



In-depth Sequence Analysis of SARS-CoV-2 Spike Protein Repudiates Mutation Mediated Adaptive Selection in the Virus

Pramita Chowdhury*, Bijurica Chakraborty* and Sanghamitra Sengupta

Department of Biochemistry, University of Calcutta, Kolkata, West Bengal, India

*Corresponding Author: Pramita Chowdhury, Bijurica Chakraborty, Department of Biochemistry, University of Calcutta, Kolkata, West Bengal, India.

DOI: 10.31080/ASMI.2020.03.0729

Received: September 14, 2020

Published: November 18, 2020

© All rights are reserved by **Pramita Chowdhury, Bijurica Chakraborty., et al.**

Abstract

Increasing transmissibility and pathogenicity of SARS-CoV-2 has often been attributed to mutations accumulating in RNA genome of SARS-CoV-2 ancestral type originated in China. In this study, using *in silico* structural modeling and molecular docking, we demonstrate that observed low frequency non-synonymous amino acid alterations do not improve the efficacy of interaction between SARS-CoV-2 spike glycoprotein and host's ACE2 receptor. The major spike haplotype detected corresponds to the ancestral Wuhan isolate which displays the most favorable interaction with ACE2. Although the sequences are different, length and number of predicted B-cell epitopes of SARS-CoV-2 spike protein, identified using BepiPred 2.0, are comparable to that of SARS-CoV. However, the above data differ from the ones from HCoV-NL63. Analysis of population genetic parameters implies relaxed purifying selection to be predominant force shaping the genetic landscape of SARS-CoV-2 spike protein, till date. These findings provide important clues in designing therapeutic strategies to contain the pandemic.

Keywords: SARS-CoV-2; Spike; ACE2; Epitopes; Purifying Selection

Introduction

In December 2019, Chinese public health officials notified the World Health Organization (WHO) about an unknown virus causing disease with pneumonia-like symptoms in Wuhan [1]. The contagion was soon recognized to be a member of coronavirus (CoV) family, a group of zoonotic pathogen and named as severe acute respiratory syndrome coronavirus 2 or SARS-CoV-2 [2]. The disease spread rapidly within three months from its appearance to more than 200 countries as this virus is capable of rapid human-to-human transmission. As of May 06, 2020, over 3.66 million cases of COVID-19 have been reported, with > 2,57,000 deaths worldwide. In India, there are 67,700 active COVID-19 cases with more than 2,215 deaths according to the latest figures updated by the Ministry of Health. Currently, seven CoV species are confirmed that act as human pathogens [3]. Four of them namely HCoV-229E, NL63, OC43, and HKU1 cause non-severe acute respiratory syndrome. Other human CoVs include SARS-CoV, MERS-CoV (Middle East Respiratory Syndrome) which caused severe outbreaks in the years 2002 and 2012 respectively. Similar to SARS-CoV and MERS-CoV, SARS-CoV-2 infection in its early stages were observed in family clusters and among hospital personnel in Wuhan, China [4]. SARS-

CoV, MERS-CoV, and SARS-CoV-2 were believed to have originated in bats, but the intermediate host from bat to human is still unknown [5]. The source of SARS-CoV-2 outbreak is linked to seafood wet market in Wuhan, which served as the epicenter of COVID-19 outbreak prior to January 1st, 2020 [6].

The RNA genome of SARS-CoV-2 encodes six major ORFs [7]. Of these, spike glycoprotein plays the key role in determining viral tropism and mediates binding of the virion to host receptor, ACE2, as a homo-trimer [8]. Spike protein is composed of two domains, S1 and S2. S1 contains a receptor binding domain (RBD). The C-terminal of RBD binds to ACE2 following which, the S2 domain induces fusion of the virion to the host cell membrane, through the exposure of a highly conserved fusion peptide which is activated by proteolytic cleavage in conserved S2 [9,10]. In addition, protease priming, ionic interactions (H⁺ and Ca²⁺) induces conformational changes in the spike protein that eventually dictate the launch of the virion onto ACE2 [11]. Binding of spike protein to ACE2 receptor is a critical step as it is believed to correlate with viral infectivity and governs clinical outcomes [12]. Like SARS-CoV-2, SARS-CoV and HCoV-NL63 use ACE2 human receptor for cellular entry [13].

