

DOI:10.12113/202002001

2019 新型冠状病毒 S 蛋白可能存在 Furin 蛋白酶切位点

李 鑫^{1,2*}, 段广有^{3*}, 张 伟¹, 施劲松⁴
陈嘉源², 陈舜梅⁵, 高 山^{2**}, 阮吉寿^{1**}

(1.南开大学 数学科学学院, 天津 300071; 2.南开大学 生命科学学院, 天津 300071;
3.齐鲁师范学院 生命科学学院, 济南 250200; 4.东部战区总医院, 南京 210016;
5.昆明医科大学 分子临床医学研究院, 昆明 650500)

摘要:2019年12月,中国武汉报道了2019新型冠状病毒(2019 novel Coronavirus, 2019-nCoV)引起的肺炎。基于基因组信息,我们前期研究结果显示2019-nCoV与SARS冠状病毒虽然同属于Beta冠状病毒B亚群(BB冠状病毒),但两种病毒差异较大,这一结果与两者临床症状差异一致。前期研究还发现了BB冠状病毒存在大量的可变翻译,并从分子水平揭示了BB冠状病毒变异快、多样性高的特点。本研究在国际上首次报道Beta冠状病毒S蛋白上的一个重要突变,这个突变使2019-nCoV具有了一个可供Furin蛋白酶切的位点,是大部分Beta冠状病毒(特别是SARS和SARS样(SARS-like)冠状病毒)所不具有的。我们的一个结论是这个突变有可能增强了2019-nCoV侵染细胞的效率,进而使其传播力显著大于SARS冠状病毒。由于这个突变,2019-nCoV的感染机制也会不同于SARS等大部分Beta冠状病毒,而与鼠肝炎冠状病毒、HIV、埃博拉病毒和一些禽流感病毒的感染机制更相似。我们意外发现一些禽流感病毒也可以通过突变获得Furin蛋白酶切位点,这说明自然突变可以引入Furin酶切位点。除此之外,在2019-nCoV的S蛋白中插入的“CGGCGG”序列编码两个精氨酸,然而“CGG”对于宿主(人)来说是蛋白质翻译的稀有密码子。我们的另一个结论是引入Furin蛋白酶切位点的插入突变中包含的“CGG”是传播到人之前形成的;2019-nCoV的中间宿主应该是密码子“CGG”相对使用频率更高的哺乳动物。使用我们提供的密码子“CGG”相对频率表结合2019-nCoV检测阳性的动物样品信息可以准确地确定2019-nCoV的中间宿主。对这个重要突变的后续研究将为揭示2019-nCoV传播力强的原因,以及为药物、抗体和疫苗的开发等工作奠定基础。

关键词:冠状病毒;Furin蛋白酶;SARS-CoV-2;中间宿主;禽流感

中图分类号:Q93 **文献标志码:**A **文章编号:**1672-5565(2020)02-103-06

A furin cleavage site was discovered in the S protein of the 2019 novel coronavirus

LI Xin^{1,2*}, DUAN Guangyou^{3*}, ZHANG Wei¹, SHI Jinsong⁴
CHEN Jiayuan², CHEN Shunmei⁵, GAO Shan^{2**}, RUAN Jishou^{1**}

(1. School of Mathematical Sciences, Nankai University, Tianjin 300071, P.R.China;
2. College of Life Sciences, Nankai University, Tianjin 300071, China;
3. School of Life Sciences, Qilu Normal University, Jinan 250200, China;
4. National Clinical Research Center of Kidney Disease, Jinling Hospital, Nanjing University School of Medicine, Nanjing 210016, China;
5. Institute of Molecular and Clinical Medicine, Kunming Medical College, Kunming 650500, China)

Abstract:The 2019 novel Coronavirus (2019-nCoV) has caused the pneumonia outbreak in Wuhan (a city of China). In our previous study, the analytical results showed that both 2019-nCoV and SARS coronavirus belong to Betacoronavirus subgroup B (BB coronavirus), but have large differences, which are consistent with the differences in the clinical symptoms of two related diseases. The most important finding was that the alternative translation of Nankai CDS could produce more than 17 putative proteins, which may be responsible for the host adaption. The

收稿日期:2020-02-04;修回日期:2020-02-22.

