

精要速览系列

Instant Notes

GENETICS (THIRD EDITION)

遗传学 (第三版)



· 中译版 ·

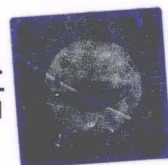
Hugh Fletcher, Ivor Hickey
& Paul Winter

编著

张博等译



科学出版社
www.sciencep.com



精要速览系列

Genetics

Third Edition

遗传学

(第三版, 中译版)



科学出版社

北京

内 容 简 介

“精要速览系列”(Instant Notes Series)丛书是国外教材“Best Seller”榜的上榜教材。该系列结构新颖,视角独特;重点明确,脉络分明;图表简明清晰;英文自然易懂,被国内多所重点院校选用作为双语教材。

第三版在第二版基础上进行修订,对人类基因研究、RNA 调控、组蛋白修饰、表观遗传学、人类进化等新近研究进行重点补充和调整,其他各章节也进行了修订。

本书适合普通高等院校生命科学、医学、农学等相关专业使用,也可作为双语教学参考教材使用。

G. I. Hickey, H. L. Fletcher, P. Winter

Instant Notes in Genetics, 3rd edition

© 2007 by Taylor & Francis Group

ISBN0-4153-7619-x

All Right Reserved. Published by arrangement with Taylor & Francis Books Ltd, 2 & 4 Park Square, Milton Park, Abingdon, OX14 4RN, UK.

Licensed for sale in the Mainland of China only, booksellers found selling this title outside the Mainland of China will be liable to prosecution. Copies of this book sold without a Taylor & Francis sticker on the cover are unauthorized and illegal.

本授权版本图书仅可在中国大陆范围内销售,中国大陆范围以外销售者将受到法律起诉。本书封面贴有 Taylor & Francis 防伪标签,未贴防伪标签属未获授权的非法行为。

图书在版编目(CIP)数据

遗传学:第3版/(英)弗莱彻(Fletcher, H. L.)等编著;张博等译. —北京:科学出版社,2010.9

(精要速览系列)

书名原文:Genetics

ISBN 978-7-03-028755-7

I. ①遗… II. ①弗…②张… III. ①遗传学-双语教学-高等学校-教材 IV. ①Q3

中国版本图书馆 CIP 数据核字(2010)第 164015 号

责任编辑:单冉东 席 慧 / 责任校对:邹慧卿
责任印制:张克忠 / 封面设计:耕者设计工作室

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

双青印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2001年5月第 一 版 开本:787×1092 1/16

2006年8月第 二 版 印张:19 1/2

2010年9月第 三 版 字数:460 000

2010年9月第一次印刷 印数:1—3 000

定价:53.00 元

(如有印装质量问题,我社负责调换)

译者名单

第三版翻译人员名单(按贡献大小排列)

张 博	高宇英	黄 鹏	肖 安
高 巍	徐琳杰	夏赭丹	梁 巍
王展翔	张 莹	陈 辉	

第二版翻译人员名单(按章节顺序排列)

谢 雍	周 晶	黄可峻	董 姝
李 辉	刘宇峰	毛贤军	沈宇峰
王显花	姚 蒙	张爱京	张 弩
顾惠娟	许元富	熊冬生	许小兰
林 焱	李光金	严维耀	任兆瑞
秦 焱			

前言

自本书第二版出版以来,人们对遗传学的研究取得了一些重要的进展。虽然基因组测序不再属于前沿领域,但是从中所获得的数据促进了诸多的研究工作。人类基因组仅有 20,000 个基因被鉴定出来,估计上限为 25 000 个基因;这远远低于先前的预测,将人类跟大多数其他的多细胞生物紧紧地联系在一起。人类的基因数目似乎比卷心菜的还要少。基因组比对结果还显示,有 67~83 个在黑猩猩体内依然活跃的基因,由于在进化过程中产生突变而在人类中成为没有活性的假基因。更令人惊异的是,人类基因组的某些部分与黑猩猩的差异比其余部分大得多。这表明当祖先群体向这两个不同的物种趋异进化时经历了很长一段时间,期间发生了一些遗传交换。

另一个新进展是发现 RNA 在控制染色质结构方面的作用,以及组蛋白 N 端的共价修饰在调节染色质结构方面的重要性。某些组蛋白修饰(如位点特异性甲基化)能够促使核小体紧密堆积,导致形成致密的染色质,造成转录装置难以接近,从而使其中的基因的表达受到限制。双链 RNA 能够引导特异的 DNA 序列发生这种修饰,使 CpG 序列(在植物中为 CpNpG)中的胞嘧啶被甲基化,并且这种修饰与组蛋白的去乙酰化相关。随着用以鉴定组蛋白特异修饰的单克隆抗体不断增多,人们对所谓的“组蛋白密码”的认识将会越来越清晰。

为了反映上述变化,我们减少了介绍基因组测序技术的篇幅,同时将表观遗传学移到了 A 单元中的“分子遗传学”部分,因为这样看起来更为恰当。人类的进化作为单独的一个单元出现。人类在地球上定居与迁移的地理学越来越清楚,而随着类人猿基因组的信息逐渐明朗,人们才刚刚开始认识人类与类人猿之间在分子水平上的差异。

有很多审稿人对本书的前一个版本提出了评论和修改建议,我们在此表示感谢。在力所能及的范围内,我们尽可能采纳了他们的建议。我们希望至少有一些审稿人对某些改动会感到满意。

