

特约评述

DOI: 10.12211/2096-8280.2022-024

基于无细胞体系的生物合成代谢模块设计、构建与快速途径原型

唐士茗, 胡纪元, 郑穗平, 韩双艳, 林影
(华南理工大学, 生物科学与工程学院, 广东 广州 510006)

摘要: 随着代谢工程及合成生物学技术的发展, 化学品高效生物合成与绿色制造成为可能。高效生物合成体系的设计与构建是绿色生物制造的核心, 其理论体系建立及关键技术突破, 将为实现绿色生物制造领域高效生产及资源与环境可持续发展提供有力支撑。本文借助代谢途径模块设计的案例, 探讨化合物生物合成过程中潜在通用模块设计原则、设计工具, 以及基于无细胞蛋白合成体系的代谢模块快速构建及测试的方法, 将突破生物合成途径多基因、多模块“设计-构建-测试”(Design-Build-Test cycle, DBT cycle) 高效循环迭代的技术瓶颈。结合机器学习方法的应用, 将使“设计-构建-测试”向“设计-构建-测试-学习”(Design-Build-Test-Learn cycle, DBTL cycle) 进一步延伸, 对高效合成模块的“精准-鲁棒性”设计与构建、推动合成生物学科学与技术发展具有重要意义。

关键词: 生物合成模块; 无细胞蛋白合成; 无细胞快速原型; 机器学习

中图分类号: Q816 **文献标志码:** A

Designing, building and rapid prototyping of biosynthesis module based on cell-free system

TANG Shiming, HU Jiyuan, ZHENG Suiping, HAN Shuangyan, LIN Ying

(School of Biology and Biological Engineering, South China University of Technology, Guangzhou 510006, Guangdong, China)

Abstract: Bio-based chemical production has drawn attention regarding the realization of a sustainable society. With the development of metabolic engineering and synthetic biology, it is possible to make more efficient biosynthesis and scale-up commercial production of useful metabolites by metabolic pathway modification. Usually, some microorganisms are utilized as platforms by increasing the expression of desired genes and/or decreasing, that of undesired genes. Precise control of natural metabolism is, however, still challenging due to the complicated regulatory architecture at the levels of transcription, translation, and post-translation. Hence, various strategies of design and construction of biosynthesis modules have been proposed to optimize and expedite the design-build-test cycles of

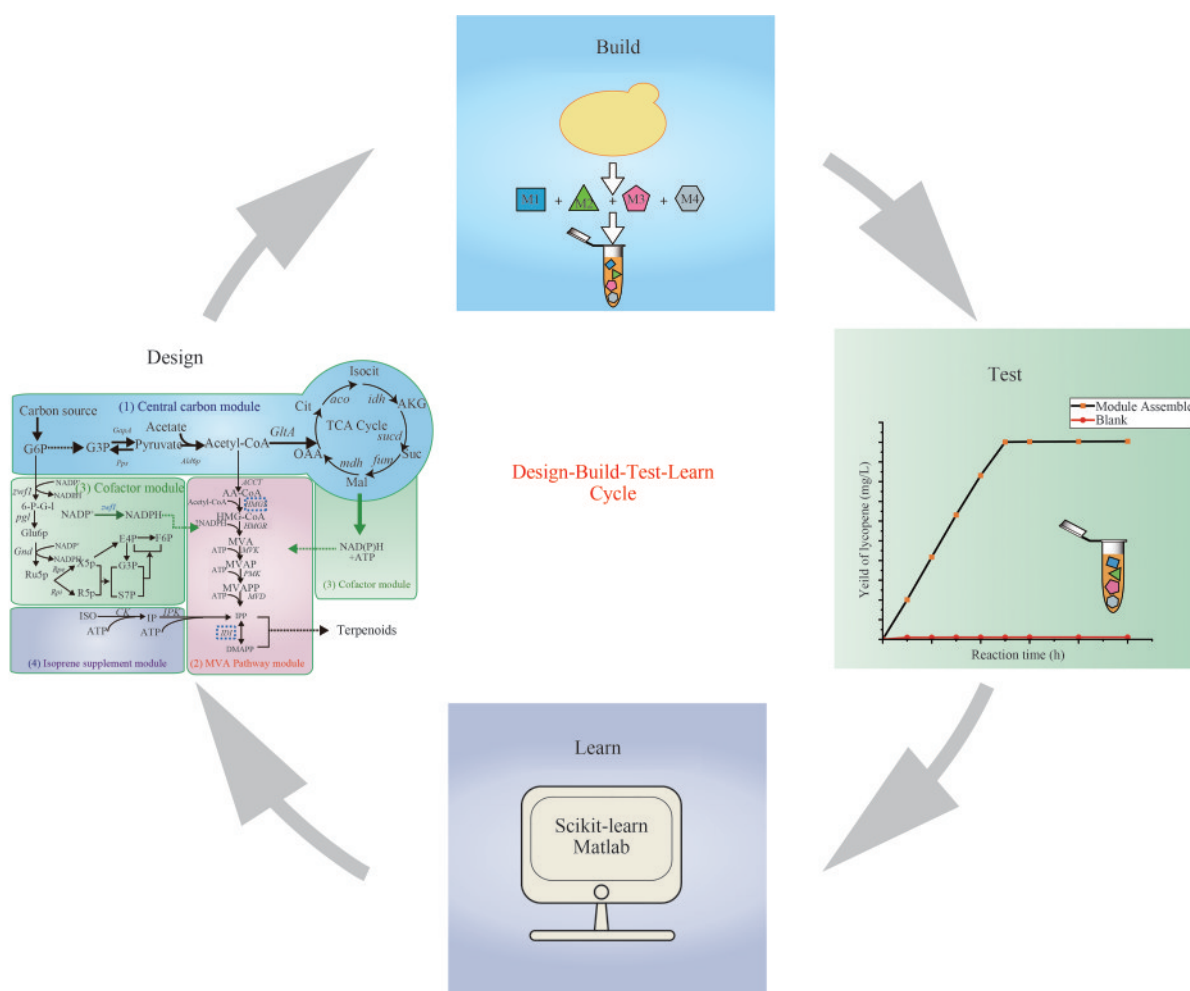
收稿日期: 2022-04-21 修回日期: 2022-09-06

基金项目: 国家重点研发计划 (2018YFA0901700)

引用本文: 唐士茗, 胡纪元, 郑穗平, 韩双艳, 林影. 基于无细胞体系的生物合成代谢模块设计、构建与快速途径原型[J]. 合成生物学, 2022, 3(6): 1250-1261

Citation: TANG Shiming, HU Jiyuan, ZHENG Suiping, HAN Shuangyan, LIN Ying. Designing, building and rapid prototyping of biosynthesis module based on cell-free system[J]. Synthetic Biology Journal, 2022, 3(6): 1250-1261

developing biosynthetic system for renewable chemical synthesis *in vitro* to avoid laborious and expensive ways for the optimization of metabolism pathway, strain and biocatalysts for each new product. In this review, we discuss the strategies of modules design and their rapid prototyping based on cell-free protein synthesis for assembling biosynthetic pathway *in vitro*. Biosynthetic modules could be sets of enzymes that catalyze a cascade reaction for a specific purpose or chemical, containing the conversion of starting materials to intermediary metabolites, biosynthesis of target products from the intermediates, cofactor balance and phosphorylation. Enzymes from distinct sources can be combined to construct desired reaction cascades with fewer biological constraints in one vessel, enabling easier pathway design with high modularity. Multiple modules could then be designed by different groups for different purpose with the help of metabolic pathway database, computational design tool and some general module design rules. Cell-free protein synthesis was here utilized to build and rapidly prototype the functionality of biosynthesis pathway and module. The further application of machine learning methods might also contribute to better “precision-robustness” design and construction of these modules. This process is also called cell-free pathway engineering which has been proved to be a powerful and flexible enabling technology, providing simpler and faster engineering solutions with an unprecedented freedom of design in an open environment than cell system.



