

生物信息学在工业生物催化研究中的应用

于慧敏* 罗 晖 史 悦 孙旭东 沈忠耀

(清华大学化工系 生物化工研究所, 北京 100084)

摘 要 随着石油等不可再生资源的日益减少以及环境污染问题的日益严重,应用工业生物催化技术改造或取代传统化工工艺已经成为新世纪化学工业可持续发展的研究热点。工业生物催化技术的研究对象是生物催化剂及其催化过程。近来,利用生物信息学技术进行工业生物催化研究已经越来越受到人们的重视。随着工业生物催化技术的发展,生物信息学将直接指导并加快新型高效生物催化剂的发现及功能改造进程。

关键词 工业生物催化,生物催化剂,生物信息学

中图分类号 Q811.4 文献标识码 A 文章编号 1000-3061(2004)03-0325-07

经过数百年的发展,化学工业已经成为我国国民经济的基础支柱产业之一。然而,随着石油等不可再生资源的日益减少以及环境污染问题的日益严重,能源问题、可持续发展问题越来越引起人们的关注,应用生物催化技术改造或取代传统化工工艺已经成为新世纪化学工业可持续发展的研究热点^[1]。

1 工业生物催化的兴起

工业生物催化是生物技术应用中一个新兴和关键的领域。与传统的化学合成相比,生物催化合成具有高效、高选择性、环境友好、可以合成一些用化学方法无法生产的手性化合物及高分子等独特的优点。因此,它在医药化工、精细化工乃至传统化工等领域均有广阔的发展空间,并已经有越来越多的化工产品的生产工艺由化学法成功地向生物催化法转变,如乙醛酸、丙烯酸、壳聚糖、单甘酯、透明质酸、黄原胶以及多种氨基酸和抗生素等大宗化学品、表面活性剂、功能高分子和医药产品的生产^[1]。仅以日本 Daichi Fine Chemicals 公司 1999 年开发的外消旋泛内酯的酶法生产工艺为例,与以前的化学方法相比,生物催化法不仅具有令人非常满意的经济效益,而且具有很好的环保效应(少消耗 49% 的水和 62% 的氧,少产生 30% 的二氧化碳)^[2]。

随着公众环保意识、可持续发展意识、健康意识的加强,人们在商品取向上会越来越倾向使用生物制品和新型环保产品,因此社会对新技术、新产品,尤其是手性药物、新兴健康食品等的渴求,就要求生物催化技术能够更快更好地发展。继医药和农业之后,工业生物催化已经被广泛看作是“生物技术的第三次浪潮”,生物催化剂(工业酶)则被视为

21 世纪化学工业可持续发展的必要工具^[3]。

目前已经开发的商品酶有 200 种左右,而工业上应用的酶仅有 50 多种,这说明酶工程仍然是一门年轻的学科,有着广阔的发展前景。因此,许多国家都在致力于酶工程的研究,以提高生物催化用酶的活性、稳定性和单位产率,并逐步实现酶的工业化应用。以日本为例,日本是最早开始发展工业生物催化技术的国家之一。目前,日本有多家大学、研究机构和企业均在致力于工业生物催化工艺和催化剂的研究和开发,并在氨基酸、甜味剂、油、维生素、化学品以及药物中间体等众多产品的生物催化法生产方面处于领先水平。表 1 列出了日本近年来采用微生物酶作为生物催化剂进行工业生产的部分产品^[2]。

近年来,我国在采用工业生物催化法生产重要生物产品、生物材料和生物能源等方面的研究也取得了长足的进展。就其规模而言,我国的氨基酸(如谷氨酸)、有机酸(如柠檬酸、乳酸)等产品在世界上占有举足轻重的位置。但是,我国的技术水平尚与发达国家具有较大的差距。例如,氨基酸的整体水平国外比我国高 30%~50%,我国酶制剂中的支柱产品如糖化酶,产酶活力仍停留在 32000IU/mL 左右,而国外已经达到 50000IU/mL;由于成本偏高及质量不稳定,我国酶制剂产品难以进入国际市场;在抗生素工业中,作为我国医药工业战略品种之一的头孢菌素类抗生素的酶法工艺尚不能实现产业化。为此,在 2001 年全面启动并实施的“十五”国家科技攻关项目中,“生物化工原料和新材料”、“新型酶制剂与重大酶制剂产品”、“氨基酸新产品、新工艺”、“寡糖新产品的开发与应用”、“新型食品及饲料添加剂的开发与应用”、“重大抗生素生产新技术、新工艺”和“农用抗生素新技术、新工艺”等课题被列入“生物工程关键技术与重大产品”开发领

收稿日期 2003-09-29, 修回日期 2003-12-29。

基金项目 国家自然科学基金面上项目(No. 20206014);全国优秀博士学位论文作者专项资金资助项目。

* 通讯作者。 Tel 86-10-62788568; Fax 86-10-62770304; E-mail yuhm@tsinghua.edu.cn

© 中国科学院微生物研究所期刊联合编辑部 http://journals.im.ac.cn

域。可以说 ,工业生物催化在传统化学反应过程中的普及和推广将是化学工业的一次彻底的绿色革命^[4]。

表 1 近年来日本微生物酶的工业应用

Table 1 Recent industrial applications of microbial enzymes in Japan^[2]

