

多约束代谢网络模型的研究进展

杨雪^{1,2,3#}, 张培基^{1,2#}, 毛志涛², 赵欣^{2,3}, 王若宇², 蔡敬一², 王智文¹, 马红武²

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

3 中国科学院大学, 北京 100049

杨雪, 张培基, 毛志涛, 赵欣, 王若宇, 蔡敬一, 王智文, 马红武. 多约束代谢网络模型的研究进展. 生物工程学报, 2022, 38(2): 531-545.

YANG X, ZHANG PJ, MAO ZT, ZHAO X, WANG RY, CAI JY, WANG ZW, MA HW. Development of metabolic models with multiple constraints: a review. Chin J Biotech, 2022, 38(2): 531-545.

摘 要: 基于约束的基因组尺度代谢网络模型 (genome-scale metabolic models, GEMs) 分析已被广泛应用于代谢表型的预测。而实际细胞中代谢速率除计量学约束外, 还受到酶资源可用性和反应热力学可行性等其他因素影响, 在 GEMs 中整合酶资源约束或者热力学约束构建多约束代谢网络模型可以进一步缩小优化解空间, 提升细胞表型预测的准确性。文中主要对近年来多约束模型的研究进展进行了综述, 介绍了不同多约束模型的构建方法, 以及其在研究基因敲除影响、预测可行途径和提供代谢瓶颈信息等方面的应用。将多种约束条件进一步与代谢模型整合, 可更准确地预测细胞代谢的瓶颈和关键调控改造靶点, 从而为工业菌种代谢工程改造提供精确的设计指导。

关键词: 基因组尺度代谢网络模型; 酶约束; 热力学约束; 多约束模型; 关键酶; 瓶颈反应

Received: May 7, 2021; **Accepted:** July 26, 2021; **Published online:** July 29, 2021

Supported by: National Key Research and Development Program of China (2018YFA0900300, 2018YFA0901400); Tianjin Synthetic Biotechnology Innovation Capacity Improvement Project (TSBICIP-PTJS-001, TSBICIP-KJGG-005)

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

[#]These authors equally contributed to this work

基金项目: 国家重点研发计划 (2018YFA0900300, 2018YFA0901400); 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)

Development of metabolic models with multiple constraints: a review

YANG Xue^{1,2,3#}, ZHANG Peiji^{1,2#}, MAO Zhitao², ZHAO Xin^{2,3}, WANG Ruoyu², CAI Jingyi², WANG Zhiwen¹, MA Hongwu²

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

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

Abstract: Constraint-based genome-scale metabolic network models (genome-scale metabolic models, GEMs) have been widely used to predict metabolic phenotypes. In addition to stoichiometric constraints, other constraints such as enzyme availability and thermodynamic feasibility may also limit the cellular phenotype solution space. Recently, extended GEM models considering either enzymatic or thermodynamic constraints have been developed to improve model prediction accuracy. This review summarizes the recent progresses on metabolic models with multiple constraints (MCGEMs). We presented the construction methods and various applications of MCGEMs including the simulation of gene knockout, prediction of biologically feasible pathways and identification of bottleneck steps. By integrating multiple constraints in a consistent modeling framework, MCGEMs can predict the metabolic bottlenecks and key controlling and modification targets for pathway optimization more precisely, and thus may provide more reliable design results to guide metabolic engineering of industrially important microorganisms.

Keywords: genome-scale metabolic network models; enzymatic constraints; thermodynamic constraints; metabolic models with multiple constraints (MCGEMs); key enzyme; bottleneck reaction

1991年, Bailey^[1]在 *Science* 上发表题为“Toward a Science of Metabolic Engineering”的综述文章, 将代谢工程的研究范畴概括为采用基因工程技术重构代谢网络以提升细胞的产能。1999年, Varma和Palsson^[2]基于稳态假设(steady state assumption, SSA)和通量平衡分析原理(flux balance analysis, FBA), 运用嗜流感菌 *Haemophilus influenzae* Rd的基因组注释信息构建了第一个代谢网络模型^[3], 这标志着人类在电子计算机上(*in silico*)运用基因组尺度代谢网络模型这一优化约束模型来描述和模拟生命代谢过程的全新领域已然形成。此后,

代谢网络模型的数量呈指数增长趋势, 指导意义充分被实践证明^[4-5]。

基于约束原理的代谢网络模型作为代谢系统的数字化工具^[6-7], 因具备强大的整合组学信息和兼容算法的能力, 从出现至今, 其功能和内涵也在不断扩大。随着更多组学数据的测定和蛋白酶分析技术领域提取和测定方法的发展, 蛋白质含量^[8]、代谢物浓度^[9-10]、酶转换数 k_{cat} ^[11]以及反应的标准吉布斯自由能 $\Delta_r G_i^\circ$ 等参数陆续达到组学规模^[12]。尤其对于大肠杆菌^[13]和酿酒酵母^[14]等模式菌株, 模型、参数以及表型结果都已得到很好的累积。基于更好的数据

利用能力和更现实的表型预测结果,通过整合多组学信息来实现约束层次升级的多约束代谢网络模型开始引发关注。

本文主要对近年来发表的在计量学模型基础上整合其他约束的多约束代谢网络模型进行综述,对各类多约束模型的构建原理和方法,以及其在基因敲除结果模拟、可行途径预测和代谢瓶颈确定等方面的应用效果进行了展示。通过探讨多约束模型在关键酶和热力学瓶颈反应等靶点预测工作中的表现,说明其在精准指导代谢工程的应用实践中具有极大潜能。此外,本文对未来多约束模型的发展方向和前景进行了展望。

1 酶资源约束模型

在基于计量学的代谢网络通量平衡分析中,绝大部分反应速率都只受一个宽泛的上下限值约束,因此当底物输入速率设定值较高时会得到很大的、失真的内部反应通量值。实际细胞中每个反应的速率都受催化该反应的酶的含量和比酶活的限制,而单位质量、单位体积或单个细胞中酶的总量也需要限定在一定范围之内(常通过蛋白组学、细胞组成数据等实验数据确定)。将这些系统水平的酶资源约束以特定数学形式引入计量学模型中即可得到酶约束模型。

1.1 整合资源约束的原理进阶

2007年, Beg 等^[15]通过假定固有的细胞体积会成为酶资源的空间限制,提出了基于大分子体积约束思想的拥挤 FBA 建模方法 (flux balance analysis with molecular crowding, FBAwMC)^[16],这意味着整合系统水平的细胞资源约束开始被关注。2011年, Zhuang 等^[17]从发酵和呼吸过程对细胞膜资源的竞争角度进行约束建模,并尝试用其开发的基于膜资源经济性原则的代谢模型 (FBA with membrane economics, FBA^{ME}) 解读溢

流代谢现象。2012年, Adadi 等^[18]通过在代谢网络模型基础上整合酶动力学参数,开发了受蛋白成本限制的酶约束代谢模型 (metabolic modeling with enzyme kinetics, MOMENT),并用其研究了底物利用问题。同年, Lerman 等^[19]发布了拓扑结构更为精细的整合酶动力学参数的代谢和表达模型 (metabolism and expression, ME),用以研究细胞代谢和基因表达水平变化相关的生物学问题。在2019年赵欣等^[20]对酶约束模型的综述中,代表酶约束模型开端的 MOMENT 建模原理可被概括为:1) 将反应的酶成本简化为通量 v_i , 酶的分子量 MW_i 以及酶转换数的倒数 $1/k_{cat}$ 的乘积;2) 存在酶资源总量的上限约束,原理表达式详见该综述中的方程 (5)–(7)。