Keywords: biosynthesis module; cell-free protein synthesis; rapid prototyping; machine learning

合成生物学是在工程学思想指导下重新设计与改造天然生物系统，并融合了生物学、工程学、化学、物理学及计算机等学科的崭新学科。随着合成生物学的发展，快速、可靠地设计符合预期功能的生物合成系统正逐渐成为现实。近十年来，诞生了大量基于合成生物学思想指导化学品生产的成功案例，如柠檬烯^[1]、正丁醇^[2]、乙酰辅酶A^[3]、淀粉^[4]、阿魏酸^[5]、莱鲍迪苷^[6]、维生素B₁₂^[7]、脂肪酸^[8]等的生物合成。这些新创制的化学品绿色生物制造方法有望在未来逐步替代化学工业领域高耗能、高污染的合成方法，助力国家“碳达峰，碳中和”战略目标，为医药、农业、食品等领域创造新机遇。

规模化生物制造利用微生物与酶作为生物催化剂，利用廉价、可再生原料生产具有高附加值的酶、蛋白质或其他化合物，从传统诱变育种到代谢工程分子育种，大大加快了生物制造发展的进程。近年来，以标准化和模块化为核心准则的合成生物学的发展，为生物合成体系的精准化构建提供了理论基础。标准化指对基本生物元件如启动子、终止子、功能基因、核糖体结合位点等进行标准化定制，构建标准化元件库，以期实现即插即用式的工程化改造。而模块化则实现了代谢途径模块化区分，降低了途径的复杂度，有利于模块化的途径通量优化，通用型模块可以被移植到不同的底盘细胞中，加速工程菌株构建过程。

无细胞蛋白质合成（cell-free protein synthesis, CFPS）是基于天然的细胞提取物或纯化后的蛋白质合成组分，通过添加蛋白质合成反应底物（如DNA或mRNA模板、氨基酸、核苷酸等）、能源物质及必要的辅因子，将细胞内的蛋白质合成反应

简化为细胞外的溶液反应的方法，容易实现反应过程的标准化。加入模板基因后，在开放的环境中直接调控和操作基因转录、翻译乃至生物代谢。CFPS体系作为一种新的使能平台，近年来已被应用于生物合成途径和遗传电路的快速原型。基于无细胞蛋白质合成体系快速原型技术，通常完成生物合成模块“设计-构建-测试”的一个循环可能只需要数天的时间，大大加速了生物合成模块的开发及应用^[9]。

1 生物合成模块设计原则与特征

生物合成途径模块化策略为构建更复杂的细胞工厂提供了便利，这些细胞工厂可以被用于合成各种重要目标产物，萜类化合物便是其中之一。此类化合物广泛存在于自然界中，在医药、食品、化妆品等领域有着广泛应用。基于细胞的化学品生产仍充满许多挑战，如细胞膜屏障、产物毒性^[1, 10]以及产物合成和细胞生长之间的矛盾（表1）。随着人们对胞内代谢途径理解的逐渐深入，越来越多的科研人员尝试在细胞外重建清晰的目标产物合成途径，这种方法被描述为无细胞生物合成体系（cell-free biosynthesis system）^[11]。具体地说，是将一条完整的产物合成途径需要的所有酶在宿主菌中表达，并将粗裂解液或纯化酶在体外混合，从而在体外重建整条代谢途径^[12-26]。无细胞生物合成体系的途径构建的特点在于较强的鲁棒性和可扩展性。此外，无细胞生物合成途径的模块化设计及应用，不需考虑宿主遗传背景差异带来的限制^[27]，可以容易地实现模块化途径的即插即用。

表1 细胞生物合成与无细胞生物合成的比较

Tab. 1 Comparison of cellular and cell-free biosynthesis

衡量标准	细胞生物合成	无细胞生物合成
途径设计	胞内代谢网络复杂	自由灵活
生产效率	碳通量被用于细胞生长及产生副产物	碳通量最大程度流向目标化合物
物质传输	具有选择性屏障	直接添加底物，产物易提取
产物分离纯化	副产物较多	较容易
毒性影响	毒性物质影响细胞生长	没有生存限制
成本	较低	酶的制备和辅因子成本
成熟性	多年实践经验	初具规模

途径的模块化通常以重要代谢物或代谢分支为节点,可分为:底物输入模块、产物合成模块、通用中间模块及辅因子/能量因子再生模块等^[28-30],化合物生物合成模块化设计是合成途径设计的第一步,如何加快代谢模块的构建和测试是关键问题。Sieber课题组^[30]以葡萄糖为原料,利用启发于古菌的4步糖酵解模块将葡萄糖转化为丙酮酸,丙酮酸作为平台化合物进入乙醇合成模块,19 h内使葡萄糖合成乙醇的摩尔转化率达到57.4%;另外Sieber等将四步糖酵解模块与异丁醇合成模块进行组合,实现了葡萄糖经丙酮酸合成异丁醇的级联反应,23 h内摩尔转化率达到53%。Xie等^[31]将四步糖酵解模块与L-乳酸合成模块相结合,在无细胞生物合成体系中实现了葡萄糖经丙酮酸合成L-乳酸的反应,在优化的条件下,L-乳酸的产率达到90%。此外,这些模块化途径设计使得上游丙酮酸合成模块产生的NADH刚好被下游的产物合成模块利用,从而巧妙地实现了模块间辅酶的循环和整条途径辅酶的平衡(图1)。虽然上述纯化酶体系能将碳通量最大程度导向目标化合物,但酶的纯化成本较高,粗裂液体系则直接利用简单处理的细胞破碎液催化目标产物合成,具有比纯化酶体系更低的酶制备成本。Krutsakorn等^[2]设计了以葡萄糖为底物经16个酶催化生成正丁醇体外合

成途径,并利用简单热处理的重组大肠杆菌所释放的途径酶执行催化反应,该途径实现了辅酶的平衡和ATP/ADP的循环,同时,通过将途径分为7个模块使得途径优化更为便利,整条途径各部分的通量能够保持相对平衡,反应7 h正丁醇浓度达到3.5 mmol/L,转化率达到82%。

萜类化合物(类异戊二烯)是以异戊二烯为单位合成的系列衍生物,其生物合成代谢途径显示类异戊二烯来源于同一对底物异戊烯基二磷酸(IPP)及其同分异构体二甲基酰(DMAPP),这些前体分子起源于碳中心代谢产物乙酰辅酶A(进入甲羟戊酸MVA途径)或丙酮酸/甘油醛-3-磷酸(进入甲基赤藓糖醇-4-磷酸MEP途径),为了强化合成途径还可以重构辅因子再生模块及IUP引入模块(图2)。Bowie教授团队对大肠杆菌葡萄糖至柠檬烯的27步合成途径进行模块化划分,分为中心碳代谢模块、甲羟戊酸模块、用于辅酶平衡的分子净化阀模块,经模块化优化后的柠檬烯转化率达到88.3%。该课题组在后续工作中尝试使用粗裂解液将葡萄糖转化为柠檬烯^[32],通过建立模块化的柠檬烯合成平台,优化后的柠檬烯产率达到3.8 mg/(L·h)。精细的模块化途径划分使得无细胞生物合成体系模块与模块间的代谢通量调节更为便利。

