

李渊源, 孙琦, 杨齐, 等. 非洲猪瘟病毒 T、B 细胞表位的免疫信息学预测与分析 [J]. 畜牧与兽医, 2024, 56 (5): 99-106.

LI Y Y, SUN Q, YANG Q, et al. Immunoinformatic prediction and analysis based on T-cell and B-cell epitopes in the African swine fever virus [J]. Animal Husbandry & Veterinary Medicine, 2024, 56 (5): 99-106.

## 非洲猪瘟病毒 T、B 细胞表位的免疫信息学预测与分析

李渊源<sup>1</sup>, 孙琦<sup>1</sup>, 杨齐<sup>2</sup>, 黄转青<sup>1</sup>, 张莹<sup>1</sup>, 徐风华<sup>1\*</sup>

(1. 解放军总医院医疗保障中心药剂科药学基础研究室, 北京 100853;

2. 解放军总医院医疗保障中心药剂科药品保障室, 北京 100853)

**摘要:** 旨在通过免疫信息学方法预测非洲猪瘟病毒 (ASFV) 结构蛋白的 T、B 细胞表位, 为非洲猪瘟 (ASF) 表位疫苗的设计研制提供参考。从 NCBI 和 RCSB 数据库获取 ASFV 蛋白质序列和三维结构, 利用 IEDB、DTU Health Tech 等数据库的生物信息学工具对 ASFV 的 5 种结构蛋白 p72、p17、p49、M1249L 和 H240R 的细胞毒性 T 细胞表位、线性 B 细胞和构象 B 细胞表位进行鉴定。结果显示: 5 种蛋白质均为亲水性, 二级结构以无规则卷曲为主, 仅 M1249L 例外; 预测到抗原性良好、无毒、无致敏的细胞毒性 T 细胞优势表位 27 个, 线性 B 细胞优势表位 35 个; 预测到仅针对 p72 的构象 B 细胞优势表位 2 个。结论: ASFV 的 5 种蛋白质可能具有多个潜在 T、B 细胞表位, 其中 B 细胞表位相对占优, 5 种蛋白质中 p72 和 M1249L 最具疫苗研发前景, 可结合蛋白质相关参数信息为构建 ASF 表位疫苗提供参考。

**关键词:** 非洲猪瘟病毒; 预测; 表位; T 细胞; B 细胞; 免疫信息学

中图分类号: S852.4 文献标志码: A 文章编号: 0529-5130(2024)05-0099-08

## Immunoinformatic prediction and analysis based on T-cell and B-cell epitopes in the African swine fever virus

LI Yuanyuan<sup>1</sup>, SUN Qi<sup>1</sup>, YANG Qi<sup>2</sup>, HUANG Zhuanqing<sup>1</sup>, ZHANG Ying<sup>1</sup>, XU Fenghua<sup>1\*</sup>

(1. Pharmaceutical Sciences Research Division, Department of Pharmacy, Medical Supplies Center, PLA General Hospital, Beijing 100853, China;

2. Drug Supplies Division, Department of Pharmacy, Medical Supplies Center, PLA General Hospital, Beijing 100853, China)

**Abstract:** This study was to predict the T-cell and B-cell epitopes of African swine fever virus (ASFV) proteins by immunoinformatic methods and to provide reference for the design and development of ASF epitope vaccines. The sequence and the three-dimensional structure of ASFV proteins were obtained by searching NCBI and RCSB databases. The cytotoxic T-cell epitopes, linear B-cell, and conformational B-cell epitopes of the five structural ASFV proteins, p72, p17, p49, M1249L, and H240R, were identified using bioinformatic tools in IEDB, DTU Health Tech and other databases. The results showed that five proteins were all hydrophilic proteins. The secondary structure of the five proteins was mainly random coil, except M1249L. 27 potential epitopes of the cytotoxic T cells and 35 potential epitopes of the linear B cells were predicted, with good antigenicity, no toxicity, and no allergen. Two conformational B-cell potential epitopes only from p72 were also predicted. The five structural proteins of ASFV might have multiple potential T-cell and B-cell epitopes, among which B-cell epitopes were dominant. In the five proteins, p72 and M1249L were the most promising for designing and developing vaccines. Combined with the protein-related information, the antigen epitopes might be used to provide a research reference for the construction of ASF epitope vaccines.

**Keywords:** African swine fever virus; prediction; epitope; T cell; B cell; immunoinformatic

非洲猪瘟病毒 (African swine fever virus, ASFV), 一种在生猪之间传播的双链 DNA 病毒, 正二十面体球形对称结构, 由外膜、蛋白质衣壳、脂质

内膜、核壳、基因组 5 层结构组成, 包含 3 万余个蛋白质亚基, 其主要衣壳蛋白为结构蛋白 p72<sup>[1]</sup>。其强毒力毒株对生猪致死率接近 100%, 传染性强, 致病性高, 暂时没有特效药和商业化疫苗, 一经发现必须立即扑杀生猪<sup>[2-3]</sup>, 对于生猪以及养殖户来说, 近乎“灭顶之灾”。该疫情最早出现在非洲肯尼亚, 于 2018 年传入我国, 近 1 年的时间波及 31 个省份, 发病率最高可达 30.4%, 给我国的养猪业敲响了警

