

王 武

(无锡轻工业学院, 无锡)

近几年来, DNA 序列分析技术有了很大发展, 可以说, 分子生物学突飞猛进的发展与 DNA 序列分析技术的不断改善是分不开的。

虽然 Watson 与 Crick 早在 1953 年就测出了 DNA 的双螺旋结构, 然而直到六十年代末还没有人能对 DNA 进行序列分析。值得注意的是, 人们在六十年代中期就能够进行 RNA 序列分析。第一个被测出完整序列的是酵母丙氨酸 tRNA<sup>[1]</sup>。主要步骤是, 在一定条件下, 先用核酸酶把这个 tRNA 切成大小不同的片段, 分离出各亚片段, 再用牛胰核酸酶切开亚片段中嘧啶核苷酸与以 5'-P 与其相连的核苷酸之间的磷酸酯键, 同时核酸酶 T<sub>1</sub> 则用于切开嘌呤核苷酸与以 5'-P 与其相连的核苷酸之间的磷酸酯键。处理得到的低聚核苷酸经双向 DEAE 纸上电泳分离, 可从它们的互补排列图上读出这个 tRNA 的全部序列。这个方法的成功无疑会给予 DNA 序列分析方法一些新的启发, 然而, 三、四年后, 人们还未能找到与牛胰核酸酶、核酸酶 T<sub>1</sub> 相对应的脱氧核酸酶类。

到了六十年代末, Southern<sup>[2]</sup> 首次进行了 DNA 序列分析, 他用二苯胺降解法测定豚鼠的  $\alpha$ -卫星 DNA 不能推广于一般的 DNA, 因为卫星 DNA 序列具有高度的重复性。与此同时, Wu 和 Taylor<sup>[3]</sup> 的方法则是有些苗头的。为了分析噬菌体  $\lambda$  DNA 的粘性末端序列, 他们用了四种 <sup>3</sup>H-dATP 作为 DNA 多聚酶的底物, 以  $\lambda$  DNA 粘性末端为模板进行了体外分段接力补链合成, 最终从同位素电泳图谱上分析出了  $\lambda$  DNA 两个粘性末端各 12bp 的序列。此法在当时还不能用于长片段的序列测定, 不过它奠定了同位素分析法的基础。

以后的四、五年内, DNA 序列分析技术没有什么新的进展。一直到了 1975 年, 著名的 Sanger 博士发表了他的快速测定法<sup>[4]</sup>, 此法也被简称为“加减法”。主要步骤见图 1。先分离出一条待测 DNA 的单链, 以其为模板, 用 [ $\alpha$ -<sup>32</sup>P]-dATP 作标记, 由去除了核酸酶活力的 DNA 多聚酶 I 进行带引物体外补链合成。把酶促补链反应适当地控制在几个时间点上, 可期待“热”链(新合成带标记的链)是一组长短不等的片段混合物。把它们分成两份, 分别进行“加”、“减”法处理。

1. “减”法处理: 已知未去除 3'→5' 核酸酶活力的 DNA 多聚酶 I, 在加有三种 dNTP 的情况下进行切链和补链的净结果, 是使得新链停留在与所缺加的那种 dNTP 相应的位点之前。“减”法处理分四组进行, 每组各缺

