

· 工业发酵科学 ·

孙际宾 博士，博士生导师，中国科学院天津工业生物技术研究所副所长，中国科学院“百人计划”入选者，中国科学院特聘研究员，中国科学院科技创新交叉与合作团队牵头人，中国生物发酵产业协会理事。致力于发展工业微生物的系统与合成生物技术，以复杂生物网络研究为基础，整合组学数据，解析工业微生物的高产抗逆分子基础和代谢瓶颈，设计和创造符合工业化需求的微生物。承担了国家自然科学基金、973 计划、863 计划、中科院科技服务网络计划 (STS 计划)、中科院科技创新交叉与合作团队、天津市首批 A 类重点领域创新团队以及多个企业合作项目。已发表 SCI 论文 90 余篇，申请中国发明专利 30 余项，获中国专利授权 9 项，PCT 专利 2 项，曾获教育部提名国家技术发明奖二等奖与中国产学研合作创新成果奖等。



郑平 博士，博士生导师，中国科学院天津工业生物技术研究所研究员，天津市滨海新区“131”人才工程第二层次人才，天津市三八红旗手，天津市滨海新区第三届人大代表。长期从事细菌的代谢工程改造和系统生物学研究，致力于创建绿色、高效、低成本的生物制造体系，创新现代生物技术，提升推动传统生物及化工企业转型升级，承担国家自然科学基金面上项目、863 计划、中科院知识创新工程重要方向项目、天津市应用基础及前沿技术研究计划项目，并参与 973 项目与中科院知识创新工程项目等，已发表 SCI 论文 20 余篇，申请中国发明专利 20 余项，PCT 专利 2 项，已获中国专利授权 9 项。



面向工业生物技术的系统生物学

郑小梅^{1,2,3}，郑平^{1,2,3}，孙际宾^{1,2,3}

1 中国科学院天津工业生物技术研究所，天津 300308

2 中国科学院系统微生物工程重点实验室，天津 300308

3 中国科学院大学，北京 100049

郑小梅，郑平，孙际宾. 面向工业生物技术的系统生物学. 生物工程学报, 2019, 35(10): 1955–1973.

Zheng XM, Zheng P, Sun JB. Systems biology for industrial biotechnology. Chin J Biotech, 2019, 35(10): 1955–1973.

摘要: 工业生物技术是以微生物细胞工厂利用可再生的生物原料来生产能源、材料与化学品等的生物技术，在解决资源、能源与环境等问题方面起着越来越重要的作用。系统生物学是全面解析微生物细胞工厂及其发酵过程

Received: May 28, 2019; **Accepted:** August 5, 2019

Supported by: National Natural Science Foundation of China (Nos. 31700085, 31370113).

Corresponding authors: Jibin Sun. Tel: +86-22-84861949; Fax: +86-22-84861943; E-mail: sun_jb@tib.cas.cn

Ping Zheng. Tel: +86-22-84861945; Fax: +86-22-84861943; E-mail: zheng_p@tib.cas.cn

国家自然科学基金 (Nos. 31700085, 31370113) 资助。

网络出版时间: 2019-08-23

网络出版地址: <http://kns.cnki.net/kcms/detail/11.1998.Q.20190823.1520.001.html>

从“黑箱”到“白箱”的重要研究方法。系统生物学借助基因组、转录组、蛋白质组、代谢组以及代谢流组等多组学数据,可解析微生物细胞工厂在 RNA、蛋白与代谢物等不同水平上的变化规律与调控机制。目前,系统生物学在微生物细胞工厂的设计创建与发酵工艺优化中起着越来越重要的指导作用,许多成功应用实例不断涌现,推动着工业生物技术的快速发展。文中重点综述基因组、转录组、蛋白质组、代谢组与代谢流组以及基因组规模的网络模型等各组学技术的最新发展及其在工业生物技术尤其是菌株改造与发酵优化中的应用,并就工业生物技术中系统生物学的未来发展方向进行展望。

关键词: 工业生物技术, 系统生物学, 多组学, 菌株设计, 发酵优化

Systems biology for industrial biotechnology

Xiaomei Zheng^{1,2,3}, Ping Zheng^{1,2,3}, and Jibin Sun^{1,2,3}

1 Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

2 Key Laboratory of Systems Microbial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

3 University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: In industrial biotechnology, microbial cell factories utilize renewable resources to produce energy, materials and chemicals. Industrial biotechnology plays an increasingly important role in solving the resource, energy and environmental problems. Systems biology has shed new light on industrial biotechnology, deepening our understanding of industrial microbial cell factories and their bioprocess from “Black-box” to “White-box”. Systems-wide profiling of genome, transcriptome, proteome, metabolome, and fluxome has proven valuable to better unveil network operation and regulation on the genome scale. System biology has been successfully applied to create microbial cell factories for numerous products and derive attractive industrial processes, which has constantly expedited the development of industrial biotechnology. This review focused on the recent advance and applications of omics and trans-omics in industrial biotechnology, including genomics, transcriptomics, proteomics, metabolomics, fluxomics and genome scale modeling, and so on. Furthermore, this review also discussed the potential and promise of systems biology in industrial biotechnology.

Keywords: industrial biotechnology, system biology, multi-omics, strain design, fermentation optimization

工业生物技术以微生物细胞工厂利用可再生的生物原料来生产能源、材料与化学品等,是继医药生物技术、农业生物技术之后全球生物技术的第3次浪潮,在解决资源、能源与环境问题上起着越来越重要的作用,在建设绿色、低碳与可持续的产业经济体系上具有重大战略意义^[1]。工业生物技术的两大核心问题即微生物细胞工厂(即工业菌种)的设计构建与生物发酵过程的优化放大。对微生物细胞工厂及其发酵过程的认识与理解是解决这两大问题的关键,而系统生物学是解析工业菌种及发酵过程从“黑箱”到“白箱”的重要研究方法,是推动工业生物技术发展的重要驱动力。

系统生物学通过基因组、转录组、蛋白质组、

代谢组以及代谢流组等组学分析技术,系统解析细胞在 RNA、蛋白与代谢物等不同水平上的变化规律与调控机制,进而通过数据驱动的方法与数学模型化来模拟和认识工业细胞工厂的生命过程^[2]。近年来, DNA 测序与质谱等组学技术的发展将对生物细胞与生命过程的认识推入了系统化与数据化的时代。系统生物学在解决工业生物技术核心问题中起到越来越重要的指导作用,许多成功的应用实例不断在工业菌种的设计改造和发酵过程的优化放大中涌现,具体包括细胞工厂底盘细胞的选择与设计、新代谢途径的设计与构建、基因组规模的改造靶点鉴定、代谢途径优化、耐受性的提升、碳源利用的优化以及工业发酵工艺优化与放大等方面。

本文首先简述基因组、转录组、蛋白质组、代谢组与代谢流组等组学技术及相关数据库的最新研究,然后系统综述各组学技术或跨多组学的(Transomics)整合策略在工业生物技术中的应用,最后就工业生物技术中系统生物学的未来发展方向进行展望。

1 系统生物学的组学技术

近年来,随着不同物种在不同条件下的多组学海量数据的井喷式涌现,各组学数据库与多组学分析技术成为揭示生命过程分子机制的重要工具(表1)。本节中将简述基因组、转录组、蛋白质组、代谢组与代谢流组等相关技术及数据库的发展。

1.1 基因组

基因组(Genome)是生物体的全部遗传信息的总和,是研究认识生命过程的基础。基因组测序技术,尤其是第二代与第三代测序技术的快速发展,大大降低了基因组的测序成本,推动了基因组学的发展。以Illumina公司的HiSeq为代表的第二代测序技术^[3]具有高通量、高精度度与高覆盖度等优势,但其测序读长较短。而第三代单分子测序技术^[4]则具有10–20 kb的测序读长甚至可以跨越绝大多数的重复序列区域,但其精确度相对较低。第二代测序和第三代测序可互相取长补短,采用两个测序技术可得到更加完整和高质量的基因组数据。

基因组组装与注释是挖掘认识基因组信息的重要手段。基因组的注释关键包括基因组结构注释与功能注释。基因组结构注释预测定位各基因的物理图谱,获得基因起始密码子、内含子和外显子、基因的终止密码子等结构信息,可采用基于机器学习的从头预测或基于同源序列的比对预测。基因功能注释则多采用基于数据库中的已知序列进行同源比对的方式。常用的基因组数据库

有GenBank^[5]、Genome^[5]与GOLD^[6]等综合数据库以及物种特异性数据库如包含大肠杆菌*Escherichia coli*和谷氨酸棒杆菌*Corynebacterium glutamicum*等细菌基因组的数据库Microbesonline^[7]、酿酒酵母*Saccharomyces cerevisiae*基因组数据库SGD^[8]、真菌基因组数据库FungiDB^[9]等(表1)。除基因组的从头预测注释外,比较基因组可通过对比具有不同表型的菌株的基因组,获得与特定表型相关的单核苷酸多态性(Single nucleotide polymorphisms, SNPs)、基因的插入与缺失等。

1.2 转录组

转录组(Transcriptome)指在特定环境条件下表达的所有RNA。转录组学通过对不同条件下特定细胞内的所有RNA的定量检测,来分析各基因的表达水平^[10]。转录组定量检测技术主要包括基因表达芯片技术与基于第二代测序的RNA测序技术(RNA-seq)等。基因芯片在RNA的数量、物种灵活性、定量与序列分析等方面具有一定的局限,而RNA-seq技术由于具有高准确度、高分辨率、高灵敏度等优势,由于测序成本不断降低而被广泛应用。

