

体外分子定向进化研究进展*

徐卉芳 张先恩** 张用梅

(中国科学院武汉病毒研究所, 武汉 430071)

A. E. G. CASS

(Imperial College of Science, Technology and Medicine, South Kensington, London SW7 2AY, UK)

摘要 体外定向进化作为近几年发展起来的一种蛋白质改造新策略, 可以在未知目标蛋白三维结构信息和作用机制的情况下, 通过对编码基因的随机突变、重组和定向筛选, 获得具有改进功能或全新功能的蛋白质, 使几百万年的自然进化过程在短期内得以实现, 因而是发现新的生物活性分子和反应途径的重要方法, 已在短短几年内取得了令人瞩目的成就。

关键词 体外分子进化, 蛋白质工程, DNA shuffling, 渐进切割法产生杂和酶 (ITCHY), 基因组进化

学科分类号 Q81

大量定点基因突变实验表明, 蛋白质功能和性质的改变来自于许多小的内部修饰的积累, 这些小的修饰或突变分布于较大的序列空间内。人们试图利用已有的结构生物学信息对蛋白质进行合理设计 (rational design), 但蛋白质结构的复杂性极大地增加了合理设计的难度。更何况, 对于大多数要改造的蛋白质来说, 我们并不清楚其三维结构信息, 不能进行合理设计。而近年来发展的分子定向进化 (molecular directed evolution) 策略属于蛋白质的非合理设计范畴, 它不需要事先了解蛋白质的三维结构信息和作用机制, 而是在体外模拟自然进化的过程 (随机突变、重组和选择), 使基因发生大量变异, 并定向选择出所需性质或功能, 从而在几天或几周内实现自然界需数百万年才能完成的事情。

1 体外分子定向进化的策略

定向进化第一步是由一个靶基因或一群相关的家族基因起始创建分子多样性 (突变和/或重组); 然后对该多样性文库的基因产物进行筛选, 那些编码改进功能产物的基因被利用来继续下一轮进化; 重复这个过程直到达到目标。该进化策略有以下三个显著特征: a. 进化的每一关键步骤都受到严密控制。b. 除修饰改善蛋白质已有特性和功能外, 还可引入一个全新的功能, 来执行从不被生物体所要求的反应; 甚至为生物体策划一个新的代谢途径^[1]。c. 能从进化结果中探索蛋白质结构和功能的基本特征。

2 体外定向进化技术的演变和发展

2.1 错误倾向 PCR

错误倾向 PCR (error prone PCR)^[2] 是一种简便快速地在 DNA 序列中随机制造突变的方法。其基本原理是通过改变传统 PCR 反应体系中某些组分的浓度, 或使用低保真度的 DNA 聚合酶等, 使碱基在一定程度上随机错误引入而创造序列多样性文库。本法的关键在于选择适当的突变频率, 一般为每个基因 2~5 个碱基替换。另外, 迭代错误倾向 PCR 可使小的有益突变累积而产生大的提高。尽管这种方法已有不少成功的例子^[2,3], 但毕竟遗传变化只发生在单一分子内部, 属于无性 PCR 进化范畴。

2.2 DNA shuffling

DNA shuffling 由美国 Stemmer 于 1994 年首次提出^[4]。该法是对一组基因群体 (进化上相关的 DNA 序列或曾筛选出的性能改进序列) 进行重组创造新基因的方法。因为该法在 DNA 片段组装过程中也可能引入点突变, 所以它对从单一序列指导进化蛋白质也是有效的。其基本原理如图 1 所示。这种方法产生的多样性文库, 可以有效积累有益突变, 排除有害突变和中性突变, 同时也可实现目的蛋白多种特性的共进化^[5]。正是由于该技术包括

* 中国科学院与英国皇家学会联合资助项目 (Q174)。

** 通讯联系人。

Tel: 027-87641492, E-mail: x. zhang@pentium.whiov.ac.cn

收稿日期: 2002-01-07, 接受日期: 2002-03-04

了 DNA 重新组装的过程, 使它与以往的诱变技术有了质的不同. 当 DNA shuffling 用来重组一套进化上相关的基因时, 又被称为 family shuffling, 一个突出的例子是 shuffling 4 种头孢菌素酶基因, 酶活增加了 270~540 倍, 而单基因 shuffling 只增加了 8 倍^[6].

