

DOI:10.12113/j.issn.1672-5565.201812003

结核分枝杆菌 Rv1385 的抗原表位预测分析

黄劲^{1,2}, 王洁³, 高霄², 雷依丽², 杨光红⁴, 冯永红³, 秦莲花^{3*}

(1. 贵州医科大学 环境污染与疾病监控教育部重点实验室, 贵阳 550025; 2. 贵州医科大学 基础医学院, 贵阳 550025;
3. 同济大学 医学院附属上海市肺科医院 上海市结核病重点实验室, 上海 200433;
4. 贵州医科大学 公共卫生学院, 贵阳 550025)

摘要: 结核分枝杆菌生长缓慢, 难以获得足量的天然蛋白抗原。而利用基因工程技术制备重组蛋白抗原, 存在着表达量低、不易纯化, 以及特异性较低等不利因素。随着生物信息技术的发展, 研究者可根据生物信息学软件预测, 选择合成优势抗原肽段, 而不需克隆整个蛋白, 具有简便、经济、特异性高等特点。因此, 本研究应用生物信息学方法, 预测结核分枝杆菌 Rv1385 蛋白的抗原表位并分析其优势表位, 对了解 Rv1385 蛋白的免疫学特性及其与结核分枝杆菌致病的关系具有重要意义。首先, 从 NCBI 数据库下载 Rv1385 的氨基酸序列, 然后采用生物信息学软件 PSIPRED Server 预测蛋白质二级结构; BepiPred 1.0 Server 和 ABCpred 预测该蛋白的 B 细胞抗原表位; BIMAS、SYFPEITHI 及 NetCTLpan 1.1 Server 预测该蛋白的 CTL 表位; SYFPEITHI 和 NetMHCIIpan 3.2 Server 预测该蛋白的 Th 细胞表位。最后综合分析预测结核分枝杆菌 Rv1385 的优势候选抗原表位。结果显示, Rv1385 蛋白二级结构中, α 螺旋占 51.8%, β 折叠占 41.6%, 无规卷曲占 6.6%; 共有 4 个 B 细胞抗原位点, 位于 109~124、152~169、178~192、198~209 氨基酸区段; 3 个 CTL 表位, 位于 263~271、87~95、240~248 氨基酸区段; 5 个 Th 表位, 位于 77~91、87~101、234~247、70~84、46~60 氨基酸区段。生物信息学分析显示 Rv1385 含有潜在的 B 细胞和 T 细胞抗原表位, 为该蛋白抗原表位的进一步研究及应用奠定了基础。

关键词: 结核分枝杆菌; 生物信息学; Rv1385; 抗原性

中图分类号: R378.91+1 文献标志码: A 文章编号: 1672-5565(2019)03-161-06

Prediction and analysis of the distribution of epitopes of Rv1385 from *Mycobacterium tuberculosis*

HUANG Jin^{1,2}, WANG Jie³, GAO Xiao², LEI Yili², YANG Guanghong⁴, FENG Yonghong³, QIN Lianhua^{3*}

(1. Key Laboratory of Environmental Pollution Monitoring and Disease Control of Ministry of Education, Guizhou Medical University, Guiyang 550025, China; 2. School of Basic Medical Sciences, Guizhou Medical University, Guiyang 550025, China; 3. Shanghai Key Laboratory of Tuberculosis, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai 200433, China; 4. School of Public Health, Guizhou Medical University, Guiyang 550025, China)

Abstract: *Mycobacterium tuberculosis* (*M. tuberculosis*) grows very slowly, thus it is difficult to obtain sufficient amount of natural protein antigen. However, the preparation of recombinant protein antigen by genetic engineering technology has some disadvantages, such as low expression, difficulty in purification, and low specificity. With the development of bioinformatics, instead of cloning the whole protein, researchers can select dominant antigen peptides according to the prediction of bioinformatics software, which has the characteristics of simplicity, economy, and specificity. Therefore, this study adopts bioinformatics methods to predict the antigenic epitopes of Rv1385 in *M. tuberculosis* and analyzes its dominant epitopes, which is of great significance for understanding the immunological properties of Rv1385 and its relationship with the pathogenesis of *M. tuberculosis*. First, the amino

收稿日期: 2018-12-21; 修回日期: 2019-03-23.

