Chapter 7
Evolutionary and Structural Studies of NCoV and SARS CoV-Spike proteins and their association with ACE2 Receptor

Abstract Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2)/Novel Corona Virus Disease-19 (nCOVID-19)/COVID-19 has only been discovered recently, and so our understanding of the disease epidemiology is continuously evolving. WHO has declared it a worldwide pandemic with high morbidity and significant mortality, hence it has been announced as the global health and wealth emergency. At present there is no any specific therapy available to fight against this virus, hence the drug repositioning is the most challenging to entire scientific community. The aim of this study is to determine the mutation(s) in the sequence of the spike protein, which plays a significant role in transmission from human to human. By using bioinformatics approach first we analyzed spike protein sequence of four nearest coronavirus family that include COVID-19, bat coronavirus RaTG13, pangolian coronavirus and SARS CoV, to determine phylogenetic distance between them. The homology modeling of COVID-19 spike protein has been done by iTASSER. and the protein-protein docking with human receptor ACE2 by Frodock web based docking tool showing the less binding energy of COVID-19 (−12.7 kcal/mol) in comparison with SARS CoV (10.3 kcal/mol). Further, the superimposed structure of COVID-19 and SARS CoV viruses has been performed to find the mutational site in association with human ACE2 protein. The extensive and detailed computational analyses approaches help to identify conserved region of COVID-19 and SARS CoV. Hence, our present data might help to identify potential target site and to develop antiviral drugs/vaccine to combat this pandemic.

Keywords COVID-19 · SARS CoV · ACE2 receptor · Protein docking

7.1 Introduction

Currently, the emergency has been declared by WHO due to novel human coronavirus sporadic outbreaks in different countries. The first case of novel coronavirus (nCoV-19) has been detected in China in December, 2019, where, patients presenting with viral pneumonia like symptom caused by severe acute respiratory syndrome

© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd. 2020

A. Kumar et al., *Novel Coronavirus 2019*,
SpringerBriefs in Forensic and Medical Bioinformatics,
https://doi.org/10.1007/978-981-15-7918-9_7

coronavirus 2 (SARS-CoV-2), also known as coronavirus disease 2019 (COVID-19) zoonotic origins belongs to family coronaviridae, genus betacoronavirus [1]. These viruses mostly infect animals, including birds and mammals. In human, it can be transmitted from person to person [3, 4]. In humans, they generally cause mild respiratory infections, such as those observed in the high fever, dry cough with breathing issue. However, some recent human coronavirus infections have resulted in lethal endemics, which include the SARS and MERS (Middle East Respiratory Syndrome) endemics. Both of these are caused by zoonotic coronaviruses that belong to the same genus β-coronavirus within coronaviridae. SARS-CoV originated from Southern China and caused an endemic in 2003. A total of 8098 cases of SARS were reported globally, including 774 associated deaths, and an estimated case-fatality rate of approx. 15% [2]. The first case of MERS-CoV occurred in Saudi Arabia in 2012. Since then, a total of 2494 cases of infection have been reported, including 858 associated deaths, and an estimated high case-fatality rate of 34.4%. However no case of SARS-CoV infection has been reported since 2004, similarly MERS-CoV since 2012 (WHO 2011) remains undetected. Presently, COVID-19 has spread globally to about 200 countries. Globally 3,076,185 corona cases and 211,941 deaths has been recorded (April, 2020) [3, 4] and still counting is going on.

Novel Corona Viruses are very long (32 Kbp) positive sense single-stranded RNA viruses and their structure includes four main structural proteins: the spike, membrane, envelope protein, and nucleocapsid. Viral membrane protein and peptides involved in replication of viral genetic material play an integral part in virus host interaction [5]. The spike protein of coronavirus virion particles plays a significant role in the recognition of angiotensin-converting enzyme 2 (ACE2) [6, 7]. ACE2 belongs to the renin-angiotensin-aldosterone system (RAAS), which involves in regulating blood pressure, hypertension cardiovascular and renal diseases by regulating homeostasis of blood pressure, maintain electrolyte and inflammatory activities. The renin enzyme, mainly generated in the kidney and it cleaves angiotensinogen to angiotensin I (Ang I); the angiotensin-converting enzyme 2 (ACE2) cleaves Ang I to produce angiotensin II (Ang II), a key effector of the RAAS [9]. Due to alteration in ACE2, the catalytic function modulates RAAS activity, resulting in enhanced inflammation and vascular permeability observed in the pathogenesis of inflammatory lung disease [10].

