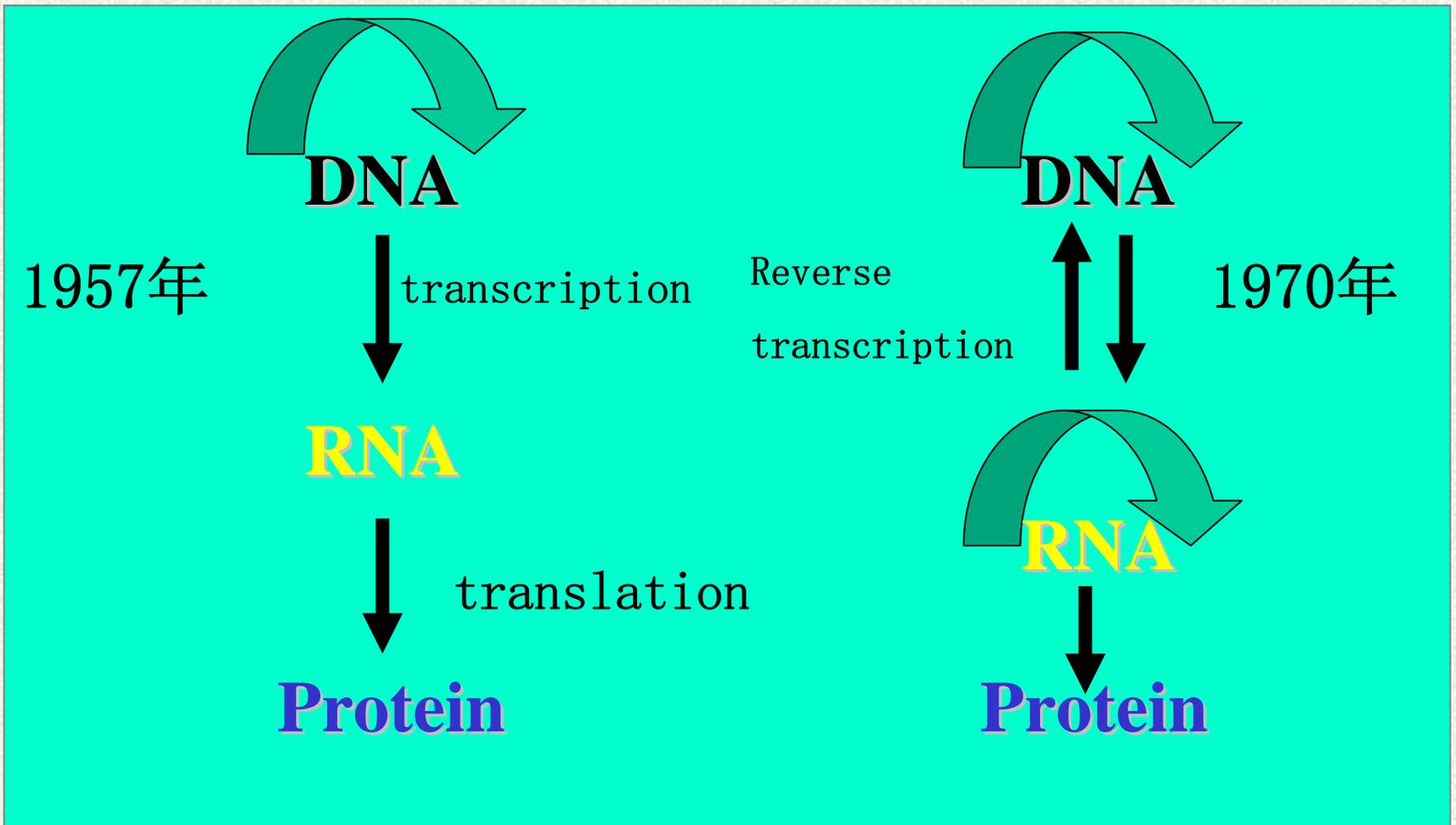




第二讲 染色体与DNA



Crick的中心法则 (central dogma)



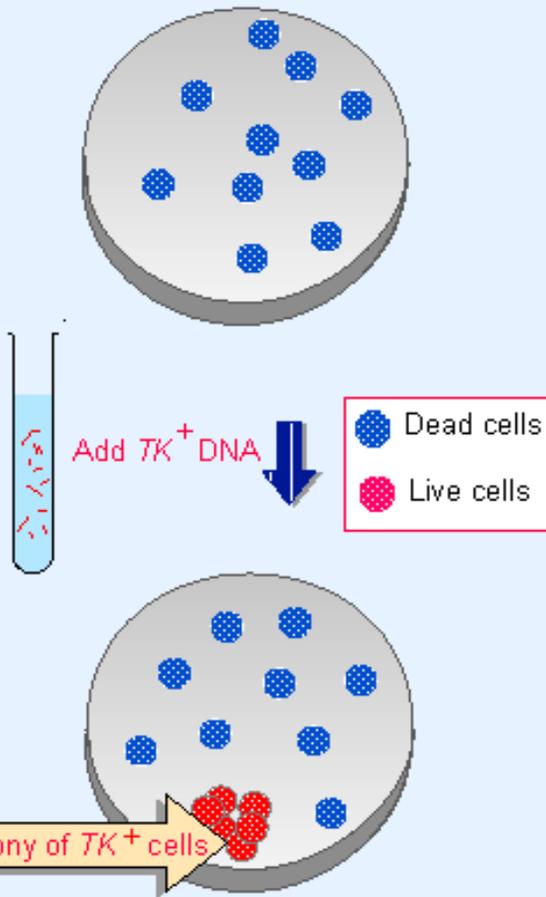


主要内容:

- 1、染色体
- 2、DNA的结构
- 3、DNA的复制
- 4、原核和真核生物DNA的复制特点
- 5、DNA的修复
- 6、DNA的转座



Cells that lack *TK* gene cannot produce thymidine kinase and die in absence of thymidine



Some cells take up *TK* gene; descendants of transfected cell pile up into a colony

DNA is the genetic material

1. Griffith于1928年发现的“转化”现象对于认识核酸具有重要作用。1952年，Hershey和Chased的噬菌体侵染细菌的实验等证明细胞与噬菌体的遗传物质都是DNA。

DNA不仅是真核生物的遗传物质，而且能够在不同物种间相互转移并保持功能活性。

Eukaryotic cells can acquire a new phenotype as the result of transfection by added DNA.

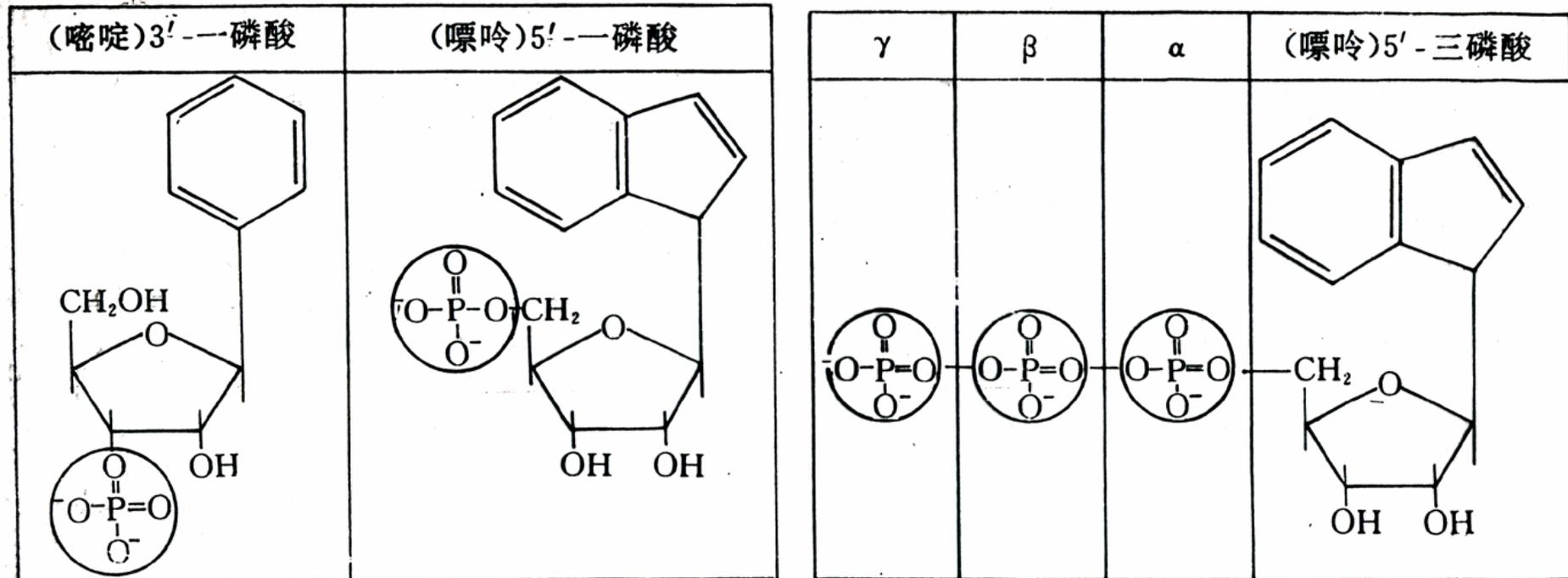


分子生物学研究已经证实，**DNA**控制了生物的性状遗传。

无论**DNA**或**RNA**，都是由许许多多多个**核苷酸**连接而成的生物大分子，而每个核苷酸又由**磷酸**、**核糖**和**碱基**3部分组成。



图2-1 核苷酸的组成与结构。核苷的5'位和3'位都可能与磷酸基团相连。





2.1 染色体

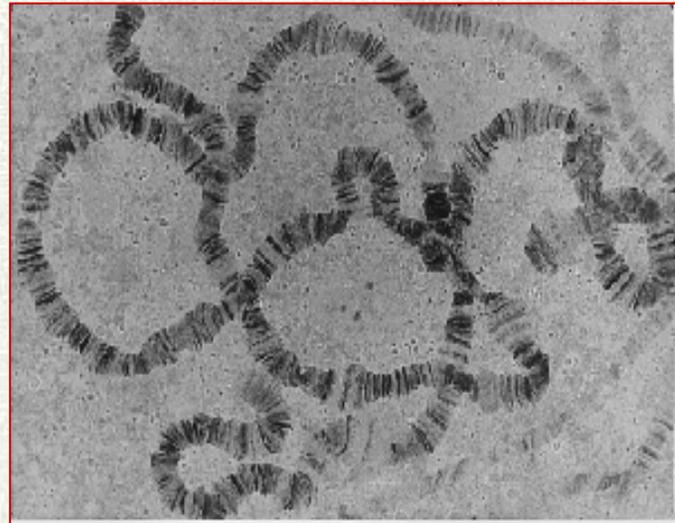
2.1.1 染色体—遗传物质的主要载体

染色体在遗传上起着主要作用，因为亲代能够将自己的遗传物质以**染色体（chromosome）**的形式传递给子代，保持了物种的稳定性和连续性。



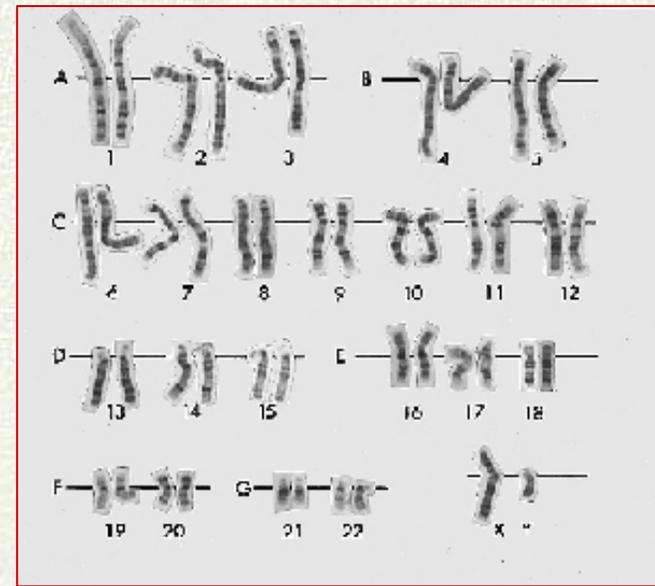
The **nucleus** of a nondividing cell (**interphase**) is filled with a **diffuse** material called **Chromatin**, so called because early microscopists found that it stained brightly with certain dyes.

摘自:Lehninger, 3rd, p35





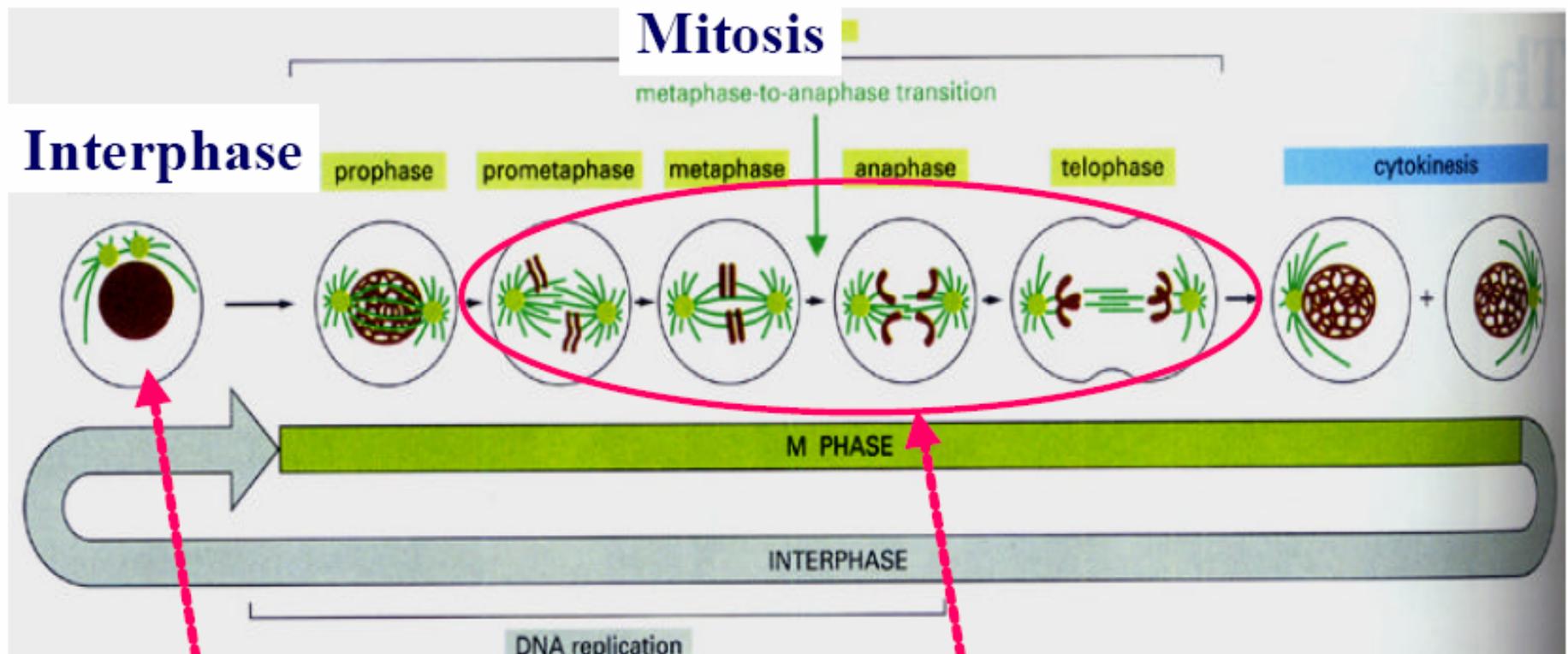
Chromatin consists of **DNA** and **proteins** bound tightly together and is the **substance of the chromosomes**, which do not condense and become individually visible until the cell is ready to divide.





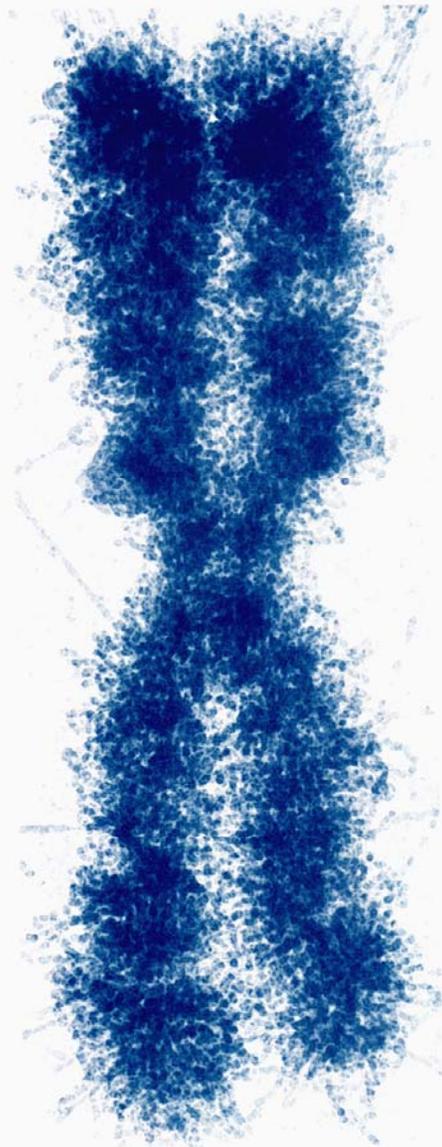
Nucleolus(核仁) is a specific region of the nucleus in which DNA is located. Before mitosis, the double-helical DNA of the **chromatin** is replicated, then in the first phase of mitosis the chromatin fibers condense into discrete bodies (**chromosomes**), each consisting of two identical **chromatids**.

In the first phase of mitosis the **chromatin** fibers condense into **chromosomes**.



Chromatin
(diffused)

Chromosome
(condensed)



**Chromosomes
consisting of two
identical
chromatids**



染色体包括DNA和蛋白质

两大部分。同一物种内每条染色体所带DNA的量是一定的，但不同染色体或不同物种之间变化很大，人X染色体有1.28亿个核苷酸对，而Y染色体只有0.19亿个核苷酸对。



染色体 1



染色体 2



染色体 3



染色体 4



染色体 5



染色体 6



染色体 7



染色体 8



染色体 9



染色体 10



染色体 11



染色体 12



染色体 13



染色体 14



染色体 15



染色体 16



染色体 17



染色体 18



染色体 19



染色体 20



染色体 21



染色体 22



染色体 X



染色体 Y

人类22对 染色体及 性染色体 显微结构 图



表2-1 人类基因组各条染色体中碱基对数量和推导的功能基因数量对照

染色体编号	长度 (Mbp)	估计的基因数
1	220	3453
2	240	2954
3	200	2427
4	186	1861
5	182	2136
6	172	2257
7	146	1831
8	146	1560
9	113	1537
10	130	1653
11	132	2185



染色体编号

长度 (Mbp)

估计的基因数

12	134	1861
13	99	1032
14	87	1283
15	80	1198
16	75	1421
17	78	1545
18	79	826
19	58	1675
20	61	986
21	33	449
22	36	835
X	128	1465
Y	19	210
总长	2907	39114



2.1.2 真核细胞染色体的组成

作为遗传物质，染色体具有如下特征：

- ①分子结构相对稳定；
- ②能够自我复制，使亲子代之间保持连续性；
- ③能够指导蛋白质的合成，从而控制整个生命过程；
- ④能够产生可遗传的变异。



着丝粒位置 英文名称 形 状

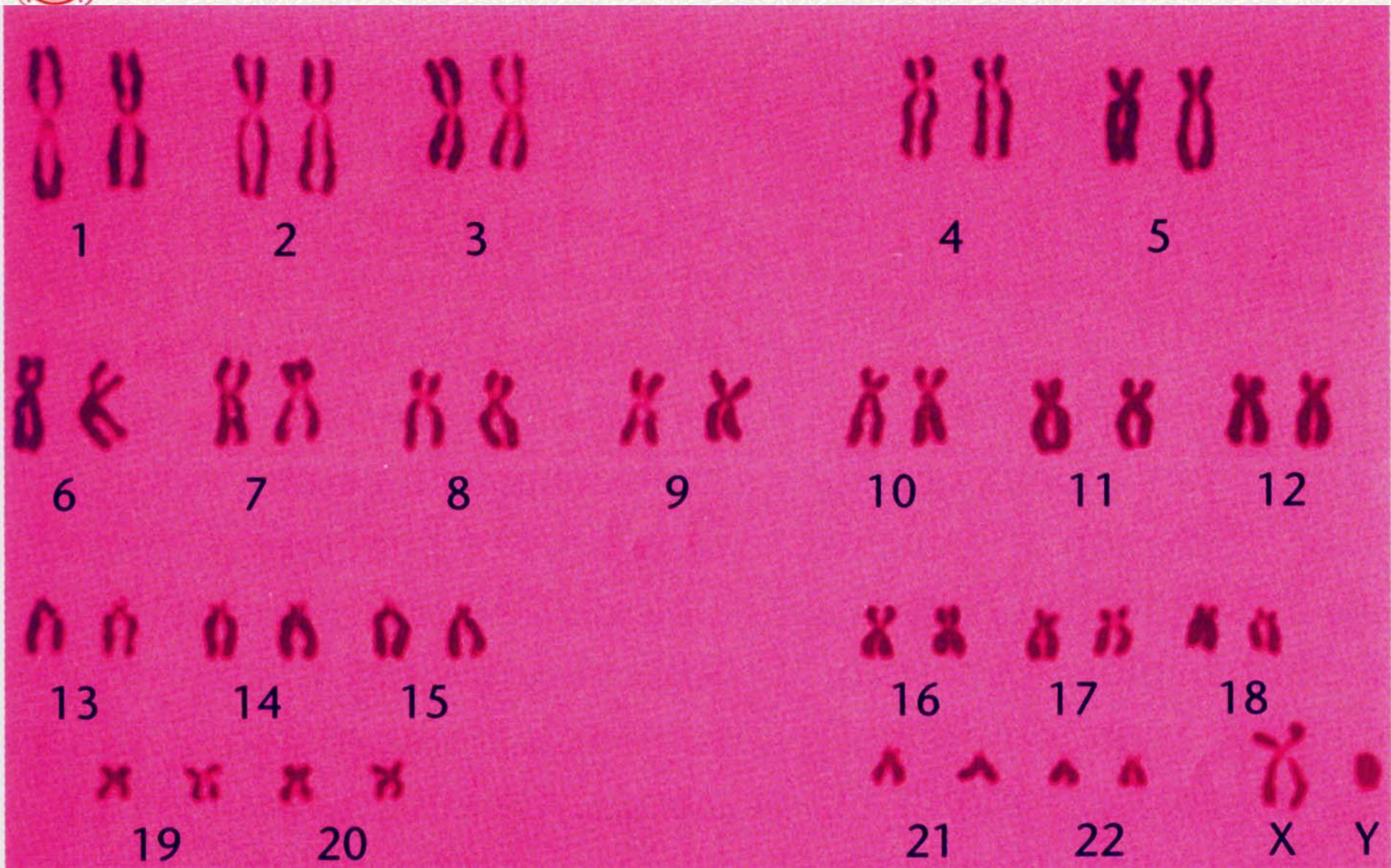
分裂中期 分裂后期

染色体的分类

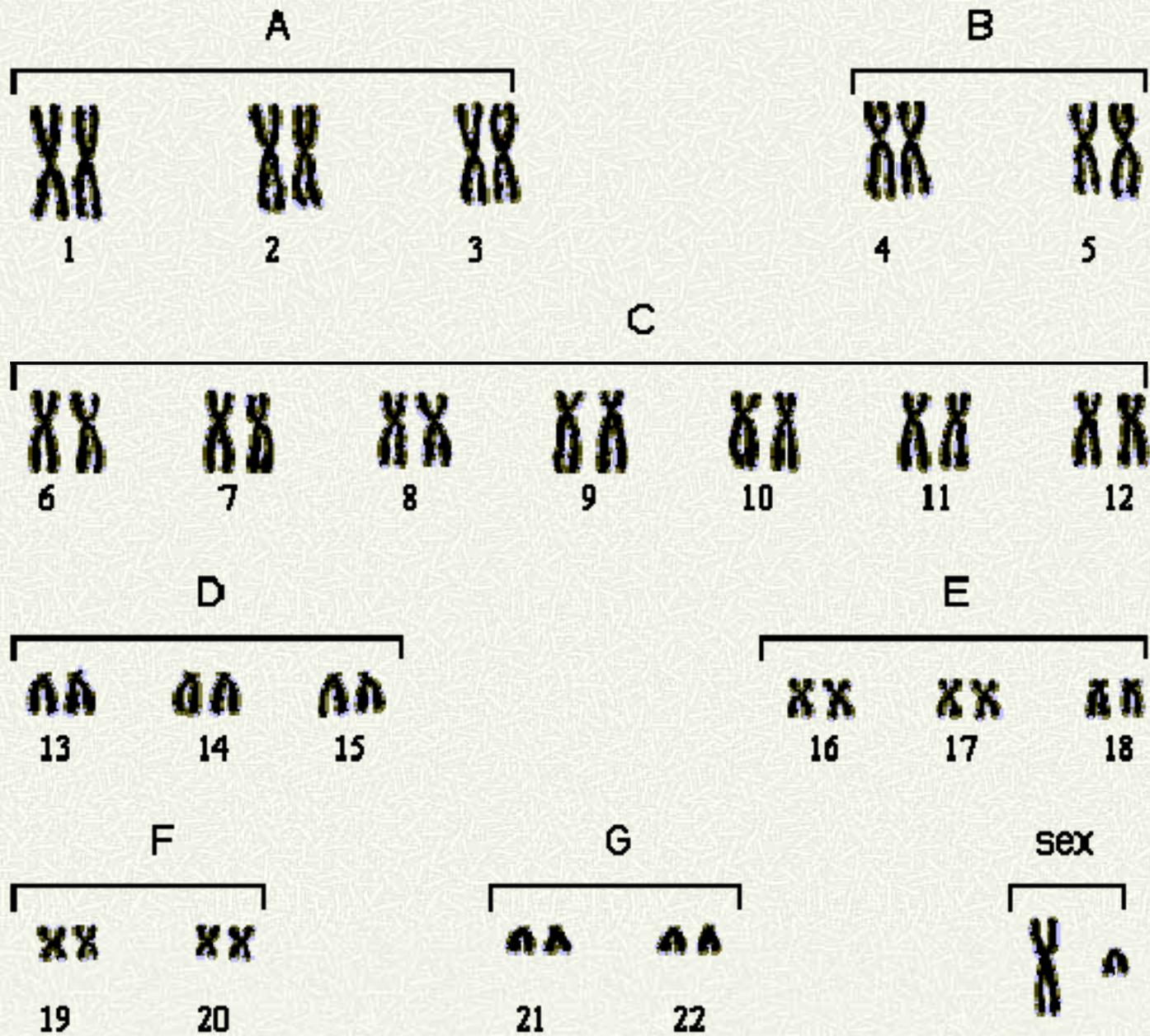
			分裂中期	分裂后期
中间	Metacentric	Centromere		
偏上	Submetacentric			
靠近末端	Acrocentric	P-arm q-arm		
末端	Telocentric			

The shorter arm, by convention, is shown above the centromere and is called the p arm (petite), while the longer arm is called the q arm.

Handwritten text in purple ink, appearing to be a list or notes. The text is written in a cursive style and is arranged in several lines, though the words are difficult to decipher due to the handwriting and the image's orientation. The text appears to be a list of items or names, possibly related to a collection or inventory.

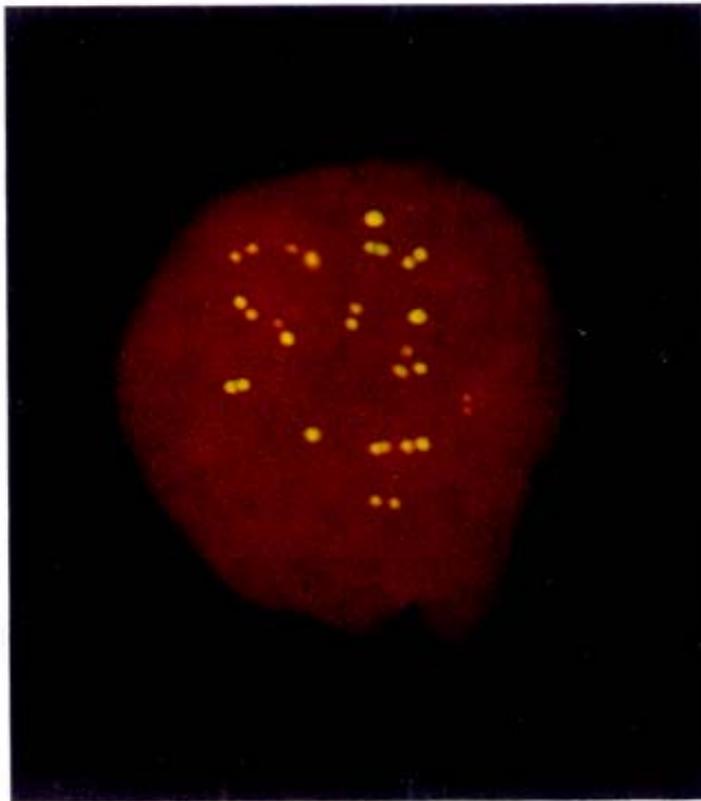


NLPE&PGE





Mitotic chromosome (有丝分裂期的染色体)



(B)



(C)

10 μm



表 2-1 主要动、植物种类的染色体数

物 种	单倍体 染色体数	物 种	单倍体 染色体数
人	23	线虫	11/12 (M/F)
猿猴	21	酵母	8 左右
牛	30	绿藻	10 左右
狗	39	洋葱	8
猪	19	大麦	7
马	32	水稻	12
小鼠	20	玉米	10
大鼠	21	龙头花	8
荷兰猪	32	番茄	12
兔	22	烟草	24
鸡	39	菜豆	11
鳄鱼	16	松树	12
蛙	13	豌豆	7
鲤鱼	52	马铃薯	24
蚕	28	红叶草	7
淡水螳	16	家蝇	6
细菌	1	果蝇	4
猫	19	狐	17



2.1.2.1 蛋白质

染色体蛋白主要分为**组蛋白**和**非组蛋白**两类。

真核细胞的染色体中，**DNA**与**组蛋白**的质量比约为**1: 1**。**DNA**、**组蛋白**和**非组蛋白**及部分**RNA**（尚未完成转录而仍与模板**DNA**相连接的那些**RNA**，其含量不到**DNA**的**10%**）组成了染色体。



组蛋白是染色体的结构蛋白，分为 H_1 、 H_2A 、 H_2B 、 H_3 及 H_4 五种，与DNA共同组成核小体。

组蛋白含有大量的赖氨酸和精氨酸，其中 H_3 、 H_4 富含精氨酸， H_1 富含赖氨酸。 H_2A 、 H_2B 介于两者之间。



组蛋白具有如下特性：

1、进化上的**极端保守性**。

不同种生物组蛋白的氨基酸组成十分相似。牛、猪、大鼠的H₄氨基酸序列完全相同，与豌豆序列相比也只有两个氨基酸的差异。



2、无组织特异性

只有鸟类、鱼类及两栖类红细胞染色体不含 H_1 而带有 H_5 ，精细胞染色体的组蛋白是鱼精蛋白。



3、肽链上氨基酸分布的不对称性

碱性氨基酸集中分布在N端的半条链上，而大部分疏水基团都分布在C端。碱性的半条链易与DNA的负电荷区结合，而另外半条链与其他组蛋白、非组蛋白结合。



4、存在较普遍的修饰作用，
如甲基化、乙基化、磷酸化及
ADP核糖基化等。

修饰作用只发生在细胞周期的特
定时间和组蛋白的特定位点上。



表2-5 真核细胞染色体上的组蛋白成分分析

种类	相对分子质量	氨基酸数目	分离难易度	保守性	染色质中比例	染色质中位置
H ₁	21 000	223	易	不保守	0.5	接头
H ₂ A	14 500	129	较难	较保守	1	核心
H ₂ B	13 800	125	较难	较保守	1	核心
H ₃	15 300	135	最难	最保守	1	核心
H ₄	11 300	102	最难	最保守	1	核心



表2-6 不同组蛋白分子中所含的碱性氨基酸比较
(占氨基酸总数的%)

碱性氨基酸	H ₁	H ₂ A	H ₂ B	H ₃	H ₄
赖氨酸	29.5	10.9	16.0	9.6	10.8
精氨酸	1.3	9.3	6.4	13.3	13.7



非组蛋白约为组蛋白总量的60%~70%，可能有20~100种（常见的有15~20种），主要包括酶类、与细胞分裂有关的收缩蛋白、骨架蛋白、核孔复合物蛋白以及肌动蛋白、肌球蛋白、微管蛋白、原肌蛋白等。



2.1.2.2. DNA

真核细胞基因组的最大特点是它含有大量的重复序列，而且功能DNA序列大多被不编码蛋白质的非功能DNA所隔开，这就是著名的“C值反常现象（C-value paradox）”。

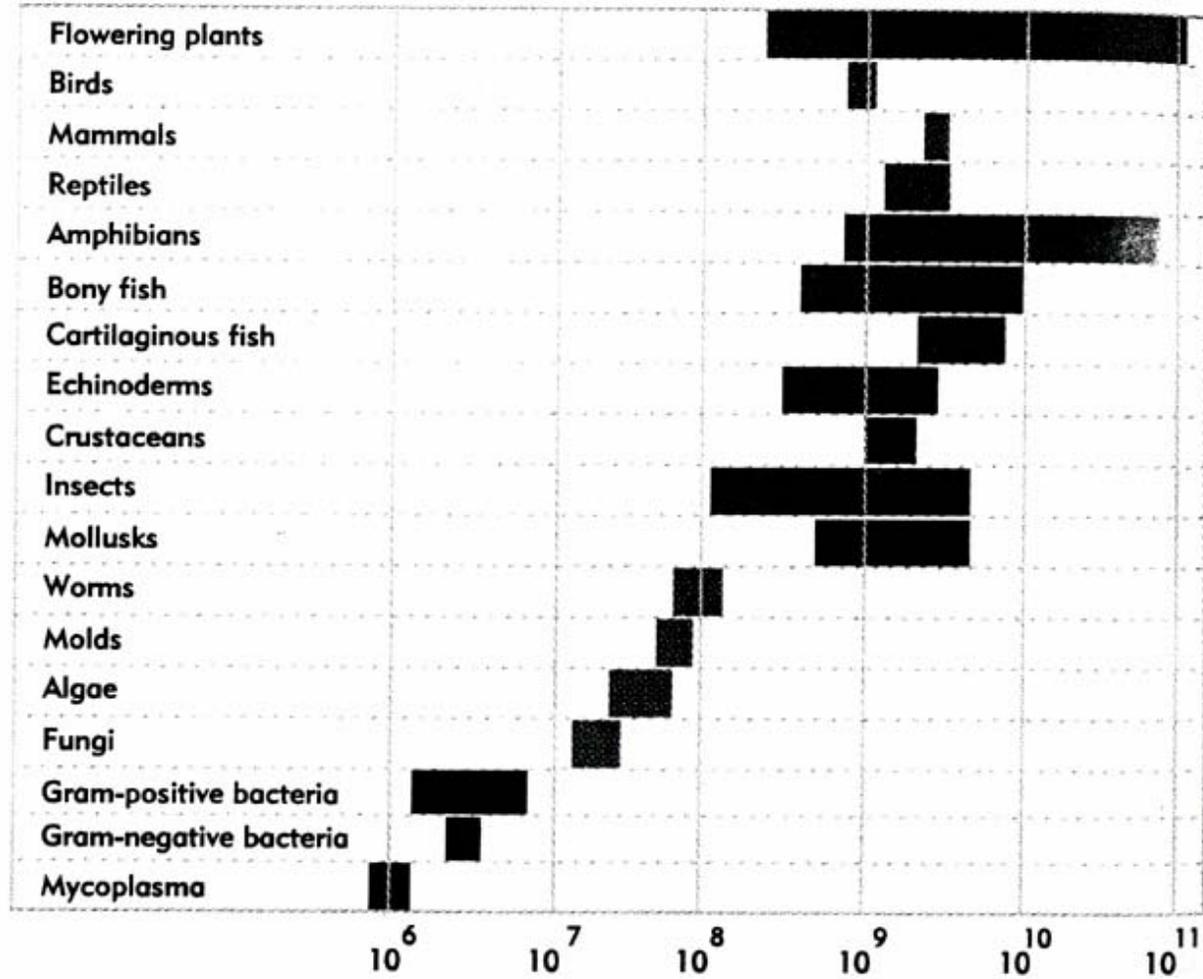
所谓C值，通常是指一种生物单倍体基因组DNA的总量。



各种生物细胞内DNA总量的比较

Figure 22.1

DNA content of the haploid genome is related to the morphological complexity of lower eukaryotes, but varies extensively among the higher eukaryotes. The range of DNA values within a phylum is indicated by the shaded area.



同类生物不同种属之间DNA总量变化很大。
从编码每类生物所需的DNA量的最低值看，生物细胞中的C值具有从低等生物到高等生物逐渐增加的趋势。



真核细胞DNA序列可被分为3类：

1、不重复序列。

在单倍体基因组里，这些序列一般只有一个或几个拷贝，它占DNA总量的**10%~80%**。不重复序列长约**750~2 000bp**，相当于一个结构基因的长度。

蛋清蛋白、蚕的丝心蛋白、血红蛋白和珠蛋白等都是单拷贝基因。

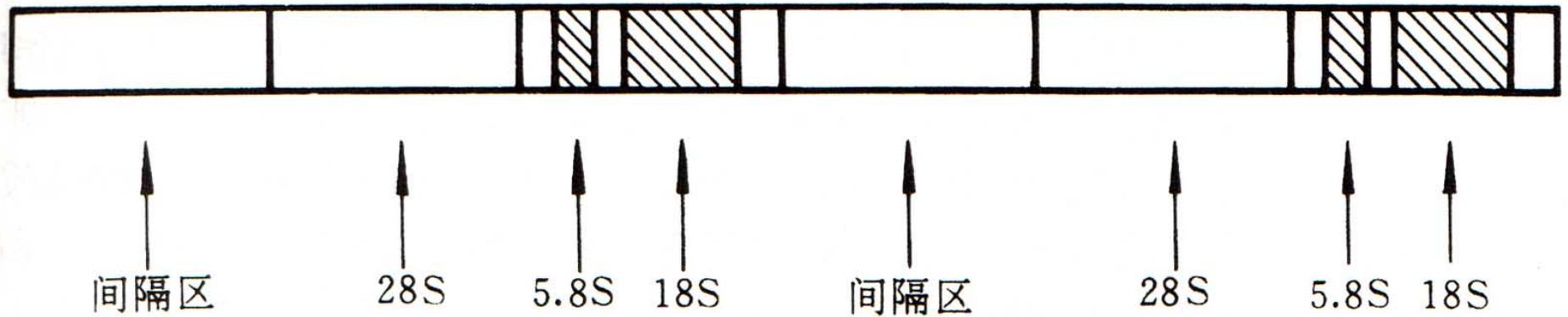


2、中度重复序列

这类序列的重复次数在 $10^1 \sim 10^4$ 之间，占总DNA的**10%~40%**，如小鼠中占**20%**，果蝇中占**15%**，各种**rRNA**、**tRNA**以及某些结构基因如组蛋白基因等都属于这一类。



图2-6 非洲爪蟾的rRNA基因结构示意图





在动物卵细胞形成过程中，**rDNA**可进行几千次不同比例的复制，产生 2×10^6 个拷贝，使**rDNA**占卵细胞**DNA**的**75%**，从而使该细胞能积累 10^{12} 个核糖体，以合成大量蛋白质供细胞分裂之需。



3、高度重复序列——卫星DNA

只存在于真核生物中，占基因组的10%~60%，由6~100个碱基组成，在DNA链上串联重复高达数百万次。因为卫星DNA不转录，其功能不详。它们是异染色质的成份，可能与染色体的稳定性有关。



Drosophila_satellite DNA repeat (several million copies)



ACAAACT, 1.1×10^7 bp, 25% genome

ATAAACT, 3.6×10^6 bp, 8% genome

ACAAATT, 3.6×10^6 bp, 8% genome

AATATAG, cryptic

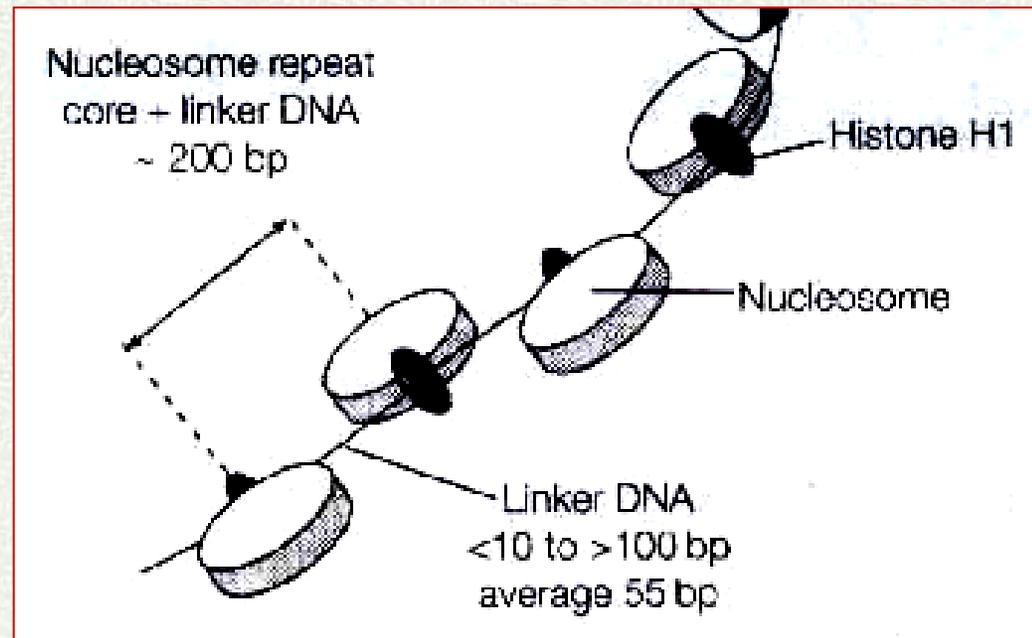
Satellites comprise more than 40% of the genome