基金项目: 国家自然科学基金 (No.8147009, No.81771692, No.81471563); 贵州省科技厅基金 (黔科合[2010]3154号); 贵州省教育厅开放课题基金 (黔教合 KY 字[2018]482).

作者简介: 黄劲, 女, 副教授, 研究方向: 病原菌与宿主互作机制. E-mail: huangjin@gmc.edu.cn.

* 通信作者: 秦莲花, 女, 副研究员, 研究方向: 结核分枝杆菌致病机制. Email: lianhuaq1013@163.com.

acid sequence of Rv1385 was downloaded from NCBI database, and then the secondary structure of the protein was predicted by bioinformatics software PSIPRED Server. The B cell epitopes were predicted by BepiPred 1.0 Server and ABCpred, the CTL epitopes were predicted by BIMAS, SYFPEITHI, and NetCTLpan 1.1 Server, and the Th cell epitopes were predicted by SYFPEITHI and NetMHCIIpan 3.2 Server. Finally, the candidate epitopes of Rv1385 were analyzed and predicted. Results showed that in the secondary structure of Rv1385, α -helix, β -sheet, and random coil accounted for 51.8%, 41.6%, and 6.6%, respectively. There were four B cell antigen sites, located in 109–124, 152–169, 178–192, and 198–209 amino acid regions, three CTL epitopes, located in 263–271, 87–95, and 240–248 amino acid regions, and five Th epitopes, located in 77–91, 87–101, 234–247, 70–84, and 46–60 amino acid regions. Bioinformatics analysis showed that Rv1385 contained potential B cell and T cell epitopes, which laid a foundation for the further research and application of the epitopes of Rv1385.

Keywords: *Mycobacterium tuberculosis*; Bioinformatics; Rv1385; Antigenic epitope

结核病是由结核分枝杆菌 (*Mycobacterium tuberculosis*, Mtb) 引起的一种与机体细胞免疫应答密切相关的慢性消耗性传染病, 宿主免疫细胞可通过其表面受体识别结核分枝杆菌抗原蛋白中特殊的化学基团, 即表位 (Epitope) 而刺激机体产生免疫应答, 但其机制和宿主免疫保护机制仍不清楚^[1]。因此, 探索结核分枝杆菌蛋白抗原表位的免疫学特性, 对结核分枝杆菌致病机制的研究和表位疫苗的研发具有重要意义^[2]。越来越多的研究显示, 结核分枝杆菌在体内生长必需的基因及其产物可作为有效的药物研发靶点和候选疫苗^[3]。

核苷酸是结核分枝杆菌生长和分化的必需物质, 主要由从头合成途径产生^[4]。其中, 乳清酸核苷-5'-磷酸脱羧酶 (Orotidine 5'-monophosphate decarboxylase, OMPdecase; Rv1385) 是嘧啶核苷酸从头合成途径最后一步反应的酶, 催化乳清酸核苷-5'-磷酸脱羧生成尿嘧啶核苷酸^[5]。尿嘧啶核苷酸是合成其他嘧啶核苷酸的前体, 对促进结核分枝杆菌的分化和增殖具有重要作用^[6]。

Rv1385 蛋白由结核分枝杆菌 *pyrF* 基因编码, 包含 274 个氨基酸残基, 富含丙氨酸、甘氨酸和缬氨酸等疏水性氨基酸, 其相对分子质量为 27.4 KD。生物信息学分析显示, Rv1385 蛋白不含信号肽, 亚细胞定位可能为细胞膜或细胞浆^[7]。迄今, 对 Rv1385 的抗原性研究尚无报道, 本研究应用生物信息学的方法对 Rv1385 的抗原性进行预测分析, 以为后续深入研究结核分枝杆菌的致病机制和新的药物靶点奠定基础。

1 材料与方法

1.1 氨基酸序列

从 NCBI 数据库检索获得结核分枝杆菌 (*Mycobacterium tuberculosis*, Mtb) H37Rv 及 15 株

M. tuberculosis 菌株的 Rv1385 蛋白氨基酸序列, 序列号分别为: AJF02736、WP_031708180.1、WP_031734789.1、WP_017487696.1、WP_118962198.1、WP_031658710.1、WP_016720430.1、WP_070892765.1、WP_055366474.1、WP_052635252.1、WP_057361138.1、WP_055375288.1、WP_116258713.1、WP_057370066.1、WP_015631297.1 和 WP_019283631.1。

