

# 基于 Matlab 对解脂耶氏酵母基因组规模代谢网络模型 仿真结果的可视化分析

简星星<sup>1</sup> 高琪<sup>1</sup> 花强<sup>1,2\*</sup>

(1. 生物反应器工程国家重点实验室 华东理工大学生物工程学院 上海 200237)

(2. 上海生物制造技术协同创新中心 上海 200237)

**摘要:**【目的】近十年来,基因组代谢网络模型迅速发展。通过构建基因组代谢网络模型进行计算机仿真模拟已成为研究生物体复杂的生理代谢不可或缺的工具。实现对仿真结果的可视化分析,可以直观地追踪模型中的代谢流向,从而更好地对仿真结果进行分析。【方法】在简要概述目前可视化方法的基础上,提出了一种基于 Matlab 实现基因组规模代谢网络模型仿真结果可视化的方法:通过 CellDesigner 预先绘制与模型相匹配的图,通过 RAVEN toolbox 中的函数于 Matlab 进行读图、并实现仿真结果的可视化。【结果】以解脂耶氏酵母基因组规模代谢网络模型 iYL619\_PCP v1.7 为对象,实现并阐明其仿真结果的可视化。【结论】通过该方法可以清晰地监测模型中的流量和流向变化,提高仿真结果的分析效率。

**关键词:** 基因组规模代谢网络模型, 解脂耶氏酵母, 可视化, CellDesigner, RAVEN toolbox

## Matlab-based visualization analysis of genome-scale metabolic network model simulation of *Yarrowia lipolytica*

JIAN Xing-Xing<sup>1</sup> GAO Qi<sup>1</sup> HUA Qiang<sup>1,2\*</sup>

(1. State Key Laboratory of Bioreactor Engineering, School of Bioengineering, East China University of Science and Technology, Shanghai 200237, China)

(2. Shanghai Collaborative Innovation Center for Biomanufacturing Technology (SCICBT), Shanghai 200237, China)

**Abstract:** [Objective] In recent decade, genome-scale metabolic network model (GSMM) has been flourishing rapidly. It has been an indispensable tool to investigate complex physiology of organisms through reconstruction of GSMMs and prediction cellular characteristics based on computational simulation. Visualization of simulation results can help to trace intuitively metabolic flux changes in the model and generate possible metabolic engineering strategies. [Methods] In this study, we first briefly summarized current methods for metabolic network visualization, followed by proposing a new method to realize visualization of GSMMs simulation based on Matlab. A pre-drawing of the metabolic network was performed first via CellDesigner software. Two functions of RAVEN toolbox were employed to integrate the raw map into Matlab and to realize visualization of simulation. [Results] As

基金项目: 国家 973 计划项目(No. 2012CB721101); 国家自然科学基金项目(No. 31200025)

\*通讯作者: Tel: 86-21-64250976; 信箱: qhua@ecust.edu.cn

收稿日期: 2014-11-03; 接受日期: 2015-01-15; 优先数字出版日期(www.cnki.net): 2015-02-10

an example, the visualization of iYL619\_PCP v1.7, a GSMM of *Yarrowia lipolytica* was realized.

**[Conclusion]** According to the maps of visualization, we could monitor clearly flux changes in pathways under different environmental and genetic conditions to analyze efficiently simulation results.

**Keywords:** Genome-scale metabolic network model, *Yarrowia lipolytica*, Visualization, CellDesigner, RAVEN toolbox

自生物体的大规模全基因组测序和海量的注释信息以及高通量的组学数据涌现以来, 基因组规模代谢网络模型得到迅速发展<sup>[1-2]</sup>。构建生物体的基因组规模代谢网络模型, 并进行计算机仿真模拟已成为研究生物体复杂的生理代谢不可或缺的工具<sup>[3-4]</sup>。

基因组规模代谢网络模型的构建及其应用广泛且大量<sup>[5]</sup>, 将仿真结果可视化是一种高效分析仿真结果的方法, 也是目前模型发展与应用迫切的需求。而实现可视化最关键是如何在低维度空间(通常是二维平面)对节点进行合理布局, 反映代谢网络中复杂的拓扑关系<sup>[6]</sup>。基于一种或结合几种算法来对节点进行自动规划布局, 实现自动可视化分析、直观展现仿真结果依然是面临的一大挑战。

由于基因组规模代谢网络模型在各方面的应用, 已有一些可视化的软件被开发, 但多是为特定生化网络模拟而设计、具有特定功能的绘图工具。Cytoscape 就是其中一个在生物分子间相互作用方面应用最为广泛的可视化工具<sup>[7-8]</sup>, 以 Cytoscape 为基础而开发的相关插件还有很多, 对调控网络的可视化应用较强大。LucidDraw<sup>[6]</sup>是与 Cytoscape 类似的一个基于 Matlab 的生化网络可视化分析工具, 主要是对分子间的相互作用、调控信息的可视化。仍无法实现在代谢网络图上直观显示模型仿真模拟的结果, 即无法在代谢网络图上直接呈现出模型仿真模拟后各反应的流量大小, 也就无法实现对模型仿真结果的可视化分析。

可视化软件 Paint4Net<sup>[9]</sup>巧妙运用 Matlab 中函数 biograph 对代谢网络模型实现自动可视化。该软件允许研究人员自定义可视化范围, 但当反应数量大于 20 时自动生成的代谢图就比较杂乱。其主要原因在于函数 biograph 对于每个节点(代谢物)在图

