

中心法则——复制和转录

李璞泽&樊一鸣

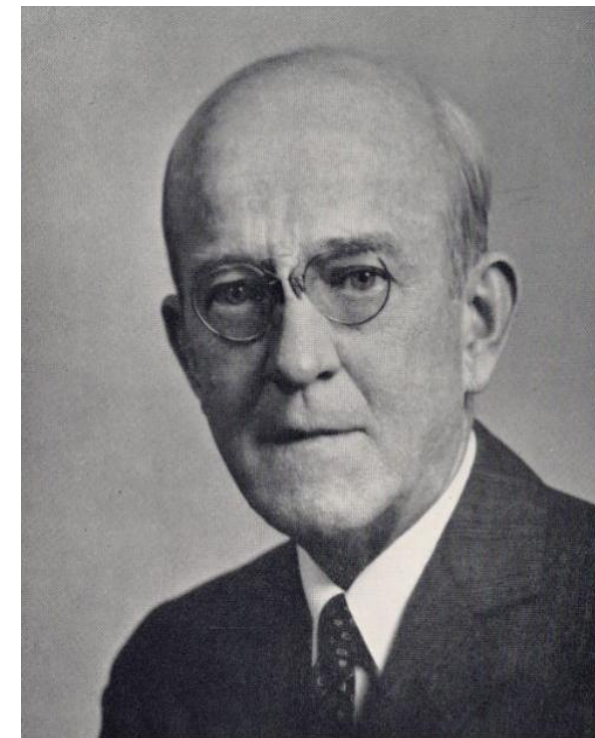
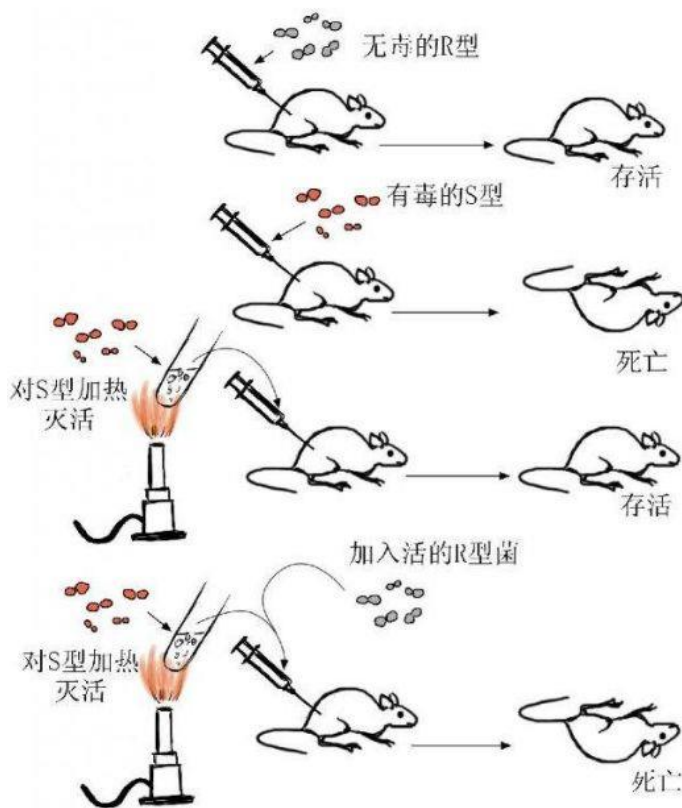
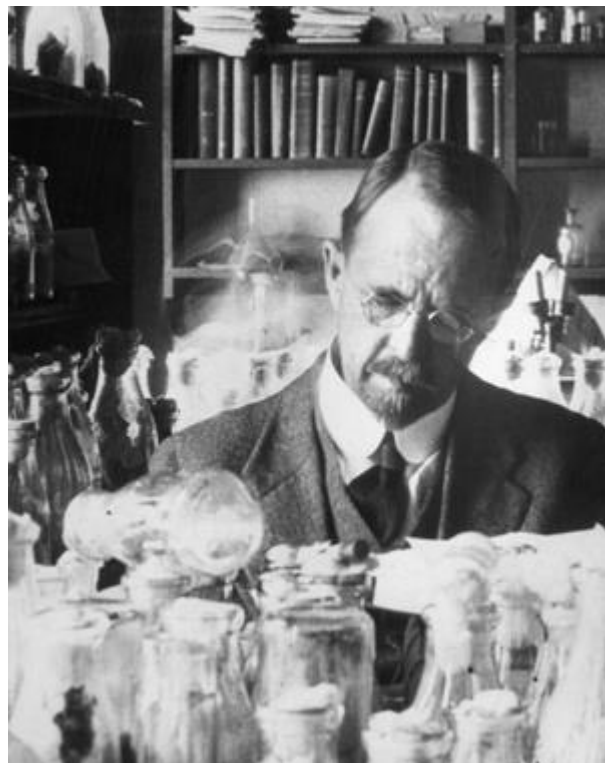
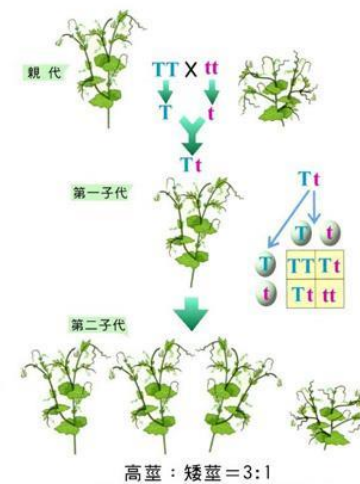
遗传信息的载体——DNA

1866 年 Mendel 开创遗传学。

20世纪初，以Morgan为代表的科学家发展了染色体的遗传学说。

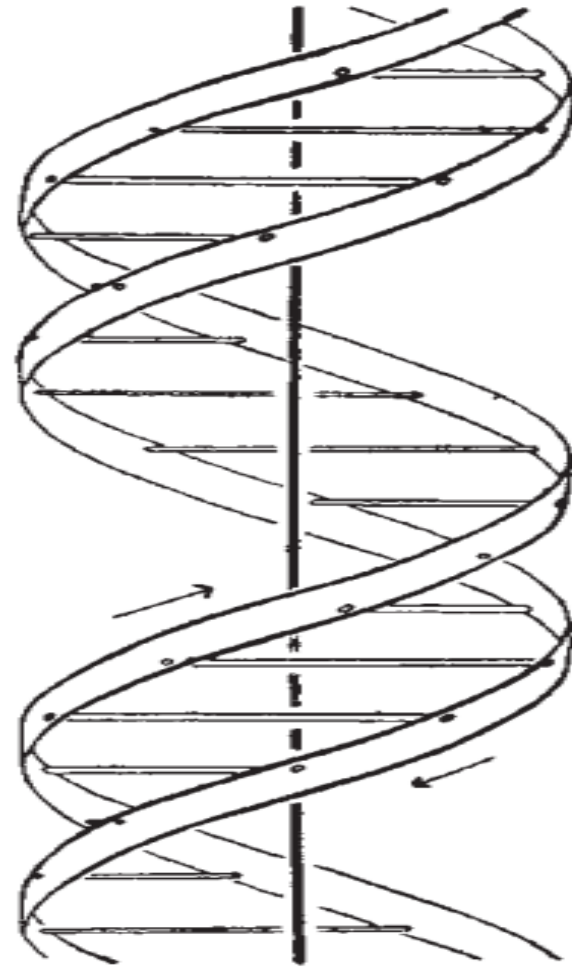
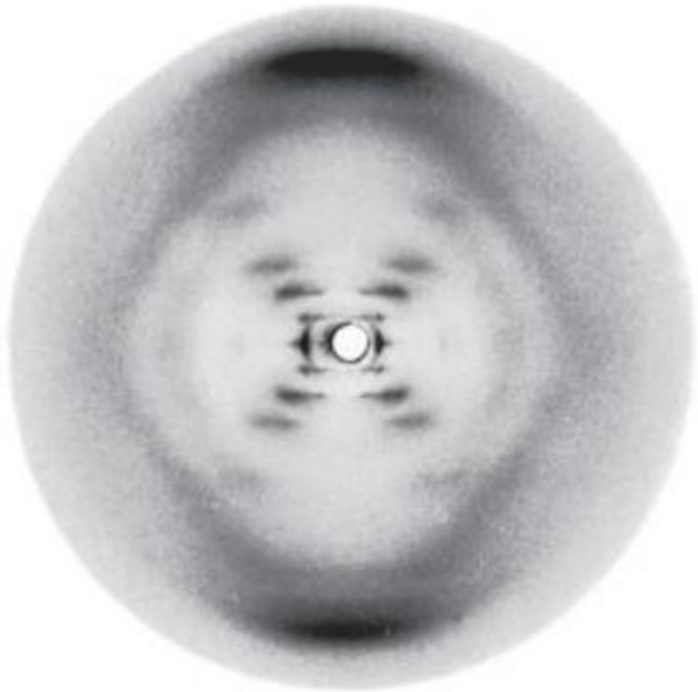
1928年，Griffith设计肺炎球菌的转化实验。