Item	Product	Year	Process
Amino acids	D-p-Hydroxy- phenylglycine	1979	Enzymatic
		1995	
	Aspartate	1984	Membrane bioreactor
		1986	Membrane bioreactor
	DOPA	1994	Enzymatic
	Hydroxyproline	1997	Enzymatic
Sweeteners	Paratinose	1984	Immobilized bioreactor
	Aspartame	1987	Enzymatic
	Lactosucrose	1990	Enzymatic
	Galacto-oligo- saccharide	1990	Enzymatic
	Maltotriose	1990	Immobilized bioreactor
	Engineered stevia sweetner	1993	Enzymatic
	Theande-oligosaccharide	1994	Immobilized bioreactor
	Trehalose	1995	Enzymatic
	Nigero-oligosaccharide	1998	Enzymatic
Oils	Physiologically functional oils	1989	Immobilized bioreactor
		1990	Enzymatic
		1998	Enzymatic
	Polyunsaturated fatty acids	1998	Enzymatic
Vitamins	Stabilized vitamin C	1990	Enzymatic
	Nicotinamide	1998	Enzymatic
	Vitamin C phosphate	1999	Enzymatic
	Pantothenate intermediate	1999	Enzymatic
Chemicals	Acrylamide	1988	Immobilized bioreactor
	Chiral epoxides	1985	Enzymatic
Pharma intermediates	Herbesser intermediate	1992	Membrane bioreactor
	Chiral alcohols	2000	Enzymatic
Others	Casein phosphopeptide	1988	Enzymatic
	Hypoallergenic rice	1991	Enzymatic
	Hypoallergenic protein	1991	Enzymatic

然而 ,由于天然蛋白质往往只能在自然条件下起到最佳功能 ,而在工业生产中常见的高温、高压或高酸碱度等条件下 ,大多数蛋白质都会失活。因此 ,要成功获得稳定、高效且专一的生物催化剂 ,就必须对蛋白质进行改造 ,使其能够在特定的条件下起到特定的功能。为了加快生物催化剂的开发进程 ,迅速抢占生物学前沿 ,应用近年来蓬勃发展的各种新兴生物技术——尤其是生物信息学技术 ,进行生物催化剂的发现和改性研究是必然的途径。

2 生物信息学的产生、发展与现状

21 世纪是生命科学的世纪 ,其里程碑就是于 1990 年 10 月启动的、耗资数十亿的著名的人类基因组计划(Human Genome Project ,HGP)^[5] ,而应人类基因组计划和生物科学迅猛发展的要求而迅速兴起的生物信息学则历史性地成为生命科学浪潮中当仁不让的弄潮儿。

生物信息学(bioinformatics)一词是由林华安博士于 1987 年提出的 ,而它的起源最早可以追溯到 20 世纪 50 年代末计

算机在生物学研究中的应用工作^[6,7]。到了 20 世纪末期 ,继多年的沉默之后 ,伴随着计算机技术和网络的革命性发展 ,生物信息学也突飞猛进地发展起来。简单地说 ,生物信息学的实质就是利用计算机科学和网络技术来解决生物学问题。

生物信息学研究建立在三方面的知识基础之上 ,分别为 :发达的、复杂的、可相互交流的数据库系统 ,强有力的创新算法和软件以及自动化的、大规模的、高通量的生物学研究方法与平台技术^[8]。而生物信息学的研究内容则大部分集中在基因组、蛋白质组、蛋白质结构以及与之相结合的药物设计上^[8-11]。

国外一直非常重视生物信息学的发展 ,欧美等发达国家在生物信息方面已有较长时间的积累。在数据库方面 ,目前 ,绝大部分的核酸和蛋白质数据由美国、欧洲和日本的 GenBank/EMBL/DDBJ 这 3 家数据库系统产生 ;在算法和软件方面 ,多序列比对、人工神经网络、Threading 方法、模拟退火算法等正在日益引起生物信息学家的关注 ,而由美国多家生

物公司开发的序列分析软件 BioEdit、多功能软件 Vector NTI Suite、引物设计软件 Oligo、蛋白质三维分子结构显示软件 RasMol、序列多重对齐软件 ClustalW 等也因其强大的功能和友好的界面而受到使用者的青睐;同时,遍布世界的国际互联网为数据库系统、创新算法和软件的迅猛发展提供了无限的空间,而一些专业的生物信息学研究机构,如美国国家生物技术信息中心(NCBI)、欧洲生物信息学研究所(EBI)和日本信息生物学中心(CIB)等,则为生物信息学的长足发展提供了强有力的支撑^[11,12]。从生物学实验技术角度来讲,一方面,具有高集成度、高并行处理能力及可自动化分析等特点的 DNA 芯片和微阵列制样技术越来越多地用于获得基因表达的功能谱、DNA 的快速测序、DNA 突变检测以及药物筛选等;另一方面,二维凝胶电泳和生物质谱等技术也越来越广泛地应用于高通量的蛋白质分离与鉴定、蛋白质的分子量、质量肽谱、突变体、翻译后修饰、配位体、活力部位、折叠和高级结构的分析与测定等等^[5,8]。

