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

Ye Feng[1,2†*], Min Qiu[3†], Liang Liu[3†], Shengmei Zou[1,2†], Yun Li[4], Kai Luo[3], Qianpeng Guo[3†], Ning Han[3], Yingqiang Sun[3], Kui Wang[3], Xinlei Zhuang[5], Shanshan Zhang[3,6], Shuqing Chen[3,5,6*], Fan Mo[3,5,7,8*]

[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; [8] Hangzhou AI-Force Therapeutics Co., Ltd., Hangzhou, China.

[†] These authors contributed equally to this work.

***Corresponding authors:** Fan Mo (fmo@prostatecentre), Shuqing Chen (chenshuqing@zju.edu.cn) or Ye Feng (pandafengye@zju.edu.cn)

Abstract

A new coronavirus SARS-CoV-2 has caused over 9.2 million infection cases and 475758 deaths worldwide. Due to the rapid dissemination and the unavailability of specific therapy, there is a desperate need for vaccines to combat the epidemic of SARS-CoV-2. An in silico approach based on the available virus genome was applied to identify 19 high immunogenic B-cell epitopes and 499 human-leukocyte-antigen (HLA) restricted T-cell epitopes. Thirty multi-epitope peptide vaccines were designed by iNeo Suite, and manufactured by solid-phase synthesis. Docking analysis showed stable hydrogen bonds of epitopes with their corresponding HLA alleles. When four vaccine peptide candidates from the spike protein of SARS-CoV-2 were selected to immunize mice, a significantly larger amount of IgG in serum as well as an increase of $CD19^+$ cells in ILNs was observed in peptide-immunized mice compared to the control mice. The ratio of IFN-γ-secreting lymphocytes in $CD4^+$ or $CD8^+$ cells in the peptides-immunized mice were higher than that in the control mice. There were also a larger number of IFN-γ-secreting T cells in spleen in the peptides-immunized mice. This study screened antigenic B-cell and T-cell epitopes in all encoded proteins of SARS-CoV-2, and further designed multi-epitope based peptide vaccine against viral structural proteins. The obtained vaccine peptides successfully elicited specific humoral and cellular immune responses in mice. Primate experiments and clinical trial are urgently required to validate the efficacy and safety of these vaccine peptides.

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

Importance

So far, a new coronavirus SARS-CoV-2 has caused over 9.2 million infection cases and

475758 deaths worldwide. Due to the rapid dissemination and the unavailability of specific therapy, there is a desperate need for vaccines to combat the epidemic of SARS-CoV-2. Different from the development approaches for traditional vaccines, the development of our peptide vaccine is faster and simpler. In this study, we performed an in silico approach to identify the antigenic B-cell epitopes and human-leukocyte-antigen (HLA) restricted T-cell epitopes, and designed a panel of multi-epitope peptide vaccines. The resulting SARS-CoV-2 multi-epitope peptide vaccine could elicit specific humoral and cellular immune responses in mice efficiently, displaying its great potential in our fight of COVID-19.

Introduction

Since December 2019, a new type of coronavirus, SARS-CoV-2 (previously named 2019-nCoV by the World Health Organization), has caused an outbreak of viral lung infections in Wuhan City, Hubei Province, China, and later thrived over 30 countries worldwide (1-4). As of 24 June 2020, there have been over 9.2 million total confirmed cases and 475758 deaths of SARS-CoV-2 infection in the ongoing pandemic (5). 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 (6), which suggested its probable origin in bats (7). The main clinical manifestations of SARS-CoV-2 patients are fever (≥38℃), 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) (8).

Currently, there is no approved therapeutics or vaccines available for the treatment of

COVID-19 (9). Due to lack of anti-viral drugs or vaccines, control measures have been relying on the rapid detection and isolation of symptomatic cases (9). 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 enables 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 (9). The genome of SARS-CoV-2 encodes the spike protein, the membrane protein, the envelope protein, the nucleocapsid protein and a few replication and 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 (10). 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) (6, 9, 11, 12). The RBD of SARS-CoV-2 binds angiotensin-converting enzyme 2 (ACE2) with 10-20-fold higher affinity than does SARS-CoV (13). 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 (11, 14-17).

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 (18). Since the development of a peptide vaccine

against the virus causing foot-and-mouth disease (19), the establishment of peptide synthesis method by Lerner et al. (20), 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. Then we selected a few candidate vaccine peptides to immunize mice. As a result, these peptides successfully elicited specific humoral and cellular immune responses, showing their potentials in the real combat against SARS-CoV-2.

Materials and Methods

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) (21).

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 (22). Prediction of linear B-cell epitopes was performed through Bepipred software (23). The antigenic sites were determined with Kolaskar method (24). The surface accessible epitopes were predicted by Emini tool (25).

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 was predicted using

netMHCpan (26) 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 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.

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.

Structural analysis

The online server swiss-model was used to predict the 3D protein structures of viral proteins and HLA molecules (27). The online server PEP_FOLD was used to predict T-cell epitopes' structures (28). To display the interaction between T-cell epitopes and

HLA molecules, T-cell epitope models were docked to HLA molecules using MDockPep (29). All predicted structures or models were decorated and displayed by the open source version of pymol program (https://github.com/schrodinger/pymol-open-source).

Immuno-stimulation of B lymphocytes

The selected 4 peptides were synthesized with the solid phase synthesis method by GenScript Biotech Company (Nanjing, China), and were mixed at an equal concentration of 1mg/ml. The immunization experiment constituted three groups, each consisting of 12 6–8 week-old female C57mice. Mice were immunized subcutaneously with 100 μL of the following compounds: Group 1, 100 μg peptide mixture & 50 μl QuickAntibody-Mouse (Biodragen, Beijing, China); Group 2, 50 μl QuickAntibody-Mouse; and Group 3, 100 μl PBS as negative control. The immunization was performed once a week and repeated four times in total.

On the 14th, 21st, 28th day after the 1st immunization, retro-orbital blood was collected from 5 randomly selected mice in each group. The sera were tested for the presence of IgG by enzyme linked immunosorbent assay (ELISA). Briefly, ELISA plates were coated with 100 μL of S protein (2 μg/mL recombinant 2019-nCoV s-trimer protein; Novoprotein, Shanghai, China). The coated plates were incubated overnight at room temperature and washed with PBS containing 0.05% Tween 20 (PBS-T). Blocking buffer (1% nonfat milk in PBS) was added and incubated for 2 h at 37 °C. After washing with PBS-T, the diluted sera (1:1000 in PBS buffer containing 1% nonfat milk and 0.05% Tween-20) were added and incubated at 37 °C for 2 h. Next the plates were washed with PBS and again incubated at 37 °C for 2 h with horseradish peroxidase conjugated Goat Anti-mouse IgG antibody (1:5000; Genscript, Nanjing, China). Detection was carried out

with O- Phenylene Diamine (OPD, 0.01%) substrate (Thermo Scientific, Waltham, US) for 30 min at 37 °C. Finally, the reaction was stopped using stop buffer (Solarbio, Beijing, China), and the absorbance was measured at 492 nm.

