bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Multi-epitope vaccine design using an immunoinformatics approach for 2019 novel coronavirus in China (SARS-CoV-2)

Ye Feng[1,2†], Min Qiu[3†], Shengmei Zou[1,2†], Yun Li[4], Kai Luo[3], Rongchang Chen[3], Yingqiang Sun[3], Kui Wang[3], Xinlei Zhuang[5], Shanshan Zhang[3,6], Shuqing Chen[3,5,6*], Fan Mo[3,5,7*]

[1] Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China; [2] Institute for Translational Medicine, Zhejiang University School of Medicine, Hangzhou, China; [3] Hangzhou Neoantigen Therapeutics Co., Ltd., Hangzhou, China; [4] Zhejiang Forest Resources Monitoring Center, Hangzhou, China; [5] College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, China; [6] Zhejiang California International Nanosystems Institute, Zhejiang University, Hangzhou, China; [7] Vancouver Prostate Centre, University of British Columbia, Vancouver, BC, Canada.

[†] These authors contributed equally to this work.

• ***Corresponding authors:** Address: 866 Yuhangtang Rd, Hangzhou 310058, China. Email: Fan Mo (mofan_hz@163.com) or Shuqing Chen (chenshuqing@zju.edu.cn)

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

A new coronavirus SARS-CoV-2, recently discovered in Wuhan, China, has caused over 74000 infection cases and 2000 deaths. Due to the rapidly growing cases and the unavailability of specific therapy, there is a desperate need for vaccines to combat the epidemic of SARS-CoV-2. In the present study, we performed an in silico approach based on the available virus genome to identify the antigenic B-cell epitopes and human-leukocyte-antigen (HLA) restricted T-cell epitopes. A total of 61 B-cell epitopes were initially identified, 19 of which with higher potential immunogenicity were used for vaccine design. 499 T-cell epitopes were predicted that showed affinity with the 34 most popular HLA alleles in Chinese population. Based on these epitopes, 30 vaccine candidates were designed and inspected against safety risks, including potential toxicity, human homologous, pharmaceutical peptides and bioactive peptides. Majority of vaccine peptides contained both B-cell and T-cell epitopes, which may interact with the most prevalent HLA alleles accounting for ~99% of Chinese population. Docking analysis showed stable hydrogen bonds of epitopes with their corresponding HLA alleles. In conclusion, these putative antigenic peptides may elicit the resistance response to the viral infection. In vitro and in vivo experiments are required to validate the effectiveness of these peptide vaccine.

Keywords: SARS-CoV-2, epitope, immunoinformatics, vaccine

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1. Introduction

Since December 2019, Wuhan City, Hubei Province, China has become the center of an outbreak of viral lung infections caused by a new type of coronavirus, SARS-CoV-2 (previously named 2019-nCoV by the World Health Organization) (Gorbalenya, 2020; Joseph T Wu*, 2020; Organization, 2020; Perlman, 2020).The outbreak has so far infected over 74000 patients, and has spread to approximately 30 countries/regions worldwide. Comparisons of the genome sequences of SARS-CoV-2 with other virus has shown 79.5% and 96% similarities at nucleotide level to SARS-CoV and bat coronaviruses, respectively (Zhou et al., 2020), which suggested its probable origin in bats (Benvenuto et al., 2020). The main clinical manifestations of SARS-CoV-2 patients are fever (≥38ºC), dry cough, low or normal peripheral white blood cell count, and low lymphocyte count, known as novel coronavirus-infected pneumonia (NCIP) or coronavirus disease 2019 (COVID19) (Huang et al., 2020).

Currently, there is no approved therapeutics or vaccines available for the treatment of COVID19 (Chen et al.). Due to lack of anti-viral drugs or vaccines, control measures have been relying on the rapid detection and isolation of symptomatic cases (Chen et al.). In this context, a safe and efficacious vaccine is urgently required. Traditional approaches for developing vaccines waste much time in isolating, inactivating and injecting the microorganisms (or portions of them) that cause disease. Fortunately, computation-based method enable us to start from analysis of viral genome, without the need to grow pathogens and therefore speeding up the entire process. Complete genome sequencing of SARS-CoV-2 has finished and paved the way for the vaccine development (Chen et al.). The genome of SARS-CoV-2 encodes the spike protein, the membrane protein, the envelope protein, the nucleocapsid protein and a few replication and

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

transcription-related enzymes. Given the lack of repairing mechanism of RNA virus replicase complex, mutations are prone to occur during virus replication. The 4% nucleotide difference of the virus isolated from Rhinolophus to that from human suggests that SARS-CoV-2 mutates rapidly to achieve the host conversion (Jiayuan Chen, 2020). Like SARS-CoV, SARS-CoV-2 uses its receptor binding domain (RBD) on the spike protein to bind to the host's angiotensin-converting enzyme 2 (ACE2) (Chen et al.; Dong et al., 2020; Tian et al., 2020; Zhou et al., 2020). In terms of binding capacity, the RBD of SARS-CoV-2 is much stronger than SARS-CoV and is considered to be between HKU3-4 which cannot bind to ACE2 receptor and rSHC014 which presents the largest binding energy (Gralinski and Menachery, 2020; Wrapp et al., 2020; Zhao et al., 2020). Consequently, the SARS-CoV-2 vaccine can be developed targeting the structural proteins, and in particular, the RBD region, following the strategy for the SARS-CoV vaccine development (Babcock et al., 2004; Buchholz et al., 2004; He et al., 2005; Saif, 1993; Tian et al., 2020).

