Epitope-Based Potential Vaccine Candidate for Humoral and Cell-Mediated Immunity to Combat Severe Acute Respiratory Syndrome Coronavirus 2 Pandemic

Bratin Kumar Das and Debashree Chakraborty*

Cite This: J. Phys. Chem. Lett. 2020, 11, 9920−9930

ABSTRACT: The emergence of severe acute respiratory syndrome from novel Coronavirus (SARS-CoV-2) has put an immense pressure worldwide where vaccination is believed to be an efficient way for developing hard immunity. Herein, we employ immunoinformatic tools to identify B-cell, T-cell epitopes associated with the spike protein of SARS-CoV-2, which is important for genome release. The results showed that the highly immunogenic epitopes located at the stalk part are mostly conserved compared to the receptor binding domain (RBD). Further, two vaccine candidates were computationally modeled from the linear B-cell, T-cell epitopes. Molecular docking reveals the crucial interactions of the vaccines with immune-receptors, and their stability is assessed by MD simulation studies. The chimeric vaccines showed remarkable binding affinity toward the immune cell receptors computed by the MM/PBSA method. van der Waals and electrostatic interactions are found to be the dominant factors for the stability of the complexes. The molecular-level interaction obtained from this study may provide deeper insight into the process of vaccine development against the pandemic of COVID-19.

The sudden outbreak of febrile respiratory syndrome caused by a novel β-coronavirus (2019-nCoV) has created a global catastrophe where death rate is increasing every day.1−3 The rapid propagation of novel-coronavirus or SARS-CoV-2 has outcompeted the past epidemics caused by Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) and Middle East Respiratory Syndrome Coronavirus (MERS-CoV),4 and therefore, it has been declared as a global pandemic by the WHO. As of now, more than 29 million confirmed cases and 924 000 deaths are caused by SARS-CoV-2. The genome of SARS-CoV-2 is a single-stranded positive-sense RNA and reported to be largest viral genome until date.5,6 The genome entry in the host cell is guided by crown-shaped glycosylated spike protein (S-protein) located at the envelope surface. The S-protein is a trimeric class I fusion protein, which exists in a metastable prefusion conformation and undergoes a dramatic structural rearrangement to fuse into the host cell membrane.7−9 The ectodomain of S protein includes the receptor binding S1-sub unit and the trimeric membrane fusion S2 or stalk domain.10−12 The S1 subunit receptor-binding domain (RBD) specifically interacts with the host receptor known as angiotensin-converting enzyme 2 or ACE 2.13 At the moment of binding of the S1 domain to a host-cell receptor, the perfusion trimer is destabilized, which results in shedding of the S1 subunit, and transition of the S2 subunit to a stable post fusion conformation.7,10,14 The study of crucial activity of the S-protein can therefore provide a breakthrough in vaccine design and development compared with other structural proteins of SARS-CoV-2 such as nucleocapsid phosphoprotein (N-protein) or envelop protein (E-protein).15

Many efforts have been made for the discovery of antiviral drugs against SARS-CoV-2,16−19 but no such licensed therapeutic are available in the market until date. Therefore, the development of an effective treatment strategy for the pandemic is a research priority. Moreover, the design of a novel vaccine against viruses using kits and related antibodies is time-consuming and expensive.20,21 Previously, numerous approaches including the whole virus, viral-DNA, subunit, and virus-like particles have been employed in developing epitope-based vaccines against SARS and MERS virus.22−27 These epitopes can be prepared by chemical synthesis techniques and are easier in quality control.28,29 There are evidence which supports that in silico predictions are helpful in successful production of commercially important vaccines.30 However, the structural modifications, delivery systems, and adjuvants are the additional requirements in the formulation because of low immunogenicity caused by their structural complexity and low molecular weight.31 Recently, a set of B

Received: September 18, 2020
Accepted: October 29, 2020
Published: November 11, 2020
and T cell epitopes from the highly conserved region in SARS-CoV-2 were identified, which may help in developing vaccine candidates. However, very little is known about the dynamic stability and affinity of the predicted epitope toward the interacting domain of antibody and T-cell receptors (TCRs), which is crucial for validating and improving the efficacy of predicted vaccines. In this respect, we apply a combination of immuno-informatic approach to identify potent epitopes to design the vaccine candidates followed by computational chemistry analysis to check their effectiveness. With the help of molecular docking, MD simulations and free-energy calculations, an analysis of all the important interactions necessary to give stability to the immuno-receptor complexes have been performed.

T-cells are known to recognize and activate defense responses against viral infection; B-cells on the other hand can have antibody reactions which help in recovering extreme respiratory infection. Therefore, we have done a detailed analysis of the viral antigens to predict B-cell, T-cell linear epitopes located at the S-protein of SARS-CoV-2, evaluated their immunogenicity, and designed chimeric vaccines. The conservation of all B- and T- cell epitopes were assessed across most of the isolates and coronavirus species from different parts of the globe. Furthermore, we carried out in silico cloning of the linear vaccine construct to design a recombinant plasmid that can help in expressing the vaccines in E. coli expression system.

