

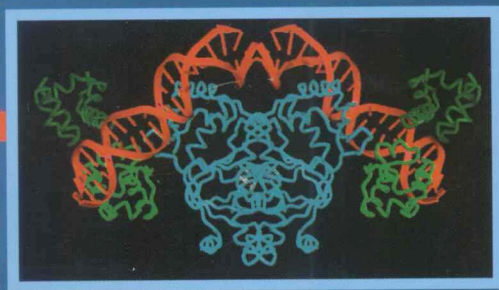


“十二五”普通高等教育本科国家级规划教材

# 分子生物学精要

赵亚华 编著

*Essentials  
of Molecular  
Biology*



科学出版社

“十二五”普通高等教育本科国家级规划教材

# 分子生物学精要

赵亚华 编著

科学出版社

## 内 容 简 介

本书从分子生物学的定义出发,以 DNA 和 RNA 这两类生物大分子为主线,由浅入深地介绍了这些大分子及基因与基因组的结构与功能,DNA 和 RNA 的复制、转录和转录后的剪接、编辑,原核细胞与真核细胞基因表达调控等内容,既全面系统地阐述了分子生物学的基本理论,又列举了学科发展的前沿进展。本书的特色是按照叙述的顺序在每个重要知识点后都附有小结框,便于教师在教学中掌握要点,学生在学习中抓住重点、强化记忆,有利于复习考试。

本书可作为综合性大学、医学院校、师范院校和农林院校生命科学本科生、研究生的分子生物学教材,也可作为相关专业教师、研究人员等的参考书。

### 图书在版编目(CIP)数据

分子生物学精要/赵亚华编著. —北京:科学出版社,2013

“十二五”普通高等教育本科国家级规划教材

ISBN 978-7-03-037759-3

I. ①分… II. ①赵… III. ①分子生物学-高等学校-教材 IV. ①Q7

中国版本图书馆 CIP 数据核字(2013)第 123976 号

责任编辑:席 慧 杨晓庆 郝晨扬/责任校对:桂伟利

责任印制:阎 磊/封面设计:迷底书装

**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>

**骏杰印刷厂印刷**

科学出版社发行 各地新华书店经销

\*

2013年6月第一版 开本:787×1092 1/16

2013年6月第一次印刷 印张:21 3/4

字数:558 000

定价:39.00元

(如有印装质量问题,我社负责调换)

**版权所有,侵权必究**

## 前 言

在生命科学教学、研究与人才培养的全部过程中,分子生物学是十分重要的专业基础课之一。随着相关学科理论与技术研究的不断深入,分子生物学的研究领域不断扩大,学科进展也非常迅速。因此,全面系统地掌握分子生物学的知识非一日之功,而配备一本好的教科书,使教师和学生能够在有限的教学时间内学习这门课程的核心知识,无疑是非常重要的。

本书的编写工作从2003年开始,经历了10年。随着学科的发展,曾在2004、2006、2011年分别出版了第一版、第二版和第三版。2004年编写出版的《分子生物学教程》第一版,随着学科的发展,原书的内容已不能满足教学和参考的需要。作者于2006年在第一版基础上参考了国内外的一些优秀教科书,博采众长,并根据平时的教学积累和学生用书需求,撰写了第二版《基础分子生物学教程》。第二版出版后,在经历了6年教学实践的基础上,作者继续参阅近年来国内外同类教材和书籍,仔细比较,汲取精华,增加了较多的量化描述,枚举了很多国际上的经典实例,对许多图表做了细致明确的标示和修改。在此基础上,全面修订改版形成了第三版。在第三版中对原来的章节做了许多必要的充实并加入了大量实例,如增加了对非编码RNA的重要性认识,核糖体开关、RNA干涉和microRNA的研究进展,两种新的碱基切除修复机制,几个选择性剪接的重要例子,转录及转录激活因子的调控等内容。

为了使第三版《分子生物学教程》在教与学的过程中能够更加突出和浓缩重点,使教师和学生掌握分子生物学的精华,强化对经典理论和实例的理解,以及对重要知识点的记忆,作者在第三版的基础上,删减了一些难度较大且抽象的内容和例子,如噬菌体基因的表达调控、甾体激素对基因转录的调控,以及蛋白质生物合成一章(与生物化学有部分交叉)等,并且在每章前面增加了导读和重要知识点说明,章后增加了关键术语,从而形成了本书。

分子生物学的内容十分丰富而繁杂,本书的特点是将最基本且重要的科学内容汇集于一体,取材新颖,条理清晰,语言精练,概念陈述准确,知识点全面,信息量大,实例丰富,具有很强的可读性、可参考性和适用性,不啻为指导学生快速掌握分子生物学基础知识的理想教材。由于分子生物学发展迅速,资料浩瀚,且国内外从事这方面工作的专家学者众多,关注的侧重点各有不同,本书难以面面俱到,疏漏和不足之处在所难免,敬请广大读者批评指正,作者不胜感激。

本书的出版得到了教育部高等教育司,科学出版社,华南农业大学教务处、生命科学学院,以及许多同行的大力支持,作者在此一并表示衷心的感谢!

