Design of an Epitope-Based Peptide Vaccine against the Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2): A Vaccine Informatics Approach

Arbaaz Khan[#1], Aftab Alam[#2], Nikhat Imam[2,3], Mohd Faizan Siddiqui[4] and Romana Ishrat*[2]

[1]Department of Computer Science, JMI University, New Delhi-110025.
[2]Center for Interdisciplinary Research in Basic sciences, JMI University, New Delhi-110025.
[3]Institute of Computer Science & Information Technology, Department of Mathematics, Magadh University, Bodh Gaya-824234 (Bihar, India).
[4]International Medical Faculty, Osh State University, Osh City, 723500, Kyrgyz Republic (Kyrgyzstan)

Abstract

The recurrent and recent global outbreak of SARS-COV-2 has turned into a global concern which has infected more than one million people all over the globe, and this number is increasing in hours. Unfortunate no vaccine or specific treatment is available, which make it more deadly. An immunoinformatics approach has shown significant breakthrough in peptide-based epitope mapping and opens the new horizon in vaccine development. In this study, we have identified a total of 15 antigenic peptides (including T and B Cells) in the surface glycoprotein of SARS-CoV-2 which found to be 100% conserved with other SARS coronaviruses. Furthermore, the population coverage analysis has found that $CD4^+$ T-cell peptides showed higher cumulative population coverage over to $CD8^+$ peptides in the 16 different geographical regions in the world. Notably, only 09 out of 15 peptides *(LTDEMIAQY*, *IRASANLAA*, *FGAISSVLN*, *VKQLSSNFG*, *FAMQMAYRF*, *FGAGAALQ*, *VITPGTNTS*, *WTAGAAAYY* and *QTQTNSPRRARS*) that have $80\% - 90\%$ identity with experimentally identified epitopes of different organisms including SARS-CoV and this will likely be beneficial for a quick progression of the vaccine design. Moreover, docking analysis suggested that these peptides are tightly bound in the groove of HLA molecules which can induce the T-cell response. Overall this study allows us to determine potent peptide antigen targets in surface glycoprotein on intuitive grounds which open up a new horizon in COVID-19 research. However, this study needs experimental validation by in vitro and in vivo.

Keywords: COVID-19; Vaccinomics; Allergenicity; Immunogenicity; Epitope prediction; T & B cell; Molecular docking; Population Coverage.

***Correspondence**
Romana Ishrat, Ph.D.
Centre for Interdisciplinary Research in Basic Science
Jamia Millia Islamia University
New Delhi-110025, India
romanaishrat@gmail.com
romana05@gmail.com

Note: #=Equal Contribution

Introduction

History is witness to the fact that whenever a major disease comes, it brings tremendous destruction with itself and these destructions are physical and economical. Previously, Cholera, Smallpox, Bubonic plague and Influenza are some of the most brutal killers in human history and killed millions of people all over the world. Nowadays, the world is scared of COVID-19, and its outbreak continues to spread from China to all over the world, and we do not yet know when it will stop. It is a contagious disease caused by SARS family virus named *SARS-COV-2* which has not been previously identified in humans. The initial symptoms of COVID-19 are coughing, throat infection, fever, breath shortness, muscles pain and fatigue; most of the individuals (about 80%) recover without any special treatment but older people and a person with pre-existing medical conditions (cardiovascular disease, diabetes, lungs disease, and cancer) are more likely to develop serious illness. The incubation period for the infection with *SARS-CoV-2* ranges from $2\ to\ 14$ days after exposure(1).

In just four months, the COVID-19 that originally emerged from the Wuhan province of China poses a major public health and governance challenges. The cases have now spread in 210 countries across all the continents. Till $10\ April\ 2020$, more than $1.7\ Million$ infected cases with $100{,}000+$ deaths around the globe have been reported**.** At present, no vaccine or specific treatment is available; however many drug companies are claiming to a developed vaccine like India-based pharmaceutical firm Zydus Cadila has recently start DNA vaccine development program against the COVID-19(2). Moreover, the Health and Human Services (HHS), USA, Sanofi Pasteur and Johnson & Johnson have taken the initiative to develop vaccine and therapeutics to use against COVID-19(3),(4). In a similar way, the Taiwan-based company "Medigen Vaccine Biologics Corp (MVC)" recently publicized that it has teamed up with the US National Institutes of Health (NIH) to develop a vaccine against Covid-19(5).Currently, the scientist of University of Oxford have opened up their COVID-19 vaccines for clinical trials (6).

