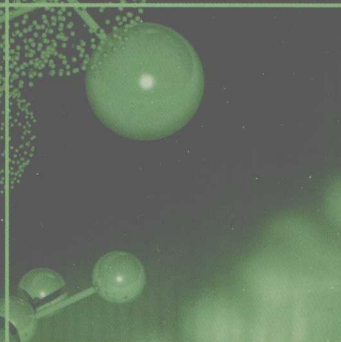
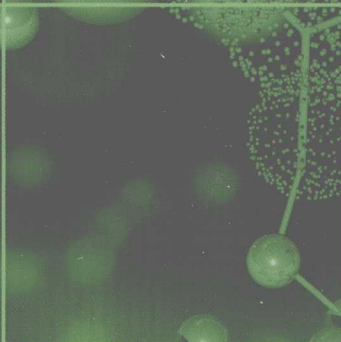
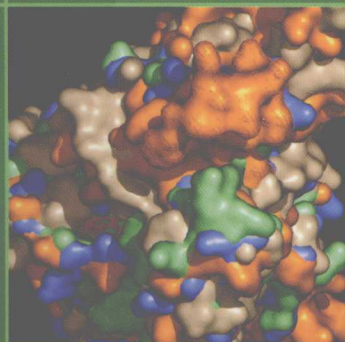
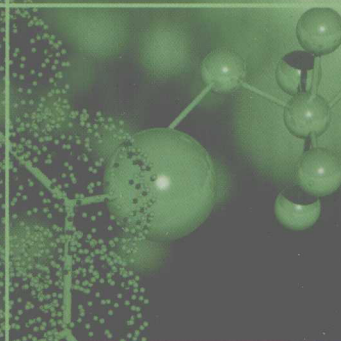
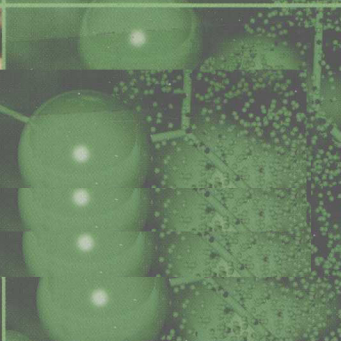
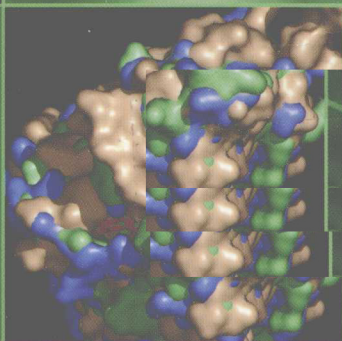
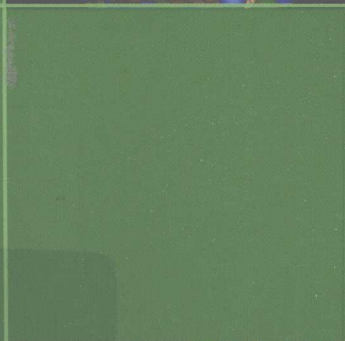
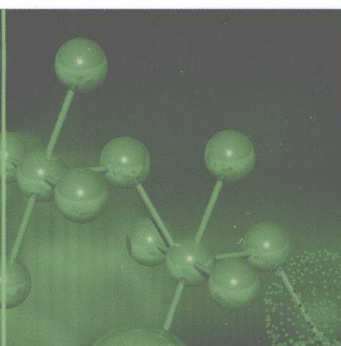
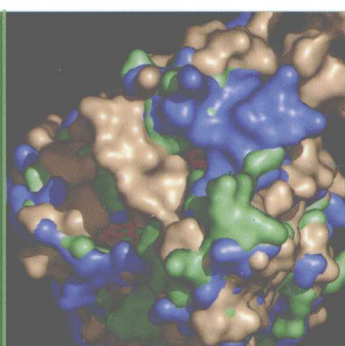


