 doi:10.20944/preprints202002.0359.v1

In silico Identification of novel B Cell and T cell epitopes of Wuhan coronavirus (2019-nCoV) for effective multi epitope-based peptide vaccine production

Muhammad Asif Rasheed[1,#], Sohail Raza[2,#*], Ali Zohaib[3], Tahir yaqub[2], Masood Rabbani[2], Muhammad Ilyas Riaz[2] , Muhammad Awais[1], Amna Afzal[1]

[1] Department of Biosciences, COMSATS University Islamabad, Sahiwal Campus, 57000, Sahiwal, Pakistan

[2] Department of Microbiology, University of Veterinary and Animal Sciences, Lahore 54000, Pakistan

[3]Department of Health Care Biotechnology, Atta-ur-Rehman School of Applied Biosciences(ASAB), National University of Science and Technology, Islamabad 44000, Pakistan

*__Correspondence to__:

Dr. Sohail Raza
Assistant Professor,
Department of Microbiology,
University of Veterinary and Animal Sciences,
Lahore 54000, Pakistan
Tel: +923006524483
Email: sohail.raza@uvas.edu.pk

# Equal contributing authors.

Running title: Identification of epitopes in coronavirus

 doi:10.20944/preprints202002.0359.v1

Abstract

During December 2019, a novel coronavirus named as 2019-nCoV, has emerged in Wuhan, China. The human to human transmission of this virus has also been established. Untill now the virus has infected more than seven thousand people and has spread to fifteen countries. The World Health Organization (WHO) has declared 2019-nCoV as global health emergency due to its outburst well beyond China. There is need to develop some vaccines or therapeutics to control or prevent 2019-nCoV infections. The bottleneck with current conventional approaches is that these require longer time for vaccine development. However, computer assisted approaches help us to produce effective vaccine in short time compared with conventional methods. In this study, bioinformatics analysis was used to predict B cell and T cell epitopes of surface glycoprotein of 2019-nCoV that could be suitable to trigger significant immune response**.** The sequence of surface glycoprotein was collected from the database and analyzed to identify the immunogenic epitope. Both B cell and T cell epitopes were analyzed so the predicted epitopes can stimulate both cellular and humoral immune responses. We predicted 13 B cell and 05 T cell epitopes that later on were joined with GPGPG linker to make a single peptide. This computational approach to design a multi epitope peptide vaccine against emerging 2019-nCoV allows us to find novel immunogenic epitopes against the antigen targets of surface 2019-nCoV surface glycoprotein. This multi epitope peptide vaccine may prove effective to combat 2019-nCoV infections.

Keywords: Novel coronavirus, Wuhan, Vaccine, Epitopes, Peptide

Introduction

 doi:10.20944/preprints202002.0359.v1

With the advent of twentieth century, novel viruses start to emerge posing global health issues with increasing impact. One of the recent examples of such viruses is the 'Wuhan coronavirus (2019-nCoV)'. This zoonotic virus has been found to have potential to infect humans. This virus was firstly indentified in Wuhan, China in individuals that were exposed to the seafood market (1). The infection of this virus leads to respiratory damage with signs and symptoms of fever, nasal discharge, frequent coughs and sore throat. Fatal cases involved acute respiratory distress, pneumonia, and multipleorgan failure. The 2019-nCoV virus infection appears to be milder than other human corona viruses (SARS and MERS) in terms of case fatality, severity, and transmission (2). To date, more than thousand cases with more than fifty deaths of 2019-nCoV infection have been reported from Wuhan, China. Initially, there were limited evidences of human to human virus transfer because no health care personals in the hospitals were infected (3). After few days, as these cases start to increase, officials admit the horizontal transfer of 2019-nCoV and spread through nasal secretions of the infected person (4, 5). In the seafood market of Wuhan, along with other sea animals, chicken, bats, snake and other wild animals were also sold. In the beginning, it was predicted that spike protein of the 2019-nCoV contribute its transfer from snakes to humans (6). Till now, there are no evidences available for snake infections by this 2019-nCoV (7).

The coronavirus belongs to the family *Coronaviridae*, based on their genetic properties divided into four genera including *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus.* The corona viruses are enveloped, non-segmented, containing single stranded positive sense RNA genome. The genome size ranges from 26 kbp to 32 kbp, largest of all RNA viruses. The corornaviruses are 80 to 120 nm in diameter and spherical or pleomorphic in shape (8). The genome encodes for four structural proteins: the spike protein (S), Membrane protein (M), Nucleocapsid protein (N), and envelop protein (E) (9). For most of the coronaviruses, all structural proteins are required to produce complete virus particle; however few coronaviruses do not require full structural protein to produce complete infectious virus particle (10). Each protein not only plays its role in the structure of the virus but also plays part in other aspects of viral replication. The S protein is involved in the viral attachment and fusion with host cell to facilitate viral entry into the cell (11). The N protein involved in the nucleocapsid formation and host cellular response during viral infection (12). The M protein is involved in structure of virus and has central role in t assembly of the coronavirus (13). The E

 doi:10.20944/preprints202002.0359.v1

protein has an important role in assembly of the virion (10). Among these, S protein playes an important role in the production of immunity against coronaviruses.

