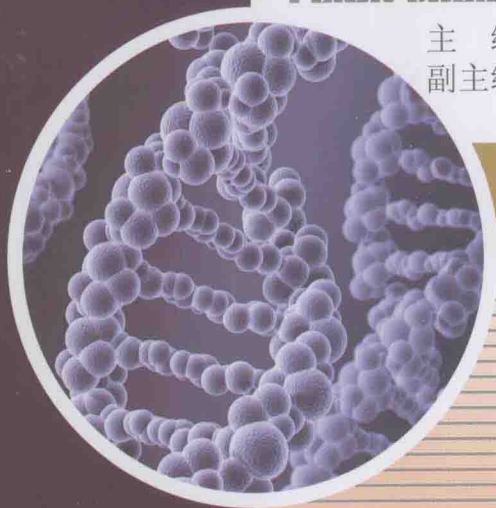


分子生物

反应与技术研究

FENZI SHENGWU FANYING YU JISHU YANJIU

主 编 冀玉良 兰 喜 王庆东
副主编 刘富平 尹艳丽 武安泉
翟焕趁



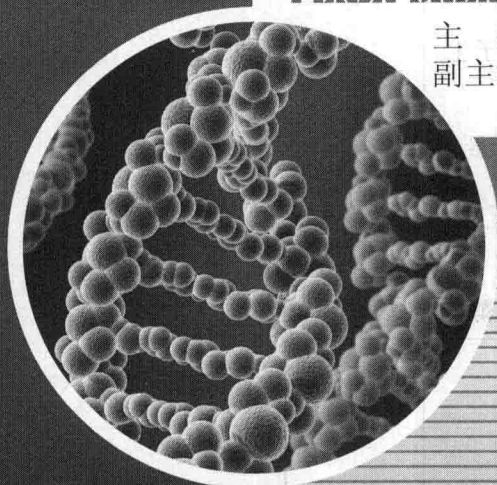
中国水利水电出版社
www.waterpub.com.cn

分子生物

反应与技术研究

FENZI SHENGWU FANYING YU JISHU YANJIU

主 编 冀玉良 兰 喜 王庆东
副主编 刘富平 尹艳丽 武安泉
翟焕趁



中国水利水电出版社
www.waterpub.com.cn

内 容 提 要

全书围绕分子生物反应与技术展开讨论,主要内容包括导论,遗传物质的分子本质,DNA、RNA、蛋白质的生物合成,转座子,生物基因表达的调控,生物发育的分子调控,细胞的信息传递,分子遗传技术,分子生物基本操作技术,分子生物技术的应用等内容,通过对基本分子生物反应分析而逐步展开对分子生物反应相关技术进行研究。

本书内容丰富,取材新锐,文字表述简明扼要,是一本适合于分子生物反应研究爱好者阅读的实用性强的学术著作,对相关院校专业的师生和相关领域的研究人员来说也是一本颇为有益的参考书。

图书在版编目(CIP)数据

分子生物反应与技术研究/冀玉良,兰喜,王庆东

主编. —北京:中国水利水电出版社,2014.3

ISBN 978-7-5170-1824-7

I. ①分… II. ①冀…②兰…③王… III. ①分子生物学—研究 IV. ①Q7

中国版本图书馆 CIP 数据核字(2014)第 051450 号

策划编辑:杨庆川 责任编辑:杨元泓 封面设计:马静静

书 名	分子生物反应与技术研究
作 者	主 编 冀玉良 兰 喜 王庆东 副主编 刘富平 尹艳丽 武安泉 翟焕趁
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
经 售	电话:(010)68367658(发行部)、82562819(万水) 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 27 印张 691 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	89.00 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

分子生物学是从分子水平研究生命本质的一门新兴边缘学科。该学科以核酸和蛋白质等生物大分子的结构及其在遗传信息和细胞信息传递中的作用为研究对象,是当前生命科学中发展最快并正在与其他学科广泛交叉与渗透的重要前沿领域。分子生物学的发展为人类认识生命现象带来了前所未有的机会,也为人类提高意识进化至新的层面创造了极为广阔的前景。

进入 21 世纪,分子生物学知识获得了快速更新,已深入到了生命科学的各个分支中,如生理学、病理学、发育生物学、神经生物学、免疫学等学科,取得了许多令人瞩目的成就,已经成为推动生物科学持续有力发展的主要动力和重要的技术保障,同时也深刻影响着人类的生活和社会发展。可以说,掌握和了解分子生物反应的原理与技术,成为了从事该领域乃至其他相关学科研究工作的一个十分重要的前提。

尽管生命体的结构和生活方式千姿百态,但是从分子水平上观察却都凸显出生命过程内在的统一。正是这种统一使分子生物学发展成为一门独立的学科。本书在结合各参编教师多年来的实践和经验,同时还吸取了近年来相关著作优点的基础上,用 13 章的内容对分子生物反应与技术进行了研究。