2.1.2.3 染色质和核小体

由DNA和组蛋白组成的染色质纤维细丝是许多核小体连成的念珠状结构。

**“Beads on a string”
structure**





染色质DNA的 T_m 值比自由DNA高，说明在染色质中DNA极可能与蛋白质分子相互作用。

在染色质状态下，由DNA聚合酶和RNA聚合酶催化的DNA复制和转录活性大大低于在自由DNA中的反应。



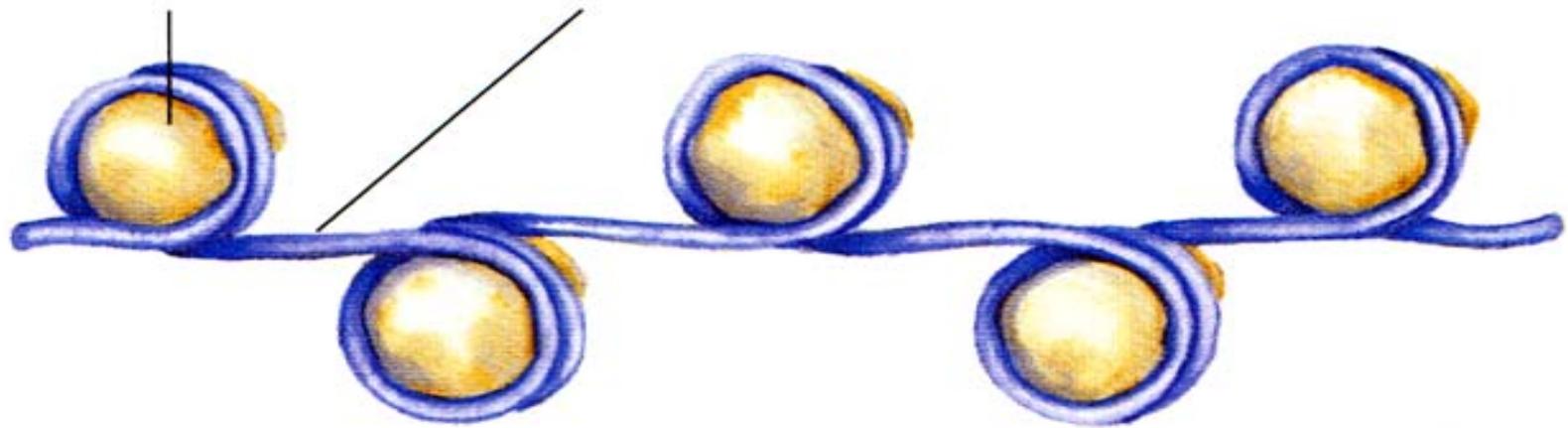
DNA酶I (DNaseI) 对染色质DNA的消化远远慢于对纯DNA的作用。

用小球菌核酸酶处理染色质以后进行电泳，便可以得到一系列片段，这些被保留的**DNA**片段均为**200bp**基本单位的倍数。



核小体结构示意图

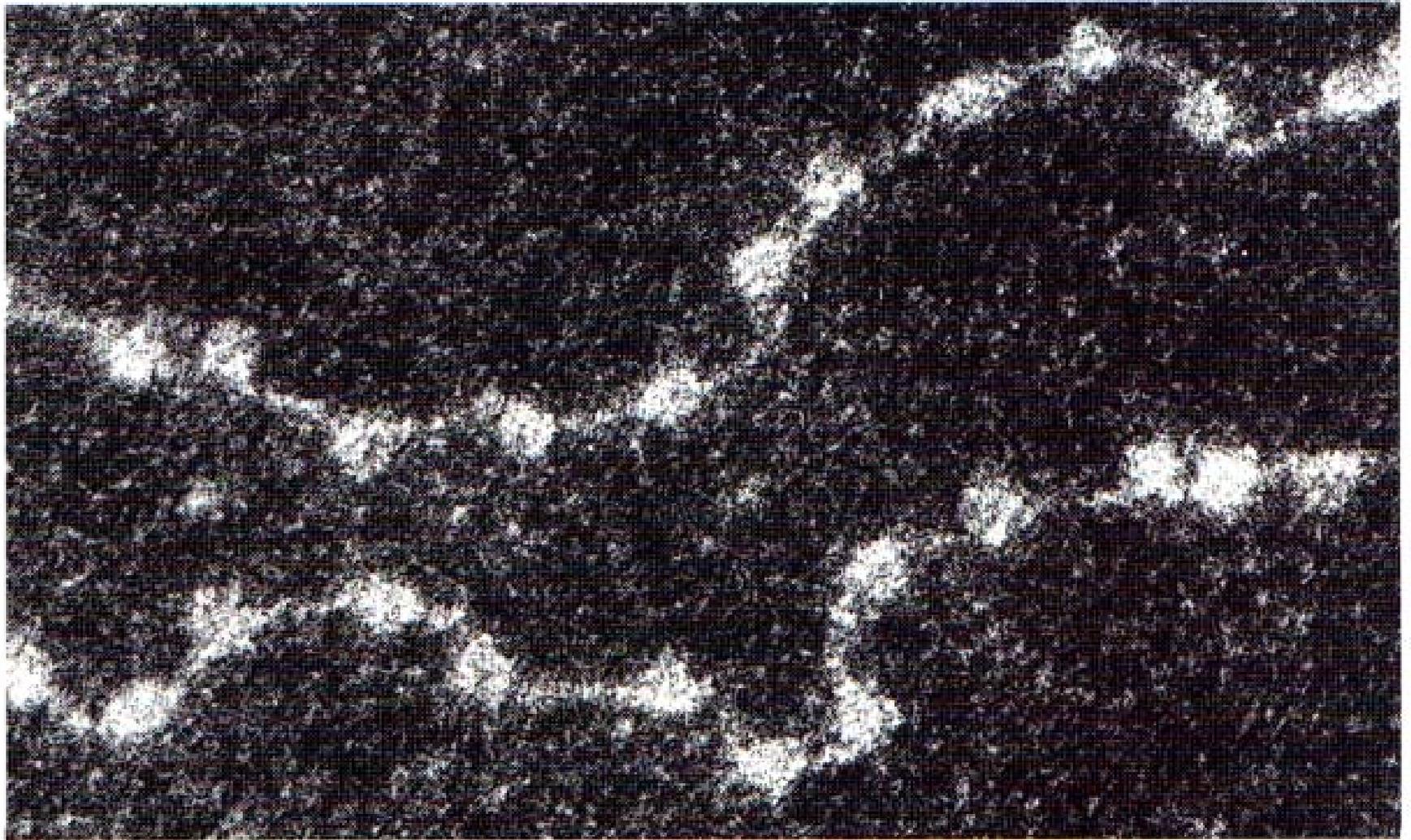
核小体中的组蛋白核心结构 核小体“接头”之间的 DNA



(a)



电镜下看到的染色质结构

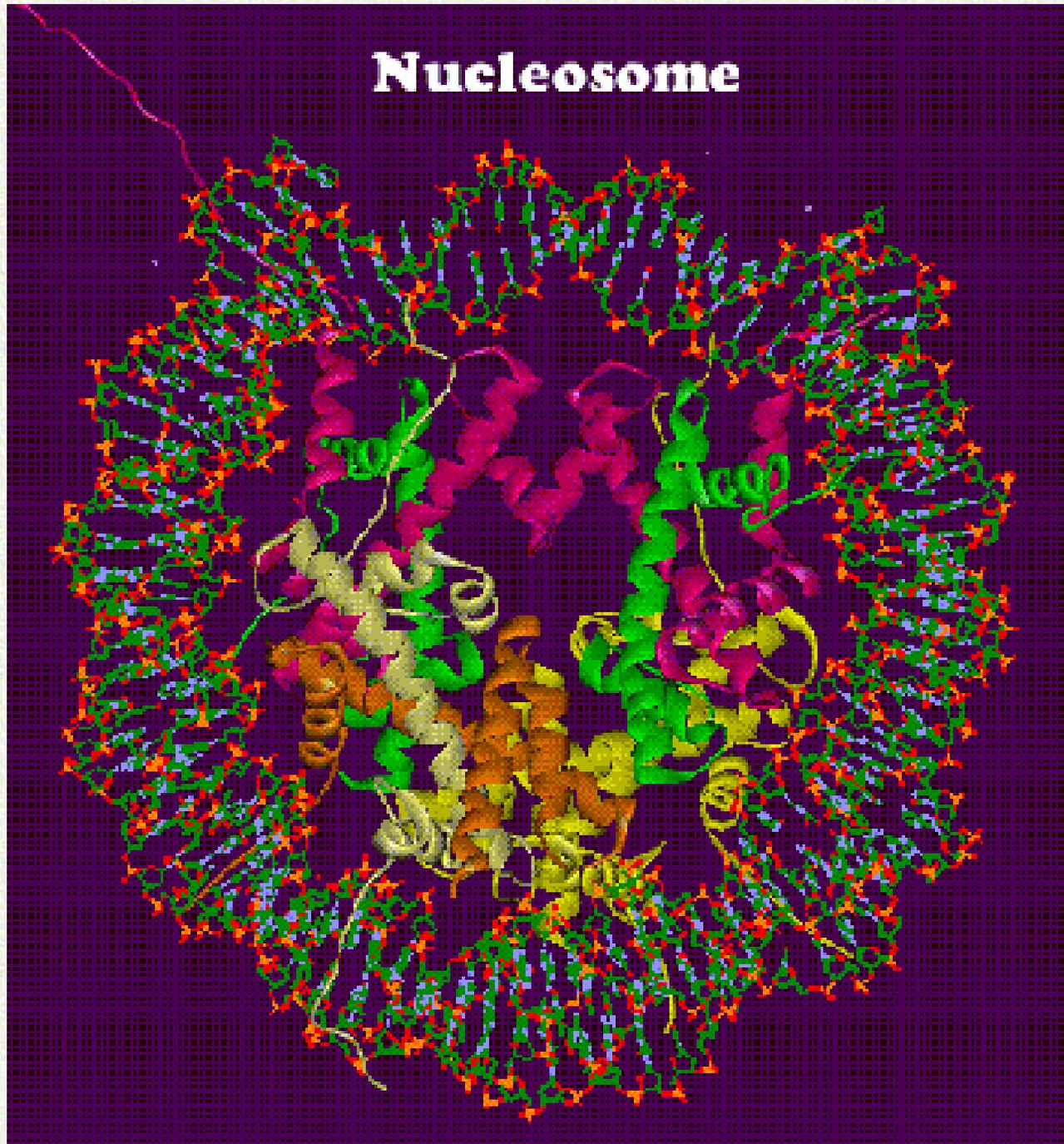


50 nm

(b)



Nucleosome



Nucleosome (核小体) 是染色质的基本结构单位，由~200 bp DNA和组蛋白八聚体组成

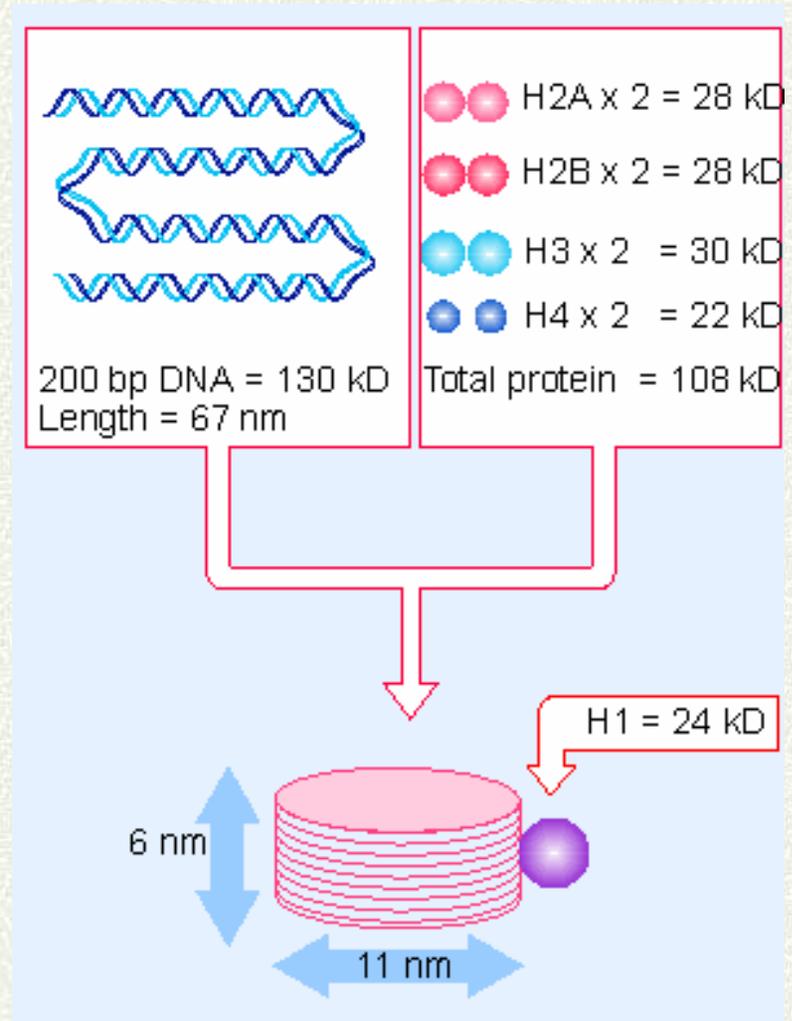
DNA + Histone octamer (组蛋白八聚体)

→ Nucleosome core (核小体核心 146bp) +

H1 → Chromatosome (染色小体 166bp) +

linker DNA → Nucleosome (核小体)

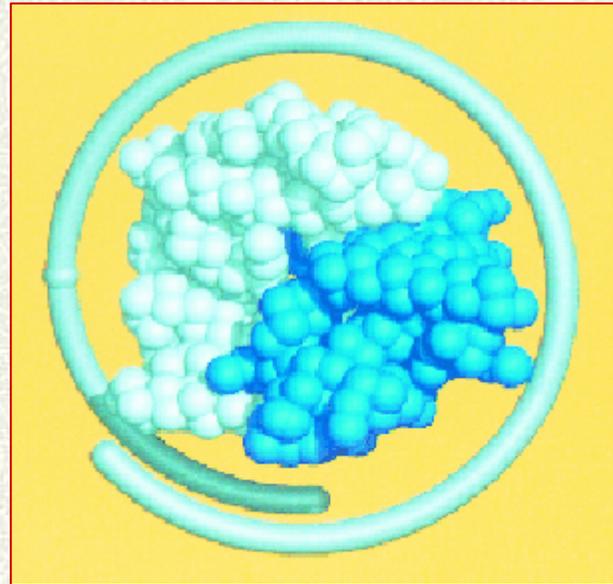
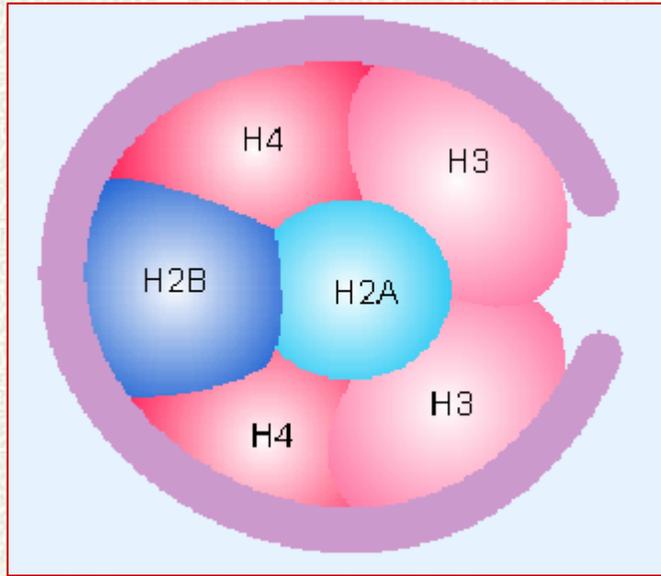
(~200 bp of DNA)



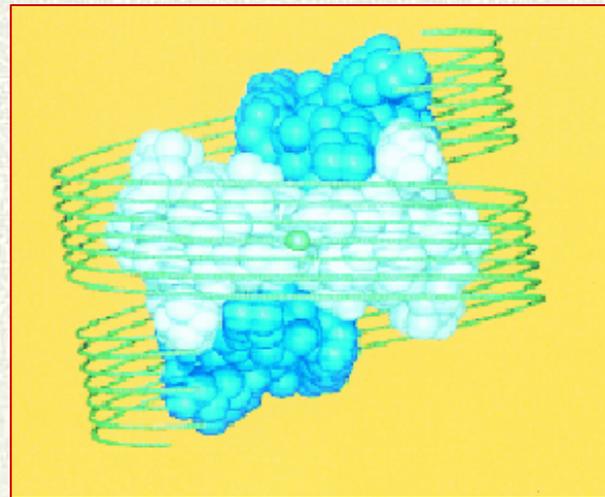
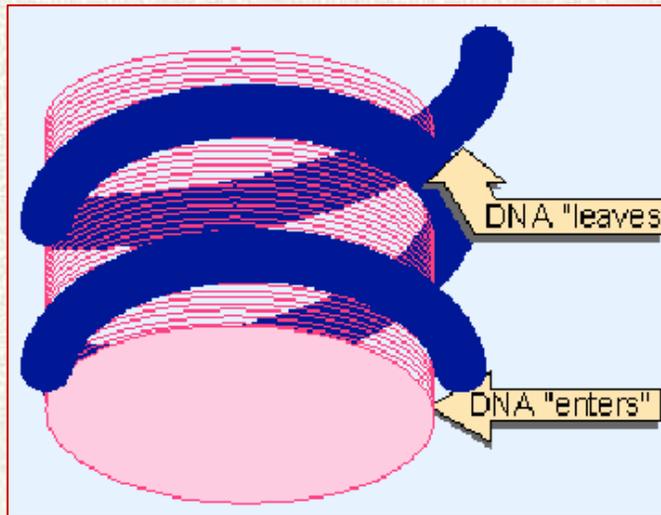


Histone octamer (组蛋白八聚体)

Nucleosome core



Top view



Side view

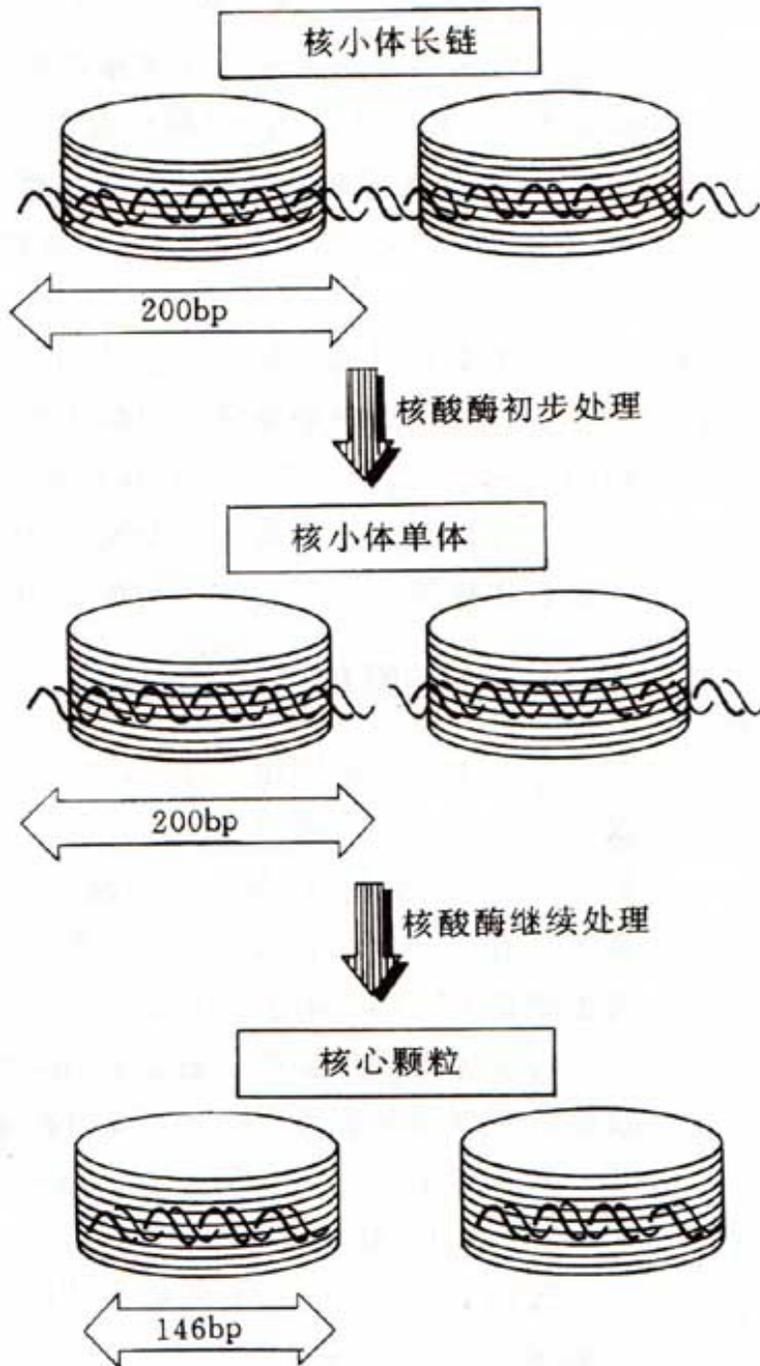
核小体单元的产生

核小体由H2A、H2B、H3、H4各两个分子生成的八聚体和由大约**200bp DNA**组成。八聚体在中间，DNA分子盘绕在外，而每个核小体只有一个H1，分布在核小体的外面。

核心颗粒包括组蛋白八聚体及与其结合的**146bp DNA**，该序列绕在八聚体外面**1.75圈**，每圈约**80bp**。

由许多核小体构成了连续的染色质DNA细丝。

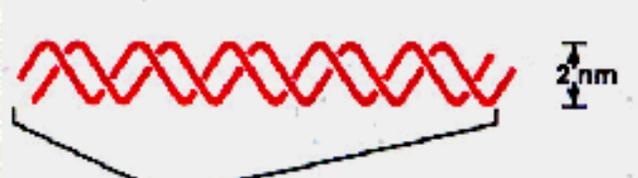
NLPE&PGE



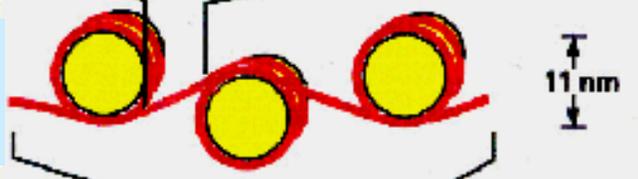


From DNA to chromosome

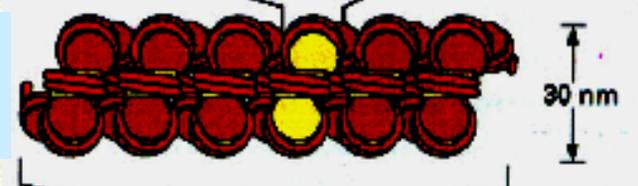
Short region of DNA double helix



“Beads on a string” form of Chormatin



30-nm chromatin fiber of packed nucleosomes



Section of chromosome in an extended form



Condensed section of Metaphase chromosome

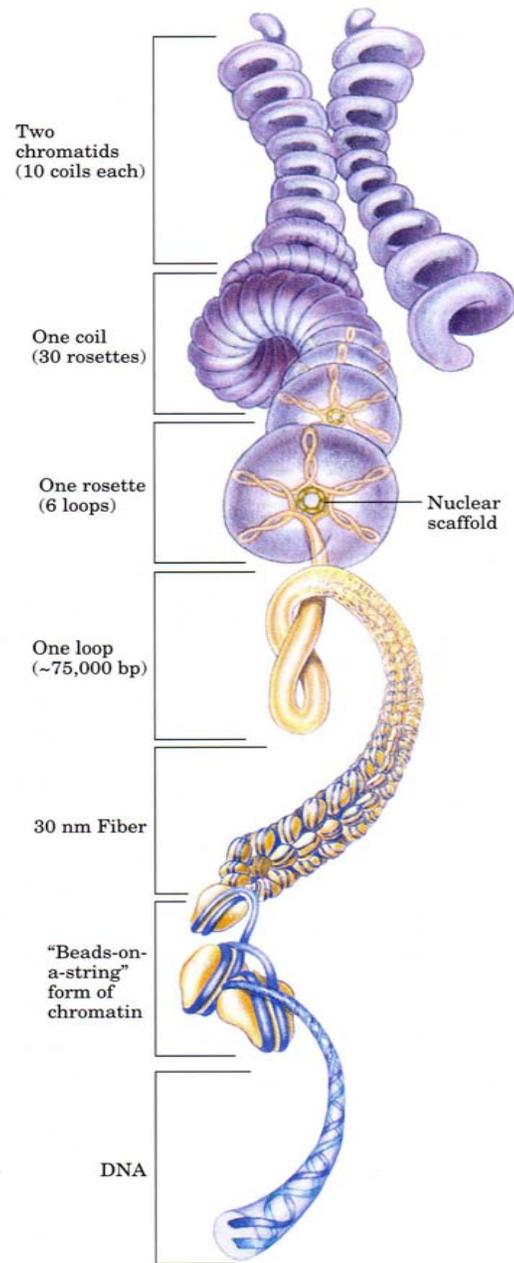


Entire metaphase chromosome





染色体DNA结构示意图



两段各含10个螺旋的染色质

一个螺旋中包含30个莲座状结构

每个莲座状结构中都有6个环状DNA

每个环状结构中含有75000 bp

30 nm结构: Solenoid (螺线管)

染色体DNA的念珠状结构

双链DNA



在核小体中，DNA盘绕组蛋白八聚体核心，使分子收缩 $1/7$ 。

人中期染色体中含 6.2×10^9 碱基对，其理论长度应是200cm，这么长的DNA被包装在46个 $5 \mu\text{m}$ 长的圆柱体（染色体）中，其压缩比约为 10^4 。

分裂间期染色质比较松散，压缩比大约是 $10^2 \sim 10^3$ 。



染色体形成过程中长度与宽度的变化

				宽度增加	长度压缩
第一级	DNA+组蛋白	↔	核小体	5倍	7倍
第二级	核小体	↔	螺线管	3倍	6倍
第三级	螺线体	↔	超螺旋	13倍	40倍
第四级	超螺线体	↔	染色体	2.5-5倍	5倍
				500-1000倍	8400倍 (8000-10000)



2.1.3 原核生物基因组

原核生物的基因组很小，大多只有一条染色体，且DNA含量少，如大肠杆菌DNA的相对分子质量仅为 4.6×10^6 bp，其完全伸展总长约为1.3mm，含4000多个基因。



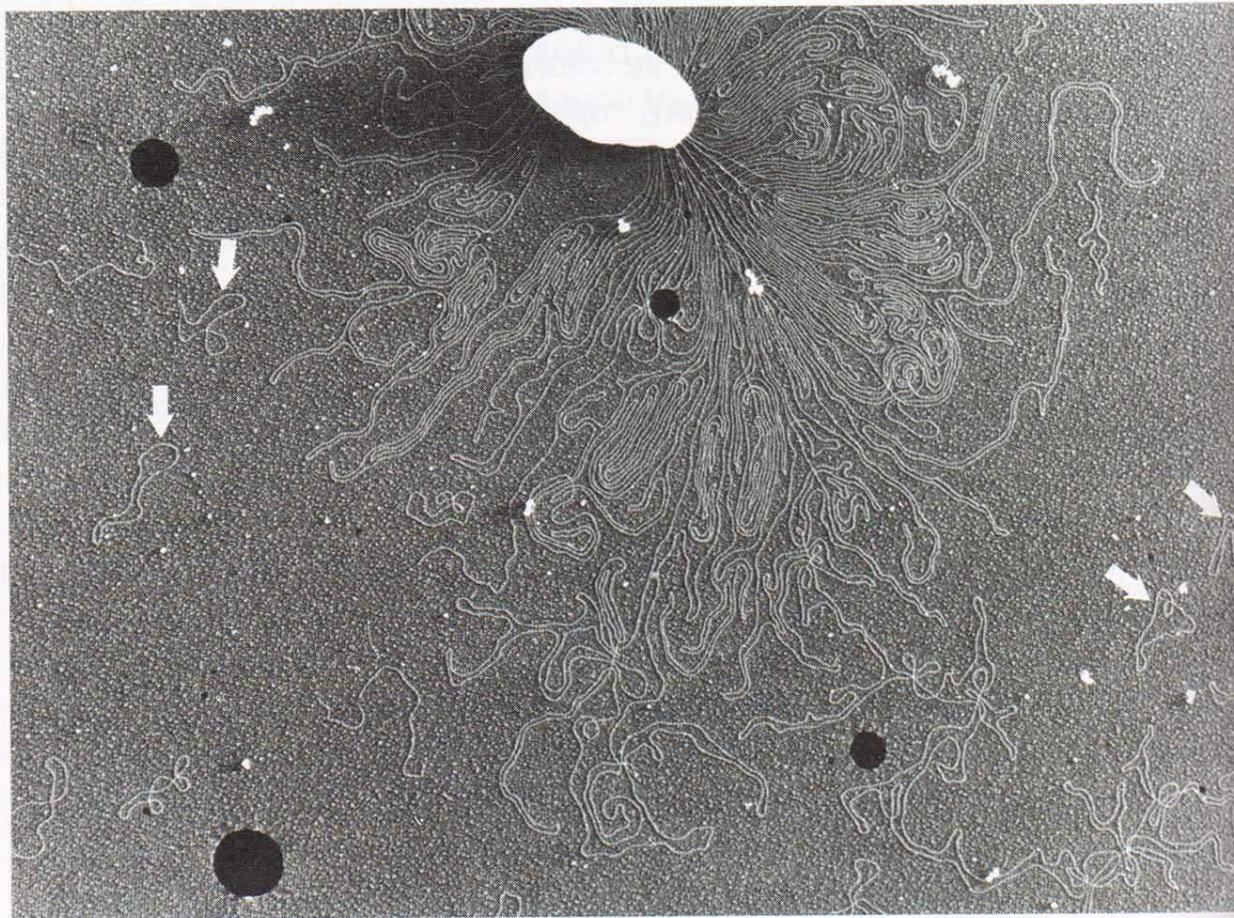
原核生物基因主要是单拷贝基因，只有很少数基因（如rRNA基因）以多拷贝形式存在；

整个染色体DNA几乎全部由功能基因与调控序列所组成；

几乎每个基因序列都与它所编码的蛋白质序列呈线性对应状态。



大肠杆菌细胞中基因组DNA的电镜显微照片



细菌DNA是一条相对分子量在 10^9 左右的共价、闭合双链分子，通常也称为染色体。

箭头处为环状质粒DNA。



原核细胞DNA特点:

1、结构简炼。

原核DNA分子的绝大部分是用来编码蛋白质的，只有很小一部分控制基因表达的序列不转录。

如在 Φ X174中不转录部分只占4%左右 (217/5386), T4 DNA中占5.1% (282/5577)。



2、存在转录单元

原核生物**DNA**序列中功能相关的**RNA**和蛋白质基因，往往丛集在基因组的一个或几个**特定部位**，形成转录单元并转录产生含多个**mRNA**的分子，称为**多顺反子mRNA**。



3、有重叠基因。

一些细菌和动物病毒存在重叠基因，同一段DNA能携带两种不同蛋白质的信息。

1973年，Weiner和Weber在研究一种大肠杆菌RNA病毒时发现，有两个基因从同一起点开始翻译，一个在400bp处结束，而在3%的情况下，翻译可一直进行下去直到800bp处碰到双重终止信号时才停止。



1977年，Sanger正式发现了**重叠基因**：

Φ X174感染寄主后共合成**9**个蛋白质，相对分子质量约 **2.5×10^5** ，相当于**6 078**个核苷酸，而病毒DNA本身只有**5 375**个核苷酸。Sanger在弄清 Φ X174 DNA的全部核苷酸序列及各个基因的起迄位置和密码数目以后发现，**9**个基因中有些是重叠的。

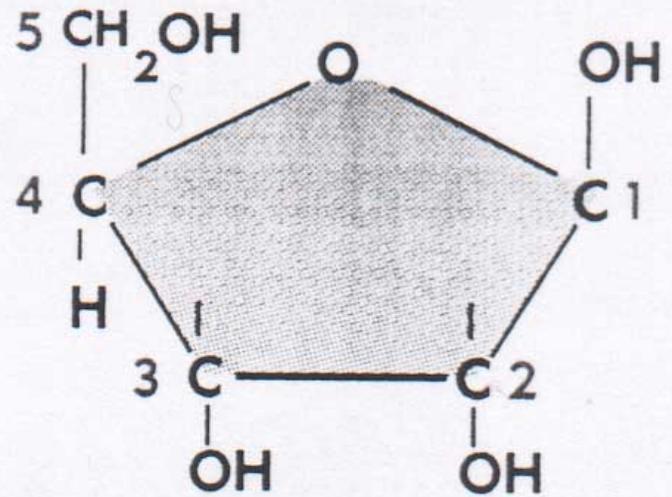
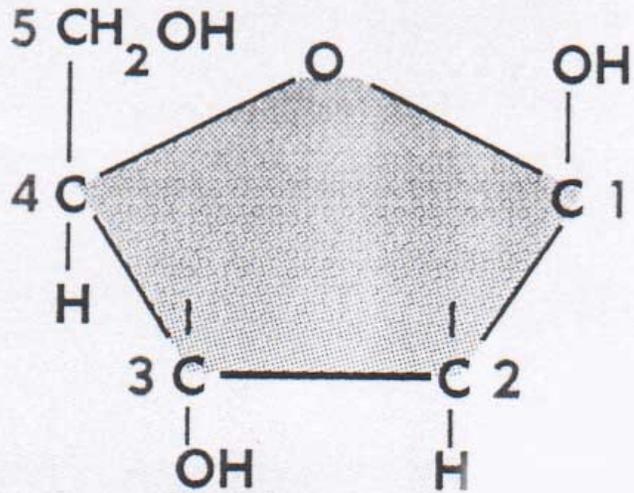


2.2 DNA的结构

2.2.1 DNA的一级结构

DNA又称**脱氧核糖核酸**，是 **deoxyribonucleic acid**的简称，它是一种高分子化合物，其**基本单位**是**脱氧核苷酸**。

在**DNA**分子中，**磷酸**和**脱氧核糖**是不变的，而**4种含氮碱基**即**腺嘌呤**（**A**）、**鸟嘌呤**（**G**）、**胞嘧啶**（**C**）和**胸腺嘧啶**（**T**）是可变的。



2-脱氧核糖（左）和核糖（右）结构式



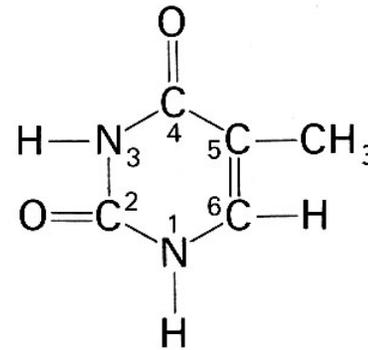
组成DNA分子的碱基只有4种：

胸腺嘧啶 (T)

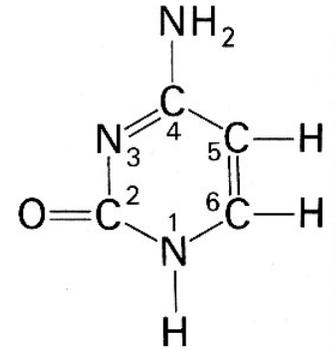
胞嘧啶 (C)

腺嘌呤 (A)

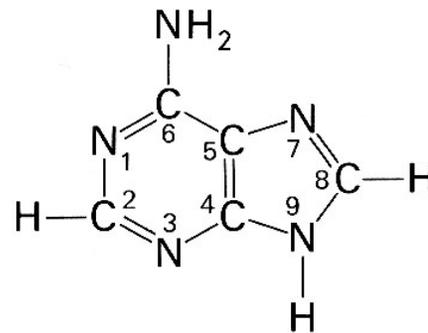
鸟嘌呤 (G)



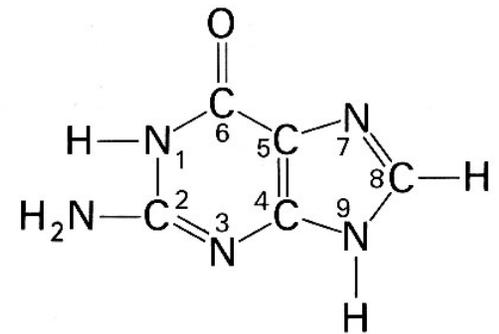
Thymine
(T)



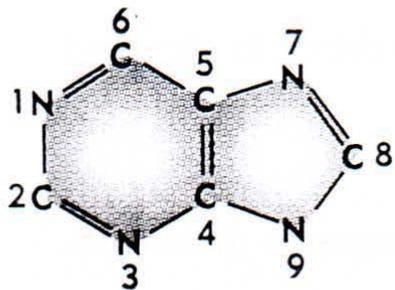
Cytosine
(C)



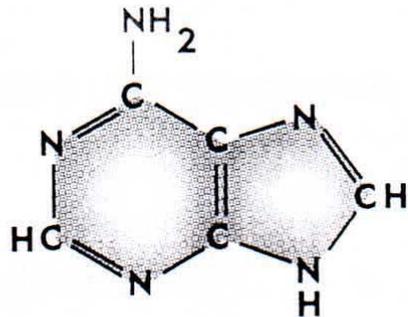
Adenine
(A)



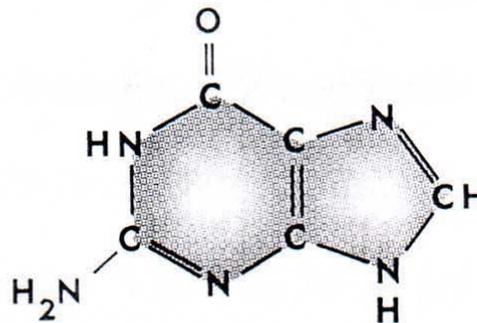
Guanine
(G)



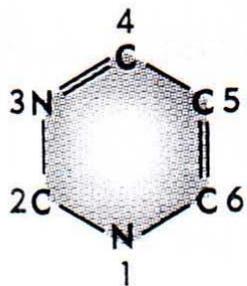
嘌呤



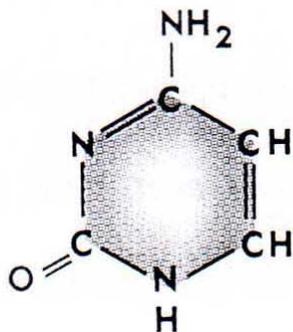
腺嘌呤



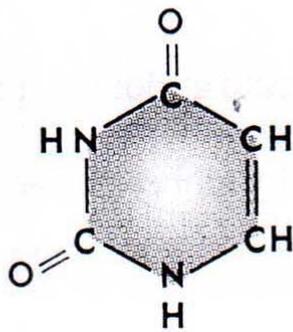
鸟嘌呤



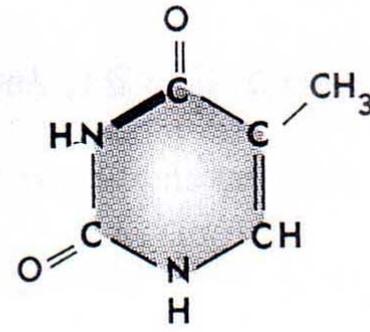
嘧啶



胞嘧啶



尿嘧啶

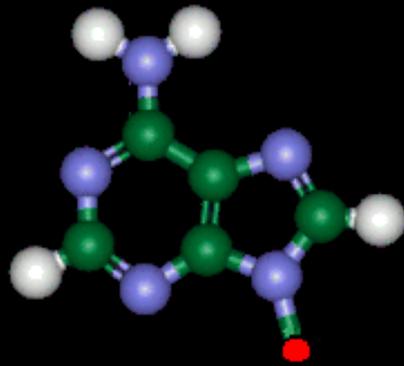


胸腺嘧啶

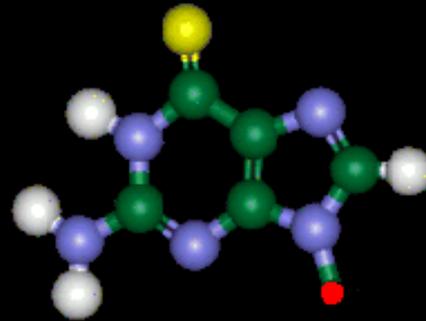
图2-2 组成DNA和RNA分子的五种含氮碱基的结构式



Purines

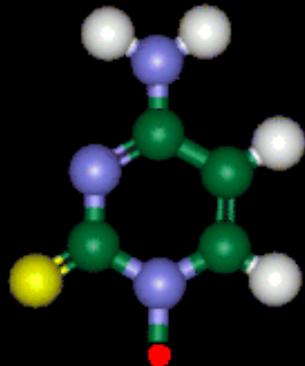


Adenine

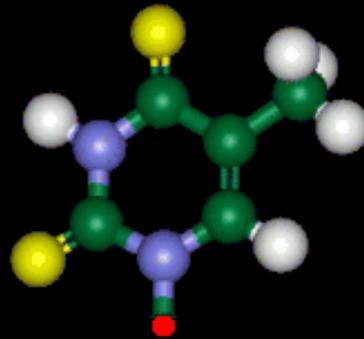


Guanine

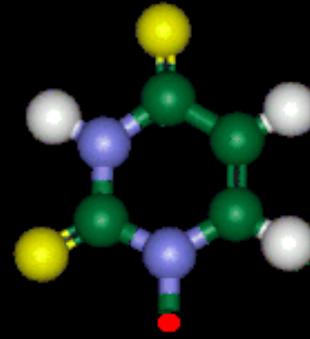
Pyrimidines



Cytosine



Thymine



Uracil

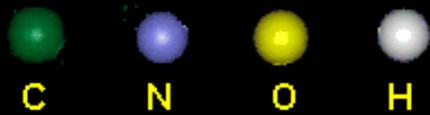
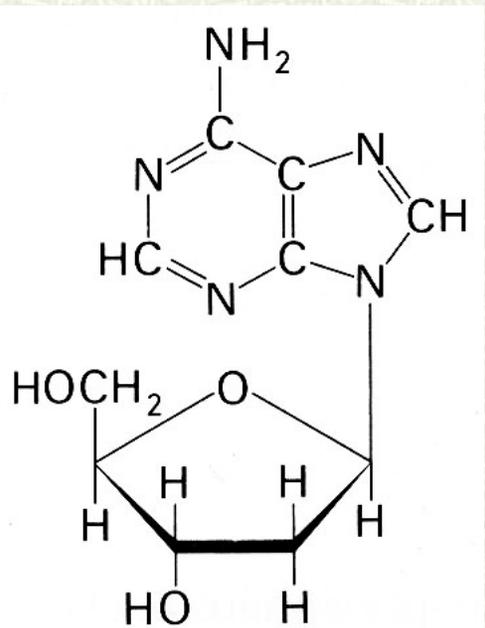


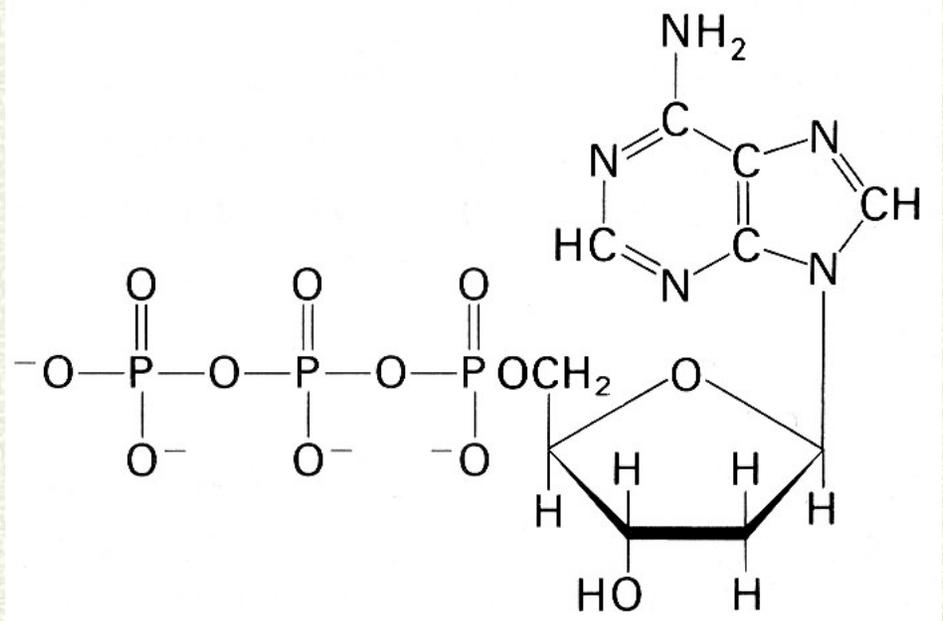


表2-7 碱基、核苷和核苷酸

碱基	核苷	核苷酸	RNA	DNA
腺嘌呤 (adenine)	腺苷 (adenosine)	腺苷酸 (adenylic acid)	AMP	dAMP
鸟嘌呤 (guanine)	鸟苷 (guanosine)	鸟苷酸 (guanylic acid)	GMP	dGMP
胞嘧啶 (cytosine)	胞苷 (cytidine)	胞苷酸 (cytidylic acid)	CMP	dCMP
胸腺嘧啶 (thymine)	胸苷 (thymidine)	胸苷酸 (thymidylic acid)		dTMP
尿嘧啶 (uracil)	尿苷 (uridine)	尿苷酸 (uridylic acid)	UMP	