作 者

2012年12月于广州五山

|                                       |    |
|---------------------------------------|----|
| 前言                                    |    |
| <b>第 1 章 绪论</b> .....                 | 1  |
| 1.1 分子生物学的概念 .....                    | 1  |
| 1.2 分子生物学研究的主要内容 .....                | 2  |
| 1.2.1 基因与基因组的结构与功能 .....              | 2  |
| 1.2.2 DNA 的复制、转录和翻译 .....             | 2  |
| 1.2.3 基因表达调控的研究 .....                 | 2  |
| 1.2.4 DNA 重组技术 .....                  | 3  |
| 1.2.5 结构分子生物学 .....                   | 3  |
| 1.3 分子生物学与生物化学之间的关系 .....             | 4  |
| 1.4 分子生物学发展的历程 .....                  | 4  |
| 1.4.1 人类对 DNA 和遗传信息传递的认识阶段 .....      | 4  |
| 1.4.2 重组 DNA 技术的建立和发展阶段 .....         | 5  |
| 1.4.3 重组 DNA 技术的应用和分子生物学的迅猛发展阶段 ..... | 5  |
| 1.4.4 对功能基因与基因组的研究阶段 .....            | 7  |
| 1.5 21 世纪分子生物学发展的趋向 .....             | 8  |
| 1.5.1 功能基因组学 .....                    | 8  |
| 1.5.2 蛋白质组学 .....                     | 9  |
| 1.5.3 生物信息学 .....                     | 10 |
| 1.6 怎样学好分子生物学 .....                   | 11 |
| 关键术语 .....                            | 12 |
| 思考题 .....                             | 12 |
| <b>第 2 章 核酸的结构与功能</b> .....           | 13 |
| 2.1 细胞内的遗传物质 .....                    | 13 |
| 2.1.1 DNA 是主要的遗传物质 .....              | 13 |
| 2.1.2 RNA 也是遗传物质 .....                | 14 |
| 2.2 核酸的化学组成与共价结构 .....                | 14 |
| 2.2.1 核酸的化学组成 .....                   | 14 |
| 2.2.2 多聚核苷酸的结构 .....                  | 16 |
| 2.3 DNA 的高级结构 .....                   | 18 |
| 2.3.1 双螺旋模型的特征 .....                  | 18 |
| 2.3.2 DNA 高级结构的其他形式 .....             | 19 |
| 2.4 真核生物的染色体及其组装 .....                | 25 |
| 2.4.1 真核生物的染色体 .....                  | 25 |
| 2.4.2 染色体中的蛋白质 .....                  | 26 |
| 2.4.3 核小体的形成 .....                    | 27 |
| 2.4.4 染色质的高级结构 .....                  | 28 |
| 2.5 RNA 的结构与功能 .....                  | 30 |
| 2.5.1 RNA 的结构特点及与 DNA 的区别 .....       | 31 |
| 2.5.2 RNA 在细胞中的分布 .....               | 31 |
| 2.5.3 细胞中 RNA 分类概述 .....              | 32 |
| 2.6 核酸的变性、复性与分子杂交 .....               | 37 |
| 2.6.1 核酸的变性 .....                     | 37 |
| 2.6.2 核酸的复性与分子杂交 .....                | 39 |
| 关键术语 .....                            | 42 |
| 思考题 .....                             | 42 |
| <b>第 3 章 基因与基因组的结构和功能</b> .....       | 43 |
| 3.1 基因的概念 .....                       | 43 |
| 3.1.1 基因与 DNA 的关系 .....               | 44 |
| 3.1.2 基因与多肽链的关系 .....                 | 45 |
| 3.2 基因的命名 .....                       | 46 |
| 3.3 基因组 .....                         | 47 |
| 3.3.1 基因组的概念 .....                    | 47 |
| 3.3.2 基因及基因组的大小与 C 值悖理 .....          | 48 |
| 3.4 病毒及其基因组 .....                     | 49 |
| 3.4.1 病毒基因组的一般特点 .....                | 49 |
| 3.4.2 病毒的核酸 .....                     | 50 |
| 3.4.3 噬菌体基因组 .....                    | 51 |
| 3.4.4 逆转录病毒及其基因组 .....                | 53 |

|                                   |     |                            |     |
|-----------------------------------|-----|----------------------------|-----|
| 3.5 细菌基因组 .....                   | 59  | 关键词 .....                  | 115 |
| 3.5.1 细菌基因组的一般特点 .....            | 59  | 思考题 .....                  | 115 |
| 3.5.2 细菌的染色体基因组 .....             | 59  | <b>第5章 DNA的损伤、修复和基因突变</b>  |     |
| 3.6 真核生物基因组 .....                 | 61  | .....                      | 116 |
| 3.6.1 真核生物基因组的特点 .....            | 61  | 5.1 DNA的损伤 .....           | 116 |
| 3.6.2 真核生物基因组结构 .....             | 61  | 5.1.1 DNA分子的自发性损伤 .....    | 116 |
| 3.6.3 真核生物基因组的序列异质性 .....         | 66  | 5.1.2 物理因素引起的DNA损伤 .....   | 118 |
| 3.6.4 线粒体基因与基因组结构 .....           | 75  | 5.1.3 化学因素引起的DNA损伤 .....   | 118 |
| 3.6.5 叶绿体基因与基因组结构与功能 .....        | 76  | 5.2 DNA的修复 .....           | 119 |
| 3.6.6 人类基因组简介 .....               | 78  | 5.2.1 直接修复 .....           | 119 |
| 关键词 .....                         | 86  | 5.2.2 切除修复 .....           | 120 |
| 思考题 .....                         | 86  | 5.2.3 错配修复 .....           | 122 |
| <b>第4章 DNA的复制</b> .....           | 87  | 5.2.4 重组修复 .....           | 123 |
| 4.1 DNA复制概述 .....                 | 87  | 5.2.5 易错修复和SOS反应 .....     | 124 |
| 4.1.1 DNA复制的基本概念 .....            | 89  | 5.3 基因突变 .....             | 126 |
| 4.1.2 复制方向 .....                  | 91  | 5.3.1 基因突变的类型 .....        | 126 |
| 4.1.3 复制方式 .....                  | 92  | 5.3.2 DNA诱变剂 .....         | 127 |
| 4.1.4 DNA复制的酶体系 .....             | 94  | 5.3.3 诱变剂和致癌剂的检测 .....     | 129 |
| 4.1.5 DNA的半不连续复制 .....            | 99  | 5.3.4 基因突变的后果 .....        | 129 |
| 4.1.6 DNA合成的保真性 .....             | 100 | 关键词 .....                  | 130 |
| 4.1.7 DNA拓扑异构酶 .....              | 101 | 思考题 .....                  | 130 |
| 4.2 DNA复制的机制 .....                | 102 | <b>第6章 DNA的重组与转座</b> ..... | 131 |
| 4.2.1 大肠杆菌复制的起始 .....             | 102 | 6.1 同源重组 .....             | 132 |
| 4.2.2 真核生物DNA复制起点的结构 .....        | 103 | 6.1.1 同源重组的分子模型 .....      | 132 |
| 4.2.3 大肠杆菌DNA复制的延伸 .....          | 104 | 6.1.2 同源重组的酶学分子机制 .....    | 134 |
| 4.2.4 复制的终止 .....                 | 107 | 6.1.3 酵母的减数分裂重组 .....      | 137 |
| 4.3 真核生物DNA的复制 .....              | 108 | 6.1.4 异源双链与基因转换 .....      | 138 |
| 4.3.1 真核生物的DNA聚合酶 .....           | 108 | 6.1.5 细菌的基因转移与DNA重组 .....  | 139 |
| 4.3.2 真核生物染色体端粒的复制 .....          | 109 | 6.2 特异位点重组 .....           | 141 |
| 4.4 原核细胞DNA复制的调控 .....            | 112 | 6.3 DNA的转座 .....           | 144 |
| 4.4.1 大肠杆菌染色体DNA的复制调控 .....       | 112 | 6.3.1 转座子的概念 .....         | 144 |
| 4.4.2 ColE1质粒DNA的复制调控 .....       | 113 | 6.3.2 转座子的分类 .....         | 144 |
| 4.4.3 单链DNA噬菌体的复制调控 .....         | 113 | 6.3.3 转座子的转座机制 .....       | 146 |
| 4.4.4 $\lambda$ 噬菌体DNA的复制调控 ..... | 113 | 6.3.4 转座子转座的基本特征 .....     | 149 |
| 4.5 真核生物DNA复制调控概述 .....           | 114 | 6.3.5 转座子转座引起的遗传学效应 .....  | 149 |
|                                   |     | 6.3.6 真核生物的转座子 .....       | 150 |
|                                   |     | 6.4 逆转录转座子 .....           | 154 |
|                                   |     | 6.4.1 逆转录转座子的概念 .....      | 154 |

