

DOI:10.13350/j.cjpb.220701

• 论著 •

新冠病毒 N 基因编码蛋白的生物信息学分析^{*}

孔令启¹,丁一²,屈艳琳^{3**}

(1. 济宁市第一人民医院小儿外科,山东济宁 272000;2. 济宁医学院基础医学院;3. 济宁医学院形态学实验室)

【摘要】 目的 利用生物信息学方法分析预测新冠病毒 N 基因编码蛋白的结构及抗原表位,为研究新冠病毒靶向药物和疫苗提供理论依据。 方法 通过 NCBI 获得新冠病毒核衣壳蛋白的核苷酸序列和氨基酸序列,采用 ORF Finder、ProtParam、SOPMA、DTU Health Tech、TMHMM-2.0、Conserved domains、Cell-PLoc 2.0、NetPhos 3.1 Servera、Net-NGlyc、SWISS MODE、Immunomedicine Group、IEDB、SYFPEITHI、BLAST、Uniprot 等生物信息学软件对新冠病毒 N 蛋白的开放阅读框、理化性质、亲疏水性、信号肽、跨膜区、结构域、亚细胞定位、磷酸化及糖基化位点、二级结构、三级结构、抗原决定簇、B 细胞与 T 细胞抗原表位、蛋白同源性及相互作用蛋白进行预测分析。 结果 新冠病毒核衣壳蛋白是由 419 个氨基酸组成的亲水蛋白,分子式为 $C_{1970} H_{3137} N_{607} O_{629} S_7$,相对分子质量为 45.625×10^3 ,理论等电点为 10.07,脂肪族氨基酸指数为 52.93,平均亲水系数为 -0.971;该蛋白二级结构以无规则卷曲居多占 55.13%,无信号肽和跨膜区,存在 57 个磷酸化位点,5 个糖基化位点。预测该蛋白有 16 个抗原决定簇,8 个 B 细胞优势抗原表位,16 个 CTL 细胞优势表位,13 个 Th 细胞表位,与 SARS-CoV 同源性最高,主要与 6 种蛋白相互作用发挥功能。 结论 新冠病毒 N 基因编码蛋白含有多个 B、T 细胞优势抗原表位,与 SARS-CoV 同源性高,可为冠状病毒疫苗的研发提供和理论基础。

【关键词】 新冠病毒;核衣壳蛋白;生物信息学;抗原表位

【中图分类号】 R373.1

【文献标识码】 A

【文章编号】 1673-5234(2022)07-0745-06

[*Journal of Pathogen Biology*. 2022 Jul.;17(7):745-750,756.]

Bioinformatics analysis of protein encoded by SARS-CoV-2 N gene

KONG Ling-qi¹, DING Yi², QU Yan-lin³ (1. Department of Pediatric Surgery, Jining No. 1 People's Hospital, Jining 272000, Shandong, China; 2. Basic Medical College of Jining Medical College; 3. Morphology laboratory, Jining Medical College)

【Abstract】 **Objective** The structure and antigenic epitopes of the protein encoded by novel coronavirus N gene were analyzed and predicted by bioinformatics methods, providing theoretical basis for the study of novel Coronavirus targeted drugs and vaccines. **Methods** The nucleotide and amino acid sequences of novel Coronavirus nucleocapsid protein were obtained by NCBI. ORF Finder, ProtParam, SOPMA, DTU Health Tech, TMHM-2.0, Cell-PLoc 2.0, NetPhos 3.1 Servera, NetNGlyc, SWISS MODE, Immunomedicine Bioinformatics software such as Group, IEDB, SYFPEITHI, BLAST and Uniprot were used to identify phosphorylation and glycosylation sites in subcells of the transmembrane domain of novel coronavirusN open reading frame. The homology and interaction of epitope proteins of B cells and T cells were predicted.

Results Novel Coronavirus nucleocapsid protein is a hydrophilic protein composed of 419 amino acids. The molecular formula is $C_{1970} H_{3137} N_{607} O_{629} S_7$, the relative molecular weight is 45.625×10^3 , the theoretical isoelectric point is 10.07, the aliphatic amino acid index is 52.93, and the average hydrophilic coefficient is -0.971. There were 57 phosphorylation sites and 5 glycosylation sites in the secondary structure of the protein, and there was no signal peptide or transmembrane region, accounting for 55.13% of the total. It was predicted that the protein had 16 epitopes, 8 dominant epitopes of B cells, 16 dominant epitopes of CTL cells and 13 Th cell epitopes. The protein had the highest homology with SARS-CoV, and mainly interacted with 6 proteins to function. **Conclusion** Bioinformatics analysis showed that nucleocapsid protein is a hydrophilic protein with good thermal stability, and contains multiple phosphorylation sites and glycosylation sites to mediate related signaling pathways, which plays an important role in the development of the disease. The Novel coronavirusN gene encodes a protein contains multiple B and T cell dominant antigen epitopes and has high homology with SARS-CoV. It can provide theoretical basis for the research and development of coronavirus vaccine.

