Indian Journal of Biochemistry & Biophysics
Vol. 59, May 2022, pp. 536-551

Structure-function relationships among selected human coronaviruses

Esther Jamir[1,2], Kikrusenuo Kiewhuo[1,2], Lipsa Priyadarsinee[1,2], Himakshi Sarma[1], Selvaraman Nagamani[1,2] & G Narahari Sastry[1,2]*

[1]Advanced Computation and Data Sciences Division, CSIR-North East Institute of Science and Technology, Jorhat-785 006, Assam, India

[2]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad-201 002, Uttar Pradesh, India

Received 15 February 2022; *revised 19 May 2022*

Identifying the key proteins among different types of human disease-causing coronaviruses is essential for the molecular mechanism and thereby designing potential drug molecules. Eight selected proteins of seven types of disease-causing coronaviruses, *viz.*SARS-CoV-2 (severe acute respiratory syndrome coronavirus2), SARS-CoV (severe acute respiratory syndrome coronavirus), MERS-CoV (middle east respiratory syndrome coronavirus), Human coronavirus OC43, Human coronavirus HKU1, Human coronavirus 229E and Human coronavirus NL63, were chosen for the comparison. Further, an attempt has been made to explore the most important host-pathogen interactions with a special focus on spike (RBD) protein region as this region deemed to be functionally most important. Epitope region was also identified which helps in the design of epitope-based vaccines. The structural comparison carried out among the seven types of human coronaviruses has revealed the molecular level details on the similarity among this series. This study has facilitated the identification of the important residues in the studied proteins which control the key functions such as viral replication and transmission. Thus, exploring the protein space in the family of coronaviruses, provide valuable insights into the molecular basis associated with the role of proteins and viral infections, which is expected to trigger the identification of the drug targets for coronaviruses infections, in a rational way.

Keywords: Hotspot residues, Multiple sequence alignment, Protein-protein interaction, Seven human coronaviruses

Coronaviruses are enveloped, single-stranded, positive-sensed RNA viruses belonging to the family Coronaviridae with genomes ranging from 26 to 32 kb in length (Table S1).Several known strains of coronaviruses such as OC43, HKU, 229E5, and NL63 are pathogenic to humans and associated with mild common cold symptoms. These viruses are responsible for ~5-30% of common colds.In the past two decades, three notable coronaviruses (Severe Acute Respiratory Syndrome Coronavirus, SARS-CoV; Middle East Respiratory Syndrome Coronavirus MERS-CoV and Severe Acute Respiratory Syndrome Coronavirus 2, SARS-CoV-2) have emerged and caused severe clinical symptoms, including acute respiratory distress syndrome. These three viruses are prone to infect different human body parts resulting to acute lung injury, septic shock, multi-organ failure with high case fatality ratio (CFR)[1-10].

The coronaviruses have been categorized into four genera namely alpha, beta, gamma and delta (Table 1)[6]. Among these viruses alpha and beta coronaviruses seek more attention due to their ability to cross animal-human barriers and thus they become major human pathogens[7]. The global pandemic caused by SARS-CoV was reported in November 2002, in Foshan, China was the first serious concern worldwide with its devastating outcome[11].

The second dreadful HCoV is MERS CoV and it was emerged a decade later in April 2012 and the first incidence was occurred in Jordan. The most recent laboratory confirmed MERS CoV patients were reported in Riyadh on 28th March 2020[12]. Recently, the pandemic caused by SARS-CoV-2 first occurred in Wuhan, China in December 2019 and spread across the globe. The spread of SARS-CoV2 has been declared as a public health emergency of International concern[11]. All these viruses are seasonal, generate short-term immunity and reinfections are observed within a year[13]. All HCoVs can cause severe complications in the elderly and immune compromised individuals[14]. Despite the fact that all HCoVs share similarities in viral replication, they differ in their accessory proteins, host receptor, incubation period and pathogenicity[8]. The SARS-CoV-2 transmissibility is higher than other human coronaviruses, and thus it poses a threat to the populated communities[15].

*Correspondence:
E-mail: gnsastry@gmail.com
Suppl. Data available on respective page of NOPR

Table 1 — The four different coronavirus families and their sub types. The human disease-causing coronaviruses has been highlighted in bold

Genera	Species	Genetic material	Transcription regulatory sequences
Alphacoronavirus	Transmissible gastroenteritis virus (TGEV)	RNA	CUAAAC
	Feline coronavirus	RNA	CUAAAC
	Canine coronavirus	RNA	CUAAAC
	Human coronavirus 229E	RNA	CUAAAC
	Human coronavirus NL63	RNA	CUAAAC
	Miniopterus bat coronavirus 1	RNA	CUAAAC
	Miniopterus bat coronavirus HKU8	RNA	CUAAAC
	Porcine epidemic diarrhea virus	RNA	CUAAAC
	Rhinolophus bat coronavirus HKU2	RNA	CUAAAC
	Scotophilus bat coronavirus 512	RNA	CUAAAC
Betacoronavirus	Bovine Coronavirus,	RNA	CUAAAC
	Human coronavirus OC43)	RNA	CUAAAC
	Hedgehog coronavirus 1	RNA	NA
	Human coronavirus HKU1	RNA	CUAAAC
	Middle East respiratory syndrome-related coronavirus	RNA	NA
	Murine coronavirus	RNA	NA
	Pipistrellus bat coronavirus HKU5	RNA	ACGAAC
	Rousettus bat coronavirus HKU9	RNA	ACGAAC
	Severe acute respiratory syndrome–related coronavirus (SARS-CoV)	RNA	ACGAAC
	Severe acute respiratory syndrome–related coronavirus (SARS-CoV2)	RNA	ACGAAC
	Tylonycteris bat coronavirus HKU4	RNA	ACGAAC
Gammacoronavirus	Avian coronavirus	RNA	CUUAACAA
	Beluga whale coronavirus SW1	RNA	AAACA
Deltacoronavirus	Bulbul coronavirus HKU11	RNA	ACACCA
	Porcine coronavirus HKU15	RNA	ACACCA

*NA = Not Available

In this study, an attempt was made to understand the sequence and structural level difference among the seven HCoVs which may shed light on why the transmissibility and infection rate is higher in case of SARS-CoV-2 as compared to other HCoVs. It is well known that the receptor binding domain of SARS-CoV-2 spike protein is getting mutated and subsequently it increases the binding affinity of spike protein with the host receptor ACE2 (Angiotensin-converting enzyme 2)[16,17]. Other than the spike protein, we considered seven other proteins that may affect viral replication and transcription. A comparative study of eight (spike, NSP3, NSP5, NSP12, NSP13, NSP14, NSP15, NSP16) selected proteins of human coronaviruses (HCoVs) have been performed at sequence level as well as at structural level[18-20]. Multiple sequence alignment has been carried out to analyze the sequence identity/similarity and mutations[21,22]. The wild and mutant amino acids were analyzed using PredictSNP[23]. We also predicted B-cell epitope for 8 selected proteins of seven HCoVs. Identification of B-cell epitopes play an important role in different biomedical applications including disease control, diagnostics, and vaccine development[24,25]. The difference in the binding pattern and binding energy of spike with its respective host receptor of the seven HCoVs was also identified and studied in detail. In addition, the relative binding affinity of host receptor and spike proteins of all the HCoVs was compared using prodigy server (https://wenmr.science.uu.nl/prodigy/). Among the seven types of human coronaviruses, spike protein of SARS-CoV-2, SARS-CoV and NL63 have common host receptor ACE2 (Angiotensin-converting enzyme 2), MERS binds toDPP4 (Dipeptidyl peptidase 4), 229E binds with APN (Amino peptidase N) while spike of OC43 and HKU1 binds to the sialoglycan-based receptors with 9-O-acetylated sialic acid (9-O-Ac-Sia). Thus, analyzing different coronavirus proteins and their interacting partner in human will shed light in identifying hotspot residues. Moreover, the advancement in computational drug discovery and understanding the manifestation of non-covalent interactions in biological systems expedite the drug discovery process[26-29].

Over all in this study, by investigating the eight important selected proteins of 7 HCoVs may shed

light in understanding the difference among them at sequence and structural level, difference in the binding patterns of spike with its receptors. Our observation in the current study may help to identify potential inhibitors against the selected proteins and to disrupt the host protein and spike protein interactions. Figure 1 represents the workflow of the current study.

Material and Methods

Data Collection

Retrieval of protein sequence and protein structure

The sequences of selected proteins NSP3, NSP5, NSP12, NSP13, NSP14, NSP15, NSP16 and spike of the 7 HCoVs, *viz*. OC43, HKU1, 229E, NL63, SARS-CoV2, SARS-CoV and MERS-CoV, were obtained from NCBI (http://www.ncbi.nlm.nih.gov/)[30]. The information on retrieved protein sequence is given in (Table S2).

The 3D structuresof the selected proteins NSP3, NSP5, NSP12, NSP13, NSP14, NSP15, NSP16 and spike of the 7 (HCoV) *viz.* OC43, HKU1, 229E, NL63, SARS-CoV2, SARS-CoV and MERS-CoV, were retrieved from RCSB PDB databank (https://www.rcsb.org/)[31]. For the unavailable 3D crystal structures, the proteins were modelled using protein sequences in SWISS-MODEL (https://swissmodel. expasy.org/)[32]. The modelled proteins were further validated using Ramachandran (RC) plot available at the Structure Analysis and Verification Server (SAVES) server (https://saves.mbi.ucla.edu/)[33]. The information on PDB ID of the retrieved 3D protein is given in (Table S2).

