

蛋白质定向进化的研究进展*

王晓玥^{1, 2, 3)**} 王白云^{1, 2, 3)**} 王智文^{1, 2, 3)***} 陈涛^{1, 2, 3)} 赵学明^{1, 2, 3)}

(¹ 天津大学化工学院, 天津 300072; ² 天津大学系统生物工程教育部重点实验室, 天津 300072; ³ 天津化学化工协同创新中心, 天津 300072)

摘要 在工业生物催化过程和生物细胞工厂构建方面, 蛋白质定向进化被广泛地应用于酶的分子改造。蛋白质定向进化不仅可以针对某一目的蛋白进行改造, 还可以改善代谢途径、优化代谢网络、获得期望表型细胞。为了获得更高效的突变效率, 快捷、高通量的筛选方法, 提高蛋白质定向进化的效果, 研究者不断开发蛋白质体内、体外进化方法, 取得了新的进展和应用。本文介绍了最近发展的蛋白质定向进化技术的原理、方法及特点, 总结了突变文库的筛选方法和蛋白质定向进化的最新应用, 最后讨论了蛋白质定向进化存在的挑战和未来发展方向。

关键词 蛋白质定向进化, 随机进化, 半理性进化, 理性进化, 筛选

学科分类号 Q5, Q7, Q81

DOI: 10.16476/j.pibb.2014.0059

1859年, 达尔文在《物种起源》中提出了生物进化论学说, 揭示了生物在遗传、变异、生存斗争和自然选择中是不断发展变化的。近代生物学的发展使人们认识到蛋白质进化大多是由于某些点突变或修饰的积累^[1], 而在自然条件下, 蛋白质性质或功能的改变通常需要很长时间。早期人们采取了各种诱变手段以期能够对蛋白质性能进行改造, 如最早在1982年, Winter等^[2]报道了通过基因定位诱变获得改进的酪氨酸 tRNA 合成酶。1983年, Ulmer^[3]在《科学》(Science)上提出蛋白质工程(protein engineering)一词, 指按照人们的需求, 通过对蛋白质信息和功能的分析, 设计改造蛋白质分子。但当人们试图基于蛋白质结构对其进行改造时, 由于结构生物信息的匮乏以及蛋白质结构的复杂性, 合理设计进行得十分困难, 蛋白质工程的发展受到了严重制约^[4]。

蛋白质定向进化的发展拓宽了蛋白质工程的设计范围, 可在未知目标蛋白质结构信息和作用机制的情况下对蛋白质进行改造^[5]。蛋白质定向进化发展初期, 科学家主要将其用于筛选和控制(或影响)所需表型。20世纪中期, 蛋白质定向进化被引入实验室用于再现和研究自然进化进程。近年, 定向进化更多地被用来改善蛋白质性能, 改进蛋白质药物的稳定性、半衰期、免疫原性, 开发酶的新底物利用以及改进或拓展新的代谢途径等^[6]。最近的研究

表明, 蛋白质定向进化被成功地应用在代谢途径的关键酶设计、新底物催化功能的开发、创造全新功能蛋白质以及筛选和鉴别预期功能蛋白质中^[6-7], 在代谢工程和合成生物学领域发挥了重要作用。

1 蛋白质定向进化策略

蛋白质定向进化的本质是构建分子多样性文库以及从文库中筛选到性状有改进的突变体^[8], 根据文库构建原理的不同, 可分为随机进化、半理性进化和理性进化三种策略, 其大致思路均为由某一靶基因或一族相关的家族基因起始, 通过对编码基因进行突变或重组, 创建分子多样性文库; 筛选文库获得能够编码改进性状的基因, 作为下一轮进化的模板; 在短时间内完成自然界中需要成千上万年的进化^[9], 从而获得具有改进功能或全新功能的蛋白质。表1总结了最新发展的蛋白质定向进化方法及其特点。

* 国家重点基础研究发展计划(973)(2011CBA00804, 2012CB725203), 国家自然科学基金(21206112, 21390201), 国家高技术研究发展计划(863)(2012AA02A702, 2012AA022103)和天津大学自主创新基金(1308)资助项目。

** 共同第一作者。

*** 通讯联系人。

Tel: 022-27406582, E-mail: zww@tju.edu.cn

收稿日期: 2014-05-13, 接受日期: 2014-07-13

Table 1 New options of protein directed evolution and the advances and challenges

表 1 蛋白质定向进化的最新方法、优势及存在的问题

	随机进化	半理性突变	理性进化
方法	TRINS 序列饱和突变技术 (sequence saturation mutagenesis, SeSaM) 重叠延伸 PCR (prolonged overlap extension PCR, POE-PCR)	CASTing 定点饱和突变 (gene site saturation mutagenesis, GSSM) ISM 盒式诱变 (Cassette mutagenesis)	计算机建模 (Computational Modeling) 从头设计 (<i>De novo</i>)
优势	增加突变多样性 针对每个或多个密码子 增强生物鲁棒性	避免终止密码子 减小文库大小 增加阳性突变比例	指导和模拟实验 实现从头设计 优化反应途径
发展方向	突变多个连续密码子 有益突变的评价标准 (功能、结构、突变率)	同源模型的选择 突变区域的确定	计算机算法的优化 (能量函数、力场等)
劣势	突变文库通常较为庞大, 筛选困难	文库大小随突变点的增多而成指数级增长, 突变体超多 10^6 时很难筛选	大大减少了实验需进行的筛选工作, 但通常得到的蛋白质活性不高

1.1 随机进化

蛋白质分离纯化和结构解析的相关技术及软件的快速发展, 产生了海量的蛋白质结构信息. 然而截至目前, 蛋白质有效突变点的预测仍然十分困难, 因而随机进化 (random evolution) 仍是十分有效

