

• 合成生物技术 •

基于图论的代谢网络中流通代谢物处理新方法

高雅杰^{1,2}, 袁倩倩², 杨雪², 毛志涛², 余文童², 刘浩¹, Igor Goryanin^{2,3}, 马红武²

1 天津科技大学 生物工程学院, 天津 300457

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

3 University of Edinburgh, School of Informatics, Edinburgh, UK

高雅杰, 袁倩倩, 杨雪, 毛志涛, 余文童, 刘浩, Igor Goryanin, 马红武. 基于图论的代谢网络中流通代谢物处理新方法. 生物工程学报, 2022, 38(4): 1554-1564.

GAO YJ, YUAN QQ, YANG X, MAO ZT, YU WT, LIU H, Igor Goryanin, MA HW. A graph-theory-based method for processing of currency metabolites in metabolic networks. Chin J Biotech, 2022, 38(4): 1554-1564.

摘要: 使用图论方法对基因组尺度代谢网络进行途径搜索是目前途径设计中最常用的方法之一。然而, 由于流通代谢物 (如 H^+ 、 H_2O 、 CO_2 和 ATP 等) 的影响, 图论搜索得到的途径在生物学上经常是不可行的。虽然前人提出了一些方法对流通代谢物进行处理, 但均存在一些问题, 目前还没有标准的处理方法。文中基于流通代谢物在反应中作为转移磷酸等功能基团转移载体的特性, 提出了一种对流通代谢物成对处理的方法, 并将其按照其在功能基团转移中发挥的作用进行优先级排序, 以保证每个反应都有对应的代谢物连接, 进而通过程序实现自动化处理。与已有其他方法处理结果的比较表明, 这种流通代谢物成对排序处理的方法可以排除掉生物学不可行的途径, 从而提高途径搜索的准确性, 为利用图论方法进行途径设计和可视化提供更好的支撑。

关键词: 代谢网络模型; 图论方法; 流通代谢物; 途径分析

Received: June 1, 2021; **Accepted:** August 27, 2021; **Published online:** September 13, 2021

Supported by: National Key Research and Development Program of China (2018YFA0900300); Tianjin Synthetic Biotechnology Innovation Capacity Improvement Project, China (TSBICIP-PTJS-001, TSBICIP-KJGG-005)

Corresponding author: MA Hongwu. Tel: +86-22-24828735; E-mail: ma_hw@tib.cas.cn

基金项目: 国家重点研发计划 (2018YFA0900300); 天津市合成生物技术创新能力提升行动项目 (TSBICIP-PTJS-001, TSBICIP-KJGG-005)

A graph-theory-based method for processing of currency metabolites in metabolic networks

GAO Yajie^{1,2}, YUAN Qianqian², YANG Xue², MAO Zhitao², YU Wentong², LIU Hao¹, Igor Goryanin^{2,3}, MA Hongwu²

1 College of Biotechnology, Tianjin University of Science & Technology, Tianjin 300457, China

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

3 School of Informatics, University of Edinburgh, Edinburgh, UK

Abstract: Graph-theory-based pathway analysis is a commonly used method for pathway searching in genome-scale metabolic networks. However, such searching often results in many pathways biologically infeasible due to the presence of currency metabolites (e.g. H^+ , H_2O , CO_2 , ATP etc.). Several methods have been proposed to address the problem but up to now there is no well-recognized methods for processing the currency metabolites. In this study, we proposed a new method based on the function of currency metabolites for transferring of functional groups such as phosphate. We processed most currency metabolites as pairs rather than individual metabolites, and ranked the pairs based on their importance in transferring functional groups, in order to make sure at least one main metabolite link exists for any reaction. The whole process can be done automatically by programming. Comparison with existing approaches indicates that more biologically infeasible pathways were removed by our method and the calculated pathways were more reliable, which may facilitate the graph-theory-based pathway design and visualization.

Keywords: metabolic network model; graph theory; currency metabolites; pathway analysis

基因组尺度代谢网络模型 (genome-scale metabolic network models, GSMNs) 是理解复杂网络进化和生物功能关系的有力工具^[1-2]。基于计量学约束的通量平衡分析方法^[3]是目前最常用的代谢网络功能分析方法,近年来多篇综述文章对应用相关方法进行代谢网络模型构建^[1,4], 细胞表型预测^[4]、代谢途径设计和代谢改造靶点预测^[4-5], 以及添加更多约束^[6]或者整合多组学数据^[7]以获得更准确预测结果等方面进行了详细阐述。相比之下,近几年应用图论的方法对代谢网络进行的研究工作较少。而研究 GSMNs 途径最主流的两类底层方法则为基于图论的拓

扑方法和基于约束的计量学方法。图论方法进行代谢网络分析的优势是便于对代谢网络的全局结构特征进行分析,如无标度、模块化^[8]、Bowtie 连接性^[9-10]等。应用图论的途径搜索方法可以更快地搜索两个代谢物间转化的最短途径^[11-14], 计算需求显著低于基于计量学模型的最优途径计算。此外,图论表示也是代谢网络可视化很重要的步骤,很多可视化工具可以直接由图连通结构自动布局节点位置生成网络连接图^[15-16]。图论方法应用于代谢网络分析时一个主要问题是流通代谢物处理的问题^[17]。由于流通代谢物在代谢网络中的高度连通性,往往会

