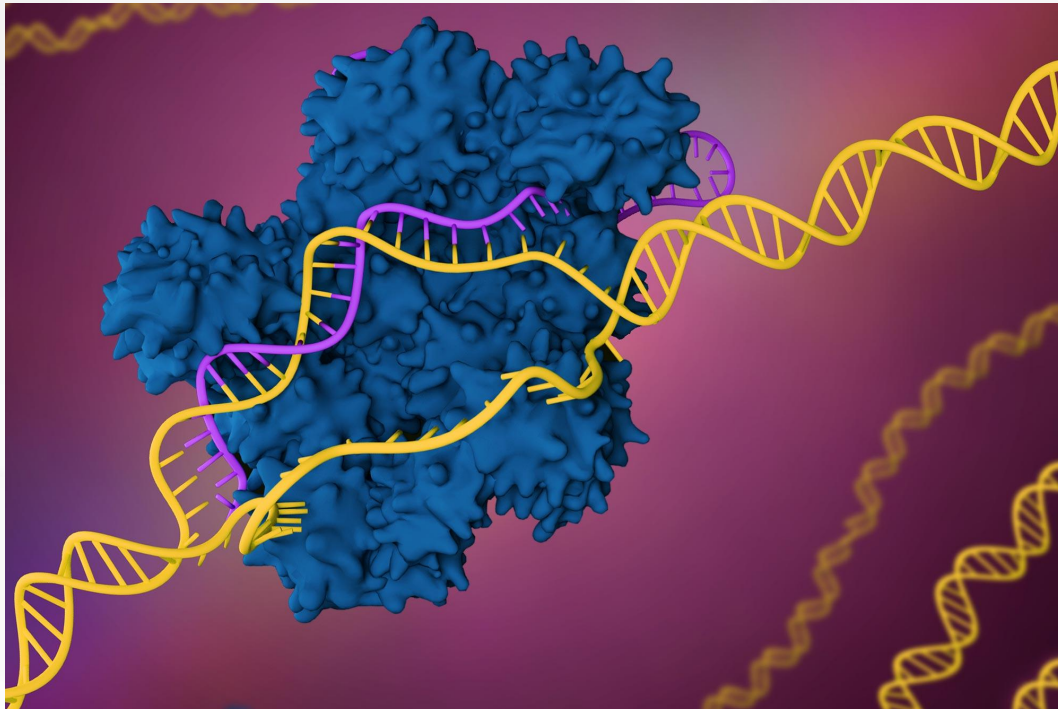


第三章 转录、转录后加工

Transcription and post-transcription processing



黎婕昕, 副教授

中山大学药学院, 406室

E-mail: lijixin3@mail.sysu.edu.cn

目录

01

A General Overview

02

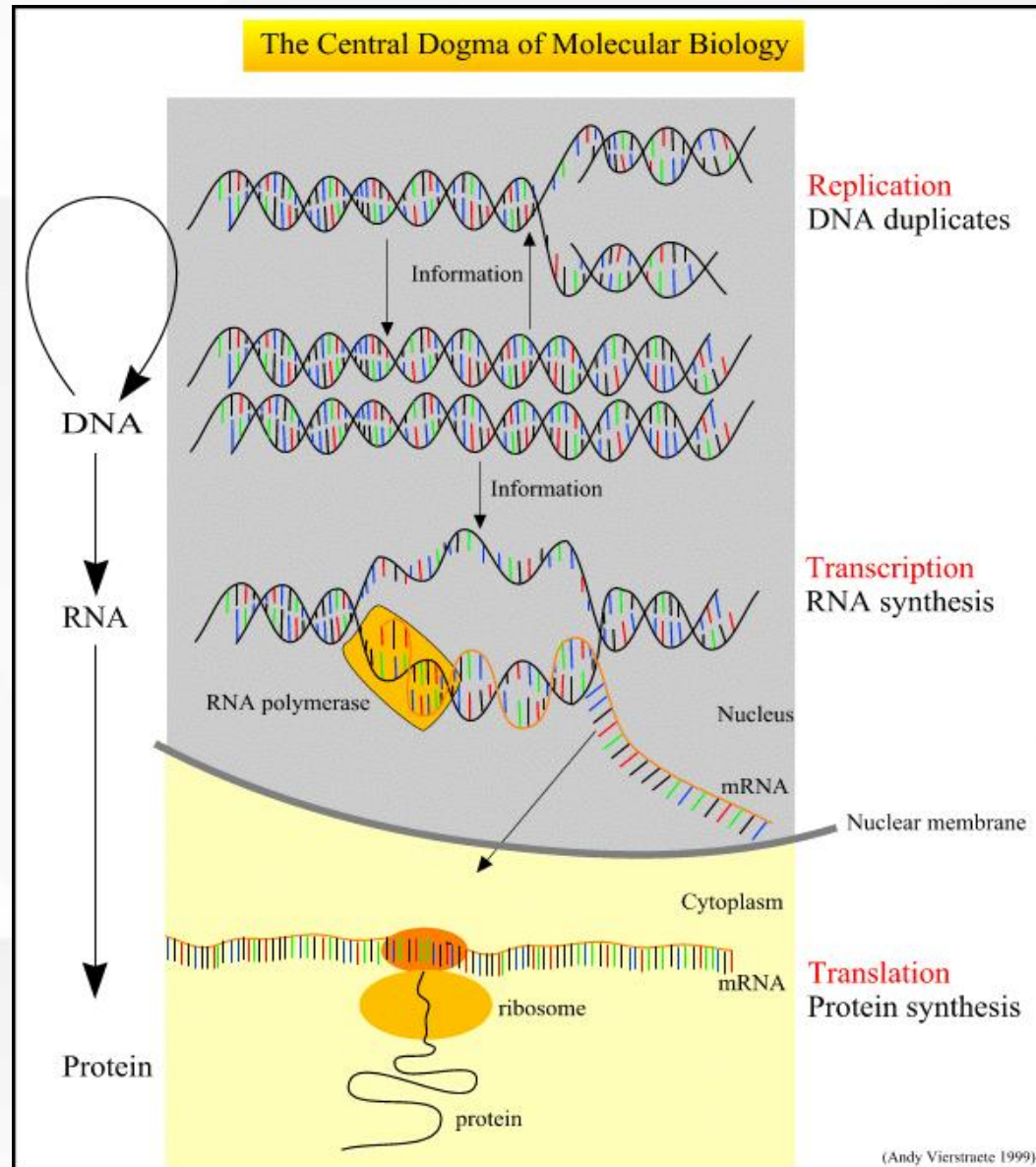
DNA Structures Related to Transcription

03

Transcription in *Prokaryotes* and *Eukaryotes*

04

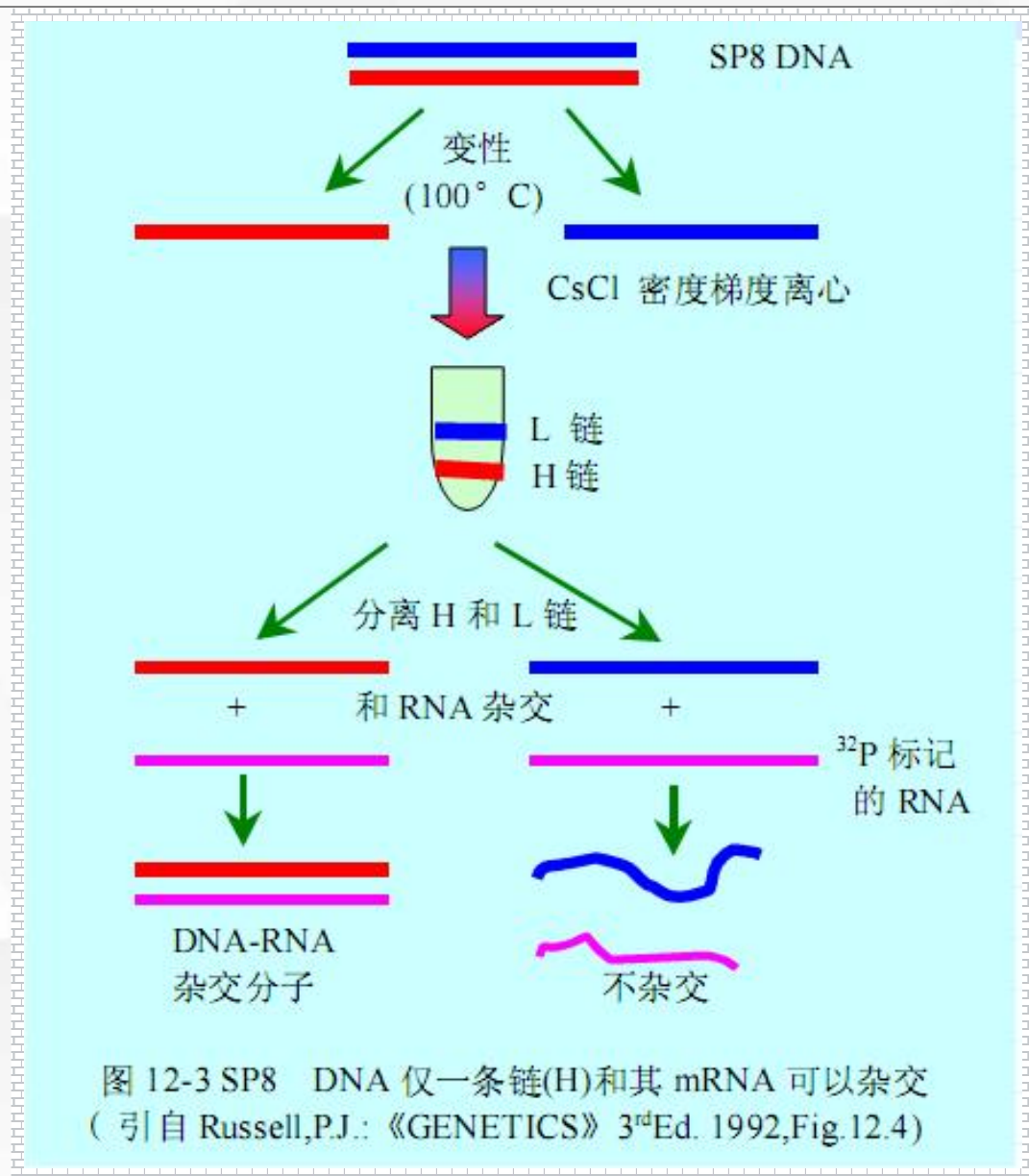
Post-transcription Processing





- **转录 (transcription):** 是指以DNA为模板，在依赖于DNA的RNA聚合酶催化下，以4种NTP (**ATP、CTP、GTP和UTP**) 为原料，合成RNA的过程。
 - 是遗传信息由DNA向RNA传递的过程；
 - 是**基因表达**的第一步，也是最关键的一步；
 - mRNA, tRNA, rRNA
 - 在有些病毒中，RNA也可以指导合成RNA。
 - 以dsDNA中的一**条单链作为转录模板** (杂交实验所证实)

杂交实验证实转录以**一条DNA单链**为模板⁵



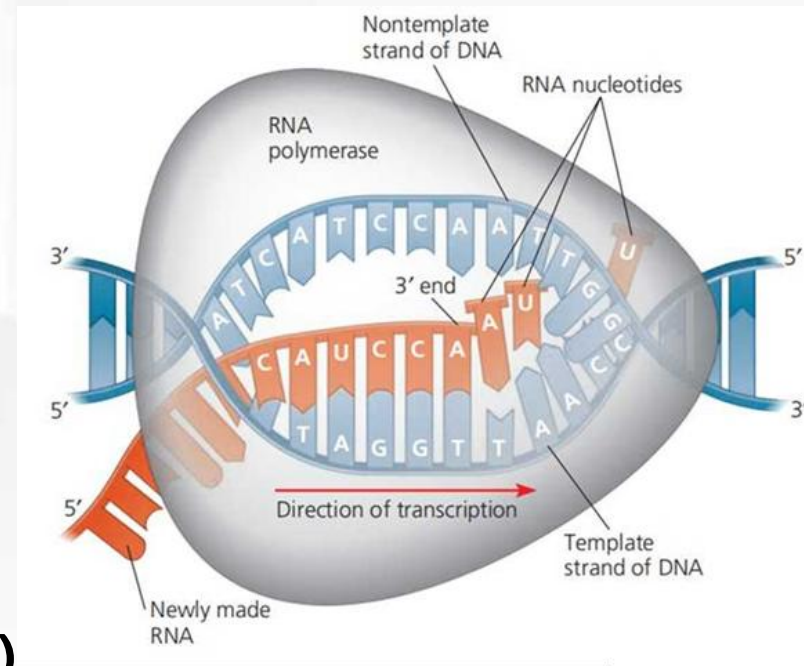
1.1 概述

6

- 合成反应可表示为：



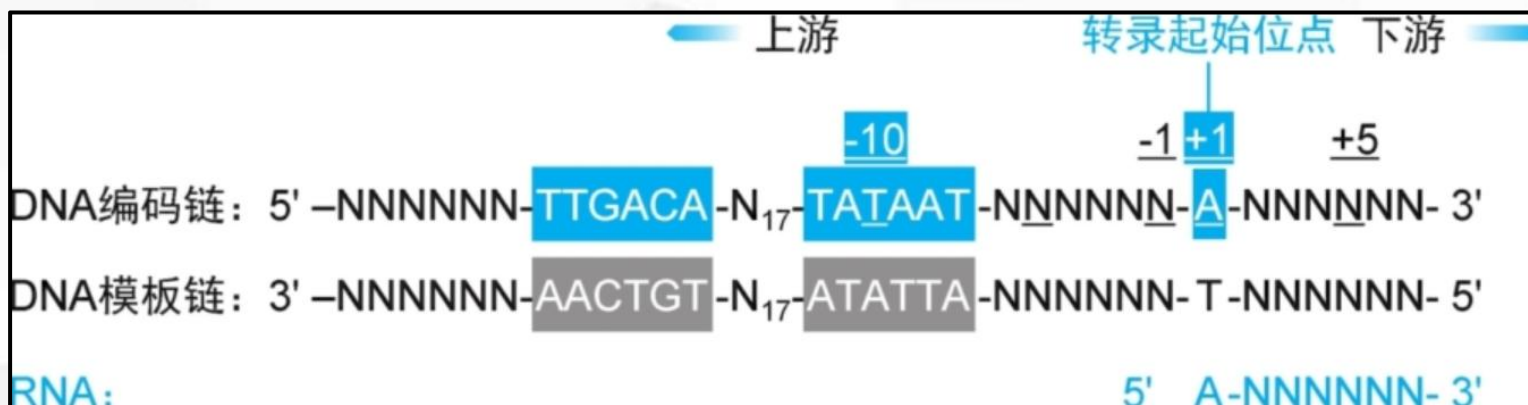
- ✓ 依赖DNA的**RNA聚合酶**作用下进行：
 - A = U、C ≡ G 合成RNA分子
- ✓ 转录合成RNA链的方向与模板链极性方向相反：
 - RNA链转录**合成方向**为**5' → 3'**；
 - 模板单链DNA的极性方向为3' → 5'，而非模板单链的极性方向与RNA链相同，均为5' → 3'。（**书写**）



