



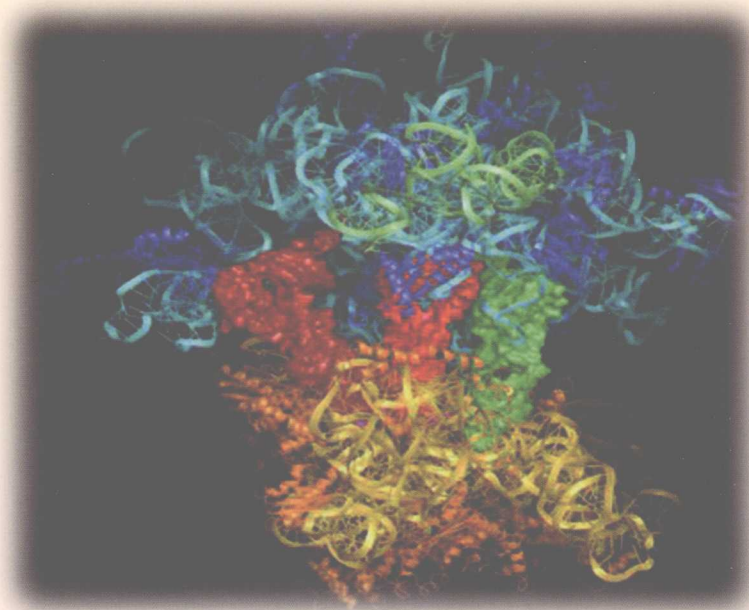
中国科学院教材建设专家委员会规划教材
全国高等医学院校规划教材

供临床、预防、基础、口腔、麻醉、影像、药学、检验、
护理、法医等专业使用



医学分子生物学

德 伟 欧 芹 主编



科学出版社
www.sciencep.com

中国科学院教材建设专家委员会规划教材
全国高等医学院校规划教材

案例版™

供临床、预防、基础、口腔、麻醉、影像、药学、检验、护理、法医等专业使用

医学分子生物学

主 编 德 伟 欧 芹

副主编 陈建业 张吉林 万福生 葛银林

编 者 (以姓氏笔画为序)

万福生 南昌大学医学院

王 杰 沈阳医学院

王晓华 广州医学院

杜培革 北华大学

杨红英 昆明医学院

李艳丽 南京医科大学

肖建英 辽宁医学院

宋玉国 北华大学

张吉林 北华大学

张春举 南京中医药大学

陈建业 川北医学院

欧 芹 佳木斯大学医学院

易光辉 南华大学医学院

咎玉玺 新乡医学院

葛银林 青岛大学医学院

科 学 出 版 社

北 京

郑 重 声 明

为顺应教育部教学改革潮流和改进现有的教学模式,适应目前高等医学院校的教育现状,提高医学教学质量,培养具有创新精神和创新能力的医学人才,科学出版社在充分调研的基础上,引进国外先进的教学模式,独创案例与教学内容相结合的编写形式,组织编写了国内首套引领医学教育发展趋势的案例版教材。案例教学在医学教育中,是培养高素质、创新型和实用型医学人才的有效途径。

案例版教材版权所有,其内容和引用案例的编写模式受法律保护,一切抄袭、模仿和盗版等侵权行为及不正当竞争行为,将被追究法律责任。

图书在版编目(CIP)数据

医学分子生物学:案例版/德伟,欧芹主编. —北京:科学出版社,2008
中国科学院教材建设专家委员会规划教材·全国高等医学院校规划教材
ISBN 978-7-03-021619-9
I. 医…II. ①德…②欧…III. 医药学:分子生物学-医学院校-教材 IV. Q7
中国版本图书馆 CIP 数据核字(2008)第 049051 号

策划编辑:胡治国 / 责任编辑:胡治国 / 责任校对:郑金红
责任印制:刘士平 / 封面设计:黄 超

版权所有,违者必究。未经本社许可,数字图书馆不得使用

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>

天时彩色印刷有限公司印刷

科学出版社发行 各地新华书店经销

*

2008 年 6 月第 一 版 开本:850×1168 1/16

2008 年 6 月第一次印刷 印张:16

印数: 1—5 000 字数:527 000

定价:39.00 元

(如有印装质量问题,我社负责调换〈环伟〉)

前 言

本教材借鉴案例教学法的教学模式,即以案例为教材,让教育对象通过阅读、分析和思考以及相互间进行讨论和交流,以提高思维推理和处理问题能力的教学过程。在教材中增加临床真实案例或标准化案例,并结合理论知识进行分析、归纳,加强基础与临床、临床与实践的联系,丰富教学内容,充分调动学生学习的积极性、主动性和创造性,激发学生进行不断学习的内在动机和热情,给学生创造了思考问题的条件,把学到的分子生物学理论灵活地运用到临床实际中,从而增强学生分析问题、解决问题和适应临床实际工作的能力。案例教学在医学教育中,是培养高素质、创新型和实用型医学人才的有效途径。本教材是在传统教材的基础上应用案例教学法的原理进行的新的尝试,在教材编写过程中对教材内容的选择和组织还缺乏经验,还存在一些待完善的地方,我们将在今后的工作中不断改进。

分子生物学是在分子水平上研究生命本质、生命活动和生命现象的科学,其理论和技术已在医学领域得到广泛应用,是医学院校学生的一门必修课。分子生物学理论和技术是生命科学中发展最快,与临床医学有着密切的联系,不断更新分子生物学的新理论及技术,进而丰富医学科学内容,是医学分子生物学今后有待解决的问题。医学院校学生学习分子生物学既要掌握分子生物学的基础理论和基本技术,也要掌握分子生物学在医学领域的应用和研究进展。

教材第一篇主要对 DNA、RNA 和蛋白质等生物大分子的结构和功能,遗传信息的传递及调控,细胞信号传导及其分子机制等分子生物学的经典理论结合教学难点案例进行了深入浅出的论述,便于学生掌握分子生物学基本理论和概念。

第二篇重点介绍了临床常见病如肿瘤、心血管病、遗传病、代谢病、感染性疾病的分子生物学基础及其分子机制,以临床案例形式提出问题,帮助学生从分子生物学的角度加深对上述疾病发病机制的认识,从基因的突变、基因的多态性和个体基因与环境相互作用等方面探讨各种致病因子对疾病的影响,指导临床诊断和治疗实践。

第三篇介绍了临床医学分子生物学常用技术,如疾病基因检测方法(DNA、RNA 定性、定量分析、基因芯片分析)、疾病基因的克隆、诊断以及基因工程药物与疫苗的研发及临床应用。近些年来,人们发现并克隆了多种人类肿瘤相关抗原基因,逐步了解了肿瘤抗原的识别和 T 细胞活化机制,并已能成功地大量培养和扩增树突状细胞(dendritic cell, DC),从而使以 DC 为基础的抗肿瘤主动免疫治疗有了新的发展。

随着近代医学的发展,越来越多的分子生物学的理论和技术应用于疾病的预防、诊断和治疗,从分子水平探讨各种疾病的发生发展的机制,也已成为当代医学研究的共同目标。近年来,特别是人类基因组研究计划和人类基因组学的序列分析图的完成,RNA 组学、蛋白质组学等研究领域的开拓,代谢组学、药物基因组学及跨学科的发展和整合,信息化、规模化和整体化的研究方法已应用于医学科学研究的各个领域。对恶性肿瘤、心血管疾病和神经系统等重大疾病发病机制进行了分子水平的研究,取得了一些成果。可以相信,随着分子生物学理论和技术的高速发展,必将给临床医学的诊断和疾病的治疗做出更大的贡献。