Study Guide and Problem Solving for Molecular Biology

分子生物学 学习指南与习题解析

主编 杨荣武

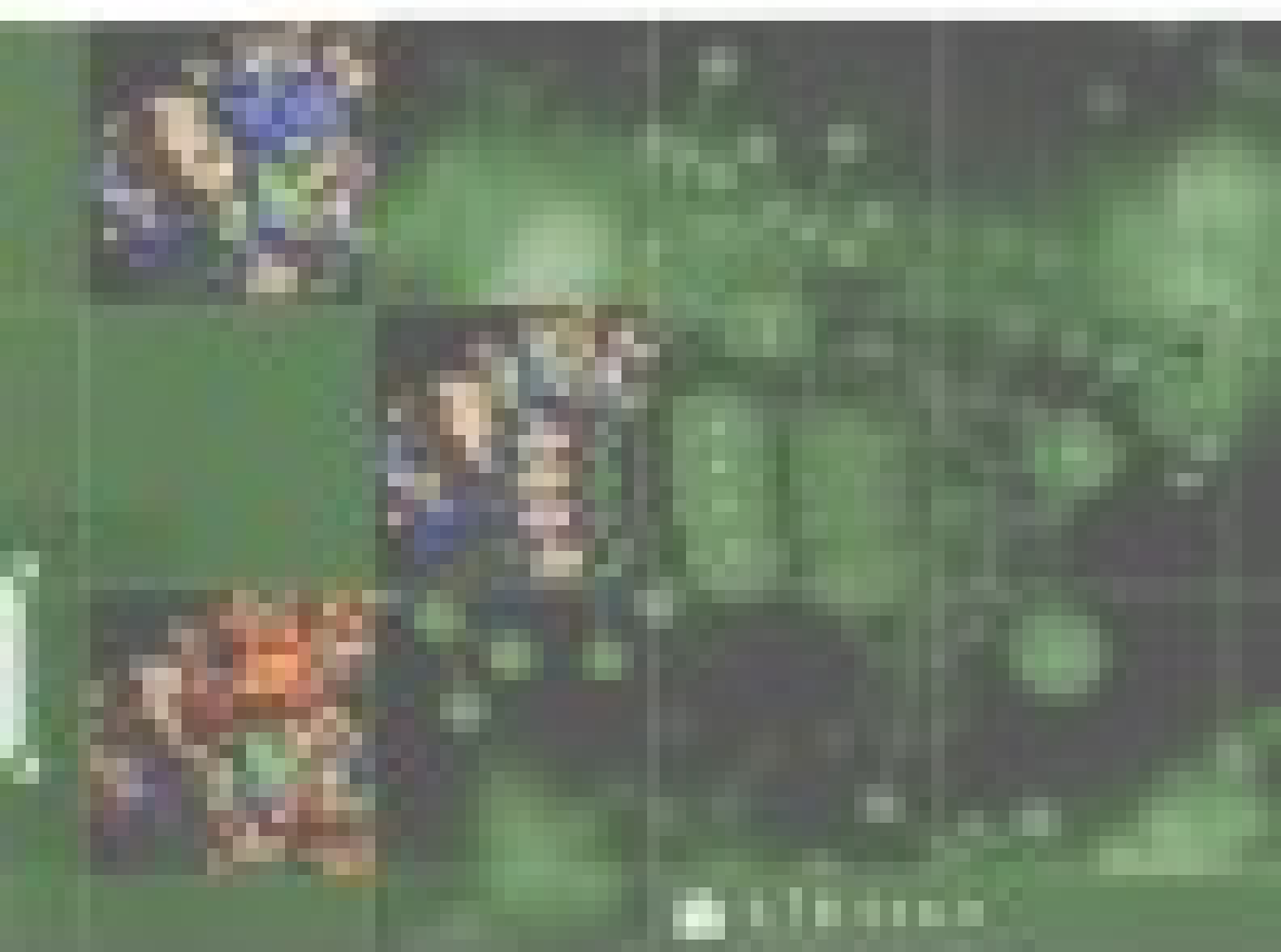


高等教育出版社
HIGHER EDUCATION PRESS

第 1 版 2013 年 12 月第 1 次印刷 第 2 版 2015 年 12 月第 2 次印刷

分子生物学 学习指南与习题解析

第 2 版 第 2 次印刷



人民邮电出版社

分子生物学 学习指南与习题解析

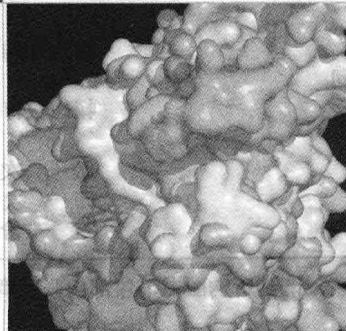
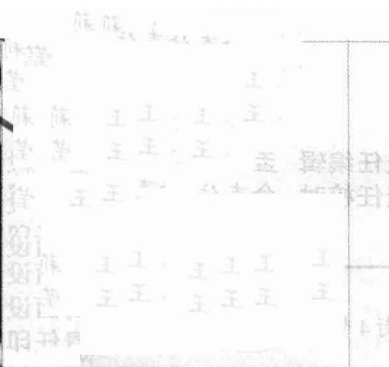
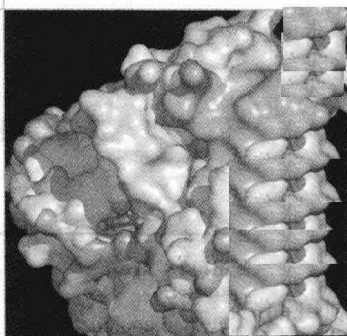
Fenzi Shengwuxue Xuexi Zhinan yu Xiti Jixi

主 编 杨荣武

副 主 编 郑典元 臧宇辉

参编人员 (按姓氏拼音排序)

刘新建 戚金亮 邵世光 杨荣武
郑典元 郑伟娟 朱 煜



高等教育出版社·北京
HIGHER EDUCATION PRESS BEIJING

内容简介

本书以本科生物科学类专业“分子生物学”课程的教学大纲为指导,按照杨荣武主编的《分子生物学》和朱玉贤等编著的《现代分子生物学》(第3版)体系编排,为这两部教材的学习指导书。全书共分为六章,每章的内容包括:本章要求、本章内容提要、本章疑难点剖析及学习方法、典型例题解析、习题和习题答案及解析。疑难点剖析及学习方法部分精心设计有大量表格,以供比较、记忆;习题包括名词解释、填空题、是非题、选择题(单选和多选)和问答题等题型,问答题包括简答题、实验分析题和实验设计题。书后还附有近几年南京大学攻读硕士及博士学位研究生入学分子生物学考试试题,为读者自测之用。

本书内容丰富,不仅提供了许多设计新颖、富有启发性的习题及其解题技巧,还针对性地剖析了同学们在学习经常遇到的难点和疑点,同时介绍了一些好的学习和记忆方法。适合作为高等院校生物科学类各专业的师生学习分子生物学的参考用书,还特别适合准备报考研究生的学生自我检测和复习之用。

图书在版编目(CIP)数据

分子生物学学习指南与习题解析/杨荣武主编. —北

京:高等教育出版社,2010.6

ISBN 978-7-04-029135-3

I. ① 分… II. ① 杨… III. ① 分子生物学-高等学校-教学参考资料 IV. ① Q7

中国版本图书馆 CIP 数据核字(2010)第 081809 号

策划编辑 王 莉
版式设计 王 莹

责任编辑 孟 丽
责任校对 俞声佳

封面设计 张 楠
责任印制 陈伟光

责任绘图 尹 莉

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

经 销 蓝色畅想图书发行有限公司
印 刷 北京市鑫霸印务有限公司

开 本 850×1168 1/16
印 张 20.75
字 数 640 000

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.landracom.com>
<http://www.landracom.cn>
畅想教育 <http://www.widedu.com>

版 次 2010 年 6 月第 1 版
印 次 2010 年 6 月第 1 次印刷
定 价 36.00 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 29135-00

前言

分子生物学是在分子水平上研究生物体的结构与功能的一门学科。现在它不仅是一门学科,还成为生命科学研究中一种十分重要的工具。生命科学因为有了它才可能在较短的时间内取得了那么多的突破。正因为如此,学好它对于每一个从事生命科学和相关科学研究的人都特别重要。然而,从我身边接触到的人反映来看,许多人认为难,甚至对它有点心存畏惧。但我个人认为,这门学科的内容应该是整个生命科学中逻辑性最强的部分,因此,如果方法得当,还是比较容易学好、学透的。要学好它,最重要的两点应该是整合和比较。因为这门学科的全部内容围绕分子生物学的“中心法则”展开的,死记的内容并不多,学习的目的就是要搞清楚一个“中心”(DNA)和两个“基本点”(RNA 和蛋白质)之间内在的分子逻辑关系。在学习中,需要将它们作为一个整体来看,一环套一环,同时注意科学家们在揭示一些重要发现的时候所使用的各种技术和手段。例如,在 DNA 复制的一般特征这一节,基本上每一要点的背后都有着设计巧妙的实验,从 Meselson 和 Stahl 设计的被誉为“生命科学中最美丽的实验”即证明 Watson 和 Crick 提出的半保留复制的实验,到 Okazaki 设计的证明 DNA 复制是半不连续性的实验,每一个实验设计得都很精彩。我们在学习的时候,除了要搞清楚这些实验的结果与得出的结论之间的逻辑关系之外,更需要理解这些实验设计的思路,以便为将来自己做科研的时候提供宝贵的借鉴。而老师在授课的时候要多多介绍相关的背景故事,而同学更需要理解基于这些实验设计衍生出来的分子生物学真谛。比较和归纳是在学习这些内容的时候最重要的方法之一。此外,学习这门课程还需要在如何选择好教材、如何做好笔记、如何记忆、如何充分使用网络资源等方面掌握好一些窍门。总之,归纳起来有以下几点:

1. 注意将原核系统和真核系统进行比较

无论是原核生物还是真核生物,都在进行 DNA 复制、转录、转录后加工、翻译等基本的分子事件,两类生物在这些事件上既有相同之处,也有许多差异。在学习的时候,时刻要注意将两大系统进行比较。例如:在学习 DNA 复制的时候,注意将原核细胞内的 DNA 聚合酶 I、II、III、IV、V 和真核生物的 DNA 聚合酶 α 、 β 、 γ 、 δ 、 ϵ 进行比较,将原核 DNA 聚合酶 III 的 β 滑动钳和真核 DNA 聚合酶 δ 的 PCNA 滑动钳进行比较;在学习转录的时候,需要将两者的启动子结构和 RNA 聚合酶的结构与功能进行比较;在学习转录校对的时候,注意将原核细胞中的 GreA、GreB 和真核细胞内的 TFIIIS 进行比较;在学习 DNA 甲基化的时候,要注意原核生物与真核生物在甲基化的位点和功能上是不同的;在学习弱化子机制的时候,要注意这种机制是原核系统特有的,真核系统没有。如果能这样去学习的话,那所有的内容就活了,将它们串在一起理解要比孤立地记忆强得多!