However, Scientists are endeavouring to find drugs against COVID-19. There is many evidence thus for has revealed many drug candidate including natural compounds, western medicines and traditional Chinese medicines (TCN) that may have potential efficacy against SARS-CoV-2. Morevere, other antiviral drugs like interferon α (IFN-α), lopinavir/ritonavir, chloroquine phosphate, ribavirin, and arbidol are recommended as the tentative treatments COVID-19 induced Pneumonia(7). Currently, there is no specific treatment/medicine or vaccine to cure COVID-19. So there is an urgent need to develop the new vaccines/drugs against COVID-19. In this way, integration of computational techniques provides a novel approach for integrating immunoinformatics approach for the development of vaccines. These methods had earlier been used in the development of new vaccines against diseases like Dengue(8), Malaria(9), influenza(10), Multiple sclerosis(11), and tumor(12). However, these approach generally work through the identification of MHC-1 & II molecules and T-Cell epitopes ($CD8^+$ & $CD4^+$)(13), which particularize the selection of the potential vaccine agents related to the transporter of Antigen Presentation (TAP) molecules(14),(15).

In the current study, we have predicted epitope based vaccine candidates against the SARS-CoV-2 using systematic immunoinformatics approach. We have considered surface glycoprotein of SARS-CoV-2 due to its specific functions. Our work can save the time needed to screen a large number of possible epitopes compared with experimental techniques and also guide the experimental work with high confidence of finding the desired results in vaccine development.

Materials and methods

1. Sequence Retrieval and Analysis

The surface glycoprotein sequence (ID:- QHO62112.1) was obtained from NCBI gene bank database (https://www.ncbi.nlm.nih.gov/gene/) in the FASTA format. We also obtain the secondary structure by using PSIPRED tool (16). It is well known fact that the β-sheet region of the protein is more antigenic(17).

2. Identification of $CD8^+$ T-Cell Peptides (Epitopes)

The identification of $CD8^+$ peptides, The NETCTL_1.2(18) server was used at set threshold value 0.95 to sustain the sensitivity and specificity of 0.90 and 0.95 respectively. We used all the expanded 12 MHC-I super types (including *A1, A2, A3, A24, A26, B7, B8, B27, B39, B44, B58 and B62)* and incorporate the peptide prediction of $MHC\ class - I$ binding molecules and Proteasomal C-terminal cleavage with TAP transports efficiency. These results were accomplished by weighted TAP transport efficiency matrix and then MHC-I binding and proteasomal cleavage efficiency were used to obtain the total scores and translated it into the sensitivity and specificity. We selected peptides as the epitope candidate on the basis of overall score. Further, we check the peptides binding to $MHC\ class - I$ molecules by using MHC-I binding predictions tool(19). The predicted output were given in units of IC_{50} nM (IC_{50} values $< 50\ nM$ = high affinity, $IC_{50} < 500\ nM$ moderate affinity and $IC_{50} < 5000\ nM$ low affinity)(16). We chose those peptides (epitopes) only which have binding affinity $\leq 400\ nm$ for the further analysis. Next, we check the probability of selected peptides for naturally processed and tied with MHC molecule by using MHC-NP tool(20).

3. Identification of $CD4^+$ T-Cell Peptide (Epitope)

$CD4^+$ T Cells play vital role in adaptive immune system and assist the other immune cell's activity. Besides, they indispensable in of B-Cell antibody class switching and affinity

maturation(21). The $CD4^+$ T-cell peptides prediction emphases on the peptide-binding procedure by MHC class II molecules(22). In our study, we have predicted a total of 162 possible peptides with IC50≤100 ((IC_{50} values $< 50\ nM$ = high affinity, $IC_{50} < 500\ nM$ moderate affinity and $IC_{50} < 5000\ nM$ low affinity) by using MHC-II binding predictions tool(23).

4. B-Cell Peptides (Epitopes) Identification

The identification of B-cell peptides (Epitopes) in surface glycoprotein of SARS-CoV-2 was accomplished to find the potent antigen which gives an assurance of humoral immunity. Here, we used Antibody Epitope Prediction tool(24) to find the B-cell antigenicity with classical propensity scale methods including *Bepipred Linear Epitope Prediction 2.0*(25), *Chou & Fasman Beta-Turn Predictio*n(26), *Emini Surface Accessibility Prediction*(27), *Karplus & Schulz Flexibility Prediction*(28), *Kolaskar & Tongaonkar Antigenicity*(29) and *Parker Hydrophilicity Prediction*(30). We selected that region in sequences (as the epitopic region) which were commonly predicted as the epitopic region (above the threshold value) by all the above classical method.

5. Epitope Conservancy and Immunogenicity Analysis

The epitope-conservation defined the degree of resemblance betwixt the peptide and query sequences. Hence, we used Epitope Conservancy Analysis (ECA) tool(31) to find conservancy of our peptides among the other SARS coronaviruses, Bat SARS like coronaviruses and Bat coronaviruses. Additionally, the immunogenicity prediction can reveal the degree of efficiency for individual peptides to produce an immunogenic response. In our study, we used Class I Immunogenicity(32) tool to predict the immunogenicity of the MHC class-I binding peptides.

