Molecular modelling predicts SARS-CoV-2 ORF8 protein and human complement Factor 1 catalytic domain sharing common binding site on complement C3b

Jasdeep Singh[1], Sudeshna Kar[2], Seyed Ehtesham Hasnain[1] and Surajit Ganguly[3 #]

[1] Jamia Hamdard Institute of Molecular Medicine, Jamia Hamdard, Hamdard Nagar, New Delhi, 110062, India.

[2] Host-Microbe Interaction and Molecular Carcinogenesis Laboratory, Jamia Hamdard Institute of Molecular Medicine, Jamia Hamdard, Hamdard Nagar, New Delhi, 110062, India.

[3] Neurobiology and Drug Discovery (NDD) Laboratory, Jamia Hamdard Institute of Molecular Medicine (JH-IMM), Jamia Hamdard, New Delhi 110062

[#] Corresponding Author:

Dr. Surajit Ganguly,

Jamia Hamdard Institute of Molecular Medicine,

School of Interdisciplinary Studies and Technologies,

Jamia Hamdard (Hamdard University),

Hamdard Nagar, New Delhi, 110062, India.

email: surajit.ganguly@jamiahamdard.ac.in

Phone: +919999797944

Abstract

Function of Open Reading Frame 8 (ORF8) protein of SARS-CoV-2 is still unclear. Here, we predict the functional role of ORF8 of SARS-CoV-2 in the context of host-pathogen relationship and the impact of mutations acquired during transmission. Mutational entropy analysis of 1042 ORF8 sequences of SARS-CoV-2 reveals a remarkable conservation among all isolates with high propensity of mutations only at amino-acid positions S24, V62, and L84. Search for structural homolog of ORF8 protein identified human complement factor 1 (F1; PDB ID: 2XRC) with 48% similarity to the C-terminus serine-protease domain. Comparative protein-protein interaction modelling predicts ORF8 binding with human complement C3b, an endogenous substrate of F1. ORF8 appears to bind via overlapping F1-interacting region on C3b (Chain B) with higher binding energy than F1-C3b complex. However, introduction of natural mutations on ORF8 reduced the binding energy. Thus, ORF8 can potentially disrupt complement activation by competing with F1 for C3b binding.

Key Words: Covid-19, SARS-CoV-2, coronavirus, ORF8, Complement C3b, innate immunity, Factor F1.

1. Introduction

The world is currently facing an unprecedented pandemic caused by the transmission of a novel coronavirus, named as severe acute respiratory syndrome coronavirus 2 or SARS-CoV-2 [1]. The epicentre of the outbreak was first detected in Wuhan, China and Lu et al first described the manifestation of the disease as "pneumonia of unknown etiology" in December 2019 [2]. Latest WHO data suggest that the confirmed cases are mounting dramatically, with more than 3.5 million confirmed cases and about 250,000 deaths reported worldwide as of first week of May, 2020. Following the identification of the virus, the early phase of clinical characterization of the disease reveal that a substantial number of the patients progresses towards acute respiratory distress syndrome (ARDS) eventually leading to multi-organ failure [3, 4]. In order to fight this outbreak, WHO has announced a R&D blue print to accelerate research addressing critical questions that might lead to an intervention of this devastating outbreak and prevent future such pandemics. Towards this end, various countries, including China, India, USA, Italy, and others have managed to isolate and sequence the virus genome and have been engaged in unravelling information about the virus at a rapid pace.

The SARS-CoV-2 is a positive-stranded RNA virus with a genome size of about 29 Kb. NCBI database has now more than 1000 SARS-CoV-2 genome sequences. The genome of SARS-CoV-2, similar to that of other CoVs, is organized from 5' end as open reading frame comprising of 1ab (ORF1ab), spike (S), ORF3, envelope (E), membrane (M), ORF6 to ORF8, and nucleocapsid (N) [5]. It is believed that the origin of human SARS-CoV non-structural ORF8 (Open reading frame 8) protein is from Rhinolophus species SARSr-Rs-BatCoVs reservoir [6 - 9]. However, during human to human transmission of SARS-CoV at the time of the 2003 outbreak, a 29 nucleotide deletion was detected in ORF8, leading to generation of two truncated polyeptides, ORF8a and ORF8b [10]. Though the full length ORF8 has been

suggested to activate protein folding machinery in the host cell, their role in host-pathogen relationship is largely unknown [11].

In our search for functional role of ORF8, we observed that it has partial sequence conservation with HIV-1 (Human Immunodeficiency Virus) gp-120 protein and C-terminal domain human complement Factor 1 (F1; PDB ID: 2XRC). Interestingly, both gp120 and F1 interacts with human C3b complement to regulate host immune responses [12, 13]. C3b is central player in humoral immunity of hosts and precursor of potent components involved in opsonisation, pathogen tagging and phagocytic apoptosis [14]. In case of coronavirus infections, Gralinsk et al [15] have shown direct role of C3 complement systems in SARS-CoV mediated respiratory dysfunctions, validating that its associated pathologies are predominantly immune driven . Recent clinical studies on SARS- and MERS-CoV (Middle East respiratory syndrome) infections also pointed towards innate immune evasion mechanisms adopted by CoVs mediating delay in innate immune responses [16]. However, the underlying molecular mechanisms are still unclear.

Here, using protein-protein interaction modelling, we predict ORF8 binding with human complement C3b, the endogenous substrate for F1. Interestingly, the binding interface of ORF8 on C3b appears to overlap with F1-interacting region. Moreover, the ORF8 binds with C3b, releasing higher global energy than F1-C3b complex. However, incorporation of common naturally-occurring mutations on ORF8 (mORF8) seems to perturb the interactions with C3b. Thus, our results suggest that the 121-amino acid long wild-type ORF8 of SARS-CoV-2 can potentially compete with F1 for binding to C3b, with a possibility of hindering complement activation in the host.

