Wuhan 2019 coronavirus spike protein may contain a Furin protease cleavage site

Authors: Li Xin, Duan Guangyou, Zhang Wei, Shi Jinsong, Chen Jiayuan, Chen Shunmei, high mountain, Ruan Jishou, alpine, Ruan Jishou

Date: 2020-02-14T00:00:00+00:00

Abstract

Abstract: In December 2019, pneumonia caused by the 2019 novel coronavirus (2019-nCoV) was reported in Wuhan, China. Based on genomic information, our previous studies demonstrated that although 2019-nCoV and SARS coronavirus both belong to Beta coronavirus subgroup B (BB coronavirus), the two viruses differ substantially—a finding consistent with differences in their clinical manifestations. Previous studies also revealed extensive alternative translation in BB coronavirus and elucidated at the molecular level the characteristics of rapid mutation and high diversity. This study reports, for the first time internationally, an important mutation in the S protein of BB coronavirus; this mutation endows 2019-nCoV with a furin protease cleavage site absent in all other BB coronaviruses (including SARS and SARS-like coronaviruses) except murine hepatitis coronavirus. This mutation may enhance the efficiency of viral cell entry, thereby conferring transmissibility significantly greater than that of SARS coronavirus. Due to this mutation, the packaging mechanism of 2019 coronavirus also differs from most other Beta coronaviruses such as SARS, and may resemble that of murine hepatitis coronavirus, HIV, Ebola virus, and some avian influenza viruses. As an unexpected finding, some avian influenza viruses can also acquire furin protease cleavage sites through mutation. Follow-up studies of this important mutation will lay the foundation for elucidating the mechanisms underlying the high transmissibility of 2019-nCoV and for developing drugs, antibodies, and vaccines.

Full Text

A Furin Cleavage Site Was Discovered in the S Protein of the Wuhan 2019 Novel Coronavirus

Xin Li[12] 3, **Wei Zhang**[1], **Jinsong Shi** , **Jiayuan Chen**[2], **Shunmei Chen** , **Shan Gao**[2], ***Jishou Ruan***[1]

[1]School of Mathematical Sciences, Nankai University, Tianjin 300071, P.R.China
[2]School of Life Sciences, Nankai University, Tianjin 300071, P.R.China
[3]School of Life Sciences, Qilu Normal University, Jinan, Shandong 250200, P.R.China
National Clinical Research Center of Kidney Disease, Jinling Hospital, Nanjing University School of Medicine, Nanjing, Jiangsu 210016, P.R.China
Institute of Molecular and Clinical Medicine, Kunming Medical College, Kunming, Yunnan 650500, P.R.China

Abstract

In December 2019, a pneumonia outbreak caused by the 2019 novel coronavirus (2019-nCoV) was reported in Wuhan, China. Based on genomic analysis, our previous research demonstrated that although both 2019-nCoV and SARS coronavirus belong to Betacoronavirus subgroup B (BB coronavirus), they exhibit substantial genetic differences consistent with their divergent clinical manifestations. Our earlier studies also identified extensive alternative translation events in BB coronaviruses, revealing their high mutation rates and genetic diversity at the molecular level. The present study reports, for the first time, a critical mutation in the spike (S) protein of BB coronaviruses. This mutation introduces a furin protease cleavage site in 2019-nCoV—a feature absent in all other BB coronaviruses (including SARS and SARS-like viruses) except for mouse hepatitis coronavirus. This mutation likely enhances the efficiency of viral cell entry, conferring significantly greater transmissibility to 2019-nCoV compared to SARS coronavirus. Consequently, the packaging mechanism of 2019-nCoV may differ from that of most other betacoronaviruses and instead resemble those of mouse hepatitis coronavirus, HIV, Ebola virus, and certain avian influenza viruses. As a serendipitous finding, some avian influenza viruses can also acquire furin cleavage sites through mutation. Further investigation of this important mutation will elucidate the mechanisms underlying 2019-nCoV's high transmissibility and provide a foundation for drug, antibody, and vaccine development.

Keywords: coronavirus; furin protease; 2019-nCoV; HIV; avian influenza

Introduction

In December 2019, a pneumonia outbreak associated with the 2019 novel coronavirus (2019-nCoV) emerged in Wuhan, China. Based on genomic analysis, we identified four key findings: (1) 2019-nCoV and severe acute respiratory syndrome (SARS) coronavirus both belong to Betacoronavirus subgroup B (BB coronavirus), yet exhibit substantial genetic divergence that aligns with their distinct clinical presentations; (2) compared to SARS coronavirus, 2019-nCoV

demonstrates lower virulence but higher transmissibility; (3) phylogenetic analysis supports a bat origin of 2019-nCoV from Chinese horseshoe bats; and (4) BB coronaviruses undergo extensive alternative translation, exhibiting rapid mutation and high genetic diversity.

