

• 综 述 •

基于图论和约束优化的异源代谢途径设计方法及应用

余文童^{1,2}, 袁倩倩², 马红武², 王智文¹

1 天津大学 化工学院, 天津 300350

2 中国科学院天津工业生物技术研究所 生物设计中心, 天津 300308

余文童, 袁倩倩, 马红武, 王智文. 基于图论和约束优化的异源代谢途径设计方法及应用. 生物工程学报, 2022, 38(4): 1390-1407.

YU WT, YUAN QQ, MA HW, WANG ZW. Graph-based and constraint-based heterologous metabolic pathway design methods and application. Chin J Biotech, 2022, 38(4): 1390-1407.

摘 要: 构建高产高附加值产品的微生物细胞工厂是代谢工程的研究目标之一, 设计高效的产品合成途径是实现这一目标的重要方式。不同微生物底盘因其代谢能力有所差异, 故而可以利用的底物和生产的产品范围有限。为了扩大其生产能力, 需要进行代谢途径从无到有的设计。传统代谢工程基于经验进行异源途径设计的方式低效且无法确保结果的全面性, 而系统生物学技术的蓬勃发展, 为途径设计提供了新的方法。异源代谢途径设计方法主要包括基于图论的途径设计和基于约束优化的途径设计, 这些方法利用包含丰富代谢反应信息的数据库, 通过图论搜索策略或约束算法, 计算出多条产品合成途径, 进而筛选出最优途径指导菌株改造。本文将对途径设计所依赖的数据库、途径设计算法及其在代谢工程中的应用进行详细介绍, 并讨论现有算法在实际应用中存在的不足和挑战, 为研究者选择不同方法进行产品合成途径设计提供指导。

关键词: 途径设计; 代谢网络; 代谢工程; 图论方法; 约束方法

Received: November 7, 2021; **Accepted:** January 11, 2022; **Published online:** January 13, 2022

Supported by: National Key Research and Development Program of China (2018YFA0901400); National Natural Science Foundation of China (21908239)

Corresponding authors: MA Hongwu. E-mail: ma_hw@tib.cas.cn

WANG Zhiwen. E-mail: zww@tju.edu.cn

基金项目: 国家重点研发计划项目 (2018YFA0901400); 国家自然科学基金 (21908239)

Graph-based and constraint-based heterologous metabolic pathway design methods and application

YU Wentong^{1,2}, YUAN Qianqian², MA Hongwu², WANG Zhiwen¹

¹ School of Chemical Engineering and Technology, Tianjin University, Tianjin 300350, China

² Biodesign Center, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

Abstract: It is among the goals in metabolic engineering to construct microbial cell factories producing high-yield and high value-added target products, and an important solution is to design efficient synthetic pathway for the target products. However, due to the difference in metabolic capacity among microbial chassis, the available substrate and the yielded products are limited. Therefore, it is urgent to design related metabolic pathways to improve the production capacity. Existing metabolic engineering approaches to designing heterologous pathways are mainly based on biological experience, which are inefficient. Moreover, the yielded results are in no way comprehensive. However, systems biology provides new methods for heterologous pathway design, particularly the graph-based and constraint-based methods. Based on the databases containing rich metabolism information, they search for and uncover possible metabolic pathways with designated strategy (graph-based method) or algorithm (constraint-based method) and then screen out the optimal pathway to guide the modification of strains. In this paper, we reviewed the databases and algorithms for pathway design, and the applications in metabolic engineering and discussed the strengths and weaknesses of existing algorithms in practical application, hoping to provide a reference for the selection of optimal methods for the design of product synthesis pathway.

Keywords: pathway design; metabolic network; metabolic engineering; graph-based methods; constraint-based methods

气候变化和化石资源的日益枯竭,促使许多化学品的生产开始从传统的化学制造走向绿色可持续的生物制造。为满足这一需求,代谢工程选取遗传背景清晰的微生物,通过重定向细胞内的代谢流,将底物最大程度地转化为目标产物从而构建高效细胞工厂,以实现化学品和材料,如生物燃料、天然产物和药物等高效生产^[1-2]。然而,不同菌株之间的代谢能力不同,导致在很多情况下,所选宿主菌株并没有利用某些碳源或合成某些产品的能力。为了使其获得这种能力,需要对底物利用或产品合成途径

进行从无到有的设计。传统代谢工程基于经验和文库,通过手动查阅的方式获得相关转化途径,并将其中异源部分引入到相应的底盘菌株中以实现底物利用或产品合成^[3]。但是由于传统的方法并未系统考虑菌株的代谢能力,也未对所有异源的已知反应信息进行全面有效的挖掘,导致只能对途径进行局部设计,无法得到最优的途径。随着系统生物学的不断发展和生物数据的不断积累,研究者们开始从系统水平开展异源途径的理性设计,开发了多种途径设计方法,通过对数据库的全面挖掘,快速高效

地进行异源途径设计。

异源途径设计包含异源的底物利用途径设计和异源的产品合成途径设计,其设计方法主要包括两类:基于图论 (graph-based methods) 的拓扑学方法 (如 MRE^[4], BPFinder^[5]) 和基于约束优化 (constraint-based methods) 的计量学方法 (如 OptStrain^[6], optStoic^[7])。借助这些方法,可以实现从无到有和从有到优两个途径的设计目标^[8]。图论方法通过将代谢网络表示为图,利用相关算法搜索底物到产物的所有可行途径。图论方法的优势在于计算速度快,能够产生大量的预测结果。缺点是往往只能预测线性途径;预测的途径无法达到能量和还原力平衡;只能定性的进行合成途径从无到有的设计,无法定量评估不同途径的转化率,不能提供途径的得率信息;因流通代谢物的存在,导致很多途径并无生物学意义,需要进一步的途径评估和筛选^[9]。基于约束的方法将代谢网络表示为化学计量矩阵,通过线性规划的方式搜索解空间^[10]。相比于图论方法,基于约束的方法避免了不考虑计量学约束的缺点,可以预测出复杂的分支途径,得到碳流、能量和辅因子平衡的代谢途径,同时还可以对途径进行定量分析,评估不同途径对底物的转化率情况,获得不同得率的产品合成途径,指导途径选择,实现途径设计从无到有向从有到优的跨越。但是基于约束方法的计算结果准确性在很大程度上取决于代谢网络质量的高低。对于一个包含数万反应的代谢网络而言,其内错综复杂的反应关系往往会形成一些无效循环问题,导致计算结果出现错误。因此高质量的代谢网络是约束方法计算结果可靠的有效保证。此外,直接在反应数量庞大的代谢网络上进行约束方法求解,在计算性能上也面临着巨大挑战^[11]。

综上所述,基于图论和基于约束的异源

途径设计方法各有优劣,尽管二者均存在不足之处,但可以在系统水平上设计出依据经验无法得到的新途径,加速并提升异源途径设计过程,因此已被广泛应用于代谢工程途径设计中^[12-13]。在实际应用中需要根据途径设计目标以及计算数据集的大小选择合适的方法,也可以将二者结合进行途径设计^[14]。两种方法的实现依托于可靠的酶反应数据库和可行的途径设计方法,因此,本文将对异源途径设计所依赖的数据库,以及近年来开发的基于图论和约束优化的代谢途径设计方法进行综述。

1 途径设计依赖的数据库

异源代谢途径设计依赖的已知生化反应数据库主要有 BiGG^[15]、KEGG^[16]、MetaCyc^[17]和 BRENDA^[18]等,详见表 1。KEGG 和 MetaCyc 是包含多物种已知生化反应及代谢途径的综合知识库,它们为途径设计提供了丰富的反应和代谢物信息,其中 MetaCyc 在反应数量、反应方向、文献出处和学术免费等方面优于 KEGG。BiGG 数据库整合了已发表的 35 个物种的 108 个高质量的基因组尺度代谢网络模型 (genome-scale metabolic network models, GSMNs),涵盖大肠杆菌、恶臭假单胞菌和酿酒酵母等常见模式微生物中的 28 302 个反应,为特定底盘生物的途径设计提供了良好的数据基础。借助 BiGG 数据库中的 GSMNs,可以将反应数据与基因组数据相结合,以发现菌株实现新途径所需的异源反应,并且能够通过预测产品得率等信息对途径进行评估。然而,不同数据库间反应和代谢物的 ID 并不一致,导致在整合不同来源的反应和代谢物信息时,面临着 ID 匹配问题。为此,研究者们陆续构建了 BKM-react^[19]、MetRxn^[20]、Rhea^[21]、MetaNetX^[22]等数据库,将不同数据来源的反应和代谢物的 ID 归一化,试图解决此类