A, C, G, T。缺 A 的组里, 修饰后的“热”链是一系列终止在“A”前一位的混合片段。同理类推于缺 C, G, T 组。

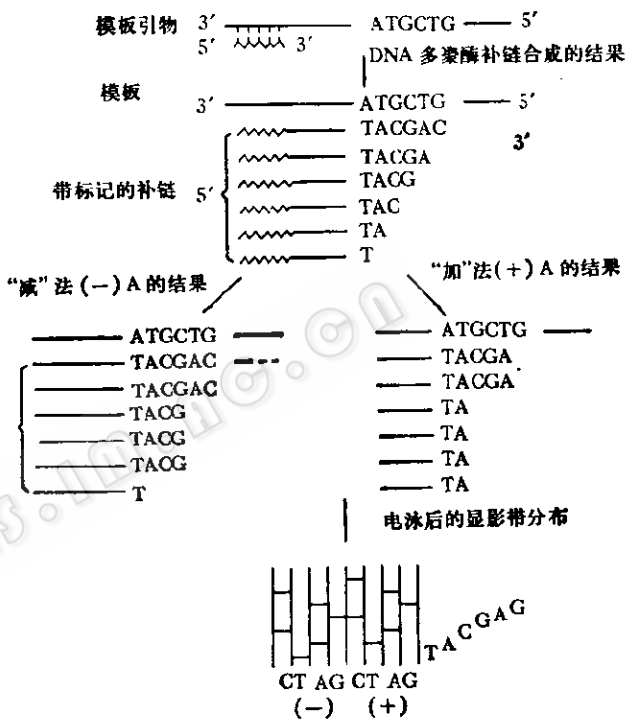


图 1 “加减法”DNA 序列分析的基本原理

2. “加”法处理: 已知带有 3'→5' 核酸酶活力的 T<sub>1</sub> DNA 多聚酶, 在单种 dNTP 存在下进行切链和补链的净结果, 是使得新链以该种核苷酸为末端。同样, “加”法处理也分四组进行, 分别只加 A, C, G, T。这样 A 组里的新链都以 A 为 3'-末端。同理类推于 C, G, T 组。

接着对以上八组反应物分别进行变性处理, 把“热”链从模板上分离下来, 再分别置于含有 8M 尿素的聚丙烯酰胺凝胶上进行电泳分离, 分离的结果能使得仅为一个核苷酸之差的片段分离开。经同位素自显影后, 从 X 光底片上的感光带排列中可以互补地读出待测 DNA 的序列。“加”和“减”法用以互相证实对方的结果, 提高可靠性。在当时情况下, 可一次测出长约 50bp 的 DNA 片段。该法虽有突破, 但就任何一段未

表 1 化学法断链处理的主要过程

处理组别	碱基修饰试剂	修饰作用	修饰后处理	断裂的碱基
Ia	硫酸二甲酯	嘌呤甲基化	先冷, 稀 HCl, 后热碱处理断链	A
Ib	硫酸二甲酯	嘌呤甲基化	先 pH7, 加热, 后热碱处理断链	A + G
IIa	胍	打开嘧啶环	嘧啶中加热断链	T + C
IIb	胍 + 2M NaCl	打开胞嘧啶环	嘧啶中加热断链	C

表 2 双脱氧法的分组处理系统

组别	DNA 样品	标记物*	dNTP 底物	ddNTP 底物	追加 dNTP	末端核苷酸
A	单链带引物	[ $\alpha$ - $^{32}$ P]dATP	dCTP, dGTP, dTTP	ddATP	dATP	A
C	单链带引物	[ $\alpha$ - $^{32}$ P]dATP	dATP, dGTP, dTTP	ddCTP	dATP	C
G	单链带引物	[ $\alpha$ - $^{32}$ P]dATP	dATP, dCTP, dTTP	ddGTP	dATP	G
T	单链带引物	[ $\alpha$ - $^{32}$ P]ATP	dATP, dCTP, dGTP	ddTTP	dATP	T

\* 标记物可用任一种 dNTP, 不过底物系统也应作相应的调整。

知序列的 DNA, 要互补上合适的一小段引物仍是很费事的, 再者, 下手的 DNA 仍要求是单链的。

一年多后, 哈佛大学的 Maxam 和 Gilbert 发表了另一种化学断链法<sup>[17]</sup>。此法仍采用  $^{32}$ P 标记和凝胶电泳分离读带技术。下手的 DNA 可以是单链的, 也可以是双链的。 $^{32}$ P 不是被掺入到链中, 而是被标在末端。如果用 [ $\gamma$ - $^{32}$ P] ATP 作标记, 可用核苷酸磷酸激酶把  $^{32}$ P 标在 5'-端, 如用 [ $\alpha$ - $^{32}$ P]dATP 作标记, 则可用末端转移酶把  $^{32}$ P 标在 3'-端。待测的 DNA 若是双链的, 标记后用一种限制性内切酶将其切成两段, 以保证标记只标在片段的单侧。纯化后的待测片段分成四份, 分别进行化学断链处理, 主要处理过程见表 1。