2015年, Basan 等^[21]通过构建粗粒化的蛋白资源分配模型 (coarse-grained model of proteome allocation) 对细胞的溢流代谢进行了重新阐述,此后蛋白资源限制是导致溢流代谢发生真正原因的观点获得广泛认可,在代谢网络模型内部寻找碳源、氮源和氧气^[22]等限制因素的传统模型分析方法被挑战。2017年, Sánchez 等^[23]发布了基于酿酒酵母代谢网络模型构建酶约束模型的自动化流程 (genome-scale model to account for enzyme constraints using kinetics and omics, GECKO)。该研究团队采用 GECKO 方法进行了酿酒酵母生长过程中的溢流代谢分析,对比了发酵和呼吸过程中能量合成的酶成本的差异^[24],以及通过构建酶约束模型 ecYeast8 提升新版本酵母代谢网络模型 Yeast8 的表型预测能力^[25]。至此,基于酶约束模型,已可进行溢流代谢 (overflow metabolism)^[26]、底物层级利用^[27-28]和能量效率^[24]等多种生物学现象的描述和预测。

1.2 酶约束模型构建方法的发展

GECKO 模型发布的同时,研究人员还将

构建过程和源代码也一同公开,这使得酶约束模型在其他物种中的迁移变得容易。2020年, Ye 等^[29]和 Chen 等^[30]同期发表的大肠杆菌 ec_iML1515 和凝结芽孢杆菌 ec-iBag597 两个酶约束模型,即均为采用自动化建模工具 GECKO toolbox 完成。两个团队的工作肯定了 GECKO 方法贡献的同时,还进一步说明了酶约束模型在途径酶资源分布和能量合成效率分析等生物学问题研究中的适用性。

2020年, Bekiaris 和 Klamt^[31]采用大肠杆菌 iJO1366 模型为底板,对 2012年 Adadi 等基于大肠杆菌 iAF1260 构建的 MOMENT 酶约束模型进行了重构。这项工作沿用了 MOMENT 和 GECKO 方法的构建原理,吸收了 GECKO 在参数获取环节的自动化方法和筛选规则。作者从模型拓扑结构上对 GECKO 进行了简化,并在细节的处理上进行了改动,如对同工酶催化的反应进行了去冗余,使用有效酶成本概念合并了总酶池和平均饱和度的乘积项等。此外,作者还开发了利用蛋白组或者转录组数据对酶约束模型进行动力学参数校正的自动化方法 (automatic integration of protein allocation constraints in metabolic networks, AutoPACMEN toolbox),并将其重构酶约束模型的方法命名为精简的 MOMENT (short MOMENT, sMOMENT)。值得推崇的是,这是继 GECKO 发布后,人们可以直接使用的又一个自动化酶约束模型构建工具。使用该方法构建的酶约束模型,因仅需添加一个总酶量相关的合并反应 R_{pool} , 反应数量较 GECKO 方法大幅减少,其模型拓展原理如图 1 所示。

1.3 酶约束模型的原理局限性

从模型结构和实现原理来看,这类模型还可以整合更多的数据^[11],如: 1) 通过整合热力学约束,对反应和途径的可行性进一步限制;

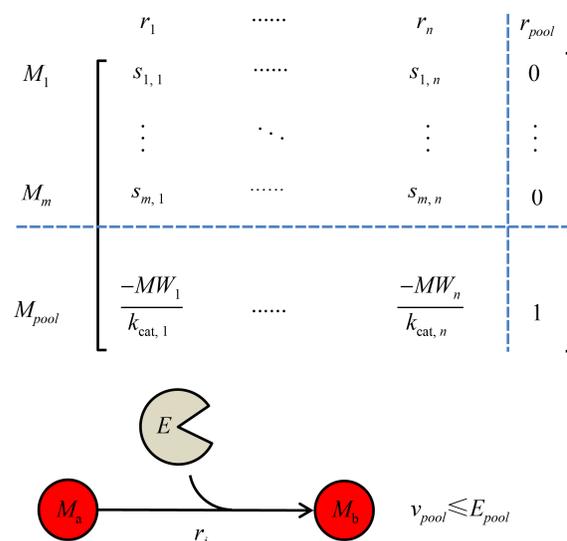


图 1 sMOMENT 方法中计量矩阵的扩充原理^[31]
Figure 1 Augmentation of the stoichiometric matrix with the sMOMENT approach. M_{pool} is the enzyme pool pseudo-metabolite and R_{pool} the enzyme-pool-delivering pseudo-reaction. r_i stands for reaction i , M_j for metabolite j ; r is the number of reactions, m is the number of metabolites. The principles refer to the work performed by Bekiaris and Klamt^[31].

2) 通过整合代谢物组学信息对酶的饱和度和真实催化效率进行校正; 3) 通过整合基因调控关系,如代谢物对酶的激活和抑制关系,加强模型的关联能力。通过上述约束的集成可以整合更多已知数据和信息,尽可能缩小模型预测的解空间,提升模型的预测能力,从而为代谢工程领域提供更具实践价值的指导方案。

2 基于热力学原理的约束模型

在计量学代谢网络模型分析中,研究人员需要基于化学反应的自由能变化确定其可逆性进而对通量施加上限或下限约束。而由于热力学相关参数,如 $\Delta_r G_i^\circ$ 、代谢物浓度和酶的平衡常数 K_{eq} 等信息的缺失情况普遍存在,人们难免会对反应的可逆性做出误判,或因难以判断而

不得不采取某种存在风险的默认做法,如统一设置为可逆反应。这可能导致由模型预测得到的途径不具有生物学可行性,如出现能量和还原力的无限循环问题^[32]。此外,单个反应的热力学可行性和通过网络求解得到的包含多个反应的代谢途径的热力学可行性也是两个不同问题。因此,通过评估单个反应的热力学可行性来预先设置反应可逆性的方法,本身即存在不合理性。反应在途径中的可行性,以及途径在网络中的可行性,均应在网络水平的模型中进行考察。如何将热力学约束和基因组规模代谢网络结合提高途径预测的可靠性是人们一直致力于解决的问题。

2.1 整合热力学信息的代谢网络分析

早在 2007 年, Henry 等^[33]就提出了 TMFA (thermodynamics-based metabolic flux analysis) 方法,采用混合整数线性规划 (MILP) 实现了热力学约束在代谢网络模型 FBA 计算过程中的整合,通过剔除热力学不可行的解空间,提升了代谢通量分布的预测准确性^[34]。TMFA 是首个将热力学原理以约束形式整合于代谢网络的工作,尽管其仅探讨了解空间和通量可变性等现象,但其约束实现方式和系统化的约束整合思想对后续研究,尤其是代谢网络模型的构建过程产生了深远影响^[35-36]。2019 年,相隔 12 年后,同课题组的 Salvy 等将 TMFA 方法拓展为基于 Python 和 MATLAB 两种编程语言的自动化分析工具,分别命名为 pyTFA 和 matTFA^[37],进一步为人们应用 TMFA 方法提供便利。

2.2 关注热力学可行性的途径优化方法

2014 年, Noor 等^[38]提出了针对代谢途径热力学驱动力优化的方法 MDF (max-min driving force)。对于一条确定的途径, MDF 方法中酶成本的公式被简化为通量 v 和可逆性因子 η^{rev} 的比值,在此条件下,可推演出当所有酶的成