1.2 Rv1385 的二级结构预测

应用在线软件 PSIPRED Server (<http://bioinf.cs.ucl.ac.uk/psipred/>) 预测该蛋白的二级结构。

1.3 Rv1385 的 B 细胞抗原表位预测

应用在线软件 BepiPred 1.0 Server (<http://www.cbs.dtu.dk/services/BepiPred-1.0/>) 和 ABCpred (http://crdd.osdd.net/raghava/abcpred/ABC_submission.html) 预测该蛋白的 B 细胞抗原表位。

1.4 Rv1385 的 CTL 抗原表位预测

应用 BIMAS (https://www-bimas.cit.nih.gov/molbio/hla_bind/)、SYFPEITHI (<http://www.syfpeithi.de/bin/MHCServer.dll/EpitopePrediction.htm>) 及 NetCTLpan 1.1 Server (<http://www.cbs.dtu.dk/services/NetCTLpan/>) 预测该蛋白的 CTL 表位。

1.5 Rv1385 的 Th 抗原表位预测

应用 SYFPEITHI (<http://www.syfpeithi.de/bin/MHCServer.dll/EpitopePrediction.htm>) 和 NetMHCIIpan 3.2 Server (<http://www.cbs.dtu.dk/services/NetMHCIIpan/>) 预测该蛋白的 Th 细胞表位。

1.6 Rv1385 氨基酸序列的同源性预测

应用 Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo>) 对该蛋白的氨基酸序列进行同源性分析。

2 结果与分析

2.1 Rv1385 的二级结构

PSIPRED Server 结果见图 1, 其中 α 螺旋(H)、

β 折叠(E)、无规卷曲(C)分别占 51.8%、41.6%和 6.6%。

2.2 Rv1385 的 B 细胞抗原表位

BepiPred 1.0 Server 预测结果显示 95~101、109~124、152~169、178~192、198~209、218~236、249~260 等氨基酸区段可能具有抗原表位;ABCpred 预测具有抗原表位可能性最大的位点依次为 109~124、94~109、120~135、189~204、195~210、147~162、65~80、168~183 位点。综合两种软件预测结

果的分值(≥0.65),同时结合 Rv1385 蛋白的二级结构预测结果无规卷曲存在的区段进行分析,推测 B 细胞抗原表位较有可能位于 109~124、152~169、178~192、198~209 等氨基酸区段。

2.3 Rv1385 的 CTL 表位

应用 3 种生物信息学软件预测,分值较高的 CTL 表位结果见表 1,综合分析,推测可能性较高的 CTL 表位存在于 OMPdecase 蛋白的下述氨基酸区段:263~271、87~95、240~248。

