及展览会论文集. 北京：北京大学出版社，1993：B5
4 杨伯宇，朱大模，洪群发等. 科学通报，1994；39（14）：1299
5 Beavis R C, Chait B T. Anal Chem, 1990; 62: 1836
6 Weinberger S R. Oral presentation at the Pittsburgh conference and exposition on analytical chemistry and applied spectroscopy. Atlanta: 1993; March 3—8
7 Schaer M, Boernsen K O, Gassmann E. Rapid Commun Mass Spectrom, 1991; 5: 319
8 Weinberger S R. 见：孙德中编. 第五次北京分析测试学术报告会及展览会论文集. 北京：北京大学出版社，1993：B33

Matrix-Assisted Laser Desorption Lonization Mass Spectrometry and Its Applications in Biochemistry Research. Yang Boyu, Zhu Damo, Zhang Yukui (*Dalian Institute of Chemical Physics, Chinese Academy of Sciences, National Chromatographic Research and Analysis Centre, Dalian* 116011, China).

Abstract Matrix-assisted laser desorption ionization mass spectrometry (MALD1/MS) is a new technique developed in recent years. It can be broadly applied to many aspects of modern life science. Basic principles of this technique are discribed, some important research results in our group are also discribed, for example, accurate protein molecular weight determination and the separation and identification of protein mixtures. The feasibilities and problems of MALD1/MS applied to fast protein sequencing and DNA sequencing are also discussed.

Key words matrix-assisted laser desorption ionization mass spectrometry, molecular weight determination, separation and identification

PCR 产物中无错误碱基拷贝最低比率的计算

方悦群 王秉瑞
（卫生部兰州生物制品研究所，兰州 730046）

摘要 分析了 PCR 过程中带有错误碱基拷贝的量变过程，得出不同循环（n）后不同类型拷贝数的计算通式并以逐次代入方式归纳出 PCR 产物中无错误碱基拷贝最低比率（R）和有效循环数（N），拷贝酶促合成链长（H）及错配率（f）的关系式 $R_n=(1-Hf/2)^{N-1}(1-Hf)$，对 PCR 技术制备表达用 DNA 片段有指导意义.

关键词 多聚酶链反应，碱基错配，基因表达

耐热 DNA 聚合酶有较高的碱基错配率（如 2×10^{-4}）[1—3]. 理论上讲，每次 PCR 循环中不仅前一循环后已带有错误碱基的拷贝（以下称错误拷贝）数量会加倍，而且还会产生新的错误拷贝. 随着循环数的增加，DNA 双链中不含任何错误碱基的拷贝（以下称无错误拷贝）比例在产物中越来越小，这使 PCR 产物不适用于基因表达. 本文试图找到 PCR 产物中无错误拷贝比率与扩增循环数，拷贝链长和错配率之间的数学关系.

1 PCR 扩增的拷贝的分类

将所有 PCR 扩增拷贝（双链）分为 a. 无错误拷贝；b. 单错误拷贝，指模板链无错误碱基而复制时互补链按错配率发生了碱基错配的拷

收稿日期：1994-07-13，修回日期：1995-02-20

贝；c. 双错误拷贝，指模板链含一个以上错误碱基，复制时互补链按碱基互补原则掺入错误碱基的拷贝（也按错配率发生错配，但已不必考虑）.

2 PCR 产物中无错误拷贝比率的计算原理

2.1 计算的前提条件

a. 扩增的拷贝的碱基链长小于错配率倒数，否则计算无意义. 错配率倒数的含义为每个循环中随机产生一个碱基错配所需要的酶促合成碱基数；b. 每个循环中每个拷贝复制最多产生一次错配. 此前提条件是合理的，因为首先，由于条件 a 和概率原因，这符合大多数实际情况；其次，在错配率一定时（即每个循环发生的碱基错配总数一定），此条件使碱基错配在拷贝间分布呈最大限度分散，计算的无错误拷贝比率最低，这对评价和安排筛选无错误拷贝克隆是有益的；c. 不考虑回复突变. 发生回复突变将提高无错误拷贝比率，亦有利于实验.

2.2 计算原理

以 f 代表错配率；设 n 循环后 T_n 为拷贝总数，C_n 为无错误拷贝总数，D_n 为双错误拷贝总数，S_n 为单错误拷贝总数，R_n 为无错误拷贝在总拷贝数中的比率，计算通式分别为

T_n： $T_n = C_n + D_n + S_n$

另外，以 a 为扩增前拷贝数，不考虑扩增效率（下同）时

$$T_n = 2^n a \quad \cdots\cdots(1)$$

C_n： $C_n = T_n - D_n - S_n \quad \cdots\cdots(2)$

R_n： $R_n = C_n/T_n = (T_n - D_n - S_n)/T_n$

$$= 1 - D_n/T_n - S_n/T_n \quad \cdots\cdots(3)$$

D_n：根据 DNA 复制原理，$n-1$ 循环后的一个双错误拷贝在 n 循环复制时产生两个双错误拷贝；$n-1$ 循环后的一个单错误拷贝在 n 循环复制时产生一个双错误拷贝，这是 n 循环后的双错误拷贝的全部来源，

$$D_n = 2D_{n-1} + S_{n-1} \quad \cdots\cdots(4)$$

S_n：因每个循环中每个拷贝最多产生一次碱基错配（2.1b.），故循环中发生在所有无错误碱基模板的互补链上的碱基错配总数在数值上等于该循环后的单错误拷贝数 S_n（最大值，下同），从而可由前者来求 S_n. 设 E_n 为 n 循环中发生在所有无错误碱基模板的互补链上的碱基错配总数，拷贝链长为 h（碱基数），错配率 f 此时应为 E_n 与循环中所有无错误碱基模板的互补链的碱基合成总数之比，可表示成