第 1 章为导论,主要就分子生物学的起源与发展、概念和基本内容进行了概述。第 2 章和第 3 章的内容包括遗传物质的分子本质,以及基因、基因组与基因组学。第 4 章至第 6 章讲述了 DNA、RNA、蛋白质的生物合成反应。第 7 章至第 10 章分别对转座子、生物基因表达的调控、生物发育的分子调控、细胞的信息传递等内容进行了描述。第 11 章至第 13 章则主要从实际应用出发,详细阐述了分子遗传技术、分子基本操作技术和分子生物技术的应用。本书的编写力求做到概念明确,逻辑严密,条理清楚,语言简练,文字流畅,理论联系实际,在贯彻基础性、系统性、科学性等原则的同时,又极力注重本书的实用性。

全书由冀玉良、兰喜、王庆东担任主编,刘富平、尹艳丽、武安泉、翟焕趁担任副主编,并由冀玉良、兰喜、王庆东负责统稿,具体分工如下:

第 1 章、第 3 章:冀玉良(商洛学院);

第 8 章、第 9 章、第 13 章:兰喜(中国农业科学院兰州兽医研究所);

第 2 章、第 11 章:王庆东(郑州大学);

第 6 章、第 10 章:刘富平(海南大学);

第 7 章、第 12 章:尹艳丽(河南工业大学);

第 4 章:武安泉(周口师范学院);

第 5 章:翟焕趁(河南工业大学)。

在本书的编写中,参阅了许多相关著作和文献资料,吸收了他们的很多观点,在此一并谨向这些相关作者表示由衷的感谢,恕这里不一一列举了。

分子生物学的原理和研究方法极其丰富且发展十分迅速的,由于篇幅所限,书中难以一一详细论述探讨。同时,囿于编者水平,书中难免有不妥之处,敬请读者和专家批评指正,我们将不胜感激。

编 者

2014 年 1 月

目 录

第 1 章 导论	1
1.1 分子生物学的定义	1
1.2 分子生物学的发展	1
1.3 分子生物学研究的主要内容	12
第 2 章 遗传物质的分子本质	19
2.1 遗传物质	19
2.2 核酸结构	24
2.3 核酸的变性和复性	45
2.4 核酸的研究技术	50
2.5 核酸的序列测定	55
第 3 章 基因、基因组与基因组学	59
3.1 基因	59
3.2 基因组	71
3.3 基因组学	93
第 4 章 DNA 生物合成反应	102
4.1 DNA 的复制	102
4.2 DNA 的逆转录	116
4.3 DNA 的损伤与修复	120
4.4 DNA 的突变	130
4.5 DNA 的重组和转座	137
第 5 章 RNA 生物合成反应	144
5.1 DNA 转录	144
5.2 生物 RNA 的合成	149
5.3 生物 RNA 的转录后加工	175
5.4 RNA 生物合成的选择性抑制	184
5.5 内含子与外显子	187
第 6 章 蛋白质生物合成反应	192
6.1 蛋白质构象	192
6.2 参与蛋白质合成的物质	194
6.3 遗传密码	202
6.4 生物的蛋白质合成	206
6.5 蛋白质翻译后的修饰	214
6.6 蛋白质翻译后的输送	219
6.7 蛋白质合成的抑制剂	227
6.8 蛋白质的降解	228

第 7 章 转座子	231
7.1 转座子概述	231
7.2 转座子模型与基因表达	251
7.3 反转录转座子	257
7.4 转座的分子机制	267
第 8 章 生物基因表达的调控	271
8.1 生物基因表达调控概述	271
8.2 原核生物的基因表达调控	272
8.3 真核生物的基因表达调控	291
8.4 基因表达调控异常与疾病	304
第 9 章 生物发育的分子调控	307
9.1 概述	307
9.2 各物质对生物发育的分子调控	311
第 10 章 细胞的信息传导	317
10.1 细胞信息传导概述	317
10.2 G-蛋白关联受体的信息调控	319
10.3 通过酶关联细胞表面受体进行信号调控	329
10.4 小分子信号调控	331
10.5 细胞对信号的反应	332
第 11 章 分子遗传技术	334
11.1 分子遗传基本技术	334
11.2 DNA 文库的构建与筛选	346
11.3 遗传物质的分子水平分析	350
第 12 章 分子基本操作技术	361
12.1 核酸的分离、纯化、检测和杂交	361
12.2 聚合酶链式反应	365
12.3 凝胶电泳技术	369
12.4 DNA 序列分析	377
12.5 基因表达分析	379
第 13 章 分子生物技术的应用	389
13.1 分子标记	389
13.2 生物芯片	392
13.3 基因治疗	395
13.4 DNA 指纹图谱	402
13.5 DNA 重组技术鉴别人类基因	405
13.6 利用抑制基因表达分析生物学过程	413
13.7 转基因动物和植物	420
参考文献	425

第 1 章 导 论

1.1 分子生物学的定义

分子生物学(Molecular Biology)是在分子水平上研究生命的重要物质(注重于核酸、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。