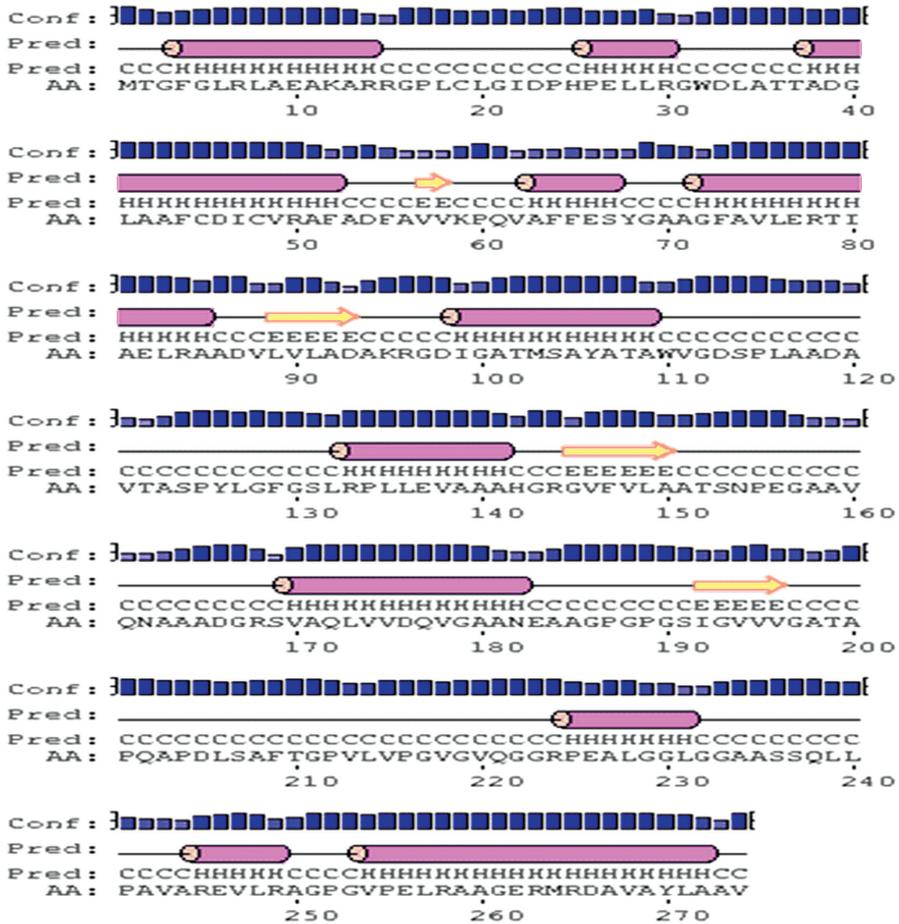


图 1 Rv1385 二级结构预测结果

Fig.1 Predicted secondary structure of Rv1385

表 1 Rv1385CTL 表位预测结果

Table 1 Predicted CTL epitopes of Rv1385

表型分类	BIMAS		SYFPEITHI		NetCTLpan	
	起始位置	氨基酸序列	起始位置	氨基酸序列	起始位置	氨基酸序列
HLA-A2	263	RMRDAVAYL	75	VLERTIAEL	263	RMRDAVAYL
			263	RMRDAVAYL		
HLA-A3	88	VLVLADAKR	51	AFADFAVVK	/	/
	7	RLAEAKARR	87	DVLVLADAK		
HLA-B7	12	KARRGPLCL	200	APQAPDSA	124	SPYLFGSL
	240	LPAVAREVL	240	LPAVAREVL	240	LPAVAREVL

2.4 Rv1385 的 Th 表位

SYFPEITHI 和 NetCTLpan 3.2 Server 在线软件预测,分值较高的 Th 表位结果见表 2,综合 2 种软件

结果推测 Th 表位可能位于 77~91、87~101、234~247、70~84、46~60。

表 2 Rv1385Th 表位预测结果

Table 2 Predicted Th epitopes of Rv1385

表型分类	SYFPEITHI		NetCTLpan	
	起始位置	氨基酸序列	起始位置	氨基酸序列
HLA-DRB1 * 0101	225	PEALGGLGGAASSQL	51	AFADFAVVKPQVAFF
	77	ERTIAELRAADV LVL	77	ERTIAELRAADV LVL
HLA-DRB1 * 0301	17	PLCLGIDPHPELLRG	170	VAQLVVDQVGAANEA
	87	DVLVLADAKRGDIGA	87	DVLVLADAKRGDIGA
HLA-DRB1 * 0401	70	AAGFAVLERTIAELR	142	HGRGVFVLAATSNEP
	17	PLCLGIDPHPELLRG	70	AAGFAVLERTIAELR
HLA-DRB1 * 0701	228	LGGLGGAASSQLLPA	234	AASSQLLPAVAREVL
	96	RGDIGATMSAYATAW		
HLA-DRB1 * 1101	70	AAGFAVLERTIAELR	70	AAGFAVLERTIAELR
HLA-DRB1 * 1501	38	ADGLA AFCDICVRAF	46	DICVRAFADFAVVKP
	46	DICVRAFADFAVVKP		

2.5 Rv1385 的氨基酸序列同源性分析

运用 NCBI 的 Blast 对 *M. tuberculosis* complex 数据库的 101 株菌株的 Rv1385 进行氨基酸序列比对,结果显示一致性为 81.32%~100%。然后从中选择一致性较低的 16 株 *M. tuberculosis* 菌株进行同源性分析,结果见图 2。

3 讨论

结核病仍然是全球性健康性疾病,据 2018 年世界卫生组织的结核病报告报道,目前全世界有近 1/3 人口感染结核分枝杆菌,每年大约有 180 万人死于结核病^[8]。尽管自 1990 年以来,结核病的发病率和新增病例有所下降,但随着多重耐药和广泛耐药结核菌株的增多,以及结核和人类免疫缺陷病毒 (Human Immunodeficiency Virus, HIV) 共同感染患者的出现,给结核病的防治带来巨大挑战^[9]。

