

# 一种基于代谢网络分析最小化基因组的方法及其在大肠杆菌中的应用

汤彬彩<sup>1,2</sup>, 郝彤<sup>1,2,4</sup>, 袁倩倩<sup>1,2,3</sup>, 陈涛<sup>1,2</sup>, 马红武<sup>1,2,3</sup>

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

4 天津师范大学生命科学学院, 天津 300387

汤彬彩, 郝彤, 袁倩倩, 等. 一种基于代谢网络分析最小化基因组的方法及其在大肠杆菌中的应用. 生物工程学报, 2013, 29(8): 1173-1184.

Tang BC, Hao T, Yuan QQ, et al. Genome minimization method based on metabolic network analysis and its application to *Escherichia coli*. Chin J Biotech, 2013, 29(8): 1173-1184.

**摘要:** 最小生命体的合成是合成生物学研究的重要方向。最小化基因组的同时而又不对细胞生长产生影响是代谢工程研究的一个重要目标。文中提出了一种从基因组尺度代谢网络模型出发, 通过零通量反应删除及对非必需基因组删除计算获得基因组最小化代谢网络模型的方法, 利用该方法简化了大肠杆菌经典代谢网络模型 *iAF1260*, 由起始的 1 260 个基因简化得到了 312 个基因, 而最优生物质生成速率保持不变。基因组最小化代谢网络模型预测了在细胞正常生长的前提下包含最少基因的代谢途径, 为大肠杆菌获得最小基因组的湿实验设计提供了重要参考。

**关键词:** 基因组尺度代谢网络, 基因组最小化, 组合算法, 大肠杆菌

**Received:** March 26, 2013; **Accepted:** May 7, 2013

**Supported by:** National Basic Research Program of China (973 Program) (Nos. 2012CB725203, 2011CBA00804), National Natural Science Foundation of China (Nos. NSFC-21106095, NSFC-21176182).

**Corresponding author:** Tong Hao. Tel: +86-22-23766672; E-mail: haotong@tju.edu.cn

Hongwu Ma. Tel: +86-22-24828735; E-mail: ma\_hw@tib.cas.cn

国家重点基础研究发展计划 (Nos. 2012CB725203, 2011CBA00804), 国家自然科学基金 (Nos. NSFC-21106095, NSFC-21176182) 资助。

# Genome minimization method based on metabolic network analysis and its application to *Escherichia coli*

Bincai Tang<sup>1,2</sup>, Tong Hao<sup>1,2,4</sup>, Qianqian Yuan<sup>1,2,3</sup>, Tao Chen<sup>1,2</sup>, and Hongwu Ma<sup>1,2,3</sup>

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 Biotechnology, Chinese Academy of Sciences, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

4 College of Life Sciences, Tianjin Normal University, Tianjin 300387, China

**Abstract:** The minimum life is one of the most important research topics in synthetic biology. Minimizing a genome while at the same time maintaining an optimal growth of the cells is one of the important research objectives in metabolic engineering. Here we propose a genome minimization method based on genome scale metabolic network analysis. The metabolic network is minimized by first deleting the zero flux reactions from flux variability analysis, and then by repeatedly calculating the optimal growth rates after combinatorial deletion of the non-essential genes in the reduced network. We applied this method to the classic *E. coli* metabolic network model --iAF1260 and successfully reduced the number of genes in the model from 1 260 to 312 while maintaining the optimal growth rate unaffected. We also analyzed the metabolic pathways in the network with the minimized number of genes. The results provide some guidance for the design of wet experiments to obtain an *E. coli* minimal genome.

**Keywords:** genome-scale metabolic network, genome minimization, combination algorithm, *Escherichia coli*

随着 2010 年 5 月美国生物学家文特尔等在 *Science* 杂志发表了关于由化学合成的基因组控制的细胞的文章<sup>[1]</sup>, 人工合成生命的研究引起了广泛的关注。构建最小生命体一直是人工合成生命的重要内容之一, 文特尔的工作再次引发了研究者们对最小基因组的研究兴趣。

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, 在最小基因组中, 敲除任何基因后细胞都会死亡。在不同培养条件下, 得到的最小基因组可能是不一样的, 这与生物本身的合成能力有关, 所以最小基因组是相对而言的。在代谢工程研究领域中, 人们关心更多的是如何得到一个最小化的基因组而同时保持细胞的生长不受影响。

目前对最小基因组的研究在湿实验方面已经取得了一些进展, 1995 年起研究者开始对生殖支原体进行研究<sup>[2]</sup>, 因为生殖支原体是能够自主复制并生存的最小细胞生物, 其基因组只有 580 kb, 2007 年 Lartigue 等将生殖支原体整个基因组转移到掏空的山羊支原体里<sup>[3]</sup>并使之成活, 2008 年 Gibson 等人工合成整个生殖支原体<sup>[4]</sup>, 这些工作体现着研究者们对获得人工合成最小生命的不断探索。除了“自下而上”地从头合成基因组外, 另一个研究思路是从已有物种出发, 通过“自上而下”地逐步敲除基因, 去除细胞中的冗余基因, 得到最小基因组。例如, Posfai 等通过大片段基因敲除成功将 *Escherichia coli* K-12 基因组减少了 15%<sup>[5]</sup>。同时这一小基因组大肠杆