20 世纪初期,现代科学技术迅速发展起来。当时的研究者们已经在细胞学、遗传学、生物化学和物理学等领域取得了大量的研究成果。到了 20 世纪中期,科学家们不仅研究豌豆、果蝇、玉米等材料,也对一些简单的生物,如细菌、噬菌体等作了深入的研究。所谓的简单只是与高等生物相比较而言,这些细菌和噬菌体其实也是相当复杂的。科学家们发现:在大多数生物中,DNA 是主要的遗传信息载体;DNA 的结构使它的复制与修复近乎完美;DNA 的线状结构编码了蛋白质的三维结构。所有这些研究成果表明:那些控制简单生物的基本生物学原则对那些复杂的生物同样适用。从此,分子生物学真正发展起来。

从广义上来说,蛋白质及核酸等生物大分子的结构与功能的研究都属于分子生物学的范畴,也就是从分子水平阐明生命现象和生物学规律,如蛋白质的结构、功能和运动,酶的作用机理和动力学,膜蛋白的结构、功能和跨膜运输等,都属于分子生物学的内容。但是有些题目一般不属于分子生物学内容,如代谢中的某些反应,如果这些反应由反应物和产物的浓度来调节,一般就认为是典型的生物化学反应。另外,细胞结构与各种细胞成分的组织则属于细胞生物学。

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。特别是核糖核酸(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。

本书的内容也侧重于狭义分子生物学的概念。

1.2 分子生物学的发展

科学领域中任何一门学科的形成和发展,一般很难准确地说明它是何时、何人创始的。分子生物学的产生和发展,同其他学科一样,经历了漫长而艰辛的过程,逐步走向成熟而迅速发展的道路。

1.2.1 分子生物学的产生背景

自从有了人类文明史,就有了人们对生命现象的记载与描述,就有了人们对大自然、对生命现象的观察与思考。分子生物学的形成经历了漫长的历史发展过程:由形态结构→细胞结构→分子结构;由定性分析→定位分析→定量分析;由表现型→基因型→表现型,不断深化、不断成熟。

生命的形成经历了两个大的阶段:化学进化阶段与生物进化阶段。恩格斯说:“生命的起源,必然是通过化学的途径实现的。”奥巴斯也指出:“地球上生命的起源是碳氢化合物,和由它们形成的多分子系统进化过程中的有规律事件。”洪荒时代的地球是一个无生命的世界,被一氧化碳、二氧化碳、氢气、氮气、氨气、甲烷、硫化氢等组成的还原性大气笼罩。随着地球上化学反应的活跃,物质开始了化学进化。无机化合物借地热、放电、紫外线、宇宙射线等能量合成了简单的有机化合物。随着有机化合物的蓄积与有序组合,在某种特定环境中出现了原始生命。物质进化从此由化学进化步入了生物进化阶段。

纵观生命科学史,根据已有的研究与证据,关于生命现象与本质的研究可分为四个阶段:生命物质形成阶段,细胞形成阶段(单细胞生物),多细胞生物形成阶段,人类——高等智能生物的形成阶段。对生命研究的层次与内容也随之不断地多元化、精细化、量化、系统化与整合化。

生物学最早是研究动植物的形态、解剖和分类,偏重于宏观的描述。从早期对自然界生物的观察与描述,发展到研究其结构、机能以及各种生命过程。自从1839年Schwann和Schleiden证明动物和植物都是由细胞组成之后,生物学进入细胞水平的研究。由于细胞学的研究得到迅速发展,遗传学原理也得到揭示,生理学和生物化学随之兴起。以细胞为主要材料,人们对细胞的化学组成的了解日益深化,对构成细胞的生物大分子(主要是蛋白质及核酸)在生命活动中所起的作用有了深刻的认识。这些促使生物学的探索逐渐进入了亚细胞水平与分子水平。

1938年的一份洛克菲勒基金会年度报告(Report of the Rockefeller Foundation)中,首次出现“molecular biology”一词。当时,洛克菲勒基金支持了Bernal和Crowfoot发表的第一张胃蛋白酶晶体的X线衍射图谱有关研究,以及Astbury和Bell关于DNA的X线晶体图谱所揭示的DNA结构像“一叠钱币”的研究。这一系列的研究工作已经开始应用相当精细的技术进行了生命活动的定量研究,研究的内容已经涉及生命活动的精细过程。人们的目光已经注意到生命现象的深层次问题——大分子生命物质的分子结构与功能的相关性。1945年,William Astbury正式使用“分子生物学”这一术语,并将分子生物学定义为生物大分子的化学和物理结构的研究。

1869年,德国学者Miescher首次从莱茵河鲑鱼精子中提取了DNA。1871年,Miescher又从白细胞核中分离出DNA。1910年,Kossel第一次分离获得了核酸类物质的组成单位——核苷酸。19世纪末20世纪初,Mendel和Morgan等人根据长期实验研究结果,已开始认识到生物遗传的分子基础。一系列实验已经证明一切生命现象和生物性状都与细胞内核酸、蛋白质等生物大分子的基本化学反应密切相关,只是这些生物大分子是以何种特殊的结构形式与作用机制来决定生命现象和生物性状的还不清楚。

