

专论与综述

克隆 DNA 的体外定位诱变

陈 乃 用

(中国科学院微生物研究所, 北京)

克隆 DNA 体外定位诱变就是利用寡核苷酸为诱变引物,按照设计好的方案,在已知 DNA 顺序中准确地改变蛋白质氨基酸编码中一个或几个碱基,以研究蛋白质结构与功能的关系。同样,定位诱变还用于研究基因调节中转录、转译、RNA 加工等和核苷酸顺序改变的关系,准确找出基因调控的关键部位,以及在 DNA 顺序中增加必要的顺序或限制酶切位点,删除不需要的顺序等。这种对 DNA 顺序任意进行设计和改造的方法是不能利用传统遗传诱变方法实现的。因为传统的诱变方法是对整个生物进行随机诱变处理,不仅诱变效率低,而且只能凭借表型鉴别突变体,对于那

些难以用表型区别的突变是不适用的。而这些缺点正是体外定位诱变法的长处。

体外定位诱变要求:(1)克隆到载体上的待诱变 DNA 片段已测序,(2)以单链 DNA 为模板,如 M13 类单链噬菌体等,(3)人工合成的寡核苷酸为诱变引物。近年来,随着 M13 类载体在 DNA 测序中广泛应用和核苷酸自动合成仪逐渐普及,上述条件在国内一般分子生物学研究单位已不难办到。

定位诱变的基本方法是通过单个引物构成:单点诱变,多点诱变,插入诱变和缺失诱变等(图 1)。

定位诱变的基本原理^[1]如图 2:(1)把诱变引物

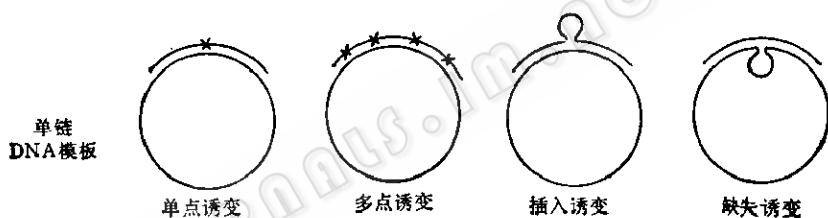


图 1 寡核苷酸引物用于单点、多点、插入和缺失诱变的示意图

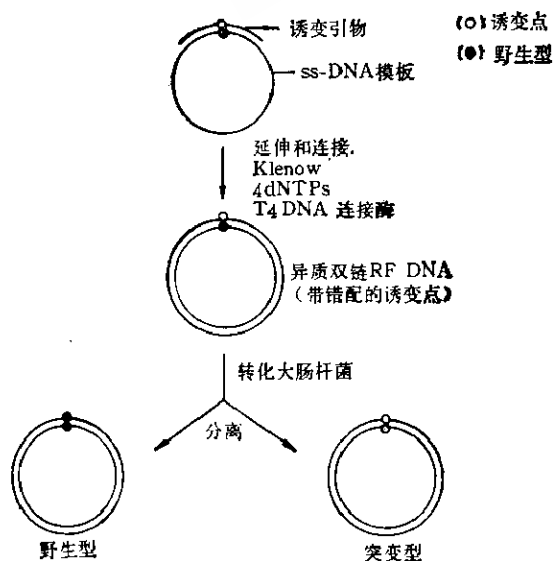


图 2 克隆 DNA 体外定位诱变基本原理

结合到野生型 ss-DNA 模板上,(2)借助于 DNA 聚合酶、DNA 连接酶和四种 dNTPs 使诱变引物沿着模板延伸并连接成异质双链 DNA,新合成链中含有诱变点,(3)异质双链 DNA 转化大肠杆菌后分离出野生型和突变型子代。从理论上讲,野生型和突变型在子代中应各占一半。但实际上,突变型最多只有百分之几。突变型子代所占比例很小的主要原因是:(1)在引物延伸、连接反应混合物中,除了异质双链 DNA 外,还混杂相当多未结合引物的单链 DNA 模板和引物延伸不完全的局部双链 DNA 分子^[2,3]。这两类 DNA 分子转化后都产生野生型子代,大大增加了子代中野生型的比例;(2)大肠杆菌中存在甲基定向的错配 DNA 修复系统^[4],专门修复非甲基化的 DNA。细胞中新合成的 DNA 链尚未甲基化,其中错配的位置就会优先被修复,从而防止了突变的发生。同样,体外定位诱变的 DNA 链也是未甲基化的,在细胞中也会被修复,因此转化子代中出现的主要是野生型。

