

新冠病毒刺突糖蛋白结构与功能的生物信息学分析及原核表达



开放科学
(OSID)

徐本锦^{1*}, 宣焱^{1#}, 杜淼^{1#}, 刘玲^{1*}, 吴惠文², 宋彬好²,
侯艳香¹, 陈利荣¹, 刘晓梁²

(1. 山西医科大学 汾阳学院 医学检验系, 汾阳 032200; 2. 山西医科大学 汾阳学院 科技中心, 汾阳 032200)

摘要: SARS-CoV-2是一种高致病性且传播迅速的病原体,通过刺突糖蛋白(Spike glycoprotein, S蛋白)识别宿主细胞表面的受体来实现入侵和感染。对S蛋白进行系统的生物信息学分析和原核表达,有助于深入理解S蛋白的功能和阐明该蛋白介导病毒感染的分子机制。本文采用Protparam、Pfam、TMHMM、ExPASy-ProtScale、PSORT II、SignalP、UniProt、NetPhos 3.1、NetNGlyc 1.0、NetOGlyc 4.0和BLAST等生物信息学软件和数据库对S蛋白的理化性质、亚细胞定位、翻译后修饰及相互作用网络等生物学特性进行了系统分析。利用Clustal X2和MEGA7.0软件对该蛋白进行了基于氨基酸序列的同源性分析和系统进化分析。最后,通过分子克隆技术构建重组表达载体pET-22b-S并进行原核表达。结果显示,S蛋白由1273个氨基酸组成,分子量141.2 kD,等电点6.24,有两个卷曲螺旋结构,一个跨膜螺旋区,疏水性较强。S蛋白包含刺突受体结合结构域和S2糖蛋白结构域,主要分布于宿主细胞的内质网膜和细胞膜,含有136个潜在的磷酸化位点和20个可能的糖基化位点。与SARS-CoV-2 S蛋白序列一致性最高的是SARS冠状病毒、SARS冠状病毒WH20和蝙蝠冠状病毒HKU3,均为76%。SARS-CoV-2与SARS冠状病毒和蝙蝠冠状病毒聚为一大支,提示它们可能具有共同的祖先。S蛋白主要在细菌裂解液离心之后的沉淀中表达,这为后续的结构分析和疫苗研发奠定了基础。S蛋白在SARS冠状病毒和蝙蝠冠状病毒之间保守性较高,提示其在病毒入侵过程中具有重要功能。SARS-CoV-2与SARS冠状病毒和蝙蝠冠状病毒可能具有共同的祖先。本研究为SARS-CoV-2 S蛋白的表达纯化、结构与功能研究提供了重要的数据基础,有助于全面揭示S蛋白的生物学功能,同时为设计和筛选靶向S蛋白的新型抗病毒药物提供了科学依据。

关键词: 新冠病毒(SARS-CoV-2);刺突糖蛋白(S蛋白);生物信息学分析;原核表达

中图分类号: Q71 **文献标识码:** A **文章编号:** 1000-8721(2022)01-0041-16

DOI: 10.13242/j.cnki.bingduxuebao.004090

收稿日期: 2021-06-12; **接受日期:** 2021-08-17

基金项目: 山西医科大学汾阳学院引进人才启动金项目(项目编号:2020A01),题目:RelA介导大肠杆菌严谨反应的分子机制研究;山西省高等学校科技创新项目(项目编号:2020L0749),题目:大肠杆菌核糖体再循环因子RRF的功能研究;国家级大学生创新创业训练计划项目(项目编号:202117114001),题目:新冠病毒Nsp1基因的克隆、表达载体构建及其对大肠杆菌生长的影响;山西省高等学校大学生创新创业训练计划重点项目(项目编号:S202117114008),题目:SARS-CoV-2核壳蛋白过表达影响细菌蛋白翻译的分子机制;吕梁市科技计划项目(项目编号:2020SHFZ29),题目:吕梁市食源性“超级耐药菌”MRSA的遗传多态性与溯源研究;吕梁市科技计划项目(项目编号:2020XGZX104),题目:新冠肺炎疾病转归预测和风险评估的模型构建与评估。

#共同第一作者: 徐本锦,宣焱,杜淼

作者简介: 徐本锦(1987-),男,博士,讲师,从事基因工程与病原分子生物学研究,E-mail:benjin_ibprnalab@sina.com;宣焱(2000-),男,本科生,从事医学检验技术研究,E-mail:965113920@qq.com;杜淼(2000-),男,本科生,从事医学检验技术研究,E-mail:1115871615@qq.com

***通讯作者:** 刘玲(1983-),女,硕士,助教,从事基因工程与病原分子生物学研究,E-mail:ll772x@sxmu.edu.cn;徐本锦(1987-),男,博士,讲师,从事基因工程与病原分子生物学研究,E-mail:benjin_ibprnalab@sina.com

2019年底,中国武汉出现了新型冠状病毒引起的肺炎疫情,该病毒致病性强、传播迅速,给国内外人民生命健康安全带来了巨大威胁,引发了严重的全球公共卫生事件^[1]。患者出现类似于2003年的SARS-CoV和2012年的MERS-CoV感染引起的症状,包括发热、头痛、干咳、呼吸困难和肺炎等^[2-3]。2020年2月,国际病毒分类委员会将新型冠状病毒命名为“SARS-CoV-2”。

SARS-CoV-2属于冠状病毒科, β 冠状病毒属^[4],与2003年引起非典的SARS-CoV病毒具有较高的相似性,二者序列一致性高达79%^[5]。SARS-CoV-2的基因组约30 kb(图1),编码9860个氨基酸^[6]。其膜上有刺突糖蛋白(Spike glycoprotein, S蛋白)、包膜蛋白和膜蛋白,膜内有衣壳蛋白和单股正链RNA^[5,7]。SARS-CoV-2通过S蛋白识别宿主细胞表面的受体:血管紧张素转化酶2(Angiotensin converting enzyme 2, ACE2)来实现入侵和感染^[1,8-9]。尽管SARS-CoV-2与SARS-CoV属于同

一科同一属,但是 SARS-CoV-2 的传染性更强,比 SARS-CoV 给人民造成的损失更大^[10-11],主要是因前者的 S 蛋白对人体细胞表面的 ACE2 具有比后者更强的亲和力^[12]。

本研究利用生物信息手段对 SARS-CoV-2 S 蛋白的理化性质、亚细胞定位、翻译后修饰及蛋白相互作用网络等生物学特性进行了系统分析。此外,构

建了 S 蛋白重组表达载体并进行了原核表达。研究结果为深入解析 S 蛋白的生物学功能、阐明 S 蛋白介导 SARS-CoV-2 感染人类的分子机制奠定了基础,为靶向 S 蛋白的新型抗病毒药物筛选与应用提供了线索和依据,同时为病毒-宿主相互作用理论提供了新的见解。



图1 新冠病毒(SARS-CoV-2)的基因组结构^[13]

Figure 1 Genome structure of SARS-CoV-2^[13]

材料与amp;方法

1 材料

SARS-CoV-2 S 蛋白的序列从 UniProt 网站 (<https://www.uniprot.org/>) 下载 (Entry: P0DTC2), pET-22b 为本实验室保存, DNA marker 购自大连宝生物, *Nde*I 和 *Xho*I 购自 NEB, 胶回收试剂盒、Protein Ruler II、质粒抽提试剂盒及 *E.coli* 感受态细胞 (Top10 和 BL21) 购自北京金全生物技术有限公司, 氨苄和 IPTG 购自国药集团, NaCl 购自天津欧博凯化工有限公司, 琼脂粉购自北京酷来博有限公司, 酵母提取物和胰蛋白胨购自 OXOID 公司。

2 蛋白的理化性质分析

利用在线网站 (<https://web.expasy.org/protparam/>) 对 S 蛋白的等电点、氨基酸组成、分子量、不稳定指数和消光系数等理化性质进行预测; 利用 ExPASy 网站 (<https://web.expasy.org/protscale/>) 进行亲/疏水性预测; 用 TMHMM Server v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) 进行跨膜螺旋预测; 通过 ExPASy 网站 (https://embnet.vital-it.ch/software/COILS_form.html) 对 S 蛋白进行卷曲螺旋分析。

3 信号肽和亚细胞定位分析

通过在线工具 SignalP (<http://www.cbs.dtu.dk/services/SignalP/>) 对 S 蛋白进行信号肽预测; 利用在线网站 PSORT II (<https://psort.hgc.jp/form2.html>) 对该蛋白进行亚细胞定位分析。

4 磷酸化和糖基化位点预测及相互作用网络分析

用 NetPhos 3.1 Server (<http://www.cbs.dtu.dk/services/NetPhos/>) 对 S 蛋白进行磷酸化位点预测;

用 NetNGlyc 1.0 Server (<http://www.cbs.dtu.dk/services/NetNGlyc/>) 进行 N-糖基化位点进行预测; 用 NetOGlyc 4.0 Server (<http://www.cbs.dtu.dk/services/NetOGlyc/>) 进行 O-糖基化位点预测; 通过 UniProt 网站 (<https://www.uniprot.org/uniprot/P0DTC2>) 进行二元相互作用网络分析。

5 S 蛋白结构域预测及受体结合结构域分析

利用在线软件 Pfam (<http://pfam.xfam.org/search/sequence>) 预测 S 蛋白的结构域。利用 NPS @ SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) 对 S 蛋白的受体结合结构域进行二级结构分析, 构象状态数为 4 (Helix, Sheet, Turn, Coil), 相似性阈值为 8, 窗口宽度 17。