在适当的处理条件下, 每组处理样品中, 断链都是随机发生的, 由此可从每组反应物中得到一系列带标记的、长短不等的断裂片段。最后通过类似的热变性处理, 凝胶电泳, 同位素自显影等步骤, 可以从四组互补图谱上读出 DNA 序列。

Maxam 和 Gilbert 方法的缺点是  $^{32}$ P 只标在 DNA 片段的末端, 要使所有的 DNA 分子都标上, 初始剂量要大, 而且此法涉及到多步的同位素操作。不过这个方法以不需要引物, 可直接用双链 DNA 为其优点, 至今仍为世界上许多实验室所沿用。

1977 年, Sanger 博士也在他的“加减法”基础上改革了 DNA 序列分析法, 这就是著名的“双脱氧”法<sup>[18]</sup>。此法仍用单链 DNA 为模板进行带引物补链合成。不仅用了 dNTP, 而且还用了 ddNTP (2', 3'-双脱氧核苷三磷酸) 作为 DNA 多聚酶 I Klenow 片段 (无核酸酶活力) 的底物。这个方法的关键是 2', 3'-ddNTP 仍能按碱基配对原则被掺到新链中去, 不过由于 3' 位也脱氧, 掺入后再也不能与下一个核苷酸形成磷酸酯键, 而

迫使新链的合成停止。双脱氧法也把样品分四组进行体外补链反应, 大致情况如表 2 所示。

通过调整底物 dNTP 和 ddNTP 之间的用量, 可以从各组反应物中得到一系列长短不等的“热”链。以后的步骤与“加减法”相同。与“加减法”相比, 该法不仅简单, 且精确度高。同位素用量比“化学断链法”为少, 但缺点是仍需单链 DNA 和特定的底物。

为解决这些问题, Messing 等人发展了一种新的“M13 克隆/双脱氧顺序”方法<sup>[17, 18]</sup>。该法的关键是利用经遗传工程改造过的噬菌体 M13 DNA 作为克隆待测 DNA 的载体。M13 在 *E. coli* 中 DNA 为双链, 释放到培养液后, DNA 是单链的, 克隆后的 DNA 片段随 M13 DNA 的复制而复制, 大量扩增后的单链拷贝很容易得到。为解决满足各种限制性内切位点的克隆、克隆后的检出以及标准引物位点的建立等问题, M13 DNA 被改造成 M13 MP 系统 (图 2)。6.5kb 的 M13 DNA 中只有一段 507bp 部分 (IS 片段) 被外源 DNA 插入后仍不失其自身的活性。在 IS 片段中先插入一段 *E. coli* 乳糖操纵子的调节基因和 Z 基因的前半部分, M13 宿主的 F 质粒的缺失过的 Z 基因能与之发生互补作用 ( $\alpha$ -互补), 互补结果使得 *E. coli* 能产生完整的基因产物—— $\beta$ -半乳糖苷酶。这种改造过的 M13 mp 噬菌体能在经典的 IPTG-Xgal 平板上显出蓝色噬菌斑。进一步的改造是在 M13 mp 的 Z 基因 N 端部分插入一段人工合成的多接点 (Polylinker) 片段, 以便于各种限制性末端的外源 DNA 都可插入。多接点片段仅长几十个碱基对, 它的插入不破坏  $\alpha$ -互补作用, 然而当外源 DNA 再插到多接点片段后,  $\alpha$ -互补作用就被破坏了, 结果克隆有外源 DNA 的 M13mp 在 IPTG-Xgal 平板上显出无色噬菌斑, 很容易鉴别出克隆子。M13 mp 的结

