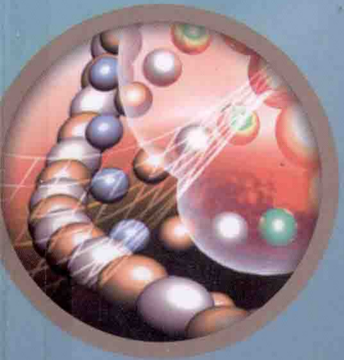
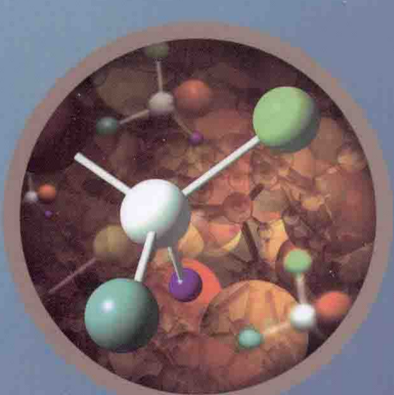
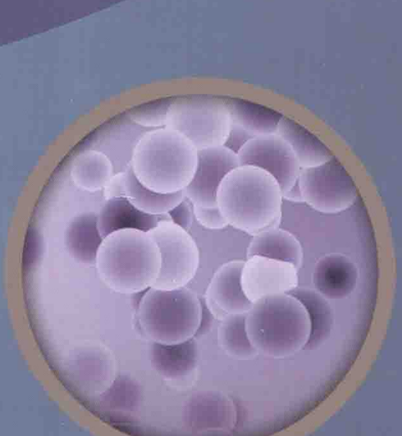
FENZI SHENGWUXUE LILUN
YU FANGFA YANJIU

分子生物学理论
与方法研究

主　编　张来军　孔　芳　马　军
副主编　邢立群　李安华　杨立国　黄先忠　周晓晶





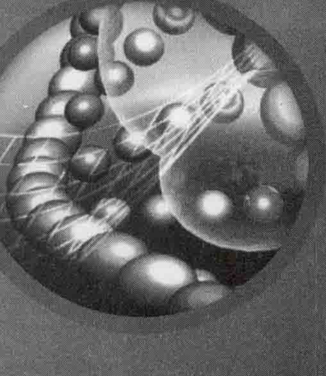
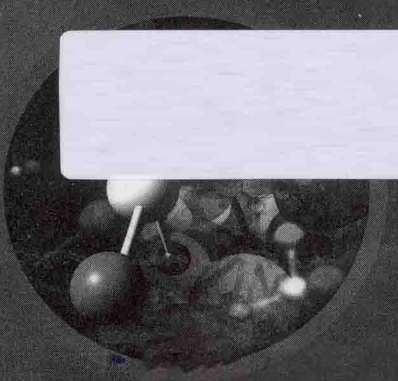
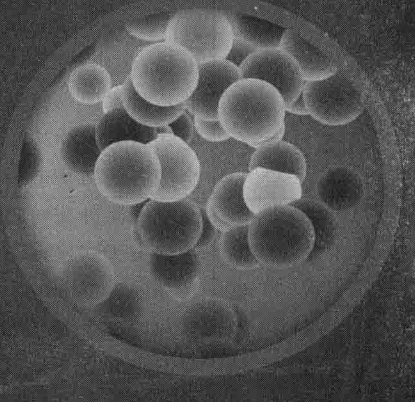


中国水利水电出版社
www.waterpub.com.cn

FENZI SHENGWUXUE LILUN
YU FANGFA YANJIU

分子生物学理论与方法研究

主　编　张来军　孔　芳　马　军
副主编　邢立群　李安华　杨立国　黄先忠　周晓晶

中国水利水电出版社
www.waterpub.com.cn

内 容 提 要

本书主要从分子水平入手，以生物分子学、遗传学等为理论基础，运用基因技术等，分别介绍了分子生物学的起源与发展、概念、内容；遗传物质的分子本质、遗传过程；DNA、RNA 的生物合成；细胞的信号传递；癌分子生物学；原核、真核生物的基因表达以及分子遗传技术等内容。本书可作为生物学、生物分子学、遗传学等领域业内专家及学者的参考资料，也可以作为生物学相关专业学生的理论参考用书。

图书在版编目(CIP)数据

分子生物学理论与方法研究/张来军，孔芳，马军主编. --北京：中国水利水电出版社，2014. 1

ISBN 978-7-5170-1743-1

Ⅰ. ①分… Ⅱ. ①张…②孔…③马… Ⅲ. ①分子生物学 Ⅳ. ①Q7

中国版本图书馆 CIP 数据核字(2014)第 024097 号

策划编辑：杨庆川　责任编辑：杨元泓　封面设计：崔　蕾

书　名	分子生物学理论与方法研究
作　者	主　编　张来军　孔　芳　马　军 副主编　邢立群　李安华　杨立国　黄先忠　周晓晶
出版发行	中国水利水电出版社 (北京市海淀区玉渊潭南路 1 号 D 座 100038) 网址：www. waterpub. com. cn E-mail：mchannel@263. net(万水) sales@waterpub. com. cn 电话：(010)68367658(发行部)、82562819(万水)
经　售	北京科水图书销售中心(零售) 电话：(010)88383994、63202643、68545874 全国各地新华书店和相关出版物销售网点
排　版	北京鑫海胜蓝数码科技有限公司
印　刷	三河市天润建兴印务有限公司
规　格	184mm×260mm　16 开本　24.5 印张　627 千字
版　次	2014 年 6 月第 1 版　2014 年 6 月第 1 次印刷
印　数	0001—3000 册
定　价	86.00 元

凡购买我社图书，如有缺页、倒页、脱页的，本社发行部负责调换

版权所有 · 侵权必究

前　言

分子生物学是研究生物大分子的结构特征和功能及其规律的科学，即在分子水平上揭开生物世界神秘的面纱，阐述生物体在生长、发育、分化过程中各种生物大分子的相互作用、细胞信息传导、基因表达及其调控的机制，从而使人类由被动地响应自然界逐步转向主动地顺应、改造和利用自然界的科学。由于生命体十分复杂和精细，即使是组成生命的一个细胞，也远比一台最庞大、最复杂、最精密的机器复杂和精细得多。因此，在整体水平和细胞水平上研究生物学困难相当大，当我们超越细胞去研究组成细胞的生物大分子的功能时，分子生物学就诞生了。因此，只有在我们充分认识各种生物大分子的功能，认识环境与基因互作、基因之间互作、蛋白质之间互作、信号途径的交叉互作等之后，我们才能更好地在细胞水平和机体水平上研究和认识生命科学。

