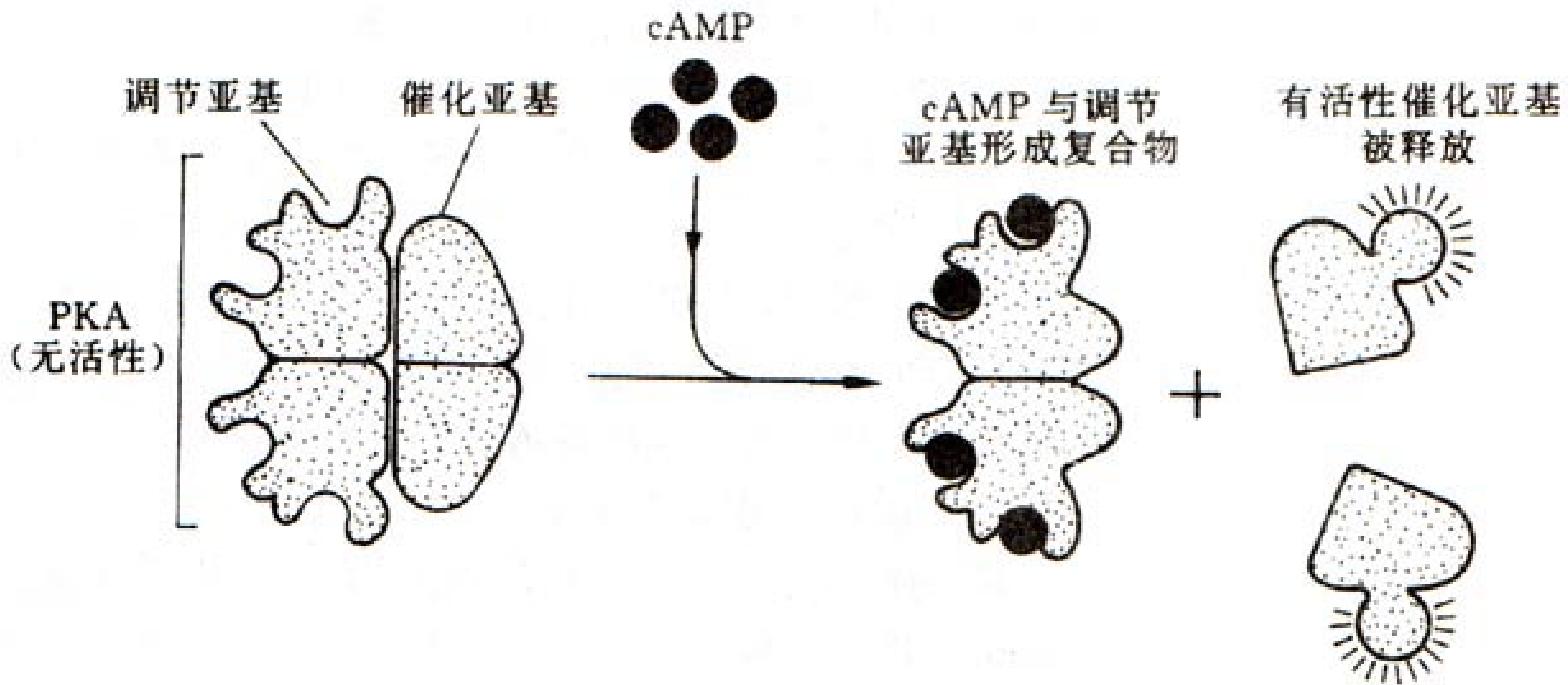


图 7-31 cAMP 调控的 A 激酶活性



糖原代谢时，激素与其受体在肌肉细胞外表面相结合，诱发细胞质cAMP的合成并活化A激酶，后者再将活化磷酸基团传递给无活性的磷酸化酶激酶，活化糖原磷酸化酶，最终将糖原磷酸化，进入糖酵解过程并提供ATP（图7-30）。

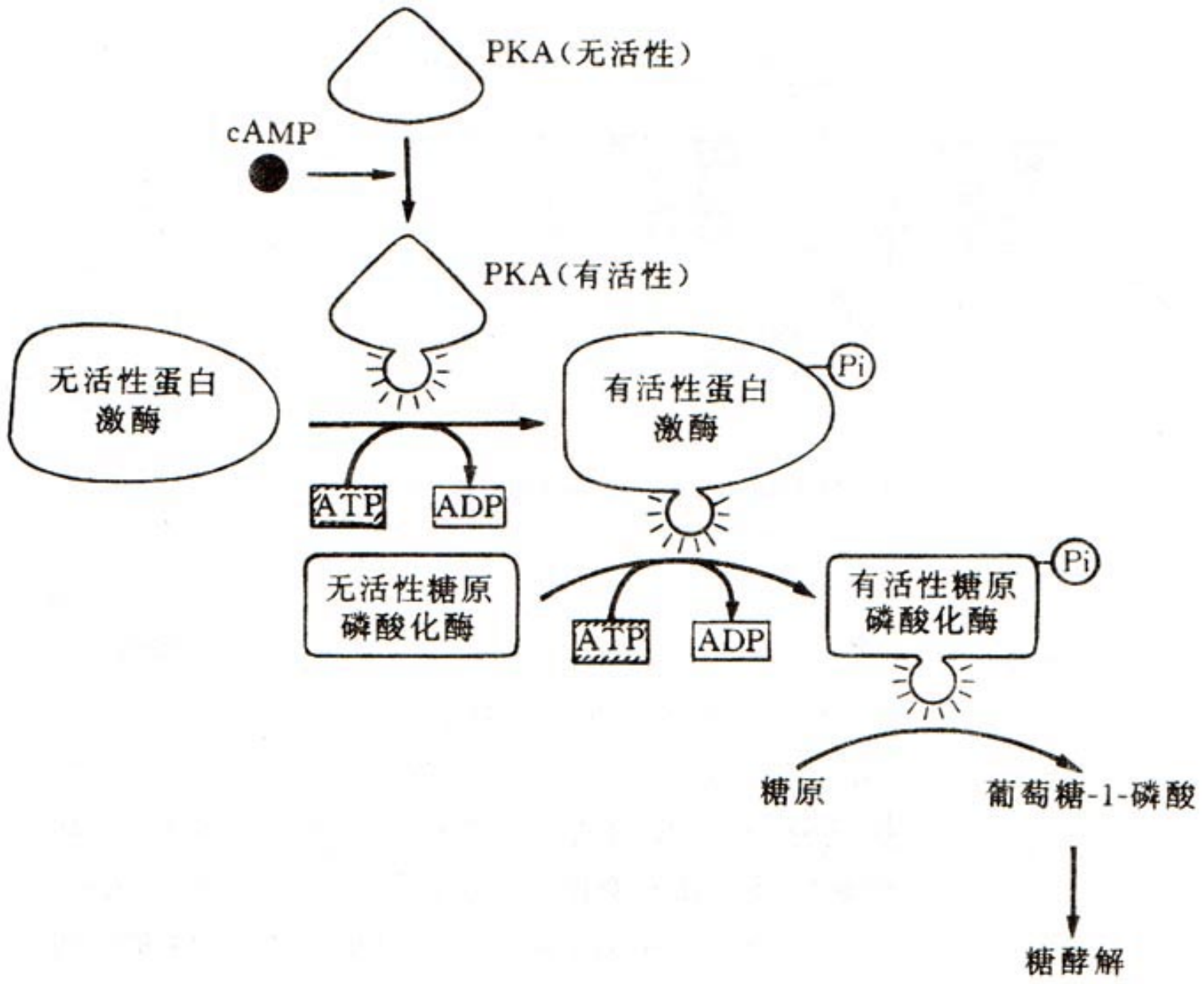


图 7-32 cAMP 活化糖原磷酸化酶示意图



2、C激酶

该蛋白激酶活性是依赖于 Ca^{2+} 的，所以称为C激酶（PKC）。

IP3引起细胞质 Ca^{2+} 浓度升高，导致C激酶从胞质转运到靠近原生质膜内侧处，并被DAG和 Ca^{2+} 的双重影响所激活。DAG提高了C激酶对于 Ca^{2+} 的亲合力。



C激酶是一个 7.7×10^4 的蛋白质，主要实施对丝氨酸、苏氨酸的磷酸化，它具有一个催化结构域和一个调节结构域。

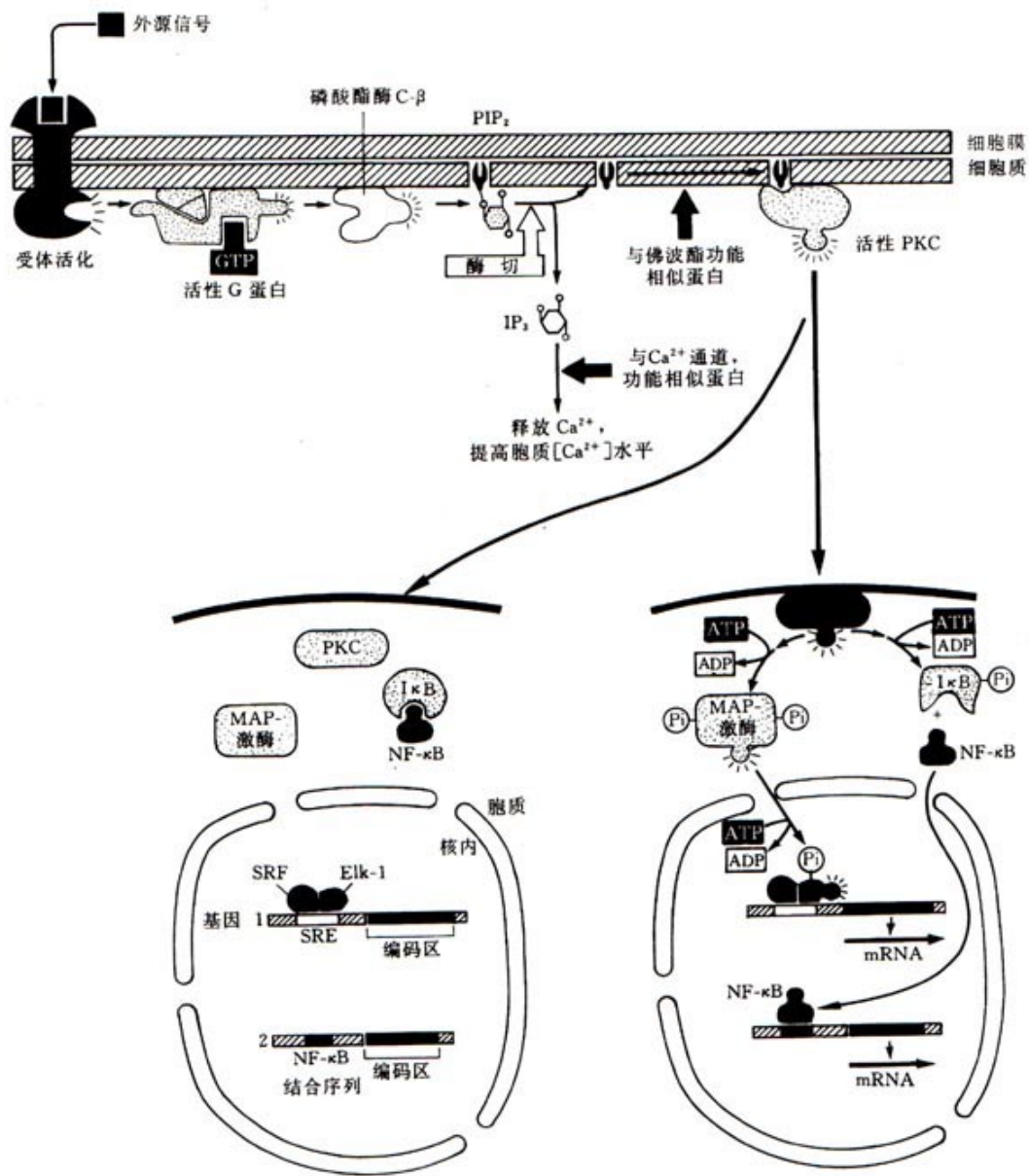


图 7-34 C 激酶信号传递及基因表达示意图



NF- κ B是广泛存在于各种细胞中的转录调控因子, 调节与细胞的应急反应, 如免疫应答、炎症反应及细胞抗凋亡作用相关的基因转录。

NF- κ B是一个二聚体, 由NF- κ B / Rel家族成员组成, 与靶基因增强子的特定序列 (5'-GGGACTTTC-3') 相结合, 所有成员都包含一个300氨基酸的Rel同源区 (RHR, Rel homology region)。p65/p50; p50/p50, p52/p52