1944年，Avery等通过生物化学方法体外证明DNA的转化活性。

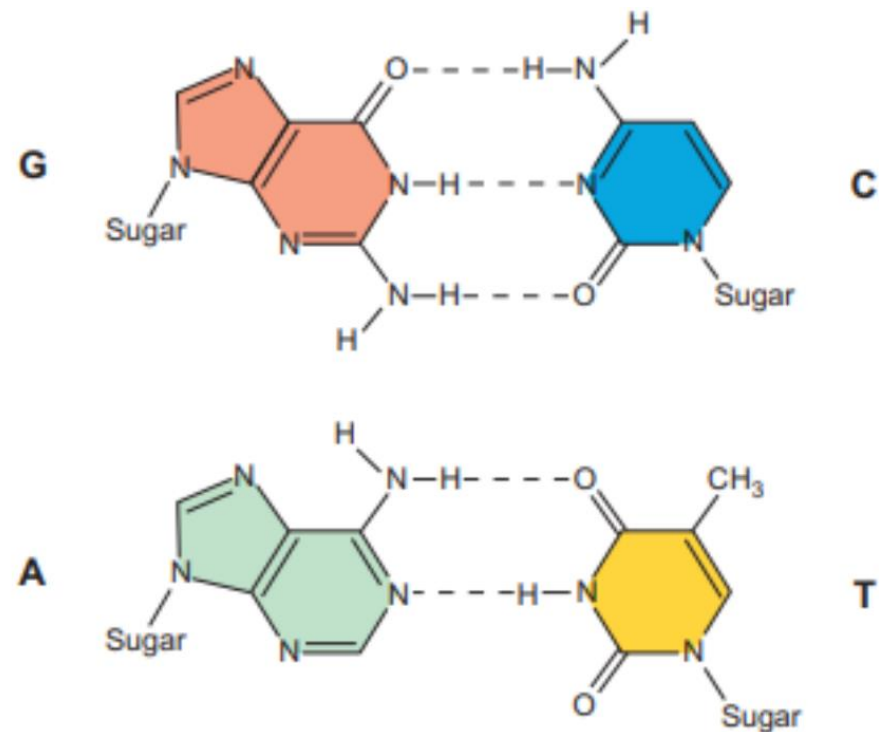
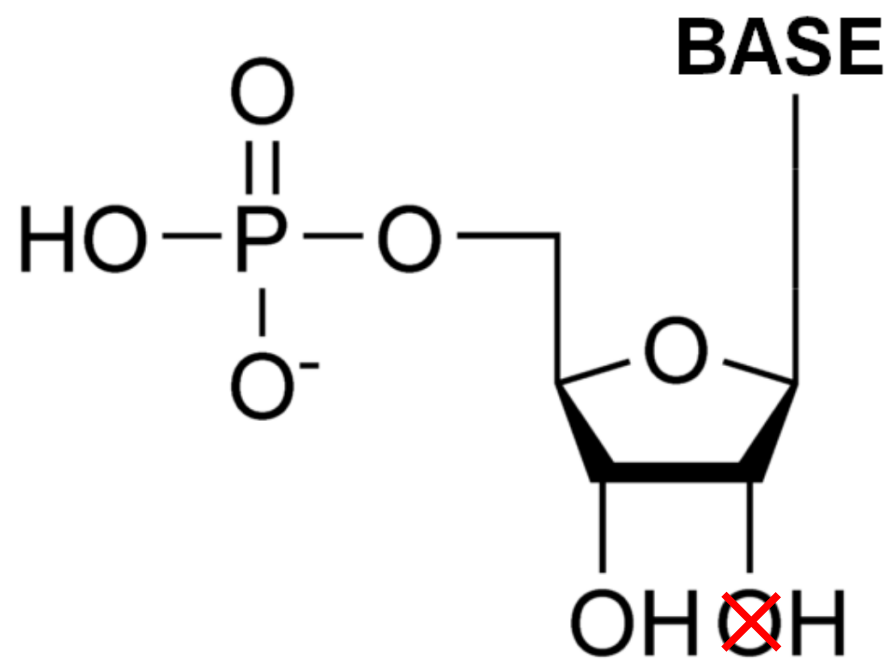


DNA的双螺旋结构

1953年，James D Watson和Francis HC Crick在Franklin和Wilkins对DNA进行X射线衍射实验的基础上提出DNA的双螺旋结构。

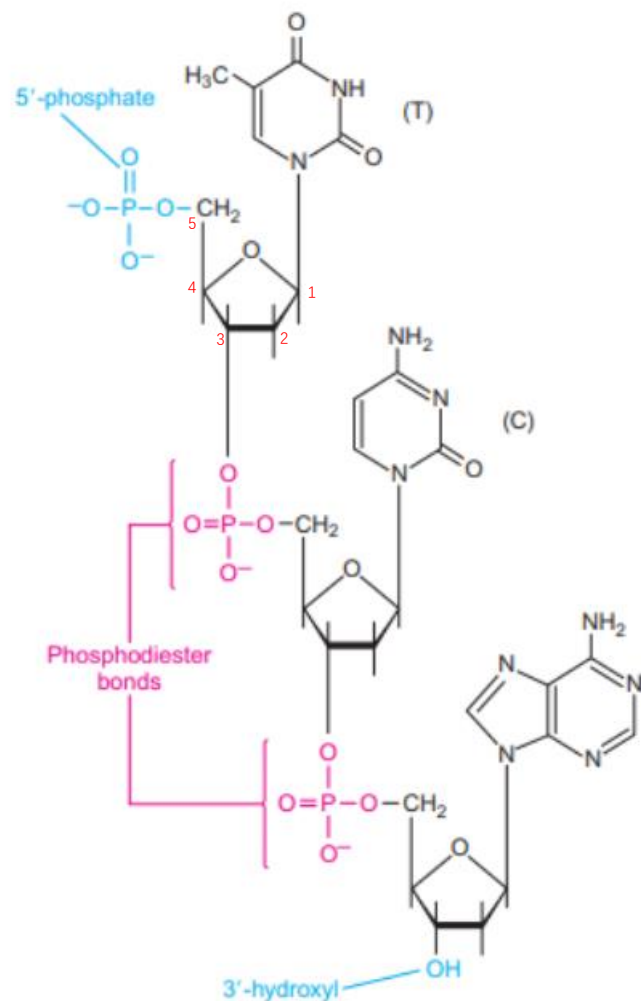


构成DNA链的单体——脱氧核糖核苷酸



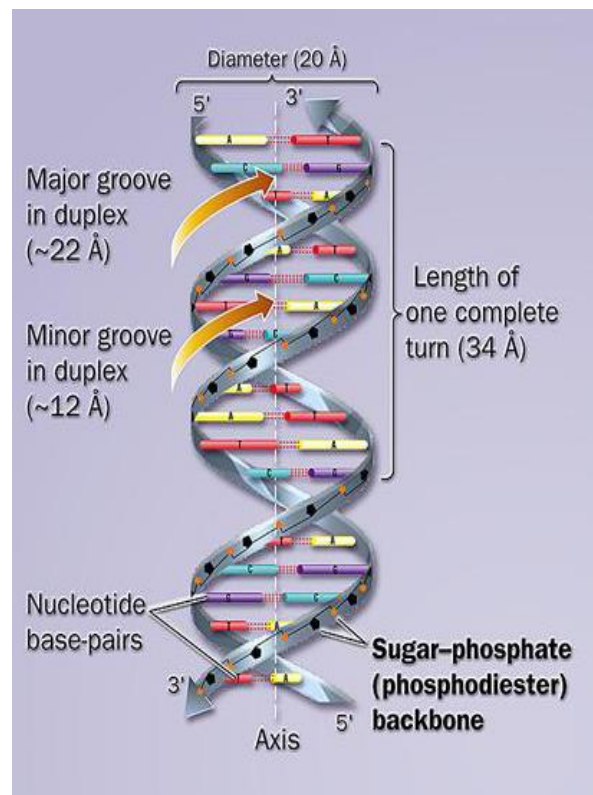
碱基通过氢键互补配对
A=T； C≡G

脱氧核糖核苷酸通过磷酸二酯键形成DNA单链



DNA序列的**方向性**：5'末端磷酸基团→3'末端羧基

5'-TCA-3'
3'-AGT-5'