表 1 途径设计使用的数据库

Table 1 Databases used for pathway design

Databases	Website	Description
BiGG	http://bigg.ucsd.edu/	A database with a set of stardized identifiers, which integrates 108 genome-scale metabolic network models
KEGG	https://www.kegg.jp/	An integrated database consisting of eighteen original databases in four categories. The database contains 541 pathways, 18 749 metabolites, 11 467 reactions and 7 276 genomic information
MetaCyc	https://metacyc.org/	A curated database of experimentally elucidated metabolic pathways from all domains of life. It contains 2 879 pathways, 16 861 metabolites, 16 986 reactions from 3 185 different organisms
BRENDA	https://www.brenda-enzymes.org/	A database of enzyme functional data, which contains over 5 million data for about 9 000 enzymes from 13 000 organisms
Rhea	https://www.rhea-db.org/	An expert-curated knowledgebase consisting of 11 638 metabolites and 13 362 reactions information, which using the ChEBI to describe metabolites
MetaNetX	https://www.metanetx.org/	An integrated database consisting of information from BiGG, ModelSEED, BioCyc and other databases, which using MNXref to describe reactions and metabolites
MetRxn	http://ec2-54-191-241-202.us-west-2.compute.amazonaws.com/MetRxnTest.php	An integrated database consisting of information from BRENDA, KEGG, BioCyc and BKM-react
BKM-react	http://bkm-react.tu-bs.de/	An integrated and non-redundant biochemical reactions database containing known enzyme-catalyzed and spontaneous reactions from BRENDA, KEGG, MetaCyc and SABIO-RK

问题。相比单一数据库，这些复合数据库包含更为丰富的生化反应信息，为途径设计提供了更多可能。

2 基于图论的途径设计

2.1 基于图论方法进行代谢途径设计的原理概述

基于图论的方法源自拓扑学中的图理论，通过将代谢网络表示为图，利用相关算法预测底物到产物的所有可行途径。代谢网络常见的图表示形式见图 1，主要有如下 4 类：(1) 代谢物图：节点表示代谢物，边表示连接两个代谢物的反应；(2) 反应图：节点表示反应，如果一个代谢物既是节点反应 R1 的产物，又是节点反应 R2 的底物，则存在一条边连接 R1 和 R2；(3) 二分图：节点表示代谢物和反应，边将代谢物和其参与的反应连接起来；(4) 超

图：边表示反应，用于连接参与此反应的所有代谢物^[23]。

2.2 基于图论的代谢途径设计方法及应用

图论算法以搜索点到点的距离为目标，从而发现特定距离（即反应步骤）内所有可能的途径。表 2 列出目前已发表的主要的图论方法，其中 LPAT^[24]、MRE^[4]、PATH^{cre8[25]}、BPFinder^[5]提供了在线网站以供生物学家使用，减少了软件使用学习的技术壁垒。利用图论方法进行途径设计时，因流通代谢物（如：ATP、H₂O、NADH）在代谢网络中具有高连通性，导致预测结果出现大量通过流通代谢物连接的无生物学意义的途径，不利于后续的筛选和分析，研究者们针对流通代谢物提出了不同的处理方式并开发了相关图论算法。本文对已有图论方法进行总结，依据对流通代谢物处理方式分为 5 类，并介绍每种方法的优缺点。

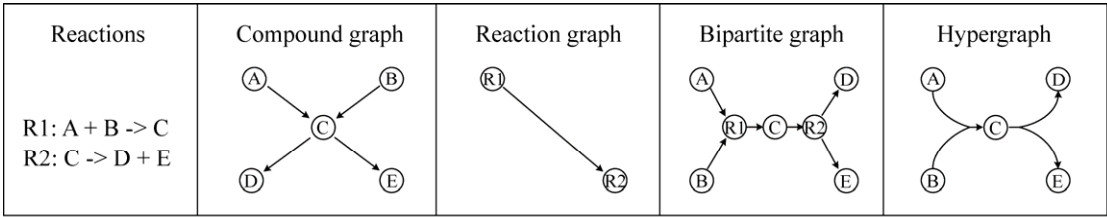


图 1 代谢网络的图形表示
Figure 1 Graphical representation of metabolic networks.

表 2 基于图论的代谢途径设计方法
Table 2 Graph-based metabolic pathway design methods

Type of dealing with currency metabolites	Methods	Database	Search algorithm	Network representation	Pathway ranking	Access	References
Removal	FMM	KEGG	Breadth-first search (BFS)	Compound graph	Compare pathway across organisms	Closed access	[26]
	PATH ^{cre8}	KEGG	Yen's loopless k-shortest path with A* heuristic	Bipartite graph	Thermodynamics, compound toxicity, product consumption, enzyme copy number	Open access web server: https://www.cbrc.kust.edu.sa/pathcre8/	[25]
Weighted graph	Croes et al.	KEGG	DFS	Bipartite graph	Weight of pathway	No access	[27]
Structure similarity	PHT	KEGG	BFS with HOHL (higher-order horn logic)	Compound graph	Structure similarity and pathway length	Closed access	[28]
Atom mapping	ReTrace	KEGG	Heuristic search	Bipartite graph	Atom conservation and pathway length	Free download	[29]
	LPAT	KEGG	BPAT-M	Bipartite graph	Atom conservation and pathway length	Open access web server: http://metabolicpaths.kavrakilab.org/#lpatrun	[24]
Combination methods	Faust et al.	KEGG	Jimenez's k-shortest paths	Bipartite graph	Wilcoxon paired signed-rank test and pathway length	Closed access	[30]
	MRE	KEGG	Yen's loopless k-shortest path	Compound graph	Thermodynamics and genes from host organism	Open access web server: https://www.cbrc.kust.edu.sa/mre/	[4]
	AGPathFinder	KEGG	CAGTG	Compound graph	Reaction thermodynamics and compound similarity	Closed access	[31]
	BPFinder	KEGG	BPFinder	Compound graph	Compound similarity, thermodynamic feasibility and conserved atom groups	Open access web server: http://114.215.129.245:8080/atomic/	[5]

2.2.1 流通代谢物移除

移除流通代谢物是基于图论进行途径设计时,减少设计的无生物学意义途径常用的方法。2003年, Ma 等^[32]就提出了在利用图论方法进行途径设计时需要去除流通代谢物,并提供了流通代谢物的集合。之后陆续开发的图论方法也都不同程度地进行了流通代谢物的删除,如 Chou 等^[26]开发的基于 KEGG 数据库的在线途径预测工具 FMM,该工具将 KEGG 数据库中的生化反应表示为代谢物图,并手动删除经过流通代谢物连接的途径,但当流通代谢物作为主要代谢物时,其途径予以保留,以此避免相关合成途径的丢失。Chou 等^[26]以大肠杆菌为底盘,利用此工具模拟设计植物次级代谢物柚皮素的合成途径,预测了 4-香豆酸到柚皮素的合成途径,以大肠杆菌 4-香豆酸为中间体,引入 3 种外源酶 (4-香豆酸辅酶 A 连接酶,查尔酮合酶和查尔酮异构酶),可以实现柚皮素的合成,这与 Watts 等的实验报道一致^[33]。但这种通过手动删除流通代谢物连接的过程太过繁琐,且常常会因网络过于复杂而漏删无意义的连接。为此,研究人员通常会预先定义一个流通代谢物列表,然后在生成代谢网络图的时候自动删除对应的流通代谢物节点,而前期的研究方法也为流通代谢物列表的确定提供了很好的基础。

2020 年, Motwalli 等^[25]开发了 PATH^{cre8}工

具,该工具根据 KEGG 数据库中代谢物参与反应的数目筛选出 43 个流通代谢物,之后在将代谢网络转化为代谢物图时直接去除这些流通代谢物以减少无意义的连接。作者成功预测了异戊二烯合成途径和可卡因降解途径,说明了 PATH^{cre8} 的有效性。利用该工具进行途径设计时,仅需输入底物、产物和宿主即可。但是该方法存在一个默认值为 11 的终止参数用于限制预测的途径长度,即预测的途径中不能超过 5 步反应和 6 个代谢物。这也是 Motwalli 在文章中分两步预测异戊二烯的原因所在。这个参数导致该方法无法直接用于预测合成途径较长的产品。此外,PATH^{cre8} 预先确定的流通代谢物列表范围过窄,导致其预测结果仍然存在错误连接。如图 2 所示,以大肠杆菌为底盘菌株,预测苯丙氨酸到白藜芦醇的合成途径,PATH^{cre8} 预测的最优途径中碳酸氢根参与了途径连接,而实际上其并不作为主代谢物参与反应,且大多数情况下,碳酸氢根也应被定义为流通代谢物。所以,在使用 PATH^{cre8} 时,用户还需根据计算结果确定额外需要删除的流通代谢物,然后在参数输入界面额外添加这类需要删除的流通代谢物。虽然通过这种方式可以一定程度上减少无生物学意义途径的产生,但是该方法还存在无法识别主次代谢物的局限。如图 2 所示,合成途径最后一步反应 R01614: 3 malonyl-CoA+

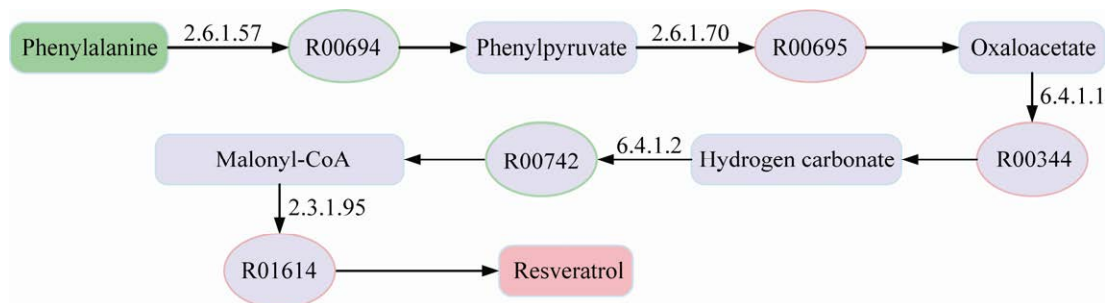


图 2 PATH^{cre8} 预测的苯丙氨酸合成白藜芦醇的最优途径

Figure 2 Optimal pathway for the synthesis of resveratrol with phenylalanine predicted by PATH^{cre8}.

