

网络出版时间: 2020/5/29 13:02 网络出版地址: <http://kns.cnki.net/kcms/detail/34.1065.r.20200527.1642.001.html>

◇基础医学研究◇

日本斑点热立克次体安徽 120 株 *ompB* 基因及其编码蛋白的生物信息学和 B 细胞抗原表位分析

刘伯玉¹, 罗婉蓉^{1*}, 陈振¹, 余冬阳¹, 高越¹, 朱禹¹, 洪颖^{1,2}, 杨莉¹, 柳燕¹

摘要 目的 对从我国临床患者血液中新发现和分离出的日本斑点热立克次体安徽 120 株 *ompB* 基因及其编码蛋白进行系统的结构与功能预测, 为深入了解其在立克次体致病机制中的作用、研发斑点热临床诊断试剂和疫苗提供理论基础。方法 基于日本斑点热立克次体安徽 120 株 *ompB* 基因及其编码的氨基酸序列, 应用生物信息学方法预测其信号肽、跨膜区、疏水性等性质和可能的 B 细胞抗原表位。结果

日本斑点热立克次体安徽 120 株 *ompB* 基因在 3 297 ~ 3 311 处存在连续 15 个碱基的缺失, 除此之外, 该分离株还有 4 个碱基突变。该基因编码的氨基酸数目为 1 651 个, 蛋白质分子量为 167.69 ku, 共由 20 种氨基酸组成, 其中含量较多的氨基酸是 Gly; 理论等电点 pI 为 5.15, 消光系数为 58 805 或 58 680, 蛋白不稳定指数为 7.15, 脂肪指数为 88.86, 总平均亲水性 (GRAVY) 为 0.06, 为疏水性蛋白; Signal peptide (Sec/SPI) Likelihood 为 0.567 7, 有信号肽序列; 存在一个 Pfam 区域和自转运体区, 分别与细菌黏附宿主和蛋白质转运有关; 蛋白二级结构中以无规卷曲为主, 其次为 β -折叠和 α -螺旋, 分别占到 53.73%、30.83% 和 15.45%。同源建模蛋白的三级结构中, 第 1 308 ~ 1 651 氨基酸部分在蛋白表面形成桶状结构, 可能具有丝氨酸蛋白酶的作用; 蛋白分子内抗原表位较丰富, 共有 159 个可能的抗原表位, 预测分数最高 (0.94) 的为 1 269 ~ 1 284 位序列。结论 成功分析和预测了日本斑点热立克次体安徽 120 株 *ompB* 基因及其编码蛋白的组成、蛋白质的二级结构、三级结构和 B 细胞抗原表位, 为研究日本斑点热立克次体的致病机制、临床诊断试剂和亚单位疫苗的研发提供了理论支持。

关键词 日本斑点热立克次体安徽 120 株; *ompB*; B 细胞抗原表位; 生物信息学

中图分类号 R 376.3

文献标志码 A 文章编号 1000-1492(2020)06-0819-06

2020-02-22 接收

基金项目 “十三五”国家科技重大专项(编号:2018ZX10711001); 国家自然科学基金(编号:81571963, 81772203)

作者单位: ¹ 安徽医科大学微生物学教研室, 合肥 230032

² 马鞍山疾病预防控制中心, 马鞍山 243000

作者简介: 刘伯玉, 男, 博士, 讲师;

柳燕, 女, 教授, 硕士生导师, 责任作者, E-mail: yanliu@ahmu.edu.cn

* 对本文具有同等贡献

doi: 10.19405/j.cnki.issn1000-1492.2020.06.001

斑点热 (spotted fever, SF) 是由斑点热群立克次体 (spotted fever group rickettsiae, SFGR) 引起的一组以急性发热和皮疹为主要临床症状的疾病的总称, 主要包括: 落矶山斑点热、日本斑点热、北亚热、昆士兰热等, 是一种自然疫源性疾病, 分布范围广且人群普遍易感^[1-2]。近年来, 随着人们社会活动的扩大以及旅游业的蓬勃发展, SF 在美洲、欧洲和亚洲等新发病例逐年增加。我国至今已在十几个省、市、自治区证实有斑点热立克次体感染, 并在蜱、鼠等中检测到斑点热立克次体 DNA^[3-5]。