Numerous RNA viruses invade host cells through membrane fusion mediated by viral fusion proteins. The SARS coronavirus spike protein, HIV envelope glycoprotein (Env), influenza hemagglutinin (HA), and Ebola glycoprotein (GP) all belong to the class I viral fusion protein family [**?**]. Both HIV and SARS coronavirus share similarities as enveloped viruses that enter cells via membrane fusion, requiring proteolytic cleavage of their fusion proteins into receptor-binding and membrane-fusion domains prior to infection. However, they differ critically in their processing: HIV gp160 is cleaved by furin protease during intracellular packaging, yielding separate gp120 (receptor-binding) and gp41 (membrane-fusion) subunits on the virion surface, whereas the SARS coronavirus spike protein remains intact with its S1 (receptor-binding) and S2 (membrane-fusion) domains fused. SARS coronavirus employs two entry pathways: direct membrane fusion at the cell surface when proteases like trypsin are present to cleave the S protein into S1 and S2, or endocytosis followed by lysosomal cathepsin-mediated cleavage. Studies show the direct membrane fusion pathway is 100- to 1000-fold more efficient than the endocytic route [**?**]. Except for mouse hepatitis coronavirus (MHV), most betacoronaviruses, including SARS, lack furin cleavage sites at the S1-S2 junctional region.

Many mechanisms of SARS coronavirus infection remain unclear, limiting applications in drug, vaccine, and antibody development. Large-scale genomic studies, particularly on S protein variations, can deepen our understanding of BB coronavirus infection mechanisms and reveal unique features of 2019-nCoV, providing foundations for prevention and treatment. Building on our previous work, we serendipitously identified a potential furin cleavage site in the 2019-nCoV genome. This finding suggests fundamental differences in infection pathways between 2019-nCoV and SARS coronavirus, indicating that 2019-nCoV may utilize a packaging mechanism similar to HIV and other viruses. This discovery provides theoretical support for treating 2019-nCoV infections, as numerous existing drugs for related viruses (e.g., HIV) could be repurposed in combination therapies with immunosuppressants to improve efficacy or reduce side effects.

Materials and Methods

Our previous study utilized 13 BB coronavirus genome sequences (GenBank: JX993987, JX993988, GQ153539, GQ153540, GQ153542, DQ071615, DQ412042, DQ412043, AY515512, AY572034, AY274119, MN908947, and MG772934). In this study, these sequences were grouped by host into five categories for further analysis: SARS (AY274119), civet (AY515512 and AY572034), 2019-nCoV (MN908947), Chinese horseshoe bat group from Zhoushan, Zhejiang

(MG772934), and other bat groups (eight additional bat-derived sequences excluding MG772934). Additionally, BGI provided 29 2019-nCoV genome sequences for result verification. Sequence multiple alignment was performed using the online tool ClustalW2, while data processing, statistical analysis, and visualization were conducted using R v2.15.3 [**?**]. Protein secondary structure prediction employed PSIPRED v4.0 [**?**] with default parameters.

Results and Analysis

Using NCBI BLAST to compare S1 and S2 nucleotide sequences between 2019-nCoV (MN908947) and SARS coronavirus (AY274119), we found 66.4% identity in S1 and 80.1% identity in S2—a substantial difference. Examination of the amino acid sequence at the S1-S2 junctional region revealed an "RRAR" motif (Figure 1A [Figure 1: see original paper]), which matches the furin recognition pattern "RXXR" [**?**]. Alignment of the junctional nucleotide sequences between 2019-nCoV and SARS showed that this variation resulted from a 12-nucleotide insertion (Figure 1B). BLAST searching the "CGGCGG"core sequence extended by 15 bp at both 5' and 3' ends against the NCBI NT database suggested possible bacterial origins, necessitating verification to exclude false positives from sequencing or assembly errors. We confirmed the finding through: (1) identifying three or more submitted 2019-nCoV genome sequences in NCBI GenBank supporting the "CGGCGG" insertion; (2) searching all betacoronaviruses (excluding 2019-nCoV) for "RRAR" motifs in the S protein junctional region, finding only mouse hepatitis coronavirus contains a similar furin cleavage site ("RRARR"); and (3) confirming via protein secondary structure prediction that "RRAR" does not participate in folded structures.

Based on these results, the 2019-nCoV S protein likely acquired a furin cleavage site through mutation, suggesting its packaging mechanism may resemble that of mouse hepatitis coronavirus rather than most other betacoronaviruses like SARS. Previous studies show MHV S protein can be cleaved by furin-like proteases during intracellular packaging, releasing virions with non-fused S1 and S2 subunits [**?**]. Notably, the newly acquired "RRAR" motif in 2019-nCoV includes a final arginine that corresponds to a trypsin cleavage site (R667) in SARS coronavirus S protein (Figure 1A), while the primary trypsin site R797 remains conserved in 2019-nCoV. Moreover, R667 aligns with the furin cleavage site in MHV S protein. Experimental evidence indicates that engineered introduction of furin cleavage sites at R667 or R797 in SARS coronavirus S protein can enhance membrane fusion capacity [**?**]. Thus, 2019-nCoV uniquely possesses both furin (R685) and trypsin (R815) cleavage sites compared to SARS coronavirus, likely facilitating more frequent direct membrane fusion entry and higher infection efficiency.