通常，NF- κ B与抑制因子 I κ B相结合，以非活性形式存在于细胞中。当细胞受到肿瘤坏死因子、白介素-1及细菌脂多糖等上游调节因子作用时，I κ B蛋白被迅速磷酸化并被降解，释放NF- κ B并向核内迁移，激活靶基因转录。

I κ B主要包括I κ B α ， I κ B β ， I κ B γ ， I κ B ϵ 以及Bc1-3， NF- κ B1的前体p105， NF- κ B2的前体p100。



每个I κ B家族成员分子中均含有6或7个ankyrin重复序列，I κ B分子凭借这些重叠的螺旋结构与NF- κ B分子中的RHR区域结合，屏蔽NF- κ B的核定位序列，使之滞留于胞质中。

新合成的I κ B α 进入核内与NF- κ B结合，使其从DNA上解离，I κ B α 分子中的核输出序列NES (nuclear export signal) 再促使NF- κ B重新回到细胞质中。

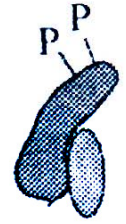


LT- β
BLys

TNF α
IL-1
LPS



IKK α IKK β



Re1B:p100



p50:p65:1kB

泛素化及蛋白酶体
介导的降解



Re1B:p52



p50:p65

靶基因

固有免疫

炎症

细胞存活



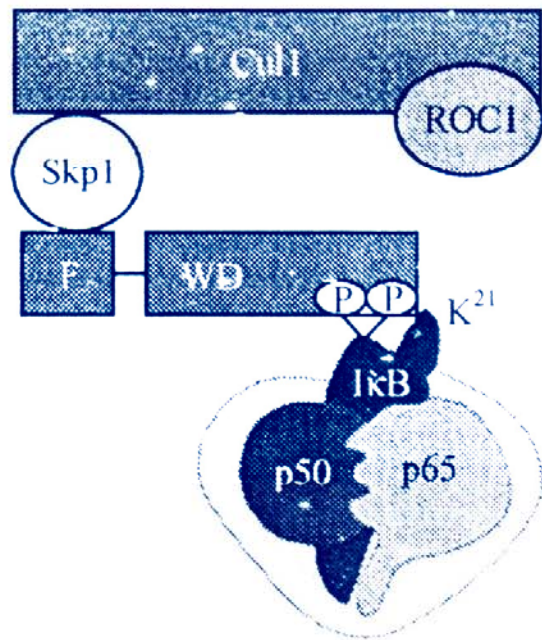
I κ B α 在激酶复合物IKK的作用下被磷酸化，磷酸化发生在I κ B α 分子中Ser32和Ser36两个位点。

IKK复合物由活性亚基IKK α ，IKK β 和调节亚基IKK γ 组成。

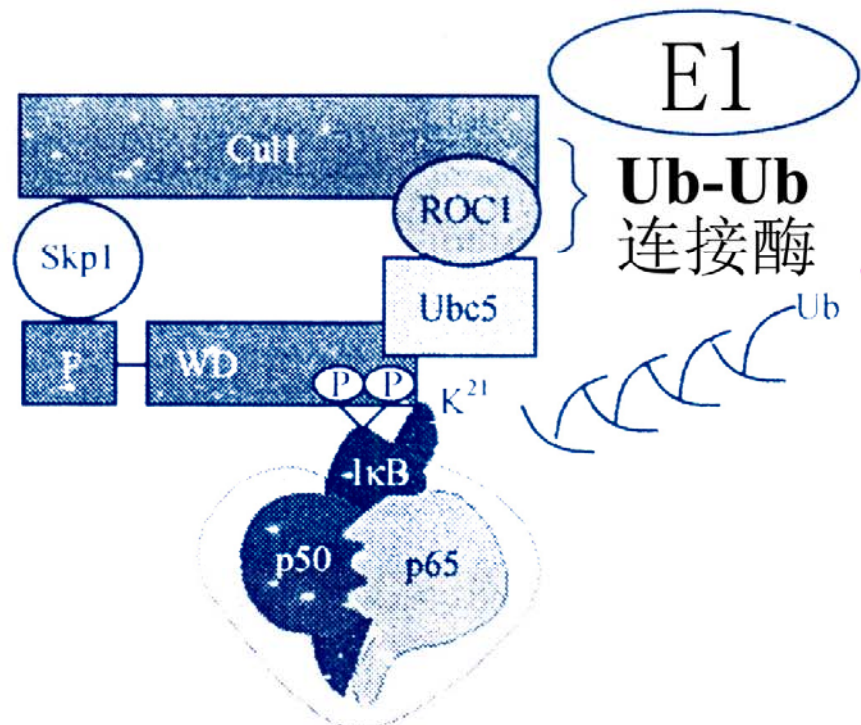
磷酸化后的I κ B α 被泛素连接酶复合物SCF (Skp-1 Cu11 F-box) 识别，I κ B α 分子中的Lys21及Lys22被泛素化，被26S蛋白酶体识别并迅速降解。



E3RS 1kB



1kB 分子的结合



泛素化



3、CaM激酶及MAP激酶

Ca²⁺的细胞学功能主要通过钙调蛋白激酶（**CaM-kinase**）来实现的，它们也是一类丝氨酸/苏氨酸激酶，但仅应答于细胞内**Ca²⁺**水平。



MAP 激 酶（ **mitogen-activated proteinkinase**， **MAP-kinase**， 又 称 为 **extracellular-signal-regulated kinase**， **ERKS**） 活 性 受 许 多 外 源 细 胞 生 长、 分 化 因 子 的 诱 导。 **MAP-激 酶** 的 活 性 取 决 于 该 蛋 白 中 仅 有 一 个 氨 基 酸 之 隔 的 酪 氨 酸、 丝 氨 酸 残 基 是 否 都 被 磷 酸 化。



能同时催化这两个氨基酸残基磷酸化的酶被称为**MAP-激酶-激酶**，它的反应底物是**MAP激酶**。**MAP-激酶-激酶**本身能被**MAP-激酶-激酶-激酶**所磷酸化激活，后者能同时被**C激酶**或酪氨酸激酶家族的**Ras**蛋白等激活，从而在信息传导中发挥功能。

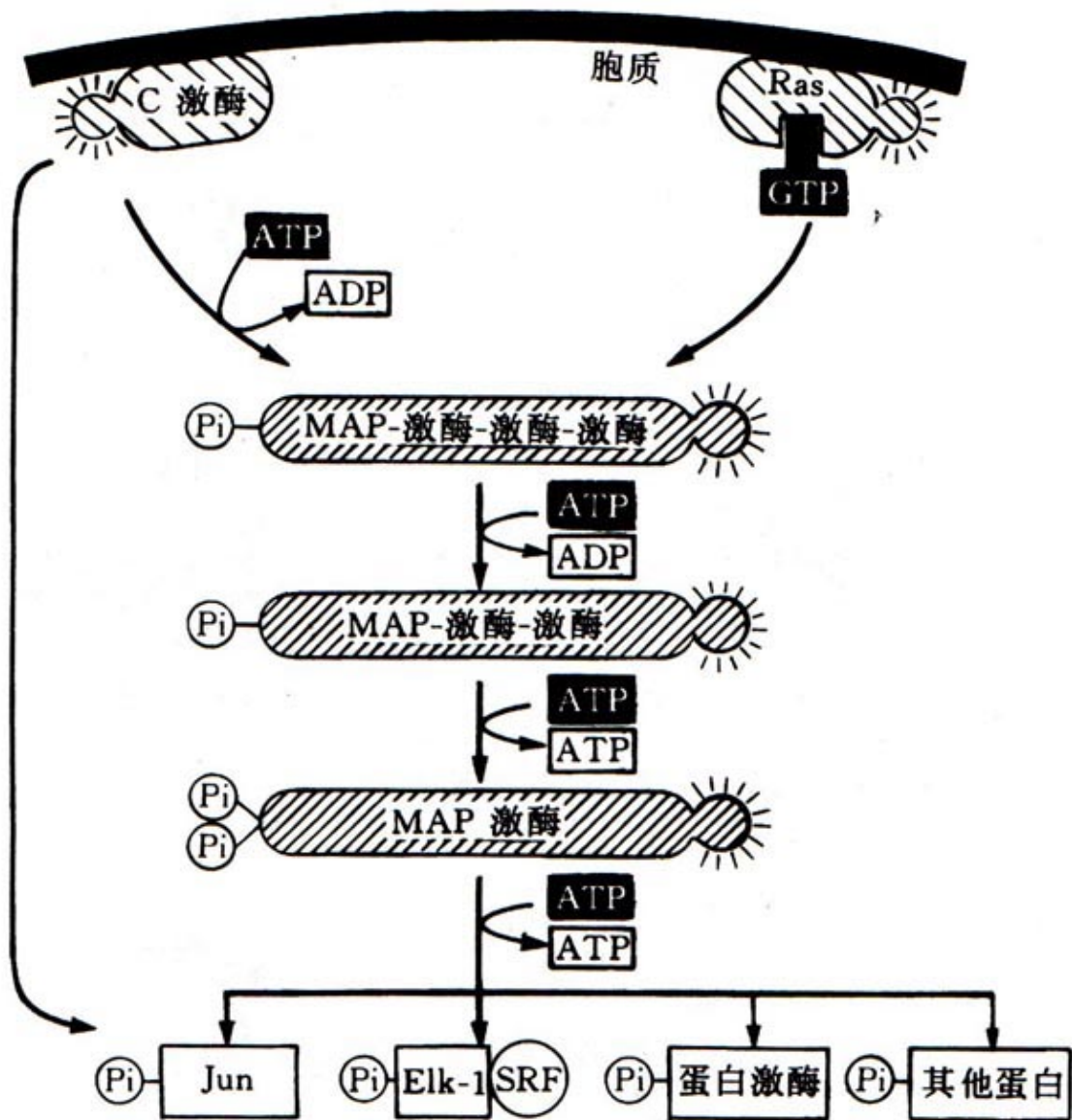


图7-35 Ras蛋白和C激酶激活丝氨酸/苏氨酸“瀑布式”磷酸化(级联磷酸化),引起相关生理反应图示



4、酪氨酸蛋白激酶（PTK）途径

包括跨膜受体家族与胞质非受体家族两大类。

受体类由胞外结合配体结构域、跨膜结构域和细胞质激酶结构域组成，其PTK活性受胞外结构域与配体的调节。配体与受体结合可诱导受体蛋白的二聚化，将受体胞质区酪氨酸残基磷酸化。



