

# 第 1 章 绪 论

20 世纪 70 年代初，在生命科学发展史上发生了一个伟大的事件，美国科学家 S.Cohen 第一次将两个不同的质粒加以拼接，组合成一个杂合质粒，并将其引入大肠杆菌体内表达。这种被称为基因转移或 DNA 重组的技术立即在学术界引起了很大的震动。很多科学家深刻认识到这一发现所包含的深层含义以及将会给生命科学带来的巨大变化，惊呼生命科学一个新时代的到来 并且预言 21 世纪将是生物科学的世纪。由于基因转移是将不同的生命元件按照类似于工程学的方法组装在一起，生产出人们所期待的生命物质，因此也被称为基因工程。基因工程的出现使人类跨进了按照自己的意愿创建新生物的伟大时代。虽然从它的诞生至今不足 40 年，但这一学科却获得了突飞猛进的发展。本书将着重介绍基因工程学的基本原理以及它的一些应用。

## 1.1 基因与基因工程

### 1.1.1 基因的概念

人们对基因的认识经历了长时间的发展过程，而且随着生命科学的发展，基因的概念还在不断深化。

1866 年，遗传学的始祖孟德尔(G. J. Mendel)在他的豌豆杂交实验论文中，将控制性状的遗传因素称为遗传因子，并且用大写字母代表显性性状，用小写字母表示隐性性状。虽然当时孟德尔对遗传的物质基础一无所知，但事实上他所讲的遗传因子已经形成了基因的雏形。时至今日，在遗传学的分析上我们还经常用这些字母来表示所分析的基因。

1909 年，丹麦的遗传学家 W. L. Johanssen 首次提出用“gene”来代替孟德尔的遗传因子（我国著名遗传学家谈家桢先生首先将 gene 翻译为基因），提出了基因型与表现型的区别，指出前者是一个生物的基因成分，后者是这一基因表现的性状。当时提出的基因概念仅仅是一代表遗传性状符号的改变，并未涉及遗传的物质概念。

1910 年以后，美国遗传学家以果蝇为材料进行杂交实验，第一次把代表某一个性状的特 定基因与某一特定染色体上的特定位置联系起来，发现了连锁交换定律。摩尔根(T. H. Morgan)提出了遗传粒子理论，认为基因是一粒一粒在染色体上呈直线排列的，且互不重叠，就像连在线上的佛珠一样。摩尔根理论的重要性在于基因已不再是一个抽象的符号，而是与染色体紧密相关的一个实体。

20 世纪 40 年代初，物理学家和化学家把研究方向转移到对基因本质问题的探讨上。1944 年 Avery 等首次证实遗传的物质基础是 DNA，把基因位于染色体上的理论进一步推进到基因位于 DNA 上。1953 年 Watson 和 Crick 提出了 DNA 双螺旋结构模型 这时人们接受了基因是具有一定遗传效应的 DNA 片段的观念。

1955年, Benzer 基于 T4 噬菌体的顺反互补试验, 提出了顺反子的概念。过去人们认为基因是三合一体, 即既是一个功能单位, 也是一个突变单位和一个交换单位。Benzer 通过研究证实, 一个基因内部的许多位点可以发生突变, 并且可以在这些位点之间发生交换, 说明一个基因并不是一个突变单位或一个交换单位。实际上顺反子要比突变单位或重组单位大得多。一个顺反子内部可以发生突变或重组, 即包含着许多突变子和重组子。到此为止, 已经从功能单位的意义上把顺反子和基因统一起来了, 顺反子实际上成为基因的同义词。

20 世纪 60 年代 法国遗传学家 F. Jacob 和 J. Monod 在研究细菌基因调控中证实: 基因是可分的, 功能上是有差别的, 即既有决定合成某种蛋白质的结构基因, 又有编码阻遏或激活结构基因转录和合成蛋白质的调节基因, 还有其他无翻译产物的基因。操纵基因的发现修正了一个基因就有一条多肽, 或决定一个蛋白质功能的结构单位的说法, 同时也提出了顺反子代替基因概念的不确切性。

20 世纪 70 年代以后 人们陆续发现了断裂基因、重叠基因、跳跃基因 使对基因的认识更进一步深化。

科学家在比较 DNA 序列与相应的 mRNA 序列以后发现: 一个基因往往由几个互不相邻的段落组成; 它的内部还包含一段或几段最终不相应出现在成熟 mRNA 中的片段的基因, 这些不相应出现在成熟 mRNA 中的片段称为内含子, 而相应出现在成熟 mRNA 中的片段则称为外显子; 有时一个基因可被几百个至几千个碱基所间隔, 经过转录加工后在成熟的 mRNA 中被除去。在珠蛋白基因、卵清蛋白基因、rDNA、tRNA 等的基因中均发现这种间隔的片段。

1977 年, F. Sanger 在测定噬菌体  $\Phi$ X174 全部核苷酸序列时发现 D 基因中包含着基因 E。基因 E 的第一个密码子从基因 D 的中央一个密码子 TAT 的中间开始 因此 两个部分重叠的基因所编码的两个蛋白质大小不等, 氨基酸的组成也不同。

可移动遗传因子 mobile genetic element 的发现动摇了基因是带有一定遗传信息的稳定结构的观念, 使人们认识到也有跳跃的遗传因子。

综上所述, 我们对基因的认识可以肯定以下几点: 基因是实体, 它的物质基础是 DNA 或 RNA)。基因是具有一定遗传效应的 DNA 分子中的特定核苷酸序列。基因是遗传信息传递和性状分化发育的依据。基因是可分的。根据基因的产物可将其分为编码蛋白质的基因 (包括编码酶和结构蛋白的结构基因以及编码作用于结构基因的阻遏蛋白或激活蛋白的调节基因) 无翻译产物的基因 (如转录成为 RNA 以后不再翻译成为蛋白质的转运核糖核酸 tRNA 基因和核糖体核酸 rRNA 基因) 以及不转录的 DNA 区段 (如启动区、操纵基因等)。概括说来, 基因是一个含有特定遗传信息的核苷酸序列, 它是遗传物质的最小功能单位。

### 1.1.2 基因工程与生物工程的关系

生物工程亦称生物技术, 是 20 世纪 70 年代初在分子生物学、细胞生物学和遗传学基础上发展起来的一个新兴领域。它主要包括以下 5 个方面。

(1) 基因工程 对不同生物的遗传物质——基因 在体外进行剪切、组合和拼接 使遗传物质重新组合 然后通过载体 质粒、噬菌体或病毒等 转入微生物、植物或动物细胞内 进行无性繁殖, 并使所需要的基因在细胞中表达, 产生出人类所需要的产物或建组成新的生物类型。

(2) 细胞工程 包括细胞融合、细胞大规模培养以及植物组织培养快速繁殖技术。细胞融合技术是指将两种不同种类的细胞, 通过化学、生物学或物理学手段使之融合在一起, 从而产

生出兼备两个亲本的遗传特性的新的细胞。细胞大规模培养技术是以工业化生产为目的，摆脱气候、产地、季节的限制，从大量培养的细胞中获得药物或其他有用物质。植物组织培养快速繁殖技术是利用植物细胞的全能性由扩增的细胞分化再生成植株，这样就有可能用细胞器官和组织的再生苗来代替种子实生苗，无限地扩大繁殖系数。

