

· 论 著 ·

## 冠状病毒突刺糖蛋白的免疫表位预测分析\*

焦娟<sup>1</sup>, 姜海明<sup>2</sup>, 李珊珊<sup>1</sup>, 张立新<sup>1</sup>, 刘杰<sup>1△</sup>

1. 中国人民解放军总医院第七医学中心检验科, 北京 100700; 2. 滨州医学院烟台附属医院重症医学科, 山东烟台 264100

**摘要:**目的 通过对冠状病毒突刺糖蛋白的抗体免疫表位进行深度分析, 为冠状病毒疫苗的研发提供可靠的科学依据。方法 基于目的蛋白一级序列、二级结构、高级结构和 B 细胞免疫表位预测, 分析新型冠状病毒突刺糖蛋白的抗体免疫表位信息。应用 PHD 和 Jpred4 等软件, 预测目的蛋白的  $\alpha$  螺旋、 $\beta$  折叠、转角以及卷曲结构; 应用在线软件 Bepipred Linear Epitope Prediction 2.0 预测目的蛋白的 B 细胞表位。结果 结合蛋白质数据库中突刺糖蛋白的高分辨三维结构, 进一步筛选优势抗原表位区段, 得到突刺糖蛋白的优势 B 细胞免疫表位。结论 筛选出突刺糖蛋白的优势 B 细胞免疫表位, 为进一步通过动物实验筛选出高效表位及冠状病毒疫苗的研制和检测奠定基础。

**关键词:** 冠状病毒; 突刺糖蛋白; 二级结构; B 细胞表位

DOI: 10.3969/j.issn.1673-4130.2022.06.011

中图法分类号: R446.5

文章编号: 1673-4130(2022)06-0688-04

文献标志码: A

## Immune epitopes analysis of Spike protein of Coronavirus\*

JIAO Juan<sup>1</sup>, JIANG Haiming<sup>2</sup>, LI Shanshan<sup>1</sup>, ZHANG Lixin<sup>1</sup>, LIU Jie<sup>1△</sup>

1. Department of Clinical Laboratory, the Seventh Medical Centre, Chinese PLA General Hospital, Beijing 100700, China; 2. Intensive Care Unit, Yantai Hospital Affiliated of Binzhou Medical University, Yantai, Shandong 264100, China

**Abstract: Objective** The purpose of this article is to analyze the immune epitopes of the Spike protein of Coronavirus, and to provide reliable scientific basis for the development of Coronavirus vaccine. **Methods** The Spike protein was analyzed based on their primary sequence, secondary structure, 3D structure and B-cell immune epitopes prediction. PHD and Jpred4 software were used to predict the alpha helix, random coil, beta turn and extended strand of the target protein. Bepipred Linear Epitope Prediction 2.0 was used to predict B-cell Epitopes of the target protein. **Results** Combining with high-resolution three-dimensional protein structures of Spike protein in the protein data bank (RCSB), the predominant antigen epitopes of Spike proteins were further screened, and then obtain the dominant B-cell immune epitopes. **Conclusion** The dominant B cell immune epitopes of Spike protein has been screened, and this work has paved the road for further screening of efficient epitopes through animal experiments.

**Key words:** coronavirus; Spike protein; secondary structure; B cell epitope

冠状病毒是一类 RNA 病毒, 通过人的呼吸道或接触传播, 感染该病毒后常见临床表现有发热, 以及咳嗽、气促和呼吸困难等呼吸道症状, 严重者可导致肺炎、严重急性呼吸综合征、肾衰竭, 甚至死亡。

近期, 新型冠状病毒(SARS-CoV-2)疫苗相继问世(灭活疫苗、腺病毒载体疫苗和 mRNA 疫苗等), 其作用机制即通过激活人自身的免疫系统来对抗病毒, 其主要的靶点为分布于病毒表面的突刺糖蛋白(Spike 蛋白)。部分学者在 SARS-CoV-2 致病机制方

面的基础研究也做出了巨大的努力, 从最初的 SARS-CoV-2 基因组的测定完成<sup>[1]</sup>, 到 SARS-CoV-2 全病毒三维结构的解析<sup>[2]</sup>, 都显示 Spike 蛋白是 SARS-CoV-2 的主要致病因素: Spike 蛋白与人体细胞表面血管紧张素转换酶 2(ACE2)的结合介导了 SARS-CoV-2 对人体细胞的入侵<sup>[3-4]</sup>。因此 SARS-CoV-2 的 Spike 蛋白已经成为人们诊断、治疗、预防 SARS-CoV-2 感染的主要靶点。本研究拟通过生物信息学软件, 通过横向比较和分析 SARS-CoV-2 的 Spike 蛋白的二级结

