Scouting the Receptor Binding Domain of COVID-19: A Comprehensive Immunoinformatics Inquisition

CURRENT STATUS: POSTED

Zaira Rehman
National University of Sciences and Technology, Islamabad, Pakistan

Ammad Fahim
National University of Medical Sciences

ammad.fahim@numspak.edu.pk Corresponding Author

Muhammad Faraz Bhatti
National University of Sciences and Technology, Islamabad, Pakistan

mfbhatti@asab.nust.edu.pk Corresponding Author

DOI: 10.21203/rs.3.rs-25501/v1

SUBJECT AREAS
Virology

KEYWORDS
COVID-19, Coronaviruses, Epitope mapping, Human ACE2 receptor
Abstract
The December of 2019 witnessed emergence of worldwide outbreak by a novel strain of coronavirus termed COVID-19 with sequence similarity of overall 96.2% with BatCoV RaTG13 (coronavirus isolated from bat) and 94% sequence identity with Severe Acute respiratory syndrome Virus (SARS-CoV) that resulted in outbreak in 2002-2003. There is no therapeutic or preventive strategy like vaccine developed so far to overcome infection.

The receptor binding domain (RBD) of COVID-19 for any potential vaccine epitopes were explored. The structure of RBD of COVID-19, BatCoV RaTG13 and bACE2 were chalked through homology modeling followed by molecular docking and structural validation. A comprehensive immunoinformatics approach mapped conserved peptide sequence on COVID-19 RBD for their B-, Helper T- & Cytotoxic T-cell epitope profile. The recognized epitopes were further studied and validated for their docking interaction with MHC-I and MHC-II alleles.

Through immune-informatics approaches the study identified conserved B- and T-cell epitopes in RBD. The B-cell epitopes lying within the receptor binding motif, LFRKSN and SYGFQPT I were found to be highly antigenic. Among T-cell epitopes, the epitope CVADYSVLY and FTNVYADSF were antigenic and exhibited affinity for maximum number of MHC-I alleles. The T cell epitopes YRLFRKSNL, YYAWNRKRI displayed affinity for maximum number of MHC-II alleles. The docking analysis of the epitopes with MHC proteins revealed strong interactions of T-cell epitopes with MHC-I and MHC-II alleles. The overlapping epitope among B- and T-cells was YRLFRKSNL. The deployment of these epitopes in potential vaccine against COVID-19 may help in sweeping the COVID-19 infectious spread.

Introduction:
The ongoing epidemic of novel coronaviruses named by WHO as COVID-19 also known as SARS CoV2[1], has recalled the memories of Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) epidemic [2,3]. Before the emergence of SARS epidemic, coronaviruses were not a notable pathogenic entity for causing epidemics or pandemics, although they have been noted in the literature since 1960’s [4]. The SARS coronavirus (CoV) and MERS-CoV were other most
famous coronaviral strains known for crossing the species barrier and infecting humans, while 229E, NL63(alpha), OC43 and HKU1 (beta) strains of coronaviruses caused common cold in humans [5]. The SARS-CoV which surfaced during 2002-2003 in China, infected almost 8000 individuals and caused 774 causalities in 37 countries [6,7], while MERS-CoV emerged in 2012 and responsible for infecting 2494 humans and 858 deaths [8,9,2]. The ongoing epidemic of COVID-19 has affected more than 1.9 million patients with 121787 deaths till 14th April 2020 [10]. The detailed viral sequence of COVID-19 points it to contain a similar genetic organization with SARS and MERS, harboring six ORFs, and belongs to the same family as that of SARS [11]. The genome of coronaviruses spans from 27-35 kb, stuffed inside a nucleocapsid protein (N) enclosed by envelope. There are 3 structural proteins associated with viral envelope: membrane (M), envelope (E) and spike (S) proteins [12]. The S proteins is vitally imperative for viral entry but also impel tissue tropism, host range as well as host immune responses [13]. The S protein constitutes S1 and S2 subunits where S1 harbors receptor binding domain (RBD) while S2 aids in virus-cell fusion. Interestingly, the S gene of coronaviruses have been noticeable for the most recurring recombination breakpoints in SARS-CoV [14]. In case of SARS CoV, the main entry receptor is angiotensin-converting enzyme2 (ACE2) [15] while in case of MERS-CoV, the primary entry receptor is dipeptidyl peptidase 4 (DPP4, also known as CD26) [16]. Comparatively, the genome of SARS-CoV and MERS-CoV although harbors significant similarity, they potentially harbors variation, particularly from the antigenic response they might generate with reference to RBD [17].

Receptor binding protein have been previously targeted for vaccinable solutions for preventing full blown infections [18,19]. However, from the evolutionary standpoint, coronaviral RBD is reported as a hypervariable region [20]. There is no approved treatment or vaccine available to tackle COVID-19 thus far. Contextually, vaccinomics approaches by delving into immunogenomics of COVID-19 RBD can lead to potential vaccine candidates. Immunoinformatics driven in-depth analysis can aid in identifying repertoire of viral antigenic epitopes with linear discontinuous profiles, about their virulence and immunogenicity [21]. A comprehensive immunoinformatics data mining of COVID-19 RBD can increase our understanding for its antigenic profile. Encouragingly, epitope prediction
analysis were reported earlier for SARS and MERS strains [22,19] and quite recently for COVID-19 [23-25].

Numerous studies have reported Bats as the primary reservoirs for SARS and MERS viruses [26-28,14,29], however rodent origin has also reported [13,30]. A recently reported strain derived from bats, the BatCoV RaTG13 contains more than 96% homology with COVID-19 and more than 93% homology with its S protein, rendering it to be most closely related to COVID-19 reportedly till date [11]. Anatomically, conservative peptide sequence of coronaviral RBD compared with the closest known zoonotic coronaviral strain can provide better potential vaccine candidates for human testing.

After the emergence of COVID-19 pandemic, SARS CoV surface protein has been repeatedly utilized for identification of potential vaccine epitopes for COVID-19 [24,23,25]. However, there has also been simultaneous speculation regarding potential existence of cross resistance epitopes between SARS CoV2 and SARS CoV [17,31,32]. Earlier, the computational probing of protein structures for respiratory infections by employment of docking methods have added useful information regarding stereochemical properties, virus binding mediated host receptor conformational transformation and binding preferences [33-37]. Contextually, viral receptor interactions were considered valuable in the instances of picornaviruses, influenza, HIV and coronaviruses [38-40,20]. For the current COVID-19 induced infection, the basic reproduction number for viral transmissibility ($R_0$) as per various estimates is around 1.1-5.5 [41,42]. Since this threshold points towards very high infectivity rate, it is pertinent to target the viral binding regions with vaccines to prevent infection.

Vaccinable peptide sequence for epitope based vaccine in case of Alphaviruses, Hepatitis B & C, HIV, HPV and Influenza viruses for recognition of potent immunogens have shown propitious results [43-46]. Several studies reported on SARS and MERS CoV strains provided useful information regarding the potential epitopes retained by these strains [47-54]. While, the data in context pertinent to COVID-19 is insufficient. The global COVID-19 pandemic has sparked rigorous R&D activity for COVID19 vaccine development, and in a matter of just more than 4 months, various potential vaccine candidate are in the preclinical and clinical development phase [55]. However, the clinical behavior of COVID-19 infecting people around the world but with varied clinical symptomatology, ranging from
completely asymptomatic to rapidly progressing lethal respiratory insufficiency demands for utilization of more and rapid novel technology platforms with more vaccinable options against COVID-19 [55,56].

Therefore, this study reports key findings regarding COVID-19 RBD for its variable and conservative RBD residues in comparison with BatCoV RaTG13 strain, which can be considered as immunogenic epitopes for potential multi epitope vaccine candidate for COVID-19 in the backdrop of its binding orientation.

Materials And Methods:
To identify the presence of antigenic epitopes within the RBD of spike glycoprotein of COVID-19, in-silico analysis was performed (Figure I). The antigenicity of RBD was determined through Vaxijen v2.0 [57].

