

医用分子遗传学

主编 童普德

上海中医药大学出版社

医用分子遗传学

主 编	童普德		
编 委	王海南	王晓玲	刘成海
	吴小江	陆振虞	陆 雄
	周建锋	官 斌	姚明忠
	胡旭东	莫启忠	秦红友
	钱汝红		
主 审	丁镛发		

上海中医药大学出版社

图书在版编目 (CIP) 数据

医用分子遗传学/主编童普德. —上海: 上海中医药大学出版社, 2001.8

ISBN 7-81010-563-9

I. 医... II. 童... III. 分子遗传学—研究生—教材 IV. Q75

中国版本图书馆 CIP 数据核字 (2001) 第 25442 号

责任编辑 / 钱静庄

技术编辑 / 宓国华

责任校对 / 葛兴棣

封面设计 / 王 磊

出 版 / 上海中医药大学出版社

(200032) 中国上海零陵路 530 号

发 行 / 新华书店上海发行所

印 刷 / 江苏省丹阳市教育印刷厂

版 次 / 2001 年 8 月第 1 版

印 次 / 2001 年 8 月第 1 次印刷

开 本 / 787 × 1092 1/16

字 数 / 359 千

印 张 / 14.75

印 数 / 1-3 000 册

ISBN 7-81010-563-9

R·534 定价: 29.50 元

编写说明

我校 1988 年起为研究生开设“医学分子遗传学”课程。10 多年来，随着分子生物学的长足进展、人类基因组计划的提出与实施以及对后基因时代的无限遐想，同学们学习热情高涨。他们急切需要了解、掌握分子遗传学基本原理、基本方法与基本技术，以启迪思路、培养技能，更好地为中医药事业承前启后、开拓创新。教学实践告诉我们，中医院校研究生在主修中医药专业的基础上，需要博采众长，需要“为我所用”，所以选修科目众多，但学时却有限。因此，编写一本合适的教材成为教学的基本保证。感谢市教委教材基金给我们提供了这样宝贵的机会，也感谢上海第二医科大学陆振虞教授为本教材撰写“基因诊断和基因治疗”这一重要章节。特别要提及的是我校丁镛发教授冒着盛夏酷暑审读并指导了本书的编写，在此深表谢意。

本教材共 12 章：一至七章为基础理论，八至十一章为专题内容，十二章为教学实验。全书编写分工如下，第一章一至二节：童普德，三至四节：王晓玲；第二至三章：童普德；第四章：童普德、王晓玲；第五章：胡旭东；第六章：姚明忠、童普德；第七章：童普德、吴小江；第八章：陆振虞；第九章：刘成海、周建锋；第十章：刘成海、王海南；第十一章：宫斌、莫启忠、秦红友；第十二章：童普德、吴小江、姚明忠、童普德、钱汝红、陆雄。

编 者

2000 年 12 月

目 录

第一章 基因—DNA—染色体	(1)
第一节 基因概念	(1)
第二节 基因结构	(8)
第三节 基因组	(14)
第四节 人类基因组计划	(20)
第二章 DNA 的生物合成	(23)
第一节 DNA 半保留复制的特点及复制体构成	(23)
第二节 其他形式的复制	(39)
第三章 RNA 的生物合成	(45)
第一节 转录概述	(45)
第二节 RNA 聚合酶	(47)
第三节 DNA 转录过程	(49)
第四章 转录后 RNA 的加工修饰	(59)
第一节 mRNA 前体 (pre-mRNA) 的加工	(59)
第二节 rRNA 和 tRNA 的加工	(70)
第五章 蛋白质的生物合成	(74)
第一节 参与蛋白质生物合成的主要生物大分子	(74)
第二节 蛋白质在核糖体上的翻译过程	(82)
第三节 翻译后的加工、修饰及分泌	(90)
第四节 翻译水平的调控	(98)
第六章 基因表达的调控	(100)
第一节 原核生物基因表达的调控	(100)
第二节 真核生物基因表达的调控	(105)
第七章 基因工程	(113)
第一节 目的基因的获取	(113)
第二节 基因工程工具酶	(115)
第三节 常用载体	(120)
第四节 目的基因与载体的连接及重组体导入细胞	(126)
第五节 阳性重组体的筛选与鉴定	(131)
第六节 聚合酶链式反应 (Polymerase Chain Reaction, PCR) 技术	(132)
第八章 基因诊断和基因治疗	(137)
第一节 基因诊断	(137)
第二节 基因治疗	(145)
第九章 癌基因与抑癌基因	(161)

第一节 肿瘤的发生机制	(161)
第二节 肿瘤的基因治疗	(171)
第十章 细胞信号转导	(177)
第一节 信号、受体与关键酶	(177)
第二节 重要的细胞跨膜信号转导通路	(185)
第三节 信号转导异常的常见疾病	(189)
第十一章 转基因动物及其应用	(191)
第一节 目的基因的克隆和分离	(191)
第二节 目的基因的结构分析	(195)
第三节 转基因动物的制备	(196)
第四节 转基因动物在医学科学研究中的应用	(200)
第十二章 教学实验	(205)
【核酸提取技术】	
实验一 真核细胞 DNA 提取	(205)
实验二 真核细胞 RNA 提取（一步法）	(206)
实验三 紫外分光光度法定量抽提物	(207)
【电泳技术】	
实验四 酯酶同工酶的聚丙烯酰胺凝胶电泳分离	(208)
实验五 琼脂糖凝胶电泳分离 DNA 片段	(210)
【聚合酶链式反应（PCR）技术】	
实验六 PCR 性别鉴定	(212)
实验七 RT-PCR 技术	(214)
【杂交技术】	
实验八 Northern 印迹杂交	(216)
实验九 非放射性标记探针原位杂交检测石蜡切片中的 mRNA	(219)
实验十 蛋白印迹杂交技术（Western blot）	(220)
【DNA 损伤测定】	
实验十一 非程序 DNA 合成（UDS）的测定	(222)
实验十二 凋亡细胞梯状条带电泳鉴定	(225)

