

生物科学参考资料

第九集

科学出版社

内 容 简 介

本书收集了有关遗传工程(主要是基因工程)文章共17篇。包括基因分离,限制性核酸内切酶和DNA连接酶在遗传工程中的应用,DNA体外合成与遗传工程研究,DNA体外重组的载体,遗传工程与基因表达等综述性的研究进展;与农业、医药和国防有关的根瘤菌、高等植物、抗菌素、哺乳动物等方面有关遗传工程的研究新动向;以及减少制备重组质体DNA中的可能危险等也作了介绍。可供生物学、遗传学、生物化学、生物物理、细胞学、微生物学、植物学、动物学及医药工业、农业等有关的科研单位和大专院校师生参考。

生 物 科 学 参 考 资 料

第 九 集

*

科 学 出 版 社 出 版

北京朝阳门内大街137号

中 国 科 学 院 印 刷 厂 印 刷

新华书店北京发行所发行 各地新华书店经售

*

1978年12月第 一 版 开本:787×1092 1/16

1978年12月第一次印刷 印张:9

印数:0001—15,450 字数:185,000

统一书号:13031·887

本社书号:1258·13—6

定 价: 0.95 元

目 录

遗传工程的研究现状和趋势·····	1
基因的分离·····	17
限制性核酸内切酶及其在遗传工程中的应用·····	26
DNA 连接酶研究的新进展·····	37
DNA 体外合成与遗传工程研究·····	44
DNA 体外重组的载体——质体和病毒·····	55
遗传工程与基因表达·····	71
根瘤菌固氮作用的遗传与调节·····	79
目前高等植物遗传工程的载体研究·····	85
放线菌的基因工程——抗菌素产生菌育种的一个新方向·····	91
用遗传重组体对哺乳动物初步所得结果·····	101
真核基因在原核生物中的表现·····	104
遗传工程的技术和潜力·····	106
真核 DNA 在大肠杆菌中遗传功能的表现·····	109
无性繁殖的酵母 DNA 在大肠杆菌中的功能表现·····	117
λ 噬菌体经 <i>EcoRI</i> 酶切后所有片段在大肠杆菌中的无性繁殖·····	127
减少制备重组质体 DNA 中的可能危险·····	129
附录 I 利用体外 DNA 重组技术转移基因·····	132
附录 II 限制酶的类型·····	134

遗传工程的研究现状和趋势

所谓遗传工程,简言之,即应用酶学的方法将不同来源的遗传物质在体外合成重组 DNA 分子,这种重组 DNA 分子在细胞内可以增殖而繁衍下去成为一个纯系。这一过程一般在自然界演化中是不会发生的。遗传工程的研究是分子生物学发展的一个新的领域。如果从体外成功地合成重组 DNA 分子算起,它的历史不超过五年,而且其中由于考虑到生物学上的潜在危险,很多工作曾停止了一年多的时间。我们应该注意到由于近代遗传学和生物化学成果的积累和运用,这个领域在短期内发展很快。在这里生物科学工作者首次将遗传物质进行周密地设计和人工操纵,为研究基因,特别是真核细胞基因的结构和功能提供一个有效的手段,同时为改变生物特性开辟了远景。

根据一些有关遗传工程研究的资料和近年来在这方面的主要研究内容和趋势作下列的介绍。

一、基因载体的设计

在进行遗传工程的过程中为了使体外合成的 DNA 重组分子在细胞内繁衍下去,因此构成这个重组 DNA 分子的一个组成部分将是一个复制子。这个复制子负责运载我们所需研究的 DNA 顺序叫作基因载体。如果基因载体在细胞内的复制率高,那末由此引入细胞的 DNA 顺序的扩增率也大,所以选择复制率高的复制子作为基因载体是实施遗传工程的一个重要环节。但是,更重要的是将那些在操作过程中混杂的 DNA 片段进行淘汰,于是选择就成为必要的条件。遗传工程的设计者考虑到除了基因载体本身具有选择性的遗传标记外,当它载有某个 DNA 顺序时也必须显示出一个特殊的信号,便于我们选择。只有这样,才能将异性群的 DNA 经过增殖成为同质 DNA 供研究或应用。

目前作为原核细胞的基因载体仅有两类复制子:λ 噬菌体和细菌质体(Plasmid)。在英、法都趋向于应用 λ 噬菌体的衍生型作为基因载体。它的优点在于 λ 噬菌体的遗传知识丰富,而且利于基因的扩增,不易引起生物学上的潜在危险。在改建 λ 噬菌体成为基因载体时要考虑的问题有:(1) λ 噬菌体 DNA 有 5 个 *EcoRI* 切点(图 1)如何将其减少成 1 或 2 个切点作为引进外源 DNA 顺序的地方;(2) 如何使形成的 λ-DNA 和外源 DNA 的重组分子具有选择性状;(3) 如何使 λ DNA-外源 DNA 的重组分子在感染细胞中有效地扩增;(4) 如何应用 λ 噬菌体的基因调控系统有利于外源 DNA 顺序的表达。

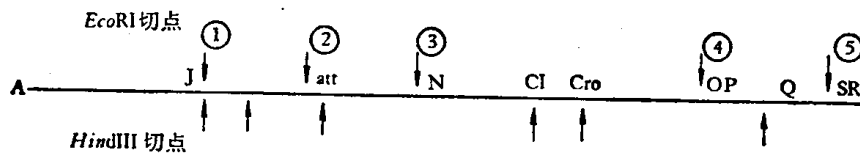


图 1

已知 J—N 这一段是 λ 噬菌体增殖不重要的部分,可以除去,因此可使这个部位保持两个切点作为容纳外源 DNA 之用,而其他切点都可以去掉使其不受限制性内切酶 *EcoRI* 的作用。要做到这一点可以从缺失型 b538 着手。因为这个缺失型只留下 3⁺, 4⁺, 5⁺, *EcoRI* 切点(第 1, 2 切点均在缺失部分), 将这个 b538 缺失型不断交替地在大肠杆菌(*E. Coli*) K 和大肠杆菌 K/R1 中繁殖,测定其噬斑。当其限止比率,大肠杆菌 K 滴定数/大肠杆菌 K/R1 滴定数 = 1 时,表示这个噬菌体已失去其所有 *EcoRI* 的切点,成了抗 *EcoRI* 的类型。然后将这类抗 *EcoRI* 的 b538 型与对 *EcoRI* 呈正常型的 λ 噬菌体杂交,就可以获得具有 1⁺, 2⁺ *EcoRI* 切点,但失去 3, 4, 5 *EcoRI* 切点的 λ 型,即 3[°], 4[°], 5[°] *EcoRI* 切点型。如果将抗 *EcoRI* 的 b538 缺失型和另一缺失型 h803⁺ CI_{ts} 4⁺ 5⁺ Sam 相杂交,可以获得具有第 4⁺ 切点 b538 型噬菌体。由于切点 4⁺ 与 CI_{ts} 相连锁,因此 CI_{ts} 可以用作选择因子。用前一种 λ 型作为基因载体时,外源 DNA 可以插入噬菌体 DNA 的第 1、第 2 切点之间,也就是外源 DNA 顺序置换切点 1—2 之间的 λ 片段(图 2)。