\* 基金项目: 陆军部后勤科研项目(CLJ7J021)。

作者简介: 焦娟, 女, 主管技师, 主要从事生物化学与分子生物学研究。△ 通信作者, E-mail: fhaalj@163.com。

本文引用格式: 焦娟, 姜海明, 李珊珊, 等. 冠状病毒突刺糖蛋白的免疫表位预测分析[J]. 国际检验医学杂志, 2022, 43(6): 688-691.

构、三级结构和 B 细胞表位,探讨 Spike 蛋白潜在的高效免疫表位,为 SARS-CoV-2 检测试剂盒、疫苗制备、治疗药物的研发提供可靠的生物学信息和实验基础。

## 1 材料与方法

**1.1 蛋白结构和序列的选择** 根据已公布的蛋白结构数据库(<https://www.rcsb.org/>),比对 Spike 蛋白相关的结构信息,下载对应的结构信息文件(pdb 文件),其登录号为 6XR8/6XRA(SARS-CoV-2),5X58(SARS-CoV),5X59(MERS);同时,下载相对应的氨基酸序列。结合下载的序列,比对 GenBank 数据库,分别下载各蛋白的完整蛋白序列,其登录号分别为 YP\_009724390, P59594.1, AGN70929.1。

**1.2 二级结构预测** 抗原免疫表位主要位于蛋白质表面,具有较强的亲水性,含有或邻近柔性区域,易于形成典型的二级结构等。因此分析目的蛋白质的二级结构,对了解其免疫表位分布具有重要的参考价值。本研究应用 PHD([https://npsa-prabi.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_phd.html](https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_phd.html))和 Jpred4(<http://www.compbio.dundee.ac.uk/jpred/>)等在线软件分别预测目的蛋白质的  $\alpha$  螺旋、 $\beta$  折叠、转角以及卷曲等二级结构的组成。

**1.3 B 细胞表位预测** 根据目的蛋白的蛋白序列应用在线软件 Bepipred Linear Epitope Prediction 2.0(<http://tools.iedb.org/bcell/>)分别预测目的蛋白的 B 细胞表位,该软件运用随机森林算法对目的蛋白的表位和非表位氨基酸进行预测,同时利用自身表位数据库对预测序列进行多重优化<sup>[5]</sup>。本研究同时比对目的蛋白的二级结构和三维结构的数据,进一步筛选优势抗原表位区段,得出 Spike 蛋白的优势 B 细胞免疫表位。

## 2 结果

**2.1 氨基酸序列分析** Spike 蛋白的结构信息非常丰富:目前在蛋白质结构数据库中,Spike 蛋白的高分辨结构有上百个,均为冷冻电子显微解析的结构。这些结构信息显示,Spike 蛋白有多种构象形式:侵染细胞前状态、侵染细胞后状态和与中和抗体结合状态等,这些状态为分析 Spike 蛋白的免疫表位提供了丰富的分析线索。本文以 Spike 蛋白三维结构为基础模型,以多种预测软件为辅助,分析和研究严重急性呼吸综合征冠状病毒(SARS-CoV)、SARS-CoV-2 和中东呼吸综合征冠状病毒(MERS-CoV)共 3 种冠状病毒 Spike 蛋白的 B 细胞表位信息。首先在 GeneBank 数据库中下载相关蛋白的一级序列信息:SARS-CoV 的 Spike 蛋白序列(YP\_009724390)与 SARS-CoV-2 的 Spike 蛋白(P59594.1)具有 76% 同一性,与 MERS-CoV 的 Spike 蛋白具有 35% 的同一性(AGN70929.1)。

**2.2 Spike 蛋白二级结构分析** 运用 PHD 和 Jpred4

两种在线软件,对 SARS-CoV、SARS-CoV-2、MERS-CoV 的 Spike 蛋白二级结构进行预测。SARS-CoV 的 Spike 蛋白  $\alpha$  螺旋主要分布于 5~11、51~56、409~419、729~769、797~805、849~861、871~1 018、1 129~1 140、1 151~1 184、1 195~1 213 等肽段。 $\beta$  折叠分布于 146~166、173~178、181~190、198~230、241~258、277~279、309~377、421~428、442~450、465~484、552~582、584~601、610~646、693~724、1 032~1 058、1 075~1 087、1 100~1 115、1 213~1 219、1 247~1 255 等肽段。自由卷曲位于  $\alpha$  螺旋和  $\beta$  折叠之间。见表 1。

