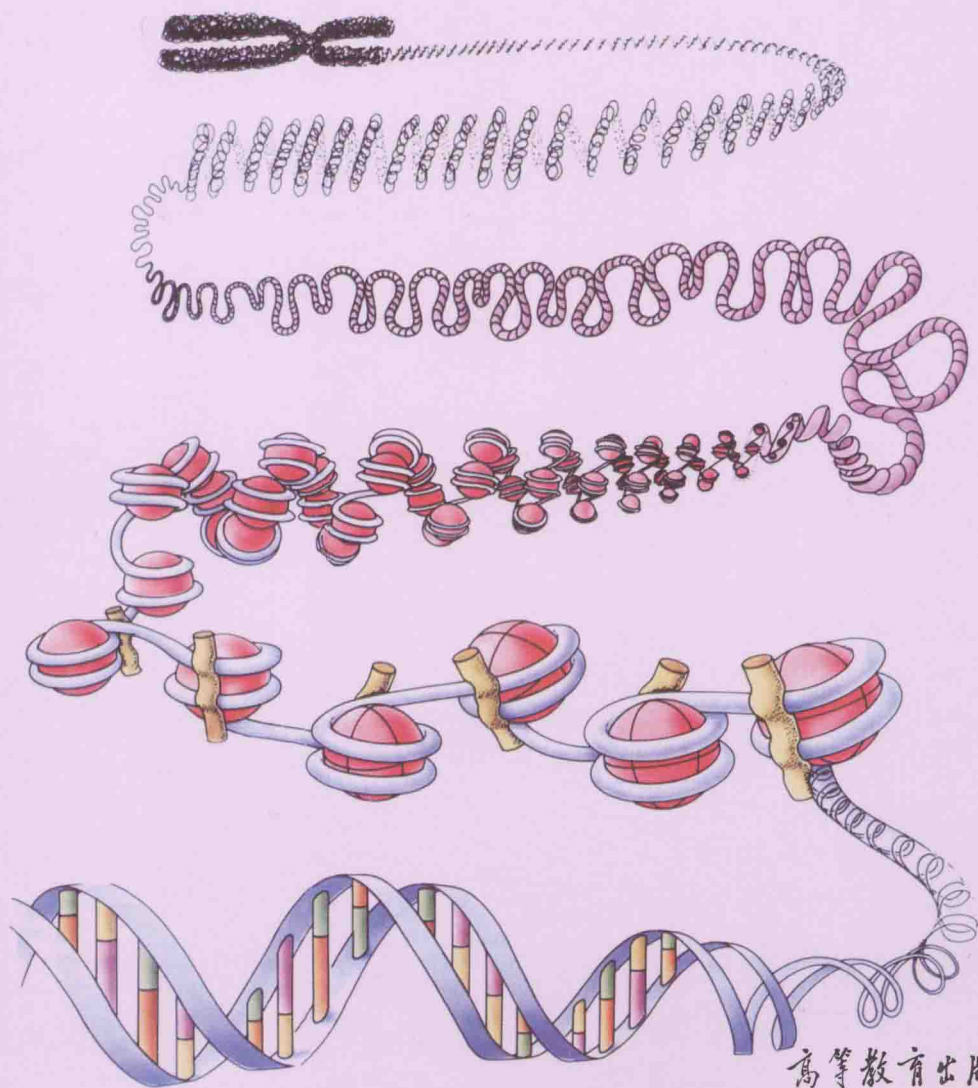


高等学校生物技术、生物工程系列教材

Gene Engineering

基因工程

朱旭芬 吴 敏 向太和 编著



高等教育出版社

J I Y I N G O N G C H E N G

基因工程

朱旭芬 吴 敏 向太和 编著

高等教育出版社·北京

内容提要

本书系统阐述了基因工程的基本理论与实用技术,以基因工程四大要素(工具酶、载体、供体、受体)为基础,以基因工程操作流程为主线,涵盖了核酸提取、靶基因克隆、基因重组、基因诱导表达、表达蛋白质的纯化、动植物转基因,以及基因工程应用等内容。此外,还介绍了支撑该学科发展的一些常用和最新的技术,如各种电泳技术、第1~3代测序技术、噬菌体展示技术、酵母双杂交技术、RNA干扰技术、基因靶向技术,以及基因组、转录组、蛋白质组的基因表达谱的研究技术等前沿热点。

与书配套的数字课程(<http://abook.hep.com.cn/40848>)提供以下拓展资源:与教材内容相关的彩图及拓展阅读资料,各章思考题的解题要点与思路,中英文名词及缩写,菌株学名,参考文献,相关网站等。

