

《生物工程讲座》(II)

——基因工程的理论及应用(一)

基因的结构与分离

崔桂芳

(中国科学院上海生物化学研究所)

自从 40 年代证实生物的遗传物质就是核酸以来,对于核酸如何实现遗传功能的研究,一直广泛地吸引着科学家浓厚的兴趣。50 年代提出的“一个基因,一个多肽”的概念,由基因型和表型统一的角度明确了基因的定义,将基因视为携带一定生命功能信息的单元。全套基因活动的总合,可以导致生物个体由出生至死亡一系列的生物化学反应,表现出生命的种种活动形式。而由父系和母系结合的遗传信息,便将生命的特征传给下一代。细胞的分裂、分化,生物的生长、发育,和性状遗传等生命基本过程,都是基因活动的结果。要了解生命规律,并利用这些规律为人类服务,基因的研究则具有特别重要的意义。需要分离纯化特定的基因,以分析其结构,探讨其功能。

以下先介绍基因的基本结构,然后再介绍如何利用基因工程技术来分离和纯化特定的基因。

一、基因的结构

复杂的真核生物和较简单的原核生物,它们的细胞染色体中的核酸都是 DNA。真核生物细胞中,除染色体外,在线粒体和叶绿体中也有遗传物质 DNA。它们只占总 DNA 的 0.1%。原核生物在染色体外,常有分子量较小的 DNA,如质粒。遗传物质是 DNA 的,还包括许多动物病毒和噬菌体。少数植物病毒的基因组核酸成分也是 DNA。

以 RNA 作为遗传物质的,只见于病毒,包

括动物病毒、类病毒和一些噬菌体。

从功能上分析,基因可分为两大类:一类叫作结构基因,另一类叫作调控基因。它们在结构上各有特点。

(一) 结构基因的结构 结构基因是指 DNA 中能够转录产生信使 RNA 的部分。其中能再通过翻译产生多肽链的部分称为编码区。不论真核或原核生物,它们的结构基因的编码区,都是通过密码子来表达的。所谓密码子,就是一组连续的三个核糖核苷酸。在翻译过程中,每种密码子可以对应于一种氨基酸。因此,在结构基因中一段固定排列的核苷酸,可以最终产生一条相应的多肽链。

转录和翻译都有一定的起始点和终止点。在基因组上由特殊的核苷酸顺序所决定。翻译的起始点和终止点都很明确。起始信号是密码子 AUG (基因上相应为 ATG),终止信号是密码子 UAA, UAG 和 UGA (基因上相应为 TAA, TAG 和 TGA)。绝大多数氨基酸是用不止一种密码子来表达的。实际上,不同的生物和细胞器,在多种密码子均可代表一种氨基酸的情况下,往往侧重使用其中一些密码子来表达某一特定的氨基酸。

在基因转录生成的信使 RNA 上,除了编码区,在其 5' 端还有一个核糖体结合位点(在原核位于起始密码子前方-3 至-11 个核苷酸处,称为 SD 顺序),它有利于翻译的进行。真核信使 RNA 的 5' 端,还需加上个 7-甲基鸟苷三磷酸的“帽子”,翻译才能开始。

真核生物的结构基因有一个显著的特点,即它们的编码区大多数是不连续的。由于一些作用尚不清楚的核苷酸区段的插入,它们被分隔开。这种区段叫做插入顺序,或内含子,编码区称为外显子。插入顺序随着编码区一同转录,生成信使 RNA 前体,但在转录后的加工过程中,即被除去。血红蛋白的基因有两个插入顺序。胶原蛋白的基因有甚至超过 40 个之多的插入顺序。令人惊异的是,插入顺序的总长度往往超过编码区的总长度。如肌红蛋白基因的插入顺序的长度竟比编码区大出 20 倍。卵白蛋白基因中的插入顺序占结构基因的 85%。但是,组蛋白基因内却无插入顺序,这是真核基因中少见的例外情况。显然,插入顺序的存在必然有着至今还不甚了解的特殊意义。这是很有趣味的研究课题。

由真核信使 RNA (mRNA) 反转录所得的互补性 DNA (cDNA), 没有插入顺序。它们在以重组 DNA 形式转化至原核细胞中后,可以避免转录后不能加工去除这类顺序的不利情况,而能够翻译出蛋白质。但在真核细胞中,若没有这种插入顺序,就没有成熟的 mRNA 送入细胞浆,也就不能产生蛋白质。

