

doi:10.3969/j.issn.1672-5565.2014.01.10

基于约束建模法的结核菌 H37Rv 代谢网络分析

杜志成¹, 关鹏¹, 黄德生^{2*}

(1. 中国医科大学公共卫生学院流行病学教研室, 沈阳 110001;

2. 中国医科大学基础医学院数学教研室, 沈阳 110001)

摘要: 基于结核分枝杆菌国际标准强毒株 H37Rv 菌株的基因组尺度代谢网络模型 iNJ661 进行分析, 以寻找代谢网络中培养基的关键成分和必要基因。该研究在 Matlab 平台上利用 COBRA 工具箱, 采用基于约束的建模方法进行动态生长模拟、解空间抽样在酶活性水平上的具体化和基因删除模拟实验。结果发现培养基成分中铵盐、三价铁盐、磷酸盐、硫酸盐、甘油等可影响 H37Rv 的生长; 培养基中去除磷酸盐后十种酶均在不同程度上受到抑制, 其中丙糖磷酸异构酶、3-磷酸甘油醛脱氢酶、磷酸甘油酸变位酶、烯醇酶受限明显。通过基因删除得出 188 个必要基因以及非必要基因中的 16 个致死基因对。基于约束建模分析可初步了解结核杆菌 H37Rv 菌株代谢网络的性质, 可为后续相关研究提供参考和借鉴。

关键词: 代谢网络; 系统生物学; 建模; 结核杆菌

中图分类号: R378.91+1; TP319 **文献标志码:** A **文章编号:** 1672-5565(2014)-01-060-05

Metabolic network analysis of Mycobacterium tuberculosis H37Rv with constraint-based models

DU Zhicheng¹, GUAN Peng¹, HUANG Desheng^{2*}

(1. Department of Epidemiology, School of Public Health, China Medical University, Shenyang 110001, China;

2. Department of Mathematics, College of Basic Medical Sciences, China Medical University, Shenyang 110001, China)

Abstract: To explore the key medium components and essential genes of tuberculosis metabolic network, we used the genome-scale metabolic network model iNJ661 to analyze Mycobacterium tuberculosis H37Rv. The study was conducted on the COBRA toolbox for Matlab. Simulation of dynamic growth, solution space sampling at the level the enzyme activity and gene deletion analysis were performed with the help of constraint-based modeling methods. Ammonium, ferric iron, phosphate, sulfate and glycerol of the medium components could affect the growth of H37Rv. After removing phosphate from the medium, ten enzymes were inhibited in varying degrees, among which triosephosphate isomerase, glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate mutase and enolase were constrained significantly. 188 essential genes and 16 pairs of synthetic lethal genes among nonessential genes were obtained by simulating deletion studies. Constraint-based modeling and analysis can now be used for the preliminary understanding of the nature of Mycobacterium tuberculosis H37Rv strain metabolic network and could provide the reference for future related research.

Keywords: Metabolic network; System biology; Modeling; Mycobacterium tuberculosis

世界卫生组织的发布的《2012 年全球结核病报告》显示, 在 2011 年里, 有 870 万新发病例, 估计有 140 万人死于结核病, 其中包括 50 万名妇女, 该病

已经成为全世界妇女的主要杀手之一^[1]。多重药物耐受结核病和广泛药物耐受结核病的出现^[2]提示新的抗结核病药物靶标相关研究迫在眉睫。

收稿日期: 2013-08-30; 修回日期: 2013-10-20.

基金项目: 国家自然科学基金资助项目 (71073175)。

作者简介: 杜志成, 男, 在读硕士, 研究方向: 数据挖掘; E-mail: dgdzc@hotmail.com.

* 通信作者: 黄德生, 教授, 博士生导师, 研究方向: 人工智能和生物信息学; E-mail: dshuang@mail.cmu.edu.cn.

