

# 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

孙 晋 武

(中国科学院微生物研究所, 北京)

自从 1977 年 Sanger 的链中止法<sup>[1]</sup>和 Maxam-Gilbert 的化学降解法<sup>[2]</sup>问世以来, DNA 序列分析工作出现了一个飞跃。这两种方法各有千秋, 相互补充, 已成功地解决了不少长达数千对核苷酸的序列问题。如本文中所述的噬菌体 M13 的 DNA 序列, 就是综合了这两种方法所得到的结果<sup>[3]</sup>。Sanger 的链中止法虽然较为简便、快速和准确, 但是由于这一方法需要单链 DNA 作为模板, 以及每 200—300 个碱基就需要有一个专门与其 3'-末端互补的引物, 这对于一般的实验室来讲, 是相当困难的, 因而限制了这一方法的广泛使用。除英国和世界上少数一些有条件的实验室外, 大多数实验室仍旧乐于采用较为费事的 Maxam-Gilbert 法。

近年来, 人们对 Sanger 的方法作了一些改

进, 采用了经改造的噬菌体 M13 作为克隆运载体, 可以方便地得到单链 DNA 模板<sup>[4-6]</sup>, 另外又合成了一些适用于这种模板的通用引物<sup>[7-9]</sup>, 从而使这一方法易于被广大实验室采纳。

本法的基本原理是: M13 是一种对雄性 *E. coli* 专一的单链 DNA 噬菌体。Messing 等人已经用遗传工程的手段, 把乳糖操纵子 (*Lac*) 的一部分插入到该噬菌体中, 然后再把一些含有某些限制性内切酶识别和切割位点的核苷酸片段接进 *Lac* 中, 从而将 M13 改造成可以进行克隆的运载体。如果把一段未知序列的 DNA 插入到上述的核苷酸片段的一定部

---

本文在整理过程中, 得到了美国耶鲁大学分子生物物理和生物化学系 Konigsberg, W. H. 实验室, 特别是 Lin, T. C. 博士的热情指导和帮助。特此感谢。

位中去,再让其感染 *E. coli*, 这样就可以得到大量含有该待测 DNA 序列片段的单链模板。*Lac* 的插入,有助于对这种 M13 衍生物进行识别。另外,人们又设法合成了一些相邻于待测 DNA 插入部位的,与 *Lac Z* 基因互补的一段通用引物,这样就可以采用常规的 Sanger 链中止法,从放射自显影图谱上直接读出这段待测 DNA 的序列。

整个过程可以用图 1 来表示,下面就各步进行详细的说明。

一、噬菌体 M13 (图 1-1)

M13 是一种纤维状的对雄性 *E. coli* 专一的单链 DNA 噬菌体,由 6407 个核苷酸所组成。它的一些基本特性总结于表 1。它共包含有十个基因,各个基因的功能可见表 2。

图 2 是经改造后的 M13 基因图。由此图可见:在基因 II 和 IV 之间,有一段不编码蛋白的 507 个碱基对所组成的基因间隙(Intergenic Space, IS)。根据目前的知识,认为外源 DNA 只有插入到这一区域,才不影响该噬菌体的生长。M13 DNA 的复制可分为三个连续的阶段(图 3)。第一个阶段是噬菌体起感染作用的单