(3)酶工程 包括酶的生产应用、酶和细胞的固定化以及酶的分子修饰技术。酶是生物体内产生的具有催化作用的蛋白质。其催化效率超出化学催化千百倍，而且是在常温、常压下进行，专一地催化某一反应。所谓酶工程，就是在一定的生物反应器中，利用酶的催化作用将相应的原料转化成有用物质的技术。

(4)微生物发酵工程 包括菌种选育、菌体生产利用、代谢产品的生产利用以及微生物机能的利用技术。微生物发酵工程是利用微生物的特定性状，通过现代化工程技术，生产有用物质或直接应用于工业生产的一种技术体系。

(5)生化工程 包括生物反应器设计制造、传感器的研制以及产物的分离提取和精制技术。

以上 5 个方面的工程技术系统是相互依赖、相辅相成的，但在这些技术系统中，基因工程占主导地位。因为，只有用基因工程改造过的微生物和细胞，才能真正按照人们的意愿进行工程设计，产生出特定的生物工程产品。而微生物发酵工程又常常是基因工程的基础和必备条件。生化工程是其他生物工程技术转化为生产力时所必不可少的重要环节。正是由这 5 个工程技术系统，共同组成了现代生物工程学。从图 1-1 我们可以更清楚地理解它们之间的关系。

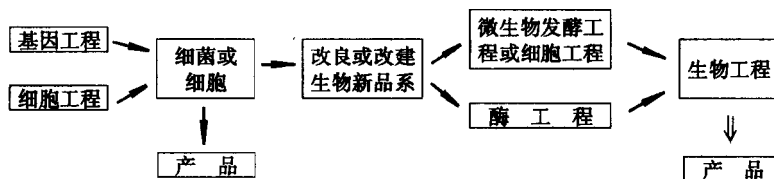


图 1-1 基因工程与生物工程的关系

## 1.2 基因工程的操作和应用

### 1.2.1 基因工程的操作流程

一个完整的基因工程流程一般包括目的基因的获得、载体的制备、重组体的制备、基因的转移、基因的表达、基因工程产品的分离提纯等过程。

传统的基因工程操作是将真核生物细胞的基因在原核生物细胞内表达。这一过程包括真核细胞基因的分离、目的基因与载体分子在细胞外的重组、重组体分子转化原核细胞等环节（图 1-2）。

### 1.2.2 基因工程的应用

尽管基因工程出现后的一段时间内带给人们的是猜疑和恐惧，但它还是以迅猛的速度发展。实践表明，基因工程会给人类带来难以估量的经济效益和社会效益。特别是对人类所面

临的能源、粮食、人口、环境和疾病等日趋严重的社会问题，基因工程正在并且将要发挥越来越大的作用。

### 1.2.2.1 基因工程与农业

基因工程在农业中的应用主要包括提高植物光合作用效率、扩展植物的固氮能力、生产转基因植物和转基因动物等。

(1) 光合作用 光合作用是指绿色植物将大气中的二氧化碳转化为碳水化合物，并向周围环境释放氧气的过程。这一过程是在叶绿体中进行的，叶绿体是绿色植物特有的细胞器，其功能与生存是受核基因组控制的。

绿色植物光合作用的产物约占植物干重的95%以上，它是地球上一切动物（包括大多数微生物）的生命源泉，同时也是人类社会的主要物质和能量的来源。然而，地球上的植物利用太阳能的效率相当低。据统计，农作物的产量还不到转变为生物量的太阳能的5%。因此提高光合作用效率具有重要意义。应用基因工程技术，已经克隆了许多种参与光合作用的基因并分析了光对基因表达的调节作用。目前的工作主要有如下两个方面：一是深入研究在 $\text{CO}_2$ 的固定反应中起关键作用的二磷酸核酮糖羧化酶（RuBisCo）以便提高其与 $\text{CO}_2$ 的亲合力，以及取消或减少光呼吸的竞争反应。实验表明，通过交换 RuBisCo 亚基的基因，将不同来源的基因导入同一种植物，形成具有异源亚基的 RuBisCo 基因或是采用定点突变技术改变 RuBisCo 的活性增加其同 $\text{CO}_2$ 的亲合力；甚至用更为有效的突变基因，取代正常的 RuBisCo 基因等办法，将有可能提高植物对 $\text{CO}_2$ 的固定效率。二是提高光能吸收及转化效率。实验表明，通过在不同植物之间交换光系统的组合，或是利用体外定点突变技术改变光系统的组分并实现优化组合，便有可能使其转能效率达到最佳的水平，从而提高光合作用的效率。

(2) 固氮作用 固氮作用通常是指豆科植物（如蚕豆、豌豆和三叶草等）将空气中的氮（ $\text{N}_2$ ）转变为氨（ $\text{NH}_3$ ）的过程。它是通过与其共生的根瘤菌属 *Rhizobium* 细菌实现的。

大多数植物都需要大量的可溶性氮才能很好生长。虽然地球上每年由微生物固定的 $\text{N}_2$ 的总量约200 Mt左右，但只有豆科植物能与根瘤菌共生固氮。如果禾谷类作物及其他非豆科植物都能够具有天然固氮的能力，或转变为根瘤菌的宿主，那么，在农业生产上将节省大量的化肥，具有重大意义。要使普通的非固氮植物的细胞从遗传上转变为具有固氮功能的特殊细胞必须具备如下5个条件：①根瘤菌的全部 nif 基因都能在同一植物细胞中适当地表达；②固氮酶复合体能正确地加工和组装；③具有一个厌氧的环境；④提供足够的 ATP；⑤提供 NADPH。这是一项十分复杂而艰巨的工作。总的来说，目前主要有如下两种方法：一种是用带有 nif 基因的质粒转化植物细胞的叶绿体，从而有可能使用正常的原核信号进行表达，而不

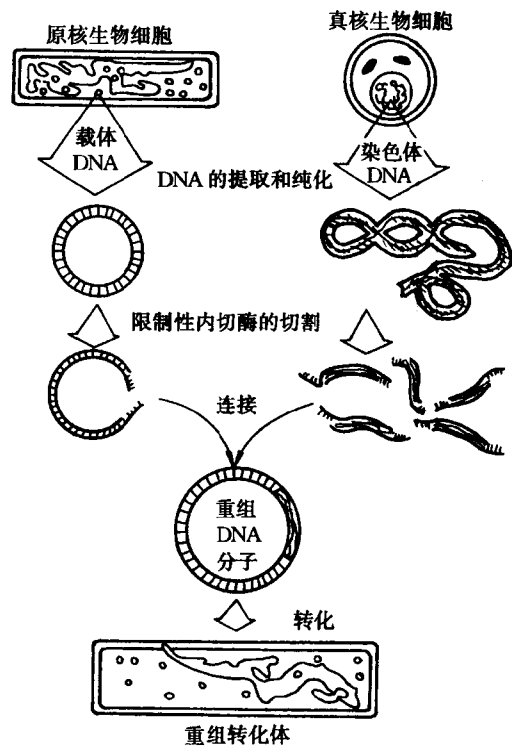


图 1-2 典型的重组 DNA 实验

必将 17 个 *nif* 基因都置于植物细胞核基因组启动子的控制之下；另一种是把豆科植物的固氮基因转移到其他植物中，使其对固氮菌的感染产生相应的反应。到目前为止，已有许多植物的根瘤蛋白基因 *nodulin gene* 被克隆出来 而且还建立了一种三叶草 (*Lotus corniculatus*) 的根瘤形成模型。