第一章 基因—DNA—染色体

现今，核酸是遗传物质虽然早已成了不争的事实，但是科学永远为探索者提供着新的舞台。人们过去往往关注核酸分子上的编码基因，对非编码序列却研究不够。现在看来，核酸中大部分序列并不是编码基因，而且就大部分编码基因内部组成而言，其中不能转录成成熟 mRNA 的内含子也属于非编码序列，它们在长度上往往比外显子长，并且可以含有重要信息。例如，据人基因组计划预示，人类基因总数并不是预计的 10 万多个，而仅仅为 3~4 万个左右，95% 以上的序列竟是非编码序列。显然，定义基因是 DNA 分子上一段有生物学效应的片段时，还必须明白基因是 DNA 分子上有一定组织结构的功能片段，因为缺乏启动序列、终止序列及其他非编码序列，仅存在主体序列的结构基因是一事无成的。单个基因如此，基因群乃至整个基因组组成更是如此。因此，从基因重排、拼接、编辑、重组等角度来研究人类基因数目远比蛋白质（数十万种）少的原因，从基因组学角度研究人类基因功能，从蛋白质组学角度研究遗传信息表达与医学的复杂关系是十分必要的。

本章是对基因（基因组）、DNA 和染色体构成的简述。

第一节 基因概念

一、基因是具有生物学效应（编码蛋白质）的一段 DNA 片段

（一）DNA 是遗传物质的实验证据

1. 肺炎球菌的转化现象

肺炎球菌（pneumococcus）中，外膜由粘多糖荚膜包被、菌落光滑而闪光的为 S 型，无荚膜包被、菌落粗糙的为 R 型。粘多糖是水溶性的，有抗原性，能使小鼠因败血症死亡。但当把无毒性的 R 型菌和加热杀死的 S 型菌一同注入小鼠，几天后小鼠致病死亡（图 1-1），且从小鼠血液发现了 S 型菌。Griffith 将无毒性的 R 型菌在加热杀死的 S 型菌作用下，转变为有毒性的 S 型菌称为转化（transformation）。

显然，被加热杀死的 S 型菌中必定有一种物质起转化作用。1944 年 Avery 等不仅重新证实了上述转化现象，并且从 S 型菌的抽提物中确认转化物质是 DNA。他们将 S 型菌的成分分离后，分别去转化 R 型菌，最后发现只有 S 型菌的 DNA 能起转化作用。为了排除抽提物中混有的微量蛋白质起作用的可能性，Avery 等分别用 DNA 酶和蛋白水解酶处理抽提的 DNA，结果经 DNA 酶处理的抽提物不起转化作用，经蛋白酶处理的抽提物转化率不受影响。到 1949 年，用连续提纯的方法将蛋白质污染降至 0.02% 以下，结果是 DNA 纯度越高，转化率就越高，充分证明起转化作用的是 S 型菌的 DNA 而不是蛋白质。后来知道，S 型菌的 DNA 结构在加热中并没遭到破坏，它们在 R 型菌中得以表达而产生 S 型菌的性状。并且知道，DNA 作为遗传物质引起感受态细胞的转化具有一定的普遍性。

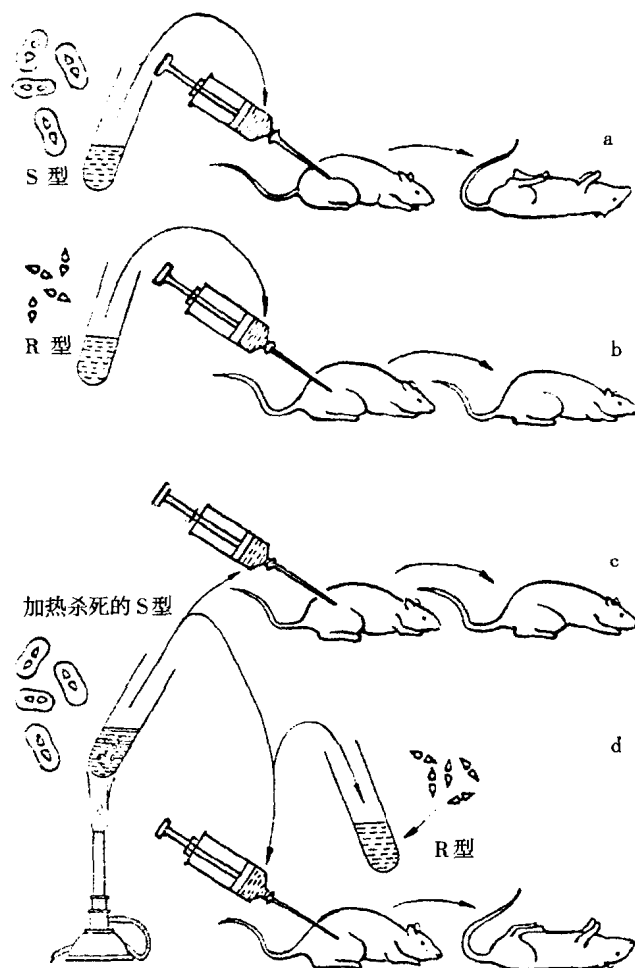


图 1-1 肺炎球菌的转化

a S型肺炎球菌注入小鼠能使小鼠致病死亡。b R型肺炎球菌注入小鼠不能致病。c 加热杀死的S型肺炎球菌注入小鼠不能致病。d R型肺炎球菌和加热杀死的S型肺炎球菌一同注入小鼠能使小鼠致病死亡。