6 多序列比对及系统进化树构建

通过 UniProt 网站的 BLAST 功能 (<https://www.uniprot.org/blast/>) 检索与 SARS-CoV-2 S 蛋白序列相似性较高的 12 个病毒 S 蛋白序列, 使用 Clustal X2 对上述 S 蛋白进行多序列比对; 利用 MEGA7.0 基于邻接法 (Neighbour-Joining, NJ) 经 1500 次置信度构建进化树。

7 表达载体 pET-22b-S 构建

对空载体 pET-22b 和 S 基因片段同时进行 *Nde*I 和 *Xho*I 的双酶切, 用 Quick Gel Extraction Kit 回收酶切片段, 将目的片段与载体链接后转化 Top10 感受态细胞, 菌落 PCR 验证。然后将重组质粒转化 BL21, 进行蛋白诱导表达。

8 S 蛋白的诱导表达

S 蛋白的诱导表达参照文献执行^[14], 制样完成后用 12% 的 SDS-PAGE 进行检测。

结 果

1 S蛋白的理化性质分析

SARS-CoV-2 S蛋白由1273个氨基酸组成(表1),负电荷氨基酸(天冬氨酸+谷氨酸)有110个,其中谷氨酸48个(3.8%),天冬氨酸62个(4.9%);带正电荷氨基酸(赖氨酸+精氨酸+组氨酸)有120个,其中赖氨酸61个(4.8%),组氨酸17个(1.3%),精氨酸42个(3.3%)。含量最高的是亮氨酸(8.5%),其次是丝氨酸(7.8%)、苏氨酸和缬氨酸(均为7.6%)。S蛋白的分子量为141.2 kD,等电点6.24,原子总数为19710,分子式为 $C_{6336}H_{9770}N_{1656}O_{1894}S_{54}$;当波长为280 nm时,水溶液中的消光系数为 $148960 M^{-1}cm^{-1}$;体外条件下在哺乳动物网织红细胞中的半衰期为30 h,脂溶系数为84.67,总平均亲水系数(GRAVY)为-0.079;不稳定指数为33.01,说明该蛋白稳定。

2 S蛋白卷曲螺旋分析

卷曲螺旋(Coiled coil)分析表明,在3种不同的

表1 新冠病毒(SARS-CoV-2)S蛋白的氨基酸组成

Table 1 Amino-acid composition of the spike glycoprotein of SARS-CoV-2

Amino acids	Number	Frequency (%)	Amino acids	Number	Frequency (%)
Ala (A)	79	6.2	Leu (L)	108	8.5
Arg (R)	42	3.3	Lys (K)	61	4.8
Asn (N)	88	6.9	Met (M)	14	1.1
Asp (D)	62	4.9	Phe (F)	77	6.0
Cys (C)	40	3.1	Pro (P)	58	4.6
Gln (Q)	62	4.9	Ser (S)	99	7.8
Glu (E)	48	3.8	Thr (T)	97	7.6
Gly (G)	82	6.4	Trp (W)	12	0.9
His (H)	17	1.3	Tyr (Y)	54	4.2
Ile (I)	76	6.0	Val (V)	97	7.6

窗口宽度(14, 21, 28)共检测到两个卷曲螺旋结构(Score>0.1),分别位于残基Q949~I993位和残基V1176~L1203位(图2)。

3 S蛋白跨膜结构预测

跨膜结构预测显示,S蛋白存在一跨膜螺旋区,位于残基W1214位~T1273位,因此S蛋白是一个跨膜蛋白(图3)。

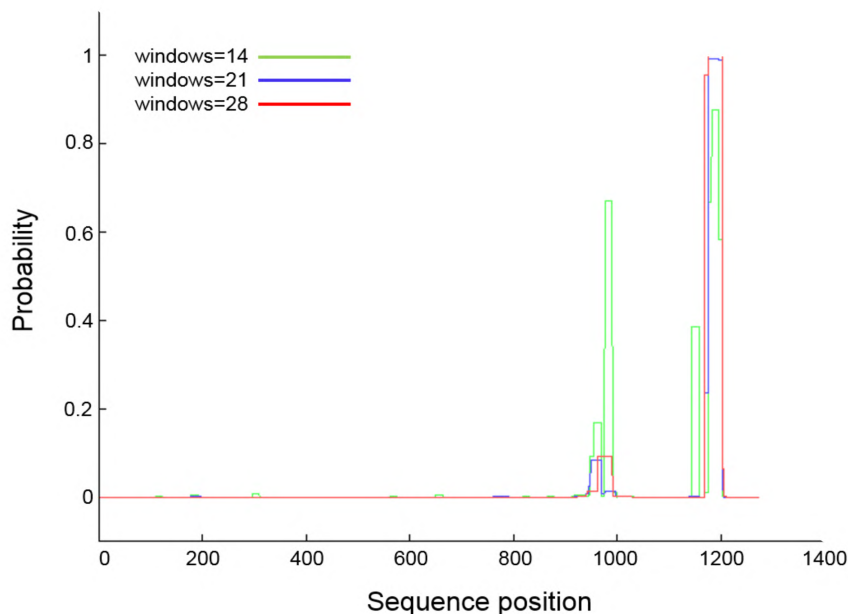


图2 新冠病毒(SARS-CoV-2)S蛋白的卷曲螺旋分析

Figure 2 Coiled coil analysis of the spike glycoprotein of SARS-CoV-2

4 S蛋白亲/疏水性分析

亲/疏水预测结果显示,S蛋白的残基N679亲水性最强,Score值为-2.589,L7疏水性最强,Score

值为3.222,亲水氨基酸与疏水氨基酸数目差异不大。从图4可以看出,S蛋白亲水性较弱,疏水性较高。

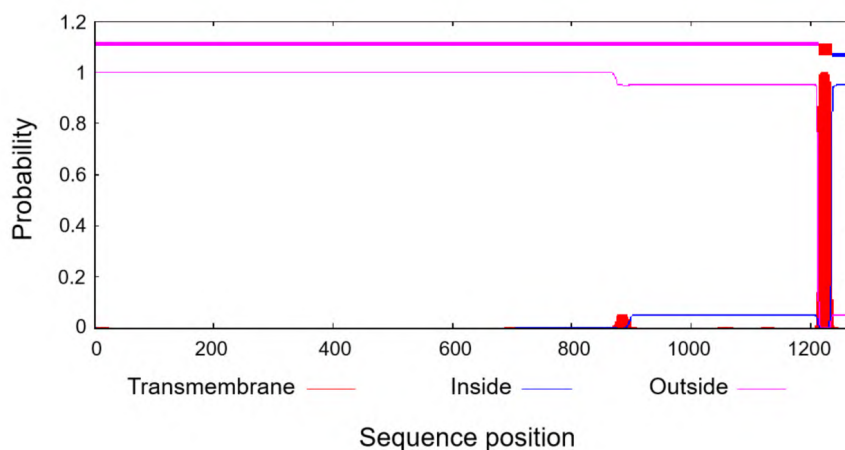


图3 新冠病毒(SARS-CoV-2)S蛋白的跨膜结构预测

Figure 3 Prediction of the transmembrane structures of the spike glycoprotein of SARS-CoV-2

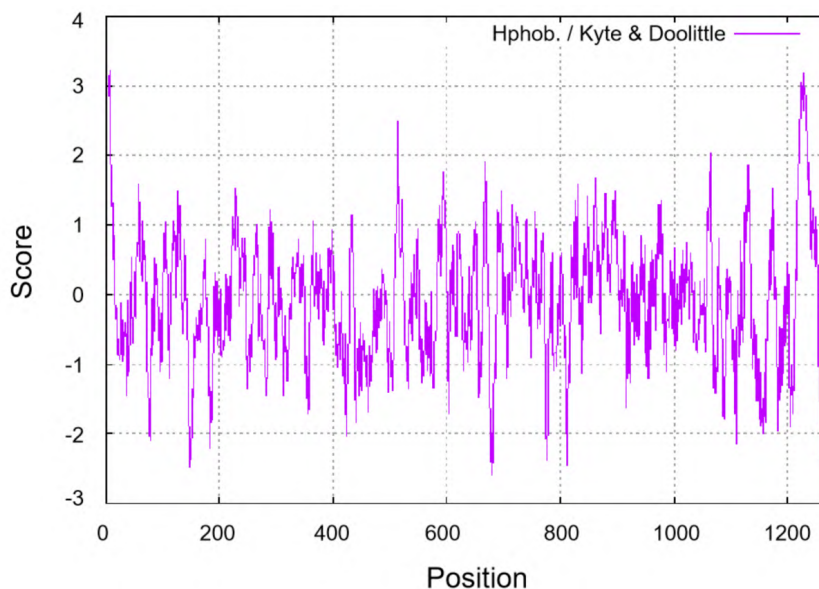


图4 新冠病毒(SARS-CoV-2)S蛋白的亲疏水性分析

Figure 4 Hydrophilic/hydrophobic properties of the spike glycoprotein of SARS-CoV-2

5 S蛋白亚细胞定位和信号肽预测

亚细胞定位分析显示,S蛋白主要分布于宿主细胞的内质网膜(39.1%)、细胞膜(21.7%)、线粒体和高尔基体(均为13.0%),在细胞外(4.3%)、细胞质(4.3%)和液泡(4.3%)中也有分布。信号肽预测显示,S蛋白具有信号肽的概率为0.9689,远高于阈值0.5,说明该蛋白是分泌型蛋白。此外,信号肽切割位点位于C15和V16之间(SQC-VN),概率为0.5353(图5)。

6 S蛋白的常见突变位点

通过检索数据库发现,目前已有多个 SARS-

CoV-2 S蛋白的突变位点被报道(表2)。

7 S蛋白磷酸化位点预测

磷酸化位点预测结果显示,S蛋白共有136个可能的磷酸化位点,其中68个丝氨酸磷酸化位点,45个苏氨酸磷酸化位点,以及23个酪氨酸磷酸化位点(图6)。通常情况下取阈值为0.5,磷酸化势能越高,置信度越高。

