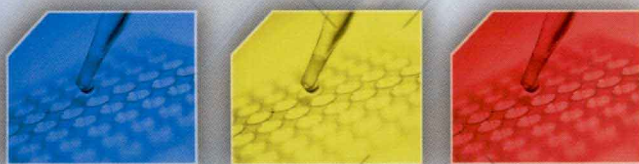


国家“十一五”重点图书出版项目
——生物医学实验技术系列丛书

JIYINZUXUE

基因组学

宋方洲◎主编



国家“十一五”重点图书出版项目
——生物医学实验技术系列丛书

基因组学

主 编:宋方洲

副主编:卜友泉 彭惠民

编 委:(按姓氏拼音排序)

卜友泉(重庆医科大学)	郭风劲(重庆医科大学)
马永平(重庆医科大学)	彭惠民(重庆医科大学)
宋方洲(重庆医科大学)	汤 华(重庆医科大学)
武向梅(重庆医科大学)	易发平(重庆医科大学)
朱慧芳(重庆医科大学)	周 兰(重庆医科大学)
张雪梅(重庆医科大学)	

军事医学科学出版社
· 北 京 ·

图书在版编目(CIP)数据

基因组学/宋方洲 主编. -北京:军事医学科学出版社,2011.9

ISBN 978-7-80245-775-1

I. ①基… II. ①宋… III. ①基因组—研究生—教材

IV. ①Q343.1

中国版本图书馆CIP数据核字(2011)第139588号

出 版:军事医学科学出版社

地 址:北京市海淀区太平路27号

邮 编:100850

联系电话:发行部:(010)66931051,66931049,81858195

编辑部:(010)66931127,66931039,66931038,

86702759,86703183

传 真:(010)63801284

网 址:<http://www.mmssp.cn>

印 装:北京宏伟双华印刷有限公司

发 行:新华书店

开 本:787mm×1092mm 1/16

印 张:19

字 数:314千字

版 次:2011年9月第1版

印 次:2011年9月第1次

定 价:48.00元

本社图书凡缺、损、倒、脱页者,本社发行部负责调换

内容简介

本书是在当前生命科学迅猛发展以及我国研究生教育变革的背景下,为适应研究生培养课程体系和教学内容改革的新形势而编写的。编者均为长期工作在科研教学一线的教师。全书内容包括十二章,其中第一章至第五章的内容主要为基因组学的理论基础,第六章至第十二章的内容主要为基因组学研究的相关技术与方法。

本书不仅可以作为高等院校生物医学相关专业研究生和高年级本科生的教材使用,而且对于从事生物医学相关专业的研究人员、青年教师和学生也颇具参考价值。

前 言

生命科学实际上就是围绕其重要的基石——中心法则,进行生命科学3个层次的研究,即从DNA→RNA→蛋白质,从而形成了3个层面的科学:即基因组学、转录组学和蛋白质组学,这是目前生命科学研究最重要、最热门和最前沿的3个研究领域。

为了跟踪和把握生命科学发展的特点和趋势,适应国家教育教学改革及国家对研究生培养的要求,重庆医科大学率先进行了研究生培养课程体系和教学内容的改革,将研究生的专业基础课和专业课课程归纳合并为基因组学、蛋白质组学、生物信息学和分子病理学4门生命科学前沿课程,并集中学校的优势资源实行课程负责人制,依托国家级重点学科(临床检验诊断学)和省部级重点学科(生物化学与分子生物学、遗传学等),组成教学团队。领衔主讲的成员都是工作在教学科研第一线的教师,大多都具有博士学位和出国留学经历,具有开拓和创新精神,工作富有成效,教师将丰富的科研实践知识和创新体验融入到教学中,提高了教学质量、增强了教学效果,使得重庆医科大学的研究生课程体系改革和研究生教育上了一个新的台阶。

基因组学是生命科学中心法则中最核心、最基础的部分,是中心法则的上游,是研究生物全套基因的组成、基因组内各基因的精确结构、相互关系及其表达调控的科学。

但非常遗憾的是,到目前为止,在国内外还没有一本适合研究生层次的《基因组学》的教材,所以我们在承担重庆医科大学的研究生基因组学课程的基础上,汇总了各任课教师讲课的内容,经过了3轮修改、补充和完善,编写成了这本《基因组学》教材。