2. T_2 噬菌体的侵染试验

噬菌体结构简单，外壳为蛋白质，内部是核酸。兼有溶源周期和溶菌周期的称为温和噬菌体（图 1-2），只有溶菌周期的则为裂性噬菌体。

1952 年，因 DNA 中尚含有少量蛋白质污染而对转化物质依然有争论时，Hershey 等用实验证明 T_2 一类裂性噬菌体在世代延续中只有 DNA 是连续的，蛋白质是在 DNA 遗传信息控制下合成的。Hershey 等分别用 ^{32}P 和 ^{35}S 标记两部分 T_2 噬菌体， ^{32}P 渗入 DNA， ^{35}S 渗入外壳蛋白。再分别将标记的噬菌体加入液体培养的大肠杆菌中，侵染一定时间后，通过振荡把细菌和附在细菌表面的噬菌体残留部分分开。离心后，细菌沉至管底，未进入细菌的噬菌体残留部分留在上层。分别测定底部细菌和上层培养液的放射性后发现：在用 ^{32}P 标记噬菌体时，放射性出现在底层细菌而不在上层培养液内；在用 ^{35}S 标记噬菌体时，放射性

出现在上层培养液中而不在底层细菌中，说明侵染细菌时进入寄主的只是 DNA，而蛋白质外壳则留在外面。噬菌体 DNA 能借用细菌复制机制，大量复制噬菌体 DNA，并翻译出噬菌体蛋白质，最终组装成 T_2 子代。既然 DNA 是世代间的连续物质，那么遗传物质只能是 DNA 而不是蛋白质。

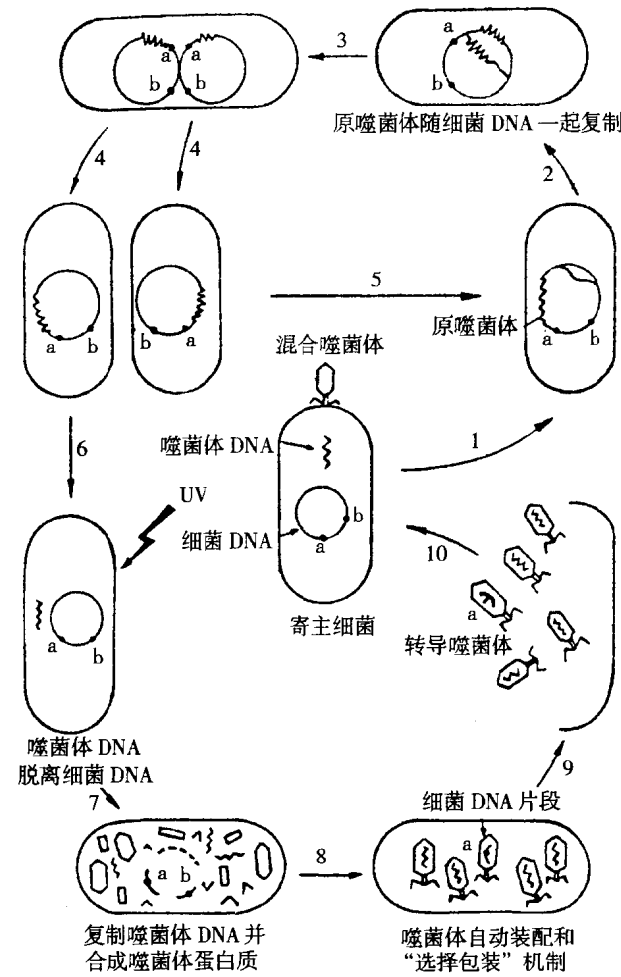


图 1-2 温和噬菌体溶源周期和溶菌周期

其他一些实验也证实了上述结论。例如，转导试验验证实在噬菌体侵染细菌延续世代过程中，可能会将前一寄主的 DNA 片段带至后一寄主，使后者呈现前者的某种遗传性状（这个过程称为转导），直接证实了性状由 DNA 片段所决定。20 世纪末叶，方兴未艾的转基因技术是人们对 DNA 片段决定性状的这一基本原理的最直接的证明。例如，我国将大豆根瘤中的固氮基因转入禾本类或其他作物，使受体株获得固氮性状。将人白蛋白基因序列转入转基因牛，使牛奶中含有人白蛋白产物。近年来，与人类白内障基因高度同源的小鼠白内障隐性位点也已定位，基因决定性状的也为基因治疗提供了广阔的前景。