An ideal vaccine may contain both B-cell epitopes and T-cell epitopes, with combination of which vaccine is able to either induce specific humoral or cellular immune against pathogens efficiently (Purcell et al., 2007). Since the development of a peptide vaccine against the virus causing foot-and-mouth disease (Adam et al., 1978), the establishment of peptide synthesis method by Lerner et al. (Lerner et al., 1981), along with the advent of a peptide vaccine design combining T-cell and B-cell epitopes, has accelerated the vaccine development. In the present study, we followed this in silico approach to identify the potential B-cell and T-cell epitope(s) from the spike, envelope and membrane proteins that could be used as promising vaccines against SARS-CoV-2.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2. Materials and Methods

2.1. Data retrieval

The genome sequence of SARS-CoV-2 isolate Wuhan-Hu-1 was retrieved from the NCBI database under the accession number MN908947. Gene and protein sequences were acquired according to the annotation. In particular, the RBD region for the spike protein was referred to as the fragment from 347 to 520 amino acid (aa) (Lu et al., 2020).

2.2. B-cell epitope prediction

The online tool in IEDB (Immune-Epitope-Database And Analysis-Resource) was used for the analysis of the conserved regions of the candidate epitopes (Vita et al., 2015). Prediction of linear B-cell epitopes was performed through Bepipred software (Jespersen et al., 2017). The antigenic sites were determined with Kolaskar method (Chen et al., 2007). The surface accessible epitopes were predicted by Emini tool (Almofti et al., 2018).

2.3. T-cell epitope prediction

The sequences of structural proteins were split into small fragments with a length of 9aa; their binding affinity with the 34 most prevalent HLA alleles (>1% in the Chinese population according to data obtained from the website for China Marrow Donor Program (CMDP)) was predicted using netMHCpan (Hoof et al., 2009) and our in-house prediction software iNeo-Pred, respectively. iNeo-Pred was trained on a large immune-peptide dataset, and achieved a better performance in predicting binding affinity of epitopes to specific HLA alleles. Only the epitopes

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

predicted by both tools were selected. Next, for each epitope, a HLA score was calculated based on the frequencies of binding HLA alleles in Chinese population, which will be used as metrics to select better candidates for downstream analysis.

2.4. Vaccine peptide design

The vaccine peptides were designed by our in-house tool iNeo-Design. First, the selected B-cell epitopes and their adjacent T-cell epitopes were bridged to form candidate peptides with length no more than 30aa. Meanwhile, to facilitate the peptide synthesis, vaccine peptide sequences were optimized based on their hydrophobicity and acidity. To minimize the safety risk, peptides that contained toxicity potential, human homologous region (full-length matches and identity > 95%), or bioactive peptide were discarded.

Besides the vaccine peptides containing both B-cell epitopes and T-cell epitopes, iNeo-Design also utilized all predicted T-cell epitopes to generate T-cell epitopes-only vaccine peptides. For each vaccine candidate, the epitope counts and HLA score reflecting the population coverage were calculated. Vaccine candidates with the higher epitope counts and HLA score were considered to be preferable for the downstream analysis.

2.5. Structural analysis

The online server swiss-model was used to predict the 3D protein structures of viral proteins and HLA molecules (Waterhouse et al., 2018). The online server PEP_FOLD was used to predict T-cell epitopes' structures (Lamiable et al., 2016). To display the interaction between T-cell epitopes and HLA molecules, T-cell epitope models were docked to HLA molecules using MDockPep (Xu et al., 2018). All predicted structures or models were decorated and displayed by

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the open source version of pymol program (https://github.com/schrodinger/pymol-open-source).

Ethics

N/A.

Data availability

N/A.

3. Results and Discussion

3.1. Prediction of B-cell epitopes

During the immune response against viral infection, B-cell takes in viral epitopes to recognize viruses and activates defense responses. Recognition of B-cell epitopes depends on antigenicity, accessibility of surface and predictions of linear epitope (Fieser et al., 1987). A total of 61 B-cell epitopes were predicted, which seemed preferentially located within certain regions of each gene (Figure 1; Figure 2; Table S1). Only 19 epitopes were exposed on the surface of the virion and had a high antigenicity score, indicating their potentials in initiating immune response. Therefore, they were considered to be promising vaccine candidates against B-cells. Among the 19 epitopes, 17 were longer than 14 residues and located in the spike protein that contained RBD and functioned in host cell binding (Table 1). The average Emini score for the 19 epitopes was 2.744, and the average for Kolaskar (antigenicity) score was 1.015. Two epitopes were located within the RBD region, while the one with the highest Kolaskar score (1.059), 1052-FPQSAPH-1058, was located at position 1052aa of the spike protein.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3.2. Prediction of T-cell epitopes

The immune response of T-cell is considered as a long lasting response compared to B-cell where the antigen might easily escape the antibody memory response (Black et al., 2010). Moreover, the CD8+ T and CD4+ T-cell responses play a major role in antiviral immunity. It is therefore important to design vaccines that can induce T-cell's immune response (Sesardic, 1993). A total of 499 T-cell epitopes were predicted on the spike protein (378 epitopes), the membrane protein (90 epitopes) and the envelop protein (31 epitopes); 48 of the 378 epitopes for the spike protein were located in the RBD region (Figure 1; Table 2; Table S2). There is no preference in certain genes or regions for T-cell epitope generation; no biased distribution of T-cell epitopes among HLA types were observed either. Among all T-cell epitopes, the epitope 869-MIAQYTSAL-877 in the spike protein was predicted to be able to bind to 17 HLA alleles. Most of the HLA alleles included in the present study were covered by these vaccine candidates, which suggested a wide population coverage.