The methodology of the exploration of potential epitopes from the S-protein, vaccine design and their validation are discussed in Supporting Information 1 (SI-1). All protein molecules are protonated at the biological pH of 7.0. Then the chimeric vaccine and peptide epitopes are docked at the antigen binding domain of respective immune-receptors by Hex software. We carried out atomistic MD simulations of all the systems using GROMACS-2016.5. Amber 99-SB force-field was employed due to its better balance in β-sheet and helicity propensity compared to other force fields. Water molecules were used to solvate the receptor-peptide systems because of their compatibility with AMBER force field. All the systems underwent a 50 000 step energy minimization by steepest-descent algorithm to remove the steric clash. Leapfrog integrator algorithm was used to integrate the equation of motions with a time step of 2 fs. LINC-α algorithm was applied to constrain all the bonds in the peptide molecule, and SETTLE algorithm was employed to constrain the geometry of water molecules. The systems were equilibrated in canonical ensemble (NVT) followed by isothermal–isobaric ensemble (NPT) for 5 and 10 ns respectively by restraining the solute heavy atoms. Next, the restraints were removed, and the protein molecules were allowed to move freely during the production run for 100 ns. The temperature and pressure of the system was maintained employing Velocity rescale (τf = 0.1 ps) and Parrinello–Rahman coupling algorithm (τp = 0.2 ps). The cutoff for short-range electrostatic and van der Waals interactions was assigned to 1.2 nm for the particle-mesh Ewald method.

The amino-acid sequences of the spike-proteins for twenty-six different coronavirus species (SI-1) were considered for multiple sequence analysis (MSA) to figure out the conserved amino-acid region and the variable region which differentiate the SARS-CoV-2 from other classes of coronavirus. MSA reveals that the C-terminal end is highly conserved compared to the N-terminal end of the amino acid sequence of spike protein retrieved from different coronavirus species (Figure S1). This C-terminal perfusion or shaft domain (S2-domain) is thus common in all coronavirus spices for genome transfer while the receptor binding domain (RBD or S1-domain) is unique in SARS-CoVid2 (Figure 1A). The phylogenic tree constructed from MSA indicates that the spike protein of β-coronavirus species (SARS-CoVid, MERS-CoVid) has maximum structural similarity with the γ-coronavirus (Figure S-2). The antibody-mediated defense responses or humoral immunity against viral infection is believed to be guided by B-cell epitopes. The linear B-cell epitopes of S-glycoprotein sequence (accession no: MT-415323) were predicted by a set of physicochemical parameters, such as the exposed surface propensity by Emin et al., hydrophilicity by Parker et al., flexibility by Karplus et al., antigenic propensity by Kolaskar et al., β-turn propensity by Chou et al. and so on, implemented in the immune epitope database and analysis.
resource (IEDB) tool\textsuperscript{57} and are documented in Tables S11–S15 and in S10. The predicted B-cell and T-cell epitopes with high antigenicity scores are shown in Table 1.

The size of the predicted linear B-cell epitopes is found to be varied between 7mer-15mer which is believed to be of optimum size. The hot-spot residues which are considered as the key to the antigenicity of the epitopes, are found to be mostly polar in nature which may contribute in stable hydrogen bonds to the residues present in the antigen binding region of the antibodies. In Figure 1B, we showed the location of linear B-cell epitopes on the single chain of SARS-CoV-2 spike protein. It is evident from Figure 1B that most of the highly conserved B-cell epitopes are located at the RDB domain and the shaft region of the S-protein. The peptide 808 DPSKPSKRSF 817 and 600 PGQTGKI 607 are found to be common in Emini-surface accessibility, BepiPred linear epitope and Chou–Fasman methods indicating their greater importance as linear B-cell epitopes.

The 3-D structure of the spike protein was generated by homology modeling implemented in Swiss Model\textsuperscript{58} using PDB ID: 6VIB and 6VXX as template. The conformational epitopes of the spike protein of SARS-CoV-2 in their closed and open state were determined by Disctope\textsuperscript{59,60} (V-1.1) server and are documented in Tables S11, S12. It can be seen that the conformational epitopes of spike protein are located almost in the same position (RDB region) in their both closed and open state (Figure S3A, B). The locations of the conformational epitopes on the tertiary structure of S-protein and open state (Figure S1C, B). The locations of the conformational epitopes on the tertiary structure of S-protein are mainly found at 405–427 and 439–505 residue stretches and shown in Figure 1C. Based on the location of the epitopes, the domains are marked as R1, R2, and R3. It can be noted that in both closed and open state, all antibody binding regions (R1, R2, and R3) have extended antiparallel $\beta$-sheet or $\beta$-barrel structure which is an important factor for antigenicity (Figure 1D).

The identification of CD8+ and CD4+ T-cell epitopes are known to be important for eliciting cell-mediated immunity or generation of memory B-cell against viral infections.\textsuperscript{61} Such peptide epitopes are generally presented by the Major Histocompatibility complex (MHC) -class I and MHC-class II molecules expressed on the surface of helper T-cell (T\textsubscript{H}).\textsuperscript{61,62} We employed IEDB server to identify and validate the T-cell peptide epitopes of the S-protein of SARS-CoV-2 which can bind with MHC-I and MHC-II receptors. All MHC-I and based epitope-peptides obtained from IEDB T-cell epitope prediction tools are documented in Tables S13–S16 and that of MHC-II are documented in Tables S17 and S18.