值得一提的是,真核细胞中还有一些基因是多次重复的,可以重复百次、千次。这类基因多是不编码蛋白质的。其中有的转录生成 rRNA, 5SRNA 和 tRNA。编码蛋白质的重复基因只有组蛋白基因。此外,有些单拷贝基因,在特殊情况下,会通过基因放大的形式,而成为多拷贝的。如果蝇卵泡外壳蛋白和二氢叶酸还原酶 (DHFR) 等的基因,在大量需要其蛋白质时,就以分支复制或滚环复制的方式,将基因的拷贝数增加数十倍。此外,基因组里还有很多连转录也不转录的重复顺序,它们的作用还不甚清楚,还不能明确称为基因。

(二) 调控基因的结构 调控基因是在结构基因之外,决定基因是否转录的功能单位。转录由 RNA 聚合酶促成。这种酶能识别调控基因上的特殊位点,并在一定部位与之结合而开始转录。原核上的结合区叫启动子,它位于

转录起始点上游 10 个核苷酸左右处。其中 6 个核苷酸的顺序相当保守,为 TATAAT (称为 Pribnow 盒)。特点是 TA 多,这可能因为它有利于解链。在 -35 个核苷酸左右处,顺序也比较保守,为 TTGACA,可能是 RNA 聚合酶的识别位点。真核生物有三种 RNA 聚合酶,分别可转录出 rRNA, mRNA, 和小分子的 5SRNA 及 tRNA。转录时,所识别的起始点不同。使转录生成的 mRNA 的聚合酶 II, 起始信号区似原核,也富有 TA (称为 TATA 盒或 Hogness 盒),但距离转录开始点 5' 端较远,在 -28 核苷酸左右处。使转录生成 5SRNA 的聚合酶 III 则很特殊,它的结合位点不在转录开始点的 5' 端,而是在转录区之内。

在调控基因元件中,有一种功能特殊的核苷酸顺序,叫作“强化子”,它对结构基因的转录有增强的效应。免疫球蛋白,胰岛素和糜蛋白酶的基因,以及病毒中都发现有它的存在。这是一段短的 DNA 顺序,有保守的核心部分。它们不论在启动子前方或后方,也不论与基因是同向或反向排列,都有提高转录效率的作用。

原核转录物,在终止区 GC 多,它们可自成

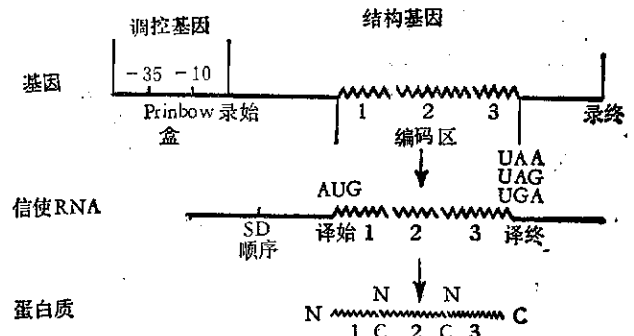


图1 原核基因结构示意图

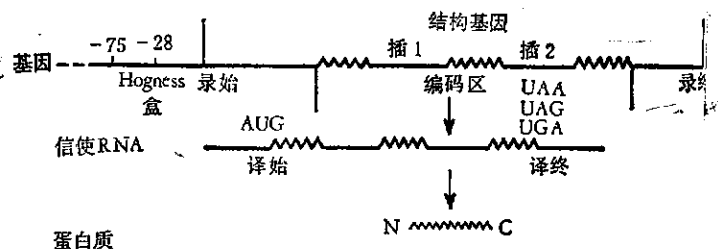


图2 真核基因结构示意图

互补链而形成发夹结构,后面还有多个U。这种结构的特点是可以阻止酶的继续转录(见图1)。真核结构(见图2)。

二、基因的分离

分离纯化某特定基因,这在DNA重组技术出现以前是很难想象的事。困难之一是,核酸是一种庞大的生物大分子。如大肠杆菌,其染色体DNA的分子量即有 10^9 道尔顿左右。估计带有约2000个基因。由其中分出某个单拷贝基因当然并非易事。人DNA的总分子量则有 10^{12} 道尔顿之多。很明显,其中某一特定的单拷贝基因在总DNA中所占的比例,更是非常之小,约为百万分之一。然而,更难的是,细胞里所有的基因都是DNA的片段,它们的物理、化学性质非常接近。因此,要借一般的物理、化学方法将其分离开,几乎是不可能的事。这就是基因的分离纯化研究迟迟未能打开困境的原因。

