

第一篇

分子生物学技术

人类基因组的结构和功能

FRIEDHELM HILDEBRANDT

核酸结构

“分子生物学”一词系指核酸生物学。分子生物学方法是将自然发生的机制用于体外实验，例如转录和转化。大多数分子生物学技术基于两个基本原理：

- 核酸修饰酶的应用
- 核酸两条互补链之间的杂交（碱基配对）。

在聚合酶的帮助下，DNA修饰酶通常用于特异位点（限制酶消化）切割DNA、重组DNA片段（连接）或扩增和测定DNA序列。杂交方法允许人们用确定的探针通过碱基配对对感兴趣的核酸片段进行特异的选择。为检测相应的杂交片段，利用放射性或非放射性方法标记探针

单倍体人类基因组的大小（即它的复杂度）约为 3×10^9 个核苷酸
图 1.1 表明中期染色体电子显微镜图：用去污剂处理过的仅留下从染色体骨架散落的带有DNA的蛋白支架残迹。这个结构为我们在研究个体核苷酸序列时所论述的大小给了一个印象，就一条完整人染色体的结构而言，它约有 10^8 个核苷酸。下面，从最小结构成分——核苷酸开始讨论人类基因组结构，接着依照增加的复杂度依次描述DNA双螺旋结构、核小体、染色体结构和全套染色体。最后概述人类基因组的一些功能

核苷酸

一个单核苷酸由一分子含氮碱基、一分子糖和一分子磷酸组成

- 含氮碱基（图 1.2A）：嘌呤：腺嘌呤（A）；鸟嘌呤（G）；嘧啶：胞嘧啶（C）；胸腺嘧啶（T）[RNA：尿嘧啶（U）]；罕见的碱基：dA, m⁵C
- 糖：2'-脱氧核糖（DNA），核糖（RNA）
- 一磷酸残基，二磷酸残基，三磷酸残基

Friedhelm Hildebrandt, Universitäts-Kinderklinik, Mathildenstrasse 1, Freiburg, 79106, Germany (phone +49 - 761 - 270 - 4301; fax +49 - 761 - 270 - 4533; e-mail hildebra@kkl200.ukl.uni-freiburg.de)

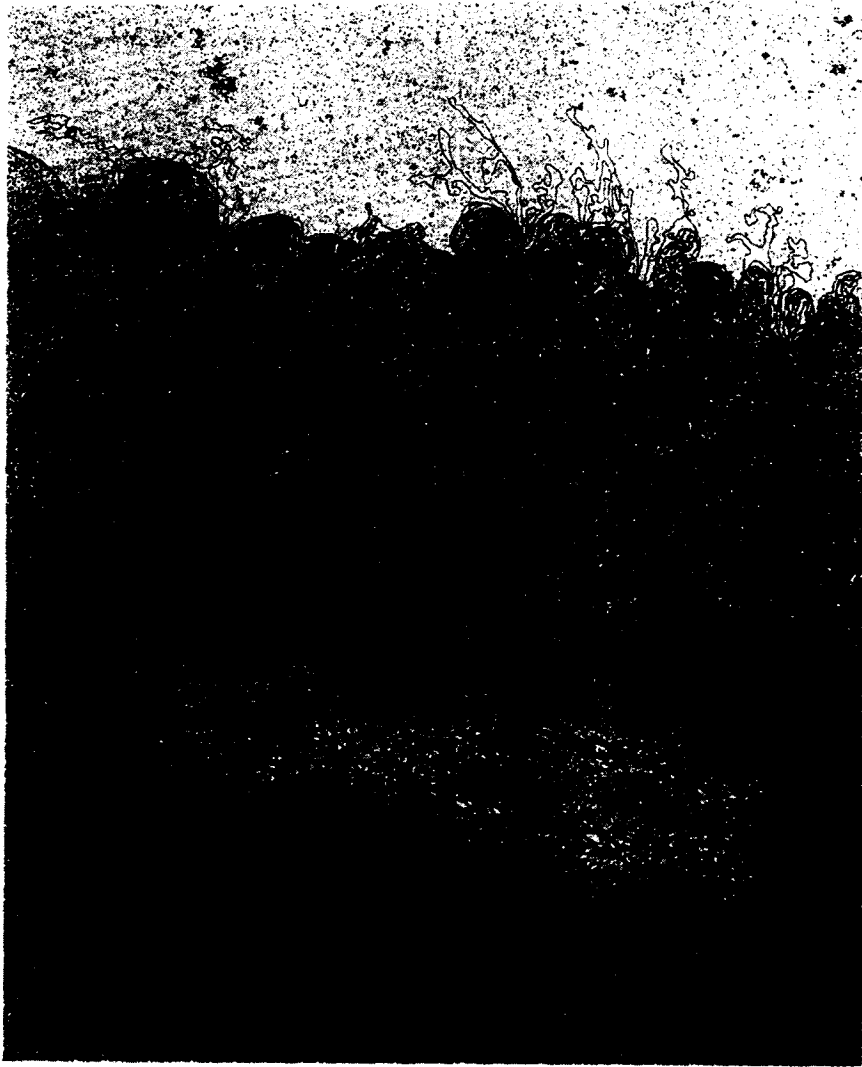


图 1.1 中期染色体的电子显微图，用去污剂处理后仅留下从染色体骨架中散落的带有 DNA 的蛋白支架残迹（从参考文献 1 修改）。

图 1.2B 表明核苷酸成分图解：含氮碱基 + 糖 = 核苷（即，腺嘌呤 + 核糖 = 腺苷）

核苷 + 磷酸 = 核苷酸 [即，腺苷 + 单磷酸 = 单磷酸腺苷 (AMP)，腺苷 + 二磷酸 = 二磷酸腺苷 (ADP)，腺苷 + 三磷酸 = 三磷酸腺苷 (ATP)]。

单核苷酸是构成核酸（DNA、RNA）的基本结构单位。核酸的命名见表 1.1。

表 1.1 核酸的命名

碱基	核苷	核苷酸	缩写	
			RNA	DNA
腺嘌呤	腺苷	腺苷酸	AMP	dAMP
鸟嘌呤	鸟苷	鸟苷酸	GMP	dGMP
胞嘧啶	胞苷	胞苷酸	CMP	dCMP
胸腺嘧啶	胸苷	胸苷酸	dTMP	
尿嘧啶	尿苷	尿苷酸	UMP	