目 录

前言

缩略词

A 分子遗传学	1
A1 DNA 结构	1
A2 基因	5
A3 遗传密码	9
A4 基因转录	11
A5 转移 RNA	18
A6 核糖体 RNA	22
A7 信使 RNA	25
A8 翻译	30
A9 DNA 复制	37
A10 原核生物基因表达调控	42
A11 真核生物基因表达调控	47
A12 表观遗传学与染色质修饰	53
B 基因组	59
B1 染色体	59
B2 细胞分裂	66
B3 原核生物基因组	69
B4 真核生物基因组	73
B5 DNA 突变	79
B6 诱变剂和 DNA 修复	84
B7 重组	91
B8 噬菌体	97
B9 真核生物的病毒	102
B10 基因组分析	107
C 遗传机制	111
C1 基础孟德尔遗传学	111
C2 孟德尔遗传学(续)	117
C3 减数分裂和配子发生	123
C4 连锁	129
C5 细菌间的基因转移	134
C6 真核生物细胞器中的基因	140
C7 数量遗传	143
C8 性别决定	151
C9 性别和遗传	155
C10 近交	158
C11 概率	162

C12 适合度检验:卡方和精确度检验	166
D 群体遗传学与进化	170
D1 简介	170
D2 通过自然选择的进化	173
D3 群体中的基因:哈迪-温伯格平衡	178
D4 遗传多样性	185
D5 新达尔文进化论:选择作用于等位基因	189
D6 染色体在进化中的变化	193
D7 物种和物种形成	202
D8 多倍性	208
D9 进化	212
D10 人类的进化	217
E 重组 DNA 技术	221
E1 DNA 研究技术	221
E2 RNA 研究技术	228
E3 DNA 克隆和转染	231
E4 生物信息学	239
F 人类遗传学	245
F1 遗传病	245
F2 遗传筛选	253
F3 基因与癌症	256
F4 基因治疗	260
G 遗传学的应用	263
G1 遗传学在法医学中的应用	263
G2 生物技术	268
G3 转基因学	273
G4 伦理学问题	277
进一步阅读文献和网站的使用	283
索引	290

A 分子遗传学

A1 DNA 结构

要 点

核 苷 酸

DNA 是由核苷酸单体构成的链状聚合物,每个核苷酸由一个糖基、一个碱基和一个磷酸基团组成。DNA 的糖基是 2'-脱氧核糖,含有 5 个碳原子,分别称为 1'、2'……碳原子。DNA 的碱基分为 4 种:腺嘌呤和鸟嘌呤含有两个碳氮杂环,属于嘌呤碱;胸腺嘧啶和胞嘧啶只含有一个碳氮杂环,属于嘧啶碱。这些碱基与脱氧核糖的 1'碳原子相连,构成核苷。糖基的 5'碳原子可以与一个、两个或三个磷酸基团相连,构成一个核苷酸。核苷酸能够以单个分子或多聚体(如 DNA 或 RNA)的形式存在。

DNA 多聚核 苷酸

分别对应 4 种碱基的核苷三磷酸相连接而形成了 DNA 多聚核苷酸链。核苷酸聚合时,其中的两个磷酸基团脱落,余下的 5'磷酸基与下一个核苷酸的 3'羟基形成磷酸二酯键。多聚核苷酸的一端为游离的 5'磷酸基(5'端),另一端为游离的 3'羟基(3'端)。遗传信息由碱基的顺序编码,其阅读方向为从 5'到 3'。多聚核苷酸链很长,可以产生 4^n (n 为核苷酸数)种不同的序列。

双 螺 旋

DNA 分子由两条互相缠绕成双螺旋结构的多聚核苷酸链组成。其中的糖基和磷酸基团构成 DNA 分子的骨架,碱基则面向内部重叠排列。两条多聚核苷酸链走向相反。DNA 双螺旋结构是右旋的,每 10 对碱基构成一个螺旋周期。双螺旋结构的大沟可与蛋白质相互作用。目前已发现了各种不同的 DNA 结构,包括具有左旋结构的 Z-DNA。

碱基互补配对

两条 DNA 链上碱基之间形成的氢键维持了双螺旋结构的稳定性。由于受到双链内的空间限制,碱基配对只能发生在嘌呤和嘧啶之间,其中 A 只能与 T 配对,G 只能与 C 配对。这就是碱基互补配对原则。碱基配对的限制意味着两条链上的碱基顺序是相互关联的,一条链的顺序决定了另一条链的顺序。这样能够使遗传信息在 DNA 复制和基因表达的过程中得以维持。加热、化学物质或酶的作用能够导致 DNA 碱基间氢键的断裂,使双螺旋的两条链分开变成单链。

RNA 的结构

在 RNA 中,胸腺嘧啶被尿嘧啶取代,2'脱氧核糖被核糖取代。RNA 通常以单链多聚核苷酸的形式存在,但在互补的序列之间有可能产生短片段的碱基配对。

相关主题

基因转录(A4)

DNA 突变(B5)

DNA 复制(A9)

核 苷 酸

DNA 携带细胞复制所必需的遗传信息的能力与 DNA 分子的结构密切相关,DNA 是由被称为核苷酸(nucleotide)的单体组成的长链状 DNA 分子聚合物,因而被称为多聚核苷酸。每个核苷酸含有三个部分:糖基、碱基(base)和磷酸基团。DNA 所含的糖是一种叫做 2'-脱氧核糖的五碳糖,它的 2'碳原子上的羟基被氢原子取代(图 A1-1)。五碳糖上



