生 物 工 程 学 报
Chinese Journal of Biotechnology
http://journals.im.ac.cn/cjbcn
DOI: 10.13345/j.cjb.150118

January 25, 2016, 32(1): 1−13

综 述

基于代谢网络预测菌种基因改造靶点方法的研究进展

李培顺[1,2,3]，马红武[3]，赵学明[1,2]，陈涛[1,2]

1 天津大学化工学院生物工程系，天津 300072
2 天津大学教育部系统生物工程重点实验室，天津 300072
3 中国科学院天津工业生物技术研究所 中国科学院系统微生物技术重点实验室，天津 300308

李培顺, 马红武, 赵学明, 等. 基于代谢网络预测菌种基因改造靶点方法的研究进展. 生物工程学报, 2016, 32(1): 1–13.
Li PS, Ma HW, Zhao XM, et al. Predicting genetic modification targets based on metabolic network analysis–a review. Chin J Biotech, 2016, 32(1): 1–13.

摘 要：高产特定产品的人工细胞工厂的构建需要对野生菌株进行大量的基因工程改造，近年来随着大量基因组尺度代谢网络模型的构建，人们提出了多种基于代谢网络分析预测基因改造靶点以使某一目标化合物合成最优的方法。这些方法利用基因组尺度代谢网络模型中的反应计量关系约束和反应不可逆性约束等，通过约束优化的方法预测可使产物合成最大化的改造靶点，避免了传统的通过相关途径的直观分析确定靶点的方法的局限性和主观性，为细胞工厂的理性设计提供了新的思路。以下结合作者的实际研究经验，对这些菌种优化方法的原理、优缺点及适用性等进行详细介绍，并讨论了目前存在的主要问题和未来的研究方向，为人们针对不同目标产品选择合适的方法及预测结果的可靠性评估提供了指导。

关键词：基因组尺度，代谢网络，菌种优化，系统生物学，代谢工程

Received: March 19, 2015; **Accepted:** May 20, 2015
Supported by: National Basic Research Program of China (973 Program) (Nos. 2012CB725203, 2011CBA00804), National High Technology Research and Development Program of China (863 Program) (No. 2012AA022103), Applied Basic Research Program of Tianjin (No. 12JCYBJC33000).
Corresponding authors: Hongwu Ma. Tel: +86-22-24828735; E-mail: ma_hw@tib.cas.cn
Tao Chen. Tel: +86-22-27406770; E-mail: taochen@tju.edu.cn
国家重点基础研究计划 (973 计划) (Nos. 2012CB725203, 2011CBA00804)，国家高技术研究发展计划 (863 计划) (No. 2012AA022103)，天津市应用基础及前沿技术研究计划项目 (No. 12JCYBJC33000) 资助。

Predicting genetic modification targets based on metabolic network analysis–a review

Peishun Li[1,2,3], Hongwu Ma[3], Xueming Zhao[1,2], and Tao Chen[1,2]

1 *Department of Biochemical Engineering*, *School of Chemical Engineering & Technology*, *Tianjin University*, *Tianjin* 300072, *China*
2 *Key Laboratory of Systems Bioengineering*, *Ministry of Education*, *Tianjin University*, *Tianjin* 300072, *China*
3 *Key Laboratory of Systems Microbial Biotechnology*, *Tianjin Institute of Industrial Biotechnology*, *Chinese Academy of Sciences*, *Tianjin* 300308, *China*

Abstract: Construction of artificial cell factory to produce specific compounds of interest needs wild strain to be genetically engineered. In recent years, with the reconstruction of many genome-scale metabolic networks, a number of methods have been proposed based on metabolic network analysis for predicting genetic modification targets that lead to overproduction of compounds of interest. These approaches use constraints of stoichiometry and reaction reversibility in genome-scale models of metabolism and adopt different mathematical algorithms to predict modification targets, and thus can discover new targets that are difficult to find through traditional intuitive methods. In this review, we introduce the principle, merit, demerit and application of various strain optimization methods in detail. The main problems in existing methods and perspectives on this emerging research field are also discussed, aiming to provide guidance to choose the appropriate methods according to different types of products and the reliability of the predicted results.

Keywords: genome-scale, metabolic network, strain optimization, systems biology, metabolic engineering

生物技术正在被广泛地应用于一些工业产品例如抗生素、生物能源、生物材料、大宗化学品等的生产。生物法相对于传统的化学法具有反应条件温和、节约能源、转化率高和环境友好等优势。为了进行高效生产，通常需要对工业微生物进行一系列基因改造以提高目标产物的生产速率及得率。代谢工程主要任务之一就是识别工业微生物基因改造的靶点以达到高产目标化合物的目的。早期基因改造靶点的确定主要依靠对局部代谢途径的分析和实验的经验。随着系统生物学和合成生物学的发展，在基因组测序和注释海量数据的基础上各种微生物的基因组尺度代谢网络模型相继被重构出来[1-5]。它们已经成为研究生物代谢系统不可缺少的工具，并广泛应用于预测能提高目标化合物生产的微生物代谢工程改造靶点。基于约束的表型模拟方法，比如通量平衡分析 (FBA)[6-7]，能计算出特定网络中使目标产品得率最大的最优途径，从而确定基因过表达靶点。相比之下，通过代谢网络分析确定基因敲除靶点要更困难一些。

最早用来预测基因敲除靶点的方法是最小化代谢调整 (MOMA) 算法[8]，它假设经过基因敲除的突变菌在达到稳态过程中会受到最小化代谢调整的影响而使新代谢通量分布与野生型菌株的通量分布之间的欧几里得距离最小，由它求得的解为满足约束条件的次优解。MOMA 算法预测到的基因敲除靶点成功应用于毕赤酵母 *Pichia pastoris* 的基因工程改造以高效生产胞质人类超氧化物歧化酶 (hSOD)。靶点乙醇脱氢