On the 28th day after the 1st immunization, 6 randomly selected mice were euthanized. The inguinal lymph nodes (ILNs) were harvested and processed into single cell suspensions. The cells were stained with Zombie Aqua (Biolegend, San Diego, US), APC-conjugated anti-mouse CD19 antibody (Biolegend, San Diego, US), PerCP/Cyanine5.5- conjugated anti-mouse CD95 (Fas) antibody (Biolegend, San Diego, US) and FITC-conjugated anti-mouse GL7 antibody (Biolegend, San Diego, US). The stained cells were resuspended in 500 μl PBS and subsequently processed by the Aria II flow cytometry instrument (BD, Franklin Lakes, US).

Immuno-stimulation of T cells

The design of immunization experiment was similar to that for the B cells, but the injecting compounds were different: Group 1, 100 μg peptide mixture & 10 μg granulocyte-macrophage colony stimulating factor (GM-CSF; Novoprotein, Shanghai, China); Group2, 10 μg GM-CSF; Group 3, 100 μl PBS as negative control.

On the 14th and 28th day after the 1st immunization, 3 randomly selected mice were euthanized, respectively. The relative proportions of T cells in the splenocytes and ILN lymphocytes were analyzed by the Aria II flow cytometry instrument (BD, Franklin Lakes, US). Briefly, spleen and ILN were harvested and processed into single cell suspensions. Splenocytes (1×10^6/well) and ILN lymphocyte (1×10^6/well) were cultured overnight with the peptide mixture (5μg/ml) or in RPMI-1640 alone (negative control). Cells were stained with Zombie Aqua (Biolegend, San Diego, US),

PerCP/Cyanine5.5-conjugated anti-mouse CD8a antibody (Biolegend, San Diego, US), PE/Cy7- conjugated anti-mouse CD4 antibody (Biolegend, San Diego, US) and APC-conjugated anti-mouse IFN-γ antibody (Biolegend, San Diego, US). The stained cells were resuspended in 500μl PBS for flow cytometry analysis.

The IFN-γ-secreting T lymphocytes were also quantified on 6 randomly selected mice using an ELISPOT kit (Dakewe, Shenzhen, China). Briefly, 100 μL of PBS were added to 96-well plates pre-coated with an anti-IFN-γ mAb. 2×10^5/well cells were incubated in duplicate with 5μg/mL of peptide or medium alone (negative control) for 16 h in a 37 °C humidified incubator with 5% CO_2. Splenocytes stimulated with phorbol myristate acetate (PMA) served as positive control. After removing cells and washing with buffer (PBS with 0.1% Tween 20), 1:100 diluted biotinylated anti-IFN-γ were added and incubated for 90 min at 37°C. After each incubation step, the plates were washed three times with buffer. Next, after 1 h of incubation with streptavidin-alkaline phosphatase conjugate (1/5000 in PBS-0.1% Tween 20), the plates were developed with a solution of 5-bromo-4-chloro-3-indolylphosphate (BCIP)-nitroblue tetrazolium until red spots appeared. Tap water was used to stop the reaction, and the plates were dried in air overnight. Individual spots were counted under a CTL-ImmunoSpot®S6 FluoroSpot (Cellular Technology, Kennesaw, USA).

Statistical analysis

Comparisons were analyzed by one-way analysis of variance (ANOVA). P value less than 0.05 was considered significant.

Results

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 (30). 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.

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 (31). 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 (32). 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 was 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.

Multi-epitope vaccine design

Based on the 19 B-cell epitopes and their 121 adjacent T-cell epitopes, 17 candidate vaccine 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 analyzed 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 envelope 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 (33).

Interaction of predicted peptides with HLA alleles

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

Humoral immune responses to SARS-CoV-2 S protein

Based on the above immunoinformatics analysis, 4 designed vaccine peptides, namely P9,

P12, P14 and P15, were chosen as the candidates for the downstream validation experiments because of their relatively higher counts of B-cell and T-cell epitopes and the higher frequencies of their epitopes' corresponding HLA alleles (Table 3). We immunized mice by subcutaneous injection of the mixture of these synthesized peptides plus QuickAntibody (an adjuvant for stimulating B cells). Mice injected with QuickAntibody only or PBS were studied as controls. The immunization was performed once a week and repeated four times in total.

To evaluate whether these peptides induce B cells to produce specific antibody against the S protein, an ELISA assay was conducted to detect IgG in the sera of mice. Fourteen days after the 1st immunization, the amount of IgG showed little difference between the peptide-treated mice and the controls (Figure 4a), suggesting that two weeks were not long enough to elicit humoral immune response. Since the 21st day after the 1st immunization, however, the expression of IgG had risen to the plateau in the peptide-treated mice and was remarkably higher than that in the control mice ($p<0.05$; Figure 4a). Germinal centers (GCs) are the main sites for the production of high-affinity, long-lived plasma cells and memory B cells. On the 28th day after the 1st immunization, we collected ILN and stained cells with antibody of GL7 and FAS. Flow cytometry showed that there were much more B cells activated in the peptides-treated mice than in the mice injected with adjuvant only (Figure 4b); and the numbers of rapidly proliferated B cells ($CD19^{+}/FAS^{+}/GL7^{+}$) from GC in ILNs of the peptides-treated mice were significantly higher than that in the control groups, demonstrating that there were increased GCs induced by peptide vaccines (Figure 4c-d). In future, a viral neutralization study is further required for demonstrating that the designed peptide vaccines can

efficiently activate specific humoral immune responses to the S protein of SARS-CoV-2.

Cellular immune responses to SARS-CoV-2 S protein

In parallel, we also immunized mice with peptides plus GM-CSF, an adjuvant that induced the development of monocytes, neutrophils and dendritic cells. Mice injected with GM-CSF only or PBS were used as controls. On both the 14th and the 28th day after the 1st immunization, the ILNs were collected. We found that the ratios of IFN-γ-secreting cells to both $CD4^+$ and $CD8^+$ T cells in the peptides-treated mice were significantly higher than that in the control groups (Figure 5), suggesting the activation of T cells by the peptide vaccines. Notably, the ratio of IFN-γ-secreting cells seemed to reach its plateau on the 14th day, since no further significant increase of this ratio was observed on the 28th day. It was likely that, in the absence of the virus, the repeated T cell stimulation led to depletion or transfer of T cells in ILNs; accordingly, the ratio of IFN-γ-secreting cells in the ILNs kept relatively stable afterwards.