（二）DNA 作为遗传物质的结构基础

1953 年 Watson 和 Crick 建立的 DNA 双螺旋模型（图 1-3），以无可辩驳的事实说明了

DNA 作为遗传物质在结构上是最合理的。

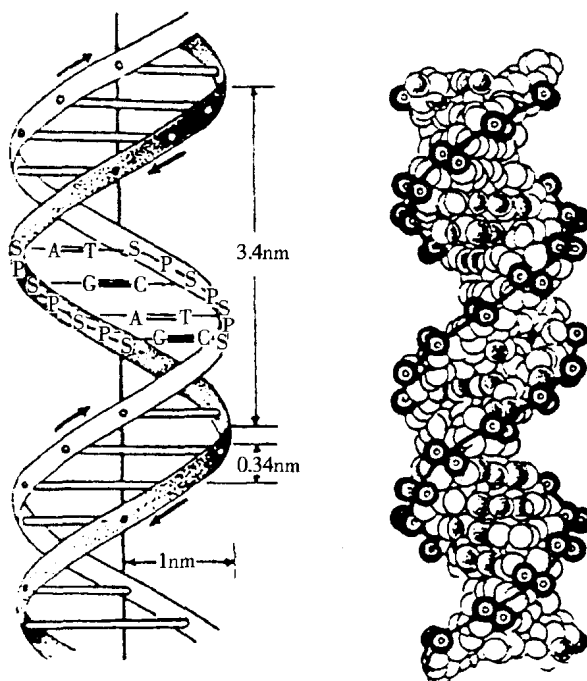


图 1-3 DNA 分子的双螺旋模型

1. 遗传物质的一般特性

遗传物质一般必须具备下列特性：①能贮存遗传信息，并将之传递给子代。②这种信息传递是精确的，极少发生差错。③遗传物质应具有物理的和化学的稳定性，从而不会丧失较多的信息。④具有一定的变异能力，这是物种进化的基础。

2. DNA 的结构合理性

DNA 作为遗传物质，其结构合理性体现在：

(1) DNA 的碱基序列所携带的信息量大，为 4^n (n 为碱基数)。

(2) 双螺旋结构保证了复制传代的精确性，一旦发生复制差错还有修复机制来纠正。

(3) 分子结构的理化稳定性体现在：①戊糖中 C-C 键除了在非常高的浓度下与强酸作用外，能经受各种化学攻击。②连接碱基与戊糖的 N-糖苷键除了在极端 pH 下，一般是稳定的，只有当碱基受到修饰，如嘌呤环第 7 位氮原子甲基化才大大减弱化学稳定性，这就是氮或硫芥子气的作用方式。③DNA 内部疏水性碱基堆积成一个巨大的圆柱体，将水完全排斥在碱基部位以外，有害化合物无法攻击内部的带电基团。④戊糖 2 位碳的脱氧消除了 -OH 在碱性条件下对磷酸二酯键的断裂作用。⑤DNA 链可发生低频率的自发突变。

(三) 基因概念

基因的概念起始于孟德尔的豌豆杂交试验 (1865 年)，在孟德尔之前，有人曾把遗传性状的传递看作是双亲性状的融合过程，这种掺和模式甚至为达尔文所接受。孟德尔的开创性研究在于发现能遗传的相对性状是由一对遗传因子，即基因控制的，相对性状取决于

两个基因之间的显隐性关系，并提出了著名的遗传分离律和自由组合律。例如花色是一种表现型，而决定开黄花的 YY 或 Yy 和决定开白花的 yy 都是基因型。由于基因型为 Yy 的植株开黄花，所以决定开黄花的 Y 是显性基因，y 则是隐性基因。Y 和 y 这两个等位基因在形成生殖细胞时必然要分离，即进入不同的生殖细胞，这就是分离律。后来知道，等位基因之间的显隐性关系并非十分绝对，上述 Yy 一类的杂合型基因型有的可产生中间性状，而且显隐性关系是严格地针对某一相对性状而言的。例如， $Hb^A Hb^S$ 是一种杂合型基因型，就贫血症而言， $Hb^A Hb^A$ 和 $Hb^A Hb^S$ 都表现为正常人， $Hb^S Hb^S$ 为患者，所以 Hb^A 对 Hb^S 为完全显性；但是就红细胞形态而言， $Hb^A Hb^A$ 在缺氧条件下是正常的圆盘形， $Hb^S Hb^S$ 是不正常的镰刀形，而 $Hb^A Hb^S$ 则是中间性的，则表现为不完全显性。再就血红蛋白在电泳中的迁移性而言，在 pH6.9 的磷酸缓冲液中， $Hb^A Hb^A$ 的蛋白移向负极， $Hb^S Hb^S$ 的蛋白移向正极， $Hb^A Hb^S$ 的蛋白混合物则分别移向两极，表现为共显性。这个例子同时也说明基因是多效的。

摩尔根不仅把孟德尔的遗传颗粒演绎为基因，并且全面陈述了基因学说的内容，包括提出了连锁交换律。摩尔根时代的实验证明，性状可以因决定它们的基因发生交换而分开。因此，基因是控制遗传性状的功能单位，是携带遗传信息的结构单位，是排列在染色体上的突变、交换和重组的单位。

Beadle 和 Tatum 等以诱发链孢霉基因突变的方法研究基因和酶的关系，得出“一个基因一个酶”的假说（图 1-4），揭示了基因通过酶来控制代谢，从而控制性状。

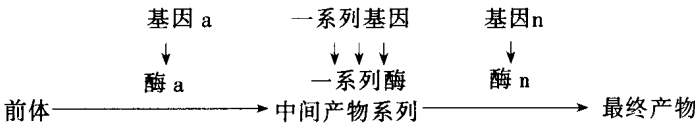


图 1-4 一个基因一个酶假说

（人类基因组计划预示，基因间的相互作用和基因内部的变化都十分复杂，上述简单的假说正趋消失）

从根本上动摇摩尔根的基因不能经交换而再分割概念的首推 Edward B. Lewis，他用实验证明基因内也可以发生重组。例如，d 和 e 是果蝇染色体上 bx 基因内的 2 个位点，如果 d 和 e 在同一条染色体上，野生型等位位点 + 和 + 在另一条染色体上，de/++ 称为顺式排列，杂合体表现为野生型，而反式排列的 d+/+e 表现为突变型。突变型中两条同源染色体的同一基因序列都是有缺陷的，但是表现型相同的两个突变型杂交后可出现 de/++ 顺式排列的子代，从而证明重组发生在基因内的不同位点上，基因内的位点缺陷通过重组而纠正。后来 Yanofsky 等用大肠杆菌中控制色氨酸合成酶形成的基因，Benzer 用 T₄ 噬菌体都证明了核苷酸才是重组交换的基本单位。