Among class I viral fusion proteins, HIV gp160 (GenBank: NC_001802.1) and Ebola virus GP (GenBank: NC_002549.1) contain furin cleavage sites "REKR"

and "RKIR," respectively, while among betacoronaviruses, only mouse hepatitis coronavirus has the "RRARR" furin site. Generally, SARS and other betacoronaviruses differ in packaging and infection mechanisms from viruses possessing furin cleavage sites. Serendipitously, we found that some influenza viruses can also acquire furin cleavage sites through mutation, predominantly avian influenza strains (Table 1). The evolutionary acquisition of a furin cleavage site in 2019-nCoV clearly impacts its biological functions, particularly infectivity. Subsequent studies will advance our understanding of viral fusion protein function and membrane fusion mechanisms, as well as deepen knowledge of betacoronavirus packaging and infection pathways.

Table 1. Furin Cleavage Sites in Influenza Viruses

Virus	Accession Number	Subtype	Host	Furin Site Position	Furin Site Sequence
Influenza A	MN653237	Canine	Great crested grebe	···	···
Influenza A	MH988772	Cormorant	··	···	···
··· (additional rows)	...	...	...	...	...

The first column shows NCBI GenBank accession numbers; the fourth column indicates the furin cleavage site position in the amino acid sequence; the fifth column shows the furin cleavage site sequence.

Conclusions

This study reaches three main conclusions: (1) The 2019-nCoV S protein likely contains a furin protease cleavage site, suggesting a packaging mechanism similar to mouse hepatitis coronavirus but distinct from most other betacoronaviruses including SARS; (2) This altered packaging mechanism may confer higher efficiency in cell entry, potentially explaining 2019-nCoV' s greater transmissibility compared to SARS coronavirus; and (3) Some avian influenza viruses can similarly acquire furin cleavage sites through mutation to enhance cell entry efficiency.

The altered packaging mechanism of 2019-nCoV implies that numerous existing antiviral drugs—particularly those targeting mouse hepatitis coronavirus, HIV,

Ebola virus, and avian influenza—could be repurposed in combination therapies with immunosuppressants to improve treatment efficacy or reduce side effects. Current antiviral strategies primarily target viral proteins (e.g., RNA polymerase) or host proteins (e.g., furin protease). Unlike most virus-targeted approaches, Jishou Ruan and colleagues at Nankai University propose host-targeted drug screening to circumvent viral mutation issues. Based on clinical data from Tianjin University of Traditional Chinese Medicine First Affiliated Hospital for heart failure treatment, we have identified a drug combination that effectively inhibits furin protease activity with minimal side effects.

Acknowledgments

We thank Professors Quan Chen, Deling Kong, Wenjun Bu, Tao Zhang, Dawei Huang, Yanqiang Liu, Qiang Zhao, and Bingjun He from the School of Life Sciences at Nankai University for their long-term support of our bioinformatics research. We also acknowledge the voluntary contributions of Yibo Xuan and colleagues from Hebei Normal University.

References

1. B.J. Bosch, R. van der Zee, C.A. de Haan and P.J. Rottier, *The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex.* Journal of virology, 2003. 77(16): p. 8801-8811.
2. S. Matsuyama, M. Ujike, S. Morikawa, M. Tashiro and F. Taguchi, *Protease-mediated enhancement of severe acute respiratory syndrome coronavirus infection.* Proceedings of the National Academy of Sciences, 2005. 102(35): p. 12543-12547.
3. GAO Shan, OU JianHong, XIAO Kai. *R language and Bioconductor in bioinformatics applications* (Chinese Edition) [M]. 2014, Tianjin: Tianjin Science and Technology Translation Publishing.
4. Buchan,D.W.A. and Jones,D.T. (2019) *The PSIPRED Protein Analysis Workbench: 20 years on.* Nucleic Acids Res., 47, W402–W407.
5. S. Molloy, P. Bresnahan, S.H. Leppla, K. Klimpel and G. Thomas, *Human furin is a calcium-dependent serine endoprotease that recognizes the sequence Arg-XX-Arg and efficiently cleaves anthrax toxin protective antigen.* Journal of Biological Chemistry, 1992. 267(23): p. 16396-16402.
6. C.A. de Haan, K. Stadler, G.-J. Godeke, B.J. Bosch and P.J. Rottier, *Cleavage inhibition of the murine coronavirus spike protein by a furin-like enzyme affects cell-cell but not virus-cell fusion.* Journal of virology, 2004. 78(11): p. 6048-6054.

7. S. Belouzard, V.C. Chu and G.R. Whittaker, *Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites.* Proceedings of the National Academy of Sciences, 2009. 106(14): p. 5871-5876.

Note: Figure translations are in progress. See original paper for figures.