

doi:10.3969/j.issn.1000-484X.2022.24.004

新冠病毒核衣壳蛋白结构与功能的生物信息学分析及原核表达^①

徐本锦 范蕾^{②③} 杜森^② 宣焱 李璟^③ 李卓禧^③ 陈晓聪^③ 吴惠文^④ 门杰^③ 刘玲^⑤
杨娅男 汤文婷^③ 寇妍祺^③ 余骏骁^③ (山西医科大学汾阳学院医学检验系,汾阳 032200)

中图分类号 Q812 文献标志码 A 文章编号 1000-484X(2022)24-2963-08

[摘要] 目的:对SARS-CoV-2核衣壳(N)蛋白进行系统生物信息学分析和原核表达,阐明其功能和病毒RNA复制机制。方法:采用UniProt、NetPhos 3.1、IEDB和Blast等生物信息学软件对N蛋白的理化性质、蛋白互作网络、翻译后修饰、同源性 & 系统进化关系等生物学特性进行系统分析。构建原核表达质粒pET-22b-N并表达蛋白。结果:N蛋白是由419个氨基酸组成的碱性蛋白,分子量为45.62 kD,等电点为10.07,其N端有1个RNA结合结构域,C端有1个二聚体界面,有57个可能的磷酸化修饰和25个潜在糖基化修饰,15个T细胞抗原表位和11个线性B细胞抗原表位,无信号肽和跨膜螺旋,是一种亲水性蛋白。N蛋白中无规则卷曲占比最高(54.42%),其次为 α -螺旋(22.20%)。与N蛋白序列一致性最高的是BtRs- β -冠状病毒/HuB2013和蝙蝠冠状病毒Rp/Shaanxi2011,进化分析显示SARS-CoV-2与蝙蝠非典型冠状病毒YNLF_31C聚为一支。原核表达后发现,细菌裂解液离心后的沉淀中有大量N蛋白表达,为后续蛋白纯化和结构分析提供了依据。结论:N蛋白在SARS冠状病毒和蝙蝠冠状病毒间高度保守,为SARS-CoV-2 N蛋白结构与功能研究提供了基础,为靶向N蛋白的抗病毒药物设计提供了依据。

[关键词] 新冠病毒;核衣壳蛋白;生物信息学分析;载体构建;原核表达

Bioinformatics analysis for structure and function of nucleocapsid protein of SARS-CoV-2 and its prokaryotic expression

XU Benjin, FAN Lei, DU Miao, XUAN Yan, LI Jing, LI Zhuoxi, CHEN Xiacong, WU Huiwen, MEN Jie, LIU Ling, YANG Yanan, TANG Wenting, KOU Yanqi, YU Junxiao. Department of Medical Laboratory, Fenyang College of Shanxi Medical University, Fenyang 032200, China

[Abstract] **Objective:** To expound function and replication mechanism of viral RNA of nucleocapsid (N) protein of SARS-CoV-2 by systematic bioinformatics analysis and prokaryotic expression. **Methods:** Bioinformatics software such as UniProt, NetPhos 3.1, IEDB and Blast were used to systematically analyze physical and chemical properties, protein-protein interaction network, post-translational modification and homology and phylogenetic relationship of N protein. Prokaryotic expression plasmid pET-22b-N was constructed and the protein was expressed. **Results:** N protein was a basic protein composed of 419 amino acids, with a molecular weight of 45.62 kD and an isoelectric point of 10.07, which has an RNA binding domain in N-terminal, a dimer interface in C-terminal, 57 possible phosphorylation modification and 25 potential glycosylation modification, 15 T-cell epitopes and 11 linear B-cell epitopes, was a hydrophilic protein without signal peptide and transmembrane helix. Irregular coil accounted for highest proportion of N protein (54.42%), followed by α -helix (22.20%). BtRs- β -CoV/HuB2013 and Bat coronavirus Rp/Shaanxi2011 had highest sequence identity with N protein. Phylogenetic analysis showed that SARS-CoV-2 and bat SARS-like coronavirus YNLF_31C clustered into one branch. After prokaryotic expression, it was found that N protein was mainly expressed in the precipitation after centrifugation of bacterial lysate, which laid a foundation for subsequent protein purification and structural analysis. **Conclusion:** N protein is highly conserved between SARS coronavirus and bat coronavirus, which provides an important basis for study of structure and function of SARS-CoV-2 N protein, and provides a basis for design of antiviral drugs targeting N protein.

[Key words] SARS-CoV-2; Nucleocapsid protein; Bioinformatics analysis; Vector construction; Prokaryotic expression

①本文受山西省高等学校科技创新项目(2020L0749);国家级大学生创新创业训练计划项目(202117114001);山西省高等学校大学生创新创业训练计划重点项目(S202117114008);山西医科大学汾阳学院引进人才启动金项目(2020A01);吕梁市科技计划项目(2020SHFZ29, 2020XGZX104);山西医科大学汾阳学院大学生创新创业训练计划项目(FDC202111)资助。

②共同第一作者。

③山西医科大学汾阳学院基础医学部,汾阳 032200。

④山西医科大学汾阳学院科技中心,汾阳 032200。

⑤通信作者,E-mail:ll772x@sxmu.edu.cn。

作者简介:徐本锦,男,博士,讲师,主要从事基因工程和蛋白质结构与功能方面的研究,E-mail:bj0726@sxmu.edu.cn。

2019年12月,武汉暴发了新冠病毒肺炎疫情,患者出现发热、胸闷、呼吸急促等症状,重症患者因急性呼吸窘迫综合征(acute respiratory distress syndrome, ARDS)、多器官功能衰竭而死亡^[1-2]。该病毒是20世纪以来第3种可感染人类的高致病性病原^[3]。2020年2月11日,国际病毒分类委员会将其命名为“SARS-CoV-2”^[4]。据WHO报道,截至2021年8月23日,全球已有211 730 035例确诊病例。