日本斑点热归属于斑点热群, 其病原体为日本斑点热立克次体, 一种以蜱为主要传播媒介的感染性疾病。1984 年首次在日本一名患者血中分离得到^[6]。目前, 该疾病已在多个国家发现, 包括中国、日本、韩国、菲律宾和泰国等^[7-8]。2013 年, 课题组在安徽省大别山区 1 例发热患者血液标本中分离到 1 株日本斑点热立克次体近缘株安徽 120 株, 简称安徽 120 分离株; 这是首次在我国发现并成功分离的人感染日本斑点热立克次体^[9]。

立克次体表面蛋白是存在于菌体表面的一大类蛋白, 目前已知有 20 多种, 包括 OmpA (sca0)、OmpB (sca5)、GeneD (sca4)、 β -peptide、Adr2、Sca1、Sca12、Sca13 等多种蛋白^[10]; 不同种的立克次体表面蛋白种类也有差异, 同种立克次体表面蛋白在核苷酸和氨基酸序列也可存在差异, 从而决定了其致病性、免疫性也存在不同。研究^[11]表明 OmpB 能通过其效应结构域促进立克次体吸附到宿主细胞表面, 并与其表面受体 (Ku-70) 特异性结合, 介导细菌进入宿主细胞; 同时 OmpB 蛋白还可刺激宿主产生特异、保护性抗体, 在恢复期通过体液免疫杀灭游离的病原体, 也可对再次入侵的病原体起到杀灭作用^[12]。因此 *ompB* 基因及其编码蛋白的研究对于深入理解立克次体的致病机制、临床诊断试剂和亚单位疫苗研发等均有重要意义。

考虑到安徽 120 株的地区特异性,在遗传进化上可能不同于国外分离菌株。为更好地了解安徽 120 分离株的生物学特点,该研究拟采用生物信息学方法,综合多种分析策略对其外膜蛋白基因 *ompB* 的结构、编码蛋白以及 B 细胞抗原表位进行探讨,为研究日本斑点热立克次体的致病机制与研发临床诊断试剂和亚单位疫苗等提供理论支持,为日本斑点热疾病的防制提供依据。

1 材料与方法

1.1 安徽 120 分离株 *ompB* 基因的克隆 病原体的分离和鉴定 *ompB* 基因序列的 PCR 扩增、克隆和测序,参照课题组先前发表的文章^[9,13]。

1.2 *ompB* 基因序列的 NCBI 递交 所得的测序结果进行初步注释后,将基因序列提交 NCBI。并与 NCBI 数据库上公布的日本斑点热立克次体其他分离株的 *ompB* 序列做同源比对。

1.3 *ompB* 基因的生物信息学分析及 B 细胞表位的预测 参考已发表文献^[14]的方法,分析安徽 120 分离株的 *ompB* 基因序列及其编码蛋白质。过程如下:采用 Bioedit V7.0 和 NCBI 的 BLAST 程序分析 *ompB* 基因序列特征和组成。采用 NCBI 的 ORF Finder 在线工具、Bioedit 和 ExPASy 的 ProtParam tool 在线工具软件来预测 *ompB* 编码蛋白质的理化性质,如分子量、等电点、氨基酸组成、稳定性和摩尔消光系数等;采用 ExPASy 服务器的 ProtScale 程序分析 *ompB* 编码蛋白的疏水性;采用 SignalP 5.0 对 *ompB* 编码蛋白进行信号肽预测;采用 TMHMM 在线工具预测 *ompB* 编码蛋白跨膜区;采用 SMART 服务器对 *ompB* 编码蛋白功能结构域分析;采用 GOR4 服务器和 SSPro 4.0 服务器对 *ompB* 编码蛋白二级结构进行预测;采用 SWISS-MODEL 同源建模 *ompB* 编码蛋白的三级结构;参考文献^[15-16]采用 Laser-gene 软件的 Protean 模块并同时使用在线工具 ABCpred Prediction Server 预测 *ompB* 编码蛋白的 B 细胞线性表位。

1.4 统计学处理 采用 SPSS 20.0 对数据进行统计分析。构成比的比较采用 χ^2 检验, $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 安徽 120 分离株 *ompB* 序列特征 通过 T-A 克隆得到的 *ompB* 碱基序列后,利用 Bioedit 软件对 *ompB* 基因序列的组分进行了分析,发现 *ompB* 基因