本教材遵循了编写教材应具有的知识性、系统性、科学性、前瞻性与实用性的原则。“授人以鱼不如授人以渔”,考虑到研究生已经过4~5年本科阶段的学习,初步具备生命科学的基本知识,要深入掌握生命科学的前沿知识,才能凝炼出本学科最需要深入研究的课题。而且要完成研究生,尤其是博士生的研究课题,必须要有好的思路和实验设计以及切实可行的先进技术路线和方法。

本课程的主要教学目标:使研究生掌握基因组学的基本概念;理解基因组学重要的基本原理;了解基因组学研究前沿;开阔视野、把握世界生命科学的潮流,为占领生命科学的制高点打下坚实的基础,使研究生能联系本人专业或研究方向提出与基因组学相关的问题,并学会找到研究这些问题的方法。通过深入浅出的讲解,以及师生良好的互动,调动学生学习的积极性,启

发学生的潜能,有利于联系临床实践开展研究生课题的选题与研究。

本课程的教学内容体系共分为 12 章。实际上可以归纳为两个部分,第一部分为基因组学的理论基础;第二部分为基因组学研究的相关技术与方法。

本教材在编写上除了内容上注意特色和优势外,在编写形式上也有所创新,第一章没有采用常规的“绪论”,而是用“概论”,旨在从一开始就能开阔学生视野,把握整个学科的全貌,跟踪学科的前沿,而在应用技术的章节标题都加入了“……技术的应用与发展”,这样更能够使学生学习、了解和掌握该门技术的实用性及其新的发展,使学生能够站在一个更高的层面上来学习了解这门科学。

本教材最主要的参考文献是杨金水老师编著的《基因组学》、贺林院士编著的《解码生命——人类基因组计划和后基因组计划》等,很多理念、构思、结构都是从他们的书籍中受益或得到启发的,甚至很多内容和图表都是引用他们的,在此,特向他们表示崇高的敬意和衷心的感谢。同时,因为在学习的过程中,可能存在对他们的内容理解不到位或不深刻的地方,还请他们多多包涵和原谅。

由于本教材是汇总了各任课教师的讲课内容,经过反复修改、补充、完善而成,我们尽可能保留了各位专家教授的特点和个性。因编者水平有限,时间仓促,又是第一本研究生层次的《基因组学》教材,不当、遗漏或错误之处在所难免,恳请同行专家和同学们批评指正,以便再版和重新印刷时修改完善。

编 者

2011 年 5 月于重庆

目 录

第 1 章 基因组学概论	(1)
1 基因组学概况	(1)
2 基因组学发展的历史沿革	(6)
3 后基因组学产生的背景	(18)
4 后基因组学的主要研究内容与发展趋势	(20)
第 2 章 基因组多态性	(26)
1 限制性片段长度多态性	(26)
2 小卫星 DNA	(27)
3 微卫星 DNA	(30)
4 单核苷酸多态性	(33)
5 表达序列标签和序列标签位点	(35)
6 表观遗传学	(36)
第 3 章 基因组作图	(42)
1 基因组遗传图	(43)
2 基因组物理图	(45)
3 基因组序列图	(50)
4 基因图谱	(51)
第 4 章 疾病基因组学	(56)
1 疾病基因组研究概述	(56)
2 疾病基因组研究策略	(58)
3 单基因病的基因组学研究	(63)
4 复杂性疾病的基因组学研究	(70)
5 肿瘤基因组学研究	(75)
第 5 章 基因组的进化与分子系统学	(83)
1 基因组进化的分子基础	(83)
2 基因组进化的模式	(84)
3 基因组与生物进化	(90)
第 6 章 重组 DNA 技术及其应用	(120)
1 重组 DNA 技术	(120)