DNA 一级结构

在图 1.2D 中，三磷酸分子与腺苷五碳糖的第 5 位碳原子的羟基缩水成键结合形成的三磷酸残基与另一个核苷酸戊糖环第 3 位碳原子上的羟基相互作用，并与焦磷酸裂解后的磷酸二酯键连接，这个过程产生能量供其反应所需。反应结果表明于图 1.2C。核酸合成从 5'端开始向 3'端进行。如惯例，核苷酸序列从左边 5'端到右边 3'端编写（图 1.2C）。在双链 DNA 中，这个规则用于上排链的编写。

DNA 二级结构

在自然状态下，一个完整的 DNA 分子结构呈双链。Waston 和 Crick 于 1953 年提出的 DNA 二级结构为右手双螺旋模型，由两股反向平行的脱氧多核苷酸链组成，即两股链走向相反，一股链中的磷酸二酯键是 5'→3'走向，而另一股链中的则为 3'→5'走向（图 1.3A）。核糖与磷酸二酯键连接构成 DNA 分子的骨架，这个结构类似 2 个互相盘旋上升的楼梯。两股长链上的碱基是互补配对的。C 和 G 通过 3 个氢键连接，而 A 和 T 之间仅通过 2 个氢键连接（图 1.3A）。因为核糖和碱基之间的糖苷键实际上并不是平行排列的，所以双螺旋显示出一个大沟和一个小沟（图 1.3B）。DNA 结合蛋白，如限制酶和转录因子，主要通过与大沟相互作用识别特异的碱基（图 1.3C）。

双螺旋在水溶液中最常见的构象称 B-DNA（图 1.3D）。每转角有 10 个碱基对。因为碱基的糖苷键能够自由旋转，所以双螺旋会发生构象变化。在 A 型结构中大沟是短的，在罕见的 Z 型结构中，双螺旋转换为左手螺旋（图 1.3D）。DNA 的构象依赖于它的水合作用状态和核苷酸序列。不同的构象状态最可能有调节功能。扫描隧道电镜可直接观察 DNA 双螺旋（图 1.3E）。

