High Level of Similarity in Amino Acid Sequence of Surface Proteins of SARS and SARS-Cov-2

Wenfa Ng*

Department of Biomedical Engineering, National University of Singapore

*Corresponding author: Wenfa Ng, Department of Biomedical Engineering, National University of Singapore, Email: ngwenfa771@hotmail.com ; ngwenfa@nus.edu.sg

Abstract

Biological molecules are related to one another. Such evolutionary connectedness means that we could probe the origins of one molecule based on some characteristics (usually sequence) of a related molecule. Such conceptual tools have been codified as sequence alignment and phylogenetic tree software used by bioinformatician on a daily basis to search for species relationships at the sequence level. This work uses sequence alignment to probe for the relatedness of surface proteins of SARS and SARS-CoV-2, with the aim of gaining an understanding of possible sequence and structure conservation of the surface proteins (S, N, E, and M) and their implications in clinical diagnostics and treatment. Results revealed high level of similarity of all surface protein amino acid sequence for both SARS and SARS-CoV-2. This implies that this set of surface proteins have evolved under tight constraints, and may be selected for by a common natural host of both coronaviruses. In addition, SARS and SARS-CoV-2, as judged by sequence conservation of surface proteins, are related viruses possibly belonging to the same virus family. Given that sequence conservation implies similar protein structure, diagnostics and treatment developed for SARS should be readily translatable to SARS-CoV-2 if the protein in question is a viral surface protein. Of biggest surprise in the work is the revelation that E and M protein exhibit very high level of sequence conservation across SARS and SARS-CoV-2 which speaks of their essentiality to the pathogenesis and function of the coronaviruses. Such conservation implies that both proteins may be targets for therapeutic and diagnostic development in anticipation of future coronavirus outbreak from the same virus family. Overall, sequence alignment is used in this work to reveal the high level of conservation of surface proteins across SARS and SARS-CoV-2 at the amino acid sequence level. Such conservation implies relatedness between the two coronaviruses, but more importantly, point to avenues for which the biotechnology and pharmaceutical industries could exploit for diagnostic and therapeutic development.

Keywords: Sequence Alignment; Sequence Conservation; SARS; SARS-Cov-2; Evolutionary Relationship

Introduction

Sequence alignment is one of the primary tools in the toolkit of bioinformaticians and disease epidemiologists in understanding the origins and species relatedness of new pathogens circulating in different parts of the world. Originally developed to understand how one nucleotide sequence or amino acid sequence is related to another, sequence alignment, whether at the pairwise or multiple sequence level, helps illuminate the evolutionary ingrained

marks that separate one species from another at the protein level. This work utilises the primary tool of sequence alignment to uncover possible conservation in sequence of the different surface proteins of SARS-CoV and SARS-CoV-2. Proteins investigated include spike protein (S), membrane glycoprotein (M), nucleocapsid protein (N), and envelope protein (E). These proteins dot the surface of the coronavirus, and are thus, hugely important given that binding between one or more of these surface proteins and receptors on human cells would initiate cell entry of the virus. At a deeper level, understanding the conservation in sequence of different surface proteins of SARS and SARS-CoV-2 hold multiple lines of implications.

Firstly, similarity in sequence implies that particular proteins on the surface of SARS and SARS-CoV-2 are related. This means that SARS and SARS-CoV-2 may be evolutionary related, be in the same virus family, or share the same natural host, where, in the latter, the host exerts evolutionary pressure to select for particular shape and sequence of viral surface proteins. Secondly, from the disease diagnostic and treatment perspective, similarity in sequence, and thus, shape of surface proteins of SARS and SARS-CoV- 2 suggests that technologies and diagnostics developed for detecting SARS could be applied, with minimal alterations, to the detection of SARS-CoV-2. Finally, conservation in sequence and structure of surface proteins of SARS and SARS-CoV-2 confirm that the two viruses use similar routes to enter human cells, and suggest that treatments for SARS could be repurposed, at a different efficiency level, for treating SARS-CoV-2 infection.

