

文章编号: 1004-0374(2011)09-0853-07

# 全基因组范围代谢网络的构建和最小基因组研究

房柯池, 王 晶\*

(中国科学院心理研究所心理健康院重点实验室, 北京 100101)

**摘要:** 全基因组范围代谢网络 (genome-scale metabolic network, GSMN) 的构建是合成生物学研究的一个重要研究手段。通过整合各种组学数据和借助计算机进行模拟分析, 将基因型与表型的关系进行定量关联, 从而为从全局的角度探索和揭示生物代谢机制, 进而对生物进行合理的重新设计和工程改造提供了有效的框架。该方法在最小基因组研究中也具有突出的优势, 通过计算机辅助的基因组最小化模拟与分析, 能够系统鉴定微生物基因组基因的必需性。迄今为止, 已有近百个基因组范围的代谢网络发表, 覆盖的生物包括原核生物、真核生物和古生生物, 并广泛应用于医药、能源、环境、工业和农业等多个领域, 展现出了广阔的应用前景。将对全基因组范围代谢网络构建的方法、应用, 特别是其在最小基因组研究中的应用作简要的综述。

**关键词:** 全基因组范围代谢网络; 最小基因组; 合成生物学

**中图分类号:** Q621.33; Q812      **文献标志码:** A

## Genome-scale metabolic network reconstruction and minimal genome research

FANG Ke-Chi, WANG Jing\*

(Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China)

**Abstract:** Genome-scale metabolic network reconstruction has become indispensable framework for systematically exploring and understanding biological system by integrating various -omics data and taking advantages of computational simulation and analysis. The complex metabolic properties of living organisms can be accurately mimicked, predicted, controlled, and even re-designed by genome-scale metabolic network. This method has an outstanding merit in the minimal genome research, which is one of the major subjects of synthetic biology. So far, nearly 100 genome-scale metabolic networks have been published covering prokaryotes, archaea and eukaryotes. Genome-scale metabolic reconstruction has a broad spectrum of basic and practical applications, including biological discovery, metabolic engineering, phenotypic behavior, and bacteria evolution, and has been successfully applied in medicine, energy, environmental, industrial, agricultural, and other fields. In this paper, the methods of genome-scale metabolic network reconstruction, its application, as well as minimal genome research will be reviewed briefly.

**Key words:** genome-scale metabolic network; minimal genome; synthetic biology

### 1 全基因组范围的代谢网络 (genome-scale metabolic network, GSMN)

随着测序技术和高通量生物技术的高速发展, 越来越多的生物体完成了全基因组测序, 基因功能被注释, 庞大的组学 (-omics) 数据源源不断的产生, 以及丰富的生物数据库资源的开发, 其中也蕴含了海量的生物信息, 亟待系统地、合理有效地整合这

些数据, 并进行模拟分析和数据解释。GSMN 的构建, 为从全局的角度探索和揭示生物代谢机制提供了一个有效的框架, 很好地将基因型与表型的关系

收稿日期: 2011-05-07; 修回日期: 2011-08-05

基金项目: 国家重点基础研究发展计划 (“973” 项目) (2011CBA00802)

\*通信作者: E-mail: wangjing@psych.ac.cn

定量地关联起来。GSMN 在基因组序列的基础上, 结合基因功能注释信息, 把基因编码蛋白 (enzyme) 所催化的生化反应 (biochemical reaction) 构建为一个代谢网络, 反应了基因 - 蛋白质 - 生化反应 (gene-protein-reaction, GPR) 之间的相互作用, 从而有效地转化为数学模型在计算机上 (*in silico*) 进行模拟、分析, 并用实验数据加以验证, 提出假设。GSMN 充分利用了各种组学数据, 从系统的角度, 由生物体内各种分子的鉴别以及相互作用的研究到代谢通路、网络、模块, 最终完成整个生命活动的路线图。GSMN 构建的目标是, 得到一个尽可能接近真实生物系统的理论模型, 能够对生物体以及生物现象做出精确的预测、控制, 甚至重新设计, 它是合成生物学研究中的一个重要的研究手段。GSMN 方法在诸多生物学基本问题的研究中发挥着重要作用, 如探索生命起点, 即维持细胞生命所需要包含的一组必需基因的最小基因组, 研究物种进化等; GSMN 在生产生活方面也有广泛的实际应用, 已经成功地在医药、能源、环境、工业和农业等多个领域展现出巨大的应用前景和发展潜力。

