

DOI:10.16781/j.0258-879x.2020.05.0465

• 专题报道 •

严重急性呼吸综合征冠状病毒 2 主要蛋白质分子的氨基酸变异分析

秦照玲¹, 罗利^{2*}

1. 海军军医大学(第二军医大学)海军医学系生物医学防护教研室, 上海 200433

2. 上海大学生命科学学院, 上海 200444

[摘要] **目的** 应用生物物理学方法鉴定分析严重急性呼吸综合征冠状病毒 2 (SARS-CoV-2) 中主要蛋白质分子的关键氨基酸变异。**方法** 通过氨基酸序列同源比对、突变氨基酸残基分类、蛋白质三维结构重建和氨基酸残基静电相互作用测量, 以同源性最高的蝙蝠冠状病毒 RaTG13 为参照, 进行 SARS-CoV-2 中主要蛋白质分子的关键氨基酸变异分析。**结果** 初步分析确定 SARS-CoV-2 中 RNA 依赖的 RNA 聚合酶 (RdRp)、核糖核酸外切酶 (ExoN)、尿苷酸特异性核糖核酸内切酶 (NendoU) 和刺突蛋白 (S 蛋白) 上至少发生了 10 处影响静电相互作用的氨基酸变异, 这些变异可能影响蛋白质分子的空间构象及其生物学功能。**结论** 初步确定了 SARS-CoV-2 中主要蛋白质分子的关键氨基酸变异, 为理解 SARS-CoV-2 的遗传特性、致病性和流行病学特征提供了有用线索。

[关键词] 严重急性呼吸综合征冠状病毒 2; 蛋白质类; 氨基酸变异; 蛋白质构象; 盐键; 斥力

[中图分类号] R 373.1 **[文献标志码]** A **[文章编号]** 0258-879X(2020)05-0465-09

Analysis of amino acid variations of major proteins from severe acute respiratory syndrome coronavirus 2

QIN Zhao-ling¹, LUO Li^{2*}

1. Department of Biomedical Defense, Faculty of Naval Medicine, Naval Medical University (Second Military Medical University), Shanghai 200433, China

2. School of Life Sciences, Shanghai University, Shanghai 200444, China

[Abstract] **Objective** To identify key amino acid variations of major proteins from severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by biophysical methods. **Methods** Through amino acid sequence alignment, classification of variant amino acid residues, three-dimensional structure reconstruction of proteins, and electrostatic interaction analysis of amino acid residues, the key amino acid variations of major proteins from SARS-CoV-2 was analyzed with RaTG13, the bat coronavirus with the highest homology, as the reference. **Results** At least ten amino acid variations that affect the possible electrostatic interactions were identified in RNA-dependent RNA polymerase (RdRp), exoribonuclease (ExoN), uridylyte-specific endoribonuclease (NendoU), and spike (S) protein from SARS-CoV-2. These variations may affect the spatial conformation and biological functions of the proteins. **Conclusion** The key amino acid variations of the major proteins from SARS-CoV-2 have been preliminarily identified, providing clues for understanding the genetic, pathogenic and epidemiological characteristics of the virus.

[Key words] severe acute respiratory syndrome coronavirus 2; proteins; amino acid variations; protein conformation; salt bond; repulsion

[Acad J Sec Mil Med Univ, 2020, 41(5): 465-473]

2020年3月11日(当地时间), WHO将新型冠状病毒肺炎(coronavirus disease 2019, COVID-19)列为全球性大流行病, 目前全球已有200多个国家和地区先后暴发了COVID-19疫情。截至2020年4月15日, COVID-19已造成全球180多万人感染, 死亡11万多例^[1]。在COVID-19

疫情出现初期, 研究人员应用高通量测序技术迅速鉴定出引发此次疫情的病原体为一种新型冠状病毒, 它与严重急性呼吸综合征冠状病毒(severe acute respiratory syndrome coronavirus, SARS-CoV)的基因组序列具有约80%的同源性, 随后成功分离出该病毒^[2-3]。国际病毒分类委员会将该病

[收稿日期] 2020-03-25

[接受日期] 2020-04-19

[作者简介] 秦照玲, 博士, 讲师. E-mail: zhlingqin@smmu.edu.cn

*通信作者(Corresponding author). Tel: 021-61335321, E-mail: luoli@shu.edu.cn

毒命名为严重急性呼吸综合征冠状病毒 2 (severe acute respiratory syndrome coronavirus 2, SARS-CoV-2)。研究发现 SARS-CoV-2 具有很强的传染性,致病性似乎不及 SARS-CoV,病毒基因组经多代传播后仍保持了较高的稳定性,变异较慢^[4,7]。这些遗传特性暗示 SARS-CoV-2 与它的祖先病毒相比发生了一些重要的遗传变异,但究竟是哪些病毒蛋白发生了较大变异,变异蛋白又涉及哪些关键的氨基酸残基,目前除了刺突蛋白 (spike protein, S 蛋白) 外,其他蛋白则缺少相关研究报道。

已有研究指出 SARS-CoV-2 与蝙蝠冠状病毒 RaTG13 (bat CoV RaTG13) 的基因组序列同源性最高 (达 96.2%)^[3],与穿山甲冠状病毒基因组同源性也达到 90%^[8-9]。与 SARS-CoV 相似,这些冠状病毒的基因组构成高度保守,编码相似的结构和非结构蛋白 (non-structural protein, Nsp),包括开放阅读框 (open reading frame, Orf) 1ab [编码木瓜样蛋白酶 (papain-like protease, PL^{pro})、3C 样蛋白酶 (3C-like protease, 3CL^{pro})、RNA 依赖的 RNA 聚合酶 (RNA-dependent RNA polymerase, RdRp)、Nsp7、Nsp8、Nsp9、核糖核酸外切酶 (exoribonuclease, ExoN)、尿苷酸特异性核糖核酸内切酶 (uridylylate-specific endoribonuclease, NendoU) 和 Nsp13 等 Nsp, 主要与病毒 RNA 合成相关]、S 蛋白、Orf3a、包膜蛋白 (envelope protein, E 蛋白)、膜蛋白 (membrane protein, M 蛋白)、Orf6、Orf7、Orf8 和核衣壳蛋白 (nucleocapsid protein, N 蛋白) 等^[10-11]。位于冠状病毒颗粒表面的 S 蛋白含有受体结合区 (receptor-binding domain, RBD), 主要负责与靶细胞表面的受体结合及随后的膜融合,是病毒入侵宿主细胞的关键蛋白^[12]。研究显示,与 SARS-CoV S 蛋白相比, SARS-CoV-2 S 蛋白的 RBD 发生了较大变异,包括受体结合位点的氨基酸突变及 4 个小肽片段插入,但是这些变异并不影响受体结合构象的形成^[13]。与 bat CoV RaTG13 S 蛋白相比, SARS-CoV-2 S 蛋白的 RBD 发生了 19 个氨基酸残基的替换,但这些氨基酸残基变异的重要程度并不清楚^[2]。与某些穿山甲冠状病毒的 S 蛋白相比, SARS-CoV-2 S 蛋白的 RBD 高度保守,提示 SARS-CoV-2 S 蛋白的 RBD 可能来自于某些穿山甲冠状病毒^[9,14]。冷冻电镜研究显示, SARS-CoV-2 的 S 蛋白由于 PRRA 序列的