菌还具有一些意想不到的特性：比如高效的电转化率，准确的复制重组基因和稳定的质粒复制等。这揭示了在大肠杆菌中并非所有的基因都是必需的，通过大量删除基因，可以获得具有新性能的微生物。根据这一思路，必需基因<sup>[6-7]</sup>的确定是获得最小基因组的重要依据。但最小基因组中除了这些必需基因外还需要包括其他的基因，这些基因虽然单独敲除不会对细胞产生影响，但两个或多个同时敲除时则可能导致细胞死亡。如何找出这些非必需但又需要包含在最小基因组中的基因是目前研究中面临的一项重要挑战。而基因组规模代谢网络则有可能为解决这一问题提供可能性。

基因组尺度的代谢网络模型指的是以注释信息及基因序列为基础，利用基因-蛋白质-反应的关系为主线，成功模拟代谢物在生物机体内的过程的一种包含大量信息的模型，到 2012 年 6 月为止已经构建完成 135 个基因组尺度代谢网络模型，涉及 64 种原核生物、19 种真核生物和 4 种古生菌 (GSMNDB: <http://synbio.tju.edu.cn/GSMNDB/gsmndb.htm>)。国内外已经有数篇关于基因组尺度代谢网络模型重构以及模拟方法的综述<sup>[8-12]</sup>，模型在模拟基因敲除、靶点药物发现、扰动实验预测以及物种进化研究方面都有着重要作用<sup>[13]</sup>。

虽然最小基因组的研究在湿实验方面已经有了一些进展，但是到目前为止还没有一个完整的关于最小基因组的代谢网络模型和计算方法，Yizhak 等<sup>[14]</sup>注重研究基因敲除与合成生物量之间的关系，Pa 等<sup>[15]</sup>构建的模型只是对于模型的简化，并没有得到一个不能删除任何基因的最小

基因组模型。在本文中我们按照“自上而下”的研究思路，通过对基因组尺度代谢网络模型的逐步精简第一次获得最小基因组代谢网络模型。

## 1 方法

### 1.1 通量平衡分析方法

在基因组规模代谢网络模型分析中较为广泛应用的是基于约束的最优化方法，这一方法主要由决策变量、约束条件和目标函数三部分组成，本文采用其中的通量平衡分析 (Flux balance analysis, FBA)<sup>[16-17]</sup>方法进行模拟计算，在 FBA 中，决策变量就是模型中所有反应的通量值，不同的通量分配方案决定着不同的代谢状态，限制条件是实现最优化模拟必须满足的基本条件，包括等式限制和不等式限制，等式限制条件基于拟稳态条件下所有中间代谢物满足物质守恒的假设，即代谢物的生成与消耗相等，而不等式限制基于模型中每个反应速度的极限值。约束条件表示为数学形式如下：

$$\begin{aligned} \text{s.t. } & S \cdot v = 0 \\ & v_{\min} \leq v_i \leq v_{\max}, \quad i=1,2,\dots,n \end{aligned}$$

其中， $v_{\min}$  和  $v_{\max}$  分别表示反应速度的最小和最大值，在本文中，在没有特殊限制条件下，对于可逆反应，其最小值和最大值分别为 -1 000 和 1 000 mmol g/(DW·h)，而对于不可逆反应，其最小值和最大值则分别为 0 和 1 000 mmol g/(DW·h)。目标函数在数学上可以表示为决策变量的解析函数 (即一系列决策变量的线性或非线性组合)，根据不同的模拟目的，目标函数既可以只有一个，也可以同时有多个，本文的目标是在保持细胞生长不受影响的情况下将基因组最小化，因此

以生物量的最大速率为目标函数。在本文中当一个基因被敲除后,生物合成量的模拟值小于最初模型的生物合成量,这样的基因被定义为必需基因。我们利用的计算工具是 Matlab 平台下运行的 COBRA<sup>[18]</sup>工具包。

## 1.2 通量可变性分析

通量可变性分析 (Flux variability analysis, FVA)<sup>[19]</sup>是在 FBA 基础上发展的通量分析方法。FBA 计算出来的最优解不是唯一的,有多种通量分布可以使目标反应达到最优值。FVA 就是在保证目标反应在最优值的情况下得到每个反应的通量的可能最大值和最小值,从而得到每个反应的通量变化范围。我们定义那些通量变化范围总是大于 0 或总是小于 0 的反应为达到最优解的必需反应,因为这些反应对每个最优解都是必须要存在的。这里要说明的是必需反应对应的基因并非都是必需基因,因为一个反应可能有多个同工酶基因对应。除了必需反应外的其他反应定义为非必需反应,这些非必需反应可以单个敲除但不能同时敲除。