In general, RNA viruses are characterized by a high mutation rate probably due to limited fidelity of RNA dependent RNA polymerase, that replicates viral genome. It is speculated that genome variability may enable viruses to heighten its virulence, escape host immunity and develop drug resistance. The present report thus explores the evolutionary constraints structuring the genetic landscape of spike glycoprotein of SARS-CoV-2. The possible functional ramification of the observed spike mutants in relation to forming complexes with ACE2 and determining Antigenic properties of the virus are also evaluated. Our study indicates the possibility that relaxed purifying selection carves the genetic architecture of SARS-CoV-2. The study renounces the popular believe of functional association of spike mutations with increasing SARS-CoV-2 virulence and host immune response.

Materials and Methods

To carry out the above, 162 sequences encoding gene of spike glycoprotein of SARS-CoV-2 circulating globally during December, 2019 to February, 2020 were downloaded from NCBI [14]. Spike gene sequences of BatCoV-RaTG13 (MN996532.1), SARS-CoV (FJ588686.1) and HCoV-NL63 (NC_005831.2) glycoprotein were downloaded from NCBI. The sequences were aligned with SARS-CoV-2 isolate Wuhan-Hu-1 (MN908947.3), the first viral genome sequenced completely, to identify sequence variants. Genetic parameters for haplotype and nucleotide diversities, test for neutrality (Tajima's D, Fu and Li's D and F statistics) were estimated using DnaSP v6 [15]. Signature of natural selection was evaluated by estimating non-synonymous and synonymous substitution rates using MEGA vX [16]. McDonald-Kreitman (MK) test was performed to compare inter- and intra-specific nucleotide changes between human SARS-CoV-2 and BatCoV-RaTG13 spike protein sequences [17]. Spike protein sequences of SARS-CoV-2 (6VYB: identical to Hap1 encoded protein) and SARS-CoV (6ACK) used for homology modeling were obtained from the RCSB protein database [18]. Besides, amino acid sequences of spike protein for BatCoV-RaTG13 (QHR63300.2) and HCoV-NL63 (YP_003767.1) were extracted from NCBI. Homology based structural models of viral spike proteins for SARS-CoV-2, SARS-CoV, HCoV-NL63 and BatCoV-RaTG-13 were built using the template crystal structures with PDB codes 6VYB, 6ACK, 6VXX and 5SZS respectively using SWISS-MODEL [19]. The models were subjected to further optimization by addition of hydrogen atoms to preserve global folding and minimize torsional strain using Accelrys-DS viewer 2.0 (accelrys.com). To study protein-protein interaction, crystal structure of ACE2 (PDB code 2AJF) were downloaded from the RCSB. The global binding energy for interaction between the spike protein and human ACE2 receptor was evaluated using Patch Dock [20,21] and FireDock [22,23]. To carry out this, 10 best models of interaction between

ligand and receptor were constructed using PatchDock. Refinement of all docked poses was carried out to derive the best model with most favorable global energy using Fire Dock server. To study the binding of spike variants with ACE2, AutoDock (autodock.scripps.edu/) and Accelrys-DS viewer 2.0 software were employed. The coordinates of the side chains of each mutant residue were incorporated in the homotrimer of Hap1 encoded protein structure using AutoDock and successive structural refinement was carried out for energy minimization, as before. A comparison of binding energies of two alternative forms for all 13 missense and 6 silent mutations was conducted using AutoDock by considering 31 peptides each of length 12 residues including the variant amino acid. The resulting docking data were processed and analyzed using PRODIGY software [24] followed by evaluation of global binding affinity of the variant peptide with ACE2 receptor. Linear B cell epitopes of SARS-CoV-2 spike protein was predicted using BepiPred 2.0 assigning a threshold of 0.55 that corresponded to a specificity cutoff of 80% [25]. The threshold score of 0.55 was chosen to slightly increase the specificity without compromising the sensitivity as compared to a default optimum set up of 0.5. A comparison of residue numbers and distribution for predicted epitopes was made among the observed haplotype (Hap1-13) encoded SARS-CoV-2 spike proteins and also for spikes of SARS-CoV, HCoV-NL63 and BatCoV-RaTG13. To compare the region of epitope preference for a certain CoV species with ancestral SARS-CoV2 spike predicted epitopes as query sequence and to account for the percentage query cover BLAST (<https://blast.ncbi.nlm.nih.gov/>) analysis was employed.