8 S蛋白糖基化位点预测

糖基化位点预测结果显示,S蛋白有17个潜在的N-糖基化位点,分别是N17、N61、N74、N122、N149、N165、N234、N282、N331、N343、N603、

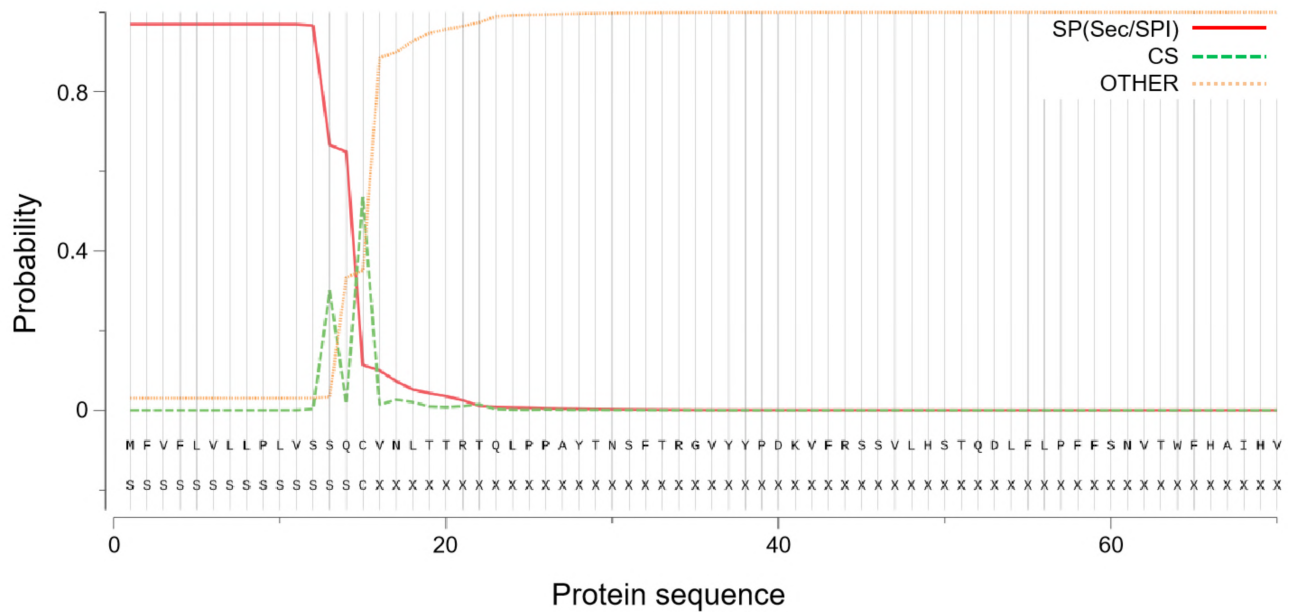


图5 新冠病毒(SARS-CoV-2)S蛋白的信号肽预测

Figure 5 Signal-peptide prediction of the spike glycoprotein of SARS-CoV-2

表2 新冠病毒(SARS-CoV-2)S蛋白的突变位点

Table 2 Mutation sites of the spike glycoprotein of SARS-CoV-2

Mutation type	Mutation sites	Impact on functions	Length
Point mutation	N234Q, L452R, A475V, V483A	Increased resistance to neutralizing antibodies ^[15]	1
Point mutation	N331Q, N343Q	Reduced viral infectivity ^[15]	1
Point mutation	F490L	Increased resistance to neutralizing antibodies and human covalent sera neutralization ^[15]	1
Point mutation	Q493N/Y, N501T	Reduced host ACE2-binding affinity <i>in vitro</i> ^[16]	1
Point mutation	H519P	Increased resistance to human covalent sera neutralization ^[15]	1
Point mutation	D614G	Increase of pseudotyped VSV particle production <i>ex vivo</i> ^[17, 18]	1
Point mutation	679 ~ 684(NSPRRA- IL)	About 50% stronger entry efficiency in Vero E6 cells while 30% weaker entry efficiency in hACE2-expressing BHK cells ^[9]	6
Point mutation	681 ~ 684(PRRRA-RRRK)	Optimized cleavage by host furin ^[19]	4
Deletion	682 ~ 684(RRA)	Complete loss of cleavage by host furin ^[19]	3

N616、N717、N801、N1098、N1134和N1194(图7),3个潜在的O-糖基化位点,分别是S673、T678和S686。

9 S蛋白结构域分析

结构域分析显示,S蛋白包含2个有意义的保守结构域,分别是刺突受体结合结构域(Spike receptor binding domain)和冠状病毒S2糖蛋白(Coronavirus S2 glycoprotein)结构域(图8),前者位

于P330-E583,后者位于C671-L1270。刺突受体结合结构域(S1亚基)的HMM为212,E-value为6.6e-75,Bit score为251.5;冠状病毒S2糖蛋白(S2亚基)的HMM为605,E-value为1.4e-266,Bit score为885.9。

10 S蛋白相互作用分析

相互作用分析显示,S蛋白与19个人源或鼠源蛋白存在二元相互作用(图9),分别是它本身

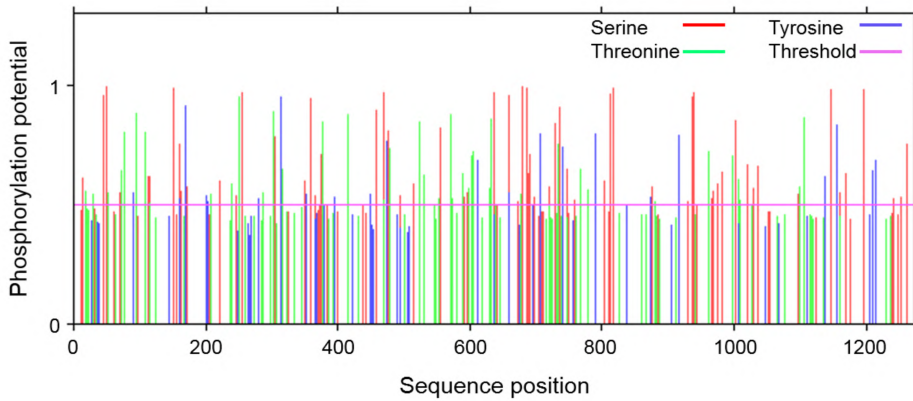


图 6 新冠病毒(SARS-CoV-2)S蛋白的磷酸化位点

Figure 6 Phosphorylation sites of the spike glycoprotein of SARS-CoV-2

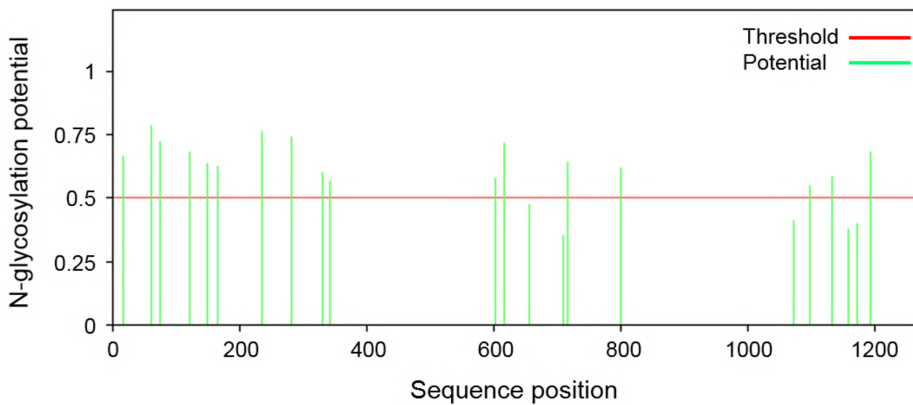


图 7 新冠病毒(SARS-CoV-2)S蛋白的N-糖基化位点预测

Figure 7 Prediction of the N-glycosylation sites of the spike glycoprotein of SARS-CoV-2



图 8 新冠病毒(SARS-CoV-2)S蛋白的结构域分析

Figure 8 Domain analysis of the spike glycoprotein of SARS-CoV-2

(SPIKE_SARS2), 神经肽-1(NRP1), 血管紧张素转化酶2(ACE2), 糖基化跨膜蛋白Basigin(BASI), 肉碱O-棕榈酰转移酶1(CATL1), Furin蛋白酶(FURIN), 酪氨酸蛋白激酶受体(UFO), C型凝集素结构域家族10成员A(CLC10-2), 半乳糖凝集素-3(LEG3), 半乳糖凝集素-8(LEG8), 甲型肝炎病毒细胞受体1(HAVR1), C型凝集素结构域家族10成员A(CLC10), 唾液酸结合Ig样凝集素10(SIG10), 低密度脂蛋白受体(LDLR), 表皮生长因子受体(EGFR)。

巨噬细胞甘露糖受体1(MRC1), CD209抗原(CD209), 半乳糖凝集素-7以及C型凝集素结构域家族4成员M(CLC4M)。

11 S蛋白受体结合结构域分析

SARS-CoV-2 S蛋白的受体结合结构域(receptor-binding domain, RBD)位于L387-E516,共130个残基^[20]。本研究首先对SARS-CoV-2和SARS-CoV S蛋白的RBD进行了二级结构预测。结果显示,两种RBD的二级结构均以无规则卷曲为主,比例约为57%;其次是β折叠,SARS-CoV-2略高于SARS-CoV;接着是β转角,SARS-CoV略高于SARS-CoV-2;比例最低的是α螺旋,二者均不到4%。总体来看,SARS-CoV-2和SARS-CoV的RBD二级结构组成差异不大(图10)。