导致图结构与真实生物转化关系不符, 途径计算时得到大量没有生物学意义的假阳性途径, 网络可视化的结果也与生物学家熟悉的代谢途径差异很大^[18], 使得人们不得不依赖耗时的手绘绘图工具以得到高质量的代谢图^[19]。因此, 应用图论方法对代谢网络进行分析时, 流通代谢物的处理一直是研究者们不可避免的问题。

流通代谢物最常用的处理方法是直接从网络中去除^[20-21], 但流通代谢物作为主代谢物参与的途径也一并被去除, 以 ATP 为例, 其与 ADP 一起转移磷酸基团时作为流通代谢物应该被去除, 但是当 ATP 参与核苷酸合成时就是主代谢物, 如果被直接去除将缺失核苷酸的合成途径。因此, 很多流通代谢物不能简单地定义并从网络中排除, 而是由其在具体反应中的功能决定的。针对这一问题, Ma 等提出了将流通代谢物按照在反应中的功能成对去除^[17], 如用于转移氨基的 α -酮戊二酸 (2-oxoglutarate, akglu) 和谷氨酸 (L-glutamate, glu) 代谢物对, 转氢的 NADH 和 NAD 代谢物对, 以及转磷的 ATP 和

ADP 代谢物对。这些代谢物对在反应中的功能只是将一些基团转移而并不参与反应中主代谢物的碳骨架转化, 因此可以被成对去除。该工作手动对几千个反应中的常见流通代谢物对进行处理, 并未对整个代谢网络中所有可能的流通代谢物对进行分析与流程化处理。随后, Rahman、Heath、Croes 和 Blum 等, 相继提出了依据代谢物的原子结构信息^[22-23]和对代谢物途径进行加权重^[11,24]以及两者相结合^[25]的方法计算最短途径, 从而自动筛选含有流通代谢物的连接途径。表 1 中对这些方法的特点和问题进行了简要比较。这些方法中有些并没有将不合理的流通代谢物去除, 有些则存在定义不严谨导致去除不完整或错误地去除了本来应该是主代谢物的连接。到目前为止, 图论方法计算代谢途径中还没有广泛认可的标准方法处理流通代谢物。

2020 年 Motwalli 等开发了搜索异源生物最优合成途径的 PATH^{cre8} 数据资源工具^[26], 利用该工具网站进行途径搜索, 给定底物和目标产品, 可以同时搜索到 IntEnz^[27]、Rhea^[28]、

表 1 流通代谢物不同处理方法的比较

Table 1 Comparison of different methods for dealing with currency metabolites

Methods	Description	References
Remove currency metabolites nodes	Highly connected metabolites are identified as currency metabolites and removed from the network. The problem is that a metabolite (e.g. ATP) can be a currency metabolite in one reaction but a main metabolite in another	[23]
Remove currency metabolite links	Determine which links should be removed in each reaction based on their functions (e.g. for transferring H), need time consuming manual examination of every reaction	[24]
Weighted metabolic network graph	Highly connected currency metabolites have high weights and thus less likely to be found by a lightest pathway search method. Suitable for biologically meaningful pathway search but not network visualization	[27]
Atom mapping	Avoid meaningless links without atom transfer but many links through currency metabolites still exist (e.g. carbon atom transfer to CO ₂). Require metabolite structure information which are not available in many published metabolic models	[25-26]
Weighted atom mapping graph	Combination of the above two methods, improved pathway search results but not suitable for network visualization	[28]
Pair and rank method	Automatic process of each reaction to determine which metabolite links should be removed based on predefined and ranked currency metabolite pairs. The reduced graph can be used for pathway search as well as visualization	This study

BiGG^[29]和 ModelSEED^[30]不同数据库的途径。应用上述网站进行丙酮酸 (pyruvate, pyr) 到天冬氨酸 (L-aspartate, asp) 的图论搜索, 得到基于 4 个数据库以及 PATH^{cre8} 优化网络后的途径结果, 如图 1A、1B、1C、1D 和 1E 所示。从图 1A–D 中可以看出, 因流通代谢物没有在数据库的反应集中去除, 导致 H⁺、H₂O 及 CO₂ 等不作为碳骨架的代谢物错误地出现在途径中。而 PATH^{cre8} 工具在对网络进行优化时, 为确保途径的可行性, 排除了在许多反应中大量消耗或者生成的流通代谢物, 因此, 如图 1E 中所示的途径, 有效避免了上述 4 个途径出现的问题。再利用 PATH^{cre8} 搜索葡萄糖到 pyr 的途径, 如图 1F 所示, 分析经过排序筛选的最优途径 (左侧紫色路线), 尽管去除了流通代谢物的连接, 但是通过反应 R06620

(D-glucose+ubiquinone<=>D-glucono-1,5-lactone+ubiquinol)^[31], 得到了葡萄糖 (glucose) 与泛醇 (ubiquinol) 的途径连接, 而泛醌 (ubiquinone) 与 ubiquinol 在此反应中应该作为转移氢离子的载体不参与碳骨架的连接。由此可知, 该工具对需要去除的辅因子整理不够全面, 没有去除在反应中起转移功能基团的流通代谢物对。因此, 本文基于之前的研究, 提出对流通代谢物进行组对并排序去除的处理方法, 该方法在程序中可实现自动化去除流通代谢物, 得到更加符合生物学意义的代谢物途径图。本研究中整理的流通代谢物对及流通代谢物成对排序去除代码公布于 github 网站 (<https://github.com/tibbdc/GraphCM>, 其中 GraphCM 是本研究开发的程序名, 即基于图论的流通代谢物处理方法) 供研究者使用。