国内对生物信息学领域也越来越重视。在一些著名院士和教授的带领下,在各自领域取得了一定成绩,有的在国际上还占有一席之地。然而,由于国内的生物信息学研究起步较晚,因此在数据库、网络、算法与软件、研究机构、生物学实验技术平台乃至应用等方面均有较大的差距。

3 生物信息学在生物催化领域的应用

由于生物信息学独特的桥梁作用和整合作用,因此它不仅仅是一门科学学科,更是一种重要的研究开发工具。M. Vihinen 在 2001 年提出了一个“洋葱”结构模型,对基因和蛋白质功能研究的实验方法和生物信息学方法进行了形象的描述(图 1)。他认为,研究人员进行科学研究的实验过程就像“剥洋葱”一样,由外而内,由实验到理论,而生物信息学方法则不同,它是由理论到实验,由内而外来进行^[10]。因此,充分利用互联网上的免费生物信息学资源来指导生物学实验研究,将为传统的以实验为基础的生物科学指出明确的理论方向,从而大大加快研究的进程。有鉴于此,近 10 年来,生物信息学已经在医药、农业、微生物、寄生虫学、糖生物学以及食品和健康等领域均开辟了非常广泛的发展空间^[13-17]。

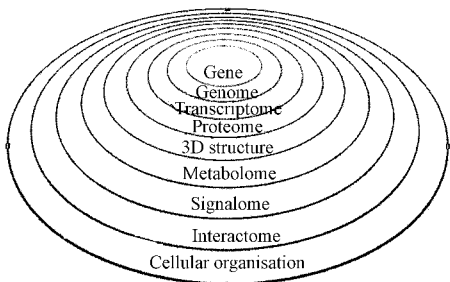


图 1 进行基因和蛋白质功能研究的实验和生物信息学方法的“洋葱”结构模型

Fig.1 The onion model of experimental and bioinformatics methods to study the function of genes and proteins^[10]

可以预见,随着工业生物催化技术的发展,生物信息学技术在生物催化领域的应用也必将异军突起,直接指导新型高效生物催化剂的发现和改造进程,从而发挥举足轻重的作用。

3.1 应用生物信息学分析指导生物催化剂基因发现与获取

生物信息学技术在生物催化领域应用的第一个重要方向就是对生物催化剂基因的发现或获取进行直接指导。基于因特网上基因组和基因序列数据库中公布的基因数据的快速增长和比较基因组学的迅速发展,从大量的数据资源中寻找新的基因,发掘其蕴涵的生物学信息,从而为新型高效生物催化剂的发现服务目前已经成为现实^[18]。

众所周知,传统的工业酶的发现是基于微生物多样性基础上的大规模菌株筛选。以原手性羧基化合物进行生物还原生产化学纯乙醇为例,为得到性能较好的目标菌株,日本的研究人员对上千株细菌、酵母菌或霉菌等进行了筛选^[19]。然而,一般而言,通过传统筛选方式进行菌株筛选的过程不仅工作量大而且周期长,往往还会发生在大量的菌株筛选后仍然一无所获的情况。例如,在进行半合成头孢菌素母核—7-氨基头孢烷酸(7-ACA)的生物法生产工艺研究中,Binder 等人从 5000 株菌中才筛选出一株具有戊二酰基-7-氨基头孢烷酸(GL-7-ACA)酰化酶活性的假单胞菌 BL072。与此类似,本课题组为了获得具有 GL-7-ACA 酰化酶活性的菌株,对土壤中上千株假单胞菌进行了选择性培养和筛选,历时半年多,但最终以失败告终。有鉴于生物信息学技术的迅速发展,本课题组改变思路,直接调取土壤中经假单胞菌选择性培养获得的菌群的基因组 DNA,然后利用互联网上免费的生物信息学数据库资源和生物信息学软件,对已经公开的 5 组 GL-7-ACA 酰化酶的基因序列进行了分析,最终选定 EMBL 核酸序列数据库中 3 株假单胞菌的 GL-7-ACA 酰化酶基因序列,分别设计引物,对 GL-7-ACA 酰化酶基因进行 PCR 扩增,最终成功获得了 3 个 GL-7-ACA 酰化酶基因片段。将这 3 个基因序列与 EMBL 核酸数据库中其它假单胞菌 KAC-1, C427, GK16 和 130 的 GL-7-ACA 酰化酶基因序列进行比较,结果表明,本课题组获得的 DNA 片段确为 GL-7-ACA 酰化酶基因,且与文献报道的这几个基因序列具有很高的同源性(~ 96%)。对应蛋白质序列的同源性则达到 97% ~ 98%。更重要的是,经过基因的生物信息学序列比对分析,发现在所得的几个 GL-7-ACA 酰化酶基因中,与活性中心或酶的后加工有关位点的氨基酸没有突变。目前,这 3 个基因序列已向 EMBL 核酸数据库提交,得到的序列号分别为 AY311487、AY311488 和 AY311489^[20]。同样利用数据库检索、软件分析与引物设计等生物信息学手段,本课题组还成功地从一株基因组序列未知的诺卡氏菌中直接扩增获得了腈水合酶基因片段(AY168347),并进行了活性表达^[21]。Kim 对利用国际互联网上的生物信息学工具进行未知基因的引物设计的实施策略进行了总结^[22]。