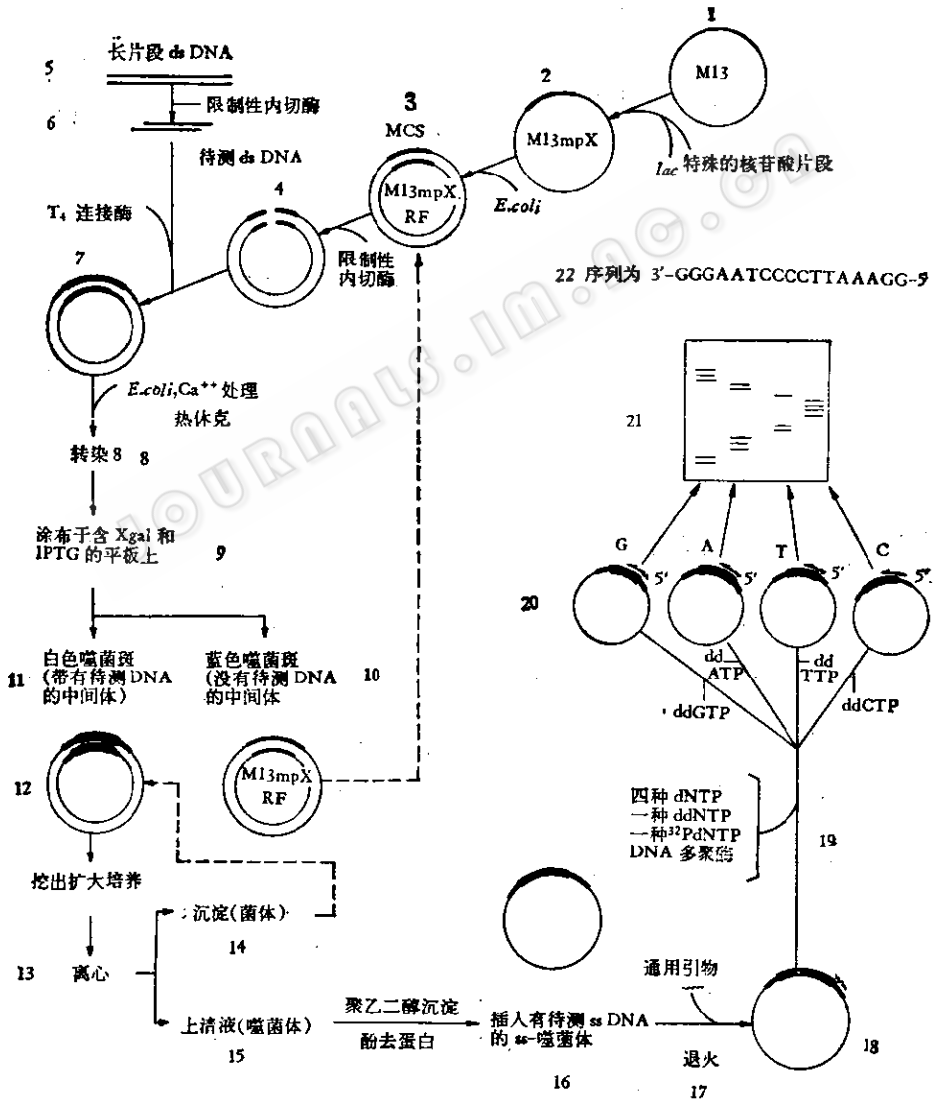


图 1 采用改造过的 M13 作为克隆载体进行 DNA 序列测定的示意图

链 DNA (+链)进入寄主细胞,而成为亲本的双链复制形式 (Replicative Form, RF), 即 SS  $\rightarrow$  RF 阶段。这个阶段依赖于寄主的复制系统。第二阶段是复制形式的扩增, RF  $\rightarrow$  RF。这是由噬菌体基因 2 编码的蛋白所起始的, 每个寄主细胞大约可产生 300 个这样的分子。第三阶段则由 DNA 的复制形式转变为子代噬菌体的单链形式, 即 RF  $\rightarrow$  SS。此单链 DNA 被基因 5 产生的结合蛋白所包裹, 随后在通过寄主细胞膜时改为基因 8 所编码的外壳蛋白而释放出去。

M13 与其它一些纤维状的噬菌体, 如 fd 和 fi 等十分相像, 它们不是通过将寄主细胞裂解, 而是随着寄主细胞不断地生长和分裂而从感染的细胞中释放出来的<sup>[3,10]</sup>。不过, 由于感染过的细胞生长速度比正常细胞慢, 如正常的 *E. coli*, 繁殖一代所需时间为 15—20 分钟, 而感染了 M13 后, 则需 30—60 分钟 这样就会在平板上出现不同于正常的噬菌斑。