2. 注意将两种不同的分子机制进行比较

细胞内的很多分子机制是很相似的,这就需要我们在学习的时候,将相关联的分子机制放在一起去领会、理解。如 DNA 复制和 DNA 转录,两者有很多共同的特点,例如都需要解链,合成的方向都是从 $5' \rightarrow 3'$,都遵循 Watson 和 Crick 碱基配对原则。当然,在意识到这些共同特点的时候,也不能忽视它们的差别,比如, DNA 复制需要引物, RNA 不需要, DNA 聚合酶通常具有自我校对能力, RNA 聚合酶没有校对能力。这里更要明白为什么会有这些差异,为什么允许有这些差异。

3. 以“中心法则”为核心,“碱基互补配对”和“蛋白质与核酸之间的相互作用”为主线,巧妙地利用“外因与内因的关系”的理论,全面理解分子生物学的机制

分子生物学的核心内容是所谓的“中心法则”,即生物体内的三种生物大分子——DNA、RNA 和蛋白质之间的关系。其中涉及遗传信息的复制、损伤修复、重组、转录、逆转录、转录后加工和翻译等。这些过程总是涉及蛋白质和核酸分子之间的相互作用和碱基互补配对,因此,掌握蛋白质和核酸分子之间相互作用的规律以及碱基互补配对的原则对于深入理解分子生物学的各种机制和原理至关重要。另外,细胞内的很多机制都可以使用哲学的“外因”和“内因”之间的关系原理进行理解,掌握这一点非常重要。例如,理

解 DNA 复制为什么具有固定的起点? 这涉及 DNA 复制起始区和复制起始蛋白之间的相互作用, 在这里可以将 DNA 复制起始区看成“内因”, 复制起始蛋白(大肠杆菌为 DnaA 蛋白)看成“外因”。按照“内因”和“外因”之间的关系原则, 即“内因”是变化的根据, “外因”是变化的条件, “外因”需要通过“内因”起作用, DNA 复制区所具有的特殊序列是 DNA 复制具有固定起点的根本原因, 即“内因”, 但仅有它是不够的, 还需要识别这种特殊序列的蛋白质, 它就是“外因”, 正是它们之间的相互作用才使得 DNA 复制从固定的起点开始。

4. 注意掌握各种研究方法的原理及其应用

分子生物学的发展与研究方法的进步是分不开的, 而反过来它的发展又促使人们提出和发明新的研究手段。两者之间相互依存, 相互促进。因此, 在学习各章节内容的时候, 对于分子生物学家在研究各种分子机理时所使用的方法要充分理解。例如, 对参与 DNA 复制的各种蛋白质和酶的鉴定主要是利用 DNA 复制突变体的互补和体外复制系统的重建两种方法。互补的原理是利用某种野生型的蛋白质去恢复特定的 DNA 复制缺陷突变体的复制功能, 从而确定参与复制的蛋白质。重建的原理是在较为简单的体外复制系统(如 SV40 病毒复制系统)中, 先人为去掉某种成分, 致使复制不能正常进行, 然后, 将复制系统中逐一添加分步收集的可能参与复制的蛋白质抽提物, 看是否能够恢复复制活性, 从而确定复制蛋白质。有时, 添加的蛋白质可能来自于其他物种, 这样可以从中找到同源的或同工的蛋白质。为了方便理解重建的原理, 这里可以打一个比方加以说明。假定你的一台电脑坏了一个部件而不能运转, 那么如何迅速找到是哪一个部件有毛病呢? 这时可以用类似重建的手段来确定: 首先弄一台运转正常的电脑, 将它的各个部件拆开, 那么, 来自这台正常电脑内的所有部件都应该是正常的(相当于野生型蛋白质)。然后, 将坏掉的电脑逐一取出一个部件(如内存条或主板), 再用正常电脑的相应部件取而代之。如果某一个部件经过替换以后, 坏的电脑恢复正常了, 这就等于找到了坏的部件(相当于突变型蛋白质)。这两种方法对于参与其他过程(如信号转导、转录、转录后加工、翻译、细胞周期的调控等)的蛋白质的鉴定也很有帮助。例如, 为了找到人细胞内参与细胞周期的某一种蛋白质, 先是将酵母细胞内某一种与细胞周期有关的蛋白质突变, 这样的酵母的细胞周期肯定会有异常。然后, 将正常的人细胞内的各种可能与细胞周期有关的蛋白质导入到突变的酵母细胞中, 如果其中的某一组分加入以后, 酵母的细胞周期恢复正常, 那么这种导入的蛋白质就是人细胞内的一种与细胞周期有关的蛋白质。

5. 选择好合适的教材和参考书

目前市场上有各种各样的分子生物学教材和参考书, 如何选择适合自己的教材和参考书对于培养学习兴趣, 学好本学科十分重要。我个人认为, 最好能准备中英文教材各一本和一本学习指南与习题解析。中文教材可以选择由我主编的《分子生物学》或北京大学朱玉贤教授等编著的《现代分子生物学》(第 3 版), 中文教材在价格上比较便宜, 在理解上也更加容易。英文的原版教材可选择 Weaver 的 *Molecular Biology*、Watson 的 *Molecular Biology of the Gene* 或者 Lewin 的 *Genes* 系列。英文原版教材的特点是新、印刷精美, 图表多为彩图, 通常还有配套的多媒体光盘, 方便于自学。阅读一本好的英文教材, 不仅有助于提高自己的专业英语水平, 而且更能加深对各章节内容的理解; 至于学习指南与习题解析, 最好带有学习方法介绍和例题详细分析的, 本书就是一个很好的选择。

6. 学会做笔记

首先有一点必须强调, 上课时同学们的主要任务是听老师讲课而不是做笔记, 因此在课堂上要集中精力听讲, 一些不清楚的内容和重要的内容可以笔录下来, 以便课后复习和向老师求教。当然, 条件好的同学可以买来录音笔, 将老师的上课内容录下来, 以供课后消化。现在国外很多大学直接将课程录制下来, 提供 MP3 和 MP4 下载, 如果国内的大学也能这样, 那就更方便了。另外, 老师的讲稿大都作成了幻灯片, 同学们可从老师那里得到拷贝。如果事先将老师的课件打印出来, 然后在打印稿上做笔记, 效果会更好。

7. 多使用图表、比较和联想的方法来辅助理解和记忆所学过的内容

重点学习复制、转录和翻译的基本过程, 并从必要条件、所需酶蛋白和特点等方面对三个过程进行比较, 在理顺它们的基本框架后, 就应全面、系统、准确地掌握教材的基本内容, 并且找出共性, 抓住规律。