SARS-CoV-2为单股正链RNA病毒,是目前已知的第七种冠状病毒^[5-6]。SARS-CoV-2含有2个大的开放阅读框(ORF1a和ORF1b),二者在不同冠状病毒中高度保守,编码4种同源结构蛋白:刺突蛋白(S)、包膜蛋白(E)、膜蛋白(M)、核衣壳蛋白(N)及8种辅助蛋白^[1,4,6-9]。

N蛋白是SARS-CoV-2的重要组成,是一种高免疫原性蛋白,在感染过程中大量表达^[10-12]。N蛋白具有较高的保守性和RNA分子伴侣活性,是干扰素拮抗剂和病毒编码的RNA干扰抑制因子^[13-14]。N蛋白与病毒基因组RNA相互缠绕形成核衣壳,在病毒RNA合成过程中发挥重要作用^[15]。此外,N蛋白还参与病毒mRNA转录和复制,组织细胞骨架和免疫调节,调控细胞代谢和细胞周期,诱导感染后体液和细胞免疫应答^[5-6,13,16-18]。因此,N蛋白可作为病毒检测的标志性蛋白,已被广泛用于疫苗研制和血清学检测^[19-20]。但目前有关SARS-CoV-2 N蛋白的报道较少,急需对该蛋白的更新认识。

本研究对SARS-CoV-2 N蛋白进行了系统生物信息学分析及原核表达,为阐明N蛋白在SARS-CoV-2感染宿主细胞中的作用机制奠定了基础,为靶向该蛋白的抗病毒药物筛选提供了依据。

1 材料与方法

1.1 材料 pET-22b为本实验室保存,蛋白和DNA marker购自宝生物公司;E. coli感受态细胞(Top10和BL21)、质粒提取及胶回收试剂盒购自全式金有限公司;Xho I与Nde I购自NEB;氨苄西林、IPTG、氯化钠等购自国药集团。

1.2 方法

1.2.1 表达载体pET-22b-N构建 酶切N基因片段和空载体pET-22b,回收酶切后的片段,链接载体与N基因片段后转化感受态细胞Top10,将验证后的重组载体转化感受态细胞BL21,IPTG诱导蛋白表达。

1.2.2 N蛋白诱导表达 37℃、220 r/min培养11.5 h,采用终浓度为1 mmol/L的IPTG诱导,25℃、160 r/min继续培养8 h,10%SDS-PAGE检测诱导前后目的蛋白表达^[21]。

1.2.3 生信分析 依据文献中的网站对N蛋白的理化性质等特性进行分析^[21]。

1.2.4 多序列比对与进化分析 UniProt网站(<https://www.uniprot.org/blast/>)下载与N蛋白序列相似度较高的12种病毒蛋白,分别用Clustal X2和MEGA7.0执行多序列比对和进化树构建。

2 结果

2.1 理化性质 N蛋白是由419个氨基酸(共19种)组成的碱性蛋白,不含半胱氨酸。含负电荷氨基酸(Asp+Glu)36个,其中Asp 24个,Glu 12个;正电荷氨基酸(Arg+Lys+His)64个,其中Arg 29个,Lys 31个,His 4个。含量最多的为甘氨酸(10.30%),其次为丙氨酸和丝氨酸(均为8.80%,表1)。N蛋白分子量为45.62 kD,等电点为10.07,分子式为C₁₉₇₁H₃₁₃₇N₆₀₇O₆₂₉S₇,消光系数为43 890 L/(mol·cm),不稳定指数为55.09;在哺乳动物细胞中的半衰期为30 h,在大肠杆菌体内>10 h,脂肪族系数为52.53,总平均亲水系数为-0.971。

2.2 跨膜结构预测 预测结果显示,SARS-CoV-2 N蛋白不存在跨膜螺旋区(图1),不属于跨膜蛋白。

2.3 亲水/疏水性分析 亲水/疏水性预测结果显示,D371亲水性最强,Score值为-3.556;A220和L221疏水性最强,Score值为2.322;亲水残基数远多于疏水残基数(图2)。因此推测SARS-CoV-2 N蛋白为亲水性蛋白。

表1 新冠病毒N蛋白的氨基酸组成

Tab. 1 Amino acid composition of SARS-CoV-2 N protein

Aminoacids	Number	Frequency (%)	Amino acids	Number	Frequency (%)
Ala(A)	37	8.80	Leu(L)	27	6.40
Arg(R)	29	6.90	Lys(K)	31	7.40
Asn(N)	22	5.30	Met(M)	7	1.70
Asp(D)	24	5.70	Phe(F)	13	3.10
Gln(Q)	35	8.40	Pro(P)	28	6.70
Glu(E)	12	2.90	Ser(S)	37	8.80
Gly(G)	43	10.30	Thr(T)	32	7.60
His(H)	4	1.00	Tyr(Y)	11	2.60
Ile(I)	14	3.30	Val(V)	8	1.90
Trp(W)	5	1.20	-	-	-

2.4 功能位点预测 SARS-CoV-2 N 蛋白含有 1 个 RNA 结合结构域,位于 N 末端 A50-G175 位,该结构域包含 12 个 RNA 结合位点,分别为 A50-A55、R107、Y109、Y111、R149、A156 和 E174。此外,C 末端 P258-A359 位还存在 1 个二聚体界面,包含 51 个二聚体相互作用位点,分别为 P258、Q260-A264、V270、F274、R277、G284-F286、L291、T296、W301、I304-Q306、A308-I320、M322、V324、G328-D341、N345、F346、L353 和 I357。

2.5 磷酸化修饰预测 结果显示,SARS-CoV-2 N 蛋白共有可能的磷酸化修饰 57 个,其中苏氨酸位点 22 个,丝氨酸位点 31 个及酪氨酸位点 4 个(图 3)。上述位点及对应激酶如表 2 所示,通常取 0.5 为阈值,磷酸化势能越高置信度越高。