本研究对SARS-CoV-2与SARS-CoV的RBD做了序列比对分析(图11)。结果显示,SARS-CoV-

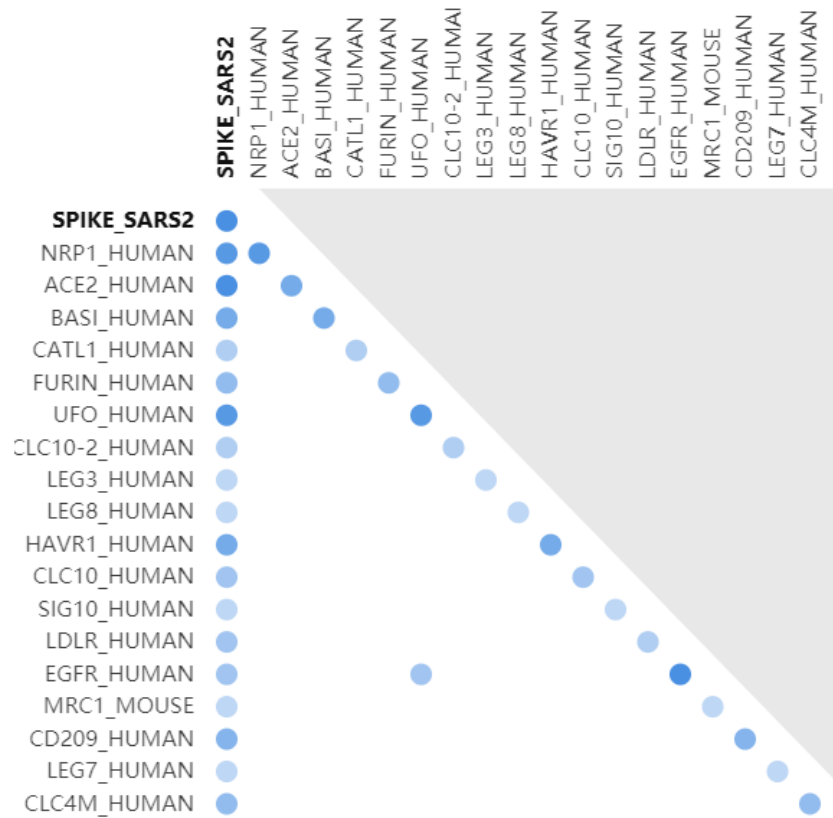


图9 新冠病毒(SARS-CoV-2)S蛋白的二元相互作用

Figure 9 Binary interaction of the spike glycoprotein of SARS-CoV-2

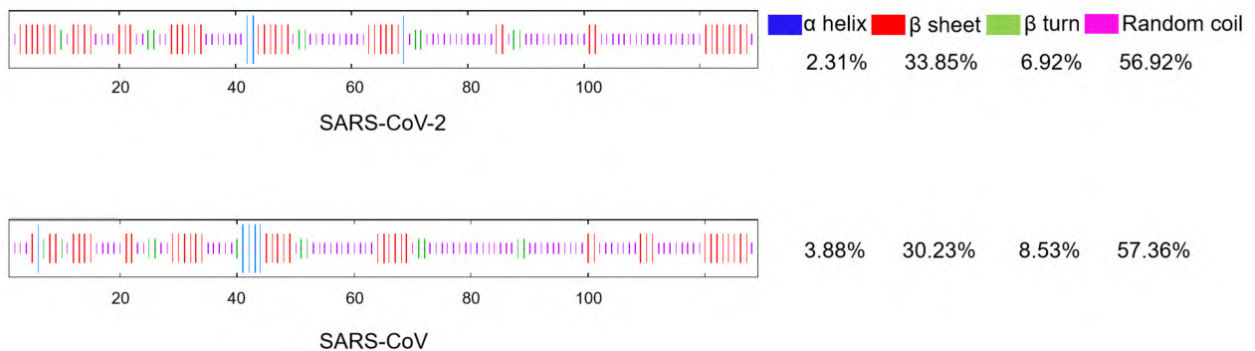


图10 SARS-CoV-2与SARS-CoV的RBD二级结构分析

Figure 10 Secondary structure of the receptor-binding domains of SARS-CoV-2 and SARS-CoV

2与SARS-CoV的RBD序列一致性较高,完全相同的氨基酸残基(*表示)有87个,占RBD残基总数的67%。

12 S蛋白序列同源性分析

利用UniProt网站的BLAST功能找到与SARS-CoV-2 S蛋白序列一致性较高(>70%)的12个病毒蛋白,分别是SARS冠状病毒S蛋白,蝙蝠冠状病毒279/2005 S蛋白,蝙蝠冠状病毒Rp32004 S蛋白,蝙蝠冠状病毒HKU3 S蛋白,蝙蝠非典型冠状病毒YNLF_31C S蛋白,SARS冠状病毒BtRf-

BetaCoV/JL2012 S蛋白,SARS冠状病毒BJ182-4 S蛋白,SARS冠状病毒Rs_672/2006 S蛋白,蝙蝠冠状病毒BM48-31/BGR/2008 S蛋白,SARS冠状病毒WH20 S蛋白,SARS冠状病毒TJF S蛋白,以及蝙蝠冠状病毒Rp/Shaanxi2011 S蛋白。多序列比对显示,S蛋白的氨基酸序列保守性很高,13个蛋白的氨基酸序列完全一致的残基(*表示)有748个,占58.8%,性质特别相近的氨基酸残基(:表示)有188个,占14.8%,性质微弱相近的氨基酸残基(.表示)有77个,占6.0%。与SARS-CoV-2 S蛋白序列

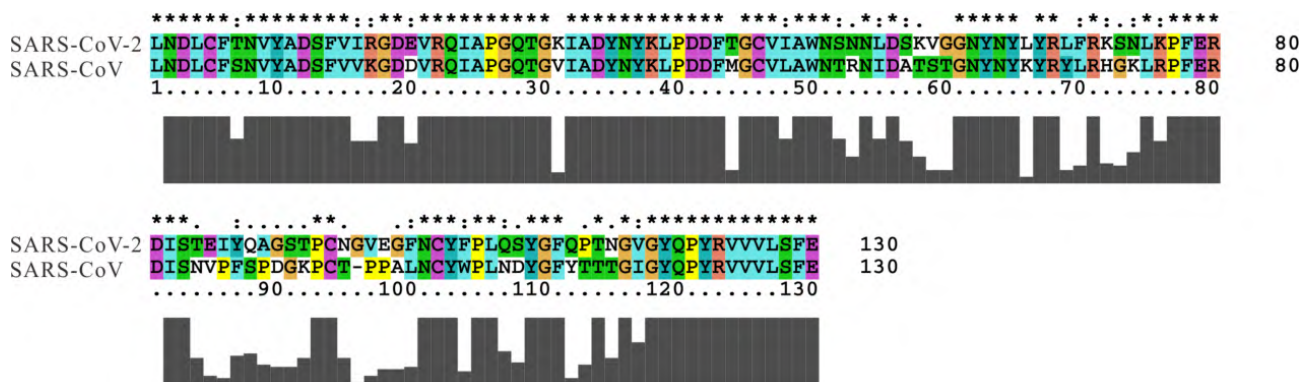


图 11 SARS-CoV-2与SARS-CoV的RBD序列分析

Figure 11 Sequence analysis of the receptor-binding domains of SARS-CoV-2 and SARS-CoV

一致性最高的是SARS冠状病毒、SARS冠状病毒WH20和蝙蝠冠状病毒HKU3,均为76%;其次是SARS冠状病毒TJF,序列一致性达75.9%,再次是SARS冠状病毒BJ182-4,序列一致性为75.8%(图12-1和图12-2)。

13 系统进化分析

使用MEGA7.0对包括SARS-CoV-2在内的13个病毒S蛋白进行多序列比对并构建系统进化树。结果显示,蝙蝠冠状病毒Rp32004、蝙蝠冠状病毒HKU3、蝙蝠冠状病毒279/2005、SARS冠状病毒Rs_672/2006、蝙蝠冠状病毒Rp/Shaanxi2011、蝙蝠非典型冠状病毒YNLF_31C和SARS冠状病毒BtRf-BetaCoV/JL2012聚为一支,置信度为100,说明它们在进化上相距较近,可能来源于同一祖先。SARS冠状病毒、SARS冠状病毒BJ182-4、SARS冠状病毒WH20和SARS冠状病毒TJF聚为一支,置信度为100,说明它们也可能来源于同一祖先。SARS-CoV-2与上述两支同属一大支,暗示它们可能具有共同的祖先。此外还发现,蝙蝠冠状病毒BM48-31/BGR/2008与上述病毒相距较远,单独聚为一支(图13)。

14 pET-22b-S原核表达载体构建

14.1 pET-22b空载体的检测及酶切验证

pET-22b空载体大小约5500 bp,载体图谱如图14A,对其进行NdeI或XhoI的单酶切及NdeI和XhoI的双酶切,然后用1.8%琼脂糖凝胶电泳检测酶切前、后的产物(图14B),大小均符合预期。

14.2 新冠病毒(SARS-CoV-2)S蛋白基因片段的酶切验证及菌落PCR

对新冠病毒(SARS-CoV-2)S蛋白基因片段进行NdeI和XhoI的双酶切验证,结果符合预期(图15A)。通过菌落PCR扩增S蛋白基因序列,用

1.8%琼脂糖凝胶电泳进行检测(图15B),大小也符合预期。

14.3 重组质粒pET-22b-S的酶切验证与转化

将重组质粒进行NdeI和XhoI的双酶切,用1.8%琼脂糖凝胶电泳进行检测,大小符合预期(图16A)。将重组质粒转化大肠杆菌感受态细胞Top10,涂布LB平板(Amp⁺),37℃培养12h(图16B)。挑取单克隆后抽提质粒,转化BL21感受态细胞,为蛋白表达做准备(图16C)。