A result from sequence alignment analysis using the swalign function in MATLAB Online strongly suggests high level of similarity in the amino acid sequence of different surface proteins in SARS and SARS-CoV-2. Specifically, most regions of each surface protein aligned well between the variant in SARS and SARS-CoV- 2 with differences at two to three amino acids at a stretch. This suggests that the surface proteins of SARS and SARS-CoV-2 have similar shape, which means that antibodies targeting SARS surface proteins should also find a similar purpose when applied to SARS-CoV-2. More importantly, SARS and SARS-CoV-2 are related to each other judging from the similarity in their surface proteins' amino acid sequence and structures. But, given the unique presence of ORF10 in SARS-CoV-2 genome, SARS and SARS-CoV-2 are likely in the same virus family, share the same natural host, but SARS-CoV-2 did not evolve from SARS. This work highlights from the sequence conservation perspective that high level of similarity in sequence does not suggests that, at the virus or microbe level, one virus evolve from another. In essence, evolution could exert micro-level effects in selecting for similar solutions for problems such as finding a ligand to bind with high affinity to a human receptor for gaining entry into host cells.

Materials and Methods

Genome sequences of SARS and SARS-CoV-2 were obtained from National Center for Biotechnology Information (NCBI)'s Genbank. The annotated genome sequences were parsed into gene database comprising gene identifier, gene function, and gene sequence using in-house MATLAB genome analysis software. Subsequently, gene sequence of each gene in the respective genome of SARS and SARS-CoV-2 were translated into amino acid sequence that forms the basis of this sequence alignment analysis. Amino acid sequence of spike (S), membrane glycoprotein (M), nucleocapsid (N) and envelope E protein of SARS and SARS-CoV-2 were consolidated into respective FASTA file for analysis by the swalign algorithm in MATLAB Online. Sequence alignment results were depicted in the seqalignviewer app of MATLAB Online.

Results and Discussion

Spike protein (S) is the primary mode by which SARS and SARS-CoV-2 bind to and gain entry into human cells. Specifically, spike proteins of both coronaviruses bind to the ACE2 receptor of human cells with the spike protein of SARS-CoV-2 binding the receptor with stronger affinity with a purported slightly different structure [1]. Figure 1 shows the sequence alignment results of amino acid sequence of spike proteins of the two coronaviruses. Results reveal that both spike proteins show high level of sequence conservation at most regions of the protein sequence. However, there are differences in amino acid residue at various locations in the protein sequence, and this may result in altered shape of the protein with corresponding changes in its binding affinity to ACE2 receptor of human cells. Overall, high level of similarity of amino acid sequence of spike proteins of SARS and SARS-CoV-2 suggests that both viruses share the same virus family, are related, and likely have the same natural host. More importantly, similarity in amino acid sequence suggests similar protein structure, which means that antigen rapid tests for SARS could be applied to detect SARS-CoV-2, and drugs that interfere with the function of S protein in SARS could be used to treat SARS-CoV-2 infection.

Wenfa Ng. High Level of Similarity in Amino Acid Sequence of Surface Proteins of SARS and SARS-Cov-2. Int J Biochem Physiol 2021, 6(2): 000198.

