

张靖鹏, 江锦秀, 林裕胜, 等. 丝状支原体山羊亚种特异性蛋白基因 Mmc-3740 生物信息学分析[J]. 福建农业学报, 2018, 33(9):883—887.
ZHANG J P, JIANG J X, LIN Y S, et al. Bioinformatics of Specific Gene Mmc-3740 of *Mycoplasma mycoides* subsp. *capri* [J]. *Fujian Journal of Agricultural Sciences*, 2018, 33 (9): 883—887.

丝状支原体山羊亚种特异性蛋白基因 Mmc-3740 生物信息学分析

张靖鹏, 江锦秀, 林裕胜, 游伟, 胡奇林*

(福建省农业科学院畜牧兽医研究所, 福建 福州 350013)

摘要: 为研究丝状支原体山羊亚种 Mmc-3740 基因编码蛋白的结构与功能, 运用生物信息学软件, 对该基因编码蛋白的二级结构、跨膜区域、信号肽等生物信息进行预测及分析。结果显示, 该基因编码蛋白拥有 α 螺旋、延伸链、随机卷曲结构; 有信号肽, 亚细胞定位在细胞外或者细胞质内, 可能为分泌蛋白, 具有较高的免疫原性, 抗原表位偏高的区域主要集中在氨基酸序列的第 28~51、60~67、80~130 位处; 蛋白分子式为 $C_{607} H_{1000} N_{184} O_{261} S_5$, 分子量 15.21 kD, 等电点 (PI 值) 为 4.42, 在微生物体内半衰期较短, 为不稳定蛋白。该结果为 Mmc-3740 编码蛋白的进一步研究提供基础数据。

关键词: 丝状支原体山羊亚种; 特异性基因; 生物信息学

中图分类号: S 852.62

文献标识码: A

文章编号: 1008—0384 (2018) 09—883—05

Bioinformatics of Specific Gene Mmc-3740 of *Mycoplasma mycoides* subsp. *capri*

ZHANG Jing-peng, JIANG Jin-xiu, LIN Yu-sheng, YOU Wei, HU Qi-lin*

(Institute of Animal Husbandry & Veterinary Medicine, Fujian Academy of Agricultural Sciences, Fuzhou, Fujian 350013, China)

Abstract: Structure and function of Mmc-3740 encoded protein of *Mycoplasma mycoides* subsp. *capri* were studied based on the bioinformatics of the secondary structure, transmembrane region and signal peptide of the gene. The results showed that the protein was an alpha helix with extended random coil, and that the signal peptide located either extracellular or in cytoplasmic with high antigen index and immunogenicity. The area that showed high antigen index and immunogenicity concentrated in the regions of 28—51, 60—67 and 80—130 amino acid sequences. The molecular formula of the protein was $C_{607} H_{1000} N_{184} O_{261} S_5$ with a molecular weight of 15.21 kD and an isoelectric point of 4.42. The half-life of Mmc-3740 in microorganisms appeared to be short and unstable.

Key words: *Mycoplasma mycoides* subsp. *capri*; Mmc-3740 gene; bioinformatics

丝状支原体山羊亚种 (*Mycoplasma mycoides* subsp. *capri*, Mmc) 是丝状支原体簇的成员之一, 可引起羊支原体性肺炎 (*Mycoplasma pneumoniae* of sheep and goats, MPG)、山羊传染性无乳症 (*Caprine contagious agalactia*) 等^[1]。羊支原体肺炎在山羊和绵羊中广泛流行, 是一种危害严重的疾病^[2], 在羊群中传播迅速, 发病率为 22%~30%, 有的可高达 60%~80%, 死亡率为 15%~30%^[3], 给羊饲养业造成了严重的经济损失。该病临床表现为高热、咳嗽、纤维素性肺炎和胸膜

炎^[4], 其临床症状和病理变化与山羊支原体山羊肺炎亚种 (*M. capricolum* subsp. *capripneumoniae*, Mccp) 引起的山羊传染性胸膜肺炎相似。Mmc 与 Mccp 及山羊支原体山羊亚种 (*Mycoplasma capricolum* subsp. *capricolum*, Mcc) 同属丝状支原体簇 (mycoplasma mycoides cluster, Mm Cluster), 其基因组同源性很高, 理化特性相似, 血清学上存在交叉反应, 因此不易通过血清学方法进行准确鉴定^[5]。目前 Mmc 的检测方法主要包括支原体分离培养^[6]、PCR^[7]、qRT-PCR^[8] 等。因