基因序列的书写规则

7

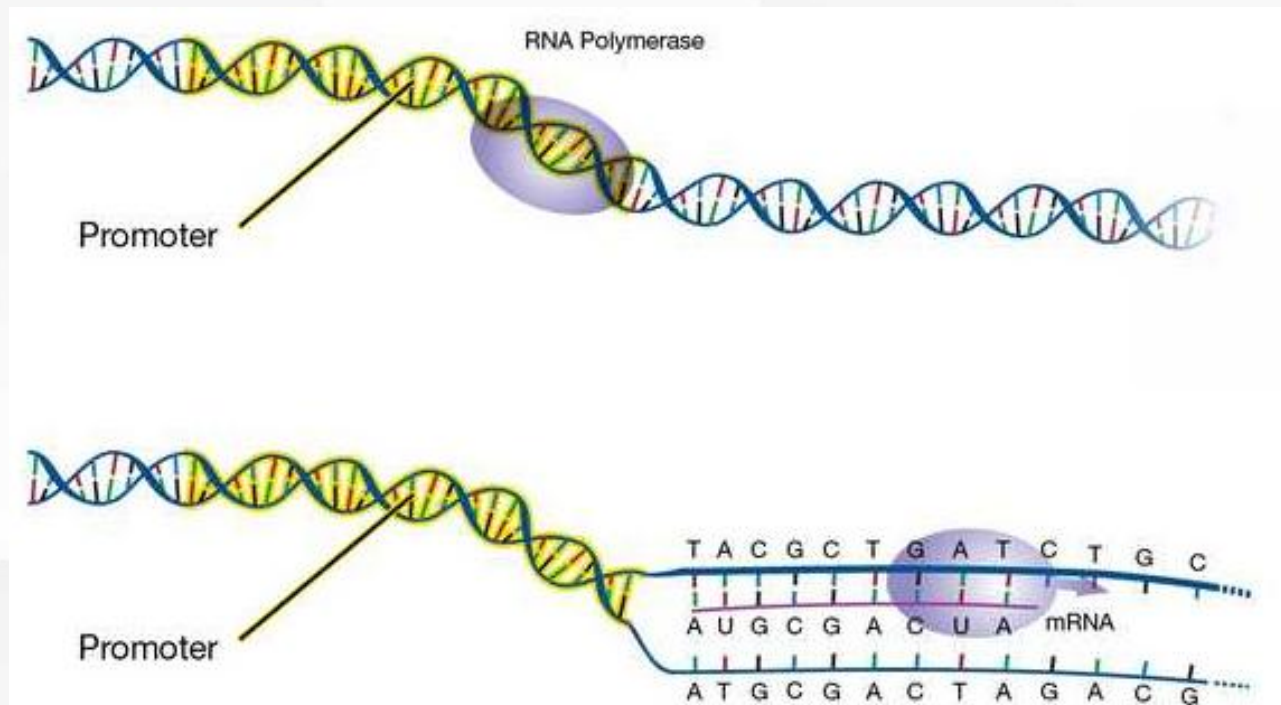
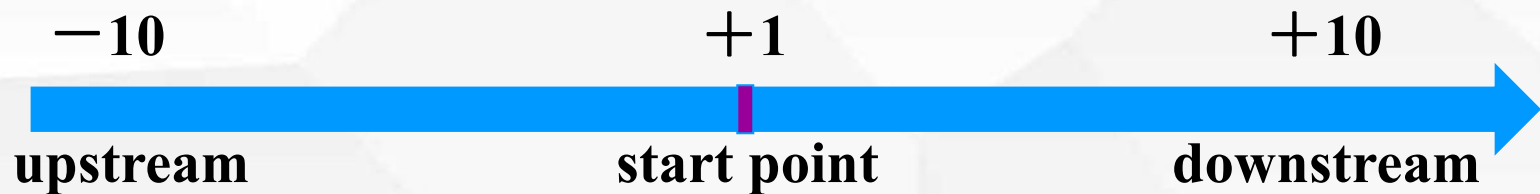
- I. 因为DNA双链的序列是互补的，所以只要给出一股链的序列，另一股链的序列也可推出。因此，为了避免繁琐，**书写DNA序列时只写出一股链。**
- II. 因为DNA编码链与转录产物RNA的核苷酸序列一致，只是RNA中以U取代了DNA中的T，所以为了方便解读遗传信息，**一般只写出编码链。**
- III. 通常将编码链上位于**转录起始位点的核苷酸编为+1号**；转录进行的方向称为下游（downstream），核苷酸依次编为+2号、+3号等；相反方向称为上游（upstream），核苷酸依次编为-1号、-2号等，**没有0号。**



基因序列的书写规则

8

- 转录起点即转录原点记为 +1，其上游记为负值，下游记为正值。



»» 1.1 概述- 转录的特征

9

- **选择性转录：**不同组织细胞、或同一细胞在机体不同的生长发育阶段、根据生存条件和代谢需要对基因进行选择转录；
 - 大肠杆菌通常只有约5%的基因处于高表达状态，成人个体每种组织一般只表达10%~20%的基因；（复制：全部DNA）
- **不对称转录：**DNA每个基因的转录区都只有一股链可被转录；
 - 作为转录的模板DNA链又称**反义链**（模板链，template strand）；而另一条链由于与RNA序列相同又称**有义链**（编码链，coding strand）。
- **连续性转录：**一个RNA分子从头到尾由一个RNA聚合酶分子催化合成；
- **转录后加工：**RNA聚合酶转录合成的RNA称为RNA前体（pre-RNA）、初级转录产物（primary transcript），大多数需要经过加工才能成为成熟RNA。

性质	模板链	编码链
其他名称	负链、反义链、非编码链	正链、正义链、有义链
可以转录	一定	不一定
转录后加工产物	mRNA、tRNA、rRNA	反义RNA
转录范围	整个转录区	可以是转录区局部
转录时序	先	后



- RNA聚合酶 (RNA Pol) : 全称**DNA依赖的RNA聚合酶** (DNA-dependent RNA polymerase, DDRP) , 又称**转录酶**。
 - RNA聚合酶催化RNA的转录合成, 是参与转录的关键要素之一。
 - 原核生物和真核生物的RNA聚合酶有其共同特点, 但在结构、组成和性质等方面不尽相同。

1.2 RNA Pol与DNA Pol的异同

13

① 原料:

- RNA - 核糖核苷三磷酸 (NTP)
- DNA - 脱氧核糖核苷三磷酸 (dNTP)

② 合成酶:

- RNA聚合酶 - RNA
- DNA聚合酶 - DNA

③ 合成方向及引物:

- 均为5' → 3'
- RNA合成不需引物, 而DNA复制需引物

④ 模板:

- 转录时只有一条DNA链为模板
- 复制时两条链都可作为模板

⑤ 忠实性:

- 转录的忠实性略低于复制。
- RNA聚合酶缺乏自我校对机制。

RNA pol和DNA pol 的不同

- (1) RNA pol没有校对功能;
- (2) 能起始新的RNA链。

1.2 转录与复制的异同

14

项目	转录	复制
聚合酶	RNA聚合酶	DNA聚合酶
DNA模板	基因组局部（转录区，选择性转录） 转录单链（模板链，不对称转录）	基因组全部复制双链 （半保留复制）
原料	NTP	dNTP
起始	启动子	复制起点
引物	不需要	需要
碱基配对原则	dA-rU, dT-rA, dG-rC, dC-rG	dA-dT, dT-dA, dG-dC, dC-dG
错配率	$10^{-4} \sim 10^{-5}$ （保真性低）	$10^{-6} \sim 10^{-8}$ （保真性高）
连续性	连续	不连续
终止	终止子	终止区
产物	单链RNA	双链RNA
后加工	有	无

1.2.1 原核生物RNA聚合酶

15

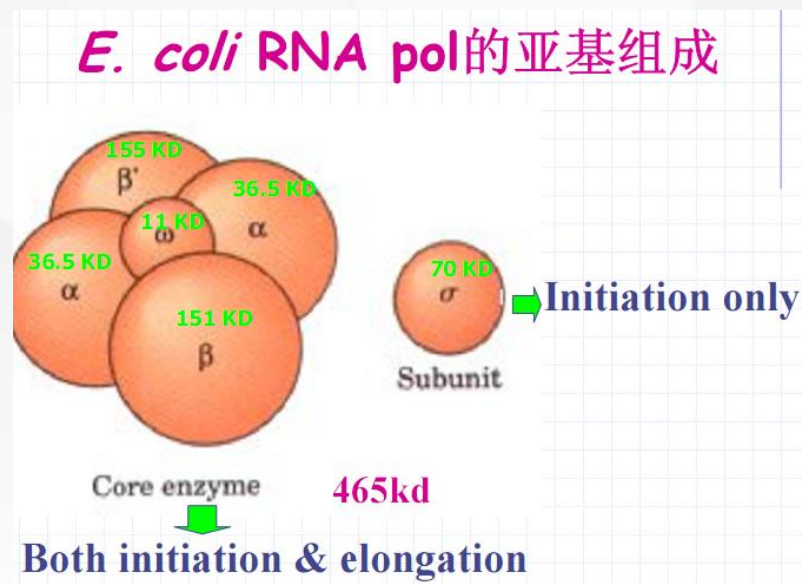
A. 原核生物的RNA聚合酶:

全酶 (holoenzyme) $\alpha_2\beta\beta'\omega\sigma$
5种亚基 (六聚体)

σ 因子

核心酶 $\alpha_2\beta\beta'\omega$: 4种亚基 (五聚体)

- σ 因子: 识别启动子, 提供起始信号;
 - 无催化功能;
 - σ 因子单独存在时不能与DNA模板结合;
- 核心酶: 催化特异的NTP形成单核苷酸;
 - 与DNA模板非特异性结合。

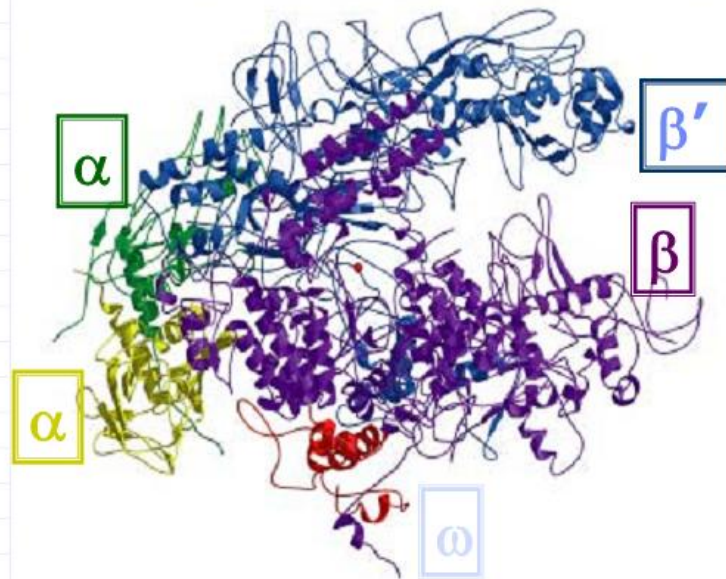


1.2.1 原核生物RNA聚合酶

16

亚基	功能	基因
α	启动RNA聚合酶组装，通过CTD直接识别并结合上游启动子元件，与某些激活蛋白结合	<i>rpoA</i>
β	含活性中心，催化形成磷酸二酯键	<i>rpoB</i>
β'	结合DNA模板	<i>rpoC</i>
ω	促进RNA聚合酶组装，参加某些转录调控	<i>rpoZ</i>
σ^{70}	与核心酶构成全酶后直接识别并结合启动子元件	<i>rpoD</i>