Copyright© Wenfa Ng.
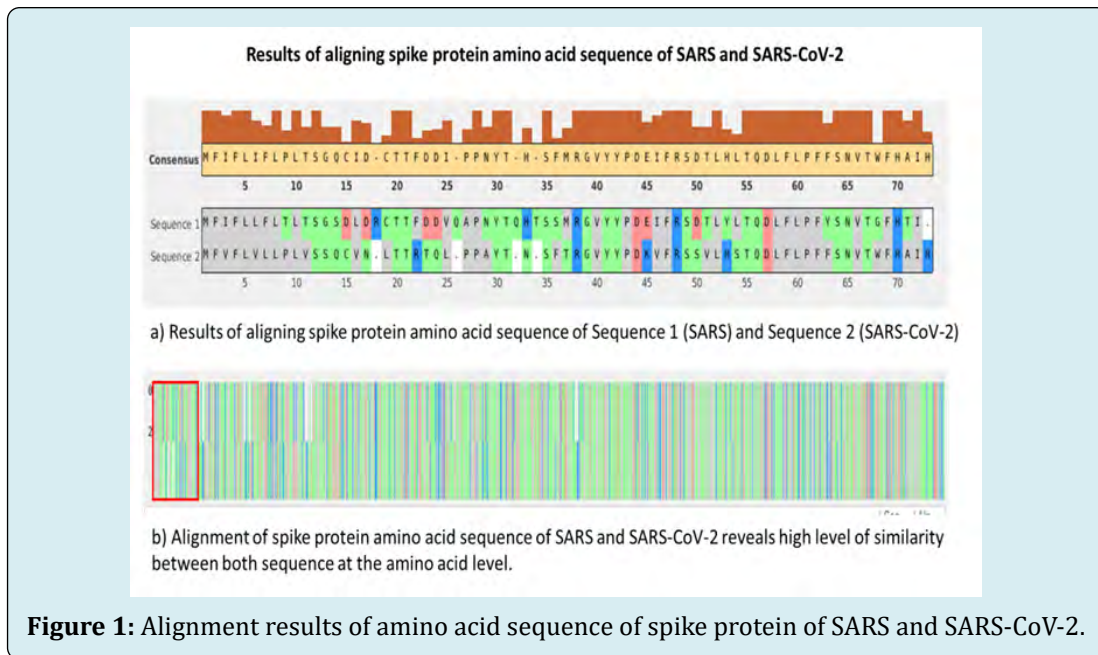
Figure 1: Alignment results of amino acid sequence of spike protein of SARS and SARS-CoV-2.

Similar to spike protein, the nucleocapsid protein (or N protein) is another surface protein on SARS and SARS-CoV-2. Investigating the sequence conservation of this protein in SARS and SARS-CoV-2 holds relevance because many of the antigen rapid tests for SARS-CoV-2 target this protein. Figure 2 shows the alignment results of the amino acid sequence of N protein of SARS and SARS-CoV-2. Results reveal that, except for sporadic locations in the amino acid sequence of the N protein, most regions of the protein show high level of similarity which suggests that the nucleocapsid protein of SARS and SARS-CoV-2 have similar structure. This is important for understanding the evolutionary relationship between SARS and SARS-CoV-2 from the perspective of N protein structure. Overall, the data suggests that the protein structure of nucleocapsid protein in SARS and SARS-CoV-2 should be similar, which implies that antigen rapid tests for SARS could be applied to detect SARS-CoV-2. Both viruses should be in the same family and belongs to the same natural host.
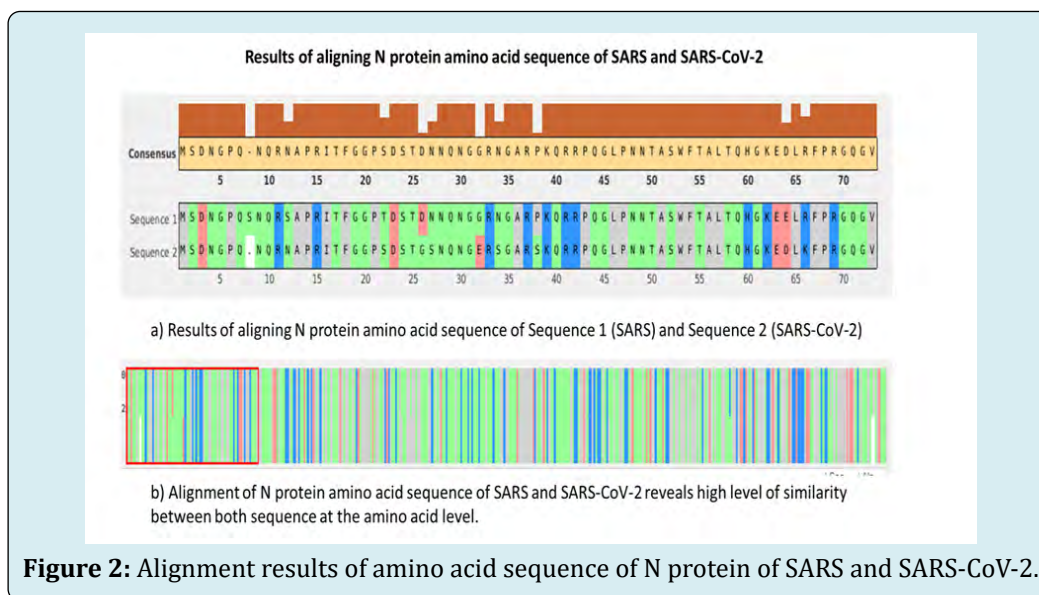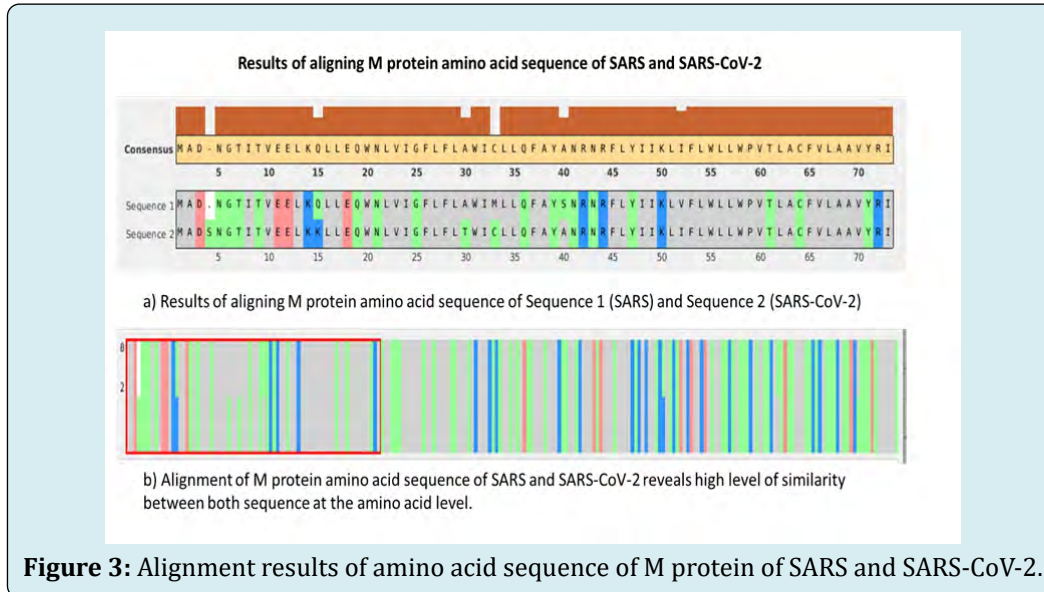