### 1.1 全基因组范围代谢网络的构建方法

全基因组代谢网络的构建方法已经相对成熟, 广泛地对近百种生物构建了 GSMNs, 表 1 代表性列举了在真菌、古生菌和真核生物中已发表的模式生物的 GSMNs。关于构建方法, 国外已经有了数篇综述<sup>[15-17]</sup>, 本文作者将结合自己的研究经历 (构

建致病菌 *Burkholderia cenocepacia* 全基因组代谢网络<sup>[18]</sup>), 概括地阐述 GSMN 的构建方法。

针对不同的研究, 网络有不同的定义, 在全基因组代谢网络构建研究中, 关于代谢网络的描述是: 每一个代谢物分子 (metabolite) 是一个节点 (node); 连接节点之间的边 (edge) 是生化反应, 边的方向与生化反应进行的方向一致; 每个反应由特定的酶催化 (protein->reaction), 而每个特定的酶 / 蛋白质又由相应的基因编码 (gene->protein)。因此, 酶连接了基因、生化反应之间的相互作用。

GSMN 的构建是一个循环反复的过程, 包括以下几个阶段: 初始建模 (initial modeling); 模型完善 (model gap-filling); 基于模型的模拟分析和初步预测 (model-based analysis and preliminary prediction); 模型验证 (model validation); 进一步完善模型 (remodeling)。后四个阶段循环反复, 直到得到一个准确、可靠的模型。首先, 初始建模从特定生物体的基因组序列出发, 结合基因功能注释和各种生物数据库资源 (例如 GenBank<sup>[19]</sup>、KEGG<sup>[20]</sup>、UniProt<sup>[21]</sup>、BRENDA<sup>[22]</sup>、TCDB<sup>[23]</sup>、TransportDB<sup>[24]</sup>、BiGG<sup>[25]</sup>), 确定该生物体全部基因 - 蛋白质 - 生化反应 (GPR) 之间的相互作用。通常初始模型是不完整的, 存在一些网络缺口 (gap), 导致构建的代谢网络模型不能够行使正常的功能, 比如, 不能合成细胞生长所需要的各个组成部分 (DNA、RNA、氨基酸、脂类、cofactor, 等), 需要进一步挖掘生物数据库和文献

表1 真菌、古生菌和真核生物中已发表的模式生物的GSMN

Organism	Model	No. of genes/reactions/metabolites	Reference
<b>Bacteria</b>			
<i>Escherichia coli</i>	iJR904	904/931/625	Reed <i>et al.</i> , 2003 [1]
<i>Staphylococcus aureus</i>	iSB619	619/724/571	Becker and Palsson, 2005 [2]
<i>Bacillus subtilis</i>	iYO844	844/1245/988	Oh <i>et al.</i> , 2007 [3]
<i>Mycobacterium tuberculosis</i>	GSMN-TB	726/849/739	Beste <i>et al.</i> , 2007 [4]
<i>Pseudomonas putida</i>	iJN746	746/950/911	Nogales <i>et al.</i> , 2008 [5]
<i>Salmonella typhimurium</i>	iRR1083	1083/1087/744	Raghunathan <i>et al.</i> , 2009 [6]
<i>Clostridium thermocellum</i>	iSR432	432/577/525	Roberts <i>et al.</i> , 2010 [7]
<b>Archaea</b>			
<i>Methanosarcina barkeri</i>	iAF692	692/619/558	Feist <i>et al.</i> , 2006 [8]
<i>Halobacterium salinarum</i>	None	490/711/557	Gonzalez <i>et al.</i> , 2008 [9]
<b>Eukaryotes</b>			
<i>Saccharomyces cerevisiae</i>	iFF708	708/1175/584	Förster <i>et al.</i> , 2003 [10]
<i>Homo sapiens</i>	Recon 1	1496/3311/2712	Duarte <i>et al.</i> , 2007 [11]
<i>Aspergillus nidulans</i>	iHD666	666/1213/732	David <i>et al.</i> , 2008 [12]
<i>Arabidopsis thaliana</i>	AraGEM	1419/1567/1748	de Oliveira Dal'Molin <i>et al.</i> , 2010 [13]
<i>Mus musculus</i>	iMM1415	1415/3726/2774	Sigurdsson <i>et al.</i> , 2010 [14]