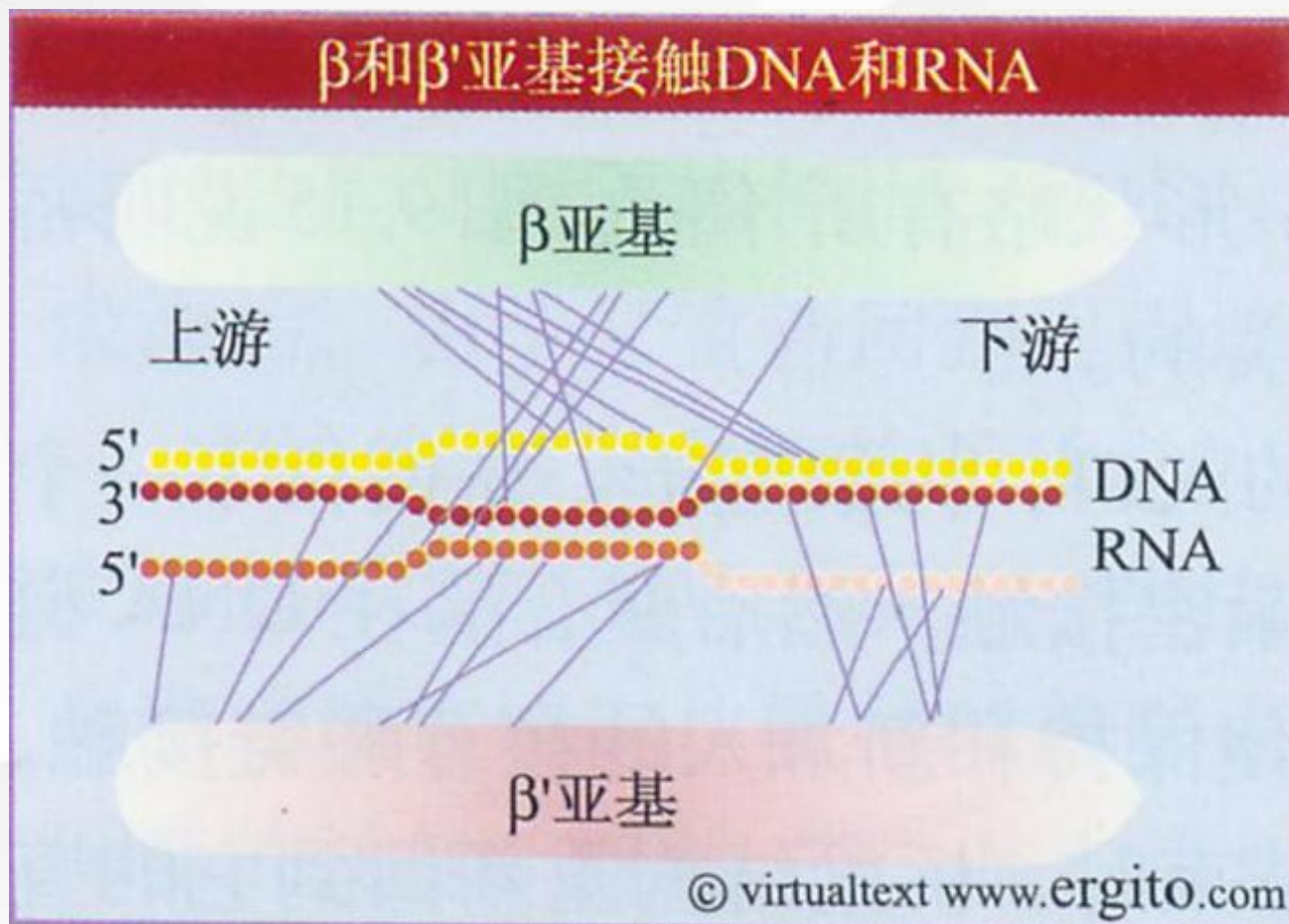
The core enzyme alone synthesizes RNA



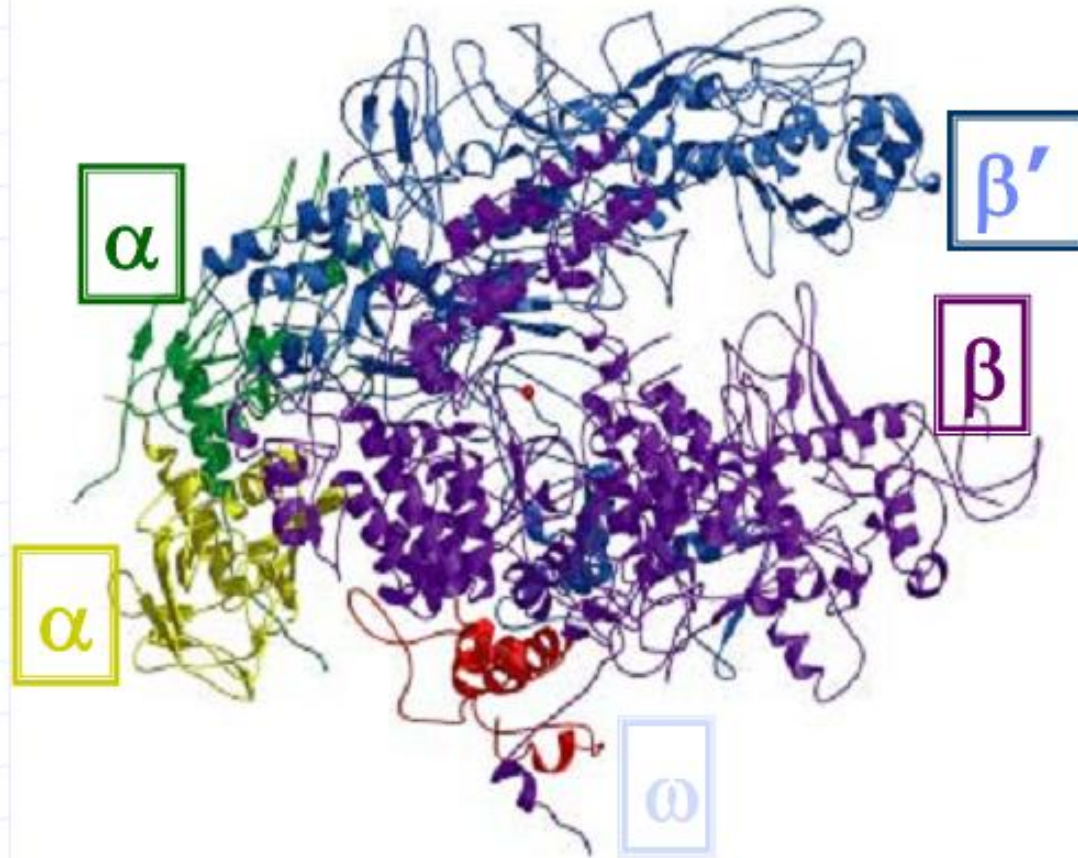
β和β'亚基的功能

17

- RNA Pol大概覆盖~40bp DNA区域
- 转录泡: ~17bp



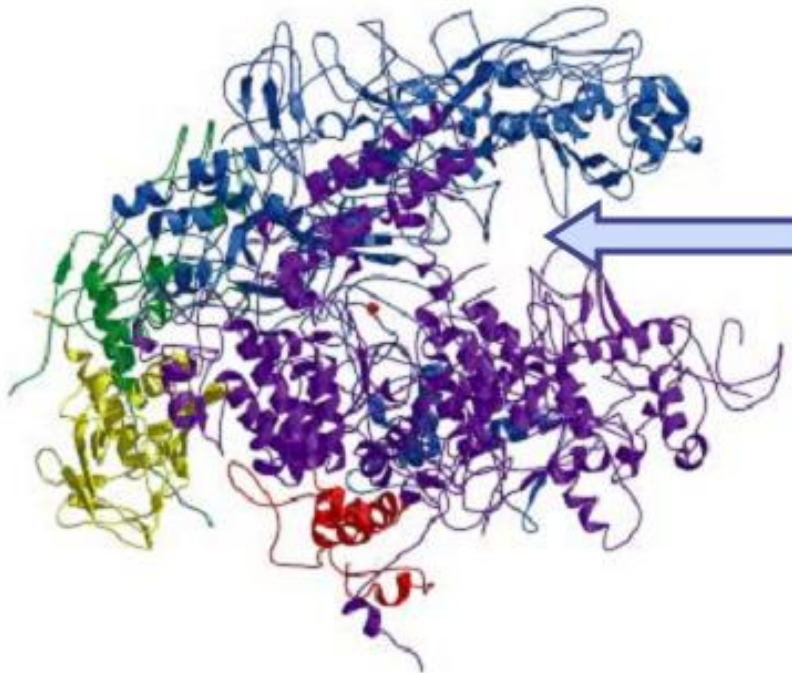
The core enzyme alone synthesizes RNA



RNA聚合酶的结构

19

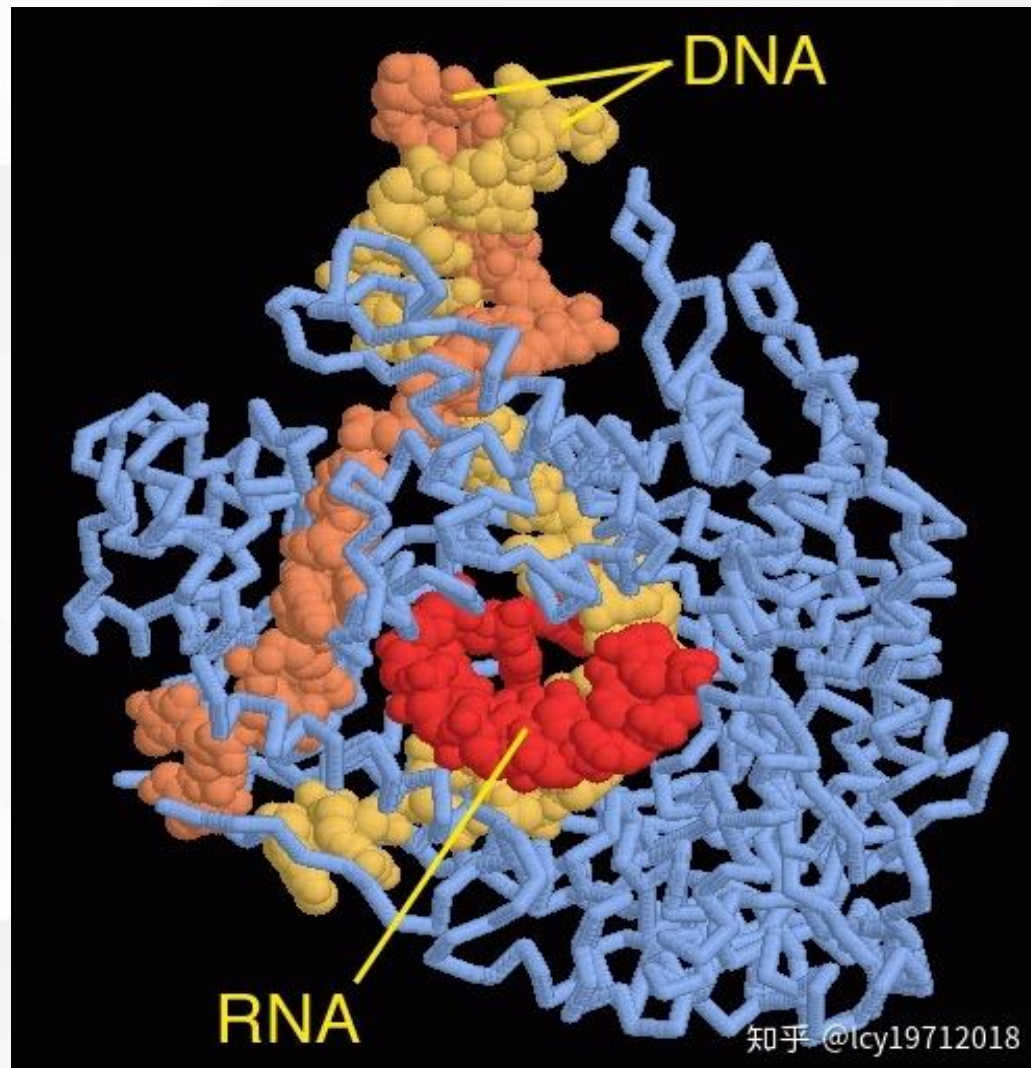
“Crab claw” shape of RNAP
(The shape of DNA pol is___)



Active center cleft

RNA聚合酶的结构

20



*E. coli*中不同的 σ 因子可识别不同的启动子

Gene	Factor	Use	-35 Sequence	Separation	-10 Sequence
<i>rpoD</i>	σ^{70}	general	TTGACA	16-18 bp	TATAAT
<i>rpoH</i>	σ^{32}	heat shock	CCCTTGAA	13-15 bp	CCCGATNT
<i>rpoE</i>	σ^E	heat shock	not known	not known	not known
<i>rpoN</i>	σ^{54}	nitrogen	CTGGNA	6 bp	TTGCA
<i>fliA</i>	σ^F	flagellar	CTAAA	15 bp	GCCGATAA

1.2.2 真核生物RNA聚合酶

B. 真核生物的RNA聚合酶:

● RNA聚合酶I:

- 不受 α -鹅膏蕈碱的抑制, 大于 10^{-3}mol/L
- 存在于核仁中, 合成5.8S rRNA、18S rRNA和28S rRNA;

● RNA聚合酶II:

- 对 α -鹅膏蕈碱最为敏感, $10^{-8} \sim 10^{-9}\text{mol/L}$
- 存在于核质中, 合成hnRNA、snRNA;

● RNA聚合酶III:

- 对 α -鹅膏蕈碱中度敏感, 在 $10^{-4} \sim 10^{-5}\text{mol/L}$ 时表现抑制;
- 存在于核质中, 合成tRNA, 5S rRNA。

- **线粒体和叶绿体：**

- **发现少数RNA聚合酶，分子量小，活性低**
- **mt mRNA, rRNA, tRNA；能被利福霉素或利福平抑制（与原核生物RNA聚合酶相似）**
- **由核基因编码，在细胞浆中合成后运送至细胞器中**

1.2.3 真核生物三种RNA聚合酶比较

24

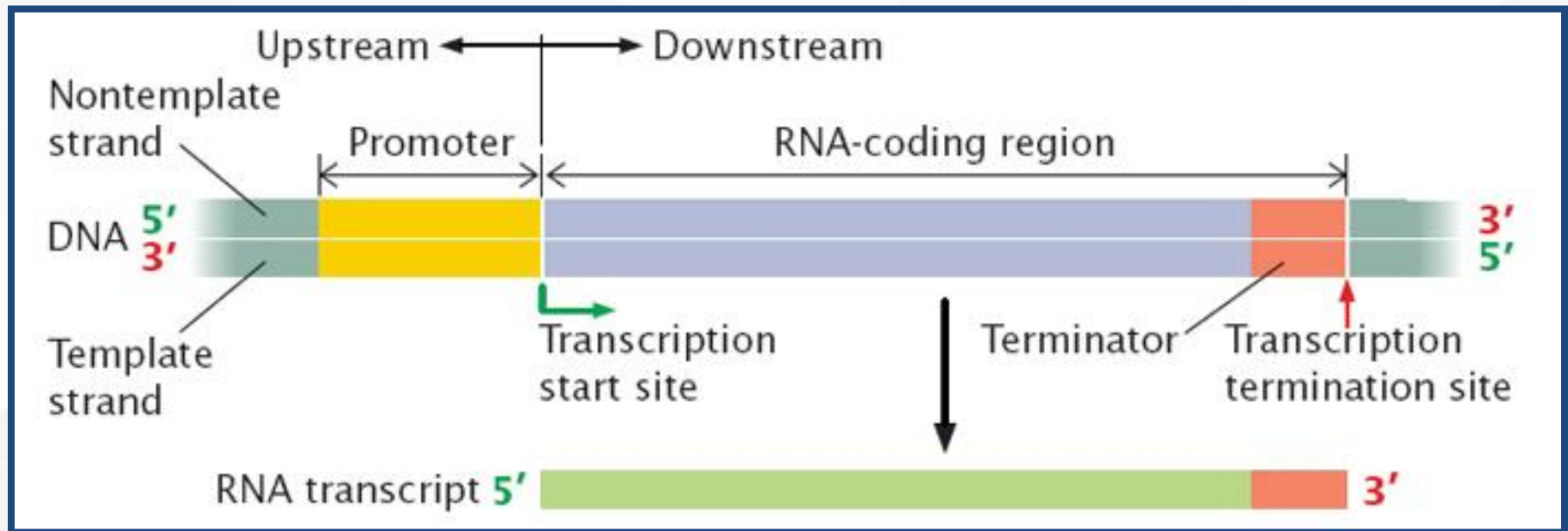
酶	位置	产物	活性比较	对 α -鹅膏蕈碱的敏感性
RNA聚合酶 I	核仁	rRNA	50-70%	不敏感
RNA聚合酶 II	核质	hnRNA	20-40%	敏感
RNA聚合酶 III	核质	小分子RNA, tRNA	10%	有种属特异性

2. 与转录相关的DNA结构 - 转录单位

25

□ 从启动子 (promoter)到终止子(terminator)称为**转录单位 (transcription unit)**;

- ① 启动子：调控转录起始的DNA序列
- ② 转录起始位点：开始转录的位点，多为A或G
- ③ 终止子：终止转录的DNA信号，多为回文序列



2. 与转录相关的DNA结构 - 转录单位

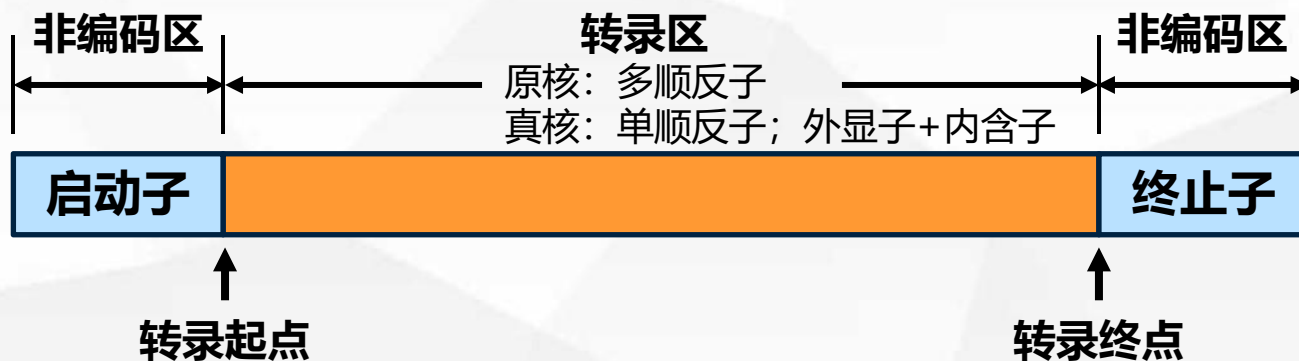
26

① 启动子

□ **Promoter**: 是指**DNA分子上**被RNA聚合酶识别并结合形成起始转录复合物的区域，具有方向性；它还包括一些调节蛋白因子的结合。

② 终止子

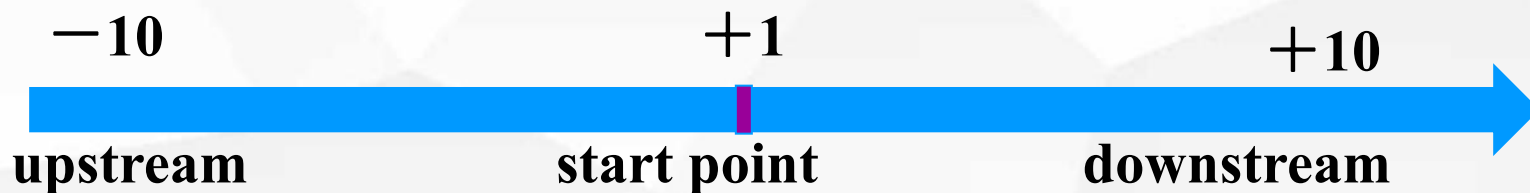
□ **Terminator**: 给予RNA聚合酶转录终止信号的**DNA序列**。



非翻译区?

③ 转录起始位点

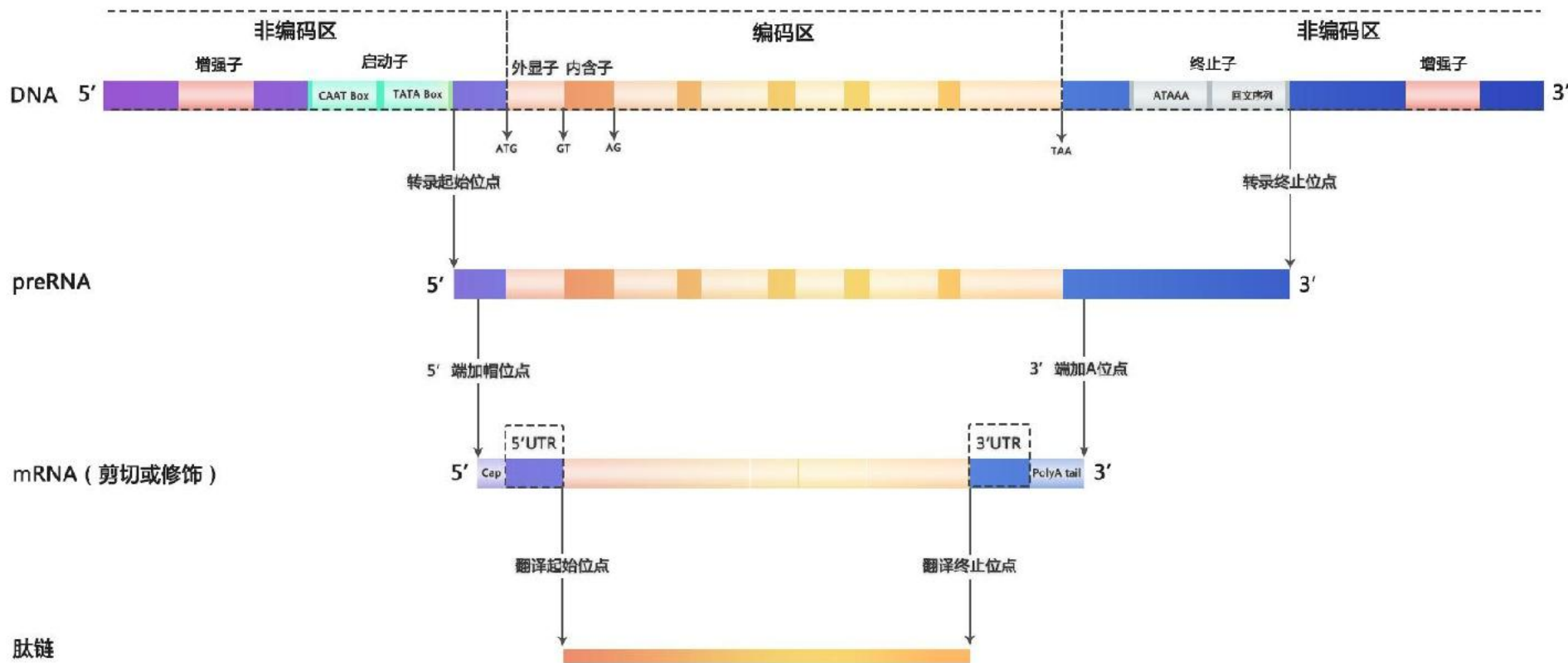
- transcription start site (TSS): 是指一个基因的5'端转录的第一个碱基。
 - 它是与新生RNA链第一个核苷酸相对应DNA链上的碱基，通常为**一个嘌呤 (A或G)**；
 - 通常把转录起始位点前即5'末端的序列称为上游，而把其后即3'末端的序列称为下游。
- **转录起点即转录原点记为 + 1，其上游记为负值，下游记为正值。**





