

# Spike 蛋白的生物信息学研究分析

郑柳凤, 伊嘉豪, 梁定发, 叶远浓

(贵州医科大学大健康学院生物信息学与生物医学大数据挖掘实验室, 贵阳 550025)

**摘要:** 本研究通过挖掘国内外文献, 系统地分析了 Spike (S) 蛋白的作用机制和生物信息学方面的研究进展。总结目前国内外科学家利用生物信息学领域工具开发 Spike 蛋白的药物、疫苗等方面的研究。此外, 还对现有的 Spike 蛋白数据资源进行了系统整理。总结分析了现有国内外与 Spike 蛋白相关的生物信息学工具及数据库, 最后对病毒 Spike 蛋白在该领域的研究提出了展望。本综述旨在为研究 Spike 蛋白的生物学家、病毒学家等研究人员提供系统的 Spike 蛋白研究参考, 以促进 Spike 蛋白机制研究、有效药物和疫苗研发。

**关键词:** 生物物理学; Spike 蛋白; 综述; 冠状病毒; SARS-CoV-2; 生物信息学; 数据库

中图分类号: Q67

文献标识码: A

文章编号: 1674-2850(2022)03-0288-07

开放科学(资源服务)标识码(OSID):



## Analysis of bioinformatics research on Spike protein

ZHENG Liufeng, YI Jiahao, LIANG Dingfa, YE Yuannong

(Bioinformatics and Biomedical Big data Mining Laboratory, School of Big Health, Guizhou Medical University, Guiyang 550025, China)

**Abstract:** This study systematically analyses the mechanism of Spike(S) protein and bioinformatics research progress by mining domestic and foreign literatures. This paper summarizes the current research on use of bioinformatics tools by scientists at home and abroad to develop drugs and vaccines research of Spike protein. In addition, the existing Spike protein data resources are systematically organized. Related bioinformatics tools and databases are summarized, and the research prospects of Spike protein in this field are proposed. This review aims to provide systematic Spike protein research references for biologists, virologists and other researchers who study Spike proteins. It promotes the research of Spike protein mechanism and the development of effective drugs and vaccines.

**Key words:** biophysics; Spike protein; review; coronavirus; SARS-CoV-2; bioinformatics; database

## 0 引言

Spike (S) 蛋白是 I 类病毒融合蛋白, 在病毒入侵细胞过程中, 介导病毒与宿主细胞受体结合, 介导病毒包膜与宿主细胞膜之间的膜融合, 从而影响病毒在细胞中的感染倾向<sup>[1-3]</sup>。Spike 蛋白是病毒进入宿主细胞必需且重要的媒介, 从而感染宿主。Spike 蛋白在翻译后被裂解为 S1 受体结合亚基和 S2 膜融合亚基, 病毒首先通过 S1 亚基与宿主细胞表面的受体结合, 然后其 S2 亚基起到融合病毒和宿主细胞膜的作用<sup>[1,4]</sup>。在病毒入侵过程中 Spike 蛋白不断地发生不可逆的由融合前构象向融合后构象转变, 以释放能量, 激活融合物被触发的多样性, 使融合机制更为复杂。因此 Spike 蛋白决定了病毒的宿主选择范围和宿主特异性, 同时也是相关病毒疫苗设计的关键靶点<sup>[5-6]</sup>。因此, 国内外众多研究者对 Spike 蛋白进行了各层次的研究。