|                                   |     |                                       |     |
|-----------------------------------|-----|---------------------------------------|-----|
| 6.4.2 逆转座子的结构和作用机制 .....          | 154 | 7.9.3 RNA 聚合酶的抑制物 .....               | 208 |
| 6.4.3 逆转座子的生物学意义 .....            | 157 | 关键术语 .....                            | 208 |
| 关键术语 .....                        | 159 | 思考题 .....                             | 208 |
| 思考题 .....                         | 159 | <b>第 8 章 RNA 转录后的剪接与加工</b> ...        | 209 |
| <b>第 7 章 RNA 的转录</b> .....        | 160 | 8.1 原核生物 RNA 的转录后加工 .....             | 209 |
| 7.1 RNA 转录概述 .....                | 160 | 8.1.1 原核生物 rRNA 前体的加工 .....           | 209 |
| 7.1.1 RNA 转录的一般特点 .....           | 160 | 8.1.2 原核生物 tRNA 前体的加工 .....           | 210 |
| 7.1.2 原核生物和真核生物基因转录的<br>差异 .....  | 162 | 8.1.3 原核生物 mRNA 前体的加工 ...             | 212 |
| 7.2 启动子的结构与功能 .....               | 162 | 8.2 真核生物 RNA 的加工 .....                | 213 |
| 7.2.1 启动子的结构 .....                | 163 | 8.2.1 真核生物 tRNA 前体的转录后加工<br>.....     | 213 |
| 7.2.2 启动子的功能 .....                | 163 | 8.2.2 真核生物 rRNA 前体的转录加工<br>.....      | 215 |
| 7.3 细菌的 RNA 聚合酶 .....             | 166 | 8.2.3 细胞核 mRNA 前体剪接概述 ...             | 216 |
| 7.3.1 RNA 聚合酶概述 .....             | 166 | 8.2.4 细胞核 mRNA 前体剪接的机制<br>.....       | 219 |
| 7.3.2 大肠杆菌的 RNA 聚合酶 .....         | 166 | 8.2.5 RNA 的自我剪接 .....                 | 228 |
| 7.3.3 T7 RNA 聚合酶 .....            | 168 | 8.2.6 核酶 .....                        | 234 |
| 7.3.4 $\sigma$ 因子的结构与功能 .....     | 168 | 8.3 真核生物 mRNA 前体的选择性剪接<br>.....       | 235 |
| 7.3.5 核心聚合酶的结构与功能 .....           | 170 | 8.4 反式剪接 .....                        | 239 |
| 7.3.6 RNA 聚合酶全酶的结构与功能<br>.....    | 172 | 8.5 mRNA 5'端加帽 .....                  | 241 |
| 7.3.7 原核生物 RNA 的转录过程 .....        | 176 | 8.6 mRNA 3'端的多聚腺苷酸化 ...               | 243 |
| 7.4 真核生物的 RNA 聚合酶及其转录<br>.....    | 182 | 8.7 RNA 的编辑 .....                     | 245 |
| 7.4.1 真核生物基因转录的特点 .....           | 182 | 8.7.1 RNA 编辑的机制 .....                 | 246 |
| 7.4.2 真核生物基因转录的实验系统<br>.....      | 183 | 8.7.2 RNA 编辑的类型 .....                 | 248 |
| 7.4.3 研究真核生物基因转录起点的技术<br>.....    | 184 | 8.7.3 RNA 编辑的生物学意义 .....              | 249 |
| 7.4.4 真核生物基因转录的 RNA 聚合酶<br>.....  | 185 | 8.8 RNA 的再编码 .....                    | 250 |
| 7.5 真核生物基因转录的启动子 .....            | 188 | 8.9 RNA 干涉 .....                      | 251 |
| 7.6 类型 II 基因转录的转录因子 .....         | 193 | 8.9.1 RNA 干涉概述 .....                  | 251 |
| 7.7 类型 II 基因转录起始复合物的装配<br>.....   | 199 | 8.9.2 RNA 干涉的分子机制 .....               | 252 |
| 7.8 类型 I 和 III 基因转录的转录因子<br>..... | 201 | 8.9.3 RNAi 在异染色质形成和基因沉默<br>中的作用 ..... | 254 |
| 7.9 RNA 转录的抑制 .....               | 205 | 8.9.4 短干涉 RNA 的扩增 .....               | 255 |
| 7.9.1 嘌呤和嘧啶类似物 .....              | 206 | 8.9.5 RNA 干涉的几个重要特征 .....             | 256 |
| 7.9.2 DNA 模板功能的抑制物 .....          | 206 | 8.9.6 RNA 干涉的生理意义 .....               | 257 |
|                                   |     | 8.9.7 RNA 干涉的应用 .....                 | 257 |
|                                   |     | 关键术语 .....                            | 258 |