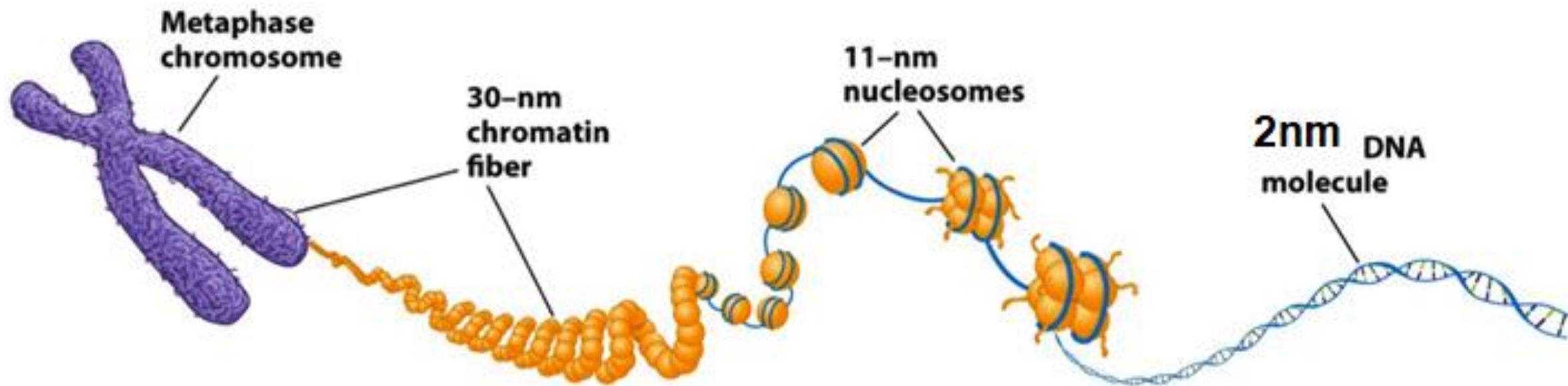
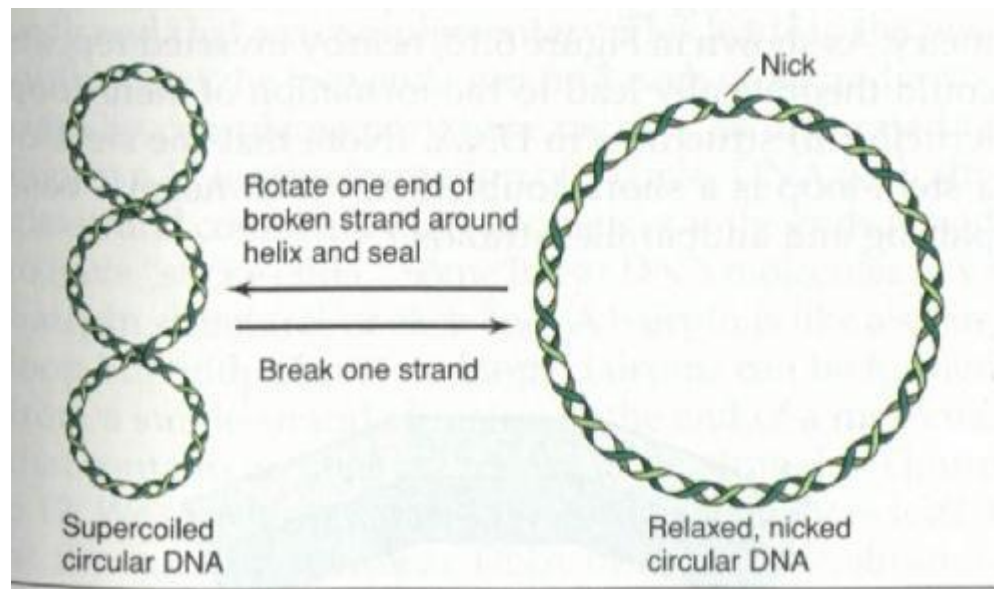
双螺旋为**右手螺旋**；
磷酸、戊糖交替形成骨架；
碱基在内部互补配对