中只能出现一个, 对于某些代谢物参与的反应较多时, 连接这些代谢物的反应连线就会很多, 因此严重影响各节点的布局和连线的绘制。

多功能生化网络建模工具 CellDesigner<sup>[10]</sup>以系统生物学标记语言(SBML)为模型(绘制的图文件)描述基础, 且可以输出多个版本的 SBML 格式(SBML level1 version2、SBML level2 version1、SBML level2 version2、SBML level2 version3、SBML level2 version4), 也可以输出“SBGN-ML”和“Image”格式。其中格式“SBML level2 version3”的 CellDesigner 图可以借助 RAVEN toolbox<sup>[11]</sup>在 Matlab 中对图进行读取, 并能将模型的仿真模拟结果在图上显示, 实现对模型仿真结果的可视化分析。

BiGG 数据库是一个系统生物学模型构建交流平台, 包含有 10 个基因组规模代谢网络模型(*E. coli* iJR904, *E. coli* iAF1260, *E. coli* textbook, *H. sapiens* Recon1, *S. cerevisiae* iND750, *M. barkeri* iAF692, *H. pylori* iT341, *S. aureus* iSB619, *M. tuberculosis* iNJ661, *P. putida* iJN746)以及与模型相对应的代谢网络图<sup>[12]</sup>。这些图都是通过商业软件 Simpheny 依据预先设计好的布局人工绘制。这些模型和图(有些模型由于反应数量较多, 因此有多个图文件)均可免费下载, 通过 COBRA toolbox<sup>[13]</sup>可以实现模型仿真结果的可视化。其可视化图是以灰度、线条的颜色来表征反应流量的大小, 流量为 0 的反应则不显示线条。缺憾是数据库中模型的数量和所涉及的物种较少, 同时还存在模型与图中反应命名不一致等情况, 这样会影响可视化后反应流量在图中的正常显示。因此在使用这些模型和图进行仿真模拟前, 需要先将命名不统一的反应找出、修改使其命名统一, 否则可能导致不正确的分析结果。

本文是以产油脂模式酵母——解脂耶氏酵母

的基因组规模代谢网络模型 iYL619\_PCP v1.7 实现对仿真结果的可视化分析为例, 介绍一种基于 Matlab 实现基因组规模代谢网络模型仿真结果可视化的方法。模型 iYL619\_PCP v1.7 源自本实验室构建的解脂耶氏酵母基因组规模代谢网络模型 iYL619\_PCP<sup>[14]</sup>, 相比最初的代谢网络模型 iYL619\_PCP, iYL619\_PCP v1.7 有部分反应的修改和更新。

## 1 方法

### 1.1 相关软件和算法

绘图是用软件 CellDesigner (CellDesigner v4.3) 完成的; 在其之前可以先用自动绘图软件 Paint4Net (Paint4Net v1.3)按途径局部绘图进行参照。本研究中使用 COBRA toolbox 2.0 在 Matlab [版本 7.12.0.635(R2011a)]上对模型 iYL619\_PCP v1.7 进行优化, 所用的 solver 包括'glpk' (v4.35)和 Tomlab/CPLEX (v7.9)。读图和可视化都使用的是 RAVEN toolbox (RAVEN v1.07)中的函数。

IdealKnock (未发表)是由本实验室最新开发的一个敲除算法, 针对目标产物最大化生产提出代谢改造策略(敲除反应和所敲除反应所对应的基因); 在这改造策略的基础上还可继续 OptKnock<sup>[15]</sup>运算, 找到更优的改造策略。本文中的改造策略就是 IdealKnock 与 OptKnock 联用所提出的。

### 1.2 CellDesigner 和 Paint4Net

通常情况下想要直接根据模型本身绘制与模型关联的代谢网络图是很困难的, 即利用 CellDesigner 直接绘制模型的代谢网络图几乎不可能。本研究是参考 Paint4Net 自动生成的图例, 用 CellDesigner 手动绘制代谢网络图。

Paint4Net 是一个基于 Matlab 的自动绘图软件, 直接加载路径, 在 Matlab 的命令窗口输入指令就可对目标反应自动绘图<sup>[9]</sup>。Paint4Net 自动生成的目标反应途径的图, 如图 1 所示, 可以保存为 Matlab 格式备用。

