

快。

#### 4 挑战和展望

双向电泳面临的挑战主要有: 提高蛋白溶解性; 增加显色灵敏度和在不影响分辨率情况下的上样量; 改善双向电泳结果的分析处理性能, 使其高通量、高精度; 建立完整的双向电泳数据库并整合解析所得信息。另外双向电泳只是蛋白质组研究开展的一个切入点, 而对结构蛋白质组, 药物蛋白质组等全面深入的研究必将是多学科共同发展的结果。质谱的检测灵敏线已从 50 pmol/L 降低到 100 fmol/L。蛋白质组研究需建立自动化平台, 如机械手切胶, 胶被切成 1mm<sup>3</sup> 小块, 并直接送入 96 孔板, 而后自动消化胶, 并送入质谱检测并对结果进行自动分析。与传统富集分离蛋白的色谱方法相比, 双向电泳技术可以认为是一种快速简便的分离纯化手段。即使在微流体, 蛋白质芯片蓬勃兴起的今天, 蛋白质的有效富集扩增仍是一个难题。

#### 参 考 文 献

1 News Natures, 1995; 378: 653

- 2 Wasinger VC *et al* Electrophoresis, 1995; 16: 1090
- 3 Blackstock WP *et al* Trend Biolo Technol, 1999; 17: 121
- 4 Neubauer G *et al* Proc Natl Acad Sci USA, 1997; 94: 385
- 5 Klöse J, Human Genetik, 1975; 26: 211
- 6 O Farrel PH, J Biochem, 1975; 20: 4007
- 7 Gasparic V *et al* 1975 Swedish patent No. 7514049
- 8 Bjellqvist B *et al* J Biochem Biophys Methods, 1982; 6: 317
- 9 Klöse J *et al* Electrophoresis, 1995; 16: 1034
- 10 O Farrel PZ *et al* Cell, 1977; 12: 1133
- 11 Gorg A *et al* Electrophoresis, 1995; 16: 1079
- 12 Banks RE *et al* Electrophoresis, 1999; 20: 689
- 13 Molloy MP *et al* Electrophoresis, 1999; 20: 701
- 14 Lollo BA *et al* Electrophoresis, 1999; 20: 854
- 15 U.S. Patent Application Serial, No 09/017. 284
- 16 Yan JY *et al* Electrophoresis, 1999; 20: 723
- 17 Gorg A *et al* Electrophoresis, 1999; 20: 712
- 18 Godova ZI *et al* Electrophoresis, 1999; 20: 952
- 19 Prasad SC *et al* Electrophoresis, 1999; 20: 1065
- 20 Young D S *et al* Clin Chem, 1982; 28: 737
- 21 Aderson L *et al* Clin Chem, 1984; 30: 1897
- 22 Charrier JP *et al* Electrophoresis, 1999; 20: 1075
- 23 Bobring C *et al* Electrophoresis, 1999; 20: 971
- 24 Jungblut PR *et al* Electrophoresis, 1999; 20: 2100 (2000-06-26 收稿)

## 体外突变新技术: DNA Shuffling 技术

杨海杰综述 夏宁邵审阅

厦门大学细胞生物学与肿瘤细胞工程重点实验室 (厦门, 361005)

**摘要** 体外 DNA shuffling 技术是一种快速、高效并且实用的分子定向进化技术。和以往的基因突变技术不同, 它是一种高通量的突变筛选技术, 可以对靶序列进行多次重组和选择, 利用该技术建立的突变文库 (嵌合文库) 具有容量大、多样性好等优点, 且更易于实现有益突变的积累。DNA shuffling 技术已广泛应用于酶、药物蛋白及其它功能蛋白的改造, 并取得了相当大的成功, 显示了该技术在蛋白质改造和分子育种方面的潜在应用价值。

**关键词** DNA shuffling; 基因突变; 蛋白质改造; 嵌合文库

基因突变技术是研究基因功能和改良蛋白质性能的一种重要手段。传统上应用于基因突变的方法主要有两种: 一是致错 PCR (error-prone PCR) 技术; 一是盒式突变 (cassette mutation) 技术, 即寡核苷酸介导的诱变技术。二者均利用 PCR 技术, 对克隆化的靶序列进行随机突变或定点突变。这两种技术均存在不足之处: 致错 PCR 所用的酶聚合能力差, 利用该技术所产生的点突变并非完全随机, 且中