DNA的高级结构

一级结构是碱基的排列顺序

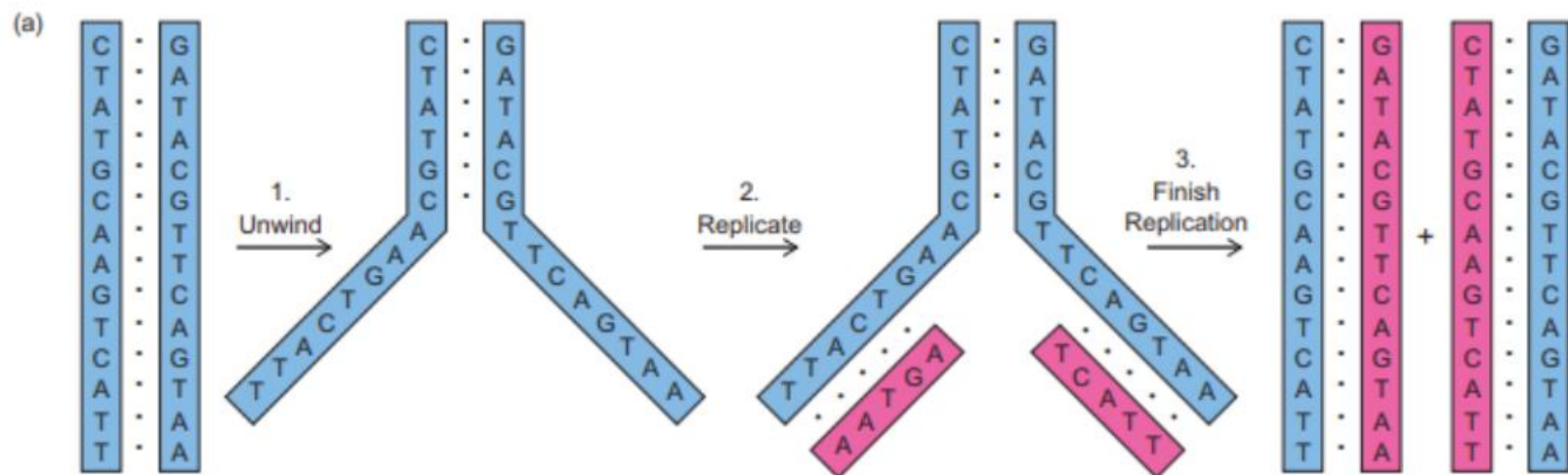
二级结构是双螺旋结构

更高级结构是超螺旋和染色体的包装



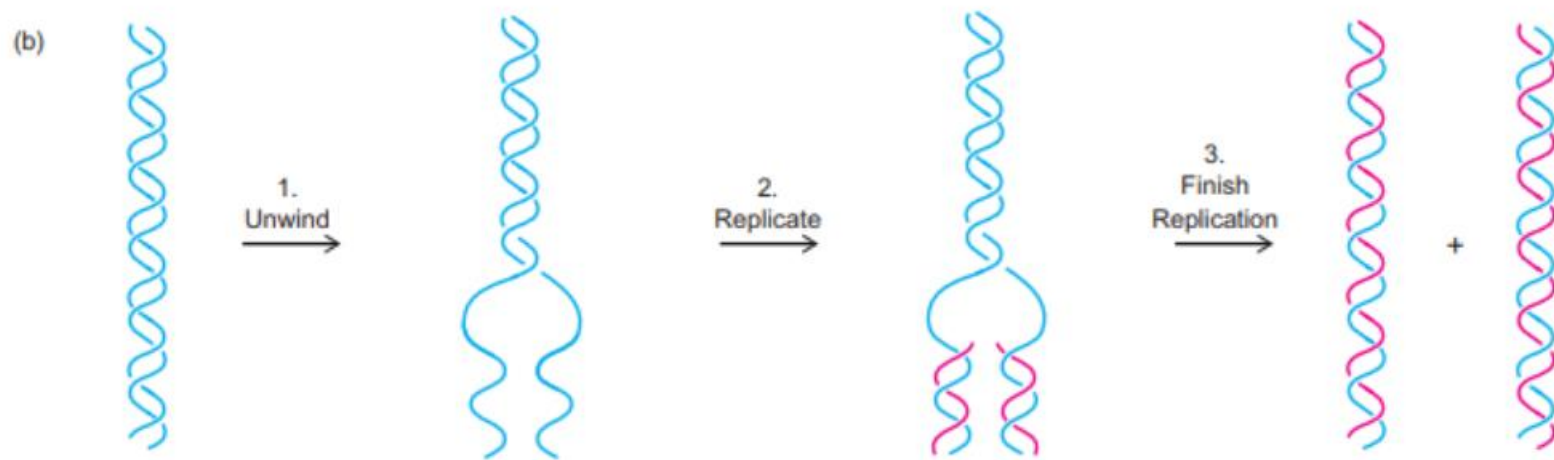
DNA的半保留复制

在提出双螺旋结构的同年（1953），Watson 和 Crick立刻提出了DNA双链复制的可能方式。

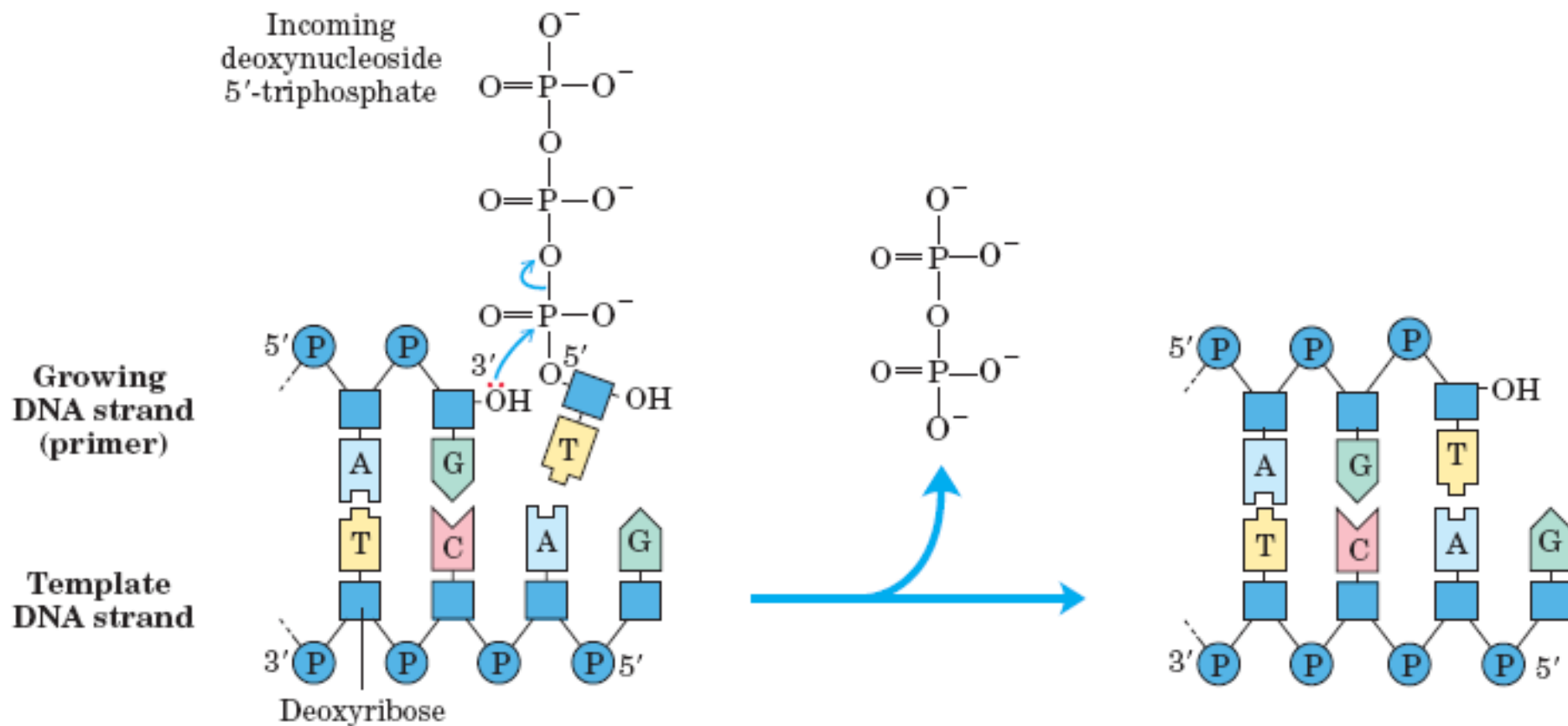


两条链各自可作为模板，依据碱基互补配对原则，形成新的DNA双链。

这种复制方式同时提供了基因突变的分子基础。



DNA复制的化学基础

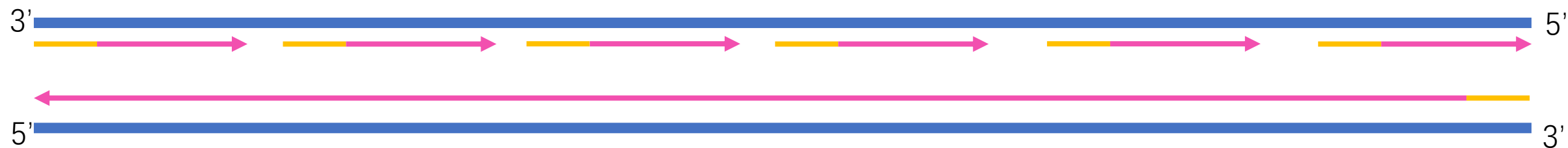


- 真核生物进行复制延伸主要依靠的酶是**DNA聚合酶III**;
- DNA聚合酶III**只能在已有的核酸链末端添加新的脱氧核苷酸**, 而不能从第一个脱氧核苷酸开始合成;
- DNA的复制**具有方向性**, 只能由5'端向3'端进行延伸。

半不连续复制



- 一条链按照5'→3'方向连续复制;
- 另一条链按照局部5'→3'方向不连续复制, 需要RNA引物



半不连续复制



- 一条链按照5'→3'方向连续复制；
- 另一条链按照局部5'→3'方向不连续复制，需要RNA引物，形成冈崎片段。



半不连续复制



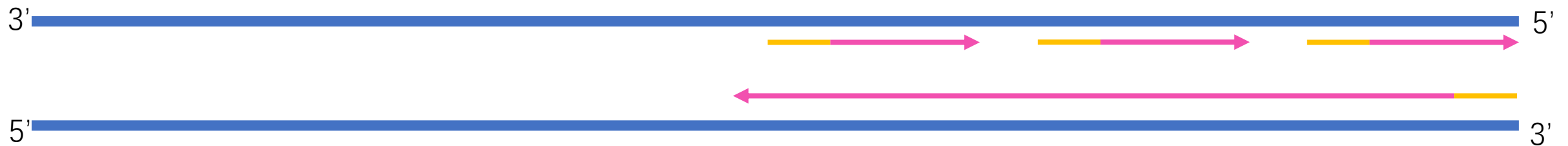
- 一条链按照5'→3'方向连续复制；
- 另一条链按照局部5'→3'方向不连续复制，需要RNA引物，形成冈崎片段。



半不连续复制



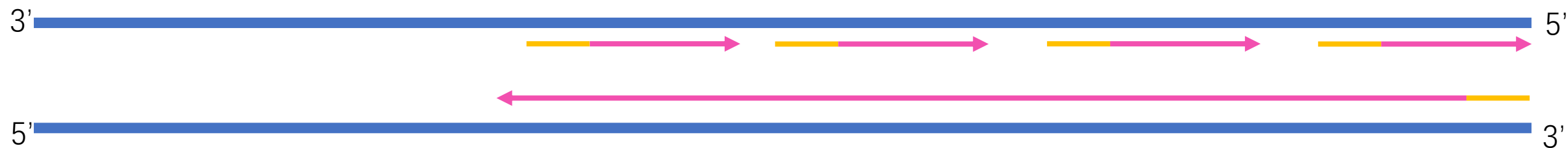
- 一条链按照5'→3'方向连续复制；
- 另一条链按照局部5'→3'方向不连续复制，需要RNA引物，形成冈崎片段。



半不连续复制



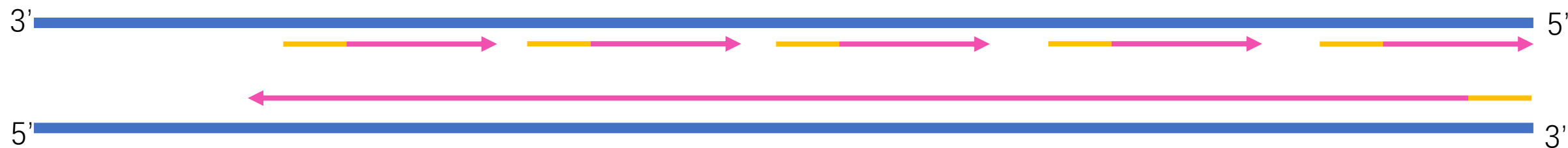
- 一条链按照5'→3'方向连续复制；
- 另一条链按照局部5'→3'方向不连续复制，需要RNA引物，形成冈崎片段。



半不连续复制



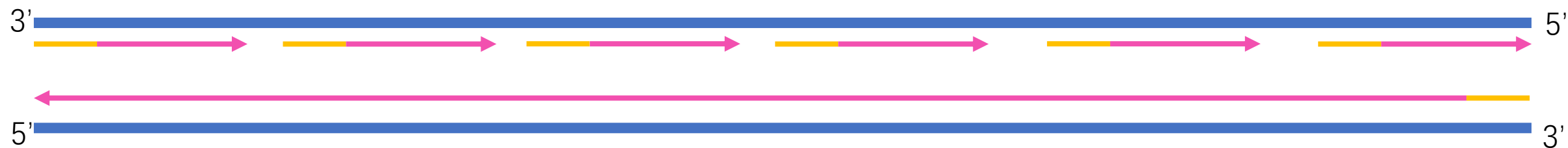
- 一条链按照5'→3'方向连续复制；
- 另一条链按照局部5'→3'方向不连续复制，需要RNA引物，形成冈崎片段。



半不连续复制



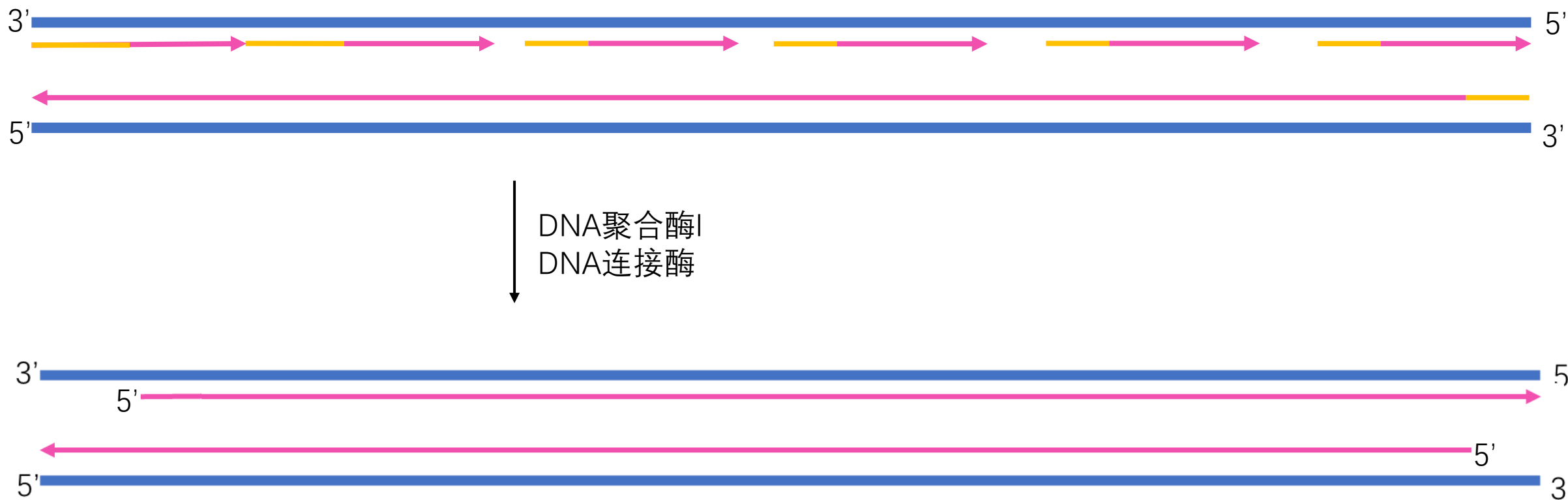
- 一条链按照5'→3'方向连续复制；
- 另一条链按照局部5'→3'方向不连续复制，需要RNA引物，形成冈崎片段。



冈崎片段的连接

DNA聚合酶I：同时具有聚合酶活性和外切酶活性，实现切口向3'方向平移；

DNA连接酶：连接切口



结果：完成了两条链的DNA复制，但两条模板链3'末端均有一部分序列在复制中丢失。在真核生物染色体中，丢失的部分是一端叫做“**端粒**”的序列。

端粒和端粒酶

端粒(telomere): 真核生物染色体线性DNA分子末端的结构。

结构特点:

- 由末端单链DNA序列和蛋白质构成。
- 末端DNA序列是多次重复的TxGy短序列，人 TTAGGG，互补链CyAx短序列。存在TG单链区段。端粒DNA长度从几十到几千bp不等。
- 大部分体细胞端粒不可延长，端粒序列随分裂次数而逐渐缩短。



功能:

- 维持染色体的稳定性和DNA复制的完整性

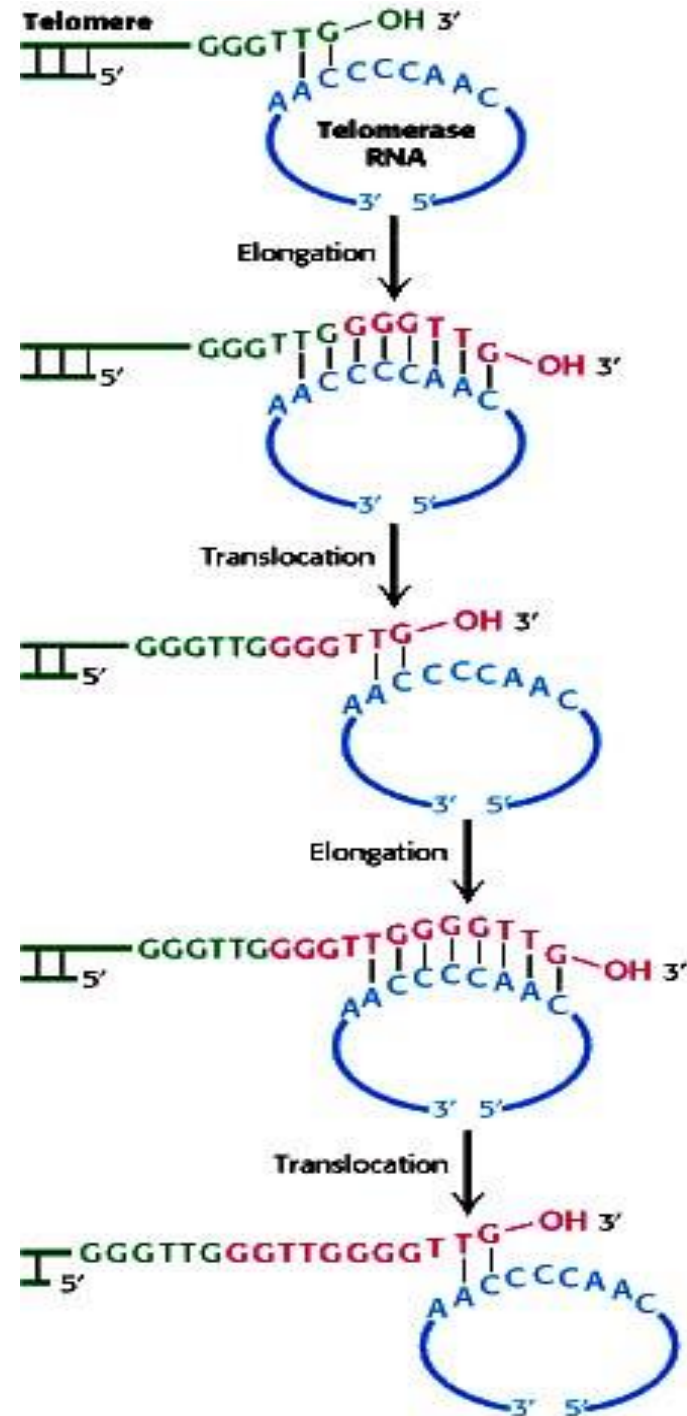
端粒和端粒酶

端粒酶

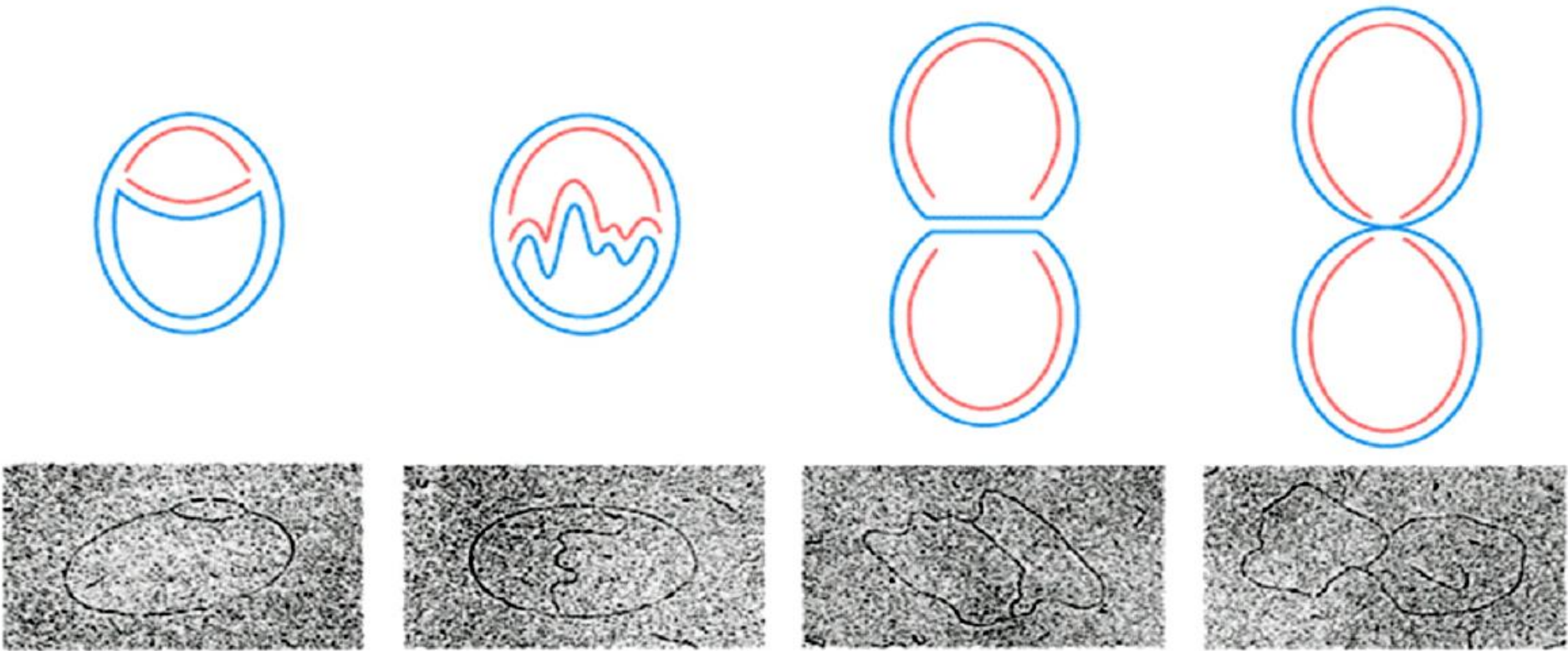
作用：延长端粒；

存在：大部分体细胞中端粒酶被抑制，而在**造血细胞、干细胞和生殖细胞**中端粒酶活性较高；绝大多数**肿瘤**中端粒酶活性很高；

结构：逆转录酶，以自身携带的RNA序列为模板知道DNA的合成。



原核生物的拟核和质粒为环状DNA双链，不存在端粒，也不存在复制末端的序列丢失问题。



复制过程中的错误和修复

点突变的结果

- 错义突变：由于密码子的突变导致蛋白质中原来的氨基酸被另一种氨基酸取代为错义突变
- 无义突变：当氨基酸密码子突变为终止密码子时为无义突变。它导致翻译提前结束而产物失活。
- 沉默突变：突变影响非必需DNA或突变对基因功能的影响可忽略。
- 回复突变：突变克服第一次突变造成的影响。

框移突变

正常

5'G C A G U A C A U G U C

丙 缬 组 缬

缺失C

5'G A G U A C A U G U C

谷 酪 蛋 丝

转录

基因的转录：

主要的酶：RNA聚合酶

催化的反应： $(\text{NMP})_n + \text{NTP} \rightarrow (\text{NMP})_{n+1} + \text{PPi}$

底物：DNA模版；NTP（ATP，UTP，GTP，CTP）

聚合方向：5'→3'