核酸序列总长度为 5 280 碱基对;分子质量为 1 601.6 ku(单链状态)和 3 198.3 ku(双链状态);GC 含量是 37.20%,AT 的含量是 62.80%。A、C、G 和 T 核苷酸的摩尔分数见图 1。同时,将所得的测序序列进行注释后,提交到 NCBI,并与 NCBI 数据库上公布的日本斑点热立克次体其他分离株的 *ompB* 序列做同源比对,安徽 120 分离株的 *ompB* 基因(GenBank: KY364904.1)在 3 297 ~ 3 311 处与 NCBI 上的日本斑点热立克次体(Sequence ID: AP017602.1)序列比较,两序列的同源性为 99%。安徽 120 株存在连续 15 碱基的缺失,除此之外,该分离株还有 4 个碱基突变,分别位于: T3204G、C3170T、A414C 和 A292G。

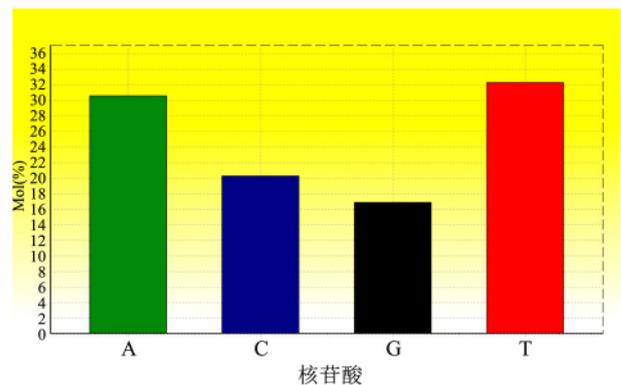


图 1 日本斑点热安徽 120 分离株 *ompB* 基因序列各核苷酸摩尔分数分布

2.2 *ompB* 编码蛋白氨基酸组成 利用 NCBI 网站所提供的 ORF Finder 在线工具和 Bioedit 软件,结合 NCBI 网站的蛋白质数据库,预测了 *ompB* 基因编码的氨基酸数目为 1 651 个,蛋白质分子量为 167.69 ku。所含的氨基酸含量较多的氨基酸是甘氨酸(Gly)(212 个,12.84%),其次为 Ala 和 Asn(均为 196 个,11.87%)。其余氨基酸组成分别为:半胱氨酸(Cys)(2 个,0.12%),Asp(75 个,4.54%),Glu(23 个,1.39%),Phe(70 个,4.24%),His(11 个,0.67%),Ile(119 个,7.21%),Lys(57 个,3.45%),Leu(108 个,6.54%),Met(15 个,0.91%),Pro(30 个,1.82%),Gln(62 个,3.76%),Arg(22 个,1.33%),Ser(91 个,5.51%),Thr(195 个,11.81%),Val(133 个,8.06%),色氨酸(Trp)(2 个,0.12%),Tyr(32 个,1.94%),见图 2。

2.3 *ompB* 编码蛋白理化性质分析 利用 Geneva 大学的 ExPASy 服务器(<https://web.expasy.org/protparam/>)中的 ProtParam tool 对 *ompB* 编码蛋白的

