Northeast Journal of Contemporary Research, December 2022. Vol. 9 No. 1, pp 1-7
ISSN 2349-3089

Mutational analysis of severe acute respiratory syndrome coronavirus 2 spike (S) protein

Debashree Saikia
Department of Biotechnology,
Darrang College, Tezpur-784001, Assam, India

ABSTRACT

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a newly emerged coronavirus responsible for coronavirus disease 2019 (COVID-19); it become a pandemic since March 2020 and is spreading rapidly to people across the globe. SARS-CoV-2 attaches to the host cell surface to initiate the interaction between the receptor-binding domain (RBD) of its spike glycoprotein (S) and the human Angiotensin-converting enzyme (hACE2) receptor. SARS-CoV-2 mutates frequently, which challenges the antiviral development. Mutations can cause viruses to better evade host's immune systems, treatments and vaccines. Analysis of mutation of spike protein of SARS-CoV-2 isolated from different countries is important to understand the biology and pathogenicity of virus. In this study, mutational analysis of SARS-CoV-2 structural spike (S) protein was done by downloading 1311 sequences with high coverage genome of SARS-CoV-2 using the GISAID database. Moreover, docking study of spike protein with human ACE2 was done to understand whether the mutation helps the protein to bind more efficiently and hence becomes potentially more infectious. Further study mutation hot-spots will be helpful in designing suitable drugs and other therapeutics for the treatment of SARS-CoV-2.

Keywords: SARS-CoV-2, COVID-19, Mutation, Structural spike (S) protein, GISAID database

INTRODUCTION

The emergence of rapidly spreading variants of SARS-CoV-2, the causative agent for COVID-19, threatens to prolong an already devastating pandemic. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the newest member of *Coronaviridae* family. After the COVID-19 (SARS-COV-2) infection broke out suddenly in Wuhan, China, it had spread across more than 200 countries worldwide affecting 70,476,836 people and causing 1,599,922 deaths as of 15th December 2020 (https://covid19.who.int/). The World Health Organization (WHO) declared this as a public health emergency of international concern (PHEIC) on 30th January 2020 and a pandemic on 11th March. Till date 500,186,525 infections and 6,190,349 fatalities occurred due to this corona virus pandemic as per WHO record. SARS-CoV-2 has some genetic similarity with MERS CoV (Middle East Respiratory Syndrome Coronavirus) and SARS CoV (Severe Acute respiratory Syndrome Coronavirus). Although there are some resemblances, a detailed study of the proteins helps us understand their differences in pathogenicity (Mousavizadeh and Ghasemi, 2020).

*Corresponding author's Email: debashree.saikia@gmail.com

The genome of SARS-CoV-2 shares about 80% identity with that of SARS-CoV and is about 96% identical to the bat coronavirus BatCoV RaTG13 (Zhou *et al*., 2020). Coronaviruses are enveloped, non-segmented, positive sense single-stranded RNA virus genomes in the size ranging from 26-32 kb, the largest known viral RNA genome. The virion has a nucleocapsid composed of genomic RNA and phosphorylated nucleocapsid (N) protein, which is buried inside phospholipid bilayers and covered by two different types of spike proteins: the spike glycoprotein homotrimer (S) that can be found in all CoVs, and the hemagglutinin esterase (HE) that exists in some CoVs. The membrane (M) protein (a type III transmembrane glycoprotein) and the envelope (E) protein are located among the S proteins in the virus envelope.

Sarbecoviruses express a large (approximately 140 kDa) glycoprotein termed spike protein (S, a homotrimer), which mediates binding to host cells via interactions with the human receptor angiotensin converting enzyme 2 (ACE2) (Letko *et al*, 2020; Wrapp, 2020; Walls *et al*., 2020). Coronavirus entry into host cells is mediated by the transmembrane spike (S) glycoprotein that forms homotrimers protruding from the viral surface. The spike glycoprotein (S protein) on the virion surface mediates receptor recognition and membrane fusion. During viral infection, the trimeric S protein is cleaved into S1 and S2 subunits and S1 subunits are released in the transition to the postfusion conformation. S1 contains the receptor binding domain (RBD), which directly binds to the peptidase domain (PD) of angiotensin -converting enzyme 2 (ACE2), whereas S2 is responsible for membrane fusion. When S1 binds to the host receptor ACE2, another cleavage site on S2 is exposed and is cleaved by host proteases, a process that is critical for viral infection. The S protein is highly immunogenic with the receptor-binding domain (RBD) being the target of many neutralizing antibodies (Berry *et al*., 2010). Individuals infected with coronaviruses typically mount neutralizing antibodies and a neutralizing response has been demonstrated for SARS-CoV-2 in an individual case from day 9 onwards against spike protein (Huang *et al*., 2020; Haveri *et al*., 2020). Therefore, analysis of mutation of spike protein of SARS-CoV-2 isolated from different countries is important to understand the biology and pathogenicity of virus. As on 12th May 2020, more than 15000 whole genomes of SARS-CoV-2 isolated from different countries have been deposited in the GISAID database.