酶基因 *adh2* 的敲除使 hSOD 的产量提高了 20%，并且没有减少细胞生长。另外，该方法预测到的敲除靶点成功地指导了大肠杆菌 *Escherichia coli* 代谢工程改造以过量生产番茄红素、缬氨酸及聚乳酸等产品。MOMA 算法需要先选择几个基因或基因组合进行敲除然后评估其对生长和产物合成的影响。对包含上千个基因的代谢网络多基因组合靶点的计算非常困难，因此 MOMA 更多用于在确定敲除靶点后评估其对代谢通量分布的影响，并不适用于从一复杂网络中找出使产物生成最大化的敲除靶点组合。针对这一问题人们提出了一些优化目标化合物生产的敲除靶点预测方法，例如 OptKnock[9]、OptGene[10]等。大多数方法是通过细胞生长和产物合成的双层优化找到将生长和产物合成相偶联的最优基因敲除策略，已成功用于指导大肠杆菌和酿酒酵母 *Saccharomyces cerevisiae* 的代谢工程改造以优化部分重要生化产品的生产。本文将结合作者的研究工作，重点针对这类方法详细阐述其基本原理、特征及优缺点，并介绍几种在优化中常用的软件。

1 OptKnock 及双层优化基本思路

2003 年 Burgard 等[9]提出了第一个基于基因组尺度代谢网络模型进行敲除靶点预测的双层优化方法 OptKnock。该方法利用双层混合整型线性规划 (MILP) 来进行优化计算，如图 1 所示。该双层结构中内层问题是生物量最大化时的约束线性规划问题，包括化学计量学、反应通量上下限、热力学、反应敲除等约束条件。外层问题则是以目标化合物合成最大化为目标的最优化问题，确定哪些反应敲除可以在满足内层优化的同时也可使产物生成最大化。

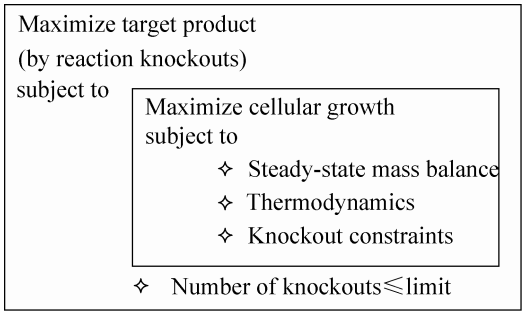

图 1 OptKnock 方法的双层结构

Fig. 1 The bilevel structure of OptKnock.

为了减少计算时间和提高预测敲除靶点的效率，OptKnock 方法预测的敲除靶点数目一般限制在 5 个以内，并且在进行计算之前需要对基因组尺度代谢网络进行预处理[11]，主要包括以下步骤：1) 通过通量可变性分析 (Flux variability analysis，FVA)[12]，除去模型中通量始终为零的反应，并缩小反应的通量变化范围。2) 为了减小候选敲除靶点的范围，需除去以下 4 类反应：a) 干、湿实验中对于生物量生长是必需的；b) 非基因关联的反应、自发反应和扩散反应，实际上这些反应在湿实验中是无法被敲除的；c) 特定子系统的反应，比如细胞膜生物合成、膜脂质代谢、无机离子运输等；d) 对于偶联反应子集，只对其中一个反应进行分析，因为敲除该子集中任何一个反应都是等效的。

基于大肠杆菌代谢网络模型，利用 OptKnock 方法以琥珀酸、乳酸及 1,3-丙二醇等产品为目标进行优化计算，能得到使生物量生长和目标产品合成相偶联的靶点组合。以琥珀酸为例，OptKnock 预测到的敲除靶点基因组合为乙醇脱氢酶、乳酸脱氢酶、磷酸葡萄糖异构酶和丙酮酸甲酸裂解酶。敲除这些靶点后在无氧条件最大化细胞生长时必须生成琥珀酸，即

通过基因敲除使琥珀酸成为细胞生长时的一种强制性产物，并且不会导致细胞无法生长。对此可以由图 2 中琥珀酸合成速率与生长速率间的关系予以说明。图中虚线为野生菌在设定生长速率为最大生长速率范围内任一值时通过 FBA 计算得到的琥珀酸最大和最小生成速率。其最小速率始终为 0，而最大速率则随着生长速率的增加而不断减小，这表明此时琥珀酸生成和生长是一种竞争关系，因此琥珀酸生成对生长不利，细胞最优化生长的结果将使得副产物琥珀酸生成为 0。而在敲除 OptKnock 计算得到的 4 个靶点基因后再进行计算的结果如图中实线所示。此时最优生长速率降低，并且在生长速率为 0.052 h^{-1} 到 0.09 h^{-1} 范围内琥珀酸最小生成速率随生长速率的增大而增大，在最优生长速率时琥珀酸最大和最小生成速率曲线交汇到一点 (圆圈所示)。这意味着要最大化细胞生长就必须要生成琥珀酸，琥珀酸生产与细胞生长相偶联。因此对突变株细胞自身调控优化生长

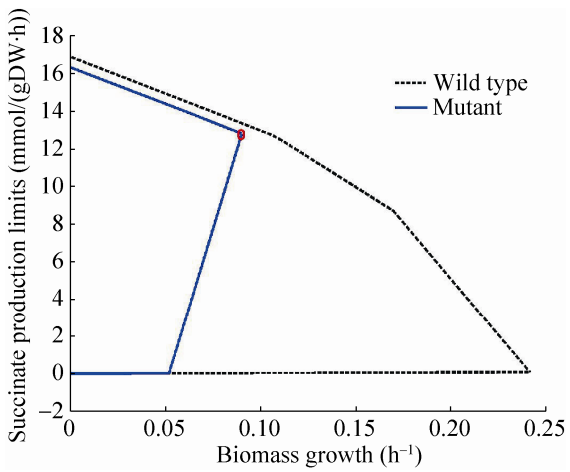

图 2 琥珀酸最大和最小生产速率随生长速率的变化
Fig. 2 Maximal and minimal succinate production rates at different gowth rates.