【Key words】 SARS-CoV-2;nucleocapsid phosphoprotein;bioinformatics;antigen epitopes ***

* 【基金项目】 国家自然科学基金项目(No. 82000979)。

** 【通讯作者】 屈艳琳,E-mail:2753167672@qq.com

【作者简介】 孔令启(1988-),男,山东济宁人,硕士,医师,主要从事微生物与肿瘤致病机制的相关研究。E-mail:konglingqi37@126.com

新型冠状病毒肺炎(COVID-19)已在全球流行,截止2022年2月15日超过4亿人感染,其传播速度快,感染范围广,对世界公共卫生安全带来严峻挑战^[1]。新冠病毒(SARS-CoV-2)主要感染人体呼吸、消化、中枢神经等系统,通过各种病毒蛋白逃逸免疫系统的识别,最终导致患者多器官衰竭^[2]。目前临床主要运用抗药物、痊愈患者的血浆或血清进行治疗,但是存在药物靶向性较差、患者再次感染的风险,因此研制针对SARS-CoV-2特异性抗原的疫苗是控制疫情及降低重症率的有效手段^[3]。

新型冠状病毒是一种单正链RNA病毒,其基因组编码4种结构蛋白:刺突蛋白(S)、核衣壳蛋白(N)、包膜蛋白(E)、膜蛋白(M),9种辅助蛋白(ORF)及16种非结构蛋白(nsp1-16)。SARS-CoV-2进入人体后,通过表面的S蛋白与ACE2受体结合来感染细胞^[4],S蛋白诱导产生的抗体可与新冠抗原接触更直接,目前大多数疫苗是基于S蛋白研发的^[5],但S蛋白多位点易突变,削弱了疫苗的免疫效果^[6]。N蛋白是一种高度保守的结构蛋白,在病毒的复制、释放、包装等过程中起关键作用,还调节宿主细胞周期,增强病毒感染力^[7],依据N蛋白设计疫苗是一种新的研究方向。本研究拟对新冠病毒N蛋白进行抗原性分析,探究其作为候选疫苗的可行性。

材料与方法

1 N蛋白氨基酸序列及其编码基因序列的获取

登录NCBI数据库(<https://www.ncbi.nlm.nih.gov/>),搜索并下载SARS-CoV-2N蛋白的氨基酸序列和核苷酸序列。N基因登录号为43740575,编码蛋白登录号为YP_009724397。通过NCBI查询N蛋白由419个氨基酸组成,其序列如下:

```
MSDNGPQNQRNAPRITFGGSDSTGSNQNG
ERSGARSKQRRPQGLPNNTASWFTALTQHGKE
DLKFPRGQGVPIINTSSPDDQIGYYRRATRRIR
GGDGKMKDLSRKWYFYIYLGTPPEAGLPYGAN
KDGIIWVATEGALNTPKDHGTRNPANNAIVL
QLPQGTTLPKGFYAEGSRGGSQASSRSSRSRNS
SRNSTPGSSRGTSPARMAGNGDAALALLLLDR
LNQLESKMSGKGQQQQGQTVTKSAAEASKK
PRQKRTATKAYNVTQAFGRRGPEQTQGNFGD
QELIRQGTDYKHWPQIAQFAPSASAFFGMSRIG
MEVTPSGTWLYTGAIKLDDKDPNFKDQVILL
NKHIDAYKTFPPTEPKDKKKKADETQALPQ
RQKKQQTVTLLPAADLDDFSKQLQQSMSSAD
STQA。
```

2 方法

2.1 预测N基因开放阅读框的数目 在ORF Finder(<http://www.ncbi.nlm.nih.gov/orffinder>)网址中输入N基因核苷酸序列来预测开放阅读框的数目。

2.2 预测N蛋白的理化性质及亲疏水性 运用protparam(<https://www.expasy.org/prot-param>)预测N蛋白的分子量、理论等电点、氨基酸组成、消光系数、半衰期等理化性质以及蛋白的亲疏水性。

2.3 信号肽、跨膜结构及结构域的预测 采用Conserved domains (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>)预测蛋白结构域,运用<http://services.healthtech.dtu.dk/service.php?SignalP-5.0>和<http://services.healthtech.dtu.dk/service.php?TMHMM-2.0>分析N蛋白的信号肽和跨膜结构。

2.4 预测N蛋白的空间结构 运用SOPMA软件(http://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl)预测蛋白的二级结构,SWISS MODEL(<https://swissmodel.expasy.org/>)预测三级结构。

2.5 预测N蛋白磷酸化和糖基化位点及亚细胞定位

运用NetPhos 3.1 Server(<http://www.cbs.dtu.dk/services/NetPhos/>)预测磷酸化位点,NetNGlyc (<http://www.cbs.dtu.dk/services/NetNGlyc/>)预测糖基化位点,Cell-Ploc 2.0(<http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2/>)分析N蛋白在宿主细胞的定位。

2.6 N抗原蛋白的优势B细胞表位预测 通过IEedb在线网站(https://www.iedb.org/home_v3.php)筛选线性优势B细胞表位。选取柔韧性大于平均值的表位,并结合表面可及性>1的表位进一步筛选,综合线性、β转角、抗原性、亲水性均大于平均值的表位筛选出优势表位。

2.7 N抗原蛋白的T细胞表位预测 综合IEedb数据库(https://www.iedb.org/home_v3.php)和SYFPEITHI(www.syfpeithi.de/)对N蛋白的优势T细胞表位进行预测。预测CTL细胞表位时,选取IEedb得分>0.5,SYFPEITHI得分>20的表位。预测Th细胞表位选取IEedb得分前20%,SYFPEITHI>20的表位作为优势表位。