因为**Src**亚家族非受体酪氨酸激酶以及许多**PTK**作用底物分子都存在被称为**Src**同源结构域（**Src homology domain, SH**）。



在正常细胞中，**pp60^{c-Src}**上的**Tyr527**被磷酸化并以头尾相连的缔合方式折入同一分子的**SH2**结构域中，从而抑制了**PTK**活性。**pp60^{c-Src}**与靶分子**PI₃激酶**，**rasGAP**等的缔合作用均涉及**pp60^{c-Src}SH2**结构域与**PI₃激酶**及**rasGAP**分子上酪氨酸磷酸化区的结合。

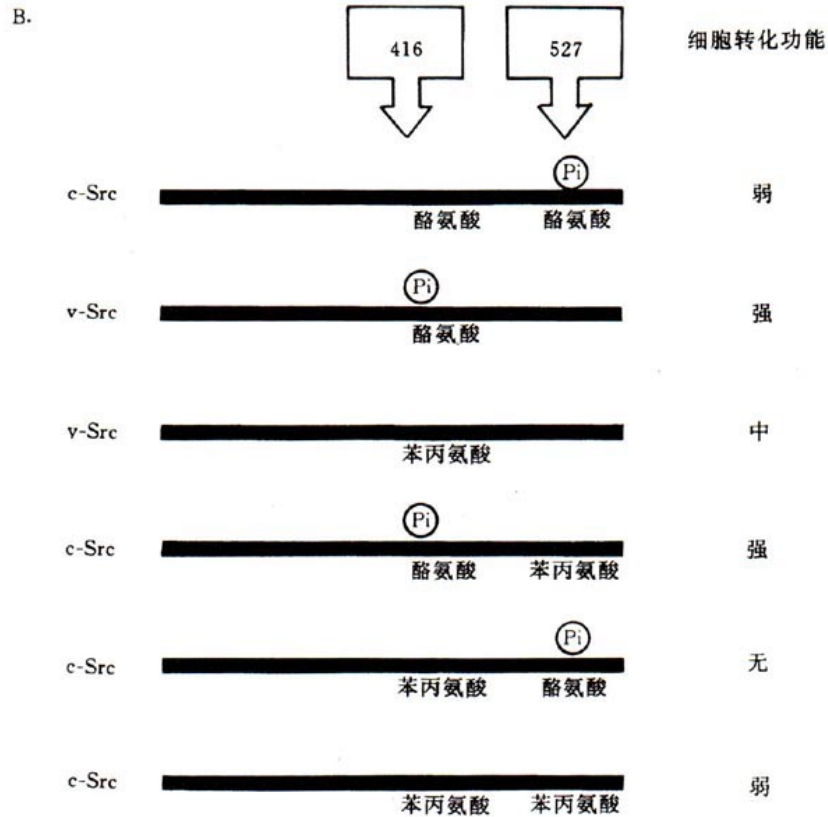
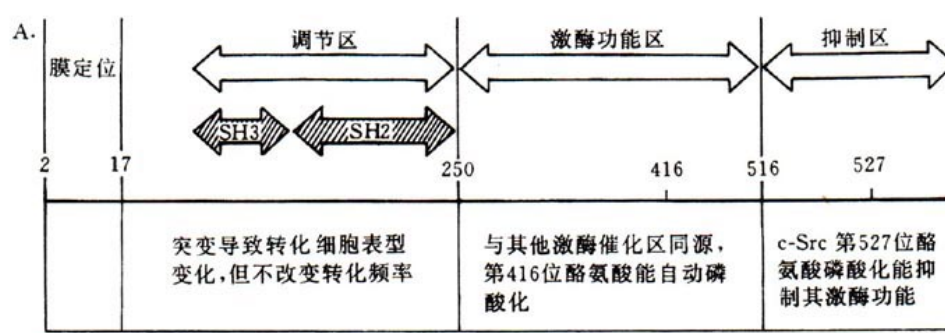
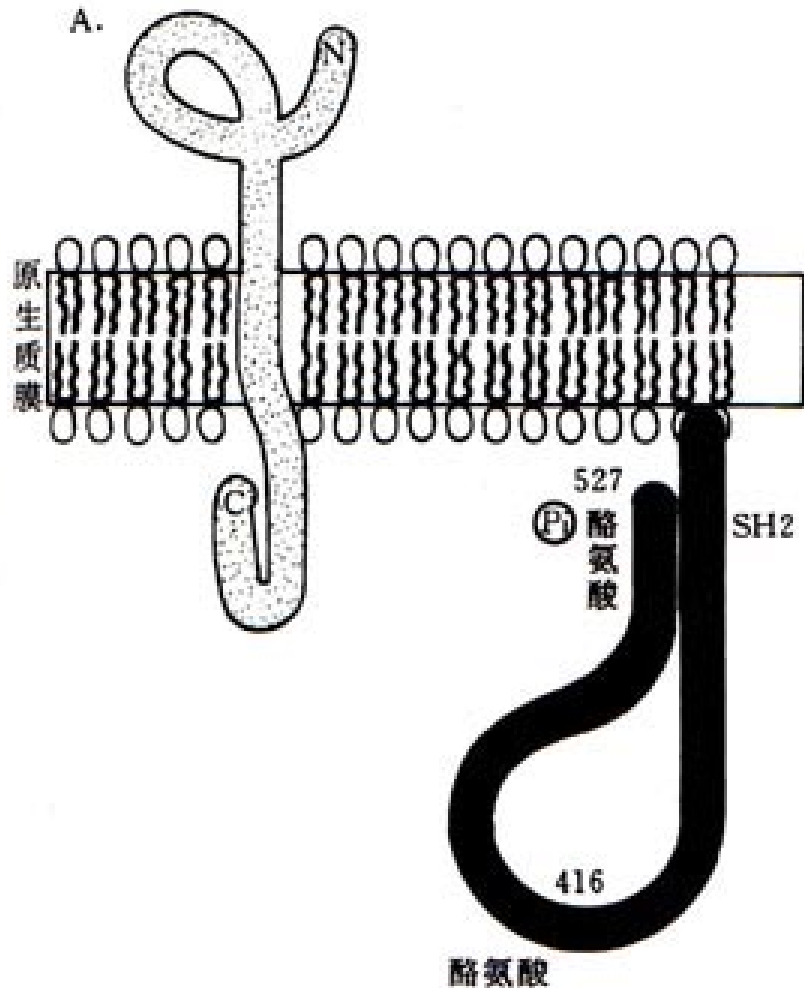
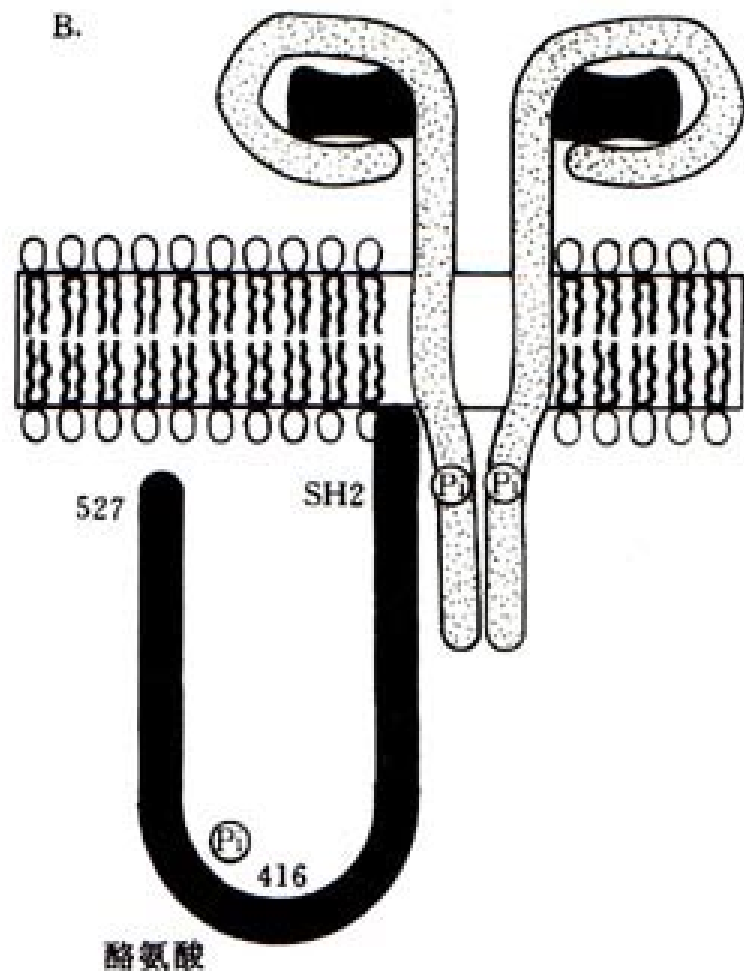


图 7-36 c-Src 蛋白功能区划分 (A) 及第 416 位、527 位酪氨酸残基磷酸化或点突变以后对该蛋白质功能的影响 (B)



Src 激酶无活性



Src 激酶被活化

图 7-37 Src 激酶活化途径 A. 第 527 位酪氨酸被磷酸化后能与 SH2 区相结合，阻碍了受体分子与之结合；B. 第 527 位酪氨酸去磷酸化后，SH2 区与膜受体结合，第 416 位酪氨酸被磷酸化，Src 获得酶活性



Cell (2003) 114:663-671.

Cell (2003) 113:127-137.

科学通报 (2003) 48:1893-1911.

Ubiquitination:

**Sumoylation: Small Ubiquitin-related
Modifier (SUMO)**

Acetylation: Mainly to Lysine residues.

氧化还原电位与共价修饰对基因表达的调控.

Nature (2003) 423:769-773; 773-777.

Cell (2003) 112:11-17.