的同时也就优化了琥珀酸的生成。这正是 OptKnock 等双层优化方法进行靶点预测的生物学基础。对应用 OptKnock 方法针对一些目标产品，如琥珀酸、乳酸、1,3-丙二醇等预测得到的部分敲除靶点人们已经进行了湿实验验证，能够显著提高目标化合物的产量[13]。

2 OptKnock 衍生的菌种优化方法

2.1 预测敲除靶点的方法

2.1.1 OptGene

由于基因组尺度代谢网络的复杂性、冗余性和双层结构的混合整型线性规划问题求解面临着巨大的挑战，导致 OptKnock 方法搜索到敲除靶点的计算时间过长，甚至搜索不到结果。基于此，Patil 等[10]利用遗传算法 (GA) 提出了一种新的菌种优化方法 OptGene。遗传算法是一种通过模拟达尔文生物进化论的原理搜索最优解的方法，能够在相对较短的时间内为超大问题提供近似最优解，但是很多时候无法得到全局最优解，而是得到在预设的时间或者突变次数达到之后的局部最优解，并且 OptGene 较 OptKnock 没有速度上的提升。不过，OptGene 相对 OptKnock 的优势是它既能够进行反应水平的敲除，也能够进行基因水平的敲除。因此，它不需要像 OptKnock 一样排除自发反应，只需要确定哪些基因是可以敲除的。另外，OptGene 还能优化包含非线性目标函数的问题。

基于酿酒酵母代谢网络模型，利用 OptGene 方法以琥珀酸、甘油及香草醛为目标进行优化计算，预测到的靶点能将生物量生长和这些目标产品的生产进行偶联。

2.1.2 RobustKnock

随着基因组尺度代谢网络维度的增大，代

谢网络中会存在与目标化合物最优合成相竞争的途径，它们的存在能够使代谢通量偏离目标化合物，从而导致较低的生产速率，甚至得不到目标产物。例如，OptKnock 方法基于大肠杆菌基因组尺度代谢网络模型 iJO1366[14]以乳酸为目标进行优化计算，由于竞争性途径的存在，预测到的靶点会导致乳酸的生产速率依然为 0。2009 年，Tepper 等提出了能够识别代谢网络中竞争性途径的菌种优化方法 RobustKnock[15]。与 OptKnock 不同，RobustKnock 外层问题是最大化目标化合物最小生成速率的优化问题，从而识别到的敲除靶点能够使目标化合物的生产成为生物量生长时的一种强制性产物。RobustKnock 与 OptKnock 优化方法预测敲除靶点的差异可以由图 3 中的简单例子说明，该图中简化的网络模型由 5 个代谢物 (M_1–M_5) 组成，V_{uptake}、$V_{biomass}$、$V_{product}$ 和 $V_{by\text{-}product}$ 分别表示底物、生物量、目标产物和副产物的速率，由 M_1 到 M_3 的反应计量系数为 2∶1，即 2 分子 M_1 生成 1 分子 M_3，其他反应的计量系数关系都是 1∶1。当以生物量生长速率 $V_{biomass}$ 为目标对

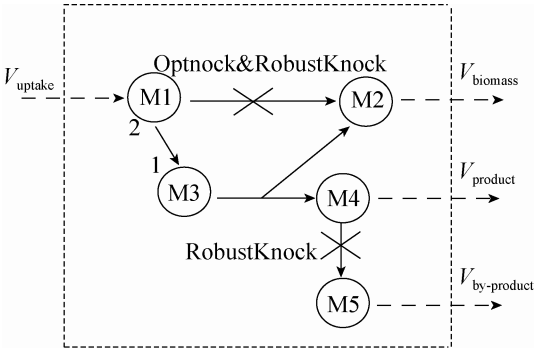

图 3　RobustKnock 与 OptKnock 预测敲除策略的差异
Fig. 3　Difference of knockout strategies for RobustKnock and OptKnock.

该网络模型进行通量平衡分析时，通量分布结果是生物量最优生长速率 $V_{biomass}$ 等于 V_{uptake}，而目标产物生成速率 $V_{product}$ 为 0，即此时代谢通量只经过反应 $M_1 \rightarrow M_2$。OptKnock 基于该网络模型以 $V_{product}$ 为目标进行优化计算，预测到的敲除靶点 (反应 $M_1 \rightarrow M_2$) 能够使目标化合物的最大生产速率达到 $0.5 \times V_{uptake}$，但是当敲除反应 $M_1 \rightarrow M_2$ 并以生物量生长速率为目标进行通量平衡分析时，目标化合物的生产速率可能为 0，因为此时代谢通量会流向竞争性产物 M_5。而应用 RobustKnock 基于该网络模型进行优化计算预测到的敲除靶点则包括 $M_1 \rightarrow M_2$ 和 $M_4 \rightarrow M_5$ 两个反应，能够使目标产物的最小生产速率达到 $0.5 \times V_{uptake}$，从而保证了目标产物的生成与生物量生长相偶联。

基于大肠杆菌代谢网络模型，应用 RobustKnock 方法以乙酸、甲酸、延胡索酸、乙醇、琥珀酸和乳酸这些产品为目标进行优化计算，预测到的靶点都能将生物量生长和这些目标产品合成进行偶联，这表明与 OptKnock 方法相比，RobustKnock 方法预测的敲除靶点更具有鲁棒性。

2.1.3　OptSwap

微生物代谢过程中氧化还原酶 (如脱氢酶, 还原酶) 通常与辅因子 NAD (H) 和 NADP (H) 这两个主要流通代谢物其中的一种有结合特异性。正是氧化还原酶的这种结合特异性才导致代谢功能的分工：将 NAD^+ 还原为 NADH 的酶能驱动氧化磷酸化，而将 $NADP^+$ 还原为 NADPH 的酶能驱动合成代谢反应，这样生物系统才能调控资源流向能量生成或合成代谢。因此，优化辅因子生成的策略在菌种优化方法中也应该

