

单分子 PCR 产物错误率分析*

王国华¹⁾ 吕军鸿¹⁾ 雷晓玲¹⁾ 李海阔²⁾ 李民乾¹⁾ 陈润生³⁾ 方海平¹⁾ 胡钧^{1,2)}**

(¹⁾中国科学院上海应用物理研究所, 上海 201800; (²⁾上海交通大学 Bio-X 生命科学研究中心, 上海 200030;

(³⁾中国科学院生物物理研究所, 北京 100101)

摘要 碱基错配致使 PCR 扩增产物中存在突变序列. 大量模板 PCR 扩增时突变序列所占的比例较低, 对随后进行的 PCR 产物分析影响不大, 但当对微量甚至单个模板 DNA 扩增时, 情况则完全不同. 对单分子 PCR 产物的错误率进行了理论分析, 结果表明: 根据实验目的和条件, 选择忠实性不同的聚合酶是十分关键的.

关键词 单分子 PCR, 错误率, 忠实性, 模板数

学科分类号 Q503

体外 DNA 扩增时, 聚合酶可能引发的碱基错配使聚合酶链式反应 (polymerase chain reaction, PCR) 的产物出现错误, 因此正确地计算出产物的错误率成为一个重要的问题. 在常规的分子生物学实验中, 模板数通常很大 (一般为 10^5 个以上), 如果使用忠实性低的耐热 Taq 酶 (掺入错误率为 $10^{-4}/\text{bp}$), 在扩增效率为 70% 时, 将 200 bp 长的片段扩增一百万倍, 产物错误率为 33%, 而使用忠实性高的 pfu 聚合酶 (掺入错误率为 $7 \times 10^{-7}/\text{bp}$), 同样实验效率下将 500 bp 长的片段扩增一百万倍, 产物错误率仅为 0.7%^[1]. 可见, 大量模板下的 PCR 产物中突变序列通常只占其中一小部分, 大多数情况下不足以影响随后进行的实验分析^[2-6].

最近, 基因组测序和法医诊断学的需求, 极大地促进了以少量甚至单个 DNA 分子为模板的单分子 PCR 扩增 (single molecule PCR amplification) 实验的发展^[8-11]. 在这类实验中, 模板数很少, 聚合酶导致的突变一旦发生, 在随后的循环中都会随着原来的序列一起呈指数扩增, 并且由于还可能产生新的突变, 使得突变序列随扩增循环数的增加而不断加大其所占比例, 从而可能严重影响最终的实验结果. 因此, 事先对单分子 PCR 扩增产物的错误率进行论证分析是十分必要的, 并且也已引起人们越来越多的重视.

1 研究方法与实验环境

众所周知, PCR 是一个扩增产物随着扩增循环数的增加而呈指数上升的过程. 扩增过程中产物片段数遵循如下公式:

$$T = O \times (1 + k)^N \quad (1)$$

其中 T 为产物片段数, O 为初始模板数, k 为扩增效率, N 为扩增循环数. 在常规 PCR 实验中, 大量初始模板数下, 一次扩增循环后正确链的概率遵循泊松分布的未击中事件的概率, 则进行 N 次循环后产物错误率 F 即为

$$F = 1 - e^{-L \times f \times N} \quad (2)$$

其中 L 为片段长度, f 为聚合酶的掺入错误率.

虽然扩增过程不变, 但是在小样本情况下, 泊松分布不再适用, 公式 (2) 不能直接用于计算单分子 PCR 产物的错误率. 所以在我们的计算中, 当扩增产物到达 10^5 以前用随机数进行模拟, 也就是说利用计算机产生足够多的随机数, 通过判断产生的随机数是否大于所使用的聚合酶掺入错误率来决定一个碱基是否发生了突变, 从而计算某一循环的产物错误率. 当产物的量达到 10^5 以后则可继续采用公式 (2), 以简化计算过程. 这里选择 10^5 作为计算转折点, 是因为常规实验下以 10^5 为扩增起点, 我们希望两种情况下的计算能够尽量接近. 扩增循环次数 N 仍然遵循公式 1 的规律 (这里不考虑没有扩增产物的前两个循环).

