

文章编号:2095-7386(2019)03-0027-05

DOI:10.3969/j.issn.2095-7386.2019.03.006

结核分枝杆菌 Rv3248c 生物信息学 分析及同源建模

周放¹, 项杰², 占卫红¹, 王培东¹, 何俊才¹, 余晓丽¹, 张可¹

(1. 武汉轻工大学 生物与制药工程学院, 湖北 武汉 430023; 2. 武汉市医疗救治中心检验科, 湖北 武汉 430023)

摘要:运用生物信息学方法对结核分枝杆菌 Rv3248c 进行结构功能分析及同源建模。通过 NCBI、BLAST、ProtParam 和 Interproscan 等生物信息学方法对 Rv3248c 进行同源性、理化性质、保守序列、跨膜、信号肽二级结构和三级结构预测分析。结果显示 Rv3248c 的氨基酸序列与腺苷高半胱氨酸水解酶高度一致, 该蛋白包含 495 个氨基酸, 没有跨膜及信号肽结构, 其二级结构 α 螺旋占 43%, 无规则卷曲占 41.8%, β 折叠占 15.2%, 利用同源建模法构建该蛋白的三级结构, 并对该模型进行质量评估。研究结果有望对结核病的诊断和治疗提供参考。为进一步研究该蛋白质的结构与功能打下基础。

关键词:结核分枝杆菌; Rv3248c; 生物信息学; 同源建模

中图分类号:R 378.911

文献标识码:A

Bioinformatics analysis of the *Mycobacterium tuberculosis* Rv3248c gene and its protein of homologous modeling

ZHOU Fang¹, XIANG Jie², ZHAN Wei-hong¹, WANG Pei-dong¹, HE Jun-cai¹, YU Xiao-li¹, ZHANG Ke¹

(1. School of Biology and Pharmaceutical Engineering, Wuhan Polytechnic University, Wuhan 430023, China;

2. Department of Clinical Laboratory, Wuhan Medical Treatment Center, Wuhan 430023, China)

Abstract: To analyze *Mycobacterium tuberculosis* Rv3248c and its encoding protein. To predict its structure and function with bioinformatics methods. According to the amino acid sequences from NCBI of protein encoded by the *Mycobacterium tuberculosis* Rv3248c from NCBI databases, use BLAST, ProtParam, Interproscan and other online bioinformatics tools to predict the protein's homology, biochemical properties, conserved sequence, transmembrane, signal peptide, secondary structure and tertiary structure. The amino acid sequences of the protein encoded by Rv3248c is highly similar to adenosylhomocysteinase, the protein contains 495 amino acids and no, transmembrane and signal peptide structure, the percent of the alpha helix is 43%, coil is 41.8%, beta sheet is 15.2%, use homology modeling method to construct the tertiary structure of the protein performance good in quality evaluation. The result is to establish foundation for future research on the structures and function of Rv3248c, and to make preparation for the diagnosis and treatment of tuberculosis.

Key words: *Mycobacterium tuberculosis*; Rv3248c; bioinformatics analysis; homology modeling

收稿日期:2019-04-29.

作者简介:周放(1993-),男,硕士研究生,E-mail:674505235@qq.com.

通信作者:张可(1977-),女,讲师,E-mail:kzhang@whpu.edu.cn.

1 引言

结核分枝杆菌是一种古老的细菌,能引起结核病。据2018年最新WHO报道,2017年全球估计有1千万人感染结核,其中160万人因结核病死亡,结核病每年会夺走4500多人生命^[1]。

2008年Manchi C. M. Reedy报道了来自结核分枝杆菌Rv3248c(SahH)的第一个原核晶体结构^[2],2015年Lingaraja Jena、Shamsi Yari等人研究表明SahH与耐多药结核病异烟肼耐药有关,SahH能与异烟肼结合,从而影响异烟肼抑制和杀灭结核杆菌的能力^[3-4]。2016年Dziadek B等人研究表明SahH与人的IL-8结合,构建过量产生SahH的突变菌株,促进病原体进入中性粒细胞。该蛋白在早期调节结核杆菌感染中,可能促使病原体与靶细胞的附着^[5]。前期资料研究显示,Rv3248c具有作为结核病诊断标志和抗结核治疗药物靶点的潜在价值。

