

特约评述

DOI: 10.12211/2096-8280.2020-028

基于合成生物学策略的酶蛋白元件规模化挖掘

张建志, 付立豪, 唐婷, 张嵩亚, 朱静, 李拓, 王子宁, 司同

(中国科学院深圳先进技术研究院, 深圳合成生物学创新研究院, 中国科学院定量工程生物学重点实验室, 广东 深圳 518055)

摘要: 生物制造以人工生物体系为催化剂合成工业化学品、药物和功能材料, 具有低碳循环、绿色清洁等特征。酶蛋白是构建生物催化系统的重要功能单元, 然而, 由于缺乏准确预测序列-功能关系的方法, 目前酶的理性设计仍面临巨大挑战。因此, 需要利用合成生物学工程化的思路 and 手段, 从自然界中大规模挖掘新的酶蛋白元件, 相关研究不但可以为开发工业酶制剂和构建细胞合成代谢提供优质元件, 而且有利于快速获得酶蛋白序列-结构-功能间的对应关系, 为建立预测与设计模型提供基础。本文针对酶元件工程化挖掘的关键技术进行综述: 介绍了计算机辅助设计的算法和软件, 用于将数据库中海量的酶蛋白序列按照实验目的进行聚类分析和优先化排序; 总结了规模化合成组装、异源表达和功能筛选酶蛋白元件的高通量实验技术; 讨论了如何综合利用计算与实验手段, 系统性探索酶家族成员的催化性能。未来, 通过综合计算机辅助设计、自动化合成生物构建、高通量测试等方法, 设计和建设高度集成的工程化研究平台, 成为实现对酶蛋白资源进行系统化的研究和挖掘的重要方向。

关键词: 酶; 计算机辅助设计; 高通量技术; 蛋白表达; 合成生物学

中图分类号: Q55 **文献标志码:** A

Scalable mining of proteins for biocatalysis *via* synthetic biology

ZHANG Jianzhi, FU Lihao, TANG Ting, ZHANG Songya, ZHU Jing, LI Tuo, WANG Zining, SI Tong

(CAS Key Laboratory of Quantitative Engineering Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, Guangdong, China)

Abstract: Biomanufacturing provides a sustainable alternative to traditional petrochemical processes in producing chemicals, drugs, and functional materials. Enzymes are cores for creating catalytic biosystems with diverse functions. Due to the lack of predictive models for enzyme functions, however, rational design is still challenging. On the other hand, next-generation sequencing reveals millions of diverse natural enzymes, of which only a tiny fraction have been experimentally characterized. Synthetic biology applies engineering principles to study, engineer, and create biological systems. Through standardization and modularization, synthetic biology enables large-scale prototyping of enzyme

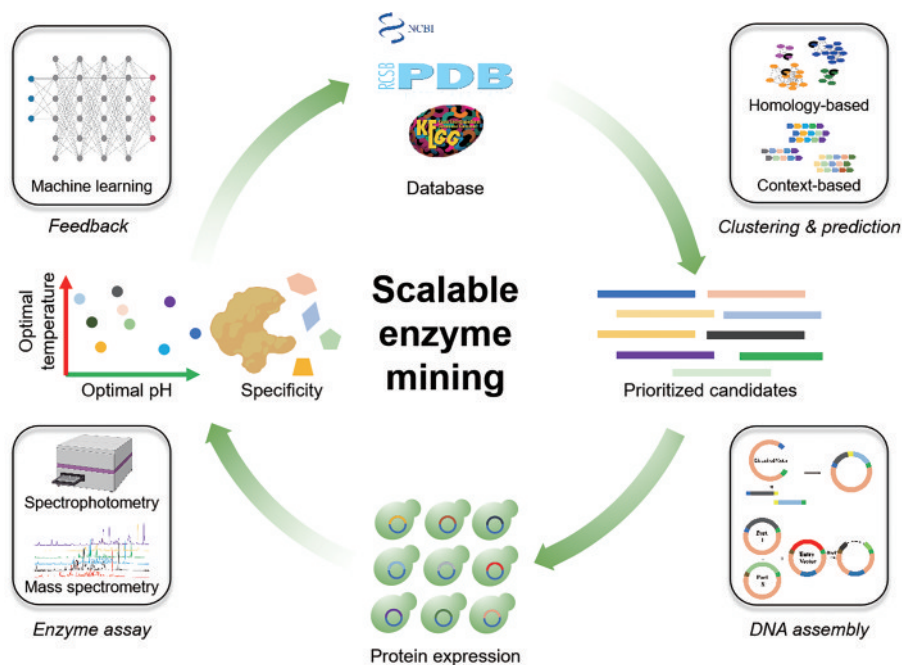
收稿日期: 2020-03-17 修回日期: 2020-04-29

基金项目: 深圳合成生物学创新研究院主题项目 (ZTXM20190002)

引用本文: 张建志, 付立豪, 唐婷, 张嵩亚, 朱静, 李拓, 王子宁, 司同. 基于合成生物学策略的酶蛋白元件规模化挖掘[J]. 合成生物学, 2020, 1(3): 319-336

Citation: ZHANG Jianzhi, FU Lihao, TANG Ting, ZHANG Songya, ZHU Jing, LI Tuo, WANG Zining, SI Tong. Scalable mining of proteins for biocatalysis *via* synthetic biology[J]. Synthetic Biology Journal, 2020, 1(3): 319-336

sequences, which not only helps to identify efficient biocatalytic parts, but also accelerates quantitative understanding of sequence-structure-function relationship. Here we review recent advances in scalable mining of enzymes *via* synthetic biology. We firstly introduce computational tools for functional clustering and prioritization of promising sequences from enormous genome/protein databases, followed by experimental approaches for high-throughput cloning, expression, and characterization of selected candidates. We then discuss the applications of such tools in systematic studies of enzyme (super) families. We conclude with future perspectives in creating integrated synthetic biology foundries to accelerate enzyme mining.



Keywords: enzyme; computer-aided design (CAD); high-throughput technologies; recombinant protein expression; synthetic biology

酶 (enzyme) 是一种重要的生物催化剂, 可以促进生物体内化学反应在生理条件下高效和特异地进行, 其化学本质是蛋白质或RNA^[1]。酶蛋白具有催化效率高、专一性强、作用条件温和、环境友好等特点, 被用于开发工业酶制剂和构建微生物合成代谢途径^[2-4], 在医药、化工、能源、材料等领域有广泛应用^[5-7]。例如, 研究者从罂粟、褐鼠、假单胞菌、黄连和花菱草中挖掘酶蛋白, 构建了包含18个酶催化步骤的合成途径, 实现了抗癌化合物诺司卡品在酿酒酵母中的高效生产^[8-9]。酶蛋白具有天然多样性, 截至2019年, NCBI (National Center for Biotechnology Information) 数据库中中共有约 4×10^8 种特异性的蛋白序列; 且随着

DNA 测序技术的快速发展使得蛋白序列数量每24个月增加约1倍^[10-12]。在此背景下, 对酶蛋白资源进行规模化的挖掘和开发变得极为重要和迫切。

传统酶资源挖掘一般采用菌种筛选、功能宏基因组等技术。菌种筛选以催化活性为导向, 首先对天然微生物进行培养与筛选, 再通过分子生物学技术克隆与鉴定相关基因, 或者对微生物蛋白组进行提取、分离与组学分析, 鉴定活性酶蛋白; 功能宏基因组方法基于活性或DNA序列相似性, 对表达宏基因组DNA片段的克隆进行筛选, 用于挖掘环境中的未培养微生物基因资源。两类方法都依赖实体样品, 研发周期长, 成本高, 具有盲目性。

相较于传统方法,合成生物学为酶蛋白的规模化研究提供了新的思路。基于工程学理念,合成生物学对生物体进行有目标的“设计—构建—测试—学习”,获得具有特定功能的人工生命系统。就酶催化系统而言,合成生物学采用“自下而上”的策略,首先针对目标生化反应,从数据库中识别酶蛋白的氨基酸序列;接着根据底盘细胞适配原则,设计并合成酶元件、调控元件的DNA序列,利用标准化DNA组装方法构建蛋白表达模块或合成代谢通路;最后在底盘细胞中进行转化、表达与功能表征。与传统方法相比,基于合成生物学进行酶蛋白挖掘具有诸多优点,如不受生物实体的限制,可以利用生信方法针对性确定研究对象,提高新颖元件的开发效率;DNA合成可以进行序列工程化改造,从密码子偏好、亚细胞定位等方面提高酶元件与底盘的适配性;基于标准化底盘、流程和实验条件开展研究,有利于积累优质元件与数据,为人工智能指导的酶元件挖掘等研究方法奠定基础。

利用合成生物学对酶蛋白进行工程化挖掘,需要开发高通量的计算和实验方法。目前缺乏从序列准确预测功能的方法,往往需要筛选大量候选序列,才能够识别自然界中催化目标生化反应的酶蛋白;同时,天然酶元件往往需要进行工程改造才能达到特定指标,如催化效率、专一性、底盘细胞适配性等。因此,需要基于合成生物学理念开发工程化、标准化的技术和流程,从海量测序数据中识别、合成、表征天然酶元件。针对酶元件的识别,传统方法基于同源序列比对、蛋白质家族结构域比对等原理进行酶功能注释。但是,序列相似性与功能之间并不总是完全对应,例如序列同源性达到98%的蛋白质可能具有不同的生化功能^[13],而序列同源性很低的蛋白质也可能具有相同的生化功能^[14-15]。因此,需要开发新的生物信息方法,对酶蛋白从序列、三维结构、进化关系和蛋白互作等多个层面进行综合性分析,根据实验目的优先化排序用于功能验证的候选序列。另外,不同来源酶蛋白在底盘细胞中进行异源表达时,可能存在密码子偏好不同、无法折叠、稳定性低、辅酶因子及底物缺失等适配问题,需要进行工程化实验设计(design of experiment,

DOE),对DNA序列、调控元件、DNA组装方法、蛋白表达条件等因素进行系统性探索和优化。综上所述,对上百个候选序列的功能表征和对实验条件的系统性优化,对研究平台的操作通量和自动化水平提出了新的需求。