In terms of the distribution of the predicted epitopes against different HLA haplotypes, no significant differences were observed among different HLA haplotypes (Table S3). There were 287, 208 and 195 epitopes predicted to be able to bind to HLA-A, HLA-B and HLA-C haplotypes, respectively. For the most popular five HLA types (HLA-A*11:01, HLA-A*24:02, HLA-C*07:02, HLA-A*02:01 and HLA-B*46:01), the counts for epitopes with binding affinity were 51, 49, 115, 48 and 58.

3.3. Multi-epitope vaccine design

Based on the 19 B-cell epitopes and their 121 adjacent T-cell epitopes, 17 candidate vaccine

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

peptides that contained both B-cell and T-cell epitopes were generated by our in-house software iNeo-Design. Most of the 17 candidate vaccine peptides contained one B-cell epitopes, except for AVEQDKNTQEVFAQVKQIYKTPPIKDFGG, which involved two B-cell epitopes and eight T-cell epitopes, and AKNLNESLIDLQELGKYEQYIKWPWYIWKK, which contained two B-cell epitopes and 6 T-cell epitopes. By comparison, the vaccine peptide FKNLREFVFKNIDGYFKIYSKHTPINLV had the largest count of T-cell epitopes, whereas the vaccine peptide SYGFQPTNGVGYQPYRVVVLSFELLHAPAT showed the highest HLA score, indicating their wide population coverage and promising efficacy.

In addition to the vaccine candidates involved both B-cell and T-cell epitopes, we also analysed the entire 499 core T-cell epitopes to generate another 102 vaccine peptides containing T-cell epitopes only. Based on both the epitope counts and HLA score, we eventually selected 13 T-cell epitopes-only vaccine peptides.

Taken together, a total of 30 peptide vaccine candidates were designed (Table 3). 26 of them were from the spike protein, two from the membrane protein and two from the envelop protein. Five peptides were located in the RBD region, indicating they were likely to induce the production of neutralizing antibody. The 30 vaccine peptides covered all structural proteins that may induce immune response against SARS-CoV-2 in theory; and the multi-peptide strategy we applied would better fit the genetic variability of the human immune system and reduce the risk of pathogen's escape through mutation (Skwarczynski and Toth, 2016).

3.4. Interaction of predicted peptides with HLA alleles

To further inspect the binding stability of T-cell epitopes against HLA alleles, the T-cell epitopes

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

involved in the above designed vaccine peptides were selected to conduct an interaction analysis. Figure 3 illustrated the docking results against the most popular HLA types for the two epitopes from vaccines peptide 25 and 27 (Table 3; Table 4), which showed relatively higher HLA score. The MDockPep scores were between -148 ~ -136, indicating that the predicted crystal structures were stable. All epitopes were docked inside the catalytic pocket of the receptor protein. In particular, the epitope 1220-FIAGLIAIV-1228 from the spike protein possessed 2-5 stable hydrogen bonds with the HLA alleles; the epitope 4-FVSEETGTL-12 from the envelop protein possessed 4~5 stable hydrogen bonds (Table 4). Taken together, the epitopes included in our vaccine peptides can interact with the given HLA alleles by in silico prediction.

4. Conclusions

Vaccine design using an in silico prediction method is highly appreciated as it selects specific epitopes in proteins than conventional methods. In the present study, this reverse vaccinology approach was adopted to identify surface-exposed peptides, instead of targeting the whole pathogen which is obviously less efficient and effective. Immunogenic regions of antigenic epitopes of all proteins encoded by SARS-CoV-2 were screened to identify potential vaccine candidates. As a result, 17 vaccine peptides involving both T-cell epitopes and B-cell epitopes as well as 13 vaccine peptides involving T-cell epitopes only were successfully designed. These multi-epitope vaccines provided excellent candidates for the development of vaccines against SARS-CoV-2. Various formulation of vaccine can be rapidly manufactured, such as peptide, DNA and mRNA. However in vitro and in vivo trials are required to achieve the effectiveness of these

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

vaccine peptides.

Acknowledgements

This work was supported by Zhejiang Provincial Natural Science Foundation of China.

Author contributions

SC and FM conceived and designed the project. MQ, SZ, KL, RC, YS, KW, XZ and SZ analysed the data. YF and YL wrote the initial draft. All authors revised and approved the final manuscript.