由于 M13 是纤维状的, 比较适合不同大小外源 DNA 的插入; 另外, M13 本身是单链的, 对感染过的 *E. coli* 发酵液的离心上清液进行

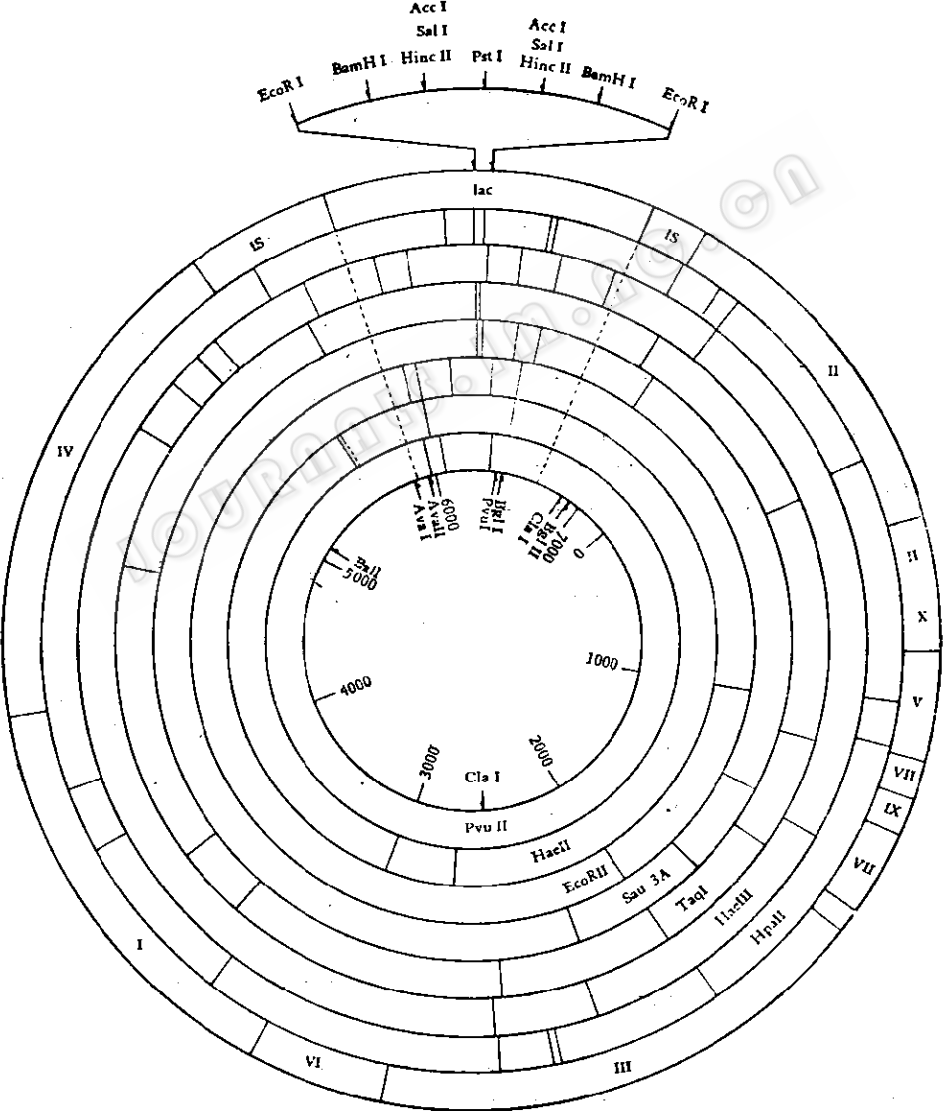


图2 M13 mp7 复制形式 DNA 的限制性内酶切点图

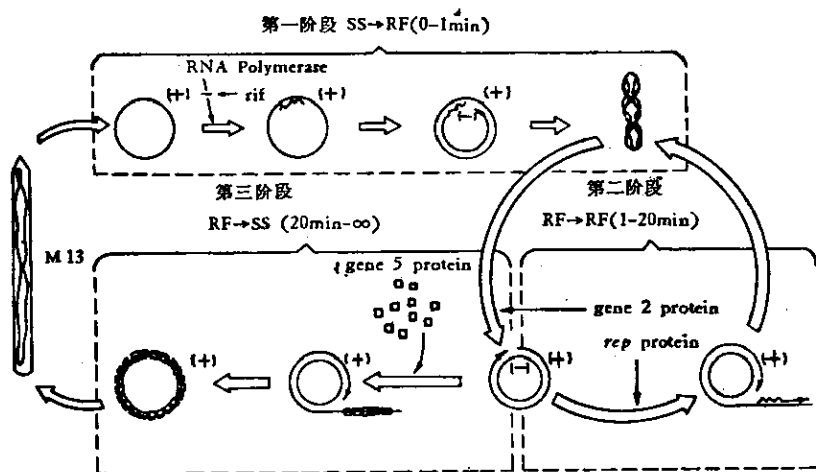


图3 M13 DNA 复制的三个阶段

表1 M13 病毒粒子和它的 DNA 特性

病毒粒子	DNA(单链)
形状: 纤维状	形状: 环状
质量: $16.4 \times 10^6$ 道尔顿	质量(钠盐): $2.11 \times 10^6$ 道尔顿
$10^{14}$ 个的重量: 2.7mg	$10^{14}$ 个的重量: 0.35mg
长度: 895nm	核苷酸数: 6407
直径(外径): 9nm	碱基组成
(内径): 2.5nm	
外壳蛋白数: 2710	
(基因8编码)	
附着蛋白数: 4—5	
(基因3编码)	
	A 24.58%
	T 34.67%
	G 20.52%
	C 20.23%