1949年,Chargaff发现一个重要规律,即在任何来源的DNA中, $G=C$, $A=T$,且 $A+G=C+T$ 。在不同来源的DNA分子中, $(A+T)/(G+C)$ 的比值不同。1953年,Watson和Crick共同提出了脱氧核糖核酸的双螺旋模型,即DNA分子是以磷酸糖链为主链反向平行的双股螺旋结构,DNA上的四种碱基按Chargaff规律依靠氢键构成双股磷酸糖链之间的碱基对。这个模

型为揭开遗传信息的复制和转录的秘密铺平了道路。同年, Sanger 利用纸层析和纸电泳技术第一次揭示了生物大分子蛋白质激素——胰岛素的一级结构, 开创了生物大分子序列分析的新纪元。

Watson 和 Crick 于 1953 年提出了 DNA 反向平行双螺旋的结构模型, 生物学研究转向以生物大分子为目标, 分子生物学开始形成了独立的学科。这是人类对生物界认识的一个不断深入的过程。

生物学发展到今天, 以分子生物学为特征的现代生物学从生物表现型研究发展到基因型研究, 从基因的遗传与稳定性研究发展到基因的突变与转移性研究, 从零星的单个基因研究发展到基因组学研究, 从功能基因研究发展到基因表达调控研究。生物学研究从宏观到微观, 从现象深入到本质, 从结构联系到功能, 终于形成了生物学的带头学科——分子生物学。

现代生物学研究已经证明, 生物进化的外部原因在于自然环境选择的结果, 而内部的动力则是基因的突变、转移与重组。分子生物学研究已经实现了“基因→表型”的飞跃, 使生物学研究由“表型→基因”的研究模式提升为“表型→基因→表型”的研究路途。这一良性循环的研究模式大大加速了生命科学发展的进程。因此, 现代生物学又称之为反向生物学, 分子生物学就是现代生物学的灵魂。

1.2.2 分子生物学的发展历程

归纳分子生物学的产生背景, 可将分子生物学的发展分为三个阶段: ①19 世纪后期到 20 世纪 50 年代初的准备和酝酿阶段; ②50 年代初到 70 年代初的现代分子生物学的建立和发展阶段; ③70 年代后至今的初步认识生命本质并开始改造生命的深入发展阶段。

1. 准备和酝酿阶段

在这一阶段的两大重点是: 确定了蛋白质是生命的主要基础物质; 确定了生物遗传的物质基础是 DNA。

20 世纪二三十年代的确认了自然界有 DNA 和 RNA 两类核酸。但由于当时对核苷酸和碱基的定量分析不够精确, 长期认为 DNA 结构只是“四核苷酸”单位的重复, 不具有多样性, 不能携带更多的信息。当时对携带遗传信息的候选分子更多的是考虑蛋白质。A. Kossel 关于细胞化学尤其是蛋白质和核酸方面的研究取得突破, T. H. Morgan 发现染色体在遗传中的作用, 他们分别获得 1910 年和 1933 年诺贝尔生理学或医学奖。40 年代以后的实验事实使人们对核酸的功能和结构有了正确的认识。1944 年 O. T. Avery 等证明了肺炎球菌转化因子是 DNA; 1952 年 S. Furber 等的 X 射线衍射分析阐明了核苷酸并非平面的空间构象, 提出了 DNA 是螺旋结构; 1948~1953 年 E. Chargaff 等用新的层析和电泳技术分析组成 DNA 的碱基和核苷酸量, 提出了 DNA 碱基组成 A—T、G—C 的 Chargaff 规则, 为碱基配对的 DNA 结构认识打下了基础。

2. DNA 的半保留复制

1953 年 Watson 和 Crick 提出的 DNA 双螺旋结构模型(DNA double helix model)是现代分子生物学诞生的里程碑。DNA 双螺旋结构发现的深刻意义在于: ①确立了核酸作为信息分子的结构基础; ②提出碱基配对是核酸复制、遗传信息传递的基本方式; ③确定了核酸是遗传的物质基础, 为认识核酸与蛋白质的关系及其生命中的作用打下了最重要的基础。

在 Watson 和 Crick 提出 DNA 的双螺旋结构模型的时候,就对 DNA 的复制过程进行了预测,按照 DNA 的双螺旋结构模型,DNA 分子由两条反平行的 DNA 单链组成,两条链上的碱基按照碱基互补配对原则,通过氢键相连,一条链上的 G 只能与另一条链上的 C 配对,A 只能与 T 配对,即一条链上的碱基排列顺序决定了另一条链上的碱基排列顺序,或者说,DNA 分子中的每一条链都含有合成它的互补链所需要的全部信息。因此,Watson 和 Crick 认为,DNA 复制时,两条互补链的碱基对之间的氢键首先断裂,双螺旋解开,两条链分开,分别作为模板,按照碱基互补配对原则合成新链,每条新链与其模板链组成一个子代 DNA 分子,子代 DNA 分子具有与亲代 DNA 分子完全相同的碱基排列顺序,即携带了相同的遗传信息。在这个过程中,一个亲代 DNA 分子通过复制产生了两个相同的子代 DNA 分子,每个子代 DNA 分子中一条链来自亲代 DNA(模板),一条链来自新合成的 DNA,这种方式被称为半保留复制(semiconservative replication)。后来由 Meselson 和 Stahl 设计的实验结果与按照半保留复制机制预期的结果完全一致。

3. 基因与蛋白质之间确立关系

DNA 通过半保留复制机制精确地自我复制,从而将遗传信息传递给子代,保证了遗传的稳定性。那么 DNA 又是如何发挥作用,从而决定的不同生物特定的性状呢?