1983年, Biggin, Gibron 和洪国藩在 Sanger 博士的实验室里研究出两种新方法<sup>[9]</sup>, 结合起来克服了上

表 3 几种常见的用于 DNA 序列分析的电脑程序

程序名称	程序处理内容
NUMINFO	程序 NUMSEQ 的使用说明与练习
NUMSEQ	为输入的 DNA 序列编号, 译出编码的氨基酸序列
RESTINFO	程序 BACHREST 的使用说明与练习
BACHREST	查出所有的限制性内切位点
COMREST	查出商业上已供应的限制性内切酶的切点
DIGINFO	程序 DIGEST 的使用说明与练习
DIGEST	计算序列上各限制性内切片段的长度
COMPINFO	程序 COM 的使用说明和练习
COMP	计算 DNA 序列中四种碱基的组成
DOMINFO	以下四种程序的说明书
SLOHOM	查出 DNA 序列之间的相同部分 (打印出)
GRAHOM	查出 DNA 序列之间的相同部分 (图表示出)
PROTHOM	查出蛋白质序列之间的相同部分 (打印出)
PROHOM	查出蛋白质序列之间的相同部分 (图表示出)
GEL	根据 DNA 分子量标准物, 计算出未知 DNA 片段的长度

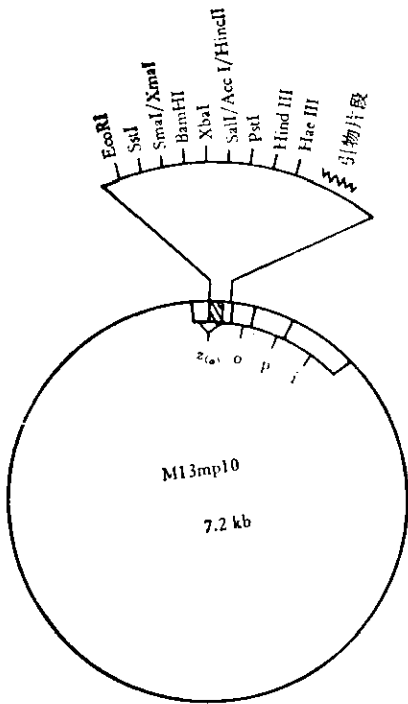


图 2 M13mp10 DNA 结构  
 □ 插入的乳糖操纵子基因部分  
 ▨ Polylinker

构使得引物问题也迎刃而解, 由于外源片段总是插在多接点上, 以与其相邻的部位作为标准引物的互补部位, 任何待测片段都可以使用同一种引物。单链 DNA 和引物问题解决后, 以双脱氧法进行序列分析。目前 M13 克隆双脱氧顺序法已成为最常用的方法之一。

不管是化学断链法还是“双脱氧”法都离不开用  $^{32}\text{P}$  作标记。 $^{32}\text{P}$  标记法有几个缺点: 一是半衰期仅 14 天, 难以保存; 二是  $^{32}\text{P}$  衰变会造成磷酸二酯键断裂; 三是其  $\beta$ -射线的能量高, 加上散射的原因, 致使 X 光底片上显出的带比实际电泳带为宽, 对分辨率有所影响; 再者  $\beta$ -硬射线对工作人员带来一定的危害性。