The 2019-nCoV is the seventh of the family that infects humans. During genomic analysis this virus formed different clade in sarbecovirus subgenus, subfamily orthocoronavirinae. The 2019-nCoV is different from other known human coronaviruses as it was isolated using human airway epithelial cells from the bronch-alveolar lavage of the patients admitted in the hospital. (14). The closest linkage of 2019-nCoV is with two SARS like coronaviruses that are transferred through bats. Coronaviruses of the other species are genetically different from 2019-nCoV, reflecting that new coronavirus did not originate from other animal hosts (3).

Presently, no vaccine or therapeutic agents are available to treat and protect from human coronavirus infections. Development of effective treatment and vaccine is a research priority to control emerging diseases. The immunoinformtics, a novel approach to find out the effective ways to control the diseases, is these a better tool (15). The conventional approaches for vaccine development not only require more time but also ineffective to control of viral diseases spread through RNA viruses. Hence, the immunoinformatics based approach has become much popular with the advancement in the genomic and protein sequence databases (16). Therefore, the main aim of this study was to predict T-cell and B-cell epitopes present in the surface glycoprotein of novel 2019-nCoV that could be used to develop promising vaccine against 2019-nCoV.

Materials and Methods

Retrieval of sequence

The sequence of surface glycoprotein of the isolated corona virus was retrieved from NCBI protein database (https://www.ncbi.nlm.nih.gov/nuccore/1798172431). The accession number of the protein is QHD43416.1.

Prediction of B cell epitopes and joining

The linear B cell epitopes are used to induce defensive immunity (17) while B cell epitopes are chunk of proteins or other molecules by which the antibodies bind. Hidden Markov models (HMMs) were used for the prediction of the location of Linear B cell epitopes. In order to predict

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 February 2020 doi:10.20944/preprints202002.0359.v1

B cell epitopes in the retrieved protein, Immune Epitope Database (IEDB: http://tools.iedb.org/main/) was used. The epitopes were predicted by B cell linear epitope prediction 2.0. The epitopes were selected and joined according to the length of the predicted epitopes and by remove the redundancy. All the epitopes between length of 10 to 30 amino acids were selected. A GPGPG spacer was used to join the selected epitopes. GPGPG had been selected based on certain properties (18).

Prediction of T cell epitopes and joining

T cell epitopes exist on the outside of an antigen-presenting cell. Just like B cell epitopes prediction, IEDB resource analysis was used to predict T cell epitopes. IEDB recommended 2.22 method was used for prediction. More specifically, each field (except sequences and alleles) was satisfied with default recommended settings for prediction and selection of best peptides. Alleles were selected according to HLA targeting efficiency. A*24-02, A*68-01, A*68-02, B*39-01 and A*32-02 were used for the analysis as they were used previously by the researchers (19). T cell epitopes were filtered according to the two main approaches. The first includes the selection of all peptides with IC50 value less than 500 nm. The second approach was to select peptides with percentile ranks below 1% (20). At the end, epitopes were arranged according to their positions and redundancies in the epitopes were removed. Then resultant epitopes were joined by using the GPGPG spacer to prevent the formation of junctional epitopes.

Structure prediction for the protein

The structure of the target protein was not available in protein databank (PDB) database. Hence, homology modeling approach was used. This method can provide good models if the sequence identity is greater than 75% between the target protein and the template protein (21). The BLAST result against PDB database highlighted that the protein has 75.12% identity with 95% query coverage against PDB protein (6ACC). Homology modeling was performed using swiss model (https://swissmodel.expasy.org/) webserver using 6ACC template. The .pdb file was downloaded for further analysis

Highlighting epitopes on the structure

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 February 2020 doi:10.20944/preprints202002.0359.v1

The positions of predicted epitopes on 3D structure of predicted protein was observed using Pymol (22). The residues on the structure were selected accoding to the filtered epitopes and the epitopes were highlighted in red spheres.