考虑。King 等开发的 OptSwap[16]方法能通过预测氧化还原酶辅因子交换位点和敲除靶点来优化目标产物的生产。该方法是基于 RobustKnock 的双层 MILP 问题，并且添加氧化还原酶辅因子特异性交换作为约束条件，结合辅因子交换和反应敲除来获得优化目标产物的菌种改造靶点。

基于大肠杆菌基因组尺度代谢网络模型 iJO1366，OptSwap 以丙氨酸、琥珀酸、乙酸及乳酸这些产品为目标进行优化计算，预测到的改造靶点能将生物量生长和这些产品的生成相偶联。针对丙氨酸和乳酸进行优化计算时，同时敲除 3 个反应和交换 1 个氧化还原酶辅因子结合位点才能将这两种产物的生成与生物量生长相偶联，而只敲除 4 个或以下反应时都不能将两者偶联起来。针对琥珀酸和乙酸的优化计算表明，同时考虑辅因子交换位点和敲除靶点的菌种优化策略比仅仅考虑敲除靶点的效果要好。OptSwap 方法拓展了能与生物量生长相偶联的目标产品的范围，但是比 RobustKnock 计算要求更高[16]。

2.1.4 GDBB

为了解决双层优化方法计算时间过长的问题，2012 年 Egen 等利用分支定界 (Truncated branch and bound) 算法提出了一种新的菌种优化方法 GDBB[17]。与 OptKnock 相比，该方法能够在更短时间内计算得到更多敲除数的基因敲除策略，但是它只能找到优化问题的近似最优解，很多时候无法得到全局最优解。

基于大肠杆菌基因组尺度代谢网络模型 iAF1260，利用 GDBB 以乙酸和琥珀酸为目标进行优化计算，预测到的靶点能将生物量生长和目标产物的生产进行偶联，并且近似最优的敲除策略分别在 81 s 和 345 s 时被找到。

2.1.5 其他预测敲除靶点的方法

与 OptGene 方法的求解方式不同，GDLS[18]方法采用局部搜索的启发式算法，在求解空间中进行有效、低复杂度的多路径搜索，这样可以降低计算时间。有研究表明经过基因改造的突变菌事实上在达到稳态过程中会受到最小化代谢调整的影响，从而使调整之后的代谢通量分布与野生型菌株的通量分布之间的改变最小，而并非以生物量生长为目标大幅调整通量分布。2013 年 Ren 等基于此提出了一种新的双层优化方法 MOMAKnock[19]，它的内层问题是以最小化代谢调整为目标时的二次规划问题。结果表明该方法能在最小化代谢调整通量分布情况下找出具有鲁棒性的敲除靶点。另外，Xu 等开发的 ReacKnock[20] 方法采用 Karush-Kuhn-Tucker (KKT) 算法求解双层混合整型线性规划问题，这为双层规划问题的求解提供了一种新思路。最近，Zhang 等结合 LTM (Logic transformation of model) 方法与预测反应敲除的 OptKnock 双层优化方法开发了 OptGeneKnock[21]，并且利用分支定界算法进行求解，不仅能直接进行基因水平的预测，还能较快地计算得到使目标化合物过量生产的近似最优的敲除策略。

2.2 预测敲除、上调和下调靶点的方法

2.2.1 OptReg

由于微生物菌种的基因改造手段并不局限于基因的添加和完全敲除，而且研究表明基因表达的上调或下调会对胞内代谢产生重要影响[22]，因此有必要开发能够识别基因上调或下调靶点的菌种优化方法。2006 年，Pharkya 等开发了一种新的菌种优化方法 OptReg[23]，该方法

在OptKnock基础上进一步整合了反应通量的调节机制，能够同时识别令目标产物过量生产的上调、下调和敲除靶点，但是由于拓展了识别基因改造靶点的范围，导致双层规划问题含有更多变量和非线性关系，进一步增加了转化为可求解的单层优化问题的难度，计算上面临着巨大挑战。假设某一反应 V_j 在模型中通量范围为[0, 1 000]，而依据通量可变性分析 (FVA) 确定的反应上下限为[$V_{j,L}$, $V_{j,U}$]。在OptReg方法中，当限制 V_j 在 ($V_{j,U}$, 1 000] 区间时，表示该反应被上调；当限制 V_j 在 (0, $V_{j,L}$) 区间时，表示该反应被下调；当限制 V_j 为 0 时，则表示该反应被敲除，如图 4 所示。

基于大肠杆菌代谢网络模型，OptReg 方法以乙醇为目标进行优化计算，预测到的靶点能将生物量生长和乙醇进行偶联，研究表明菌种改造过程中反应敲除和调节之间存在着协同作用。

2.2.2 OptORF

由于多功能酶、多亚基酶复合物及同工酶的存在，基因与反应之间的关系并非总是一一映射，导致基于反应敲除的突变菌株有时很难在实验过程中被构建。此外，微生物细胞代谢还会受到转录调控网络的影响。而大多数菌种优化方法是基于反应敲除，而且没有考虑转录调控网络。为了克服以上缺点，2010 年 Kim 等开发了 OptORF[24]，该方法能利用基因组尺度代谢与转录调控整合网络模型去识别代谢基因敲除、过表达和转录因子敲除的菌种改造靶点。OptORF 方法应用于一个简化的代谢与转录调控整合网络的实例如图 5 所示。网络中底物 S 通过中间代谢物 M1 或 M2 能被转化为生物量 B。假设反应 R2 能将 M1 转化为目标产物 P1 和 0.08 B，而反应 R5 能将 M2 转化为副产物 P2 和 0.12 B。当底物 S 存在，转录因子 TF1 呈激活状态，由此激活基因 G3 和 G5 的表达，并抑制 G1A 的表达。最大化生物量生长时，代谢通量会流过反应 R3 和 R5，因此会生成副产物 P2，而不会生成目标产物 P1。以 P1 为目标产品，利用 OptORF 方法对该网络进行优化计算，预测到的基因改造靶点包括敲除基因 G3 和 G4 (关联反应 R3、R4) 或 G5 和 G6 (关联反应 R5)。由于基因 G1A 的表达被转录因子 TF1 所抑制，导致反应 R1 无法进行。因此，OptORF 在预测到上述敲除靶点的同时还会识别到过表达基因 G1A。除此之外，OptORF 也能预测到转录因子敲除靶点 TF1，从而解除它对基因 G1A 的抑制作用，同时使基因 G3 和基因 G5 失活。