|                               |     |                                |     |
|-------------------------------|-----|--------------------------------|-----|
| 思考题 .....                     | 258 | 10.3.2 异染色质化 .....             | 294 |
| <b>第9章 原核生物基因表达的精细调控</b>      |     | 10.3.3 组蛋白对基因活性的影响 .....       | 294 |
| .....                         | 260 | 10.3.4 组蛋白的乙酰化-去乙酰化 .....      | 295 |
| 9.1 基因表达调控概述 .....            | 260 | 10.3.5 活性染色质对 DNase 的敏感性 ..... | 297 |
| 9.2 原核基因表达调控的若干概念 .....       | 261 | 10.3.6 非组蛋白 .....              | 298 |
| 9.2.1 细菌细胞对营养的适应 .....        | 261 | 10.3.7 核基质与基因活化 .....          | 300 |
| 9.2.2 顺式作用元件和反式作用因子 .....     | 262 | 10.3.8 基因的丢失 .....             | 301 |
| 9.2.3 结构基因和调节基因 .....         | 262 | 10.3.9 基因的扩增 .....             | 301 |
| 9.2.4 操纵基因和阻遏蛋白 .....         | 263 | 10.3.10 染色体基因的重排 .....         | 302 |
| 9.2.5 组成型蛋白和调节蛋白 .....        | 263 | <b>10.4 DNA 序列水平上的调控</b> ..... | 305 |
| 9.2.6 操纵子 .....               | 263 | 10.4.1 DNA 甲基化 .....           | 305 |
| 9.2.7 原核生物基因对调控作用做出反 .....    | 264 | 10.4.2 DNA 甲基化与转录抑制 .....      | 306 |
| 应的类型 .....                    | 264 | 10.4.3 甲基化影响 DNA 与蛋白质的 .....   | 306 |
| 9.2.8 小分子效应物的作用 .....         | 264 | 相互作用 .....                     | 306 |
| 9.3 乳糖操纵子的调控 .....            | 265 | <b>10.5 真核生物基因转录水平的调控</b>      |     |
| 9.3.1 乳糖操纵子的负调控 .....         | 265 | .....                          | 307 |
| 9.3.2 乳糖操纵子的正调控 .....         | 268 | 10.5.1 真核生物与原核生物转录调控的区别 .....  | 307 |
| 9.4 阿拉伯糖操纵子的调控 .....          | 271 | 10.5.2 真核生物基因基础转录的调节 .....     | 308 |
| 9.5 色氨酸操纵子的两级调控 .....         | 273 | 10.5.3 真核生物转录调控的顺式作用元件 .....   | 309 |
| 9.5.1 色氨酸操纵子的阻遏调控 .....       | 273 | <b>10.6 真核基因表达的转录后水平调控</b>     |     |
| 9.5.2 色氨酸操纵子的衰减系统 .....       | 274 | .....                          | 318 |
| 9.6 受双启动子调控的半乳糖操纵子 .....      | 278 | 10.6.1 酪蛋白 mRNA 的稳定性 .....     | 318 |
| 9.7 细菌的应急反应 .....             | 280 | 10.6.2 转铁蛋白受体 mRNA 的稳定性 .....  | 318 |
| 9.8 正调控系统和负调控系统 .....         | 281 | <b>10.7 转录因子对基因表达的调控</b> ..... | 319 |
| 9.9 受多重启动子调控的操纵子 .....        | 284 | 10.7.1 细胞对转录因子的调控 .....        | 319 |
| 9.10 重叠基因的调控作用 .....          | 285 | 10.7.2 转录激活因子的类型与结构 .....      | 324 |
| 9.11 细菌中 DNA-蛋白质的相互作用 .....   | 287 | 10.7.3 转录激活因子结合 DNA 的结构 .....  | 327 |
| 关键术语 .....                    | 289 | 基序 .....                       | 327 |
| 思考题 .....                     | 289 | 10.7.4 调控蛋白对特异 DNA 序列的识别 ..... | 333 |
| <b>第10章 真核生物的基因表达调控</b> ..... | 290 | .....                          | 333 |
| 10.1 真核基因表达调控的特点 .....        | 290 | 关键术语 .....                     | 334 |
| 10.2 真核细胞基因表达调控的不同层 .....     | 292 | 思考题 .....                      | 334 |
| 次 .....                       | 292 | <b>主要参考文献</b> .....            | 335 |
| 10.3 DNA 染色体水平的调控 .....       | 293 | <b>索引</b> .....                | 337 |
| 10.3.1 染色质结构状态的调控 .....       | 293 |                                |     |



**小结框** 分子生物学是研究核酸和蛋白质等生物大分子的结构与功能，并从分子水平上阐述它们间相互作用关系及基因表达调控机制的科学，是人类由被动适应自然界转向主动改造和重组自然界的学科。

## 1.2 分子生物学研究的主要内容

分子生物学是研究所有生物学现象的分子基础。从这个意义上说，分子生物学包括了地球上所有的生物学。因此，对分子生物学研究内容的界定是困难的。随着科学技术突飞猛进地向纵深发展，各学科之间的划分越来越细，在长期的科学研究与实践中，对生物学中的许多分支与科目，分子生物学家并未将它们列为分子生物学研究的范畴。例如，某些生物学反应就像一个标准的化学反应一样通过酶和产物浓度被调节，对这些反应调节的研究就属于生物化学的范围。但如果一个酶催化的反应是通过酶基因或酶分子结构的改变而被调节，则属于分子生物学的内容。这类类似于把细胞内化学成分的排列和结构的研究称为细胞生物学。但当人们分离到了昆虫和某些原虫的行为突变体后，就要对其进行分子生物学分析。因此，分子生物学与生物学其他各分支之间的界限越来越不明显了。尽管分子生物学涉猎的范围十分广泛，研究内容也包罗万象，但是按照狭义分子生物学定义，我们可将现代分子生物学的主要研究内容概括为以下几个大的方面。

### 1.2.1 基因与基因组的结构与功能

基因的研究一直是影响整个分子生物学发展的主线。在不同的历史时期对基因的研究有不同的内容，20世纪50年代以前，主要从细胞染色体水平上进行研究，是基因的染色体遗传学内容；50年代之后，主要从DNA大分子水平上进行研究，属于基因的分子生物学阶段。近20多年来，由于重组DNA技术的不断完善和应用，人们已经改变了从表型到基因型的传统研究基因的途径，而能够直接从克隆目的基因出发，研究基因的功能及其与表型的关系，使基因的研究进入了反向生物学阶段。在这个历程中，对基因与基因组的微细及高级结构与功能的研究始终是分子生物学研究内容最基础最重要的部分。

### 1.2.2 DNA的复制、转录和翻译

这一方面研究的重点是DNA或基因怎样在各细胞系统相关的酶与蛋白质等因子作用下，按照中心法则进行自我复制、转录、逆转录和翻译。同时，对mRNA分子进行各种剪接、加工修饰、编辑，以及对新生多肽链折叠成有功能的空间结构的分子机制研究。

### 1.2.3 基因表达调控的研究

基因表达的实质是遗传信息的转录和翻译。在生物个体的生长、发育和繁殖过程中，遗传信息的表达按照一定的时序发生变化（时序调节的表达）；并且，随着内外环境的变化而不断地加以修正（环境调控表达）。