**作者简介:** 郑柳凤 (1996—), 女, 硕士研究生, 主要研究方向: 生物信息学

**通信联系人:** 叶远浓, 副教授, 主要研究方向: 智能生物信息学. E-mail: yyn@gmc.edu.cn

## 1 Spike 蛋白的研究现状

### 1.1 Spike 蛋白的发现

自从 COMPANS 等<sup>[7]</sup>在流感病毒中鉴定出 Spike 糖蛋白后,越来越多的研究人员展开对病毒 Spike 蛋白的功能研究。WAGNER 等<sup>[8]</sup>和 McSHARRY 等<sup>[9]</sup>在水泡性口炎病毒 (vesicular stomatitis virus, VSV) 的研究中证实了 Spike 蛋白作为糖蛋白主要与富含质膜的光滑膜有关,与宿主细胞膜存在特异性的相互作用。GAROFF 等<sup>[10]</sup>和 KIELIAN 等<sup>[11]</sup>分别发现塞姆利基森林病毒 (Semliki Forest virus, SFV) 的 Spike 糖蛋白通过病毒膜与核衣壳密切接触,从而介导了病毒包膜和含胆固醇的靶膜之间的膜融合。此外, HELENIUS 等<sup>[12]</sup>研究发现 SFV 的 Spike 蛋白八聚体能够分别与人的主要组织相容性抗原 (human lymphocyte antigen-A, HLA-A; human lymphocyte antigen-B, HLA-B) 和小鼠的主要组织相容性抗原 (histocompatibility 2-K, H2-K; histocompatibility 2-D, H2-D) 结合。

### 1.2 冠状病毒的 Spike 蛋白

2002 年由冠状病毒 (SARS-CoV) 引起的重症急性呼吸综合征 (severe acute respiratory syndrome, SARS) 爆发后,冠状病毒频繁感染鸟类和哺乳动物,引发一系列病毒感染性疾病大流行。WANG 等<sup>[5]</sup>发现,在冠状病毒中, Spike 蛋白是一种重要的病毒抗原,介导宿主细胞受体结合并帮助病毒入侵宿主,提出了基于 Spike 蛋白的疫苗研发和治疗的方法。2019 年新型冠状病毒 (SARS-CoV-2) 引发新型冠状病毒肺炎 (corona virus disease 2019, COVID-19) 疫情的全球大爆发,生物学家们从该病毒的结构、功能和进化关系等一系列研究发现,主要作用靶标是 Spike 蛋白<sup>[4-6]</sup>, SARS-CoV-2 的 Spike 蛋白与人类 ACE2 复合物受体结合进入人体细胞引发感染<sup>[13-17]</sup>。冠状病毒能够通过相对容易的突变和重组来适应新的环境,能有效地改变宿主范围和组织特异性<sup>[4,18]</sup>。为有效地预防和控制由冠状病毒引起的感染性疾病,研究人员针对病毒进行疫苗研发。由于 COVID-19 疫情的爆发和感染范围之广,迫切需要生物信息学方法加入疫苗研发,通过生物信息学技术快速地、更全面地进行分子机制的识别与分析,以确定抗病毒治疗潜在的蛋白质靶点,并预测现有抗病毒药物对病毒的治疗效果<sup>[5,19]</sup>。

## 2 Spike 蛋白的生物信息学研究

### 2.1 生物信息学在 Spike 蛋白的药物、疫苗方面的研究

高俊等<sup>[20]</sup>分析了 105 株 SARS-CoV 的 Spike 蛋白受体结合域 (receptor binding domain, RBD) 的氨基酸序列的变异特征,同时比对了 4 株果子狸源性的 pcSARS-CoV 和 3 株人的 huSARS-CoV 氨基酸序列,揭示 Spike 蛋白在从动物向人传播过程中经历了很强的位点选择压力,大多数氨基酸的替换对感染力影响不大,但是对于 Spike 蛋白和 ACE2 结合的亲合力有影响。YURINA 等<sup>[21]</sup>在 Uniprot 数据库中检索了 9 个冠状病毒家族的 Spike 蛋白氨基酸序列,采用 Clustal 进行序列比对分析,发现 Spike 蛋白的胞外结构域存在一个高度保守的区域,对该区域进行进一步分析,发现预测的 Spike 蛋白的 B 细胞和 T 细胞表位有成为冠状病毒疫苗的候选靶点的潜力。HE 等<sup>[22]</sup>将 IEDB 数据库中 SARS-CoV-2 的 66 个 Spike 蛋白五肽表位的序列进行比对,通过采用 1 218 株 Spike 蛋白的单核苷酸多态性 (single nucleotide polymorphism, SNP) 分析获得 66 个 SARS-CoV-2 的 Spike 蛋白的潜在免疫相关肽。鉴定出 52 个易突变位点,通过突变和免疫信息学分析,筛选出 SARS-CoV-2 的 Spike 蛋白最佳疫苗候选基因表位。HSIEH 团队通过设计 100 种 Spike 蛋白结构,发现高表达的融合前稳定 Spike 蛋白结构将会加速新冠肺炎的疫苗研发和血清诊断<sup>[23-24]</sup>。Spike 蛋白作为 SARS-CoV-2 疫苗研发的主要靶点, SARS-CoV-2 的三聚体 Spike 糖蛋白结构在