插入,导致形成一个新的 RRAR 蛋白酶剪切位点,并在三维结构上形成一个新的结构单元,极大提高了其与受体的结合能力,这也可能是 SARS-CoV-2 具有较强传染性的原因^[15]。

由于多数 SARS-CoV 编码蛋白的三维结构已被解析,有了较多、较好的研究数据与分子模型,而且 SARS-CoV-2 中主要蛋白质分子与 SARS-CoV 编码蛋白具有较高的序列保守性,因此本研究应用蛋白质三维结构重建技术,分析 SARS-CoV-2 中主要蛋白质分子变异氨基酸残基电荷的相互作用变化,从而确定影响病毒蛋白构象和功能的特定氨基酸残基,以期为更好地理解 SARS-CoV-2 的遗传特性提供线索,也为后续临床治疗药物的研发提供理论参考。

1 材料和方法

1.1 蛋白质氨基酸序列同源性分析 所有的冠状病毒基因组序列均来自 GenBank [美国国立生物技术信息中心 (National Center for Biotechnology Information, NCBI)], Orf1ab 包含的功能蛋白应用 CD Search (NCBI) 进行确认分析,采用 NCBI 的 BLAST、COBALT 软件和欧洲生物信息研究所 (European Bioinformatics Institute, EBI) 的 Clustal Omega 软件进行序列同源性分析。本研究使用的基因组序列包括 SARS-CoV CV7 (登录号: DQ898174)、SARS-CoV-2 (登录号: MN908947) 和 bat CoV RaTG13 (登录号: MN996532)。

1.2 蛋白质三维结构重建 SARS-CoV 蛋白质三维结构模板均来自 NCBI 的蛋白质数据库 (Protein Database, PDB), SARS-CoV-2 和 bat CoV RaTG13 主要蛋白质分子三维结构重建均在 SWISS-MODEL 服务器上完成。蛋白质三维结构重建的主要参数: GMQE (global model quality estimation) 和 QMEAN (qualitative model energy analysis)。GMQE 为整体模型质量评估值,合理范围为 0~1; QMEAN 为模型能量值定量分析值,合理值在 0 附近^[16]。

1.3 酸性与碱性氨基酸的电荷相互作用分析 应用 PyMOL 软件确定带电氨基酸残基侧链氨基和羧基的氮氧原子间的距离 (10 Å 以内),判断蛋白质中突变氨基酸及空间上邻近氨基酸残基的电荷相互作用强弱。异种电荷相互吸引,形成盐键;同种电荷相互排斥,产生斥力。在特定的 pH 范围内,酸性

氨基酸天冬氨酸(D)和谷氨酸(E)带负电,碱性氨基酸赖氨酸(K)和精氨酸(R)带正电,而组氨酸(H)质子化后也带正电。电荷之间的作用力与原子间距离的平方成反比。

2 结果和讨论

2.1 SARS-CoV-2 与 bat CoV RaTG13 中的蛋白质分子氨基酸变异分析 以 SARS-CoV CV7 株编码的主要蛋白作为出发序列,在 GenBank 数据库中进行同源比对,发现这些蛋白与蝙蝠冠状病毒(bat CoV RaTG13 除外)编码蛋白相比变异率为 0.2%~11.6%,其中 Orf8 的变异率最高,达到

11.6%。由于目前病毒学界已经确定 SARS-CoV 来源于中华菊头蝠冠状病毒,以此作为参考,对 SARS-CoV-2 编码的蛋白作进一步细致的变异率分析。我们发现这些蛋白与 SARS-CoV 编码蛋白相比,变异率为 6.6%~41.3%;而与蝙蝠冠状病毒(bat CoV RaTG13 除外)编码的蛋白相比,变异率降低,为 1.3%~19.7%。有趣的是,SARS-CoV-2 编码的蛋白与 bat CoV RaTG13 编码蛋白相似度较高,变异率均不超过 5.0%。而 bat CoV RaTG13 编码的蛋白与其他蝙蝠冠状病毒编码蛋白相比,变异率为 0.4%~19.8%。见图 1。