基因表达的调控主要发生在转录水平和翻译水平上。原核生物的基因组和染色体结构都比真核生物的简单，转录和翻译在同一时空内发生，基因表达调控主要发生在转录水平。真

核生物有细胞核结构，转录和翻译过程在时间和空间上都被分隔开，且在转录和翻译后都有复杂的分子信息加工过程，其基因表达的调控发生在各种不同的水平，主要表现在对上游调控序列、信号转导、转录因子及 RNA 剪接等多个方面。

#### 1.2.4 DNA 重组技术

分子生物学的核心是遗传信息的结构、传递和控制，在这个过程中 DNA 重组技术是不可缺少的手段之一。DNA 重组技术是 20 世纪 70 年代初兴起的一门科学技术。应用此技术能将不同的 DNA 片段进行人为的重组和定向连接，并指定在特定的受体细胞中与载体同时复制和表达，产生影响受体细胞的新的遗传性状。严格地说，DNA 重组技术并不完全等于基因工程，这是因为后者还包括其他能使生物细胞基因组结构发生改变的体系。DNA 重组技术是核酸化学、遗传学、细胞学、病毒学、蛋白质化学、酶工程及微生物学等长期深入研究的结果，反过来，这些学科的发展又以 DNA 重组技术作为重要手段而进行。在这个过程中，限制性内切核酸酶、DNA 连接酶及其他工具酶的发现与应用是这一技术得以建立的关键。

作为分子生物学研究的内容之一，DNA 重组技术的主要目的是：①用于大量生产某些在正常细胞代谢中产量很低的多肽，如激素、抗生素、酶类及抗体等，提高产量，降低成本，使许多有价值的多肽类物质得到广泛的应用。例如，用于治疗艾滋病的基因工程白细胞介素 12 可有效地阻止病情发展，恢复人类免疫缺陷病毒（HIV）携带者的免疫系统和功能。②用于定向改造某些生物的基因组结构，使它们所具备的特殊功能更符合人类生活的需要，其经济价值能成百上千倍地提高。例如，一种含有分解各种石油成分的重组 DNA 超级细菌能快速分解石油，可用来恢复被石油污染的海域和土壤。③DNA 重组技术用于进行基础研究，已经成为研究分子生物学领域一切基础性问题的技术方法和常规武器。

#### 1.2.5 结构分子生物学

任何一个生物大分子当它在发挥生物学功能时都必须具备两个前提：一是必须拥有特定的空间结构；二是在它发挥生物学功能的过程中必定存在着结构和构象的变化。结构分子生物学的发展就是研究生物大分子特定的空间结构及结构的运动变化与其生物学功能关系的科学。它包括结构的测定、结构运动变化规律的探索和结构与功能相互关系三个方向的研究。近年来，科学家不断研究发现了大量新的生物大分子静态结构，并逐步深化对其动态功能的认识，已经能够研究运动时间最短达  $10^{-15} \sim 10^{-12}$  s，运动幅度最小为 0.1pm 的分子运动。由单一分子研究深入到对复合体乃至多亚基、多分子复合体研究，使对多种生物大分子联合起来的复杂结构与相互作用功能的认识与研究成为可能。对核糖体的结构与功能的研究就是这方面的一个例子。目前，研究生物大分子空间结构及其运动规律的手段主要是 X 射线衍射技术、二维和多维核磁共振成像、电子和中子衍射、电镜三维重组及各种波谱学的技术方法。

结构分子生物学在今后仍然是生命科学发展的基础学科。在这一领域中，仍需要生物学家、生物化学家、物理学家、化学家及计算机和工程学的专家们的共同努力。

**小结框** 分子生物学主要研究内容包括基因与基因组的结构与功能, DNA 的复制、转录和翻译, 基因表达调控, DNA 重组技术, 结构分子生物学等。

### 1.3 分子生物学与生物化学之间的关系

当代生命科学的一大特点是几乎所有关于生命的分支学科均已被分子生物学渗透, 由此涉及的许多难以穷尽的方方面面难以界定。实际上, 分子生物学与生物化学之间的关系是非常紧密而难以区分的。但随着这两门学科的研究向纵深发展, 内容越来越多, 科学家们不得不将它们做相对的划分。分子生物学的定义如前所述, 它从分子水平上研究生命现象, 生物化学是从分子水平上研究生命现象的化学本质; 从学科范畴讲, 分子生物学包括了生物化学; 但在研究的基本内容上, 如在遗传信息流 DNA→mRNA→蛋白质的传递过程中, 许多内容又属于生物化学的范围, 等等。因此, 分子生物学与生物化学这两门学科是“你中有我”、“我中有你”, 不能截然分开。但这两门学科的研究方向和研究方法与手段也显示出了明显的区别。

在研究方向上, 分子生物学主要是研究蛋白质、核酸和其他大分子的结构与功能, 以及它们之间的相互作用, 着重解决细胞中的信息传递和代谢调节等问题。而生物化学主要研究大分子物质的组成、性质结构与功能及其在生命活动中的代谢转化等动态的过程, 包含大量有机小分子的参与。因此, 分子生物学与生物化学虽然在研究内容上有相同之处, 但在研究方向上, 分子生物学的着重点是大分子的结构与功能, 而生物化学则以生物分子的动态代谢转化为主。

在研究方法上, 分子生物学是以化学和物理学的方法研究大分子结构, 采用生物化学与遗传学相结合的方法探索其功能, 解决大分子结构与功能及其代谢调节的关系。而生物化学主要采用生物化学与化学及生理学的方法, 探索生命的化学过程, 解决分子转化与能量转换的问题。所以, 分子生物学与生物化学在分离、纯化生物分子时, 或许采用同样的方法, 但在分别探索其研究的问题时, 却采用了许多不同的手段。

### 1.4 分子生物学发展的历程

分子生物学在人类文明史上的光辉成就, 以前所未有的速度推动着生物学的发展, 使整个生物学的面貌发生了巨大的变化。由于无数分子生物学家的不懈追求与刻苦研究, 人类现在不但能从分子水平上认识核酸的结构与功能, 以及复制、转录、翻译、剪接、加工、修饰等的详细过程, 而且已经测知了许多重要生物的基因组及其结构与功能, 真正从分子水平上对这些基因控制的生长、发育和变异等一系列生物学问题有了更深入的了解, 获得了令人振奋的结果。为了使学习和认识上的条理更加清晰, 在此将其发展过程简单地概括为 4 个阶段。