进入 21 世纪，分子生物学知识获得了快速更新。它已深入到了生命科学的各个分支，尤其是生理学、病理学、发育生物学、神经生物学、免疫学等学科，并极大地推动着这些学科的发展。它的原理和研究方法已广泛应用于医药、农林、食品和工程等各个领域，并取得了许多令人瞩目的成就。分子生物学是当代生物学乃至自然科学中迅速发展的学科之一。它将完全改变生命科学的面貌，也将深刻影响人类的生活和社会发展。

全书共 13 章。第 1 章为绪论，介绍分子生物学的起源与发展、概念和基本内容；第 2 章至第 10 章分别介绍了遗传物质的分子本质、基因与基因组、DNA 的生物合成、RNA 的生物合成、蛋白质的生物合成、原核生物基因表达调控、真核生物基因表达调控、细胞的信号传递和癌分子生物学；第 11 章至第 13 章则主要从实际应用出发讨论了分子遗传技术、分子生物学方法和分子生物学的应用三个方面。本书在编写过程中力求做到概念明确，理论讲述逻辑严密、条理分明，书中内容详尽、便于阅读，深度和广度适宜，文字通俗流畅，言语简练，注重理论联系实际，注重实用，极力贯彻基础性、系统性、科学性等原则，保证逻辑性和系统性。希望本书的出版可以帮助读者获得简明而正确的概念和理解，准确地掌握分子生物学的基本理论和学科前沿知识，了解开展分子生物学理论研究的方法，适当地联系生活实际应用，以增强理解和兴趣，并因此获得思维、逻辑与分析的启迪和创新能力的提升。

在编写过程中参考了一些前人的成果和论述，同时也得到了许多朋友和家人的支持，在此表示衷心的感谢。分子生物学的内容极为丰富且发展迅速，由于篇幅所限，书中不可能一一详细论述探讨，有许多内容也必须精减。水平有限，书中难免有不妥之处，还请读者多批评指正，以使不断完善。

编者

2013 年 12 月

目　　录

第 1 章　绪论 …… 1

1.1　分子生物学的起源与发展 …… 1
1.2　分子生物学的概念 …… 7
1.3　分子生物学内容 …… 8

第 2 章　遗传物质的分子本质 …… 14

2.1　遗传物质 …… 14
2.2　核酸的结构 …… 19
2.3　核酸的变性和复性性质 …… 39
2.4　核酸的研究方法 …… 41
2.5　核酸的序列测定 …… 46

第 3 章　基因与基因组 …… 50

3.1　基因的概念 …… 50
3.2　基因的类型 …… 54
3.3　基因组 …… 57
3.4　基因组学 …… 70

第 4 章　DNA 的生物合成 …… 77

4.1　DNA 复制的基本原理 …… 77
4.2　原核生物 DNA 的复制 …… 85
4.3　真核生物 DNA 的复制过程 …… 86
4.4　病毒 DNA 的复制 …… 88
4.5　DNA 的损伤和修复 …… 91
4.6　DNA 的突变 …… 96
4.7　DNA 的重组和转座 …… 98

第 5 章　RNA 的生物合成 …… 107

5.1　转录的基本原理 …… 107
5.2　转录相关的调控元件 …… 108
5.3　原核生物 RNA 的合成 …… 118
5.4　真核生物 RNA 的合成 …… 127
5.5　RNA 病毒 RNA 的合成 …… 134

此为试读，需要完整PDF请访问：www.ertongbook.com

5.6 原核生物 RNA 的转录后加工 …… 137
5.7 真核生物 RNA 的转录后加工 …… 140
5.8 内含子与外显子 …… 148
5.9 RNA 生物合成的抑制 …… 152

第 6 章 蛋白质的生物合成与功能结构 …… 155

6.1 参与蛋白质合成的物质 …… 155
6.2 原核生物蛋白质的合成 …… 161
6.3 真核生物蛋白质的合成 …… 165
6.4 蛋白质翻译后修饰 …… 169
6.5 蛋白质的靶向转运 …… 172
6.6 蛋白质的结构与功能 …… 176

第 7 章 原核生物基因表达调控 …… 190

7.1 基因表达的组成性、可诱导性、可阻遏性 …… 190
7.2 基因表达的正调控与负调控 …… 191
7.3 操纵子 …… 193
7.4 λ 噬菌体调控级联 …… 204
7.5 基因表达的翻译调控 …… 211
7.6 翻译后调控机制 …… 216

第 8 章 真核生物基因表达调控 …… 218

8.1 DNA 水平调控 …… 218
8.2 真核细胞转录的分子调控 …… 223
8.3 转录后水平调控 …… 227
8.4 翻译水平的调控 …… 239

第 9 章 细胞的信号传递 …… 244

9.1 信号转导基本概念 …… 244
9.2 G-蛋白与跨膜信号传递 …… 247
9.3 酪氨酸蛋白激酶及其相关受体 …… 259
9.4 小分子信号调控 …… 260
9.5 细胞对信号的反应 …… 261

第 10 章 癌分子生物学 …… 263

10.1 癌发生的分子基础 …… 263
10.2 原癌基因和癌基因 …… 264
10.3 肿瘤抑制基因 …… 269
10.4 癌症发生的遗传学说 …… 273

第 11 章　分子遗传技术 …… 277

11.1　用于鉴定、扩增和克隆基因的一些基本技术 …… 277
11.2　构建和筛选 DNA 文库 …… 289
11.3　DNA、RNA 和蛋白质的分子水平分析 …… 292
11.4　基因和染色体在分子水平的分析 …… 297

第 12 章　分子生物学方法 …… 305

12.1　核酸杂交 …… 305
12.2　基因工程技术 …… 309
12.3　DNA 测序和基因组测序 …… 314
12.4　基因表达分析 …… 326
12.5　LPL 内含子突变检测技术 …… 334
12.6　聚合酶链式反应 …… 338

第 13 章　分子生物学的应用 …… 342

13.1　利用 DNA 重组技术鉴别人类基因 …… 342
13.2　基因治疗 …… 349
13.3　人类疾病的分子诊断 …… 369
13.4　DNA 指纹图谱 …… 371
13.5　在细菌中生产真核生物蛋白质 …… 375
13.6　转基因动物和植物 …… 376

参考文献 …… 383

年，Hershey 和 Chase 又用同位素示踪技术证明 T_2 噬菌体感染大肠杆菌。主要是核酸进入细菌内，而病毒外壳蛋白留在细胞外。烟草花叶病毒的重建实验证明，病毒蛋白质的特性由 RNA 决定，即遗传物质是核酸而不是蛋白质。至此，DNA 作为遗传物质才被普遍地接受。