Generally, the potential vaccine candidates comprise of either inactivated or live attenuated or subunit viruses, or DNA or RNA vaccines. Mutation becomes a very important factor in determining the sustainability of a vaccine. High mutation rate of a virus or its proteins sometimes makes the vaccine less effective after a period of time. In this study, analysis of the mutations in SARS-CoV-2 structural spike (S) protein, the most important viral protein required for entry into host cells, was done and also try to understand the regions in its genome where mutation is playing an important role so that it may help us to understand the dynamics of its evolution and guide us in designing a sustainable vaccine.

MATERIALS AND METHODS

Retrieval and curation of SARS-CoV-2 full genome sequence of Indian isolates

1311 sequences have been downloaded with high coverage genome of SARS-CoV-2 on 17th February 2021 using the GISAID database (https://www.gisaid.org/). Different criteria were applied such as gene size mismatch, internal stop codons, deletion, and ambiguous nucleotides other than A/T/G/C to study substitution polymorphism in the genome. The SARS-CoV-2 isolate Wuhan-Hu-1 genome consisting of 29,903 bases have been used as reference genome (NC_045512.2). The coding region (CDS) of each gene that encodes for structural, nonstructural, and accessory proteins including the regions of untranslated (UTR) ends of the genome have been extracted and used as query sequence to perform local BLAST and the filtered

Table 1. Preparation of Consensus sequences

Strains	N1	N2	N3	N4	N5	N6	N7	N8
1	A	G	T	C	A	A	G	A
2	A	A	T	C	T	A	C	C
3	A	G	C	C	A	A	G	C
4	A	G	T	G	A	A	G	G
5	G	G	T	C	A	A	A	C
6	A	T	C	T	T	A	G	C
7	A	G	T	G	A	A	G	C
8	A	G	C	C	T	A	G	T
9	T	G	T	C	A	A	A	C
10	A	G	T	T	A	A	G	C
Count of A	8	1	0	0	7	10	2	1
Count of G	1	8	0	2	0	0	7	1
Count of C	0	0	3	6	0	0	1	7
Count of T	1	1	7	2	3	0	0	1
CON (Most frequent nucleotide)	A	G	T	C	A	A	G	C

Table 2. Strain wise analysis the number of mutations

Strains	N1	N2	N3	N4	N5	N6	N7	N8	No. of Mutations
CON	A	G	T	C	A	A	G	C	0
1	A	G	T	C	A	A	G	A	1
2	A	A	T	C	T	A	C	C	3
3	A	G	C	C	A	A	G	C	1
4	A	G	T	G	A	A	G	G	2
5	G	G	T	C	A	A	A	C	2
6	A	T	C	T	T	A	G	C	4
7	A	G	T	G	A	A	G	C	1
8	A	G	C	C	T	A	G	T	3
9	T	G	T	C	A	A	A	C	2
10	A	G	T	T	A	A	G	C	1

sequence of 1059 strains were finally retained out of 1311 strains. Filtration of sequence and polymorphism analysis was performed using several computer programs written in Python language. A procedure of finding a consensus sequence and inferring polymorphism employed in this study is provided as Table 1 & Table 2.