2. Materials and Methods

2.1. Viral isolates and Sequence data: ORF8 protein sequences from NCBI databases were used for the analysis to evaluate the amino acid (AA) conservation. The analysis includes 1042 ORF8 sequences of SARS-COV-2 available in NCBI from isolates across the hotspots of the infection worldwide. The sequence ID of each are provided along with multiple sequence alignment image in Supplementary Figure S1. A consensus ORF8 protein sequence, deduced from SARS-CoV-2 genome sequences, was also used to perform multiple sequence alignment with the following sequences: ORF8a [AAP33705.1 of SARS coronavirus Frankfurt 1, AAQ94068.1 SARS coronavirus AS and AAP82972.1 of SARS coronavirus Shanhgai LY] and 8b protein of SARS-COV and the ORF8 protein from AAP33706.1 of SARS coronavirus Frankfurt 1, AAQ94069.1 of SARS coronavirus AS, AAP82973.1 of SARS coronavirus Shanhgai LY. The hypothetical protein from SARS coronavirus Rs_672/2006 of Rhinolophus species (ACU31043.1) was used as a reference to identify the divergent amino acids.

2.2. Alignment of ORF8 protein sequences: The retrieved ORF8 protein sequences from protein database were aligned using MUSCLE algorithm in SeaView [17]. Shanon entropy was calculated using HIV sequence database entropy tool (https://www.hiv.lanl.gov/content/sequence/ENTROPY/).

2.3 Identification of structural homologs of ORF8

To find sequence based nearest structural homologs of ORF8, we subjected 121 AA consensus ORF8 sequence to blastp (protein-protein) and PSI blast (position specific iterated) queries against PDB database. Alignments were generated using expect threshold of 10 and BLOSUM62 scoring matrix with conditional compositional score matrix adjustment. PSI blast was run at 0.005 threshold.

2.4. Methods for homology modelling and protein-protein docking: Modelling of ORF8 and mutant ORF8 (mORF8; triple mutations S24L, V62L and L84S) was performed using I-TASSER (Iterative Threading ASSEmbly Refinement) with human complement factor 1 (F1; PDB id: 2xrc) as template. The stereo chemical quality of resulting model was verified using Ramachandran and ProSA analysis. Without assigning any prior binding site (unbiased), protein-protein dockings were performed employing Patchdock using clustering RMSD of 4.0 and subsequently top 20 resulting solutions were refined using Firedock (18 - 21). The global energy of top 20 ranked solutions by FireDock were calculated from contribution of attractive and repulsive vander waals forces, atomic contact energy (ACE) and contribution of the hydrogen bonds. Interaction analysis was carried out using PDBPisa (https://www.ebi.ac.uk/pdbe/pisa/).

2.4. Molecular dynamics simulations and setup. MD simulations were carried using gromacs v2016. Simulations of F1-C3b and ORF8-C3b complexes were carried at 300 K in a cubical box with 10 nm spacing from box edges. Ionic concentrations was kept at 0.1 M NaCl followed by system minimization by Steepest descent protocol. Equilibration and production runs (25 ns) were carried using parameters detailed in our previous reports [22, 23]. Images were constructed using PyMol while data was analysed using standard gromacs tools.

3. Results

3.1. Computational analysis of SARS-CoV-2 ORF8 Sequence

A single polypeptide ORF8 of SARS-like coronaviruses was found in bats (Rhinolophus species) as host Reservoir. It remained as a single protein on transfer to humans and in the early phase of the SARS-COV epidemic in 2003. In the middle phase of the transmission, however, ORF8 was mutated and generated two protein variants (ORF8a and ORF8b) as a result of a deletion of 29 nucleotides [10]. This variant of ORF8 dominated in the late phase as well,

without regaining the lost 29 nucleotides. This led us to search for similar variants in the currently prevailing SARS-COV-2 strains throughout the world.

Our amino acid conservation analysis and multiple sequencing alignment (Figure 1A and Supplementary Figure S1) using 1042 ORF8 protein sequences of SARS-CoV-2 showed high sequence conservation among clinical isolates. Shannon entropy, which provides a measures of amino acid (AA) divergence, revealed that the highest frequency of single AA variation among the ORF8 of SARS-CoV-2 sequences exists only in 3 AAs at positions 24, (S24L), 62 (V62L) and 84 (L84S) (Figure 1A). This observation, also confirmed by multiple sequence alignment (Supplementary Figure. S1), suggested that the divergence in the ORF8 of SARS-COV-2 is limited to three residues, even in the current stage of human transmission. This was further validated by analysis of ORF8 from 4431 isolates retrieved from an open-source platform (nextrain.org) demonstrating minimal variation entropy in ORF8 compared to other ORFs, highlighting its conservation during SARS-CoV-2 transmission from Dec 2019 to April 2020 (Supplementary Figure S2). The highest frequency S84L mutation appeared in early January 2020 and geographically dominant in Asian and North American sub-continents followed by South America and Europe (Supplementary Figure S3).

Next, we compared consensus SARS-CoV-2 ORF8 protein with SARS-CoV ORF8 mutant sequences and the ORF8 hypothetical protein (GenBank ID # ACU31043.1) of the SARS virus from the horseshoe bat reservoir. The sequence alignment showed a significant conservation of the AAs throughout the protein between the SARS-CoV-2 and the horseshoe bat derived ORF8 (Figure 1B). In addition, it was noted that the ORF8a (39 AA) of SARS-CoV aligned with the N-terminal region of the ORF8 from SARS-CoV-2 and ORF8b (84 AA) aligned with the C-terminus region (Figure 1B). It is also interesting to note that the residues S24, V62 and L84, which appears to be susceptible to mutations (Figure 1A), are unique for SARS-CoV-2 (Figure 1B). Previously, it was reported that the fragmentation of the ORF8 due

to the deletion of 29 nucleotides has reduced the replication rate of SARS-CoV in the human host as compared to the wild type ORF8 [10]. Hence, it can be inferred that the preservation of the intact ORF8 protein in the SARS-CoV-2 might have contributed to the robustness of its replication.

