

DOI:10.13350/j.cjpb.230402

• 论著 •

SARS-CoV-2 结构蛋白 S 和 N 的生物信息学 比较分析及应用研究^{*}

闫静静¹, 迟晓妍², 卢佳琪¹, 薛庆节^{1**}, 王晖^{1**}

(1. 济宁医学院基础医学院, 山东济宁 272067; 2. 济宁医学院法医学与医学检验学院)

【摘要】 目的 采用生物信息学方法分析新型冠状病毒(SARS-CoV-2)重要结构蛋白 S 和 N。将 S 基因和 N 基因融合并构建重组质粒, 重组质粒与骨架质粒在 293T 细胞中包装后获得重组腺病毒, 为利用其融合基因进行疫苗研发提供理论和实验依据。 **方法** 从 NCBI 数据库中获取 S、N 蛋白的基因组序列和氨基酸序列, 采用生物信息学分析工具 SOPMA、IEDB、BLAST、Immunomedicine Group、UniProt 预测并分析 S 和 N 蛋白的理化性质、二级结构和三级结构、B 细胞表位和 T 细胞表位、抗原决定簇以及相互作用蛋白。分别以 S-pcDNA 和 N-pcDNA 质粒为模板, 以 S-F1、S-F2 和 N-F3、和 N-F4 为引物 PCR 扩增目的基因; 通过 T4 DNA 连接酶连接获得 S 基因和 N 基因的融合基因, 以连接后的融合基因为模板, 以 S-F1 和 N-F4 为引物采用重叠 PCR 扩增融合基因 S-N。将融合基因 S-N 与穿梭表达载体 pDC316-mCMV-EGFP 连接, 获得重组质粒, 然后与骨架质粒 pBHGlox(delta)E1, 3Cre 共转染 HEK293T 细胞, 获得重组腺病毒。

结果 S 蛋白由 1 269 个氨基酸组成, 分子式为 $C_{6325}H_{9752}N_{1654}O_{1886}S_{54}$, 原子总数为 19 671, 理论等电点 pI 为 6.35; 该蛋白含有一个跨膜长度为 23 个氨基酸序列的跨膜螺旋(1210-1232), 二级结构以无规则卷曲(占 43.89%)为主, 属于稳定的疏水性蛋白; N 蛋白由 419 个氨基酸组成, 分子式为 $C_{1976}H_{3150}N_{608}O_{629}S_7$, 原子总数为 6 370, 理论等电点 pI 为 10.09; 二级结构以无规则卷曲(占 58.23%)为主, 属于不稳定的疏水性蛋白。选择出 S 蛋白最具优势的抗原表位 4 个, 选择出 N 蛋白最具优势的抗原表位 3 个; 选出 S 和 N 蛋白 Th 表位各 2 个优势序列; 选出 CTL 优势 S 蛋白表位 HLA-A 2 个, 选出 CTL 优势 N 蛋白表位 HLA-A 3 个, HLA-B 2 个; 从 S 蛋白抗原决定簇选择 3 个优势抗原决定簇, 从 N 蛋白选择出 2 个优势的抗原决定簇; 获得 10 个与 S 蛋白相互作用的蛋白, 6 个与 N 蛋白相互作用的蛋白。在此基础上, 成功构建了融合基因 S-N 重组腺病毒并在 HEK293T 细胞中表达。 **结论** S 和 N 蛋白理化性质、免疫原性、抗原决定簇及同源性等均符合疫苗的基本要求, 为 SARS-CoV-2 结构蛋白 S 和 N 融合基因重组疫苗的研究奠定了基础。

【关键词】 新型冠状病毒; 理化性质; 生物信息学分析; 重组腺病毒

【中图分类号】 R373

【文献标识码】 A

【文章编号】 1673-5234(2023)04-0377-08

[*Journal of Pathogen Biology*. 2023 Apr;18(4):377-384.]

Comparative bioinformatics analysis of structural proteins s and N of SARS-CoV-2 and their application

YAN Jing-jing¹, CHI Xiao-yan², LU Jia-qi¹, XUE Qing-jie¹, WANG Hui¹ (1. School of Basic Medical, Jining Medical University, Jining 272067, Shandong, China; 2. School of Forensic and Laboratory Medicine, Jining Medical University) ^{***}

【Abstract】 Objective SARS-CoV-2 important structural protein S and N were analyzed by bioinformatics method. The S and N genes were fused and the recombinant plasmid was constructed by experimental methods. Then the recombinant plasmid and the skeleton plasmid were packaged in HEK293T cells to obtain the recombinant adenovirus, which provided a theoretical and experimental basis for vaccine research by using the fusion gene. **Methods** The genomic and amino acid sequences of the S and N proteins were obtained from NCBI database, the bioinformatics analysis tools SOPMA, IEDB, BLAST, Immunomedicine Group, UniProt to predict and analyse the physicochemical properties, secondary structure, tertiary structure, B cell and T cell epitopes, antigenic determination clusters, and interacting proteins. S-pcDNA and N-pcDNA plasmids were used as templates, S-F1, S-F2, N-F3 and N-F4 were used as primers to amplify the target genes by PCR. The S and N genes were ligated by T4 DNA ligase to get the fused gene after ligating, and the fusion gene S-N was amplified by overlapping PCR with S-F1 and N-F4 primers, the fusion gene S-N was ligated

^{*} **【基金项目】** 山东省重点研发计划项目(No. 2018GSF118137); 济宁市重点研发计划项目(No. 2019SMNS020); 济宁医学院重点项目(No. 16008); 贺林院士工作站重点项目(No. JYHL2019ZD03); 省级大创项目(No. S202110443023); 济宁医学院大创项目(No. cx2022023z)

^{**} **【通讯作者】** 王 晖, E-mail: 3052w@163.com; 薛庆节, E-mail: qjxue9797@126.com

【作者简介】 闫静静(2002-), 女, 山东梁山人, 本科在读, 主要研究方向: 微生物分子生物学; E-mail: 676370420@qq.com
闫静静和迟晓妍为并列第一作者。

with the shuttle expression vector pDC316-mCMV-EGFP to obtain the recombinant plasmid, then the pDC316-mCMV-EGFP-S-N was cotransfected into 293T cells with pBHGlox(delta) E1,3Cre to obtain recombinant adenovirus. **Results**

