

几种定向进化技术的比较及文库构建策略*

贾向东 陈德富** 陈喜文 郭少影

(南开大学生物化学与分子生物学系 天津 300071)

摘要 蛋白质定向进化在蛋白质工程中取得了令人瞩目的成就,其核心技术——随机突变库构建技术已成为近年来体外定向进化研究的热点。在概述定向进化基本原理基础上,对几种随机突变技术进行了介绍、分类和比较,并对突变库的特征及构建策略予以分析和描述。

关键词 定向进化 随机突变库 碱基错配 DNA 重组

纷繁复杂的蛋白质构成了生命的全部,它们在漫长的自然进化过程中产生和发展,自然界赋予了其精美的结构和独特的功能。但随着生物技术的发展,人们已不满足于自然进化的恩赐,希望将漫长的蛋白质改造过程缩短到数年、数月、甚至数天,以满足人们特殊的需要。随机突变技术和定向筛选技术所组成的定向进化技术的出现,是实现这一愿望的有效途径,其原理是对目的基因进行突变、重组、表达和筛选,在试管内模拟蛋白质的自然进化过程,获取人们需要的、具有特定性质和功能的蛋白质^[1]。该技术近年来得到了迅猛发展,据统计,仅 Nature(含系列)杂志每年以近 100 篇(2001 年 99 篇,2002 年 94 篇)的篇幅报道其最新的研究成果(<http://www.nature.com>)。鉴于目前报道的随机突变技术种类繁多,为便于选择,我们根据其突变机理的差异对它们进行分类、概述和比较。

1 基于 PCR 错配而实现的随机突变

1.1 高错误倾向 PCR

高错误倾向 PCR (error prone PCR, EP-PCR) 是利用低保真度 Taq DNA 聚合酶,或者改变 PCR 反应体系的条件,在新链 DNA 聚合过程中随机引入错配碱基,经多轮 PCR 扩增,构建序列多种多样的突变库。由于不改变基因长度,突变频率控制在适度范围,能有效地获得有益突变体^[2]。尽管 EP-PCR 只能产生点突变,但其基本原理已广泛应用于蛋白质定向进

化改造中。

1.2 一分子 PCR

日本 Ohuchi 等^[3]首次建立了一分子 PCR (single molecular PCR, SM-PCR) 技术,将 DNA 模板稀释到每个 PCR 反应体系中仅含一个 DNA 模板分子,属于一种极端条件下的高错误倾向 PCR。实验利用绿色荧光蛋白基因为模式分子,在 Taq 聚合酶标准错配率为 0.8×10^{-5} /碱基/循环时,由于反应条件偏离得到 2.5×10^{-5} /碱基/循环的错配频率,提高了突变效率。另外,如果采用高保真度 DNA 聚合酶进行 SM-PCR,结合体外无细胞蛋白质表达系统,可以对突变库实现快速高通量筛选,在定向进化中也具有良好的应用前景^[4]。

2 基于 DNA 同源重组而实现的随机突变

美国人 Stemmer 基于 DNA 同源重组原理于 1994 年首先提出了 DNA 洗牌技术,随后由该技术发展起来的新的随机突变技术不断被报道。

2.1 有性 PCR

有性 PCR (Sexual PCR) 是对一组相关基因用 DNase I 或超声波进行切割产生随机大小的 DNA 片段,再用无引物 PCR 将其连接成为接近目的基因长度的 DNA 分子,再利用基因两端序列为引物扩增全长基因^[5]。突变库经历了 DNA 片段的重新组装,与高错误倾向 PCR 有本质的不同。该方法由多个亲本参与进化,引入了重组、缺失、重复等多种突变类型,并且可以迅速积累有益突变,使表达蛋白的平均活性明显提高。

2.2 随机引物体外重组

随机引物体外重组 (random priming *in vitro*

收稿日期:2003-06-02 修回日期:2003-10-22

* 教育部留学回国人员科研启动基金资助[教外司留(2003)406]

** 通讯作者,电子信箱: chendefu@nankai.edu.cn

recombination, RPR) 是用随机引物在 DNA 模板上扩增 DNA 片段,再用类似有性 PCR 的方法组装成全长基因^[6]。该方法对模板 DNA 需求量小,并可对较短的 DNA 分子进行优化,在扩增 DNA 片段时可同时引入错配碱基,产生点突变,获取比有性 PCR 更广泛的突变库。

2.3 交错延伸

交错延伸(staggerd extension process, StEP) 是在 PCR 反应体系中,加入一组相关亲本 DNA 作模板,在随后的多轮变性和短暂复性、延伸过程中,使延伸片段在不同的具有部分同源序列的 DNA 模板间跳转,最终形成全长杂合基因突变库^[7]。该方法的基因重组程度可通过控制反应条件和时间予以调节,仅需在一个反应管中进行,简化了有性 PCR 操作。