图 1 中显示的是 11 个反应自动绘制后的结果。

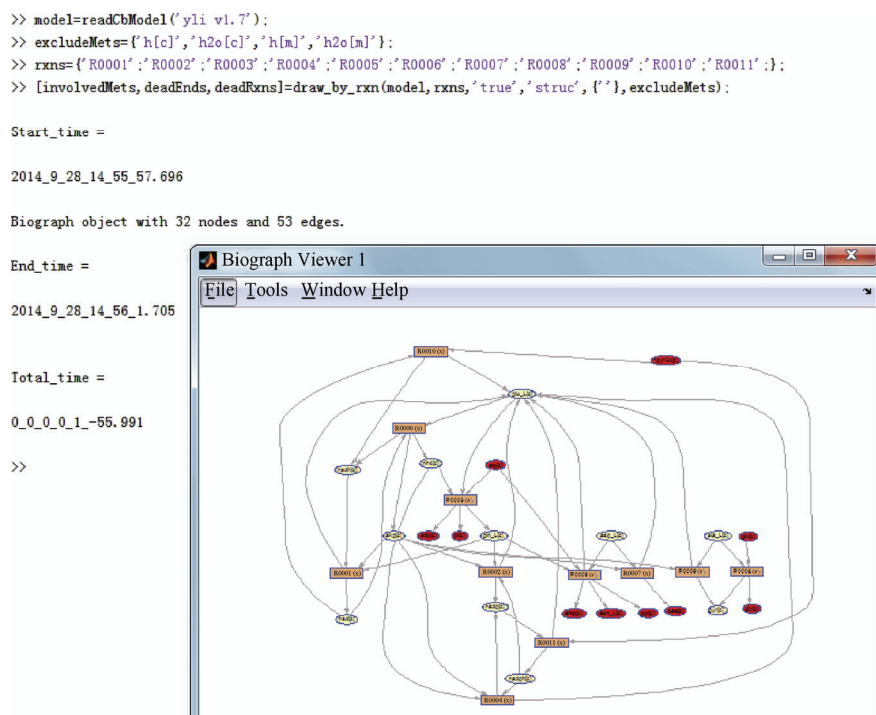


图 1 利用 Paint4Net 在 Matlab 中自动生成目标反应途径的代谢网络图

Figure 1 Automatic drawing using Paint4Net software in Matlab

Paint4Net 通常只能清晰地绘制小于 20 个反应的代谢网络。将目标代谢网络模型中的反应按途径(subSystems)分类, 借助 Paint4Net 自动绘制出每个途径草图, 然后根据该草图用 CellDesigner 更为精致地绘制出与模型相匹配的代谢网络图。为了清晰直观地反映模型中网络结构、代谢流量及便于对仿真结果的分析, 绘制图谱时也应该按单个途径绘制, 并注明某些代谢物在其他途径中的存在情况, 方便在分析时的查找和分析代谢反应的流向。CellDesigner 的操作界面如图 2 所示。

### 1.3 代谢网络绘制及可视化具体操作步骤

(1) 将模型中的反应按途径分类: 于 Matlab 读取模型, 并将反应、反应名称、反应途径(rxnns、rxnNames、subSystems)复制到 Excel 表格中, 按反应途径列排序, 将每个途径的反应归类。

(2) Paint4Net 按途径自动绘图: 将 Paint4Net 加载路径后, 利用函数 draw\_by\_rxn 将每个途径的反应自动绘图。当反应数量太多时可以多次绘图, 最后于 CellDesigner 中再整合成一条完整的途径。

Paint4Net 一次绘图, 图中每个代谢物仅出现一个。由于某些小分子物质(例如:  $H_2O$ 、 $H^+$ 、ATP、ADP、 $NAD^+$ 、NADH、 $NADP^+$ 、NADPH 等)参与的反应很多, 因此连接它们的线条也会相当密集杂乱, 这也是为什么 Paint4Net 不能用于对整个模型直接自动绘图的原因。在不影响反映反应的代谢关系时, 可以使用函数 excludeMets 在自动绘图前先将某些小分子物质排除(例如:  $H_2O$ 、 $H^+$ 。本文的示例图中参与反应的  $H_2O$ 、 $H^+$ 就预先被排除)。

(3) 根据 Paint4Net 用 CellDesigner 手动绘图: 根据 Paint4Net 图中所反映的各反应间的连接关系, 以及模型中反应的反应式(COBRA 函数 printRxnFormula 可以列出反应在模型中的反应式)于 CellDesigner 将反应绘制出来。其中“Simple Molecule”代表参与反应的反应物和生成物; “Generic Protein”代表催化反应的基因, 也即反应 ID (rxns); “State Transition”是用于连接主反应物和主生成物的有向线条, 若是可逆反应可以双击线条选择“true”, 线条即变成双向。对于其他参与的反应

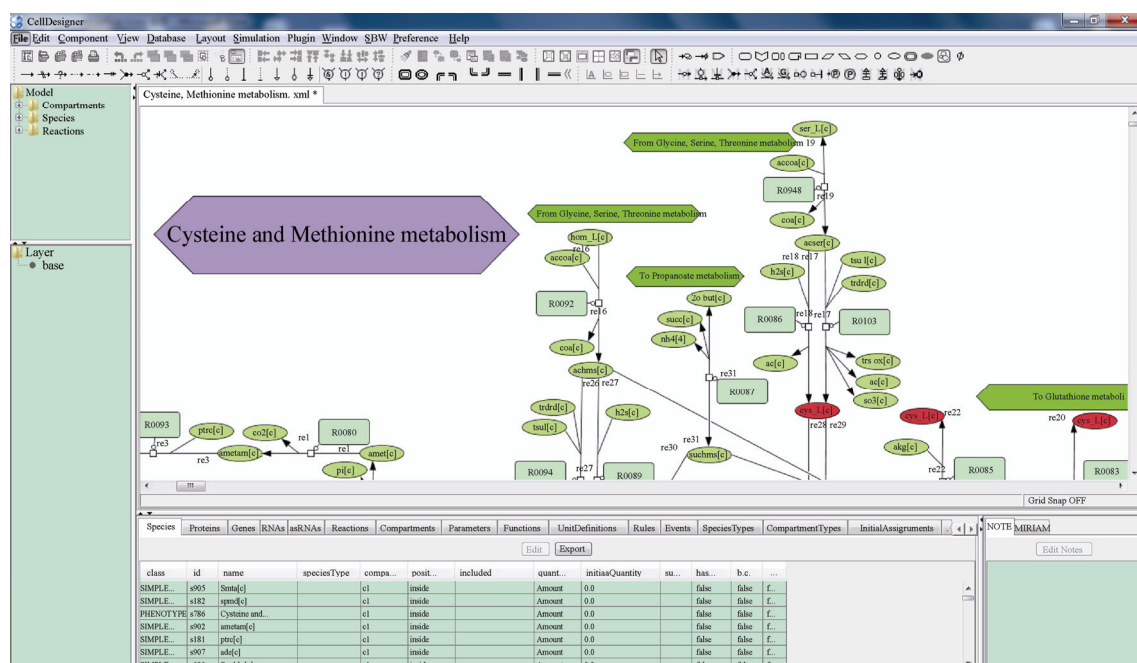


图 2 CellDesigner 绘图操作界面  
Figure 2 Screenshot of CellDesigner



物和生成物，可以通过“Add Reactant”和“Add Product”与线条连接加入到反应中；对于运输反应则用“Transport”代替“State Transition”，若可逆运输反应同样双击线条选择“true”，线条即变成双向。