实验使用的软件是 Matlab 6.0, 假设扩增效率均为 70%. 我们选用了两种典型的聚合酶, 忠实性低而常用的耐热 Taq 酶 (掺入错误率为 $2 \times 10^{-4}/\text{bp}$ ^[1]) 和忠实性最高的 pfu 酶 (掺入错

* 国家自然科学基金重点项目 (10335070).

** 通讯联系人.

Tel: 021-59552394, Fax: 021-59552394

E-mail: junhu22@hotmail.com

收稿日期: 2003-06-30, 接受日期: 2003-09-28

误率为 $7 \times 10^{-7}/\text{bp}$), 通过计算模板数从 1 到 10 之间变化时, 分别扩增 200 bp 和 500 bp 长的片段, 产物达到 10^{12} 时的突变序列总数, 对单分子 PCR 产物的错误率进行讨论.

2 结果与分析

考虑到单个模板 PCR 与大量模板 PCR 扩增实验的情况不同, 我们首先对不同情况下的突变进行了定义. 当 PCR 被用于研究异源群体中的稀有分子, 例如对野生型基因的突变检测, 寻找 SNPs 等, 如果聚合酶诱导的突变序列掩盖了稀有 DNA 序列, 那么扩增产物将与实际结果大相径庭^[2], 这种情况下我们将每一条扩增链看作一个整体, 突变的定义针对整条链而言, 即只要含有错配碱基的序列就定义为突变序列.

但是在另一些实验中, 如为野生型基因测序而进行的 PCR 扩增, 聚合酶诱导的错配碱基是随机分布在整条链上的, 即使一些扩增产物中可能会含有个别错配碱基, 但是最终通过对所有产物片段的比较仍然可以得到一致序列^[1]. 这种情况下扩增

产物将不再作为一个整体被处理, 突变的定义针对单个碱基而言, 它演变为突变碱基的定义. 因此我们根据不同的实验目的对单分子 PCR 产物的错误率分别进行分析.

2.1 针对整条链的突变

研究异源群体中稀有分子的 PCR 实验中, 突变的定义是针对整条链而言的, 这里又分两种情况进行计算. 一种是在起始循环中是否发生错配, 完全按照产生的随机数进行模拟, 这基本代表了真实情况下的产物错误率; 第二种情况我们则假设起始循环中确有一个模板发生错配, 这时的 PCR 错误率代表了较为极端情况下的结果.

在随机错配情况下, 从表 1 中所示的 PCR 错误率计算结果可以看出, 模板从 1 到 10 的变化对产物错误率的影响不大, 并且因为计算对结果的精确度要求不高, 所以, 可以认为单个模板和稀少模板的 PCR 实验得到的结果相等, 并且这时的错误率非常接近于大量模板时的情况, 可见突变序列总数受模板数的影响很小.

Table 1 The error rate of single molecule PCR using Taq and pfu polymerase

Templates number	1	2	3	4	5	6	7	8	9	10
Taq polymerase	0.5048	0.4727	0.4833	0.4739	0.4847	0.4671	0.4675	0.4814	0.4662	0.4745
pfu polymerase	0.0109	0.0105	0.0109	0.0105	0.0109	0.0107	0.0105	0.0109	0.0113	0.0106

如果起始循环中确有一个模板发生错配, 错配碱基会随着以后的循环一起被扩增^[1], 这样的事件遵循二项分布, 这时产物的错误率应该大于随机错配假设下的错误率. 计算结果表明 (表 2), 在起始循环中发生突变的概率不是很高, 也就是说在相对少的情况下第一个扩增循环就产生突变序列, 特别是当实验中使用忠实性较高的聚合酶时. 但

是, 如果使用的是忠实性低的聚合酶, 例如 Taq 酶, 起始循环中一旦错配发生, 即使增加模板数到 10 个, 最终的产物错误率也会非常高, 这时只有更换忠实性高的聚合酶才能使错误率降低. 所以在这类实验中虽然可以用单个分子作为模板, 但是在实验中必须使用忠实性较高的聚合酶, 才能保证得到好的实验结果.

