

文章编号: 1004-0374(2011)09-0860-09

生物元件的挖掘、改造与标准化

王钱福, 严 兴, 魏 维, 张 磊, 钱昌丽, 周志华*

(中国科学院上海生命科学研究院植物生理生态研究所, 中科院合成生物学重点实验室, 上海 200032)

摘 要: 生物元件是合成生物学中的三大基本要素之一, 是合成生物学的基石。现阶段, 生物元件的挖掘、鉴定和改造仍然是合成生物学领域的重要研究方向之一。合成生物学与基因工程和代谢工程最显著的差别在于能够将大量的生物元件进行快速、随意的组装, 而实现这一目标的前提是将生物元件标准化。目前, 已经有大量基因组被解析, 通过这些基因组数据库的注释与功能验证, 并借助于各种生物信息学软件预测启动子、终止子、操纵子、转录因子和转录因子结合位点、核糖体结合位点以及蛋白质编码区等部件, 为合成生物学提供丰富的生物元件信息资源。随着元基因组技术的兴起, 大量未培养微生物中的基因和基因簇信息被解析, 使得我们可以从占自然界中实际存在微生物总数 99% 的未知微生物中挖掘更多的生物元件。另外, 生物元件可以从自然界分离出来, 也可以对天然生物元件进行修饰、重组和改造后得到新的元件。酵母是异源蛋白表达的通用宿主和生物基产品生产的细胞工厂, 但其本身可用的启动子非常有限, 近年来各国学者在酵母启动子改造和文库构建方面做了很多工作, 该文也将概述酵母启动子改造和在合成生物学研究领域中的应用方面的研究进展。

关键词: 合成生物学; 生物元件; 基因组; 元基因组; 启动子; 文库构建

中图分类号: O621.33; Q812 **文献标志码:** A

Screening, modification and standardization of biological parts for synthesis biology

WANG Qian-Fu, YAN Xing, WEI Wei, ZHANG Lei, QIAN Chang-Li, ZHOU Zhi-Hua*

(Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China)

Abstract: Biological part is the foundation for synthetic biology. Seeking, defining and modifying of new biological parts are attracting common attentions. The most significant difference between synthetic biology and genetic engineering or metabolic engineering must be that the interchangeable biological parts are used and assembled quickly in synthetic biology, which had better to be standardized previously. With more and more genomes of different organisms being sequenced, parts such as promoters, terminators, operons, transcriptional factors and transcriptional factor binding sites, ribosome binding sites and protein encoding regions could be annotated by different bioinformatics tools. Owing to the development of metagenomic technology, vast information about the genes and gene clusters of the remaining 99% uncultured microorganisms in nature were explored and a lot of new biological parts could be mined from natural resources directly. Of course, the biological part originated from nature need be modified, reconstructed and improved before it shows highly compatible in a synthetic module or biological system. *Saccharomyces cerevisiae* was widely used as the general host for heterologous proteins expression as well as cell factory for producing biobased products. However, only a few promoters could be applied in *S. cerevisiae*.

收稿日期: 2011-05-23; 修回日期: 2011-06-08

基金项目: 中科院生命科学基础前沿研究所专项(KSCX2-EW-J-12)

*通信作者: E-mail: zhzhou01@sibs.ac.cn

Many efforts to modify promoters and establish promoter mutant libraries for *S. cerevisiae* have been tried recently. A brief introduction of studies on promoter evolution of *S. cerevisiae* and its application in synthetic biological field are also included in this manuscript.

Key words: synthetic biology; biological parts; genome; metagenome; promoter; library construction

合成生物学是按照一定的规律和已有的信息设计和构建新的生物元件、装置和系统, 或者重新设计已有的天然生物系统 (<http://syntheticbiology.org/>)。合成生物学的核心在于设计、改造、重建或制造生物元件、生物反应系统、代谢途径与过程, 乃至创造具有生命活动能力的细胞和生物个体, 为生命起源、生物进化、复杂疾病、干细胞等生物学难题的研究开拓新的蹊径, 为解决人类发展在环境、资源、能源等方面面临的若干重大挑战提供新技术与方法^[1]。合成生物学旨在将工程原理应用到生物系统来改变生物技术^[2]。在短短不到十年的时间里, 合成生物学领域已经产生了很多技术应用, 为药品生产^[3-4]、生物基产品的生产^[5]和基因治疗^[6-7]提供了新的途径。除此之外, 合成生物学在替代能源生产方面也展现出巨大的应用潜力^[8]。

生物元件、装置和系统是合成生物学的三大基本要素。生物元件是指遗传系统中最简单、最基本的生物积块 (BioBrick), 是具有特定功能的氨基酸或者核苷酸序列, 可以在更大规模的设计中与其他元件进一步组合成具有特定生物学功能的生物学装置 (Device)^[9-10]。早在 2003 年, 美国麻省理工大学相关合成生物学实验室就成立了标准生物元件登记库 (Registry of Standard Biological Parts)(<http://partsregistry.org/>), 专门收集各种满足标准化条件的生物元件。全世界的科学家都可以提交自己设计的模块供其他人获取, 以便设计更加复杂的系统。目前标准生物元件登记库有 15 000 多个生物元件 BioBrick(<http://partsregistry.org/cgi/partstodb/Statistics.cgi>), 既包括启动子、转录单元、质粒骨架、接合转移元件、转座子、蛋白质编码区等 DNA 序列, 也包括核糖体绑定点和终止子等 RNA 序列以及蛋白质结构域。标准生物元件登记库中, 每一个生物学元件, 例如启动子、核糖体结合位点 (RBS) 都以载体形式保存^[11]。

虽然有越来越多的生物元件被发现, 但很多生物元件都没有详细定性, 标准生物元件库中的元件大部分都是 iGEM 参赛者提供的, 其中很多生物元件都没有经过系统的定性, 也无法保证品质。生物