(4) 将多个途径组合成大图：将所绘制的相关途径放置在同一 CellDesigner 图上组合成一张大图，有利于可视化后的查找和分析。本文中模型中的代谢途径组合成了 5 张大图：Carbohydrate metabolism、Amino acid metabolism、Lipid metabolism、Nucleotide metabolism、Cofactor-vitamin-energy metabolism。最终没有将所有反应组合成一张图的原因，一是整合后数据太大，不便于操作；二是整合后图幅会更加大，反而不利于查找和可视化结果分析。

(5) 在各大图中加入相应的运输反应：将每张大图图中所有代谢物所涉及的运输反应从所有运输反应中找出，并将其加入相应的大图、分布在区间上。本文示例模型中只有 3 个分区(Extracellular、Cytoplasm、Mitochondria)，因此运输反应只有胞外到细胞质、细胞质到线粒体两种。但是在可视化之前，需要将图中所有“Generic Protein”的 Species Notes 中写入反应 ID(rxns)，使其与模型中的反应一一对应。

(6) 通过 RAVEN toolbox 实现可视化：RAVEN toolbox 中的函数 `constructPathwayFromCelldesigner` 可以将 SBML 格式的 CellDesigner 图读入 Matlab 的工作区。而 RAVEN toolbox 中函数 `drawMap`，可以将模型的仿真结果于图上呈现，实现仿真结果的可视化分析。并且可以同时两种条件下的仿真结果可视化，同一反应在两种条件下的流量值运用公式  $\text{Log}_{10}(|\text{condition A}|/|\text{condition B}|)$  得到一个位于  $[-1,1]$  区间的值，这个值就对应了一种颜色作为该“Generic Protein”的填充颜色。其中这个值越接近 1，颜色越接近绿色；这个值越接近 -1，颜色越接近红色。两个流量值的变化越大，其填充颜色也会越深。图 3 概述了代谢网络的绘制和可视化的具体操作步骤。

## 2 结果

模型 iYL619\_PCP v1.7 共有 619 个基因，1 139 个反应：其中交换反应 125 个、胞外向细胞质的运输反应 129 个、细胞质向线粒体的运输反应 105 个、酶促反应 780 个。780 个酶促反应共有 59 条途径(subSystems)。通过 CellDesigner 绘制与模型 iYL619\_PCP v1.7 相对应的图后共组合了 5 张大图。这 5 张大图分别命名为：Carbohydrate metabolism、Amino acid metabolism、Lipid metabolism、Nucleotide

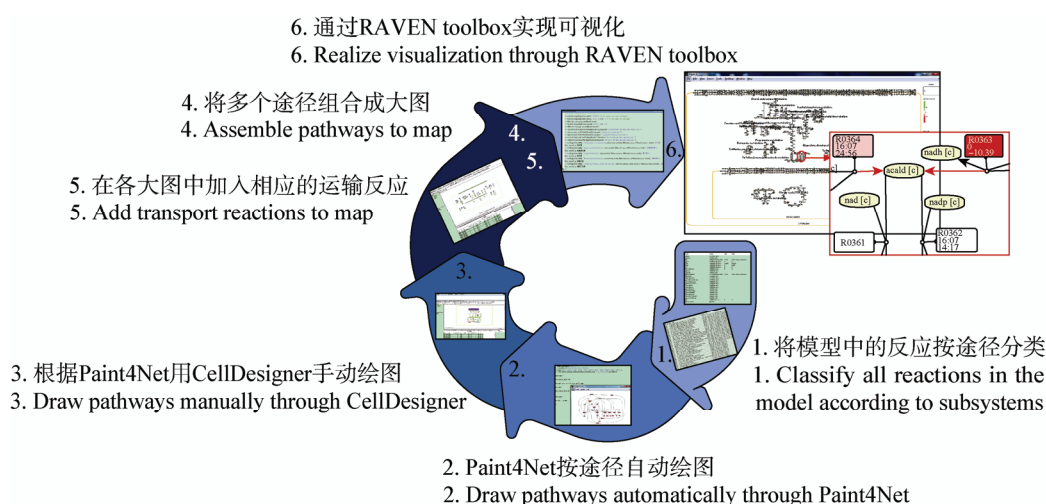


图 3 绘图、可视化步骤概述图

Figure 3 Overview of drawing and visualization

metabolism、Cofactor-vitamin-energy metabolism。为了展示可视化结果和可视化分析的优势,下面以两个仿真例子进行介绍。

## 2.1 可视化例 1

模型以琥珀酸在胞外与细胞质的交换反应(R1033)为目标函数,改变氧气交换反应(R0978)的下限为-1 000 作为条件 A、下限为-10 作为条件 B,然后在两种条件下分别对目标函数最大化,得到两组反应流量分别是 solutionA.x、solutionB.x。通过 RAVEN toolbox 中函数 construct Pathway From Celldesigner 读取模型 iYL619\_PCP v1.7 的 CellDesigner 图,然后通过函数 drawMap 将 solutionA.x、solutionB.x 于图上呈现,实现仿真结果的可视化。当反应的流量值小于默认值  $1 \times 10^{-7}$  (默认值可设置)时,看作反应流量为 0,图上的反应框中不显示。同一反应中的两组流量会自行比较:  $\text{Log}_{10}(|\text{condition A}|/|\text{condition B}|)$ ,得到一个值对应一种颜色填充反应框。也就是说,反应框中两个流量有变化,则会有填充颜色;两个流量大小相等没有变化,则是白色,且颜色越深两个流量变化越大。