比较转录组学通过对不同条件下基因表达水平的比对分析,来研究不同条件下的响应变化与不同菌株之间的差异。时间序列转录组数据可以动态地再现细胞内基因的表达变化,变化趋势一致的基因可能会有比较一致的功能。时间序列的转录组数据可展示不同生命过程中基因的动态变化,这些变化信息蕴含着丰富的调控关系,有助于我们更加深刻地认识微生物内部基因的动态关系,理解代谢过程。转录组数据库收集了不同物种不同条件下的转录组数据,如GEO^[11]、ArrayExpress^[12]与M3D^[13]等为转录组的整合分析提供了宝贵的数据资源(表1)。

表 1 各组学数据库

Table 1 Data repositories associated with omics datasets and metabolic engineering interest

Name	OMICS	Website	Description	Reference
GenBank	Genomics	www.ncbi.nlm.nih.gov/genbank	Genetic sequence database, an annotated collection of all publicly available DNA sequences	[5]
Genome	Genomics	https://www.ncbi.nlm.nih.gov/genome	Organizes information on genomes including sequences, maps, chromosomes, assemblies, and annotations	[5]
GOLD	Genomics	www.genomesonline.org	Catalogs genome-scale sequencing projects	[6]
Microbesonline	Genomics	http://www.microbesonline.org/	Contains 3 707 genomes, including 1 752 bacteria, 94 archaea, and 119 eukaryotes	[7]
SGD	Genomics	www.yeastgenome.org	A database for <i>S. cerevisiae</i> specifically	[8]
FungiDB	Genomics	fungidb.org/fungidb/	An integrated genomic and functional genomic database for the kingdom Fungi	[9]
GEO	Transcriptomics	www.ncbi.nlm.nih.gov/geo	A public functional genomics data repository	[11]
ArrayExpress	Transcriptomics	www.ebi.ac.uk/arrayexpress	Functional genomics data	[12]
M3D	Transcriptomics	http://m3d.mssm.edu/	A resource of microbial gene expression data	[13]
PRIDE	Proteomics	www.ebi.ac.uk/pride/archive	Mass-spectrometry-based proteomics specifically	[17]
GPM	Proteomics	www.thegpm.org	Tandem mass-spectrometry-based proteomics specifically	[18]
PeptideAtlas	Proteomics	www.peptideatlas.org	A multi-organism, publicly accessible compendium of peptides identified in a large set of tandem mass spectrometry proteomics experiments	[19]
Uniprot	Proteomics	www.uniprot.org	Provides proteomes for species with completely sequenced genomes	[20]
MitoMiner	Proteomics	http://mitominer.mrc-mbu.cam.ac.uk	A mitochondrial proteomics database of mammals, zebrafish and yeasts	[21]
Plasma Proteome Database	Proteomics	http://www.plasmaproteomedatabase.org	Plasma Proteome Database as a resource for proteomics research	[22]
METLIN	Metabolomics	metlin.scripps.edu	A repository for mass spectral metabolite data	[29]
MassBank	Metabolomics	www.massbank.jp	A mass spectral database of experimentally acquired high resolution MS spectra of metabolites	[57]
HMDB	Metabolomics	http://www.hmdb.ca/	A freely available electronic database containing detailed information about small molecule metabolites found in the human body	[32]
Golm Metabolome Database	Metabolomics	http://gmd.mpimp-golm.mpg.de/	Facilitates the search for and dissemination of reference mass spectra using GC-MS	[30]
MetaboLights	Metabolomics	www.ebi.ac.uk/metabolights	Contains protocols, publications, raw data, and meta data for a wide variety of metabolomics studies	[31]
CeCaFDB	Fluxomics	www.cecafdb.org	A collection of 100+ ¹³ C-flux publications, mostly regarding prokaryotes	[38]
KEGG	Metabolic pathway	www.kegg.jp	One of the most complete and widely used databases containing metabolic pathways (372 reference pathways) from a wide variety of organisms (>700)	[46]

待续

				续表 1
WikiPathways	Metabolic pathway	www.wikipathways.org	An open, collaborative platform for capturing and disseminating models of biological pathways for data visualization and analysis	[47]
MetaCYC	Metabolic pathway	metacyc.org	A metabolic pathway-specific database which describes the enzymes involved in over 400 metabolic pathways across 158 different organisms	[48]
BioCYC	Metabolic pathway	biocyc.org	A collection of 371 individual species pathway/genome databases	[49]
LIGAND	Metabolic reaction	www.genome.ad.jp/dbget/ligand.html	A repository for enzyme-catalyzed reactions	[50]
BRENDA	Metabolic reaction	www.brenda-enzymes.org	A repository for enzyme-catalyzed reactions	[51]

1.3 蛋白质组

蛋白质组 (Proteome) 指在特定环境条件下表达的所有蛋白质。蛋白质组学对不同环境条件下表达的所有蛋白质进行定性和定量分析, 系统研究蛋白质的加工、修饰以及蛋白质间的相互作用。随着蛋白质组的发展, 有许多不同方法用于蛋白质组的研究, 如蛋白芯片、二维聚丙烯酰胺凝胶电泳-质谱 (Two-dimensional polyacrylamide gel electrophoresis/Mass spectrometry, 2D-PAGE/MS) 与液质联用 (Liquid chromatography-tandem mass spectrometry, LC-MS/MS) 等。近年来, 质谱相关技术的迅速发展, 大大促进了蛋白质的鉴定与定量分析。一系列定量蛋白质组技术发展起来, 如细胞培养氨基酸稳定同位素标记 (Stable isotope labeling with amino acids in cell culture, SILAC)^[14]、同位素亲和标签 (Isotope coded affinity tag, ICAT)^[15]与同位素标记相对和绝对定量标记 (Isobaric tags for relative and absolute quantitation, iTRAQ)^[16]等, 为全面、系统地定性和定量分析复杂细胞蛋白质组提供了有效的技术手段。

目前, 常用蛋白质组数据库主要包括基于质谱的蛋白质组数据库如 PRIDE^[17]、GPM^[18]与 PeptideAtlas^[19]等, 综合蛋白质组数据库 Uniprot^[20]

与特定蛋白质组数据如线粒体蛋白质组 MitoMiner^[21]与膜蛋白质组 Plasma Proteome Database 等^[22](表 1)。这些数据库资源也为蛋白质组的鉴定与定量分析提供了有力的支持。

1.4 代谢组与代谢流组

与转录组与蛋白质组相似, 代谢组也是在特定生理条件下细胞所有代谢物的集合, 尤其是相对分子质量为 1 000 Da 以下的小分子代谢物。代谢组学主要系统研究代谢物的变化, 揭示生命过程中细胞代谢调控变化。由于基因突变或环境扰动可能不会引起转录和翻译的变化, 但会导致酶活性和代谢物浓度的变化。因此, 与转录组和蛋白质组相比, 代谢组更能反映出细胞代谢的动态变化^[23]。质谱相关技术如液质联用、气质联用 (Gas chromatography-Mass spectrometry, GC-MS/MS) 与毛细管电泳质谱 (Capillary electrophoresis-Mass spectrometry, CE-MS/MS) 以及核磁共振技术 (Nuclear magnetic resonance spectroscopy, NMR) 的发展, 大大推动了代谢组研究。

可靠的胞内代谢物组数据是认识细胞真实代谢状态的直接证据, 对于提高菌株代谢改造和工业发酵过程优化具有重要意义。但与 RNA 与蛋白质相比, 许多代谢组的半衰期非常短, 代谢组样品制备会直接影响代谢组数据质量。由于不同微

生物在代谢物谱上差异,针对不同微生物需要进行代谢组样品制备的优化,尤其是针对复杂的胞内代谢组^[24-28]。代谢组的样品制备需对细胞淬灭与代谢物提取等步骤进行优化。常用细胞淬灭方法有冷甲醇淬灭、冷乙醇淬灭与液氮速冻等,如冷甲醇或冷乙醇等对细胞造成损伤时,可考虑采用快速取样与液氮速冻相结合的方法。代谢物提取有热乙醇、热水、冷氯仿-甲醇、冷甲醇、冷乙醇等,由于不同提取剂的提取效率不同,为获得完整的胞内代谢物信息,也往往需要测试不同代谢物提取方法。

代谢组数据分析往往包括峰识别与匹配、归一化处理、小分子代谢物的鉴定与定量分析等。液相质谱数据库 Metlin^[29]和 MassBank、气相质谱数据库 Golm Metabolome Database^[30]以及综合质谱数据库 MetaboLights^[31]和 HMDB^[32]等可用于小分子代谢物的鉴定(表 1)。目前,代谢组分析的挑战主要在于不同微生物的样品方法优化、非靶向代谢组中的结构鉴定、复杂未知代谢物的结构鉴定以及不同代谢物的绝对定量等^[33]。