收稿日期: 2018—03—22 初稿; 2018—08—23 修改稿

作者简介: 张靖鹏 (1988—), 男, 硕士研究生, 主要从事动物传染病研究 (E-mail: 870063543@qq.com)

* 通讯作者: 胡奇林 (1963—), 男, 硕士, 研究员, 主要从事动物传染病学和免疫学研究 (E-mail: hql562713@163.com)

基金项目: 国家重点研发计划项目 (2016YFD0500906); 福建省农业科学院科技创新团队建设项目 (2017-09); 福建省农业科学院青年人才创新基金 (YC20180-3)

此,寻找Mmc特异性抗原,以建立特异的血清学检测方法成为关注的焦点。目前已取得了一定的进展,已报道特异性基因LPPA基因^[9]和由10个ORF串联而成的麦芽糊精/麦芽糖(maltodextrin/maltose)基因簇^[10],但未见后续研究报道。近年来,国内外完成了Mmc、Mccp和Mcc等全基因组序列的测序分析,为新基因的寻找与分析奠定了基础。

生物信息学对基因的预测有着实验生物学无法比拟的快速和方便的优势,因此备受研究者的青睐^[11]。本研究前期通过对Mmc与丝状支原体簇的其他成员的比对发现Mmc-3740编码的氨基酸序列与丝状支原体簇的其他成员间相似性低,是Mmc所特有的基因,但其结构和功能还不清楚。因此,本研究利用生物信息学软件对Mmc-3740所编码蛋白的二级结构、疏水性、跨膜区域、信号肽等生物信息进行预测及分析,以期为该蛋白功能的深入研究提供参考。

1 材料与方法

1.1 材料

登录美国国立生物技术信息中心网站(<https://www.ncbi.nlm.nih.gov/>)获得Mmc、Mccp、Mmc LC型、Mcc的全基因组序列。

1.2 方法

1.2.1 Mmc特异性片段的筛选 利用DNAStar软件,对获得的全基因组序列进行比对寻找相似性低的区域,结合NCBI的BLAST功能,筛选出Mmc特异的基因片段。

1.2.2 Mmc-3740基因编码蛋白二级结构预测 应用SOPMA(https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html)预测蛋白质的二级结构。

1.2.3 Mmc-3740基因编码蛋白信号肽的预测 参考任冬梅^[12]的方法,利用SignalP 4.1(<http://www.cbs.dtu.dk/services/SignalP-4.1/>)和Sosui(Sosui:<http://harrier.nagahama-i-bio.ac.jp/sosui/mp1/mp1.cgi>)预测信号肽。

1.2.4 跨膜区的预测 利用DAS在线预测软件(<http://mendel.imp.ac.at/sat/DAS/DAS.html>)、TMHMM在线预测软件(<http://www.cbs.dtu.dk/services/TMHMM/>)及Sosui在线预测软件分析Mmc-3740基因编码蛋白跨膜区。

1.2.5 亚细胞定位的预测 应用CELLO version2.5(<http://cello.life.nctu.edu.tw/>)对Mmc-3740基因编码蛋白进行亚细胞定位。

1.2.6 抗原表位的预测 应用IEBD(http://www.iedb.org/home_v3.php)预测Mmc-3740基因编码蛋白的抗原表位。

1.2.7 其他信息的预测 结合ProtParam(<http://web.expasy.org/protparam/>)与Sosui分析Mmc-3740基因编码蛋白的氨基酸等电点、相对分子质量等生物信息。

2 结果与分析

2.1 Mmc-3740基因特异性

通过NCBI与DNASTAR软件的比对,发现Mmc-3740基因位于丝状支原体山羊亚种95010株(登录号:FQ377874.1)全基因组序列的第485 951~486 385 bp位,长度435 bp,为完整的开放阅读框架(Open Reading Frame, ORF),编码141个氨基酸。其编码的氨基酸序列只与Mmc的PG3株相似,相似性78%,而与丝状支原体簇的其他成员间没有相似性,为Mmc所特有的基因片段。

2.2 Mmc-3740编码蛋白二级结构预测

利用SOPMA在线软件分析,发现Mmc-3740基因编码蛋白拥有 α 螺旋、延伸链、随机卷曲结构,分布如图1,其中 α -螺旋占20.13%,主要位于其编码氨基酸序列的第1~8、14~19、47~50、137~144位,延伸链占12.08%,主要位于该氨基酸序列的第20~23、55~58、81~83、101~103位,随机卷曲占67.79%,主要位于该氨基酸序列的第24~46、59~70、71~80、85~100、104~136位(表1)。