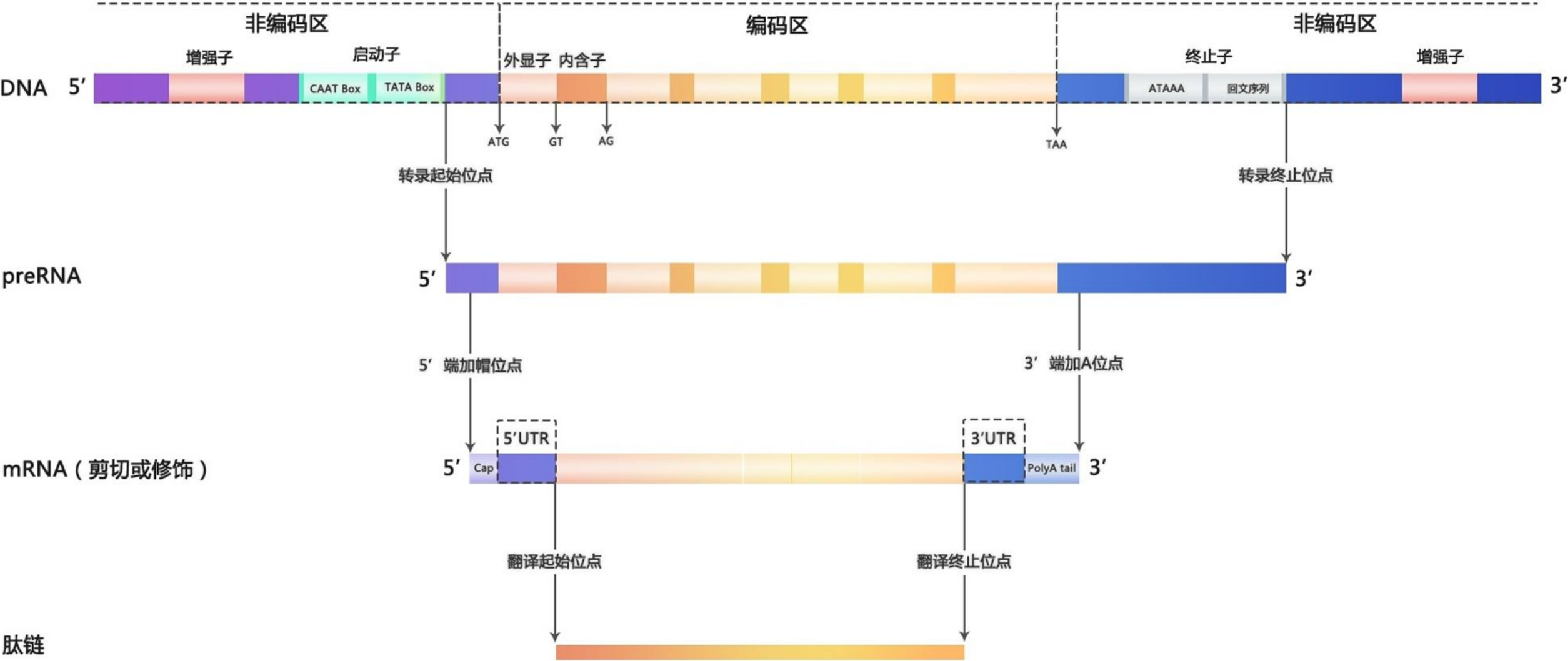
催化特点：不需要引物；起始于启动子；通常以GTP或ATP开始； Mg^{2+}

无外切酶活性：缺乏具校正功能的3'→5'外切活性

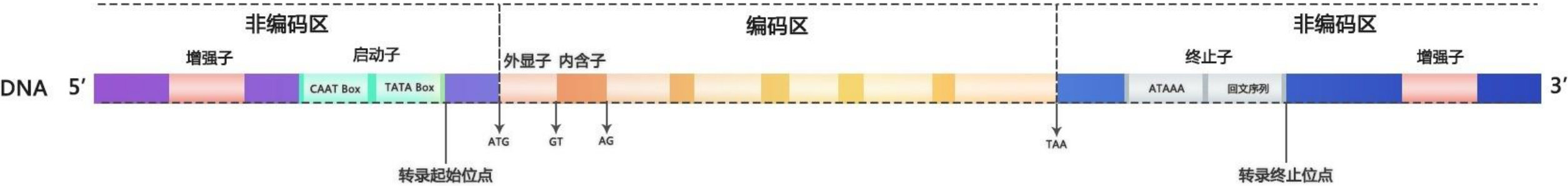
模板：DNA双链中仅一股单链作转录模板。



真核生物基因的结构



真核生物的转录起始



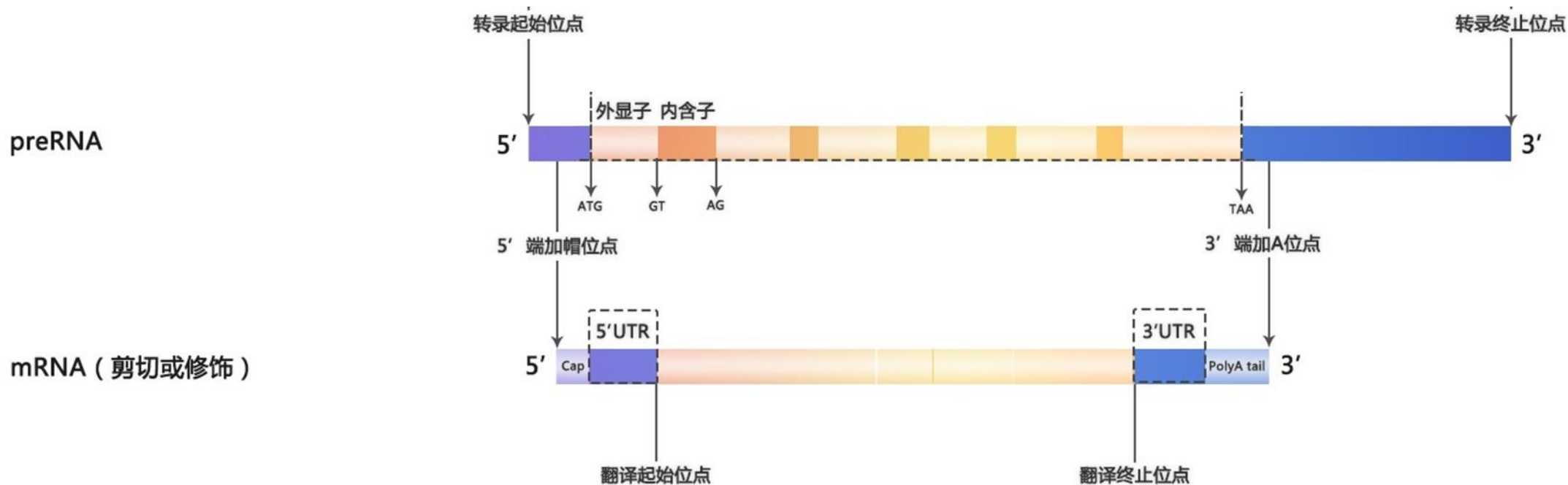
编码区(Coding Region)是具有翻译成蛋白潜力的DNA序列，非编码区(Non-coding Region)不能翻译形成蛋白质，但发挥重要的调控作用。非编码区所占的序列比例远远大于编码区。

启动子(promoter): 位于编码区上游，RNA聚合酶和转录因子识别和组装的区域，转录起始于启动子3'端的转录起始位点(Transcription Starting Site, TSS); 不同启动子的基础转录效率差异很大; 存在共有碱基序列。