2020年3月初首次报道<sup>[15,25]</sup>。2020年, McCARTHY等<sup>[9]</sup>比对了从世界多地收集的约15万个S基因序列,发现SARS-CoV-2 Spike蛋白的反复缺失可能导致抗体逃逸。生物信息学方法能高表达的加快Spike蛋白研究,以促进Spike蛋白在生物学和病毒学领域的研究<sup>[8~10]</sup>。

## 2.2 现有的Spike数据资源

截至2020年9月4日和截至2022年1月12日,NCBI<sup>[26]</sup>蛋白数据库中收录Spike蛋白序列及相关注释信息分别为59 685, 1 130 525。其中物种来源最多和增长最快的是病毒,截至2020年9月4日和2022年1月12日物种来源病毒分别为40 528, 1 083 278(如表1所示),占比分别为67.90%、95.82%,历时16个月以26.73倍速度增长。由于Spike蛋白本身作为糖蛋白在病毒入侵细胞中发挥的重要作用,尤其在冠状病毒中主要靶标是Spike蛋白,同时,本次覆盖全球的新型冠状病毒肺炎疫情的大流行,造成了这种数据产出。进一步对以上数据分析发现,Spike蛋白数据来源复杂,各种数据库存储形式多样,即使在NCBI中也是以文本形式存储,不利于数据的直接使用。

表1 NCBI收录的Spike蛋白数据物种来源和数据库来源列表

Tab. 1 Species sources and database sources of Spike protein data collected from NCBI

截至2020年9月4日(59 685)				截至2022年1月12日(1 130 525)			
物种来源		数据库来源		物种来源		数据库来源	
动物	2 210	PDB	2 346	动物	4 484	PDB	5 563
植物	1 011	RefSeq	4 976	植物	1 189	RefSeq	23 769
真菌	143	ProtKB/Swiss-Prot	678	真菌	281	ProtKB/Swiss-Prot	735
原生生物	86	DDBJ	2 685	原生生物	118	DDBJ	4 107
细菌	14 772	EMBL	5 285	细菌	39 854	EMBL	7 925
古生菌	28	GenBank	43 685	古生菌	111	GenBank	1 087 538
病毒	40 528	PIR	18	病毒	1 083 278	PIR	18
其他	907	其他	12	其他	1 210	其他	870

## 3 Spike蛋白的生物信息学工具和数据库

### 3.1 国内外Spike蛋白的生物信息学工具和数据库

目前,包括NCBI、GISAID<sup>[27]</sup>、UniProt<sup>[28]</sup>在内的大型生物信息数据库存储了Spike蛋白一级信息,但由于各数据库存储数据格式和内容不同,以及存储的数据量也不一样,造成Spike蛋白信息的重复冗余和数据不能直接融合统一等问题。为此,研究者们通过对Spike蛋白的序列二次分析和挖掘,构建了针对某一领域或某种功能的二级数据库或分析工具(如见表2所示)。2020年1月22日中国国家生物信息中心(China National Center for Bioinformatics, CNCB)发布的2019年新型冠状病毒资源(2019 novel coronavirus resource, 2019nCoV-2)<sup>[29]</sup>,这是基于SARS-CoV-2的病毒信息而开发的数据库。2019nCoV-2结合不同数据库和平台全面整合了所有公开可用的SARS-CoV-2分离株的序列和临床信息。2019nCoV-2收集有关基因组序列、变异、单倍型和文献的最新信息,仅在数据下载模块中存储部分SARS-CoV-2的相关Spike蛋白序列信息。Ensembl<sup>[30]</sup>数据库是为了从人类基因组计划中获取数据而开发的,现已发展成为一个用于生成、整合和传播参考基因组的大型数据库。针对SARS-CoV-2,该数据库团队于2020年5月推出了Ensembl COVID-19子库。Ensembl COVID-19使用对准参照基因组覆盖序列比对以作进化树的方法,提供基因注释和最高置信度的变异数据。数据库仅提供部分SARS-CoV-2的相关Spike蛋白基因组