2.4 临时模板随机嵌合

临时模板随机嵌合(random chimera-genesis on transient templates, RACHITT) 技术与以上介绍的 DNA 洗牌技术不同,它是将随机切割的 DNA 片段杂交到另一个同家族 DNA 临时模板上, DNA 片段经过在模板上重新排序、修剪、缺口连接等形成全长 DNA 新链,消化掉临时模板,获得高度重组的基因文库^[8]。其优点是重组率高,可以获得小于 5 bp 的重组片段,且突变子具有较高的继承性,从而得到更多的活性克隆。

上述几种基于同源重组的突变技术都源自 DNA 洗牌的基本原理,一般认为是 DNA 洗牌的不同形式,它们的区别列于表 1。

表 1 几种基于同源重组的随机突变技术比较

突变技术	优点	缺点
Sexual PCR	多个亲本基因参与进化,迅速积累有益突变,消除有害突变	亲本要求较高同源性
RPR	对模板 DNA 需求量少,方便引入点突变	需随机引物,扩增片段有局限性
StEP	操作简单,在同一反应管中进行,无需 DNA 片段纯化过程	最适 PCR 反应条件难以把握
RACHITT	重组率高,可获得小于 5 bp 的重组片段,降低 PCR 反应引入的有害突变	λ 核酸外切酶不完全切割影响文库质量,事前需获得单链 DNA 模板和重组片段

3 基于非同源重组而实现的随机突变

由于 DNA 洗牌固有的同源重组特性,在非同源或低同源基因重组上的应用受到限制,因此近年来发展出了许多基于非同源重组产生基因变异的方法,也成为蛋白质定向进化的重要手段。

3.1 渐进式切割产生杂合酶

渐进式切割产生杂合酶(incremental truncation for the creation of hybrid enzymes, ITCHY) 是利用核酸外切酶 III 对基因片段进行消化,通过控制切割速度和时间,或者通过切割前随机插入 α phosphothioate 修饰的核苷酸的方法,产生依次相差一个碱基的一系列 DNA 片段,随后将一个基因的 5'-随机片段与另一个基因的 3'-随机片段连接,产生杂合基因文库^[9]。该技术可以在无同源性或低同源性的两个基因间产生重组,突破了 DNA 洗牌仅能在高同源基因间重组的限制。尽管 ITCHY 引入的移码突变降低了文库平均活性,但核酸外切酶渐进式切割产生 DNA 片段变异的原理被广泛应用。

3.2 SCRATCHY 文库

SCRATCHY 是将 ITCHY 和 DNA 洗牌结合起来的一种方法,首先使用 ITCHY 在两个低同源性基因间建立广泛的杂合基因文库,随后该文库被作为亲本基因进行 DNA 洗牌^[10]。SCRATCHY 突破了单一 DNA 洗牌较高同源性要求的限制,可在低同源性基因间产生多次交叉重组,甚至在高同源性基因应用上也可构建平均每个序列比单独 DNA 洗牌多 1.5 个交叉点的突变库。该方法比单一 DNA 洗牌有更广阔的应用范围。

3.3 随机多重重组 PCR

随机多重重组 PCR (random multi-recombinant PCR, RM-PCR) 首先是将需要重组的 DNA 片段进行随机平末端连接,再用一组特异引物扩增 DNA 片段二聚体,二聚体混合物经过一次无引物 PCR 连接成不同长度的 DNA 分子,最后电泳回收所需长度的片段^[11]。它实现了多个非同源 DNA 片段的随机重组,并通过对二聚体混合物的选择限定,建立了两个不同的基因文库:随机洗牌文库实现了 DNA 片段的随机组合;选择拼接文库则在不改变 DNA

片段排列次序情况下产生了片段的随机缺失。

3.4 外显子洗牌

Kolkman 等^[1]于 2001 年建立了以外显子为单元进行自由重组的外显子洗牌 (exon shuffling) 技术, 首先用嵌合寡核苷酸引物分别扩增需要参与洗牌的外显子或外显子组, 混合预扩增产物, 经过无引物 PCR 反应连接成不同组装形式的完整长度的 DNA 分子, 形成外显子洗牌文库。通过控制参与洗牌的外显子范围, 可产生不同特性的突变库, 它可以加入更多的理性设计, 甚至可以完全避免点突变, 因此可应用于医药等某些特殊领域。