## 1.3 组合删除法

在代谢网络中从代谢物 A 到 B 可能有多个不同的代谢途径,如图 1 所示,从 A 到 B 可以通过 A-X-B、A-Y-Z-B 和 A-M-N-O-B 三条路线实现,显然 A-X-B 途径最短。因而需要考虑的问题是如何保证在精简的代谢网络中从代谢物 A 到 B 需要通过的途径是基因 X 而不是 Y-Z 或者 M-N-O。我们采用的方法是将参与从 A 到 B 所有途径的基因 {X,Y,Z,M,N,O} 进行分步组合模拟删除,首先进行单基因敲除实验,发现可被敲除的基因数目即非必需基因数目为 6 个,第一步从

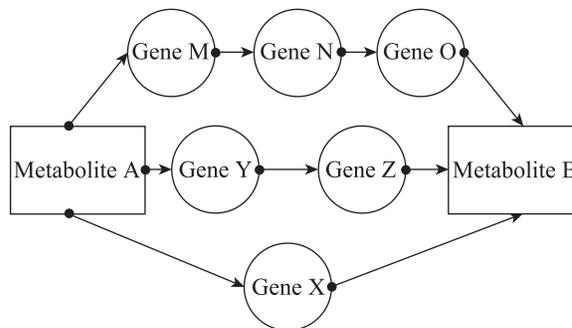


图 1 从代谢物 A 到 B 的代谢途径示例

Fig. 1 Metabolic pathways from metabolite A to B.

这 6 个基因里取出 6 个基因组合模拟删除,看能否得到代谢物 B,只有一种取法  $C(6,6)=1$ ,很明显不能,那我们就从 6 个基因里选出 5 个基因进行组合,共有  $C(5,6)=6$  种选法,每种选法都进行一次模拟删除,看能否得到代谢物 B,在本例中当选取的组合为 {Y,Z,M,N,O} 的时候,代谢物 A 能够生成 B,最后保留的基因数目就为 6 (所有非必需基因数目)-5 (选取组合包含基因的数目)=1 (非必需基因保留的数目),我们将这种思想运用到整个模型上,以葡萄糖为底物,生物量合成为目标函数,利用程序对非必需基因进行逐步递减组合模拟删除,在保证生物量不变的情况下,保留删除基因最多的反应途径在模型中,最后得到基因组最小化的代谢网络模型。

## 1.4 基因组简化的步骤

整个基因组简化步骤如图 2 所示,主要包括以下步骤:

1) 对初始代谢网络模型进行 FVA 分析,其中有很多反应的最大通量和最小通量都为 0,即这些反应与细胞最优生长无关,因而首先将这些反应从网络中删除。

2) 处理同工酶基因。在模型中有很多同工

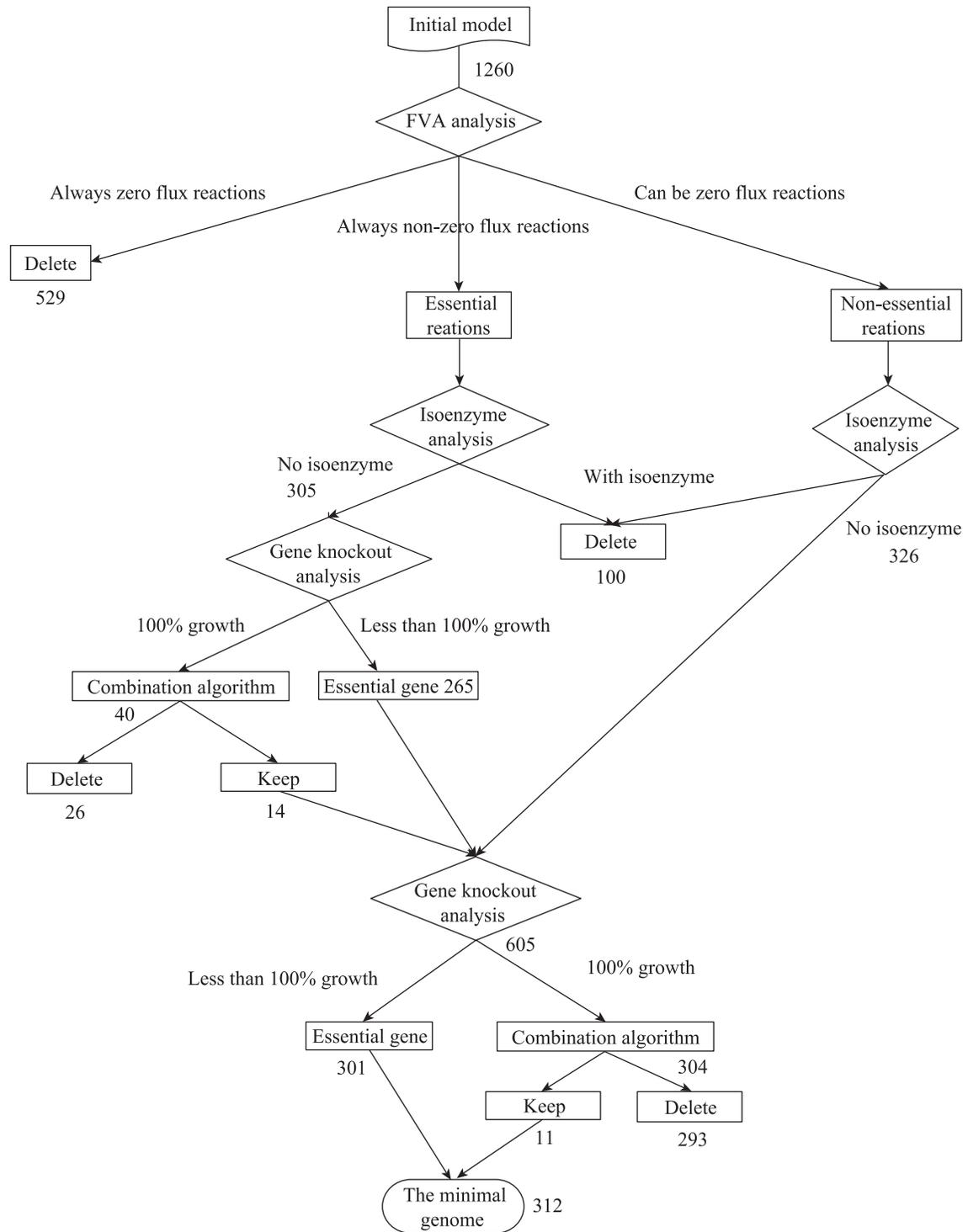


图 2 基因组最小化的步骤 (图中数字为各步处理后的基因数)

Fig. 2 Procedure of genome minimization. The numbers of genes after every step are shown in the diagram.

酶基因催化同一反应的情况,这些基因只需存在一个(组),反应就能够进行,因而多余的基因可以去除。需要注意的是在这些同功能基因中,有些同时催化多个反应,其中有些反应是菌体生长所必需的,因而简单地去除每个同功能基因可能导致菌体的死亡,因而我们仅将只催化单个(一类)反应的同功能基因去除,而同时催化多个反应的同功能基因则保留下来进行进一步计算。

3) 经过上述两步我们得到了部分缩减的代谢网络模型,通过 FVA 分析模型中的反应分为必需反应和非必需反应两部分,在此基础上我们将该模型中的基因分为两个子集:保留基因集和候选基因集。所有必需反应中的基因构成保留基因集,要提出的是有的必需反应是有几个同功能基因催化的,只要保证该必需反应能够进行,多余的基因是可以被删除的,所以保留基因集并不等于所有基因都保留,除了保留基因以外的基因则为候选基因集。

4) 保留基因集的基因组最小化,利用单基因敲除函数找出必需反应部分可以敲除的基因,通过组合删除方法进行筛选,找到必需反应部分对应的最小基因组。

5) 候选基因集的基因组最小化,在 4 的基础上,同样利用单基因敲除函数找出可以被敲除的基因,因为非必需基因数目较多,但由于 4 步骤中对应必需反应的非必需基因已经在前一步进行了单独处理,大部分必需基因得到保留,因而在做组合删除的时候运算量大大减少。最后得到的新模型就是该物种模型的基因组最小化代谢网络模型。

## 2 结果与分析

### 2.1 由代谢网络分析进行大肠杆菌基因组最小化

大肠杆菌是目前了解得最清楚、分析得最透彻的微生物之一,是遗传学研究、生化研究和代谢模拟研究的重要平台,大肠杆菌作为遗传重组领域广泛应用的宿主之一,在医药和发酵工业中可用于生产重组蛋白、氨基酸和其他化学品。因为大肠杆菌在动物肠道和环境中进化其部分基因组对某些用途是不需要的,甚至可能是有害的,因此,通过删除尽可能多的基因片段,构建遗传稳定的菌株,可使大肠杆菌具有健壮的代谢行为,同时也可以添加具有实际用途的基因。

目前大肠杆菌的基因组尺度代谢网络模型共有 4 个,包括 *iJE660*<sup>[20]</sup>、*iJR904*<sup>[21]</sup>、*iAF1260*<sup>[22]</sup> 和 *iJO1366*<sup>[23]</sup>, *iAF1260* 在必需基因预测及与实验结果的拟合度上准确率较好,所以本文采用 *iAF1260* 模型作为出发模型,培养条件设定为最小培养基。葡萄糖为唯一碳源,其通量设定为  $-8 \text{ mmol g}(\text{DW}\cdot\text{h})$ 。最优生物合成速率为  $0.7367 \text{ mmol g}(\text{DW}\cdot\text{h})$ 。

通过前文方法部分介绍的步骤,我们对基因组尺度代谢网络模型分步骤进行简化,各步骤结果(图 2)如下:

1) 我们首先删除模型中的经过 FVA 分析后的 0 通量反应,在 2 382 个反应中共有 1 535 个反应在 FVA 结果中的代谢通量为 0,最后将 0 通量反应删除后,得到 847 个反应,对应 731 个基因。我们将由这些反应构成的新模型称为 *model1*,将其转化为新的 SBML 文件(数据未显示),以便于进行下一步的计算。

2) 优化网络, 去除仅催化一个或一类反应的同功酶基因, 这样的基因共有 100 个, 因而得到的模型 (model2) 包含 847 个反应, 对应 631 个基因, 将其重新转化为新的 SBML 文件 (数据未显示), 以便于进行下一步计算。

3) 通过对 FVA 结果的分析, model2 模型中必需反应部分包含 366 个反应, 对应有 305 个基因, 利用单基因敲除函数, 找出非必需基因 40 个 (具体数据作为附件, 可在网络版中下载。下文简称“附件”), 对这 40 个基因进行逐步递减组合模拟删除, 最后确定其中 26 个基因 (附件) 敲除后对菌体生长没有影响, 这部分基因敲除后整个模型包含 841 个反应, 605 个基因。此时必需反应部分中无任何基因可以被敲除, 规模达到最小。我们将由这些反应构成的新模型称为 (model3), 将其转化为新的 SBML 文件 (数据未显示), 以便于进行下一步的计算。

4) 将 model3 的非必需反应部分进行基因组最小化, 同样利用单基因敲除函数找出非必需基因, 其数目为 304 个 (附件), 对这些非必需基因, 我们采用组合法进行删除, 在保证合成生物量不变的情况下, 依次从里面取出 304 个基因、303 个基因、302 个基因……, 以此类推, 进行模拟删除, 当取出 293 个基因 (附件) 删除的时候得

到一种组合方式, 满足条件, 所以非必需基因保留的最小值就是  $304-293=11$  (个), 最终得到含有 403 个反应, 对应 312 个基因的模型 model4, 将该模型转化为 SBML 文件 (数据未显示)。

5) 将新模型 model4 再次进行单基因敲除实验, 发现可以敲除的基因数目为 0, 这验证了利用这一方法得到的确实是最小基因组。利用该基因组最小化的代谢网络模型, 在葡萄糖输入设定为  $-8 \text{ mmol g}/(\text{DW}\cdot\text{h})$  时求得最优生物质合成速率为  $0.7367 \text{ mmol g}/(\text{DW}\cdot\text{h})$ , 与原始模型相同, 表明基因组精简后, 菌体生长不受影响。整个模型简化过程中各个步骤的结果如表 1、图 2、图 3 及附件所示。

从表 1、图 3 可以看出在简化基因组的过程中, FVA 分析和非必需反应部分删除的基因和反应占有删除的基因和反应数目的大部分, 这符合生物学的特征, 最后整个基因组缩小为原来的 24.7%, 这也为湿实验寻求敲除目标基因部分提供参考。

## 2.2 最小基因组代谢网络模型途径分析

为了进一步了解得到的基因组最小化代谢网络模型的特征, 我们对模型中反应的代谢途径分布进行了分析。模型中的 403 个反应分布在 32 个代谢途径中, 具体分布情况如表 2 所示。

表 1 基于代谢网络模型的大肠杆菌基因组最小化结果

Table 1 Minimization of *E. coli* genome based on metabolic network analysis

	iAF1260	FVA screening	Deletion of same functional genes	Essential reactions screening	Non-essential reactions screening
Reactions	2 382	847	847	841	403
Genes	1 260	731	631	605	312
Metabolite	1 668	595	595	594	407

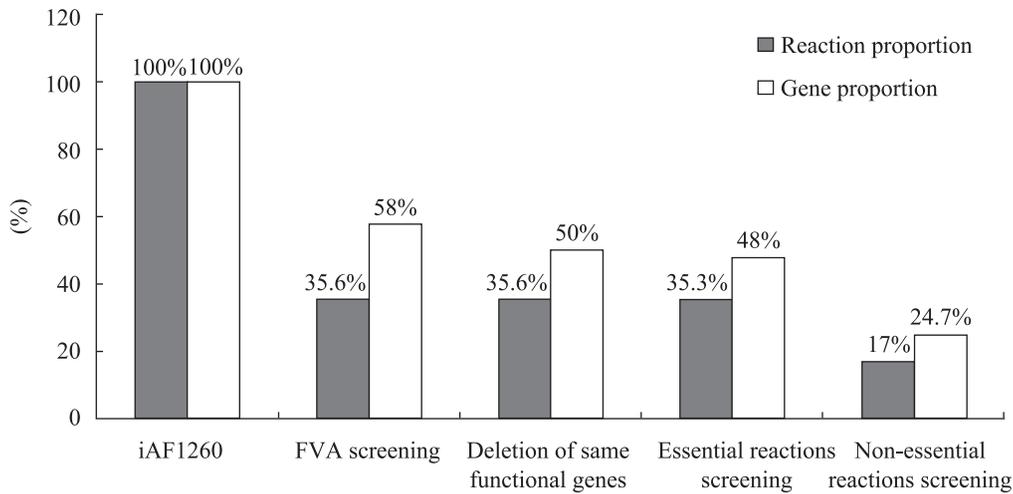


图3 每个步骤后的基因数目和反应数与初始模型比较结果

Fig. 3 Comparison of the gene number and the reactions number with the initial model after every step.