本书条理清晰,内容新颖,具有较好的系统性、先进性与教学适用性,可供高等学校生命科学相关专业的师生作为教材使用,也可供研究人员参考。

著者 向太和 编 吴 袁琳来

图书在版编目(CIP)数据

基因工程 / 朱旭芬, 吴敏, 向太和编著. — 北京: 高等教育出版社, 2014.9
ISBN 978-7-04-040848-5

I. ①基… II. ①朱… ②吴… ③向… III. ①基因工程—高等学校—教材 IV. ①Q78

中国版本图书馆CIP数据核字(2014)第207492号

策划编辑 王 莉

责任编辑 高新景

封面设计 张 楠

责任印制 韩 刚

出版发行 高等教育出版社
社 址 北京市西城区德外大街4号
邮政编码 100120
印 刷 涿州市京南印刷厂
开 本 787mm×1092mm 1/16
印 张 14.5
字 数 500千字(含数字课程)
购书热线 010-58581118

咨询电话 400-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.landaco.com>
<http://www.landaco.com.cn>
版 次 2014年9月第1版
印 次 2014年9月第1次印刷
定 价 28.00元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换
版权所有 侵权必究
物 料 号 40848-00

东北·琳琳教育出版

数字课程 (基础版)

基因工程

朱旭芬 吴敏 向太和 编著

登录方法:

1. 访问<http://abook.hep.com.cn/40848>
2. 输入数字课程用户名 (见封底明码)、密码
3. 点击“进入课程”

账号自登录之日起一年内有效, 过期作废
使用本账号如有任何问题
请发邮件至: lifescience@pub.hep.cn



基因工程

朱旭芬 吴敏 向太和 编著

用户名

密码

验证码

9337

进入课程



系列教材

内容介绍

纸质教材

版权信息

联系方式



基因工程实验指导 (第2版)

朱旭芬 编著

基因工程数字课程与纸质教材一体化设计, 紧密配合。数字课程内容包括以下各部分: 与教材内容相关的彩图及知识扩展资料、教学与实践案例PPT、各章思考题的解题要点与思路、中英文名词及缩写、菌株学名、相关文献和网站等。充分运用多种形式媒体资源, 极大地丰富了知识呈现形式, 拓展教材内容, 提升课程教学效果。

高等教育出版社

<http://abook.hep.com.cn/40848>

前言

21 世纪是生命科学的世纪,基因工程作为一门发展较快、影响较大、最具有生命力、引人瞩目的前沿学科之一,其技术已经渗透到医学、药学、农学、食品、环保、能源、材料和工程技术各个领域,成为现代生物学研究的核心技术,不断推动着生命科学的发展,驱动着人类社会生活发生巨大的变化。

本教材以基因工程四大要素(工具酶、载体、供体、受体)为基础,以基因工程操作流程为主线,按照核酸提取、靶基因克隆、DNA 酶切与连接、重组子的转化与筛选、靶基因的诱导表达、目标蛋白的纯化,以及基因表达的研究进行基因工程原理的描述,尽可能将一个完整的基因工程知识体系介绍给读者。

此外,本教材还介绍了支撑该学科发展的一些技术,如各种电泳技术(核酸凝胶电泳、毛细管凝胶电泳、变性凝胶电泳、聚丙烯酰胺凝胶电泳、脉冲场凝胶电泳、蛋白双向电泳和双向荧光差异凝胶电泳),第 1 代、第 2 代(454 技术、Solexa 技术和 SOLiD 技术)、第 3 代(HeliScope 测序、单分子实时 SMRT 测序、荧光共振能量转移 FRET 测序、纳米孔单分子读取技术, Ion Torrent 测序)测序技术,噬菌体展示,酵母双杂交,基因靶向, RNA 干扰技术以及基因组、转录组、蛋白质组的基因表达谱的研究技术等。

根据本科生专业课程学习时间减少,学时缩短的现状,本教材对一些基因工程原理与基础知识进行必要的阐述,第一、二章工具酶与载体的内容,可以穿插在“基因-酶切-连接-转化-筛选-鉴定”中介绍。结合我们长期的教学与科研经验以及对基因工程原理与实践的领悟,书中采用 200 多幅图示及 30 多张表格进行辅助说明。此外,本教材纸质内容精炼,其配套数字课程中包含:与教材相关的彩图及拓展阅读材料,各章思考题的解题要点与思路,中英文名词及缩写,菌株学名,相关网站等,以方便学生对知识的理解,满足他们进一步学习和工作的需要。

本教材得到了浙江大学生物学实验教学中心探究性教学项目以及浙江省药用植物种质改良与质量控制技术重点实验室、杭州师范大学“攀登工程”计划的出版资助。在教材编写中,我们参考了相关书籍与互联网的资料,并得到

高等教育出版社王莉老师的大力支持与帮助,高新景编辑为本书出版做了大量工作,在此一并表示最真诚的谢意!