**Deoxyadenosine
(A nucleoside)**



**Deoxyadenosine 5'-triphosphate
(dATP)
(A nucleotide)**

Base

Nucleoside

Nucleotide

Adenine

(Deoxy)adenosine

(d)A (mono, di-, tri) phosphate

Guanine

(Deoxy)guanosine

(d)G (mono, di-, tri) phosphate

Thymine

(Deoxy)thymidine

(d)T (mono, di-, tri) phosphate

Cytosine

(Deoxy)cytidine

(d)C (mono, di-, tri) phosphate

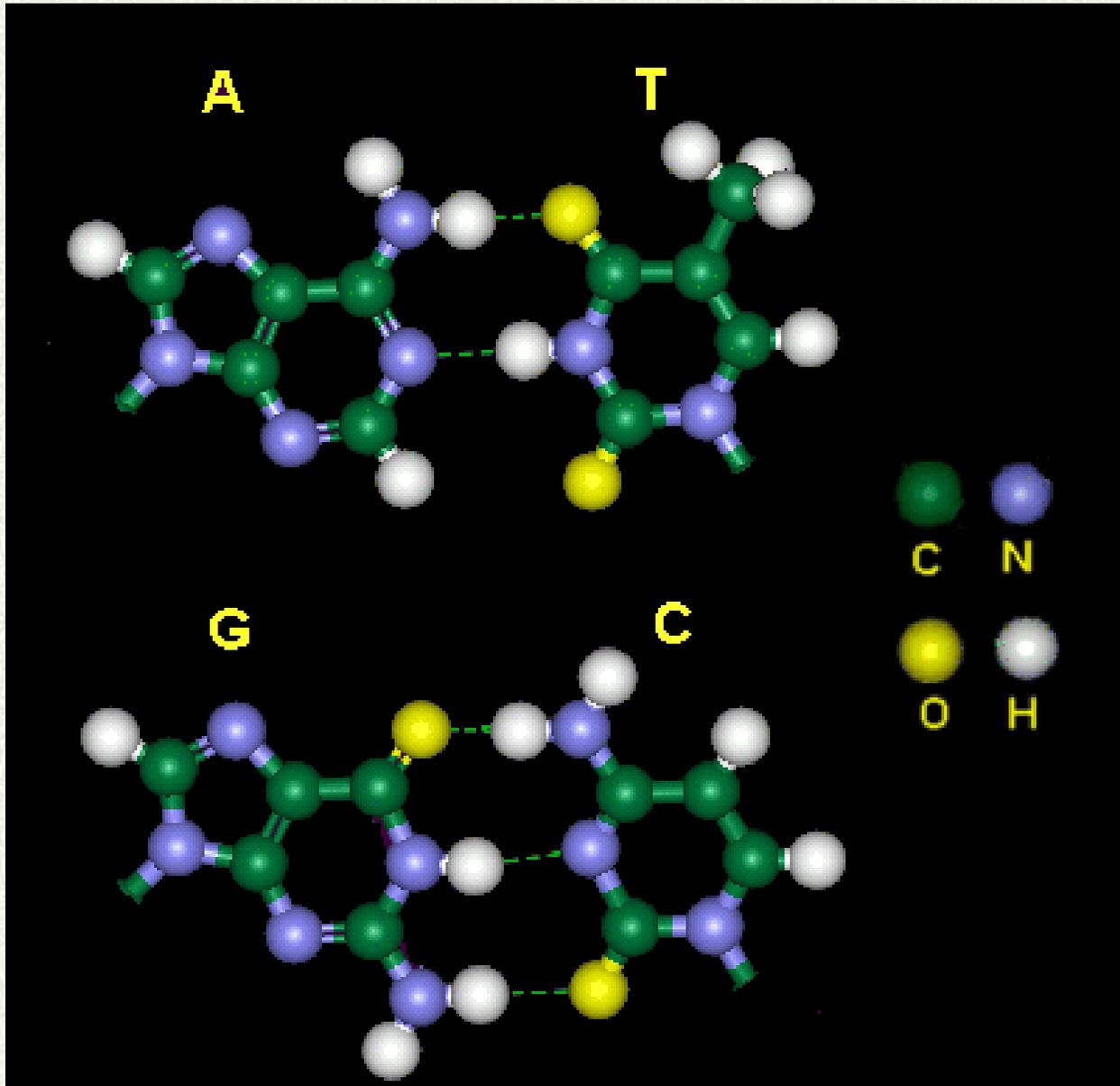


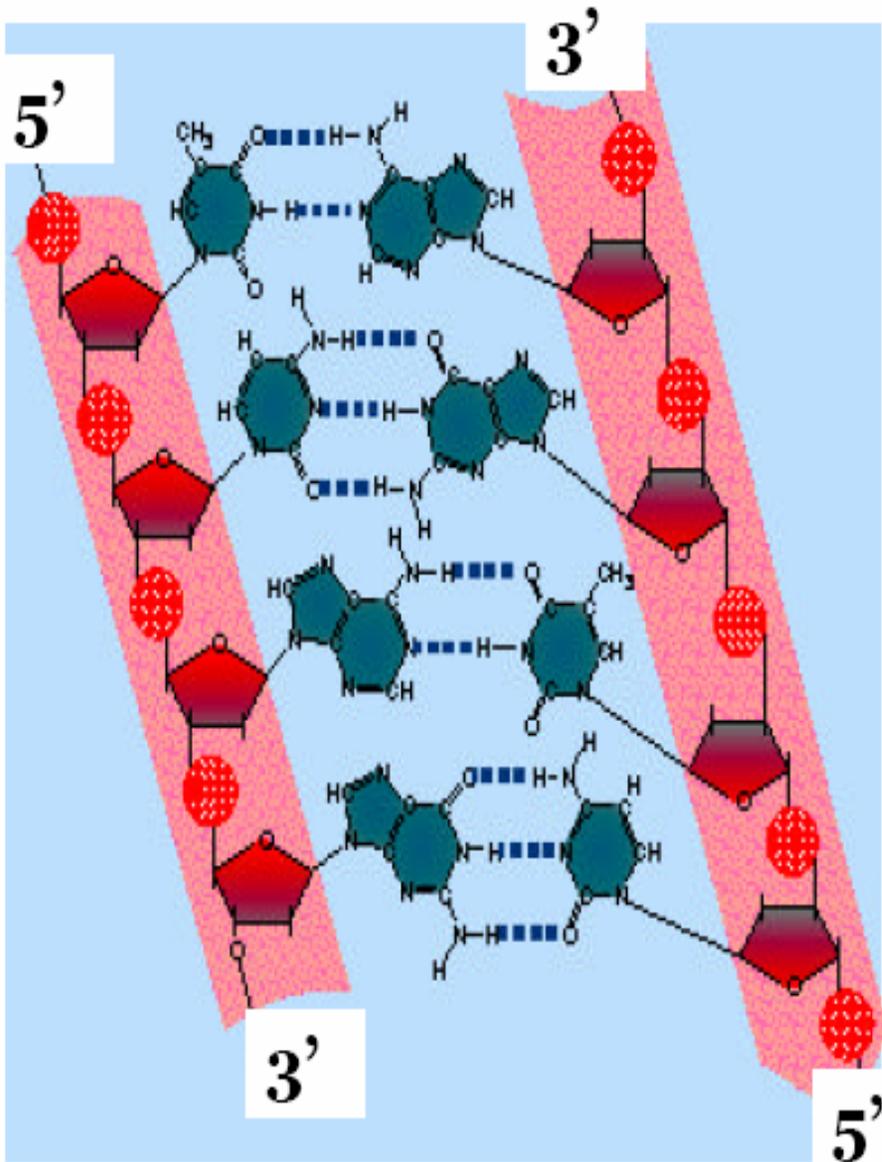
在DNA分子中，**嘌呤**永远与**嘧啶**配对，而且**腺嘌呤**（A）只能与**胸腺嘧啶**（T）配对，**鸟嘌呤**（G）只能与**胞嘧啶**（C）配对。

如一条链上某一碱基是C，另一条链上与它配对的碱基必定是G。碱基之间的这种一一对应的关系叫**碱基互补配对原则**。



Base pairing





DNA链的基本特点是：

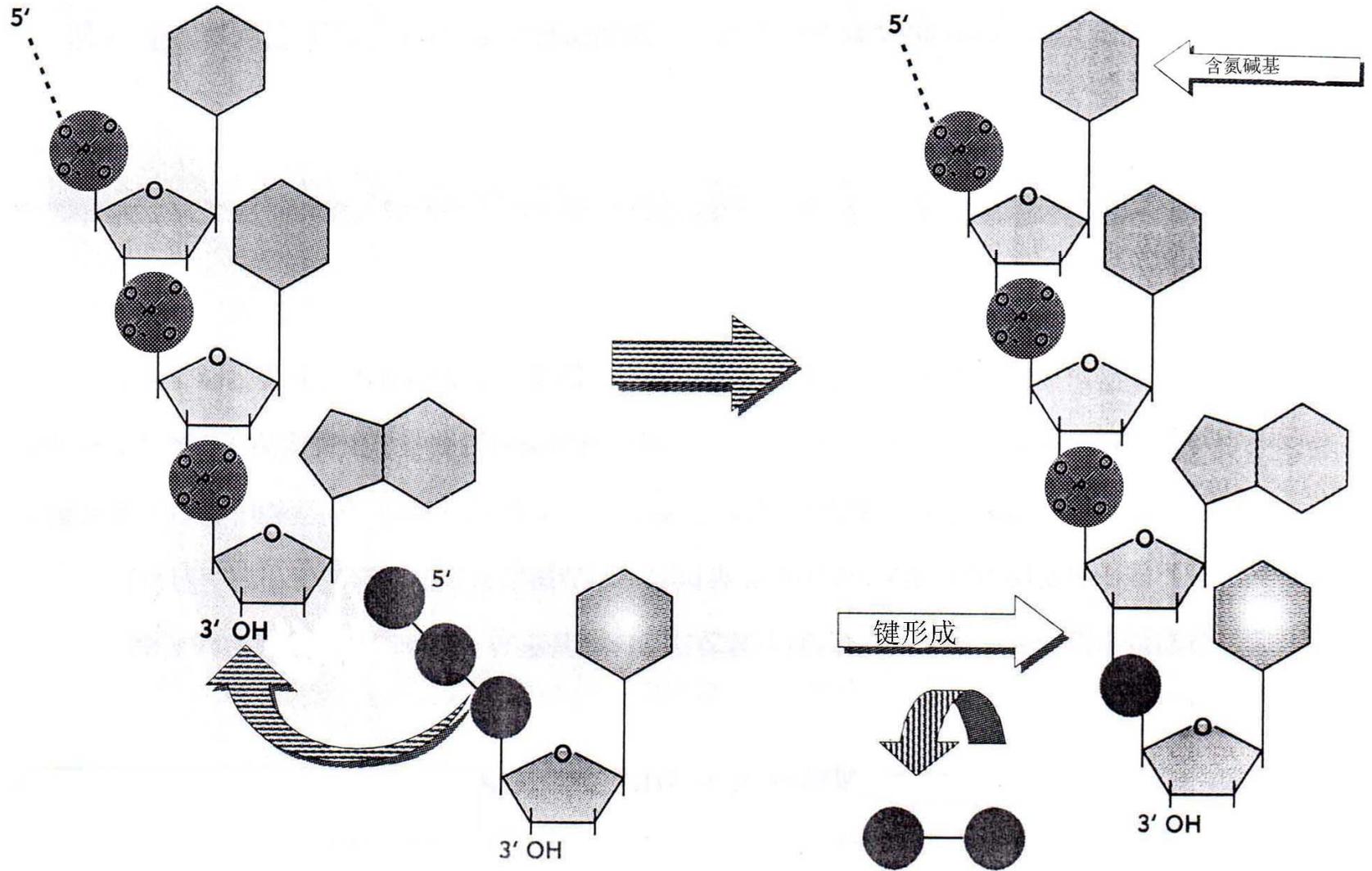
1、DNA是由两条互相平行的脱氧核苷酸长链盘绕而成的。

2、DNA分子中的脱氧核糖和磷酸交替连接，排在外侧，构成基本骨架，碱基排列在内侧。

3、两条链上的碱基通过氢键相结合，形成碱基对。



多聚核苷酸链中新生磷酸糖苷键的产生过程



单核苷酸5'-磷酸基团向核酸链的3'-OH发起进攻



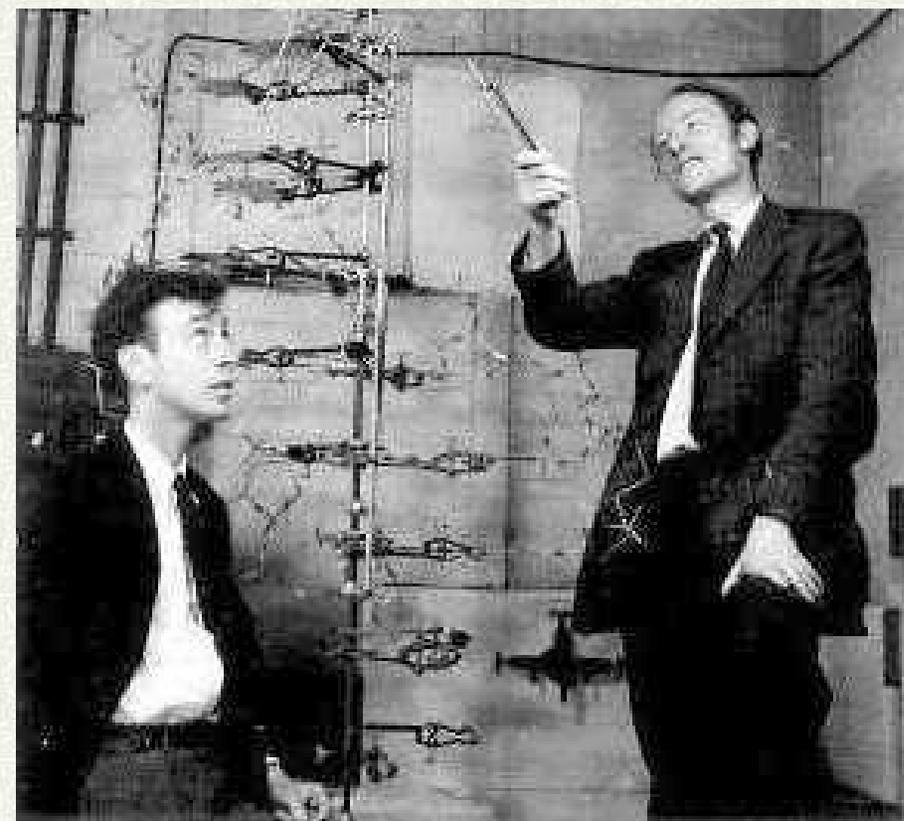
2. 2. 2 DNA的二级结构

DNA的一级结构是指4种核苷酸的连接及其排列顺序，表示了该**DNA**分子的化学构成。

DNA不仅具有严格的化学组成，还具有特殊的高级结构，它主要以有**规则的双螺旋**形式存在。



DNA double helix (DNA双螺旋结构)



- **Watson and Crick , 1953**
- **The genetic material of all organisms except for some viruses**
- **The foundation of the molecular biology**

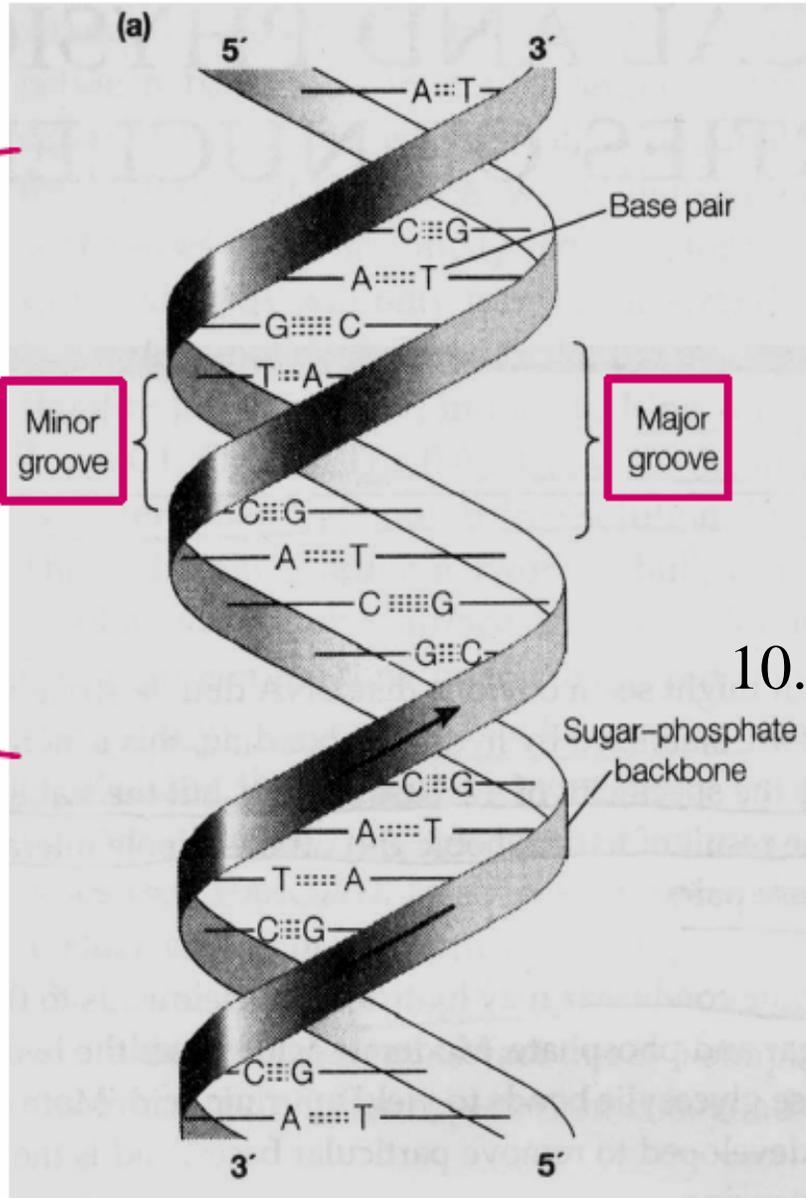


DNA的二级结构是指两条多核苷酸链**反向平行盘绕**所生成的**双螺旋结构**。

通常情况下，**DNA的二级结构**分两大类：一类是**右手螺旋**，如**A-DNA**和**B-DNA**；另一类是**左手螺旋**，即**Z-DNA**。



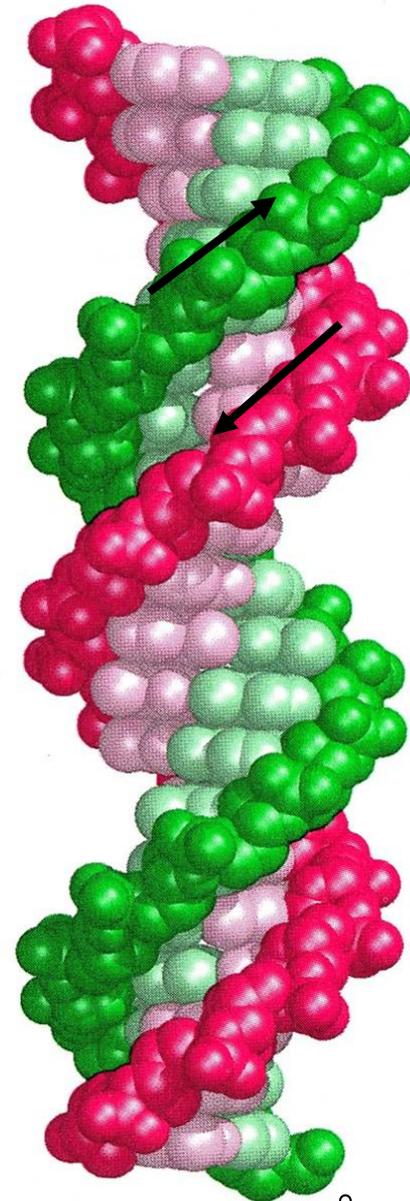
DNA的反向平行双螺旋结构



Rise
 3.4 \AA

Pitch
 34 \AA
 34 \AA

10.4 bp/turn



Minor
Groove

Major
Groove

Width 20 \AA

NLPE&PGE



图2-11 A-, B-, Z-型DNA三种结构比较

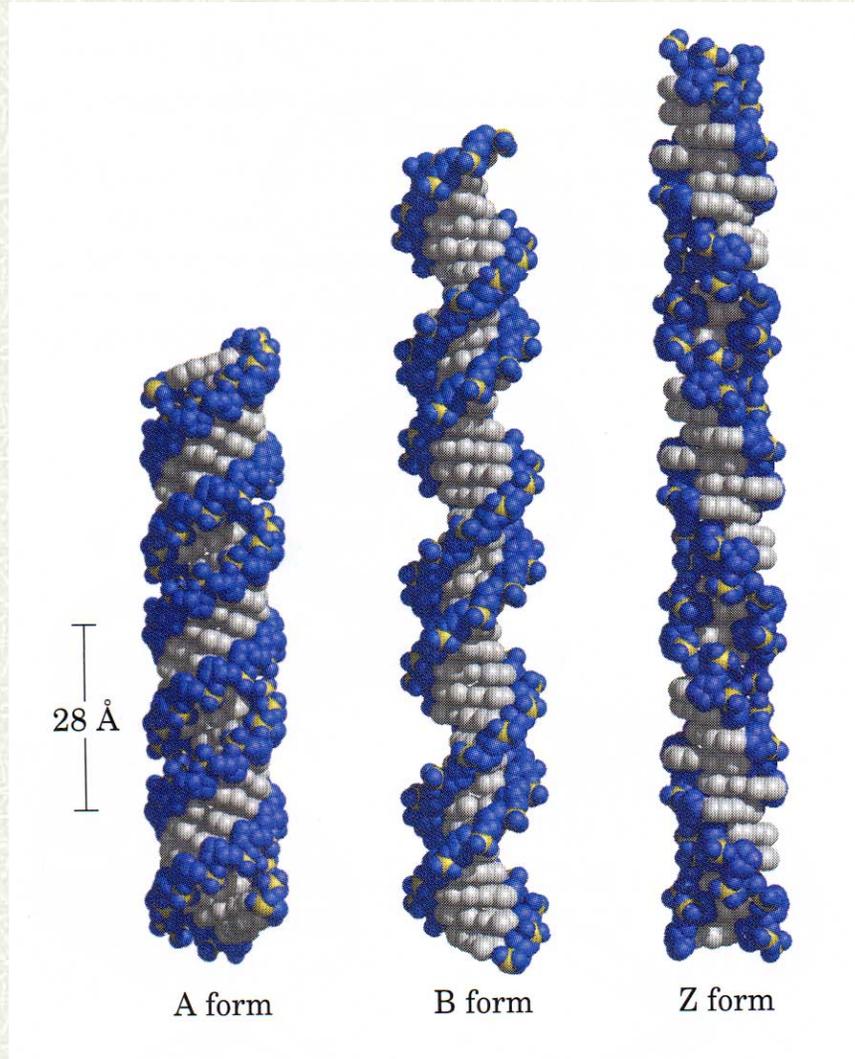




表2-8 不同螺旋形式DNA分子主要参数比较

双螺旋	碱基倾角（度）	碱基间距（nm）	螺旋直径（nm）	每轮碱基数	螺旋方向
A-DNA	20	0.26	2.6	11	右
B-DNA	6	0.34	2.0	10	右
Z-DNA	7	0.37	1.8	12	左



2.2.2.2 DNA二级结构中左手螺旋——**Z-DNA**的研究

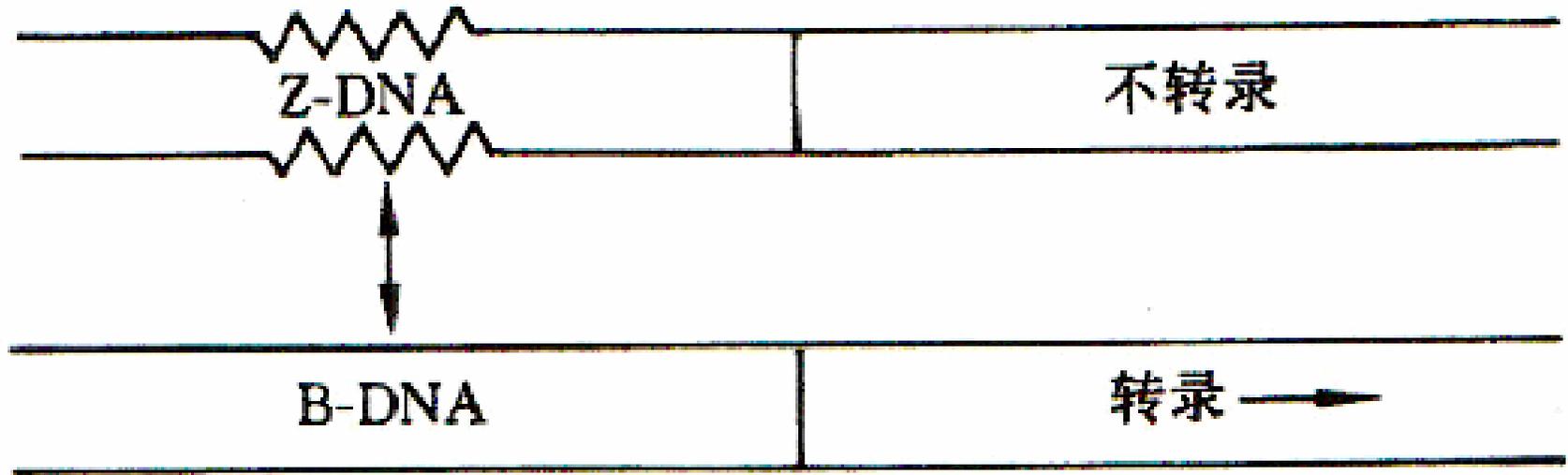
B-DNA是最常见的DNA构象，**A-DNA**和**Z-DNA**可能具有不同的生物活性。



Z-DNA调控基因转录的两个模式

A. 邻近调控(转录区与调节区相邻)

控制区-启动子

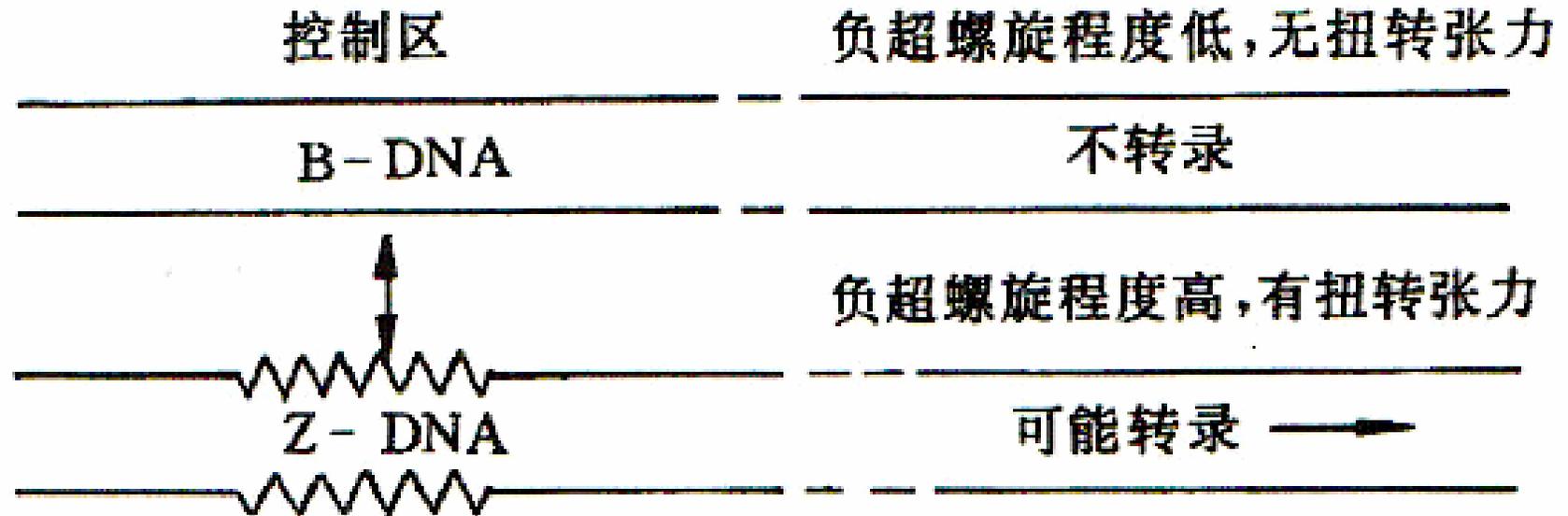


在邻近调控系统中，与调节区相邻的转录区被Z-DNA抑制，只有当Z-DNA转变为B-DNA后，转录才得以活化。



Z-DNA调控基因转录的两个模式

B. 远距离调控(转录区可在上千个碱基对以外)



在远距离调控系统中，Z-DNA可通过改变负超螺旋水平，决定聚合酶能否与模板链相结合而调节转录起始活性。



2.2.3 DNA的高级结构

DNA的高级结构是指DNA双螺旋进一步扭曲盘绕所形成的特定空间结构。超螺旋结构是DNA高级结构的主要形式，可分为正超螺旋与负超螺旋两大类，它们在特殊情况下可以相互转变，如：

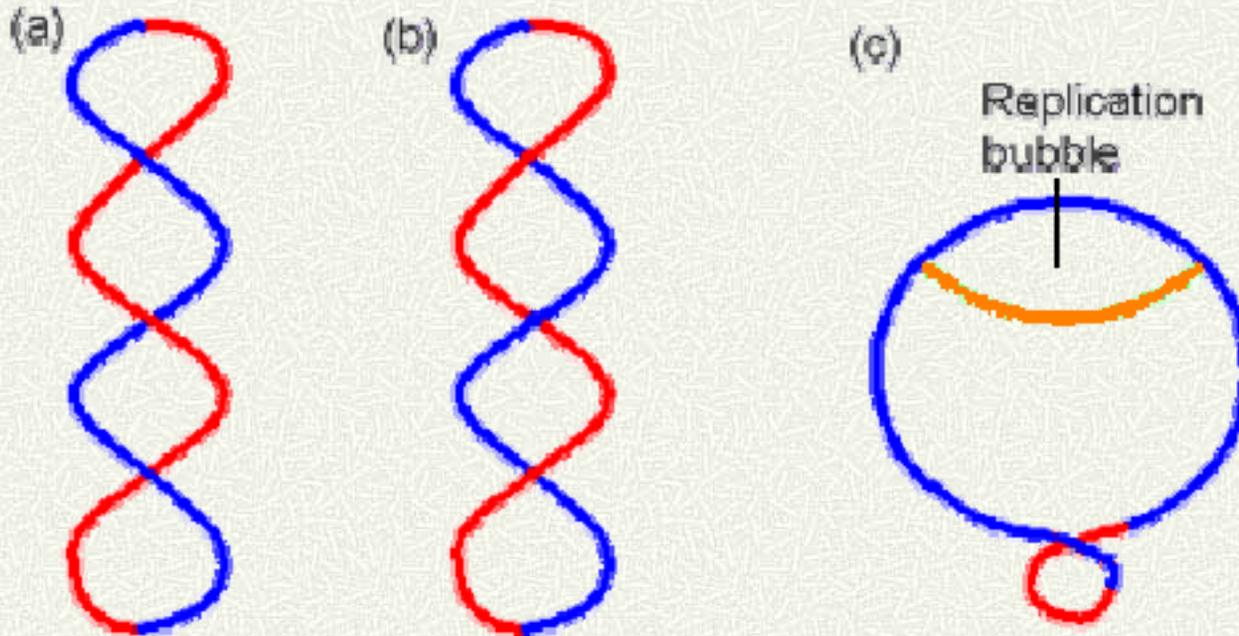
拓扑异构酶
溴乙锭

拓扑异构酶
溴乙锭

负超螺旋 —— 松弛DNA —— 正超螺旋



The structure of supercoils



(a) Positive supercoils - the front segment of a DNA molecule cross over the back segment from left to right.

(b) Negative supercoils.

(c) The positive supercoil in bacteria during DNA replication.

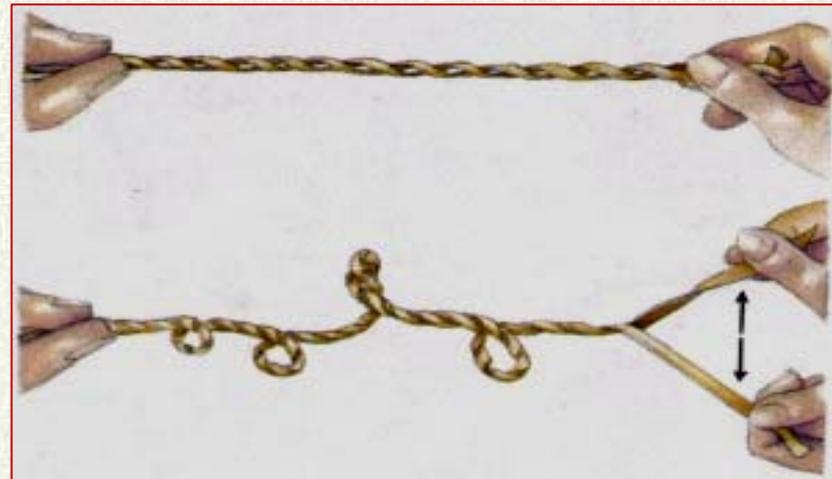
 **Topoisomerases**(拓扑异构酶) exist in cell to regulate the level of supercoiling of DNA molecules. The major role of topoisomerases is to prevent DNA tangling.

Type I topoisomerase: produces transient single-strand breaks in DNA, removes supercoils from DNA one at a time

Type II topoisomerase: produces transient double-strand breaks, removes supercoils two at a time.

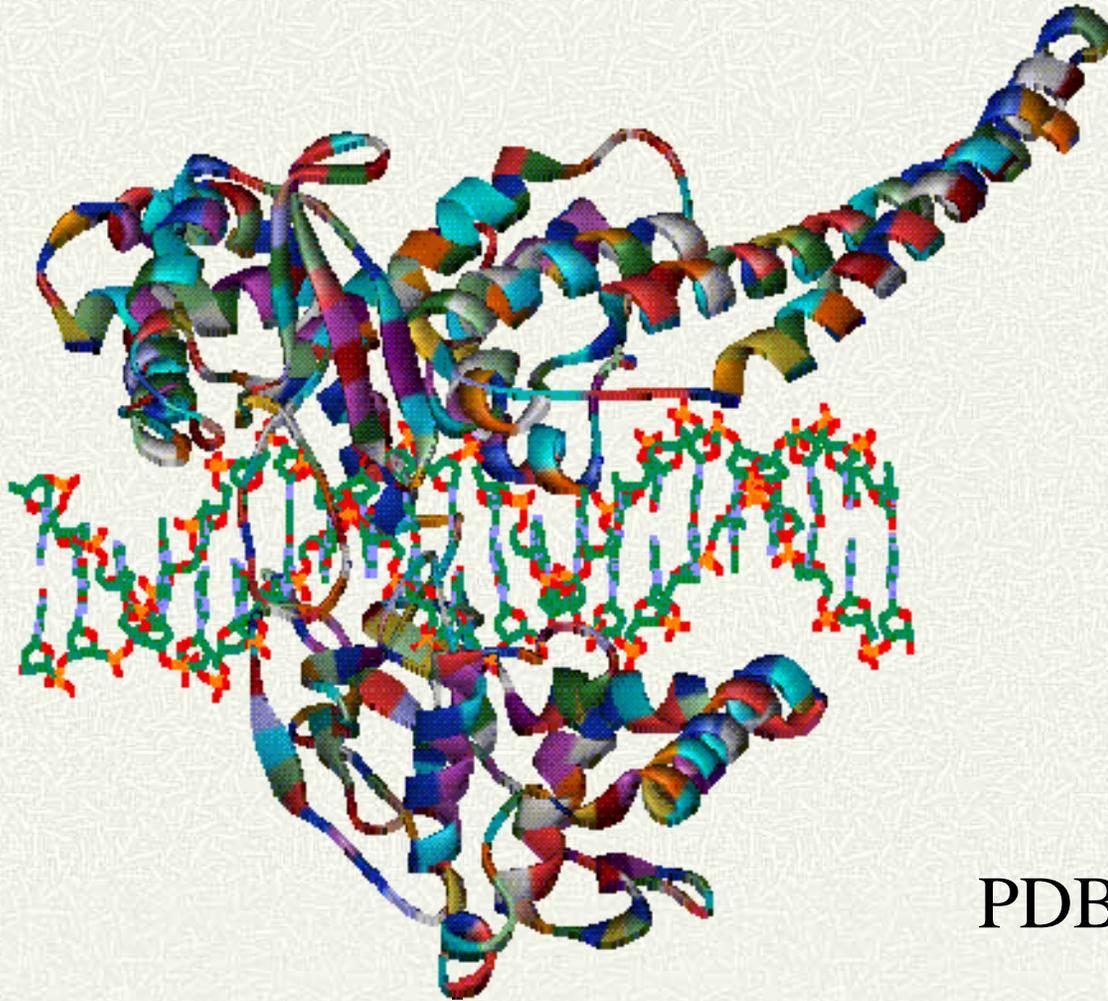
Gyrase: to remove the positive supercoils during DNA replication, (2) to introduce negative supercoils (one supercoil for 15-20 turns of the DNA helix) so that the DNA molecule can be packed into the cell.

Ethidium bromide (intercalator): locally unwinding of bound DNA, resulting in a reduction in twist and increase in writhe.





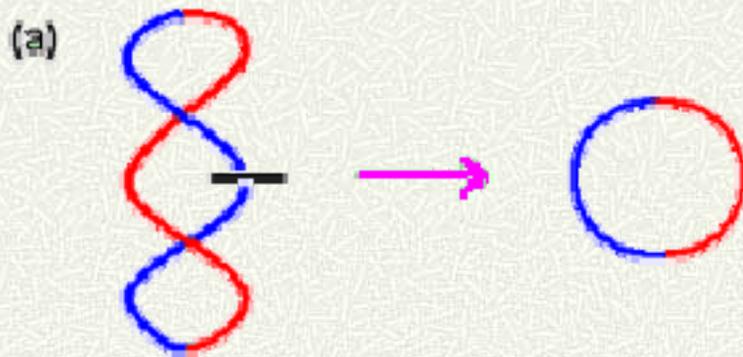
Structure of the Topo I/DNA complex



PDB ID = 1A36



The function of topo II



(a) To remove supercoils. This involves a double-strand break (indicated by a short line), allowing the tangled segment to pass through. The break is then resealed.



(b) To remove catenanes. The topo II makes a double-strand break in one DNA molecule (the blue one), allowing the other molecule to pass through. The break is then resealed.



研究**细菌质粒DNA**时发现，**天然状态下该DNA以负超螺旋为主**，稍被破坏即出现开环结构，两条链均断开则呈线性结构。

在电场作用下，相同分子质量的**超螺旋DNA比线性DNA迁移率大**，**线性DNA又比开环的DNA迁移率大**，以此可判断质粒结构是否被破坏。



DNA的复制

(DNA Replication)



2.3 DNA的复制

生命的遗传实际上是染色体DNA自我复制的结果，而染色体DNA的自我复制主要是通过**半保留复制(Semi-conservative)**来实现的，是一个以亲代DNA分子为模板合成子代DNA链的过程。

由于DNA是遗传信息的载体，因此亲代DNA必须以自身分子为模板来合成新的分子——准确地复制成两个拷贝，并分配到两个子代细胞中去，才能真正完成其遗传信息载体的使命。



DNA polymerases are enzymes that synthesize a daughter strand(s) of DNA (under direction from a DNA template). May be involved in repair or replication.

DNases are enzymes that attack bonds in DNA.

Endonucleases cleave bonds within a nucleic acid chain; they may be specific for RNA or for single-stranded or double-stranded DNA.

Exonucleases cleave nucleotides one at a time from the end of a polynucleotide chain; they may be specific for either the 5' or 3' end of DNA or RNA.

Parental strands of DNA are the two complementary strands of duplex DNA before replication.

Replication fork is the point at which strands of parental duplex DNA are separated so that replication can proceed.

Ribonucleases are enzymes that degrade RNA. Exo(ribo)nucleases work progressively, typically degrading one base at a time from the 3' end toward the 5' end.

Endo(ribo)nucleases make single cuts within the RNA chain.

RNA polymerases are enzymes that synthesize RNA using a DNA template (formally described as DNA-dependent RNA polymerases).

RNAases are enzymes that degrade RNA.

Semiconservative replication is accomplished by separation of the strands of a parental duplex, each then acting as a template for synthesis of a complementary strand.