信息, 来填补网络缺口, 构建出完整的代谢网络 (具体方法在其他研究中已有一些介绍<sup>[26-27]</sup>)。其次, 将 GSMN 转化为数学模型用计算机进行模拟分析, 并产生初步预测, 比如, 模拟环境变化对生物体的代谢特性的改变, 改变培养基中碳、氮、磷、硫等基本元素的组成, 分析在特定环境下的生长能力; 模拟遗传扰动, *in silico* 单基因或多基因敲除, 预测细胞生长所需要的必需基因和非必需基因。然后, 利用实验数据验证模型预测的结果, 常用的实验数据有: BIOLOG 表型数据<sup>[28]</sup>、<sup>13</sup>C 流分析、蛋白质组学数据、高通量基因敲除等实验数据。根据模型预测与实验结果不一致的情况, 进一步改进模型, 再次实验验证……直到得到一个预测值可靠、准确的模型。

## 1.2 全基因组代谢网络的分析方法

构建的全基因组代谢网络 (即完整的 GPR 集合), 需要将其转换成数学模型, 以便在计算机上模拟并量化分析。现代生物模型的建立需要满足新的标准: 模型针对特定的生物、数据驱动、容易测量、能够整合各种组学数据、能够解释生物学功能的不确定性。目前, 用于建立细胞过程的模型方法有动力学模型<sup>[29-30]</sup>、控制论<sup>[31-32]</sup>、随机方法<sup>[33-34]</sup>。虽然上述方法能够提供有用的结果, 但是由于需要大量的参数和计算复杂性, 还很难用于全基因组水平的网络建模。然而, 基于约束的分析方法, 不寻求找到问题的单一解, 而是找到一个可行解的集合, 不需要动力学参数, 是到目前为止唯一一个可行的能够用于建立全基因组水平代谢模型的方法。基于约束的分析方法有: 流平衡分析方法 (flux balance analysis, FBA)<sup>[35]</sup>、表型相平面分析 (phenotype phase plane analysis, PhPP)<sup>[36]</sup>、混合整数线性规划 (mixed integer linear programming, MILP)<sup>[37]</sup>、通量变异性分析 (flux variability analysis, FVA)<sup>[38]</sup>、基元模式 (elementary modes)<sup>[39]</sup>、极端通路 (extreme pathways)<sup>[40]</sup> 等。其中, FBA 方法应用得最为广泛<sup>[41]</sup>, 而其他方法也多为在 FBA 方法基础上的演变。FBA 应用线性最优化算法 (linear programming), 假定细胞内的物质在单位时间内处于拟稳态, 即代谢物既没有累积, 也没有缺失。通过将代谢网络转化为化学计量矩阵, 加入热力学参数、代谢能力约束、细胞生长环境约束等, 设定目标函数 (如细胞最优生长速率), 利用最优化原理, 预测细胞内所有生化反应的流值, 最终得到最佳的稳态代谢网络的代谢流分布, 使得目标函数取得最优值。

FBA 分析的数学描述如下:

$$\begin{aligned} & \max : c^T \cdot v \\ & s.t. \begin{cases} \sum_{j=1}^C S_{ij} \cdot v_j = 0 & \forall i \in \text{compounds} \\ v_{\min} \leq v_j \leq v_{\max} & \forall j \in \text{reactions} \end{cases} \end{aligned}$$

其中,  $S$  是化学计量矩阵,  $S_{ij}$  是矩阵中第  $i$  行第  $j$  列的值, 代表代谢网络中代谢物  $i$  参与第  $j$  个生化反应的系数;  $v$  是变量向量, 由代谢网络中所有生化反应的流值组成,  $v_{\min}$  和  $v_{\max}$  分别是对生化反应  $v_j$  最小、最大的边界约束;  $c^T$  是由 0 和 1 组成的常量向量, 代表每个生化反应是否被设置为目标函数。通常目标函数设置为生物质 (biomass) 最大化<sup>[42]</sup>, 即求得细胞最优生长速率, biomass 的组成及各个成分所占的比重由实验测得。以上这些等式约束和不等式约束共同定义了一个可行的流分布空间, 利用线性规划算法即可求得满足所有约束条件的最优目标函数值, 并获得完整的代谢流分布。目前, 已有专门用于代谢网络基于约束的模拟分析工具, 如 COBRA (constraints based reconstruction and analysis)<sup>[43]</sup>, 运行在 Matlab 环境下。