从表2可以看出在基因组最小模型中,氨基酸代谢和脂类代谢包含的反应数最多,分别为112个和80个,这表明大肠杆菌的生长过程中,氨基酸、脂类代谢具有最为重要的作用,这与氨基酸和脂类物质在大肠杆菌细胞组成中占有较高比例相符。另外,辅因子代谢包含77个反应,在整个网络中也具有很大的比重,这表明辅因子是合成生物量不可缺少的重要成分。另外要注意的一点是交换反应在模型同样具有重要的作用,任何基因组最小化的模型都是相对的,当培养基条件发生改变时,模型包含的交换也会发生改变,交换反应越多时,表明外界提供的营养物质越多,则菌体自身合成的物质越少,从而模型所包含的反应数和基因数都会相应的减少。

另外,从碳源的利用上来看,在基因组最小化模型模拟中乳酸的合成途径完全被删除,其他副产物乙酸、琥珀酸合成途径通量也为零,表明底物充分用于生物量的合成,去除了分流到其他副产物而产生的碳源消耗,该结果显示了基因组

最小化代谢网络模型对生物代谢途径的优化。

### 2.3 与湿实验结果的比较

湿实验方面,Posfai等<sup>[5]</sup>通过大片段基因敲除成功将*E. coli* K-12 MG1655基因组减少了15%,构建了MDS40、41、42和43菌株,Kolisnychenko等将大肠杆菌<sup>[24]</sup>基因组减少了8.1%,构建了MDS12菌株,Hashimoto等<sup>[25]</sup>将大肠杆菌基因组减少了29.7%构建了 $\Delta 16$ 菌株,但其类核组织表现异常,生长速率较低,在M9培养基上无法生长。同时Mizoguchi等<sup>[26]</sup>通过无痕敲除方法删除了103个候选区域,共删除1080个基因,有84个删除突变株的生长行为与野生株类似,将这些删除累积起来构建了一个基因组减小菌株——MGF-01。Hirokawa等<sup>[27]</sup>在MGF-01的基础上继续删减得到结构稳定,没有任何营养缺陷性的,可以更高效生长的菌株——DGF-327和DGF-298。我们综合考虑选用MDS43和DGF-298菌株与模型相比较,因为这两种菌株的生产速率与野生型类似,且稳定性强,电转化效率高。

表 2 大肠杆菌基因组最小化代谢网络模型的代谢途径分布

Table 2 Metabolism pathways of the minimized metabolic network in *E. coli*

Subsystem	Metabolic system	Number of reactions
Tyrosine tryptophan and phenylalanine metabolism	Amino acid metabolism	18
Transport outer membrane porin	Amino acid metabolism	18
Valine leucine and isoleucine metabolism	Amino acid metabolism	15
Threonine and lysine metabolism	Amino acid metabolism	12
Arginine and proline metabolism	Amino acid metabolism	12
Histidine metabolism	Amino acid metabolism	10
Methionine metabolism	Amino acid metabolism	8
Cysteine metabolism	Amino acid metabolism	7
Glycine and serine metabolism	Amino acid metabolism	4
Alanine and aspartate metabolism	Amino acid metabolism	4
Murein biosynthesis	Amino acid metabolism	2
Glutamate metabolism	Amino acid metabolism	2
Cell envelope biosynthesis	Lipids metabolism	55
Lipopolysaccharide biosynthesis recycling	Lipids metabolism	11
Glycerophospholipid metabolism	Lipids metabolism	10
Membrane lipid metabolism	Lipids metabolism	4
Cofactor and prosthetic group biosynthesis	Cofactor	72
Folate metabolism	Cofactor	5
Purine and pyrimidine biosynthesis	Nucleic acid metabolism	22
Nucleotide salvage pathway	Nucleic acid metabolism	21
Exchange	Exchange reactions	22
Citric acid cycle	Energy metabolism	8
Oxidative phosphorylation	Energy metabolism	6
Anaplerotic reactions	Energy metabolism	2
Glycolysis gluconeogenesis	Carbohydrate metabolism	10
Pentose phosphate pathway	Carbohydrate metabolism	8
Alternate carbon metabolism	Carbohydrate metabolism	4
Pyruvate metabolism	Carbohydrate metabolism	1
Inorganic ion transport and metabolism	Transport reactions	15
Transport inner membrane	Transport reactions	9
Transport outer membrane	Transport reactions	2
Unassigned		4