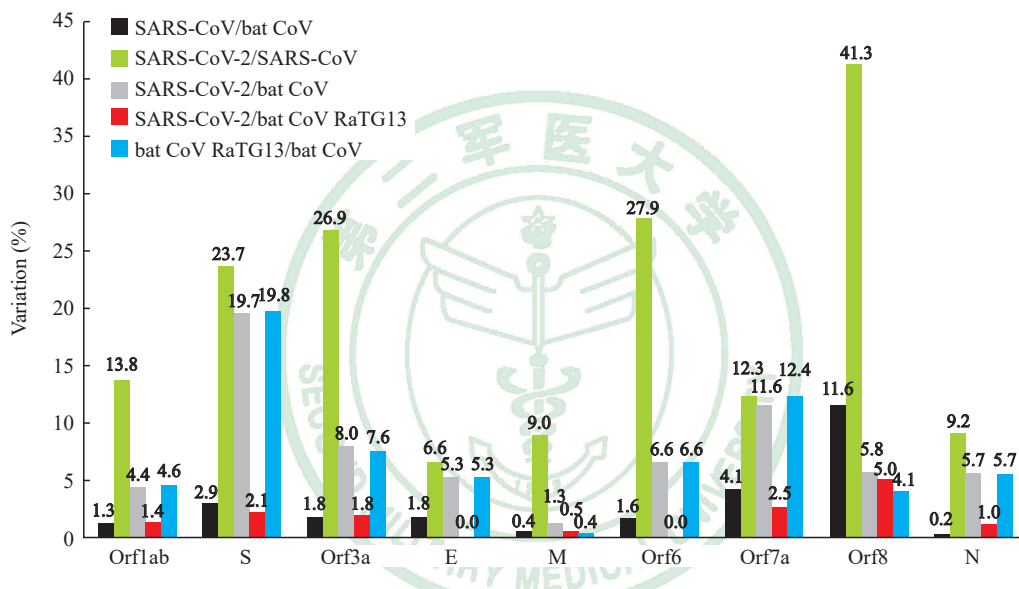


图 1 SARS-CoV、蝙蝠冠状病毒和 SARS-CoV-2 中主要蛋白质分子的氨基酸变异分析

Fig 1 Analysis of amino acid variations of major proteins from SARS-CoV, bat CoV and SARS-CoV-2

The numbers are percentage of the amino acid variations. SARS-CoV: Severe acute respiratory syndrome coronavirus; SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; bat CoV: Bat coronavirus; Orf: Open reading frame; S: Spike protein; E: Envelope protein; M: Membrane protein; N: Nucleocapsid protein

为进一步确认比对结果,我们对 SARS-CoV-2 和 bat CoV RaTG13 Orf1ab 编码的蛋白单独进行了氨基酸变异分析,发现这些蛋白与 bat CoV RaTG13 对应的蛋白相比,变异率不超过 10.29%;而且在 Orf1ab 上存在 2 个高度变异的区域(Nsp1-Macro, 变异率为 4.34%~10.29%; Nsp3-NAR, 变异率为 2.21%~3.08%),其中 DUF3655 变异率高达 10.29%,其余区域的变异率不超过 0.72% (图 2)。氨基酸序列比对分析显示 SARS-CoV-2 编码的蛋白与 bat CoV RaTG13 编码的蛋白高度同源,这与之前文献报道的关于 SARS-CoV-2 与

bat CoV RaTG13 基因组高度同源^[3]的观点一致。根据 SARS-CoV 编码蛋白的氨基酸变异率分析结果,结合现有的冠状病毒基因组数据,我们认为 bat CoV RaTG13 可能是 SARS-CoV-2 的祖先病毒。

为进一步分析确定编码蛋白中关键的氨基酸变异,我们将 SARS-CoV-2 和 bat CoV RaTG13 中主要蛋白质分子的氨基酸变异进行了更为细致的区分。将氨基酸错义突变区分为近似突变和非近似突变两类。所谓近似突变,即突变前后氨基酸的基本性质未变,蛋白质构象也不会发生明显变化,如电性(谷氨酸 E-天冬氨酸 D、精氨酸 R-赖氨酸 K)

或极性(天冬酰胺 N- 谷氨酰胺 Q, 丝氨酸 S- 苏氨酸 T- 半胱氨酸 C- 酪氨酸 Y, 苯丙氨酸 F- 异亮氨酸 I- 亮氨酸 L- 缬氨酸 V- 丙氨酸 A) 相同的氨基酸之间的替换, 也包括性质相似氨基酸之间的转换如苯丙氨酸 F- 酪氨酸 Y、苏氨酸 T- 丙氨酸 A 的替换。非近似突变即突变后蛋白质构象发生明显的变化, 功能出现异常, 如特殊性质氨基酸的插入或缺失或

不同电性的氨基酸的转换。根据这一界定, 重新统计 2 种病毒中的主要蛋白(共统计 28 个蛋白)的氨基酸非近似变异率不超过 5.88%, 其中 DUF3655 的非近似变异率最高, 为 5.88%; 共有 14 个蛋白没有发生非近似突变, 另有 11 个蛋白分子的非近似变异率低于 1.00% (图 3)。

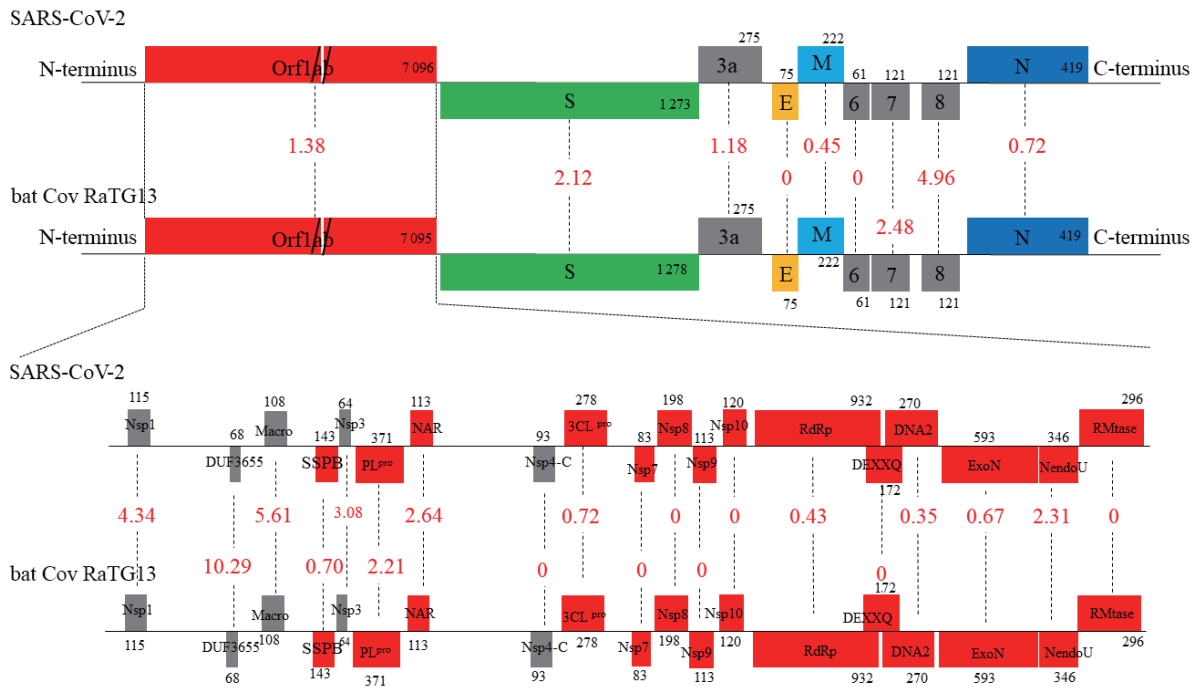


图 2 SARS-CoV-2 和 bat CoV RaTG13 中主要蛋白质分子的氨基酸变异图谱

Fig 2 Amino acid variation map of major proteins from SARS-CoV-2 and bat CoV RaTG13