3.2 Homology search with SARS-CoV-2 ORF8

Using the 121 AA long deduced consensus sequence (wild-type; the sequence with most prevalent amino acids as S at 24, V at 62 and L at 84 across 1024 genome sequences) from the SARS-CoV-2 ORF8 alignment, we searched for homologous proteins in human protein database. BLASTp analysis (with expect threshold of 10 as described in the Methods section) of the consensus ORF8 protein sequence showed significant homology with the F1 protein (human complement factor 1, PDB id: 2XRC) (Supplementary Data S4). The consensus SARS-CoV-2 ORF8 protein sequence showed 48% similarity, with 25% identity with C-terminal domain (500 to 557) of F1 (Figure 2A). According to the NCBI Conserved Domain database, F1 protein is a trypsin-like serine protease with the domains as described in Figure 2B. The catalytic domain extends from 322 to 554 AA residues after the zymogen cleavage site between R (321) and I (322) residues. This catalytic domain is subdivided into an active site, composed of consensus triad residues - H (363), D (411), and S (507), and a substrate binding site, comprising residues D (501), S (527) and G (529). The endogenous substrate of F1 is the complement C3b which has been known to play a central role in complement activation cascade [24]. The C-terminal domain of F1 showed structural RMSD (root-mean-square deviation) of 7.3 Å with the modelled ORF8 protein. In addition, the ORF8 protein lacks the consensus serine protease catalytic triad as in the F1 catalytic domain (Figure 2B). Taken together, it appears that ORF8 is restricted in its function as only a binding ligand, devoid of any catalytic role. Interestingly, PSI (position specific iteration) BLAST of ORF8 highlighted gp120 (glycoprotein120) of HIV1 as top hit (E value~0.4) with 50% identity in the 18% of

total sequence (Supplementary Data S5). The partially conserved region of HIV gp120 has already been shown to interact with C3b [12, 13]. However at this point of time, the correlation between complement modulation and similarity of ORF8-gp120 could not be established.

3.3. Interaction modelling of ORF8 with Complement C3B

In order to check the ability of ORF8 of SARS-CoV-2 to compete with F1 protein for its binding with C3b (PDB Entry - 2I07), we applied an unbiased protein-protein docking paradigm where no pre-defined interaction sites were assigned prior to the docking.. Interestingly, the global energy for top ranked solution of C3b-ORF8 docking (-23.16) was slightly higher than C3b-F1 docking (-20.85). The top ranked solutions from both dockings showed partial overlap of C3b binding site for ORF8 and F1 (Table 1 and 2, Figure 3A and B). ORF8 was found to share the same binding region on chain B (between residues E769 and Y1483) of C3b where F1 binds with highest affinity. The interaction network of F1-C3b and ORF8-C3b are outlined in Table 1 and 2.

In order to investigate the impact of mutations in ORF8, as identified in Figure 1, on binding with C3b, the mutations S24L, V62L and L84S were incorporated in the ORF8 sequence and the interactions between the mutant ORF8 (mORF8) with C3b were studied. Surprisingly, in the top ranked solution for mORF8-C3b docking, mORF8 was bound to a site different from F1 and wild-type ORF8 binding sites (Global energy -15.82) (Figure 4A and B). The docking solution at the proximity of F1 and ORF8 binding site yielded a much lesser global energy score (-2.9) (Figure 4A and C). Although both ORF8 and mORF8 were structurally conserved (RMSD~0.2 Å; Figure 4A), the mutations have clearly perturbed mORF8 binding with C3b. It appears that the preferential interactions of mORF8 with C3b are displaced from F1-C3b binding interface, indicating that mORF8 is less likely to compete for the same binding site with F1 on C3b.

To understand the dynamics of F1/ORF8 interactions with C3b, we performed MD simulations of both complexes for 25 ns. Variation in H-bond interaction network for both systems showed higher number of H-bonds/polar contacts between ORF8-C3b (21.09 H-bonds per time frame) compared to F1-C3b (18.45 H-bonds per time frame) (Figure 5). This further reinforced our docking outcomes of comparatively higher affinity binding of ORF8 to the same site where F1 binds. Thus, it appears that ORF8 possesses the capability of competing with F1 for a common binding pocket on C3b. Therefore, binding of ORF8 has the possibility of locking the C3b-cofactor complex into a conformation that might hinder the access to F1 binding for subsequent proteolytic cleavage between residues R (1303) and S (1321) (Figure 3B) or other sequential cleavage sites on C3b CUB (C1r/C1s, Uegf, Bmp1) domains as described previously [24]. However, in spite of having higher affinity for a shared docking region on C3b, the concentration of ORF8 in the host is more likely to dictate the physiological impact of the ORF8-C3b complex.

4. Discussion

SARS-CoV-2 is phylogenetically more related to SARS-CoV than MERS-CoV [25]. However, the novel coronavirus shares pathological manifestations with other CoVs in inducing cytokine storms and vigorous pro-inflammatory responses. The current study highlights immunomodulatory potential of ORF8 protein of SARS-CoV-2. Using a combination of MD simulations and an unbiased (with no prior input for binding sites) protein-protein docking approach, we have compared interactions between ORF8 and F1 protein for binding with C3b, highlighting a possible mechanism of complement evasion adopted by SARS-CoV-2. We show that the acquired mutations in ORF8, as observed in figure 1A, can negatively modulate its interactions with C3b, leading to loss of binding at F1 binding interface. Apparent lower C3b affinity due to mutations in ORF8, as in mORF8 (Figure 4), might impact host immune responses and influence the outcome of infection. Moreover, the binding results

with the mutant mORF8, reinforces our findings that the wild-type ORF8 and F1 shares energetically favourable overlapping binding region on C3b. Thus, in spite of the absence of experimental correlations, the current work presents interesting insights into functional role of ORF8 at host-viral interface and warrants further investigation.