70年代初期,基因工程方法的建立,是个爆炸性的飞跃,它完全扭转了基因研究的处境。这个技术的关键就是,可以建立特定基因片段的单克隆,从而通过大量复制,能够十分方便地拿到一定数量的纯基因。本文介绍三种利用基因工程技术获取单一基因的方法。

(一) 由基因库筛选目标基因 这是一种由基因组分离基因所使用的方法。基因组都是线性大分子,某特定基因在其中只占极小的部分。因此,可以先将此大分子分解成许多较小的片段,再由其中找到带有目标基因的片段。所谓基因库就是:利用DNA重组技术,将这些片段和作为载体的噬菌体DNA或质粒DNA连接,再在大肠杆菌中复制。由于每个载体各携带一个片段复制后产生的每一个噬斑或菌株就只含有一种重组载体,称为一个克隆。由全部基因片段所产生的单个克隆的总合,就叫作基因库。通过适当的筛选办法,可以从其中挑选出带有目标基因的单一克隆,后者便是提供纯基因片段的较好来源。

这个方法的好处,一是可以利用单个克隆

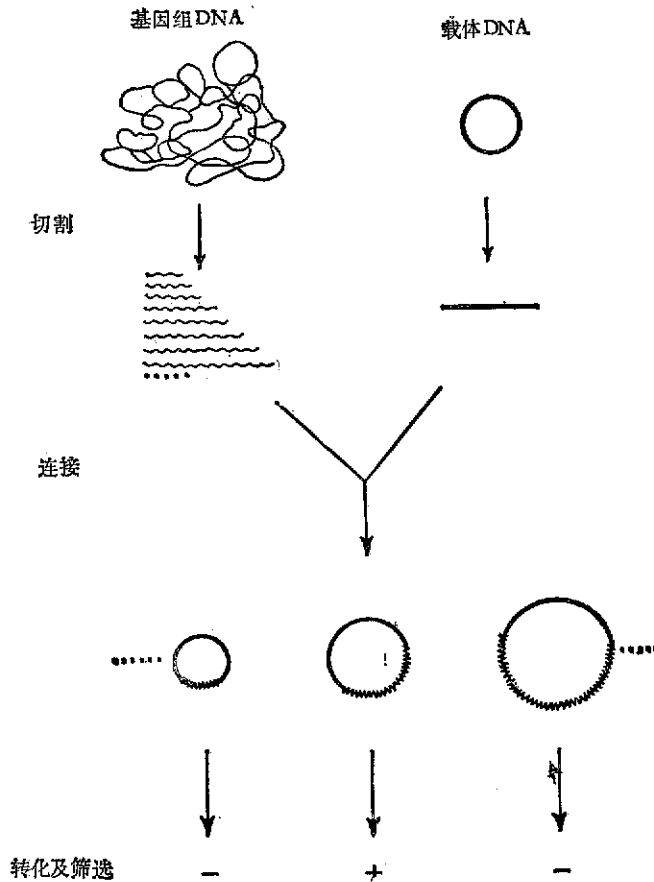


图3 基因库制备示意图

能将不同片段分开的特点,方便地进行筛选,从而收集大量足供研究的材料。二是筛出的基因片段,就是基因在染色体基因组中的结构。

方法的全过程包括:基因组DNA的切割,运载体DNA的切割,外源DNA片段与运载体DNA的连接,重组DNA的包装,重组体的转染或转化,和克隆的筛选等(见图3)。

1. 基因组DNA的切割 这个程序是由限制性核酸内切酶来完成的。这种酶的切点都有很强的专一性。它们各有特定的识别顺序,通常为4至6个核苷酸。因此,对某一特定的基因组DNA来说,经某一特定酶作用后,所得到的片段不但数目固定,每个片段的长短也是固定的。例如,在制备大肠杆菌或噬菌体λ的基因库时,基因组DNA经酶完全水解后,情况就是这样。