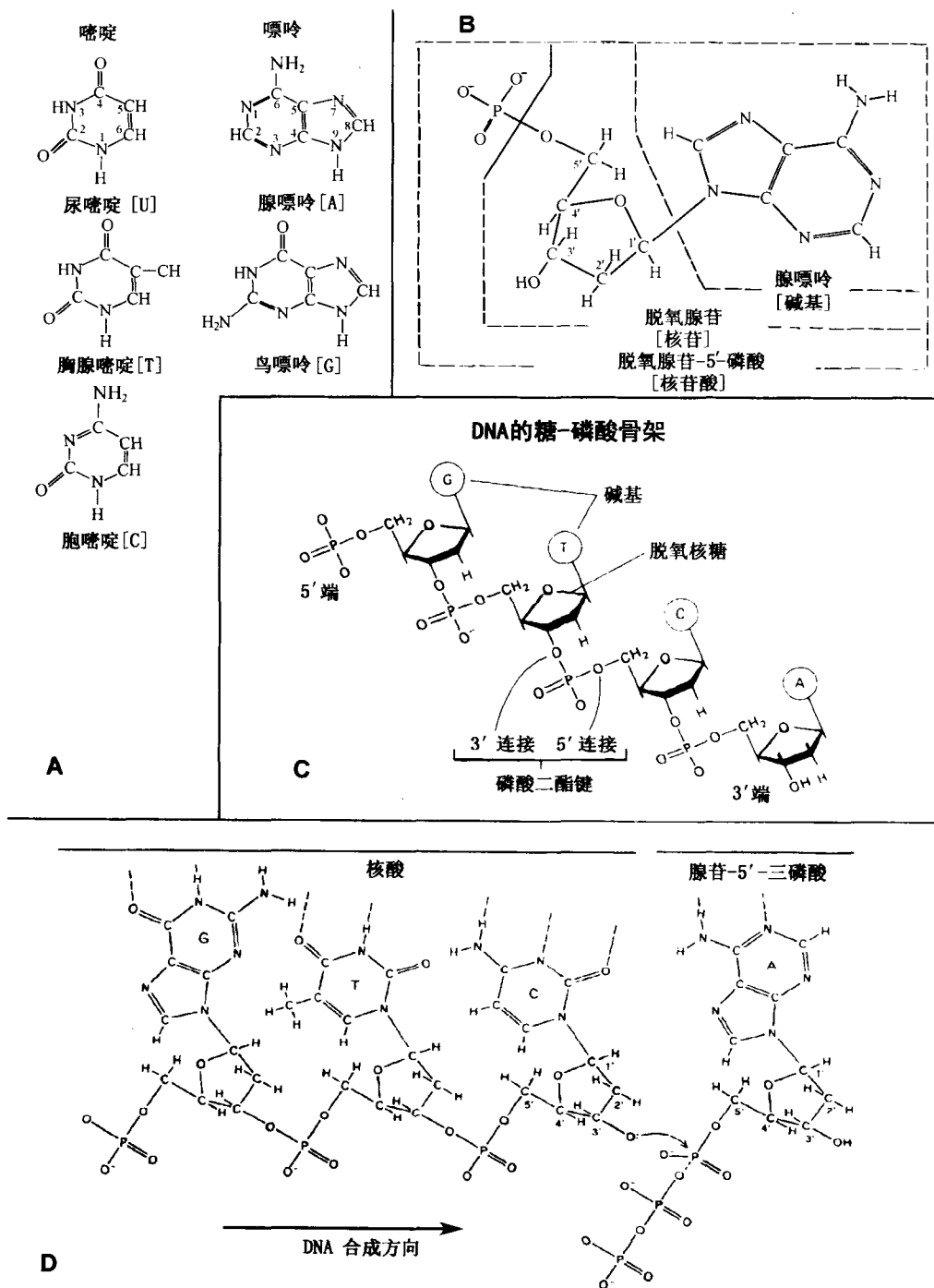


图 1.2 DNA 一级结构。详见原文 (从参考文献 1, 2 和 4 修改)。

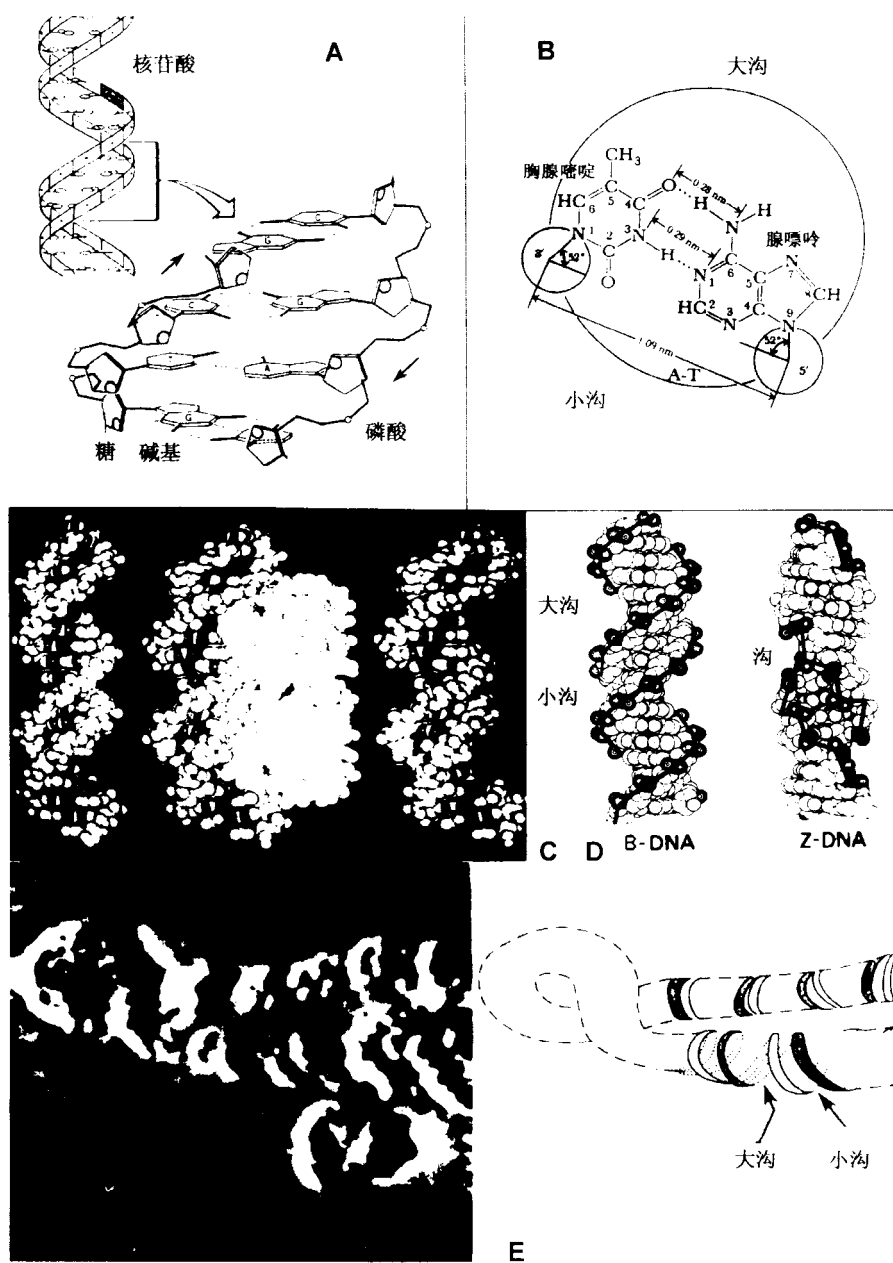


图 1.3 DNA 二级结构详见原文（从参考文献 2、3、6 和 7 修改）

DNA 三级结构

如果一段双链 DNA 是未缠绕的，则超螺旋的三级结构即将出现（在普通实验中，一个过度弯曲的电话线作为超螺旋的形象例子）。图 1.4A 中，以细菌质粒的松环状双链 DNA 为例，说明负超螺旋结构。如

果质粒 DNA 被切割，导致的线状 DNA 是未缠绕的，然后两链端共价再连接，结果形成能量上最优构象：右手超螺旋或左手螺旋。大多自然状态下的 DNA 是以这种负超螺旋形式存在。如果仅单链被切割，那么负超螺旋消失并再形成松环状 DNA。这些结构的电镜图表明于图 1.4B。