S protein of SARS-CoV-2 is composed of 1269 amino acids, the formula is $C_{6325}H_{9752}N_{1654}O_{1886}S_{54}$, the total number of atoms is 19 671, and it's theoretical isoelectric point (pI) is 6. 35. The protein contains a transmembrane helix (1210-1232) of 23 amino acid sequences in length. It is a stable hydrophobic protein with mainly uncoiled (43. 89%). The N protein consists of 419 amino acids with the formula $C_{1976}H_{3150}N_{608}O_{629}S_7$ and a total number of atoms of 6 370. It's theoretical isoelectric point (pI) is 10. 09. It's protein is mainly unregularly coiled (58. 23%), which is unstable hydrophobic protein. In this research, 4 dominant epitopes of S protein, 3 dominant epitopes of N protein, 2 dominant sequences of Th epitope of S and N proteins were selected. At the same time, 2 dominant CTL epitope HLA-A of S protein, 3 HLA-A and 2 HLA-B with N protein were selected. 3 dominant antigenic determinants were selected from S protein antigenic determinants and 2 from N protein. 10 proteins interacting with S proteins and 6 with N were obtained. At last, the recombinant adenovirus was constructed and expressed in HEK293T cells successfully. **Conclusion** In our research, the physical and chemical properties of protein, immunogenicity, antigen determination cluster and homology of SARS-CoV-2 S and N proteins all met the basic requirements of the vaccine, the recombinant adenovirus vaccine was successfully packaged in HEK293T cells, which laid the foundation for the vaccine research with fusion gene of SARS-CoV-2 structural proteins S and N.

【Key words】 SARS-CoV-2; physical and chemical properties; bioinformatics analysis; recombinant adenovirus

新型冠状病毒(SARS-CoV-2)属于 β 冠状病毒属,由16个非结构蛋白、4个结构蛋白、9个辅助蛋白组成^[1],其基因长度约为30 kb,其编码结构蛋白和辅助蛋白的编码区约为10 kb,编码的结构蛋白主要为刺突蛋白S、膜蛋白M、包膜蛋白E和核衣壳蛋白N,其中刺突蛋白S是与病毒入侵密切相关的重要蛋白^[2-4]。S蛋白由S1和S2两个功能亚单位组成,远端S1亚基提供一个受体结合位点与宿主细胞受体结合,S2亚基主要负责病毒与细胞膜的融合。因此,S刺突蛋白在SARS-CoV-2对人体宿主细胞的吸附和入侵过程中发挥重要的作用。

N蛋白是SARS-CoV-2的重要组成部分^[5-6],具有高免疫原性,在感染过程中大量表达^[7]。N蛋白与SARS-CoV-2基因组缠绕形成核衣壳,在子代病毒的复制过程中发挥重要作用^[8]。另外,N蛋白还具有RNA分子伴侣活性,是人体干扰素的拮抗剂^[9-10]。N蛋白包含两个RNA结合域,周围是内在无序区域,无序区域的磷酸化促进蛋白质的转录功能,但其潜在的机制尚不清楚。N蛋白还参与病毒mRNA的转录和复制,影响机体细胞骨架的形成和免疫调节^[11],调控其细胞代谢和细胞周期,诱导感染后机体的体液和细胞免疫反应^[12-13]。因此,N蛋白可作为病毒检测的标志性蛋白,已被广泛应用于血清学检测^[14]。

本研究对S蛋白和N蛋白进行生物信息学预测,对2种蛋白的理化性质、结构、B细胞及T细胞抗原表位、抗原决定簇、相互作用蛋白进行比较和分析,并在此基础上构建融合基因S-N并与穿梭表达载体pDC316-mCMV-EGFP连接,获得重组质粒,然后与骨架质粒pBHGlox(delta)E1,3Cre共转染HEK293T

细胞,获得重组腺病毒,为SARS-CoV-2结构蛋白S和N融合基因疫苗研究奠定基础。

材料与方法

1 S和N氨基酸序列及其编码基因序列的获取

获取S蛋白的基因组序列(ID:43740568)和氨基酸序列,查询出SARS-CoV-2 S蛋白由1269个氨基酸组成,氨基酸序列为:MFVFLVLLPLVSSQCVNLTT RTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQD LFLPFFSNVTWFHAI GTNGTKRFDNPVLPFNDG VYFASTEKSNIIRGWIFGTTLD SKTQSL LIVNNA TNVVIKVCEFQFCNDPFLGVYHKNNKSWMESE FRVYSSANNCTFEYVSQPFLMDLEGKQGNFKN LREFVFKNIDGYFKIYSKHTPINLVRDLPQG FSA LEPLVDLPIGINITRFQTLLALHRSYLT PGDSSSG WTAGAAAYYVGYLQPRTFLLKYNENG TITDAV DCALDPLSETKCTLKSFTVEKGIYQTSNFRVQP TESIVRFPNITNLC PFGEVFNATRFASVYAWN R KRISNCVADYSVLYNSASFSTFKCYGVSP TKLND LCFTNVYADSFVIRGDEV RQIAPGQTGKIADYN YKLPDDFTGCVIAWNSNNLDSKVGGNYNLYR LFRKSNLKP FERDISTEIYQAGSTPCNGVEGFNC YFPLQSYGFQPTYGVGYQPYRVVVL SFELLHAP ATVCGPKKSTNLVKNKCVNFNFNGLTGTG VLT ESNKKFLPFQQFGRDIDDTDAVRDPQTLE ILDI TPCSFGGVSVITPGTNTSNQVAVLYQGVNCTEV PVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGA EHVNNSECDIPIGAGICASYQTQTNSHRRAR SV ASQSIIAYTMSLGAENSVAYSNNNSIAIPINFTISVT

TEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFQAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAA
LQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFN
SAIGKIQDLSSTASALGKLQDVVNQNAQALN
TLVKQLSSNFGAISSVLNDILARLDKVEAEVQIDR
LITGRLQSLQTYVTQQLIRAAEIRASANLAATKM
SECVLGQSKRVDFCGKGYHLMSPQASAPHGTVF
LHVTYVPAQEKNFTTAPAICHDGKAHFPREGVF
VSNGTHWVFTQRNFYEPQIITHTNTFVSGNCDV
VIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSP
DVDLGDISGINASVVNIQKEIDRLNEVAKNLNE
SLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVM
VTIMLCCMTSCCCLKGCCSCGSCCKFDEDDSE
PVLKGVKLYHT。

获取其 N 蛋白的基因组序列(ID:43740575)和氨基酸序列,查询出 SARS-CoV-2 N 蛋白由 419 个氨基酸组成,氨基酸序列为:MSDNGPQNQRNAPRITFG
GPSDSTGSNQNGERSGARSKQRRPQGLPNNTAS
WFTALTQHGKEDLKFRGQGVPIINTNSSPDDQIG
YYRRATRRIRGGDGKMKDLSPRWYFYLLGTGPE
AGLPYGANKDGIWVATEGALNTPKDHIGTRNL
ANNAIIVLQLPQGTTLPKGFYAEGRGGSQASS
RSSSRSRNSSRNSTPGSSKRTSPARMAGNGGDAA
LALLLLDRLNQLESKMSGKGQQQQGQTVTKKS
AAEASKKPRQKRTATKAYNVTQAFGRRGPEQT
QGNFGDQELIRQGTDYKHWPQIAQFAPSASAFF
GMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKD
QVILLNKHIDAYKTFPPTPEPKDKKKKKADETQA
LPQRQKKQQTVTLLPAADLDDFSKQLQQSMSS
ADSTQA。

2 S 和 N 蛋白的生物信息学分析

利用在线软件 ProtParam (<https://web.expasy.org/protparam>) 分析 S 和 N 蛋白的理化性质, 包括氨基酸数目、分子式、氨基酸组成、分子质量、等电点等; 利用在线软件 ProtScale (<http://web.expasy.org/scale>) 分析 SARS-CoV-2 S 和 N 蛋白的亲/疏水性。利用在线软件 (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) 分析 S 和 N 蛋白的二级结构。利用 IEDB 软件 (https://www.iedb.org/home_v3.php) 综合预测 B 细胞表位。输入氨基酸序列, 通过 Karplus & Schulz 分析其柔韧性、Kolaskar & Tongaonkar 预测抗原性、Emili 预测抗原表面可及性、Parker 预测亲水性及

Chou&Fasman 预测氨基酸编码蛋白质的 β 转角,对 B 细胞表位进行联合预测分析。 β 转角与无规则卷曲多出现在蛋白表面,为凸出疏松结构,与抗体更有利结合,成为 B 细胞抗原表位可能性较大。另外选取亲水性与表面可及性良好、柔韧性强、抗原性强、卷曲和转角含量较多的区域,作为候选 B 细胞表位。利用 IEDB 软件(https://www.iedb.org/home_v3.php)预测 Th 淋巴细胞和 T 淋巴细胞表位。利用在线软件 Immunomedicine Group(<http://imed.med.ucm.es/Tools/antigenic.pl>)预测 S 和 N 蛋白抗原决定簇。利用 UniProt 数据库(<https://www.uniprot.org/>)预测 S 和 N 蛋白的相互作用蛋白。

3 融合基因 S-N 重组腺病毒的构建与表达

3.1 引物设计 根据 S 基因序列设计扩增目的基因需要的引物序列,在 5' 端引入限制性内切酶 Pst I 识别序列,并在其下游引物加入 Linker(Gly₄Ser)₃:(5'-GCTGCCGCCACCGCCGCTTCCGCCACCGCCGCTTCCACCGCCACC-3');S-F1:CGCCTGCAGATGTTCTGCTTCTGGTCCTGCTGCC;S-F2:GCTGCCGCCACCGCCGCTTCCGCCACCGCCGCTTCCACCGCCACCGGTGATCAGGCTCTCGTTC。

根据 N 基因序列设计扩增目的基因需要的引物序列,在 5' 端引入限制性内切酶 HindⅢ 识别序列,并在其上游引物加入 Linker(Gly₄Ser)₃;N-F3:GCTGC CGCCACCGCCGCTTCCGCCACCGCCGCTTCCAC CGCCACCATGTCTGATAATGGACCC;N-F4:GC GAAGCTTTGATTAGGCCTGAGTTGAGTCAGC ACTGC。S、N 基因融合示意图及腺病毒的穿梭表达质粒与骨架质粒模式图见图 1。

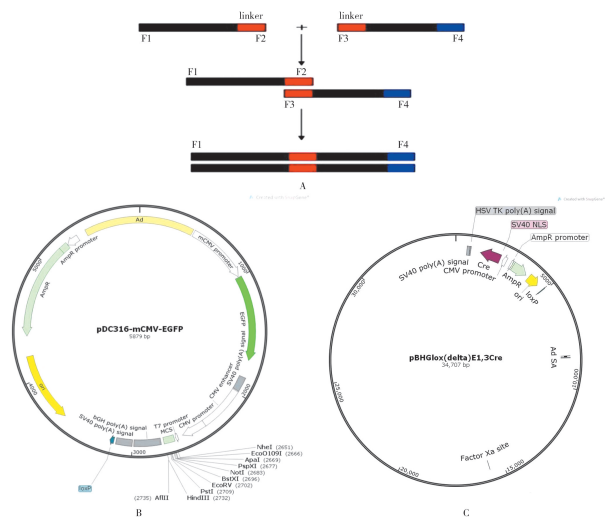
3.2 目的基因 S 和 N 扩增 分别以 S-pcDNA 和 N-pcDNA 质粒为模板,以 S-F1 和 S-F2、N-F3 和 N-F4 为引物 PCR 扩增目的基因。取 PCR 产物进行 1% 琼脂糖凝胶电泳,在紫外透视仪下切出符合大小的片段,使用琼脂糖凝胶 DNA 回收试剂盒进行回收。