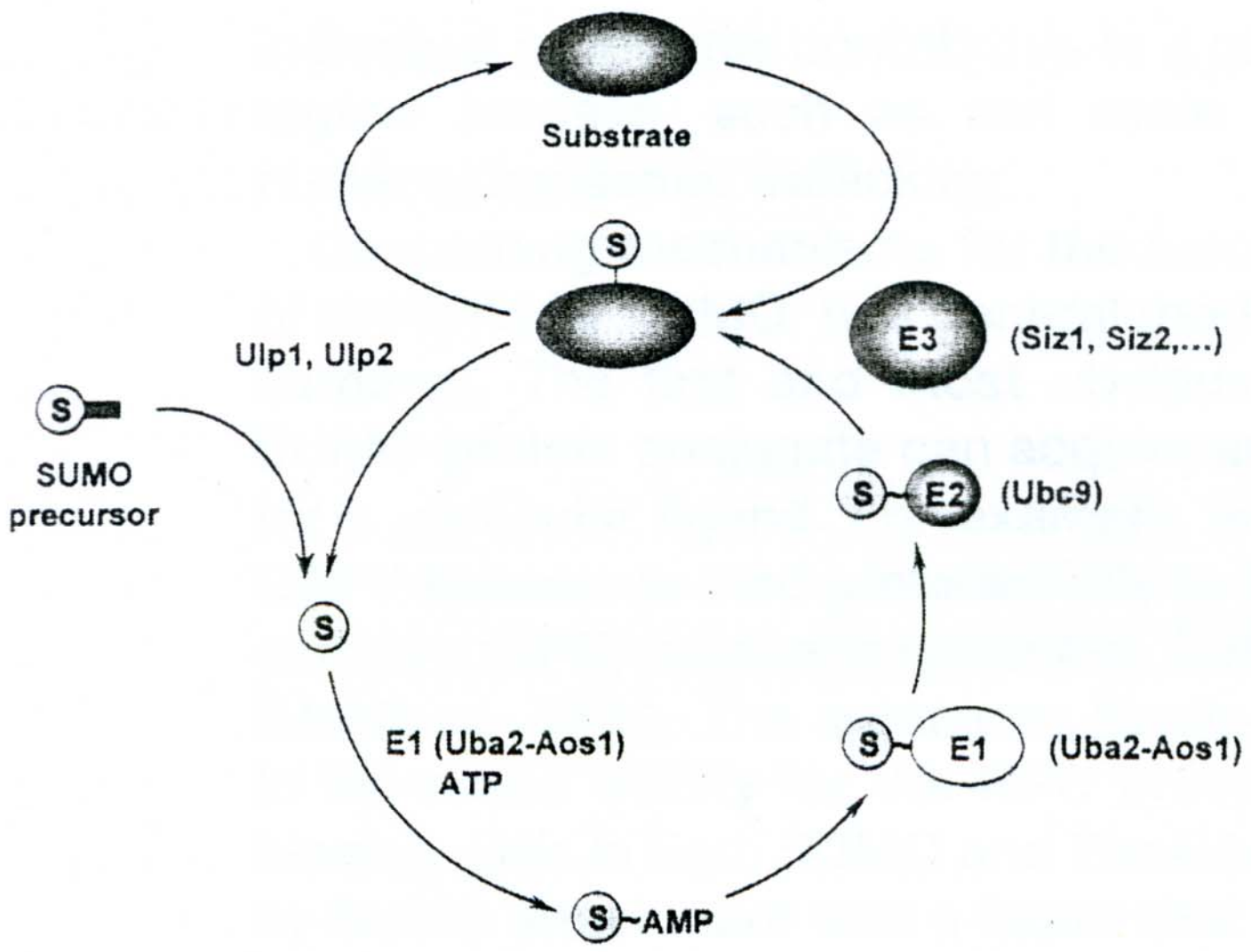


Figure 1. The SUMO Cycle

Enzyme names are from *S. cerevisiae*, where most of the components were first described.

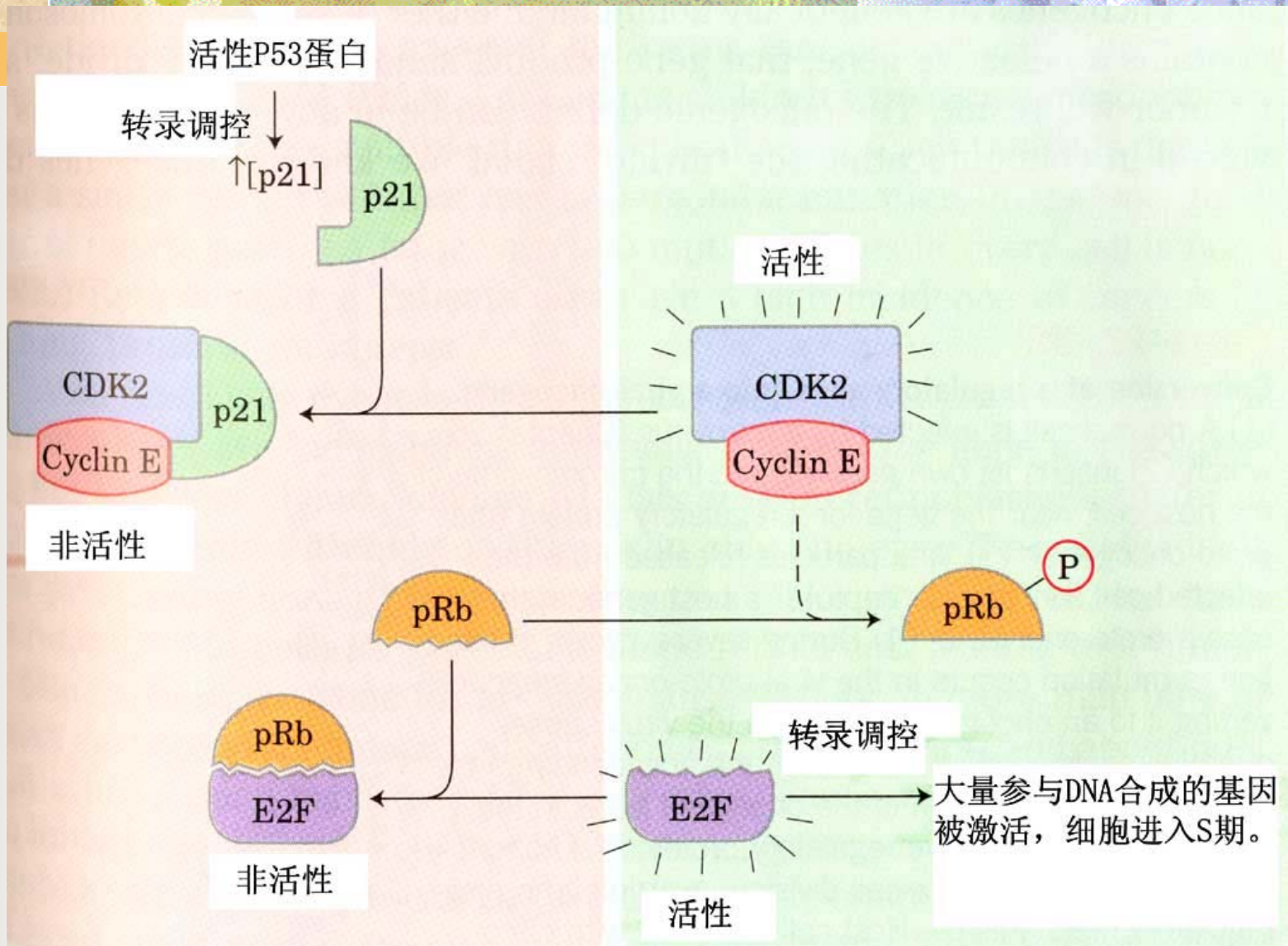


图7-35 pRb蛋白的磷酸化和去磷酸化过程控制了细胞分裂的周期性。



细胞通过 P^{53} 及 P^{21} 蛋白控制 **CDK**
(**cyclin-dependent protein kinase**)
活性，调控细胞分裂的进程。



P²¹ 蛋白过量时，大量周期蛋白（cyclin）E-CDK2复合物与P²¹蛋白相结合，使CDK2丧失磷酸化P^{Rb}蛋白的功能。

没有被磷酸化的P^{Rb}蛋白与转录因子E2F相结合并使后者不能激活与DNA合成有关的酶，导致细胞不能由G1期进入S期，细胞分裂受阻。



如果细胞中**P⁵³**基因活性降低，**P²¹**蛋白含量急剧下降，周期蛋白**E-CDK2**复合物就能有效地将**P^{Rb}**蛋白磷酸化。

此时，**P^{Rb}**蛋白不能与**E2F**相结合，后者发挥转录调控因子的作用，激活许多与**DNA**合成有关的基因表达，细胞从**G1**期进入**S**期，开始分裂。



CDK活性受双重调控

- (1) 没有周期蛋白，**CDK**无活性。
- (2) 随着周期蛋白的合成和积累，逐步形成周期蛋白-**CDK**复合物。
- (3) **CDK**上的酪氨酸位点被磷酸化，掩盖了其**ATP**结合位点，**ATP**不能有效地与之相结合，**CDK**仍然无活性。



(4) **CDK**蛋白T-环中的苏氨酸位点被磷酸化，并将其酪氨酸位点上的磷酸基团去掉，**CDK**蛋白才能发挥其生物学活性。

(5) 同时，**CDK**蛋白还能使细胞中的磷酸酯酶磷酸化，以加速脱去自身酪氨酸位点上的磷酸基团。



(6) 有生物活性的周期蛋白-**CDK**复合物能将**DBRP (Destruction Box Recognizing Protein)** 磷酸化, 激活泛素连接酶, 把大量泛素加到周期蛋白上并使之迅速降解, **CDK**失活, 新的周期开始。



7.4.2 激素及其影响

许多类固醇激素（如雌激素、孕激素、醛固酮、糖皮质激素和雄激素）以及一般代谢性激素（如胰岛素）的调控作用都是通过起始基因转录而实现的。



类皮质激素（肾上腺类固醇）

醇）

糖皮质激素能提高血糖含量，并具有抗炎能力。

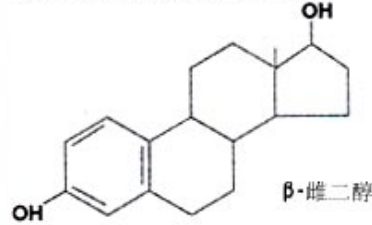


无机皮质激素参与维持细胞内无机盐和水分的平衡

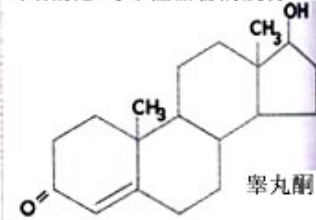


固醇类性激素

雌激素参与雌性特征的发育

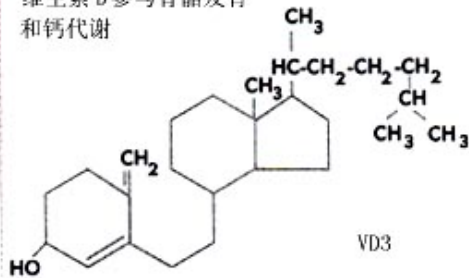


雄激素参与雄性器官的发育

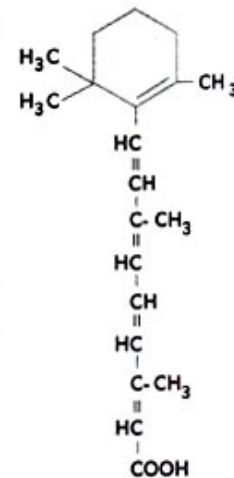


发育与形态建成

维生素D参与骨骼发育和钙代谢

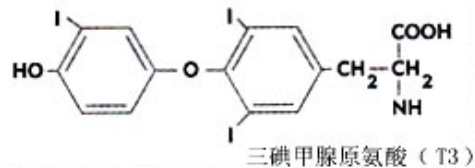


视黄酸参与形态建成



甲状腺激素

甲状腺激素调控了体内的基础代谢水平



反式视黄酸

图7-37 几种常见的疏水性小分子激素的结构式。



体内存在的许多糖皮质类激素应答基因都有一段大约**20bp**的顺式作用元件（激素应答元件，简称**HRE**），该序列具有类似增强子的作用，其活性受激素制约。靶细胞中含有大量激素受体蛋白，而非靶细胞中没有或很少有这类受体。