由于DNA分子由两条多核苷酸链组成，两条链上的碱基——G只能与C相配对，A只能与T相配对，所以，两条链是互补的，一条链上的核苷酸排列顺序决定了另一条链上的核苷酸排列顺序。



DNA的半保留复制

(**semi-conservative replication**)

DNA在复制过程中，每条链分别作为模板合成新链，产生互补的两条链。这样新形成的两个**DNA**分子与原来**DNA**分子的碱基顺序完全一样。因此，每个子代分子的一条链来自亲代**DNA**，另一条链则是新合成的，这种复制方式被称为**DNA**的半保留复制。

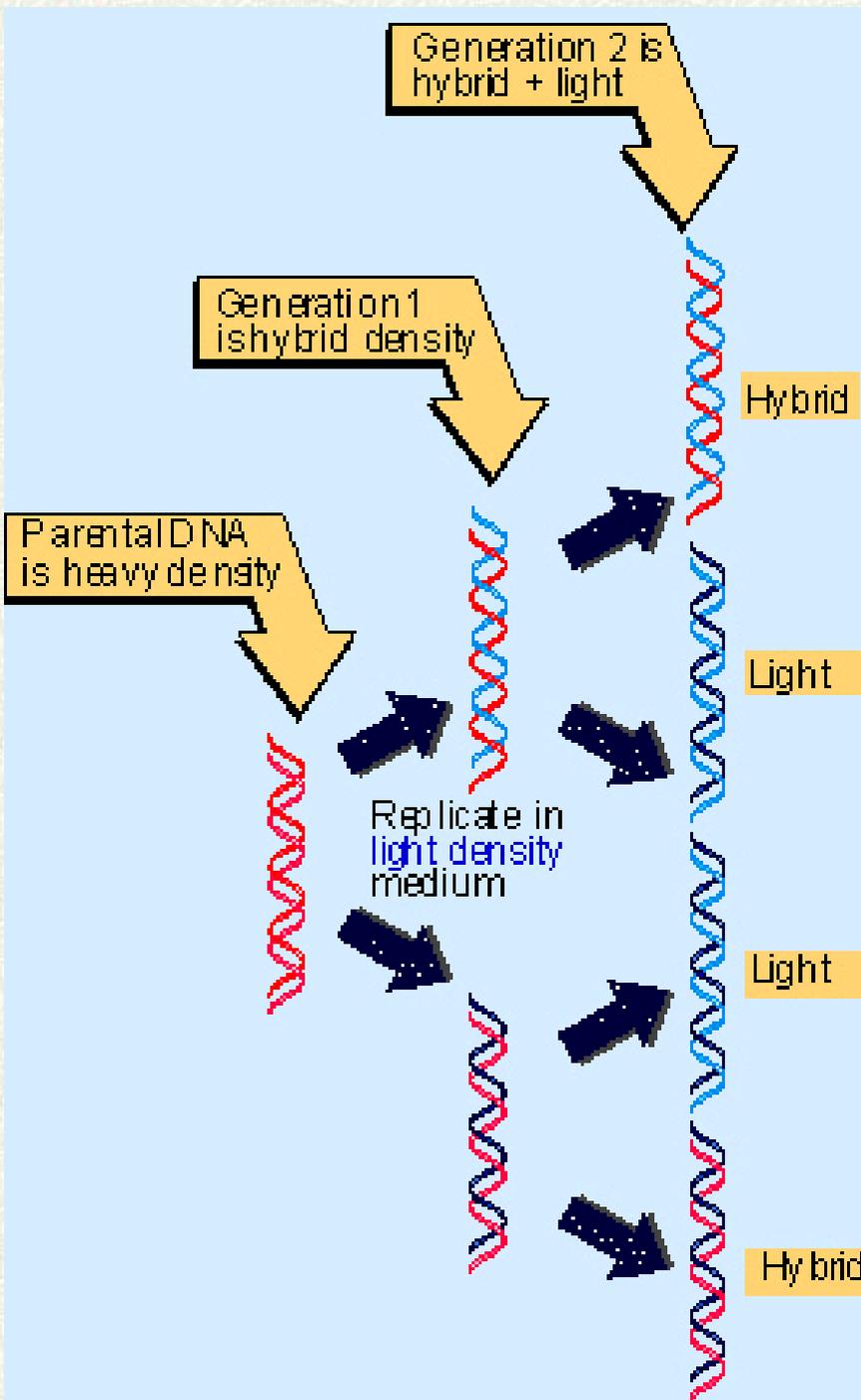
DNA的半保留复制

Semi-conservative mechanism

^{15}N labeling experiment

1958年，Meselson 和 Stahl研究了经 ^{15}N 标记3个世代的大肠杆菌DNA，首次证明了DNA的半保留复制。

DNA的半保留复制保证了DNA在代谢上的稳定性，与DNA的遗传功能相符合。





Lagging strand synthesis

Previous fragment

Last fragment

Single strand

Parental DNA



Leading strand synthesis

Nucleotides added continuously to 3' end

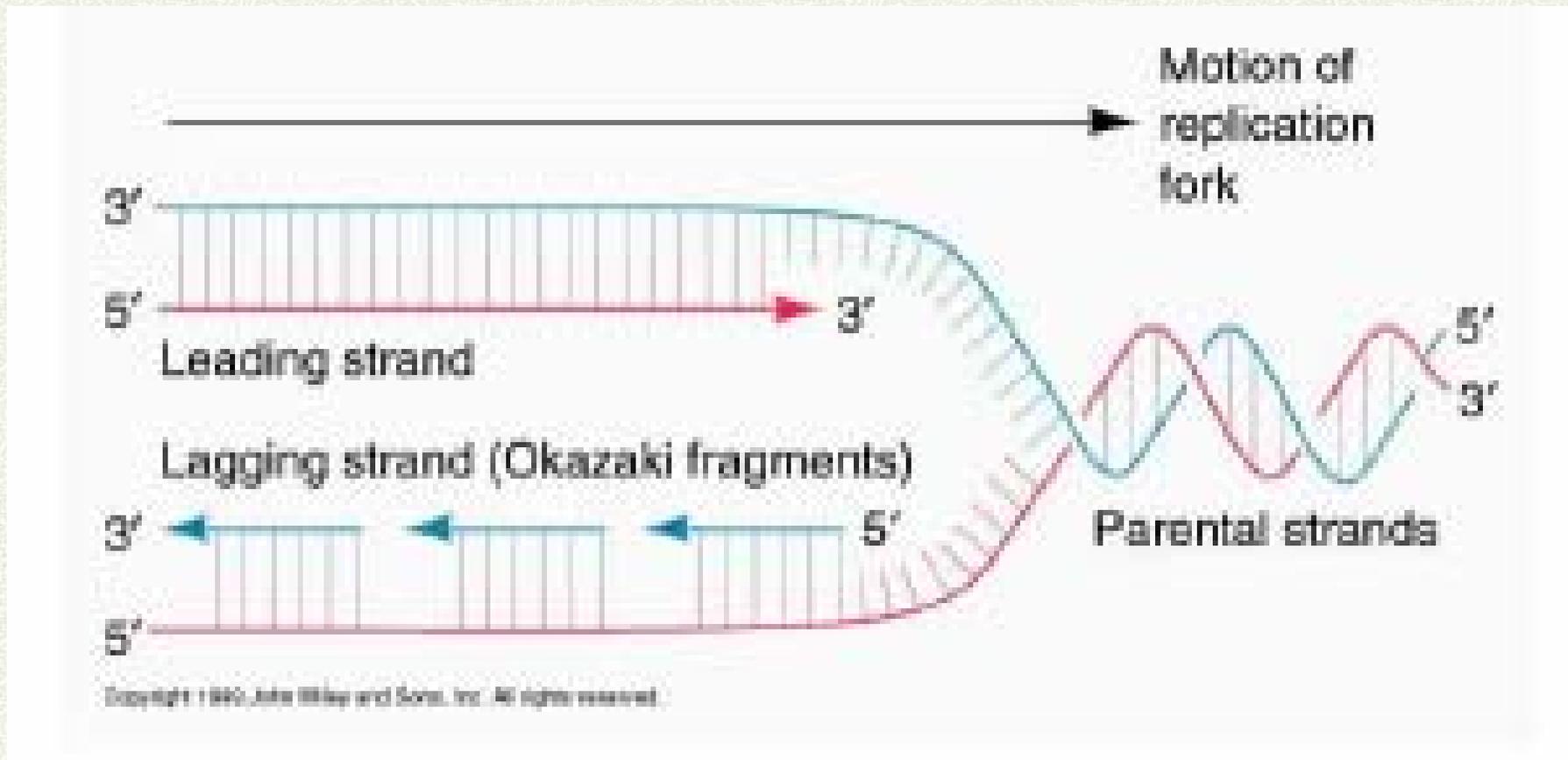


DNA的半不连续复制

(semidiscontinuous replication)

由于DNA双螺旋的两条链是反向平行的，因此在复制叉附近解开的DNA链一条是5'→3'方向，另一条是3'→5'方向，两个模板极性不同。而所有已知DNA聚合酶的合成方向都是5'→3'，两条链无法同时进行复制。为了解释DNA的等速复制现象，日本学者冈崎（Okazaki）等提出了DNA的半不连续复制模型。

半不连续复制 Semi-discontinuous replication: 冈崎片段 Okazaki fragment



[³H] Thymidine pulse-chase labeling and alkaline sucrose gradient: discovery of semi-discontinuous replication

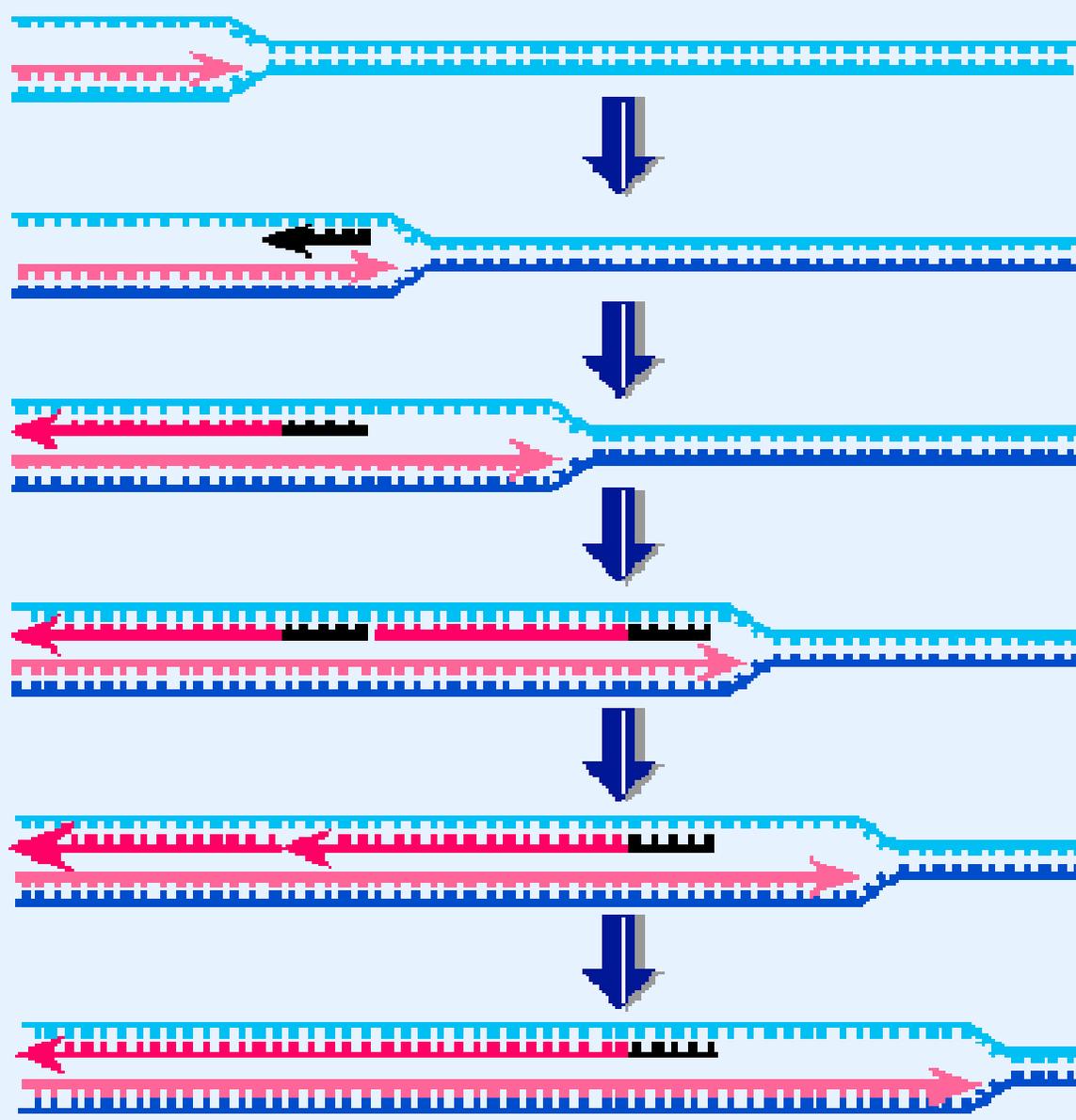


1、用 ^3H 脱氧胸苷短时间标记后提取DNA，得到不少平均长度为2-3kb DNA片段。

2、用DNA连接酶温度敏感突变株进行实验，在连接酶不起作用的温度下，有大量小片段累积，说明复制过程中至少有一条链首先合成较短的片段，然后再生成大分子DNA。

3、前导链的连续复制和滞后链的不连续复制在生物界是有普遍性的，因而称之为DNA的半不连续复制。

Synthesis of Okazaki fragments requires priming, extension, removal of RNA, gap filling, and nick ligation



Primase
synthesizes RNA

DNA polymerase III
extends RNA primer
into Okazaki fragment

Next Okazaki
fragment is
synthesized

DNA polymerase I
uses nick translation
to replace RNA primer
with DNA

Ligase seals the nick



2.3.2 复制的起点、方向和速度

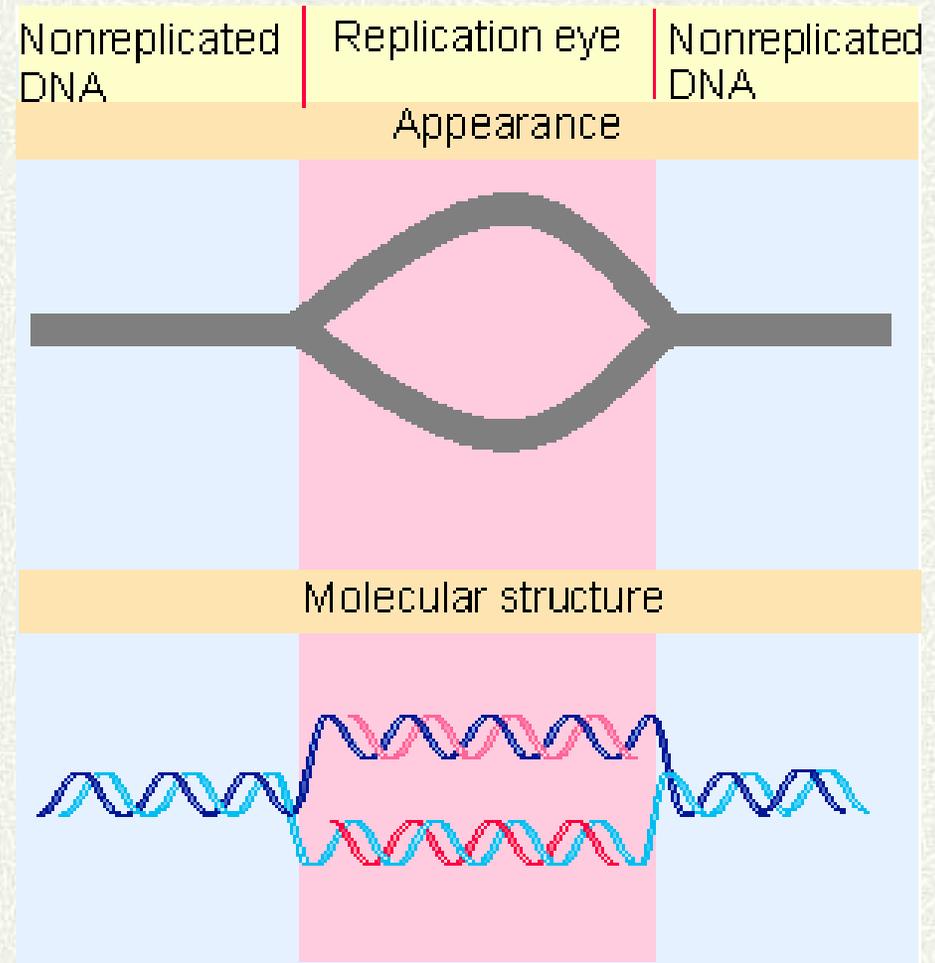
复制时，双链DNA要解开成两股链分别进行，所以，复制起点呈叉子形式，被称为**复制叉 (Replication fork)**。

DNA的复制是由固定的起始点开始的。一般把生物体的复制单位称为**复制子 (replicon)**，一个复制子只含一个复制起点。



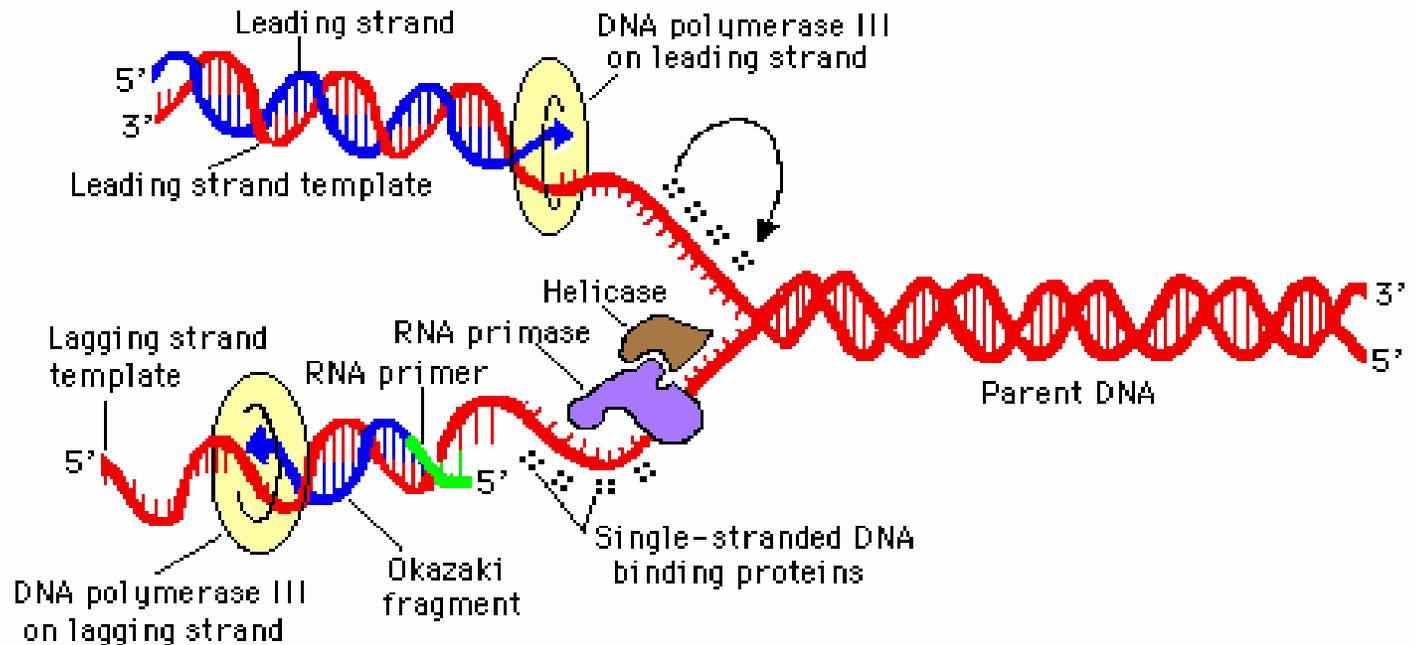
Replicon is a unit of the genome in which DNA is replicated; contains an origin for initiation of replication.

Replication fork is the point at which strands of parental duplex DNA are separated so that replication can proceed.





Collaboration of Proteins at the Replication Fork





A General Model for DNA Replication

1. The DNA molecule is **unwound** and prepared for synthesis by the action of DNA gyrase, DNA helicase and the single-stranded DNA binding proteins.
2. A free 3'OH group is required for replication, but when the two chains separate no group of that nature exists. **RNA primers** are synthesized, and the free 3'OH of the primer is used to begin replication.
3. The replication fork moves in one direction, but DNA replication only goes in the 5' to 3' direction. This paradox is resolved by the use of Okazaki fragments. These are short, discontinuous replication products that are produced off the **lagging strand**. This is in comparison to the continuous strand that is made off the **leading strand**.
4. The final product does not have RNA stretches in it. These are removed by the **5' to 3' exonuclease** action of Polymerase I.
5. The final product does not have any gaps in the DNA that result from the removal of the RNA primer. These are filled in by the **5' to 3' polymerase** action of DNA Polymerase I.
6. DNA polymerase does not have the ability to form the final bond. **This is done by** *NLPE & PGE*



DNA Replication Bubble

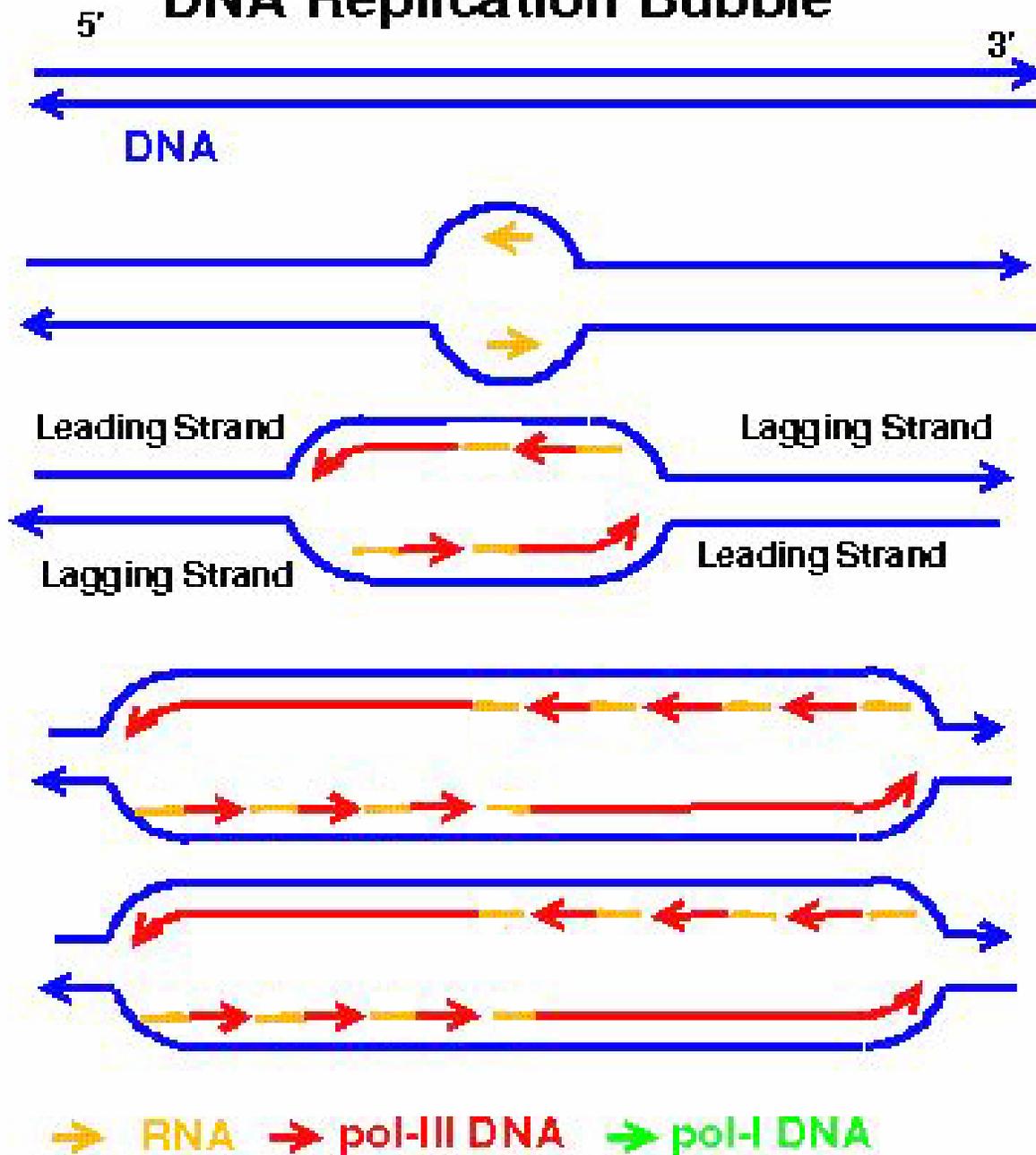




表2-9 部分生物复制子的比较

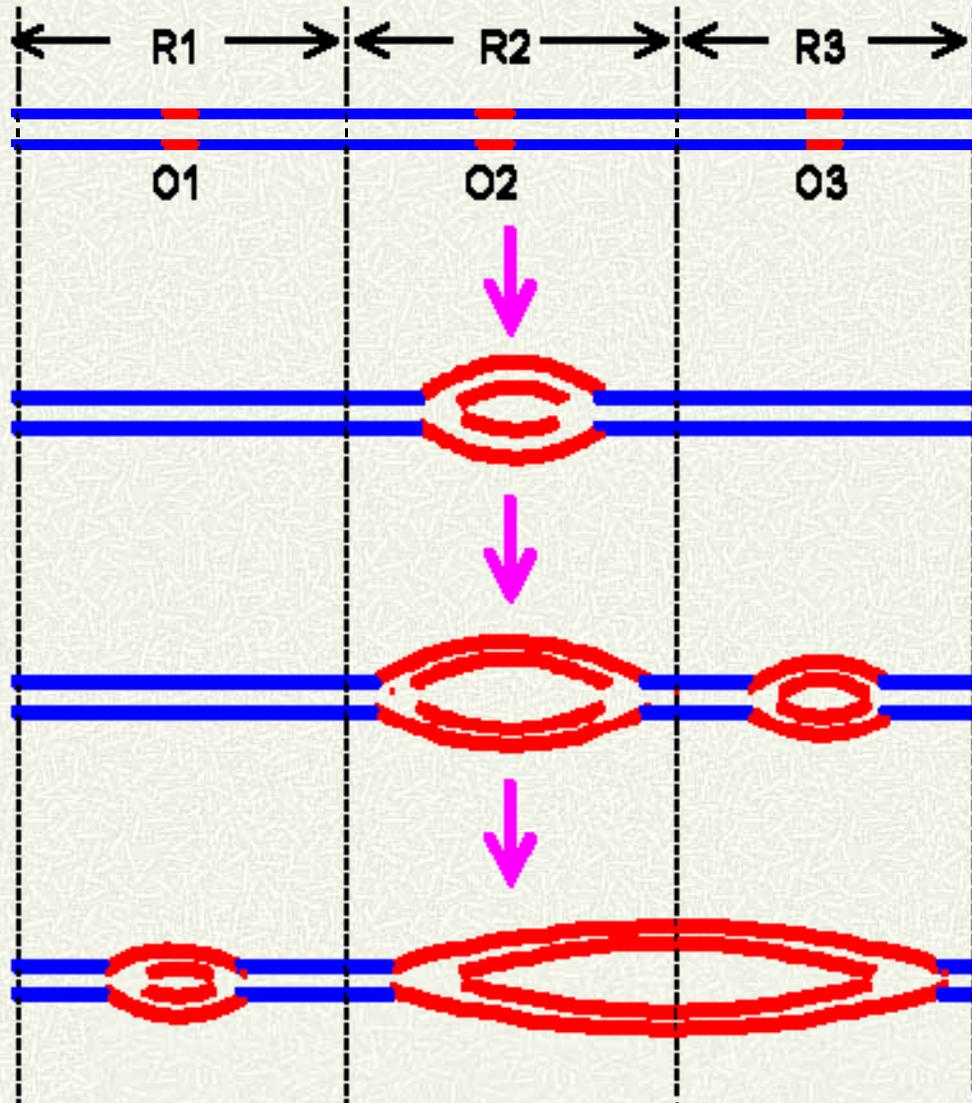
物种	细胞内复制子数目 (个)	平均长度/kb	复制子移动速度 (bp·min ⁻¹)
大肠杆菌	1	4 200	50 000
酵母	500	40	?
果蝇	3 500	40	2 600
蟾蜍	15 000	200	500
蚕豆	35 000	300	?



细菌、病毒和线粒体的DNA分子都是作为单个复制子完成复制的；真核生物基因组可以同时多个复制起点上进行双向复制，也就是说它们的基因组包含多个复制子。



Multiple eukaryotic replicons

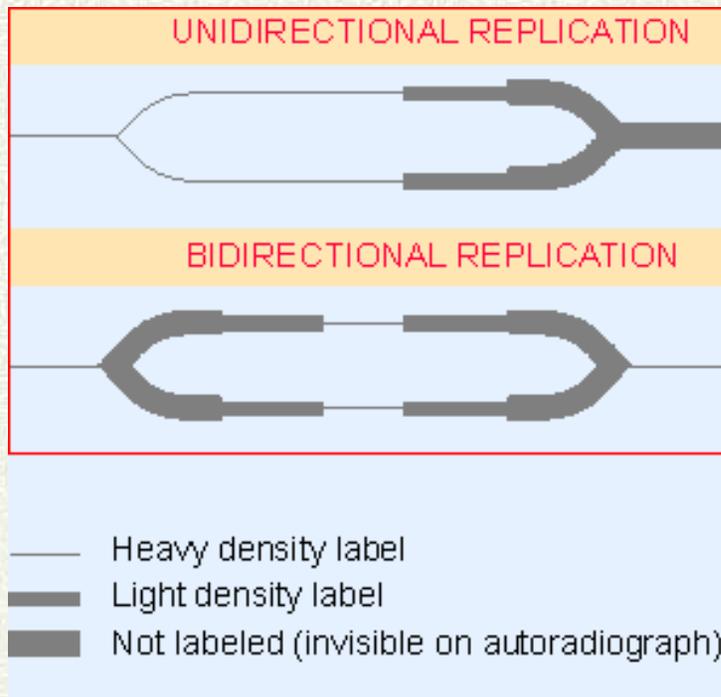


O1, O2, and O3 are replication origins, each serving a region called **replicon** (R1, R2, and R3).

DNA replication involves unwinding of the double helix. New strands are synthesized by DNA polymerases using the old strands as template. Unwinding of a DNA molecule looks like a "fork" growing in one direction. The region being replicated looks like a bubble called the "**replication bubble**" (in red).



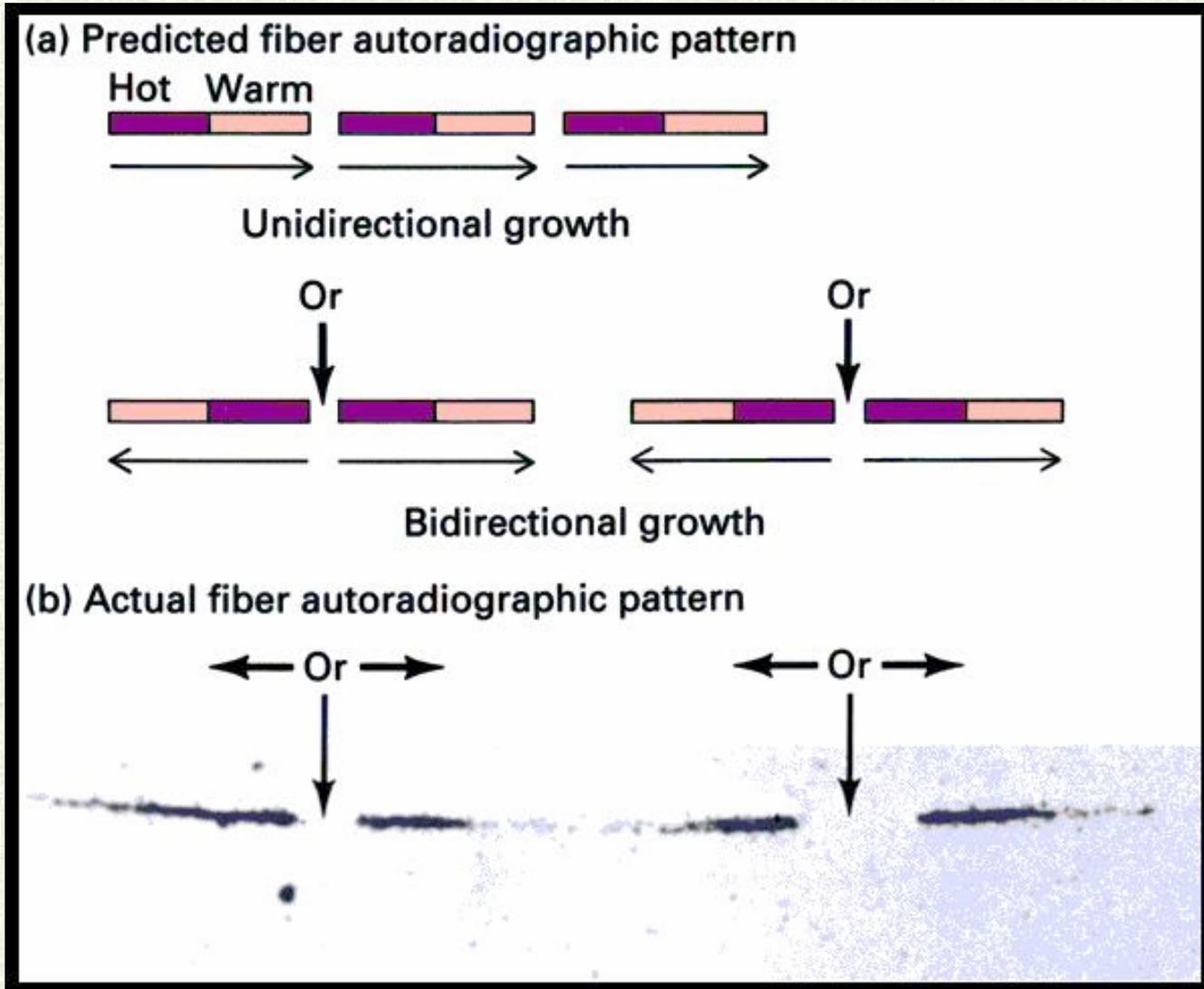
实验结果表明，无论是原核生物还是真核生物，DNA的复制主要是从固定的起始点以双向等速复制方式进行的。复制叉以DNA分子上某一特定顺序为起点，向两个方向等速生长前进。



放射自显影技术表明染色体复制子为双向复制。

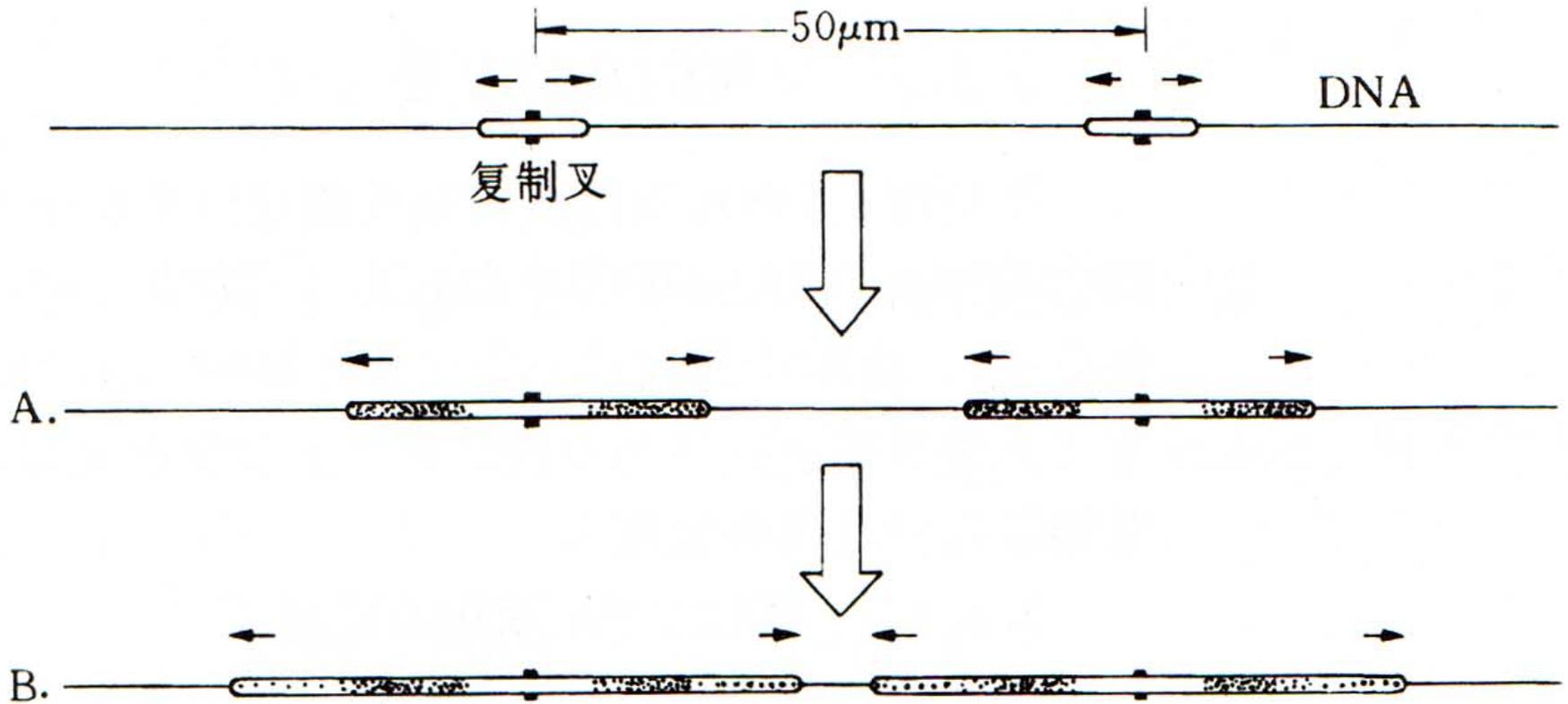


Pulse with radiolabeled nucleotide; chase with cold nucleotide. Then do autoradiography





放射性标记实验证明DNA的复制是从固定的起始点双向等速进行的



- A. DNA仅用 $[^3\text{H}]$ 胸腺嘧啶标记10 min 即压X光片；
- B. DNA在 $[^3\text{H}]$ 标记10 min 后又继续非标记合成10 min，导致X光片上的复制叉变长，银颗粒分布广而稀少。



2.3.3 复制的几种主要方式

2.3.3.1 线性DNA双链的复制

2.3.3.2 环状DNA双链的复制

(1) θ 型

(2) 滚环型(rolling circle)

(3) D-环型(D-loop)

环状DNA双链的复制- θ 型

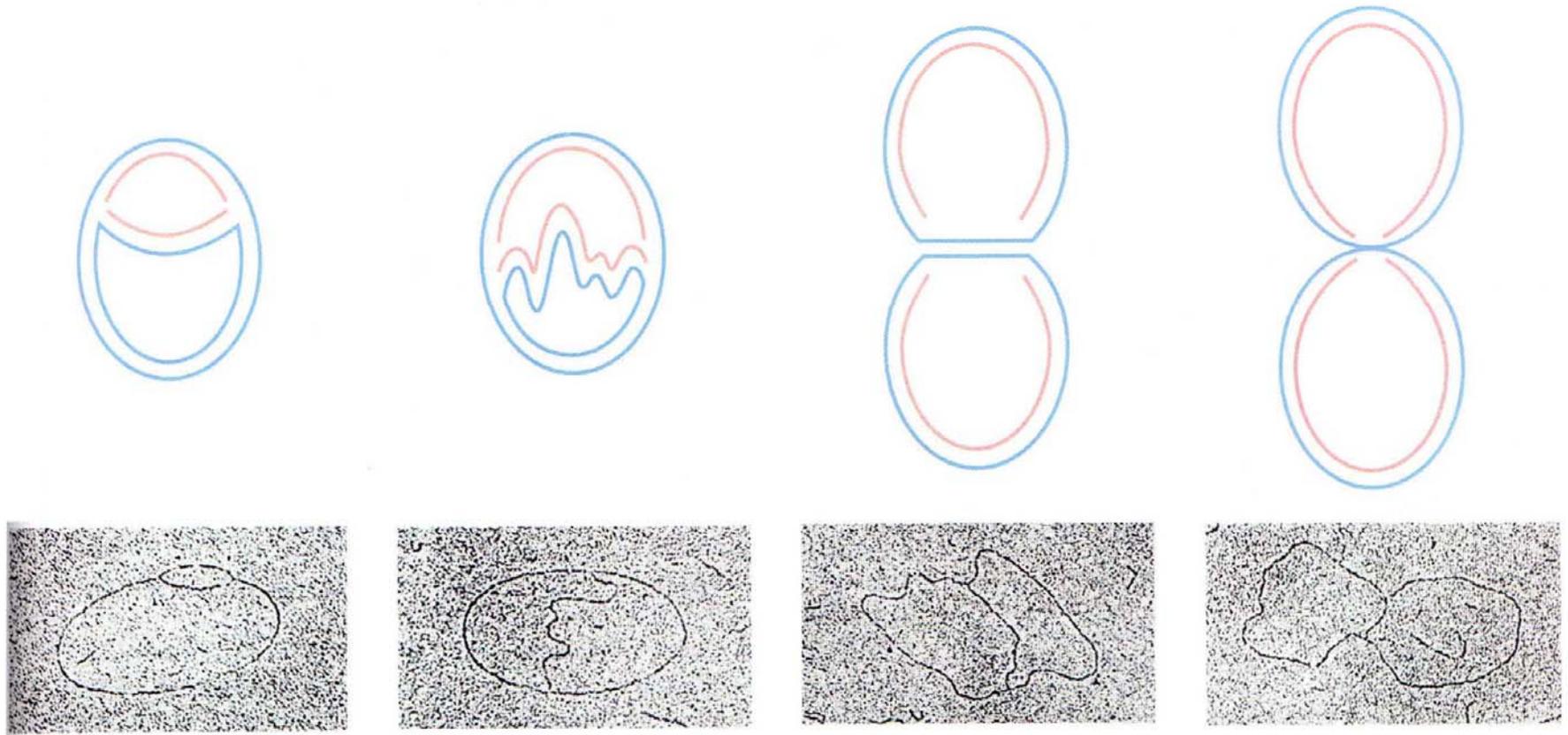
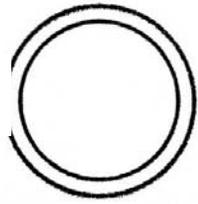


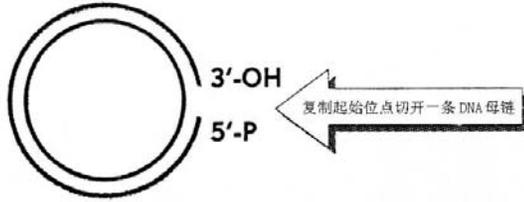
图2-19 大肠杆菌质粒DNA双向复制的模式与实验验证。上： θ 形复制模式图；下： ^3H 标记质粒DNA合成图。



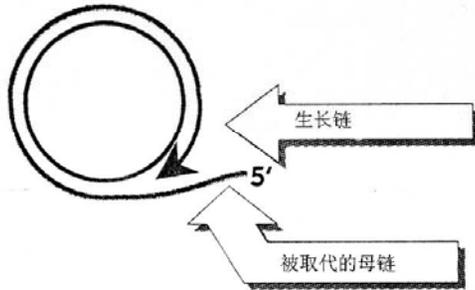
双链环状DNA



复制起始于某一条DNA链上



新生DNA链延伸，不断取代母链



复制一圈，产生单位长度的线性DNA链



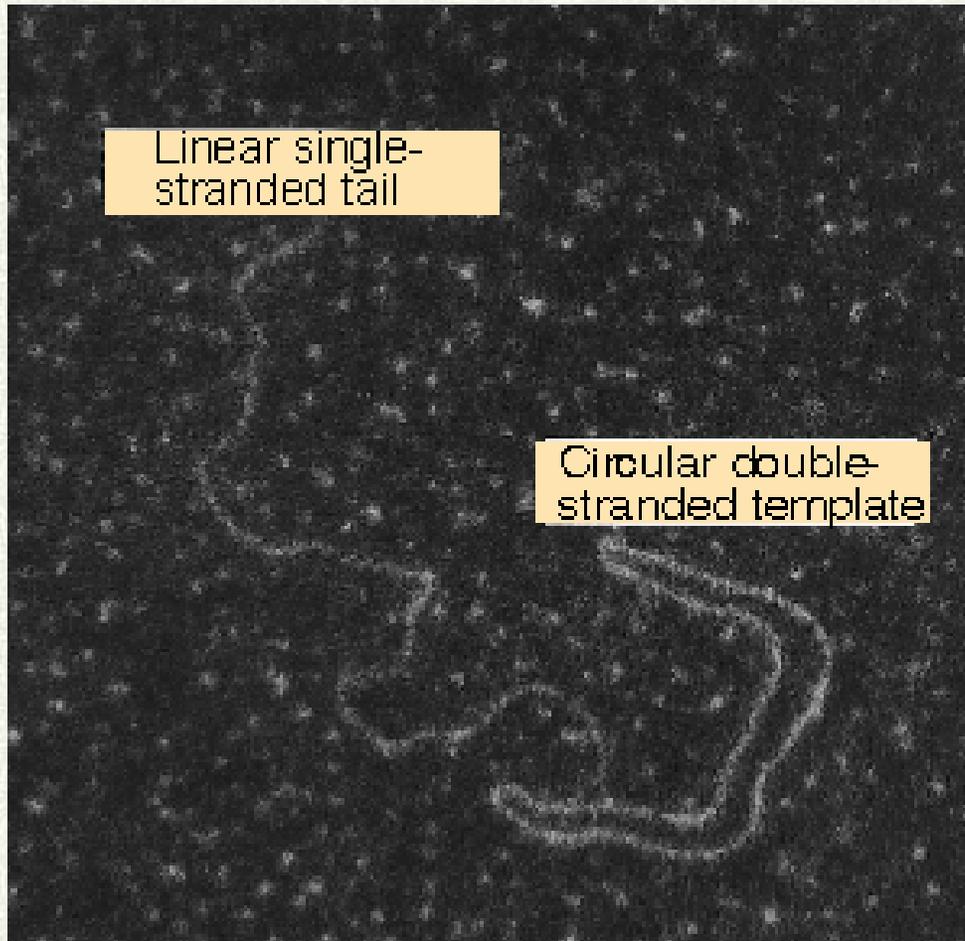
若复制继续进行，可产生多单元线性DNA链



图2-20 环状DNA可以通过滚环式复制产生多单元单链DNA



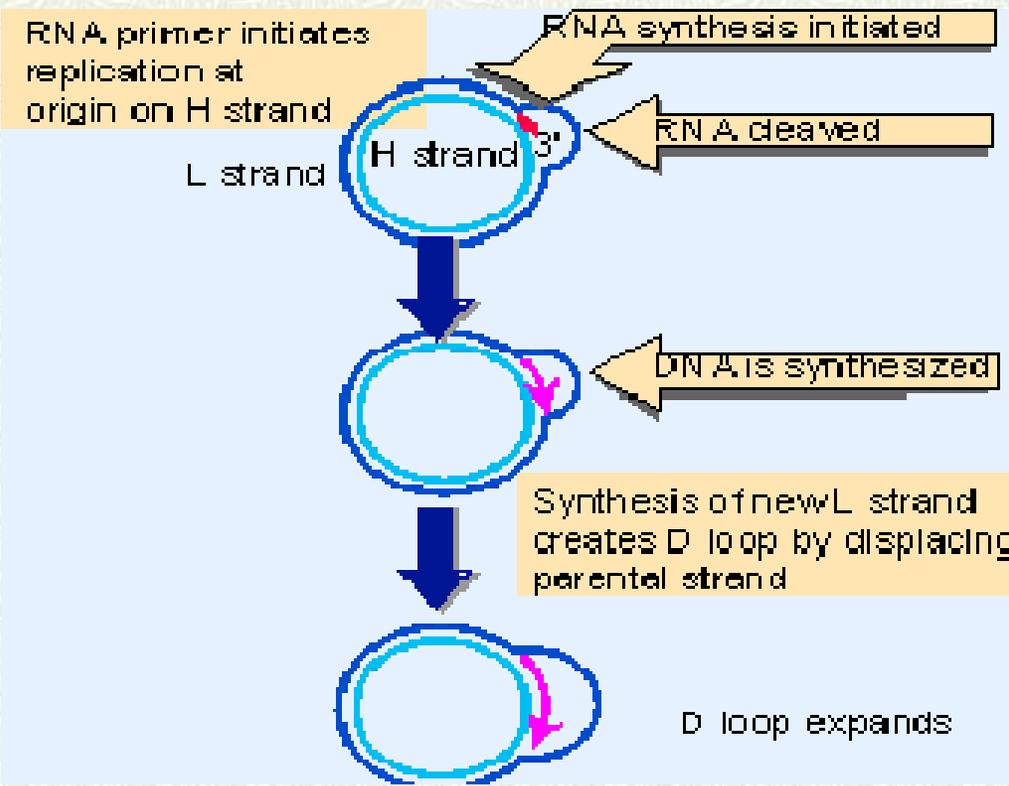
Rolling circles produce multimers of a replicon



A rolling circle appears as a circular molecule with a linear tail by electron microscopy



D loops



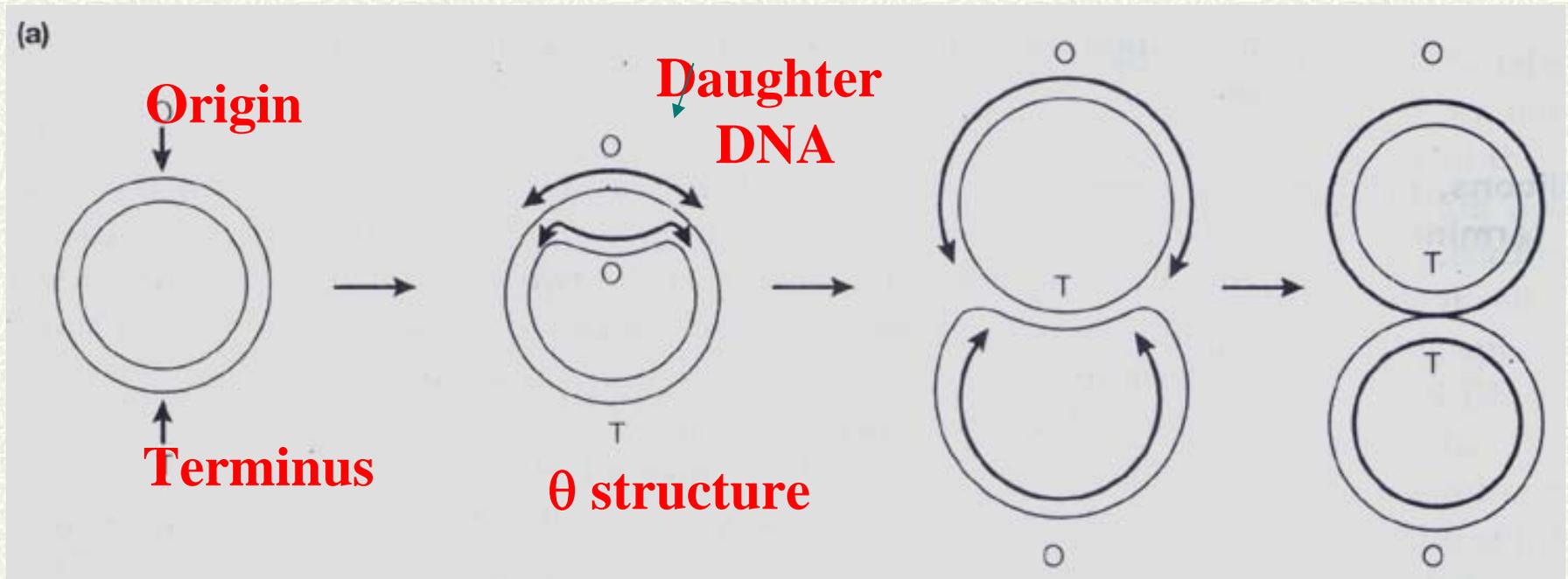
D loop 是单向复制的特殊方式。首先在动物线粒体中被发现。一条短RNA与一条DNA互补，取代了该区域原来的互补的DNA链，使得两条链的合成高度不对称，一条链上迅速合成出互补链，另一条链则为游离的单环。