的蛋白质定向进化手段. 最近发展的随机进化方法主要分为体外随机进化 (其技术流程如图 1 所示) 和基于重组的体内随机进化, 前者包括非重组的体外随机进化和基于重组的体外随机进化. 表 2 显示了不同进化方法的具体技术及其特点, 其中非重组的

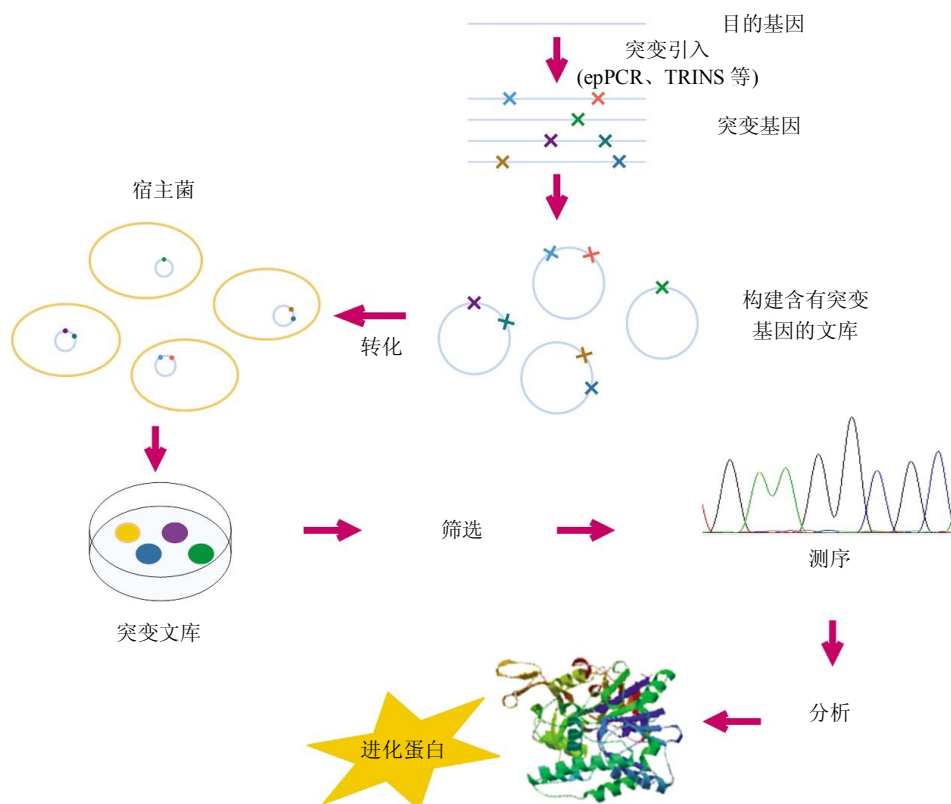
Fig. 1 The main experimental steps of an *in vitro* random directed evolution process

图 1 蛋白质体外随机进化的一般流程

体外随机进化, 如易错 PCR(error-prone polymerase chain reaction, epPCR)等的方法和应用已有详细综

述^[10], 本文将重点介绍最新发展的基于重组的体外随机进化及体内随机进化.

Table 2 Research progress of random evolution strategies

表 2 随机进化的方法及其特点

	非重组的体外随机进化	基于重组的体外随机进化	基于重组的体内随机进化
方法	epPCR SeSaM	DNA 重组 StEP TRINs	重组工程 MAGE MIPE
改进性质	蛋白活性 pH/ 热稳定性 底物抑制 催化效率	pH/ 热稳定性 反馈抑制 蛋白活性 产物得率 底物选择性	催化效率
目标蛋白	青霉素 G 酰化酶 ^[11] 麦芽糖淀粉酶 ^[12] cAMP 受体蛋白 ^[13] 吡啶-β-1, 4-x 木聚糖酶 ^[14]	纤维二糖水解酶 ^[15] cAMP 受体蛋白 ^[16] 酯酶 ^[17] 转氨酶 ^[18] 胺氧化酶 ^[19]	核酸酶 ^[20] 红色荧光蛋白 ^[21]

1.1.1 基于重组的体外随机进化

1994 年, Willem^[22]提出了一种基于重组的体外随机进化方法: DNA 重组(DNA shuffling). 与 epPCR 相比, DNA shuffling 可被用来进行多个同源基因的重组, 且由于该法在片段组装过程中有可能引入点突变, 因此也可用以指导单一序列的进化. Zhao 等^[23]提出交错延伸(staggered extension process, StEP)方法, 即通过多轮 PCR 反应, 进行

不同模板的转换, 最终得到含有多个亲本信息的产物. StEP 改进了 DNA shuffling 实验周期长、有益突变率低等问题.

最近发展的串联重复插入(tandem repeat insertions, TRINS)是一种通过滚环复制将原始基因以串联重复序列的形式(有时不只一个重复)插入到目的基因中的方法. 图 2d 展示了 TRINS 的基本流程^[8]. 尽管 TRINS 局限于使用特定的短序列片段,