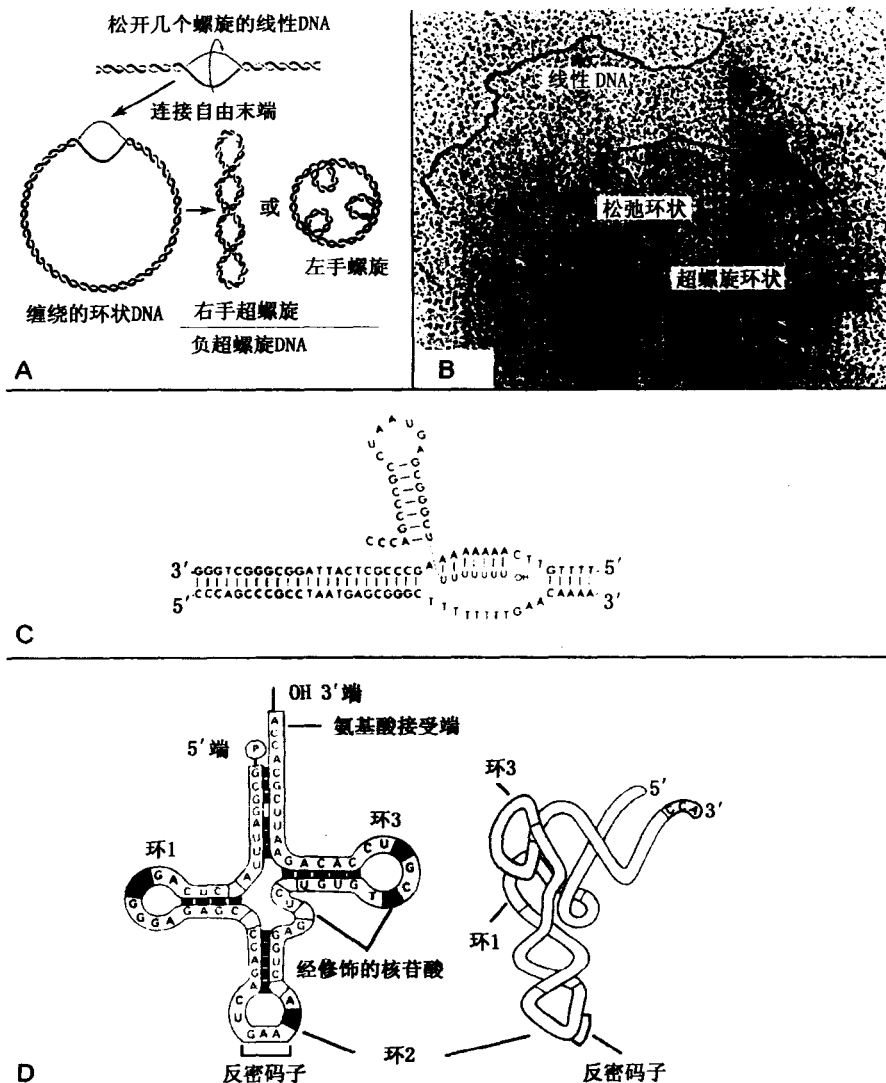


图 1.4 DNA 三级结构和转移 RNA 结构。详见原文（从参考文献 1, 2, 3, 5 修改）。

RNA 结构

RNA 与 DNA 之间最重要的结构差异强调于表 1.2。在相同的单链核酸分子内经碱基配对 RNA（和单链 DNA）能自动形成能量上较稳定的二级结构。图 1.4C 表明 RNA 链呈“发夹”式结构的一个例子，这种结构可能现于显示二重对称的核苷酸序列中。如果一个序列跟随着逆互补

(例如：AACGG-CCGTT) 那么该序列的二重对称则存在。“发夹结构”可作为原核细胞的转录终止子而发挥作用 (图 1.4C)。转移 RNA (tRNA) 的二级结构广泛呈 L 状分子 (图 1.4D)。

表 1.2 DNA 与 RNA 一些特征的差异

	DNA	RNA
最丰富的自然发生	双链	单链
戊糖	2'-脱氧核糖	核糖
核苷酸	脱氧腺苷酸	腺苷酸
	脱氧胞苷酸	胞苷酸
	脱氧鸟苷酸	鸟苷酸
	脱氧胸苷酸	尿苷酸 (罕见)
化学稳定性	室温下相当稳定	室温下相当不稳定 (可被 RNA 酶和碱水解)

染色质组织

染色质结构

染色体 DNA 与组蛋白有关。每两个分子组蛋白 H2A, H2B, H3 和 H4 (图 1.5B) 形成一个多聚体蛋白复合体, 称核小体 (图 1.5A)。双链 DNA 围绕核小体外围转 2 圈, 相当于 146 个碱基对的长度。电镜下, 连续串排的核小体酷似所谓“串珠”或“10 nm 纤丝” (图 1.5B, C) 在高盐浓度下, 10 nm 纤丝凝集形成“30 nm 纤丝” (图 1.5D, E)。组蛋白 H1 分子组织染色质形成螺线管 (solenoid), 它是以花束样结构排列的 6 个核小体的聚合体 (图 1.5E, F)。

染色体

有丝分裂期间, 在非组蛋白影响下, 发生染色质高度凝聚并且导致中期染色体 (metaphase chromosome) 形成 (图 1.6A)。细胞遗传学技术可使代表染色体上富含 AT 区域 (黑) 或富含 CG 的区域 (亮) 的特异染色体带型着色。图 1.6B 中, 每条人染色体的左半表示中期染色体染色, 而在右半则表示较高分辨度, 因它在前中期可见。虽然这种带型是种属特异的, 但在整个进化过程是高度保守的。

图 1.1 表示染色体的电镜图谱 (用去污剂处理过)。虽然长环 DNA 从支架流出, 但蛋白支架的残迹仍可见。

人类基因组的大小

- 人类染色体为二倍体, 46 条, 性染色体 XY (即 22 对均一常染色体:

- 22条来自父亲，22条来自母亲，加上2条性染色体：XY或XX)。
- 单倍体人类基因组的长度是 3×10^9 碱基对 (bp)。
 - 人类基因组中约有 8 万个表达基因。
 - 图 1.7 表明用碱基对标示人类基因组 DNA 结构的大小。为不同大小的结构实体指出研究其分子结构的适合的方法学。