德 伟 欧 芹

2008 年 3 月 10 日

目 录

前言

第一篇 分子生物学基础

第 1 章 基因与基因组的结构与功能 1

第一节 DNA 的结构与功能 1

第二节 RNA 的结构与功能 6

第三节 基因组的结构与功能 8

第 2 章 蛋白质的结构与功能 14

第一节 蛋白质的结构与功能 14

第二节 蛋白质的合成与加工 19

第三节 蛋白质组学 25

第 3 章 基因表达调控 28

第一节 原核生物基因表达调控 28

第二节 真核生物基因表达调控 34

第 4 章 细胞信号转导与分子机制 45

第一节 细胞信号转导分子及其作用 45

第二节 受体介导的信号转导途径 53

第三节 细胞信号转导的相互联系 58

第四节 细胞代谢异常影响信号转导的机制 与疾病 59

第 5 章 细胞增殖、分化与凋亡的分子机制 ... 63

第一节 细胞增殖的分子机制 63

第二节 细胞分化的分子调控机制 66

第三节 细胞凋亡的分子调控机制 69

第 6 章 细胞免疫分子生物学 79

第一节 细胞因子及受体 79

第二节 免疫球蛋白家族 82

第三节 免疫细胞表面分子 83

第四节 人类白细胞抗原 86

第二篇 临床疾病案例与分子生物学

第 7 章 遗传性疾病的分子机制 92

第一节 遗传性疾病的临床特征 92

第二节 单基因遗传病 94

第三节 多基因遗传病 103

第 8 章 肿瘤性疾病细胞的分子机制 108

第一节 肿瘤的临床特征 108

第二节 肿瘤细胞增生的分子机制 110

第三节 血管生成的分子机制 118

第 9 章 感染性疾病的分子机制 123

第一节 感染性疾病的临床特征 123

第二节 病原菌致病的分子机制 124

第三节 病毒致病的分子机制 128

第四节 病原微生物基因组 132

第 10 章 炎症的分子机制 135

第一节 炎症性疾病的临床特征 135

第二节 细胞炎性反应的因子 137

第三节 炎症反应的分子机制 138

第四节 炎症反应相关的信号传导机制 ... 143

第 11 章 心血管疾病的分子机制 147

第一节 心血管疾病的分类和特征 147

第二节 动脉粥样硬化的分子机制 149

第三节 心肌肥厚的分子机制 155

第四节 心律失常的分子机制 160

第 12 章 内分泌及代谢病的分子机制 165

第一节 内分泌代谢病的临床特征 165

第二节 内分泌代谢病的分子机制 169

第 13 章 免疫性疾病的分子机制 179

第一节 免疫性疾病的临床特征 179

第二节 免疫性缺陷的分类 180

第三节 自身免疫性疾病的分子机制 186

第 14 章 衰老的分子机制 193

第一节 衰老的临床特征 193

第二节 衰老的分子机制 195

第三篇 医学分子生物学常用技术及应用

第 15 章 基因的检测及克隆 203

第一节 基因检测的方法 203

第二节 基因克隆的基本程序 205

第三节 基因克隆和 DNA 分析在医学上

的应用	208	第二节 基因治疗的临床应用	224
第 16 章 基因诊断	210	第三节 基因治疗的存在问题	228
第一节 基因诊断的基本原理和方法	210	第 18 章 基因工程药物与疫苗	231
第二节 基因诊断应用	211	第一节 基因工程药物的种类和临床	
第 17 章 基因治疗	217	应用	231
第一节 基因治疗的基本原理和过程	217	第二节 基因工程疫苗的种类和临床应用	233
中英对照名词索引	239		

第一篇

分子生物学基础

第1章 基因与基因组的结构与功能

基因(gene)的概念是19世纪提出的,当时对其化学本质及功能并没有真正了解。直到1944年,Avery通过实验才证实基因是由DNA组成的。随着分子生物学的迅猛发展,人们对基因概念的认识也逐步深化,基因是负责编码RNA或一条多肽链的DNA片段,包括编码序列、编码序列外的侧翼序列及插入序列。作为分子生物学研究领域的主要内容之一,基因将生物化学、遗传学、细胞生物学等多学科融合到一起,成为揭示生命奥秘的重要环节。

基因组(genome)泛指一个细胞或病毒的全部遗传信息。基因组学(genomics)旨在阐明基因组结构及其与功能的关系,基因与基因之间相互作用以及破译相关的遗传信息。20世纪末,各种生物全基因组序列测定的完成,特别是人类基因组计划(human genome project, HGP)的顺利实施推进了这一学科的迅速发展。

第一节 DNA 的结构与功能

脱氧核糖核酸(deoxyribonucleic acid, DNA)是一切生物的遗传物质,担负着生命信息的储存和传递功能,并且在生长、遗传、变异等一系列重大生命现象中起决定性的作用。

一、DNA 的基本结构

DNA是由数量众多的脱氧核苷酸通过3',5'-磷酸二酯键连接而成的生物大分子,无分支结构。四种脱氧核苷酸可以任意排列,因此形成了各种特异性的DNA片段。

DNA是高分子化合物,具有复杂的空间构象。DNA的一级结构是指四种脱氧核苷酸的排列顺序,两个末端分别称为5'末端和3'末端,DNA链是有方向性的,从5'端到3'端。DNA分子中的显著特点就是 C_2' 位上无自由羟基,这是DNA作为主要

遗传物质极其稳定的根本原因。DNA的二级结构是指核酸的立体空间结构,即Watson与Crick建立的DNA双螺旋结构模型,其特点如图1-1所示。DNA在双螺旋结构基础上通过扭曲、折叠所形成的特定三维构象称为三级结构。两端开放的DNA双螺旋分子在溶液中以处于能量最低的状态存在,此为松弛态DNA(relaxed DNA)。但如果DNA分子的两端是固定的,或者是环状分子,当双螺旋缠绕过分或缠绕不足时,双螺旋旋转产生的额外张力就会使DNA分子发生扭曲,以抵消张力,这种扭曲称为超螺旋(supercoils)。超螺旋如果与右手双螺旋的顺时针方向相同,过度缠绕导致绕双螺旋的张力更大,使结构更加紧张,称为正超螺旋;相反则称负超螺旋。

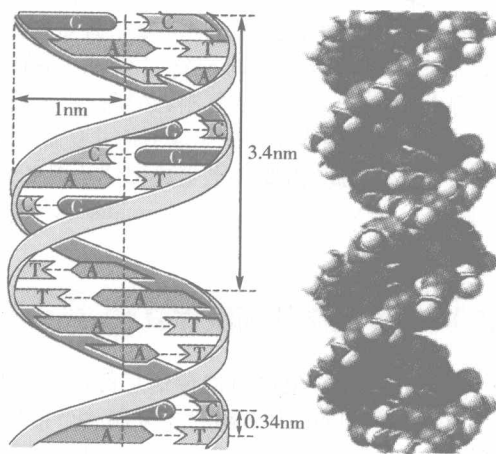


图1-1 DNA右手双螺旋

真核细胞DNA以非常致密的形式存在于细胞核内,在细胞生活周期的大部分时间里以染色质的形式出现,在细胞分裂期形成染色体。染色质的基本单位是核小体,它是直径为 $10\text{nm} \times 6\text{nm}$ 的组蛋白核心和盘绕其上的DNA所构成,核小体串联成的微丝进一步旋转折叠、压缩形成纤维状结构和棒状结构,最后形成棒状的染色体。

二、DNA 的复制

复制(replication)是指遗传物质的传代,它以母链 DNA 为模板(template)、dNTP 为原料,按碱基配对原则合成子链 DNA 的过程。其化学本质是酶促生物细胞内单核苷酸的聚合。各种酶和蛋白质因子的参与是聚合能够迅速、准确完成的保证。

(一) DNA 复制的基本规律

1. 半保留复制 子代细胞的 DNA,一股单链从亲代完整地接受过来,另一股单链则完全重新合成。按半保留复制的方式,两个子细胞的 DNA 都和亲代 DNA 碱基序列一致,保证了遗传信息稳定的传递。

2. 双向复制 原核生物的染色体和质粒,真核生物的细胞器 DNA 都是环状双链分子。它们都在一个固定的起点开始,分别向两侧进行复制,形成两个复制叉,称为单点双向复制;真核生物基因组庞大而复杂,由多个染色体组成,每个染色体又有多个起始点,称为多点双向复制。即每个起始点产生两个移动方向相反的复制叉,复制完成时,复制叉相遇并汇合连接。习惯上把两个相邻起始点之间的距离定为一个复制子(replicon),它是独立完成复制的功能单位。

3. 复制的半不连续性 当复制叉向前移动时双链 DNA 不断解链形成模板,同时合成出两条新的互补链,以边解链边复制方式进行。但所有已知 DNA 聚合酶只能以 $5' \rightarrow 3'$ 方向催化新链的合成,所以在复制时,领头链的合成方向和复制叉前移方向一致,而随从链的合成方向与复制叉前移方向相反,不能顺着解链方向连续复制,必须待模板链解开至足够长度,然后从 $5' \rightarrow 3'$ 生成引物并复制子链。延长过程中亦如此,需要再次生成引物而延长,这种复制方式称为半不连续复制(semidiscontinuous replication)。

(二) DNA 复制的高度保真性

如何保持 DNA 复制时遗传信息传递的完整、准确,并维持序列的整体连续性? 复制除严格按照碱基配对规律进行外,还依赖于酶学的机制等来保证复制的保真性。

1. DNA 聚合酶(DNA pol)对模板的识别作用

DNA pol 对模板的依赖性,是子链与母链能准确配对。只有在形成正确碱基对的情况下,引物的 $3'-OH$ 处于最佳位置上引入的核苷三磷酸的 α 磷酸才能发生催化反应,形成 $3',5'$ -磷酸二酯键。