Sequence Retrieval and Multiple Sequence Alignment:
Protein sequences of spike glycoprotein of COVID-19 (reported till 31st March, 2020) were retrieved from National Center for Biotechnology Information (NCBI). The sequences from China, Australia, USA, Taiwan, India, Pakistan, Nepal, Italy, Sweden, Brazil, Vietnam, Spain, Colombia, Peru and Japan were selected for analysis. The sequence of spike glycoprotein from bat coronavirus (BatCoV RaTG13) (GISAID accession no. EPI_ISL_402131) and SARS coronavirus ZJ02 (Accession No. ABB29898) were used as reference for comparison. Sequence analysis was performed to ascertain the changes in the receptor binding domain of spike glycoprotein. Multiple sequence alignment was performed using Clustal X. The consensus sequence of COVID-19 was used as input for epitope prediction.

B-cell Epitope Prediction:
For identification of B-cell epitopes, Immune Epitope Database (IEDB) and BepiPred-2.0 were used [58,59]. The consensus sequence of COVID-19 was used to predict B-cell epitopes through Karplus & Schulz Flexibility, Kolaskar & Tongaonkar Antigenicity, Chou & Fasman Beta-Turn, Emini Surface Accessibility and BepiPred Linear Epitope Prediction methods. Kolaskar & Tongaonkar method predicted the antigenicity on the basis of amino acid abundances in naturally occurring epitopes as well as their physicochemical properties. The default threshold was set to 1.00 for antigenicity
determination [60]. Emini surface accessibility method predicts the surface accessibility of epitopes as the surface accessible peptides recognized by the immune system [61]. Chou & Fasman Beta-Turn method was used to predict the antigenic regions exhibiting beta turn as the beta turns are usually hydrophilic in nature and highly accessible [62]. Karplus & Schulz Flexibility method predicts those antigens that are exhibiting flexible amino acids in nature as flexibility is correlated with antigenicity [63]. Bepipred prediction method based on Hidden Markov model predict linear epitopes in protein [64]. The B-cell epitopes were also predicted using ElliPro. This method identifies linear and discontinuous epitopes in protein structure [65]. **T-cell Epitope Prediction:**

In vaccine development, cytotoxic T-lymphocyte (CTL) epitopes play an important role. Hence the T-cell epitopes were identified that have the ability to bind with major histocompatibility complex class I (MHC-I) and class II (MHC-II). CTL epitopes were identified through NetCTL 1.2 server [66]. The Immune epitope database (IEDB) and NetMHC 4.0 server was used to predict the binding of epitopes with MHC-I. NetMHC 4.0 predicts the binding affinity through artificial neural network (ANN) by schooling 81 distinct HLA-A, -B, -C and -E human MHC alleles [67]. T-helper cell epitopes were predicted through IEDB and NetMHCII 2.3 server [68]. The epitopes were predicted having high affinity towards HLA-DR, -DQ, and –DP. For all the T-cell epitopes the threshold for predicting strong binding affinity with MHC-I and II was set to be 500nM.

**B- and T-cell Epitopes Feature Profiling:**

B- and T-cell epitopes were further scrutinized for their enzyme digestion, toxicity, hydrophobicity, and physiochemical properties. The digestion of peptides with enzymes is an important parameter in vaccine development as the peptides that are digested by many enzymes are usually rendered unstable. Hence the digestion of peptides by different enzymes was predicted through protein digest server. AntiangioPred was used to predict the mutation and other physicochemical properties of peptides. ClanTOX predicted toxicity of peptides [69]. Antigenicity of peptides were predicted through Immunomedicine group server. For a peptide to be antigenic the threshold is 1.0 [60].

**Human proteome analysis for nonhuman homologues:**

To avert autoimmunity, vaccine contenders were screened for human and nonhuman homologues.
The nonhuman homologues were yielded by scrutinizing selected epitopes sharing <30% identity with human proteome, via BLASTp analysis.

**Docking of T-cell epitopes with MHC-I and MHC-II Alleles:**

The peptides that were showing affinity for maximum number of MHC-I and MHC-II alleles were selected for interaction analysis. The structure of peptides were modelled through PEPFOLD server [70] followed by energy minimization. In case of MHC-I, the common allele between the peptide was selected for docking. Hence, the crystal structure of human HLA-A*0101 was downloaded from pdb (PDB ID: 6AT9: resolution= 2.9 Å). Same criterion was followed for MHC-II alleles and for that the crystal structure of HLA-DRB5*01:01 (PDB ID: 1FV1: resolution= 1.9 Å) as also downloaded from pdb. Both the structures after ligand removal underwent protonation followed by energy minimization by AMBER99 force field. In order to analyze the inhibitory potential of peptides the docking of MHC-I specific peptides were executed with the HLA-A*0101 alleles while the MHC-II specific peptides were docked with HLA-DRB5_0101. The docking studies were accomplished with induced fit docking protocol implemented in Molecular Operating Environment version 2016.08. By using, Triangle Match as placement method with London-dG scoring and GBVI/WSA dG re-scoring function, 50 poses of each peptide were generated. Interaction analysis was done using Molecular Operating Environment (MOE) version 2016.08 and PyMOL.

**Interaction analysis of BatCoV RaTG13 with bACE2:**

After identification of vaccine epitopes, we further explored whether these vaccine epitopes harbor any important residues that involved in binding of COVID-19 with hACE2 and BatCoV RaTG13 with bACE2. The interactions of COVID-19 with hACE2 has recently been reported by Lan et al., 2020 [71]. The interaction analysis of BatCoV RaTG13 with bACE2 was performed in the current study. To our knowledge, the structure of RBD of BatCoV RaTG13 as well as bACE2 has not been determined yet. Hence the structure of both the proteins was determined through homology modeling using Modeller V9.23. BatCoV RaTG13 was modeled using SARS-COV as template (pdb ID: 2GHV). While bACE2 was modeled using Human Angiotensin Converting Enzyme (pdb ID: 1R42) as template. The generated model was subjected to model evaluation and structural validation via Ramachandran plot, PROSA,
ERRAT, Qmean, and MolProbity. Ramachandran plot calculates the presence of amino acid residues in allowed, favored and outlier regions on the basis of torsional angles (\(\Psi\) and \(\Phi\)) of amino acids [72]. PROSA reveal the quality of model by estimating any error in the models. It also calculates the score of model on the basis of experimentally reported (X-ray, and NMR) structures of proteins [73]. Qmean (Qualitative Model Energy Analysis) apprises the geometry of protein structure by measuring the torsion angles on three consecutive amino acid residues [74]. MolProbity evaluates the protein structure by assessing its geometry [75]. ERRAT gauges the quality of model by analyzing the statistics between non-bonded interactions and different type of atoms and compared these values with the extremely refined structures [76]. The best model was then subjected to energy minimization using AMBER99 forcefield implemented in Molecular Operating Environment (MOE) version 2015.10. Docking of BatCoV RaTG13:bACE2 were performed using HADDOCK web server [77]. Analysis of protein-protein interactions were performed through pdbSum [78] and PyMOL v2.3.

Results:

**Multiple Sequence Alignment:**

Receptor binding domain of COVID-19 is 192 amino acids long (within position 330-522 amino acids) lying in S1 region of spike glycoprotein. When comparing the receptor binding motif with the BatCoV RaTG13 there was variation between the two virus strains (Fig II). The important changes were observed at position 439 (Lys→Asn), 440 (His→Asn), 441 (Ile→Leu), 443 (Thr→Lys), 445(Glu→Val), 449 (Phe→Tyr), 459 (Ala→Ser), 478 (Lys→Thr), 483 (Gln→Val), 484 (Thr→Glu), 486 (Leu→Phe), 490 (Tyr→Phe), 492 (Ile→Leu), 493 (Tyr→Gln), 494 (Arg→Ser), 498 (Tyr→Gln), 501 (Asp→Asn), 505 (His→Tyr). The changes at positions 441, 486, 492, 493, 498, and 505 may not have any obvious effect on binding due to similar properties of amino acids.

