

# 蛋白质定向进化的研究进展

肖志壮 刘梦海 汪天虹 曲音波

(山东大学微生物技术国家重点实验室 济南 250100)

**摘要** 定向进化是改造蛋白质分子的一种有效的新策略。主要是在实验室里模拟自然进化过程,通过由易错 PCR、致突变菌株诱变等方法对编码蛋白质的基因进行随机诱变,由 DNA 改组、随机引导重组和交错延伸等方法进行突变基因体外重组,设计高通量筛选方法来选出需要的突变株。它不仅可快速产生工业上有用的新酶,而且对研究蛋白质的结构与功能的关系具有非常重要的意义。

**关键词** 定向进化 随机诱变 DNA 改组 随机引导重组 交错延伸技术

经过长时期的自然选择,自然界产生的无数蛋白质表现出非常理想的生物学功能<sup>[1]</sup>。酶的催化效率和专一性是大多数传统工业催化剂无法与之相比的,因而,酶在各个领域中的应用越来越普遍,并将继续迅速发展至下个世纪<sup>[2]</sup>。然而,利用酶来设计工业工艺的化学工程师们却经常遇到这样的难题:酶经过数亿年的进化,在生物体内能完成非常特定的生物学功能。但当它们处于机体外复杂的化学体系中时,一些特征则不是人们所期望的,如产物抑制<sup>[3]</sup>。生物工艺通常需要下列类型的酶:长时间内保持活性稳定的酶、在极端环境中保持高活性的酶或可接受不同底物(包括自然界不存在的底物)的酶等<sup>[4]</sup>。

具有新的功能和特性的蛋白质可以通过两条途径获得:一是从大量未知的自然种系中寻找;二是改造现有已知的天然蛋白质或酶。对于在自然进化过程中没有经过选择的特性来说,后一条途径可能更合适<sup>[5]</sup>。八十年代初期,蛋白质工程技术应运而生。其基本概念是首先分析研究蛋白质的三维空间结构,搞清结构与功能之间的关系,然后采用定点突变(site-directed mutagenesis)改变蛋白质分子中个别氨基酸残基,产生新性状的蛋白质。因而,这种尝试被称为理性设计(rational design)。然而,这类方法虽获得了一定的进展,但仅适用于蛋白质三维结构已弄清,并且对结构与功能的关系比较清楚的蛋白质<sup>[2]</sup>。蛋白质的结构和功能的相互关系非常复杂。目前,我们对这类关系的理解仍然非常浮浅<sup>[3]</sup>。事实上,

通过不同途径(如随机诱变和大量筛选)成功地改造了的酶,经常有一些氨基酸的变化是在根据蛋白质模型预测的区域之外<sup>[2]</sup>。当我们对所研究的蛋白质分子结构了解很少时,定点突变就更是束手无策了<sup>[6]</sup>。于是人们开始考虑似乎是非理性的方法——定向进化(Directed evolution)<sup>[7]</sup>。

## 1 蛋白质定向进化的基本思路

自然进化是在整个有机体繁殖和存活的过程中自发出现的一个非常缓慢的过程<sup>[4]</sup>。自然选择使进化向有利于生物适应生存环境的方向发展。环境的多样性和适应方式的多样性决定了进化方向的多样性。我们可以在实验室中模仿自然进化的关键步骤——突变、重组和筛选,在较短时间内完成漫长的自然进化过程,有效地改造蛋白质<sup>[1]</sup>,使之适于人类的需要。与自然进化不同的是,这种策略具有明确的人为设定的目标,只针对特定蛋白质的特定性质,因而被称为定向进化(directed evolution)。定向进化技术使我们可以在对蛋白质分子了解很少的情况下,寻求自然环境中并不需要的功能。由于定向进化不需要蛋白质的结构信息,因之被称为非理性设计。