4 基于 DNA 长度改变而实现的随机突变

高错误倾向 PCR 以及基因重组产生的随机突变并没有充分模拟自然进化过程中的全部变异类型, 基于 DNA 片段长度改变的基因突变方法也成为蛋白质定向进化技术的重要补充。

4.1 DNA 片段的随机重复与缺失

荷兰 Pikkemaat 等^[12]报道了一种产生 DNA 片

段随机缺失与重复的方法, 它利用质粒为载体, 同一基因的 3'-端和 5'-端分别用核酸外切酶 BAL-31 渐进式切割, 随后将两组随机长度 DNA 片段进行连接, 在连接点上产生了 DNA 片段的重复或缺失。该方法可以在基因全长范围内随机产生重复或缺失, 通过调整切割程度, 还可以理性地控制变异区段的大小和位点。

4.2 随机插入/缺失突变

随机插入/缺失突变 (random insertion/deletion mutagenesis, RID) 的步骤是, 先将目的基因环化并降解为单链, 用 Ce(IV)-EDTA 随机切开成线性单链 DNA, 并在其一端连接准备插入的 DNA 片段, PCR 扩增后再经环化、酶切等操作, 获取突变库^[13]。该方法独特的地方是, 删除位点随机, 且外源 DNA 片段插入位点和删除位点相同, 插入和删除片段大小可以控制。通过酶切位点特异性, 还可以对特定部位进行突变, 非常适合小范围变异的产生。

上述几类随机突变技术在方法、应用上各有独特之处, 为便于比较, 将它们的区别列于表 2。

表 2 几类随机突变技术的比较

突变技术	原理及主要技术	突变类型	特点	应用
高错误倾向 PCR	PCR 过程中的碱基错配, 使用低保真 DNA 聚合酶	点突变	操作简单, 由单个基因创建低水平高丰度基因文库	应用范围最广, 可用于单一起始亲本基因的进化, 还可与其他进化技术很好地配合使用
DNA 洗牌	随机切割, 同源序列重组, 无引物 PCR	缺失、重复、颠倒、重组等多种形式	操作比较简单, 由多个同源基因重组创建突变库, 良性突变比例较高	适合对高同源性的同家族、同功能蛋白质改造, 还可以对由高错误倾向 PCR 得到的一组良性克隆进行新一轮进化和筛选
非同源重组	核酸外切酶渐进切割, 嵌合引物扩增 DNA 片段, 无引物 PCR	缺失、重复、颠倒、重组、移码等多种形式	突破 DNA 洗牌高同源性限制, 产生更多交叉重组, 但容易产生移码突变, 降低文库活性	适于基因序列非同源或低同源性的蛋白质改造, 应用嵌合引物可以融入更多理性设计
DNA 长度改变	核酸外切酶渐进切割, 应用限制性内切酶和 DNA 连接酶	特定片段的插入、缺失、重复	在基因局部产生变异, 可以控制变异方式和水平	适用于基因序列和蛋白质空间结构有一定了解的蛋白质改造

5 突变库的分析及构建策略

5.1 DNA 洗牌文库与随机点突变文库

DNA 洗牌文库分析以及数学模型的建立对控制反应条件和预测重组结果具有重要意义。Joern 等^[14]用探针杂交分析方法对 DNA 洗牌文库每个基因平均含有的交叉点数量、文库对亲本基因的偏爱性以及交叉点产生的位置偏爱性进行了分析和实验证实。目前更为完善的热动力学数学模型可以追踪全部 DNA 片段在整个 DNA 洗牌循环中的状态, 预测所有反应条件对 DNA 洗牌过程的影响^[15]。

实验结果与模拟分析表明, DNA 洗牌的反应条件对突变库的特性有显著的影响(见表 3)。

高错误倾向 PCR 得到的随机点突变库的适合度和丰度由引入错配碱基数量的多少来调节, 温和的突变产生高适合度、低丰度的文库, 剧烈的突变产生低适合度、高丰度的文库。大多数点突变带来中性和有害突变, 只有少数点突变产生有益的基因克隆, 但随机点突变库突变位点范围广泛, 具有产生更多新颖克隆的潜力。

DNA 洗牌文库具有很强的继承性, 其活性特征围绕在亲本基因周围, 而随机点突变库的特征是

表 3 DNA 洗牌反应条件对突变库的影响

影响因素	交叉点数量	文库平均适合度	文库丰度	其它说明
亲本数量增加	增加	降低	增高	亲本数量的增加使获得个别高适合度克隆的潜力增强
亲本同源性提高	增加	提高	降低	交叉点偏好发生于高同源区段两侧, 亲本同源性降低大大提高文库亲本基因的偏好性
平均 DNA 片段长度减小	增加	降低	增高	过小的 DNA 片段要求较低的退火温度, 引起非同源错配, 产生混乱的基因文库
退火温度降低	增加	降低	增高	过低的退火温度引起 DNA 片段的非同源重组, 迅速降低文库适合度