与代谢组相比,代谢流组更侧重胞内代谢流量的分析。目前,多采用¹³C-标记物来分析不同特定条件下细胞内的碳流分布,通常包括¹³C-标记物存在下的菌株培养、胞内代谢物的同位素分布检测、¹³C-标记辅助下的代谢途径与代谢流分析等步骤^[34]。代谢流分析(¹³C-Metabolic flux analysis, MFA)往往需要同位素数据处理与相应分析软件,如 OpenFLUX2^[35]、13CFLUX2^[36]与 INCA^[37]等。目前,代谢流分析数据库 CeCaFDB 收集了 100 多个基于¹³C-标记的代谢流数据^[38](表 1)。在代谢流分析相关研究方面也有许多比较优秀的综述,如 2009 年花强与杨琛^[39]系统综述了代谢流比率分析的研究方法及其在代谢改造中的应用。再如 Guo 等^[34]对¹³C-MFA 相关分析方法及其在不同工业菌株的应用进行了详细的综述。

1.5 基因组规模的生物网络模型

各组学数据的不断丰富为从系统的角度全面解析生命过程奠定了基础。但如何解析海量多组学数据仍然是系统生物学面临的挑战。全基因组规模的生物网络是解决这一问题的有效策略。目前,多个重要微生物不同层次的生物网络已经建立,尤其是基因组规模的代谢网络模型、转录调控网络模型与蛋白互作网络等。

目前,许多重要工业微生物,如大肠杆菌^[40]、酿酒酵母^[41-43]与黑曲霉 *Aspergillus niger*^[44-45]的基因组规模的代谢网络模型已建立。在代谢网络模型的构建中,首先可根据代谢途径相关的数据库如 KEGG^[46]、WikiPathways^[47]、MetaCyc^[48]、BioCyc^[49]、LIGAND^[50]与 BRENDA^[51]等对基因组进行解析,获得“基因-酶-通路-反应”的关系,进而搭建代谢网络框架(表 1)。然后,利用代谢流平衡分析(Flux balance analysis, FBA)方法^[52],结合自动化和人工校正,获得计量学的代谢网络模型。后续,可通过整合如时间序列的多组学数据,进一步建立基于动力学的代谢网络模型^[53]。基因组规模的代谢网络动力学模型,可以微分方程等数学方程刻画细胞过程,对细胞遗传因素与环境因素的预测更加可靠,将会成为未来的重要发展趋势。

转录调控网络模型是全局研究基因表达与调控的有效方法。利用时间序列的转录组数据或针对不同条件下的转录组数据进行基因共表达模式的整合分析,是构建转录调控网络模型常用策略。比如, Schape 等^[54]通过对 155 组不同条件下的黑曲霉转录组数据基于斯皮尔曼相关性分析,构建了基因组规模的基因共表达网络以及 9 500 个基因的共表达子网络,为基因组规模的基因功能研究提供有力的数据支撑。

目前大部分工业微生物还没有可靠的蛋白质相互作用网络。Kludas 等^[55]利用机器学习的算法,以从 GOLD 数据库中提取的酿酒酵母蛋白互作网

络为训练集,对里氏木霉 *Trichoderma reesei* 的蛋白-蛋白互作关系进行了预测分析,获得了里氏木霉的蛋白互作网络,其阳性率可达 75%^[55]。利用机器学习的模拟有望成为根据已知网络模型获得未知网络模型的重要方法。

1.6 多组学分析

近年来,基于各组学的系统生物技术的快速发展,为我们认识改造微生物成为细胞工厂提供了强有力的指导。全基因组测序与细胞转录、翻译与代谢等各个层次的定量分析是破解其遗传密码、进行基因组育种、基因信息挖掘与基因组规模建模的基础。基于多组学的系统分析,使我们可以从全局上理解细胞的代谢与调控机制,成为系统代谢工程的学习-设计-创建-测试 (Learn-Design-Build-Test, LDBT) 研究策略中的重要环节 (图 1)。

各组学数据是认识细胞代谢与生命过程的重要研究策略,但值得注意的是组学研究应是问题驱动的。在研究伊始,就需要提出明确的研究目

标与拟解决的问题,然后根据相应问题来选择相应的研究方法。比如,转录组与蛋白质组都是分析胞内基因表达水平的有效策略,但细分来看,转录组的通量更高、成本相对更低,是分析基因转录水平差异的首选。但转录组无法预测可能在翻译或翻译后修饰上存在限制的相关基因,针对这一问题,蛋白质组尤其是修饰蛋白质组有成为更为有效的方法。同时,由于蛋白质组还与蛋白质的稳定性相关,因此不同基因的转录组与蛋白质组可能并不是一一对应的。但也可利用两者之间的差异,通过转录组与蛋白质组的整合,可获得在翻译或翻译后修饰等方面的蛋白信息。

基于质谱的蛋白质组与代谢组分析,往往涉及非靶向与靶向的问题。非靶向分析更倾向于利用如主成分分析 (Principal component analysis, PCA) 等发现新的靶点,从检测与分析上,就要求全以获得更为完整的胞内蛋白或代谢物信息。靶向分析则更多地用于已知途径中不同蛋白或代谢物的分析。

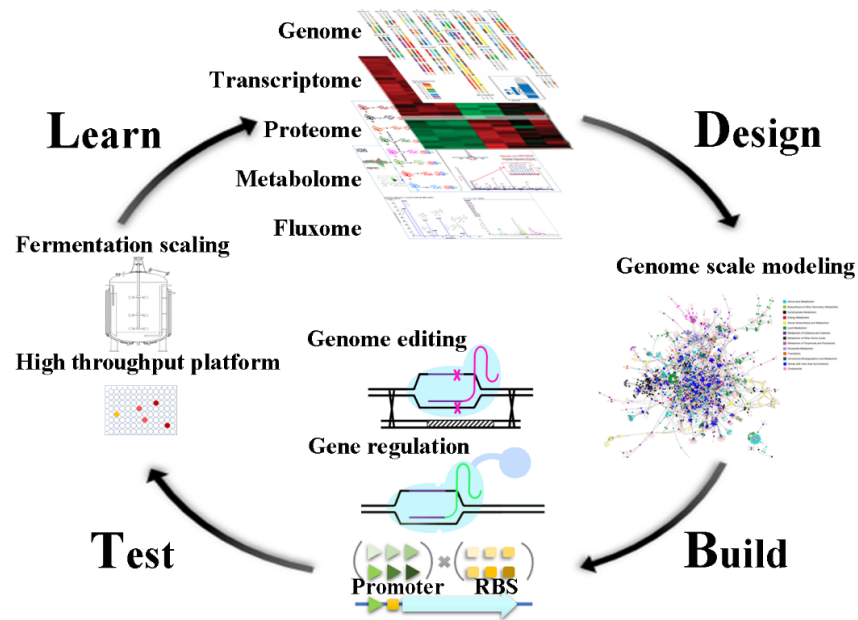


图 1 系统生物学驱动的合成生物学细胞工厂创建优化策略^[101]

Fig. 1 LDBT (Learn-Design-Build-Test) cycle of synthetic biology to develop industrial cell factory driven by systems biology study^[101].

由于细胞是一个复杂的整体,跨多组学数据的整合分析是全面解析细胞不同层面分子机制的重要研究方法。比如,Zhu等^[56]利用基因组、比较转录组与比较蛋白质组等多组学分析,揭示氮源缺乏是导致产油真菌圆红冬孢酵母 *Rhodospiridium toruloides* 油脂积累的重要原因,为系统阐释该菌株的全局调控奠定了基础,也为该菌株的理性改造提供新的靶点。

2 系统生物学在工业生物技术中的应用

工业生物技术可利用微生物细胞工厂以可再生的生物原料来生产目标能源、材料与化学品,是石油化学品工业生产的高效替代途径。理论上,微生物细胞工厂具有生产任一代谢网络中的代谢中间物的潜力,但目前绝大部分的天然微生物的生产能力是非常有限的。因此,基于系统生物学的认识与设计,利用合成生物学的手段来最终实现微生物细胞工厂的构建与优化,不断提升产量、转化率与生产强度这三大发酵指标,是目前工业生物技术的重要研究策略^[58-60]。系统生物学在微生物细胞工厂的设计改造以及工业发酵优化中具有重要的指导作用,许多成功的应用实例不断涌现。本节将根据具体实例分析系统生物学在工业生物技术中的应用。

2.1 基于系统生物学的细胞工厂设计改造

2.1.1 细胞工厂底盘细胞的选择与设计

重要模式微生物如大肠杆菌与酵母等由于清晰的遗传背景与高效的遗传操作,目前被开发拓展为可用于许多不同产品生产的工业底盘细胞^[60]。一些传统工业发酵菌株,如高产氨基酸的谷氨酸棒杆菌、高产有机酸的黑曲霉、高产丁醇的梭菌 *Clostridium* sp.、高产脂类物质的解脂酵母 *Yarrowia lipolytica* 以及高产多酮类化合物的放线菌等,由于具有强工业鲁棒性、高生产强度与低发酵成本

等优势,也成为非常理想的细胞工厂底盘^[2,61]。另外,一些具有独特生化特性的微生物,如蓝藻 *Cyanophyta*、微藻 *Microalgae* 和甲烷菌 *Methanobacteria* 等一碳底物利用菌株^[62],如嗜热毁丝霉 *Myceliophthora thermophila* 与嗜热侧孢霉 *Sporotrichum thermophila* 等嗜热菌^[63],也是底盘细胞的候选菌株。