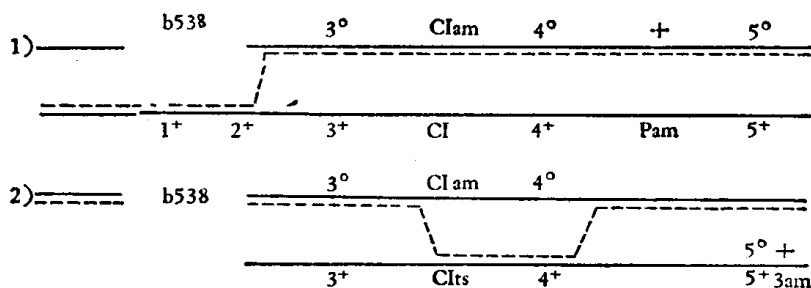


图 2

由于在 J—N 之间发生缺失的部分对这个噬菌体的增殖是不重要的,但因缺失而造成 DNA 分子的残留长度不足以装配蛋白壳时,就影响噬菌体的正常形成。相对的,如果外源 DNA 顺序插入后,增加了 λ -DNA 的长度,就可以恢复它的正常发育。因此在设计时可使 λ 噬菌体在 *EcoRI* 第 1 到第 2 切点之间发生缺失,但保留 2⁺ 和 3⁺ 切点,同时在 4[°]—5[°] 切点之间,在对噬菌体增殖不重要的 nin-5 区造成缺失。这样的 λ 噬菌体当经 *EcoRI* 作用后,由于 2⁺—3⁺ 切点之间的片段(一般称为 C 片段)的脱落,就不能形成噬斑。但当这个 C 片段被外源 DNA 片段所置换时,就可以形成正常的噬斑(图 3)。这样就可以作为选择 DNA 重组分子的条件。

除了采用这些选择特性外,可以利用基因的互补作用或分子杂交的方法选择 DNA 重组分子。最近 Kourilsky 等应用免疫化学方法探测在噬斑中 DNA 重组分子的存在。原理是先将基因的产物——蛋白质,做成抗体,再将组成这个抗体的免疫球

蛋白做成免疫球蛋白抗体,同时使后者与过氧化酶或葡萄糖氧化酶连接成为酶标记的免疫球蛋白抗体。在探测 DNA 重组分子时,即根据双抗体反应进行,当基因的产物蛋白质和它的抗体结合发生沉淀后,再用酶标记的免疫球蛋白抗体,根据葡萄糖氧化酶或过氧化酶的反应进行显示。

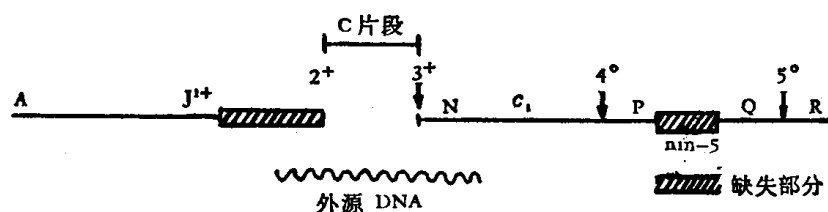


图 3

应用 λ 噬菌体的基因调控系统,提高外源基因的扩增和表达,Edinburgh 大学的 Brammar 等做了一些工作。他们将 λ 噬菌体的 S, R 基因突变为 S^- , R^- 或者将 Q 突变为 Q^- ,由于 Q^- 使 S, R, J, A 基因不能转录,结果延迟大肠杆菌的裂解,相对地使 λ -DNA 与外源 DNA 的重组分子扩增。当外源基因不带有启动基因 p 的条件下,可以利用 λ 噬菌体 p 的作用使基因表达。例如当色氨酸操纵子插入 λ 噬菌体后,由于 λ 噬菌体 p 的强大启动作用能合成过量的色氨酸合成酶。

在原核细胞中,细菌质体是较普遍地被用作基因载体的。理想的细菌质体应该容易从共价闭环形式转变为单位长度形式,含有 1—2 个遗传标记作为选择用。因此,这类质体只含有 1 个限制性内切酶的作用点,而且这个切点不在作为选择标记的 DNA 顺序之中。目前常用的细菌质体都是根据这样的要求而选择来的,如抗四环素质体 (pSC101) (含有耐四环素 T_e^r 标记,分子量为 5.8×10^6 道尔顿),RSF1010 (耐链霉素 Sm^r , 分子量为 5.5×10^6 道尔顿),大肠杆菌素源因子 (ColE1) (决定大肠杆菌素的合成和耐药性 Col^r , 分子量为 4.2×10^6 道尔顿)等。质体在大肠杆菌细胞内的数目和染色体数目之比一般为 1—2。但是 ColE1 有强力的扩增性,在大肠杆菌的正常生长情况下,每个细菌细胞约含有 24 个 ColE1 拷贝,但当有氯霉素存在时,由于染色体复制遭到抑制,而 ColE1 却不受影响,结果 ColE1 的拷贝数可增加到每个细菌细胞含有 1,000—3,000 个。由于这个特点,ColE1 被选作为理想的基因载体。可是由于它的选择标记 Col^r 不太显著,所以遗传工程师设计将它和其他细菌质体如 pSC101 经内切酶作用构成杂种质体 pSC134 (分子量为 10×10^6 道尔顿)。这样的杂种质体具有 ColE1 的扩增性和 pSC101 的 T_e^r 标记。又如 pCR1,是 ColE1 和另一个细菌质体 pSC105 (含有 Km^r 标记)经 *EcoRI* 处理后形成的杂种质体。由于 pSC105 上只有一个 *HindIII* 切点,而且正居 Km^r 标记的 DNA 顺序之中,因此当 pCR1 作为基因载体,利用 *HindIII* 切点作为引进外源 DNA 顺序的插入口时, Km^r 可以用作反选择因子,而 Col^r 作为选择因子有效地选择 DNA 重组分子。

在改建质体方面,例如上面谈到的利用 *EcoRI* 的作用合成杂种质体作为基因载体,这样的杂种质体结果将含有两个 *EcoRI* 作用点。怎样使多增的,不需要的一个切

点去除? Covey, C., Richardson, D. 和 Corbon, J. 等做了下列的实验。他们将杂种质体 pML2 (ColE1 的衍生质体, 在 ColE1 和 K_m' DNA 片段连接之处有两个 *Eco*RI 切点) 用 *Eco*RI 适当处理使只有 1 个切点发生裂解, 然后用 λ 5'-外核酸酶处理除去因 *Eco*RI 裂解作用后产生的单链 5' 末端部分。这样的线状质体 DNA 可用凝胶电泳将其分离(图 4)。