* 作者简介:李鑫,男,副研究员,研究方向:蛋白质结构. E-mail:lix1980@nankai.edu.cn;

段广有,男,教授,研究方向:生物信息学. E-mail:guangyou.duan@qlnu.edu.cn.

** 通信作者:高山,男,副教授、硕导,研究方向:生物信息学. E-mail:gao_shan@mail.nankai.edu.cn;

阮吉寿,男,教授、博导,研究方向:生物信息学. E-mail:jsruan@nankai.edu.cn.

genotyping of 13 viruses using the 17 putative proteins revealed the high mutation rate and diversity of BB coronavirus. The present study for the first time (on January 21st, 2020) reported a very important mutation in the Spike (S) proteins of Betacoronavirus. By this mutation, 2019-nCoV acquired a cleavage site for furin enzyme in its S protein, which is not present in the S proteins of most other Betacoronavirus (e.g. SARS coronavirus). This cleavage site may increase the efficiency of virus infection into cells, making 2019-nCoV has significantly stronger transmissibility than SARS coronavirus. The infection mechanism of 2019-nCoV may be changed to being more similar to those of MHV, HIV, Ebola virus (EBoV) and some avian influenza viruses, other than those of most other Betacoronavirus (e.g. SARS coronavirus). In addition, we unexpectedly found that some avian influenza viruses acquired a cleavage site for furin enzyme by the similar mutation as 2019-nCoV. Therefore, the natural mutation can result in a short insertion to form a cleavage site for furin enzyme. The cleavage site for furin enzyme in 2019-nCoV contains the "CGGCCG" sequence encoding two arginine (R) residues. "CGG", however, is a rare codon for human. So we concluded that these two codons were present in the 2019-nCoV-like Betacoronavirus before they transmitted into human and the intermediate host(s) are mammals with a high relative frequency of "CGG" usage. We provide a relative frequency table of "CGG" usage in mammals to help identify the intermediate hosts of 2019-nCoV. Future studies of this mutation will help to reveal the stronger transmissibility of 2019-nCoV and lay foundations for vaccine development and drug design of, but not limited to 2019-nCoV.

Keywords: Coronavirus; Furin enzyme; SARS-CoV-2; Intermediate host; Avian influenza virus

2019年12月,中国武汉报道了2019新型冠状病毒(2019 novel Coronavirus, 2019-nCoV)引起的肺炎。基于2019-nCoV的基因组信息和公开的临床数据,我们发现^[1]:(1)2019-nCoV与严重急性呼吸综合征(Severe Acute Respiratory Syndrome, SARS)冠状病毒(SARS Coronavirus, SARS-CoV)同属Beta冠状病毒B亚群(BB冠状病毒),但两种病毒差异较大,这一结果与两者临床症状差异一致;(2)与SARS冠状病毒相比,2019-nCoV虽然毒力较弱,但传播力更强;(3)溯源分析的结果支持2019-nCoV源自蝙蝠;(4)BB冠状病毒存在大量的可变翻译,而且具有变异快、多样性高的特点。

已知多种RNA病毒都要通过自身的膜融合蛋白与靶细胞结合(膜融合)从而进入细胞。其中,SARS冠状病毒S蛋白(Spike protein)、HIV包膜糖蛋白(Envelope glycoprotein, Env)、埃博拉病毒糖蛋白(Glycoprotein, GP)和流感病毒血凝素(Hemagglutinin, HA)等属于I类膜融合蛋白^[2]。比较HIV病毒与SARS冠状病毒,相同之处有:都是包膜病毒,都可以通过膜融合途径(各自的其它途径不做进一步讨论)进入细胞,都需要其膜融合蛋白被细胞蛋白酶切割,分割为受体结合结构域和膜融合结构域。两种病毒的不同之处在于:HIV病毒的gp160在细胞内包装过程中被Furin蛋白酶切割,分泌出的病毒颗粒表面的gp120(负责与受体结合)和gp41(负责膜融合)是分开的两个亚基;而SARS冠状病毒分泌出的病毒颗粒表面的S蛋白(Spike Protein)中的S1(负责与受体结合)和S2(负责膜融