部分反应的流量如表 1 所示。在碳代谢的可视化图中发现在条件 A 下氧气的运输反应(R0978)的流量是 21.88 mmol/(g DCW·h);条件 B 下氧气运输反应的流量是 10.00 mmol/(g DCW·h),说明条件 B

氧气是一个限定因数,且野生型(模型没有任何改变)对氧气的最大需求的流量是 21.88 mmol/(g DCW·h)。且两种条件下琥珀酸的运输反应(R1033)向胞外运输的流量分布是-28.03、-22.52 mmol/(g DCW·h),说明琥珀酸的生产受到了氧气影响。而 R1165 是乙醇的运输反应,两种条件下的流量分别是 0、-10.39 mmol/(g DCW·h);R0363 是乙醇合成反应,两种条件下的流量分别是 0、-10.39 mmol/(g DCW·h),说明在氧气供应充足时菌株不产乙醇,而受限时才产生大量乙醇运输到胞外,伴随产生更多  $\text{CO}_2$  (R1083),具体可视化结果如图 4 所示。在条件 B 下,柠檬酸合成反应(R0786)流量下降,说明由于氧气受限导致 TCA 受到抑制。

## 2.2 可视化例 2

模型 iYL619\_PCP v1.7 通过 IdealKnock 敲除算法(未发表)计算琥珀酸高产的反应敲除策略,结果显示敲除 66 反应(对应需敲 15 组基因)可以提高琥珀酸交换反应(R1033)的流量;再基于这 66 个反应与 OptKnock<sup>[15]</sup>联用,得到一组最优解:敲除模型中的 8 个反应(R0001、R0002、R0004、R0092、R0274、R0362、R0483、R0598),琥珀酸交换反应(R1033)流变化分析(FVA)得到最低流量和最高流量分别是 12.80 mmol/(g DCW·h)和 12.80 mmol/(g DCW·h),生物量(Biomass)反应(R2000)的流量是  $0.02 \text{ h}^{-1}$ 。

表 1 两种条件下部分反应的流量

Table 1 Fluxes comparison of part of metabolic reactions under two conditions (mmol/(g DCW·h))

反应 ID Reaction ID	反应名 Reaction name	条件 A Condition A	条件 B Condition B
R0978	O <sub>2</sub> transport (diffusion)	21.88	10.00
R1033	Succinate transport via proton symport	-28.03	-22.20
R1165	Ethanol reversible transport	0	-10.39
R0363	Ethanol:NAD <sup>+</sup> oxidoreductase	0	-10.39
R1083	CO <sub>2</sub> transporter via diffusion	-7.86	-9.13
R1084	CO <sub>2</sub> transport (diffusion), mitochondrial	-23.93	-16.12
R0786	Citrate hydroxymutase	11.97	7.09

注: 每个反应在模型中的方向定义为正方向,即可视化图中的红色箭头方向。流量值为正说明反应沿正方向进行,为负则相反。

Note: The direction of each reaction in the model is positive by default, i.e. the direction of red arrow on the visualization maps. The flux value is positive, indicating the reaction along the positive direction, reverse the opposite.

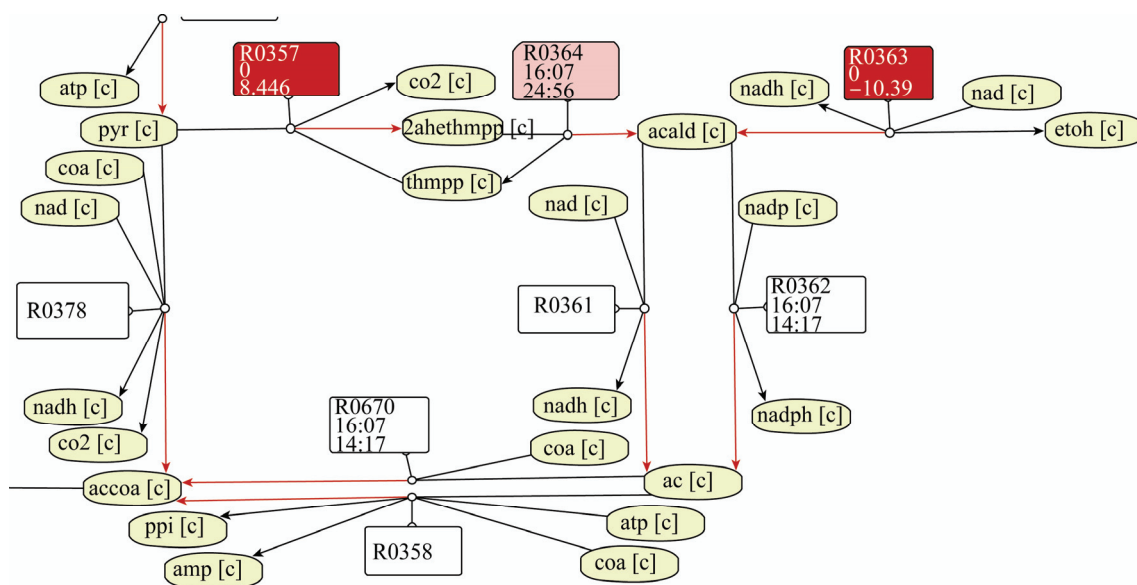


图4 两种条件下丙酮酸生产乙醇的情况

Figure 4 Ethanol production from pyruvate under two conditions

对上面的敲除策略进行可视化仿真结果分析,具体如图5所示。直接对模型敲除上述8个反应,模型的目标函数始终是生物量反应(R2000),然后对目标函数最大化得到一组流量 solutionB.x,其中琥珀酸交换反应(R1033)的流量是 12.80 mmol/(g DCW·h)、生物量反应(R2000)的流量是 0.02 h<sup>-1</sup>;作为对照不敲除反应、以生物量反应(R2000)为目标函数,对目标函数最大化得到一组流量 solutionA.x,琥珀酸交换反应(R1033)的流量是 0.01 mmol/(g DCW·h)、生物量反应(R2000)的流量是 0.04 h<sup>-1</sup>。表2显示了部分反应两种条件下的流量。