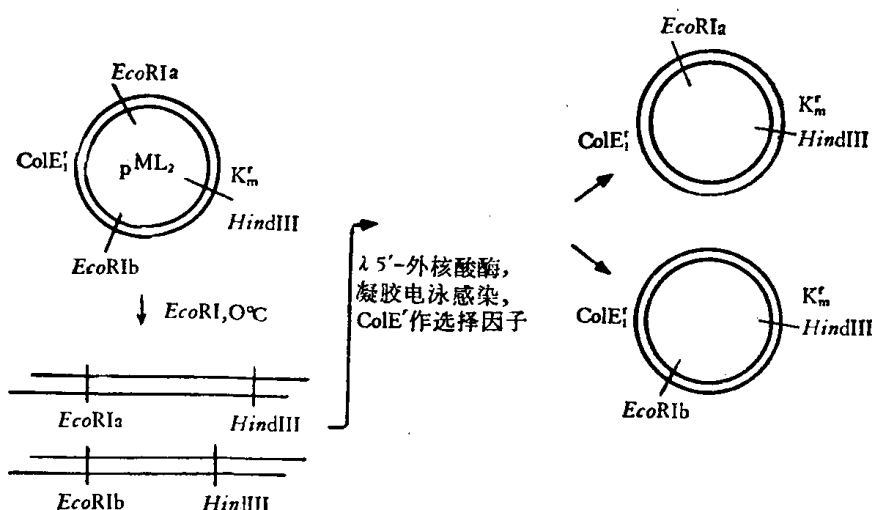


图 4

与此相对的工作是引进限制性内切酶的作用点 DNA 顺序, 据 Berg 在 Harden 会议上的介绍, 他们合成 *Eco*RI 切点顺序 $\begin{matrix} \text{TGAATTCA} \\ \text{|||||} \\ \text{ACTTAAGT} \end{matrix}$, 8 核苷酸顺序。用末端转移酶的接尾法使与 pMB9 相结合, 使 pMB9 获得两个 *Eco*RI 切点(图 5)。

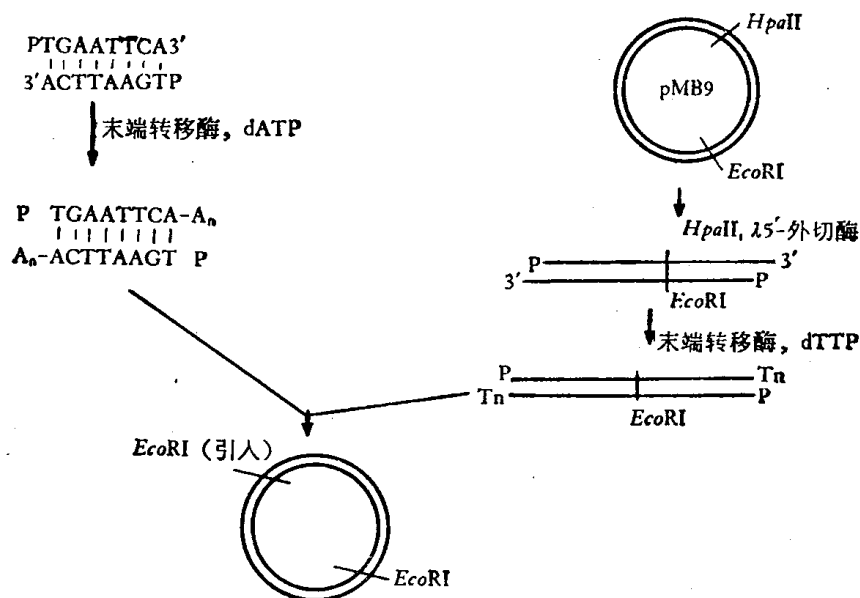


图 5

最后想谈谈真核细胞的基因载体问题。将外源 DNA 顺序移入到真核细胞内并使之增殖,就必须选择一种能在真核细胞内增殖的复制子。致癌的猿猴病毒 (SV40) 是一个理想的真核基因载体。一般采用含有 DNA 复制点的无效 SV40 DNA 片段和外源 DNA 片段进行共价连接,做成 DNA 重组分子用以感染猿猴细胞。Ganem 等利用一些严重缺失型 SV40 突变株,例如 d_3 作为载体。这是由一个具有复制起点的 SV40 片段 d_1 重复连接成的五聚体。每个 d_1 片段包括 880 b. p. 占正常型 SV40 DNA 全长的 17.5%,含有一个 *Hind*III 和 *Hpa*II 切点(图 6)。

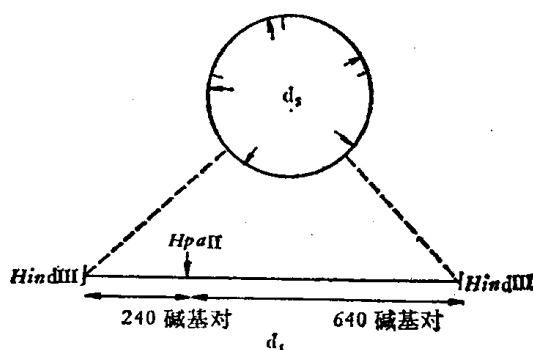


图 6

Ganem 等将 λ DNA 的 *Eco*RI 裂解片段中,分离出 *Eco*RI 片段 B (图 7),

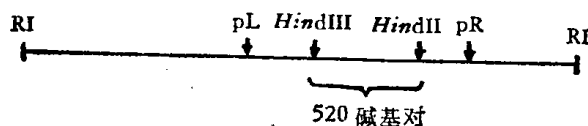


图 7

再将这 *Eco*RI 片段 B 经 *Hind*III 作用,分离出含有 520 碱基对 (b.p.) 长的 λ DNA 作为引入 d_3 的外源 DNA 顺序。当 d_3 的 *Hind*III 裂解片段与上述的 λ DNA 520 碱基对片段经过退火处理和连接酶的作用即可形成 DNA 重组分子 λ - d_1 , 全长为 1400 碱基对。将 λ - d_1 附加少量正常型 SV40 作为辅助因子,感染猿猴细胞即可分离出它的三聚体 (λ - d_3), 大小等于 SV40 DNA 的 84%。

最近 Berg 等直接分离含有复制起点的 SV40 片段。他们将两类缺失型,如 0.54—0.55 缺失型和 0.70—0.73 缺失型形成异源双链分子 (Heteroduplex) (图 8)。然后应用 SI 内核酸酶,将其不配对部分切下,就可以获得具有复制起点的片段,占 SV40 DNA 全长的 13%,这种片段可作为基因载体用。

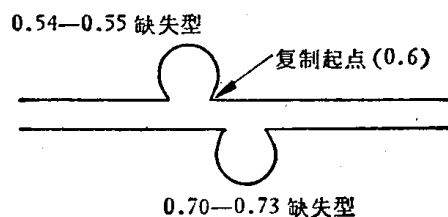


图 8

基因经载体转入受体细胞后的表达,特别是真核基因的表达存在许多问题。有些可能是在设计建成 DNA 重组分子时发生的问题。例如 Berg 等在建立 SV40- λ pgal⁺ 重组 DNA 分子时,由于引入聚(dA:dT)顺序,而使转录仅限于 SV40 DNA 片段部分(图 9)。