## 1.3 全基因组范围代谢网络模型的应用

全基因组代谢网络模型为系统水平理解生物体的生理代谢特性, 进而对生物进行合理的重新设计和工程改造提供了有效的框架, 如基于模式生物大肠杆菌 GSMN<sup>[1,44]</sup> 的应用可分为以下五大类<sup>[45]</sup>。(1) 用于代谢工程研究。通过 *in silico* 菌株设计 (模拟基因/代谢途径的敲除或引入), 使其产生比野生型更好的特性, 能够生产高附加值产物。(2) 基因功能注释的完善。利用 GSMN 包含的已有的生化信息和遗传信息, 并结合其他实验数据 (如 BIOLOG), 能够发现功能注释不正确或者缺少功能注释的基因。(3) 预测特定环境下细胞内部的生理状态。模拟外部环境的变化对细胞表型的影响, 分析表达的基因、参与反应的代谢物以及代谢途径有哪些变化。(4) 分析和揭示生物网络固有的特性。如在各种环境下, 耦合表达的反应集合。(5) 进化研究。利用大肠杆菌的 GSMN 作为一个物种特异的遗传和代谢知识数据库, 和其他物种相应的 GSMN 进行比较, 分析适应性进化事件、水平基因转移和最小代谢网络的进化等。

对于其他微生物, 结合其生物特性, 可在多个不同的领域发挥其固有的代谢特性优势, 展现出巨大的应用前景和发展潜力, 下面将举例说明。首先,

基于 GSMN 方法对细菌进行工程改造,使其成为一个细胞工厂,能够生产高附加值产品,如在以下三大方面的应用:(1)生产生物聚合物,例如基于构建的 *Pseudomonas putida* 全基因组代谢网络进行改造,使其提高 poly-3-hydroxyalkanoates (PHA) 的产量<sup>[5,46,47]</sup>,替代传统的生产塑料需要使用石油化学产品的方法,降低对不可再生资源的依赖;(2)生产生物燃料<sup>[48-49]</sup>,例如利用 *Clostridium acetobutylicum*、*Methanosarcina maripaludis*、*Acinetobacter baylyi* 等微生物的代谢特性对其改造,提高乙醇、丁醇、氢气、甲烷等可再生的生物新能源<sup>[8,50-53]</sup>的生产,取代以石油为中心的能源工业体系,具有重要的战略意义;(3)生物制药,例如美国 Keasling 教授领导的研究小组成功地改造大肠杆菌生产抗疟疾药物青蒿素<sup>[54-56]</sup>,解决了传统的从植物提取耗时、成本高的问题,大大降低了制造费用,推动了这一药品的广泛使用。另一方面,应用 GSMN 方法改造微生物进行环境污染的治理,清除环境中很难被降解的毒性污染物,如利用 *Geobacter metallireducens* 和 *Geobacter sulfurreducens* 能够降解 Fe(III) 的代谢特性<sup>[57-58]</sup>,修复铀、钷、镉和钒等化学元素造成的环境污染。在医药健康方面,GSMN 可以用于致病菌潜在药物靶点的识别、抗菌药物的研发和疫苗的改良。目前,人类的基因组代谢模型已经构建完成<sup>[11,59]</sup>,越来越多的致病菌的 GSMN 也已经构建(包括 *Staphylococcus aureus*、*Mycobacterium tuberculosis*、*Salmonella typhimurium*、*Pseudomonas aeruginosa* 和 *B. cenocepacia* 等多种致病菌<sup>[2,4,6,18,60-63]</sup>),通过预测致病菌的必需基因/酶,并筛选出与人类基因组不同源的基因和蛋白质作为潜在药物靶点(如 *B. cenocepacia* 的 GSMN, iKF1028, 预测了 45 个潜在的药物靶点<sup>[18]</sup>),将为抗菌药物的研发、疫苗的研制等提供重要的基础。

## 2 基于全基因组代谢网络的最小基因组研究

最小基因组研究是合成生物学中重要的研究内容之一。无论是从头合成基因组 DNA,还是对现有通路进行改造,都需要一个合适的载体细胞或“底盘”(chassis),以实现特定的生物学功能或研究用途。理想的载体细胞应是复杂度低、稳定、强壮的,具有精简、健壮基因组结构,即“最小”基因组,以降低研究问题的复杂度,提高对所设计系统的可控性和可操作性。

