

DOI:10.13350/j.cjpb.220808

• 论著 •

鸟分枝杆菌 MAV_2043 基因编码蛋白的生物信息学分析*

陈瑞枫, 马枫茜, 李祥芳, 丁寿鹏, 高婧华, 吴利先**

(大理大学基础医学院微生物学与免疫学教研室, 云南大理 671000)

【摘要】 目的 预测鸟分枝杆菌 MAV_2043 基因编码蛋白的结构和功能。方法 利用 UniproKB 数据库、PredictProtein 服务、SWISS-MODEL 服务、ProtParam 蛋白分析工具、Protcale、TMHMM-2.0 服务和 SignalP-5.0 服务、ProtCompB 方法和 IEDB 数据库等生物信息学预测软件对 MAV_2043 编码蛋白的二级和三级结构、理化性质、疏水性分析、信号肽、跨膜区、亚细胞定位及免疫表位进行分析预测;运用 UniprotKB 数据库的 Blast 工具对氨基酸序列进行比对,采用 MEGA11 软件构建进化树,进行系统进化分析。结果 MAV_2043 基因全长 618 bp,共编码 205 个氨基酸。MAV_2043 基因编码蛋白与鸟分枝杆菌亚种起源相同,该蛋白与结核分枝杆菌的[Cu-Zn]超氧化物歧化酶同源性较高。该蛋白为稳定的亲水性蛋白,无跨膜区、有信号肽;预测该蛋白亚细胞定位在细胞膜上,二级结构以 β 折叠为主;三级结构模型显示该蛋白可形成同源二聚体,有两个 Cu^+ 配体的结合空间位置。结论 生物信息学预测鸟分枝杆菌 MAV_2043 基因编码蛋白可能为[Cu-Zn]超氧化物歧化酶,是鸟分枝杆菌中的重要毒力因子,可为鸟分枝杆菌靶向治疗提供新思路。

【关键词】 鸟分枝杆菌; MAV_2043; [Cu-Zn]超氧化物歧化酶; 生物信息学

【中图分类号】 R378.911

【文献标识码】 A

【文章编号】 1673-5234(2022)08-0907-05

[*Journal of Pathogen Biology*. 2022 Aug.;17(8):907-911.]

Bioinformatics analysis of *Mycobacterium avium* protein encoded by gene MAV_2043

CHEN Rui-feng, MA Feng-xi, LI Xiang-fang, DING Shou-peng, GAO Jing-hua, WU Li-xian (Department of Microbiology and Immunology, Basic Medical College, Dali University, Yunnan 671000, China)

【Abstract】 **Objective** To predict the structure and function of the protein encoded by gene MAV_2043 in *Mycobacterium avium*. **Methods** Bioinformatics prediction software such as uniproKB database, PredictProtein service, SWISS-MODEL service, ProtParam protein analysis tool, Protcale, TMHMM-2.0 service and SignalP-5.0 service, ProtCompB method and IEDB database were used for the secondary and tertiary structures, physicochemical properties, hydrophobicity analysis, signal peptide, transmembrane region, subcellular localization and immune epitopes of the protein encoded by gene MAV_2043 were analyzed and predicted; the Blast tool of uniprotKB database was used to compare the amino acid sequences, and software MEGA11 was used to construct an evolutionary tree and perform phylogenetic analysis. **Results**

The gene MAV_2043 is 618 bp long and encodes 205 amino acids, of which threonine accounts for 14.1%, and the molecular formula of the encoded protein is $C_{885}H_{1423}N_{249}O_{290}S_9$. The protein encoded by the gene MAV_2043 has the same origin as *Mycobacterium avium* subsp. *avium*, and the protein is also homologous with the [Cu-Zn] superoxide dismutase of *Mycobacterium tuberculosis*. The protein is a stable hydrophilic protein with an instability index of 33.93 and a grand average of hydropathicity score of -0.061. The protein has no transmembrane region, has signal peptide, has some antigenic epitopes, predicted that the protein subcellularly localizes to the cell membrane, and the secondary structure is dominated by β -folding, accounting for 22.93%. The tertiary structure model shows that the protein could form a homodimer with two Cu^+ ligand binding space positions. **Conclusion** Bioinformatics predicts that the protein encoded by the gene MAV_2043 of *Mycobacterium avium* may be [Cu-Zn] superoxide dismutase, which is an important virulence factor in *Mycobacterium avium* and may provide new ideas for targeted therapy of *Mycobacterium avium*.