表2 噬菌体 M13 的基因及其功能

基因	功能	所编码的蛋白质质量 ( $\times 10^3$ 道尔顿)	每个病毒粒子或感染细胞的拷贝数
1	形态发生	35	4—5
2	RF 的复制, 切口	34	
3	附着蛋白 (A 蛋白)	43	
4	形态发生	50	
5	单链结合蛋白	9.8	160,000
6	形态发生 (D 蛋白)	12	5
7	形态发生 (C 蛋白)	3.5	10(包括基因9产生的)
8	主要的外壳蛋白(B 蛋白)	5.2	2710
9	次要的外壳蛋白(C 蛋白)	3.3	10(包括基因7产生的)
X	未知	12	
IS	基因间隙		

抽提, 每 ml 可得到 5—10  $\mu$ g 单链 DNA, 这就为大量制备单链 DNA 模板提供了方便; 而侵入到 *E. coli* 后的 M13 复制形式, 其 DNA 又呈双链形式, 便于各类限制性内切酶的识别和

切割, 也便于连接酶将外源 DNA 接进去。除此之外, M13 还有其它一些我们随后将会提到的优点, 因此人们将它选择作为运载工具不是没有道理的。

## 二、M13 的改造(图 1-2)

野生型的 M13, 本身不具备  $\beta$ -半乳糖苷酶的基因, 即为 *Lac*<sup>-</sup>型, 而某些 *E. coli* 的突变株(如 K12 JM101 和 JM103), 其  $\beta$ -半乳糖苷酶的结构基因是有缺失的(缺少其中的  $\alpha$ -肽的结构基因), 同样为 *Lac*<sup>-</sup>型。这类寄主细胞在受到野生型 M13 的感染后, 在含有 IPTG 和 X-gal 的琼脂平板上呈现为白色的噬菌斑\*, 说明没有  $\beta$ -半乳糖苷酶的产生。

Messing 等用 DNA 重组的办法, 把 *Lac* 的启动-操纵基因和为  $\beta$ -半乳糖苷酶头 145 个氨基酸编码的结构基因插入到 M13 的 IS 区域中, 如果此时再将其感染到上述  $\beta$ -半乳糖苷酶结构基因缺失的某些 *E. coli* 菌系中去, 则会补足寄主细胞基因的缺失, 恢复  $\beta$ -半乳糖苷酶的产生(这一过程称之为  $\alpha$ -互补), 使这些细胞表现出 *Lac*<sup>+</sup>型来, 因而在含有 IPTG\* 和 X-gal\*\* 的平板上出现蓝色的噬菌斑<sup>[11-12]</sup>。

\* IPTG: Isopropylthio- $\beta$ -galactoside, 是一种  $\beta$ -半乳糖苷酶的诱导物。

\*\* X-gal: 5-bromo-4-chloro-3-indolyl- $\beta$ -9-galactoside, 是  $\beta$ -半乳糖苷酶的人工底物。 $\beta$ -半乳糖苷酶在 IPTG 的诱导下, 可以作用于 X-gal, 产生蓝色。这是鉴定  $\beta$ -半乳糖苷酶存在的一种方法。



图4 M13 不同衍生物的 DNA 序列

另外，他们又在此 *Lac* 中插入了一段具有不同限制性内切酶识别和切割位点的寡核苷酸片段，从而构成了一些改造过的 M13 的衍生物。根据插入物的不同，分别将其中几种常用的命名为 M13mp2, M13mp5, M13mp7, M13mp8, 和 M13mp9。为了简便起见，我们在这里将其统称为 M13mpX。图 2 是 M13mp7 的基因图。由图中可以看到 *Lac* 和寡核苷酸片段在 M13 中的插入部位。起先，他们插入的片段只具备少量单调的切割位点。如在 M13mp2 中只有 *EcoRI* 一种位点，M13mp5 也只有 *Hind III* 和 *Eco RI* 二种位点，不利于广泛、灵活地应用。后

来的 M13mp7 则在 M13mp2 中又插入了 426P 的一段，而具有 *Eco RI*, *Sal I*, *Acc I*, *Bam HI*, *Hinc II* 及 *Pst I* 多种内切酶的识别和切割位点，适用于多克隆位点的目的 (multipurpose cloning site, 简称 MCS)，并有商品出售。现又制成 M13mp8 和 M13mp9，这是即将投入使用的一对核苷酸排列序列正好相反的新产品。各种 M13 衍生物的插入片段示于图 4。经这番改造后的 M13 衍生物，由于插入的片段很小，因此并不影响 *Lac* 的表达，仍能产生  $\beta$ -半乳糖苷酶来。

(未完待续)