但在制备真核基因库时,多采取部分酶解,也即是不完全酶解的方法。因为,从 DNA 中四种碱基随机分布的几率计算,能识别 6 个核苷酸的酶,所切片段的长度为 4Kb,而能识别 4 个的则为 256 Bp。如果 DNA 链中的每个识别顺序都被酶切,基因的完整性很可能受到损伤。因此,为了提高基因保留完整的可能性,就采取部分酶解的办法。可以推知,用要求 4 个核苷酸顺序的酶,如 Mbol, AluI, HaeIII, 进行部分酶解,比用要求 6 个核苷酸顺序的酶,由于酶切位点的分布更广泛,所产生片段的类别就更多。

2. 制备基因库的载体 常用的载体是 λ 噬菌体和野生型质粒的衍生物,是经人工特别改造和组建的 λ 或质粒。它们带有一些限制酶的独特识别位点。在此种切口处,可将带有同种酶切口的外源基因片段插入。经连接扩增后,还可用原酶切出。

λ 噬菌体适用于插入较长的片段,一般可插入 15—20Kb。如 Charon 4A,只保留 λ 原来两个 EcoRI 酶切点,可在其间置入经 EcoRI 酶切割的外源基因组 DNA 片段。质粒载体多用于携带较小的 DNA 片段。它们带有多种酶的单切口,如在 pBR 322 上, EcoRI, PstI, BamHI 和 HindIII 等酶都只有一个切口。此外,质粒上还带有抗药性基因,有利于以后重组质粒的筛选。又如 PUC9,是一种以乳糖操纵子作为筛选标志的质粒,在操纵子内部插入一段人工合

成的寡聚核苷酸,其中包含着多种酶的单切口。

还有一种兼有质粒抗药性标志和 λ 噬菌体 VNA 粘性末端的载体,叫做粘性质粒 (Cosmid)。它的特点是可以插入长达 45Kb 左右的外源 DNA 片段。这对克隆具有很大基因组的真核基因片段非常有利。这种载体以噬菌体的形式进入寄主菌,再以质粒的形式进行复制(见图 4)。

3. 外源 DNA 和载体 DNA 的连接 这种连接靠连接酶完成。制备噬菌体基因库时,为了减少无效连接,提高阳性筛选率,可先将酶切所得片段混合物用电泳,或作蔗糖密度梯度超离心,按大小粗粗分开,只选用其中长度适宜的部分,与载体连接。如 Charon 4A 或 Cosmid,只有分别装入 15—20Kb 或 35—45Kb 的片段时,才能重新与外壳蛋白组装成具有转染能力的噬菌体。过大或过小的片段,都不仅不能重新组装,而且消耗有用的载体。因此,预先将之除去是完全必要和有利的。制备质粒基因库时,较小片段易于插入。此时如欲取得较大的片段,可将小片段预先除去。

4. λ 噬菌体的重新包装 用 λ 重组 DNA 作库时,将外源 DNA 片段接入 λ DNA 载体之后,必须将重组 DNA 用 λ 噬菌体外壳蛋白包装起来,形成完整的噬菌体颗粒。感染大肠杆菌,可将重组的 λ DNA 引入寄主菌,建成一个 λ 基因库。 λ 噬菌体的蛋白外壳有很多组