$\text{p-coumaroyl-CoA} \rightleftharpoons 4 \text{ CoA} + \text{resveratrol} + 4 \text{ CO}_2$, 其中作为主代谢物的应该是 p-coumaroyl-CoA, 而通过删除流通代谢物并不能避免图中的错误连接, 因此, 直接去除流通代谢物的方法在产生正确途径的同时也伴随着生成大量错误的途径, 需要用户进一步甄选正确途径。

2.2.2 代谢网络加权

虽然去除流通代谢物可以在一定程度上减少无生物学意义途径的发生, 但因流通代谢物的定义不详细, 导致流通代谢物的确定往往基于经验, 不够全面。为了避免预先定义流通代谢物, Croes 等^[27]提出了代谢网络的加权方法。将代谢网络表示为一个加权图, 对其中的每个代谢物分配一个权重参数, 其值为代谢物在网络中参与的反应数目。然后再通过搜索算法, 找到权重最小的一条或多条途径。Croes 等建立了两套用于比较的数据集, 一套为 KEGG 数据库和 aMAZE^[34]中的 56 条已知途径, 另一套为 EcoCyc^[35]数据库和大肠杆菌中 92 条已知途径。利用这两套数据集和搜索算法, 比较了原始网络、删除流通代谢物网络和加权网络在途径设计方面的准确性, 最终发现基于加权网络进行途径设计的结果准确性最高。但是此类加权方法也存在一定的局限: 首先, 流通代谢物没有从网络中删除, 只是赋予高权重, 使其出现的概率降低, 导致其仍会出现在部分途径中; 其次, 一些中心代谢物 (如丙酮酸) 在此网络中也会被分配较高权重, 但其并非流通代谢物, 这就会造成经过这些代谢物的反应不易在途径中发生。因此, 简单地根据代谢物参与反应的数目附加权重的方式只能在一定程度上解决流通代谢物问题。

2.2.3 代谢物结构相似性

与加权网络的目的类似, 研究人员也可以借助代谢物结构信息来避免预先定义流通代谢

物。2005 年, Rahman 等^[28]基于 KEGG 数据库提出了借助化学结构信息的途径设计工具 PHT (pathway hunter tool)。他们使用来自 CDK (chemistry development kit)^[36]的指纹算法, 将二维的化学结构信息转换为一维的二进制流, 通过指纹进行快速的相似性搜索。同时, 利用指纹, 通过结合代谢物的原子量和 Tanimoto^[37]算法得到归一化评分函数来计算两个分子之间的相似性, 从而避免了途径中的假阳性连接, 使得到的结果更具生物学意义。使用广度优先搜索算法, 他们成功地找到了大肠杆菌中葡萄糖到丙酮酸的最短途径, 说明了方法的有效性。这类方法也存在一个缺点, 它们依赖于分子结构数据, 而在现有数据库中, 并非所有代谢物都有结构信息, 这就导致会因缺乏结构信息而丢失部分合成途径, 但整体设计结果的准确性相较于前两种策略有了很大的提升。

依据代谢物的结构信息, 研究者搭建了反应物对数据库 RPAIR^[38], 用于描述反应中底物到产物之间的原子转移。该数据库基于 KEGG 数据库反应数据, 使用分子对齐的化学结构比较方法, 提取得到了 9 125 个反应物对信息, 并将其分为 5 个类型: (1) main-pairs: 描述底物的主要变化; (2) cofac-pairs: 描述氧化还原酶辅因子的变化; (3) trans-pairs: 描述转移酶的转移基团变化; (4) ligase-pairs: 描述用于连接酶的三磷酸核苷的消耗; (5) leave-pairs: 描述用于分解酶和水解酶等酶的无机化合物的分离和添加。根据此数据库, 后续基于 KEGG 数据库开发的途径设计方法就能在很大程度上避免了无生物学意义的连接。

2.2.4 原子映射

随着原子映射数据的完善和自动生成正确原子映射的计算工具^[39-40]的发展, 原子映射逐渐成为流通代谢物处理的常用方法, 其表示方

式如图 3 所示。如在代谢物 C00074 和 C00022 之间存在的碳原子映射关系为 C00074:C1->C00022:C1、C00074:C2->C00022:C2、C00074:C3->C00022:C3，借助这种映射关系，即可追踪每一步反应过程中的原子转移，有利于得到最相关代谢物的连接方式。

利用原子映射也可以在一定程度上解决上述方法无法获得分支途径的局限。ReTrace 是第一个利用原子映射信息查找分支途径的工具^[29]，该工具基于 Arita^[41]的 ARM 方法原则，即认为一个有生物学意义的途径应该至少将一个原子从底物转移到产物中。ReTrace 在 Arita 方法基础上进行了改进，通过确保底物到产物之间原子转移数量最大，来发现两者之间的最短合成途径，并将这些最短途径进行组合以发现分支途径和高得分途径。利用此方法，研究人员在里氏木霉 (*Trichoderma reesei*) 中重构了精氨酸和赖氨酸合成途径^[42]。

2011 年，Heath 等^[24]基于 KEGG 数据库开发了在线工具 LPAT。该工具先利用原子映射识别线性代谢途径，然后将这些线性途径合并为分支途径。利用此工具，Heath 成功预测了葡萄

糖-6-磷酸合成水苏糖的途径，并且和 ReTrace 预测结果进行了比较，发现 ReTrace 预测的合成途径中存在着多余的反应，这些反应主要是由于代谢物之间产生循环导致。相比流通代谢物的直接删除，利用原子映射获得的结果更为准确。利用 LPAT 进行途径设计时需要输入底物、产物和碳转移数目。其中碳转移数目默认为算法自动识别的最大数，也可人为定义数目，适当降低碳转移数目有助于发现更好的途径设计结果。当使用 LPAT 预测苯丙氨酸合成白藜芦醇的途径时，可得到 12 条合成途径，这些途径都有生物学意义，其中最优途径如图 4 所示。但该网站的数据已经很久未更新，导致无法获得 KEGG 数据库新收录的一些途径，并且此工具没有基于特定底盘菌株，导致将预测到的途径应用于特定底盘菌株时，需要额外判别其中的异源部分。

2.2.5 组合方法

相较于单一的处理方法，组合方法在途径设计中更受欢迎且效果更佳。如 Faust 等^[30]在 Croes 等^[27]的方法进行了改进，通过结合加权网络和 RPAIR 数据库中包含的 KEGG 反应中反应

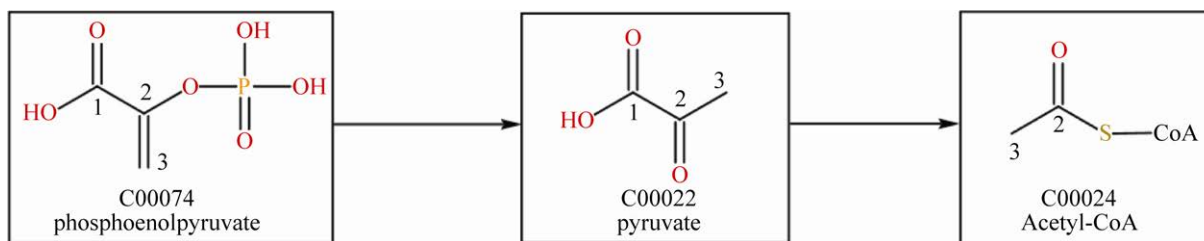


图 3 原子映射示例图

Figure 3 Example of atom mapping. Only the mapping information for carbon atoms is displayed.

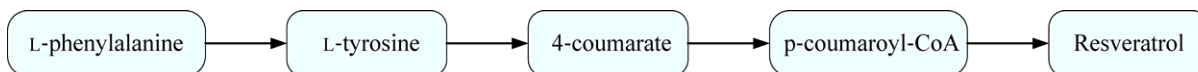


图 4 LPAT 预测的苯丙氨酸合成白藜芦醇的最优途径

Figure 4 Optimal pathway for the synthesis of resveratrol with phenylalanine predicted by LPAT.