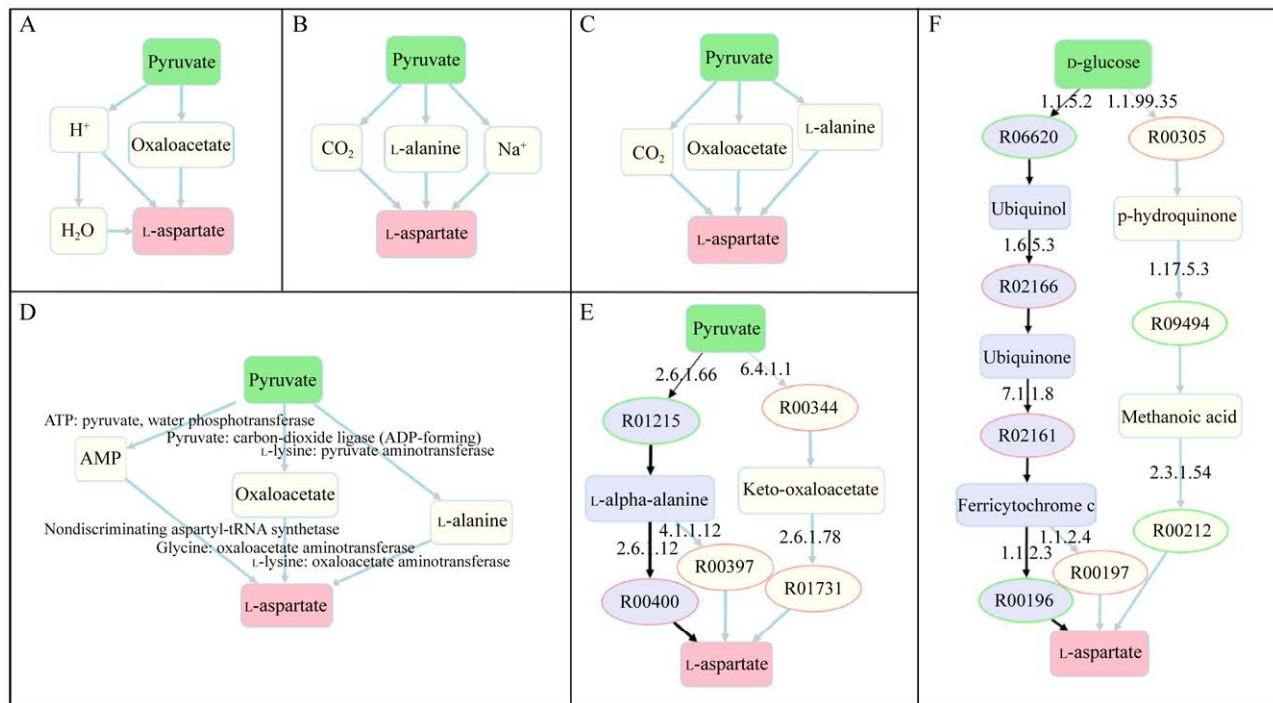


图 1 基于不同代谢反应数据库的图论途径搜索结果^[26]

Figure 1 Pathways found by graph based methods from different metabolic databases. A–E were three searched pathways from pyruvate to L-aspartate obtained from IntEnz, BiGG, Rhea, ModelSeed databases and based on PATH^{cre8} tool^[26], respectively. F is three searched pathways from D-glucose to L-aspartate using the PATH^{cre8} tool, in which the purple pathway on the left is the optimal pathway after ranking and screening.

1 材料与amp;方法

1.1 数据库选择及模型下载

BiGG 模型数据库^[29] (<http://bigg.ucsd.edu/>) 是高质量基因组尺度代谢模型的集中存储库, 包含 35 种不同生物的 108 个高质量基因组尺度代谢网络模型。本研究选择模型质量最高且研究最为普遍的模式菌株——大肠杆菌 (*Escherichia coli*) 模型作为研究对象。

BiGG 数据库中包含大肠杆菌 58 个已发表模型, *Escherichia coli* str. K-12 substr. MG1655 菌株最后更新的模型是 2017 年发表的 iML1515^[32], 模型中包含 1 877 个代谢物、2 712 个反应、1 516 个基因。在 BiGG 数据库中下载 iML1515 的 SBML (systems biology markup language)^[33] (<http://sbml.org/>) 格式文件作为我们方法研究的底盘菌株模型。

1.2 工具包选择

选取 Python 编程语言进行脚本的书写, 选用基于 Python 的工具包 COBRApy (constraints-based reconstruction and analysis for Python)^[34] 的相应函数语句读取 SBML 格式模型中的代谢物和反应等所有信息, 再使用 Pandas 工具包的函数语句将其转化为 Excel 格式文档, 以便人工查阅模型信息。