The active form of human F1 has been shown to be involved in regulation of complement system by binding to the C3b/cofactor and C4b/cofactor complexes [24, 26, 27]. The F1 is first activated by the cleavage of its inactive zymogen form, leading to binding with its substrates, like C3b-cofactor and C4b-cofactor complexes. F1 binding induces favourable structural orientation of the serine protease active site, triggering sequential proteolysis of C3b and C4b producing opsonins. This F1 mediated degradation of C3b and C4b into fragments triggers amplification of immune response cascade, including the adaptive immune system [26, 27]. In this context, the homology of the SARS-COV-2 ORF8 protein with the human F1 substrate binding domain is intriguing. It is possible that ORF8 protein can directly interact with active complement factors like C3b and C4b, thereby rapidly blocking generation of membrane attack complexes required for the viral lysis and at the same time, limiting the generation of C3b and C4b degradation products and suppressing the induction of adaptive immune response. The blocking of complement activation at the C3b stage has been demonstrated for Herpes Simplex Virus (HSV) as well [28]. It appears that the HSV-1 surface protein gC plays a role in attenuating C3b-mediated activation of the complement system. This complement evasion strategy is known to be adopted by Vaccinia Virus also, secreting a complement control protein (VCP) that mediates complement inactivation [29]. Our results supporting ORF8 binding to C3b and shielding the F1 cleavage sites in the CUB domain of C3b with a possibility of limiting the generation of C3b proteolytic products, predicts a similar mechanism for SARS-CoV-2.

Previously, SARS-CoV has been shown to bind to a mannose binding lectin (MBL), which is known to be a component of innate immunity [30]. MBL is believed to participate in attenuating the virion fusion in the host cells. However, MBL opsonisation is not sufficient for viral neutralisation [28, 31]. It appears that additional molecular reinforcements in the form of C3b proteolytic products, like ic3b, c3dg and c3d, are also necessary to be deposited on the virion particles. Thus, our results bring forth the possibility of ORF8 mediated protection of the serine –protease R-S (arginine-serine) cleavage sites on C3b and provide a critical mechanistic lead for complement evasion. Thus, complement inactivation might represent an important survival strategy for the virus preventing host detection and destruction by eliciting robust IgG memory responses. In this context, it is to be noted that the mutations in ORF8 sequence, S24L, V62L and L84S acquired during the course of transmission, appear to disrupt the interactions in the F1 binding region on C3b (Fig 4) with a possibility of impacting the outcome of the infection in the host. However, the significance of these mutations on the host-virus relationship requires experimental validation.

5. Conclusion:

Recognition and elimination of viral particles by the host complement system has been known since 1930 [32]. During the adoption of humans as host, viruses have also evolved unique strategies to subvert the complement mediated innate immunity. Most human viruses, including Astroviruses, Togaviruses, Orthomyxoviruses and Paramyxoviruses, have been demonstrated to block complement activation [33]. It is advantageous for the viruses to evade complement activation. So, it is not surprising that SARS-CoV-2 might potentially evolve a strategy to suppress the host innate defence mechanism. However, the prediction of the ORF8 as a binding partner of the complement-cofactor complexes needs to be demonstrated at the cellular level. This work may serve as an important lead for elucidation of the function of ORF8 in SARS-CoV-2 pathogenesis and for a possible antiviral target in near future. However, as we

prepare this manuscript, a pre-print [34] has emerged that claims to have identified a strain in 8 patients in Singapore with 382-nt deletion covering almost the entire ORF8 of SARS-CoV-2. The information about the extent of prevalence of this mutated strain is still not available. Due to an immense host adoption pressure, it is possible though that the full length or a partial stretch of nucleotides of ORF8 may be shed by the virus during the course of rapid transmission, giving rise to a mutated version of SARS-CoV-2 as observed during the previous SARS-CoV outbreak.

Financial Support and sponsorship: None

Conflict of interest: None

Acknowledgement: The authors acknowledge the sacrifices of all those who are fighting the Covid-19 pandemic. JS acknowledges financial support under Young Scientist scheme of Dept. of Health Research (DHR), Ministry of Health and Family Welfare, Government of India. SEH is a JC Bose National Fellow of the Department of Science & Technology (DST), Government of India. SEH is a Robert Koch Fellow of the Robert Koch Institute. Berlin, Germany. SK acknowledges SERB/DST for core research support. SG thanks research support from DHR, ICMR and SERB/DST.

References

1. A. E. Gorbalenya, S. C Baker, R. S. Baric, et al., The species Severe acute respiratory syndrome-related coronavirus: Classifying 2019-nCoV and naming it SARS-CoV-2, Nature Microbiology 5 (2020): 536-544.

2. H. Lu, C. W Stratton, Y. W Tang, Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle, J. Med. Virol. 92 (2020): 401-402.

3. N. Chen, M. Zhou, X. Dong, et al., Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study, Lancet 395 (2020): 507-513.

4. R. Lu, X. Zhao, J. Li, et al., Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding, Lancet 395 (2020): 565-74.

5. Y. Guan, B. J. Zheng, Y. Q. He, , et al., Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China, Science 302 (2003):276–278. http://dx.doi.org/10.1126/science.1087139.

6. S. K. P. Lau, Y. Feng, H. Chen, et al., Severe Acute Respiratory Syndrome (SARS) Coronavirus ORF8 Protein Is Acquired from SARS-Related Coronavirus from Greater Horseshoe Bats through Recombination, J. Virol. 89 (2015): 10532–10547. Published online 2015 Aug 12. doi: 10.1128/JVI.01048-15

7. J. Yuan, C. C. Hon,., Y. Li, et al., Intraspecies diversity of SARS-like coronaviruses in Rhinolophus sinicus and its implications for the origin of SARS coronaviruses in humans, J. Gen. Virol. 91 (2010): 1058-1062.

8. H. Singh, J. Singh, M. Khubaib, et al, Mapping the genomic landscape and diversity of COVID-19 based on >3950 clinical isolates of SARS-CoV-2: Likely origin and transmission dynamics of isolates sequenced in India, Indian Journal of Medical Research, 2020 (In Press)

9. J. A. Sheikh, J. Singh, H. Singh, et al., Emerging genetic diversity among clinical isolates of SARS-CoV-2: Lessons for today, Infection, Genetics and Evolution, (Available online 24 April 2020, 104330, In Press)

10. D. Muth, V. M. Corman, Roth, H. et al., Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission, Sci Rep 8 (2018)**:** 15177. https://doi.org/10.1038/s41598-018-33487-8

11. D. X. Liu, T. S. Fung, K. K. Chong, A. Shukla, R. Hilgenfeld, Accessory proteins of SARS-CoV and other coronaviruses, Antiviral Res. 109 (2014): 97-109. doi: 10.1016/j.antiviral.2014.06.013. Epub 2014 Jul 1.