图 A1-1 2'-脱氧核糖的结构。R=H

的碳原子分别用 1~5 编号,数字后缀以撇号('),称为 **prime**,以与位于碱基上的碳相区别。这种编号很重要,因为它能够反映核苷酸中的其他成分与糖基相连的情况。

每个核苷酸含有 4 种碱基中的一种:腺嘌呤(**adenine**)、鸟嘌呤(**guanine**)、胸腺嘧啶(**thymine**)或胞嘧啶(**cytosine**)(图 A1-2)。这些碱基都是含有碳氮杂环的复杂分子。腺嘌呤和鸟嘌呤含有两个碳氮杂环,都称为嘌呤(**purine**),而胞嘧啶和胸腺嘧啶则含有单环,都称为嘧啶(**pyrimidine**)。嘌呤的第 9 位氮原子或嘧啶的第 1 位氮原子以共价键与糖基的 1' 碳原子相连,从而将碱基连接到糖基上。带有碱基的核糖称为核苷(**nucleoside**)(图 A1-3a)。

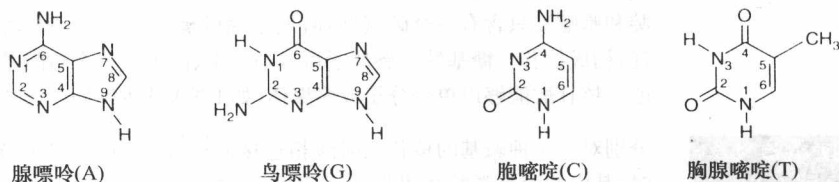


图 A1-2 DNA 中的碱基

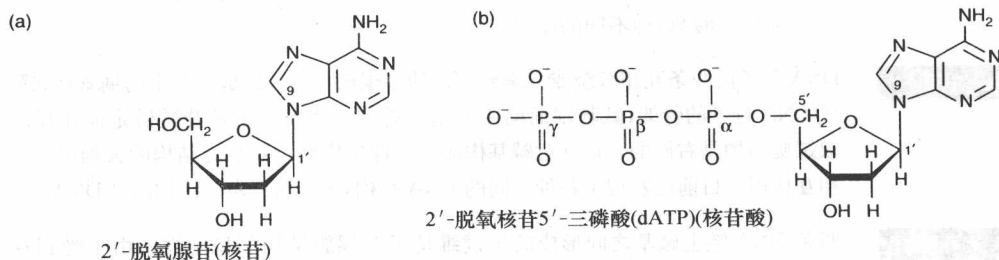


图 A1-3 (a)核苷的结构;(b)核苷酸的结构

核苷酸中的磷酸基团(PO_4)与糖基的 5' 碳原子相连(图 A1-3b)。含有磷酸基的核苷称为核苷酸,它可以带有 1~3 个磷酸基团,分别被称为 α 、 β 和 γ ,其中 α 与糖基直接相连。核苷酸在细胞中可以以单体形式存在(例如,三磷酸核苷作为一种能量载体为酶促反应提供能量,因而在细胞中具有重要作用),也可以聚合在一起成为核酸(DNA 或 RNA)。

DNA 多聚核苷酸

核苷三磷酸相互连接可形成多聚核苷酸。有 4 种核苷酸参与形成 DNA 多聚核苷酸,它们分别是:2'-脱氧腺苷 5'-三磷酸(dATP 或 A)、2'-脱氧胸苷 5'-三磷酸(dTTP 或 T)、2'-脱氧胞苷 5'-三磷酸(dCTP 或 C)和 2'-脱氧鸟苷 5'-三磷酸(dGTP 或 G)。在聚合时, β 和 γ 磷酸基脱落,核苷酸单体通过余下的磷酸基连接在一起。在聚合反应中,一个核苷酸的 5'-磷酸基与下一个核苷酸的 3'-碳原子形成共价键,该 3'-碳原子上的羟基被去除。所形成的共价键称为 3'-5' 磷酸二酯键[3'-5' **phosphodiester bond** (C-O-P)](图 A1-4)。多核苷酸链的一端是一个游离的 5'-三磷酸基,称为 5' 端;另一端是一个游离的 3'-羟基,称为 3' 端。这样就赋予了 DNA 分子极性,可以将 DNA 分子的走向描述为从 5'→3' 或 3'→5'。

遗传信息是由 DNA 多聚核苷酸中的碱基顺序编码的。序列信息通常按 5'→3' 的方向书写(聚合酶按这个方向复制 DNA)。多聚核苷酸序列可以很长,对核苷酸的数目和排列顺序均无限制。对于一条多聚核苷酸而言,其最多可能有 4^n 种碱基序列, n 表示核苷酸的数目。这是一个巨大的数目,如一条仅含有 6 个碱基的多核苷酸链可以有 $4^6 = 4096$ 种不同的序列。