Results and Discussion

An inspection of 162 sequences representing the gene for spike glycoprotein of SARS-CoV-2 revealed the presence of 18 nucleotide variations of which 13 caused amino acid replacements. Eight of these non-synonymous changes mapped to RBD. Eighteen variable sites produced 18 haplotypes that encoded 13 distinct spike mutant proteins. The one encoded by Hap1.1 was considered to be the ancestral type because of its very high abundance. This and Hap1.6 which also represented ancestral type spike protein were originated in China (Table 1). In addition, spike proteins encoded by Hap3 and Hap5 which differed from the ancestral spike protein (alias Hap1) by Y28N and N74K respectively were detected in China. Spike variants encoded by Hap2, Hap4 and Hap6-13 harbored non-synonymous variants namely L5F, H49Y, F157L, G181V, S221W, S247R, R408I, G476S, D614G, F797C, and A930V, which presumably originated outside China (Figure 1A). Eleven spike proteins (encoded by Hap3-13) differed from their ancestral counterpart by a single amino acid change except for Hap2 which had two alterations (C13T/L5F and G1426A/G476S) existing in linkage disequilibrium.

HAP ID	Variant position		Binding affinity (Kcal/mol) Ancestral, derived	Amino acid sequence in haplotype	Frequency	Global energy (Kcal/mol)
	Base	Amino Acid				
Hap 1.1	Ancestral		Reference	LYHNFGSSRGDFA	0.736	-21.06
Hap 1.2	C870T	D290D	-17.2	LYHNFGSSRGDFA	0.006	-21.06
Hap 1.3	G1422A	Q474Q	-12.5	LYHNFGSSRGDFA	0.025	-21.06
Hap 1.4	C1623T	F541F	-16.4	LYHNFGSSRGDFA	0.006	-21.06
Hap 1.5	C2472T	N824N	-15.1	LYHNFGSSRGDFA	0.031	-21.06
Hap 1.6	A2763G	K921K	-14.0	LYHNFGSSRGDFA	0.012	-21.06
Hap 2	C13T	L5F	NA	FYHNFGSSRSDF A	0.012	-12.00
	G1426A	G476S	-12.9, -14.8			
Hap 3	T82A	Y28N	-10.5, -12.5	LNHNFGSSRGDFA	0.006	-15.91
Hap 4	C145T	H49Y	-15.9, -12.8	LYYNFGSSRGDFA	0.006	-15.66
Hap 5	T222A	N74K	-12.0, -14.3	LYH K FGSSRGDFA	0.006	-11.65
Hap 6	C471A	F157L	-10.4, -11.8	LYHNLGSSRGDFA	0.006	-6.92
Hap 7	G542T	G181V	-11.5, -12.1	LYHNFVSSRGDFA	0.006	-5.40
Hap 8	C662G	S221W	-10.5, -10.2	LYHNF G WSRGDFA	0.006	-3.18
Hap 9	T741G	S247R	-10.0, -13.1	LYHNFG S RRGDFA	0.006	4.65
Hap10	G1223T	R408I	-12.6, -15.4	LYHNFG S IGDFA	0.006	-3.18
Hap11	A1841G	D614G	-9.5, -10.8	LYHNFGSS R GGFA	0.105	-15.66
Hap12	T2390G	F797C	-12.9, -14.5	LYHNFGSS R GD C A	0.006	-18.28
Hap13	C2789T	A930V	-12.8, -15.4	LYHNFGSS R GD F V	0.006	-15.48

Table 1: Haplotypes representing SARS-CoV-2 spike proteins and predicted affinities (residue-wise and global) of spike isoforms for human ACE2 receptor.

The amino acids indicated in bold letters denote missense alterations.