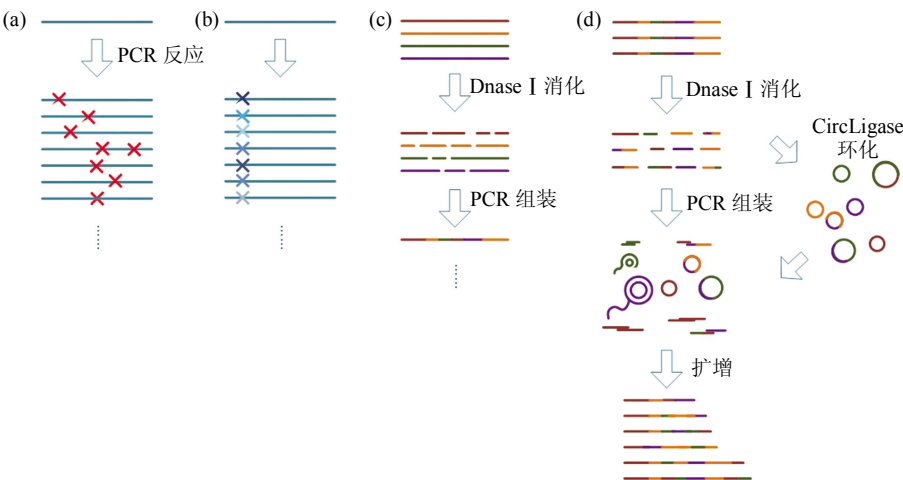


Fig. 2 Routine of different random evolution

图 2 不同随机进化的基本技术流程

(a) 易错 PCR, 改变 PCR 反应条件, 产生随机突变. (b) 点饱和和突变, 将模板 DNA 的特定位点突变为所有 20 种氨基酸. (c) DNA 重组, 将一组同源基因用核酸酶 I 消化, 得到的随机产物互为引物和模板进行 PCR 扩增, 当来源不同的片段之间相互形成模板时, 即发生重组. (d) 串联重复插入, 将一组基因用 CircLigase 消化, 再使用连接酶将其连接成环, 并以此作为模板进行 PCR 反应, 不同的串联重复序列发生随机连接, 得到目的文库.

但它能够鉴定蛋白质中的特定区域^[24], 并且 TRINS 通过模拟自然进化中的复制插入(insertion-by-duplication)机理, 避免了文库的过分膨胀, 当高通量筛选条件受限时, TRINS 将发挥十分重要的作用。

1.1.2 基于重组的体内随机进化

基于重组的体外随机进化方法均是通过 PCR 获得, 仅能改造某一个基因。早期发展的基于重组的体内随机进化方法, 如重组工程(recombineering)虽然能够十分有效地达到蛋白质进化的目的^[25], 但也难以对代谢途径或网络进行操作。近年迅速发展的体内重组方法则能够平行操作多个基因, 同时进化代谢网络中的多个关键蛋白酶, 从而在短时间内获得表型优良的突变株。

2009 年, Harris 等^[26]提出多元基因组工程(multiplex automated genome engineering, MAGE), 能够对多个基因进行修饰, 达到进化代谢途径甚至整个细胞性能的目的。MAGE 基于 ssDNA 同源重组的原理, 将合成的 ssDNA 导入细胞并定位在染色体的特定位点, 使得到的每个细胞都含有一组不同的突变, 重复进行此过程可以快速且连续地获得丰富的序列多样性异质群, 最后鉴定细胞基因型和表型。优化条件后, 使用 MAGE 技术可在 2~2.5 h 内对目标细胞群体引入至少 30% 的基因修饰。Harris 等用长度为 90-mer 的核糖体结合位点(RBS)文库(DDRRRRRDDDD; D=A, G, T; R=A, G)对 1-脱氧-D-木酮糖-5-磷酸还原异构酶合成途径(DXP 合成途径)中 20 种基因的 RBS 区域进行置换, 优化不同基因的表达水平。同时, 为提高 DXP 合成途径通量, 将其溢流代谢途径中 4 个基因(*ytjC*、*fdhF*、*aceE*、*gdhA*)的开放阅读框(ORF)分别引入了两个无义突变。经过 35 轮 MAGE 循环, 从多达 150 亿种突变体文库中筛选到了将番茄红素的产能提高 5 倍的突变株。

MAGE 技术大大提高了基因组进化效率, 且其进化对象广, 可以对基因组的任一序列进行突变。然而由于 MAGE 技术需要高重组效率, 其应用范围目前仅局限于大肠杆菌(*E. coli*)中。针对这一问题, 本课题组提出了多元质粒工程(multiplex iterative plasmid engineering, MIPE)技术^[21]。参考 MAGE 技术的原理, 以单链 ssDNA 介导的 λ -red 重组, 可以同时修改质粒上的多个位点。通过酶切共筛选技术, 提高了多片段重组效率, 6 bp 的酶切位点可以被反复利用, 无限循环。MIPE 技术的突

变过程和筛选过程严格分开, 这使它能够应用到没有高效重组系统的宿主中。我们利用 MIPE 技术对红色荧光蛋白(red fluorescent protein, RFP)上的 7 个关键位点进行突变, 仅用 3 天时间, 获得了不同光谱特性和不同光强的 RFP 突变体, 这显示了 MIPE 技术在蛋白质定向进化中的巨大应用潜力。

1.2 半理性进化

尽管随机进化策略十分有效, 但仍存在突变文库大、阳性突变少、难以筛选等问题。半理性进化(semi-rational evolution)策略则借助了生物信息学方法, 在分析大量的蛋白质序列比对信息, 二级结构数据, 甚至是在同源建模得到目的蛋白三维空间构象的基础上更有针对性地对蛋白质进行改造, 不但提高了阳性突变率, 而且大大缩小了突变文库容量, 更易于筛选。

半理性进化的关键是通过计算机模拟获得潜在的有益突变位点, 再利用适当的饱和突变技术构建合适的突变文库, 表 3 所示为不同的计算机算法以及结构分析方法。此外, 对于结构较为复杂的蛋白质, 可将其分为不同的结构单元, 并在其内部独立进化, 组合筛选最佳进化单元, 得到完整蛋白质。

Table 3 Computational algorithm and mutation library construction method for semi-rational evolution
表 3 半理性进化的计算机算法和文库构建方法