在考虑实验方案时必须注意蛋白质的以下特点:(1)蛋白质的顺序空间是巨大的。一种由 300 个氨基酸串联而成的蛋白质可能有  $20^{300}$  种排列方式,其中绝大多数排列方式没有反应功能,尤其是我们想要的功能。(2)大多数突变是有害的,有利突变稀少,而有利突变的组合就更加罕见。特点(1)说明通过完全随机的取样方式很难得到我们所需的东西。蛋白质的进化最好是选择一个性状接近于你所想要

本项目由国家自然科学基金(39970392)资助。

的蛋白质作为起点。特点(2)则告诉我们理想的状态是在每一蛋白质序列中每次只产生一个氨基酸突变<sup>[1]</sup>。

定向进化需要(1)产生包含大量带有微量有利突变的突变体的文库;(2)突变体应能在适当的微生物(如大肠杆菌或酵母菌)体内进行功能的表达<sup>[4]</sup>;(3)必须有一种灵敏的筛选方法,能反映出由一个氨基酸的置换而引起的预期性状的较小提高<sup>[3]</sup>。

定向进化能否解决一个特定的问题,在很大程度上取决于自然进化对这一性状所改造的状态。如果一种特性(如催化活性)已经处于自然选择的长期压力之下,在实验室中获得进一步提高的几率就会比较小。一般说来,未经过自然选择的性能容易提高。对于生物功能从不需要的特征,如提高酶对一种非天然底物的活性余地很大<sup>[4]</sup>。

## 2 定向进化的常用方法

蛋白质定向进化通常分三步进行:基因随机诱变、体外重组和筛选。每一步都可以有多种方法。

### 2.1 随机诱变

#### 2.1.1 通过易错 PCR(error-prone PCR)产生随机突变

所谓易错 PCR 是指在体外扩增基因时使用适当条件,使扩增的基因出现少量碱基误配,引起突变。易错 PCR 的关键是控制 DNA 的突变频率。如果 DNA 的突变频率太高,产生的绝大多数酶将失去活性;如果突变频率太低,野生型的背景太高,样品的多样性则较少。对于每一 DNA 序列来说,合理的碱基突变数是 1—3。易错 PCR 产生的 DNA 突变频率一般为 0.25—20 个置换/1000 个碱基。理想的碱基置换率和易错 PCR 的最佳条件则依赖于随机突变的目标 DNA 片段的长度<sup>[9]</sup>。

#### 2.1.2 化学诱变剂介导的随机诱变

体外随机诱变也可在 65℃ 下直接用羟胺处理带有目的基因片段的质粒,然后用限制性内切酶切下突变了基因片段,再克隆到表达载体中进行功能的筛选<sup>[14]</sup>。

#### 2.1.3 由致突变菌株(mutator strain)产生随机突变

美国 Stratagene 公司构建了一株 DNA 修复途径缺陷的大肠杆菌突变株 XL1-Red,它体内的 DNA 突变率比野生型高五千倍。将带有要突变基因的质粒转化到 XL1-Red 菌株内复制过夜,在此过程中会产生随机突变。每两千个碱基中通常约有一个碱基置换。将带有突变过的基因的质粒转化到表达系统中

进行筛选<sup>[15]</sup>。

连续几代的随机突变和筛选的方法是从每一代中挑选出一个最佳的突变体作为下一代的亲本。通过积累氨基酸的正突变,可加快进化的进程<sup>[8]</sup>。

### 2.2 基因体外重组

在每一代诱变中经常出现几个阳性突变。如果只选一株突变株继续诱变,便遗漏了其它所有未被选作为下一代的亲本的已经鉴定出来的阳性突变体。这些突变体包含大量有用的信息,应加以合理利用。要通过诱变将这些有利突变整合到一起是困难的<sup>[1,8,9]</sup>。再者,随机诱变除了产生回复突变外,没有其他机制删除有害突变。而连续诱变方法同时也积累有害突变,导致一个进化的谱系提前终止<sup>[8]</sup>。计算机模拟分析表明,仅靠点突变来改变蛋白质经常显得过于缓慢,而且不能产生继续进化所需的整块的变化<sup>[11]</sup>。因此,需要利用重组机制来克服随机诱变方法的不足,更好地完成蛋白质的改造。