没有方向性。两种文库对同一基因的两个独立活性的适合度典型分布如图 1 所示, 图中实线圈代表基因初始活性范围, 虚线圈代表活性 A 和 B 都得

到提高的活性范围^[16]。如果对蛋白质多个特性同时进行筛选, DNA 洗牌文库的潜力明显高于随机点突变库。

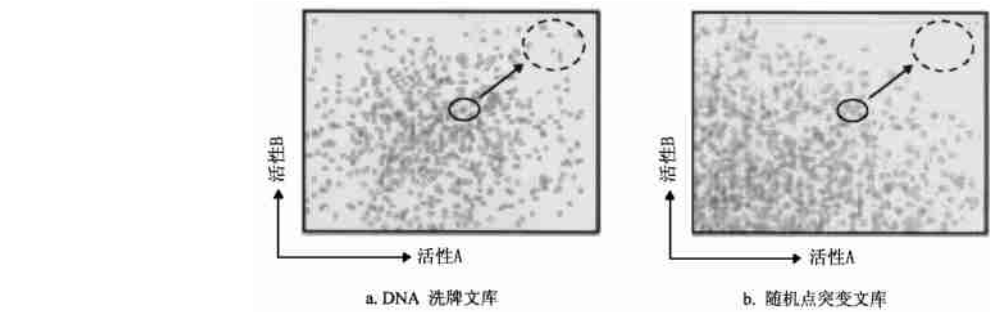


图 1 DNA 洗牌文库和随机点突变文库对两个独立活性适合度的典型分布图

5.2 突变库构建策略

在实际应用中, 希望得到中等适合度和丰度的突变库, 适合度与丰度相平衡, 适合度与蛋白质筛选通量相匹配。虽然不同的蛋白质有不同的要求, 但一般来说, 随机点突变控制突变频率在每个序列 2~ 3 个碱基或 1 个氨基酸残基变异的水平, 可以获得适当活性的基因文库^[17]; DNA 洗牌文库则可以依据实践经验和数学模型控制反应条件得到适当的突变频率。

由单一亲本基因起始进行定向进化的典型策略有两种, 一种是运用随机点突变技术构建起始基因文库, 选择最好的克隆进入下一轮随机点突变^[18]; 另一种是运用随机点突变技术构建起始基因文库后, 选择适合度高于某一域值的所有良性克隆, 以此为亲本构建 DNA 洗牌文库。前者构建的突变库大部分克隆适合度低于亲本基因, 需要高通量筛选方法予以配合, 定量筛选最好克隆也需较大工作量, 同时有害突变容易积累。后者则可以在起始突变库的基础上, 将多数良性克隆并行向前进化, 避免丢掉局部的良性突变, 同时快速消除有害突变, 积累有益突变。

可以应用家族 DNA 洗牌方法构建突变库^[19]。DNA 洗牌最终获得高适合度克隆后, 可采用回交的策略, 将其与某一原始亲本基因再进行一次 DNA 洗牌, 消除无贡献突变和影响其它特性的有害突变^[20]。

5.3 突变库构建中的理性设计

蛋白质理性设计与定向进化有着本质的区别, 但两者的结合似乎有着更强大的威力^[21]。随着晶体 X 射线衍射及核磁共振技术的发展, 结合计算机辅助设计, 人们已有相当的能力对蛋白质进行理性改造, 理性设计的思想也逐渐融入蛋白质的定向进化之中。

建立非同源重组基因文库多数是以蛋白质结构同源为依据的, 需要依靠更多的理性设计为指导。如外显子洗牌技术, 可以通过对参与洗牌的外显子进行理性筛选、对嵌合引物进行理性设计等, 控制突变库的特性^[11]。一些高丰度、低水平突变库的筛选会受到相应筛选技术通量的限制, 通过对蛋白质结构的理性思考和计算, 可以区分有利于进化的可能位点, 有目的地进行突变, 从而大大降低文库丰度, 减轻筛选工作量^[22]。

总之, 构建突变库的方法多种多样, 各有特点,

在实际应用中需要根据具体问题因地制宜地加以选择和灵活应用。蛋白质体外定向进化的另一重要环节——高通量文库筛选技术也在不断改进,尤其是体外无细胞翻译系统的发展和应用更推动了蛋白质体外定向进化技术的快速发展。蛋白质体外定向进化技术已成功地应用于酶活力的提高、药物蛋白特性的改进、新代谢途径的获取等众多领域,它把人类的智慧融入蛋白质世界的进化和发展之中,必将创造出更多的新颖蛋白质造福于人类。