If random, transversions (purine–pyrimidine changes) should be observed twice as often as transitions (purine to purine or pyrimidine to pyrimidine changes) solely due to the accessible mutations. In the present study observed ratio of ti/tv was close to 1.3 to 1.4 across the coding regions of structural and accessory genes and ti/tv of 2.8 was observed for genes encoding non-structural proteins. Previously it was reported that RNA virus polymerases incorporate transition mutations at a frequency of 10^{-5} and transversions mutations at a frequency of 10^{-6} to 10^{-7} and the structural basis for fidelity of nucleotide selection is understood best for the RdRP of Poliovirus (3Dpol) (Castro *et al*., 2005). In fact, a previous report showed that RNA-dependent RNA polymerase of Influenza virus can make fewer trans-

versions than transitions (Pauly and Lauring, 2015).

In this study, computational structural biology methods was used to analyze the role of mutations in the severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) variant, and the infectivity and immune escape properties.

Molecular Docking study

The review and comparative assessment of sequences of S protein among available SARS-CoV-2 genomes in the GISAID database showed three major mutations viz., H49Y, D614G, and T573I. Multiple crystallized 3D structures of S protein can be found in the Protein Data Bank (PDB) (Walls *et al*., 2020). However, to capture conformational changes in the 3D atomistic model, it is necessary to perform molecular dynamics (MD) simulations (Sikora *et al*., 2020). To determine if the mutated S proteins could affect its 3D structure conformations, here MD simulations were employed for full length atomistic models of S protein 3D structure and its mutant D614G. Autodock 4.2 software was used for docking calculations and docking results were analyzed using Autodock Tools 1.5.6 (Morris *et al*., 2009).

RESULTS

Coronavirus entry into host cells is mediated by the transmembrane spike (S) glycoprotein that forms homotrimers protruding from the viral surface (Tortorici and Veesle, 2019). S protein comprises two functional subunits, responsible for binding to the host cell receptor (S1 subunit) and fusion of the viral and cellular membranes (S2 subunit). For many Coronaviruses, S is cleaved at the boundary between the S1 and S2 subunits, which remain non-covalently bound in the pre-fusion conformation. This region is reported to be the most potent and indispensable for viral attachment and entry into host system (Walls *et al*., 2020). The miss-sense mutations in S protein those it was found are mostly single point mutations with few double and triple mutations (Table 3).

Any alteration could impact the structural and functional conformation of the protein which makes it more infectious. Although the clinical significance of the observed mutations is not readily available, our findings in Indian patients lay the ground work for India to understand the impact of SARS-CoV2 mutations on disease severity, host immune response, vaccine development and serological response.

The Spike (S) glycoprotein helps the virus to attach with ACE2 (Angiotensin-converting enzyme 2) and TMPRSS2 (Transmembrane serine protease 2). It is of length 1273 amino acids. Mul-

Table 3. Mutational Analysis of India (25th July 2020 to 17th February 2021)

Months	0	1	2	3	4	5	6	Total strains
July(2020)	50	33	10	2	1	0	1	97
August(2020)	211	159	89	24	4	3	0	490
September(2020)	71	146	56	25	9	2	1	310
October(2020)	11	10	2	3	3	0	0	29
November(2020)	17	14	10	4	2	0	0	47
December(2020)	6	10	24	8	6	0	0	54
January(2021)	3	4	14	3	4	4	0	32

tiple sequence alignment of SARS-CoV-2 S-proteins revealed unique mutations. Total 1059 isolates showed at least one mutation in the S protein sequences (Table 2).

In this study, the docking results are incorporated to make a comparative study of the interaction energy between the host proteins and the SARS-CoV-2 viral proteins. That is to check whether the mutation helps the protein to bind more efficiently and hence becomes potentially more infectious. In this regard, the spike glycoprotein becomes the most important viral protein as it is involved in direct contact with the host ACE-2 receptor which helps the virus to enter the human cells. The 3-dimensional structures were searched for both the wild and mutated variety of the spike glycoprotein in the protein data bank. The crystal structure of the SARS-CoV-2 spike protein bound to the human ACE2 receptor (PDB code: 6M0J) and the structure of the human ACE2 receptor (PDB code: 1R42) were downloaded from the Protein Data Bank.