注释和变异数据。2019nCoV、Ensembl COVID-19 都基于冠状病毒中的 SARS-CoV-2 设计, 这些数据库包含病毒 Spike 蛋白的基因组序列信息。

表 2 最近发表的 Spike 蛋白相关生物信息学工具和数据库

Tab. 2 Recently published Spike protein related bioinformatics tools and databases

工具或数据库名	URL	描述
2019nCoV	<a href="https://bigd.big.ac.cn/ncov/">https://bigd.big.ac.cn/ncov/</a>	2019 新型冠状病毒资源, 集成注释和分析数据, 非冗余全球 SARS-CoV-2 基因组的完整集合
Ensembl COVID-19	<a href="https://covid-19.ensembl.org">https://covid-19.ensembl.org</a>	SARS-CoV-2 基因组, 基因集、多个变体集及注释是数据库
CoV3D	<a href="https://cov3d.ibbr.umd.edu">https://cov3d.ibbr.umd.edu</a>	高分辨率的冠状病毒蛋白结构的数据库
SCoV2-MD	<a href="https://submission.gpermd.org/covid19/home/">https://submission.gpermd.org/covid19/home/</a>	SARS-CoV-2 蛋白质组动态和变异影响预测的数据库
RCSB PDB	<a href="http://RCSB.org">http://RCSB.org</a>	生物信息学蛋白质数据库, 3D 大分子结构数据库
CATH	<a href="https://www.cathdb.info">https://www.cathdb.info</a>	识别蛋白质结构中的结构域, 分类进化家族, 提供结构和功能注释
ESC	<a href="http://clingen.igib.res.in/esc/">http://clingen.igib.res.in/esc/</a>	SARS-CoV-2 免疫逃逸变异数据库
VirusViz	<a href="http://gmql.eu/virusviz/">http://gmql.eu/virusviz/</a>	病毒核苷酸和氨基酸变异的比较分析和有效可视化

CoV3D<sup>[31]</sup>是一个冠状病毒蛋白质结构的数据库, 存储部分冠状病毒 (SARS-CoV、MERS-CoV、SARS-CoV-2) 蛋白结构信息, 为用户提供可视化 Spike 糖蛋白结构的功能。SCoV2-MD<sup>[32]</sup>使用分子动力学 (molecular dynamics, MD) 方法进行的模拟 SARS-CoV-2 蛋白质组的原子, 以研究病毒蛋白的结构-动力学-功能关系, 可以组合所有指标构建预定义或自定义分数, 以评估不断变化的变体对蛋白质结构和功能的影响, 以提供 SARS-CoV-2 3D 蛋白质组的高分辨率结构信息。结构生物信息学蛋白质数据库 (research collaboratory for structural bioinformatics protein data bank, RCSB PDB)<sup>[33]</sup>是应用全球蛋白质数据库, 免费开源提供 3D 大分子结构数据, 与 40 多个外部生物数据资源集成的 PDB 数据。针对最近发布的 COVID-19 相关的 SARS-CoV-2 蛋白和宿主细胞蛋白的结构, 包括来源 SARS-CoV-2 的 Spike 蛋白。CATH<sup>[34]</sup>识别蛋白质结构中的结构域, 并将其分类为进化超家族, 从而提供功能注释 (包括 GO、EC 和分类信息) 和蛋白质 3D 结构的最新和系统的结构及其分类信息的数据库。因而包含 Spike 蛋白的 3D 结构和以 FunFam 为主要的注释信息。CoV3D、SCoV2-MD 是分别基于部分冠状病毒和 SARS-CoV-2 开发的, 而 RCSB PDB、ESC 是针对蛋白质结构设计, 这些数据库都存储不同功能的蛋白质 3D 结构和注释信息, 其中包含部分 Spike 蛋白。