随着测序技术和高通量生物技术的迅速发展,大量的生物体已经完成全基因组测序、基因功能注释。研究人员结合生化、生理数据,系统、合理、有效地重建了多种生物体的基因组尺度代谢网络(Genome-scale metabolic network, GSMN)(包括金黄色葡萄球菌 iSB619、结核分枝杆菌 iNJ661、巴克红曲霉 iAF692 和绿脓杆菌 iJN746 等),从而得以在计算机上进行相关的模拟预测以供实验室和临床研究参考^[3-4]。本研究采用基于约束的建模方法对结核分枝杆菌的基因组尺度代谢网络 iNJ661^[5]进行重构和仿真分析以进一步探索该代谢网络的属性。

1 材料与方法

1.1 数据获取

结核分枝杆菌国际标准强毒株 H37Rv 菌株的 GSMN 模型 iNJ661 于 2013 年 4 月 18 日从 BIGG (<http://bigg.ucsd.edu/>) 下载,格式为 Systems Biology Markup Language (SBML)^[6]。它包含了 661 个基因、540 个基因编码蛋白、826 个代谢物和 1 025 个代谢反应(其中 939 个发生在细胞内的)。同时从 KEGG (<http://www.genome.jp/kegg>)、PubMed (www.ncbi.nlm.nih.gov/pubmed)、TubercuList (<http://genolist.pasteur.fr/TubercuList>) 获得详细的关于结核杆菌 H37Rv 菌株的反应、蛋白、基因组等多源化信息。

1.2 COBRA 软件

约束的重建和分析软件 COBRA^[7](Constraint-based reconstruction and analysis)是在美国加利福尼亚大学圣地亚哥分校 Palsson 和 Herrgard 领导下开发的,它是一个运行在 Matlab (Mathworks Inc., Natick, MA, USA) 环境下专门用于代谢网络基于约束的模拟分析工具,主要功能包括:通量平衡分析(Flux balance analysis, FBA)、通量平衡几何分析、必要基因分析法、最小代谢调整分析、碳十三追踪、差异填充(Gap Filling)、代谢工程和计算机模拟可视化。该软件的最新版本可以从 Palsson 小组的主页免费下载(<http://systemsbiology.ucsd.edu/>)。

1.3 通量平衡分析

FBA^[8-9]是最基本的一种约束建模分析方法,

可以计算细胞的生长情况。它可以使用化学计量矩阵(S)进行稳态分析($S_v=0$)处理一些系统性问题,其中 S 是代谢网络的化学计量矩阵,矩阵中的每一行表示一个代谢物,每一栏表示一个代谢反应,每个元素表示该代谢物在该代谢反应中的化学计量系数。当未施加约束条件时,代谢网络的通量可分布于解空间内的任意点,若设定反应边界值作为约束条件后,即可得到一个有限的解空间。最后给定某目标函数,用线性规划的方法寻找其最优解。本研究进行的最优生长率、模拟动态生长情况、基因删除、功能都通过 FBA 过程实现。FBA 过程只返回最优解,从而使有关整个解决方案的空间未能体现。为解决该问题,COBRA 软件使用抽样对解空间进行描述。

1.4 OptKnock 过程^[10]

类似“生物工程”,COBRA 软件通过对整个代谢网络进行分析,设定某个代谢物为目标产物和允许删除反应的数量为约束条件,获得为使目标产物产量最大,何种或哪些特定反应可删除的解决方案。

2 结果

2.1 结核杆菌 H37Rv 菌株的 GSMN 模型 iNJ661 描述

模型中每个代谢反应均有一个范围从 0 到 4 的得分来反映可用的信息和证据,其中 4 分为最高,代表有实验的生化数据来支持。一个模型反应集合的平均得分水平代表该模型的真实性和 iNJ661 的平均得分为 2.31 分。该模型基因-蛋白质-反应(Gene-protein-reaction, GPR)相互联系的比例占 76.76%,表明该模型完整性较好。应用 Gap Filling^[11]过程对该模型进行检测,仅发现 37 个不能被生成的代谢物和 48 个缺乏所需反应物的代谢物,这将有待于研究人员进行完善。

2.2 iNJ661 的生长表型

iNJ661 模型提供了 87 个可供调整的培养基成分,其预设的培养基为 Middlebrook 7H9。结核杆菌生长缓慢且难于培养,选择或配制一种促进结核杆菌生长的培养基显得尤其重要。模拟菌株在 Middlebrook 7H9^[12]、Youmans^[13]、CAMR^[14] 三种培养基生长,培养基成分和对应的生长率见表 1。

表1 Middlebrook 7H9、Youmans 和 CAMR 培养基成分和生长率

Table 1 Composition and biomass accumulation of Middlebrook 7H9, Youmans and CAMR