【Key words】 *Mycobacterium avium*; MAV_2043; [Cu-Zn] superoxide dismutase; bioinformatics

* **非结核分枝杆菌(nontuberculous mycobacteria, NTM)为机会性致病菌,在艾滋病患者中感染率高于非艾滋病患者。随着艾滋病在全球蔓延,艾滋病合并NTM病呈快速增多趋势,而HIV感染者中的非结核分枝杆菌病大多由鸟分枝杆菌复合群(*Mycobacterium avium* complex, MAC)感染引起。MAC可侵犯人体肺脏、神经、淋巴结、骨骼、关节、皮肤和软组织等

多部位,并可引起全身播散性疾病。病原体与人体宿

* 【基金项目】 国家自然科学基金项目(No. 81760357)。

** 【通讯作者】 吴利先, E-mail: w_lixian@163.com

【作者简介】 陈瑞枫(1996-),男,山西大同人,在读研究生。主要研究方向:分枝杆菌致病相关。
E-mail: CRF1006@Hotmail.com

主之间的相互作用是复杂的,但细菌表面分子与机体在相互作用的几个方面具有重要意义。免疫系统有效地与病原体接触,需要识别可接近的目标,这些目标往往是表面暴露的分子。近年对鸟分枝杆菌(*M. avium*)的发病机制研究取得重大进展,但该细菌感知其环境并与其相互作用的机制仍不十分清楚。附着、运动、分子运输和结合均高度依赖于暴露在表面界面的蛋白质的功能。表面蛋白可能在与吞噬细胞的相互作用中发挥重要作用。表面蛋白也是先天和适应性免疫反应的主要目标。有研究表明,MAV_2043 基因编码的蛋白质是鸟分枝杆菌的重要毒力因子,在巨噬细胞感染和中性粒细胞感染中具有重要意义^[1]。本研究采用多种蛋白质分析工具对 MAV_2043 基因编码蛋白质进行生物信息学分析,预测其结构和功能,为进一步了解该蛋白在鸟分枝杆菌致病机制中的作用及作为候选疫苗抗原提供理论基础。

内容与方法

1 MAV_2043 基因核苷酸序列及编码蛋白氨基酸序列信息的获取

通过 UniProt 的知识库(UniProtKB)及 NCBI 的 Gene 数据库获取鸟分枝杆菌(菌株 104)MAV_2043 基因的核苷酸序列以及该基因编码蛋白的氨基酸序列相关信息。

2 蛋白二、三级结构预测及绘图

采用 PredictProtein 服务^[2] 和 SWISS-MODEL 服务^[3] 在线分析该蛋白的二级结构及三级结构并对其进行建模。

3 蛋白特性、跨膜结构及信号肽预测分析

通过 ProtParam 工具^[4] 对该蛋白的分子质量、等电点、氨基酸和原子组成、消光系数、半衰期、不稳定系数、脂肪族氨基酸指数等进行预测分析;运用 ProtScale^[4] 分析蛋白的亲疏水性;通过 TMHMM-2.0 服务^[5] 和 SignalP-5.0 服务^[6] 分析和预测蛋白的跨膜结构及信号肽。

4 序列比对分析

使用 UniProtKB 中自带的 Blast 工具,对 MAV_2043 基因编码的氨基酸序列进行比对,比对出同源性较高的氨基酸序列,将比对后的氨基酸序列运用 MEGA11 软件构建进化树。

5 蛋白亚细胞定位

运用 ProtCompB 方法^[7] 对该蛋白的亚细胞定位进行分析。

6 蛋白免疫表位分析

采用 IEDB 数据库预测该蛋白的 CD4⁺ T 细胞免疫原性^[8] 及 B 细胞抗原表位^[9]。

结 果

1 MAV_2043 基因及其编码蛋白的基本信息

通过预测和比对,鸟分枝杆菌 MAV_2043 基因编码蛋白为[Cu-Zn]超氧化物歧化酶([Cu-Zn] superoxide dismutase, [Cu-Zn] SOD),该蛋白条目为 A0A0H3A1X7,基因编码为 ABK68060.1,氨基酸总数为 205 个。MAV_2043 基因全长 618 bp,在基因组 CP000479.1 中的位置为 2040139-2040756。

