



Sanger 链中止法的新进展——以 M13 为克隆的运载体进行 DNA 序列分析 (续)

孙 晋 武

(中国科学院微生物研究所, 北京)

三. M13mpX 复制形式(RF)的制备 (图 1-3)

虽然 M13mpX 是单链的噬菌体,一旦它感染到寄主细胞后,便成为双链超螺旋的复制形式(图 3)。这样既便于限制性内切酶在其 MCS 处切开,也能通过连接酶的作用把待测 DNA 片段接进去,构成一个重组体。

通常采用的制备 RF 的方法是: 将 M13mpX 按一定比例接入 *E. coli* 中,用蛋白胨-酵母膏培养基,37℃ 振荡通气培养数小时,离心得细胞沉淀,用溶菌酶和 Sarcosyl 破碎细胞,离心得上清液,酚抽提去蛋白,乙醇沉淀 DNA,再用蔗糖密度梯度离心,或用 RPC-5 柱层析法便可制得 RF。

四、重组体的制备(图 1-3—7)

这是一个将待测双链 DNA 克隆到上述的 M13mpX 复制形式中去的过程。在此过程中应考虑如下一些问题:

1. 由于 Sanger 法一次只能读出 200—400 个核苷酸的排列顺序,虽然待测 DNA 片段克隆到 M13mpX RF 中去时,会出现机率相等而方向相反的两种连接方式,但核苷酸数目超过 600 的片段仍旧难于一次读出来。另外,一般说来,大于 1000 个碱基对的 DNA 片段,在插入到 M13mpX 之后,通常并不很稳定^[13],因此必须事先用限制性内切酶将其切成大小合适(约为 300—500 个碱基对)的小片段。有时待测 DNA 片段虽然不大,但为了验证某些疑难之处,也需要对其再行切割,分别克隆。为此目

的,有两种可采用的方针:一是如果在你对待测 DNA 有个基本的了解,而该片段又不很大的情况下,你可选用合适的内切酶将其切成 2—3 个片段,经序列分析,根据重叠部分很容易将它们连接起来。另一种是所谓的“榴弹散射法”(Shotgun)^[15,13],就是你对样品的情况并不清楚,而该片段可能又很大时,则可通过一些内切酶的任意切割,得到许多小片段,进行克隆。然后分别进行 DNA 序列分析,最后则需应用电子计算机,根据重叠交叉和一些可能性的分析,把它们组合起来。根据文献报道,如果在一切准备工作就绪的情况下,有时用此方法,一天就可以测得 1000 个核苷酸的序列^[13]。

2. 在实际工作中,经常是将待测 DNA 片段克隆进 M13mpX 中,先对这段 DNA 进行序列测定,然后再用一、二种限制性内切酶,将测过的一小段 DNA 去掉,用连接酶把剩余部分连接起来,便可继续测定下去。不过在此过程中要注意,选用的内切酶一定要合适,切口不能太多,不然会使问题变得复杂化;另外也不能在 M13 上有切口,表 3 中列出了在 M13 中不存在的内切酶切点。我们可以在此表中加以选择。

3. 一般说需要有 1μg 以上的经纯化的待测 DNA 样品。如果待测 DNA 片段较大,所需样品量相应也有所增加。常规的提纯方法有聚丙烯酰胺和琼脂糖凝胶电泳法:样品经电泳分离,根据溴化乙锭所显示的位置,可采用一系列