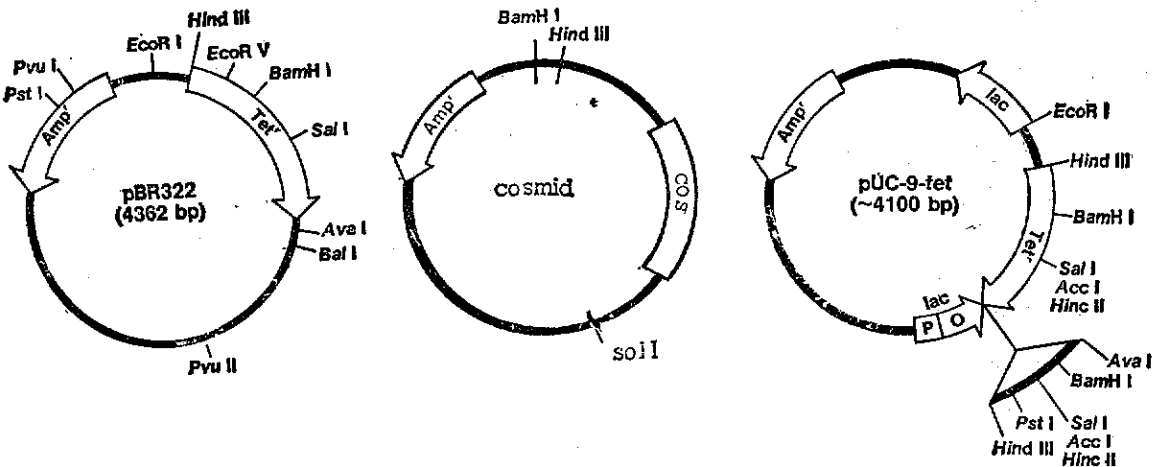


图 4 几种常用的制备载体

份,分别组成头部和尾部。头部装入 λ DNA 后,再加上组装好的尾部,便成为有感染活性的完整的噬菌体。

在重组 λ DNA 的包装中,所用的蛋白质提取液来自两种大肠杆菌变异株。两种提取液单独使用,都不能完成包装任务。将二者一同使用,则必要的包装蛋白组份齐备,就可将重组的 λ DNA 包装起来。

在正式包装前,需用标准 λ DNA 试包,以确定蛋白提取液的效果。每微克 λ DNA 应可产生 10^7-10^8 个噬斑。蛋白提取液分别制备好以后,需小量分批低温冰冻保存。使用时放冻中小心融解,立即加入重组 DNA。

5. 基因库的质量 基因库的质量是指它所包含的遗传信息量而言。一个库达到什么指标才算合格呢? 这与对象染色体 DNA 的大小,可能存在多少基因,和基因的重复程度或拷贝数等有关。通常以能产生多少噬斑或克隆菌株数来表示。

如利用 pBR322 建立一个大肠杆菌的基因库,因大肠杆菌的基因组为 3×10^6 Bp,若酶切片段长度平均为 3Kb,则至少需有 10^3 个克隆株,才能得到一个比较完整的基因库。这样的库,仅仅保证大肠杆菌染色体 DNA 都分段被克隆,而不能说所有基因都是完整的。所以,有时需要作另一种酶切片段的基因库,以拿到完整的特定基因。

2 噬菌体载体主要用于高等生物的动、植物细胞的基因库。为了尽量保证染色体 DNA 全部得到克隆,而且基因保留完整,总采用部分酶解如前所述。所以同一区段就可能在很多不同种类的片段中重复出现。因此,需要有大量克隆才足以构成一个完全的库。

关于供筛选所需最低重组子数,可根据几率计算公式推算如下:

$$N = \frac{\ln(1-P)}{\ln(1-f)}$$

其中 N 代表尽可能保证筛选到阳性片段所需重组子数(以噬斑数表示)

P 为要求存在的几率(以百分数表示)

f 等于单一重组子中,目标基因片段在基因组中所占的比例(按 Bp 计)

例如,人的基因组为 3×10^9 Bp,酶切片段长度为 17Kb,而要求的存在几率为 99% 时,则

$$N = \frac{\ln(1-0.99)}{\ln\left(1 - \frac{1.7 \times 10^4}{3 \times 10^9}\right)} \\ = 8.1 \times 10^5$$

也即是说,这个库所含重组子的种类不得低于 8.1×10^5 个。

6. 重组子的筛选 制备好基因库,仅提供了能获取某特定基因的材料。关键问题是如何由其中选出带有所要求基因的 DNA 片段。由噬斑筛选目标基因片段,常用探针杂交法。由质粒筛选时,除探针杂交法,还可根据载体的抗药性标志,显色标志,和营养型标志等进行筛选。

探针是一个 DNA (或 RNA) 片段,它的核苷酸顺序与所要筛的基因相同或相近,因此可以与目标基因形成互补杂交链。探针可来自 mRNA 反转录的 cDNA,人工合成的寡聚核苷酸,纯化病毒的 DNA 或 RNA,或其它生物体的近似基因。它还必须带有能够很方便检出的标志,如放射性同位素标志 (^{32}P 或 ^{125}I)。这样,一旦与目标基因形成杂交链,就很容易找到它们的位置。所以探针必须很纯,很专一。