通过 RAVEN toolbox 中函数 construct Pathway From Celldesigner 和 drawMap 于 Matlab 中读图和将流量(solutionA.x 和 solutionB.x)于图上呈现,实现可视化。

由图5可见,敲除8个反应后琥珀酸交换反应(R1033)的流量由敲除前的 0.01 mmol/(g DCW·h)增长到 12.80 mmol/(g DCW·h),虽然生物量反应(R2000)流量有所下降,由敲除前的 0.04 h<sup>-1</sup>降到 0.02 h<sup>-1</sup>,但琥珀酸流量提高还是较为明显的。根据图5B可以看出敲除反应之后琥珀酸的大部分流量来自细胞质中丁酸代谢途径——R0118、R0889的

流量都是 12.71 mmol/(g DCW·h)):由α-酮戊二酸(akg[c])生成琥珀酸半醛(succsal[c]),再生成琥珀酸(succ[c]);图5C显示在敲除反应后TCA循环的流量:柠檬酸生成异柠檬酸(R0786),再生成α-酮戊二酸(R0787)的流量分别是 12.81 mmol/(g DCW·h)、12.73 mmol/(g DCW·h),生成了大量的α-酮戊二酸;α-酮戊二酸再由线粒体运输到细胞质的反应(R1092)流量为 12.72 mmol/(g DCW·h)。细胞质中的大量α-酮戊二酸如图5B所示,经R0118、R0889转化为大量的琥珀酸。生成的大量琥珀酸再经R1033运输到胞外。

敲除的8个反应包括有R0001、R0002、R0004、R0092、R0274、R0362、R0483、R0598。其中R0001、R0002、R0004都出现在丙氨酸、天冬氨酸和谷氨酸代谢途径(Alanine, aspartate and glutamate metabolism)中,被敲除是因为这3个反应都需要消耗α-酮戊二酸,会影响琥珀酸的生产;R0362被敲除避免了乙酸的生成或积累,也保证丙酮酸可以大量转化为α-酮戊二酸;R0483是磷酸戊糖途径(Pentose phosphate pathway)中前期的一个反应,被敲除可以保证糖酵解途径大量生产丙酮酸,为α-酮戊二酸、琥珀酸的生产提供前期物质;R0092