在本文中,笔者以结核分枝杆菌Rv3248c为研究对象,通过生物信息学方法对其进行功能分析及同源建模,为进一步研究该蛋白质的结构与功能奠定基础。

2 实验方法

2.1 同源性分析

通过NCBI提供的BLAST工具对该蛋白的氨基酸序列进行序列比对,因查询序列为氨基酸序列,故需采用BLAST工具中的BLASTP并选择nr数据库对该蛋白序列进行序列比对。BLASTP比对后,当数据库中搜索到多个显著相似的序列时,目的序列是否与数据库中的被检索到序列真正有关联,可通过双序列比对分析,找出最匹配的序列,采用BLAST2sequences进行双序列比对,进一步判断两条序列的相似程度。

2.2 理化性质分析

应用在线软件ProtParam(<http://web.expasy.org/protparam/>)对该蛋白进行理化性质分析,主要包括对该蛋白质的氨基酸数量、分子量、等电点、氨基酸组成、理论分子式、原子总数、不稳定系数等等基本性质的分析结果^[6]。对于该蛋白的亲疏水性分析应用在线软件ProtScale(<http://web.expasy.org/protscale/>),输入该蛋白质的氨基酸序列或者该蛋白质的GI号,使用Kyte&Doolittle法调整窗口大小为19(默认值为9),对其的亲疏水性

进行预测分析。

2.3 保守序列分析

使用Interproscan(<http://www.ebi.ac.uk/interpro/scan.html>)在线工具输入该蛋白质的氨基酸序列或者该蛋白质的GI号,对该蛋白的保守结构域、保守位点进行预测,并分析其保守结构域、保守位点的基本性质与功能。

2.4 跨膜分析

应用在线软件TMHMM 2.0(<http://www.cbs.dtu.dk/services/TMHMM/>)输入该蛋白质的氨基酸序列或者该蛋白质的GI号,选择默认参数对该蛋白的拓扑结构进行预测分析。

2.5 信号肽分析

应用在线软件Signal P 4.1(<http://www.cbs.dtu.dk/services/SignalP/>)输入该蛋白质的氨基酸序列或者该蛋白质的GI号,参数选取革兰氏阳性菌和图形输出,其余参数选择默认参数,对该蛋白的信号肽进行预测分析。

2.6 三级结构预测

利用生物信息学来预测蛋白质的三级结构一般有三种方法:同源建模法、线串法和从头算法^[7]。在本文中,笔者采用同源建模的方法对蛋白质的三级结构进行预测,应用SWISS-MODEL(<http://swissmodel.expasy.org/>)在线工具输入该蛋白质的氨基酸序列或者该蛋白质的GI号,在线对其三级结构进行建模并对模型质量进行评估^[8],利用蛋白质结构可视化工具Swiss-PdbViewer对用SWISS-MODEL在线工具构建的三级结构进行分析、优化,利用蛋白质结构检验工具PDBsum Generate对构建的模型进行质量评估。

3 实验结果

3.1 同源性分析结果

通过BLASTP在线工具寻找到多条与该蛋白相似的序列,且这多条序列的打分值、序列一致性、覆盖度及E值均很高,利用Smart Blast找到五条最匹配的序列。选择打分值最高、氨基酸序列同一性最高的序列,即腺苷高半胱氨酸酶的氨基酸序列,其GI号为686042721,将其与Rv3248c的氨基酸序列进行双序列比对,并进一步分析其相似性,结果表明:两条序列的氨基酸序列基本一样,高度相似,由此可推知Rv3248c可能为腺苷高半胱氨酸酶。

3.2 理化性质分析结果

经过 ProtParam 在线软件对 Rv3248c 进行理化性质分析可知,该蛋白质氨基酸数目为 495 个,理论分子量大小为 54 323. 6,理论等电点为 5. 07,带负电荷残基(Asp+Glu)73 个,带正电荷残基(Arg+Lys)51 个,含量最多的氨基酸是丙氨酸和甘氨酸,所占比例分别为 9. 3%和9. 1%,含量最少的氨基酸为半胱氨酸,所占比例为 0. 6%,理论分子式为