We also quantified the peptide specific lymphocytes in mice spleen on the 28th day using the enzyme-linked immunospot (ELISPOT) assay of IFN-γ secreting. In terms of either the ratio of IFN-γ-secreting lymphocytes in splenocytes or the total number of IFN-γ-secreting lymphocytes in spleen, mice immunized with peptide vaccines had significantly higher ratios than that in the control groups. This finding was overall consistent with the flow cytometry results of ILN cells, suggesting that the lymphocytes were activated and might recirculate to gather in the spleen after 4-week vaccination.

Discussion

Since the outbreak of COVID-19, numerous researchers have contributed in the

development of safe and effective vaccines for COVID-19 to increase the chances of success. Meanwhile, World Health Organization (WHO) has encouraged scientists to test all candidate vaccines until they fail. Currently, the developing candidate vaccines include inactivated virus vaccine, recombinant protein vaccine, viral vector vaccine, peptide vaccine, DNA and RNA vaccines.

In comparison, the involvement of the optimization of cell and virus culture process increase the complexity and thus preparation time of inactivated virus vaccine and viral vector vaccine. Also, for the development of inactivated virus vaccine, besides Good Manufacturing Practices (GMP) system, extremely high manufacture standard is required to avoid medical accidents due to the failure of complete inactivation of virus toxicity. Differently, the development of our peptide vaccine is much simpler, including only two major steps: sequence design through reverse vaccinology approach, and peptide synthesis.

Until now, most recombinant protein vaccine candidates have focused on spike protein, which although can provide good safety, is probably unable to stimulate strong T cell immune response (34). Therefore, effective adjuvants are usually added when use this kind of vaccines. Similar to recombinant protein vaccines, most viral vector vaccine candidates are based on the expression of spike protein (35). Although the application of viral vector could enhance the vaccine's delivery efficiency, the body's immune response to vector itself might interfere with the immune response to target epitopes, and thus compromise the efficacy of this kind of vaccines. Unlike the vaccine candidates discussed above, our peptide vaccine candidate is composed of 30 peptides from not only spike protein, but also membrane protein and envelope protein of SARS-CoV-2, containing

both B-cell epitopes and T-cell epitopes to induce specific humoral and cellular immune response against SARS-CoV-2 more efficiently.

On the other hand, although promising in preclinical or Phase I and II clinical trials, no mRNA vaccine has been formally approved by FDA (36). In other words, more clinical trials should be conducted to prove the efficacy of mRNA vaccines. Moreover, the batch production, manufacturing process, and industrial expansion of mRNA vaccines are still under exploration. For DNA vaccines, besides the same concerns for mRNA vaccines, the possibility of the integration of foreign DNA sequence to human genome has brought uncertainty in their development. Different from these nucleic acid vaccines, peptide vaccine has relatively more mature manufacturing process. And the successful launch of previous peptide vaccines has already demonstrated the safety and efficacy of peptide-based vaccines.

In this study, our SARS-CoV-2 multi-epitope peptide vaccine could elicit specific humoral and cellular immune responses in mice, displaying its great potential in the fight of COVID-19. In future, more experiments will be conducted to validate the efficacy of our SARS-CoV-2 peptide vaccine.

Authors' contributions

SC and FM conceived and designed the project. YF, FM, MQ, SZ, KL, YS, KW, XZ and SZ analyzed the data. LL and QG conducted in-vivo validation. YF, YL and NH wrote the initial draft. All authors revised and approved the final manuscript.

Acknowledgement

The authors would like to thank Dr. Jing Guo for the suggestion on the in-vivo experiments.

Ethics approval and consent to participate

Ethical approval for the animal experiments was obtained from Zhejiang Chinese Medical University Laboratory Animal Research Center.

Competing interests

The authors declare that they have no competing interests.

References

1. Organization WH. 2020. Surveillance case definitions for human infection with novel coronavirus (nCoV): interim guidance v1, January 2020. . Organization WH, World Health Organization, Geneva. https://apps.who.int/iris/handle/10665/330376.
2. Perlman S. 2020. Another Decade, Another Coronavirus. New England Journal of Medicine doi:10.1056/NEJMe2001126.
3. Joseph T Wu* KL, 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 doi:10.1016/S0140-6736(20)30260-9.
4. Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, Haagmans BL, Lauber C, Leontovich AM, Neuman BW, Penzar D, Perlman S, Poon LLM, Samborskiy D, Sidorov IA, Sola I, Ziebuhr J. 2020. Severe acute respiratory syndrome-related coronavirus: The species and its viruses – a statement of the Coronavirus Study Group. bioRxiv doi:10.1101/2020.02.07.937862:2020.02.07.937862.
5. Anonymous. Coronavirus COVID-19 Global Cases by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6. Accessed Retrieved 18 April.
6. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, Si H-R, Zhu Y, Li B, Huang C-L. 2020. Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. doi:10.1101/2020.01.22.914952.

396 7. Benvenuto D, Giovannetti M, Ciccozzi A, Spoto S, Angeletti S, Ciccozzi M. 2020. The
397 2019-new coronavirus epidemic: evidence for virus evolution. Journal of Medical Virology.
398 8. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X. 2020. Clinical
399 features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet.
400 9. Chen Y, Liu Q, Guo D. Emerging coronaviruses: genome structure, replication, and
401 pathogenesis. Journal of Medical Virology.
402 10. Jiayuan Chen JS, Tungon Yau ,Chang Liu , Xin L, Qiang Zhao, Jishou Ruan , Gao Shan.
403 2020. Bioinformatics analysis of the Wuhan 2019 human coronavirus genome. doi:18. 5.
404 10.12113/202001007.
405 11. Tian X, Li C, Huang A, Xia S, Lu S, Shi Z, Lu L, Jiang S, Yang Z, Wu Y. 2020. Potent
406 binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human
407 monoclonal antibody. Emerging microbes & infections 9:382-385.
408 12. Dong N, Yang X, Ye L, Chen K, Chan EW-C, Yang M, Chen S. 2020. Genomic and
409 protein structure modelling analysis depicts the origin and infectivity of 2019-nCoV, a new
410 coronavirus which caused a pneumonia outbreak in Wuhan, China. bioRxiv.
411 13. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, Graham BS, McLellan
412 JS. 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation.
413 Science 367:1260-1263.
414 14. He YX, Lu H, Siddiqui P, Zhou YS, Jiang SB. 2005. Receptor-binding domain of severe
415 acute respiratory syndrome coronavirus spike protein contains multiple
416 conformation-dependent epitopes that induce highly potent neutralizing antibodies.
417 Journal of Immunology 174:4908-4915.
418 15. Babcock GJ, Esshaki DJ, Thomas WD, Ambrosino DM. 2004. Amino acids 270 to 510 of
419 the severe acute respiratory syndrome coronavirus spike protein are required for
420 interaction with receptor. Journal of Virology 78:4552-4560.
421 16. Saif LJ. 1993. Coronavirus Immunogens. Veterinary Microbiology 37:285-297.
422 17. Buchholz UJ, Bukreyev A, Yang LJ, Lamirande EW, Murphy BR, Subbarao K, Collins PL.
423 2004. Contributions of the structural proteins of severe acute respiratory syndrome
424 coronavirus to protective immunity. Proceedings of the National Academy of Sciences of
425 the United States of America 101:9804-9809.
426 18. Purcell AW, McCluskey J, Rossjohn J. 2007. More than one reason to rethink the use of
427 peptides in vaccine design. Nature Reviews Drug Discovery 6:404-414.
428 19. Adam K-H, Kaaden O, Strohmaier K. 1978. Isolation of immunizing cyanogen
429 bromide-peptides of foot-and-mouth disease virus. Biochemical and biophysical research
430 communications 84:677-683.
431 20. Lerner RA, Green N, Alexander H, Liu F-T, Sutcliffe JG, Shinnick TM. 1981. Chemically
432 synthesized peptides predicted from the nucleotide sequence of the hepatitis B virus