1950 年：Chargaff 以不同来源 DNA 碱基组成的精确数据推翻了四核苷酸论，提出了 Chargaff 规则，即 DNA 的碱基组成有一个共同的规律，胸腺嘧啶的摩尔含量总是等于腺嘌呤的摩尔含量，胞嘧啶的摩尔含量总是等于鸟嘌呤的摩尔含量，即[A]=[T]和[G]=[C]。

1951 年：Pauling 和 Corey 应用 X-射线衍射晶体学理论研究了氨基酸和多肽的精细空间结构，提出了两种有周期规律性的多肽结构学说，即 α-螺旋和 β-叠理论。

1953 年：这是开创生命科学新时代的第一年，具有里程碑意义的是 Watson 和 Crick 发表了“脱氧核糖核酸的结构”的著名论文，他们在 Franklin 和 Wilkins X-射线衍射研究结果的基础上，推导出 DNA 双螺旋结构模式，开创了生物科学的新纪元。同年，Sanger 历经 8 年的研究，完成了第一个蛋白质——胰岛素的氨基酸全序列分析。

随后，1954 年 Gamnow 从理论上研究了遗传密码的编码规律；1956 年 Volkin 和 Astrachan 发现了 mRNA（当时尚未用此名）；1958 年，Hoagland 等发现了 tRNA 在蛋白质合成中的作用；Meselson 和 Stahl 应用同位素和超离心法证明 DNA 的半保留复制；Crick 提出遗传信息传递的中心法则。

1960 年：Marmur 和 Dofy 发现了 DNA 的复性作用，确定了核酸杂交反应的专一性和可靠性；Rich 证明 DNA—RNA 杂交分子与核酸间的信息传递有关，开创了核酸实际应用的先河。与此同时，在蛋白质结构研究方面，Kendrew 等得到了肌红蛋白 0.2 nm 分辨率的结构，Perutz 等得到了血红蛋白 0.55 nm 分辨率的结构。

1961 年：这是分子生物学发展不平凡的一年。Jacob 和 Monod 提出操纵子学说，发表了蛋白质合成中遗传调节机理的论文，此论文被誉为是分子生物学中文笔优美的经典论文之一。同年，Brenner 等获得 mRNA 的证据；Hall 和 Spiegelman 证明 T_2DNA 和 T_2 专一性 RNA 的序列互补；Crick 等证明了遗传密码的通用性。

1962 年：Arber 提出第一个证据，证明限制性核酸内切酶的存在，导致以后对该类酶的纯化，并由 Nathans 和 Smith 应用于 DNA 图谱和序列分析。

1965 年：Holley 等采用重叠法首先测定了酵母丙氨酰-tRNA 的一级结构，为广泛、深入地研究 tRNA 的高级结构奠定了基础。

1967 年：Gellert 发现了 DNA 连接酶，该酶将具有相同粘末端或者平末端的 DNA 片段连接在一起。同年，Philips 及其同事确定了溶菌酶 0.2 nm 分辨率的三维结构。

1970 年：Temin 和 Baltimore 几乎同时发现了反转录酶，证实了 Temin 1964 年提出的“前病毒假说”。在劳氏肉瘤病毒（RSV）感染以后，首先产生的是含有 RNA 病毒基因组全部遗传信息的 DNA 前病毒，子代病毒的 RNA 是以前病毒的 DNA 为模板进行合成的。反转录酶已成为目前分子生物学研究中的一个重要工具。

1972～1973 年：重组 DNA 时代到来。Berg、Boyer 和 Cohen 等创建了 DNA 克隆化技术，在体外构建成具有生物学功能的细菌质粒，开创了基因工程新纪元。与此同时，Singer 和 Nicolson 提出生物膜结构的液态镶嵌模型。

1975 年：Southern 发明了凝胶电泳分离 DNA 片段的印迹法；Gruustein 和 Hogness 建立了克隆特定基因的新方法；O'Farrell 发明了双向电泳分析蛋白质的方法，为分子生物学的深入发

展创造了重要的技术条件；Blobel 等报导了信号肽。

1976 年：Bishop 和 Varmus 发现动物肿瘤病毒的癌基因来源于细胞基因（即原癌基因）。1977 年：Berget 等发现了“断裂”基因；Sanger、Maxam 和 Gilbert 创立了“酶法”“化学法”测定 DNA 序列的方法，标志着分子生物学研究新时代的到来。

1979 年：Solomon 和 Bodmer 最先提出至少 200 个限制性片段长度多态性（RFLP）可作为连接人整个基因组图谱之基础。

1980 年：Wigler 等通过与某个选择性标志物共感染，从而把非选择性基因导入哺乳动物细胞；Cohen 和 Boyer 获得一项克隆技术的美国专利。

1981 年：Cech 等发现四膜虫 26S rRNA 前体的自我剪接作用，随后又证明前体中的居间序列（intervening sequence，IVS）有五种酶的活力。几乎在同时，Altman 从纯化的 RNase P 中，证明催化 tRNA 前体成熟的催化剂是 RNase P 中的 RNA。具有催化作用 RNA（ribozyme）的发现，促进了 RNA 研究的飞速发展。

1982 年：Prusiner 等在感染搔痒病的仓鼠脑中发现了朊病毒（prion）。

1983 年：Herrera-Estrella 等用 Ti 质粒作为转基因载体转化植物细胞获得成功。

1984 年：McGinnis 等发现果蝇、非洲爪蟾等同源异形基因中的同源异形盒（homeobox）的核苷酸序列；Schwartz 和 Cantor 发明了脉冲梯度凝胶电泳法；Simons 和 Kleckner 等发现了反义 RNA。

1985 年：Saiki 等发明了聚合酶链式反应（PCR）；Sinsheimer 首先提出人类基因组图谱制作计划的设想；Smith 等报导了 DNA 测序中应用荧光标记取代同位素标记的方法；Miller 等发现 DNA 结合蛋白的锌指结构。

1986 年：Dryja 等发现成视网膜细胞瘤（Rb）基因是一种抑癌基因；Robin 等采用 X-光晶相学，证实了 DNA 结合蛋白的螺旋—转角—螺旋结构。

1987 年：Mirkin 等在酸性溶液的质粒中发现三链 DNA；Burke 等用酵母人工染色体（YAC）作载体克隆了大片段 DNA；Hoffman 等确定了 Dnchenne 肌肉萎缩病灶的蛋白产物是萎缩素（dystrophin）；Hooper 等和 Kuehn 等分别用胚基细胞进行哺乳动物胚的转基因操作，取得重大进展。