性突变多, 此外致错 PCR 靶序列的长度有限, 大都不超过 0.5~1 kb; 盒式突变的突变效能往往取决于寡核苷酸序列的数目及长短, 而且还会出现统计学上的瓶颈效应, 如可能会放弃目前效果可能不是最好, 但却有着良好的进化前景的突变体。同时计算机模拟分析表明, 仅仅通过点突变, 对欲获得连续性序列进化所需的大量序列变化来说, 则显得太缓慢了。以上事实表明这两种方法仅适用于单循环的序

列精调,无法进行反复突变和连续筛选,无法实现突变的优势积累效应。与此相比,近几年发展起来的新体外突变技术——DNA shuffling (DNA 重排) 技术则在一定程度上满足了以上要求。

### 1 技术的建立

大多数生物体进化都来自于自然选择和性繁殖,其中在真核生物性重组便是生物进化的一种主要方式。早在 1990 年,Marion 和 Meyerhans 等人就已经发现在体外 PCR 中存在的 DNA 重组现象,并指出该重组现象是由 DNA 断裂或出现切口引起的。1994 年美国的 Stemmer 等以 LacZ  $\alpha$  基因为实验材料,用 DNase I 进行消化,得到 10~70 bp 大小的随机片段集合,在不加入引物的情况下,利用 PCR 得到了全长的 LacZ  $\alpha$  产物,从而以巧妙的设计验证了体外 PCR 中存在的 DNA 重组现象<sup>[1]</sup>。Stemmer 将该技术称为体外 DNA shuffling 技术,并于 1998 年申请了专利。由于该方法一定程度上模拟了生物体自然进化过程中减数分裂期等位基因间的 DNA 片段互换(有性交换,sexual exchange),故 DNA shuffling 又被称为有性 PCR (sexual PCR)。

### 2 技术的原理

DNA shuffling 技术是一种利用多次、反复选择和重组以获得具有预期性状的多聚核苷酸的方法。其原理(图 1)大致如下:先将由多种同源核苷酸序列组成的模板(初级文库)进行随机片段化,所产生的随机片段集合包含了至少 1000 种以上寡核苷酸,这些寡核苷酸来自不同的同源序列,具有不同的 3 末端,这一点是下一步随机重排的前提条件。利用 PCR 技术在不加入引物的情况下对这些随机化片段进行重新排布:在起初的几轮循环中,具有互补 3 末端的来自不同模板序列的寡核苷酸互为引物,扩增出一段较长的核苷酸产物;接着下一轮循环时,以上轮产物为模板,又将产生一段更长的核苷酸产物,随着扩增数递增,模板的不断变化,产物长度亦在不断递增,如此,经过多轮扩增,最终将获得多种低拷贝的全长的重排产物,这些集合的重排产物又被称为嵌合文库(突变文库)。对嵌合文库进行筛选,选择改良的突变体组成下一轮 shuffling 的模板(次级文库),重复上述步骤进行多次重排和筛选,直到最终获得性状比较理想的突变体。