最小基因组是指在最适宜的条件下,维持细胞

生长繁殖所必需的最小数目的基因<sup>[64]</sup>,因此,最小基因组研究的核心是确定必需基因集合。当然,这个“最小”的概念是相对的,并不是绝对的,由于研究目标的不同,最小基因组的组成可能有所不同。目前,针对基因组最小化和必需基因的研究主要有两种策略:比较基因组学方法和实验方法。然而,比较基因组学的方法忽略了必需功能的基因在物种之间的变异,而实验的方法对于可由代偿途径或基因多复制等补充其功能的必需功能基因则显得无能为力。因此,单靠这两种方法都无法得到一个全面的可以支持载体细胞生存或行驶功能的最小基因组。全基因组代谢网络模型的研究在预测必需基因和非必需基因方面具有突出优势,弥补了上述方法的不足,为从整体水平认识细胞提供了可能。GSMN 方法通过在计算机上模拟细胞生长环境,逐个敲除基因,修改相应 GPR 逻辑关系,重新设置相应生化反应的约束条件,再运行 FBA 算法,如果敲除基因后的模型导致目标函数为 0,即细胞不能够生长,该基因被预测为必需基因,反之为非必需基因。模型预测得到的必需基因集合对基因敲除实验中必需基因和可删除片段的筛选具有指导意义,也可以与实验结果相互验证。以全基因组代谢网络模型为基础开展计算机辅助的基因组最小化操作,进行必需基因的 *in silico* 预测,结合比较基因组学方法及基因组的大片段删除实验技术,将使得构建理想的具有“最小”基因组的底盘成为可能。

## 3 结语和展望

全基因组范围代谢网络的构建,能够整合、利用各种组合数据,从系统的角度,由生物体内各种分子的相互作用的研究到代谢通路、网络、模块,最终对整个生命活动的路线图有了完整的理解,进而对生物代谢功能进行重新设计和合理的工程改造,为开展合成生物学的系统研究与应用提供理论基础和生物信息平台。但是,目前仍然存在一些问题有待解决。首先,GSMN 方法有很强的物种特异性,不同的生物体具有不同的代谢特性,基因-蛋白质-化学反应之间的相互作用、代谢通路、细胞结构具有很大差异,因此,每个生物体的 GSMN 模型都需要独立的构建,不存在通用的 GSMN 模型。然而,GSMN 构建过程依赖人工完成和校对,需要花费大量的时间和人力。如图 1 所示,尽管随着测序技术的发展,完成全基因组测序的生物越来越多,但是,与完成 GSMN 构建的生物个数相比,两者