Due to lack of sufficient and accurate information about this virus, the computational approach offers a method to test hypotheses of new acknowledged target site receptor with viral spike protein; hence the efficiency of viral infection is strongly dependent on virus-associated protein-protein interactions. Various metabolites are associated with protein-protein interfaces and describe the type of chemical changes occurring between ligand and target site receptor. Thus, the rationale behind this study to determine the mutation and potential drug targets to evaluate the energetic profile of the interaction between the COVID-19 spike protein and the human cell receptor ACE2.

7.2 Materials and Methods

7.2.1 Collection of Sequences

Coronavirus family spike protein sequences were retrieved from National Center for Biotechnology Information (NCBI) protein sequence data base (https://www.ncbi.nlm.nih.gov/protein) with their reference number such as COVID-19 (YP_009724390.1), Bat Coronavirus RatG13 (QHR63300.2), Pangolian coronavirus (QIQ54048.1) and SARS CoV (ACU31032.1).

7.2.2 Phylogenetic Analysis

Mega 6.0 software has been used to construct the phylogenetic tree to establish the relationship between these four coronavirus family. Alignment of the full-length coronavirus spike proteins was performed by MUSCLE with default parameters. The neighbor joining (NJ) tree was computed from the pairwise phylogenetic distance matrix creation [11].

7.2.3 Protein Structure Homology Modeling by ITASSER

I-TASSER (Iterative Threading ASSEmbly Refinement) is a hierarchical approach to protein structure and function prediction. It first identifies structural templates from the PDB by multiple threading approach LOMETS, with full-length atomic models constructed by iterative template-based fragment assembly simulations. To create structural models of the full length COVID-19 spike protein the SARS-CoV spike glycoprotein (5XLR) has been used as a template for modeling [12, 13].

7.2.4 Protein-Protein Docking

The protein–protein complexes from the predicted structure of NCoV and ACE2 human receptor were (PDB id 1R42) downloaded from protein data bank. Further, we use FRODOCK software web based user-friendly protein–protein docking server for interaction between the viral spike protein and ACE2 receptor using molecular docking [14].

7.2.5 *Structural and Functional Analysis of NCoV 19 and SARS-CoV Using Human ACE2*

NCoV 19 spike protein with ACE2 human binding receptor generated by FRODOCK and SARS-CoV with human ACE2 binding receptor (PDB id 3D0G) [15] were downloaded from the Protein Data Bank. Then, binding patterns and affinity estimations for the interaction between the viral spike and ACE2 receptor were performed using Mol Star tools for web molecular graphics [16].

7.3 Results

Through bioinformatics analysis phylogenetic studies reveal that Novel CoV belongs to a group containing SARS-CoV family. The spike glycoprotein is approximately 97% similar to bat coronavirus, 90% to pangolin coronavirus and 80% closest to SARS CoV shown in Fig. 7.1.

The 3D structures of the COVID-19 spike protein (QHD43416.1.pdb) and SARS-CoV (3D0G) interacting with the receptor binding domain (RBD) site in human ACE2 were analyzed by Frodock web based protein-protein docking tool showing (−12.7 kcal/mol) and (−10.3 kcal/mol) respectively as shown in Tables 7.1 and Fig. 7.2a, b. The interaction pattern between the viral spike proteins is quite similar in COVID-19 as well as in SARS CoV. In the case of COVID-19 total 38 amino acid

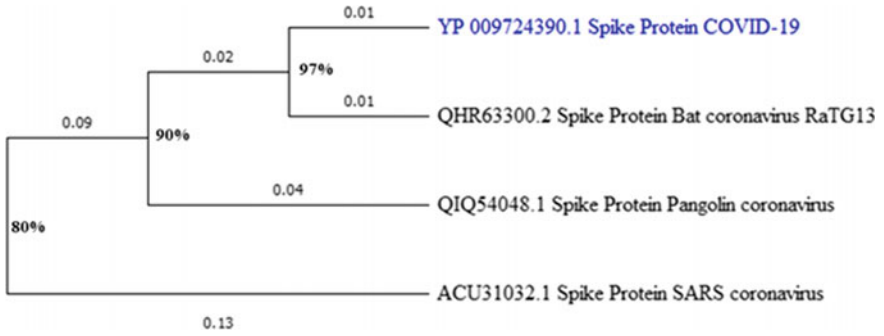

Fig. 7.1 Showing the Phylogenetic tree of Wuhan COVID-19 Spike glycoprotein sequences in context to nearest corona virus families drawn by MEGA 6.0

