Sequence Analysis Indicates that 2019-nCoV Virus Contains a Putative Furin Cleavage Site at the Boundary of S1 and S2 Domains of Spike Protein

Z. Galen Wo PhD
USA

Abstract

The infectious 2019-nCoV virus, which caused the current novel coronavirus pneumonia epidemic outbreak, possesses a unique 4-Amino Acid insert at the boundary of the two subdomains (S1 and S2) of Spike protein based on multiple protein sequence alignment with the large SARS and SARS-related virus family. Using Bat CoV_RaTG13 Spike protein as reference (sharing 97% aa identity) the 4-amino acid insert can be identified as PRRA (AA position 681-684). The effect of the 4-AA insertion is the presence of a furin signature sequence motif (PRRARSV) at the boundary of S1 and S2 domains of spike protein. This sequence motif consists the required Arg residue for P1 and P4 position of Furin site. In addition, it contains Arg at P3 site as well as Ser at P1' site of furin motif. This sequence motif matches Aerolysin furin site in FurinDB and was predicted to be moderately strong (score 0.62) by ProP, a protease cleavage site prediction program. This finding suggests that the infectious 2019-nCoV virus, unlike SARS viruses, may be processed via cellular furin recognition and cleavage of the spike protein before host cell membrane fusion and entry. This putative furin site in spike protein of 2019-nCoV virus, if proven to be functional, suggests the potential of looking into agents inhibiting furin as therapeutic mean for the treatment of the novel coronavirus pneumonia.

Key Words: 2019-nCoV, Spike Protein, Furin, Cleavage, SARS CoV-2, Infectivity

The novel coronavirus pneumonia epidemic caused by the infectious 2019-nCoV challenges the scientific community worldwide to study this new virus in order to understand its origin, molecular and cellular mechanism, transmission and pathogenicity (Ref 1, 2, 3, 4). The Spike protein of coronavirus is known to be crucial for host cell receptor binding, membrane fusion and cellular entry. 2019-nCoV Spike protein has been studied in recent reports (Ref 6). Comparison of Spike protein of 2019-nCoV and the reported bat CoV_RaTG13 (sharing 96% genome nt identity) showed high degree of conservation (97% aa identity), yet 2019-nCoV spike protein contains a 4-amino acid insert (position 681-684) which doesn't exist in CoV_RaTG13. Therefore, a detailed analysis was carried out regarding this 4-aa insert in 2019-nCoV Spike protein.

Using 2019-nCoV Spike Protein as query BlastP against NCBI non-redundant protein sequences (nr) database and tBlastn against Nucleotide Collection (nr/nt) both returned results which clearly show that 2019-nCoV Spike protein possesses a 4-amino acid insert in a relatively conserved region of Spike protein (Figure 1). This 4-amino acid insert does not exist in any of the SARS and SARS-related viruses. Note the 4-amino acid gap is not all aligned in the multiple alignment of SARS-related Spike proteins with some shifted one amino acid to the left. Using Bat CoV_RaTG13 Spike protein as reference the 4-amino acid insert can be identified as PRRA (aa position 681-684).

This insertion occurred at the boundary of the two subdomains of Spike protein. The first (S1) is for receptor binding and the second (S2) is for fusion with host cell membrane and virus entry (Figure 2). Each of the two domains folds as an independent unit (Ref 7). And it's known coronavirus activation requires the cleavage of the two domains. While type 2 and type 3 coronavirus (including influenzas and HIV) cleavage involves furin-dependent mechanism, SARS virus cleavage is not Furin dependent as SARS virus doesn't possess furin recognition motif and the cleavage is likely via endosome/ lysosomal Cathepsins as well as other proteases (Ref 7,8). However, with the insertion of 4-amino acid (PRRA) in 2019-nCoV spike protein a putative Furin recognition motif (PRRARSV) started to emerge.

The consensus recognition sequence for furin protease is X-Arg-X-Lys/ Arg-Arg-Y where X is a hydrophobic amino acid and the first Arg and the final Arg are most crucial. Furin cleaves between the final Arg and the Y residue which is Ser with many viruses (Ref 9, 10). In addition, the ability for furin to cleave a protein substrate depends on its tertiary structure as well as on the amino acids immediately surrounding the cleavage site. The motif (PRRARSV) generated by the 4-aa insert in 2019-nCoV Spike protein meets the general requirement for furin cleavage. To further assess the likelihood, this sequence motif was utilized as a query to search FurinDB which contains all known furin cleavage site motifs (Ref 11, 12). The search returned one hit – bacteria Aerolysin (aerA; Protein ID: 113485) providing supporting evidence that the furin sequence motif at the S1/ S2 site of 2019-nCoV may be functional (Figure 3a and 3b).

The furin sequence motif at S1/ S2 site is missing in all SARS/ SARSr virus Spike protein, raising an interesting question on whether this furin sequence motif in 2019-nCoV is entirely novel or