15 S蛋白的诱导表达检测

挑取两个BL21单克隆(红色框和青色框),IPTG诱导后用12%的SDS-PAGE检测,结果显示:与诱导前相比,诱导后S的表达量明显升高(4泳道高于1泳道,10泳道高于7泳道)。为了优化S蛋白的纯化条件,本研究分别对诱导前、后全菌、上清和沉淀中S蛋白的表达情况进行了检测。结果显示(图17),S蛋白主要在沉淀中表达(3和6泳道,9和12泳道),因此在进行蛋白纯化时优先考虑从沉淀纯化。

讨论

新冠肺炎疫情是国际关注的公共卫生事件,SARS-CoV-2的出现和迅速蔓延正在危及全球人民生命健康和经济发展^[2,21]。研究发现,S蛋白以三聚体的形式排列在SARS-CoV-2膜表面,对病毒入侵宿主细胞至关重要^[12]。因此,S蛋白是疫苗、治疗性抗体和诊断的关键靶点^[12,22]。

理化性质分析显示,SARS-CoV-2 S蛋白由1273个氨基酸组成,分子量141.2 kD,等电点6.24,消光系数为148960 M⁻¹cm⁻¹,性质较稳定,疏水性较高。该蛋白含有两个卷曲螺旋,是一个分泌型跨膜

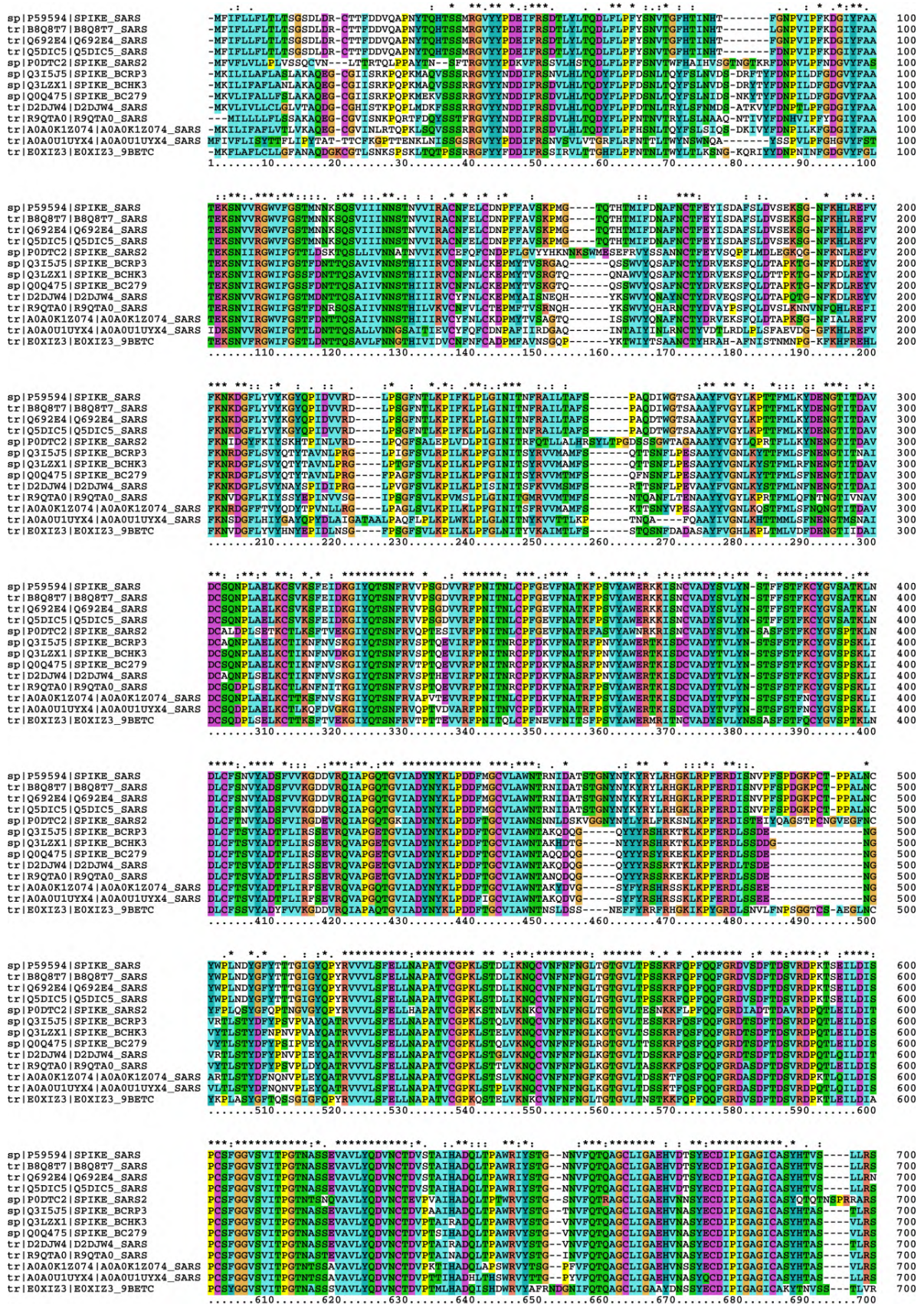


图 12-1 新冠病毒(SARS-CoV-2)与其它病毒S蛋白的多序列比对

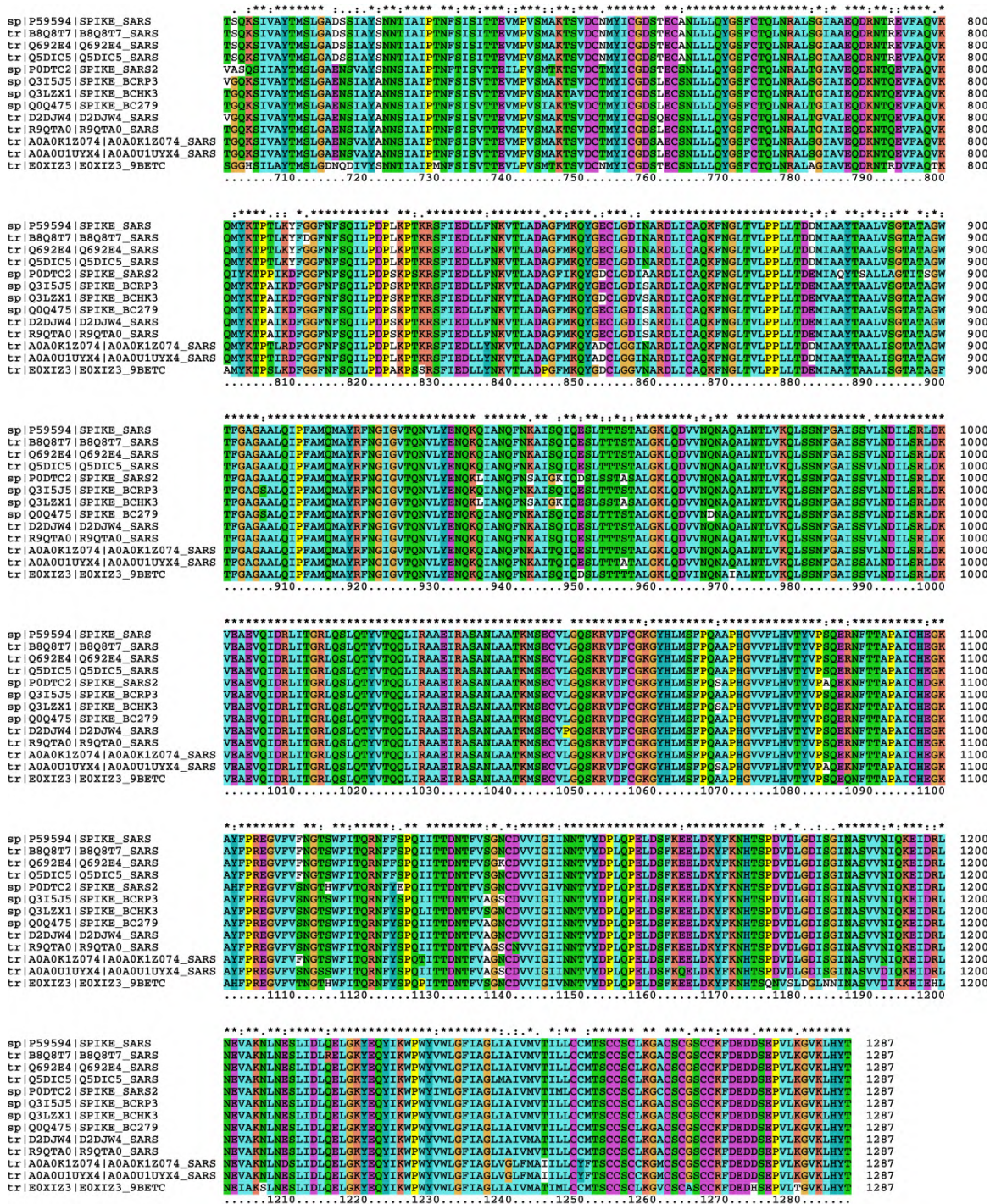


图 12-2 新冠病毒(SARS-CoV-2)与其它病毒S蛋白的多序列比对

注: spP59594: SARS 冠状病毒; trB8Q8T7: SARS 冠状病毒 BJ182-4; trQ692E4: SARS 冠状病毒 T1J; trQ5DIC5: SARS 冠状病毒 WH20; spP0DTC2: SARS 冠状病毒 2 型; spQ3I5J5: 蝙蝠冠状病毒 Rp32004; spQ3LZX1: 蝙蝠冠状病毒 HKU3; spQ0Q475: 蝙蝠冠状病毒 279/2005; trD2DJW4: SARS 冠状病毒 Rs_672/2006; trR9QTA0: 蝙蝠冠状病毒 Rp/Shaanxi2011; trA0A0K1Z074: 蝙蝠非典型冠状病毒 YNLF_31C; trA0A0UIUYX4: SARS 冠状病毒 BtRf-BetaCoV/JL2012; trE0XIZ3: 蝙蝠冠状病毒 BM48-31/BGR/2008

Figure 12 Multiple sequence alignment of the spike glycoprotein between SARS-CoV-2 and other viruses.