元件的性能会随不同的底盘细胞类型, 或者不同的实验室条件而改变, 而不是总表现相同的功能, 发挥功能的时间也不总是相同的^[12]。2010 年 12 月 12 日在夏威夷举行的生物技术工业组织环太平洋地区会议上, 合成生物学工程研究中心的创建人 Drew Endy 和 Vivek Mutalik 发起了国际发展生物技术开放基金, 由美国国家自然科学基金 (NSF)、劳伦斯伯克利国家实验室 (LBNL) 和生物砖基金 (BBF) 联合支持, 用于开发新的生物元件并对已经存在的元件进行详细定性描述。Drew Endy、Vivek Mutalik 以及生物砖基金的执行主席 Holly Million 强调该开放基金的目的是开发上千个可免费使用的标准 DNA 元件, 推动合成生物学的研究进程, 缩短研发周期并降低研发成本。

为了使合成物生物学像其他工程学设计那样, 可以在电脑帮助的模拟设计下构建新的生物系统, 英国学者 Cooling 等^[13] 创建了一个标准虚拟生物元件库 (SVPs), 为合成生物学提供模块化建模元件库。最近华盛顿大学的研究者又创建了一个标准生物元件知识库 (SBPkb), 用于查询和检索生物元件用于合成生物学研究和应用^[14]。SBPkb 将标准生物元件知识库中所有元件的信息转换成可以利用合成生物学元件语义框架进行运算的信息。该框架称作为合成生物学公开语言语义 (SBOL-semantic), 是合成生物学公开语言的一部分, 用于描述通用的合成生物学实体, 其目的是提高所有生物元件的分散和交换。

与自然界中巨大的生物元件资源相比, MIT 标准生物元件登记库中的现有库存仅仅是沧海一粟。现阶段如何鉴定与建立更多的可通用型生物元件依然是合成生物学的重要任务^[15-16]。生物元件可以从自然界分离出来, 也可以对天然生物元件进行修饰、重组和改造后得到新的元件。

1 基因组信息中挖掘生物元件

生物元件主要来源于自然界, 因此从自然界中分离是获得生物元件的主要途径之一。在大规模基因组测序前, 从自然界中分离生物元件主要是运用

传统的分子克隆方法或基于保守序列的 PCR 方法。随着越来越多的基因组信息得到解析, 可以通过生物信息学预测直接从基因组中得到各种生物元件。目前已有超过 1 000 个微生物的全基因组得到了分析, 对这些基因组数据库的注释与功能验证将为合成生物学提供丰富的生物元件信息资源。

通过对基因组中的功能蛋白、转录和翻译特征序列分析, 可以得到丰富的启动子、核糖体结合位点、蛋白质编码序列以及终止子等生物元件资源。利用基因组信息, 结合各种生物信息学软件和在线网站可以预测各种生物元件, 并对元件信息进行详细分析。利用 FPROM 可从人类基因组预测启动子序列, TSSG 和 TSSW 可预测人的 PolII 启动子区域和转录开始位点^[17-19]。TSSP 可预测植物启动子序列 (<http://linux1.softberry.com/berry.phtml?topic=tssp&group=programs&subgroup=promoter>)。PromH(G)、PromH(W) 和 Promoter2.0 软件可从真核基因组预测启动子序列^[20-21]。SCOPE 可以预测部分原核生物和真核生物的调控 DNA 基序, 该方法利用 BEAM、PRISM 和 SPACER 三种程序分别预测非退化基序、退化基序和两者都有的基序^[22-26]。BPROM(<http://www.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb>) 可预测原核生物的启动子序列。同样多个软件可从基因组信息预测终止子序列, 如通过 TransTermHP 可预测原核生物的终止子^[27-28]。最新建立的 WebGesTer DB 是目前已知的最大的转录终止子资料库, 其中包括来源于 1 060 个细菌基因组序列和 798 个质粒序列信息的百万个终止子^[29]。RibEx 可预测终止子和一些细菌中保守的转录调控序列^[30]。RNA 序列家族信息可通过 Rfam 信息库^[31], tRNAs 和 snoRNAs 信息可通过 tRNAscan-SE、snoscan 和 snoGPS 进行查找^[32]。BDB 是一个预测序列特异性 DNA 结合转录因子的数据库, 可从已经公布的 929 个完整基因组序列预测其中的转录因子^[33]。TFSEARCH 可预测真核生物转录因子结合位点 (<http://www.cbrc.jp/research/db/TFSEARCH.html>), Tfsitescan(<http://www.ifti.org/cgi-bin/ifti/Tfsitescan.pl>) 和 TESS(<http://www.cbil.upenn.edu/cgi-bin/tess/tess>) 可预测原核和真核生物转录因子结合位点。rVISTA 2.0 是一个比较分析非编码序列调控潜力的工具, 它整合了转录因子预测、序列比对和聚类分析, 用于确认进化上保守和出现在基因组序列中特定分子构型中的非编码 DNA 区域序列^[34]。利用 operonHMM 方法可在没有实验数据的

情况下预测细菌基因组中的操纵子模型^[35]。可利用神经网络预测大肠杆菌和其他生物的核糖体结合位点^[36-37]。蛋白质结构域可通过多种方法进行预测, 可从已有的基因组资源挖掘更多的具有特定功能的新结构域。Pfam 是一个庞大的蛋白质家族和结构域数据库^[38], 最新版本 (25.0) 中收集了 12 273 个蛋白家族信息。Smart 是一个确认和注释蛋白质结构域的重要在线工具^[39], 最新的版本中已经收集了 886 个蛋白质结构域信息。