表 1 Rv3248c 氨基酸组成

氨基酸	百分比	氨基酸	百分比	氨基酸	百分比	氨基酸	百分比
Ala	9. 3%	Gln	2. 4%	Leu	8. 5%	Ser	3. 6%
Arg	4. 2%	Glu	7. 5%	Lys	6. 1%	Thr	6. 9%
Asn	4. 2%	Gly	9. 1%	Met	3. 0%	Trp	6. 9%
ASP	7. 3%	His	2. 6%	Phe	2. 8%	Tyr	2. 4%
Cys	0. 6%	Ile	5. 7%	Pro	4. 0%	Val	7. 9%

3.3 保守序列分析结果

经过 Interproscan 在线工具对该蛋白进行保守序列分析,发现该蛋白属于腺苷高半胱氨酸酶家族,可能为腺苷高半胱氨酸酶。对 Rv3248c 进行保守性分析,发现其具有一个结合 NAD⁺因子的保守结构域和两个保守位点。用于结合 NAD⁺因子的保守结构域序号为 IPR015878,是位于 253—415 氨基酸之间的序列,这段序列富含甘氨酸且位于该蛋白氨基酸序列的中间区域,表明位于该蛋白中间区域且富含甘氨酸的氨基酸偏好于结合 NAD⁺辅因子。两个保守位点的序号为 IPR020082,分别位于 92—106 氨基酸和 275—291 氨基酸之间,位于 92—106 氨基酸之间的保守位点位于该蛋白的 N 端,位于 275—291 氨基酸之间富含甘氨酸的保守位点用于结合 NAD⁺因子。

3.4 跨膜分析结果

经过在线工具 TMHMM 软件对 Rv3248c 进行跨膜分析可知,该蛋白质 1—495 位氨基酸均位于膜外,无跨膜结构。

3.5 信号肽分析结果

笔者应用在线软件 Signal P 对 Rv3248c 进行信号肽分析结果可知,该蛋白 max. C 值为 0. 123, max. Y 值为 0. 113, max. S 值为 0. 136, mean S 值为 0. 108,其中 mean S 值用来判断是否为分泌蛋白,若 mean S 值大于 0. 5,则预测为分泌蛋白,存在信号肽^[10],该蛋白 mean S 值小于 0. 5,没有信号肽,不是分泌蛋白或者跨膜蛋白,符合该蛋白没有跨

C2407H3794N656O739S18,原子总数为 7 614。不稳定系数为 26. 12,可认定该蛋白为稳定蛋白^[9]。应用在线工具 ProtScale 分析该蛋白的亲疏水性,结果表明该蛋白含两个典型的疏水区和六个典型的亲水区,疏水区分布在 75—84,228—234 位氨基酸之间,亲水区分布在 133—148,175—181,198—212,246—262,433—454,481—486 位氨基酸区域,亲疏水分析表明该蛋白为亲水性蛋白。

膜结构的事实。

3.6 二级结构预测结果

蛋白质的二级结构主要包括 α 螺旋、β 折叠、β 转角以及无规则卷曲等,利用 SSPro 在线工具对 Rv3248c 进行二级结构预测分析,由此可知其中 α 螺旋与无规则卷曲所包含的氨基酸数目所占比例较大,分别是 43%和 41. 8%,而 β 折叠所占比例较少,只占 15. 2%,其中 α 螺旋主要分布在 26—53、71—83、158—173、184—200、220—232、285—297、308—316、417—437 氨基酸之间,无规则卷曲分布较集中的区域有 53—61、108—129、173—184、269—275、317—323、480—495 氨基酸之间,β 折叠分布较多的区域为 62—68、276—282、332—337、378—402 氨基酸之间,Rv3248c 蛋白二级结构的组成如表 2 所示。

表 2 Rv3248c 的二级结构组成

二级结构	α 螺旋	β 螺旋	无规则卷曲
个数	213	75	207
比例	43%	15. 2%	41. 8%

3.7 三级结构预测结果

笔者利用 SWISS—MODEL 的自动建模方式对 Rv3248c 进行三级结构预测,将预测结果经蛋白质结构可视化工具 Swiss—PdbViewer 优化,得出结果如图 1—图 3 所示,其中黑色区域表示 α 螺旋,白色区域表示 β 折叠,灰色区域表示无规则卷曲。从蛋白质三级结构可以看出该蛋白质是一同源四聚体,自动建模系统所选择的模板为 3dhy. 1. A,