Sequence analysis of the seven types of human coronaviruses

Multiple sequence alignment

The multiple sequence alignment (MSA) of selected 8 proteins of the 7 HCoVs were performed using Clustal omega program available at EBI (European Bioinformatics Institute) server https://www.ebi.ac.uk/Tools/msa/clustalo/[34]. Clustal omega is a new MSA tool that uses Hidden Markov Model (HMM) profile-profile comparison techniques for the alignments between three or more sequences and generates a guide tree along with identity percentage matrix[35-37]. The aligned sequences were exported and then analyzed to construct phylogenetic tree using freely available phylogenetic tree tool in Vipr (https://www.viprbrc.org/)[38].

Mutational analysis of SARS-CoV2 selected proteins

The NSP3, NSP5, NSP12, NSP13, NSP14, NSP15, NSP16 and spike of SARS-CoV and SARS-CoV2 proteins sequences were given as input in Predict SNP server (https://loschmidt.chemi.muni.cz/predictsnp1/)[39] in order to perform mutational analysis. All the sequences were aligned in the server to identify the mutated residues present in SARS-CoV2.

B-cell epitope prediction

The linear B cell epitope of NSP3, NSP5, NSP12, NSP13, NSP14, NSP15, NSP16 and spike proteins among 7 HCoVs was predicted using ABCpred (http://crdd.osdd.net/raghava/abcpred/)[40]and BepiPred 2.0 (http://www.cbs.dtu.dk/services/BepiPred-2.0/)[41] online servers and the common predicted linear B cell epitopes from two methods were considered for the further analysis.

Protein structural analysis of the seven types of coronaviruses

Superimposition of proteins

TM-align program (https://zhanglab.ccmb.med.umich.edu/TMalign/)[42] was used to perform the structural comparison of the proteins present in the 7 HCoVs. Random structural similarity was determined by calculating TM-score and root-mean-square deviation (RMSD. The cut-off for TM-score is between 0.0 and 0.3 and for root-mean-square

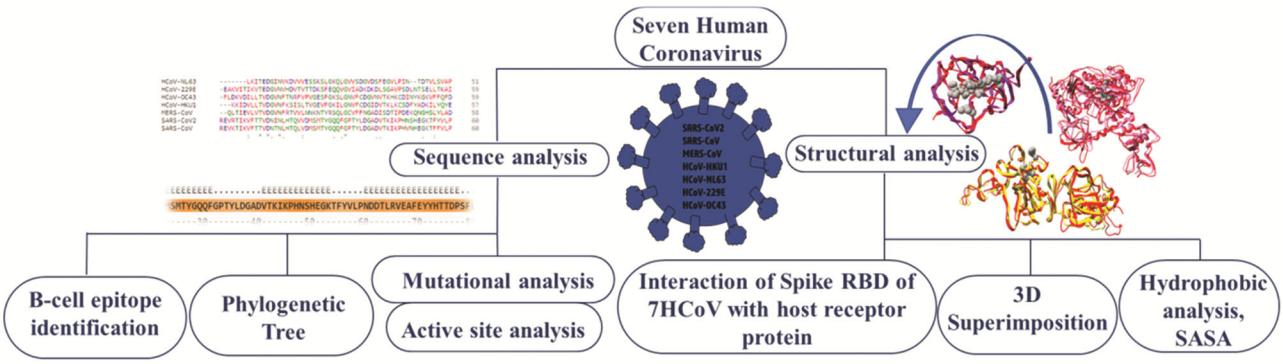

Fig. 1 — Schematic representation of the work flow for structural and functional analysis of seven types of Coronaviruses

deviation (RMSD) the cut-off is ≥ 5.0 Å. Further the superimposed protein structures were analyzed using UCSF chimera (https://www.cgl.ucsf.edu/chimera/) software[43].

Hydrophobic analysis and solvent accessibility surface area (SASA) calculations

The hydrophobicity defines the physicochemical unit i.e., a surface, any molecule, chemical groups, proteins etc. PyMoL software was used to analyze the hydrophobic and hydrophilic clusters in the surface of eight different SARS-CoV2 proteins where a python file was used to highlight the hydrophobic (red colour) and hydrophilic surface (white colour) of the protein. In addition, the protein PDB files were taken as templates of the selected proteins for the solvent accessibility surface area calculations. The web server GETREAD (http://curie.utmb.edu/getarea.html) was used for the SASA calculation, using a radius of 1.4 Å, default radii, and atomic solvent parameters.

Host receptor protein and pathogen spike RBD interaction

Interaction profile analysis

The host-pathogen interaction profile was analyzed to study the protein-protein interactions between the viral spike receptor binding domain and host receptor. The PDB complex of spike RBD of SARS-CoV2 with host ACE2, SARS-CoV-ACE2, HCoV-NL63-ACE2, MERS-CoV-DDP4 and HCoV-229E-APN were subjected to PDBsum (https://www.ebi.ac.uk/thornton-srv/databases/pdbsum)[44] online server to identify the bonds and non-bonded contacts between the protein-protein interaction complexes. In addition, binding affinity between the complexes were calculated using Prodigy server (https://wenmr.science.uu.nl/prodigy/)[45,46]. In case of HCoV-OC43 and HCoV-HKU1, host receptor is sialoglycan-based with 9-O-acetylated sialic acid. Spike RBD directly interacts with 9-O-acetylated sialic acid. The binding energy between them was calculated using AutoDock and prodigy server and 2D interaction analysis of the protein-ligand complex were analyzed in Discovery studio visualizer.

Hot-spot residues in viral spike RBD and host receptor protein

The hotspot residues identification between the viral spike RBD and host receptor protein was carried out using three computational servers *i.e*., KFC Server (https://mitchell-web.ornl.gov/KFC_Server/index.php)[47,48] SpotOn server (http://alcazar.science.uu.nl/cgi/services/SPOTON)[49,50] and Robetta Server (https://robetta.org)[51,52]. The common hotspots residues which are predicted by all the three servers are highlighted and the potential hotspots were also analyzed in the protein-protein interaction profile.

Per-residue energy contribution

The per-residue energy contribution was analyzed using online server pyDockEneRes (https://life.bsc.es/pid/pydockeneres)[53] in order to identify the residues contributing to the binding energy of the protein-protein complex. The calculation was done for different complexes that includes spike RBD of SARS-CoV2 with host ACE2, SARS-CoV-ACE2, HCoV-NL63-ACE2, MERS-CoV-DDP4 and HCoV-229E-APN.

Results and Discussion

Structural details of the selected proteins

Non-Structural Protein 3 (NSP3)

NSP3 is a multi-domain and the largest protein of the Coronavirus (CoVs) genome. NSP3 plays a pivotal role in the replication-transcription complex. The investigation of the three-dimensional structures of NSP3 domains by X-ray crystallography and/ or nuclear magnetic resonance (NMR) spectroscopy have shown the existence of eight domains in all CoVs- the Glu-rich acidic domain, a macrodomain, the ubiquitin-like domain 1 (Ubl1), ubiquitin-like domain 2 (Ubl2), the papain-like protease 2 (PL2pro), the NSP3 ectodomain, the domains Y1, and CoV-Y of unknown functions. The dominant active sites of the protein include L165, D167, R169, Y271, Q272, K160, Y267 and V246[54-56].

Non-Structural Protein 5 (NSP5)

NSP5 serves as the main protease (3Chemotrypsin like protease (3Cl-pro)) for proteolytic processing of the replicase polyproteins (pp1a and pp1ab). This protein is also referred to as 'main protease', or Mpro because of its active role in gene expression and replicase processing. The active amino acids of this protein are Y269, Q189, L4, N142, T26, C145, H41, T25, L27, V3, E240, H246, P108, P132, I200, Q107, N180, R105, and F185[57-59].

Non-Structural Protein 12 (NSP12)

NSP12 is an RNA dependent RNA polymerase which is involved in the transcription and replication process by catalyzing the synthesis of viral RNA of CoVs. The addition of the NSP7 and NSP8 co-factors greatly stimulates the polymerase activity of this protein. This protein has two metal-binding sites, one in the nidovirus-unique extension and the other in the

finger's domain. The active sites of the protein include A-558, T556, R553, S682, T680, and D760[60-62].

Non-Structural Protein 13 (NSP13)

NSP13 is a potent viral interferon antagonist and it interacts with the viral RNA-dependent RNA polymerase NSP12, which is known to stimulate the helicase activity because of mechano-regulation. Besides this role, NSP13 also has RNA 5′ triphosphatase activity within the same active site, which aids in the formation of the viral 5′ mRNA cap. NSP13 also contains the domains responsible for residues responsible for nucleotide binding and hydrolysis. Active sites of this protein are E375, Q404, D374, S289, K288, and R567[63,64].

Non-Structural Protein 14 (NSP14)

NSP14 involved in the exoribonuclease and methyl transferase activity. The C terminal region plays a role in the messenger RNA capping in the viral life cycle. The N terminal domain is important for proofreading exoribonuclease (ExoN) which is required for high-fidelity replication of the virus. The dual function of the virus, which is absent in other RNA viruses, helps in the genome expansion of the virus. The active sites of the virus are E92, H268, D273, D90, and E(Q)191[65,66].

Non-Structural Protein 15 (NSP15)

NSP15 is a uridine specific endoribonuclease and has a poorly defined role in replication. The endoribonuclease property of the protein is a characteristic of all CoVs. This evidence indicates its importance in the viral infection, yet a defined function of the protein is still unclear. The cleavage specificity of the protein has been shown experimentally, which demonstrates the cleavage of RNA substrates 3′ of uridines. The active sites of the protein include T341 (340), H250 (249), G248, H235, E340, L246, P355, S294, K290, G258, V292, G247, D240, Q245, P344, L346, and Y343[67,68].

Non-Structural Protein 16 (NSP16)

NSP16 is a 2′-O-methyltransferase which contributes mainly in immune evasion. NSP16 mimics the human protein CMTr1 and forms a complex with NSP10 to activate its enzymatic activity. Nsp16 enables the transfer of a methyl group from its S-adenosylmethionine (SAM) cofactor to the 2′ hydroxyl of ribose sugar of viral mRNA[18,19]. This methylation process is essential for translation efficiency and camouflages the mRNA thus recognized by intracellular pathogen recognition receptors such as IFIT and RIG-I. The active sites are N43, Y47, H69, D99, N101, D114, D130, K170[69-71].