(3)转基因植物 转基因植物是指将克隆到的一些编码特殊性状的基因，通过生物、物理和化学等方法，导入到受体植物细胞，然后进行组织培养而培育出再生植株。它是农业生物技术的主要内容，为农业育种提供了一条十分有用的途径。人们可以在一定范围内根据自己的意愿来改造植物的一些性状，从而获得高产、稳定、优质和抗逆性强的品种。自从 20 世纪 80 年代初首次从细菌中分离到一些分解抗生素的基因，并转移到植物细胞中获得第一株能在抗生素培养基上生长的植株以来，已有近百种转基因植物相继问世，这些植物有烟草、番茄、马铃薯、矮牵牛、胡萝卜、向日葵、油菜、苜蓿、亚麻、甜菜、棉花、芹菜、荷花、黄瓜、拟南芥、大白菜、大豆、水稻、玉米、莴苣、红豆以及稗麦等等。

应用转基因植物技术 不但可以培育出抗病毒、抗真菌、抗虫害、抗逆性、抗镉或抗除草剂的植物，而且可以获得雄性不育植株或增加种子的营养价值，此外，还可以应用转基因植物技术使植物果实变硬便于储运，或改变花卉的颜色提高观赏价值，以及在转基因植物中生产一些医药上应用的多肽。

(4)转基因动物 转基因动物是指用实验方法导入的外源基因在染色体基因组内稳定整合并能遗传给后代的一类动物。自从 20 世纪 80 年代初美国首次将大鼠生长激素基因导入小鼠受精卵的雄原核中 获得了个体比对照组大一倍的转基因‘超级鼠’后 转基因昆虫、猪、鱼、兔、羊和牛等相继问世，不但为动物基因工程育种提供了新的途径，而且可以作为生物反应器生产各种有用的蛋白质，特别是医用活性肽。此外，还可以通过转基因技术培育抗病的转基因家畜，使其免遭传染病的危害。

(5)产生次生代谢产物 植物提供了全世界 25% 的药物资源 并产生出化学物质 如生物碱等) 及生物物质 (如各种必需氨基酸等)。应用基因工程技术，结合植物组织培养方法有可能对其编码的药物基因进行改造，以提高有效成分的合成效率并确保其生物活性，甚至可以生产出崭新性质的植物生化药物。

#### 1.2.2.2 基因工程与工业

基因工程在工业中的应用主要包括纤维素的开发利用、酿酒工业、食品工业、制药工业和新型蛋白质的生产等方面。

(1)纤维素的开发利用 纤维素是植物的主要组成部分。据估计，全世界的纤维素资源总量约为  $7 \times 10^{11}$  t 而每年经绿色植物光合作用合成的纤维素又可高达  $4 \times 10^{10}$  t 因此 纤维素被认为是地球上数量最丰富的有机物质。从化学分子结构上看，纤维素是一种无水葡萄糖的线性多体分子，其重复单位叫纤维二糖。纤维素完全降解后的产物葡萄糖是食品、燃料和化学原料的重要来源。

由于纤维素通常是以不溶性的纤维成晶状排列形式存在的，再加上纤维素分子与其他多糖 (如半纤维素和果胶等) 结合在一起，而其外又包裹上木质素，致使消化酶分子难以接近，因此，纤维素是很难降解的。植物材料中纤维素的天然降解主要是由丝状真菌发酵引起的。现在已从细菌和丝状真菌中克隆出了各种纤维素分解酶的基因。如果通过基因工程方法，把这些基因导入酿酒酵母 (*S. cerevisiae*)，并使酿酒酵母具备分泌纤维素酶的能力，那么就有可能

将纤维素降解成葡萄糖，再发酵成酒精，从而实现酒精生产流程一步化的新工艺。

(2)酿酒工业 酿酒酵母不仅是一种在酿酒工业中广泛使用的发酵微生物，而且是一种很有用的基因操作菌株之一。如果把面包酵母 (*S. diastaticus* 基因组中编码淀粉  $\alpha$ -1,4-葡萄糖苷酶的 DEX 基因引入到酿酒酵母细胞，产生出一种新的酵母菌株，就可以克服酿酒酵母不能发酵糊精 (含 22% 的碳水化合物) 的缺点，生产出碳水化合物含量低、味道好的优质啤酒。如果把能够降解具有极高相对分子质量的分枝糊精 (branched dextrans) 的淀粉酶基因导入酿酒酵母，则可进一步改良啤酒的质量。还有，如果把木瓜蛋白酶基因引入酿酒酵母，则可保持啤酒的清晰度。此外，还可以应用体外突变技术主动改变这些酶的特性，使其稳定性增加。总之，在酿酒工业中，基因工程技术是大有可为的。

(3)食品工业 在食品工业中，干酪的生产离不开凝乳酶对乳蛋白-酪蛋白的切割。凝乳酶是从哺乳小牛的第四个胃中提取的，很不经济。现在已经将小牛的凝乳酶基因克隆出来，并在酿酒酵母中实现了表达，生产出高产量的、具有全部天然活性的凝乳酶，它能够使牛奶凝固。

干酪生产过程中的废物乳清含有 4%~5% 的乳糖、少量蛋白质、大量矿物质和维生素。其中的乳糖若被降解成葡萄糖和半乳糖，便能被酿酒酵母发酵，生产出酒精和单细胞蛋白质。现已把克鲁维酵母乳酸菌 *Kluyveromyces lactis* 的  $\beta$ -半乳糖苷酶基因和乳糖透性酶基因转移到酿酒酵母中，虽然仍需改进表达这两种酶的方法，但经过遗传变异后的酵母最终能够用乳清做底物生产出燃料酒精、饮用酒精和生物饮料，从而达到底物综合利用的目的。

此外，近年来也开始利用基因工程技术提高酿酒过程废弃酵母的经济价值。其办法是将带有高度调节性能启动子的表达载体导入酿酒酵母，使其在发酵过程中关闭，而当废弃酵母重新悬浮在诱导培养基时开动起来，从而生产出需求量很大的蛋白质，例如凝乳酶和血清蛋白等。

(4)制药工业 传统的制药工业，要么依靠化学合成，要么从自然界中筛选药物产生菌，然后通过发酵分离提取获得，这两者都费时费力。应用基因工程技术，不但可以提高药物的产量，而且可以创造药物新品种。目前已商品化生产的基因工程药物有各种抗生素和多肽药物，多达百种以上。我国已能自行生产基因工程干扰素、红细胞生成素 (EPO)、白介素和心钠素等。

(5)新型蛋白质的生产 利用基因工程技术，不仅能生产真核基因编码的蛋白质，而且还能够生产出新型的蛋白质。其最简单的途径是，利用定点突变技术重新设计酶分子的结构，以增加酶的稳定性，改变酶作用底物的稳定性，以及把有关联的酶构成一种多酶复合物，这对酶制剂工业生产显然具有相当重要的意义。