#### 1.4.1 人类对 DNA 和遗传信息传递的认识阶段

1928 年, Griffith 做的肺炎双球菌的转化试验奠定了 DNA 是遗传物质的理论基础。1944 年, Avery 等用生物化学和物理化学手段对 Griffith 的肺炎双球菌转化试验做了进一步分析, 证实 DNA 是生物的遗传物质。这一重大发现打破了长期以来许多生物学家认为的只

有像蛋白质那样的大分子才能作为细胞遗传物质的观点，在遗传学上树立了 DNA 是遗传信息载体的理论。

1950 年，Chargaff 提出了 DNA 碱基组成的等比例规律。与此同时 Hotchkiss 对 Avery 的转化物做了纯化，进一步证实了高纯度的 DNA 是遗传物质。

1952 年，Hershey 和 Chase 用同位素示踪技术，将 T2 噬菌体侵染大肠杆菌 (*Escherichia coli*) 细胞，证实了主要是核酸进入细菌体内，而病毒外壳蛋白留在细胞外，且进入菌体的 DNA 能利用细菌的生命过程合成噬菌体自身的 DNA 和蛋白质，并能自我组装成与亲代完全相同的子代噬菌体。烟草花叶病毒的重建实验也证明，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。

1953 年是开创生命科学新时代具有里程碑意义的一年，Watson 和 Crick 发表了“脱氧核糖核酸的结构”著名论文，提出了 DNA 双螺旋结构模型，为人类充分揭示遗传信息的传递规律奠定了坚实的理论基础。同年，Sanger 历经 8 年，完成了第一个蛋白质——胰岛素的氨基酸全序列分析。

1954 年，Crick 在前人研究工作基础上，提出了中心法则理论；Gamnow 从理论上研究了遗传密码的编码规律；1961 年 Johann 和 Matthaei 做了一个突破性试验，破译了第一批遗传密码，对正在兴起的分子生物学研究起了重要推动作用，将永载史册。

1957 年，Kornberg 在大肠杆菌中发现了 DNA 聚合酶 I，这是能在试管中合成 DNA 的第一种核酸聚合酶，这种酶已被广泛用于制备标记的 DNA 探针。

1958 年，Meselson 和 Stahl 通过经典的同位素<sup>15</sup>N 标记 DNA 的氯化铯 (CsCl) 密度梯度超速离心试验对 DNA 复制的 3 种可能的机制进行了辨别。

1940~1965 年，Jacob 和 Monod 等历经大量试验验证，提出了著名的原核基因表达操纵子学说。

#### 1.4.2 重组 DNA 技术的建立和发展阶段

1967 年，Gellert 发现了 DNA 连接酶；Philips 等确定了溶菌酶 2Å 分辨率的三维结构。

1970 年，Smith 和 Wilcox 等分离到第一种限制性内切核酸酶。同年，Temin 和 Baltimore 在肿瘤 RNA 病毒中发现了逆转录酶，证实了 Temin 1964 年提出的前病毒假说。逆转录酶已成为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。Boyer 和 Berg 等发展了重组 DNA 技术，并完成了第一个细菌基因的克隆，开创了基因工程的新纪元。

1975 年，Southern 发明了 DNA 片段印迹法；Gruenstein 和 Hogness 建立了克隆特定基因的方法；O'Farrell 发明了双向电泳的蛋白质分析方法；Blohel 等提出了信号肽假说。

1975~1977 年，Sanger、Maxam 和 Gilbert 发明了 DNA 序列测定技术。

1977 年，Berget 等发现了断裂基因；第一个全长 5387bp 的噬菌体  $\Phi$ X174 基因组序列测定完成。

#### 1.4.3 重组 DNA 技术的应用和分子生物学的迅猛发展阶段

1979 年，Solomon 和 Bodmer 最先提出至少 200 个限制性内切核酸酶片段长度多态性 (RFLP) 可作为连接人类基因组图谱的基础。

1980年, Wigler 等将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 RNA 前体的自我剪接作用, 具有催化作用 RNA (酶性核酸) 的发现促进了 RNA 研究的飞速发展; 同年, Palmiter 等获得转基因小鼠; Sprading 等培育出转基因果蝇。

1982年, Prusiner 等在感染瘙痒病的仓鼠脑中发现了朊病毒 (prion); 同年, 第一个由基因工程生产的药物——胰岛素在美国和英国获准使用; Sanger 及其合作者完成  $\lambda$  噬菌体全长 48 502bp 的基因组 DNA 全序列测定。

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

1989年, Greider 等在纤毛原生动物中发现了端粒酶是以内源性 RNA 为模板的逆转录酶; Hiatt 等首次报道了在植物中亦可产生单克隆抗体。同年, 美国专利商标局宣布将接受基因工程植物和基因工程动物方面的专利申请。第一个具有专利权的用于医药研究的动物——杜邦肿瘤鼠诞生。

1990~1992年, 转基因玉米及转基因小麦诞生, 谷物基因工程开始变为现实。1992年, 欧洲共同体各国 35 家实验室发表第一个真核生物染色体 (酵母 3 号染色体) DNA 全序列, 共 315kb。

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

1991年, 由欧洲共同体组织 17 个国家 35 个实验室的 147 位科学家, 手工测序完成了第一条完整染色体的测序工作。

1994年, 日本科学家在 *Nature Genetics* 上发表了水稻基因组遗传图; Wilson 等完成了线虫 3 号染色体连续的 2.2Mb 的测定, 预示着百万碱基规模的 DNA 测序时代的到来; 澳大利亚学者 Wilkins 提出了“蛋白质组”的概念。

1995年, Zimmerly 等发现 II 类内含子的结构及剪接机制。

1996年, Dietrich 等绘制了小鼠基因组的完整遗传图谱。

1997年, Wilmut 等首次不经过受精, 用成年母羊的体细胞遗传物质成功获得克隆羊——Dolly; Willard 等首次构建了人染色体 (HAC); Salishury 等发现 DNA 的一种新的结构形式——四显性组合, 是基因交换期间 DNA 连接的一种方式。

1998年, 第一个多细胞真核生物线虫基因组被揭示; 美国科学家 James 等用 ES 和 EG 细胞建立了胚胎干细胞系, 为研究胚胎干细胞的发育和利用胚胎干细胞治疗疾病提供了全新的途径; Doyle 等勾画出了清晰的第一个钾离子通道结构, 这是分子生物学史上的一个重要进展。