2.6 糖基化修饰预测

2.6.1 N-糖基化修饰预测 结果显示,N 蛋白有 2 个潜在的 N-糖基化位点,分别为 N47 和 N269(图 4)。

2.6.2 O-糖基化修饰预测 结果显示,N 蛋白有 23 个潜在的 O-糖基化位点,分别为 S33、S176、S180、S183、S184、S186、S187、S188、S190、S193、S194、

S197、T198、S201、S202、S206、T247、T271、T379、S410、S412、S413 和 T417(图 5)。

2.7 信号肽预测 蛋白有无信号序列的判断方法是通过预测 N 端前 70 个氨基酸中是否有潜在的酶切位点,结果显示,N 蛋白不含信号序列(图 6)。

2.8 二级结构分析 结果显示,N 蛋白含有 α -螺旋 96 个(22.20%),延伸链 66 个(16.47%), β -转角 34 个(6.92%),无规则卷曲 223 个(54.42%)。见附图 1 (www.immune99.com)。

表 2 N 蛋白磷酸化修饰及对应激酶
Tab. 2 Phosphorylation sites of N protein and corresponding kinases

Site	Score	Kinase	Site	Score	Kinase	Site	Score	Kinase
S21	0.571	unsp	S187	0.991	unsp	T271	0.597	DNAPK
S23	0.543	unsp	S188	0.997	unsp	T282	0.906	unsp
S26	0.523	unsp	S190	0.992	unsp	T296	0.687	unsp
S33	0.824	unsp	S193	0.997	unsp	Y298	0.743	unsp
S37	0.973	unsp	S194	0.995	unsp	S310	0.636	unsp
T49	0.532	cdc2	S197	0.986	unsp	T325	0.955	unsp
T57	0.676	PKC	T198	0.856	unsp	S327	0.555	PKA
S79	0.998	unsp	S201	0.964	unsp	T332	0.544	CKI
Y87	0.512	EGFR	S202	0.879	unsp	Y360	0.535	unsp
T91	0.988	unsp	T205	0.818	PKC	T362	0.537	unsp
S105	0.996	unsp	S206	0.995	unsp	T366	0.511	PKC
T115	0.788	PKC	S232	0.662	PKC	T379	0.637	DNAPK
T141	0.970	unsp	S235	0.987	unsp	T391	0.647	PKA
Y172	0.587	EGFR	T245	0.905	PKC	T393	0.501	cdc2
S176	0.507	cdc2	T247	0.847	PKC	S410	0.976	unsp
S180	0.984	unsp	S250	0.990	unsp	S412	0.956	unsp
S183	0.956	unsp	S255	0.696	PKC	S413	0.508	cdc2
S184	0.987	unsp	T263	0.736	PKG	S416	0.852	unsp
S186	0.980	unsp	T265	0.737	PKA	T417	0.641	DNAPK

Note: Unsp. Undetermined kinase.

图 1 新冠病毒 N 蛋白的跨膜结构预测
Fig. 1 Transmembrane structure prediction of SARS-CoV-2 N protein

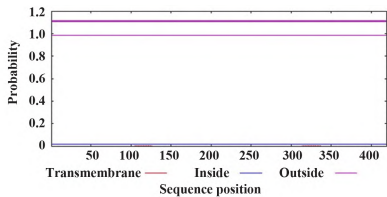


图 2 新冠病毒 N 蛋白亲水/疏水性分析
Fig. 2 Hydrophobic/hydrophobic analysis of SARS-CoV-2 N protein

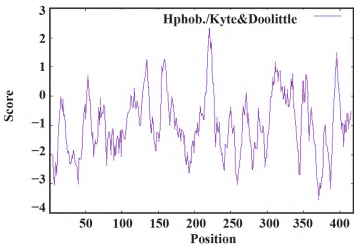


图 4 新冠病毒 N 蛋白 N-糖基化修饰
Fig. 4 N-glycosylation site of SARS-CoV-2 N protein

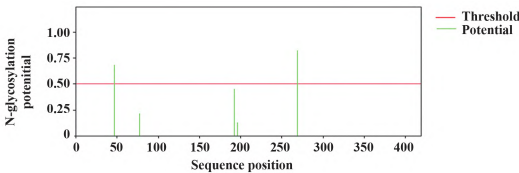


图 3 新冠病毒 N 蛋白磷酸化修饰
Fig. 3 Phosphorylation sites of SARS-CoV-2 N protein

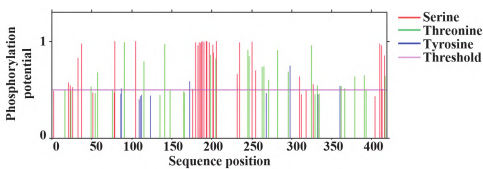
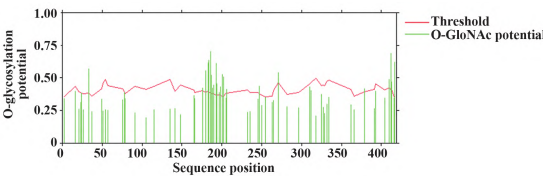


图 5 新冠病毒 N 蛋白 O-糖基化修饰
Fig. 5 O-glycosylation site of SARS-CoV-2 N protein



2.9 三级结构分析 通过 SWISS-MODEL 数据库将目的蛋白与已有蛋白进行序列比对,选择相似度高或同源蛋白进行自动比对建模,预测未知蛋白的三级结构,SARS-CoV-2 N 蛋白三级结构预测结果如图 7。