本均相同时,途径的效率最高。因此,该方法将途径中热力学驱动力最低的反应的驱动力最大化作为优化目标,对参数层的依赖降低。具体来说, MDF 方法的核心原则遵循木桶原理,在满足代谢物浓度约束的解空间内,寻找可使途径中热力学短板反应的驱动力最高的浓度分布,由此预测途径的热力学瓶颈,并指导该环节的优化策略制定,原理如图 2 所示。在 MDF 算法提出后,其研究团队还将该功能整合到了自行基于基团贡献和化合物贡献理论^[39]开发的用于计算反应热力学参数的 eQuilibrator^[12]网站工具中,使人们可以非常方便地利用该工具对途径的热力学驱动力进行优化计算。MDF 方法将途径中的反应进行关联后,为反应的热力学可行性判断提供了更合理的约束,由此实现了途径中瓶颈反应^[40] (指途径中的热力学可行性最差的反应,在 MDF 原理内核中以其可行性

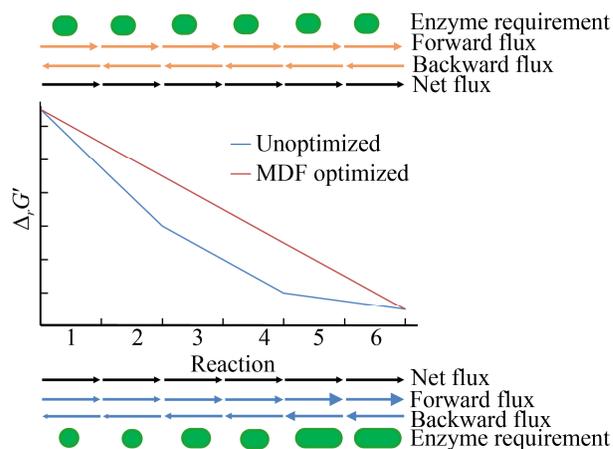


图 2 两种途径的驱动力和酶成本对比示意图^[38]
Figure 2 Schematic comparison between two pathways. Each pathway starts and ends with the same compounds, employs six enzymes and carries the same net flux. The kinetic parameters of all enzymes in both pathways, as well as enzyme and metabolite concentrations, are assumed to be identical. The principles refer to the work performed by Noor et al.^[38].

提升作为优化目标,是途径可行性的短板因素)以及限制性代谢物(指对于途径的热力学水平发挥限制性作用的代谢物,其在途径达到热力学最优水平时,浓度不再具备可变性。在实际应用中,可根据其影响的方向,选取促进积累或加快消耗的方式^[41]提升途径的热力学可行性的预测能力。该方法对于代谢工程实践的指导能力,已被大量途径筛选和优化工作所证明^[41-43]。

2.3 网络规模的热力学优化模型

不同于 TMFA 将热力学作为约束条件,上文提及的 MDF 方法是将热力学驱动力最优化作为目标函数。由此,前者需要代谢物组学作为已知的参数层输入信息,而 MDF 则将可产生热力学途径最优效果的代谢物浓度分布作为预测结果输出,使用户摆脱了目前普遍存在的代谢物组学信息,特别是代谢物绝对浓度信息难以准确获取的困境,因此一经提出,即得到了迅速推广。如 2018 年,Asplund-Samuelsson 等^[44]提出的途径批量化设计流程(prospecting optimal pathways with python, POPPY)和 Hadicke 等^[45]发表的网络热力学驱动力最优途径计算方法(identification of MDF optimal pathways in genome scale networks, OptMDFpathway),均为基于 MDF 方法发展而来的工作流程或分析工具。二者分别基于 Python 和 MATLAB 平台,均代码开源。特别指出的是,包括 POPPY 在内的一些工作流程需要预先通过基元模式分析^[42,44,46]等手段来确定途径结构和通量分布,再与后续的途径热力学驱动力评估环节进行衔接。相比之下,OptMDFpathway 算法的与众不同之处在于,它以 MDF 算法作为内核,将热力学优化目标关联到代谢网络的 FBA 求解过程。

3 整合多约束的代谢网络模型

目前基于酶约束的代谢网络模型主要整合

了 k_{cat} 和蛋白组学信息^[23],而尚未完成对代谢组学、热力学水平和调控信息等关键数据的整合。尽管相比于计量学模型,这类模型已可更准确地描述多种生物学现象,但代谢改造靶点级预测精度的成功案例^[47]仍非常有限,因此,通过整合更多约束向多约束模型发展^[48],可增强其对实践的指导能力。

3.1 热力学因素与酶成本的关联

上述 MOMENT、GECKO 和 sMOMENT 等酶约束模型的构建原理均基于简化的酶成本表达方式,具体来说,仅用酶转换数 k_{cat} 这一理想参数表征酶的比活性,缺少对于底物和产物的浓度对酶活性影响的考量,因此完全忽略了饱和度、可逆性和调控关系等受代谢物浓度影响的因素。实际上,考虑代谢物浓度影响的酶成本表达式会更加复杂。代谢物浓度和酶成本之间可以通过可逆的米氏方程(Reversible Michaelis-Menten)中的平衡常数 K_{eq} 及亲和常数 K_{m} 进行关联,也可通过热力学第二定律进行连接,因此,热力学可行性可被整合在酶成本的计算过程中。以最简单的仅包含单底物和单产物的反应机制为例,连接方式见方程(1)–(9)^[49]。

可逆的米氏方程:

$$v(s, p, E) = E \cdot \frac{k_{\text{cat}}^+ \cdot \frac{s}{K_s} - k_{\text{cat}}^- \cdot \frac{p}{K_p}}{1 + \frac{s}{K_s} + \frac{p}{K_p}} \quad (1)$$

稳态条件下,存在:

$$k_{\text{cat}}^+ \cdot \frac{s'}{K_s} - k_{\text{cat}}^- \cdot \frac{p'}{K_p} = 0 \quad (2)$$

因此,根据定义, K_{eq} 可表示为:

$$K_{\text{eq}} = \frac{p'}{s'} = \frac{k_{\text{cat}}^+ \cdot K_p}{k_{\text{cat}}^- \cdot K_s} \quad (3)$$

此时,方程(1)可被表述为:

$$v(s, p, E) = E \cdot k_{\text{cat}}^+ \cdot \frac{\frac{s}{K_s}}{1 + \frac{s}{K_s} + \frac{p}{K_p}} \cdot \left(1 - \frac{1}{K_{\text{eq}}} \cdot \frac{p}{s} \right) \quad (4)$$

反应的吉布斯自由能定义:

$$\Delta_r G' = \Delta_r G'^{\circ} + RT \cdot \ln \left(\frac{p}{s} \right) \quad (5)$$

其中, 标准吉布斯自由能的表达式为:

$$\Delta_r G'^{\circ} = -RT \cdot \ln(K_{\text{eq}}) \quad (6)$$

因此, 平衡常数 K_{eq} 可表示为:

$$K_{\text{eq}} = e^{-\frac{\Delta_r G'^{\circ}}{RT}} \quad (7)$$

因此, 方程 (4) 可表述为:

$$v(s, p, E) = E \cdot k_{\text{cat}}^+ \cdot \frac{\frac{s}{K_s}}{1 + \frac{s}{K_s} + \frac{p}{K_p}} \cdot \left(1 - e^{-\frac{\Delta_r G'}{RT}} \right) \quad (8)$$

最终, 可得到考虑热力学因素的酶成本公式:

$$E(s, p, v) = v \cdot \frac{1}{k_{\text{cat}}^+} \cdot \frac{1 + \frac{s}{K_s} + \frac{p}{K_p}}{\frac{s}{K_s}} \cdot \frac{1}{1 - e^{-\frac{\Delta_r G'}{RT}}} \quad (9)$$

此处, v 表示反应通量 (mol/L·s), 是底物浓度 s 、产物浓度 p 和酶量 E 的函数; K_s 和 K_p 分别表示正反应方向和逆反应方向酶对底物的亲和力常数 (mol/L); s' 和 p' 分别表示反应达到平衡时, 底物和产物的浓度 (mol/L); k_{cat}^+ 和 k_{cat}^- 分别表示酶在正、逆两个方向的转换数 (/s); 酶量 E 的单位为 mol/L; 反应标准吉布斯自由能 $\Delta_r G'^{\circ}$ 和反应吉布斯自由能 $\Delta_r G'$ 的单位均为 kJ/mol。