收稿日期: 2023-03-09; 修回日期: 2024-03-17

基金项目: 国家自然科学基金面上项目 (82171814)

第一作者: 李渊源, 男, 博士研究生

\* 通信作者: 徐风华, 研究员, 主要从事药物新剂型和新型递药

系统研发, E-mail: xufh@301hospital.com.cn。

钟<sup>[4-5]</sup>。研究人员已经先后尝试灭活疫苗、减毒活疫苗、DNA 疫苗、重组疫苗和亚单位疫苗等,但是均未完全成功。尽管 ASFV 的 3D 精细结构已经被破译,但是大多数 ASFV 基因功能尚未知晓,由于病毒的强变异性以及致弱毒株的生物安全隐患等问题限制了灭活以及减毒活疫苗的应用,针对 ASFV 的特效防护和及时控制依然任重而道远。

表位亚单位疫苗通过利用病毒表面的抗原表位,诱发生猪产生免疫保护,本身不含病毒核酸,生物安全性更高,正逐渐获得研究人员的关注。抗原表位是抗原分子中决定抗原特异性的特殊化学基团,T 细胞表位作为线性表位,与主要组织抗原复合物(major histocompatibility complex, MHC)结合,并能被 CD8<sup>+</sup>(MHC-I)和 CD4<sup>+</sup>(MHC-II) T 细胞受体识别<sup>[6]</sup>; B 细胞表位则被 B 细胞表面受体或抗体特异性识别并结合,既包括线性表位,也有构象表位,位于抗原分子表面,激活后引起免疫应答,在疫苗研制、单克隆抗体和诊断试剂开发等方面扮演了重要角色。随着生物信息学技术的不断发展,依托免疫信息学工具预先对抗原蛋白质或基因信息进行严格的预测或筛选,挖掘和识别 T、B 细胞的表位序列,能够较大程度地提高抗原蛋白质中优势表位的发现效率,免疫信息学已成为表位预测分析不可或缺的工具,能够高效快速地筛选抗原蛋白序列中的候选表位,对研究设计表位疫苗具有重要意义<sup>[7-8]</sup>。本研究旨在揭示 ASFV 结构蛋白的 T 细胞和 B 细胞表位肽的生物学功能,探讨其抗原性,结合免疫信息学,从序列筛选和表位预测的角度,为 ASF 疫苗的研发提供参考途径和可靠资料。

## 1 材料与方法

### 1.1 蛋白质序列和性质

从美国国家生物信息中心以 Fasta 格式获取 ASFV 分离株的基因组序列(<https://www.ncbi.nlm.nih.gov/protein/?term=ASFV>),序列号:NC\_001659.2,检索其 5 种主要的结构蛋白,包括 4 种衣壳蛋白(p72、p17、p49 和 M1249L)和 1 种五邻体蛋白(H240R)<sup>[9]</sup>,使用 ExPasy Sever 中在线分析工具 ProtParam(<https://web.expasy.org/protparam/>)分析每种蛋白的分子量、等电点、半衰期、稳定性、亲水性等理化性质,并使用 ProtScale(<https://web.expasy.org/protscale/>)对蛋白质具体位置上氨基酸的亲/疏水情况进一步分析,相关参数选择默认值<sup>[10]</sup>。

### 1.2 蛋白质二级结构预测

分别使用 SOPMA([https://npsa-prabi.ibcp.fr/NPSA/npsa\\_sopma.html](https://npsa-prabi.ibcp.fr/NPSA/npsa_sopma.html))和 PSIPRED 4.0(<http://bioinf.cs.ucl.ac.uk/psipred/>)在线服务器预测分析 5 种

蛋白质 p72、p17、p49、M1249L 和 H240R 的二级结构,包括  $\alpha$ -螺旋、 $\beta$ -折叠、 $\beta$ -转角和无规则卷曲。

### 1.3 T 细胞表位预测

使用生物信息学工具 NetMHCpan3.0(<http://tools.iedb.org/processing/>)和 NetMHCpan 4.1(<https://services.healthtech.dtu.dk/service.php?NetMHCpan-4.1>)预测 p72、p17、p49、M1249L 和 H240R 这 5 种蛋白质中潜在的细胞毒性 T 淋巴细胞(cytotoxic T lymphocyte, CTL)表位<sup>[11-12]</sup>。选取 75 个猪白细胞抗原(swine leukocyte antigen, SLA) I 类等位基因,其中 SLA-1 等位基因 23 个,SLA-2 等位基因 26 个,SLA-3 等位基因 21 个,SLA-6 等位基因 5 个,设定肽段长度通常为 8~14 个氨基酸,通过抗原肽与 MHC-I 类分子间结合力、结合蛋白酶体、抗原加工相关蛋白转运,以 IC<sub>50</sub><500 nm 和亲和力预测分数排名前 2%作为潜在多肽疫苗候选表位的评价标准。通过免疫表位数据库 IEDB 中 MHC-I Immunogenicity 工具(<http://tools.iedb.org/immunogenicity/>)进一步考察候选表位的免疫原性,筛选免疫原性得分超过 0.2 的表位作为 CTL 表位的预测结果<sup>[13]</sup>。结合 SOPMA 或 PSIPRED 4.0 预测二级结构的结果,避开  $\alpha$ -螺旋、 $\beta$ -折叠等较难作为表位的区域。