在基因组序列分析中寻找和发现新基因的关键是高效的序列比对算法,一些搜索程序如 BLAST 是免费的强大搜索工具。Nevill-Manning 等人于 1998 年开发了一种称为

EMOTIF 的程序,用于在基因组数据库中对蛋白质家族的高特异性功能域进行搜索。通过对酵母基因组中未知功能蛋白质的 833 个阅读框进行 EMOTIF 分析,他们推断出了 100 多个新蛋白质的功能。而这些蛋白质功能的确定则对新工业生物催化剂的发现和应用具有重大价值^[23]。Chen 等人则通过对表达序列标签(Expressed Sequence Tags, ESTs)数据库的挖掘,发现了与人雌激素受体相关的两种新型核受体基因(*hERRb2* 和 *hERRc2*),并推导获得了这两种核受体的氨基酸序列和蛋白质序列,确定了这两种核受体蛋白在某些基因疾病发生时的胞内位置^[24]。Zharkov and Grollman 应用生物信息学方法对可以作为催化剂的蛋白质家族和超家族的氨基酸残基保守性进行了定量分析,分析结果阐明了蛋白质中氨基酸残基的功能相似性和相异性,并为定点突变实验设计提供了理论起点^[25]。Wishart 等人应用生物信息学方法对磷脂酶同族蛋白(*PTEN*)和肌肉微管蛋白(*MTMR*)的相关蛋白进行了研究。他们通过对人、苍蝇和线虫的全基因组数据库的 BLAST 搜索、多序列比对、进化树分析和蛋白质的结构与功能预测、基因功能和表达分析,发现并定位了 3 个新的人 *MTMR* 基因,并在果蝇和线虫基因组中发现了至少 6 个完全不同的 *MTMR* 基因,更重要的是,从底物结合特性、信号传导途径等方面分析了 *PTEN* 和 *MTMR* 相关蛋白的结构特性和生理功能,从理论的高度揭示了 *PTEN* 和 *MTMR* 相关蛋白与人类疾病的关系,从而为生物学动物模型实验提供了明确指导^[26]。Marrs 等人认为,相应于各种新兴生物学技术的进步,应用生物信息学分析来指导生物多样性筛选、基因组序列分析、定向进化和噬菌体展示等实验方法是当前工业酶发现的新途径^[27]。

正如 Searls 在 2000 年所指出的,生物信息学工具对于基因和药物的发现,对于海量数据资源——包括表达序列标签(ESTs)、微生物基因组序列、模式生物序列、多态性、基因表达数据、蛋白质组学数据等等——向工业应用的转化,都起到了并将继续起到非常关键的作用^[28]。

3.2 应用生物信息学分析指导生物催化剂的改性研究

生物信息学技术在生物催化领域应用的另一个重要方向就是对现有生物催化剂的功能改造进行直接指导。

在工业生物催化过程中,有时一个目标产品的获得需要经过多个酶的多步协同催化作用,如类胡萝卜素的生物法代谢合成^[29]。在这种情况下,生物催化剂的改性研究首先要以代谢网络分析为导向,确定代谢途径中的瓶颈催化步骤,从而确定需要进行改性的目标酶。从这个意义上讲,代谢工程技术在工业生物催化研究中的应用也具有广泛的空间^[29]。

对于目标工业酶的改性研究,最基本的思路 and 措施有两个,分别为理性设计和定向进化。

3.2.1 生物催化剂的理性设计:对于理性设计,顾名思义,就是从理论出发,对生物催化剂进行改性设计,再进行生物学实验,对设计结果进行实践和检验。Gustafsson 等人在 2003 年的一篇综述文章中指出,将生物信息学领域的数据挖

掘工具应用于催化剂设计,进行肽和蛋白质序列与活性的关系分析和新肽、新蛋白质设计,将为昂贵的序列分析和不可靠的高通量筛选等提供强有力的替代方法^[30]。例如,通过序列对齐统计分析对肽或蛋白质的单个氨基酸进行替换模拟,再进行实验验证,真菌植酸酶的热稳定性可以提高到 30℃ 以上,通过局部最小平方回归方法对多个氨基酸的替换引起的肽和蛋白质活性的变化进行数学关联和模拟预测来指导实验,神经激肽对神经肽底物 P 的亲性和乙酰胆碱酯酶在人 COS-1 细胞上的表达活性都可以大大提高。生物信息学方法同样可以用来指导多个氨基酸的交互作用研究^[30]。