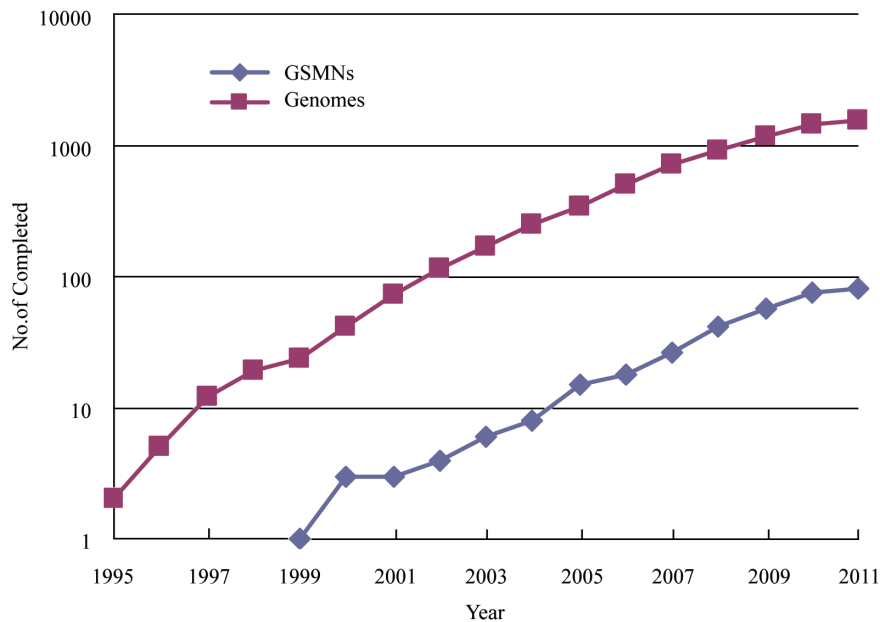


图1 已完成全基因组测序的物种个数和GSMN的个数统计

之间仍有一个鸿沟。目前, 虽然有一些辅助的数据库和软件, 但是并不能全自动地、用计算的方法完成基因组代谢网络的构建。其次, GSMN 进行 FBA 分析是基于拟稳态假设, 而真实生物体的代谢过程是动态的, 即代谢物在不同的时间内是有消耗或者积累的。目前, 已有关于代谢网络动态模型的建立<sup>[65-67]</sup>, 但只是建立小范围的动态模型。有待进一步优化、开发相关算法, 建立全基因组的动态代谢网络模型。最后, 生物体的代谢过程与转录调控、信号转导机制密切相关, 为了更加精确地做出预测, 更加真实地模拟生物体生命活动的路线图, 需要在代谢网络基础上, 整合转录调控网络和信号转导网络。目前, 已有模式生物 *Escherichia coli* 和 *Saccharomyces cerevisiae* 整合的调控网络、信号网络和代谢网络模型<sup>[68-70]</sup>, 但是, 对于其他非模式生物, 研究还不是很广泛, 缺乏生物实验数据, 阻碍了转录调控网络和信号转导网络的研究及整合; 但随着 GSMN 构建和分析技术的不断完善, 将有越来越多生物的 GSMN 被系统构建, 这些技术的创新与发展将为我们认识生命, 并进行有目的改造与优化提供重要的基础。

#### [参 考 文 献]

[1] Reed JL, Vo TD, Schilling CH, et al. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol*, 2003, 4(9): R45

[2] Becker SA, Palsson BO. Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation. *BMC Microbiol*, 2005, 5: 8

[3] Oh YK, Palsson BO, Park SM, et al. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *J Biol Chem*, 2007, 282(39): 28791-9

[4] Beste DJ, Hooper T, Stewart G, et al. GSMN-TB: a web-based genome-scale network model of *Mycobacterium tuberculosis* metabolism. *Genome Biol*, 2007, 8(5): R89

[5] Nogales J, Palsson BO, Thiele I. A genome-scale metabolic reconstruction of *Pseudomonas putida* KT2440: iJN746 as a cell factory. *BMC Syst Biol*, 2008, 2: 79

[6] Raghunathan A, Reed J, Shin S, et al. Constraint-based analysis of metabolic capacity of *Salmonella typhimurium* during host-pathogen interaction. *BMC Syst Biol*, 2009, 3: 38

[7] Roberts SB, Gowen CM, Brooks JP, et al. Genome-scale metabolic analysis of *Clostridium thermocellum* for bioethanol production. *BMC Syst Biol*, 2010, 4: 31

[8] Feist AM, Scholten JC, Palsson BO, et al. Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*. *Mol Syst Biol*, 2006, 2: 2006. 0004

[9] Gonzalez O, Gronau S, Falb M, et al. Reconstruction, modeling & analysis of *Halobacterium salinarum* R-1 metabolism. *Mol Biosyst*, 2008, 4(2):148-59

[10] Förster J, Famili I, Fu P, et al. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res*, 2003, 13(2): 244-53

[11] Duarte NC, Becker SA, Jamshidi N, et al. Global reconstruction of the human metabolic network based on geno-