Though the 3-dimensional structure for wild spike protein was found, unfortunately the complete structure of the mutated D614G spike protein was unavailable. The only structure available lacked the receptor binding domain (PDB ID: 6XS6). Considering the fact that the receptor binding domain in the spike protein plays the key role in the binding of human ACE-2 and viral spike protein, it was at first apprehensive that docking the ACE-2 with an incomplete spike protein will not give the correct results. Thus, the ultimate goal of comparison remains unfulfilled due to the lack of protein structures in the database. Although, there is incompleteness in the pdb files, the blind docking was performed and the docking structures of both wild type (WT) as well as mutated Spike protein (D614G) with human ACE2 are given below in **Figure 1**.

D614G showed a more compact cluster distribution than the other mutated proteins, suggesting a reduction in conformational mobility due to single residue mutation; additionally, it showed a dissimilar conformer distribution along the subspace in comparison to the others mutated proteins systems, this behavior suggests that the trajectory sampled different regions of the phase space with different minima and small energy barrier, in this sense, D614G mutant affects the structural behavior of the protein.

DISCUSSION

The SARS-CoV-2 infection caused over 271.9

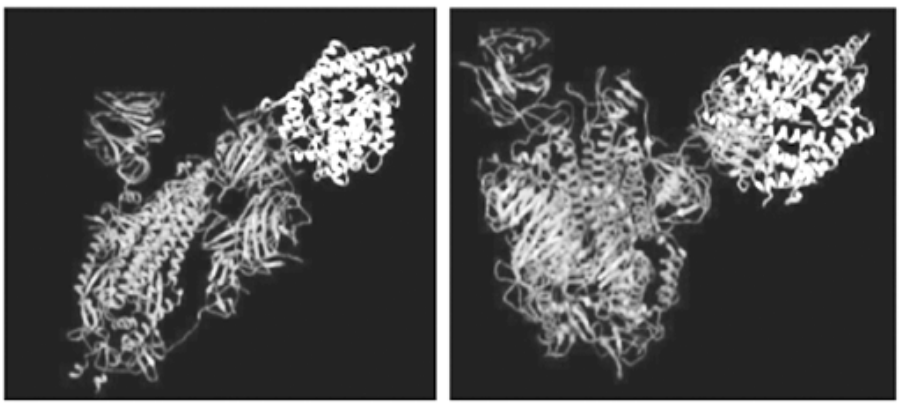

Figure 1. Docking Image of Spike Protein with human ACE2. Left image is of Wild Protein Docking and Right Image is of mutated protein (D614G spike protein) docking.

million confirmed cases of Coronavirus Disease 2019 (COVID-19), including over 5.3 million deaths. Despite mitigating efforts across the world, emerging variants continue to threaten the health of individuals, burden healthcare support, and destabilize the economy.

A significant mutation is observed in the Spike glycoprotein at position 614, where an aspartic acid (polar) is changed to a glycine (nonpolar). The mutation is in fact destabilizing the native spike protein (having both the S1–S2 domains) which may eventually influence its cleavage. This mutation lies close to the S1–S2 junction of the spike protein and it was found out that the point mutation has developed an additional cleavage site for elastase 2. From previous studies on coronavirus, it has been found that proteolysis at several points of the spike glycoprotein is essential for its entry inside the cell (Belouzard *et al*., 2009). So, it can be concluded from the data that the generation of a novel protease site at the vicinity of the S1–S2 junction has helped the virus enter the host cell more efficiently. This indicates that the mutated protein may be increasing the potential of the virus to attach with host receptors and undergo cleavage.

The genome of SARS-CoV-2 for mutations prevalent around the world was surveyed. From this study it can be stated that mutations in the proteins of SARS-CoV-2 are slow yet steady. It was also observed how the wild and mutated spike proteins tussled with each other and ultimately the mutated protein became more widespread.

The spike protein amino acid change D614G was noted to be increasing in frequency in April 2020 and to have emerged several times in the global SARS-CoV-2 population, and the coding sequence exhibits a high dN/dS ratio, suggesting positive selection at the codon position 614 (Korber, B. *et al*, 2020).Subsequent studies indicated that D614G confers a moderate advantage for infectivity (Hou *et al*, 2020; Yurkovetskiy *et al*, 2020) and transmissibility (Volz, E. *et al*, 2020). Several other spike mutations of note have now arisen. The extent to which mutations affecting the antigenic phenotype of SARS-CoV-2 will enable variants to circumvent immunity conferred by natural infection or vaccination remains to be determined. However, there is growing evidence that mutations that change the antigenic phenotype of SARS- CoV-2 are circulating and affect immune recognition to a degree that requires immediate attention. Consequently, mutations that affect the antigenicity of the spike protein are of particular importance. Prediction of the mutational pathways by which a virus such as SARS-CoV-2 will evolve is extremely challenging. Nonetheless, there is a rapidly expanding knowledge base regarding the effect of SARS-CoV-2 spike mutations on antigenicity and other aspects of virus biology. The integration of these data and emerging SARS-CoV-2 sequences has the potential to facilitate the automated detection of potential variants of concern at low frequency that is, before they are spreading widely.