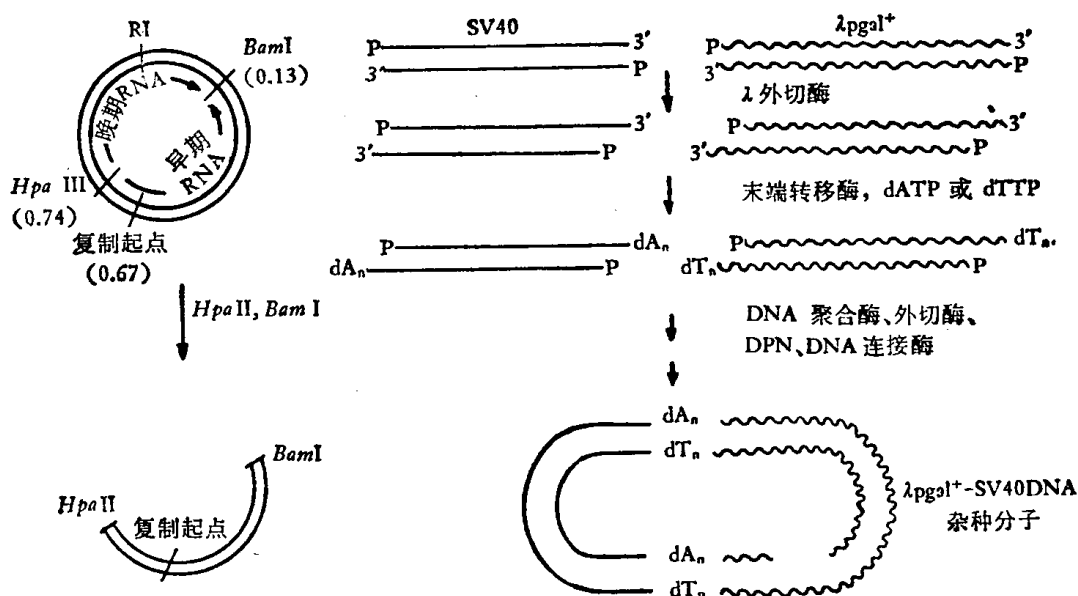


图 9

又如 Cohen 等在研究哺乳动物线粒体 DNA (简称 mit-DNA) 在原核细胞内表达时,他们将老鼠的 mit-DNA 和质体 pSC101 做成杂种分子,当感染大肠杆菌后,发现大肠杆菌 RNA 多聚酶对杂种分子 DNA 的转录不是从 pSC101 部分开始,而是优先从 mit-DNA 开始。

将真核基因经载体送入大肠杆菌后的表达问题,除了上述的例子可能在建立 DNA 重组分子时发生的干扰外,由于原核细胞转录和翻译系统的不同,真核基因的表达受到影响。将真核基因经遗传工程的手段转入大肠杆菌细胞后,至今还没有证明能合成由这个基因所决定的正常蛋白质分子。最近 Davis 等应用面包酵母 DNA 的 *Eco*RI 裂解片段与载体 λ -DNA 结合形成重组分子,以此感染一种不能合成咪唑甘油磷酸酯脱水酶的大肠杆菌组氨酸缺陷型 *His*⁻,发现有互补现象。作者应用分子杂交的方法证明经感染而进入受体细菌的 λ DNA-酵母 DNA 的重组分子确乎含有酵母 DNA 片段。随着 Carbon 等人又证明当酵母 DNA 片段经载体 *ColE1* 转入一些大肠杆菌的缺陷型(缺失性突变型)如亮氨酸缺陷型 *leu*⁻ 和色氨酸缺陷型 *trpAB*⁻,同样产生互补现象。说明酵母细胞中决定上述的咪唑甘油磷酸酯脱水酶, β -异丙基苹果酸脱氢酶以及色氨酸合成酶的基因可以转入大肠杆菌中而表达信息。这是首次成功地使真核基因在原核细胞中正常表达。

二、真核基因的结构和表达

应用遗传工程的手段研究真核基因的结构和作用正在广泛展开。在分子生物学的发展中,科学工作者逐渐注意到发育生物学或发育遗传学上的一些基本问题,例如什么是真核基因?真核基因在染色体 DNA 上是怎样排列和怎样表达的?遗传发育在染色体上怎样密码的?这里就涉及到基因表达的顺序性调控问题。这些基本问题在目前可以说并没有搞清楚。拿具体的实验例子来说,果蝇 (*Drosophila melanogaster*) 的基因组由 4 个染色体组成,X 染色体含有 DNA 39,000 千碱基对 (Kb, 下同) 第二染色体为 60,000 千碱基对,第三染色体为 63,000 千碱基对,第四染色体为 6,000 千碱基对,总共为 165—170,000 千碱基对。这些染色体分成 5,000 多个染色粒或称染色粒单位。每个染色粒单位平均含有 26 千碱基对 DNA。根据细胞遗传学实验,每个染色粒单位中只含有一个基因,所谓一个染色粒一个基因。多余 DNA 顺序的作用是什么?遗传工程对这类复杂的问题提供了足以着手研究的途径。Hogness 将果蝇染色体 DNA 用 *EcoRI* 进行部分裂解,获得 DNA 片段,称为 Dm 片段长达 14 千碱基对左右。这些 Dm 片段再与同样经 *EcoRI* 裂解的质体 pSC101,通过 DNA 连接酶处理,即形成共价键的 Dm-pSC 杂种分子,简称 PDm。将其感染大肠杆菌,增殖成含有不同 Dm 片段的 PDm 杂种分子。Hogness 等又建立了一个筛选 PDm 菌落的方法。这个方法的主要步骤是使含有 PDm 杂种分子的菌落培养在铺在琼脂平板上的硝酸纤维滤膜上,用复印法做好参考菌落平板,放置在 2—4℃ 作为备用。将生长在硝酸纤维膜上的菌落,从培养板底部灌入 0.5 N NaOH,经洗涤后,再用蛋白酶处理使其自溶,发生 DNA 变性,使固定在滤膜上成为“DNA-印迹”。然后用与 Dm 顺序相互补的标记 ³²P-RNA 或 ³H-RNA 与之进行分子杂交,根据 X 光自身显影肯定其杂交结果。于是从参考菌落平板上挑取相应菌落进行培养,提取不同的 PDm 分子做试验,测定其 DNA 顺序的重复程度。如果将各种 PDM 做成的标记 c-RNA (如 ³H-cRNA) 与果蝇的多线染色体进行原位杂交,可以测定这些 Dm 在染色体上的分布。据 Hogness 报道有两类 Dm 片段:一类是 DNA 顺序不呈重复性的,限于一个染色粒区,分布在各个染色体上。另一类的 DNA 顺序呈不同程度的重复性,不在染色粒区内。其中有一个 Dm 分布在染色体中心的异染色质部分。还有一个 Dm 分布在核仁部位,根据分子杂交试验证明这个 Dm 含有密码 18S 和 28S rRNA 的 DNA 顺序。这是应用遗传工程研究真核细胞染色体结构的结果。作者对果蝇染色体 DNA 顺序发生串联式的重复提出了推论。认为这是由于染色体间不平均重组的结果。

另外一个例子是关于组蛋白基因的研究。因为组蛋白基因的活动在细胞周期和生物的发育时期所表现出的一些调节现象,所以许多实验室在研究真核基因组的结构和表达时都以组蛋白基因作为研究模型。加以组蛋白基因的产物, mRNA 的提取和分离目前已不存在什么问题,利用遗传工程作为研究途径更显示其优越性。瑞士 Zurich 大学分子生物研究所的 Schaffner 利用海胆 (*Psammechinus miliaris*) 为材料进