Middlebrook 7H9	Youmans	CAMR	
ammonium	oxygen	L-alanine	
calcium	L-asparagine	L-arginine	
chloride	citrate	L-asparagine	
citrate	glycerol	L-aspartic acid	
copper	water	L-glutamic acid	
ferric iron	octadecanoate (Tween)	L-glycine	
L-glutamate	phosphate	L-isoleucine	
magnesium	sulfate	L-leucine	
oxygen	magnesium	L-serine	
phosphate	potassium	L-phenylalanine	
sodium	bicarbonate	pyruvate	
sulfate	ferric iron (minimal amount)	glycerol	
water		octadecanoate (Tween)	
D-glucose			
glycerol			
Biomass Accumulation	0.052 2	0.031 4	0.058 0

分别向模型中加入不同氨基酸和碳源来观察其对菌株生长率的影响,结果显示在加入丝氨酸(Ser)、天冬氨酸(Asn)、谷氨酸(Glu)足够让菌株快速生长;以麦芽糖(Maltose)和海藻糖(Trehalose)作为碳源加入到培养基中,菌株生长最快(图1和图2)。

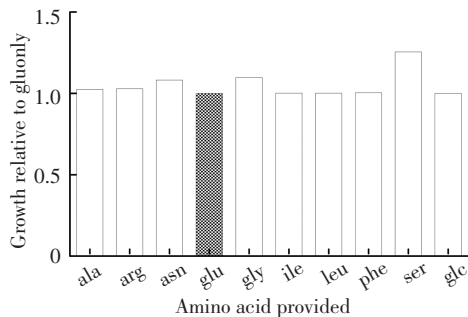


图1 单独加入不同氨基酸时 iNJ661 的生长率

Fig. 1 Biomass accumulation of iNJ661 model when different amino acids added singly

注:所有的数值都以谷氨酸(Glu)作为1进行了标化。

Notes: All values have been standardized to glutamic acid as 1.

同理,通过改变预设培养基中每个成分来观察相应的生长率,结果显示去掉铵盐(Ammonium)、三价铁盐(Ferric iron)、磷酸盐(Phosphate)、硫酸盐(Sulfate)、甘油(Glycerol),iNJ661生长将会受到阻碍甚至无法生长(表2),这提示以上5种代谢物的转运体和激酶都可能成为药物的靶标,达到抑菌甚至杀菌的效果。

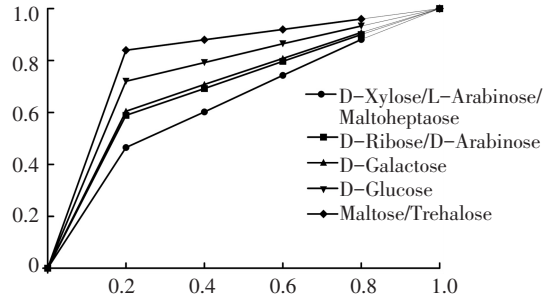


图2 不同的碳源下 iNJ661 的生长率

Fig. 2 Biomass accumulation of iNJ661 model when different carbon sources added

表2 单独改变 Middlebrook 7H9 每一成分对应的 iNJ661 生长率
Table 2 Corresponding biomass accumulation of iNJ661 model with single change in Middlebrook 7H9

Deletion	iNJ661 Growth	
	Biomass Accumulation	
ammonium	0.048 4	
calcium	0.052 2	
chloride	0.052 2	
citrate	0.052 2	
copper	0.052 2	
ferric iron	0	
L-glutamate	0.052 2	
magnesium	0.052 2	
oxygen	0	
phosphate	0	
sodium	0.052 2	
sulfate	0	
water	0.052 2	
D-glucose	0.026 6	
glycerol	0	

2.3 解空间抽样

在模拟 H37Rv 菌株生长表型的步骤中,已经发现从培养基中去掉磷酸盐会使模型的生长受到明显的阻碍,现通过抽样直方图来观察一系列酶的活性(图3)。结果显示十种酶都在不同程度上受到了抑制,其中丙糖磷酸异构酶(Triosephosphate isomerase, TPI)、3-磷酸甘油醛脱氢酶(Glyceraldehyde-3-phosphate dehydrogenase, GAPD)、磷酸甘油酸变位酶(Phosphoglycerate mutase, PGM)、烯醇酶(Enolase, ENO)受限明显。