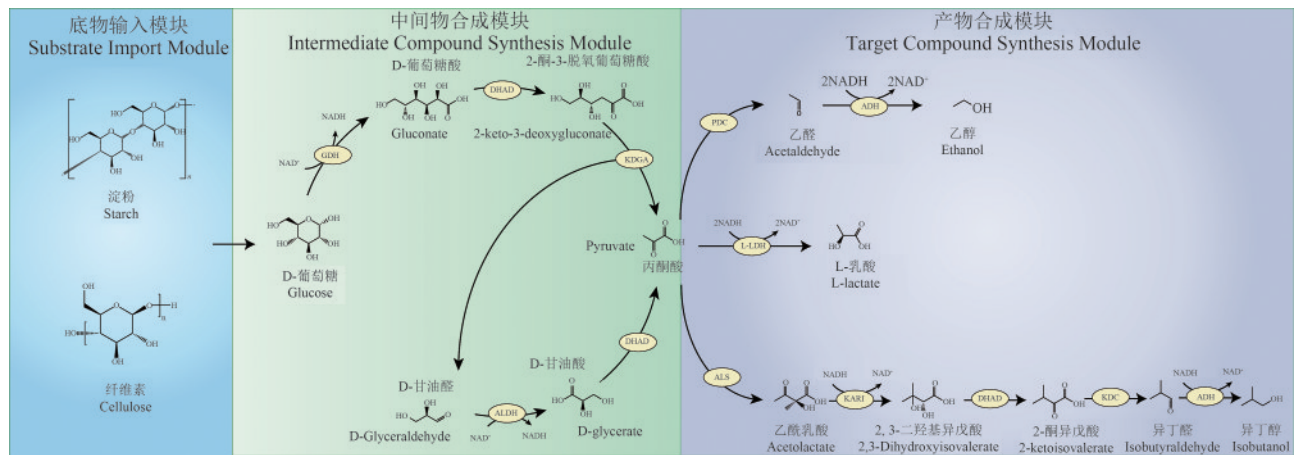


图1 淀粉生物基无细胞生物合成途径模块化设计

GDH—葡萄糖脱氢酶; DHAD—二羟酸脱水酶; KDGA—2-酮-3-脱氧葡萄糖醛缩酶; ALDH—甘油醛脱氢酶; L-LDH—L-乳酸脱氢酶; PDC—丙酮酸脱羧酶; ADH—乙醇脱氢酶; ALS—乙酰乳酸合成酶; KARI—酮醇酸还原异构酶; KDC—2-酮酸脱羧酶

Fig. 1 Modular design of cell-free biosynthetic pathway from starch-base materials

GDH—glucose dehydrogenase; DHAD—dihydroxy acid dehydratase; KDGA—2-keto-3-desoxy gluconate aldolase;

ALDH—glyceraldehyde dehydrogenase; L-LDH—L-lactate dehydrogenase; PDC—pyruvate decarboxylase;

ADH—alcohol dehydrogenase; ALS—acetolactate synthase; KARI—ketolacid reductoisomerase; KDC—2-ketoacid decarboxylase

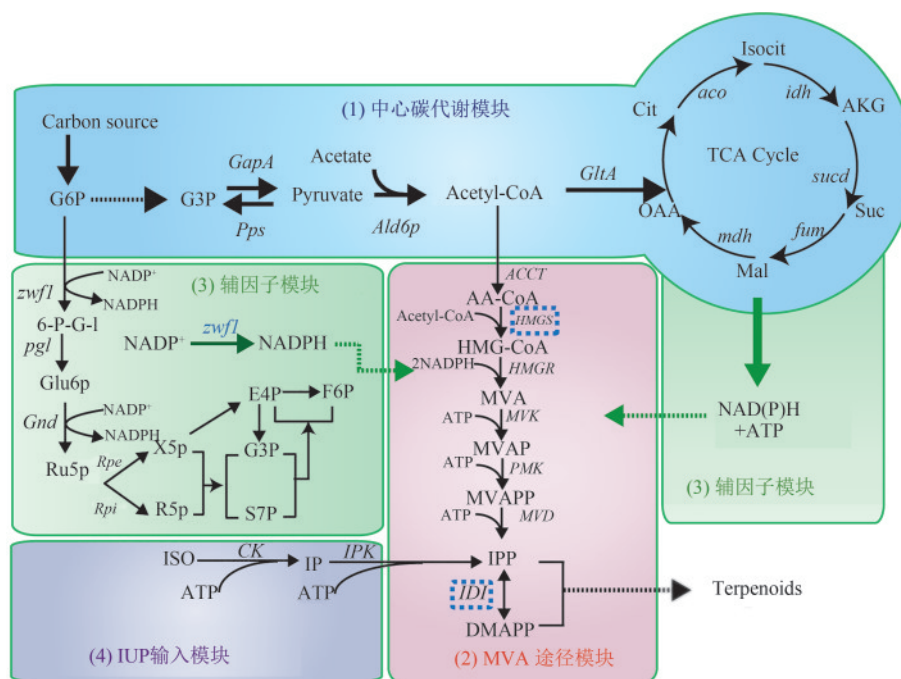


图2 酵母萜类化合物合成模块设计

Fig. 2 Modular design of terpenoids biosynthetic pathway in yeast

辅因子模块设计对无细胞生物合成的能量循环及辅酶再生显得尤为重要，体系中辅因子的消耗或不平衡减缓了反应，最终停止整个级联反应。为了克服这一问题，研究人员开发了辅因子再生和平衡模块，如ADP/ATP和NAD(P)⁺/NAD(P)H等。方翎教授^[33]在乳-N-二糖的合成途径中巧妙地设计ATP再生模块。乳-N-二糖的合成第一步通过半乳糖激酶的催化作用使半乳糖磷酸化，反应需要添加ATP，所生成的半乳糖-1-磷酸与N-乙酰氨基葡萄糖在乳-N-二糖磷酸化酶催化作用下合成乳-N-二糖，反应释放1分子磷酸。为了解决磷酸产物抑制的问题，在丙酮酸氧化酶的作用下消化磷酸生成乙酰磷酸，再利用乙酰激酶作用，将磷酸加到ADP上，构建了ADP/ATP循环再生模块。由此，一方面解决了磷酸对乳-N-二糖磷酸化酶的抑制，另一方面实现了反应途径中ATP的再生。

为了实现辅酶再生，通常需要在途径设计中确保途径上下游的辅酶化学计量平衡。Sieber课题组^[30]通过巧妙的设计，在上游模块将2 mol NAD⁺还原为NADH，再在下游模块将2 mol NADH氧化以实现辅酶再生，但采用这种方法严格限制了途径设计的灵活性。Bowie教授^[34]引入由PDH^{NADH}（具有还原NAD⁺活性）、PDH^{NADPH}（具有还原