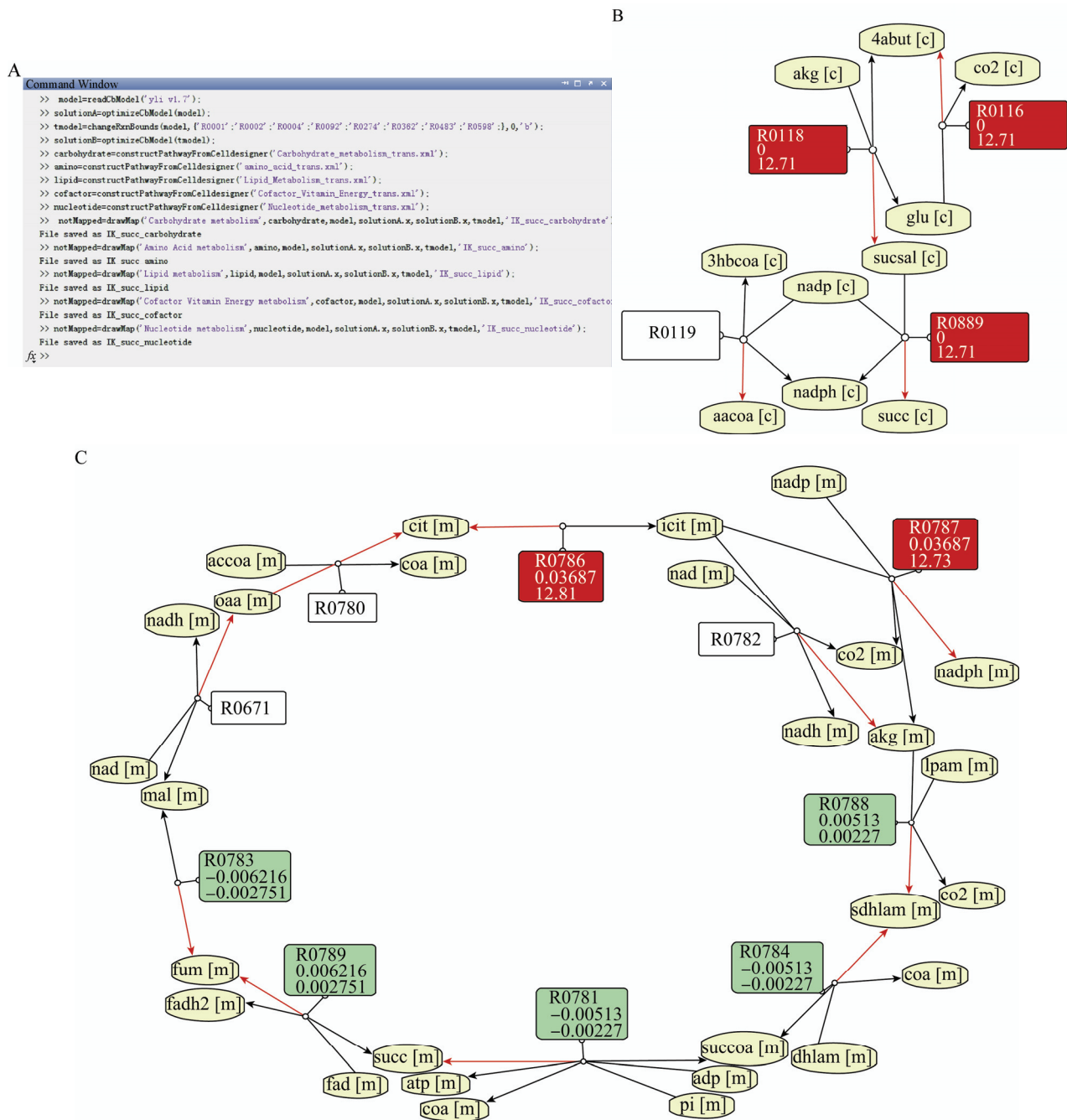


图5 可视化输入的指令和可视化结果的局部截屏

Figure 5 Screenshots of commands for visualization and some results

注: A: Matlab 命令窗口输入的可视化指令; B: 模型 iYL619\_PCP v1.7 中‘Butanoate metabolism’的可视化; C: 模型 iYL619\_PCP v1.7 中‘Citrate cycle (TCA cycle)’的可视化。

Note: A: Commands for visualization on Matlab command window; B: Visualization of Butanoate metabolism in iYL619\_PCP v1.7; C: Visualization of Citrate cycle (TCA cycle) in iYL619\_PCP v1.7.



表 2 两种条件下部分反应的流量  
Table 2 Fluxes comparison of part of metabolic reactions under two conditions

反应 ID Reaction ID	反应名 Reaction name	条件 A Condition A	条件 B Condition B
R1033	Succinate transport via proton symport	-0.01	-12.80
R2000	Biomass	0.04	0.02
R0889	Succinate_semialdehyde:NADP <sup>+</sup> oxidoreductase	0	12.71
R0118	4_aminobutanoate:2_oxoglutarate aminotransferase	0	12.71
R0786	Citrate hydroxymutase	0.04	12.81
R0787	Isocitrate: NADP <sup>+</sup> oxidoreductase (decarboxylating)	0.04	12.73
R0781	Succinate: CoA ligase (ADP_forming)	-0.01	-0.00
R0789	Succinate: (acceptor) oxidoreductase	0.01	0.00
R0978	O <sub>2</sub> transport (diffusion)	26.46	33.82
R0362	Acetaldehyde: NADP <sup>+</sup> oxidoreductase	30.90	0
R0483	6_phospho_D_gluconate: NADP <sup>+</sup> 2_oxidoreductase (decarboxylating)	0	0
R0598	3_Oxopropanoate: NADP <sup>+</sup> oxidoreductase (CoA_malonylating)	9.29	0
R0001	L_glutamate: NAD <sup>+</sup> oxidoreductase (transaminating)	0	0
R0002	L_Glutamate: NADP <sup>+</sup> oxidoreductase (transaminating)	0	0
R0004	L_Glutamate: NADP <sup>+</sup> oxidoreductase (deaminating)	0	0
R0092	Acetyl_CoA: L_homoserine O_acetyltransferase	0	0
R0274	5,10_Methylenetetrahydrofolate: glycine hydroxymethyltransferase	-4.77	0
R0279	3_Phospho_D_glycerate: NAD <sup>+</sup> 2_oxidoreductase	4.88	0.05
R1092	2_oxodicarboylate transporter, mitochondrial	0.03	12.72

是消耗乙酰辅酶 A 的反应, 敲除后可以保证柠檬酸的合成; R0274 是甘氨酸、丝氨酸和苏氨酸代谢途径(Glycine, serine and threonine metabolism)中的一个反应, R0274 和 R0294 都可以实现丝氨酸和甘氨酸的相互转化, R0274 被敲除后该途径中反应流量整体下降, 其中利用前体物质 3-磷酸甘油酸(3gp[c])的反应(R0279)的流量由敲除前的 4.88 mmol/(g DCW·h)降为 0.05 mmol/(g DCW·h), 在不影响该途径氨基酸供给的前提下, 保证糖酵解途径中的 3-磷酸甘油酸可以大量转化为丙酮酸。

### 3 讨论

实现模型仿真结果的可视化之前需要通过 CellDesigner 绘制与模型相匹配的图。绘制的过程同时也是模型中代谢末端(Metabolic dead-ends)和代谢漏洞(Metabolic gaps)的发现过程。参考最新文献或数据库信息对模型进行补充和完善, 也可以提高模型预测的准确性。

基于模型仿真预测通常通过 FVA (Flux

variability analysis)分析来确定各反应的流量变化范围。此方法是通过 FBA (Flux balance analysis)分析得到的一组最优解, 这组最优解是不唯一的。所以进行仿真结果的可视化分析时, 对于颜色较浅(流量变化不大)的反应, 为确保分析的准确可以参考该反应在两种条件下的 FVA 流量变化范围。