Figure 2: Alignment results of amino acid sequence of N protein of SARS and SARS-CoV-2.

Membrane glycoprotein (or M protein) is another surface protein that dots the viral surface of SARS and SARS-CoV-2. Unlike spike and nucleocapsid protein, M protein is less studied, and is not a major target for the development of RT-PCR or antigen rapid test or antibody test. Figure 3 shows the alignment result of the amino acid sequence of M protein

Wenfa Ng. High Level of Similarity in Amino Acid Sequence of Surface Proteins of SARS and SARS-Cov-2.
Int J Biochem Physiol 2021, 6(2): 000198.

Copyright© Wenfa Ng.

of SARS and SARS-CoV-2. Results reveal very high level of similarity in amino acid sequence of M protein of both coronaviruses. This suggests that M protein may be critical for virus function, and thus, there is evolutionary pressure to maintain its amino acid sequence and protein structure.

Indeed, there has been research describing the essential role of M protein in aiding the association of other viral structural proteins [2]. Hence, M protein may be a suitable target for development of neutralizing antibodies for treating SARS and SARS-CoV-2 infection.

Figure 3: Alignment results of amino acid sequence of M protein of SARS and SARS-CoV-2.

Envelope protein is another surface protein on SARS and SARS-CoV-2. Similar to M protein, E protein is less studied. Figure 4 shows the alignment of amino acid sequence of E protein of SARS and SARS-CoV-2. Results reveal very high level of similarity and sequence conservation of E protein on SARS and SARS- CoV-2. E protein is currently not a major target of diagnostics for SARS or SARS-CoV-2. But, high level