合)仍是融合状态。SARS冠状病毒通过两种方式侵染宿主细胞,在细胞表面有蛋白酶(如胰蛋白酶Trypsin)时,S蛋白被切割成为S1和S2两个亚基,进而与宿主细胞膜融合进入细胞;否则,SARS冠状病毒通过胞吞途径进入宿主细胞,而后S蛋白被溶酶体内的组织蛋白酶切割(不再进一步讨论)。研究显示前一种方式(直接膜融合)的侵染效率是后者的约100到1000倍^[3]。除了鼠肝炎冠状病毒(The Mouse Hepatitis coronavirus, MHV)等少数外,SARS等其它大部分Beta冠状病毒的S1与S2之间的交界区(Junctional region)没有被Furin蛋白酶切割的位点(Cleavage site),即Furin蛋白酶切位点。

有关SARS冠状病毒感染过程中的很多机制尚不明确,因此不能为药物、疫苗和抗体开发等应用提供更多信息。基于大量基因组数据的比较研究,特别是针对S蛋白区域的变异研究,不仅可以有助于深入了解BB冠状病毒的感染机制,而且有助于揭示2019-nCoV感染的特点,为病毒防控以及治疗奠定基础。本研究在前期工作基础上,无意间发现了2019-nCoV的S蛋白可能存在Furin蛋白酶切位点。这一发现暗示了2019-nCoV可能在感染机制上与SARS冠状病毒有较大差异,转而与HIV等其它病毒的感染机制更相似。不同病毒采取相同的感染机制,提示了治疗相关病毒(例如HIV)的大量药物都可以考虑以“老药新用”的方式进行组合,与免疫抑制剂等联合用药,以提高治疗效果。

1 数据与方法

在前期研究中,我们共使用 13 条 BB 冠状病毒基因组序列, (GenBank: JX993987、JX993988、GQ153539、GQ153540、GQ153542、DQ071615、DQ412042、DQ412043、AY515512、AY572034、AY274119、MN908947 和 MG772934)。在本研究中,13 条序列根据其宿主分为五组,用于进一步研究,这五组命名为 SARS (AY274119)、果子狸 (AY515512 和 AY572034)、2019-nCoV (MN908947)、来自浙江舟山的蝙蝠群体 (MG772934) 和其它蝙蝠群体 (MG772934 之外 8 条来自蝙蝠的序列)。在本研究中,序列多重比对使用在线软件 ClustalW2,数据处理、统计与作图使用软件 R v2.15.3^[4],蛋白质二级结构预测使用软件 PSIPRED v4.0^[5],所有参数采用默认值。密码子偏好性分析使用 NCBI GenBank 数据库 (截止 2019 年 12 月 12 日) 中符合物种 (灵长类、脊椎动物、哺乳动物和啮齿动物任何一个) 的数据,去除密码子数少于 5 000 的物种,再加上 2019-nCoV 和蝙蝠冠状病毒两大类数据。密码子偏好性分析中,密码子相对使用频率 (以下简称相对频率) 指的是一个密码子使用次数占其对应氨基酸所有密码子使用次数和的百分比。

2 结果与分析

使用 NCBI 工具 Blast, 比对 2019-nCoV (MN908947) 与 SARS 冠状病毒 (AY274119) 之间 S1 和 S2 核酸序列之间同一度 (Identity), 结果显示 S1 在两种病毒之间同一度是 66.4%, 而 S2 是 80.1%, 两个同一度差异很大。于是观察 S1 与 S2 之间的交界区的氨基酸序列, 无意间发现了“RRAR”序列 (见图 1A), 该序列符合 Furin 酶切位点的识别模式“RXXR”^[6]。比对 2019-nCoV 与 SARS 之间交界区的核酸序列, 发现变异来源是插入了 12 个碱基, 最重要的插入是“CGGCGG” (见图 1B)。将以“CGGCGG”为中心向 5' 和 3' 端各扩展 15 bp 得到的序列比对到 NCBI NT 数据库, 发现“CGGCGG”也有可能来自细菌。因此, 必须通过以下方式排除“CGGCGG”来自测序或拼接错误: (1) 检索 NCBI Genbank 数据库, 找到三条以上已提交的 2019-nCoV 基因组序列支持“CGGCGG”; (2) 检索 NCBI Genbank 数据库, 从所有的 Beta 冠状病毒 (2019-