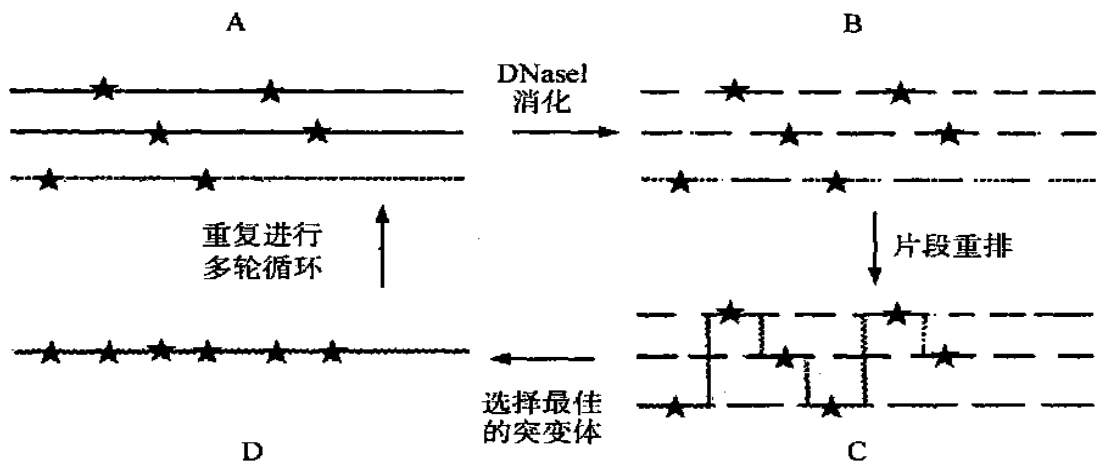


图 1 DNA shuffling 原理示意图

由多个含不同点突变的同源基因组成的模板集合(初级文库)经 DNase I 消化为片段集合。(B)为简化起见,所有的点突变均被认为是有益的。(C)模板在不断变化,随机小片段重排生成全长基因,一个功能改善、含 5 个点杂交(竖线表示)的重组基因被筛选出来。(D)将选择到的改良的突变体混合在一起,作为下一次 DNA shuffling 的模板(次级文库),进行多轮突变和筛选。

利用随机化片段的重新排列组合产生多样化的突变文库是 DNA shuffling 技术的精髓所在,这一点有点类似 assembly PCR<sup>[2]</sup>。其优点是:可以在体外实现同源序列间的重组,并可能实现那些对进化有利的点突变的优化组合,最终获得性状改良比较理想的突变体。

### 3 具体方法

#### 3.1 目的性状的确立和模板的来源

目的蛋白的改造必须针对一定的表型特征,即要针对特定的目的性状,包括结构和功能方面的特性。如荧光蛋白的荧光强度、最大波长及酶的催化能

力、热稳定性、底物的专一性等。此外,一些DNA 调控序列,如启动子、5'UTR、3'UTR 等也可以作为靶序列进行改造以改善其调控功能。模板可以由多个同源序列或同源基因(如家族基因)组成初级文库(primary library),也可以单个基因组成。后者在DNA shuffling 之前要先利用其它突变方法对单个基因进行突变,或在DNA shuffling 的同时借助聚合酶产生的点突变以产生多样性的突变产物。同源靶序列的长度在 50 bp~ 50 kb 之间,一般通过 PCR 获得,反应后应除去多余的引物,以提高重组效率。

### 3.2 随机片段化处理

初级文库(模板)用 DNase I 或多种内切酶处理后,获得随机片段集合。随机片段大小应在 20~ 500 bp 之间为宜。并对片段集合进行纯化,除去未消化的模板,以避免原序列的污染。

### 3.3 无引物的 PCR 扩增

在不加引物的情况下,将以上片段集合置于适当的 PCR 系统中进行 PCR 扩增。该步要求片段集合有较高的总量和浓度,同时还要求 PCR 反应有较高的循环数,以保证产生多样性足够好的嵌合文库(突变文库)。

### 3.4 加入引物的 PCR 扩增

上一步的 PCR 产物按一定比例稀释,同时加入两外侧引物,选用较少的循环数,如 15 轮,进行 PCR 扩增,便可获得大量的全长目的基因,即放大的嵌合文库(突变文库)。有时 PCR 产物中还可能存在全长目的基因的线形多拷贝聚合体,如 Stemmer 等对全长 2.7 kb 的 pUC19 质粒进行 DNA shuffling 时,得到的产物大多在 20 kb 以上,但用适当的内切酶消化后,即得到均一的全长目的基因<sup>[1]</sup>。

### 3.5 克隆、筛选、分析及多轮筛选

PCR 产物用适当的内切酶处理,克隆到表达载体中并转入宿主细胞(如大肠杆菌,噬菌体等)中进行表达,在特定的选择压力下,选择目的性状改良的突变体;将选择到的突变体混合在一起,作为下一次 DNA shuffling 的模板(次级文库),选择压力不断提高,重复 1~ 5 步骤进行多次选择和重组,直到最终得到性状改良明显的理想的突变体为止。

在进行 DNA shuffling 时,最重要的是选择合适的筛选模型,只有选择的模型恰当,才会减少工作量,快速地实现预期目标。如欲提高酶的热稳定性,应以不断递增的温度为选择压力,通过突变产物(酶)对底物的活性变化来筛选热稳定性改良的酶;如果改造对象是细菌的酶,还可以选择该酶缺陷的细菌为宿主细胞进行筛选。