NADP⁺活性）以及NoxE（具有氧化NADH活性）组成的“purge valve”模块。在该模块中，当NADPH浓度低时，NADP⁺/NADPH平衡可通过PDH^{NADPH}及下游产物合成途径氧化NADPH实现，当NADPH浓度高时，NAD⁺/NADH平衡可通过PDH^{NADH}及NoxE实现。“purge valve”模块的引入完美避开了辅酶再生所需的化学计量平衡限制，并成功应用于聚羟基丁酸酯^[34]、异戊二烯^[34]、柠檬烯^[1]、蒎烯^[1]、桉烯^[1]合成途径。

2 无细胞生物合成模块化设计方法

利用无细胞生物合成体系进行化学品合成，首先需要获得目标化合物的合成途径，常用的方法是基于文献挖掘，通过组合不同文献报道的底物到重要节点化合物，以及重要节点化合物到产物的合成模块获得新途径^[31, 35]，这种方法获得的途径可靠性高，但耗时耗力，需要对相关领域有较深刻的了解。随着生物信息学和信息技术的快速发展，越来越多的计算机辅助途径设计工具走进了科研人员的视野^[36]，这些工具包括从代谢途径数据库中获取反应的软件，如PathComp、

Pathway hunter tool^[37]及FMM^[38]等,用户只需要按照要求输入起始化合物和目标化合物,即可得到若干条可能的化学品合成途径,或基于反应规则指导的途径设计工具,如Retropath^[39]和XTMS^[40]。类似的工具还有许多,Wang等已进行了详细的综述^[36]。基于软件的路径设计工具效率较高,但可能存在不可靠途径,需要仔细筛查以选择合适的目标产物合成途径。近年来,一些组合算法例如comb-FBA也被应用于化学品合成途径设计并成功用于CO₂合成淀粉的途径^[4, 41]。

随着机器学习和人工智能浪潮的到来,AI辅助的途径设计正逐渐成为可能,机器学习基本流程包括:数据收集、数据预处理、模型训练、模型评价。数据收集是获得蛋白质序列、结构等与其对应标签信息的过程;数据预处理阶段将序列编码为数值型特征向量,并对数据进行标准化和归一化。目前在生物领域比较常用的机器学习算法包括支持向量机、人工神经网络、逻辑回归、决策树、随机森林等。Segler等^[42]提出了三种神经网络与MCTS相结合的深度神经网络算法框3N-MCTS,利用该框架选择、拓展、展示、更新的不断迭代得到全新的化学品合成途径,但该策略是基于化学反应规则来设计合成途径,如何将其应用于生物合成领域并匹配可靠的酶元件仍然面临挑战。

机器学习除了用于辅助途径设计外,还可以被应用于无细胞生物合成途径的酶元件改造和途径优化。目前,机器学习已经被应用于改善酶的热稳定性^[43-48]、改变酶的立体选择性^[49]、提高酶的催化性能^[50]、预测蛋白质溶解度^[51]及预测酶的EC编号^[52]。另外,机器学习可被应用于无细胞生物合成体系途径优化,Michael C. Jewett等^[53]使用神经网络模型对乙酰辅酶A到丁醇的合成途径进行了优化,经过神经网络模型预测和无细胞生物合成体系快速评价,使得丁醇效价达到了初始组的2.5倍。机器学习辅助的策略将会极大加速无细胞生物合成体系“设计-构建-测试-学习”循环,为更多化学品的无细胞生物合成提供可能。

对无细胞生物合成体系的途径设计而言,当获得若干条候选途径后,需要对整条途径的热力学可行性、途径长度、酶的可获得性等因素进行

综合评价以减少实验压力。热力学评价以标准吉布斯自由能变化为指标,当值小于零时,认为反应可以自发正向进行,Ron Milo课题组^[54]建立了最大最小驱动力指标(MDF)的热力学评价尺度,可以对整条反应途径的热力学可行性进行评估,MDF值越大代表目标途径在热力学上更有利,当MDF值小于0时代表途径不可行。另外,Ron Milo课题组开发的eQuilibrator在线工具(<http://equilibrator.weizmann.ac.il>)可以方便地计算途径中每一步反应的热力学自由能和发现途径中存在的热力学瓶颈^[55]。从途径长度的角度考虑应优先选择较短的途径,短的途径意味着更低的成本和更简化的体系。另外,所选途径中的每一步反应需要有相匹配的酶元件,且容易异源表达和纯化。与细胞生物合成相比,无细胞生物合成途径设计不需要考虑模块与宿主代谢网络的匹配及化合物毒性等问题,使得途径评价更加简单高效。

3 无细胞生物合成途径原型与优化

CFPS在开放的环境中,可实现直接调控和操作基因转录、翻译及代谢^[9],CFPS激活生物合成不受细胞膜和细胞壁的限制。与细胞合成运行模式类似,这种在体外同时实现基因表达及代谢功能作用的方法称为无细胞代谢途径原型(cell-free pathway prototyping)。然而,与细胞合成体系不一样,CFPS^[56-58]将细胞生长和蛋白质合成分开以减轻细胞的代谢负担,增强了生产力。由于CFPS的开放性,CFPS可以绕过传质限制及解决蛋白质毒性的耐受问题。此外,CFPS无细胞膜限制,更加容易地操作、调控与优化。

近年来,CFPS体系已发展成为一个使能平台,应用于生物合成途径和遗传电路的快速原型^[59]。在CFPS中使用线性DNA(如PCR产物)作为转录的模板^[60],容易实现高通量、快速原型设计^[61]。与细胞体内操纵转录和翻译机器策略不同,CFPS在相匹配的转录-翻译反应条件下,代谢途径中基因表达水平可以通过调节模板DNA浓度与比例^[62]、调节CFPS的理化性质^[63]、调节CFPS的组分及组分浓度^[64]、加入翻译调节因子^[65]等方式来实现,并在代谢途径和遗传电路的构建和优

化方面取得成功^[62, 66], 该应用体系的建立可使筛选的一个循环从几天减少到几个小时。因此, 代谢途径构建和基因电路测试的组合快速原型, 打开了独特的通路去加速“设计-构建-测试”循环, 加速了代谢工程和合成生物学的发展。

此外, 无细胞体系的快速途径原型作为代谢工程模块化的重要工具^[67], 该工具通过利用纯酶系统或细胞提取物系统(图3), 实现在细胞外代谢途径快速原型和模块化代谢途径优化^[68-73], 再回到细胞内组装并合成相应的化合物, 大大地提高了代谢合成体系的“设计-构建-测试”循环的效率^[74-75]。该工具运用于通用代谢模块的构建和平台化合物合成的优化, 进一步拓展模块适配性及通用性机制的研究, 更加具有吸引力和挑战性。

基于纯酶系统的快速途径原型, 首先将途径酶进行重组表达并纯化, 然后对纯化后的途径酶

进行代谢途径组装并投入底物进行途径验证。该类原型的优点在于其代谢网络清晰, 可对反应条件和代谢通量进行精确控制^[67]。Korman等^[1]设计了一个由27种酶组成的以葡萄糖为底物体外合成单萜纯酶系统, 经过优化, 柠檬烯、蒎烯、桉烯的摩尔转化率分别可达 $88.3\% \pm 4.6\%$ 、 $103.9\% \pm 8.1\%$ 和 $94.5\% \pm 4.7\%$, 但基于纯酶系统的快速途径原型存在成本高、辅因子再生困难的问题^[67, 76]。