	计算机算法		基于结构分析
	基于三维结构或同源建模分析	基于性质	
突变位点获得	SCHEMA ^[27] HotSpot Wizard ^[28] ProSAR ^[29]	QSAR ^[30] ASRA ^[31]	SCOPE ^[32]
文库构建方法	GSSM ^[33] ISM ^[34] CASTing ^[35]		外显子重排
改进性质	对映选择性 ^[36] 热稳定性 ^[37] 底物特异性 ^[38]		配合基-受体

1.3 理性进化

理性进化策略主要是在计算机中(*in silico*)完成^[39]。通过计算机建模(computational modeling)预测蛋白质活性位点, 考察某基因突变对目标蛋白稳定性、折叠以及与底物结合的影响^[40], 从而对蛋白质进化进行设计指导和模拟筛选实验, 提高实验的成功率。

在代谢工程中, 虽然反应能垒对途径的影响十分重要^[41], 但随机进化和半理性进化策略均不能直接解决这一问题, 从头设计则能够对这些因素加以考虑. 从头设计首先根据量子力学建模得到目的催化反应^[42], 额外考虑某个具有高能垒的反应, 推测其过渡态, 定位所需的催化侧链并结合反应的优化过渡态^[43]. 利用 QM/MM 模型^[44]分析得到包含过渡态和涉及结合、催化的蛋白质功能团的理论蛋白质突变文库, 再利用 Rosetta、ORBIT、PyMol 等软件, 在大量稳定的蛋白支架中搜索能够支持这些理想活性位点的蛋白质主链群, 最终经过优化处理的基因序列即可进行实验验证^[45]. 这种方法能够在数百个潜在的模板酶侧链中自动挑选合适的算法搜索催化侧链过渡态, 并将其定位到合适位置^[46-47], 在一些研究中, 单一蛋白质或最佳活性位点也可以进行手动选择^[48-49].

从头设计对于酶进化轨迹的分析有助于增强人们对蛋白质自然进化的理解, 从实验中获得的反馈信息也可以更好地辅助完善理性设计. 此外, 利用计算机预测可以建立代谢途径、分析代谢网络, 从而使用工程学方法选出重要的酶加以进化, 以达到整个途径或网络的优化目的.

1.4 蛋白质定向进化文库的筛选策略

传统的筛选策略是根据表型观察的筛选, 通过对细胞的生长率、生存率或底物消耗、产物生成速率等的观测筛选出目的菌株, 例如琼脂平板克隆筛选或粗酶裂解液的酶活检测等方法. 然而, 表型观察筛选法不能定量且对微小变化不灵敏, 因此传统筛选方法在很大程度上限制了突变文库大小以及筛选能够达到的通量. 后续发展的各类表面展示技术(surface display)及流式细胞分选技术(flow cytometer, FCM)^[50]大大提高了筛选通量. 表面展示技术具有高效、灵敏等特点, 其中以噬菌体表面展示(phage display)^[51]和酵母表面展示(yeast surface display)^[52]最为常用, 为蛋白质突变体库高通量筛选, 尤其在抗体蛋白的研究中提供了很好的选择. 各类展示技术的原理大体相似, 即将表达的多肽以融合蛋白形式展现在核糖体、病毒或细胞表面, 并使其保持相对独立的空间结构和生物活性, 借以研究多肽的性质、相互识别和作用, 筛选特定功能的多肽结构, 实现蛋白质的定向进化. 流式细胞分选技术能够实现单细胞的依次高速通过激光聚焦监测点, 经激发光激发后产生荧光信号, 根据光信号的变化来判断该细胞的大小、形态以及荧光强度, 并可以将需要

得到的细胞亚群从中分选出来. 除此之外, 由于 FCM 进行筛选时, 若标记底物与文库结合, 则相应的细胞就被带上荧光标记, 而未结合的荧光底物则可通过多次洗涤去除^[49]. 因此, FCM 不仅可以根据荧光强度分选出目标蛋白, 还可以经过多轮筛选达到富集的效果. 近年, 微流控技术(droplet microfluidic technology)^[53]受到许多学者的关注. 虽然与 FCM 相比其筛选速度约低一个数量级, 且体系组装复杂、稳定性较差, 但其功能灵活多样、价格低廉, 因而有潜力得到广泛应用^[54].

2 蛋白质定向进化的应用

2.1 改善酶催化效率

蛋白质定向进化可以有效地提高酶活性^[55]、解除抑制作用^[56]. 2012 年, Liao 等^[55]利用 epPCR 方法对植酸酶进行进化后, 与出发菌相比, 突变酶的比活力增强了 61%, 酶与底物亲和力增加了 53%, 催化效率提高了 84%. Chen 等^[57]将蛋白质分子动力学与定向进化结合, 在统计天冬氨酸激酶(AK3)家族的序列信息后, 确定了与天冬氨酸激酶运动和耦合作用相关的残基. 随后对这些残基进行点突变, 新得到了 12 个对赖氨酸抑制效果不敏感的 AK3 突变株, 以及 6 株减弱了天冬氨酸脱氢酶 I - 高丝氨酸脱氢酶(AK1-HD1)变构抑制作用的突变株.

2.2 提高酶稳定性

蛋白酶的稳定性主要体现在对 pH 值的稳定性和热稳定性两方面^[58-60], 其对生物催化效率有着重要影响. 酶稳定性的提高是蛋白质定向进化的重要应用之一.

2012 年, Liu 等^[61]为获得耐酸性的地衣芽孢杆菌 α -淀粉酶(BLA), 利用 epPCR 对 BLA 进行了定向进化, 并得到了两个重要的突变位点: T353I 和 H400R. 相比较与野生菌株, 单一位点突变株 Thr353Ile、His400Arg 以及双位点突变株 Thr353Ile/His400Arg 在 pH 4.5 的条件下的 K_{cat}/K_m 值分别提高了 3.5 倍、6.0 倍和 11.3 倍, 且 Thr353Ile/His400Arg 突变株更耐受低 pH 条件, 同时其热稳定性没有明显改变.