利用探针筛选的过程是:将基因库在平盘上扩增,将扩增的噬斑或经过排列的菌株全部复印到硝酸纤维滤膜上,使它们所含的 DNA 通过碱的作用释放出来并变性,烘干固定后,再与探针进行杂交。经过保温与洗涤后,探针上的放射性标志保留在目标基因片段上。利用放射自显影技术,间接在培养平盘上定出相应的阳性噬斑或菌株的位置。选出的阳性噬斑,需经二筛或三筛皆为阳性,并且在不同稀释度下,噬斑数与阳性点数完全一致时,方可确认其为独立的真正阳性点,而不是混合的或假的阳性点。阳性克隆扩增后,可将目标基因由载体用原来的酶切出。通过凝胶电泳进行分析,或由凝胶中回收,供进一步研究使用。

探针也可不用放射性同位素标记。如有一种探针是利用酶标显色,来代替同位素示踪。这种方法没有同位素废物处理问题,也没有同位素化合物的保存期受半衰期限制的问题,所以是颇有前途的一种新方法。但是目前在灵敏度上,似还不如放射性标记探针。

显色的方法是,使作为探针用的 DNA 或 RNA 片段,在碱基 T 或 U 上与生物素结合。使碱性磷酸酶或过氧化物酶与核酸链上的抗生物素结合。这两种结合物可以连在一起成为一个复合物。它遇到适当的酶的底物,即可产生颜色反应。当探针与目标基因在滤膜上杂交后,在目标基因的位置上就显出了颜色。

质粒载体在组建时,一般有意地保留了二种抗药性基因。在这种基因上往往还存在有质粒的单一酶切位点。这样,在此酶切口处插入外源 DNA 片段时,由于抗药基因被破坏,质粒就失去原有的抗药性。转化后,在含有另一种抗生素的培养盘上,可以筛去许多未被转化的细菌。再进一步与对照抗菌素盘比较,可以区分只含空载质粒的菌落,而选出带有插入片段的克隆。对这些克隆,再进行探针杂交法筛选,即可得到有目标基因的克隆株(见图 5)。

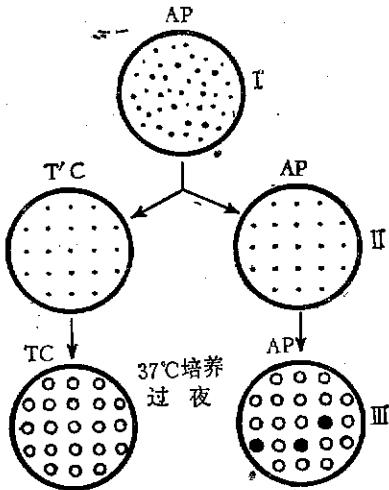


图 5 抗药性筛选法 AP = AP' 系氨基青霉素;

T_c = T_c' 系四环素; I——外源片段插入 T_c 基因、转化后铺在 AP 盘上; II——将单个菌落排出,按相同顺序分别点在 T_c' 及 AP₂ 盘上; III——○ 为含空载质粒的菌株, ● 为有插入片段的菌株。

对于携带有大肠杆菌乳糖操纵子的载体,可用 lac⁻ 寄主菌筛选。在培养盘上,加入半乳糖苷酶的底物, X-gal (5-溴-4-氯-3-吡啶-β-D-半乳糖苷)。若是空载的载体,即形成蓝色克隆株或噬斑。若此段基因被外源 DNA 隔断,则只产生白色克隆株或噬斑。因此,可以根据颜色找出带有重组子的载体。IPTG (异丙基-β-D-硫代半乳糖苷) 是半乳糖苷酶的诱导物,与 X-gal 同时加入,可大大提高鉴别率。

利用适当的营养缺陷型寄主菌,亦可提供良好的筛选条件。称为互补筛选。方法是将被基因库转化的寄主菌,在不完全培养基上铺盘,则由能生长的菌株中,可筛到带有营养缺陷型所缺基因的克隆。

(二) 由 cDNA 库获取目标基因

cDNA 是以 mRNA 为模板,经反转录而制成的 DNA 链。将它与质粒载体连接,转化后得到克隆。由某一特定 mRNA 制备物所得到的 cDNA 克隆的总和,就叫做 cDNA 库。cDNA 库法也是常用的获取特定基因的途径。