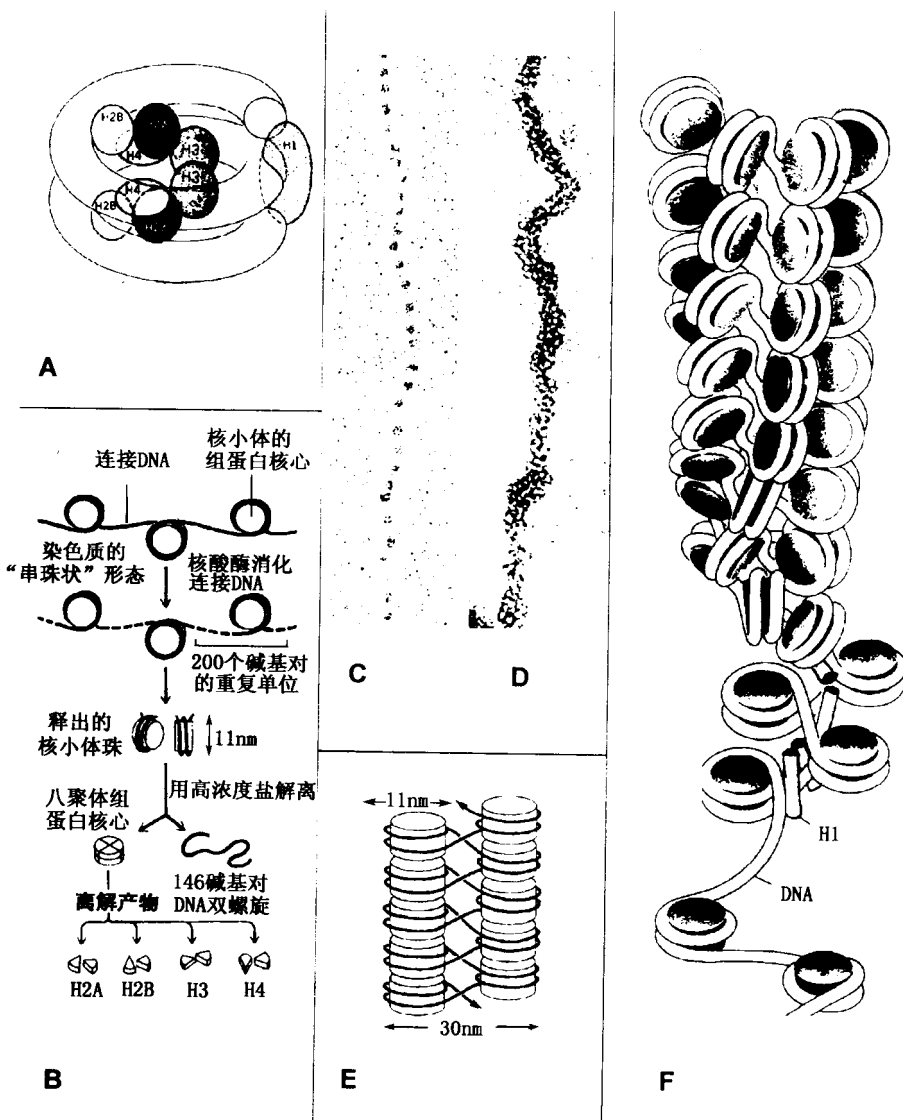


图 1.5 染色质结构。详见原文（从参考文献 1, 3 和 4 修改）。

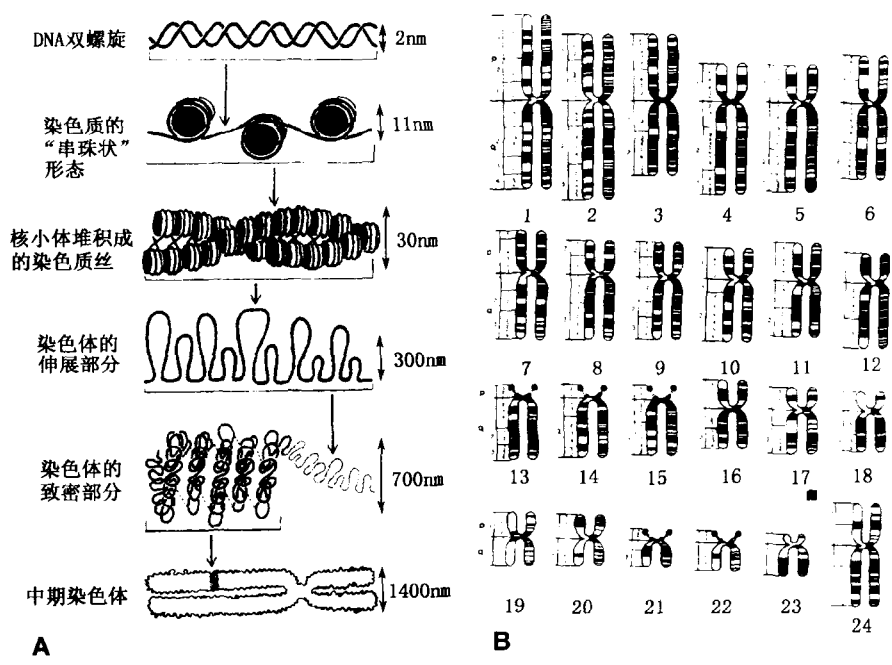


图 1.6 染色体结构。详见原文（从参考文献 1 修改）

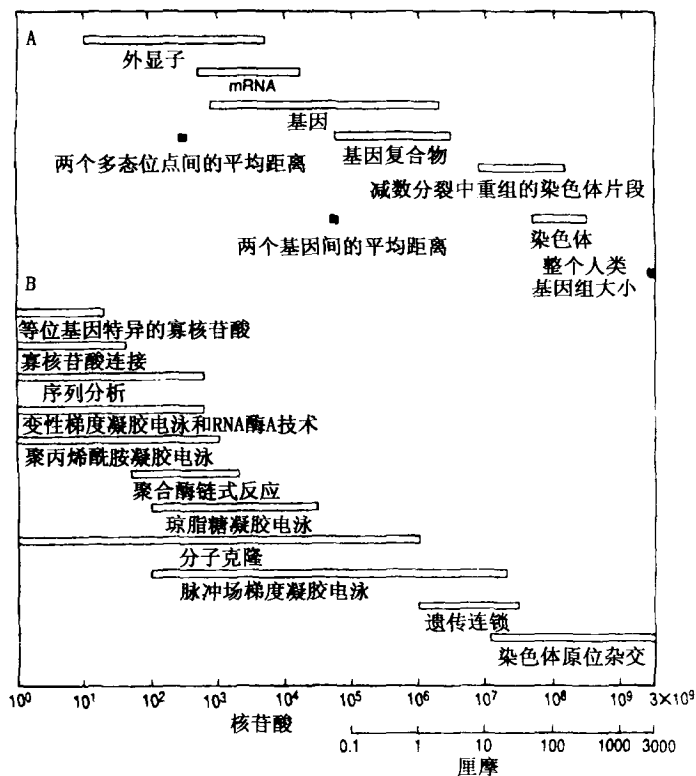


图 1.7 诊断技术和人类基因组。(A) 人类基因组的信息单位大小范围：核苷酸或厘米。(B) 各种诊断技术相应的应用范围（从参考文献 7 修改）。

人类基因组的功能

复制

有丝分裂期间的细胞分裂和减数分裂期间的配子发生均要求单倍体染色体优先复制（加倍）。所谓二价染色体有 46 条，性染色体 XY。每个二价体由 2 个姐妹染色单体在着丝粒处连接组成。复制发生于细胞周期的 S 期。