The peptides with highest antigenicity and high affinity are listed in Table 1, and their location at the S-protein of SARS-CoV-2 is depicted in Figure 1E,F. The size of the peptides for MHC binding are found to vary from 8mer to 15 mer. It is evident from Figure 1 that peptide epitope 781 VFAQVKQIY 789 corresponding to MHC-I and 960 NTLVKQLSSNFGA 972 associated with MHC-II are located at the stem part of the S-protein. Unlike, the epitopes 567 RDIADTTDAV 576, 747 VRFPNITNL 754, (related to MHC-I), and 116 SLLIVNNATNVVIK 125 (related to MHC-II) are found to be located at the surface (i.e., the receptor binding domain (RBD) of S-protein). The location of this peptides shares the binding domain of Angiotensin-Converting Enzyme 2 (ACE2) receptor which make them more interesting. These selected peptides are used to design multiepitope vaccine for cell mediated immunity.

The knowledge regarding the linear-epitope-conservation is necessary to design vaccines capable of inducing adaptive immune response for all coronavirus species. It is found that all the epitopes considered here contain at least two conserved residues (Figure S1). Epitopes with three identical or similar residues are considered to be highly conserved. Among all the selected B-cell epitopes in Table 1, 412 PGQTGKI 418, 567 RDIADTTDAV 576, 747 VRFPNITNL 754, (related to MHC-I), and 116 SLLIVNNATNVVIK 125 are found to be highly conserved (Figure S1). In case of the T-cell epitope, 116 SLLIVNNATNVVIK 125, 781 VFAQVKQIY 789 and 960 NTLVKQLSSNFGA 972 are highly conserved. It is evident from Figure 1B that most of the highly conserved B-
cell epitopes are located at the stalk or shaft region of the S-protein except 411PGQTGK518 and 1057PHGVFELHTTVYPAPA1070. Similarly, in the case of T-cell epitopes, highly conserved 116SLLIVNATNVK129, 781VFAQVKQIY785 epitopes are located at the shaft of the S-protein. The rest of the epitopes are less conserved and mostly located at the RBD region of S-protein. This indicates that amino-acid composition of RDB of SARS-CoV-2 S-protein is different from other coronavirus species.

Further, we have designed two separate vaccines from the epitopes predicted in Table 1, and named as Vac-COVID-B, Vac-COVID-T respectively, which can be used as a combination for both humoral and cell-mediated immunity. The linear B-cell epitopes and T-cell epitopes are linked by the GPGPG linker to avoid the formation of junctional epitope. The linear vaccine from the B-cell epitope exhibited 12 hydrogen bond interactions with the variable region of the antibody. The residues Gly38, Gly 50, Glu 41, Gly76, Gly 54, Thr56, Tyr32, Ser31 of Vac-COVID-B exhibited hydrogen bond with Ser56, Tyr58, Tyr52, Ser53, Tyr33, Tyr94 of Vh and Tyr97, Ala99, Ile59 and Lys58 of Vl respectively. The interaction of conformational epitopes with the 7BZ5 at the molecular level is depicted in Figure 3B-D. The conformational epitope R1 (Arg 102, Ser98, Phe140, and Leu242) was found to make hydrogen bond with Tyr100, Tyr32, Tyr 33, and Tyr94 of the antibody molecule (Figure 3B). Similarly, Gly28, Tyr32, Asn92, Tyr58, Tyr52, Tyr33 of 7BZ5 formed hydrogen bonds with Tyr449, Ser494, Tyr453, Arg453, Arg403, Asp405, and Lys417 of R2 (Figure 3C).

In addition, Tyr 94 of the antibody and Tyr 505 of R2 showed π-π interaction which is crucial for the antibody-epitope stabilization. In case of R3, no nonbonded interaction was found (Figure 3D) except one π-π stacking interaction between Tyr 32 and Phe562. The T-cell peptide epitopes were also found to be stabilized by the hydrogen bond formed with the peptide presenting groove of the MHC-receptors. The TCRs are known to recognize the antigens in pieces that are presented by MHC molecules. Therefore, the interaction of conformational epitopes with the 7BZ5 at the molecular level is depicted in Figure 3B–D. The structural quality of the vaccines is assessed by Z-score and Ramachandran plot (Figure S4). The Z-score of the vaccines made of B-cell and T-cell epitopes are −3.43 and −4.85, respectively, which confirm the reliability of our model. Additionally, 95.42% and 92.7% amino acid residues are found to be in the favored region in the Ramachandran plot. Further, the antigenicity of the two vaccines are found to be 0.73 and 0.58 (score >0.4 is considered to be antigenic).

Next, it will be interesting to study the interaction of the multi-epitope vaccines designed with the immune-cell receptor for eliciting stable immune response. We considered the structure of the humanized antibody (7BZ5) as the immune receptor for B-cell epitope, where the antigen binding region of 7BZ5 is formed by a variable region of light chain (Vl) and heavy chain (Vh) (Figure S6A). The antigen binding pocket of the MHC-I molecule (2GTZ, 5WWJ, 6IEX, 5IM7) is formed by the interaction of α1 and α2 domain of α-chain (Figure S6B), whereas the antigen presenting domain of MHC-II (2SEB, 6BIY) molecules is formed by the association of α1 and β1 domains (Figure S6C). However, the peptide presenting platform is similar for both the MHC class-I and class-II molecules. The detailed interaction of the chimeric vaccine with the variable region of 7BZ5 is depicted in Figure 3A by molecular docking. The linear vaccine from the B-cell epitope exhibited 12 hydrogen bond interactions with the variable region of the antibody. The residues Gly38, Gly 50, Glu 41, Gly76, Gly 54, Thr56, Tyr32, Ser31 of Vac-COVID-B exhibited hydrogen bond with Ser56, Tyr58, Tyr52, Ser53, Tyr33, Tyr94 of Vh and Tyr97, Ala99, Ile59 and Lys58 of Vl respectively.