2.10 抗原表位预测

2.10.1 T 细胞抗原表位预测 IEDB 预测结果显示,N 蛋白共有 15 个 T 细胞抗原表位,分别为 G5-S21、F66-N77、D81-D98、K100-L113、G116-G120、K127-D128、I130-N140、A220-L230、K257-T271、Y298、H300-W301、A305-A336、D340、H356 和 K361-K369 (图 8)。

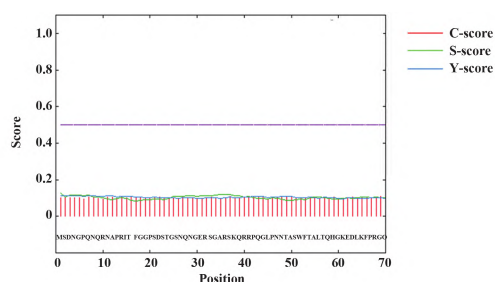


图 6 新冠病毒 N 蛋白信号肽预测

Fig. 6 Prediction of SARS-CoV-2 N protein signal peptide

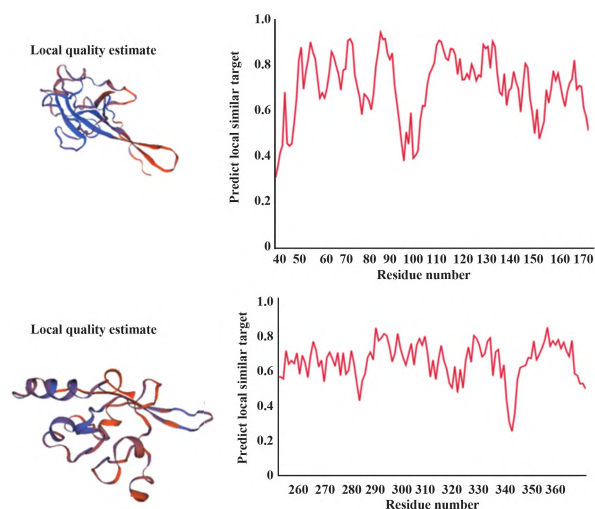


图 7 新冠病毒 N 蛋白三级结构预测及同源蛋白相似性波形图

Fig. 7 Tertiary structure prediction of SARS-CoV-2 N protein and similarity waveform of its homologous proteins

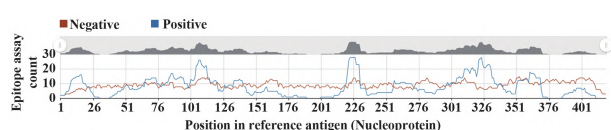


图 8 新冠病毒 N 蛋白 T 细胞抗原表位预测

Fig. 8 Prediction of T cell epitopes of SARS-CoV-2 N protein

2.10.2 B 细胞抗原表位预测 IEDB 预测结果显示,取阈值为 0.5,N 蛋白共有 11 个线性 B 细胞抗原表位,分别为 N4-I15、F17-N48、H59-S105、A119-K127、G137-Q163、T165-D216、R226-A267、R276-K299、D343-D348、D358-D402 和 S404-S416。根据表位分布图最终确定 N 蛋白的优势抗原表位区段为 I15~D216(图 9)。

2.11 相互作用网络分析 N 蛋白与 6 种蛋白存在二元相互作用,分别为 70 kD 热休克蛋白 1A、抗病毒天然免疫应答受体 RIG- I、信号转导和转录激活因子 1- α/β 、信号转导和转录激活因子 2、蛋白酶体激活物复合体亚单位 3 和富含丝氨酸和精氨酸的蛋白特异性激酶 1(图 10)。

2.12 同源性分析 UniProt 网站下载与 SARS-CoV-2 N 蛋白序列相似度较高 12 个 N 蛋白序列,分别为 SARS 冠状病毒、SARS 冠状病毒 PUMC03、SARS 冠状病毒 PUMC02、BtRs- β -冠状病毒/

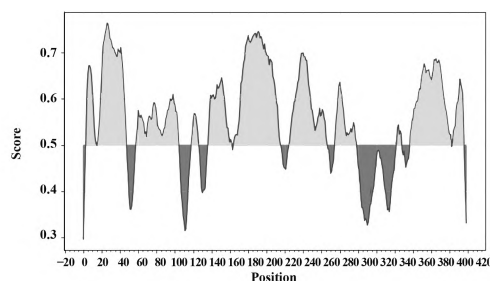
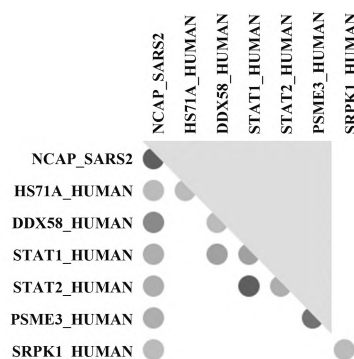


图 9 新冠病毒 N 蛋白 B 细胞抗原表位预测

Fig. 9 Prediction of B cell epitopes of SARS-CoV-2 N protein



Note: NCAP. SARS-CoV-2 nucleocapsid protein; HS71A. Heat shock 70 kD protein 1A; DDX58. Antiviral innate immune response receptor RIG- I ; STAT1. Signal transducer and activator of transcription 1- α/β ; STAT2. Signal transducer and activator of transcription 2; PSME3. Proteasome activator complex subunit 3; SRPK1. SRSF protein kinase 1.