表1 Mmc-3740编码蛋白二级结构预测结果

Table 1 Predicted Mmc-3740 secondary structure

结构	α -螺旋	3_{10} -螺旋	π -螺旋	β -折叠桥	延伸链	β -转角	弯曲区	随机卷曲	模糊区	其他
碱基数	30	0	0	0	18	0	0	101	0	0
占比/%	20.13	0.00	0.00	0.00	12.08	0.00	0.00	67.79	0.00	0.00

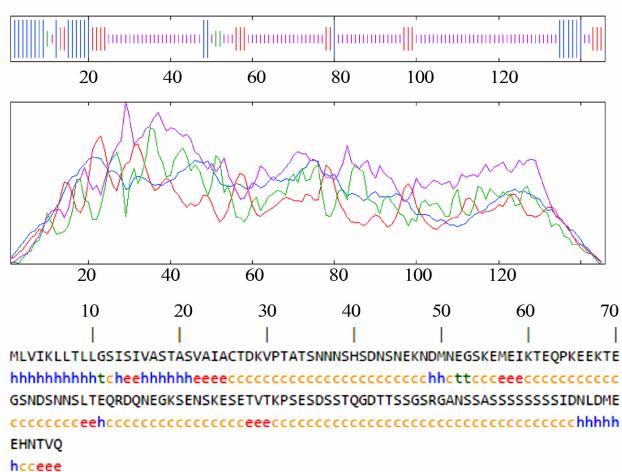


图 1 Mmc-3740 编码蛋白 α 融合、延伸链、随机卷曲结构分布

Fig. 1 Distribution of alpha helix, extension chain and random coil structure of Mmc-3740

2.3 信号肽的预测结果

利用 SignalP4.1 和 Sosui 预测信号肽, SignalP4.1 预测结果 (图 2) 显示, D 值为 0.655, 信号肽的切割位点位于氨基酸序列第 23~24 位。Sosui 预测 (图 3) 有信号肽, 位于氨基酸序列的第 1~23 位, 二者预测结果一致, 因此推测 Mmc-3740 蛋白是分泌蛋白。

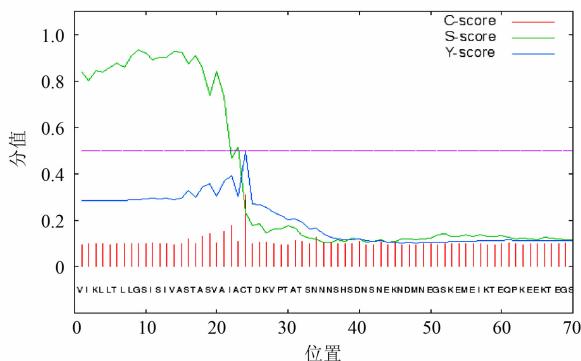


图 2 signalP4.1 预测结果

Fig. 2 SignalP4.1 prediction

No.	SHORT_ID (10)	seg.Length	Type	Signal region	1st. TM region	Topology of TM (size of domain)	Seq of TM region	ID
0001	None	149.a.a	Sec	1-23				None

图 3 SOSUJ 预测结果

Fig. 3 SOSUI prediction

2.4 跨膜区域预测

利用 DAS 在线预测软件和 TMHMM 在线预测，结果（图 4）发现，Mmc-3740 蛋白存在跨膜区，预测的跨膜区域为氨基酸序列的第 6~25、7~22、129~134 位。而 TMHMM 的预测结果（图 5）显示该蛋白没有跨膜区，均分布在膜外。利用 Sosui 进行预测（表 2），结果显示该蛋白为一种膜蛋白，有跨膜区域，位于氨基酸序列的第 4~26 位。