2. DNA pol 对碱基配对的选择作用 DNA pol 能够根据模板链上的核苷酸选择正确 dNTP 掺入

到引物的 $3'$ 末端。DNA pol 对正确配对的和错误配对的 dNTP 的亲和力不同,将其聚合至引物 $3'-OH$ 的速度亦不同。在新的磷酸二酯键形成之前,dNTP 结合到聚合位点上,碱基对之间先形成氢键。DNA pol 能够识别正确与错误的配对,错误配对的 dNTP 将被排斥出聚合位点。

3. DNA pol 的即时校读功能 当碱基对产生不正确配对,其出错率水平由校正效率来决定。原核生物的 DNA-pol I 和真核生物的 DNA-pol δ 的 $3' \rightarrow 5'$ 外切酶活性都很强。当发生不正确核苷酸被添加到引物链时,错配 DNA 改变了 $3'-OH$ 和引入核苷酸的几何构象,降低了核苷酸的添加速度,而增加了 DNA 聚合酶 $3' \rightarrow 5'$ 外切核酸酶的活性,将错配的核苷酸从引物链的 $3'$ 端除去,同时利用 $5' \rightarrow 3'$ 聚合酶活性补回正确配对,DNA 合成继续进行。

4. DNA 修复系统 平均而言,DNA pol 每添加 10^5 个核苷酸就会插入 1 个不正确的核苷酸。即时校读功能将不正确配对碱基的发生概率降低到 10^{-8} 。错误概率仍比通常在细胞中观察到的实际突变率(10^{-10})高很多。细胞修复系统是 DNA 复制高度忠实性的重要因素。此外,复制起始必须利用引物,也是确保 DNA 复制忠实性的重要机制之一。

(三) 真核生物 DNA 复制的特点

(1) DNA 复制仅出现在细胞周期的 S 期,且只能复制一次。DNA 与组蛋白形成核小体,复制后还要重新形成核小体。DNA 复制时,需要克服亲代染色质中组蛋白的影响,故复制叉前进速度慢。但真核生物染色体 DNA 复制在多个复制点上(相距约 $5 \sim 300\text{kb}$)同时进行双向复制,所以从总体上可以快速进行合成。

(2) 端粒的形成:真核生物染色体 DNA 是线性结构,两端靠常规复制机制无法填补新合成 DNA 链 $5'$ 端引物去除后留下的空隙,且剩下的模板 DNA 单链 $3'$ 端如不填补成双链,就会被核内 DNase 酶水解,将产生下一轮 DNA 复制缩短的染色体(图 1-2)。

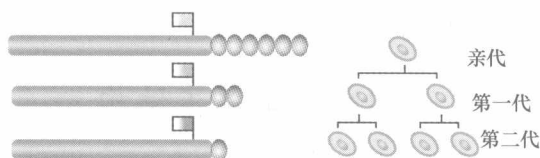


图 1-2 染色体末端复制问题

然而,染色体在正常生理状况下复制,可以保持其应有长度的。它通过染色体末端特殊结构——端粒(telomere)解决了复制问题,并可稳定染色体末端结构,防止染色体间末端的连接,补偿清除引物后造成的空缺。

端粒酶是由 RNA 和蛋白质组成的一种核糖核

蛋白(RNP),能识别和结合端粒序列。端粒酶能够延伸其 DNA 底物的 3'端。与一般 DNApol 不同是不需要内源 DNA 模板指导,而是以自身 RNA 组分为模板(这一 RNA 序列能与染色体的 3'sDNA 互补),染色体的 3'端 sDNA 作引物,通过“爬行模型”的机制维持染色体的完整(图 1-3)。

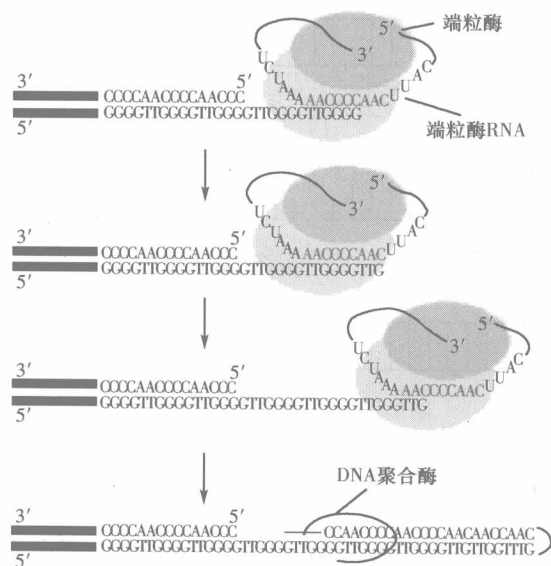


图 1-3 端粒酶催化作用的爬行模型

由于端粒酶的存在,端粒一直保持着一定的长度。在缺乏端粒酶活性时,细胞连续分裂将使端粒不断缩短,短到一定程度即引起细胞生长停止或凋亡。组织培养的细胞证明,端粒在决定细胞的寿命中起重要作用。主要的肿瘤细胞中均发现存在端粒酶活性升高,因此设想端粒酶可作为抗癌治疗的靶位点。

三、DNA 的损伤与修复

DNA 损伤(DNA damage)是指 DNA 分子上碱基的改变或表型功能的异常变化,也称为 DNA 突变(mutation)。遗传物质保持代代持续传递依赖于把突变概率维持在低水平上,活细胞需要成千上万个基因正确行使职能。生殖细胞系中高频突变将摧毁物种,体细胞中高频突变将摧毁个体。再者,某些突变会产生一些疾病包括遗传病、肿瘤及有遗传倾向的疾病。其中少数已知其遗传缺陷所在,如血友病是凝血因子基因的突变,白化病患者则是缺乏酪氨酸酶的基因所致。人类基因组完成核苷酸的测序后,疾病相关基因的检出和研究是后基因组学的重要内容。

但是,如果继承的遗传物质具有绝对的忠实性,将失去驱动进化所需的基因变异,那么新的物种,包括人类都不可能出现。因此,生命和生物多样性依赖于突变与突变修复之间的良好平衡。

案例 1-1

患者,男,12岁,因手足反复出水疱12年来院就诊。患儿出生后不久即见手足出现水疱,以关节和摩擦部位为著,皮损疼痛明显,夏季重(图 1-4),冬季稍缓解,愈后无瘢痕。体格检查:患儿营养发育好。心肺正常,肝脾未及肿大。皮肤科情况:肘膝关节外侧、掌跖关节背侧见多个厚壁清亮水疱,尼氏征阴性。组织病理:基底细胞空泡变性,表皮真皮间水疱形成。透射电镜显示表皮基底细胞层出现裂隙。家族史:家系四代中共有患者7人,每代均有患者。



图 1-4 患者足部水疱

诊断:单纯性大疱性表皮松懈症(EBS)。

问题与思考

1. 单纯性大疱性表皮松懈症发病的分子机制?
2. 该病确诊的依据?

(一) DNA 突变的类型

很多因素能导致 DNA 损伤。环境中紫外线与电离辐射等物理因素及烷化剂等化学因素是造成 DNA 损伤的主要原因,同时,机体内也会出现 DNA 的自发性损伤。根据 DNA 分子结构改变情况的不同,突变主要有以下几种类型:

1. 错配 DNA 上某一碱基的置换,使子代多聚核苷酸突变位置上核苷酸与模板 DNA 对应位置上核苷酸不配对,这种碱基错配又称为点突变(point mutation)。分为两类:①转换(transition)是同型碱基间的改变;②颠换(transversion)是异型碱基间的改变。点突变发生在基因的编码区,可导致蛋白质一级结构改变而影响其功能。若发生在简并性密码子的第三位,可能不会导致蛋白质的改变。

2. 缺失、插入和框移 插入与缺失若出现在编码区,可导致编码特异性蛋白的基因读码框发生移动,称为移码突变,造成蛋白质氨基酸排列顺序发生改变,翻译出的蛋白质可能完全不同,致使其功

能改变。3 或 3 的整数倍核苷酸的插入或缺失, 不引起移码突变。

3. 重排 DNA 分子内较大片段的交换, 称为重组或重排。移位的 DNA 可在新位点上颠倒方向反置(倒位), 也可以在染色体之间发生交换重组。

案例 1-1 相关提示

(1) EBS 是一种常染色体显性遗传病, 主要由于编码角蛋白 5 或角蛋白 14 的基因突变所致。通过对其 KRT5 和 KRT14 基因检测发现 KRT5 的第 2 外显子第 596 位, 碱基由腺嘌呤突变为胞嘧啶, 导致相应蛋白的结构及功能异常, 最终基膜带连接功能受损而引发表皮真皮交界处水疱形成。