**B-Cell epitopes within Receptor Binding Domain:** Continuous B-cell epitopes were predicted using B-cell epitope prediction methods on IEDB server. The Kolaskar & Tongaonkar method predicted 11 antigenic epitopes in the receptor binding domain (Table I) which can prompt B-cell responses. Surface accessibility analysis revealed 4 epitopes with surface accessibility (Table II).
Flexibility of epitopes is a measure of antigenicity [53]. The flexible epitopes in RBD were at positions 352-362, 380-392, 400-410, 421-433, 434-451, 454-473, 472-487, 495-506. Beta turns are the more flexible regions of the protein. According to Chou & Fasman predictions the beta turn epitopes were at positions 437-443, 455-468, 422-428, 495-500. Linear epitopes determined by Bepipred method are shown in Table 3. On the basis of consensus of all the methods, the peptides that can induce B-cell response were placed at positions 423-428, 455-461, 494-500. The 423-428, 455-461, 494-500, were the peptides that may prompt B-cell responses as predicted by ElliPro method. The mapping of epitopes on three dimensional structure of protein is shown in Fig III.

Cytotoxic T-cell Epitope Prediction:

The default setting in the NetCTL server was used to predict T cell epitopes. On the basis of highest combinatorial scores, five epitopes were opted for subsequent analysis (Table IV). On the basis of NetCTL scores the peptides with the highest score was CVADYSVLY. Further analysis of all the five peptides for their binding with MHC-I showed that peptide CVADYSVLY illustrated binding with maximum MHC-I alleles (HLA-A*26:01, HLA-A*01:01, HLA-A*30:02, HLA-B*35:01, HLA-A*11:01, HLA-B*15:01, HLA-A*68:01, HLA-A*03:01, HLA-B*53:01, HLA-C*07:01). The next peptide showing the binding with maximum number of alleles was FTNVYADSF.

Helper T-cell Epitope Prediction:

A total of 9 peptides were predicted which exhibited strong affinity for MHC-II alleles (Table V). Among these the peptide YRLFRKSNL and VYAWNRKRI reflected affinity for maximum number of alleles. YRLFRKSNL held strong affinity with large number of MHC-II allele including: DRB1_0103, DRB1_0701, DRB1_0801, DRB1_0802, DRB1_1602, DRB4_0103, DRB1_1001, DRB1_1101, DRB1_1501, DRB4_0103, DRB5_0101.

B- and T-cell Epitopes Feature Profiling:

To identify the best epitope for vaccine construction, different features of T-cell epitopes were determined (Table VI). The identified epitopes didn’t show any homology with human proteins, didn’t exhibited mutations and predicted to be non-toxic. The peptides which were digested by fewer enzymes have been considered good potential vaccine candidates (Table VI). Antigenicity of the
peptides depicted that CTL specific peptides can be antigenic except ERDISTEiY. In case of helper T-cell epitopes, FELLHAPAT, TGCVIAWS, and VLYNSASFS were highly antigenic while other peptides were less antigenic. In case of B-cell epitopes, all the 3 peptides were antigenic. Interaction Analysis of CTL Epitopes with MHC-I Specific Alleles:

Two peptides (CVADYSVLY, FTNVYADSF) with the maximum number of bindings with HLA alleles were selected for interaction analysis. PEPFOLD created 5 models for each peptide and the model with the best score was selected for further analysis. HLA-A*0101 was selected as representative allele for interaction analysis. The docking of HLA-A with the top ranking peptide resulted in generation of 50 complexes for each peptide. The peptide 3 (CVADYSVLY) binds with HLA with binding energy of -11.3 KJ/mol. Detailed interaction analysis revealed that peptide is having H-bonding interactions with Asp-115, Arg-156, Tyr-99 and Asn-77 (Fig IV-a). Peptide 4 (FTNVYADSF) binds with the HLA-A with an energy value of -12.6 KJ/mol. The peptide is exhibiting H-bonding interactions with Glu-63, Arg-156, Gln-155, Asn-77, Thr-143, and Lys-146 (Fig IV-b).

Interaction Analysis of HTL Epitopes with MHC-II Specific Alleles:

Two of the peptides (YRLFRKSNL, VYAWNRKRI) that showed affinity with maximum number of MHC-II alleles were selected for interaction analysis. Docking of peptides with HLA-DRB5 depicted both the peptides to be binding with strong affinity. The Peptide 6 and peptide 5 with maximum number of alleles were used to study affinity of these peptides with DRB5_0101. Peptide 5 (YRLFRKSNL) bound with the energy of -10.2KJ/mol while peptide 6 (VYAWNRKRI) bound with the energy of -11.3 KJ/mol. The residues Asp-70, Asn-62, Arg-71, Glu-55 of HLA-DRB5 were having H-bonding interactions with peptide 5 (Fig IV-c). Interaction analysis revealed that Ser-53, Asp-11, Asp-30, and Asp-70 of HLA-DRB5 were exhibiting H-bonding interactions with peptide 6 (Fig IV-d).

Interaction analysis of BatCoV RaTG13 with bACE2:

The interactions of BatCoV RaTG13 with bACE2 was done using HADDOCK. 155 different complexes of BatCoV RaTG13:bACE2 were generated that clustered into 12 groups. Table S1 is showing the Z scores of all the seven clusters, size of each cluster, RMSD from the overall lowest-energy structure,
and energy values of electrostatic, Van der Waals, and de-solvation. The cluster with the best HADDOCK score (-178.9±3.6) was further used for analysis. Detailed interaction analysis showed that 26 residues of bACE2 and 9 residues of BatCoV RaTG13 were present at the interface. These residues were involved in 12 H-bonded, 2 Salt bridges and 157 non-bonded contacts. The detailed interactions are shown in table VII and Fig V.0 V. The important residues of batCoV RaTG13 that were involved in interactions with bACE2 are Lys-417, Leu-455, Phe-456, Asn-487, Tyr-489, Asp-501, His-505. The important conserved residues of BatCoV RaTG13 that involved in interactions and overlapping with epitopes are Leu-455, Phe-456. These residues are lied in one B-cell epitope (455-LFRKSN-461) and T-cell epitope (453-YRLFRKSNL-461). The important interacting residues of BatCoV RaTG13 with bACE2 and a comparison with COVID-19:hACE2 is shown in Fig VI.

Discussion:
Research on various coronavirus species earlier have been continually reported since the last one and a half decade after the emergence of SARS, for annotating signatures and virulence factors [79]. Viral entry receptors are crucial in viral life cycle, sustenance and egress [80]. Realizing their particular tissue tropism further augment their importance for therapeutic targeting and restricting viral entry into cell which can abolish infectivity altogether.

In case of SARS and MERS-CoV, spike protein and specially S1 regions has been the prime focus in developing immune strategies against these strains [52]. Similar strategy can be employed by investigating the S protein for identifying immune epitopes against SARS CoV2. Vaccines against SARS CoV2 can serve as one of the most promising modes of containing COVID19 pandemic. To date no reliable treatment options available for COVID19, so logically vaccine against it is a much needed entity. As the global burden of infectivity by COVID19 pandemic is increasing every day, computational biology aided vaccine design for SARS CoV2 with removal of unnecessary antigenic load and screened allergic response can provide the characteristic immune response required for preventing COVID19 infection. Similar to SARS and MERS- CoV, S1 region of CVOID-19 harbors the RBD which is involved in entry of virus into host cell. Hence, the identification of antigenic peptides within RBD can be a good strategy to forestall infection. In the current study, after identifying the
important receptor binding residues on the RBD for any antigenic peptides were probed. The present study focused on deriving immunogenic epitopes capable of triggering both humoral and cell mediated immune response, on the basis of high degree of comparative sequence similarity of RBD from BatCoV RaTG13 with SARS CoV2. Previously, immunogenic epitopes for SARS CoV2 by immunoinformatics method have been reported in comparison with SARS CoV [24,23,25]. Using this approach may yield rather specific epitopes against COVID-19. Contextually, cross reactivity of SARS CoV antibodies against SARS CoV2 epitopes is also under debate and is providing useful information against potential SARS CoV2 host immune response [32,17,31].