参考文献

- [1] Kolkman J A, Stemmer W P C. Directed evolution of proteins by exon shuffling. *Nature Biotechnology*, 2001, 19: 423~ 428
- [2] Leung D W, Chen E, Goeddel D V. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique*, 1989, 1: 11~ 15
- [3] 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/translation. *Nucleic Acids Res*, 1998, 26: 4339~ 4346
- [4] Suang R, Kawarasaki Y, Imaede T, et al. High throughput, cloning independent protein library construction by combining single molecule DNA amplification with *in vitro* expression. *J Mol Biol*, 2002, 318: 395~ 405
- [5] Zhao H, Arnold F H. Optimization of DNA shuffling for high fidelity recombination. *Nucleic Acids Res*, 1997, 25: 1307~ 1308
- [6] Shao Z, Zhao H, Giver L, et al. Random priming *in vitro* recombination: an effective tool for directed evolution. *Nucleic Acids Res*, 1998, 26: 681~ 683
- [7] Zhao H, Giver L, Shao Z, et al. Molecular evolution by staggered extension process (SMP) *in vitro* recombination. *Nature Biotechnology*, 1998, 16: 258~ 261
- [8] Coco W M, Levinson W E, Crist M J, et al. DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nature Biotechnology*, 2001, 19: 354~ 359
- [9] 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: e16
- [10] Lutz S, Ostermeier M, Moore G L, et al. Creating multiple crossover DNA libraries independent of sequence identity. *Proc Natl Acad Sci USA*, 2001, 98: 11248~ 11253
- [11] Tsuji T, Onimaru M, Yanagawa H. Random multi recombinant PCR for the construction of combinatorial protein libraries. *Nucleic Acids Res*, 2001, 29: e97
- [12] Pikkemaat M G, Janssen D B. Generating segmental mutations in haloalkane dehalogenase: a novel part in the directed evolution toolbox. *Nucleic Acids Res*, 2002, 30: e35
- [13] Murakami H, Holsaka T, Sisido M. Random insertion and deletion of arbitrary number of bases for codon based random mutation of DNAs. *Nature Biotechnology*, 2002, 20: 76~ 81
- [14] Joem J M, Meinhold P, Arnold F H. Analysis of shuffled gene libraries. *J Mol Biol*, 2002, 316: 643~ 656
- [15] Maheshri M, Schaffer D V. Computational and experimental analysis of DNA shuffling. *Proc Natl Acad Sci USA*, 2003, 100: 3071~ 3076
- [16] Kurtzman A L, Govindarajan S, Vahle K, et al. Advances in directed protein evolution by recursive genetic recombination: applications to therapeutic proteins. *Biotechnology*, 2001, 12: 361~ 370
- [17] Arnold F H, Wintrodde P L, Miyazaki K, et al. How enzymes adapt: lessons from directed evolution. *Trends in Biochemical Sciences*, 2001, 26: 100~ 106
- [18] You L, Arnold F H. Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide. *Protein Eng*, 1996, 9: 77~ 83
- [19] Cramer A, Raillard S A, Bermudez E, et al. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature*, 1998, 391: 288~ 291
- [20] Stoop A A, Eldering E, Dafforn T R, et al. Different structural requirements for plasminogen activator inhibitor 1 (PAI1) during latency transition and proteinase inhibition as evidenced by phage displayed hypermutated PAI1 libraries. *J Mol Biol*, 2001, 305: 773~ 783
- [21] Bomscheuer U T, Pohl M. Improved biocatalysts by directed evolution and rational protein design. *Chemical & Biology*, 2001, 5: 137~ 143
- [22] Voigt C A, Mayo S L, Arnold F H, et al. Computational method to reduce the search space for directed protein evolution. *Proc Natl Acad Sci USA*, 2001, 98: 3778~ 3783

Comparison of Several Directed Evolution Techniques and Strategy of Library Construction

Jia Xiangdong Chen Defu Chen Xiwen Guo Shaoying

(Department of Biochemistry and Molecular Biology Nankai University Tianjin 300071)

Abstract The directed evolution has been applied successfully in the field of protein engineering. Its key technique through constructing gene library by random mutation has recently become a focus in the directed evolution area. we summarized the principle of directed evolution and introduced several random mutation techniques and classified them according to their mutation principles. Characteristics of mutant libraries were compared and strategy of library construction was also discussed.

Key words Directed evolution Random mutant library Base mismatch DNA recombination