限于作者的知识与学术水平与能力,书中一定有疏漏与错误,敬请同行专家与读者批评指正,以便不断地改进和完善。

朱旭峰

郑重声明

高等教育出版社依法对本书享有专有出版权。任何未经许可的复制、销售行为均违反《中华人民共和国著作权法》，其行为人将承担相应的民事责任和行政责任；构成犯罪的，将被依法追究刑事责任。为了维护市场秩序，保护读者的合法权益，避免读者误用盗版书造成不良后果，我社将配合行政执法部门和司法机关对违法犯罪的单位和个人进行严厉打击。社会各界人士如发现上述侵权行为，希望及时举报，本社将奖励举报有功人员。

反盗版举报电话 (010) 58581897 58582371 58581879

反盗版举报传真 (010) 82086060

反盗版举报邮箱 dd@hep.com.cn

通信地址 北京市西城区德外大街4号 高等教育出版社法务部

邮政编码 100120

短信防伪说明

本图书采用出版物短信防伪系统，用户购书后刮开封底防伪密码涂层，将16位防伪密码发送短信至106695881280，免费查询所购图书真伪。

反盗版短信举报

编辑短信“JB, 图书名称, 出版社, 购买地点”发送至10669588128

短信防伪客服电话

(010) 58582300

目 录

绪论	1
一、基因工程研究进展	1
二、遗传物质的性质与存在方式	2
三、基因及基因表达	5
第一章 工具酶	12
第一节 限制与修饰酶	12
一、限制性内切酶	12
二、甲基化酶	16
第二节 DNA 连接酶	17
第三节 聚合酶	19
一、DNA 聚合酶	19
二、RNA 聚合酶	22
第四节 核酸酶	23
第五节 核酸末端修饰酶	25
第六节 其他酶	27
第二章 分子克隆载体	29
第一节 质粒载体	29
一、概述	29
二、几种重要的质粒	31
第二节 噬菌体载体	31
一、 λ 噬菌体	31
二、黏粒	35
三、丝状噬菌体	36
四、噬菌粒	36
第三节 真核生物载体	38
一、酵母载体	38
二、植物细胞载体	40
三、动物细胞载体	40

第四节 人工染色体	41
第三章 核酸提取与基因文库	
构建	46
第一节 核酸的制备	46
一、核酸的提取	46
二、核酸的检测	48
第二节 基因文库的构建	53
一、基因文库	54
二、构建基因文库的载体	55
三、基因组文库的构建	58
四、cDNA 文库的构建	60
五、微量样品 cDNA 文库的构建	66
六、基因组文库与 cDNA 文库的适用范围与差别	67
第四章 靶基因的克隆策略	68
第一节 靶基因的克隆	68
一、靶基因的部分片段获取	68
二、PCR 合成法	69
三、核酸探针筛选	74
四、基因功能筛选	76
五、图位克隆法	81
六、人工化学合成法	83
第二节 克隆基因的序列分析	84
一、DNA 序列测定	84
二、基因序列分析	85
第五章 体外 DNA 重组	91
第一节 DNA 重组	91
一、载体的选择	91

二、靶基因的获取	92	二、动植物细胞表达体系	132
三、DNA 的体外重组	92	第七章 表达蛋白的纯化	134
第二节 重组 DNA 导入宿主	97	第一节 蛋白质纯化	134
一、宿主的选择	97	一、细胞破碎	135
二、重组 DNA 导入宿主	101	二、破胞后处理	137
第三节 阳性重组体的筛选与 鉴定	104	三、蛋白质浓缩	139
一、平板筛选法	104	四、蛋白质分离	140
二、电泳筛选法	106	第二节 蛋白质分析	146
三、PCR 筛选法	107	一、SDS - PAGE 分析	146
四、测序确认	107	二、Western 杂交	148
五、报告基因检测法	107	三、蛋白质含量测定	149
六、核酸探针筛选法	107	四、蛋白质浓缩与贮存	150
七、基因定点突变	108	第八章 转基因	152
第六章 靶基因的诱导表达	111	第一节 植物转基因	152
第一节 外源基因表达涉及的 因素	111	一、植物转基因的载体	152
一、正确的阅读框	111	二、植物转基因的受体	157
二、靶基因的有效转录	111	三、DNA 重组	157
三、靶基因转录的终止	114	四、外源 DNA 导入植物细胞	158
四、mRNA 的有效翻译	115	五、植物转基因筛选与植株 分化	161
五、密码子的利用与偏爱	116	第二节 动物转基因	161
六、翻译后的修饰加工及表达 蛋白的分泌	116	一、表达载体	162
七、蛋白质大小与融合标签	117	二、哺乳动物细胞	169
第二节 大肠杆菌表达体系	117	三、外源基因导入动物细胞	170
一、表达载体	118	四、外源基因的诱导表达	174
二、表达用的宿主菌	121	第九章 基因组、转录组和 蛋白质组的研究	175
第三节 外源基因在大肠杆菌中的 表达	123	第一节 基因组	175
一、表达重组载体的构建	123	一、测序文库的构建	176
二、外源基因的诱导表达	124	二、基因组测序	176
三、大肠杆菌中外源基因的 表达类型	125	三、测序后数据组装	184
四、外源基因表达的检测	129	四、内洞修补、拼接	185
第四节 外源基因在真核生物中的 表达	129	五、全基因组注释	185
一、酵母表达体系	129	第二节 转录组	187
		一、微阵列技术	189
		二、SAGE 和 MPSS 技术	189