433 genome elicit antibodies reactive with the native envelope protein of Dane particles.
434 Proceedings of the National Academy of Sciences 78:3403-3407.
435 21. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X,
436 Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y,
437 Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, Chen W, Shi W,
438 Tan W. 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus:
439 implications for virus origins and receptor binding. Lancet
440 doi:10.1016/S0140-6736(20)30251-8.
441 22. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, Wheeler DK,
442 Gabbard JL, Hix D, Sette A. 2015. The immune epitope database (IEDB) 3.0. Nucleic
443 acids research 43:D405-D412.
444 23. Jespersen MC, Peters B, Nielsen M, Marcatili P. 2017. BepiPred-2.0: improving
445 sequence-based B-cell epitope prediction using conformational epitopes. Nucleic acids
446 research 45:W24-W29.
447 24. Chen J, Liu H, Yang J, Chou K-C. 2007. Prediction of linear B-cell epitopes using amino
448 acid pair antigenicity scale. Amino acids 33:423-428.
449 25. Almofti YA, Abd-elrahman KA, Gassmallah SAE, Salih MA. 2018. Multi Epitopes Vaccine
450 Prediction against Severe Acute Respiratory Syndrome (SARS) Coronavirus Using
451 Immunoinformatics Approaches. American Journal of Microbiological Research 6:94-114.
452 26. Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, Buus S, Nielsen M. 2009.
453 NetMHCpan, a method for MHC class I binding prediction beyond humans.
454 Immunogenetics 61:1.
455 27. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de
456 Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. 2018. SWISS-MODEL: homology
457 modelling of protein structures and complexes. Nucleic Acids Res 46:W296-W303.
458 28. Lamiable A, Thevenet P, Rey J, Vavrusa M, Derreumaux P, Tuffery P. 2016. PEP-FOLD3:
459 faster de novo structure prediction for linear peptides in solution and in complex. Nucleic
460 Acids Res 44:W449-54.
461 29. Xu X, Yan C, Zou X. 2018. MDockPeP: An ab-initio protein-peptide docking server. J
462 Comput Chem 39:2409-2413.
463 30. Fieser TM, Tainer JA, Geysen HM, Houghten RA, Lerner RA. 1987. Influence of protein
464 flexibility and peptide conformation on reactivity of monoclonal anti-peptide antibodies with
465 a protein alpha-helix. Proceedings of the National Academy of Sciences 84:8568-8572.
466 31. Black M, Trent A, Tirrell M, Olive C. 2010. Advances in the design and delivery of peptide
467 subunit vaccines with a focus on toll-like receptor agonists. Expert review of vaccines
468 9:157-173.
469 32. Sesardic D. 1993. Synthetic peptide vaccines. Journal of medical microbiology

470 39:241-242.
471 33. Skwarczynski M, Toth I. 2016. Peptide-based synthetic vaccines. Chem Sci 7:842-854.
472 34. Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. 2020. A Sequence
473 Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune
474 Responses to SARS-CoV-2. Cell Host Microbe 27:671-680.e2.
475 35. Zhu F, Li Y, Guan X, Hou L, Wang W, Li J, Wu S, Wang B, Wang Z, Wang L, Jia S, Jiang
476 H, Wang L, Jiang T, Hu Y, Gou J, Xu SB, Xu JJ, Wang X, Wang W, Chen W. 2020. Safety,
477 tolerability, and immunogenicity of a recombinant adenovirus type-5 vectored COVID-19
478 vaccine: a dose-escalation, open-label, non-randomised, first-in-human trial. Lancet
479 395:10.
480 36. Jackson N, Kester K, Casimiro D, Gurunathan S, DeRosa F. 2020. The promise of mRNA
481 vaccines: a biotech and industrial perspective. NPJ Vaccines 5.
482
483

Figure legends

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

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

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

Fig. 4. Humoral immune responses to SARS-CoV-2 S protein. (a) Comparison of humoral response among groups of mice injected with PBS (marked in red), QuickAntibody (in green) and Peptide + QuickAntibody (in blue), respectively. The level of IgG was measured by ELISA. (c) The percentage of B cells ($CD19^+$ cells) in live cells. (c) The percentage of GC cells ($FAS^+/GL7^+$ cells) in $CD19^+$ cells. (d) Flow cytometry showing the larger number of $FAS^+/GL7^+$ cells in the peptides-treated mice. PBS, Q, P+Q represent mice injecting with PBS, mice with QuickAntibody, and mice with peptide vaccines plus QuickAntibody, respectively. *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$; ns, not significant.

Figure 5. Specific T cell activation by SARS-CoV-2 S protein. (a) and (b) The IFN-γ-secreting T lymphocytes in $CD4^+$ cells on the 14th and the 28th day after the 1st immunization, respectively. (c) and (d) The IFN-γ-secreting T lymphocytes in $CD8^+$ cells on the 14th and the 28th day after the 1st immunization, respectively. (e) and (f) The percentage of IFN-γ-secreting cells in $CD4^+$ cells and $CD8^+$ cells, respectively. PBS, G, P+G represent mice injected with PBS, GM-CSF, and peptide vaccines plus GM-CSF, respectively. *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$.

Figure 6. Quantification of the IFN-γ-secreting lymphocytes in mice spleen by ELISPOT. (a), Responses of splenocyte to DMSO (negative control), PMA (positive control) and the peptide mixture. (b) The number of IFN-γ-secreting cells per 100,000 splenocytes. (c) Total number of IFN-γ-secreting cells in spleen. PBS, G, P+G represent mice injected with PBS, GM-CSF, and peptide vaccines plus GM-CSF, respectively. ***, $p<0.001$.

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
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 B4 are located within the RBD region.