当考虑调控因素、协同效应和多底物等问题时, 酶反应的机制方程的复杂程度^[50]会增加, 但热力学与酶成本在原理层上的关联性不会消失。因此, 在约束模型的发展过程中, 已有一

些方法尝试将热力学整合在酶成本的表达式中进行最优化求解。这些方法普遍从热力学驱动力差的反应将耗费更多酶成本这一认识出发, 或单独将酶成本, 或将其与代谢物的总负载水平最小化作为目标函数, 用以研究酶成本、代谢物浓度水平和通量之间的关系, 用以优化途径或者阐释次优表型的存在意义。如在 2013–2015 年间陆续被提出的代谢拔河现象分析方法 (metabolic tug-of-war, mTOW)^[51]、途径酶成本最小化方法 (enzyme cost minimization, ECM)^[38] 和低酶成本次优途径求解方法 (cost reduced sub-optimal FBA, corsoFBA)^[52] 等。这些方法在原理公式中实现了代谢物浓度和热力学因素与酶成本的关联, 通过探讨次优代谢过程来阐释生物学目标多样性, 对基因组尺度约束模型的原理发展和算法开发具有很好的借鉴价值。

以 ECM 方法为例, 将其酶成本的影响因素划分为六类因素的乘积, 如方程 (10) 所示。其中, 酶成本 q 是代谢物浓度 C (x 为 C 的自然对数, 即 $x = \ln C$) 和反应通量 v 的函数, 其与 v 成正比, 与酶的负担因子 h_E (包含酶的组成、结构维持成本、表达和修饰成本、使用寿命和酶活特异性等多种因素的影响) 成正比, 而与酶的转换数 k_{cat} 、可逆性因子 η^{rev} 、饱和度因子 η^{sat} 以及调控因子 η^{reg} 等 4 个因式成反比。

$$q(x, v) = \sum_l h_{E_l} E_l(x, v_l) = \sum_l h_{E_l} \cdot v_l \cdot \frac{1}{k_{\text{cat}, l}^+} \cdot \frac{1}{\eta_l^{\text{rev}}(x)} \cdot \frac{1}{\eta_l^{\text{sat}}(x)} \cdot \frac{1}{\eta_l^{\text{reg}}(x)} \quad (10)$$

可以看出, ECM 方法在原理层面充分考虑了酶活性的影响因素。但是, 由于真实的情况往往过于复杂, 公式中的很多参数无法确定, 实际使用中必须要进行适当简化。比如, 对于负担因子 h_E , 实际使用过程中, ECM 仅将其以蛋白分

子量 MW 表示。同理,上述 3 个因子也往往需要根据可用参数情况进行简化处理。因此, Noor 等在发布 ECM 方法时,根据参数需求的复杂度不同,将酶成本公式划分为 5 个层次(如表 1),提示用户可以根据参数获取能力选择适合研究需求的原理复杂度水平。此外,这种粗粒化、模块式的原理组装形式,也便于用户自行重组设计后使用,即使用方法包括但并不局限于表 1 中提及的 5 种形式,如上文提及的 MDF 方法可看作 EMC2 忽略 h_E 和 k_{cat} 后的简化形式。

3.2 热力学和酶约束的整合模型

整合热力学因素可以将代谢物浓度及其对可逆性的影响体现出来,因此可以在很大程度上弥补当前酶约束模型普遍存在的过度简化造成的约束松弛、预测失真和研究角度单一等问题。2020 年,借助 ME 模型的自动化建模工具 COBRAME^[53], Salvy 和 Hatzimanikatis 将 TMFA 方法应用于大肠杆菌的 ME 模型中,开发了整合热力学约束的 ME 模型的建模方法 (expression and thermodynamics flux models, ETFL),并用于研究资源限制下的生长表型和考察新的约束层次带来的解空间压缩效果^[54]。2021 年,本课题组发表了可将热力学约束与酶资源约束进行整合的自动化建模框架 (enzymatic and

thermodynamic GEMs, ETGEMs)^[55],并在框架中实现了多种途径靶点的分析算法。尽管 ME 和 MOMENT 模型均整合了蛋白组学数据和酶的动力学参数信息,构建原理相似,但由于底层的计量学框架差异很大,二者的应用角度相去甚远。具体来说,ME 模型更侧重于研究表型背后的生命机制和大分子运转成本^[56],而 MOMENT 则更聚焦于提供指导代谢过程相关的通路和靶点优化方案。因此,ETFL 和 ETGEMs 方法的研究目标和应用场景差异也由此产生,即 ETGEMs 更侧重于对代谢工程实践中的途径设计和靶点预测等工作的指导能力。

3.3 多约束模型的应用研究

整合蛋白资源约束的酶约束模型,考虑热力学可行性约束的热力学模型,以及将酶资源可用性和热力学可行性进行集成的多约束模型,均可通过更加严格的可行性判断标准,实现解空间压缩和通量再分配,并且拓展最优方案的观察角度,由此带来传统 GEMs 不具备的预测效果和应用可能性。

3.3.1 研究基因敲除的影响

2019 年,Massaiu 等^[47]采用 GECKO 原理,仅为枯草芽孢杆菌 *Bacillus subtilis* 的 *iYO844* 模型中的 17 个中心代谢反应添加了酶资源约

表 1 ECM 方法中复杂度可选的简化酶成本公式

Table 1 Simplified enzyme cost functions in ECM method

Levels	Expression formula	Annotation
EMC0	$q(v) = \sum_l v_l$	Only consider the flux of reactions
EMC1	$q(v) = \sum_l h_{E_l} \cdot v_l \cdot \frac{1}{k_{cat,l}^+}$	Consider the catalytic ability of enzymes
EMC2	$q(x,v) = \sum_l h_{E_l} \cdot v_l \cdot \frac{1}{k_{cat,l}^+} \cdot \frac{1}{\eta_l^{rev}(x)}$	Consider the reversibility of reactions
EMC3	$q(x,v) = \sum_l h_{E_l} \cdot v_l \cdot \frac{1}{k_{cat,l}^+} \cdot \frac{1}{\eta_l^{rev}(x)} \cdot \frac{1}{\eta_l^{sat}(x)}$	Consider the substrate saturation of enzymes
EMC4	$q(x,v) = \sum_l h_{E_l} \cdot v_l \cdot \frac{1}{k_{cat,l}^+} \cdot \frac{1}{\eta_l^{rev}(x)} \cdot \frac{1}{\eta_l^{sat}(x)}$	More detailed enzymatic kinetic mechanism

束, 即提升了 *ec_iYO844* 模型在多聚谷氨酸 (poly- γ -glutamate) 合成过程中的基因敲除预测准确性。其实现原理在于, *ec_iYO844* 可对效率差的 α -酮戊二酸脱羧酶 (α -oxoglutarate decarboxylase) 进行辨别, 因此无需占用基因敲除的反应位, 而让真正需要被敲除的反应得以暴露出来。而初始的 *iYO844* 模型则不具备对低酶活性反应的识别能力, 必须经过敲除才能使其参与的旁路失效, 从而因错误占用了基因敲除名额, 造成预测脱靶。尽管确切来说, 该工作并未实现真正的基因组尺度枯草芽孢杆菌酶约束模型的构建, 但从原理的差异上为酶约