ESC<sup>[35]</sup>是 SARS-CoV-2 免疫逃逸变体数据库, 精确整合了文献中人工整理的 SARS-CoV-2 变体。其包含 5 258 个变体, 提供 SARS-CoV-2 逃逸变体的广泛注释信息。其中 2 068 个独特的变体映射到 Spike 蛋白、ORF1ab 和 ORF3a, 通过实验和计算分析 2 060 个变体与编码 Spike 蛋白的基因有关。Virus Viz 是一个用于病毒核苷酸和氨基酸变体的比较分析和有效可视化的生物信息学工具, 它主要储存 SARS-CoV-2 变体, 该工具还支持其他病毒种类 (SARS-CoV、MERS-CoV、登革热、埃博拉病毒), 尤其在 Spike 蛋白变体分析效果很好。这些数据库收录了部分 Spike 蛋白的单一信息。目前还没有一个较为全面的 Spike 蛋白信息数据库, 因此, 建立一个全面的、专门存储基于所有病毒的 Spike 蛋白的数据库是必要的和必需的。同时, VirusViz<sup>[36]</sup>、CATH 分别用于 Spike 蛋白的可视化和 FunFam 注释。国内外研究人员<sup>[37-39]</sup>开发的算法和计算模型也可用于 SARS-CoV-2 的 Spike 蛋白变异和表位预测。但目前, 还没有开发一个专门用于病毒 Spike 蛋白结构、突变和表位预测及计算的生物信息学工具, 这也要求生物信息研究人员亟需开发针对 Spike 蛋白的分析工具。

### 3.2 Spike 蛋白数据库建立概况

目前, 没有一个专门的 Spike 蛋白数据库建立, 课题组针对这个问题, 做了大量工作, 已建立了一个

专门收集 Spike 蛋白及核酸信息的数据库 (the database of viruses Spike protein, DVSP)。该数据库是病毒 Spike 蛋白数据库, 用户可以通过网址 (<http://yeyn.group:90/>) 免费访问。DVSP 包含两部分病毒的刺突蛋白及刺突蛋白翻译前核苷酸数据分别是 35 579 个和 35 175 个, 其中, 来源 SARS-CoV-2 的 16 027 个刺突蛋白和 15 623 个刺突蛋白翻译前核苷酸信息。

数据库中以表格形式提供数据浏览, 数据表格结构组成为每个刺突蛋白及核酸两种数据对应相结合, 是为了给更多研究者在使用数据库时能同时获取两种类型的信息, 即刺突蛋白信息及翻译前核苷酸信息。数据库还为用户提供浏览、搜索、BLAST 功能。为研究刺突蛋白生物性质的生物学家们提供数据基础, 能提高他们的工作质量, 挖掘更多的机制研究、研发更有效的抗病毒药物和疫苗等。

## 4 结论与展望

本文指出了 Spike 蛋白在病毒感染细胞中的重要作用; 同时, 受本次世界范围内的 COVID-19 疫情影响, 其测序和注释数据沉积的速度可能会持续并呈指数增长。生物和生物医学家们更偏向采用高通量蛋白质组学技术来更好地理解基础分子生物学和疾病发生过程, 丰富的蛋白质组学数据使研究人员能够构想并提出新的、复杂的生物学问题<sup>[40]</sup>。此外, 更广泛的数据参考能降低实验结果的偏好性和增加研究结果的普适性。因此, 有必要开发一个专门的、持续更新的 Spike 蛋白信息库。

目前, 与 Spike 蛋白相关的生物信息学工具和数据库是基于蛋白、病毒或其他研究目的开发的, 不适用于病毒 Spike 蛋白研究, 具有片面性和局限性。如 CoV3D 数据库是为存储部分冠状病毒(SARS-CoV、MERS-CoV、SARS-CoV-2) 的蛋白结构而开发的, 包含与这 3 种病毒相关的 Spike 蛋白结构和序列信息。建立丰富且非冗余的 Spike 蛋白数据库为研究人员提供参考价值, 能基于 Spike 蛋白而提出新的、更复杂的生物学问题。开发具有功能多样性且高效的生物信息学工具, 以便能够精准且快速的预测和计算 Spike 蛋白结构、变异序列等<sup>[41]</sup>。