Table 2 The error rates of single molecule PCR using Taq and pfu polymerase when one template happens mismatch in the first cycle

Templates number		1	2	3	4	5	6	7	8	9	10
Taq	<i>P</i> /%	2.0	3.9	5.8	7.5	9.2	10.9	12.4	13.9	15.3	16.7
polymerase	<i>F</i> /%	73.5	64.5	57.6	55.8	53.6	52.3	51.8	51.1	50.5	50.1
pfu	<i>P</i> /%	0.03	0.07	0.10	0.14	0.17	0.21	0.24	0.28	0.31	0.35
polymerase	<i>F</i> /%	50.5	33.7	20.9	17.6	13.5	11.0	10.1	8.7	7.7	6.9

P is the probability that one template happens mismatch; *F* is the error rate of single molecule PCR.

2.2 针对单个碱基的突变

为基因测序进行的 PCR 扩增过程中, 单个碱基的错误不足以影响整个链的质量, 这时的错误率是指所有片段相同位置上错配碱基的概率. 我们假设扩增实验在片段不同位置上的作用相同, 那么每个碱基位置上发生突变的概率相等.

这里同样分两种假设进行讨论. 在第一种假设下, 是否发生错配完全按照产生的随机数进行模拟, 计算结果表明: 使用忠实性较低的 Taq 酶时, 以单个分子为模板, 单碱基的错误率为 17.9%; 如果更换忠实性较高的聚合酶, 碱基错误率会更低.

在起始循环中确有一个模板上发生突变的情况下, 如果使用忠实性低的 Taq 酶, 虽然单个碱基的错误率会超过 50%, 但是发生错配的概率却极低, 仅为 2×10^{-4} error/bp. 如果更换忠实性稍高的聚合酶, 或者略微增加初始模板个数, 不但发生突变的概率大大下降而且碱基错误率也下降到 50% 以下. 因此, 在为测序进行的单分子 PCR 实验中, 起始循环发生错配的概率极低, 可以不予考虑, 而略微优化实验条件, 都可以得到非常理想的实验结果. 我们初步的单分子 PCR 扩增实验结果也已证实了这一点.

2.3 实验条件的影响

从上面的计算可以看出, 实验中使用的聚合酶是非常重要的因素, 表 3 是在 500 bp 链长情况下, 序列从单个模板扩增到 10^{12} 时, 用不同的聚合酶诱导带来的错误率变化, 影响效果非常显著.

Table 3 The error rate of single molecule PCR using different polymerase

f/bp^{-1}	10^{-7}	10^{-6}	10^{-5}	10^{-4}
F	0.0016	0.0164	0.1525	0.8091

其他实验条件例如扩增效率、模板长度、pH 值、 Mg^{2+} 离子浓度等, 对结果也会有一定程度的影响, 限于篇幅, 不再一一讨论.

3 结 论

计算结果表明, 大量模板下 PCR 产物的错误率基本上不会影响实验结果. 而单分子 PCR 实验, 针对不同的实验目的和条件, 错误率对实验结果的影响不能一概而论.

当突变的定义针对整条链时, 也就是在单个碱

基突变检测等一类实验中, 如果使用忠实性较低的酶, 在起始循环中一旦发生错配, 实验产物将不能用作下一步的分析, 所以这类实验中应该尽量使用忠实性高的聚合酶, 以保证实验质量; 但当突变的定义针对单个碱基时, 例如在测序等实验中, 除了非常极端的情况下错误率会超过 50% 以外, 即使使用了忠实性低的酶, 实验结果也很理想.

同时从表 1 中可以看出, 相对于大量模板下产物的错误率, 少量模板 PCR 的错误率并未大幅增加, 说明模板数对产物错误率的影响很小, 所以一般情况下可以用单个模板分子进行 PCR 扩增. 当然实验条件会对错误率产生直接影响, 为了能够得到好的实验结果, 应该努力优化 PCR 反应条件, 比如使用忠实性高的酶, 努力提高扩增效率.