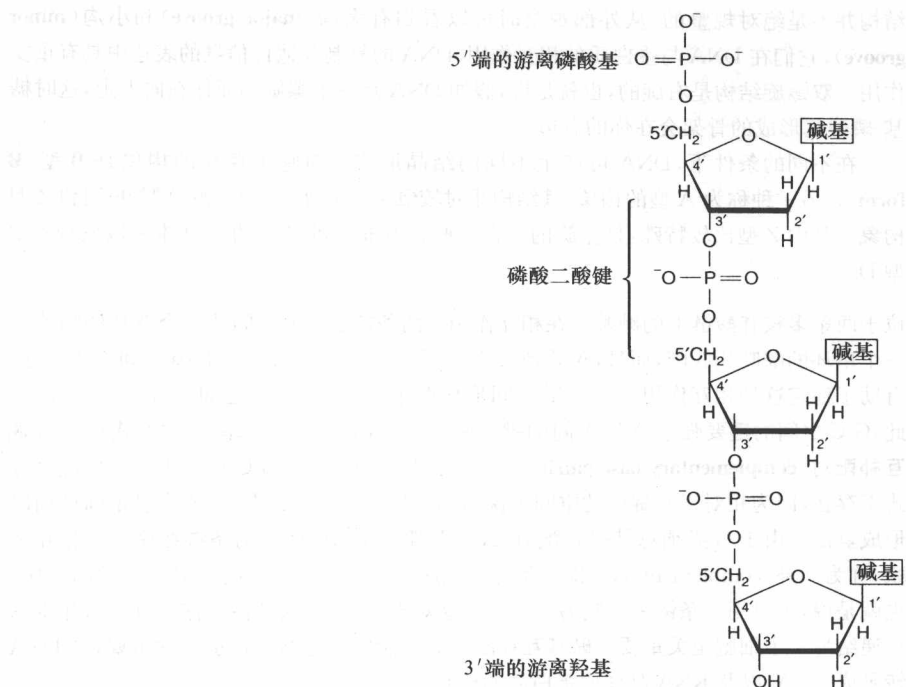


图 A1-4 DNA 多聚核苷酸中将核苷酸连接起来的磷酸二酯键

双螺旋

DNA 具有与众不同的、特征性的三维空间结构,称为双螺旋(double helix)(图 A1-5)。Watson 和 Crick 于 1953 年在剑桥工作时从 Franklin 和 Wilkins 拍摄的 X 射线衍射图上发现了 DNA 的这种结构。DNA 的两条多核苷酸链相互缠绕形成双螺旋结构。其中的糖基和磷酸基在双螺旋的外侧形成脊柱或骨架,扁平的碱基像碟子一样堆叠在一起,面向双螺旋的中心。

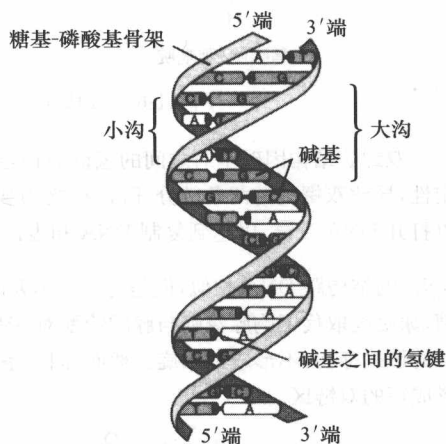


图 A1-5 双螺旋

双螺旋的 X 射线衍射图显示出重复性的衍射条带,反映了 DNA 结构的规律性。双螺旋的每 10 个碱基对旋转一周,高度约是 $34\text{\AA}$,因此每个碱基对之间的距离是 $3.4\text{\AA}$ 。双螺旋的直径是 $20\text{\AA}$ 。双螺旋呈反向平行(anti-parallel),其中一条链的方向为 $5'\rightarrow 3'$,另一条为 $3'\rightarrow 5'$ 。只有反向平行的多核苷酸链才能形成稳定的双螺旋。双螺旋

结构并不是绝对规整的,从外部观察时可以看到有大沟(major groove)和小沟(minor groove),它们在 DNA 与蛋白质的相互作用、DNA 的复制及遗传信息的表达中具有重要作用。双螺旋结构是右旋的,也就是说,假如 DNA 是一个螺旋梯而你在向上走,这时糖基-磷酸基形成的骨架会在你的右边。

在不同的条件下,DNA 可产生不同的结晶形式。细胞中存在的构象是 B 型(B form)。另一种称为 A 型的构象,其结构相对较致密。另外还有 C 型、D 型、E 型和 Z 型构象。其中 Z 型比较特殊,是左旋的。在一些含有非标准结构的染色体区域发现过 Z 型 DNA。

碱基互补配对

位于两条多核苷酸链上的碱基存在相互作用。两条链之间的空间使一个双环的嘌呤和一个单环的嘧啶相匹配,而且,A 总是与 T 配对,G 总是与 C 配对。碱基之间形成氢键,有助于稳定这种相互作用。A 与 T 之间形成两个氢键,G 与 C 之间形成三个氢键,因此,G-C 之间的键要强于 A-T 之间的键。两条 DNA 链之间碱基配对的方式称为**碱基互补配对(complementary base pairing)**,极其重要(图 A1-6)。G-C 与 A-T 之外的配对方式不存在,因为相对于螺旋内的空间来说,它们要么太大、要么太小,或难以正确排列以形成氢键。由于 G 必须总是跟 C 配对、A 总是跟 T 配对,使得两条链在序列上相互关联,称为互补(complementary),即一条链的顺序可以用来预测和决定另一条链的顺序。也就是说,可以用一条链来复制另一条链。这对于在细胞分裂后保留遗传信息,并将它传递给另一个细胞至关重要。碱基互补配对对于遗传信息的表达也非常重要,是 DNA 转录成 RNA,以及 RNA 翻译成蛋白质的核心。