本文作者对利用合成生物学方法高通量挖掘酶元件的研究进展进行综述。介绍了相关的生信工具,已有文献对多序列比对(multiple sequence alignment, MSA)、基因注释、同源建模(homologous modeling)等内容进行了系统总结^[16-17],本文着重关注聚类分析、可溶性预测、杂泛性预测等算法。总结了对候选酶元件进行合成组装、异源表达、功能筛选的高通量实验技术。接着,讨论了规模化挖掘技术在酶家族系统性研究方面的应用。搭建自动化合成生物学平台开展酶蛋白高通量研究是未来的趋势。

1 计算机辅助设计用于高通量酶元件挖掘

高通量挖掘酶元件的一般流程如图1所示^[18]。随着测序技术的飞速发展,大量(宏)基因组和转录组得到解析,从中可以预测得到大量蛋白序列。与蛋白质结构域数据库进行同源比对可以实现酶功能的初步注释^[19]。同时,代谢多样性提供了丰富的生物催化资源,各类初级和次生代谢产物合成途径的预测和分析,也为酶元件的挖掘提供了帮助。但是,目前很难从酶序列出发准确预测其催化活性、底物选择性、可溶表达等关键性质,需要开发新的算法对候选元件进行优先化排序,提高功能筛选的成功率,利用最少的实验资源探索相似酶蛋白的功能多样性。下面将从酶的聚类分析(clustering)和实验验证的优先化排序两个方面进行介绍。

1.1 酶蛋白聚类分析

各种类型的蛋白质资源数据库为酶元件的高通量挖掘提供了宝贵的材料,如表1所示。截至2020年2月,蛋白质资源数据库UniProt中储存了约18亿条蛋白质序列。其中大部分功能注释是通过与CDD(conserved domain database)^[20],

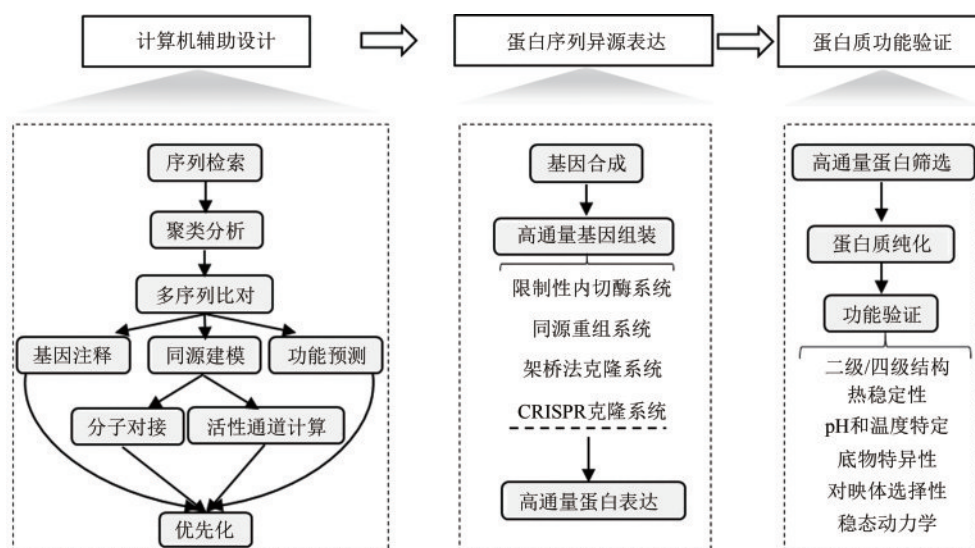


图1 酶蛋白资源高通量挖掘流程图

Fig. 1 Workflow for high-throughput enzyme mining

Pfam^[21], CATH^[22] 和 FIGfams^[23] 等数据库中蛋白的结构域序列进行同源性比对而获得^[19]; 但至少50%的注释并不精确甚至是错误的^[24]。因此, 需要整合不同层次的功能注释工具和实验数据, 帮助科研人员从序列、结构、进化和蛋白互作等多个层面对候选酶元件进行综合分析。其中, 酶蛋白聚类分析可以利用可视化手段最大程度地利用文献中实验数据对未知蛋白质进行多个维度的功能注释^[25]。下面将重点介绍用于酶蛋白聚类分析的两个方法: 序列相似性网络分析工具 (sequence similarity network, SSN)^[26], 以及侧重蛋白质结构比对的 CATH 分析工具 (Class, Architecture, Topology, Homologous superfamily)^[22]。

SSN 是一种可显示同源蛋白之间两两序列相似性关系的多维网络^[27]。网络中每个蛋白质由一个节点 (node) 表示; 如果两个蛋白质间具有超过序列相似性 (blast *E*-value) 的指定阈值, 则代表它们的两个节点通过一条边 (edge) 连接; 每个簇 (cluster) 中的节点至少有一条边与簇中的其他节点相连。选择合适的阈值是成功构建 SSN 的关键。SSN 中的“节点属性”包含每个节点的各种信息, 比如系统分类, TrEMBL/SwissProt、PDB 和 GO 等数据库信息的链接^[19, 28-29]。这些信息有助于用户对 SSN 进行分析, 通过设定合适的阈值将不同节点聚类。与多序列比对和系统进化树 (phylogenetic tree) 方法相比, SSN 能够快速分析更大的蛋白

质序列集, 并能同时评估正交信息。例如, 映射到序列相似性的功能多样性, 有助于对酶家族中共享显著序列相似性的子群如何关联进行推测。SSN 分析也可揭示序列相似性较低但具有相似结构或功能的远同源蛋白之间的关系, 从而指导研究者优先探索序列空间内的未知区域^[27, 30]。最初, Barber 等^[31] 利用 Python 语言编写了 Pythoscape 工具用于生成 SSN; 但该工具需要基于服务器集群在 Unix 环境下执行命令行脚本, 并未得到广泛的应用。为了推广 SSN 的应用, 研究者开发了用户友好的 EFI-EST 网络工具 (enzyme similarity tool)^[25, 32], 可以通过浏览器访问 (<https://efi.igb.illinois.edu/efi-est/>), 使得用户不需要专业编程知识即可快速生成 SSN, 其结果可以在开源的 Cytoscape 软件中进行可视化分析。

CATH 工具可以揭示蛋白质结构域间的进化关系。基于空间结构的相似性, CATH 识别 wwPDB 数据库 (Worldwide Protein Data Bank) 中的蛋白质结构域, 并将其聚类形成结构域超家族 (domain superfamily)^[22]。由于仅有少数蛋白质 (10 万余种) 具有实验测定的三维空间结构信息, CATH-Gene3D 工具将具有代表性的结构域转换成隐性 Markov 模型, 形成基于一级序列的“指纹图谱”库; 利用这一方法, 共预测得到 43 万个蛋白结构域, 构成 6000 余个结构域超家族。CATH 进一步按照 Class (二级结构的组成)、Architecture

(二级结构形成的形状) 和 Topology (二级结构连接的顺序) 的三个层次, 分析结构域超家族间的进化关系^[33]。另外, FunTree 数据库进一步将酶蛋白的功能信息, 如生化反应物的结构、酶催化动力学数据等, 与 CATH 结构域超家族分类进行整合, 从而综合分析序列、结构、功能和进化关系。该数据库目前已包含 2340 个结构域超家族、7 万个结构域和 40 万个代表性序列^[34]。

表1 酶元件资源数据库及链接

Tab. 1 Enzyme resource database and their corresponding links

蛋白质序列数据库	基本信息
Uniprot https://www.uniprot.org/	信息最丰富、资源最广的蛋白质数据库。由 Swiss-Prot、TrEMBL 和 PIR-PSD 三大数据库的数据组合而成
NCBI https://www.ncbi.nlm.nih.gov/	
Conserved Domains	蛋白质保守结构域数据库,是关于蛋白质功能单元注释的资源
Protein	包含 GenBank, RefSeq, TPA, SwissProt, RIP, PDB 等数据库中的序列
Protein cluster	包括完整的原核生物基因组和叶绿体基因组编码的 RefSeq 蛋白质序列
Structure	数据来源于 PDB 的蛋白质结构数据库,并将结构数据链接到书目信息、序列数据库和 NCBI 的 Taxonomy 中,运用 3D 结构浏览器和 Cn3D,可以从 Entrez 获得分子间相互作用的图像
PDB http://www.rcsb.org/	生物大分子(如蛋白质和核酸)数据库,包括由全世界生物学家和生物化学家上传的蛋白质或核酸的 X 射线晶体衍射或者 NMR 核磁共振结构数据。从 PDB 的网站上,可以通过蛋白质的编号查找到相应的 3D 结构,并通过 PyMol、RasMol、Chimera、VMD、Swiss-PdbViewer 等软件查看、编辑
InterPro http://www.ebi.ac.uk/interpro/	整合了 CATH、CDD、PRINTS、Pfam 等多个数据库,并去掉冗余数据,对蛋白质家族预测、结构域和结合位点预测进行注释
Brenda https://www.brenda-enzymes.org/	包括酶促反应、特异性、动力学参数、结构和稳定性等蛋白质功能数据及基因组序列信息
BKMS-react http://bkm-react.tu-bs.de/	包括酶促反应、动力学参数、实验条件和代谢途径等信息
EzCatDb http://ezcatdb.cbrc.jp/EzCatDB/	包括酶促反应、辅因子、中间代谢产物、催化活性结构域和结构等信息
M-CSA ^b https://www.ebi.ac.uk/thornton-srv/m-csa/browse/	对催化残基、辅助因子和反应机理的注释
FireProt https://loschmidt.chemi.muni.cz/fireprotodb/	包括序列突变后导致的热稳定性变化数据
ProTherm https://www.iitm.ac.in/bioinfo/ProTherm/	包括序列突变后导致的热稳定性变化数据
eSOL http://tanpaku.org/tp-esol/index.php?lang=en	基于蛋白质翻译和离心条件推测蛋白质溶解度
SoluProtMut https://loschmidt.chemi.muni.cz/soluprotmutdb/	包括序列突变后导致的溶解性变化数据
TargetTrack http://dx.doi.org/10.5281/zenodo.821654	蛋白质溶解性数据库
ProtaBank https://protabank.org/	包括各种突变后的蛋白质序列信息
hPDB http://hackage.haskell.org/package/hPDB	可动态展示生物大分子立体结构
KEGG https://www.kegg.jp/	整合基因组、化学和系统功能信息的数据库。把从已经完整测序的基因组中得到的基因目录与更高级别的细胞、物种和生态系统水平的系统功能关联起来