3.3 重叠 PCR 获得融合基因

通过 T4 DNA 连接酶连接获得 S 基因和 N 基因,将所需材料混匀后置于 16℃ 金属浴过夜,得到连接后的融合基因,以连接后的融合基因为模板,以 S-F1 和 N-F4 为引物通过重叠 PCR 扩增融合基因 S-N。扩增产物经 1% 琼脂糖凝胶电泳分离后在紫外透视仪下切出符合大小的片段,使用琼脂糖凝胶 DNA 回收试剂盒进行回收。

3.4 重组质粒的构建与筛选鉴定 利用 Pst I 和 HindIII 对融合基因 S-N 与穿梭表达质粒 pDC316-mCMV-EGFP 分别进行双酶切。将回收的融合基因 S-N 和穿梭表达质粒 pDC316-mCMV-EGFP (图 1B) 以 T4 DNA 连接酶连接 (金属浴 16 ℃ 过夜), 所得连

接产物转化感受态细胞 *E. coli* Stbl3。挑取抗性菌落接种于 LB 液体培养基(氨苄抗性(50 $\mu\text{g/ml}$),于 37 $^{\circ}\text{C}$ 、160 r/min 摇床培养 12 h,使用质粒小提试剂盒提取质粒,进行双酶切鉴定、PCR 鉴定,并送上海生工公司测序,测序无误后将重组质粒命名为 S-N-pDC316。



A S、N 基因融合示意图 B 腺病毒的穿梭表达质粒 pDC316-mCMV-EGFP 腺病毒骨架质粒 pHGlox(delta)E1.3Cre

图 1 S、N 基因融合示意图及腺病毒的穿梭表达质粒与骨架质粒模式图

A The fusing schematic diagram of S and N gene B Shuttle expression plasmid of adenovirus C Adenovirus backbone plasmid pHGlox(delta)E1.3Cre

Fig. 1 The fusing schematic diagram of S and N genes and the pattern diagram of shuttle plasmid and backbone plasmid of adenovirus

结果

1 蛋白的理化性质

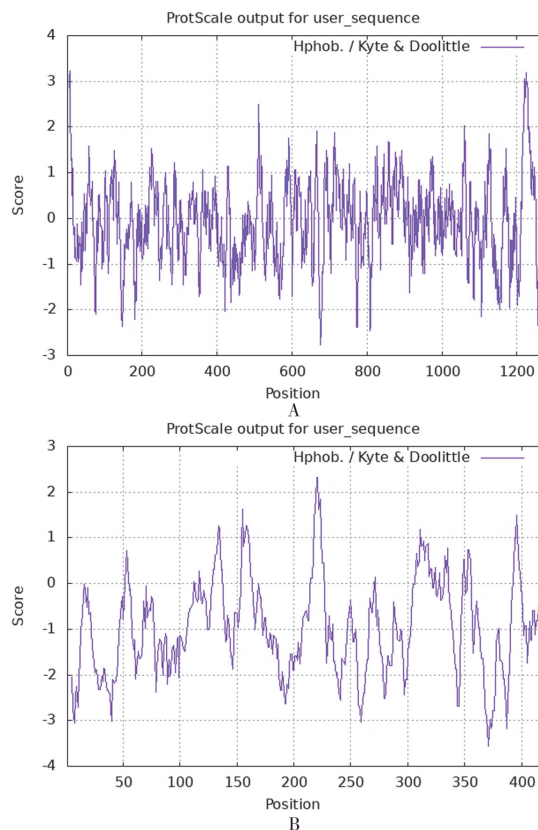
预测 S 蛋白理论等位点为 6.35,相对分子质量为 140.87220×10^3 ,由 1 270 个氨基酸构成。当蛋白 N 端的残基为 Met 时,该蛋白在哺乳动物网状细胞中的半衰期为 30 h,在大肠埃希菌体内的半衰期 >10 h。S 蛋白平均疏水性为 -0.074(表 1);ProtScale 预测 S 基因编码蛋白为疏水性蛋白,第 676 位氨基酸亲水性最高为 -2.767,第 7 位为氨基酸疏水性得分最高为 3.222(图 2A)。

表 1 S 蛋白的理化性质预测
Table 1 Physicochemical properties of S protein

S 蛋白的理化性质 Physicochemical properties of S protein	预测结果 Predicted result
氨基酸数量	1270
相对分子质量	140 872.20
理论等位点	6.35
分子式	$\text{C}_{6325}\text{H}_{9752}\text{N}_{1654}\text{O}_{1886}\text{S}_{54}$
不稳定指数	32.82
脂肪族氨基酸指数	84.95

N 蛋白理论等位点为 10.09,相对分子质量为 45.71286×10^3 ,由 419 个氨基酸构成。当蛋白 N 端

的残基为 Met 时,该蛋白在哺乳动物网状细胞中的半衰期为 30 h,在大肠埃希菌体内的半衰期 >10 h。N 蛋白的平均疏水性为 -0.967(表 2);ProtScale 预测 N 基因编码蛋白为疏水性蛋白,第 371 位氨基酸亲水性最高为 -3.556,第 220 位、第 221 位为氨基酸疏水性得分最高为 2.322(图 2B)。



A S 蛋白的亲(疏)水性分析 B N 蛋白的亲(疏)水性分析
图 2 S 和 N 蛋白的亲(疏)水性分析

A Hydrophilicity and hydrophobicity analysis of S protein B Hydrophilicity and hydrophobicity analysis of N protein