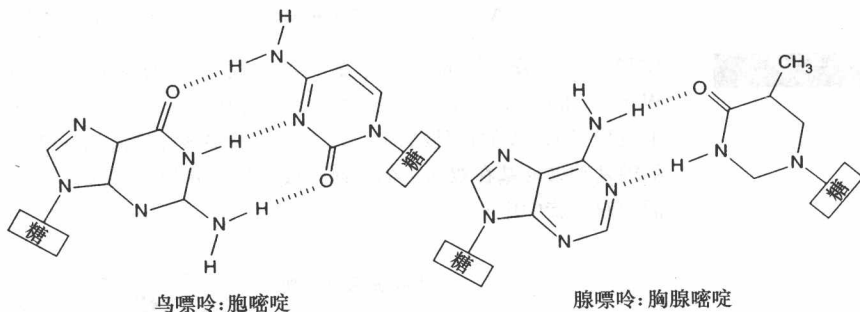


图 A1-6 碱基互补配对。虚线表示氢键

双螺旋结构因碱基对之间的氢键而稳定,但是受热或某些化学物质可破坏这种稳定性,导致双螺旋的两条链分开,这称之为**变性(denaturation)**。在细胞内,特定的酶可以打开 DNA 双链,以达到复制 DNA 和表达遗传信息的目的。

RNA 的结构

RNA 的结构跟 DNA 相似,但是有一些重要的区别。在 RNA 中,核糖取代了 2'-脱氧核糖,尿嘧啶取代了胸腺嘧啶与腺嘌呤配对(图 A1-7)。此外,RNA 通常以单链多聚核苷酸的形式存在,不形成双螺旋。然而,同一条 RNA 链上的互补部分也会产生碱基配对,形成短的双链区。

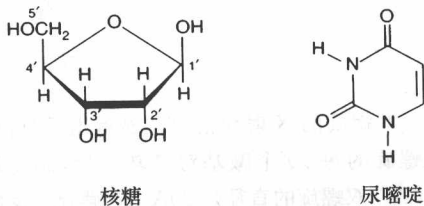


图 A1-7 核糖和尿嘧啶的结构

A2 基因

要 点

基因的结构

基因是遗传信息的基本单位,它对应于 DNA 上一个不连续的区段,编码一个多肽的氨基酸序列^①。人类细胞约含有 25 000 个基因,分布于 23 条染色体上。基因是不连续的,被非编码的基因间 DNA 分开。遗传信息由模板链编码,用于指导 RNA 分子的合成^②。两条 DNA 链都能成为模板链。DNA 分子具有庞大的储存遗传信息的能力。

基因家族

有些基因成簇排列,称为操纵子或多基因家族。操纵子存在于细菌中,含有被共同调控、功能相关的基因。多基因家族存在于高等生物中,它含有一组相同或类似的基因,但是这些基因并不受协同调控。简单的多基因家族含有相同的基因,以满足细胞对该基因产物的大量需要。复杂的多基因家族则含有非常类似的基因,这些基因往往编码功能相关的蛋白质。

基因表达

编码于基因中的生物信息通过基因表达而得以表现。在这个过程中,基因首先指导合成 RNA,然后由 RNA 指导合成蛋白质。基因表达的中心法则指出:(遗传)信息总是由 DNA 传递给 RNA,再由 RNA 传递给蛋白质。细胞的功能通过许多蛋白质的协同作用而实现。基因的表达保证了蛋白质在合适的位置和合适的时间被合成。

基因启动子

基因表达是受到高度调节的。细胞中并非所有的基因都处于活化状态,不同类型的细胞表达不同的基因。基因的表达受到位于编码区上游的一段 DNA 序列调控,称为启动子。它能够与 RNA 聚合酶及有关的转录因子相结合,从而启动 RNA 分子的合成。

内含子和外显子

基因中的编码序列被分割成一系列的片段,称为外显子;它被称为内含子的非编码序列分开。内含子通常占据大部分的基因序列,其大小和数量因不同的基因而异。在蛋白质合成之前,内含子将被从 RNA 转录本中去除,这个过程称为剪接。细菌的基因中通常不存在内含子。

假 基 因

某些基因的拷贝含有在进化过程中获得的序列错误,使它们无法产生蛋白。这些基因被称为假基因,它们代表原始基因的进化遗迹。珠蛋白假基因即为一例。

相关主题

原核生物基因表达调控(A10)

原核生物基因组(B3)

真核生物基因表达调控(A11)

真核生物基因组(B4)

表观遗传学与染色质修饰(A12)

注①:关于基因的概念,更为普遍的看法是:基因是指一段特殊的 DNA,这一段 DNA 能够作为一个完整的功能单位指导合成一条多肽链或一个 RNA 分子。

注②:应为:遗传信息由有义链编码,与之互补的为模板链,该链用于指导 RNA 分子的合成。

基因的结构

生物体复制自身所需的生物信息存在于它的 DNA 之中。这些信息编码在 DNA 的碱基序列之中,组成大量的基因,每个基因包含合成一种多肽的信息*。从物理学的角度看,基因是 DNA 上一个不连续的区段**,其碱基顺序决定了一个多肽的氨基酸顺序。基因的大小可以从少于 100 个碱基对到几百万个碱基对不等。在高等生物中,基因存