2013 年, Zheng 等^[62]用 epPCR 对天然坎皮纳斯类芽孢杆菌(*Paenibacillus campinasensis*)家族 -11 木聚糖酶(XynG1-1)进行了随机突变, 并从第二轮突变库中获得了最适 pH 值显著提高(pH 7.0~9.0)的突变株 XynG1-1B43(V90R/P172H). 随后, 利用

定点突变方法进行单氨基酸置换(D16Y),同时引入一个二硫键(T84C-T182C),这使得最终突变株 XynG1-1B43cc16 (V90R/P172H/T84C-T182C/D16Y)的最适温度提高 10℃,达到 60℃~70℃.

2.3 改变底物特异性

木糖是自然界中第二大丰富的可再生能源,利用木糖发酵生产乙醇有重大的经济潜能,但主要用于生产乙醇的酿酒酵母不能够有效利用木糖.将树干毕赤酵母中的木糖还原酶(PsXR)和木糖醇脱氢酶(PsXDH)引入酿酒酵母,经过两步氧化还原反应可以使木糖转变为木酮糖,从而被利用.但由于 PsXR 和 PsXDH 均特异性依赖 NADPH^[63],因此这种改造导致了细胞内还原力不平衡,降低了乙醇得率并生成大量副产物木糖醇.为消除胞内的氧化还原不平衡,人们试图将依赖 NADPH 的 PsXR 和 PsXDH 改造为 NADH 依赖性.

PsXDH 的底物特异性改造^[64]取得了一定的效果,而对 PsXR 的改造没有显著效果. Liang 等^[65]为解决这一问题,分析了 PsXR 的序列信息,发现其与纤细念珠菌木糖还原酶(CtXR)有 76%的同源结构,而 CtXR 的结构以及其与辅酶的相互作用已经被研究得十分详细,且 PsXR 与 CtXR 的序列比对结果也显示了作用于辅酶的氨基酸残基十分保守. Liang 等利用同源建模获得了 6 个 PsXR 中参与辅酶相互作用的重要的残基.将每一轮突变得到的最佳突变株作为下一轮突变的模板,经过几轮点饱和突变,最终得到了特异性常数之比(K_{cat}/K_m)^{NADH}/ $(K_{cat}/K_m)^{NADPH}$ 达 42 的突变菌株.虽然突变株中 PsXR 对 NADH 和 NADPH 的亲和力都有所减弱,但已经成功改变了 PsXR 对辅因子的结合特异性.

2.4 创造新型催化活性蛋白酶

随着合成生物学的快速发展,设计目标生物体性状,甚至构建新型蛋白质成为了蛋白质定向进化的重要发展方向.

乙二醛酶 II 是乙二醛酶系统的重要组成部分,催化 S-D- 乳酰谷胱甘肽水解生成 D- 乳酸和谷胱甘肽,其在结构上与其他金属水解酶如 β 内酰胺酶以及芳基硫酸酯酶水解酶有相似之处.为在现有的蛋白支架上改变酶的催化活性, Park 等^[66]同时对金属水解酶的催化活性位点和结合位点进行了分析,并对其活性位点的环状区域进行了碱基的插入或删除,从而将 β 内酰胺酶活性引入到了乙二醛酶 II 的 $\alpha\beta/\beta\alpha$ 金属水解酶支架上.最终得到的进化蛋白 evMBL8 完全不表现水解酶活性,但具有 β 内

酰胺酶活性,且其催化头孢噻肟水解的(K_{cat}/K_m)^{app} 值为每秒 1.8×10^2 mol/L,表达 evMBL8 的 *E. coli* 对头孢噻肟的耐药性提高了 100 倍.尽管 evMBL8 的 β 内酰胺酶活性低于天然 β 内酰胺酶,但此案例仍成功显示了蛋白质新功能可以通过定向进化的方法获得.除此之外,蛋白质定向进化不仅能够成功构建酶的新型催化功能,还可以获得非天然的催化反应,如催化具有非天然底物的反应等^[67].且结合蛋白质随机进化策略可以进一步提高其催化效率.

3 展 望

随着蛋白质定向进化的快速发展,其已成为代谢工程与合成生物学不可或缺的工具,在人类健康、环境、能源等多个方面均具有重要的应用.定向改进蛋白酶可以改善细胞代谢途径,优化细胞性能.目前,定向进化取得的成果大多为改进单个蛋白,因而其在代谢工程中的应用也多局限在某些已知代谢通路的个别蛋白质中.近年来,基于重组的体内定向进化技术得到了很快的发展,有望实现在代谢途径、网络和细胞水平上进行改造.除了前文提及的 MAGE、MIPE 等体内重组技术有望成为蛋白质定向进化的新技术外,最近发展的 Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated Cas9 (CRISPR/Cas9)^[68] 基因组定点编辑技术也具有体内蛋白质定向进化应用潜力. CRISPR/Cas9 应用宿主范围广,成功地应用于人类细胞、斑马鱼和小鼠以及细菌的基因组精确修饰,修饰类型包括基因定点 InDel 突变、基因定点敲入、多位点同时突变和小片段的缺失^[69].因此, CRISPR/Cas9 理论上能够实现基因内多个位点的定点修饰或饱和突变.由于其突变效率高、制作简单及成本低的特点,将会是应用前景广阔的基因组定点改造分子工具和蛋白质定向进化新技术.

无论定向进化某个特定蛋白还是细胞全局,筛选效率均是其最大的限制因素,尽管理性设计能够对结果做出一定的预测,降低了蛋白质定向进化对筛选效率的依赖性,但仍需繁复的实验验证.并且对于新型功能蛋白质的获得而言,从头设计方法得到的酶活性较低,如设计磷酸丙糖异构酶及分支酸变位酶十分困难等^[70-71].因而,除了发展先进的定向进化技术外,开发高突变效率的进化策略和易行通用的高通量筛选方法仍是未来蛋白质定向进化亟待解决的问题.