Benzer 研究了几百个 T₄ 噬菌体 rII 基因突变株的顺反效应，提出了顺反子概念。由于许多酶由 2 种或 2 种以上肽链组成，而且有的蛋白质根本就不是酶，因而把顺反子和肽链对应起来，即“一个顺反子一条多肽链”。提出顺反子概念旨在强调功能基因内可以突变和重组，但是顺反子毕竟没有包括一些非转录的调控序列，随着对非编码序列研究的深入，从分子概念来认识基因的结构与功能显得越来越有必要。

总之，1953 年 Watson 和 Crick 的 DNA 模型揭示了基因的分子本质，遗传信息就贮存在

DNA 分子的碱基排列之中,基因就是具有生物学功能的一段 DNA 片段。1961 年 Jacob 和 Monod 细菌操纵子模型揭示了调控序列对结构基因的作用以及多顺反子 mRNA 的存在。1965 年 Nirenberg 和 Matthai 破译了三联体遗传密码,使得遗传信息如何从 DNA 经由 mRNA 传至多肽链、构成蛋白质有了合理的解释,DNA 如何通过转录和翻译编码多肽链大体得以阐明。1977 年 Gilbert 报道真核细胞 DNA 中存在内含子,这些存在于基因内外的非编码序列究竟有什么作用引起人们的极大关注。可以相信,随着人类基因组计划的完成和后基因时代新的研究思路的实施,人们对基因概念一定会有更新的认识。今天可以这样认为:基因是 DNA 分子上的一个具有生物学效应(编码蛋白质)的片段,它主要指编码功能蛋白质多肽链的核酸序列,也包括转录出 rRNA、tRNA 等 RNA 的核酸序列,保证转录及转录后加工的序列和其他各种非编码序列,如内含子,调控序列,5'、3'端的非编码序列等。

二、基因传递规律与染色体传递规律的相应性

(一) 从 DNA 到染色体

按过去观点,人体每个细胞的 DNA 含有 3×10^9 bp 的核苷酸,以一个基因平均为 1 000bp 计算,至少包括 5~10 万个基因,这些染色质 DNA 连接起来,总长度可近 2m。如此长的 DNA 要存放在直径大致为 $5 \sim 10 \mu\text{m}$ 的细胞核内并行使各种功能,如何折叠引人注目。

染色体是处于细胞分裂期的高度凝缩了的染色质,一旦分裂结束,它又回复成染色质形态(图 1-5)。染色体是遗传信息的载体,由 DNA、RNA、组蛋白和非组蛋白按 1:0.05:1:0.5~1.5 比例组成。按照单线学说,每一染色体上包含一个 DNA 分子,细胞分裂中期的 X 型染色体中才含有 2 个尚未分开的 DNA 分子。

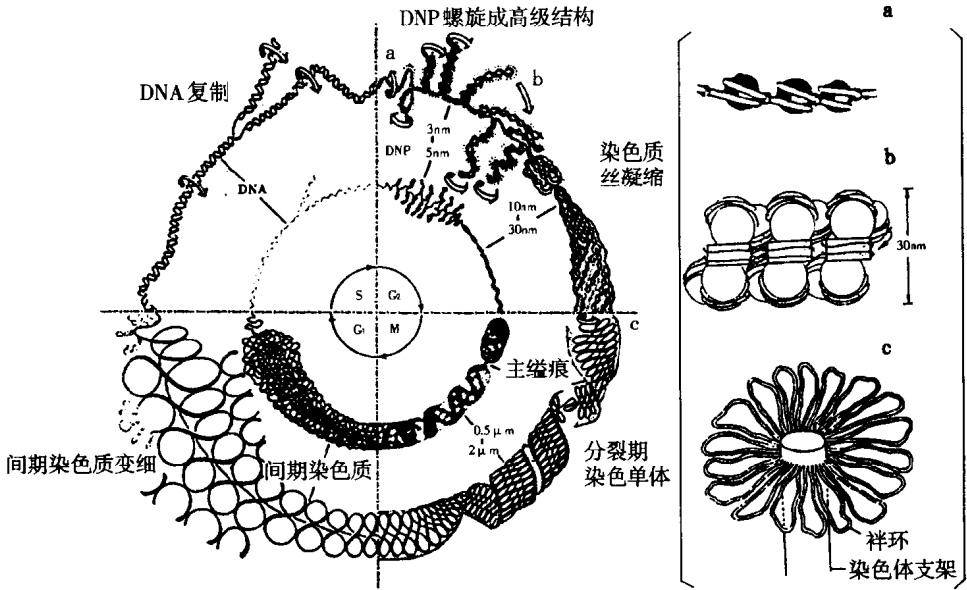


图 1-5 DNA 和染色体形态在细胞周期中的变化

a 串珠状核小体。b 螺线管螺旋(侧面观)。c 染色体袢环结构。

按照 Marsden 和 Laemmli 最先提出的袢环模型(loop model),从 DNA 到染色体大致经历下面几种结构:

1. 核小体 (nucleosome) 及其串联体

核小体由组蛋白 H_2A 、 H_2B 、 H_3 、 H_4 的各 2 个分子组成八聚体核心，约 146bp 的 DNA 绕核心 1.75 圈（酵母为 165bp、鼠肝细胞为 196bp），约 60~100bp 的 DNA 与 H_1 组蛋白组成核心连接区域，不同生物的核心连接区碱基数不同（20~200bp）。许多核小体串连成念珠状，如图 1-5a 所示，念珠状核小体是染色质一级结构。