原核生物和真核生物 DNA的复制特点



2.4.1 原核生物DNA的复制特点



大肠杆菌基因组以**双链环状DNA分子**的形式存在，其DNA复制的**中间产物**可形成一个 θ ，复制从定点开始**双向等速**进行，复制起始区位于其遗传图的84min附近。复制起始后，两个复制叉在距起始点 180° 处会合。

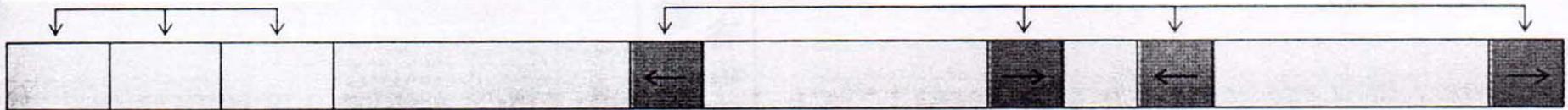


大肠杆菌DNA复制起始点

(*oriC*) 保守序列分布图

3X13bp直接重复序列

DnaA结合位点, 4 x 9bp



保守序列

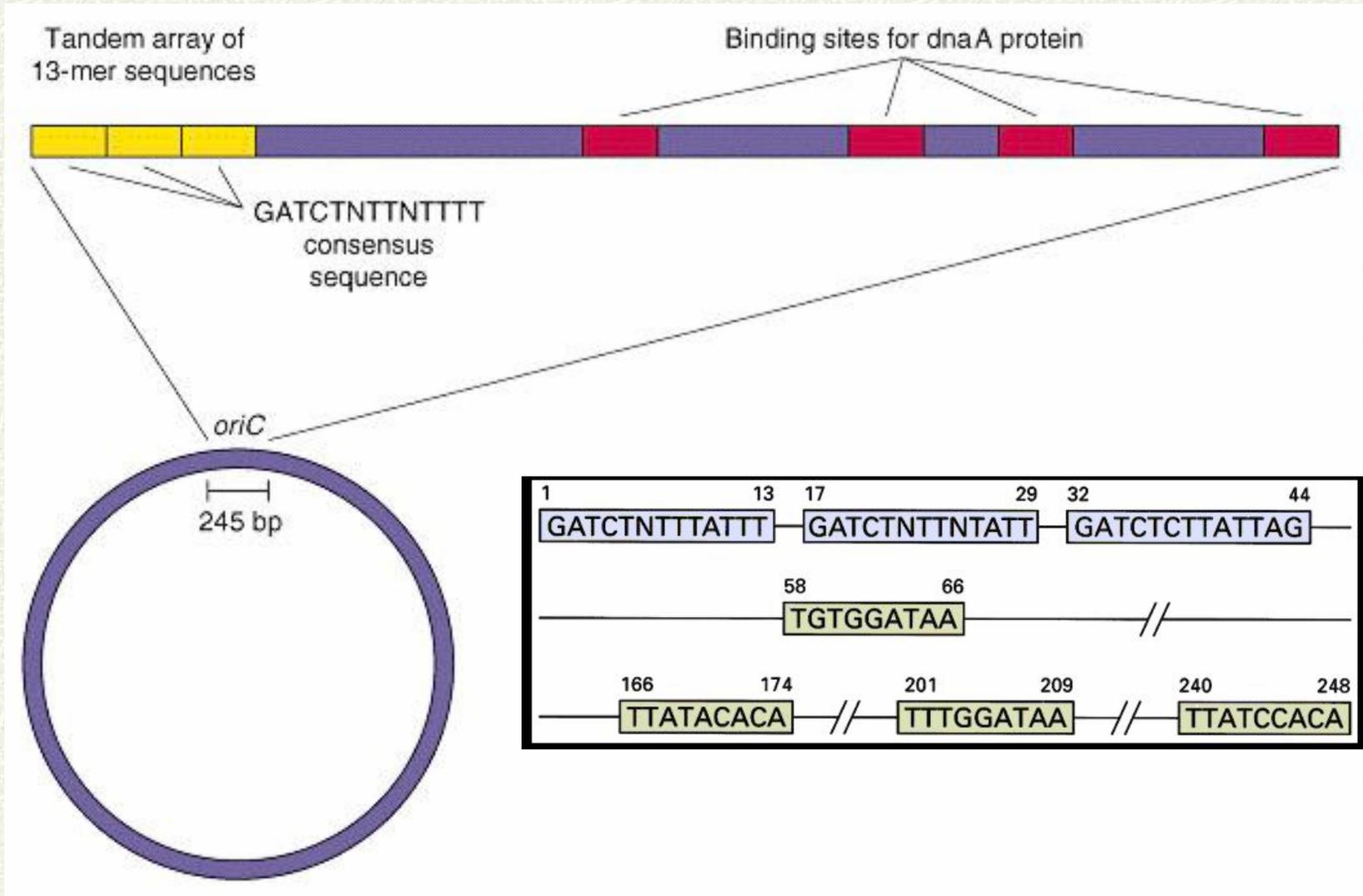
保守序列

GATCTNTTNTTT

TTATCCACA

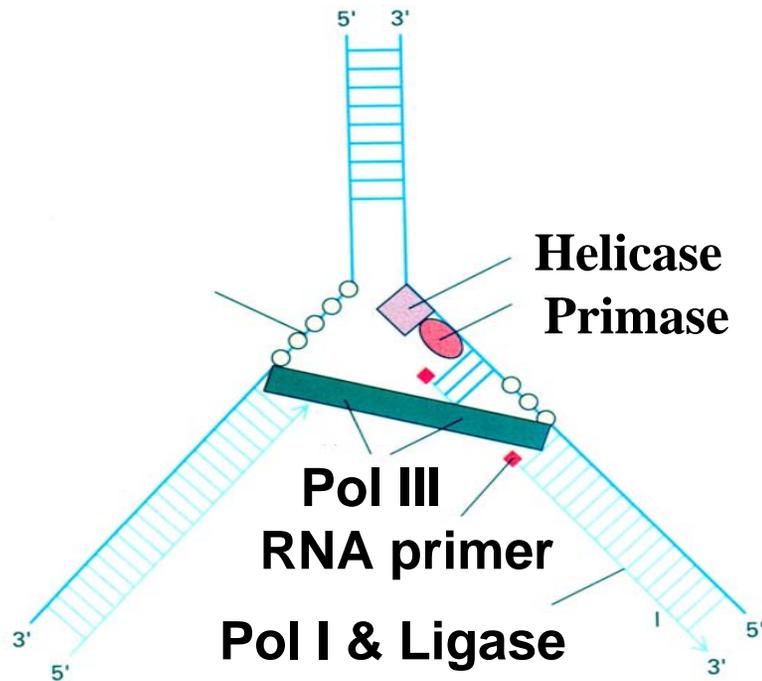


OriC in E. coli chromosomal DNA





2.4.1.1 DNA双螺旋的解旋



首先，在**拓扑异构酶I**的作用下解开负超螺旋，并与**解链酶**共同作用，在复制起点处**解开双链**。解链过程中必需有**SSB**蛋白来稳定已解开的单链，以保证该局部结构不会恢复成双链。

接着，由**引发酶**等组成的引发体迅速作用于两条单链**DNA**上。不论是**前导链**还是**滞后链**，都需要一段**RNA引物**以开始子链**DNA**的合成。

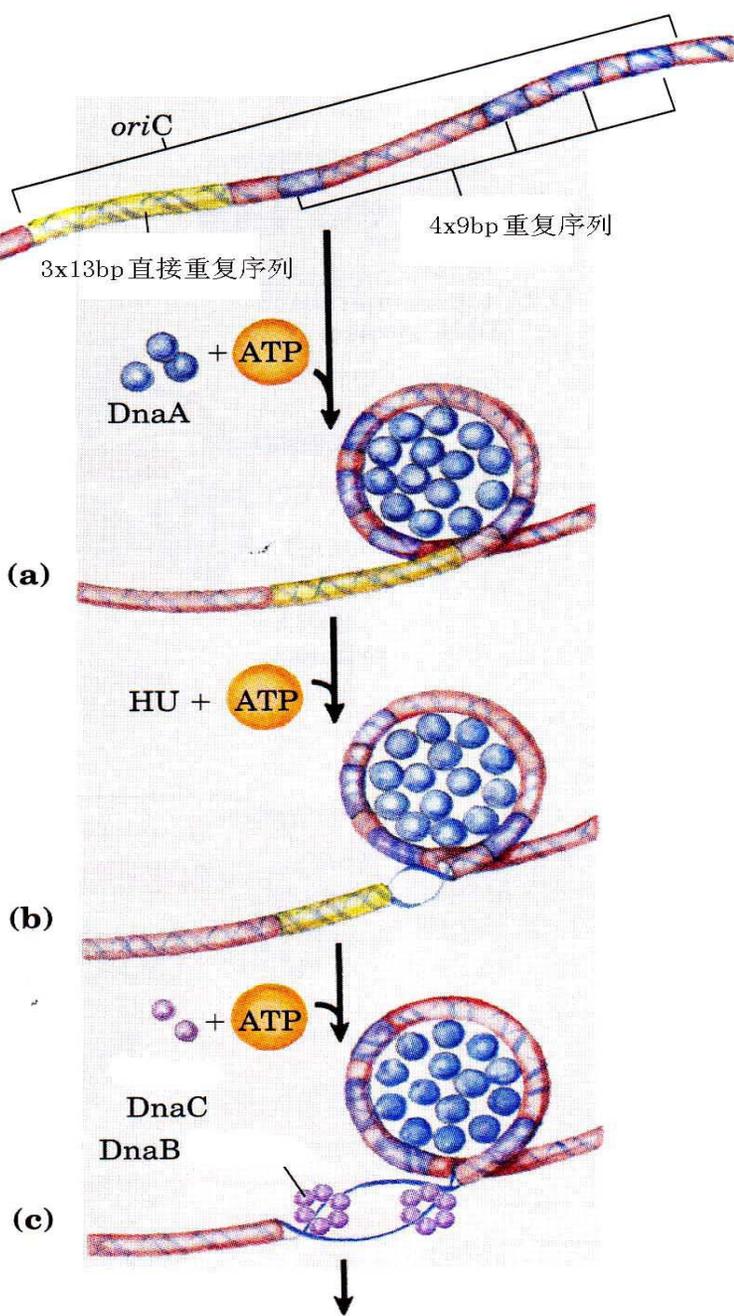
由大肠杆菌oriC复制起始点处引发的DNA复制过程

大约20个DnaA蛋白在ATP的作用下与oriC处的4各9 bp保守序列相结合。

在Hu蛋白和ATP的共同作用下，DnaA复制起始复合物使3x13 bp直接重复序列变形，形成开链。

DnaB（解链酶）六体分别与单链DNA相结合（需要DnaC的帮助）进一步解开DNA双链。

与引物结合，起始DNA复制。





(1) 单链结合蛋白 (SSB蛋白)

T4噬菌体的32 kDa蛋白可以结合单链DNA，它还能使双链DNA在远低于解链温度时分开。

大部分原核SSB蛋白与DNA结合时还表现出协同效应：如第一个SSB蛋白结合到DNA上去的能力为1，第二个SSB结合能力则可高达 10^3 。SSB蛋白的作用是稳定单链DNA。

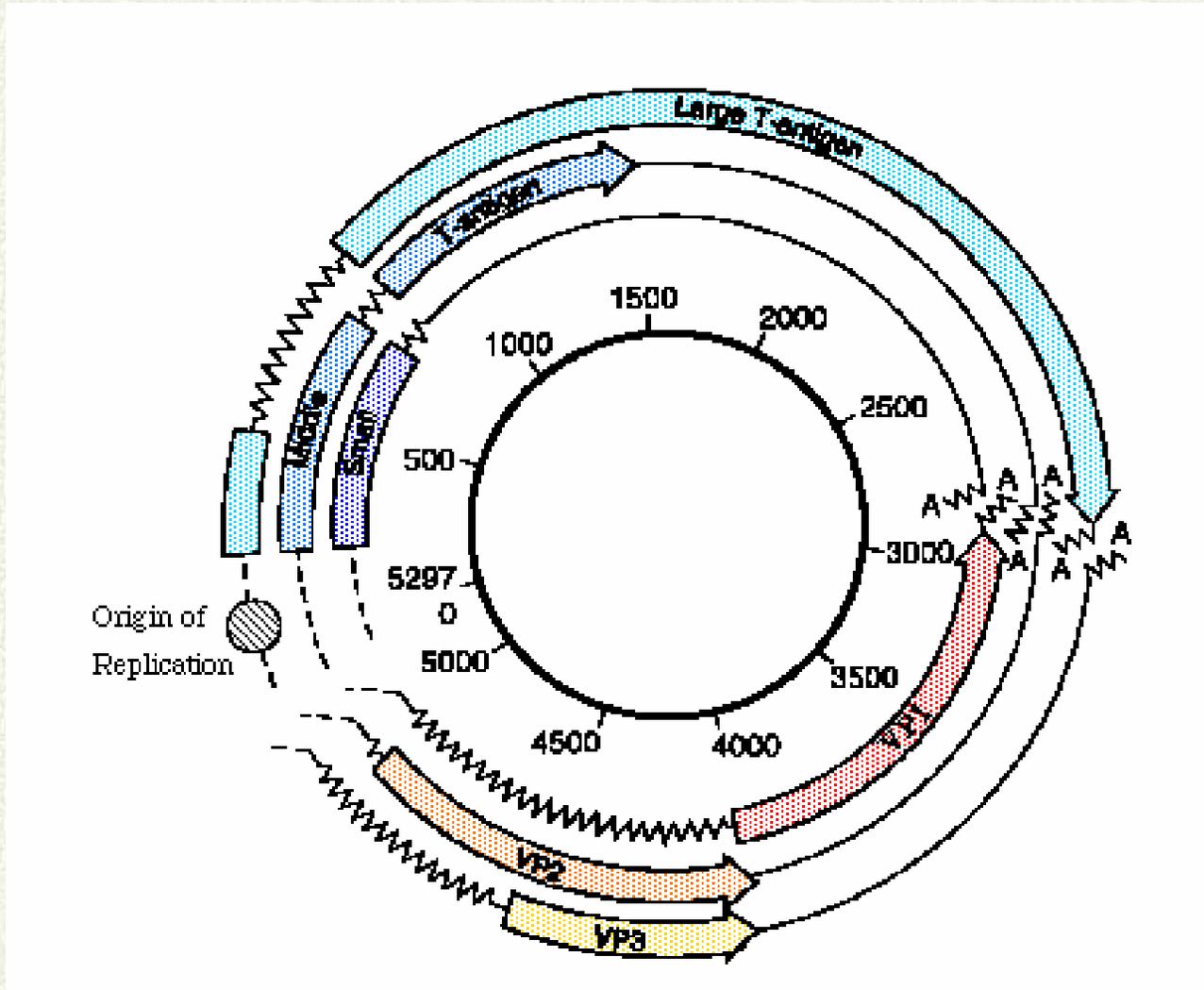


(2) DNA解链酶 (DNA helicase)

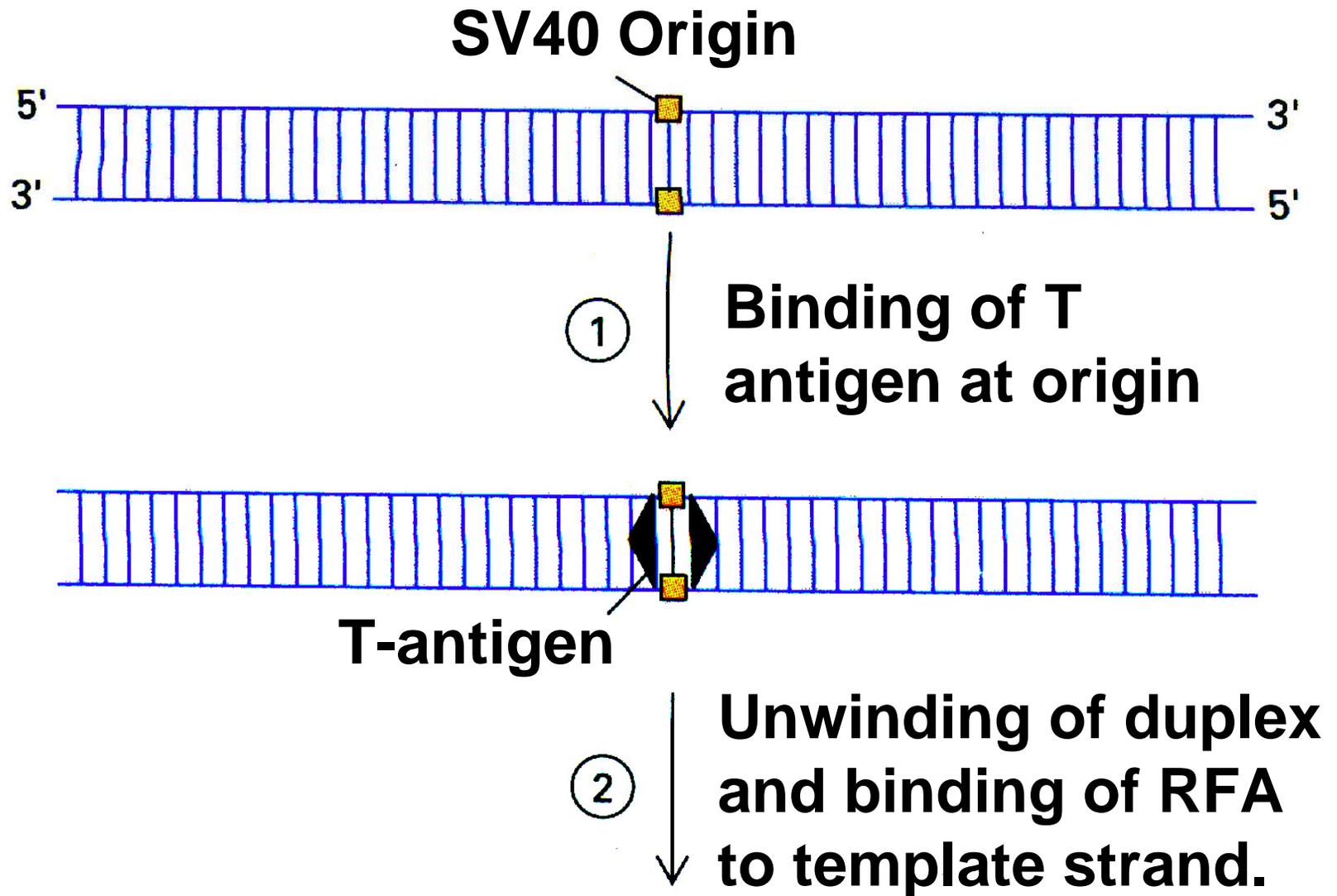
大部分DNA解链酶都沿滞后链模板的5'→3'方向并随着复制叉的前进而移动，只有Rep蛋白沿前导链模板的3'→5'方向移动。因此，Rep蛋白和其他DNA解链酶分别在DNA的两条母链上协同作用，**解开双链DNA**。



SV40: a polyomavirus that is very useful for studying eukaryotic replication enzymes

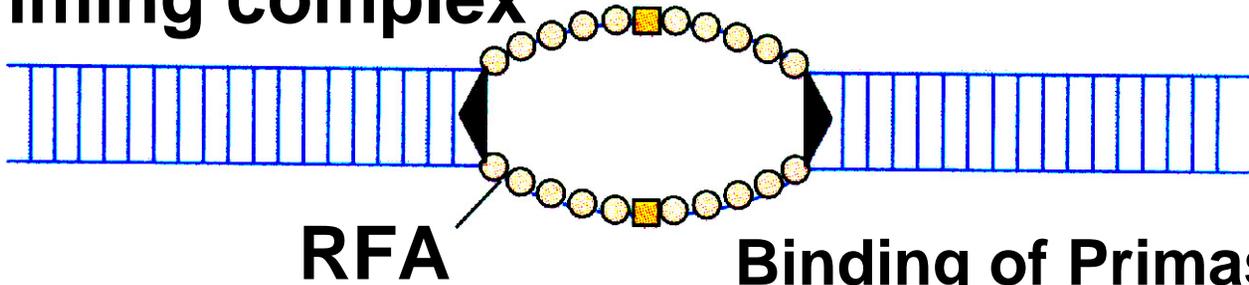


Replication of SV40 DNA





Formation of prepriming complex



RFA

Binding of Primase-Pol α

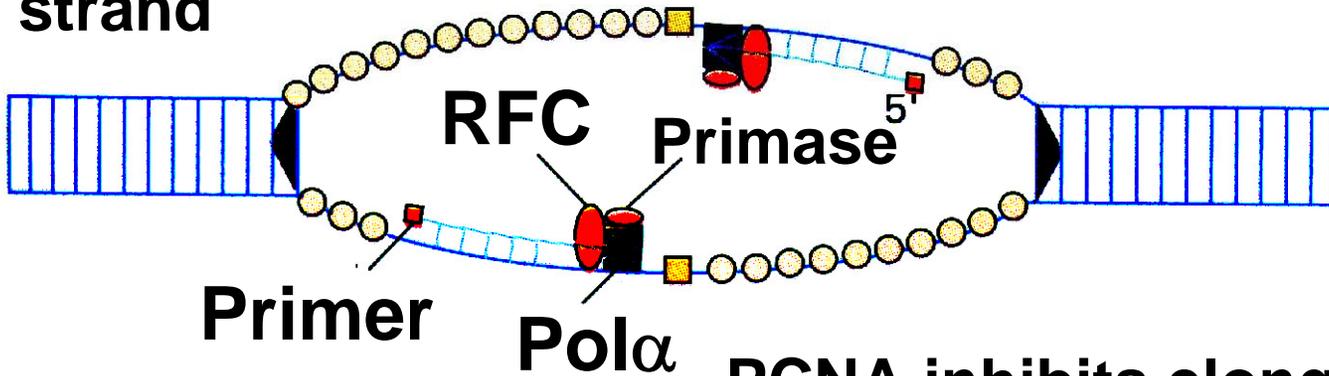
Primer synthesis

Binding of RFC

DNA synthesis

3

Synthesis of primer and 5' end of leading strand



RFC

Primase

5'

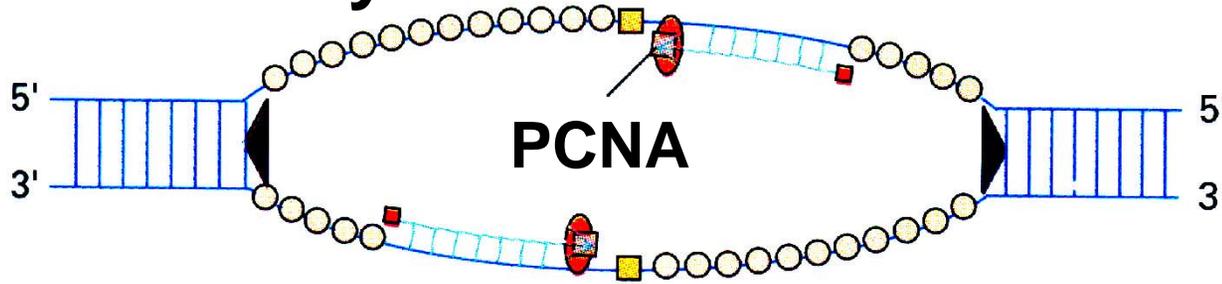
Primer

Pol α

PCNA inhibits elongation and displaces primase-Pol α

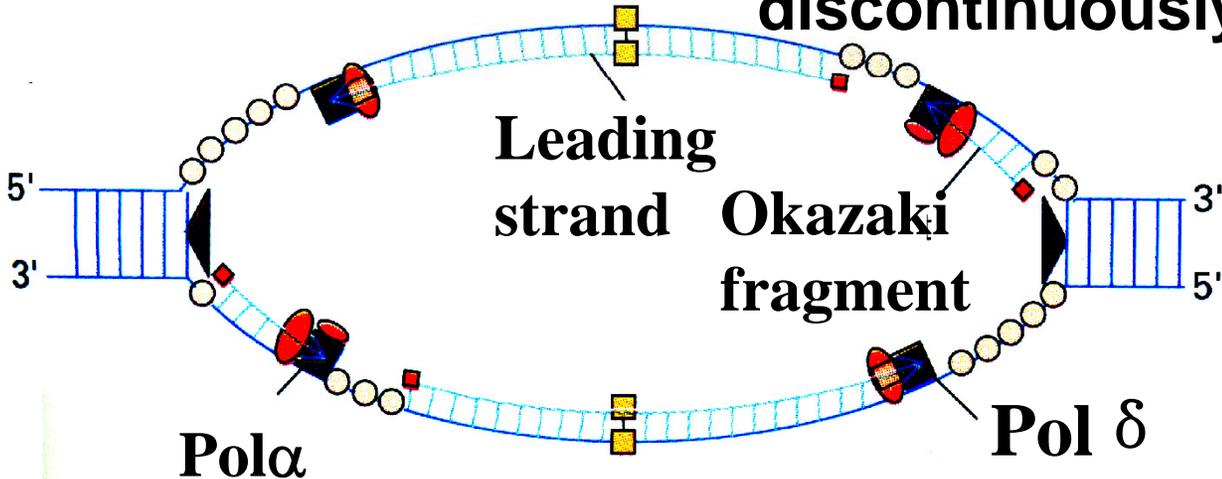
4

Interruption of leading strand synthesis



Binding of Pol δ to PCNA/RFC complex and leading strand synthesis; Binding of primase-Pol α and RFC to lagging strand and synthesis of lagging strand discontinuously

5





2.4.1.3 复制的引发和终止

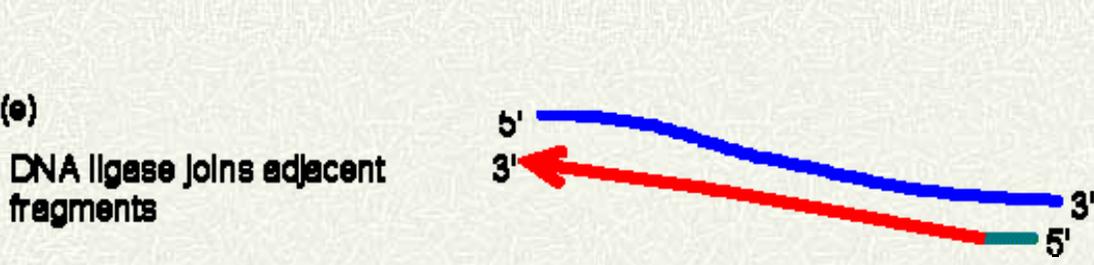
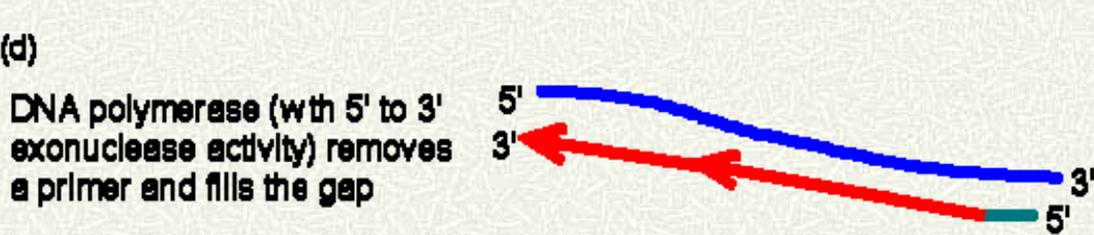
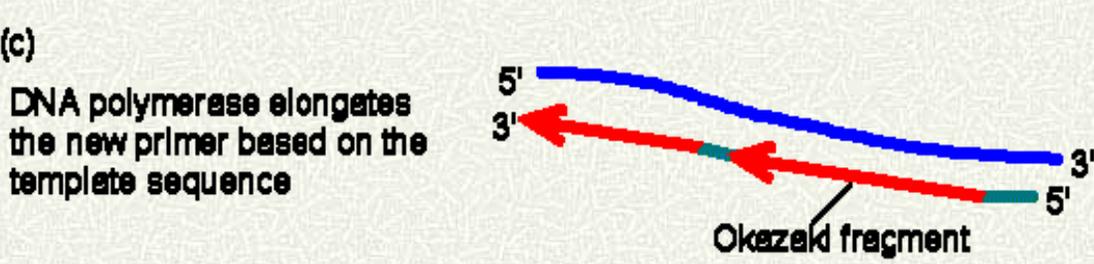
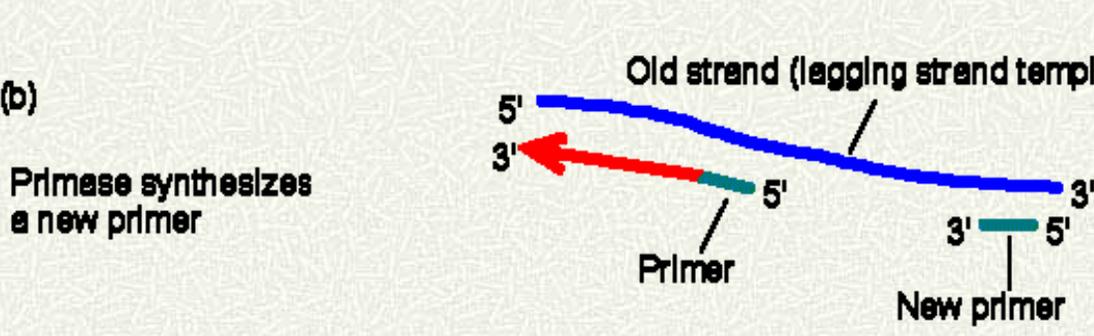
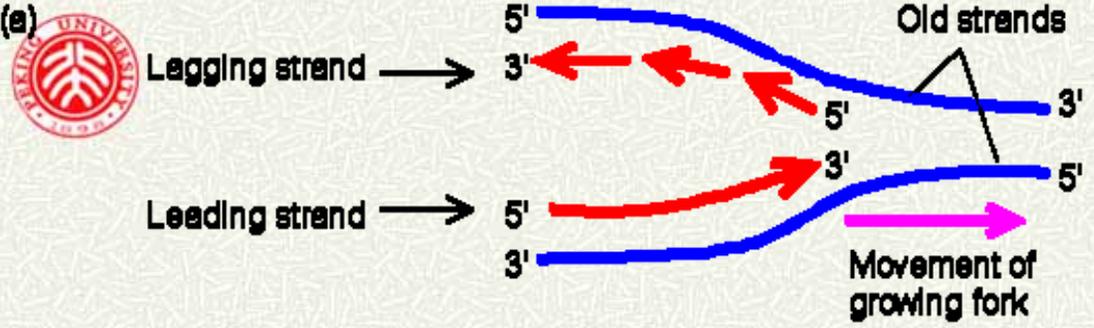
DNA聚合酶只能延长已存在的DNA链，而不能从头合成DNA链，那么，新DNA的复制是怎样开始的呢？

研究发现，DNA复制时，往往先由RNA聚合酶在DNA模板上合成一段RNA引物，再由DNA聚合酶从RNA引物3'端开始合成新的DNA链。

滞后链的引发过程比较复杂，需要多种蛋白质和酶的协同作用，还牵涉到冈崎片段的形成和连接。



Steps in the synthesis of the lagging strand



(b) 首先引物酶合成约10核苷酸大小的新引物。两个引物间的距离在细菌中约为1000-2000个核苷酸，在真核细胞中约为100-200个核苷酸。

(c) DNA polymerase III以5'-3'方向延伸该新的引物，直到遇到邻接引物的5'端。这个新合成的DNA片段就是**Okazaki fragment**。

(d) 在 *E. coli*中，DNA polymerase I 具有 5' 到 3' 外切酶的活性，被用来去除引物。

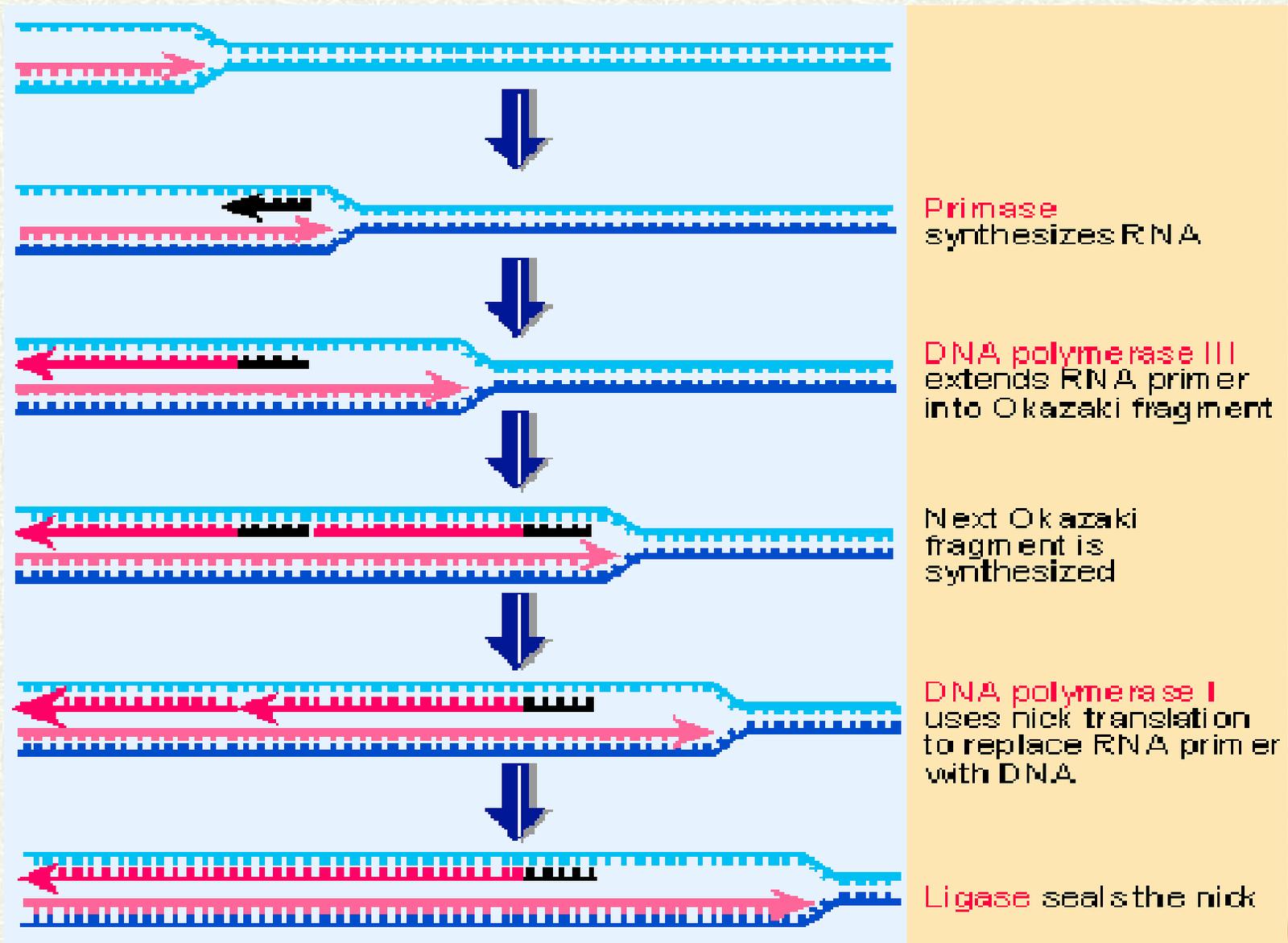
(e) DNA 连接酶连接相邻的冈崎片段。整个滞后链的合成重复步骤 (b) 到(e)。



滞后链的引发过程由**引发体**（**primosome**）来完成。引发体由**6种蛋白质** n 、 n' 、 n'' 、**Dna B**、**C**和**I**共同组成，只有当引发前体（**preprimosome**）把这**6种蛋白质**合在一起并与**引发酶**（**primase**）进一步组装后形成引发体，才能发挥其功能。



Synthesis of the lagging strand





引发体在滞后链分叉的方向上前进，并在模板上间断性引发生成滞后链引物--RNA短链，再由**DNA聚合酶III**作用合成DNA，直至遇到下一个引物或冈崎片段为止。

由**RNase H**降解RNA引物并由**DNA聚合酶I**将缺口补齐，再由**DNA连接酶**将两个冈崎片段连在一起形成大分子DNA。

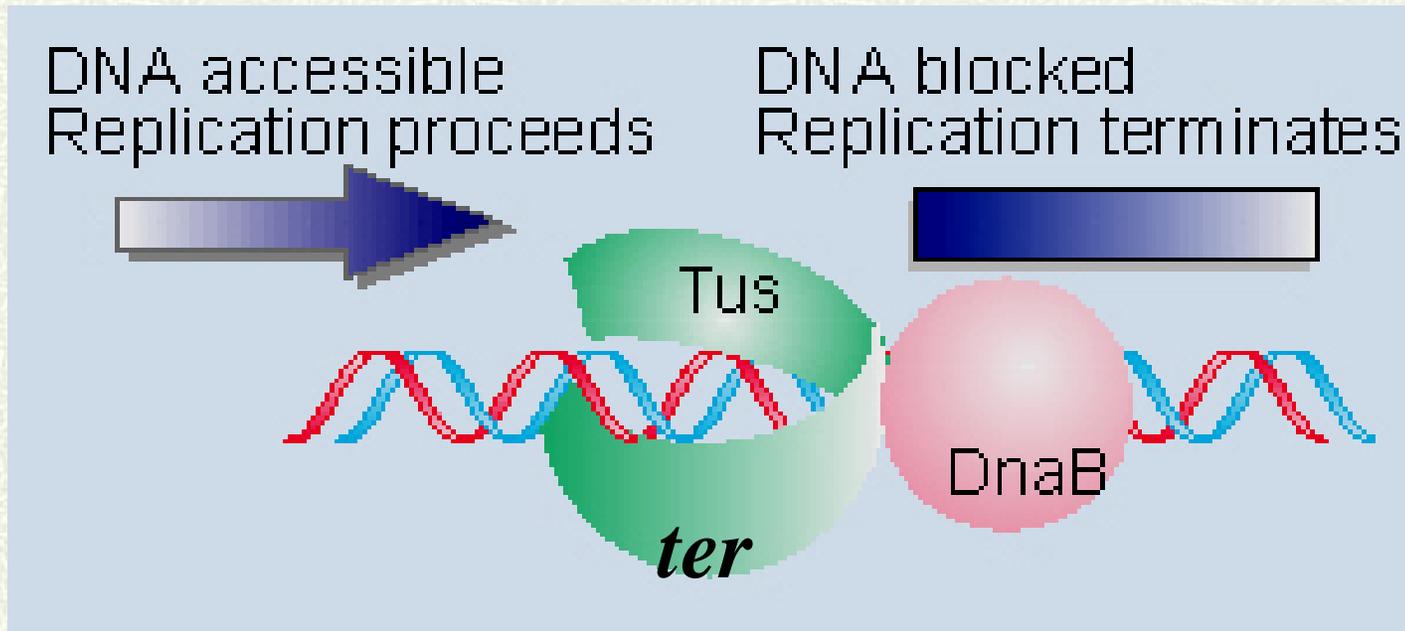


链的终止需要Tus蛋白参与

当复制叉前移，遇到20bp重复性终止子序列（Ter）时，Ter-Tus复合物能阻挡复制叉的继续前移，等到相反方向的复制叉到达后在DNA拓扑异构酶IV的作用下使复制叉解体，释放子链DNA。



链的终止 (**Termination**)



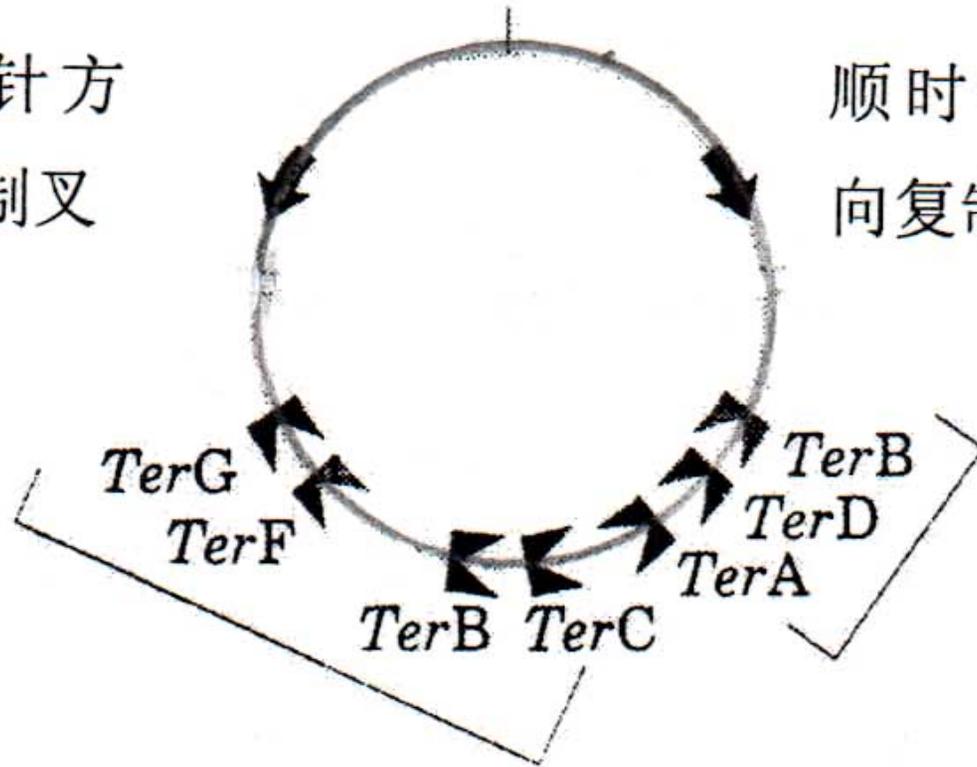
- **Terminus:** containing several terminator sites (*ter*) approximately 180° opposite *oriC*.
- **Tus protein:** *ter* binding protein, an inhibitor of the DnaB helicase



复制起始位点

逆时针方向
复制叉

顺时针方向
复制叉



顺时针方向
复制叉陷阱

逆时针方向
复制叉陷阱

大肠杆菌中DNA复制的终止



2.4.1.4 DNA聚合酶

大肠杆菌中主要有DNA聚合酶**I**、**II**、**III**。



Arthur Kornberg



- Discovered an enzyme in *E. coli* that could catalyse the synthesis of DNA
- He called it DNA Polymerase
 - Now known as **DNA polymerase I**
 - Four other DNA polymerase enzymes in *E. coli*
 - 1959 - Nobel prize in Physiology or Medicine

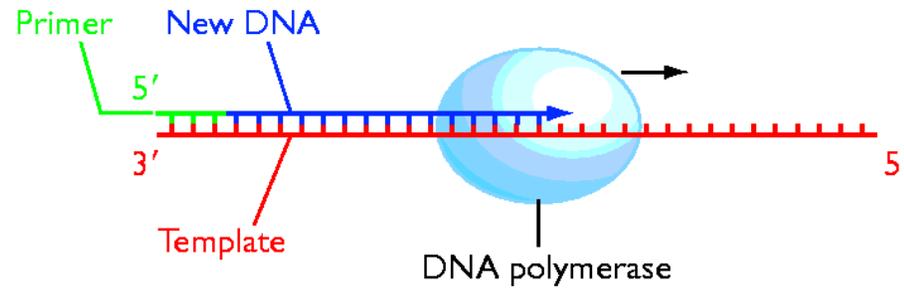


表2-10 大肠杆菌DNA聚合酶I、II和III的性质比较

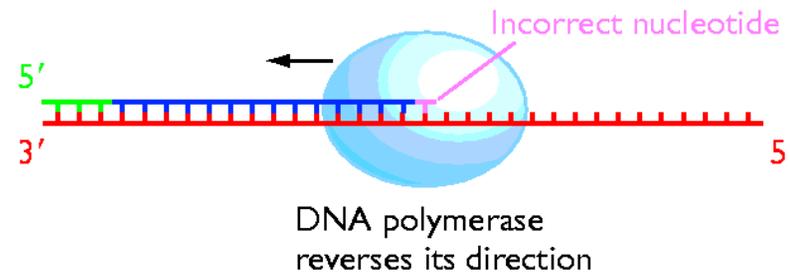
性 质	聚合酶I	聚合酶II	聚合酶III
3' → 5'外切	+	+	+
5' → 3'外切	+	-	-
新生链合成	-	-	+
相对分子质量 (×10 ³)	103	90	900
细胞内分子数	400	?	10~20
生物学活性	1	0.05	15
已知的结构基因	Pol (A)	Pol (B)	Pol (C) (dnaE, N, Z, X , Q等)