对的注释,实现了更为精准的预测。Faust 选取大肠杆菌、酿酒酵母和智人中的 55 条已知途径作为参考,选取这些途径的底物和产物作为输入,对 KEGG 代谢网络进行 3 种不同参数(是否删除流通代谢物、是否根据代谢物参与反应数分配权重、是否根据 RPAIR 注释分配权重)的组合处理,在处理后的网络中利用开发的算法进行途径搜索,并将搜索得到的途径与参考途径进行对比,最终发现 RPAIR 注释和代谢物加权的结合可以显著提升途径设计结果。

2016 年, Kuwahara 等^[4]开发的在线途径设计工具 MRE,通过预先删除 101 种流通代谢物并考虑 RPAIR 化学转化类型(默认 main、cofac、trans)来实现流通代谢物的处理。该工具利用 KEGG 数据库信息,通过加权的方式构建宿主依赖的代谢网络,即通过波尔曼兹分布计算每个反应转化前体的概率,并将此概率的对数值作为权重赋予到代谢物图的边上,从而在进行途径设计时可以区分本源和外源反应。利用该工具进行途径设计时,需要输入的必需参数包括底物、产物和宿主。该工具也提供了高级选项用于用户定义计算结果,包括合成途径中的最大反应数目(上限 20)、计算结果中的最大合成途径数目(上限 500)、RPAIR 类型和

需要删除的代谢物列表。在高级选项默认设置下,利用此工具预测了大肠杆菌中苯丙氨酸合成白藜芦醇的途径,所得到的最优途径与 LPAT 预测结果一致。但其也存在无法正确判别反应中的主次代谢物的问题。图 5 展示了预测结果中的次优途径,可以看出 MRE 和 PATH^{cre8} 都在反应 R01614 中无法判断主要的代谢物转化,这是因为它们都没有考虑原子映射信息,从而说明了原子映射在途径设计结果准确性方面的重要作用。

之前基于原子映射数据进行途径设计时,往往需要预定义追踪的目标原子,而这会导致无法找到那些无目标原子转移的替代途径。为此, Huang 等^[31]开发了 AGPathFinder 方法,通过原子团的映射数据进行途径搜索,并引入反应热力学信息和代谢物结构信息作为约束,使得计算所得途径更具生物学意义。之后, Huang 等^[5]又在此基础上进行了改进,提出了 BPFinder,用于发现合成途径中的分支途径。该方法通过跟踪代谢网络中原子团的运动来枚举线性代谢途径,并将线性原子团保守途径合并为分支途径。通过使用化合物相似性、热力学可行性和保守原子团的综合信息对候选的分支途径进行排名。利用 BPFinder 进行途径设计时,需要输

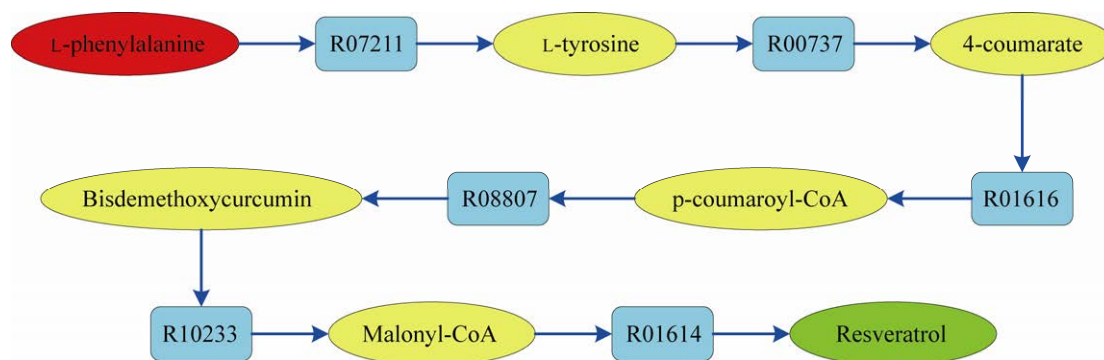


图 5 MRE 预测的苯丙氨酸合成白藜芦醇的次优途径

Figure 5 Suboptimal pathway for the synthesis of resveratrol with phenylalanine predicted by MRE.

入底物、产物、最小原子团数目和合并策略。其中最小原子团数目和合并策略的确定并不容易,根据 Huang 等在文章中的描述,一般最小原子团数目可以设置为 2 或 3,对于合并策略而言,可以默认网站设置,即优先选择非重叠合并,在非重叠合并无解的情况下选择重叠合并。

在设置最小原子团数目为 2,默认合并策略的情况下,利用 BPFinder 预测苯丙氨酸合成白藜芦醇的途径,可以得到 14 个解,其中最优途径如图 6A 所示,可以看出乙酸作为主要代谢物参与了途径连接。而实际上,反应 R01308 ($4\text{-hydroxybenzoate} + \text{acetate} + \text{NADH} + \text{H}^+ \rightleftharpoons 4\text{-coumarate} + 2\text{H}_2\text{O} + \text{NAD}^+$) 和 R01426 ($\text{benzoate} + \text{acetate} + \text{NADH} + \text{H}^+ \rightleftharpoons \text{trans-cinnamate} + 2\text{H}_2\text{O} + \text{NAD}^+$) 是两个水解反应,其中乙酸只是副产物,所以该方法采用的原子团映射还存在部分问题。此外,预测出的第 4 条途径如图 6B

所示,可以看出这条途径相比最优途径更具有生物学意义,但是反应 R00691 ($\text{L-arogenate} \rightleftharpoons \text{L-phenylalanine} + \text{H}_2\text{O} + \text{CO}_2$) 实际上并不能走固碳方向,这说明该方法没有对 KEGG 数据库的反应方向进行有效的修正。总而言之,利用 BPFinder 进行途径设计,在大多数情况下还是可以为用户提供准确有效的合成途径。

3 基于约束的途径设计

3.1 基于约束优化方法进行代谢途径设计的原理概述

约束方法基于代谢网络的计量学平衡,能够设计满足物质、能量、还原力平衡的最优合成途径。通过将代谢网络表示为化学计量矩阵,使用线性规划的方式搜索解空间,计算满足特定约束条件下的产品最优合成途径。最普遍使用

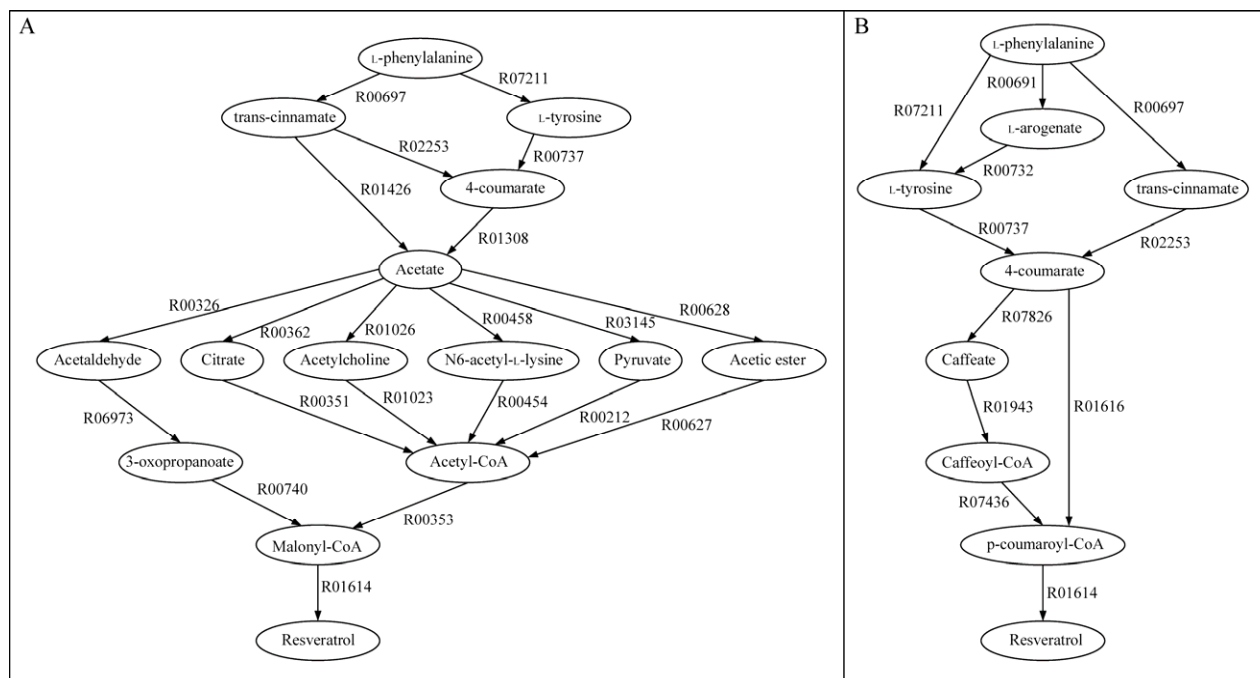


图 6 BPFinder 预测的苯丙氨酸合成白藜芦醇的途径

Figure 6 Pathway for the synthesis of resveratrol with phenylalanine predicted by BPFinder. (A) The part of optimal pathway. (B) The fourth pathway in the results.