Declarations of interest: none

References

Adam, K.-H., Kaaden, O., Strohmaier, K., 1978. Isolation of immunizing cyanogen bromide-peptides of foot-and-mouth disease virus. Biochemical and biophysical research communications 84, 677-683.

Almofti, Y.A., Abd-elrahman, K.A., Gassmallah, S.A.E., Salih, M.A., 2018. Multi Epitopes Vaccine Prediction against Severe Acute Respiratory Syndrome (SARS) Coronavirus Using Immunoinformatics Approaches. American Journal of Microbiological Research 6, 94-114.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

224 Babcock, G.J., Esshaki, D.J., Thomas, W.D., Ambrosino, D.M., 2004. Amino acids 270 to 510
225 of the severe acute respiratory syndrome coronavirus spike protein are required for interaction
226 with receptor. J Virol 78, 4552-4560.

227 Benvenuto, D., Giovannetti, M., Ciccozzi, A., Spoto, S., Angeletti, S., Ciccozzi, M., 2020. The
228 2019-new coronavirus epidemic: evidence for virus evolution. Journal of Medical Virology.

229 Black, M., Trent, A., Tirrell, M., Olive, C., 2010. Advances in the design and delivery of peptide
230 subunit vaccines with a focus on toll-like receptor agonists. Expert review of vaccines 9,
231 157-173.

232 Buchholz, U.J., Bukreyev, A., Yang, L.J., Lamirande, E.W., Murphy, B.R., Subbarao, K.,
233 Collins, P.L., 2004. Contributions of the structural proteins of severe acute respiratory
234 syndrome coronavirus to protective immunity. P Natl Acad Sci USA 101, 9804-9809.

235 Chen, J., Liu, H., Yang, J., Chou, K.-C., 2007. Prediction of linear B-cell epitopes using amino
236 acid pair antigenicity scale. Amino acids 33, 423-428.

237 Chen, Y., Liu, Q., Guo, D., Emerging coronaviruses: genome structure, replication, and
238 pathogenesis. Journal of Medical Virology.

239 Dong, N., Yang, X., Ye, L., Chen, K., Chan, E.W.-C., Yang, M., Chen, S., 2020. Genomic and
240 protein structure modelling analysis depicts the origin and infectivity of 2019-nCoV, a new
241 coronavirus which caused a pneumonia outbreak in Wuhan, China. bioRxiv.

242 Fieser, T.M., Tainer, J.A., Geysen, H.M., Houghten, R.A., Lerner, R.A., 1987. Influence of
243 protein flexibility and peptide conformation on reactivity of monoclonal anti-peptide antibodies

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

with a protein alpha-helix. Proceedings of the National Academy of Sciences 84, 8568-8572.

Gorbalenya, A.E., 2020. Severe acute respiratory syndrome-related coronavirus – The species and its viruses, a statement of the Coronavirus Study Group. bioRxiv, 2020.2002.2007.937862.

Gralinski, L.E., Menachery, V.D., 2020. Return of the Coronavirus: 2019-nCoV. Viruses 12, 135.

He, Y.X., Lu, H., Siddiqui, P., Zhou, Y.S., Jiang, S.B., 2005. Receptor-binding domain of severe acute respiratory syndrome coronavirus spike protein contains multiple conformation-dependent epitopes that induce highly potent neutralizing antibodies. J Immunol 174, 4908-4915.

Hoof, I., Peters, B., Sidney, J., Pedersen, L.E., Sette, A., Lund, O., Buus, S., Nielsen, M., 2009. NetMHCpan, a method for MHC class I binding prediction beyond humans. Immunogenetics 61, 1.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet.

Jespersen, M.C., Peters, B., Nielsen, M., Marcatili, P., 2017. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic acids research 45, W24-W29.

Jiayuan Chen, J.S., Tungon Yau ,Chang Liu , Xin L, Qiang Zhao, Jishou Ruan , Gao Shan.,

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2020. Bioinformatics analysis of the Wuhan 2019 human coronavirus genome.

Joseph T Wu*, K.L., Gabriel M Leung, 2020. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet.

Lamiable, A., Thevenet, P., Rey, J., Vavrusa, M., Derreumaux, P., Tuffery, P., 2016. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. Nucleic Acids Res 44, W449-454.

Lerner, R.A., Green, N., Alexander, H., Liu, F.-T., Sutcliffe, J.G., Shinnick, T.M., 1981. Chemically synthesized peptides predicted from the nucleotide sequence of the hepatitis B virus genome elicit antibodies reactive with the native envelope protein of Dane particles. Proceedings of the National Academy of Sciences 78, 3403-3407.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., Chen, J., Meng, Y., Wang, J., Lin, Y., Yuan, J., Xie, Z., Ma, J., Liu, W.J., Wang, D., Xu, W., Holmes, E.C., Gao, G.F., Wu, G., Chen, W., Shi, W., Tan, W., 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet.

Organization, W.H., 2020. Surveillance case definitions for human infection with novel coronavirus (nCoV): interim guidance v1, January 2020. , in: Organization, W.H. (Ed.). World Health Organization, Geneva.

Perlman, S., 2020. Another Decade, Another Coronavirus. New England Journal of Medicine.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

284 Purcell, A.W., McCluskey, J., Rossjohn, J., 2007. More than one reason to rethink the use of
285 peptides in vaccine design. Nat Rev Drug Discov 6, 404-414.