1988 年：Landschalz 等在对 CyC3（细胞色素 C 基因调节蛋白）、癌基因产物（MyC、V-jun、V-fos）和 CBP（CCAAT 盒结合蛋白）的研究过程中，发现了结合区亮氨酸序列的周期性，提出 DNA 结合蛋白的亮氨酸拉链结构模型；同年，Whyfe 等证明癌的发生是癌基因的激活和抑癌基因失活的结果。

1989 年：Greider 等首先在纤毛原生动物中发现了端粒酶（telomerase）是以内源性 RNA 为模板的反转录酶；Hiatt 等首次报导了在植物中亦可产生单克隆抗体。

1990 年：人类基因组计划（HGP）全面正式启动；Simpson 等发现了对 mRNA 前体编辑起指导作用的小分子 RNA（guide RNA）；Sinclair 等在人类 Y 染色体上发现了新的性别决定基因 SRY 基因。

1991 年：由欧洲共同体（EC）组织 17 个国家 35 个实验室的 147 位科学家，以手工测序为主要手段，首先完成了第一条完整染色体（酵母 3 号染色体）的 315 kb 的测序工作；Hake 等首次报导在植物中发现含有同源异形盒基因；Blackburn 等提出调节聚合序列[通式为 $(T/A)_m G_n$，m＝124，n＝1～8]的单链 DNA 可形成分子内或分子间的四螺旋结构，起着稳定染色体的作用。

1993 年：Jurnak 等在研究果胶酸裂解酶时，发现一种新的蛋白质结构——平行 J3 螺旋(parallel phelix)；Yuan 等在哺乳类细胞内发现一种参与调节细胞凋亡并具有剪切作用的蛋白质——IL-1β 转换酶(interlukin-1β-converting enzyme，ICE)。

1994 年：日本科学家在《Nature Genetics》上发表了水稻基因组遗传图；Wilson 等用 3 年时间完成了线虫(C. elegans)3 号染色体连续的 2.2 Mb 的测定，预示着百万碱基规模的 DNA 序列测定时代的到来。

1995 年：Cuenoud 等发现了具有酶活性的 DNA；Tu 等在 *E. coli* 中发现了具有转运与信使双功能的 RNA——10 Sa RNA。

1996 年：Lee 等首次报导了酵母转录因子 GCN4 中的氨基酸片段能自动催化合成自我复制的肽；洪国藩等采用"指纹—锚标"战略构建了高分辨率的水稻基因组物理图谱，DNA 片段的长度为 120 kb；Goffeau 等完成了酵母基因组 DNA 全序列(1.25×10^7 bp)的测定。

1997 年：Wilmut 等首次不经过受精，用成年母羊体细胞的遗传物质，成功地获得克隆羊——多莉(Dolly)；Willard 等首次构建了人染色体(HACs)；Salishury 等发现 DNA 一种新的结构形式——四显性组合，这可能是基因交换期间 DNA 联结的一种方式。

1998 年：Renard 等用体细胞操作获得克隆牛——Marguerife，再次证明从体细胞可克隆出遗传上完全相同的哺乳动物；Gene Bank 公布了最新人的"基因图谱 98"，代表了 30 181 条基因定位的信息；Venter 对人类基因组计划提出新的战略——全基因组随机测序，毛细管电泳测序仪启动。

从以上所述分子生物学的发展中，可以看出 20 世纪以核酸的研究为核心，带动着分子生物学向纵深发展。50 年代的双螺旋结构，60 年代的操纵子学说，70 年代的 DNA 重组，80 年代的 PCR 技术，90 年代的 DNA 测序都具有里程碑的意义，将生命科学带向一个由宏观到微观再到宏观，由分析到综合的时代。

1.1.2 21 世纪分子生物学的发展趋势

原定 2005 年完成人类基因组 DNA 测序的计划，已提前 5 年完成。当前，人类基因组研究的重点正在由"结构"向功能转移，一个以基因组功能研究为主要研究内容的"后基因组"(post-genomics)时代已经到来。它的主要任务是研究细胞全部基因的表达图式和全部蛋白图式，或者说"从基因组到蛋白质组"。于是，分子生物学研究的重点似乎又将回到蛋白质上来，生物信息学也应运而生。随着新世纪的到来，生命科学又将进入这样一个新时代，学习分子生物学的青年学生，应该了解一些本学科特征和发展趋向。

1. 功能基因组学

遗传学最近的定义是，对生物遗传的研究和对基因的研究。功能基因组学(functional genomics)是依附于对 DNA 序列的了解，应用基因组学的知识和工具去了解影响发育和整个生物体的特定序列表达谱。以酿酒酵母(S. cervisiae)为例，它的 16 条染色体的全部序列已于 1996 年完成，基因组全长 12 086 kb，含有 5 885 个可能编码蛋白质的基因，140 个编码 rRNA 基因，40 个编码 snRNA 基因和 275 个 tRNA 基因，共计 6 340 个基因。功能基因组学是进一步研究这 6 000多个基因，在一定条件下，譬如酵母孢子形成期，同时有多少基因协同表达才能完成这一发育过程，这就需要适应这一时期的全套基因表达谱(gene expression pattern)。要解决如此复杂的问题就必须在方法学上有重大的突破，创造出高效快速地同时测定基因组成千上万个基因活

动的方法。目前用于检测分化细胞基因表达谱的方法，有基因表达连续分析法(serial analysis of gene expression，SAGE)、微阵列法(microarray)、有序差异显示(ordered differential display，ODD)和 DNA 芯片(DNA chips)技术等。今后，随着功能基因组学的深入发展，将会有更新更好的方法和技术出现。

功能基因组亦包括了在测序后对基因功能的研究。酵母有许多功能重复的基因，常分布在染色体的两端，当酵母处于丰富培养基条件时，这些基因似乎是多余的，但环境改变时就显示出其功能。基因丰余现象实际上是对环境的适应，丰余基因的存在为进化适应提供了可选择的余地。基因组全序列还保留了基因组进化的遗迹，提示基因重复常发生在近中心粒区和染色体臂中段。

当前，研究者已把酵母基因组作为研究真核生物基因组功能的模式，计划建立酵母基因组 6 000多个基因的单突变体文库(single mutant library)，并可用于其他高等真核生物基因组之"基因功能作图"。

总之，功能基因组学的任务，是对成千上万的基因表达进行分析和比较，从基因组整体水平上阐述基因活动的规律。核心问题是基因组的多样性和进化规律，基因组的表达及其调控，模式生物体基因组研究等。这门新学科的形成，是在后基因组时代生物学家的研究重点从揭示生命的所有遗传信息转移到在整体水平上对生物功能进行研究的重要标志。