Posfai 等<sup>[5]</sup>通过构建的基因组缩减菌株 MDS43 共敲除了 743 个基因, *iAF1260* 模型中包含其中的 82 个基因, 这 82 个基因在基因组最小化模型并没有被全部去除 (附件), 其中 5 个基因仍然保留在模型中, 这 5 个基因是 b0339 (*cynt*)、b2976 (*glcb*)、b2979 (*glcd*)、b4467 (*glcf*) 和 b4468 (*glce*)。b0339、b2976 两个基因在原有模型当中都具有同功酶, 即处于“or”的关系, 由于组合删除的随机性, 导致删除了与其同功能的基因, 而保留了这两个基因, 这一点对于菌体生长并没有影响。而剩下的 3 个基因 (b2979、b4467 和 b4468) 构成一酶复合物催化反应 GLYCTO2 (Glycolate+Q8=Glyoxylate+Q8H2)。该反应可以将代谢网络中某些反应过程生成的乙醇酸转化为乙醛酸, 进而通过乙醛酸循环转化为生物质合成需要的前体。我们注意到大肠杆菌代谢网络模型中还有另外两个基因 (b1033、b3553) 催化乙醇酸和乙醛酸间的转化但所用辅酶为 NADH 或 NADPH (R\_GLYCLTD: Glycolate + NAD(P) $\rightleftharpoons$  Glyoxylate+NAD (P)H2), 但该反应被定义为不可逆且反应方向为从乙醛酸到乙醇酸。当我们将该反应改为可逆后进行基因组最小化计算得到的最小基因组中就不包括 b2979、b4467 和 b4468 而是只包括 b1033 或 b3553, 基因个数也从 312 个减少到了 310 个。这表明大肠杆菌中很可能反应 R\_GLYCLTD 也是可逆的, 因此敲除 b2979、b4467 和 b4468 这 3 个基因对生长没有影响。

我们还将得到的最小基因组与 Hirokawa 等<sup>[27]</sup>构建的 DGM-298 菌株进行了比较, DGF-298 一共删除了 1 429 个基因, 其中 *iAF1260* 模型中包含被实验证明可被删除的基因数为 298 个 (附件), 同样这 298 个基因在基因组最小化模

型中并没有被全部去除, 其中有 12 个基因仍保留在最后的模型当中, 分别是 b0149 (*mrcb*)、b0339 (*cynt*)、b0381 (*ddlA*)、b0521 (*ybcf*)、b1415 (*alda*)、b2097 (*fbab*)、b2708 (*gutq*)、b2797 (*sdab*)、b2979 (*glcd*)、b3744 (*asna*)、b4094 (*phnn*)、b4467 (*glcf*) 和 b4468 (*glce*) (附件)。其中 b0149、b0339、b0381、b4383、b0521、b2097、b2708 和 b2797 七个基因是因为同功能基因, 即“or”关系的选择而被保留在模型中。b2979、b4467 和 b4468 三个基因在上一部分已经说明是因为模型中一些反应的可逆性设置的问题。b1415 和 b3744 两个基因所催化的反应分别是  $\text{gcald}+\text{h}_2\text{o}+\text{nad} \rightleftharpoons \text{glyclt}+\text{nadh}$  和  $\text{asp-L}+\text{atp}+\text{nh}_4 \rightleftharpoons \text{amp}+\text{asn-L}+\text{ppi}$ , 这两个反应在模型中以葡萄糖为唯一碳源时为必需反应。而 Hirokawa 等采用的是富培养基, 可以从培养基中直接得到天门冬酰胺等而不需要从葡萄糖生成, 因此在其湿实验中可以敲除而在我们的模拟计算中保留在最小基因组中。

从上面的结果可以看出, 当需要敲除某些基因的时候, 可以利用得到的最小基因组模型反馈于湿实验, 确定哪些基因可以尝试敲除, 哪些途径可以得到保留, 模拟的结果为湿实验的设计提供了有价值的参考。

### 3 结论与展望

现代生物技术的高速发展, 特别是组学技术的兴起, 高通量数据的迅速积累, 促进了生物代谢网络模型的快速发展, 使网络构建的理论逐渐完善, 质量大幅度提升, 应用范围同样也显著扩大, 在此基础上, 本文第一次提出了基于基因组尺度代谢网络模型, 通过模型简化构建基因组最

小化代谢网络模型的方法,该方法通过零通量反应删除和对非必需基因的组合删除模拟计算,确定能够计算的基因组最小化代谢网络模型。该方法依托于超级计算机的计算能力,运用 Matlab 中的 COBRA 工具包,以大肠杆菌的 *iAF1260* 模型为例,将模型从最初的 1 260 个基因缩减到 312 个基因,模型简化后生物量合成速率保持不变。通过该方法得到的基因组最小化代谢网络模型中氨基酸代谢及脂类代谢仍为生物体生长的主要需求,符合生物代谢规律,证明了基于非必需基因组合的删除方法的合理性。