真核生物编码基因的结构

28

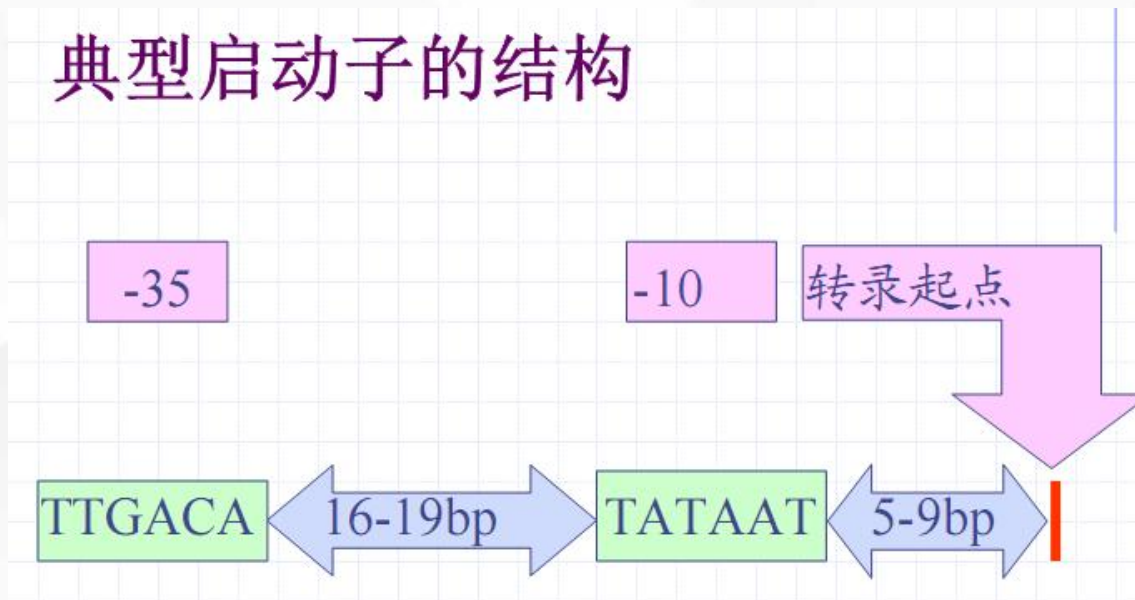


2.1.1 原核生物的启动子

A. 原核生物的启动子:

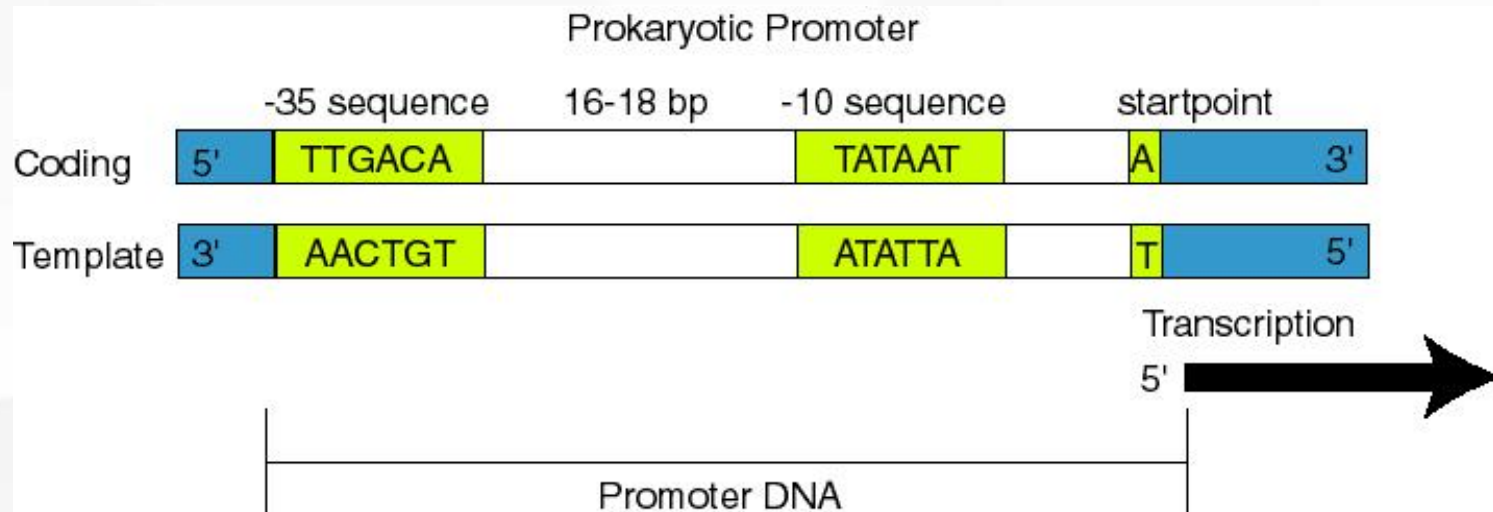
- 位于-70~+30区, 长度40~70 bp
- 含有三段保守序列, 具有高度保守性和一致性:
 - Sextama盒 (-35) } 核心启动子 (core promoter)
 - Pribnow盒 (-10) }
 - 转录起始位点 (+1)

典型启动子的结构



2.1.1 原核生物的启动子

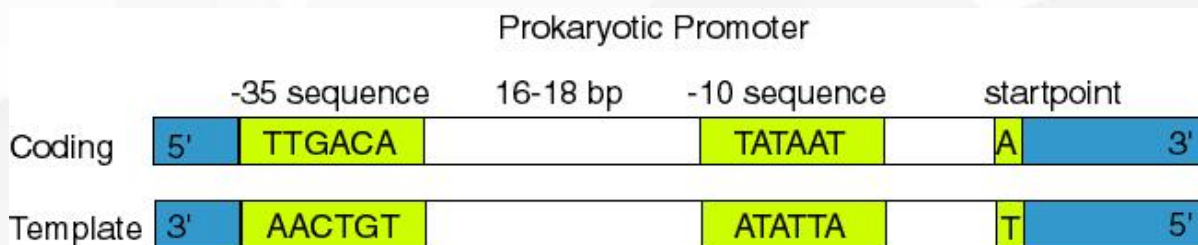
- 原核生物转录启动的4个关键区域：
 - 转录起始点：多数情况下为嘌呤，A为转录起始点
 - -10区：Pribnow盒
 - -35区：Sextama盒
 - -10与-35之间的序列



2.1.1 原核生物的启动子

① -35区/Sextama盒:

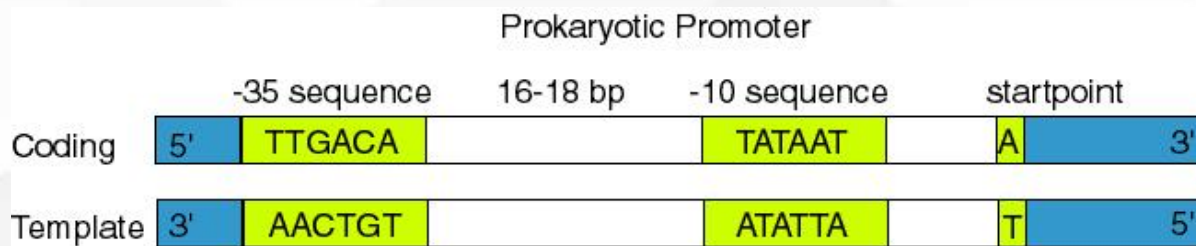
- 其保守序列为**TTGACA**，中心位于-35核苷酸处
 - 大多数启动子中共有序列为： $T_{82}T_{84}G_{78}A_{65}C_{54}A_{45}$
 - 重要性：很大程度上决定了**启动子的强度**
 - 位置在不同启动子中略有变动
- RNA聚合酶的 **σ 因子**可以识别并初始结合该位点，所以又称为**RNA聚合酶识别/辨认位点**。
- RNA聚合酶全酶的**松弛（初始）结合位点**



2.1.1 原核生物的启动子

② -10区/Pribnow盒/TATA盒:

- 其一致序列为**TATAAT**，中心位于TSS上游约-10处；
 - 一致序列：T₈₀A₉₅T₄₅A₆₀A₅₀T₉₆
 - 几乎在目前已知的所有启动子中均存在
 - 位置范围：-4到-13
- RNA聚合酶依靠σ因子识别的牢固结合位点，因而又称**RNA聚合酶(核心酶)的结合位点**；在RNA聚合酶的作用下首先解链。



» 2.1.1 原核生物的启动子

33

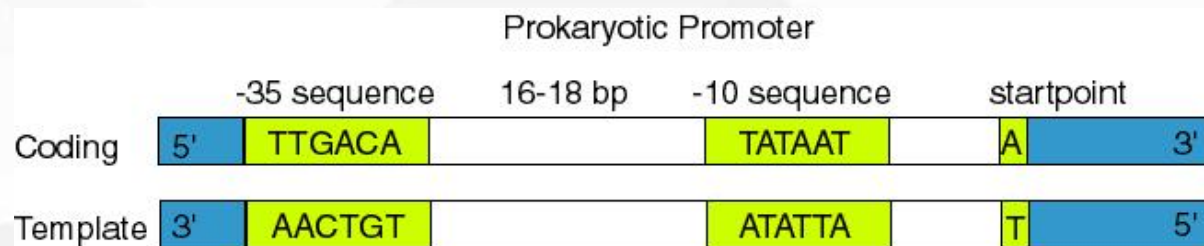
- -10区序列对转录的效率影响：
 - $T_{80}A_{95}T_{45}A_{60}A_{50}T_{96}$
 - **TATAAT**→**AATAAT**，转录效率下降，称为下降突变 (down mutation) ；
 - 可能由于的TA堆集能要小于→AA的堆积能，所以突变后双链比突变前要多消耗能量，所以影响转录效率；
 - **TATGTT**→**TATATT**，转录效率上升，为上升突变 (up mutation) ；
 - 上升突变可能是由于TG→TA不仅堆集能降低了，而且氢键也减少了，所以比突变前更易打开双链，转录效率也会提高。

» 2.1.1 原核生物的启动子

34

③ -10区与-35区之间的距离：

- -35区和-10区之间的距离在绝大多数原核生物启动子为16到18 bp。该区域的碱基序列并不重要，但该**距离的长短**是至关重要的；
- 适宜的距离可以为RNA聚合酶提供合适的空间结构，便于转录的起始。



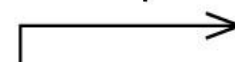
➤➤ -10区与-35区之间的距离为RNA聚合酶提供合适的空间结构

RNA Polymerase

σ

- 10 box

Transcription



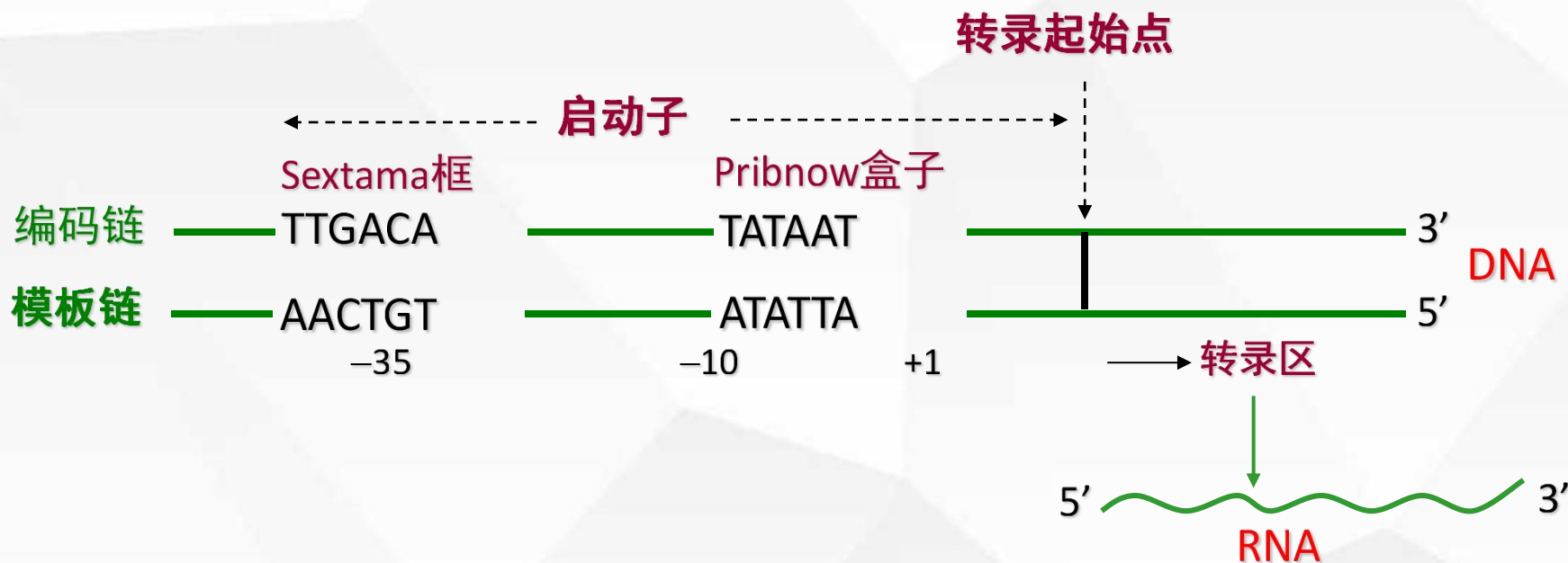
TTGACa
AACTGt

TAtAaT
ATaTtA

Unwound region

2.1.1 原核生物的启动子 – 小结

36



»» 2.1.2 真核生物的启动子

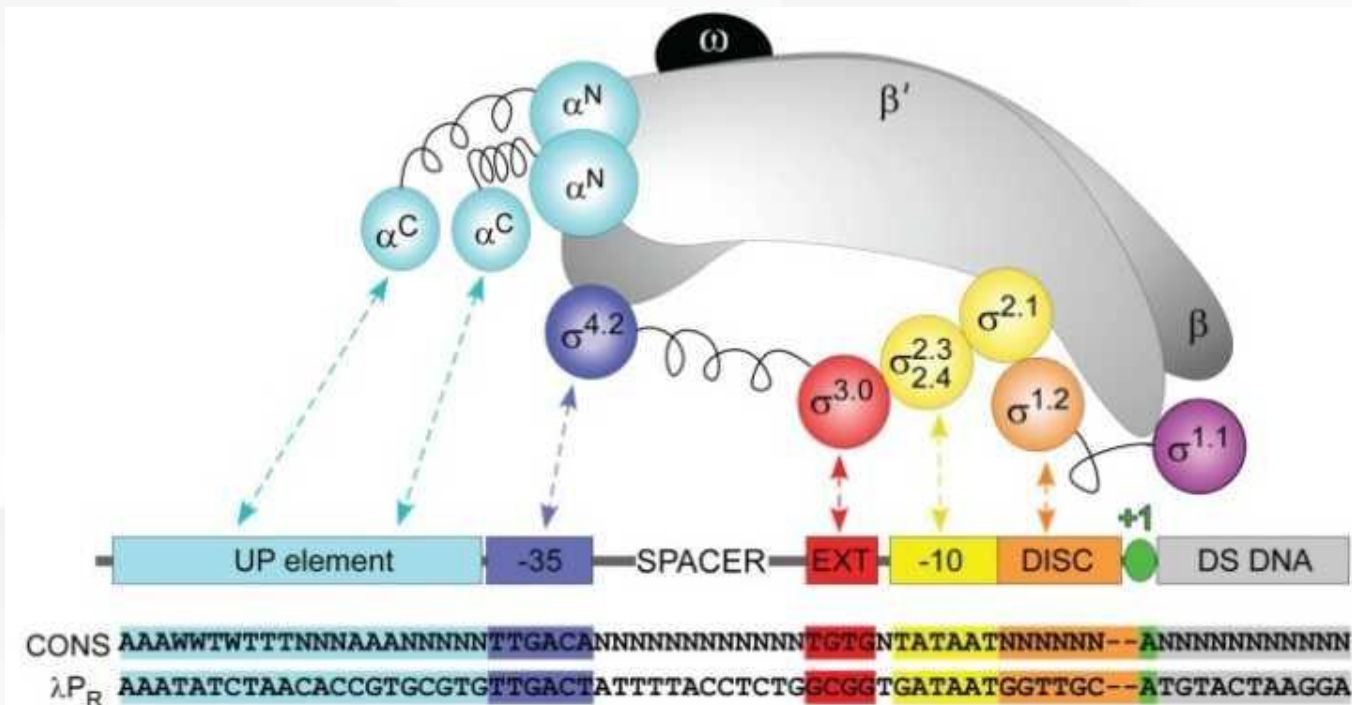
37

- 真核生物有三种RNA聚合酶，每一种都有自己的启动子类型。
 - RNApol I: 核仁 活性所占比例最大
转录rRNA (5.8S、18S、28S)
 - RNApol II: 核质
主要负责 hnRNA、snRNA 的转录
 - RNApol III: 核质
负责 tRNA、5S rRNA、Alu序列和部分 snRNA
- 目前在线粒体和叶绿体内发现少数 RNApol (活性较低)
 - 线粒体的RNA聚合酶，催化合成线粒体mRNA、tRNA和rRNA;