表 3 在 M 13中不存在的限制性内切酶的识别和切割位点

<i>Acc</i> I B (GTCTAC)	<i>Acc</i> I C (GTATAC)	<i>Acc</i> I D (GTAGAC)
<i>Asu</i> II (TTCGAA)	<i>Ava</i> II B (GGTCC)	<i>Ava</i> III (ATGCAT)
<i>Avr</i> II (CCTAGG)	<i>Bcl</i> I (TGATCA)	<i>Bst</i> E II A (GGTAACC)
<i>Bst</i> E II B (GGTCACC)	<i>Bst</i> E II C (GGTGACC)	<i>Bst</i> E II D (GGTTACC)
<i>Fnu</i> 4H I B (GCCGC)	<i>Hae</i> I A (AGGCCT)	<i>Hgi</i> A I B (GAGCTC)
<i>Hgi</i> A I D (GAGCAC)	<i>Hind</i> III (AAGCTT)	<i>Hpa</i> I (GTTAAC)
<i>Kpn</i> I (GGTACC)	<i>Mst</i> I (TGCGCA)	<i>Sac</i> I (GAGCTC)
<i>Sac</i> II (CCGCGG)	<i>Sma</i> I (CCCGGG)	<i>Sph</i> I (GCATGC)
<i>Xba</i> I (TCTAGA)	<i>Xho</i> I (CTCGAG)	<i>Xma</i> III (CGGCCG)

不同的方法,将 DNA 分离提纯出来^[14,15];或用低熔点琼脂糖凝胶电泳法,可以直接得到。另外也可采用 RPC-5 柱(逆相层析法)^[16],此法较为简便快速,柱可反复使用,样品回收率也较高,所得样品稍加处理便可直接应用,这对较大的样品制备是相当方便的。

表 4 在 M13mp7 DNA 中潜在的一些克隆位点

M13 mp7 限制性内切酶位点	可以克隆的片段
<i>Bam</i> HI G ⁺ GATCC	<i>Bam</i> HI G ⁺ GATCC <i>Bgl</i> II A ⁺ GATCT <i>Bcl</i> I T ⁺ GATCA <i>Sau</i> 3A N ⁺ GATCN <i>Xho</i> II Pu ⁺ GATCPy
<i>Eco</i> RI G ⁺ AATTC	<i>Eco</i> RI G ⁺ AATTC <i>Eco</i> RI* N ⁺ AATTN
<i>Pst</i> I CTGCA ⁺ G 以 G 结尾	<i>Pst</i> I CTGCA ⁺ G 任何以 C 结尾的片段
<i>Sal</i> I G ⁺ TCGAC	<i>Sal</i> I G ⁺ TCGAC <i>Xho</i> I C ⁺ TCGAG
<i>Acc</i> I GT ⁺ CGAC	<i>Acc</i> I GT ⁺ CGAT <i>Cla</i> I AT ⁺ CGAT <i>Hpa</i> II NC ⁺ CGN <i>Taq</i> I NT ⁺ CGAN
<i>Hinc</i> II GTC ⁺ GAC	对任何片段都有较低的效果 如: <i>Bal</i> I TGG ⁺ CCA <i>Hae</i> III NGG ⁺ CCN <i>Itha</i> I NNGCG ⁺ CNN NNN NN NNN <i>Bal</i> 31 ⁺

4. 为了要把纯化到的待测 DNA 片段克隆到 M13mpX 复制形式的 MCS 中去,首先要用限制性内切酶将 MCS 部位切开。待测 DNA 片段和切开后的 M13 mpX 复制形式必须具有互补的粘性末端,两者才能在连接酶的作用下得以重组。M13mp7, M13mp8 和 M13mp9 提供了多种限制性内切酶的识别和切割位点,以便于得到与待测 DNA 片段取得互相匹配的粘性末端。如待测 DNA 具有 GAATTC 这

样的末端,则我们可选用 *Eco* RI 去切割 M13 mpX。但有时情况并非如此简单。例如,有时我们的待测 DNA 片段是个平头末端,那末我们就要借助于一小段人工合成的所谓“分子连接物”(molecular linker),通过磷酸化而接到此待测 DNA 上去。如 *Eco* RI 连接物引入了 *Eco* RI 的切割位点,用此内切酶切割便可得到相应的粘性末端。还有一种情况是待测 DNA 片段的末端虽然并非平头,但无法与 M13mpX 的切口匹配,这就必须对此末端进行改造。可先用 *E. coli* polymerase I 将其末端切平,然后再采用上述的接上分子连接物的办法。