将模型的仿真结果于图上直观地呈现, 对两种不同条件下的反应流量还会进行对比:  $\text{Log}_{10}(|\text{condition A}|/|\text{condition B}|)$  得到一个位于区间  $[-1, 1]$  的值对应可视图的右上角的灰度条上的一种颜色, 该颜色作为该反应框的填充颜色来表征流量的变化, 其颜色的深浅表征流量变化的大小。而当两种条件下的流量值相差大于 10 倍时, 甚至可能相差的倍数是无穷, 根据以上公式计算的结果会大于 1 或者小于 -1, 此时反应框的颜色是按 1 或者 -1 进行处理。当条件 A 的流量绝对值大于条件 B 流量的绝对值时, 对应灰度条  $[0, 1]$  区间呈绿色; 相反则对应灰度条  $[-1, 0]$  区间呈红色。但当两组流量一个为正值一个为负值时不适用以上计算灰度的公式,

此时反应框内不会有填充颜色,而框的颜色变橙色来区分。另外,当反应的流量值小于默认值时定义该反应流量为 0,也就不会出现在反应框中;但当反应框中另一个流量值不为 0 时,此时反应框中才会出现 0 流量。

此方法只对两种不同条件下的途径流量进行比较,是由公式  $\text{Log}_{10}(|\text{condition A}|/|\text{condition B}|)$  决定的:运用该公式计算每个反应的流量变化情况,决定该反应框的填充颜色。对两种条件下进行比较,也是考虑到一种条件作为实验组,另一种条件作为对照组。若实现一种条件下可视化、各反应的流量间纵向比较(基于所有反应流量的排序)来定义反应框的填充颜色基于代码也是可以实现。若需要对多种条件下( $\geq 3$ )实现仿真结果的可视化分析,可以基于此方法实现两两的比较(实验组、对照组)。基于此方法实现同时对多种条件( $\geq 3$ )下的仿真结果进行可视化分析目前还无法实现,最大的挑战还在于如何定义反应框的填充颜色。

图中反应框的颜色,表征了两种条件下反应流量的变化,也便于从其他反应中区分和查找:在可视化后的图上基于反应框填充颜色的不同追踪反应流量的变化,并直观监测模型中的物质流方向和流量大小,便于分析改造策略致使目标物质产量提升的原因,提高对仿真结果的分析效率。

BiGG 数据库上的模型和图也可以实现可视化。与本文所介绍的方法相比,其只是对单个条件下的仿真结果可视化,并以灰度、线条的颜色来表征该条件下所有反应的流量大小,不会显示具体的流量值。当对仿真前后两种条件下的流量进行对比分析时,需要进行两次可视化并对两次可视化的图进行比较分析。本文介绍的方法则可以同时对两组流量可视化,显示流量值、并以颜色灰度表征两种条件下反应流量的变化大小,便于区分、查找和分析。

对于可视化的发展若能基于一种或者结合几种有效的算法对模型中的反应和代谢物进行自动或半自动规划布局、绘图,取代人工绘制,将可以提高绘图效率。若有相关软件可以实现按途径地局

部自动绘图,并允许人工编辑、将局部图组合成大图保存,也可以取代人工绘图这一费时过程。另外,若能开发一款读图软件,直接对可视化后的图进行扫描识别,可以人为控制显示所需的代谢网络,对于反应在图上快速定位和可视化分析都将是很有意义的。

对于可视化过程若能加入更多的自动分析代码,对反应的流量进行更全面的比较和分析、并以不同的方式表征在可视化后的图上,便于可视化分析时的查找和区分,将进一步提高仿真结果的分析效率。

## 参考文献

- [1] Notebaart RA, van Enckevort FH, Francke C, et al. Accelerating the reconstruction of genome-scale metabolic networks[J]. BMC Bioinformatics, 2006, 7: 296
- [2] Francke C, Siezen RJ, Teusink B. Reconstructing the metabolic network of a bacterium from its genome[J]. Trends in Microbiology, 2005, 13(11): 550-558
- [3] Kitano H. Systems biology: a brief overview[J]. Science, 2002, 295(5560): 1662-1664
- [4] Kitano H. Computational systems biology[J]. Nature, 2002(420): 206-210
- [5] McCloskey D, Palsson BO, Feist AM. Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*[J]. Molecular Systems Biology, 2013, 9(1): 661
- [6] He S, Mei J, Shi G, et al. LucidDraw: efficiently visualizing complex biochemical networks within MATLAB[J]. BMC Bioinformatics, 2010, 11: 31
- [7] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks[J]. Genome Research, 2003, 13(11): 2498-2504
- [8] Smoot ME. Cytoscape 2.8: new features for data integration and network visualization[J]. Bioinformatics, 2011, 27(3): 431-432
- [9] Kostromins A, Stalidzans E. Paint4Net: COBRA Toolbox extension for visualization of stoichiometric models of metabolism[J]. BioSystems, 2012, 109(2): 233-239
- [10] Funahashi A, Matsuoka Y, Jouraku A, et al. CellDesigner 3.5: A Versatile Modeling Tool for Biochemical Networks[J]. Proceedings of the IEEE, 2008, 96(8): 1254-1265
- [11] Agren R, Liu LM, Shoaie S, et al. The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for *Pencillium chrysogenum*[J]. PLoS Computational Biology, 2013, 9(3): e1002980
- [12] Schellenberger J, Park JO, Conrad TM, et al. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions[J]. BMC Bioinformatics, 2010, 11: 213
- [13] Schellenberger J, Que R, Fleming RM, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0[J]. Nature Protocols, 2011, 6(9): 1290-1307
- [14] Pan P, Hua Q. Reconstruction and In silico analysis of metabolic network for an oleaginous yeast, *Yarrowia lipolytica*[J]. PLoS One, 2012, 7(12): e51535
- [15] Burgard AP, Pharkya P, Maranas CD. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization[J]. Biotechnology and Bioengineering, 2003, 84(6): 647-657