真核生物RNA聚合酶的结构

38

- 人的3种细胞核RNA聚合酶都含2个大亚基（如RNA聚合酶II的RPB1和RPB2）、2个类 α 亚基和1个类 ω 亚基，分别与大肠杆菌核心酶的 β' 和 β 、2个 α 亚基和 ω 亚基同源。
- RNA聚合酶II的大亚基RPB1含有C端结构域（CTD），参与RNA合成、后加工、转运的调控。在启动转录时，它必须保持去磷酸化状态；然而转录一旦启动，它必须被磷酸化，才能使转录进入延伸阶段。



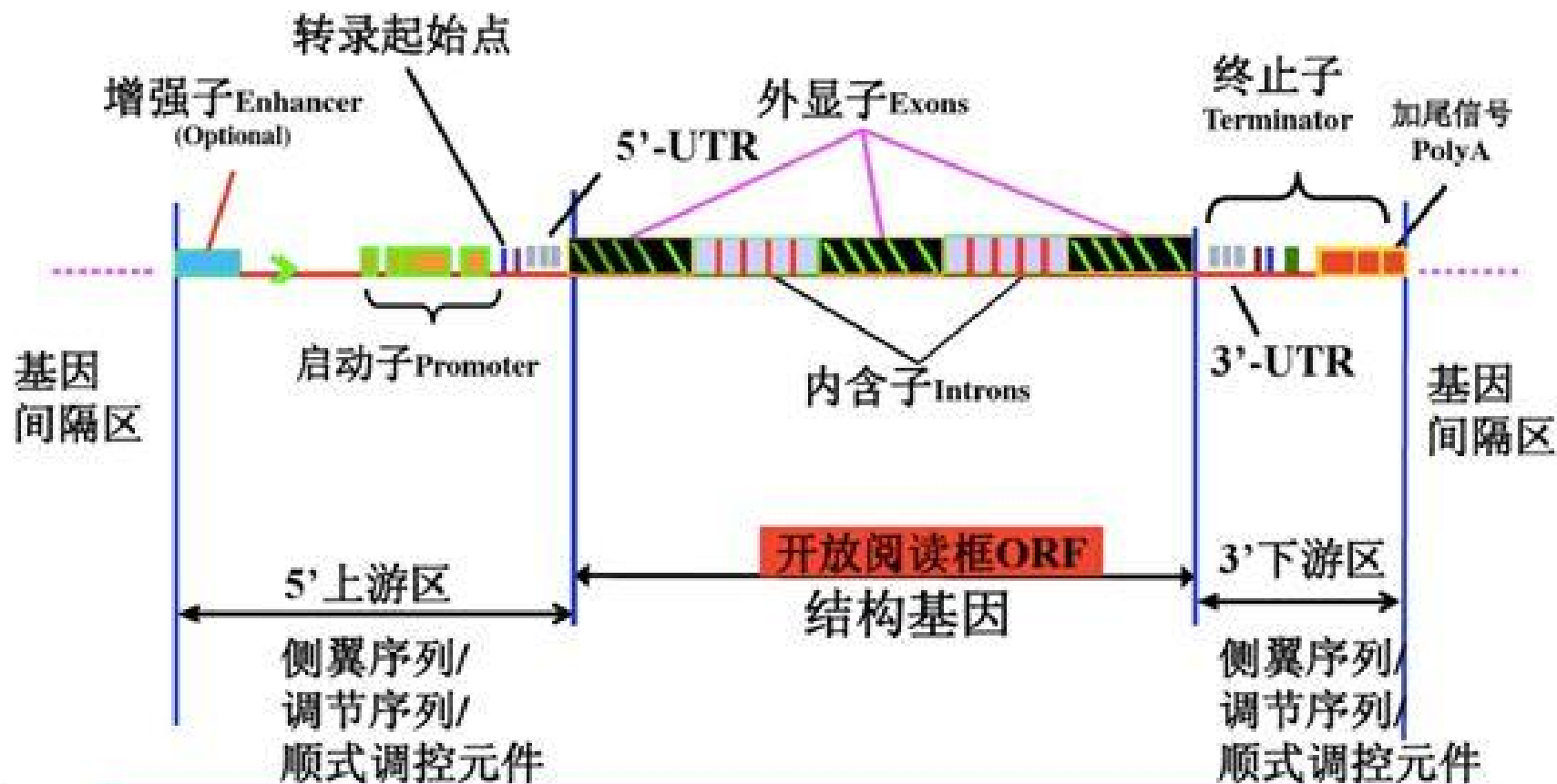
»» 2.1.2 真核生物的启动子

□ 顺式作用元件(cis-acting element):

- 基因旁侧**DNA序列**，如启动子、增强子、调控序列和可诱导元件等
- 作用：参与基因表达调控
- 本身不编码任何蛋白质，仅提供一个作用位点，要与反式作用因子相互作用而起作用。

□ 反式作用因子(trans-acting factor):

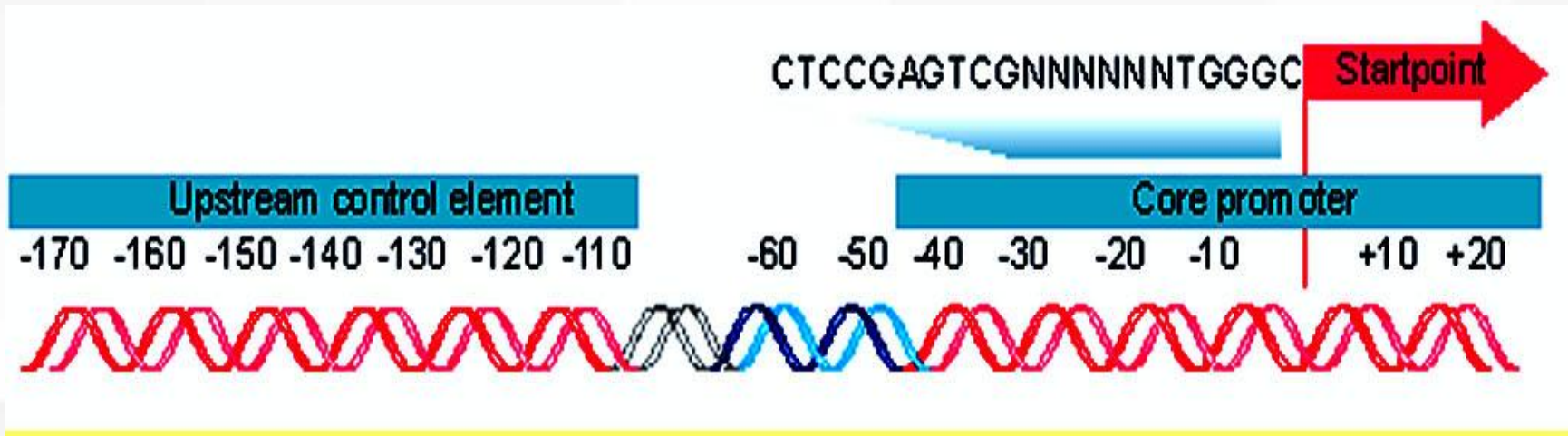
- 能直接或间接识别，结合转录上游区段DNA的**蛋白质**
 - 其中，直接或间接结合RNA聚合酶的，则称**转录因子** (transcriptional factor, TF);
 - 相应于RNA pol I、II、III的TF，分别称为TFI、TFII、TFIII。
- **TATA结合蛋白 (TBP)**：是唯一能识别并结合TATA盒的转录因子。



2.1.2.1 RNA聚合酶I 识别的启动子

41

- RNA聚合酶 I 的启动子主要由两部分组成的，其结构为：
 - **核心启动子** (core promoter) 位于 - 45至 + 20的区域内，足以使转录起始。
 - **上游调控元件** (upstream control element) 位于 - 180至 - 107，可提高转录起始效率。



»» 2.1.2.1 RNA聚合酶I 识别的启动子

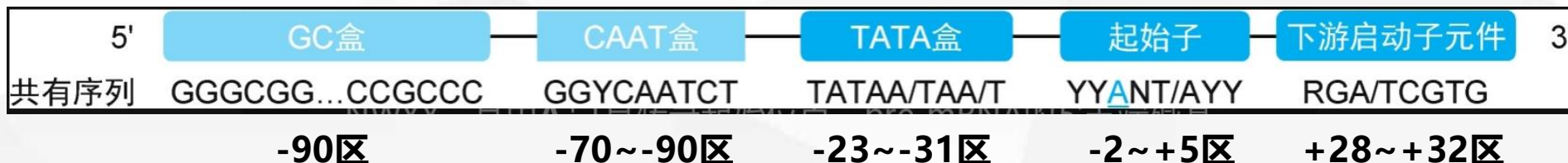
42

- RNA聚合酶I 活性最高
- 细胞内能促进rRNA基因转录的因素：
 - 专一性的聚合酶I启动子：无竞争；
 - rDNA串联重复基因组织位于很小的核仁区：增加启动子浓度；
 - 串联重复基因：使聚合酶I在转录第二个基因前不需活化；
 - 转录起始因子牢固地结合于DNA上。

2.1.2.2 RNA聚合酶II 识别的启动子

43

- RNA聚合酶II主要负责hnRNA和部分核内小rRNA (snRNA) 的基因转录，其启动子结构最为复杂。
- RNA聚合酶II单独并不能起始转录，必须和其他的辅助因子共同作用形成转录起始复合物才能起始转录。
- 包含两类元件：
 - 核心启动子元件 (core promoter element, CPE) : -40~60nt, 包括TATA盒、起始子、下游启动子元件(DPE) 等, 位于转录起始位点的上游和下游, 功能是确定转录起始位点。
 - 上游调控元件 (upstream promoter element, UPE) 包括GC盒、CAAT盒, 功能是控制转录启动效率。



2.1.2.2 RNA聚合酶II 识别的启动子

44

- **结构最复杂；大多数位于转录起始点上游，多个短序列元件：**

I. 起始子 (initiator, Inr) : 以TSS为中心

- 位于 - 2~+5处；
- 哺乳动物共有序列是YY A⁺1 NWY (A⁺1是转录起始位点) ；
- pre-mRNA的5'末端碱基通常是嘌呤，特别是腺嘌呤；
- 起始子是通用转录因子TF II D特定TAF亚基的识别结合位点。