8. 充分利用网络资源

网上有各种免费的教学资源,有条件的同学可经常去浏览,跟踪和了解最新的进展。也可以去一些BBS或Blog站点,与网友一起交流学习的体会和对一些热点问题进行讨论。

以上就学习分子生物学的方法谈了自己的看法,但需要指出的是,每一个人学习这门课的基础、目的和条件可能有差别,所以最好是结合自己的特点总结出最适合自己的学习方法。总之,只要同学们勤于思考、方法得当、多做题目和实验,学好并考好分子生物学是完全可能的。

杨荣武

南京大学生命科学学院

2009/10/18

目 录

1 基因、基因组、基因组学和

基因的化学本质 1

- 一、本章要求 1
- 二、本章内容提要 2
- 三、本章疑难点剖析及学习方法 5
- 四、典型例题解析 10
- 五、习题 11
- 六、习题答案及解析 28

2 DNA 的复制、重组、损伤、

修复以及突变 45

- 一、本章要求 45
- 二、本章内容提要 47
- 三、本章疑难点剖析及学习方法 50
- 四、典型例题解析 53
- 五、习题 56
- 六、习题答案及解析 79

3 DNA 转录、逆转录及其

转录后加工 102

- 一、本章要求 102
- 二、本章内容提要 103
- 三、本章疑难点剖析及学习方法 106
- 四、典型例题解析 109
- 五、习题 111
- 六、习题答案及解析 133

4 翻译及其后加工 153

- 一、本章要求 153
- 二、本章内容提要 154
- 三、本章疑难点剖析及学习方法 156
- 四、典型例题解析 159

五、习题 161

六、习题答案及解析 179

5 基因表达的调控 197

- 一、本章要求 197
- 二、本章内容提要 198
- 三、本章疑难点剖析及学习方法 201
- 四、典型例题解析 205
- 五、习题 206
- 六、习题答案及解析 229

6 基因工程以及其他现代

分子生物学技术 249

- 一、本章要求 249
- 二、本章内容提要 250
- 三、本章疑难点剖析及学习方法 251
- 四、典型例题解析 256
- 五、习题 258
- 六、习题答案及解析 273

南京大学 2009 年攻读硕士学位研究

生入学考试试题 290

南京大学 2008 年攻读硕士学位研究

生入学考试试题 293

南京大学 2007 年攻读硕士学位研究

生入学考试试题 296

南京大学 2006 年攻读硕士学位研究

生入学考试试题 297

南京大学 2005 年攻读硕士学位研究

生入学考试试题 298

南京大学 2004 年攻读硕士学位研究

生入学考试试题 302

南京大学 2003 年攻读硕士学位研究

生入学考试试题	309	入学考试试题	317
南京大学 2002 年攻读硕士学位研究		南京大学 2008 年春季攻读博士学位	
生入学考试试题	312	研究生入学考试试题	318
南京大学 2001 年攻读硕士学位研究		主要参考书目	320
生入学考试试题	315	鸣谢及后记	321
南京大学 2000 年攻读硕士学位研究生			

基因、基因组、基因组学和基因的化学本质

一、本章要求

1. 掌握以下名词的含义:基因、基因组(genome)、基因组学、蛋白质组(proteome)、蛋白质组学、转录组(transcriptome)、互作组(即相互作用物组, interactome)、生物信息学(bioinformatics)、连锁图(linkage map)、单核苷酸多态性(SNP)、限制性酶切片段长度多态性(RFLP)、DNA 限制性酶切图谱(物理图谱)、种属间印迹(zoo blotting)、C 值、N 值、K 值、C 值矛盾、N 值矛盾、K 值矛盾、单拷贝 DNA、中度重复序列、高度重复序列、基因组计划、卫星 DNA。
2. 了解近一个世纪以来基因概念所发生的变化;了解一个转录单位与现代基因概念之间的关系以及可读框(ORF)与基因概念之间的关系。
3. 掌握原核生物的基因与真核生物的基因至少有三个主要的差别。
4. 了解如何将基因和蛋白质分成家族和超家族。
5. 了解必需基因和基因丰余(gene redundancy)的概念;了解基因串联重复和基因簇产生之间的关系;了解串联重复的基因移位到不同染色体的两种机制。
6. 掌握什么是假基因及其形成的机制。
7. 知道生物随着复杂性的提高具有跨膜和胞外结构域的蛋白质数目出现猛增的趋势。
8. 了解高度重复序列是如何组织的;了解基因重复(gene duplication)如何导致一种生物进化出具有新功能的蛋白质的机制。
9. 了解珠蛋白基因在发育过程中是如何进行差异表达的。
10. 知道不等交换能够增加或减少一个基因簇内的基因数目。
11. 了解哺乳动物卫星 DNA 的类型以及如何从短的核苷酸序列进化产生。
12. 了解基因组、转录组、蛋白质组、代谢组、脂质组和互作组之间的关系。
13. 了解如何使用经典的传递遗传学获得连锁图谱。
14. 了解限制性酶切图谱和序列图谱与连锁图谱之间的差别。
15. 了解 DNA 编码区以外的各种序列的多态性。
16. 掌握如何使用凝胶电泳的方法显示 DNA 的多态性。
17. 了解如何对多态性作图以及如何通过谱系来进行跟踪。
18. 掌握如何使用多态性定位致病基因。
19. 了解 C 值和 C 值复杂性之间的关系;能够应用特定的例子描述 C 值矛盾并能解释 C 值矛盾产生的原因。
20. 区分卫星 DNA、小卫星 DNA 和微卫星 DNA。
21. 知道真核生物基因组内的三类不同的重复序列。
22. 了解一种生物内的转座子如何成为生物进化和新物种产生的驱动力。
23. 了解可读框的概念以及如何使用它来评价基因组的组织。
24. 掌握种属间印迹对于鉴定一个基因组内重要基因的重要性。