### 1.4 B 细胞表位预测

采用两种常用的连续/线性 B 淋巴细胞(linear B lymphocyte, LBL)表位预测工具 ABCpred([https://webs.iitd.edu.in/raghava/abcpred/ABC\\_submission.html](https://webs.iitd.edu.in/raghava/abcpred/ABC_submission.html))和 Bepipred 2.0(<http://tools.iedb.org/bcell/>)综合判断。设置 ABCpred 服务器预测阈值为 0.85,氨基酸肽段长度设置为 16<sup>[14]</sup>。设置 Bepipred 2.0 预测阈值为 0.5,肽段长度为 5~30 个氨基酸<sup>[15]</sup>。LBL 表位经上述两种预测工具筛选后,通过 iBCE-EL 预测服务器(<http://www.thegleelab.org/iBCE-EL/>)继续验证<sup>[16]</sup>。

从蛋白质资源数据库 UniProt(<https://www.uniprot.org/>)<sup>[17]</sup>和 RCSB 蛋白质数据库(<https://www.rcsb.org/>)<sup>[9]</sup>以 PDB 格式获取 ASFV 中 5 种蛋白质(p72、p17、p49、M1249L 和 H240R)的三维结构信息。使用 Discotope 2.0(<https://services.healthtech.dtu.dk/service.php?DiscoTope-2.0>)和 ElliPro(<http://tools.iedb.org/ellipro/>)综合判断预测不连续/构象 B 淋巴细胞(conformational B lymphocyte, CBL)表位。设置 Discotope 2.0 鉴定阈值为 1.9<sup>[18]</sup>。设置 ElliPro 预测最低阈值为 0.5,不连续最大距离设置为 6 埃<sup>[19]</sup>。

### 1.5 表位理化性质评价

利用在线服务器 Vaxigen v2.0(<http://www.ddg-pharmfac.net/vaxijen/>)进一步考察候选 CTL、LBL 表

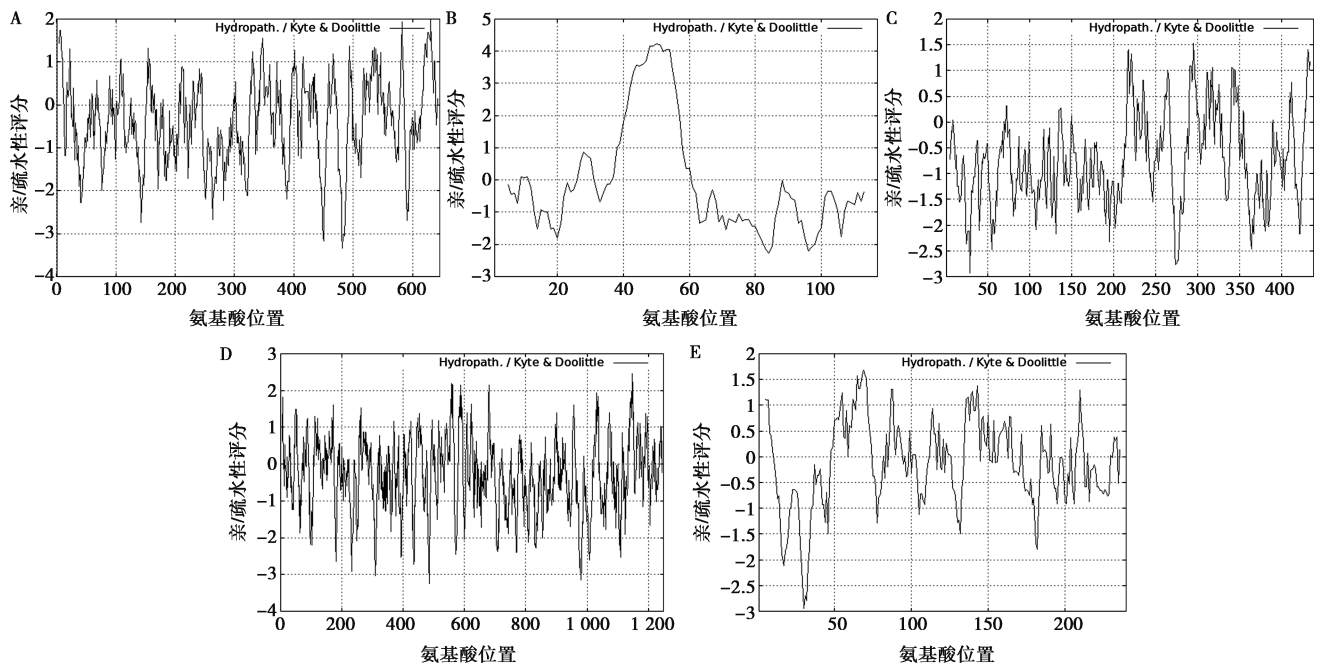
位的抗原性, 设置 Vaxigen 服务器抗原性评价阈值为 0.4, 选择目标微生物为病毒模型<sup>[20]</sup>。为避免注射后的不良反应, 针对已经筛选出的线性表位评价毒性和致敏性, 毒性评价采用在线服务器 ToxinPred (<https://webs.iitd.edu.in/raghava/toxinpred>), 设置毒性评价阈值为 0<sup>[21]</sup>。致敏性评价采用在线服务器 AllerTOP v2.0 (<http://www.ddg-pharmfac.net/AllerTOP/>)<sup>[22]</sup>。