行了下列的工作。在分离和制备组蛋白 mRNA 中,先提取标记的 ^3H -9SRNA,根据离体翻译系统的实验证明 5 种组蛋白 mRNA 的存在。在提取密码 5 种组蛋白基因时,他们利用质体 pSC101, pCR1 或 λ 噬菌体作为载体将海胆染色体 DNA 片段经 *Eco*RI 作用形成 DNA 杂种分子,在大肠杆菌中进行增殖。然后提取 DNA 使与上述的组蛋白 mRNA 杂交,证明其是否含有密码组蛋白的 DNA 顺序。这样就可以获得长达 6—7 千碱基对的组蛋白-DNA (简称 hist-DNA) 片段。一般,运用这种 DNA 杂种分子的扩增办法,估计在含有不同种 DNA 片段的大肠杆菌细胞中有 1/2,000—1/3,000 的机率获得这种 6—7 千碱基对 hist-DNA。将这 6 千碱基对 hist-DNA 片段经内切酶 *Hind*II, *Hind*III 和 *Eco*RI 混合处理,得到 5 个片段,由于这 5 个片段都能与上述的 ^3H -9S RNA 杂交,说明每一片段均有密码组蛋白 DNA 顺序,而且这些内切酶的切点均不在结构基因之中。作者应用片段的重叠法将这 6 千碱基对 hist-DNA 进行物理定位,也就是确定它 5 个片段的顺序(图 10)。然后将 5 个片段分别与 5 种 hist-mRNA 进行分子杂交,肯定 5 种组蛋白基因的位置。

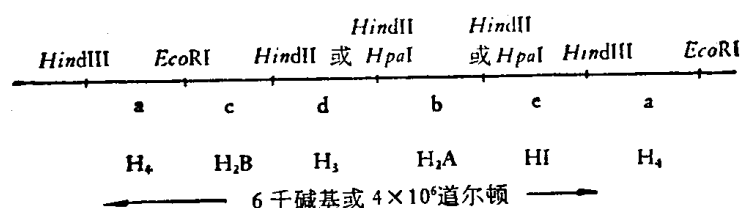


图 10

作者进一步研究这些组蛋白基因的转录方向。他们采用外核酸酶处理,再以组蛋白基因 H_4 mRNA 来确定 H_4 基因在 6 千碱基对片段上的方向。要知 H_4 基因在 *Hind*III 片段 (6 千碱基对) 上可以有两种动向:

- (1) $5' \rightarrow a \ c \ d \ b \ e \rightarrow 3'$
 $3' \leftarrow \text{—————} 5'$
- (2) $5' \rightarrow e \ b \ d \ c \ a \rightarrow 3'$
 $3' \leftarrow \text{—————} 5'$

在 *Eco*RI 片段, H_4 基因的两种可能动向:

- (3) $5' \rightarrow c \ d \ b \ e \ a \rightarrow 3'$
 $3' \leftarrow \text{—————} 5'$
- (4) $5' \rightarrow a \ e \ b \ d \ c \rightarrow 3'$
 $3' \leftarrow \text{—————} 5'$

由于外切酶的作用从 5' 末端开始,如果 H_4 基因在 5' 末端,则将被切除,但在 3' 末端便无影响。实验证明,外核酸酶对 *Hind*III 片段作用后仍保持与 H_4 -mRNA 杂交的部分,说明 H_4 不在 5' 末端而是在 3' 末端,因此模型(2)是正确的。如果以 *Eco*RI 片段用外核酸酶处理,再用 ^3H -mRNA 进行杂交试验,发现已失去 H_4 顺序,说明模型(4)是确实的。

总结试验结果说明组蛋白 DNA 6 千碱基对片段上组蛋白基因的极性顺序将是 $H_I \rightarrow H_2A \rightarrow H_3 \rightarrow H_2B \rightarrow H_4$ 。由于证明组蛋白 mRNA 仅与 6 千碱基对 hist-DNA 分子的一条单链相杂交,说明密码顺序同样在一条链上呈极性倾向,转录的方向将是: $H_4 \rightarrow H_2B \rightarrow H_3 \rightarrow H_2A \rightarrow H_I$ 。前面提到 6 千碱基对片段包含 5 个组蛋白基因,成为一个单位在染色体上重复地串联着。但对这 5 个组蛋白基因 2 千碱基对 DNA 就足以包括了。多余的 DNA 顺序是什么?有什么作用?根据 6 千碱基对片段的 T_m 测定 ($1/2$ 为 83°C , $1/2$ 为 90.3°C),推测 GC 含量为 37% 和 52%。而 50% 左右的组蛋白 DNA 将是 AT。在海胆组蛋白 mRNA 中证明 GC 较高,因此 AT 部分将不属于结构基因内部。再根据组蛋白 DNA 部分变质图谱 (Partial denaturation map),证明 GC 区与 AT 区交叉存在,GC 区相当于基因所在,AT 区是间隙区。接近 H_4 基因处是一末端间隙区。Kedes 利用海胆 (*S. Purpuratus*) 为材料得到同样的结果。根据组蛋白-DNA 经限制性内切酶作用的裂解片段与组蛋白 mRNA 进行分子杂交的结果证明组蛋白基因之间存在着间隙区。有趣的是,果蝇的组蛋白-DNA 以 4 千碱基对片段成为一个单位在染色体重复地串联着,根据分子杂交发现与海胆的组蛋白 mRNA 呈同源性顺序。从这里可以看出真核生物的演化关系。

三、病毒和线粒体基因组的物理定位和基因定位

要知道在一些肿瘤病毒中决定引起肿瘤性状的基因所在,利用不同的限制性内切酶分析它的基因组的结构,进行物理定位和基因定位是一个重要的手段。英国 Glasgow 大学的病毒研究所开展了关于腺病毒 (Adenovirus) 和疱疹病毒 (Herpes Virus) 的基因定位工作。在腺病毒方面, J. Williams 等利用对某些限制性内切酶呈不同切点分布的温度敏感性 T_r 突变型进行杂交,分离其重组子 (T_r^+),用内切酶分析其基因组的裂解片段的变化,这样就可以从物理图谱断定其重组的部位,同时可以测定 t_r ,也就是引起细胞转化的基因所在。例如腺病毒 $Ad_2 t_r$ 温度敏感型只能在 32.5°C 在人体细胞内增殖而引起转化,但在 38.5°C 就没有作用。而另一种温度敏感型 $Ad_2^+ ND1 t_r$, 是 Ad_2 和 SV40 的杂种病毒,这种病毒在 Ad 病毒基因组的一端发生 5.5% 的缺失。在这缺失部分插入相当于 SV40 基因组 10% 的 SV40 DNA 顺序。因此这两类 T_r 病