(2) 依据患儿临床表现、组织病理及电镜观察, 致病基因的定位和克隆均符合单纯性大疱性表皮松懈症。遗传性皮肤病是由于遗传物质的改变所引起的, 通常具有上下代之间呈垂直传递或家族聚集性以及终身性的特征。

(二) DNA 损伤的修复

细胞内存在众多的机制来修复 DNA 损伤, 主要分为四种: ①直接修复; ②切除修复; ③重组修复; ④SOS 修复。以切除修复为主, 其修复包括两个过程: 一是由细胞内特异的酶找到 DNA 损伤部位并切除含损伤结构的核酸链; 二是修复合成并连接。如果只有单个碱基缺陷, 像修复细胞 DNA 中碱基(如 C)自发脱氨基产生的异常碱基(如 U), 则以碱基切除修复(base excision repair, BER)方式进行修复。如果 DNA 损伤造成 DNA 螺旋结构较大变形, 则需要以核苷酸切除修复(nucleotide excision repair, NER)方式进行修复(图 1-5)。较高等细胞中核苷酸切除修复的原理与大肠埃希菌中的基本相同, 但是对损伤的检测、切除和修复系统更为复杂。

四、DNA 与蛋白质相互作用

DNA 与蛋白质相互作用是指 DNA 结合蛋白(DNA binding protein)在它们各自靶基因的转录控制区域与序列特异性 DNA 元件相结合, 从而发挥其调控转录的功能。目前已知, 真核细胞有许多不同的结构基序, 如锌指、亮氨酸拉链、螺旋-环-螺旋和同源结构域等, 对 DNA 进行特殊的识别。序列特异性结合可能是由碱基所显示的不同结合模式识别来完成, 或者是识别主链的不同构象。许多序列特异的 DNA 结合蛋白是以二聚体的形式存在, 这样增加了 DNA-蛋白相互作用的灵敏性和特异性, 并且产生协同结合的效应, DNA 结合蛋白的二聚化还增加识别的多样性和可调控的程度。

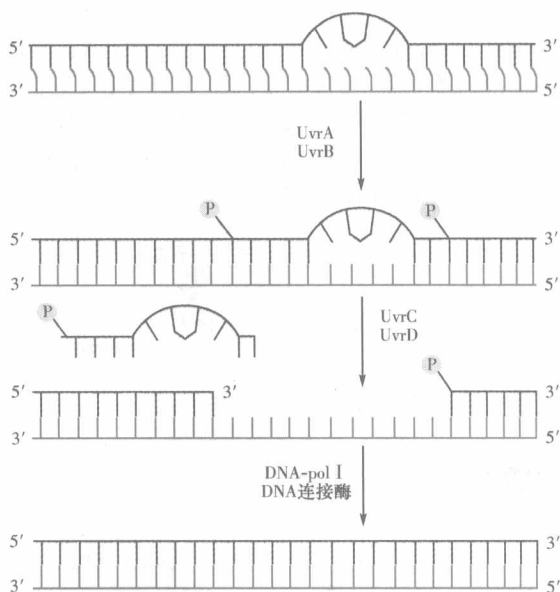


图 1-5 核苷酸切除修复方式

(一) DNA 与蛋白相互作用基本理论

DNA 结合蛋白识别整个 DNA 螺旋分子的几何结构, 通过接触大沟或小沟, 连接不变的核糖磷酸主链或碱基同 DNA 相互作用。然而, 螺旋的构象不仅可由外部环境造成(如氢键), 还可能是局部碱基堆积的相互作用造成其多样性, 这种作用为顺序特异。局部结构不仅影响了碱基同螺旋轴之间的关系, 而且还影响了螺旋的周期和大沟, 小沟。碱基序列也决定了 DNA 内在的可弯曲程度, 局部的变化可通过决定在蛋白结合时成键原子间的空间组成和 DNA 可改变的能力大小, 从而控制了蛋白和 DNA 之间的空间化学关系。

与双链 DNA(dsDNA)结合的蛋白可分三类: ①与 DNA 末端相互作用的蛋白(如 DNA 连接酶, 外切酶); ②围绕 DNA 或以深的狭缝结合 DNA 的蛋白(如 DNA 聚合酶, 拓扑异构酶); ③与 DNA 双螺旋的表面作用的蛋白。前两类包括了许多 DNA 加工的酶, 而后一类成员最多, 包括大多数转录因子、限制性内切酶、DNA 包装蛋白、位点特异的重组酶和 DNA 修复酶。因而, 后一类不仅包括普通和序列特异的 DNA 结合蛋白, 还包括识别非正常核酸结构(如损伤碱基等)的蛋白。还有一些蛋白同 ssDNA 作用, 如 RecA、单链 DNA 结合蛋白等。

(二) DNA 结合蛋白识别的一般方式

与双螺旋表面相互作用的蛋白通常具有一种十分符合大沟的基序, 通过形成稳定结合的埋藏接触面使接触区最大化。侵入大沟对于序列特异的蛋白结合是十分重要的, 有助于通过直接结合碱基进行序列识别。对于这些相互作用来讲, 结合大沟

比结合小沟更适合,因为它更大可容纳 α -螺旋、 β -折叠和环状结构。与磷酸主链的结合允许普遍性识别,可以稳定同大沟结合的结构。

1. α -螺旋在 DNA 识别过程中的作用 大多数 DNA 结合蛋白都靠大小和形状适于和大沟相识别的 α -螺旋来识别 DNA(图 1-6)。氨基酸侧链的化学多样性和柔韧性以及螺旋轴的旋转使 α -螺旋具有大量与特异性 DNA 序列相结合的潜在识别表面。晶体研究也表明,DNA 大沟中所谓的识别 α -螺旋在沟中相对于 DNA 骨架轴的位置,在不同的调控蛋白中差别很大。但因无证据能说明分离的 α -螺旋能进行独立的识别,故推断是由其他蛋白质因子介导的。与磷酸骨架的相互作用对螺旋在其位点上的合适定位是必需的。在 DNA 识别过程中,也有许多例子使用 β -片层的 DNA 结合蛋白,包括原核 Met 和 Arc 抑制因子和真核 P53 蛋白。而且,许多蛋白质除了利用大沟相互作用外,还利用肽段上可变的伸展区,伸展回折并与小沟相接触,增强识别的特异性。

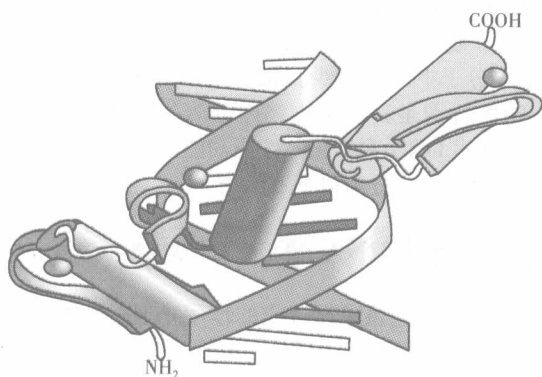


图 1-6 α -螺旋识别及结合 DNA

2. 大沟、小沟的特异性 尽管 DNA 的形状和构象在它的大小和扭角上发生微小序列依赖性变化,特异性受到碱基上化学基团向大沟和小沟序列特异性凸出的影响。

当 A/T 或 C/G 之间氢键作为将 DNA 反向平行链相连的碱基对时,这四个碱基暴露出多价的或不参与配对的化学基团,他们能与调控蛋白识别表面上的氨基酸侧链或肽骨架相结合。但只有大沟和小沟上氢键供体和受体的排列随 DNA 序列发生显著变化时,化学基团才能影响特异性。Seeman 等根据实验推断,大沟中每个碱基对的化学基团都具有独特的三维模式,多个碱基识别位点就表现出联合性的、更为精细的排列,排列的独特性随位点的增大而增强。这一变化在大沟比小沟更为明显。此外,碱基对还含有某些非结合型的化学基团。大沟似乎是序列特异性的主要靶位点。

虽然大多数 DNA 结合蛋白都偏好大沟,但有些蛋白质能将大沟和小沟识别作用相结合,这些蛋

白质通过同源结构域螺旋-转角-螺旋或 HTH 基序和大沟相接触,通过延伸的氨基端臂与小沟相接触。在原核生物中,Hin 重组酶既利用小沟,又利用大沟发生接触作用。而有些蛋白质则排他性地识别小沟。

氨基酸侧链或主链酰胺基和羰基参与同一系列碱基或磷酸骨架的序列特异性相互作用。但是,在 DNA 结合蛋白家族内部,可能更倾向于与嘌呤形成氢键且识别序列可能具有某种程度的保守性,比如同源结构域家族或锌指家族。大量研究已开始涉及改变特异性的锌指,提示至少一个转录因子家族中氨基酸-碱基对相互作用的一般规律。对 DNA 结合相互作用的进一步研究会揭示更为清楚的 DNA 序列识别模型。