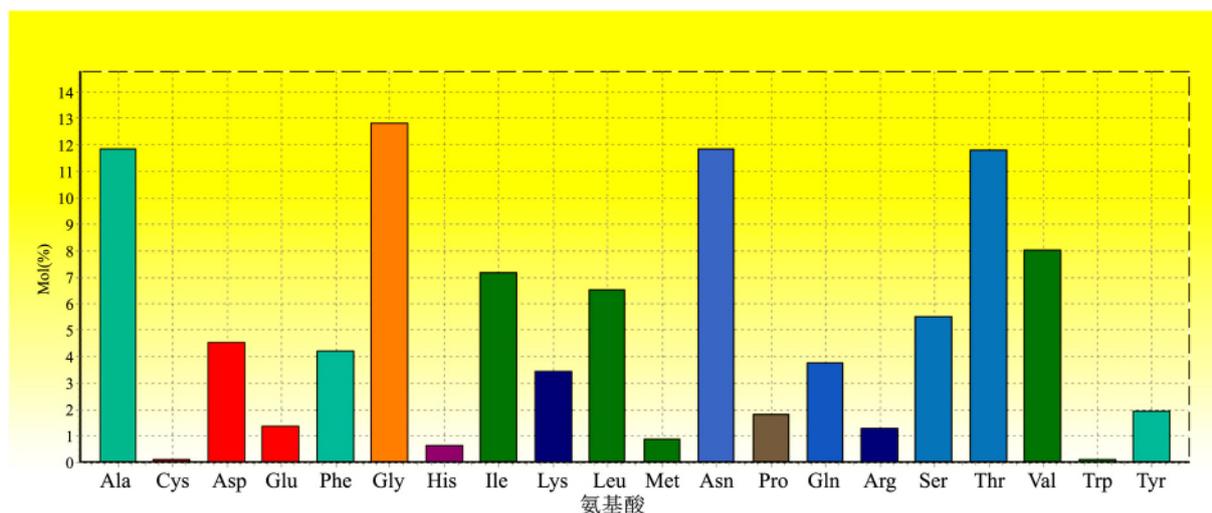


图2 日本斑点热安徽120分离株 *ompB* 编码蛋白各氨基酸摩尔分数分布

理化参数进行了预测。该蛋白的原子组成为 $C_{7312}H_{11658}N_{2056}O_{2424}S_{17}$,总原子数目为 23 467 个。理论等电点 pI 为 5.15 ,在组成 *ompB* 编码蛋白的 20 种氨基酸中 ,Gly 所占的比例最高 ,达到 12.84% ,而 Cys 和 Trp 的比例最低 ,均为 0.12% 。这些预测结果与 Bioedit 软件分析结果相一致。此外 ,课题组还得到 *ompB* 编码蛋白的消光系数为 58 805 (假设所有成对的半胱氨酸残基形成胱氨酸) 或 58 680 (假设所有的半胱氨酸残基被还原) ,且蛋白不稳定指数为 7.15 ,脂肪指数为 88.86 ,总平均亲水性 (grand average of hydropathicity ,GRAVY) 为 0.06 ,根据 Gu-ruprasad 方法表明 *ompB* 编码蛋白较为稳定 (蛋白质的不稳定系数 >40 ,表明蛋白质在溶液中性质不稳定) ,且有疏水性 (GRAVY 值在 2 与 -2 之间 ,正值说明蛋白为疏水蛋白 ,反之为亲水蛋白) 。

2.4 *ompB* 编码蛋白疏水性分析 组成 *ompB* 编码蛋白质的 20 种氨基酸各自带有不同极性的侧链基团 ,其疏水的相互作用分析是进行蛋白质高级结构预测、功能域划分等必要过程。本文采用 ExPASy 服务器的 ProtScale 程序 (<https://web.expasy.org/protscale/>) 分析 *ompB* 编码蛋白氨基酸序列的疏水性。从结果中可得 ,多肽链第 508 位的 Gly (G) 具有最高的分值 2.556 ,第 1 427 位的 Tyr (Y) 和 1 428 位的 Lys (K) 具有最低的分值 -3.011。总体来看 ,疏水性氨基酸均匀分布在整個肽链中 ,且多于亲水性氨基酸 ,因此整个多肽链表现为疏水性 ,无明显的亲水区域 ,可认为 *ompB* 编码蛋白是疏水性蛋白 ,见图 3。

2.5 *ompB* 编码蛋白信号肽分析 使用 SignalP

5.0 server (<http://www.cbs.dtu.dk/services/SignalP/>) 的结合了深度神经网络、条件随机分类和迁移学习方法 ,能对信号肽进行更准确的预测。*ompB* 编码蛋白经分析后 ,可以看出 Signal peptide (Sec/SPI) Likelihood 为 0.567 7 ,存在信号肽序列 ,见图 4。