Fig. 2 Hydrophilicity/hydrophobicity analysis of S/N protein

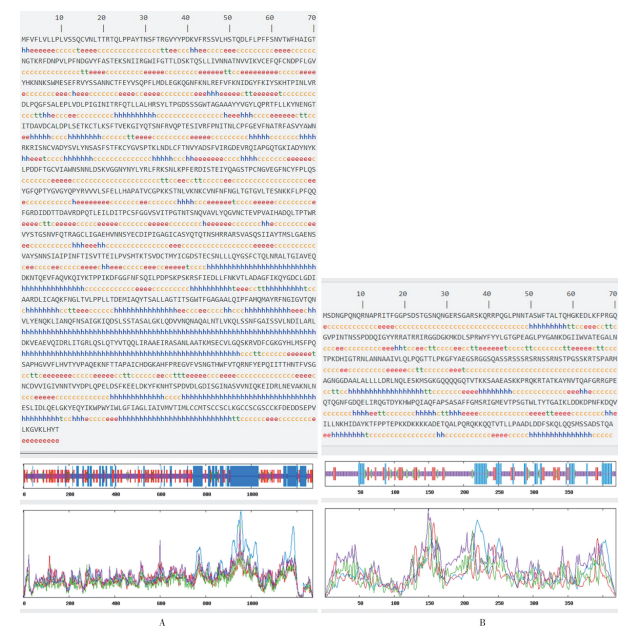
表 2 N 蛋白的理化性质预测
Table 2 Physicochemical properties of N protein

N 蛋白的理化性质 Physicochemical properties of N protein	预测结果 Predicted results
氨基酸数量	419
相对分子质量	45712.86
理论等位点	10.09
分子式	$\text{C}_{1976}\text{H}_{3150}\text{N}_{608}\text{O}_{629}\text{S}_7$
不稳定指数	55.42
脂肪族氨基酸指数	53.46

2 蛋白的二级结构

利用在线分析软件 SOPMA 预测 S 蛋白的二级结构, α -螺旋(Hh)385 个,占 30.34%; β -转角(Tt)44 个,占 3.47%; β -折叠(Ee)283 个,占 22.30%;无规则卷曲(Cc)557 个,占 43.89%(图 3A)。N 蛋白的二级结构中 α -螺旋(Hh)78 个,占 18.62%; β -转角(Tt)30 个,占 7.16%; β -折叠(Ee)67 个,占 15.99%;无规则

卷曲(Cc)244个,占58.23%(图3B)。



A S蛋白二级结构预测 B N蛋白二级结构预测(蓝、红、绿、紫分别代表 α -螺旋、 β -折叠、 β -转角和无规则卷曲)

图3 S和N抗原蛋白二级结构预测

A Secondary structure prediction of S B Secondary structure prediction of N

Fig. 3 Secondary structure prediction of S/N

3 B细胞表位

利用IEDB软件分析预测SARS-CoV-2 S的可及性,再结合S蛋白的线性表位、 β -转角、亲疏水性、抗原性4个方面综合分析筛选出B细胞优势表位,其中最具优势的抗原表位为25-31、419-421、670-681、769-775氨基酸区段(表3)。

表3 S蛋白优势B细胞表位预测

Table 3 Prediction of dominant B cell epitopes of S protein		
序号	氨基酸位置	氨基酸序列
Order	The position of amino acid	The sequence of amino acid
1	25aa-31aa	PPAYTNS
2	419aa-421aa	YKL
3	670aa-681aa	YQTQTNSHRRAR
4	769aa-775aa	EQDKNTQ

根据N蛋白的柔韧性、表面可及性、线性表位、 β -转角、亲疏水性、抗原6个方面筛选B细胞表位,其中最具优势的抗原表位为4-10、295-300、384-390氨基酸区段(表4)。

表4 N蛋白优势B细胞表位预测

Table 4 Prediction of dominant B cell epitopes of N protein		
序号	氨基酸的位置	氨基酸序列
Order	The position of amino acid	The sequence of amino acid
1	4aa-10aa	NGPQNQR
2	295aa-300aa	GTDYKH
3	384aa-390aa	QRQKKQQ

4 Th和CTL细胞表位

利用IEDE软件在线预测SARS-CoV-2 S和SARS-CoV-2 N抗原蛋白的限制性Th细胞表位,筛选其中表型具有最多的表位;在分析SYFPEITHI的预测结果后,对于表型分值评分大于20的氨基酸序列进行选择,以作为候选Th表位结果。综合分析以上结果,候选出S和N蛋白Th的表位结果各2个优势序列(表5、表6)。

表5 S蛋白Th细胞表位分析

Table 5 S protein Th cell epitope analysis				
MHC-II类分子亚型	起始氨基酸位置	氨基酸序列	IEDB得分	SYFPEITHI得分
MHC-II molecular subtype	Startposition	The sequence of amino acid	IEDB Score	SYFPEITHI Score
HLA-DRB1 * 01:01	2	FVFLVLLPLVSSQCV	0.34	32
HLA-DRB1 * 07:01	711	PINFITSVITETLPV	0.45	32

表6 N蛋白Th细胞表位分析

Table 6 N protein Th cell epitope analysis				
MHC-II类分子亚型	起始氨基酸位置	氨基酸序列	IEDB得分	SYFPEITHI得分
MHC-II molecular subtype	Startposition	The sequence of amino acid	IEDB Score	SYFPEITHI Score
HLA-DRB1 * 11:01	83	QIGYYRATRRIRGG	0.39	25.5
HLA-DRB1 * 11:01	81	DDQIGYYRATRRIR	0.42	24.0

利用IEDE软件在线预测SARS-CoV-2 S和SARS-CoV-2 N抗原蛋白的CTL细胞表位,分析9个氨基酸长度的HLA-A、HLA-B两种限制性CTL表位,筛选HLA-A与HLA-B表型进行预测的分值较高的潜在CTL表位数目及选择每一种表型分值均较高的多肽序列,候选出CTL优势S蛋白表位HLA-A为2个;候选出CTL优势N蛋白表位HLA-A为3个,HLA-B为2个(表7、表8)。

表7 S蛋白CTL细胞表位分析

Table 7 S protein CTL cell epitope analysis				
MHC-II类分子亚型	起始氨基酸位置	氨基酸序列	IEDB得分	SYFPEITHI得分
MHC-II molecular subtype	Startposition	The sequence of amino acid	IEDB Score	SYFPEITHI Score
HLA-A	66	DGYFKIYSK	0.55	25
	265	FASVYAWNR	0.45	30

