

O- β -GlcNAcylation, Chloroquine and 2-Hydroxybenzohydrazine may hamper SARS-CoV-2 entry to human via inhibition of ACE2 phosphorylation at Ser787 but also induce disruption of virus-ACE2 binding

Waqar Ahmad^{1,2}, Khadija Shabbiri² Nazarul Islam³

¹Department of Biochemistry, College of Medicine and Health Sciences, UAE University, Al Ain UAE

²The University of Queensland, Brisbane-4072, Australia

³Institute of Chemical Sciences, University of Peshawar, Pakistan

Email addresses:

WA: waqar.ahmad@uqconnect.edu.au

waqar.ahmad@uaeu.ac.ae

waqarchemist123@yahoo.com

KS: k_shabbiri@yahoo.com

k.shabbiri@uq.edu.au

NI: islanaz@yahoo.com

*Corresponding author

Dr. Khadija Shabbiri, k_shabbiri@yahoo.com

The University of Queensland, Brisbane-4072, Australia

Dr. Waqar Ahmad, waqar.ahmad@uaeu.ac.ae

Department of Biochemistry, College of Medicine and Health Sciences, UAE University, Al Ain UAE

ABSTRACT

The novel coronavirus COVID- 19 disease is extremely contagious and has been spread worldwide. First COVID-19 case was identified in December, 2019 and within three months, more than one million affected cases and over 65,000 deaths have been reported. SARS-coronavirus 2

(SARS-CoV-2) also known as 2019-nCoV is a causative agent of COVID-19 disease and belongs to the SARS CoV (Severe Acute Respiratory Syndrome corona virus) family. The SARS-CoV-2 enters the human body by binding its viral surface spike protein with the host angiotensin-converting enzyme 2 (ACE2) receptors and cause infection. To prevent the virus entry and its transmission in the human body, we focused on the two domains of ACE2: i) the N-terminal extracellular binding domain (18-740 residues) reported for coronavirus spike interaction, and ii) the C-terminal cytoplasmic region (762-805 residues) to prevent the virus transmission. Therefore, we proposed: i) inhibition of receptor binding domain (RBD) of SARS-CoV-2 and human ACE2 protein may prevent the virus entry to the host and ii) inhibition of phosphorylation at Ser-787 of ACE2 protein may prevent the transmission of the virus in the COVID-19 patients. In the past, the critical role of Ser 787 in human ACE2 protein has been experimentally verified in SARS-CoV transmission, that upon binding to the receptor, SARS- CoV induces CKII- mediated phosphorylation of ACE2 at Ser-787 that in-turn facilitate virus entry to host cells, followed by replication and activation of ACE2, initiates downstream signaling leading to lung fibrosis. Therefore, in this study, we have suggested post-translational modification (PTM) O- β -GlcNAcylation, and two compounds Chloroquine and 2-hydroxybenzohydrazine might share the common pathways to prevent the COVID-19 infection in human. The addition of O- β -GlcNAcylation at same or neighboring Ser/ Thr residues results in phosphorylation inhibition and a change in protein structural and functional confirmations. Thereby, using neural networking methods, we have identified Ser/ Thr residues in ACE2 that are potential sites for phosphorylation and / or O- β -GlcNAcylation. Molecular docking showed that UDP-GlcNAc has more binding affinity with Ser-787 than the phosphoryl group. Moreover, chloroquine and 2-hydroxybenzohydrazine also showed great potential to bind at Ser-787 that may result in inhibition of Ser-787 phosphorylation and downstream signaling. Furthermore, O- β -GlcNAcylation, chloroquine and 2-hydroxybenzohydrazine showed their high affinity at ACE2-SARS-CoV-2 receptor binding domain that may prevent the entry of SARS-CoV-2 into human body. In conclusion, inhibition of human ACE2 phosphorylation at Ser-787 and ACE2-SARS-CoV-2 binding domain could be promising targets against SARS-CoV-2 infection.

Keywords: SARS-CoV-2, COVID-19, SARS-CoV, ACE2, spike protein, phosphorylation, O- β -GlcNAcylation, molecular docking, chloroquine, 2-hydroxybenzohydrazine

1. Introduction

Recent outbreak of novel coronavirus (SARS-CoV-2/ COVID-19) resulted in more than 1,569,849 (1,353,361 confirmed) cases while 92,191 (79,235 confirmed) deaths from Dec 2019 to 9th April, 2020 (<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>) and (<https://www.worldometers.info/coronavirus/>). The coronavirus attack resulted in severe respiratory infection that could lead to patient's death in 3-4% cases due to weak immune response. This situation could be worse in near future if there is no treatment available for this infection.

SARS-CoV-2 is an enveloped virus structurally related to other severe acute respiratory syndromes- related coronaviruses. It has 96% genomic identity with chine's horseshoe bat coronavirus. It belongs to realm "Riboviria" and family coronaviridae (Li, 2016; Lu et al., 2020; Peiris et al., 2004). Coronaviruses contains four important structural proteins, including spike, envelope, membrane and nucleocapsid. Spike protein of coronavirus is responsible for the viral attachment and entry to the human cells via human ACE2 receptors (Li, 2016; Shang et al., 2020). Chen et al showed that interaction of spike protein induced human ACE2 phosphorylation by casein kinase II followed by coronavirus infusion and activation of ERK1/2 that in turn activated fibrosis-associated chemokine ligand 2 (CCL2) regulation (Chen et al., 2010). CCL2 regulation has been found to induce severe lung inflammatory disorders followed by progressive respiratory failure and patient's death (Chen et al., 2010; Song et al., 2018). The inhibition of spike protein interaction with human ACE2 could be considered as an important target to develop a vaccine against SARS infections.

Despite phosphorylation, ACE2 may also experience modification by *O*- β -GlcNAc. It has been reported previously that *O*-GlcNAcylation could interfere with viral pathogenesis in host cells (Ahmad et al., 2011a; Ahmad et al., 2011b). During *O*- β -GlcNAcylation, one molecule of *N*-acetylglucosamine (UDP-GlcNAc) is attached to a Ser or Thr residue by the enzyme *O*-GlcNAc transferases (OGT) (Kamemura et al., 2002). All *O*- β -GlcNAc modified proteins are also phosphoproteins and exhibits both global and site-specific reciprocal relationship (Kamemura and Hart, 2003). *O*- β -GlcNAcylation of proteins results in temporary conformational changes that modulate many of their functions. Residues where *O*- β -GlcNAc and phosphorylation compete for each other are known as Yin Yang sites (Zachara and Hart, 2002). These Yin-Yang sites can be

predicted using computer-assisted neural network-based programs and analyzed to predict the modification potential and functional importance of each site.

Recent reports from various hospitals around the globe highlighted the importance of malaria- drug chloroquine (CQ) for the treatment of coronavirus infection in human. Chloroquine has been proposed as a possible treatment for the coronavirus infection (Alam and Lee, 2016; Avital et al., 1994; Keyaerts et al., 2009; Kundin et al., 1964) . In the past, Kundin WD et al (1964) found Benzol hydrazides and hydrazines derivatives effective against the influenza virus in the chick embryo lung tissue cultures (Kundin et al., 1964), and these compounds also have been suggested previously as anticancer agents (Alam and Lee, 2016). Therefore, in this study, we have used Chloroquine, 2-hydroxybenzohydrazide and O-GlcNAcylation as docking ligands, and we found that these compounds might inhibit the ACE2 phosphorylation and suppressing the interaction of coronavirus spike glycoprotein with human ACE2.

2. Materials and methods

In the present study, the proteins of interest were: i) Human Angiotensin-converting Enzyme-related Carboxypeptidase enzyme (ACE2), ii) SARS-CoV-2 coronavirus prefusion (2019-nCoV) spike glycoprotein (PDB ID: 6VSB), and iii) SARS coronavirus (SARS-CoV) spike receptor-binding domain (RBD) complexed with its receptor (ACE2) (PDB ID: 2AJF), used to determine their interactions with each other and with the docked ligands. The ligands UDP- GlcNAc (PubChem CID: 445675), Chloroquine (PubChem CID: 2719), 2-hydroxybenzohydrazide (PubChem CID: 13637) and Phosphate were used for docking experiments.

2.1. Sequence Alignment of human ACE2 protein

The FASTA sequence of human ACE2 was retrieved from the Uniprot sequence database with the entry name ACE2_human and accession number Q9BYF1 (Wu et al., 2006). A homology search of the ACE2 protein sequences of all organisms in the NCBI database was done with the BLASTP algorithm using default parameters (Altschul et al., 1997). A total of 100 hits were obtained for human ACE2 with an *E*-value of zero. Sequences based on isoforms, predicted, hypothetical, synthetic, and unnamed proteins were neglected to ensure reduced redundancy in selection. Out of 100 retrieved sequences, fourteen were selected representing major mammalian groups. The

Uniprot accession numbers for fourteen selected sequences (Fig. 1) were Human (Q9BYF1), *R. pearsonii* (Pearson horseshoe bat, B0LW5_9CHIR), *R. macrotis* (Big eared bat, E2DHI9_9CHIR), *R. pusillus* (least horseshoe bat, E2DHI9_9CHIR), *R. sinicus* (Chinese rufous bat, E2DHI7_9CHIR), Gorgo (G3QWX4), Pantr (A0AA2J8KU96), Panpa (A0A2R9BkD8), Nomle (GIRE79), Macne (AOA2K6D1N8), Macmu (B6DUE2), Rat (Q5EGZ1) and Mouse (Q8R010). ClustalW2 (Thompson et al., 1994) (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) was used for multiple alignments of all the sequences of ACE2 to get the conservation status of Ser and Thr residues.

2.2. Post-translational modifications prediction

In this study, well documented prediction programs (Qian and Sejnowski, 1988; Saraswathi et al., 2012; Thompson et al., 1994) have been used to determine the post-translational modifications (phosphorylation and O- β -GlcNAcylation) on human ACE2 protein.

2.3. Prediction of phosphorylation residues and related kinases

Phosphorylation potential for human ACE2 was predicted for every Ser, Thr and Tyr residues using NetPhos 3.1 (<http://www.cbs.dtu.dk/services/NetPhos/>) with the minimum threshold value of 0.5 (Blom et al., 1999). For further confirmation, DISPHOS 1.3 (<http://www.dabi.temple.edu/disphos/>) and KinasePhos (<http://kinasephos.mbc.nctu.edu.tw/>) were used to predict possible phosphorylation sites. Kinases are involved in phosphorylation on residues, therefore Kinase specific phosphorylation sites in human ACE2 were predicted by NetPhosK 1.0 (<http://cbs.dtu.dk/services/NetPhosK>) (Blom et al., 2004) and KinasePhos (<http://kinasephos.mbc.nctu.edu.tw/>) (Huang et al., 2005). These programs predict kinase specific acceptor substrates, including Ser, Thr and Tyr.

2.4. Prediction of O- β -GlcNAcylated residues

O- β -GlcNAc modification potential sites were predicted by using YinOYang 1.2 (<http://www.cbs.dtu.dk/services/YingOYang/>) (Gupta and Brunak, 2002), OGTSITE

(<http://csb.cse.yzu.edu.tw/OGTSite/index.php>) (Kao et al., 2015) and GlycoEP (<http://crdd.osdd.net/raghava/glycoep/>) (Chauhan et al., 2013).