三、转录组测序技术	191	一、医药卫生	206
第三节 蛋白质组	193	二、农牧业、食品工业	208
一、双向电泳	194	三、环境保护	208
二、生物质谱技术	197	第二节 转基因的安全性分析	209
三、SILAC 技术	198	一、剖析与转基因产品安全性	
四、iTRAQ 技术	199	有关的十大事件	209
五、蛋白质芯片技术	200	二、转基因的安全性问题	212
六、酵母双杂交技术	200	三、转基因的安全性管理	214
七、噬菌体展示技术	201	参考文献	216
八、基因靶向	201	索引	219
第十章 基因工程应用与思考	206		
第一节 基因工程的应用	206		

绪论

基因工程(gene engineering)是以分子遗传学为理论基础,以分子生物学和微生物学的现代方法为手段,将不同来源的基因按预先设计的蓝图,在体外构建重组 DNA 分子,然后导入宿主细胞,以改变生物原有的遗传特性,获得新品种、生产新产品。

一、基因工程研究进展

1953 年 Watson 和 Crick 在总结 Wilkins 等人研究工作的基础上,提出了 DNA 双螺旋结构的立体模型,设想核酸所携带的遗传信息以密码的形式储存于核苷酸排列顺序中,通过 DNA 半保留复制传给下一代,并通过指导蛋白质合成而表现出各种遗传性状。为此 Watson、Crick 和 Wilkins 获得 1962 年诺贝尔生理学或医学奖。

1956 年 Kornberg 分离并纯化了 DNA 聚合酶;1957 年 Ochoa 和 Kornberg 人工合成了 DNA 和 RNA,于 1959 年获得诺贝尔生理学或医学奖。1958 年 Crick 在综合了 20 世纪 50 年代有关遗传信息流向资料的基础上,首次提出了遗传信息的中心法则(central dogma),即储存于 DNA 分子中遗传信息通过 RNA 分子传递给蛋白质分子(DNA→RNA→蛋白质)。1966 年 Nirenberg、Ochoa 和 Khorana 等人破译了遗传密码,建立了遗传三联密码图。Nirenberg 由于在破译 DNA 遗传密码方面的贡献,获得了 1968 年诺贝尔生理学或医学奖。

1967 年世界上 5 个实验室几乎同时报道发现了 DNA 连接酶,特别是 1970 年 Khorana 等发现 T4 DNA 连接酶可高效连接不同的 DNA 片段。1962 年 Arber 等提出限制与修饰假说,于 1968 年发现 I 型限制性内切酶。1970 年 Smith 和 Wilcox 等从流感嗜血杆菌(*Haemophilus influenzae* Rd)分离出第 1 个 II 型核酸内切酶(*Hind* II)^①,1971 年 Nathaus 首次使用 II 型核酸内切酶完成了对基因的切割;3 人同时获得 1978 年诺贝尔生理学或医学奖。

1970 年美国 Temin 和 Baltimore 在 RNA 肿瘤病毒中发现反转录酶(reverse transcriptase),如劳斯氏肉瘤病毒(Rous sarcoma virus,RSV)能以 RNA 分子为模板,在反转录酶的作用下,合成 DNA 互补链。1971 年 Crick 修改了中心法则(图 1)。Temin 和 Baltimore 获得 1975 年诺贝尔生理学或医学奖。



图 1 中心法则

1972 年美国斯坦福大学 Berg 等将猴病毒 40 的基因组 SV40 DNA 和 λ 噬菌体 DNA 分别用内切酶切割,再用 T4 DNA 连接酶将两个片段进行连接,完成了世界上第一个 DNA 体外重组的实验,获得了 SV40 与 λ DNA 重组的杂种 DNA 分子^②。

1973 年斯坦福大学 Cohen 等将含有卡那霉素抗性基因的 *E. coli* R6-5 质粒与含有四环素抗性基因的 pSC101 质粒进行连接,在体外构建出含有四环素和链霉素两个抗性基因的重组质粒,并导入 *E. coli* 中,获得了双抗性的转化子。1974 年 Cohen 等又将非洲爪蟾核糖体基因与 pSC101 质粒重组,转化 *E. coli*,并在菌体内成功转录出相应的 mRNA,第一次成功进行基因克隆实验,标志着基因工程的诞生。

1978 年和 1981 年,美国科学家 Altman 和 Cech 由于分别发现某些 RNA 具有生物催化(酶)的功能,将其称为核酶(ribozyme),共享 1989 年的诺贝尔化学奖。1975 年 Sanger, Maxam 和 Gilbert 发明 DNA 快速测序技术;1979 年 Smith 发明了寡聚核苷酸基因定点突变技术,获得 1993 年诺贝尔化学奖。1980 年噬菌体 Φ X174(5368 碱基对)完成全部测序,成为第一个测定的基因组。