nCoV 除外) 的 S 蛋白的交界区中搜索“RRAR”模式, 发现只有在鼠肝炎等少数冠状病毒中存在 Furin 酶切位点, 而在所有 SARS 和 SARS 样冠状病毒中都不存在 Furin 酶切位点; (3) 通过蛋白质二级结构预测确定“RRAR”未参与折叠配对。最后, 使用华大基因公开的 148 条 2019-nCoV 基因组序列确定了所有 2019-nCoV 的 S 蛋白都包括完全一样的 Furin 酶切位点。

基于以上结果, 2019-nCoV 的 S 蛋白可能因为突变引入了 Furin 蛋白酶切位点, 其感染机制可能与鼠肝炎冠状病毒的感染机制更相似, 而不同于 SARS 等其它大部分 Beta 冠状病毒。前期实验结果表明, 鼠肝炎冠状病毒的 S 蛋白在细胞内包装过程中可被 Furin 样 (Furin-like) 蛋白酶切割, 从而分泌出 S1 和 S2 呈非融合状态的病毒颗粒^[7]。另一方面, 2019-nCoV 的 S 蛋白中新增的“RRAR”中最后一个 R 恰好对应 SARS 冠状病毒 S 蛋白中的一个胰蛋白酶切位点 R667 (图 1A), 而 SARS 冠状病毒 S 蛋白中的另外一个胰蛋白酶切位点 R797 (相比 R667 更主要) 在 2019-nCoV 的 S 蛋白中对应 R815。另外, R667 恰好对应鼠肝炎冠状病毒 S 蛋白中的 Furin 酶切位点。有实验表明如果在 SARS 冠状病毒 S 蛋白的交界区的 R667 或 R797^[8] 人为加入 Furin 酶切位点, 可以增强 S 蛋白的膜融合能力。2019-nCoV 同时具备 Furin (R685) 和胰蛋白酶切位点 (R815), 分别对应 SARS 冠状病毒的胰蛋白酶切位点 (R667 和 R797)。这种改变将使 2019-nCoV 更多地通过直接膜融合的方式侵染细胞, 故进入细胞的效率更高。

在第 I 类膜融合蛋白中, HIV (GenBank: NC_001802.1) 的 gp160 和埃博拉病毒 (GenBank: NC_002549.1) GP 的 Furin 酶切位点分别是“REKR”和“RKIR”, 而鼠肝炎冠状病毒的 Furin 酶切位点是“RRARR”。总体上, SARS 等 Beta 冠状病毒与具有 Furin 酶切位点的其它病毒的感染机制不同。作为一个意外发现, 一些流感病毒的 HA 也可以通过突变获得一个 Furin 蛋白酶切位点, 而且这些流感病毒大部分是禽流感 (见表 1)。综合分析 NCBI GenBank 全库的 Beta 冠状病毒、HIV、埃博拉病毒和流感病毒中的 Furin 酶切位点后, 我们推断 Furin 酶切位点对病毒感染等生物学功能产生较大影响, 后续研究将有助于我们提高对这几类病毒的膜融合蛋白功能以及病毒膜融合机制的认识, 也有助于我们对 Beta 冠状病毒的感染机制的深入研究。