此模板表示的蛋白质为腺苷高半胱氨酸酶,建模范围 10—495 位氨基酸,序列相似度为 61%,序列的同一性 100%。SWISS-MODEL 软件对构建好的模型进行质量评估,质量评估分为全局法和局部法,全局法是基于模型整体对其进行评估打分,利用 GMQE(Global Model Quality Estimation)值对模型的质量进行评估。局部法中对每个氨基酸残基均进行评估打分,用 QMEAN4 的值表示模型的质量高低。Rv3248c 的三级结构质量评估结果为 GMQE(Global Model Quality Estimation)值为 0.99, QMEAN4 值为 0.76,综合 GMQE 和 QMEANscore4 值表明模型质量较高,模型构建合理。

为进一步评估模型的质量高低,将构建好的模型提交至基于 PROCHECK 程序的在线蛋白质结构检验工具 PDBsum Generate,得到模型评估结果。对 Rv3248c 三级结构评估中,其在四个区所占比例分别为 88.8%,11.0%,0.1%,0.1%,说明该模型的质量较好,可进一步优化。G-Factors 参数用于蛋白质结构罕见度评估,G-Factors <-0.5 表明该结构不常见,G-Factors <-1.0 说明该结构十分罕见。对 Rv3248c 三级结构评估中,G-Factors 分值为 0.19,可认为该模型结构属于正常范围。综合 SWISS-MODEL 软件和蛋白质结构检验工具 PDBsum Generate 对模型质量的评估结果,可认为所构建的模型合理。

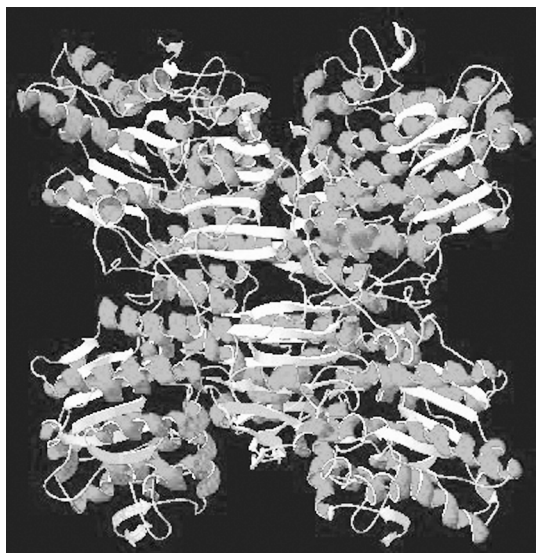


图 1 Rv3248c 三级结构

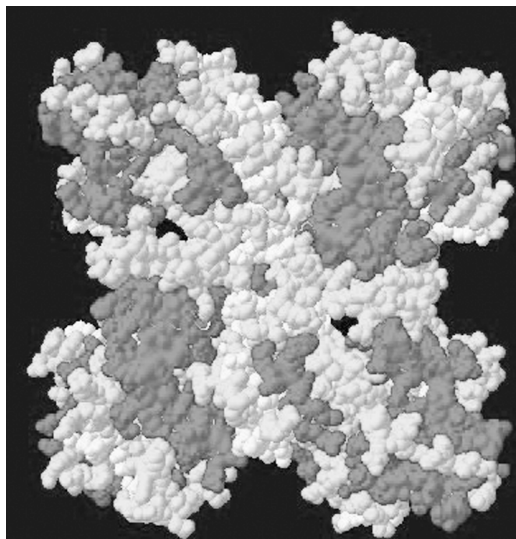


图 2 空间填充模型



图 3 Rv3248c 四聚体结构

4 结束语

结核分枝杆菌 Rv3248c 由 495 个氨基酸组成,分子量为 54 323.6,理论等电点为 5.07,为酸性蛋白酶。不稳定系数为 26.12,认定该蛋白质稳定。根据亲疏水性分析可知该蛋白为亲水性蛋白。跨膜和信号肽分析表明该蛋白是位于细胞质内的非分泌蛋白。保守性分析可知该蛋白为高度保守蛋白,含有一个结合 NAD⁺辅因子的结构域和两个保守位点。该蛋白的二级结构以 α 螺旋和无规则卷曲为主,三级结构建模中与模板相似度为 61%,大于 30%表明选择的模板合理,构建的三级结构置信度高,结果表明该蛋白为一同源四聚体,利于与底物结合。本研究结果有望对结核病的诊断和治疗中提供