图 10 新冠病毒 N 蛋白的二元相互作用分析

Fig. 10 Binary interaction analysis of SARS-CoV-2 N protein

HuB2013、蝙蝠非典型冠状病毒 WIV1、蝙蝠冠状病毒 Rp/Shaanxi2011、蝙蝠冠状病毒 Rp3/2004、蝙蝠非典型冠状病毒 YNLF_31C、SARS 冠状病毒 WH20、蝙蝠冠状病毒 Cp/Yunnan2011、蝙蝠冠状病毒 HKU3 和蝙蝠冠状病毒 279/2005。多序列比对结果显示,与 SARS-CoV-2 N 蛋白序列相似度最高的为 BtRs- β -冠状病毒/HuB2013 和蝙蝠冠状病毒 Rp/Shaanxi2011 (均为 90.3%),其次为 SARS 冠状病毒、SARS 冠状病毒 PUMC02 和 SARS 冠状病毒 PUMC03 (均为 89.8%)。相似度最低的是蝙蝠冠状病毒 HKU3 (88.8%)。此外,13 条序列中完全相同(*表示)的残基有 284 个(67.8%);性质及其相似(:表示)和性质微弱相近(.表示)的残基均为 19 个(4.5%),见附图 2 (www.immune99.com)。

2.13 进化分析 采用 MEGA7.0 软件对包括 SARS-CoV-2 在内的 13 种病毒 N 蛋白进行多序列比对并构建进化树,结果显示,SARS-CoV-2 与蝙蝠非典型冠状病毒 YNLF_31C 亲缘关系最近,二者聚为一支,置信度为 24,提示其可能具有共同祖先;其次,SARS 冠状病毒与蝙蝠非典型冠状病毒 WIV1 聚为一支,置信度为 20;蝙蝠冠状病毒 Cp/Yunnan2011 与蝙蝠冠状病毒 HKU3 聚为一支,置信度为 49;BtRs- β -冠状病毒/HuB2013 与蝙蝠冠状病毒 Rp/Shaanxi2011 聚为一支,置信度为 25;蝙蝠冠状病毒 Rp3/2004 与蝙蝠冠状病毒 279/2005 聚为一支,置信度为 36(图 11)。

2.14 原核表达载体的构建

2.14.1 空载体的酶切验证 pET-22b 空载体图谱如图 12A,大小约 5 500 bp。经 *Nde* I、*Xho* I 单酶切及双酶切后用琼脂糖凝胶电泳检测,结果显示大小符合预期(图 12B)。

2.14.2 目的片段的酶切及 PCR 验证 对 N 基因片段进行 *Xho* I 与 *Nde* I 双酶切(图 13A);PCR 扩增 N 蛋白编码序列,电泳检测显示,条带大小正确(图 13B)。

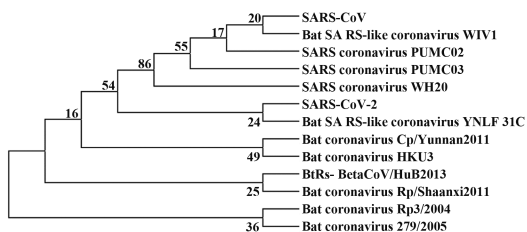
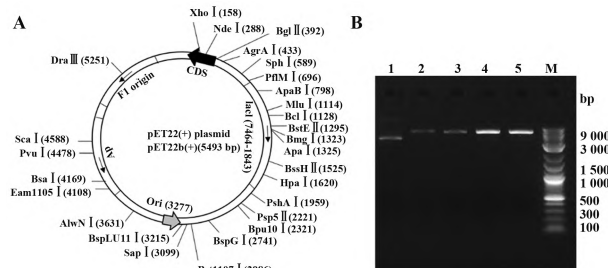


图 11 新冠病毒 N 蛋白的系统进化树

Fig. 11 Phylogenetic tree of SARS-CoV-2 N protein

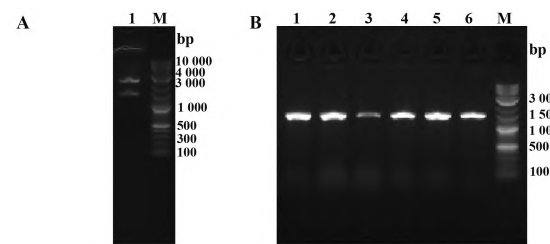
2.14.3 重组质粒酶切验证及转化 将 pET-22b-N 进行双酶切,电泳检测结果显示,大小正确(图 14A)。将 pET-22b-N 转化 Top10 感受态细胞,在 LB 平板(Amp^r)上于 37 °C 培养 12~13 h(图 14B 左);



Note: A. Map of pET-22b vector; B. Electrophoretic detection of empty vector pET-22b by restriction endonuclease digestion. 1. No digestion; 2. *Nde* I digestion; 3. *Xho* I digestion; 4. *Nde* I and *Xho* I digestion; M. 2-log DNA marker.

图 12 pET-22b 空载体检测及酶切验证

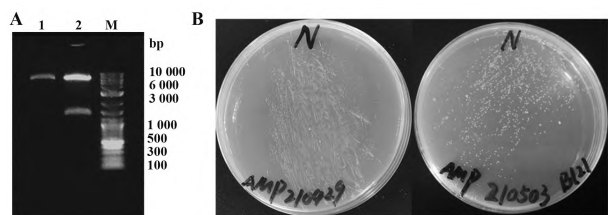
Fig. 12 Detection and restriction endonuclease digestion of pET-22b empty vector



Note: A. Detection of double restriction endonuclease digestion products of N protein coding gene; M. 2-log DNA marker; 1. *Nde* I and *Xho* I digestion products of N protein coding gene; B. Colony PCR validation; 1~6. Colony PCR products.

图 13 N 蛋白基因片段的酶切及菌落 PCR 验证

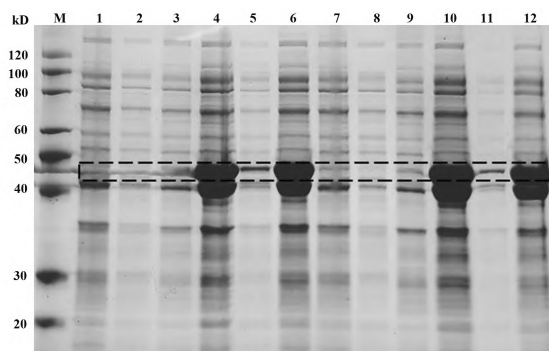
Fig. 13 Restriction endonuclease digestion and colony PCR validation of N protein gene fragment



Note: A. Verification of recombinant plasmid by *Nde* I and *Xho* I digestion; 1. Recombinant plasmid was not digested; 2. *Nde* I and *Xho* I digestion of recombinant plasmid; M. 2-log DNA marker; B. Recombinant plasmid was transformed into competent cells, Top10 on the left and BL21 on the right.