2.8 N蛋白的同源性分析及抗原决定簇预测 利用BLAST(<https://blast.ncbi.nlm.nih.gov/>)进行同源蛋白的分析,首先选择ProteinBLAST,输入氨基酸序列FASTA格式,在Organism一栏中输入ssRNA-Virus进行预测运用Immunomedicine Group(<http://imed.ucm.es/Tools/antigenic.pl>)在线网站对N蛋白的抗原决定簇进行分析。

2.9 预测N的相互作用蛋白 运用Uniprot生信软

件对 N 蛋白的相互作用蛋白进行预测,搜索栏中输入 SARS-CoV-2 N,选择 UniProtKB 数据库进行查询预测。

结 果

1 N 基因开放阅读框的数目

N 基因位于全基因组(NC_045512.2)的 28 274-29 533 位,总长度为 1 260 bp,N 基因总共有 6 个开放阅读框(ORF),其中 ORF1 为最长的开放阅读框,起始密码子为 ATG,终止密码子为 TAA,编码 419 个氨基酸,与目的蛋白全长一致,表明 N 蛋白全长可完全表达(图 1)。

2 N 蛋白的理化性质及亲疏水性

运用 protparam (<https://www.expasy.org/prot-param>)预测新冠病毒 N 蛋白由 419 个氨基酸组成,包含 19 种氨基酸,其中以甘氨酸(10.3%)、丙氨酸(8.8%)、丝氨酸(8.8%)谷氨酰胺(8.4%)所占比例较高,无半胱氨酸、吡咯赖氨酸和硒代半胱氨酸,带负电荷残基数目(天冬氨酸+谷氨酸)为 36,带正电荷残基数目(精氨酸+赖氨酸)为 60,分子式为 C₁₉₇₀H₃₁₃₇N₆₀₇O₆₂₉S₇,相对分子质量为 45.625×10³,理论等电点为 10.07。当蛋白的 N 端为 Met 时,该蛋白在哺乳动物网织红细胞中的半衰期为 30 h,在酵母菌中的半衰期>20 h,在大肠埃希菌中的半衰期>10 h。N 蛋白消光系数为 43890,280nm 处的吸光度为 0.962。不稳定系数为 55.09,为较稳定蛋白,脂肪族氨基酸系数为 52.93,热稳定性较好,平均亲水系数为-0.971,为亲水蛋白。

3 信号肽、跨膜结构及结构域的预测

Conserved domains 预测结果显示 N 蛋白在 50-175 氨基酸序列存在 N 端结构域(NTD),在 258-359 位存在 C 端结构域(CTD)(图 2),N 端和 C 端结构域对于病毒基因的结合都是必需的,C 端结构域通过形成低聚物,驱动核衣壳的形成和病毒的组装。通过在线网站 Signal-5.0 和 TMHMM-2.0 预测 N 蛋白无信号肽(图 3)和跨膜结构(图 4)。