25. 了解细胞质遗传与细胞器 DNA 之间的关系。
26. 了解线粒体基因组和叶绿体基因组的大致结构以及它们与细菌基因组之间的关系。
27. 知道线粒体或叶绿体内的蛋白质并不都是由它们自己的基因组编码。
28. 知道解释线粒体和叶绿体起源的内共生学说。
29. 掌握从已知的基因组序列中获得有用信息的方法。
30. 对专性寄生菌、自由生活的细菌、单细胞真核生物、多细胞真核生物和具有复杂组织/器官的动物、植物所具有的基因数目有所了解。
31. 了解几种模式生物(细菌、酵母、果蝇、线虫、拟南芥和小鼠)的基因组的相对大小。
32. 了解人类基因组的大小、基因的数目以及人类基因组计划完成的基本概况。
33. 掌握以下名词的含义: A 型、B 型、Z 型双螺旋, 大沟, 小沟, 负超螺旋, 正超螺旋, T_m , Chargaff 规则, 增色效应(hyperchromic effect), 减色效应(hypochromic effect), H-DNA, Watson-Crick 碱基对, 核酶(ribozyme), 分子杂交, DNA 的复杂度和 $Cot_{1/2}$ 。
34. 掌握五种最常见的碱基(A、G、T、U、C)的化学结构以及嘌呤环和嘧啶环的编号规则。
35. 了解碱基、核苷和核苷酸的化学结构及其理化性质, 特别是碱基的“互变异构”性质。
36. 掌握 DNA 和 RNA 在结构上和功能上的异同, 并能够解释为什么 DNA 分子上需要用 T 代替 U 以及用脱氧核糖代替核糖。
37. 了解几种常见的 RNA 以及近几十年来发现的一些特殊的 RNA 的名称和功能。
38. 了解核酸一级结构、二级结构和三级结构的定义。
39. 掌握核酸一级结构的表示方法以及其测定的主要手段和原理。
40. 掌握 DNA 双螺旋结构的主要内容、支持 DNA 双螺旋结构的主要实验证据及其生物学意义; 了解大沟对于 DNA 与蛋白质相互作用的重要性。
41. 掌握 A 型、B 型和 Z 型双螺旋结构之间的异同以及形成 A 型和 Z 型双螺旋所需要的条件。
42. 能够画出 AT、AU 和 GC 碱基对之间的氢键。
43. 了解几种 DNA 非标准二级结构(H-DNA、P-DNA 及错配滑动 DNA)及其可能具有的生物学功能。
44. 能够区分双螺旋结构中的 Watson-Crick 碱基对和三螺旋结构中的 Hoogsteen 碱基对; 能够解释为什么 DNA 的三链结构需要两条嘧啶链和一条嘌呤链。
45. 了解 RNA 的各种可能的二级结构以及几种重要的三级结构; 能够将 GU 碱基对视为 RNA 分子上正常的碱基对。
46. 掌握 DNA 的正超螺旋和负超螺旋形成的原因; 能够使用 $L = T + W$ 预测 DNA 过度缠绕或缠绕不足的后果; 知道何种超螺旋有利于 DNA 复制和转录。
47. 了解 DNA 的变性、复性及其杂交等性质; 知道能够影响 DNA 变性和复性的各种因素。
48. 了解影响 DNA 的 T_m 的各种因素。
49. 了解各级水平的染色质包装: DNA 双螺旋、10 nm 纤维、30 nm 纤维、成环结构域、间期染色质和中期染色体。
50. 掌握 H3-H4 四聚体和 H2A-H2B 二聚体是如何组装成核小体核心颗粒的; 了解核小体结构如何导致 DNA 形成超螺旋; 掌握组蛋白 N 端尾巴如何从核心颗粒伸出来并与其他蛋白质发生相互作用的。
51. 了解如何使用微球菌核酸酶和 DNA 酶 I 来研究核小体定位。
52. 了解各种常见的分离和纯化核酸以及定量测定核酸的方法。
53. 能够区分朊病毒和类病毒。

二、本章内容提要

基因是遗传物质(主要是 DNA)分子上最基本的功能单位。现在人们不仅能从表型上去研究基因, 而

且能够先克隆基因,再深入研究其结构和功能,并且发现了诸如断裂基因、重复序列、重叠基因、转座子和假基因等以各种形式存在的基因。真核生物的基因一般以单顺反子的形式存在,编码的是单基因产物;而原核生物的基因绝大多数以多顺反子的形式存在,其转录物为多顺反子 RNA,可同时编码两种甚至数种基因产物。原核生物的基因和真核生物的基因都可以分为编码区和非编码区。绝大多数真核生物其编码区被非编码区所打断,称为断裂基因;但是绝大多数原核生物的基因编码区则是连续的。断裂基因的大小与外显子关系不大,主要由内含子的大小和数目决定。内含子越多,片段越长,则其基因越大。真核生物中,从酵母到人类,基因的平均大小略有增加。根据基因密度可推知基因组中基因的数目。生物从低等到高等,基因数目逐渐增加。真核生物基因组中来源相同、结构相似、功能相关的一组基因称为基因家族。根据家族成员在染色体上的分布形式,基因家族又分为在染色体上串联重复存在的基因簇和在染色体上散在分布的散在基因家族。除了能编码蛋白质的基因家族外,染色体上还存在大量无转录活性的重复 DNA 序列。根据重复单位的大小,一些高度重复序列又分为卫星 DNA、小卫星 DNA 和微卫星 DNA。

基因组一般是指某一物种单倍体细胞内的全套染色体 DNA。基因组可分为原核生物基因组、真核生物核基因组和细胞器基因组。原核生物基因组简单,只有一个染色体组成,其类核外可能有质粒 DNA 存在。细胞器基因组主要指真核生物的线粒体和叶绿体基因组,编码细胞器所有的 tRNA、rRNA 以及部分蛋白质。真核生物的核基因组比较复杂,可分为非重复序列、中度重复序列和高度重复序列。每个基因组中各成分所占比率不同,但较大的基因组非重复序列部分较少。复杂度描述了其中独特序列的长度,重复频率描述了每个序列重复的次数。C 值矛盾描述了真核基因组中编码潜力和 DNA 含量并非一致。大多数结构基因位于非重复 DNA 内。非重复 DNA 的复杂度比总基因组能更好地反映生物的复杂程度,非重复 DNA 最大复杂度达 2×10^9 bp。人类基因组计划自 1990 年启动到 2005 年已基本结束,人类基因组的全部测序工作已经完成,人类的遗传图谱和物理图谱已成功绘制。在绘制遗传图谱时,主要用到的遗传标记有 RFLP、STR 和 SNP 等。物理图谱的绘制实际上是对 DNA 序列两点间的实际距离的测定,用到的遗传标记主要为 STS。利用全基因组的“鸟枪法”测序策略,还得到了数十几种模式生物的全基因组序列。

基因组学是指研究基因组的科学,它包括结构基因组学和功能基因组学。目前对基因组的研究进入了功能基因组学时代。功能基因组学以高通量、大规模实验方法以及统计与计算机分析为特征。基因芯片、基因表达系列分析、定位克隆等都是功能基因组学的主流研究方法。蛋白质组学是功能基因组学的一个分支,包括结构蛋白质组学和功能蛋白质组学,采用的主要研究方法有双向(二维)电泳、质谱分析、酵母双杂交和蛋白质芯片等。生物信息学诞生于生物、计算机、数学等多个领域的交叉,其研究目标是揭示和理解基因组学获得的大量数据中所包含的生物学意义。

核苷酸由核苷和无机磷酸组成。核苷由碱基和(脱氧)核糖通过 β -N 糖苷键连接而成。碱基是含有 N 原子的嘌呤或嘧啶的衍生物。核酸分子中代表遗传信息的是碱基序列。生物体内最常见的嘧啶碱基是胞嘧啶(C)、尿嘧啶(U)和胸腺嘧啶(T),嘌呤碱基是腺嘌呤(A)和鸟嘌呤(G)。U 通常只存在于 RNA,T 一般只存在于 DNA。碱基几乎不溶于水,在溶液中能发生酮式(氨基式)—烯醇式(亚氨基式)的互变异构和酸碱解离,并具有强烈的紫外吸收,其最大吸收波长为 260 nm。

核苷中的 D-核糖或 2-脱氧-D-核糖都以呋喃型环状结构存在,其中的 β -N 糖苷键由戊糖的异头 C 与嘧啶碱基的 N1 或嘌呤碱基 N9 形成。在核苷中,碱基在糖苷键上的旋转受到空间位阻的限制。嘧啶核苷通常为反式构象,自由的嘌呤核苷更容易形成顺式构象,但 DNA 和 RNA 螺旋中的嘌呤核苷主要为反式构象。由于核糖部分的高度亲水性,核苷的水溶性要比自由的碱基高得多。核苷在碱性条件下较稳定,嘧啶核苷能抵抗酸水解,嘌呤核苷很容易发生酸水解。