12. H. Stoiber, C. Ebenbichler, R. Schneider, J. Janatova, M. P, Dierich, Interaction of several complement proteins with gp120 and gp41, the two envelope glycoproteins of HIV-1, AIDS 9(1995), 19-26. doi:10.1097/00002030-199501000-00003

13. Stoiber H, Schneider R, Janatova J, Dierich MP. Human complement proteins C3b, C4b, factor H and properdin react with specific sites in gp120 and gp41, the envelope proteins of HIV-1. Immunobiology. 193(1995): 98-113. doi:10.1016/s0171-2985(11)80158-0

14. M. C. Carroll, Complement and humoral immunity, Vaccine. 26 Suppl 8(2008): I28-33. doi: 10.1016/j.vaccine.2008.11.022.

15. L. E. Gralinski, Ti. P. Sheahan, T. E. Morrison, et al, Baric Complement Activation Contributes to Severe Acute Respiratory Syndrome Coronavirus Pathogenesis mBio 9 (2018) e01753-18; DOI: 10.1128/mBio.01753-18

16. M. Kikkert, Innate Immune Evasion by Human Respiratory RNA Viruses, J. Innate Immun. 12 (2020): 4-20. doi:10.1159/000503030

17. M. Gouy, S. Guindon, O.r Gascuel, SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building, Molecular Biology and Evolution 27 (2010), 221–224, https://doi.org/10.1093/molbev/msp259

18. D. Duhovny, R. Nussinov, H. J. Wolfson, Efficient Unbound Docking of Rigid Molecules. In Gusfield et al., Ed. Proceedings of the 2'nd Workshop on Algorithms in Bioinformatics(WABI) Rome, Italy, Lecture Notes in Computer Science 2452, pp. 185-200, Springer Verlag, 2002

19. D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, H. J. Wolfson,. PatchDock and SymmDock: servers for rigid and symmetric docking, Nucl. Acids. Res. 33 (2005): W363-367,

20. N. Andrusier, R. Nussinov and H. J. Wolfson, FireDock: Fast Interaction Refinement in Molecular Docking, Proteins 69(2007):139-159.

21. E. Mashiach, D. Schneidman-Duhovny, N. Andrusier, R. Nussinov, H. J. Wolfson. FireDock: a web server for fast interaction refinement in molecular docking, Nucleic Acids Res. (2008), 36(Web Server issue):W229-32.

22. J. Singh, M. I. Khan, S. P. Singh Yadav, et al, L-Asparaginase of Leishmania donovani: Metabolic target and its role in Amphotericin B resistance, Int J Parasitol Drugs Drug Resist. 7(2017): 337-349. doi:10.1016/j.ijpddr.2017.09.003

23. T. Bansal, E. Chatterjee, J. Singh, et al, Arjunolic acid, a peroxisome proliferator-activated receptor α agonist, regresses cardiac fibrosis by inhibiting non-canonical TGF-β signalling, J Biol Chem. 2017;292(40):16440-16462. doi:10.1074/jbc.M117.788299

24. F. Forneris, J. Wu, X. Xue, et al., Regulators of Complement Activity Mediate Inhibitory Mechanisms Through a Common C3B-Binding Mode., EMBO J. 35 (2016): 1133-1149.

25. S. Jamal; J. Singh, J. A. Sheikh, et al., Molecular Analyses of Over Hundred Sixty Clinical Isolates of SARS-CoV-2: Insights on Likely Origin, Evolution and Spread, and Possible Intervention, Preprints 2020, 2020030320 (doi: 10.20944/preprints202003.0320.v1).

26. P. Roversi, S. Johnson, J. J. E. Caesar, et al., Structural basis for complement factor I control and its disease-associated sequence polymorphisms, Proc. Natl. Acad. Sci. U S A. 108(2011): 12839–12844. Published online 2011 Jul 18. doi: 10.1073/pnas.1102167108

27. D. Ricklin, G. Hajishengallis, K. Yang, J. D, Lambris, Complement: a key system for immune surveillance and homeostasis, Nat. Immunol. 11(2010): 785-97. doi: 10.1038/ni.1923. Epub 2010 Aug 19.

28. K. A. Stoermer, T. E. Morrison, Complement and viral pathogenesis, Virology 411 (2011): 362-373. doi:10.1016/j.virol.2010.12.045

29. G. J. Kotwal, S. N. Isaacs, R. McKenzie, M. M. Frank, B. Moss, Inhibition of the complement cascade by the major secretory protein of vaccinia virus, Science 250 (1990): 827–830

30. W.K. Ip, K.H. Chan, H.K. Law, et al, Mannose-binding lectin in severe acute respiratory syndrome coronavirus infection. J. Infect. Dis. 191 (2005): 1697–1704.

31. A. Fuchs, T.Y. Lin, D.W. Beasley, et al, Direct complement restriction of flavivirus infection requires glycan recognition by mannose-binding lectin. Cell Host Microbe 8 (2010): 186–195.

32. S. R. Douglas, W. Smith, A study of vaccinal immunity in rabbits by means of in vitro methods, Br. J. Exp. Path. 11 (1930): 96–111.

33. P. Agrawal, R. Nawadkar, H. Ojha, et al., Complement Evasion Strategies of Viruses: An Overview, Front. Microbiol. 8 (2017): 1117. Published 2017 Jun 16. doi:10.3389/fmicb.2017.01117

34. Y. C. F. Su, D. E. Anderson, B. E. Young, et al., Discovery of a 382-nt deletion during the early evolution of SARS-CoV-2, bioRxiv 2020.03.11.987222; doi: https://doi.org/10.1101/2020.03.11.987222

Table 1: Protein protein interaction network between pre-simulated F1 and C3b. Residues involved in inter-molecular H-bonds and salt-bridge formation are indicated.