286 Saif, L.J., 1993. Coronavirus Immunogens. Vet Microbiol 37, 285-297.

287 Sesardic, D., 1993. Synthetic peptide vaccines. Journal of medical microbiology 39, 241-242.

288 Skwarczynski, M., Toth, I., 2016. Peptide-based synthetic vaccines. Chemical science 7,
289 842-854.

290 Tian, X., Li, C., Huang, A., Xia, S., Lu, S., Shi, Z., Lu, L., Jiang, S., Yang, Z., Wu, Y., 2020.
291 Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human
292 monoclonal antibody. bioRxiv.

293 Vita, R., Overton, J.A., Greenbaum, J.A., Ponomarenko, J., Clark, J.D., Cantrell, J.R., Wheeler,
294 D.K., Gabbard, J.L., Hix, D., Sette, A., 2015. The immune epitope database (IEDB) 3.0.
295 Nucleic acids research 43, D405-D412.

296 Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T.,
297 de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., 2018. SWISS-MODEL:
298 homology modelling of protein structures and complexes. Nucleic Acids Res 46, W296-W303.

299 Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.-L., Abiona, O., Graham, B.S.,
300 McLellan, J.S., 2020. Cryo-EM Structure of the 2019-nCoV Spike in the Prefusion
301 Conformation. bioRxiv, 2020.2002.2011.944462.

302 Xu, X., Yan, C., Zou, X., 2018. MDockPeP: An ab-initio protein-peptide docking server. Journal
303 of computational chemistry 39, 2409-2413.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

304 Zhao, Y., Zhao, Z., Wang, Y., Zhou, Y., Ma, Y., Zuo, W., 2020. Single-cell RNA expression
305 profiling of ACE2, the putative receptor of Wuhan 2019-nCov. bioRxiv.

306 Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B.,
307 Huang, C.-L., 2020. Discovery of a novel coronavirus associated with the recent pneumonia
308 outbreak in humans and its potential bat origin. bioRxiv.

309

310

311

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

312 **Figure legends**

313 **Fig. 1.** Distribution of B-cell and T-cell epitopes. The outermost circle (light blue) stands for the
314 T-cell epitope count. The 2rd outer circle stands for Emini (in red) and Kolaskar (in green) score
315 used to evaluate the B-cell epitopes. The 3rd circle marked the name of the viral proteins. The
316 4th-6th circles stands for HLA-A (in blue), HLA-B (in green), and HLA-C (in yellow) scores; the
317 points closer to the center indicates a lower score.

318

319 **Fig. 2.** Locations of the recognized B cell epitopes on the viral spike protein (a), envelop protein
320 (b) and membrane protein (c). The transparent cartoon models display the predicted 3D structure;
321 the colorful balls marks the position of the recognized epitopes.

322

323 **Fig. 3.** Interaction between the predicted peptides (by yellow sticks) and different HLA alleles (by
324 green cartoons). Amino acids were labeled adjacent to the contact sites. Table 3 displays the
325 detailed docking information.

326

327

328

329

330

331

332

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

333

334

335

336

337

338

339

Table 1. B-cell epitope candidates

Epitope	Protein	Start	End	Peptipe	Emini	Kolaskar
B1	Spike	19	43	TTRTQLPPAYTNSFTRGVYYPDKVF	6.424	1.028
B2	Spike	90	99	VYFASTEKSN	1.573	1.019
B3	Spike	206	209	KHTP	2.463	1.002
B4	Spike	405	430	DEVRQIAPGQTGKIADYNYKLPDDFT	5.81	1.001
B5	Spike	494	507	SYGFQPTNGVGYQP	1.553	1.02
B6	Spike	671	688	CASYQTQTNSPRRARSVA	3.531	1.027
B7	Spike	771	782	AVEQDKNTQEVF	2.342	1.011
B8	Spike	787	799	QIYKTPPIKDFGG	1.465	1.006
B9	Spike	805	816	ILPDPSKPSKRS	4.69	1.019
B10	Spike	1052	1058	FPQSAPH	1.381	1.059
B11	Spike	1068	1091	VPAQEKNFTTAPAICHDGKAHFPR	1.063	1.03
B12	Spike	1108	1123	NFYEPQIITTDNTFVS	1.039	1.007

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

B13	Spike	1135	1151	NTVYDPLQPELDSFKEE	6.183	1.011
B14	Spike	1153	1172	DKYFKNHTSPDVDLGDISGI	1.399	1.007
B15	Spike	1190	1193	AKNL	1.087	1.005
B16	Spike	1203	1209	LGKYEQY	2.512	1.035
B17	Spike	1255	1265	KFDEDDSEPVL	2.654	1.003
B18	Spike	63	70	KNLNSSRV	3.471	1.002
B19	Spike	173	176	SRTL	1.504	1.011

Note: Epitopes B4 and B5 are located within the RBD region.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 2. Distribution of T-cell epitopes among three structural proteins

Protein	Count of T-cell Epitope	No. of epitope per residue	Epitope overage	HLA Types Count
Spike	378	0.297	93.01%	33
Membrane	90	0.405	96.00%	31
Envelop	31	0.413	94.14%	32