#### 1.2.2.3 基因工程与环境保护

基因工程在环境监测与净化领域的研究与应用中已经发挥了重大作用，预示着十分光明的前景。

(1)环境监测 已有报道，可应用基因探针检测出水中，特别是饮用水中的病毒。使用一个特定的核酸片段 (DNA 或 RNA 作为探针 使之同被检测的病毒 DNA 互补的碱基结合 从而把病毒检测出来。这种方法的特点是快速灵敏。传统的方法检测一次需耗时数天或几周，准确性也差；而用探针只需不到一天，且能在 1 000 升水中检测出 10 个病毒来。现在已有用于检测 10 种病毒的不同探针。使探针同从环境中分离出来的细菌 DNA 杂交 从而确定某种有危害性细菌的存在。有人应用 DNA 探针在 400 个不同地方检测沙门氏杆菌获得成功，说

明在鉴定带菌者及预防流行病方面可应用基因跟踪法。

(2)环境污染净化 有报道称,把4种不同假单孢杆菌的质粒重组成一个“超级质粒”由OCT降解辛烷、己烷、癸烷)、XYL(降解二甲苯和甲苯)、CAM分解樟脑和NAH降解萘构成一个质粒并送入细菌获得了“超级菌”。这种“超级菌”能在原油中迅速繁殖因为它代谢碳氮化合物的活性比任何一种含单个质粒的细菌都强大得多。这种“超级菌”能够在浮游过程中除去污染了水面的石油,几小时就可以降解2/3的烃类物质,而天然菌则需耗时一年以上才能达到同样的效果。把嗜油酸单孢杆菌的耐汞基因转移入腐臭假单孢杆菌中,该菌能把剧毒的汞化物吸收到细胞内,还原成金属汞,然后可通过气化的方法从菌体中回收金属汞,此法可用于净化汞污染。有报道称,用以下方法消除镉污染:把原在中国仓鼠中的屏蔽基因(即可将重金属离子排去的基因)植入一种十字花科植物——芜菁的体内,此种植物便可将土壤中的镉留在植物根部,阻止它到达植物的茎、叶、果实部位。这对保护人、畜的健康很有好处。关于净化农药DDT残留问题可从抗DDT的害虫中分离出抗DDT基因,然后将此基因转移入细菌体内将这种“超级菌”投入到土壤中可把农田中残留的DDT降解掉。

关于分解四氯联苯乙烷(TGE,它可引起癌症和肝病),一种假单孢杆菌(*Pseudomonas cetacia*)能产生一种酶可使TGE分解成单盐离子和二氧化碳。另外,科学家还培育出可作为降解氯化物溶剂、表面活性剂、硫化物和多氯联苯的新菌种。已发现一种细菌的突变株能够“吃”矿山产生的含氟废物。还开发出一种磷的浓缩(phostrip)技术即利用微生物浓缩磷可清除废水中99.3%的磷。已分离出一种微生物能清除地下储罐溢出的原油。在加入营养物和空气后在6个月内该微生物能把溢油分解成二氧化碳和水。生物降解法比污染土壤的填埋法可节省75%的费用。运用基因工程的方法已构建了能分解煤炭中硫成分的微生物和能降解四氯联苯乙烷的微生物。

还有一些科学家通过基因重组构建新的生物杀虫剂以取代化学农药。现在已有多种生物杀虫剂进入了大田试验。

#### 1.2.2.4 基因工程与医学

基因工程在医学中的应用极其广泛,除上述利用转基因植物生产生化药物和基因工程多肽药物外,还可以利用基因工程技术生产疫苗并进行诊断和治疗疾病等。

(1)基因工程疫苗的研制与生产 以基因工程疫苗为主体的新型疫苗的研制是现代生物技术热点之一,其主要对象是:①不能或难以培养的病原体,如乙肝(HBV)、丙肝(HCV)、戊肝(HEV)、EB病毒(EBV)、巨细胞病毒(CMV)、人乳头瘤病毒(HPV)、麻风杆菌、疟原虫和血吸虫等。有潜在致癌性或免疫病理作用的病原体,前者如I型嗜人T淋巴细胞病毒(HTLV-I)、人免疫缺陷病毒(HIV)、单疱疹病毒(HSV),后者如呼吸道合胞病毒(RSV)、登革热病毒(DGV),可能还有肾综合症出血热病毒(HFRSV)。③常规疫苗效果差如霍乱和痢疾或反应大,如百日咳和伤寒等疫苗。可大大节约成本、简化免疫程序的多价疫苗,如以痘苗病毒、腺病毒、卡介苗或沙门氏菌属为载体的多价活疫苗。此外,利用基因工程技术还可能为目前尚无有效疫苗的某些疾病(例如艾滋病)生产出有效的疫苗。目前已商品化生产的基因工程疫苗共达数十种。基因工程疫苗研制的新进展是双特异性抗体和多价卡介苗(BCG)。

(2)基因诊断 基因诊断是指在基因水平上对疾病的诊断。其主要特点是:特异性强、灵敏度高、简便和快速。应用DNA探针技术可以对遗传病、传染病(包括艾滋病)、心血管疾病、癌症和职业病等进行基因诊断。此外,目前还发展出DNA指纹图分析法和限制性片段长

度多态性 RFLP) 基因连锁分析法两种新技术, 分别广泛应用于法医领域和基因定位与诊断等。

(3)基因治疗 基因治疗一般是指将正常的外源基因导入生物体靶细胞内, 以弥补所缺失的基因、关闭或降低异常表达的基因, 以达到治疗某种疾病的目的。当前, 应用基因工程技术治疗的疾病有遗传病、恶性肿瘤、心血管疾病、糖尿病和传染病(包括艾滋病)等。基因治疗研究虽然已经取得很大进展, 但总的来看, 现在仍处于探索性阶段。

#### 参考文献

- 1 陈永青 王文华 . 微生物遗传学导论 . 上海: 复旦大学出版社, 1990
- 2 卢圣栋 . 现代分子生物学实验技术 . 北京: 高等教育出版社, 1993
- 3 贾士荣 . 农业生物技术进展与展望 . 合肥: 中国科学技术大学出版社, 1993
- 4 阮力 汪垣 强伯勤 . 新型疫苗研究的现状与展望 . 北京: 学苑出版社, 1992
- 5 吴冠云 方福德 . 基因诊断技术及应用 . 北京: 北京医科大学中国协和医科大学联合出版社, 1992
- 6 Booth, W. Of Mice, Oncogenes and Rifkin. Science, 1988, 239: 341~345
- 7 Singer M, Soll D. Guidelines for DNA Hybrid Molecules. Science, 1973, 181: 1114~1116
- 8 Paul V K (ed.). Genetic Engineering Applications for Industry. Park Ridge, New Jersey: Noyes Data Corporation, 1981

## 第 2 章 生物大分子

生命是物质进化到达一定阶段以后的产物，在此之前存在于地球上的矿物质并不能导致生命的出现，只有当核苷酸、脂肪酸、氨基酸、单糖等形成之后，生命的诞生才具备了条件。核酸（DNA 和 RNA）、蛋白质是机体的两类最基本的生物大分子。它们都属于信息分子，DNA 是遗传信息的原初载体，蛋白质是遗传信息的表现者，也就是 DNA 代表信息，蛋白质代表由此信息所规定的功能，两者相互依存。在 DNA 和 mRNA 之间，遗传信息的传递是通过碱基互补实现的；在 mRNA 和多肽链之间，由 tRNA 作为中介使 mRNA 的核苷酸顺序转变为氨基酸顺序。