Annotations of predicted proteins are mainly from the server of CD search in National Center for Biotechnology Information (NCBI). The red numbers are percentage of the amino acid variations. SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; bat CoV: Bat coronavirus; Orf: Open reading frame; S: Spike protein; E: Envelope protein; M: Membrane protein; N: Nucleocapsid protein; Nsp: Non-structural protein; DUF: Protein of unknown function; SSPB: Single-stranded poly(A)-binding domain; PL^{pro}: Papain-like protease; NAR: Nucleic acid-binding domain; 3CL^{pro}: 3C-like protease; RdRp: RNA-dependent RNA polymerase; DNA2: Superfamily I DNA and/or RNA helicase; DEXXQ: DEXXQ-box helicase domain of Upf1-like helicase; ExoN: Exoribonuclease; NendoU: Uridylate-specific endoribonuclease; Rmtase: Cap-0 specific (nucleoside-2'-O)-methyltransferase

2.2 基于 SARS-CoV-2 与 bat CoV RaTG13 中蛋白质分子三维结构的变异氨基酸电荷相互作用分析 为了更加准确地界定 SARS-CoV-2 中发挥关键作用的氨基酸残基, 我们重建了 SARS-CoV-2 和 bat CoV RaTG13 主要编码蛋白的三维结构(表 1)。这些蛋白三维结构的 GMQE 值为 0.73~0.98; 除了 ExoN 和 S 蛋白, 其他蛋白的 QMEAN 值均约为 0。尽管这 2 种病毒的 ExoN 和 S 蛋白的 QMEAN 值偏离了 0 点, 但是数值非常接近, 其结构具有可比性。

2.2.1 RNA 合成相关酶的氨基酸变异对电荷相互作用的影响 通过氨基酸序列比对分析, 发现负责 SARS-CoV-2 与 bat CoV RaTG13 基因组 RNA 合成的 Nsp 非常保守, 变异率较低或根本没有变异(图 3)。RdRp 在冠状病毒 RNA 复制过程中发挥了关键作用, 其催化效率及保真性直接关系到病毒 RNA 合成的数量和变异速度^[11]。SARS-CoV-2 与 bat CoV RaTG13 的 RdRp 蛋白中共有 4 个氨基酸残基发生变异, 其中 2 个位点属于近似变异(R-K、V-I), 而第 185 位和第 719 位的变异属于非近似变异。

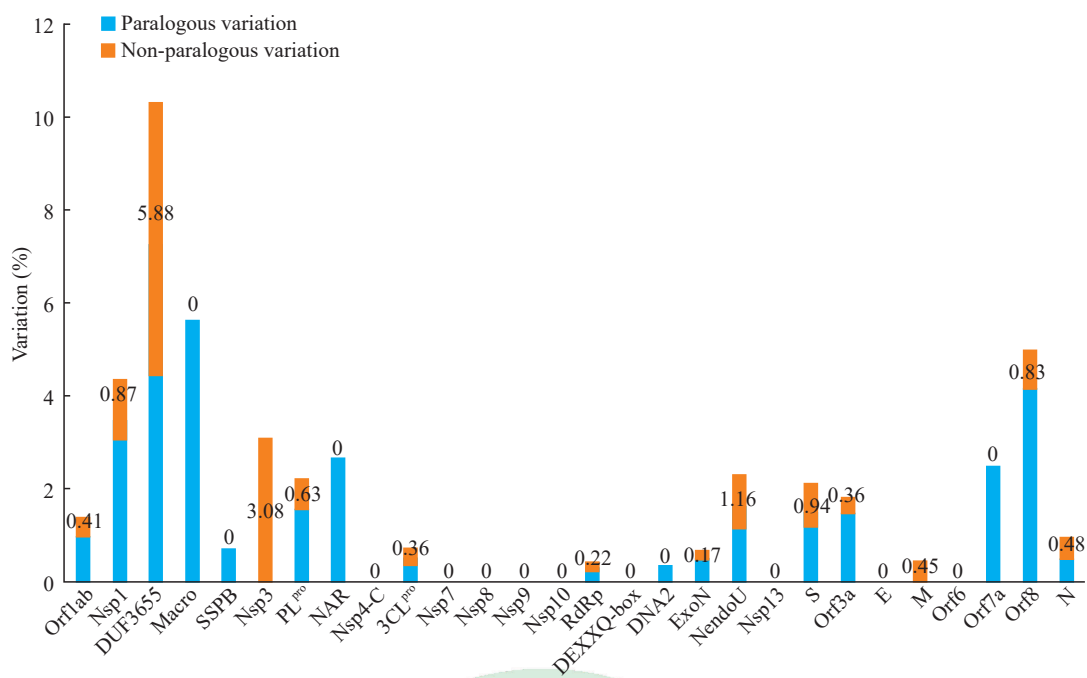


图 3 SARS-CoV-2 和 bat CoV RaTG13 中主要蛋白质分子的非近似氨基酸变异分析

Fig 3 Analysis of non-paralogous amino acid variations of major proteins from SARS-CoV-2 and bat CoV RaTG13

The numbers are percentage of the non-paralogous amino acid variations. SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; bat CoV: Bat coronavirus; Orf: Open reading frame; Nsp: Non-structural protein; DUF: Protein of unknown function; SSPB: Single-stranded poly(A)-binding domain; PL^{pro}: Papain-like protease; NAR: Nucleic acid-binding domain; 3CL^{pro}: 3C-like protease; RdRp: RNA-dependent RNA polymerase; DEXXQ-box: DEXXQ-box helicase domain of Upf1-like helicase; DNA2: Superfamily I DNA and/or RNA helicase; ExoN: Exoribonuclease; NendoU: Uridylate-specific endoribonuclease; S: Spike protein; E: Envelope protein; M: Membrane protein; N: Nucleocapsid protein