为了克服上述缺点,可以用蔗糖梯度离心法或凝

胶电泳法分离异质双链 DNA, 或用硝酸纤维素滤膜分离法^[17]除去 ss-DNA, 以及选用 *mutL*、*mutS* 或 *mutH* 突变的大肠杆菌为转化宿主, 阻止错配 DNA 的修复^[4], 增加突变体的比例。但更有效的方法还是采取一些措施阻止野生型子代生长。例如作者和 Paulus^[12]在研究枯草杆菌天冬氨酸激酶 II β 亚单位的转译起始密码子 ATG 时, 曾利用野生型 DNA 顺序中唯一的 *NcoI* (C↓CATGG) 切点破坏野生型 RF DNA, 使诱变率提高到 90%。这个措施十分有效, 但应用范围显然是有限的。因为在诱变点上具有单一限制酶切点的机会是不会很多的。下面介绍的五个改进方法, 各具特色, 虽然需要增加一些实验步骤, 但都能有效地提高诱变率。其中聚合酶链反应 (PCR) 法是新发展的一类新技术。

(一) 缺口双链 DNA 法^[11]

在重组 DNA 研究用的载体中, 如果某些主要基因带有无意义突变, 这类载体只能在含有无意义突变抑制基因 (Su^+) 的宿主中生长, 在 Su^- 宿主中是致死的。本方法就是利用 M13mp9 等带有琥珀突变 (UAG) 的载体为模板, 用 Su^- 宿主选择突变型子代, 单点诱变率可达 60—70%。缺口双链 DNA 法的要点如下 (图 3): (1) 把待诱变的 DNA 片段克隆到含琥珀突变的 M13 类载体上, 制取 ss-DNA 模板; (2) 把不含琥珀突变的 M13 载体的 RF DNA 酶切成线

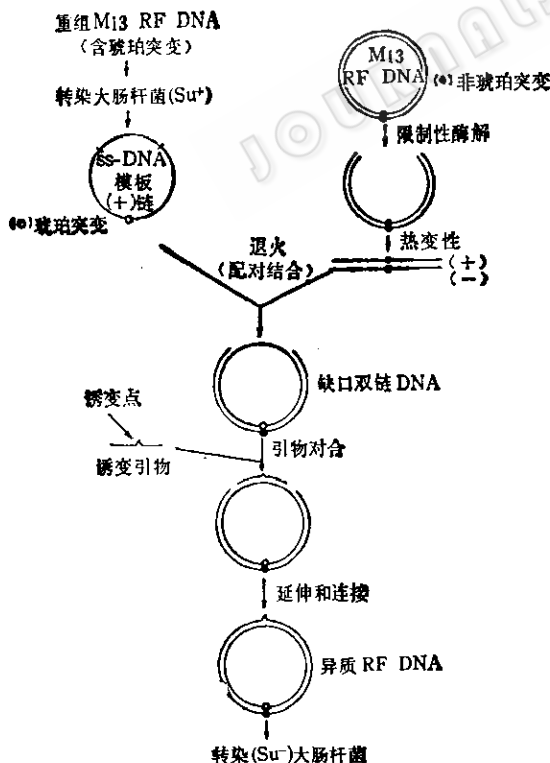


图3 缺口双链 DNA 法

型 DNA, 加热变性后与 ss-DNA 模板配对结合, 形成带缺口的双链 DNA, 其中克隆 DNA 处在单链“窗口”的位置; (3) 把诱变引物结合到上述单链“窗口”的待诱变部位, 再经 DNA 延伸、连接反应形成异质双链 DNA; (4) 最后用没有琥珀抑制基因 (Su^-) 的宿主选择突变型子代。