* 译者注:也有不编码多肽的基因,如编码 rRNA、tRNA、miRNA 等的基因。

** 译者注:基因应视为一段连续的 DNA 序列。

在于一系列很长的称为**染色体(chromosome)**的DNA分子中*。人类中有20 000~25 000个基因,分布于23条染色体上。基因的分布是不连续的,彼此之间被一些看起来不含有任何有用信息的序列分开,这些序列称为**基因间DNA(intergenic DNA)**。基因间DNA很长,如在人类中基因序列占到总DNA的不到30%。DNA的双链中只有一条携带着生物信息,这条DNA链被称为**模板链(template strand)**,可以用来合成一条与其互补的RNA分子,RNA则指导多肽的合成。另一条DNA链称为**非模板链(nontemplate strand)**。DNA双螺旋的两条链都有成为模板链的可能,不同的基因可以存在于不同的链上。除模板链和非模板链外,还存在其他一些用于描述双螺旋链的术语,包括有义(sense)/反义(antisense)、编码(coding)/非编码(noncoding),其中反义链和非编码链相当于模板链。

DNA分子储存遗传信息的能力巨大。对于一个具有 n 个碱基的DNA分子来说,4个碱基随机组合可产生 4^n 种不同的排列。即使是很短的DNA分子,它可能出现的不同序列的数目也是很大的。虽然实际上包含有用信息的序列是有限的,但不可否认它编码信息的能力仍是极大的。

基因家族

大多数基因是随机地排列在染色体上,但是也有一些基因成群或成簇地组织在一起。主要有两种成簇存在的基因,一种是**操纵子(operon)**(图A2-1),另一种是**多基因家族(multigene family)**(图A2-2)。

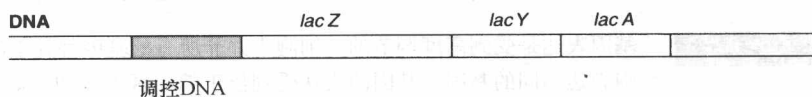


图 A2-1 乳糖操纵子。三个基因(*lacZ*、*lacY* 和 *lacA*)排列在一起并受到共同调节

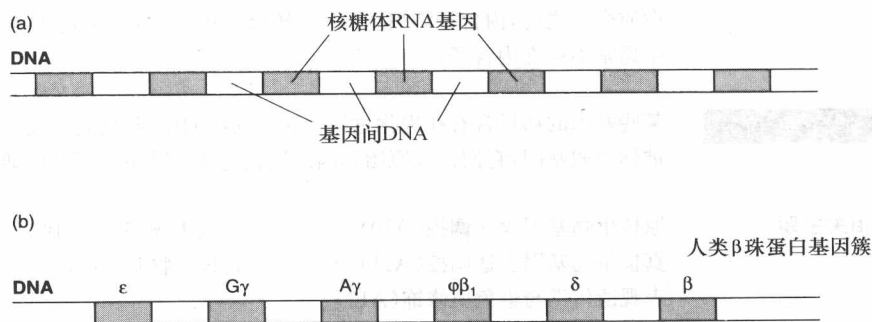


图 A2-2 多基因家族。(a)简单的(多基因家族);(b)复杂的(多基因家族)

操纵子是在细菌中发现的基因家族,它含有协同调控的基因,编码功能上密切相关的蛋白质。以 *E. coli* 的乳糖(*lac*)操纵子为例,它含有3个基因,编码细菌分解乳糖所必需的酶。当乳糖为能量来源时,就需要乳糖操纵子所编码的这3个蛋白质。由于存在于同一个操纵子中,使得这些基因可同时表达或关闭,从而使生物体能够有效地利用资源(图A2-1)。

高等动物中不存在操纵子,其基因簇以多基因家族的形式存在。与操纵子不同的是,多基因家族中的基因是相同的或非常相似的,但是并没有协同调控的关系。多基因家族中基因成簇存在可能反映了需要相应基因的多个拷贝,这种多拷贝是在进化过程

* 译者注:DNA≠染色体。

中通过基因重复来实现的。有些多基因家族成簇分布在不同的染色体上,这可能是在进化过程中发生了 DNA 重排,使多基因簇分开而造成的。多基因家族可以是简单的或复杂的。**简单的多基因家族 (simple multigene family)** (图 A2-2a) 中的基因是相同的。例如, 5S 核糖体 RNA 基因, 在人类中存在着约 2000 个成簇的 5S rRNA 基因的拷贝, 这反映了细胞对该基因产物的大量需求。**复杂的多基因家族 (complex multigene family)** (图 A2-2b) 中的基因十分类似但并不相同。例如, 珠蛋白基因家族编码一系列多肽 (α 、 β 、 γ 、 ϵ 和 χ 珠蛋白), 它们彼此之间仅相差几个氨基酸。珠蛋白可相互结合成复合物, 再结合一个称为血红素的辅助因子, 形成成体和胚胎形式的能携带氧的血红蛋白 (hemoglobin)。

基因表达

DNA 分子中的生物信息存在于其碱基顺序中。基因表达是将这些信息展示在细胞中的过程。这种利用信息的过程称为**中心法则 (central dogma)**, 由 Crick 首先提出, 它指出遗传信息是由 DNA 传给 RNA, 再传给蛋白质 (图 A2-3)。在基因表达过程中, DNA 分子通过指导合成一条序列互补的 RNA 链拷贝自身的信息。这个过程称为**转录 (transcription)**。然后, RNA 指导多肽的合成, 该肽链的氨基酸序列由 RNA 的碱基序列决定。这个过程称为**翻译 (translation)**。蛋白质的氨基酸顺序决定了它的三维结构, 进而决定了该蛋白质的功能。中心法则指出遗传信息的传递只能是单向的, 即从 DNA 到 RNA, 再到蛋白质, 而不能反向传递。这个法则的一个例外是在反转录病毒中发现的。这种病毒含有一种酶, 称为**反转录酶 (reverse transcriptase)**, 能够以 RNA 为模板合成 DNA。细胞的功能, 以至整个生物体的功能, 依赖于很多不同的蛋白质之间的相互协调。基因中携带的生物信息作为一套指令, 指导蛋白质在合适的时间和合适的地点被合成。