To assess the efficiency of virus entry to host cell, global binding energy of each of the 13 spike proteins with ACE2 was enumerated. Our analysis revealed that the ancestral spike protein (alias Hap1) displayed the most favorable interaction with ACE2 as shown by the lowest global binding energy (Figure 1B and table 1). This questions the validity of hypothesis that mutations in the gene encoding spike may improve virus entry inside host cell. This also casts a doubt on the assumption that D614G eases the entry of the virus into host cell and serves as the basis of a selective sweep in the virus sequences. Global binding energy of spike proteins with ACE2 for other human CoVs with varied pathogenicity was estimated (Figure 1C). Estimated energy of interaction of HCoV-NL63-spike to ACE2 was lower (more favorable) compared to corresponding spike proteins of SARS-CoV and SARS-CoV-2. This might explain the undisrupted circulation of HCoV-NL63 in human since its first isolation [26,27]. The difference in RBD sequences of spike protein of HCoV-NL63 from SARS-CoV (82.9%) and SARS-CoV-2 (71%) could explain the difference in predicted binding energies [28]. The altered primary structure of the pro-

tein may lead to a topological difference in spike homotrimer and hence an alteration in the binding complex as indicated by a study showing that the number of residues contacting ACE2 is lower in HCoV-NL63-spike (n = 23) compared to those SARS-CoV-2-spike (n = 30) and SARS-CoV-spike (n = 24) [28]. The higher binding affinity of HCoV-NL63-spike to ACE2 could thus be attributed to the difference in the three-dimensional surface areas lining the interface of spike-ACE2 complexes [28]. This is also supported by our observation that the estimated residue wise binding energy of a mutant amino acid was lower than the corresponding reference amino acid for 10 out of 13 mutation in spike protein. Despite this, the ancestral spike protein which harbored ancestral amino acid at all 13 positions attained the lowest global binding energy when bound to the receptor (Table 1). The relatively stronger binding affinity in SARS-CoV-2 compared to SARS-CoV might be related to the rapid spread of COVID-19 through respiratory droplets [29]. Together, our analysis revealed that the spike protein of the ancestral variety which originated in China could engage into ACE2 with the highest affinity compared to its other low frequency counterparts circulating in the human population, globally.

To elaborate the repertoire of antigenic epitopes presented by spike proteins in CoV mediated infection, linear B cell epitopes were predicted for SARS-CoV-2, SARS-CoV and HCoV-NL63 (Table 2). Majority of the epitope sequence overlap with a query cover of 100% for SARS-CoV-2 observed haplotypes, although a limited number of additional stretch of residues were detected for Hap 2 (405-417: DEVRQIAPGQTGK), Hap 5 (404-419: GDEVRQIAPGQTGKIA), Hap 6 (482-489: GVEGFNCY, 617-622: CTEVPV), Hap 7 (404-419: GDEVRQIAPGQTGKIA), Hap 8 (406-413: EVRQIAPG, 487-492: NCYFPL, 1110-1114: YEPQI), Hap 10 (405-419: DEVIQIAPGQTGKIA) and Hap 12 (482-489: GVEGFNCY, 617-622: CTEVPV). This finding thus failed to provide support to the speculation that presence of mutations in viral proteins was responsible for virus mediated fatality. The epitope sequences predicted from Hap1 of SARS-CoV-2 and SARS-CoV demonstrated 75% similarity in query cover, while that between Hap1 and BatCoV-RaTG13 displayed 99% similarity (data not shown). On the other hand, a very low query cover (25%) was observed for epitopes of SARS-CoV-2 and HCoV-NL63 spike protein epitopes attributed to sequence dissimilarity between these two CoV species. The number and length of predicted epitopes of SARS-CoV-2 (#20.23 ± 0.83; 11.05 ± 0.71) and SARS-CoV (#19; 11.00 ± 5.77) were higher than that of HCoV-NL63 (#18; 9.55 ± 2.93)]. The cumulative lengths of linear B cell

epitopes occupied 17.91%, 16.84% and 12.68% of the corresponding spike protein sequences in SARS-CoV-2, SARS-CoV and HCoV-NL63 respectively. However, one should be cautious to associate density of the epitopes with the virulence of the virus.

Finally we compared the nucleotide diversity and mismatch distribution based on 162 spike gene sequences from SARS-CoV-2. Watterson's θ (0.00083 ± 0.00027) was 5.9 times higher than nucleotide diversity (π) estimate (0.00014 ± 0.00002) due to excess low frequency variants including singletons, which was also reflected in statistically significant negative Tajima's D^{**} (-2.25175) and statistically significant D^* (-4.63509) and F^* (-4.45814). The pairwise mismatch distribution profile of spike gene sequences did not deviate from that expected under neutral evolution (Figure 2). A lower ($dn/ds < 1$) ratio implied a signature of purifying selection which was further supported by MK test. This statistical test is typically applied to detect signature of adaptive evolution by comparing diversity within a species with the divergence between species at non-synonymous and synonymous sites. A significantly low ratio of nonsynonymous to synonymous variation observed between species (SARS-CoV-2 and BatCoV-RaTG13) compared to that observed in SARS-CoV-2 virus population ($34/225 < 13/5$, $p < 0.00001$) rejected any possibility of adaptive selection in the present data set.