- mic and bibliomic data. *Proc Natl Acad Sci USA*, 2007, 104(6): 1777-82
- [12] David H, Ozcelik IS, Hofmann G, et al. Analysis of *Aspergillus nidulans* metabolism at the genome-scale. *BMC Genomics*, 2008, 9:163
- [13] de Oliveira Dal'Molin CG, Quek LE, Palfreyman RW, et al. AraGEM, a genome-scale reconstruction of the primary metabolic network in *Arabidopsis*. *Plant Physiol*, 2010, 152(2): 579-89
- [14] Sigurdsson MI, Jamshidi N, Steingrimsdottir E, et al. A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1. *BMC Syst Biol*, 2010, 4: 140
- [15] Thiele I, Palsson BO. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc*, 2010, 5(1): 93-121
- [16] Feist AM, Herrgard MJ, Thiele I, et al. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol*, 2009, 7(2): 129-43
- [17] Durot M, Bourguignon PY, Schachter V. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiol Rev*, 2009, 33(1): 164-90
- [18] Fang K, Zhao H, Sun C, et al. Exploring the metabolic network of the epidemic pathogen *Burkholderia cenocepacia* J2315 via genome-scale reconstruction. *BMC Syst Biol*, 2011, 5(1): 83
- [19] Benson DA, Karsch-Mizrachi I, Lipman DJ, et al. GenBank. *Nucleic Acids Res*, 2009, 37(Database issue): D26-31
- [20] Kanehisa M, Goto S, Hattori M, et al. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res*, 2006, 34(Database issue): D354-7
- [21] The universal protein resource (UniProt). *Nucleic Acids Res*, 2008, 36(Database issue): D190-5
- [22] Schomburg I, Chang A, Ebeling C, et al. BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res*, 2004, 32(Database issue): D431-3
- [23] Saier MH, Jr., Tran CV, Barabote RD. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res*, 2006, 34(Database issue): D181-6
- [24] Ren Q, Chen K, Paulsen IT. TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Res*, 2007, 35(Database issue): D274-9
- [25] Schellenberger J, Park JO, Conrad TM, et al. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, 2010, 11: 213
- [26] Reed JL, Patel TR, Chen KH, et al. Systems approach to refining genome annotation. *Proc Natl Acad Sci USA*, 2006, 103(46): 17480-4
- [27] Orth JD, Palsson BO. Systematizing the generation of missing metabolic knowledge. *Biotechnol Bioeng*, 2010, 107(3): 403-12
- [28] Bochner BR, Gadzinski P, Panomitros E. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res*, 2001, 11(7): 1246-55
- [29] Tomita M, Hashimoto K, Takahashi K, et al. E-CELL: software environment for whole-cell simulation. *Bioinformatics*, 1999, 15(1): 72-84
- [30] Loew LM, Schaff JC. The Virtual Cell: a software environment for computational cell biology. *Trends Biotechnol*, 2001, 19(10): 401-6
- [31] Varner J, Ramkrishna D. Mathematical models of metabolic pathways. *Curr Opin Biotechnol*, 1999, 10(2): 146-50
- [32] Guardia MJ, Gambhir A, Europa AF, et al. Cybernetic modeling and regulation of metabolic pathways in multiple steady states of hybridoma cells. *Biotechnol Prog*, 2000, 16(5): 847-53
- [33] Arkin A, Ross J, McAdams HH. Stochastic kinetic analysis of developmental pathway bifurcation in phage  $\lambda$ -infected *Escherichia coli* cells. *Genetics*, 1998, 149(4): 1633-48
- [34] McAdams HH, Arkin A. Simulation of prokaryotic genetic circuits. *Annu Rev Biophys Biomol Struct*, 1998, 27: 199-224
- [35] Edwards JS, Covert M, Palsson BO. Metabolic modelling of microbes: the flux-balance approach. *Environ Microbiol*, 2002, 4(3): 133-40
- [36] Edwards JS, Ramakrishna R, Palsson BO. Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnol Bioeng*, 2002, 77(1): 27-36
- [37] Hatzimanikatis V, Emmerling M, Sauer U, et al. Application of mathematical tools for metabolic design of microbial ethanol production. *Biotechnol Bioeng*, 1998, 58(2-3): 154-61
- [38] Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng*, 2003, 5(4): 264-76
- [39] Schuster S, Dandekar T, Fell DA. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol*, 1999, 17(2): 53-60
- [40] Schilling CH, Letscher D, Palsson BO. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol*, 2000, 203(3): 229-48
- [41] Orth JD, Thiele I, Palsson BØ. What is flux balance analysis. *Nat Biotechnol*, 2010, 28(3): 4
- [42] Feist AM, Palsson BO. The biomass objective function. *Curr Opin Microbiol*, 2010, 13(3): 344-9
- [43] 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): 727-38
- [44] 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: 121
- [45] Feist AM, Palsson BO. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat Biotechnol*, 2008, 26(6): 659-67
- [46] Puchalka J, Oberhardt MA, Godinho M, et al. Genome-scale reconstruction and analysis of the *Pseudomonas*