### Figure 2. Tertiary structure of chimeric vaccines made of (A) B-cell linear epitopes (B) T-cell linear epitopes. The amino acid sequence of the corresponding vaccines is shown below. The adjuvant is shown in red; the adjuvant linker is shown in green, and the epitope linkers are indicated in blue.
The docking predicted interaction of the vaccine-receptor complex in motion were assessed by all atom MD-simulation studies. In order to check the reproducibility of our MD results, we carried out another set of simulation and provided Figure 3. Nonbonded interaction of the vaccines and conformational epitopes with the immune receptor. (A) The residues of the linear vaccine involved in the formation of hydrogen bond with 7BZ5. The vaccine is shown in blue color. The interaction of (B) R1 (red), (C) R2 (green), and (D) R3 (brown) with the variable region of 7BZ5. The T-cell epitope peptide presented at the peptide presenting groove of (E) HLA-A*02:01, (F) HLA-A*24:01, (G) HLA-B*40:01, (H) HLA-B*58:01, (I) DRB1*04:01, (J) DRB1*07:01. The interacting residues of the receptor are shown in violet color, whereas the residues of the vaccine or the discontinuous epitope are marked in pink color.
in Supporting Information 2 (Figure S20, S21) and Supporting Information 3 (Table S21). The time evolution of the RMSD, RMSF, radius of gyration (rGyr) of B-cell linear vaccine and epitopes located at R1, R2, R3 region are illustrated in Figure 4. The RMSD of the linear vaccine (Vac-COVID-B) was found to increased approximately 0.7 nm up to initial 8 ns and remain stabilize until 40 ns. The RMSD fluctuation and rGyr profile of Vac-COVID-B near 60–90 ns fluctuation indicate minor secondary structure modification of Vac-COVID-B during the course of simulation (Figure 4A,B). The RMSF profile of the designed vaccine is shown in Figure S7A. The residue stretches 39–48, 57–73 were found to highly fluctuate during the course of simulation because of the interaction with the antibody molecule. The RMSF value of the residues located at antigen binding domain of 7BZ5 is found to be less than 0.3, which helped the chimeric vaccine to get stabilized at the immune complex (Figure S8). The average number of hydrogen bonds between the antibody and the vaccine is calculated to be 9 and remain intact throughout the simulation trajectory (Figure 4E, Figure S9). The docking predicted residue pairs Gly54-Tyr97, Gly54-Tyr52, Thr56-Tyr33 have hydrogen bond occupancy values of 59.4%, 13.7%, 11.3%. The highest hydrogen bond occupancy of 69.7% is found between Tyr58 of the linear vaccine and Pro51 of 7BZ5, which newly evolved during the course of simulation (Table S20A). The structural stability of the conformational epitopes located at R1, R2, R3 is assessed in Figure 4C,D. The RMSD and rGyr profile of the conformational epitope located at R1 and R3 is found to be stable, whereas that of R2 is highly unstable which indicates the conformational change of the epitope during the course of the simulation. The C-terminal and the N-terminal end of all the epitopes have greater fluctuation because of solvent exposure (Figure S7B–D). The average number of hydrogen bonds between the epitope located at R1, R2, and R3 are 2, 7, and 6, respectively. The docking predicted residues pairs were not found to make hydrogen bonds during the course of simulation; rather, the new hydrogen bonds evolved during the course of simulation with a lesser hydrogen bond
occupancy percentage (Table S20B). In the case of R2, the docking predicted residue pair Asn92-Tyr453 showed the highest hydrogen bond occupancy of 99.9% (Table S20C). The epitope located at R3 was found to make stable hydrogen bond with Asn32 and Asn92 with H-bond occupancy of 55.9% and 51.6%, respectively (Table S20D). The number of hydrogen bond profile of R1 is fluctuating compared with other conformational epitopes, whereas the number of hydrogen bond is found to be increasing between the antibody and R3 with respect to simulation time to stabilize the immune complex (Figure S9B−D). The conformational free-energy landscape of the antibody and antigen binding is depicted in Figure S10 to confirm the adequate sampling of the immune complexes. The solvent accessible surface and RMSD of the chimeric vaccine, B-cell conformational epitopes are considered as reaction coordinate. It is evident from FEL graphs that R1 is trapped in a deep minimum, which indicates the single binding conformation throughout the simulation. In the case of other antigens, two prominent binding conformations along with others are present, which evolved during the course of simulation.

We have done MD simulation of individual peptide epitopes with their corresponding HLA compounds. The dynamic nature of the T-cell epitopes with their respective HLA-complex are assessed in the Figure 5. The RMSD of the peptides correspond to MHC-I increased up to 20 ns and stabilized with an average value of 0.45 nm for the rest of the simulation. The peptide associated with MHC-II is found to be stable for the last 30 ns of simulation. Further, we calculated the time evolution of the distance between the peptide and the surface of the binding pocket of MHC molecules (as shown in Figure S11). The distance between the surface of the peptide epitope and the platform of antigen presenting domain was found to be stable throughout the simulation, indicating that the peptides did not diffuse from the peptide presenting grooves. The solvent accessible surface area of the peptide epitopes are found to be stabilized with their initial value,
which confirms the constant solvent exposure at the peptide presenting groove (Figure S12).