Enriched analysis of the protein analysis

Enriched analysis of the protein was performed using ProtParam webserver (https://web.expasy.org/protparam/). This tool calculates various physical and chemical parameters for a given protein stored in Swiss-Prot or TrEMBL or a user-entered protein sequence. This tool calculates various physical and chemical properties for a given protein. Calculated Characteristics include theoretical pI, molecular weight, atomic composition, estimated half-life, extinction coefficient, aliphatic index, grand average of hydropathicity (GRAVY), and instability index. .

Results:

Structural Analysis of novel 2019-nCoV surface glycoprotein

The physicochemical properties of novel 2019-nCoV surface glycoprotein are predicted using Protparam which exhibited that it has 141.17847 kDa molecular weight, with 1273 amino acids. This reflects good antigenic nature of this protein. The isoelectric point (PI) of this protein was 6.24 that predicts it's almost neutral in nature. A total of 110 amino acid predicted to be negatively charged; however all 103 are positively charged amino acids. The instability index of this protein was 33.01 that categorize that this as stable protein. The formula of the protein that denotes Carbon (C), Hydrogen (H), Nitrogen (N), Oxygen (O), and Sulphur (S) is $C_{6336}H_{9770}N_{1656}O_{1894}S_{54}$. There are total numbers of 19710 atoms in the protein while the extinction coefficient of the protein is 148960. The predicted aliphatic index of the protein is 84.67. The grand average of hydropathicity (GRAVY) of this protein is -0.073. The estimated half life of this protein is 30 hrs in mammalian raticulocytes, >20 hrs in yeast and > 10 hrs in *Escherichia coli*.

Prediction of B cell epitopes and joining

The graph of predicted B cell epitopes based on average, minimum and maximum score of the amino acid residues is given in **Figure 1 (A).** The average score for the graph was -0.066, minimum score was -0.001 while maximum score was 2.291 as shown in **Figure 1 (A)**. The

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 February 2020 doi:10.20944/preprints202002.0359.v1

identified B cell epitopes are shown in **Table 1**. After filtering, total 13 epitopes were predicted based on the length of the epitope. The longest epitope has length of 27 amino acids with starting and ending positions of 461 and 487 respectively. Moreover, the smallest epitope has length of 10 amino acids with starting and ending positions of 44 and 53 respectively. The antigencity index was predicted using Kolaskar & Tongaonkar Antigenicity prediction tool and the threshold value was 1.041 **Figure 1(B)**. The plot of surface accessibility epitope was predicted using default parameters to Emini surface accessibility prediction having threshold value of 1.00 **Figure 1 (C)**. The hydrophilicity of the predicted epitopes was calculated using Parker hydrophilicity prediction tool having default parameters with threshold value 1.238 **Figure 1(D)**. These epitopes were joined by GPGPG spacer as shown in **Figure 2**. The total length of the B cell epitope after joining by GPGPG spacer is 295 as shown in the **Figure 2**.

Prediction of T cell epitopes and joining

Just like prediction of B cell epitopes, the prediction of T cell epitopes was performed using IEDB web server. Altogether five T cell epitopes were predicted after filtration of the data as mentioned in **Figure 3** and **Table 2**. As mentioned in material and method, the filtration of the predicted data was performed on the basis of percentile rank and IC50. All peptides with IC50 value less than 500 nm were selected. Moreover, the peptides with percentile ranks below the 1% were selected. After filtration of the data, the epitopes mentioned in Table 2 were combined using GPGPG spacer as shown in **Figure 3**. The total length of the T cell epitope after joining by GPGPG spacer is 85 as shown in the **Figure 3.** Moreover, among the predicted T cell epitopes, the shortest epitope length is 10 while longest epitope length is 14.

Structure prediction for the protein and highlighting the epitopes on the structure

The structure of the protein was predicted by homology modeling. The predicted structure is shown in **Figure 4**. The predicted epitopes using IEDB web server were highlighted on the structure of the protein using Pymol. Altogether, 13 predicted epitopes are highlighted on the structure of the protein as shown on **Figure 5**.

Discussion:

Emerging and pandemic zoonotic diseases have become an emerging public health concern for the world. Zoonotic diseases have also posed a threat to global health security. Almost 60 % of infectious diseases and 75 % of emerging infectious diseases are zoonotic in origin. Infectious diseases are responsible for 15.8 % deaths globally and 43.7 % deaths in Low-income countries. Each year, zoonotic diseases are account for 2.7 million human deaths worldwide (23, 24). Along with human deaths, zoonotic diseases also lead to heavy economic losses to the world specially; low income countries. For instance, last epidemic of Ebola in 2014 was responsible for 11316 deaths and $2.2 billion in terms of economic losses; moreover rabies is responsible for almost 59000 human deaths and $8.6 billion economic losses in single year (24, 25).