在底盘细胞的设计中,通过重塑中心代谢途径来提高某一关键中间代谢物的合成流量,然后以这一中间代谢物为基础前体物质,用于后续一系列代谢物的生产。如 Jouhten 等^[64]利用 FBA、MiMBI 与代谢通量变化分析 (Flux variability analysis, FVA) 3种代谢流分析算法,重新设计了酿酒酵母的中心代谢流分布,构建了乙酰 CoA 的底盘细胞。而乙酰 CoA 作为参与 34 个代谢反应的重要中间代谢物,可用于一系列工业价值产品的生产,如可用于生物柴油与生物医药的脂类物质、可用于抗生素与抗癌药物合成的多酮类化合物以及可用于化妆品与食品添加剂的异戊二烯类物质等^[64-65](表 2)。

2.1.2 新代谢途径的设计与构建

由于涉及关键酶的筛选与选择,目标特定化学品的代谢途径设计是一个较为复杂的问题。传统途径设计多基于文献分析或基于 KEGG、MetaCyc 与 BRENDA 等代谢途径数据库的已知代谢途径分析,设计难度大,设计能力有限。目前,基于有机化学和生物化学的原理,通过计算机辅助设计,以基因组技术挖掘催化特定反应的酶的编码基因,实现快速非自然途径的创建^[33]。结合高效的 DNA 组装技术与基因组编辑技术,可在底盘细胞异源构建涉及多步反应的复杂化合物的合成,如天然代谢产物与微生物的异源合成。

Galanie 等^[66]借助基因组新酶挖掘,设计了阿片药物 (Opioid) 的完整合成途径,通过在酿酒酵

母底盘细胞中,整合来源于植物、动物、酵母与细菌的 21 个二甲基吗啡 (Thebaine) 合成相关与 23 个氢可酮 (Hydrocodone) 合成相关的关键酶,成功获得阿片药物细胞工厂,实现二甲基吗啡与氢可酮的异源合成。Liu 等^[67]通过对灯盏花基因组的系统挖掘,成功筛选鉴定到灯盏花素合成途径中的 P450 酶 EbF6H 和糖基转移酶 EbF7GAT 这两个关键酶,通过在酿酒酵母底盘细胞中异源构建灯盏花素合成途径,成功获得灯盏花素合成的细胞工厂,灯盏花素含量达到百毫克级。该研究对

将传统农业种植生产方式转变为规模化工业发酵生产路线,具有非常好的借鉴意义。

另外, Fang 等^[68]通过系统解析维生素 B12 好氧合成途径中钴螯合与腺苷钴啉醇酰胺磷酸的合成机理,将来源于 5 种不同细菌中的 28 个基因在大肠杆菌底盘细胞中成功组装与基因表达调控,成功在大肠杆菌中实现了维生素 B12 的从头合成,转化率达 307 mg/g DCW,发酵周期大大缩短,仅为目前工业水平的 1/10,为复杂化学品的从头合成与传统发酵菌株升级换代提供了很好的示范。

表 2 系统生物学在工业生物技术中的应用实例

Table 2 Recent applications of system biology in industrial biotechnology

Goal	Strategy	Strain	Product	Effect	Reference
Microbial cell factory engineering					
Chassis selection and design	<i>In silico</i> flux analysis	<i>Saccharomyces cerevisiae</i>	Acetyl-CoA chassis	Platform cell for the production of 29 chemicals	[64–65]
Novel metabolic pathways construction	Genomics	<i>S. cerevisiae</i>	Opioids		[66]
	Genomics	<i>S. cerevisiae</i>	Breviscapine	108 mg/L scutellarin; 185 mg/L apigenin-7-O-glucuronide	[67]
Genome-wide metabolic engineering targets identification	Genomics	<i>Escherichia coli</i>	Vitamin B12	↑250-fold (yield)	[68]
	Genomics+ transcriptomics	<i>Aspergillus niger</i>	Citric acid	↑5-folds (titer, 109 g/L)	[69–70]
Metabolic flux optimization	Proteomics	<i>E. coli</i>	Sesquiterpene	↑3-folds (titer, >500 mg/L)	[71]
	Proteomics+PCA	<i>E. coli</i>	Limonene	↑40% (titer)	[72]
	Proteomics+ metabolomics	<i>E. coli</i>	Polyphenol pinosylvin	↑2.2-folds (titer)	[73]
	Transcriptomics+ <i>in silico</i> flux modelling	<i>E. coli</i>	6-deoxyerythronolide (6dEB)	↑3-folds (titer, 210 mg/L)	[74]
Stress tolerance improvement	Metabolomics	<i>E. coli</i>	1-butanol	↑22% (titer, 18 g/L)	[75]
	Fluxome	<i>Corynebacterium glutamicum</i>	Lysine	↑60% (yield, 0.24 mol/mol)	[79]
	Fluxome	<i>Basfia succiniciproducens</i>	Succinate	↑44% (yield, 2.5 mol/mol)	[77]
Carbon utilization	Genomics	+ <i>E. coli</i>	Isopropanol	Improve isopropanol tolerance	[84]
	Transcriptomics	<i>C. glutamicum</i>	Succinate	↑38% (titer, 152 g/L)	[86]
	Proteomics+ metabolomics	<i>E. coli</i>	3-hydroxypropionate (3HP)	↑2-folds (titer, 39 g/L)	[87]
Bioprocess optimization	Transcriptomics	<i>Penicillium chrysogenum</i>	Arabinose	↑450-folds	[89]
	Metabolome	<i>C. glutamicum</i>	Arabinose		[90]
Bioprocess optimization	Gaussian-model	<i>E. coli</i>	Flavan-3-ols	↑65% (titer, 40.7 mg/L)	[91]
	Artificial neural network models	<i>Debaryomyces nepalensis</i>	Xylitol		[92]
	Fluxome	<i>Ashbya gossypii</i>	Riboflavin	↑45% (titer, 18 g/L)	[93]

2.1.3 基因组规模的改造靶点鉴定

代谢途径中限速节点的发现是细胞工厂的设计优化的关键。传统代谢工程中,往往基于先验知识的试错或大规模的文库筛选来发现代谢瓶颈。但这种策略在菌株改造靶点的发现也往往会进入一个瓶颈。基于各组学及其整合分析则成为从全基因组规模上发现代谢瓶颈的有效策略。

比较组学是发现代谢瓶颈的有效策略。殷娴等^[69]利用比较基因组与比较转录组分析了柠檬酸高产黑曲霉菌株与低产的退化菌株之间的差异,发现柠檬酸转运蛋白的候选基因可能是柠檬酸高产的重要因素。Steiger 等^[70]通过基因组挖掘的方式也发现了相同的柠檬酸转运蛋白编码基因,利用 Tet-on 诱导表达系统来过表达该转运蛋白后,可使柠檬酸的产量较出发菌株提高了 5 倍。Redding-Johanson 等^[71]利用选择反应监测-质谱 (Selected-reaction monitoring-Mass spectrometry, SRM-MS) 对倍半萜烯 (Amorpha-4,11-diene) 的大肠杆菌工程菌株进行定量蛋白质组分析,发现来源于酿酒酵母的甲羟戊酸激酶 (Mevalonate kinase, MK) 与磷酸甲羟戊酸激酶 (Phosphomevalonate kinase, PMK) 可能是倍半萜烯代谢途径中的限速节点。通过对 MK 与 PMK 的密码子优化与强启动子替换,倍半萜烯的产量提高 3 倍,达到 500 mg/L 以上。

另外,PCA 分析也是发现代谢节点的有效方法。Alonso-Gutierrez 等^[72]对 27 个具有不同柠檬烯产量的大肠杆菌中的 9 个甲羟戊酸合成关键酶的靶向定量蛋白质组数据进行 PCA 分析,鉴定出高产菌株中的表达强化的关键酶,进一步强化关键酶的表达后,柠檬烯的产量提高 40%。

多组学整合分析可用于发现翻译后修饰等调控所导致的代谢瓶颈。多酚银松素具有抗菌与抗真菌的活性,丙二酰 CoA 是其重要的前体,需要外源添加浅蓝菌素 (Cerulein) 来获得,但其

高水平反而抑制生产。为解析其抑制机制, Xu 等^[73]进行蛋白质组与代谢组分析,发现胞内高水平的丙二酰 CoA 与胞内蛋白的丙二酰化相关联。进一步分析发现 4-香豆酸 CoA 连接酶 (4-coumarate-CoA ligase, 4CL) 与芪合成酶 (Stilbene synthase, STS) 丙二酰化后活性丧失,这可能是浅蓝菌素抑制银松素合成的关键。将 STS 的丙二酰化位点 Lys113 与 Lys161 突变为精氨酸,避免其丙二酰化后,突变菌株可耐受更高水平的丙二酰 CoA,银松素的产量提高了 2.2 倍。由此可见,在菌株设计改造中,如翻译后修饰等不同调控水平上的限制因素分析是非常重要的。

2.1.4 代谢途径优化

除在限速代谢节点的预测外,各组学及基因组规模的代谢网络模型等在整体代谢途径的优化上也发挥着越来越大的作用,许多成功的例子不断被报道。

Meng 等^[74]通过转录组与代谢流模型的整合分析,来预测在 6-脱氧赤酮酸内酯 B (6-deoxyerythronolide B, 6dEB) 合成中不同代谢模块的作用。研究通过弱化磷酸戊糖途径与核苷酸代谢中的所有 25 个基因,显著提高 6dEB 的生产,与出发菌株相比产量提高了 3 倍,达到 210 mg/L。值得注意的是,不同组学分析的预测结果可能会存在差异,比如在比较转录组中发现上调的基因,在代谢流模型可能预测为下调。这提示我们转录组数据反映的是胞内基因的表达水平,会影响胞内代谢状态,但并不能直接代表代谢水平。