Spike glycoprotein (S)

Spike glycoprotein (S) is aclove-shaped trimer which contains three segments- ectodomain, a single-pass transmembrane anchor, and a short intracellular tail. This protein plays an important role in binding host receptor and mediating membrane fusion and virus entry into host cells[67]. Spike mediates the virus into the host cell by binding to a receptor on the host cell surface through its S1 subunit. The S2 subunit of the protein then fuses viral and host membrane. The recognition of a variety of receptors aids in the viral attachment. The coronavirus spike. The active sites of the protein are Y453, Q496, Q498, N501, Y449, S494, Q493, G496, T500, Y505, F497, R403, Y495, L455, R403, Y453, G502, S494, Q406, Y505, K417, F497, Q409, Y453, D405, Q406, Y495, G496, E406, V860, D830, V951, Q949, Q957, and Y1007[72,73].

Sequence analysis of the seven types of human coronaviruses

Sequence similarity and evolutionary relationship

In this study, eight different proteins (amino acid sequences) have been considered (NSP3, NSP5, NSP12, NSP13, NSP14, NSP15, NSP16 and Spike) from 7 HCoVs. Evolutionarily, most of the genes are highly conserved among different organisms as shown in (Fig. S1A-H). No significant changes were observed among SARS-CoV and SARS-CoV2 sequences across different organisms. The MERS-CoV sequences have less genetic change with SARS-CoV and SARS-CoV2 sequences. Similarly, HCoV-HKU1 and HCoV-OC43 organisms has less genetic changes among eight sequences. The HCoV-NL63 and HCoV-229E sequences have high sequence similarity. However, these two organisms have less sequence identity with other organisms for the selected eight different sequences. The phylogenetic analysis of eight proteins with seven different organisms is depicted in (Fig. 2). The percentage of identity matrix across different eight gene sequences with 7 types of HCoVs has been shown in (Table 2).

Mutational analysis of SARS-CoV and SARS-CoV2 selected proteins

Mutational analysis of protein NSP3, NSP5, NSP12, NSP13, NSP14, NSP15, NSP16 and spike of SARS-CoV and SARS-CoV2 was carried out to

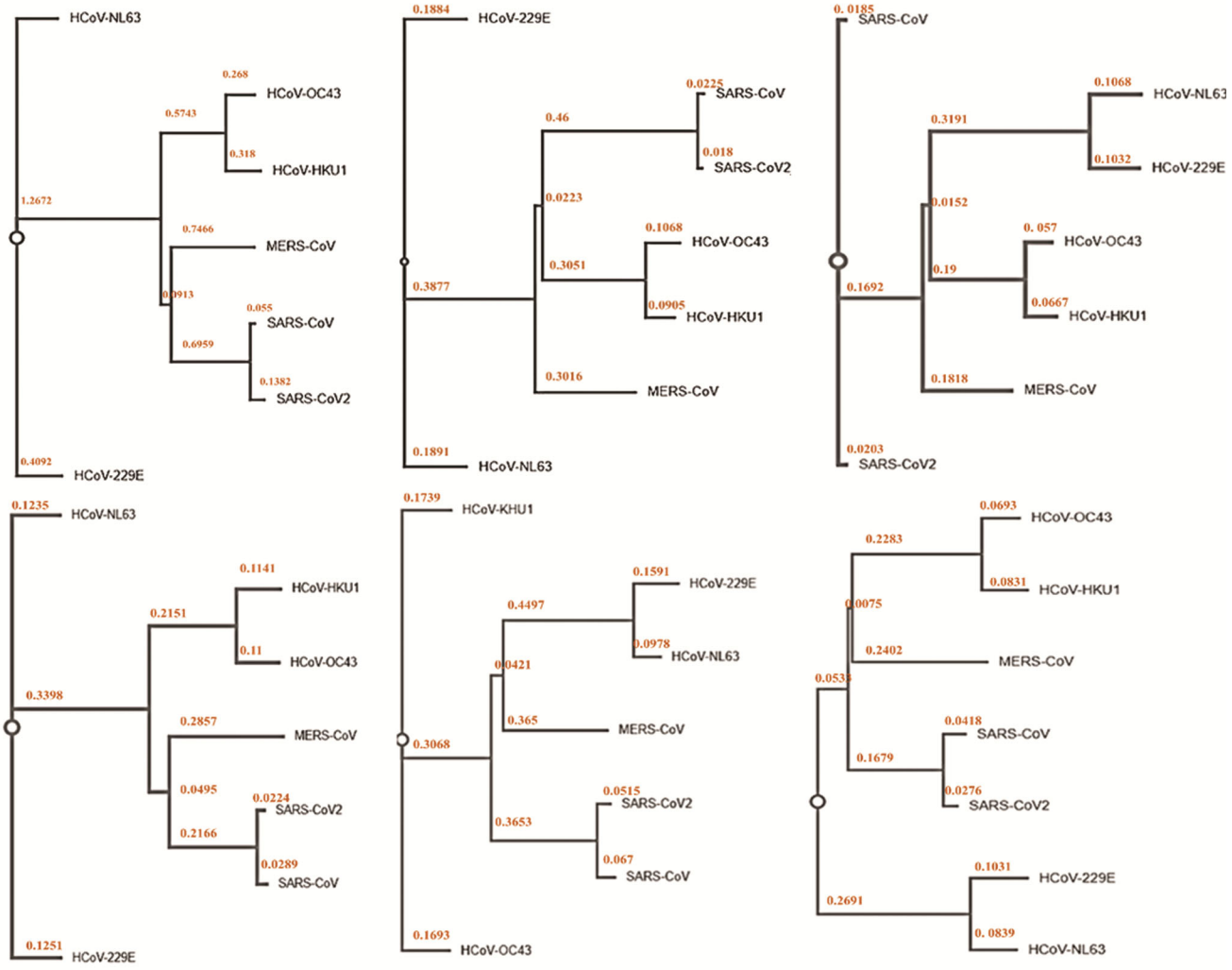

Fig. 2 — Phylogenetic analysis of (i) NSP3; (ii) NSP5; (iii) NSP12; (iv) NSP13; (v) NSP14; (vi) NSP15; (vii) NSP16; and (viii) Spike proteins of all seven types of coronaviruses with the distance matrix score generated using Virus Pathogen Resource (ViPR) server

identify the change in the amino acid sequence of SARS-CoV2 and identify the mutated residues and predicted its effect on mutation (neutral or deleterious). The residues were considered deleterious which were predicted by more than three mutational analysis tools as shown in (Table S3). In NSP3 protein, among the list of mutations, G100N, S115A, A130P, S156C, N263S, T275K, R285C, T309N mutations were predicted to be deleterious. For NSP5, H134F and T285A were predicted to be deleterious mutations by only one serve and in the case of NSP12, the mutation Y768F was predicted to be deleterious among the other mutations. Similarly, for NSP13, I57V mutation was predicted to be deleterious by only one server. In NSP14, D90E and E128P were predicted to be deleterious mutation and it was also observed that the residue D90E was found to be an active site residue which is in the conserved region of the aligned sequence as shown in (Fig. S1E). In the case of NSP15, the predicted deleterious mutations were L122F, S118P and S258F and A209C for NSP16. For spike protein, the mutations T95I, N501Y, Y505H, N764K, N856K, N969K were predicted to be deleterious out of which Y505H is an active site residue and in addition, among the mutations, G496S which is predicted to be neutral was also observed to be an active site residue. The presence of deleterious mutations in SARS-CoV2 strains shows how the amino acids have been replaced to enhance the viral replication and transmission, thus leading to severity of the disease caused by SARS-CoV2 as compared to SARS-CoV. Interestingly, the

Table 2 — The percentage identity matrix (in %) of eight proteins within the seven types of coronaviruses (*i.e.*, SARS-CoV2, SARS-CoV, MERS-CoV, HCoV-NL63, HCoV-229E, HCoV-OC43 and HCoV-HKU1)