Current results from analysis of RBD of COVID-19 revealed that 423-YKLPDD-428, 455-LFRKSN-461, 494-SYGFQPT-500 are the B-cell epitopes that are highly antigenic. The two epitopes lie in receptor binding motif and also muddled in interactions with hACE2 [71]. The Leu-455, Phe-456, Ser-494, and Thr-500 were the conserved residues among all the COVID-19 sequences hooked in interactions with hACE2. In case of T-cell epitopes, several epitopes were identified from IEDB and NetCTL and the peptides manifesting affinity with maximum number of MHC-I/MHC-II alleles were further selected for analysis. Previously, it has been reported that peptides showing affinity with maximum number of HLA alleles could be very antigenic [54,53]. In case of CTL epitopes, two epitopes (peptide3: CVADYSVLY, peptide4: FTNVYADSF) were found be highly antigenic and also showed strong binding with HLA-A allele. HLA are polymorphic proteins with variable expression in different population. Therefore, a vaccine which is suited for all population without inciting any autoimmunity met by for T-cell epitopes by HLA selectivity is crucial for an effective vaccine candidate [81]. In case of HLA-A*0101, the already docked peptide in crystal structure (PDB ID: 6AT9) has shown interactions with Tyr-7, Phe-9, Phe-33, Val-34, Tyr-59, Glu-63, Asn-77, Tyr-84, Ile-97, Tyr-99, Thr-143, Lys-146, Trp-147, Gln-155, Arg-156, Tyr-159 of HLA-A [82]. Peptide 3(CVADYSVLY) has paraded interactions with Arg-156, Tyr-99 and Asn-77 while peptide 4 (FTNVYADSF) has shown interactions with maximum number of amino acids (Glu-63, Arg-156, Gln-155, Asn-77, Thr-143, and Lys-146). In case of helper T-cell epitopes, YRLFRKSNL, VYAWNRKRL illustrated affinity with maximum number of MHC-II alleles. One of the MHC-II allele HLA-DRB5*0101 was used to study possible interactions of peptide with HLA-DR. In case of HLA-
DRB5*0101 (PDB ID: 1FV1) the crystallized structure has displayed interactions with Asp-9, Phe-12, Glu-55, Met-23, Ser-53, Asn-62, Asp-66, Asn-69, Asp-11, Tyr-13, Asp-70, Arg-71, Tyr-78, Asn-82 [83].

When we studied peptide 5 (YRLFRKSNL) Asn-62, Arg-71, and Glu-55 were residues important for binding of peptide. In case of peptide 6 (VYAWNRKRI), the Ser-53, Asp-11, and Asp-70 were involved in interactions with peptide. The interaction as well as binding energy data depicted these 2 peptides to be binding with HLA-DRB5 and can be used in vaccine construction. The peptide 453-YRLFRKSNL-461, present within the receptor binding motif and Leu-455 and Phe-456 illustrated interactions with hACE2 [71] as well as bACE2. On the basis of different properties of selected B- and T-cell these epitopes, it can be observed that these epitopes were 100% conserved from the reported data till date, as predicted by conservation analysis. These peptides also did not exhibited homology with any human protein hence may not incite any autoimmunity. These peptides did not display any toxicity. The digesting enzyme data showed these peptides to be indigestible by many enzymes and hence are safer to use (Table V). The overlapping B- and T-cell epitopes are LFRKSN and YRLFRKSNL. Hence, the presence of YRLFRKSNL in vaccine construct along with other peptides may enhance the efficacy of vaccine. The results proposed are preliminary and further in-vitro and clinical testing on animals are required for the proposed vaccinable targets of COVID-19. Based on the study results, the predicted epitopes harbor attractive capability to be considered for ascertaining therapeutic potency. The identified B & T cell based epitopes can be deployed for potential peptide based vaccine to prevent further coronaviral infectious spread.

Cumulatively, the detailed interaction analysis of BatCoV RaTG13 have shown that hydrophobic and charged residues have been involved in binding with bACE2. Analogous pattern has been observed in binding of COVID-19 with hACE2 receptor [71]. The important residues of batCoV RaTG13 that were involved in interactions with bACE2 are Lys-417, Leu-455, Phe-456, Asn-487, Tyr-489, Asp-501, His-505. The residues at these positions in COVID-19 were also involved in interactions with hACE2 [71]. Among them Lys-417, Leu-455, Phe-456, Asn-487, and Tyr-489 are the conserved residues (Figure VI).

Designing vaccine against viral infection such as COVID19 is a tricky entity. On one hand, it has to be
ensured that the vaccinable epitopes hold enough antigenic potential to mount a befitting yet specific immune response so as to rapidly clear the infection if the need arises. While on the other hand, the host immune response should not be large enough to ensure chronic inflammation which in case of COVID19 can significantly deteriorate lung infection considering lung as an organ is highly sensitive to inflammatory changes posed by surge of cytokine response [84]. Mutations in the viral genome are capacitating coronaviruses to breach species barrier repeatedly. As the coronaviruses harbor an error-prone RNA dependent RNA polymerase, it may engender recombination events with mutational diversity, concocting therapeutic challenges and survival advantage to virus [79]. This has been the case observed in SARS-2004-2004 epidemic [85,86]. It is of grave concern that COVID-19 has the potential to cause pandemics while considering Ro estimates. As bats are considered primary hosts for coronavirus species, it will be interesting to scrutinize how bats evade viral entry as previous studies have identified bats evolving mechanism for defying interferon pathway activation by the STING interferon pathway [87].

Conclusion:
The current study proposed potential multi epitopes for vaccine development against COVID-19. The potential vaccine epitopes have been rigorously screened for multiple HLA, B-Cell, CTL & HTL epitopes thus augmenting its capability in inducing both humoral and cellular immune responses. Furthermore, it can be co-opted with adjuvant treatment in further enhancing viral disease control. The epitopes were further screened and validated for 100% conservancy with non-overlapping human proteome thus additionally reducing the chances of autoimmunogenic side effects. The molecular docking of epitopes with HLA alleles were further validated for their mode of binding patterns and analysis of binding energy affinities.

Declarations

Acknowledgments:
We would like to acknowledge the Shanghai Public Health Clinical Center and School of Public Health, Central Hospital of Wuhan, Huazhong University of Science and Technology, Wuhan Center for Disease Control and Prevention, National Institute for Communicable Disease Control and Prevention,
Chinese Center for Disease Control, University of Sydney Australia for releasing the sequence of COVID-19 genome and Department of Chemistry, COMSATS, Abbottabad, for computational lab support.

Declaration of Interest statement:

The authors declare no conflict of interest

Funding:

The authors received no external funding for this research.

References

1. Coronavirus latest: WHO officially names disease COVID-19. Nature. https://www.nature.com/articles/d41586-020-00154-w. Accessed on 11th January 2020

2. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med 367 (19):1814-1820. doi:10.1056/NEJMoa1211721

3. Drosten C, Gunther S, Preiser W, van der Werf S, Brodt HR, Becker S, Rabenau H, Panning M, Kolesnikova L, Fouchier RA, Berger A, Burguiere AM, Cinatl J, Eickmann M, Escriou N, Grywna K, Kramme S, Manuguerra JC, Muller S, Rickerts V, Sturmer M, Vieth S, Klenk HD, Osterhaus AD, Schmitz H, Doerr HW (2003) Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med 348 (20):1967-1976. doi:10.1056/NEJMoa030747 NEJMoa030747 [pii]

4. Fouchier RA, Kuiken T, Schutten M, van Amerongen G, van Doornum GJ, van den Hoogen BG, Peiris M, Lim W, Stohr K, Osterhaus AD (2003) Aetiology: Koch's postulates fulfilled for SARS virus. Nature 423 (6937):240. doi:10.1038/423240a423240a [pii]

5. Wang Y, Sun J, Zhu A, Zhao J (2018) Current understanding of middle east respiratory
syndrome coronavirus infection in human and animal models. J Thorac Dis 10 (Suppl 19):S2260-S2271. doi:10.21037/jtd.2018.03.80jtd-10-S19-S2260 [pii]