参 考 文 献

- [1] 徐卉芳, 张先恩, 张用梅. 体外分子定向进化研究进展. 生物化学与生物物理进展, 2002, **29**(4): 518–522
Xu H F, Zhang X E, Zhang Y M. Prog Biochem Biophys, 2002, **29**(4): 518–522
- [2] Winter G, Fersht A R, Wilkinson A J, *et al.* Redesigning enzyme structure by site-directed mutagenesis: tyrosyl tRNA synthetase and ATP binding. Nature, 1982, **299**(5885): 756–758
- [3] Ulmer K M. Protein engineering. Science, 1983, **219**(4585): 666–671
- [4] Bornscheuer U T, Huisman G W, Kazlauskas R J, *et al.* Engineering the third wave of biocatalysis. Nature, 2012, **485**(7397): 185–194
- [5] Cobb R E, Chao R, Zhao H. Directed evolution: past, present, and future. AICHE J, 2013, **59**(5): 1432–1440
- [6] Rollié S, Mangold M, Sundmacher K. Designing biological systems: systems engineering meets synthetic biology. Chemical Engineering Science, 2012, **69**(1): 1–29
- [7] Zhao H. Directed evolution of novel protein functions. Biotechnol Bioeng, 2007, **98**(2): 313–317
- [8] Kipnis Y, Dellus-Gur E, Tawfik D S. TRINS: a method for gene modification by randomized tandem repeat insertions. Protein Eng Des Sel, 2012, **25**(9): 437–444
- [9] Chen Z, Zeng A P. Protein design in systems metabolic engineering for industrial strain development. Biotechnol J, 2013, **8**(5): 523–533
- [10] Ruff A J, Dennig A, Schwaneberg U. To get what we aim for——progress in diversity generation methods. FEBS J, 2013, **280**(13): 2961–2978
- [11] Balci H, Ozturk M T, Pijning T, *et al.* Improved activity and pH stability of *E. coli* ATCC 11105 penicillin acylase by error-prone PCR. Appl Microbiol Biotechnol, 2014, **98**(10): 4467–4477
- [12] Ben Mabrouk S, Ayadi D Z, Ben Hlima H, *et al.* Thermostability improvement of maltogenic amylase MAUS149 by error prone PCR. J Biotechnol, 2013, **168**(4): 601–606
- [13] Chong H, Geng H, Zhang H, *et al.* Enhancing *E. coli* isobutanol tolerance through engineering its global transcription factor cAMP receptor protein (CRP). Biotechnol Bioeng, 2013, **111**(4): 700–708
- [14] Wang Y, Feng S, Zhan T, *et al.* Improving catalytic efficiency of endo-beta-1, 4-xylanase from *Geobacillus stearothermophilus* by directed evolution and H179 saturation mutagenesis. J Biotechnol, 2013, **168**(4): 341–347
- [15] Dana C M, Saija P, Kal S M, *et al.* Biased clique shuffling reveals stabilizing mutations in cellulase Cel7A. Biotechnol Bioeng, 2012, **109**(11): 2710–2709
- [16] Zhang H, Chong H, Ching C B, *et al.* Random mutagenesis of global transcription factor cAMP receptor protein for improved osmotolerance. Biotechnol Bioeng, 2012, **109**(5): 1165–1172
- [17] Yu X W, Wang R, Zhang M, *et al.* Enhanced thermostability of a *Rhizopus chinensis* lipase by *in vivo* recombination in *Pichia pastoris*. Microb Cell Fact, 2012, **11**(1): 1475–2859
- [18] Savile C K, Janey J M, Mundorff E C, *et al.* Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. Science, 2010, **329**(5989): 305–309
- [19] Li T, Liang J, Ambrogelly A, *et al.* Efficient, chemoenzymatic process for manufacture of the Boceprevir bicyclic [3.1.0] proline intermediate based on amine oxidase-catalyzed desymmetrization. J Am Chem Soc, 2012, **134**(14): 6467–6472
- [20] Mosberg J A, Gregg C J, Lajoie M J, *et al.* Improving lambda red genome engineering in *Escherichia coli* via rational removal of endogenous nucleases. PLoS One, 2012, **7**(9): e44638
- [21] Li Y, Gu Q, Lin Z, *et al.* Multiplex iterative plasmid engineering for combinatorial optimization of metabolic pathways and diversification of protein coding sequences. ACS Synth Biol, 2013, **2**(11): 651–661
- [22] Stemmer W P. Rapid evolution of a protein *in vitro* by DNA shuffling. Nature, 1994, **370**(6488): 389–391
- [23] Zhao H, Giver L, Shao Z, *et al.* Molecular evolution by staggered extension process (StEP) *in vitro* recombination. Nat Biotechnol, 1998, **16**(3): 258–261
- [24] Davids T, Schmidt M, Bottcher D, *et al.* Strategies for the discovery and engineering of enzymes for biocatalysis. Curr Opin Chem Biol, 2013, **17**(2): 215–220
- [25] Zhang Y, Buchholz F, Muirers J P P, *et al.* A new logic for DNA engineering using recombination in *Escherichia coli*. Nat Genet, 1998, **20**(2): 123–128
- [26] Wang H H, Isaacs F J, Carr P A, *et al.* Programming cells by multiplex genome engineering and accelerated evolution. Nature, 2009, **460**(7257): 894–898
- [27] Damborsky J, Brezovsky J. Computational tools for designing and engineering biocatalysts. Curr Opin Chem Biol, 2009, **13**(1): 26–34
- [28] Pavelka A, Chovancova E, Damborsky J. HotSpot Wizard: a web server for identification of hot spots in protein engineering. Nucleic Acids Res 2009, **37**(Web Server issue): W376–W383
- [29] Fox R J, Davis S C, Mundorff E C, *et al.* Improving catalytic function by ProSAR-driven enzyme evolution. Nat Biotechnol, 2007, **25**(3): 338–344
- [30] Hellberg S, Sjoström M, Wold S. The prediction of bradykinin potentiating potency of pentapeptides. An example of a peptide quantitative structure-activity relationship. Acta Chem Scand B, 1986, **40**(2): 135–140
- [31] Feng X, Sanchis J, Reetz M T, *et al.* Enhancing the efficiency of directed evolution in focused enzyme libraries by the adaptive substituent reordering algorithm. Chemistry-A European J, 2012, **18**(18): 5646–5654
- [32] O'Maille P E, Bakhtina M, Tsai M-D. Structure-based combinatorial protein engineering (SCOPE). J Mol Biol, 2002, **321**(4): 677–691
- [33] DeSantis G, Wong K, Farwell B, *et al.* Creation of a productive, highly enantioselective nitrilase through gene site saturation mutagenesis(GSSM). J Am Chem Soc, 2003, **125**(38): 11476–11477
- [34] Reetz M T, Krebs G P L. Challenges in the directed evolution of stereoselective enzymes for use in organic chemistry. Comptes Rendus Chimie, 2011, **14**(9): 811–818
- [35] Reetz M. Directed evolution of enantioselective enzymes as