2.5. Prediction of Yin-Yang sites

The potential phosphorylation and O- β -GlcNAcylation sites on ACE2 protein were predicted by YinOYang 1.2 server. The predicted Yin-Yang sites are with the highly uneven threshold that is adjusted in accordance with amino acid surface accessibility. False-negative (FN) Yin-Yang sites were determined by coupling modification potential and conservation status of Ser/ Thr residues described in previous studies (Ahmad et al., 2006; Ahmad et al., 2011c). NetSurfP-2.0 (<http://www.cbs.dtu.dk/services/NetSurfP/>) (Klausen et al., 2019) was used to predict the surface accessibility of every Ser and Thr residue of human ACE2 3D structure. PaleAle 5.0 (<http://distilldeep.ucd.ie/paleale/>) (Kaleel et al., 2019) was used to predict the available area of Ser and Thr residues for ligand interaction. Yin-Yang sites were also predicted on 3D structure of ACE2 protein through docking experiments with the addition of the phosphate group, and UDP-GlcNAc individually using SwissDock server (<http://www.swissdock.ch/>) (Grosdidier et al., 2011).

2.6. Human ACE2 protein 3D structure prediction

The experimentally determined crystalline protein structure of native human Angiotensin-converting Enzyme-related Carboxypeptidase (ACE2) is available in the protein data bank (PDB ID: IR42) with 1-615 residues. To fully understand the post-translational modifications on Ser and Thr residues and their surface accessibility, a complete 3D structure of ACE2-Human with 805 residues was required. Therefore, the complete 3D structure of ACE2-Human was predicted using I-TASSER server (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>) (Zhang, 2008). The best predicted ab-initio models of ACE2-Human were selected based on their 1) similarity with the experimentally verified human ACE2 protein structure, and 2) chemical correctness of protein's backbone angles using Ramachandran plots. Molprobability (<http://molprobability.biochem.duke.edu/>) and RAMPAGE (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>) (Davis et al., 2004) servers were used to perform geometrical evaluations and Ramachandran plot's formation.

2.7. 3D structure prediction of SARS-CoV-2 (2019-nCoV) receptor-binding domain (RBD)

The SARS-CoV-2 (2019-nCoV) spike (PDB ID: 6VSB) is a trimer protein. Therefore, FASTA sequence of its chain A was only selected for sequence alignment with the FASTA sequence of Chain E of dimer SARS-CoV spike protein (PDB ID: 2AJF). The FASTA sequences downloaded from RCSB Protein Data Bank to do a comparison between the residues of the two proteins in the binding domain regions. EMBOSS Water (https://www.ebi.ac.uk/Tools/psa/emboss_water/) was used for the sequence alignment of SARS-CoV-2 and SARS CoV spike proteins.

The experimentally determined cryo-electron microscopy structure of the SARS-CoV-2 spike in the perfusion conformation (PDB ID: 6VSB) lacks the residues from TYR453-LEU492 in its structural conformation (Wrapp et al., 2020). Therefore, SwissProt was used to predict the complete structural model of SARS-CoV-2 by selecting two templates: 2019-nCoV (SARS-CoV-2) spike protein (PDB ID: 6VSB), and SARS-CoV spike protein (PDB ID: 2AJF). The best ab-initio model was selected for SARS-CoV-2 (2019-nCoV) spike protein and structurally aligned with the SARS-CoV spike of the SARS-CoV spike- ACE2 receptor complex (PDB ID: 2AJF) for the comparative study at spike-receptor binding domain. The structural alignments of the proteins were done in UCSF Chimera.

2.8. Protein structure preparation for docking analysis

To prepare the predicted 3D structures of ACE2 or SARS-CoV-2 spike proteins individually, the highly scored I-Tasser predicted 3D protein structure of human ACE2, and the highly scored Swiss Model predicted structure of SARS-CoV-2 spike protein respectively was selected. The structures were opened in UCSF Chimera 10.1 (Resource for Biocomputing, Visualization, and Informatics, University of California San Francisco, CA, USA). Hydrogen, and Gasteiger charges were added to the structures using Dock Prep, a built-in tool in UCSF Chimera and saved the structure in pdb format for docking. Swissdock and Autodock Vina, a built-in tool of UCSF Chimera was used for docking.

To prepare the experimentally determined structure of SARS-CoV-2 (2019-nCoV) spike glycoprotein, the Cryo-EM structure of COVID-19 spike glycoprotein with a single RBD up (PDB ID: 6VSB) was downloaded in pdb format from RCSB PDB (<https://www.rcsb.org/structure/6VSB>) and opened in UCSF Chimera 10.1 to delete the nonstandard residues and chains B and C. The SARS-CoV-2 spike is a trimer protein; therefore,

only one chain, i.e. chain A was retained for Dock Prep preparation in UCSF Chimera to add and merge hydrogens and Gasteiger charges and saved in pdb format for further docking procedure.

The pdb format of crystal structure of SARS-CoV spike RBD complexed with its receptor (ACE2) (PDB ID: 2AJF) was downloaded from RCSB Protein Data Bank (<https://www.rcsb.org/structure/2ajf>) and opened in UCSF Chimera 10.1 to delete the nonstandard residues and chain B of bound receptor (ACE2) protein, and chain F of the SARS-coronavirus spike protein. In this spike-receptor structure, chain A of ACE2 receptor protein and chain E of SARS spike protein was retained for Dock Prep preparation in UCSF Chimera to add and merge hydrogens and Gasteiger charges and saved in pdb format for further docking procedure via online SwissDock server.

2.9. Ligands preparation for Docking Analysis

The ligands UDP- GlcNAc (PubChem CID: 445675), Chloroquine (PubChem CID: 2719) were fetched by their PubChem CIDs into the UCSF Chimera and saved in pdb format. On the other side, the structure of phosphate was created in the Open Bable converted into Mol format, hydrogens were added and then converted into MOL2 format. The individual pdb files of UDP- GlcNAc and Chloroquine were also opened in Open Bable and converted into Mol format, hydrogens were added and then converted into MOL2 formats. All MOL2 files were then subjected to DockPrep in UCSF Chimera at default parameters for preparing the ligands. The energy minimization and dock preparation of 2 -hydroxybenzohydrazide (PubChem CID: 13637) ligand was done in UCSF Chimera and saved as Mol2 file.

2.10. Docking Analysis:

In SwissDock program, MOL2 files of the ligands were submitted along with the prepared proteins (pdb files) for flexible docking at default parameters. The predicted docking files generated from SwissDock were visualized and analyzed in UCSF Chimera 10.1. While for docking ligand in the binding pocket of the protein, Auto Grid program of Chimera with a grid marker having dimensions of 30x30x30 was used in AutoDock Vina. To determine the surface accessibility of the residues in the binding pockets, UCFS Chimera X version 0.93 was used to generate the hydrophobicity surfaces. Our approach to select a docking poses in each analysis comprised of:

Fig 1: Multiple alignments of fourteen mammalian sequences of ACE2 protein. The consensus residues are highlighted by an asterisk (Ser: red highlighted, Thr: green highlighted), conserved substitutions by double dots (Ser/Thr: yellow highlighted), and semi-conserved substitutions by single dot (Ser/ Thr: blue highlighted). CLUSTALW was used to align the sequences.

Absolutely conserved, conserved substitutions and semi-conserved Ser and Thr residues within each subtype were determined among the fourteen mammalian species. Fig. 1 showed Ser-2, 4, 10, 19, 43, 44, 47, 70, 77, 105, 109, 124, 128, 155, 257, 317, 331, 409, 411, 502, 511, 545, 547, 563, 602, 611, 617, 623, 626, 645, 646, 680, 692, 707, 721, 783, and 804 (37/54); and Thr-20, 27, 52, 129, 276, 282, 294, 324, 334, 347, 362, 365, 371, 414, 434, 449, 453, 496, 571, 519, 548, 567, 608, 686, 730, 763, and 803 (27/39) are highly conserved in mammals. While Ser-170, 420, 507, and 740 and Thr-15, 118, 122, and 125 were found to be conserved substitutes. On the other side, Ser-3, 113, 167, 218, 280, 425, 607, 709, 776 and 787; and Thr-55, 78, 92, 229, 445, 593, 698, and 798 showed semi-conserved substitutions and mainly substituted by Ala (alanine) and Asn (asparagine) residues.

3.2. 3D Modelling of ACE2 protein

In this study, we predicted the complete 3D structure of ACE2 protein comprises of 805 amino acids using I-Tasser server to examine the post-translational modifications, surface accessibility (buried and exposed) of the residues for compounds to be interacted to inhibit the binding of corona virus spike protein.

The server used the available experimentally determined ACE2 3D template (PDB ID: 1R42) (1-615AA) and predicted five possible models (Fig S1). All the predicted models were aligned with currently available partial 3D structure, and model with 100% identity and low geometrical errors was chosen for further study (named as M1-ACE2, Fig 2).

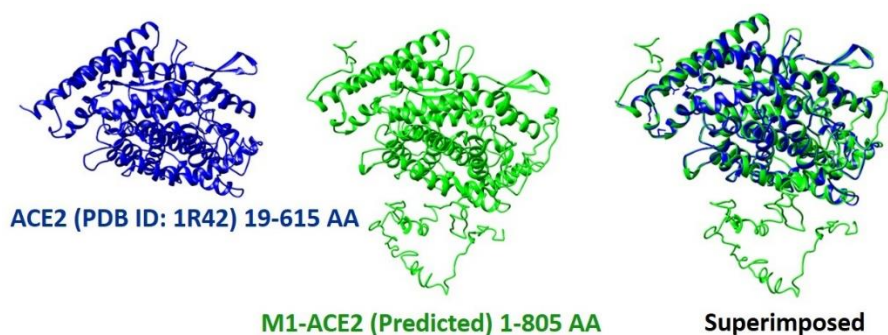


Fig 2: Complete 3D structural model of human ACE2. The complete model (AA: 1-805) was retrieved from I-TASSER using already available ACE2 template with PDB ID: 1R42 (AA: 19-615).

We constructed Ramachandran plots against each predicted model (Fig S1). The Ramachandran plot analysis estimates phi and psi torsion angles based on experimentally known data and standard geometrical properties of protein structure (Hollingsworth and Karplus, 2010). For the first time, we proposed these five predicted 3D structures and submitted in the PMDB Protein Model Database (<http://srv00.recas.ba.infn.it/PMDB/main.php?submission=false>) (Castrignano et al., 2006) in pdb file extensions with PMDB IDs: PM0082979, PM0082975, PM0082976, PM0082977, and PM0082978, respectively. These submitted structures can be accessed from PMDB by entering their IDs.

3.3. High potential of phosphorylation and O- β -GlcNAcylation modifications on ACE2 protein

ACE2 has been found to be phosphorylated and being a main target in coronavirus infection. ACE2 phosphorylation could mediate host-pathogen responses (Chen et al., 2010).