6. Chan-Yeung M, Xu RH (2003) SARS: epidemiology. Respirology 8 Suppl:S9-14. doi:10.1046/j.1440-1843.2003.00518.x

7. Peiris JS, Guan Y, Yuen KY (2004) Severe acute respiratory syndrome. Nat Med 10 (12 Suppl):S88-97. doi:nm1143 [pii]10.1038/nm1143

8. Lee J, Chowell G, Jung E (2016) A dynamic compartmental model for the Middle East respiratory syndrome outbreak in the Republic of Korea: A retrospective analysis on control interventions and superspreading events. J Theor Biol 408:118-126. doi:10.1016/j.jtbi.2016.08.009S0022-5193(16)30243-0 [pii]

9. Lee JY, Kim YJ, Chung EH, Kim DW, Jeong I, Kim Y, Yun MR, Kim SS, Kim G, Joh JS (2017) The clinical and virological features of the first imported case causing MERS-CoV outbreak in South Korea, 2015. BMC Infect Dis 17 (1):498. doi:10.1186/s12879-017-2576-510.1186/s12879-017-2576-5 [pii]

10. Coronavirus (COVID-19) Cases statistics. https://www.worldometers.info/coronavirus/coronavirus-cases/ Accessed on 14th April 2020

11. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. doi:10.1038/s41586-020-710.1038/s41586-020-2012-7 [pii]

12. Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, Meng J, Zhu Z, Zhang Z, Wang J, Sheng J, Quan L, Xia Z, Tan W, Cheng G, Jiang T (2020) Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. Cell Host
13. Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, Liu W, Bi Y, Gao GF (2016) Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. Trends Microbiol 24 (6):490-502. doi:S0966-842X(16)00071-8 [pii]10.1016/j.tim.2016.03.003

14. Hu B, Zeng LP, Yang XL, Ge XY, Zhang W, Li B, Xie JZ, Shen XR, Zhang YZ, Wang N, Luo DS, Zheng XS, Wang MN, Daszak P, Wang LF, Cui J, Shi ZL (2017) Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. PLoS Pathog 13 (11):e1006698. doi:10.1371/journal.ppat.1006698PPATHOGENS-D-17-00265 [pii]

15. Jeffers SA, Tusell SM, Gillim-Ross L, Hemmila EM, Achenbach JE, Babcock GJ, Thomas WD, Jr., Thackray LB, Young MD, Mason RJ, Ambrosino DM, Wentworth DE, Demartini JC, Holmes KV (2004) CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus. Proc Natl Acad Sci U S A 101 (44):15748-15753. doi:0403812101 [pii]10.1073/pnas.0403812101

16. Meyerholz DK, Lambertz AM, McCray PB, Jr. (2016) Dipeptidyl Peptidase 4 Distribution in the Human Respiratory Tract: Implications for the Middle East Respiratory Syndrome. Am J Pathol 186 (1):78-86. doi:10.1016/j.ajpath.2015.09.014S0002-9440(15)00576-3 [pii]

17. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D (2020) Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell. doi:S0092-8674(20)30262-2 [pii]10.1016/j.cell.2020.02.058

18. Jiang S, Lu L, Liu Q, Xu W, Du L (2012) Receptor-binding domains of spike proteins of emerging or re-emerging viruses as targets for development of antiviral vaccines. Emerg Microbes Infect 1 (8):e13. doi:10.1038/emi.2012.1

19. Zhu X, Liu Q, Du L, Lu L, Jiang S (2013) Receptor-binding domain as a target for
18

developing SARS vaccines. J Thorac Dis 5 Suppl 2:S142-148. doi:10.3978/j.issn.2072-1439.2013.06.06jtd-05-S2-S142 [pii]

20. Du L, Yang Y, Zhou Y, Lu L, Li F, Jiang S (2017) MERS-CoV spike protein: a key target for antivirals. Expert Opin Ther Targets 21 (2):131-143.
doi:10.1080/14728222.2017.1271415

21. Davies MN, Flower DR (2007) Harnessing bioinformatics to discover new vaccines. Drug Discov Today 12 (9-10):389-395. doi:S1359-6446(07)00135-3 [pii]10.1016/j.drudis.2007.03.010

22. Srivastava S, Kamthania M, Singh S, Saxena AK, Sharma N (2018) Structural basis of development of multi-epitope vaccine against Middle East respiratory syndrome using in silico approach. Infect Drug Resist 11:2377-2391.
doi:10.2147/IDR.S175114idr-11-2377 [pii]

23. Baruah V, Bose S (2020) Immunoinformatics-aided identification of T cell and B cell epitopes in the surface glycoprotein of 2019-nCoV. J Med Virol 92 (5):495-500.
doi:10.1002/jmv.25698

24. Ahmed SF, Quadeer AA, McKay MR (2020) Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. Viruses 12 (3). doi:E254 [pii]10.3390/v12030254v12030254 [pii]

25. Bhattacharya M, Sharma AR, Patra P, Ghosh P, Sharma G, Patra BC, Lee SS, Chakraborty C (2020) Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach. J Med Virol.
doi:10.1002/jmv.25736

26. Kuehn BM (2013) More evidence emerges that bats may have spread SARS. JAMA 310 (20):2138. doi:10.1001/jama.2013.2834951785455 [pii]
27. Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH, Wong SS, Leung SY, Chan KH, Yuen KY (2005) Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc Natl Acad Sci U S A 102 (39):14040-14045. doi:10.1073/pnas.0506735102

28. Kan B, Wang M, Jing H, Xu H, Jiang X, Yan M, Liang W, Zheng H, Wan K, Liu Q, Cui B, Xu Y, Zhang E, Wang H, Ye J, Li G, Li M, Cui Z, Qi X, Chen K, Du L, Gao K, Zhao YT, Zou XZ, Feng YJ, Gao YF, Hai R, Yu D, Guan Y, Xu J (2005) Molecular evolution analysis and geographic investigation of severe acute respiratory syndrome coronavirus-like virus in palm civets at an animal market and on farms. J Virol 79 (18):11892-11900. doi:10.1128/JVI.79.18.11892-11900.2005

29. Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, Mazet JK, Hu B, Zhang W, Peng C, Zhang YJ, Luo CM, Tan B, Wang N, Zhu Y, Cramer G, Zhang SY, Wang LF, Daszak P, Shi ZL (2013) Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature 503 (7477):535-538. doi:10.1038/nature12711

30. Forni D, Cagliani R, Clerici M, Sironi M (2017) Molecular Evolution of Human Coronavirus Genomes. Trends Microbiol 25 (1):35-48. doi:S0966-842X(16)30133-0

31. Yuan M, Wu NC, Zhu X, Lee CD, So RTY, Lv H, Mok CKP, Wilson IA (2020) A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV. Science. doi:eabb7269

32. Tian X, Li C, Huang A, Xia S, Lu S, Shi Z, Lu L, Jiang S, Yang Z, Wu Y, Ying T (2020) Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. Emerg Microbes Infect 9 (1):382-385. doi:10.1080/22221751.2020.1729069
33. Bakkers MJ, Zeng Q, Feitsma LJ, Hulswit RJ, Li Z, Westerbeke A, van Kuppeveld FJ, Boons GJ, Langereis MA, Huizinga EG, de Groot RJ (2016) Coronavirus receptor switch explained from the stereochemistry of protein-carbohydrate interactions and a single mutation. Proc Natl Acad Sci U S A 113 (22):E3111-3119. doi:10.1073/pnas.1519881113 [pii]

34. Xu X, Chen P, Wang J, Feng J, Zhou H, Li X, Zhong W, Hao P (2020) Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. Sci China Life Sci. doi:10.1007/s11427-020-1637-5 [pii]

35. Kannan S, Kolandaivel P (2016) Computational studies of pandemic 1918 and 2009 H1N1 hemagglutinins bound to avian and human receptor analogs. J Biomol Struct Dyn 34 (2):272-289. doi:10.1080/07391102.2015.1027737