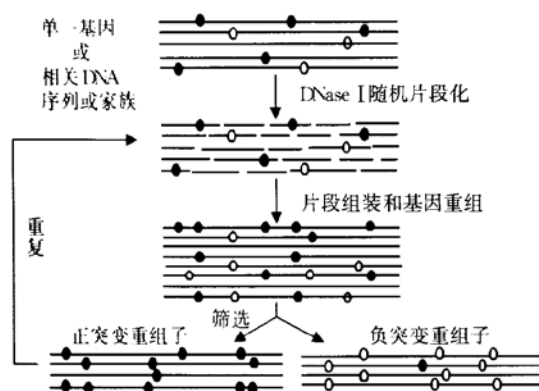


Fig.1 DNA shuffling procedure

图1 DNA shuffling 原理

○: 负突变表型; ●: 正突变表型.

2.3 StEP 重组

StEP (staggered extension process) 重组由 Zhao 等^[7]提出, 是一种简化的 DNA shuffling 技术. 它不是由短片段组装全长基因, 而是在 PCR 样反应中, 将含不同点突变的模板混合; 随之进行多轮变性、短暂复性/延伸反应, 在每一轮 PCR 循环中, 那些部分延伸的片段可以随机地杂交到含不同突变的模板上继续延伸, 由于模板转换而实现不同模板间的重组, 这样重复直到获得全长基因片段, 重组的程度可以通过调整反应时间和温度来控制. 该方法在单一试管中进行, 是定向进化的又一创新.

2.4 随机引物体外重组法

随机引物体外重组法 (random-priming *in vitro* recombination, RPR)^[8]是用一套随机序列引物, 产生互补于模板不同位置的短 DNA 片段库 (由于碱基错配, 这些短 DNA 片段也含有少量的点突变). 然后进行类似于 DNA shuffling 的全长基因装配反应, 获得多样性文库. 与 DNA shuffling 相比, RPR 具有以下优点: a. 可用单链 DNA 或 mRNA 为模板. 且对模板量要求少, 大大降低了亲本组分, 便利筛选. b. 克服了 DNA shuffling 中 DNase I 所具有的序列偏爱性, 保证了子代全长基因中突变和交叉的随机性. c. 片段组装体系与片段

合成体系缓冲系统可兼容, 组装前无需纯化操作. d. 随机引发 DNA 合成不受模板 DNA 长度的限制, 便利了小肽的改造.

2.5 过渡模板随机嵌合生长

过渡模板随机嵌合生长 (random chimeragenesis on transient templates, RACHITT)^[9]是与 DNA shuffling 概念上明显不同的、改进的基因家族重组技术. 它不包括热循环、链转移或交错延伸反应, 而是将随机切割的基因片段杂交到一个临时 DNA 模板上进行排序、修剪、空隙填补和连接 (图2). 其中的悬垂切割步骤使短片段 (比 DNase 消化片段还短) 得以重组, 明显提高了重组频率; 如果在片段重组前后采用错误倾向 PCR 还可引入额外点突变. Coco 等^[9]首次报道此法改造二苯并噻吩单加氧酶, 产生的嵌合文库平均每个基因含 14 个交叉, 重组水平比 DNA shuffling 类方法 (1~4 个交叉) 高出几倍; 并且可在短至 5 bp 的序列同一区内产生交叉. 这种高频率、高密度的交叉水平是 DNA shuffling 所难以达到的.