束的添加可促进代谢网络模型敲除靶点预测能力提供了实验支持。

3.3.2 预测生物学可行途径

以基于 ETGEMs 方法构建的大肠杆菌的整合酶和热力学约束模型 EcoETM 为例, 由于约束层次的添加, 带来了途径预测结果更现实的可能。如图 3 所示, 使用 EcoETM 模型预测的精氨酸合成途径 (图 3B), 在通量水平和途径结构上与传统的代谢网络模型 *iML1515* 的预测结果 (图 3A) 存在明显差异。具体来说, 当以葡萄糖摄入速率 5 mmol/(g DW·h) 为基准时, 得率从 *iML1515* 模型预测的 4.28 mmol/(g DW·h)

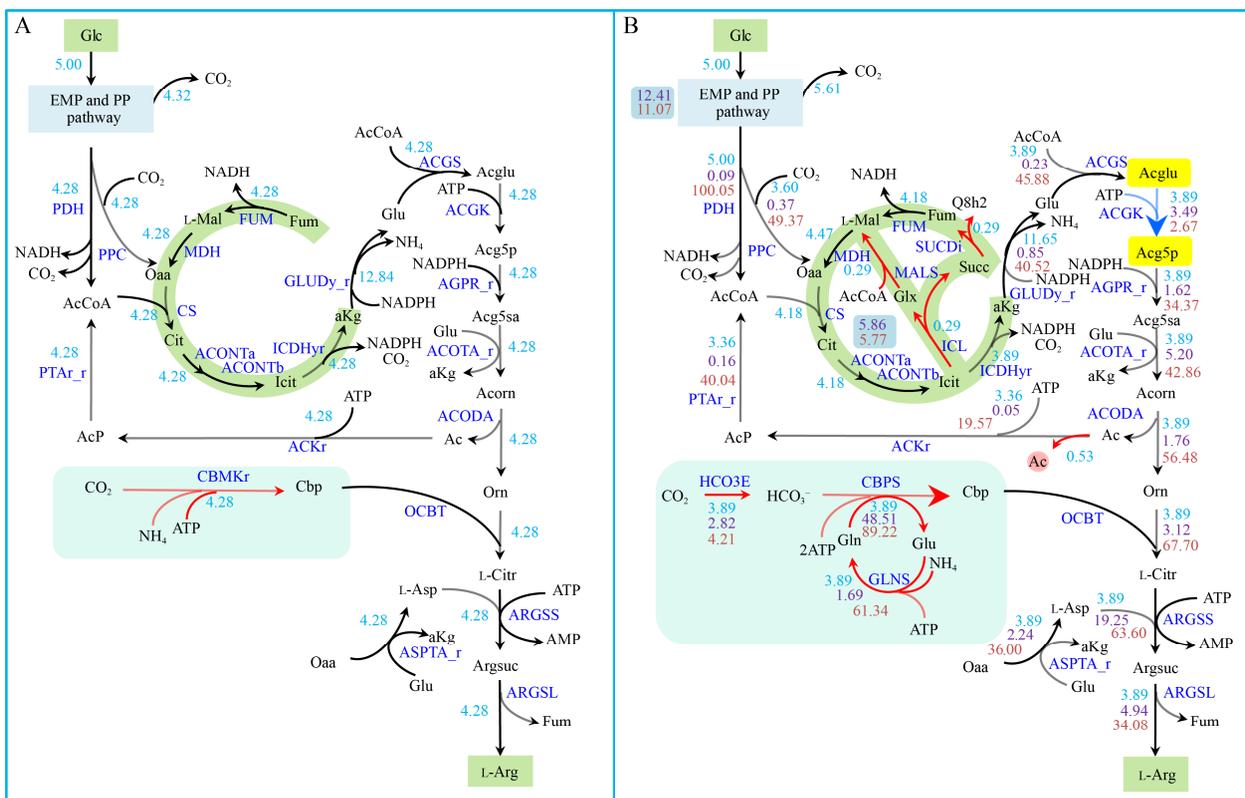


图 3 模型预测的 L-精氨酸合成途径^[55]

Figure 3 Model predicted pathways for L-Arg synthesis^[55]. Predicted by *iML1515* (A) and EcoETM (B). Shown are: the structural change of the pathway (red arrow); the reaction with the highest enzyme cost (red thick arrow); the thermodynamic bottleneck reaction (blue arrow); the limiting metabolite (yellow background); and the simplified pathways of EMP and PP (navy-blue background). The unit of the flux value is mmol/(g DW·h) (blue, on the top); the unit of the enzyme cost is mg/g DW (purple, in the middle); and the unit of the maximum thermodynamic driving force is in kJ/mol (orange, at the bottom).

降低至 3.89 mmol/(g DW·h), 途径中氨甲酰磷酸 (carbamoyl phosphate, Cbp) 的合成过程发生切换, 不具备热力学可行性的由氨甲酰磷酸激酶催化的高得率反应 (carbamate kinase reaction, CBMKr) 得到有效排除, 并以更具现实性的由氨甲酰磷酸合酶 (carbamoyl-phosphate synthetase, CBPS) 催化的反应替代。同时, 合成途径中发生了乙酸的溢流现象。上述预测结果, 均可通过文献予以证实^[57-59]。

3.3.3 提供代谢瓶颈信息

从图 3 中的途径信息可以发现, 约束的添加除造成了途径切换, 即途径结构和得率水平的改变以外, 还增加了结果的输出层次。相对于传统的代谢网络模型仅能给出途径的通量分布信息, 多约束模型由于采用了更多的输入数据并整合了多种求解算法, 可为用户提供更多角度的途径信息。具体来说, 除通量分布信息外, 还提供了途径的酶成本分布, 以及反应的最佳热力学驱动力水平分布信息。此外, 还可通过代谢物浓度可变性确定途径中的限制性代谢物, 以及根据热力学驱动力和酶成本分布信息确定途径中的瓶颈反应和关键酶^[60] (途径中占用酶成本居多的反应, 其酶活性的提升对途径的通量影响非常明显, 通常为过表达和酶的定向进化等策略的实施对象) 等信息。同样, 途径中的关键酶 CBPS^[61-62]和 ARGSS、热力学瓶颈步骤 AGCK 和限制性代谢物乙酰谷氨酸^[63-64]等信息均可通过文献证实。这些信息可为人们应用酶的定向进化^[65-66]、基因编辑^[67]和解调控^[68-69]等手段优化途径提供了依据, 并最终服务于代谢产品合成效率的提升。

4 总结与展望

文中提及的酶约束模型, 热力学约束模型, 以及多约束模型, 均对传统 GEMs 进行了组学

数据的扩充。具体来说, 当前的酶约束模型完成了网络规模的 k_{cat} 参数和蛋白组学数据与计量学框架的整合, 热力学约束模型考虑了代谢组学和网络规模的 $\Delta_r G^\circ$ 参数数据, 而整合热力学和酶约束的多约束模型则集成了上述组学或网络规模的数据。

鉴于多约束模型尚处于发展初期阶段, 存在很大的研究空间和应用潜能, 本节特对在约束模型的原理发展和构建方法方面具有代表性的工作进行了梳理, 详见图 4, 以便于读者理解各类约束模型 (含多约束模型) 的起源、发展和关联。此外, 本节也将从多约束模型的原理完善、数据获取和校正, 以及指导代谢工程的应用实践 3 个方面进行了展望, 以期促进多约束模型的开发、完善和应用。