2 重组 DNA 技术的基本步骤	(124)
3 重组 DNA 技术的发展	(143)
4 重组 DNA 技术在医学领域的应用	(144)
第 7 章 分子杂交与印迹技术	(155)
1 分子杂交与印迹技术简介	(155)
2 探针的种类及其制备	(159)
3 常用分子杂交与印迹技术简介	(163)
第 8 章 PCR 技术及其相关技术的发展和應用	(173)
1 PCR 技术简介	(173)
2 基于 PCR 的相关技术	(181)
3 荧光定量 PCR	(194)
4 PCR 技术的临床应用	(198)
第 9 章 转基因动物技术的应用与发展	(202)
1 转基因动物概述	(202)
2 转基因动物的发展状况与趋势	(206)
3 转基因动物技术	(207)
4 转基因动物制备	(214)
5 转基因动物的应用与未来	(221)
第 10 章 生物芯片的应用与发展	(228)
1 生物芯片的分类	(229)
2 DNA 芯片的概念和基本原理	(229)
3 DNA 芯片的种类	(233)
4 数据分析	(234)
5 基因芯片技术在医学领域的应用	(236)
第 11 章 新基因功能研究的策略与方法	(251)
1 新基因功能研究的基本思路	(251)
2 新基因功能研究的主要策略与方法	(253)
3 一些需要注意的问题	(259)
第 12 章 RNAi 技术的应用与发展	(262)
1 RNAi 的发现	(262)
2 RNAi 的发生机制	(264)
3 RNAi 作用的特点	(266)
4 RNAi 研究的技术方法	(269)
5 RNAi 技术的应用	(281)
6 siRNA 与 miRNA 的比较	(285)
术语索引	(288)

第 1 章

基因组学概论

生命科学实际上就是围绕其重要的基石——中心法则,而进行生命科学三个层次的研究,即从 DNA→RNA→蛋白质,从而形成了三个层面的科学,即基因组学、转录组学和蛋白质组学,这是目前生命科学研究最重要、最热门,也是最前沿的三个研究领域。

1 基因组学概况

1.1 “组学”的概念

DNA 结构的发现与中心法则的建立和完善,以及随后发明的 DNA 重组和分子克隆等技术,使人类获得了崭新的干预生物进化和优化生物某些功能的有效手段和途径。现在,生命科学的研究继续在分析和综合两个层面上不断深入和拓展。虽然传统的基于相对独立的研究小组或实验室的研究活动仍然是主要的工作方式,但随着生命科学研究内涵的丰富和范围的扩大,多个实验室间的合作研究成为当前的新潮流,大规模的跨单位、跨地区、跨国家的联合研究亦呈现强劲的势头。此外,由于复杂系统理论和非线性科学的发展,生命科学研究正从局部观向整体观转变,从线性单向思维走向复杂性思维,从单一学科为主转变为多学科的交叉融合。随着计算生物学和系统生物学等新兴学科的不断涌现,生命科学大综合和大发展的时期即将或已经来临。在此过程中,“组学”的概念应运而生。

研究细胞、组织或是整个生物体内某种分子(DNA、RNA、蛋白质、代谢物或其他分子)的所有组成内容,称之为“组学”(- omics)。“ - omics”术语代表了对怎样思考生物学和生物体系工作方式的重新定位。组学研究包括

对基因组及基因产物(转录子组和蛋白质组)的系统生物学研究,随后必然要上升到细胞机制、分子机制和系统生物学的水平。由于基因组的信息是用来发现和解释具有普遍意义的生命现象和它们的变化、内在规律和相互关系,因此,学科的交叉合作就成为必然,同时,基因组的复杂性必然导致多学科的介绍和介入,例如,各生物学科、医学、药学、计算机科学、化学、数学、物理学、电子工程学、考古学和地质学等。

从科学的历史来看,形成一门学科并非一件容易的事。但在人类基因组计划(human genome project, HGP)实施的短短几年间,各种“组学”迅速在生命科学界蔓延。最早出现的是与 DNA 相关的“基因组学”(genomics),人类基因组计划完成以后产生了“后基因组学”(post-genomics),随后又形成了许多与各种生物大分子或小分子相关的“组学”。

在生命科学基础研究中,最活跃的前沿主要包括:分子生物学、细胞生物学、神经生物学、生态学等学科,并由此引申出相应系统性、整体性的组学,诸如:基因组学(genomics)、结构基因组学(structural genomics)、功能基因组学(functional genomics)、转录组学(transcriptomics)、蛋白质组学(proteomics)、比较基因组学(comparative genomics)、代谢组学(metabonomics)、糖组学(glycomics)等重要领域。

本章仅对基因组学、后基因组学等几种重要组学进行简要阐述。

1.2 基因、基因组及基因组学的基本概念

1.2.1 基因(gene)