的方法是通量平衡分析 (flux balance analysis, FBA^[10]) (图 7), 它以代谢网络计量矩阵和反应通量上下限作为约束, 以生长或某一产物生成速率为优化目标, 搜索满足最大化产品生成时的可行途径。这类方法因可以直接预测出碳流、能量和辅因子平衡的代谢途径, 且得到的途径数目远远少于图论方法, 更适合于代谢途径设计。

3.2 基于约束的代谢途径设计方法及应用

基于约束进行产品合成异源途径设计的方法可以根据其是否基于底盘菌株分为两类: 基于特定底盘的途径设计方法和基于数据库的途径设计方法。其中基于底盘菌株的方法需要以高质量的底盘菌株代谢网络模型为基础, 然后

在其中引入数据库来源的异源反应实现途径搜索。目前已经开发了多种基于约束算法的异源途径设计方法 (表 3), 这些方法在实际应用中取得了一定成效。

3.2.1 基于数据库的代谢途径设计

2015 年, Chowdhury 等^[17]基于 MetRxn 数据库开发了 optStoic 算法, 该算法旨在设计与产品转化整体化学计量相匹配的从头合成途径, 通过两步程序实现: 首先, 通过元素平衡和电荷平衡、热力学可行性等约束识别最佳的整体化学计量 (optStoic); 然后通过网络化学计量、反应方向性和交换反应通量等约束最小化反应数目 (minRxn) 或总代谢通量 (minFlux),

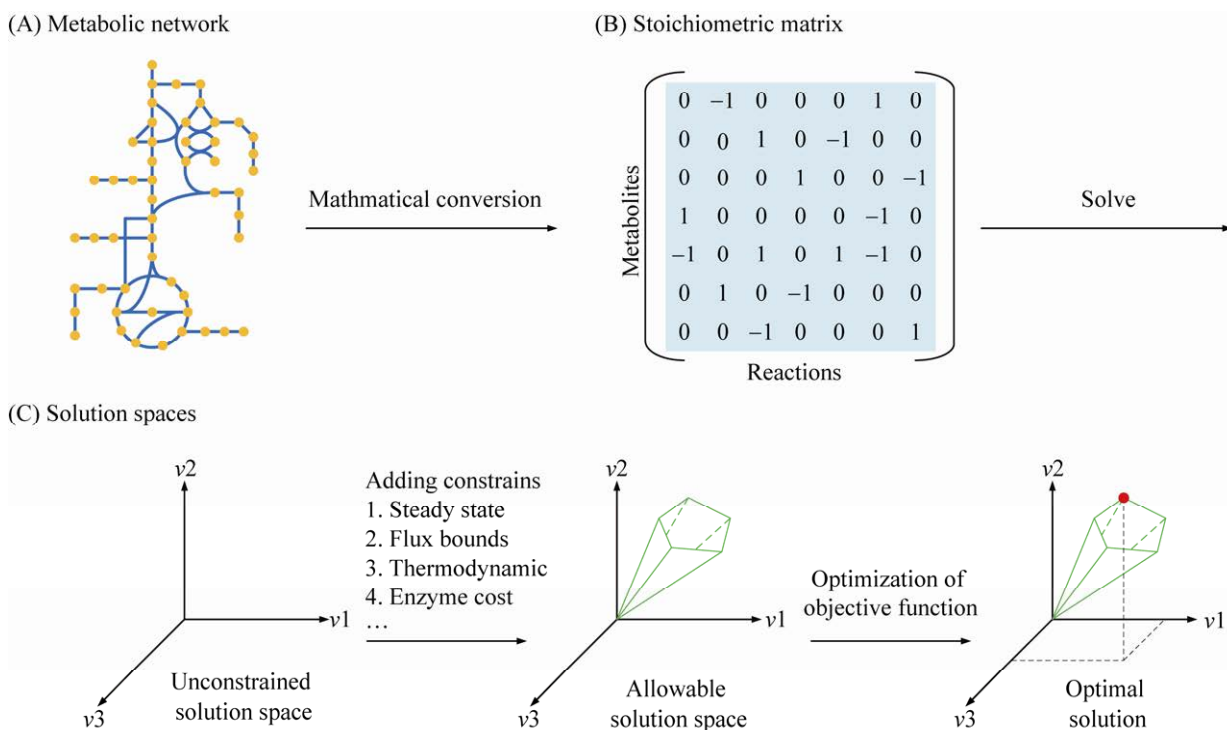


图 7 基于约束方法进行途径设计的流程^[43]

Figure 7 Workflow of pathway design based on constraint methods^[43]. (A) Integrate metabolic reactions information from one or more species to construct metabolic network. (B) Transfer metabolic network to stoichiometric matrix (S matrix), where rows represent metabolites, columns represent reactions and numbers represent metabolite coefficients in reactions. (C) The flux distribution can be obtained by using the S matrix and the objective function of the model. Adding additional constraints can reduce the solution space in which objective function optimization can be performed to find one or more optimal solutions (red dot).

表 3 基于约束的理性途径设计方法

Table 3 Constraint-based rational pathway design methods

Methods	Database	Optimization features	References
OptStrain	KEGG	Maximize product theoretical yield, minimize the number of non-native reactions and maximize growth-coupled production	[6]
SimOptStrain	KEGG	Maximize growth-coupled production	[44]
optStoic	MetRxn	Optimize overall stoichiometry and minimize the number of reactions (minRxn) or the total metabolic flux (minFlux)	[7]
PathTracer	BiGG	Minimize the number of overall transfers and elemental transfers, maximize the overall transfer scores, and minimize pathway length	[45]
comb-FBA	MetaCyc, ATLAS ^[46]	Maximize product yield and minimize the total flux	[47]

以满足第一步得到的最佳的整体化学计量。其中 minRxn 需要直接解决 MILP 问题,这在计算上具有挑战;而 minFlux 则是在第一次迭代计算时解决 LP 问题,后续迭代再解决越来越复杂的 MILP 问题,相比 minRxn,其计算时间得到了有效缩减。Chowdhury 利用此方法发现了 6 条将葡萄糖完全转化为乙酸的合成途径,其中包括了 Bogorad 等构建的 NOG 途径^[48]。此外,还利用此方法设计甲醇与 CO₂ 共利用生产二碳化合物(乙酸、3-羟基丁酸、2-酮异戊酸和间苯三酚)和甲烷转化为乙酸的热力学可行途径,在预测结果中均得到了实验验证的已知途径和新途径。

2019 年,本研究组基于 MetaCyc 数据库和 ATLAS 数据库开发了 comb-FBA 算法^[47],该算法通过组合迭代的方式进行多次 pFBA 计算,从而获得多条产品合成途径。利用该方法,本研究组模拟计算了甲醇同化途径。当仅用 MetaCyc 数据库时,可以得到已被报道的步骤短、无能耗的途径 MCC (图 8A);当使用 MetaCyc 和 ATLAS 数据库时,可通过途径热力学和动力学分析筛选出了 3 条最优途径,进行实验验证,最终体外构建了新的一碳同化途径 GAA 途径 (图 8B),并经过过程优化实现了 88% 的碳得率。当使用图论方法进行甲醇同化途径设计时,其计算结果的最优解仅为 MCC 途径中

的线性合成部分 (图 8C),无法获得中间产物和副产物的回收利用途径,这也体现了约束方法在途径设计上可以提供全局最优解的优势。

3.2.2 基于特定底盘的代谢途径设计

相比基于数据库的代谢途径设计,基于底盘的代谢途径设计方法在指导底盘菌株设计方面更有优势。如 OptStrain^[6]通过在底盘菌株大肠杆菌模型中引入 KEGG 这一包括多物种代谢反应信息的数据库来指导异源代谢途径设计,实现产品得率最大化。该算法通过两步优化程序来寻找最优合成途径:最大化目标产品通量和最小化异源反应数目。然后将得到的最优异源途径添加到宿主底盘的代谢模型中,利用 OptKnock^[49]识别并敲除与目标产品竞争的反应,从而进一步扩大目标产品的通量。他们将此方法应用于大肠杆菌产香兰素中,实现了大肠杆菌高产香兰素的策略,最终葡萄糖产香兰素的得率达到 6.79 mmol/(g DW·h)。随后, Kim 等^[44]开发了 SimOptStrain 算法,以解决 OptStrain 可能会因没有同时评估反应添加和敲除而失去更高产量的策略,从而扩大了途径设计的范围。但是 OptStrain 和 SimOptStrain 都是基于混合整型线性规划 (mixed-integer linear programming, MILP) 方法进行求解,所以在对大规模网络进行计算时,其消耗的计算资源较多,计算时间较长。

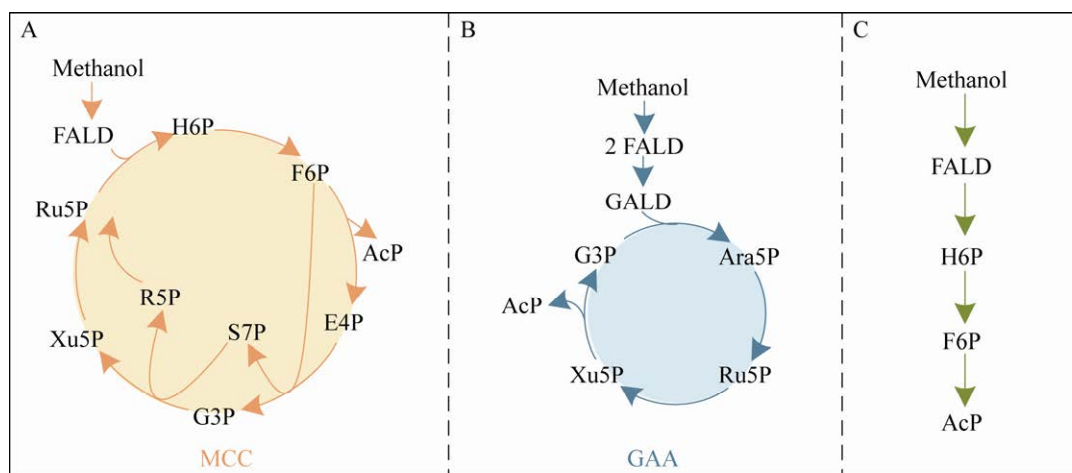


图 8 基于不同方法的甲醇同化途径预测结果

Figure 8 Prediction results of methanol assimilation pathways based on different methods. (A) Using comb-FBA based on MetaCyc. (B) Using comb-FBA based on MetaCyc and ATLAS. (C) Graph-based methods. FALD: formaldehyde; H6P: hexulose 6-phosphate; F6P: fructose 6-phosphate; AcP: acetyl-phosphate; E4P: erythrose 4-phosphate; S7P: sedoheptulose 7-phosphate; G3P: glyceraldehyde 3-phosphate; R5P: ribose 5-phosphate; Xu5P: xylulose 5-phosphate; Ru5P: ribulose 5-phosphate; GALD: glycolaldehyde; Ara5P: arabinose 5-phosphate.