图 1 N 基因开放阅读框分析

Fig. 1 Open reading frame analysis of N gene

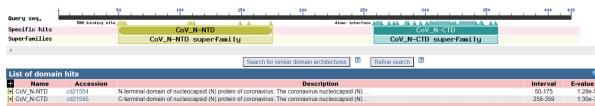


图 2 N 蛋白结构域预测

Fig. 2 Prediction of N protein domain

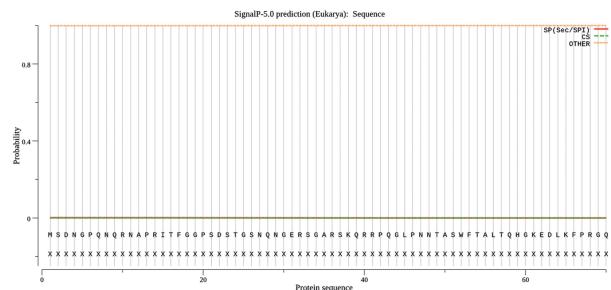


图 3 N 蛋白信号肽

Fig. 3 Prediction of N protein signal peptide

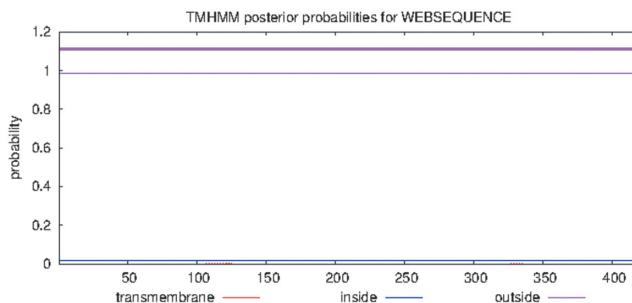


图 4 跨膜结构预测

Fig. 4 transmembrane structure

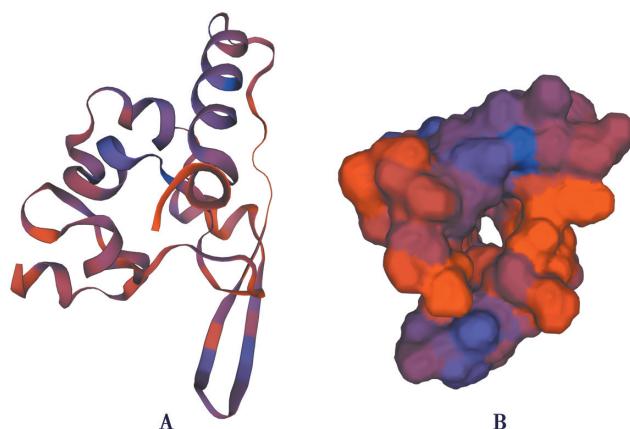
4 N 蛋白的空间结构

通过 SOMPA 对蛋白的二级结构进行预测,无规则卷曲占 55.13%, α 螺旋占 21.24%,延伸片段占 16.71%, β 转角占 6.92%(图 5),无规则卷曲占比高,说明该蛋白有较好的抗体嵌合性。SWISS MODEL 预测三级结构,条带模型和表面模型见图 6。



图 5 N 蛋白二级结构

Fig. 5 The Secondary structure of N protein



A 飘带模型 B 表面模型

A Ribbon model B Surface model
Fig. 6 Tertiary structure of N protein

5 N蛋白磷酸化和糖基化位点及亚细胞定位

运用NetPhos 3.1Server预测,阈值设定为0.5时N蛋白共有57个潜在磷酸化位点(图7),其中丝氨酸位点有31个,苏氨酸位点有22个,酪氨酸位点有4个,这些位点可能介导体内相关信号通路中关键分子的去磷酸化过程。

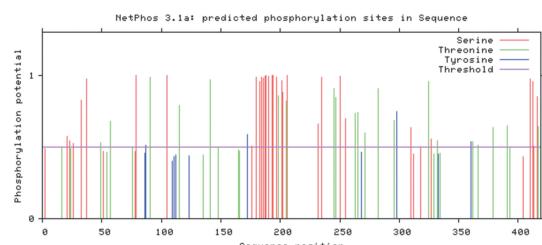


图7 N蛋白磷酸化位点
Fig. 7 Analysis of potential sites of N protein phosphorylation

NetNGlyc预测N蛋白糖基化位点有5个,分别在47、77、192、196、269位点,其中269(82.13%)、47(67.98%)、192(45.04%)位点糖基化概率较高(图8)。糖基化位点易形成体积较大的聚糖分子,不利于抗体的结合,使病毒逃避免疫系统的识别^[8],因此设计疫苗时应尽量避开糖基化潜在序列,防止免疫原性的降低。通过Cell-Ploc 2.0分析N蛋白的亚细胞定位主要位于宿主细胞的内质网发挥作用(图9)。

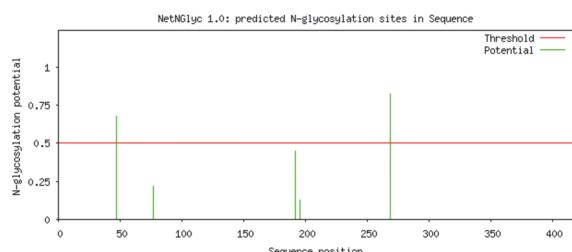


图8 糖基化潜在位点分析
Fig. 8 glycosylation

Virus-mPLoc Computation Result

Query protein	Predicted location(s)
query protein1	Host endoplasmic reticulum.

图9 N蛋白亚细胞定位分析
Fig. 9 Subcellular localization analysis of N protein

6 N抗原蛋白的B细胞表位预测

通过IEDB在线网站(https://www.iedb.org/home_v3.php)预测N蛋白的优势B细胞表位。蛋白的可塑性与抗体的结合度相关,可塑性得分越高的位点与抗体的嵌合度越高,因此选择可塑性大于平均值的序列,再结合线性表位、抗原性、β转角、亲水性三个性质进行筛选,均选取评分高于平均值的表位,最终得出B细胞的优势表位有8个(表1)。

表1 N蛋白B细胞优势表位预测
Table 1 Prediction of dominant B cell epitopes of N protein

序号 NO.	起始氨基酸位置 Epitope Starting Position	氨基酸序列 Epitope Sequence	终止氨基酸位置 Epitope Ending Position
1	39	QRRPQGL	45
2	181	QASSRSS	187
3	192	NSSRNST	198
4	238	GQQQQGQ	244
5	363	FPPTEPK	369
6	372	KKKKADE	378
7	396	PAADLDD	402
8	404	SKQLQQS	410

7 N抗原蛋白的T细胞表位预测

运用IEDB筛选出N蛋白的综合优势T细胞表位(图10)。利用IEDB和SYFPEITHI综合预测CTL和Th的优势表位。其中CTL细胞优势表位有9个(表2),通过查阅文献主要对DRB1*01:01、DRB1*03:01、DRB1*07:01等人类HLAⅡ等位基因进行Th细胞优势表位的预测主要^[9],综合分析结果有13个优势表位(表3),两个表格综合分析N蛋白T细胞的优势表位在216-225、266-274、299-309、325-334、352-361位点。

8 N蛋白的抗原决定簇表位预测及同源蛋白分析

Immunomedicine Group(imed.med.ucm.es/Tools/antigenic.pl)在线网站分析N蛋白的平均抗原倾向性为0.9871,共有16个抗原决定簇表位,主要位于氨基酸序列的105-114、216-226、266-272、332-338、346-362、402-410等位点(图11),结果显示该蛋白抗原倾向指数较高,易形成抗原决定簇,且相关位点也为B或T细胞优势表位,表明N蛋白免疫原性较强。运用BLAST(<https://blast.ncbi.nlm.nih.gov/>)对目的蛋白的氨基酸序列进行分析,结果显示有4个同源性较高的蛋白(图12),其中SARS-CoV和Bat SARS-