2. 蛋白质组学

蛋白质组(proteome)对不少人来说，目前还是一个比较陌生的术语。它是在 1994 年由澳大利亚 Macguarie 大学的 Wilkins 等首先提出的，随后，得到国际生物学界的广泛承认。他们对蛋白质组的定义为："蛋白质组指的是一个基因组所表达的全部蛋白质"(proteome indicates the proteins expressed by a genome)；"proteome"是由蛋白质一词的前几个字母"prote"和基因组一词的后几个字母"ome"拼接而成。

蛋白质组学是以蛋白质组为研究对象，研究细胞内所有蛋白质及其动态变化规律的科学。蛋白质组与基因组不同，基因组基本上是固定不变的，即同一生物不同细胞中基因组基本上是一样的，人类的基因总数约是 6～10 万个。单从 DNA 序列尚不能回答某基因的表达时间、表达量、蛋白质翻译后加工和修饰的情况，以及它们的亚细胞分布等。这些问题可望在蛋白质组研究中找到答案，因为蛋白质组是动态的，有它的时空性、可调节性，进而能够在细胞和生命有机体的整体水平上阐明生命现象的本质和活动规律。蛋白质组研究的数据与基因组数据的整合，亦会对功能基因组的研究发挥重要的作用。

蛋白质组由原定义一个基因组所表达的蛋白质，改为细胞内的全部蛋白质，比较更为全面而准确。但是，要获得如此完整的蛋白质组，在实践中是难以办到的。因为蛋白质的种类和形态总是处在一个新陈代谢的动态过程中，随时发生着变化，难以测准。所以，1997 年，Cordwell 和 Humphery-Smith 提出了功能蛋白质组(functional proteome)的概念，它指的是在特定时间、特定环境和实验条件下基因组活跃表达的蛋白质。与此同时，中国生物科学家提出了功能蛋白质组学(functional protemics)新概念，把研究定位在细胞内与某种功能有关或在某种条件下的一群蛋白质。

功能蛋白质组只是总蛋白质组的一部分，通过对功能蛋白质组的研究，既能阐明某一群体蛋白质的功能，亦能丰富总蛋白质数据库，是从生物大分子(蛋白质、基因)水平到细胞水平研究的重要桥梁环节。

此为试读，需要完整PDF请访问：www.ertongbook.com

无论是蛋白质组学还是功能蛋白质组学，首先都要求分离亚细胞结构、细胞或组织等不同生命结构层次的蛋白质，获得蛋白质谱。为了尽可能分辨细胞或组织内所有蛋白质，目前一般采用高分辨率的双向凝胶电泳。一种正常细胞的双向电泳图谱通过扫描仪扫描并数字化，运用二维分析软件可对数字化的图谱进行各种图像分析，包括分离蛋白在图谱上的定位，分离蛋白的计数、图谱间蛋白质差异表达的检测等。一种细胞或组织的蛋白质组双向电泳图，可得到几千甚至上万种蛋白质，为了适应这种大规模的蛋白质分析，质谱已成为蛋白质鉴定的核心技术。从质谱技术测得完整蛋白质的相对分子质量、肽质谱(或称肽质量指纹，pepetide massfingerprint)以及部分肽序列等数据，通过相应数据库的搜寻来鉴定蛋白质。此外，尚需对蛋白质翻译后修饰的类型和程度进行分析。在蛋白质组定性和定量分析的基础上建立蛋白质组数据库。

从提出蛋白质组的概念到现在短短几年后的 1997 年构建成第一个完整的蛋白质组数据库——酵母蛋白质数据库(yeast protein database，YPD)，进展速度极快，新的思路和技术不断涌现，蛋白质组学这门新兴学科，在今后的实践中将会不断完善，充实壮大，发展成为后基因组时代的带头学科。

3. 生物信息学

HGP 大量序列信息的积累，导致了生物信息学(Bioinformatics)这门全新的学科的产生，对 DNA 和蛋白质序列资料中各种类型信息进行识别、存储、分析、模拟和转输。它常由数据库、计算机网络和应用软件三大部分组成。

国际上现有 4 个大的生物信息中心，即美国生物工程信息中心(GenBank)和基因组序列数据库(GSDB)，欧洲分子生物学研究所(EMBL)和日本 DNA 数据库(DDBJ)。这些中心和全球的基因组研究实验室通过网站、电子邮件或者直接与服务器和数据库联系而获得的搜寻系统，使得研究者可以在多种不同的分析系统中对序列数据进行查询，利用和共享巨大的生物信息资源。

随着 DNA 大规模自动测序的迅猛发展，序列数据爆炸性地积累，HGP 正式启动之时，就与信息科学和数据库技术同步发展，收集、存储、处理了庞大的数据，生物信息学逐步走向成熟，在基因组计划中发挥了不可取代的作用。建立的核苷酸数据库，已存有数百种生物的 cDNA 和基因组 DNA 序列的信息。在已应用的软件中，有 DNA 分析、基因图谱构建、RNA 分析、多序列比较、同源序列检索、三维结构观察与演示、进化树生成与分析等。

在蛋白质组计划中，由于蛋白质组随发育阶段和所处环境而变化，mRNA 丰度与蛋白质的丰度不是显著相关，以及需要经受翻译后的修饰，因而对蛋白质的生物信息学研究，在内容上有许多特殊之处。现在建立的数据库，有蛋白质序列、蛋白质域、二维电泳、三维结构、翻译后修饰、代谢及相互作用等。而通用的软件，主要包括蛋白质质量＋蛋白质序列标记、模拟酶解、翻译后修饰等。

当今的潮流是利用生物信息学研究基因产物——蛋白质的性质并估计基因的功能。传统的基因组分析是利用一系列方法来得到连续的 DNA 序列的信息，而蛋白质组连续系 (proteomic cortigs)则源于多重相对分子质量和等电范围，由此来构建活细胞内全部蛋白质表达的图像。氨基酸序列与其基因的 DNA 序列将被联系在一起，最终与蛋白质组联系在一起，从而允许人们研究不同条件下的细胞和组织。

1.2 分子生物学的概念

分子生物学(molecular biology)是在分子水平上研究生命的重要物质(注重核酸、蛋白质等生物大分子)的化学与物理结构、生理功能及其结构与功能的相关性,揭示复杂生命现象本质的一门现代生物学。它是定量地阐明生物学规律(遗传进化规律、分化发育规律、生长衰老规律等),透过生命现象揭示生命本质的一门学科,也是当今生命科学中最具活力的一门学科。