(三) DNA-蛋白质相互作用的化学基础

1. 化学键

(1) 氢键:暴露于大沟侧缘的碱基供氢或受氢能力不同,与不同氨基酸残基侧链之间形成氢键的趋势也不同。例如,Glu 和 Asp 残基侧链能作为氢受体,可以与 C 和 A 形成氢键;Ser、Thr 和 Cys 残基侧链含有一OH 或—SH 基团,它们既可供氢,又可受氢,故而可与各种碱基建立氢键联系;Arg 和 Lys 只能供氢,所以不能与 C 联系。

(2) 疏水键:暴露于 DNA 大沟侧缘 T 的—CH₃是疏水的,原则上疏水性氨基酸残基 Ala、Val、Leu、Ile 等均可建立疏水键联系;C 的环内两个—CH₂—也可与疏水氨基酸残基侧链相互作用,但力量较 T 的—CH₃弱。Ala 因侧链较短,不能与 C 形成疏水键。某些氨基酸的侧链,如 Ser、Gln、His 也能与 T 形成疏水键。

(3) 离子键:一些荷电氨基酸可依据其荷电性质选择相反电荷的核苷酸碱基,通过离子键相互作用。例如 Arg 和 Lys 与 G;Asp 和 Glu 与 C 之间的离子键联系。

实际上,可能一个氨基酸残基与某种碱基相互接触时形成两个相同或不同的化学键;另一种情况是,尽管理论上某种氨基酸可与几种碱基相互作用,但实际只结合某一种碱基。这种复杂多变的情况与 DNA 和蛋白质的空间结构、位置有关,与相互作用的化学基团局部环境、状态也有关。

2. 单体和二聚体 DNA 识别单元的重复是自然界用以设计 DNA 结合蛋白最广泛采用的策略之一。采用这种重复的方法有多种:①二聚化或另一个更高级寡聚体的形成;②DNA 识别单元的多聚体化。因为亲和力与结合自由能呈指数相关($\Delta G = -RT \ln K_d$),通过使识别单元的数目加倍,加倍结合能量,使二聚体相对于单体或含有串联识别单元的单体相对于含单一单元的单体亲和力呈指数

增加。

许多蛋白质通过异源二聚化增加其调节的多样性,其中每个单体识别半位点的一个。由于异源二聚体中每个单体都有不同的序列偏好,异源二聚体可以识别含有非对称性半位点的位点。或者,异源二聚体中的两个组分具有相同的 DNA 结合特异性,但具有独特的调节特性。前一种情况的一个很好例子是真核 Jun 蛋白,它既能以同源二聚体(Jun-Jun)形式与 DNA 结合,又能以异源二聚体(Jun-Fos, Jun-CREB)形式与 DNA 结合。Jun 同源二聚体与 AP-1 启动子元件微弱地结合,而 Jun-Fos 异源二聚体则与 AP-1 启动子元件紧密结合。Jun 和其他家族成员的异源二聚体也能与一系列位点相结合。

单一多肽中 DNA 结合基序的多聚化是用于增强特异性的另一种有效方法。多聚蛋白质可以含有多个相同的识别单元,也可以含有多个具有不同结构的单元。如在 Zif268 中,三个 C—C—H—H 指识别双螺旋上相似的串联序列基序。锌指恰能通过形成围绕双螺旋的“C”形结构将 Zif268 的识别螺旋定位到大沟中。

3. 序列特异识别的结合能 蛋白质与特异 DNA 序列的选择性结合主要依赖两种类型的相互作用提供结合能。第一类相互作用是多肽链与碱基对之间的直接接触,第二类相互作用是多肽链中的碱性氨基酸残基与戊糖-磷酸骨架之间的电荷联系。序列特异的选择性结合主要在于第一类结合,即多肽链与 DNA 大沟暴露的碱基之间通过氢键和 van der Waal 力所建立的直接联系;同时,序列依赖的 DNA 弯曲或变形又可使一个特殊的结合位点被限定在最有利的构象,这种构象辅助多肽链与碱基对的定向连接。因此,多肽链中某些氨基酸与碱基对之间的定向连接属二级反应,多肽链与碱基对的识别、序列依赖的 DNA 弯曲或变形这两级效应调节着亲合位点的亲合性。

虽然多肽链与碱基对之间的直接定向相互作用可产生一定的可利用结合能,但这些能量不足以维系在平均 6~15bp 长的结合位点上所形成复合物的稳定性,额外需要的结合能则来自多肽链中带正电荷的碱性残基与带负电荷的 DNA 戊糖-磷酸骨架之间的电荷作用。当 DNA 结合某种蛋白质时,肽链与 DNA 骨架之间的这种电荷作用也会形成一种空间约束力,限定或维持 DNA 的特殊构型。通过这种电荷力实现的 DNA-蛋白质相互作用具有一个重要特性,这就是:只要 DNA 具备适当的骨架构型,蛋白质就可以结合 DNA 的任何区域,这就是蛋白质所具有的、不依赖序列的、与 DNA 双螺旋相互作用的能力。一种 DNA 结合蛋白的序列选择性主要是由其序列依赖与序列不依赖相互作用的结合能差决定的。

第二节 RNA 的结构与功能

细胞中还含另一类核酸,即核糖核酸(ribonucleic acid, RNA),成熟的 RNA 主要存在于细胞质中,少量位于细胞核。RNA 分子一般比 DNA 小得多,通常是单链线型分子,但可自身回折在碱基互补区形成局部短的双螺旋结构,而非互补区则膨出成环。RNA 的 C_{2'} 位羟基是游离的,它使 RNA 的化学性质不如 DNA 稳定,能产生更多的修饰组分,使 RNA 主链构象因羟基(或修饰基团)的立体效应而呈现出复杂、多样的折叠结构,这是 RNA 能执行多种生物功能的结构基础。RNA 的功能其一是信息分子,其二是功能分子。它能传递储存于 DNA 分子中的遗传信息,参与初始转录产物的转录后加工,并且在蛋白质生物合成中发挥着重要作用。

一、RNA 的分类与结构特点

组成 RNA 分子的基本单位是四种核苷酸: AMP、GMP、CMP 和 UMP,通过 3', 5'-磷酸二酯键连接而成的多聚核苷酸链。除此之外,在有些 RNA 分子中尚含有少量的稀有碱基和稀有核苷。生物体特别是真核生物体内, RNA 的种类、大小、结构都比 DNA 多样化。按照功能的不同和结构特点, RNA 主要分为以下三大类。

(一) 信使 RNA 的结构与功能

DNA 决定蛋白质合成的作用是通过这类特殊的 RNA 实现的,这很像一种信使作用,因此,这类 RNA 被命名为信使 RNA(messenger RNA, mRNA)。

mRNA 仅占细胞总 RNA 的 3%~5%,但其种类最多。mRNA 一般不稳定,代谢活跃,半衰期短,更新迅速。原核 mRNA 的半衰期只有 1 分钟至数分钟, mRNA 的降解紧随着蛋白质翻译过程发生。而真核 mRNA 可达数小时,甚至 24h。有些寿命较长,如人红细胞内的珠蛋白 mRNA 可达数周。

真核生物 mRNA 结构的最大特点往往以一个较大相对分子质量的前体形式在细胞核内合成,经加工成熟后再运送到细胞质内。特别是当初始转录物长达 20~30 个核苷酸时,在 5'-端形成 m⁷G-5' ppp5'-N-3' 帽子结构。3'-端大多数加一段由数十个至百余个腺苷酸连接而成的多聚 A 尾[poly(A)] 结构。两种特殊结构共同负责 mRNA 从核内向胞质的转位、mRNA 的稳定性维系以及翻译起始的调控。

mRNA 的功能是转录核内 DNA 遗传信息的碱基排列顺序,并携带至细胞质,指导蛋白质合成中的氨基酸排列顺序。一条完整的 mRNA 包括 5'-非翻译区、编码区和 3'-非翻译区。

(二) 转运 RNA 的结构与功能

转运 RNA (transfer RNA, tRNA) 约占总 RNA 的 15%, 是细胞内分子量最小的 RNA, 目前已完成一级结构测定的 tRNA 有 100 多种, 由 70~90 个核苷酸组成。细胞内 tRNA 的种类很多, 每一种氨基酸都有其相应的一种或几种 tRNA。所有 tRNA 均有以下类似的结构特点:

1. tRNA 分子含有较多的稀有碱基 约占所有碱基的 10%~20%, 如 DHU、 ψ 等。

2. tRNA 二级结构呈三叶草形 存在着一些能局部互补配对的区域呈茎状, 中间不能配对的部分则膨出形成环, 形状类似三叶草形 (cloverleaf pattern)。尽管不同的 tRNA 核苷酸组分和排列各异, 但其 3' 端无例外地都含有—CCA—OH 序列, 这是接受氨基酸的特定部位。

3. tRNA 的三级结构呈倒 L 形 这种结构比较稳定, 半衰期均在 24h 以上。氨基酸臂与 T ψ C 臂形成一个连续的双螺旋区, 构成字母 L 下面的一横。而 DHU 臂与反密码臂及反密码环共同构成字母 L 的一竖。此外, DHU 环中的某些碱基与 T ψ C 环及额外环中的某些碱基之间形成额外的碱基对, 是维持三级结构的重要因素。

4. tRNA 的功能 tRNA 在蛋白质生物合成过程中具有转运氨基酸和识别密码子的作用。不仅如此, 它在蛋白质生物合成的起始过程、DNA 逆转录合成及其他代谢调节中起重要作用。

(三) 核糖体 RNA 的结构与功能

核糖体 RNA (ribosomal RNA, rRNA) 是细胞内含量最多的 RNA, 约占总 RNA 的 80% 以上。rRNA 与核糖体蛋白 (ribosomal protein) 共同构成核糖体 (ribosome) 或称核蛋白体。核糖体的组成见表 1-1, 均由易于解聚的大、小两个亚基构成。

表 1-1 核糖体的组成

原核生物(以大肠埃希菌为例)		真核生物(以小鼠肝为例)	
小亚基	30S	40S	
rRNA	16S	1 542 个核苷酸	18S 1 874 个核苷酸
蛋白质	21 种	占总量的 40%	33 种 占总量的 50%
大亚基	50S	60S	
rRNA	23S	2 940 个核苷酸	28S 4 718 个核苷酸
	5S	120 个核苷酸	5.8S 160 个核苷酸
			5S 120 个核苷酸
蛋白质	36 种	占总量的 30%	49 种 占总量的 35%

各种 rRNA 的碱基顺序均已测定, 并据此推测

出了二级结构和空间结构。数种原核生物的 16SrRNA 的二级结构颇为相似, 形似 30S 小亚基。真核生物的 18SrRNA 的二级结构呈花状, 形似 40S 小亚基, 其中多个茎环结构为核糖体蛋白的结合和组装提供了结构基础。

rRNA 的主要功能是与多种蛋白构成核糖体, 为多肽链合成所需要的 mRNA、tRNA 以及多种蛋白因子提供了相互结合的位点和相互作用的空间环境, 在蛋白质生物合成中起着“装配机”的作用。

二、基因转录

转录 (transcription) 是指以 DNA 为模板, 在 DNA 依赖性的 RNA 聚合酶催化下, 以 NTP 为原料合成一条 RNA 的过程。转录是基因表达的第一步, 也是最关键的步骤。

(一) 转录的基本特性

细胞根据不同的发育时期、生存条件和生理需要, 只启动部分基因转录。能转录出 RNA 的 DNA 区段, 称为结构基因 (structural gene)。DNA 双链中只能有一股链按碱基配对规律指引转录生成 RNA, 这股单链称为模板链 (template strand), 相对的另一股单链则称为编码链 (coding strand)。在这段 DNA 双链上, 一股链用作模板指导转录, 另一股链不转录, 而且模板链并非总是在同一单链上。这种选择性的转录称为不对称转录 (asymmetric transcription)。

(二) 真核生物转录的特点

1. 转录起始 真核生物与原核生物的起始有一个显著的区别, 即在真核生物启动子上发生的转录起始涉及许多转录因子。这些转录因子通过识别 DNA 顺式作用元件作用位点 (cis-acting site) 而起作用的, 然而结合 DNA 并不是转录因子的唯一作用方式, 它还可以通过识别另一种因子而起作用, 或识别 RNA pol, 或和其他几种蛋白质一起组成转录起始复合体。真核生物细胞中的转录分别由三种 RNA pol (I、II、III) 催化, 起始时需要转录因子, 在随后的转录过程中则不再需要。对于所有真核生物 RNA pol 而言, 都是先由转录因子结合到启动子上形成一种结构, 以此作为 RNA pol 识别的靶标。

RNA pol I 和 RNA pol II 的启动子基本上都位于转录启动点的上游, 而 RNA pol III 的部分启动子则位于转录启动点的下游。每一种启动子均包含一组特征性的短保守序列, 能被相应的转录因子识别。

2. 转录延伸 当转录起始复合物形成后,

笔记栏

RNA pol 即开始依碱基配对关系,按模板链的碱基序列,从 5'→3'方向逐个加入核糖核苷酸。与原核生物不同的是真核生物的 DNA 结合于核小体,基因被激活表达时调控区的核小体不存在了,组蛋白或者与 DNA 脱离,或者当激活蛋白和转录因子与相应的顺式作用元件结合时,组蛋白被推开。但是当 RNA pol 通过结构基因进行转录时可以转移核小体,核心组蛋白被移位到 RNA pol 的后面,仍与同一 DNA 分子结合。这一过程可能是通过一种缠绕机制完成的。

3. 转录终止 真核生物转录终止的机制,目前了解尚不多,且三种 RNA pol 的转录终止方式不完全相同。RNA pol I 转录出 rRNA 前体 3'末端后,继续向下游转录超过 1 000 个碱基,此处有一个 18bp 的终止子序列,可被辅助因子识别辅助转录终止。RNA pol III 转录模板的下游存在一个终止子,可能有与原核生物不依赖 ρ 因子的终止子相似的结构和终止机制。RNA pol II 转录终止是和转录后修饰密切相关的。

第三节 基因组的结构与功能

无论简单的病毒或复杂的高等动植物细胞,都有一整套决定生物基本特征和功能的遗传信息。不同生物的基因组所储存的遗传信息量有着巨大的差别,其结构也各有特点。

一、基因组的结构特点与功能

不同生物的基因组大小及复杂性不同,生物的复杂性与基因组内的基因数量有关。进化程度越高,基因组越复杂。

(一) 病毒基因组的结构特点与功能

病毒(virus)是最简单的生物,外壳蛋白包裹着里面的遗传物质—核酸。根据基因组的核酸类型,可将其分为 DNA 病毒和 RNA 病毒。但是病毒的 DNA 复制及基因表达往往依赖于宿主细胞的系统。

(1) 每种病毒只含一种核酸,通常是 DNA 分子或 RNA 分子。核酸的结构可以是双链或单链结构,也可以是闭合环状或线状分子等不同类型。

(2) 病毒的基因组很小,所含遗传信息量较小,只能编码少数的蛋白质。但不同病毒基因组大小差异又很大。最小的仅 3kb,编码 3 种蛋白质。最大的可达 300kb 以上,具有几百个基因。

(3) 病毒常常具有重叠基因的结构,即同一 DNA 序列可以编码 2 种或 3 种蛋白质分子。一个基因可以完全在另一基因内,或部分重叠,重叠基因使用共同的核苷酸序列,但转录成的 mRNA 有

不同的阅读框架(open reading frame, ORF)。有些重叠基因使用相同的 ORF,但起始密码子或终止密码子不同。基因重叠现象的存在,表明病毒能够利用有限的核酸序列储存更多的遗传信息以提高在进化过程中的适应能力。

(4) 病毒基因组的大部分序列编码蛋白质,基因之间的间隔序列非常短,只占基因组的一小部分。

(5) 噬菌体基因组中无内含子,基因是连续的。但感染真核细胞的病毒基因组中具有内含子。除正链 RNA 病毒之外,真核细胞病毒的基因先转录成 mRNA 前体,再经过剪接,成为成熟 mRNA。

(二) 细菌基因组的结构特点与功能

原核生物常以大肠埃希菌为代表。这类生物能自我繁殖,具有复杂的细胞结构和代谢过程。所有原核生物的遗传物质都是 DNA,基因组在 0.6×10^6 (如支原体)到 8×10^6 (如固氮菌)之间,所包含的基因从几百个到数千个不等。目前,很多基因的功能尚不清楚,但可在其他生物中找到这些基因的同源序列,提示这些基因可能具有非常保守的生物学功能。

(1) 细菌基因组通常仅由一条环状双链 DNA 分子组成,在细胞中与蛋白质结合以复合体的形式存在。细菌染色体 DNA 在细胞内形成一个致密区域,即类核(nucleoid),类核与细胞质之间无核膜结构。基因组中通常只有 1 个 DNA 复制起点。在细菌中,除了染色体 DNA 外,还具有自主复制能力双链环状质粒存在。