#### 2.2.1 限制性酶切-再连接介导的基因重组(recombining the genes by restriction and religation)

利用限制性内切酶对欲重组的两个以上 DNA 序列进行酶切,然后在连接酶的作用下随机重新连接起来,可实现 DNA 分子间的重组。实验证明,简单的酶切-再连接可进一步提高对硝基苯酯酶的活性<sup>[9]</sup>。但如果两个正突变处于同一段限制性酶切片段上,则无法将它们重组于同一序列中。

#### 2.2.2 DNA 改组(DNA shuffling)

1994 年,Stemmer 等首先使用 DNA 改组完成体外重组<sup>[11]</sup>。DNA 改组是将一群密切相关的序列(如多种同源而有差异的基因或一组突变基因文库)在 DNase I 的作用下随机酶切成小片段,这些小片段之间有部分的碱基序列重叠,它们通过自身引导 PCR (self-priming PCR) 延伸,并重新组装成全长的基因,这一过程被称为再组装 PCR(reassembly PCR)。该过程所产生的相关序列间的交换归因于模板的转换<sup>[5]</sup>。DNA 改组过程中也可能引入新的点突变。因此,仅 DNA 改组就可以有效地从单一基因序列开始进行蛋白质的定向进化<sup>[8]</sup>。

DNA 改组具有以下有用的特征:(一)它可以利用现存的有利突变,快速积累不同的有利突变。(二)重组可伴随点突变同时发生。(三)可以删除个体中的有害突变和中性突变<sup>[1,3,11]</sup>。

但是,DNA 改组过程中伴随的较高点突变频率会严重阻碍正突变组合的发现。由于绝大多数突变是有害的,有利突变的重组和稀少有利点突变会被

有害突变的负背景所掩盖。当对四种进化的了的 pNB esterase 基因用 Taq DNA 聚合酶进行重组时,90 % 的克隆在筛选时不表现活性。为了降低突变频率,使用一种具有校正功能的 DNA 聚合酶——Pwo,结果 80 % 的克隆保持活性<sup>[8]</sup>。采用 DNA 聚合酶 Pfu 进行 DNA 改组,所伴随的点突变频率仅为 0.05 %,可以重组仅有 12bp 之隔的突变,并使 95 % 以上的克隆保持活性<sup>[16,17]</sup>。

### 2.2.3 随机引导重组(random-priming recombination, RPR)

1998 年,Arnold 提出了一种有效的新方法——随机引导重组(RPR)。RPR 的原理是:用随机序列的引物来产生互补于模板序列不同部分的大量的 DNA 小片段。由于碱基的错误掺入和错误引导,这些 DNA 的小片段中也因之而含有少量的点突变。DNA 小片段之间可以相互同源引导和重组。在 DNA 聚合酶作用下,经反复的热循环可重新组装成全长的基因,克隆到表达载体上,随后筛选。与 DNA 改组相比,RPR 技术具有以下优点:(一)RPR 可直接利用单链 DNA 或 mRNA 作模板。(二)DNA 改组利用 DNase I 随机切割双链 DNA 模板,在 DNA 片段重新组装成全长序列之前,DNase I 必须去除干净。一般说来,RPR 技术使基因的重新组装更容易。(三)合成的随机引物长度一致并缺乏序列的偏向性,保证了点突变和交换在全长的后代基因中的随机性。(四)随机引导的 DNA 合成不受 DNA 模板长度的影响,这给小肽的改造提供了机会。(五)所需亲代 DNA 比 DNA 改组所需的量少 10—20 倍<sup>[12]</sup>。