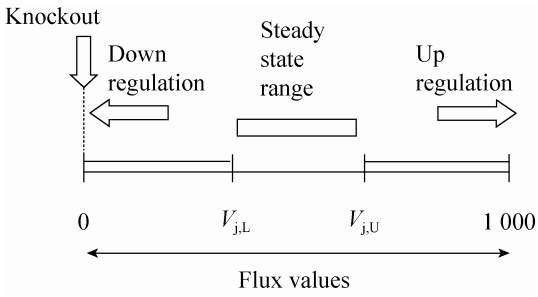

图 4 OptReg 方法中反应上调、下调及敲除的定义[23]
Fig. 4 Definitions of up/down regulation and knockout in OptReg[23].

基于大肠杆菌代谢与转录调控整合网络模型 iMC1010^{v2}[25]，OptORF 方法以乙醇为目标产品进行优化计算，预测到的改造靶点 (比如敲除基因 *pgi* 和过表达基因 *edd*) 能将生物量生长和乙醇的生成进行偶联。与 OptKnock 方法比较结果表明，基于反应敲除的策略通常要求敲除更多相关联的基因以除去被识别的反应，并且当考虑转录调控时还可能会导致致死性生长表型。

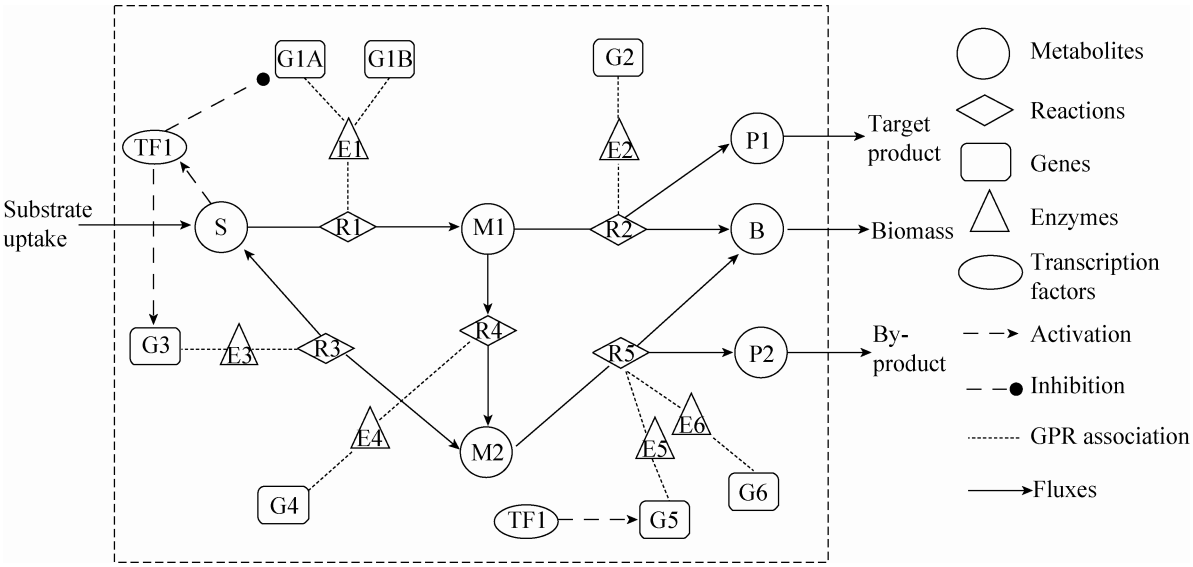

图 5　OptORF 应用于简化的代谢与转录调控整合网络

Fig. 5　Application of OptORF to an example metabolic and regulatory network.

2.2.3　OptForce

上述的双层菌种优化算法在设计时均假设基因改造后的代谢网络模型会根据最大化细胞目标的思想去分配代谢通量。而 OptForce[26]采用逆向的计算方法，假设当目标产物生产速率达到某一预设值时，突变型代谢网络中反应的通量范围与野生型进行比较，进而确定那些通量范围偏差 (上调、下调或变为 0) 较大的反应作为基因改造的靶点。与 OptReg 方法一样，OptForce 确定反应通量范围也是基于 FVA 算法，依次以代谢网络中每个反应为优化目标，计算满足目标产物生产速率和其他约束条件下每个反应通量的最大值和最小值。

2.2.4　其他预测改造靶点的方法

2004 年 Pharkya 等提出了一种整合的菌种优化方法 OptStrain[27]，该方法首先通过构建反应数据库为微生物细胞添加最优外源途径从而赋予该微生物生产某一非天然化合物的能力，然后在加入外源合成途径的代谢网络中通过 OptKnock 算法预测到使该目标化合物过量生产的敲除靶点。