从广义的角度来说,生命体中一切相关物质的结构、功能、变化及其规律都是分子生物学研究的内容,如蛋白质的结构、运动和功能及生物催化剂的作用机制和动力学,膜蛋白结构功能和膜运输,核酸的结构和功能研究等。总之,生物化学中所涉及的一切大大小小的生物组成成分及其各种物质的分子结构、代谢过程及作用机制都属于分子水平的生物学研究内容。

从狭义概念来描述分子生物学这一学科的定义与研究范畴,则是偏重于生物大分子——核酸(或基因),主要研究脱氧核糖核酸(deoxyribonucleic acid,DNA)的复制、转录、翻译和基因表达调控的过程。同时涉及与重要调控过程有关的蛋白质和非编码核糖核酸(noncoding RNA,ncRNA)结构与功能的分子生物学研究。尤其是,近年来连续五年被美国 Science 杂志评为年度科技十大突破之一的核糖核酸(ribonucleic acid,RNA)方面的研究,包括小干扰 RNA (small interference RNA,siRNA)和微小 RNA(microRNA,miRNA)在内的小分子 RNA 对生物分化、发育、细胞周期、凋亡、印迹、应激等的调控功能研究,以及核酶(ribozyme)催化蛋白质生物合成等功能的研究。2000 年,RNA 组学(RNomics)正式提出,受到全球科学界的关注。基因组学(genomics)、核糖核酸组学(RNomics)、蛋白质组学(proteomics)和代谢组学 (smetabolismics)已成为以功能基因组学为核心的、不可分割的分子生物学内容,组成了系统生物学(system biology),在后基因组时代为破解生命之谜作出贡献。

20 世纪初,人们重新发现与证实了孟德尔遗传定律,从此,生物学沿着正确的基本原理与法则突飞猛进。同时,与之相关的化学、数学、物理学与计算机等基础学科的理论、技术及其各项新成果向生命科学不断渗透,推动了生物大分子结构与功能的研究。核酸、蛋白质、生物催化剂(蛋白酶、核酶)、多糖等大分子物质的分子结构、理化性质、生理功能、作用机制以及结构与功能之间的关系等方面研究都有大量的文献资料积累和重大的理论与技术突破。尤其是随着核酸化学研究的进展,揭示出核酸化学的许多规律之后,1953 年 Watson 和 Crick 共同提出了脱氧核糖核酸的双螺旋模型。这个模型的建立为揭开遗传信息的复制和转录奠定了基础,也是分子生物学学科形成的奠基石。

生物性状遗传信息的传递方式是分子生物学的核心问题。1958 年,Crick 提出了生物遗传的"中心法则",认为生物性状遗传信息是以单向不可逆的方式传递的,即 DNA(自我复制)RNA→Protein 的单向不可逆的生物信息传递方式。其核心说明基因是连续的 DNA 序列,生命世界是 DNA-蛋白质的世界。1970 年,Temin 和 Baltimore 在 RNA 肿瘤病毒中发现了 RNA 逆转录酶,病毒 RNA 分子通过其编码的逆转录酶将病毒 RNA 分子转换成为与其互补的单链 DNA。这一发现使单向不可逆方式传递的中心法则受到挑战。为此,1971 年,Crick 对他自己所提出的"中心法则"作了补充。与此同时,人们已注意到蛋白质生物合成过程中新生肽的活性形成问题,如它们是如何通过自身内在的信息及其周围的微环境(如分子伴侣、折叠酶等)的相互作用而产生具有完全生物活性的蛋白质的?是否存在多肽折叠的密码?线性多肽折叠形成具有特定三维

空间结构活性蛋白的机制是什么？只有这些相关研究的突破，才能深入了解蛋白质空间结构的形成与其功能表达之间的关系，才能最终完整地表述生物性状遗传信息传递的全过程。

20 世纪 70 年代末，Sharp 和 Robert 发现了断裂基因。1981 年，Cech 发现了核酶。上述修改了的“中心法则”再次受到挑战。尤其是 RNA 研究的新发现不断挑战“中心法则”的定义。一个由 RNA 组成的调控 DNA 遗传信息的网络已经显露出来。在 DNA 一蛋白质的生命世界中，古老的 RNA 不仅不是“垃圾”，而且仍然占有一席之地，与 DNA、蛋白质共同构成生命世界生物遗传的“中心法则“将不断得以补充与完善。

围绕着生物遗传“中心法则”研究的不断深入与突破，核酸的分子生物学得到了异乎寻常的迅速发展，从而确立了分子生物学在生命科学中的地位与价值。

1.3 分子生物学内容

分子生物学的研究对象主要是核酸和蛋白质这两种生物大分子。目前普遍认为，在所有生物中，①构成生物大分子的单体都是相同的，它们具有共同的核酸语言和共同的蛋白质语言；②生物体内一切有机大分子的建成都遵循共同的规则；③生物大分子单体（核苷酸、氨基酸）组成和排列方式的不同是产生功能差异的基础，不同的生物大分子之间的互作是造成物种特性差异的根本原因。

核酸和蛋白质的结构分析及遗传物质和遗传信息传递规律的研究对分子生物学学科的发展起到了巨大的推动作用，在这些研究的基础上，分子生物学在基因组学和功能基因组学、基因的表达和调控、蛋白质组学（proteome）和基因工程（DNA 重组技术）等方面取得了非常卓越的成就。

1.3.1 结构分子生物学

生物大分子，特别是蛋白质和核酸结构和功能的研究，是分子生物学在分子水平上研究生命现象本质的基础。所谓的分子水平，指的是那些携带遗传信息的核酸和在遗传信息传递及细胞内、细胞间通讯过程中发挥着重要作用的蛋白质等生物大分子。这些生物大分子均具有较大的分子质量，由简单的小分子核苷酸或氨基酸排列组合以蕴藏各种信息，并且具有复杂的空间结构以形成精确的相互作用系统，由此构成生物的多样化和生物个体精确的生长发育和代谢调节控制系统。阐明这些复杂的结构及结构与功能的关系是分子生物学的主要任务。

要了解一种生物大分子的功能，通常要先研究其结构。例如，对 DNA 的结构的研究使认识基因突变（gene mutation）、染色体复制（chromosome replication）和遗传重组（genetic recombination）成为可能；tRNA 分子的三维结构的研究解释了 DNA 储存的遗传信息如何翻译到蛋白质的氨基酸顺序中；对初始转录本 mRNA 和成熟 mRNA 排列顺序的比较让我们认识到 RNA 加工在基因表达过程中的重要性；对 DNA 结合蛋白的激活结构域和 DNA 结合结构域的诠释展现了生物大分子间相互作用的主要方式。