在制备 cDNA 库的过程中,由于在细胞中占很大比例但并不转录的 DNA 片段,和大量转录而产物并非 mRNA 的 rRNA, tRNA, 以及 5s RNA, 几乎可以完全除去,所以,供筛选的对象就相对集中了许多。加之,在有些特殊组织中,或某一组织在一定特殊时期,其中的某种特定 mRNA 可在总 mRNA 中占优势。这就又大大提高其 cDNA 的阳性筛选率。如网质红细胞是制造珠蛋白的细胞,其珠蛋白 mRNA 的含量可高达 90%。用它作起始原料制成的 cDNA 库,自然容易得到珠蛋白的 cDNA。又如,由肝脏 cDNA 库筛选白蛋白 cDNA 是有利的。脑组织中 mRNA 的种类特别多,由其 cDNA 库可筛取一些脑所特有的 cDNA。

cDNA 库的制备,可分为以下几个步骤:

1. mRNA 的抽提: mRNA 很容易遭到破坏,所以在抽提及保存 mRNA 阶段,要采取特别措施,以保证 mRNA 的完整性。整个操作过程,要在低温进行,并要绝对防止内源及外源

核糖核酸酶的破坏作用。为此,可将器皿预先用药物或高温处理,避免皮肤的直接或间接接触,及迅速加酚等。生物组织一旦由生物体分出,需立即丢入干冰中,然后在干冰中捣碎。再加酚及匀浆溶液匀浆,以抽出其中的核酸。由于 mRNA 一般具有多聚 A 的 3' 端,故可以利用寡聚 dT 纤维素柱,将其与其它种核酸成份分开。一次柱分离,往往不能充分去掉 rRNA 等杂质。因此,最好能反复上柱 2—3 次。产物在酒精中于 -70°C 保存,使用时再离心溶解。

2. cDNA 的制备: 由 mRNA 制备 cDNA, 除需要反转录酶和四种脱氧核苷酸底物外, 还需要有一个引物。这引物即是能与 mRNA 尾端的多聚 A 成互补的寡聚 dT。 它可促进反转录的开始。这样就合成出一个杂交双链。将此杂交双链用碱处理, RNA 链被破坏。留下的是与其互补的 cDNA 单链。再以此单链为模板, 用反转录酶或 DNA 聚合酶 I 合成出一个双链的 cDNA。合成完毕, 用 S_1 核酸酶将两端的单链部分去掉, 就得到一个完整的双链 cDNA 了。

3. cDNA 的克隆: 双链 cDNA 可通过 DNA 重组技术与载体连接。 然后对易感受态寄主细菌进行转化。 在每个被转化的细菌中, 只带有一种质粒。不同的重组质粒携带不同的外源 cDNA 片段。将转化菌铺盘培养后, 可得到个别分离的菌落, 称为单克隆。由其中可以得到目标基因的 cDNA 片段。

4. 目标片段的筛选: 可先利用抗药性标志, 或 lac 标志进行粗筛。然后用探针杂交法筛。原理和操作步骤同上节, 不再详述。

必须指出, 由 cDNA 库筛出的片段, 对高等生物来说, 并不反映真正基因的顺序。只有 RNA 病毒例外。cDNA 基本上只反映由基因产生的成熟 mRNA 的顺序, 主要是编码顺序部分。真核基因中非转录区的调控顺序和转录区的插入顺序就完全无从了解。不过, 这样得到的纯 cDNA 片段, 可以作为极好的探针, 再从基因库中筛选天然的基因。此外, 这种片段

在原核细胞中表达时, 不受插入顺序的妨碍。这对生产真核生物的基因工程产物, 反而是求之不得的。

(三) 人工合成基因 人工合成基因也是取得目标基因的重要手段。这种技术是逐步发展起来的。Khorana 自 50 年代起, 开始用磷酸二酯法合成 DNA 片段。1972 年完成了酵母丙氨酸 tRNA 结构基因的合成。1979 年完成了大肠杆菌酪氨酸校正基因的合成。由于存在着溶解度, 回收率, 和片段分离等问题, 曾认为合成基因的工作, 十分困难, 要普遍应用是不现实的。但技术不断改进。1981 年, 利用合成仪, 由 67 个小片段合成出共有 514 个碱基对的 α -干扰素基因。近年来, 由于 DNA 固相自动合成仪的出现, 合成面貌大为改观。速度达到每加长一个核苷酸, 只需 30 分钟左右, 产率在 97% 以上。自动合成仪使操作手续大大简化。因此, 合成的基因在基因工程中的地位, 就愈来愈重要了。