Goodson 等^[17]最近用此法定位诱变一种抗癌药物, 重组抗白细胞素 (interleukin 2) 的糖基化位点, 以带有两个琥珀突变的重组 M13mp9 的 ss-DNA 为模板, 用不含琥珀突变的 M13mp19 RF DNA 制备缺口双链 DNA。最后转化时用了两株 Su^- 大肠杆菌。异质双链 DNA 先转化 HB2154 (Su^- *mutL*)^[18], 使诱变顺序能以复制又不被修复; 随后再在转化的平板上铺一层上层洋菜, 内含 HB2151 (Su^-)^[19] 细胞以得到稳定的突变体, 诱变率达到 87% (21/24)。

(二) *EcoK*/*EcoB* 循环选择诱变法 (双引物法) (图 4)

EcoK 和 *EcoB* 都是 I 型限制酶, 二者的识别顺序只相差一个碱基如下:

EcoK AACNNNNNNNGTGC

EcoB TGANNNNNNNNTGCT.

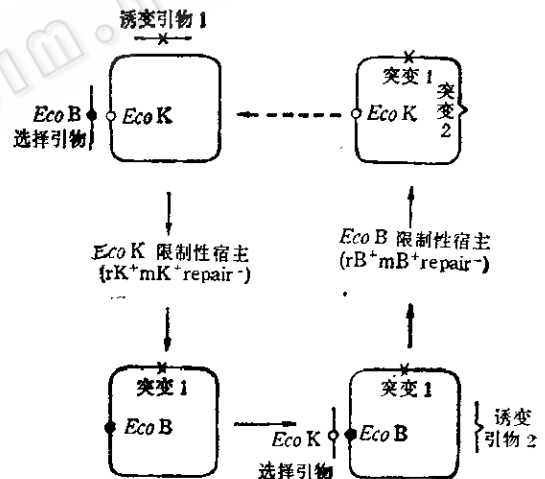


图4 *EcoK*/*EcoB* 循环选择诱变法 (双引物法)

Carter 等^[20]利用这个特点设计了一种双引物法。两个引物中, 一个是选择性引物, 带有 *EcoB* (或 *EcoK*) 识别位点, 另一个是诱变引物, 带有诱变点。此法要求构建一个特殊的载体, 内含单个 *EcoK* 识别位点。Carter 等把含有 *EcoK* 位点的短片段插到 M13mp19 载体上取名 M13K19, 这种载体在 *EcoK* 限制性宿主 ($rK^+ mK^+$, repair⁻) 中是不能生存的。当两个引物同时结合到模板上, 新合成的诱变链中的选择性位点从 *EcoK* 变成 *EcoB*, 也就是说, 突变型子代可以在 *EcoK* 限制性宿主中生长, 而野生型则不能, 单点诱变效率和琥珀突变选择效率相似, 都在 60—70% 左右。此法的

另一个优点是新突变体中含有 *EcoB* 位点, 还可以进行第二轮诱变, 用 *EcoB* 限制性宿主选择第二轮突变体(图4)。这个性质对需要在同一片段内进行多次定位诱变的研究是很有用的。

(三) 含尿嘧啶单链 DNA 模板法^[9-11]

所谓含尿嘧啶单链 DNA 模板就是在 DNA 模板中的少数胸腺嘧啶残基(T)为尿嘧啶(U)所取代。在大肠杆菌 *dut⁻ ung⁻* 菌株中可以制备这种模板。在 *dut⁻* 突变体中没有脱氧尿苷酸酶, 细胞内的 dUTP 浓度高, 可以和 dTTP 竞争并参入到 DNA 中; *ung⁻* 突变体缺少尿嘧啶-N-糖基化酶, 这种酶的作用是降解 DNA 中的尿嘧啶。因此在 *dut⁻ ung⁻* 突变体中, dUTP 可以取代 dTTP 参入 DNA 的一些部位, 并且不被清除掉。任何常用的 M13 类载体都可以在 *dut⁻ ung⁻* 宿主中培养制备出含尿嘧啶的 DNA 模板, 这种模板中通常每个基因组含有 20—30 个尿嘧啶残基。

在实验条件下含尿嘧啶的 ss-DNA 模板和正常模板没有区别, 模板中 dUMP 和 dTMP 的编码相同, 不会造成诱变。模板中的尿嘧啶也不会妨碍 DNA 的体外合成。