表7-12 三种激素应答序列成分分析

激 素	元 件	序 列
糖皮质激素	GRE	TGGTACANNNTGT TCT
雌激素	ERE	GGTCANNNTGTCC
甲状腺素	TRE	CAGGGACGTGAC CGCA



固醇类激素的受体蛋白分子有相同的结构框架，包括保守性极高并位于分子中央的**DNA**结合区，位于**C**端的激素结合区和保守性较低的**N**端。如果糖皮质激素受体蛋白激素结合区的某个部分丢失，就变成一种永久型的活性分子，（图7-38）。



多肽激素胰高血糖素接触靶细胞时，首先与受体结合，并激活膜上的腺苷酸环化酶，使之以 Mg^{2+} -ATP为底物生成环腺苷酸和焦磷酸。细胞内cAMP浓度升高，导致蛋白激酶活性增强，特定酶系的磷酸化水平及酶活性都得到改善，促进糖原最终分解为葡萄糖-1-磷酸（图7-39）。

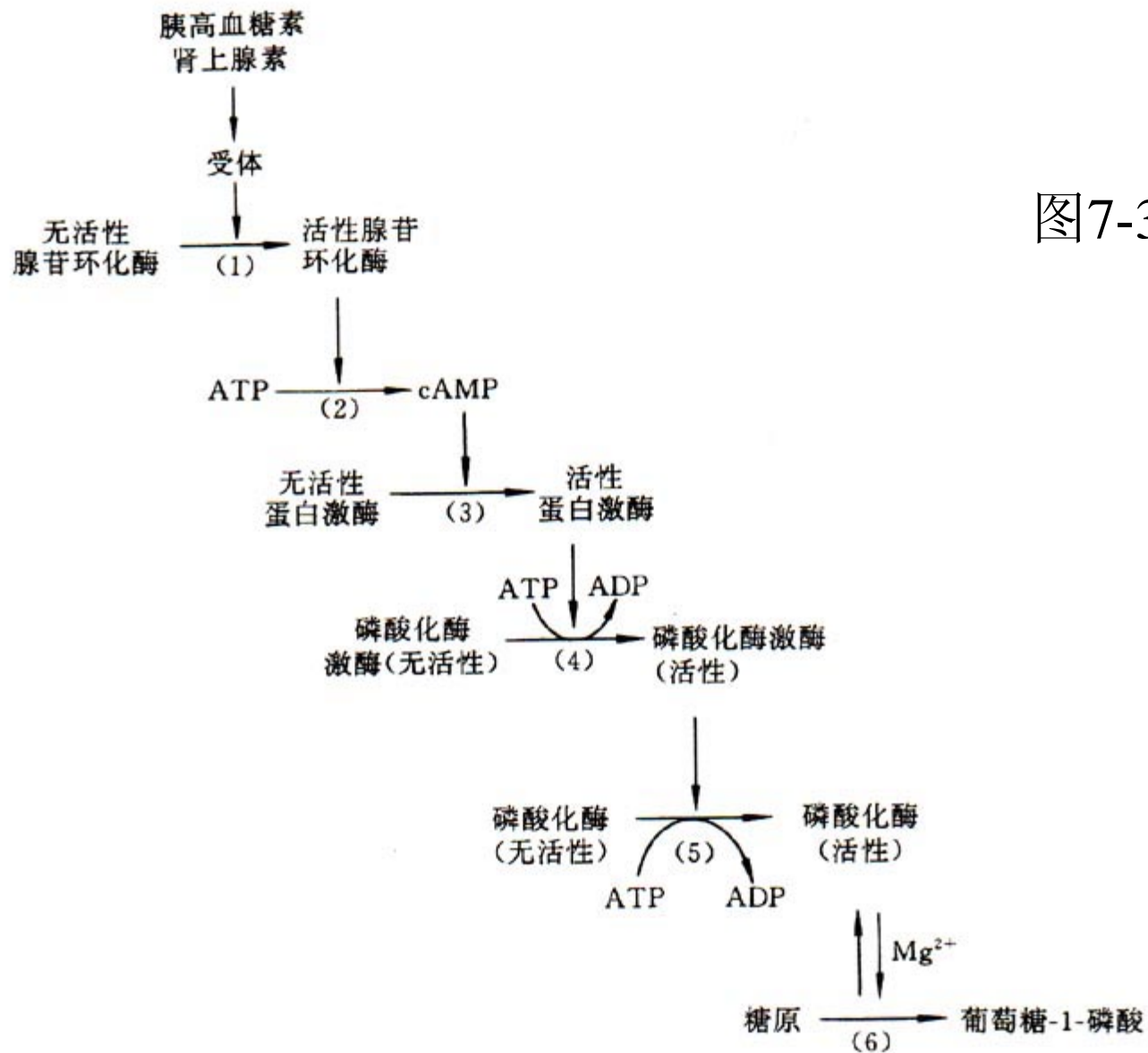


图7-39

图7-40 胰高血糖素激活糖原分解连锁反应



7.4.3 热激蛋白诱导的基因表达

能与某个（类）专一蛋白因子结合，从而控制基因特异表达的**DNA**上游序列称为应答元件（**response element**）。



主要有热激应答元件（**heat shock response element, HSE**）

糖皮质应答元件（**glucocorticoid response element, GRE**）

金属应答元件（**metal response element, MRE**）。

应答元件与细胞内专一的转录因子相互作用，协调相关基因的转录。



表7-13 特异转录因子与各类应答元件的相互作用

调控因子	应答元件	DNA序列	结合蛋白	相对分子质量
热激	HSE	CNNGAANNTC CNNG	HSF	9.3×10^4
镉	MRE	CGNCCCGGNC NC	?	?
佛波酯	TRE	TGACTCA	AP1	3.9×10^4
血清	SRE	CCATATTAGG	SRF	5.2×10^4



许多生物在最适温度范围以上，能受热诱导合成一系列热休克蛋白（**heat shock protein**）。受热后，果蝇细胞内**Hsp70 mRNA**水平提高**1000**倍，就是因为热激因子（**heat shock factor, HSF**）与**hsp70**基因**TATA**区上游**60bp**处的**HSE**相结合，诱发转录起始。



表7-14 主要HSP家族及其生理功能

家族	主要成员	主要生理功能	免疫应答中的可能作用
HSP90	HSP90, HSP83	促进甾醇激素受体与激素的结合及与DNA结合, 调节激酶磷酸化活性。	抗肿瘤, 提高自身免疫性能
HSP70	HSP70, BiP, Dnat, Hsc70, Grp78	参与蛋白质折叠和去折叠、蛋白转位及多聚复合物的组装。	免疫球蛋白装配, 类抗原加工, 病原体抗原及自身免疫性诱导。
小分子 HSP	Hsp27, Gro23, HSP16	参与蛋白质折叠和去折叠及多聚复合物的装配。	病原体抗原及自身免疫性诱导。
泛素	泛素	蛋白质降解。	类抗原加工, 淋巴细胞回巢, 自身免疫性诱导。

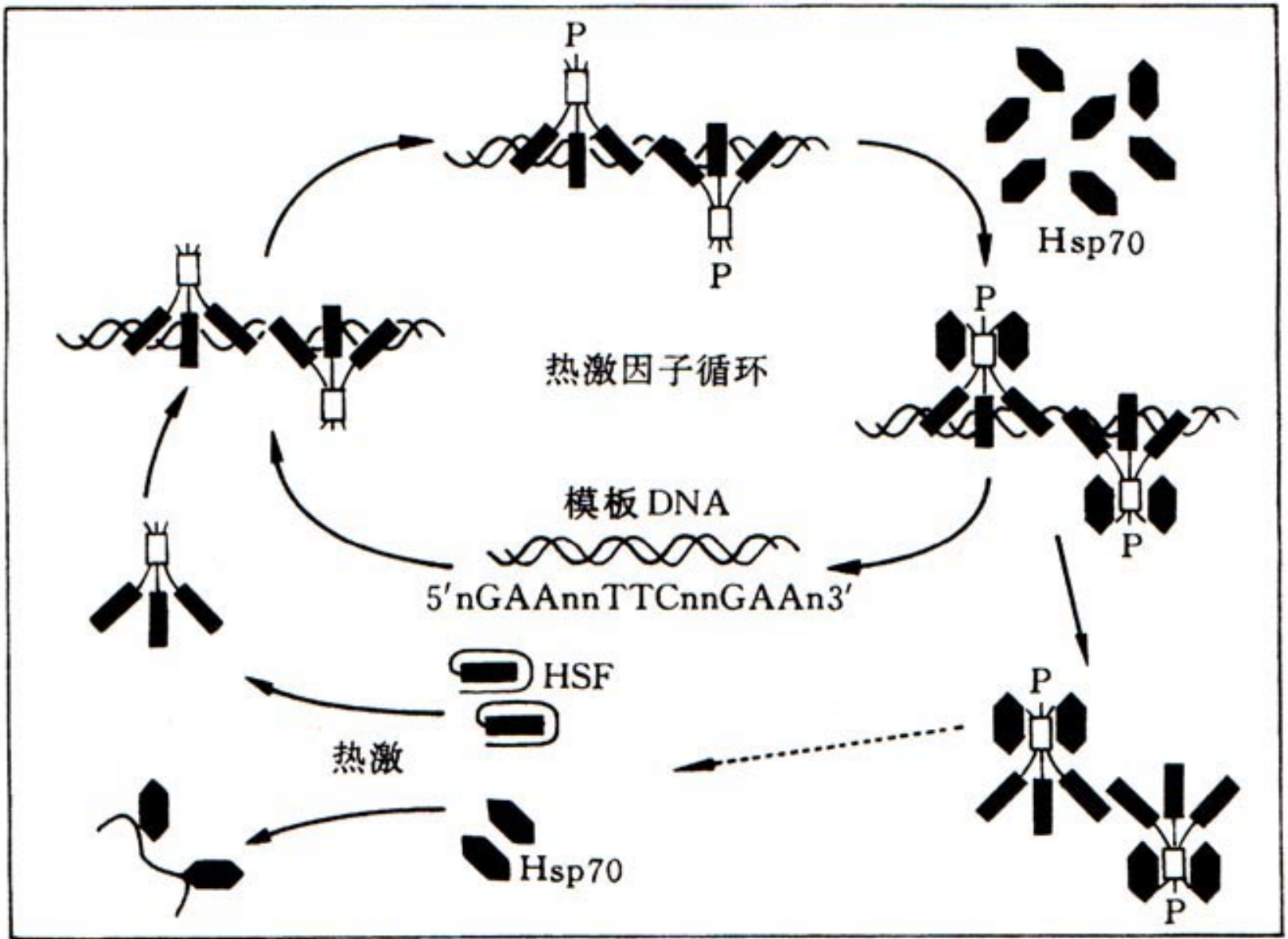


图7-40

图7-41 热激蛋白调控的基因表达机制



不受热或其他环境胁迫时，**HSF**主要以单体的形式存在于细胞质和核内。单体**HSF**没有**DNA**结合能力，**Hsp70**可能参与了维持**HSF**的单体形式。

受热激或其他环境胁迫时，细胞内变性蛋白增多，与**HSF**竞争结合**Hsp70**，从而释放**HSF**，使之形成三体并输入核内。**HSF**的三体能与**HSE**特异结合，促进基因转录。