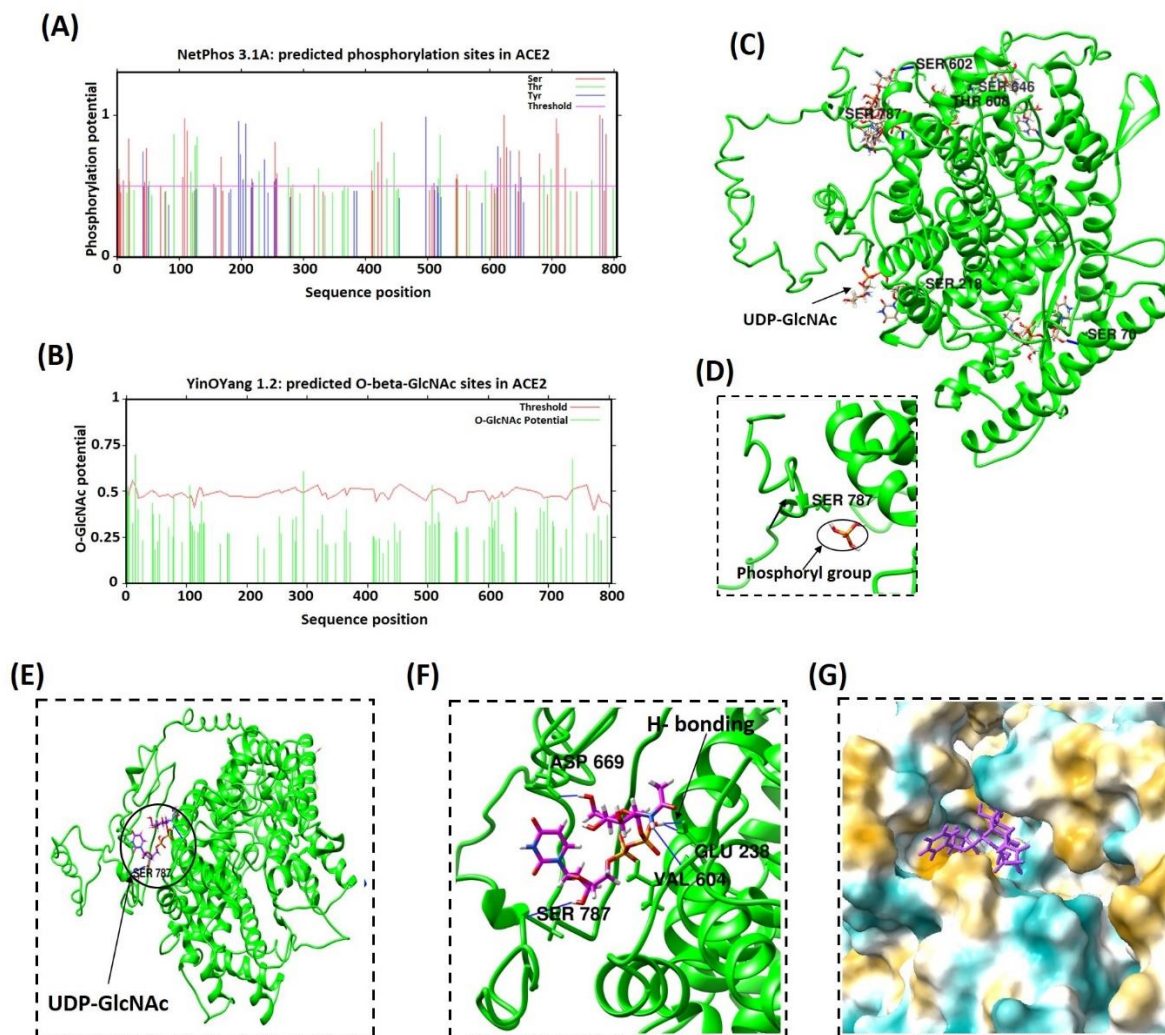


Fig 3: Potential phosphorylation and O-β-GlcNAcylation Ser and Thr residues and predicted YinYang residues in human ACE2. (A) Phosphate modification potential at Ser/ Thr residues in the human ACE2. The light gray colored line shows the threshold for phosphorylation modification. The blue, green and red lines specify phosphorylation potential for Ser, Thr and Tyr, respectively. This graph was generated using NetPhos 3.1a server. (B) Ser/ Thr residues with potential for O-β-GlcNAcylation modification. Green vertical lines show probable O-GlcNAc potential of Ser and Thr residues in whole sequence. Light blue wavy line indicates the threshold for O-GlcNAc modification. This graph was generated using YinOYang 1.2 server. (C) Predicted Yin-Yang sites showing docking of UDP GlcNAc within 3D structure of ACE2 protein. Six Yin-Yang residues were predicted including Ser-70, 218, 602, 646, 787, and Thr-608. (D) Presence of the phosphoryl group in the vicinity of Ser-787 of ACE2. (E-G): The interacting ligand UDP-GlcNAc is shown in purple color with its hetero colored atoms. (E) An overview of UDP-GlcNAc binding with human ACE2. (F) Cavity zoom of UDP-GlcNAc and ACE2 binding showing strong hydrogen bonding of UDP-GlcNAc with surrounding amino acid residues Glu-238, Val-604, Asp-669 and Ser-787. The UDP-GlcNAc has strong binding affinity with the human ACE2 with binding energy -8.66 kcal/mol. (G) Hydrophobic surface view of UDP-GlcNAc with ACE2 binding (golden: hydrophobic, white: intermediate, cyan: hydrophilic). ChimeraX was used to create the surface diagram from UDP-GlcNAc and ACE2 interaction. The predicted docking position is located inside the hydrophobic-polar pocket. Polar-hydrophilic residues His-239, Lys-596, Val-604, Gly-608, Asp-669 and Ser-787 are in the proximal position to the UDP-GlcNAc.

The phosphorylation potentials of the Ser and Thr residues and kinases involved for phosphorylation were shown in Table 1, whereas graphical representation of potential phosphorylation residues was shown in Fig 3A. Forty-seven (47) Ser/ Thr residues were predicted potential phosphorylation sites. However, among these 23 were buried or less accessible residues for modifications. The accessible potential phosphorylated sites were comprised of 24 Ser (2, 3, 19, 47, 70, 105, 106, 109, 113, 124, 128, 218, 254, 317, 420, 425, 602, 617, 626, 646, 709, 721, 783, 787) and 20 Thr (15, 20, 27, 55, 92, 118, 122, 125, 276, 324, 365, 434, 445, 496, 548, 593, 686, 698, 763, 803) residues. A number of kinases may be implicated in phosphorylation of Ser and Thr residues. Almost each kinase predicted was involved in phosphorylation of two or more residues (Table 1).

Both Ser and Thr residues have found to be involved in O- β -GlcNAc modification (Audagnotto and Dal Peraro, 2017; Blom et al., 2004; Xin and Radivojac, 2012). In our study, prediction of O- β -GlcNAcylation sites showed ACE2 had high potential for O- β -GlcNAc modification (Table 1, Fig 3B) and we suggested 37 putative O- β -GlcNAc sites including Ser-2, 3, 5, 105, 106, 124, 128, 317, 420, 425, 545, 607, 617, 626, 692, 709, 721, 783, 787, 804 and Thr-15, 20, 55, 125, 294, 324, 334, 347, 365, 371, 445, 496, 686, 698, 730, 798, 803. All these sites were highly conserved in mammals except Ser-5 and 106 (Fig. 1). Ser-3, 425, 607, 787, and Thr-55, 445, 698, and 798 were semi-conserved substitution (substituted by glycine and/ or asparagine in some mammals), while Ser-420 and Thr-15 and 125 were semi-conserved substitutions (substituted by alanine in rat and mouse) (Fig. 1).

Both phosphorylation and O- β -GlcNAcylation could compete for Ser and Thr linkage on the same or neighboring residues. These residues are known as Yin-Yang sites. Yin-Yang sites determined by the neural networking software could be false-positive or false-negative due to the residues which are completely buried or not accessible for phosphorylation and O- β -GlcNAcylation modifications. To minimize the false predictions, UDP-GlcNAc sugar was docked with human ACE2' 3D structure, and we found a total of thirteen Ser (47, 70, 218, 511, 602, 611, 617, 626, 646, 787) and Thr (276, 371, 608) residues as Yin-Yang sites in ACE2 protein. Most of the residues showed high potential for the O-GlcNAcylation using neural networks except Ser-47, 70, 602, 646 and 787, and Thr-276 and 371. These sites are known as false-negative Yin-Yang sites. We proposed six ACE2 residues, including Ser-70, 218, 602, 646, 787 and Thr-608 as possible Yin-Yang sites based on the highest energy release during docking UDP-GlcNAc with

ACE2 3D structure (Fig 3C, Table 2). Hydrogen bonding and energy released during binding of UDP-GlcNAc with these Yin-Yang sites were shown in Table 2. UDP-GlcNAc was strongly bound ($\Delta G = -8.775$) through five hydrogen bonds with Ser 787 (2.504 Å), Asp-669 (1.871 Å), Arg-671 (2.471 Å), Hsd-239 (2.071 Å) and Glu-238 (1.753 Å) residues in the target binding pocket of ACE2 receptor protein (Fig 3. E-F, Table 2). Fig 3D showed presence of a phosphoryl group in the vicinity of Ser-787 with maximum energy released (-2.1 kcal/mol). These results showed that ACE2 has high potential for O- β -GlcNAcylation binding upon competition with phosphorylation modification at the same or neighboring residues.

3.4. Chloroquine and 2-hydroxybezohydrazide may alter downstream signaling of human ACE2 protein by inhibiting phosphorylation

Recent reports from clinics caring coronavirus- infected patients showed a positive effect of malaria drug chloroquine (CQ) on eliminating/ reducing the SARS-CoV-2 symptoms. Therefore, in this study, we tested CQ along with another compound known as 2-hydroxybezohydrazide (HBH) to observe the inhibitory effect on human receptor protein ACE2. The conformation of CQ in Fig 4 (A, B, C) shows that it could be able to form two hydrogen bonds with the residues of Ser 787 (2.702 Å) and Hsd 788 (2.831 Å) of ACE2 receptor with high binding affinity ($\Delta G = -7.111$ kcal/mol).

In the past (1964), 2-hydroxybenzohydrazide (HBH) was effectively used to inhibit the growth of influenza virus in chick embryo lung tissue culture (Kundin et al., 1964). The derivatives of HBH were found to have anti-microbial activities and anti- influenza activity (Alam and Lee, 2016; Kundin et al., 1964). In our study, HBH was found to bind with the ACE2 receptor via two hydrogen bonds with Ser-787 (2.385 Å) and the neighboring residue Lys-788 (2.108 Å) as shown in Fig 4 (D, E, F). These findings showed that both CQ and HBH might compete with the phosphorylation on Ser 787 and results in alteration in the structural and functional conformation of ACE2 receptor protein, thus inhibits the SARS-CoV-2 spike interaction and signaling to proceed with the cascade of the infection.