1999年, Cate 等第一次绘制完整核糖体晶体结构, 并揭示了很多精细结构; 国际人类基因组计划 (HGP) 研究小组完成了人的第 22 号染色体测序工作。

2000年, 国际 HGP 联合研究小组完成了人类第 21 号染色体的测序, 从原定预计的 2003 年 6 月提前到 2001 年 6 月; 果蝇和拟南芥的基因组测序完成; 由 Coleman 领导的研究小组完成的世界首例克隆猪在苏格兰诞生; 中国超级杂交水稻基因组计划正式启动; 2000 年

3月,美国塞莱拉基因组公司宣布完成果蝇基因组测序;2000年12月,英国、美国等国科学家宣布绘出拟南芥基因组的完整图谱,这是人类首次全部破译一种植物的基因序列,此专题在国际知名刊物上有多篇重要文章发表。

#### 1.4.4 对功能基因与基因组的研究阶段

2001年, *Nature* 和 *Science* 同时发表了 HGP 全序列。Hartwell、Hunt 和 Nurse 因为对细胞周期调控因子的研究取得重大突破而分享了诺贝尔生理学或医学奖。

2002年4月,以杨焕明为首的中国科学家在 *Science* 上发表了水稻(籼稻)全基因组框架序列图。

2003年,美国、日本、英国、法国、德国、俄罗斯和中国科学家共同宣布 HGP 序列图绘制成功, HGP 目标基本完成;中国科学家宣布启动“中华人类基因组单体型图”计划;中国、美国分别测定出非典型肺炎病毒的基因图谱。

2004年,以色列学者 Ciechanover、Hershko 和美国的 Rose 发现泛素蛋白调节的蛋白质降解机制。

2005年,由美国、中国、日本等国的200多位学者参加的“国际人类基因组单体型图计划”取得重要成果,公布了人类基因组“差异图”,发现了100多万个常见 SNP 位点,标定了单体型“模块”在 DNA 链上的“边界”,并划分了基因组上包含最常见 DNA 变异的10个区域。利用这份“差异图”,在对糖尿病、阿尔茨海默病和癌症等疾病的研究中,将患者与健康人全基因组的 SNP 进行比较,更高效地寻找与疾病相关的基因变异。

2006年,美国 Fire 和 Mello 发现了 RNAi 干扰机制,他们因在 RNAi 干扰机制研究中的贡献获得了诺贝尔生理学或医学奖,目前该机制已被广泛用作研究基因功能的一种手段,有助于诊断和治疗遗传病等多种疾病;英国 Wellcome Trust Sanger Institute 和美国杜克大学人类遗传学中心等研究机构的研究人员联合在 *Nature* 上发表人类最后一条染色体——1号染色体基因测序。在人体全部22对常染色体中,1号染色体包含的基因数量最多,破译难度最大。

2007年,美国犹他大学人类遗传学研究所 Capecchi、北卡罗来纳州大学 Smithies 和英国 Evans 对胚胎干细胞的研究使其获得诺贝尔生理学或医学奖。

2008年,日本、美国和美籍华裔科学家因研究绿色荧光蛋白作出特殊贡献而获奖;*Science* 评出的2008年十大科学进展中,细胞重编程被评为第一位,这项研究几乎在一夜之间开启了一个生物学的新领域,它有望成为挽救生命的医学上的重大进步。细胞重编程质疑细胞单向发育的理论,认为已分化的细胞在特定条件下能被逆转,恢复到全能性状态,或形成胚胎干细胞系,或进一步发育成一个新个体。通过插入回拨细胞发育时钟的基因,科学家正在深入了解疾病和研究细胞如何决定其命运的生物学。

2009年, Ramakrishnan、Steitz 及 Yonath 3位科学家因对核糖体结构和功能的研究而获得诺贝尔化学奖;3位美国学者 Blackburn、Greider 及 Szostak 因发现了由染色体根冠制造的端粒酶,以及这种染色体的自然脱落物将引发衰老和癌症,而被授予诺贝尔生理学或医学奖;中国上海药物研究所药物发现与设计中心(DDDC)罗成副研究员与美国弗吉尼亚联邦大学医学院 Spiegel 首次发现了鞘氨醇磷酸酯(S1P)是一种组蛋白去乙酰化酶(HDAC)生理调节因子。

2010年,中国复旦大学学者发现乙酰化作用机制,由于人体80%疾病与代谢有关,揭开代谢的奥秘就等于找到了制服疾病的秘密,为肝病、肿瘤等代谢疾病的药物研发提供了开拓性的思路;美国密歇根大学学者发现RNA分子三维结构不是由复杂的化学相互作用决定,而仅仅取决于几何学特性,这一规则有可能使人们重新认识RNA结构的三维构象,并有利于合理掌握RNA的结构从而控制其活性,利用RNA分子设计小分子药物,以及设计根据使用者指定的方式而改变结构的RNA传感器等。

2011年,美国科学家通过抑制被致癌基因所编码的P53蛋白的表达,成功地将人体的皮肤细胞变为脑细胞,表明癌细胞和干细胞可能有相同起源;美国俄亥俄州立大学学者创造出一种新技术,能同时观察细胞中的基因数量和蛋白质表达,此技术能更透彻地了解癌症及其他复杂疾病;美国科学家找到了DNA的新碱基,即5-甲酰胞嘧啶、5-羧基胞嘧啶,并在人体胚胎干细胞和实验鼠器官染色体组的DNA中发现了这两个碱基的踪迹。

2012年,美国宾夕法尼亚大学Tanakaa等发现,microRNA-34在果蝇中可抑制衰老相关疾病的发生;台湾大学与美国约翰霍普金斯大学合作证实有两种蛋白质可调控长寿基因(AMPK)表达。

现代分子生物学的研究成就还包括以下几个方面的重要突破:①对人类基因组计划的实施取得了多项重大成果,建立了多种不同类型和层次的文库;各类分子探针如限制性内切核酸酶片段长度多态性、扩增片段长度多态性(AFLP)、随机扩增片段长度多态性(RAPD)等的应用;聚合酶链反应的普及应用和日益高度自动化并由机器人操作的核酸自动测序仪,以及大容量先进的计算机分析存储技术,等等。②在互联网上注册公布的有详细注释的许多生物,如噬菌体、大肠杆菌、酵母、果蝇、线虫、小鼠、水稻、拟南芥等的全基因组序列,以及大量在过去认为根本无法得到的蛋白质序列,为研究者提供了共享生物分子信息资源的绝好平台。③对肿瘤、艾滋病、心血管疾病、高血压及糖尿病等的防治,从分子机制上已取得可喜的成果。④农作物抗病虫害、动植物品种和品质改良、抗病毒病等方面已逐步走向实用阶段。