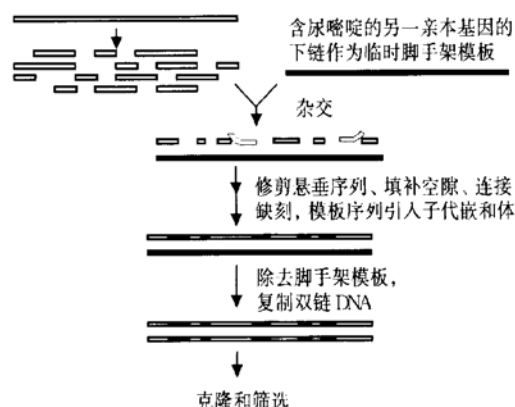


Fig 2 The general procedure of RACHITT^[9]

图2 RACHITT 法基本程序^[9]

2.6 酵母增强组合文库

酵母增强组合文库 (combinatorial libraries enhanced by recombination in yeast, CLERY) 是一个真核基因家族 shuffling 策略^[10]. 其原理是将体外 DNA shuffling 程序与接下来的酵母体内重组结合起来, 构建高丰度低亲本水平的重组文库: 直接以含目的基因的质粒作为体外 shuffling 的模板; shuffling 后基因产物与线性化的酵母表达载体共转化酵母细胞启动体内重组事件, 同时表达功能性重组子直接用于筛选. Truan 等^[10]用该法重组人细胞色素 P450 1A1 和 1A2, 所得文库含 86% 的嵌合基

因,大大提高了文库丰度;另外,用单链 DNA 做 shuffling 的模板可进一步降低文库中的亲本比率.该法为蛋白质结构/功能研究、多组分真核复合酶活力的调控提供了一个新的有力工具.

2.7 渐增切割法产生杂和酶

尽管 DNA shuffling 介导的重组已成为定向进化创建高质量序列多样性的重要工具,但它们通常不能用于重组同源性低于 70%~80% 的序列.而自然界大多数同源序列具有比这个数目低得多的相似性.为了解决这个问题, Benkovic 研究组建立了渐增切割法产生杂和酶 (incremental truncation for the creation of hybrid enzymes, ITCHY) 及 Thio-ITCHY (α -硫代磷酸核苷酸介导的 ITCHY) 方法来产生不依赖于 DNA 序列同源性的重组文库. ITCHY^[11] 的基本原理是:控制核酸外切酶 III 的切割速度不大于 10 碱基/min,然后间隔很短时间连续取样终止反应,以获得一组依次有一个碱基缺失的片段库;然后将基因 A 的一组随机长度的 5' 端片段与基因 B 的一组随机长度的 3' 端片段随机融合产生杂合基因文库.但由于该方案冗长繁琐, Lutz 等^[12] 又提出了 Thio-ITCHY 方法:首先将待重组的两基因串联在同一载体上,然后通过 PCR 反应或单链模板引伸反应将 α -硫代磷酸核苷酸随机引入产物中,由于其抗核酸外切酶 III 的切割,接下来的切割反应会在此处随机终止,连接切割产物即产生随机交叉文库.该法操作便利,同时也可引入随机点突变.此外, ITCHY/Thio-ITCHY 不依赖 DNA 序列同源性,可在非序列同一区内产生活性融合子.迭代 ITCHY 或对其文库进行 shuffling 还可以创造多个交叉.

2.8 不依赖序列同源性的蛋白质重组

尽管上述 ITCHY 可以进行不依赖于序列同源性的重组,但它是一个随机长度片段与另一个随机长度片段相融合,结果文库中基因长度不再保守,子代中功能杂合子的比率很低.要想提高功能杂合子所占比例,须使交叉发生在具有相似结构环境的位置,但由于缺乏足够的序列相似性,同源重组不能发挥作用,因此 Sieber 等^[13] 提出了不依赖序列同源性的蛋白质重组 (sequence homology-independent protein recombination, SHIPREC) 方法:通过琼脂糖凝胶电泳回收单基因长度随机片段,保证了子代嵌合体长度的保守性,使交叉保持了适当的序列匹配,主要发生在结构上相关的位点,交叉点处的两个氨基酸仍处于它们在亲本蛋白质结构中的位置,

从而提高了文库中阳性克隆的比例.该法可以在低序列同一性甚至无序列同一性的同源蛋白质间创造具有单交叉的杂合蛋白质组合文库.迭代 SHIPREC 可以产生多个交叉.

以上阐述了定向进化创建序列多样性文库的各种新方法.但任何一种方法都不是万能的,在实际应用中,应针对具体问题,选择合适的方法或方法组合,以达到事半功倍的效果.