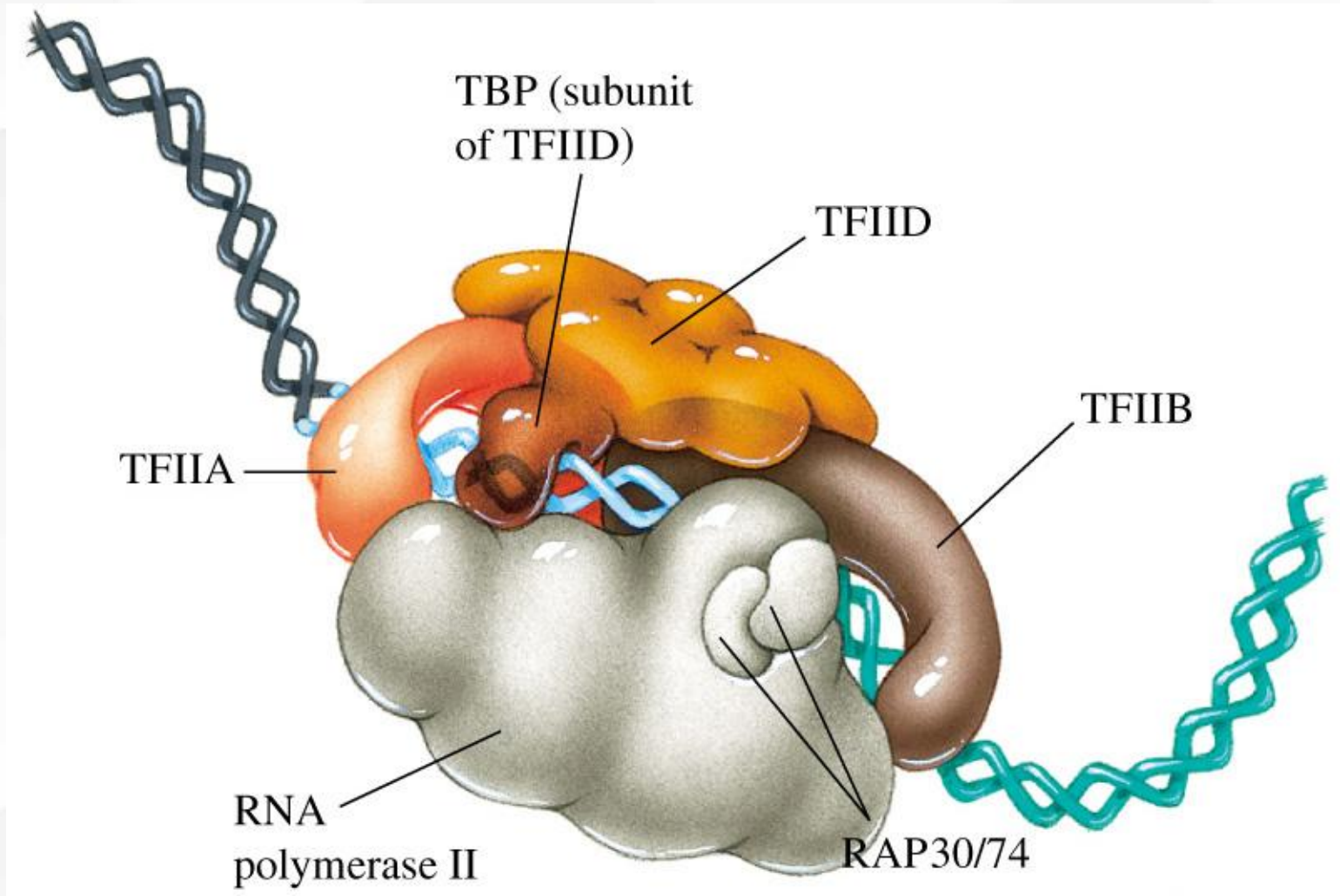
II. TATA框 (又称Hogness盒):

- 位于 - 25~-31处 (酵母TATA盒位于-90区) ；
- 共有序列是TATAAAA；
- 是转录因子TBP (TATA结合蛋白) 的识别结合位点；
- 定位转录起始点的功能 (类似原核的Pribnow框) ；
- TATA是绝大多数真核基因的正确表达所必需的，常与起始子共存。



2.1.2.2 RNA聚合酶II 识别的启动子

45



III. CAAT框 (CAAT box) :

- 位于 - 70~-90处;
- 一致序列为GGC/TCAATCT;
- 是转录因子C/EBP、CTF/NF1、NF-Y的结合位点;
- 前两个 G 的作用十分重要 (转录效率);
- 影响启动子的效率、频率, 不影响启动子的特异性 (距转录起始点的距离, 正反方向)。

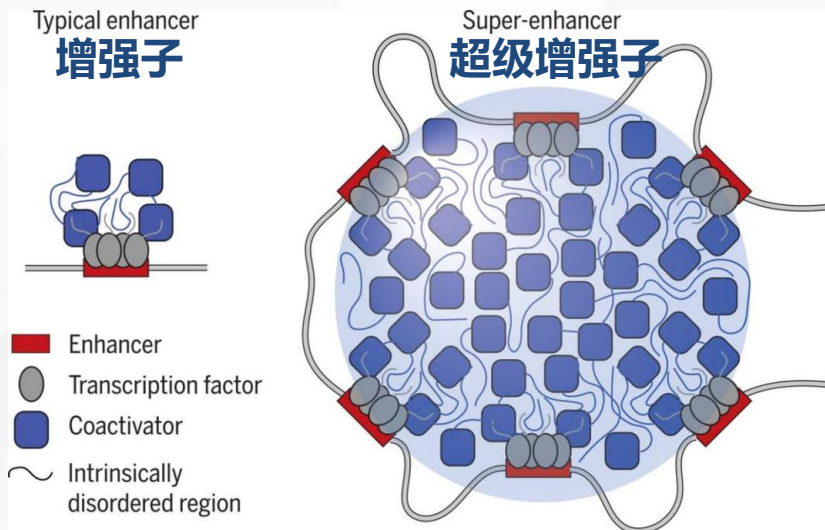
IV. GC框 (GC box) :

- 位于 - 90附近, 较常见的成分
- 包含两段共有序列: GGGCGG 和CCGCCC (互为反向重复序列)
- 是转录因子Sp1的结合位点;
- 作用是控制转录启动效率。



V. 增强子 (enhancer) :

- 增强效应十分显著
- 增强效应与其位置和取向无关
- 大多为重复序列
- 增强效应有严密的组织特性和细胞特异性
- 没有基因专一性
- 许多增强子还受外部信号的调控

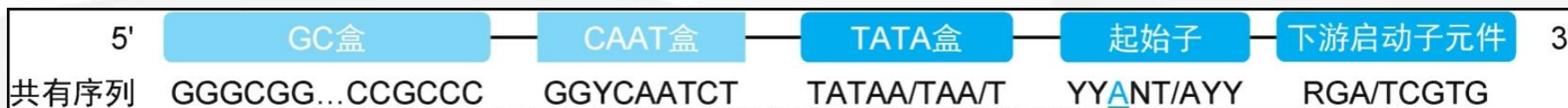


VI. 其他元件:

- 八聚核苷酸元件 (octamer element OCT元件) : 一致序列为 ATTTGCAT
- KB元件: 一致序列为 GGGACTTTCC
- ATF元件: 一致序列为 GTGACGT
- 还有一些位于起始点下游的相关元件。

VII. 不同元件的组合情况:

- **在不同的启动子中, 这些元件的组合情况是不同的。**
- 例如:
 - 蛋白基因启动子中, 30%只含起始子, 30%含起始子+TATA盒, 25%含起始子+DPE, 15%含起始子+TATA盒+DPE;
 - 猿猴空泡病毒40 (SV40) 的早期启动子中有6个GC框。





在H2B的启动子中:

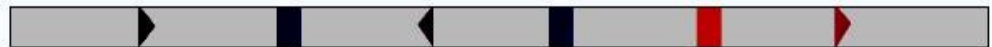
- 2个OTC元件
- 2个CAAT框
- 1个TATA框

Figure 20.17 Promoters contain different combinations of TATA boxes, CAAT boxes, GC boxes, and other elements.

SV40 early



Thymidine kinase



Histone H2B



-140 -120 -100 -80 -60 -40 -20



Types of module

▶
Octamer

◀
CAAT

■
GC

■
TATA

» 2.1.2.3 RNA聚合酶III 识别的启动子

50

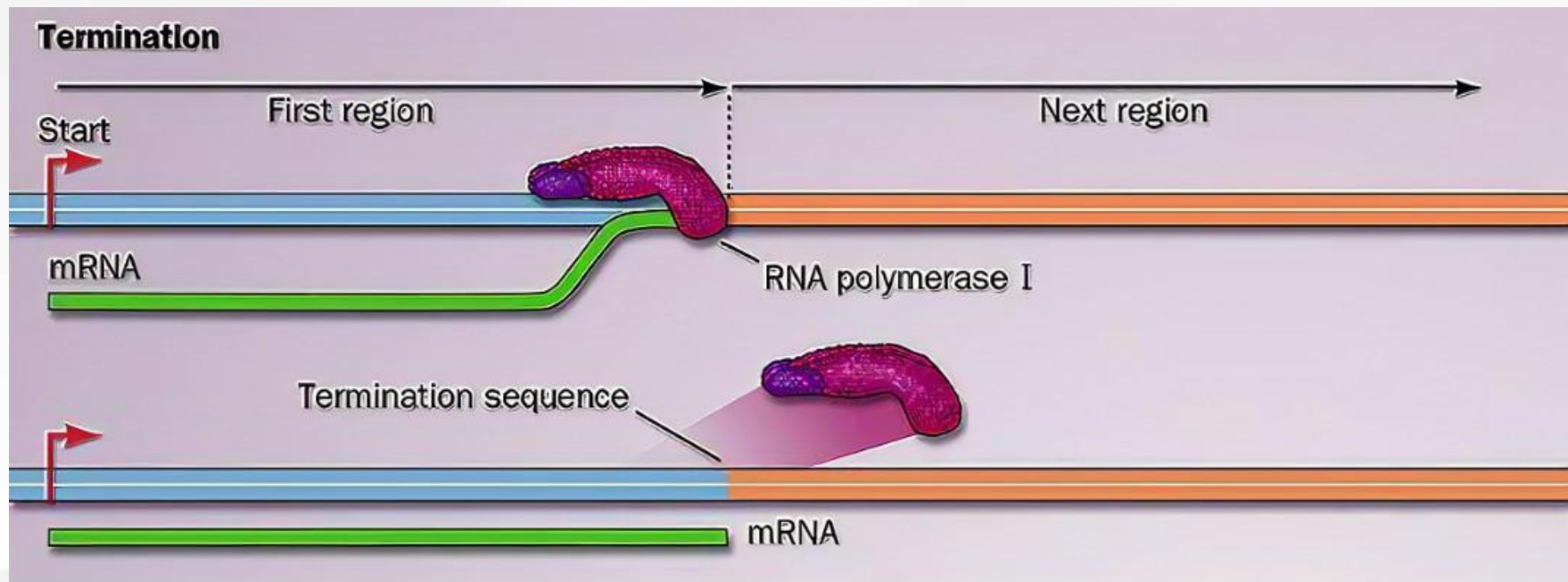
- **RNApol III启动子的结构:**
 - RNA聚合酶III转录5S rRNA、tRNA和部分snRNA。这三种启动子的结构是不同的;
 - RNA聚合酶III也**必须和其他的辅助因子**共同作用, 才能识别不同的启动子。

- 原核和真核生物启动子结构异同?

» 2.2 终止子

52

- **终止子 (Terminator) : 给予RNA聚合酶转录终止信号的DNA序列。**
- **终止子位于转录区下游，最后才被转录，所以编码RNA前体的3'端。**



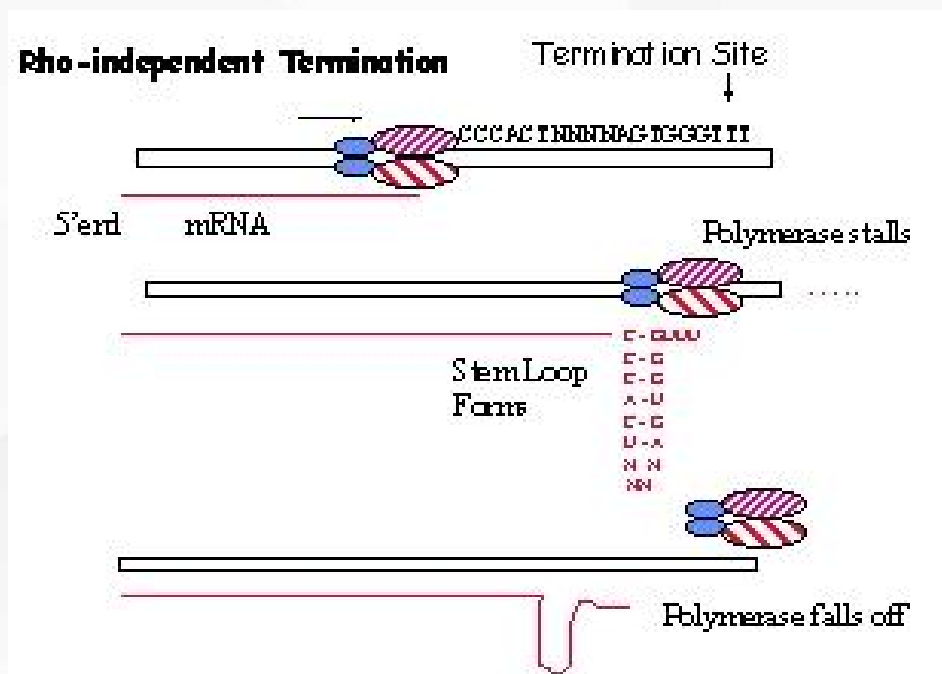
»» 2.2.1 原核生物的终止子

53

- 分为两类：
 - 一类是**不依赖 ρ 因子的终止子**/简单终止子 **(强)**
 - 另一类是**依赖 ρ 因子的终止子** **(弱)**
 - ρ 因子 (Rho) 是一种蛋白性终止因子;
 - 高度保守的终止因子, 存在于几乎所有的原核生物。

A. 不依赖 ρ 因子的终止子

- 启动子由DNA序列来提供信号，但真正起终止作用的不是DNA序列本身，而是转录生成的RNA。
- 这种终止方式也称为内部终止 (intrininsic termination)
- 大肠杆菌中的大多数基因采用内部终止。