表 1 重建三维结构的 SARS-CoV-2 和 bat CoV RaTG13 中主要蛋白质分子

Tab 1 Major proteins with three-dimensional structure reconstruction from SARS-CoV-2 and bat CoV RaTG13

| Protein | Template | GMQE ^a | QMEAN ^b | Oligo state | Ligand |
|-------------------|----------|-------------------|--------------------|--------------|----------------------|
| SARS-CoV-2 | | | | | |
| PL ^{pro} | 5tl6.1.A | 0.95 | -0.14 | Monomer | 1 × Zn ²⁺ |
| Nsp3 | 2acf.1.A | 0.87 | -0.43 | Monomer | |
| RdRp | 6nur.1.A | 0.83 | -0.72 | Monomer | 1 × Zn ²⁺ |
| ExoN | 5nfy.1.A | 0.86 | -3.15 | Monomer | 2 × Zn ²⁺ |
| NendoU | 2h85.1.A | 0.98 | 0.13 | Homo-hexamer | |
| S | 6acc.1.A | 0.73 | -3.63 | Homo-trimer | |
| bat CoV RaTG13 | | | | | |
| PL ^{pro} | 5tl6.1.A | 0.95 | 0.01 | Monomer | 1 × Zn ²⁺ |
| Nsp3 | 2acf.1.A | 0.84 | -0.78 | Monomer | |
| RdRp | 6nur.1.A | 0.83 | -0.70 | Monomer | 1 × Zn ²⁺ |
| ExoN | 5nfy.1.A | 0.85 | -3.17 | Monomer | 2 × Zn ²⁺ |
| NendoU | 2h85.1.A | 0.98 | -0.06 | Homo-hexamer | |
| S | 6acc.1.A | 0.73 | -3.65 | Homo-trimer | |

^a: Confidence value is 0-1; ^b: Confidence value is approximately 0^[16]. SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; bat CoV: Bat coronavirus; PL^{pro}: Papain-like protease; Nsp3: Non-structural protein 3; RdRp: RNA-dependent RNA polymerase; ExoN: Exoribonuclease; NendoU: Uridylate-specific endoribonuclease; S: Spike protein; GMQE: Global model quality estimation; QMEAN: Qualitative model energy analysis

根据三维结构中侧链基团氮氧原子之间的距离测定,判断 bat CoV RaTG13 RdRp 蛋白氨基端 D185 与 R184 之间可以形成强盐键;而 SARS-CoV-2 中 N185 则无法形成盐键(图 4A~4C)。bat CoV RaTG13 RdRp 蛋白羧基端质子化的 H719 可以与 D717 形成一个盐键,而 SARS-CoV-2 中的 Y719 则无法与 D717 形成盐键(图 4D~4F)。因此,与 bat CoV RaTG13 相比, SARS-CoV-2 RdRp 蛋白靠近两端部分由于氨基酸变异各消除了一个盐键。而在 SARS-CoV RdRp 三维结构(PDB:6nur)中, D198(对应 bat CoV RaTG13 中的 D185)与 R197

形成一个强盐键;其 Y719(对应 SARS-CoV-2 中的 Y719)无法与 D717 形成盐键。有意思的是,在 bat CoV RaTG13 中, K268(对应 SARS-CoV 中的 C281)与 R272 通过正电排斥更靠近 E265,从而形成 K268-E265 强盐键(对应 SARS-CoV 中的 E278-R285 弱盐键)。因此,与 SARS-CoV 相比, SARS-CoV-2 RdRp 蛋白由于氨基酸变异在氨基端消除了一个盐键,再形成了一个强盐键。由此可见, SARS-CoV-2 RdRp 蛋白两末端部位的盐键发生了明显的变化,可能会影响其构象和催化活性。

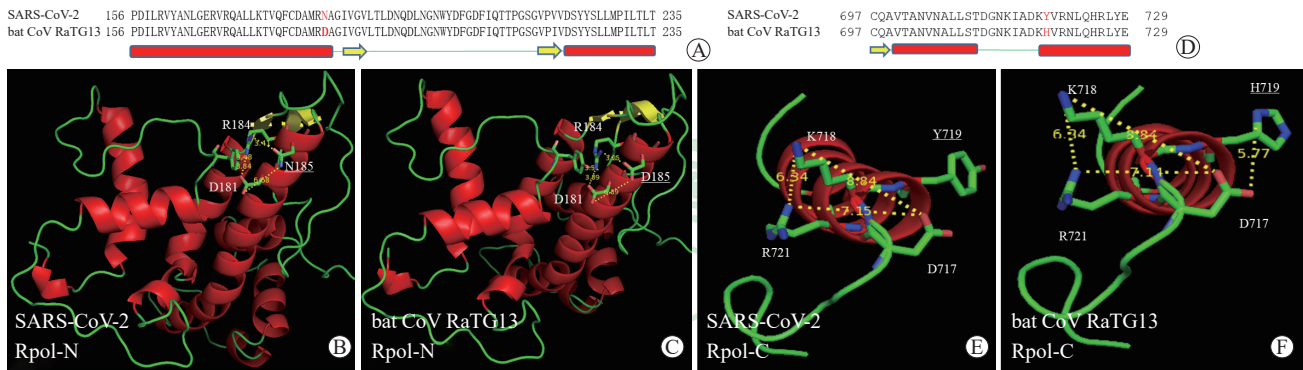


图 4 SARS-CoV-2 和 bat CoV RaTG13 RdRp 蛋白中关键氨基酸变异及电荷相互作用

Fig 4 Key amino acid variations and electrostatic interactions in RdRp from SARS-CoV-2 and bat CoV RaTG13

A, D: Amino acid sequences shown in the three-dimensional structure. The variant amino acid residues are in red. Green lines, yellow arrows and red bars indicate loops, β -folds and α -helices, respectively. B, C, E, F: Partial three-dimensional structure of RdRp. N185, D185, Y719 and H719 are the variant amino acid residues. SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; bat CoV: Bat coronavirus; RdRp: RNA-dependent RNA polymerase; Rpol-N: N-terminus of RdRp; Rpol-C: C-terminus of RdRp