2 利用元基因组技术挖掘生物元件

随着微生物分子生态学技术的发展, 使人们认识到自然界中的大部分微生物是不可培养的, 可培养的微生物仅占自然界中实际存在微生物的 1%, 这些未培养微生物中必然蕴藏着巨大的生物元件资源。近年来发展起来的元基因组研究技术, 可以让我们走出纯培养微生物的局限, 从自然界, 特别是极端环境中获得更丰富的生物资源, 包括合成生物学所需要的各种生物元件。利用元基因组技术, 一方面通过测序可以得到未培养生物中的基因信息, 获得许多未知的生物元件信息; 另外一方面可以通过功能筛选, 得到许多具有各种用途的功能基因。

21 世纪以来, 元基因组学的研究越来越受到关注, 不断有新的报道。应用 DNA 序列导向和功能导向的元基因组筛选方法, 可以筛选并鉴定具有某种功能的新基因或基因簇, 是获得功能蛋白生物元件的重要途径。如对河底沉积物的富集培养物元基因组测序后, 从中找到了以 NO₂ 为电子受体来氧化甲烷的新代谢途径^[40]。目前各国科学家利用元基因组技术, 从未培养生物中分离到许多有重要工业应用价值的酶基因, 如酯酶基因^[41-43]、脂肪酶基因^[44-45]、蛋白酶基因^[41,46] 以及从冰川冰的元基因组文库中筛选到 DNA 聚合酶 I 基因^[47]。同时, 随着单细胞全基因组扩增技术和流式细胞仪分选技术的日益成熟, 已经可以利用单细胞扩增和分析未培养微生物的基因组^[48-49]。从这些未培养微生物的元基因组, 我们可以挖掘到大量新的生物元件。

近年来随着化石原料的不断枯竭以及工业原料与粮食争地矛盾的凸显, 新一代工业生物技术将逐步以木质纤维素类生物质代替淀粉、糖类等粮食类生物质和化石原料。因此, 与木质纤维素利用和炼制密切相关的纤维素酶受到普遍关注^[50]。而纤维素酶在生产生物燃料应用方面的研究也是各国生物技术领域研究的热点, 其对关键技术的解决更是全球

生物技术领域知识产权争夺的重要制高点。美国能源部 (DOE) 为了寻求解决环境修复、再生能源及气候变化等问题的途径和策略, 率先启动了“Genome To Life”计划, 并完成了对极端酸性条件下生物膜中微生物^[51]、海洋微生物^[52]和白蚁肠道微生物^[53]的元基因组测序工作。目前各国学者已经从厌氧高温木头堆^[54]、盐湖和火山口湖床沉淀物^[55]、土壤^[56,57]、牛瘤胃^[58]、兔子盲肠^[59]等各种生境通过元基因组测序和筛选得到许多纤维素基因。尤其牛瘤胃作为一个高效的纤维素酶降解系统, 对该系统中微生物群落的元基因组解析发现大量的糖基水解酶基因, 大大增加了 CAZy 数据库中纤维素酶基因的数量^[60-61]。我们中科院合成生物学重点实验室也以具有高效降解纤维素的厌氧沼液系统和白蚁肠道为基质, 建立了元基因组文库, 从中分离到很多与木质纤维素降解相关的新酶基因^[62]。

3 生物元件的改造及其在合成生物学中的应用

生物元件可以从自然界分离出来, 也可以被修饰、重组和改造后得到新的生物元件。启动子是专一地与 RNA 聚合酶结合并决定转录开始位置的元件, 控制基因转录的起始时间和表达程度。启动子是合成生物学中一个非常重要的调控元件, 各种遗传线路^[63]、基因开关^[64]、基因振荡^[65]和细胞群体系统脉冲发生器^[66]等功能的发挥大都是需要通过启动子的调控来实现的。

最近, 台湾学者创建了一个集成的酿酒酵母启动子特性知识库, 称作酵母启动子图谱 (YPA)。YPA 整合了启动子的九种特性, 包括启动子序列、转录起始位点、3' 和 5' 非编码区、TATA 框、转录因子结合位点、核小体占用率、DNA 弯曲性、转录因子结合、转录因子敲除表达和转录因子之间相互作用。酵母启动子图谱设计利用联合方式呈现其资料, 即仅当对启动子的各种特性进行综合考虑时才能揭示许多重要的观察值。例如, 刚性 DNA 能阻止被核小体包裹, 从而使得转录因子更易接近刚性 DNA 区域的转录因子结合位点。结合核小体占用率、DNA 弯曲性、转录因子结合、转录因子敲除表达和转录因子结合位点等资料, 可以帮助我们确定其中真正发挥功能的转录因子结合位点。在酵母启动子图谱中, 各种启动子特性被整合进一个集中的、有组织的平台中。研究者可以方便地观察其感兴趣的转录因子结合位点是否被核小体占有或者位于刚性 DNA 区域, 也可以知道下游基因的表达是否响应

相关转录因子的敲除。跟其他酵母启动子资料库相比, 酵母启动子图谱不仅仅集合了转录因子结合位点, 而且结合了许多其他启动子特性, 这些信息将对生物学家研究基因的转录调控有帮助^[67]。

对基因功能的研究往往需要通过改变基因的表达水平来考察不同表达水平下产生的结果。原则上, 目的基因可以通过特定调控实现连续不同的表达水平。传统的方法中, 对基因功能的研究是在两个极端的状态: 基因敲除和高表达。但是, 基因功能在代谢优化和控制分析方面的研究需要基因表达呈现一种连续的、增量微小的表达水平, 这就需要基因在一系列连续的强弱不同的启动子调控下表达^[68]。而作为通用蛋白表达宿主和合成生物学中的细胞工厂的酵母, 其本身可用的启动子极其有限, 因此需要对现有的启动子进行改造和人工合成新的启动子。随着蛋白定向改造技术的日益成熟, 可以将蛋白定向改造技术引入到酵母启动子的改造中。通过定向改造技术, 建立包括各种不同强度启动子的文库, 是实现基因精确调控的有利遗传工具。目前, 已经对毕赤酵母和酿酒酵母中的启动子进行了改造, 通过构建启动子文库, 得到了一系列不同强度的启动子元件。