1980 年 Palmiter 和 Brinster 成功获得第一个转基因小鼠。1983 年通过农杆菌介导法培育出了转基因烟草。1985 年美国科学家 Mullis 发明了 PCR 仪,与第一个设计定点突变的 Smith 共享了 1993 年诺贝尔化学奖。此后,在常规 PCR 基础上又发展了多种特殊的 PCR。1990 年“人类基因组计划”(HGP)正式启动,英、法、德、日等国家先后加入了 HGP 计划,1999 年 9 月中国承担了 1% 的测序工作。1997 年英国爱丁堡大学罗斯林(Roslin)研究所胚胎学家 Wilmut 领导的科研小组利用一只成年绵羊的乳腺细胞克隆出一只羊羔多利(Dolly)。2000 年 6 月人类基因组草图绘制完毕,即“工作框架图”。此外, RNA 干扰技术的建立,在基因治疗、功能基因组学等研究等方面具有重要的应用价值,获得 2006 年诺贝尔生理学或医学奖。基因打靶技术等获得 2007 年诺贝尔生理学或医学奖。

二、遗传物质的性质与存在方式

1. DNA 性质

(1) 碱基配对(base pairing) 碱基配对是 DNA 分子结构的重要特性,GC 碱基对有 3 对氢键,AT 碱基对仅有 2 对氢键^③。

(2) DNA 变性 将双链 DNA 或天然 DNA 加热时,两条链之间的结合力受到破坏,两条链分离。另外,高 pH 如在 pH 大于 11.3 时,所有的氢键断裂,DNA 完全变性。这种由于外部条件导致双链 DNA 形成单链的过程称为 DNA 变性(denaturation)。把 DNA 达到 50% 变性的温度称为融解温度(melting temperature, T_m)^④,与 DNA 的 G + C 含量成正比。基因工程中 DNA 变性常用加热或碱水解。DNA 变性后双链的亲水力减少,碱基的电荷性能改变,造成对紫外线的吸收增加,即增色反应(hyperchromicity)^⑤。

(3) 复性与杂交 变性的 DNA 在一定条件下,两条互补单链重新配对成原有双链 DNA,即为复性(renaturation)^⑥。如加热后的 DNA 让其慢慢冷却,则会发生复性。如果加热后迅速冷却置于冰浴,DNA 则呈单链的变性状态。当复性 DNA 由不同的两条同源单链互配形成杂合双链分子,称为杂交(hybridization)^⑦。

2. DNA 二级结构

DNA 二级结构分 3 种构型:B 型 DNA、A 型 DNA 和 Z 型 DNA^⑧。

B 型 DNA 是经典的 Watson-Crick 结构,二级结构相对稳定。DNA 分子由两条互补但方向相反的多核苷酸长链组成双螺旋,通过氢键的引力把两条多核苷酸长链连在一起。B 型 DNA 沿中心轴旋转一周为 10 个核苷酸,且有大沟和小沟之分。

A 型 DNA 是 B 型 DNA 的重要变构形式,其分子形状与 RNA 的双链区和 DNA/RNA

杂交分子很相近。在低温条件下,DNA 可被诱导形成 A 型。A 型也为右旋,但结构更加紧密。当 B 型 DNA 所处环境的相对湿度小于 70% 时可转变为 A 型。A 型 DNA 的螺距仅为 2.8 nm,每个螺距含 11 个碱基对。

Z 型 DNA 是左手螺旋,也是 B 型 DNA 的变构形式。Z 型 DNA 中两条由 dC 和 dG 交替形成的多聚核苷酸反向平行排列,通过碱基配对联结并缠绕成 Z 型 DNA。在左旋双螺旋 DNA 中,脱氧胞苷为反式构象,而脱氧鸟苷取顺式构象。这种顺、反式使得主链上的糖环的取向也交替变化,从而使 Z 型 DNA 主链呈锯齿状(zigzag)。

表 1 3 种构型 DNA 的主要特征

生物体内一般以右旋 B 型为主,少量以左旋 Z 型存在。研究表明 Z 型 DNA 构象与基因的转录活性有关。3 种构型 DNA 中,大沟的特征在遗传信息表达中起关键作用,此外,沟的宽窄及深浅也直接影响碱基对的暴露程度,从而影响调控蛋白对 DNA 信息的识别。B 型 DNA 是活性最高的 DNA 构象;变构后的 A 型 DNA 仍有活性;变构后的 Z 型 DNA 活性明显降低。此外,DNA 双螺旋能进一步扭曲盘绕形成特定的高级结构,DNA 二级结构变化在 DNA 复制与转录中具有重要的生物学意义。

知识扩展 1 DNA 的压缩折叠

3. 遗传密码的性质

遗传密码是指 DNA 链上各个核苷酸的特定排序。每个密码子(codon)由 3 个核苷酸决定,是负载遗传信息的基本单位,具有以下特性:

(1) 通用性(universal property) 自然界所有的生命形式公用一本密码图,除极少数例外,三联密码子与氨基酸之间的对应关系见图 2。由于这一通用性,使不同生物基因之间