DNA 由 4 种核苷酸（A, T, C, G）组成，核苷酸的排列顺序代表了核酸的语言。蛋白质则由 20 种基本  $\alpha$ -氨基酸组成。氨基酸的排列顺序代表了蛋白质的语言，两者存在着对应关系（图 2-1）。多肽链的氨基酸排列顺序并不能直接表现出功能，只有当多肽链折叠成特定的三维结构后才表现出功能。但多肽链的氨基酸顺序包括了它折叠的全部信息。

多肽链是怎样折叠成三维结构的呢？

### 2.1 蛋白质的结构

蛋白质的结构可分为多个层次，即一级结构、二级结构、三级结构和四级结构。

#### 2.1.1 蛋白质的一级结构

蛋白质的一级结构即多肽链的氨基酸顺序。多肽链是蛋白质结构的基础，它是由  $\alpha$ -氨基酸按一定的顺序通过肽键连接而成的。肽键是由一个氨基酸的  $\alpha$ -羧基与另一个氨基酸的  $\alpha$ -氨基间失水缩合而成的（图 2-2）。

基本的蛋白质氨基酸有 20 种，由于  $\alpha$ -碳上连接基团的不同决定了它所组成的多肽链的构象和所能表现的功能不同（表 2-1）。

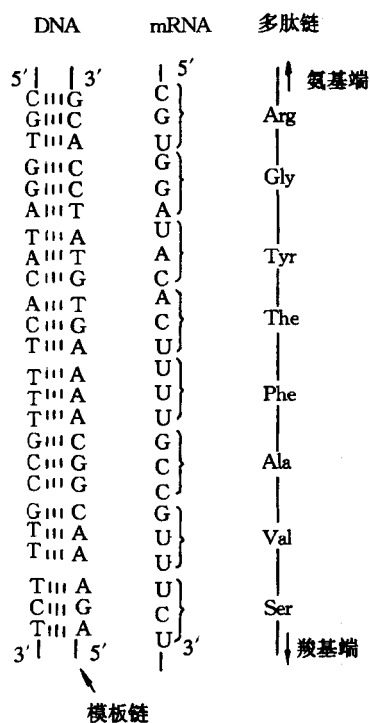


图 2-1 核苷酸和氨基酸顺序间的对应关系



酸残基或多肽链中的大部分残基。氢键是稳定二级结构的主要作用力。蛋白质的二级结构有  $\alpha$ -螺旋、 $\beta$ -折叠和  $\beta$ -转角三种基本类型。它们广泛存在于球状蛋白质内（图 2-3）。 $\alpha$ -螺旋中每个残基  $C_\alpha$  的成对二面角  $\Phi$  和  $\Psi$  各自取一数值， $\Phi = -60^\circ$ ， $\Psi = -45^\circ \sim 50^\circ$ 。多肽主链成螺旋走向，螺旋的半径为 0.25 nm 每周螺旋含 3.6 个氨基酸残基，沿螺旋轴方向上升 0.54 nm 每个残基绕轴旋转  $100^\circ$  沿轴上升 0.15 nm。 $\alpha$ -螺旋中，上下两圈主链的  $\text{N}-\text{H}$  和  $\text{C}=\text{O}$  间都形成了氢键；在螺旋内部存在范德华力，R基则突出于螺旋的外侧。 $\alpha$ -螺旋在动力学上是稳定的构象。 $\beta$ -折叠是蛋白质中第二种最常见的二级结构。两条或多条几乎

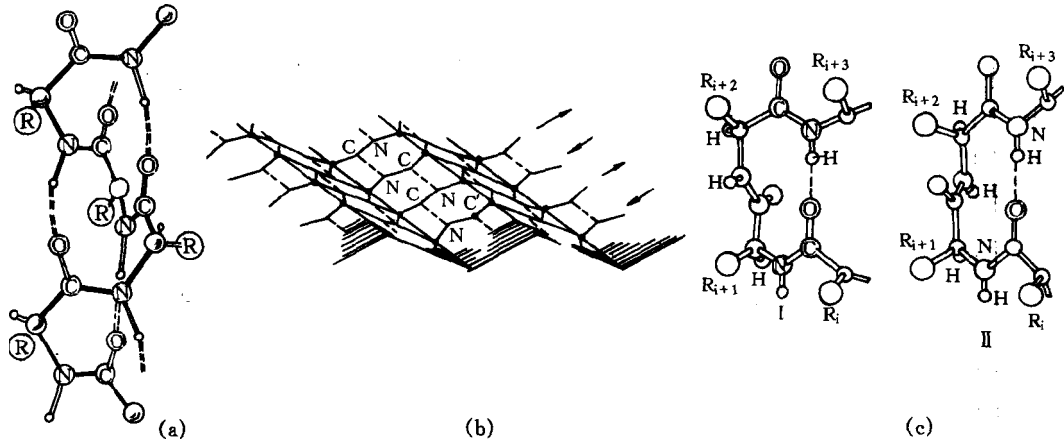


图 2-3 蛋白质的二级结构  
(a) $\alpha$ -螺旋;(b) $\beta$ -折叠;(c) $\beta$ -转角

完全伸展的多肽链侧向聚集在一起，相邻肽链主链上的  $-\text{NH}$  和  $\text{C}=\text{O}$  之间形成有规则的氢键，这样的多肽构象就是  $\beta$ -折叠片。 $\beta$ -折叠有平行和反平行两种，其中以反平行较稳定。 $\beta$ -转角是一种位于多肽链折叠处的结构。 $\beta$ -转角含 4 个氨基酸残基，它为第一个氨基酸残基的  $\text{C}=\text{O}$  与第一个残基的  $\text{NH}$  之间的氢键所稳定。 $\beta$ -转角有 I 型和 II 型两种，它们的主要区别在于第二肽链的空间取向。

在许多球状蛋白质中，我们常可观察到一些二级结构的组合形式，即超二级结构，最常见的超二级结构有  $\alpha\alpha$ 、 $\beta\alpha\beta$  和  $\beta\beta\beta$  三种图式（图 2-4）。

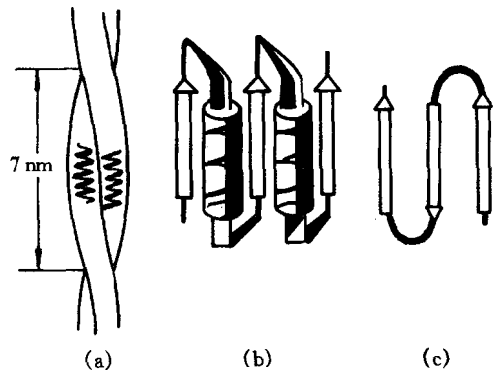


图 2-4 蛋白质的超二级结构  
(a) $\alpha\alpha$ ;(b) $\beta\alpha\beta$ ;(c) $\beta\beta\beta$

### 2.1.3 蛋白质的三级结构

形成二级结构之后，多肽链还可进一步折叠成三维球状结构，它们具有独立的结构和功能，称为三级结构。有些较大蛋白质的多肽链会折叠成相对独立的结构域的三维结构。结构

域是多肽链的一个部分，但已具有完整球状蛋白质的特征，从这一意义上来说它是自立的。它们可以通过有限的蛋白酶解从多肽链上切割下来而不改变其性质。有些蛋白质的结构域相似，有些则有明显差异。现在已经证明，结构域常对应于基因中各自的外显子（exon）。结构域的组合就形成了蛋白质的完整三级结构。