Table 7.1 Binding affinity (ΔG) and dissociation constant (Kd) predicted values for the interaction between viral spike and ACE2 receptor

Protein-protein complex (viral spike/ACE2)	ΔG (kcal/mol)
SARS-CoV-2	−12.7
SARS-CoV	−10.3

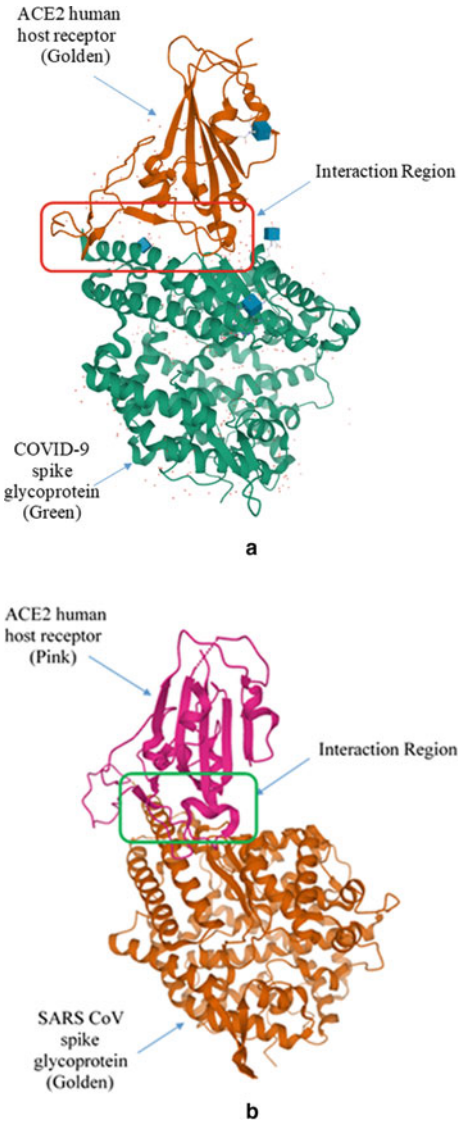

Fig. 7.2 Showing interacting binding sites **a** COVID-19 spike glycoprotein (Green) with human ACE2 receptor (Golden), **b** SARS-CoV spike glycoprotein (golden) with human ACE2 receptor (pink)

Table 7.2 Illustration of Interaction between COVID-19 and SARS-CoV spike protein with human ACE2 receptor

Protein-protein complex (viral spike/ACE2)	ΔG (kcal/mol)
SARS-CoV-2	−12.7
SARS-CoV	−10.3

residues interact with ACE2 receptor while in the case of SARS CoV total 35 amino acid residues with 22 residues of ACE2 receptor illustrated in Table 7.2.

Human ACE2 receptor COVID-19 (SARS-CoV 2) spike protein residues SARS-CoV spike protein residues

S19 A475 R426
Q24 N487 N473
T27 F456 L443, Y475
F28 Y489 Y475
D30 K417 Y442
K31 E484, F490, Q493 Y442, Y475
H34 L455, Q493 Y440, Y442, N479
E35 Q493
E37 Y505 Y491
D38 Y449, G496, Q498 Y436, G482, Y484
Y41 Q498, T500, N501 Y484, T487
Q42 G446, Y449, Q498 Y475
L45 C498, T500 T486
L79 F486 L472
M82 F486 L472
Y83 N487, Y489 N473
N330 T500 T486
K353 G496, N501 G502, Y505 Y481, G482, T487, G488, Y491
G354 G502, Y505 G488
D355 T500, G502 T487
R357 T500 G488
R393 T500 I489
Q325 R426, I489
E329 R426

In Fig. 7.3, we present a space-fill superimposed structure of spike protein showing high similarity with SARS CoV structure. The overall structure covers approx. 80% of the residues in the full-length sequence, with several important residues having in grey color, whereas red colour on superimposed structure showing mutated region and cartoon structure in cyan color representing ACE2 interaction, our following analysis based on Global Initiative on Sharing All Influenza Data (https://www.gisaid.org/hcov-19-analysis-update/) [17].