CoV Rp 3/2004 病毒的 N 蛋白与 SARS-CoV-2 N 蛋白的同源性高, 分别为 90.52% 和 90.26%, 基于 N 蛋白设计的疫苗还可预防多种冠状病毒引发的疾病。

表 2 N 蛋白 CTL 细胞优势表位分析
Table 2 Analysis of dominant epitopes of N protein CTL

序号 NO.	MHC I 类 分子亚型 I molecular subtype	起始 氨基酸位置 Epitope Starting Position	氨基酸 序列 Epitope Sequence	SYFPEITHI 得分 SYFPEITHI score (≥20)	Iedb 得分 Iedb score (0.500-1)
1	HLA-A	53	FTALTQHKG	21	0.583
2		249	KSAAEASKK	23	0.675
3		311	ASAFFGMSR	23	0.795
4		290	ELIRQGTDY	25	0.520
5	HLA-B1	159	LQLPQGTTL	21	0.794
6		238	GQQQQGQTV	23	0.852
7		352	LLNKHIDAY	22	0.871
8	HLA-B3	45	LPNNTASWF	20	0.738
9		299	KHWPQIAQF	20	0.883
10		355	KHIDAYKTF	20	0.843
11		395	LPAADLDDF	22	0.639
12	HLA-B5	79	SPDDQIGYY	23	0.590
13		216	DAALALLLL	22	0.676
14		266	KAYNVTQAF	25	0.926
15		325	TPSGTWLTY	22	0.833
16		403	FSKQLQQSM	22	0.501

表 3 N 蛋白 Th 细胞优势表位分析
Table 3 Analysis of dominant epitopes of N protein in Th cells

序号 NO.	MHC II 类 分子亚型 II molecular subtype	起始 氨基酸位置 Epitope Starting Position	氨基酸 序列 Epitope Sequence	SYFPEITHI 得分 SYFPEITHI score (≥20)	Iedb 百分位得分 percentage (0-20)
1	HLA-DRB1 * 01:01	106	PRWYFYYLGTGPEAG	29	5.9
2		154	NAAIVLQLPQGTTL	24	8
3		357	IDAYKTFPPTEPKKD	26	17
4	HLA-DRB1 * 03:01	219	AALALLLDRLNQLE	32	4.1
5		334	TGAIKLDDKDPNFKD	28	3.4
6	HLA-DRB1 * 04:01	50	ASWFTALTQHGKEDL	28	17
7		288	DQELRQGTDYKHWP	20	12
8		317	MSRIGMEVTPSGTWL	20	11
9	HLA-DRB1 * 07:01	257	KPRQKRATKAYNVT	22	13
10		304	IAQFAPSASAFFGMS	30	7.1
11		326	PSGTWLTYTGAIKLD	24	0.58
12	HLA-DRB1 * 11:01	84	IGYYRATRRIRGGD	20	0.42
13	HLA-DRB1 * 15:01	221	LLLDRLNQLESKMS	20	8.8

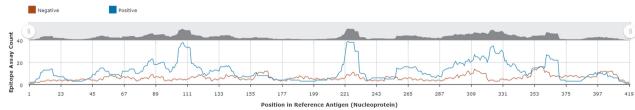


图 10 N 蛋白优势 T 细胞表位分析
Fig. 10 Analysis of dominant T cell epitopes of N protein

9 预测 N 蛋白的相互作用蛋白

通过 Uniprot 软件预测显示 N 蛋白主要与新冠病毒 M 蛋白、人类 G3BP1/2、STAT1/2 蛋白相互作

用,且蛋白自身可二聚化(图 13)。N 蛋白的 CTD 结构域可发生自身二聚化^[10],与 M 蛋白相互作用来识别病毒 RNA 的顺式作用元件,从而促进病毒的复制及包装。人类 G3BPs 蛋白是 RIG-I 信号通路的关键分子,与 RIG-I 协同诱导重要的抗病毒细胞因子 IFN-β 的产生^[11]。STAT1 是 IFN 信号通路下游的关键转录因子,经过激酶 JAK1 的催化激活,与 STAT2 和 ISGF3/IRF9 结合形成三聚体,进入细胞核后 ISGF3 与 IFN 刺激反应元件(ISRE)结合以激活 IFN 刺激基因(ISG)的转录,该基因驱动细胞处于抗病毒状态^[12]。因此,推测 N 蛋白通过与上述蛋白的相互作用来抑制宿主细胞的抗病毒反应。