- putida* KT2440 metabolic network facilitates applications in biotechnology. PLoS Comput Biol, 2008, 4(10): e1000210
- [47] Sohn SB, Kim TY, Park JM, et al. *In silico* genome-scale metabolic analysis of *Pseudomonas putida* KT2440 for polyhydroxyalkanoate synthesis, degradation of aromatics and anaerobic survival. Biotechnol J, 2010, 5(7): 739-50
- [48] Senger RS. Biofuel production improvement with genome-scale models: The role of cell composition. Biotechnol J, 2010, 5(7): 671-85
- [49] Wijffels RH, Barbosa MJ. An outlook on microalgal biofuels. Science, 2010, 329(5993): 796-9
- [50] Lee J, Yun H, Feist AM, et al. Genome-scale reconstruction and *in silico* analysis of the *Clostridium acetobutylicum* ATCC 824 metabolic network. Appl Microbiol Biotechnol, 2008, 80(5): 849-62
- [51] Senger RS, Papoutsakis ET. Genome-scale model for *Clostridium acetobutylicum*: Part I. Metabolic network resolution and analysis. Biotechnol Bioeng, 2008, 101(5): 1036-52
- [52] Stolyar S, Van Dien S, Hillesland KL, et al. Metabolic modeling of a mutualistic microbial community. Mol Syst Biol, 2007, 3: 92
- [53] Mao S, Luo Y, Zhang T, et al. Proteome reference map and comparative proteomic analysis between a wild type *Clostridium acetobutylicum* DSM 1731 and its mutant with enhanced butanol tolerance and butanol yield. J Proteome Res, 2010, 9 (6): 3046
- [54] Chang MC, Keasling JD. Production of isoprenoid pharmaceuticals by engineered microbes. Nat Chem Biol, 2006, 2(12): 674-81
- [55] Chang MC, Eachus RA, Trieu W, et al. Engineering *Escherichia coli* for production of functionalized terpenoids using plant P450s. Nat Chem Biol, 2007, 3(5): 274-7
- [56] Martin VJ, Pitera DJ, Withers ST, et al. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. Nat Biotechnol, 2003, 21(7): 796-802
- [57] Mahadevan R, Bond DR, Butler JE, et al. Characterization of metabolism in the Fe(III)-reducing organism *Geobacter sulfurreducens* by constraint-based modeling. Appl Environ Microbiol, 2006, 72(2): 1558-68
- [58] Sun J, Sayyar B, Butler JE, et al. Genome-scale constraint-based modeling of *Geobacter metallireducens*. BMC Syst Biol, 2009, 3: 15
- [59] Ma H, Sorokin A, Mazein A, et al. The Edinburgh human metabolic network reconstruction and its functional analysis. Mol Syst Biol, 2007, 3: 135
- [60] Jamshidi N, Palsson BO. Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the *in silico* strain iNJ661 and proposing alternative drug targets. BMC Syst Biol, 2007, 1: 26
- [61] Perumal D, Samal A, Sakharkar KR, et al. Targeting multiple targets in *Pseudomonas aeruginosa* PAO1 using flux balance analysis of a reconstructed genome-scale metabolic network. J Drug Target, 2011, 19(1): 1-13
- [62] Heinemann M, Kummel A, Ruinatscha R, et al. *In silico* genome-scale reconstruction and validation of the *Staphylococcus aureus* metabolic network. Biotechnol Bioeng, 2005, 92(7): 850-64
- [63] Lee DS, Burd H, Liu J, et al. Comparative genome-scale metabolic reconstruction and flux balance analysis of multiple *Staphylococcus aureus* genomes identify novel antimicrobial drug targets. J Bacteriol, 2009, 191(12): 4015-24
- [64] Koonin EV. How many genes can make a cell: the minimal-gene-set concept. Annu Rev Genomics Hum Genet, 2000, 1: 99-116
- [65] Mahadevan R, Edwards JS, Doyle FJ, 3rd. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. Biophys J, 2002, 83(3): 1331-40
- [66] Uygun K, Matthew HWT, Huang Y. DFBA-LQR: An optimal control approach to flux balance analysis. Ind Eng Chem Res, 2006, 45(25): 8554-64
- [67] Sauro HM. Network dynamics. Methods Mol Biol, 2009, 541: 269
- [68] Thiele I, Jamshidi N, Fleming RM, et al. Genome-scale reconstruction of *Escherichia coli*'s transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization. PLoS Comput Biol, 2009, 5(3): e1000312
- [69] Covert MW, Knight EM, Reed JL, et al. Integrating high-throughput and computational data elucidates bacterial networks. Nature, 2004, 429(6987): 92-6
- [70] Lee TI, Rinaldi NJ, Robert F, et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. Science, 2002, 298(5594): 799-804