对代谢网络进行途径分析时使用 NetworkX^[35] 网络分析包。首先, 将 SBML 格式文件中复杂的生物反应转化为简单的点到点通

过反应连接的关系对, 整个生物体中所有的这些关系对组合到一起就形成了图论模型。随后, 基于图论模型使用 NetworkX 就可以得到以代谢物作为点、反应作为边的有向途径图。表 2 中提供了本研究中所用工具的功能、版本以及下载地址, 更加详细的各个软件包版本要求以及安装步骤可参见 github 网站中的 README.md 文件 (<https://github.com/tibbdc/GraphCM/blob/main/README.md>)。

1.3 读取数据

使用 Excel 文件存储所有流通代谢物以及流通代谢物对, 从 Excel 中使用 Pandas 中的函数语句导出作为不同功能的流通代谢物和流通代谢物对, 并写成 List 列表, 用于后面网络的分析。

1.4 函数语句

分析代谢网络主要使用 COBRApy、NetworkX 和 Pandas 三个主流的工具包, 本文研究中所用工具包中的函数语句及其实现的功能如表 3 所示。

2 结果与分析

2.1 流通代谢物功能分析及处理

首先, 利用“model.reactions”函数提取大肠 iML1515 模型中的反应信息 (排除交换反应以及菌体组成反应), 将反应转化成反应物和生成物对。结果显示 H⁺、H₂O、O₂、Pi 等流通代谢物在模型中出现的频率和连接度较高, 达 9 985 对

表 2 本研究中所用编程语言及计算工具的详细信息

Table 2 Details of the programming languages and computing tools used in this study

Names	Category	Function	Version	Download link
Python	Programming language	Programming language	3.8.5	https://www.python.org/getit/
COBRApy	Package	Read model information	0.20.0	http://opencobra.sourceforge.net/
NetworkX	Package	Graph based network analysis	2.5	https://networkx.org/
Pandas	Package	Data analysis and manipulation tool	1.1.4	https://pandas.pydata.org/

表 3 文中所用函数语句及其实现的功能

Table 3 The name of the function statement used in this paper and its implementation function

Function names	Package	Function
Model.reaction	COBRAPy	Reaction information
Model.metabolite	COBRAPy	Metabolite information
Reaction.reactants	COBRAPy	Reactants in reaction
Reaction.products	COBRAPy	Products in reaction
Metabolite.elements	COBRAPy	Metabolite elemental composition
Reaction.reversibility	COBRAPy	Determine reaction reversibility
MultiDiGraph()	NetworkX	Create a directed graph
Add_edge	NetworkX	Directed graph add edge
All_shortest_paths	NetworkX	Calculate shortest path
Read_excel	Pandas	Read the excel file
ExcelWriter	Pandas	Create excel writable file
DataFrame	Pandas	Tabular data structure

连接,但在反应中只是起到辅助因子的作用,并不进行实质的碳流交换,所以这些不含碳的小分子流通代谢物(其完整的列表详见https://github.com/tibbdc/GraphCM/tree/main/data/external/special_currency_metabolites.csv)参与的连接对应该做删除处理。之后,使用图论方法对处理后的代谢网络搜索 pyr 到 asp 的最短途径(图 2 P1-P3)。如图 2 所示,P2 和 P3 途径中出现 pyr 直接到 akg 和丙氨酸(L-alanine, ala)直接到 glu 的错误连接。分析发现造成错误的原因是在可逆反应 ALATA_L (2-oxoglutarate+L-alanine \rightleftharpoons L-glutamate+pyruvate)中,akg 和 glu 均被当作主代谢物。而实际上这两个代谢物是作为氨基转移载体出现在反应的两侧实现 ala 和 pyr 间的转氨,并没有发生 akg 和 pyr 及 pyr 和 glu 间的碳转移,所以 akg 和 glu 在此反应中应该作为转氨的流通代谢物对去除。我们对大肠杆菌 iML1515 模型中同时存在 akg 和 glu 的 20 个生化反应进行分析,发现仅反应 GLUDy (glu_L_c+h2o_c+nadp_c \rightleftharpoons akg_c+h_c+nadph_c+nh4_c)中 akg 和 glu 应该作为主干代谢物保留。在其他 19 个反应中均为提供转氨功能的流通代谢物,应去除以避免上述错误。

另外,还发现直接去除之后,很多连接对被错误地删除或者是保留。如表 4 所示, GUACYC 反应中,将 gtp 作为流通代谢物去除,会使 gtp 到 35cgmp 的代谢物对缺失; NTPP6 反应中,直接将 atp 和 amp 都作为流通代谢物去除,但其实在此反应中没有其他代谢物的存在, atp 与 amp 的连接应该作为代谢物对保留。还有谷草转氨 (GF6PTA) 和谷氨酰胺转氨 (PTRCTA) 反应, akg 和 glu 以及 gln 和 glu 出现在转移反应的两侧,只是为了提供氨基并不作为碳骨架转移,应该作为流通代谢物被去除。

为此本文作者 Ma 等^[17]在代谢网络全局结构研究中提出流通代谢物不应单独以代谢物来定义,而应该根据在反应中的作用来定义。因此,我们根据流通代谢物通常作为某些离子或者功能基团的载体,成对出现在反应的两侧这一特性,将这些代谢物按照不同功能组成流通代谢物对(currency metabolite-pairs, CMPs),如转移氢(NADH/NAD)、转移磷酸根(ATP/ADP)、转移氨基(AKG/GLU)、转移一碳单位(METHF/THF)、转移乙酰基团(AcCoA/Ac)等,然后成对地在反应中去除。