如下原理适用于基因组 DNA 的复制（图 1.8A）：

- 基因组 DNA 的复制为半保留复制，即各新形成的双链 DNA 由一条新合成子代链和一条原亲代链组成，原亲代链在复制过程充当碱基配对的模板。

当含有 DNA 聚合酶（来自粒细胞的 DNA-Pol. III）的多酶复合物结合到双链 DNA 分子上并拆开双链时，复制开始。DNA 聚合酶（图 1.8B）利用原来的亲代链的核苷酸序列作为合成互补链的模板，子代链经顺序添加与亲代链互补的三磷酸脱氧核糖核苷（dNTP）而延长。发现磷酸二酯键位于 dNTP 的 5'-磷酸基末端与新合成的子代链的 3'-羟基末端之间。聚合酶总是要求单核苷酸引物的存在，以提供 3'-羟基末端组并开始聚合作用。聚合酶仅能以 5'→3' 方向进行合成新链，因为两个模板链呈逆方向的。因此，复制导致在一条模板链上的前导链（leading strand）和另一条链上的后随链（lagging strand）（图 1.8C）。因为复制在两条链上以相同方向进行，但子代链对后随链的聚合作用如在前导链是从 5'→3' 方向进行，故后随链不得不落后一段时间。短 RNA 引物作为结合后随链的聚合酶的引物（图 1.8C）。从这些引物开始聚合所谓的冈崎片段。为最终连接单个冈崎片段，在展示有外切核酸酶活性的 DNA 聚合酶的作用下，DNA 替换了 RNA 引物。最后，连结酶连接冈崎片段。

- 复制导致一个移动的复制叉（图 1.8C），如电镜所证明的（图 1.8D）。
- 复制错误率约为 $1/10^8 \sim 1/10^{12}$ 个核苷酸。这个相对低的错误率是由于半保留复制和 DNA 修复酶类的存在。

图 1.9A 表明真核细胞的代表——酵母菌的周期。细胞周期的进展受所谓“细胞分裂周期基因”（cell division cycle genes）调节。例如，图 1.9B 表明在有丝分裂中期，直到成熟促进因子（MPF）的 **cdc2** 激酶成分被激活，细胞周期蛋白浓度才增加。然后当细胞分裂时，细胞周期蛋白降解并且 MPF 活性降低，最后细胞周期重复。

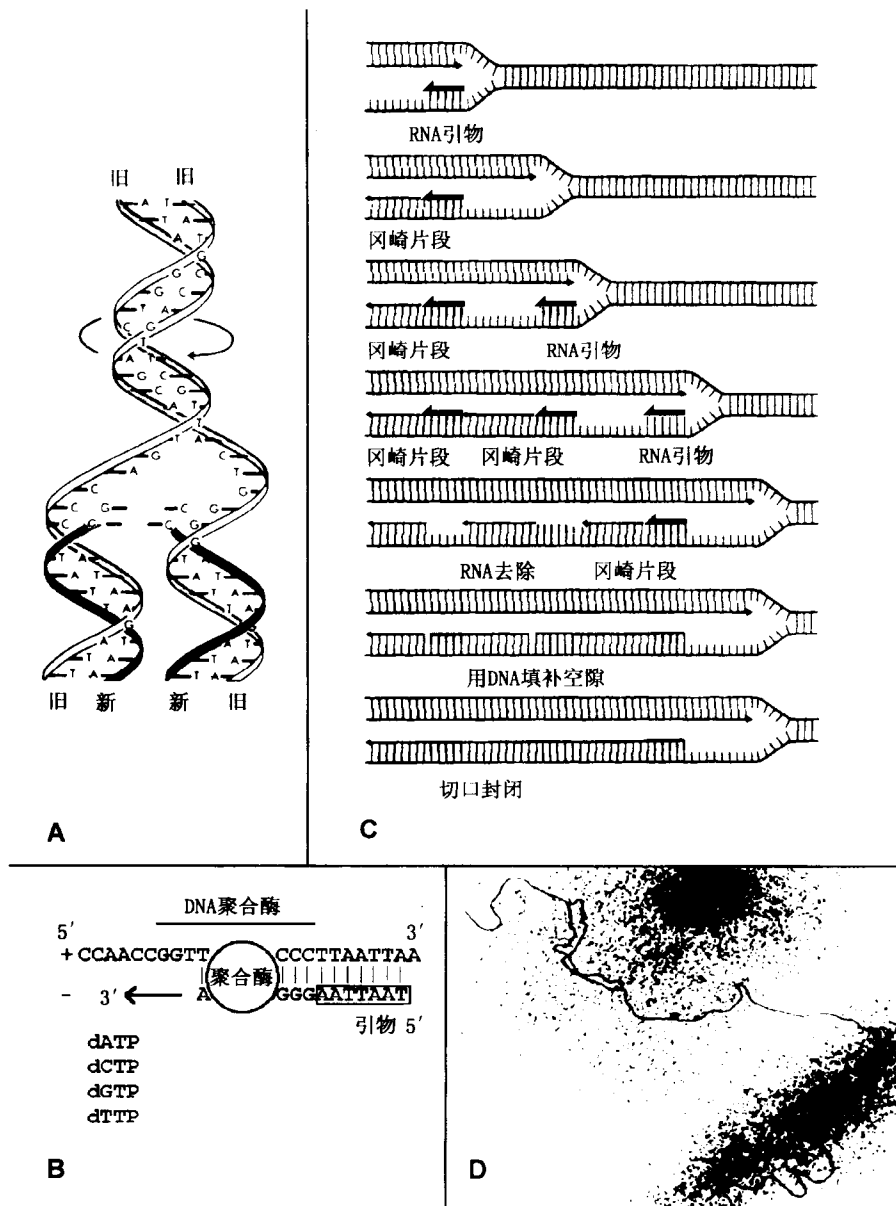
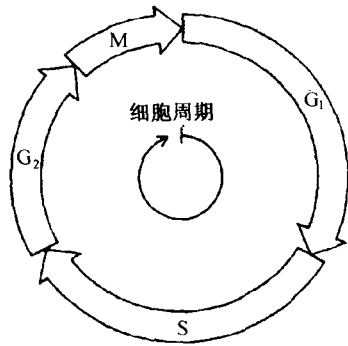