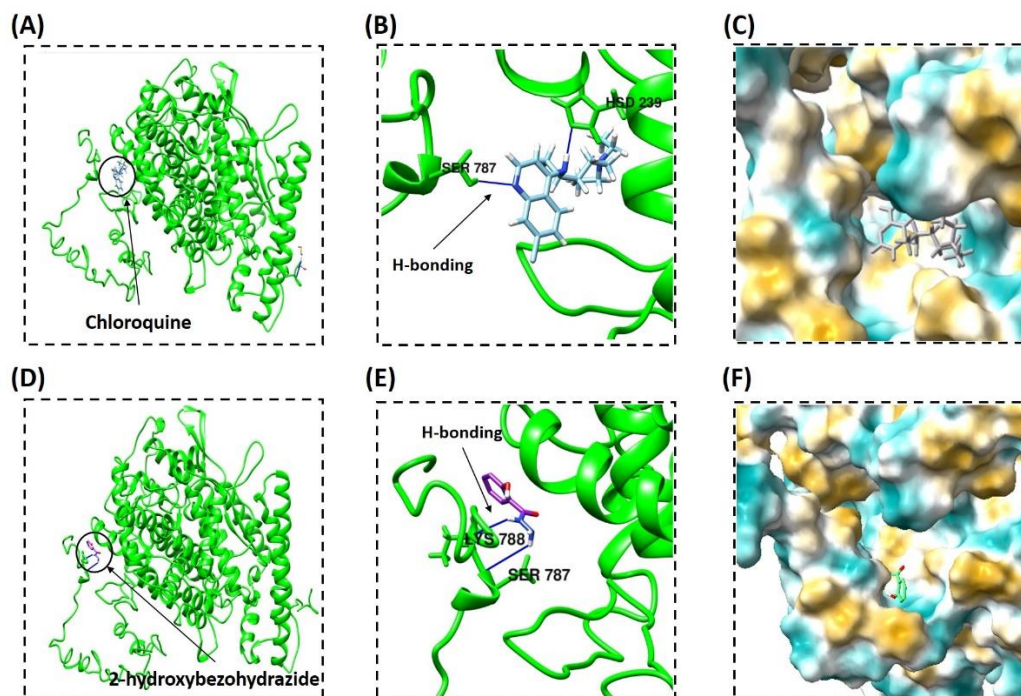


Fig 4: Molecular docking of chloroquine and 2-hydroxybenzohydrazine with human ACE2. Protein structure has been shown in rainbow colors. Blue lines indicate hydrogen bond's formation of ligands with the protein. (A-C): The interacting ligand CQ is shown in gray color with its hetero colored atoms. (A) An overview of CQ with human ACE2 showing its exact position in the whole molecule. (B) Cavity zoom of CQ-ACE2 binding. Hydrogen bonding of CQ with surrounding amino acid residues Ser-787 and Hsd-239. The CQ has strong binding affinity with the human ACE2 with binding energy -7.11 kcal/mol. (C) Hydrophobic surface view of CQ-ACE2 binding (golden: hydrophobic, white: intermediate, cyan: hydrophilic). ChimeraX was used to create surface diagram from CQ-ACE2 interaction. Fig 6 (D-F): The interacting ligand HBH is shown in parrot green color with its hetero colored atoms. (A) An overview of HBH with human ACE2 showing its position in the whole molecule. (B) Cavity zoom of HBH-ACE2 binding. Hydrogen bonding of HBH with surrounding amino acid residues Ser-787 and Lys-788. The HBH has strong binding affinity with the human ACE2 with binding energy -5.3 kcal/mol. (C) Hydrophobic surface view of HBH-ACE2 binding (golden: hydrophobic, white: intermediate, cyan: hydrophilic). ChimeraX was used to create a surface diagram from HBH-ACE2 interaction.

3.5. Sequence and structural alignment of SARS-CoV-2 and SARS-CoV Receptor binding domains

During SARS- infections, the receptor binding domain of SARS virus interacts with N- terminal of human ACE2 receptor. Therefore, in this study, the amino acid sequence and 3D structural alignment of SARS-CoV-2 spike and SARS-CoV RBD had been done (Fig5A, B) for comparative analysis. The sequence alignment of SARS-CoV-2 RBD (AA 323-502) with SARS-CoV-RBD (AA 323-501) resulted in 72.2% identity and 80% similarity with 1 gap (Fig 5A). The Sequence alignment showed following mutations in residues from SARS-CoV RBD to SARS-CoV-2 RBD: T424S, R425N, I427L, A429S, T430K, S431V, T432G, K438L, L443F, H445K, G446S, K447N, N457T, V458Em P459I, F460Y, S461Q, P462A, D463G, G464S, K465T, T468N, P469G, P470E,

A471G, L472F, W476F, N479Q, D480S, Y484Q, T485P, T487N, and I489V (Fig 5A, C). Some of these mutations resulted in changes of the propensity of the residues, which may affect the interaction of the SARS-CoV-2 spike RBD with the receptor. The Y442L mutation changes the propensity from polar to hydrophobic, there was no change in the propensity of polarity at mutations: Y479Q, Y484Q, T487N, and L472F. The mutation at R426N changes propensity of the residue from charged to polar (Fig 5C). These propensity changes may alter the interactions between ACE2 and SARS-CoV-2 spike RBD and can be used as possible drug targets.

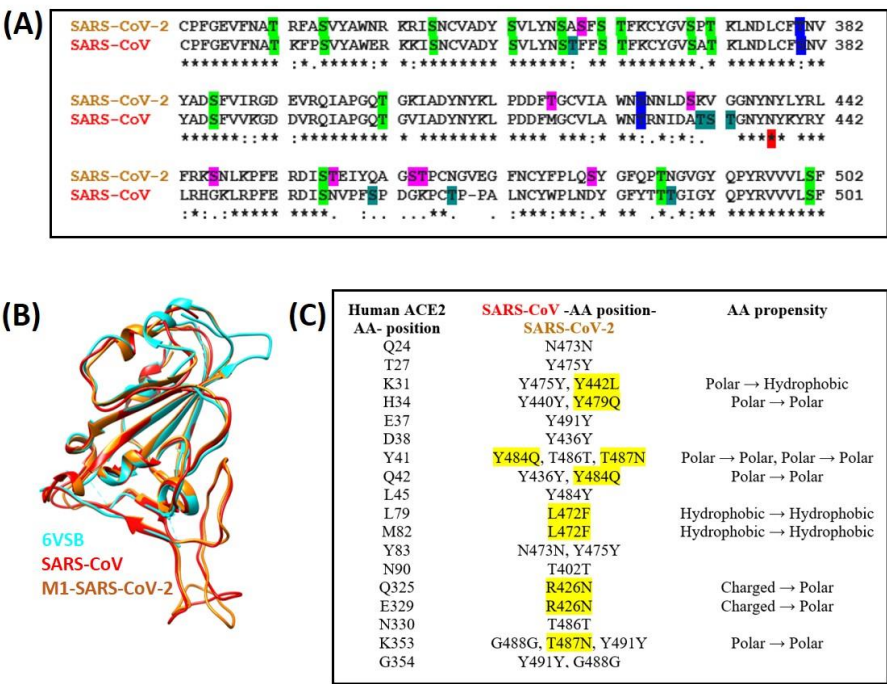


Fig 5: Sequence alignments, homology modeling, and characterization of SARS-CoV-2 RBD. (A) Amino acid alignment for RBD of SARS-CoV-2 and SARS-CoV. Highlighted residues: green= conserved Ser/ Thr between two sequences, blue= conserved substitution, pink= new Ser/ Thr residues introduced in SARS-CoV-2, sea green= Ser/ Thr residues in SARS-CoV replaced by other amino acids in SARS-CoV-2. (B) Homology model of RBD of SARS-CoV-2 was predicted using Swissdock. The model selected as possible complete structure of SARS-CoV-2 RBD contained already experimentally determined templates (6VSB and 2AJF) of RBD and named M1-SARS-CoV-2. (C) Detailed binding interface of predicted SARS-CoV-2 RBD with ACE2 N- terminal.

We predicted the complete 3D structure of SARS-CoV-2 spike to include the missing residues Tyr 453-Leu 492 of RBD of the experimentally determined SARS- CoV-2 spike (PDB ID: 6VSB). Two predicted models (model 1 and model2) for SARS- Cov-2 spike RBD from SwissProt were submitted in the PMDB Protein Model Database in pdb file extensions with IDs: PM0082982 and PM0082983, respectively (Fig 5, Fig S2). The highly scored Model 1 of SARS- Cov-2 spike RBD was selected for further studies. While finalizing this manuscript, a non- refined x-ray structure of SARS-CoV-2 RBD interacting ACE2 was recently published (PDB ID: 6LZG,

<https://www.rcsb.org/structure/6LZG>). Our predicted structure showed maximum alignment with the 6LZG (Fig S3. Figure 5B showed the structural alignment of highly scored predicted Model 1 of SARS-CoV-2, experimentally determined SARS-CoV-2 spike structure (6VSB), and SARS-CoV spike RBD (PDB ID: 2AJF).

3.6. O- β -GlcNAcylation modification in the ACE2 and SARS-CoV-2 RBD complex

SARS-CoV-2 RBD with ACE2 receptor complex was prepared in Chimera by replacing the SARS-CoV spike with SARS-CoV-2 spike in SARS-CoV RBD-ACE2 receptor complex (PDB ID: 2AJF). We found four hydrogen bonds between SARS-CoV-2 and ACE2 residues in RBD [Tyr436-Gln42 (2.228Å), Thr487-Tyr41 (3.021Å), Gly489-Lys353 (2.907Å) and Tyr492-Glu37 (3.143Å)], as shown in Fig 6A.

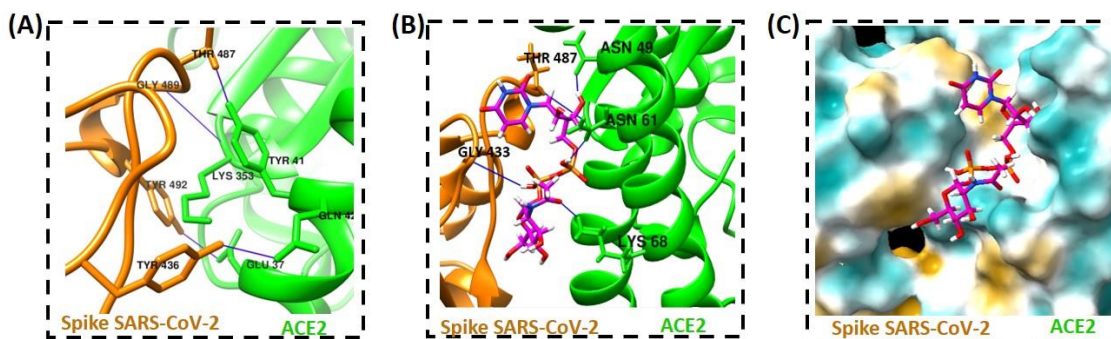


Fig 6: Interactions between ACE2 with SARS-CoV-2 and SARS-CoV at RBD and molecular docking with UDP-GlcNAc. ACE2 structure has been shown in green color while SARS-CoV-2 in golden and SARS-CoV in red. Blue lines indicate hydrogen bonds formation of ligands with the protein. The interacting ligand UDP-GlcNAc has been shown in magenta color with its hetero colored atoms. (A) An overview of hydrogen bonding between SARS-CoV-2 RBD with human ACE2. (B) Cavity zoom of UDP-GlcNAc molecule docked in the vicinity of SARS-CoV-2 and ACE2. (C) Hydrophobic surface view of UDP-GlcNAc with SARS-CoV-2 RBD with human ACE2 (golden: hydrophobic, white: intermediate, cyan: hydrophilic). ChimeraX was used to create the surface diagram for the interaction.