H-bonds	
C:ARG 430 [NH2]	B:GLU 769 [OE1]
C:SER 256 [O]	B:TYR1428 [HH]
C:GLU 270 [O]	B:TYR1447 [HH]
C:GLN 373 [OE1]	B:GLN 983 [NE2]
C:THR 377 [O]	B:ARG 979 [NH2]
C:ARG 388 [O]	B:ARG 979 [NH2]
C:ASP 447 [OD2]	B:TYR1482 [HH]
C:THR 448 [OG1]	B:TYR1483 [HH]
Salt bridges	
C:ARG 430 [NH2]	B:GLU 769 [OE1]
C:GLU 470 [OE1]	B:LYS1360 [NZ]

Table 2. Protein-protein interaction network between ORF8 and C3b. Residues involved in inter-molecular H-bonds and salt-bridge formation are indicated.

H-Bonds	
A:ILE 58 [N]	B:TYR1482 [OH]
A:ASP 107 [N]	B:GLU1487 [OE2]
A:SER 21 [O]	B:HIS1349 [NE2]
A:SER 24 [OG]	B:ASN 939 [ND2]
A:GLU 106 [OE1]	B:ASN1484 [ND2]
A:ASP 107 [OD2]	B:LYS1360 [N]
A:ASP 113 [O]	B:GLU1486 [N]
Salt-bridges	
A:LYS 2 [NZ]	B:ASP 832 [OD2]
A:HIS 28 [NE2]	B:GLU1486[OE1]
A:ASP 113 [OD1]	B:LYS1478 [NZ]
A:ASP 113 [OD2]	B:LYS1478 [NZ]

Figure Legends

Figure 1. A. Shannon entropy as a measure of variation protein sequence alignment of 1042 sequences of ORF8 (121 residues). The amino acid variation plot is used to calculate the entropy at each position in a sequence set. **B.** Multiple sequence alignment between (1) Three ORF8a (1-39 amino acid) of SARS-CoV; (2) Three ORF8b (1-39 amino acid) of SARS-CoV; (3) The consensus sequence of ORF8 (1-121 amino acid) of SARS-CoV-2 with major prevalent distribution of amino acids at positions - 24 serine (S), 62 valine (V) and 84 leucine (L); and (4) SARS coronavirus ORF8 of a Rhinolophus species. The shades of blue represents the identical residues (dark blue) and similar residues (light blue).

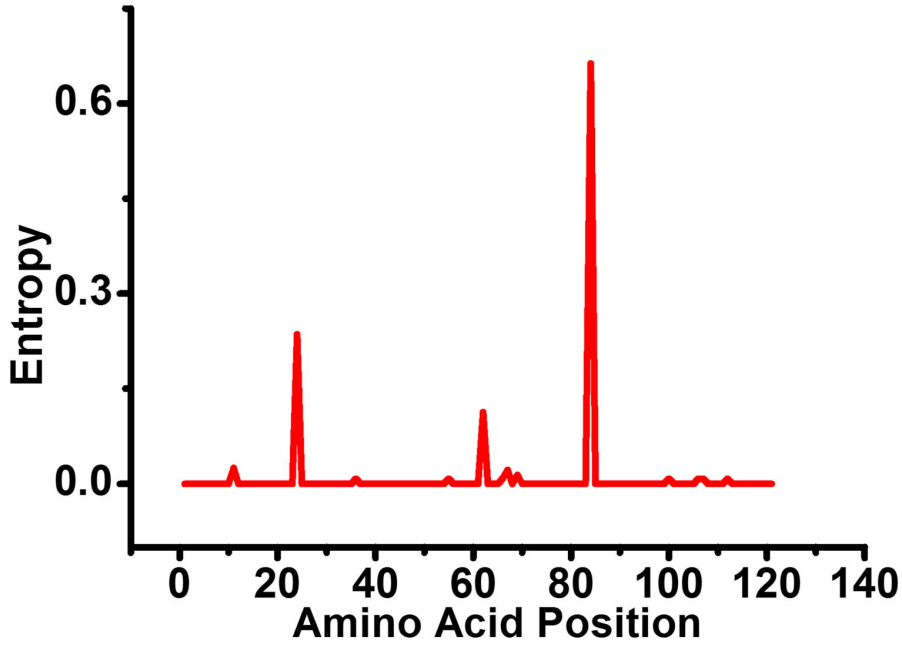
Figure 2. (A) Sequence alignment between human complement Factor 1 (PDB Sequence ID 2XRC_A) amino acid number 500 to 557 and amino acid number 58 to 115 of the consensus ORF8 protein from SARS-CoV-2. The residues are 25% identical and 48% similar (positives) **(B)** Cartoon showing the homology mapping of ORF8 on F1 alpha chain (pdb 2XRC_A / gi 339961198).

Figure 3. Modelling and Protein-protein docking of ORF8/F1 complement with C3b human complement. **(A)** Superposition of modelled ORF8 (Red) upon the human F1 complement (Cyan) as template (PDB id: 2xrc). **(B)** Highest ranked model obtained from protein-protein docking of F1 complement (Cyan) and human C3b complement (Grey). **(C)** Highest ranked model obtained from protein-protein docking of ORF8 (Red) and human C3b complement (Grey). Blue spheres indicate Arg-Ser (RS) protease cleavage sites encompassing C3b CUB (C1r/C1s, Uegf, Bmp1) domain near the docked complexes. The arrows define outcomes from protein-protein docking of C3b with human F1complement (Cyan arrow) and ORF8 (Red arrow).

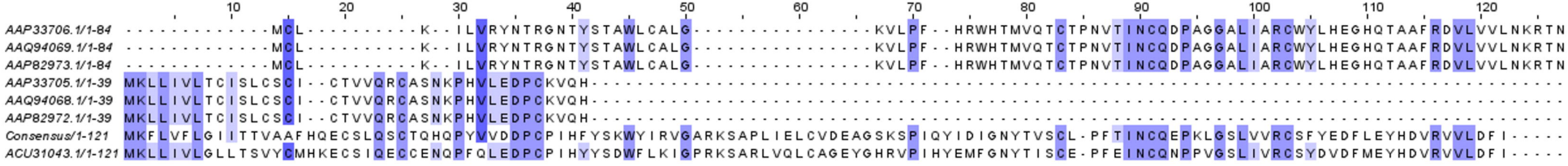
Figure 4. Modelling and Protein-protein docking of mORF8 with C3b human complement. **(A)** Modelled ORF8 (Brown) showing three high entropy mutations; S24L, V62L and L84S (Green Sticks) **(B)** Highest ranked model obtained from protein-protein docking of mORF8 (Brown) and human C3b complement (Grey). **(C)** Low energy model obtained from protein-protein docking of mORF8 (Brown) and human C3b complement (Grey) at F1 binding site. The arrows (Black) define outcomes from protein-protein docking of C3b with mORF8 at its two different sites.