续表

蛋白质序列数据库	基本信息
PMD http://www.proteinmicroarray.cn/	蛋白质突变数据库
Pfam http://pfam.xfam.org/	蛋白质家族数据库,蛋白质家族以多序列比对和隐马尔科夫模型的形式表示
FIGFAMs http://blog.theseed.org/servers	基于人工注释的来源于不同细菌,古菌等生物的蛋白质序列构成,同功能蛋白质组合为相应的集合
PRINTS http://bioinf.man.ac.uk/dbbrowser/PRINTS	蛋白质 Motif 数据库
CATH http://cathdb.info/	以蛋白质结构域层次组织的包含类型(C)、架构(A)、拓扑学(T)和同源超家族的数据库
SCOP http://scop.mrc-lmb.cam.ac.uk/	蛋白质结构分类
CCDC https://ccdc.cam.ac.uk/	剑桥晶体中心数据库
SWISS-3DIMAGE http://us.expasy.org/sw3d/	蛋白质及其他生物分子的三维图像
BioMagResBank http://www.bmrb.wisc.edu/	蛋白质、氨基酸和核苷酸的核磁共振数据库
SWISS-MODEL Repository http://swissmodel.expasy.org/repository/	自动产生蛋白质模型的数据库
PROTOMAP https://www.scripps.edu/cravatt/protomap/	Swiss-Prot 蛋白质自动分类系统
iProClass http://pir.georgetown.edu/iproclass/	蛋白质分类数据库
TIGRFAM http://www.tigr.org/TIGRFAMs/	蛋白质家族数据库
OWL http://www.bioinf.man.ac.uk/dbbrowser/OWL/	非冗余蛋白序列数据库,由 SWISS-PROT、PIR、GenBank(由其编码序列翻译而成的氨基酸序列)和 NRL-3D 一级序列数据库集合而成
3DID http://3did.irbbarcelona.org	包括 3D 结构已知的蛋白质的互作信息
DOMINE http://manticore.niehs.nih.gov/cgi-bin/Domine	蛋白质结构域互作数据库
PiSite http://pisite.hgc.jp	以 PDB 为基础,可在蛋白序列中搜寻互作位点
Binding MOAD http://www.BindingMOAD.org	提供蛋白质-配体晶体结构数据信息
Phospho.ELM http://phospho.elm.eu.org	蛋白质磷酸化位点数据库
STITCH http://stitch.embl.de/	蛋白质-化合物作用网数据库
Reactome http://www.reactome.org	人体生命活动路径与过程数据库,提供生化过程网络图,并对参与其中的蛋白质分子有详细注解,与其他数据库如 UniPort、KEGG、OMIM 等建立了广泛的交叉应用
UniHI http://www.unihi.org	人体蛋白质-蛋白质相互作用数据库
Bionemo http://bionemo.bioinfo.cnio.es	包括与生物降解代谢相关的蛋白质、基因数据,包括蛋白序列、结构域、结构;基因序列、调控元件、转录单元等信息。除此之外还包括生物降解的代谢路径图、相关生化反应等

续表

蛋白质序列数据库	基本信息
CAZy http://www.cazy.org/	碳水化合物活性酶数据库, 包括能够合成或者分解复杂碳水化合物的酶类
LED http://www.led.uni-stuttgart.de/	脂肪酶工程数据库, 整合了脂肪酶的结构域序列信息
MEROPS https://www.ebi.ac.uk/merops/	肽酶数据库, 同时也包括一些蛋白质类肽酶抑制剂的信息
PLANT-PIs http://plantpis.ba.itb.cnr.it/	植物蛋白酶抑制剂及相关基因信息
PROSPER https://prosper.erc.monash.edu.au	基于序列预测 24 个不同蛋白酶家族的催化底物和切割位点
CPDB 140.113.15.73/~lab/deprecated/cpdb-2009/	利什曼原虫中半胱氨酸蛋白酶注释数据库
SABIO-RK http://sabio.h-its.org/	包括生化反应动力学参数的数据库

1.2 实验验证的优先化

聚类分析对海量蛋白质数据库进行初步筛选之后^[35], 仍存在大量候选序列, 远远超出实验验证的能力^[18]。例如, 软海绵素 B 合成途径包含 47 步不同的酶催化反应, 假设每个步骤候选酶数量为 100 个, 则需要构建和筛选的代谢途径组合可达到约 10⁹⁴ 个^[36]。因此, 需要探索优先化标准与算法对候选酶进行排序, 提高功能验证实验的效率^[2, 4]。

酶蛋白的可溶性表达和底物杂泛性 (promiscuity) 是优先化排序的重要标准。可溶性表达是对酶进行功能表征的前提, 目前已有一些基于能量计算、机器学习和进化分析的算法对这一性质进行预测^[37]。例如, Wilkinson-Harrison 溶解度模型可以预测蛋白序列在大肠杆菌中可溶性表达的概率^[38]。Vanacek 等^[18]对脱卤酶家族进行生信分析及优先化选择之后, 筛选得到 20 个候选蛋白序列进行表达实验, 最终有 60% 的蛋白质在大肠杆菌中实现可溶性表达, 与模型预测得到的理论值一致。另外, 需要根据实验目的选择具有不同底物杂泛性的酶。体外生物催化体系通常使用单一底物, 使用高底物杂泛性酶不影响反应专一性, 并且有利于将同一种酶应用于不同反应体系, 缩短研发周期和成本。与此相反, 高底物杂泛性的酶可能会在复杂的体内环境中产生副反应,

从而消耗能量、辅酶、关键前体等细胞资源, 或者导致毒性物质的积累, 因此构建细胞合成代谢途径需要优先化选择专一性较高的酶^[5, 39]。目前, 研究者已开发了一系列的算法用来预测酶的底物杂泛性, 例如: 基于酶学分类对新型化学结构和反应预测的 BNICE^[40], 基于理化性、单肽/二肽分布、分子量、等电点、氨基酸序列信息的 SVM^[41], 基于分子电性参数、立体参数、疏水参数、取代基等参数的 QSAR^[42] 和基于蛋白质三维结构信息的 BioGPS^[43] 等算法。目前, 未见系统性研究比较各类预测算法的效率、准确度等性能; 未来酶蛋白的规模化挖掘实验可以为算法评估提供数据支持。

构建生物合成途径时还应考虑其他因素。例如, 大部分算法都倾向于选择最少反应步骤的代谢途径, 以减少参与酶的数量; 通过预测每一步催化反应的 ΔG 来计算构建整条合成途径的热力学可行性; 避免毒性中间产物的产生以及维持代谢流的平衡等^[34]。目前, 研究者已经开发出多种算法用于构建合成途径中酶的优先化选择。例如, GEM-Path 可以预测在不同溶氧和生长条件下目的产物的产量^[44]; RetroPath 则是将合成途径中每一步催化反应所需候选酶的数量和酶的底物杂泛性等因素作为计算参数, 从而指导每一步催化反应所需酶的挖掘^[45]; Genomatica 等使用 SimPheny Biopathway Predictor

预测得到了10000个不同的、包含4~6个基因的1,4-丁二醇合成途径,并结合酶催化反应步骤数、热力学可行性和产量等标准进行优先化排序,缩减实验量,最终得到了具有工业化价值的菌株^[35]。

2 酶蛋白挖掘的高通量实验技术

由于无法从氨基酸序列准确预测酶的催化功能,需要对大量候选序列进行实验表征,筛选出与目标性能最为接近的酶元件。另一方面,由于候选酶蛋白往往来自于不同物种,需要系统性优化密码子、调控元件、细胞培养条件等实验方案,提高酶蛋白与异源底盘的适配性。为此,研究者开发了一系列高通量技术用于酶蛋白的规模化表征,以及实验方案的组合式优化^[2, 4]。下面分别就分子克隆、蛋白质表达和功能筛选等三个环节进

行介绍。

2.1 高通量分子克隆

高效可靠的DNA组装方法对于实现候选酶元件的规模化克隆至关重要,并且可以针对表达载体、调控元件、融合蛋白等因素进行组合式优化,提高异源表达的成功率^[46]。下面将介绍满足根据不同实验需要的高通量分子克隆方法(表2)。针对酶元件DNA序列的优化策略,如密码子优化、酶切位点移除等,已有文献进行系统性总结^[47],本文不再复述。

2.1.1 限制性内切酶克隆系统(restriction enzyme-based cloning)

近几十年,限制性内切酶克隆方法已广泛地应用于分子生物学实验当中,包括BioBrick、

表2 基因高通量克隆、组装方法

Tab. 2 Methods for high-throughput gene cloning and assembly

分类	原理	特点
限制性内切酶克隆系统	限制性内切酶使基因片段和载体产生黏性末端,通过T4连接酶组装	1. 普通的双限制性内切酶克隆操作简单,但需考虑目的基因序列中酶切位点序列,适合单基因克隆; 2. 稀有限制性内切酶 <i>SgfI</i> 和 <i>PmeI</i> 使其具备高通量操作的可行性; 3. <i>Golden Gate</i> 可同时实现多片段酶切与模块化组装,但也需要移除目的序列中的 <i>BsaI</i> 及 <i>BsmBI</i> 等识别位点; 4. <i>MASTER</i> 只能识别甲基化的4bp位点,无需考虑目的基因序列是否含有 <i>MspJI</i> 识别位点序列
同源重组克隆系统	基于基因片段与载体两端的同源序列进行克隆组装	1. 操作简单、效率高、快速、稳定,适合单基因或多基因组装; 2. 无限制性内切酶和连接酶的参与; 3. 无需考虑目的基因序列内部的酶切位点; 4. 存在重复序列时难以组装
基于寡核苷酸的架桥法克隆系统	基于热稳定的连接酶和寡核苷酸进行DNA组装	在相似的反应条件下(12个DNA片段的组装),连接反应的保真性与酵母胞内DNA组装方法相近,并高于CPEC和Gibson等方法
基于CRISPR的克隆系统	基于II类type V CRISPR系统和 <i>Taq</i> DNA连接酶开发的DNA无缝拼接方法	仅需 <i>CpfI</i> 一种内切酶进行切割,经过条件优化或蛋白改造,连接效率可达到70%以上,并可实现21kbp长度的DNA组装