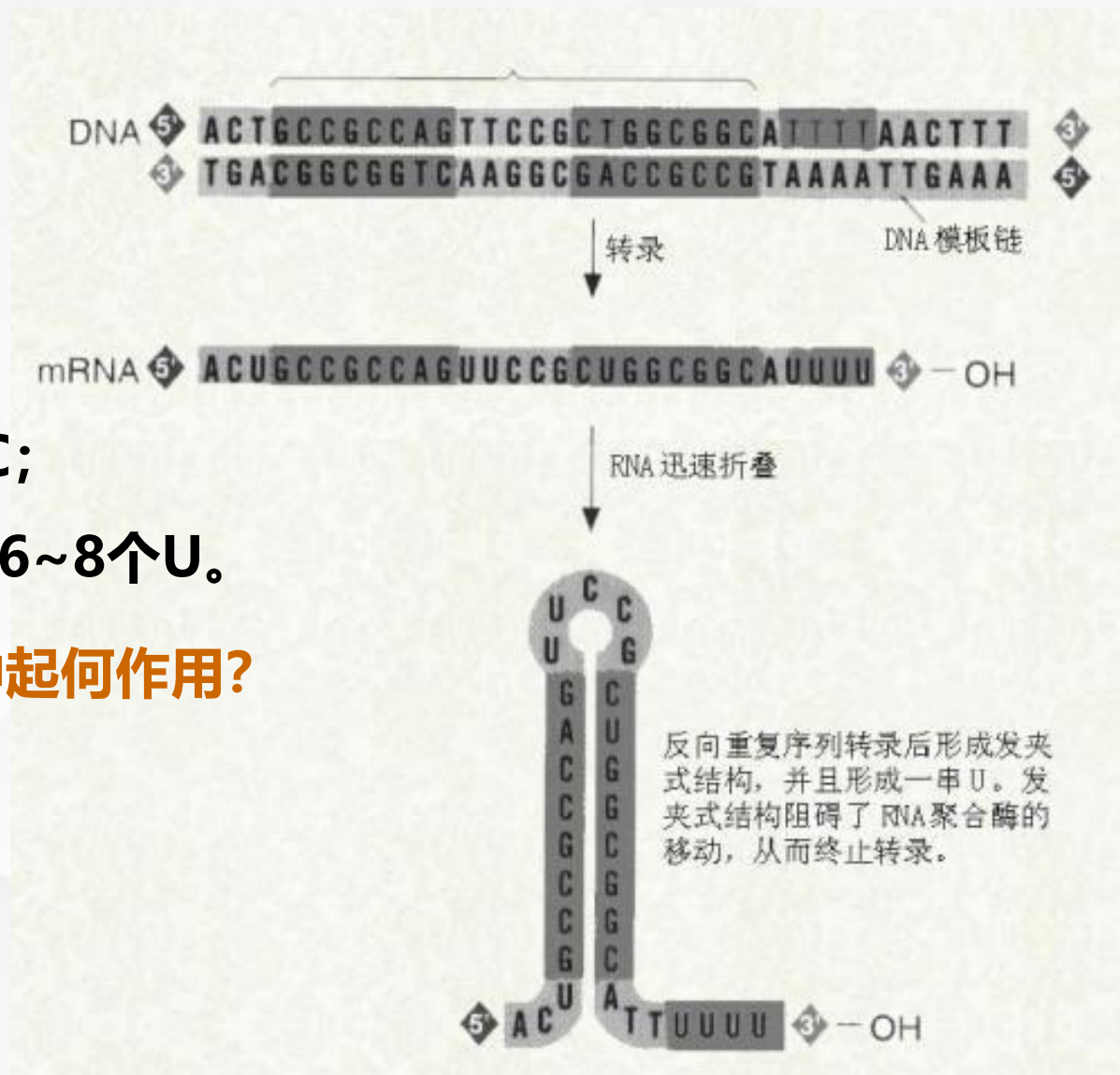
2.2.1 原核生物的终止子

55

● 强终止子的结构特点：

- 有回文结构存在；
- 茎的区域内富含G-C；
- 强终止子3' 端上有6~8个U。

以上结构特点在终止中起何作用？



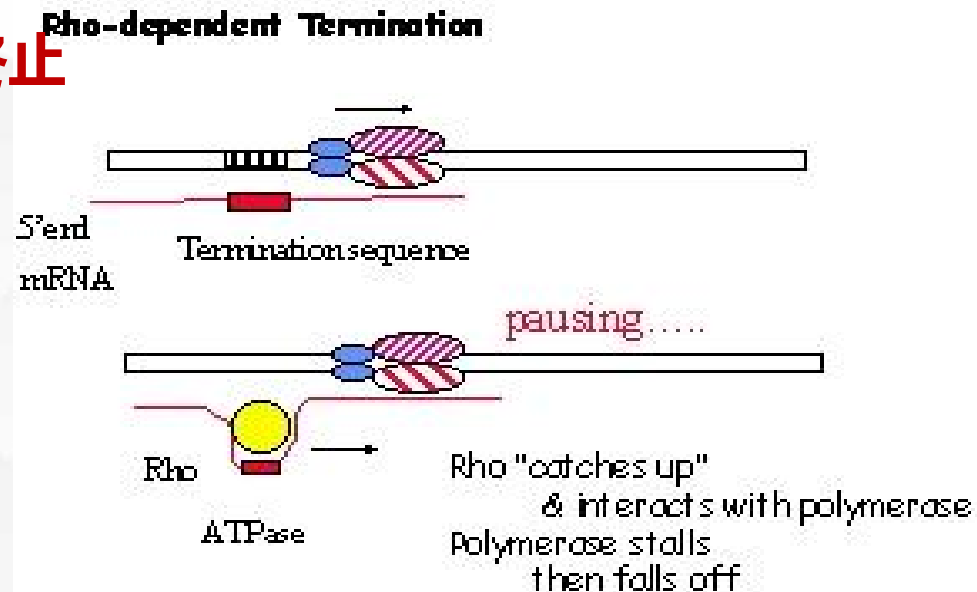
2.2.1 原核生物的终止子

56

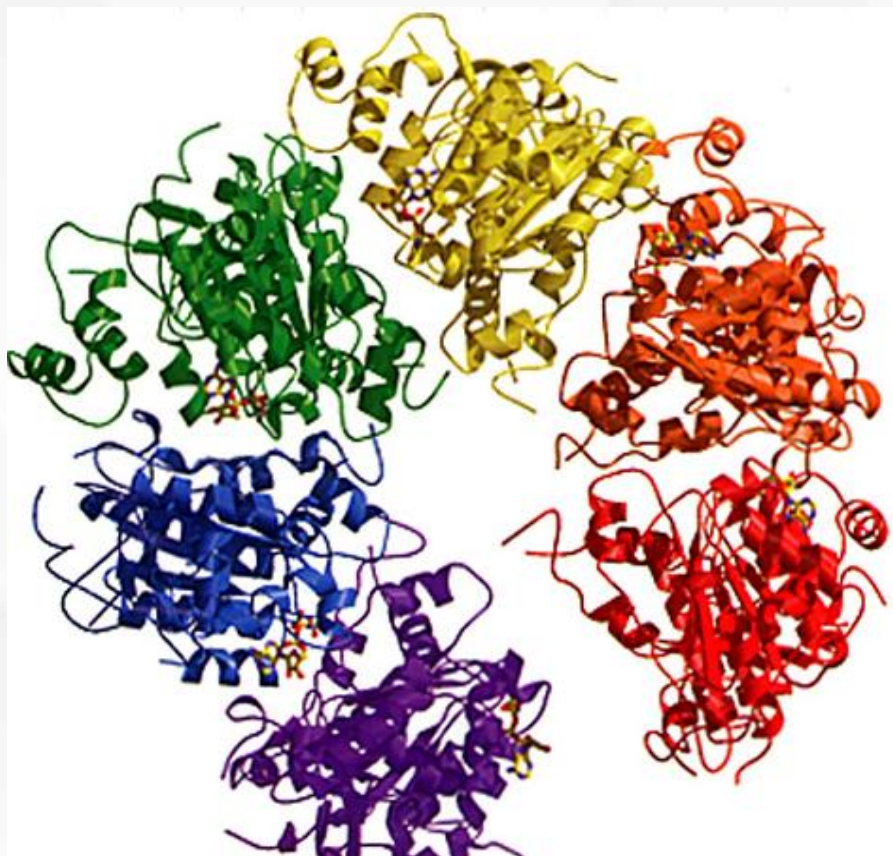
B. 依赖 ρ 因子的终止子

- ρ 因子依赖的终止大约占20-30%
- **结构特点:**
 - 反向重复序列中的 G / C 对含量较少;
 - 发夹结构末端没有固定特征。

- **依靠与 ρ 的共同作用而实现终止**

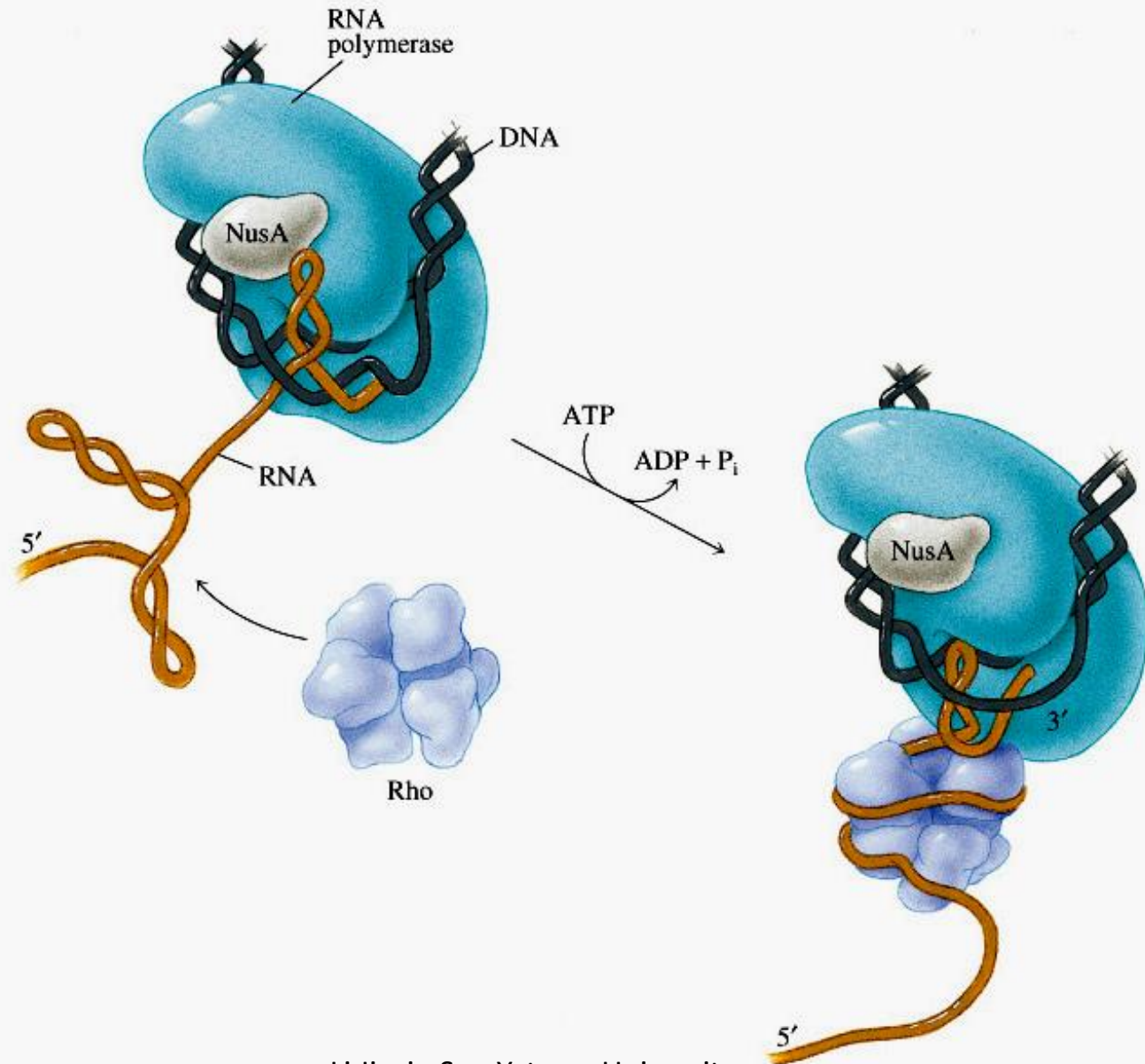


- ρ因子是一个同源六体蛋白，具有ATP酶和解链酶的活性。
 - ATP酶活性：提供能量使ρ因子追赶RNA聚合酶，到达终止子区
 - 解链酶活性：能够催化RNA/DNA和RNA/RNA双螺旋的水解。



大肠杆菌中依赖 ρ 因子的转录终止

58





»» 2.2.2 真核生物的终止子

60

- 真核生物的RNA聚合酶有3种，其终止方式也有所不同：
 - **RNA Pol I**的终止需要**转录终止因子TTF-1**与rRNA基因下游的终止子结合，导致聚合酶暂停。然后由PTRF（Pol I和转录物**释放因子**）介导转录复合物的解离。
 - **RNA Pol II**的终止研究最多，它主要用**加尾信号**作为转录终止信号。当mRNA中转录出聚腺苷酸化信号AAUAAA后，会募集一系列蛋白因子，切割mRNA并添加poly A尾，然后才释放RNAP，转录终止。
 - **RNA Pol III**转录的基因都比较小，所以终止子相对简单。其模板链中有一段oligo (dA)（多聚A），转录出**多聚U**后二者的结合较弱，使复合物不稳定。比较特殊的是，其非模板链的oligo (dT) 也会参与终止过程，可以促进聚合酶暂停和转录本释放等。

- 转录和复制的异同点
- 转录单位?
- 原核和真核生物启动子结构异同?
- 强终止子的结构与功能关系如何对应?