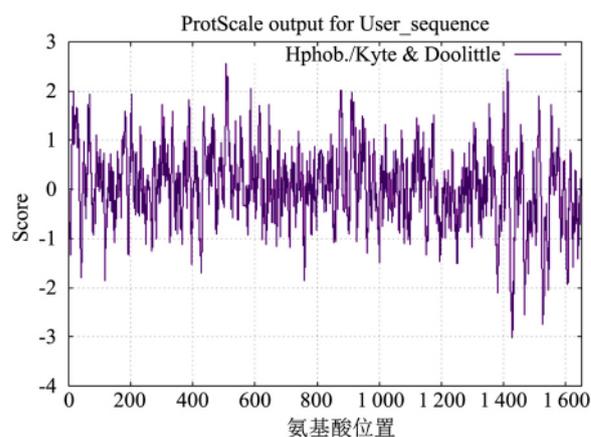


图3 ProtScale 程序分析 *ompB* 编码蛋白的疏水性结果

2.6 *ompB* 编码蛋白跨膜区分析 *ompB* 编码蛋白为跨膜蛋白 ,是一类结构独特的蛋白质 ,发挥着重要的生物学功能。使用 TMHMM 在线工具 (<http://www.cbs.dtu.dk/services/TMHMM/>) 对 *ompB* 编码蛋白跨膜结构域进行确证 ,从图 5 可以看出 ,蛋白在第 12 位到第 34 位氨基酸为跨膜区域。

2.7 *ompB* 编码蛋白功能结构域分析 使用 SMART 服务器 (http://smart.embl.de/smart/set_mode.cgi?NORMAL=1) 分析 *ompB* 编码蛋白的结构功能域。*ompB* 编码蛋白在第 704 位和第 926 位

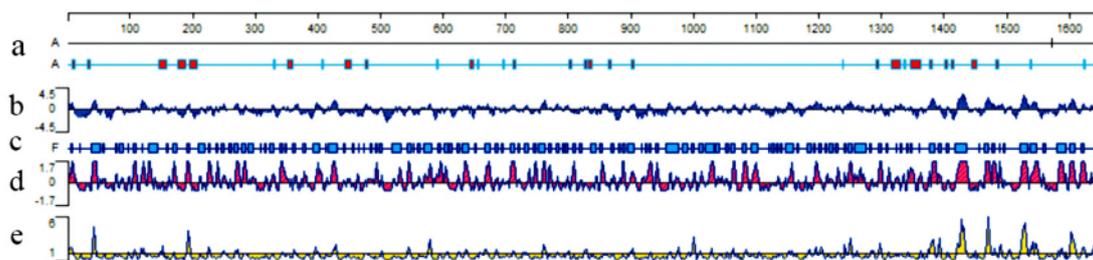


图7 *ompB* 编码蛋白二级结构和抗原表位预测(Protean 模块)

a: α -Helix prediction (Garnier-Robson and Chou-Fasman); b: Hydrophilicity prediction (Kyte-Doolittle); c: Flexible regions prediction (Karplus-Schulz); d: Surface probability prediction (Emini); e: Antigenic prediction (Jameson-Wolf)

3 讨论

本文所使用的 *ompB* 基因序列来自 2003 年本课题组在安徽省大别山区的患者急性期血液样本中分离到的一株日本斑点热立克次体近缘株, 通过对其 *ompA*、*ompB*、*geneD*、16S *rRNA*、17kD 基因进行 PCR 测序分析表明, 该菌为一株新的日本斑点热立克次体, 相关研究^[9] 成果已在国际权威杂志发表。本文通过 NCBI 的 BLAST 程序对 *ompB* 基因进行了详细分析, 发现安徽 120 分离株的 *ompB* 基因在 3 297 ~ 3 311 处与 NCBI 上的日本斑点热立克次体 (Sequence ID: AP017602.1) 序列比较, 存在连续 15 碱基的缺失, 除此之外, 该分离株还有 4 个碱基突变, 表明安徽 120 株不同于国外已经报道的日本斑点热立克次体, 具有重要的研究价值, 也提醒了国内临床医生在诊断不明原因发热疾病时, 需留意日本斑点热立克次体的感染可能。