(2) 功能上相关的几个基因成簇地串联排列于染色体上,连同其上游的调控区(包括启动序列和操纵序列)以及下游的转录终止信号共同组成一个基因表达单位,即操纵子(operon)结构,在同一启动序列控制下,操纵子可转录出多顺反子 mRNA。

(3) 基因组中的基因密度非常高,基因组序列中编码区所占的比例较大,只有一小部分是不翻译的。不翻译区域称间隔区,其中也包含一些基因表达调控 DNA 序列。

(4) 细菌的结构基因无重叠现象和内含子,其基因序列是连续的,因此在转录后 mRNA 不需要剪接加工。但编码 rRNA 的基因往往是多拷贝的。

(5) 基因组 DNA 有多个具有各种功能的识别区域,如复制起始区、复制终止区、转录启动子、转录终止区等特殊序列,并且含有反向重复序列。

(6) 细菌基因组中的可移动成分能产生转座现象。

(三) 真核生物基因组的结构特点与功能

真核生物从单细胞的酵母到高等哺乳动物,都

有一个共同特点,具有完整核膜的结构,使细胞核与细胞质分隔开。细胞之间也有很大差异,已分化为多种细胞类型。不同类型的细胞,执行着不同功能,复杂的功能也表现在基因组的复杂性。

(1) 基因组结构庞大,远远大于原核生物基因组,例如人的单倍体基因组 DNA 约为 3.3×10^9 bp,而大肠埃希菌的基因组只有 4.6×10^6 bp。

(2) 基因组由染色体 DNA 和染色体外 DNA (即线粒体 DNA, mitochondrial DNA, mtDNA) 组成,线性染色体 DNA 位于细胞核内,每个染色体 DNA 有很多复制起始位点(ori)。mtDNA 是闭环双链分子,结构紧凑、几乎没有重复序列。mtDNA 上某些基因可以重叠,而且没有内含子。

(3) 基因组中非编码序列多于编码序列,可占总 DNA 量的 95% 以上。在这些非编码序列中,一部分是基因的内含子、调控序列等。另一部分便是重复序列(repeat sequences),其功能主要与基因组的稳定性、组织形式以及基因的表达调控有关。根据在基因组中重复序列出现的频率不同,DNA 序列分为 ① 高重复序列 DNA (highly repetitive DNA) 重复次数可高达数百万次,这种序列可集中在某一区域串联排列。典型的高重复序列 DNA 有卫星 DNA(satellitic DNA) 和反向重复 DNA (inverted repeat DNA)。② 中重复序列 DNA 重复次数为 $10 \sim 10^5$,散在分布于基因组中,约占基因组 DNA 总量的 35%,常与单拷贝基因间隔排列,有一部分是编码 rRNA、tRNA、组蛋白及免疫球蛋白的结构基因。另外一些可能与基因的调控有关。③ 单一序列 DNA 指重复极少或不重复的序列。在单倍体基因组里这些序列包括编码蛋白质和酶的结构基因以及基因的间隔序列,这些序列一般只有一个或几个拷贝,占基因组 DNA 总量的 40%~80%。

(4) 基因组中存在多基因家族、假基因和断裂的基因。多基因家族(multigene family)是指核苷酸序列或编码产物的结构具有一定程度同源性的基因,其编码产物常常具有相似的功能,另外还有一种基因家族,是由多基因家族及单基因组成的更大的基因家族,它们的结构有程度不等的同源性,但是它们的功能不一定相同,称为基因超家族(gene super family)。假基因(pseudogene)是指与某些有功能的基因结构相似,但不能表达基因产物的基因。这些基因起初可能是有功能的,在基因复制时编码序列或调控元件发生突变,或是插入了 mRNA 逆转录的 cDNA,缺少基因表达所需要的启动子序列,变成了无功能的基因。假基因在高等哺乳动物基因组中是一种普遍的现象,许多的多基因家族中部分成员为假基因。断裂基因是指结构基因不连续,内部存在许多不编码蛋白质的间隔序列,内含子与外显子相间排列,转录时一同被转录下来。真核生物的蛋白质基因往往以单拷贝存在,转录产物为单顺反子 mRNA。

二、基因组学

基因组学是以分子生物学技术,电子计算机技术和信息技术为研究手段,以生物体内基因组的全部基因为研究对象,从整体水平上探索全基因组在生命活动中的作用及其内在规律和内外环境影响机制的科学。基因组学的研究内容主要包括: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)和比较基因组学(comparative genomics)。

(一) 结构基因组学

结构基因组学代表基因组分析的早期阶段,以建立生物体高分辨率遗传图谱、物理图谱和大规模测序为基础。

1. 遗传学图谱又称连锁图谱(linkage map)

遗传学图谱的建立为基因识别和完成基因定位创造了条件。遗传学图谱即以已知性状的基因座位和多种分子标记的座位,经过计算连锁的遗传标记之间的重组频率,来确定它们之间相对距离,将编码该特征性状的基因定位于染色体的特定位置。遗传学图谱上的连锁距离用厘摩(cM)来表示,cM 值越高,表明两点之间距离越远,重组率 1% 即为 1cM(1cM 大约相当于 100 万个碱基的长度)。人类基因组遗传学图谱的绘制需要应用多态性标记。人的 DNA 序列上平均每几百个碱基会出现一些变异(variation),这些变异通常不产生病理性后果,并按照 Mendel 遗传规律由亲代传给子代,从而在不同个体间表现出不同,因而被称为多态性(polymorphism)。多态性标记主要有三种:

(1) 限制性片段长度多态性(restriction fragment length Polymorphism, RFLP):是第 1 代标记,当用限制性内切酶特异性切割 DNA 链,由于 DNA 的点突变而产生不同长度的等位片段,可用凝胶电泳显示多态性,用于基因突变分析、基因定位和遗传病基因的早期检测等方面。

(2) DNA 重复序列的多态性标记:有小卫星 DNA 多态性和微卫星的 DNA 多态性等多种。① 小卫星 DNA 重复序列(minisatellite)是指基因组 DNA 中有数十到数百个核苷酸片段的重复,重复的次数在人群中有高度变异,总长不超过 20 kb 是一种遗传信息量很大的标记物,可以用 Southern 杂交或 PCR 法检测。② 微卫星 DNA 重复序列(microsatellite)或短串联重复(short tandem repeats, STR)多态性,是基因组中由 1~6 个碱基的重复产生的,以 CA 重复序列的利用度为最高。微卫星 DNA 重复序列在染色体 DNA 中散在分布,其数量可达 5 到 10 万,是目前最有用的遗传标记。第二代

DNA 遗传标记多指 STR 标记。

(3) 单核苷酸多态性标记 (single nucleotide polymorphism, SNP): 被称为“第三代 DNA 遗传标记”。这种遗传标记的特点是单个碱基的置换,与第一代的 RFLP 及第二代的 STR 以长度的差异作为遗传标记的特点不同,而且 SNP 的分布密集,每千个核苷酸中可出现一个 SNP 标记位点。这些 SNP 标记以同样的频率存在于基因组编码区或非编码区,存在于编码区的 SNP 约有 20 万个。称为编码 SNP(coding SNP, cSNP)。每个 SNP 位点通常仅含两种等位基因,其变异不如 STR 繁多,但数目比 STR 高出数十倍到近百倍,因此被认为是应用前景最好的遗传标记物。

2. 物理图谱 (physical map) 用全部染色体 DNA 或分离的 24 条染色体,可分别建立 YAC 库。染色体 DNA 太长,必须先切成一个个大小不同的片段,每个片段建立 YAC 克隆。每个 YAC 克隆再利用易于测定的序列标记位点 (sequencing tagged site, STS) 来识别。STS 是一段 300~500bp 的已知序列,它们在染色体上有一定的位置。构建的 YAC 克隆都含有某些已知的 STS,克隆之间还有部分重叠,即一个 STS 同时出现在两个以上的 YAC 克隆中,构成重叠群。通过杂交,将这些重叠的相邻的 YAC 克隆分别定位在染色体的不同区域。整个基因组被这些相邻的 YAC 克隆群所定位、排布,这称为物理图谱。

3. 基因组测序 随着遗传图谱和物理图谱绘制的完成,基因组测序工作已成为结构基因组学重要的研究内容。只有完成了物种基因组的 DNA 序列测定,才有可能在碱基水平上破译生物体的遗传信息。

(1) 鸟枪法:采用超声波处理或酶解的方法,将待测的 DNA 片段随机的切成大小不同的小片段并制成亚克隆。分别测定核苷酸序列后,通过计算机处理各片段核苷酸序列的资料,最终将重叠的序列拼接,直接得到待测基因组的完整序列。或者在获得一定的遗传和物理图谱信息的基础上,先对待测 DNA 作限制性内切酶谱分析,有目的地选择酶解片段进行克隆,然后测序,就可以极大地减少测序的工作量。