表8 N蛋白CTL细胞表位分析

Table 8 N protein CTL cell epitope analysis				
MHC-II类分子亚型	起始氨基酸位置	氨基酸序列	IEDB得分	SYFPEITHI得分
MHC-II molecular subtype	Startposition	The sequence of amino acid	IEDB Score	SYFPEITHI Score
HLA-A	181	QASSRSSSR	0.29	16
	269	NVTQAFGR	0.25	18.5
	311	ASAFFGMSR	0.16	23
HLA-B	386	QKKQQTIVTL	0.15	21
	395	LPAADLDDF	0.15	20

5 抗原决定簇

利用Immunomedicine Group网站预测S蛋白的氨基酸序列有63个抗原决定簇,该蛋白的平均抗原性

倾向为 1.0418; N 蛋白的氨基酸序列有 15 个抗原决定簇, 该蛋白的平均抗原性倾向为 0.9876。经综合分析, 从 S 蛋白氨基酸序列的 63 个抗原决定簇选择出 3 个优势的抗原决定簇(表 9), 从 N 蛋白氨基酸序列的 15 个抗原决定簇选择出 2 个优势的抗原决定簇(表 10)。

表 9 S 蛋白氨基酸序列的 3 个抗原决定簇
Table 9 3 antigenic determinants of the amino acid sequence of S protein position

序号 Order	起始氨基酸位置 Start position	氨基酸序列 Sequence	终止氨基酸位置 End position
1	33	TRGVYYPDKVFRSSVLHSTQDLFLPFF SNVTWFH	66
2	1034	KRVDFCGKGYHLMSFPQSAPHGVVFL HVTYVPA	1066
3	1210	WYIWLGFIAGLIAIVMVTIMLCCMTSC CSCLKGCCSCGSCCK	1251

表 10 N 蛋白氨基酸序列的 2 个抗原决定簇
Table 10 2 antigenic determinants of the amino acid sequence of N protein position

序号 Order	起始氨基酸位置 Start position	氨基酸序列 Sequence	终止氨基酸位置 End position
1	298	YKHWPQIAQFAPSASAF	314
2	346	FKDQVILLNKHIDAYKT	362

6 相互作用蛋白

利用 UniProt 数据库预测与 S 蛋白相互作用蛋白有 CD81、MX1、ISG15、IFI6、IFI35、IFITM2、OAS2、IFIT1、IFIT3、CD19(图 4A); N 蛋白相互作用蛋白有 EDY31957.1、EDY31956.1、EDY31959.1、EDY31954.1、EDY31955.1、EDY31960.1(图 4B)。

7 目的基因扩增和鉴定

PCR 扩增 S 和 N, 经 1% 琼脂糖凝胶电泳鉴定目的基因片段大小与理论值相同(图 5A); 利用重叠 PCR, 以 S-F1 和 N-F4 为前后引物扩增融合基因 S-N, 经 1% 琼脂糖凝胶电泳鉴定, 扩增产物大小与理论值大致相同(图 5B)。

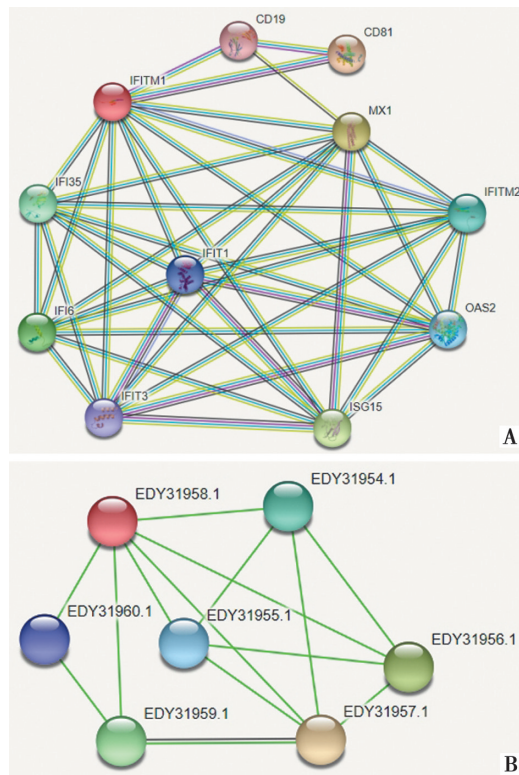
8 融合基因的构建及重组腺病毒的检测

将 PCR 鉴定正确的重组质粒送至上海生工公司测序, 测序结果与已知 SARS-CoV-2 的 S、N 基因序列完全一致; 重组腺病毒用 HEK293T 细胞传代后于荧光显微镜下可看到绿色荧光, 表明重组腺病毒能表达目的蛋白(图 6)。

讨论

虽然世界各国已有新冠病毒疫苗投入应用, 但是由于新冠病毒的易变异性, 疫苗的应用效果均不尽如

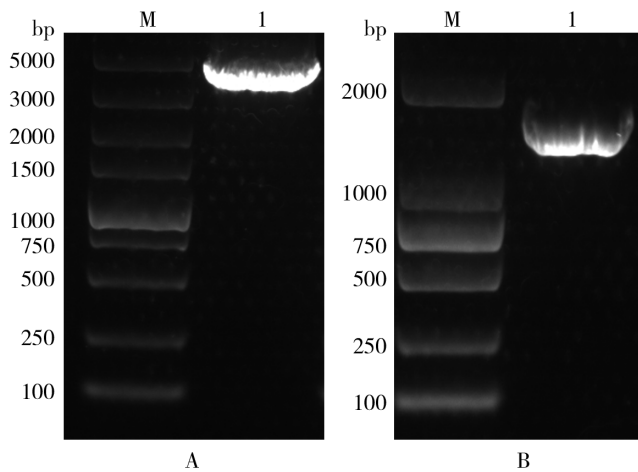
意, 全世界感染人数与死亡人数不断创历史新高, 新冠病毒继续在全球肆虐。本研究利用生物信息学方法综合分析显示 SARS-CoV-2 重要的刺突蛋白 S 为大分子、稳定性的疏水蛋白, 利用其制成疫苗有利于抵抗各种因素的影响, 保持其生物活力。其脂肪族系数为 84.95, 热稳定性好, 制成疫苗后便于储存。S 蛋白的二级结构中 α -螺旋(Hh)385 个, 占 30.34%; β -转角(Tt)44 个, 占 3.47%; β -折叠(Ee)283 个, 占 22.30%; 无规则卷曲(Cc)557 个, 占 43.89%。该蛋白结构较稳定, 易与抗体嵌合, 易形成潜在的抗体表位。S 蛋白可通过磷酸化调节基因表达, 以及细胞周期、分化、代谢、凋亡等多种生物过程, 通过磷酸化和脱磷酸化参与信息传递, 在大多数信号通路中发挥重要作用。另外, SARS-CoV-2 中谷氨酰胺、天冬酰胺、亮氨酸、苯丙氨酸和丝氨酸氨基酸的存在增强了 ACE2 的结合^[15]。感染人体细胞时, 新型冠状病毒通过表面蛋白 S 与人类细胞受体结合, 打开进入细胞通道而进入细胞, 成为病毒感染细胞最关键的一步。SARS-CoV-2 与 ACE2 的结合主要依靠 S 蛋白中一些重要氨基酸残基完成, 该病毒中谷氨酰胺、天冬酰胺、亮氨酸、苯丙氨酸和丝氨酸氨基酸的存在增强了 ACE2 的结合^[16]。这些氨基酸序列在 SARS-CoV-2 中较为保守, 其他冠状病毒不存在该氨基酸序列。