为了扩展解空间,同时加快约束方法计算速度和简化结果分析,Tervo 等^[45]尝试将图论方法与约束方法结合,提出了 MapMaker 和 PathTracer 方法。其中 MapMaker 用于确定碳转移,方便结果分析;PathTracer 用于途径设计,可以找到两个代谢物间最短最可行的途径。利用此方法,他们对大肠杆菌的代谢工程潜力进行了研究,即大肠杆菌可以产生哪些有价值的化学品及所需的本源和外源反应^[14]。通过整合 KEGG 数据库和 iJO1366 模型中的反应,构建了一个有 5 166 个代谢物和 6 992 个反应的复合模型。在此模型中利用 FBA 计算了葡萄糖到 5 166 个代谢物的合成途径,发现其中 2 527 个代谢物可以获得合成途径,其中包括 750 个大肠杆菌可以天然合成的代谢物以及需要引入异源反应才能合成的 1 777 个代谢物。然后通过 MapMaker 和 PathTracer 分析 1 777 种外源产品中 279 种具有商业价值产品的合成途径和前体,发

现异戊烯基二磷酸酯 (isopentenyl diphosphate)、丙酮酸和草酰乙酸是这些产品生产最常用的前体,由它们生成高经济价值产品时需要引入的异源途径最短。

我们采用上述基于约束的 OptStrain 方法,以大肠杆菌为底盘菌株,计算从葡萄糖到白藜芦醇的合成途径,如图 9 所示。该途径与已报道的大肠杆菌利用葡萄糖合成白藜芦醇的途径一致^[50-51]。相比于图论方法,基于约束的异源途径设计方法可以设计得到产品得率最高的途径,并且可以设计出从源头底物(如葡萄糖)到终端产品的合成途径。

4 总结与展望

随着自动化设计-构建-测试-学习 (design-build-test-learn) 流程^[52]的提出,异源途径设计作为流程基础的一部分在代谢工程领域得以广泛应用。与早期代谢工程基于经验的异源途径

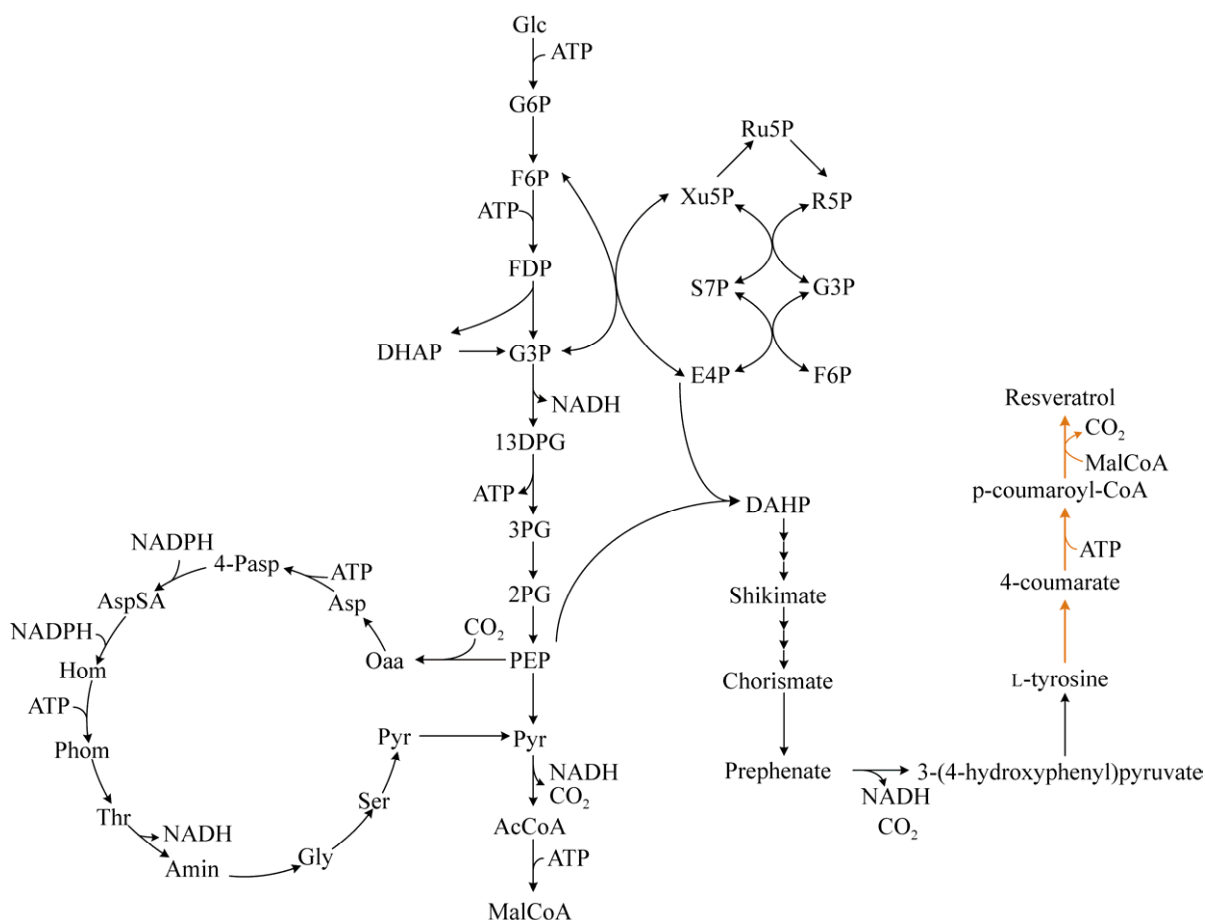


图9 OptStrain 预测的葡萄糖合成白藜芦醇的途径

Figure 9 Pathway for the synthesis of resveratrol with glucose predicted by OptStrain. The heterologous reactions are marked in orange.

设计相比, 基于代谢网络开发的异源途径设计算法以其快速、综合和系统的优势而备受青睐。近年来, 异源途径设计算法的发展十分迅速, 在预测底物利用和产品合成途径上取得了显著成功, 本节对途径设计算法在实际应用场景中的选择提供了指导流程, 详见图 10, 以便于读者可以根据实际需求选择最合适的方法进行途径设计。此外, 本节也对途径设计结果准确性和其在实际应用中存在的问题和挑战进行了总结和展望。

首先, 基于图论的异源途径设计方法在结果准确性方面主要受到流通代谢物处理的影

响。现有的流通代谢物处理方法中, 原子映射方法越来越受到关注, 通过原子映射可以最大程度地处理反应中主次代谢物的对应关系, 并且在合并线性途径得到分支途径方面也具有一定优势, 因此原子映射在后续的流通代谢物处理中将受到广泛应用。但现有数据库中许多反应都缺乏原子映射信息, 这就导致利用原子映射数据处理流通代谢物仍存在不足。而基于机器学习自动生成正确的原子映射信息有望解决这一问题, 如 AMLGAM^[53]等工具。此外, 应用图论方法进行途径设计时, 还存在如下局限: 第一, 图论方法往往是从某个合成前体如丙酮