- catalysts for organic synthesis. *Adv Catalysis*, 2006, **49**: 1–69
- [36] Reetz M T, Wang L W, Bocola M. Directed evolution of enantioselective enzymes: iterative cycles of CASTing for probing protein-sequence space. *Angewandte Chemie*, 2006, **118**(8): 1258–1263
- [37] Reetz M T, Carballeira J D, Vogel A. Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angew Chem Int Ed Engl*, 2006, **45**(46): 7745–7751
- [38] Chockalingam K, Chen Z, Katzenellenbogen J A, *et al.* Directed evolution of specific receptor-ligand pairs for use in the creation of gene switches. *Proc Natl Acad Sci USA*, 2005, **102**(16): 5691–5696
- [39] Kries H, Blomberg R, Hilvert D. *De novo* enzymes by computational design. *Curr Opin Chem Biol*, 2013, **17**(2): 221–228
- [40] Simons K T, Bonneau R, Ruczinski I, *et al.* Ab initio protein structure prediction of CASP III targets using ROSETTA. *Proteins: Structure, Function, and Bioinformatics*, 1999, **37**(S3): 171–176
- [41] Mulholland A. Computational enzymology: modelling the mechanisms of biological catalysts. *Biochem Soc Trans*, 2008, **36**(1): 22–26.
- [42] Zhang X, DeChancie J, Gunaydin H, *et al.* Quantum mechanical design of enzyme active sites. *J Org Chem*, 2008, **73**(3): 889–899
- [43] Zanghellini A, Jiang L, Wollacott A M, *et al.* New algorithms and an *in silico* benchmark for computational enzyme design. *Protein Sci*, 2006, **15**(12): 2785–2794
- [44] Acevedo O, Jorgensen W L. Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions. *Acc Chem Res*, 2009, **43**(1): 142–151.
- [45] Marcheschi R J, Groneberg L S, Liao J C. Protein engineering for metabolic engineering: Current and next-generation tools. *Biotechnology Journal*, 2013, **8**(5): 545–555
- [46] Siegel J B, Zanghellini A, Lovick H M, *et al.* Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science*, 2010, **329**(5989): 309–313
- [47] Rothlisberger D, Khersonsky O, Wollacott A M, *et al.* Kemp elimination catalysts by computational enzyme design. *Nature*, 2008, **453**(7192): 190–195
- [48] Privett H K, Kiss G, Lee T M, *et al.* Iterative approach to computational enzyme design. *Proc Natl Acad Sci USA*, 2012, **109**(10): 3790–3795
- [49] Korendovych I V, Kulp D W, Wu Y, *et al.* Design of a switchable eliminase. *Proc Natl Acad Sci USA*, 2011, **108**(17): 6823–6827
- [50] Ruff A J I, Dennig A, Wirtz G, *et al.* Flow cytometer-based high-throughput screening system for accelerated directed evolution of P450 monooxygenases. *ACS Catalysis*, 2012, **2**(12): 2724–2728
- [51] Kushwaha R, Downie A B, Payne C M. Uses of phage display in agriculture: sequence analysis and comparative modeling of late embryogenesis abundant client proteins suggest protein-nucleic acid binding functionality. *Comp Math Methods Med*, 2013, **2013**: 470390
- [52] Yi L, Gebhard M C, Li Q, *et al.* Engineering of TEV protease variants by yeast ER sequestration screening (YESS) of combinatorial libraries. *Proc Natl Acad Sci USA*, 2013, **110**(18): 7229–7234
- [53] Brouzes E, Medkova M, Savenelli N, *et al.* Droplet microfluidic technology for single-cell high-throughput screening. *Proc Natl Acad Sci USA*, 2009, **106**(34): 14195–14200
- [54] Guo M T, Rotem A, Heyman J A, *et al.* Droplet microfluidics for high-throughput biological assays. *Lab Chip*, 2012, **12**(12): 2146–2155
- [55] Liao Y, Zeng M, Wu Z F, *et al.* Improving phytase enzyme activity in a recombinant phyA mutant phytase from *Aspergillus niger* N25 by error-prone PCR. *Appl Biochem Biotechnol*, 2012, **166** (3): 549–562
- [56] Bissaro B, Saurel O, Arab-Jaziri F, *et al.* Mutation of a pH-modulating residue in a GH51 α -l-arabinofuranosidase leads to a severe reduction of the secondary hydrolysis of transuranosylation products. *Biochim Biophys Acta*, 2014, **1840**(1): 626–636
- [57] Chen Z, Rappert S, Sun J, *et al.* Integrating molecular dynamics and co-evolutionary analysis for reliable target prediction and deregulation of the allosteric inhibition of aspartokinase for amino acid production. *J Biotechnol*, 2011, **154**(4): 248–254
- [58] Koyama Y, Hidaka M, Nishimoto M, *et al.* Directed evolution to enhance thermostability of galacto-N-biose / lacto-N-biose I phosphorylase. *Protein Eng Des Sel*, 2013, **26**(11): 755–761
- [59] Steffler F, Guterl J-K, Sieber V. Improvement of thermostable aldehyde dehydrogenase by directed evolution for application in Synthetic Cascade Biomanufacturing. *Enzyme Microb Technol*, 2013, **53**(5): 307–314
- [60] Bora L, Gohain D, Das R. Recent advances in production and biotechnological applications of thermostable and alkaline bacterial lipases. *J Chem Technol Biotechnol*, 2013, **88**(11): 1959–1970
- [61] Liu Y H, Hu B, Xu Y J, *et al.* Improvement of the acid stability of *Bacillus licheniformis* α amylase by error-prone PCR. *J Appl Microbiol*, 2012, **113**(3): 541–549
- [62] Zheng H, Liu Y, Sun M, *et al.* Improvement of alkali stability and thermostability of *Paenibacillus campinasensis* Family-11 xylanase by directed evolution and site-directed mutagenesis. *J Ind Microbiol Biotechnol*, 2014, **41**(1): 153–162
- [63] Watanabe S, Saleh A A, Pack S P, *et al.* Ethanol production from xylose by recombinant *Saccharomyces cerevisiae* expressing protein-engineered NADH-preferring xylose reductase from *Pichia stipitis*. *Microbiology*, 2007, **153**(9): 3044–3054
- [64] Watanabe S, Kodaki T, Makino K. Complete reversal of coenzyme specificity of xylitol dehydrogenase and increase of thermostability by the introduction of structural zinc. *J Biol Chem*, 2005, **280**(11): 10340–10349
- [65] Liang L, Zhang J, Lin Z. Altering coenzyme specificity of *Pichia stipitis* xylose reductase by the semi-rational approach CASTing. *Microb Cell Fact*, 2007, **6**(36): 1–11
- [66] Park H-S, Nam S-H, Lee J K, *et al.* Design and evolution of new catalytic activity with an existing protein scaffold. *Science*, 2006, **311**(5760): 535–538
- [67] Brustad E M, Arnold F H. Optimizing non-natural protein function