A S 相互作用蛋白 B N 相互作用蛋白

图 4 SARS-CoV-2 相互作用蛋白

A Protein interactions of S B Protein interactions of N
Fig. 4 Protein interactions of SARS-CoV-2

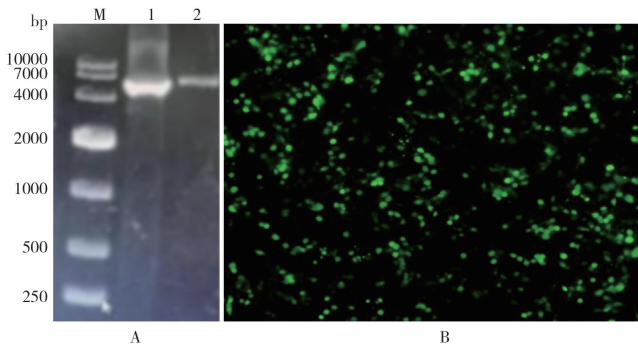


A S 基因 PCR 产物 1% 琼脂糖凝胶电泳分析 M DNA 标志物 (DL5000) 1 S 基因 PCR 产物 B N 基因 PCR 产物 1% 琼脂糖凝胶电泳分析 M DNA 标志物 (DL3000) 1 N 基因 PCR 产物

图 5 目的基因扩增和鉴定

A 1% agarose gel electrophoresis of PCR products of S gene M DNA molecular weight standard (DL5000) 1 S gene PCR product; B 1% agarose gel electrophoresis of PCR products of N gene M DNA molecular weight standard (DL3000) 1 N gene PCR product

Fig. 5 Amplification and identification of target gene



A S-N 融合基因 1% 琼脂糖凝胶电泳分析 M DNA 标志物 (DL10000) 1 S-N 融合基因 PCR 产物 B 重组腺病毒转染的 HEK293T 细胞 (48h, 200×)

图 6 融合基因的构建及重组腺病毒的检测

A 1% agarose gel electrophoresis of S-N gene M DNA molecular weight standard (DL10000) 1 PCR product of S-N fusion gene B HEK293T cells were transfected with recombinant adenovirus (48h, 200×)

Fig. 6 Construction of fusion gene and detection of recombinant adenovirus

N 蛋白由 419 个氨基酸组成, 相对分子质量为 45.71286×10^3 , 原子总数为 6370, 为大分子蛋白, 抗原性较强, 但其为不稳定疏水蛋白, 如与稳定的 S 蛋白融合制备疫苗效果可能会更好。N 蛋白的二级结构中 α -螺旋 (Hh) 占约 19%, β -转角 (Tt) 占约 7%, β -折叠 (Ec) 占约 16%, 无规则卷曲 (Cc) 占 58%。该蛋白结构易与抗体嵌合, 容易形成潜在抗体表位, N 蛋白中央无序区域的磷酸化促进了蛋白质的转录功能, 但其潜在的机制尚不清楚^[17]。分析发现 SARS-CoV-2 的 N 蛋白与病毒 RNA 一起形成生物分子缩合物, 基于多

价 RNA 蛋白质和蛋白质-蛋白质的相互作用, 未经修饰的 N 蛋白形成部分有序的凝胶样凝聚物和离散的 15nm 颗粒。与 S 蛋白相比, N 蛋白是一个在有效抑制病毒后代的增殖, 疾病防治领域很有前途的结构蛋白。有研究证实经过筛选的 N 蛋白抗体可以像 S 蛋白抗体一样激活机体产生针对病毒感染的初次免疫应答, 且在二次免疫应答效果上优于 S 蛋白抗体^[12,17]。测序表明, N 蛋白具备更高突变频率, 使其可以刺激宿主产生比 S 蛋白抗体更强的免疫反应。

本研究采用生物信息学方法对 SARS-CoV-2 S 和 N 蛋白进行了预测分析, 蛋白的理化性质、免疫原性、抗原决定簇及同源性等均符合疫苗的基本要求。构建了融合基因 S-N 并与穿梭表达载体 pDC316-mCMV-EGFP 连接获得重组质粒, 与骨架质粒 pBHGlox (delta)E1, 3Cre 共转染 HEK293T 细胞获得重组腺病毒, 为 SARS-CoV-2 结构蛋白 S 和 N 融合基因疫苗的研究奠定了基础。