$f = E_n/$(循环中无错误碱基模板总数 $\times h$)

其中：

循环中无错误碱基模板总数 = 循环中模板总数 − 循环中有错误碱基模板总数

循环中模板总数显然等于拷贝总数 T_n；$n-1$ 循环后的单错误拷贝和双错误拷贝在 n 循环中分别成为等量的和两倍的有错误碱基模板，因此

循环中有错误碱基模板总数 $= 2D_{n-1} + S_{n-1} = D_n$(参考公式(4))，

则： $f = E_n/(T_n - D_n)h$

$E_n = (T_n - D_n)(hf)$

$S_n = E_n = (T_n - D_n)(hf)$

或 $= T_n(hf) - D_n(hf) \quad \cdots\cdots(5)$

将公式（5）代入公式（3）并整理，

$$R_n = (1 - D_n/T_n)(1 - hf) \quad \cdots(6)$$

由于以上是以 2.1 *b*. 为计算基础的，故 S_n 为最大值，而 R_n 为最低值.

3 无错误拷贝最低比率计算公式的建立

在以上基础上将 $n=1$，2，3…依次代入(1)，(4)，(5)，(6) 式后整理归纳.

$n = 1$：

当 $n = 0$ 时，$D_0 = 0$；$S_0 = 0$；$T_1 = 2a$；

$D_1 = 2D_0 + S_0 = 0$

$S_1 = T_1(hf) - D_1(hf)$

$= 2a(hf) - 0 \times (hf) = 2a(hf)$

$R_1 = (1 - D_1/T_1)(1 - hf)$

$= (1 - 0/T)(1 - hf) = 1 - hf$

$n = 2$：

$T_2 = 2^2 a$

$D_2 = 2D_1 + S_1$

$$=2\times 0+2a(hf)=2a(hf)$$

$$S_2=T_2(hf)-D_2(hf)$$

$$=2^2a(hf)-2a(hf)(hf)$$

$$=2^2a(hf)-2a(hf)^2$$

$$R_2=(1-D_2/T_2)(1-hf)$$

$$=[(1-2a(hf)/2^2a](1-hf)$$

$$=(1-hf/2)(1-hf)$$

$n=3$：

$$T_3=2^3a$$

$$D_3=2D_2+S_2$$

$$=2^3a(hf)-2a(hf)^2$$

$$S_3=T_3(hf)-D_3(hf)$$

$$=2^3a(hf)-2^3a(hf)^2+2a(hf)^3$$

$$R_3=(1-D_3/T_3)(1-hf)$$

$$=[1-2\times hf/2+(hf/2)^2](1-hf)$$

$n=4$：

$$\cdots R_4=[1-3\times hf/2+3\times(hf/2)^2-(hf/2)^3](1-hf);$$

$n=5$：

$$\cdots R_5=[1-4\times hf/2+6\times(hf/2)^2-4\times(hf/2)^3+(hf/2)^4](1-hf)$$

…… ……

不同 n 的 R_n 表明，R_n 式前一因子中各项为扬晖三角系数，即为一个二项式，可简化为 $1-hf/2$ 的 $n-1$ 次方与 $1-hf$ 的乘积，得

$$R_n=(1-hf/2)^{n-1}(1-hf) \quad \cdots(7)$$

4 R_n 计算公式参数校正

n 的校正：由于 PCR 扩增效率问题，在公式（7）计算中采用实际循环数 n 是不准确的，这里采用有效循环数 N[12]，即无论 PCR 效率如何，都把最终产物量 Y 对最初模板量 A 的倍数表示成

$$Y/A=2^N$$

此表达式与 Saiki 的考虑扩增效率 X 时的 PCR 扩增数学表达式 $Y/A=(1+X)^n$ 的数学效应一致，限于篇幅，不赘述两式的关系和使用 N 的理由．则

$$N=3.32\lg(Y/A) \quad \cdots\cdots(8)$$

其中 Y 和 A 可通过实验测得．N 与 n 的关系可由效率系数 α 表明

$$\alpha=N/n \text{ 或 } n=N/\alpha \quad \cdots\cdots(9)$$

h 的校正：PCR 引物由 DNA 合成仪合成（化学合成），故引物的链长部分不存在碱基错配问题．设 H 为扩增片段的酶促合成链长（碱基数），P_1 和 P_2 为两条引物长度（碱基数），这样，一对引物所确定的一种双链拷贝的正负链平均酶促合成链长为（拷贝长时可忽略不计）