The peptide associated with HLA-B*40:01 and DRB1*04:01 is found to have less fluctuation in MHC-I and MHC-II based epitopes, respectively. MHC-I based peptides mostly fluctuate at their N-terminal end whereas; MHC-II based peptides have higher fluctuation at the C-terminal end (Figure 5C,D). The RMSF profile of the binding domain of MHC-I and MHC-II molecules are below 0.3 which helps to stabilize the peptide epitopes inside the peptide presenting pocket (Figure S13).

The average numbers of hydrogen bonds formed between MHC-I based TCRs T-cell epitopes are 4, 8, 6, and 7, respectively. Thus it is evident from Figure S14 that the number of MHC-I based TCRs T-cell epitopes are 4, 8, 6, and 7, respectively. MHC-I based peptides have higher hydrogen bond occupancy of Gln8-Asn77 (87.8%), Tyr10-Tyr80 (89.9%) Tyr10-Thr143 (83.0%) residue pair stabilization of the peptide VFAQVKQIY at the binding domain of HLA-A*02:01-RDIADTTDAV complex (Table S20E). The residue pairs Leu10-Trp147, Phe12-Tyr84 of HLA-A*24-02- TKRFDPNVLPG complex and Pro5-Tyr159, Arg3-Glu63 of HLA-B*40:01- VRFPNITNL has higher hydrogen bond occupancy which stabilizes the peptide-ligand at the peptide presenting domain (Table S20F, G). The higher hydrogen bond occupancy of Gln8-Asn77 (87.8%), Tyr10-Tyr80 (89.9%) Tyr10-Thr143 (83.0%) residue pairs stabilization of the peptide VFAQVKQIY at the binding domain of HLA-A*02:01- (Table S20E). The residue pairs Leu10-Trp147, Phe12-Tyr84 of HLA-A*24-02- TKRFDPNVLPG complex and Pro5-Tyr159, Arg3-Glu63 of HLA-B*40:01- VRFPNITNL has higher hydrogen bond occupancy which stabilizes the peptide-ligand at the peptide presenting domain (Table S20F, G).

The higher hydrogen bond occupancy of Gln8-Asn77 (87.8%), Tyr10-Tyr80 (89.9%) Tyr10-Thr143 (83.0%) residue pairs stabilization of the peptide VFAQVKQIY at the binding domain of HLA-A*02:01- (Table S20E). The residue pairs Leu10-Trp147, Phe12-Tyr84 of HLA-A*24-02- TKRFDPNVLPG complex and Pro5-Tyr159, Arg3-Glu63 of HLA-B*40:01- VRFPNITNL has higher hydrogen bond occupancy which stabilizes the peptide-ligand at the peptide presenting domain (Table S20F, G).

The free energy of binding ($\Delta G_{\text{bind}}$) is believed to be an important thermodynamic quantity to assess the favorable protein–protein interaction as well as their affinity for accurate modeling of biological systems. In this regard, we calculated the binding free energy of the epitopes at the binding domain of immune cell receptors in implicit solvent model by the end-point free-energy method such as MM/PBSA. It is found that both the MHC-II based peptides have remarkably high binding affinity of $\sim$453.59 kJ/mol with the variable region of 7BZ5 which confirms the thermodynamic stability of the complex. The $\Delta G_{\text{elec}}$ term ($\sim$469.00 kJ/mol) between the residues of vaccine and the humanized antibody was found to have important contribution toward the stability of the complex (Table2). The affinity of the epitope located at R1 is the lowest compared with other discontinuous epitopes because of favorable van der Waals energy ($\sim$198.54 kJ/mol). The lower affinity of R2 and R3 is due to higher $\Delta G_{\text{elec}}$ penalty and positive electrostatic interactions respectively. It is found that both the MHC-II based peptides 960NTLVKQLSSNFAGA and 116SLLVIVNATNVKIK exhibit the lowest binding energy of $\sim$493.66 kJ/mol and $\sim$538.71 kJ/mol, respectively. Among the MHC-I peptide epitopes, 78TKRFDPNVLPG associated with HLA-A*24:02 has the lowest binding energy of $\sim$430.3 kJ/mol. All the peptide epitopes related to MHC-molecules have considerably low Gibbs free energy, indicating the stable epitope-TCR complex. It is evident that the electrostatic and van der Waals free-energy terms of the T-cell peptide epitopes are crucial for their stability. The hydrogen bond observed during our MD study is mainly responsible for favorable electrostatic energy contribution, which is dominating the positive solvation free-energy term for the stable interaction of the epitopes at the immune cell receptors. Further, we calculated the free-energy contribution of residues located at the antigen binding domain of immune cell receptors (Figures S17 and S18). The docking predicted residues and residues with higher hydrogen bond occupancy percentage was found to contribute lower energies which stabilize the immune complexes. In the case of chimeric