本研究运用了 Bioedit、ORF finder、ProtParam tool、ProScale、SignalP、TMHMM 等对 *ompB* 编码蛋白的氨基酸组成、蛋白质的理化性质、疏水性、信号肽、蛋白跨膜区、功能结构域、二级结构、三级结构等进行了分析。综合来看, *ompB* 编码基因的氨基酸数目为 1 651 个, 蛋白质分子量为 167.69 ku, 共用 20 种氨基酸组成, 其中含量较多的氨基酸是 Gly, 其次为 Ala 和 Asn。该蛋白的原子组成为 C₇₃₁₂ H₁₁₆₅₈ N₂₀₅₆ O₂₄₂₄ S₁₇, 总原子数目为 23 467 个。理论等电点 pI 为 5.15, 消光系数为 58 805 或 58 680, 蛋白不稳定指数为 7.15, 脂肪指数为 88.86, 总平均亲水性 (GRA-VY) 为 0.06, 为疏水性蛋白; Signal peptide (Sec/SPI) Likelihood 为 0.567 7, 存在信号肽序列; 蛋白在第 704 位和第 926 位氨基酸之间存在一个 Pfam 区域, 与细菌黏附宿主的功能相关; 在第 1 367 位和第

1 641 位氨基酸之间有一自转运体区, 与细菌的蛋白质转运有关; 蛋白二级结构中以无规卷曲为主, 其次为 β -折叠和 α -螺旋, 分别占到 53.73%、30.83% 和 15.45%。无规则卷曲和 β -折叠是 *ompB* 编码蛋白最大的结构元件, 而 α -螺旋则散布于整个蛋白质中; 同源模建蛋白的三级结构中, 第 1 308 ~ 1 651 氨基酸部分在蛋白表面形成桶状结构, 可能具有丝氨酸蛋白酶的作用。

此外, 本研究采用了 Protean 模块和 ABCpred Prediction Server 预测了 *ompB* 编码蛋白的 B 细胞线性表位。从该蛋白的二级结构、亲水性、表面可及性、柔韧性及抗原指数 5 个方面进行预测分析。从结果可以看出, *ompB* 编码蛋白具有较强的抗原性, 分子内抗原表位较丰富, 共有 159 个可能的抗原表位, 预测分数最高 (0.94) 的为 1 269 ~ 1 284 位序列, 其次为 103 ~ 168 位的序列 (Score: 0.93); 这些结果表明 *ompB* 编码蛋白 B 表面抗原的多肽较多, 在制备亚单位疫苗时, 应当选择蛋白的柔韧性区域, 因为该区域易发生折叠和扭曲的几率高, 形成表位的可能性较大, 易与抗体进行结合^[15]; 当然除了结合生物信息学分析结果外, 选择的抗原表位肽段制备的疫苗效果如何, 需要临床试验来证实。

参考文献

- [1] Parola P, Paddock C D, Raoult D. Tick-borne rickettsioses around the world: emerging diseases challenging old concepts [J]. Clin Microbiol R 2005, 18(4): 719-56.
- [2] Guo L P, Jiang S H, Liu D, et al. Emerging spotted fever group rickettsiae in ticks, northwestern China [J]. Ticks Tick Borne Dis 2016, 7(6): 1146-50.
- [3] 牛东升, 陈香蕊. 斑点热的实验室诊断技术研究进展 [J]. 传染病信息 2008, 21(1): 33-6.
- [4] 左双燕, 唐琨, 郑元春, 等. 黑龙江林区野鼠中斑点热群立克次体的核酸检测与序列分析 [J]. 中南大学学报: 医学版,