事实上,在 M13mpX 的 MCS 上,与各种限制性内切酶作用得到的粘性末端所对应的可以克隆的片段并不止一种,表 4 是以 M13mp7 为例显示这一对应关系。在设计实验时,也可参考该表加以考虑。

五、单链 DNA 模板的制备 (图 1-7—16)

上述 DNA 的重组过程只以一定的概率发生,而且其数量是极小的,为此需要将重组体挑出来进行扩大培养,以获得足够量的、高纯度的单链 DNA 模板。

首先是将上述 M13mpX 重组体转染到新鲜的、处于对数生长期的、事先经 Ca^{++} 处理过以增加膜通透性的感染细胞,如 *E. coli* JM 103 上去。为增加转染效率,可将混合液从 0℃ 突然升温到 42℃,放置 2—3 分钟,进行所谓的热休克 (heat shock),然后再涂布于含有 IPTG, X-gal 的 YT 培养基(蛋白胨和酵母膏)的平板上,37℃ 培养 6—12 小时后,便可见到在大量的蓝色噬菌斑中有个别的白色斑点存在,这

是因为待测 DNA 片段插入到 M13mpX 中后，破坏了 *Lac* 的表达，因而不能产生 β -半乳糖苷酶。利用这一明显的颜色反应，我们很容易地将重组过后的 M13mpX 挑选出来。再转接到 YT 液体培养基中，37℃ 扩大培养，待菌体长到对数生长后期，离心除去寄主细胞，（此寄主细胞含有插入了待测 DNA 片段的 M13mpX 复制形式，可以转接到新鲜培养基中继续培养，便可不断释放出单链噬菌体来以供序列分析之用，如图 1 中从 14 可回到 11 去），而包含有待测 DNA 序列的单链噬菌体便存在于上清液中。在上清液中加入聚乙二醇-6000，使最终浓度为 2—4%，另外加入最终浓度为 0.5 M 的 NaCl，放置一段时间，噬菌体便会沉淀下来。离心得到噬菌体沉淀，再用酚抽提法去除噬菌体的蛋白部分，最后用乙醇把单链 DNA 沉淀下来，经含有溴化乙锭的琼脂糖凝胶电泳鉴定，如果纯度合格，便可用于随后的序列分析中去。

六、通用引物（图 1-17）

Sanger 链中止法除了要有一单链 DNA 模板外，还必需要有一段能与该模板互补的引物，这样在 DNA 多聚酶的作用下，以此单链 DNA 为模板，从该引物的 3' 末端开始，便可以四种 dNTP 为材料合成新的与模板互补的 DNA 链了。根据前面的介绍，现在有了统一的运载工具 M13mpX，对于它的 DNA 序列，人们已经完全搞清楚了^[3]。现在，我们只要把一段待测 DNA 插入到此 M13mpX 的某一特定区域中去，不论你插入的是什麼，都不影响其余部分的 DNA 序列。因此只要有一段能与离此插入部位的 3' 末端靠近的 *Lac z* 基因上一小段 DNA 序列互补的小片段，就可适合于各种 DNA 插入片段的测定工作了。这样的小片段称之为“通用引物”（Universal primer）。一般说来，通用引物除应满足上述的基本要求外，还应当保证与模板能进行退火而构成稳定的双链结构的前提下，尽可能地短一些。这样，可从凝胶电泳的结果读出的序列数也就会相应地得到增加。根据这一原则，Messing 等人首先由酶解

