

代谢途径分析研究进展及其应用

丁德武^{1,2}, 陆克中¹, 须文波², 方康年¹

DING De-wu^{1,2}, LU Ke-zhong¹, XU Wen-bo², FANG Kang-nian¹

1. 池州学院 数学与计算机科学系, 安徽 池州 247000

2. 江南大学 信息工程学院, 江苏 无锡 214036

1. Department of Mathematics and Computer Science, Chizhou College, Chizhou, Anhui 247000, China

2. School of Information Technology, Jiangnan University, Wuxi, Jiangsu 214036, China

E-mail: dwding@yahoo.com.cn

DING De-wu, LU Ke-zhong, XU Wen-bo, et al. Advance in metabolic pathway analysis and its application. Computer Engineering and Applications, 2009, 45(33): 193-196.

Abstract: Due to the absence of detailed kinetic parameters, one of the most important methods, network topology based metabolic pathway analysis has been developed. The paper firstly introduces the brief history of network based on pathway analysis. Then, two such network based methods are emphasized: Elementary modes and extreme pathways, with their similarities, differences and applications. At last, extreme pathways are employed to analyze PHB metabolism of *B. thuringiensis*. Except for "historical" biochemical pathways, the results also suggest that there are some novel pathways.

Key words: metabolic network; metabolic pathway; elementary modes; extreme pathways

摘要: 由于缺乏详尽的热力学参数, 基于网络拓扑的代谢途径分析是现阶段最重要的途径分析方法。首先简单地介绍了基于网络的途径分析方法发展简史, 随后着重说明了基于凸分析的两种代谢途径分析方法: 基元模式和极端途径及它们之间的异同, 并详细阐述了这两种方法的具体应用。最后, 使用极端途径分析了苏云金杆菌的 PHB 代谢。结果表明, 除了传统的途径, 还存在一些新颖的途径可用于 PHB 合成。

关键词: 代谢网络; 代谢途径; 基元模式; 极端途径

DOI: 10.3778/j.issn.1002-8331.2009.33.062 **文章编号:** 1002-8331(2009)33-0193-04 **文献标识码:** A **中图分类号:** TP339

近年新兴的复杂网络理论在生物学、物理学、社会学等多种不同的学科中都有着广泛的应用^[1-6]。然而事实上, 早在 20 世纪 80 年代, 就有人开始使用相关的数学理论来进行网络与代谢途径分析^[7]。但是, 这样形成的代谢途径只能有一个底物和一个产物, 并不适用于分析实际的代谢网络^[8]。后来, 有人开发出改进的算法, 但是在计算代谢途径之前需要先将代谢物分成四类, 这限制了该方法的使用^[8-9]。为了得到更通用的方法, Fell 等用线性代数理论计算代谢途径^[10]: 认为代谢网络对应的化学计量系数矩阵的一组线性无关解就是代谢网络的一组途径, 该组途径能够线性组合从而形成代谢网络的所有可能途径。这种方法存在两个主要问题: (1) 由于组成系数矩阵的线性无关解不是惟一的, 因而对应的途径组也不是惟一的; (2) 由于这些线性无关解可正可负, 而实际上有些反应是不可逆的, 这就导致了有些无关解实际上并不能作为途径。自凸分析被用于定义代谢途径开始, 代谢途径分析方法得到了长足发展, 逐渐形成了目前国际上普遍承认的两种代谢途径分析方法: 基元模式(EFMs)^[11-12]

和极端途径(EPs)^[13-14]。

将着重介绍基元模式和极端途径这两种代谢途径分析方法以及它们之间的异同, 并将详细阐述这两种方法的具体应用。最后, 将使用极端途径分析苏云金杆菌的 PHB 代谢。

1 两种代谢途径分析方法

1.1 基元模式及其应用

一个基元模式^[11-12]可以简单地理解为一个包含最少酶的集合, 这些酶可以使得该模式能够在稳定的代谢通量分布状况下朝一定方向进行。该定义包括了 3 个基本条件:

(1) 稳定状态, 这保证了内部代谢物没有积累;

(2) 可行性, 这要求该基元模式中所有不可逆反应的速率必须大于零, 即必须满足反应的热力学要求;