Emerging viruses poses a great challenge to medicine and science because little is known about them before they emerge. Human coronaviruses are one of the most important human viruses that pose great threat to global health. During past two decades, many coronaviruses caused serious problems in both animals and humans (26). In 2003, severe acute respiratory syndrome coronavirus (SARS-CoV) epidemic was lead to 299 deaths among 1755 infected people with death ratio of 6:1. The horseshoe bats were later traced as primary reservoir of SARS-CoV (27). Later on in 2012, Middle East respiratory syndrome coronavirus (MERS-CoV) was discovered from the patients died due to mysterious fatal disease (28). The dromedary camels were found to be the source of MERS-CoV spread in humans (29). Recently, in December 2019, a novel human corona virus (2019-nCoV) is discovered in Wuhan, China. More than 1000 people are infected with this virus with more than hundred deaths reported. Scientists have started working on therapeutic, vaccine and diagnostic kits for 2019-nCoV. Conventional approaches to make vaccine production and diagnostics require handsome amount of time. However, modern use of information technology in the discipline of virology has become very helpful to develop effective vaccine and diagnostics in a short time.

Bioinformatics is a good option to be used in development of vaccines and diagnostics for newly emerged viruses. The use of this approach can able to reduce the time and cost. Understanding the epitope/antibody interaction is the key to constructing potent vaccines and effective diagnostics. The surface glycoprotein of human corona viruses is the primary target for neutralizing antibodies thus a good target for vaccine development and diagnostics. The surface glycoprotein of 2019-nCoV consists of 1273 amino acids. The surface glycoprotein of 2019-

 doi:10.20944/preprints202002.0359.v1

nCoV is relatively smaller than MERS-CoV, has 1353 amino acids and comparatively closer to SARS-CoV, has 1255 amino acids (30). This predict that 2019-nCoV is closer to SARS-CoV moreover; recent studies also confirm that 2019-nCoV is 75% to 80 % genetically similar to SARS-CoV (5, 14). The surface glycoprotein mainly involve in the entry of the virus into the cell during virus replication cycle (31). The surface glycoproteins of coronaviruses are able to produce neutralizing antibodies, and block virus entry or neutralize the viral infection (32, 33).

B cell epitopes are the region of amino acids recognized by the B cell receptor or specific antibodies (34). The B cell epitope was predicted using Immune Epitope Database. After filtering, a total of 13 epitopes were predicted in surface glycoprotein of 2019-nCoV based on the length of the epitopes. The B cell epitopes were joined by the GPGPG linkers. The linkers are having a prominent role in the epitope based vaccine. The addition of GPGPG decreases the binding affinity around the core binding regions. Moreover, rich GPGPG linker is associated with beta turns with no effect on secondary and tertiary structure of the protein. The linkers not only prevent the immune processing of antigen but also prevent the generation of junctional epitopes. The linkers also helps presentation of selected epitopes through HLA-II (35, 36).

T cell epitope is a short linear sequence of amino acids of a specific antigen that is recognized by CD4 or /and CD8+ T cell receptors. Moreover, epitope has the capacity to stimulate cytotoxic and long lasting immune response. Therefore, T cell epitope mapping is important for the designing of effective epitope based vaccine and for affective therapeutics (37). After analysis of surface glycoprotein of 2019-nCoV, five T cell epitopes were predicted with the shortest epitope length was 10 while longest epitope length was 14. The epitopes were liked with GPGPG linker to make a peptide.

In this study the antigenic epitopes of the novel 2019-nCoV surface glycoprotein was predicted using bioinformatics analysis. By using this approach, we are able to reduce the time and cost of the production of effective diagnostics and vaccines against novel 2019-nCoV. The predicted epitopes should be tested for the vaccines, therapeutics and diagnostics in future studies. This analysis will be helpful to develop effective low cost epitope based vaccines and diagnostics to diagnose and control imminent novel 2019-nCoV challenge.

Authors Contributions Statement: Conceptualization: MAR, SR Data Mining and Analysis: MAR, IR, AA, AZ Paper writing: MAR, SR, MA, AZ Revision and correction: SR, TY, MR, AZ

Conflict of Interest Statement:

The authors declare no conflict of interest.