1902 年,Archibald Garrod 在研究黑尿病(alkaptonurea)时发现这种疾病符合孟德尔隐性遗传规律,因此推测这种疾病很可能是由一个基因变异失活而引起的。病人的主要症状是尿液中黑色素的积累,Garrod 据此认为该病是由某条生化代谢途径中的某种中间产物的异常积累而引起的,他假设这种异常积累是因为转化该中间产物的酶失活而造成的。在此基础上 Garrod 提出了“一个失活的基因产生一种失活的酶”的假设,即一个基因一种酶假说(one-gene/one-enzyme hypothesis),

George Beadle 和 E. L. Tatum 通过链孢菌在突变菌株中找到了催化某一代谢反应的失活的酶,并证明在突变菌株中只有一个基因发生了突变。也就是说,一个突变的基因产生一种失活的酶,或者干脆不产生产物。这就是一个基因一种酶假说。“一个基因一种酶”的假说在 1957 年第一次得到了很好的实验证明。V. M. Ingram 研究了镰刀形细胞贫血症(sickle cell anemia)的血红蛋白和正常血红蛋白的氨基酸序列后发现,镰刀形细胞贫血症患者的 β 珠蛋白同正常野生型之间仅有一个氨基酸的差别,即在 β 珠蛋白的氨基端第六位缬氨酸取代了正常的谷氨酸。这表明基因的突变会直接影响到它所编码的蛋白质多肽链的结构,从而为“一个基因一种酶”的假说提供了有利的证据。然而,这个假说并不完全正确,这是因为:

- 1)许多酶分子可以由数条多肽链组成,而一个基因只能产生一条多肽链。
- 2)许多基因负责产生非酶蛋白质。
- 3)一些基因的产物并不是蛋白质,而是 RNA。因此,后来有人主张将“一个基因一种酶”假说修正为“大多数基因含有产生一条多肽链的信息”。

4. 中心法则的建立

所谓中心法则是指生物体可以在 DNA 模板上合成 RNA,再以 RNA 为模板、在适应体参与下、氨基酸按密码顺序合成肽链,使得遗传信息从 DNA 流向 RNA 再流向蛋白质的规律。

在中心法则之中,编码蛋白质的基因中所蕴含的信息通过转录(transcription)和翻译(translation)两个相关联的过程得到表达。

关于基因的转录,1955年 Brachet 用洋葱根尖和变形虫做实验,发现若加入 RNA 酶,则蛋白质合成会停止,Hall 和 Spiegelman 关于 T2 噬菌体 DNA-RNA 的杂交实验也发现,蛋白质合成的模板是 RNA,1958 年 Crick 提出著名的中心法则。1960 年 Weiss 和 Hurwitz 两个小组分别发现了 RNA 聚合酶,随后在真核生物中分离出多种 RNA 聚合酶,在原核生物和真核生物中分离出多种与转录相关的酶和蛋白质,并对这些酶和蛋白质的结构和功能进行了深入的研究,搞清楚了转录的基本过程。

关于基因的翻译,1954 年 Gamow 推测遗传密码是三联体,1961 年 Crick,Barrett 和 Brenner 等用插入和缺失突变证实了遗传密码为三联体。同年,Brewner,Jacob 和 Meselson 发现细菌的 mRNA,Nirenberg 开始用人工合成的核苷酸同聚物作为 mRNA 破译遗传密码。1964 年 Khorana 通过合成的核苷酸重复共聚物破译密码子,Nirenberg 等通过三联体结合实验破译密码子。1966 年遗传密码的破译工作基本结束,Crick 绘制了密码表,并提出了摆动学说。

1970 年前后,Howard Temin 和 David Baltimore 分别从致癌 RNA 病毒——劳氏肉瘤病毒(rous sarcoma virus)和鼠白血病病毒(murine leukemia virus, MuLV)中发现了逆转录酶(reverse transcriptase, RT)。逆转录酶的发现揭示了生物遗传中存在着由 RNA 形成 DNA 的过程,遗传信息不仅可以从 DNA 流向 RNA,也可以从 RNA 流向 DNA,进一步发展和完善了“中心法则”。

5. 基因工程的兴起和发展

基因工程的兴起是以对 DNA 结构和功能的深入研究,和一些工具酶的发现为基础的。1964 年 Holliday 提出了 DNA 重组模型,1967 年不同的实验室同时发现了 DNA 连接酶。

从 20 世纪 60 年代末开始,限制性内切酶等工具酶的发现,DNA 测序技术的建立,质粒载体、病毒载体的利用,终于在 70 年代中期诞生了基因工程。人们可以任意地将 DNA 基因元件切割、组建,并使指定的基因在不同的细胞中工作。