### 2.2.4 交错延伸技术(staggered extension process, STEP)

1998 年,Arnold 等人又建立了一种新的体外重组方法——交错延伸技术:用 PCR 同时扩增多个拟重组的模板序列时,在热循环中大大缩短退火与聚合酶催化延伸的时间。在每一循环中,不断延长的片段根据序列的互补性与不同模板退火,并进一步延伸。反复重复直到全长序列形成。由于模板的转换,大多数多核苷酸含有不同亲本的序列信息。该技术已成功地重组了由易错 PCR 产生的 5 个热稳定的枯草杆菌蛋白酶 E 的突变体,得到了热稳定性进一步提高的重组酶<sup>[13]</sup>。

### 2.3 筛选

在突变体文库产生之后,筛选的条件确定蛋白质预期特征的进化方向,这被称为定向进化的第一定律。建立有效的搜索蛋白质文库的方法是影响定

向进化成功与否的关键。当突变体酶可赋予宿主细胞生长或存活的优势时,很容易搜寻含有  $10^6$  以上个蛋白质突变体的文库。然而,这类情况并不多见。当感兴趣的功能可产生可见的信号时,简单的肉眼可见的筛选方法被广泛地应用。例如,菌落分泌出有活性的蛋白酶可在含有酪蛋白的琼脂平板上产生清晰的水解圈,其大小与水解活性成正比。另外,易被观察的菌落表型也可用于快速筛选。尽管根据颜色或水解圈形成的筛选方法快速高效,但也有明显的局限性。它是非定量的,而且经常对性状的微小变化不灵敏。高通量 96-孔板适合自动定量筛选,能鉴定出酶对底物的特异性、热稳定性、对映体选择性及其他特定性状等<sup>[7]</sup>。

## 3 成功的例子

短短几年中,用定向进化的方法改造工业用酶的成功例子已有几十个,它们在提高酶在非天然环境(如有机溶剂)中的活性、耐热稳定性、非天然底物的特异性、对映选择性及表达水平等方面都有成功。本文略举其中有代表性的几例:Reetz (1997) 将脂酶水解对硝基-2-甲基癸酯(p-nitrophenyl 2-methyldecanoate)的对映体选择性提高了 40 倍;You 和 Arnold (1996)使用易错 PCR 提高枯草杆菌蛋白酶 E 的表达水平和在有机溶剂中的活性,使酶的总活力提高了 500 倍;Zhang 等人(1996)改造  $\alpha$ -半乳糖苷酶,其底物特异性提高了 1000 倍;Zhao 和 Arnold (1998)在保持酶活力不变的条件下将枯草杆菌蛋白酶 E 的作用温度提高了 17 度;最为显著的是,Stemmer (1994)使  $\alpha$ -内酰胺酶对新的底物 cefotaxime 的活性提高了 32000 倍<sup>[10]</sup>。

## 4 蛋白质定向进化的理论意义

定向进化不仅在工业、医疗方面对蛋白质的改造具有重要的意义,而且,它也非常适合于基础理论的研究。对蛋白质分子进化得到的突变体进行分析,为研究蛋白质的结构和功能的关系提供了新的工具<sup>[5]</sup>。天然蛋白质序列中往往有很多是随机遗传漂移,而不是功能性的适应。例如,从适应不同环境的物种中提取出的蛋白质经常有几十个氨基酸的不同,其中只有一小部分是与特定功能差异有关的。大量的中性突变只能干扰科学家辨别分子的规律。在实验室的进化实验中,种系很清楚,而且突变主要是适应性的。Arnold 等人使用高保真 DNA 改组成

(转第 56 页)

## Studies on the detection of HBV DNA from human serum by PCR-ELISA Technique

Zhan Qun-shan   Chen Yu-bao\*   Lin Yu-bin   Xiao Wen-yuan

(Beijing MDC Biotech Co., Ltd. 100012)

**Abstract** A sensitive, accurate method is important for early diagnosis of patients with HBV infected, so we established ELISA assay for the detection of PCR amplified sequence of HBV DNA in human serum. Through this method, 268 sera from HBV positive or negative patient were detected, the result showed that the ELISA assay based on PCR product is more sensitive and more specific than electrophoresis assay. This method in clinical application was discussed in details in the paper.