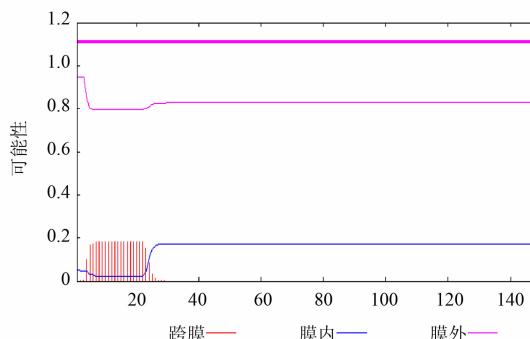


图 4 DAS 预测结果

Fig. 4 DAS prediction

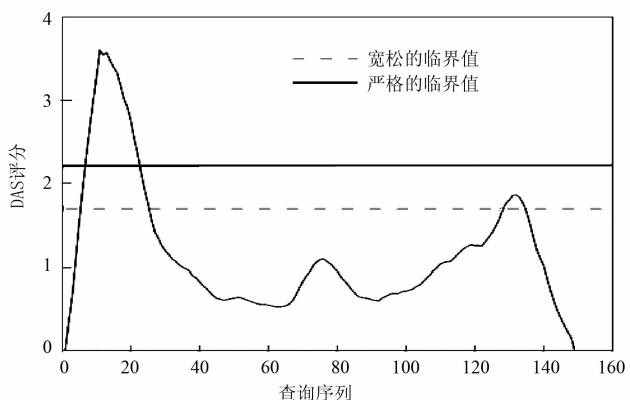


图 5 TMHMM 的预测结果

Fig. 5 TMHMM prediction

表 2 SOSUI 预测结果

Table 2 SOSUI prediction

N 端	跨膜区	C 端	类型	长度	结 果
4	LLTLLGGSISIVAS TASVAlACTD	26	一级 结构	23	该氨基酸序列是一种膜蛋白,其中有 1 个跨膜螺旋

2.5 亚细胞定位

CELLO 预测结果（表 3）可以看出 Mmc-3740 蛋白的亚细胞定位在细胞外或者细胞质内，可能是一种分泌型蛋白。

表3 CELLO 预测结果
Table 3 CELLO prediction

CELLO 预测定位	数值
细胞外	1.902 *
细胞质	1.295 *
周质	0.993
外膜	0.451
内膜	0.359

注:标有 * 号为可信值。

2.6 抗原表位预测

IEBD 预测 (http://www.iedb.org/home_v3.php) 结果显示 Mmc-3740 蛋白抗原指数平均值为 1.081, 最高 2.294, 偏高的区域主要集中于氨基酸序列的第 28~51、60~67、80~130 位(图 6)。因此, 预测该蛋白有 3 个抗原性较好的区段, 其中第 80~130 位为抗原性最强的区段。推测该蛋白可引起机体的免疫应答。将其进行体外表达后, 可用于 Mmc 检测方法的建立。

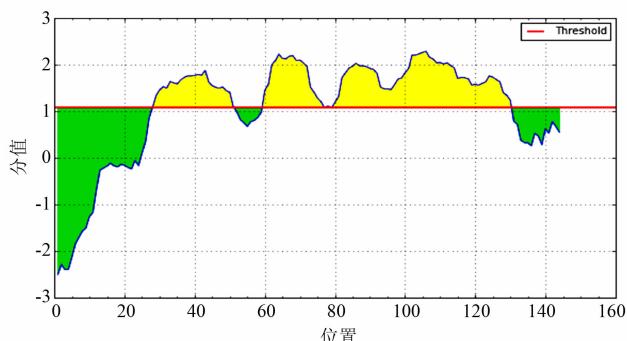


图 6 IEBD 预测抗原表位结果

Fig. 6 IEBD prediction

2.7 其他信息预测

结合 Sosui 和 ProtParam (<https://web.expasy.org/cgi-bin/protparam/protparam>) 分析得出, Mmc-3740 编码蛋白分子式: C₆₀₇H₁₀₀₀N₁₈₄O₂₆₁S₅, 总原子数量 2 057, 分子质量 15 211.99, 等电点 (PI 值) 4.42。氨基酸组成如表 4, 带负电的氨基酸残基 (Asp+Glu) 25 个, 带正电的氨基酸残基 (Arg+Lys) 12 个。平均亲疏水性: -1.121, 为亲水蛋白。半衰期预测: 10 h (大肠杆菌体内), 20 h (酵母体内), 30 h (网织红细胞内)。不稳定指数为 48.43, 该蛋白质为不稳定蛋白。