除了上述由 OptKnock 衍生过来的菌种优化方法，还有一些基于 FBA 或 FVA 的代谢网络分析预测改造靶点的方法，比如通量响应分析 (Flux response analysis)[28]、FSEOF[29]、FVSEOF[30]和 RobOKoD[31]等。其中，通量响应分析、FSEOF 和 FVSEOF 方法仅用来预测过表达靶点。通量响应分析方法通过系统地考察目标产物反应通量随代谢网络模型中其他反应通量的改变的变化趋势来确定反应之间的相互关系，从而识别出能够提高目标产物产量的过表达靶点。类似地，FSEOF 通过系统地考察代谢网络模型中那些随目标产物反应通量的增加而增加的反应作为过表达靶点。另外，FVSEOF 方法利用 FVA 算法考虑了基于约束问题的多种最优解和代谢网络中各反应速率的范围，克服

了 FSEOF 方法在这方面的缺陷。同时该方法通过添加反应簇群 (GR) 约束来模拟细胞的生理状态，有效提高了预测准确性。而 RobOKoD 最近由 Stanford 等[31]提出，该方法利用 FVA 针对生物量生长和目标化合物最大化时的反应进行分级评价，找出能够使目标化合物过量生产的敲除、过表达和抑制靶点。

3 双层优化方法存在的问题

3.1 双层结构及求解问题

微生物代谢网络的行为会受到内部细胞目标的调控，而一般情况下内部细胞目标会与目标产物的过量生产相竞争。不过，基于基因组尺度代谢网络模型，双层菌种优化方法能够通过靶点的改造，比如反应或基因的敲除、上调及下调，将内部细胞目标与部分目标产物的生产偶联起来，从而实现目标产物的过量生产。这类方法主要通过双层混合整型线性或非线性规划来实现。正是由于此类双层结构的特点和基因组尺度代谢网络的复杂性及冗余性，导致双层菌种优化方法直接求解是不可行的。因此，若要求解它们都是先利用线性规划的对偶理论将内层的原始问题转化为对偶问题，然后再将双层问题转化为标准的单层混合整型线性规划问题。而这类单层问题能利用各种数学优化软件中 MILP 求解器进行求解。由于求解空间的维度极大 (例如从 150 个候选目标反应集合中搜索 10 个敲除靶点，其组合数超过 10^{15}，实际上远远不止如此)，导致求解时间随敲除数的增加而呈指数型增长。尽管 GDBB 方法能够在较短时间内给出计算结果，但是不能得到全局最优解。大多数双层菌种优化方法能预测到的改造靶点数目一般不会超过 5 个。

3.2 解的特点及问题

双层优化方法试图通过识别令目标化合物过量生产的改造靶点来实现代谢网络的重新布局，将生物量生长与目标产物相偶联，使目标化合物的生产成为生物量生长时的一种强制性副产物，即该产物的生成是平衡细胞生长的还原力和能量需求的唯一途径，通过这类方法预测到的改造靶点不会使细胞无法生长而只是平衡生长与产物生成的需求。但是并非所有的代谢物其生成都可以和细胞生长偶联，在某些情况下，虽然用双层优化方法可以计算得到敲除靶点，但对改造后的代谢网络进行模拟的结果却表明该敲除策略并不能使产物和生长相偶联。例如，我们应用 OptKnock 针对聚羟基脂肪酸脂 (PHB) 进行优化计算，得到的一组敲除靶点基因组合为乙酸激酶、琥珀酰辅酶 A 合成酶和磷酸丙糖异构酶。然而敲除这些靶点后进行计算并不能得到类似图 2 针对琥珀酸计算的结果，而是如图 6 所示。在敲除改造的菌株中生长速率最大时 PHB 生成速率为 0，并没有与生

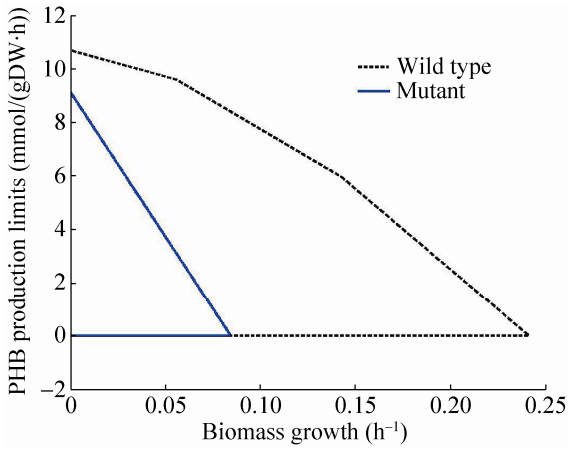

图 6 PHB 最大和最小生产速率随生长速率的变化
Fig. 6 Maximal and minimal PHB production rates at different gowth rates.

长偶联。我们还针对苏氨酸、核黄素等产品进行了计算，发现也无法得到真正将产物生成和细胞生长相偶联的敲除结果，因此这类双层优化方法仅适用于一些可以与生长相偶联的代谢物的敲除靶点预测，以同时实现产物合成和细胞生长的最大化，对于无法与生长相偶联的代谢物则需要从其它角度寻找新的敲除靶点预测方法。