Artificial neural networks showed high protentional for O- β -GlcNAcylation on SARS-CoV-2 spike RBD at Ser346, 370, 464 and Thr380, 465, 487 residues. While, the docking studies showed high affinity ($\Delta G = -9.001$ kcal/mol) of UDP-GlcNAc to ACE2 with four hydrogen bonds at Asn49 (1.935 Å), Asn61 (2.574 Å), Lys68 (2.160 Å), Gly433 (1.788 Å) in the vicinity of Thr487 of the SARS-CoV-2 in RBD (Fig 6B, Table 2). Our results showed, the binding position of UDP-GlcNAc in the RBD may alter the conformation and thus inhibit the binding of the SARS-CoV-2 spike with the human ACE2 receptor to prevent the entry into the host cells. The binding pocket of this predicted docking position in RBD (Fig 56) had other polar-hydrophilic residues of

SARS-CoV-2 spike protein including Val-432, Gly-433, Tyr-436, Ser-481 and Thr-487 in the vicinity of UDP-GlcNAc, and polar-hydrophilic residues of ACE2 receptor including Leu-39, Asp-38, Gln-42, Leu-45, Ala-46, Glu-57, Asn-61, Asn-64 and Lys-68, also supported by Li et al (Li et al., 2005). Moreover, in this study, these residues have been found in the interactions between SARS-CoV-2 spike RBD and ACE2 complex in various conformations, and binding of UDP-GlcNAc may disrupt these interactions to inhibit the binding of SARS-CoV-2 spike binding with the human ACE2 receptor.

3.7. Chloroquine and 2-hydroxybezohydrazide have potential to inhibit the binding at RBD of SARS-CoV-2 -human ACE2 receptor complex

In this study, CQ and HBH have also been found to interact with the RBD of SARS-CoV-2 spike and ACE2 receptor complex. One conformation showed CQ binding via one hydrogen bond with the Glu35 residue (1.869Å) of ACE2 with high binding affinity ($\Delta G = -8.581$ kcal/mol) in a binding pocket interacting with the RBD of SARS-CoV-2 spike (Fig 7A). The other conformation of CQ showed its binding via one hydrogen bond with the Glu471 residue (1.809 Å) of RBD of SARS-CoV-2 spike in a binding pocket interacting with the ACE2 with high binding affinity ($\Delta G = -8.441$ kcal/mol) (Table 2). Both CQ conformations showed that these interactions in either way may disrupt the binding of SARS-CoV-2 spike with the human ACE2 receptor to inhibit the virus entry into the host cell.

Similar to CQ, HBH also showed high binding affinity ($\Delta G = -6.301$ kcal/mol) with the Glu471 residue (1.997Å) of SARS-CoV-2 spike via one hydrogen bond, and with the Glu35 residue (2.318Å) of ACE2 via second hydrogen bond in the RBD cavity of SARS-CoV-2 spike and ACE2 receptor complex (Fig 7B, Table 2). Various conformations of HBH in the RBD cavity showed that either HBH binds with the SARS-CoV-2 spike RBD or the ACE2 receptor or both may disrupt the structural and functional conformation of the SARS-CoV-2 spike and ACE2 complex, thus inhibit the entry of virus into the host cell.

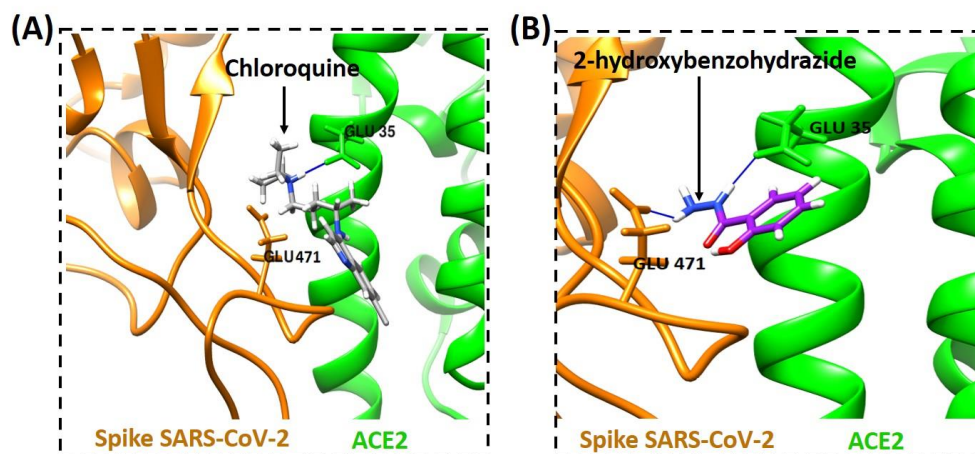


Fig 7: Molecular docking of chloroquine and 2-hydroxybenzohydrazide at SARS-CoV-2 RBD interacting with human ACE2. ACE2 protein structure has been shown in green color while SARS-CoV-2 in golden colors. Blue lines indicate hydrogen bond's formation of ligands with the protein. (A): The interacting ligand CQ is shown in gray color with its hetero colored atoms in the cavity zoom. Hydrogen bonding of CQ with surrounding amino acid residue Glu-35 (ACE2). (B) The interacting ligand HBH is shown in purple color with its hetero colored atoms. (A) An overview of HBH with human ACE2 showing its position in the whole molecule. Hydrogen bonding of HBH with surrounding amino acid residues Glu-471 (SARS-CoV-2) and Glu-35 (ACE2).

4. Discussion

The current global outbreak of coronavirus SARS-CoV-2 demands efficient therapeutics against this infection. Due to limited clinical and experimental data, researchers are focusing on computational analysis to examine the SARS-CoV-2 structure and function, and its interaction with the host protein to find the possible cure. In this study using computational techniques, we proposed a mechanism and compounds to inhibit the SARS-CoV-2 spike interaction with human ACE2 receptor, entry into host cells, and downstream signaling. Chen et al (2010) showed that upon binding with ACE2 receptor, SARS-CoV induces ACE2 phosphorylation at Ser-787 via activation of CKII (Chen et al., 2010). This process does not only facilitate SARS entry to host cells for reproduction but also induces CCL2 signaling to cause lung fibrosis and severe infection (Chen et al., 2010). Our results proposed that inhibition of phosphorylation of human ACE2 receptor at Ser-787 might block downstream signaling of spike protein of SARS-CoV-2. This will in turn reduce virus replication and ACE2 downstream signaling to control the COVID-19 disease spread.

To inhibit the phosphorylation of ACE2 at Ser-787, we have proposed two inhibition mechanisms i) inhibition through PTMs and (ii) inhibition through drug binding. Post translational modifications (PTMs) at specific residues cause geometrical and conformational changes in proteins to alter their functions (Duan and Walther, 2015; van der Laarse et al., 2018; Xin and Radivojac, 2012). Interestingly, both phosphorylation and O- β -GlcNAc compete for the same or neighboring Ser/ Thr residues and change protein conformation and its functional process. These Ser/ Thr residues are known as Yin-Yang sites (Leney et al., 2017). ACE2 has been found to be phosphorylated at two residues Ser-680 and Ser-787. The phosphorylation at Ser-787, present at C-terminal of ACE2 is responsible for its downstream signaling upon SARS interaction (Chen et al., 2010). Ser-787 is a conserved residue except in few mammals, including mouse and rat as shown by multiple alignments.

The first mechanism is to inhibit the ACE2 phosphorylation with the O- β -GlcNAc at Ser-787 by O-GlcNAc transferase (OGT). (Lu et al., 2020). Unlike phosphorylation, it is usually difficult to detect O- β -GlcNAc on proteins due to several technical and experimental limitations (Li et al., 2019; Zachara et al., 2015). To overcome this difficulty, several neural network methods have been used to predict O- β -GlcNAc on Ser/ Thr residues based on available experimental data and consensus sequences (Wang et al., 2018). Our results showed that ACE2 has high potential for O- β -GlcNAcylation, which can inhibit the downstream signaling of SARS-COV-2 by reducing the ACE2 phosphorylation at Ser-787. The neural network methods usually predict PTMs based on the secondary structure and homology of proteins of interest (Audagnotto and Dal Peraro, 2017; Blom et al., 2004).

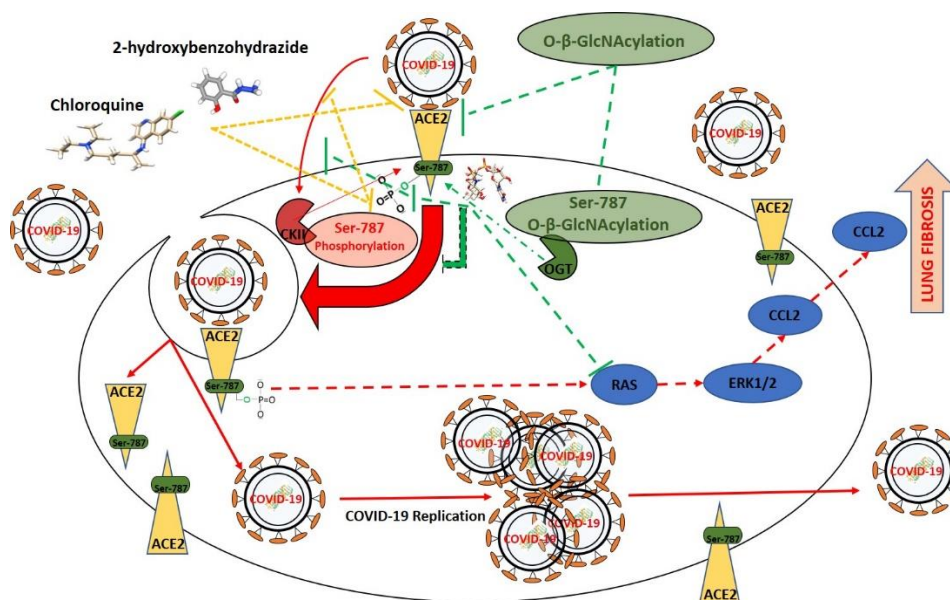


Fig 8: Schematic diagram showing the impact of inhibition of Ser-787 phosphorylation either by UDP-GlcNAc modification via O-GlcNAc transferase (OGT) or by using specific inhibitor's chloroquine or 2-hydroxybenzohydrazide. SARS-CoV-2 infection induces phosphorylation on Ser-787 residue of ACE2 via casein kinase II (CKII) followed by host cell entry and induction of lung fibrosis via activation of ERK1/2 and CCL2 pathways. UDP-GlcNAc or CQ or HBH could compete with the phosphoryl group on Ser-787 residue and inhibits Ser-787 phosphorylation. Inhibition of Ser-787 phosphorylation could lead to reduced binding of viral spike protein with human ACE2 followed by reduced SARS-CoV-2 host-cell entry, replication and infection. This could lead to deactivation of ERK1/2 and CCL2 pathway thus resulting in reduced lung injury after coronavirus infection.

To add further confidence in our results, we predicted O-β-GlcNAcylation sites on 3D structure of proteins as well. We identified six potential accessible yin yang Ser/ Thr residues that could induce ACE2 conformational and functional changes upon phosphorylation and O-β-GlcNAc interplay. Interestingly, in our study, UDP-GlcNAc is found to be three times stronger than the phosphoryl group in binding with Ser/ Thr residues of ACE2. These results strengthened our hypothesis that UDP-GlcNAc might be able to inhibit Ser-787 phosphorylation of ACE2 to stop the viral infection spread in the body. The previous reports also found that SARS CoV virus was less efficient to infect mouse and rat species despite maximum homology in ACE2 structures (Li et al., 2004; Sheahan et al., 2020; Shereen et al., 2020; Wang et al., 2020). Based upon our results, we suggest this could be due to the mutation of Ser-787 in the mouse and rat with glycine as it is difficult for the glycine to be phosphorylated.