通过比较 SARS-CoV-2 与 bat CoV RaTG13 中 ExoN 蛋白的氨基酸序列,发现有 4 个氨基酸残基出现变异,包括 3 个近似突变(V-I、I-V 和 R-K)和 1 个非近似突变(L-R)(图 5A)。根据三维结构中侧链基团氮氧原子间的距离判断:在 bat CoV RaTG13 ExoN 蛋白中,质子化的 H281 与 E282 形成一个较强的盐键,而 R257 所带的正电荷通过正电斥力使 H281 更靠近 E282,增强二者形成的盐键;而在 SARS-CoV-2 中, L257 取代了 R257,因此不会对 H281 和 E282 之间的盐键产生增强作用(图 5B、5C)。而在 SARS-CoV ExoN 蛋白结构(PDB:5nfy)中,对应的 H283-E284 盐键由于质子化的 H260 的正电斥力而加强。因此, SARS-CoV-2 ExoN 蛋白中的 E282 形成的盐键最弱,可能影响该酶的分子构象和生物化学功能。

通过分析 SARS-CoV-2 和 bat CoV RaTG13 中

NendoU 蛋白氨基酸序列,发现变异率高达 2.31% (图 2),非近似氨基酸变异率达到 1.16% (图 3),是该区域变异率最高的一个蛋白。在 SARS-CoV-2 中这些非近似的氨基酸变异包括 D128、D183、Q201 和 S243 等 4 个氨基酸残基(图 6A)。D128 通过负电斥力增强 D124 与 R126 形成的盐键, D183 与 R138 形成一个强盐键(图 6B)。而在 bat CoV RaTG13 中, N128 无法增强 D124 和 R126 之间的盐键, N183 也无法与 R138 形成盐键(图 6C)。在该病毒中 K201 与 E202 形成盐键, R243 通过正电斥力增强质子化的 H242 与 D239 之间的盐键(图 6C)。但是,在 SARS-CoV-2 中 Q201 不能与 E202 形成盐键, S243 也不能增强质子化的 H242 与 D239 之间的盐键(图 6B)。因此,与 bat CoV RaTG13 相比, SARS-CoV-2 NendoU 蛋白中的盐键发生了明显变化。在 SARS-CoV NendoU 蛋白

三维结构 (PDB:2h85) 中, 质子化的 H21 与 D18 形成强盐键 (在 SARS-CoV-2 和 bat CoV RaTG13 中则为 Q21, 无法形成盐键); D201、E203 和 D204 形成一个负电斥力环 (在 SARS-CoV-2 和 bat CoV RaTG13 中, 该区域只有 E202, 无法形成负电斥力环); K266 与 D241 形成强盐键 (在 SARS-CoV-2 和 bat CoV RaTG13 中, 对应的 E264、E266 与 D267 形成一个负电斥力环)。因此, 与 SARS-CoV 相比, SARS-CoV-2 NendoU 蛋白中电荷相互作用也发生了明显变化, 这可能影响其构象和生物学功能。

简而言之, 与 bat CoV RaTG13 相比, SARS-CoV-2 的 RdRp、ExoN 和 NendoU 3 个蛋白上的 7 处氨基酸发生变异 (表 2), 可能影响了这些 RNA 合成相关酶的构象和功能, 从而改变病毒 RNA 复制的速率和保真性。

SARS-CoV-2 249 GNLQSNHDL^YCQVHGNAHVASCD^AIMTRCLAVHECF 284
 bat CoV RaTG13 249 GNLQSNH^DRYCQVHGNAHVASCD^AIMTRCLAVHECF 284

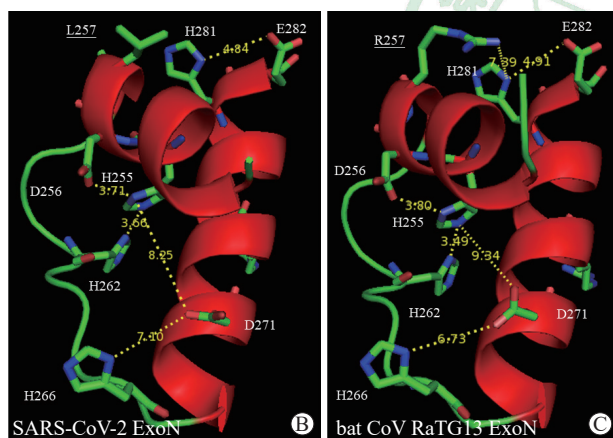


图 5 SARS-CoV-2 和 bat CoV RaTG13 ExoN 蛋白中关键氨基酸变异及电荷相互作用

Fig 5 Key amino acid variations and electrostatic interactions in ExoN from SARS-CoV-2 and bat CoV RaTG13
 A: Amino acid sequences shown in the three-dimensional structure. The variant amino acid residues are in red. Green lines and red bars indicate loops and α -helices, respectively. B, C: Partial three-dimensional structure of ExoN. L257 and R257 are the variant amino acid residues. SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; bat CoV: Bat coronavirus; ExoN: Exoribonuclease

2.2.2 S 蛋白的氨基酸变异对电荷相互作用的影响 S 蛋白是冠状病毒中最关键的结构蛋白之一, 由 S1 亚基和 S2 亚基组成。其中, S1 的羧基端含有一个 RBD, 可以与宿主细胞表面受体如血

管紧张素转化酶 2 (angiotensin-converting enzyme 2, ACE2) 结合, 从而改变蛋白构象, 诱导 S2 亚基中的疏水融合肽暴露, 介导病毒与宿主细胞发生膜融合^[12]。因此, S 蛋白是冠状病毒侵入宿主细胞的“钥匙”。在 bat CoV RaTG13 三聚化的 S 蛋白中, S1/S2 切割位点处的 E670 和 D672 之间由于负电斥力保持 8.2~8.4 Å 距离 (在 3 个单体中分别为 8.20、8.23 和 8.38 Å)。而在 SARS-CoV-2 S 蛋白中, S680 处 (位于 S2 亚基的氨基端) 插入了 PRRA 肽段, 这一插入导致 S1/S2 切割位点处的 E661 和 D663 之间的距离明显缩短为 6.8~7.0 Å (在 3 个单体中分别为 6.79、6.81 和 6.98 Å), 说明这 2 个氨基酸残基间的负电斥力明显增强, 从而影响该区域的蛋白质构象, 推测很可能会影响 S1 和 S2 亚基的剪切、ACE2 受体的识别与结合或膜融合。