应用生物信息学方法进行理性设计研究的最广泛应用就是生物催化剂的定点突变。在生物信息学技术的指导下,当前已经对枯草杆菌蛋白酶、二氢叶酸还原酶、胰蛋白酶、核糖核酸酶、T4 噬菌体溶菌酶以及阿朴脂蛋白等许多种类的蛋白质进行了定点突变改造。例如,Matthews 等人研究了 Gly 和 Pro 对 T4 噬菌体溶菌酶(164 个氨基酸残基)稳定性的影响,结果发现,去除 Gly 或引入 Pro 的突变都使突变体的变性温度提高,从而增强了蛋白质的稳定性^[31]。Duan 等人对一种高密度脂蛋白—阿朴脂蛋白 M(*apoM*)进行了敏感序列搜索和“穿针引线”(threading)比较建模,成功预测了 *apoM* 的三维结构,进而预测并通过定点突变验证了 Asn135 的特殊功能^[32]。Li 等人对细胞色素 P450BM-3 进行的定点突变研究表明, Phe87Gly 的定点突变可以使得 (R)(+)-3-氯苯乙烯环氧化物的光学纯度从原来的 61% 提高到 94.6%^[33]。同样通过细胞色素 P450 酶的理性设计, Jones 等人极大提高了多氯化苯类物质的氧化转化速率和底物耦合效率^[33]。在本课题组的研究中,我们成功应用生物信息学的序列比对分析,对诺卡氏菌脲水合酶 α 亚基的一个氨基酸残基进行了定点突变,突变后的脲水合酶在重组大肠杆菌中的表达活性达到了较高的水平,而突变前的脲水合酶在重组大肠杆菌中不能进行活性表达。

应用生物信息学方法进行生物催化剂的分子设计也越来越受到人们的重视。以 T4 噬菌体溶菌酶为例,Matthews 等人经过仔细的生物信息学分析和最小化计算设计,成功地在原来不存在二硫键、只有 Cys54 和 Cys97 两个半胱氨酸的 T4 噬菌体溶菌酶中引入了 3 对二硫键。除原来的 Cys97 外,其它 5 个半胱氨酸分别由 Ile3、Ile9、Thr21、Thr142 和 Leu164 突变而来。实验证明,所有的突变体在氧化态下均比天然蛋白质更稳定^[31]。

生物信息学分析同样可以用于指导融合蛋白的构建进程。以本课题组 7-ACA 的生物法生产工艺研究为例,我们首先应用生物信息学工具进行了同源基因搜索和序列比对分析,从而在互联网上的蛋白质结构数据库 PDB 中分别获得了与本研究克隆得到的 D-氨基酸氧化酶和 GL-7-ACA 酰化酶具有较高同源性的氧化酶和酰化酶的三维结构数据,进而利用生物信息学软件 Vector NTI Suite 6.0 对这两个酶的空间结构进行了模拟考察,并对双酶融合蛋白的构建策略进行了

设计,最终成功构建了同时具有 D-氨基酸氧化酶和 GL-7-ACA 酰化酶活性的融合蛋白。

由此可见,在生物信息学分析的指导下,对生物催化剂进行目标明确的分子设计与改造,再进行实验验证,是一种卓有成效的策略,在生物催化剂的改造过程中必将收到事半功倍的效果。

3.2.2 生物催化剂的定向进化:定向进化是近年来兴起的进行蛋白质功能改造的有效手段,它在某种程度上就是自然进化过程的仿真。与理性设计不同,定向进化不是一个自上而下的设计方法,而是通过随机突变产生一个巨大的变异基因库,然后通过高通量的筛选方法获得具有性质改善的变异型。通过进一步比较突变前后生物催化剂基因的序列、活性与功能的变化,进行“反向工程”(reverse-engineering)研究,还将有助于从根本上揭示蛋白质的催化机理和折叠机制、蛋白质一级序列与三级结构的关系以及结构与功能的关系等。

应用定向进化技术可以成功地对工业酶的专一性、活性、结构稳定性、热稳定性以及其它性能进行有效的改造^[34]。例如,Miyazaki 等对嗜寒菌的枯草杆菌蛋白酶进行了 3 轮定向进化,得到菌株 3-2G7,它的蛋白酶在 60℃ 的稳定性比野生菌株提高了 500 倍,甚至超过了同源的嗜高温菌的枯草杆菌蛋白酶。Wintrode 等进一步从菌株 3-2G7 出发,再经过 5 轮的定向进化,使得稳定性提高到原野生菌株的 1200 倍,且比同源的嗜高温菌高 20 倍^[35]。

在应用定向进化技术对生物催化剂进行改造的过程中,生物信息学技术的作用也不容忽视。首先,将要进行定向进化改造的目标基因的获得常常需要通过生物信息学工具来获得;其次,在各种定向进化方法中,如 DNA family Shuffling,也常常需要首先使用生物信息学工具对将要进行重排的目的基因进行同源性分析,从而增大实验成功的可能性。