Notes: spP59594: SARS-CoV; trB8Q8T7: SARS-CoV BJ182-4; trQ692E4: SARS-CoV T1J; trQ5DIC5: SARS-CoV WH20; spP0DTC2: SARS-CoV-2; spQ3I5J5: Bat coronavirus Rp3/2004; spQ3LZX1: Bat coronavirus HKU3; spQ0Q475: Bat coronavirus 279/2005; trD2DJW4: SARS-CoV Rs_672/2006; trR9QTA0: Bat coronavirus Rp/Shaanxi2011; trA0A0K1Z074: Bat SARS-like coronavirus YNLF_31C; trA0A0UIUYX4: BtRf-BetaCoV/JL2012; trE0XIZ3: Bat coronavirus BM48-31/BGR/2008

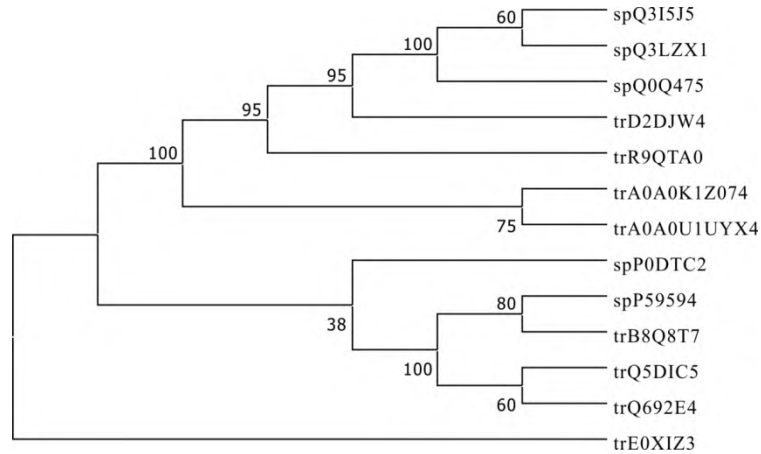


图13 基于新冠病毒(SARS-CoV-2)S蛋白序列构建的新冠病毒(SARS-CoV-2)与其它病毒的系统进化树

注:Entry number同图12

Figure 13 Phylogenetic tree of SARS-CoV-2 and other viruses based on the amino-acid sequence of spike glycoproteins

Notes:Entry numbers are the same as those in Figure 12

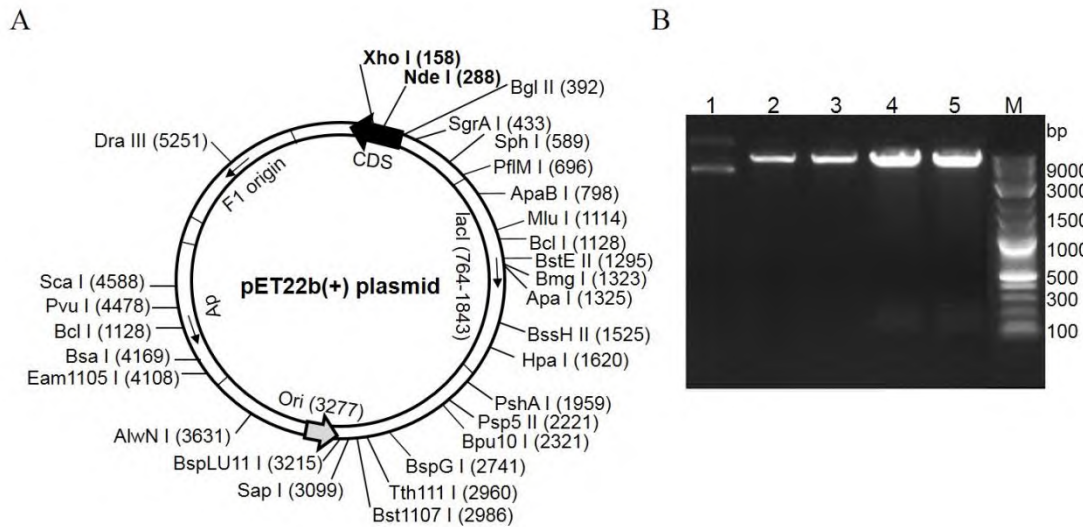


图14 pET-22b空载体检测及酶切验证

注:A. pET-22b的载体图谱;B. pET-22b空载体的酶切产物电泳检测;1:未酶切,2:*NdeI*单酶切,3:*XhoI*单酶切,4-5:*NdeI*和*XhoI*双酶切,M:2-log DNA marker

Figure 14 Detection and restriction endonuclease digestion of the pET-22b empty vector

Notes:A. Map of the pET-22b vector; B. Electrophoretic detection of the empty vector pET-22b by restriction endonuclease digestion; 1: No restriction endonuclease digestion. 2: *NdeI* digestion. 3: *XhoI* digestion. 4-5: *NdeI* and *XhoI* digestion. M: 2-log DNA marker

蛋白,主要分布于宿主细胞的内质网膜和细胞膜,有136个潜在的磷酸化位点,其中68个丝氨酸磷酸化位点,45个苏氨酸磷酸化位点,以及23个酪氨酸磷酸化位点。有报道称S蛋白的高度糖基化可使其结构更加稳定,增强了逃避机体免疫的能力^[23]。本研究结果显示,SARS-CoV-2 S蛋白具有17个潜在的N-糖基化位点和3个可能的O-糖基化位点,这与其较强的感染和传播能力相一致。

结构域分析显示,S蛋白包含2个保守结构域,

分别是刺突受体结合结构域(S1)和冠状病毒S2糖蛋白结构域(S2)。受体结合结构域(RBD)是S1亚单位中的关键功能成分,负责ACE2与SARS-CoV-2的结合^[9]。S蛋白三聚体的电镜结构揭示了RBD在开放和闭合状态下不同的构象状态和功能^[9,12]。我国科学家王新泉等解析了SARS-CoV-2 S蛋白的RBD与人受体ACE2复合物的晶体结构^[20],鉴定到S蛋白RBD和受体ACE2的相互作用位点,发现SARS-CoV-2 RBD的ACE2结合模式与SARS-

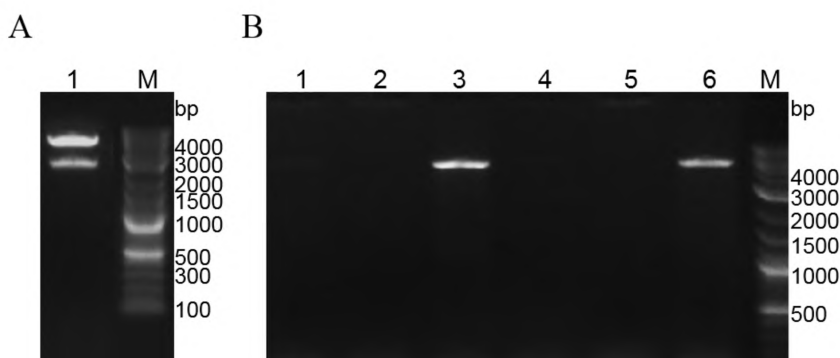


图 15 新冠病毒(SARS-CoV-2)S蛋白基因片段的酶切及菌落PCR验证
注:A. S蛋白基因的双酶切产物电泳检测;1:S蛋白基因的双酶切,M:2-log DNA marker;B. 菌落PCR验证;1-6:菌落PCR扩增产物,其中3和6泳道为阳性克隆

Figure 15 Restriction endonuclease digestion and colony PCR validation of S glycoprotein gene fragments

Notes:A. Detection of double-restriction endonuclease digestion products of the S glycoprotein coding gene by agarose gel electrophoresis; 1: double-restriction endonuclease digestion products of the S glycoprotein coding gene. M: 2-log DNA marker; B. Colony PCR validation; 1 - 6: Colony PCR products, among them, lane-3 and lane-6 were positive clones

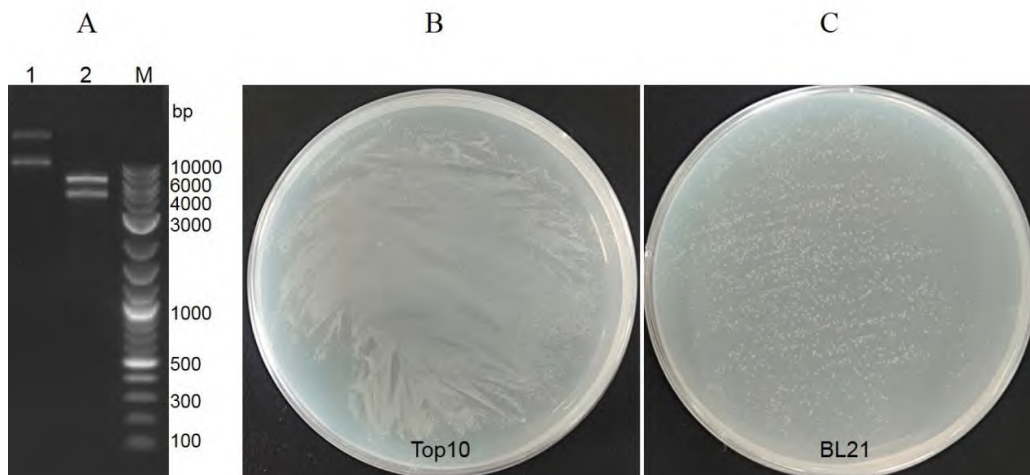


图 16 重组质粒 pET-22b-S 的酶切验证与转化

注:A. 重组质粒的双酶切验证;1:重组质粒未酶切,2:重组质粒双酶切,M:2-log DNA marker;B. 重组质粒转化 Top10感受态细胞;C. 重组质粒转化 BL21感受态细胞