这个方法之所以受重视, 不仅因为速度快, 产品纯, 和产率高, 它还可以在仅知蛋白质中一部分氨基酸顺序的情况下, 由遗传密码推算出其基因中相应部分的核苷酸排列顺序。按此顺序人工合成的多核苷酸链, 可以模拟真正基因的部分 DNA 链。用之作为探针, 可由基因库或 cDNA 库中筛选目标基因。

合成编码蛋白质的结构基因时, 要考虑密码的简并性质。在制订合成计划时, 使其中的密码子符合寄主细胞所偏重使用的密码子。这对此基因在这种寄主细胞中的表达, 可能十分有利。

此外, 在蛋白质结构、功能关系的研究中, 可以任意改变合成基因中的某个核苷酸, 使其产生的蛋白质中的氨基酸发生变异, 以观察其影响。如 β -干扰素结构中有三个半胱氨酸。将其 17 位的半胱氨酸(密码子为 TGT) 改变成丝氨酸(密码子为 AGT) 后, 发现其生物活性不但受影响, 而且比天然的 β -干扰素更为稳定。它在 -70°C 保存 150 天活性不变, 而天然的干扰素活性则逐渐下降。这对人工合成高

稳定性、高生物活性的蛋白质分子,有极为实际的意义。这种通过基因改造,取得蛋白质分子改造产物的技术,现在称之为蛋白质工程。

当然,通过基因合成,还可以合成出具有特殊启动、调控功能的调控基因元件,以及具有其它功能的基因元件。

合成基因时,总是先合成一些较短的寡聚核苷酸片段。它们在部分核苷酸排列上,存在着交错的碱基配对关系。因此可以先形成一些带有突出末端的短双链。这些短双链上的末端,再通过碱基配对和连接酶的作用,就可由小片段结合成为大片段。大片段借助载体,能得到克隆。经顺序分析证明正确无误后,既可直接作为探针使用,可进一步组装成大的合成基因。

DNA 片段的合成方法,由磷酸二酯法,改进为磷酸三酯法,又进一步发展为亚磷酸三酯法。另一方面,由手工的液相合成,发展到自动的固相合成。反应速度,效率,和回收率都有显著提高。

现以固相亚磷酸三酯法为例,说明如下:本方法是以硅胶作为固相载体。第一个单体是个核苷。它通过脱氧核糖上的 3'-OH,与载体连接。以后按 3' 至 5' 方向延伸,逐个将单体加上去。这与体内的合成方向正好相反。

核苷酸单体都事先制备好。它们在碱基和

核糖上带有保护基,可控制反应发生的位置。如在碱基 A 上为 N⁶-苯甲酰, C 上为 N⁴-苯甲酰, G 上为 N²-异丁酰。T 不需保护。核糖的 5'-OH 用 DMT (二甲氧三苯甲基) 保护。它的 3'-OH 与亚磷酸酰胺酯结合,后者提供了有反应活性的磷酸基。在参与反应之前,需先经活化处理。

合成的第一步是,除去与固相载体相连接的第一个核苷上的 5'-DMT,使可起反应的 5'-OH 暴露出来。第二步是,将第二个单体 3'-OH 上的亚磷酸酰胺酯基活化,然后与第一个核苷上的 5'-OH 偶联。再用 I₂ 将亚磷酸三酯氧化成磷酸三酯。最后是,封闭第一单体上未反应的 5'-OH,以免下一步错接,产生错误的顺序。

以上步骤,可用固相 DNA 合成仪,在计算机控制下,自动连续进行。逐渐加长的 DNA 链,在达到预定长度后,先将磷酸三酯中的甲基脱去,使成为天然的磷酸二酯键。然后用浓氨水将 DNA 链由载体上切下。再除去碱基上的保护基,和最后一个 5'-DMT,以恢复天然的碱基和 5'-OH。

由载体上回收的产物,可用 20% 聚丙烯酰胺凝胶电泳,或高压液相层析,收得纯化的最终产物。