- 2013 ,38(5):443-7.
- [5] Li W, Liu L, Jiang X, et al. Molecular identification of spotted fever group Rickettsiae in ticks collected in central China [J]. *Clin Microbiol Infect* 2009 ,15 Suppl 2:279-80.
- [6] Uchida T, Uchiyama T, Kumano K, et al. *Rickettsia japonica* sp. nov., the etiological agent of spotted fever group rickettsiosis in Japan [J]. *Int J Syst Bacteriol* ,1992 ,42(2):303-5.
- [7] Lu Q, Yu J, Yu L, et al. *Rickettsia japonica* Infections in Humans, Zhejiang Province, China, 2015 [J]. *Emerg Infect Dis* , 2018 ,24(11):2077-9.
- [8] Gaywee J, Sunyakumthorn P, Rodkvamtook W, et al. Human infection with *Rickettsia* sp. related to *R. japonica*, Thailand [J]. *Emerg Infect Dis* 2007 ,13(4):657-9.
- [9] Li J, Hu W, Wu T, et al. Japanese spotted fever in eastern china, 2013 [J]. *Emerg Infect Dis* 2018 ,24(11):2107-9.
- [10] Qi Y, Xiong X, Wang X, et al. Proteome analysis and serological characterization of surface-exposed proteins of *Rickettsia heilongjiangensis* [J]. *PLoS One* 2013 ,8(7):e70440.
- [11] Chan Y G, Cardwell M M, Hermanas T M, et al. Rickettsial outer-membrane protein B (rOmpB) mediates bacterial invasion through Ku70 in an actin, c-Cbl, clathrin and caveolin 2-dependent manner [J]. *Cell. Microbiol.* 2009 ,11(4):629-44.
- [12] Osterloh A. Immune response against rickettsiae: lessons from murine infection models [J]. *Med Microbiol Immunol* ,2017 ,206(6):403-17.
- [13] 宗佳, 刘伯玉, 陈振, 等. 日本斑点热立克次体近缘株重组 OmpB 蛋白抗原性研究 [J]. *安徽医科大学学报* ,2017 ,52(12):1768-72.
- [14] 俞海洋, 陈伟, 倪德群, 等. 肠道病毒 71 型六安地区分离株 VP1 基因的生物信息学分析及 B 细胞表位预测 [J]. *安徽医科大学学报* 2012 ,47(3):233-40.
- [15] 郑小莉, 张艺, 谢晓东, 等. 鼠 Mincle 蛋白 B 细胞抗原表位预测及多克隆抗体制备 [J]. *免疫学杂志* 2011 ,27(8):683-7.
- [16] 李江英, 白雪娟, 梁艳, 等. 用 DNASTar 软件预测 Rv1410c 结核分枝杆菌蛋白抗原表位 [J]. *细胞与分子免疫学杂志* , 2015 ,31(4):474-7.

Analysis of bioinformatics and B cell epitope of *ompB* gene and its encoded protein from *R. Japonica* strain Anhui 120

Liu Boyu, Luo Wanrong, Chen Zhen et al

(Dept of Microbiology, Anhui Medical University, Hefei 230032)

Abstract Objective To systematically predict the structure and function of the *ompB* gene and its encoded protein from the newly discovered *R. Japonica* Strain Anhui 120 isolated from the blood of clinical patients in China, so as to provide a theoretical basis for the role in pathogenic mechanisms of rickettsia, the development of diagnostic reagents and vaccines of clinical spotted fever. **Methods** Based on the *ompB* gene and its encoded amino acid sequence of *R. Japonica* strain Anhui 120, bioinformatics method was used to predict the signal peptide, transmembrane region, hydrophobicity and possible B cell epitopes. **Results** There were 15 deletions of consecutive bases of the 3 297 ~ 3 311 in the *ompB* gene of *R. Japonica* strain Anhui 120. In addition, the isolate also had 4 base mutations. The number of amino acids encoded by the gene was 1 651, the molecular weight of the protein was 167.69 ku, and there were 20 kinds of amino acids. The amino acid with more content was Gly. The theoretical isoelectric point pI was 5.15, the extinction coefficient was 58 805 or 58 680, the unstable index was 7.15, the fat index was 88.86, the total average hydrophilicity (GRAVY) was 0.06 and it was hydrophobic protein, the Signal peptide (Sec/SPI) Likelihood was 0.5677 with signal peptide sequence. There was a Pfam region and autotransporter region, which was associated with bacterial adhesion to host and protein transport respectively. The main secondary structure of the protein was random coils, followed by β -sheet and α -helix, accounting for 53.73%, 30.83% and 15.45%, respectively. In the tertiary structure of homologous model protein, the amino acid of 1 308 ~ 1 651 formed a barrel-like structure on the surface of the protein, which might have the action of serine protease. The antigenic epitope in the protein molecule was rich, and there were 159 possible epitopes. The highest predicted score (0.94) was the sequence of 1 269 ~ 1 284. **Conclusion** The composition of the *ompB* gene and its encoded protein, the secondary structure, tertiary structure and B cell epitope of *R. Japonica* strain Anhui 120 were successfully analyzed and predicted, which provides the theoretical support for the development of pathogenic mechanisms, clinical diagnostic reagents, and subunit vaccines of *R. Japonica*.

Key words *R. Japonica* strain Anhui 120; *ompB* gene; B cell epitope; bioinformatics