当反应中反应物和产物对都为 CMPs 时，成对去除 CMPs 之后，会使反应缺失，可能会导致途径断裂。例如，反应 NTPP6 中，atp 和 amp 为一对 CMPs，直接去除之后，就缺失了反应 NTPP6。为了确保每个反应至少有一个代谢物对连接，本研究根据大量的分析总结 and 数据库查阅 (如 MetaCyc^[36])，判断反应中的哪个 CMPs 是主要基团转移对，将其进行排序，再按照优先级顺序去除反应中的 CMPs。例如，根据 MetaCyc 中许多核苷二磷酸应该作为受

体，而许多核糖和脱氧核糖核苷三磷酸往往应该作为供体的原理^[36]，在反应 NDPK5: ATP+dGDP \rightleftharpoons ADP+dGTP 中，将 ATP 作为供体、dGDP 作为受体。因此 ATP 和 ADP 作为反应中主要的基团转移对，应该排在 dGDP 和 dGTP 连接对前面，按照 CMPs 的优先级去除，最后留下反应真正进行转化的 dGDP 到 dGTP 代谢物对。排序好的完整流通代谢物对列表可从 https://github.com/tibbdc/GraphCM/tree/main/data/external/pairs_of_currency_metabolites.csv 中获取。

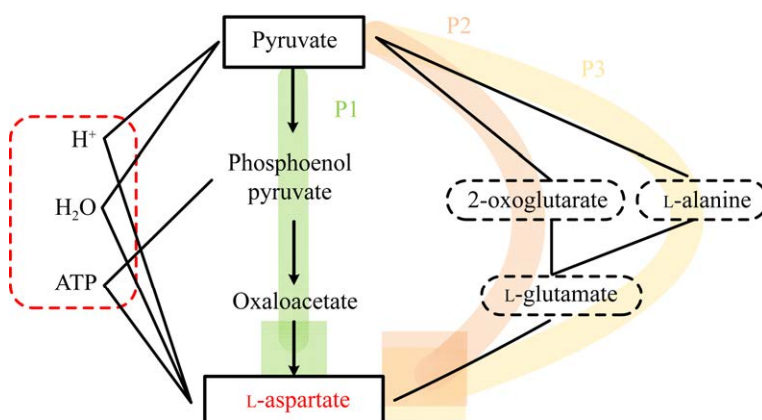


图 2 流通代谢物不同处理方法得到的丙酮酸到天冬氨酸的途径

Figure 2 Pathways from pyruvate to L-aspartate upon different treatment of currency metabolites. The left side represents the pathway from pyruvate to L-aspartate through currency metabolites; P1, P2 and P3 represent the pathway from pyruvate to L-aspartate by directly removing the currency metabolites in the metabolic network, where P1 represents the main pathway from pyruvate to L-aspartate in the metabolic network.

表 4 直接去除流通代谢物造成的缺失或者冗余的连接对示例

Table 4 Examples of missing or redundant connection pairs caused by direct removal of currency metabolites

Reaction ID	Reaction names	Reaction equations	Missing or redundant connection pairs
GUACYC	Guanylate cyclase	$gtp_c \rightarrow 35cgmp_c + ppi_c$	$gtp_c \rightarrow 35cgmp_c$
NTPP6	Nucleoside triphosphate pyrophosphorylase	$atp_c + h_2o_c \rightarrow amp_c + h_c + ppi_c$	$atp_c \rightarrow amp_c$
GF6PTA	Glutamine fructose 6-phosphate transaminase	$f6p_c + gln_L_c \rightarrow gam6p_c + glu_L_c$	$gln_L_c \rightarrow gam6p_c / f6p_c \rightarrow glu_L_c$
PTRCTA	Putrescine transaminase	$akg_c + ptrc_c \rightarrow 4abutn_c + glu_L_c$	$akg_c \rightarrow 4abutn_c / ptrc_c \rightarrow glu_L_c$

2.2 流通代谢物处理后获得图论模型

基于图论方法对代谢网络进行途径分析时, 需要将代谢网络模型转化为以代谢物作为点、反应作为点与点之间连接边的有向图论模型, 流程如图 3 所示。提取代谢网络 *iML1515* 中的反应信息, 首先去除不含碳的小分子流通代谢物, 再将如 ATP、NADH 等含碳的代谢物按照成对以及优先级排序后的顺序进行去除,

最后至少保留一个连接的代谢物对。最后, 将所有代谢物连接对存储在有向图库中, 这个有向图库便形成了我们所说的图论模型, 对图论模型进行途径搜索可以得到代谢物之间的途径图, 运用可视化工具也可以画出代谢网络的全局连接图 (该方法的脚本文件网址为 https://github.com/tibbdc/GraphCM/blob/main/notebook/graph_currency_metabolites.ipynb)。