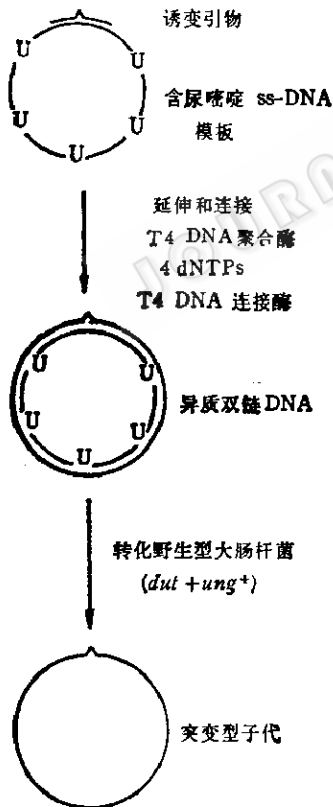


图5 含尿嘧啶单链 DNA 模板法

当诱变引物沿着含尿嘧啶模板合成新链时(图5), 底物中只加四种 dNTPs, 不含尿嘧啶, 新链合成

后可以外加尿嘧啶-N-糖基化酶除去模板链上的尿嘧啶; 较为简便的方法是把异质 RF DNA 转染野生型(*dut⁺ ung⁺*) 大肠杆菌, 由细胞内的尿嘧啶-N-糖基化酶使尿嘧啶脱落, 在模板链上形成无嘧啶(AP)位点, 这种 AP 位点是致死的损害。由于模板链是没有生物活性的, 转染子代主要来自诱变的互补链, 单点诱变效率可达 50—80%。此法除了需要一株 *dut⁻ ung⁻* 菌株制备含尿嘧啶模板外, 别无其他要求, 是一种较好的定位诱变方法。Bio-Rad 公司根据此法配成 Muta-Gene 体外诱变试剂盒^[12]。

Venkitaraman^[13] 曾用此法在克隆基因中插入 90 个碱基的 DNA 片段, 引物长度 120 个碱基。他用聚合连续性高的 DNA 聚合酶“Sequenase”代替 T4 DNA 聚合酶或 Klenow 片段, 结果在 148 个噬菌斑中得到 9 个突变体(6%), 而用后面两种 DNA 聚合酶的没有得到突变体。

(四) 脱氧硫代核苷酸保护法

Taylor 等^[14]报道, 有些限制酶如 *AvaI*, *AvaII*, *BanII*, *HindII*, *NciI*, *PstI* 和 *PvuI* 等不能切割含硫代磷酸的 DNA 链。如果当引物沿着 ss-DNA 模板合成新链时参入一种硫代核苷酸, 这条新链就不能为上述限制酶切开, 而不含硫代磷酸的模板链仍然可以被切开, 酶解的结果是在异质 RF DNA 的模板链上形成若干断口, 断口经核酸外切酶 III (*Exo III*) 处理, 模板链的部分或全部被消化, 而新合成的诱变链则受到保护。因此, 子代主要是由诱变链衍生的。

NciI 是其中较理想的限制酶^[15], 它的识别顺序是 $CC^{\downarrow}G^{\downarrow}GG$ 。如果在引物延伸反应中用 dCTP_s 代替 dCTP, 新合成链中所有的 *NciI* 以及 *SmaI* ($CCC^{\downarrow}GGG$) 等切点都可以受到保护。[M13 类载体的多克隆位点 (multiple cloning site) 上都有 *SmaI* 切点]。脱氧胞苷-5-[α -硫代]三磷酸 (dCTP_s) (图6) 是 dCTP 的 α 磷酸的氧为硫所取代。这种类似物在引物延伸时可以正常参入, 对碱基配对没有影响, 也不会引起诱变。