(3) 不可分, 即该基元模式中不能包含有其他的基元模式。

由于把可逆反应分成两个方向相反的不可逆反应将会增加基元模式的数目, 在计算过程中还需要消除没有生物学意义

基金项目: 国家自然科学基金(the National Natural Science Foundation of China under Grant No.60474030); 安徽省自然科学基金(the Natural Science Foundation of Anhui Province of China under Grant No.KJ2009B233Z); 池州学院引进硕士研究生项目(No.XYK200809)。

作者简介: 丁德武(1984-), 男, 助教, 主要研究领域为计算智能、计算系统生物学; 陆克中(1976-), 男, 讲师, 主要研究领域为生物信息学、网络安全; 须文波(1946-), 男, 博士, 教授, 博导, 主要研究领域为轻工信息技术、生物信息学等。

收稿日期: 2009-06-25 **修回日期:** 2009-07-30

的二元环。因此,基元模式不把可逆反应分成两个方向相反的不可逆反应。另外,区分外部代谢物和内部代谢物在基元模式的计算中非常重要的,这两者必须在计算基元模式之前进行分类。其中,外部代谢物是指有一定缓冲浓度的代谢物;而内部代谢物是指在稳定状态下达到平衡的代谢物,即该代谢物生成的速率和消耗的速率相等。

目前,基元模式已经成为系统生物学、生物技术和代谢工程中的一个重要理论工具被广泛使用,主要应用有识别没有被途径利用的反应、识别酶子集、计算网络的冗余性、预测优化途径和优化生长速率、计算最小切割子集、简单预测转录调控、标注未知孤独基因功能、探测热力学可行性等^[8,15-21]。

(1)识别那些没有连接到代谢网络上的反应,这些反应的反应物或者产物不被代谢网络模型的其他部分产生或者消耗,这也同时说明了代谢网络中存在着较弱的结构特征,这些反应现在已经通过基因组标注得到了证实^[15]。

(2)分析基于基因组规模的代谢网络的酶子集(Enzyme subset)。酶子集是指那些在网络中不得不在一起操作的,并且结构上彼此需要的反应组,或者也可以称着基于基因组规模的代谢网络中系统性相关的反应组。这些酶子集(相关的反应组)在理解代谢网络的调控结构上具有重要意义^[16]。

(3)研究途径的冗余性。途径的冗余性是用来描述有相同的底物或者相同的产物的系统独立的途径数量的,这是网络结构鲁棒性的一种定量描述方法^[17]。

(4)细胞功能包括生长和调节功能都被用基元模式进行了分析,例如 Stelling 等人用基元模式计算了使用两种不同底物的理论转录比例,预测结果与实验结果一致^[17]。

(5)Liao 等人计算了 E.coli 的核心代谢网络的基本反应模式,然后利用这些基本反应模式来分析用糖类生成芳香族氨基酸前体(3-脱氧阿拉伯庚酮糖酸-7-磷酸, DAHP)的最优化产率和次优化产率,设计了相应的 E.coli 工程菌并且与计算预测值一致^[18]。

(6)Klamt 等借助基元模式的概念引入了一个概念,叫最小切割子集(Minimal Cut Set, MCS)。最小切割子集是指在代谢网络中一个最小的即数量不可进一步减少的反应组,该反应组的所有反应失活将会导致该代谢网络中某一个功能的失效。通过使用最小切割子集的方法,一系列潜在的应用都能够被很容易地实现,包括网络证实、表型预测、评价结构的刚性和鲁棒性、代谢网络结构分析和药物靶点识别等^[19]。

(7)Forster 等人还利用基元模式分析了酿酒酵母核心代谢网络的代谢数据,标注了一些未知孤独基因的功能^[20]。

(8)Beard 等人的研究表明,基元模式所表示的内部循环在热力学上是可行的^[21]。

1.2 极端途径及其应用

极端途径^[13-14]的概念与基元模式很接近。极端途径除了满足基元模式定义之外,还有几点与基元模式不一样:

(1)具有系统独立性,系统独立性是指极端途径集合 $\{P_1, P_2, \dots, P_n\}$ 中没有任何一条途径可以写成其他途径的非负线性组合。系统独立性和线性代数中的向量的独立性是不完全一样的,后者不一定是非负线性组合,而前者必须是非负线性组合。