## 4 应用

DNA shuffling 技术可以应用于蛋白质工程来改善蛋白质的结构和功能,提高蛋白质应用的有效性。目前该技术主要应用于酶的改良、药物蛋白的优化及其它领域。

### 4.1 酶的改良

酶的改良包括酶的诸多特性的改良,如酶的催化能力、酶的热稳定性、酶的专一性及可溶性表达等。

4.1.1 提高酶的催化能力 野生型的  $\beta$  内酰胺酶水解头孢噻污的能力低,在大肠杆菌中头孢噻污对该酶的最低抑制浓度(MIC)仅为  $0.02 \mu\text{g}/\text{ml}$ 。Stemmer 等人利用体外 DNA shuffling 技术来筛选高活性的 TEM-1  $\beta$  内酰胺酶<sup>[3]</sup>。第 3 轮得到的突变体 ST-1 的 MIC 达到  $320 \mu\text{g}/\text{ml}$ ,比原来的酶提高了 16000 倍,此外 ST-1 在大肠杆菌中的表达量比 wt  $\beta$  内酰胺酶也提高了 2 倍。而利用致错 PCR 法对 wt  $\beta$  内酰胺酶进行定向进化的对照组经过 3 轮选择后,最大的 MIC 值仅为  $0.32 \mu\text{g}/\text{ml}$ 。GUS 用甲醛或戊二醛固定组织后很容易失活,限制了 GUS 在组织固定时的应用。Ichiro 等人以野生型 GUS 基因为模板筛选到一高活性的克隆 GUS<sup>[4]</sup>。对 wt-GUS 和 GUS<sup>+</sup> 的细菌抽提物用戊二醛进行处理,结果发现:wt-GUS 用 0.04% 的戊二醛处理后,GUS 酶的活性丧失  $99.6 \pm 0.24\%$ ,GUS<sup>+</sup> 用 0.2% 的戊二醛处理相同时间后,仍保持  $78.1 \pm 0.69\%$  的活性。应用 DNA shuffling 技术,得到了对固定剂不敏感的 GUS<sup>+</sup>,扩展了 GUS 染色技术的应用范围和能力,同时使得多染色成为可能。

4.1.2 提高酶的热稳定性 在较高温度时(如 39 $^{\circ}\text{C}$ ),FLP 的活性很低,限制了该酶在基因工程中的应用。Frank 等人利用 DNA shuffling 技术来改善重组酶 FLP 的热稳定性,8 轮选择后得到一热稳定性理想的突变体 FLPe,42 $^{\circ}\text{C}$  时在 *E. coli* 内仍具较高的活性,此外 FLPe 在体外及哺乳动物细胞内的热稳定性也有很大程度的提高,同时还意外地发现 FLPe 易于纯化,纯度可达 95%,而野生型 FLP 的纯化则十分困难<sup>[5]</sup>。

4.1.3 改变酶对底物的专一性 Zhang 等运用 DNA shuffling 技术,将 LacZ  $\beta$  半乳糖苷酶改造为高效的  $\beta$  岩藻糖苷酶<sup>[6]</sup>。以 LacZ  $\beta$  半乳糖苷酶为模板,经过 7 轮筛选后得到的改良酶和野生型  $\beta$  半乳糖苷酶相比,对邻硝基苯酚岩藻糖/邻硝基半乳糖苷的相对专一性提高了 300 倍,改良的岩藻糖苷酶对

底物的  $K_{cat}/K_m$  值比大肠杆菌自身的岩藻糖苷酶提高了 10~20 倍。同时值得一提的是: 该酶的分子量为 109 kD, 利用其它突变方法很难以对如此大的蛋白进行改造, 而 109 kD 的蛋白仍在高通量突变技术——DNA shuffling 技术的正常改造范围之内。

#### 4.2 药物蛋白的优化

DNA shuffling 还可以用来优化药用蛋白, 提高药用蛋白的临床应用价值。目前该技术已广泛应用于如细胞因子、基因工程单链抗体等。

$\alpha$ -干扰素是螺旋束超基因家族细胞因子的一员, 虽然这些蛋白在许多疾病治疗上具有药用价值, 但用于治疗, 这些蛋白未必达到了最优化。Chang 等人将 20 多个人类 IFN- $\alpha$ (氨基酸同源性在 85%~98%) 基因作为初级文库进行 DNA shuffling, 筛选到 3 个突变体 IFN-CH 2 1-2 3。其中 IFN-CH 2 3 抗病毒活性为 Hu-IFN- $\alpha$ 1 的 185 倍, Hu-IFN-2 $\alpha$  的 285000 倍及小鼠自身活性最高的 Mu-IFN- $\alpha$ 4 的 3.5 倍<sup>[7]</sup>。