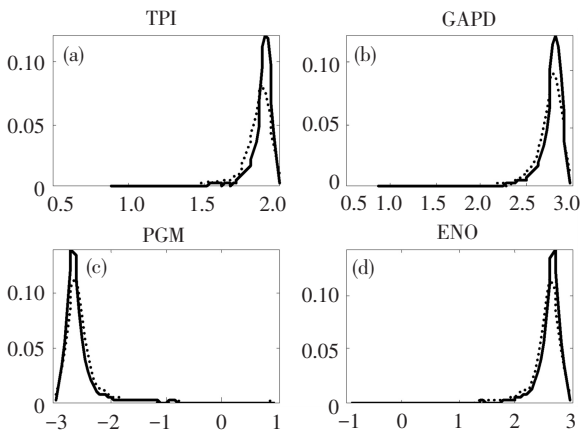


图 3 培养基中去掉磷酸盐前后不同反应通量下蛋白酶浓度的抽样直方图

Fig. 3 Sample histogram of proteinase concentration under different reaction flux in medium with phosphate vs. without phosphate

注: 虚线指示从培养基中去掉磷酸盐。

Notes: The dotted lines indicate phosphate is removed from the medium.

2.4 基因删除

随着多重耐药菌株的不断出现, 识别更多靶标迫在眉睫。通过 FBA 过程完成 iNJ661 的基因删除学习, 结果显示模型包含 188 个必要基因, 其中 *ilvD* (Singh V, et al. 2011)、*rmlA* (Qu H, et al. 2007)、*inhA* (Banerjee A, et al. 1994)、*kasA* (Lee W, et al. 2011)、*embC/embA/embB* (Goude R, et al. 2008/ Amin A, et al. 2008/ Li W, et al. 2010) 等已被研究证实。另外在非必要基因的双致死性基因学习中获得 16 个基因对 (表 3)。

表 3 iNJ661 模型非必要基因中 16 个致死基因对

Table 3 16 pairs of non-essential genes in iNJ661 model whose simultaneous deletion is lethal

ID	Name	ID	Name
1	Rv1373 Rv0295c	9	fabD fabD2
2	tal rpe	10	plsC Rv2182c
3	fabG1 fcoT	11	suhB impA
4	gmdA gea	12	pgsA3 pgsA2
5	ilvA thrC	13	thyX thyA
6	ilvA thrB	14	pyrH cmk
7	argG arcA	15	accE5 accD3
8	argH arcA	16	hpt deoD

2.5 OptKnock 过程

研究发现结核杆菌的致病性主要与占细胞壁干

重 60% 以上的脂质成分 (如 Phthiocerol dimycocerosate, PDIM) 密切相关^[15]。本文应用 OptKnock 过程, 将原有的培养基葡萄糖供给调整为 $10 \text{ mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$, 分别将目标设定为 PPDIM (Phenol phthiocerol dimycocerosate) 和 PDIM 以求其最大产量 (表 4), 结果显示抑制 ACGS 一个反应可以让 PPDIM 产量最大化, 抑制 ACDS、GLCS1、ICL 三个反应可让 PDIM 产量最大。通过以上模拟实验, 提示可能短时间获得更多的 PPDIM 和 PDIM, 以进行其致病和毒力损伤机制研究, 不仅加深对病原体-宿主互作网络中的理解, 还有望为结核病的治疗带来新的希望与突破。

表 4 OptKnock 过程

Table 4 OptKnock process model whose simultaneous deletion is lethal

Metabolites	Knockouts reactions	Enzyme	Max ν biomass	
			Growth (h^{-1})	Production ($\text{mmol} \cdot \text{h}^{-1}$)
Wild	"Completed Network"		0.052 2	
PPDIM	ACGS	N-acetylglutamate synthase	0.052 2	0.413 5
	ACGS	N-acetylglutamate synthase		
PDIM	GLCS1	glycogen synthase		
	ICL	Isocitrate lyase	0.052 2	0.445 9

3 讨论

本研究通过仿真分析发现 iNJ661 模型的动态生长情况与 H37Rv 菌株的实验数据相一致。以丝氨酸、天冬氨酸、谷氨酸作为氨基酸来源, 以麦芽糖、海藻糖作为碳源, 菌株增长明显。去除培养基中的铵盐、三价铁盐、磷酸盐、硫酸盐和甘油, 菌株的生长将受限。解空间抽样分析在蛋白酶水平解释关键的营养目标 (如磷酸盐) 对菌株生长的影响, 后续研究若设计相关的酶抑制剂很有可能达到抑菌甚至杀菌的效果。