3.1 毕赤酵母中启动子改造和文库构建

醇氧化酶启动子 (*AOXI*) 是毕赤酵母中最强的诱导型启动子, Hartner 通过敲除和复制启动子 *AOXI* 中可能的转录因子结合位点建立了由 46 个 *AOXI* 突变子组成的启动子文库, 得到了一系列强度不同的启动子, 其活性范围为野生型启动子活性的 6% 到 160%。另外, 研究表明至少有 12 个顺式调控元件与 *AOXI* 启动的高水平表达有关, Hartner 等^[69]基于敲除分析, 通过在启动子基本元件上结合顺式调控元件构建了由新的、短的人工突变启动子组成的文库。利用新合成的 *AOXI* 突变启动子, 不仅提高了异源蛋白的产量和质量, 而且将作为一个基本的工具箱, 用于在代谢工程中精细调控基因表达和在合成生物学中顺序诱导蛋白表达。

Qin 等^[70]利用易错 PCR 技术, 对组成型启动子 *GAP* 进行突变, 利用荧光蛋白作为标记, 创建了一个由 33 个突变子组成的, 具有功能的启动子文库。相对于 *GAP* 来说, 33 个突变子的活性范围是野生型的 0.6%~19.6 倍, 并呈现由高到底的线性化分布。为了证实该文库更广泛的适用性, 利用 β -半乳糖苷酶和蛋氨酸腺苷转移酶 (MAT) 基因作为报告基因, 研究了启动子强度对基因在转录水平和蛋

白表达水平上的影响。通过研究异源蛋氨酸腺苷转移酶基因表达水平对细胞生长和 S-腺苷甲硫氨酸 (SAM) 产生的影响来考察启动子文库的实用性。依靠文库中启动子强度的宽泛性, 通过确定最适的蛋氨酸腺苷转移酶活性来获得最大的 S-腺苷甲硫氨酸产量。可以利用启动子文库实现基因表达的精确调控以及定量评估基因表达水平与生理学参数的相关性。

3.2 酿酒酵母中启动子改造和文库构建

过去研究特定基因表达对表型和代谢流的影响, 采用的传统方法是通过对野生型基因、敲除和高表达三种条件下的变化情况来研究的。但是, 基因的高表达或完全敲除往往不能完全反应某种表型或合成途径的真实信息, 期望的表型通常需要基因的适度表达而不是高拷贝表达, 基于这种方法的基因功能研究是不完整的^[71-74]。通过调节目的基因在一个宽的范围内的连续不同的表达水平, 才能实现酿酒酵母中基因表达的精确调控。很多方法可以实现目的基因的梯度表达, 如利用各种不同强度的天然启动子和拷贝数不同的质粒调控基因表达水平^[72,75]以及通过改变诱导物的浓度来调控诱导型启动子调控下基因的表达水平等^[76], 但这些方法都有其局限性。近年来, 利用基于定向进化和基因改组的蛋白工程技术^[77-78], 人工构建了包括不同强度启动子的文库, 从而实现基因在一个宽的范围内的调控。

Alper 等^[79]首先构建了噬菌体 P_L - λ 启动的突变文库, 在保证在单细胞水平平均一性的情况下, 以 GFP 为报告基因研究了启动子强度对荧光产生水平的影响, 表明启动子强度的增加和荧光强度的增强之间呈线性关系; 进一步利用 RT-PCR 从转录水平考察了启动子强度对 GFP 基因转录的影响, 表明启动子在转录水平调控 GFP 基因的表达; 另外, 以 cat 为报告基因, 根据对氯霉素的抗性研究表明, 随着启动子强度的增加, 菌株对氯霉素的抗性也增强。该研究从三个角度评估了启动子对基因表达的影响, 并显示了启动子文库宽的调控范围和 P_L - λ 调控不依赖于下游基因的组成型特性。将这些启动子整合到染色体中, 可以进行遗传控制的精确定量评估。利用该启动子文库评估磷酸烯醇式丙酮酸羧化酶表达水平对生物量的影响, 表明最大生物量的产生需要最适的磷酸烯醇式丙酮酸羧化酶表达水平, 随着启动子强度的增加, 生物量达到一个最高值, 但当强度继续增加时, 会产生负面影响, 生物量会

降低; 同样, 以脱氧木酮糖磷酸合成酶为目的基因, 研究其表达水平对番茄红素产生的影响时也表明最大量番茄红素的产生需要最适的脱氧木酮糖磷酸合成酶表达水平。表明利用文库中不同强度的启动子, 确定基因的最佳表达水平, 从而得到最期望的表型。但是, 在工程化的产番茄红素菌株中, 番茄红素产量与脱氧木酮糖磷酸合成酶表达水平成线性增长关系, 脱氧木酮糖磷酸合成酶是其限速步骤, 表明启动子调控的基因型-表型关系跟菌株的遗传背景相关^[79]。为了拓宽利用该方法在其他研究领域的应用, 他们也酿酒酵母的 *TEF1* 启动子利用易错 PCR 方法进行改造, 建立了突变启动子库, 得到了一系列强度不同的突变启动子。在此基础上, Nevoigt 等^[74]进一步对 *TEF1* 突变库的 11 个启动子特性进行了研究。研究表明, 突变启动子的活性是野生型活性的 8%~120%; 11 个突变启动子控制的报告基因的表达水平差异与碳源的利用无关, RT-PCR 证实这些差异是由转录水平的变化引起的, 而这种变化是由启动子强度的变化产生的。除了 CEN/ARS 型质粒的启动子库外, 他们也创建了启动子置换盒库, 能使其下游的酵母基因整合到基因组中, 通过基因表达水平观察其详细的基因型-表型关系。为了阐述这个方法的适用性, 分别用 *TEF1* 突变库中五个不同强度的启动子替代了酿酒酵母的 *GPD1* 启动子, 用于分析甘油 3 磷酸脱氢酶活性对甘油产量和生物量的影响。表明在甘油 3 磷酸脱氢酶的活性是野生型的两倍的范围内, 甘油 3 磷酸脱氢酶的表达水平是酿酒酵母中甘油产生的限速步骤, 在此范围内甘油的产量随着甘油 3 磷酸脱氢酶活性的增加而增加; 而且表明, 在一个适度范围内高表达甘油 3 磷酸脱氢酶将不会降低生物量。