1965 年,瑞士微生物遗传学家 Werner Arber 首次从理论上提出了生物体内存在着一种具有切割基因功能的限制性内切酶(restriction enzyme, RE),并于 1968 年成功分离出 I 型限制性内切酶;1970 年,Hamilton O. Smith 分离出了 II 型限制性内切酶;同年,Daniel Nathans 使用 II 型限制性内切酶首次完成了对基因的切割。他们的研究成果为人类在分子水平上实现人工基因重组提供了有效的技术手段。

1975 年,Frederick Sanger 发明了确定 DNA 分子一级结构的末端终止法(酶法);1977 年,Walter Gilbert 发明了 DNA 一级结构测定的化学断裂法。其中,Sanger 发明的 DNA 测序法至今仍被广泛使用,经过改良之后的末端终止法是分子生物学研究中最基本、最常用的技术之一。Berg 把两个不同来源的 DNA 连接在一起并发挥其应有的生物学功能,证明了完全可以在体外对基因进行操作。

1973 年,S. N. Cohen 和 H. W. Boyer 等人将大肠杆菌中两种不同特性的质粒片段用内切酶和连接酶进行剪切和拼接,获得了第一个重组质粒,然后通过转化技术将它引入大肠杆菌细胞中进行复制,并发现它能表达原先两个亲本质粒的遗传信息,从而开创了遗传工程的新纪元。在此基础上,Boyer 于 1976 年成功地运用 DNA 重组技术生产出人的生长激素;1978 年,美国哈佛大学的科学家利用 DNA 重组技术生产出胰岛素;1980 年,瑞士和美国科学家利用 DNA 重组技术生产出干扰素。从此引发了 70 年代末、80 年代初的基因工程工业化的热潮。现代生物工程由

此崛起,它包括基因工程、细胞工程、酶工程与发酵工程、蛋白质工程等。到 20 世纪末,全世界已有 50 多个国家和地区拥有生物工程企业、生物工程产品不少于 160 种。这些最新成果已经对人类健康、生命质量、农业生产及其产品的加工产生了积极而深远的影响。

通过观察分子生物学的发展过程,可看到分子生物学是生命科学范围发展最为迅速的一个前沿领域,推动着整个生命科学的发展。至今分子生物学仍在迅速发展,新成果、新技术不断涌现,这也从另一方面说明分子生物学发展还处在初级阶段。虽然分子生物学已建立的基本规律给人们认识生命的本质找出了光明的前景,但分子生物学的历史还短,积累的资料还不够,对核酸、蛋白质组成生命的许多基本规律还在探索当中,目前还不能彻底搞清楚基因产物的功能、调控、基因间的相互关系和协调,因此,分子生物学还要经历漫长的研究道路。

1.2.3 分子生物学的发展基础

分子生物学是一门交叉学科,与其他学科互相促进,互相渗透。回顾分子生物学的发展历史,许多学科的成果都成为分子生物学发展的基础。分子生物学发展的历史充满着科学家的艰苦付出,也有很多逸闻趣事。

1. 遗传学基础

1859 年,达尔文发表了《物种起源》,提出了适者生存的进化理论。他认为,动植物在长期的生命过程中会发生一些微小的变化,其中一些动植物积累了这些变化,对环境更加适应,从而得到更好的生存与繁衍。

1856—1864 年,孟德尔(Mendel)的著名实验开创了现代遗传学,他提出了遗传的分离定律和独立分配定律,还提出了遗传因子的概念。

1903 年,Sutton 自己的工作解释了孟德尔的实验,并推测遗传因子是细胞中染色体的一个部分。尽管 Sutton 的试验并没有直接证明遗传的染色体理论,但非常重要,因为 Sutton 第一次把遗传现象与细胞学联系在一起。

1910 年以后,摩尔根(Morgan)等人用果蝇作试验时,发现有些基因在染色体上的距离近些,有些则远些。他们根据大量的试验构建了遗传图谱,并提出了连锁遗传规律。

1915 年,摩尔根证实了遗传的染色体基础。

不同的基因可能会位于同一条染色体上产生遗传的连锁现象,然而连锁通常并不完全。Janssens 提出了染色体交换理论,染色体在减数分裂过程的联会阶段发生断裂,然后交叉连接起来,造成了不完全连锁。

1931 年,Barbara McClintock 用玉米为材料证实了染色体的断裂重接。

一旦遗传的规律被阐明,就可以解释生物的变异现象和进化理论。但是,每个基因上发生的变化都很小,这些小的变化足以产生新的物种吗?Wright 等人认为:由于地球的年龄很大,又由于选择的压力很温和,一些小的有利的性状足以积累起来形成新的物种。到 20 世纪 40 年代,生物学家 Huxley、遗传学家 Dobzhansky、古生物学家 Simpson 和鸟类学家 Mayr 从各自的研究结果出发,都证实经典遗传学与进化论确实是一致的。