Flexi、Golden Gate等克隆方法。在使用常规的限制性内切酶克隆方法时，插入片段和目标载体都需要经过限制性内切酶处理，操作时需要避免目的基因内部有使用的内切酶识别位点序列，因此较难实现多基因的平行操作。稀有限制性内切酶 *Sgf I* 和 *Pme I* (*Sgf I*: GCGATCGC; *Pme I*: GTTTAAAC) 的发现及 Flexi 克隆系统的开发，使内切酶切系统具有了高通量应用的可行性。*Sgf I* 和 *Pme I* 识别位点序列在许多模式生物中出现频率很低（人 1.2%，鼠 1.2%，酿酒酵母 2.96%，拟南芥 2.4%，大肠杆菌 6.35%），而且该方法可以将基因序列在不同的 Flexi 载体之间实现简单转换，不需要重新测序^[48]。Nagase 等^[49] 使用 Flexi 系统实现了人类基因组中的 1929 个基因的高通量克隆，证明了该系统在操作中的良好应用性。

Golden Gate 组装方法基于 II s 型限制内切酶在同一反应体系中进行酶切和质粒组装^[48]。II s 型限制内切酶，如 *Bsa I*、*BsmB I* 等，在其识别序列的外侧进行切割产生 4bp 的黏性末端。因此，只需在相邻片段上合理地设计 4bp 的互补序列，就可进行无痕组装。理论上，4bp 的序列可以组成 256 种不同互补区，而且不依赖于相邻片段之间的同源性，因此可实现多片段快速组装，且不受重复序列的影响；并且，通过多种限制性内切酶的替换使用，可以实现标准化基因元件的多轮逐级组装（如调控和催化元件、单一酶蛋白表达框、代谢通路等）。值得注意的是，4bp 接口序列及其组合对于组装的效率、准确性等有很大影响^[50]，需要利用经验和算法优化接口序列设计（如 <http://cuba.genomefoundry.org/home>）。基于这一组装方法，赵惠民团队^[51] 开发了类转录激活因子感受器核酸酶（transcription activator-like effector nucleases, TALEN）的自动化组装方法，可以一步组装 15 个 DNA 片段，基于机器人平台每天可以合成超过 400 个 TALEN 蛋白对，成功率超过 96%。此外，研究者已开发了多个基于 Golden Gate 的自动化 DNA 组装方法^[52-54]。与 Flexi 方法类似，该组装方法的一个限制因素是需要移除目的基因序列的 *Bsa I*、*BsmB I* 等酶切位点。相较于 *Bsa I* 或 *BsmB I*，*Sap I* 的识别位点序列出现的频率更低，具有更好的通用性。此外，赵国屏课题组^[55] 在 2013 年建立了一种

新型 Golden Gate 组装方法，称为 MASTER 连接法（methylation-assisted tailorable ends rational ligation method）。该方法使用同时具有 II m 型和 II s 特性的内切酶 *MspJ I*，*MspJ I* 只能识别甲基化的 4bp 位点，¹⁴CNNR（R=A 或 G），并在识别位点外侧进行切割，因此不受到目的序列是否包括 *MspJ I* 酶切位点的限制。

2.1.2 同源重组克隆系统(sequence homology-based cloning)

同源重组克隆组装依赖于载体和连接片段两端的同源序列。该方法简便高效，既可用于单片段的克隆，也可用于多片段与载体的组装，且不受酶切位点的限制。该系统主要包含 Gibson、Gateway、Echo Cloning、Creator 等体外酶法组装^[48]，以及利用酵母高效同源重组机制的 DNA assembler 等胞内组装方法^[56]。其中，Gateway 方法在高通量克隆组装实验中应用最为广泛，该方法利用 λ 噬菌体与大肠杆菌的染色体之间发生的位点特异性的重组整合 [attB-attP (BP 反应) \rightarrow attL-attR (LR 反应)]，克隆效率可达 95% 以上。其最主要优势是完成入门克隆之后，目的基因序列可以通过 LR 反应高效、简便地连接至其他目的载体^[57]。但是，当目的基因序列的长度超过 3000bp 时，连接效率会降低。另外，研究者也基于 Gibson assembly^[52-53]、DNA assembler^[58-59] 等技术开发了高通量的 DNA 组装方法。这一类克隆系统的共同局限，是在组装含有重复性序列的片段时，容易由于序列内部的非特异性同源重组导致组装失败。

2.1.3 基于寡核苷酸的架桥法克隆系统(bridging oligo)

De Kok 等^[60] 基于连接酶循环反应（ligase cycling reaction, LCR）开发了新型的 DNA 序列组装方法。通过设计与相邻 DNA 的两端序列互补的单链桥接寡核苷酸，在较低温度下进行退火，从而使上游片段的 3' 端与下游片段的 5' 端连接，将 2 个 DNA 片段组装成单个的线性片段。在接下来的循环中，以组装好的线性片段为模板来组装互补链，通过多次热循环，可以将线性 DNA 片段组装成环形质粒，并转化大肠杆菌感受态进行扩增。De Kok 等利用该方法成功地将 12 个 DNA 片段连接至 20kb 的载体上^[60]。

2.1.4 基于CRISPR的克隆系统

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) 是存在于原核生物中的获得性免疫系统。Cpf1 蛋白属于 II 类 V 型 CRISPR 系统; 相较于 Cas9 蛋白, Cpf1 具有相似的基因编辑效率、蛋白质更小、具有较低的脱靶效应和较好的可操作性。由单个 crRNA (CRISPR RNA) 引导, Cpf1 结合在富含胸腺嘧啶的 PAM (protospacer adjacent motif) 位点相邻并与 crRNA 互补的 DNA 序列, 切割靶标 DNA 互补链的 23 位和非互补链的 18 位, 从而形成 5 nt 黏性末端。Cpf1 的切割位点受 crRNA 中靶向 spacer 序列长度的影响: 当 spacer 序列长度大于等于 20bp 时, Cpf1 倾向于切割非互补链的 18 位; spacer 序列长度小于 20bp 时, Cpf1 倾向于切割非互补链的 14 位, 从而形成 8 nt 的长黏性末端。基于这一特性, 研究者开发了 CCTL (Cpf1-assisted Cutting and Taq DNA ligase-assisted Ligation) 方法用于大 DNA 片段的体外编辑。利用 17 nt 长度的 crRNA spacer, 研究人员成功将放线菌紫红素合成基因簇 *act II -orf4* 基因的启动子进行了原位替换, 效率达到 70% 以上^[61]。进一步, 研究者对 Cpf1 进行了改造, 使其可以识别 60 种不同的 PAM 位点, 将其靶向范围扩大了 4 倍, 大大增加了 Cpf1 可以靶向的序列范围^[62]。

2.2 高通量蛋白表达

酶蛋白表达是一个复杂、高成本和耗时的过程。蛋白质表达的底盘系统包括原核细胞 (如大肠杆菌和枯草芽孢杆菌)、真核细胞 (如酵母、昆虫细胞和哺乳动物细胞) 以及无细胞 (cell-free) 蛋白合成系统。本文重点介绍大肠杆菌表达系统和无细胞表达系统。

大肠杆菌表达系统是经典的重组蛋白表达系统, 具有操作简单快捷、培养周期短、成本低廉、遗传背景清楚等优势, 使其成为研究者首选的蛋白表达系统^[63]。传统的大肠杆菌蛋白表达、纯化和检测的实验操作都由研究人员亲自动手完成, 不仅烦琐, 还耗费大量的等待时间。随着自动化技术成本的降低以及智能程度的提高, 采用自动化机器替代人工, 使得大肠杆菌蛋白表达过程可

高通量、自动化地进行, 而且操作更精准、更标准化。目前, 国内外已搭建起若干个全自动高通量筛选平台。例如, 德国格赖夫斯瓦尔德大学拥有一套全自动化高通量筛选平台用于大规模的蛋白质工程筛选。它以 96 微孔板等作为实验载体, 通过自动化操作系统 Momentum 和 Agilent VWorks 9 软件对实验过程进行编程, 自动化进行重组质粒转化、细胞培养、蛋白质表达与纯化以及蛋白质性质表征等^[64]。

此外, 无细胞蛋白质合成体系逐渐发展为一项快速、高效的体外合成蛋白质的技术手段。该技术是以外源的 DNA 或 mRNA 为模板, 通过补充底物和能量物质, 在细胞抽提物提供的酶系作用下完成蛋白质的体外表达^[65-66]。细胞抽提物可以来源于大肠杆菌、兔网织红细胞、昆虫细胞和哺乳动物细胞等。该系统能够以 PCR 产物作为线性模板, 因此无需烦琐的克隆、连接、转化、细胞裂解和蛋白质提取等步骤。且由于该系统不存在活细胞, 可用来表达在胞内系统中难于表达的毒性蛋白质。传统的胞内蛋白质表达系统, 从基因克隆、质粒转化、宿主菌培养、目标蛋白质表达和蛋白质分离纯化, 一般需要 2~3 周左右的时间。无细胞蛋白质合成体系, 只需要简单的基因扩增、无细胞蛋白质表达以及简单的分离纯化等步骤, 整个过程一般只需要 3~4 天, 表达周期大大缩短并节约相应的经济成本^[67]。例如, Nakano 等^[68]开发了一种完全在体外进行的蛋白突变体库构建与筛选的高通量方法。利用无细胞蛋白质合成体系、PCR 体外扩增、微孔板反应并结合 ELISA 筛选, 对抗人血清蛋白的单链可变片段进行了体外筛选, 成功筛选到具有抗体结合功能的蛋白质片段。商业化的 96 孔和 384 孔板是最为常用的较高通量的实验载体, 能方便地使用与之配套的酶标仪等仪器进行检测。然而, 由于受到微孔板孔数的限制, 商业化微孔板中的无细胞蛋白表达难以满足通量更大的研究, 因此, Angenendt 等^[69]定制了一种孔最大容量为 1.5 μL 的 1536 孔规格的微孔板。进一步, 利用绿色荧光蛋白与 β -半乳糖苷酶作为检测蛋白, 研究者成功将微孔板中无细胞反应体系缩小到 100 nL, 而且反应体系的浓度稀释 10 倍后反应依然能进行。