Coronaviruses		Proteins							
		NSP3	NSP5	NSP12	NSP13	NSP14	NSP15	NSP16	SPIKE
SARS-CoV2	SARS-CoV	82.91	96.08	96.24	99.83	95.05	88.7	93.29	73.65
	MERS-CoV	30.57	50.83	71.61	71.4	63.41	51.47	66.11	30.76
	HCoV-OC43	28.48	48.18	66.56	68	58.88	48.08	66.11	31.55
	HCoV-HKU1	29.87	49.17	66.99	65.33	58.77	48.97	63.42	30.4
	HCoV-NL63	22.56	44.37	59.07	61.72	54.42	45.4	58.59	27.1
	HCoV-229E	19.64	41.39	58.75	60.07	53.59	44.21	57.58	28.54
SARS-CoV	SARS-CoV2	82.91	96.08	96.24	99.83	95.05	88.7	93.29	73.65
	MERS-CoV	31.53	51.82	72.15	71.4	63.41	51.47	66.11	32.75
	HCoV-OC43	29.8	48.51	66.67	68	58.88	48.08	66.11	31.82
	HCoV-HKU1	29.87	48.51	65.91	65.33	58.77	48.97	63.42	30.87
	HCoV-NL63	22.56	43.71	58.96	61.72	54.42	45.4	58.59	26.43
	HCoV-229E	22.55	40.73	58.75	60.07	53.59	44.21	57.58	29.33
MERS-CoV	SARS-CoV2	30.57	50.83	71.61	71.4	63.41	51.47	66.11	30.76
	SARS-CoV	31.53	51.82	72.15	71.4	63.41	51.47	66.11	32.75
	HCoV-OC43	29.57	53.47	68.18	68.51	57.53	48.68	62.54	33.72
	HCoV-HKU1	30.2	55.12	68.07	66.5	57.03	46.92	62.21	33.31
	HCoV-NL63	22.18	48.68	58.7	62.59	52.06	49.41	56.71	27.42
	HCoV-229E	22.99	49.01	57.95	61.91	51.84	47.06	54.36	29.13
HCoV-OC43	SARS-CoV2	28.48	48.18	66.56	68	58.88	48.08	66.11	31.55
	SARS-CoV	29.8	48.51	66.67	68	58.88	48.08	66.11	31.82
	MERS-CoV	29.57	53.47	68.18	68.51	57.53	48.68	62.54	33.72
	HCoV-HKU1	60.4	82.51	88.69	88.89	80.38	71.66	86.29	65.31
	HCoV-NL63	23.4	43.38	56.07	58.72	50.98	44.08	53.36	28.69
	HCoV-229E	22.71	44.04	56.18	58.49	52.34	42.6	54.7	29.31
HCoV-HKU1	SARS-CoV2	29.19	49.17	66.99	65.33	58.77	48.97	63.42	30.4
	SARS-CoV	29.87	48.51	65.91	65.33	58.77	48.97	63.42	30.87
	MERS-CoV	22.56	55.12	68.07	66.5	57.03	46.92	62.21	33.31
	HCoV-OC43	60.4	82.51	88.69	88.89	80.38	71.66	86.29	65.31
	HCoV-NL63	21.56	43.71	55.86	57.69	53.44	45.27	55.37	28.73
	HCoV-229E	22.96	45.36	55.64	56.81	52.63	44.08	55.37	29.58
HCoV-NL63	SARS-CoV2	22.18	44.37	59.07	61.72	54.42	45.4	58.59	27.1
	SARS-CoV	22.56	43.71	58.96	61.72	54.42	45.4	58.59	26.43
	MERS-CoV	22.18	48.68	58.7	62.59	52.06	49.41	56.71	27.42
	HCoV-OC43	23.4	43.38	56.07	58.72	50.98	44.08	53.36	28.69
	HCoV-HKU1	21.56	43.71	55.86	57.69	53.44	45.27	55.37	28.73
	HCoV-229E	50.37	70.53	81.77	87.09	78.63	78.78	83.67	63.04
HCoV-229E	SARS-CoV2	19.64	41.39	58.75	60.07	53.59	44.21	57.58	28.54
	SARS-CoV	22.55	40.73	58.75	60.07	53.59	44.21	57.58	29.33
	MERS-CoV	22.99	49.01	57.95	61.91	51.84	47.06	54.36	29.13
	HCoV-OC43	22.71	44.04	56.18	58.49	52.34	42.6	54.7	29.31
	HCoV-HKU1	22.96	45.36	55.64	56.81	52.63	44.08	55.37	29.58
	HCoV-NL63	50.37	70.53	81.77	87.09	78.63	78.78	83.67	63.04

identified deleterious residues CYS156 of SARS-CoV2 protein NSP3 was also found in the B-cell epitope region of SARS-CoV2and this can also serve as a potential target for inhibiting the viral replication. The superimposed structures of all the proteins of 7 HCoVs are depicted in (Fig. S2A-H) and the hydrophobicity map of all the proteins are depicted in (Fig. S3A-H). The corresponding Solvent Accessible Surface Area and protein structure alignment results are reported in (Tables S4 and S5), respectively.

B-cell epitopes prediction

For each protein involved in the 7 HCoVs, epitope identification was carried out where a total of 20 amino acids length epitopes were predicted based on their score calculated from specificity and sensitivity towards the B-cell. From the predicted linear epitope, it was observed that these epitopes are also present in the active site residues found in the conserved regions of the aligned protein sequence (Fig. S1 A-H**)**.The epitope residues for each protein which are found to

be in the conserved region in the aligned sequences of protein NSP3 are ASP165 (SARS-CoV2, SARS-CoV, MERS-CoV), ASP148 (HCoV-NL63), ASP156 (HCoV-229E), ASP164 (HCoV-OC43) and ASP160 (HCoV-HKU1). For NSP5 protein, the epitope at conserved regions is LEU27 (SARS-CoV2 SARS-CoV, MERS-CoV, HCoV-NL63, HCoV-229E, HCoV-OC43 andHCoV-HKU1) and in NSP12 the residues are SER682 (SARS-CoV2), SER684 (SARS-CoV), SER683 (MERS-CoV), SER677 (HCoV-NL63), SER676 (HCoV-229E), SER678 (HCoV-OC43) and SER678 (HCoV-HKU1). In the case of NSP13, the identified epitopes at conserved region are SER289 (SARS-CoV2 SARS-CoV), SER288 (MERS-CoV), SER290 (HCoV-NL63, HCoV-229E), SER288 (HCoV-OC43) and HCoV-HKU1). Similarly, the identified epitopes of NS14 are also present in the conserved regions which are GLU92 (SARS-CoV2, SARS-CoV, MERS-CoV, HCoV-229E), GLU88 (HCoV-NL63), GLU91 (HCoV-OC43, HCoV-HKU1). In NSP15, the identified epitopes at conserved regions are HIS249(SARS-CoV2, SARS-CoV), HIS246 (MERS-CoV, HCoV-NL63), HIS250 (HCoV-229E),HIS278 (HCoV-

OC43) and HIS277 (HCoV-HKU1) and in Spike the residues are GLN957 (SARS-CoV2), GLN931 (SARS-CoV), GLN1023 (MERS-CoV), GLN1015 (HCoV-NL63), GLN832 (HCoV-229E), GLN1038 (HCoV-OC43) and GLN1034 (HCoV-HKU1)as shown in (Table 3). These residues are also among the active site residues of SARS-CoV2 which are highlighted to be in the conserved regions. The residues present in the conserved regions are considered to be fundamentally important for understanding the biology and pathogenicity of virus and its replication[74]. The fact that the identified epitopes being present in the conserved region show its importance in terms of their specificity and sensitivity towards being recognized by antibody[75,76]. This study indicates that the predicted epitope residues play an important role in the viral replication and thus inhibitors can be developed based on the linear B-cell epitopes.

Spike RBD interaction with specific host receptor protein

Interaction profile analysis

The interaction between the viral spike receptor binding domain and host receptor of SARS-CoV2 with ACE2, SARS-CoV with ACE2, MERS-CoV with DPP4, HCoV-NL63 with ACE2, HCoV-229E with APN, HCoV-OC43 and HCoV-HKU1 with 9-O-acetylated sialic was analyzed to identify the interaction profile which are involved in viral genome entry into the host. If we compare all the binding energy of spike RBD with its corresponding receptor, the binding energy of SARS-CoV-2 spike RBD-ACE2 (-11.9 kcal/mol) is more as compared to SARS-CoV (-10.9 kcal/mol) and MERS (-9.9 kcal/mol), however NL63 CoV belong alpha coronavirus, spike RBD-ACE2 has shown more binding energy (-11.4 kcal/mol) than SARS-CoV and MERS-CoV as shown in (Table 4). The residues involved in SARS-CoV2 and ACE2 are highlighted, among which 17 residues of SARS-CoV2 interacts with 20 residues of ACE2 sharing 10 hydrogen bonds and 101 non-contact bonds as shown in (Table 4 and Fig. S4). In SARS-CoV and ACE2, the number of interacting residues is 16 and 20, respectively with 12 H-bond and 84 non bonded contacts shared between the complex (Table 4 and Fig. S4). In the case of HCoV-NL63 and ACE2, 13 residues and 15 residues are interacting respectively with 4 H-bonds and 57 non-bonded contacts (Table 4 and Fig. S4). The number of residues interacting between MERS-CoV and DDP4 are 17 and 19 respectively with 5 H-bonds (Table 4 and Fig. S4). Similarly, in HCoV-229E and APN, there are 13 and 11 residues respectively which are interacting in the protein-protein interface with 7 H-bonds (Table 4 and Fig. S4). In the case of HCoV-HKU1 and HCoV-OC43 interacting with 9-O-acetylsialic acid it was observed that 12-13 numbers of residues are shared between the host and spike with 6 and 4 H-bonds, respectively.

Hot-spot residues in viral spike RBD and host receptor protein

The hotspot residues present at the interface of the protein-protein interaction plays an important role in understanding the binding mechanism of the host-pathogen interaction[77,78]. Three computational methods (KFC, SpotOn and Robetta server) were used for the identification of hotspot residues in the protein-protein interacting interfaces. The predicted hotspots from KFC, SpotOn and the hotspot residues with $\Delta\Delta G$ value >1 kcal/mol or 1kcal/mol from Robetta are highlighted in red circles in (Fig. 3). From the study carried out, it can be observed that the predicted hotspot residues are identified among the interacting residues of the spike RBD and host receptor protein. The common predicted hotspots

Table 3 —The linear B-Cell epitopes predicted for NSP3, NSP5, NSP12, NSP13, NSP14, NSP15, NSP16 and spike protein of SARS-CoV2, SARS-CoV, MERS-CoV, HCoV-NL63, HCoV-229E, HCoV-OC43 and HCoV-HKU1 coronaviruses. The predicted epitopes are based on the BCPred score and the epitope percentage is given by the average generated from BepiPred server