代谢组分析可能更有助于设计优化胞内的代谢流分布。Ohtake 等^[75]利用野生大肠杆菌菌株与敲除菌株的比较代谢组分析,发现当敲除 *pta* 途径后会引入 CoA 代谢失衡,而引起其他副产物丙酮酸与丁酸甲酯的生成,进一步分析发现由乙醇脱氢酶 AdhE2 所催化的丁酰 CoA 到正丁醛的合成流量降低可能是限速步骤。通过乙醇脱氢酶

AdhE2 表达的精细调控, 提高胞内 CoA 水平后, 显著降低副产物的生成, 最终实现大肠杆菌中的正丁醇产量提升 22%, 达到 18.3 g/L。

^{13}C -同位素标记的代谢流分析可利用质谱与核磁检测计算各胞内代谢物的 ^{13}C -同位素分布情况, 进而定量计算胞内各酶催化下的代谢通量, 被认为是最能反映胞内真实代谢状态的组学分析工具, 也不断应用于许多不同的工业生产菌株的认识与改造中。如 Lu 等^[76]利用同位素标记的代谢组与 ^{13}C 代谢流分析相结合, 比较分析了黑曲霉糖化酶高产菌株 DS03043 和野生型 CBS 513.88 的代谢差异。研究发现在高产菌株中, ATP/AMP 的比率更高, 高浓度 ATP 抑制了葡萄糖-6-磷酸异构酶的活性, 使碳流更多地流向氧化磷酸化途径, 导致 TCA 循环的碳流减少, 从而降低了如柠檬酸与草酸等副产物的积累。再如 Lange 等^[77]利用基于 GC-MS 的 ^{13}C -代谢流分析了产琥珀酸巴斯夫菌 *Basfia succiniciproducens* 胞内从蔗糖到琥珀酸的代谢通量。通过对 460 种代谢物的同位素分布检测, 准确定量了完整的中心代谢通量, 发现一个新的果糖激酶 RbsK 参与到蔗糖的分解产物果糖的磷酸化活化。通过对 PTS 葡萄糖转运系统中 FruA 的弱化与果糖激酶 RbsK 的强化, 可有效降低 PTS 系统造成的碳流损失, 从而大大提高了琥珀酸的转化率, 转化率达到 2.5 mol/mol, 接近 3 mol/mol 的理论转化率。另外, Yang 等^[78]发现甲基磷酸赤藓糖醇途径 (Methylerythritol phosphate pathway, MEP) 与甲羟戊酸途径 (Mevalonate pathway, MVA) 两个途径的同时过表达, 相对单独过表达时, 可分别使大肠杆菌的异戊二烯产率提高了 20 倍和 3 倍。进一步利用 ^{13}C -代谢流分析, 发现两途径协同过表达时, 可使 MEP 途径与 MVA 途径的流量分别增强 4.8 倍和 1.5 倍, 表明还原力的平衡与 ATP 供给在两途径协同平衡中起重要作用。两途径的协同作用可使大肠杆菌的异戊二烯

的产量达到 24 g/L, 葡萄糖转化率达到 0.267 g/g。

另一个利用代谢流分析的成功例子是谷氨酸棒杆菌以甘露醇为原料生产赖氨酸^[79]。甘露醇是新一代非粮原料海草中的主要成分。研究发现, 当以甘露醇为原料时, 碳流多流向糖酵解途径, 磷酸戊糖途径代谢流量较低, 使 NADPH 的供给严重不足。研究首次通过在胞内引入异源的果糖激酶来重新分配碳流, 从而增加磷酸戊糖途径的代谢流量, 但发现提升效果较低。为另寻他路, 通过表达参与糖酵解途径中的 NADP-依赖的甘油醛脱氢酶来提高 NADPH 还原力的供给, 成功解决了还原力供给不足的问题。

2.1.5 细胞工厂耐受性的提升

细胞工厂对发酵终产物或毒性中间代谢物的耐受性是制约工业发酵性能提升的重要因素。如果压力响应机制或毒性机制是清晰的, 则可通过相应的策略来缓解压力对细胞的胁迫, 提升菌株的耐受性^[80]。比如, 芳香族或烷烃等疏水化合物可直接插入至细胞膜中, 进而影响细胞膜的完整性与流动性引起细胞损伤。针对这一问题可利用细胞膜组分修饰来缓解, 如过表达细胞膜脂类顺反异构酶与去饱和酶等, 增加细胞膜中环丙基脂肪酸或改变脂肪酸链长度等。有的压力胁迫会激活胞内的未折叠蛋白应激反应 (Unfolded protein response, UPR) 而导致胞内蛋白质的错误折叠, 因此过表达分子伴侣也是较为常用的方式。比如在大肠杆菌或丙酮丁醇梭菌中过表达分子伴侣 GroESL 可显著提高正丁醇等的耐受性^[81-82]。再如, Zhu 等^[83]发现如 5-氨基乙酰丙酸 (5-aminolevulinic acid, 5-ALA) 等氨基酮类物质的积累也会产生活性氧等而损失细胞, 过表达过氧化氢酶 KatG/KatE 或超氧化物歧化酶 SodA/SodB/SodC 等可有效缓解氧化应激, 提高菌株的耐受性并增强 5-ALA 的产量。一些毒性化合物的外排蛋白如 RND 外排泵或压力调控转录因子 MetR 与 SoxS 的改造也有助

于提高菌株的耐受性。

当耐受性机制尚不清楚时,则可采用定向适应性进化的策略,利用连续传代富集获得耐受性显著提高的菌株。然后,利用比较组学的手段来发现引起耐受性提高的分子机制,再通过对底盘菌株的理性设计改造,获得耐受性不断提高的菌株^[60]。比如异丙醇对细胞有一定的毒性,会抑制细胞生长并降低产量。Horinouchi 等^[84]通过在异丙醇压力条件对大肠杆菌进行适应性进化,获得了耐受性提升的菌株。借助比较基因组,鉴定出 *relA*、*marC*、*proQ*、*yfgO* 与 *rraA* 基因上的突变可有助于耐受性的提高,进一步利用比较转录组分析,发现氨基酸合成代谢、铁离子稳态维持以及能量代谢等与耐受性有关。Wang 等^[85]利用基因组复制元件 DnaQ 突变体辅助的连续进化技术 (Genome replication engineering assisted continuous evolution, GREACE) 将大肠杆菌在赖氨酸终点发酵液中进行了定向进化,获得了赖氨酸产量显著提升的菌株 RS3,其赖氨酸产量可达到 155 g/L,比出发菌株 MU-1 提升 14.8%。进一步通过比较基因组与比较代谢组的分析,发现菌株 RS3 中 *SpeB*、*AtpB* 与 *SecY* 等基因的点突变,可能有助于提高细胞完整性与增强赖氨酸合成的碳流。

靶标产品的积累会对细胞产生一定的压力,从而限制菌株发酵性能的不不断提升。如谷氨酸棒杆菌发酵生产琥珀酸时,琥珀酸的不断积累会反馈抑制葡萄糖的吸收、细胞适应性以及琥珀酸生产。Chung 等^[86]利用基于 DNA 芯片的转录组分析发现随着琥珀酸的积累,转运途径与转录调控的基因表达发生显著变化。通过过表达一些转录下调的基因时可提高葡萄糖利用速率,但过表达中心代谢的关键酶对生产并无促进。然而,过表达转录因子 NCgl0275 可提高细胞适应性,并使琥珀酸生产大幅提高 38%,产量达 152 g/L。另外,蛋白质组与代谢组的整合也用于通过分析大肠杆

菌的静止细胞 (Q-cell, quiescent cell),来揭示 Q-cell 独特的代谢调控与耐受性^[87]。Q-cell 由质子载体吡啶诱导而生,可使细胞不再生长,但维持较高的代谢活性。通过比对细胞正常生长下与处于 Q-cell 的蛋白质组,发现加入吡啶后,压力响应蛋白的蛋白水平显著提高。这些压力蛋白将维持 Q-cell 耐受胞内的代谢失衡。利用这种代谢失衡引起的乙酰 CoA 与磷酸烯醇式丙酮酸的积累,可驱动羟基丙酸 (3-hydroxypropionate, 3HP) 的合成,使 3HP 的产量提高 2 倍,达到 39 g/L。

2.1.6 细胞工厂碳源利用的优化

目前,葡萄糖与蔗糖等是工业发酵中的主要碳源。但随着粮食危机的风险,许多替代性的非粮碳源如秸秆纤维素以及甲醇与甲酸等一碳化学品的利用也成为研究热点。与菌株耐受性提升的策略相似,替代性碳源利用菌株也可通过适应性进化的方式来获得,然后再借助多组学技术来分析在碳源利用中的关键酶。比如 Tuyishime 等^[88]利用适应性进化成功获得甲醇利用率显著提升的菌株,进一步利用比较基因组的方法发现胞内 NAD^+/NADH 、甲醇的吸收与耐受等是影响甲醇利用的关键因素。在碳源利用中,底物的转运是重要的影响因素。Bracher 等^[89]采用比较转录组分析了产黄青霉在葡萄糖、阿拉伯糖与乙醇 3 种不同碳源条件下的基因表达差异,预测出 5 个潜在的阿拉伯糖转运蛋白,通过验证发现, *PcAraT* 可回补酵母中 *GAL2* 的缺失,并可进一步提高阿拉伯糖的吸收与利用。由于该转运蛋白的高亲和力与专一性、在复杂碳源中仍保持较高的活性等特点,有望在生物炼制中具有较好的应用。