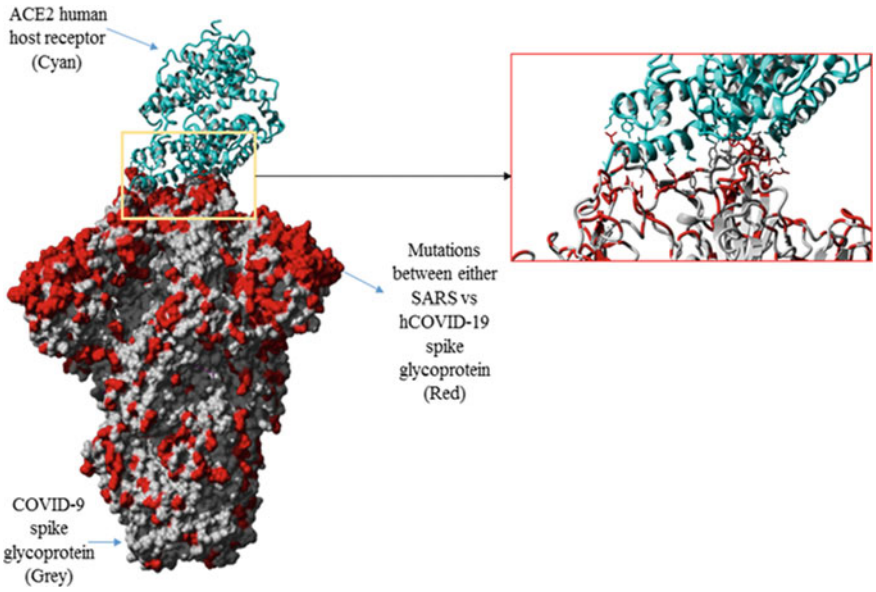

Fig. 7.3 Superimposed structure of viral spike glycoprotein of COVID-19 (grey) and SARS CoV with ACE2 (cyan) and a complex structure also showing mutational site (red) either in COVID-19 or SARS CoV [17]

7.4 Discussion

The ongoing COVID-19 pandemic makes us painfully realize that our current situation is only based on isolation from society and maintain hygienic condition to protect us from this unpredictable behavior of this virus. Earlier the outbreaks of SARS CoV in 2003 and MERS-CoV in 2012 has provide us extensive research efforts, but unfortunately there were no any drug which protect us from any type zoonotic coronavirus. Earlier, strains of these viruses were not much highly spreadable and unable to infect worldwide. But, this virus has very fluctuating and unstable strain due to their epitopic nature of the virus that make a challenges to worldwide scientific community to develop antiviral therapeutics. Due to this nature of virus earlier there was no any prototype of drug for coronavirus was progressed. After 17 years of SARS CoV epidemic and current COVID-19, emerging coronaviruses are very similar with their infection site. Therefore, our approach based on systematic comparison with these two viruses and associated mutation.

In this study, we presented a bioinformatics based methodology for systematic interaction and identification of similarity between COVID-19 and SARS CoV viral spike protein in association with ACE2 receptor binding domain. The host ACE2 has been proved by many studies to be the specific receptor for the Spike RBD of SARS-CoV [18]. The latest research shows that the host receptor of COVID-19 is consistent with SARS-CoV, exhibiting that the Spike RBD sequence of COVID-19

is similar to SARS-CoV RBD and there are important interactions between several key amino acid residues of RBD receptor-binding motif and ACE2 [8].

However, our data show the differences of interaction with considerably less favorable binding energies between these viruses with ACE2 [19], as shown in this Table 7.1. Thus, the loops observed in the spike protein of COVID-19 in the amino acids mutation (substitutions or deletion) considered as key factor with ACE2 binding [20]. Mutations in the spike protein could change the physiochemical activity of a virus which increases viral pathogenesis [21]. The presence of square bracket in (Fig. 7.2a, b) around the interaction region might be promoting the interface with ACE2 receptor, which illustrate the binding to this receptor and interaction between amino acid residues (Table 7.2).

Interestingly, mutational site could play an important clue to determine the host receptor specificity [22] for the viral spike protein which is responsible for increasing infection and viral spreading. The comparative studies to determine the impact COVID-19 and SARS CoV mutational site are quite similar with respect to receptor binding domain, which helpful and required to predict possible zoonotic event in the future as well as develop therapeutics. However, the present data might play a significant role into develop antiviral drugs and vaccines to stop the COVID-19 disease with unpredictable death.