2 蛋白的二、三级结构

鸟分枝杆菌中[Cu-Zn] SOD 的蛋白二级结构预测结果如图 1 第一、三行所示,蓝色表示螺旋结构(Helix),主要包括 α -螺旋,占蛋白质总长度的 7.80%;红色表示延展结构(strand),主要指 β -折叠,占蛋白质总长度 22.93%;其他结构(other)用黄色指示,共占蛋白质总长度的 69.27%。[Cu-Zn]SOD 蛋白的三级结构建模如图 2 所示,与模板 1pzs. 1. A 的序列一致性为 65.19%。该模板来自结核分枝杆菌(*Mycobacterium tuberculosis*, MTB)的[Cu-Zn] SOD 的晶体结构,QMEANDisCo Global 评估分为 0.75 \pm 0.05(分数范围为 0~1,分数越大模型质量越高)。建模显示该蛋白为同型二聚体,有两个 Cu⁺ 配体。

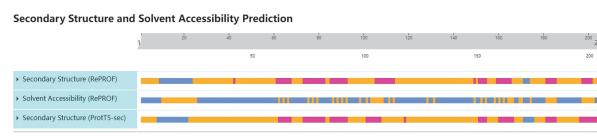


图 1 [Cu-Zn] 超氧化物歧化酶的蛋白二级结构建模
Fig. 1 Results of protein secondary structure modeling of [Cu-Zn] superoxide dismutase

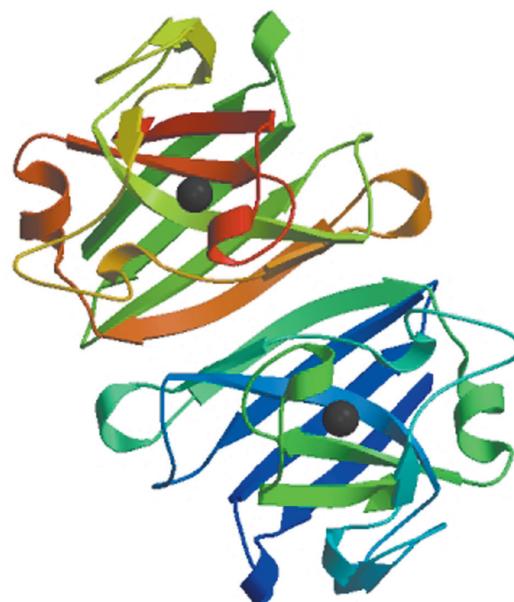


图 2 [Cu-Zn] 超氧化物歧化酶的蛋白三级结构预测
Fig. 2 Results of protein tertiary structure prediction of [Cu-Zn] superoxide dismutase

3 蛋白的理化性质、跨膜区及信号肽

该蛋白的分子式为 $C_{885}H_{1423}N_{249}O_{290}S_9$, 分子质量为 20.480 07 ku, 等电点(pI)为 6.13。总数 205 个氨基酸中丙氨酸 Ala(A)25 个, 占 12.2%; 甘氨酸 Gly(G)28 个, 占 13.7%; 苏氨酸 Thr(T)29 个, 占 14.1%。其余氨基酸含量占比均低于 10.0%。带负电荷的残基总数(Asp + Glu)为 15 个, 带正电荷的残基总数(Arg+Lys)为 12 个。该蛋白在水溶液中吸光度(A)值 280 nm 处测量时, 假设所有成对的 Cys 残基都形成胱氨酸, 消光系数为 4720; 假设所有的 Cys 残基都被还原, 消光系数为 4470, 单位为 $M^{-1}cm^{-1}$ 。该蛋白在大肠埃希菌体内半衰期大于 10 h, 不稳定指数 33.93, 为稳定蛋白。脂肪族氨基酸指数(Aliphatic index)为 71.95, 脂肪族氨基酸指数越大则对蛋白质热稳定性提高越有利。蛋白的总平均亲水性(Grand average of hydropathicity)得分为 -0.061, 图 3 显示在第 17 位氨基酸处有最大得分, 在 192 位氨基酸处得分最小, 综合分析认为该蛋白质为亲水性蛋白质。TM-HMM 预测该蛋白跨膜螺旋数为 0, 说明此蛋白为非跨膜蛋白(图 4)。SignalP 分析显示, 在氨基酸序列 24 号位和 25 号位之间有一个裂解位点 MTA-CS, 概率为 99.75%(图 5); 蛋白质类型为脂蛋白信号肽(Lipoprotein signal peptide (Sec/SPII)), 概率为 99.76%, 说明该蛋白有极大可能产生信号肽。