应用生物信息学工具对定向进化后的突变基因库进行比对分析,可以确定一些能够高度容忍各种氨基酸替换的位点。进一步利用生物信息学分析可以发现,在定向进化过程中,引起各种有益的功能改变的基因突变通常会发生在这些位点。由此出发,我们可以通过计算机分析,确定所有的(当然也是有限的)能够高度容忍氨基酸替换的位点,然后面向全基因序列的“饱和突变”(saturation mutagenesis)就成为可能,此时多个突变的相互作用也可以被发现^[36]。Gray 等人就利用这种蛋白质的饱和突变技术,使得从红球菌中获得的盐烷烃脱卤酶的热稳定性极大地改善,抵抗变性的能力大大增强,而酶的活性却没有降低。将突变后的酶固定在矾土上,在 55℃ 的生物反应器中对 1,2,3-三氯丙烷进行酶催化反应,转化为 2,3-二氯丙醇。结果表明,突变后的固定化酶的操作稳定性提高了 25 倍,半衰期也延长了 18 倍^[36]。因此,将生物信息学分析、随机的定向进化和理性的定点突变等方法结合起来进行生物催化剂的功能改造,往往会获得更好的效果。

总之,由于各种公共数据库、创新软件和算法以及大规模、高通量生物学实验平台等生物信息学技术的迅猛发展,

使得从核酸和蛋白质序列信息的汪洋大海中迅速、准确地获取生物催化剂的相关信息,并利用这些信息进行高效的蛋白质分子设计与改性成为可能。因此,为了满足生物催化技术高速发展的迫切需要,以生物信息学技术为工具,以互联网上的公共数据库为基础,对生物催化剂进行分子设计和改造,借以改善生物催化剂的稳定性、专一性、底物和产物耐受性等物理和化学性质,具有非常广阔的发展和应用空间。总体而言,不论是本课题组的研究成果,还是其他科研工作者的实践经验,都使人深深感受到了生物信息学工具在生物催化领域的巨大应用潜力。为此,每一个生物相关领域的科学工作者,都不可不关注生物信息学的发展,不可不掌握生物信息学的基本工具。

4 在生物催化领域应用生物信息学工具的基本策略

针对当前生物催化领域的研究现状,利用生物信息学技术不断加快新生物催化剂的识别与开发进程的研究已经越来越紧迫。概括而言,可以按照如下策略应用生物信息学工具进行工业生物催化研究。

4.1 对于新生物催化剂的发现研究

利用生物信息学工具来发现新生物催化剂需要对因特网上已经存在的大量公共数据库、搜索工具以及生物信息学软件等具有较深入的了解并能熟练应用,以从大量的数据资源中识别新的具有工业应用价值的生物催化剂基因。如果具有很好的数学和计算机知识基础,能够自行开发基因识别及序列比对分析的算法和程序,在该领域的开创性研究将会更有优势。

4.2 对于现有生物催化剂的功能改造

对目前工业生产上已经成功应用的工业酶进行定点突变或定向进化研究,进一步提高其活性、操作稳定性和单位产率,是目前工业生物催化研究的主要方向之一。生物信息学技术在该领域的具体应用主要体现在设计引物调取基因,对基因和蛋白质的序列进行分析,了解新基因的碱基组成与分布、翻译后蛋白质的疏水性、跨膜区等基本性质,确定基因序列的限制性酶切位点、可读框架和蛋白质的结构功能域,并在基因序列同源性分析的基础上进一步进行蛋白质的结构与功能预测,提出蛋白质改造的分子设计方案,研究蛋白质的催化机制,等等,从而大大加快生物催化剂的改造进程。生物信息学技术的上述应用同样适用于定点突变或定向进化后筛选获得的高活性、高稳定性或高底物产物耐受性的优选生物催化剂,通过比较改造前后基因和蛋白质序列的变化以及由该变化所直接引起的蛋白质结构和功能的改变,可以在理论上进一步揭示蛋白质一级序列、蛋白质折叠与功能的对应关系。

无论是新生物催化剂基因的发现,还是现有生物催化剂的改造研究,都需要经常用到下面的几个非常重要的生物信息学数据库(表 2)。

表 2 Web 中一些重要的生物信息学数据库
Table 2 Some important databases in the Web^[10, 11, 17]

Databases	World Wide Web
GenBank Nucleotide and protein sequences	Http://www.ncbi.nlm.nih.gov
EMBL Nucleotide sequences	Http://www.embl-heidelberg.de
PIR Protein sequences	Http://www.nbrf.georgetown.edu/pir/
SWISS-PROT Protein sequences	Http://www.expasy.org
SWISS-PROT mirror site in Peiking University	Http://cn.expasy.org/
The Protein Data Bank (PDB)	Http://www.rcsb.org/pdb/