单链抗体(scFv)基因分子比较小, 故越来越广泛地应用于临床诊断及治疗疾病。但由于以往利用噬菌体展示筛选 scFv 缺少选择压力, 故不能保证 scFv 折叠性、产量、稳定性同时实现优化。Sabine 等利用 DNA shuffling 技术, 获得的 scFv (4d5Flu) 突变体热稳定性热动力学常数比原 scFv 提高了 4 kcal/mol, 在 *E. coli* 中的产量提高了 2 倍, 与抗原的结合常数则提高了 20 倍<sup>[8]</sup>。

#### 4.3 其它

绿色荧光蛋白(GFP)基因作为一种荧光标记基因, 已广泛地应用于基因调控的研究。但在真核生物中 GFP 的表达效果欠佳, 无法获得整个细胞的荧光信号。Stemmer 等人以合成的 GFP 基因('wtGFP')为模板, 进行 3 轮筛选, 分别得到 cycle2 和 cycle3 两个改良体。结果显示: 在 *E. coli* 中 cycle2 的荧光强度分别比 Clontech 公司商品化的 GFP 和 'wtGFP' 提高了 22 倍和 7 倍, cycle3 的荧光强度则分别提高了 44 倍和 15~17 倍。cycle2 和 cycle3 克隆斑的生长速度比 'wt' GFP 提高了 2~3 倍。在哺乳动物细胞中表达, 发现 cycle3 的整个细胞的荧光信号是 'wt' GFP 的 45 倍。对 cycle3 和 'wt' GFP 在 *E. coli* 的可溶性表达情况进行了比较, 发现 cycle3 主要以可溶性形式存在, 而后者主要以包含体形式存在<sup>[9]</sup>。

Xirodimas 等人用多个编码对热变性耐受的 p53 蛋白的基因作为靶序列进行 DNA shuffling, 使 P53 的热稳定性大大提高, 且证实了 P53 中心区的

20 个氨基酸(101~120)与热稳定性相关<sup>[10]</sup>。Juarez 等利用该技术使反式激活蛋白 B<sub>3</sub>N ifA 与增强子的结合力提高<sup>[11]</sup>。此外, DNA shuffling 技术还被用来改善细菌的解毒能力。Cramer 利用该技术改造砷抗性操纵子, 使其在固体培养基和液体培养基中抗砷的 MIC 分别提高了 63 倍和 14 倍<sup>[12]</sup>; Stemmer 改造钙抗性操纵子, 使其在细菌中的解毒能力提高了 7 倍。

## 5 改进

DNA shuffling 技术的建立源于同源序列间的重排, Ostermeyer 等人对此进行了改进, 建立了递减法建立杂交酶技术(incremental truncation for creation of hybrid enzymes, ITCHY), 使得非同源性序列间也可以发生重排。以核酸外切酶 III 代替 DNase I 对靶序列进行消化, 由于核酸外切酶 III 的 3' 外切核酸酶作用, 因此得到的随机片段文库(incremental truncation libraries, ITLs)理论上包括了靶序列 DNA 的单碱基对删除的各种情况, 使得在较低的复性温度下可以实现非同源区的融合。

人和鼠 L-1 $\beta$  基因同源区段平均长度为 41 bp, 不适宜采用常规的 DNA shuffling 技术。为此 Stemmer 对 DNA shuffling 做了改进: 在进行无引物的 PCR 扩增时, 前 15 轮用 klenow 酶替代 Taq 酶, 后 22 轮用 Taq 酶。随机对 9 个克隆测序, 发现人和鼠 L-1 $\beta$  基因间的重组共发生了 17 处, 其中一个重组建立在两个碱基同源的基础上, 表明只要重组位点两侧存在的同源性, 异源重组也是可能实现的<sup>[11]</sup>。

在进行 DNA shuffling 时, 一般使用 ds-DNA 做模板, 所产生的背景比较严重, 嵌合率比较低。为了提高家族基因 DNA shuffling 时的嵌合形成率, Heita 等人用 ss-DNA 替代 ds-DNA 作为模板进行 DNA shuffling<sup>[15]</sup>。实验中制备了儿茶酚 2, 3-二氧酶同源基因 nahH 和 LyIE 的 ss-DNA, xyIE 链同 nahH 链互补。运用 ss-DNA 作模板得到的嵌合基因(nahH 和 LyIE 间的杂交体)比率高达 14%, 而以 ds-DNA 作为模板的嵌合发生率低于 1%, 且利用 ss-DNA 做模板得到的嵌合基因所编码的蛋白热稳定性更好。