根除结核病的一个重要问题是需要长期的药物治疗^[10-11]。其原因之一是结核病的病原菌 (*Mycobacterium tuberculosis*, Mtb) 能够进入一种非复制的持留状态,导致对传统抗结核药物的敏感性降低^[12]。因此,需要新的药物来缩短疗程,同时,急切需要开发新的疫苗控制结核病的发展。为了实现这个目的,对结核分枝杆菌的生理功能研究迫在眉睫,找到一些对结核分枝杆菌在体内生长必需的基因及其产物以开发新的药物靶点和设计疫苗成为当务之急^[3]。核苷酸是结核分枝杆菌生长和分化的必需物

质,同时核苷酸还可参与到结核分枝杆菌的其他代谢过程,如糖代谢、脂代谢和氨基酸代谢,在结核分枝杆菌的增殖、DNA 损伤修复和细胞周期等生理生化过程中发挥重要作用^[4]。嘧啶核苷酸主要通过从头合成途径生成,包含 6 个连续的酶促反应过程^[13],简述如下:氨基甲酰磷酸合成酶 II 催化谷氨酰胺、 HCO_3^- 和 ATP 生成氨基甲酰磷酸;天冬氨酸氨基甲酰转移酶催化天冬氨酸与氨基甲酰磷酸生成氨基甲酰天冬氨酸;二氢乳清酸酶催化氨基甲酰天冬氨酸脱水环化形成二氢乳清酸;二氢乳清酸脱氢酶催化二氢乳清酸生成乳清酸;乳清酸磷酸核糖转移酶催化乳清酸与磷酸核糖焦磷酸反应生成乳清酸核苷-5'-磷酸;乳清酸核苷-5'-磷酸脱羧酶 (Orotidine 5'-monophosphate decarboxylase, OMPdecase; Rv1385) 催化乳清酸核苷-5'-磷酸脱羧生成尿嘧啶核苷酸。分枝杆菌中 OMPdecase 由 274 个氨基酸残基组成,相对分子质量 27.4 KD^[14]。在真核细胞中,编码最后两步反应的 OMPdecase 与乳清酸磷酸核糖转移酶的基因发生融合,使得上述两种酶位于同一条多肽链上,即尿嘧啶核苷酸合成酶^[15-16],包含 480 个氨基酸残基。由于一级结构的差异,结核分枝杆菌和人 OMPdecase 的氨基酸序列同源性较低,产生交叉反应的概率较低,该酶可作为新的特异性结核诊断及抗结核疫苗的候选靶标。随着生物信息学的快速发展,生物信息学分析软件已成为预测抗原表位的常用工具,受到越来越多的科研工作者的青睐^[17]。本研究运用生物信息学方法,预测分析 Rv1385 的