代谢组分析也在碳源利用的优化中发挥重要作用。Kawaguchi 等^[90]利用胞内代谢组分析了谷氨酸棒杆菌对葡萄糖与阿拉伯糖的代谢利用。与许多微生物一样,谷氨酸棒杆菌也存在葡萄糖抑制效应,会优先利用葡萄糖,而抑制阿拉伯糖的利用。

代谢组分析发现磷酸烯醇式丙酮酸与丙酮酸的开关是优化阿拉伯糖利用的关键。过表达丙酮酸激酶后,谷氨酸棒杆菌可同时利用葡萄糖与阿拉伯糖,尤其是当敲除阿拉伯糖抑制转录因子 AraR 时,可提高阿拉伯糖的利用。这一研究使谷氨酸棒杆菌可以在 PTS 系统存在下同时利用已糖与戊糖。

2.2 基于系统生物学的工业发酵优化放大

2.2.1 工业发酵工艺优化

多组学及基因组规模的代谢网络模型除在菌株的设计改造中发挥着越来越重要的作用外,在工业发酵优化中也逐渐得到应用。

发酵条件与工艺如温度、pH、溶氧与底物补料等都是目标产品发酵性能的重要影响因素,针对特定的工业菌株,往往需要优化其发酵工艺以达到最优的发酵性能。由于发酵过程的复杂性,许多统计方法与数学模型不断用于发酵优化。Jones 等^[91]利用数据驱动的高斯模型 (Empirically-derived scaled-Gaussian model) 对大肠杆菌共培养下的黄酮发酵过程进行模拟与优化。首先利用最小二乘回归 (Least squares regression) 对 21 个参数进行拟合,然后利用该模型优化如菌株间的适配度、碳源浓度、温度、诱导时间以及接种比例等发酵参数,以获得最优的黄酮类物质黄烷 (Flavan-3-ols) 的发酵条件。利用该模型优化的条件,可使黄烷的产量提高 65%,达到 40.7 mg/L,是单菌发酵的 970 倍。另外,神经网络模型 (Artificial neural network models, ANN) 也是发酵过程动态仿真与优化的重要方法。Pappu 等^[92]利用 ANN 模型以 339 个实验数据为训练集,来系统评估了发酵 pH、温度与溶氧等条件对卡森德巴利酵母 *Debaryomyces nepalensi* 木糖醇生产的影响。

代谢流分析可反映胞内真实代谢状态,除在代谢改造中,在发酵工艺优化中也被不断应用。Schwechheimer 等^[93]利用 GC/MS、LC/MS 与 1D-/2D-NMR 等多种质谱与核磁的方法系统检测

了棉阿舒囊霉 *Ashbya gossypii* 在生长与维生素 B2 等核黄素生产等不同阶段的胞内代谢物的 ¹³C 分布,细致刻画了核黄素代谢流分析。研究发现,甲酸的跨膜流量变化可能是核黄素合成初期的代谢瓶颈。通过分段补加少量甲酸与丝氨酸的补料方式则可解决一碳原料瞬时供给不足的问题,最终通过增加胞内甘氨酸、丝氨酸与甲酸可使核黄素的产量提高 45%。

2.2.2 工业发酵工艺放大

在工业发酵过程,发酵过程的逐级放大也是非常重要的环节。如何有效从实验室规模放大到工业实际生产规模,还是较大的挑战。工业发酵的规模往往较大,充分的搅拌是发酵的关键。不充分的搅拌可能会导致传质传氧传热的不均一,使微生物细胞所处的微环境不断变化,如底物、毒性副产物、溶氧、温度、pH 及 CO₂ 等,最终导致生物量、转化率与生产速率的不均一,从而形成复杂的细胞响应与代谢变化。

针对这一问题,发酵过程的多组学整合分析以及整合流体力学与细胞代谢动力学的分析可能成为有效的解决方法^[94]。通过对同一细胞工厂在不同发酵规模上的比较组学分析,可揭示出在逐级放大过程中的宏观发酵表型与微观细胞代谢的变化,有助于解决发酵放大中的瓶颈问题。另外,针对同一发酵过程中的各组学数据的整合分析及整合模型建立也有助于加深在发酵过程中细胞内各不同分子水平上的动态变化。同时,¹³C-代谢流分析也是检测不同发酵规模中细胞代谢变化的有效手段。de Jonge 等^[95]利用间歇性葡萄糖补料工艺与 ¹³C-代谢流相结合的方法,分析了产黄青霉的青霉素发酵过程中的碳源充足与饥饿条件下的变化,发现不同条件下有的胞内代谢物的浓度出现 100 倍的变化。这表明工业发酵放大过程,搅拌所带来的底物供给非均一化,会引起局部微环境中细胞代谢的剧烈变化,这可能是导致放大效果难以

成线性的重要原因。流体动力学模型(Computational fluid dynamics, CFD)也是预测不同发酵规模下的发酵条件变化的有力工具^[2],可通过对发酵罐内的传热与传质等不同参数的动态模拟来发现在发酵工艺控制与放大中的限速步骤。

3 总结与展望

随着以 DNA 测序技术与质谱技术为代表的组学技术与基因组规模生物网络的快速发展,系统生物学在工业生物技术的菌株改造与发酵过程优化中起着越来越重要的作用。结合许多工业菌株中高效基因组编辑工具^[96-99]、基因精细调控技术^[100]以及适应性进化,系统生物学的发展也将不断加速细胞工厂的改造与发酵优化(图 1)^[101]。

在细胞工厂的设计与优化中,各组学及多组学的整合分析与基因组规模的生物网络模型将是未来的重要研究方向。利用整合多种不同条件的组学数据的生物网络,可不断推进我们对微生物在基因密码、转录调控、翻译调控、翻译后修饰与代谢调控等各个不同层次上的认识,使基于数据模拟与仿真的细胞工厂设计成为现实,尤其是在复杂天然产物的人工合成上具有独特的优势^[102]。此外,针对不同物种的基因组信息挖掘与新酶从头设计改造也是细胞工厂的设计与优化的重要方向。

在工业发酵过程优化中,利用系统生物学的研究策略正在不断加深工业发酵过程中工业菌种的细胞代谢动态变化。各组学数据与一系列宏观发酵表型数据以及发酵罐的流体动力学数据^[2]的整合分析,也是未来从微观到宏观、从分子水平到发酵水平系统认识工业发酵过程,并实现高效放大优化的有效策略。

REFERENCES

[1] Chen F, Ding CJ, Chen YW, et al. A study on global

trends of industrial biotechnology and the perspectives in China. *World Sci-Tech R & D*, 2018, 40(2): 133-148 (in Chinese).

陈方,丁陈君,陈云伟,等. 工业生物技术领域国际发展态势及我国发展前景展望. *世界科技研究与发展*, 2018, 40(2): 133-148.

- [2] Campbell K, Xia JY, Nielsen J. The impact of systems biology on bioprocessing. *Trends Biotechnol*, 2017, 35(12): 1156-1168.
- [3] Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*, 2016, 17(6): 333-351.
- [4] Roberts RJ, Carneiro MO, Schatz MC. The advantages of SMRT sequencing. *Genome Biol*, 2013, 14: 405.
- [5] Sayers EW, Cavanaugh M, Clark K, et al. GenBank. *Nucleic Acids Res*, 2019, 47(D1): D94-D99.
- [6] Mukherjee S, Stamatis D, Bertsch J, et al. Genomes OnLine database (GOLD) v.7: updates and new features. *Nucleic Acids Res*, 2019, 47(D1): D649-D659.
- [7] Dehal PS, Joachimiak MP, Price MN, et al. MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res*, 2010, 38(D4): D396-D400.
- [8] Cherry JM, Hong EL, Amundsen C, et al. Saccharomyces genome database: the genomics resource of budding yeast. *Nucleic Acids Res*, 2012, 40(D1): D700-D705.
- [9] Basenko EY, Pulman JA, Shanmugasundram A, et al. FungiDB: an integrated bioinformatic resource for fungi and oomycetes. *J Fungi*, 2018, 4(1): 39.
- [10] Velculescu VE, Zhang L, Zhou W, et al. Characterization of the yeast transcriptome. *Cell*, 1997, 88(2): 243-251.
- [11] Clough E, Barrett T. The gene expression omnibus database//Mathé E, Davis S, Eds. *Statistical Genomics: Methods and Protocols*. New York, NY: Humana Press, 2016: 93-110.
- [12] Athar A, Füllgrabe A, George N, et al. ArrayExpress update - from bulk to single-cell expression data. *Nucleic Acids Res*, 2019, 47(D1): D711-D715.
- [13] Faith JJ, Driscoll ME, Fusaro VA, et al. Many microbe microarrays database: uniformly