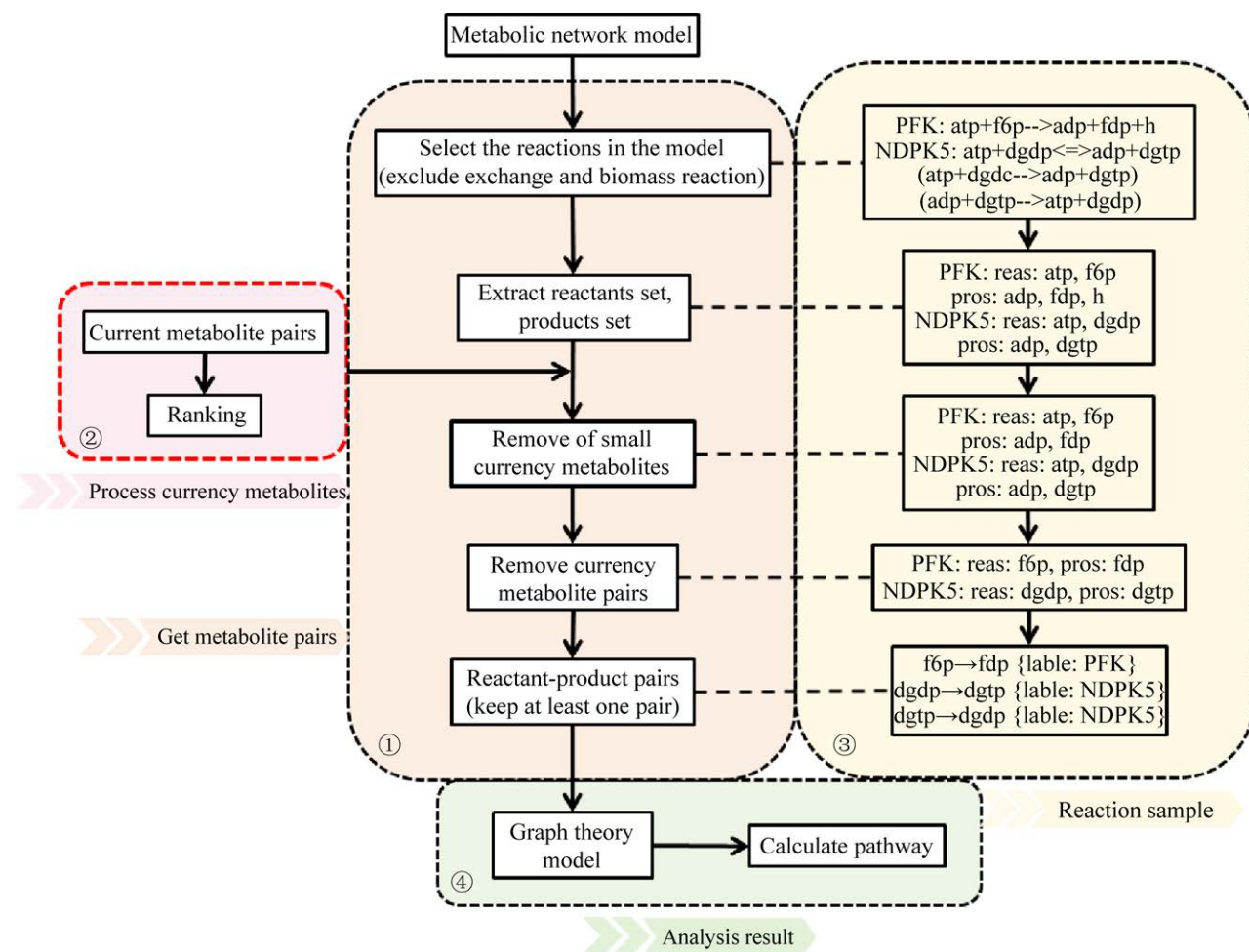


图 3 代谢网络转化为图论模型流程图

Figure 3 The flow chart of transforming metabolic network model into graph theory model. ① represents the process of transforming the metabolic network model into reactant-product pairs; ② represents the pairing and ranking of currency metabolites in the network; ③ represents the example of splitting reversible and irreversible reaction into metabolite pairs respectively; ④ represents the pathway search analysis of graph theory model.

需要注意的是, 针对代谢网络中的可逆反应, 本研究使用 COBRApy 工具包中的“reaction.reversibility”函数进行识别, 并将每个可逆反应拆分为两个不可逆反应^[14], 如图 2 示例中的反应 NDPK5, 拆分为两个不可逆反应, 然后对两个反应同时去除流通代谢物, 得到如表 5 中所示的代谢物对。

对包含 1 877 个代谢物、2 712 个反应的 *i*ML1515 模型进行上述的流通代谢物处理之后, 得到 1 713 个含碳代谢物之间的 3 526 个代谢物对的连接图。但是在模型中会存在两种或者两种以上的反应可能具有相同的代谢物与代谢物之间的连接。例如, 缬氨酸 (valine) 合成的最后一个反应可以由两种不同的酶催化: IlvE (glutamate+2-ketovaline \rightleftharpoons valine+ketoglutarate) 和 AvtA (alanine+2-ketovaline \rightleftharpoons valine+pyruvate), 这两个反应分别使用 glu 和 ala 作为氨基基团供体, 但主要代谢连接对都是 2-ketovaline \rightarrow valine。因此对于这种不同酶催化得到相同代谢物的连接, 保留一个即可。去除重复连接后, 连接数