	Protein	Position	Epitope	BCPred Score	Epitope percentage (Average)
	NSP3	151	LILAYCNKTVGELGDVRETM	0.89	0.470
SARS-CoV2	NSP5	10	SGKVEGCMVQVTCGTTTLNG	0.90	0.416
	NSP12	674	YVKPGGTSSGDATTAYANSV	0.99	0.492
	NSP13	270	QKVGMQKYSTLQGPPGTGKS	1.00	0.438
	NSP14	78	EAIRHVRAWIGFDVEGCHAT	0.84	0.476
	NSP15	246	GGLHLLIGLAKRFKESPFEL	0.65	0.446
	NSP16	65	MRVIHFGAGSDKGVAPGTAV	0.99	0.448
	Spike	951	VVNQNAQALNTLVKQLSSNF	0.76	0.443
	NSP3	159	TVGELGDVRETMTHLLQHAN	0.73	0.498
SARS-CoV	NSP5	10	SGKVEGCMVQVTCGTTTLNG	0.90	0.415
	NSP12	676	YVKPGGTSSGDATTAYANSV	0.99	0.474
	NSP13	270	QKVGMQKYSTLQGPPGTGKS	1.00	0.438
	NSP14	80	IRHVRAWIGFDVEGCH	0.86	0.474
	NSP15	247	GLHLMIGLAKRSQDSPLKLE	0.78	0.468
	NSP16	65	MRVIHFGAGSDKGVAPGTAV	0.99	0.448
	Spike	919	SLTTTSTALGKLQDVVNQNA	0.73	0.399
	NSP3	162	APDDASRLLHTVLAKAELCC	0.83	0.474
MERS-CoV2	NSP5	20	VTCGSMTLNGLWLDNTVWCP	0.99	0.383
	NSP12	670	CGGGYYVKPGGTSSGDATTA	0.99	0.489
	NSP13	274	YSKYVTVQGPPGTGKSHFAI	0.99	0.402
	NSP14	91	VEGAHASRNACGTNVPLQLG	0.86	0.518
	NSP15	241	TLGGLHLLIGLYKKQQEGHI	0.74	0.455
	NSP16	67	VIHFGAGSDKGIAPGTSVLR	0.99	0.467
	Spike	1012	FTTTNEAFQKVQDAVNNNAQ	0.75	0.432
	NSP3	142	SKGQKGDAEEALSKLSEYLI	0.67	0.504
HCoV-NL63	NSP5	19	RVCYGSTVLNGVWLGD	0.86	0.391
	NSP12	663	YSNGGFYFKPGGTTSGDAST	0.99	0.484
	NSP13	271	QLIGKQKITTIQGPPGSGKS	1.00	0.428
	NSP14	87	VESAHVCGDNIGTNVPLQVG	0.83	0.500
	NSP15	245	LHLLISQFRLSKMGVLKADD	0.79	0.448
	NSP16	68	HYGAGSDKGVAPGTTVLKRW	0.99	0.488
	Spike	1015	QDVVNQQGSALNHLTSQLRH	0.73	0.447
	NSP3	153	DKGDAEDTLNKLSYLANEA	0.69	0.489
HCoV-229E	NSP5	17	VVRVCYGNTVLNGLWL	0.80	0.385
	NSP12	669	YFKPGGTTSGDATTAYANSV	0.99	0.490
	NSP13	271	QLIGKQRITTIQGPPGSGKS	1.00	0.404
	NSP14	90	DVEGAHVTGDNVGTNVPLQV	0.95	0.516
	NSP15	238	YGDVSKTTLGGLHLLISQVR	0.69	0.372
	NSP16	64	MRVLHLGAGSDYGVAPGTAV	0.66	0.449
	Spike	831	IQDVVNQQGNSLNHLTSQLR	0.78	0.457
	NSP3	161	DPADSRDFLRVVFSQVDLTG	0.85	0.484
HCoV-OC43	NSP5	19	SVTYGNMTLNGLWLDD	0.82	0.401
	NSP12	665	CGGCYYVKPGGTSSGDATTA	0.99	0.505
	NSP13	269	QHIGMKRYCTVQGPPGTGKS	0.99	0.425
	NSP14	91	EGAHATRDSIGTNFPLQLGF	0.80	0.491
	NSP15	269	FNQKIIGGLHLLIGLYRRQQ	0.79	0.398
	NSP16	69	HLGAGSEKGVAPGSAVLRQW	0.85	0.399
	Spike	1025	QGFDATNSALVKIQAVVNAN	0.65	0.398
	NSP3	149	LAKGHFKFDPSDATDFIRV	0.90	0.490
HCoV-HKU1	NSP5	18	SVTYGSMTLNGLWLDD	0.87	0.396
	NSP12	665	CGGCYYVKPGGTSSGDATTA	0.99	0.495
	NSP13	269	QHIGMKRYCTVQGPPGTGKS	0.99	0.421
	NSP14	78	AIKRVRGWVGFDVEGAHATR	0.85	0.491
	NSP15	272	VIGGLHLLIGLFRRKKKSNL	0.66	0.397
	NSP16	63	VNMRVLHLGAGSDKEVAPGS	0.65	0.425
	Spike	1029	ALAKIQSVVNSNAQALNSLL	0.77	0.4257

The presence of active sites in the epitope region is shown in boldface

Table 4 — The interacting interface statistics of viral spike receptor binding domain and host proteins of SARS-CoV2 with ACE2, SARS-CoV with ACE2, MERS-CoV with DPP4, HCoV-NL63 with ACE2 and HCoV-229E with APN, HCoV-OC43 and HKU1 with 9-O-acetylsialic acid

HCoV	PDB ID	Host receptor	Binding energy (kcal/mol)	Protein-protein/ligand interaction analysis				
				Interacting residues	Interface area (A)	Salt bridges	Hydrogen Bonds	Non-bonded contacts
SARS-CoV-2	6M0J	ACE2	−11.9	SAR-CoV-2: 17 ACE2: 20	825 863	1	10	101
SARS-CoV	3SCI	ACE2	−10.9	SARS-CoV: 16 ACE2: 20	894 940	1	12	84
MERS-CoV	4L72	DPP4	−9.9	MERS: 17 DPP4: 19	292 1028	2	5	78
NL63	3KBH	ACE2	−11.4	NL63: 13 ACE2: 15	627 707	-	4	57
229E	6ATK	APN	−9.0	229E: 13 APN: 11	512 482	1	7	66
OC43	6NZK	9-O-acetylsialic acid	−7.2	−12	-	-	4	9
HKU1	5I08	9-O-acetylsialic acid	−7.0	−13	-	-	6	11

*ACE2 (Angiotensin-converting enzyme 2); APN (Aminopeptidase N); DPP4 (Dipeptidyl peptidase 4)

from the three severs are highlighted in boldface in (Tables S6-S10). It can be observed that the hotspot predicted by all the three servers in SARS-CoV2 (Chain E) and ACE2 (Chain A) complex are GLN498E, THR500E and LYS353Aand in SARS-CoV (Chain E) and ACE2 (Chain A) complex the common residues are TYR475E, THR486E and THR27A. For MERS-CoV (Chain B) with DPP4 (Chain A) complex only ARG336A residue was predicted by all the three servers and no common residue was found in HCoV-NL63 (Chain E) and ACE2 (Chain A) complex. Lastly, in HCoV-229E (Chain E) and APN (Chain B) complex CYS317E and ASN319E were among the common hotspot predicted by all the three servers.

From the study it can be observed that most of the common hotspot amino acids are PHE, TYR, TRP, ARG, ASN and VAL. Studies have reported some of them to be potential hotspot residues involved in the spike RBD and host receptor interaction during the viral entry[79,80]. There are various bonded and non-bonded contacts between the protein-protein interaction and hydrogen bond is one of the most important contacts in predicting and profiling the protein-protein interface[81,82]. Among the hotspots, the residues forming H-bond between SARS-CoV2 (Chain E) and ACE2 (Chain A) complex are ASN487E, THR500E, TYR449E, GLN24A, TYR83A, LYS353A, TYR41A and GLN42A and active site residues present on spike are TYR505, ASN501, THR500, GLN498 and TYR449.Similarly,in SARS-CoV (Chain E) and ACE2 (Chain A) complex the hotspot residues areTHR486E, ASN473E, TYR484E, ARG426E, LYS353A, TRY41A, GLN42A and GLU329A.In the case of MERS-CoV (Chain B) with DPP4 (Chain A) complex, GLU513B, ARG336A, LEU294A, LYS267A and ALA291A residues are found to be involved in the host-spike hydrogen bond formation. In the interaction between HCoV-NL63 (Chain E) and ACE2 (Chain A) complex, the residues involved in hydrogen bond are found to be SER540E, TYR498E, THR324A and LYS353A and in HCoV-229E (Chain E) and APN (Chain B) complex, ASN319E, PHE318E and ASP288B were observed to be involved in the hydrogen bond formation in the protein-protein interaction. There are reports on previously identified hotspots for SARS-CoV2 spike protein such as L455, F456, F486, Y489, N501, Q49, Y505, K417, Y449 and Y473[83-86] which are important for viral RBD interaction with host receptor. In this regard, our study also shows the presence of these residues as hotspots predicted by two or three of the servers. It is also suggested that conserved polar residues consist of hotspots[84] and some of the polar residues such as GLN498E, THR500, ASN501E (SARS-CoV2), THR486E (SARS-CoV), THR27A (ACE2), ASN501B, SER557B (MERS-CoV), GLN286A, THR288A (DDP4), THR324A (ACE2), CYS317E and ASN319E (HCoV-229E) are identified as potential hotspots. The residues in the conserved regions could have critical role in the binding of the spike RBD with host protein. The predicted hotspots