表 1 3 种病毒 Spike 蛋白二级结构组成(%)

病毒	$\alpha$ 螺旋	$\beta$ 折叠	自由卷曲
SARS-CoV	23.7	28.6	47.7
SARS-CoV-2	21.7	28.6	49.7
MERS-CoV	23.7	26.3	50.0

SARS-CoV-2 的 Spike 蛋白  $\alpha$  螺旋主要分布于 361~371、747~788、815~823、867~879、887~1 035、1 147~1 159、1 175~1 123、1 134~1 151 等肽段。 $\beta$  折叠分布于 2~8、36~50、61~77、90~108、116~145、167~177、190~206、262~288、298~329、376~381、395~403、418~425、432~437、450~455、507~517、538~577、583~588、595~600、608~613、589~695、711~729、1 048~1 075、1 093~1 105、1 120~1 132、1 151~1 156、1 190~1 195 等肽段。自由卷曲位于  $\alpha$  螺旋和  $\beta$  折叠之间。

MERS-CoV 的 Spike 蛋白  $\alpha$  螺旋主要分布于 107~112、225~231、277~281、304~318、405~418、430~436、806~837、846~856、887~895、908~912、941~954、963~1 110、1 230~1 241、1 262~1 285、1 297~1 316 等肽段。 $\beta$  折叠分布于 2~10、30~35、64~70、84~89、98~104、116~123、129~131、162~172、179~188、202~206、236~242、262~271、287~296、321~337、360~377、399~405、440~449、464~471、478~483、567~576、603~664、670~676、713~718、754~765、779~804、1 121~1 127、1 134~1 151、1 202~1 213、1 317~1 320 等肽段。自由卷曲位于  $\alpha$  螺旋和  $\beta$  折叠之间。

**2.3 B 细胞表位预测** 运用 Bepipred Linear Epitope Prediction 2.0 在线软件,对上述病毒 Spike 蛋白 B 细胞表位完成预测,阈值设定为 0.5。SARS-CoV 的 Spike 蛋白 B 细胞表位区段分布于 16~38、60~84、134~148、173~181、203~212、236~249、317~351、355~412、427~473、505~519、602~630、642~651、663~689、768~796、811~827、1 089~1 100、1 115~1 156、1 234~1 248 等肽段。SARS-CoV-2 的 Spike 蛋白 B 细胞表位区段分布于 13~37、59~81、138~

156、177~189、206~221、250~260、306~322、370~393、409~425、440~450、458~501、516~536、617~632、634~644、656~666、672~690、695~710、786~800、806~816、828~842、1 108~1 118、1 133~1 172、1 198~1 206、1 252~1 263、1 275~1 306 等肽段。MERS-CoV 的 Spike 蛋白 B 细胞表位区段分布于 16~38、60~84、134~148、173~181、203~212、236~249、317~351、355~412、427~473、505~519、602~630、642~651、663~689、768~796、811~827、1 089~1 100、1 115~1 156、1 234~1 248 等肽段。

### 2.4 结合蛋白三维结构的 B 细胞表位区段的筛选

以上的表位信息结果均源于软件的预测,这些结果很可能与实际情况差异较大,因此本研究将以上 B 细胞表位区段预测信息,在高分辨的蛋白三维结构和二级结构基础上,同时结合 B 细胞抗原表位结构和生化特点,进一步人工筛选 SARS-CoV (5X59)<sup>[6]</sup>、SARS-CoV-2(6XR8)<sup>[3]</sup>、MERS-CoV(5X58)<sup>[6]</sup> Spike 蛋白的 B 细胞表位区段,见表 2~4。

表 2 SARS-CoV(5X59)Spike 蛋白的 B 细胞表位区段

起始区段	终止区段	二级结构	长度(bp)	是否位于蛋白表面
16	38	卷曲-折叠	23	是
60	84	卷曲-折叠	25	是
134	148	卷曲-折叠	15	是
203	212	卷曲	10	是
236	249	卷曲-折叠	14	是
317	351	卷曲-折叠-螺旋	35	是
355	412	卷曲-折叠-螺旋	58	是
427	473	卷曲	47	是
505	519	卷曲	15	是
602	630	卷曲-折叠	29	是
642	651	卷曲-折叠	10	是
663	689	卷曲-折叠	27	是
768	796	卷曲-折叠	29	是
811	827	密度丢失	17	是
1 089	1 100	卷曲	12	是
1 115	1 156	密度丢失	42	是