2. 螺线管

在 Mg^{2+} 作用下，念珠状核小体螺旋化形成直径为 30nm 的螺线管（图 1-5b）。螺线管为外径 30nm，内径 10nm，螺距 11nm 的中空管状结构，其横切面由 6 个核小体绕成一圈。螺线管是染色质的二级结构。

3. 染色体支架与袢环结构

30nm 的螺线管折旋成一个个袢环，它们依附在由非组蛋白的纤维网组成的染色体支架上（图 1-5c），由中央向四周散射。袢环可以像绳索一样自绕，6 个自绕的袢环构成一个玫瑰花结 (rosettes)，30 个玫瑰花结缠绕成一个直径更大的螺旋，一个染色单体包含 10 个这样的螺旋。袢环结构与超螺线管结构一样，都还是一种推测。但是从 DNA 凝缩成染色体，长度上大致可压缩 1 万倍左右。

(二) 基因传递与染色体传递的相应性

既然染色体是遗传物质的载体，那么基因和染色体在遗传传递规律上应该是一致的，这种相应性主要体现在下面几个方面：

1. 来源

决定相对性状的一对等位基因分别来源于 2 个亲本，含有等位基因的一对同源染色体也分别来源于 2 个亲本。

2. 遗传传递

在形成生殖细胞时，一对等位基因必然分离而进入不同生殖细胞，即表现遗传分离律；而一对同源染色体也必然分离而进入不同的子代细胞，即表现为减数分裂中同源染色体的减数行为。对于 2 对等位基因而言，2 个非等位基因必需组合进同一生殖细胞，这种组合是随机的，即表现遗传的自由组合律；而对于 2 对同源染色体而言，2 条非同源染色体必需组合进同一生殖细胞，这种组合也是随机的。如果 2 对等位基因在同一对同源染色体上，则同一染色体上的 2 个非等位基因是连锁传递的，即使发生交换与重组，其频率也是相当低的，这与连锁群的数目与同源染色体的对数一致相吻合。

3. 交换与重组

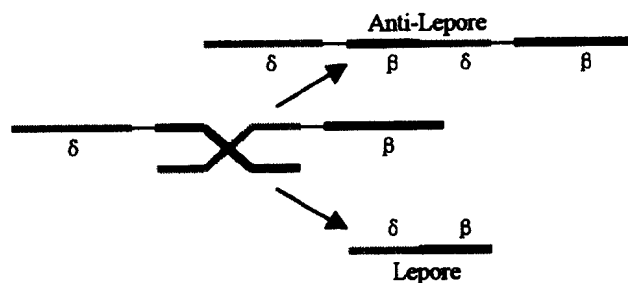


图 1-6 等位片段的不等位交换形成 β 地中海贫血

交换和重组发生在 2 个等位基因序列之内时可能会产生顺反效应，发生在 2 个等位基因之外时产生性状分离，发生在 2 个同源染色体之间时产生非亲本型染色体结构重排。图 1-6 显示同源染色体之间等位片段的不等交换会导致 β 地中海贫血，这种交换和重组可以发生在 β 基因内外的许多位点。如果发生在 β 基因之内，则在 2 条同源染色体上都形成融合基因（如图中 $\beta\delta$ 和 $\delta\beta$ ），最终产生融合蛋白而致病。如果发生在 β 基因之外，则形成一条染色体上 β 基因的缺失，造成遗传缺陷。

第二节 基因结构

一、基因结构的有序性

（一）编码序列与非编码序列

图 1-7 反映了真核细胞较典型的基因结构，若干个非编码的内含子将基因分隔成相应数目的具有编码作用的外显子，这就是存在于大多数真核细胞中的断裂基因（split gene）的结构形式。除了组蛋白、 α -干扰素和酵母一些蛋白外，所有高等真核生物的蛋白编码基因都含有内含子，只是数量不同、大小不一。但是不同生物的同基因中，内含子数目、位置往往相同。断裂基因当然还包括 5' 和 3' 端的非翻译区段。其实，非编码序列不仅内容十分丰富，而且所占的比重也大。以抗肌萎缩蛋白基因为例，其外显子序列只占全部序列的 6%，而非编码区如果包括基因间隔序列的话可达 90% 以上。不久前完成序列解读的果蝇基因组含有 2.5 亿碱基对，其中编码序列占百分之三十几。内含子按其序列特征和剪接方式大体分为 4 大类，即 I 类、II 类、一般核 mRNA 前体和核 tRNA 类（内含子保守序列及其剪接机制将在第 4 章介绍）。各内含子中都含有各自的保守序列，例如，I 类内含子含有高度保守的 P、Q、R、S 序列，核 mRNA 前体内含子含有高度保守的剪接分支点序列及 5'、3' 端序列等。但是各类内含子上的剪接信号却是相同的，它们遵循 GU-AG 规则，即内含子以 GT (GU) 为 5' 起始端，以 AG 为 3' 终止端，只有少数以 GC-AG 等排列。现已清楚，原核生物也有内含子，例如 T_4 噬菌体 *td*、*nrdB*、*sunY* 基因和枯草杆菌噬菌体 SPO₁ DNA 聚合酶基因，以及蓝细菌 tRNA^{Leu} 前体中都有 I 类内含子。