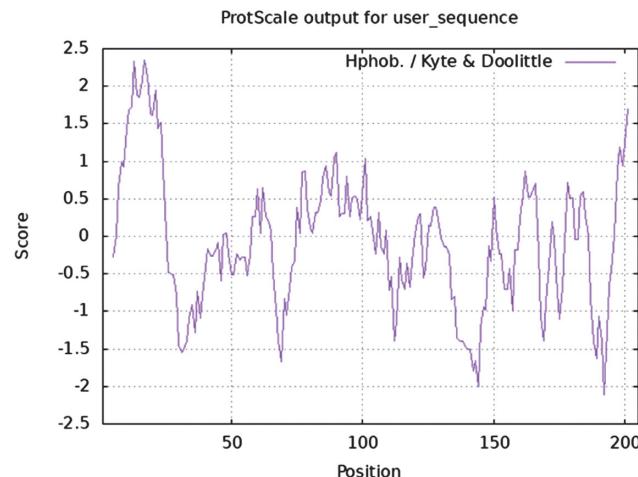


图 3 [Cu-Zn]超氧化物歧化酶的亲疏水性预测

Fig. 3 Predicted hydrophobicity of [Cu-Zn] superoxide dismutase

4 蛋白系统进化树

经氨基酸序列比对后, 鸟分枝杆菌 104 菌株的 [Cu-Zn] SOD 与鸟分枝杆菌亚种(*Mycobacterium avium* subsp. *avium*)、副结核分枝杆菌(*Mycobacterium paratuberculosis*)、未分类的分枝杆菌(unclassified *Mycobacterium*)、莎斯喀彻温分枝杆菌(*Mycobacterium saskatchewanense*)、结核分枝杆菌均比对出高相似度氨基酸序列, UniProtKB 中蛋白质条目分别为 X8AZS1、Q73Y27、I2AC51、A0A1X2BIS5、Q9AGW2、P9WGE8, 序列相似覆盖区域分别为 100.0%、99.5%、88.3%、81.2%、56.1%、54.7%, 用 MEGA11 软件构建系统进化树, 结果见图 6, 表明 [Cu-Zn] SOD 与鸟分枝杆菌亚种中的蛋白质起源于同一物种, 具有高度同源性。

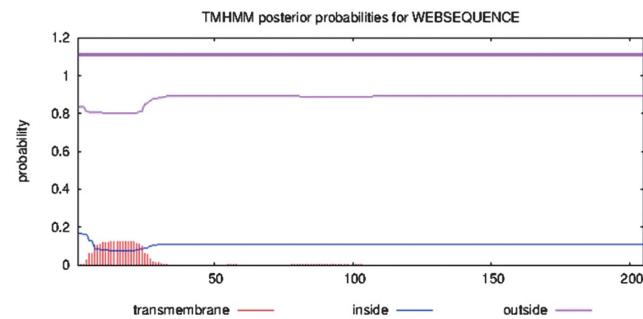


图 4 [Cu-Zn]超氧化物歧化酶的跨膜区预测
Fig. 4 Predicted results for the transmembrane region of [Cu-Zn] superoxide dismutase

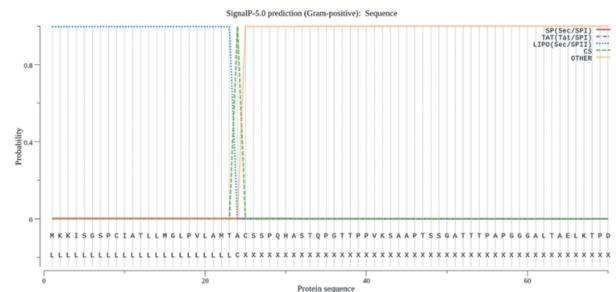


图 5 [Cu-Zn]超氧化物歧化酶的信号肽预测
Fig. 5 Signal peptide prediction results for [Cu-Zn] superoxide dismutase