【参考文献】

- [1] Gorkhali R, Koirala P, Rijal S, et al. Structure and function of major SARS-CoV-2 and SARS-CoV proteins[J]. Bioinform Biol Insights, 2021, 15(15): 1-32.
- [2] 闻玉梅. 冠状病毒的致病性及防控[J]. 微生物与感染, 2020, 15(1): 2-5.
- [3] Hulswit RJ, de Haan CA, Bosch BJ. Coronavirus spike protein and tropism changes[J]. Adv Virus Res, 2016(96): 29-57.
- [4] Xu X, Chen P, Wang J, et al. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission[J]. Science China (Life Sciences), 2020, 63(3): 457-460.
- [5] 王坤, 张玲, 王超男, 等. 新型冠状病毒 N 蛋白优势表位区段抗原的克隆表达及其诊断价值初步评价[J]. 生物技术通讯, 2020, 31(6): 670-674.
- [6] Gordon DE, Hiatt J, Bouhaddou M, et al. Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms[J]. Science, 2020, 370(6521): 1-25.
- [7] Liu SJ, Leng CH, Lien S, et al. Immunological characterizations of the nucleocapsid protein based SARS vaccine candidates[J]. Vaccine, 2006, 24(16): 3100-3108.
- [8] Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus; implications for virus origins and receptor binding[J]. The Lancet, 2020, 395(10224): 565-574.
- [9] Zuniga S, Cruz Jazmina LG, Sola I, et al. Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription[J]. J Virol, 2010, 84(4): 2169-2175.
- [10] Cui L, Wang H, Ji Y, et al. The nucleocapsid protein of coronaviruses acts as a viral suppressor of RNA silencing in mammalian cells[J]. J Virol, 2015, 89(17): 9029-9043.
- [11] Surjit M, Liu B, Chow Vincent TK, et al. The nucleocapsid protein of severe acute respiratory syndrome-coronavirus inhibits the activity of cyclin-cyclin-dependent kinase complex

- and blocks S phase progression in mammalian cells[J]. J Biol Chem, 2006, 281(16):10669-10681.
- [12] Xiang F, Wang X, He X, et al. Antibody detection and dynamic characteristics in patients with COVID-19[J]. Clin Infect Dis, 2020, 71(8):1930-1934.
- [13] Ni L, Ye F, Cheng ML, et al. Detection of SARS-CoV-2-Specific humoral and cellular immunity in COVID-19 convalescent individuals[J]. Immunity, 2020, 52(6):971-977.
- [14] Ahmed SF, Quadeer AA, McKay MR. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies[J]. Viruses, 2020, 12(3):254.
- [15] Sairaj S, Madhavan N. Structural proteins in severe acute respiratory syndrome Coronavirus-2[J]. Arch Med Res, 2020, 51(6):482-491.
- [16] Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor[J]. Cell, 2020, 181(2):271-280.
- [17] Carlson CR, Asfaha JB, Ghent CM, et al. Phosphoregulation of phase separation by the SARS-CoV-2 N protein suggests a biophysical basis for its dual functions[J]. Molecular Cell, 2020, 80(6):1092-1103.

【收稿日期】 2022-11-13 【修回日期】 2023-02-01

(上接 376 页)

- [4] Lu S, Wang S, Grimes-Serrano JM, et al. Current progress of DNA vaccine studies in humans[J]. Expert Rev Vaccines, 2008(7):175-191.
- [5] 黎强, 丁伟, 刘高文, 等. 狂犬病毒 M 蛋白多克隆抗体的制备与鉴定[J]. 昆明理工大学学报(自然科学版), 2022, 47(4):90-96, 117.
- [6] Chavda VP, Pandya R, Apostolopoulos V. DNA vaccines for sars-cov-2: Toward third-generation vaccination era[J]. Expert Rev Vaccines, 2021, 20:1549-1560.
- [7] Yu J, Tostanoski LH, Peter L, et al. DNA vaccine protection against sars-cov-2 in rhesus macaques[J]. Science, 2020, 369:806-811.
- [8] Schirmbeck R, Bhm W, Ando K, et al. Nucleic acid vaccination primes hepatitis B virus surface antigen-specific cytotoxic T lymphocytes in nonresponder mice[J]. J Virol, 1995, 69:5929-5934.
- [9] Fuller JT, Macklin M, Drape B, et al. Gene-gun-mediated DNA immunization with hbsag: Efficacy in small and large animals[J]. Vaccines, 1997, 157-162.
- [10] Hu Z, Ban H, Zheng H, et al. Protein phosphatase 1 catalyzes hbv core protein dephosphorylation and is co-packaged with viral pregenomic rna into nucleocapsids[J]. PLoS Pathog, 2020, 16:e1008669.
- [11] Pujol FH, Navas MC, Hainaut P, et al. Worldwide genetic diversity of hbv genotypes and risk of hepatocellular carcinoma[J]. Cancer Lett, 2009, 286:80-88.
- [12] Shouval D, Roggendorf H, Roggendorf M. Enhanced immune response to hepatitis b vaccination through immunization with a pre-s1/pre-s2/s vaccine[J]. Med Microbiol Immunol, 2015, 204:57-68.
- [13] Barrera A, Guerra B, Notvall L, et al. Mapping of the hepatitis b virus pre-s1 domain involved in receptor recognition[J]. J Virol, 2005, 79:9786-9798.
- [14] 王炜晓, 赵丹莹, 常军亮, 等. 乙型肝炎病毒截短 L 蛋白的表达及其与 S 蛋白生物信息学对比分析[J]. 中国病原生物学杂志, 2022, 17(4):373-379, 387.
- [15] Progress AA. challenges and opportunities in fish vaccine development[J]. Fish Shellfish Immunol, 2019, 90:210-214.
- [16] Ruiz S, Tafalla C, Cuesta A, et al. In vitro search for alternative promoters to the human immediate early cytomegalovirus (ie-cmv) to express the g gene of viral haemorrhagic septicemia virus (vhsv) in fish epithelial cells[J]. Vaccine, 2008, 26:6620-6629.
- [17] Roos AK, Moreno S, Leder C, et al. Enhancement of cellular immune response to a prostate cancer DNA vaccine by intradermal electroporation[J]. Mol Ther, 2006, 13:320-327.
- [18] 涂亦娴, 张馨玉, 王宾, 等. 猪瘟病毒 E2 基因真核表达载体表达效率和免疫效果的比较[J]. 中国农业大学学报, 2005, 10(6):37-41.

【收稿日期】 2022-09-21 【修回日期】 2022-12-01