核苷酸是核苷的戊糖羟基的磷酸酯,包括核糖核苷酸和脱氧核苷酸。自然界的核苷酸多为核苷-5'-磷酸。核苷酸有核苷单磷酸、核苷二磷酸和核苷三磷酸,分别用 NMP(rNMP)、NDP(rNDP)和 NTP(rNTP)表示;脱氧核苷单磷酸、脱氧核苷二磷酸和脱氧核苷三磷酸分别用 dNMP、dNDP 和 dNTP 表示。核苷酸的构象不是固定不变的,具有一定的柔性。

核酸是由多个核苷酸通过 3',5'-磷酸二酯键相连的多聚物,分为 RNA 和 DNA。

核酸的一级结构是指构成核酸的多聚核苷酸链上的所有核苷酸或碱基的排列顺序。每一条线形多聚核苷酸链都具有不对称的 5'端和 3'端。书写核酸一级结构的惯例是 5'端→3'端(从左到右)。核酸一级结构的意义是储存生物体的遗传信息。

DNA 的二级结构主要是各种形式的螺旋,特别是 B 型双螺旋,此外还有 A 型双螺旋、Z 型双螺旋和三链螺旋等。其中最主要的形式为 Watson 和 Crick 于 1953 年提出的 B 型双螺旋,其核心内容是:DNA 由两条呈反平行的多聚核苷酸链组成,它们相互缠绕形成右手双螺旋;两条链通过 AT 碱基对和 GC 碱基对互补结合在一起;碱基对位于双螺旋的内部,并垂直于暴露在外的脱氧核糖磷酸骨架。相邻碱基对之间的疏水作用和范德华力对双螺旋的稳定起一定的作用,双螺旋的表面含有大沟和小沟;相邻碱基对距离为 0.34 nm,相邻核苷酸之间的夹角约 36°。螺旋直径为 2 nm,每一转完整的螺旋含有 10 bp,其高度为 3.4 nm。

在特定的条件下,双链 DNA 可以从 B 型转变成 A 型或 Z 型,但在正常的细胞环境中能够存在的只有 B 型和 Z 型。引起 DNA 双链构象改变的因素有碱基组成和序列、盐的种类、盐浓度和相对湿度。低湿度下,DNA 可形成 A 型双螺旋。DNA 与 RNA 形成的杂交双链为 A 型双螺旋。嘌呤、嘧啶相间排列的 DNA 在高的盐浓度下可形成左旋的 Z-DNA。而体内 m⁵C 上的甲基化有利于 B 型向 Z 型的转变。体内 Z-DNA 的形成可能与基因表达调控有关。

支持 DNA 为双螺旋结构的证据有 X 线衍射数据、Chargaff 法则和碱基的互变异构性质。稳定双螺旋的因素有氢键、碱基堆积力和阳离子或带正电荷的化合物对磷酸基团的中和,其中起决定性作用的是碱基堆积力。

三链螺旋结构即 H-DNA,它是 DNA 的非标准二级结构,其形成需要至少 DNA 的一条链全部由嘌呤核苷酸组成。在细胞内,H-DNA 经常出现在 DNA 复制、转录和重组的起始位点或调节位点。

DNA 的三级结构为超螺旋,分为正超螺旋和负超螺旋,其中正超螺旋为左手超螺旋,因 DNA 双螺旋过度缠绕引起,负超螺旋为右手超螺旋,因 DNA 双螺旋缠绕不足引起。超螺旋 DNA 可以通过连环数、扭转数、缠绕数和比连环差等几种参数定量地表示。负超螺旋 DNA 很容易解链,故有利于 DNA 的复制、重组和转录。DNA 在复制和转录过程中会形成正超螺旋,但细胞内存在的 DNA 拓扑异构酶可将它们及时清除。

RNA 的二级结构取决于碱基组成,有多种形式。双链 RNA 为 A 型双螺旋。多数 RNA 只有一条链,其二级结构主要由链内碱基的互补性决定,最常见的是茎环(发夹)结构。tRNA 的二级结构像三叶草,含有四个环和四个茎。按照从 5'→3'的顺序,四个环依次是 D 环、反密码子环、可变环和 T ψ C 环。四个茎依次是受体茎、D 茎、反密码子茎和 T ψ C 茎。rRNA 分子上都有大量链内互补的序列,因此 rRNA 高度折叠,不同物种的同一类型的 rRNA 上存在十分保守的折叠样式。

RNA 的三级结构主要有假节结构、“吻式”发夹结构和发夹环突触结构等。tRNA 的三级结构是倒 L 形结构。在这种结构中,氨基酸受体茎位于“L”的一端,而 D 环和 T ψ C 环形成“L”的角,反密码子环位于另一端。这样的结构将 tRNA 的两个功能端很好地分割开。rRNA 天然的三级结构在核糖体内与蛋白质结合在一起,蛋白质对其三级结构有重要的影响。

真核基因组 DNA 呈高度折叠,并与一些特殊的蛋白质结合形成染色质。染色质的蛋白质有组蛋白和非组蛋白。染色质在细胞核为高度压缩的无定形的长纤维,有丝分裂的中期呈高度浓缩。染色质分为常染色质和异染色质,前者在间期浓缩程度很低,染色淡,具有转录活性,而后者浓缩程度高,染色深,一般无转录活性。核小体为染色质的一级结构。

染色体是基因组的结构单位,它由一个 DNA 分子和与它相结合的蛋白质组成。原核生物含有一条连续的环状染色体。真核生物染色体为线形,有多个,其每一条染色体都含有自主复制序列、着丝粒和端粒。真核生物染色体的包装经历了核小体→30 nm 纤维/螺线管→环→玫瑰花瓣→染色体的过程。

DNA 的功能是作为生物体的主要遗传物质。而 RNA 主要的生物功能有:充当 RNA 病毒的遗传物

质;作为核酶;参与翻译;作为引物,参与 DNA 复制;参与 RNA 前体的后加工,参与基因表达的调控,参与蛋白质共翻译定向和分拣,以及参与雌性哺乳动物体细胞内一条 X 染色体的失活。

核酸的性质包括紫外吸收、酸碱解离、变性、复性和水解。其中紫外吸收与碱基有关,酸碱解离与核苷酸有关,变性和复性与核酸的二级结构的破坏和恢复有关,水解与磷酸二酯键裂解有关。

核酸的变性是指核酸受到极端的 pH、热或离子强度的降低等因素或特殊的化学试剂的作用,其双螺旋解链成单链的过程,其中并不涉及共价键断裂。核酸变性时,紫外吸收和浮力密度升高,黏度降低,生物活性可能发生变化,其中紫外吸收增加的现象称为增色效应。当各种变性因素不复存在的时候,变性时解开的互补单链全部或部分恢复到天然双螺旋结构的现象称为复性。热变性 DNA 一般经缓慢冷却后即可复性,此过程称为退火。伴随着 DNA 复性发生的变化是与其变性时发生的正好相反,其中紫外吸收减少的现象称为减色效应。这些性质的变化可用做检测核酸变性或复性的指标。