减少到 3 189 个。与之前未处理流通代谢物、直接去除流通代谢物以及仅仅成对去除流通代谢物并未将流通代谢物对排序处理得到的连接对结果进行对比, 结果如表 6 所示。从表中我们可以看出, 成对并排序去除流通代谢物后, 由于去除了一些冗余的连接对, 使得连接对的数量明显减少, 这使得有向网络更加符合实际, 也更加有利于途径分析。

2.3 图论模型途径搜索结果分析

利用流通代谢物不同处理方法得到的有向图, 计算 pyr 到 asp 步长分别为 2、3 和 4 的途径个数, 结果如表 6 所示。从表中可以看出, 流通代谢物经过成对并排序的方法去除之后, 得到的途径个数减少, 途径更加精练合理, 结果更加准确。观察结果发现, 去除流通代谢物之后, pyr 到 asp 的最短途径长度为 3, 而图 1 通过不同数据搜索得到的最短路径长度为 2, 原因是大肠杆菌 *i*ML1515 代谢网络中缺少二氧化碳连接酶, 没有丙酮酸直接生成草酰乙酸的反应。另外, 对流通代谢物对进行排序之后, pyr 到 asp

表 5 可逆反应拆分示例

Table 5 Examples of splitting reversible reaction

Reaction ID	Reaction equation	Lower bounds	Upper bounds	Reactant-product pairs
NDPK5	atp_c+dgdp_c \rightleftharpoons adp_c+dgtp_c	-1 000	1 000	-
NDPK5	atp_c+dgdp_c \rightarrow adp_c+dgtp_c	0	1 000	dgdp_c \rightarrow dgtp_c
NDPK5	adp_c+dgtp_c \rightarrow atp_c+dgdp_c	0	1 000	dgtp_c \rightarrow dgdp_c

表 6 流通代谢物不同处理得到的代谢物对连接数

Table 6 Number of connections of metabolite pairs obtained by different treatment of currency metabolites

Treatment of currency metabolites	Number of connections between metabolites	Number of connections between metabolite pairs after removal of duplicates	pyr-asp pathway numbers		
			2	3	4
Untreated	14 021	8 602	7	1 387	254 280
Directly to remove	4 036	3 477	0	22	68
In pairs to remove	4 731	3 989	0	0	15
Pair and rank to remove	3 526	3 189	0	1	16

Annotation: “pyr”: pyruvate, “asp”: L-aspartate.

的途径个数比未进行排序时多了 1 个。分析得到的途径发现, 直接去除 CMPs 而并未排序时, 由于连接对缺失, 图论搜索时未得到 pyr 到 asp 的主干途径 P1 (图 2)。这进一步说明将成对的流通代谢物进行排序对于得到符合生物学的途径具有重要意义。

综上对途径分析的结果可知, 通过本研究开发的流通代谢物处理方法与其他方法相比, 进行途径搜索时得到的途径个数更少 (表 6), 准确度更高, 更有利于筛选正确的途径, 并且在计算结果中避免了一些由于可逆反应的存在导致无效途径出现 (图 2 中的 P2 和 P3) 的问题。

3 讨论

图论方法经常被用于研究代谢网络的途径, 但是由于流通代谢物处理不当导致得到的代谢途径不够准确。本文中提出了根据流通代谢物在反应中的意义, 将流通代谢物组成流通代谢物对去除。然后, 按照转移不同功能基团载体的作用分为不同类型, 确定每一类型中实现相应功能基团的转化的主要流通代谢物对。最后, 按照优先级顺序依次去除反应中的流通代谢物, 得到更加准确的有向图的方法。通过本研究对流通代谢物处理方法的改进, 将代谢网络模型转化为有向的图论模型后, 使用图论方法搜索得到的途径避免了流通代谢物的连接, 预测结果更加符合生物学意义, 同时也可以更加准确地得到代谢网络的途径, 为更好地观察代谢网络的整体结构奠定基础。

REFERENCES

- [1] 张晨阳, 武耀康, 徐显皓, 等. 工业微生物代谢网络模型的研究进展及应用. 生物工程学报, 2021, 37(3): 860-873.
Zhang CY, Wu YK, Xu XH, et al. Current status and future perspectives of metabolic network models of industrial microorganisms. *Chin J Biotech*, 2021, 37(3): 860-873 (in Chinese).
- [2] Gu C, Kim GB, Kim WJ, et al. Current status and applications of genome-scale metabolic models. *Genome Biol*, 2019, 20(1): 121.
- [3] Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol*, 2010, 28(3): 245-248.
- [4] 叶超, 徐楠, 陈修来, 等. 应用代谢网络模型解析工业微生物胞内代谢. 生物工程学报, 2019, 35(10): 1901-1913.
Ye C, Xu N, Chen XL, et al. Application of metabolic network model to analyze intracellular metabolism of industrial microorganisms. *Chin J Biotech*, 2019, 35(10): 1901-1913 (in Chinese).
- [5] 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 Chinese Acad Sci*, 2018, 33(11): 1166-1173 (in Chinese).
- [6] 赵欣, 杨雪, 毛志涛, 等. 基于酶约束的代谢网络模型研究进展及其应用. 生物工程学报, 2019, 35(10): 1914-1924.
Zhao X, Yang X, Mao ZT, et al. Progress and application of metabolic network model based on enzyme constraints. *Chin J Biotech*, 2019, 35(10): 1914-1924 (in Chinese).
- [7] 周静茹, 刘鹏, 夏建业, 等. 基于约束的基因组规模代谢网络模型构建方法研究进展. 生物工程学报, 2021, 37(5): 1526-1540.
Zhou JR, Liu P, Xia JY, et al. Advances in the development of constraint-based genome-scale metabolic network models. *Chin J Biotech*, 2021, 37(5): 1526-1540 (in Chinese).
- [8] Lacroix V, Cottret L, Thébault P, et al. An introduction to metabolic networks and their structural analysis. *IEEE/ACM Trans Comput Biol Bioinform*, 2008, 5(4): 594-617.
- [9] Ma HW, Zeng AP. The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics*, 2003, 19(11): 1423-1430.
- [10] Barabási AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet*, 2004, 5(2): 101-113.
- [11] Croes D, Couche F, Wodak SJ, et al. Metabolic path finding: inferring relevant pathways in biochemical networks. *Nucleic Acids Res*, 2005, 33(web server