B 细胞和 T 细胞抗原表位,分析结果将为下一步寻找新的抗原靶点,为结核早期快速诊断和候选疫苗

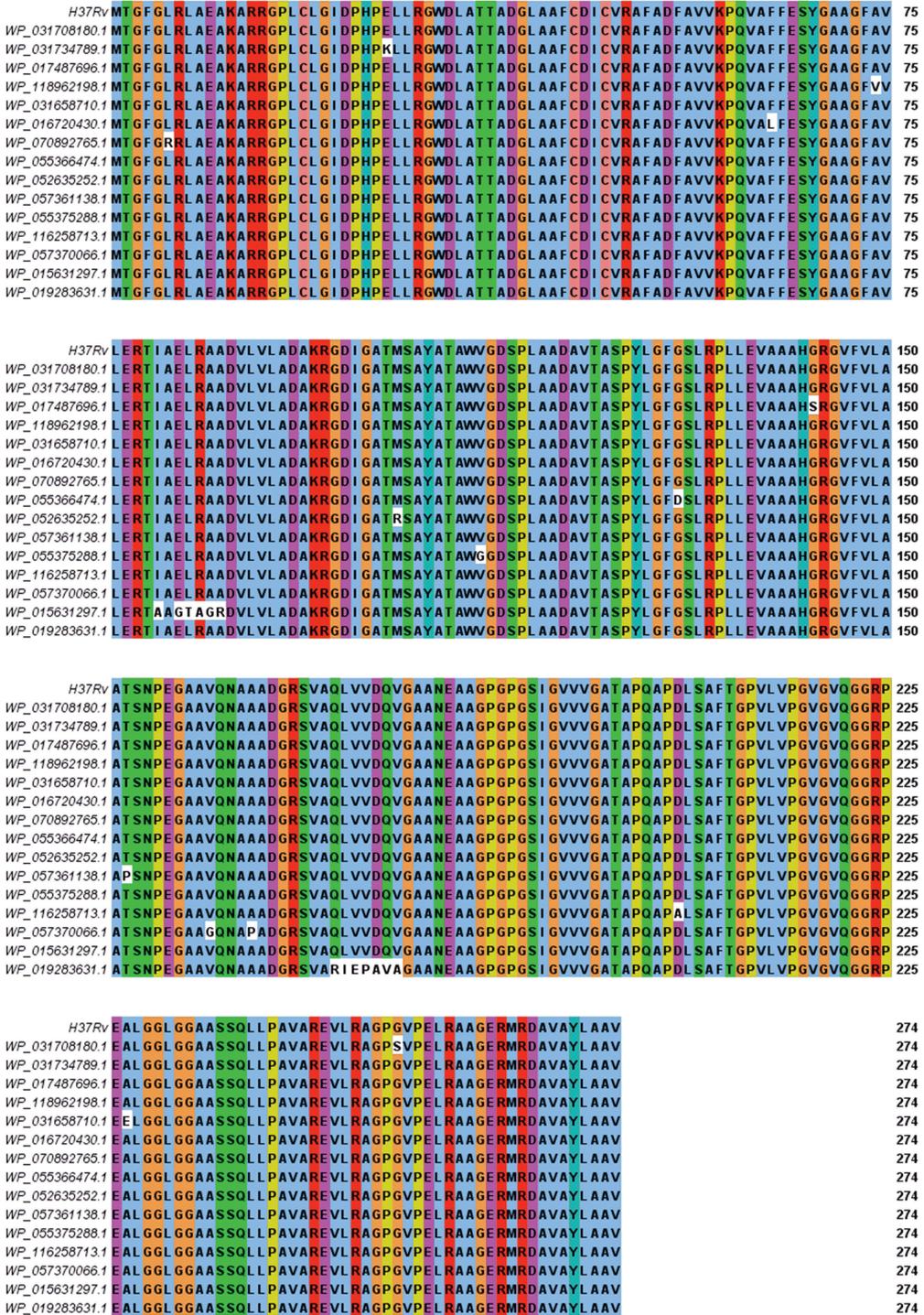


图 2 不同菌株 Rv1385 氨基酸序列的同源性分析

Fig. 2 Multiple sequence alignment of amino acid sequences of Rv1385 from different strains

4 结 论

Rv1385 含有潜在的 B 细胞和 T 细胞抗原表位, 本文结果为结核特异诊断和多表位结核疫苗的研究奠定了基础。

参考文献(References)

[1] 谢龙祥, 党艺方, 谢建平. 翻译后修饰与结核分枝杆菌 耐药调控网络[J]. 生物工程学报, 2018, 34(8): 1279-1287. DOI: 10.13345/j.cjb.170530.