1. Stanley Perlman. Another Decade, Another Coronavirus. *Th e new england journal o f medicine* (2020). doi: DOI: 10.1056/NEJMe2001126.
2. Hui DS, E IA, Madani TA, Ntoumi F, Kock R, Dar O, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - The latest 2019 novel coronavirus outbreak in Wuhan, China. *Int J Infect Dis* (2020) 91:264-6. Epub 2020/01/19. doi: 10.1016/j.ijid.2020.01.009. PubMed PMID: 31953166.
3. Ning Dong XY, Lianwei Ye, Kaichao Chen, Edward Wai-Chi Chan, Mengsu Yang, Sheng Chen Genomic and protein structure modelling analysis depicts t 1 he origin and infectivity of 2019-nCoV, a new coronavirus which caused a pneumonia outbreak in Wuhan, China. *bioRxiv* (2020). doi: https://doi.org/10.1101/2020.01.20.913368
4. Parry J. China coronavirus: cases surge as official admits human to human transmission. *The british Medical Journal* (2020). doi: doi: 10.1136/bmj.m236.
5. Jasper Fuk-Woo Chan* SY, Kin-Hang Kok*, Kelvin Kai-Wang To*, Hin Chu*, Jin Yang, Fanfan Xing, Jieling Liu, Cyril Chik-Yan Yip, Rosana Wing-Shan Poon, Hoi-Wah Tsoi, Simon Kam-Fai Lo, Kwok-Hung Chan, Vincent Kwok-Man Poon, Wan-Mui Chan, Jonathan Daniel Ip,Jian-Piao Cai, Vincent Chi-Chung Cheng, Honglin Chen, Christopher Kim-Ming Hui, Kwok-Yung Yuen. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet* (2020):30154-9.
6. Wei Ji WW, Xiaofang Zhao, Junjie Zai , Xingguang Li. Homologous recombination within the spike glycoprotein of the newly identified coronavirus 2019-nCoV may boost cross-species transmission from snake to human. *Journal of Medical virology* (2020). doi: https://doi.org/10.1002/jmv.25682.
7. Cyranoski ECaD. Why snakes probably aren't spreading the new China virus. *Nature* (2020). doi: doi: 10.1038/d41586-020-00180-8.
8. Fung TS, Liu DX. Human Coronavirus: Host-Pathogen Interaction. *Annu Rev Microbiol* (2019) 73:529-57. Epub 2019/06/22. doi: 10.1146/annurev-micro-020518-115759. PubMed PMID: 31226023.
9. Schoeman D, Fielding BC. Coronavirus envelope protein: current knowledge. *Virol J* (2019) 16(1):69. Epub 2019/05/28. doi: 10.1186/s12985-019-1182-0. PubMed PMID: 31133031; PubMed Central PMCID: PMC6537279.
10. Ruch TR, Machamer CE. The coronavirus E protein: assembly and beyond. *Viruses* (2012) 4(3):363-82. Epub 2012/05/17. doi: 10.3390/v4030363. PubMed PMID: 22590676; PubMed Central PMCID: PMC3347032.

11. Kirchdoerfer RN, Cottrell CA, Wang N, Pallesen J, Yassine HM, Turner HL, et al. Pre-fusion structure of a human coronavirus spike protein. *Nature* (2016) 531(7592):118-21. Epub 2016/03/05. doi: 10.1038/nature17200. PubMed PMID: 26935699; PubMed Central PMCID: PMC4860016.
12. McBride R vZM, Fielding BC. The coronavirus nucleocapsid is a multifunctional protein. *Viruses* (2014) 6(8):2997-3018.
13. Masters PS. The molecular biology of coronaviruses. *Adv Virus Res* (2006) 66:193-292. Epub 2006/08/01. doi: 10.1016/S0065-3527(06)66005-3. PubMed PMID: 16877062.
14. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* (2020). Epub 2020/01/25. doi: 10.1056/NEJMoa2001017. PubMed PMID: 31978945.
15. Raza S, Siddique K, Rabbani M, Yaqub T, Anjum AA, Ibrahim M, et al. In silico analysis of four structural proteins of aphthovirus serotypes revealed significant B and T cell epitopes. *Microb Pathog* (2019) 128:254-62. Epub 2019/01/08. doi: 10.1016/j.micpath.2019.01.007. PubMed PMID: 30616000.
16. Oany AR, Emran AA, Jyoti TP. Design of an epitope-based peptide vaccine against spike protein of human coronavirus: an in silico approach. *Drug Des Devel Ther* (2014) 8:1139-49. Epub 2014/09/05. doi: 10.2147/DDDT.S67861. PubMed PMID: 25187696; PubMed Central PMCID: PMC4149408.
17. Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Res* (2006) 2:2. Epub 2006/04/26. doi: 10.1186/1745-7580-2-2. PubMed PMID: 16635264; PubMed Central PMCID: PMC1479323.
18. Livingston B, Crimi C, Newman M, Higashimoto Y, Appella E, Sidney J, et al. A rational strategy to design multiepitope immunogens based on multiple Th lymphocyte epitopes. *Journal of immunology* (2002) 168(11):5499-506. doi: 10.4049/jimmunol.168.11.5499. PubMed PMID: 12023344.
19. Hertz T, Oshansky CM, Roddam PL, DeVincenzo JP, Caniza MA, Jojic N, et al. HLA targeting efficiency correlates with human T-cell response magnitude and with mortality from influenza A infection. *Proc Natl Acad Sci U S A* (2013) 110(33):13492-7. doi: 10.1073/pnas.1221555110. PubMed PMID: 23878211; PubMed Central PMCID: PMC3746844.
20. Fleri W, Paul S, Dhanda SK, Mahajan S, Xu X, Peters B, et al. The Immune Epitope Database and Analysis Resource in Epitope Discovery and Synthetic Vaccine Design. *Frontiers in immunology* (2017) 8:278. doi: 10.3389/fimmu.2017.00278. PubMed PMID: 28352270; PubMed Central PMCID: PMC5348633.
21. Rodriguez R, Chinea G, Lopez N, Pons T, Vriend G. Homology modeling, model and software evaluation: three related resources. *Bioinformatics* (1998) 14(6):523-8. doi: 10.1093/bioinformatics/14.6.523. PubMed PMID: 9694991.
22. WL D. Pymol: an open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography* (2002) 40:82–92.
23. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. *Nature* (2008) 451(7181):990-3. Epub 2008/02/22. doi: 10.1038/nature06536. PubMed PMID: 18288193; PubMed Central PMCID: PMC5960580.
24. Salyer SJ, Silver R, Simone K, Barton Behravesh C. Prioritizing Zoonoses for Global Health Capacity Building-Themes from One Health Zoonotic Disease Workshops in 7 Countries, 2014-2016. *Emerg Infect Dis* (2017) 23(13). Epub 2017/11/21. doi: 10.3201/eid2313.170418. PubMed PMID: 29155664; PubMed Central PMCID: PMC5711306.