- normalized Affymetrix compendia with structured experimental metadata. *Nucleic Acids Res*, 2008, 36(D4): D866–D870.
- [14] Hoedt E, Zhang GA, Neubert TA. Stable isotope labeling by amino acids in cell culture (SILAC) for quantitative proteomics//Woods AG, Darie CC, Eds. *Advancements of Mass Spectrometry in Biomedical Research*. Cham: Springer, 2014: 93–106.
- [15] Xiao Z, Veenstra TD. Comparison of protein expression by isotope-coded affinity tag labeling//Vlahou A, Ed. *Clinical Proteomics: Methods and Protocols*. Totowa, NJ: Humana Press, 2008: 181–192.
- [16] Gan CS, Chong PK, Pham TK, et al. Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). *J Proteome Res*, 2007, 6(2): 821–827.
- [17] Vizcaíno JA, Côté RG, Csordas A, et al. The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res*, 2013, 41(D1): D1063–D1069.
- [18] Craig R, Cortens JP, Beavis RC. Open source system for analyzing, validating, and storing protein identification data. *J Proteome Res*, 2004, 3(6): 1234–1242.
- [19] McCord J, Sun Z, Deutsch EW, et al. The PeptideAtlas of the domestic laying hen. *J Proteome Res*, 2017, 16(3): 1352–1363.
- [20] The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*, 2019, 47(D1): D506–D515.
- [21] Smith AC, Robinson AJ. MitoMiner v3.1, an update on the mitochondrial proteomics database. *Nucleic Acids Res*, 2016, 44(D1): D1258–D1261.
- [22] Nanjappa V, Thomas JK, Marimuthu A, et al. Plasma proteome database as a resource for proteomics research: 2014 update. *Nucleic Acids Res*, 2014, 42(D1): D959–D965.
- [23] Urbanczyk-Wochniak E, Luedemann A, Kopka J, et al. Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. *EMBO Rep*, 2003, 4(10): 989–993.
- [24] Zhang QQ, Zheng XM, Wang Y, et al. Comprehensive optimization of the metabolomic methodology for metabolite profiling of *Corynebacterium glutamicum*. *Appl Microbiol Biotechnol*, 2018, 102(16): 7113–7121.
- [25] Zheng XM, Yu JD, Cairns TC, et al. Comprehensive improvement of sample preparation methodologies facilitates dynamic metabolomics of *Aspergillus niger*. *Biotechnol J*, 2019, 14(3): 1800315.
- [26] Meinert S, Rapp S, Schmitz K, et al. Quantitative quenching evaluation and direct intracellular metabolite analysis in *Penicillium chrysogenum*. *Anal Biochem*, 2013, 438(1): 47–52.
- [27] Kim S, Lee DY, Wohlgemuth G, et al. Evaluation and optimization of metabolome sample preparation methods for *Saccharomyces cerevisiae*. *Anal Chem*, 2013, 85(4): 2169–2176.
- [28] Schädel F, David F, Franco-Lara E. Evaluation of cell damage caused by cold sampling and quenching for metabolome analysis. *Appl Microbiol Biotechnol*, 2011, 92(6): 1261–1274.
- [29] Guijas C, Montenegro-Burke JR, Domingo-Almenara X, et al. METLIN: a technology platform for identifying knowns and unknowns. *Anal Chem*, 2018, 90(5): 3156–3164.
- [30] Kopka J, Schauer N, Krueger S, et al. GMD@CSB.DB: the golm metabolome database. *Bioinformatics*, 2005, 21(8): 1635–1638.
- [31] Kale NS, Haug K, Conesa P, et al. MetaboLights: an open-access database repository for metabolomics data. *Curr Protoc Bioinformatics*, 2016, 53(1): 14.13.1–14.13.18.
- [32] Wishart DS, Feunang YD, Marcu A, et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res*, 2018, 46(D1): D608–D617.
- [33] Presnell KV, Alper HS. Systems metabolic engineering meets machine learning: a new era for data-driven metabolic engineering. *Biotechnol J*, 2019, 14(9): 1800416.
- [34] Guo WH, Sheng JY, Feng XY. ¹³C-metabolic flux analysis: an accurate approach to demystify microbial metabolism for biochemical production. *Bioengineering*, 2015, 3(1): 3.
- [35] Shupletsov MS, Golubeva LI, Rubina SS, et al.

- OpenFLUX2: ^{13}C -MFA modeling software package adjusted for the comprehensive analysis of single and parallel labeling experiments. *Microb Cell Fact*, 2014, 13: 152.
- [36] Weitzel M, Nöh K, Dalman T, et al. 13CFLUX2-high-performance software suite for ^{13}C -metabolic flux analysis. *Bioinformatics*, 2013, 29(1): 143–145.
- [37] Young JD. INCA: a computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics*, 2014, 30(9): 1333–1335.
- [38] Zhang ZD, Shen T, Rui B, et al. CeCaFDB: a curated database for the documentation, visualization and comparative analysis of central carbon metabolic flux distributions explored by ^{13}C -fluxomics. *Nucleic Acids Res*, 2015, 43(D1): D549–D557.
- [39] Hua Q, Yang C. Application of metabolic flux ratio analysis in metabolic engineering—a review. *Chin J Biotech*, 2009, 25(9): 1303–1311 (in Chinese).
花强, 杨琛. 代谢流量比率分析及其在代谢工程中的应用. *生物工程学报*, 2009, 25(9): 1303–1311.
- [40] Chandrasekaran S, Price ND. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA*, 2010, 107(41): 17845–17850.
- [41] Chi BF, Tao SH, Liu YL. Physiologically shrinking the solution space of a *Saccharomyces cerevisiae* genome-scale model suggests the role of the metabolic network in shaping gene expression noise. *PLoS ONE*, 2015, 10(10): e0139590.
- [42] Österlund T, Nookaew I, Bordel S, Nielsen J. Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. *BMC Syst Biol*. 2013, 7: 36.
- [43] Förster J, Famili I, Fu P, et al. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res*, 2003, 13(2): 244–253.
- [44] Sun JB, Lu X, Rinas U, et al. Metabolic peculiarities of *Aspergillus niger* disclosed by comparative metabolic genomics. *Genome Biol*, 2007, 8(9): R182.
- [45] Lu HZ, Cao WQ, Ouyang LM, et al. Comprehensive reconstruction and *in silico* analysis of *Aspergillus niger* genome-scale metabolic network model that accounts for 1210 ORFs. *Biotechnol Bioeng*, 2017, 114(3): 685–695.
- [46] Kanehisa M. The KEGG database//Bock G, Goode JA, Eds. 'In Silico' Simulation of Biological Processes. Chichester, West Sussex: John Wiley, 2002, 247: 91–101.
- [47] Slenter DN, Kutmon M, Hanspers K, et al. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res*, 2018, 46(D1): D661–D667.
- [48] Caspi R, Billington R, Fulcher CA, et al. The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res*, 2018, 46(D1): D633–D639.
- [49] Karp PD, Ouzounis CA, Moore-Kochlacs C, et al. Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res*, 2005, 33(19): 6083–6089.
- [50] Goto S, Nishioka T, Kanehisa M. LIGAND: chemical database of enzyme reactions. *Nucleic Acids Res*, 2000, 28(1): 380–382.
- [51] Schomburg I, Chang A, Hofmann O, et al. BRENDA: a resource for enzyme data and metabolic information. *Trends Biochem Sci*, 2002, 27(1): 54–56.
- [52] Lakshmanan M, Koh G, Chung BKS, et al. Software applications for flux balance analysis. *Brief Bioinform*, 2014, 15(1): 108–122.
- [53] Gomez JA, Barton PI. Dynamic flux balance analysis using DFBAIab//Fondi M, Ed. *Metabolic Network Reconstruction and Modeling: Methods and Protocols*. New York, NY: Humana Press, 2018: 353–370.
- [54] Schäpe P, Kwon MJ, Baumann B, et al. Updating genome annotation for the microbial cell factory *Aspergillus niger* using gene co-expression networks. *Nucleic Acids Res*, 2019, 47(2): 559–569.
- [55] Kludas J, Arvas M, Castillo S, et al. Machine