除了基因的组成型表达以外, 建立一个易于控制、可进行条件诱导或抑制的基因表达系统是基础研究和工业规模生物技术应用必不可少的工具。然而, 诱导剂的毒性、昂贵的价格和诱导剂介导的多效性影响等特点限制了这些诱导型启动子的应用。通过启动子的改造, 得到具有特定调节特性的适配化启动子将使进行基因表达的复杂调控成为可能。德国学者应用随机突变和多级流式细胞筛选, 分离酿酒酵母氧响应性启动子 *DAN1* 的突变子。分离到两个与野生型启动子相比, 在非严格厌氧条件下可诱导的突变子, 这可通过在细胞生长过程中氧气的消耗而实现。这使得酵母发酵中基因表达的诱导可仅仅通过细胞生长过程中氧气的消耗而完成。

而且, 无论是在严格厌氧还是微厌氧环境下, 工程化的启动子的最大表达量远远高于未突变 DAN1 启动子的表达量^[80]。

利用模块化的组件构建具有可预测行为的工程化基因网络是合成生物学的主要目标之一^[81-82]。然而, 由于缺乏合适的组件和构建网络通常需要大量反复测试工作, 使得构建预期功能的基因网络受到阻碍。美国波士顿大学学者发表了一种利用启动子文库结合杂交模型用于指导预测酿酒酵母基因网络构建的方法。利用芯片模型耦合了由多元化组件组成的文库, 这些多元化组件是通过随机非必需序列合成的。他们在酿酒酵母中合成了调节启动子文库, 并利用该文库构建具有各种预期输入输出特性的前馈环网络, 通过该研究证实了上述方法的有效性。为了拓宽该方法的应用范围, 他们合成了一个具有计时器功能的基因网络, 并可通过组件的选择进行更改。运用该基因网络控制酵母沉淀的时间, 表明该设计的即插即用性质, 可方便地应用到生物技术产业中^[83]。

启动子文库不仅应用在基因的最优化表达中, 而且在基因网络的构建中发挥重大作用。利用蛋白质改造方法构建启动子文库是一种非常重要的遗传改造工具, 已经广泛应用于代谢工程、功能基因组学、合成生物学和各种生物技术研究。

3.3 生物元件的标准化与组装

合成生物学对生物元件的组装提出了很高的要求, 需要建立能够对大量生物元件进行组装的方便、快捷的方法, 最终目标是实现生物元件组装的自动化。实现这一目标的前提之一是将生物元件标准化。标准化元件在机械、电子和计算机工程等工业领域早有广泛的应用。由于标准化元件的应用, 使得不同功能的元器件和不同公司的产品能够方便地集成, 从而使工业界能够生产出复杂而可靠的产品^[14]。与此类似, 标准化生物元件的使用可以让不同实验室构建的标准生物元件都按照相同的规则进行组装, 从而可以避免大量的重复劳动, 从而缩短合成复杂的生物装置或者生命系统所需的时间。他们将这些标准化元件以及由它们相互连接所组成的标准生物模块称为“生物积块”(BioBrick)。

美国麻省理工大学标准生物元件登记库(Registry of Standard Biological Parts)已在生物元件的标准化方面开展了许多基础研究, 包括建立新载体体系和对生物元件进行标准化处理, 在生物元件两端接上统一的“接口”, 这些标准化元件以及由

它们相互连接所组成的标准生物模块被称为“生物积块”(BioBrick)。MIT 的生物元件(BioBrick)的组装方法包括标准组装和分层组装(Layered assembly)两种方法^[84]。标准组装方法主要利用限制性内切酶酶切和连接: 每个标准生物元件的两端都含有 2 个规定的酶切位点, 通常是 *EcoRI* 和 *XbaI* 酶切位点作为序列的前缀, *SpeI* 和 *PstI* 酶切位点作为序列的后缀。经过这些同尾酶酶切后产生的黏性末端, 可以让两个不同的标准元件进行连接, 连接后产生的新元件的两端也具有标准生物元件所要求具备的酶切位点。分层组装主要利用 gateway 技术产生带有生物元件的组装载体, 再利用 2ab 技术将两个带有不同生物元件的组装载体进行组装([http://2008.igem.org/Team:UC_Berkeley/Layered Assembly](http://2008.igem.org/Team:UC_Berkeley/LayeredAssembly))。以上两种组装方法产生的拼接产物本身也符合标准生物元件的要求, 也属于标准生物元件, 所以可以按照相同的方法与其他的标准元件再次进行连接, 然后分阶段, 逐次完成对目标序列的拼接。

以上两种组装方法, 因为每次都只组装两种元件, 因此称为二元组合法(binary assembly)。除了上述的两种方法之外, 二元组合法还包括其他一些方法^[85-87], 但是并不直接利用 BioBrick 进行组装。除了二元组合法, 还有多元组合法, 目前主要是利用酵母菌同源重组的功能, 将多个生物元件进行一次体内拼装, 合成含有多个基因的代谢途径, 例如 DNA assembler 等方法^[88]。Gibson 等^[89]利用类似的方法, 在酵母菌内已经成功合成了生殖支原体的基因组。除了体内重组外, SLIC 是一种不依赖于序列和连接反应的克隆方法, 可利用同源重组通过一个反应实现多个 DNA 片段在体外的有效重组^[90]。