双链 DNA 热变性是在很窄的温度内发生的,与晶体在熔点时突然熔化的情形相似,因此 DNA 也具有“熔点”,用 T_m 表示。 T_m 实际是 DNA 的双螺旋有一半发生热变性时相应的温度。DNA 的 T_m 受到 DNA 的均一性、G-C 含量、离子强度和其他变性因素的影响。某种 DNA 均一性越高,其 T_m 范围就越窄;在溶剂条件固定的前提下, T_m 的高低取决于 DNA 分子中的 G-C 的含量;DNA 溶液中的离子强度越高, T_m 就越高;容易形成氢键的试剂能破坏碱基对之间的氢键,因此可以降低 DNA 的 T_m 。某些蛋白质因为能够稳定 DNA 单链状态也能降低 T_m 。

RNA 的 T_m 较为复杂,双链 RNA 的 T_m 与 DNA 相近。单链 RNA 内的双螺旋区域有限,因此变性性质变化程度不及 DNA,其 T_m 较低、变性曲线较宽。

影响 DNA 复性的因素有温度、离子强度、DNA 浓度和 DNA 序列的复杂度等。

在研究 DNA 序列对复性速度的影响时,将温度、离子强度、核酸片段大小等其他因素均给以固定,以不同程度的核酸分子重新缔合部分取对数对 Cot 作图,以此来研究 DNA 序列复杂度。在标准条件下测得的复性率达 0.5 时的 Cot 值称为 $Cot_{1/2}$,它与 DNA 序列复杂度成正比。原核生物 DNA 的 $Cot_{1/2}$ 可代表基因组的大小及基因组中碱基序列的复杂度。真核基因组中因含有许多不同程度的重复序列,其 Cot 曲线由若干个 S 形曲线叠加而成。

核酸杂交是一种利用核酸分子变性和复性的性质,将来源不同的核酸片段,按照碱基互补配对规则形成异源双链的技术,它既可以在液相中也可以在固相中进行。Southern 印迹、Northern 印迹和 DNA 芯片都是利用了核酸杂交的性质。

酸、碱和酶均可导致核酸水解。核酸分子内的糖苷键和磷酸二酯键对酸的敏感性不同,糖苷键 > 磷酸二酯键,嘌呤糖苷键 > 嘧啶糖苷键。RNA 的磷酸二酯键对碱极敏感,很容易水解成 2'-或 3'-核苷酸的混合物。核酸可受到多种不同酶的作用而发生水解,不同的酶对底物的专一性、水解的方式和磷酸二酯键的断裂方式并不相同。

核酸的化学合成与多肽的人工合成的原理和基本步骤相似,都需要对核苷酸单体上的活性基团进行保护或去保护,对成键基团进行活化,一般都使用固相合成法等,但合成的方向跟生物合成的反应正好相反,为 $3' \rightarrow 5'$ 。

核酸的分离、纯化的基本原理也和蛋白质差不多,有沉淀、电泳和层析等方法。紫外分光光度法可用于对核酸进行定性和定量分析。

测定 DNA 序列的基本方法有末端终止法、化学断裂法和焦磷酸法等。以在末端终止法基础上发展起来的 DNA 序列自动化分析使用最为广泛。

三、本章疑难点剖析及学习方法

本章内容中关于基因组特别是人类基因组的主要信息和特征需要有所了解,为了方便同学们的记忆,下面使用图表的形式对它们进行总结,参看图 1-1 和表 1-1:

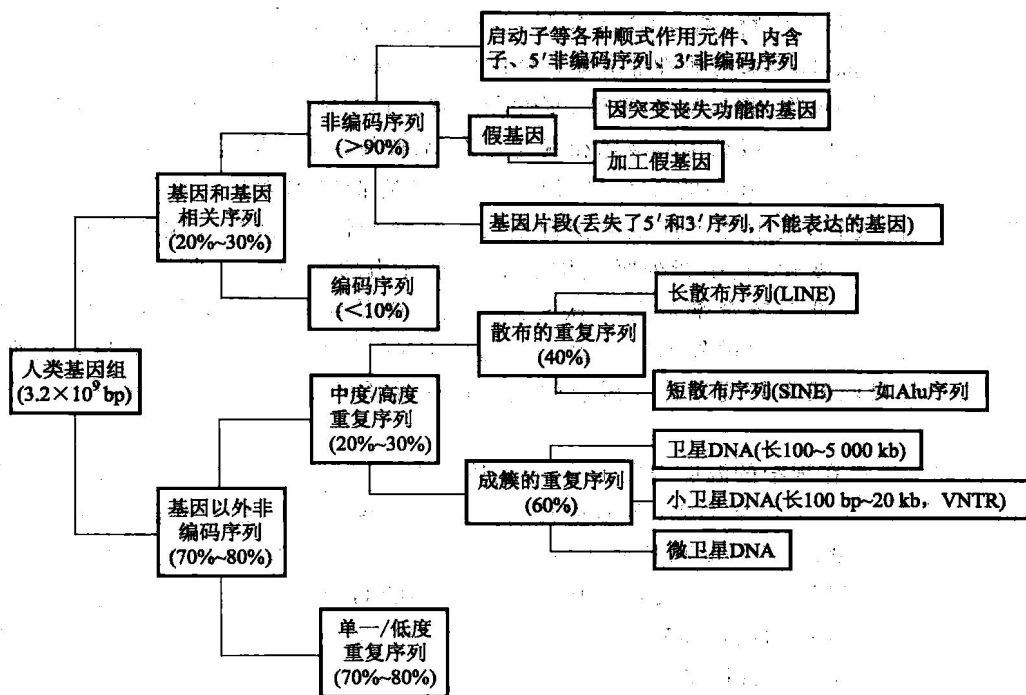


图 1-1 人类基因组的主要特征

表 1-1 人类基因组的主要信息

项 目	性 质
DNA 大小(C 值)	3.2×10^9 bp
蛋白质基因的数目	约 20 500 个
基因数目最大的染色体	1 号染色体
基因数目最少的染色体	Y 染色体
最大的基因	2.4×10^6 nt
基因的平均大小	27 000 nt
一个基因含有的最少的外显子数目	1
一个基因含有的最多的外显子数目	178
基因平均含有的外显子数目	10.4
最大的外显子的大小	17 106 nt
外显子的平均大小	145 nt
最大的内含子的大小	71 000 nt
内含子的平均大小	3 413 nt
假基因的数目	约 20 000
蛋白质外显子(编码区)占基因组的百分比	1.5%
其他高度保守的序列(包括 mRNA 两端 UTR、结构和功能性 RNA、保守的蛋白质结合位点)	3.5%
高度重复序列的比例	约 50%

另外,原核生物和真核生物在基因组结构与功能上的差异,也需要我们有所了解,具体参看表 1-2:

表 1-2 原核细胞与真核细胞在基因组结构与功能上的比较

特 征	原 核 细 胞	真 核 细 胞
DNA 量(信息量)	少	多
DNA 分子数	1	≥ 2
DNA 分子结构	环状	线状
基因组数	1 个	≥ 2 (核基因组、线粒体基因组和叶绿体基因组)
基因数	几千	大约几万
大量的重复序列	—	+
基因中插入内含子	少见	普遍
DNA 与组蛋白结合	几乎裸露,不与或 与少量类组蛋白结合	与 5 种组蛋白结合
核小体和染色质结构	—	+
DNA 复制的明显周期性	—	+
基因表达的调控	通常是负调控,功能 相关基因组成操纵子	通常是正调控,很少 形成操纵子结构
转录与翻译的时空关系	偶联	核内转录,细胞质内翻译
细胞复制与分裂 (DNA 传递与分配)	无丝分裂	有丝分裂或减数分裂