(2)不区分内部代谢物和外部代谢物,而区分交换反应和内部反应,根据这种区分,极端途径被分为3类。

(3)所有可逆的内部反应都必须被分解成两个方向相反的不可逆反应,这样规定是为了限制所有内部通量必须是非负的。

与基元模式一样,极端途径也已经被广泛应用于红血球细胞^[22],大肠杆菌^[23],嗜血性流感杆菌^[24]和幽门螺杆菌^[25]等代谢途径的分析,主要用于预测最小底物组成、识别没有被途径利用的反应、识别酶子集、计算网络的冗余性、预测优化途径和优化生长速率、预测基因修饰后结果、简单预测转录调控等^[8,22-25]。

1.3 两种方法的比较

如果网络中所有与环境交换的反应都不可逆时,有生物学意义的基元模式集合和极端途径集合是一致的。但是,通常一个网络的交换反应中都有可逆反应存在,这样网络中的极端途径集合就是基元模式集合的子集^[26-28]。图1给出了一个简单的网络范式的基元模式和极端途径比较,可以看出基元模式1-3分别对应三个极端途径,而基元模式4可以由前两个极端途径组合得到,这简单地说明了极端途径集合就是基元模式集合的子集。

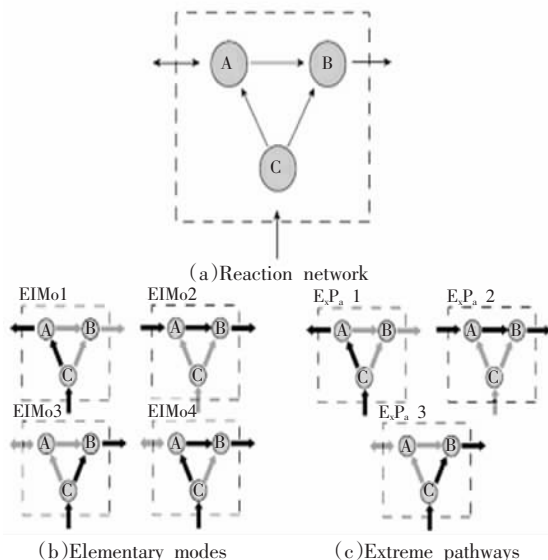


图1 一个网络范式的基元模式和极端途径比较

(a)具有3个代谢物,3个内部反应和3个交换反应的网络范式;(b)该网络的四个基元模式;(c)该网络的三个极端途径

2 极端途径分析实例:PHB 代谢

2.1 PHB 简介

如前所述,基元模式和极端途径是两个非常相似的方法,但是由于极端途径满足系统独立性的要求,一般来说所得极端途径的数量比基元模式要少,所以这里选取极端途径方法分析了一种非常重要的可降解生物材料聚羟基丁酸酯(poly-beta-hydroxybutyrate, PHB)的代谢。PHB 广泛存在于自然界的生物体中,一般作为能源物质用于合成细胞物质或为降解体外有机物质提供能量,目前已广泛应用于生物医药、生物材料、环境保护、工农业生产等领域。然而通常情况下,细胞内的 PHB 含量很低,生产成本非常高,因此有必要对 PHB 的生物合成过程进行优化控制。

2.2 数据集

选用 KEGG 数据库中信息相对比较完善的苏云金杆菌代谢网络的丁醇代谢途径作为 in silico 模型来执行本次分析,表1列出了其中涉及的所有代谢物缩写(限于篇幅,代谢物全称和代谢反应列表未给出,感兴趣的读者可通过邮件从作者处获得),按照极端途径的计算要求将代谢物分成了三类(见表2)。

其中主要的交换物(primary exchanges)是指与外界环境交换的代谢物, 而流通代谢物(currency exchanges)是指 ADP, ATP, NADH 等辅因子^[14]。

此外, 排除了明显的孤立反应, 如酶 aldehyde dehydrogenase(EC 1.2.1.3), 4-aminobutyrate aminotransferase(EC 2.6.1.19)催化的反应。随后, 可以得到该模型的化学计量矩阵, 并画出了相应的代谢图(图 2)。

表 1 苏云金杆菌代谢网络丁醇代谢途径中的所有 29 个代谢物列表

代谢物分类	代谢物缩写
P.Ex.	3Hb, 33Hbb, 3H3mCoA, 3Hbt, 1Bt, PYR, FOR, Ac, CoA.
Int.	3HbCoA_R, Bt, Btt, BtP, BtCoA, Act, AcCoA, CrCoA, AaCoA, PHB.
C.Ex.	ADP, ATP, NAD, NADH, NADP, NADPH, PI, H, H ₂ O.