图 14 重组质粒 pET-22b-N 的酶切验证与转化

Fig. 14 Verification and transformation of recombinant plasmid pET-22b-N



Note: M. Protein Marker; 1~3. Whole cell lysate, centrifugation supernatant of whole cell lysate and centrifugation precipitation of whole cell lysate before IPTG induction; 4~6. Whole cell lysate, centrifugation supernatant of whole cell lysate and centrifugation precipitation of whole cell lysate after IPTG induction; 7~9. Whole cell lysate, centrifugation supernatant of whole cell lysate and centrifugation precipitation of whole cell lysate before IPTG induction; 10~12. Whole cell lysate, centrifugation supernatant of whole cell lysate and centrifugation precipitation of whole cell lysate after IPTG induction.

图 15 N 蛋白诱导表达检测

Fig. 15 Detection of induced expression of N protein

将验证后的重组质粒转化 BL21, 37 °C 培养 13 h (图 14B 右)。

2.15 蛋白诱导表达 挑取 2 个单菌落, IPTG 诱导后 10% 聚丙烯酰胺凝胶电泳检测, 结果显示: 诱导后 N 蛋白表达含量比诱导前显著增多(泳道 4 高于泳道 1, 泳道 10 高于泳道 7)。为进一步优化 N 蛋白提纯条件, 分别对诱导前/后的全菌、上清、沉淀中 N 蛋白表达进行检测, 结果如图 15(黑色虚线框): 诱导后的 N 蛋白大量表达于沉淀中(泳道 6、12), 因此在提纯该蛋白时考虑从沉淀进行。

3 讨论

N 蛋白是病毒检测的重要靶点和药物作用靶标, 保守性高, 是 SARS-CoV-2 的重要抗原, 参与病毒基因组包装和病毒颗粒释放^[22-23]。血清学诊断发现, SARS 患者血清中针对 N 蛋白的特异性抗体比 SARS-CoV 其他结构蛋白抗体具有更高的敏感性和持久性^[24-25]。此外, N 蛋白抗体在感染早期具有较高特异性^[26]。因此, 建立大量表达 N 蛋白的方法并对其进行系统生物信息学分析有助于深入了解该蛋白的功能和病毒 RNA 复制机制。

本研究显示, SARS-CoV-2 N 蛋白呈碱性, 富含正电荷, 有助于其与病毒基因组 RNA 结合^[27-28]。N 蛋白亲水性较强, 无信号肽, 不属于分泌型蛋白。

二级结构分析发现, N 蛋白以无规则卷曲为主 (54.42%), 与钟琦等^[29]报道一致, 为抗病毒药物研发与新冠肺炎患者诊断提供了参考。

抗原表位分析发现, N 蛋白有 15 个 T 细胞表位和 11 个线性 B 细胞表位, 其中 I15-D216 是其优势 B 细胞表位。近期研究显示, N 蛋白存在较高频率的 R203K 和 G204R 双位点变异, 该变异破坏了 N 蛋白整体结构稳定性和灵活性^[30]。尽管 N 蛋白被认为是 SARS-CoV-2 疫苗开发和诊疗的重要靶点, 但其持续进化的特点易产生传染性更强的突变株, 导致病毒与已有抗体亲和力下降, 给疫苗研发带来巨大挑战^[3]。因此, 需进一步进行免疫信息学分析, 同时持续监测和追踪 SARS-CoV-2 N 蛋白演变。

糖基化是重要的蛋白翻译后修饰过程, 高度亲水的糖基对蛋白理化性质和生理功能具有重要影响。磷酸化是调节和控制蛋白活力和功能的最基本、最普遍, 也是最重要的机制。本研究显示, N 蛋白有 2 个可能的 N-糖基化修饰和 23 个 O-糖基化修饰, 有 31 个丝氨酸磷酸化修饰, 22 个苏氨酸磷酸化修饰, 以及 4 个酪氨酸磷酸化修饰。RAHMAN 等^[30]对 N 基因突变频率较高的 20 例确诊病例与感染率的关系进行了系统研究, 发现 N 蛋白磷酸化修饰位点 S250、S255、S310、S325、S327、T141、T247、T263、T265、T271 和 T362 存在 2~3 种单点变异类型, 分别为 S250F/P、S255F/A/P、S310I/C/N、S325I/R/A、S327L/P、T141I/P/A、T247I/A、T263I/A、T265I/A、T271I/A 和 T362I/K。此外, S193、S194、S201、S202、T205 和 S206 等位点还存在不同形式的氨基端缺失变异类型。O-糖基化修饰位点 S180 是 N 蛋白突变形式最多的位点, 具有 6 种氨基酸变异形式。以上变异可能影响 SARS-CoV-2 的传播和侵染能力, 但课题组并未发现 N 蛋白突变频率与 SARS-CoV-2 感染率显著相关。

N 蛋白通过其 N-末端结构域 (N-NTD) 与病毒 RNA 形成复合物, 在病毒复制周期中发挥重要作用, 使该结构域成为重要的药物靶点^[31-33]。本研究功能位点分析显示, SARS-CoV-2 N 蛋白含有 1 个 RNA 结合结构域, 位于 NTD A50-G175 位, 包含 12 个 RNA 结合位点, 与 KANG 等^[5]结果一致, 发现 N 蛋白的 N47-A50 残基具有高度灵活性且向外伸展, 打开了 RNA 结合口袋, 有助于同病毒基因组 RNA 高级结构结合。RAHMAN 等^[30]发现, N 蛋白 A55S、P67T、