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
Envelope	31	0.413	94.14%	32

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
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
P27	Envelope	4	33	FVSEETGTLIVNSVLLFLAFVVFLKKKKKK	11	0	1.133
P28	Envelope	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.

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	Envelope	4	FVSEETGTL	HLA-A*02:06	0.052	-147.7	PHE-1,VAL-2,SER-3,GLU-4,THR-6
e	Envelope	4	FVSEETGTL	HLA-B*46:01	0.102	-140.2	PHE-1,SER-3,GLU-4,THR-6,THR-8
f	Envelope	4	FVSEETGTL	HLA-C*07:02	0.152	-136.7	PHE-1,GLU-4,THR-8,LEU-9

Figure 1

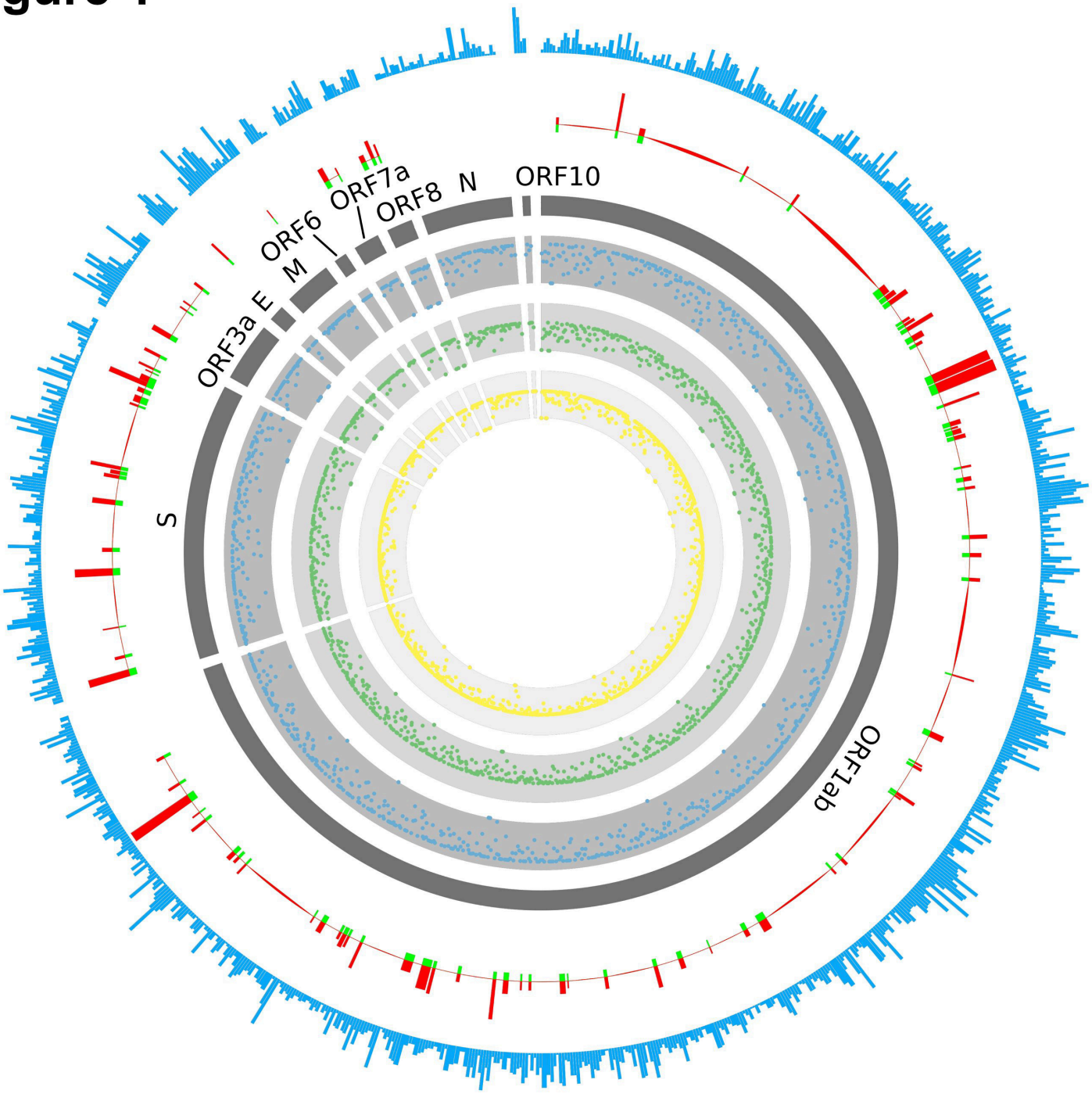


Figure 2

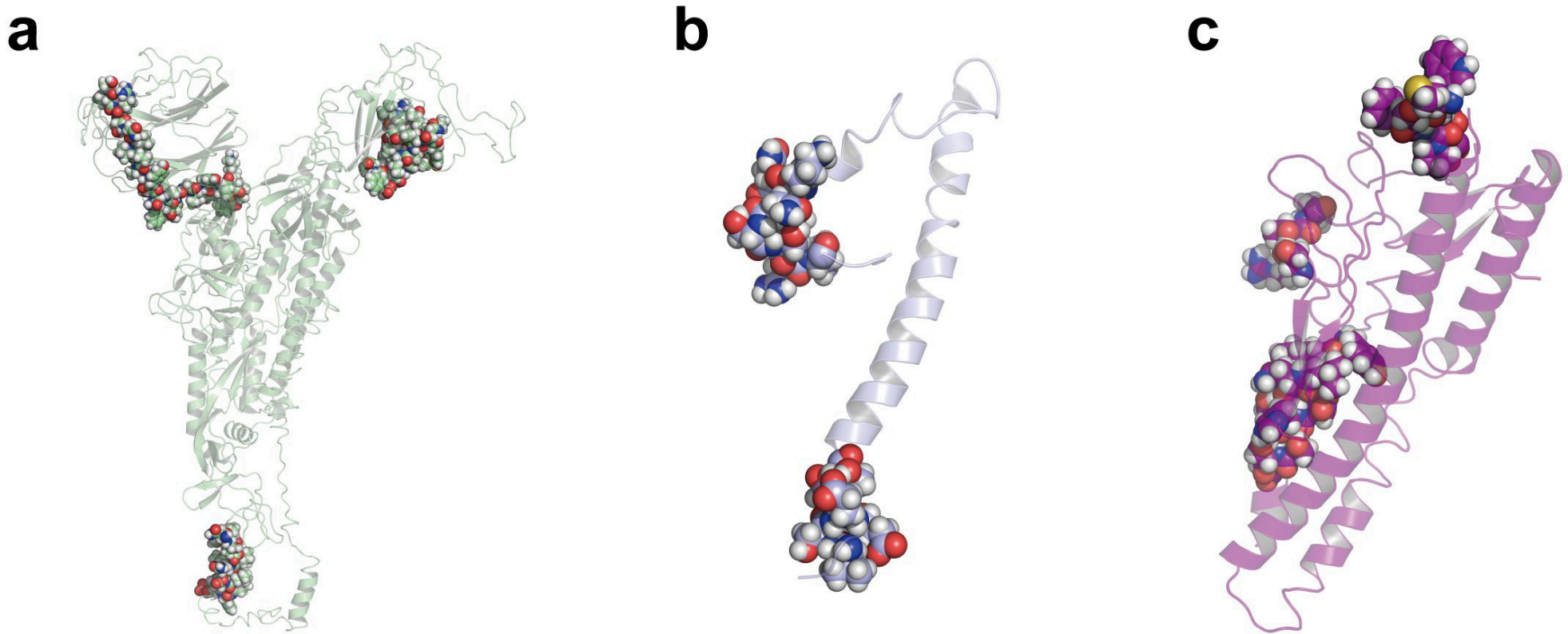


Figure 3

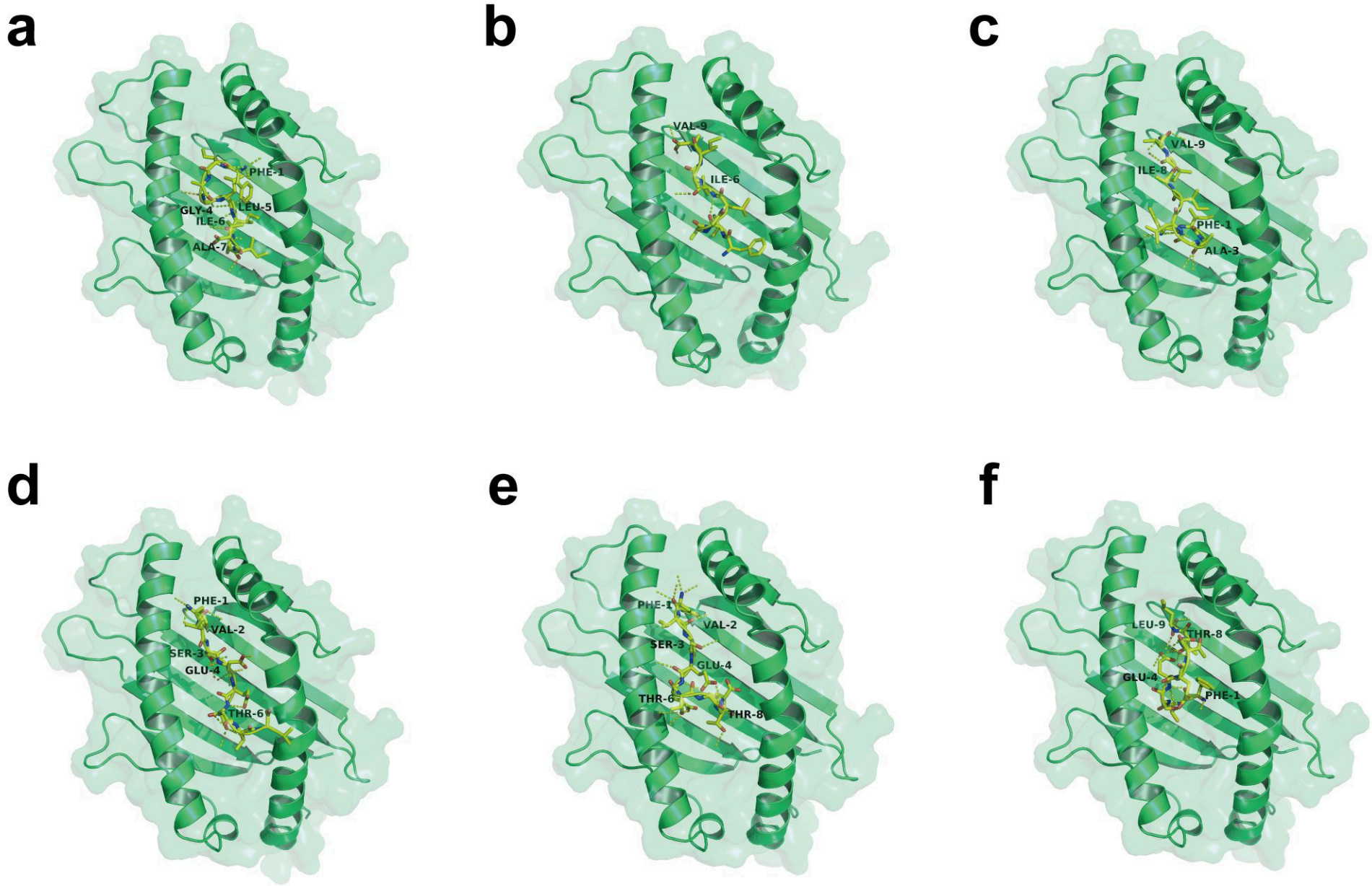

a
PHE-1
GLY-4
LEU-5
ILE-6
ALA-7
b
VAL-9
ILE-6
c
VAL-9
ILE-8
PHE-1
ALA-3
d
PHE-1
VAL-2
SER-3
GLU-4
THR-6
e
PHE-1
VAL-2
SER-3
GLU-4
THR-6
THR-8
f
LEU-9
THR-8
GLU-4
PHE-1