36. Tran-To Su C, Ouyang X, Zheng J, Kwoh CK (2013) Structural analysis of the novel influenza A (H7N9) viral Neuraminidase interactions with current approved neuraminidase inhibitors Oseltamivir, Zanamivir, and Peramivir in the presence of mutation R289K. BMC Bioinformatics 14 Suppl 16:S7. doi:10.1186/1471-2105-14-S16-S71471-2105-14-S16-S7 [pii]

37. Pan D, Xue W, Wang X, Guo J, Liu H, Yao X (2012) Molecular mechanism of the enhanced virulence of 2009 pandemic influenza A (H1N1) virus from D222G mutation in the hemagglutinin: a molecular modeling study. J Mol Model 18 (9):4355-4366. doi:10.1007/s00894-012-1423-2

38. Skehel JJ, Wiley DC (2000) Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu Rev Biochem 69:531-569. doi:69/1/531 [pii]10.1146/annurev.biochem.69.1.531

39. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA (1998)
Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature 393 (6686):648-659. doi:10.1038/31405

40. Rossmann MG, He Y, Kuhn RJ (2002) Picornavirus-receptor interactions. Trends Microbiol 10 (7):324-331. doi:S0966842X02023831 [pii]10.1016/s0966-842x(02)02383-1

41. Chen J (2020) Pathogenicity and Transmissibility of 2019-nCoV-A Quick Overview and Comparison with Other Emerging Viruses. Microbes Infect. doi:S1286-4579(20)30026-5 [pii]10.1016/j.micinf.2020.01.004

42. Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, Lou Y, Gao D, Yang L, He D, Wang MH (2020) Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. Int J Infect Dis. doi:S1201-9712(20)30053-9 [pii]10.1016/j.ijid.2020.01.050

43. Sominskaya I, Skrastina D, Dislers A, Vasiljev D, Mihailova M, Ose V, Dreilina D, Pumpens P (2010) Construction and immunological evaluation of multivalent hepatitis B virus (HBV) core virus-like particles carrying HBV and HCV epitopes. Clin Vaccine Immunol 17 (6):1027-1033. doi:10.1128/CVI.00468-09CVI.00468-09 [pii]

44. Rodrigues RL, Menezes GL, Saivish MV, Costa VGD, Pereira M, Moreli ML, Silva RAD (2019) Prediction of MAYV peptide antigens for immunodiagnostic tests by immunoinformatics and molecular dynamics simulations. Sci Rep 9 (1):13339. doi:10.1038/s41598-019-50008-3 [pii]

45. Stanekova Z, Vareckova E (2010) Conserved epitopes of influenza A virus inducing protective immunity and their prospects for universal vaccine development. Virol J 7:351. doi:10.1186/1743-422X-7-3511743-422X-7-351 [pii]

46. Jabbar B, Rafique S, Salo-Ahen OMH, Ali A, Munir M, Idrees M, Mirza MU, Vanmeert M,
Shah SZ, Jabbar I, Rana MA (2018) Antigenic Peptide Prediction From E6 and E7 Oncoproteins of HPV Types 16 and 18 for Therapeutic Vaccine Design Using Immunoinformatics and MD Simulation Analysis. Front Immunol 9:3000. doi:10.3389/fimmu.2018.03000

47. Dutta NK, Mazumdar K, Lee BH, Baek MW, Kim DJ, Na YR, Park SH, Lee HK, Kariwa H, Mai le Q, Park JH (2008) Search for potential target site of nucleocapsid gene for the design of an epitope-based SARS DNA vaccine. Immunol Lett 118 (1):65-71. doi:10.1016/j.imlet.2008.03.003 [pii]

48. Liu SJ, Leng CH, Lien SP, Chi HY, Huang CY, Lin CL, Lian WC, Chen CJ, Hsieh SL, Chong P (2006) Immunological characterizations of the nucleocapsid protein based SARS vaccine candidates. Vaccine 24 (16):3100-3108. doi:S0264-410X(06)00087-9 [pii]

49. Zhao J, Huang Q, Wang W, Zhang Y, Lv P, Gao XM (2007) Identification and characterization of dominant helper T-cell epitopes in the nucleocapsid protein of severe acute respiratory syndrome coronavirus. J Virol 81 (11):6079-6088. doi:JVI.02568-06 [pii]

50. Zhou M, Xu D, Li X, Li H, Shan M, Tang J, Wang M, Wang FS, Zhu X, Tao H, He W, Tien P, Gao GF (2006) Screening and identification of severe acute respiratory syndrome-associated coronavirus-specific CTL epitopes. J Immunol 177 (4):2138-2145. doi:177/4/2138 [pii]

51. Gupta V, Tabiin TM, Sun K, Chandrasekaran A, Anwar A, Yang K, Chikhlikar P, Salmon J, Brusic V, Marques ET, Kellathur SN, August TJ (2006) SARS coronavirus nucleocapsid immunodominant T-cell epitope cluster is common to both exogenous recombinant and endogenous DNA-encoded immunogens. Virology 347 (1):127-139. doi:S0042-6822(05)00783-X [pii]
52. Wang L, Shi W, Joyce MG, Modjarrad K, Zhang Y, Leung K, Lees CR, Zhou T, Yassine HM, Kanekiyo M, Yang ZY, Chen X, Becker MM, Freeman M, Vogel L, Johnson JC, Olinger G, Todd JP, Bagci U, Solomon J, Mollura DJ, Hensley L, Jahrling P, Denison MR, Rao SS, Subbarao K, Kwong PD, Mascola JR, Kong WP, Graham BS (2015) Evaluation of candidate vaccine approaches for MERS-CoV. Nat Commun 6:7712. doi:10.1038/ncomms8712 [pii]

53. Shi J, Zhang J, Li S, Sun J, Teng Y, Wu M, Li J, Li Y, Hu N, Wang H, Hu Y (2015) Epitope-Based Vaccine Target Screening against Highly Pathogenic MERS-CoV: An In Silico Approach Applied to Emerging Infectious Diseases. PLoS One 10(12):e0144475. doi:10.1371/journal.pone.0144475PONE-D-15-41720 [pii]

54. Tahir Ul Qamar M, Saleem S, Ashfaq UA, Bari A, Anwar F, Alqahtani S (2019) Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: an immune-informatics study. J Transl Med 17(1):362. doi:10.1186/s12967-019-2116-8 [pii]

55. Thanh Le T, Andreadakis Z, Kumar A, Gomez Roman R, Tollefsen S, Saville M, Mayhew S (2020) The COVID-19 vaccine development landscape. Nat Rev Drug Discov. doi:10.1038/d41573-020-00073-5 [pii]

56. Kang S, Peng W, Zhu Y, Lu S, Zhou M, Lin W, Wu W, Huang S, Jiang L, Luo X, Deng M (2020) Recent Progress in understanding 2019 Novel Coronavirus associated with Human Respiratory Disease: Detection, Mechanism and Treatment. Int J Antimicrob Agents:105950. doi:S0924-8579(20)30100-X [pii]10.1016/j.ijantimicag.2020.105950

57. Doytchinova IA, Flower DR (2007) Vaxijen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics 8:4. doi:1471-2105-8-4 [pii]10.1186/1471-2105-8-4

58. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, Wheeler DK,
Gabbard JL, Hix D, Sette A, Peters B (2015) The immune epitope database (IEDB) 3.0. Nucleic Acids Res 43 (Database issue):D405-412. doi:10.1093/nar/gku938gku938 [pii]

Jespersen MC, Peters B, Nielsen M, Marcatili P (2017) BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res 45 (W1):W24-W29. doi:10.1093/nar/gkx3463787843 [pii]

Kolaskar AS, Tongaonkar PC (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett 276 (1-2):172-174. doi:0014-5793(90)80535-Q [pii]10.1016/0014-5793(90)80535-q

Emini EA, Hughes JV, Perlow DS, Boger J (1985) Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J Virol 55 (3):836-839

Chou PY, Fasman GD (1978) Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol Relat Areas Mol Biol 47:45-148. doi:10.1002/9780470122921.ch2