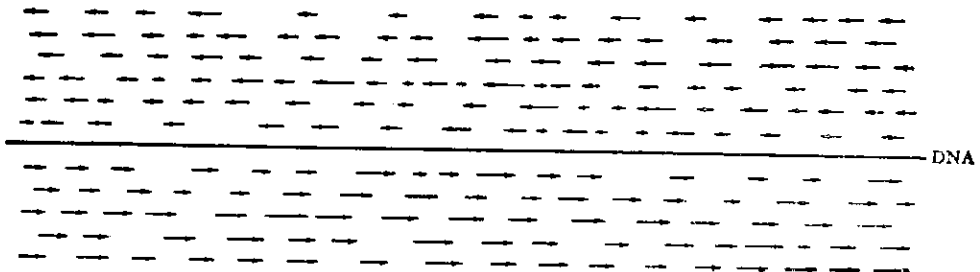


图 3 DNA 序列分析霰弹法的示意图

图中的短箭头表示每个被分析的亚片段; 中线上方的箭头表示分析的是 (+) 链, 下方的箭头表示分析的是 (-) 链; 中线代表由电脑排出的完整序列

述的不足之处。他们发明了以  $[\alpha\text{-}^{35}\text{S}]\text{-dATP}$  代替了  $[\alpha\text{-}^{32}\text{P}]\text{dATP}$  作为标记物进行“双脱氧”法序列分析。 $^{35}\text{S}$  的半衰期为 12 周,其放射出的  $\beta$ -射线的能量远比  $^{32}\text{P}$  的为低。这样既延长了同位素试剂的存放期,提高了电泳显影带的分辨率,也减轻了射线对人体的危害性。文章<sup>[9]</sup>中介绍的一种新的梯度凝胶薄层制作法,与常规的凝胶薄层相比,在同样的操作条件下可多读带 20~30%。 $[\alpha\text{-}^{35}\text{S}]\text{dATP}$  标记法与梯度凝胶电泳相结合使用,可一气从长 50cm 的凝胶薄层上读出 300 bp 左右的 DNA 序列。一块薄层上可同时测 7 个类似长短的 DNA 片段,这样一次成功的电泳可测出共 2000 bp 左右的 DNA 序列。目前世界上许多实验室已开始采用这些方法。

为建立基因文库的需要,各种生物大量的 DNA 序列有待于测定,尽管我们有了上述的种种方法,但要先摸清一段长达几万碱基对以上的 DNA 序列的限制性内切图谱,再分割成约几百碱基对长的片段分别进行测定,是很花时间的。目前解决这个问题的有效策略是利用“霰弹效应”的原理进行大批量无序亚片段的测定。具体的方法有很多种,如 Deininger 的超声波断链-霰弹测定法<sup>[10]</sup>。这个方法使用了超声波对长至几万甚至几十万碱基对的 DNA 片段进行随机断链处理得到的长短不一的短片段经分离后,取出相当于长几百碱基对的片段,克隆到 M13 mp 载体上,挑出无色的噬菌斑,检查插入的 DNA 片段的大小,然后随机地挑出许多长短不一的片段进行“双脱氧”法分析。多亏有电脑相助,分析结果输入电脑后,片段之间重复的部分很快就被查出,于是整个大片段的序列也就随之排出了(图 3)。此外,还有外核酸酶 III-霰弹测定法、DNaseI-

霰弹测定法<sup>[11]</sup>等,所采用的策略都是一样的。

值得一提的是电脑开始在分子生物学领域大显身手,光 DNA 序列分析资料处理就用到了许多种程序(表 3)。普通的 APPLE II 等都可用来进行这些程序的运算处理。

电脑程序在 DNA 序列分析技术方面的应用大大加速了分子生物学研究的进展。

### 参 考 文 献

- [1] Holley, R. et al.: *Science*, **147**: N3664, 1462—1465, 1966.
- [2] Southern, E. et al.: *Nature*. **227**: 22, 1970.
- [3] Wu, R. et al.: *J. Mol. Biol.*, **57**: 491, 1971.
- [4] Sanger, F. et al.: *J. Mol. Biol.*, **94**: 441—448, 1975.
- [5] Maxam, A. and W. Gilbert: *Proc. Natl. Acad. Sci. USA*. **74**: 560—564, 1977.
- [6] Sanger, F. et al.: *Proc. Natl. Acad. Sci. USA*, **74**: 5463—5467, 1977.
- [7] Messing, J. et al.: *Proc. Natl. Acad. Sci. USA*, **74**: 3642—3646, 1977.
- [8] Messing, J. et al.: In the Third Cleveland Symposium on Macromolecules: Recombinant DNA. ed. A. Walton, Elsevier, Amsterdam, 143—153, 1981.
- [9] Biggin, M. et al.: *Proc. Natl. Acad. Sci. USA*, **80**: 3963—3965, 1983.
- [10] Deininger, P.: *Anal. Biochem.*, **129**: 216—223, 1983.
- [11] Anderson, S. et al.: *Nucleic Acids Res.*, **9**: 3015—3027, 1981.