SARS-CoV-2 128 NGQVDFLNARNGVLLI^TEGSVKGLQ^PSVGPQASLNGVTLI^GEAVKTFQFNYYKVDGVQQLPETVFTQSRNLQ^EFKPKRS 207
 bat CoV RaTG13 128 DGQVDFLNARNGVLLI^TEGSVKGLQ^PSVGPQASLNGVTLI^GEALKTQFNYYKRVNGVQQLPETVFTQSRNLQ^EFKPKRS 207
 SARS-CoV-2 208 QMEIDFLELAMDEFI^ERYKLEGYAF^EHYVGF^SHSLSLGGHLLI^GLAKRFKES^PFELEDFIP^MDSTVKNYF^ITDAGTGS 287
 bat CoV RaTG13 208 QMEIDFLELAMDEFI^ERYKLEGYAF^EHYVGF^SHSRQLGGHLLI^GLAKRKS^PLELEDFIP^MDSTVKNYF^ITDAGTGS 287

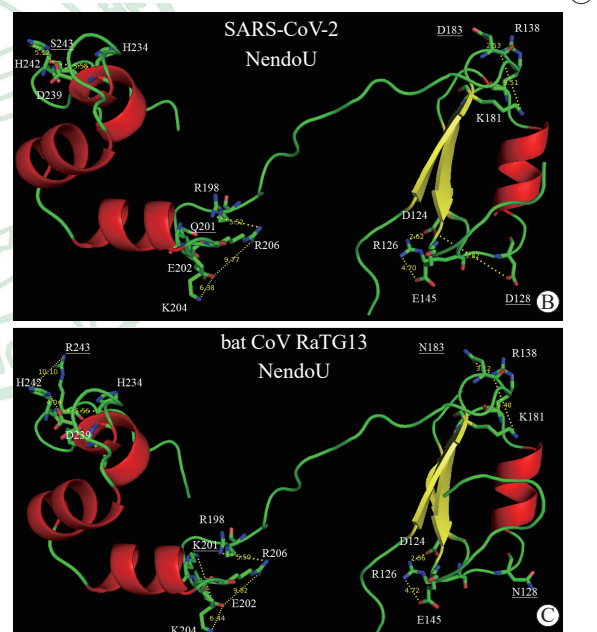


图 6 SARS-CoV-2 和 bat CoV RaTG13 NendoU 蛋白中关键氨基酸变异及电荷相互作用

Fig 6 Key amino acid variations and electrostatic interactions in NendoU from SARS-CoV-2 and bat CoV RaTG13
 A: Amino acid sequences shown in the three-dimensional structure. The variant amino acid residues are in blue. Green lines, yellow arrows and red bars indicate loops, β -folds and α -helices, respectively. B, C: Partial three-dimensional structure of NendoU. S243, Q201, D183, D128, R243, K201, N183 and N128 are variant amino acid residues. SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; bat CoV: Bat coronavirus; NendoU: Uridylate-specific endoribonuclease

表2 SARS-CoV-2 中主要 Nsp 分子的氨基酸变异位点

Tab 2 Amino acid variation sites of Nsp from SARS-CoV-2

| Protein | SARS-CoV-2 | bat CoV RaTG13 | SARS-CoV |
|---------|-------------------------|-------------------------|------------------------|
| RdRp | N185 | D185- R184 | D198- R197 |
| | Y719 | H719- D717 | Y719 |
| | K268- R272-E265 | K268- R272-E265 | C281- R285-E278 |
| ExoN | L257- H281-E282 | R257- H281-E282 | H260- H283-E284 |
| NendoU | Q21 | Q21 | H21- D18 |
| | D128- D124-R126 | N128- D124-R126 | E130- D126-R128 |
| | D183- R138 | N183 | D185 |
| | Q201 | K201- E202 | D201- E203-D204 |
| | S243 | R243- H242-D239 | S241 |
| | E264- E266-D267 | E264- E266-D267 | K266- D241 |
| | PL ^{pro} | K232- E203-R183 | Q232- E203-R183 |
| Nsp3 | D66- E63-H85/H93 | H66- E63-H85/H93 | D73- E70-H100 |
| | H90- K89 | N90- K89 | K95- K96 |

Bold letters indicate amino acid residues that may have electrostatic interactions with the first residue. SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; bat CoV: Bat coronavirus; SARS-CoV: Severe acute respiratory syndrome coronavirus; RdRp: RNA-dependent RNA polymerase; ExoN: Exoribonuclease; NendoU: Uridylate-specific endoribonuclease; PL^{pro}: Papain-like protease; Nsp: Non-structural protein

进一步对 SARS-CoV-2 和 bat CoV RaTG13 中 S 蛋白的氨基酸序列进行比对分析,发现二者的 RBD 共有 19 个氨基酸残基发生变异,其中有 9 个氨基酸残基涉及电荷变化。这些氨基酸变异中有 1 个位于受体结合位点的左侧,其他均位于右侧。根据三维结构中侧链基团氮氧原子间的距离判断: bat CoV RaTG13 S 蛋白的 RBD 含有 3 个强盐键 (D451-R453、D510-质子化的 H514、D414-R417),且该区域含有较多的碱性氨基酸残基,因此这些盐键的作用力受正电斥力作用而增强。而在 SARS-CoV-2 S 蛋白的 RBD,其受体结合位点右侧只形成了 2 个强盐键 (R403-E406、D405-R408),而且由于含有的碱性氨基酸残基较少,盐键的正电排斥增强效应减弱;另外,其左侧则形成了一个全新的盐键 (E516-质子化的 H519)。由此可见,与 bat CoV RaTG13 相比, SARS-CoV-2 S 蛋白受体结合位点两侧的盐键发生了明显变化,这必然影响其与受体的结合能力。已有研究报道,与 SARS-CoV 相比, SARS-CoV-2 S 蛋白的受体结合位点处有多个氨基酸残基发生突变,但是受体结合的三维构象并没有改变^[13]。最近的一项研究显示, SARS-CoV-2 S 蛋白与 ACE2 受体的结合能力远高于 bat CoV RaTG13 (超过 10 倍)^[15]。我们的分析结果也支持上述研究结果^[15],提示