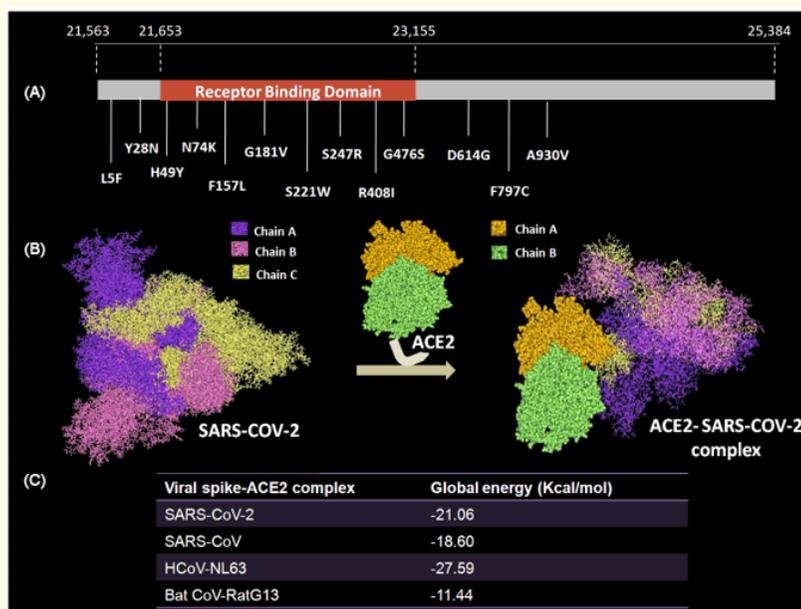


Figure 1: Interaction model between SARS-CoV-2 spike protein and the human ACE2 receptor (A) Genetic map of SARS-CoV-2 spike protein highlighting positions of non-synonymous and synonymous amino acid changes. (B) A molecular docking model of SARS-CoV-2 spike protein and human ACE2 receptor. (C) Comparison of spike-ACE2 interactions of two human CoVs and a Bat-CoV based on global energy estimates.