7.5 Conclusion

Drug/vaccine development against the COVID-19 is a challenging for scientific community worldwide due to their frequent recombination events. We need explore this study on system biology to accelerate the structural and functional details of the life cycle of the COVID-19 and their mode of action. Again, as a preventive measure and strict observation of viral changes in different hosts for the prediction of an event is important aspect. Based on the current research progress, ACE2 is considered as the host potential target site for the treatment of coronavirus infection to block COVID-19 from entering host cells.

References

1. Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395:497–506
2. World-Health-Organization (2003) Update 49—SARS case fatality ratio, incubation period. Available online: https://www.who.int/csr/sars/archive/2003_05_07a/en/. Accessed 31 Jan 2020
3. World-Health-Organization (2020a) Middle East respiratory syndrome coronavirus (MERS-CoV). Available online: https://www.who.int/emergencies/mers-cov/en/. Accessed 31 Jan 2020

4. World Health Organization (WHO) (2020b) Novel Coronavirus (COVID-19) situational reports. Available online https://www.who.int/emergencies/diseases/novelcoronavirus-2020/situation-reports/98
5. Jyothsna G, Kumar A, Kashyap A, Saxena AK, Sanyal A (2020) In search of novel coronavirus 19 therapeutic targets. Helix 10(02):01–08. Retrieved from https://helixscientific.pub/index.php/Home/article/view/98
6. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, Si H-R, Zhu Y, Li B, Huang C-L, Chen H-D, Chen J, Luo Y, Guo H, Jiang R-D, Liu M-Q, Chen Y, Shen X-R, Wang X, Zheng X-S, Zhao K, Chen Q-J, Deng F, Liu L-L, Yan B, Zhan F-X, Wang Y-Y, Xiao G-F, Shi Z-L (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270–273
7. Letko M, Munster V (2019) Functional assessment of cell entry and receptor usage for lineage B β-coronaviruses, including 2019-CoV. bioRxiv 2020, No. 2020.01.22.915660
8. Wan Y, Shang J, Graham R, Baric RS, Li F (2020) Receptor recognition by novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS. J Virol 94:e00127e20
9. Jia H (2016) Pulmonary angiotensin-converting enzyme 2 (ACE2) and inflammatory lung disease. Shock 46:239–248
10. Burrell LM, Johnston CI, Tikellis C, Cooper ME (2004) ACE2, a new regulator of the renin-angiotensin system. Trends Endocrinol Metab 15:166–169
11. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5):1792–1797
12. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER Suite: protein structure and function prediction. Nat Methods 12:7–8
13. Huang X, Pearcs R, Zhang Y (2020) Computational design of peptides to block binding of the SARS-CoV-2 spike protein to human ACE2. BioRxiv Prinprint https://doi.org/10.1101/2020.03.28.013607
14. Ramírez-Aportela E, López-Blanco JR, Chacón P (2016) FRODOCK 2.0: fast protein-protein docking server. Bioinformatics 32(15):2386–2388
15. Li F (2008) Structural analysis of major species barriers between humans and palm civets for severe acute respiratory syndrome coronavirus infections. J Virol 82 (14) 6984-6991
16. Sehnal D, Rose AS, Koča J, Burley SK, Velankar S (2018) Mol: towards a common library and tools for web molecular graphics. In: Proceedings of the workshop on molecular graphics and visual analysis of molecular data (MolVA'18). Eurographics Association, 2018. Goslar, DEU, 29–33
17. GISAID (2020) Genomic epidemiology of novel coronavirus. Available at https://nextstrain.org/ncov 2020 Last accessed 9 March 2020
18. Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH et al (2013) Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature 503:535e8
19. Agrawal P, Singh H, Srivastava HK et al (2019) Bench marking of different molecular docking methods for protein-peptide docking. BMC Bioinform 19:426
20. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q (2020) Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. Science, epub ahead of print
21. Shang J, Wan Y, Liu C, Yount B, Gully K, Yang Y et al (2020) Structure of mouse coronavirus spike protein complexed with receptor reveals mechanism for viral entry. PLoS Pathog 16(3)
22. Gupta D, Kumar A (2013) Prospects for drug designing: similar conserved interactions of Bim with MCL-1 AND BCL-2