有些蛋白质仅含一条多肽链，为单体蛋白。有许多蛋白质含两个或两个以上的亚基，为寡聚蛋白如血红蛋白由 4 个亚基组成，天冬氨酸转氨甲酰酶由 12 个亚基组成等。在这些蛋白中，亚基的空间关系和缔和称为四级结构。寡聚蛋白普遍存在于机体中。

## 2.2 核酸的结构

人们经过长期的研究发现遗传特性由一种称为基因的因子决定，而基因则线性地分布于染色体上，染色体是由核酸和蛋白质组成的。直到 1944 年 Avery 等的肺炎球菌的转化实验才证明了遗传物质是 DNA 而不是蛋白质。1952 年 Hershey 和 Chase 的 T2 噬菌体对大肠杆菌的感染试验进一步证明了这一点。1953 年 Watson 和 Crick 提出了 DNA 双螺旋结构模型，它使许多遗传现象从分子水平上得到了充分和合理的解释，成为生物学发展中的一个里程碑。

### 2.2.1 核苷酸

核酸是一种线形多聚核苷酸，它的基本结构单位是核苷酸。核苷酸由三部分组成：戊糖、碱基和磷酸。在核苷酸分子中，戊糖和碱基缩合成糖苷酸键而形成核苷，核苷中的戊糖羟基被磷酸酯化，就形成核苷酸。核苷酸分为核糖核苷酸与脱氧核糖核苷酸两类，不同的核苷酸用碱基的第一个字来命名（表 2-2）。

表 2-2 核苷酸的组成和名称

核酸	核苷酸(缩写)	核 苷	碱 基
RNA	腺苷酸(A,AMP)	腺 苷	腺嘌呤
	鸟苷酸(G,GMP)	鸟 苷	鸟嘌呤
	胞苷酸(C,CMP)	胞 苷	胞嘧啶
	尿苷酸(U,UMP)	尿 苷	尿嘧啶
DNA	脱氧腺苷酸(A,dA,dAMP)	脱氧腺苷	腺嘌呤
	脱氧鸟苷酸(G,dG,dGMP)	脱氧鸟苷	鸟嘌呤
	脱氧胞苷酸(C,dC,dCMP)	脱氧胞苷	胞嘧啶
	脱氧胸苷酸(T,dT,dTMP)	脱氧胸苷	胸腺嘧啶

注：核苷酸现已成为包括核糖核苷酸和脱氧核糖核苷酸在内的类名，在核酸分子中，它们都以 A,G,T,C 表示；在游离状态下，脱氧核糖核苷酸则分别以 dA,dG,dT 和 dC 表示。

### 2.2.2 DNA 的结构

#### 2.2.2.1 DNA 的一级结构

DNA 的一级结构是由数量庞大的 4 种脱氧核糖核苷酸即脱氧腺嘌呤核苷酸、脱氧鸟嘌呤

呤核苷酸、脱氧胞嘧啶核苷酸和脱氧胸腺嘧啶核苷酸)，通过 3',5'-磷酸二酯键连接起来的直线形或环形多聚体。由于脱氧核糖中 C-2'上不含羟基,C-1 又与碱基相连接,唯一可以形成的键是 3',5'-磷酸二酯键 所以 DNA 没有侧链。图 2-5 表示 DNA 多核苷酸链的一个小片段。多核苷酸的碱基可简化为如图 2-5 的右方 也可缩写为 pGpCpTpA 和 pGCTA。按规定,多核苷酸链以连接于 5'-羟基的磷酸开始,以脱氧核糖的 3'-羟基终止。

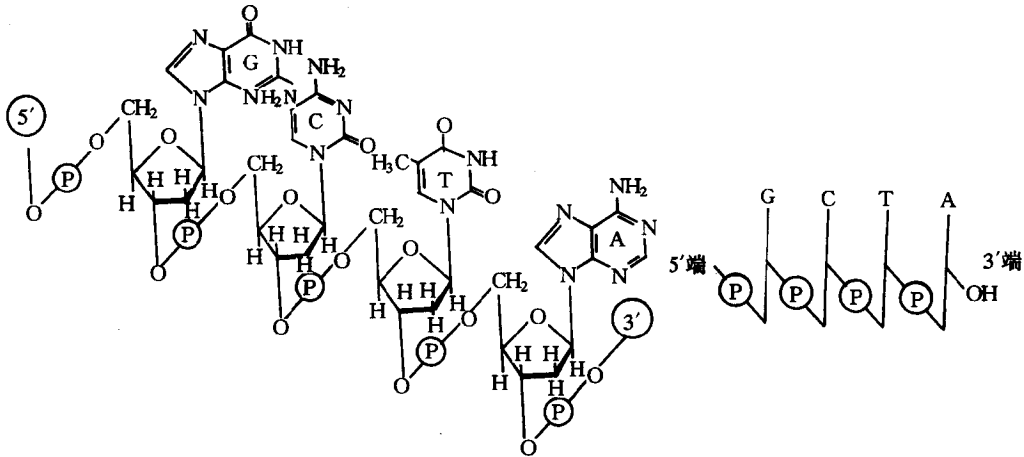


图 2-5 多核苷酸链片段

#### 2.2.2.2 DNA 的双螺旋结构

Watson 和 Crick 于 1953 年提出了 DNA 的双螺旋结构模型,后人的许多工作证明这个模型基本是正确的。其模型要点如下。

DNA 分子是由两条反向平行的多核苷酸链围绕同一个中心轴构成的双螺旋结构。多核苷酸链的方向取决于核酸间的磷酸二酯键的走向,习惯上以 3'→5'为正向 两条链都是右手螺旋。链之间的螺旋形成凹槽,一条较深,一条较浅。

②两条链的碱基层叠于螺旋内侧,碱基平面与纵轴相垂直,碱基之间堆积距离为 0.34 nm。磷酸基与脱氧核糖在外侧,彼此通过 3',5'-磷酸二酯键相接 形成 DNA 的骨架。糖环平面与纵轴平行。双螺旋的平均直径为 2 nm,两个核苷酸之间夹角为 36°。因此沿中心轴每旋转一周有 10 个核苷酸。每一转的高度 即螺距 为 3.4 nm 如图 2-6。

两条核苷酸链依靠碱基之间形成的氢键相联系而结合在一起。一条链上的嘌呤碱基必须与另一条链上的嘧啶碱基相匹配,A 与 T 相结合,其间形成两个氢键;G 与 C 相结合 其间形成 3 个氢键 因此 G 与 C 之间连接更为稳定一些。

DNA 双螺旋结构是稳定的,这里起作用的主要有 3 种力量。第一种是互补碱基对之间的氢键;第二种是由于芳香族碱基的  $\pi$  电子之间相互作用而引起的碱基堆积力;第三种使 DNA 分子稳定的力是磷酸残基上的负电荷与介质中的阳离子之间形成离子键,与 DNA 结合的离子如  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Mg}^{2+}$  和  $\text{Mn}^{2+}$  在细胞内大量存在。