In our study, the sequence alignment of SARS-CoV-2 spike with SARS- CoV spike protein showed mutation at the residues Y442L, R426N and T487N may change SARS-CoV-2 binding

with the ACE2 receptor. The comparative analysis also revealed that the other mutated residues in the SARS-CoV-2 spike have same propensity of the side chain as found in the SARS-CoV (Fig 4). The binding position of UDP-GlcNAc in the RBD in the vicinity of Thr 487 may alter the conformation and thus inhibit the binding of the SARS-CoV-2 spike with the human ACE2 receptor to prevent the entry into the host cells (Fig5, table 2). Our results have also been supported by the previous studies that Ser481 and Thr487 of SARS-CoV were responsible to induce SARS-CoV binding with human ACE2 receptor (Li et al., 2005). Therefore, we can hypothesize based upon our results that O- β -GlcNAcylation in the RBD zone of SARS-CoV-2 spike complex with human ACE2 may disrupt the interaction of the two proteins and ultimately inhibit the viral entry into the host cells.

The other method to inhibit the interaction of SARS-CoV-2 with the ACE2 at the receptor binding domain and /or to inhibit the ACE2 phosphorylation at Ser-787 are by the compound's binding to the receptor ACE2. Therefore, in this study, we examine the role of chloroquine (CQ) and 2-hydroxybezohydrazide (HBH) as probable inhibitors against SARS infection. CQ is a well-known anti-malarial drug and has been proposed recently as being efficient against SARS-CoV-2 infection too (Liu et al., 2020). Meanwhile, HBH has been found to reduce influenza infection (Alam and Lee, 2016; Kundin et al., 1964). Interestingly, we found these inhibitors showed their strong binding with Ser-787 of the ACE2 protein, thus blocking the access of the phosphoryl group at the sites or inactivate phosphatases involved to block the viral infection. Furthermore, our results also showed that both CQ and HBH have strong binding with Glu35 of ACE2, a critical residue to bind with the SARS-CoV-2 RBD. Moreover, both CQ and HBH were also able to bind at the Glu471 residue of SARS-CoV-2 spike RBD, important for binding with ACE2. The interaction of CQ and HBH either with the ACE2 or the SARS-CoV2 may inhibit the viral entry into the host cells.

In conclusion, our results showed that: i) inhibition of the receptor binding domain (RBD) of SARS-CoV-2 and interacting binding pocket of human ACE2 receptor may prevent the virus entry to the host, and ii) inhibition of phosphorylation at Ser-787 of ACE2 protein may prevent the transmission of the virus in the COVID-19 patients. We highlighted the significant importance of the inhibition of ACE2 phosphorylation at Ser-787, and the inhibition of SARS-CoV-2 RBD and interacting binding pocket of human ACE2 receptor either by molecular level via O- β -GlcNAcylation or by the use of CQ and HBH inhibitors (Fig 8) to block the entry or downstream

signaling of SARS-CoV-2 infection. These findings suggest the potential targets to develop therapeutics to control the SAR-CoV-2 (2019-nCoV) infection.

Author Contributions

W. A was responsible for the study design. W.A, K. S and N. I collected and analyzed the data.

Declaration of interest

We author have no interests to declare.

Funding resources

N.A

Supplementary data

The supplementary data has been provided with the submitted manuscript.

5. References

- Ahmad, I., Hoessli, D.C., Walker-Nasir, E., Rafik, S.M., Shakoory, A.R., Nasir ud, D., 2006. Oct-2 DNA binding transcription factor: functional consequences of phosphorylation and glycosylation. *Nucleic acids research* 34, 175-184.
- Ahmad, W., Shabbiri, K., Ijaz, B., Asad, S., Nazar, N., Nazar, S., Fouzia, K., Kausar, H., Gull, S., Sarwar, M.T., Shahid, I., Hassan, S., 2011a. Serine 204 phosphorylation and O-beta-GlcNAC interplay of IGFBP-6 as therapeutic indicator to regulate IGF-II functions in viral mediated hepatocellular carcinoma. *Virol J* 8, 208.
- Ahmad, W., Shabbiri, K., Ijaz, B., Asad, S., Sarwar, M.T., Gull, S., Kausar, H., Fouzia, K., Shahid, I., Hassan, S., 2011b. Claudin-1 required for HCV virus entry has high potential for phosphorylation and O-glycosylation. *Virol J* 8, 229.
- Ahmad, W., Shabbiri, K., Nazar, N., Nazar, S., Qaiser, S., Shabbir Mughal, M.A., 2011c. Human linker histones: interplay between phosphorylation and O-beta-GlcNAc to mediate chromatin structural modifications. *Cell division* 6, 15.
- Alam, M.S., Lee, D.U., 2016. Synthesis, biological evaluation, drug-likeness, and in silico screening of novel benzylidene-hydrazone analogues as small molecule anticancer agents. *Arch Pharm Res* 39, 191-201.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* 25, 3389-3402.
- Audagnotto, M., Dal Peraro, M., 2017. Protein post-translational modifications: In silico prediction tools and molecular modeling. *Comput Struct Biotechnol J* 15, 307-319.
- Avital, A., Godfrey, S., Maayan, C., Diamant, Y., Springer, C., 1994. Chloroquine treatment of interstitial lung disease in children. *Pediatr Pulmonol* 18, 356-360.
- Blom, N., Gammeltoft, S., Brunak, S., 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of molecular biology* 294, 1351-1362.

- Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S., Brunak, S., 2004. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* 4, 1633-1649.
- Castrignano, T., De Meo, P.D., Cozzetto, D., Talamo, I.G., Tramontano, A., 2006. The PMDB Protein Model Database. *Nucleic acids research* 34, D306-309.
- Chauhan, J.S., Rao, A., Raghava, G.P., 2013. In silico platform for prediction of N-, O- and C-glycosites in eukaryotic protein sequences. *PLoS One* 8, e67008.
- Chen, I.Y., Chang, S.C., Wu, H.Y., Yu, T.C., Wei, W.C., Lin, S., Chien, C.L., Chang, M.F., 2010. Upregulation of the chemokine (C-C motif) ligand 2 via a severe acute respiratory syndrome coronavirus spike-ACE2 signaling pathway. *J Virol* 84, 7703-7712.
- Davis, I.W., Murray, L.W., Richardson, J.S., Richardson, D.C., 2004. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic acids research* 32, W615-619.
- Duan, G., Walther, D., 2015. The roles of post-translational modifications in the context of protein interaction networks. *PLoS Comput Biol* 11, e1004049.
- Grosdidier, A., Zoete, V., Michielin, O., 2011. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic acids research* 39, W270-277.
- Gupta, R., Brunak, S., 2002. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 310-322.
- Hollingsworth, S.A., Karplus, P.A., 2010. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomolecular concepts* 1, 271-283.
- Huang, H.D., Lee, T.Y., Tzeng, S.W., Horng, J.T., 2005. KinasePhos: a web tool for identifying protein kinase-specific phosphorylation sites. *Nucleic acids research* 33, W226-229.
- Kaleel, M., Torrisi, M., Mooney, C., Pollastri, G., 2019. PaleAle 5.0: prediction of protein relative solvent accessibility by deep learning. *Amino Acids* 51, 1289-1296.
- Kamemura, K., Hart, G.W., 2003. Dynamic interplay between O-glycosylation and O-phosphorylation of nucleocytoplasmic proteins: a new paradigm for metabolic control of signal transduction and transcription. *Progress in nucleic acid research and molecular biology* 73, 107-136.
- Kamemura, K., Hayes, B.K., Comer, F.I., Hart, G.W., 2002. Dynamic interplay between O-glycosylation and O-phosphorylation of nucleocytoplasmic proteins: alternative glycosylation/phosphorylation of THR-58, a known mutational hot spot of c-Myc in lymphomas, is regulated by mitogens. *The Journal of biological chemistry* 277, 19229-19235.
- Kao, H.J., Huang, C.H., Bretana, N.A., Lu, C.T., Huang, K.Y., Weng, S.L., Lee, T.Y., 2015. A two-layered machine learning method to identify protein O-GlcNAcylation sites with O-GlcNAc transferase substrate motifs. *BMC bioinformatics* 16 Suppl 18, S10.
- Keyaerts, E., Li, S., Vijgen, L., Rysman, E., Verbeeck, J., Van Ranst, M., Maes, P., 2009. Antiviral activity of chloroquine against human coronavirus OC43 infection in newborn mice. *Antimicrob Agents Chemother* 53, 3416-3421.
- Klausen, M.S., Jespersen, M.C., Nielsen, H., Jensen, K.K., Jurtz, V.I., Sonderby, C.K., Sommer, M.O.A., Winther, O., Nielsen, M., Petersen, B., Marcatili, P., 2019. NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning. *Proteins* 87, 520-527.
- Kundin, W.D., Robbins, M.L., Smith, P.K., 1964. Inhibitory activity of benzoyl hydrazides and hydrazine on the growth of influenza virus in chick embryo lung tissue culture. *Experientia* 20, 438-439.
- Leney, A.C., El Atmioui, D., Wu, W., Ovaa, H., Heck, A.J.R., 2017. Elucidating crosstalk mechanisms between phosphorylation and O-GlcNAcylation. *Proceedings of the National Academy of Sciences of the United States of America* 114, E7255-E7261.
- Li, F., 2016. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol* 3, 237-261.
- Li, F., Li, W., Farzan, M., Harrison, S.C., 2005. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* 309, 1864-1868.