4 菌种优化软件及工具包

早期的菌种优化方法，比如 OptKnock、OptStrain、OptReg 以及后来的 OptORF，都是基于通用代数建模系统 (The general algebraic modeling system，GAMS) 这一计算平台来实现运算，但都没有提供源代码。随后在 Matlab 平台的基础上，一些菌种优化方法相继实现运算。其中，基于约束的重构及分析工具，COBRA (Constraints based reconstruction and analysis)[32-33] 工具箱是目前应用最为广泛的分析软件，2007 年由 Palsson 课题组设计，不仅能实现基因组尺度代谢网络的构建，也能进一步利用代谢网络对微生物的表型进行模拟、分析及预测。COBRA 工具箱在菌种优化设计方面集成的算法有 MOMA、OptKnock、OptGene、GDLS 等。另一个基于 Matlab 用于代谢网络半自动化构建、分析和可视化的工具包是 RAVEN[34]，可以读写 SBML 和 Excel 格式的模型文件，集成的算法有 FBA、MOMA 等。此外，GDLS、GDBB、RobustKnock 和 OptSwap 等方法也是基于 Matlab 平台实现运算，并提供了工具包。2013 年，Palsson 研究组基于 Python 开发了 COBRApy[35] 工具包。该工具包支持基本的 COBRA 方法，具有面向对象的形式，很适合处理像代谢和基因表达这类复杂的生物过程，并且在数据读写以及 FBA 计算速度上更有优势。另一个基于 Python 用于代谢网络构建和分析的工具包是 Metano[36]，集成的算法有 FBA、FVA、MOMA 等，不仅可以读取 SBML 格式的文件，也可以读写 txt 文件。2013 年，Gelius-Dietrich 等基于 R 开发了 Sybil[37]工具包，集成的算法有 FBA、FVA、MOMA、ROOM 等，能够快速地进行基因组尺度计算分析，并且具有面向对象的形式，很容易被用户拓展。但是，COBRApy、Sybil、Metano 和 RAVEN 等工具包只集成有 FBA 和几种常用的代谢网络构建及分析方法，而预测靶点的方法只有 MOMA，并没有类似 OptKnock 这种双层优化方法。另外，OptFlux[38]是一种开放型资源和模块化的软件，并且是第一个整合了菌种优化任务的工具，其中集成有 OptKnock 和 OptGene 方法，具有用户友好性，可以让用户在可视化界面中执行某种操作，不需要在计算平台上输入指令，但是它并不能正确地预测到敲除靶点，因此有很大的局限性。常用的菌种优化软件及工具包如表 1 所示。虽然目前已有以上几种菌种优化的工具包及软件被开发，并为代谢工程改造微生物工厂提供指导，但是无论是计算时间及可行性，还是预测的准确性都存在很大问题，亟待进一步改进和开发新型的工具。

5 菌种优化方法总结及展望

近几年，菌种优化方法的研究发展非常迅速，不仅能基于基因组尺度代谢网络预测到令目标化合物过量生产的基因改造靶点，而且相

表 1 常用的菌种优化软件和工具包

Table 1 Software and toolboxes used for strain optimization

Software and toolboxes	Website	Platform based	Description
COBRA	http://opencobra.sourceforge.net /openCOBRA/Welcome.html	Matlab	Quantitative prediction of cellular metabolism with constraint-based models
COBRApy	http://opencobra.sourceforge.net/	Python	An object-oriented framework designed to meet the computational challenges
OptFlux	http://www.optflux.org		An open-source and modular software to support in silico metabolic engineering tasks
RobustKnock	http://www.cs.technion.ac.il/~tomersh /methods.html	Matlab	An implementation of RobustKnock
OptSwap	http://online.liebertpub.com/doi/abs /10.1089/ind.2013.0005	Matlab	An implementation of OptSwap
GDLS	http://crab.rutgers.edu/~dslun/gdls/index.html	Matlab	Using an efficient, low-complexity local search approach to identify favorable genetic designs
GDBB	http://crab.rutgers.edu/~dslun/gdbb/index.html	Matlab	Using an efficient truncated branch and bound approach to identify favorable genetic designs
Sybil	http://CRAN.R-project.org/package=sybil	R	Designed to address large scale questions in reasonable time frames based on R
Metano	http://metano.tu-bs.de/	Python	An open-source software toolbox for analyzing the capabilities of metabolic networks
RAVEN	http://www.sysbio.se/BioMet	Matlab	Allowing for semi-automated reconstruction, analysis and visualization of genome-scale models

关技术、算法、各种软件工具都有了很大的进步。然而目前仍然存在一些关键的问题暂时无法解决。

首先，双层优化方法只能预测到能够与生物量生长相偶联的目标产物的改造靶点，以实现生物量生长和产物合成这两个目标的同时优化，而大部分代谢物实际上无法与生物量生长相偶联，所以针对这些代谢物该类方法无法预测到有效的改造靶点。

其次，现有的菌种优化方法很少利用基因组尺度代谢网络与其他组学网络整合在一起的生物大网络。在细胞内存在着很多其他的重要机制，如转录调控、信号转导等，如果整合这些组学网络形成全细胞网络，将大大提高菌种优化方法对生物表型的预测能力，并成为代谢工程决策的有力武器。

再次，大多数双层菌种优化方法求解时间过长。由于基因组尺度代谢网络的复杂性及冗余性，导致菌种优化方法在计算上面临着巨大的挑战。尽管 GDBB 能够在较短的时间内给出计算结果，却不能得到全局最优解。然而，随着并行计算技术的发展，相信未来能够解决更大维度的问题，搜索到更多敲除数的改造策略。

最后，目前大多数菌种优化方法是基于反应敲除来实现代谢网络的重新布局。然而，细胞内存在复杂的基因-蛋白质-反应 (GPR) 映射关系，这样从基因改造层面有时很难实现代谢网络重新布局这一目标，甚至会导致产量降低或致死性表型。不过，OptGene、OptORF 和 OptGeneKnock 等方法能够直接做基因水平的预

测，解决了这个问题。

随着细胞整合网络、高性能计算机和并行计算技术的不断发展，菌种优化方法预测的结果将更准确，更符合生物学意义，计算速度更快，这将会使基因改造的范围变广，更趋向于理性，大大拓展能够与生物量生长相偶联的化合物范围。

REFERENCES

[1] Edwards JS, Palsson BO. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc Natl Acad Sci USA, 2000, 97(10): 5528–5533.

[2] Reed JL, Vo TD, Schilling CH, et al. An expanded genome-scale model of *Escherichia coli* K-12 (*i*JR904 GSM/GPR). Genome Biol, 2003, 4(9): R54.

[3] Feist AM, Henry CS, Reed JL, et al. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol, 2007, 3(1): 121.