		第二位核苷酸					
		U	C	A	G		
第一位核苷酸(5'端)	U	UUU} 苯丙氨酸 (Phe) UUC UUA} 亮氨酸 (Leu) UUG	UCU} 丝氨酸 (Ser) UCC UCA UCG	UAU} 酪氨酸 (Tyr) UAC UAA} 终止密码 UAG	UGU} 半胱氨酸 (Cys) UGC UGA} 终止密码 UGG 色氨酸 (Trp)	U	第三位核苷酸(3'端)
	C	CUU} 亮氨酸 (Leu) CUC CUA CUG	CCU} 脯氨酸 (Pro) CCC CCA CCG	CAU} 组氨酸 (His) CAC CAA} 谷氨酰胺 (Gln) CAG	CGU} 精氨酸 (Arg) CGC CGA CGG	C	
	A	AUU} 异亮氨酸 (Ile) AUC AUA} 甲硫氨酸 (Met) 或起始密码 AUG	ACU} 苏氨酸 (Thr) ACC ACA ACG	AAU} 天冬酰胺 (Asn) AAC AAA} 赖氨酸 (Lys) AAG	AGU} 丝氨酸 (Ser) AGC AGA} 精氨酸 (Arg) AGG	A	
	G	GUU} 缬氨酸 (Val) GUC GUA GUG	GCU} 丙氨酸 (Ala) GCC GCA GCG	GAU} 天冬氨酸 (Asp) GAC GAA} 谷氨酸 (Glu) GAG	GGU} 甘氨酸 (Gly) GGC GGA GGG	G	

图 2 三联密码图

的转移和表达成为可能。

(2) 简并性(degeneracy) 一个氨基酸由一个以上的密码子编码。密码图上除甲硫氨酸(Met)和色氨酸(Trp)只有一个密码子外,编码其他氨基酸的密码子都有2个或2个以上,最多有6个,如亮氨酸(Leu)、丝氨酸(Ser)和精氨酸(Arg)(表1)。

表1 遗传密码的简并性

同义密码数	编码的氨基酸	合计
6	Leu, Ser, Arg	18
4	Gly, Pro, Ala, Val, Thr	20
3	Ile	3
2	Phe, Tyr, Cys, His, Gln, Glu, Asp, Asn, Lys	18
1	Met, Trp	2
编码氨基酸密码		61
终止密码		3

蛋白质合成中,氨基酸通过相应的 tRNA 携带,即 tRNA 分子上的反密码子与编码氨基酸密码子的碱基配对,这种配对要求密码子的前两个核苷酸形成精确的配对,而在第三位上可以不同,即简并性。通过改变基因序列中的核苷酸,可以不使其编码的氨基酸发生变化,从而产生或删除必要的限制性内切酶的酶切位点,也可以进行蛋白质已知氨基酸序列设计,合成具有各种可能性编码的寡核苷酸探针,用于基因的分离和鉴定。

(3) 偏倚性(bias) 密码子的专一性主要由前两个碱基决定。在蛋白质合成中,对简并密码子的使用频率不同。如 UUU 和 UUC 都编码苯丙氨酸(Phe),但在高表达的蛋白质中 UUC 的使用频率明显高于 UUU。一般密码子的第一、二位碱基是 AU,那么第三位碱基将尽量使用 GC,反之亦然。在基因工程中常选择性使用“高频密码”,增加基因中常用密码子的数量,提高外源基因在宿主细胞中的表达水平;另外通过密码子中间的碱基性质,可以判断所编码的氨基酸是亲水性还是疏水性氨基酸,即第二位碱基 U 是编码疏水性氨基酸的密码子;第二位碱基 A 是编码亲水性氨基酸的密码子;而第二位碱基 G 或 C 是编码疏水性亲水性居中的氨基酸。

(4) 不重叠性(non-overlapping)和阅读方向性(reading directivity) 对于特定的多肽链,在蛋白质合成过程中密码子是不重叠的,即从起始密码开始,严格地按照3个碱基决定一个氨基酸的方式依次顺读。密码阅读的方向与 mRNA 编码的方向一致,从5'→3'进行(图3)。

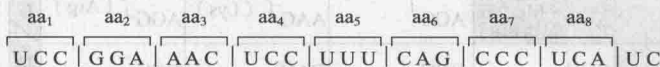


图3 密码子阅读的方向性

(5) 摆动性(wobble) 密码子与反密码子配对,有时会出现不遵从碱基配对规律的情况,称为遗传密码的“摆动”现象(表2)。

表 2 密码子与反密码子配对的“摆动”现象

反密码子的 5'端碱基	密码子的 3'端碱基
C	G
A	U
U	A 或 G
G	C 或 U
I	U、C 或 A

(6) 特殊性(particularity) 在酵母、无脊椎动物、脊椎动物的线粒体(mitochondrion), 以及支原体中的遗传密码与遗传密码图有一些偏离(表 3), 如终止密码子(UGA)变为编码色氨酸(Trp); 在哺乳类细胞线粒体中的 AGA 与 AGG(精氨酸 Arg)变成终止密码子; 哺乳类、果蝇和酵母线粒体中的 AUA(异亮氨酸 Ile)变为甲硫氨酸(Met); 酵母线粒体中的 CUA(亮氨酸 Leu)和果蝇线粒体中的 AGA(精氨酸 Arg)分别改变为苏氨酸(Ser)和丝氨酸(Thr)。遗传密码的某些偏离可能是生命演化过程中的产物。