2. 物理学与生物化学基础

几乎就在孟德尔的遗传定律刚刚发现以后,遗传学家就开始思考基因的化学结构,以及基因是

如何工作的。但是在很长时间内都没有实质性的进展。因为当时核酸和蛋白质的结构都不清楚。

1927年,Muller和Stadler分别独立地发现了X射线可以诱导突变,由突变的频率可以估计一个基因的大小。以后的时间里,很多科学家都发现基因的突变可以影响到细胞中的蛋白质,由此发现基因与蛋白质之间具有一定的关系。此时,Beadle和Edward基于对红色面包霉的研究,提出了一个基因一个酶的假说。

但是,蛋白质的结构是什么?基因的结构是什么?基因是如何工作的?决定遗传信息的究竟是蛋白质还是基因?经典遗传学无法提供有效的研究手段来研究基因的化学本质,因此需要借助其他学科的方法。

1928年,Griffith证明灭活的致病性肺炎链球菌中的某些成分可以转化非致病性的肺炎链球菌。

1944年,Avery鉴定出了这种成分的化学本质,这是来自于致病性肺炎链球菌的DNA。Avery由此证明了DNA是遗传信息的载体。

1952年,Hershey和Chase用噬菌体也证明了DNA是遗传信息的载体。

一个多世纪以前,很多科学家认为研究生物学一定要用完整的活细胞,因为他们相信细胞具有某种“活力”,一旦细胞破碎,“活力”就会丧失,研究就无法进行。一旦科学家决定打开一个活细胞,分子生物学就诞生了。

1949年,生物化学家Chargaff研究了DNA的化学组成,提出了Chargaff规则。当蛋白质的结构开始用X射线分析时,一些科学家也在用X射线分析DNA的结构。

到了20世纪50年代,Wilkins和Franklin获得了DNA纤维结构高质量的X射线衍射图谱。这个图谱显示DNA是螺旋结构,而且可能由2条或3条多核苷酸链组成。1952年,有机化学家发现DNA的多核苷酸链间的连键是 $3' \rightarrow 5'$ 磷酸二酯键。

1953年,Watson和Crick推导出了DNA的双螺旋结构,并推测了DNA分子自我复制的机制。DNA双螺旋结构的发现是一场伟大的革命,从此DNA不再神秘,它与丙酮酸、甘油等一样,都是化学分子,都可以在实验室进行研究。

当Watson和Crick推导出了DNA的双螺旋结构的时候,Kornberg正在研究嘧啶和嘌呤核苷酸的合成。Kornberg和同事们发现了嘌呤和嘧啶核苷酸的从头合成途径以及这些途径中的很多酶。

1956年,Kornberg等人发现DNA合成的前体是四种脱氧三磷酸核苷,随后又发现了合成DNA的酶(现在我们知道这是DNA聚合酶I),并用纯化的DNA聚合酶合成了具有侵染性的噬菌体 Φ X174的DNA。

1958年,Meselson和Stahl发现DNA在复制的过程中两条链要分开,子代DNA分子中只有一条链是新合成的,一条链保留在子代分子中,就是半保留复制。

1968年,冈崎(Okazaki)提出DNA半不连续复制的模型。

双螺旋的发现彻底终结了DNA是否是遗传物质的争论。但是,DNA不可能是蛋白质合成的直接模板。因为在细胞中,活跃地合成蛋白质的位置上不存在DNA。况且真核生物的DNA位于细胞核中,而蛋白质合成是在细胞质中。那么,一定还有另外一种分子,它从DNA那里得到了遗传信息,然后移动到细胞质中作为合成蛋白质的模板。这种分子很可能就是第二类核酸——RNA。Casporsson和Brachet已经在细胞质中发现了大量的RNA。

但是,RNA的核苷酸顺序如何转变成蛋白质中的氨基酸顺序呢?一开始人们想象可能是

RNA 折叠起来形成一个疏水的空穴,空穴的形状恰好与一种特殊的氨基酸契合。但是 Crick 认为这根本不可能。因为从化学性质上看, RNA 中碱基上的基团更可能与水溶性的基团反应。其次,即使某些 RNA 片段能够形成疏水的空穴,这样的结构也无法区分 Gly 与 Ala,或者 Val 与 Ile。因为这 2 对氨基酸的侧链非常相似,各自只差一个甲基。在 1955 年, Crick 推测一定存在一种接头分子(adaptor),这种分子既可以识别核酸,又能够连接氨基酸,它很可能也是一种 RNA。Crick 在 1956 年又提出了中心法则:遗传信息可以自我复制,可以从 DNA 流向 RNA,然后再流向蛋白质,这就构成了中心法则的基本内容。

1953 年, Zamecnik 等发展了在体外用无细胞系统合成蛋白质的体系。几年以后,他们又发现在合成蛋白质以前,氨基酸首先要连接到 tRNA 上;同时还发现了催化这个反应的是氨酰-tRNA 合成酶。tRNA 就是 Crick 预言的接头分子, tRNA 的一端用反密码子与蛋白质合成的模板识别,另一端连接氨基酸。