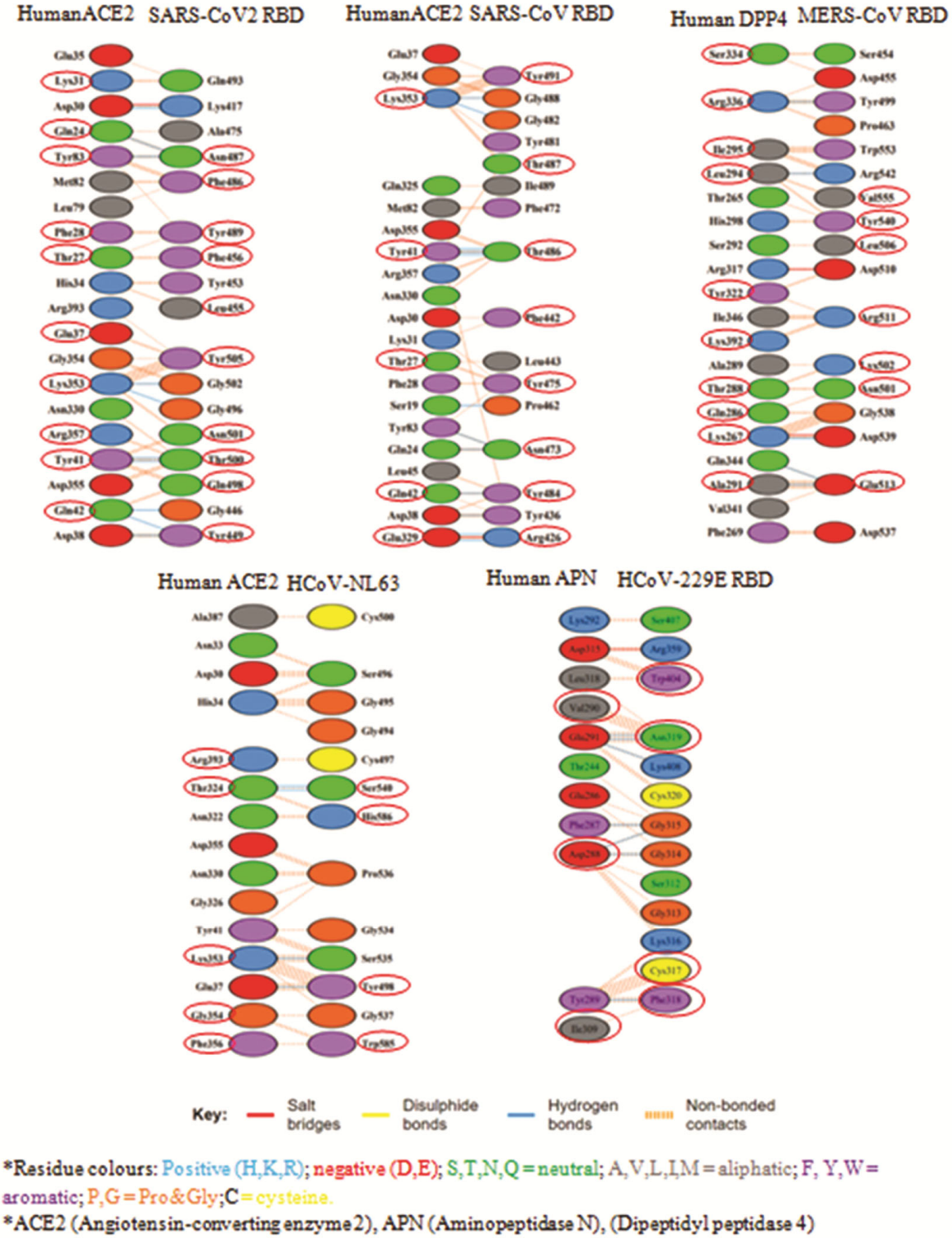

Fig. 3 — The protein-protein interactions between the viral spike receptor binding domain and host proteins. The interactions are represented for SARS-CoV2 with ACE2, SARS-CoV with ACE2, MERS-CoV with DPP4, HCoV-NL63 with ACE2, HCoV-229E with APN which are involved in viral genome entry into the host. The diagrammatic representation indicates the interaction residues between the proteins

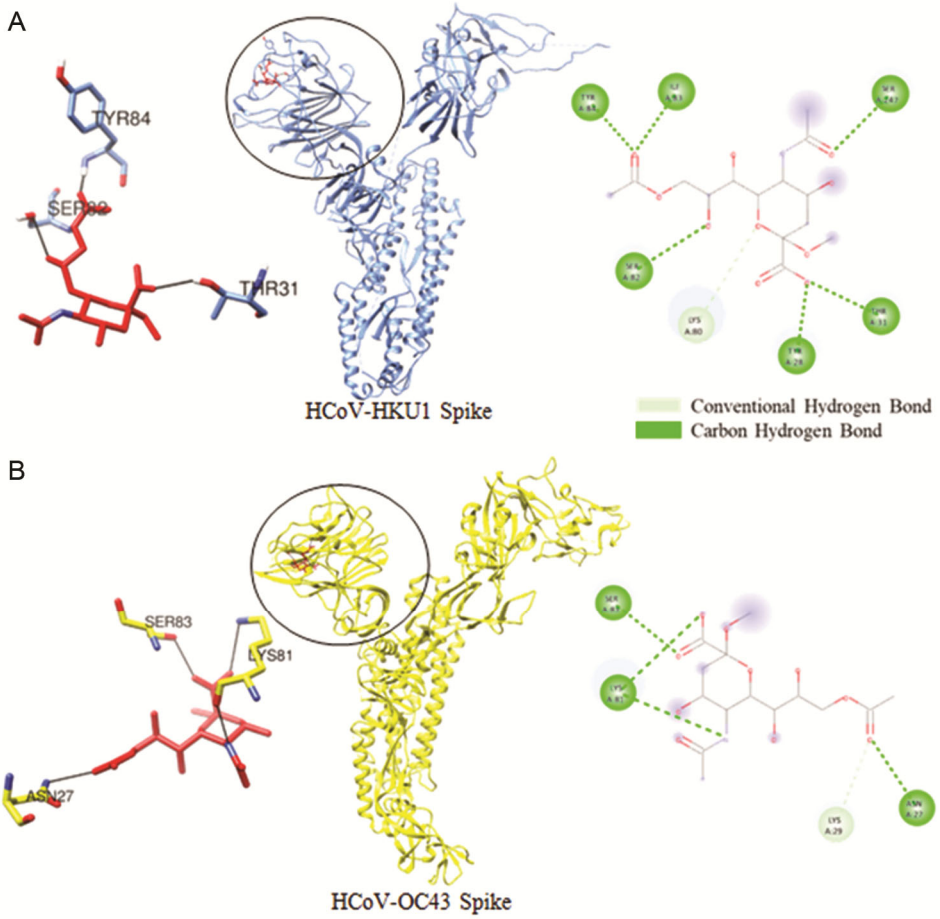

Fig. 4 — The interaction of 9-O-acetylated sialic acid with spike (RBD) protein of (i) HCoV-HKU1, (ii) HCoV-OC43. The 3D representation shows theHCoV-HKU1 and HCoV-OC43 spike protein binding pockets where the ligand binds with a binding energy of −7.0 kcal/mol and −7.2 kcal/mol respectively and residues forming hydrogen bonds. In addition, the 2D interaction analysis shows the residues forming conventional and carbon hydrogen bond analyzed on Discovery studio

and the presence of hotspots in conserved regions and also forming hydrogen bond in the interacting interface of the protein-protein interaction indicates the significant impact of these hotspot residues on the binding of the spike RBD with human receptor.

Additionally, we analyzed the HCoV-HKU1 and HCoV-OC43 interaction with 9-o-acetylated sialic acid. It was observed that HCoV-HKU1 and OC43 binds to 9-o-acetylated sialic acid with binding energy of −7.0 and −7.3 kcal/mol respectively predicted by prodigy server. For HKU1 the 2D interaction analysis shows that the receptor binds at the binding site SER247, TYR28, THR31, ILE83, TYR84 and SER82 forming conventional hydrogen and carbon hydrogen bond as shown in (Fig. 4). We have also carried outdocking using Autodock Vina where the binding score was found to be −5.6 kcal/mol and −6.0 kcal/mol respectively. For OC43 the 2D interaction shows residues SER83, LYS81, ASN27 and LYS29 forming conventional hydrogen and carbon hydrogen bond between the spike RBD and host 9-o-acetylated sialic acid as shown in (Fig. 4). Among the two complexes (HCoV-HKU1 and HCoV-OC43), the residue SER is found to be the common residue which in involved in the interaction of spike RBD and host receptor binding. Here, SER being a polar residue and with suggestions of conserved polar residues constituting of hotspots[87], this indicates the importance of SER being a potential hotspot that can be targeted for inhibiting the interaction between these complexes. The prediction of hotspot and interacting residues can play an important role in identifying the interface area of the host-pathogen interaction and small molecules can be designed to bind to the hotspot residues leading to blockage of protein-protein complex thus inhibiting the viral entry into the host.

Per-residue energy contribution

Per-residue energy contribution was calculated in order to identify the residues that are involved in contributing the maximum energy in the protein-protein interaction. In SARS-CoV2 (Chain E) and ACE2, residues PHE456E (−6.964) and THR27A (−7.772) have maximum energy contributing to the interaction. In SARS-CoV and ACE2 complex, TYR484E (−9.070) and MET82A (−6.953) have contributed to the maximum energy. In HCoV-NL63 and ACE2 complex, TRP585E (−8.410) PHE356A (−5.902) contributed the maximum energy. Residues TYR540B (−8.418) and LEU294A (−5.791) has the maximum energy contributing to the host pathogen interaction in case of MERS-CoV and DDP4. Similarly, HCoV-229E and APN complex also showed residues TRP404 (−8.067) ASP315 (−8.222) contributing maximum energy in the interaction.