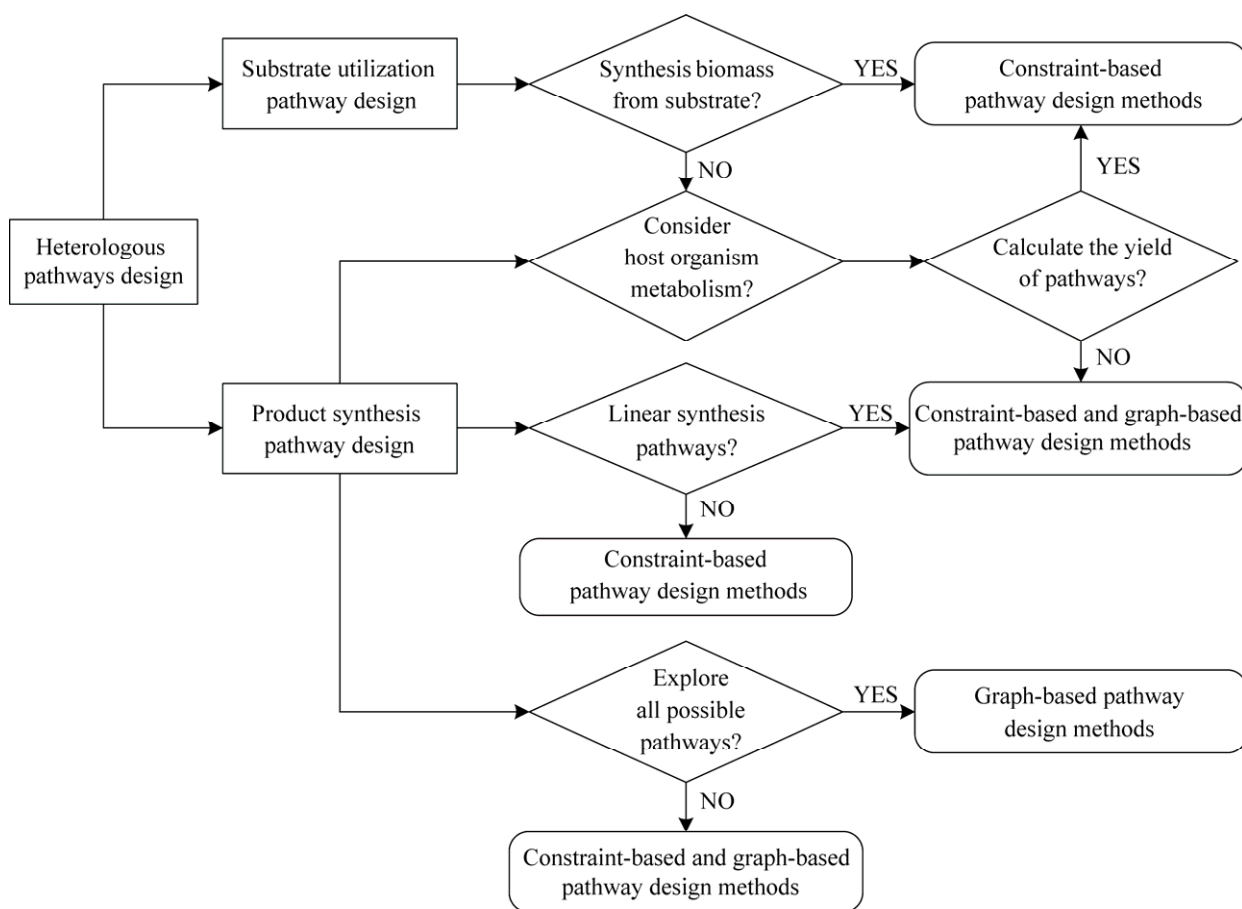


图 10 指导途径设计方法选择的流程图

Figure 10 Flow chart of guiding the selection of pathway design methods.

酸等出发找可能的途径,如果直接从葡萄糖等底物出发计算,会得到无数可能解,且大多数解在生物学上不可行;第二,针对复杂途径,如多前体、长途径、涉及共底物和副产物等情况,图论方法的计算结果可靠性较低。相比之下,约束方法在解决这些问题方面具有一定优势。

其次,基于约束的异源途径设计方法主要受到模型质量的限制。不管是单一物种的基因组尺度代谢网络模型还是跨物种的复合网络模型,都会因模型内部反应彼此交互而产生各种错误,这些错误会使得预测出来的产品合成途径存在碳源无限供给、能量和还原力无限生成等问题。因此,需要对代谢网络模型进行一定

程度的质控,以提高其质量。为此,研究人员提出了各种模型质量提高的方法。如 Yuan 等^[54]指出低质量的模型不仅导致途径预测结果可能存在大量错误,也会导致完全错误的定量预测结果,并提出通过检查能量和还原力无限生成错误来修正模型。另外,本研究组参与了由丹麦技术大学诺和诺德基金会生物可持续性中心发起模型质量检测控制倡议,提出一套由社区维护的标准化代谢模型质量自动测试流程——MEMOTE^[55]。这些都为模型质量的提升提供了很大助力。

再次,这两类方法在异源途径设计时都会产生很多可能的合成途径。因此需对得到的合

成途径进行全面的评估和排名,以便于途径的筛选和验证。常见的途径评估标准有:产品得率、途径长度、外源反应数目、热力学可行性、酶资源可用性、宿主兼容性和中间代谢物毒性等。对于图论方法而言,还可通过反应化学计量对途径进行评估和筛选。虽然现有方法也采用了部分标准进行途径评估,但都不够系统全面。因此,如何制定切实可行的组合策略进行途径评估,是之后异源途径设计需要解决的问题,一些研究者也对此提出了相关建议^[56-58]。

另外,异源途径从理论设计走向实际构建,还需要解决酶的选择问题。当为底盘菌株设计出异源的底物利用或产品合成途径后,如果途径中的外源反应在其他物种中存在,则只需从亲缘关系近的物种中提取相关酶进行实验转化即可。但如果外源反应天然无法发生,则需要对旧酶的改造和新酶的挖掘。蛋白质工程在这过程中发挥着重要作用,它通过序列比对和分子模拟等方法对现有酶的活性和专一性进行改变,从而设计新酶填补空白^[59]。随着人工智能技术的发展,未来有望加速这一过程^[60-61]。

最后,异源途径设计方法还可以与其他优化算法进行结合,进一步获得代谢网络优化的工程策略。主要思路为:首先,借助异源途径设计算法得到目标产品异源合成所需反应;然后将这部分反应整合进底盘菌株的代谢网络模型之中;最后在整合得到的模型基础上进行其他优化算法的实现。当与菌株优化算法(如OptKnock等)结合时,可以在产品能够合成的基础上,得到合成途径优化的改造靶点,通过过表达、弱化或敲除相关途径节点,以实现产品得率的进一步提升。除了这种静态设计外,还可以结合动态代谢工程中的相关算法实现途径动态调控策略的挖掘。现有的主流动态控制策略主要分为3种:两阶段代谢转化、持续代谢控制

和群体行为控制。其中,两阶段代谢转化是目前广泛使用且能与途径设计方法紧密关联的控制策略^[62]。目前,基于两阶段代谢转化开发的优化算法有dFBA(dynamic flux balance analysis)^[63]、MoVE(metabolic value enumerator)^[64]等。通过这些算法即可实现生长与生产的解耦,为动态代谢工程实验提供指导策略。

REFERENCES

- [1] 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.
- [2] Jiang T, Li CY, Teng YX, et al. Recent advances in improving metabolic robustness of microbial cell factories. *Curr Opin Biotechnol*, 2020, 66: 69-77.
- [3] Kim SM, Peña MI, Moll M, et al. A review of parameters and heuristics for guiding metabolic pathfinding. *J Cheminform*, 2017, 9(1): 51.
- [4] Kuwahara H, Alazmi M, Cui XF, et al. MRE: a web tool to suggest foreign enzymes for the biosynthesis pathway design with competing endogenous reactions in mind. *Nucleic Acids Res*, 2016, 44(W1): W217-W225.
- [5] Huang YR, Xie YS, Zhong C, et al. Finding branched pathways in metabolic network via atom group tracking. *PLoS Comput Biol*, 2021, 17(2): e1008676.
- [6] Pharkya P, Burgard AP, Maranas CD. OptStrain: a computational framework for redesign of microbial production systems. *Genome Res*, 2004, 14(11): 2367-2376.
- [7] Chowdhury A, Maranas CD. Designing overall stoichiometric conversions and intervening metabolic reactions. *Sci Rep*, 2015, 5: 16009.
- [8] 马红武, 陈修来, 袁倩倩, 等. 面向生物合成的代谢工程策略设计. *中国科学院院刊*, 2018, 33(11): 1166-1173.
Ma HW, Chen XL, Yuan QQ, et al. Design of metabolic engineering strategies for biosynthesis of valuable products. *Bull Chin Acad Sci*, 2018, 33(11): 1166-1173 (in Chinese).
- [9] Jeffries JG, Seaver SMD, Faria JP, et al. A pathway for every product? Tools to discover and design plant metabolism. *Plant Sci*, 2018, 273: 61-70.
- [10] Orth JD, Thiele I, Palsson BØ. What is flux balance