表 3 遗传密码的特殊性

密码子	常用	交替使用	交替使用地方
AGA、AGG	Arg	终止密码、Ser	一些动物线粒体, 原生动物
AUA	Ile	Met	线粒体
CGG	Arg	Trp	植物线粒体
CUU、CUC、CUA、CUG	Leu	Thr	酵母线粒体
AUU	Ile	起始密码子	一些原核生物
GUG	Val	起始密码子	一些原核生物
UUG	Leu	起始密码子	一些原核生物
UAA、UAG	终止密码	Glu	一些原核生物
UGA	终止密码	Trp、硒代半胱氨酸	线粒体、支原体、 <i>E. coli</i>

起始密码子 AUG, 原核 mRNA 将其译作甲酰甲硫氨酸(f-Met), 真核生物则译作甲硫氨酸(Met)。除 AUG 外, 以 GUG 作为起始密码子占 8%, 而以 UUG 作为起始密码子只有 1%。终止密码子有 UAA、UAG 和 UGA, 其中 UAA 终止密码子效率最高, UAG 最低, 且紧随着三种密码子之后的是 U。

近来研究发现, 除线粒体使用一组密码子其含义有别于核基因外, 原先被认为仅用作终止信号的无义密码子在某些情况下编码特定的氨基酸, 如 UGA 编码硒代半胱氨酸(selenocysteine, Se-Cys)(图 4)。在某些古菌和细菌中, UAG 编码第 22 种天然氨基酸吡咯赖氨酸(pyrrrolysine, Pyl)(图 5)。

三、基因及基因表达

基因组(genome)是细胞或生物体的全套遗传物质。就细菌和噬菌体而言, 其基因组是指拟核遗传物质所含的全部基因, 而二倍体真核生物的基因组则是指维持配子正常功能的一套染色体及其所携带的全部基因。如人的基因组包括一套 23 + 1 条染色体、含有

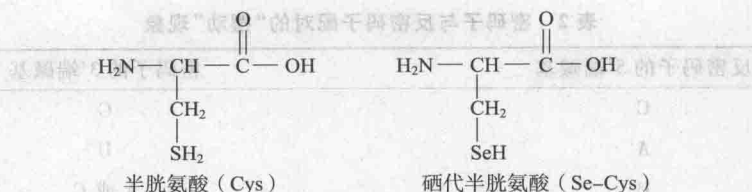


图4 半胱氨酸与硒代半胱氨酸的分子结构

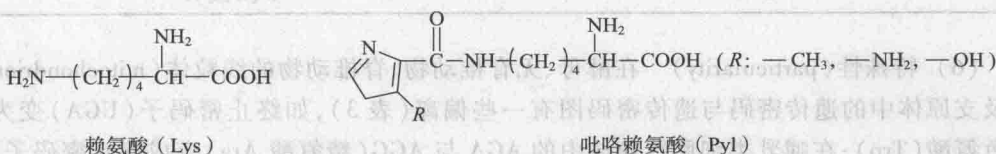


图5 赖氨酸与吡咯赖氨酸的分子结构

3×10^9 个碱基对的 DNA。20 世纪 60 年代发现了插入序列,开始认识到基因组中某些成分的位置并非一成不变,同种生物的不同个体间基因组大小或基因组数目也不是绝对固定的,甚至由于基因组结构变化还会导致功能的变化。

1. 基因

基因(gene)是生物体内一切具有自主复制能力的遗传单位,是负载特定生物遗传信息的 DNA 或 RNA 分子片段,是细胞中所有 RNA 及蛋白质分子的“蓝图”。

(1) 移动基因(movable gene) 1951 年美国冷泉港实验室(cold spring harbor laboratory)科学家 McClintock 在玉米研究中发现,玉米种子的颗粒颜色可通过杂交使其改变,这种改变可通过非杂交方法恢复。她认为在玉米基因组上存在可移动的控制因子,在基因组上穿梭巡逻,当它任意插到玉米染色体的靶位点上,就可控制所在位置基因的表达,使基因开启或关闭。且此控制因子又可从插入部位重新被删除,使基因的表达恢复原来的状态。McClintock 在 20 世纪 50 年代创立了移动基因理论,但由于传统的观念及当时对分子生物学知识的贫乏,此观点没能被接受,直到 1961 年 Jacob 和 Monod 的乳糖操纵子模型和控制基因理论发表,才被人们重视。20 世纪 60 年代末 Shapiro 研究 *E. coli* 高效突变时,在细菌中发现了插入序列,从而使“移动基因”被大家所接受,直到 1983 年 McClintock 才获得诺贝尔生理学或医学奖。