Furin 酶切位点(见表 1)中精氨酸 R 偏好使用密码子“AGA”,这恰好与禽类宿主“AGA”相对频率高对应。因此,我们提出中间宿主必须满足的条件是从

该动物提取的病毒的 S 蛋白应该含有 Furin 酶切位点和密码子“CGG”。

表 1 流感病毒获得 Furin 酶切位点
Table 1 Furin cleavage sites in influenza viruses

Virus ^a	Type	Subtype	AAposition ^b	Furin ^c	Host
MN653237	Influenza A	H3N2	342	RQTR	Canine
MH363669	Influenza A	H3N2	342	RQTR	Canine
MH266392	Influenza A	H3N2	342	RQTR	Canine
MH988772	Influenza A	H5N1	343	RKKR	Great created grebe
MH988771	Influenza A	H5N1	343	RKKR	Cormorant
DQ864717	Influenza A	H5N1	343	RKKR	Goose
FJ602810	Influenza A	H5N1	343	RKKR	Great black-headed gull
EU401796	Influenza A	H5N1	343	RKKR	Peacock
MH988774	Influenza A	H5N2	339	RETR	Teal
M24457	Influenza A	H7N1	339	RKKR	Chicken
AJ493216	Influenza A	H7N1	340	RVRR	Turkey
EF470587	Influenza A	H7N3	343	RMTR	Chicken
AY303631	Influenza A	H7N3	346	RETR	Chicken
CY015065	Influenza A	H7N3	338	RRRR	Turkey
AY338459	Influenza A	H7N7	338	RRRR	Chicken
Z12617	Influenza A	H7N7	339	RRKR	Fowl
Z47199	Influenza A	H7N7	339	REKR	Chicken
M17735	Influenza A	H7N7	339	REKR	Chicken
MK552554	Influenza A	H9N2	335	RSSR	Duck
LC208508	Influenza A	H9N2	335	RSSR	Duck
MN918143	Influenza A	H9N2	335	RSSR	Chicken

注:第 1 列使用 NCBI GenBank 数据库的 Accession Number;第 4 列是 Furin 酶切位点在氨基酸序列中的位置;第 5 列是 Furin 酶切位点的序列。
Notes: (a)The Accession Numbers in the NCBI GenBank database.(b)The positions of cleavage sites for furin enzyme. (c)The sequences of cleavage sites for furin enzyme.

3 结 论

1)2019-nCoV 的 S 蛋白可能存在 Furin 蛋白酶切位点,从而导致 2019-nCoV 的感染机制不同于 SARS 等大部分 Beta 冠状病毒,而与鼠肝炎冠状病毒、HIV、埃博拉病毒和一些禽流感病毒的感染机制更相似;

2)由于感染机制的改变,2019-nCoV 获得了更高的进入细胞的效率,这可能是其传播能力大于 SARS 冠状病毒的一个原因;

3)一些禽流感病毒也可以通过突变获得一个 Furin 蛋白酶切位点,这说明自然突变可以引入 Furin 酶切位点;

4)包含“CGGCGG”的插入突变是传播到人之前形成的,2019-nCoV 的中间宿主应该来自一个“CGG”相对频率更高的动物,特别是鹿、狐和鼠;

5)我们提出中间宿主必须满足的条件是从该动物提取的病毒的 S 蛋白应该含有 Furin 酶切位点和密码子“CGG”;

6)使用我们提供的“CGG”相对频率表结合 2019-nCoV 检测阳性的动物样品信息可以准确地确定 2019-nCoV 的中间宿主。

4 讨 论

不同病毒采取相同的感染机制,提示了现有的大量抗病毒药物(特别是抗鼠肝炎冠状病毒、HIV、

埃博拉病毒和禽流感的)都可以考虑以“老药新用”的方式进行组合,与免疫抑制剂等联合用药,以提高治疗效果。在已有抗病毒药物中,有以病毒蛋白(如RNA合成酶)为靶点的药物,也有以宿主(人)蛋白为靶点(如Furin蛋白酶)的药物。根据BB冠状病毒变异快的特点,南开大学阮吉寿等提出同时考虑以宿主(人)蛋白为靶点进行药物筛选或设计,这样可以有效规避病毒变异的影响。我们根据天津中医药大学第一附属医院的心衰治疗数据,选定了一组治疗心衰的药物组合,可以很好地抑制Furin蛋白酶活性,同时副作用较小。下一步的研究还应该考虑从病毒和宿主细胞膜融合角度对2019-nCoV、鼠肝炎冠状病毒、流感、HIV和埃博拉病毒进行比较研究,找到膜融合的共性与差异,从干扰膜融合角度进行药物设计。