随着合成生物学的发展, 具有相似功能的生物元件会越来越多, 那么如何对这些元件进行挑选, 以达到系统或者装置的设计目标, 这是一个非常重要的问题。一般方法是利用这些元件分别合成目标序列, 然后来比较它们的功能。但是, 在目前的条件下, 元件的数量越来越多, 合成的步骤也越来越复杂, 往往不可能针对所有候选元件进行实验。因此, 利用计算机辅助设计(computer-aided design, CAD)来帮助生物元件的选择, 这已成为合成生物学一个重要的研究方向。Ellis 等^[83]就是通过构建数学模型, 然后根据计算结果来选择适当强度的启动子, 从而合成了所需的基因回路, 达到了预期的实验目标。目前, 已经有一些计算机软件来帮助建

立数学模型, 然后根据生物元件的生物学数据(动力学参数和启动子强度等)来预测合成的系统是否符合设计的目标, 从而帮助生物元件的选择, 计算机辅助设计大大加快了合成生物学的研究速度^[91-96]。

随着合成生物学的发展, 需要合成的目标序列包括的元件数量越来越多, 序列也越来越长, 因此需要对目标序列的合成途径进行优化, 以减少合成的时间、降低合成的成本。Densmore等^[84]开发的算法可以高效地利用元件库中已经存在的元件, 以及目标序列中间的重复序列和多个目标序列之间存在的共有序列, 从而优化单个目标序列或者多个目标序列(之间存在共有序列)的合成途径, 大幅减少合成过程中所需要的阶段和步骤, 从而加快合成的速度和降低合成的成本。

[参 考 文 献]

- [1] Endy D. Foundations for engineering biology. *Nature*, 2005, 438: 449-53
- [2] Andrianantoandro E, Basu S, Karig DK, et al. Synthetic biology: new engineering rules for an emerging discipline. *Mol Syst Biol*, 2006, 2: 2006.0028
- [3] Withers ST, Keasling JD. Biosynthesis and engineering of isoprenoid small molecules. *Appl Microbiol Biotechnol*, 2007, 73: 980-90
- [4] Weber W, Schoenmakers R, Keller B, et al. A synthetic mammalian gene circuit reveals antituberculosis compounds. *Proc Natl Acad Sci USA*, 2008, 105: 9994-8
- [5] Basu S, Gerchman Y, Collins CH, et al. A synthetic multicellular system for programmed pattern formation. *Nature*, 2005, 434: 1130-4
- [6] Lu TK, Collins JJ. Dispersing biofilms with engineered enzymatic bacteriophage. *Proc Natl Acad Sci USA*, 2007, 104: 11197-202
- [7] Lu TK, Collins JJ. Engineered bacteriophage targeting gene networks as adjuvants for antibiotic therapy. *Proc Natl Acad Sci USA*, 2009, 106: 4629-34
- [8] Lee SK, Chou H, Ham TS, et al. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Curr Opin Biotechnol*, 2008, 19: 556-63
- [9] Benner SA, Sismour AM. Synthetic biology. *Nat Rev Genet*, 2005, 6 (7): 533-43
- [10] Purnick PE, Weiss R. The second wave of synthetic biology: from modules to systems. *Nat Rev Mol Cell Biol*, 2009, 10 (6): 410-22
- [11] Shetty RP, Endy D, Knight TF. Engineering BioBrick vectors from BioBrick parts. *J Biol Eng*, 2008, 2: 5
- [12] Kwok R. Five hard truths for synthetic biology. *Nature*, 2010, 463(7279): 288-90
- [13] Cooling MT, Rouilly V, Misirli G, et al. Standard virtual biological parts: a repository of modular modeling components for synthetic biology. *Bioinformatics*, 2010, 26(7):925-31
- [14] Michal G, Cesar R, Deepak C, et al. Standard biological parts knowledgebase. *PLoS One*, 2011, 6(2): e17005
- [15] David S, Michael BE. Reconstruction of genetic circuits. *Nature*, 2005, 438: 443-8
- [16] Canton B, Labno A, Endy D. Refinement and standardization of synthetic biological parts and devices. *Nat Biotechnol*, 2008, 26 (7):787-93
- [17] Ghosh D. Status of the transcription factors database (TFD). *Nucleic Acids Res*, 1993, 21(13): 3117-8
- [18] Wingender E. Recognition of regulatory regions in genomic sequences. *J Biotechnol*, 1994, 35: 273-80
- [19] Hannenhalli S, Levy S. Promoter prediction in the human genome. *Bioinformatics*, 2001, 17: S90-6
- [20] Solovyev VV, Shahmuradov IA. Prom H: Promoters identification using orthologous genomic sequences. *Nucleic Acids Res*, 2003, 31(13): 3540-5
- [21] Knudsen S. Promoter 2.0: for the recognition of Pol II promoter sequences. *Bioinformatics*, 1999, 15(5): 356-61
- [22] Carlson JM, Chakravarty A, DeZiel CE, et al. SCOPE: A web server for practical *de novo* motif discovery. *Nucleic Acids Res*, 2007, 35: W259-64
- [23] Chakravarty A, Carlson JM, Khetani RS, et al. A novel ensemble learning method for *de novo* computational identification of DNA binding sites. *BMC Bioinformatics*, 2007, 8: 249
- [24] Chakravarty A, Carlson JM, Khetani RS, et al. SPACER: robust identification of cis-regulatory elements with non-contiguous critical residues. *Bioinformatics*, 2007, 23(8): 1029-31
- [25] Carlson JM, Chakravarty A, Khetani RS, et al. Bounded search for *de novo* identification of degenerate cis-regulatory elements. *BMC Bioinformatics*, 2006, 7: 254-85
- [26] Carlson JM, Chakravarty A, Gross RH. BEAM: A beam search algorithm for the identification of cis-regulatory elements in groups of genes. *J Comput Biol*, 2006, 13: 686-701
- [27] Ermolaeva MD, Khalak HG, White O, et al. Prediction of transcription terminators in bacterial genomes. *J Mol Biol*, 2000, 301(1): 27-33
- [28] Kingsford CL, Ayanbule K, Salzberg SL. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol*, 2007, 8(2): R22
- [29] Mitra A, Kesarwani AK, Pal D, et al. WebGeSTer DB--a transcription terminator database. *Nucleic Acids Res*, 2011, 39: D129-35
- [30] Abreu-Goodger C, Merino E. RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements. *Nucleic Acids Res*, 2005, 33: W690-2
- [31] Gardner PP, Daub J, Tate JG, et al. Rfam: updates to the RNA families database. *Nucleic Acids Res*, 2009, 37: D136-40
- [32] Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res*, 2005, 33:

- W686-9
- [33] Wilson D, Charoensawan V, Kummerfeld SK, et al. DBD-taxonically broad transcription factor predictions: new content and functionality. *Nucleic Acids Res*, 2008, 36: D88-92
- [34] Loots GG, Ovcharenko I. rVISTA 2.0: evolutionary analysis of transcription factor binding sites. *Nucleic Acids Res*, 2004, 32: W217-21
- [35] Bergman NH, Passalacqua KD, Hanna PC, et al. Operon prediction for sequenced bacterial genomes without experimental information. *Appl Environ Microbiol*, 2007, 73 (3): 846-54
- [36] Bisant D, Maizel J. Identification of ribosome binding sites in *Escherichia coli* using neural network models. *Nucleic Acids Res*, 1995, 23(9):1632-9
- [37] Oliveira MF, Madureira DQ, Ferrari LI, et al. Ribosome binding site recognition using neural networks. *Genet Mol Biol*, 2004, 27(4): 644-50
- [38] Finn RD, Mistry J, Tate J, et al. The Pfam protein families database. *Nucleic Acids Res*, 2010, 38: D211-2
- [39] Letunic I, Doerks T, Bork P. SMART 6: recent updates and new developments. *Nucleic Acids Res*, 2009, 37: D229-32
- [40] Ettwing KF, Butler MK, Le Paslier D, et al. Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature*, 2010, 464: 543-8
- [41] Lämmle K, Zipper H, Breuer M, et al. Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. *J Biotechnol*, 2007, 127: 575-92
- [42] Mayumi D, Akutsu-Shigeno Y, Uchiyama H, et al. Identification and characterization of novel poly(DLlactic acid) depolymerases from metagenome. *Appl Microbiol Biotechnol*, 2008, 79: 743-50
- [43] Chu X, He H, Guo C, et al. Identification of two novel esterases from a marine metagenomic library derived from south China sea. *Appl Microbiol Biotechnol*, 2008, 80: 615-25
- [44] Jeon JH, Kim JT, Kim YJ, et al. Cloning and characterization of a new cold-active lipase from a deep-sea sediment metagenome. *Appl Microbiol Biotechnol*, 2009, 81: 865-74
- [45] Suenaga H, Ohnuki T, Miyazaki K. Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. *Environ Microbiol*, 2007, 9: 2289-97
- [46] Waschowitz T, Rockstroh S, Daniel R. Isolation and characterization of metalloproteases with a novel domain structure by construction and screening of metagenomic libraries. *Appl Environ Microbiol*, 2009, 75: 2506-16
- [47] Simon C, Herath J, Rockstroh S, et al. Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. *Appl Environ Microbiol*, 2009, 75: 2964-8
- [48] Stepanauskas P, Sieracki ME. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc Natl Acad Sci USA*, 2007, 104(21): 9052-7
- [49] Woyke T, Xie G, Copeland A, et al. Assembling the marine metagenome, one cell at a time. *Plos One*, 2009, 7 (4): e5299
- [50] Lee SK, Chou H, Ham TS, et al. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Curr Opin Biotech*, 2008, 19: 556-63
- [51] Tyson GW, Chapman J, Hugenholtz P, et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, 2004, 428: 37-43
- [52] Venter JC, Remington K, Heidelberg J, et al. Environment genome shotgun sequencing of the Sargasso Sea. *Science*, 2004, 304: 66-74
- [53] Warnecke F, Lnginbuhl P, Ivanova N, et al. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature*, 2007, 450: 560-5
- [54] Healy FG, Ray RM, Aldrich HC, et al. Direct isolation of functional genes encoding cellulases from the microbial con sortie in a thermophilic, anaerobic digester maintained on lignocelluloses. *Appl Microbiol Biotechnol*, 1995, 43: 667-74
- [55] Rees HC, Grant S, Jones B, et al. Detecting cellulose and esterase enzyme activities encoded by novel genes present in environmental DNA libraries. *Extremophiles*, 2003, 7(5): 415-21
- [56] Voget S, Leggewie C, Uesbeck A, et al. Prospecting for novel biocatalysts in a soil metagenome. *Appl Environ Microbiol*, 2003, 69(10): 6235-42
- [57] Voget S, Steele HL, Streit WR. Characterization of a metagenome-derived halotolerant cellulose. *J Biotechnol*, 2006, 126: 26-36
- [58] Ferrer M, Golyshina OV, Chernikova TN, et al. Novel hydrolase diversity retrieved from a metagenomic library of bovine rumen microflora. *Environ Microbiol*, 2005, 7:1996-2010
- [59] Feng Y, Duan CJ, Pang H, et al. Cloning and identification of novel cellulose genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulose. *Appl Microbiol Biotechnol*, 2007, 75: 319-28
- [60] Brulc JM, Antonopoulos DA, Miller MEB, et al. Gene-centric metagenomics of the fiber-adherentbovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc Natl Acad Sci USA*, 2009, 106: 1948-53
- [61] Hess M, Sczyrba A, Egan R, et al. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science*, 2011, 331 (6016): 463-7
- [62] Liu N, Yan X, Zhang M, et al. Microbiome of fungus-growing termites: a new reservoir for lignocellulase genes. *Appl Environ Microbiol*, 2011, 77(1): 48-56
- [63] Hasty J, McMillen D, Collins JJ. Engineered gene circuits. *Nature*, 2002, 420 (14): 224-30
- [64] Michalowski CB, Short MD, Little JW. Sequence tolerance of the phage PRM promoter: implications for evolution of gene regulatory circuitry. *J Bacteriol*, 2004, 186: 7899-999
- [65] Elowitz MB, Leibler S. A synthetic oscillatory network of