终止子(terminator): 位于转录产物末端，是提供转录停止信号的的DNA序列。

增强子(enhancer): 通过与转录因子相互作用提高转录效率，其位置不固定。

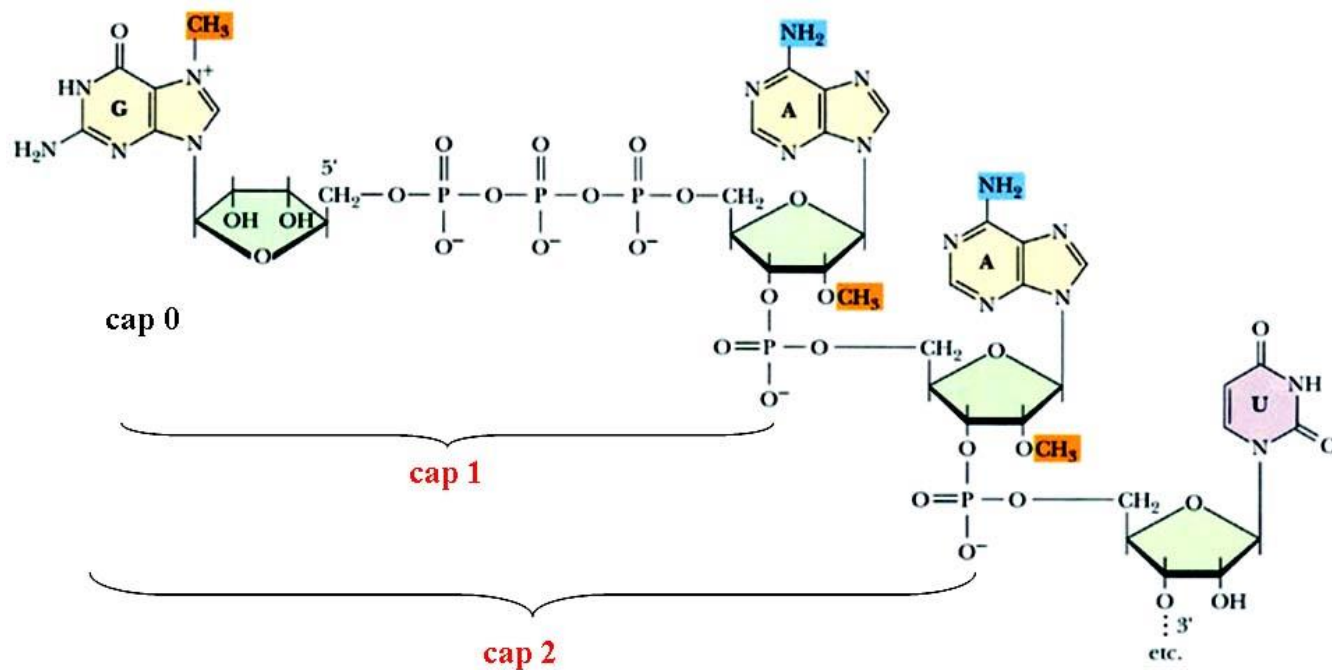
真核生物的转录终止和新生RNA链的加工



- 5'端加帽子
- 3'端加尾巴
- 内含子(intron)的剪除和外显子(exon)的拼接

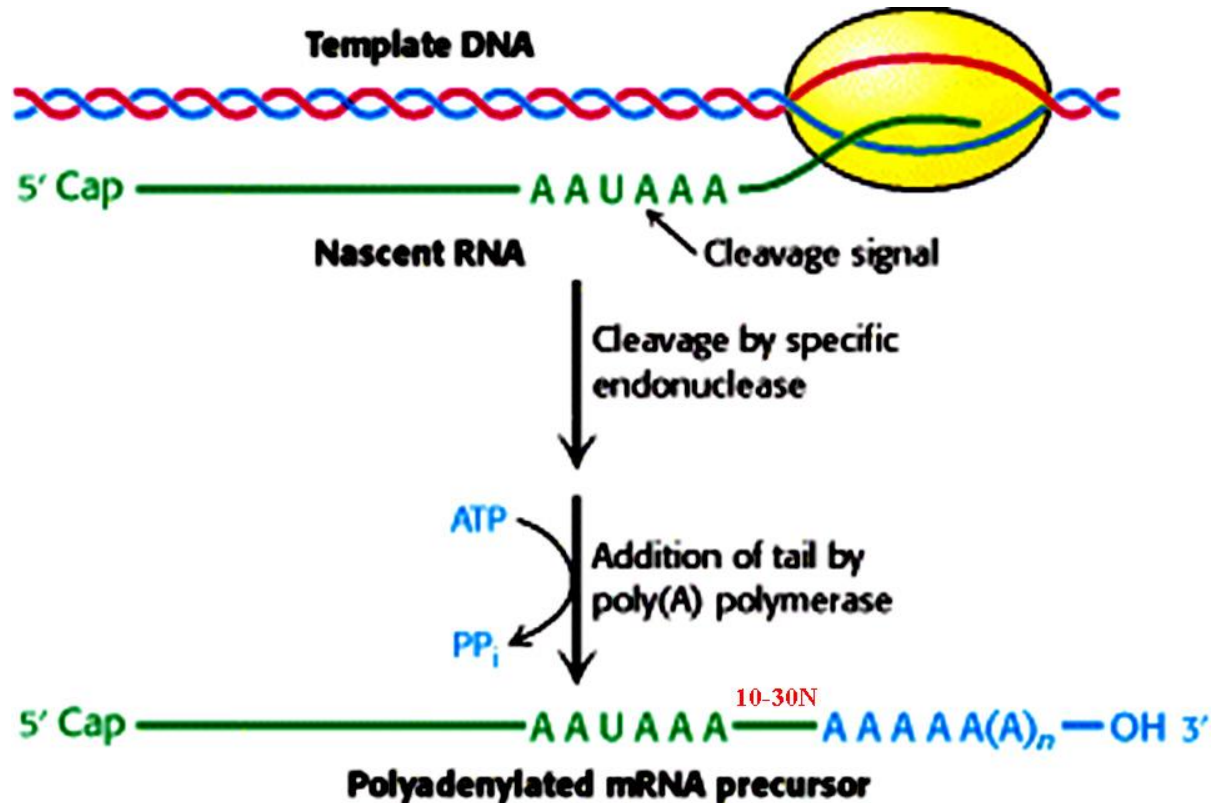
5' 加帽

- 在转录完成前就已经进行加帽;
- 结构: **m7GpppN(m)pN(m)pN--**
- 作用: 能与蛋白结合, 推测在翻译起始中参与mRNA与核糖体的结合启动翻译, 以及对mRNA 起稳定作用。
- 过程:
 - 5'端三磷酸与GTP通过5',5'-三磷酸酯键缩合形成5'帽子;
 - 鸟嘌呤N-7甲基化
 - 前两个核糖核苷酸的戊糖2号碳原子有可能被甲基化。

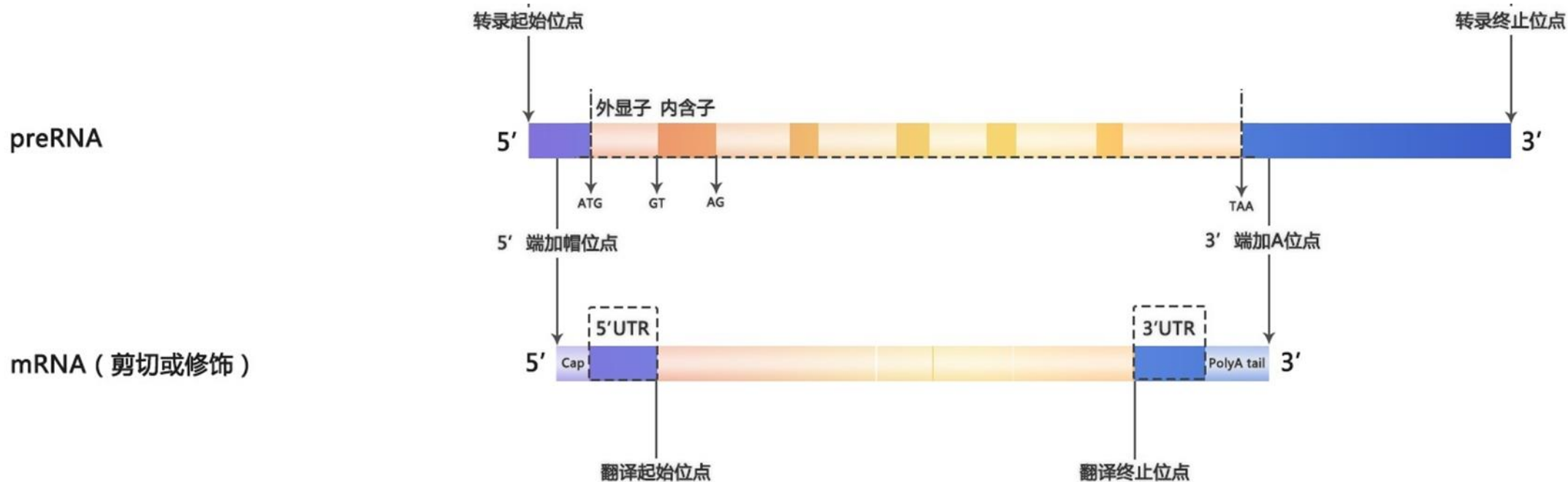


3' 加尾

- 过程：在初级转录产物3'端的特定位点(polyA site, PAS) 识别并对转录产物进行切割， polyA聚合酶直接在切口后加上20~200个腺嘌呤核糖核苷酸。
- PAS位点：人类为AAUAAA； 序列位于切割位点上游10-30个核苷酸处
- 作用：防止外切酶切割损伤转录产物



剪接



- 真核生物编码基因多为**断裂基因**，编码蛋白的DNA序列(外显子, exon)被若干段长度不一的内含子(intron)分隔成多段。不同生物通过多种方式在初级转录产物中剪除内含子，并将外显子拼接成完整连续的蛋白质编码序列。
- 翻译区域(coding dna sequence, CDS)指成熟mRNA上从起始密码子到终止密码子的一段序列。
- CDS两端的非翻译区域(untranslational region, UTR)含有调控作用。

开放阅读框(open reading frame, ORF)

- CDS是已知的一个基因上确确实实翻译成蛋白质的区段。
- ORF则是指，任意一段序列，只要起于ATG止于终止密码子，都可以叫做ORF。ORF是一种预测，而不是一种已知的翻译区。即随意写下一段DNA序列，只要以三个碱基为单位能找到ATG和终止子，就可以称作ORF，这段ORF甚至可能不是一段真正存在的DNA序列，但是它仍然是ORF。

```
1.  ATG CAA TGG GGA AAT GTT ACC AGG TCC GAA CTT ATT GAG GTA AGA CAG ATT TAA
2.  A TGC AAT GGG GAA ATG TTA CCA GGT CCG AAC TTA TTG AGG TAA GAC AGA TTT AA
3.  AT GCA ATG GGG AAA TGT TAC CAG GTC CGA ACT TAT TGA GGT AAG ACA GAT TTA A
```

上面这段序列采取不同的读码框，得到三种不同的ORF，但在生物体内能真正翻译成蛋白质的只有一种。

1.2 RNA的选择性剪接研究

RNA-seq还可用于研究全基因组水平的RNA选择性剪接，即用不同剪接方式(选择不同剪接位点组合)从一个mRNA前体产生不同的mRNA剪接异构体的过程。

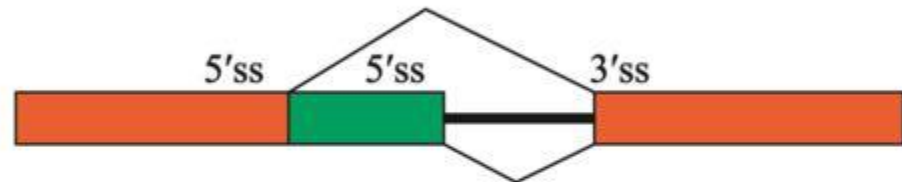
一般将选择性剪接分为如下几类：

- ① 内含子保留(intron retention);
- ② 5'选择性(alternative donor)剪接;
- ③ 3'选择性(alternative acceptor)剪接;
- ④ 外显子遗漏型(exon skipping)剪接;
- ⑤ 相互排斥性剪接。