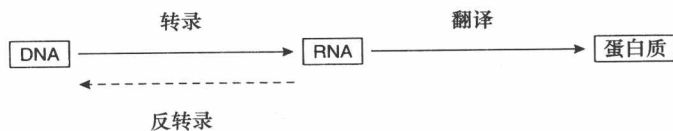


图 A2-3 中心法则(上)和反转录(下)

基因启动子

基因所携带的生物信息的表达是受到高度调控的, 并非细胞中 DNA 上所有的基因都会被表达, 不同的细胞会表达不同的基因。细胞中活跃基因的总体情况决定了该细胞的特性及它在生物体中的功能。例如, 在肌肉细胞中有活性的基因不同于在血细胞中有活性的基因。基因的表达是由一段位于编码序列上游的 DNA 序列调控的, 这段序列称为**启动子 (promoter)**。启动子中的保守序列可以被 RNA 聚合酶和其他称为**转录因子 (transcription factor)** 的相关蛋白质识别并结合, 从而启动基因的转录。细胞中某个基因的表达由其启动子的序列及该启动子与 RNA 聚合酶及转录因子的结合能力来决定。

内含子和外显子

基因的一个更为令人惊讶的特点是, 在高等生物中, 编码信息通常被分割成一系列的 DNA 序列片段, 称为**外显子 (exon)**, 它们被不含有用信息的**内含子 (intron)** 隔开 (图 A2-4)。每个基因中所含有的内含子的数目变化很大, 可以从 0~50 多个不等。外显子和内含子的长度也有变化, 但通常内含子要比外显子长, 占了整个基因序列的大部分。在基因内的生物信息被用于合成蛋白质之前, 内含子必须通过一个称为**剪接 (splicing)** 的过程从 RNA 中除去, 以保持外显子和编码序列的连续性。内含子是高等生物独有的特征, 通常不会出现在细菌中。

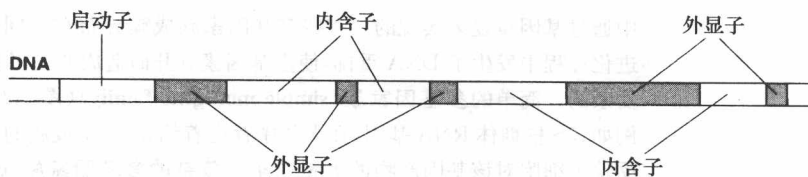


图 A2-4 内含子和外显子

假基因

有一些基因与别的基因相似,但检查它们的碱基序列后会发现存在错误,使其不可能包含有用的生物信息。这些基因被称为假基因(pseudogene)(见 B4),它们源于进化中 DNA 序列出现的错误或突变的基因,导致其生物信息产生混乱,以至于无法指导蛋白质合成。因此,假基因是进化的遗迹。在进化过程中,碱基发生变化导致生物信息丧失之后,会导致随后发生更快的变化,最终导致假基因与原来的基因出现根本的差别,如珠蛋白基因簇中的一些珠蛋白假基因。

A3 遗传密码

要 点

遗传表达

遗传信息由一系列基因的 DNA 碱基序列所编码。基因表达作为一个术语,是用来描述细胞如何将遗传信息解码,以合成细胞活动所必需的蛋白质。基因表达指合成互补的 RNA 链,这条 RNA 的序列决定了蛋白质的氨基酸序列。基因的 DNA 序列与多肽的氨基酸序列呈现线性相关。

遗传密码

20 种氨基酸由 64 种碱基三联体编码,称为密码子。大多数氨基酸具有一种以上的密码子。遗传密码的这种简并性能够减少突变的影响。编码同一个氨基酸的密码子称为同义密码子,它们通常在第三个碱基上有差异,这个碱基位置被称为“摆动”位置。AUG 编码甲硫氨酸,同时也是起始密码子。有三种终止密码子:UAG、UGA 和 UAA。

阅读框

如果可以选择不同位置的碱基开始读码,任何 DNA 序列都可以有三种不同的读码方式。每一种读码方式产生的一套密码子序列称为一个阅读框。起始密码子决定哪一种阅读框真正成为蛋白编码序列,其他阅读框往往包含终止密码子,无法用于指导蛋白质合成。阅读框是一段由起始密码子和终止密码子所限定的范围内的密码子序列。

遗传密码的
通用性

遗传密码普遍适用于所有生物,同样的密码子编码同一个氨基酸。但是,在线粒体基因组和一些单细胞生物中存在一些与标准密码子用法不同的例外。

相关主题

转移 RNA(A5)

翻译(A8)

遗传表达

DNA 携带有生物体复制自身所需的遗传信息,这些遗传信息编码在碱基序列中,并被组织为一系列的基因。基因表达是指细胞对所携带的遗传信息进行解码,并利用这些信息合成可执行各种细胞功能的蛋白质的过程。在基因表达过程中,遗传信息通过合成 RNA 分子从 DNA 拷贝到 RNA,其碱基序列与 DNA 模板互补。随后,RNA 指导蛋白质合成,蛋白质的氨基酸序列由 RNA 碱基序列决定。对每个基因而言,其 DNA 序列与其编码的多肽中的氨基酸序列呈共线性关系,即 DNA 编码链的 5'→3'的碱基顺序决定了其所编码的多肽从氨基端到羧基端的氨基酸序列。