以上介绍的是 B 型 DNA 的结构 此外还有 A 型和 C 型。这里不再一一介绍。

#### 2.2.2.3 DNA 的三级结构

双链 DNA 多数以线形存在,某些病毒、线粒体、叶绿体及某些细菌中的 DNA 为双链环

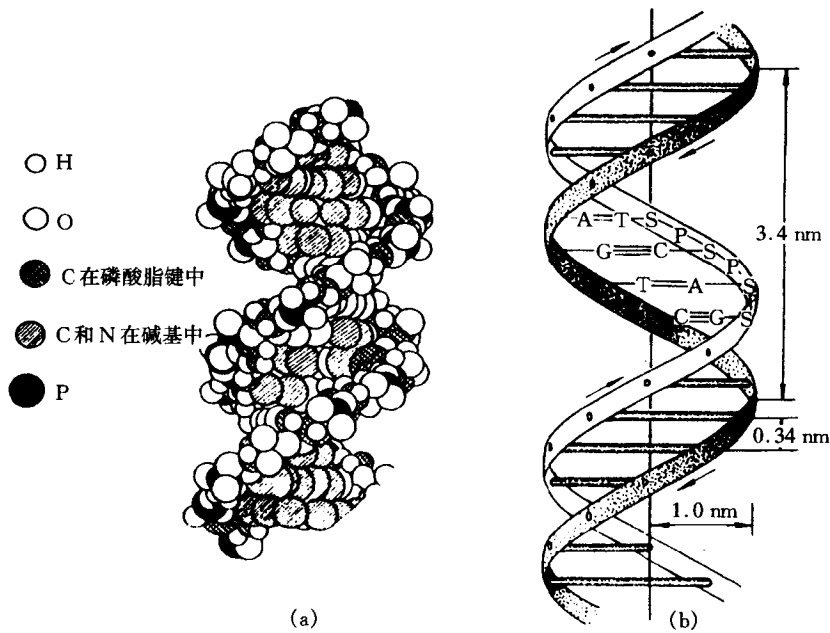


图 2-6 DNA 分子双螺旋结构模型及其图解  
(a)DNA 分子双螺旋结构模型 ;(b)模型的图解

形。在细胞内，这些环形 DNA 可以进一步扭曲折叠成“超螺旋”的三级结构(图 2-7)。

一些病毒的 DNA 和真核细胞染色质中的 DNA 是双螺旋线形分子。染色质 DNA 的结构复杂。双螺旋 DNA 先盘绕蛋白形成核粒(超螺旋)，许多核粒(或称核小体)由 DNA 链连在一起构成念珠状结构，念珠状结构进一步盘绕构成更复杂、更高层次的结构。据估计，人的

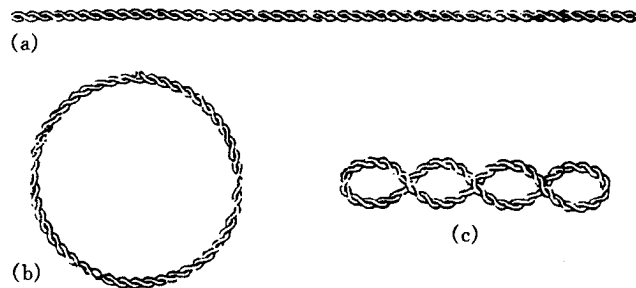


图 2-7 DNA 三级结构模式图

(a)直线型双螺旋结构;(b)开环型结构;(c)共价闭环超螺旋型结构  
DNA 大分子在染色质中反复折叠盘绕，其压缩约 8 000~10 000 倍。

### 2.2.3 RNA 的结构

RNA 也是无分支的线形多聚体核糖核苷酸，所含的 4 种基本碱基是腺嘌呤、鸟嘌呤、胞嘧啶和尿嘧啶。RNA 的碱基组成不像 DNA 那样具有严格的 A=T,G=C 的规律。RNA 的结构只有在局部构成双螺旋结构。动物、植物和微生物细胞内都含有 3 种主要 RNA 即核糖体 RNA(rRNA)、转运 RNA(tRNA)和信使 RNA(mRNA)。此外真核细胞中还有少量核内小 RNA(snRNA)。

#### 2.2.3.1 tRNA

tRNA 是传递氨基酸的 RNA。所有的 tRNA 不论其来自动物、植物还是微生物都具有

许多结构上的共同点。

相对分子质量在 25 左右 由 70~90 个核苷酸组成，沉降系数在 4 S 左右。

碱基组成中有较多的稀有碱基。

③ 3'末端都为... CpCpAOH，用来接受活化的氨基酸，此末端称为接受末端。

④ 5'末端大多为 pG... 也有 pC... 的。

⑤ tRNA 的二级结构都呈三叶草形 tRNA 二级结构如图 2-8。三叶草结构由氨基酸臂、二氢尿嘧啶环、反密码环、额外环和 TψC 环等 5 部分组成。

⑥ tRNA 的三级结构的形状像一个倒写的 L 字母(图 2-9)。

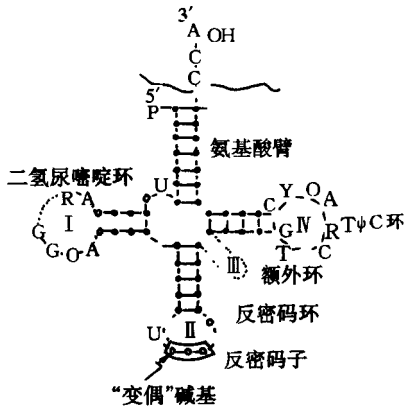


图 2-8 tRNA 三叶草形二级结构通式

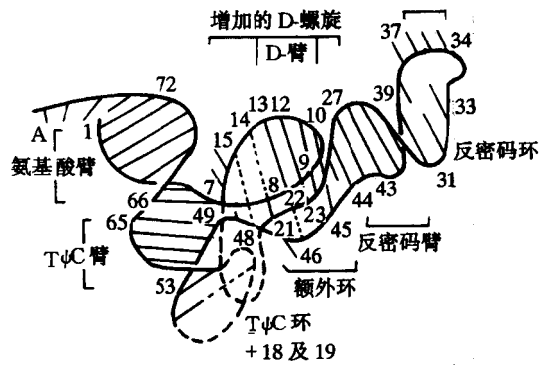
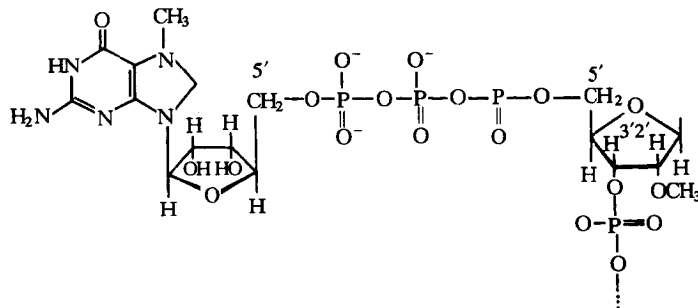


图 2-9 酵母苯丙氨酸 tRNA 的三级结构

### 2.2.3.2 mRNA

mRNA 在 DNA 和蛋白质之间起媒介的作用。即 mRNA 的功能在于把 DNA 模板链上的碱基序列转录为 RNA 分子上的碱基序列(mRNA) 再从 mRNA 上的碱基序列通过合成蛋白质的机构获得氨基酸的序列。绝大多数真核细胞 mRNA 在 3'末端有一段长约 200 个残基转录后添加上去的多聚腺苷酸。原核生物的 mRNA 则一般无多聚腺苷酸。多聚腺苷酸的结构与 mRNA 从核转移至胞质的过程有关，也与 mRNA 的半衰期有关。新合成的 mRNA 的多聚腺苷酸较长 衰老的 mRNA 分子的多聚腺苷酸较短。真核细胞 5'末端为一特殊结构(又称帽状结构)：