SARS-CoV-2 S 蛋白中 3 处关键的氨基酸变异可能影响了该病毒与受体的相互作用。

最近,有研究指出穿山甲冠状病毒与 SARS-CoV-2 的 S 蛋白高度同源 (氨基酸序列一致性达到 97%),推断 SARS-CoV-2 可能是这株穿山甲冠状病毒与 bat CoV RaTG13 基因组重组的结果^[9,14]。但是,穿山甲冠状病毒的 S1/S2 区域并无上述的 PRRA 短肽插入,而这个插入序列却曾在流感病毒和中东呼吸综合征冠状病毒 (Middle East respiratory syndrome coronavirus, MERS-CoV) 中出现^[9,14],暗示 SARS-CoV-2 S 蛋白的进化可能更为复杂,也许发生了不止一次的重组事件。

2.2.3 蛋白酶的氨基酸变异对电荷相互作用的影响 PL^{pro} 是 SARS-CoV 及其他冠状病毒 Orf1ab 氨基端 3 个 Nsp (Nsp1, Nsp2, Nsp3) 剪切成熟所必需的蛋白水解酶^[17]。对 SARS-CoV-2 和 bat CoV RaTG13 中 PL^{pro} 蛋白进行氨基酸序列比对分析,发现二者之间有 7 个氨基酸变异,其中关键的一个变异位于第 232 位氨基酸残基上。根据 PL^{pro} 蛋白三维结构中侧链基团氮氧原子间的距离判断,该变异位点附近的 E203 与 R183 可以形成一个较强的盐键。在 SARS-CoV-2 中, K232 则通过正电斥力增强 R183 与 E203 间的盐键;在 bat CoV RaTG13 中, Q232 则无增强作用。不过,盐键强弱的变化 (甚至包

括 PL^{pro} 蛋白近羧基端的 T-I 变异) 似乎并没有影响蛋白酶催化中心氨基酸残基的相互作用。因此, SARS-CoV-2 PL^{pro} 蛋白的氨基酸变异对蛋白酶构象及催化功能的影响可能并不明显。

3CL^{pro} 是对 SARS-CoV 及其他冠状病毒的 Nsp 进行成熟切割的主要蛋白酶, 在 Orf1ab 上至少有 11 个切割位点^[16]。从重建的高度保守的蛋白质三维结构来看, SARS-CoV-2 与 bat CoV RaTG13 中 3CL^{pro} 蛋白之间有 2 个氨基酸变异 (P68T 和 N152T), 它们不影响氨基酸残基之间的电荷相互作用, 也不影响蛋白酶的活性中心。因此, 这 2 个变异位点可能不会改变 3CL^{pro} 蛋白酶的构象和功能。

总之, 本研究通过分析 SARS-CoV-2 和 bat CoV RaTG13 之间主要蛋白质分子的氨基酸变异, 发现 SARS-CoV-2 编码的 RdRp、ExoN、NendoU 和 S 蛋白上至少发生了 10 处关键的氨基酸变异, 这些变异改变了相关氨基酸残基之间的电荷相互作用, 可能影响到相应蛋白的空间构象及病毒的遗传表型 (如病毒 RNA 复制校正功能、致病性、与细胞受体的亲和力及传染性等)。

[参考文献]

- [1] World Health Organization. Coronavirus disease (COVID-19) outbreak situation[EB/OL]. (2020-04-15) [2020-04-15]. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>.
- [2] WU F, ZHAO S, YU B, CHEN Y M, WANG W, SONG Z G, et al. A new coronavirus associated with human respiratory disease in China[J]. *Nature*, 2020, 579: 265-269.
- [3] ZHOU P, YANG X L, WANG X G, HU B, ZHANG L, ZHANG W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin[J]. *Nature*, 2020, 579: 270-273.
- [4] LI Q, GUAN X, WU P, WANG X, ZHOU L, TONG Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia[J]. *N Engl J Med*, 2020, 382: 1199-1207.
- [5] LU R, ZHAO X, LI J, NIU P, YANG B, WU H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding[J]. *Lancet*, 2020, 395: 565-574.
- [6] CHAN J F, YUAN S, KOK K H, TO K K, CHU H, YANG J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster[J]. *Lancet*, 2020, 395: 514-523.
- [7] HUANG C, WANG Y, LI X, REN L, ZHAO J, HU Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China[J]. *Lancet*, 2020, 395: 497-506.
- [8] LAM T, SHUM M, ZHU H, TONG Y, NI X, LIAO Y, et al. Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China[J/OL]. *bioRxiv*, 2020. doi: 10.1101/2020.02.13.945485.
- [9] XIAO K, ZHAI J, FENG Y, ZHOU N, ZHANG X, ZOU J, et al. Isolation and characterization of 2019-nCoV-like coronavirus from Malayan pangolins[J/OL]. *bioRxiv*, 2020. doi: 10.1101/2020.02.17.951335.
- [10] HILGENFELD R, PEIRIS M. From SARS to MERS: 10 years of research on highly pathogenic human coronaviruses[J]. *Antiviral Res*, 2013, 100: 286-295.
- [11] SNIJDER E J, DECROL E, ZIEBUHR J. The nonstructural proteins directing coronavirus RNA synthesis and processing[J]. *Adv Virus Res*, 2016, 96: 59-126.
- [12] LI F. Structure, function, and evolution of coronavirus spike proteins[J]. *Annu Rev Virol*, 2016, 3: 237-261.
- [13] XU X, CHEN P, WANG J, FENG J, ZHOU H, LI X, et al. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission[J]. *Sci China Life Sci*, 2020, 63: 457-460.
- [14] WU A, NIU P, WANG L, ZHOU H, ZHAO X, WANG W, et al. Mutations, recombination and insertion in the evolution of 2019-nCoV[J/OL]. *bioRxiv*, 2020. doi: 10.1101/2020.02.29.971101.
- [15] WRAPP D, WANG N, CORBETT K, GOLDSMITH J A, HSIEH C L, ABIONA O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation[J]. *Science*, 2020, 367: 1260-1263.
- [16] WATERHOUSE A, BERTONI M, BIENERT S, STUDER G, TAURIELLO G, GUMIENNY R, et al. SWISS-MODEL: homology modelling of protein structures and complexes[J]. *Nucleic Acids Res*, 2018, 46(W1): W296-W303.
- [17] HILGENFELD R. From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design[J]. *FEBS J*, 2014, 281: 4085-4096.

[本文编辑] 商素芳