- with directed evolution. *Curr Opin Chem Biol*, 2011, **15**(2): 201–210
- [68] Jiang W, Bikard D, Cox D, *et al.* RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol*, 2013, **31**(3): 233–239
- [69] 方 锐, 畅 飞, 孙照霖, 等. CRISPR/Cas9 介导的基因组定点编辑技术. *生物化学与生物物理进展*, 2013, **40**(8): 691–702
- Fang R, Chang F, Sun Z L, *et al.* *Prog Biochem Biophys*, 2013, **40**(8): 691–702
- [70] Richter F, Leaver-Fay A, Khare S D, *et al.* *De novo* enzyme design using Rosetta3. *PLoS One*, 2011, **6**(5): e19230
- [71] Lu Y, Yeung N, Sieracki N, *et al.* Design of functional metalloproteins. *Nature*, 2009, **460**(7257): 855–862

The Research Progress of Protein Directed Evolution*

WANG Xiao-Yue^{1,2,3)**}, WANG Bai-Yun^{1,2,3)**}, WANG Zhi-Wen^{1,2,3)***}, CHEN Tao^{1,2,3)}, ZHAO Xue-Ming^{1,2,3)}

⁽¹⁾ School of Chemical Engineering & Technology, Tianjin University, Tianjin 300072, China;

⁽²⁾ Key Laboratory of Systems Bioengineering Ministry of Education, Tianjin University, Tianjin 300072, China;

⁽³⁾ Collaborative Innovation Center of Chemical Science and Engineering(Tianjin), Tianjin 300072, China)

Abstract Protein directed evolution is widely used to improve enzymes, particularly for industrial biocatalytic processes and construction of cell factories. It is an efficient and powerful tool to improve and optimize natural proteins in order to generate robust biocatalysts for practical applications. In addition, optimization of metabolic pathways, regulation of functional regulatory systems, and development of desired complex phenotypes in industrial host organisms have all been achieved by way of protein directed evolution. Numerous *in vivo* and *in vitro* methods have been developed for the efficient evolutionary effects, especially in high mutation rate and rapidly high-throughput screening capabilities. Some of the methods have only recently been applied for general use and are just beginning to find greater application. In this review, we summarize some of the new methods for mutant libraries generation, including random evolution, semi-rational evolution and rational evolution. And current state-of-the-art screening techniques in protein directed evolution are also reviewed. Advancements are discussed with respect to the state of the art in diversity generation and high-throughput screening capabilities. Meanwhile, limitations and remaining challenges are also pronounced.

Key words protein directed evolution, random evolution, semi-rational evolution, rational evolution, screening

DOI: 10.16476/j.pibb.2014.0059

* This work was supported by grants from The National Basic Research Program of China (2011CBA00804, 2012CB725203), The National Natural Science Foundation of China (21206112, 21390201), Hi-Tech Research and Development Program of China (2012AA022103, 2012AA02A702) and The Innovation Foundation of Tianjin University (1308).

**These authors contributed equally to this work.

***Corresponding author.

Tel: 86-22-27406582, E-mail: zww@tju.edu.cn

Received: May 13, 2014 Accepted: July 13, 2014