Vihinen M, Torkkila E, Riikonen P (1994) Accuracy of protein flexibility predictions. Proteins 19 (2):141-149. doi:10.1002/prot.340190207

Larsen JE, Lund O, Nielsen M (2006) Improved method for predicting linear B-cell epitopes. Immunome Res 2:2. doi:1745-7580-2-2 [pii]10.1186/1745-7580-2-2

Ponomarenko J, Bui HH, Li W, Fusseder N, Bourne PE, Sette A, Peters B (2008) ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics 9:514. doi:10.1186/1471-2105-9-5141471-2105-9-514 [pii]

Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M (2007) Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinformatics 8:424. doi:1471-2105-8-424 [pii]10.1186/1471-2105-8-424

Andreatta M, Nielsen M (2016) Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics 32 (4):511-517.
68. Jensen KK, Andreatta M, Marcatili P, Buus S, Greenbaum JA, Yan Z, Sette A, Peters B, Nielsen M (2018) Improved methods for predicting peptide binding affinity to MHC class II molecules. Immunology 154 (3):394-406. doi:10.1111/imm.12889

69. Naamati G, Askenazi M, Linial M (2009) ClanTox: a classifier of short animal toxins. Nucleic Acids Res 37 (Web Server issue):W363-368. doi:10.1093/nar/gkp299gkp299 [pii]

70. Maupetit J, Derreumaux P, Tuffery P (2009) PEP-FOLD: an online resource for de novo peptide structure prediction. Nucleic Acids Res 37 (Web Server issue):W498-503. doi:10.1093/nar/gkp323gkp323 [pii]

71. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, Zhang Q, Shi X, Wang Q, Zhang L, Wang X (2020) Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. doi:10.1038/s41586-020-2180-5 [pii]

72. Sheik SS, Sundararajan P, Hussain AS, Sekar K (2002) Ramachandran plot on the web. Bioinformatics 18 (11):1548-1549. doi:10.1093/bioinformatics/18.11.1548

73. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35 (Web Server issue):W407-410. doi:gkm290 [pii]10.1093/nar/gkm290

74. Benkert P, Tosatto SC, Schomburg D (2008) QMEAN: A comprehensive scoring function for model quality assessment. Proteins 71 (1):261-277. doi:10.1002/prot.21715

75. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB, Jain S, Lewis SM, Arendall WB, 3rd, Snoeyink J, Adams PD, Lovell SC, Richardson JS, Richardson DC (2018) MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci 27 (1):293-315.
76. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2 (9):1511-1519. doi:10.1002/pro.5560020916

77. van Zundert GCP, Rodrigues J, Trellet M, Schmitz C, Kastritis PL, Karaca E, Melquiond ASJ, van Dijk M, de Vries SJ, Bonvin A (2016) The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J Mol Biol 428 (4):720-725. doi:S0022-2836(15)00537-9 [pii]10.1016/j.jmb.2015.09.014

78. Laskowski RA, Jablonska J, Pravda L, Varekova RS, Thornton JM (2018) PDBsum: Structural summaries of PDB entries. Protein Sci 27 (1):129-134. doi:10.1002/pro.3289

79. Cui J, Li F, Shi ZL (2019) Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol 17 (3):181-192. doi:10.1038/s41579-018-0118-910.1038/s41579-018-0118-9

80. Marsh M, Helenius A (2006) Virus entry: open sesame. Cell 124 (4):729-740. doi:S0092-8674(06)00182-6 [pii]10.1016/j.cell.2006.02.007

81. Maenaka K, Jones EY (1999) MHC superfamily structure and the immune system. Curr Opin Struct Biol 9 (6):745-753. doi:S0959-440X(99)00039-1 [pii]10.1016/s0959-440x(99)00039-1

82. Toor JS, Rao AA, McShan AC, Yarmarkovich M, Nerli S, Yamaguchi K, Madejska AA, Nguyen S, Tripathi S, Maris JM, Salama SR, Haussler D, Sgourakis NG (2018) A Recurrent Mutation in Anaplastic Lymphoma Kinase with Distinct Neoepitope Conformations. Front Immunol 9:99. doi:10.3389/fimmu.2018.00099

83. Li Y, Li H, Martin R, Mariuzza RA (2000) Structural basis for the binding of an immunodominant peptide from myelin basic protein in different registers by two HLA-DR2 proteins. J Mol Biol 304 (2):177-188. doi:10.1006/jmbi.2000.4198S0022-
2836(00)94198-6 [pii]

84. Liu Q, Zhou YH, Yang ZQ (2016) The cytokine storm of severe influenza and development of immunomodulatory therapy. Cell Mol Immunol 13 (1):3-10. doi:10.1038/cmi.2015.74 [pii]

85. Graham RL, Baric RS (2010) Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. J Virol 84 (7):3134-3146. doi:10.1128/JVI.01394-09 [pii]

86. Dawson P, Malik MR, Parvez F, Morse SS (2019) What Have We Learned About Middle East Respiratory Syndrome Coronavirus Emergence in Humans? A Systematic Literature Review. Vector Borne Zoonotic Dis 19 (3):174-192. doi:10.1089/vbz.2017.2191

87. Xie J, Li Y, Shen X, Goh G, Zhu Y, Cui J, Wang LF, Shi ZL, Zhou P (2018) Dampened STING-Dependent Interferon Activation in Bats. Cell Host Microbe 23 (3):297-301 e294. doi:S1931-3128(18)30041-6 [pii]10.1016/j.chom.2018.01.006

Tables
Table I: B-cell epitope antigenicity analysis by Kolaskar & Tongaonkar.

| No. | Start position | End position | Peptide               | Peptide length |
|-----|----------------|--------------|-----------------------|----------------|
| 1   | 334            | 341          | NLCPFGEV             | 8              |
| 2   | 347            | 353          | FASVYAW              | 7              |
| 3   | 358            | 372          | ISNCVADYSVLNNSA      | 15             |
| 4   | 374            | 385          | FSTFKCYGVSPF        | 12             |
| 5   | 387            | 404          | LNDLCFTNVYADSFV     | 18             |
| 6   | 407            | 412          | VRQIAP               | 6              |
| 7   | 429            | 436          | FTGCVIAW             | 8              |
| 8   | 455            | 461          | LFRKSNL              | 6              |
| 9   | 423            | 428          | YKLPDD               | 5              |
| 10  | 470            | 478          | TEIYQAGST            | 9              |
| 11  | 494            | 500          | SYGFOPT              | 6              |

Table II: Prediction of epitopes on basis of surface accessibility by Emini method.
| No. | Start position | End position | Peptide         | Peptide length |
|-----|----------------|--------------|-----------------|----------------|
| 1   | 419            | 428          | ADYNYKLPDD      | 10             |
| 2   | 437            | 442          | NSNNLD          | 6              |
| 3   | 455            | 468          | LFRKSNLPFDERDI  | 14             |
| 4   | 495            | 500          | YGFOPT          | 6              |

**Table III:** Epitope prediction by Bepipred method.

| No. | Start position | End position | Peptide          | Peptide length |
|-----|----------------|--------------|------------------|----------------|
| 1   | 382            | 385          | VSPT             | 4              |
| 2   | 407            | 420          | VRQIAFGQTGKIAD   | 14             |
| 3   | 423            | 428          | YKLPPD          | 6              |
| 4   | 439            | 447          | NNLDSKVGG       | 9              |
| 5   | 461            | 463          | LKP              | 3              |
| 6   | 466            | 467          | RD               | 2              |
| 7   | 469            | 483          | YQAGSTPCNGV     | 11             |
| 8   | 473            | 506          | YGFOPTNGGYQ     | 12             |