Epitope sequence ‡	SARS-CoV-2													SARS-CoV ACU31032.1		H-CoV-NL63 YP_003767.1		
	Hap1	Hap 2	Hap 3	Hap 4	Hap 5	Hap 6	Hap 7	Hap 8	Hap 9	Hap 10	Hap 11	Hap 12	Hap 13	Sequence	Start-End	Sequence	Start-End	
	Map positions: Start-End																	
CVNLTTR TQLPPAY TN	15-30	15-30	16-28*	15-30	15-30	15-30	15-30	15-31*	15-30	15-30	15-30	15-30	15-30	TKPQPLMDK FSSRR	24-38	NA		
NVTWFH AIHVS NGT	61-76	61-76	62-75*	61-76	70-74*	67-74*	61-76	67-76*	61-76	61-76	61-76	61-76	61-76	FLPFD	59-64	NQSTSVY	52-58	
NA													NA		TTDFLSNASS	112-123		
VYYHKNN KSWME	143-154	142-161*	143-162*	143-154	143-162*	141-154*	142-154*	142-156*	143-154	142-154*	143-154	142-155*	143-154	SNEQHY	147-152	NA		
KQGNFK	182-187	183-187*	183-187*	182-187	183-187*	181-187*	182-187	183-187*	182-187	182-187	182-187	182-187	182-187	NA				
INLVRDLP QGFS	210-221	210-220*	210-221	210-221	210-221	210-218*	210-221	210-218*	210-221	210-221	210-221	210-221	210-221	DIPRGLPV	208-215			
LTPGDSS GW	249-258	249-259*	249-258	249-258	249-258	250-258	249-259*	249-258	249-259*	249-259*	249-259*	249-259*	249-259*	NA			SDVNCNGYQH N	280-290
NA													TNRC	324-328	LNFSAN		300-305	
FNATRFAS	342-349	341-352*	342-349	342-349	342-349	342-349	344-349*	341-352*	342-350*	344-349*	342-349	342-350*	342-349	NA				
SFSTFKCY GVSP NDL	373-390	371-378*	373-390	373-379*	NA	374-390*	NA	373-390	373-390	NA	373-390	374-390*	373-390	TFKCYGVSPS	367-376	NA		
#	NA	405-417*	NA	404-419*	NA	404-419*	406-413*	NA	405-419*	NA	NA	NA	NA	GVIAD	407-411			
NLDSKVG G	440-447	440-448*	440-447	440-447	440-448*	440-448	440-448*	439-448*	440-447	440-448*	440-447	440-447	440-447	LKPFERDLTS DENGVR	447-462			
KSNLKP ERDISTEI	458-475	460-475*	458-475	458-475	454-474*	461-465*	454-474*	458-475	458-475	454-474*	458-475	461-465*	458-475			ANFLDDN	461-467	
YQA						471-478*						470-478						
#	NA					482-489*	NA	487-492*	NA			482-489*	NA	NA		NA		
NA													NAPATVCGPK LST	491-503	CKPHQVNISL N	500-510		
ATVCGPK KSTNL	522-533	519-532*	522-533	522-533	522-533	523-533*	522-533	525-533*	522-533	522-533	522-533	523-534*	522-533	KRFQS	529-533	KSGSPGDS	532-539	
NA													NASSEVAVLY QDV	575-587	NA			
NA													QLTPAWRVY STGVNVF	600-615	NSITGVYPVPS GIREFS	604-620		
#	NA					617-622*	NA					617-622*	NA	NA		NA		
PTWRVYS TGSNVFQ	631-644	631-644	631-644	631-644	631-644	634-644*	631-644	630-644*	631-644	631-644	634-642*	634-644*	631-644	HTASTLRSVG QKSIVAYTMS LGAENSIA	647-674	NA		
NSYEC	658-662	NA	658-662	658-662	658-662	658-662	658-662	NA	658-662	658-662	658-662	NA	658-662					
YQTQNS PRRARSV AS	674-689	673-689*	674-689	673-689*	674-689	674-689	673-689*	673-689*	674-689	673-689*	674-689	674-689	674-689					
TMSLGAE NSVAYS	696-708	696-708	696-708	697-708*	697-708*	696-709*	696-709*	696-708	696-708	697-708*	696-708	696-709*	696-708	NA		RYGLQNLL	702-709	
NA													YKTPAIKDFG GFNF	757-770	NA			
NA													QILPDPKPT	772-781				
KTPPIKDF G	790-798	790-798	790-798	790-798	790-798	790-798	790-798	790-798	789-798*	790-798	790-798	790-797*	790-798	NA		SNAPSLANVTS FGD	837-850	
DPSKPS	808-813	808-813	808-813	808-813	808-813	808-813	808-813	808-814*	808-813	808-813	808-813	808-813	808-813	NA		PQRNIRSSRIA G	858-869	
NA													NA		LGTVDV	886-891		
#	NA							1110-1114	NA					TVYDPLQP	1104-1111	KSQSNRYGF	1100-1108	

Citation: Pramita Chowdhury, Bijurica Chakraborty, et al. "In-depth Sequence Analysis of SARS-CoV-2 Spike Protein Repudiates Mutation Mediated Adaptive Selection in the Virus". *Acta Scientific Microbiology* 3.12 (2020): 16-23.

								*										
TVYDPLQ P	1136 - 1143	1136 - 1143	1136 - 1143	1136 - 1143	1136- 1143	1136- 1143	1136 - 1143	1136 - 1143	1136- 1143	1136 - 1143	1136 - 1143	1136 - 1143	1136 - 1143	KYFKNHTSPD VDLGD	1122- 1137			
KYFKNHT SPVDLGD DI	1154 - 1169	1154 - 1169	1154 - 1169	1154 - 1169	1154- 1169	1154- 1169	1154 - 1169	1154 - 1169	1154- 1169	1154 - 1169	1154 - 1169	1154 - 1169	1154 - 1169	NA	NA			
NA															MFQPRLPVL	1178- 1186		
NA															VELHTVIP	1205- 1212		
NA															KYVKPNF	1230- 1236		
FDEDD	1256 - 1260	1256 - 1260	1256 - 1260	NA				1256- 1260	NA				NA					
NA																	STKLPYYEF	1342- 1350

Table 2: Predicted linear B cell epitopes of human corona virus (CoV) species spike protein. Epitope probability cutoff: ≥ 0.55, Epitope length: ≥ 5 residues.