Figure 16 Verification and transformation of recombinant plasmids

Notes:A. Verification of recombinant plasmid by double-restriction endonuclease digestion; 1: No digestion, 2: Double restriction endonuclease digestion, M: 2-log DNA marker; B. The recombinant plasmid was transformed into Top10 competent cells; C. The recombinant plasmid was transformed into BL21 competent cells

CoV 几乎相同, SARS-CoV-2 RBD 中参与结合 ACE2 的残基高度保守且与 SARS-CoV RBD 的残基具有相似的侧链性质。这种结构和序列上的相似性强烈表明 SARS-CoV-2 和 SARS-CoV RBDs 之间的趋同进化。该研究阐明了 S 蛋白介导 SARS-CoV-2 感染细胞的结构基础及分子机制, 为治疗性抗体药物开发以及疫苗的设计奠定了坚实基础。本研究对 SARS-CoV-2 和 SARS-CoV S 蛋白的 RBD

分析显示, 两种 RBD 的二级结构均以无规则卷曲为主, 其次是 β 折叠和 β 转角, 比例最低的是 α 螺旋。总体来看, SARS-CoV-2 和 SARS-CoV 的 RBD 二级结构组成差异不大, 序列一致性较高, 完全相同的氨基酸残基有 87 个, 占 RBD 残基总数的 67%, 这与它们识别相同的人体细胞表面受体相一致^[8, 24]。然而一些研究表明^[25], 抗 SARS-CoV-RBD 的小鼠单抗和多抗不能与 SARS-CoV-2 的 S 蛋白相互作用,

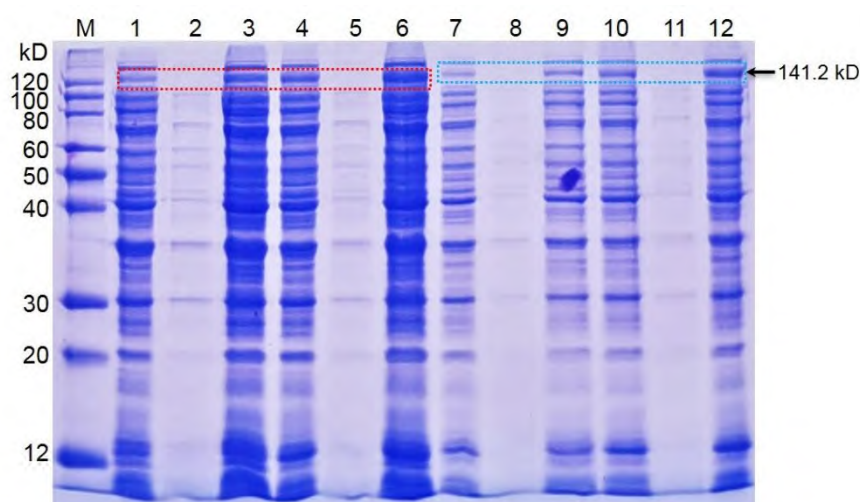


图 17 新冠病毒(SARS-CoV-2)S蛋白的诱导表达检测

注:M:蛋白分子量标准;1-3:诱导前全菌、上清、沉淀;4-6:诱导后全菌、上清、沉淀;7-9:诱导前全菌、上清、沉淀;10-12:诱导后全菌、上清、沉淀;红色框和青色框表示S蛋白(141.2 kD)的电泳条带位置

Figure 17 Detection of induced expression of the S glycoprotein

Notes:M: Protein marker; 1-3: Whole-cell lysate, centrifugation supernatant and centrifugation precipitation before IPTG induction; 4-6: Whole-cell lysate, centrifugation supernatant and centrifugation precipitation after IPTG induction; 7-9: Whole-cell lysate, centrifugation supernatant and centrifugation precipitation before IPTG induction; 10-12: Whole-cell lysate, centrifugation supernatant and centrifugation precipitation after IPTG induction; The red and cyan boxes indicate the electrophoretic band position of S protein (141.2 kD)

揭示了二者的抗原性差异。相互作用分析显示,S蛋白与神经肽-1,血管紧张素转化酶2,Furin蛋白酶等19个人源或鼠源蛋白存在二元相互作用。

多序列比对显示,S蛋白的氨基酸序列保守性较高,被分析的13个蛋白序列中完全一致的残基占58.8%,性质特别相近占14.8%。与SARS-CoV-2 S蛋白序列一致性最高的是蝙蝠冠状病毒 HKU3、SARS冠状病毒和 SARS冠状病毒 WH20(76%),其次是SARS冠状病毒 TjF(75.9%),再次是SARS冠状病毒 BJ182-4(75.8%)。进化分析显示,SARS-CoV-2与蝙蝠冠状病毒和SARS冠状病毒聚为一大支,暗示它们可能具有共同的祖先。值得注意的是,Zhou P等^[1]的报道称蝙蝠冠状病毒 RaTG13与SARS-CoV-2的亲缘关系最近,二者的S基因有93.1%以上的序列同源性。虽然本研究未对蝙蝠冠状病毒 RaTG13做进化分析,但得到了与该研究部分相似的结论。

本研究成功构建了SARS-CoV-2 S蛋白的原核表达载体,并检测了S蛋白的表达特性。结果显示,S蛋白主要在细菌裂解液离心之后的沉淀中表达,说明其可能容易形成包涵体,所以优先考虑从沉淀中去纯化。从沉淀中纯化的操作步骤相对繁琐,需保持蛋白的活性和不被降解。

综上所述,本研究对S蛋白进行了生物信息学

分析、过表达载体构建和蛋白诱导表达检测。研究结果为深入了解SARS-CoV-2 S蛋白的生物学功能、阐明S蛋白介导SARS-CoV-2入侵宿主细胞的分子机制奠定了基础,为靶向S蛋白的新型抗病毒药物筛选与应用提供了线索,同时为病毒-宿主相互作用理论提供新的见解。

参考文献:

- [1] Zhou P, Yang X, Wang X, Hu B, Zhang L, Zhang W, Si H, Zhu Y, Li B, Huang C, Chen H, Chen J, Luo Y, Guo H, Jiang R, Liu M, Chen Y, Shen X, Wang X, Zheng X, Zhao K, Chen Q, Deng F, Liu L, Yan B, Zhan F, Wang Y, Xiao G, Shi Z. A pneumonia outbreak associated with a new coronavirus of probable bat origin[J/OL]. *Nature*, 2020, 579(7798): 270-273. DOI: 10.1038/s41586-020-2012-7
- [2] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China[J/OL]. *Lancet*, 2020, 395(10223): 497-506. DOI: 10.1016/S0140-6736(20)30183-5
- [3] Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang

- B, Xiang H, Cheng Z, Xiong Y, Zhao Y, Li Y, Wang X, Peng Z. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China [J/OL]. *JAMA*, 2020, 323 (11) : 1061-1069. DOI: 10.1001/jama.2020.1585
- [4] Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, Xia J, Yu T, Zhang X, Zhang L. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study [J/OL]. *Lancet*, 2020, 395 (10223) : 507-513. DOI: 10.1016/S0140-6736(20)30211-7
- [5] Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y, Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, Chen W, Shi W, Tan W. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding [J/OL]. *Lancet*, 2020, 395 (10224) : 565-574. DOI: 10.1016/S0140-6736(20)30251-8
- [6] Chen L, Liu W, Zhang Q, Xu K, Ye G, Wu W, Sun Z, Liu F, Wu K, Zhong B, Mei Y, Zhang W, Chen Y, Li Y, Shi M, Lan K, Liu Y. RNA based mNGS approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak [J/OL]. *Emerg Microbes Infect*, 2020, 9 (1) : 313-319. DOI: 10.1080/22221751.2020.1725399
- [7] Xu X, Chen P, Wang J, Feng J, Zhou H, Li X, Zhong W, Hao P. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission [J/OL]. *Sci China Life Sci*, 2020, 63 (3) : 457-460. DOI: 10.1007/s11427-020-1637-5
- [8] Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens T S, Herrler G, Wu N H, Nitsche A, Müller M A, Drosten C, Pöhlmann S. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor [J/OL]. *Cell*, 2020, 181 (2) : 271-280. DOI: 10.1016/j.cell.2020.02.052
- [9] Walls A C, Park Y J, Tortorici M A, Wall A, McGuire A T, Velesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein [J/OL]. *Cell*, 2020, 181 (2) : 281-292. DOI: 10.1016/j.cell.2020.02.058
- [10] Lee N, Hui D, Wu A, Chan P, Cameron P, Joynt G M, Ahuja A, Yung M Y, Leung C B, To K F, Lui S F, Szeto C C, Chung S, Sung J J. A major outbreak of severe acute respiratory syndrome in Hong Kong [J/OL]. *N Engl J Med*, 2003, 348 (20) : 1986-1994. DOI: 10.1056/NEJMoa030685
- [11] Peiris J S, Lai S T, Poon L L, Guan Y, Yam L Y, Lim W, Nicholls J, Yee W K, Yan W W, Cheung M T, Cheng V C, Chan K H, Tsang D N, Yung R W, Ng T K, Yuen K Y, SARS study group. Coronavirus as a possible cause of severe acute respiratory syndrome [J/OL]. *Lancet*, 2003, 361 (9366) : 1319-1325. DOI: 10.1016/s0140-6736(03)13077-2
- [12] Wrapp D, Wang N, Corbett K S, Goldsmith J A, Hsieh C L, Abiona O, Graham B S, McLellan J S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation [J/OL]. *Science*, 2020, 367 (6483) : 1260-1263. DOI: 10.1126/science.abb2507
- [13] Chan J F, Kok K H, Zhu Z, Chu H, To K K, Yuan S, Yuen K Y. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan [J/OL]. *Emerg Microbes Infect*, 2020, 9 (1) : 221-236. DOI: 10.1080/22221751.2020.1719902
- [14] 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/OL]. *基因组学与应用生物学*, 2020, 39 (12) : 5606-5616. DOI: 10.13417/j.gab.039.005606
- [15] Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, Zhao C, Zhang Q, Liu H, Nie L, Qin H, Wang M, Lu Q, Li X, Sun Q, Liu J, Zhang L, Li X, Huang W, Wang Y. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity [J/OL]. *Cell*, 2020, 182 (5) : 1284-1294. DOI: 10.1016/j.cell.2020.07.012
- [16] Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, Geng Q, Auerbach A, Li F. Structural basis of receptor recognition by SARS-CoV-2 [J/OL]. *Nature*, 2020, 581 (7807) : 221-224. DOI: 10.1038/s41586-020-2179-y
- [17] Korber B, Fischer W M, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi E E, Bhattacharya T, Foley B, Hastie K M, Parker M D, Partridge D G, Evans C M, Freeman T M, de Silva T I, Sheffield COVID-19 Genomics Group, McDanal C, Perez L G, Tang H, Moon-Walker A, Whelan S P, LaBranche C C, Saphire E O, Montefiori D C. Tracking changes in SARS-CoV-2 spike: Evidence that D614G increases infectivity of the COVID-19 virus [J/OL]. *Cell*, 2020, 182 (4) : 812-827. DOI: 10.1016/j.cell.2020.06.043