2.9 文库筛选策略

一旦多样性文库构建好后,定向进化实验成败与否的关键则在于“要有针对目的改造特征的高效筛选方案”.近几年,与创建多样性重组文库的发展相适应,在高流通量和超高流通量筛选技术方面也取得另人瞩目的成就^[14~16].其中核糖体展示技术与 mRNA 展示技术^[14],由于在体外无细胞翻译体系中进行,不受细胞转化效率的限制,大大提高了文库容量和筛选通量 ($10^{12} \sim 10^{14}$).它们的原理是通过筛选靶蛋白-核糖体-mRNA 三元复合物或靶蛋白-mRNA 二元复合物,将基因型与表型直接偶联起来,并利用 mRNA 的可复制性,使靶基因(蛋白)得到有效富集.利用此法进化单链抗体可变区 scFv 片段,获得亲和力提高 40 倍的变异株^[14].

其他高通量筛选方法,如细胞表面展示技术和噬菌体表面展示技术;将靶活力与转录信号相偶联的三杂交体系^[15];以发光信号为指示的反射增进系统^[16];利用荧光信号的荧光共振能量转换仪 (FRET),结合荧光激活细胞筛选仪 (FACS),每小时可筛选 60 000 个细胞^[15].最近, Ghadessy 等^[17] 还报道了一种隔室化自复制技术,通过一个只复制其自身基因的聚合酶组成的简单反馈环将体外进化与筛选有机结合起来,使阳性克隆得到快速、高效的富集,获得热稳定性提高 11 倍,肝素抗性提高 130 多倍的 Taq DNA 聚合酶突变株.

在硬件设备方面, 1536 和 3456 孔板以及多通道多波长检测仪的出现、每秒钟分配几千滴皮升级样品的非接触式压电配样仪的问世^[16] 等都大大提高了样品处理速度.

3 体外定向进化的应用及发展前景

体外定向进化大大加速了人类改造蛋白质原有功能和开发新功能的步伐.至 2000 年底就已有 30 多种药用蛋白质经定向进化被改进.目前,已报道的定向进化成就主要在以下几方面:生物分子活性

和稳定性的改善^[6]、在非天然环境下的活性和稳定性^[3]、抑制剂或抗生素抗性^[4]、新功能的开发和底物范围及底物特异性的改变^[15]、新型疫苗和药物分子的发现^[14]、抗体亲和力的提高^[14]、新的代谢途径的开发^[15]，甚至预测自然进化中新突变的出现^[18]等。本文作者对大肠杆菌碱性磷酸酶进行体外定向进化，获得比野生型酶活力提高 40 倍的突变株（待发表）。美国加州理工学院 Arnold 教授领导的研究组是定向进化领域的权威实验室，他们不仅建立了一些技术和方法^[7,8,13]，而且在阐明分子进化的过程和机制方面也做出了贡献^[5]。Stemmer 本人也携带技术注册了一家公司——Maxygen (<http://www.maxygen.com>)，大量制造工业和医疗用途的功能改进产品。该公司最近还提出所谓全基因组 shuffling 的概念，并正在进行改造微生物全基因组的工作。也许，不久的将来，以该技术为支撑，改造动植物基因组也将成为可能。

总之，体外分子定向进化策略证明进化可以发生在自然界，也可以发生在试管中。它的出现开创了蛋白质工程的新纪元，也加速了我们对生命过程的认识和理解。同时这项技术也正带来巨大商机。