$$H=[2h-(P_1+P_2)]/2$$

$$=h-(P_1+P_2)/2 \quad \cdots\cdots(10)$$

根据以上 N 和 H，公式（7）应改为

$$R_n=(1-Hf/2)^{N-1}(1-Hf) \quad \cdots\cdots(11)$$

5 无错误拷贝最低比率 R 的意义

公式（7）或（11）表明，循环数（n 或 N），拷贝链长（h 或 H）和错配率（f）越大，PCR 产物中无错误拷贝最低比率（R）越小．这一公式可用于 a．根据 R 判断筛选一个 PCR 产物中无错误拷贝的难易程度，如 $R=0.1$，即意味着每 10 个拷贝中仅有一个（至少）是不带错误碱基的；b．设定一个便于筛选的比率（R）来安排 PCR 循环数．从公式（11）中可得

$$N=[\lg R_n-\lg(1-Hf)]/\lg(1-Hf/2)+1 \quad \cdots\cdots(12)$$

再通过预实验和公式（8）和（9）求得特定反应体系的效率系数后安排实际扩增循环数．

6 总 结

本文以逐次代入方式归纳出和总结了 PCR 产物中无错误拷贝最低比率等几个有用的公式：

公式(11) $R_n=(1-Hf/2)^{N-1}(1-Hf)$

公式(8)和(9) $N=3.32\lg(Y/A)$ 和 $n=N/\alpha$ 或 $\alpha=N/n$

公式(10) $H=h-(P_1+P_2)/2$

公式(12) $N=[\lg R_n-\lg(1-Hf)]/\lg(1-Hf/2)+1$

以上公式对运用 PCR 技术制备表达用基因片段有指导意义．

（下转第 287 页）

调作用. 这些结果表明，蛋白激酶C (PKC) 的活性与 TGF α 的生成有关，PKC 对 TGF α mRNA 的上调作用和 TGF α 自分泌作用可能存在不同的调节通路[7]. TPA 与胃癌细胞 TGF α mRNA 表达关系的研究目前尚未见报道.

参 考 文 献

1 Delarco J, Todaro G T. Proc Natl Acad Sci USA, 1978; **75**: 4001
2 王凯华. 实验生物学报, 1983; **16** (3): 257
3 Chomezynski P, Sacchi N. Anal Biochem, 1987; **162**: 156
4 Coffey R J, Shipley G D, Moses H L. Cancer Res, 1986; **46**: 1164
5 Malden L T, Novak V, Burgess A W. Int J Cancer, 1989; **43**: 380
6 葛振龙，周爱儒. 生物化学杂志，1992; **8** (4): 468
7 Klein S B, Fisher G J, Jensen T C *et al.* J Cell Physiol, 1992; **151**: 326

Investigation of Expression of TGF α mRNA in Human Gastrointestinal Carcinoma. Qiao Bo, Sun Congmei, Zhou Airu (*Department of Biochemistry, Beijing Medical University, Beijing* 100083, China).

Abstract The expression of transforming growth factor α (TGF α) in human gastrointestinal carcinoma by methods of molecular hybridization and cell FIT cell-cycle analysis was investigated. The results show that the levels of TGF α mRNA are higher in gastric carcinoma tissues, colonic carcinoma tissues and rectal carcinoma tissues than in the corresponding control tissues. After treating with 12-O-tetradecanoylphorbol-13-acetate (TPA, 100μg/L, 4h), the level of TGF α mRNA in gastric carcinoma cell line (MGc80-3) increases, and percentage of S phase cells is markedly increased after 6 hours. These results suggest that TGF α plays an important role in gastrointestinal carcinoma tissues and expression of TGF α mRNA can be regulated by TPA.

Key words transforming growth factor α, human gastrointestinal carcinoma, 12-O-tetradecanoylphorbol-13-acetate

(上接第 277 页)

参 考 文 献

1 Saiki R K, Gelfand D H, Stoffel S *et al.* Science, 1988; **239**: 487
2 Keohavong P, Thilly W G. Proc Natl Acad Sci USA, 1989; **86**: 9253
3 Ennis P D, Zemmour J, Salter R D *et al.* Proc Natl Acad Sci USA, 1990; **87**: 2833

The Calculation of the Minimum of Non-Base-Mismatch Copy Proportion of the PCR Products. Fang Yuequn, Wang Bingrui (*Lanzhou Institute of Biological Products, Ministry of Public Health, Lanzhou* 730046, China).

Abstract The quantity change of the copies with mismatch bases in PCR products was analyzed. The general formulas of numbers of copies in different modes after certain cycles (n) were gained and the math relation between the minimum (R_n) of the non-base-mismatch copy proportion of PCR products and the efficiency cycle numbers (N), the copy chain length (H) catalyzed by Taq polymerase and the mismatch frequency (f) was concluded by way of substituting step by step: $R_n = (1 - Hf/2)^{N-1}(1 - Hf)$. The formula has a guidance sense for preparing the gene segments used for the gene expression by PCR technique.

Key words polymerase chain reaction, base mismatch, gene expression