## 2 结果与分析

### 2.1 蛋白质序列和性质

由 NCBI 数据库检索获得 p72 (pB646L) 蛋白序

列号为 NP\_042775.1, p17 (pD117L) 蛋白序列号为 NP\_042800.1, p49 (pB438L) 蛋白序列号为 NP\_042769.1, M1249L 蛋白序列号为 NP\_042754.1, H240R 蛋白序列号为 NP\_042812.1。通过 ProtParam 对 5 种蛋白质的理化性质进行分析, M1249L 的氨基酸数量最多, 为 1 249 个, 其分子量、原子总数也是最多的; 多数蛋白质结构被预测为不稳定, 只有 p72 为稳定; 5 种蛋白质预测半衰期均超过 10 h, 均为亲水性蛋白。使用 ProtScale 进一步对具体位置氨基酸的亲/疏水性进行预测, 如图 1 所示, 5 种蛋白质中亲水性氨基酸占据相对多数。



A. p72; B. p17; C. p49; D. M1249L; E. H240R。

图 1 ProtScale 预测 5 种蛋白质的亲/疏水性

### 2.2 蛋白质二级结构分析

使用 SOPMA 预测 5 种蛋白质的二级结构, 包括  $\alpha$ -螺旋、 $\beta$ -折叠、 $\beta$ -转角和无规则卷曲, 5 种蛋白质的二级结构中, 除 M1249L 外, 无规则卷曲占据多数, 为 40%~60%,  $\alpha$ -螺旋约 20%~30% (除了 M1249L),  $\beta$ -折叠为 0~30%,  $\beta$ -转角则低于 10%; 在 M1249L 中,  $\alpha$ -螺旋约 50%, 无规则卷曲仅占 23%。使用 PSIPRED 4.0 预测 5 种蛋白质的二级结

构, 包括  $\alpha$ -螺旋、 $\beta$ -折叠和无规则卷曲, 除 M1249L 外, 无规则卷曲依然占据多数, 为 40%~70%,  $\alpha$ -螺旋和  $\beta$ -折叠范围均较大,  $\alpha$ -螺旋为 0~60%,  $\beta$ -折叠为 0~50%。两种预测方法分析 5 种蛋白质二级结构组成结果如表 1 所示, 不论 SOPMA 还是 PSIPRED 4.0 预测二级结构, 不同于其他 4 种蛋白质以无规则卷曲为主, M1249L 中  $\alpha$ -螺旋占据多数, 提示该蛋白质的二级结构可能比较稳定。

表 1 5 种蛋白质二级结构组成

蛋白质名称	SOPMA								PSIPRED 4.0					
	α-螺旋		β-折叠		β-转角		无规则卷曲		α-螺旋		β-折叠		无规则卷曲	
	氨基酸数量	占据比例/%	氨基酸数量	占据比例/%	氨基酸数量	占据比例/%	氨基酸数量	占据比例/%	氨基酸数量	占据比例/%	氨基酸数量	占据比例/%	氨基酸数量	占据比例/%
p72	125	19.35	163	25.23	34	5.26	324	50.16	66	10.22	157	24.30	423	65.48
p17	35	29.92	20	17.09	4	3.42	58	49.57	55	47.01	5	4.27	57	48.72
p49	128	29.22	30	6.85	32	7.31	248	56.62	132	30.14	14	3.20	292	66.66
M1249L	620	49.64	213	17.05	124	9.93	292	23.38	719	57.57	87	6.96	443	35.47
H240R	50	20.83	72	30.00	18	7.50	100	41.67	7	2.92	116	48.33	117	48.75

### 2.3 优势 CTL 表位分析

通过 NetMHCpan 3.0 和 NetMHCpan 4.1 预测 CTL 表位, 得分越低, 其成为优势表位可能性越大, 消除冗余后, 5 种蛋白质 p72、p17、p49、M1249L 和 H240R 分别获得 CTL 候选表位 101、14、55、260、58 个, 进一步使用 MHC-I Immunogenicity 评估免疫原性, 选择得分超过 0.2 的表位, 同时除去二级结构中 α-螺旋、β-折叠区域, 5 种蛋白质分别得到 CTL

候选表位 29、5、3、58、12 个。

分别使用 Vaxigen v2.0、ToxinPred、AllerTOP v2.0 评价 CTL 候选表位的抗原性、毒性和致敏性, 筛选得到抗原性良好、无毒性、无致敏性的 CTL 候选表位合计 27 个, 5 种蛋白质分别为 12、0、0、12、3 个, 其中 3 种蛋白质的序列、具体位置和各项评分如表 2 所示。