25. Nel LH. The role of non-governmental organisations in controlling rabies: the Global Alliance for Rabies Control, Partners for Rabies Prevention and the Blueprint for Rabies Prevention and Control. *Rev Sci Tech* (2018) 37(2):751-9. Epub 2019/02/13. doi: 10.20506/rst.37.2.2838. PubMed PMID: 30747111.

26. Lau SK, Chan JF. Coronaviruses: emerging and re-emerging pathogens in humans and animals. *Virol J* (2015) 12:209. Epub 2015/12/23. doi: 10.1186/s12985-015-0432-z. PubMed PMID: 26690088; PubMed Central PMCID: PMC4687117.

27. Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* (2013) 503(7477):535-8. Epub 2013/11/01. doi: 10.1038/nature12711. PubMed PMID: 24172901; PubMed Central PMCID: PMC5389864.

28. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* (2012) 367(19):1814-20. Epub 2012/10/19. doi: 10.1056/NEJMoa1211721. PubMed PMID: 23075143.

29. Haagmans BL, Al Dhahiry SH, Reusken CB, Raj VS, Galiano M, Myers R, et al. Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation. *Lancet Infect Dis* (2014) 14(2):140-5. Epub 2013/12/21. doi: 10.1016/S1473-3099(13)70690-X. PubMed PMID: 24355866.

30. Petit CM, Melancon JM, Chouljenko VN, Colgrove R, Farzan M, Knipe DM, et al. Genetic analysis of the SARS-coronavirus spike glycoprotein functional domains involved in cell-surface expression and cell-to-cell fusion. *Virology* (2005) 341(2):215-30. Epub 2005/08/16. doi: 10.1016/j.virol.2005.06.046. PubMed PMID: 16099010.

31. Li F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol* (2016) 3(1):237-61. Epub 2016/09/01. doi: 10.1146/annurev-virology-110615-042301. PubMed PMID: 27578435; PubMed Central PMCID: PMC5457962.

32. Kim MH, Kim HJ, Chang J. Superior immune responses induced by intranasal immunization with recombinant adenovirus-based vaccine expressing full-length Spike protein of Middle East respiratory syndrome coronavirus. *PLoS One* (2019) 14(7):e0220196. Epub 2019/07/23. doi: 10.1371/journal.pone.0220196. PubMed PMID: 31329652; PubMed Central PMCID: PMC6645677.

33. Yang ZY, Kong WP, Huang Y, Roberts A, Murphy BR, Subbarao K, et al. A DNA vaccine induces SARS coronavirus neutralization and protective immunity in mice. *Nature* (2004) 428(6982):561-4. Epub 2004/04/02. doi: 10.1038/nature02463. PubMed PMID: 15024391.