表 3 SARS-CoV-2(6XR8)Spike 蛋白的 B 细胞表位区段

起始区段	终止区段	二级结构	长度(bp)	是否位于蛋白表面
13	37	卷曲-折叠	25	是
59	81	卷曲-折叠	23	是
138	156	卷曲-折叠	19	是
177	189	卷曲-折叠	13	是
206	221	卷曲-折叠	16	是
250	260	卷曲	11	是
306	322	卷曲-折叠	17	是
328	363	卷曲-折叠-螺旋	36	是
370	393	卷曲-折叠-螺旋	24	是

续表 3 SARS-CoV-2(6XR8)Spike 蛋白的 B 细胞表位区段

起始区段	终止区段	二级结构	长度(bp)	是否位于蛋白表面
409	425	卷曲-螺旋	17	是
440	450	卷曲-螺旋	11	是
458	501	卷曲-折叠	44	是
516	536	卷曲-折叠	21	是
617	632	密度丢失	16	是
634	644	卷曲-折叠	11	是
656	666	卷曲-折叠	11	是
672	690	卷曲-折叠	19	是
695	710	卷曲-折叠	16	是
786	800	卷曲-折叠	15	是
806	815	卷曲	10	是
828	842	卷曲-螺旋	15	是
1 108	1 118	卷曲-折叠	11	否
1 133	1 172	卷曲-螺旋	40	是
1 252	1 263	密度丢失	12	是
1 275	1 306	密度丢失	32	是

表 4 MERS-CoV(5X58)Spike 蛋白的 B 细胞表位区段

起始区段	终止区段	二级结构	长度(bp)	是否位于蛋白表面
19	28	卷曲	10	是
31	51	卷曲-折叠-螺旋	21	是
74	83	卷曲	10	是
92	111	卷曲-折叠-螺旋	20	是
125	142	卷曲-折叠	18	是
189	199	卷曲	11	是
212	235	卷曲-折叠-螺旋	24	是
352	369	卷曲-折叠	18	是
379	402	卷曲-折叠-螺旋	24	是
404	436	卷曲-折叠-螺旋	33	是
458	478	卷曲-折叠-螺旋	21	是
482	499	卷曲-折叠	18	是
501	523	卷曲-折叠	23	是
531	550	卷曲-折叠	20	是
580	603	卷曲-折叠	24	是
685	709	部分密度缺失	25	是
720	732	卷曲	13	是
738	777	卷曲	40	是
855	867	卷曲	13	是
876	887	密度丢失	12	是
900	917	卷曲-螺旋	18	是
1 108	1 117	卷曲-螺旋	10	是
1 191	1 203	卷曲	13	是
1 214	1 234	密度丢失	21	是
1 236	1 257	密度丢失	22	是
1 334	1 349	密度丢失	16	是

由于蛋白三维结构信息来自于蛋白晶体衍射数据,因此存在结构密度信息的部分缺失,另一方面,结构密度信息的缺失一般意味着蛋白这部分区域处于无序排列的状态,即在周边的环境中可以自由摆动或存在结构变化,说明该区域亲水性较强并位于蛋白三维结构的表面。

将 B 细胞表位信息, SARS-CoV、SARS-CoV-2、MERS-CoV 3 种病毒 Spike 蛋白的二级结构信息和三级结构信息整合在一起筛选,可提高数据分析的可信性。蛋白抗原表位区段基本分布于蛋白三维结构的表面,与二级结构有所差别。SARS-CoV Spike 蛋白抗原表位区段中卷曲-螺旋占 17.0%,卷曲-延伸占 50.0%,卷曲占 33.0%。SARS-CoV-2 Spike 蛋白抗原表位区段中卷曲-螺旋占 22.0%,卷曲-延伸占 59.0%,卷曲占 19.0%。MERS-CoV Spike 蛋白抗原表位区段中卷曲-螺旋占 25%,卷曲-延伸占 37.5%,卷曲占 37.5%。

综合分析结果发现,具有优势 B 细胞表位区段中存在一定二级结构和一定柔韧性结构组合的占比较高,比如  $\beta$  折叠。这一特点完全符合抗原蛋白 B 细胞表位的基本特征。