基于细胞提取物系统的快速途径原型可通过两种方法得以实现: 其一, 利用细胞表达酶的提取物富集法, 首先利用底盘细胞对途径酶进行过表达, 然后对底盘细胞进行破碎, 接着将富集了途径酶的细胞提取物进行混合并投入底物进行途径验证; 其二, 直接利用CFPS, 制备细胞提取物, 接着投入途径酶的基因模板, 利用细胞提取物进行途径酶的同时表达或分别表达再混合, 最

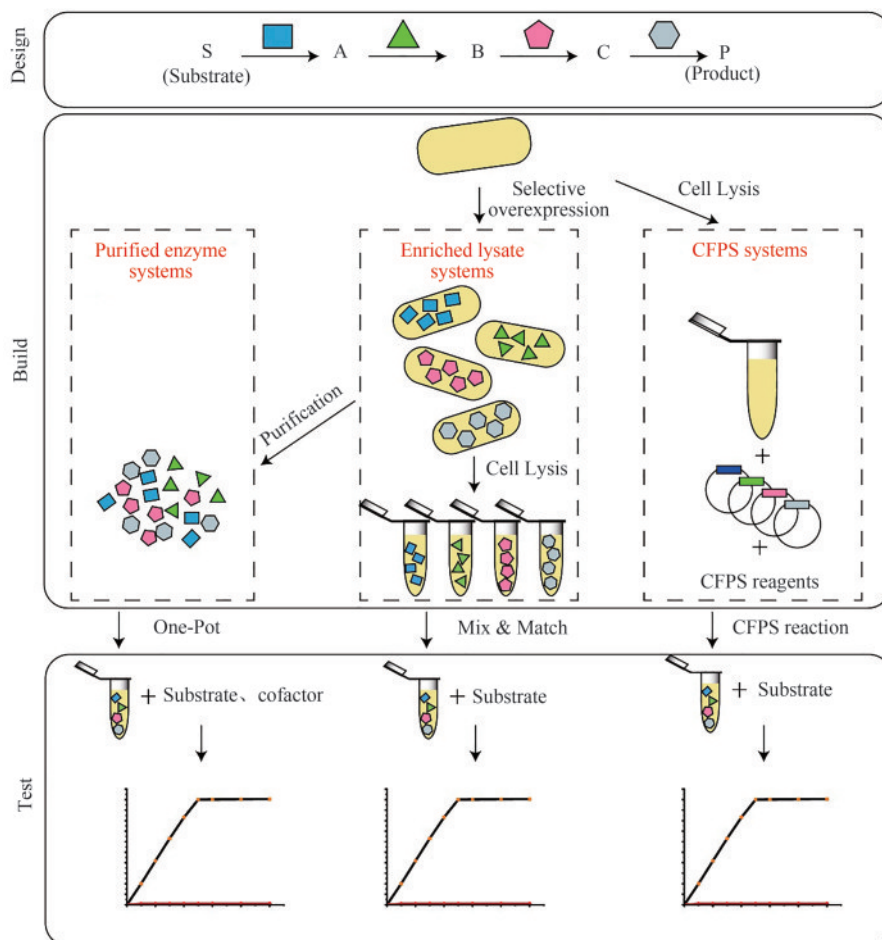


图3 基于无细胞生物合成体系的快速途径原型原理图

Fig. 3 Overview of cell-free biosynthesis pathway prototyping

后投入底物进行途径验证。该类原型的优点在于体系简便、制备成本低以及体系内存在天然代谢途径。Dudley等^[77]基于大肠杆菌CFPS系统建立了一套体外生物合成酶原型与快速优化方法(iPROBE),应用于从乙酰辅酶A到柠檬烯的9步异源酶途径优化,通过不同酶的同系物、酶浓度及辅因子条件的组合设计了超过580个反应条件,优化后反应合成柠檬烯产量提高至4.5 mmol/L(24 h),经过验证,仿生的合成效率及滴度提高效率与细胞内合成改造的效果相似。Karim等^[74]采用细胞提取物富集及CFPS系统走通了体外从葡萄糖到正丁醇的17步反应,证明了上述方法在快速生物合成途径原型和酶筛选方面的能力。所遇到的辅因子再生问题,需要通过增加辅因子再生模块加以解决。

林影课题组在建立高效毕赤酵母CFPS技术基础

上,针对萜类化合物合成复杂代谢途径进行模块化设计,搭建萜类合成模块及代谢途径快速原型体系,研究各合成模块间的匹配性及通用性,探讨萜类合成模块化设计原则,为毕赤酵母萜类合成底盘细胞构建提供基础(数据未发表),见图4。

4 挑战与展望

在化学品生物合成途径及模块设计的基础上,无细胞蛋白合成体系构建与代谢酶、生物合成模块及途径快速原型技术开发与应用已初见成效,但仍然存在以下挑战:

第一,无细胞蛋白合成体系绕开细胞生长的固有限制,将“设计-构建-测试”循环迭代的基本单元由细胞变为了细胞提取物,通过添加蛋白质

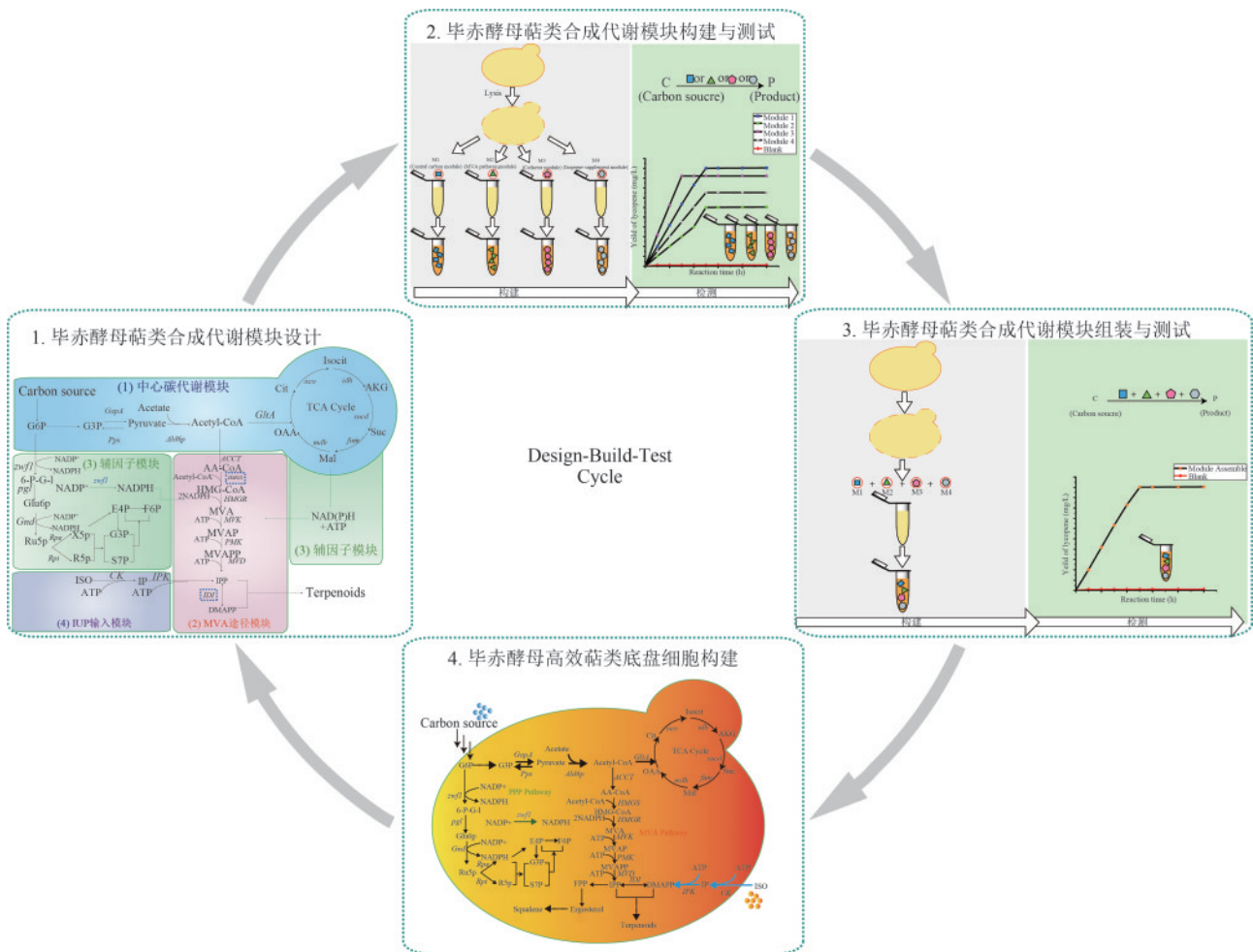


图4 酵母萜类合成模块设计及代谢途径快速原型
 Fig. 4 Design and rapid prototyping of isoprenoid synthesis modular in yeast

合成底物及酶催化的底物,可以一步实现酶同系物的快速筛选,以往体系中难以对酶表达进行定量,酶同系物筛选容易造成偏差。近期利用体外荧光蛋白分割技术构建无细胞蛋白合成定量检测方法^[78],实现了高活性和高表达途径酶的同时筛选,有利于筛选过程的定量化与标准化。

第二,随着基因组学的发展,越来越多生物物种基因组拼接完成,使得从代谢网络构建到各种类型的反应模块提炼成为可能,结合无细胞蛋白合成体系的体外代谢工程,由于其操作简单、模块化程度高、适合快速试错实验和测试,使其成为化合物生物合成途径设计和构建的重要平台技术^[28]。尽管一些已构建的辅因子或能量因子再生模块已实现了“即插即用”的尝试^[33],但当前仍存在不同代谢模块难以直接使用组装、体外基因与体内基因存在表达差异等问题亟需解决。人们尝试开发不同底盘细胞的无细胞蛋白合成体系,适应各种来源基因表达及蛋白质修饰,拓展无细胞蛋白合成体系快速原型的不同需求。

第三,无细胞蛋白质合成的合成途径/模块原型中,除了需要大力发展具有高效性、鲁棒性的原核与真核无细胞蛋白质合成体系外,遇到评测的组合和时空过于庞大时,比如途径的反应步骤较多、涉及同系酶库较大等,仍然会面临较大的试错实验压力。随着计算机技术的快速发展,机器学习作为一种强有力的工具正被逐渐应用于化学品合成途径快速原型的实验中,这种新的方式与传统的途径原型方式相比,仅需要测试部分本来训练机器学习模型,再根据模型反馈的结果择优选择预测产量较高的途径组合进行湿实验及评价,这样极大地节省测试成本,并缩短了“设计-构建-测试-学习”周期。此外,酶作为构建生物合成体系的重要功能单元,未来建立在大量蛋白质序列与其催化表型样本上的机器学习模型,能够使科研工作者真正实现输入途径酶的蛋白序列信息,便能对应酶的性能信息,通过生物合成代谢模块设计及基于无细胞蛋白合成体系的代谢途径原型,将有效加快化合物合成途径的组装及工业生产发展的进程。

近年来,生物合成途径设计、代谢模块构建及评价等方面取得了较大的进展,由于无细胞蛋

白合成体系的不断完善,使得无细胞途径原型成为酶元件筛选及代谢工程模块优化的重要工具。随着各种技术的不断加持,基于无细胞蛋白合成体系的生物合成代谢模块构建与快速原型技术将在人工生物合成体系构建,以及利用可再生资源合成大宗工业化学品、药物、功能材料和其他高附加值化合物方面大有所为。

参 考 文 献

- [1] KORMAN T P, OPGENORTH P H, BOWIE J U. A synthetic biochemistry platform for cell free production of monoterpenes from glucose[J]. *Nature Communications*, 2017, 8: 15526.
- [2] KRUTSAKORN B, HONDA K, YE X T, et al. *In vitro* production of n-butanol from glucose[J]. *Metabolic Engineering*, 2013, 20: 84-91.
- [3] LU X Y, LIU Y W, YANG Y Q, et al. Constructing a synthetic pathway for acetyl-coenzyme A from one-carbon through enzyme design[J]. *Nature Communications*, 2019, 10: 1378.
- [4] CAI T, SUN H B, QIAO J, et al. Cell-free chemoenzymatic starch synthesis from carbon dioxide[J]. *Science*, 2021, 373(6562): 1523-1527.
- [5] CHEN R B, GAO J Q, YU W, et al. Engineering cofactor supply and recycling to drive phenolic acid biosynthesis in yeast[J]. *Nature Chemical Biology*, 2022, 18(5): 520-529.
- [6] XU Y M, WANG X L, ZHANG C Y, et al. De novo biosynthesis of rubusoside and rebaudiosides in engineered yeasts[J]. *Nature Communications*, 2022, 13: 3040.
- [7] FANG H, LI D, KANG J, et al. Metabolic engineering of *Escherichia coli* for de novo biosynthesis of vitamin B₁₂[J]. *Nature Communications*, 2018, 9: 4917.
- [8] KIM H M, CHAE T U, CHOI S Y, et al. Engineering of an oleaginous bacterium for the production of fatty acids and fuels[J]. *Nature Chemical Biology*, 2019, 15(7): 721-729.
- [9] LU Y. Cell-free synthetic biology: Engineering in an open world[J]. *Synthetic and Systems Biotechnology*, 2017, 2(1): 23-27.
- [10] GRUBBE W S, RASOR B J, KRÜGER A, et al. Cell-free styrene biosynthesis at high titers[J]. *Metabolic Engineering*, 2020, 61: 89-95.
- [11] TANG S M, LIAO D C, LI X W, et al. Cell-free biosynthesis system: Methodology and perspective of *in vitro* efficient platform for pyruvate biosynthesis and transformation[J]. *ACS Synthetic Biology*, 2021, 10(10): 2417-2433.
- [12] YI T, LIM H J, LEE S J, et al. Synthesis of (R, R)-2, 3-butenediol from starch in a hybrid cell-free reaction system[J]. *Journal of Industrial and Engineering Chemistry*, 2018, 67: 231-235.