‡ denotes epitope sequences of ancestral (Hap1) SARS- CoV-2 spike protein.

* denotes any alteration of residue positions of predicted epitopes for SARS-CoV-2 Hap2-13 with ancestral type Hap 1.

and blue color were used to mark additional epitope sequences and their map positions specified by SARS-CoV-2 Hap2-13.

Red colored stretches denote epitopes that are completely identical between SARS-CoV/NL-63 and SARS-CoV-2.

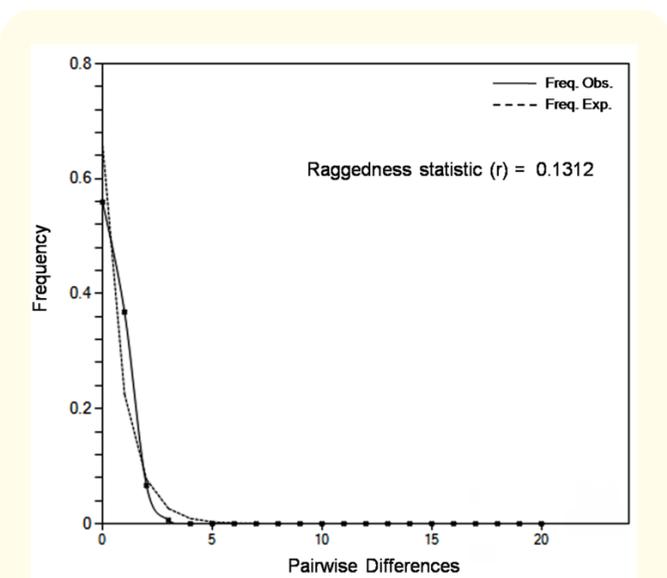


Figure 2: Mismatch distribution pattern of SARS-CoV-2 surface glycoprotein. The pattern of pairwise nucleotide differences in the SARS-CoV-2 spike protein. The black solid line indicates the observed distribution and the dotted black colored line indicates mismatch distribution pattern under neutral expectation.

Conclusion

Taken together, it may be surmised that relaxed purifying selection due to adroit maneuver of immune system and rapid transmissibility of SARS-CoV-2 due to higher R₀ factor for human-to-human

transmission shape the genetic structure of spike protein. Rather than genetic alterations mediated efficient host entry, the high viral titer in the upper respiratory tract of an infected person and shedding of the virus particles from asymptomatic and pre-symptomatic individuals through respiratory and fomite droplets are the most likely reasons behind virus transmissibility [30,31]. The present study finds the ancestral haplotype of spike glycoprotein to be the most agile form. However further enhancement of sample size and whole genome analysis is critical towards hypothesizing a more sophisticated infection strategy of SARS-CoV-2 and embracing a more refined answer for COVID-19 outburst and its remedy.

Acknowledgment

PC and BC are supported by CSIR-Direct SRF fellowships from CSIR and Swami Vivekananda Merit-cum-Means Scholarship (SVM-CM) from Government of West Bengal respectively. The authors deeply thank Dr. Srikanta Sen for his valuable suggestions that improve protein homology modeling.

Author Contributions

SSG conceived the study. SSG, PC and BC designed the study. PC and BC performed the analysis. SSG, PC and BC analyzed the data and wrote the manuscript. All authors have equally contributed and reviewed the manuscript.

Conflict of Interests

None.