of similarity between the E protein of SARS and SARS-CoV-2 suggests a strong evolutionary pressure selecting for the structure of E protein. Hence, E protein, like S protein, is likely to have an essential role in viral pathogenesis or function. Future diagnostics may look into targeting E protein for detecting SARS or SARS-CoV-2 infection.
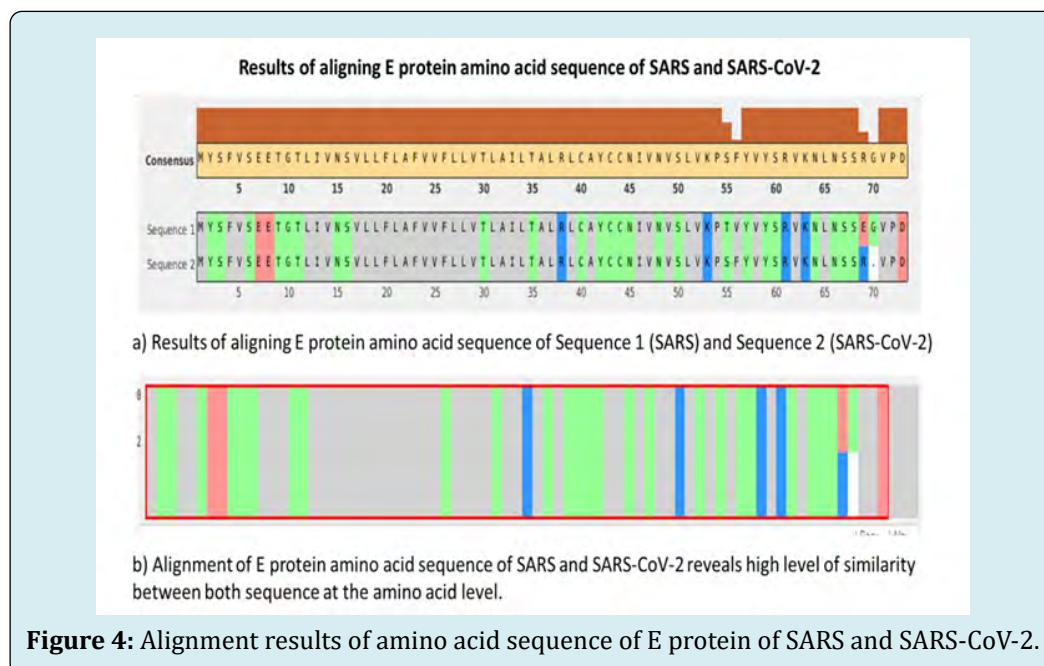
Figure 4: Alignment results of amino acid sequence of E protein of SARS and SARS-CoV-2.

Wenfa Ng. High Level of Similarity in Amino Acid Sequence of Surface Proteins of SARS and SARS-Cov-2. Int J Biochem Physiol 2021, 6(2): 000198.

Copyright© Wenfa Ng.

Conclusions

Sequence alignment remains the primary tool for bioinformaticians to discern evolutionary relationships between two proteins or nucleotide sequence. Applying sequence alignment analysis to discerning the relationship of different viral surface proteins in SARS and SARS-CoV-2 reveal high level of similarity in amino acid sequence of S, E, M, and N proteins of SARS and SARS-CoV-2. Similarity in amino acid sequence implies similar structure, which means that these viral surface proteins have evolved to bind to receptors in their natural host before the virus making the jump to infect humans. This then meant that SARS and SARS-CoV-2 share the same virus family and natural host. More importantly, similarity in protein structure of surface protein in both coronaviruses mean that diagnostics developed for SARS could be repurposed for SARS-CoV-2. Finally, unusually high level of sequence conservation between the M and E protein of SARS and SARS-CoV-2 suggests that both proteins are essential to pathogenesis or function of both coronaviruses, which from the treatment perspective, suggests that treatment for SARS could be applied for SARS-CoV-2. Such high level of conservation of M and E protein in SARS and SARS-CoV-2 also suggests that they could be targets for the development of future diagnostics of the two coronaviruses.

References

1. Xie Y, Karki CB, Du D, Li H, Wang J, et al. (2020) Spike Proteins of SARS-CoV and SARS-CoV-2 Utilize Different Mechanisms to Bind With Human ACE2. Front Mol Biosci 7: 591873.

2. Alharbi SN, Alrefaei AF (2021) Comparison of the SARS-CoV-2 (2019-nCoV) M protein with its counterparts of SARS-CoV and MERS-CoV species. J King Saud Univ - Sci 33(2): 101335.