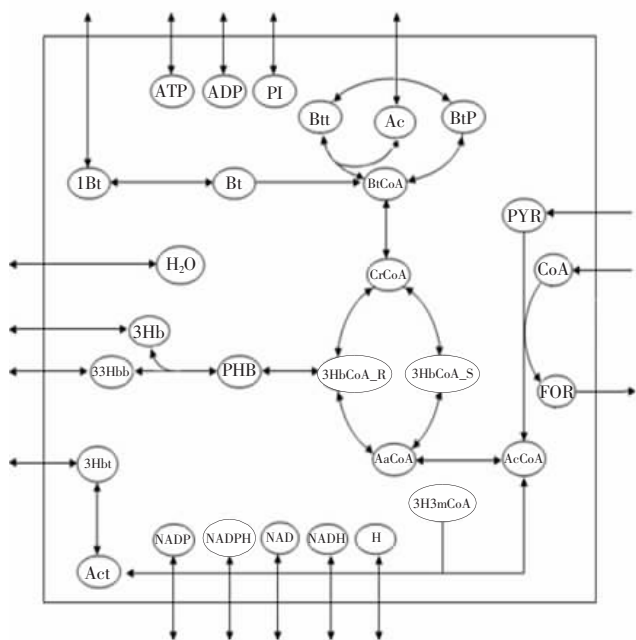


图 2 苏云金芽胞杆菌丁醇代谢途径

2.3 方法

极端途径的计算是基于如下矩阵方程进行的:

$$S \cdot v = 0 \quad (v_i \geq 0, \forall i)$$

其中 S 是化学计量矩阵, 计算极端途径的基本思想可描述如下^[14]:

(1) 初始化: 确定化学计量矩阵 S , 将一单位矩阵 E 和 S^T 相连接, 组成新矩阵 ES^T ; 找出代谢网络中的所有外部代谢物(即不限制其交换通量方向的代谢物), 把所有这些外部代谢物的代谢反应在化学计量矩阵中所对应的行移动到一个临时矩阵 $T(E)$ 中, 剩余的行组成的矩阵记为 $T(0)$ 。

(2) 找出所有的内部代谢物, 记其数量为 n_c 。

(3) 选择某个内部代谢物, 记其在化学计量矩阵中对应的列为 c , 将矩阵 $T(x-1)$ 改记为矩阵 $T(x)$, 其中 $T(x-1)$ 中 c 列值为 0 的行保持不变。

(4) 把 $T(x-1)$ 中 c 列值非 0 的任意两个符号相反的行组合, 假设这两行为 r_1, r_2 , 对应的值为 $r_{1,c}, r_{2,c}$, 则新的一行对应的值为 $r = |r_{1,c}| \times r_1 + |r_{2,c}| \times r_2$ 。

(5) 对所有在第 3 和 4 步加入 $T(x)$ 的行, 确保任一行不是 $T(x)$ 其他行的非负线性组合。可设 $A(i)$ 为第 i 行为 0 的所有元素的列标号集合, 检查任意两行对应的标号集合是否其中一个为另一个的子集, 如若 $A(i)$ 包含于 $A(j)$, 则从 $T(x)$ 中剔除 $A(i)$ 。

(6) 对在第 2 步中找出的所有内部代谢物重复第 3~5 步, 则最后的矩阵为 $T(n)$ 。

(7) 把 $T(E)$ 放在 $T(n)$ 下面, 组成一个新的矩阵。

(8) 把新得的矩阵的上部即 $T(n)$ 和下部即 $T(E)$ 进行线性变换, 使得所有外部代谢物对应的值在 $T(n)$ 中全部为 0, 则线性变换后原来 $T(n)$ 的位置就对应了所有的极端途径。

一般而言, 极端途径可以分为三类^[13]: I 类极端途径(Primary metabolic pathway)是指包含了主要代谢物(这类代谢物是代谢网络的主干, 如葡萄糖代谢中的葡萄糖)的代谢途径, 这类极端途径有外部流量穿过系统边界; II 类途径(Futile pathway)是指不包含主要代谢物的代谢途径, 这类极端途径也有外部流量, 但主要是 ATP, ADP, NADH 等流通代谢物, 它们主要起辅助作用, 以维持 I 类极端途径功能转换的顺利进行; 而 III 类途径(Reaction Cycling)则仅仅是由可逆反应组成的循环, 它们主要起调节代谢物间平衡的作用(表 2)。