- learning of protein interactions in fungal secretory pathways. PLoS ONE, 2016, 11(7): e0159302.
- [56] Zhu ZW, Zhang SF, Liu HW, et al. A multi-omic map of the lipid-producing yeast *Rhodospiridium toruloides*. Nat Commun, 2012, 3: 1112.
- [57] Horai H, Arita M, Kanaya S, et al. MassBank: a public repository for sharing mass spectral data for life sciences. J Mass Spectrom, 2010, 45(7): 703–714.
- [58] Nielsen J, Keasling JD. Engineering cellular metabolism. Cell, 2016, 164(6): 1185–1197.
- [59] Chae TU, Choi SY, Kim JW, et al. Recent advances in systems metabolic engineering tools and strategies. Curr Opin Biotechnol, 2017, 47: 67–82.
- [60] Choi KR, Jang WD, Yang D, et al. Systems metabolic engineering strategies: integrating systems and synthetic biology with metabolic engineering. Trends Biotechnol, 2019, 37(8): 817–837.
- [61] Guo YM, Zheng P, Sun JB. *Aspergillus niger* as a potential cellular factory: prior knowledge and key technology. Chin J Biotech, 2010, 26(10): 1410–1418 (in Chinese).
郭艳梅, 郑平, 孙际宾. 黑曲霉作为细胞工厂: 知识准备与技术基础. 生物工程学报, 2010, 26(10): 1410–1418.
- [62] Kalyuzhnaya MG, Puri AW, Lidstrom ME. Metabolic engineering in methanotrophic bacteria. Metab Eng, 2015, 29: 142–152.
- [63] Singh B. *Myceliophthora thermophila* syn. *Sporotrichum thermophile*: a thermophilic mould of biotechnological potential. Crit Rev Biotechnol, 2016, 36(1): 59–69.
- [64] Jouhten P, Boruta T, Andrejev S, et al. Yeast metabolic chassis designs for diverse biotechnological products. Sci Rep, 2016, 6: 29694.
- [65] Chen Y, Daviet L, Schalk M, et al. Establishing a platform cell factory through engineering of yeast acetyl-CoA metabolism. Metab Eng, 2013, 15: 48–54.
- [66] Galanie S, Thodey K, Trenchard IJ, et al. Complete biosynthesis of opioids in yeast. Science, 2015, 349(6252): 1095–1100.
- [67] Liu XN, Cheng J, Zhang GH, et al. Engineering yeast for the production of breviscapine by genomic analysis and synthetic biology approaches. Nat Commun, 2018, 9(1): 448.
- [68] Fang H, Li D, Kang J, et al. Metabolic engineering of *Escherichia coli* for *de novo* biosynthesis of vitamin B₁₂. Nat Commun, 2018, 9(1): 4917.
- [69] Yin X, Shin HD, Li JH, et al. Comparative genomics and transcriptome analysis of *Aspergillus niger* and metabolic engineering for citrate production. Sci Rep, 2017, 7: 41040.
- [70] Steiger MG, Rassinger A, Mattanovich D, et al. Engineering of the citrate exporter protein enables high citric acid production in *Aspergillus niger*. Metab Eng, 2019, 52: 224–231.
- [71] Redding-Johanson AM, Batth TS, Chan R, et al. Targeted proteomics for metabolic pathway optimization: application to terpene production. Metab Eng, 2011, 13(2): 194–203.
- [72] Alonso-Gutierrez J, Kim EM, Batth TS, et al. Principal component analysis of proteomics (PCAP) as a tool to direct metabolic engineering. Metab Eng, 2015, 28: 123–133.
- [73] Xu JY, Xu Y, Chu XH, et al. Protein acylation affects the artificial biosynthetic pathway for pinosylvin production in engineered *E. coli*. ACS Chem Biol, 2018, 13(5): 1200–1208.
- [74] Meng HL, Xiong ZQ, Song SJ, et al. Construction of polyketide overproducing *Escherichia coli* strains via synthetic antisense RNAs based on in silico fluxome analysis and comparative transcriptome analysis. Biotechnol J, 2016, 11(4): 530–541.
- [75] Ohtake T, Pontrelli S, Laviña WA, et al. Metabolomics-driven approach to solving a CoA imbalance for improved 1-butanol production in *Escherichia coli*. Metab Eng, 2017, 41: 135–143.
- [76] Lu HZ, Liu XY, Huang MZ, et al. Integrated isotope-assisted metabolomics and ¹³C metabolic flux analysis reveals metabolic flux redistribution for high glucoamylase production by *Aspergillus niger*. Microb Cell Fact, 2015, 14: 147.
- [77] Lange A, Becker J, Schulze D, et al. Bio-based succinate from sucrose: high-resolution ¹³C

- metabolic flux analysis and metabolic engineering of the rumen bacterium *Basfia succiniciproducens*. *Metab Eng*, 2017, 44: 198–212.
- [78] Yang C, Gao X, Jiang Y, et al. Synergy between methylerythritol phosphate pathway and mevalonate pathway for isoprene production in *Escherichia coli*. *Metab Eng*, 2016, 37: 79–91.
- [79] Hoffmann SL, Jungmann L, Schiefelbein S, et al. Lysine production from the sugar alcohol mannitol: design of the cell factory *Corynebacterium glutamicum* SEA-3 through integrated analysis and engineering of metabolic pathway fluxes. *Metab Eng*, 2018, 47: 475–487.
- [80] Mukhopadhyay A. Tolerance engineering in bacteria for the production of advanced biofuels and chemicals. *Trends Microbiol*, 2015, 23(8): 498–508.
- [81] Zingaro KA, Papoutsakis ET. Toward a semisynthetic stress response system to engineer microbial solvent tolerance. *mBio*, 2012, 3(5): e00308–12.
- [82] Tomas CA, Welker NE, Papoutsakis ET. Overexpression of *groESL* in *Clostridium acetobutylicum* results in increased solvent production and tolerance, prolonged metabolism, and changes in the cell's transcriptional program. *Appl Environ Microbiol*, 2003, 69(8): 4951–4965.
- [83] Zhu CC, Chen JZ, Wang Y, et al. Enhancing 5-aminolevulinic acid tolerance and production by engineering the antioxidant defense system of *Escherichia coli*. *Biotechnol Bioeng*, 2019, 116(8): 2018–2028.
- [84] Horinouchi T, Sakai A, Kotani H, et al. Improvement of isopropanol tolerance of *Escherichia coli* using adaptive laboratory evolution and omics technologies. *J Biotechnol*, 2017, 255: 47–56.
- [85] Wang XW, Li QG, Sun CM, et al. GREACE-assisted adaptive laboratory evolution in endpoint fermentation broth enhances lysine production by *Escherichia coli*. *Microb Cell Fact*, 2019, 18: 106.
- [86] Chung SC, Park JS, Yun J, et al. Improvement of succinate production by release of end-product inhibition in *Corynebacterium glutamicum*. *Metab Eng*, 2017, 40: 157–164.
- [87] Thomson NM, Shirai T, Chiapello M, et al. Efficient 3-hydroxybutyrate production by quiescent *Escherichia coli* microbial cell factories is facilitated by indole-induced proteomic and metabolomic changes. *Biotechnol J*, 2018, 13(5): 1700571.
- [88] Tuyishime P, Wang Y, Fan LW, et al. Engineering *Corynebacterium glutamicum* for methanol-dependent growth and glutamate production. *Metab Eng*, 2018, 49: 220–231.
- [89] Bracher JM, Verhoeven MD, Wisselink HW, et al. The *Penicillium chrysogenum* transporter PcAraT enables high-affinity, glucose-insensitive L-arabinose transport in *Saccharomyces cerevisiae*. *Biotechnol Biofuels*, 2018, 11: 63.
- [90] Kawaguchi H, Yoshihara K, Hara KY, et al. Metabolome analysis-based design and engineering of a metabolic pathway in *Corynebacterium glutamicum* to match rates of simultaneous utilization of D-glucose and L-arabinose. *Microb Cell Fact*, 2018, 17: 76.
- [91] Jones JA, Vernacchio VR, Sinkoe AL, et al. Experimental and computational optimization of an *Escherichia coli* co-culture for the efficient production of flavonoids. *Metab Eng*, 2016, 35: 55–63.
- [92] Pappu JSM, Gummadi SN. Modeling and simulation of xylitol production in bioreactor by *Debaryomyces nepalensis* NCYC 3413 using unstructured and artificial neural network models. *Bioresour Technol*, 2016, 220: 490–499.
- [93] Schwechheimer SK, Becker J, Peyriga L, et al. Improved riboflavin production with *Ashbya gossypii* from vegetable oil based on ¹³C metabolic network analysis with combined labeling analysis by GC/MS, LC/MS, 1D, and 2D NMR. *Metab Eng*, 2018, 47: 357–373.
- [94] Wang G, Tang WJ, Xia JY, et al. Integration of microbial kinetics and fluid dynamics toward model-driven scale-up of industrial bioprocesses. *Eng Life Sci*, 2015, 15(1): 20–29.
- [95] de Jonge L, Buijs NAA, Heijnen JJ, et al. Flux

- response of glycolysis and storage metabolism during rapid feast/famine conditions in *Penicillium chrysogenum* using dynamic ^{13}C labeling. *Biotechnol J*, 2014, 9(3): 372–385.
- [96] Zheng XM, Zheng P, Sun JB, et al. Heterologous and endogenous *U6* snRNA promoters enable CRISPR/Cas9 mediated genome editing in *Aspergillus niger*. *Fungal Biol Biotechnol*, 2018, 5: 2.
- [97] Zheng XM, Zheng P, Zhang K, et al. 5S rRNA Promoter for guide RNA expression enabled highly efficient CRISPR/Cas9 genome editing in *Aspergillus niger*. *ACS Synth Biol*, 2019, 8(7): 1568–1574.
- [98] Liu J, Wang Y, Lu YJ, et al. Development of a CRISPR/Cas9 genome editing toolbox for *Corynebacterium glutamicum*. *Microb Cell Fact*, 2017, 16(1): 205.
- [99] Wang Y, Liu Y, Liu J, et al. MACBETH: Multiplex automated *Corynebacterium glutamicum* base editing method. *Metab Eng*, 2018, 47: 200–210.
- [100] Sun DH, Chen JZ, Wang Y, et al. Metabolic engineering of *Corynebacterium glutamicum* by synthetic small regulatory RNAs. *J Ind Microbiol Biotechnol*, 2019, 46(2): 203–208.
- [101] Tong ZY, Zheng XM, Tong Y, et al. Systems metabolic engineering for citric acid production by *Aspergillus niger* in the post-genomic era. *Microb Cell Fact*, 2019, 18: 28.
- [102] Barrett CL, Kim TY, Kim HU, et al. Systems biology as a foundation for genome-scale synthetic biology. *Curr Opin Biotechnol*, 2006, 17(5): 488–492.

(本文责编 陈宏宇)