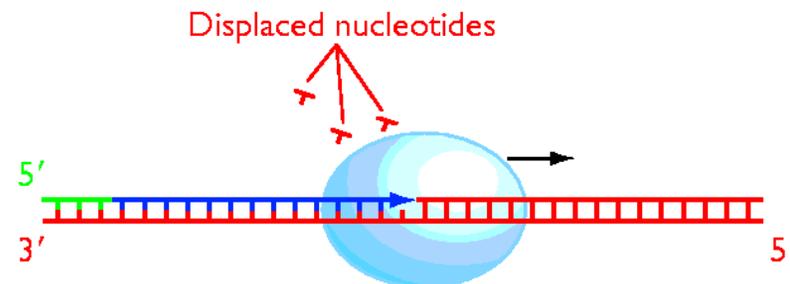
(A) 5'→3' DNA synthesis



(B) 3'→5' exonuclease activity



(C) 5'→3' exonuclease activity





DNA聚合酶的共同点:

- 1、都以**dNTP**为底物。
- 2、都需要**Mg²⁺**激活。
- 3、聚合时必须要有**模板链**和具有**3'—OH**末端的**引物链**。
- 4、链的延伸都方向为**5'→3'**。



DNA聚合酶I不是复制大肠杆菌染色体的主要聚合酶，它有5' → 3'核酸外切酶活性，保证了**DNA复制的准确性**。它也可用来除去冈崎片段5'端**RNA引物**，使冈崎片段间缺口消失，保证**连接酶将片段连接起来**。



DNA聚合酶II的活性很低，若以每分钟酶促核苷酸掺入**DNA**的转化率计算，只有**DNA聚合酶I**的**5%**，所以也不是复制中主要的酶。目前认为**DNA聚合酶II**的生理功能主要是**起修复DNA**的作用。

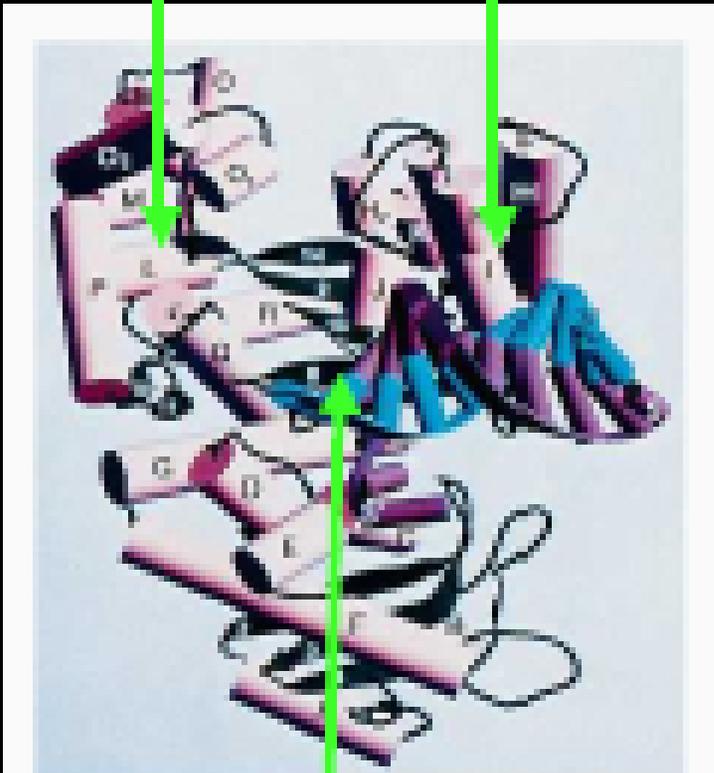


DNA聚合酶III包含有**7种**不同的亚单位和**9个**亚基，其生物活性形式为**二聚体**。它的**聚合活性较强**，为DNA聚合酶I的**15倍**，聚合酶II的**300倍**。它能在引物的**3'—OH**上以每分钟约**5万个**核苷酸的速率延长新生的**DNA链**，是大肠杆菌**DNA复制**中链延长反应的**主导聚合酶**。



X-ray structure of Klenow fragment

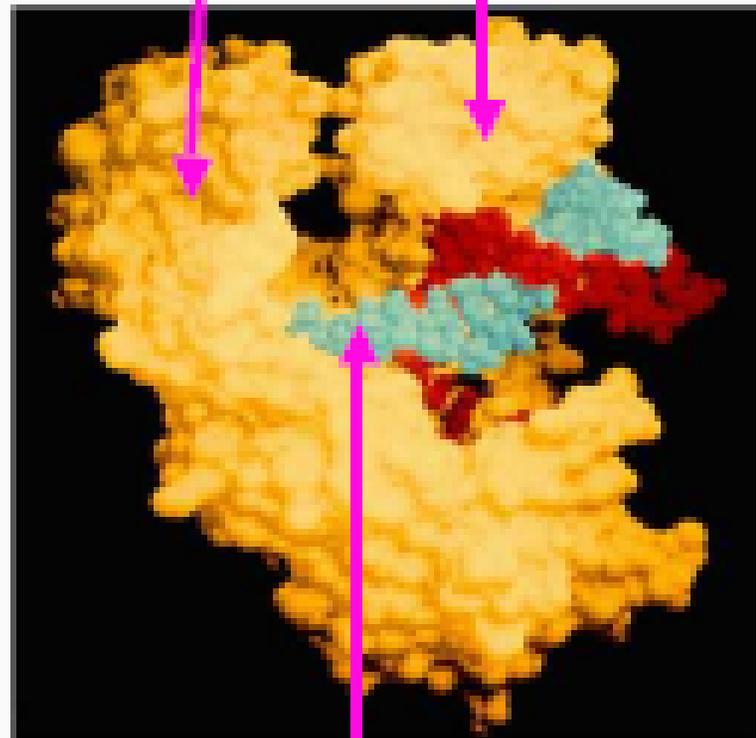
FINGERS THUMB



Courtesy of Thomas Stell, Yale University

PALM

FINGERS THUMB



Courtesy of Thomas Stell, Yale University

PALM



DNA Polymerase III

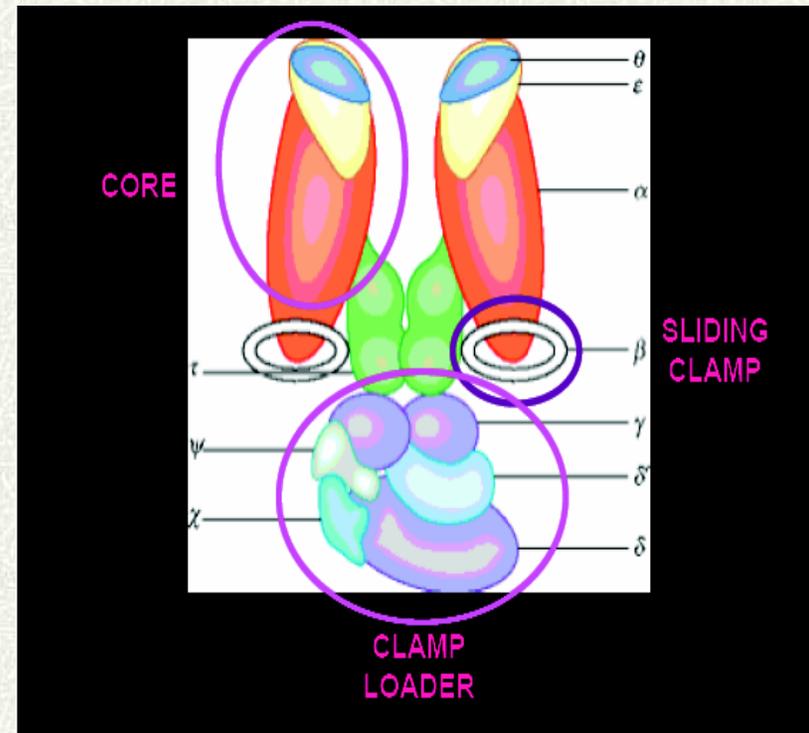
DNA polymerase III is thought to function as an **asymmetric dimer**:

One monomer consisting of the core subunits, two β subunits and two τ subunits

LEADING STRAND SYNTHESIS

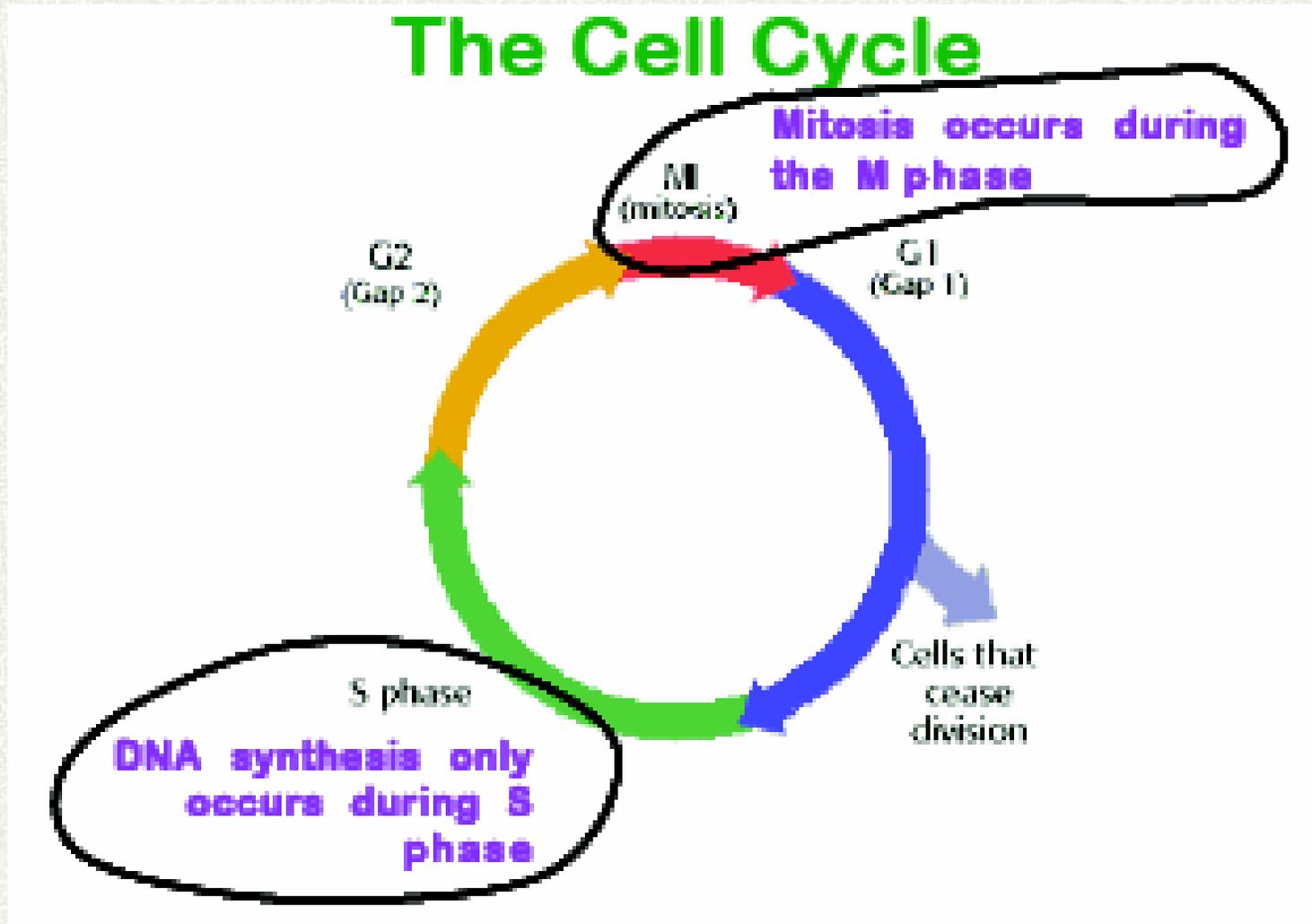
The other monomer consisting of the core subunits, two β subunits and two copies of the γ complex polypeptides

LAGGING STRAND SYNTHESIS





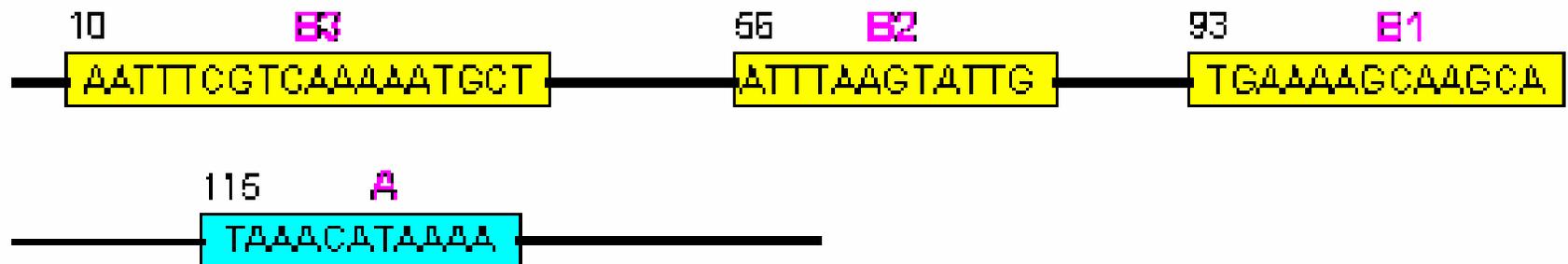
2.4.2 真核生物DNA的复制特点





真核生物DNA的复制子被称为ARS
(autonomously replicating
sequences)，长约150bp左右，包括
数个复制起始必需的保守区。

(b) Yeast replication origin (ARS1)





真核生物**DNA**复制叉的移动速度大约只有**50bp/秒**。因此，人类**DNA**中每隔**30,000-300,000bp**就有一个复制起始位点。



真核细胞中有**5种DNA聚合酶**，分别称为**DNA聚合酶** α 、 β 、 γ 、 δ 和 ϵ 。



表2-11 真核生物DNA聚合酶的特性比较

性质	DNA聚合酶 α	DNA聚合酶 β	DNA聚合酶 γ	DNA聚合酶 δ	DNA聚合酶 ϵ
亚基数	4	1	2	2-3	≥ 1
在细胞内分布	核内	核内	线粒体	核内	核内(?)
功能	DNA引物合成	损伤修复	线粒体DNA复制	主要DNA复制酶	DNA复制(?)
3' → 5' 外切	无	无	有	有	有
5' → 3' 外切	无	无	无	无	无



DNA聚合酶 α 主要参与引物合成。

DNA聚合酶 β 活性水平稳定，主要在**DNA损伤的修复**中起作用。

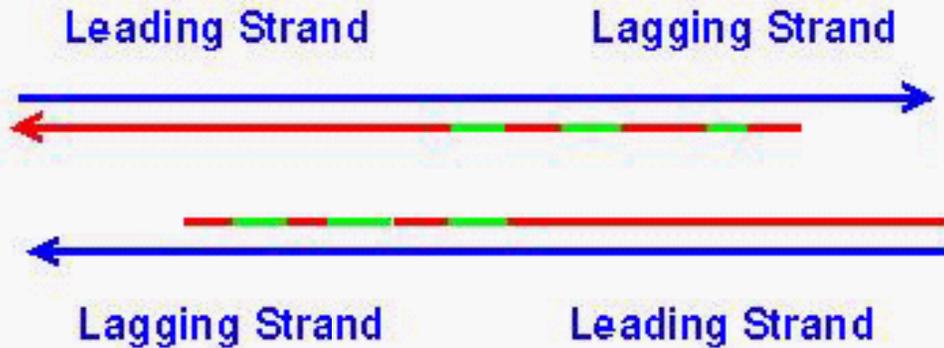
DNA聚合酶 δ 是主要负责**DNA复制的酶**。

DNA聚合酶 ϵ 的主要功能可能是在去掉**RNA引物**后**把缺口补全**。



"end-replication problem"

Synthesis of the lagging strand requires a short primer which will be removed. At the extreme end of a chromosome, there is no way to synthesize this region when the last primer is removed. Therefore, the lagging strand is always shorter than its template by at least the length of the primer. This is the so-called "**end-replication problem**".





"end-replication problem"

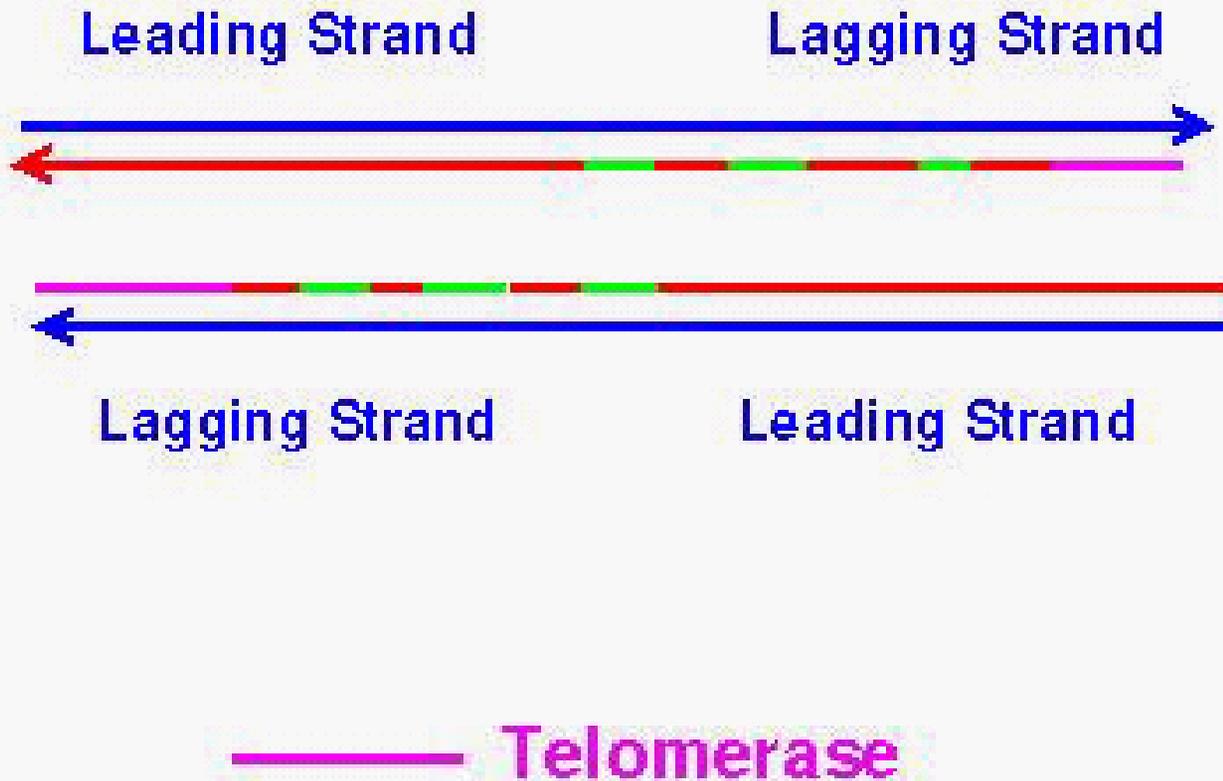
Bacteria do not have the end-replication problem, because its DNA is circular.

In eukaryotes, the chromosome ends are called telomeres which have at least two functions:

- To protect chromosomes from fusing with each other.
- To solve the end-replication problem.



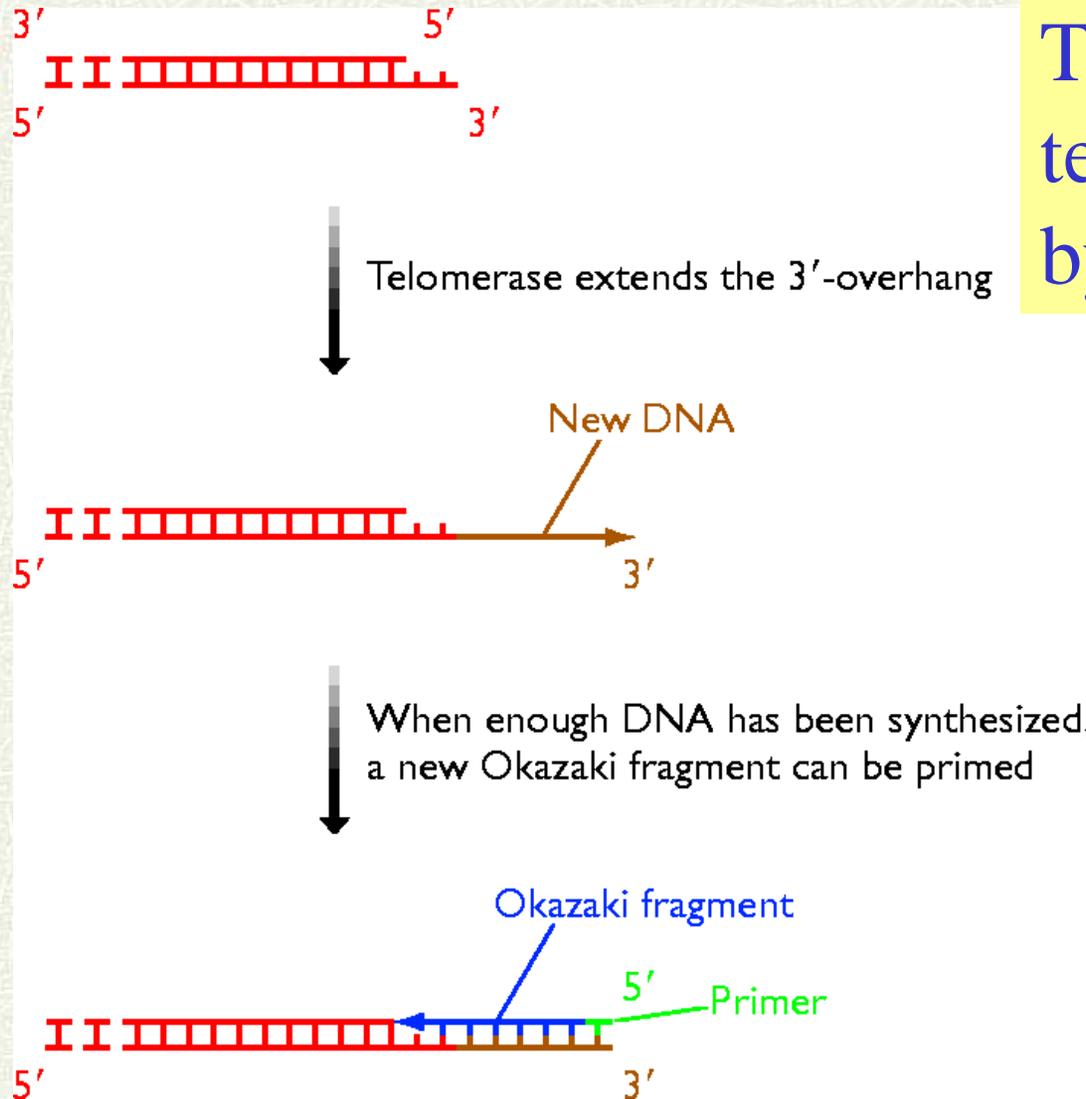
What happens in the "End"?



Telomerase is a reverse transcriptase together with a template RNA



The mechanism of telomere extension by telomerase



To extend the length of a telomere, the telomerase first extends its longer strand. Then, using the same mechanism as synthesizing the lagging strand, the shorter strand is extended.



2.4.3 DNA复制的调控

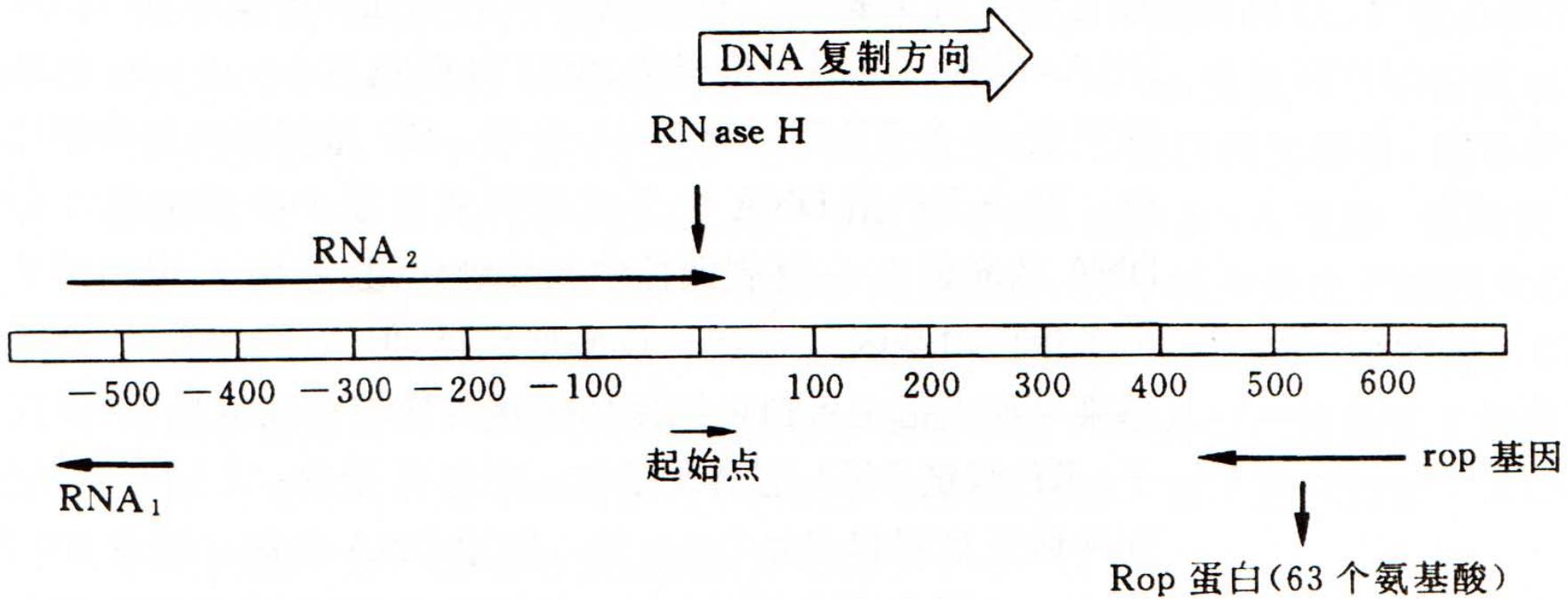
DNA链延伸的速度在同种生物的不同细胞中几乎是恒定的，只是复制叉的数量不同。

迅速分裂的细胞具较多的复制叉，而分裂缓慢的细胞复制叉较少并出现复制的间隙。复制起始频率的直接调控因子是蛋白质和**RNA**。



ColEI质粒DNA的复制调控

ColEI是一个**6646**碱基对的小质粒，在宿主细胞内有**20~30**个拷贝，其**DNA**的复制完全依靠宿主**DNA**聚合酶。质粒**DNA**编码两个负调控因子**Rop**蛋白和反义**RNA**（**RNA₁**），它们控制了起始**DNA**复制所必需的引物合成。



ColE1质粒DNA复制起始部位调控因子之间关系



引物**RNA**前体的转录起始于复制起点上游，需经**RNaseH**加工后产生有**555**个核苷酸的引物，然后由**DNA**聚合酶**I**在引物的**3'**末端起始**DNA**合成。



RNA₁的编码区在引物**RNA**编码区的5'末端，转录方向与引物**RNA**相反，因此与引物**RNA**的5'末端互补，并通过氢键配对与后者相互作用，阻止了**RNaseH**加工引物前体，使其不能转化为有活性的引物而对复制起负调控作用。**Rop**蛋白是**RNA₁**基因转录的激活因子。



大肠杆菌染色体DNA的复制调控

- 染色体的复制与分裂一般是同步的，但二者并不直接偶联。
- 在一定生长速度范围内，细胞与染色体的质量之比相对恒定。
- 复制子由起始物位点和复制起点两部分组成。起始物位点编码复制体调节蛋白，复制起点与调节蛋白相互作用并启动复制。调节蛋白通过与复制复合物的相互作用确定复制起始频率和复制方式。
- 复制起点有OriC和OriH两种。



真核细胞DNA的复制调控

- **细胞生活周期水平调控**（限制点调控）：
决定细胞停留在G1期还是进入S期(by Cyclins, Cdc (Cell Division Cycle) gene products)。
- **染色体水平调控**：决定不同染色体或同一染色体不同部位的复制子按一定顺序在S期起始复制。
- **复制子水平调控**：决定复制的起始与否，并且是高度保守的。



DNA Polymerase I

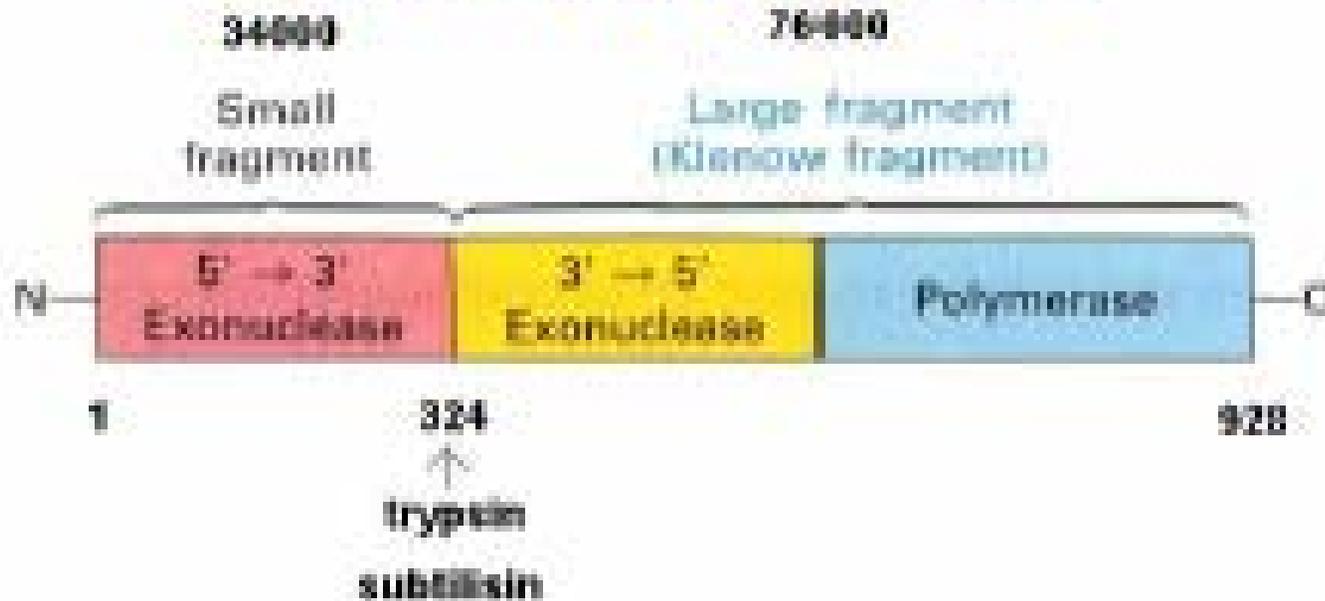
DNA polymerase I is a 928 amino-acid polypeptide (MW=103,118) encoded by the *polA* gene.

DNA polymerase I has three distinct enzymatic activities:

- a 5' → 3' polymerase activity
- a 3' → 5' exonuclease activity
- a 5' → 3' exonuclease activity



DNA Polymerase I



- Structurally, the enzyme has two domains which can be separated by proteolysis with **trypsin** or **subtilisin**
- Both enzymes cleave at the amino acid 324



supplementary

DNA Polymerase II

The polymerase subunit is 89,921 kD in size and is coded by the **polB gene (aka **dinB**)**

Synthesis of DNA polymerase II is induced during the **stationary phase of cell growth**

It is likely that this enzyme is involved in **DNA repair systems**



supplementary

DNA Polymerase III

DNA polymerase III holoenzyme is the **principal replicative enzyme** in **E. coli**

It catalyses polymerization at a high rate

The enzyme is highly processive

The enzyme is a complex of 10 polypeptides

There are two forms of the enzyme:

- **Core Enzyme**
- **Holoenzyme**



DNA Polymerase IV

Synthesis of DNA polymerase IV is also induced during the **stationary phase** of cell growth

It “competes” with DNA polymerase II

DNA polymerase IV is **an error prone DNA polymerase**

DNA polymerase IV is thought to be responsible for **50%** of the adaptive mutations that arise during stationary phase

- When **E. coli** is exposed to high levels of radiation or high levels of a mutagen, major damage to the bacterial DNA can occur
- The cell responds by inducing a special “last-resort” repair pathway called the **SOS repair pathway**
- Two **SOS genes**, **umuC** and **umuD**, code for DNA Polymerase IV (**UmuD'2C**)



DNA Polymerase V

DNA Polymerase V replicates past gaps in the DNA

DNA Polymerase V is an error prone DNA polymerase; there are no proofreading functions associated with it.

It is a relatively poor polymerase that synthesises DNA distributively

DNA Polymerase V also requires both the β subunit (sliding clamp) and the γ complex (clamp loader) of DNA polymerase III for optimal function



DNA的损伤与修复

(DNA Mutation & Repair)



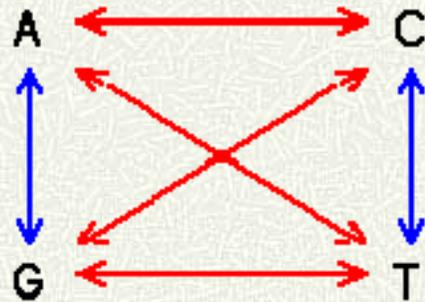
DNA的损伤 (Mutation)

- Small-scale mutations
 - a. substitution (替换)
 - b. deletion (缺失)
 - c. insertion (插入)
 - d. exon skipping
- The chromosome abnormality
 - a. Numerical abnormality: Triploidy, Monosomy, Trisomy
 - b. Structural abnormality: Two breaks in a single chromosome can cause inversion, deletion or ring structure.



The substitution mutation

(a)



(b)

Silent mutation

TGT → TGC

Cys → Cys

Missense mutation

TGT → TGG

Cys → Trp

Nonsense mutation

TGT → TGA

Cys → Stop

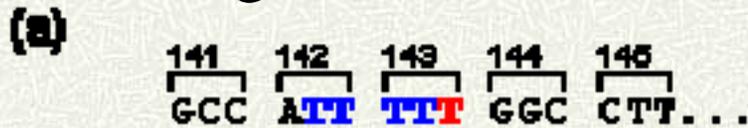
(a) Illustration of **transition (blue)** and **transversion (red)** mutations. In the transition mutation, a pyrimidine (C or T) is substituted by another pyrimidine, or a purine (A or G) is substituted by another purine. The transversion mutation involves the change from a pyrimidine to a purine, or vice versa.

(b) Examples of **silent, missense and nonsense** mutations. The silent mutation does not produce any change in the amino acid sequence, the missense mutation results in a different amino acid, and the nonsense mutation generates a stop signal.

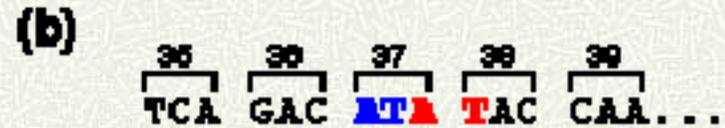


Deletion mutation

The deletion mutation involves elimination of one or more nucleotides from a DNA sequence. It may cause **frameshift**, producing a **non-functional protein**.



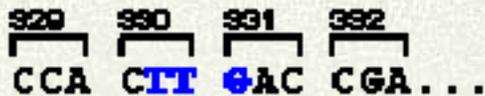
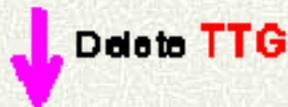
In CFTR gene



In CFTR gene



In FIX gene



In APC gene



Real examples of deletion mutations which cause diseases



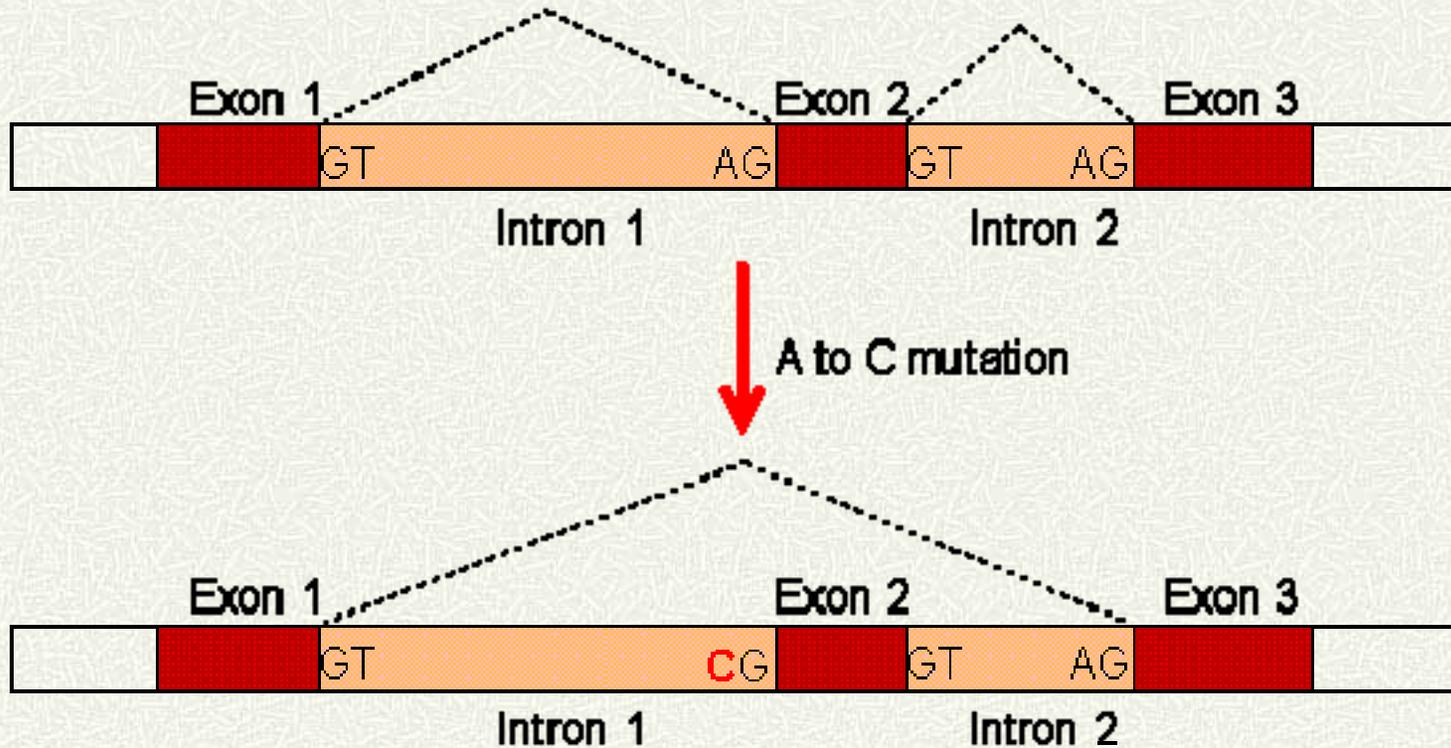
Insertion

One or more nucleotides are inserted into a sequence. If the number of inserted bases is not a multiple of 3, it will cause frameshift, resulting in serious consequences.

Disease	Gene Location	Repeat Sequence	Normal Repeat Number	Mutated Repeat Number
Huntington disease	4p16.3	CAG	9 - 35	37 - 100
Kennedy disease	Xq21	CAG	17 - 24	40 - 55
SCA1	8p23	CAG	19 - 36	43 - 81
DRPLA	12p	CAG	7 - 23	> 48
Fragile X site A	Xq27.3	CGG	6 - 54	> 200
Fragile X site E	Xq28	CCG	6 - 25	> 200
Fragile X site F	Xq28	GCC	6 - 28	> 500
Myotonic dystrophy	19q13	CTG	5 - 35	50 - 4000



Exon skipping



Splicing of an intron requires an essential signal: "GT.....AG". If the splice acceptor site AG is mutated (e.g., A to C in this figure), the splicing machinery will look for the next acceptor site. As a result, the exon between two introns is also removed.