通过对基因组尺度代谢网络模型的精简,我们可以确定对于生物生长最为重要的反应和基因,并且得到相应代谢途径、代谢系统的生物信息,通过计算机模拟得到的结果可以为湿实验的方案设计提供重要的参考。由于该方法是在基因组尺度代谢网络模型的基础上实施的,因而高质量的模型对于最终得到的基因组最小化代谢网络模型非常重要,基因组最小化代谢网络模型对于今后专一生产某种代谢物的途径设计也将提供更好思路。另外,随着网络模型的发展,基因组最小化代谢网络模型中也可能会结合调控、信号转导等信息,使基因组最小化的生物网络向更加趋近于最小生命体的方向发展。

## REFERENCES

- [1] Gibson DG, Glass JI, Lartigue C, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, 2010, 329(5987): 52–56.
- [2] Fraser CM, Gocayne JD, White O, et al. The minimal gene complement of *Mycoplasma genitalium*. *Science*, 1995, 270(5235): 397–403.
- [3] Lartigue C, Glass JI, Alperovich N, et al. Genome transplantation in bacteria: changing one species to another. *Science*, 2007, 317(5838): 632–638.
- [4] Gibson DG, Benders GA, Andrews-Pfannkoch C, et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science*, 2008, 319(5867): 1215–1220.
- [5] Posfai G, Plunkett G, 3rd, Feher T, et al. Emergent properties of reduced-genome *Escherichia coli*. *Science*, 2006, 312(5776): 1044–1046.
- [6] Juhas M, Eberl L, Glass JI. Essence of life: essential genes of minimal genomes. *Trends Cell Biol*, 2011, 21(10): 562–568.
- [7] Qiu DR. Essential genes, minimal genome and synthetic cell of bacteria: a review. *Chin J Biotech*, 2012, 28(5): 540–549 (in Chinese).  
邱东茹. 细菌必需基因、最小基因组和合成细胞. *生物工程学报*, 2012, 28(5): 540–549.
- [8] Durot M, Bourguignon PY, Schachter V. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiol Rev*, 2009, 33(1): 164–190.
- [9] Thiele I, Palsson BO. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc*, 2010, 5(1): 93–121.
- [10] Feist AM, Herrgard MJ, Thiele I, et al. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol*, 2009, 7(2): 129–143.
- [11] Park JM, Kim TY, Lee SY. Constraints-based genome-scale metabolic simulation for systems metabolic engineering. *Biotechnol Adv*, 2009, 27(6): 979–988.
- [12] Hao T, Ma HW, Zhao XM. Progress in the automatic reconstruction of genome-scale metabolic network. *Chin J Biotech*, 2012, 28 (6): 661–670 (in Chinese).  
郝彤, 马红武, 赵学明. 基因组尺度代谢网络自动重构及分析工具研究进展. *生物工程学报*, 2012, 28(6): 661–670.
- [13] Wang H, Ma HW, Zhao XM. Progress in genome-scale metabolic network: a review. *Chin J*

- Biotech, 2010, 26(10): 1340–1348 (in Chinese).  
王晖, 马红武, 赵学明. 基因组尺度代谢网络研究进展. 生物工程学报, 2010, 26(10): 1340–1348.
- [14] Yizhak K, Tuller T, Papp B, et al. Metabolic modeling of endosymbiont genome reduction on a temporal scale. *Mol Syst Biol*, 2011, 7: 479.
- [15] Pal C, Papp B, Lercher MJ, et al. Chance and necessity in the evolution of minimal metabolic networks. *Nature*, 2006, 440(7084): 667–670.
- [16] Orth JD, Thiele I, Palsson BO. What is flux balance analysis? *Nat Biotechnol*, 2010, 28(3): 245–248.
- [17] Edwards JS, Palsson BO. Metabolic flux balance analysis and the in silico analysis of *Escherichia coli* K-12 gene deletions. *BMC Bioinformatics*, 2000, 1: 1.
- [18] Schellenberger J, Que R, Fleming RM, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc*, 2011, 6(9): 1290–1307.
- [19] Muller AC, Bockmayr A. Fast thermodynamically constrained flux variability analysis. *Bioinformatics*, 2013, 29(7): 903–909.
- [20] 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.
- [21] 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): R54.
- [22] 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 Systems Biol*, 2007, 3(121): 1–18.
- [23] Orth JD, Conrad TM, Na J, et al. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Mol Systems Biol*, 2011, 7(535): 1–9.
- [24] Kolisnychenko V, Plunkett G 3rd, Herring CD, et al. Engineering a reduced *Escherichia coli* genome. *Genome Res*, 2002, 12(4): 640–647.
- [25] Hashimoto M, Ichimura T, Mizoguchi H, et al. Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. *Mol Microbiol*, 2005, 55(1): 137–149.
- [26] Mizoguchi H, Mori H, Fujio T. *Escherichia coli* minimum genome factory. *Biotechnol Appl Biochem*, 2007, 46(Pt 3): 157–167.
- [27] Hirokawa Y, Kawano H, Tanaka-Masuda K, et al. Genetic manipulations restored the growth fitness of reduced-genome *Escherichia coli*. *J Biosci Bioeng*, 2013, 1(10): 1–7.

(本文责编 郝丽芳)