除了基因组和基因组学这样的概念以外,还有其他相关的“组”及“组学”的概念需要掌握,它们总结在表 1-3 中,而它们之间的相互关系参看图 1-2:

表 1-3 各式各样的“组”与“组学”

名 称	定 义	研 究 方 法	相 关 学 科
基因组 (genome)	一种生物含有的所有 DNA 序列,包括 基因和非基因序列	大规模 DNA 序列测定和 分析	基因组学 (genomics)
蛋白质组 (proteome)	一种细胞、组织或完整生物体所拥有的 全套蛋白质	二维电泳和质谱	蛋白质组学 (proteomics)
生理组 (physiome)	在分子、细胞、组织、器官和个体的水平 上对各种动物及人类的解剖学、生物物 理及生理学的数据,加以量化、数字化及 模式化的描述	数据处理、模型建立、计算 机应用和各种生化、生理、 生物医学工程等技术及 方法	生理组学 (physiomics)
转录组 (transcriptome)	一个活细胞所能转录出来的所有 mRNA	DNA 芯片技术	转录组学 (transcriptomics)
代谢组(metabolome)	一种生物样品内所有的小分子代谢物	分离:GC、毛细管电泳、 HPLC,检测:质谱、NMR	代谢组学 (metabolomics)
脂质组 (lipidome)	一个细胞、一个组织或一个生物体内所有 的脂质	质谱、NMR、荧光光谱、双 偏振极化干涉测量分析系 统(DPI)	脂质组学 (lipidomics)
互作组 (interactome)	一个细胞内由蛋白质之间、蛋白质与其 他分子之间的相互作用而形成的网络	酵母双杂交、串联亲和纯 化(TAP)	互作组学 (interactomics)

随着对基因组研究的发展,人们发现,在同一物种内的不同个体之间表现各种形式的多态性,其中单核苷酸多态性(SNP)主要是指在基因组水平上由单个核苷酸的变异所引起的 DNA 序列多态性。它是人类可遗传的变异中最常见的一种,占有已知多态性的 90% 以上。SNP 在人类基因组中广泛存在,平均每 500~1000 bp 中就有 1 个,估计其总数可达 300 万个甚至更多。SNP 所表现的多态性只涉及单个碱基的变异,这种变异可由单个碱基的转换或颠换所引起,也可由碱基的插入或缺失所致。但通常所说的 SNP 并不包括后两种情况。目前已建立多种检测 DNA 序列多态性的方法,如表 1-4 所示。

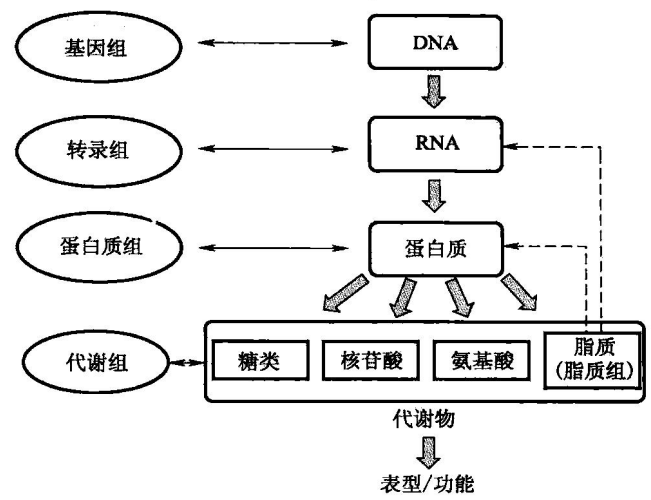


图 1-2 不同“组”之间的相互关系

表 1-4 检测 DNA 序列多态性的各种方法的原理和操纵步骤

检测手段	实 例	原 理	基本操作步骤
酶 学	限制性酶切片段长度多态性(RFLP)	由于单个碱基的缺失、重复和插入所引起限制性酶位点的变化,而导致限制性酶切 DNA 片段长度的变化	限制性酶酶切 DNA→限制性酶切片段进行凝胶电泳→转膜→变性→与标记探针杂交、洗膜→结果分析
	扩增片段长度多态性(AFLP)	首先利用可产生黏端的限制性酶酶切目的基因组序列,然后将这些长度不同的酶切片段与含有相同黏端的人工接头连接,形成模板;再在模板末端添加具有选择性核苷酸(1~3 个)的不同引物,进行 PCR 扩增(只有两端序列与选择性核苷酸配对的酶切片段被扩增);最后将扩增片段在高分辨率的测序胶上电泳,分析比较野生型和突变型基因组 DNA 在扩增片段的长度与数量的差异	限制性酶酶切基因组 DNA→酶切片段连接人工接头(有共同黏端)制取模板→模板末端添加具选择性核苷酸的不同引物→PCR 扩增→高分辨率凝胶电泳→结果分析
	裂解酶片段长度多态性(CFLP)	DNA 变性后产生的单链 DNA 片段自我折叠成带有发夹结构的构象。当 DNA 发生突变的时候,会或多或少地影响其空间构象,从而使 DNA 对特定的核酸内切酶的敏感性发生变化。这种方法所使用的核酸内切酶识别和切割的位点在单链和双链之间的连接处	DNA(野生型或突变型)热变性→冷却到事先设定的最佳温度,使得单链 DNA 自我折叠成序列依赖性的二级结构→特定的核酸内切酶消化→电泳分析酶切片段→比较野生型和突变型在酶切图谱上的差异
	随机扩增多态性 DNA(RAPD)	使用一系列长度约为 10 个碱基的随机引物,对目的基因组 DNA 进行 PCR 扩增。引物结合位点上的序列改变以及两扩增位点之间 DNA 碱基的突变均可导致扩增片段数目和长度的差异。扩增产物通过凝胶电泳分离后,经 EB 染色或放射自显影来检测扩增产物片段的多态性,这些扩增片段的多态性反映了基因组相应区域的 DNA 多态性	抽提基因组 DNA→使用随机引物进行 PCR 扩增(由于整个基因组存在众多反向重复序列,因此须对每一随机引物单独进行 PCR)→琼脂糖凝胶电泳分离→EB 染色或放射自显影来检测 DNA 片段的多态性
电 泳	单链构象多态性(SSCP)	单链 DNA 片段呈复杂的三维折叠构象,它是内部互补碱基自我配对的结果。然而,当有一个碱基发生改变时,会影响到碱基之间的配对,从而使构象发生改变。空间构象有差异的单链 DNA 分子在 PAGE 中受排阻力不同,通过非变性 PAGE,可以非常灵敏地将构象上有差异的分子分离开,进而推断该 DNA 片段中有碱基突变	PCR 扩增靶 DNA→将特异的 PCR 扩增产物变性→快速复性,使之成为具有一定构象的单链 DNA 分子→非变性 PAGE→通过放射性自显影、银染或 EB 染色,分析结果