**Key words** HBV, polymerase chain reaction, enzyme linked immunosorbent assay

\* correspondence: yubaochen @ yahoo.com

(接第 33 页)

功地区分出基因中的功能性和非功能性突变<sup>[16]</sup>。通过定向进化来研究蛋白质的结构和功能的关系, 可为蛋白质的理性设计提供理论依据。

### 参考文献

- [ 1 ] Arnold Frances H. Accounts of Chemical Research, 10/3/97: 1 - 15.
- [ 2 ] Harayama Shigeaki. TIBTECH, 1998, 16: 76 - 82.
- [ 3 ] Arnold Frances H. and Moore Jeffrey C. Advances in Biochemical Engineering/Biotechnology, 1997, 58: 1 - 14.
- [ 4 ] Arnold F. H. In IBC Directed Enzyme Evolution. San Diego. 1998.
- [ 5 ] Zhang Ji-hu, Dawes Genn and Stemmer Willem P. C. Proc. Natl. Acad. Sci. USA, 1997, 94: 4504 - 4509.
- [ 6 ] Chen Keqin and Arnold F. H. Proc. Natl. Acad. Sci. USA, 1993, 90: 5618 - 5622.
- [ 7 ] Zhao Huimin and Arnold Frances. Current Opinion in Structural Biology, 1997, 7: 480 - 485.
- [ 8 ] Moore Jeffrey C. and Arnold Frances H. Nature Biotechnology, 1996, 14: 458 - 467.
- [ 9 ] Moore Jeffrey C. and Arnold Frances H. Nature Biotechnology, 1996, 14: 458 - 467.
- [ 10 ] Stemmer W. P. C. Nature, 1994, 370: 389 - 391.
- [ 11 ] Stemmer Willem P. C. Proc. Natl. Acad. Sci. USA, 1994, 91: 10747 - 10751.
- [ 12 ] Shao Zhixin, Zhao Huimin, Arnold F. H. Nucleic Acids Research, 1998, 26(2): 681 - 683.
- [ 13 ] Zhao H, Gver L, Shao Z, Affholter JA and Arnold FH. Nat Biotechnol 1998, 16(3): 258 - 61.
- [ 14 ] Seiichi Taguchi, Akiyoshi Ozaki and Haruo Momose. Applied and Environment Microbiology, 1998, 64(2): 492 - 495.
- [ 15 ] Greener, A. and Callahan. M. Perform Random Mutagenesis. Strategies, 1994, 7: 32 - 34.
- [ 16 ] Zhao Huimin and Arnold Frances H. Proc. Natl. Acad. Sci. USA, 1997, 94: 7997 - 8000.
- [ 17 ] Zhao Huimin and Arnold Frances H. Nucleic Acids Research, 1997, 25: 1307 - 1308.

## Advances in Study on Directed Evolution of Proteins

Xiao Zhizhuang   Liu Menghai   Wang Tianhong   Qu Yinbo

(State Key Laboratory of Microbial Technology, Shandong University, Jinan 250100, P. R. China)

**Abstract** Directed evolution is a new powerful strategy for engineering proteins. It mainly mimics the process of natural evolution in the laboratory to generate random mutations in the genes coding for useful proteins by the techniques of error-prone PCR and mutator strain, then in vitro recombines positive mutations through such methods as high-fidelity DNA shuffling, random-priming recombination and staggered extension process. Finally, the desired mutants are selected out with high-throughput screening. It not only can provide rapidly new enzymes for industrial application, but also has much theoretical importance for studying the relationship between protein structure and function.

**Key words** Directed evolution, random mutagenesis, DNA shuffling, random-priming recombination, staggered extension process.