图 6 [Cu-Zn]超氧化物歧化酶的系统进化树
Fig. 6 Phylogenetic tree of [Cu-Zn] superoxide dismutase

5 蛋白亚细胞定位

经 ProtComp 分析, 由神经网络预测, 该蛋白亚细胞定位于细胞膜(Membrane), 得分为 2.0, 且该蛋白并非跨膜蛋白, 高度符合超氧化物歧化酶的功能特征, 与 McNamara 等^[1]的预测结果一致。

6 蛋白的 T、B 细胞抗原表位

运用 IEDB 数据库预测 CD4⁺ T 细胞的免疫原性, 最大综合得分阈值默认为 50, 返回 1 个可识别性较高的结果, 多肽序列为 ATLLMGLPVLAMTAC, 氨基酸起始位为 11, 终止位为 25, 综合评分 49.26, 免

疫原性评分 88.65。通过 IEDB 分析预测 B 细胞抗原表位,结果见图 7。其中,28-59、71、95-100、112-147、169-177、187-194 共 6 组氨基酸为 B 细胞优势表位,可作为抗体识别靶点。

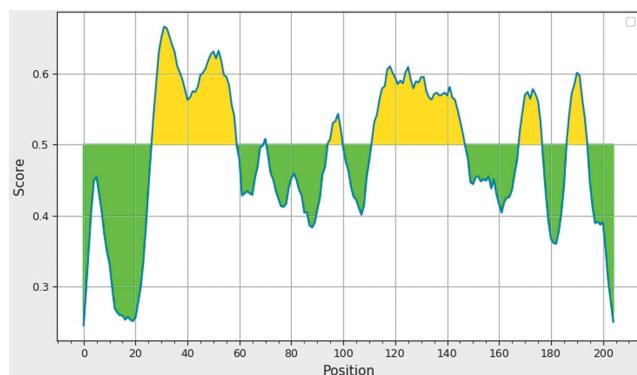


图 7 [Cu-Zn] 超氧化物歧化酶 B 细胞抗原表位预测
Fig. 7 Predicted results of [Cu-Zn] superoxide dismutase B-cell antigen epitopes

讨 论

在肺结核高负担国家中,NTM 的诊断经常被遗漏^[10],与检测 MTB 感染所用的工具相比,临幊上对 NTM 的诊断和治疗仍然较为落后。诊断结核病的方法包括分子诊断和药物敏感性试验(drug susceptibility test, DST),均具有较高的敏感性和特异性^[11]。NTM 抗原等与 MTB 多有交叉,因此可根据已获研究和证实的抗 MTB 的机制对未知的抗 NTM 机制进行相关研究,了解抗 NTM 的具体机制,从而达到降低 NTM 感染率的目标。

MTB 侵入宿主后,巨噬细胞可引发不同的细胞免疫和临床表现^[12]。巨噬细胞处于固有防御的前沿,与病原体相遇后通过触发炎症、消化细菌和诱导适应性免疫反应参与控制细菌的感染过程^[13]。然而,MTB 可破坏巨噬细胞的反应,如细胞凋亡,从而实现免疫逃逸。巨噬细胞的凋亡是宿主通过限制细菌的细胞内生长来防御 MTB 的关键部分。MTB 感染的巨噬细胞发生凋亡后,便可将 MTB 释放到细胞外环境中,从而激活树突状细胞并激发强大的适应性免疫反应^[14]。因此抗巨噬细胞杀伤对分枝杆菌的毒力至关重要。在分枝杆菌侵入宿主后,巨噬细胞成为分枝杆菌在细胞内持续存在的主要储存库^[15]。巨噬细胞等免疫细胞可释放活性氧(reactive oxygen species, ROS),通过氧化破坏微生物结构,如细胞壁、DNA 和酪氨酸残基等实现对分枝杆菌的细胞内杀灭,但强毒分枝杆菌可通过某种方式抑制这种杀灭作用^[16],也可抑制细胞凋亡并触发宿主巨噬细胞坏死以逃避先天性免疫杀伤^[17]。鸟分枝杆菌就可停留在巨噬细胞中的吞噬体内,并能够离开巨噬细胞感染其他细胞。鸟分枝杆菌具有多种