34. Yu H, Jiang LF, Fang DY, Yan HJ, Zhou JJ, Zhou JM, et al. Selection of SARS-coronavirus-specific B cell epitopes by phage peptide library screening and evaluation of the immunological effect of epitope-based peptides on mice. *Virology* (2007) 359(2):264-74. Epub 2006/10/24. doi: 10.1016/j.virol.2006.09.016. PubMed PMID: 17055022.

35. Beatriz Meza FA, Arturo Pedro Sierra-Beltrán, Javier Torres and Carlos Angulo. A novel design of a multi-antigenic, multistage and multi-epitope vaccine against helicobacter pylori: an in silico approach. *Infection, Genetics and Evolution* (2017) 49:309-17.

36. Navid Nezafat ZK, Mahboobeh Eslami, Milad Mohkam, Sanam Zandian, Younes Ghasemi. Designing an efficient multi-epitope peptide vaccine against Vibrio cholerae via combined immunoinformatics and protein interaction based approaches. *Computational Biology and Chemistry* (2016) 62:82-95.

37. Tarek A. Ahmad AEE, Laila H. El-Sayed ,. T-cell epitope mapping for the design of powerful vaccines. *Vaccine Reports* (2016) 6:13-22.

 doi:10.20944/preprints202002.0359.v1

Figure Legends

Fig 1: The graphs were generated using IEDB resource analysis by keeping the parameter at Default. The Y-axes depict correspondent score for all residues while the X axis represents their equivalent positions. (A) Plot of predicted bepipred B cell linear epitopes using threshold value 0.5 (B) Plot of Kolaskar & Tongaonkar Antigenicity prediction using online tools with threshold value 1.041 (C) The plot of surface accessibility epitope was predicted using default parameters to Emini surface accessibility prediction having threshold value of 1.00 (D) The hydrophilicity of the predicted epitopes were calculated using Parker hydrophilicity prediction tool having default parameters with threshold value 1.238

Figure 2: Joining of B cell epitopes using GPGPG spacer. A total of 13 epitopes were predicted and joined using the spacer. The longest predicted epitope has a length of 27 amino acids while the smallest epitope has length of 10 amino acids. Altogether, the total length of the B cell epitope after joining by GPGPG spacer is 295 as shown in the figure

Figure 3: Joining of T cell epitopes using GPGPG spacer. The predicted peptides are based on IC50 value less than 500 nm and percentile ranks below the 1%. The epitopes were combined using GPGPG spacer as shown in the figure. The shortest predicted T cell epitope length is 10 while longest epitope length is 14 amino acids. The total length of the T cell epitope after joining by GPGPG spacer is 85 amino acids.

Figure 4: Structure of predicted protein using homology modeling. The structure of the target protein was not available in protein databank (PDB) database. Hence, homology modeling approach was used to predict the structure using 6ACC template. The target protein has 75.12% identity with 95% query coverage against the template. The structure was predicted using swissmodel web server.

Figure 5: Highlighting B cell epitopes on the structure of the protein. Altogether, 13 predicted epitopes are highlighted on the structure of the protein as shown in the figure. The

 doi:10.20944/preprints202002.0359.v1

positions of predicted epitopes on 3D structure of predicted protein was observed using Pymol software and highlighted red in colour.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 February 2020 doi:10.20944/preprints202002.0359.v1