2.5 DNA的修复

由于染色体**DNA**在生命过程中占有至高无上的地位，**DNA**复制的准确性以及**DNA**日常保养中的损伤修复就有着特别重要的意义。



表2-12 大肠杆菌中DNA的修复系统

DNA修复系统	功能
错配修复	恢复错配
碱基切除修复	切除突变的碱基
核苷酸切除修复	修复被破坏的DNA
DNA直接修复	修复嘧啶二体或甲基化DNA



2.5.1 错配修复 (mismatch repair)

错配修复对DNA复制忠实性的贡献力达 10^2 - 10^3 ，DNA子链中的错配几乎完全能被修正。

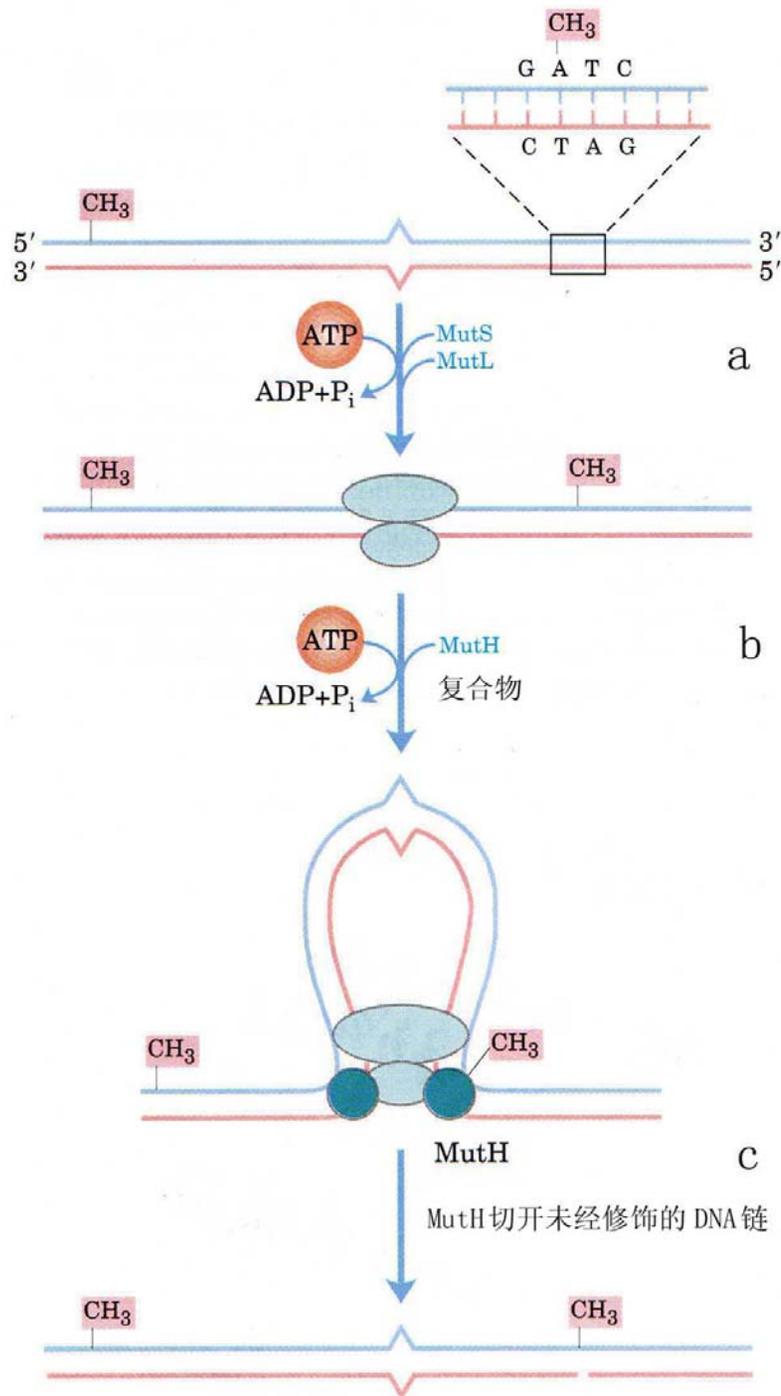


图2-25 根据母链甲基化原则找出错配碱基的示意图。

a. 发现碱基错配。
b. 在水解ATP的作用下，MutS, MutL与碱基错配位点的DNA双链相结合。

c. MutS-MutL在DNA双链上移动，发现甲基化DNA后由MutH切开非甲基化的子链。



该系统识别母链的依据来自 **Dam 甲基化酶**，它能使位于5' **GATC**序列中腺苷酸的N⁶位甲基化。一旦复制叉通过复制起始位点，母链就会在开始**DNA**合成前的几秒钟至几分钟内被甲基化。



只要两条**DNA**链上碱基配对出问题，错配修复系统就会根据“**保存母链，修正子链**”的原则，找出错误碱基所在的**DNA**链，并在对应于母链甲基化腺苷酸上游鸟苷酸的**5'**位置切开子链，再启动特定的修复途径，合成新的子链片段。

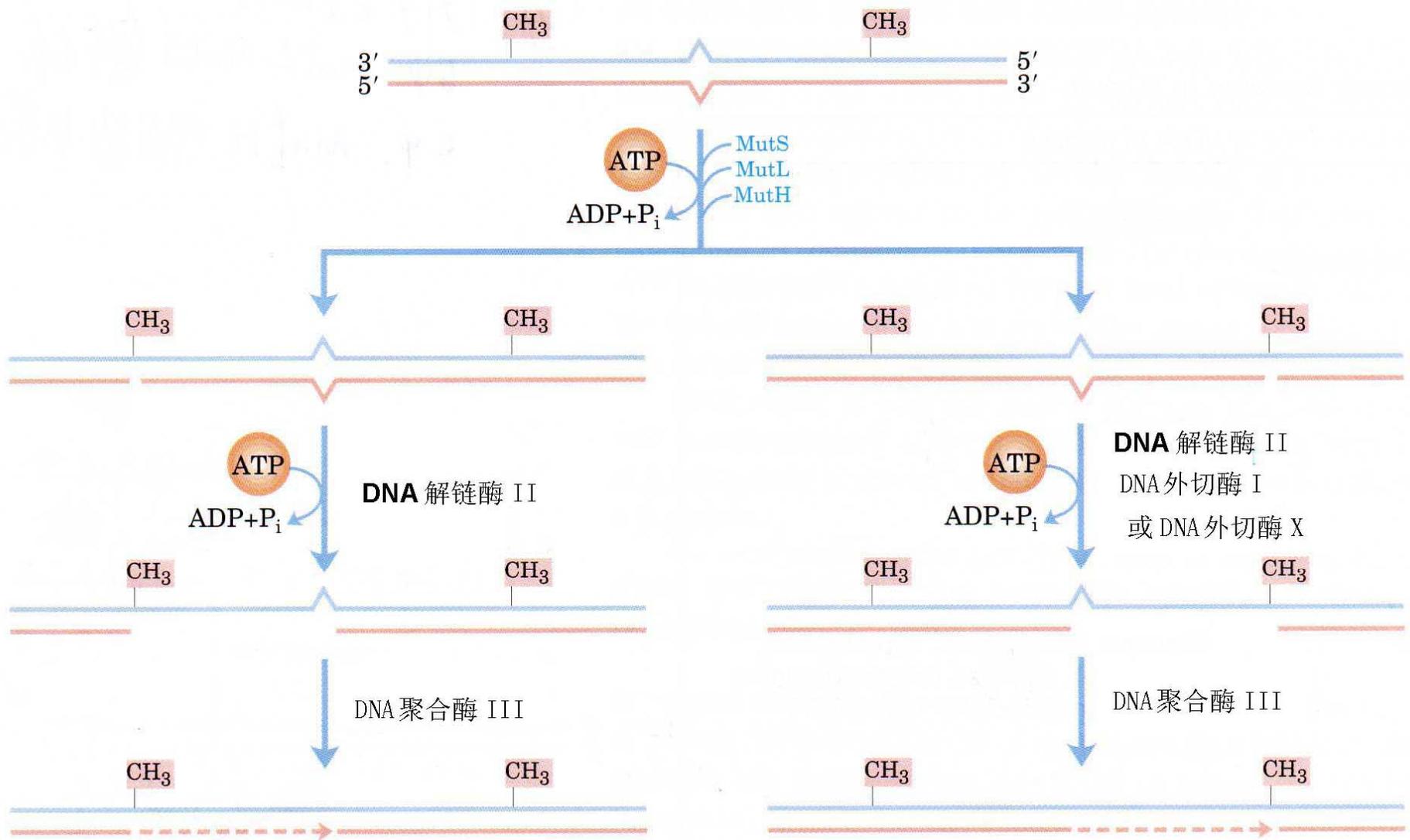
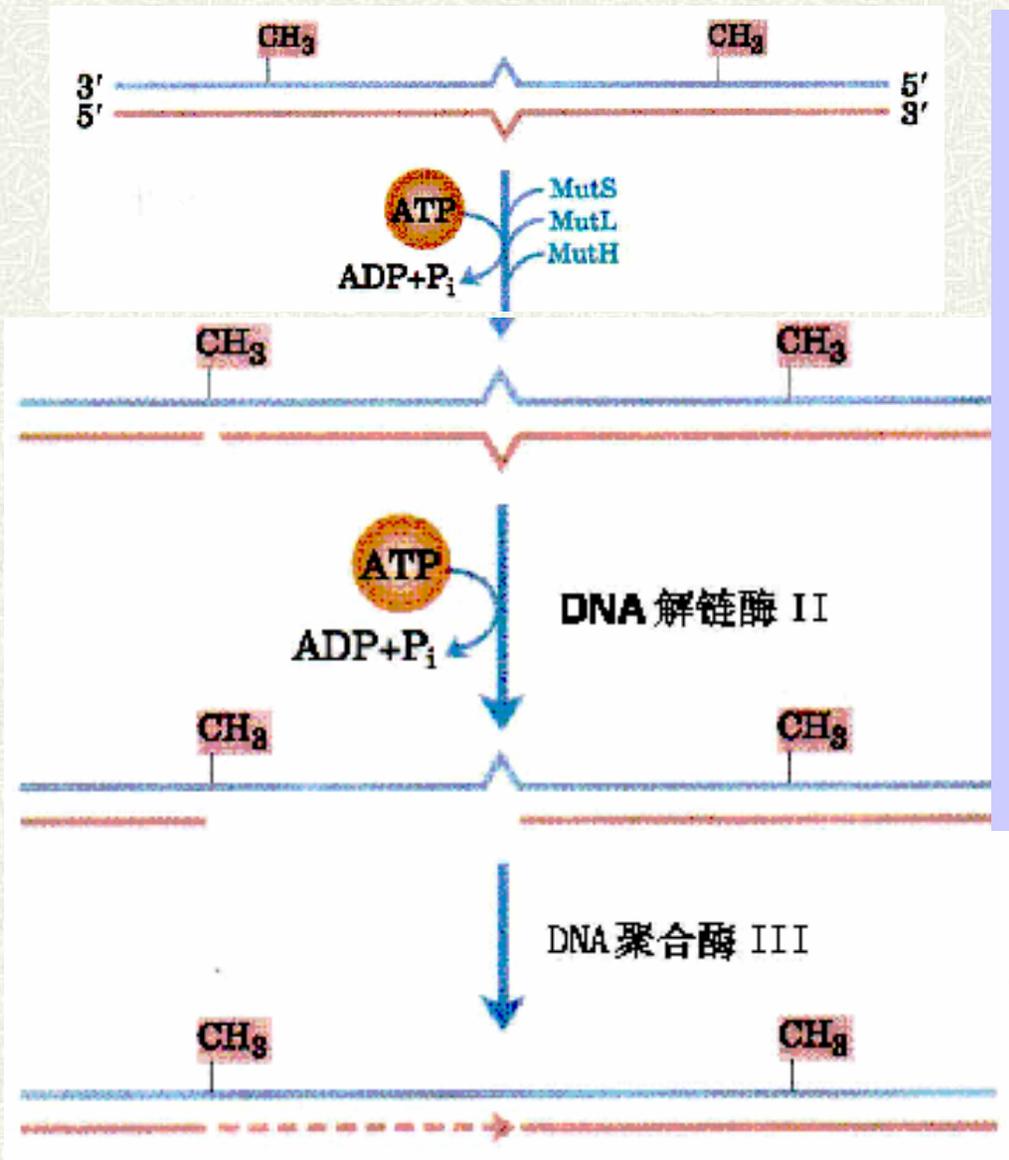
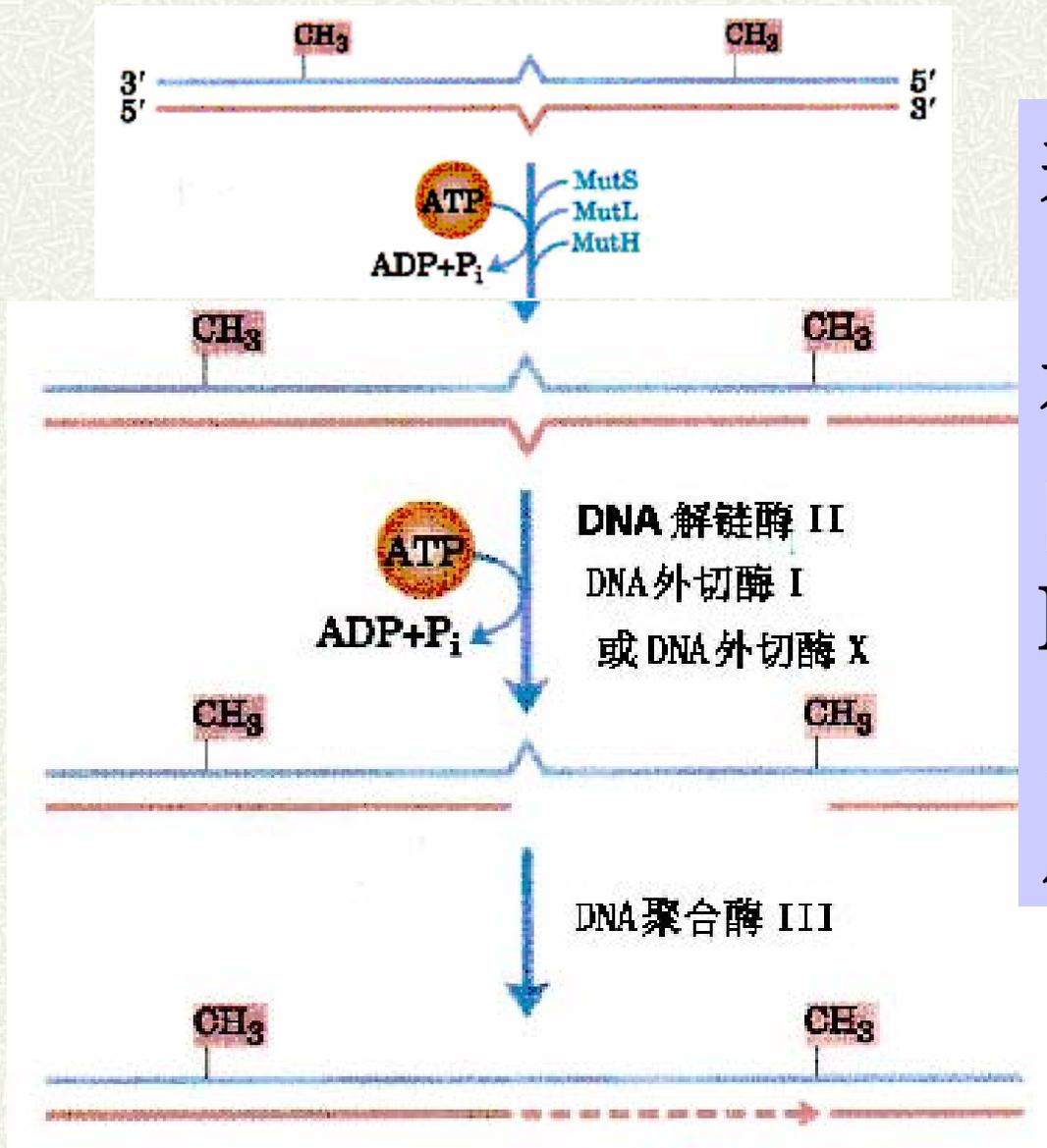


图2-26 碱基错配修复过程示意图



当错配碱基位于切口3'下游端时，在MutL-MutS、解链酶II、DNA外切酶VII或RecJ核酸酶的作用下从错配碱基3'下游端开始切除单链DNA直到错配位点，并在Pol III和SSB的作用下合成新的子链片段。



若错配碱基位于切口的5'上游端，则在DNA外切酶I或X的作用下切除单链DNA直到错配位点，再合成新的子链片段。



2.5.2 碱基切除修复 (Base-excision repair)

所有细胞中都带有能识别受损核酸位点的不同的糖苷水解酶，能特异性切除受损核苷酸上的N- β 糖苷键，形成去嘌呤或去嘧啶位点，统称为AP位点。

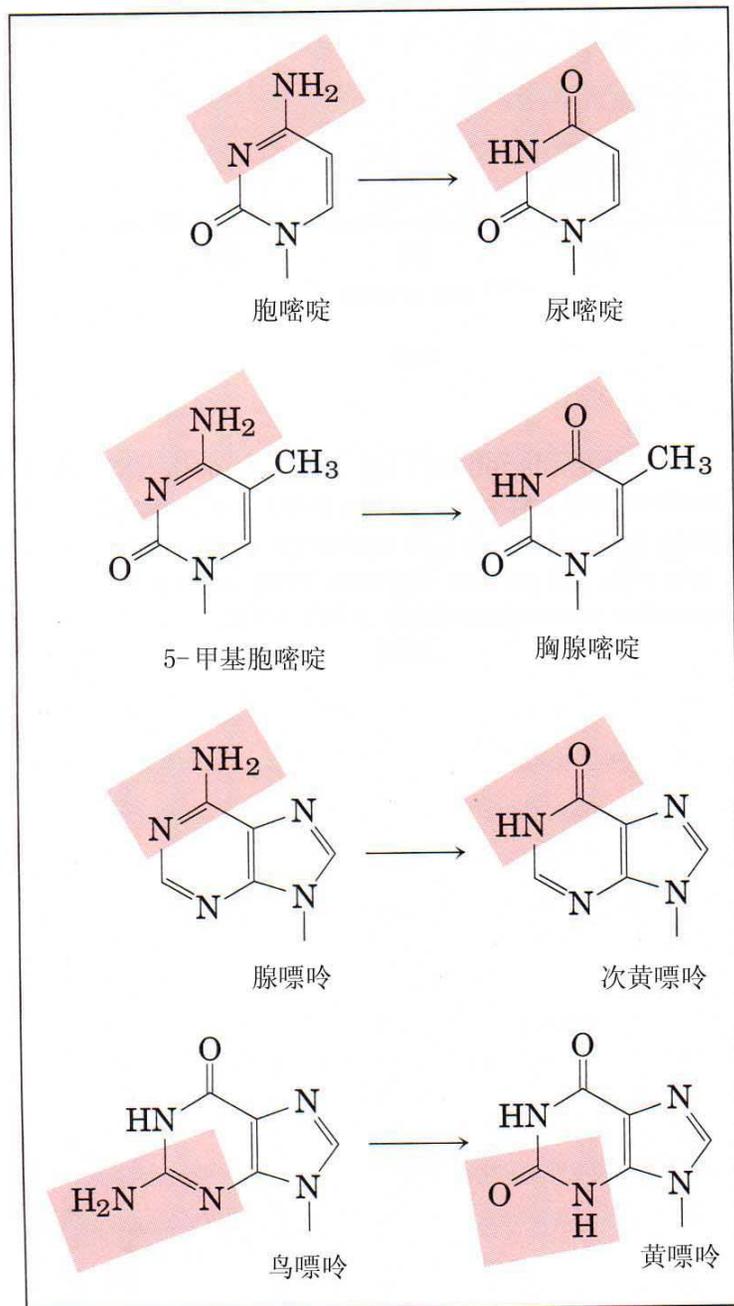


图2-27 DNA分子中常见的几种核苷酸非酶促转变反应

a. 脱氨基反应

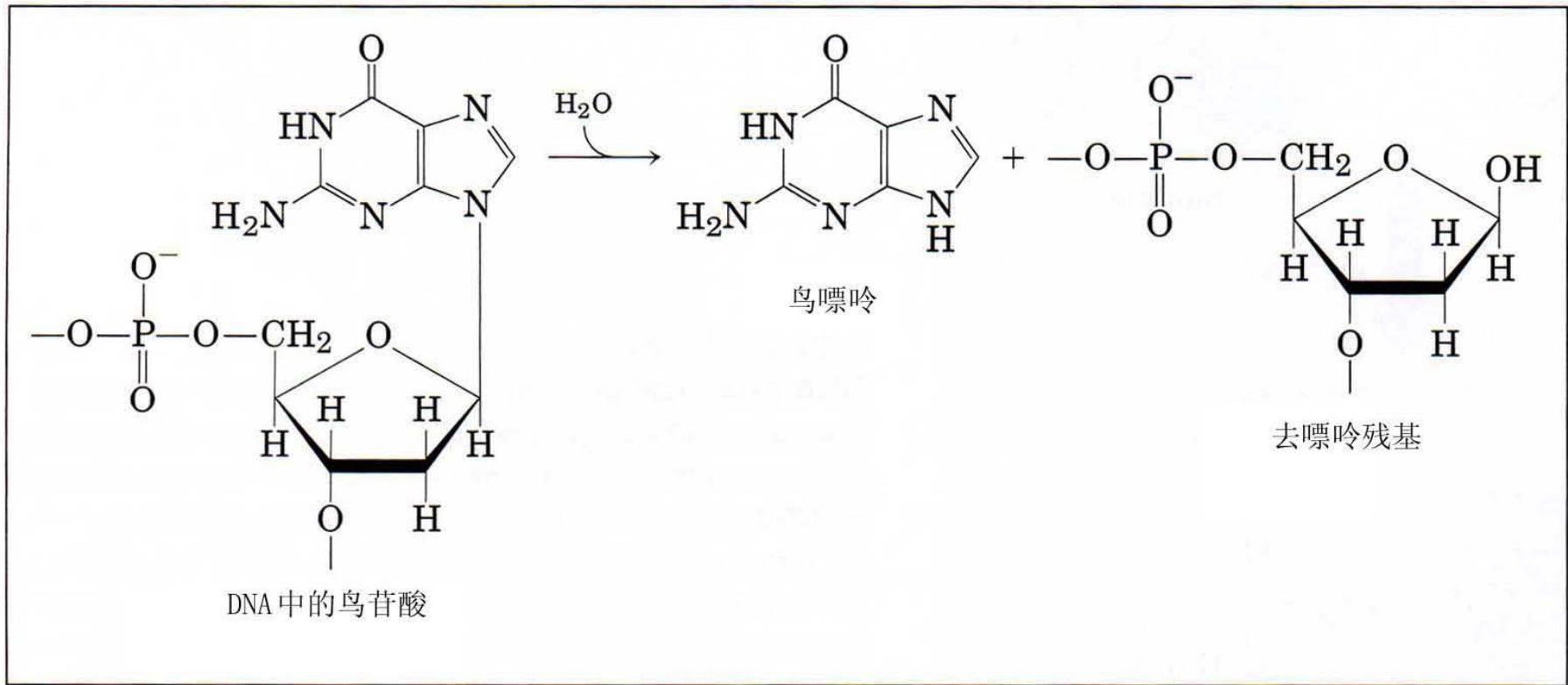
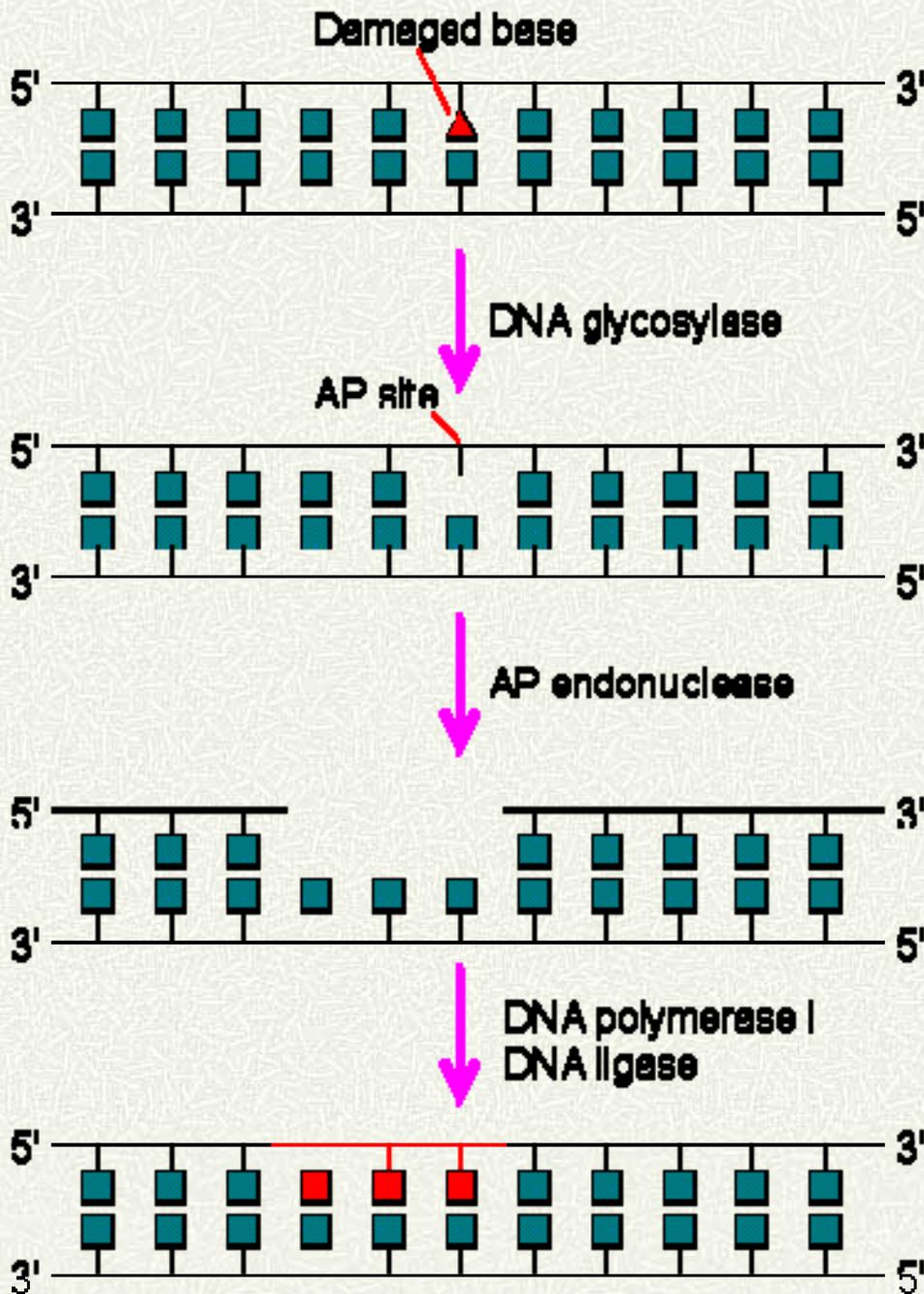


图2-27 DNA分子中常见的几种核苷酸非酶促转变反应

b. 脱嘌呤反应 (N-β糖苷键被水解)



DNA repair by base excision

DNA分子中一旦产生了AP位点，内切酶就会把受损核苷酸的糖苷-磷酸键切开，移去AP位点附近小片段DNA，并由DNA聚合酶I和DNA连接酶共同完成修复。

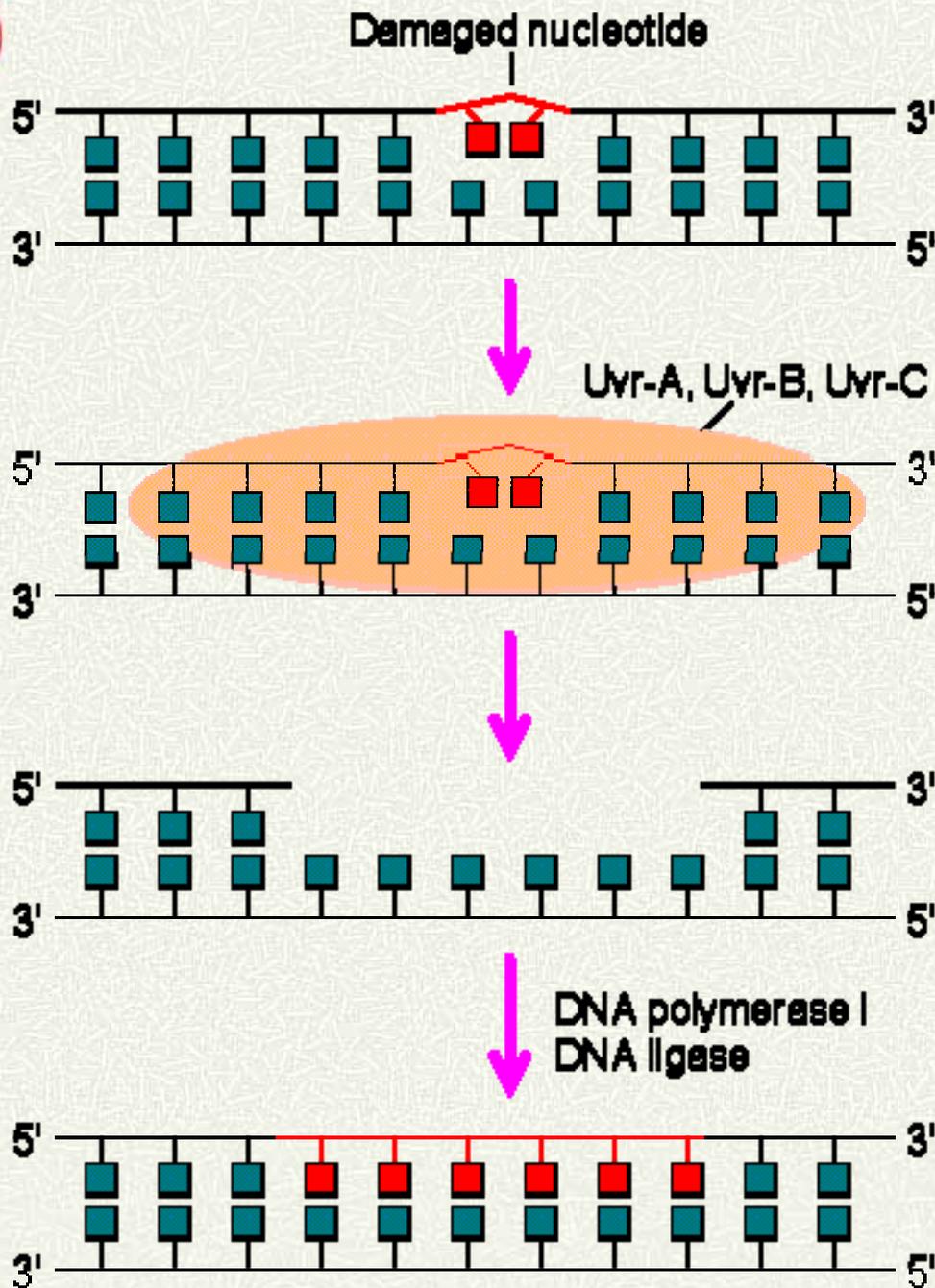


2.5.3 核苷酸切除修复 (nucleotide-excision repair)

当DNA链上相应位置的核苷酸发生损伤，导致双链之间无法形成氢键，则由核苷酸切除修复系统负责修复。



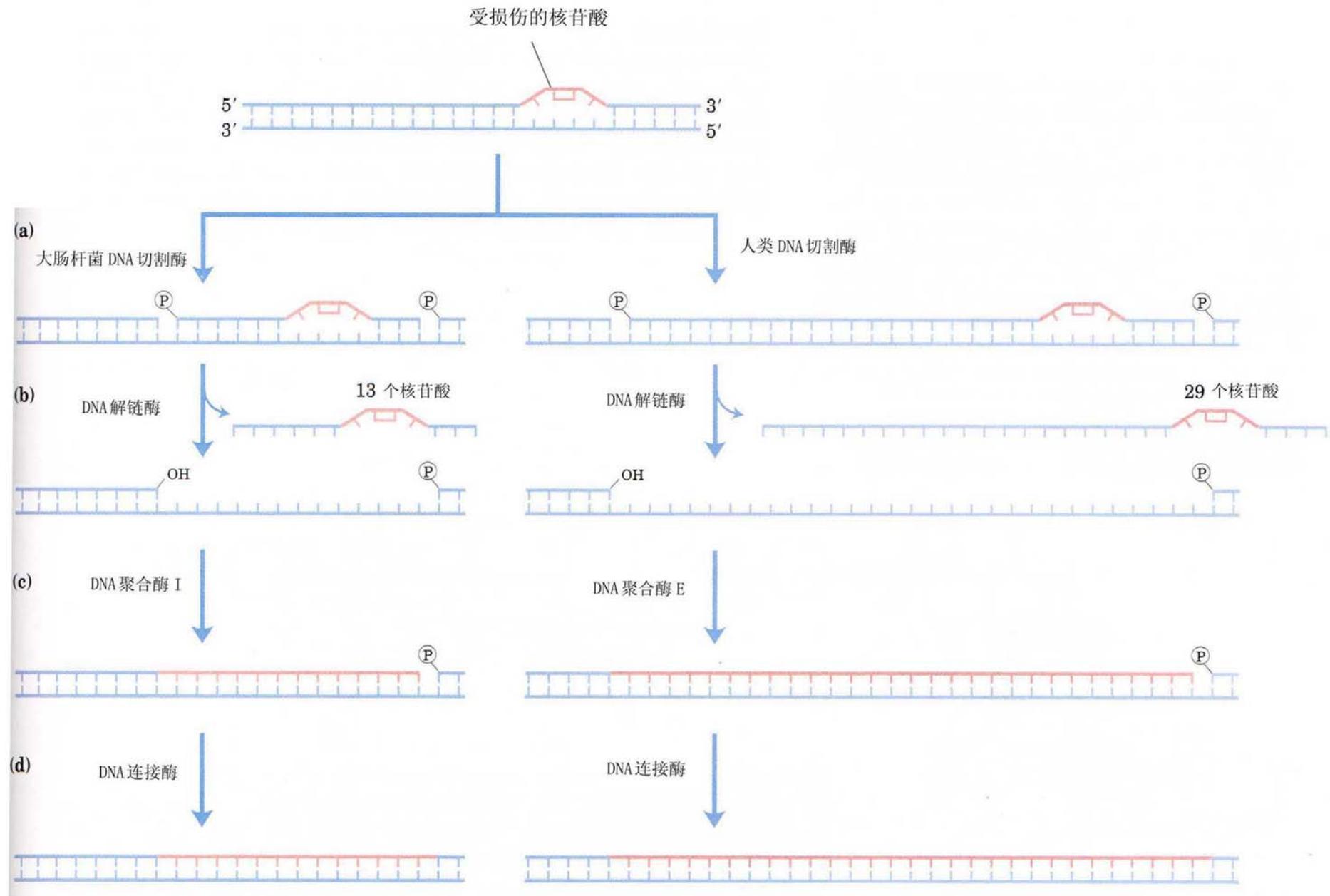
DNA repair by nucleotide excision



损伤发生后，首先由DNA切割酶 (excinuclease) 在已损伤的核苷酸5'和3'位分别切开磷酸糖苷键，产生并移去DNA小片段，然后由DNA聚合酶合成新片段，并由DNA连接酶完成修复中的最后步骤。



大肠杆菌和人类细胞中的核苷酸切除修复过程示意图





在原核生物中受损核苷酸3'端的第5位，5'端的第8位磷酸糖苷分别被DNA切割酶切开。

在人类细胞中，受损伤核苷酸3'端第6位，5'端的第23位磷酸糖苷键分别被DNA切割酶切开。



2.5.4 DNA的直接修复 (direct repair)

生物体内还存在DNA损伤直接修复而并不需要切除碱基或核苷酸的机制。

DNA光解酶 (photolyase) 能把在光下或经紫外光照射形成的环丁烷胸腺嘧啶二体及**6-4光化物 (6-4-photoproduct)** 还原成为单体。

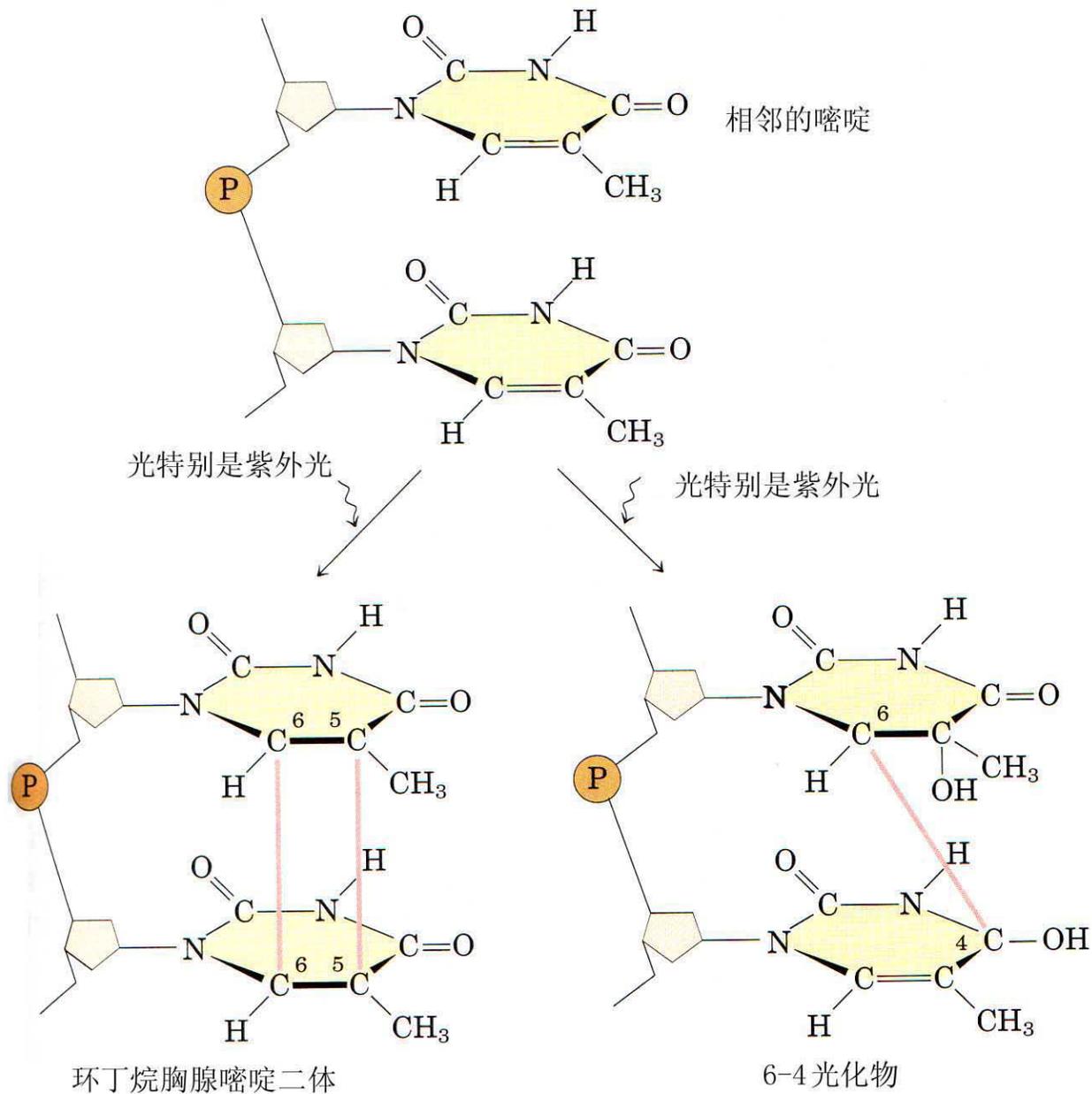
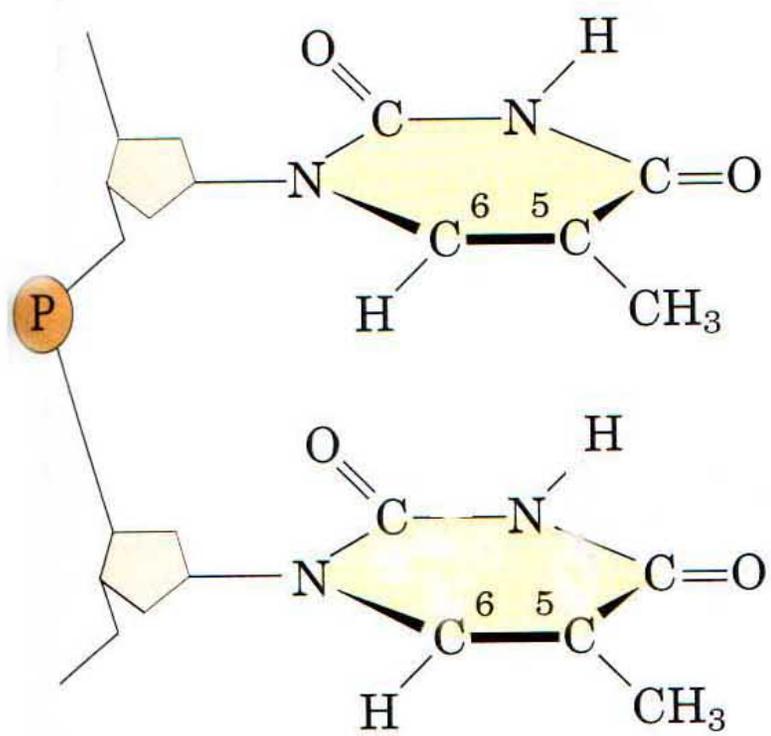
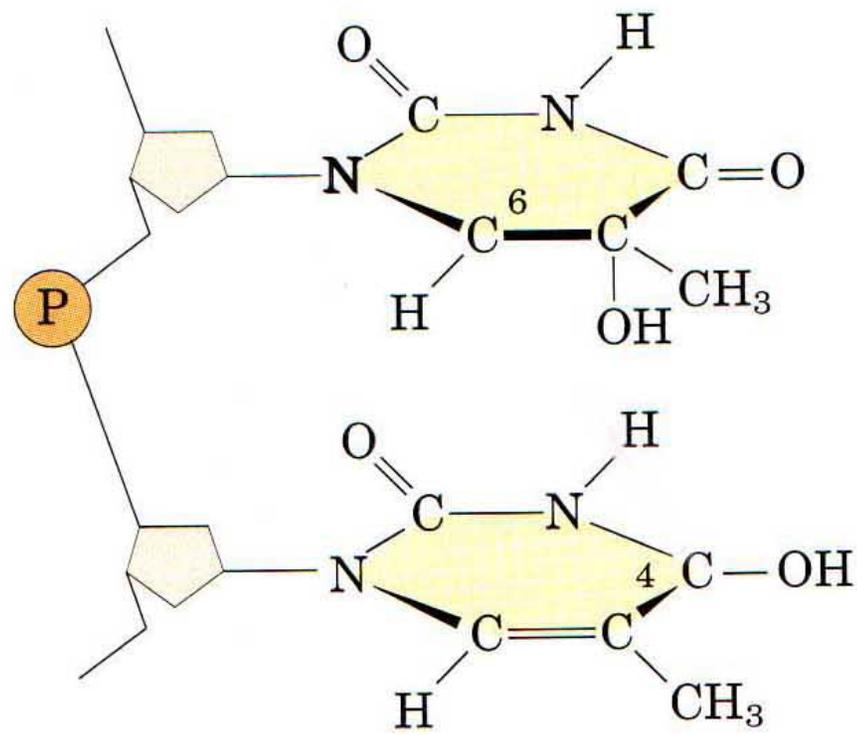


图2-29 紫外光诱发形成嘧啶二体



环丁烷胸腺嘧啶二体



6-4 光化物



此外，生物体内还广泛存在着使O⁶-甲基鸟嘌呤脱甲基化的**甲基转移酶**，以防止形成G-T配对。

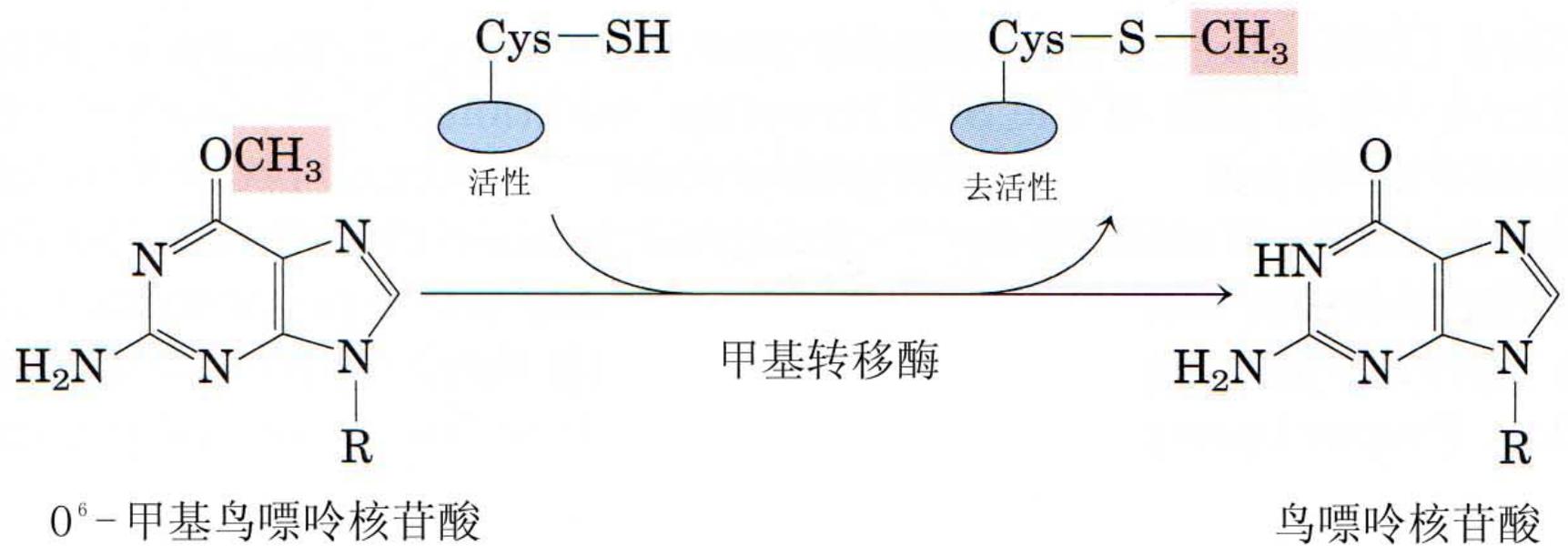


图2-30 O⁶-甲基转移酶使O⁶甲基化鸟嘌呤恢复成鸟嘌呤

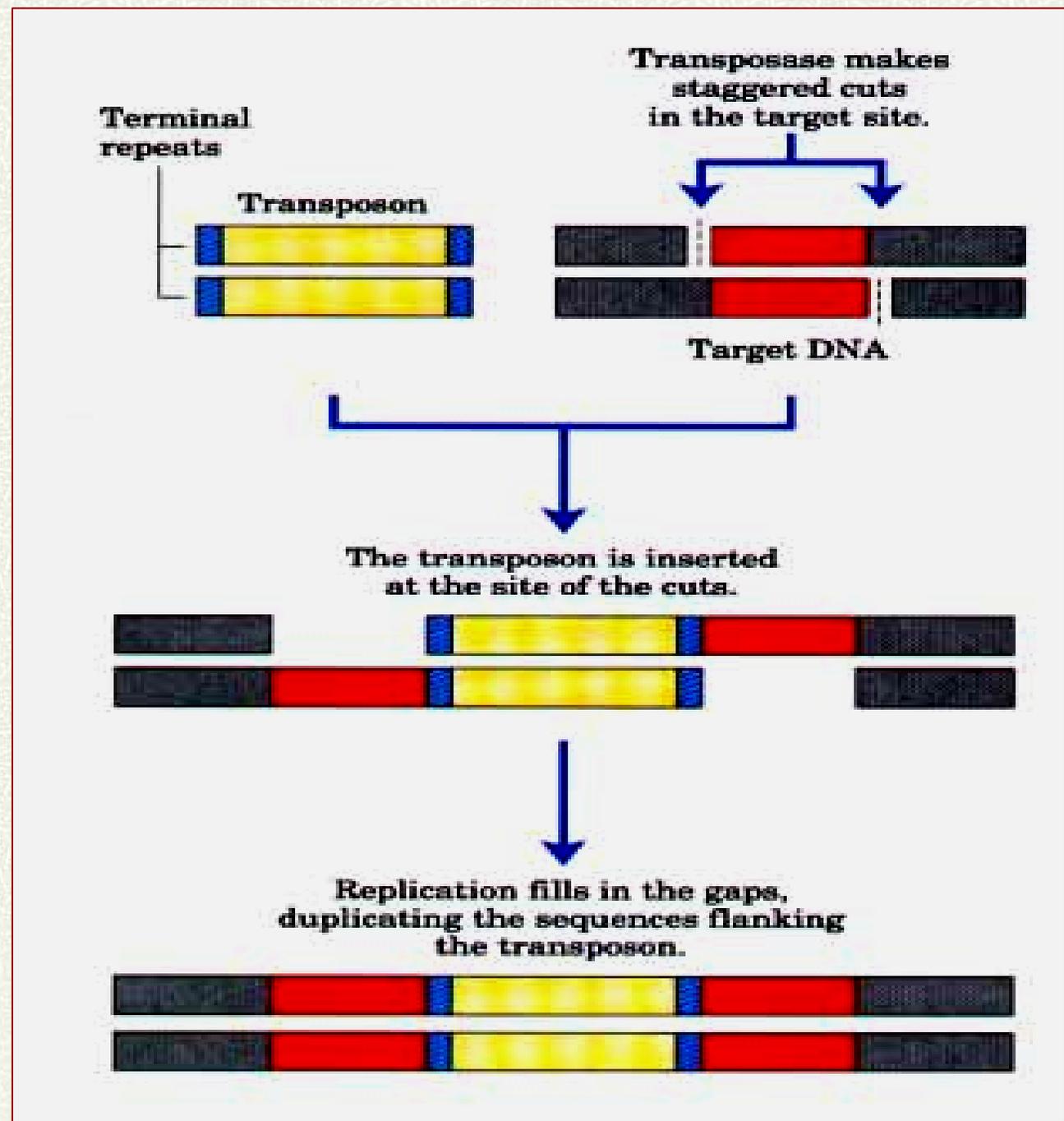


2.6 DNA的转座 (DNA Transposition)

DNA的转座，或称**移位**（**transposition**），是由**可移位因子**（**transposable element**）介导的**遗传物质重排**现象。