HSF的这种能力可能还受磷酸化水平的影响。热激后，HSF不但形成三聚体，还会迅速被磷酸化。

HSF与HSE的特异性结合，引起包括Hsp70在内的许多热激应答基因表达，大量产生Hsp70蛋白。



随着热激温度消失，细胞内出现大量游离的**Hsp70**蛋白，它们与**HSF**相结合，形成没有**DNA**结合能力的单体并脱离**DNA**。



RNA 分子杂交
试验结果

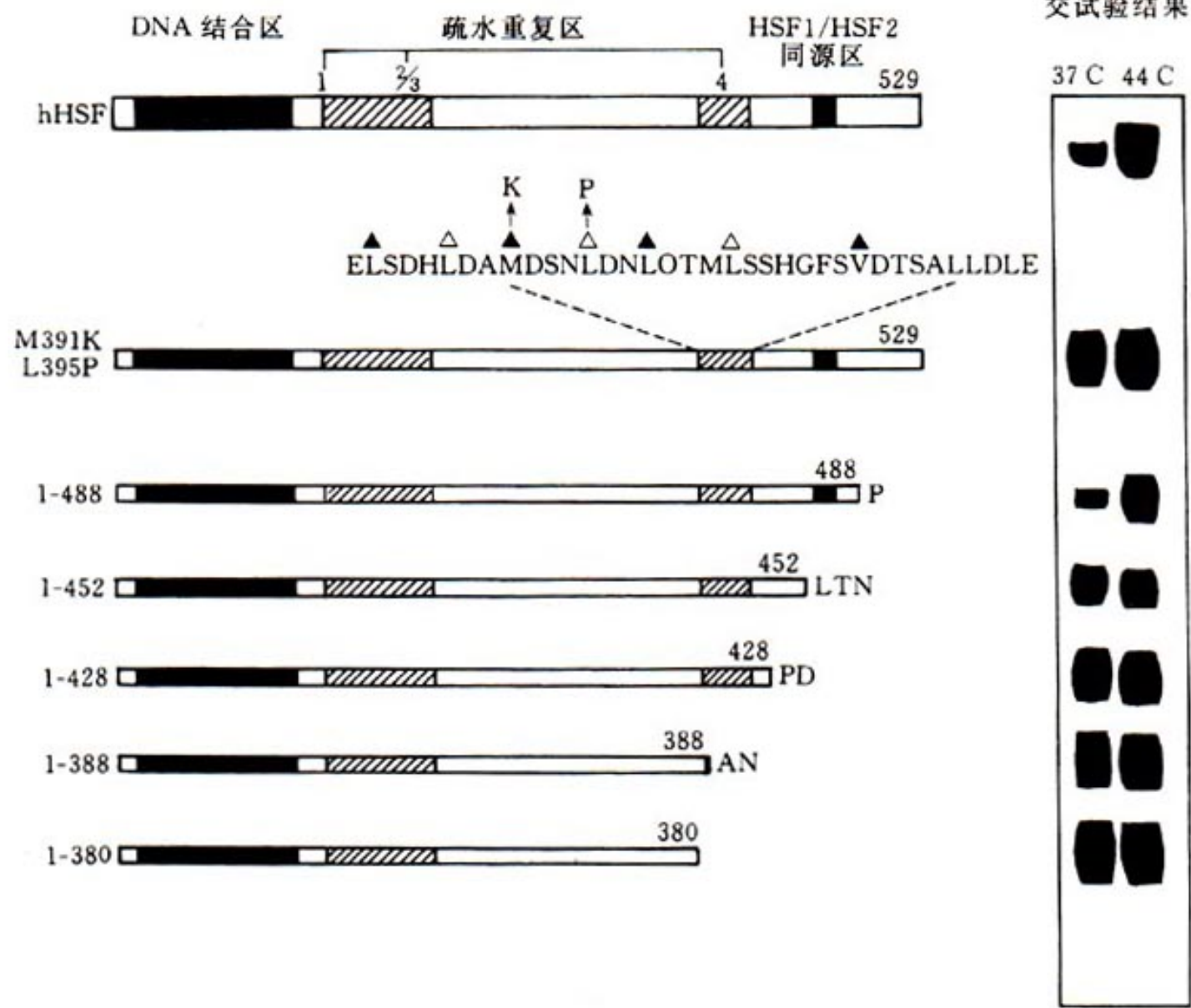


图7-41

图7-42 人热激蛋白的功能区分析



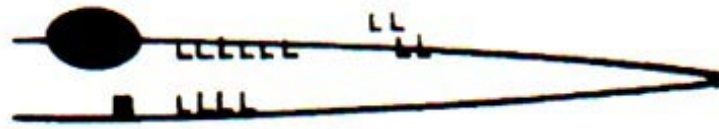
HSF具有多个可形成拉链的疏水DNA区域，其中3个位于N端，靠近DNA结合区，参与促进三体的形成。第4个拉链位于C末端，与位于第452~488位氨基酸残基之间的保守区一道参与维持HSF单体构象。



序列删除实验表明，无论去除第4个拉链区，还是更换该区甲硫氨酸³⁹¹→赖氨酸，亮氨酸³⁹⁵→脯氨酸，都会导致HSF突变体对HSE在常温下的高亲和力。删除第452~488位氨基酸，也可部分替代热激效应。



在常温下，**Hsp**蛋白第**1**和**4**号拉链发生相互作用，从而阻断了三体的形成；热激或其他环境胁迫导致内源拉链之间的解离，蛋白质伸展成长链，不同链上的拉链发生作用，形成具有**DNA**结合能力的三体。



常温



热激后

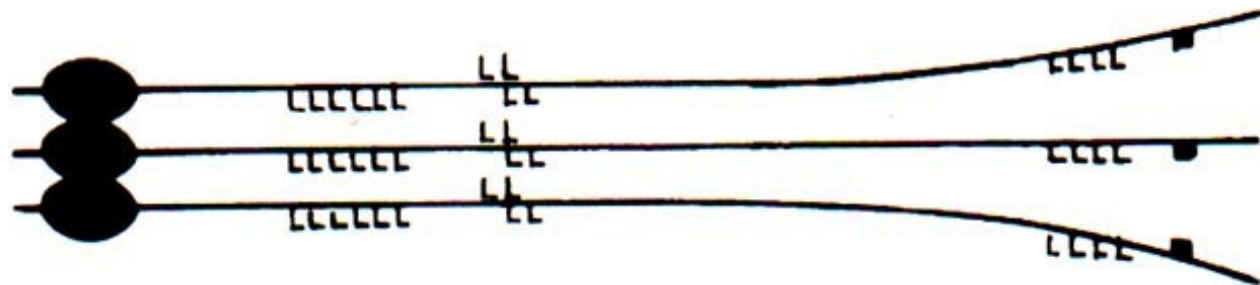


图7-42

图7-43 HSF 在常温及热激后的构象变化



7.5 其他水平上的基因调控

7.5.1 RNA的加工成熟

简单转录单位。这类基因只编码产生一个多肽，其原始转录产物有时需要加工，有时则不需要加工。这类基因转录后加工有3种不同形式。

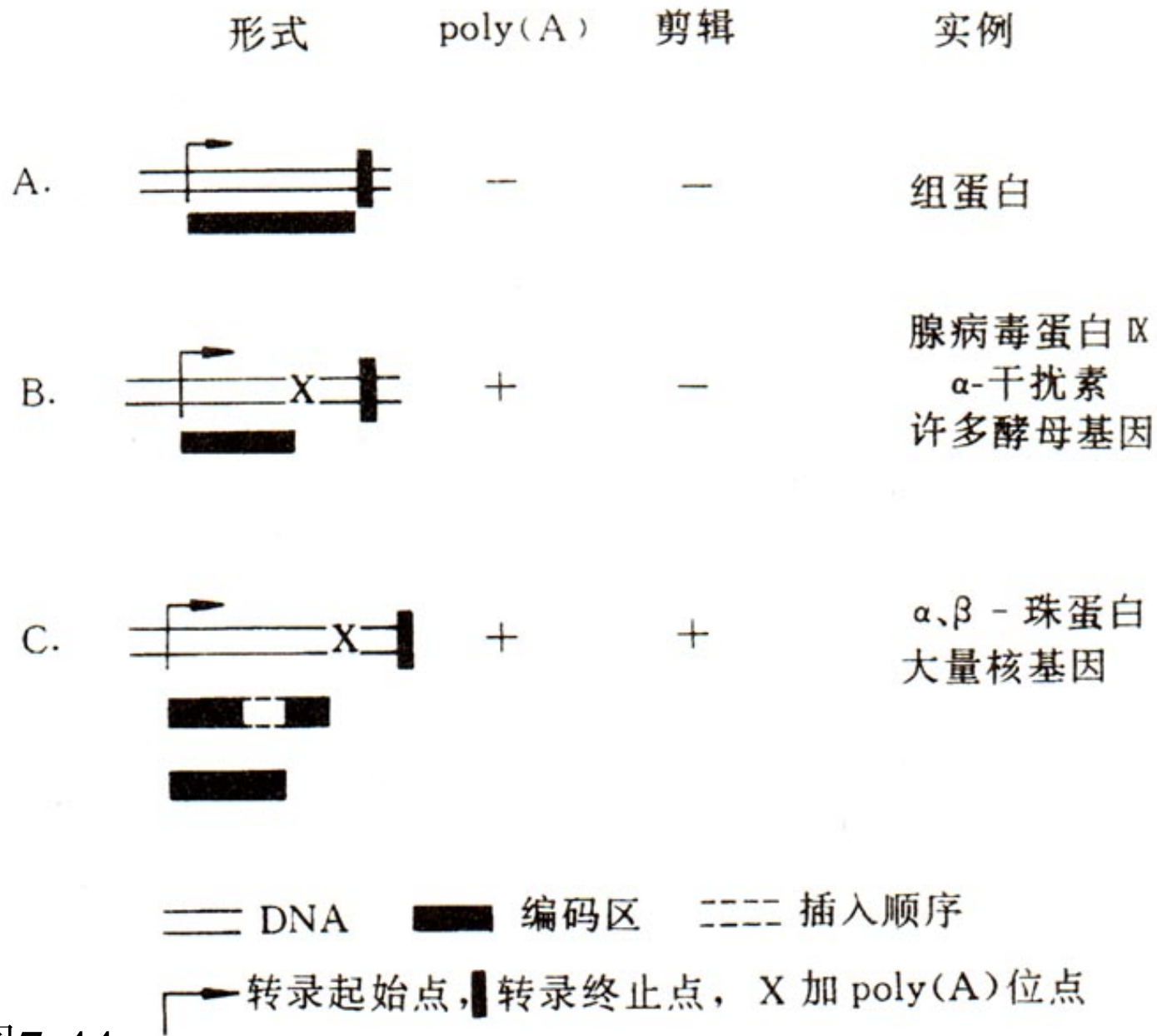


图7-44

图7-45 真核生物基因转录后加工的 3 种主要形式



复杂转录单位。含有复杂转录单位的主要是一些编码组织和发育特异性蛋白质的基因，它们除了含有数量不等的内含子以外，其原始转录产物能通过多种不同方式加工成两个或两个以上的**mRNA**。