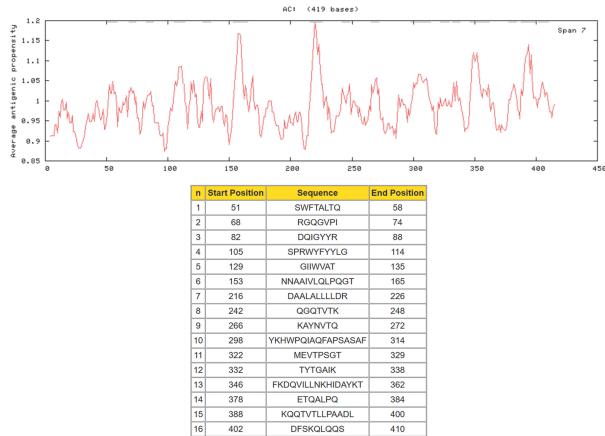


图 11 预测 N 蛋白抗原决定簇
Fig. 11 Prediction of N proteinepitopes

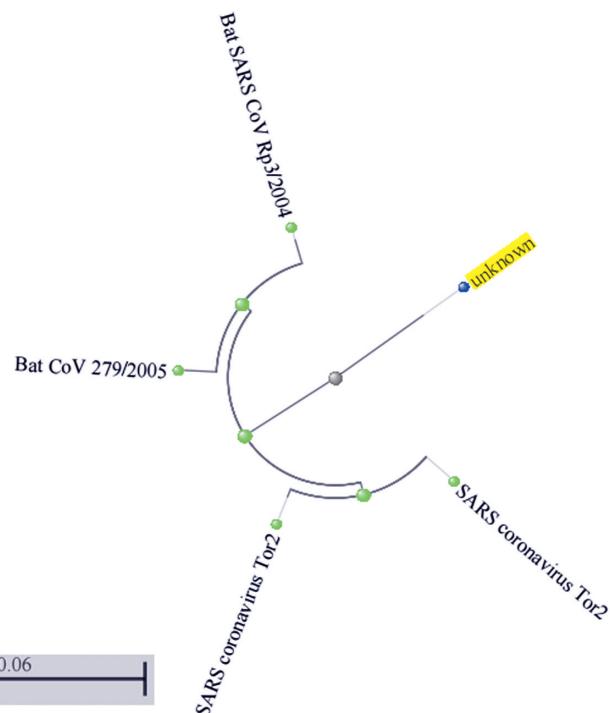


图 12 氨基酸序列预测 N 蛋白同源性
Fig. 12 Amino acid sequence predicts homology of N protein

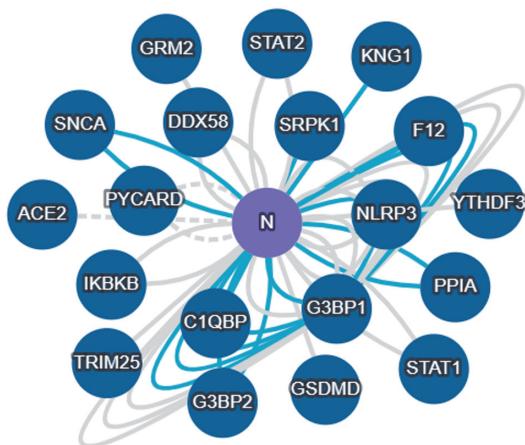


图 13 N 蛋白相互作用蛋白预测
Fig. 13 interactive proteins of N protein

讨 论

随着生物信息学技术不断发展,基因组与蛋白质数据库的持续扩充,免疫学预测工具的不断更新^[13],利用基因组学、蛋白组学、免疫组学等相关知识对未知蛋白的结构与功能进行全面的预测分析,节省了实验筛选时间,有效缩小了实验验证对象的范围,提供了丰富的理论知识以探究疾病进展的机制、开发新型靶向药物、研制免疫原性强的疫苗等,为临床治疗提供了更多元、精准的治疗方法^[14]。

新冠病毒 N 蛋白形成核衣壳与 RNA 结合,通过定位于宿主细胞内质网来介导相关病毒蛋白的表达^[15]。N 蛋白总共有三个保守域,即 N 端(NTD)、中心区域(IDR)、C 端(CTD),其中 NTD 和 CTD 是重要的功能结构域^[16],NTD 类右手拳头形状,表面主要为带正电的氨基酸残基,介导与 RNA 结合以促进病毒基因的复制与转录,IDR 富含丝氨酸和精氨酸,易发生磷酸化,促进 RNA 与蛋白的分离,CTD 通过自身二聚化与 M 蛋白相互作用,来调节宿主细胞周期,促进病毒的组装。NTD 和 CTD 保守程度高适合作为疫苗抗原与抗病毒药物的靶标^[17]。

N 蛋白除了在病毒的复制过程发挥重要作用外,还干扰先天免疫反应的相关信号通路,利于病毒逃逸免疫系统识别。干扰素(IFN)作为抵御病毒的第一道防线,其产生需要通路上游干扰素刺激基因(ISG)的激活,以及视黄酸诱导蛋白 I(RIG-I)与 G3BP 蛋白的协同作用^[18]。新冠病毒 N 蛋白主要通过抑制 ISG 上游 STAT1/2 的磷酸化来阻止 ISG 激活,干扰泛素酶 TRIM25 泛素化 RIG-I,招募 G3BP 蛋白来阻止 IFN 的生成^[19]。相关实验研究证明 N 蛋白对干扰素的产生有双重作用,即低剂量($0.25\mu\text{g}$)抑制干扰素生成,高剂量($1\mu\text{g}$)促进干扰素过度表达并引起相关炎症反应^[20]。因此,设计疫苗时应注意注射剂量,防止