蛋白质合成的模板是什么?最初有人认为是 rRNA。但是经过仔细研究之后发现这是不可能的。第一,核糖体由大、小两个亚基组成,每个亚基都含有 rRNA;第二,几乎在所有的细菌、植物和动物中,所有小亚基中的 rRNA 大小都非常相似;第三,尽管相应的 DNA 中的 AT/GC 比例不同,大、小 rRNA 的碱基组成几乎相同。这种相似性怎么可能指导合成大量的蛋白质呢?所以, rRNA 不可能是蛋白质合成的模板。现在我们知道 rRNA 是核糖体的组成成分之一。

1960 年, Brenner 及 Gross 等人发现,用噬菌体 T2 感染大肠杆菌后,细菌不再合成自己的 RNA,只从 T2 的 DNA 上转录噬菌体的 RNA。转录出来的 T2 RNA 不是与核蛋白结合形成核糖体,而是附着在核糖体上,然后沿核糖体的表面移动到可以结合氨酰-tRNA 的位置上去合成蛋白质。因为 T2 RNA 从 DNA 上接受遗传信息,又转移到核糖体上合成蛋白质,所以把它叫做信使 RNA(mRNA)。从此发现, mRNA 才是蛋白质合成的模板。

1958 年, Weiss 和 Hurwitz 的实验室分别独立地发现了依赖于 DNA 的 RNA 聚合酶。

遗传密码的破译是分子生物学的伟大成就之一。1961 年, Nirenberg 用体外蛋白质合成系统,加入人工合成 poly U 和各种同位素标记的氨基酸,得到的是多聚苯丙氨酸的肽链。这个实验证明了 UUU 是苯丙氨酸的密码子,还证明可以用合成的 RNA 链作为蛋白质合成的模板。后来他用同样的方法证明了 AAA 编码 Lys, CCC 编码 Pro。但 GGG 容易形成三股螺旋,后来使用另外的方法才证明 GGG 编码 Gly。

Nirenberg 多核苷酸磷酸化酶合成 RNA 链。这个酶合成 RNA 时不需要模板,当使用混合的核苷酸为底物时,模板的序列是随机的。虽然可以根据加入反应体系中的 NTP 的比例来估计核苷酸在密码子中的比例,但这个方法不能确定密码子的碱基顺序。

同时,有机化学家 Khorana 合成了具有确切碱基顺序的多聚核糖核苷酸。进一步证实每个密码子由三个核苷酸组成,但由于不能确定确切的起始核苷酸,还是不能确认到底哪一个密码子对应于哪一个氨基酸。

1964 年, Nirenberg 发现以三核苷酸为模板就足以使相应的氨酰-tRNA 结合到核糖体上,但是三核苷酸还是随机的片段。Khorana 将自己的方法与 Nirenberg 的方法结合,很快解读了约 50 个密码子。后来又破译了其他的密码子,并发现 3 个密码子是终止密码子, AUG 既是 Met 的密码子,也是起始密码子。这些发现也包括其他实验室的工作。

上述重要发现共同建立了以中心法则为基础的分子遗传学基本理论体系。1970 年 Temin

和 Baltimore 又同时从鸡肉瘤病毒中发现了以 RNA 为模板合成 DNA 的逆转录酶, 又进一步补充和完善了遗传信息传递的中心法则。

操纵子学说使人们开始了解基因表达调控。最初发现了原核生物基因表达调控的一些规律, 后来研究逐渐扩展到真核基因表达调控机理, 真核生物发育与细胞分化的机理。目前的研究已经使我们了解到真核生物基因表达调控机理在于: 基因的顺式调控元件与反式作用因子的相互识别与作用; 核酸与蛋白质之间的相互识别与作用; 蛋白质与蛋白质之间的相互识别与作用。

在研究基因表达调控的过程中, 细胞核内及其他小分子 RNA 具有特殊功能。还发现了具有催化活性的 RNA, 即核酶。1995 年, Guo 等发现一些短的 RNA 片段可以造成同源的 mRNA 降解, 从而调节基因表达, 叫做 RNA 沉默(RNA silence)。

3. 技术方法的基础

1967—1970 年, R. Yuan 和 H. O. Smith 等发现的限制性核酸内切酶为基因工程提供了有力的工具。

1975—1977 年, Sanger、Maxam 和 Gilbert 先后发明了 DNA 序列的快速测定法。

1985 年, Mullis 等发明聚合酶链式反应(polymerase chain reaction, PCR), 这种特定核酸序列扩增技术以其高灵敏度和特异性被广泛应用, 对分子生物学的发展起到了重大的推动作用。目前分子生物学已经从研究单个基因发展到研究生物整个基因组的结构与功能。

20 世纪 90 年代, 全自动核酸序列测定仪问世。

新技术的不断涌现促进了分子生物学的不断进步。现在, 打破种属界限, 将不同来源的 DNA 在体外重组, 从而大量获得目标蛋白已经成为一种常规的操作。

1.2.4 分子生物学的发展趋势

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. cerevisiae*)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Macquarie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 proteomics)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。

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,peptide 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中,已于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。