利用多个5'端转录起始位点或剪接位点产生不同的蛋白质。

肌球蛋白碱性轻链基因选用了不同的5'转录起始位点及剪接不同外显子产生蛋白质异构体LC1和LC3的过程。

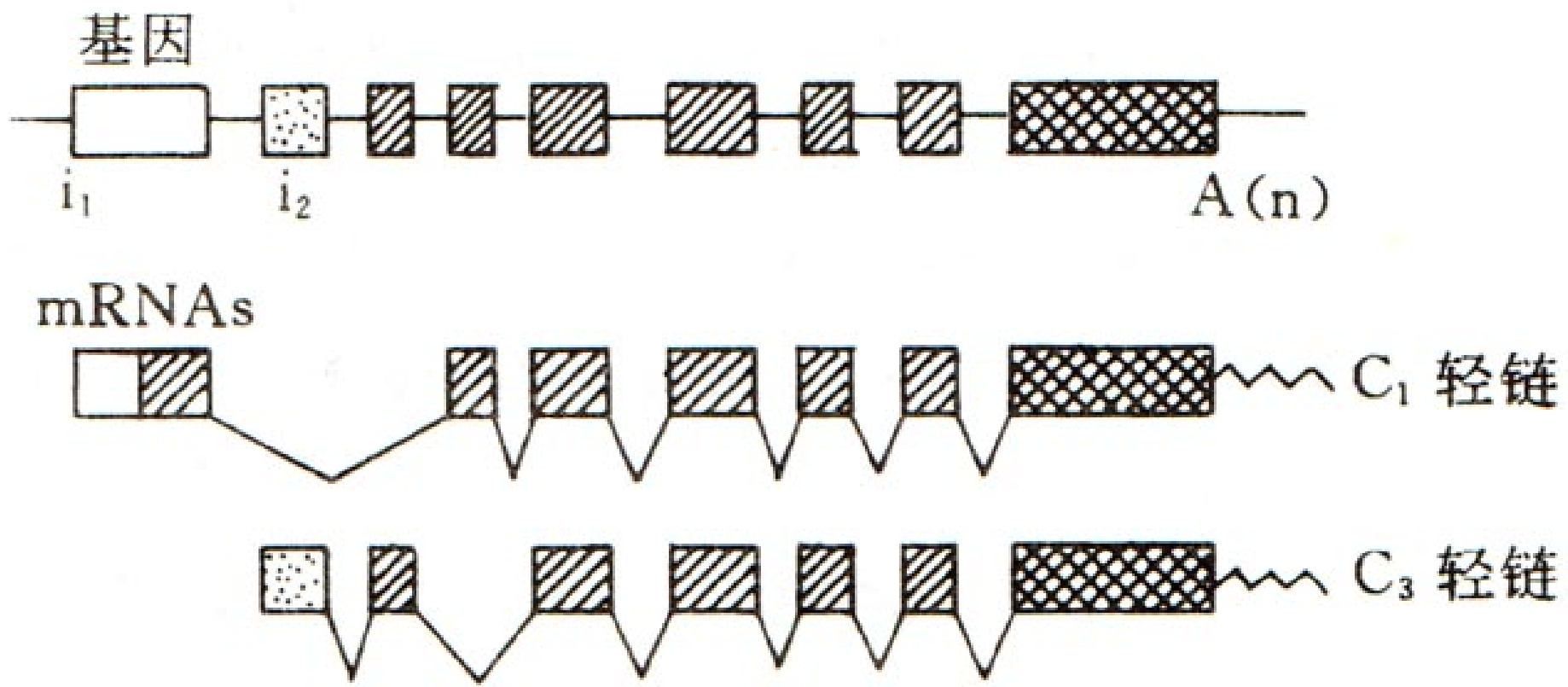


图7-46 肌球蛋白碱性轻链基因剪接的多样性

图7-45



大鼠降钙素基因编码与血钙代谢有关的蛋白质。它有一个转录起始位点和**6**个外显子，两个加多聚（A）位点。

初级转录产物以**0—1—2—3—4**号外显子连接在一起，并利用外显子4后面的多聚（A）位点，产生降钙素 mRNA。



若初级转录产物以**0—1—2—3—5—6**号外显子连接在一起，并利用外显子**6**以后的加多聚（A）位点时，就产生了编码与降钙素基因表达有关的多肽**CGRP**的**mRNA**。

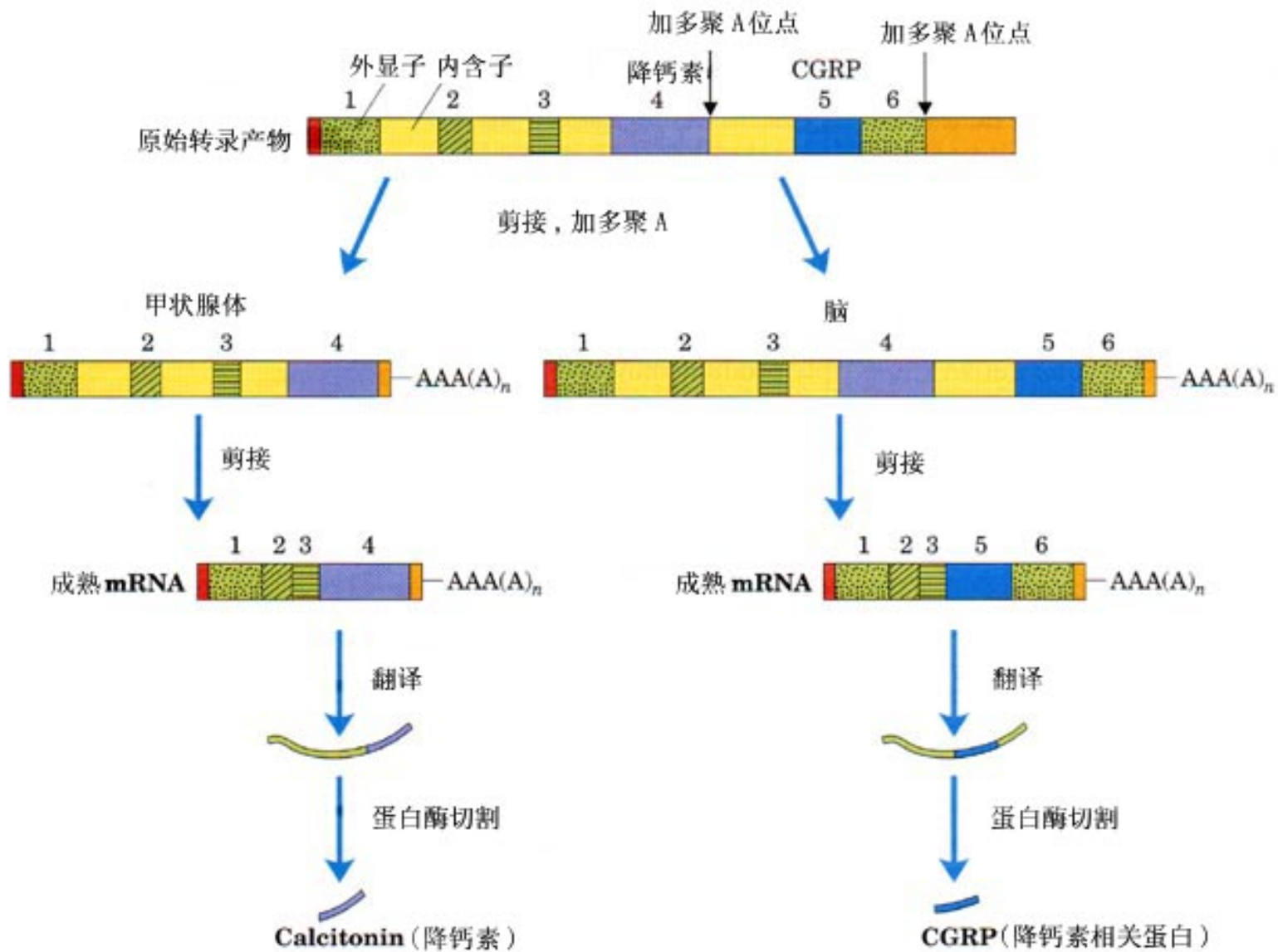
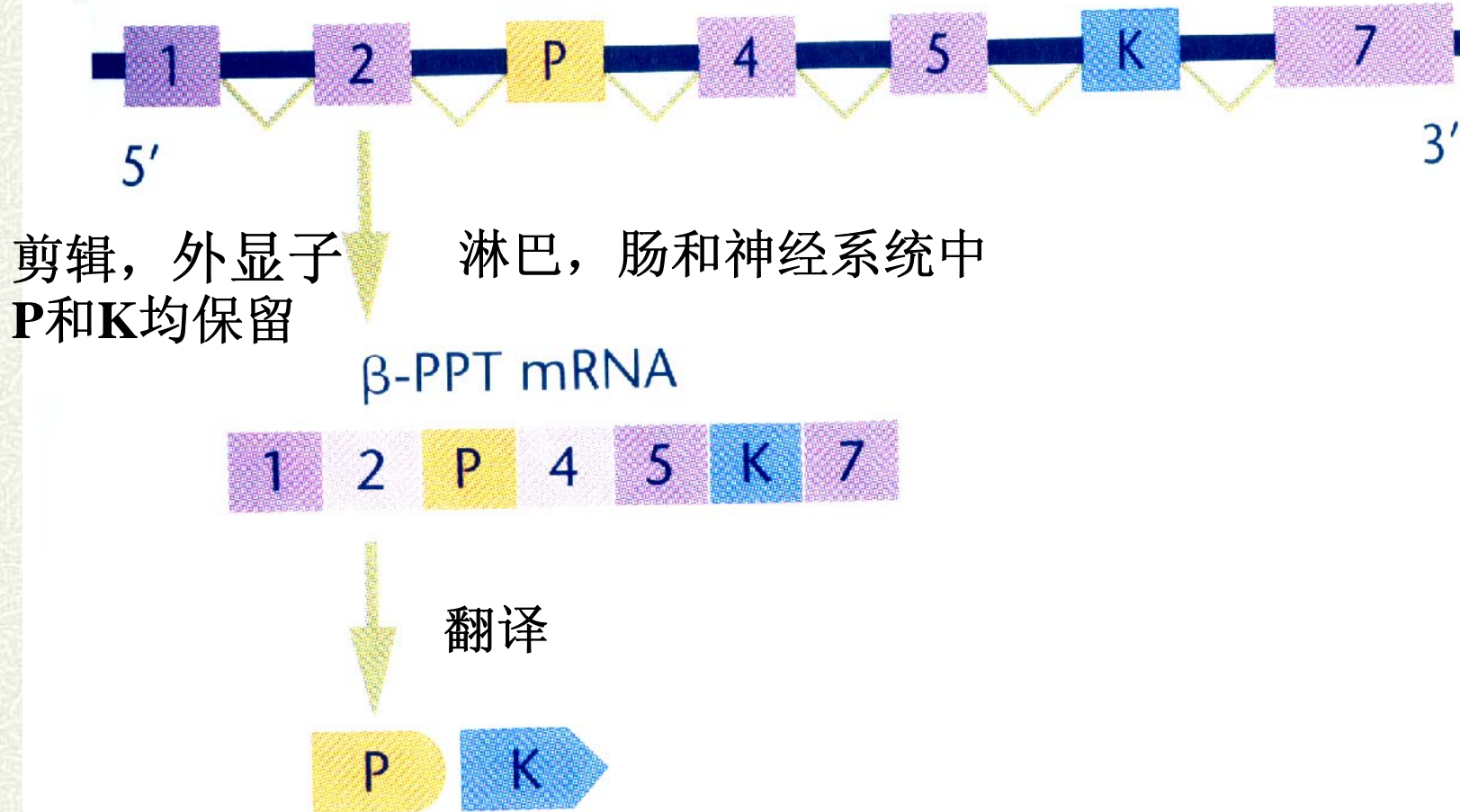