5'末端的鸟嘌呤 N<sup>7</sup> 被甲基化，鸟嘌呤核苷酸与相邻的一个核苷酸相连，形成 5',5'-磷酸二酯键 这种特殊的 5'末端有抗核酸酶水解的作用。一些病毒 RNA 也有类似的 5'末端。这

种特殊结构可能与蛋白质生物合成过程中翻译的起始有关。

### 2.2.3.3 rRNA

rRNA 即核糖体 RNA 和具有各种功能的大量蛋白质组成合成蛋白质的场所——核糖体。核糖体 RNA 共分 3 种 原核生物的核糖体由 5 S rRNA, 16 S rRNA 和 23 S rRNA 所组成 分别含有 120, 1 541 和 2 904 个核苷酸。rRNA 是稳定的分子, 其前体比成熟的 rRNA 分子大。

## 2.3 蛋白质和核酸的性质

蛋白质和核酸都是相对分子质量很大的生物大分子, 它们有许多相似的理化性质, 如都可以变性等。但由于组成的基本单元不同, 它们又存在着许多区别。

### 2.3.1 蛋白质的性质

蛋白质的相对分子质量很大 在溶液中不稳定 因此通常用渗透压法、超离心法、凝胶过滤法、聚丙烯酰胺凝胶电泳法来测定其相对分子质量。

蛋白质是两性电解质, 既能与酸作用也能与碱作用。其解离基团除肽链末端的  $\alpha$ -氨基和  $\alpha$ -羧基外, 主要是肽链上氨基酸残基侧基上的氨基、羟基、咪唑基、胍基、酚基和巯基, 在一定 pH 下这些基团解离成带电基团而使蛋白质带电。各种蛋白质有特定的等电点。在等电点蛋白质以两性离子存在, 总电荷为零, 此时最不稳定, 易聚集而沉淀析出。在一定的 pH 值溶液中, 各种蛋白质所带电荷不同, 在电场中移动的方向和速度也不同。据此可用电泳法把蛋白质混合液中各种蛋白质分开。

蛋白质的相对分子质量较大也决定了它是一种亲水胶体, 在蛋白质外可以形成水化膜, 相同颗粒的同电荷性使之不会聚集而沉淀。因此可以用渗析法来分离纯化蛋白质。它也如同一般胶体一样, 当加入适当的试剂如盐类、有机溶剂、重金属盐及某些酸类时, 可以使它产生沉淀。

天然蛋白质因受物理因素或化学因素的影响, 其分子内部原有的高度规律结构会发生变化, 致使蛋白质的理化性质和生物学性质都有所改变, 但并不破坏一级结构, 此种现象称为变性作用。能使蛋白质变性的因素很多 化学因素有强酸、强碱、尿素、胍、去污剂、苦味酸、浓乙醇等。物理因素有加热 (70~100℃)、激烈搅拌、射线、超声波等。不同蛋白质对这些因素的敏感程度也不同。蛋白质变性首先表现为失去生物活性, 如酶失去催化能力; 同时理化性质也改变了, 如溶解度降低等。变性蛋白质的一级结构没有被破坏, 破坏的只是二、三级结构, 因此蛋白质的组成和相对分子质量是不变的。在不太激烈的条件下, 变性是可逆的; 但条件激烈就成为不可逆反应。

### 2.3.2 核酸的性质

核酸的性质与其组成和结构密切相关。核酸的组成有嘌呤碱、嘧啶碱、磷酸、核糖和脱氧核糖。核酸结构特点为分子巨大, 有共轭双键、氢键、苷键和磷酸二酯键, 有烯醇式羟基、自由氨基和磷酸基, 这些结构是核酸特性的基础。

#### 2.3.2.1 DNA 变性

当温度升高时, 核酸和蛋白质的三维结构都会受到破坏。使核酸有规律的螺旋型双链结

构变成单链无规律的“线团”。从天然状态分子转变到分子变性状态分子，这一过程称为变性。将双链 DNA 或天然 DNA 加热时 两条链间的结合力受到破坏 两条链分离 因此变性 DNA 是单链的。在 pH 大于 11.3 时，可以使 DNA 对碱水解很稳定，所以这是基因工程中对 DNA 变性常选用的一种方法。

虽然 DNA 比较稳定，但这是一种动态结构，它的双链区经常被打开变成单链的泡状。双链 DNA 部分的链区被打开的现象称为呼吸。呼吸现象可能是使特定的蛋白质能与 DNA 分子发生相互作用和“阅读”它的编码的信息。由于 GC 碱基对有三对氢键，AT 碱基对仅有两对氢键 所以在 DNA 分子的富含 AT 对的区域比富含 GC 对的区域更容易发生呼吸。

### 2.3.2.2 DNA 复性与杂交

DNA 变性后，双螺旋两条链分开，如果溶液迅速冷却，两条单链继续保持分开，如图 2-10。但如果缓慢冷却，则两条链可能发生特异的重新组合而恢复成双螺旋，这一变性 DNA 恢复其原有结构和性质叫做复性或退火。重新形成的 DNA 称为复性 DNA。复性是基因工程中很有意义的特性。这可以用于显示不同生物间的遗传相关性，检测特定的 DNA，考察某些序列在特定生物体 DNA 中是否出现一次以上，确定特定碱基序列在 DNA 的位置等。

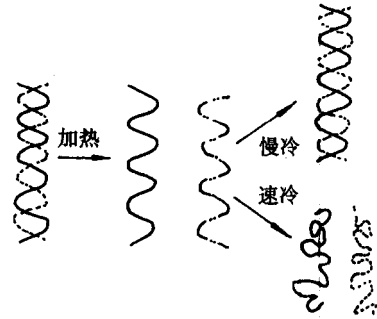


图 2-10 DNA 溶液变性及其复性

发生复性必须有两个条件：①盐浓度必须达到足以消除两条链中的磷酸基团的静电斥力，通常用 0.15 ~ 0.5 mol/L NaCl。②温度必须高到足以破坏其随机的链内氢键，但温度又不能太高，否则不能形成和维持稳定的配对。

不同来源的 DNA 形成复性 DNA 分子时，此复性称为杂交。杂交技术是研究核酸功能的重要手段。硝基纤维素滤膜与单链 DNA 结合得非常牢固，但它不会与双链 DNA 或 RNA 结合，这为测定杂交提供了重要的技术。其最重要的用途是检测 DNA 单链和 RNA 分子间的序列同源性 这可称为 DNA-RNA 杂交。这是检测从特定 DNA 分子上拷贝下来的 RNA 分子的选择方法。具体是将结合着单链 DNA 的滤膜放入含放射性的 RNA 溶液中 复性后洗涤滤膜，根据滤膜上存在的放射性 RNA 来检测杂交。RNA 变性及其复性与 DNA 变性及其复性的变化类似 但其变化程度不如 DNA 大 因为 RNA 中只有部分螺旋区。同蛋白质变性一样，核酸变性不涉及核苷酸间共价键的断裂，因而并不引起相对分子质量的降低。

## 2.4 大分子间的信息传导

1954 年 Crick 提出了遗传信息传递的规律，即中心法则：



DNA 是合成 RNA 的模板，RNA 又是合成蛋白质的模板，DNA 通过自我复制而保持遗传信息的连续性。后来还发现在逆转录酶的存在下，某些病毒可以 RNA 为模板合成单链 DNA