Figure 4

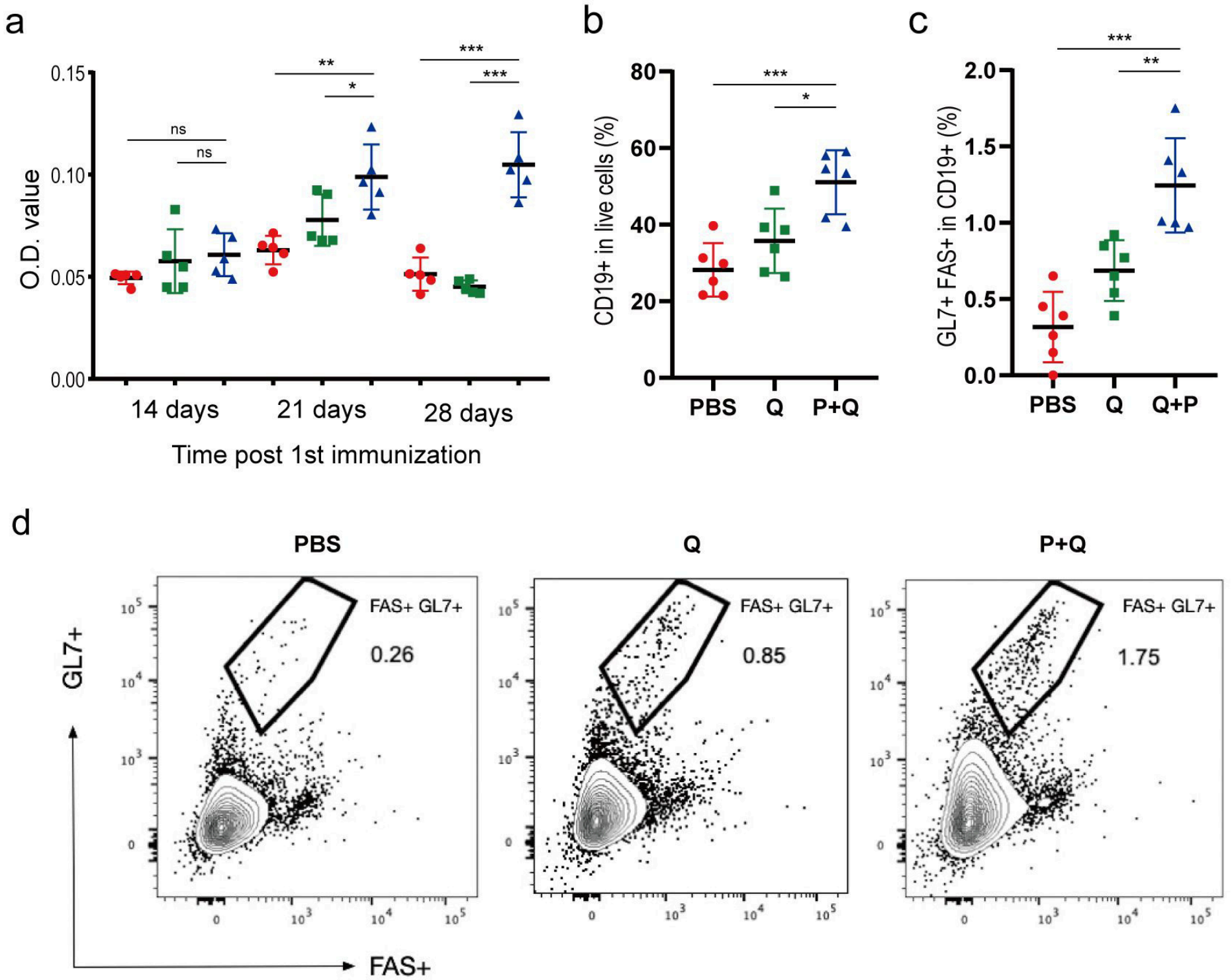
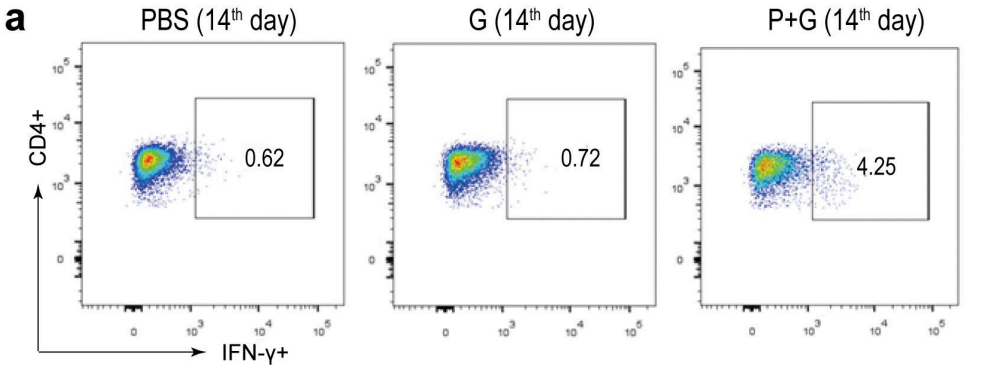
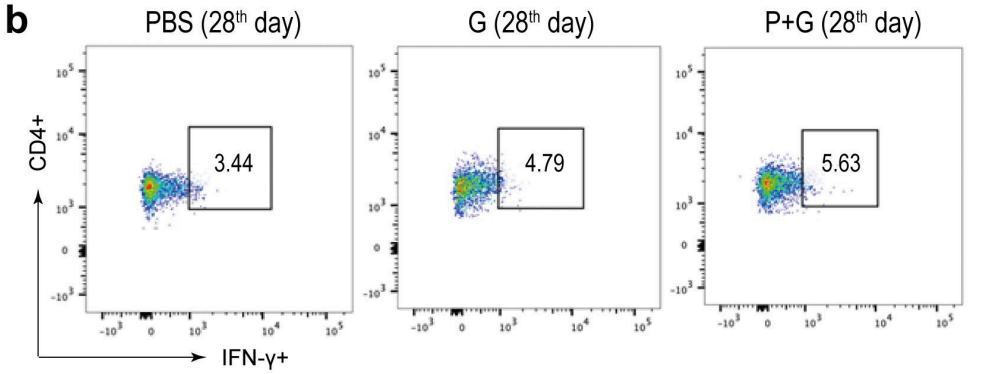
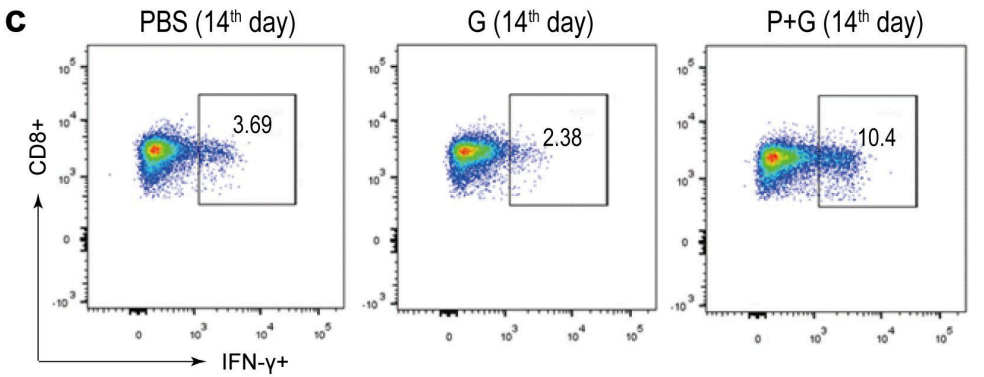
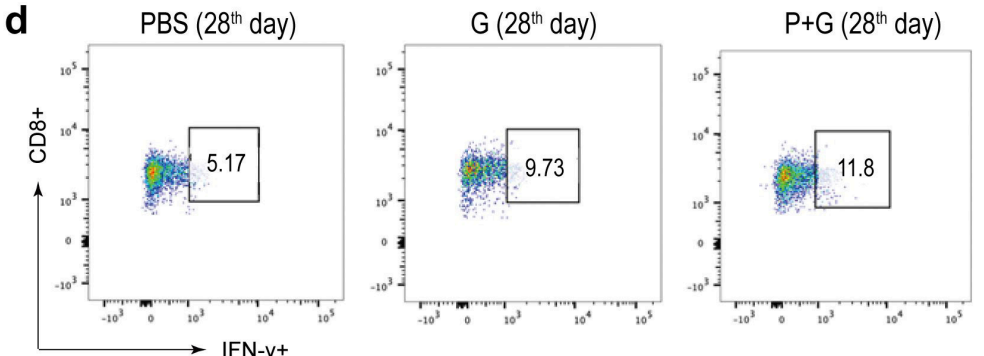