- [13] HONDA K, KIMURA K, NINH P H, et al. *In vitro* bioconversion of chitin to pyruvate with thermophilic enzymes[J]. Journal of Bioscience and Bioengineering, 2017, 124(3): 296-301.
- [14] LIU Z, ZHANG Y C, JIA X G, et al. *In vitro* reconstitution and optimization of the entire pathway to convert glucose into fatty acid[J]. ACS Synthetic Biology, 2017, 6(4): 701-709.
- [15] LI Y J, SHI T, HAN P P, et al. Thermodynamics-driven production of value-added d-allulose from inexpensive starch by an *in vitro* enzymatic synthetic biosystem[J]. ACS Catalysis, 2021, 11(9): 5088-5099.
- [16] WEI X L, LI Q Z, HU C C, et al. An ATP-free *in vitro* synthetic enzymatic biosystem facilitating one-pot stoichiometric conversion of starch to mannitol[J]. Applied Microbiology and Biotechnology, 2021, 105(5): 1913-1924.
- [17] ZHONG C, YOU C, WEI P, et al. Thermal cycling cascade biocatalysis of *myo*-inositol synthesis from sucrose[J]. ACS Catalysis, 2017, 7(9): 5992-5999.
- [18] SINGH R K, SINGH R, SIVAKUMAR D, et al. Insights into cell-free conversion of CO₂ to chemicals by a multienzyme cascade reaction[J]. ACS Catalysis, 2018, 8(12): 11085-11093.
- [19] BEER B, PICK A, SIEBER V. *In vitro* metabolic engineering for the production of α -ketoglutarate[J]. Metabolic Engineering, 2017, 40: 5-13.
- [20] LI C, ZHANG Z D, WANG J. A Thermophilic biofunctional multienzyme cascade reaction for cell-free synthesis of salvianic acid A and 3,4-dihydroxymandelic acid [J]. ACS Sustainable Chemistry & Engineering, 2019, 7(22): 18247-18253.
- [21] SUN S S, WEI X L, YOU C. The construction of an *in vitro* synthetic enzymatic biosystem that facilitates laminaribiose biosynthesis from maltodextrin and glucose[J]. Biotechnology Journal, 2019, 14(4): 1800493.
- [22] KIM T H, KANG S H, HAN J E, et al. Multilayer engineering of enzyme cascade catalysis for one-pot preparation of nylon monomers from renewable fatty acids [J]. ACS Catalysis, 2020, 10(9): 4871-4878.
- [23] OPGENORTH P H, KORMAN T P, BOWIE J U. A synthetic biochemistry module for production of bio-based chemicals from glucose[J]. Nature Chemical Biology, 2016, 12(6): 393-395.
- [24] WANG W, YANG J G, SUN Y X, et al. Artificial ATP-free *in vitro* synthetic enzymatic biosystems facilitate aldolase-mediated C—C bond formation for biomanufacturing[J]. ACS Catalysis, 2020, 10(2): 1264-1271.
- [25] BAI X, MENG D, WEI X, et al. Facile synthesis of (-)-vibioquercitol from maltodextrin via an *in vitro* synthetic enzymatic biosystem[J]. Biotechnology and Bioengineering, 2019, 116(10): 2710-2719.
- [26] DUDLEY Q M, ANDERSON K C, JEWETT M C. Cell-free mixing of escherichia coli crude extracts to prototype and rationally engineer high-titer mevalonate synthesis[J]. ACS Synthetic Biology, 2016, 5(12): 1578-1588.
- [27] LUO Y Z, LEE J K, ZHAO H M. Challenges and opportunities in synthetic biology for chemical engineers[J]. Chemical Engineering Science, 2013, 103: 115-119.
- [28] TANIGUCHI H, OKANO K, HONDA K. Modules for *in vitro* metabolic engineering: pathway assembly for bio-based production of value-added chemicals[J]. Synthetic and Systems Biotechnology, 2017, 2(2): 65-74.
- [29] XU P, GU Q, WANG W Y, et al. Modular optimization of multi-gene pathways for fatty acids production in *E. coli*[J]. Nature Communications, 2013, 4: 1409.
- [30] GUTERL J K, GARBE D, CARSTEN J, et al. Cell-free metabolic engineering: production of chemicals by minimized reaction cascades[J]. ChemSusChem, 2012, 5(11): 2165-2172.
- [31] XIE L P, WEI X L, ZHOU X G, et al. Conversion of d-glucose to l-lactate via pyruvate by an optimized cell-free enzymatic biosystem containing minimized reactions[J]. Synthetic and Systems Biotechnology, 2018, 3(3): 204-210.
- [32] DUDLEY Q M, NASH C J, JEWETT M C. Cell-free biosynthesis of limonene using enzyme-enriched *Escherichia coli* lysates[J]. Synthetic Biology, 2019, 4(1): ysz003.
- [33] DU Z Q, LIU Z Y, TAN Y S, et al. Lacto-*N*-biose synthesis via a modular enzymatic cascade with ATP regeneration[J]. iScience, 2021, 24(3): 102236.
- [34] OPGENORTH P H, KORMAN T P, BOWIE J U. A synthetic biochemistry molecular purge valve module that maintains redox balance[J]. Nature Communications, 2014, 5: 4113.
- [35] GMELCH T J, SPERL J M, SIEBER V. Optimization of a reduced enzymatic reaction cascade for the production of L-alanine[J]. Scientific Reports, 2019, 9: 11754.
- [36] 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.
- [37] RAHMAN S A, ADVANI P, SCHUNK R, et al. Metabolic pathway analysis web service (Pathway Hunter Tool at CUBIC) [J]. Bioinformatics, 2005, 21(7): 1189-1193.
- [38] CHOU C H, CHANG W C, CHIU C M, et al. FMM: a web server for metabolic pathway reconstruction and comparative analysis[J]. Nucleic Acids Research, 2009, 37(suppl_2): W129-W134.
- [39] DELÉPINE B, DUIGOU T, CARBONELL P, et al. RetroPath2.0: a retrosynthesis workflow for metabolic engineers[J]. Metabolic Engineering, 2018, 45: 158-170.
- [40] CARBONELL P, PARUTTO P, HERISSON J, et al. XTMS: pathway design in an eXTended metabolic space[J]. Nucleic Acids Research, 2014, 42(W1): W389-W394.
- [41] YANG X, YUAN Q Q, LUO H, et al. Systematic design and *in vitro* validation of novel one-carbon assimilation pathways[J].