发生免疫反应不足或过度的问题。

本研究获取新冠病毒核衣壳蛋白基因序列及氨基酸序列,利用生物信息学分析软件对蛋白的结构功能及免疫原性等进行预测。N 蛋白为亲水性蛋白,主要定位于宿主细胞的内质网,利于病毒蛋白的修饰加工,其二级结构中无规则卷曲占比高为 55.13%,有良好的抗体嵌合性,易形成潜在抗原表位,由于含有较多的脂肪族氨基酸,说明其热稳定性较高,利于疫苗的储存和运输。该蛋白在 50-175 氨基酸序列存在 N 端结构域(NTD),258-359 位存在 C 端结构域(CTD),主要参与病毒颗粒复制过程,不存在信号肽及跨膜区。预测目的蛋白潜在磷酸化位点有 57 个,位点多集中在 IDR 区域,提示 N 蛋白可能通过磷酸化和去磷酸化影响信号通路中关键分子的活性,糖基化位点预测有 5 个,可形成特异糖链结构参与分子识别、信号转导等过程,说明 N 蛋白可作为诊断新冠病毒感染的特异标志物。病毒感染期间,B 细胞介导的体液免疫不可或缺,其通过分泌多种细胞因子阻碍病毒感染^[21],因此采用 IEDB 在线网站分析 N 蛋白的 B 细胞表位,筛选可塑性好、抗原性强、β 转角密集的表位作为优势表位。先前对 SARS-CoV 疫苗效果的研究发现,细胞免疫的持续时间可达 6-11 年,而抗体数量随着时间的推移而下降^[22],通过 IEDB 和 SYFPEITHI 综合分析筛选出 N 蛋白的 CTL 细胞优势表位有 16 个,Th 细胞优势表位有 13 个,综合分析 216-225、266-274、299-309、325-334、352-361 表位可引起较强的 T 细胞反应。目的蛋白与 SARS-CoV 同源性高,因此可以借鉴先前相关冠状病毒的研究来开发新型治疗性或预防性疫苗。

新冠病毒肺炎大流行的背景下,推动了疫苗平台的快速发展,如重组腺病毒疫苗、核酸疫苗、病毒颗粒疫苗等^[23],不同平台各有优缺点,应就疫苗的安全性、有效性等要求谨慎选择。目前新冠病毒 S 蛋白疫苗已出现抗体依赖性增强(ADE)的现象^[24],促进了病毒感染,因此设计疫苗时应着重关注 N 蛋白的 T 细胞优势表位,以获得更安全、更高效的免疫力。结合本研究对 N 蛋白的生物信息预测结果,该蛋白高度保守,在病毒感染机制中发挥重要作用,且有多个 B、T 细胞优势表位,免疫原性良好,本研究希望为抗新冠病毒疫苗的研发提供理论基础,以减少病毒的传播与感染,防止疫情进一步蔓延。

【参考文献】

- [1] Majumder J, Minko T. Recent developments on therapeutic and diagnostic approaches for COVID-19[J]. Aaps J, 2021, 23(1):14.
- [2] Sharma A, Ahmad Farouk I, Lal SK. COVID-19: a review on the novel coronavirus disease evolution, transmission, detection, control and prevention[J]. Viruses, 2021, 13(2):202.

(下转 756 页)