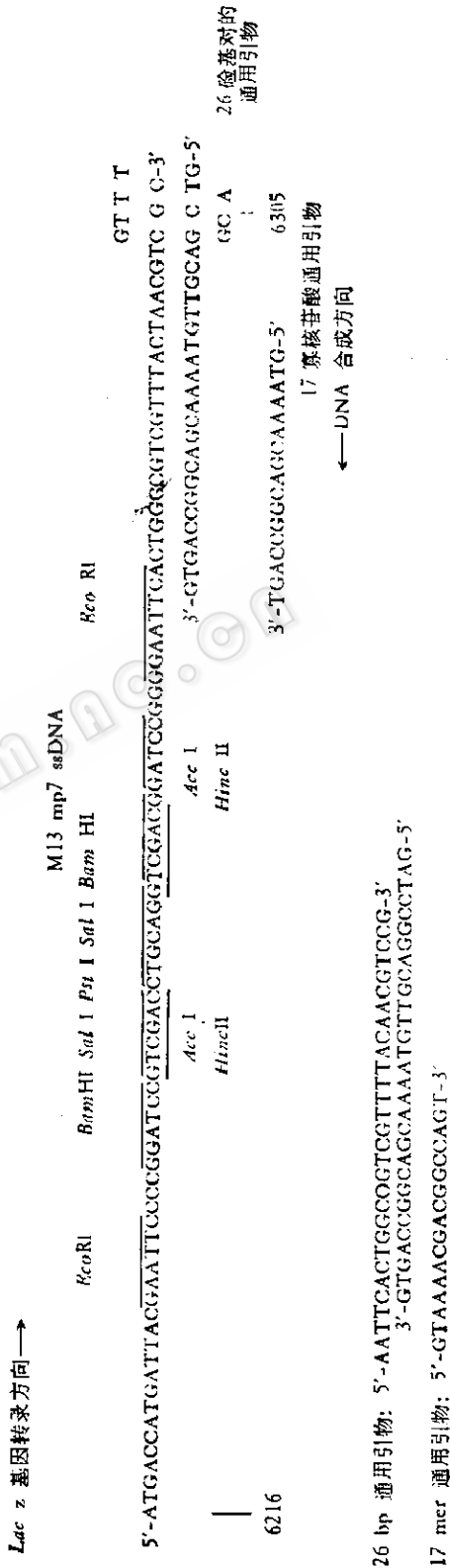


图 5 两种通用引物的 DNA 序列及其与 M13mp7 的互补关系

片段获得了 96 个碱基对的通用引物^[7], 不久 Anderson 等又制得了 26 bp 的片段^[8]。最近 Duckworth 作了进一步的改进, 用化学合成法制得了只有 17 个寡核苷酸组成的单链引物^[9], 后两种通用引物的 DNA 序列以及它们与 M13mp7 之间的互补关系见图 5。

七、DNA 序列分析 (图 1-18—22)

在上述的所有准备工作都已完成后, 便可采用 Sanger 的链中止法进行序列分析。简单说来, 整个过程可分如下几步:

1. 退火: 将单链模板与通用引物混合, 置于 100℃ 处理 3 分钟。然后转移到 67℃ 水浴中放置 5 分钟, 让其自然冷却。此时通用引物的 DNA 链便会与模板的特定区域经碱基配对而构成一小段双链, 新的 DNA 链便从引物的 3' 末端开始合成。

2. 链的合成: 反应混合物分成四组。组间的差异在于每组加入 A、T、G、C 中的一种 2'、3' 双脱氧核苷酸以中断链的合成。各组其它成份大致相同, 它们包括有: 一定比例的四种脱氧核苷酸, 其中一种应当是带有同位素标记的; 经退火的单链 DNA 模板和引物的混合物; DNA 多聚酶(采用 *E. coli* DNA polymerase I, large fragment)。反应在室温下进行 15 分钟, 然后再追加(chase)一定量的非放射性的脱氧核苷酸继续反应 15 分钟。如开始用的是 (α -³²P) dATP, 此时则应追加非放射性的 dATP。目的是使链的延伸不因开始加入的放射性 dATP 不足而过早地中断, 同时也为了使放射性在合成的长、短链中分布比较均匀。

最后, 反应以加入能使 DNA 多聚酶失活和具有解链作用的甲酰胺而中止。

3. 凝胶电泳: 一般采用 8% 或 6%、内含 7M 脲素的聚丙烯酰胺凝胶电泳。为保证样品以单链形式存在, 在加样前, 样品要在 95℃ 加热 3 分钟。加样时, 用滴管把从凝胶中扩散出来的脲素冲干净。电泳在 1200—1500 V 进行数小时。电泳过程中产生的热使凝胶表面维持在 50—60℃, 有助于保持样品以单链变性状态存在。示踪染料是溴酚蓝和 Xylene Cyanol FF