综合以上不难看出,20世纪主要以核酸研究为核心,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测序技术,以及进入21世纪以来的功能基因与以基因组和蛋白质组为代表的各种“组学”研究都是分子生物学发展的里程碑,它们将生命科学带向了一个由宏观到微观,再到宏观的伟大时代。

## 1.5 21世纪分子生物学发展的趋向

人类基因组计划已经提前完成了,人类基因组研究的重点正在由序列结构向基因功能转移,进入了以基因组功能研究为主要内容的“后基因组”(post-genomics)时代。主要任务是研究细胞内全部基因的表达图式和全部蛋白质图式,或者说是“从基因组到蛋白质组”。由此,分子生物学的重点又回到了蛋白质上来,分子生物信息学应运而生,生命科学正在进入一个全新的时代。

### 1.5.1 功能基因组学

功能基因组学(functional genomics)依赖于对DNA序列的认识。应用基因组学的知

识和工具，人们能够了解和认识影响整个生命过程的特定序列表达谱。例如，酿酒酵母 (*Saccharomyces cerevisiae*) 16 条染色体全部序列的测序工作已于 1996 年完成，基因组全长 13 040kb，含有 5885 个可能编码蛋白质的基因，140 个编码 rRNA 的基因，40 个编码 snRNA 的基因和 275 个 tRNA 基因，共计 6340 个基因。功能基因组学就是进一步研究这些基因在一定条件下，如在孢子形成期，同时有多少基因协同表达才能完成这一发育过程。这就需要研究这一时期的全套基因表达谱 (gene expression pattern)。解决如此复杂的问题必须在方法学上有重大突破，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。

功能基因组学也包括了在测序后对基因功能的研究。例如，酵母有许多功能重复的基因，分布在染色体两端，当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。基因冗余现象实际上是对环境的适应，冗余基因的存在为进化适应提供了可选择的余地。在基因组全序列中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。目前，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 6000 多个基因的单突变体文库 (single mutant library)，并用于其他高等真核生物基因组的“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比较，从基因组整体水平上阐述其活动规律。核心问题是基因组的多样性和其进化规律，基因组的表达及其调控，模式生物基因组研究，等等。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研究的重要标志。

### 1.5.2 蛋白质组学

1994 年，Wilkins 等首先提出了蛋白质组 (proteome) 的概念，随后得到国际生物学界的广泛承认。蛋白质组的定义是指一个基因组所表达的全部蛋白质的总和 (proteome indicates the proteins expressed by a genome)；“proteome”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 (proteomics) 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，一种生物体的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相同，人类基因的总数为 2 万~2.5 万个。单从 DNA 序列的信息并不能解释某个基因的表达时间、表达量、表达产物蛋白质翻译后加工和修饰等情况，以及它们在亚细胞群中的分布等。这些问题有望在蛋白质组学的研究中找到答案。因为细胞内的蛋白质组是动态的，有它的时序性、可调节性、相关性，所以能够在细胞和生物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库与基因组数据库的整合，将对功能基因组的研究发挥重要作用。

随着研究的深入和需要，科学家将蛋白质组由原来的定义，即一个基因组所表达的蛋白质，改为细胞内基因组所表达的全部蛋白质。但要获得如此完整的蛋白质组，在实践中是极其困难的。因为蛋白质的种类和形态总是处在细胞新陈代谢的动态过程之中，所以难以准确测定。所以，1997 年，Cordwell 和 Smith 提出了功能蛋白质组 (functional proteome) 的概念，功能蛋白质组指在特定时间、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，中国生物学家提出了功能蛋白质组学 (functional proteomics) 新概念，研究定位在细



胞内与某种功能有关或在某种环境条件下的一群或一套蛋白质。

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，也能丰富总蛋白质数据库，它是从生物大分子水平到细胞水平研究的重要桥梁。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，获得总蛋白质谱。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技术。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，包括被分离蛋白质在图谱上的定位、信息计数、图谱间蛋白质差异表达的检测，等等。一种细胞或组织优质的蛋白质组双向电泳图，可以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组分析，质谱技术已成为蛋白质鉴定的重要技术。由质谱技术测得完整蛋白质的相对分子质量、肽质谱及部分肽序列等数据，通过相应数据库的搜寻，从而比对鉴定目标蛋白质。最后，再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之上，建立蛋白质组数据库。

从1997年 Hodes 等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) 到目前，蛋白质组的研究进展速度极快。随着新思路 and 新技术的不断涌现，这项新技术将会不断完善，发展成为后基因组时代的重点技术手段。

### 1.5.3 生物信息学

HGP 和其他重要生物基因组的大量序列信息和功能信息的积累，以及在互联网上的注册与详细注释等网络资源，催生了生物信息学 (bioinformatics) 这门新学科。对浩如烟海的 DNA 和蛋白质分子中各种类型的信息进行识别、存储、分析、比对、注释、模拟和传输是生物信息学的主要内容，它由各类数据库、计算机网络和应用软件三大部分组成。主要在基因的染色体定位、寻找感兴趣基因的同源序列的快速搜索、基因识别和文献查寻等方面提供十分有用的帮助，目前已成为分子生物学家必不可少的工具。

国际上现有大型生物信息中心，主要是美国国家生物信息学中心 (NCBI)、基因组序列数据库 (GSDDB)、欧洲分子生物学研究所 (EMBL) 和日本 DNA 数据库 (DDBJ)，等等。这些中心和全球的基因组研究实验室通过网站、电子邮件或直接与服务器和数据库联系而获得搜寻系统，使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。

随着大规模 DNA 自动测序的迅速发展，序列数据呈爆炸性地增长，国际上许多著名实验室曾在 HGP 启动的同时，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，使生物信息学逐步走向成熟，在各基因组计划中发挥了无以取代的重要作用。所建立的核苷酸序列数据库，已存有大量生物的 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、多序列比对、同源序列检索、三维结构观察与演示、进化树生成与分析，等等。

在蛋白质组研究中，由于蛋白质组是随着生物个体发育阶段和所处的内外环境而变化的，mRNA 丰度与蛋白质的丰度并非显著相关，以及需要经历翻译后的加工修饰等，对蛋白质的生物信息学研究，在内容上有许多特殊之处。当前通用软件所建立的数据，主要有蛋白质分子质量、序列信息、结构域、二维电泳、三维结构、特殊性质、模拟酶解、翻译后修