4.1 多约束模型的方法改进

为了提升模型构建与求解的可操作性, 目前的多约束模型构建框架对于原理公式的处理方式往往存在过度简化问题。未来, 需要不断引入更多的数据和约束层次, 如借鉴 ECM 等方法中酶成本的影响因素, 通过增加可逆性、饱和度和调控等因素与酶成本之间的关联, 来修正原理中过度简化的不合理假设。与此同时, 提升对于实验来源的表型, 如通量组、代谢组和蛋白组等组学信息的利用度, 也可进一步压缩非现实的解空间, 提升预测结果的准确性。此外, 调控关系通过影响蛋白表达量和酶的活性水平对代谢能力间接施加影响, 将调控信息以适当的方式整合于代谢网络模型的求解过程, 也可作为多约束模型的发展目标之一。近年来, ECM^[70]方法中的一些粗粒化的概念, 如饱和度因子和可逆性因子分别以平均值的简化形式被运用于基因组尺度的酶约束建模 (平均饱和度系数^[71], σ) 和组学规模参数的机器学习预测过程 (平均可逆性因子^[23], η_{rev}), 相信随着

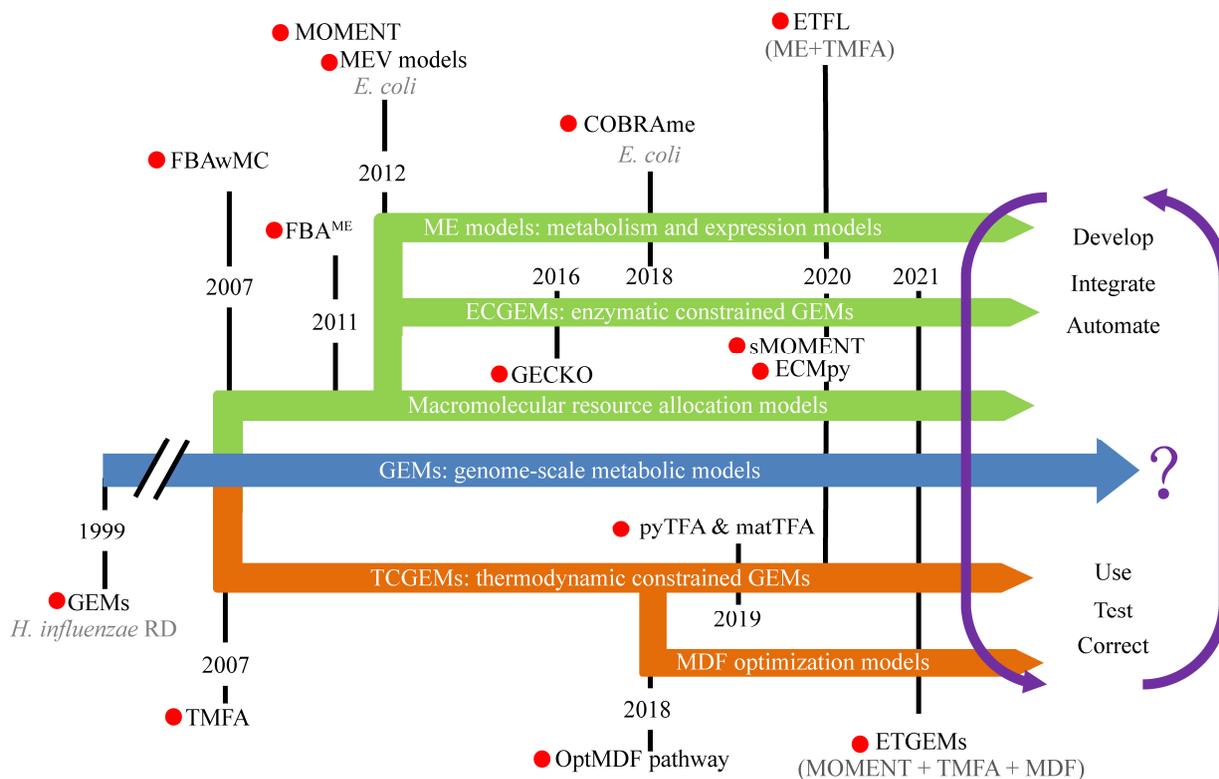


图 4 约束模型的主要发展沿革

Figure 4 The main development of constrained models. GEMs^[3]; FBAwMC^[16]; TMFA^[33]; FBA^{ME}^[17]; MOMENT^[18]; ME models^[19]; GECKO^[23]; COBRAME^[53]; OptMDF pathway^[45]; pyTFA and matTFA^[37]; sMOMENT^[31]; ECMpy^[28]; ETFL^[54]; ETGEMs^[55].

更多数据的积累和数据质量的提升,更加精细的约束模型原理完善工作将得到持续推动。与此同时,值得注意的是,不同的约束层次之间的关联一方面会提升原理的自治程度,同时关联方式的合理性,如各因素影响程度的量化和表达式的复杂程度等,也会对预测准确性产生较大的影响。例如,各因素影响程度的量化和表达式的复杂程度等如处理不当则不可避免地造成误差以非线性的方式在不同约束层次间发生传递。这也是当前一些具备很好的原理思想却仍未发展到网络水平的方法,如 mTOW、ECM 和 corsoFBA 等,所面对的共同难题。

4.2 参数的获取与校正

多约束模型可以整合多种生物学相关约

束,但同时需要大量的参数以精确地定量化表征这些约束,可靠的参数值对模型预测的准确性至关重要。因此,如何获得系统规模的准确参数值是多约束模型构建的主要瓶颈,目前的多约束模型主要是针对大肠杆菌、酿酒酵母等数据积累相对良好的模式生物。但即使针对这些模式生物也存在不同程度的数据缺失问题,在应用 ETGEMs 框架构建大肠杆菌模型 EcoETM 时,对于胞内反应,可直接获取的 $\Delta_r G_i^\circ$ 参数的覆盖度仅为 55%,缺失问题仍较为明显。这是由于 $\Delta_r G_i^\circ$ 参数的来源远不像 k_{cat} 参数那样有众多实验来源的选择,目前主要为基于 eQuilibrator 数据库的预测参数。对于酶参数的提取,便捷的自动化工具的出现已形成非常好

的发展局面, 如 GECKO^[23]和 AutoPACMEN^[31]可基于人为制定的规则及优先级从数据库中快速筛选参数, 但这些方法仍无法解决数据缺失问题。近年来, 全局特征^[72]和机器学习^[71,73]方法则基于酶的属性及测定条件对参数进行训练, 用以获取表型自洽的高覆盖度^[74]参数集合。此外, 覆盖度的满足固然重要, 参数的质量则更为直接地决定模型预测结果的准确性。需要指出的是, 即便 GECKO、AutoPACMEN 和机器学习等方法已可服务于参数获取、校正和填补环节, 但由于参数对菌种和条件存在特异性依赖^[75], 可用参数和表型信息的规模仍十分有限, 并且, 关键性参数的缺失或偏差仍会对多约束模型的预测准确性造成很大影响。因此, 实验来源的比酶活、蛋白组和代谢组等数据和通量分布等表型信息作为模型构建的源头素材和准确性校正依据^[11], 对多约束模型的发展仍具有不可替代的价值。

4.3 指导代谢工程

近年来, 计算工具已被用于新途径的系统设计^[41,43,76]。批量设计的途径, 往往需要通过酶参数水平和热力学可行性等生物学规则进行评估、筛选和优化^[77]。因此, 将这些规则以约束的形式直接整合于代谢网络的 FBA 求解过程, 可以更加高效地设计符合目标要求的途径。以多约束模型 ETGEMs 的精氨酸合成过程预测为例, 除更合理的通量分布外, 酶成本分布和热力学驱动力分布信息也可一并呈现在预测结果中。因此, 基于更好的反应重排方式, 多约束模型有望被用于为特定分子选择最适配的宿主, 或者通过引入异源反应和异源酶来促进途径的酶成本优化和热力学驱动力的提升, 从而精准指导代谢工程实践。此外, ETGEMs 或可和现有的约束模型兼容的算法相结合, 如 ptStoic^[78]和 comb-FBA^[41], 来实现高可行性途