### Table 2. Binding Affinities (kJ/mol) of the Vaccines toward the Immune Cell Receptors by MM/PBSA Method

| immune receptor complexes | $\Delta G_{\text{elec}}$ | $\Delta G_{\text{sol}}$ | $\Delta G_{\text{RASA}}$ | $\Delta G_{\text{bind}}$ |
|---------------------------|--------------------------|-------------------------|--------------------------|-------------------------|
| Ab-VAC-COVID B            | -145.07                  | -469.00                 | 345.5                    | -158.02                 | -453.59                 |
| Ab-R1                     | -198.54                  | -80.89                  | 73.3                     | -12.27                  | -218.4                  |
| Ab-R2                     | -111.79                  | -41.15                  | 103.12                   | -28.95                  | -78.77                  |
| Ab-R3                     | -75.258                  | 27.69                   | 37.84                    | -8.76                   | -18.52                  |
| TCR (MHC-I)               | HLA-A*02:01-RDIADTTDAV   | -199.89                 | -365.33                  | 367.24                  | -27.98                  | -225.95                 |
| TCR (MHC-I)               | HLA-A*24:02-TKRFDPNVLPG  | -340.64                 | -826.49                  | 779.80                  | -43.00                  | -430.3                  |
| TCR (MHC-I)               | HLA-B*40:01-VRFPNITNL    | -328.14                 | -528.54                  | 602.63                  | -40.70                  | -294.76                 |
| TCR (MHC-I)               | HLA-B*58:01-VFAQVKQIY    | -176.52                 | -384.67                  | 368.43                  | -20.62                  | -213.39                 |
| TCR (MHC-II)              | HLA-DRB1*04:01-NTLVKQLSSNFAGA | -272.99          | -652.81                  | 463.34                  | -31.19                  | -493.66                 |
| TCR (MHC-II)              | HLA-DRB1*07:01-SLLVIVNATNVKIK | -339.73              | -608.54                  | 450.9                   | -41.33                  | -538.71                 |

The $\Delta G_{\text{elec}}, \Delta G_{\text{sol}}, \Delta G_{\text{RASA}}$, and $\Delta G_{\text{bind}}$ are and are indicating the electrostatic, van der Waals, polar solvation, solvent accessible surface energies respectively.

https://dx.doi.org/10.1021/acs.jpclett.0c02846
J. Phys. Chem. Lett. 2020, 11, 9920–9930
vaccine (Vac-COVID-B), Ile S1 of V_{β} and Pro S1 of V_{λ} has maximum free-energy contribution. The amino acid residues with benzene ring (Phe 27, Tyr 32, Tyr 33, Tyr 58, Tyr 94, Tyr 100) have higher energy contribution because of stacking interactions (Figure S17). In the case of T-cell epitopes, the hydrophilic or polar amino acids located at the peptide presenting groove have higher contribution to stabilize the peptide epitopes (Figure S18).

In order to design a recombinant plasmid, we back translated the protein sequences of the vaccines and optimized the codons in the E. coli system for successful expression of linear B-cell (Vac-COVID-B) and T-cell (Vac-COVID-T) vaccines proposed in our immunoinformatic study. The size of c-DNA sequences of the vaccines made from linear B-cell, T-cell epitopes are 399 base pair (bp) and 333 bp long, respectively. The codon adaptation index (CAI) value for both the vaccines were computed to be 1.0 and the percentage of GC content for Vac-COVID-B, Vac-COVID-T are 60.65%, 56.75% respectively which are in the permissible range and hence confirmed their proficient expression in the E-coli K-12 strain. Finally, the c-DNA sequences of the vaccines were inserted computationally at the multiple cloning site (MCS) of the pUC-19 plasmid vector. The restriction map of the recombinant vector is shown in Figure S19.

To conclude, in the present article we tried to design a vaccine based on B-cell and T-cell epitopes present in spike glycoprotein of highly infective SARS-CoV-2. With the help of immunoinformatic studies, we identified the most promising epitopes which are found to be scattered on the RBD and shaft region of the spike protein. The epitopes located in the shaft region are highly conserved. The hotspot residues that are considered as key to the antigenicity are mostly found to be polar in nature, which contributes to stable electrostatic interaction with respective immune receptors. Docking calculations showed the major interaction between the immune–receptor complexes were hydrogen bond and π−π stacking interactions which were found to contribute maximum in the stability, as evident from free-energy decomposition studies. The hydrogen bond evolved during MD simulation studies the antigen and receptors was found to be the main contributor of the electrostatic energy. Vaccines designed from linear B-cell epitopes were found to exhibit a higher number of hydrogen bonds at the binding domain of the antibody compared with the conformational epitopes. All peptide epitopes corresponding to MHC-I and MHC-II showed remarkable stability because of the van der Waals and the electrostatic energy terms. The present article therefore provides deeper biophysical insights toward the stabilization of predicted vaccine candidates with immune cell receptors which will be helpful in further experimental design of potential vaccine against SARS-CoV-2.

**ASSOCIATED CONTENT**

1. Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsl.jpclett.0c02846.

Supporting Information 1: The detailed methodology of the epitope selection, Molecular docking, and MD simulation (PDF)

Supporting Information 2: (A) The tables containing immunogenicity score of the SARS-CoVid2 S-protein resides for B and T-cell epitope identification; (B) The hydrogen bond occupancies found in MD simulation (PDF)

Supporting Information 3: (A) The Multiple sequence analysis of the S-protein sequences obtained from different coronavirus species; (B) Phylogenetic analysis of S-protein from different corona virus species; (C) The graph for immunogenicity profile of S-protein; (D) The residue wise score for conformational B-cell epitopes; (E) The Z-score and Ramachandran plot for designed vaccines; (F) The structural illustration of immune receptor; (G) The RMSF profile of the B-cell chimeric vaccine and conformational epitopes; (H) The distance between the peptide epitope surface and the platform of peptide presenting groove of MHC-molecules; and (I) The restriction map of recombinant plasmid (PDF)

**AUTHOR INFORMATION**

**Corresponding Author**

Debashree Chakraborty — Biophysical and Computational Laboratory, Department of Chemistry, National Institute of Technology Karnataka, Mangalore 575025, India; orcid.org/0000-0002-0142-7941; Email: debashree@nitk.edu.in

**Author**

Bratin Kumar Das — Biophysical and Computational Laboratory, Department of Chemistry, National Institute of Technology Karnataka, Mangalore 575025, India

Complete contact information is available at: https://pubs.acs.org/10.1021/acsl.jpclett.0c02846

Notes

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

Funding from DST, SERB (CRG/2019/000578) is highly acknowledged. BKD acknowledges NIT K for SRF. The authors would like to thank Mr. Omkar Singh for his timely help and contribution in manuscript preparation. We would also like to thank Department of Chemistry, NITK Surathkal for their constant support.