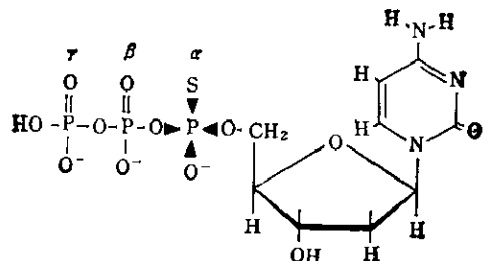


图6 脱氧胞苷-5-[α -硫代]三磷酸 (dCTP_s) 化学结构图

其他限制酶如 *Ava*I ($C\downarrow$ PyCGPuG) 和 *Ava*II ($G\downarrow$ C A CC) 等也可以选用^[16]。但是 *Ava*I, *Ava*II 切点位于 M13 多克隆位点的上游, 用核酸外切酶消化时应采用 T7 核酸外切酶 ($5'\rightarrow 3'$ 外切酶) 代替 *Exo* III ($3'\rightarrow 5'$ 外切酶)。此外, 在使用 *Ava*II 酶时还应改用 dGTP α S。上述各种组合的定点诱变效率都能达到 80—95%, 最近 Olsen 和 Eckstein^[17] 还应用此法在双链质粒 DNA 上直接进行体外定位诱变, 定点诱变率达到 70—80%。

本方法诱变效率高, 除了脱氧硫代核苷酸价格稍高外, 其他操作步骤均较简便可靠, 也是一种值得推荐的方法。Amersham 公司根据 *Nci*I/*Exo* III 系统^[17] 配制的体外诱变试剂盒已为许多实验室采纳。

(五) 聚合酶链反应 (PCR) 定位诱变法

PCR 是一项在体外大量扩增 DNA 片段的新技术。由于其方法简单、快速, 近年来发展极快。PCR 技术略加改动也适用于 DNA 片段的定位诱变, 诱变点设计在引物内(置换、缺失或插入), 也可以在引物的 5' 端添加限制酶识别顺序, 便于以后的再克隆等。由于寡核苷酸引物通常是设在模板 DNA 的两端, 若要把诱变点安排到 DNA 片段的任何部位, 必须对 PCR 扩增技术进行一些调整。现将最近发表的 PCR 定位诱变方法归纳成两类, 分别介绍如下:

1. DNA 片段的定位诱变^[18-24]

(1) 拼接法^[18,19]: 本方法采用四个引物和三组 PCR 反应, 手续略为繁琐, 但设计颇具特色。诱变分两步进行(图 7)。第一步引入诱变点: 用两组引物 A + B

和 C + D 分别在同一 DNA 模板的不同区段进行两组 PCR 反应。第一组的引物 B 和第二组的 D 含有相对应的诱变点, 并且是相互重叠的, 在两组 PCR 产物之间产生一段重叠区。第二步调整诱变点位置: 把上述两组 PCR 产物提纯后混合、变性、退火, 两组 PCR 产物能以两种方式在重叠区配对拼接成两组局部双链的 DNA 分子。但是其中只有 3'-OH 端缩进的那一组可以由 DNA 聚合酶聚合成完整的双链分子。这种拼接的频率可能不高, 但已为第三组 PCR 提供了足够的模板。

第三组 PCR 是用最远的两个引物 (C + A), 以拼接的诱变 DNA 分子为模板, 进行 15 循环 PCR 反应, 诱变率可达 100%。扩增产物可直接用于测序。

拼接法可用于单点置换、插入或缺失突变, 扩增后的差错率约为 0.026%^[19]。如果在引物 A 和 C 的 5' 端添加限制酶识别顺序, 扩增后的 DNA 片段更加便于亚克隆。此法还适用于构建嵌合基因, 以及将四个或更多 DNA 片段拼接, 构建融合蛋白等。

(2) 大引物法^[24]: 大引物法是简化的拼接法, 总共只要用三个引物和两组 PCR 反应(图 8)。

第一组 PCR 构建大引物: 用 B 和 C 两个引物, 其中引物 B 含诱变点, 以待诱变的 DNA 片段为模板, 进行 PCR 15 个循环。然后将产物提纯。这段 B—C 扩增产物就是带诱变点的大引物。