径的批量设计, 用于满足高通量实践能力提升的指导需求。值得指出的是, 目前针对多约束模型量身打造的算法设计工作仍是空白。从原理上, 多约束模型的数据层次更丰富, 可为算法设计提供更多着力角度。因此, 随着多约束模型的进一步普及和发展, 相信会有更多的算法应势而生, 更好地服务于途径设计和靶点预测等代谢工程实践的指导过程。

REFERENCES

- [1] Bailey JE. Toward a science of metabolic engineering. *Science*, 1991, 252(5013): 1668-1675.
- [2] Varma A, Palsson BO. Metabolic flux balancing: basic concepts, scientific and practical use. *Bio/Technology*, 1994, 12(10): 994-998.
- [3] Edwards JS, Palsson BO. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *J Biol Chem*, 1999, 274(25): 17410-17416.
- [4] Kim TY, Sohn SB, Kim YB, et al. Recent advances in reconstruction and applications of genome-scale metabolic models. *Curr Opin Biotechnol*, 2012, 23(4): 617-623.
- [5] 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).
- [6] Reed JL, Palsson BØ. Thirteen years of building constraint-based *in silico* models of *Escherichia coli*. *J Bacteriol*, 2003, 185(9): 2692-2699.
- [7] Ebrahim A, Lerman JA, Palsson BO, et al. COBRApy: Constraints-based reconstruction and analysis for *Python*. *BMC Syst Biol*, 2013, 7: 74.
- [8] Wang M, Weiss M, Simonovic M, et al. Pa_xDb, a database of protein abundance averages across all three domains of life. *Mol Cell Proteomics*, 2012, 11(8): 492-500.
- [9] Bennett BD, Kimball EH, Gao M, et al. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat Chem Biol*, 2009, 5(8): 593-599.
- [10] Kiparissides A, Hatzimanikatis V. Thermodynamics-

- based metabolite sensitivity analysis in metabolic networks. *Metab Eng*, 2017, 39: 117-127.
- [11] Nilsson A, Nielsen J, Palsson BO. Metabolic models of protein allocation call for the kinetome. *Cell Syst*, 2017, 5(6): 538-541.
- [12] Flamholz A, Noor E, Bar-Even A, et al. eQuilibrator—the biochemical thermodynamics calculator. *Nucleic Acids Res*, 2012, 40(database issue): D770-D775.
- [13] Monk JM, Lloyd CJ, Brunk E, et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat Biotechnol*, 2017, 35(10): 904-908.
- [14] Fernandez-Ricaud L, Warringer J, Ericson E, et al. PROPHECY—a yeast phenome database, update 2006. *Nucleic Acids Res*, 2007, 35(database issue): D463-D467.
- [15] Yang L, Yurkovich JT, King ZA, et al. Modeling the multi-scale mechanisms of macromolecular resource allocation. *Curr Opin Microbiol*, 2018, 45: 8-15.
- [16] Beg QK, Vazquez A, Ernst J, et al. Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. *PNAS*, 2007, 104(31): 12663-12668.
- [17] Zhuang K, Vemuri GN, Mahadevan R. Economics of membrane occupancy and respiro-fermentation. *Mol Syst Biol*, 2011, 7: 500.
- [18] Adadi R, Volkmer B, Milo R, et al. Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters. *PLoS Comput Biol*, 2012, 8(7): e1002575.
- [19] Lerman JA, Hyduke DR, Latif H, et al. *In silico* method for modelling metabolism and gene product expression at genome scale. *Nat Commun*, 2012, 3: 929.
- [20] 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).
- [21] Basan M, Hui S, Okano H, et al. Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nature*, 2015, 528(7580): 99-104.
- [22] Varma A, Boesch BW, Palsson BO. Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. *Appl Environ Microbiol*, 1993, 59(8): 2465-2473.
- [23] Sánchez BJ, Zhang C, Nilsson A, et al. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol Syst Biol*, 2017, 13(8): 935.
- [24] Chen Y, Nielsen J. Energy metabolism controls phenotypes by protein efficiency and allocation. *Proc Natl Acad Sci USA*, 2019, 116(35): 17592-17597.
- [25] Lu HZ, Li FR, Sánchez BJ, et al. Author correction: a consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat Commun*, 2020, 11: 5443.
- [26] Molenaar D, van Berlo R, de Ridder D, et al. Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol Syst Biol*, 2009, 5: 323.
- [27] Hermsen R, Okano H, You C, et al. A growth-rate composition formula for the growth of *E. coli* on co-utilized carbon substrates. *Mol Syst Biol*, 2015, 11(4): 801.
- [28] 赵欣. 基于酶约束的 *E. coli* 代谢网络模型构建与分析[D]. 天津: 中国科学院天津工业生物技术研究所, 2020.
Zhao X. Construction and analysis of *E. coli* metabolic network model based on enzyme constraints[D]. Tianjin: Tianjin institute of industrial biotechnology, Chinese academy of sciences, 2020 (in Chinese).
- [29] Ye C, Luo QL, Guo L, et al. Improving lysine production through construction of an *Escherichia coli* enzyme-constrained model. *Biotechnol Bioeng*, 2020, 117(11): 3533-3544.
- [30] Chen Y, Sun Y, Liu Z, et al. Genome-scale modeling for *Bacillus coagulans* to understand the metabolic characteristics. *Biotechnol Bioeng*, 2020, 117(11): 3545-3558.
- [31] Bekiaris PS, Klamt S. Automatic construction of metabolic models with enzyme constraints. *BMC Bioinformatics*, 2020, 21(1): 19.
- [32] Yuan Q, Huang T, Li P, 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.
- [33] Henry CS, Broadbelt LJ, Hatzimanikatis V. Thermodynamics-based metabolic flux analysis. *Biophys J*, 2007, 92(5): 1792-1805.
- [34] Soh KC, Hatzimanikatis V. Network thermodynamics in the post-genomic era. *Curr Opin Microbiol*, 2010, 13(3): 350-357.
- [35] Seep L, Razaghi-Moghadam Z, Nikoloski Z. Reaction lumping in metabolic networks for application with