**REFERENCES**

(1) Zheng, J. SARS-CoV-2: An Emerging Coronavirus That Causes a Global Threat. Int. J. Biol. Sci. 2020, 16 (10), 1678–1685.
(2) Sanchez, S.; Lin, Y. T.; Xu, C.; Romero-Severson, E.; Hengartner, N.; Ke, R. The Novel Coronavirus, 2019-nCoV, Is Highly Contagious and More Infectious Than Initially Estimated. medRxiv, February 11, 2020. DOI: 10.1101/2020.02.07.20021154.
(3) Lam, T. T.-Y.; Shum, M. H.-H.; Zhu, H.-C.; Tong, Y.-G.; Ni, X.-B.; Liao, Y.-S.; Wei, W.; Cheung, W. Y.-M.; Li, W.-J.; Li, L.-F. et al. Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China preprint. bioRxiv, February 18, 2020. DOI: 10.1101/2020.02.13.945485.
(4) de Wit, E.; van Doremalen, N.; Falzarano, D.; Munster, V. J. SARS and MERS: Recent Insights into Emerging Coronaviruses. Nat. Rev. Microbiol. 2016, 14 (8), $235–334$.
(5) Cui, J.; Li, F.; Shi, Z.-L. Origin and Evolution of Pathogenic Coronaviruses. Nat. Rev. Microbiol. 2019, 17 (3), 181–192.
(6) Fehr, A. R.; Perlman, S. Coronaviruses: An Overview of Their Replication and Pathogenesis. In Coronaviruses: Methods and Protocols; Maier, H. J., Bickerton, E., Britton, P., Eds.; Methods in Molecular Biology; Springer: New York, NY, 2015; pp 1–23.
(7) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Heieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM Structure of the 2019-NiCoV Spike in the Prefusion Conformation. Science 2020, 367 (6483), 1260–1263.

(8) Bosch, B. J.; van der Zee, R.; de Haan, C. A. M.; Rottier, P. J. M. The Coronavirus Spike Protein Is a Class I Virus Fusion Protein: Structural and Functional Characterization of the Fusion Core Complex. J. Virol. 2003, 77 (16), 8801–8811.

(9) Li, F. Structure, Function, and Evolution of Coronavirus Spike Proteins. Annu. Rev. Virol. 2016, 3 (1), 237–261.

(10) Walls, A. C.; Park, Y.-J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 2020, 181 (2), 281–292.e6.

(11) Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; Zhou, Q. Structural Basis for the Recognition of SARS-CoV-2 by Full-Length Human ACE2. Science 2020, 367 (6485), 1444–1448.

(12) Zhang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, B.; Li, F. Structural Basis of Receptor Recognition by SARS-CoV-2. Nature 2020, 581 (7780), 221–224.

(13) Yuan, M.; Wu, N. C.; Zhu, X.; Lee, C.-C. D.; So, R. T. Y.; Lv, H.; Mok, C. K. P.; Wilson, I. A. A Highly Conserved Cryptic Epitope in the Receptor Binding Domains of SARS-CoV-2 and SARS-CoV. Science 2020, 368 (6491), 630–633.

(14) Gui, M.; Song, W.; Zhou, H.; Xu, J.; Chen, S.; Xiang, Y.; Wang, X. Cryo-Electron Microscopy Structures of the SARS-CoV Spike Glycoprotein Reveal a Prerequisite Conformational State for Receptor Binding. Cell Res. 2017, 27 (1), 119–129.

(15) Zhang, J.; Zeng, H.; Gu, J.; Li, H.; Zheng, L.; Zou, Q. Progress and Prospects on Vaccine Development against SARS-CoV-2. Vaccines 2020, 8 (2), 153.

(16) Ghosh, R.; Chakraborty, A.; Biswas, A.; Chowdhuri, S. Evaluation of Green Tea Polyphenols as Novel Corona Virus (SARS-CoV-2) Main Protease (Mpro) Inhibitors—an in Silico Docking and Molecular Dynamics Simulation Study. J. Biomol. Struct. Dyn. 2020, 1–13.

(17) Ghosh, R.; Chakraborty, A.; Biswas, A.; Chowdhuri, S. Identification of Polyphenols from Broussonetia Papyrifera as SARS CoV-2 Main Protease Inhibitors Using in Silico Docking and Molecular Dynamics Simulation Approaches. J. Biomol. Struct. Dyn. 2020, 1–14.

(18) Misha, S. S.; Ranjan, S.; Sharma, C. S.; Singh, H. P.; Kalra, S.; Kumar, N. Computational Investigation of Potential Inhibitors of Novel Coronavirus 2019 through Structure-Based Virtual Screening. Molecular Dynamics and Density Functional Theory Studies. J. Biomol. Struct. Dyn. 2020, 0 (0), 1–13.