转座过程示意图





2.6.1 转座子的分类和结构特征

转座子 (**transposon, Tn**) 是存在于染色体DNA上可自主复制和位移的基本单位。

最简单的转座子不含有任何宿主基因而常被称为**插入序列 (insertional sequence, IS)**，它们是细菌染色体或质粒DNA的正常组成部分。一个细菌细胞常带有少于**10个IS**序列。



常见的**IS序列**都是很小的**DNA**片段（约**1kb**），**末端具有倒置重复序列**，转座时常复制宿主靶位点**4~15bp**的**DNA**形成正向重复区。

大部分**IS**序列只有一个开放读码框，翻译起点紧挨着第一个倒置重复区，终止点位于第二个倒置重复区或附近。

IS1含有两个分开的读码框，只有移码通读才能产生功能型转座酶。

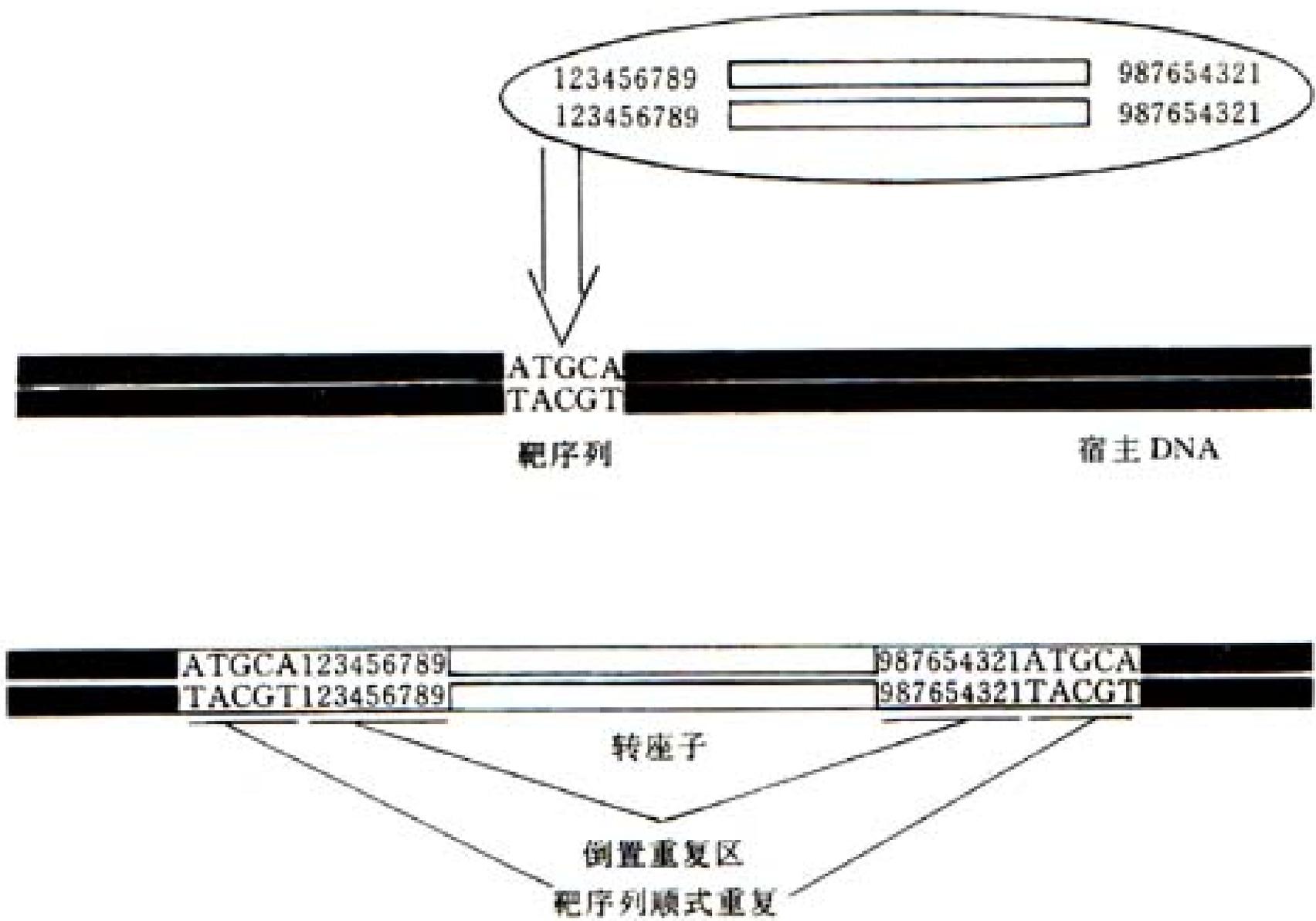


图 2-31 DNA 转座的一般模式



表2-13 IS序列的结构特征比较

	长度 (bp)	两端倒置重复区 (bp)	靶位点正向重复区 (bp)	靶位点
IS1	768	23	9	随机
IS2	1327	41	5	有热点
IS4	1428	18	11或12	AAAN20TTT
IS5	1195	16	4	有热点
IS10R	1329	22	9	NGCTNAGC N
IS50R	1531	9	9	有热点
IS903	1057	18	9	未知



复合式转座子（**composite transposon**）是一类带有某些**抗药性基因**（或其他宿主基因）的转座子，其两翼往往是**两个相同或高度同源的IS序列**。

一旦形成复合转座子，**IS序列**就不能再单独移动，因为它们的功能被修饰了，只能作为**复合体移动**。

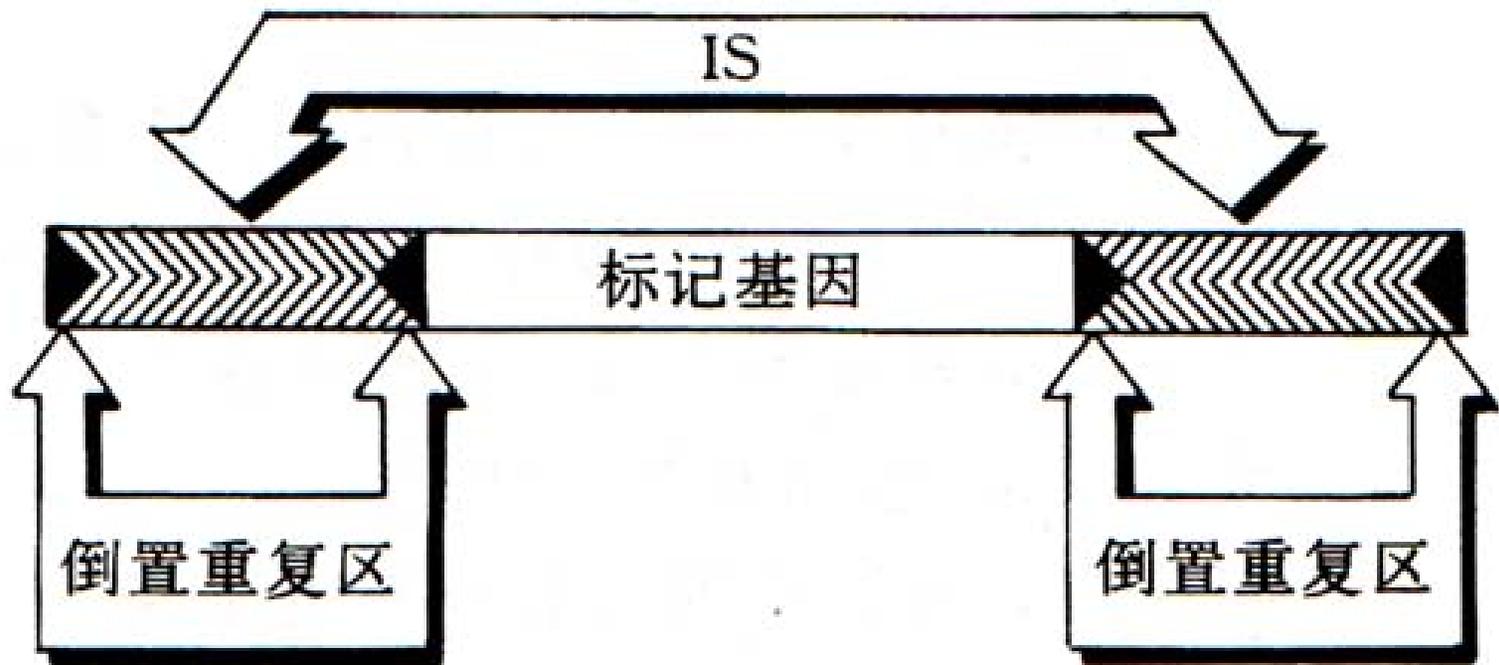


图 2-32 复合转座子的两端往往各有一个 IS 序列，在每个 IS 序列两侧各有倒置重复序列



除了末端带有IS序列的复合转座子以外，还存在一些没有IS序列的、体积庞大的转座子

(5 000bp以上) ——TnA家族。常带有3个基因，一个编码 β -内酰胺酶(AmpR)，另两个则是转座作用所必须的。

所有TnA类转座子两翼都带有38bp的倒置重复序列。

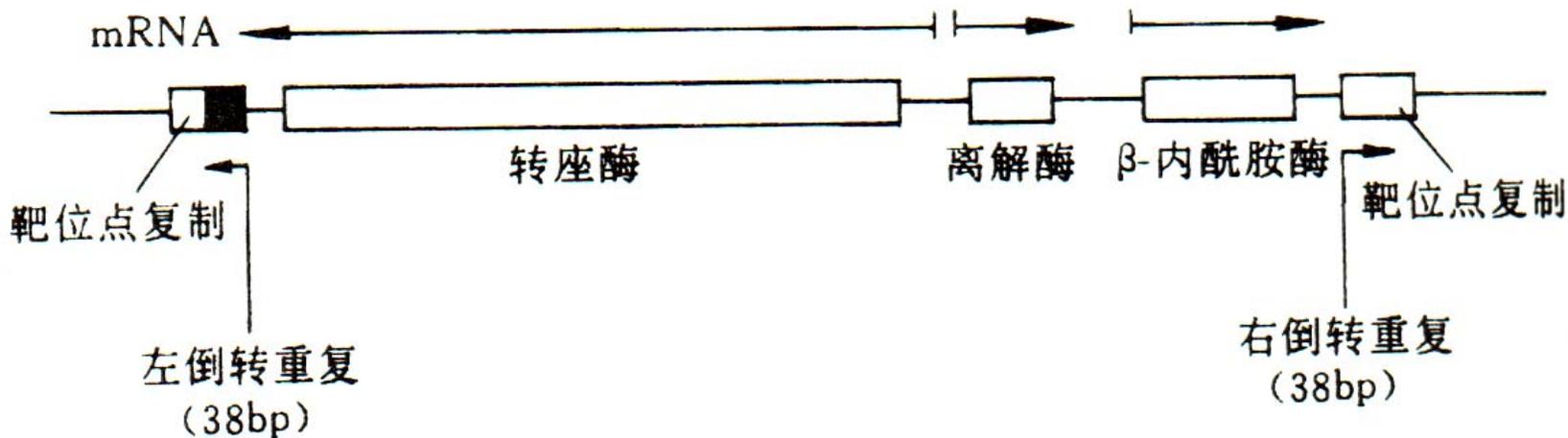


图 2-33 转座子 TnA 的结构示意图



2.6.2 转座作用的机制

转座发生时，受体分子中有一段很短的（3~12bp）、被称为**靶序列**的DNA会被复制，使插入的转座子位于两个重复的靶序列之间。

不同转座子的靶序列长度不同，但特定转座子所复制的靶序列长度是一样的。IS1两翼总有9个碱基对的靶序列，而Tn3两端总有5bp的靶序列。



在复制性转座中，整个转座子被复制，所移动的仅仅是原转座子的拷贝。**转座酶**（**transposase**）和**解离酶**（**resolvase**）分别作用于原始及复制转座子。**TnA**类主要是这种形式。

在非复制性转座中，原始转座子作为一个可移动的实体直接被移位，**IS**序列、**Mu**及**Tn5**等都以这种方式进行转座。



2.6.3 转座作用的遗传学效应

- ①转座引起插入突变，导致结构基因失活。
- ②转座产生新的基因。
- ③转座产生的染色体畸变。
- ④转座引起生物进化。由于转座作用，使某些原来在染色体上相距甚远的基因组合到一起，构建成新的表达单元，产生新的基因和蛋白质分子。

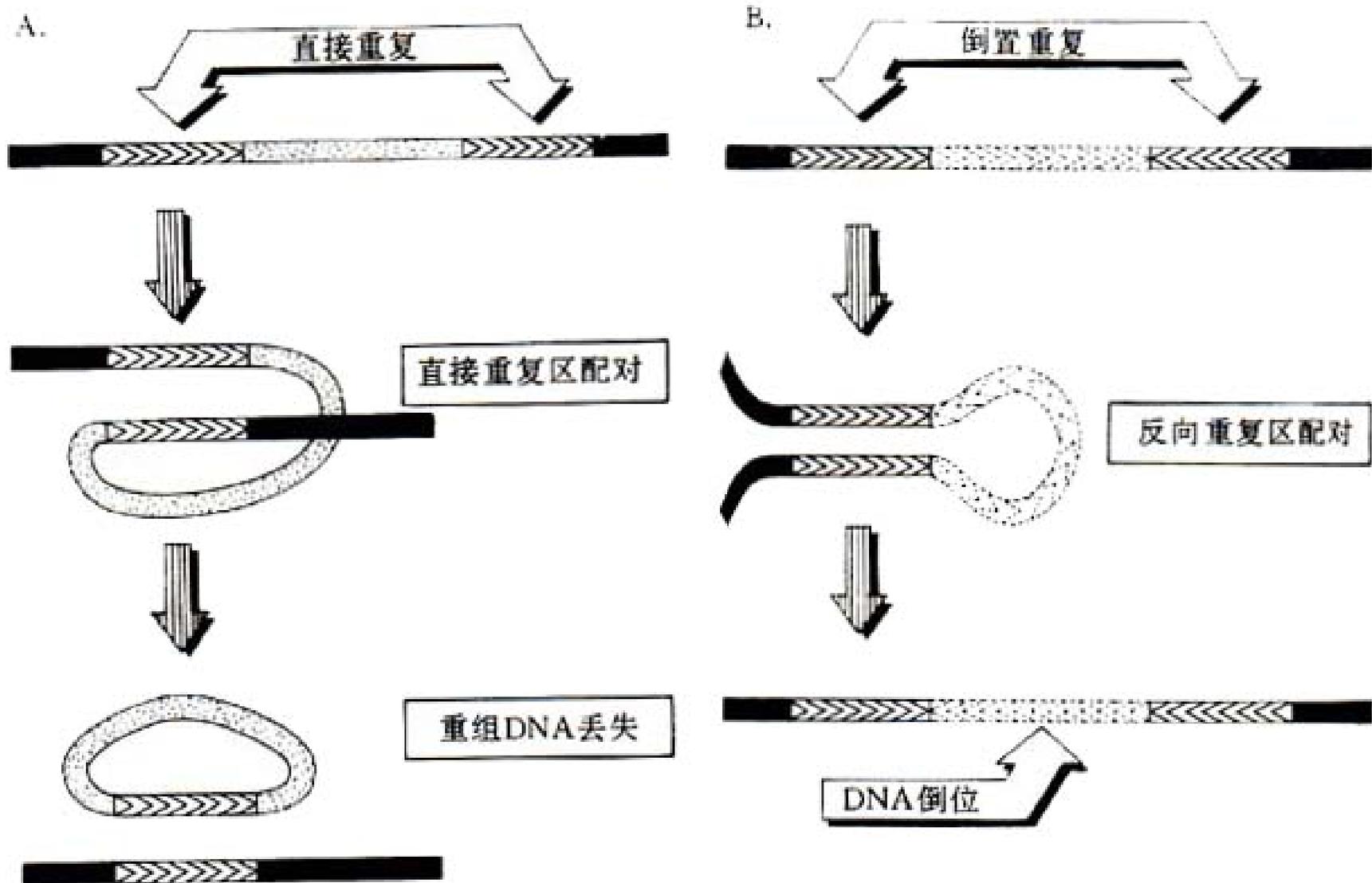


图 2-35 由转座子引起的染色体 DNA 缺失 (A) 或倒位 (B)



2.6.4 真核生物中的转座子

1 玉米中的控制因子（controlling element）- 转座子

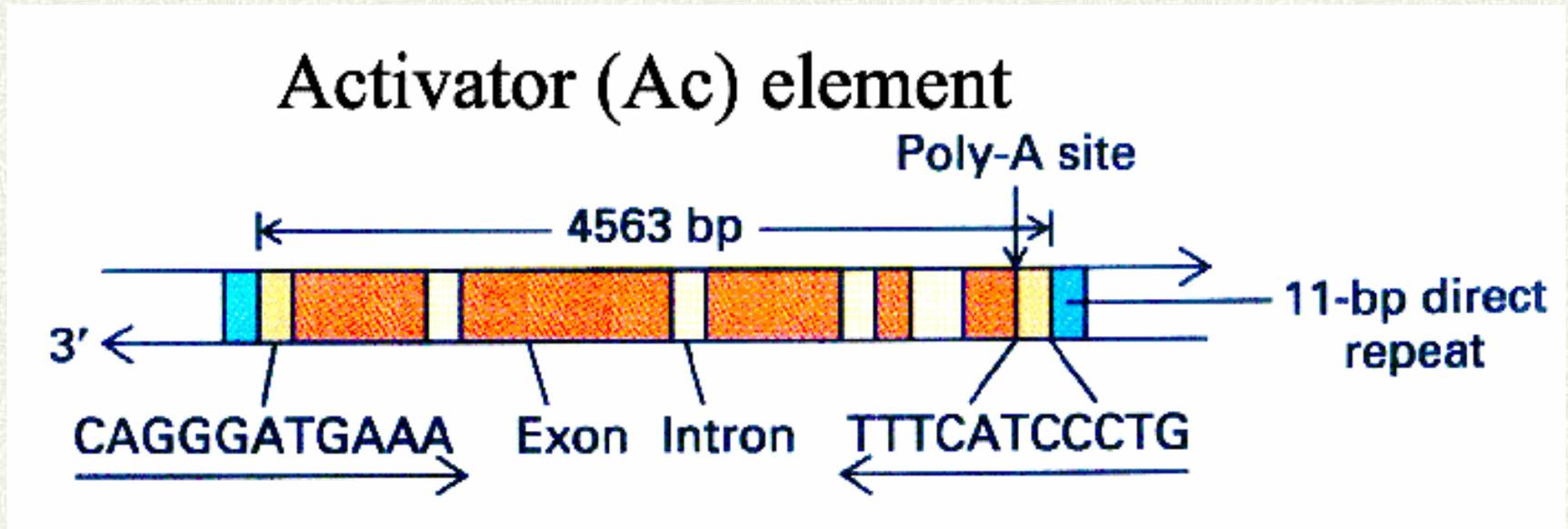
玉米中的控制因子分为两类，即自主性因子和非自主性因子。前者具有自主剪接和转座的功能；后者单独存在时是稳定的，不能转座，当基因组中存在与非自主性因子同家族的自主性因子时，它才具备转座功能。

同一家族的自主性因子能为非自主性因子的转座提供反式作用蛋白（转座酶）。



在著名的**Ac-Ds**体系中，自主转座子**Ac**长**4563bp**，转录生成**3500bp**的单一成熟**RNA**。

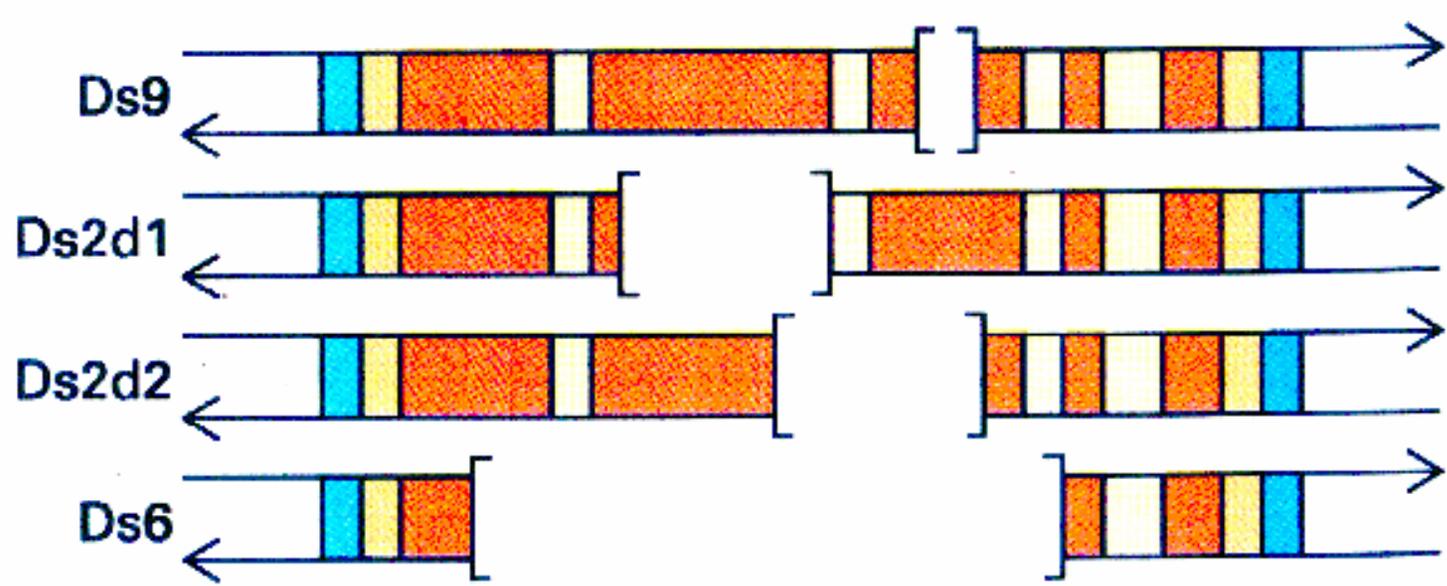
Ac转座子的两翼有**11bp**的倒转重复序列，在其靶**DNA**位点复制形成**8bp**正向重复。

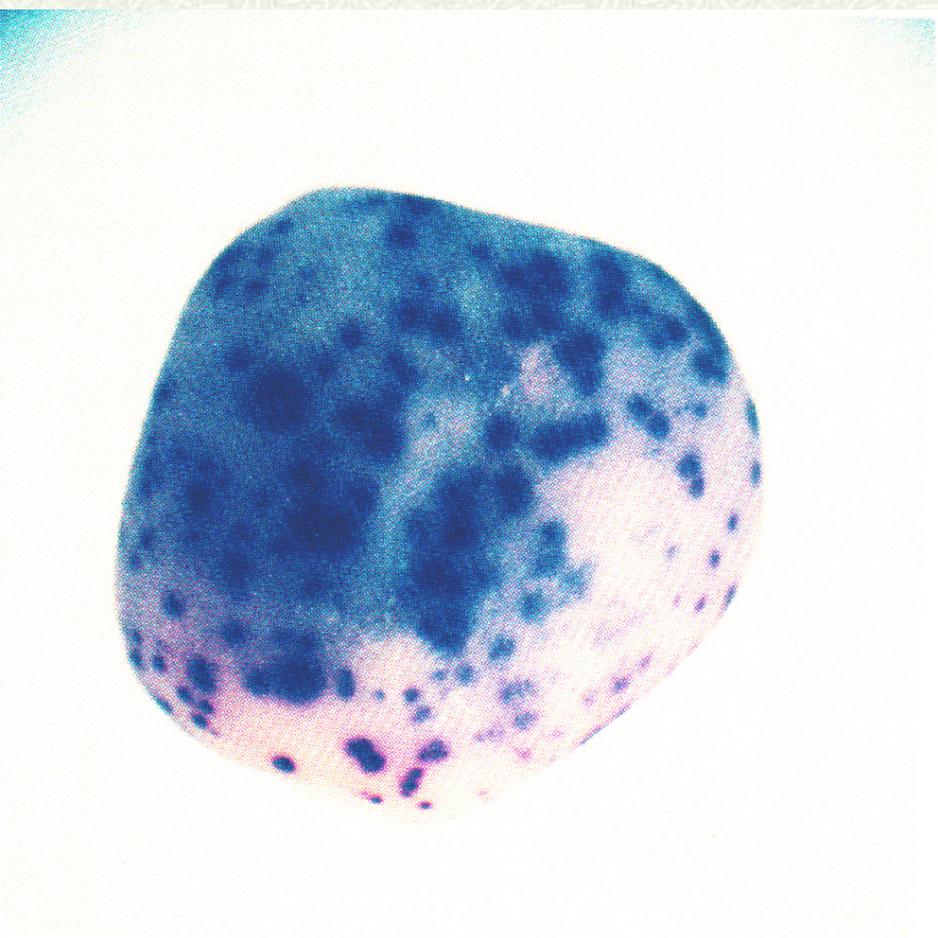
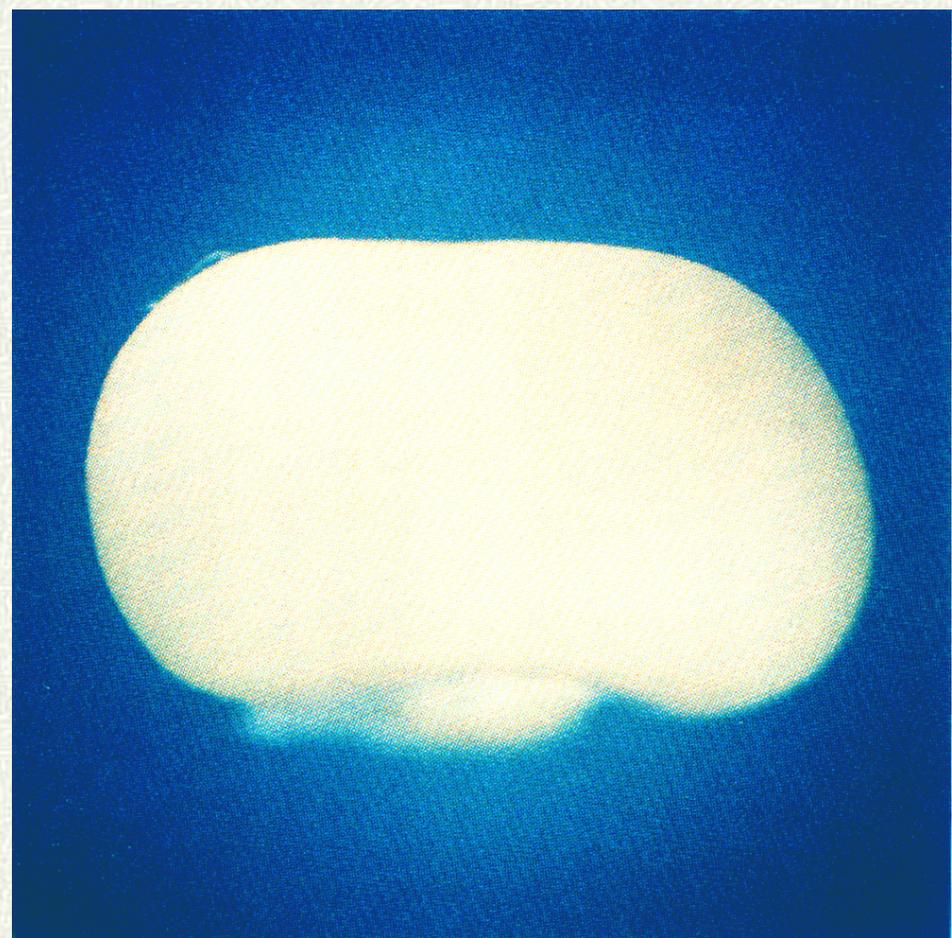




已知所有**Ds**都是**Ac**转座子的缺失突变体，其两端有完整的转座特征序列。

Dissociation (Ds) elements





NLPE&PGE

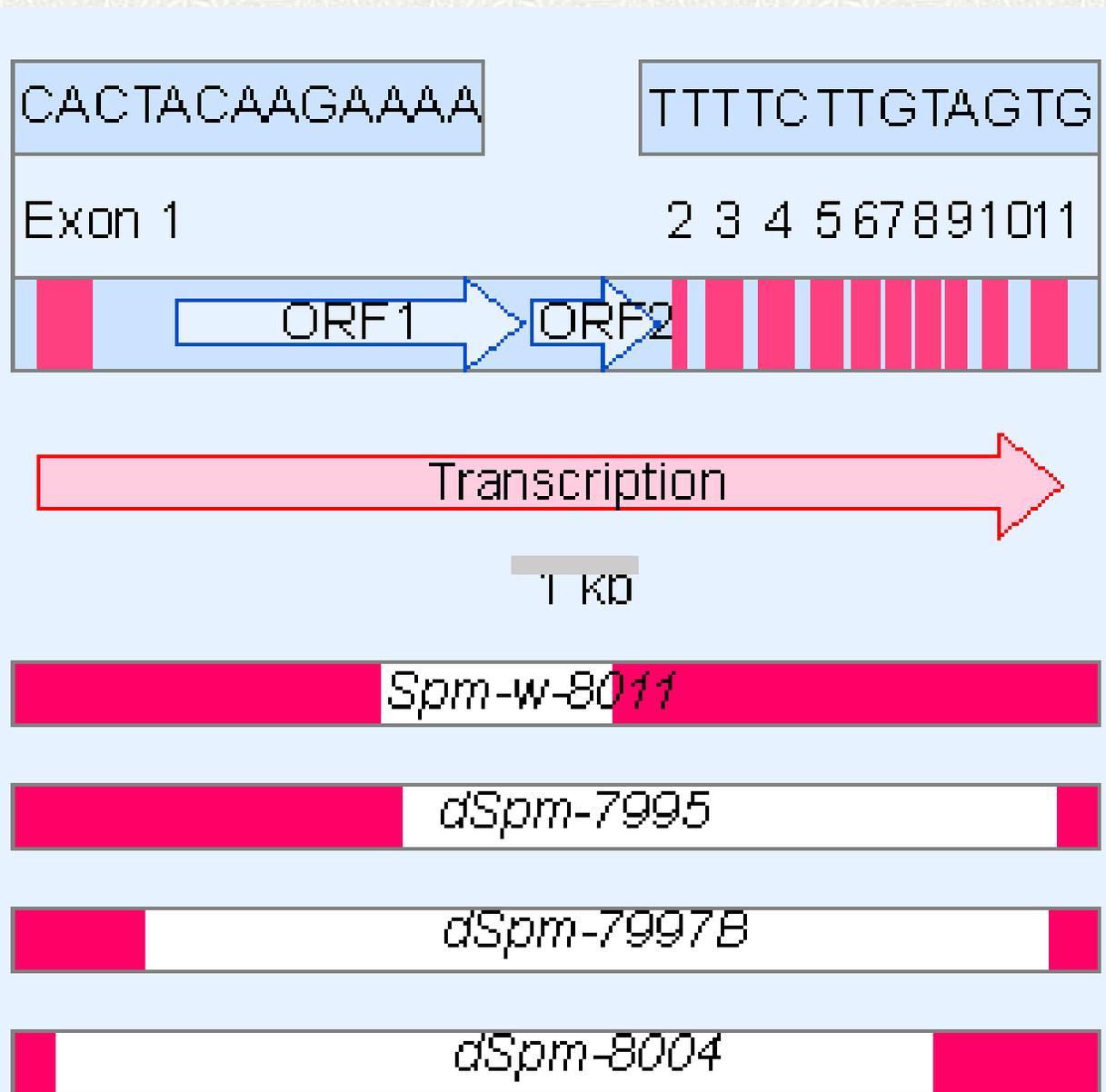


Spm (suppressor-mutator) 和 **En** (enhancer) 属于另一类转座子。

Spm和**En**几乎是完全相同的，它们只有不到**10个碱基的差异**，两端各有一个**13bp**的倒置重复序列，在其靶**DNA**位点复制形成**3bp**正向重复。



图2-36 转座子Spm/En和dSpm非自主转座子的结构比较





Spm和**En**的转录区（即**tnpA**基因）有**8300bp**，成熟**mRNA**为**2500bp**。**tnpA**基因产物的主要功能是**切割转座子序列**。另有两个开放读码框位于**tnpA**基因的内含子中，其**mRNA**的含量大约只有**tnpA**基因产物的**1%**。

由**tnpB**基因编码的这两个蛋白可能直接与转座子两端**13bp**倒转重复区相结合，有利于**DNA**的切割和转移。所有**dSpm**（**defective Spm**）都是功能型**Spm**的缺失突变体。



2 果蝇中的转座子

果蝇中的 **Copia** 转座子具有数百 bp 的末端正向重复和类似于酵母 Ty 转座子及 RNA 肿瘤前病毒的结构，能以高速度转录并产生有多聚腺苷化末端的 RNA。只具有一个长的可阅读框架，基因产物与 RNA 肿瘤病毒逆转录酶有较高的同源性。

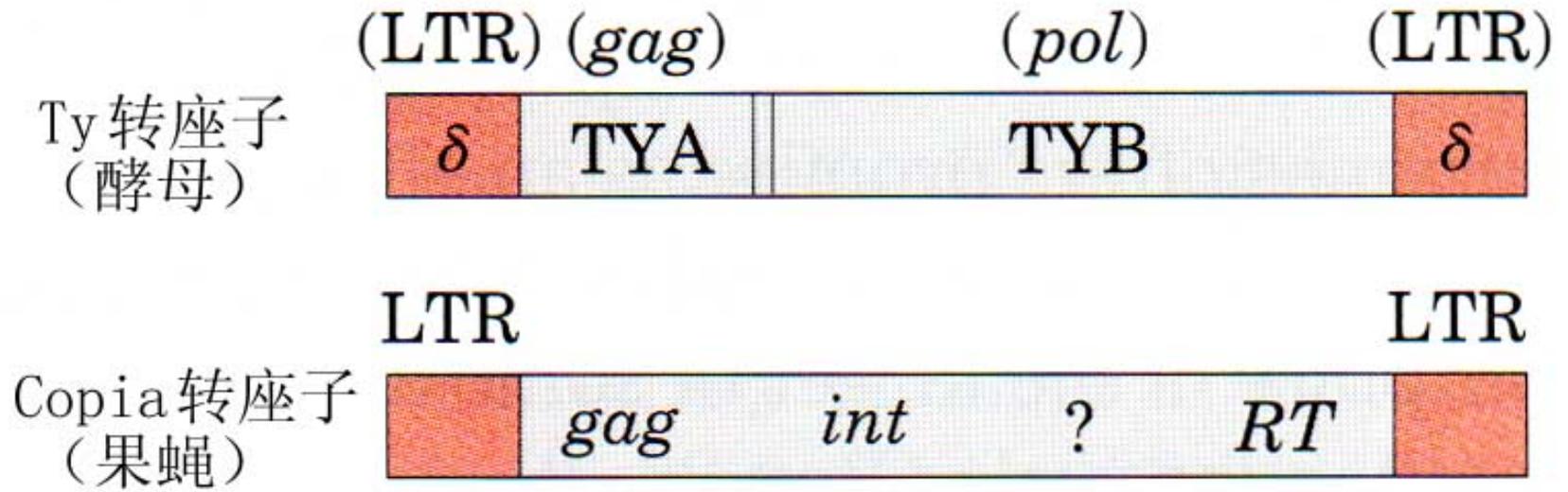
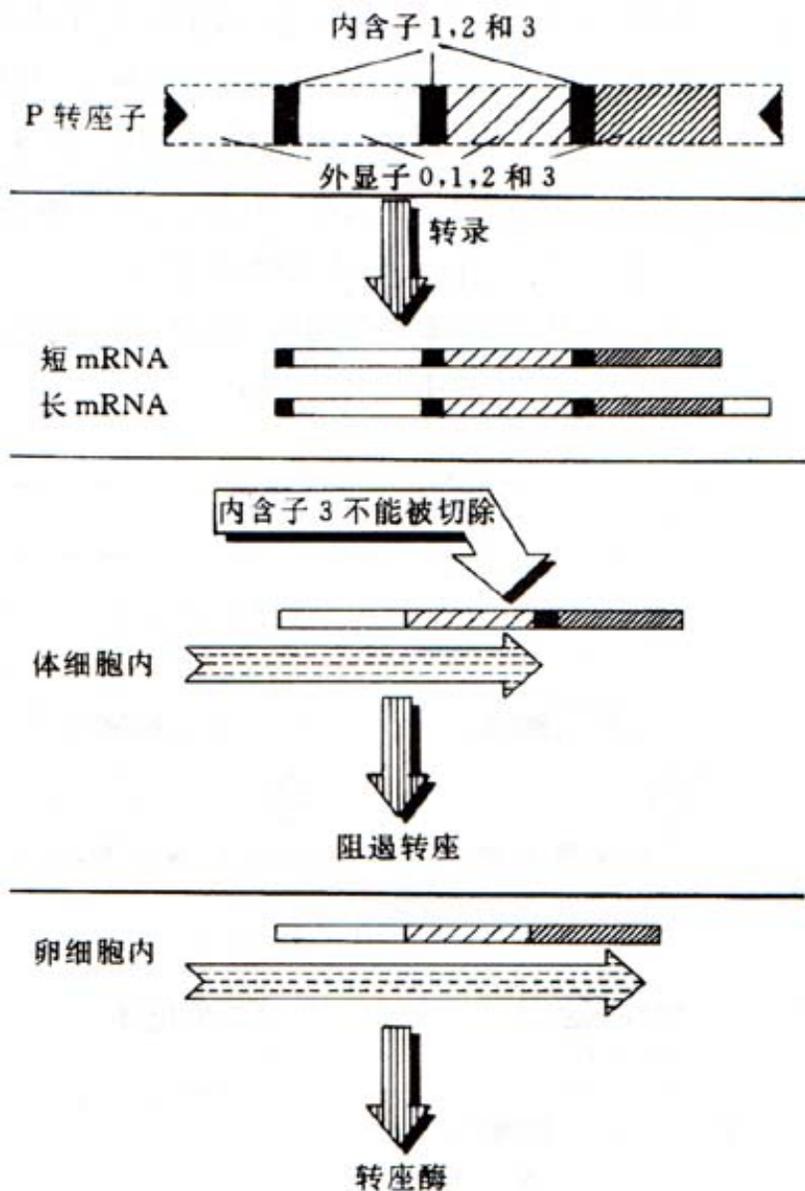


图2-37 真核生物细胞内存在不少结构上类似于反转录病毒的反转录转座子。这类转座子都编码一个与反转录病毒中的反转录酶高度同源的转座酶，左右两边都是与病毒边界序列LTR同源的重复序列。



P转座子 (P element) 属于**非复制型**，两翼都有**31bp**倒置重复序列，转座后导致靶**DNA**复制产生**8bp**正向重复序列。有实验证明，**P**转座子插入果蝇基因组**W**位点引起杂种不育。



P转座子的原初转录产物为**2.5或3.0kb**，包括外显子**0、1、2、3**及**3**个内含子。

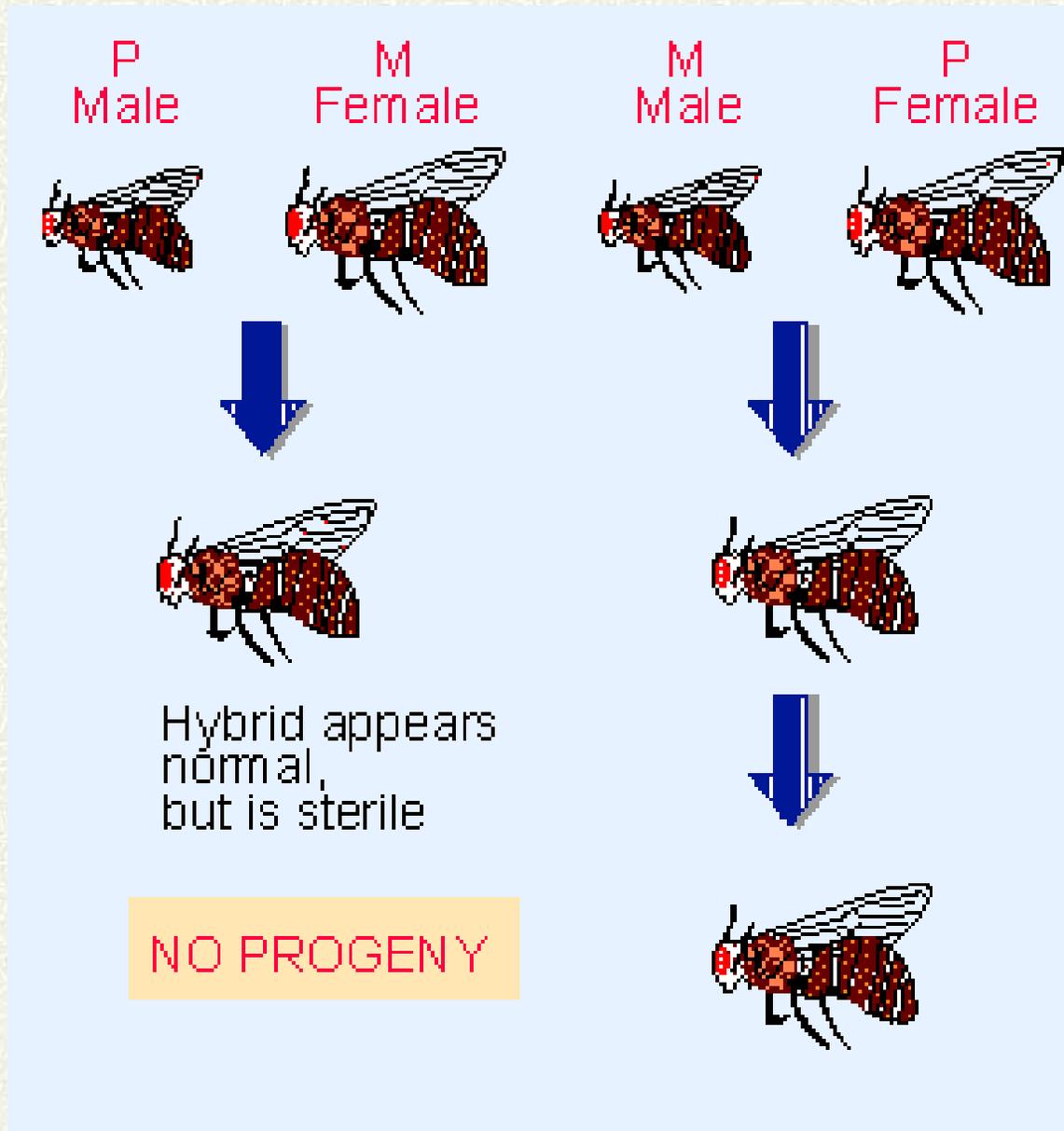
体细胞中，只有前两个内含子能被顺利切除，产生包括外显子**0~2**的功能型**mRNA**并被翻译成**一个 6.6×10^4** 的转座阻遏蛋白。

在卵细胞中，内含子**3**能被切除，所产生的成熟**mRNA**包括全部**4**个外显子并被翻译成**一个 8.7×10^4** 转座酶，导致**P转座子**转座和配子体败育。

图 2-38 果蝇转座子对杂种后代不育性的影响分析



P转座子与杂种不育





P
Male



M
Female



Hybrid appears
normal,
but is sterile

NO PROGENY

M
Male



P
Female