结构分子生物学的任务是通过阐明生物大分子的三维结构来解释细胞的生理功能。在蛋白质结构分析方面，1951 年 L. C. Pauling 等提出的 α 螺旋结构描述了蛋白质分子中肽链的一种构象；1955 年 F. Sanger 完成了胰岛素的氨基酸序列的测定；1957 年和 1959 年 J. C. Kendrew 和 M. F. Perutz 在 X 射线分析中分别应用重原子同晶置换技术和计算机技术阐明了鲸肌红蛋白和

马血红蛋白的立体结构;1965 年中国科学家合成了有生物活性的胰岛素,首先实现了蛋白质的人工合成。在核酸结构分析方面,1944 年 O. T. Avery 等研究细菌中的转化现象,证明了 DNA 是遗传物质;1953 年 J. D. Watson 和 F. H. C. Crick 提出了 DNA 的双螺旋结构,开辟了分子生物学研究的新纪元;1961 年 F. Jacob 和 J. L. Monod 提出了操纵子的概念,解释了原核基因表达的调控。到 20 世纪 60 年代中期,关于 DNA 自我复制、RNA 转录和蛋白质合成的一般性质已基本清楚,遗传和变异的规律也随之逐渐明朗。

1.3.2 遗传信息的传递规律

遗传物质可以是 DNA,也可以是 RNA。细胞的遗传物质都是 DNA,只有一些病毒和亚病毒的遗传物质是 RNA。以 DNA 为模板按照碱基互补配对的原则合成 DNA,以及以 RNA 为模板合成 RNA,都称为复制(replication)。以 DNA 为模板按照碱基互补配对的原则合成 RNA,称为转录(transcription)。转录产生的并不是成熟的 mRNA,而只是成熟 mRNA 的前体,或称为初始转录本 mRNA,它们需要进行一定的剪切和连接才能成为成熟的 mRNA,这个过程称为 RNA 加工(RNA processing)。以成熟的 mRNA 为模板在核糖体上进行蛋白质多肽链的合成,称为翻译(translation)。翻译产生的只是蛋白质前体,它需要加工、修饰、折叠和组装后,转运到适当的位置后才能发挥作用。这种从 DNA 到 RNA,再到蛋白质的遗传信息传递方向,叫做直线形中心法则(central dogma)(图 1-1A)。

以 RNA 为遗传物质的病毒称为逆转录病毒(retrovirus),在这种病毒的感染周期中,单链的 RNA 分子在逆转录酶(reverse transcriptase)的作用下,可以逆转录成单链的 DNA,然后再以单链的 DNA 为模板生成双链 DNA。在逆转录酶催化下,RNA 分子产生与其序列互补的 DNA 分子,称为互补 DNA(complementary DNA,cDNA),这个过程即为逆转录(reverse transcription)。由此可见,遗传信息并不一定是从 DNA 单向地流向 RNA,RNA 携带的遗传信息同样也可以流向 DNA。但是 DNA 和 RNA 中包含的遗传信息只是单向地流向蛋白质,迄今为止还没有发现蛋白质的信息逆向地流向核酸。另外,DNA 和 RNA 在复制、转录和翻译过程中都需要蛋白质的参与和调节,因此,蛋白质在遗传信息传递过程中起着非常重要的调控作用。这种 DNA、mRNA 和蛋白质间复杂的相互作用类似一个三角形,它是对直线形中心法则的必要补充,可形象地称之为三角形中心法则(图 1-1B)。

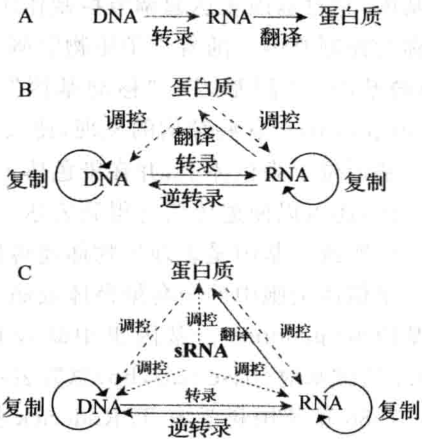


图 1-1 遗传信息的传递

A. 直线中心法则;B. 三角形中心法则;C. 圆锥形中心法则

随着对 RNA 种类的不断发现,现代分子生物学家对 RNA,尤其是小分子 RNA(small RNA,sRNA)或非编码 RNA(non-coding RNA,ncRNA)的功能有了更深入的认识。许多小分子细胞核 RNA(snRNA)与 RNA 加工有关,一些小分子细胞核仁 RNA(snoRNA)参与 rRNA 的合成,19~22 nt 的微小 RNA(miRNA)参与 DNA 和 RNA 的修饰、mRNA 的稳定性、蛋白质的合成。因此,有的科学家认为在遗传信息传递过程中,小分子 RNA 起着中心的调控作用,它们对 DNA、RNA 和蛋白质的结构与功能都有着非常重要的影响。因此,这种 DNA、mRNA、蛋

白质和 sRNA 间的复杂关系，可称为圆锥形中心法则(图 1-1C)。

虽然中心法则对遗传信息的传递方向进行了系统的概括，但还是存在一些特别现象曾对中心法则提出严重的挑战，如朊病毒的发现。朊病毒是一种蛋白质传染颗粒(proteinaceous infectious particle)，它是羊瘙痒病、疯牛病和人类的库鲁病(I～uru disease)和克一杰氏综合征(Creutzfeldt-Jacob disease，CJD)的病原体，能在寄主中传播，并在受感染的宿主细胞内产生与自身相同的分子，且实现相同的生物学功能，这意味着这种蛋白质分子也是负载和传递遗传信息的物质。但更深入的研究表明，朊病毒只是由基因编码产生的一种正常蛋白质的异构体，它进入宿主细胞后并不是自我复制，而是将细胞内基因编码产生的 PrP 蛋白(prion related protein)由正常的 PrPc 异构体转变成致病的 prpsc 异构体。因此，朊病毒并不是遗传物质。当然，不依赖核糖体的非核糖体肽合成酶(NRPS)和 RNA 编辑(RNA editing)的发现，使人们认识到遗传信息的传递规律还有待进一步的完善和发展。