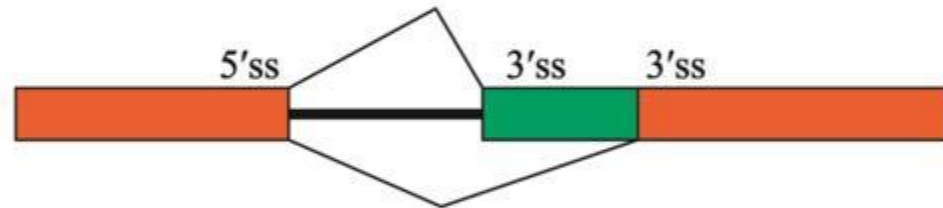
内含子保留



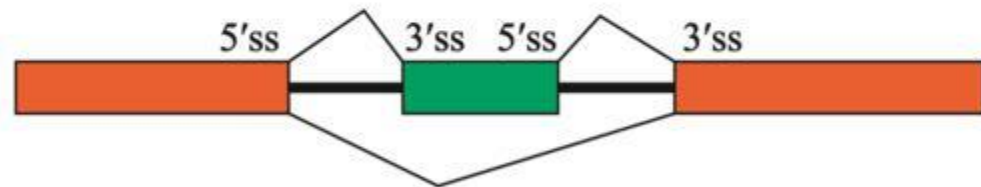
5'选择性剪接



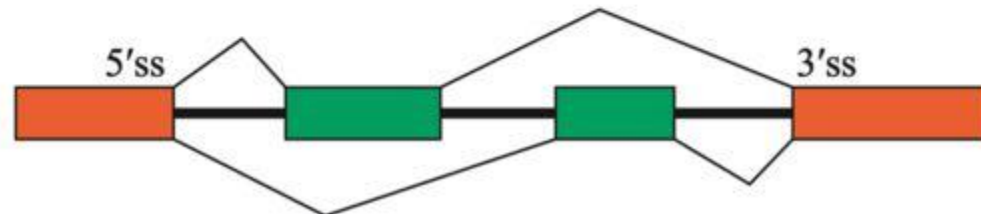
3'选择性剪接



外显子遗漏型剪接

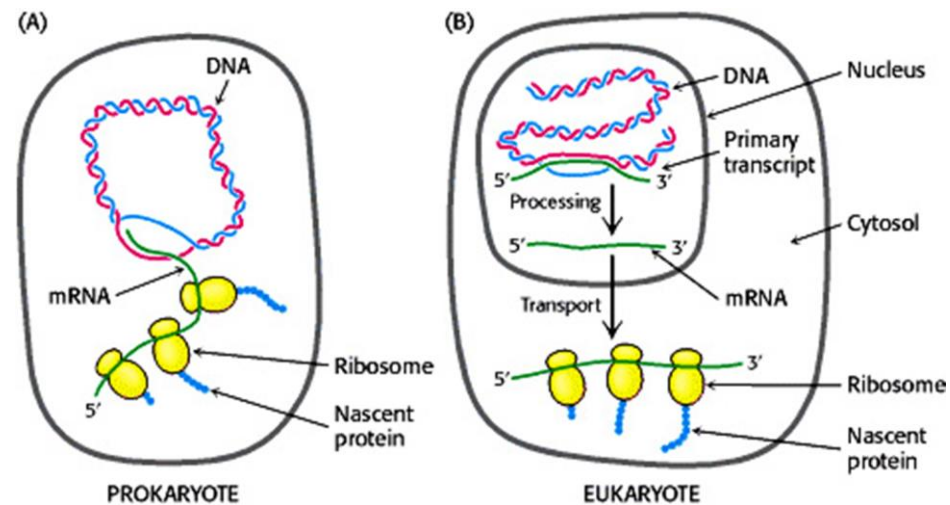
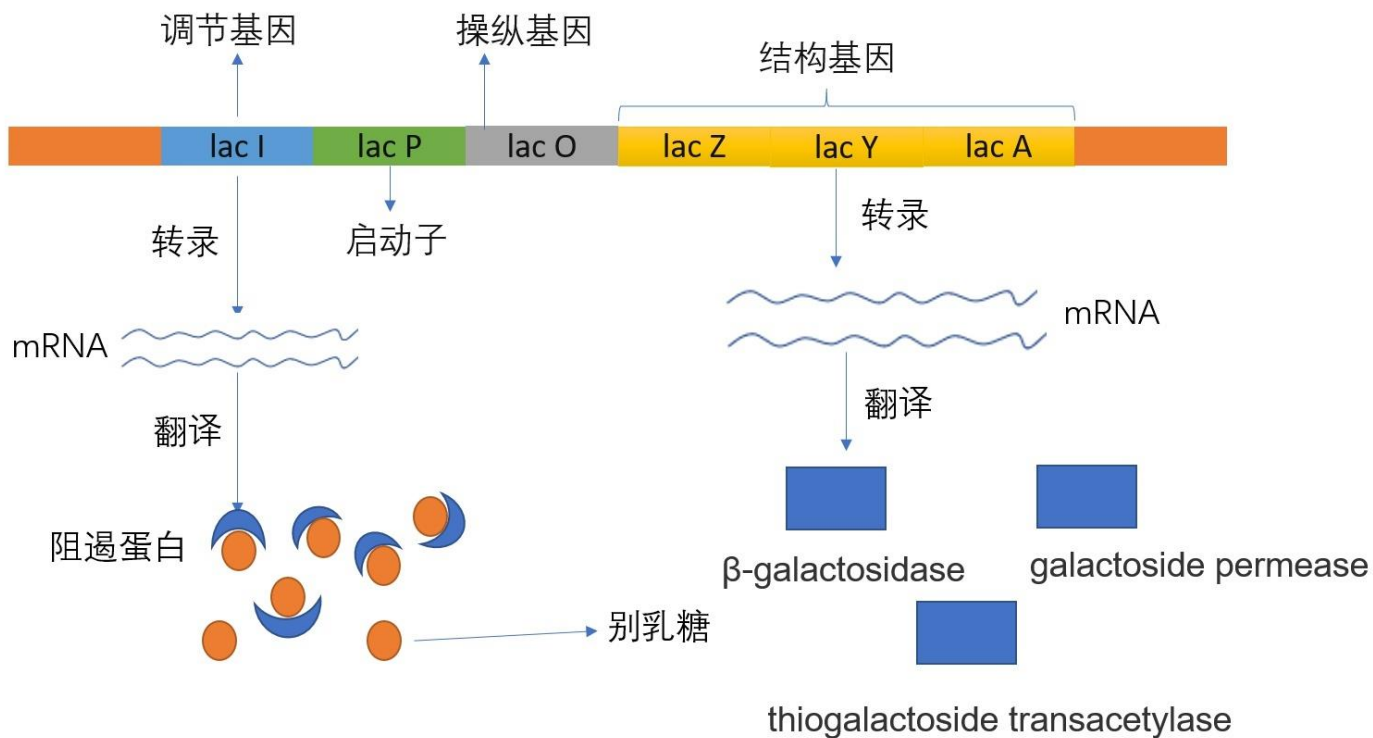


相互排斥性剪接



原核生物转录的特点

- 原核生物中多存在“**多顺反子**”的基因结构，即一个转录单位中有多个结构基因共享一套启动子和终止子，随后翻译成不同的独立蛋白。
- 原核mRNA不存在复杂的加帽和加尾过程，也没有内含子。
- 原核生物的转录和翻译在同一空间进行，无需成熟mRNA的出核步骤，转录未结束时，翻译就已经开始。与之对应的是，其mRNA的半衰期远短于真核mRNA。



非编码RNA

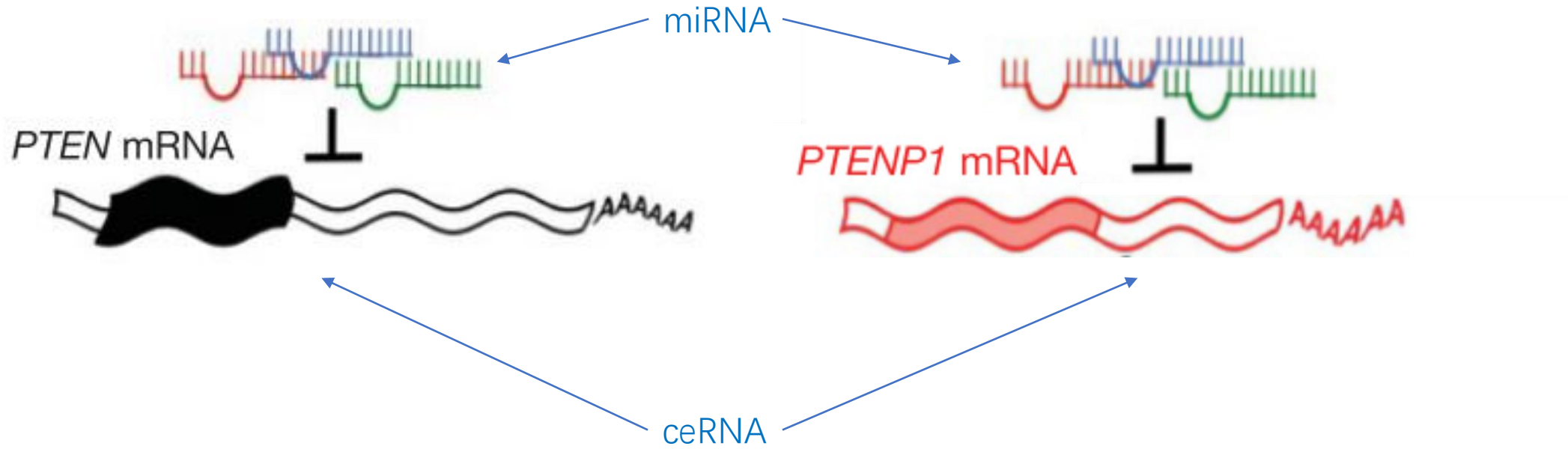
rRNA

tRNA

microRNA、siRNA、snRNA、snoRNA、circRNA、piRNA、lncRNA.....

ceRNA

miRNA & ceRNA (competing endogenous RNA)



- miRNA target sites overlap in 3' UTR