### [参考文献] (References)

- [1] BOSCH B J, van der ZEE R, de HAAN C A, et al. The coronavirus Spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex[J]. *Journal of Virology*, 2003, 77(16): 8801-8811.
- [2] WATANABE Y, ALLEN J D, WRAPP D, et al. Site-specific glycan analysis of the SARS-CoV-2 Spike[J]. *Science*, 2020, 369(6501): 330-333.
- [3] STERNBERG A, NAUJOKAT C. Structural features of coronavirus SARS-CoV-2 Spike protein: targets for vaccination[J]. *Life Sciences*, 2020, 257: 118056.
- [4] CAI Y F, ZHANG J, XIAO T S, et al. Distinct conformational states of SARS-CoV-2 Spike protein[J]. *Science*, 2020, 369(6511): 1586-1592.
- [5] WANG Q H, WONG G, LU G W, et al. MERS-CoV Spike protein: targets for vaccines and therapeutics[J]. *Antiviral Research*, 2016, 133: 165-177.
- [6] WANG F, KREAM R M, STEFANO G B. An evidence based perspective on mRNA-SARS-CoV-2 vaccine development[J]. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, 2020, 26: e924700.
- [7] COMPANS R W, KLENK H D, CALIGUIRI L A, et al. Influenza virus proteins. I. Analysis of polypeptides of the virion and identification of Spike glycoproteins[J]. *Virology*, 1970, 42(4): 880-889.
- [8] WAGNER R R, KILEY M P, SNYDER R M, et al. Cytoplasmic compartmentalization of the protein and ribonucleic acid species of vesicular stomatitis virus[J]. *Journal of Virology*, 1972, 9(4): 672-683.
- [9] McSHARRY J J, COMPANS R W, CHOPPIN P W. Proteins of vesicular stomatitis virus and of phenotypically mixed

- vesicular stomatitis virus-simian virus 5 virions[J]. *Journal of Virology*, 1971, 8(5): 722-729.
- [10] GAROFF H, SIMONS K. Location of the Spike glycoproteins in the Semliki Forest virus membrane[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1974, 71(10): 3988-3992.
- [11] KIELIAN M, HELENIUS A. pH-induced alterations in the fusogenic Spike protein of Semliki Forest virus[J]. *The Journal of Cell Biology*, 1985, 101(6): 2284-2291.
- [12] HELENIUS A, MOREIN B, FRIES E, et al. Human (HLA-A and HLA-B) and murine (H-2K and H-2D) histocompatibility antigens are cell surface receptors for Semliki Forest virus[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1978, 75(8): 3846-3850.
- [13] WANG Q H, ZHANG Y F, WU L L, et al. Structural and functional basis of SARS-CoV-2 entry by using human ACE2[J]. *Cell*, 2020, 181(4): 894-904.
- [14] COUTARD B, VALLE C, de LAMBALLERIE X, et al. The Spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade[J]. *Antiviral Research*, 2020, 176: 104742.
- [15] 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, 181(2): 281-292.
- [16] WAN Y S, SHANG J, GRAHAM R L, et al. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus[J]. *Journal of Virology*, 2020, 94(7): e00127-20.
- [17] LI W H, WONG S K, LI F, et al. Animal origins of the severe acute respiratory syndrome coronavirus: insight from ACE2-S-protein interactions[J]. *Journal of Virology*, 2006, 80(9): 4211-4219.
- [18] GRAHAM R L, BARIC R S. Recombination, reservoirs, and the modular Spike: mechanisms of coronavirus cross-species transmission[J]. *Journal of Virology*, 2010, 84(7): 3134-3146.
- [19] LEE C H, KOOHY H. In silico identification of vaccine targets for 2019-nCoV[J]. *F1000 Research*, 2020, 9: 145.
- [20] 高俊, 林勇平, 肖洪广. SARS 冠状病毒 Spike 蛋白上的 CoV 的受体结合域的研究进展[J]. *广州医学院学报*, 2006, 34 (5): 59-62.
- GAO J, LIN Y P, XIAO H G. Research progress of CoV receptor binding domain on Spike protein of SARS coronavirus[J]. *Academic Journal of Guangzhou Medical College*, 2006, 34(5): 59-62. (in Chinese)
- [21] YURINA V. Coronavirus epitope prediction from highly conserved region of Spike protein[J]. *Clinical and Experimental Vaccine Research*, 2020, 9(2): 169-173.
- [22] HE J L, HUANG F, ZHANG J H, et al. Vaccine design based on 16 epitopes of SARS-CoV-2 Spike protein[J]. *Journal of Medical Virology*, 2021, 93(4): 2115-2131.
- [23] HSIEH C L, GOLDSMITH J A, SCHAUB J M, et al. Structure-based design of prefusion-stabilized SARS-CoV-2 Spikes[J]. *Science*, 2020, 369(6510): 1501-1505.
- [24] PALLESEN J, WANG N S, CORBETT K S, et al. Immunogenicity and structures of a rationally designed prefusion MERS-CoV Spike antigen[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2017, 114(35): E7348-E7357.
- [25] WRAPP D, WANG N S, CORBETT K S, et al. Cryo-EM structure of the 2019-nCoV Spike in the prefusion conformation[J]. *Science*, 2020, 367(6483): 1260-1263.
- [26] SCHOCH C L, CIUFO S, HOTTON C, et al. NCBI Taxonomy: a comprehensive update on curation, resources and tools[J]. *Database*, 2020(2020): baaa062.
- [27] SHU Y L, McCAULEY J. GISAID: global initiative on sharing all influenza data - from vision to reality[J]. *Euro Surveillance*, 2017, 22(13): 30494.
- [28] The UniProt Consortium. UniProt: the universal protein knowledgebase[J]. *Nucleic Acids Research*, 2017, 45(D1): D158-D169.
- [29] SONG S H, MA L, ZOU D, et al. The global landscape of SARS-CoV-2 genomes, variants, and haplotypes in 2019nCoV[J].