表 2 三类极端途径间的差异

	I 类	II 类	III 类
内部通量	✓	✓	✓
主要交换通量	✓	×	×
流通交换通量	✓	✓	×

2.4 结果与讨论

计算苏云金杆菌代谢网络的丁醇代谢途径得到 33 条极端途径: 18 条 I 类途径, 2 条 II 类途径, 13 条 III 类途径, 这里主要分析了 I 类和 II 类极端途径(限于篇幅, 途径列表未给出, 感兴趣的读者可通过邮件从作者处获得)。

(1) 无效循环(FC)

一般说来, II 类极端途径称为无效循环(Futile Cycles), 但是要注意它们和由 2 个可逆反应组成的循环(即 III 类途径)有着本质的不同, 它们意味着网络中代谢物的转换并将产生焓^[2], 因此不能如 III 类途径那样直接忽略掉它们。此外, 识别这种无效循环也可用于避免计算其中损耗的通量^[11]。这两个 II 类途径表示了 CrCoA 和 AaCoA 间的可逆转换, 主要用于平衡 NAD/NADH, NADP/NADPH 比率, 这些必需的辅因子(NADH 和 NADPH)可用于维持代谢渗透性平衡。另外, 注意到第 2 个转换是非常重要的因为 PHB 的生物合成需要还原性物质。

(2) 从 AcCoA 到 PHB(AP)

计算结果还表明了传统的 PHB 合成途径(AP 组)是极端途径, 这 8 条途径包含了与 PHB 合成相关的基本通路(以 AcCoA 为底物)。其中 AP1-3 和 AP6 利用 3HbCoA_S 和 CrCoA 作为中间代谢物, 而其他的途径则没有利用该代谢物。这表明生化途径在一定的条件或需求下是可以转变的。

(3) 从 1Bt 到 PHB(BP)

尽管 AP 组表明了传统的 PHB 合成途径是极端途径, 但是计算结果还发现了大量的非传统途径(BP 组)。由于基于化学计量矩阵的代谢途径可以用于预测新奇的代谢途径, 因而这里的结论可以在代谢工程中用于优化 PHB 的合成过程。

(4) 组合(CB)

许多条件下, 可以由许多不同的代谢途径来合成所需的代谢物。这里的 CB 组就是由之前的 AP 和 BP 组合而得到的。这些组合的途径进一步反应了生物体的新陈代谢具有一定的鲁棒性, 用于对抗外部环境的变化或者其他原因如突变等。

(5) AcCoA 合成(AcCoA)

AcCoA 组显示了 3 条 AcCoA 合成途径。AcCoA1 和 AcCoA2 的不同之处是分别利用 3HbCoA_S 和 3HbCoA_R 来完成从 CrCoA 到 AaCoA 的转换, AcCoA3 是利用 PYR 作为底物。注意到, 所有这三条途径均包含了反应 $BtP \rightarrow Btt$, 因为 AcCoA 合成过程需要 ATP。

3 结束语

随着越来越多的基因组规模的代谢网络被重建, 相应的生化途径分析方法显得越来越重要。由于缺乏详尽的热力学参数, 基于约束的分析方法(如基于网络的途径分析)具有很高的实际应用价值。

使用极端途径分析了 PHB 的新陈代谢, 从结构上将所得的极端途径分为 3 类, 并分别作了相应的功能分析, 得到了一组新的 PHB 合成途径, 可以用于指导优化 PHB 的合成。

然而, 对 PHB 合成途径的分析主要是建立在丁醇代谢这一代谢子系统之上的, 为了得到更广泛的结论, 进一步的工作将需要扩充模型, 从全基因组规模的代谢网络上分析某一代谢物的代谢途径。

参考文献:

- [1] 丁德武, 刘涛, 陆克中. 复杂网络的中心化及其在代谢网络中的应用[J]. 计算机与应用化学, 2008, 25: 1508-1510.
- [2] 王林, 戴冠中. 基于复杂网络社区结构的论坛热点主题发现[J]. 计算机工程, 2008, 34: 214-217.
- [3] 王建伟, 荣莉莉. 突发事件的连锁反应网络模型研究[J]. 计算机应用研究, 2008, 25: 3288-3291.
- [4] 李敏, 陈建二, 王建新. 基于复杂网络理论的 PPI 网络拓扑分析[J]. 计算机工程与应用, 2008, 44(8): 20-22.
- [5] 雷霆, 余镇危. 基于复杂网络理论的计算机网络拓扑研究[J]. 计算机工程与应用, 2007, 43(6): 132-135.
- [6] 赵静, 俞鸿, 骆建华, 等. 应用复杂网络理论研究代谢网络的进展[J]. 科学通报, 2006, 51: 1241-1248.
- [7] Clarke B L. Stoichiometric network analysis[J]. Cell Biophys, 1988, 12: 237-253.
- [8] 何锋, 马红武, 赵学明, 等. 生物信息学用于代谢网络研究的进展与展望[J]. 化工学报, 2004, 55: 1593-1601.
- [9] Mavrouniotis M L, Stephanopoulos G, Stephanopoulos G. Computer-aided synthesis of biochemical pathways[J]. Biotechnol Bioeng, 1990, 36: 1119-1132.
- [10] Fell D A. Metabolic control analysis: A survey of its theoretical and experimental development[J]. Biochem J, 1992, 286: 313-330.
- [11] Schuster S, Dandekar T, Fell D A. Detection of elementary flux modes in biochemical networks: A promising tool for pathway analysis and metabolic engineering[J]. Trends Biotechnol, 1999, 17(2): 53-60.
- [12] Schuster S, Fell D A, Dandekar T. A general definition of metabolic

- pathways useful for systematic organization and analysis of complex metabolic networks[J]. Nature Biotechnol, 2000, 18(3): 326-332.
- [13] Schilling C H, Schuster S, Palsson B O, et al. Metabolic pathway analysis: Basic concepts and scientific applications in the post-genomic era[J]. Biotechnol Prog, 1999, 15(3): 296-303.
- [14] Schilling C H, Letscher D, Palsson B O. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective[J]. J Theor Biol, 2000, 203(3): 229-248.
- [15] Schilling C H, Covert M W, Famili I, et al. Genome-scale metabolic model of helicobacter pylori 26695[J]. J Bacteriol, 2002, 184(16): 4582-4593.
- [16] Pfeiffer T, Sanchez-Valdenebro I, Nuno J C, et al. Metatool: For studying metabolic networks[J]. Bioinformatics, 1999, 15(3): 251-257.
- [17] Stelling J, Klamt S, Bettenbrock K, et al. Metabolic network structure determines key aspects of functionality and regulation[J]. Nature, 2002, 420(6912): 190-193.
- [18] Liao J C, Hou S Y, Chao Y P. Pathway analysis, engineering, and physiological consideration for redirecting central metabolism[J]. Biotechnol Bioeng, 1996, 52: 129-140.
- [19] Klamt S, Gilles E D. Minimal cut sets in biochemical reaction networks[J]. Bioinformatics, 2004, 20(2): 226-234.
- [20] Forster J, Gombert A K, Nielsen J. A functional genomics approach using metabolomics and in silico pathway analysis[J]. Biotechnol Bioeng, 2002, 79(7): 703-712.
- [21] Beard D A, Liang S D, Qian H. Energy balance for analysis of complex metabolic networks[J]. Biophys J, 2002, 83(1): 79-86.
- [22] Wiback S J, Palsson B O. Extreme pathway analysis of human red blood cell metabolism[J]. Biophys J, 2002, 83(2): 808-818.
- [23] Schilling C H, Edwards J S, Letscher D. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems[J]. Biotechnol Bioeng, 2000, 71(4): 286-306.
- [24] Papin J A, Price N D, Edwards J S, et al. The genome-scale metabolic extreme pathway structure in Haemophilus influenzae shows significant network redundancy[J]. J Theor Biol, 2002, 215(1): 67-82.
- [25] Price N D, Papin J A, Palsson B O. Determination of redundancy and systems properties of the metabolic network of helicobacter pylori using genome-scale extreme pathway analysis[J]. Genome Res, 2002, 12(5): 760-769.
- [26] Papin J A, Stelling J, Price N D, et al. Comparison of network-based pathway analysis methods[J]. Trends Biotechnol, 2004, 22(8): 400-405.
- [27] Klamt S, Stelling J. Two approaches for metabolic pathways analysis[J]. Trends Biotechnol, 2003, 21(5): 64-69.
- [28] Papin J A, Price N D, Wiback S J, et al. Metabolic pathways in the post-genome era[J]. Trends Biochem Sci, 2003, 28(5): 250-258.