第二组 PCR 反应仍以原来的 DNA 片段为模板, 用 B—C 大引物和另一端的引物 A 进行 PCR 15 个循环。诱变产物可直接用于测序。

本实验中, 每段 DNA 只扩增 15 个循环。第一组 PCR 15 个循环构建大引物, 第二组 PCR 扩增 DNA

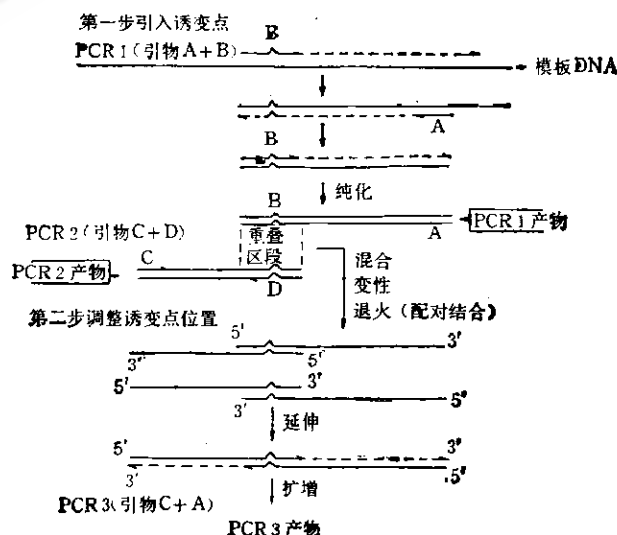


图 7 DNA 片段拼接法

两组引物 A + B 和 C + D; 其中 B 和 D 含有相对应的诱变点, 箭头表示合成方向

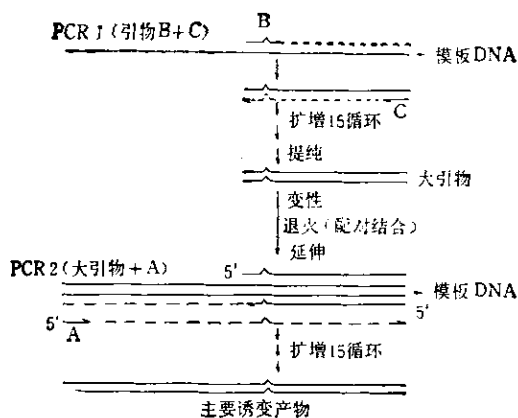


图8 PCR 大引物法

片段的其余部分。而在前述拼接法中,各段 DNA 都要经过两轮扩增。因此大引物法的总差错率会比拼接法低一些。

从大引物法还可以看到,任何 DNA 短片段均可代替寡核苷酸充当 PCR 的扩增引物,这点在实际应用上也是很有意义的。

2. 环状 DNA 的定位诱变: 环状 DNA 定位诱变法就是直接用重组质粒 DNA 为模板进行定位诱变,从而简化了诱变后还要把诱变的 DNA 片段再克隆到

载体上的手续。

(1) 重组环状 DNA 法 (RC-PCR)^[2,3]: 本方法的原理和拼接法类似,需要用四个引物和两组 PCR 反应(图 9)。

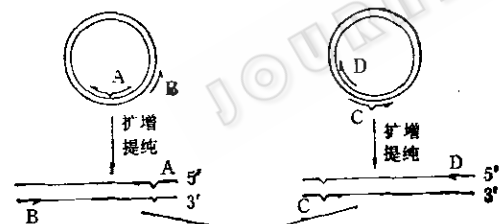
a. 引入诱变点: 设计两组引物 A + B 和 C + D, 分别和环状双链 DNA 模板进行两组 PCR 反应。引物 A 和 C 分别含对应的诱变点,并且相互是重叠的,两组引物的 5' 端都分别在模板上错开一至数个碱基,使两组 PCR 产物的“断开点”分别处在环状 DNA 模板的不同位置[如图 9(2)]。两组 PCR 反应各进行 15 个循环。

b. 重组环化: 上述两组 PCR 扩增产物经过提纯、混合、加热变性和退火后,可以重组环化成两组各带两个缺口的重组环化 DNA。这两种重组环化的 DNA 分子形成的比例都不很高,但产物不需要提纯或其他酶处理,就可以直接转化大肠杆菌。反应混合物中原来的线型 DNA 分子不会干扰转化,诱变效率 83—100%。从 PCR 扩增、重组环化到转化大肠杆菌可以在一天内完成。

重组环状 DNA 法和 DNA 片段拼接法类似,可以用于构建嵌合质粒,基因融合,或把克隆片段从一个载体定向地转移到另一个载体等。

本方法不需要限制酶和连接酶,对于拼接那些没有适当酶切位点的 DNA 片段特别有用。此外,如果把所有引物的 5' 端预先磷酸化,RC-PCR 法还可用于在体外生产大量闭环 DNA。