的混合物。在 8% 和 6% 的凝胶中, 溴酚蓝迁移

表 5 自显影结果中常见问题的分析

现象	原因	解决办法
1. 自显影是白色的一片, 看不到什么区带, 只有在靠近溴酚蓝的部位有模糊的一团黑影	(1) 没有引物或没有模板 DNA (2) 引物或模板因污染而不起作用 (诸如 Agarose, 丙烯酰胺或盐等) (3) 在退火过程中引物和模板之间没有发生杂交	(1) 检查 DNA 制品是否存在 (2) 重新纯化 DNA (3) 检查退火条件是否正确, 如高盐、退火温度太低都会影响退火效果
2. 整个凝胶区带不清晰, 无法读出序列来	(1) 样品中盐含量太高, 影响到电泳效率及分辨率 (2) 凝胶或 X-底片有毛病, 像凝胶存放时间太长, X-底片型号不对, 或在放射性自显影过程中凝胶与底片接触不好。	(1) 在进行反应前, 从 DNA 样品中去盐 (2) 检查凝胶和底片。
3. 区带排列不够规则, 似乎有二、三种序列同时存在	引物或模板不纯, 即由于有另外的引物或模板发生了退火, 同时出现一种以上的序列	在分离模板时, 一定要保证没有杂质, DNA 一起被提纯出来。如果污染太厉害, 提高“双脱氧”与“脱氧”核苷酸之间的比例, 有时能降低背景
4. 每条区带看起来像是两条, 甚至三条重叠在一起	由于污染了外切核酸酶而使引物具有不均一的 5'-末端	检查是否存在有外切核酸酶的活力
5. 凝胶上某一区域有一条或数条区带横穿各行	(1) 单链 DNA 模板有环, DNA 多聚酶的合成作用在此受到一定的阻碍 (2) 由于污染了核酸酶的作用, 在 DNA 上可能会有缺口	(1) 有时可试用互补链来进行序列测定 (2) 防止核酸酶的污染
6. 每行从上到下区带的背景很深, 而且强度分布均匀	(1) 在引物或模板制品中污染了小的 DNA 或 RNA (特别是 tRNA), 这些小的寡核苷酸能彼此或与模板 DNA 杂乱地发生退火而得到无数转录产物 (2) 由于 DNA 上杂乱的缺口所造成 (3) 反应后的样品搁置太久, 样品有所降解	(1) 增加“双脱氧”核苷酸的浓度可降低背景, 有时也可阅读每行中最深的区带而得出待测 DNA 的序列 (2) 防止核酸酶的污染 (3) 样品反应后立即进行电泳

续 表

现 象	原 因	解决办法
7. 有些行, 区带在底部很深, 而在顶部很浅, 甚至一点也没有, 或是相反的结果	(1) 可能由于 ddNTP/dNTP 的比例不合适: 此比例大, 链的中止过早, 因此在凝胶底部区带很深; 比例小, 则结果相反 (2) 放射性 dNTP/非放射性 ddNTP 的比例不对 (如用 (^{32}P)dATP, 则为 ddATP, 余此类推); 如此比例小, 则放射性主要集中在合成的小片段上; 比例大, 结果相反。	(1) 必须根据经验确定合适的 ddNTP/dNTP 之比, 此比例根据不同的 DNA 序列和不同批号的试剂, 经常会有所不同 (2) 一定要有合适的比例。为了使放射性平均分布到长短不同的链上, 通常开始只用某种放射性的脱氧核苷酸, 反应一段时间后, 再追加非放射性的这种脱氧核苷酸, 使因原料不够中断过早的链, 得以延伸下去
8. 有时在凝胶的某一区域, 区带间距变小, 甚至完全挤在一起	这是在凝胶电泳过程中, 长的单链 DNA 形成二级结构的环, 这种结构会增加凝胶中寡核苷酸的迁移率, 而使几个区带压缩到一起	(1) 采用更为变性的条件进行电泳, 如升高温度 (2) 另外可试用互补链来进行序列分析, 每条链形成环的位置通常是不同的
9. 只在靠近引物处有一小段序列, 上面是空白的一片	DNA 多聚酶活力不高	检查 DNA 多聚酶活力
10. 在凝胶的上端有大量放射性物质没有进入胶内	反应后的样品解链不好, DNA 成双链形式存在, 分子大, 进不了胶	(1) 加样前一定要将样品重新加热 (2) 采用新配制的甲酰胺 (3) 电泳时的温度应当保证