表4 Mmc-3740 编码蛋白氨基酸组成
Table 4 Amino acids of Mmc-3740

氨基酸	数量	占比/%
丙氨酸 Ala (A)	7	4.90
精氨酸 Arg (R)	2	1.40
天冬酰胺 Asn (N)	15	10.40
天冬氨酸 Asp (D)	9	6.20
半胱氨酸 Cys (C)	1	0.70
谷氨酰胺 Gln (Q)	5	3.50
谷氨酸 Glu (E)	16	11.10
甘氨酸 Gly (G)	7	4.90
组氨酸 His (H)	2	1.40
异亮氨酸 Ile (I)	6	4.20
亮氨酸 Leu (L)	6	4.20
赖氨酸 Lys (K)	10	6.90
甲硫氨酸 Met (M)	4	2.80
苯丙氨酸 Phe (F)	0	0.00
脯氨酸 Pro (P)	3	2.10
丝氨酸 Ser (S)	32	22.20
苏氨酸 Thr (T)	14	9.70
色氨酸 Trp (W)	0	0.00
酪氨酸 Tyr (Y)	0	0.00
缬氨酸 Val (V)	5	3.50
吡咯赖氨酸 Pyl (O)	0	0.00
硒半胱氨酸 Sec (U)	0	0.00

3 讨论

羊支原体性肺炎 (MPGS) 在我国多个地区发生, 发病率高、传播速度快、波及范围大, 给我国养羊业造成了重大经济损失^[13]。Mmc 作为 MPGS 的重要病原之一, 与 Mcc 等同属于丝状支原体簇的成员。该簇成员包括在遗传学和血清学上非常相近的 6 种支原体^[14], 分为 3 个亚群, 分别是山羊支原体、丝状支原体和李奇氏支原体亚群^[15], 其同源性高、血清学方面存在交叉反应。为了快速、高效地诊断 Mmc 引起的 MPGS, 寻找特异性抗原, 建立相应的检测方法成为国内外学者关注的重点。

目前已公布的 Mmc 全基因组序列只有 95 010 和 PG3 2 株, 本课题组前期通过对 Mmc 与丝状支原体簇的其他成员的基因组分析发现, Mmc-3740 编码的氨基酸序列与丝状支原体簇的其他成员间相似性低, 是 Mmc 特有的基因。同时, 经基因序列比对发现, 95 010 株 Mmc-3740 基因与 PG3 的相似性为 87%, 后续研究中用基于 95 010 株序列设计的引物可以从 Mmc 临床分离株中扩增出 Mmc-3740 基因相应长度的片段, 测序后发现相似性约为 90%。

表明 Mmc-3740 基因是 Mmc 的保守基因之一。

生物信息学预测可以破译隐藏在 DNA 序列中的遗传信息,通过对基因信息进行深入分析,预测蛋白的结构和功能,在生物、医药、农业等研究领域越来越被重视^[16]。本研究首次利用生物信息学软件对 Mmc-3740 编码蛋白进行生物信息学预测,结果显示,该基因编码蛋白除存在常见的 α 融合结构外,还有延伸链、随机卷曲结构。螺旋和折叠化学键能是能否牢固维持蛋白高级结构的重要条件,化学键能越高蛋白结构越牢固,延伸链则构成其活性位点^[17];信号肽分析发现该基因所编码的蛋白具有信号肽,新生蛋白通常在位于其 N 端的信号肽的指引下到达细胞特定区域,并由其介导跨膜转运^[18],因此推测其可能是一种分泌型蛋白。蛋白质亚细胞定位与其功能密切相关,预测蛋白的亚细胞定位,对于了解蛋白质的性质和功能具有重要意义^[19],本研究利用 CELLO 进行亚细胞定位的预测结果显示,Mmc-3740 蛋白定位于细胞外或者细胞质内,结合信号肽的预测,推测其在原核表达中能获得可溶性表达。跨膜区预测中,不同软件的预测结果不同,所用的 3 种软件都没有针对支原体的选项,因此,推测跨膜区预测软件在支原体蛋白的预测并不成熟,其结果只能作为参考。与实验验证抗原性相比,通过生物信息学方法预测抗原表位耗时短且易于操作^[20]。Mmc-3740 蛋白抗原表位预测结果显示其抗原指数较高,亲水性预测显示其表面拥有亲水基团,抗原亲水性决定 B 细胞抗原表位^[21],推测这些位点容易成为抗体识别的抗原表位区域,预测该蛋白抗原性良好,体外表达的产物可用于 Mmc 检测方法的建立或作为 Mmc 的保护性抗原,用于该病的免疫防控。稳定性预测结果显示该蛋白为不稳定蛋白。

综上所述, Mmc-3740 编码蛋白为 Mmc 所特有的抗原性良好的分泌型蛋白, 推测可用于 Mmc 特异检测方法的建立。

参考文献:

- [1] TATAY -DUALDE J, HAM P V D, FE C D L, et al. Multilocus sequence typing of *Mycoplasma mycoides*, subsp. *capri*, to assess its genetic variability in a contagious agalactia endemic area [J]. *Veterinary Microbiology*, 2016, 191:60–64.
- [2] CHUN-GUANG X U, HAO Y Q, WANG M Y, et al. Advance In Treatment And Prevention Of Mycoplasmal Pneumonia Of Sheep And Goat [J]. *Progress In Veterinary Medicine*, 2014, 35(1):81–85.
- [3] 高鹏程, 储岳峰, 赵萍, 等. 山羊支原体肺炎研究进展 [J]. 甘肃畜牧兽医, 2007, 37 (6): 43–46.
- [4] OIE A H S. Manual of diagnostic tests and vaccines for terrestrial animals (mammals, birds and bees) [J]. *Bulletin Office International Desépidémies*, 2015 (1): 1092–1106.
- [5] THIAUCOURT F, BÖLSKE G. Contagious caprine pleuropneumonia and other pulmonary mycoplasmoses of sheep and goats [J]. *Revue Scientifique Et Technique*, 1996, 15 (4): 1397–1414.
- [6] LEMCKE R M. The relationship of a caprine mycoplasma to PG3, the type strain of *Mycoplasma mycoides* subsp. *capri* [J]. *Research in Veterinary Science*, 1974, 16 (1): 119–21.
- [7] D'ANGELO A R, DI P A, DI F G, et al. Experimental infection of goats with an unusual strain of *Mycoplasma mycoides* subsp. *capri* isolated in Jordan: comparison of different diagnostic methods [J]. *Veterinaria Italiana*, 2010, 46 (2): 189–207.
- [8] 林裕胜, 江锦秀, 张靖鹏, 等. 丝状支原体山羊亚种 SYBR Green I qRT-PCR 快速检测方法的建立 [J]. 农业生物技术学报, 2017, 25 (11): 1895–1902.
- [9] MONNERAT M P, THIAUCOURT F, POVEDA J B, et al. Genetic and Serological Analysis of Lipoprotein LppA in *Mycoplasma mycoides* subsp. *mycoides* LC and, *Mycoplasma mycoides* subsp. *capri* [J]. *Clinical & Diagnostic Laboratory Immunology*, 1999, 6 (2): 224.
- [10] 徐春光, 郝永清, 郝瑞霞, 等. 丝状支原体山羊亚种 PG3 株的全基因组序列测定与分析 [J]. 微生物学通报, 2017, 44 (11): 2669–2678.
- [11] ZHANG B, PAN X, ANDERSON T A. Identification of 188 conserved maize microRNAs and their targets [J]. *Febs Letters*, 2006, 580 (15): 3753.
- [12] 任冬梅. 猪肺炎支原体 168 株疫苗候选抗原蛋白 Mhp168-116、Mhp168-341、Mhp168-322 和 Mhp168-549 的原核表达及纯化 [D]. 重庆: 西南大学, 2014.
- [13] 王华, 杨发龙, 王永, 等. 山羊支原体肺炎流行病学调查 [J]. 中国畜牧兽医, 2011, 38 (1): 210–214.
- [14] NICHOLAS R A J. Improvements in the diagnosis and control of diseases of small ruminants caused by mycoplasmas [J]. *Small Ruminant Research*, 2002, 45 (2): 145–149.
- [15] 储岳峰. 我国山羊(接触)传染性胸膜肺炎病原学、流行病学研究及灭活疫苗的研制[D]. 北京: 中国农业科学院, 2011.
- [16] 黄科, 曹家树, 吴秋云, 等. 生物信息学 [J]. 情报学报, 2002, 21 (4): 491–496.
- [17] 唐媛, 李春花, 张媛, 等. 蛋白质的二级结构预测研究进展 [J]. 现代生物医学进展, 2013, 13 (26): 5180–5182.
- [18] 彭佳师, 龚继明. 信号肽与蛋白质的分选转运 [J]. 植物生理学报, 2011, 47 (1): 9–17.
- [19] 吴燕, 袁海文, 王琦, 等. 绵羊肺炎支原体贵州株 P113 基因的克隆与生物信息学分析 [J]. 西北农业学报, 2016, 25 (3): 317–327.
- [20] 黄艳新, 鲍永利, 李玉新. 抗原表位预测的免疫信息学方法研究进展 [J]. 中国免疫学杂志, 2008, 24 (9): 857–861.
- [21] 刘争春, 王颖, 王树军, 等. 人卵巢癌相关抗原 OVA66 的 B 细胞表位预测 [J]. 实用医学杂志, 2009, 25 (4): 513–516.

(责任编辑: 张 梅)