CONCLUSION

A thorough study of the mutations that have occurred in various proteins encoded by the SARS-CoV-2 genome can also help researchers and medical personnel in designing suitable drugs and other therapeutics. Designing alternative vaccine strategies like peptide vaccines and mRNA vaccine can be boosted by this study as targeting the conserved regions of the proteins can only be done if one has sound knowledge regarding the mutation hot-spots. Computer-aided drug designing can also be improved with the help of this study. Further advanced study correlating COVID-19 symptoms with subtle mutational changes will be helpful for us to understand the virus better.

ACKNOWLEDGEMENT

Author acknowledges IQAC, Darrang College for the award of “In House Research Project, 2020” for carried out this research work.

REFERENCES

Belouzard, S., Chu, V.C., Whittaker, G.R. 2009. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites, Proc. Natl. Acad. Sci. U. S. A. 106 (14) (2009) 5871–5876, https://doi.org/10.1073/pnas.0809524106

Berry, J. D. *et al*. 2010. Neutralizing epitopes of the SARS-CoV S-protein cluster independent of repertoire, antigen structure or mAb technology. MAbs 2, 53–66.

Castro, C., Arnold, J. J., & Cameron, C. E. 2005. Incorporation fidelity of the viral RNA-dependent RNA polymerase: a kinetic, thermodynamic and structural perspective, Virus Research, 107(2) (2005) 141–149.

Haveri, A. *et al*. 2020. Serological and molecular findings during SARS-CoV-2 Infection: the First Case Study in Finland, January to February 2020 (Eurosurveillance, 2020).

Hou, Y. J. *et al*. 2020. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission *in vivo*. Science 370, 1464.

Huang, A. T. *et al*. 2020. A systematic review of antibody mediated immunity to coronaviruses: antibody kinetics, correlates of protection, and association of antibody responses with severity of disease. Preprint at medRxiv https://doi.org/10.1101/2020.04.14.20065771.

Korber, B. *et al*. 2020. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell 182, 812–827 e819.

Letko, M., Marzi, A. & Munster, V. 2020. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat. Microbiol. 5, 562–569.

Morris, G. M. *et al.*2009. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 2785–2791.

Mousavizadeh, L. and Ghasemi, S. 2020. Genotype and phenotype of COVID-19: their roles in pathogenesis, J. Microbiol. Immunol. Infect. (2020), https://doi.org/ 10.1016/j.jmii.2020.03.022. PMCID: PMC7138183.

Pauly, M.D. and Lauring, A.S. 2015. Effective lethal mutagenesis of influenza virus by three nucleoside analogs, J Virol. 89(7) 3584–3597.

Sikora, M. *et al.*2020. Map of SARS-CoV-2 spike epitopes not shielded by glycans. *BioRxiv.* https ://doi.org/10.1101/2020.07.03.186825.

Tortorici, M.A., and Veesler, D. 2019. Structural insights into coronavirus entry. Adv. Virus Res. 2019; 105; 93–116.

Volz, E. *et al*. 2020. Evaluating the effects of SARS-CoV-2 Spike mutation D614G on transmissibility and pathogenicity. Cell https://doi.org/10.1016/j.cell.2020.11.020.

Walls, A. C. *et al.* 2020. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 181, 281–292.

Walls, A.C., Park, Y.J., Tortorici, M.A., Wall, A., McGuire, A.T., Veesler, D. 2020. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein; Cell. 180; 281–292.

Wrapp, D. *et al*. 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 367, 1260–1263.

Yurkovetskiy, L. *et al*. 2020. Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. Cell 183, 739–751.e738.

Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR,3, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020 Mar; 579(7798):270-273.