## 6 优点

和传统的突变技术相比, DNA shuffling 技术具有诸多优点: 它是一种高通量的突变、筛选技术。利用该技术所建立的突变文库容量大、多样性

好,为得到理想的突变体,一般要上万个克隆进行筛选。可以进行反复筛选,通过选择压力的递增,积累有益突变,以寻求最佳候选分子。利用该技术筛选到的突变体中性突变少,且易利用分子回交除去中性突变,这对于分析对目的性状改良所必须的点突变大有帮助。可以实现同源序列间的大片段(如可能的功能区段)交换,为研究蛋白质的功能区提供了新的思路。适用的靶序列范围宽,长度可以达 50 kb。故有可能利用该技术改造病毒基因组,研制病毒活疫苗。大量的研究表明,该技术并不针对蛋白质的一种特性进化,可以同时实现目的蛋白的多种特性的共进化。和以往的对分子的合理设计相比,该技术也体现了一定的优越性。如对天冬氨酰氨基转移酶改造时,该酶的活性提高了 30 倍,分析发现其中 6 个氨基酸对活性的提高是必要的,但仅有 1 个氨基酸位于酶活性中心<sup>[16]</sup>。该研究结果表明如果仅仅依靠合理设计,则很难考虑到环境氨基酸在结构和功能上的协同作用。

DNA shuffling 技术亦有其局限性,如利用该技术所建立的突变文库并不能得到生物进化的所有材料,另外对于那些难以或无法建立筛选模型的蛋白质来说,就不能利用该技术进行改造。这就要求在进行 DNA shuffling 的同时结合其它突变技术,并且发展更多更好的筛选模型,以提高突变的有效性并拓宽 DNA shuffling 技术的应用范围。

## 7 展望

DNA shuffling 技术是创造新表型的强大发动机,近年来,体外 DNA shuffling 技术在酶、药用蛋白等领域的成功运用充分验证了它是一种快速、高效的分子定向进化手段,同时也开辟了 DNA shuffling 技术在生物分子工程领域里的新天地。目前又建立了体内 DNA shuffling 技术<sup>[17]</sup>,人们可以结合

体内和体外两种方法及对分子的合理设计,进一步提高进化的多样性和精确性。随着越来越多的筛选模型的建立和 DNA shuffling 的不断优化,该技术的应用范围也将不断拓宽:在不久的将来,人们也许会借助该技术对古生物化石的 DNA 碎片进行重排,进而获得有意义的古生物遗传物质;或者用来改造工业用酶,提高其商业价值;或提高细菌的解毒能力,用以处理环境污染问题;或改良蛋白药物,提高其临床应用价值;或改良疫苗,使其有效性提高;甚至可以用来研究蛋白的亚功能区。这一切都预示着 DNA shuffling 技术在蛋白质工程和分子育种方面的广阔的应用前景。

## 参 考 文 献

- 1 Stemmer W P *et al* Proc Natl Acad Sci U S A, 1994; 91: 10747
- 2 Stemmer W P *et al* Gene, 1995; 164: 49
- 3 Stemmer W P. Nature, 1994; 370: 389
- 4 Matsumura I *et al* Nat Biotechnol, 1999; 17: 696
- 5 Buchholz F *et al* Nat Biotechnol, 1998; 16: 657
- 6 Zhang JH *et al* Proc Natl Acad Sci U S A, 1997; 94: 4504
- 7 Chang CCJ *et al* Nat Biotechnol, 1999; 17: 793
- 8 Jung S *et al* J Mol Biol, 1994; 294, 163
- 9 Cramer A *et al* Nat Biotechnol, 1996; 14: 315
- 10 Xirodimas DP *et al* J Biol Chem, 1999; 274: 28042
- 11 Caimano MJ *et al* Infect Immun, 2000; 68: 1574
- 12 Cramer A *et al* Nat Biotechnol, 1997; 15: 436
- 13 Ostemeier M *et al* Biorg Med Chem, 1999; 7: 2139
- 14 Ostemeier M *et al* Nat Biotechnol, 1999; 17: 1205
- 15 Kikuchi M *et al* Gene, 2000; 243: 133
- 16 Yano T *et al* Proc Natl Acad Sci U S A, 1998; 95: 5511
- 17 Jirholt P *et al* Gene, 1998; 215: 471

(2000-05-11 收稿)