毒力机制以适应有害的吞噬体内环境。鸟分枝杆菌能够将巨噬细胞的表型转换为 M2 型(Alternatively activated macrophage),以逃避抗菌的细胞免疫反应。鸟分枝杆菌中还含有多种抗氧化酶,包括超氧化物歧化酶、过氧化氢酶和烷基过氧化氢还原酶等。这些机制可诱导一系列生理变化,使鸟分枝杆菌能够在巨噬细胞产生的氧化应激下存活^[18]。

鸟分枝杆菌中 MAV_2043 基因编码蛋白为[Cu-Zn] SOD 蛋白质,该基因在分枝杆菌中广泛分布^[19]。有研究发现,MTB 需要[Cu-Zn] SOD 来对外源产生的超氧物及其协同产物的毒性作用形成完全抵抗力,MTB 中的超氧化物歧化酶(SODs)可通过反应生成过氧化氢,利用过氧化氢酶-过氧化物酶(Catalase - Peroxidase, KatG)进一步解毒,以对抗巨噬细胞产生的超氧化物自由基^[20]。由此推测鸟分枝杆菌中[Cu-Zn] SOD 也可对巨噬细胞产生相似的抵抗作用。MAV_2043 基因编码的蛋白质为稳定亲水性蛋白质,无跨膜区,有一个脂蛋白信号肽,亚细胞定位于细胞膜上,而 MTB 中编码[Cu-Zn] SOD 的基因也含有一个脂蛋白结合基序,可能与细菌外膜的附着相关^[21]。多重氨基酸序列比对显示,鸟分枝杆菌中[Cu-Zn] SOD 与鸟分枝杆菌亚种中的[Cu-Zn] SOD 氨基酸序列完全相同,均与 MTB 的[Cu-Zn] SOD 有较高的同源性,表明鸟分枝杆菌中的[Cu-Zn] SOD 也非常有可能对抗超氧化物自由基,对巨噬细胞产生的氧化爆发产物进行抵抗。鸟分枝杆菌中[Cu-Zn] SOD 的二级结构中 β -折叠占比较高,约占 22.93%。蛋白质的二级结构与蛋白质 B 细胞抗原表位有一定联系,对鸟分枝杆菌[Cu-Zn] SOD 进行 B 细胞抗原表位分析,氨基酸位点 28-59、71、95-100、112-147、169-177、187-194 等 6 组 B 细胞表位。预测 CD4⁺ T 细胞的免疫原性,得到一个免疫原性评分 88.65 的多肽序列 ATLLMGLPV-LAMTAC。因此确定鸟分枝杆菌中[Cu-Zn] SOD 含有较为丰富的 T 细胞、B 细胞抗原表位,可为进一步研究该蛋白的功能和明确其在鸟分枝杆菌感染机制中作用提供理论依据,以期为筛选针对鸟分枝杆菌的诊断标志物及研制新型表位疫苗奠定基础。

目前,使用多种在线分析软件对蛋白质进行生物信息学分析已成为预测细胞蛋白质功能的重要方法。本研究对鸟分枝杆菌[Cu-Zn] SOD 进行了多维的生物信息学分析,采用多种方式从多角度进行了分析和预测,有效地规避了单一软件预测的缺陷。[Cu-Zn] SOD 是鸟分枝杆菌中的重要毒力因子,对其作用机体的机制尚不完全明确,[Cu-Zn] SOD 在人体内也有生成,且与多种疾病相关^[22],深入研究[Cu-Zn] SOD 的免疫学特性及结构功能特征,可为鸟分枝杆菌的靶向