- issue): W326-W330.
- [12] Abd Algfoor Z, Shahrizal Sunar M, Abdullah A, et al. Identification of metabolic pathways using pathfinding approaches: a systematic review. *Brief Funct Genomics*, 2017, 16(2): 87-98.
- [13] Huang Y, Zhong C, Lin HX, et al. A method for finding metabolic pathways using atomic group tracking. *PLoS One*, 2017, 12(1): e0168725.
- [14] Faust K, Croes D, van Helden J. Prediction of metabolic pathways from genome-scale metabolic networks. *Biosystems*, 2011, 105(2): 109-121.
- [15] Pavlopoulos GA, Wegener AL, Schneider R. A survey of visualization tools for biological network analysis. *BioData Min*, 2008, 1: 12.
- [16] Hari A, Lobo D. Fluxer: a web application to compute, analyze and visualize genome-scale metabolic flux networks. *Nucleic Acids Res*, 2020, 48(w1): W427-W435.
- [17] 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.
- [18] Jensen PA, Papin JA. MetDraw: automated visualization of genome-scale metabolic network reconstructions and high-throughput data. *Bioinformatics*, 2014, 30(9): 1327-1328.
- [19] Droste P, Miebach S, Niedenführ S, et al. Visualizing multi-omics data in metabolic networks with the software Omix: a case study. *Biosystems*, 2011, 105(2): 154-161.
- [20] Chazalviel M, Frainay C, Poupin N, et al. MetExploreViz: web component for interactive metabolic network visualization. *Bioinformatics*, 2018, 34(2): 312-313.
- [21] Wagner A, Fell DA. The small world inside large metabolic networks. *Proc Biol Sci*, 2001, 268(1478): 1803-1810.
- [22] Rahman SA, Advani P, Schunk R, et al. Metabolic pathway analysis web service (Pathway Hunter Tool at CUBIC). *Bioinformatics*, 2005, 21(7): 1189-1193.
- [23] Heath AP, Bennett GN, Kavraki LE. Finding metabolic pathways using atom tracking. *Bioinformatics*, 2010, 26(12): 1548-1555.
- [24] Croes D, Couche F, Wodak SJ, et al. Inferring meaningful pathways in weighted metabolic networks. *J Mol Biol*, 2006, 356(1): 222-236.
- [25] Blum T, Kohlbacher O. Using atom mapping rules for an improved detection of relevant routes in weighted metabolic networks. *J Comput Biol*, 2008, 15(6): 565-576.
- [26] 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.
- [27] Fleischmann A, Darsow M, Degtyarenko K, et al. IntEnz, the integrated relational enzyme database. *Nucleic Acids Res*, 2004, 32(database issue): D434-D437.
- [28] Morgat A, Axelsen KB, Lombardot T, et al. Updates in *Rhea*—a manually curated resource of biochemical reactions. *Nucleic Acids Res*, 2015, 43(database issue): D459-D464.
- [29] Norsigian CJ, Pusarla N, McConn JL, et al. BiGG models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res*, 2020, 48(database issue): D402-D406.
- [30] Seaver SMD, Liu F, Zhang Q, et al. The ModelSEED biochemistry database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Res*, 2021, 49(database issue): D575-D588.
- [31] Kanehisa M, Furumichi M, Sato Y, et al. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res*, 2021, 49(database issue): D545-D551.
- [32] Monk JM, Lloyd CJ, Brunk E, et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat Biotechnol*, 2017, 35(10): 904-908.
- [33] Hucka M, Finney A, Bornstein BJ, et al. Evolving a Lingua Franca and associated software infrastructure for computational systems biology: the systems biology markup language (SBML) project. *Syst Biol (Stevenage)*, 2004, 1(1): 41-53.
- [34] Ebrahim A, Lerman JA, Palsson BO, et al. COBRApy: COConstraints-based reconstruction and analysis for Python. *BMC Syst Biol*, 2013, 7: 74.
- [35] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 2003, 13(11): 2498-2504.
- [36] Caspi R, Billington R, Fulcher CA, et al. The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res*, 2018, 46(database issue): D633-D639.

(本文责编 郝丽芳)