D81Y、A119V、P122L、D128Y、L139F和D144Y变异增加了NTD结构稳定性,而E62V、D103Y、A119S、A152S、A156S、L161F和P168S变异则降低了该结构稳定性。同样,E62V、P67T、D144Y和P168S变异增加了N蛋白分子柔性,而A55S、D81Y、D103Y、A119S、A119V、P122L、D128Y、L139F、A152S、A156S和L161F变异则降低了RNA结合结构域分子柔性,进一步证明上述位点在功能上的重要性。本研究还显示,CTD第P258-A359位存在1个二聚结构域,包含51个二聚相互作用位点。RAHMAN等^[30]研究显示,H300Y、T325I、S327L、T334I、D340N改变增加了N蛋白稳定性,而Q289H、I292T、P344S、D348H改变则降低了分子稳定性。Q289H、I292T变异增加了分子灵活性,而H300Y、T325I、S327L、T334I、D340N、P344S和D348H则降低了分子柔韧性。提示这些结构域和功能位点在提高病毒转录和组装效率方面起重要作用^[34-35]。

多序列比对显示,与SARS-CoV-2 N蛋白序列相似度最高的是BtRs- β -冠状病毒/HuB2013和蝙蝠冠状病毒Rp/Shaanxi2011(均为90.3%),其次是SARS冠状病毒、SARS冠状病毒PUMC02和SARS冠状病毒PUMC03(均为89.8%)。此外,13条序列中完全相同的残基占67.8%,提示N蛋白在进化上高度保守。进化分析显示,SARS-CoV-2与蝙蝠非典型冠状病毒YNLF_31C亲缘关系最近,二者聚为一支,置信度为24,提示其可能有共同祖先。

N蛋白是SARS-CoV-2 IgM/IgG快速检测试剂卡的核心原材料,目前已有多家公司在真核系统中完成了N蛋白表达。大肠杆菌原核表达系统以其安全性好、易放大培养、周期短等优点被广泛用于基因工程药物生产^[36]。本研究构建了原核表达载体pET-22b-N,菌落PCR扩增和测序验证结果显示,载体构建正确。在大肠杆菌中表达N蛋白并研究其表达特性,结果显示,经1 mmol/L IPTG诱导,N蛋白即可大量表达,且主要在沉淀中表达。从IgM/IgG快速检测试剂卡成本考虑,大肠杆菌系统中N蛋白表达量高、速度快,将大大降低SARS-CoV-2筛查成本^[37]。

本研究有助于全面揭示N蛋白的生物学功能,为开发和设计靶向N蛋白的快速诊断方法和抗病毒药物提供了依据。

参考文献:

- [1] 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. DOI: 10.15252/embj.2020105938.
- [2] GUAN W J, NI Z Y, HU Y, *et al.* Clinical characteristics of coronavirus disease 2019 in China [J]. *N Engl J Med*, 2020, 382(18): 1708-1720. DOI: 10.1056/NEJMc2005203.
- [3] 何晓波, 虞淦军, 吴艳峰. 新型冠状病毒突变株对传染性和疾病进展及免疫保护影响的研究进展 [J]. *中国免疫学杂志*, 2021, 37(16): 2021-2028. DOI: 10.3969/j.issn.1000-484X.2021.16.021.
- [4] MALIK Y A. Properties of coronavirus and SARS-CoV-2 [J]. *Malaysian J Pathol*, 2020, 42(1): 3-11.
- [5] KANG S, YANG M, HONG Z, *et al.* Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites [J]. *Acta Pharm Sin B*, 2020, 10(7): 1228-1238. DOI: 10.1016/j.apsb.2020.04.009.
- [6] YOSHIMOTO F K. The proteins of severe acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19 [J]. *Protein J*, 2020, 39(3): 198-216. DOI: 10.1007/s10930-020-09901-4.
- [7] LE BERT N, TAN A T, KUNASEGARAN K, *et al.* SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls [J]. *Nature*, 2020, 584(7821): 457-462. DOI: 10.1038/s41586-020-2550-z.
- [8] CHEN Y, LIU Q, GUO D. Emerging coronaviruses: Genome structure, replication, and pathogenesis [J]. *J Med Virol*, 2020, 92(4): 418-423. DOI: 10.1002/jmv.25681.
- [9] GUO Y R, CAO Q D, HONG Z S, *et al.* The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak-an update on the status [J]. *Mil Med Res*, 2020, 7(1): 11. DOI: 10.1186/s40779-020-00240-0.
- [10] 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. DOI: 10.3969/j.issn.1009-0002.2020.06.007.
- [11] GORDON D E, HIATT J, BOUHADDOU M, *et al.* Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms [J]. *Science*, 2020, 370(6521): eabe9403. DOI: 10.1126/science.abe9403.
- [12] LIU S J, LENG C H, LIEN S P, *et al.* Immunological characterizations of the nucleocapsid protein based SARS vaccine candidates [J]. *Vaccine*, 2006, 24(16): 3100-3108. DOI: 10.1016/j.vaccine.2006.01.058.
- [13] ZUNIGA S, CRUZ J L, SOLA I, *et al.* Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription [J]. *J Virol*, 2010, 84(4): 2169-2175. DOI: 10.1128/JVI.02011-09.
- [14] 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. DOI: 10.1128/JVI.01331-15.
- [15] LU R, ZHAO X, LI J, *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding [J]. *Lancet*, 2020, 395(10224): 565-574. DOI: 10.1016/S0140-6736(20)30251-8.