Bibliography

1. Coronavirus disease (COVID-19) pandemic.
2. Menachery VD, et al. "Jumping species-a mechanism for coronavirus persistence and survival". *Current Opinion on Virology* 23 (2017): 1-7.
3. Weiss SR and Navas-martin S. "Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus". *Microbiology and Molecular Biology Reviews* 69.4 (2005): 635-664.
4. Gralinski LE and Menachery VD. "Return of the Coronavirus: 2019-nCoV". *Viruses* 12.2 (2020).
5. Cui J, et al. "Origin and evolution of pathogenic coronaviruses". *Nature Review Microbiology* 17.3 (2019): 181-192.
6. Li Q, et al. "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia". *The New England Journal of Medicine* 382.13 (2020): 1199-1207.
7. Zhou P, et al. "A pneumonia outbreak associated with a new coronavirus of probable bat origin". *Nature* 579.7798 (2020): 270-273.
8. Belouzard S, et al. "Mechanisms of coronavirus cell entry mediated by the viral spike protein". *Viruses* 4.6 (2012): 1011-1033.
9. Hulswit RJ, et al. "Coronavirus Spike Protein and Tropism Changes". *Advances in Virus Research* 96 (2016): 29-57.
10. Madu IG, et al. "Characterization of a highly conserved domain within the severe acute respiratory syndrome coronavirus spike protein S2 domain with characteristics of a viral fusion peptide". *Journal of Virology* 83.15 (2009): 7411-7421.
11. Heald-sargent T and Gallagher T. "Ready, set, fuse! The coronavirus spike protein and acquisition of fusion competence". *Viruses* 4.4 (2012): 557-580.
12. Mossel EC, et al. "Exogenous ACE2 expression allows refractory cell lines to support severe acute respiratory syndrome coronavirus replication". *Journal of Virology* 79.6 (2005): 3846-3850.
13. Hofmann H, et al. "Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry". *Proceedings of the National Academy of Sciences of the United States of America* 102.22 (2005): 7988-7993.
14. Basic local alignment search tool.
15. Rozas J, et al. "DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets". *Molecular Biology and Evolution* 34.12 (2017): 3299-3302.
16. Kumar S, et al. "MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms". *Molecular Biology and Evolution* 35.6 (2018): 1547-1549.
17. Mcdonald JH and Kreitman M. "Adaptive protein evolution at the Adh locus in Drosophila". *Nature* 351.6328 (1991): 652-654.
18. RCSB protein database.
19. SWISS-MODEL.
20. Duhovny D, et al. "Efficient Unbound Docking of Rigid Molecules". In: Guigó R, Gusfield D. (eds) Algorithms in Bioinformatics. WABI 2002. Lecture Notes in Computer Science, vol 2452. Springer, Berlin, Heidelberg (2002).
21. Schneidman-duhovny D, et al. "PatchDock and SymmDock: servers for rigid and symmetric docking". *Nucleic Acids Research* 33 (2015).
22. Andrusier N, et al. "FireDock: fast interaction refinement in molecular docking". *Proteins* 69.1 (2007): 139-159.
23. Mashiach E, et al. "FireDock: a web server for fast interaction refinement in molecular docking". *Nucleic Acids Research* 36 (2008): W229-232.
24. Xue LC, et al. "PRODIGY: a web server for predicting the binding affinity of protein-protein complexes". *Bioinformatics* 32.23 (2016): 3676-3678.
25. Jespersen MC, et al. "BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes". *Nucleic Acids Research* 45 (2017): W24-W29.
26. Principi N, et al. "Effects of coronavirus infections in children". *Emerging Infection Disease* 16 (2010): 183-188.
27. Raoult D, et al. "Coronavirus infections: Epidemiological, clinical and immunological features and hypotheses". *Cell Stress* 4.4 (2020): 66-75.
28. Brielle ES, et al. "The SARS-CoV-2 Exerts a Distinctive Strategy for Interacting with the ACE2 Human Receptor". *Viruses* 12.5 (2010).

29. Ortega JT, *et al.* "Role of changes in SARS-CoV-2 spike protein in the interaction with the human ACE2 receptor: An analysis". *EXCLI Journal* 19 (2020): 410-417.
30. Bai Y, *et al.* "Presumed Asymptomatic Carrier Transmission of COVID-19". *JAMA* 323.14 (2020): 1406-1407.
31. Zou L., *et al.* "SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients". *The New England Journal of Medicine* 382 (2020): 1177-79.

Assets from publication with us

- Prompt Acknowledgement after receiving the article
- Thorough Double blinded peer review
- Rapid Publication
- Issue of Publication Certificate
- High visibility of your Published work

Website: <https://www.actascientific.com/>

Submit Article: <https://www.actascientific.com/submission.php>

Email us: editor@actascientific.com

Contact us: +91 9182824667