Table 1: Predicted B cell epitopes

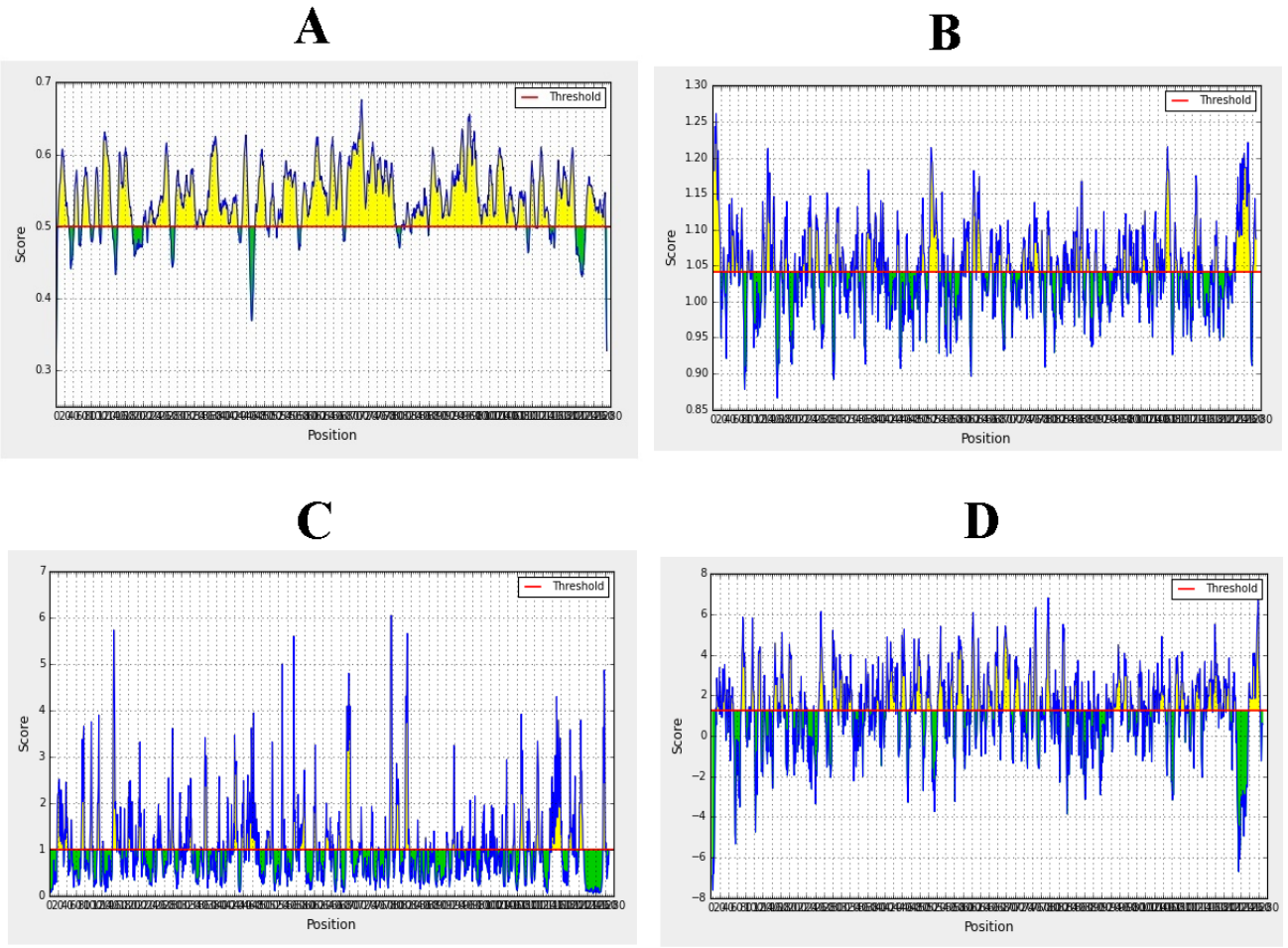
Sr No.	Start	End	Peptide	Length
1	6	30	VLLPLVSSQCVNLTTRTQLPPAYTN	25
2	44	53	RSSVLHSTQD	10
3	62	81	VTWFHAIHVSGTNGTKRFDN	20
4	90	101	VYFASTEKSNII	12
5	107	131	GTTLDSKTQSLLIVNNATNVVIKVC	25
6	327	340	VRFPNITNLCPFGE	14
7	431	447	GCVIAWNSNNLDSKVGG	17
8	461	487	LKPFERDISTEIYQAGSTPCNGVEGFN	27
9	496	508	GFQPTNGVGYQPY	13
10	807	818	PDPSKPSKRSFI	12
11	1066	1087	TYVPAQEKNFTTAPAICHDGKA	22
12	1096	1107	VSNGTHWFVTQR	12
13	1114	1139	IITTDNTFVSGNCDVVIGIVNNTVYD	26

Table 2: Predicted T cell epitopes on the basis of percentile rank and ic50

allele	start	end	length	peptide	method	Percentile Rank	ann_ic50	ann_rank
HLA-A*24:02	1	13	13	PTWRVYSTGSNVF	ann	0.57	352.54	0.57
HLA-A*68:01	14	23	10	HVTYVPAQEK	Consensus (ann/smm)	0.83	20.9	0.16
HLA-A*24:02	24	37	14	KVGGNYNYLYRLFR	ann	0.6	386.63	0.6
HLA-A*68:01	40	53	14	LCFTNVYADSFVIR	ann	0.49	55.79	0.49
HLA-A*24:02	54	67	14	AYYVGYLQPRTFLL	ann	0.17	111.15	0.17

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 February 2020 doi:10.20944/preprints202002.0359.v1

Figure 1



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 February 2020 doi:10.20944/preprints202002.0359.v1

478

479

480 **Figure 2**

481

482

483

484 **Figure 3**

485

486

487

488

489

490

491

492

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 February 2020 doi:10.20944/preprints202002.0359.v1

493

494 **Figure 4**

495

496

497

498

499

500

501

502

503

504

505

506

Figure 5

508