The per-energy residue calculated for protein-protein interaction complex gives the binding energy contribution of each residue involved in the PPI and this can also be used for identifying the hotspots[51]. Interestingly from the study carried out, the residues contributing the maximum per-residue energy are also among the predicted hotspots residues (Tables S6-S11). In SARS-CoV2 (Chain E) and ACE2 complex, PHE456E and THR27A which have non-bonded contacts are identified as hotspots by two servers, in SARS-CoV and ACE2 complex, TYR484E is identified as a hotspot residue while MET82A was identified by only one server as a hotspot which have non-bonded contacts and hydrogen bonds. Similarly, for HCoV-NL63 and ACE2 complex, residue TRP585E and PHE356A with non-bonded contacts are among the predicted hotspot. In the case of MERS-CoV and DDP4 complex, the residue TYR540B and LEU294A with bonded contacts and hydrogen bonds are predicted to be potential hotspot by two servers and also for HCoV-229E and APN complex, residues TRP404Ewas predicted as hotspot by two servers while ASP315B by only one server which has non-bonded contact and salt bridge interaction. From the study, it can be observed that the residues which are involved in contributing the maximum energy in PPI interaction are also amongst the predicted hotspot residues. These residues bind to one another through bonded, non-bonded contacts and hydrogen bonds which gives an overview of the types of interaction present in the protein-protein interface. This indicates that residues which are involved in contributing the maximum binding energy are also potential hotspots which can be a potential target for interrupting the spike RBD and host receptor interaction during the viral entry.

Conclusion

Owing to the ongoing pandemic, the urgency to understand and study the human coronaviruses and its related infection to develop effective therapeutics against SARS-CoV2, can hardly be overstated. The detailed analysis of the eight selected protein among 7 HCoVs protein sequences and structures revealed some similarity and differences. However, we could observe the differences among SARS-CoV and SARS-CoV2 interaction pattern with Spike protein.

These differences may be the reason for the increased infection and rate of replications. The close relations among SARS-CoV2, SARS-CoV and MERS-CoV clearly indicates their evolutionary relationship and their virulence when compared to other types of human coronaviruses.

Among the seven types of coronaviruses, SARS-CoV2 was observed to have undergone major mutational changes which may be one of the reasons for its higher rate of infection and transmission causing a pandemic. Among the predicted hotspots residues few of them were also found to be in the conserved regions in the aligned sequences as well as in the active site regions. This indicates the importance of the identified hotspot residues based on which small inhibitors can be developed to inhibit the viral attachment and its replication. The results obtained from the study can be utilized for exploring the viral genome evolution and fundamental mechanism of virus-host interaction and also get an insight into how each proteins functions and plays important role in the viral replication. Moreover, the detailed study on these seven types of human coronavirus will provide the basis for understanding their virological profiles which can be utilized for designing inhibitors and developing therapeutics for diseases caused by the coronaviruses.

Acknowledgement

GNS thanks J.C Bose fellowship of DST New Delhi. EJ thanks DST Inspire Fellowship, KK thanks UGC JRF Fellowship and HS thanks for an Institutional Postdoctoral Fellowship. We thank DBT center of excellence project (BT/PR40188/BTIS/137/27/2021) for the financial assistance.

Conflict of interest

All authors declare no conflict of interest.

References

1 Siegel RD, Classification of human viruses. Principles and Practice of Pediatric Infectious Diseases, (2018) 1044.

2 Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJ, Wolthers KC, Wertheim-van PD, Kaandorp J, Spaargaren J, Berkhout B, Identification of a new human coronavirus. *Nat Med*, 10 (2004) 373.

3 Vabret A, Dina J, Gouarin S, Petitjean J, Corbet S & Freymuth F, Detection of the new human coronavirus HKU1: A report of 6 cases. *Clin Infect Dis*, 42 (2006) 639.

4 Ramadan N & Shaib H, Middle East respiratory syndrome coronavirus (MERS-CoV): A review. *Germs*, 9 (2019) 35.

5 Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X & Cheng Z, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*, 395 (2020) 506.

6 Zhu Z, Lian X, Su X, Wu W, Marraro GA & Zeng Y, From SARS and MERS to COVID-19: a brief summary and comparison of severe acute respiratory infections caused by three highly pathogenic human coronaviruses. *Respir Res*, 21 (2020) 4.

7 Hu B, Guo H, Zhou P & Shi ZL, Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol*, 19 (2021) 154.

8 Liu DX, Liang JQ & Fung TS, Human Coronavirus-229E, -OC43, -NL63, and -HKU1 (Coronaviridae). (Encyclopedia of Virology, 4th Edition, Volume 2) 2021.

9 Liu J, Zheng X, Tong Q, Li W, Wang B, Sutter K, Trilling M, Lu M, Dittmer U & Yang D, Overlapping and discrete aspects of the pathology and pathogenesis of the emerging human pathogenic coronaviruses SARS-CoV, MERS-CoV, and 2019-nCoV. *J Med Virol*, 92 (2020) 494.

10 WHO G, Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). (World Health Organization) 2020.

11 Lai CC, Shih TP, Ko WC, Tang HJ & Hsueh PR, Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int J Antimicrob Agents*, 55 (2020) 105924.

12 COVID-19: Infodemic spreads faster than pandemic. (World Health Organization) 2020.

13 Edridge AW, Kaczorowska J, Hoste AC, Bakker M, Klein M, Loens K, Jebbink MF, Matser A, Kinsella CM, Rueda P & Ieven M, Seasonal coronavirus protective immunity is short-lasting. *Nat Med*, 26 (2020) 1693.

14 Channappanavar R & Perlman S, Age-related susceptibility to coronavirus infections: role of impaired and dysregulated host immunity. *J Clin Invest*, 130 (2020) 6213.

15 El Zowalaty ME & Järhult JD, From SARS to COVID-19: A previously unknown SARS-related coronavirus (SARS-CoV-2) of pandemic potential infecting humans–Call for a One Health approach. *One Health*, 9 (2020) 100124.

16 Chakraborty S, Evolutionary and structural analysis elucidates mutations on SARS-CoV2 spike protein with altered human ACE2 binding affinity. *Biochem Biophys Res*, 538 (2021) 103.

17 Chen Z, Boon SS, Wang MH, Chan RW & Chan PK, Genomic and evolutionary comparison between SARS-CoV-2 and other human coronaviruses. *J Virol Methods*, 289 (2021) 114032.

18 Bartlam M, Yang H & Rao Z, Structural insights into SARS coronavirus proteins. *Curr Opin Struct Biol*, 15 (2005) 672.

19 Li, F. Structure, function, and evolution of coronavirus Spike proteins. *Annu Rev Virol*, 3 (2016) 261.

20 Mishra AK, Gupta V & Tewari SP, *In silico* screening of some naturally occurring bioactive compounds predicts potential inhibitors against SARS-COV-2 (COVID-19) protease. *Indian J Biochem Biophys*, 58 (2021) 416.

21 Reddy CS, Vijayasarathy K, Srinivas E, Sastry GM & Sastry GN, Homology modeling of membrane proteins: a critical assessment. *Comput Biol Chem*, 30 (2006) 126.

22 Madugula SS, Nagamani S, Jamir E, Priyadarsinee L & Sastry GN, Drug repositioning for anti-tuberculosis drugs: an *in silico* polypharmacology approach. *Mol Divers*, (2021) 21.

23 Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, Brezovsky J & Damborsky J, Predict SNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput Biol*, 10 (2014) e1003440.

24 Jespersen MC, Mahajan S, Peters B, Nielsen M & Marcatili P, Antibody specific B-cell epitope predictions: leveraging information from antibody-antigen protein complexes. *Front Immunol*, 10 (2019)298.

25 Potocnakova L, Bhide M & Pulzova LB, An introduction to B-cell epitope mapping and *in silico* epitope prediction. *J Immunol Res*, (2016).

26 Srivastava HK & Sastry GN, Molecular dynamics investigation on a series of HIV protease inhibitors: assessing the performance of MM-PBSA and MM-GBSA approaches. *J Chem Info Model*, 52 (2012) 3098.

27 Badrinarayan P & Sastry GN, Virtual high throughput screening in new lead identification. *Comb Chem High Throughput Screen*, 14 (2011) 860.

28 Choudhury C, Priyakumar UD & Sastry GN, Molecular dynamics investigation of the active site dynamics of mycobacterial cyclopropane synthase during various stages of the cyclopropanation process. *J Struct Biol*, 187(2014)48.

29 Mahadevi AS & Sastry GN, Cooperativity in noncovalent interactions. *Chem Rev*, 116 (2016) 2825.

30 NCBI Resource Coordinators. Database resources of the national center for biotechnology information. *Nucleic Acids Res*, 43 (2015) 17.

31 Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN & Bourne PE, The Protein Data Bank. *Nucleic Acids Res*, 28 (2000) 242.

32 Guex N & Peitsch MC, SWISS-MODEL: an automated protein SWISSMODEL: an automated protein. *Nucleic Acids Res*, 31 (2003) 3385.

33 Pontius J, Richelle J & Wodak SJ, Quality assessment of protein 3D structures using standard atomic volumes. *J Mol Biol*, 264 (1996) 136.

34 Rodriguez-Tomé P, Stoehr PJ, Cameron GN & Flores TP, The european bioinformatics institute (EBI) databases. *Nucleic Acids Res*, 24 (1996) 12.

35 Sievers F & Higgins DG, Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci*, 27 (2018) 145.

36 Badrinarayan P & Sastry GN, Sequence, structure, and active site analyses of p38 MAP kinase: exploiting DFG-out conformation as a strategy to design new type II leads. *J Chem Inf Model*, 51 (2011) 129.

37 Badrinarayan P & Sastry GN, Specificity rendering 'hot-spots' for aurora kinase inhibitor design: the role of non-covalent interactions and conformational transitions. *PLoS One*, 9 (2014) e113773.

38 Pickett BE, Sadat EL, Zhang Y, Noronha JM, Squires RB, Hunt V, Liu M, Kumar S, Zaremba S, Gu Z & Zhou L, ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res*, 40 (2012) D5938.

39 Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, Brezovsky J & Damborsky J, Predict SNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput. Biol*, 10 (2014) e1003440.

40 Saha S & Raghava GP, Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins*, (2006) 48.

41 Jespersen MC, Peters B, Nielsen M & Marcatili P, BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res*, 45(2017) W29.

42 Zhang Y & Skolnick JT, TM-align: a Protein Structure Alignment Algorithm Based on the TM-Score. *Nucleic Acids Res*, 33 (2005) 2309.