- XIE Longxiang, DANG Yifang, XIE Jianping. Post-translational modification and regulatory network of *Mycobacterium tuberculosis* antibiotic resistance[J]. Chinese Journal of Biotechnology, 2018, 34(8): 1279–1287. DOI: 10.13345/j.cjb.170530.
- [2] BROWN E D, WRIGHT G D. Antibacterial drug discovery in the resistance era[J]. Nature, 2016, 529(7286): 336–343. DOI: 10.1038/nature17042.
- [3] DONINI S, FERRARIS D M, MIGGIANO R, et al. Structural investigations on orotate phosphoribosyltransferase from *Mycobacterium tuberculosis*, a key enzyme of the *de novo* pyrimidine biosynthesis[J]. Scientific Reports, 2017, 7(1): 1–11. DOI: 10.1038/s41598-017-01057-z.
- [4] WARNER D F, EVANS J C, MIZRAHI V. Nucleotide metabolism and DNA replication[J]. Microbiology Spectrum, 2014, 2(5): 1–20. DOI: 10.1128/microbiolspec.MGM2-0001-2013.
- [5] YUAN J, CARDENAS A M, GIBERT H F, et al. Determination of the amino acid sequence requirements for catalysis by the highly proficient orotidine monophosphate decarboxylase[J]. Protein Science, 2011, 20(11): 1891–1906. DOI: 10.1002/pro.728.
- [6] FUJIHASHI M, MNPOTRA J S, MISHRA R K, et al. Orotidine monophosphate decarboxylase – a fascinating workhorse enzyme with therapeutic potential[J]. Journal of Genetics and Genomics, 2015, 42(5): 221–234. DOI: 10.1016/j.jgg.2015.04.005.
- [7] 黄劲, 王洁, 牛雪可, 等. 结核分枝杆菌 *pyrF* 基因的生物信息学分析[J]. 中国病原生物学杂志, 2019, 14(2): 125–130, 136. DOI: 10.13350/j.cjpb.190201.
- HUANG Jin, WANG Jie, NIU Xueke, et al. Bioinformatics analysis of *pyrF* gene encoding protein orotidine 5'-monophosphate decarboxylase in *Mycobacterium tuberculosis*[J]. Journal of Pathogen Biology, 2019, 14(2): 125–130, 136. DOI: 10.13350/j.cjpb.190201.
- [8] World Health Organization. Global tuberculosis report 2018 [M]. Geneva: World Health Organization, 2018.
- [9] 袁秋露, 付玉荣, 伊正君. 结核分枝杆菌 BfrB 蛋白的生物信息学分析[J]. 中国病原生物学杂志, 2018, 13(2): 131–134. DOI: 10.13350/j.cjpb.180206.
- YUAN Qiulu, FU Yurong, YI Zhengjun. Bioinformatic analysis of Rv3841 gene from *Mycobacterium tuberculosis*[J]. Journal of Pathogen Biology, 2018, 13(2): 131–134. DOI: 10.13350/j.cjpb.180206.
- [10] 罗微. 结核分枝杆菌 RD 区在结核病诊断及疫苗中的研究进展[J]. 中华传染病杂志, 2017, 35(11): 701–704. DOI: 10.3760/cma.j.issn.1000-6680.2017.11.013.
- LUO Wei. Progress of *Mycobacterium tuberculosis* RD region in tuberculosis diagnosis and vaccine[J]. Chinese Journal of Infectious Diseases, 2017, 35(11): 701–704. DOI: 10.3760/cma.j.issn.1000-6680.2017.11.013.
- [11] GUPTA A, KAUL A, TSOLAKI A G, et al. *Mycobacterium tuberculosis*: Immune evasion, latency and reactivation[J]. Immunobiology, 2012, 217(3): 363–374. DOI: 10.1016/j.imbio.2011.07.008.
- [12] OKOMBO J, CHIBALE K. Insights into integrated lead generation and target identification in *malaria* and *tuberculosis* drug discovery[J]. Accounts of Chemical Research, 2017, 50(7): 1606–1616. DOI: 10.1021/acs.accounts.6b00631.
- [13] TANAKA Y, TERAMOTO H, INUI M. Regulation of the expression of *de novo* pyrimidine biosynthesis genes in *Corynebacterium glutamicum*[J]. Journal of Bacteriology, 2015, 197(20): 3307–3316. DOI: 10.1128/JB.00395-15.
- [14] COLE S T, BROSCHE R, PARKHILL J, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence[J]. Nature, 1998, 393(6685): 537–544. DOI: 10.1038/31159.
- [15] BEN-SAHRA I, HOWELL J J, ASARA J M, et al. Stimulation of *de novo* pyrimidine synthesis by growth signaling through mTOR and S6K1[J]. Science, 2013, 339(6125): 1323–1328. DOI: 10.1126/science.1228792.
- [16] LANE A N, FAN T W. Regulation of mammalian nucleotide metabolism and biosynthesis[J]. Nucleic Acids Research, 2015, 43(4): 2466–2485. DOI: 10.1093/nar/gkv047.
- [17] CARLOS P, ROUPIE V, HOLBERT, et al. *In silico* epitope analysis of unique and membrane associated proteins from *Mycobacterium avium* subsp. *paratuberculosis* for immunogenicity and vaccine evaluation[J]. Journal of Theoretical Biology, 2015, 384: 1–9. DOI: 10.1016/j.jtbi.2015.08.003.