表达不同来源的酶蛋白时，通常需要针对DNA/氨基酸序列、助溶融合标签、调控元件强度、细胞培养条件等进行系统性、组合式优化；因此，需要开发高通量方法评估不同实验条件下酶的可溶性表达结果。聚丙烯酰胺凝胶电泳（polyacrylamide gel electrophoresis, PAGE）和蛋白质印迹法（Western blot）是检测蛋白质是否可溶性表达的常规方法；但其过程耗时烦琐，很难进行高通量操作。Split GFP技术提供了一种操作简便、易于自动化的可溶性表达检测方法：将GFP11片段（约15个氨基酸）融合表达在目标蛋白质的氮端或者碳端；改造后的目标蛋白质与包含GFP其余序列的重组片段（GFP1~10，约200个氨基酸）在体内或体外环境相遇时，可以结合产生绿色荧光，且荧光强度与目标蛋白质的可溶表达量呈正相关^[70-71]。这一技术为快速筛选不同实验设计方案，从而优化酶元件在底盘体系中的可溶性表达提供了基础。

2.3 酶催化功能的高通量表征

2.3.1 酶反应分析方法

酶促反应动力学（enzyme kinetics）分析是对酶蛋白进行功能表征的核心手段。光谱、质谱和电化学方法是高通量酶动力学分析的常用方法。对于少数反应物具有特征光学性质的酶促反应而言，可以通过监测吸光度、荧光等信号跟踪反应进程。然而，大多数生化反应的底物或产物不产生特征的光学信号，需要通过设计酶反应分析方法（enzyme assay），如利用偶联反应、生物传感器等手段将反应物的浓度信息转化成仪器可检测的光、电等信号。例如，琥珀酸辅酶A合成过程中伴随着ATP的水解反应，形成ADP和正磷酸盐，研究者可以利用钼酸与正磷酸盐生成蓝色颜料来定量琥珀酸辅酶A合成酶的效率^[72]。另一方面，生物传感器可以利用转录因子（transcriptional factor, TF）、工程荧光蛋白、核糖核酸适配子（RNA aptamer）等功能元件识别细胞内特定代谢物，并进一步通过基因调控将目标分子的浓度信息定量转化为报告基因的表达量^[73-76]。

电化学传感器可检测电极表面生化反应导致

的电流变化，具有很高的检测灵敏度和选择性，并且可进行微型化处理。该类型的传感器主要由生物识别元件以及可将生物信号转换成电化学信号的传感器组成。研究者开发了具有96个丝网印刷电极（96 screen-printed electrodes）的系统，用于分析半乳糖氧化酶产生的 H_2O_2 ，对酶反应的检测更加灵敏、快速^[77]。电极材料是影响传感器的主要因素，根据电极材料的不同分为电流计、电位计、电导计和阻抗计。目前新型的电化学传感器采用纳米材料，如碳纳米管、金属和金属氧化物纳米颗粒、硅纳米颗粒和半导体材料纳米颗粒等^[78]。

与光学、电化学等分析方法不同，质谱（mass spectrometry, MS）基于离子的质荷比（ m/z ），可以对反应物进行无标记（label-free）的定性与定量测定，对于酶反应分析具有更好的普适性。但是，检测样品进入质谱之前需要经过耗时的色谱分离，限制了质谱筛选的通量。目前，基于新型质谱仪器设计，通过激光、微流控或声学技术将分析物直接引入质谱仪，可以在几秒钟内完成单个样品的分析。例如，基质辅助激光解析解离质谱（matrix-assisted laser desorption/ionization MS, MALDI MS）可以利用激光取样，快速分析靶板表面的酶反应阵列，具有样品制备简单、高耐盐以及广泛的生物分子覆盖性等特点^[79]。de Rond等^[80]基于MALDI MS开发了PECAN方法，完成了P450_{BM3}突变库的高通量筛选鉴定。电喷雾质谱（electrospray ionization MS, ESI MS）与微流控液滴系统联用可以进行高通量的分析与分选，在小分子分析方面相较MALDI更具优势，该技术已经应用于蛋白质工程、药物开发和诊断等研究工作^[81]。另外，基于质谱高分辨率的特点，可以开发代谢组学方法，同时对数百种代谢物进行分析^[82]。基于这一理论，将候选酶与细胞粗提物共同孵育，通过监测反应前后的质谱信号变化，可从大量代谢物中快速鉴定反应底物与产物；进一步结合化学计量学和数据库分析，实现酶的高通量筛选与功能鉴定^[83]。

2.3.2 高通量样品处理与分析技术

基于以上酶反应分析手段，可以利用自动化移液、流式分选和微流控液滴等技术提高酶功能

表征的通量和准确性。

自动化移液工作站通常由一个工作台面、移液机械臂、抓手机械臂、相关功能模块和配套电脑构成。常规的移液工作站有8通道或96通道的移液机械臂,可以实现96孔板以及384孔板的移液操作。通过合理的程序设定,可实现短时间内全自动处理大规模液体生物样本,有效提高实验的准确性、稳定性和效率。此外,在工艺流程中整合菌落挑取仪、酶标仪等功能模块,可以赋予移液工作站更加丰富、个性化的用途。例如,前述德国格赖夫斯瓦尔德大学平台通过整合移液工作站和酶标仪,发展了高通量酶活检测方法,实现了单加氧酶、转氨酶、脱卤素酶和酰基转移酶等文库大规模筛选^[84]。

细胞为酶促反应提供天然环境,并将酶蛋白与其编码基因进行物理偶联;结合流式细胞荧光分选(fluorescence-activated cell sorting, FACS)以及二代测序技术,可以快速建立酶序列-功能间的对应关系。对于细胞内代谢物或具有渗透性的胞外产物,可以将酶反应与荧光蛋白的表达、折叠或运输过程偶联^[85];针对非渗透性底物,可以利用表面展示技术将酶蛋白表达在细胞表面,并利用分子互作将反应物精巧固定于细胞表面,从而基于荧光探针和反应物的结合强度进行酶活筛选^[86]。

对于胞外游离产物,无法利用传统FACS技术进行分析,需要结合微流控液滴技术对单细胞及其周边微环境进行分析^[87]。每个细胞被分装在微流体元器件中以每秒数千滴的速度产生的独立水油小液滴中,其体积大小由通道尺寸和流体流速控制,通常从纳升到皮升不等^[88]。集成微流体分选仪可在 10^3 Hz频率下筛选高荧光液滴,并根据用户定义的分选标准施加电场,将包含单细胞的液滴转移到收集室或废物室中。微流控芯片具有高灵敏度、定量读出和高准确性等优点,通过液滴注射或液滴融合可以实现多步骤酶催化反应^[85]。结合先进的荧光标记分选技术、拉曼光谱和质谱等,可以显著提高基于液滴的微流体系统的筛选效率。此外,微流控技术也可应用于宏基因组中催化元件的高通量挖掘^[89]。

3 高通量酶元件挖掘在酶家族研究中的应用

酶家族(enzyme family)是由序列、结构、功能、进化方面具有相似性的蛋白质组成的集合。随着高通量(宏)基因组测序技术的发展,数据库中蛋白质序列的数量呈指数增长,对功能注释和预测提出了重大挑战;而对酶家族特征信息的提取,有助于充分利用已有实验数据对未知序列的功能进行归属^[90]。利用合成生物学手段系统化探索酶家族的序列-功能关系,是规模化挖掘新型酶元件的一个重要途径。

综合利用聚类分析和实验验证优先化算法,有利于提高对酶家族进行高通量挖掘的效率。卤代烷烃脱卤酶(haloalkane dehalogenases, HLD)采用水解型脱卤机制催化卤代有机化合物的碳-卤键的断裂,在环境污染物降解、手性化合物合成、分子成像等方面具有应用潜力。利用Position-Specific Iterated (PSI)-BLAST比对, Vanacek等从NCBI核酸测序数据库中识别出5661个可能的HLD^[18]。为了从数千个候选序列中快速筛选出最为相关的HLD,研究者首先采用了类似SSN的成对同源比对聚类方法,初步筛选出953个较为可信的HLD序列。接着,基于多序列比对去除了117个不完整序列和178个氨基酸水平的同义序列。对剩余的658个HLD序列,研究者制定了优先化原则——序列多样性、物种来源/生境多样性、活性位点结构多样性、与已知HLD低同源性和高可溶性等,综合利用同源结构建模、可溶性预测等生信工具,研究者选择了20个HLD蛋白进行实验表征,并最终实现了8种HLD在大肠杆菌底盘中的可溶性表达(图2)。结果显示,这8种HLD来源于不同的细菌、真核生物和古菌,表现出非常多样化的底物范围和特异性、最适温度范围($20\sim 70^{\circ}\text{C}$)和最适pH范围($5.7\sim 10$),并且发现了迄今催化活性最高的一个HLD蛋白^[18]。这项研究表明,综合利用聚类和优先化算法,有利于从酶家族中快速挖掘具有新颖功能和性能的酶元件。

许多酶家族成员虽然具有相似功能,但在序列和结构水平上具有多样性^[91]。例如,胞质谷胱甘肽转移酶(cytosolic glutathione transferase,