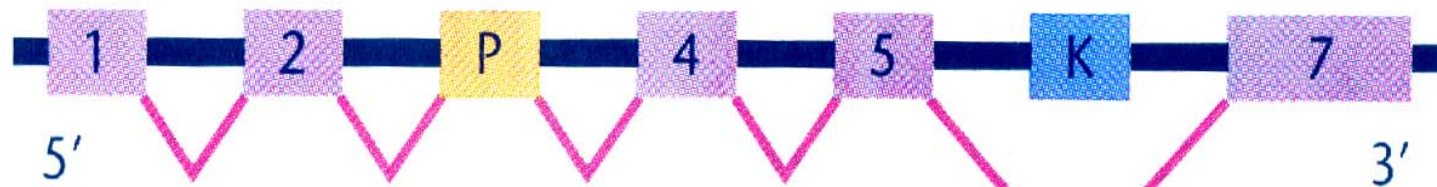
图7-46 前mRNA不同的剪接方式造成了不同组织中不同的降钙素样蛋白



(a) PPT初始转录产物



(b) PPT初始转录产物



神经系统中
只有P外显子被保留

α -PPT mRNA



翻译





mRNA的稳定性与基因表达调控

高等生物中转运铁蛋白受体（**TfR**）和铁蛋白负责铁吸收和铁解毒，其**mRNA**上存在铁应答元件（**iron response element, IRE**）。

IRE与**IRE**结合蛋白（**IREBP**）相互作用控制了这两个**mRNA**的翻译效率。细胞中能产生两个数量级的蛋白水平差异，却没有在**mRNA**水平上发现显著差异。



5'非翻译区的IRE控制了铁蛋白mRNA的翻译效率，去掉这个非翻译区IRE，可造成铁蛋白的永久性高水平翻译。

当细胞缺铁时，IREBP与IRE具有高亲和力，两者的结合有效地阻止了铁蛋白mRNA的翻译。此时，TfR mRNA上3'非翻译区中的IRE也与IREBP特异结合，有效地阻止TfR mRNA的降解，促进TfR蛋白的合成。

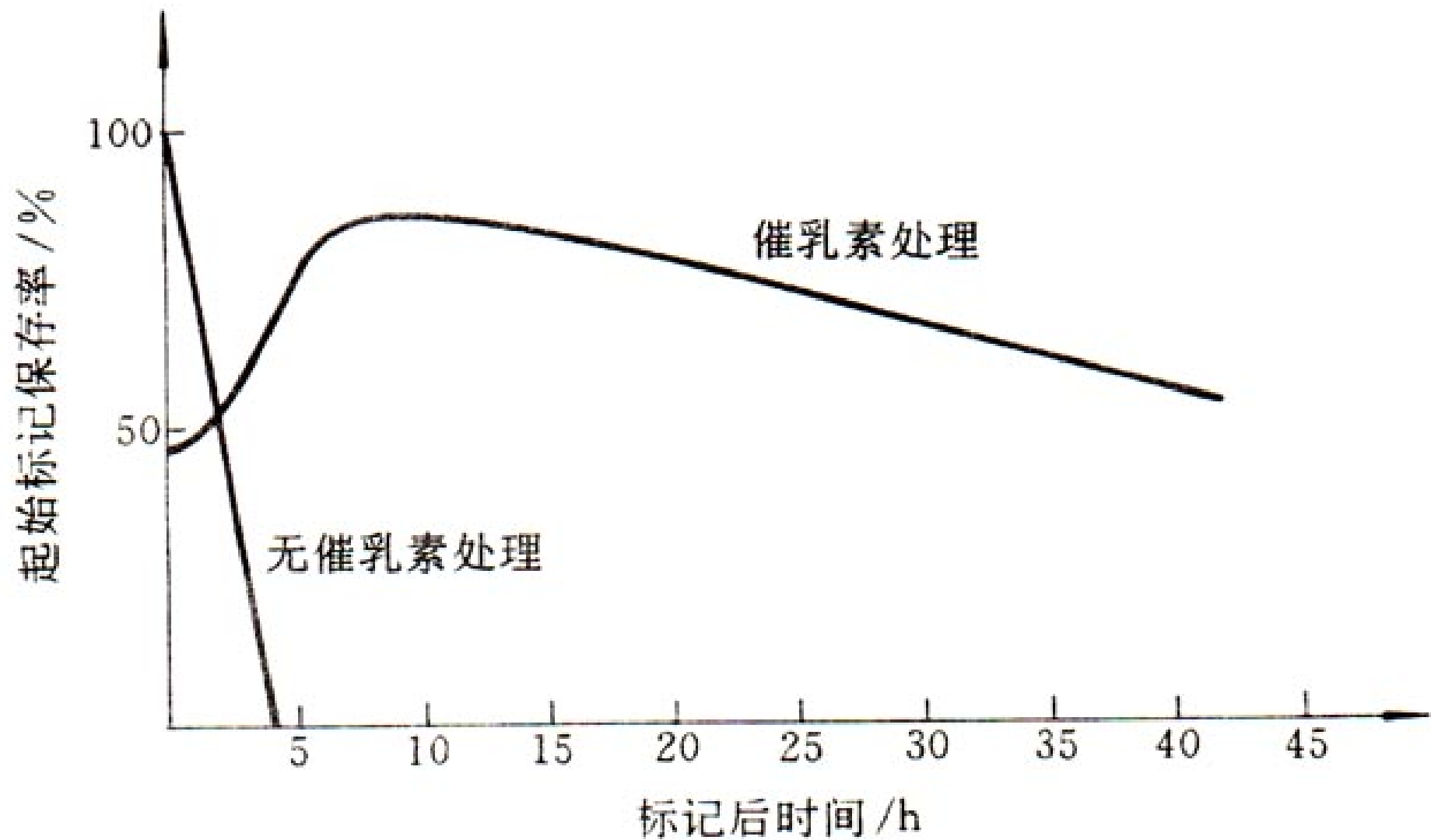


图7-49 催乳素处理能明显延缓酪蛋白 mRNA 的降解



催乳素能明显延缓酪蛋白（**casein**）**mRNA**的降解。不加入催乳素时，体系中的酪蛋白**mRNA**在1小时内降解**50%**，而加入催乳素后**40**小时，酪蛋白**mRNA**才降解**50%**。

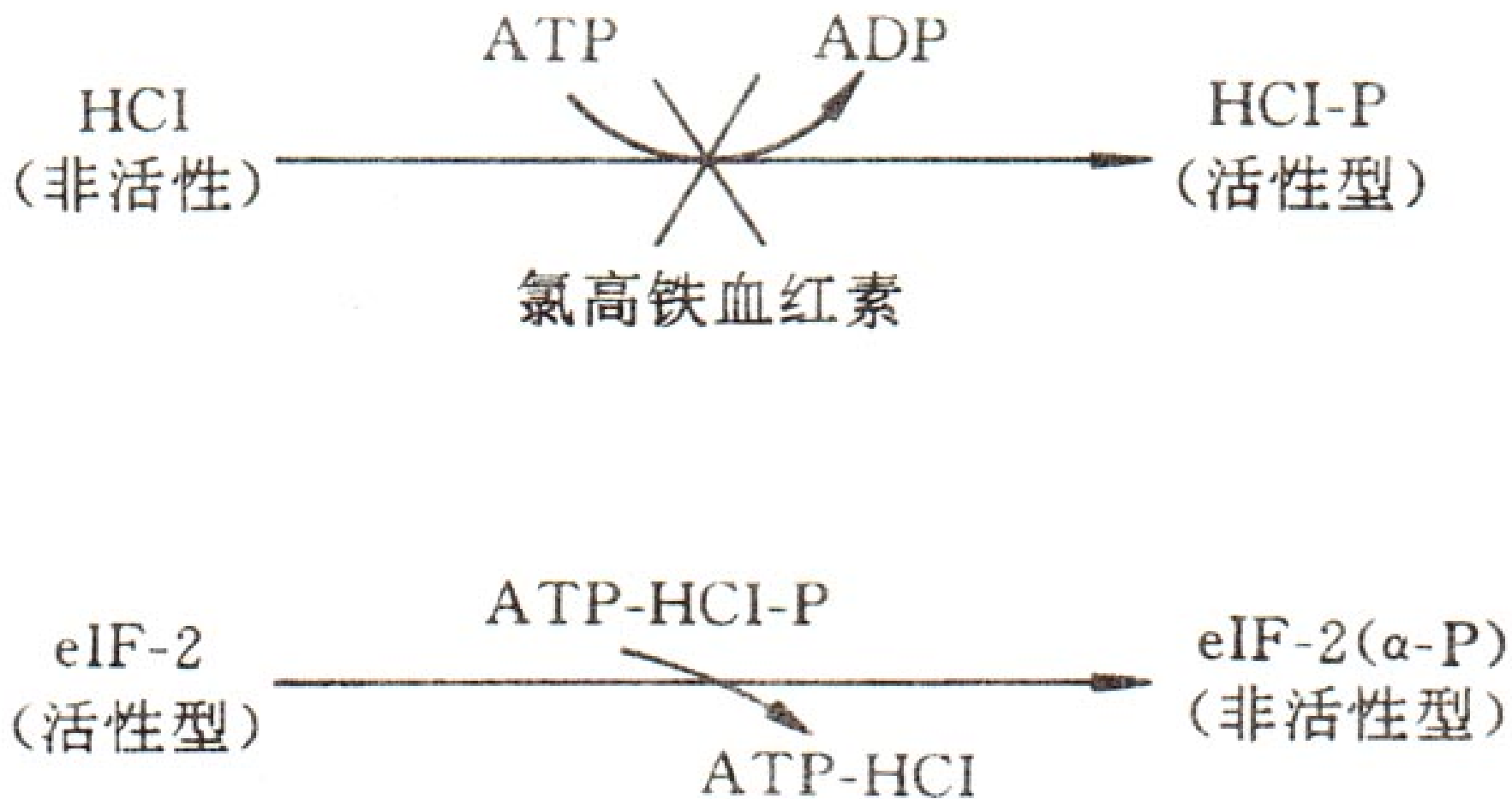


图7-50 eIF-2、HCl 及氯高铁血红素的相互关系



用兔网织红细胞粗抽提液研究蛋白质合成时发现，如果不向这一体系中添加氯高铁血红素，几分钟之内蛋白质合成活性急剧下降，直到完全消失。

没有氯高铁血红素存在时，网织红细胞粗抽提液中的蛋白质合成抑制剂就被活化，从而抑制蛋白质合成。



该抑制剂HCI受氯高铁血红素调节，是eIF-2的激酶，使eIF-2的 α 亚基磷酸化，并由活性型变成非活性型。没有生物活性的HCI也可以通过自身的磷酸化变成活性型。氯高铁血红素阻断了HCI的活化过程。