治疗提供新思路。为进一步研究[Cu-Zn] SOD的结构、功能、理化性质明确方向,也可为进一步揭示鸟分支杆菌的致病机制提供参考。

【参考文献】

- [1] McNamara M, Tzeng S C, Maier C, et al. Surface-exposed proteins of pathogenic mycobacteria and the role of cu-zn superoxide dismutase in macrophages and neutrophil survival[J]. *Proteome Sci*, 2013, 11(1):45.
- [2] Rest B, Chris S. Prediction of protein secondary structure at better than 70% accuracy[J]. *J Mol Biol*, 1993(232):584-599.
- [3] Andrew W, Martino B, Stefan B, et al. SWISS-MODEL: homology modelling of protein structures and complexes[J]. *Nucleic Acids Res*, 2018(W1):W296-W303.
- [4] Wilkins MR, Gasteiger E, Bairoch A, et al. Protein identification and analysis tools on the ExPASy server[J]. *Methods Mol Biol*, 1999(112):531-552.
- [5] Krogh A, Larsson B, Heijne GV, et al. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes [J]. *J Mol Biol*, 2001, 305(3):567-580.
- [6] Almagro A, Jose J, Tsirigos KD, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks[J]. *Nat Biotechnol*, 2019(37):420-423.
- [7] Softberry, Inc. ProtComp - Predict the sub-cellular localization of bacterial proteins [CP/OL]. [2022-2-20]. http://www.softberry.com/berry_phtml?topic=pcmb&group=programs&subgroup=proloc.
- [8] Dhanda SK, Karosiene E, Edwards L, et al. Prediction of HLA CD4 immunogenicity in human populations[J]. *Front Immunol*, 2018(9):1369.
- [9] Closter J M, Bjoern P, Morten N, et al. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes[J]. *Nucleic Acids Res*, 2017, 45(W1):W24-W29.
- [10] Sharma SK, Upadhyay V. Epidemiology, diagnosis & treatment of non-tuberculous mycobacterial diseases [J]. *Indian J Med Res*, 2020, 152(3):185-226.
- [11] Gopalaswamy R, Shanmugam S, Mondal R, et al. Of tuberculosis and non-tuberculous mycobacterial infections- a comparative analysis of epidemiology[J]. *Diagnosis Treat J Biomed Sci*, 2020, 27(1):74.
- [12] Ge G, Jiang H, Xiong J, et al. Progress of the art of macrophage polarization and different subtypes in mycobacterial infection[J]. *Front Immunol*, 2021(12):752657.
- [13] Ferraris DM, Miggiano R, Rossi F, et al. *Mycobacterium tuberculosis* molecular determinants of infection, survival strategies, and vulnerable targets[J]. *Pathogens*, 2018, 7(1):17.
- [14] Abdalla AE, Ejaz H, Mahjoob MO, et al. Intelligent mechanisms of macrophage apoptosis subversion by *Mycobacterium* [J]. *Pathogens*, 2020, 9(3):218.
- [15] Prasla Z, Sutliff RL, Sadikot RT. Macrophage signaling pathways in pulmonary nontuberculous mycobacteria infections[J]. *Am J Respir Cell Mol Biol*, 2020, 63(2):144-151.
- [16] Ganguli G, Mukherjee U, Sonawane A. Peroxisomes and oxidative stress: Their implications in the modulation of cellular immunity during mycobacterial infection[J]. *Front Microbiol*, 2019(10):1121.
- [17] Shamaei M, Mirsaeidi M. Nontuberculous mycobacteria, macrophages, and host innate immune response[J]. *Infect Immun*, 2021, 89(8):e0081220.
- [18] Abukhalid N, Islam S, Ndzeidze R, et al. *Mycobacterium avium* subsp. hominis interactions with macrophage killing mechanisms[J]. *Pathogens*, 2021, 10(11):1365.
- [19] Liao D, Fan Q, Bao L. The role of superoxide dismutase in the survival of *Mycobacterium tuberculosis* in macrophages[J]. *Jpn J Infect Dis*, 2013, 66(6):480-488.
- [20] Trivedi A, Singh N, Bhat SA, et al. Redox biology of tuberculosis pathogenesis[J]. *Adv Microb Physiol*, 2012(60):263-324.
- [21] Piddington DL, Fang FC, Laessig T, et al. Cu, Zn superoxide dismutase of *Mycobacterium tuberculosis* contributes to survival in activated macrophages that are generating an oxidative burst[J]. *Infect Immun*, 2001, 69(8):4980-4987.
- [22] Lewandowski L, Kepinska M, Milnerowicz H. The copper-zinc superoxide dismutase activity in selected diseases[J]. *Europ J Clin Invest*, 2019, 49(1):e13036.

【收稿日期】 2022-04-09 【修回日期】 2022-06-11