致谢:感谢南开大学生命科学学院陈佳、孔德领、卜文俊、张涛、黄大卫、刘燕强、赵强和贺秉军等各位老师对我们生物信息学研究的长期支持。感谢河北师范大学宣益波等同学为本文章所做的公益性劳动。感谢深圳华大基因科技有限公司公开2019-nCoV基因组序列用于验证我们结果的可靠性。

此论文于2020年1月24日提交到https://www.researchgate.net/publication/338804501_A_furin_cleavage_site_was_discovered_in_the_S_protein_of_the_2019_novel_coronavirus。同时,此论文于1月27日通过了中科院预印本平台的审核<http://www.chinaxiv.org/abs/202002.00004v1>。

参考文献(References)

[1]陈嘉源,施劲松,丘栋安,等. 2019新型冠状病毒基因组的生物信息学分析[J]. 生物信息学, 2020, 18(2): 96-102. DOI: 10.12113/202001007.
CHEN Jiayuan, SHI Jinsong, YAU Tungon, et al. Bioinformatics analysis of the 2019 novel coronavirus genome[J]. 2020, 18(2): 96-102. DOI: 10.12113/202001007.

[2]BOSCHB J, VAN DER ZEE R, DE HAAN C A M, et al. The coronavirus spike protein is a class I virus fusion pro-

tein: Structural and functional characterization of the fusion core complex[J]. Journal of Virology, 2003, 77(16): 8801-8811. DOI: 10.1073/pnas.0503203102.

- [3]MATSUYAMA S, UJIKE M, MORIKAWA S, et al. Protease-mediated enhancement of severe acute respiratory syndrome coronavirus infection[J]. Proceedings of the National Academy of Sciences, 2005, 102(35): 12543-12547. DOI: 10.1073/pnas.0503203102.
- [4]GAO Shan, OU Jianhong, XIAO Kai. R language and bioconductor in bioinformatics applications (Chinese Edition) [M]. Tianjin: Tianjin Science and Technology Translation Publishing Ltd, 2014.
- [5]BUCHAN D W A, JONES D T. The psipred protein analysis workbench: 20 years on[J]. Nucleic Acids Research, 2019, 47(1): W402-W407. DOI: 10.1093/nar/gkz297.
- [6]MOLLOY S, BRESNAHAN P, LEPPLA S H. et al. Human furin is a calcium-dependent serine endoprotease that recognizes the sequence Arg-XX-Arg and efficiently cleaves anthrax toxin protective antigen[J]. Journal of Biological Chemistry, 1992, 267(23): 16396-16402. DOI: 10.1111/j.1432-1033.1992.tb17174.x.
- [7]DE HAAN C A M, STADLER K, GODEKE G J, et al. Cleavage inhibition of the murine coronavirus spike protein by a furin-like enzyme affects cell-cell but not virus-cell fusion[J]. Journal of Virology, 2004, 78(11): 6048-6054. DOI: 10.1128/JVI.78.11.6048-6054.2004.
- [8]BELOUZARD S, CHU V C, WHITTAKER G R. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites[J]. Proceedings of the National Academy of Sciences, 2009, 106(14): 5871-5876. DOI: 10.1073/pnas.0809524106.
- [9]GUO Qian, LI Mo, WANG Chunhui, et al. Host and infectivity prediction of Wuhan 2019 novel coronavirus using deep learning algorithm[J]. BioRxiv, 2020, (2020): 1-10. DOI: 10.1101/2020.01.21.914044.
- [10]JI Wei, WANG Wei, ZHAO Xiaofeng, et al. Homologous recombination within the spike glycoprotein of the newly identified coronavirus may boost cross-species transmission from snake to human [J]. Journal of Medical Virology, 2020, 92(4): 433-440. DOI: 10.1002/jmv.22099.

[责任编辑:吴永英]