- Genomics, Proteomics & Bioinformatics, 2020, 18(6): 749-759.
- [30] de SILVA N H, BHAI J, CHAKIACHVILI M, et al. The ensembl COVID-19 resource: ongoing integration of public SARS-CoV-2 data[J]. *Nucleic Acids Research*, 2021, 50(D1): D765-D770.
- [31] GOWTHAMAN R, GUEST J D, YIN R, et al. CoV3D: a database of high resolution coronavirus protein structures[J]. *Nucleic Acids Research*, 2022, 49(D1): D282-D287.
- [32] TORRENS-FONTANALS M, PERALTA-GARCÍA A, TALARICO C, et al. SCoV2-MD: a database for the dynamics of the SARS-CoV-2 proteome and variant impact predictions[J]. *Nucleic Acids Research*, 2022, 50(D1): D858-D866.
- [33] BURLEY S K, CHARMI B, BI C, et al. RCSB protein data bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences[J]. *Nucleic Acids Research*, 2021, 49(D1): D437-D451.
- [34] SILLITOE I, BORDIN N, DAWSON N, et al. CATH: increased structural coverage of functional space[J]. *Nucleic Acids Research*, 2021, 49(D1): D266-D273.
- [35] ROPHINA M, PANDHARE K, SHAMNATH A, et al. ESC: a comprehensive resource for SARS-CoV-2 immune escape variants[J]. *Nucleic Acids Research*, 2022, 50(D1): D771-D776.
- [36] BERNASCONI A, GULINO A, ALFONSI T, et al. VirusViz: comparative analysis and effective visualization of viral nucleotide and amino acid variants[J]. *Nucleic Acids Research*, 2021, 49(15): e90.
- [37] LI W D, LI L, SUN T, et al. Spike protein-based epitopes predicted against SARS-CoV-2 through literature mining[J]. *Medicine in Novel Technology and Devices*, 2020, 8: 100048.
- [38] VASHI Y, JAGRIT V, KUMAR S. Understanding the B and T cell epitopes of spike protein of severe acute respiratory syndrome coronavirus-2: A computational way to predict the immunogens[J]. *Infection, Genetics and Evolution*, 2020, 84: 104382.
- [39] TRIVERI A, SERAPIAN S A, MARCHETTI F, et al. SARS-CoV-2 Spike protein mutations and escape from antibodies: a computational model of epitope loss in variants of concern[J]. *Journal of Chemical Information and Modeling*, 2021, 61(9): 4687-4700.
- [40] CHEN C M, HUANG H Z, WU C H. Protein bioinformatics databases and resources[J]. *Methods in Molecular Biology*, 2011, 694: 3-24.
- [41] YE Y N, HUA Z G, HUANG J, et al. CEG: a database of essential gene clusters[J]. *BMC Genomics*, 2013, 14(1): 769.

(责任编辑: 段桃)