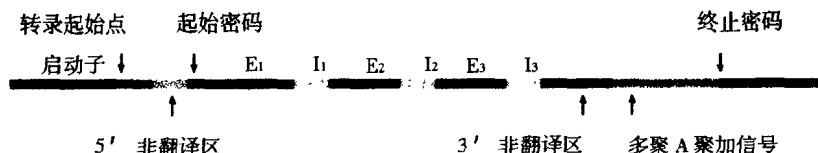


图 1-7 真核细胞基因结构 (E: 外显子, I: 内含子)

非编码序列除了内含子外还包括起信号作用或调节作用的一些序列，例如复制的起始区，转录的启动子区域，以及增强子、衰减子、活化序列、终止序列等，也包括接受调节蛋白识别和结合的序列、核基质紧密结合的序列等。以上述非编码序列中的转录启动子为例，其有序性的特征是非常明显的。启动子位于转录起始点上游，如 *E. coli* 启动子区域约 40~60bp，其中有两段共有序列（-10 和 -35 区），启动子有序性表现在 -10 和 -35 区域上的碱基高度保守，统计 100 多个启动子序列后得出 -35 区的 TTGACA 各碱基出现百分率

为 82、84、78、65、54、45，-10 区的 TATAAT 各碱基出现百分率则为 80、95、45、60、50、96，而且 2 个区域间隔为 17bp 左右，体现了相当大的序列稳定性。另外，真核生物中有一类序列，它们功能尚不明确，但在基因组中重复出现 $10^6 \sim 10^8$ 次，序列保持高度一致，称为高度重复序列。一些重复程度略低的中度重复序列也有同样的情况，序列短而保持不变。

(二) 基因重叠

基因序列的有序性不仅体现在碱基区域的排列上，而且还体现在同一序列的不同解读上，即不改变序列仅仅由于阅读框架（reading frame）的起始碱基不同，或对终止密码的通读（reading through）与否，可顺序读出不同的肽链（蛋白质），这些基因则称为重叠基因。基因重叠使基因组较小的原核细胞可经济地利用遗传物质。

在 $\Phi X 174$ 噬菌体内最早发现重叠现象（图 1-8）。 $\Phi X 174$ 是单链 DNA 病毒，全长 5 375 个碱基，欲编码 9 个蛋白质显然是勉为其难的，但是通过重叠阅读的方法，却可编码 11 个蛋白质。其重叠基因有两种不同的类型：第一种类型是，一个基因的核苷酸序列完全包含在另一个基因的核苷酸序列之中，只是它们的读码框架互不相同，如基因 B 重叠在基因 A 中，而基因 E 重叠在基因 D 中，形成基因中的基因。第二种类型是，两个基因的核苷酸序列之末端密码子发生重叠，如 A 基因终止密码子的 3 个核苷酸 TGA，与 C 基因的起始密码子 ATG 相互重叠了 2 个核苷酸；D 基因的终止密码子 TAA 与 J 基因的起始密码子 ATG 重叠了一个核苷酸（图 1-8）。从已经获得的资料分析，基因重叠可分为同相位重叠、异相位重叠、反向重叠等。阅读框架相同的同相位重叠基因所编码的蛋白质，其肽链序列相同，只是长短不一，区别在于能否通读终止密码，能读通则肽链较长。反向重叠基因的阅读框架也相同，但方向相反，其两股 DNA 都可编码。例如细菌 IS，向左编码的蛋白质较大，向右编码的蛋白质则较小，情况比较特殊。异相位重叠是指阅读框架不同，常常是上游基因的终止密码子和下游基因的起始密码子部分重叠。

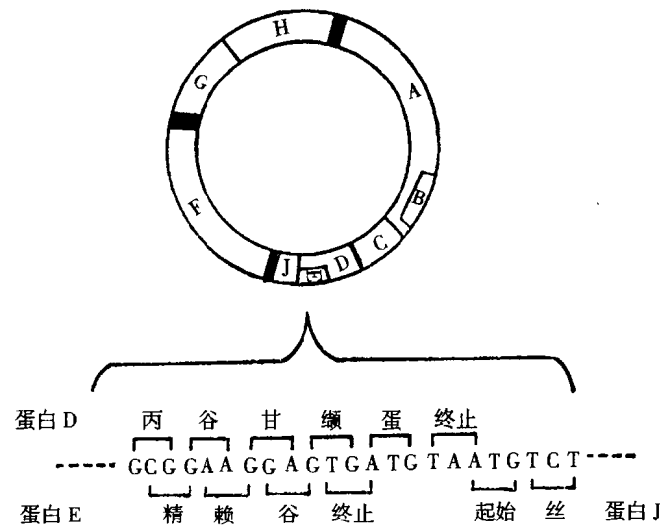


图 1-8 $\Phi X 174$ 噬菌体的基因重叠

(三) 操纵子

操纵子一般由几个功能相关的结构基因及其上游的调控序列组成，调控序列包括控制结构基因表达的操纵基因 (operator) 和启动子 (promotor)，它们又受其上游的调节基因 (regulatory gene) 的产物 (阻遏子) 控制 (图 6-1)。操纵子是原核生物中功能相关的基因联合表达的一种形式，它显示了基因群中序列的有序性与协调性。由于受控于同一调节基因，操纵子中几个结构基因的信息转录到共同的转录物中，这种转录物称为多顺反子 mRNA，是存在于原核细胞中的形式。