(19) Gutierrez-Villagomez, J. M.; Campos-García, T.; Molina-Torres, J.; López, M. G.; Vázquez-Martínez, J. Alkamides and Piperamines as Potential Antivirals against the Severe Acute Respiratory Syndrome Coronavirus. J. Phys. Chem. Lett. 2020, 11, 8008.

(20) Tahir ul Qamar, M.; Bari, A.; Adee, M. M.; Maryam, A.; Ashfaq, U. A.; Du, X.; Muneer, I.; Ahmad, H. L.; Wang, J. Peptide Vaccine against Chikungunya Virus: Immuno-Informatics Combined with Molecular Docking Approach. J. Transl. Med. 2018, 16, 298.

(21) Chen, W.-H.; Stych, U.; Hotez, P. J.; Bottazzi, M. E. The SARS-CoV-2 Vaccine Pipeline: An Overview. Curr. Top Med. Rep. 2020, 7, 61.

(22) Song, Z.; Xu, Y.; Bao, L.; Zhang, L.; Yu, P.; Qu, Y.; Zhu, H.; Zhao, W.; Han, Y.; Qin, C. From SARS to MERS, Thrusting Coronaviruses into the Spotlight. Viruses 2019, 11 (1), 59.

(23) Yong, C. Y.; Ong, H. K.; Yeap, S. K.; Ho, K. L.; Tan, W. S. Recent Advances in the Vaccine Development against Middle East Respiratory Syndrome-Coronavirus. Front. Microbiol. 2019, 10, 1781.

(24) Schindewolf, C.; Menachery, V. D. Middle East Respiratory Syndrome Vaccine Candidates: Cautious Optimism. Viruses 2019, 11 (1), 74.

(25) Prompetchara, E.; Ketloy, C.; Palaga, T. Immune Responses in COVID-19 and Potential Vaccines: Lessons Learned from SARS and MERS Epidemic. Asian Pac. J. Allergy Immunol. 2020, 38 (1), 1–9.

(26) Liu, W. J.; Zhao, M.; Liu, K.; Xu, K.; Wong, G.; Tan, W.; Gao, G. F. T-Cell Immunity of SARS-CoV: Implications for Vaccine Development against MERS-CoV. Antiviral Res. 2017, 137, 82–92.

(27) Shi, J.; Zhang, J.; Li, S.; Sun, J.; Teng, Y.; Wu, M.; Li, J.; Li, Y.; Hu, N.; Wang, H.; et al. Epitope-Based Vaccine Target Screening against Highly Pathogenic MERS-CoV: An In Silico Approach Applied to Emerging Infectious Diseases. PLoS One 2015, 10 (12), No. e014475.

(28) Gysen, H. M.; Meloen, R. H.; Barteling, S. J. Use of Peptide Synthesis to Probe Viral Antigens for Epitopes to a Resolution of a Single Amino Acid. Proc. Natl. Acad. Sci. U. S. A. 1984, 81 (13), 3998–4002.

(29) Alderete, J.; Neace, C. J. Identification, Characterization, and Synthesis of Peptide Epitopes and a Recombinant Six-Epitope Protein for Trichomonas Vaginalis Serodiagnosis. ImmunoTargets Ther. 2013, 2, 91–103.

(30) Zawawi, A.; Forman, R.; Smith, H.; Mair, I.; Jibril, M.; Albaqui, M. H.; Brass, A.; Derrick, J. P.; Else, K. J. In Silico Design of a T-Cell Epitope Vaccine Candidate for Parasitic Helminth Infection. PLoS Pathog. 2020, 16 (3), No. e1008243.

(31) Azmi, F.; Ahmad Fuad, A. A. H.; Skwarczynski, M.; Toth, I. Recent Progress in Adjuvant Discovery for Peptide-Based Subunit Vaccines. Hum. Vaccines Immunother. 2014, 10 (3), 778–796.

(32) Lin, L.; Ting, S.; Yufei, H.; Wendong, L.; Yubo, F.; Jing, Z. Epitope-Based Peptide Vaccines Predicted against Novel Coronavirus Disease Caused by SARS-CoV-2. Virus Res. 2020, 288, 19082.

(33) Lizbeth, R.-S. G.; Jazmín, G.-M.; José, C.-B.; Marlet, M.-A. Immunoinformatics Study to Search Epitopes of Spike Glycoprotein from SARS-CoV-2 as Potential Vaccine. J. Biomol. Struct. Dyn. 2020, 1–15.

(34) Kar, T.; Narsaria, U.; Basak, S.; Deb, D.; Castiglione, F.; Mueller, D. M.; Srivastava, A. P. A Candidate Multi-Epitope Vaccine from SARS-CoV-2 as Potential Vaccine. J. Immunol. 2020, 204 (5), 1592.

(35) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. J. Comput. Chem. 2005, 26 (16), 1701–1718.

(36) Abraham, M. J.; Murtola, T.; Schulz, R.; Paiva, B.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. SoftwareX 2015, 1–2, 19–25.

(37) Schindewolf, C.; Menachery, V. D. Middle East Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). J. Phys. Chem. Lett. 2020, 11, 8008.

(38) Schindewolf, C.; Menachery, V. D. Middle East Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). J. Phys. Chem. Lett. 2020, 11, 8008.

(39) Schindewolf, C.; Menachery, V. D. Middle East Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). J. Phys. Chem. Lett. 2020, 11, 8008.