基因是 1909 年丹麦植物学家 W. Johannsen 根据希腊文单词 *genos* (birth, 给予生命)创造的。

现代分子生物学的基因概念通常认为是:基因是储存和表达某一多肽链信息或 RNA 分子信息所必需的全部核苷酸序列,是生物体传递和表达遗传信息的基本单位。

1.2.2 基因组(genome)

基因组一词系由德国汉堡大学 H. Winkles 教授于 1920 年首创。最初是由 GENE 和 chromOSOME 组成为“genome”,用于表示生物的全部基因和染色体组成的概念,是指生物的整套染色体所含有的全部 DNA 序列。现在一般认为,基因组即指生物所具有的携带遗传信息的遗传物质的总和。包括所有的基因和基因间区域。如人类基因组包含了细胞核染色体(常染色体和性染色体)及线粒体 DNA 所携带的所有遗传物质,其中核基因组大约有

28.5 亿个碱基对,2~2.5 万个蛋白编码基因,这些编码区仅占整个基因组很少一部分(不到3%),而大部分为非编码区。

基因组的结构主要指核酸分子中不同的基因功能区域各自的分布和排列情况,其功能是储存及表达遗传信息。不同种类生物储存的遗传信息量迥异,其基因组的结构和组织形式也不同。在对生命现象(包括疾病)的研究中,除需要继续研究个别基因的结构和功能外,更应该了解整个基因组及其产物是如何协同作用调节生物整体或细胞活性的。DNA 重组等分子生物学技术的发明和应用极大程度上促成和推动了1990年正式启动的“人类基因组计划(HGP)”,该计划的实施和完成促进了基因组学的形成和发展。

(1) 真核生物基因组

- ①核基因组:细胞核内所有遗传物质的总和。
- ②线粒体基因组:线粒体携带遗传物质的总和。
- ③叶绿体基因组:叶绿体携带遗传物质的总和。

(2) 原核生物基因组

①染色体:是原核细胞内的主要遗传物质。

②质粒:是独立于细菌染色体的自主复制的环状双链DNA分子。能稳定地独立存在于染色体外,并传递到子代,一般不整合到宿主染色体上,通常编码毒素和耐药性等相关基因。

(3) 病毒基因组

病毒颗粒携带的遗传物质。

1.2.3 基因组学(genomics)

所谓基因组学就是对所有基因进行基因作图(包括遗传图谱、物理图谱、转录图谱)、核苷酸序列分析、基因定位和基因功能分析的一门科学,简言之,就是在基因组水平上研究基因组结构和功能的科学。基因组学研究的内容包括基因的结构、组成、存在方式、表达调控模式、基因的功能及相互作用等,是研究与解读生物基因组所蕴藏的生物全部性状的所有遗传信息的一门新的前沿科学。1986年美国科学家Thomas Roderick首次提出了基因组学的概念,但随着1990年人类基因组计划启动才开始真正有系统地研究基因组、解码生命,并由“后基因组计划”的实施推动其发展。作为当今最活跃的学科之一,基因组学已经对生物学、遗传学、医药学乃至整个人类社会产生深远影响。

1.3 基因组学的分支

根据基因组学的定义,基因组学包括结构基因组学、功能基因组学和比较基因组学。

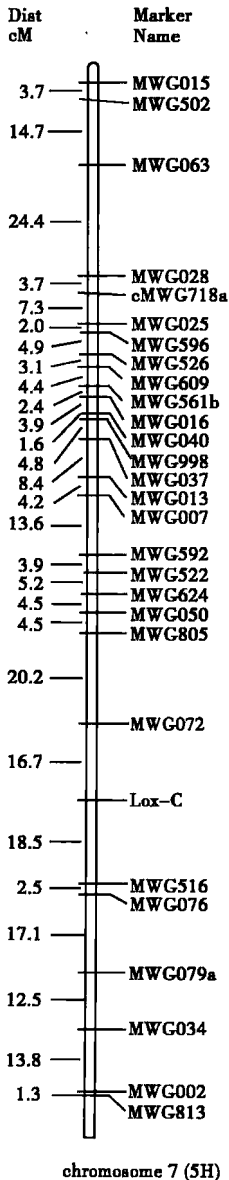


图 1-1 人第 7 号染色体的基因组图谱

1.3.1 结构基因组学

通过基因组作图、核苷酸序列分析,研究基因组结构,确定基因组成、基因定位的科学。结构基因组学的主要研究内容:

(1) 基因组测序

目前 DNA 测序每次反应仅能读取不到 1000bp 的长度。已知最小的细菌基因组为 580 000bp。人类基因组为 30 亿 bp。因此,首先将整个基因组的 DNA 分解为一些小片段,然后将这些分散的小片段逐个测序,最后将测序的小片段按序列组装。

(2) 基因组作图

在长链 DNA 分子的不同位置寻找特征性的分子标记,绘制基因组图。根据分子标记可以准确无误地将已测序的 DNA 小片段锚定到染色体的位置上。

1.3.2 功能基因组学

利用结构基因组学提供的信息和产物,在基因组系统水平上全面分析研究基因功能的科学。功能基因组学的主要研究内容:

(1) 进一步识别基因以及基因转录调控信息。

(2) 弄清所有基因产物的功能,这是目前基因组功能分析的主要层次。

(3) 研究基因的表达调控机制,研究基因在生物体发育过程以及代谢途

径中的地位,分析基因、基因产物之间的相互作用关系,绘制基因调控网络图。

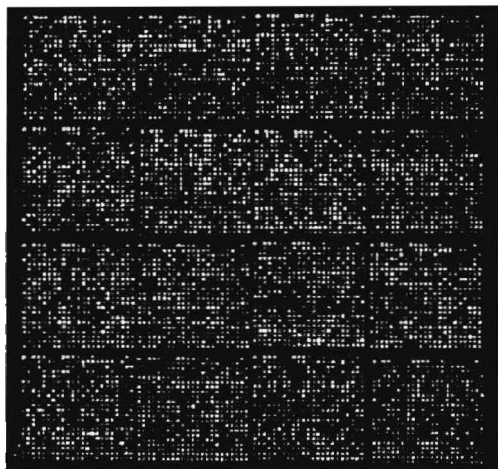


图 1-2 研究功能基因组学基因芯片图谱

1.3.3 比较基因组学

比较研究不同生物、不同物种之间在基因组结构和功能方面的亲源关系及其内在联系的学科。比较基因组学的主要研究内容:

(1)通过研究不同生物、不同物种基因组结构和功能上的相似及差异,可以勾画出一张详尽的系统进化树,而且还将显示进化过程中最主要的变化所发生的时间及特点。据此可以追踪物种的起源和分支路径。

(2)可以分析了解同源基因的功能。

(3)通过对序列差异性的研究有助于认识大自然生物多样性的产生基础。

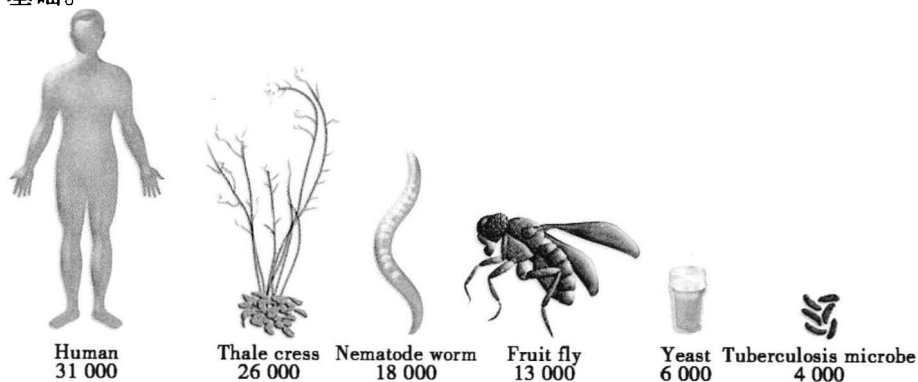


图 1-3 几种生物基因组学比较模式图

1.4 研究基因组学的意义

基因组学的研究已经涉及到生物学、医学、生物技术、制药工业、社会经济、生物进化、伦理、法律及社会等众多领域。尤其在人类疾病基因的研究方面,显现和发挥着十分重要的作用。