### 3 讨论

Spike 蛋白是 SARS-CoV、SARS-CoV-2、MERS-CoV 等冠状病毒关键性致病因子。Spike 蛋白的结构已经较为清晰,其致病的分子机制已较为明了<sup>[3-4]</sup>;首先,Spike 蛋白的受体结合域(RBD)向外伸展与人体细胞表面的 ACE2 受体结合;其次,RBD 与 ACE2 的结合启动了 Spike 蛋白自身在 S 位点的剪切作用,使 Spike 蛋白发生“脱冠”变化;最后,“脱冠”的 Spike 蛋白 HR 结构域得以暴露出来,在 HR 结构域辅助下,病毒完成和人体细胞的融合和遗传物质的注入过程。因此,Spike 蛋白 RBD 和 ACE2 识别和结合,是冠状病毒感染细胞的最关键环节,从目前已有的 Spike 蛋白-中和抗体的三维结构中,也可以证明这一结论。

在蛋白结构数据库中,SARS-CoV、SARS-CoV-2、MERS-CoV 等病毒的 Spike 蛋白和其中和抗体的结构较多,如 6NB6/7 (SARS-CoV)、6WPT (SARS-CoV-2)、7K4N (SARS-CoV-2)、6NB3/4 (MERS-CoV)等<sup>[7-9]</sup>。对这些复合体的结构分析后发现,中和抗体与 Spike 蛋白表面结合位点主要集中于 RBD 结构域和 CTD 结构域。同时本研究发现 ACE2 与 Spike 蛋白结合位点也位于 RBD 结构,且与中和抗体的结合位点重叠,这进一步证明了中和抗体在抗病毒时的作用机制。以 SARS-CoV-2 Spike 蛋白为例,在 7KNB 结构中,ACE2 结合于 449~502aa<sup>[10]</sup>;7K8W 结构中,中和抗体结合于 444~501aa<sup>[11]</sup>;本研究结果与上述研究结果相似,表明本研究分析结果具有一定的可靠性和前瞻性。

本研究同时发现 Spike 蛋白 RBD 和 CTD 结构

中,含有多个优势 B 细胞表面抗原表位的多肽区段,这为检测试剂、抗体药物的开发等提供了较好的作用靶点。然而,冠状病毒的 RBD 结构域、三维结构非常复杂,主要由内部的  $\beta$  折叠和外部的无规则卷曲所构成,说明 RBD 结构域外围表面结构十分灵活,RBD 在与 ACE2 或中和抗体结合时,可能存在局部的构象重建;同时在病毒侵染过程中 RBD 本身还会存在巨大的结构变化,使基础和临床研究难度大大增加,因此更多的结构和 B 细胞表面抗原表位信息将提供更多的研究思路和实验基础。

### 参考文献

- [1] WU F, ZHAO S, YU B, et al. A new coronavirus associated with human respiratory disease in China[J]. *Nature*, 2020, 579(7798): 265-269.
- [2] YAO H, SONG Y, CHEN Y, et al. Molecular Architecture of the SARS-CoV-2 Virus[J]. *Cell*, 2020, 183(3): 730-738.
- [3] CAI Y, ZHANG J, XIAO T, et al. Distinct conformational states of SARS-CoV-2 spike protein[J]. *Science*, 2020, 369(6511): 1586-1592.
- [4] WALLS A C, PARK Y J, TORTORICI M A, et al. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein[J]. *Cell*, 2020, 183(2): 281-292.
- [5] 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.
- [6] YUAN Y, CAO D, ZHANG Y, et al. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains[J]. *Nat Commun*, 2017, 8: 15092.
- [7] PINTO D, PARK Y J, BELTRAMELLO M, et al. Structural and functional analysis of a potent sarbecovirus neutralizing antibody[J]. *Nature*, 2020, 583(7815): 290-295.
- [8] WALLS A C, XIONG X, PARK Y J, et al. Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion[J]. *Cell*, 2019, 176(5): 1026-39. e15.
- [9] TORTORICI M A, BELTRAMELLO M, LEMPP F A, et al. Ultrapotent human antibodies protect against SARS-CoV-2 challenge via multiple mechanisms[J]. *Science*, 2020, 370(6519): 950-957.
- [10] ZHOU T, TSYBOVSKY Y, GORMAN J, et al. Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a pH-Dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains[J]. *Cell Host Microbe*, 2020, 28(6): 867-879. e5.
- [11] BARNES C O, JETTE C A, ABERNATHY M E, et al. SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies[J]. *Nature*, 2020, 588(7839): 682-687.