(2) cDNA 测序:人类基因组中发生转录表达的序列(即基因)仅占总序列的约 5%,对这一部分序列进行测定将直接导致基因的发现。由 mRNA 逆转录而来的互补 DNA 称为 cDNA,代表在细胞中被表达的基因。cDNA 测序的研究重点首先放在基因表达的短 cDNA 序列 (expressed sequence tag, EST) 测序,EST 携带完整基因的某些片段的信息,是寻找新基因、了解基因在基因组中定位的标签。比较不同条件下(如正常组织和肿瘤组织)的 EST 测序结果,可以获得丰富的生物学信息(如基因表达与肿瘤发生、发展的关系)。其次,利用

EST 可以对基因进行染色体定位。目前 cDNA 研究的热点已由 EST 转变为全长 cDNA 研究。

(3) 自动序列测定法:传统的测序方法无疑比较费时,而且测定的准确性和重复性会受到手工操作的影响。随着人类基因组计划的不断开展, DNA 测序技术得到了进一步发展。近年来建立起来的自动序列测定法,使得 DNA 测序工作标准化、规范化,极大地促进了 DNA 的结构研究。使用全自动 DNA 测序仪时,采用 4 种荧光染料标记引物或 ddNTP。由于 Sanger 测序反应产物带有不同的荧光标记物,在激光束激发时会发出 4 种不同颜色的荧光,这样一个样品的 A、G、C、T 四个测定反应产物可以在同一泳道内电泳,减少了因不同测序泳道对 DNA 片段迁移率的影响,同时激光共聚焦技术的应用,大大提高了测定的精确性,加快了检测速度。而且测定反应、灌胶、进样、电泳、扫描检测、数据分析全部实现了计算机程序控制的自动化。随后,毛细管电泳测序仪的开发,高质量的聚合酶和高度灵敏的荧光染料的出现,也使得序列测定的质量和精度不断提高。制造工艺的提高,使得以薄板凝胶系统为基础的测序仪实现了高通量产出。

(二) 功能基因组学

功能基因组学(function genomics)是研究基因组中所有基因功能的学科。它利用结构基因组学所提供的生物信息和材料,采用高通量和大规模的实验手段,结合计算机科学和统计学进行基因组功能注释(genome annotation),在整体水平上全面了解基因功能及基因之间相互作用的信息,认识基因与疾病的关系,掌握基因的产物及其在生命活动中的作用,全面系统地分析研究全部基因的功能。功能基因组学的研究主要包括:

1. 基因组功能注释 即应用生物信息学方法高通量地注释基因组所有基因编码产物的生物学功能,它是目前功能基因组学的主要研究目标。序列同源性分析、生物信息关联分析、生物数据挖掘是进行功能注释的主要生物信息学手段。研究的内容主要包括:①基因组 ORF 的识别、预测及确定基因组全部 ORF。其识别方法有两大类,一类为概率型方法,如应用 GENSCAN 评估未知 DNA 片段的编码可能性;另一类是通过同源性比较,从蛋白质数据库或 dbEST 数据库中搜寻编码区。②预测 ORF 产物的功能,采用相似聚类法寻找功能相关的保守结构域或保守模体。③非蛋白质编码区的功能注释,是功能基因组学的难点,也是新的热点。为对其进行总体研究,需构建非蛋白质编码的序列文库、大力发展比较序列和计算机分析等相关的新技术。

2. 基因表达谱的研究 某种基因的表达程度和时间是随生命活动而在不断变化和调整的。任

何一种细胞在特定条件下,所表达的基因种类和数量都有特定的模式,称为基因表达谱。它是功能基因组学研究的重要内容。研究的主要方法:①DNA 微阵列技术:是指将几百甚至几万个寡核苷酸或 DNA 片段密集地排列在硅片、玻璃片、聚丙烯或尼龙膜等固相载体上,把要研究的靶 DNA 标记后作为探针,与微阵列进行杂交,通过光电检测系统进行检测,根据杂交信号强弱及探针的位置和序列,用软件系统进行数据处理,即可确定 DNA 的表达情况,以及突变和多态性的存在。微阵列包括 DNA 芯片(DNA chip)、cDNA 微阵列等。②基因表达系统分析:不仅为定量分析全基因组表达模式提供一个良好的工具,而且大大加快了发现已知基因新功能和新基因的进展。它是以转录子(cDNA)上特定区域 9~11 bp 的寡核苷酸序列作为标签来特异性代表该转录子,然后通过连接酶将多个标签(20~60 个)随机串联并克隆到载体中,建立 SAGE 文库。通过对标签的序列分析,可获得基因转录的分布以及表达丰度情况,尤其是可检测到低丰度表达的基因,从而充分了解基因转录组的全貌。

3. 研究基因组的表达调控 一个细胞的转录表达水平能够精确而特异地反应其类型、发育阶段以及状态,因此要在整体水平识别所有基因组表达产物 mRNA 和蛋白质,以及两者之间的相互作用,绘制基因组表达在细胞发育的不同阶段和不同环境状态下基因调控网络图。

4. 研究基因组的多样性 HGP 得到的基因组序列虽然具有代表性,但是人类是一个具有多态性的群体,这决定了人生物性状的差异及对疾病的易感性,在全基因组测序的基础上进行疾病相关基因的再测序来直接识别序列变异,可以进行多基因疾病及肿瘤相关基因的研究。基因组 DNA 序列中最常见的变异形式是 SNP,在基因组中可达 300 万个。因此开展基因组多样性研究,无论对了解人类的起源、进化和迁徙,还是对生物学均会有重大的影响。

(三) 比较基因组学

比较基因组学(comparative genomics)是基于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。为充分了解人类基因组,需促进比较基因组学发展,分析各种各样的模式生物基因组。尽管模式生物基因组一般比较小,结构相对简单,但它们的核心细胞组成和生化通路在很大程度上是保守的。研究某个种属能为另一种属提供很有价值的信息。依据某种生物已知基因的知识,就能了解和分离另一生物的相关基因。比较两种远系的基因组,能领会生物学机制的普遍性和辨认实验模型所研究的复杂过程。比较两种密切相关的

基因组,更能提供基因结构与功能的细节。因此在基因组水平对不同生物体进行对照比较,将进一步加深对人类基因组结构和功能的了解,同时也可以揭示生命的起源和进化。

利用模式生物基因组与人类基因组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。另外,对植物基因组的研究也可为人类发育、衰老、病变等过程提供新的资料。

三、RNA 组学

随着基因组以及蛋白组计划的实施,国内外 RNA 研究也迅猛发展,2000 年,RNA 的研究进展被美国《科学》杂志评为重大科技突破;2001 年“RNA 干扰”作为当年最重要的科学研究成果之一,再次入选“十大科技突破”;2002 年 12 月 20 日,Science 杂志将“Small RNA & RNAi”评为 2002 年度最耀眼的明星。同时,Nature 杂志亦将 Small RNA 评为年度重大科技成功之一。2003 年,小核糖核酸的研究第四次入选“十大科技突破”,排在第四位。RNA 研究的突破性进展,是生物医学领域近 20 年来,可与人类基因组计划相提并论的最重大成果之一。

很明显 RNA 的生物功能远超出了遗传信息传递中介的范围,所以研究这些被统称为非 mRNA 小 RNA(small non-messenger RNA, snmRNAs)的时空表达情况及其生物学意义,将在全面破解生命奥秘过程中发挥重要作用。近年来随着 snmRNAs 的研究受到广泛重视,并由此产生了 RNA 组学(RNomics)的概念。RNA 组学研究同一生物体内不同种类的细胞、同一细胞在不同时间、不同状态下 snmRNAs 表达的时间和空间特异性。根据其结构和功能大约将 snmRNAs 分为:小干扰 RNA(short interfering RNA, siRNA)、微 RNA(micro RNA, miRNA)以及其他小分子 RNA(如小分子核 RNA, small nuclear RNA, sn RNA)等。

(一) siRNA

1999 年,Hamilton 等在植物基因沉默的研究中首次发现 21~25 个核苷酸双链 RNA 的出现对转基因导致基因沉默十分重要,而在转基因正确表达的植株中则未出现。这种由双链 RNA 产物高效引发的对基因表达的阻断作用被称为 RNA 干扰(RNA interference, RNAi),介导这种现象发生的小分子 RNA 称为 siRNA,这些 siRNA 一旦与 mRNA 中的同源序列互补结合,会导致 mRNA 失去功能,即不能翻译产生蛋白质,也就是使基因“沉默”了。

siRNA 通过结合并启动同源 mRNA 的降解来

笔记栏