(1) 引入诱变点



(2) 重组环化

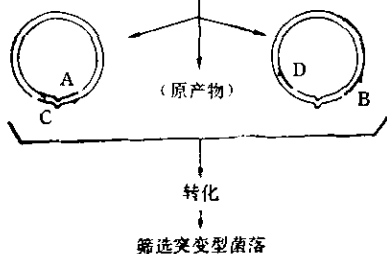


图9 PCR 重组环状 DNA 法 (RC-PCR)

两组引物 A + B 和 C + D, 箭头表示合成方向, 引物 A 和 C 带对应的诱变点, PCR 扩增产物中用粗线标出原引物的位置

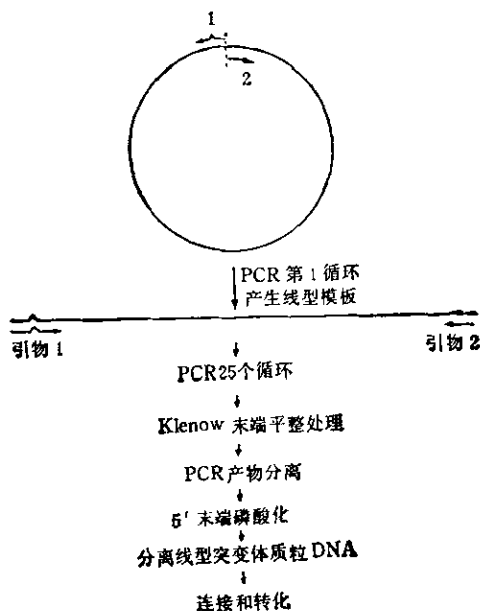


图10 反向 PCR 法示意图

背靠背的引物 1 和 2 分别处于相对应的 DNA 链上, 引物 1 含诱变点

(2) 反向 PCR 法^[26]: 本方法以完整的重组环状质粒 DNA 为模板, 用一对背靠背的引物, 它们的 5' 端紧挨着, 3' 端分别沿着相对应的环状 DNA 模板朝着相反的方向延伸(图 10)。引物可根据诱变点的位置设计在环状 DNA 的任何位置。质粒 DNA 无需特殊提纯, 小样粗提的和开环质粒 DNA 都可以用作模板。

PCR 开始的第一个循环就形成一条一端带诱变点的线型质粒分子。以后, 两个引物分别沿着线型 DNA 相对应的链延伸并连续扩增 15 个循环。扩增完成后, PCR 产物还要加 Klenow 片段进行末端平整处理, 然后提纯和进行 5' 端磷酸化, T4 DNA 连接酶平末端连接, 最后转化大肠杆菌, 所得的转化子 80% 是带有诱变点的。

反向 PCR 法是以完整的重组质粒 DNA 为模板, DNA 长度通常在 3kb 以上, 因此引物延伸的时间也需要长一些。但是聚合时间延长, Taq DNA 聚合酶往往会在扩增产物的 3' 末端添加一个额外的(不属于模板的)核苷酸。用 Klenow 片段处理就是为了除掉这个额外的碱基^[26]。另外, 反向 PCR 法对引物的纯度要求也高。用核苷酸自动合成仪合成的寡核苷酸通常是长度不一的混合物。如果引物未经纯化, 有些扩增产物的 5' 端会短少一至数个核苷酸, 环化后就在连接点造成一些缺失。Hemsley 等^[24]在测序时发现, 在 29 个环化产物的连接点上竟有 4 个样品缺失 1 至 8 个碱基对(占 14%)。因此必须使用高度纯化的寡核苷酸引物。而 RC-PCR 法对引物纯度就不要这样严格, 同时也不需要这样复杂的后处理。

3. PCR 法在定位诱变中的忠实性问题: Taq DNA 聚合酶的合成效率在最佳条件下平均每分钟合成 1 000—4 000 碱基。扩增 10 kb 的 DNA 片段是完全可以胜任的。但是在扩增完整的质粒 DNA 和长的 DNA 片段时, 效率要低些, 需要适当增加延伸时间。