- Li, W., Greenough, T.C., Moore, M.J., Vasilieva, N., Somasundaran, M., Sullivan, J.L., Farzan, M., Choe, H., 2004. Efficient replication of severe acute respiratory syndrome coronavirus in mouse cells is limited by murine angiotensin-converting enzyme 2. *J Virol* 78, 11429-11433.
- Li, Y., Xie, M., Men, L., Du, J., 2019. O-GlcNAcylation in immunity and inflammation: An intricate system (Review). *Int J Mol Med* 44, 363-374.
- Liu, J., Cao, R., Xu, M., Wang, X., Zhang, H., Hu, H., Li, Y., Hu, Z., Zhong, W., Wang, M., 2020. Hydroxychloroquine, a less toxic derivative of chloroquine, is effective in inhibiting SARS-CoV-2 infection in vitro. *Cell Discov* 6, 16.
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., Chen, J., Meng, Y., Wang, J., Lin, Y., Yuan, J., Xie, Z., Ma, J., Liu, W.J., Wang, D., Xu, W., Holmes, E.C., Gao, G.F., Wu, G., Chen, W., Shi, W., Tan, W., 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395, 565-574.
- Peiris, J.S., Guan, Y., Yuen, K.Y., 2004. Severe acute respiratory syndrome. *Nat Med* 10, S88-97.
- Qian, N., Sejnowski, T.J., 1988. Predicting the secondary structure of globular proteins using neural network models. *Journal of molecular biology* 202, 865-884.
- Saraswathi, S., Fernandez-Martinez, J.L., Kolinski, A., Jernigan, R.L., Kloczkowski, A., 2012. Fast learning optimized prediction methodology (FLOPRED) for protein secondary structure prediction. *Journal of molecular modeling* 18, 4275-4289.
- Shang, J., Wan, Y., Liu, C., Yount, B., Gully, K., Yang, Y., Auerbach, A., Peng, G., Baric, R., Li, F., 2020. Structure of mouse coronavirus spike protein complexed with receptor reveals mechanism for viral entry. *PLoS Pathog* 16, e1008392.
- Sheahan, T.P., Sims, A.C., Leist, S.R., Schafer, A., Won, J., Brown, A.J., Montgomery, S.A., Hogg, A., Babusis, D., Clarke, M.O., Spahn, J.E., Bauer, L., Sellers, S., Porter, D., Feng, J.Y., Cihlar, T., Jordan, R., Denison, M.R., Baric, R.S., 2020. Comparative therapeutic efficacy of remdesivir and combination lopinavir, ritonavir, and interferon beta against MERS-CoV. *Nat Commun* 11, 222.
- Shereen, M.A., Khan, S., Kazmi, A., Bashir, N., Siddique, R., 2020. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *J Adv Res* 24, 91-98.
- Song, W., Gui, M., Wang, X., Xiang, Y., 2018. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathog* 14, e1007236.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* 22, 4673-4680.
- van der Laarse, S.A.M., Leney, A.C., Heck, A.J.R., 2018. Crosstalk between phosphorylation and O-GlcNAcylation: friend or foe. *FEBS J* 285, 3152-3167.
- Wang, J., Cao, H., Zhang, J.Z.H., Qi, Y., 2018. Computational Protein Design with Deep Learning Neural Networks. *Sci Rep* 8, 6349.
- Wang, Z., Chen, X., Lu, Y., Chen, F., Zhang, W., 2020. Clinical characteristics and therapeutic procedure for four cases with 2019 novel coronavirus pneumonia receiving combined Chinese and Western medicine treatment. *Biosci Trends* 14, 64-68.
- Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.L., Abiona, O., Graham, B.S., McLellan, J.S., 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367, 1260-1263.
- Wu, C.H., Apweiler, R., Bairoch, A., Natale, D.A., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Mazumder, R., O'Donovan, C., Redaschi, N., Suzek, B., 2006. The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic acids research* 34, D187-191.
- Xin, F., Radivojac, P., 2012. Post-translational modifications induce significant yet not extreme changes to protein structure. *Bioinformatics* 28, 2905-2913.

- Zachara, N., Akimoto, Y., Hart, G.W., 2015. The O-GlcNAc Modification, in: rd, Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G.W., Aebi, M., Darvill, A.G., Kinoshita, T., Packer, N.H., Prestegard, J.H., Schnaar, R.L., Seeberger, P.H. (Eds.), *Essentials of Glycobiology*, Cold Spring Harbor (NY), pp. 239-251.
- Zachara, N.E., Hart, G.W., 2002. The emerging significance of O-GlcNAc in cellular regulation. *Chemical reviews* 102, 431-438.
- Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. *BMC bioinformatics* 9, 40.

Table 1: Phosphorylation, kinases involved, O- β -GlcNAc, surface accessibility and predicted YinYang residues for Ser/ Thr/ Tyr residues of human ACE2. Abbreviations: PS= residue position, CS= conservation status, DP= DisPhos, NP= NetPhos, KP= kinasePhos, OGT= OGTSITE, GEP= GlycoEP, YY= , YinOYang, PA= PaleAle

Substrate			Phosphorylation			Kinases involved		O- β -GlcNAc status			Surface Accessibility		YinYang Predictions	
Res	PS	CS	DP	NP	KP	NP	KP	OGT	GEP	YY	PA	Area	3D	True
Ser	2	*	Y	Y	-	CDC2, PKA, GSK3	IKK	-	Y	Y	E	92.3		
	3	.	Y	Y	-	PKA, CDC2, CAMII		-	Y	Y	E	52.0		
	4	*		Y	-	PKA, CDC2, CAMII			Y	Y	B	3.9		
	5	-		-	-	CDC2, CAMII, GSK3			Y		E	41.3		
	10	*		Y	-	CDC2, PKA, GSK3	MAPK		Y		B	18.2		
	19	*		Y	Y	UNSP, CKII, PKC	CKII, PKG				E	91.6		
	43	*		Y		CKII, CAMII, GSK3	CDC2				B	9.2		
	44	*		Y		DNAPK, PKA					B	14.1		
	47	*		Y		PKC, CAMII, GSK3					E	29.3	Y	
	70	*		Y		CDC2, CAMII, GSK3					E	53.7	Y	Y
	77	*		-		CAMII, GSK3, CKII	CKI				B	6.8		
	105	*		Y		CDC2, GSK3, CAMII	CDC2			Y	E	60.1		
	106	-		Y		CDC2, CAMII, GSK3	ATM		Y		E	59.7		
	109	*		Y		UNSP, PKC, CKII	CAMII, CKI				E	63.2		
	113	.		Y		UNSP, PKC, GSK3	PKA, CKI				E	31.1		
	124	*		-		GSK3, CAMII, CKI		Y	Y		E	37.1		
	128	*		Y		PKC, CKI, GSK3		Y	Y		E	41.9		
	155	*		Y		CDC2, CKII, DNAPK					B	10.1		
	167	.		Y		PKC, CDC2, CAMII					B	1.2		
	170	..		-		CDC2, GSK3, CAMII					E	31.6		
	218	.		Y		CKI, PKA, GSK3	CAMII	Y			E	64.1	Y	Y
	254	-		Y		UNSP, PKC, GSK3	PKC				E	91.9		
	257	*		Y		CDK5, GSK3, CDC2	CDC2, CDK				B	19.6		
	280	.		-		CDC2, GSK3, CAMII	PKA			Y	B	9.4		
	317	*		Y		PKA, CAMII, GSK3		Y			E	40.3		

33	1	*		-		CDC2, CAMII, CKII					B	17.8		
40	9	*		Y		DNAPK, CAMII, GSK3	MAPK	Y	Y		B	19.1		
41	1	*		-		CAMII, GSK3, CDC2	PKG, ATM			Y	B	3.8		
42	0	..		Y		PKA, CAMII, GSK3	PKA, CAMII	Y			E	92.6		
42	5	.		Y		UNSP, P38MAPK, CKII	CAMII, CDC2			Y	E	51.5		
50	2	*		-		GSK3, CAMII, CDC2			Y		B	0.6		
50	7	..		-		CKII, CAMII, GSK3			Y	Y	B	0.2		
51	1	*		-		PKG, CDC2, GSK3			Y		B	11.9	Y	
54	5	*		Y		PKA, CDC2, CAMII	PKG	Y			E	29.1		
54	7	*		Y		UNSP, CDC2, CKII	PKG				B	19.3		
56	3	*		Y		CDC2, PKA, GSK3	ATM	Y			B	20.7		
60	2	*		Y		CDC2-CAMII, GSK3	PKA				E	48.3	Y	Y
60	7	.		-		GSK3, CAMII, CKII	IKK		Y		E	27.1		
61	1	*		-		GSK3, CDK5, CAMII			Y		B	19.4	Y	
61	7	*		Y		PKC, CDC2, UNSP			Y		E	56.4	Y	
62	3	*		Y	Y	UNSP, PKA, PKC	PKG, PKC	Y			B	12.5		
62	6	*		Y		UNSP, PKC, CAMII			Y		E	74.4	Y	
64	5	*		-		CDC2, CAMII, GSK3					E	51.5		
64	6	*		Y		PKA, PKC, CAMII	PKA				E	30.1	Y	Y
68	0	*		Y		PKA, CAMII, RSK	OTHER		Y		B	24.1		
69	2	*		-		CAMII, GSK3, PKA		Y	Y		E	61.4		
70	7	*	Y	Y	Y	UNSP, PKA, CDC2	PKA, PKC	Y			B	0.4		
70	9	.	Y	Y	Y	UNSP, PKA, GSK3	PKG				E	85.8		
72	1	*		Y		UNSP, DNAPK, GSK3					E	94		
74	0	..		-		CAMII, GSK3, CDC2		Y	Y	Y	B	15.9		
77	6	.	Y	Y		UNSP, CDC2, GSK3	CKII, PKG				B	17.8		
78	3	*	Y	-		CAMII, GSK3, CDC2	PKA	Y			E	35.1		
78	7	.		Y		UNSP, CKII, GSK3	CKII, ATM				E	48.4	Y	Y
80	4	*		Y		CKI, CAMII, CKII		Y			E	54.1		
15		-		CAMII, GSK3, PKC		Y		Y	E	37.9		

Th r	20	*		-		CKII, CKI, UNSP	CKII, PKG		Y		E	90.1		
	27	*		-		CDC2, PKC, GSK3					E	89.2		
	52	*		Y		CKII, CAMII, GSK3					B	17.1		
	55	.		-		GSK3, CKII, CAMII			Y		E	49.1		
	78	.		-		CAMII, DNAPK, GSK3	OTHER	Y			B	16.8		
	92	.		Y		PKC, CAMII, P38MAPK					E	68.1		
	118	..		Y		UNSP, PKA, CAMII					E	62.3		
	122	..		Y		CAMII, GSK3, CDC2		Y			E	53.1		
	125	..		Y		UNSP, PKC, CAMII		Y	Y		E	82.5		
	129	*		Y		PKC, CDC2, CAMII	CDK		Y		B	2.5		
	229	.		Y		CKII, CKI, GSK3					B	2.8		
	276	*		Y		PKA, CAMII, GSK3					E	35.2	Y	
	282	*		Y		PKC, GSK3, CAMII	PKC				B	0		
	294	*		-		CKII, GSK3, CAMII	PKA	Y	Y	Y	E	32.9		
	324	*		Y		DNAPK, CDC2, GSK3		Y	Y		E	52.9		
	334	*		-		CAMII, CKI, CDC2			Y		E	72.9		
	347	*		-		GSK3, CAMII, PKG			Y		E	40.1		
	362	*		-		GSK3, CAMII, CDC2		Y			B	10.5		
	365	*		-	Y	PKG, PKC, CAMII	PKA	Y			E	46.9		
	371	*		-		CKII, CAMII, GSK3			Y		E	44.7	Y	
	414	*	Y	Y	Y	UNSP, P38MAPK, GSK3	MAPK				B	2.4		
	434	*		Y		CKII, GSK3, CAMII					E	27.1		
	445	.		Y		UNSP, CAMII, CDC2		Y	Y		E	38.9		
	449	*		-		GSK3, DNAPK, CDC2		Y			B	12.8		
	453	*		-		CKII, CDC2, GSK3		Y			B	4.4		
	496	*	Y	Y		UNSP, GSK3, CAMII	OTHER		Y		E	82.7		
	517	*		Y		PKG, PKA, CAMII		Y			B	0		
	519	*		Y		PKC, UNSP, DNAPK					B	6.2		
	548	*		Y		PKC, CAMII, CKII	PKG				E	93.1		
	567	*		-		GSK3, CAMII, CKII					B	0.4		