图 1.8 DNA 复制机制。详见原文（从参考文献 3 和 5 修改）

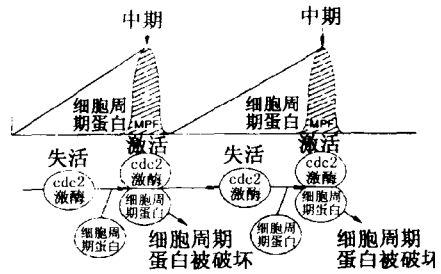
转录和翻译

DNA 复制是细胞分裂的前提并允许在生物体中生长。基因组 DNA 转录成信使 RNA (mRNA)，接着 mRNA 翻译成蛋白质，随后遗传信息和细胞代谢之间的连接。遗传信息从基因组 DNA 到蛋白质流动顺序已被 Francis Crick (1956) 在分子生物学的中心法则中详细说明 (图 1.9C，亦见图 1.10 和图 1.11B)。



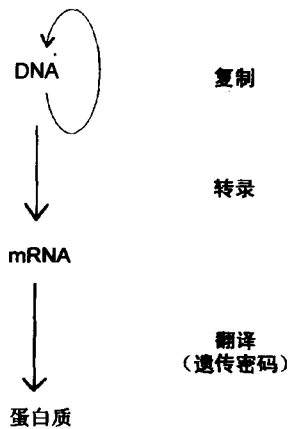
A

一个酵母细胞生活史中的时相，
M= 有丝分裂，G₁=DNA 合成前，
S=DNA 合成，G₂介于 DNA 合
成与有丝分裂之间的时期

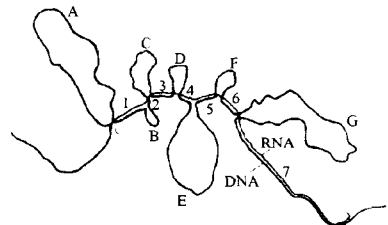


B

分子生物学的中心法则



C



D

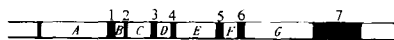


图 1.9 转录机制。详见原文 (从参考文献 4, 5 修改)。

转录

图 1.10 表明，转录从启动子开始。启动子是一段特异的 DNA 序列，它可被含 RNA 聚合酶的多酶复合体和转录因子所识别。然后，RNA 聚合酶催化 DNA 转录为信使 RNA (mRNA)。如惯例，与 mRNA 序列相同的 DNA 链称为 (+) 链或有义链，而互补链称 (-) 链或反义链。

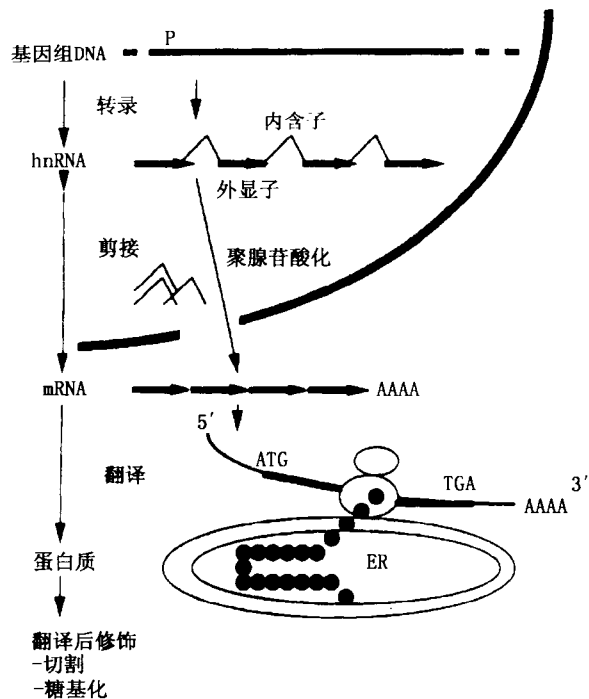


图 1.10 基因和基因产物之间的共线性方面。详见原文，P = 启动子，箭号表示外显子，钩状表示内含子。

转录最初导致未成熟异核 RNA (hnRNA) 产生。异核 RNA 含有编码氨基酸序列的外显子 (exons) 且其存在于成熟 mRNA 中。内含子 (introns) (或间插序列) 不能编码蛋白序列且缺乏成熟 mRNA。剪接体在剪接供体和剪接受体识别序列剪接内含子，临近的外显子相互连接。异核 RNA 离开核进入细胞质前，在 poly (A) 聚合酶作用下，poly (A) 尾被加到 3' 端。7-甲基鸟嘌呤保护 5' 端，构成所谓 5' 帽。这两个机制导致成熟 mRNA 高度稳定。然后成熟 mRNA 在核糖体上利用遗传密码翻译成多肽。真核细胞中的核糖体停靠在内膜网上。在一些例子中，蛋白质翻译后，接着进行修饰，例如信号肽的切割或糖基化。从转录到基因产物 (蛋白质) 的全过程称基因表达。

在原核细胞中，基因、mRNA 和多肽之间呈共线性关系。在真核细胞中，这种共线性关系由于内含子的插入而被打断。此外，这种共线性的规律可由于另一种的剪接和翻译后的修饰而打破。在电镜上应用所谓“R 环”的条件，可观察到外显子和内含子的界线 (图 1.9D)