此外,在国内的生物信息学资源网站(<http://www.bio-soft.net>)中也有大量的生物信息学分析软件和参考资料,对工业生物催化研究也将有所裨益。

5 展望

随着基因组和后基因组时代的到来,生物信息学也将逐渐成熟起来,人们对于生命现象的认识也会越来越充分,大量的数据后面隐藏着的关于生命本质的神秘面纱也必将被揭开。但是现在,生物信息学在国内、外都处于刚刚起步的阶段,生物催化技术的研究也正方兴未艾,因此,在应用生物信息学技术进行生物催化研究的过程中,必将出现这样或那样的问题。最根本的解决方法就是培养一批可以适应这种学科高度交叉的年青一代,兼有数学、生物学、计算机科学和信息学等各方面的知识,通过长期的研究工作,建立起生物信息学的理论架构并奠定生物学实验的坚实基础。尽管这需要很长的一段时间,但是我们坚信,生物信息学最终会给生物催化领域带来巨大的经济和科学回报。

正如诺贝尔奖获得者 W. Gilbert 在 1991 年曾经指出^[8]: “传统生物学解决问题的方式是实验的。现在,基于全部基因都将知晓,并以电子可操作的方式驻留在数据库中,新的生物学研究模式的出发点应是理论的。一个科学家将从理论推测出发,然后再回到实验中去,追踪或验证这些理论假设”。

REFERENCES(参考文献)

[1] Tan TW (谭天伟). Current situation and prospect of biochemical engineering. *Modern Chemical Industry* (现代化工), 2000 , 20 (12): 10 – 14

[2] Ogawa J, Shimizu S. Industrial microbial enzymes : their discovery by screening and use in large-scale production of useful chemicals in Japan. *Current Opinion in Biotechnology* , 2002 , 13 : 367 – 375

[3] Du CY (杜晨宇), Li C (李春), Zhang M (张木) *et al.* Application and development of industry biocatalysis process and biocatalysts. *Journal of Chemical Industry and Engineering* (化工学报), 2003 , 54 (4): 456 – 463

[4] Ministry of Science and Technology , P. R. China (中华人民共和国科学技术部). 2002 Development Report of Chinese Biotechnology. Beijing : Chinese Agriculture Press , 2002

[5] He I (贺林). Decoding Life-Human Genome Project and Post Genome Project. Beijing Science Press , 2001

[6] Goodman N. Biological data becomes computer literate : new advances in bioinformatics. *Current Opinion in Biotechnology* , 2002 , 13 : 68 – 71

[7] Zhong Y (钟扬), Zhang L (张亮). Concise Bioinformatics. Beijing : High Education Press , 2001

[8] Zhao GP (赵国屏). Bioinformatics. Beijing : Science Press , 2002

[9] Zhang CG (张成岗), He FC (贺福初). Methods and Practices of Bioinformatics. Beijing : Science Press , 2002

[10] M Vihinen. Bioinformatics in proteomics. *Biomolecular Engineering* , 2001 , 18 : 241 – 248

[11] Michalovich D, Overington J, Fagan R. Protein sequence analysis *in silico* : application of structurebased bioinformatics to genomic initiatives. *Current Opinion in Pharmacology* , 2002 , 2 : 1 – 7

[12] Bottomley S. Bioinformatics : guide for evaluating bioinformatic software. *Drug Discovery Today* , 1999 , 4 (5): 240 – 243

[13] Augen J. Bioinformatics and information technology : reshaping the drug discovery process. *Drug Discovery Today* , 11 (1): 39 – 40

[14] Xing T. Bioinformatics and its impact on plant science. *Trends in plant science* , 1998 , 3 (11): 450

[15] Foote SJ, Speed T, Handman E. What can bioinformatics do for parasitology research ? *Parasitology Today* , 1998 , 14 (9): 346 – 347

[16] Marchal I, Golfier G, Dugas O, Majed M. Bioinformatics in glycobiology. *Biochimie* , 2003 , 85 : 75 – 81

[17] Desiere F, German B, Watzke H *et al.* Bioinformatics and data knowledge : the new frontiers for nutrition and foods. *Trends in Food Science & Technology* , 2002 , 12 : 215 – 229

[18] Andrade MA, Sandert C. Bioinformatics : from genome data to biological knowledge. *Current Opinion in Biotechnology* , 1997 , 8 : 675 – 683

[19] Ogawa J, Shimizu S. Microbial enzymes : new industrial applications from traditional screening methods. *Focus* , 1999 , 17 : 13 – 20

[20] Luo H, Yu HM, Li Q, Shen ZY. Cloning and Expression of D – amino acid oxidase and Glutaryl – 7-aminocephalosporanic acid acylase genes in *Escherichia coli*. *Proceedings of YABEC 2003 Symposium* , 2003 : OP – I-A4

[21] Shi Y (史悦), Sun XD (孙旭东), Yu HM (于慧敏) *et al.* Bioinformatics analysis , cloning and expression of a *Nocardia* Nitrile hydratase. *Proceedings of the 4th Chinese Enzyme Engineering Symposium* , 2003 , p. 117