	593	.		Y		PKC, CDC2, GSK3					E	53.1		
	608	*		-		CDC2, GSK3, CAMII					E	60.8	Y	Y
	686	*		Y		PKC, CAMII, GSK3		Y	Y		E	25.3		
	698	.	Y	Y		CKII, UNSP, GSK3	CDK	Y			E	82.6		
	730	*		-		GSK3, CAMII, CDC2			Y		E	63.6		
	763	*		Y		PKC, GSK3, CAMII					E	52.8		
	798	.		-		CKII, CAMII, GSK3			Y		E	52.4		
	803	*	Y	Y		UNSP, CKII, GSK3		Y	Y	Y	E	72.1		

Table 2: Docking characteristics of UDP-GlcNAc, Chloroquine, and 2-hydroxybenzohydrazide with ACE2 protein and receptor binding domain of SARS-CoV-2 (SCoV-2) and ACE2 complex

<i>Docking characteristics of ligands with ACE2</i>					
Ligand	Cluster	No. of hydrogen bonds	Residues involved	Bond length (Å)	Binding energy (kcal/mol)
UDP-GlcNAc	27	2	Ser-70 Gln-102	2.789 2.128	-7.821
	33	2	Ser-218 Gln-102	2.187 2.111	-7.741
	25	4	Ser-602 Phe-805 Lys-788 Hsd-239	2.335 1.621 2.371 2.106	-9.431
	20	5	Thr-608 Thr-608 Lys-476 Arg-652 Gln-653	2.245 1.964 2.251 2.085 2.398	-9.491
	32	4	Ser-646 Thr-608 Tyr-649 Glu-479	1.922 1.918 1.791 2.519	-8.971
	8	5	Ser-787 Asp-699 Val-604 Glu-238 Glu-238	2.085 1.818 2.221 1.840 2.341	-8.729
Chloroquine	22	2	Ser-787 Hsd-239	2.702 2.831	-7.111
2-hydroxybenzo hydrazide	-	2	Ser-787 Lys-788	2.385 2.108	-5.300
<i>Docking characteristics of ligands with receptor binding domain of SARS-CoV-2 and ACE2 complex</i>					
UDP-GlcNAc	28	4	Gly-433 (SCoV-2) Asn-49 (ACE2) Asn-61 (ACE2) Lys-68 (ACE2)	1.788 1.935 2.574 2.160	-9.001
Chloroquine	15	1	Glu-35 (ACE2)	1.869	-8.581
	19	1	Glu-471 (SCoV-2)	1.809	-8.441
2-hydroxybenzo hydrazide	-	2	Glu-35 (ACE2) Glu-471 (SCoV-2)	2.318 1.997	-6.301
	-	1	Glu-35 (ACE2)	2.248	-6.181
	-	1	Val-432 (SCoV-2)	2.571	-5.298

Supplementary material

Fig S1: Homology models of human ACE2 retrieved from I-TASSER. ACE2 is 805AA protein. Five ACE2 models were predicted as possible complete ACE2 3D structure. All the models contain already experimentally determined 3D template 1R42 (1-615AA) with predicted templates (616-805AA). Best model used in this study was selected on the basis of Ramachandran plots, geometrical configurations and C-score. Model 1 was predicted as the best model in this study. (A) Comparative analysis of geometrical configuration of each model compared with the already available templates. (B) Ramachandran plot and 3D structure of template 1R42_ACE2. (C-G) Ramachandran plots and 3D structures of predicted five models. UCSF Chimera was used to visualize the models. All models represented in rainbow view.

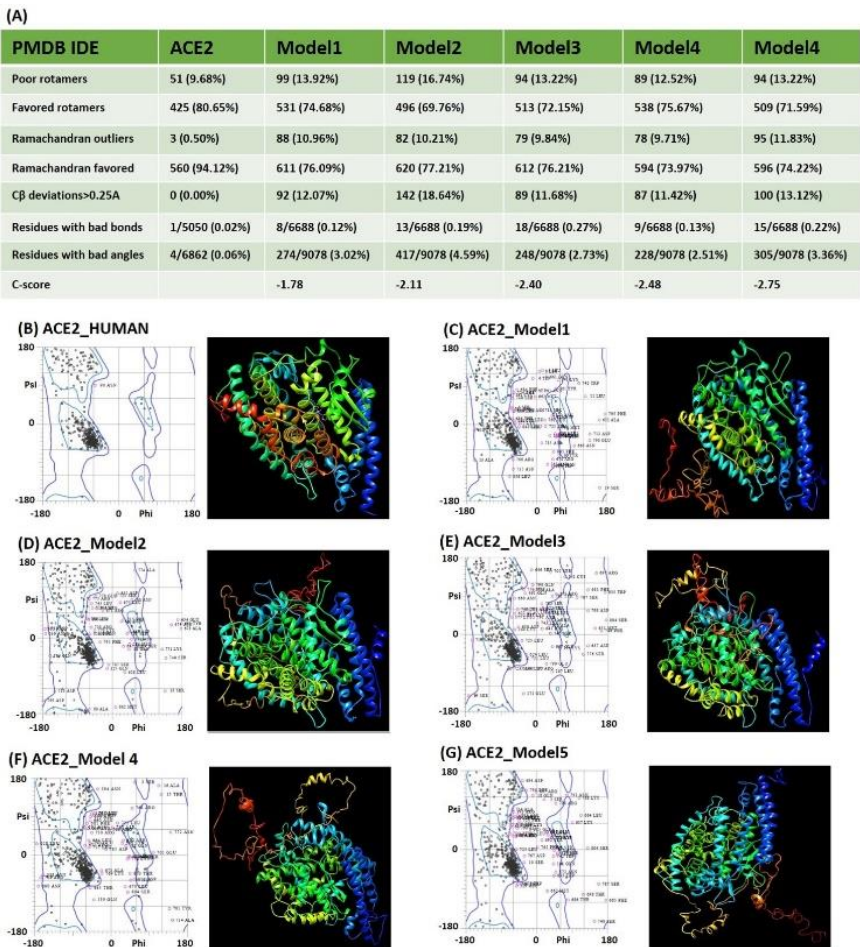


Fig S2

Homology modeling, sequence alignments and characterization of SARS-CoV-2 RDB. Homology models of RDB of SARS-CoV-2 predicted by swissdock. Two models were selected as possible complete structure of SARS-CoV-2 RDB that contained already experimentally determined templates (6VSB and 2AJF) of RDB. Model 1 was selected as the best model. (A) Predicted models and their alignments with already known available templates. (B) Comparative analysis of geometrical configurations of each model with already known available templates. (C) Ramachandran plots of already known SARS-CoV and predicted structures for SARS-CoV-2 RDB. (D) Sequence alignment of RDB of SARS-CoV-2 with SARS-CoV. Highlighted residues: green= conserved Ser/ Thr between two sequences, blue= conserved substitution, pink= new Ser/ Thr residues introduced in SARS-CoV-2, sea green= Ser/ Thr residues in SARS-CoV replaced by other amino acids in SARS-CoV-2, yellow: critical residues in binding of SARS-CoV with ACE2, red= critical residues in binding of SARS-CoV-2 with ACE2. (EI-III) Detailed binding interface of predicted SARS-CoV-2 RDB with ACE2 N- terminal.

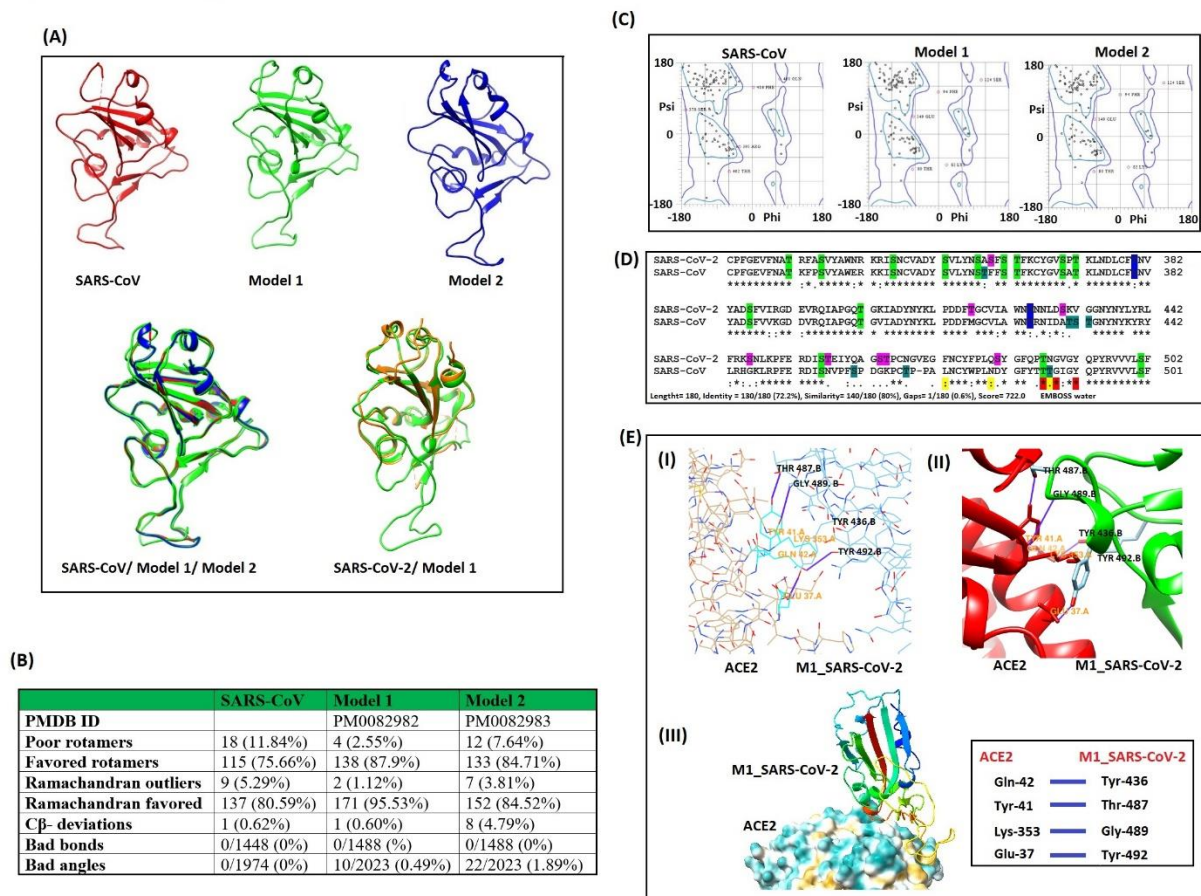



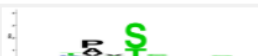




Fig S3

Predicted O-GlcNAcylation sites on SARS-CoV-2 RBD

Download Result		Input Information		
Specificity level		Medium		
Input ID		COVID		
Input Sequence		CPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKCYGVSPFT... ▶▶		
Predict Result				
Protein Name	Locations	O-GlcNAcylation Sites	Substrate Motifs	Score
COVID	24	NRKRI S NCVAD		0.4218
COVID	48	KCYGV S PTKLN		0.553
COVID	58	NDLCF T NVYAD		0.35
COVID	142	IYQAG S TPCNG		0.5608
COVID	143	YQAGS T PCNGV		0.5602
COVID	165	YGFQP T NGVGY		0.495
- top -				