的位置分别相当于 20 和 30 个寡核苷酸链所迁移的位置; 而 Xylene Cyanol FF 则分别指示 80 和 100 个碱基链的位置。利用这一点, 很容易掌握电泳的进程。有时为了在一次电泳中能读出更多的 DNA 序列, 在电泳进行一段时间后, 可在旁边加上另一组同样的样品, 继续进行电泳。电泳完毕, 将凝胶弄干, 放射自显影后便可直接进行阅读了。

八、一些容易出现的问题

DNA 序列分析从培养 *E. Coli* 和 M13, 到克隆及最后进行反应, 并从凝胶放射自显影图谱上读出结果, 是一个较长的过程。其中任何一点疏忽, 诸如微生物的污染、DNA ase 的降解作用, 试剂纯度不够或同位素材料不够新鲜等都会造成结果不够理想, 甚至全盘失败。因此最好是在每进行一步就进行适当的检查, 以免最后得不到结果, 仍不清楚问题出在什么地方。下面就自显影结果中出现的一些常见问题列表加以分析 (表 5)。

按: 此稿完成于 1982 年初, 其后, 国内外学者对此方法又作出了许多新的改进和发展, 我国学者洪国藩同志在英国剑桥大学所作出的非随机沉淀法和反向序列测定引物法便是其中较为突出的。由于作者当时尚在国外, 来不及补充, 如读者需要, 请参阅文献 [17], [18]。

参 考 文 献

- [1] Sanger, F. et al.: *Proc. Natl. Acad. Sci. USA*, 74: 5463, 1977.
- [2] Maxam, A. and N. Gilbert. *Proc. Natl. Acad. Sci. USA*, 74: 560, 1977.
- [3] Van Wezenbeek, P. et al.: *Gene*, 11: 129, 1980.
- [4] Sanger, F. et al.: *J. Mol. Biology*, 143: 161, 1980.
- [5] Messing, J. et al.: *Nucleic Acids Res.*, 9: 309, 1981.
- [6] Schreier, P. H. and R. Cortese: *J. Mol. Biology*, 129: 169—172, 1979.
- [7] Heidecker, G. et al.: *Gene*, 10: 69—73, 1980.
- [8] Anderson, S. et al.: *Nucleic Acid Res.*, 8: 1731—1743, 1980.
- [9] Duckworth, M. L. et al.: *Nucleic Acids Res.*, 9: 1691—1706, 1981.
- [10] Dehardt, D. T. et al.: *Single-Stranded DNA Phages*, Cold Spring Harbour Laboratory. New

(下转第 264 页)

(上接第 276 页)

York, 1978.

- [11] Messing, J. et al.: *Proc. Natl. Acad. Sci. USA*, 74: 3642, 1977.
- [12] Gronenborn, B. and J. Messing: *Nature*, 272: 375, 1978.
- [13] Anderson, S.: *Nucleic Acids Res.*, 9(13): 3015—3027, 1981.
- [14] Southern, E.: *Methods in Enzymology*, 68: 152—176, 1979.

- [15] Yang, R. C. A. et al.: *Methods in Enzymology*, 68: 176—182, 1979.
- [16] Wells, R. D. et al.: *Methods in Enzymology*, 65: 327—347, 1980.
- [17] Hong Guofan. *J. Mol. Biol.*, 158: 539—549, 1982.
- [18] Hong Guofan: *Bioscience Reports*, 1: 243—252, 1981.