(1) 疾病的遗传学基础。

(2) 致病基因及相关基因的克隆在基因组学研究中占据着核心位置。

(3) 对疾病的预防、诊断、治疗等有重要意义。

(4) 人类基因组计划的直接动因是要解决包括肿瘤在内的人类疾病的遗传学基础问题。

1.4.1 单基因病疾病基因研究,如血友病等

(1) 人类基因组计划使我们了解基因组序列。

(2) 现在采用定位候选克隆方法,极大地提高了发现疾病基因的效率。

(3) 定位候选克隆,通过遗传分析等方法将疾病基因定位到染色体区段上。

(4) 对人类基因组图上该区段内的基因进行功能分析,并筛选出疾病基因,为这些疾病的基因诊断和基因治疗奠定了基础。

1.4.2 多基因病疾病基因研究,如心脏病、糖尿病、癌症等

(1) 比单基因病困难,是目前疾病基因研究的重点和难点。

(2) 用比较基因表达谱的方法来识别疾病状态下基因的激活或抑制。

1.4.3 癌症基因组解剖学计划(cancer genome anatomy project, CGAP)

(1) 1996年癌症基因组解剖学计划开始。

(2) 主要由美国癌症研究所(National Cancer Institute)开展。

2 基因组学发展的历史沿革

形成比较完整的基因组学学科无非是最近一二十年的事,但它的孕育、产生和发展却经历比较长的时间。根据该门学科的发展情况,我们似乎可以将其划分为下面5个历史发展阶段。

2.1 1900年以前——前遗传学时代

(1) 1859年 Darwin C 在“On the Origin of Species”这一名著中,提出了物

种进化的自然选择学说——达尔文进化论。我们似乎可以将达尔文进化论作如下概括：

生物的来源——进化
 生物的性状——遗传
 进化的动力——选择

(2)1865年 Mendel G 发表豌豆杂交实验结果,提出了遗传学的两大遗传规律——分离规律和独立分配规律(分离规律的实质:决定一对性状的一对基因位于一对染色体上,这对基因随着染色体的分离而分离;独立分配规律的实质:决定两对形状的两对基因位于两对染色体上,它们随着染色体的分离而分离,随着染色体的组合而组合,这种组合是随机的自由组合),并认为是生物体内的遗传因子或遗传颗粒控制生物性状。但他的结论直到1900年才被得到认可和重视(被称为孟德尔规律再发现或重新发现)。

2.2 1900 ~1950 年——经典遗传学时代

标志:1900年孟德尔遗传规律再发现标志着遗传学的诞生。

孟德尔规律重新发现以后,人们开始把控制遗传性状的遗传单位称为基因(gene)。生命科学的研究工作基本上都是围绕着基因来进行。

(1)1839年细胞学说的提出。

(2)1869年发现DNA,随后RNA也被发现。

(3)1879年染色体的发现,并认为染色体最可能是DNA、RNA和蛋白质的一种。

(4)1902年染色体学说的产生,合理地解释了Mendel的实验结果。

(5)1910年发现了遗传学的第三大遗传规律——连锁遗传规律(连锁遗传规律的实质:决定两对性状的两对基因位于同一对染色体上,就会发生连锁遗传现象),并证明基因的确存在于染色体上,并呈线状排列。

(6)1944年证实了DNA是携带遗传信息、构成染色体的生物大分子。这个观点到今天仍然是正确的。

总之,自从人们认识到“基因决定生物性状”、“基因的本质就是核酸——主要是DNA”之后,就从来没有停止对基因的研究,最先是对基因结构的研究。

2.3 1950 ~1990 年——分子生物学时代(前基因组学时代)