Figure 5. MD simulations of top ranked F1-C3b and ORF8-C3b docked complexes. Evolution of H-bonds between F1 factor and C3b complement complex (Cyan) and ORF8-C3b complex (Red) over the simulation period of 25 ns.

A


Entropy
0.0
0.3
0.6
0
20
40
60
80
100
120
140
Amino Acid Position

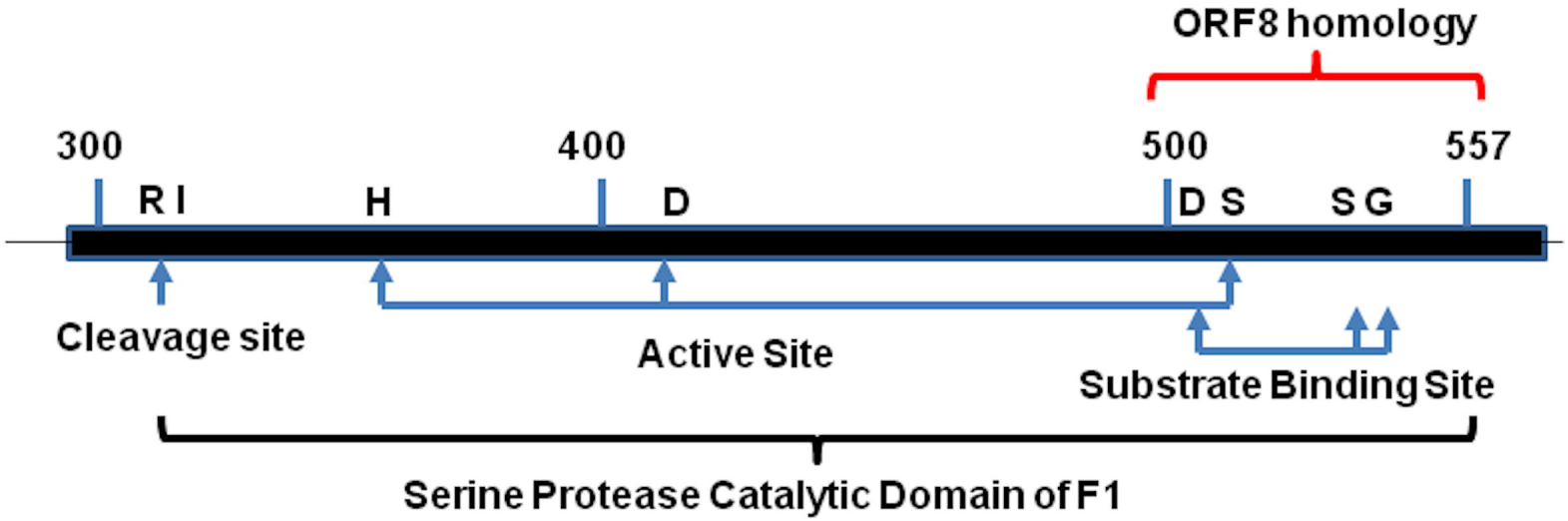
B

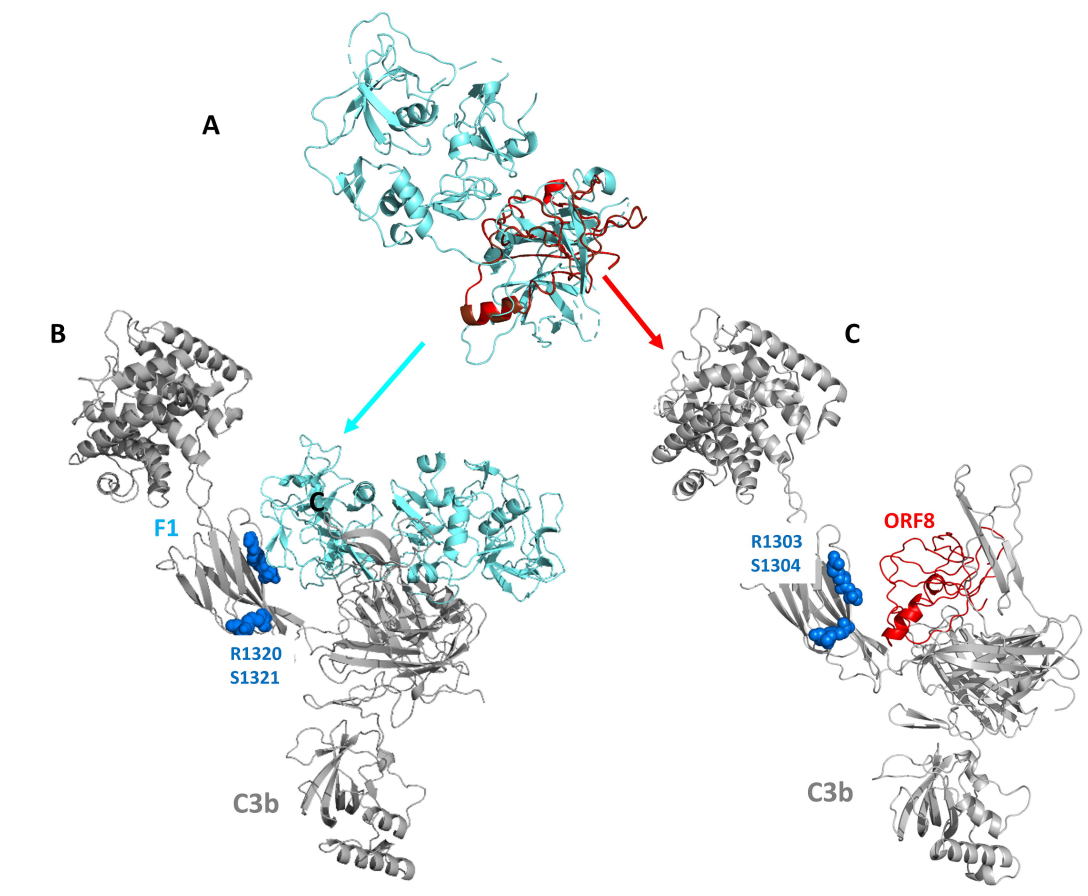

10
20
30
40
50
60
70
80
90
100
110
120
AAP33706.1/1-84
AAQ94069.1/1-84
AAP82973.1/1-84
AAP33705.1/1-39
AAQ94068.1/1-39
AAP82972.1/1-39
Consensus/1-121
ACU31043.1/1-121