由于 Taq DNA 聚合酶中没有校正读码 (proof-reading) 的 3'→5' 外切酶活性, 扩增时间愈长, 聚合出现差错的几率也愈大。不同文献报道的差错率相差很大。Saiki 等^[27]报道的差错估计值是 1/5000 bp, 或长度 239 bp 的 DNA 在 30 个循环中的差错率为 0.25%。但在改进的条件下^[28] (Mg^{++} 和 dNTPs 浓度适当调低), 每聚合 75,000 个核苷酸才出现一个差错, 或 30 循环差错率少于 0.016%。这样低的差错率从总体来看, 影响不会很大, 因为绝大多数扩增的分子是没有差错的。但是, 如果 PCR 产物是准备用于分子克隆的, 而克隆的那个分子哪怕有一点差错, 问题就可能很严重。这点正是 PCR 技术用于构建重组 DNA 的一大缺陷。为了避免出现差错, 在分子克隆时就需要同时对两或三个克隆进行测序和比较, 以确定它们的顺序是完全一致的。

在定位诱变中, 一般 PCR 只用 15 个循环, 扩增

造成的差错将不会很多。当然扩增的 DNA 片段愈长, 差错出现的几率也会愈高。例如用完整的重组环状质粒 DNA 进行定位诱变, 差错的几率就比扩增短的 DNA 片段高。

Hemsley 等^[26]在以完整质粒 DNA (约 3 kb) 为模板进行的反向 PCR 定位诱变实验中发现, 未经 Klenow 片段平整处理的 PCR 扩增产物, 环化后的连接部位额外多了一个核苷酸的占 73%, 外加的碱基对主要是 A:T。这很可能是 PCR 延伸时间用得较长 (12 分钟), 由 Taq DNA 聚合酶加在 DNA 3' 末端引起的。同样, Jones 和 Howard^[25]在用 RC-PCR 法进行定向亚克隆实验中也发现, 在三个亚克隆的连接区顺序中, 也有一个连接区的 C—C 间多了一个 T。由此可见, 在 DNA 3' 末端外加一个碱基的现象, 在扩增较长 DNA 片段时是相当普遍的。这对 PCR 扩增反应, 尤其是扩增后需要重新环化的实验中, 是一个值得重视的问题。或许在所有 PCR 扩增反应终了后都增加一项 Klenow 片段平整处理, 除掉这些额外的碱基是很有必要的。

参 考 文 献

1. Zoller M J & M Smith: *Methods in Enzymology*, 100: 468—500, 1983.
2. Osinga K A et al.: *Nucl. Acids Res.*, 11: 8595—8605, 1983.
3. Grosse F & G Krauss: *Eur. J. Biochem.*, 141: 109—114, 1984.
4. Kramer B et al.: *Cell*, 38: 879—889, 1984.
5. Chen N Y & H Paulus: *J. Biol. Chem.*, 263: 9526—9532, 1988.
6. Kramer W et al.: *Nucl. Acids Res.*, 12: 9441—9456, 1984.
7. Goodson R J et al.: *Bio/Tech.*, 8: 343—346, 1990.
8. Carter P et al.: *Nucl. acids Res.*, 13: 4431—4443, 1985.
9. Kunkel T A: *Proc. Natl. Acad. Sci. U. S. A.*, 82: 488—492, 1985.
10. Kunkel T A: In "Current Protocols in Molecular Biology", 8. 1. 1—6, John Wiley & Sons, 1987.
11. Kunkel T A et al.: *Methods in Enzymology*, 154: 367, 1987.
12. Geisselsoder J et al.: *BioTechniques*, 5: 786—791, 1987.
13. Venkitaraman A R: *Nucl. Acids Res.*, 17: 3314, 1989.
14. Taylor J W et al.: *Nucl. Acids Res.*, 13: 8749—8764, 1985.
15. Taylor J W et al.: *Nucl. Acids Res.*, 13: 8765—8785, 1985.
16. Sayers J R et al.: *Nucl. Acids Res.*, 16: 791—802, 1988.
17. Nakamaye K & F. Eckstein: *Nucl. Acids Res.*, 14: 9679—9698, 1986.
18. Higuchi R et al.: *Nucl. Acids Res.*, 16: 7351—7367,

(下转第 64 页)