表 2 3 种蛋白质优势 CTL 表位预测结果

蛋白质名称	肽段序列	长度/aa	起始位置	结束位置	IC <sub>50</sub> 评分	亲和力 百分比	免疫原性 评分	抗原性 评分	毒性	致敏性
p72	AMMITFALK	9	579	587	232.4	1.156 2	0.251 46	1.456 2	无	无
	ASAINFLLL	9	627	635	349.9	1.213 2	0.234 68	0.794 2	无	无
	KLTFGIPQY	9	89	97	420.4	0.168 4	0.211 64	0.542 0	无	无
	RAREFYISW	9	602	610	148.4	0.574 1	0.243 73	1.543 7	无	无
	RRNIRFKPWF	10	388	397	449.5	1.892 1	0.213 28	1.653 2	无	无
	SASAINFLL	9	626	634	475.3	0.389 2	0.201 50	0.655 2	无	无
	SASAINFLLL	10	626	635	303.0	1.424 9	0.203 33	0.859 4	无	无
	SQIEETHLV	9	50	58	476.0	0.027 2	0.298 81	1.089 6	无	无
	SRAREFYISW	10	601	610	336.0	0.280 0	0.288 46	1.224 2	无	无
	SVSIPFGER	9	347	355	227.8	0.458 2	0.266 72	1.512 7	无	无
	SVSIPFGERF	10	347	356	467.0	0.318 5	0.330 06	0.850 2	无	无
	VSIPFGERF	9	348	356	402.2	0.832 1	0.292 68	0.952 1	无	无
M1249L	ASLEFNTFYAF	11	922	932	377.1	1.928 0	0.372 59	0.528 9	无	无
	ASLEFNTFYAFY	12	922	933	402.0	3.066 5	0.454 59	0.469 3	无	无
	AYVTFLLEK	9	620	628	421.5	1.169 8	0.205 16	0.454 9	无	无
	FLTITIHGM	9	604	612	366.6	1.084 7	0.356 70	0.505 5	无	无
	ISPAVIEAI	9	1 029	1 037	149.9	1.769 0	0.308 61	0.749 7	无	无
	KACTFIHFGK	10	535	544	282.8	7.160 0	0.414 21	1.102 2	无	无
	KASLEFNTFYAFY	13	921	933	418.6	8.908 7	0.390 08	0.410 3	无	无
	KEFLTITI	8	602	609	430.2	0.286 1	0.222 68	0.563 7	无	无
	KNIRLIDFLF	10	43	52	397.7	1.972 4	0.327 28	1.930 2	无	无
	LIDFLFTLK	9	47	55	480.7	3.485 1	0.250 68	1.709 2	无	无
	REFALTELNTI	11	1 166	1 176	272.5	0.729 0	0.208 65	0.791 9	无	无
	REFALTELNTII	12	1 166	1 177	388.5	1.654 9	0.295 50	0.550 5	无	无

续表2

蛋白质名称	肽段序列	长度/aa	起始位置	结束位置	IC <sub>50</sub> 评分	亲和力百分比	免疫原性评分	抗原性评分	毒性	致敏性
H240R	KMTVFPFMIPFPL	13	63	75	482.7	24.579 4	0.311 82	0.761 6	无	无
	MTVFPFMIPFPL	12	64	75	349.7	11.006 1	0.274 88	0.544 7	无	无
	MTVFPFMIPFPLQK	14	64	77	403.7	10.566 3	0.208 72	0.589 4	无	无

### 2.4 优势 LBL 表位分析

分别使用 ABCpred 和 Bepipred 2.0 预测 5 种蛋白质 p72、p17、p49、M1249L 和 H240R 的 LBL 表位,按预测阈值筛选表位,除去肽段长度少于 5 和多于 30 的表位序列,5 种蛋白质分别获得了 LBL 候选表位 35、3、18、61、13 个,其预测得分平均值分别为 0.496、0.501、0.555、0.489、0.501,进一步使用 iBCE-EL 逐个验证,除去二级结构中较难形成表位

的  $\alpha$ -螺旋、 $\beta$ -折叠区域,5 种蛋白质分别获得 LBL 候选表位 34、3、17、56、11 个。

分别使用 Vaxigen v2.0、ToxinPred、AllerTOP v2.0 评价 LBL 候选表位的抗原性、毒性和致敏性,筛选得到抗原性良好、无、无致敏性的 LBL 候选表位合计 35 个,5 种蛋白质分别为 10、1、5、17、2 个,其序列、具体位置和预测值如表 3 所示。