参 考 文 献

- 1 迪芬巴赫 C W, 德维克斯勒 G S. 黄培堂等译. PCR 技术实验指南. 北京: 科学出版社, 1998. 27 ~ 31
Dieffenbach C W, Dveksler G S. translated by Huang P T. PCR Primer: A Laboratory Manual. Beijing: Science Press, 1998. 27 ~ 31
- 2 Ling L L, Keohavong P, Dias C, *et al.* Optimization of the polymerase chain reaction with regard to fidelity: modified T7, Taq, and Vent DNA polymerase. PCR Methods and Applications, 1991, **1** (1): 63 ~ 39
- 3 Eckert K A, Kunkel T A. High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. Nucleic Acids Research, 1990, **18** (13): 3739 ~ 3744
- 4 Kwok S, Kellogg D E, McKinney N, *et al.* Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. Nucleic Acids Research, 1990, **18** (4): 999 ~ 1005
- 5 Keohavong P, Thilly W G. Fidelity of DNA polymerases in DNA amplification. Proc Natl Acad Sci USA, 1989, **86** (23): 9253 ~ 9257
- 6 Brownie J, Shawcross S, Theaker J, *et al.* The elimination of primer-dimer accumulation in PCR. Nucleic Acids Research, 1997, **25** (16), 3234 ~ 3241
- 7 方悦群, 王秉瑞. PCR 产物中无错误碱基模板最低比率的计算. 生物化学与生物物理进展, 1995, **22** (3): 275 ~ 277
Fang Y Q, Wang B R. Prog Biochem Biophys, 1995, **22** (3): 275 ~ 277
- 8 Nakano M, Komatsu J, Matsuura S, *et al.* Single-molecule PCR using water-in-oil emulsion. J Biotechnology, 2003, **102** (2): 117 ~ 124
- 9 Nakano H, Kobayashi K, Ohuchi S, *et al.* Single-step single-molecule PCR of DNA with a Homo-priming sequence using a single primer and hot-startable DNA polymerase. J Biosci Bioeng, 2000, **90** (4): 456 ~ 458
- 10 Ohuchi S, Nakano H, Yamane T. *In vitro* method for the generation of protein libraries using PCR amplification of a single DNA molecule and coupled transcription. Nucleic Acids Research, 1998, **26** (19): 4339 ~ 4346
- 11 Darren G. Monckton, Jeffery's A J. Minisatellite "isoallele" discrimination in pseudohomozygotes by single molecule PCR and variant repeat mapping. Genomics, 1991, **11** (2): 465 ~ 467

Error Rate in Single Molecule PCR *

WANG Guo-Hua¹⁾, LÜ Jun-Hong¹⁾, LEI Xiao-Ling¹⁾, LI Hai-Kuo²⁾,
LI Min-Qian¹⁾, CHEN Run-Sheng³⁾, FANG Hai-Ping¹⁾, HU Jun^{1,2)} **

¹⁾ Nanobiology and Nanomedicine Laboratory, Shanghai Institute of Applied Physics, The Chinese Academy of Sciences, Shanghai 201800, China;

²⁾ Bio-X Life Sciences Research Center, Shanghai JiaoTong University, Shanghai 200030, China;

³⁾ Molecular Biology Research Center, Institute of Biophysics, The Chinese Academy of Sciences, Beijing 100101, China)

Abstract Mismatches during amplification result in some mutations in the PCR products. The simulation indicates that the rate of mutation is quite low and has no influence on the consequent sequencing analysis when template DNA molecules are abundant, but it is not the case when the PCR amplification is performed with single or few molecules as templates. The error rates of single molecule PCR under different conditions showed that it is essential to select different polymerases according to different experimental purposes.

Key words single molecule PCR, error rate, fidelity, the number of template molecule

* This work was supported by a grant from The National Natural Sciences Foundation of China (10335070).

** Corresponding author. Tel: 86-21-59552394, Fax: 86-21-59552394, E-mail: junhu22@hotmail.com

Received: June 30, 2003 Accepted: September 28, 2003