- [18] Isabel S, Graña-Miraglia L, Gutierrez J M, Bundalovic-Torma C, Groves H E, Isabel M R, Eshaghi A, Patel S N, Gubbay J B, Poutanen T, Guttman D S, Poutanen S M. Evolutionary and structural analyses of SARS - CoV - 2 D614G spike protein mutation now documented worldwide[J/OL]. *Sci Rep*, 2020, 10(1): 14031. DOI: 10.1038/s41598-020-70827-z
- [19] Hoffmann M, Kleine - Weber H, Pöhlmann S. A multibasic cleavage site in the spike protein of SARS - CoV - 2 is essential for infection of human lung cells[J/OL]. *Mol Cell*, 2020, 78(4): 779-784. DOI: 10.1016/j.molcel.2020.04.022
- [20] Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, Zhang Q, Shi X, Wang Q, Zhang L, Wang X. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor[J/OL]. *Nature*, 2020, 581(7807): 215-220. DOI: 10.1038/s41586-020-2180-5
- [21] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung K S M, Lau E H Y, Wong J Y, Xing X, Xiang N, Wu Y, Li C, Chen Q, Li D, Liu T, Zhao J, Liu M, Tu W, Chen C, Jin L, Yang R, Wang Q, Zhou S, Wang R, Liu H, Luo Y, Liu Y, Shao G, Li H, Tao Z, Yang Y, Deng Z, Liu B, Ma Z, Zhang Y, Shi G, Lam T T Y, Wu J T, Gao G F, Cowling B J, Yang B, Leung G M, Feng Z. Early transmission dynamics in Wuhan, China, of novel coronavirus - infected pneumonia[J/OL]. *N Engl J Med*, 2020, 382(13): 1199-1207. DOI: 10.1056/NEJMoa2001316
- [22] Yuan M, Wu N C, Zhu X, Lee C D, So R T Y, Lv H, Mok C K P, Wilson I A. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV[J/OL]. *Science*, 2020, 368(6491): 630-633. DOI: 10.1126/science.abb7269
- [23] Walls A C, Xiong X, Park Y J, Tortorici M A, Snijder J, Quispe J, Cameroni E, Gopal R, Dai M, Lanzavecchia A, Zambon M, Rey F A, Corti D, Veesler D. Unexpected receptor functional mimicry elucidates activation of coronavirus fusion[J/OL]. *Cell*, 2019, 176(5): 1026 - 1039. DOI: 10.1016/j.cell.2018.12.028
- [24] Yuan Y, Cao D, Zhang Y, Ma J, Qi J, Wang Q, Lu G, Wu Y, Yan J, Shi Y, Zhang X, Gao G F. Cryo-EM structures of MERS - CoV and SARS - CoV spike glycoproteins reveal the dynamic receptor binding domains[J/OL]. *Nat Commun*, 2017, 8: 15092. DOI: 10.1038/ncomms15092
- [25] Wang Q, Zhang Y, Wu L, Niu S, Song C, Zhang Z, Lu G, Qiao C, Hu Y, Yuen K Y, Wang Q, Zhou H, Yan J, Qi J. Structural and functional basis of SARS-CoV - 2 entry by using human ACE2 [J/OL]. *Cell*, 2020, 181(4): 894 - 904. DOI: 10.1016/j.cell.2020.03.045

Bioinformatics Analysis on Structure and Function of the Spike Glycoprotein of SARS-CoV-2 and Its Prokaryotic Expression

XU Benjin^{1#*}, XUAN Yan^{1#}, DU Miao^{1#}, LIU Ling^{1*}, WU Huiwen², SONG Binyu², HOU Yanxiang¹,
CHEN Lirong¹, LIU Xiaoliang²

(1. Department of Medical Laboratory, Fenyang College of Shanxi Medical University, Fenyang 032200, China; 2. Science and Technology Centre, Fenyang College of Shanxi Medical University, Fenyang 032200, China)

Abstract: Severe acute respiratory syndrome-coronavirus 2 (SARS-CoV-2) is a highly pathogenic and rapidly spreading pathogen. It can invade and infect cells by recognizing receptors on the surface of host cells with spike (S) glycoprotein. Systematic bioinformatics analysis and prokaryotic expression of the S protein can aid understanding of its function and clarify the molecular mechanism of viral infection mediated by this protein. The physicochemical properties, subcellular localization, post-translational modifications and protein-interaction network of the S protein were analyzed systematically using Protparam, Pfam, TMHMM, ExPASy-ProtScale, PSORT II, SignalP, UniProt, NetPhos 3.1, NetNGlyc 1.0, NetOGlyc 4.0, BLAST and other bioinformatics software and databases. In addition, Clustal X2 and MEGA7.0 were used to analyze the homology and phylogeny of S glycoproteins based on amino-acid sequences. Finally, recombinant expression vector of the S glycoprotein was constructed by molecular cloning technology and expressed in *Escherichia coli*. Results showed that the S glycoprotein is composed of 1,273 amino acids, with a molecular weight of 141.2 kD and an isoelectric point of 6.24. It had two coiled helical structures and one transmembrane helix region. It was a hydrophobic protein, contained a spike receptor-binding domain and S2 glycoprotein domain. This protein was distributed mainly in the endoplasmic-reticulum membrane (39.1%) and cell membrane (21.7%) of host cells, and contained 136 potential phosphorylation sites and 20 possible glycosylation sites. SARS-CoV, SARS-CoV WH20 and bat coronavirus HKU3 showed the highest sequence identity with the spike-glycoprotein sequence of SARS-CoV-2 (76%). SARS-CoV-2, SARS-CoV and bat coronavirus clustered into a large branch, suggesting that they may have a common ancestor. The S protein was expressed mainly in the precipitate after centrifugation of bacterial lysates, which lays a foundation for future structural analysis and vaccine development. The S glycoprotein was highly conserved between SARS-CoV and bat coronavirus, suggesting that this glycoprotein has a vital role in viral invasion into host cells. SARS-CoV-2 may have a common ancestor with SARS-CoV and bat coronavirus. Our study provides an important data basis for expression, purification, structural and functional analysis of the S glycoprotein of SARS-CoV-2. Our data may help to reveal the biological functions of the S glycoprotein, and provide a scientific basis for the design and screening of new antiviral drugs targeting this protein.

Key words: SARS-CoV-2; Spike glycoprotein; Bioinformatics analysis; Prokaryotic expression

Funding: The present work was supported by the Starting Fund for Talent Introduction of Fenyang College of Shanxi Medical University (No. 2020A01), title: Study on the molecular mechanism of RelA mediated *Escherichia coli* stringent response; Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (No. 2020L0749), title: Functional study on ribosome recycling factor of *Escherichia coli*; National College Students' Innovation and Entrepreneurship Training Program (No. 202117114001), title: Cloning and expression vector construction of SARS-CoV-2 Nsp1 gene and its effect on the growth of *Escherichia coli*; Key Projects of Innovation and Entrepreneurship Training Program for College Students in Shanxi Province (No. S202117114008), title: Molecular mechanism of SARS-CoV-2 nucleocapsid protein overexpression affecting bacterial protein translation; Project of Luliang City Science and Technology Program (No. 2020SHFZ29), title: Research on the genetic polymorphism and traceability of food borne super resistant bacteria MRSA in Luliang City; (No. 2020XGZX104), title: Construction and evaluation of novel coronavirus pneumonia outcome prediction and risk assessment models.

* Corresponding author: LIU Ling, E-mail: ll772x@sxmu.edu.cn; XU Benjin, E-mail: benjin_ibprnalab@sina.com