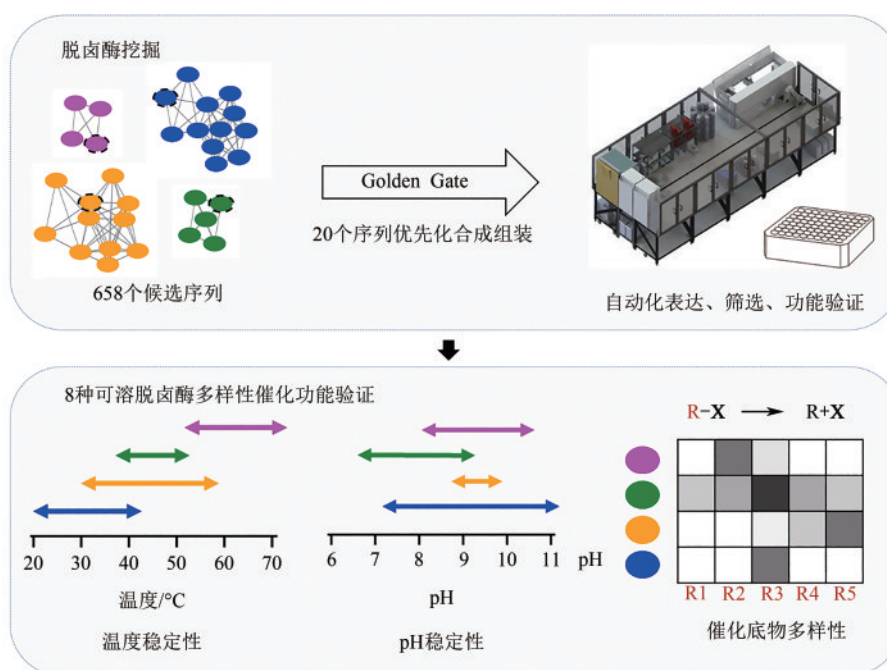


图2 脱卤酶挖掘、优先化及多样性催化功能表征

Fig. 2 Sequence analysis, prioritization and characterization for haloalkane dehalogenases (HLDs)

cytGST) 在新陈代谢和防御氧化损伤中具有关键作用；其超家族包含 13000 多个非冗余序列，催化多种反应类型（具有超过 140 个 Enzyme commission 编号），但催化机理并不完全清楚。为了系统性研究 cytGST 酶家族，Mashiyama 等^[92] 首先利用 Cd-hit 算法^[93] 将数据库中的 13493 个 cytGST 按照 50% 的序列同一性（50% sequence identity, ID₅₀）进行聚类，提取出 2190 个具有代表性的序列；针对这 2190 个 ID₅₀ 节点，研究者对衡量序列相似性的 *E*-value 进行了步进优化，并最终选定 1×10^{-25} 的 *E*-value 阈值构建 SSN 网络，包含约 30 个主要的簇和超过 60 个较小的簇或单个节点，其中有 23 个簇中至少有一个成员被实验证实有 GST 活性。在聚类分析的指导下，研究者从不同的簇或单个节点中优先选择了 857 个候选基因进行实验验证。高通量分子克隆、蛋白质表达与纯化、结晶等实验在纽约 SGX 结构基因组学研究中心（New York SGX Research Center for Structural Genomics）开展；利用自动化移液工作站等高通量手段，该研究中心可以每月制备 120 个以上的超纯蛋白质样品，每年解析 100~200 个蛋白质晶体结构^[94]。最后，共有 230 个候选序列在大肠杆菌中被成功表达纯化，其中 27 个蛋白质的 37 个 3D

晶体结构得到解析。通过筛选 175 个不同的底物，研究者发现了 82 个新的具有 GST 活性的酶，将具有 GST 活性的簇从原来已知的 23 个增加至 35 个，并发现少数成员具有非常罕见的还原性脱卤活性。研究者将新发现的蛋白质与文献报道已知具有 GST 活性的 174 个蛋白质综合分析，将其序列、结构、催化机理等信息与 SSN 网络的结构进行映射，用以生成序列-结构-功能关系的全局视图。结果表明，53% 的 cytGST 具有高度底物专一性，只与 1 个底物发生反应；而 7% 的酶可以催化至少 6 个底物的转化。作者还发现，如果通过催化机理而非序列或结构相似性建立相关网络，可以将 cytGST 家族已知 15 个催化功能中的 14 个联系起来；这一观察也再次印证了仅靠序列或结构同源性无法准确预测功能相似性的现象。这一研究是利用合成生物学方法规模化挖掘酶蛋白的经典案例，证明了采取多层次聚类分析的必要性，以及酶挖掘过程中高通量实验验证的不可替代性^[92]。

4 展 望

近年来，数据库中蛋白序列的指数性增长以

及生命科学前沿技术的快速发展,为人们提供了丰富的生物资源。如何充分利用全球共享的生物资源数据,对数据库中的酶资源进行高通量的挖掘利用,是研究者面临的重大机遇和挑战。综合上文所述,研究者开发了一系列算法、数据库、高通量实验技术,应用于酶蛋白的高通量挖掘,有效地推动了酶制剂和细胞工厂在生物制造方面的应用^[95]。

但是,目前仍然缺乏高度集成的工程化研究平台,应用合成生物学理念和技术对酶蛋白资源进行系统化的研究和挖掘。近年来,在全球范围内已建成或在建多个大型合成生物学研究基础设施,将自动化技术应用在合成生物学“设计—构建—测试—学习”的各个研究环节。这些基础设施包括美国伊利诺伊大学的 iBioFAB (Illinois Biological Foundry for Advanced Biomanufacturing)、美国能源部的 Agile Biofoundry、英国爱丁堡大学的 Edinburgh Genome Foundry (EGF) 等。2019年,8个国家的16个隶属于公共研究机构的合成生物设施于2019年成立了“全球合成生物设施联盟”(Global Biofoundry Alliance, GBA)^[96],旨在加强设施间的协作沟通,将智能制造的理念引入合成生物学。

依托这些设施平台,研究人员开发了不同程度的自动化流程进行酶蛋白挖掘。例如,本文作者在伊利诺伊大学赵惠民教授团队参与设计、搭建、运行了学术界首个全自动合成生物设施 iBioFAB^[97],开发了基于 Golden Gate assembly 的自动化 DNA 组装方法^[51],实现了大肠杆菌和酿酒酵母的自动化转化、培养、筛选等操作^[98],建立了自动化样品前处理流程对代谢物进行快速质谱表征^[99],从而对酶蛋白、代谢通路、细胞工厂等合成生物体系的高通量工程构建与优化;中国科学院天津工业生物研究所团队基于自动化平台,实现了谷氨酸棒状杆菌的高通量基因编辑^[100];如前所述,德国格赖夫斯瓦尔德大学开发了酶突变库自动化构建和筛选的流程,实现了单加氧酶、转氨酶、脱卤素酶等酶蛋白的定向进化研究^[85]。另外,中国科学院深圳先进技术研究院正在牵头建设“深圳合成生物研究重大科技基础设施”,建成后将是我国首个将软件控制、硬件设备和合成

生物学应用进行整合的大型规模化合成生物制造系统,作为开放共享平台服务我国合成生物学的科研与产业用户。通过实现生物设计、工程 DNA 合成与组装、底盘系统转化与培养、催化功能测试等环节的自动化运行,这些工程化平台可以高效集成酶蛋白高通量挖掘所需的数据库、算法、合成生物工艺和硬件仪器设备,从而实现优质酶催化元件的规模化挖掘,积累酶序列-功能关系的高质量定量数据,为实现酶蛋白的理性设计奠定基础。

在未来的研究中,基于合成生物学的酶元件挖掘研究可在药物、精细化工等高附加值分子的研究与生产等热点、难点领域发挥重要的作用。例如,参与内源性物质和药物、环境化合物等外源物质代谢的 P450 酶及其抑制剂的挖掘、筛选、功能鉴定^[101];非天然和较难获得的化合物的从头生物合成,包括对未知催化途径酶的挖掘与改造、生物途径及化学途径的整合、底盘细胞的改造、人工合成途径与底盘细胞的适配等方面^[35]。而依托工程化合成生物研究基础设施,可以计算设计并合成表征催化不同反应类型、适应不同实验条件的酶蛋白序列,从而建设包含功能特性清晰、符合组装标准的元件实体库。为了实现以上愿景,需要将生信分析与实验表征深度整合,通过界面友好的数据库、算法与网站,实现酶元件及其表征数据的查询、比对与二次计算,从而推荐已有元件或待合成表征的对象;开发自动合成生物实验技术,开展自动化高通量的元件挖掘、添加、存储和利用等操作,形成酶催化数据的信息化体系与元件共享平台。未来,相信合成生物学提供的工程化思想和能力可以大幅提高酶蛋白挖掘研究的通量与效率,从而加速生物制造理念的工业化实现。

参 考 文 献

- [1] ZAPARUCHA A, DE BERARDINIS V, VAXELAIRE-VERGNE C. Genome mining for enzyme discovery [M]// Modern biocatalysis: advances towards synthetic biological systems. The Royal Society of Chemistry, 2018: 1-27.
- [2] WANG L, DASH S, NG C Y, et al. A review of computational tools for design and reconstruction of metabolic pathways