a
O.D. value
0.15
0.10
0.05
0.00
ns
ns
**
*
***
***
14 days
21 days
28 days
Time post 1st immunization
b
CD19+ in live cells (%)
80
60
40
20
0
***
*
PBS
Q
P+Q
c
GL7+ FAS+ in CD19+ (%)
2.0
1.5
1.0
0.5
0.0
***
**
PBS
Q
Q+P
d
PBS
Q
P+Q
FAS+ GL7+
0.26
0.85
1.75
GL7+
FAS+

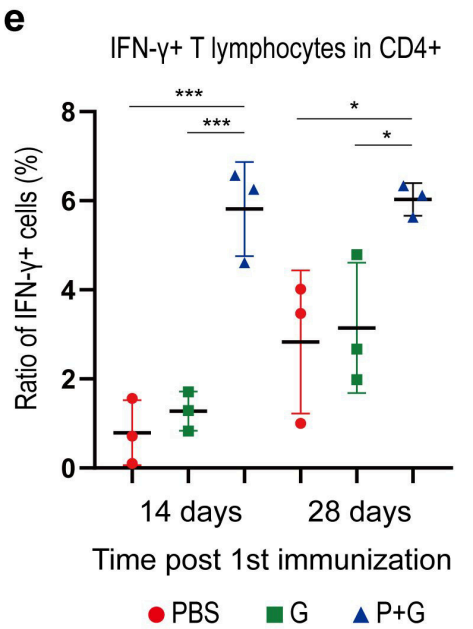
Figure 5


a
PBS (14th day)
G (14th day)
P+G (14th day)
CD4+
IFN-γ+
0.62
0.72
4.25


b
PBS (28th day)
G (28th day)
P+G (28th day)
CD4+
IFN-γ+
3.44
4.79
5.63


c
PBS (14th day)
G (14th day)
P+G (14th day)
CD8+
IFN-γ+
3.69
2.38
10.4


d
PBS (28th day)
G (28th day)
P+G (28th day)
CD8+
IFN-γ+
5.17
9.73
11.8


e
IFN-γ+ T lymphocytes in CD4+
Ratio of IFN-γ+ cells (%)
14 days
28 days
Time post 1st immunization
PBS
G
P+G

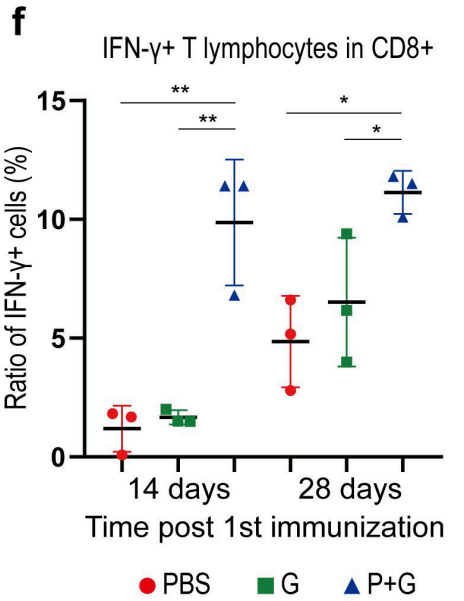
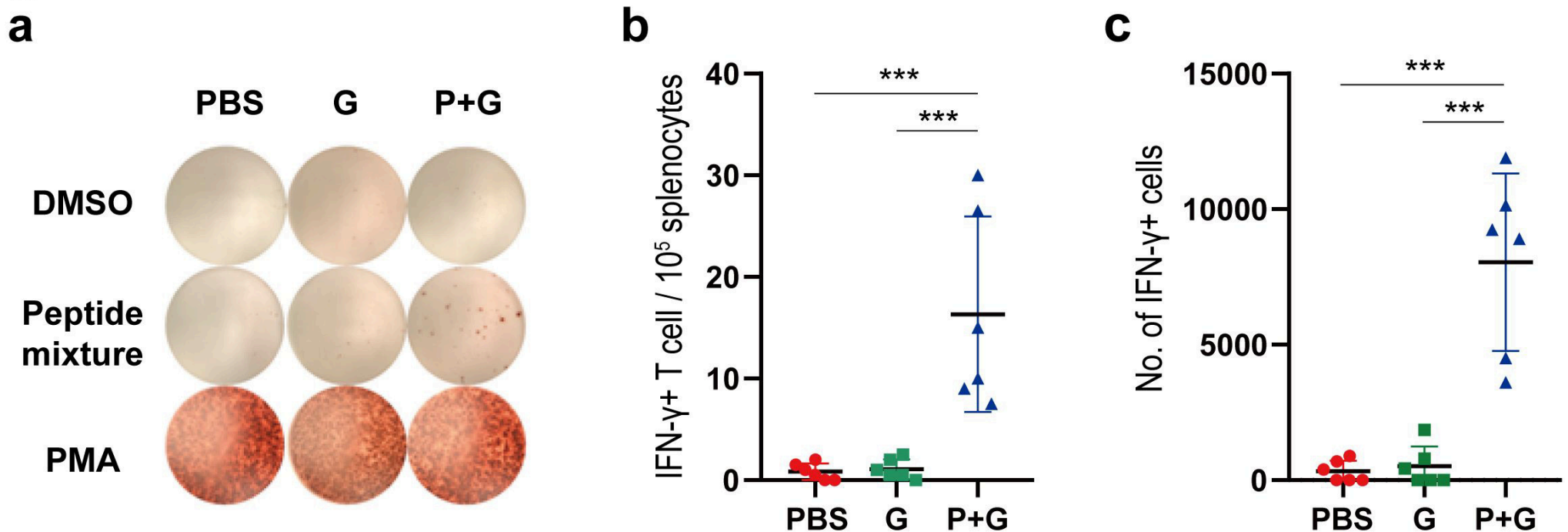

f
IFN-γ+ T lymphocytes in CD8+
Ratio of IFN-γ+ cells (%)
14 days
28 days
Time post 1st immunization
PBS
G
P+G

Figure 6


a
PBS
G
P+G
DMSO
Peptide mixture
PMA
b
IFN-γ+ T cell / 10^5 splenocytes
40
30
20
10
0
***
***
PBS
G
P+G
c
No. of IFN-γ+ cells
15000
10000
5000
0
***
***
PBS
G
P+G