基因删除学习中获得的部分必要基因与 Sasseti 等^[16] 在体的转座子定点杂交实验结果相符, 并与目前 H37Rv 部分的药物靶标相一致, 若结合与人类基因的同源性分析, 研究发现的必要基因和基因对有可能为新的抗结核病药物筛选提供提示和借鉴。OptKnock 过程给出量产致病物质 PDIM 的代谢工程途径, 进而模拟其他重要物质量产机制, 指导疫苗开发和病原微生物研究人员的工作。需要注

意的是建模进行的实验是基于现有代谢网络模型的,其阳性结果很有可能属存伪。

本研究简要提供了对结核杆菌 H37Rv 菌株进行约束建模分析的基本分析框架,可为后续相关研究提供参考和借鉴。

参考文献(References)

- [1] DIAS H M, FALZON D, FITZPATRICK C, et al. Global tuberculosis report 2012 [M]. Geneva: WHO Press, 2012, 8-28.
- [2] HINGLEY-WILSON S M, CASEY R, CONNELL D, et al. Undetected multidrug-resistant tuberculosis amplified by first-line therapy in mixed infection [J]. *Emerging Infectious diseases*, 2013, 19(7):1138-1141.
- [3] DUROT M, BOURGUIGNON P Y, SCHACHTER V. Genome-scale models of bacterial metabolism: reconstruction and applications [J]. *FEMS Microbiology Reviews*, 2009, 33(1):164-190.
- [4] SCHELLENBERGER J, PARK J O, CONRAD T M, et al. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions [J]. *BMC Bioinformatics*, 2010, 11:213.
- [5] FANG X, WALLQVIST A, REIFMAN J. Modeling phenotypic metabolic adaptations of *Mycobacterium tuberculosis* H37Rv under hypoxia [J]. *PLoS Computational Biology*, 2012, 8(9):e1002688.
- [6] HUCKA M, FINNEY A, SAURO H M, et al. SBML Forum. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models [J]. *Bioinformatics*, 2003, 19(4):524-531.
- [7] SCHELLENBERGER J, QUE R, FLEMING R M, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0 [J]. *Nature Protocols*, 2011, 6(9):1290-1307.
- [8] METRIS A, GEORGE S, BARANYI J. Modelling osmotic stress by Flux Balance Analysis at the genomic scale [J]. *International Journal of Food Microbiology*, 2012, 152(3):123-128.
- [9] RAMAN K, CHANDRA N. FLUX balance analysis of biological systems: applications and challenges [J]. *Briefings in Bioinformatics*, 2009, 10(4):435-449.
- [10] BURGARD A P, PHARKYA P, MARANAS C D. OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization [J]. *Biotechnology and Bioengineering*, 2003, 84(6):647-657.
- [11] ORTH J D, PALSSON B Ø. Gap-filling analysis of the iJO1366 *Escherichia coli* metabolic network reconstruction for discovery of metabolic functions [J]. *BMC Systems Biology*, 2012, 6:30.
- [12] PARK Y K, KOH W J, KIM S O, et al. Clarithromycin susceptibility testing of *Mycobacterium avium* complex using 2, 3-diphenyl-5-thienyl-(2)-tetrazolium chloride microplate assay with Middlebrook 7H9 broth [J]. *Journal of Korean Medical Science*, 2009, 24(3):511-512.
- [13] GADRE D V, TALWAR V. In vitro susceptibility testing of *Mycobacterium tuberculosis* strains to trifluoperazine [J]. *Journal of Chemotherapy*, 1999, 11(3):203-206.
- [14] JAMES B W, WILLIAMS A, MARSH P D. The physiology and pathogenicity of *Mycobacterium tuberculosis* grown under controlled conditions in a defined medium [J]. *Journal of Applied Microbiology*, 2000, 88(4):669-677.
- [15] DOMENECH P, REED M B. Rapid and spontaneous loss of phthiocerol dimycocerosate (PDIM) from *Mycobacterium tuberculosis* grown in vitro: implications for virulence studies [J]. *Microbiology*, 2009, 155 (Pt 11):3532-3543.
- [16] SASSETTI C M, RUBIN E J. Genetic requirements for mycobacterial survival during infection [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2003, 100(22):12989-12994.