**Table IV:** Top scoring T-cell epitopes as predicted by NetCTL and recognized by MHC-I alleles.
Table V: T-cell epitopes bind to specific MHC-II alleles.
| No. | Peptide     | MHC-II alleles          |
|-----|-------------|-------------------------|
| 1   | FELLHAPAT   | DRB1_0101               |
|     |             | DRB1_1001               |
| 2   | FNATRFAS    | DRB1_0402               |
| 3   | TGCVIAWNS   | DRB1_0403               |
|     |             | DRB3_0202               |
| 4   | FRKSNLKPF   | DRB1_0701               |
| 5   | YRLFRKSNL   | DRB1_0103               |
|     |             | DRB1_0701               |
|     |             | DRB1_0801               |
|     |             | DRB1_0802               |
|     |             | DRB1_1602               |
|     |             | DRB4_0103               |
|     |             | DRB1_1001               |
|     |             | DRB1_1101               |
|     |             | DRB1_1501               |
|     |             | DRB4_0103               |
|     |             | DRB5_0101               |
| 6   | VYAWNRKRI   | DRB1_1101               |
|     |             | DRB1_1301               |
|     |             | DRB4_0103               |
|     |             | DRB5_0101               |
|     |             | DRB1_0402               |
|     |             | DRB3_0202               |
| 7   | FERDISTEII  | DRB3_0101               |
| 8   | IRGDEVRQII  | DRB3_0101               |
| 9   | VLYNSASFS   | DRB3_0202               |
Table VI: Profiling of MHC-I, MHC-II, and B-cell specific peptides. Prediction of their toxicity, mutation, homologs... antigenicity, digestion with enzymes and other physiochemical properties as charge, weight and hydrophobicity.

| No. | Peptide         | Homology with Human Proteins (Y/N) | Mutation (Y/N) | Toxicity (Y/N) | Hydrophobicity | Charge | Mol. weight |
|-----|-----------------|------------------------------------|----------------|----------------|----------------|--------|------------|
|     | MHC-I Peptide   |                                    |                |                |                |        |            |
| 1   | NATRFASVY       | N                                  | N              | N              | -0.13          | 1.0    | 1028.24    |
| 2   | RISNCVADY       | N                                  | N              | N              | -0.20          | 0.00   | 1040.27    |
| 3   | CVADYSVLY       | N                                  | N              | N              | 0.11           | -1.00  | 1032.29    |
| 4   | FTNVYADSF       | N                                  | N              | N              | 0.03           | -1.0   | 10632.24   |
| 5   | ERDISTEIY       | N                                  | N              | N              | -0.30          | -2.0   | 1125.30    |
|     | MHC-II Peptide  |                                    |                |                |                |        |            |
|    | peptide  | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | Score |
|----|----------|----|----|----|----|----|----|----|----|----|-------|
| 1  | FELLHAPAT| N  | N  | N  | N  | 0.10 | -0.5 | 998.27 |
| 2  | FNATRFAS | N  | N  | N  | N  | -0.14 | 1.0 | 913.09 |
| 3  | TGCVIAWS | N  | N  | N  | N  | 0.11 | 0.0 | 950.20 |
| 4  | FRKSNLKPF| N  | N  | N  | N  | -0.35 | 3.0 | 1136.48 |
| 5  | YRLFRKSNL| N  | N  | N  | N  | -0.43 | 3.0 | 1196.54 |
| 6  | VYAWNRKRI| N  | N  | N  | N  | -0.03 | 3.0 | 1205.55 |
| 7  | FERDISTEI| N  | N  | N  | N  | -0.23 | -2.0 | 1109.32 |
| 8  | IRGDEVRQI| N  | N  | N  | N  | -0.38 | 0.00 | 1085.36 |
| 9  | VLYNSASFS| N  | N  | N  | N  | 0.06 | 0.00 | 987.19 |

|    | peptide  | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | Score |
|----|----------|----|----|----|----|----|----|----|----|----|-------|
| 1  | YKLPDD   | N  | N  | N  | N  | -0.34 | -1.00 | 749.89 |
| 2  | LFRKSN   | N  | N  | N  | N  | -0.44 | 2.00 | 763.97 |
| 3  | SYGFQPT  | N  | N  | N  | N  | -0.06 | 0.00 | 798.94 |

Table VII: Interactions of BatCoV RaTG13 with bat-ACE2.
| S.No | Res Name | Type of interactions | Res Name | Distance (Å) |
|------|----------|---------------------|----------|--------------|
| 1.   | Glu-24   | H-bond              | Tyr-473  | 2.73         |
| 1.   | Asp-30   | H-bond              | Lys-417  | 2.63         |
| 1.   | Thr-30   | H-bond              | His-505  | 2.78         |
|      | Lys-353  |                     |          |              |
| 1.   | Lys-35   | H-bond              | Arg-494  | 3.08         |
| 1.   | Asp-38   | H-bond              | Tyr-498  | 2.55         |
| 1.   | Gln-42   | H-bond              | Glu-445  | 2.73         |
| 1.   | Lys-61   | H-bond              | Glu-445  | 2.86         |
| 1.   | Leu-79   | H-bond              | Tyr-489  | 2.84         |
| 1.   | Asn-82   | H-bond              | Asn-487  | 3.03         |
| 1.   | Lys-31   | H-bond              | Leu-455  | 2.01         |
|      |          |                     | Phe-456  | 3.8          |
| 1.   | Gln-42   | H-bond              | Tyr-489  | 3.4          |
| 1.   | His-41   | H-bond              | Asp-501  | 3.9          |
| 1.   | Asp-30   | Salt bridge         | Lys-417  | 2.63         |
| 1.   | Lys-61   | Salt bridge         | Glu-445  | 2.77         |

**Figures**
Detailed workflow for analysis of receptor binding domain of COVID-19. The sequences of COVID-19, BatCoV RaTG13 and SARS-COV were downloaded from NCBI followed by multiple sequence alignment through ClustalX and JalView. Within the receptor binding domain, the antigenic epitopes were also determined using immune-informatics approach. Linear and conformational B-cell epitopes were determined. Linear B-cell epitopes were determined through different methods that include Bepipred, Karplus & Schulz, Kolaskar & Tongaonkar, Emini, Chou & Fasman methods. Conformational B-cell epitopes were determined through Ellipro. Cytotoxic and helper T-cell epitopes were also determined. CTL epitopes were determined through NetCTL followed by their affinity for specific MHC-I alleles through NetMHCI 4.0. Helper T-cell epitopes were identified through IEDB followed by identification of allele specificity through NetMHCII 2.3. all the B- and T-cell epitopes were then subjected to feature profiling by determining their toxicity, mutation, antigenicity, conservation,
charge, molecular weight and non-digesting enzymes. Selected T-cell epitopes were then modelled through PEPFOLD. The interaction analysis of peptides with HLA-A and HLA-DR was done through Molecular Operating Environment (MOE). Interacting residues of HLA-A and HLA-DR with peptides was determined through LigX in MOE and PyMOL.

![Multiple sequence alignment of selected COVID-19 sequences with the SARS-CoV and BatCoV RaTG13 spike protein receptor binding domain. The sequence is highlighted on the basis of % identity among them. The receptor binding motif is highlighted with red box. Numbering is according to COVID-19.](image)

**Figure 2**
Figure 3

Mapping of linear B-cell epitopes on three dimensional structure of receptor binding domain of COVID-19. Epitopes are represented by spheres.
Figure 4

Interaction analysis of peptides with HLA-A and HLA-DR. (a) CVADYSVLY:HLA-A*01:01 (b) FTNVYADSF:HLA-A*01:01. (c) YRLFRKSNL:DRB5_0101 (d) VYAWNRKRI:DRB5_0101. HLA-A protein is shown in grey while HLA-DR protein is shown in yellow. The residues involved in interactions are highlighted as stick representation.
Figure 5

Top scoring complex of BatCoV RaTG13:bACE2. BatCoV RaTG13 is shown in blue while bACE2 is shown in pink.

Figure 6

Sequence alignment of receptor binding domain of COVID-19 with BatCoV RaTG13. B-cell epitopes are highlighted with orange boxes. CTL epitopes are highlighted with blue boxes. T-helper cell epitopes are highlighted as green boxes. Green dots represent COVID-19 interacting residues with hACE2 while blue dots represent interacting residues of BatCoV RaTG13 with bACE2.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.
TableS1.docx