[4] Zhou MD, Zou W, Liu LM, et al. Reconstruction and analysis of *Bacillus megaterium* metabolic model based on literature study. Acta Microbiol Sin, 2012, 52(4): 457–465 (in Chinese).
周冒达, 邹伟, 刘立明, 等. 基于文献挖掘的巨大芽胞杆菌代谢网络模型的构建与分析. 微生物学报, 2012, 52(4): 457–465.

[5] Chai WP, Xue W, Zhang L, et al. Research on the auto-reconstruction of genome-scale metabolic network model. J Food Sci Biotechnol, 2014, 33(9): 957–965 (in Chinese).
柴文平, 薛卫, 张梁, 等. 基因组规模代谢网络模型的自动化重构. 食品与生物技术学报, 2014, 33(9): 957–965.

[6] Kauffman KJ, Prakash P, Edwards JS. Advances in flux balance analysis. Curr Opin Biotechnol, 2003, 14(5): 491–496.

[7] Price ND, Papin JA, Schilling CH, et al. Genome-scale microbial in silico models: the constraints-based approach. Trends Biotechnol, 2003, 21(4): 162–169.

[8] Segrè D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci USA, 2002, 99(23): 15112–15117.

[9] 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.

[10] Patil KR, Rocha I, Förster J, et al. Evolutionary programming as a platform for *in silico* metabolic engineering. BMC Bioinformatics, 2005, 6: 308.

[11] Feist AM, Zielinski DC, Orth JD, et al. Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*. Metab Eng, 2010, 12(3): 173–186.

[12] Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng, 2003, 5(4): 264–276.

[13] Fong SS, Burgard AP, Herring CD, et al. *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. Biotechnol Bioeng, 2005, 91(5): 643–648.

[14] Orth JD, Conrad TM, Na J, et al. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. Mol Syst Biol, 2011, 7(1): 535.

[15] Tepper N, Shlomi T. Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. Bioinformatics, 2010, 26(4): 536–543.

[16] King ZA, Feist AM. Optimizing cofactor specificity of oxidoreductase enzymes for the generation of microbial production strains—OptSwap. Ind Biotechnol, 2013, 9(4): 236–246.

[17] Egen D, Lun DS. Truncated branch and bound achieves efficient constraint-based genetic design. Bioinformatics, 2012, 28(12): 1619–1623.

[18] Lun DS, Rockwell G, Guido NJ, et al. Large-scale identification of genetic design strategies using local search. Mol Syst Biol, 2009, 5(1): 296.
[19] Ren SG, Zeng B, Qian XN. Adaptive bi-level programming for optimal gene knockouts for targeted overproduction under phenotypic constraints. BMC Bioinformatics, 2013, 14(Suppl 2): S17.
[20] Xu ZX, Zheng P, Sun JB, et al. ReacKnock: identifying reaction deletion strategies for microbial strain optimization based on genome-scale metabolic network. PLoS ONE, 2013, 8(12): e72150.
[21] Zhang C, Ji BY, Mardinoglu A, et al. Logical transformation of genome-scale metabolic models for gene level applications and analysis. Bioinformatics, 2015, 31(14): 2324–2331.
[22] Koffas MAG, Jung GY, Stephanopoulos G. Engineering metabolism and product formation in *Corynebacterium glutamicum* by coordinated gene overexpression. Metab Eng, 2003, 5(1): 32–41.
[23] Pharkya P, Maranas CD. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metab Eng, 2006, 8(1): 1–13.
[24] Kim J, Reed JL. OptORF: optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. BMC Syst Biol, 2010, 4: 53.
[25] Covert MW, Knight EM, Reed JL, et al. Integrating high-throughput and computational data elucidates bacterial networks. Nature, 2004, 429(6987): 92–96.
[26] Ranganathan S, Suthers PF, Maranas CD. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comput Biol, 2010, 6(4): e1000744.
[27] Pharkya P, Burgard AP, Maranas CD. OptStrain: a computational framework for redesign of microbial production systems. Genome Res, 2004, 14(11): 2367–2376.
[28] Song CW, Kim DI, Choi S, et al. Metabolic engineering of *Escherichia coli* for the production of fumaric acid. Biotechnol Bioeng, 2013, 110(7): 2025–2034.
[29] Choi HS, Lee SY, Kim TY, et al. *In silico* identification of gene amplification targets for improvement of lycopene production. Appl Environ Microbiol, 2010, 76(10): 3097–3105.
[30] Park JM, Park HM, Kim WJ, et al. Flux variability scanning based on enforced objective flux for identifying gene amplification targets. BMC Syst Biol, 2012, 6(1): 106.
[31] Stanford NJ, Millard P, Swainston N. RobOKoD: microbial strain design for (over) production of target compounds. Front Cell Dev Biol, 2015, 3: 17.
[32] Becker SA, Feist AM, Mo ML, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox. Nat Protoc, 2007, 2(3): 727–738.
[33] Schellenberger J, Que R, Fleming RMT, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox v2.0. Nat Protoc, 2011, 6(9): 1290–1307.
[34] Agren R, Liu LM, Shoaie S, et al. The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. PLoS Comput Biol, 2013, 9(3): e1002980.
[35] Ebrahim A, Lerman JA, Palsson BO, et al. COBRApy: COnstraints-based reconstruction and analysis for python. BMC Syst Biol, 2013, 7(1): 74.
[36] Ulas T, Riemer SA, Zaparty M, et al. Genome-scale reconstruction and analysis of the metabolic network in the hyperthermophilic archaeon *Sulfolobus solfataricus*. PLoS ONE, 2012, 7(8): e43401.
[37] Gelius-Dietrich G, Desouki AA, Fritzemeier CJ, et al. Sybil-efficient constraint-based modelling in R. BMC Syst Biol, 2013, 7(1): 125.
[38] Rocha I, Maia P, Evangelista P, et al. OptFlux: an open-source software platform for *in silico* metabolic engineering. BMC Syst Biol, 2010, 4: 45.

(本文责编　郝丽芳)