- [9] K Ito, K Akai, Nishiumi F, et al. Ability of *Ureaplasma parvum* to invade mouse sperm, fertilize eggs through infected sperm, and impair mouse sperm function and embryo development[J]. *F&S Science*, 2020, 10(3):815-819.
- [10] 宾东华, 李玲, 王磊, 等. 基于 AMPK/PPAR α 信号通路研究知柏地黄汤对解脲脲原体感染大鼠生精细胞能量代谢的影响[J]. 中华男科学杂志, 2021, 27(5):437-444.
- [11] Sweeney EL, Kallapur SG, Meawad S, et al. *Ureaplasma* species multiple banded antigen(MBA) variation is associated with the severity of inflammation In vivo and In vitro in human placentae [J]. *Front Cell Infect Microbiol*, 2017, 13(7):123-127.
- [12] 张荣荣, 贾泽玮, 李雪燕, 等. 微小脲原体 UP3-00235 基因原核载体的构建表达和生物信息学分析[J]. 中国病原生物学杂志, 2020, 15(2):183-188.
- [13] Biernat-Sudolska M, Rojek-Zakrzewska D, Bilska-Wilkoz A. *In vitro* activity of lipoic acid against *Ureaplasma urealyticum* and *Ureaplasma parvum* isolated from women with infections of the urogenital tract. A pilot study[J]. *Acta Biochim Pol*, 2020, 67(4):623-628.
- [14] Siles-Guerrero V, Cardona-Benavides I, Liebana-Martos C, et al.
- ~~~~~
- (上接 750 页)
- [3] Iqbal Yatoo M, Hamid Z, Paray OR, et al. COVID-19 - recent advancements in identifying novel vaccine candidates and current status of upcoming SARS-CoV-2 vaccines[J]. *Hum Vaccin Immunother*, 2020, 16(12):2891-2904.
- [4] Harrison AG, Lin T, Wang P. Mechanisms of SARS-CoV-2 transmission and pathogenesis[J]. *Trends Immunol*, 2020, 41(12):1100-1115.
- [5] Awadasseid A, Wu Y, Tanaka Y, et al. Current advances in the development of SARS-CoV-2 vaccines[J]. *Int J Biol Sci*, 2021, 17(1):8-19.
- [6] Harvey WT, Carabelli AM, Jackson B, et al. SARS-CoV-2 variants, spike mutations and immune escape[J]. *Nat Rev Microbiol*, 2021, 19(7):409-424.
- [7] Dong Y, Dai T, Wei Y, et al. A systematic review of SARS-CoV-2 vaccine candidates[J]. *Signal Transduct Target Ther*, 2020, 5(1):237.
- [8] Watanabe Y, Berndsen ZT, Raghwanji J, et al. Vulnerabilities in coronavirus glycan shields despite extensive glycosylation[J]. *Nat Commun*, 2020, 11(1):2688.
- [9] Paul S, Lindestam Arlehamn CS, Scriba TJ, et al. Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes[J]. *J Immunol Methods*, 2015(422):28-34.
- [10] Ye Q, West AMV, Silletti S, et al. Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein[J]. *Protein Sci*, 2020, 29(9):1890-1901.
- [11] Kim SS, Sze L, Liu C, et al. The stress granule protein G3BP1 binds viral dsRNA and RIG-I to enhance interferon- β response [J]. *J Biol Chem*, 2019, 294(16):6430-6438.
- [12] Chen K, Liu J, Liu S, et al. Methyltransferase SETD2-Mediated methylation of STAT1 Is critical for interferon antiviral activity [J]. *Cell*, 2017, 170(3):492-506.

Recent clinical relevance of mono-genital colonization/infection by *Ureaplasma parvum* [J]. *Eur J Clin Microbiol Infect Dis*, 2020, 39(10):1899-1905.

- [15] Butland G, Peregrin-Alvarez JM, Li J, et al. Interaction network containing conserved and essential protein complexes in *Escherichia coli*[J]. *Nature*, 2005, 433(7025):531-537.
- [16] Gruet A, Longhi S, Bignon C. One-step generation of error-prone PCR libraries using Gateway technology[J]. *Microb Cell Fact*, 2012, 11(7):14-18.
- [17] Zhan C, Chen L, Hu L. Neonatal *Ureaplasma parvum* meningitis complicated with subdural hematoma:a case report and literature review[J]. *BMC Infect Dis*, 2021, 21(1):268.
- [18] Kim MS, Lee DH, Kim TJ, et al. The role of *Ureaplasma parvum* serovar-3 or serovar-14 infection in female patients with chronic micturition urethral pain and recurrent microscopic hematuria[J]. *Transl Androl Urol*, 2021, 10(1):96-108.
- [19] Shatsky M, Dong M, Liu H, et al. Quantitative tagless copurification: A method to validate and identify protein-protein interactions[J]. *Mol Cell Proteomics*, 2016, 15(6):2186-2202.

【收稿日期】 2022-02-18 【修回日期】 2022-05-10

- [13] Muzio G, O'Bray L, Borgwardt K. Biological network analysis with deep learning[J]. *Brief Bioinform*, 2021, 22(2):1515-1530.
- [14] Arunkumar M, Zielinski CE. T-Cell receptor repertoire analysis with computational tools—an immunologist's perspective [J]. *Cells*, 2021, 10(12):3582.
- [15] Has ks z M, Kili S, Sara F. Coronaviruses and SARS-COV-2[J]. *Turk J Med Sci*, 2020, 50(SI-1):549-556.
- [16] Peng Y, Du N, Lei Y, et al. Structures of the SARS-CoV-2 nucleocapsid and their perspectives for drug design[J]. *Embo J*, 2020, 39(20):e105938.
- [17] Kadam SB, Sukhramani GS, Bishnoi P, et al. SARS-CoV-2, the pandemic coronavirus: Molecular and structural insights[J]. *J Basic Microbiol*, 2021, 61(3):180-202.
- [18] Oh SJ, Shin OS. SARS-CoV-2 Nucleocapsid Protein targets RIG-I-Like receptor pathways to inhibit the induction of interferon response[J]. *Cells*, 2021, 10(3):530.
- [19] Lu S, Ye Q, Singh D, et al. The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein[J]. *Nat Commun*, 2021, 12(1):502.
- [20] Zhao Y, Sui L, Wu P, et al. A dual-role of SARS-CoV-2 nucleocapsid protein in regulating innate immune response[J]. *Signal Transduct Target Ther*, 2021, 6(1):331.
- [21] Lam JH, Smith FL, Baumgarth N. B cell activation and response regulation during viral infections viral immunol[J]. *Front Immunol*, 2020, 33(4):294-306.
- [22] Lee E, Sandgren K, Duette G, et al. Identification of SARS-CoV-2 nucleocapsid and spike T-Cell epitopes for assessing T-cell immunity[J]. *J Virol*, 2021, 95(6):e02002-20.
- [23] Ura T, Yamashita A, Mizuki N, et al. New vaccine production platforms used in developing SARS-CoV-2 vaccine candidates[J]. *Vaccine*, 2021, 39(2):197-201.
- [24] Poland GA, Ovsyannikova IG, Crooke SN, et al. SARS-CoV-2 vaccine development: current status[J]. *Mayo Clin Proc*, 2020, 95(10):2172-2188.

【收稿日期】 2022-03-08 【修回日期】 2022-05-31