- transcriptional regulators. *Nature*, 2000, 403: 335-8
- [66] Basu S, Mehreja R, Thiberge S, et al. Spatiotemporal control of gene expression with pulse-generating networks. *Proc Natl Acad Sci USA*, 2004, 101(17): 6355-60
- [67] Chang DTH, Huang CY, Wu CY, et al. YPA: an integrated repository of promoter features in *Saccharomyces cerevisiae*. *Nucleic Acids Res*, 2011, 39: D647-52
- [68] Hammer K, Mijakovic I, Jensen PR. Synthetic promoter libraries-tuning of gene expression. *Trends Biotechnol*, 2006, 24(2): 53-5
- [69] Hartner FS, Ruth C, Langenegger D, et al. Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. *Nucleic Acids Res*, 2008, 36 (12): e76
- [70] Qin XL, Qian JC, Yao GF, et al. *GAP* promoter library for fine-tuning gene expression in *Pichia pastoris*. *Appl Environ Microbiol*, 2011, 77(11): 3600-8
- [71] Jin YS, Ni HY, Laplaza JM, et al. Optimal growth and ethanol production from xylose by recombinant *Saccharomyces cerevisiae* require moderate D-xylulokinase activity. *Appl Environ Microbiol*, 2003, 69: 495-503
- [72] Jones KL, Kim SW, Keasling JD. Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. *Metab Eng*, 2000, 2: 328-38
- [73] Tao L, Jackson RE, Cheng Q. Directed evolution of copy number of a broad host range plasmid for metabolic engineering. *Metab Eng*, 2005, 7:10-7
- [74] Nevoigt E, Kohnke J, Fischer CR, et al. Engineering of promoter replacement cassettes for fine-tuning of gene expression in *Saccharomyces cerevisiae*. *Appl Environ Microbiol*, 2006, 72 (8): 5266-73
- [75] Yan W, Craig EA. The glycine-phenylalanine-rich region determines the specificity of the yeast Hsp40 Sis1. *Mol Cell Biol*, 1999, 19: 7751-8
- [76] Jeppsson, M, Johansson B, Jensen PR, et al. The level of glucose-6-phosphate dehydrogenase activity strongly influences xylose fermentation and inhibitor sensitivity in recombinant *Saccharomyces cerevisiae* strains. *Yeast*, 2003, 20:1263-72
- [77] Stemmer WP. Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature*, 1994, 370: 389-91
- [78] Glieder A, Farinas ET, Arnold FH. Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. *Nat Biotechnol*, 2002, 20: 1135-9
- [79] Alper H, Fischer C, Nevoigt E, et al. Tuning genetic control through promoter engineering. *Proc Natl Acad Sci USA*, 2005, 102 (36): 12678-83
- [80] Nevoigt E, Fischer C, Mucha O, et al. Engineering promoter regulation. *Biotechnol Bioeng*, 2007, 96(3): 550-8
- [81] Guet CC, Elowitz MB, Hsing W, et al. Combinatorial synthesis of genetic networks. *Science*, 2002, 296: 1466-70
- [82] Guido NJ, Wang X, Adalsteinsson D, et al. A bottom-up approach to gene regulation. *Nature*, 2006, 439: 856-60
- [83] Ellis T, Wang X, Collins JJ. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat Biotechnol*, 2009, 27(5): 465-71
- [84] Densmore D, Hsiao TH, Kittleson JT, et al. Algorithms for automated DNA assembly. *Nucleic Acids Res*, 2010, 38 (8): 2607-16
- [85] Kodumal SJ, Patel KG, Reid R, et al. Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc Natl Acad Sci USA*, 2004, 101(17): 6355-60
- [86] Saftalov L, Smith PA, Friedman AM, et al. Site-directed combinatorial construction of chimaeric genes: general method for optimizing assembly of gene fragments. *Protein*, 2006, 64(3): 629-42
- [87] Yehezkel TB, Linshiz G, Buaron H, et al. *De novo* DNA synthesis using single molecule PCR. *Nucleic Acids Res*, 2008, 36(17): e107
- [88] Shao ZY, Zhao H, Zhao HM. DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res*, 2009, 37(2):e16
- [89] Gibson DG, Young L, Chuang RY, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*, 2009, 6(5): 343-5
- [90] Li MZ, Elledge SJ. Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. *Nat Methods*, 2007, 4 (3): 251-6
- [91] Hill A, Tomshine J, Wedding E, et al. SynBioSS: the synthetic biology modeling suite. *Bioinformatics*, 2008, 24(21): 2551-3
- [92] Chandran D, Bergmann FT, Sauro HM. TinkerCell: modular CAD tool for synthetic biology. *J Biol Eng*, 2009, 3:19
- [93] Cai Y, Wilson ML, Peccoud J. GenoCAD for iGEM: a grammatical approach to the design of standard-compliant constructs. *Nucleic Acids Res*, 2010, 38(8): 2637-44
- [94] Rialle S, Felicori L, Dias-Lopes C, et al. BioNetCAD: design, simulation and experimental validation of synthetic biochemical networks. *Bioinformatics*, 2010, 26(18): 2298-304
- [95] Weeding E, Houle J, Kaznessis YN. SynBioSS designer: a web-based tool for the automated generation of kinetic models for synthetic biological constructs. *Brief Bioinform*, 2010, 11(4): 394-402