遗传密码

遗传密码用于描述在蛋白质合成过程中如何将碱基序列翻译成不同的氨基酸。将基因的 DNA 序列分为以 3 个碱基为单位的一系列组合,每组的 3 个碱基称为一个密码子,编码一种氨基酸。DNA 和 RNA 的 4 种碱基能够组合形成 $4^3 = 64$ 种密码子,负责编码在蛋白质中发现的 20 种氨基酸(表 A3-1)。由于密码子的种类多于氨基酸的种类,因此,除了甲硫氨酸和色氨酸外,每种氨基酸都由一种以上的密码子编码,这种现象被称为遗传密码的简并性(degeneracy)或冗余(redundancy)。编码同一个氨基酸的密码子称为同义密码子,同义密码子之间往往很相似,如 ACU、ACC、ACA 和 ACG 都编码苏氨酸。同义密码子之间的差别通常发生在第三位碱基上,这个碱基位置被称为摆动位置(wobble position)。遗传密码的简并性可以减少突变的影响,这样可以减少因碱基发生改变而引起氨基酸改变的机会,避免了可能对蛋白质功能产生的有害作用。在 64 种密码子中,有 61 种编码氨基酸,而剩下的三种, UAG、UGA 和 UAA 不编码任何氨基酸,而是蛋白质合成的终止信号,称为终止密码子(termination codon 或 stop codon)。编码甲硫氨酸的 AUG 是蛋白质合成的起始信号,被称为起始密码子(initiation codon)。因此,所有多肽的合成都是以甲硫氨酸开始的,但是在某些情况下,蛋白质合成结束后该甲硫氨酸会被去掉。

表 A3-1 遗传密码

第 1 个位置(5'端)	第 2 个位置				第 3 个位置(3'端)
	U	G	A	G	
U	Phe UUU	Ser UCU	Tyr UAU	Cys UGU	U
	Phe UUC	Ser UCC	Tyr UAC	Cys UGC	G
	Leu UUA	Ser UCA	Stop UAA	Stop UGA	A
	Leu UUG	Ser UCG	Stop UAG*	Trp UGG	G
C	Leu CUU	Pro CCU	His CAU	Arg CGU	U
	Leu CUC	Pro CCC	His CAC	Arg CGC	C
	Leu CUA	Pro CCA	Gln CAA	Arg CGA	A
	Leu CUG	Pro CCG	Gln CAG	Arg CGG	G
A	Ile AUU	Thr ACU	Asn AAU	Ser AGU	U
	Ile AUC	Thr ACC	Asn AAC	Ser AGC	C
	Ile AUA	Thr ACA	Lys AAA	Arg AGA	A
	Met AUG	Thr ACG	Lys AAG	Arg AGG	G
G	Val GUU	Ala GCU	Asp GAU	Gly GGU	U
	Val GUC	Ala GCC	Asp GAC	Gly GGC	C
	Val GUA	Ala GCA	Glu GAA	Gly GGA	A
	Val GUG	Ala GCG	Glu GAG	Gly GGG	G

阅读框

起始密码子不仅标志着蛋白质合成的开始,也决定了 RNA 序列的阅读框。根据选择哪一个碱基作为密码子的起始,任何碱基顺序都有三套可能的密码子序列。在实际的蛋白质合成过程中,通常只有一套阅读框包含有用的信息,其余两种通常因为含有几个终止密码子而使它们不能用于指导蛋白质合成(图 A3-1)。从一个起始密码子开始,以一个终止密码子结束的一段连续的密码子组成一个阅读框(open reading frame, ORF)。这个特征可以用来在基因组测序中识别编码蛋白质的 DNA。

阅读框 1. 5' -AUG ACU AAG AGA UCC GG -3'
Met Thr Lys Arg Ser

阅读框 2. 5' -AUGA CUA AGA GAU CCG G -3'
终止 Leu Arg Asp Pro

阅读框 3. 5' -AUGAC UAA GAG AUC CGG -3'
Asp 终止 Glu Ile Arg

图 A3-1 根据起始密码子的不同位置,每个 DNA 可以有三种不同的阅读框

遗传密码的
通用性

遗传密码从一开始就被认为在所有生物中都是一样的,也就是说,在所有的生物中同样的密码子都编码同样的氨基酸。但是现在已经证实,密码子的编码存在变化,尽管这种现象是罕见的。例如,动物的线粒体有一个含 20 个基因的小的 DNA 基因组,在该基因组中发现了与正常编码不同的密码子。这些变化通常发生在起始密码子和终止密码子中。例如,正常情况下的终止密码子 UGA 在线粒体基因组中可编码色氨酸,而原来编码精氨酸的 AGA 和 AGG 则变成终止密码子,原来编码异亮氨酸的 AUA 则编码甲硫氨酸。人们认为这些改变得以保留下来,是由于线粒体是一个封闭的系统。除了线粒体基因组,在一些单细胞生物体中也发现了非规范使用密码子的一些例子。例如,原来作为终止密码子的 UAA 和 UAG 在某些原生动物中编码谷氨酸。

* 译者注:终止密码子 UAG 也编码硒代半胱氨酸 Secys,且在特定条件下才可翻译。