- [J]. *Synthetic and Systems Biotechnology*, 2017, 2(4): 243-252.
- [3] HATZIMANIKATIS V, LI CH, IONITA J A, et al. Exploring the diversity of complex metabolic networks [J]. *Bioinformatics*, 2005, 21(8): 1603-1609.
- [4] KUMAR A, WANG L, NG C Y, et al. Pathway design using de novo steps through uncharted biochemical spaces [J]. *Nature Communications*, 2018, 9: 184-201.
- [5] MAK W S, TRAN S, MARCHESCHI R, et al. Integrative genomic mining for enzyme function to enable engineering of a non-natural biosynthetic pathway[J]. *Nature Communications*, 2015, 6(1).
- [6] CARTER M S, ZHANG X, HUANG H, et al. Functional assignment of multiple catabolic pathways for D-apiose [J]. *Nature Chemical Biology*, 2018, 14(7): 696-705.
- [7] MORI Y, SHIRAI T. Designing artificial metabolic pathways, construction of target enzymes, and analysis of their function [J]. *Current Opinion in Biotechnology*, 2018, 54: 41-44.
- [8] LI Y R, LI S J, THODEY K, et al. Complete biosynthesis of noscapine and halogenated alkaloids in yeast [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2018, 115(17): E3922-E3931.
- [9] THODEY K, GALANIE S, SMOLKE C D. A microbial biomanufacturing platform for natural and semisynthetic opioids [J]. *Nature Chemical Biology*, 2014, 10(10): 837-844.
- [10] SAYERS E W, AGARWALA R, BOLTON E E, et al. Database resources of the National Center for Biotechnology Information [J]. *Nucleic Acids Research*, 2019, 47(D1): D23-D28.
- [11] TIAN W D, ARAKAKI A K, SKOLNICK J. EFICaz: a comprehensive approach for accurate genome-scale enzyme function inference [J]. *Nucleic Acids Research*, 2004, 32(21): 6226-6239.
- [12] CLARK K, KARSCH-MIZRACHI I, LIPMAN D J, et al. GenBank [J]. *Nucleic Acids Research*, 2016, 44(D1): D67-D72.
- [13] SEFFERNICK J L, DE SOUZA M L, SADOWSKY M J, et al. Melamine deaminase and atrazine chlorohydrolase: 98 percent identical but functionally different [J]. *Journal of Bacteriol*, 2001, 183(8): 2405-2410.
- [14] BURROUGHS A M, ALLEN K N, DUNAWAY-MARIANO D, et al. Evolutionary genomics of the HAD superfamily: Understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes [J]. *Journal of Molecular Biology*, 2006, 361(5): 1003-1034.
- [15] GLASNER M E, FAYAZMANESH N, CHIANG R, et al. Evolution of structure and function in the *o*-succinylbenzoate synthase family [J]. *Faseb Journal*, 2006, 20(5): A905-A905.
- [16] PETERHOFF D, BEER B, RAJENDRAN C, et al. A comprehensive analysis of the geranylgeranyl glyceryl phosphate synthase enzyme family identifies novel members and reveals mechanisms of substrate specificity and quaternary structure organization [J]. *Molecular Microbiology*, 2014, 92(4): 885-899.
- [17] HEINS R A, CHENG X, NATH S, et al. Phylogenomically guided identification of industrially relevant GH1 beta-glucosidases through DNA synthesis and nanostructure-initiator mass spectrometry [J]. *ACS Chemical Biology*, 2014, 9(9): 2082-2091.
- [18] VANACEK P, SEBESTOVA E, BABKOVA P, et al. Exploration of enzyme diversity by integrating bioinformatics with expression analysis and biochemical characterization [J]. *ACS Catalysis*, 2018, 8(3): 2402-2412.
- [19] BAIROCH A, APWEILER R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999 [J]. *Nucleic Acids Research*, 1999, 27(1): 49-54.
- [20] MARCHLER-BAUER A, DERBYSHIRE M K, GONZALES N R, et al. CDD: NCBI's conserved domain database [J]. *Nucleic Acids Research*, 2015, 43(D1): D222-D226.
- [21] FINN R D, BATEMAN A, CLEMENTS J, et al. Pfam: the protein families database [J]. *Nucleic Acids Research*, 2014, 42(D1): D222-D230.
- [22] SILLITOE I, LEWIS T E, CUFF A, et al. CATH: comprehensive structural and functional annotations for genome sequences [J]. *Nucleic Acids Research*, 2015, 43(D1): D376-D381.
- [23] MEYER F, OVERBEEK R, RODRIGUEZ A. FIGfams: yet another set of protein families [J]. *Nucleic Acids Research*, 2009, 37(20): 6643-6654.
- [24] SCHNOES A M, BROWN S D, DODEVSKI I, et al. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies [J]. *PLoS Computational Biology*, 2009, 5(12): e1000605.
- [25] COPP J N, ANDERSON D W, AKIVA E, et al. Exploring the sequence, function, and evolutionary space of protein superfamilies using sequence similarity networks and phylogenetic reconstructions [J]. *Methods in Enzymology*, 2019, 620: 315-347.
- [26] RODRÍGUEZ BENÍTEZ A, NARAYAN A R H. Frontiers in biocatalysis: profiling function across sequence space [J]. *ACS Central Science*. 2019, 5(11): 1747-1749.
- [27] ATKINSON H J, MORRIS J H, FERRIN T E, et al. Using sequence similarity networks for visualization of relationships across diverse protein superfamilies [J]. *PLoS One*, 2009, 4(2): e4345-e4358.
- [28] BERMAN H, HENRICK K, NAKAMURA H, et al. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data [J]. *Nucleic Acids Research*, 2007, 35: D301-D303.
- [29] ASHBURNER M, BALL C A, BLAKE J A, et al. Gene Ontology: tool for the unification of biology [J]. *Nature Genetics*, 2000,

- 25(1): 25-29.
- [30] COPP J N, AKIVA E, BABBIT P C, et al. Revealing unexplored sequence-function space using sequence similarity networks [J]. *Biochemistry*, 2018, 57(31): 4651-4662.
- [31] BARBER A E, BABBITT P C. Pythoscape: a framework for generation of large protein similarity networks [J]. *Bioinformatics*, 2012, 28(21): 2845-2846.
- [32] ZALLOT R, OBERG N, GERLT J A. The EFI web resource for genomic enzymology tools: leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways [J]. *Biochemistry*, 2019, 58(41): 4169-4182.
- [33] GETZ G, STAROVOLSKY A, DOMANY E. F2CS: FSSP to CATH and SCOP prediction server [J]. *Bioinformatics*, 2004, 20(13): 2150-2152.
- [34] SILLITOE I, FURNHAM N. FunTree: advances in a resource for exploring and contextualising protein function evolution [J]. *Nucleic Acids Research*, 2016, 44(D1): D317-323.
- [35] LIN G M, WARDEN-ROTHMAN R, VOIGT C A. Retrosynthetic design of metabolic pathways to chemicals not found in nature [J]. *Current Opinion in Systems Biology*, 2019, 14: 82-107.
- [36] KLUCZNIK T, MIKULAK-KLUCZNIK B, MCCORMACK M P, et al. Efficient syntheses of diverse, medically relevant targets planned by computer and executed in the laboratory [J]. *Chemistry*, 2018, 4(3): 522-532.
- [37] MUSIL M, KONEGGER H, HON J, et al. Computational design of stable and soluble biocatalysts [J]. *ACS Catalysis*, 2019, 9(2): 1033-1054.
- [38] WILKINSON D L, HARRISON R G. Predicting the solubility of recombinant proteins in *Escherichia coli* [J]. *Nature Biotechnology*, 1991, 9(5): 443-448.
- [39] KIM J H, KERSHNER J P, NOVIKOV Y, et al. Three serendipitous pathways in *E. coli* can bypass a block in pyridoxal-5'-phosphate synthesis [J]. *Molecular Systems Biology*, 2010, 6: 436-448.
- [40] JEFFRYES J G, COLASTANI R L, ELBADAWI-SIDHU M, et al. MINEs: open access databases of computationally predicted enzyme promiscuity products for untargeted metabolomics [J]. *Journal of Cheminformatics*, 2015, 7: 44-51.
- [41] PERTUSI D A, MOURA M E, JEFFRYES J G, et al. Predicting novel substrates for enzymes with minimal experimental effort with active learning [J]. *Metabolic Engineering*, 2017, 44: 171-181.
- [42] EKINS S. Predicting undesirable drug interactions with promiscuous proteins in silico [J]. *Drug Discov Today*, 2004, 9(6): 276-285.
- [43] FERRARIO V, SIRAGUSA L, EBERT C, et al. BioGPS descriptors for rational engineering of enzyme promiscuity and structure based bioinformatic analysis [J]. *PLoS One*, 2014, 9(10): e109354.
- [44] CAMPODONICO M A, ANDREWS B A, ASENJO J A, et al. Generation of an atlas for commodity chemical production in *Escherichia coli* and a novel pathway prediction algorithm, GEM-Path [J]. *Metabolic Engineering*, 2014, 25: 140-158.
- [45] FEHER T, PLANSON A G, CARBONELL P, et al. Validation of retroPath, a computer-aided design tool for metabolic pathway engineering [J]. *Biotechnology Journal*, 2014, 9(11): 1446-1457.
- [46] CHAO R, YUAN Y B, ZHAO H M. Recent advances in DNA assembly technologies [J]. *Fems Yeast Research*, 2015, 15(1): 1-9.
- [47] GOULD N, HENDY O, PAPAMICHAIL D. Computational tools and algorithms for designing customized synthetic genes [J]. *Frontiers in Bioengineering and Biotechnology*, 2014, 2: 41.
- [48] JIA B, JEON C O. High-throughput recombinant protein expression in *Escherichia coli*: current status and future perspectives [J]. *Open Biology*, 2016, 6(8): 160196.
- [49] NAGASE T, YAMAKAWA H, TADOKORO S, et al. Exploration of human ORFeome: high-throughput preparation of ORF clones and efficient characterization of their protein products [J]. *DNA Research*, 2008, 15(3): 137-149.
- [50] HAMEDIRAD M, WEISBERG S, CHAO R, et al. Highly efficient single-pot scarless golden gate assembly [J]. *ACS Synthetic Biology*, 2019, 8(5): 1047-1054.
- [51] CHAO R, LIANG J, TASAN I, et al. Fully automated one-step synthesis of single-transcript TALEN pairs using a biological foundry [J]. *ACS Synthetic Biology*, 2017, 6(4): 678-685.
- [52] HILLSON N J, ROSENGARTEN R D, KEASLING J D. j5 DNA assembly design automation software [J]. *ACS Synthetic Biology*, 2012, 1(1): 14-21.
- [53] KANIGOWSKA P, SHEN Y, ZHENG Y J, et al. Smart DNA fabrication using sound waves: applying acoustic dispensing technologies to synthetic biology [J]. *Jala-J. Lab. Autom*, 2016, 21(1): 49-56.
- [54] LINSHIZ G, STAWSKI N, GOYAL G, et al. PR-PR: cross-platform laboratory automation system [J]. *ACS Synthetic Biology*, 2014, 3(8): 515-524.
- [55] CHEN W H, QIN Z J, WANG J, et al. The MASTER (methylation-assisted tailorable ends rational) ligation method for seamless DNA assembly [J]. *Nucleic Acids Research*, 2013, 41(8): e93.
- [56] SHAO Z Y, ZHAO H, ZHAO H M. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways [J]. *Nucleic Acids Research*, 2009, 37(2): e16.
- [57] LANDENMARK H K, FORGAN D H, COCKELL C S. An estimate of the total DNA in the biosphere [J]. *PLoS Biology*, 2015, 13(6): e1002168.
- [58] SHAPLAND E B, HOLMES V, REEVES C D, et al. Low-cost, high-throughput sequencing of DNA assemblies using a highly multiplexed nextera process [J]. *ACS Synthetic Biology*, 2015, 4(7): 860-866.