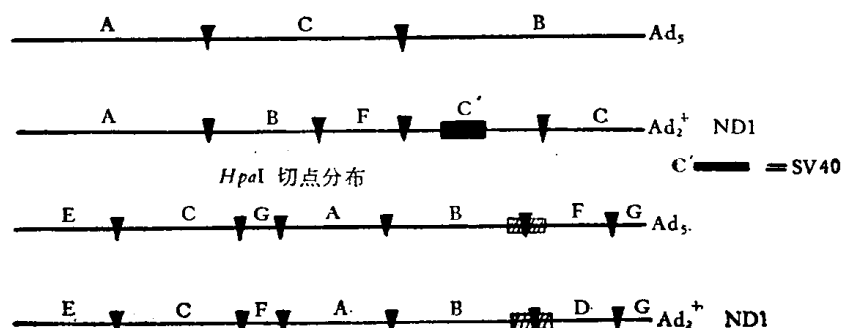


图 11

毒的 DNA 顺序是不同的。根据应用 *EcoRI*, *HpaI* 等裂解和部分裂解, 以及异源双链 (Heteroduplex) 等测定, $Ad_3 t_1^+$ 和 $Ad_2^+ ND1 t_1^+$ 型的 *EcoRI* 和 *HpaI* 切点顺序在基因组上确定下来 (图 11)。

EcoRI 切点分布:

将 $Ad_3 t_1 2$ 突变型与 $Ad_2^+ ND1 t_1 4$ 相杂交, 选择 t_1^+ 重组子, 将重组子基因组分别以 *EcoRI* 和 *HpaI* 处理, 同样将两个亲本也分别用这两种限制性内切酶处理作为对照。将裂解产物经琼脂凝胶电泳进行比较就可以断定其重组的部位和 T₁ 突变点的部位 (图 12)。 $Ad_3 t_1 2 \times Ad_2^+ ND1 t_1 4 \rightarrow t_1^+$

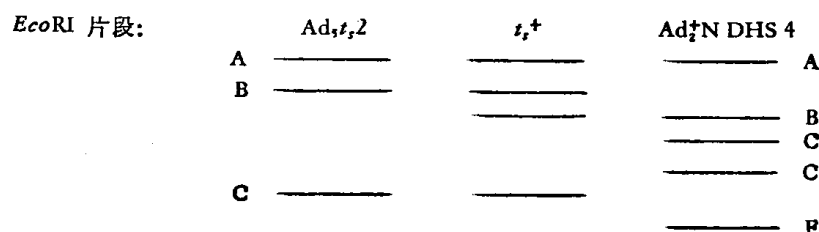


图 12

从上列 *EcoRI* 片段分析, 可以看出重组子 T_1^+ 的 A、B、C 片段与亲本 $Ad_3 t_1 2$ 的 A、B、C 片段的泳动相同。片段 A 在两个亲本基因组中相同, 但重组子产生了一个新的片段, 介乎 Ad_3 的 B 片段与 $Ad_2^+ ND1$ 的 B 片段之间。将 T_1^+ 重组子的 *HpaI* 裂解产物经琼脂凝胶电泳进行比较分析, 同样发现与亲本 Ad_3 的裂解片段相同, 但与亲本 $Ad_2^+ ND1$ 不同, 实验说明重组的发生如图 13:

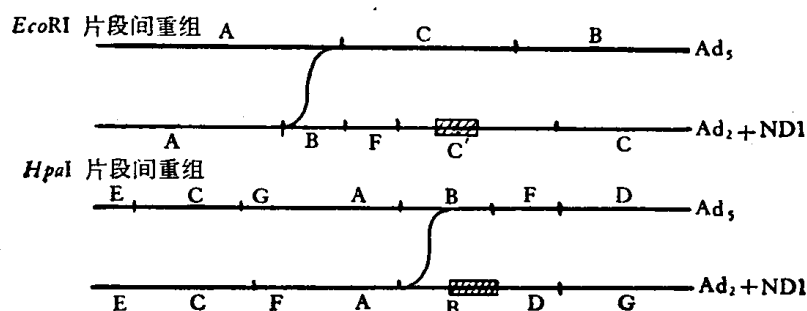


图 13

因此重组子基因组的右端来自亲本 Ad_3 的 DNA 顺序, 而左端含有亲本 $Ad_2^+ ND1$ 的 DNA 顺序。由此可以推断 $Ad_3 t_1$ 突变株的 $t_1 2$ 突变点相当于 $Ad_2^+ ND1$ 的 B、F 片段之左的 Ad_3 DNA 顺序上, 而 $Ad_2^+ ND1 t_1 4$ 突变型的 $t_1 4$ 突变点应位于 $Ad_2^+ ND1$ 的 A、B 片段之右。同样, 根据不同的 $Ad_3 t_1$ 突变型与 $Ad_2^+ ND1 t_1$ 相杂交, 将 Ad_3 不同 t_1 突变点在基因组的顺序定位下来 (图 14)。

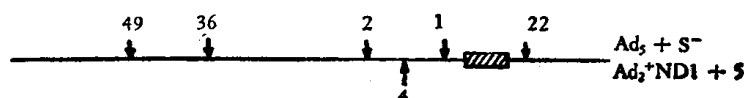


图 14

Williams 的这种方法已被广泛应用。Wilkie 就是应用限制性内切酶，异源双链等技术研究单纯疱疹病毒 (Herpes Simplex) 的物理定位。证明这个病毒的基因组含有高度重复性的 DNA 顺序，而且有些重复性顺序是反向的。他同样以 ϵ 为遗传标记，结合物理定位根据 ϵ 间的互补确定 ϵ 基因的图位。

在真核细胞中，细胞核和其他细胞器之间的遗传学关系逐渐引起重视。啤酒酵母 (*Saccharomyces cerevisiae*) 的线粒体基因组的研究，最近几年来有所突破。法国科学中心分子遗传研究室 Slonimski 做了大量的工作。一方面由于大量点变或微小缺失型突变株 mit^- 的发现。另一方面由于限制性内切酶技术的应用和线粒体 DNA (mit-DNA) 物理图谱的建立。mit-DNA 的物理图谱和线粒体基因的定位工作主要包括二个步骤：(1) 应用不同限制性内切酶将 mit-DNA 切成片段，经琼脂凝胶电泳分离，然后定位；(2) 应用标记 RNA，如标记 rRNA 或与 mit-DNA 片段相互补的标记 cRNA 进行分子杂交。在这里 rho^- 突变型就是一个有用的试验材料。因为 rho^- 突变型大多是严重的缺失性突变(有的达 99.9% 的 DNA 缺失)，而残留的顺序却不断扩增以补偿其不足，因此形成了重复性顺序。Slonimski 等获得了很多这些含有属于线粒体基因组不同部位的 rho^- 突变型。这些 rho^- 突变型的遗传信息可以通过正常型 rho^+ 的杂交而确定下来。有了一系列 $\text{rho}^- \text{mit}^-$ 的突变就可用作进行线粒体基因组物理定位和基因定位的工具。例如下图所示，一个 rho^- 突变型由于 mit-DNA 的大量缺失形成重复性顺序。经内切酶 *Hind*II 和 *Hind*III 作用后，必然产生在电泳上表示出相当于某些正常型 (rho^+) 的 DNA 片段外，还有一个新产生的接合片段(图 15)。