表 3 5 种蛋白质优势 LBL 表位预测结果

蛋白质名称	肽段序列	长度/aa	起始位置	结束位置	表位概率值	抗原性评分	毒性	致敏性
p72	AMMITFALKPREEYQP	16	579	594	0.668 6	1.618 9	无	无
	EFYISWDTDYVGSITT	16	605	620	0.956 0	0.404 6	无	无
	FGRPIVPCTKNAYRNL	16	161	176	0.946 1	0.453 5	无	无
	KPYVPVGFYKNKVRPHGTGPTLGNKL	26	65	90	0.360 8	0.490 1	无	无
	PFGRPIVPGTKN	12	160	171	0.956 0	0.773 7	无	无
	QPTHHAEIFQDRDTALPDACSSISDISP	29	499	527	0.454 5	1.274 0	无	无
	RFIPGRPSRRNIRFKP	16	380	395	0.956 0	1.224 4	无	无
	SSYIPFHYGGNAIKTP	16	559	574	0.770 0	0.515 5	无	无
	TWNISDQNPQHHRDWH	16	473	488	0.956 0	0.965 6	无	无
TWNISDQNPQHHRDWHK	17	473	489	0.956 0	0.599 9	无	无	
p17	EPHHHPVFFRKRKNS	16	83	98	0.956 0	1.051 8	无	无
p49	GGIGDSKLDSTFPKDF	16	391	406	0.851 3	0.662 3	无	无
	GIAGRGIPLGNPHVKP	16	231	246	0.956 0	1.427 5	无	无
	KLDSTFPKDFNASSVP	16	397	412	0.956 0	0.456 9	无	无
	NGYIPHKDVYNILCLA	16	206	221	0.956 0	0.722 0	无	无
	TGVPMLGPLPPKDSQH	16	260	275	0.956 0	0.756 4	无	无
M1249L	AELITTEYLNIIKKQWE	16	379	394	0.956 0	0.738 2	无	无
	DHLSLNNEEIEA	12	16	27	0.583 7	0.973 2	无	无
	EGRSYADIREGQCAPP	16	1 041	1 056	0.956 0	0.857 3	无	无
	EIGSFSPETLGYVASGA	17	147	163	0.956 0	1.086 1	无	无
	ESLRVWWGGRDEEKTS	16	759	774	0.956 0	0.539 1	无	无
	FYNIFDPNTGRADQRT	16	821	836	0.956 0	0.458 7	无	无
	GAMEGRSYADIREGQGAPPPPTSMDDPRL	29	1 038	1 066	0.724 6	0.777 2	无	无
	HAEQREKEQMILQQVD	16	306	321	0.874 0	0.690 4	无	无
	HRKLDREDQGSNISNI	16	228	243	0.956 0	0.668 6	无	无
	HRKLDREDQGSNISNIVETEEIEPEE	26	228	253	0.392 7	0.765 4	无	无
IHQNLQGLSVQDD	14	660	673	0.595 4	0.515 0	无	无	

续表3

蛋白质名称	肽段序列	长度/aa	起始位置	结束位置	表位概率值	抗原性评分	毒性	致敏性
M1249L	IQIKKDMTPSSETQP	16	991	1 006	0.956 0	0.427 5	无	无
	LTEILLGPMYDYAATV	16	678	693	0.565 5	0.503 5	无	无
	PPPTSMDDPRLMAVD	16	1 055	1 070	0.956 0	0.689 6	无	无
	QEDRTADVG	9	508	516	0.956 0	1.625 1	无	无
	TMKACTFIHFGKLV DV	16	533	548	0.928 3	0.771 1	无	无
	VVPIVSEKKKELRVRPSTR	19	88	106	0.956 0	1.175 8	无	无
H240R	LDSQEKRRHGHPFSEFKLPYGTGA	25	25	49	0.546 8	1.467 7	无	无
	TRAAPKMASKKEHQ	14	8	21	0.919 5	0.933 0	无	无

## 2.5 优势 CBL 表位分析

从 UniProt 数据库检索 ASFV 5 种蛋白质 p72、p17、p49、M1249L 和 H240R 编号分别为 P22776、Q89424、Q65165、Q65152、Q65190，仅发现 p72 有可以获取的结构信息，其余 4 种蛋白质都还没有可以获得的结构信息。

从 RCSB 数据库检索获得 p72 蛋白 pdb 标识号 6L2T，其 3D 结构如图 2 所示。通过 Discotope 2.0 预测 p72 的 CBL 表位，从 1 889 个氨基酸中获得 189 个 CBL 表位氨基酸，其中 A 链 64 个，B 链 66 个，C 链 59 个。通过 ElliPro 预测 p72 的 CBL 表位，得分越高，包含在球体模型内氨基酸残基越多，溶剂可溶性越强，成为候选表位可能性越大，p72 获得 CBL 优势表位 2 个，优势表位 1 包含 1 075 个氨基酸残基，预测评分 0.698；优势表位 2 包含 10 个氨基酸残基，预测评分 0.642，其氨基酸序列为 C: H282, C:

P284, C: E285, C: N286, C: I287, C: Q288, C: T289, C: A290, C: G291, C: K292, 2 个优势 CBL 表位的 3D 结构映射如图 3 所示。

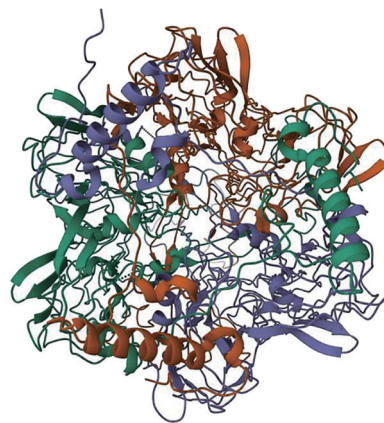
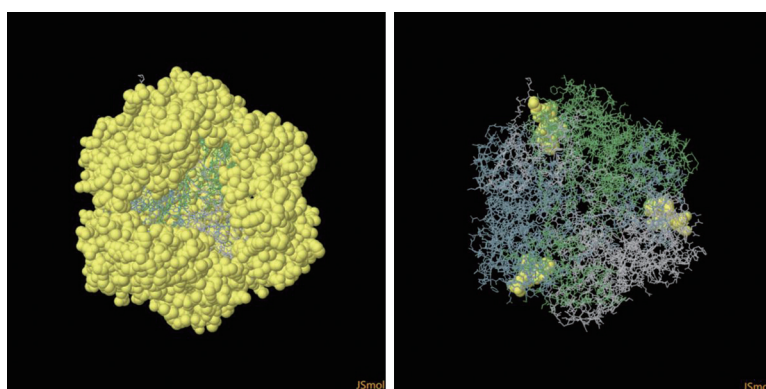