- [59] DHARMADI Y, PATEL K, SHAPLAND E, et al. High-throughput, cost-effective verification of structural DNA assembly [J]. *Nucleic Acids Research*, 2014, 42(4): e22.
- [60] De KOK S, STANTON L H, SLABY T, et al. Rapid and reliable DNA assembly *via* ligase cycling reaction [J]. *ACS Synthetic Biology*, 2014, 3(2): 97-106.
- [61] LEI C, LI S Y, LIU J K, et al. The CCTL (Cpf1-assisted cutting and taq DNA ligase-assisted ligation) method for efficient editing of large DNA constructs *in vitro* [J]. *Nucleic Acids Research*, 2017, 45(9): e74.
- [62] WANG L, WANG H, LIU H, et al. Improved CRISPR-Cas12a-assisted one-pot DNA editing method enables seamless DNA editing [J]. *Biotechnology and Bioengineering*, 2019, 116(6): 1463-1474.
- [63] DEMAİN A L, VAISHNAV P. Production of recombinant proteins by microbes and higher organisms [J]. *Biotechnology Advances*, 2009, 27(3): 297-306.
- [64] SCHMIDT S, DÖRR M, BORNSCHEUER U T. Library growth and protein expression: optimal and reproducible microtiter plate expression of recombinant enzymes in *E. coli* using MTP shakers [M]// BORNSCHEUER U T, HÖHNE M. *Protein engineering: methods and protocols*. New York: Springer New York, 2018: 145-156.
- [65] NAKANO H, YAMANE T. Cell-free protein synthesis systems [J]. *Biotechnology Advances*, 1998, 16(2): 367-384.
- [66] MARKEL U, ESSANI K D, BESIRLIOGLU V, et al. Advances in ultrahigh-throughput screening for directed enzyme evolution [J]. *Chemical Society Reviews*, 2020, 49(1): 233-262.
- [67] SWARTZ J. Developing cell-free biology for industrial applications [J]. *Journal of Microbiology and Biotechnology*, 2006, 33(7): 476-485.
- [68] RUNGPRAGAYPHAN S, NAKANO H, Yamane T. PCR-linked *in vitro* expression: a novel system for high-throughput construction and screening of protein libraries [J]. *FEBS Letters*, 2003, 540(1-3): 147-150.
- [69] ANGENENDT P, NYARSIK L, SZAFLARSKI W, et al. Cell-free protein expression and functional assay in nanowell chip format [J]. *Analytical Chemistry*, 2004, 76(7): 1844-1849.
- [70] SANTOS-ABERTURAS J, DÖRR M, BORNSCHEUER U T. Normalized screening of protein engineering libraries by split-GFP crude cell extract quantification [M]// BORNSCHEUER U T, HÖHNE M. *Protein engineering: methods and protocols*. New York: Springer New York; 2018: 157-170.
- [71] CABANTOUS S, WALDO G S. *In vivo* and *in vitro* protein solubility assays using split GFP [J]. *Nature Methods*, 2006, 3(10): 845-854.
- [72] GUO C X, HU Y L, YANG C Y, et al. Developing a colorimetric assay for Fe(II)/2-oxoglutarate-dependent dioxygenase [J]. *Analytical Biochemistry*, 2018, 548: 109-114.
- [73] DEKKER L, POLIZZI K M. Sense and sensitivity in bioprocessing-detecting cellular metabolites with biosensors [J]. *Current Opinion in Chemical Biology*, 2017, 40: 31-36.
- [74] CHEN X, ZHANG D, SU N, et al. Visualizing RNA dynamics in live cells with bright and stable fluorescent RNAs [J]. *Nature Biotechnology*, 2019, 37(11): 1287-1293.
- [75] LIM H G, JANG S, JANG S, et al. Design and optimization of genetically encoded biosensors for high-throughput screening of chemicals [J]. *Current Opinion in Chemical Biology*, 2018, 54: 18-25.
- [76] ZENG W, GUO L, XU S, et al. High-throughput screening technology in industrial biotechnology [J]. *Trends in Biotechnology*, 2020. doi:10.1016/j.tibtech.2020.01.001.
- [77] AYMARD C, BONAVENTURA C, HENKENS R, et al. High-throughput electrochemical screening assay for free and immobilized oxidases: electrochemiluminescence and intermittent pulse amperometry [J]. *Chemelectrochem*, 2017, 4(4): 957-966.
- [78] MAJDINASAB M, MITSUBAYASHI K, MARTY J L. Optical and electrochemical sensors and biosensors for the detection of quinolones [J]. *Trends in Biotechnology*. 2019, 37(8): 898-915.
- [79] SI T, LI B, COMI T J, et al. Profiling of microbial colonies for high-throughput engineering of multistep enzymatic reactions *via* optically guided matrix-assisted laser desorption/ionization mass spectrometry [J]. *Journal of the American Chemical Society*, 2017, 139(36): 12466-12473.
- [80] De ROND T, JIAN G, AMIN Z, et al. A high-throughput mass spectrometric enzyme activity assay enabling the discovery of Cytochrome P450 biocatalysts [J]. *Angewandte Chemie*, 2019, 131(30): 10220-10225.
- [81] HOLLAND-MORITZ D A, WISMER M K, MANN B F, et al. Mass activated droplet sorting (MADS) enables high-throughput screening of enzymatic reactions at nanoliter scale [J]. *Angewandte Chemie International Edition*, 2020, 59: 4470-4477.
- [82] PROSSER G A, LARROUY-MAUMUS G, De CARVALHO L P. Metabolomic strategies for the identification of new enzyme functions and metabolic pathways [J]. *EMBO Reports*, 2014, 15(6): 657-669.
- [83] SÉVIN D C, FUHRER T, ZAMBONI N, SAUER U. Nontargeted *in vitro* metabolomics for high-throughput identification of novel enzymes in *Escherichia coli* [J]. *Nature Methods*, 2017, 14(2): 187-194.
- [84] DÖRR M, FIBINGER M P C, LAST D, et al. Fully automatized high-throughput enzyme library screening using a robotic platform [J]. *Biotechnology and Bioengineering*, 2016, 113(7): 1421-1432.
- [85] LONGWELL C K, LABANIEH L, COCHRAN J R. High-throughput screening technologies for enzyme engineering [J]. *Current Opinion in Biotechnology*, 2017, 48: 196-202.
- [86] CHEN I, DORR B M, LIU D R. A general strategy for the evolution of bond-forming enzymes using yeast display [J]. *Proceedings of the National Academy of Sciences of the United States*

- of America, 2011, 108(28): 11399-11404.
- [87] QIN Y L, WU L, WANG J G, et al. A fluorescence-activated single-droplet dispenser for high accuracy single-droplet and single-cell sorting and dispensing [J]. *Analytical Chemistry*, 2019, 91(10): 6815-6819.
- [88] ZHU P A, WANG L Q. Passive and active droplet generation with microfluidics: a review [J]. *Lab On A Chip*, 2017, 17(1): 34-75.
- [89] COLIN P Y, KINTSES B, GIELEN F, et al. Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics [J]. *Nature Communication*, 2015, 6: 10008.
- [90] LOBB B, DOXEY A C. Novel function discovery through sequence and structural data mining [J]. *Current Opinion in Structural Biology*, 2016, 38: 53-61.
- [91] DAVIDSON R, BAAS B J, AKIVA E, et al. A global view of structure-function relationships in the tautomerase superfamily [J]. *Journal of Biological Chemistry*, 2018, 293(7): 2342-2357.
- [92] MASHIYAMA S T, MALABANAN M M, AKIVA E, et al. Large-scale determination of sequence, structure, and function relationships in cytosolic glutathione transferases across the biosphere [J]. *PLoS Biology*, 2014, 12(4): e1001843.
- [93] LI W, GODZIK A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences [J]. *Bioinformatics*, 2006, 22(13): 1658-1659.
- [94] KOBE B, GUSS M, HUBER T. Structural proteomics: high-throughput methods [M].//Totowa N J. *Methods in molecular biology*. London: Humana; Springer distributor. xxvi, 2008: 601 p.
- [95] VOSS H, HECK C A, SCHALLMEY M, et al. Database mining for novel bacterial beta-etherases, glutathione-dependent lignin-degrading enzymes [J]. *Applied and Environmental Microbiology*, 2020, 86(2): e02026-19.
- [96] HILLSON N, CADDICK M, CAI Y Z, et al. Building a global alliance of biofoundries [J]. *Nature Communications*, 2019, 10: 2040-2043.
- [97] CHAO R, MISHRA S, SI T, et al. Engineering biological systems using automated biofoundries [J]. *Metabolic Engineering*, 2017, 42: 98-108.
- [98] SI T, CHAO R, MIN Y H, et al. Automated multiplex genome-scale engineering in yeast [J]. *Nature Communications*, 2017, 8: 15187.
- [99] XUE P, SI T, MISHRA S, et al. A mass spectrometry-based high-throughput screening method for engineering fatty acid synthases with improved production of medium-chain fatty acids [J]. *Biotechnology and Bioengineering*, 2020, 117(7): 2131-2138.
- [100] WANG Y, LIU Y, LIU J, et al. MACBETH: Multiplex automated *Corynebacterium glutamicum* base editing method [J]. *Metabolic Engineering*, 2018, 47:200-210.
- [101] ZLOKARNIK G, GROOTENHUIS P D, WATSON J B. High throughput P450 inhibition screens in early drug discovery [J]. *Drug Discov Today*, 2005, 10(21): 1443-1450.

通讯作者: 司同(1987—),男,博士,研究员,研究方向为合成生物学。
E-mail: tong.si@siat.ac.cn



第一作者: 张建志(1988—),男,博士,助理研究员,研究方向为合成生物学、代谢工程。
E-mail: zhangjz@siat.ac.cn