43 Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC & Ferrin TE, UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*, 25 (2004)1605.

44 Laskowski RA, PDBsum: summaries and analyses of PDB structures. *J Nucleic acids Res*, 29 (2001) 222

45 Vangone A & Bonvin AM, Contacts-based prediction of binding affinity in protein-protein complexes. *Elife*, 4 (2015) e07454.

46 Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM & Vangone A, PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics*, 32 (2016) 3678.

47 Darnell SJ, LeGault L & Mitchell JC, KFC Server: interactive forecasting of protein interaction hot spots. *Nucleic Acids Res*, 36(2008) W265.

48 Zhu X & Mitchell JC, KFC2: a knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features. Proteins: *Struct Funct Genet*, 79 (2011) 2671.

49 Melo R, Fieldhouse R, Melo A, Correia JD, Cordeiro MN, Gümüş ZH, Costa J, Bonvin AM & Moreira IS, A machine learning approach for hot-spot detection at protein-protein interfaces. *Int J Mol Sci*, 17 (2016) 1215.

50 Moreira IS, Koukos PI, Melo R, Almeida JG, Preto AJ, Schaarschmidt J, Trellet M, Gümüş ZH, Costa J & Bonvin AM, SpotOn: high accuracy identification of protein-protein interface hot-spots. *Sci Rep*, 7 (2017) 11.

51 Song Y, DiMaio F, Wang RY, Kim D, Miles C, Brunette TJ, Thompson J & Baker D, High-resolution comparative modeling with Rosetta CM. *Structure*, 21 (2013) 1742.

52 Raman S, Vernon R, Thompson J, Tyka M, Sadreyev R, Pei J, Kim D, Kellogg E, DiMaio F, Lange O & Kinch L, Structure prediction for CASP8 with all-atom refinement using Rosetta. *Protein*s, 77 (2009) 99.

53 Romero-Durana M, Jiménez-García B & Fernández-Recio J, pyDockEneRes: per-residue decomposition of protein–protein docking energy. *Bioinformatics*, 36 (2020) 2285.

54 Lei J, Kusov Y & Hilgenfeld R, Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral Res,* 149 (2018) 74.

55 Erdogan T, Computational evaluation of 2-arylbenzofurans for their potential use against SARS-CoV-2: A DFT, molecular docking, molecular dynamics simulation study. *Indian J Biochem Biophys*, 58 (2021) 12.

56 Báez-Santos YM, Mielech AM, Deng X, Baker S & Mesecar AD, Catalytic function and substrate specificity of the papain-like protease domain of nsp3 from the Middle East respiratory syndrome coronavirus. *J Virol*, 88 (22014) 12527.

57 Roe MK, Junod NA, Young AR, Beachboard DC & Stobart CC, Targeting novel structural and functional features of coronavirus protease nsp5 (3CLpro, Mpro) in the age of COVID-19. *J Gen Virol*, 102 (2021) 001558.

58 Esposito S, Tagliabue C, Bosis S & Principi N, Levofloxacin for the treatment of Mycoplasma pneumoniae-associated meningoencephalitis in childhood. *Int J Antimicrob Agents*, 37 (2011) 475.

59 Pal S & Talukdar A, Compilation of potential protein targets for SARS-CoV-2: Preparation of homology model and active site determination for future rational antiviral design. *Chem Rxiv*, (2020).

60 Kirchdoerfer RN & Ward AB, Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun*, 10 (2019) 2342

61 El Hassab MA, Shoun AA, Al-Rashood ST, Al-Warhi T & Eldehna WM, Identification of a new potential SARS-COV-2 RNA-dependent RNA polymerase inhibitor *via* combining fragment-based drug design, docking, molecular dynamics, and MM-PBSA calculations. *Front Chem*, 8 (2020) 915.

62 Ma Y, Wu L, Shaw N, Gao Y, Wang J, Sun Y, Lou Z, Yan L, Zhang R & Rao Z, Structural basis and functional analysis of the SARS coronavirus nsp14–nsp10 complex. *Proc Natl Acad Sci U S A*, 112 (2015) 9441.

63 Newman JA, Douangamath A, Yadzani S, Yosaatmadja Y, Aimon A, Brandão-Neto J, Dunnett L, Gorrie-Stone T, Skyner R, Fearon D & Schapira M, Structure, mechanism and crystallographic fragment screening of the SARS-CoV-2 NSP13 helicase. *Nat Commun*, 12 (2021) 11.

64 Zhang W, Du RH, Li B, Zheng XS, Yang XL, Hu B, Wang YY, Xiao GF, Yan B, Shi ZL & Zhou P, Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes. *Emerg Microbes Infect*, 9 (2020) 389.

65 Tahir M, Coronavirus genomic nsp14-ExoN, structure, role, mechanism, and potential application as a drug target. *J Med Virol*, 93 (2021) 4264.

66 Moeller NH, Shi K, Demir Ö, Banerjee S, Yin L, Belica C, Durfee C, Amaro RE & Aihara H, Structure and dynamics of SARS-CoV-2 proofreading exoribonuclease ExoN. *BioRxiv*, (2021).

67 Pillon MC, Frazier MN, Dillard LB, Williams JG, Kocaman S, Krahn JM, Perera L, Hayne CK, Gordon J, Stewart ZD, Sobhany M, Deterding LJ, Hsu AL, Dandey VP, Borgnia MJ & Stanley RE, Cryo-EM structures of the SARS-CoV-2 endoribonuclease Nsp15 reveal insight into nuclease specificity and dynamics. *Nat Commun*, 12 (2021) 636.

68 Kumar S, Kashyap P, Chowdhury S, Kumar S, Panwar A & Kumar A, Identification of phytochemicals as potential therapeutic agents that binds to Nsp15 protein target of coronavirus (SARS-CoV-2) that are capable of inhibiting virus replication. *Phytomedicine*, 85 (2021) 153317.

69 Vithani N, Ward MD, Zimmerman MI, Novak B, Borowsky JH, Singh S & Bowman GR, SARS-CoV-2 Nsp16 activation mechanism and a cryptic pocket with pan-coronavirus antiviral potential. *Biophys J*, 120 (2021) 2889.

70 Tazikeh-Lemeski E, Moradi S, Raoufi R, Shahlaei M, Janlou MA & Zolghadri S, Targeting SARS-COV-2 non-structural protein 16: a virtual drug repurposing study. *J Biomol Struct Dyn*, 39 (2021) 4646.

71 Menachery VD, Debbink K & Baric RS, Coronavirus non-structural protein 16: evasion, attenuation, and possible treatments. *Virus Res*, 194 (2014) 199.

72 Pushkaran AC, Melge AR, Puthiyedath R & Mohan CG, A phytochemical-based medication search for the SARS-CoV-2 infection by molecular docking models towards spike glycoproteins and main proteases. *RSC Adv*, 11 (2021) 12014.

73 Li F, Structure, Function, and evolution of coronavirus spike proteins. *Annu Rev Virol*, 3 (2016) 261.

74 Mariano G, Farthing RJ, Lale-Farjat SL & Bergeron JR, Structural characterization of SARS-CoV-2: Where we are, and where we need to be. *Front Mol Biosci*, (2020) 344.

75 Modrow SU, Hahn BH, Shaw GM, Gallo RC, Wong-Staal F & Wolf H, Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: prediction of antigenic epitopes in conserved and variable regions. *J Virol*, 61 (1987) 578.

76 El-Manzalawy Y & Honavar V, Recent advances in B-cell epitope prediction methods. *Immunome Res*, 6 (2010) 9.

77 Morrow JK & Zhang S, Computational prediction of hot spot residues. *Curr Pharm Des*, 18 (2012) 1255.

78 Ibarra AA, Bartlett GJ, Hegeduüs Z, Dutt S, Hobor F, Horner KA, Hetherington K, Spence K, Nelson A, Edwards TA & Woolfson DN, Predicting and experimentally validating hot-spot residues at protein–protein interfaces. *ACS Chem Biol*, 14 (2019) 2263.

79 Mendis J, Kaya E & Kucukkal TG, Identifying Hotspots in Binding of SARS-CoV-2 Spike Glycoprotein and Human ACE2. *Biophys J*, 120 (2021) 200a.

80 Mohapatra SK & Mukhopadhyay S, Host response to SARS-CoV-2: Insight from transcriptomic studies. *Indian J Biochem Biophys*, 58 (2021) 12.

81 McDonald IK & Thornton JM, Satisfying hydrogen bonding potential in proteins. *J Mol Biol*, 238 (1994) 793.

82 Fernández A & Scheraga HA, Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. *Proc Natl Acad Sci U S A*, 100 (2003) 118.

83 Laurini E, Marson D, Aulic S, Fermeglia M, Pricl S. Computational alanine scanning and structural analysis of the SARS-CoV-2 spike protein/angiotensin-converting enzyme 2 complex. *ACS Nano*, 14 (2020) 11830.

84 Muhseen ZT, Hameed AR, Al-Hasani HM, ul Qamar MT & Li G, Promising terpenes as SARS-CoV-2 spike receptor-binding domain (RBD) attachment inhibitors to the human ACE2 receptor: Integrated computational approach. *J Mol Liq*, 320 (2020) 114493.

85 McNaughton B & Rosario PA, Computational hot-spot analysis of the SARS-CoV-2 receptor binding domain/ACE2 complex. *Chembiochem*, 22 (2020) 1200.

86 Ghorbani M, Brooks BR & Klauda JB, Critical sequence hotspots for binding of novel coronavirus to angiotensin converter enzyme as evaluated by molecular simulations. *J Phys Chem B*, 124 (2020) 10047.

87 Keskin O, Ma B & Nussinov R, Hot regions in protein-protein interactions: the organization and contribution of structurally conserved hot spot residues. *J Mol Biol*, 345 (2005) 1294.