致谢 本文撰写过程中曾与赵惠民博士进行有益的讨论。

参 考 文 献

- 1 Zhang J-H, Dawes G, Stemmer W P C. Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. *Proc Natl Acad Sci USA*, 1997, **94** (9): 4504~ 4509
- 2 Chen K, Arnold F H. Enzyme engineering for nonaqueous solvent: random mutagenesis to enhance activity of subtilisin E in polar organic media. *Bio/Technology*, 1991, **9**: 1073~ 1077
- 3 You L, Arnold F H. Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide. *Protein Eng*, 1996, **9** (1): 77~ 83
- 4 Stemmer W P C. Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature*, 1994, **370** (4): 389~ 391
- 5 Moore J C, Jin H M, Kuchner O, *et al.* Strategies for the *in vitro* evolution of protein function: enzyme evolution by random recombination of improved sequences. *J Mol Biol*, 1997, **272** (3): 336~ 347
- 6 Cramer A, Raillard S A, Bermudez E, *et al.* DNA shuffling of a family of genes from diverse species accelerates evolution. *Nature*, 1998, **391** (3): 288~ 291
- 7 Zhao H, Giver L, Shao Z, *et al.* Molecular evolution by staggered extension process (STEP) *in vitro* recombination. *Nat Biotechnol*, 1998, **16**: 258~ 261
- 8 Shao Z, Zhao H, Arnold F H. Random-priming *in vitro* recombination: an effective tool for directed evolution. *Nucleic Acids Res*, 1998, **26** (2): 681~ 683
- 9 Coco W M, Levinson W E, Crist M J, *et al.* DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat Biotechnol*, 2001, **19**: 354~ 359
- 10 Abecassis V, Pompon D, Truan G. High efficiency family shuffling based on multi-step PCR and *in vivo* DNA recombination in yeast. *Nucleic Acids Res*, 2000, **28** (20): e88
- 11 Ostermeier M, Shim J H, Benkovic S J. A combinatorial approach to hybrid enzymes independent of DNA homologv. *Nat Biotechnol*, 1999, **17**: 1205~ 1209
- 12 Lutz S, Ostermeier M, Benkovic S J. Rapid generation of incremental truncation libraries for protein engineering using α -phosphothioate nucleotides. *Nucleic Acids Res*, 2001, **29** (4): e16
- 13 Sieber V, Martinez C A, Arnold F H. Libraries of hybrid proteins from distantly related sequence. *Nat Biotechnol*, 2001, **19**: 456~ 460
- 14 Amstutz P, Forrer P, Zahnd C, *et al.* *In vitro* display technologies: novel developments and applications. *Curr Opin Biotech*, 2001, **12**: 400~ 405
- 15 Soumillion P, Fastrez J. Novel concepts for selection of catalytic activity. *Curr Opin Biotech*, 2001, **12**: 387~ 394
- 16 Hertzberg R P, Pope A J. High-throughput screening: new technology for the 21st century. *Curr Opin Chem Biol*, 2000, **4**: 445~ 451
- 17 Ghadessy F, Ong J L, Holliger P. Directed evolution of polymerase function by compartmentalized self-replication. *Proc Natl Acad Sci USA*, 2001, **98** (8): 4552~ 4557
- 18 Yano T, Kagamiyama H. Directed evolution of ampicillin-resistant activity from a functionally unrelated DNA fragment: A laboratory model of molecular evolution. *Proc Natl Acad Sci USA*, 2001, **98** (3): 903~ 907

Recent Advances in *In vitro* Molecular Directed Evolution^{*}

XU Hui-Fang, ZHANG Xian-En^{**}, ZHANG Yong-Mei

(Wuhan Institute of Virology, The Chinese Academy of Sciences, Wuhan 430071, China)

A. E. G. CASS

(Imperial College of Science, Technology and Medicine, South Kensington, London SW7 2AY, UK)

Abstract *In vitro* molecular directed evolution, based on the combination of random mutagenesis, recombination and high throughput screening, is a new strategy of improving protein properties. With this technology, the properties of the target proteins, such as specificity, catalytic activity and affinity can be tuned without the knowledge of 3-D structure information and function mechanisms. Recently, the methodology originating from error prone PCR and DNA shuffling has been greatly developed by many new techniques. Here, based on the literature review and the experience of authors, the recent developments have been reviewed.

Key words *in vitro* directed evolution, protein engineering, DNA shuffling, ITCHY, genomic evolution

^{*} This work was supported by a grant from The Joint Project Between The Royal Society of United Kingdom and The Chinese Academy of Sciences (Q174).

^{**} Corresponding author. Tel: 86-27-87641492, E-mail: x.zhang@pentium.whiov.ac.cn

Received: January 7, 2002 Accepted: March 4, 2002