图 2 RCSB 蛋白质 p72 的 3D 结构



A. p72 优势 CBL 表位 1 (1 075 个氨基酸残基); B. p72 优势 CBL 表位 2 (10 个氨基酸残基)。

图 3 ElliPro 预测 p72 优势 CBL 表位的 3D 结构映射

## 3 讨论

由于 ASFV 具有高传染性、高致死率和复杂的流行性<sup>[23]</sup>，只靠扑杀生猪很难控制疫情传播，并且成

本高、损失大，因此迫切需要一种安全、有效，可大规模应用的疫苗，控制疫情传播，减少经济损失。表位疫苗作为亚单位疫苗，以线性肽段为机体 MHC 系统所识别，使机体产生保护性免疫应答，不仅突破了

将整个基因作为目的片段的局限,同时因为表位片段较短、表达产物分子量较小,可以预留足够的空间插入合适的免疫增强因子,不存在基因整合以及重组等问题,相比于减毒活疫苗、DNA疫苗更加安全,不会产生严重的副作用<sup>[24]</sup>。

本研究从免疫信息学的角度,筛选鉴定 ASFV 5种结构蛋白 p72、p17、p49、M1249L 和 H240R 的 T、B 细胞表位,作为构成 ASFV 二十面体结构的重要衣壳蛋白,参与病毒形态发生和基因转录,可诱导产生中和抗体,产生免疫应答<sup>[9, 25]</sup>。5种蛋白均为亲水性蛋白质,共获得 CTL 优势表位 27 个,LBL 优势表位 35 个,仅针对 p72 的 CBL 优势表位 2 个,其中 5 种蛋白质 CTL 优势表位分别为 12 (44.44%)、0、0、12 (44.44%) 和 3 (11.11%) 个,LBL 优势表位分别为 10 (28.57%)、1 (2.86%)、5 (14.29%)、17 (48.57%) 和 2 (5.71%) 个。预测表位多数集中于 p72 和 M1249L,合计占 CTL 优势表位的 88.89% (24/27),LBL 优势表位的 77.14% (27/35),以及 CBL 优势表位的 100% (2/2),说明它们可被用作抗原性靶标,在 5 种蛋白质中最具疫苗设计研制前景<sup>[26]</sup>。尽管 M1249L 二级结构以  $\alpha$ -螺旋为主,但是其包含氨基酸数量较多,有足够的无规则卷曲区域可以用来筛选抗原表位。

T 细胞表位预测工具基于多肽片段与 MHC 结合、细胞内加工和转运的过程,计算得到每个多肽作为 T 细胞表位的预测评分。人类白细胞抗原分为 MHC-I 类和 MHC-II 类,分别激活人体内 CTL 和辅助 T 淋巴细胞(helper T lymphocyte, HTL)。其中 CTL 能够分泌颗粒酶、穿孔素等免疫因子中和被感染受损的宿主细胞,在 MHC-I 介导的细胞免疫中发挥关键作用。SLA,即猪的 MHC 抗原,分布于白细胞表面,能够识别 T 细胞表面的自身肽和非自身肽<sup>[27]</sup>。B 细胞表位分为 LBL 表位和 CBL 表位两种类型,LBL 表位由参与抗体结合的连续氨基酸线性延伸组成,其相互作用基于表位的一级结构,而 CBL 表位由抗原蛋白序列上相距较远的氨基酸组成,通过折叠结构聚集在空间中<sup>[28]</sup>。尽管预测 T 细胞表位相比于 B 细胞表位更加先进可靠<sup>[29]</sup>,但是对于非人类蛋白的 ASFV, T 细胞表位预测可能更受限;对于 ASFV 的这 5 种蛋白质,筛选出的 B 细胞优势表位多于 T 细胞表位。通过预测 ASFV 的 5 种蛋白质的理化性质、二级结构等参数,得到多个潜在 T、B 细胞表位,能够诱导保护性免疫应答,预测准确性良好,尤其是蛋白 p72 和 M1249L,为构建 ASF 表位疫苗提供研究依据,为设计研发效果强、安全性高的 ASF 疫苗提供潜在思路。