真核细胞中没有操纵子组合形式，但是其组蛋白基因是由 5 个功能类似的基因 (分别编码 H₁, H₂A, H₂B, H₃ 和 H₄ 5 种组蛋白) 串联组成的多基因家族 (gene family)，位于第 7 号染色体长臂上，人体基因组约含 10~40 份。上述组蛋白基因群每 5 个成一组的形式与原核生物操纵子有类似之处，但是组蛋白基因是各自转录，形成单顺反子 mRNA，这是真核生物基因群表达的特点。诚然，真核生物中还有一些结构和功能相似的基因所组成的多基因家族并不是串联重复排列的，而是散在于基因组织的各个部位，如珠蛋白、生长激素、干扰素等基因。

(四) 核糖体 RNA 基因的串联重复形式

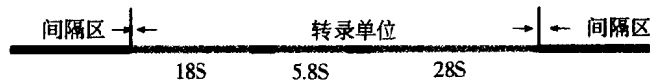


图 1-9 rRNA 基因串联重复序列

真核生物 rRNA 基因也是一种呈串联重复排列的多基因家族，由 18S、5.8S 和 28S 3 种序列依次排列成转录单位 (图 1-9)，每个转录单位的初级转录物由 47S 酶切为 45S (图 4-14)，一个 rDNA 基因簇 (gene cluster) 可包含许多转录单位，转录单位之间的间隔区随生物种类而长短不一。人体细胞可有 50~200 个 rDNA 基因簇，它们分布在能形成核仁的那些染色体上。此外，与 45S RNA 共同组成核糖体的还有 5S RNA，后者也是和一些序列组成串联重复形式，在细胞核外转录后再运送到核内构建核糖体前体物。

真核细胞内还有其他许多重复序列，它们所包含的碱基数多少不一，重复单位在个体间存在数量也不相同，但是，每个重复单位内的序列却总是一致的，它们往往是进化上的保守序列。

二、基因序列的变动性

(一) 突变 (mutation)

1. 突变概念

突变指突变体中 DNA 序列的任何改变，最常见的是一个或多个碱基的丢失、置换、插入或重排。碱基置换中，以嘧啶置换另一嘧啶或以嘌呤置换另一嘌呤的称为碱基转换；以嘧啶置换另一嘌呤或以嘌呤置换另一嘧啶的称为碱基颠换。有的突变即使仅仅是一个碱基的置换，却可致病。例如镰形细胞贫血症 (HbS) 是由于正常血红蛋白 β 链第 6 位谷氨酸密码子 GAA 变动为缬氨酸密码子 GUA 引起，这是一种颠换。而 HbC 则是由于 GAA 变动为 AAA 所致，这是一种转换，转换比较常见。非 3 的整数倍碱基的丢失或插入引起移码，这可能会全部改变该位点下游序列的编码含义。

2. 突变形成

自发突变由复制错配、碱基改变配对性质(包括丢失)等原因自发形成,诱发突变则是环境因素诱变造成。碱基配对性质改变由互变异构体引起,也可由脱氨等结构改变引起。图 1-10 显示亚硝酸诱发碱基脱氨而改变配对性质。一些主要的突变形成机理简述于表 1-1 中。

突变可能通过基因内外第二位点的突变得到回复,也可能通过机体修复机制如切除修复等来纠正。

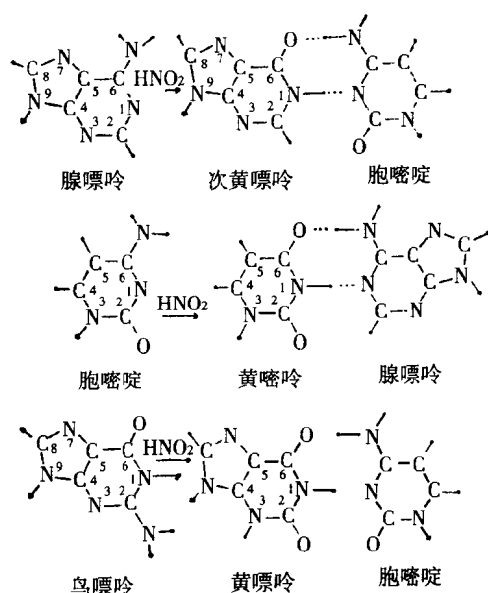


图 1-10 亚硝酸诱导突变的机理

亚硝酸可使 A 变成次黄嘌呤(I), I 只能和 C 配对从而产生 AT→GC 转换;亚硝酸也可使 C 变成 u, 并与 A 配对,从而产生 CG→AT 转换;但是 G 变成黄嘌呤(H)后,仍和 C 配对,不会产生突变。

表 1-1 突变形成方式

内外因素	机 理	结 果
复制错配	虽然机体有复制错配校正系统,但是随着分裂次数增加,错配不可避免	错配率: 大肠杆菌约 $10^{-5} \sim 10^{-6}$, 人约 10^{-10}
碱基脱氨 ($-\text{NH}_2$)	$\text{C} \rightarrow \text{U}$ $\text{A} \rightarrow \text{I}$ $\text{G} \rightarrow \text{H}$	$\text{CG} \rightarrow \text{AT}$ $\text{AT} \rightarrow \text{GC}$ 不产生突变
碱基丢失 (脱嘌呤或脱嘧啶)	糖苷键的丧失使 3'端磷酸二酯键易被水解	DNA 链断裂
碱基互变异构体	A 或 C 出现 $-\text{NH}_2$ 变为 $=\text{NH}$ 时	A = C 错配
碱基类似物 5-Bu 互变异构体	酮式与 A 配对后,再变为烯醇式,则与 G 配对	$\text{AT} \rightarrow \text{ABu} \rightarrow \text{GBu} \rightarrow \text{GC}$
紫外线照射	形成胸腺嘧啶二聚体等	影响转录、复制
吡啶橙、二氨基吡啶等	嵌入 DNA	引起移码突变