[22] Kim TD. PCR primer design : an inquiry-based introduction to bioinformatics on the World Wide Web. *Biochemistry and Molecular Biol-*

- [23] Nevill-Manning CG, Wu TD, Brutlag DL. Highly specific protein sequence motifs for genome analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 1998, **95**: 5865 – 5871
- [24] Chen F, Zhang Q, McDonald T *et al.* Identification of two hERR2-related novel nuclear receptors utilizing bioinformatics and inverse PCR. *Gene*, 1999, **228**: 101 – 109
- [25] Zharkov DO, Grollman AP. Combining structural and bioinformatics methods for the analysis of functionally important residues in DNA glycosylases. *Free Radical Biology & Medicine*, 2002, **32**(12): 1254 – 1263
- [26] Wishart MJ, Taylor GS, Slama JT *et al.* PTEN and myotubularin phosphoinositide phosphatases: bringing bioinformatics to the lab bench. *Current Opinion in Cell Biology* 2001, **13**: 172 – 181
- [27] Marrs B, Delagrave S, Murphy D. Novel approaches for discovering industrial enzymes. *Current Opinion in Microbiology*, 1999, **2**: 241 – 245
- [28] Searls DB. Using bioinformatics in gene and drug discovery. *Drug Discovery Today*. 2000, **5**(4): 135 – 143
- [29] Rohlin L, Oh MK, Liao JC. Microbial pathway engineering for industrial processes: evolution, combinatorial biosynthesis and rational design. *Current Opinion in Microbiology*, 2001, **4**: 330 – 335
- [30] Gustafsson C, Govindarajan S, Minshull J. Putting engineering back into protein engineering: bioinformatic approaches to catalyst design. *Current Opinion in Biotechnology*, 2003, **14**: 366 – 370
- [31] Lai LH(来鲁华). Structure Prediction and Molecule Design of Protein. Beijing: Peiking University Press, 1993
- [32] Duana J, Dahlback B, Villoutreix BO. Proposed lipocalin fold for apolipoprotein M based on bioinformatics and site-directed mutagenesis. *FEBS Letters*, 2001, **499**: 127 – 132
- [33] Cirino PC, Arnold FH. Protein engineering of oxygenases for biocatalysis. *Current Opinion in Chemical Biology*, 2002, **6**: 130 – 135
- [34] Arnold FH, Volkov AA. Directed evolution of biocatalysts. *Current Opinion in Chemical Biology*, 1999, **3**: 54 – 59
- [35] Miyazaki K, Wintrode PL, Grayling RA *et al.* Directed evolution study of temperature adaptation in a Psychrophilic Enzyme. *Journal of Molecular Biology*, 2000, **297**(4): 1015 – 1026
- [36] Farinas ET, Bulter T, Arnold FH. Directed enzyme evolution. *Current Opinion in Biotechnology*, 2001, **12**: 545 – 551

Application of Bioinformatics in Researches of Industrial Biocatalysis

YU Hui-Min* LUO Hui SHI Yue SUN Xu-Dong SHEN Zhong-Yao

(Department of Chemical Engineering, Institute of Biochemical Engineering, Tsinghua University, Beijing 100084, China)

Abstract Industrial biocatalysis is currently attracting much attention to rebuild or substitute traditional producing process of chemicals and drugs. One of key focuses in industrial biocatalysis is biocatalyst, which is usually one kind of microbial enzyme. In the recent, new technologies of bioinformatics have played and will continue to play more and more significant roles in researches of industrial biocatalysis in response to the waves of genomic revolution. One of the key applications of bioinformatics in biocatalysis is the discovery and identification of the new biocatalyst through advanced DNA and protein sequence search, comparison and analyses in Internet database using different algorithm and software. The unknown genes of microbial enzymes can also be simply harvested by primer design on the basis of bioinformatics analyses. The other key applications of bioinformatics in biocatalysis are the modification and improvement of existing industrial biocatalyst. In this aspect, bioinformatics is of great importance in both rational design and directed evolution of microbial enzymes. Based on the successful prediction of tertiary structures of enzymes using the tool of bioinformatics, the undermentioned experiments, *i. e.* site-directed mutagenesis, fusion protein construction, DNA family shuffling and saturation mutagenesis, etc, are usually of very high efficiency. On all accounts, bioinformatics will be an essential tool for either biologist or biological engineer in the future researches of industrial biocatalysis, due to its significant function in guiding and quickening the step of discovery and/or improvement of novel biocatalysts.

Key words industrial biocatalysis, biocatalyst, bioinformatics

Received: 09-29-2003

This work was supported by both the National Natural Science Foundation of China(No. 20206014) and the Foundation for the Author of National Excellent Doctoral Dissertation of PR China.

* Corresponding author. Tel: 86-10-62788568; Fax: 86-10-62770304; E-mail: yuhm@tsinghua.edu.cn

© 中国科学院微生物研究所期刊联合编辑部 <http://journals.im.ac.cn>