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 3. Candidate vaccine peptides

Peptide	Protein	Start	End	Vaccine peptide	Count of T Epitopes	Count of B Epitopes	HLA Score
P1	Spike	19	46	TTRTQLPPAYTNSFTRGVYYPDKVFRSS	10	1	1.086
P2	Spike	75	99	GTKRFDNPVLPFNDGVYFASTEKSNK	6	1	1.143
P3	Spike	118	143	LIVNNATNVVIKVCEFQFCNDPFLGVKK	7	0	1.179
P4	Spike	142	170	GVYYHKNNKSWMESEFRVYSSANNCTFEY	10	0	1.664
P5	Spike	186	209	FKNLREFVFKNIDGYFKIYSKHTP	8	1	1.264
P6	Spike	258	279	WTAGAAAYYVGYLQPRTFLLKYKKKKK	10	0	1.115
P7	Spike	310	337	KGIYQTSNFRVQPTESIVRFPNITNLCP	10	0	1.012
P8 *	Spike	357	386	RISNCVADYSVLYNSASFSTFKCYGVSPTK	8	0	1.318
P9 *	Spike	405	433	DEVRQIAPGQTGKIADYNYKLPDDFTGKKK	7	1	0.928
P10 *	Spike	448	472	NYNYLYRLFRKSNLKPFERDISTEI	7	0	1.625
P11 *	Spike	478	505	TPCNGVEGFNCYFPLQSYGFQPTNGVGYKK	7	0	1.413
P12 *	Spike	494	523	SYGFQPTNGVGYQPYRVVVLSFELLHAPAT	10	1	1.581

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

P13	Spike	625	652	HADQLTPTWRVYSTGSNVFQTRAGCLIG	8	0	1.214
P14	Spike	671	699	CASYQTQTNSPRRARSVASQSIIAYTMSL	8	1	1.234
P15	Spike	771	799	AVEQDKNTQEVFAQVKQIYKTPPIKDFGGK	8	2	0.952
P16	Spike	805	833	ILPDPSKPSKRSFIEDLLFNKVTLADAGFK	7	1	1.068
P17	Spike	896	923	IPFAMQMAYRFNGIGVTQNVLYENQKLI	7	0	1.625
P18	Spike	965	991	QLSSNFGAISSVLNDILSRLDKVEAEVKKK	9	0	1.012
P19	Spike	1052	1073	FPQSAPHGVVFLHVTYVPAQEK	8	1	1.532
P20	Spike	1068	1096	VPAQEKNFTTAPAICHDGKAHFPREGVFV	4	1	0.402
P21	Spike	1095	1123	FVSNGTHWFVTQRNFYEPQIITTDNTFVSK	8	1	1.236
P22	Spike	1135	1155	NTVYDPLQPELDSFKEELDKYKKKKK	2	1	0.254
P23	Spike	1153	1181	DKYFKNHTSPDVDLGDISGINASVVNIQKK	5	1	0.322
P24	Spike	1190	1217	AKNLNESLIDLQELGKYEQYIKWPWYIWKK	6	2	0.659
P25	Spike	1216	1245	IWLGFIAGLIAIVMVTIMLCKKKKKKKKK	5	0	1.394
P26	Spike	1236	1265	KKKKCCSCLKGCCSCGSCCKFDEDDSEPVL	4	1	0.520

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

P27	Envelop	4	33	FVSEETGTLIVNSVLLFLAFVVFLKKKKKK	11	0	1.133
P28	Envelop	45	70	NIVNVSLVKPSFYVYSRVKNLNSSRV	9	1	1.455
P29	Membrane	122	150	VPLHGTILTRPLLESELVIGAVILRGHLRK	9	0	1.508
P30	Membrane	173	201	SRTLSYYKLGASQRVAGDSGFAAYSRYRI	6	1	0.902