1.3.3 基因、基因组和蛋白质组

基因(gene)是 DNA 分子中含有特定遗传信息的一段核苷酸序列，它包含合成一种功能蛋白或 RNA 分子所必需的全部 DNA 序列。根据其是否具有转录和翻译功能，基因可分为三类：①编码蛋白质的基因，具有转录和翻译功能，包括编码酶和结构蛋白的结构基因及编码阻遏蛋白的调节基因；②只有转录功能而没有翻译功能的基因，包括 tRNA 基因和 rRNA 基因；③不转录的基因，它对基因表达起调节控制作用，包括启动基因和操纵基因，启动基因和操纵基因有时被统称为控制基因。随着分子生物学研究的深入，人们又发现在基因结构中存在有“断裂基因”、“重叠基因”、“假基因”、“移动基因”等。移动基因，又称跳跃基因(jumping gene)或转座子(transposon)。这些结构的发现，使人们对基因的功能有了更深入的理解。

基因位于染色体上，并在染色体上呈线性排列。基因不仅可以通过复制把遗传信息传递给下一代，还可以使遗传信息得到表达。不同个体之间在形态、发育和功能等方面的不同，都是基因差异所致。基因是表现生物体遗传性状的物质基础。

单倍体细胞中的全套染色体或病毒粒子所含的全部 DNA 分子或 RNA 分子，称为该生物体的基因组(genome)。基因组中既含有编码序列，也含有非编码序列。基因组的大小用全部 DNA 的碱基对(base pair，bp)总数表示。

1986 年美国科学家 T. Roderick 提出了基因组学(genomics)，指对所有基因进行基因组作图、核苷酸序列分析、基因定位和基因功能分析的一门科学。基因组研究主要包括以全基因组测序为目标的结构基因组学(structural genomics)、以基因功能鉴定为目标的功能基因组学(functional genomics)或后基因组(postgenome)、研究和利用模式生物基因组测序产生的大量基因组信息进行基因结构和功能分析的比较基因组学(comparative genomics)。

结构基因组学是一门通过基因作图、核苷酸序列分析确定基因组成、基因定位的科学。遗传信息在染色体上，但染色体不能直接用来测序，必须将基因组这一巨大的研究对象进行分解，使之成为较易操作的小的结构区域，这个过程就是基因作图(gene maping)。根据使用的标志和手段不同，基因作图主要有三种类型：①遗传连锁图——通过遗传重组所得到的基因在具体染色体上的线性排列图谱。利用遗传标志之间的重组频率，确定它们的相对距离，一般用厘摩(cM，即每次减数分裂的重组频率为 1%)来表示，遗传标志有 RFLP(限制性酶切片段长度多态性)、RAPD(随机引物扩增多态性 DNA)、AFLP(扩增片段长度多态性)、STR(短串联重复序列，又称

微卫星)和 SNP(单个核苷酸的多态性)等。②物理图谱——利用限制性内切核酸酶处理染色体,根据重叠序列确定片段间连接顺序的图谱。利用遗传标志之间物理距离[碱基对(bp)、千碱基(kb)或兆碱基(Mb)]确定图距,遗传标志主要采用序列标签位点(sequence tagsite,STS),染色体定位明确且可用 PCR 扩增的单拷贝序列]。③转录图谱——利用表达序列标签(expressed sequence tag,EST)作为标记所构建的分子遗传图谱。mRNA 或 cDNA 的 5′或 3′端序列称为表达序列标签,一般长 300～500 bp。

功能基因组学是利用结构基因组所提供的信息和产物,发展和应用新的实验手段,通过在基因组或系统水平上全面分析基因的功能,使得生物学研究从对单一基因或蛋白质的研究转向多个基因或蛋白质同时进行系统研究的学科。研究内容包括基因功能发现、基因表达分析及突变检测。采用的技术包括减法杂交、差示筛选、cDNA 代表差异分析、mRNA 差异显示等传统的分析技术和基因表达的系统分析、cDNA 微阵列、DNA 芯片等新型技术。

比较基因组学(comparative genomics)是在模式生物的基因组图谱和测序基础上,对已知的基因和基因组结构进行比较,来了解基因的功能、表达机理和物种进化的学科。它利用模式生物基因组与人类基因组之间编码顺序上和结构上的同源性,克隆人类疾病基因,揭示基因功能和疾病分子机制,阐明物种进化关系及基因组的内在结构。

基因是遗传信息的携带者,但全部生物功能的执行者却是蛋白质,因此仅从基因的角度来研究是远远不够的,必须研究由基因转录和翻译出蛋白质的过程,才能真正揭示生命的活动规律。蛋白质组学(proteomics)就是研究细胞内蛋白质组成及其活动规律的新兴学科。蛋白质组(proteome)是指全部基因表达的全部蛋白质及其存在方式,是一个基因、一个细胞或组织所表达的全部蛋白质成分。蛋白质组学对不同时间和空间发挥功能的特定蛋白质群体进行研究,从蛋白质水平上探索蛋白质作用模式、功能机理、调节控制及蛋白质群体内相互作用,为临床诊断、病理研究、药物筛选、药物开发、新陈代谢途径等提供理论依据和基础。但由于蛋白质具有多样性、可变性、复杂性、低表达蛋白质难以检测等特点,蛋白质组研究中要求的“全部的蛋白质成分”非常不容易达到。双向聚丙烯酰胺凝胶电泳(two-dimensional polyacrylamide gel electrophoresis, 2D PAGE)、质谱鉴定、计算机图像数据处理与蛋白质数据库是目前蛋白质组学研究的主要工具。

1.3.4 基因的表达和控制

典型的基因表达(gene expression)是指细胞在生命过程中,把储存在 DNA 序列中的遗传信息经过转录和翻译,转变成具有生物活性的蛋白质分子。生物体内的各种功能蛋白质和酶都是利用相应的结构基因编码的。rRNA、tRNA 或 microRNA 等非编码 RNA(ncRNA)的基因经转录和转录后加工产生成熟的 ncRNA,也是 ncRNA 的基因表达。

生物基因组的遗传信息并不是同时全部都表达出来的,一般情况下,只有 5%～10%的基因在高水平转录状态,部分基因处于较低水平的表达,多数基因处在沉默状态,即使蛋白质合成量比较多、基因开放比例较高的肝细胞,一般也只有不超过 20%的基因处于表达状态。

生物个体的各种组织细胞都含有个体发育、生存和繁殖的全部遗传信息,但这些遗传信息的表达是受到严格调控的,通常各组织细胞只合成其自身结构和功能所需要的蛋白质。不同组织细胞中不仅表达的基因数量不相同,而且基因表达的强度和种类也各不相同,这就是基因表达的组织特异性(tissue specificity)和细胞的差别基因表达(differential gene expression)。如果基因

此为试读,需要完整PDF请访问: www.ertongbook.com