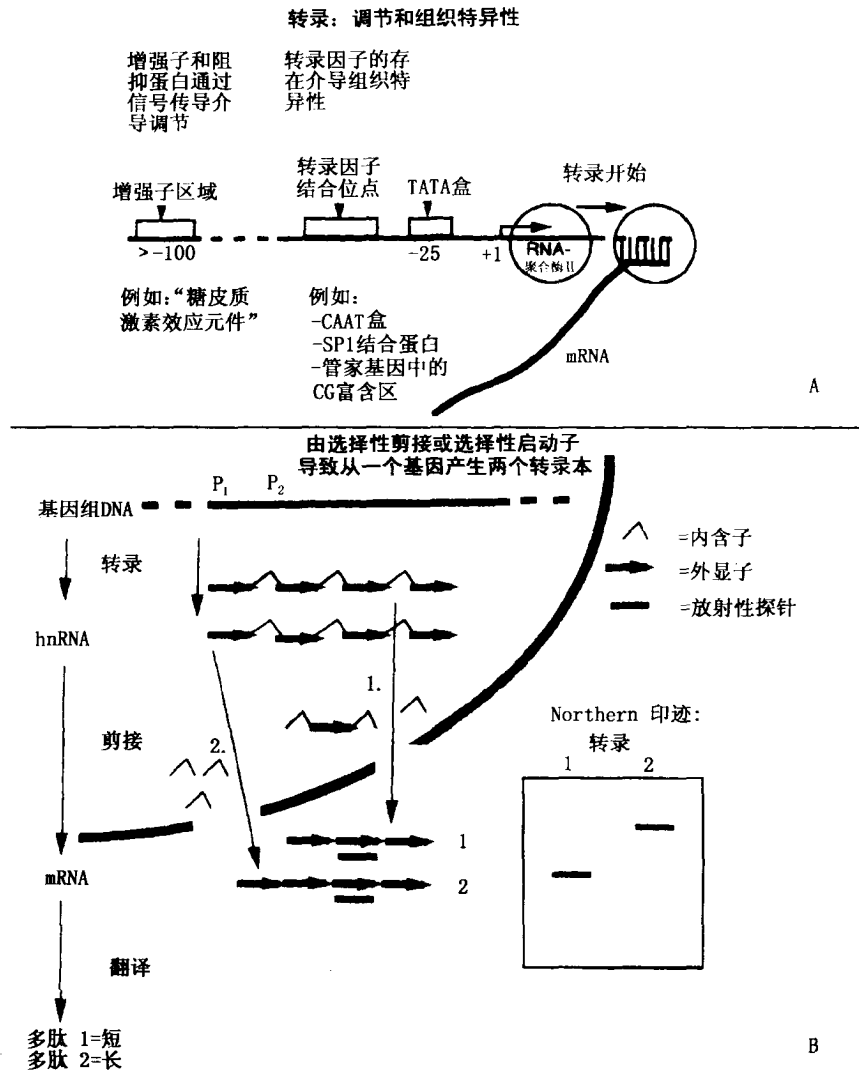


图 1.11 转录调节。详见原文。

转录的调节

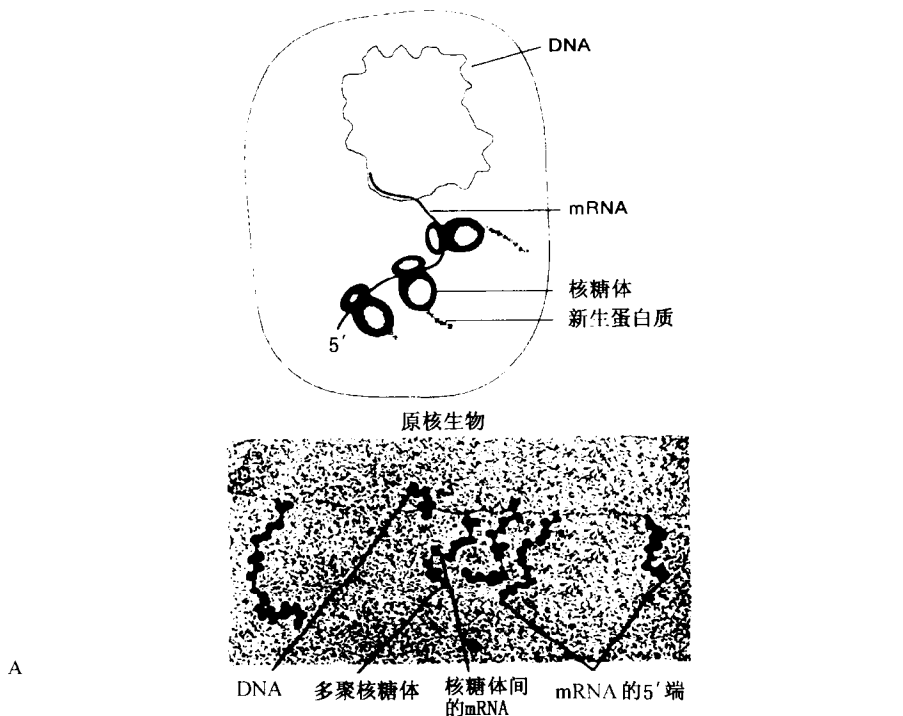
转录受转录因子（transcription factor）高度调节。转录因子是识别特异的调节 DNA 序列的蛋白分子，在大多数例子中，称增强子（enhancer）或抑制子（silencer）的那些调节序列位于转录开始位点的上游，转录开始位点在所谓 5'侧翼区。细胞中的这些蛋白因子的存在和缺乏，导致组织特异的基因表达并允许细胞以组织特异形式起作用。此外，这些因子赋予细胞有改变基因表达能力，以适应对外源信号的应答。糖皮质激素受体作为一个实例：结合类固醇激素后，细胞质受体转位到细胞核并结合糖皮质激素的效应元件（GRE）。GRE 对邻近基因具有增强子的功能。

相同基因的不同转录物

通过应用可变启动子（P₁ 和 P₂）以及可变剪接（可变剪接 1 和 2，见图 1.11B），基因产物的多样化的经济学意义在自然界中得以充分表达。

翻译和遗传密码

翻译（蛋白质生物合成）（图 1.12）在核糖体上进行（图 1.12A）
翻译期间，核糖体 A 位由含有反密码子的 tRNA 占据，其与 mRNA 上的



翻译的延伸步骤

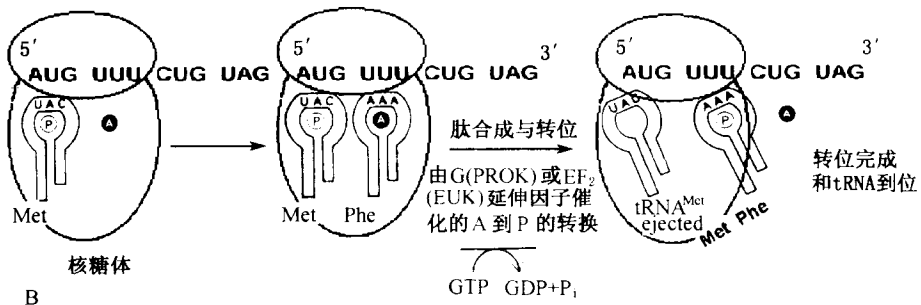


图 1.12 翻译机制。详见原文（从参考文献 2 修改）。