- thermodynamic metabolic flux analysis. *Sci Rep*, 2021, 11: 8544.
- [36] Hamilton JJ, Dwivedi V, Reed JL. Quantitative assessment of thermodynamic constraints on the solution space of genome-scale metabolic models. *Biophys J*, 2013, 105(2): 512-522.
- [37] Salvy P, Fengos G, Ataman M, et al. pyTFA and matTFA: a *Python* package and a Matlab toolbox for thermodynamics-based flux analysis. *Bioinformatics*, 2019, 35(1): 167-169.
- [38] Noor E, Bar-Even A, Flamholz A, et al. Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLoS Comput Biol*, 2014, 10(2): e1003483.
- [39] Goldberg RN, Tewari YB, Bhat TN. Thermodynamics of enzyme-catalyzed reactions—a database for quantitative biochemistry. *Bioinformatics*, 2004, 20(16): 2874-2877.
- [40] Mavrouniotis ML. Identification of localized and distributed bottlenecks in metabolic pathways. *Proc Int Conf Intell Syst Mol Biol*, 1993, 1: 275-283.
- [41] Yang X, Yuan Q, Luo H, et al. Systematic design and *in vitro* validation of novel one-carbon assimilation pathways. *Metab Eng*, 2019, 56: 142-153.
- [42] Dash S, Olson DG, Joshua Chan SH, et al. Thermodynamic analysis of the pathway for ethanol production from cellobiose in *Clostridium thermocellum*. *Metab Eng*, 2019, 55: 161-169.
- [43] Trudeau DL, Edlich-Muth C, Zarzycki J, et al. Design and *in vitro* realization of carbon-conserving photorespiration. *PNAS*, 2018, 115(49): E11455-E11464.
- [44] Asplund-Samuelsson J, Janasch M, Hudson EP. Thermodynamic analysis of computed pathways integrated into the metabolic networks of *E. coli* and *Synechocystis* reveals contrasting expansion potential. *Metab Eng*, 2018, 45: 223-236.
- [45] Hädicke O, von Kamp A, Aydogan T, et al. OptMDFpathway: identification of metabolic pathways with maximal thermodynamic driving force and its application for analyzing the endogenous CO₂ fixation potential of *Escherichia coli*. *PLoS Comput Biol*, 2018, 14(9): e1006492.
- [46] Ullah E, Yosafshahi M, Hassoun S. Towards scaling elementary flux mode computation. *Brief Bioinform*, 2020, 21(6):1875-1885.
- [47] Massaiu I, Pasotti L, Sonnenschein N, et al. Integration of enzymatic data in *Bacillus subtilis* genome-scale metabolic model improves phenotype predictions and enables *in silico* design of poly- γ -glutamic acid production strains. *Microb Cell Fact*, 2019, 18(1): 3.
- [48] Soh KC, Miskovic L, Hatzimanikatis V. From network models to network responses: integration of thermodynamic and kinetic properties of yeast genome-scale metabolic networks. *FEMS Yeast Res*, 2012, 12(2): 129-143.
- [49] Noor E, Flamholz A, Liebermeister W, et al. A note on the kinetics of enzyme action: a decomposition that highlights thermodynamic effects. *FEBS Lett*, 2013, 587(17): 2772-2777.
- [50] Chassagnole C, Noisommit-Rizzi N, Schmid JW, et al. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnol Bioeng*, 2002, 79(1): 53-73.
- [51] Tepper N, Noor E, Amador-Noguez D, et al. Steady-state metabolite concentrations reflect a balance between maximizing enzyme efficiency and minimizing total metabolite load. *PLoS One*, 2013, 8(9): e75370.
- [52] Schultz A, Qutub AA. Predicting internal cell fluxes at sub-optimal growth. *BMC Syst Biol*, 2015, 9: 18.
- [53] Lloyd CJ, Ebrahim A, Yang L, et al. COBRAME: a computational framework for genome-scale models of metabolism and gene expression. *PLoS Comput Biol*, 2018, 14(7): e1006302.
- [54] Salvy P, Hatzimanikatis V. The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models. *Nat Commun*, 2020, 11(1): 30.
- [55] Yang X, Mao ZT, Zhao X, et al. Integrating thermodynamic and enzymatic constraints into genome-scale metabolic models. *Metab Eng*, 2021, 67: 133-144.
- [56] O'Brien EJ, Lerman JA, Chang RL, et al. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol Syst Biol*, 2013, 9: 693.
- [57] Valgepea K, Adamberg K, Seiman A, et al. *Escherichia coli* achieves faster growth by increasing catalytic and translation rates of proteins. *Mol Biosyst*, 2013, 9(9): 2344-2358.
- [58] Issaly IM, Issaly AS, Reissig JL. Carbamyl phosphate biosynthesis in *Bacillus subtilis*. *Biochim Biophys Acta*, 1970, 198(3): 482-494.
- [59] Ginesy M, Belotserkovsky J, Enman J, et al. Metabolic engineering of *Escherichia coli* for enhanced arginine biosynthesis. *Microb Cell Fact*, 2015, 14: 29.

- [60] 杨雪, 张彦飞, 郑阳阳, 等. 大肠杆菌苏氨酸合成途径动力学模型的构建与分析. 生物工程学报, 2014, 30(1): 18-29.
Yang X, Zhang YF, Zheng YY, et al. Development and analysis of a kinetic model for *Escherichia coli* threonine biosynthesis. Chin J Biotech, 2014, 30(1): 18-29 (in Chinese).
- [61] Charlier D, Bervoets I. Regulation of arginine biosynthesis, catabolism and transport in *Escherichia coli*. Amino Acids, 2019, 51(8): 1103-1127.
- [62] Shen S, Zhang X, Li Z. Development of an engineered carbamoyl phosphate synthetase with released sensitivity to feedback inhibition by site-directed mutation and casting error-prone PCR. Enzyme Microb Technol, 2019, 129: 109354.
- [63] Ramón-Maiques S, Marina A, Gil-Ortiz F, et al. Structure of acetylglutamate kinase, a key enzyme for arginine biosynthesis and a prototype for the amino acid kinase enzyme family, during catalysis. Structure, 2002, 10(3): 329-342.
- [64] Utagawa T. Production of arginine by fermentation. J Nutr, 2004, 134(10 suppl): 2854S-2857S; discussion2895S.
- [65] Ding D, Li J, Bai D, et al. Biosensor-based monitoring of the central metabolic pathway metabolites. Biosens Bioelectron, 2020, 167: 112456.
- [66] Lu X, Liu Y, Yang Y, et al. Constructing a synthetic pathway for acetyl-coenzyme A from one-carbon through enzyme design. Nat Commun, 2019, 10(1): 1378.
- [67] Wang Y, Cheng H, Liu Y, et al. *In-situ* generation of large numbers of genetic combinations for metabolic reprogramming via CRISPR-guided base editing. Nat Commun, 2021, 12(1): 678.
- [68] Schramm T, Lempp M, Beuter D, et al. High-throughput enrichment of temperature-sensitive argininosuccinate synthetase for two-stage citrulline production in *E. coli*. Metab Eng, 2020, 60: 14-24.
- [69] Guo J, Man Z, Rao Z, et al. Improvement of the ammonia assimilation for enhancing L-arginine production of *Corynebacterium crenatum*. J Ind Microbiol Biotechnol, 2017, 44(3): 443-451.
- [70] Noor E, Flamholz A, Bar-Even A, et al. The protein cost of metabolic fluxes: prediction from enzymatic rate laws and cost minimization. PLoS Comput Biol, 2016, 12(11): e1005167.
- [71] Heckmann D, Lloyd CJ, Mih N, et al. Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models. Nat Commun, 2018, 9(1): 5252.
- [72] Davidi D, Noor E, Liebermeister W, et al. Global characterization of *in vivo* enzyme catalytic rates and their correspondence to *in vitro* k_{cat} measurements. PNAS, 2016, 113(12): 3401-3406.
- [73] Heckmann D, Campeau A, Lloyd CJ, et al. Kinetic profiling of metabolic specialists demonstrates stability and consistency of *in vivo* enzyme turnover numbers. PNAS, 2020, 117(37): 23182-23190.
- [74] He H, Höper R, Dodenhöft M, et al. An optimized methanol assimilation pathway relying on promiscuous formaldehyde-condensing aldolases in *E. coli*. Metab Eng, 2020, 60: 1-13.
- [75] Wortel MT, Noor E, Ferris M, et al. Metabolic enzyme cost explains variable trade-offs between microbial growth rate and yield. PLoS Comput Biol, 2018, 14(2): e1006010.
- [76] Moura M, Broadbelt L, Tyo K. Computational tools for guided discovery and engineering of metabolic pathways. Methods Mol Biol, 2013, 985: 123-147.
- [77] Bar-Even A, Noor E, Flamholz A, et al. Design and analysis of metabolic pathways supporting formatotrophic growth for electricity-dependent cultivation of microbes. Biochim Biophys Acta, 2013, 1827(8/9): 1039-1047.
- [78] Wang L, Ng CY, Dash S, et al. Exploring the combinatorial space of complete pathways to chemicals. Biochem Soc Trans, 2018, 46(3): 513-522.

(本文责编 陈宏宇)