- [16] SURJIT M, LIU B, CHOW V T, *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. DOI: 10.1074/jbc.M509233200.
- [17] 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. DOI: 10.1093/cid/ciaa461.
- [18] NI L, YE F, CHENG M L, *et al.* Detection of SARS-CoV-2-specific humoral and cellular immunity in COVID-19 convalescent individuals[J]. *Immunity*, 2020, 52(6):971-977. DOI: 10.1016/j.immuni.2020.04.023.
- [19] 巩 癰, 廖青云, 于倩倩, 等. 冠状病毒研究态势分析[J]. *中国生物工程杂志*, 2020, 40(1):21-37. DOI: 10.13523/j.cb.2002102.
- [20] AHMED S F, QUADEER A A, MCKAY M R. 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. DOI: 10.3390/v12030254.
- [21] 徐本锦, 刘 玲. 大肠杆菌翻译起始因子 IF1 的生物信息学分析、表达载体构建及蛋白纯化[J]. *基因组学与应用生物学*, 2020, 39(12):5606-5616. DOI: 10.13417/j.gab.039.005606.
- [22] 刘 彬, 秦照玲, 戚中田. 新型冠状病毒基因组结构与蛋白功能[J]. *微生物与感染*, 2020, 15(1):52-57. DOI: 10.3969/j.issn.1673-6184.2020.01.010.
- [23] ZENG W, LIU G, MA H, *et al.* Biochemical characterization of SARS-CoV-2 nucleocapsid protein[J]. *Biochem Biophys Res Commun*, 2020, 527(3):618-623. DOI: 10.1016/j.bbrc.2020.04.136.
- [24] TAN Y J, GOH P Y, FIELDING B C, *et al.* Profiles of antibody responses against severe acute respiratory syndrome coronavirus recombinant proteins and their potential use as diagnostic markers[J]. *Clin Diagn Lab Immunol*, 2004, 11(2):362-371. DOI: 10.1128/cdli.11.2.362-371.2004.
- [25] SHI Y, YI Y, LI P, *et al.* Diagnosis of severe acute respiratory syndrome (SARS) by detection of SARS coronavirus nucleocapsid antibodies in an antigen-capturing enzyme-linked immunosorbent assay[J]. *J Clin Microbiol*, 2003, 41(12):5781-5782. DOI: 10.1128/JCM.41.12.5781-5782.2003.
- [26] LEUNG D T, TAM F C, MA C H, *et al.* Antibody response of patients with severe acute respiratory syndrome (SARS) targets the viral nucleocapsid[J]. *J Infect Dis*, 2004, 190(2):379-386. DOI: 10.1086/422040.
- [27] SAIKATENDU K S, JOSEPH J S, SUBRAMANIAN V, *et al.* Ribonucleocapsid formation of severe acute respiratory syndrome coronavirus through molecular action of the N-terminal domain of N protein[J]. *J Virol*, 2007, 81(8):3913-3921. DOI: 10.1128/JVI.02236-06.
- [28] CHEN C Y, CHANG C K, CHANG Y W, *et al.* Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA[J]. *J Mol Biol*, 2007, 368(4):1075-1086. DOI: 10.1016/j.jmb.2007.02.069.
- [29] 钟 琦, 黄志鑫, 陈晓舟. 基于位点特异性打分矩阵的卷积神经网络预测 SARS-CoV-2 核衣壳蛋白的蛋白质二级结构[J]. *云南民族大学学报(自然科学版)*, 2021, 30(1):52-57. DOI: 10.3969/j.issn.1672-8513.2021.01.010.
- [30] RAHMAN M S, ISLAM M R, ALAM A, *et al.* Evolutionary dynamics of SARS-CoV-2 nucleocapsid protein and its consequences[J]. *J Med Virol*, 2021, 93(4):2177-2195. DOI: 10.1002/jmv.26626.
- [31] ZUWALA K, GOLDA A, KABALA W, *et al.* The nucleocapsid protein of human coronavirus NL63[J]. *PLoS One*, 2015, 10(2):e0117833. DOI: 10.1371/journal.pone.0117833.
- [32] GRUNEWALD M E, FEHR A R, ATHMER J, *et al.* The coronavirus nucleocapsid protein is ADP-ribosylated[J]. *Virology*, 2018, 517:62-68. DOI: 10.1016/j.virol.2017.11.020.
- [33] PAPAGEORGIOU N, LICHIERE J, BAKLOUTI A, *et al.* Structural characterization of the N-terminal part of the MERS-CoV nucleocapsid by X-ray diffraction and small-angle X-ray scattering[J]. *Acta Crystallogr D Struct Biol*, 2016, 72(Pt 2):192-202. DOI: 10.1107/S2059798315024328.
- [34] SZELAZEK B, KABALA W, KUS K, *et al.* Structural characterization of human coronavirus NL63 N protein[J]. *J Virol*, 2017, 91(11):e02503-e02516. DOI: 10.1128/JVI.02503-16.
- [35] NGUYEN T H V, LICHIERE J, CANARD B, *et al.* Structure and oligomerization state of the C-terminal region of the Middle East respiratory syndrome coronavirus nucleoprotein[J]. *Acta Crystallogr D Struct Biol*, 2019, 75(Pt 1):8-15. DOI: 10.1107/S2059798318014948.
- [36] 李生军, 张海霞, 冯若飞. 大肠杆菌表达病毒样颗粒的研究进展[J]. *中国免疫学杂志*, 2021, 37(12):1526-1532. DOI: 10.3969/j.issn.1000-484X.2021.12.024.
- [37] 鞠守勇. 新型冠状病毒(SARS-CoV-2)核衣壳蛋白(N蛋白)的表达研究[J]. *武汉职业技术学院学报*, 2020, 19(2):104-107. DOI: 10.3969/j.issn.1671-931X.2020.02.022.

[收稿 2021-07-19 修回 2021-08-31]

(编辑 周文瑜)