有益参考。

参考文献:

- [1] World Health Organization Global tuberculosis report 2018 [R]. Geneva: World Health Organization, 2018.
- [2] Reddy M C, Kuppam G, Shetty N D, et al. Crystal structures of Mycobacterium tuberculosis S — adenosyl — L — homocysteine hydrolase in ternary complex with substrate and inhibitors. [J]. Protein Sci, 2008 , 17 (12): 2134-44.
- [3] Jena L, Deshmukh S, Waghmare P, et al. Study of mechanism of interaction of truncated isoniazid-nicotinamide adenine dinucleotide adduct against multiple enzymes of Mycobacterium tuberculosis by a computational approach. [J] Int J Mycobacteriol. 2015 , 4 (4): 276-83.
- [4] Yari S, Hadizadeh Tasbiti A, Ghanei M, et al. Proteome-scale MDR-TB-antibody responses for identification of putative biomarkers for the diagnosis of drug-resistant Mycobacterium tuberculosis. [J] Int J Mycobacteriol. 2016 , 5 Suppl 1: S134-S135.
- [5] Dziadek B, Brzostek A, Grzybowski M, et al. Mycobacterium tuberculosis AtsG (Rv0296c), GlmU (Rv1018c) and SahH (Rv3248c) Proteins Function as the Human IL — 8 — Binding Effectors and Contribute to Pathogen Entry into Human Neutrophils. [J] PLoS One. 2016, 11 (2): e0148030.
- [6] 李岱容, 穆柳青, 张春燕, 等. 结核分枝杆菌 ClpC2 基因序列分析及功能预测 [J]. 重庆医科大学学报, 2014(01): 57-61.
- [7] 薛庆中. DNA 和蛋白质序列数据分析工具 [M]. 北京: 科学出版社, 2012: 105-113.
- [8] Arnold K, Bordoli L, Kopp J T. The SWISS — MODEL workspace: a web — based environment for protein structure homology modelling [J]. Bioinformatics, 2006, 22 (2): 195-201(7).
- [9] 张悦宁, 陈慧敏, 邬强. 结核分枝杆菌 rmlA 基因的生物信息学分析 [J]. 海南医学院学报, 2012(02): 145-148.
- [10] 吴祖建. 生物信息学分析实践 [M]. 北京: 科学出版社, 2010: 105-112.
- [11] Ponnaluri V K C, Estève P O, Ruse C I, et al. S — adenosylhomocysteine Hydrolase Participates in DNA Methylation Inheritance. [J]. J Mol Biol, 2018 , 430(14): 2051-2065.

(上接第 3 页)

抗氧化活性实验表明党参多糖对亚铁离子、羟基自由基和 ABTS 自由基均具有非常好的清除活性。在 CCP 浓度为 10 mg/m 时, 最大清除率分别为 92. %, 82. 3% 和 73%, 并具有明显的浓度依赖性。

参考文献:

- [1] 黄圆圆, 张元, 康利平, . 党参属植物化学成分及药理活性研究进展 [J]. 中草药, 2018, 49 (01): 239-250.
- [2] 国家药典委员会. 中华人民共和国药典 [S]. 一部. 北京: 中国医药科技出版社, 2015: 248.
- [3] 张雅君, 梁忠岩, 赵伟, 等. 党参水溶性多糖的分离、纯化及组成分析 [J]. 中国药学杂志, 2005, 40(14).
- [4] 张雅君, 梁忠岩, 张丽霞. 党参粗多糖的组成及其免疫活性研究 [J]. 西北农林科技大学学报 (自然科学版), 2012, 40(07): 199-202+208.
- [5] 张立, 李丹, 刘积平, 等. 党参多糖对铅中毒小鼠记忆障碍的影响及其作用机制 [J]. 武警医学, 2013, 24(05): 410-413.
- [6] 展锐, 邵金辉. 大枣多糖抗氧化及抗炎活性的研究 [J]. 现代食品科技, 2017, 33(12): 38-43.
- [7] Singh N, Rajini P S. Free radical scavenging activity of an aqueous extract of potato peel [J]. Food Chemistry, 2004, 85: 611-616.