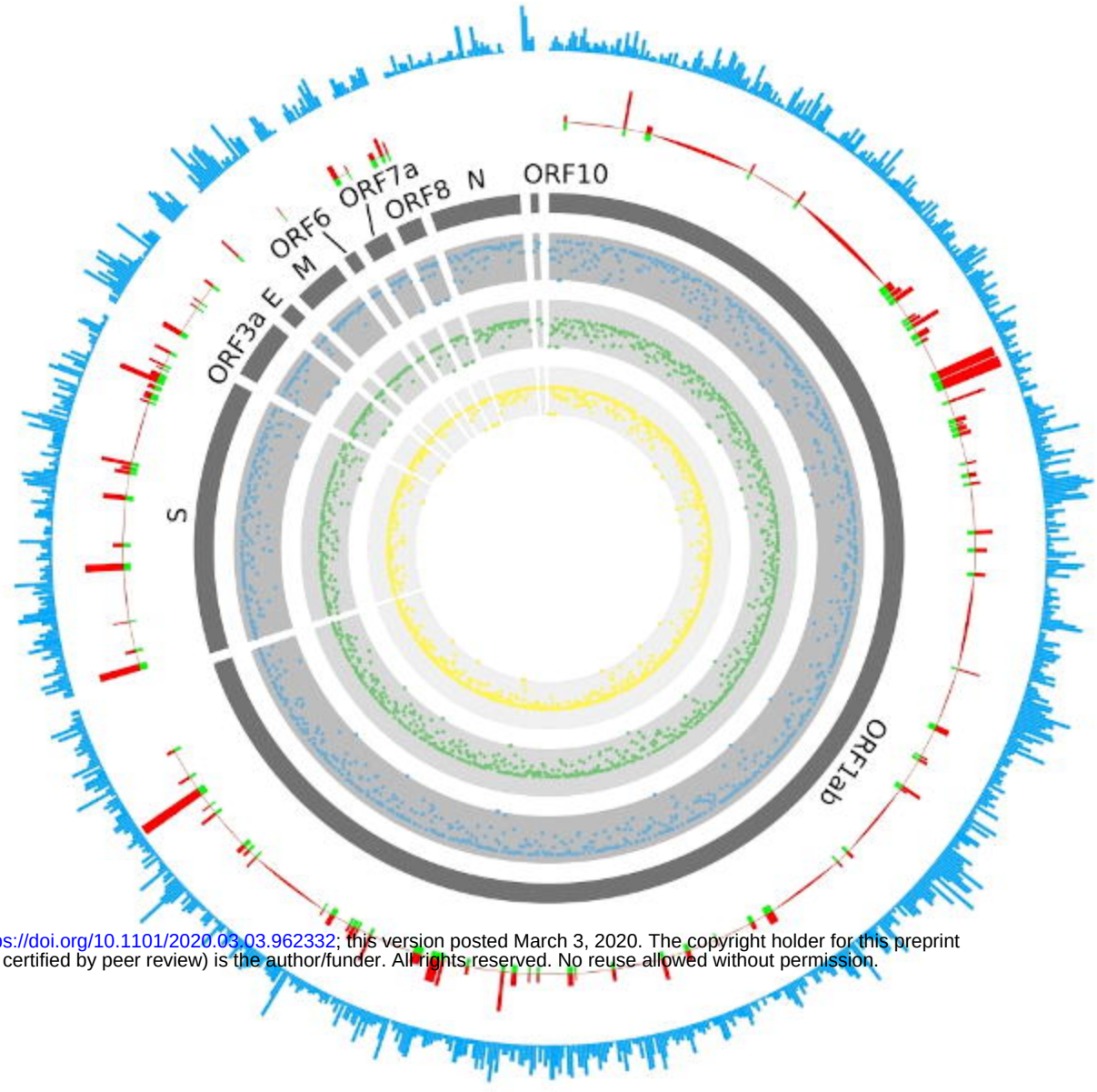
Note: Peptide labeled by asterisks (*) are located within the RBD region.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 4. Docking results for T-cell epitope P25 and P27 against three HLA types

Panel in Fig. 3	Protein	Start	Epitope	HLA type	HLA Score	ITScorePeP	Contact residues
a	Spike	1220	FIAGLIAIV	HLA-A*02:01	0.123	-144.2	PHE-1,GLY-4,LEU-5,ILE-6,ALA-7
b	Spike	1220	FIAGLIAIV	HLA-B*46:01	0.102	-138.2	ILE-6,VAL-9
c	Spike	1220	FIAGLIAIV	HLA-C*03:04	0.100	-146.6	PHE-1,ALA-3,ILE-8,VAL-9
d	Envelop	4	FVSEETGTL	HLA-A*02:06	0.052	-147.7	PHE-1,VAL-2,SER-3,GLU-4,THR-6
e	Envelop	4	FVSEETGTL	HLA-B*46:01	0.102	-140.2	PHE-1,SER-3,GLU-4,THR-6,THR-8
f	Envelop	4	FVSEETGTL	HLA-C*07:02	0.152	-136.7	PHE-1,GLU-4,THR-8,LEU-9

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



bioRxiv preprint doi: https://doi.org/10.1101/2020.03.03.962332; this version posted March 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1