## 参考文献:

- [1] LIU Q, MA B, QIAN N, et al. Structure of the African swine fever virus major capsid protein p72 [J]. *Cell Res*, 2019, 29 (11): 953-955.
- [2] GALINDO I, ALONSO C. African swine fever virus: a review [J]. *Viruses*, 2017, 9 (5): 103.
- [3] 王涛,孙元,罗玉子,等.非洲猪瘟防控及疫苗研发:挑战与对策 [J]. *生物工程学报*, 2018, 34 (12): 1931-1942.
- [4] LIU J, LIU B, SHAN B, et al. Prevalence of African swine fever in China, 2018-2019 [J]. *J Med Virol*, 2020, 92 (8): 1023-1034.
- [5] ZHOU X, LI N, LUO Y, et al. Emergence of African swine fever in China, 2018 [J]. *Transbound Emerg Dis*, 2018, 65 (6): 1482-1484.
- [6] JENSEN K K, ANDREATTA M, MARCATILI P, et al. Improved methods for predicting peptide binding affinity to MHC class II molecules [J]. *Immunology*, 2018, 154 (3): 394-406.
- [7] GRIFONI A, SIDNEY J, ZHANG Y, et al. A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2 [J]. *Cell Host Microbe*, 2020, 27 (4): 671-680.
- [8] XU W, WANG M, YU D, et al. Variations in SARS-CoV-2 spike protein cell epitopes and glycosylation profiles during global transmission course of COVID-19 [J]. *Front Immunol*, 2020, 11: 565278.
- [9] WANG N, ZHAO D, WANG J, et al. Architecture of African swine fever virus and implications for viral assembly [J]. *Science*, 2019, 366 (6465): 640-644.
- [10] WILKINS M R, GASTEIGER E, BAIROCH A, et al. Protein identification and analysis tools in the ExPASy server [J]. *Methods Mol Biol*, 1999, 112: 531-552.
- [11] REYNISSON B, ALVAREZ B, PAUL S, et al. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data [J]. *Nucleic Acids Res*, 2020, 48 (W1): W449-W454.
- [12] NIELSEN M, ANDREATTA M. NetMHCpan-3.0: improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets [J]. *Genome Med*, 2016, 8 (1): 33.
- [13] CALIS J J, MAYBENO M, GREENBAUM J A, et al. Properties of MHC class I presented peptides that enhance immunogenicity [J]. *PLoS Comput Biol*, 2013, 9 (10): e1003266.
- [14] SAHA S, RAGHAVA G P. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network [J]. *Proteins*, 2006, 65 (1): 40-48.
- [15] JESPERSEN M C, PETERS B, NIELSEN M, et al. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes [J]. *Nucleic Acids Res*, 2017, 45 (W1): W24-W29.
- [16] MANAVALAN B, GOVINDARAJ R G, SHIN T H, et al. iBCE-EL: a new ensemble learning framework for improved linear B-cell epitope prediction [J]. *Front Immunol*, 2018, 9: 1695.

- [17] The UniProt Consortium. UniProt: the universal protein knowledge-base in 2023 [J]. *Nucleic Acids Res*, 2023, 51 (D1): D523–D531.
- [18] KRINGELUM J V, LUNDEGAARD C, LUND O, et al. Reliable B cell epitope predictions: impacts of method development and improved benchmarking [J]. *PLoS Comput Biol*, 2012, 8 (12): e1002829.
- [19] PONOMARENKO J, BUI H H, LI W, et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes [J]. *BMC Bioinformatics*, 2008, 9: 514.
- [20] DOYTCHINOVA I A, FLOWER D R. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines [J]. *BMC Bioinformatics*, 2007, 8: 4.
- [21] GUPTA S, KAPOOR P, CHAUDHARY K, et al. In silico approach for predicting toxicity of peptides and proteins [J]. *PLoS One*, 2013, 8 (9): e73957.
- [22] DIMITROV I, BANGOV I, FLOWER D R, et al. AllerTOP v. 2: a server for in silico prediction of allergens [J]. *J Mol Model*, 2014, 20 (6): 2278.
- [23] SÁNCHEZ-CORDÓN P J, MONTOYA M, REIS A L, et al. African swine fever: a re-emerging viral disease threatening the global pig industry [J]. *Vet J*, 2018, 233: 41–48.
- [24] CORRAL-LUGO A, LÓPEZ-SILES M, LÓPEZ D, et al. Identification and analysis of unstructured, linear B-cell epitopes in SARS-CoV-2 virion proteins for vaccine development [J]. *Vaccines (Basel)*, 2020, 8 (3): 397.
- [25] ALEJO A, MATAMOROS T, GUERRA M, et al. A proteomic atlas of the African swine fever virus particle [J]. *J Virol*, 2018, 92 (23): e01293.
- [26] HEIMERMAN M E, MURGIA M V, WU P, et al. Linear epitopes in African swine fever virus p72 recognized by monoclonal antibodies prepared against baculovirus-expressed antigen [J]. *J Vet Diagn Invest*, 2018, 30 (3): 406–412.
- [27] CHOI N R, SEO D W, CHOI K M, et al. Analysis of swine leukocyte antigen haplotypes in Yucatan miniature pigs used as biomedical model animal [J]. *Asian-Australas J Anim Sci*, 2016, 29 (3): 321–326.
- [28] KRINGELUM J V, NIELSEN M, PADKJÆR S B, et al. Structural analysis of B-cell epitopes in antibody: protein complexes [J]. *Mol Immunol*, 2013, 53 (1/2): 24–34.
- [29] ZHONG W, RECHE P A, LAI C C, et al. Genome-wide characterization of a viral cytotoxic T lymphocyte epitope repertoire [J]. *J Biol Chem*, 2003, 278 (46): 45135–45144.