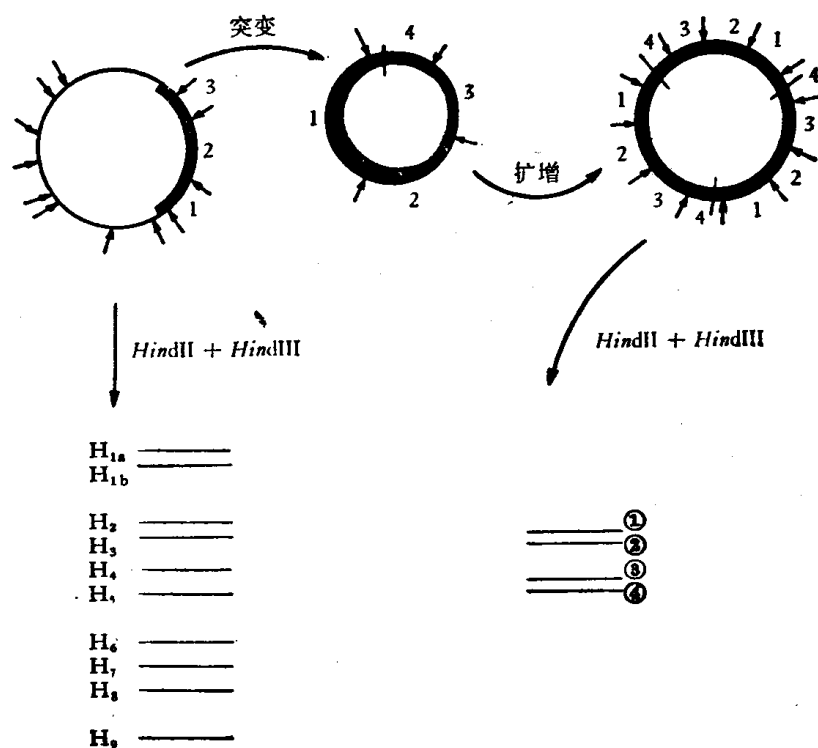


图 15

将这种 ρ^- mit-DNA 制备成 cRNA, 然后与 ρ^+ mit-DNA 进行分子杂交, 如果 ρ^- cRNA 能与 ρ^+ 的 H_2 、 H_3 、 H_4 片段相杂交, 说明这个缺失型保持了原来 mit-DNA 的 H_2 、 H_3 、 H_4 顺序。为了确定某个遗传标记的位点, 就必须选择带这个遗传标记的 ρ^- 突变型, 制备 cRNA 或 cDNA, 再与上述的 H_2 、 H_3 、 H_4 片段进行杂交试验。如果 cRNA 或 cDNA 仅能与 H_4 片段相杂交, 表示这个遗传标记在 H_4 片段上, 举一个实例 ρ^- 突变型 F12, 含有相当于 ρ^+ 型的 DNA 片段 H_4 、 H_5 、 H_6 和 H_7 。以 21S 核糖体 RNA (rRNA) 作为一探针, 从分子杂交试验证明决定这个遗传位点的部分是在 H_6 。为了确定对寡霉素 (Oligomycin) 呈抗性的遗传标记 O_1 的位点, 利用许多只含有 O_1 标记的 ρ^- 突变型如 RP6、RP15/A1 和 c111 制备成 cRNA, 经过滤杂交试验证明由 C111 制备的 cRNA 能与 F12 的 H_5 和 H_7 片段相杂交; 而根据 Southern 的纸片杂交试验, 发现 RP6 或 RP15/A1 cRNA 仅能与 F12 的 H_5 片段杂交。从这些试验结果说明遗传标记 O_1 在 H_5 片段上。利用这种方法 Slonimski 实验室将许多遗传标记在 mit-DNA 上进行定位。

法国巴黎第七大学分子生物研究所, 分子遗传研究室的 Bernardi 等对啤酒酵母线粒体基因组的组成应用遗传工程的技术进行系统分析。他们发现啤酒酵母细胞的 mit-DNA 有两个明显不同的顺序部分: (1) A + T 较多部位, (G + C 少于 5%) 呈 AT:AT 和 A:T 的重复性顺序; (2) G + C 较多部分 (平均 G + C = 32%), 组成上明显的异质性。Bernardi 认为 A + T 较多部分属于基因的间隙部分, 而 G + C 较多部分属于基因所在, 也包括调节基因作用的部位。这两部分, 大小相似 ($>1.6 \times 10^5$) 在 mit-DNA 上交替存在。应用限制性内切酶 *Hal* III 和 *Hpa* II 的作用将 mit-DNA 裂解, 发现 *Hal* 和 *Hpa* 的切点 (分别为 GGCC, CCGG) 远大于一般统计上的估计, 同时证明这些 *Hal* 切点与 *Hpa* 切点是群集在一起的, 约占 mit-DNA 的 2.6—3%, 即 $1.3—1.5 \times 10^6$ /单位基因组。应用小球菌核酸酶 (micrococcal nuclease) 将 mit-DNA 的 G + C 较多部分分离出来, 实验证明这个部分可以分为二个区: (1) *Hpa* 切点群集区; (2) G + C 群集区 (不含有 CCGG, GGCC, GCGC 顺序)。Bernardi 认为这个区不会是基因的结构部分, 也不会是基因的间隙部分。这二个区与“基因-间隙部分”紧接, 在 mit-DNA 重复存在如图 16。

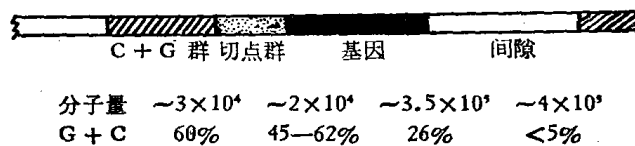


图 16

如果上述的模型是正确的话, 那末啤酒酵母的线粒体基因组在基因表达时只能形成单基因信息, 实际上确是如此, 它的间隙部位是不转录的。照例在基因组上相隔较远的基因应该协调表达。实际上也确实这样。如通气诱导即由于调节蛋白和相同顺序的操纵基因相作用的结果。从这里可以看出啤酒酵母线粒体基因组的重复性顺序的分散系统与细胞核的染色体基因相似。

Bernardi 将经 *EcoRI* 作用的 mit-DNA 片段与 pSC101 或 pCR1 相连接,然后将这种重组 DNA 感染大肠杆菌,再从大肠杆菌中分离出 mit-DNA,发现 mit-DNA 上的间隙部分的组成不同了。推测这是间隙区发生不均匀重组的结果。Bernardi 认为一些 Rho^- 突变型的自发形成可能即由于 mit-DNA 上 A-T 较多区发生内部重组的缘故。由于切点群集部分具有对称性顺序,推测与原核生物的启动基因 P 和操纵基因 O 相似,具有与蛋白质分子相结合的作用,因此线粒体基因组的切点群集部位可能属于基因调节的顺序部分。