知识扩展 2 移动基因

(2) 断裂基因(split gene) 断裂基因是指编码序列在 DNA 分子中不连续,或是被插入序列隔开。在真核细胞中,编码蛋白质的 DNA 序列只占整个基因组序列的小部分,如人类基因组中蛋白质编码的 DNA 序列只占 3%,而 90% 以上的序列并不编码蛋白质。其中一些是内含子(intron)^⑨,已发现这些 DNA 序列对包括癌症在内的人类正常基因组的修复、调控及多细胞有机体的演化至关重要。高等真核生物多数基因都有内含子。酿酒酵母(*Saccharomyces cerevisiae*)只有少数基因有内含子,裂殖酵母(*Schizosaccharomyces*)较多,*E. coli* T4 噬菌体、枯草杆菌 SPO1 噬菌体和蓝细菌少数基因含有内含子。线粒体、叶绿体基因也含内含子。不同生物的同—基因,内含子长度尽管变化很大,但数目、位置往往相同。

(3) 基因丰余(gene redundance) 真核 rRNA 基因、tRNA 基因和组蛋白基因的份数超过实际需要,称为基因丰余。①细菌中 rRNA 基因是多拷贝的,*E. coli* 有 7 份,枯草杆菌

有 10 份,各自都组成一个操纵子。基因的排列顺序为 16S—23S—5S, rRNA 基因之间以及 5S rRNA 基因下游插有 tRNA 基因。②tRNA 基因:由于遗传密码具简并性,且密码子的 3' 碱基和反密码子的 5' 碱基配对可以摇摆,所以细菌翻译 61 种密码子只需 32 种反密码子。但实际上仍然有约 60 种 tRNA 存在。③组蛋白基因由 H_1 、 H_2A 、 H_2B 、 H_3 、 H_4 共 5 个组成,首尾相连、重复排列成串。各基因之间和各重复单位之间都有间隔区分开,重复次数因物种而异。如人、鼠每单倍体基因为 10~40 份;鸡为 10 份,酵母 2 份,果蝇约 100 份,小麦 50~60 份。

(4) 重叠基因 (overlapping gene) 一般同一段 DNA 序列内不可能存在着重叠的读码基因,但在噬菌体和动物病毒中,不同基因的核苷酸序列有时是可以共用,即核苷酸序列采取彼此重叠的方式 (基因多种读框) 来增加单位 DNA 序列所携带的信息量。如 $\phi X174$ DNA 只含有 5387 个核苷酸,却编码了 11 种蛋白质,一般每个基因有 900~1500 bp。自发现 $\phi X174$ 噬菌体基因重叠后,已发现一些遗传信息重叠的情况,其中包括重叠操纵子、异相位重叠基因、同相位重叠基因、反相位重叠基因等。

异相位重叠基因中最常见的是上游基因的终止密码子和下游基因的起始密码子重叠,如基因 B 和基因 K 两基因交界处是...TGATG...,其中 A 使用两次 (图 6)。

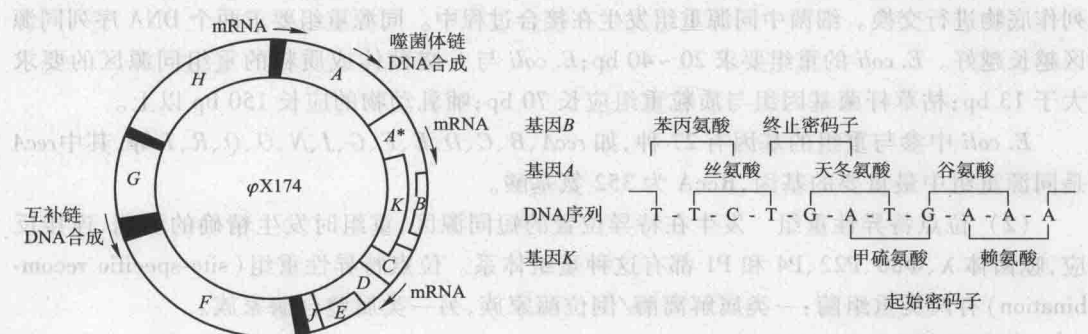


图 6 $\phi X174$ 的重叠基因

同相位重叠基因是重叠基因中最简单的一种,其编码的蛋白质只是长短不同,其氨基酸顺序完全一样。如 $Tn5$ 转座子的 P2 蛋白质比 P1 蛋白质 N 端少 40 个氨基酸, P1 是转座酶, P2 是转座酶的竞争性抑制物。

反相位重叠基因两条 DNA 都有编码功能。IS5 由右向左编码一个较大的蛋白质,由左向右编码一个较小的蛋白质。两者读框相同,方向各异,氨基酸序列不同。

这种重复利用会使 ORF 的信息容量和利用率变高,主要见于 $\phi X174$ 、SV40 和 RNA 噬菌体等小型基因组中,基因重叠可以尽量发挥其编码能力。

知识扩展 3 基因簇

2. 基因的分类

基因根据是否具有转录和翻译功能分为 4 类:

(1) 结构基因为编码蛋白质的基因 具有转录和翻译功能,即可转录形成 mRNA,并进而翻译成多肽链,构成各种结构蛋白质、催化各种生化反应的酶和激素等。

(2) 只有转录功能没有翻译功能的基因 如 rRNA 基因、tRNA 基因以及一些小分子的细胞核 RNA (snRNA) 基因。