- analysis? Nat Biotechnol, 2010, 28(3): 245-248.
- [11] Li GH, Dai SX, Han FF, et al. FastMM: an efficient toolbox for personalized constraint-based metabolic modeling. BMC Bioinformatics, 2020, 21(1): 67.
- [12] Bar-Even A, Noor E, Lewis NE, et al. Design and analysis of synthetic carbon fixation pathways. PNAS, 2010, 107(19): 8889-8894.
- [13] Trudeau DL, Edlich-Muth C, Zarzycki J, et al. Design and *in vitro* realization of carbon-conserving photorespiration. PNAS, 2018, 115(49): E11455-E11464.
- [14] Zhang XL, Tervo CJ, Reed JL. Metabolic assessment of *E. coli* as a biofactory for commercial products. Metab Eng, 2016, 35: 64-74.
- [15] King ZA, Lu J, Dräger A, et al. BiGG models: a platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res, 2016, 44(D1): D515-D522.
- [16] Kanehisa M, Furumichi M, Sato Y, et al. KEGG: integrating viruses and cellular organisms. Nucleic Acids Res, 2021, 49(D1): D545-D551.
- [17] Caspi R, Billington R, Keseler IM, et al. The MetaCyc database of metabolic pathways and enzymes-a 2019 update. Nucleic Acids Res, 2020, 48(D1): D445-D453.
- [18] Jeske L, Placzek S, Schomburg I, et al. *BRENDA* in 2019: a European ELIXIR core data resource. Nucleic Acids Res, 2019, 47(D1): D542-D549.
- [19] Lang MR, Stelzer M, Schomburg D. BKM-react, an integrated biochemical reaction database. BMC Biochem, 2011, 12: 42.
- [20] Kumar A, Suthers PF, Maranas CD. MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases. BMC Bioinformatics, 2012, 13: 6.
- [21] Lombardot T, Morgat A, Axelsen KB, et al. Updates in *Rhea*: SPARQLing biochemical reaction data. Nucleic Acids Res, 2019, 47(D1): D596-D600.
- [22] Moretti S, Du T, Tran V, Mehl F, et al. MetaNetX/MNXref: unified namespace for metabolites and biochemical reactions in the context of metabolic models. Nucleic Acids Res, 2021, 49(D1): D570-D574.
- [23] Deville Y, Gilbert D, Van Helden J, et al. An overview of data models for the analysis of biochemical pathways. Brief Bioinform, 2003, 4(3): 246-259.
- [24] Heath AP, Bennett GN, Kavraki LE. An algorithm for efficient identification of branched metabolic pathways. J Comput Biol, 2011, 18(11): 1575-1597.
- [25] Motwalli O, Uludag M, Mijakovic I, et al. PATH cre8: a tool that facilitates the searching for heterologous biosynthetic routes. ACS Synth Biol, 2020, 9(12): 3217-3227.
- [26] Chou CH, Chang WC, Chiu CM, et al. FMM: a web server for metabolic pathway reconstruction and comparative analysis. Nucleic Acids Res, 2009, 37(Web Server issue): W129-W134.
- [27] Croes D, Couche FB, Wodak SJ, et al. Inferring meaningful pathways in weighted metabolic networks. J Mol Biol, 2006, 356(1): 222-236.
- [28] Rahman SA, Advani P, Schunk R, et al. Metabolic pathway analysis web service (pathway hunter tool at CUBIC). Bioinformatics, 2005, 21(7): 1189-1193.
- [29] Pitkänen E, Jouhten P, Rousu J. Inferring branching pathways in genome-scale metabolic networks. BMC Syst Biol, 2009, 3: 103.
- [30] Faust K, Croes D, Van Helden J. Metabolic pathfinding using RPAIR annotation. J Mol Biol, 2009, 388(2): 390-414.
- [31] Huang YR, Zhong C, Lin HX, et al. A method for finding metabolic pathways using atomic group tracking. PLoS One, 2017, 12(1): e0168725.
- [32] Ma HW, Zeng AP. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics, 2003, 19(2): 270-277.
- [33] Watts KT, Lee PC, Schmidt-Dannert C. Exploring recombinant flavonoid biosynthesis in metabolically engineered *Escherichia coli*. Chembiochem, 2004, 5(4): 500-507.
- [34] Lemer C, Antezana E, Couche FB, et al. The aMAZE LightBench: a web interface to a relational database of cellular processes. Nucleic Acids Res, 2004, 32(Database issue): D443-D448.
- [35] Keseler IM, Mackie A, Santos-Zavaleta A, et al. The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. Nucleic Acids Res, 2017, 45(D1): D543-D550.
- [36] Steinbeck C, Han YQ, Kuhn S, et al. The chemistry development kit (CDK): an open-source Java library for chemo- and bioinformatics. J Chem Inf Comput Sci, 2003, 43(2): 493-500.
- [37] Xue L, Godden JW, Stahura FL, et al. Design and evaluation of a molecular fingerprint involving the transformation of property descriptor values into a binary classification scheme. J Chem Inf Comput Sci, 2003, 43(4): 1151-1157.
- [38] Kotera M, Hattori M, Oh MA, et al. RPAIR: a reactant-pair database representing chemical changes

- in enzymatic reactions. *Genome Inform*, 2004, 15(1): P062.
- [39] Akutsu T. Efficient extraction of mapping rules of atoms from enzymatic reaction data. *J Comput Biol*, 2004, 11(2/3): 449-462.
- [40] Jaworski W, Szymkuć S, Mikulak-Klucznik B, et al. Automatic mapping of atoms across both simple and complex chemical reactions. *Nat Commun*, 2019, 10(1): 1434.
- [41] Arita M. Metabolic reconstruction using shortest paths. *Simul Pract Theory*, 2000, 8(1/2): 109-125.
- [42] Jouhten P, Pitkänen E, Pakula T, et al. ^{13}C -metabolic flux ratio and novel carbon path analyses confirmed that *Trichoderma reesei* uses primarily the respirative pathway also on the preferred carbon source glucose. *BMC Syst Biol*, 2009, 3: 104.
- [43] Fang X, Lloyd CJ, Palsson BO. Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nat Rev Microbiol*, 2020, 18(12): 731-743.
- [44] Kim J, Reed JL, Maravelias CT. Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. *PLoS One*, 2011, 6(9): e24162.
- [45] Tervo CJ, Reed JL. MapMaker and PathTracer for tracking carbon in genome-scale metabolic models. *Biotechnol J*, 2016, 11(5): 648-661.
- [46] Hafner J, MohammadiPeyhani H, Sveshnikova A, et al. Updated ATLAS of biochemistry with new metabolites and improved enzyme prediction power. *ACS Synth Biol*, 2020, 9(6): 1479-1482.
- [47] Yang X, Yuan QQ, Luo H, et al. Systematic design and *in vitro* validation of novel one-carbon assimilation pathways. *Metab Eng*, 2019, 56: 142-153.
- [48] Bogorad IW, Lin TS, Liao JC. Synthetic non-oxidative glycolysis enables complete carbon conservation. *Nature*, 2013, 502(7473): 693-697.
- [49] Burgard AP, Pharkya P, Maranas CD. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng*, 2003, 84(6): 647-657.
- [50] Liu XL, Lin J, Hu HF, et al. *De novo* biosynthesis of resveratrol by site-specific integration of heterologous genes in *Escherichia coli*. *FEMS Microbiol Lett*, 2016, 363(8): fnw061.
- [51] Zhao Y, Wu BH, Liu ZN, et al. Combinatorial optimization of resveratrol production in engineered *E. coli*. *J Agric Food Chem*, 2018, 66(51): 13444-13453.
- [52] Carbonell P, Jervis AJ, Robinson CJ, et al. An automated design-build-test-learn pipeline for enhanced microbial production of fine chemicals. *Commun Biol*, 2018, 1: 66.
- [53] Litsa E, Peña MI, Moll M, et al. Machine learning guided atom mapping of metabolic reactions. *J Chem Inf Modeling*, 2019, 59(3): 1121-1135.
- [54] Yuan QQ, Huang T, Li PS, et al. Pathway-consensus approach to metabolic network reconstruction for *Pseudomonas putida* KT2440 by systematic comparison of published models. *PLoS One*, 2017, 12(1): e0169437.
- [55] Lieven C, Beber ME, Olivier BG, et al. MEMOTE for standardized genome-scale metabolic model testing. *Nat Biotechnol*, 2020, 38(3): 272-276.
- [56] Hartmann A, Vila-Santa A, Kallscheuer N, et al. OptPipe-a pipeline for optimizing metabolic engineering targets. *BMC Syst Biol*, 2017, 11(1): 143.
- [57] Schneider P, Klamt S. Characterizing and ranking computed metabolic engineering strategies. *Bioinformatics*, 2019, 35(17): 3063-3072.
- [58] Otero-Muras I, Carbonell P. Automated engineering of synthetic metabolic pathways for efficient biomanufacturing. *Metab Eng*, 2021, 63: 61-80.
- [59] Li CY, Zhang RH, Wang J, et al. Protein engineering for improving and diversifying natural product biosynthesis. *Trends Biotechnol*, 2020, 38(7): 729-744.
- [60] Kuhlman B, Bradley P. Advances in protein structure prediction and design. *Nat Rev Mol Cell Biol*, 2019, 20(11): 681-697.
- [61] Gao WH, Mahajan SP, Sulam J, et al. Deep learning in protein structural modeling and design. *Patterns (N Y)*, 2020, 1(9): 100142.
- [62] Hartline CJ, Schmitz AC, Han YC, et al. Dynamic control in metabolic engineering: theories, tools, and applications. *Metab Eng*, 2021, 63: 126-140.
- [63] Mahadevan R, Edwards JS, Doyle FJ III. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys J*, 2002, 83(3): 1331-1340.
- [64] Venayak N, Von Kamp A, Klamt S, et al. MoVE identifies metabolic valves to switch between phenotypic states. *Nat Commun*, 2018, 9(1): 5332.

(本文责编 郝丽芳)