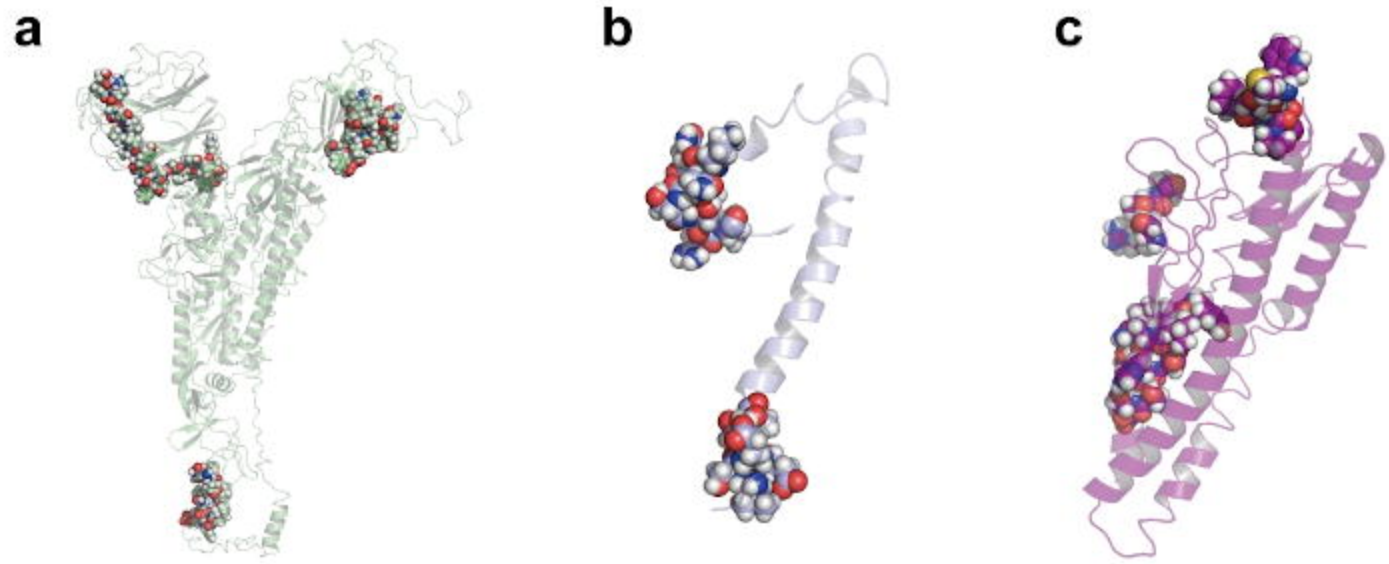


Figure 2

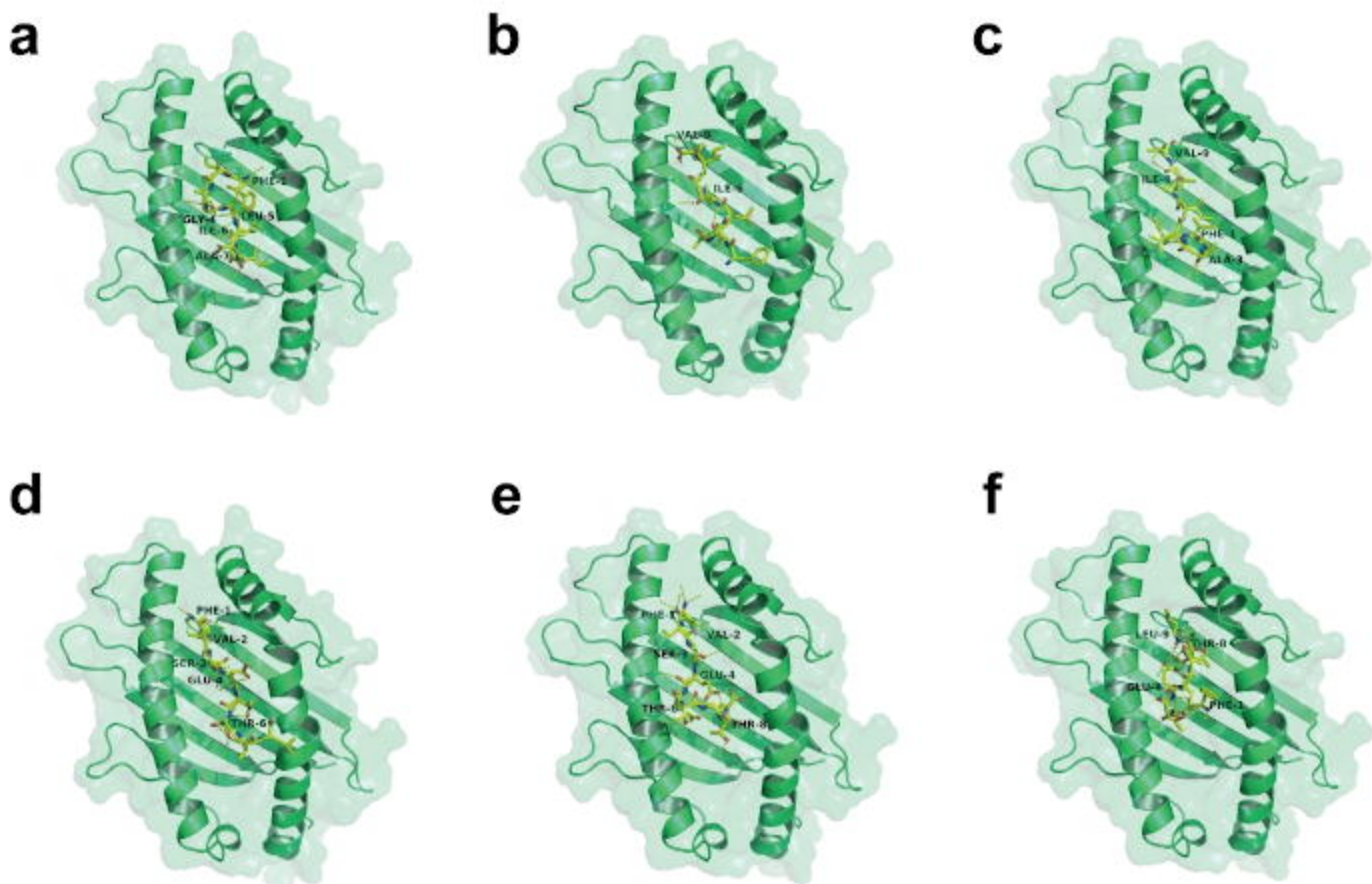


Figure 3