标志:Watson & Crick的DNA双螺旋的发现[“Nature”1953.4.25],标志

着分子生物学时代的开始。

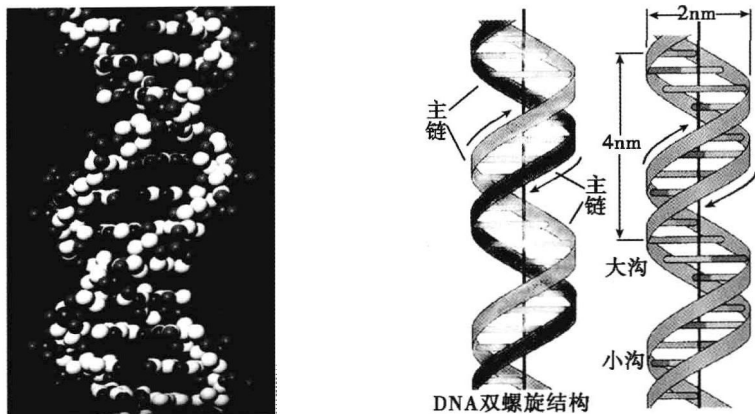


图 1-4 DNA 双螺旋结构与平面图

1953 年 Watson J, Crick F 借助于几个实验室的研究成就,根据 DNA 的 X 射线衍射图谱,提出了 DNA 双螺旋结构模型,用分子结构的特征解释了生命现象的最基本问题——基因复制的机制,从而真正开始从分子水平上研究生命活动。生物学研究也从此进入了分子生物学时代。

1953 ~ 1970 年,随着 DNA 双螺旋结构的发现和确定,带来了生物学的迅猛发展,也就是分子生物学理论和技术体系逐步形成的时期。在这一时期,分子生物学取得了一系列重大突破:

(1) mRNA、DNA 聚合酶、RNA 聚合酶的发现。

(2) DNA 半保留复制机理的阐明。

(3) 操纵子学说等的提出。

(4) 遗传密码的发现,其通用性的证明。

(5) 1966 年,64 个密码子破译。

(6) 中心法则:DNA→RNA→蛋白质建立,标志着分子生物学学科理论体系形成。

(7) 重组 DNA 技术的建立,使得分子生物学的技术体系初步形成,或者说,生命科学进入了分子生物学或前基因组学时代。

不仅如此,20 世纪 70 年代以后,分子生物学飞速发展,理论体系和技术体系得到不断完善。

2.3.1 逆转录酶的发现,修正和完善了中心法则

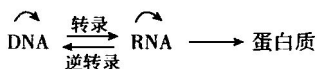


图 1-5 修正后的中心法则

2.3.2 Prion 蛋白特性的发现进一步推动了中心法则的演变

20 世纪 80 年代初期,非常厉害的疯牛病在欧洲大陆肆虐。后来,人们发现以疯牛病和人类克雅病(Creutzfeldt - Jakob disease, CJD)为代表的可传染性海绵状脑病(transmissible spongiform encephalopathies, TSE)是一类致死性神经系统疾病,其感染因子是一种与正常细胞膜蛋白(PrP^C)一级结构完全相同,但高级结构、理化特性不同的 PrP^{Sc} 或 prion 蛋白(阮蛋白)。目前的研究提示,这种不含有核酸的 prion 蛋白似乎具有自身复制的能力,也就是说这种蛋白本身就储备了生命活动必需的遗传信息。这就提示我们,在自然界中可能还存在着一种全新的微小生命体——一种无核酸的微小生物体,从而改变人类对整个生物界的认识。即蛋白质不仅可体现生物学功能,而且可储存遗传信息。而核酸成分可能在 prion 复制过程中完全缺如,这将有力地挑战目前的“生物中心法则”。PrP^C 蛋白高级结构的改变不仅可造成蛋白质功能的变化,在某种意义上还是遗传信息传递的方式。

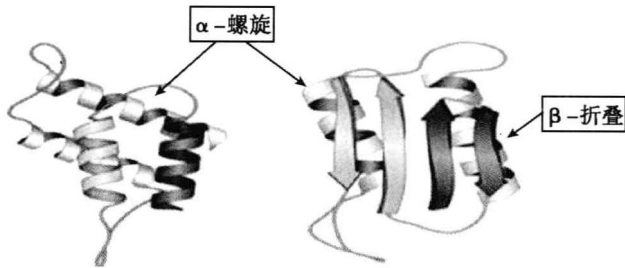


图 1-6 正常细胞膜蛋白与阮(prion)蛋白结构图

左:正常细胞膜蛋白 PrP^C(Cellular prion protein)

右:致病蛋白 PrP^{Sc}(Scrapie - associated prion protein)

根据上面这些推理,比较完善或期望的中心法则似乎应该作如下描绘:

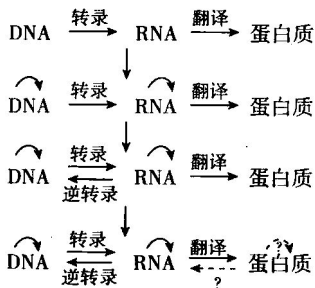


图 1-7 中心法则的发展与演变