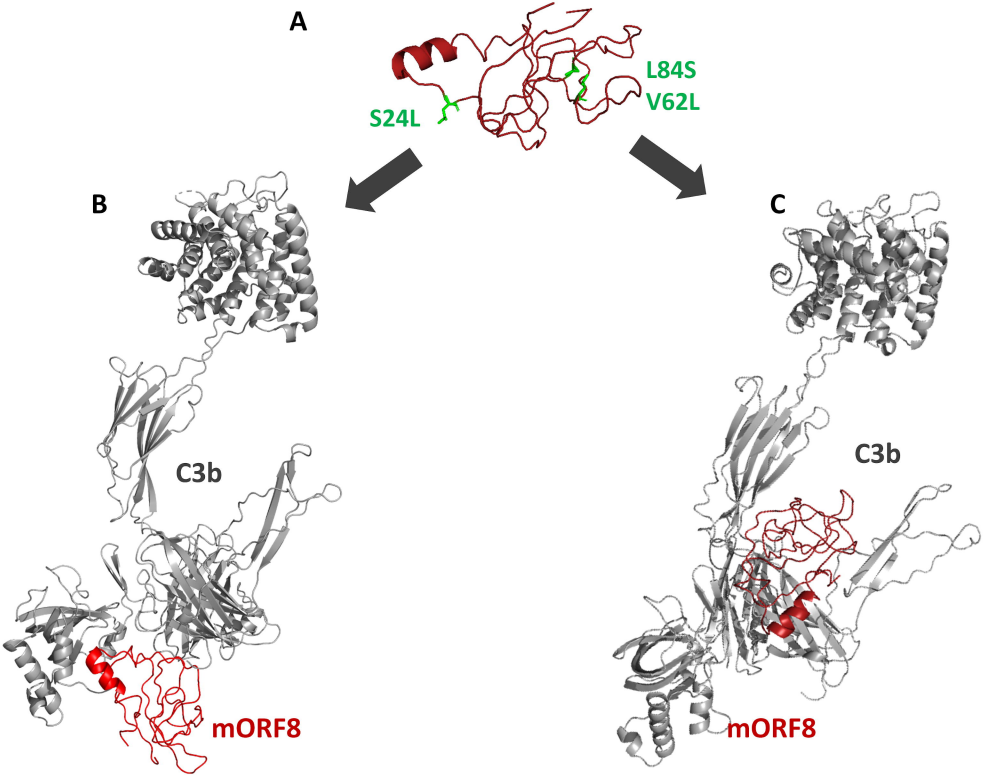
A.

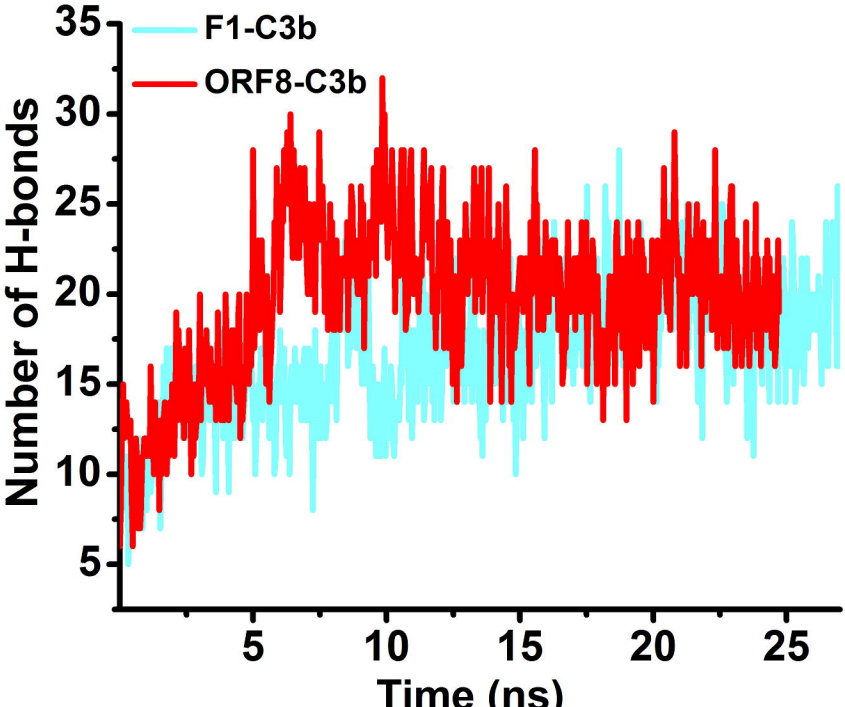
```
F1     500  IDACKGDSGG--PLVCMDANNVTYVWGVVSWGENCGKPEFPGVYTKVANYFDWISYHVGR   557
            I+ C  ++G   P+  +D  N  Y    + +  NC +P+   +  + + Y D++ YH  R
ORF8    58  IELCVDEAGSKSPIQYIDIGN--YTVSCLPFTINCQEPKLGSLVVRCSFYEDFLEYHDVR   115
```

B.




A
B
C
F1
C
R1320
S1321
C3b
R1303
S1304
ORF8
C3b


A
S24L
L84S
V62L
B
C3b
mORF8
C
C3b
mORF8


F1-C3b
ORF8-C3b
Number of H-bonds
35
30
25
20
15
10
5
5
10
15
20
25
Time (ns)