四、关于固氮基因的问题

将固氮基因通过遗传工程从细菌细胞转移到非豆科植物,使一些主要的农作物获得固氮的能力,这是人们的理想。负责将外源基因输送到真核细胞内的载体目前研究得很少。较多的工作限于 SV40 或它的衍生病毒。细菌的固氮系统除固氮酶外,还包括电子传递和能量供应等问题。虽然 1971 年英国 Sussex 大学 Dixon 等成功地将肺炎克氏杆菌 (*Klebsiella Pneumoniae*) 的固氮性状通过质体 R144drd 转移到大肠杆菌,使后者获得固氮能力,这是一个突破。但是决定固氮性状的遗传结构到目前为止还不是很清楚的。研究固氮基因操纵子的结构和固氮基因的表达将是进行遗传工程的先决条件。最近一、二年来关于这方面的工作有了一些进展。Dixon 等选择一个致育性较高的 RP4 质体作为固氮基因 *Nif* 的载体,在细菌内进行重组。他们先组成 F'_{nif} 质体。将含有 R144drd 的肺炎克氏杆菌 HF, 与大肠杆菌 SB1824, $His^- Hfr$ (F' 插入近 *his D* 位置)杂交,挑选稳定的 $Nif^+ his^+$ 杂交子(称 uNF43),这个杂交子再与大肠杆菌 $his^- rec A$ 杂交,以 his^+ 为选择因子即获得 $his^- Nif^- recA/F' his^+ Nif^+$ 。

在建立 RP41 ($Km^R Te^R Carb^R his^+ Nif^+$) 的过程,先将 RP4 ($Carb^R Km^R Te^R$) 引入大肠杆菌 SB 1801 ($\Delta Gnd, his, Shi A/F' met^+ Gnd^+ his^+ Nif^+ Shi A^+$ 即 FN68),然后再将杂交子 SB 1801 (RP41) (FN68) 与大肠杆菌 uN F502 (his^-) 杂交,以 $his^+ Km^R$ 为选择因子即获得含有 $Km^R Te^R Carb^R His^+ Nif^+$ 的杂交子。这个杂交子对质体 P 属有专一性的噬菌体 P_{RR1} 呈敏感性,但对 F 专一性的噬菌体 MS2 失去敏感性,说明 FN68 与 RP4 在细菌体内形成重组分子,一方面仍保持 RP4 那样的 P 非亲和性及移位子 (Tn) 顺序,同时又带有染色体 $his^+ Nif^+$ 部分。最后将这种杂交子与大肠杆菌 JC 5466 ($recA, his^- Nif^- Spc^R$) 杂交,即获得大肠杆菌 JC 5466 (RP41)。RP41 的分子量为 59×10^6 道尔顿。

应用这种转移频率较高和寄主范围较广的质体 RP4, Sussex 的科学家们试验 *Nif* 基因在亲缘关系较远的一些细菌中的转移和表达;同时利用 *Nif* 突变点的互补试验,确定固氮基因在肺炎克氏杆菌基因组上的顺序。他们将含有 Nif^+ 的 DNA 片段经 RP41 送入农杆菌 (*Agrobacterium tumefaciens*, 也叫根癌病土壤杆菌), 苜蓿根瘤菌 (*Rhizobium meliloti*) 和棕色固氮菌 (*Azotobacter vinelandii*)。试验的目的在探测,如果农杆菌获得固氮基因,是否可以通过它使一些非豆科植物获得固氮的能力。根瘤

菌不能自生固氮是否由于它的 *Nif* 基因不能表达？棕色固氮菌不需嫌氧而能固氮，是否它的 *Nif* 基因的结构或调节系统不同？在第一项试验中，证明 *Nif*⁺ 可以转入农杆菌而增殖，但不能在农杆菌中表达。在第二项试验中同样证明 RP41 转入根瘤菌，但 *Nif*⁺ 不能表达。当 RP4 转入棕色固氮菌 (*Nif*⁻ *Nal*^R) 以后，后者就获得肺炎克氏杆菌 *Nif*⁺ 基因，而且不需嫌氧能表达固氮作用，说明棕色固氮菌中有一个调节去氧的过程。

在基因互补方面，主要在建立质体——*Nif*/染色体——*Nif*⁻ 部分杂交子。不同的质体——*Nif*⁻ 即 RP41-*Nif*⁻ 是从一株缺失整个 *his*⁻ 和 *Nif*⁻ 操纵子的突变型 UNF107 中分离出来的，然后通过接合将其输送到不同的染色体——*Nif*⁻ 突变型，观察这种 RP4-*Nif*⁻/*Nif*⁻ 部分杂交子的固氮能力。根据互补试验有下列的 *Nif* 基因已被肯定下来，在染色体上的顺序是：

his D-*Nif* B-*nif* A-(*nif* L)-*Nif* F-*Nif* E-*Nif* K-*Nif* D-*Nif* H

nif B: 决定在 *mo*-Fe 蛋白组成中，*mo*-辅因子的合成或者它的插入。

nif A: 决定调节蛋白。因为呈多效应作用，它本身的产物不是固氮酶的多肽，但对 *nif* E、D、H 等都能互补，作者认为它的产物可能是 *nif* 基因转录时必需的活化因子。

nif L: 与 *nif* A 相似，可能属于另一类调节基因。

nif F: 决定电子传递系统的蛋白质。

nif E: 可能是决定调节蛋白质。

nif K、*nif* D: 决定 *mo*-Fe 蛋白的合成，可能是分别决定二个亚基的结构基因。

nif H: 决定 Fe-蛋白多肽的结构基因。

nif 基因在染色体上怎样以操纵子形式存在的，至今还很不清楚。Kennedy 等在美国 Brookhaven 举行的固氮遗传工程会议上谈到根据应用 P1 噬菌体转导进行 *nif* 定位，可将 *nif* 基因分成两个图位群。一群，它与 *his* D₂ 的共转率为 53—72%，另一群它与 *His* D₂ 的共转率为 25—33%。两者之间相差 20%。按公式 $f = (1 - d/L)^3$ ， f = 共转频率， d = 两个共转基因间的距离， L = 转导 DNA 片段的长度(估计 P1 = 80×10^3 碱基对)。测得 d 值后构成下列图距(图 17)：

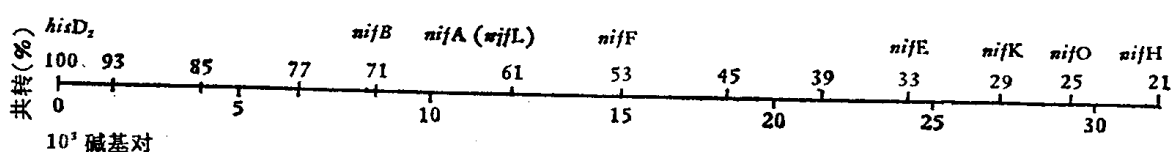


图 17

从上列图谱可以看出在连接 *His* 操纵子的 *Nif* 群中，一群靠近 *his* D₂ 约占 7 千碱基对，而另一群离 *His* D₂ 较远，占约 8 千碱基对，两者之间的距离约 8 千碱基对。Kennedy 认为这两个 *nif* 群的间距 8 千碱基对 DNA 片段可能不包括固氮的遗传信息。

Sussex 的 Ausubel 和 Cannon 等人进行了关于固氮基因的遗传工程的研究。他们提取 RP41，用 Distamycin A 处理以减少 *Eco*RI 的切点，然后将 *Eco*RI 处理 RP41 和