- Metabolic Engineering, 2019, 56: 142-153.
- [42] SEGLER M H S, PREUSS M, WALLER M P. Planning chemical syntheses with deep neural networks and symbolic AI[J]. Nature, 2018, 555(7698): 604-610.
- [43] LI Y Q, FANG J W. PROTS-RF: a robust model for predicting mutation-induced protein stability changes[J]. PLoS One, 2012, 7(10): e47247.
- [44] YANG Y, UROLAGIN S, NIROULA A, et al. PON-tstab: Protein variant stability predictor. importance of training data quality[J]. International Journal of Molecular Sciences, 2018, 19(4): 1009.
- [45] JOKINEN E, HEINONEN M, LÄHDESMÄKI H. mGPFusion: predicting protein stability changes with Gaussian process kernel learning and data fusion[J]. Bioinformatics, 2018, 34(13): i274-i283.
- [46] GIOLLO M, MARTIN A J M, WALSH I, et al. NeEMO: a method using residue interaction networks to improve prediction of protein stability upon mutation[J]. BMC Genomics, 2014, 15(Suppl 4): S7.
- [47] LI G Y, QIN Y C, FONTAINE N T, et al. Machine learning enables selection of epistatic enzyme mutants for stability against unfolding and detrimental aggregation[J]. ChemBioChem, 2021, 22(5): 904-914.
- [48] WANG X F, GAO P, LIU Y F, et al. Predicting thermophilic proteins by machine learning[J]. Current Bioinformatics, 2020, 15(5): 493-502.
- [49] WU Z, KAN S B J, LEWIS R D, et al. Machine learning-assisted directed protein evolution with combinatorial libraries[J]. Proceedings of the National Academy of Sciences of the United States of America, 2019, 116(18): 8852-8858.
- [50] FOX R J, DAVIS S C, MUNDORFF E C, et al. Improving catalytic function by ProSAR-driven enzyme evolution[J]. Nature Biotechnology, 2007, 25(3): 338-344.
- [51] KHURANA S, RAWI R, KUNJI K, et al. DeepSol: A deep learning framework for sequence-based protein solubility prediction[J]. Bioinformatics, 2018, 34(15): 2605-2613.
- [52] LI Y, WANG S, UMAROV R, et al. DEEPRe: sequence-based enzyme EC number prediction by deep learning[J]. Bioinformatics, 2018, 34(5): 760-769.
- [53] KARIM A S, DUDLEY Q M, JUMINAGA A, et al. *In vitro* prototyping and rapid optimization of biosynthetic enzymes for cell design[J]. Nature Chemical Biology, 2020, 16(8): 912-919.
- [54] NOOR E, BAR-EVEN A, FLAMHOLZ A, et al. Pathway thermodynamics highlights kinetic obstacles in central metabolism[J]. PLoS Computational Biology, 2014, 10(2): e1003483.
- [55] FLAMHOLZ A, NOOR E, BAR-EVEN A, et al. eEquilibrator—the biochemical thermodynamics calculator[J]. Nucleic Acids Research, 2012, 40(D1): D770-D775.
- [56] WHITTAKER J W. Cell-free protein synthesis: the state of the art[J]. Biotechnology Letters, 2013, 35(2): 143-152.
- [57] CARLSON E D, GAN R, HODGMAN C E, et al. Cell-free protein synthesis: applications come of age[J]. Biotechnology Advances, 2012, 30(5): 1185-1194.
- [58] GREGORIO N E, LEVINE M Z, OZA J P. A user's guide to cell-free protein synthesis[J]. Methods and Protocols, 2019, 2(1): 24.
- [59] SILVERMAN A D, KARIM A S, JEWETT M C. Cell-free gene expression: an expanded repertoire of applications[J]. Nature Reviews Genetics, 2020, 21(3): 151-170.
- [60] SCHINN S M, BROADBENT A, BRADLEY W T, et al. Protein synthesis directly from PCR: progress and applications of cell-free protein synthesis with linear DNA[J]. New Biotechnology, 2016, 33(4): 480-487.
- [61] SUN Z Z, YEUNG E, HAYES C A, et al. Linear DNA for rapid prototyping of synthetic biological circuits in an *Escherichia coli* based TX-TL cell-free system[J]. ACS Synthetic Biology, 2014, 3(6): 387-397.
- [62] JIANG L H, ZHAO J R, LIAN J Z, et al. Cell-free protein synthesis enabled rapid prototyping for metabolic engineering and synthetic biology[J]. Synthetic and Systems Biotechnology, 2018, 3(2): 90-96.
- [63] GAN R, JEWETT M C. A combined cell-free transcription-translation system from *Saccharomyces cerevisiae* for rapid and robust protein synthesis[J]. Biotechnology Journal, 2014, 9(5): 641-651.
- [64] HODGMAN C E, JEWETT M C. Optimized extract preparation methods and reaction conditions for improved yeast cell-free protein synthesis[J]. Biotechnology and Bioengineering, 2013, 110(10): 2643-2654.
- [65] XU H L, LIU W Q, LI J. Translation related factors improve the productivity of a streptomyces-based cell-free protein synthesis system[J]. ACS Synthetic Biology, 2020, 9(5): 1221-1224.
- [66] TABUCHI T, YOKOBAYASHI Y. Cell-free riboswitches [J]. RSC Chemical Biology, 2021, 2(5): 1430-1440.
- [67] KARIM A S, JEWETT M C. Cell-free synthetic biology for pathway prototyping[J]. Methods in Enzymology, 2018, 608: 31-57.
- [68] KORMAN T P, SAHACHARTSIRI B, LI D, et al. A synthetic biochemistry system for the *in vitro* production of isoprene from glycolysis intermediates[J]. Protein Science, 2014, 23(5): 576-585.
- [69] ZHU F Y, ZHONG X F, HU M Z, et al. *In vitro* reconstitution of mevalonate pathway and targeted engineering of farnesene overproduction in *Escherichia coli*[J]. Biotechnology and Bioengineering, 2014, 111(7): 1396-1405.
- [70] CHEN X X, ZHANG C Q, ZOU R Y, et al. Statistical experi-

- mental design guided optimization of a one-pot biphasic multi-enzyme total synthesis of amorpho-4, 11-diene[J]. *PLoS One*, 2013, 8(11): e79650.
- [71] RODRIGUEZ S B, LEYH T S. An enzymatic platform for the synthesis of isoprenoid precursors[J]. *PLoS One*, 2014, 9(8): e105594.
- [72] TOOGOOD H S, CHEALLAIGH A N, TAIT S, et al. Enzymatic menthol production: one-pot approach using engineered *Escherichia coli*[J]. *ACS Synthetic Biology*, 2015, 4(10): 1112-1123.
- [73] SIEBELS I, NOWAK S, HEIL C S, et al. Cell-free synthesis of natural compounds from genomic DNA of biosynthetic gene clusters[J]. *ACS Synthetic Biology*, 2020, 9(9): 2418-2426.
- [74] KARIM A S, JEWETT M C. A cell-free framework for rapid biosynthetic pathway prototyping and enzyme discovery[J]. *Metabolic Engineering*, 2016, 36: 116-126.
- [75] KAY J E, JEWETT M C. Lysate of engineered *Escherichia coli* supports high-level conversion of glucose to 2, 3-butanediol[J]. *Metabolic Engineering*, 2015, 32: 133-142.
- [76] BUJARA M, PANKE S. In silico assessment of cell-free systems[J]. *Biotechnology and Bioengineering*, 2012, 109(10): 2620-2629.
- [77] DUDLEY Q M, KARIM A S, NASH C J, et al. *In vitro* prototyping of limonene biosynthesis using cell-free protein synthesis[J]. *Metabolic Engineering*, 2020, 61: 251-260.
- [78] 林影, 袁清焱, 梁书利. 定量检测无细胞蛋白质合成系统目

的蛋白产量及筛选高催化活性酶蛋白的方法. CN114019169 A [P].2022-02-08.

LIN Y, YUAN Q Y, LIANG S L. A method of quantification of protein production in cell-free protein synthesis and screening of enzymes with high catalytic activity, CN114019169A [P]. 2022-02-08.



通讯作者: 林影(1962—),女,博士,教授。研究方向为酶学与酶工程、合成生物学等,国家重点研发计划“合成生物学”专项项目负责人。

E-mail: feylin@scut.edu.cn



第一作者: 唐士茗(1995—),男,博士研究生。研究方向为酶工程、合成生物学。

E-mail: 202010108586@mail.scut.edu.cn