6. Peptide Structural Modelling and Retrieval of HLA-Molecules

The structure of $CD^{+}8$ and $CD^{+}4$ T-Cell peptides were modelled with an online server PEP-FOLD 3.5(33). The structure of HLA molecules including HLA-B*53:01 (PDB ID: *1A1M*)(34), HLA-B*44:03(PDB ID: *4JQX*)(35) and HLA-DRB1*01:01(1AQD)(36) were we retrieved from the Protein Data Bank (PDB)(37). The structure of peptides and HLA molecules are depicted in Figure 1.

7. Population Coverage Analysis (PCA)

The Population Coverage Analysis tool (38) gives the idea about probable response of each peptide in different countries of the world based on peptide-HLA data genotypic frequency. Our $CD8^{+}$ T-cell (three peptides) and $CD4^{+}$ T-cell (six peptides) peptides and their respective HLA alleles were used for PCA. We selected 16 geographical areas (15 countries and 21 different ethnicities grouped into 16 different geographical areas.) considered were *East Asia, Northeast Asia, South Asia, Southeast Asia, Southwest Asia, Europe, East Africa, West Africa, Central Africa, North Africa, South Africa, West Indies, North America, Central America, South America and Oceania*.

8. Predicted Peptides vs. Epitope Database and Peptide Screening for Autoimmune, Allergic and Toxic Response

In this section, we used predicted peptides (including T-cell and B–cell peptides) to search for homologous epitopes at 80% to 90% identity in Epitope database(39),(40). This study helps to check our peptide identity with human proteome because there is a chance of an autoimmune response due to any kind of molecular mimicry between the predicted peptides (epitope) and the human proteome. Next, we checked the allergic and toxic nature of predicted peptides using AlgPred(41) and ToxinPred(42) tools respectively.

9. Molecular Docking Studies

In this study, molecular docking studies produced important information regarding the orientation-pattern of the peptides in the binding groove of the HLA-molecules as well as which residues are actively involved in the binding. In current, we used **DINC 2.0**(43) online server for docking of peptides. In our study, we have selected three $CD8^+$ T-cell peptides (*LTDEMIAQY, WTAGAAAYY* and *WMESEFRVY)* and five $CD4^+$ T-cell peptides (*IRASANLAA, FGAISSVLN, VKQLSSNFG, FAMQMAYRF* and *FGAGAALQ*) for docking against the HLA molecule including HLA-B*53:01, HLA-B*44:03 and HLA-DRB1*01:01. The finest binding modes are chosen through multiple parallel docking runs, and the corresponding fragments are expanded by addition of few atoms. Now these extended fragments go for next level of docking, at this stage a new subset of flexible degrees of freedom (DoFs) is generated. These flexible DoFs include few "new" atoms (red) and few "old" atoms (blue). This progression remains until the full peptide has been reassembled in its original structure.

Results

1. Sequence Retrieval and Structure Prediction

Viral glycoproteins have a major role in its pathogenesis. The main goal of viral infection is to recognise and bind a receptor on cell surface of the host. It is considered as the surface glycoprotein play important role in immunity and infection. We retrieved envelope surface glycoprotein of SARS-CoV-19 from the of NCBI gene bank database (ID: QHO62112.1). The sequence is 1273 residues long. Additionally, we predicted the secondary structure of the protein by using RaptroX tool(44) which contains α-helix(19%), β-sheets(31%), random coil(48%) and disordered region (4%). The solvent accessibility of a protein sequence (residues) is the surface area of the residue that is accessible to a solvent; it plays a key role in spatial arrangement and

packing of the protein. So we calculated it and found that 36% region is exposed, 34% is buried and 29% is medium region in the sequence. The complete details are given in Supplementary Figure-1.

2. CD8$^+$ T-cell Peptides

The NETCTL server predicted several peptides in SARS-CoV-19 surface glycoprotein but only nine most potent peptides were chosen which have high combinatorial score. We considered only those alleles of MHC-I class for which the peptides showed higher binding affinity ($IC50 \leq 400nm$). Proteasomes played a key role to cleaving the peptide bonds and converting the protein into peptide. The total score of each peptide–HLA interaction was taken into consideration and a greater score meant greater processing efficiency. The three peptides *LTDEMIAQY* (P1), *WTAGAAAYY* (P2) and *WMESEFRVY* (P3) among nine were found to bind with most of the MHC Class-I molecules, including *HLA-B*15:01*, *HLA-B*53:01*, *HLA-A*68:02*, *HLA-B*44:03* and *HLA-B*57:01* but peptides P1, P2 and P3 had maximum probable value of 0.8203, 0.8185 and 0.7539 for the *HLA-B*53:01*, *HLA-B*44:03* and *HLA-B*44:03* respectively. These peptides (P1, P2 and P3) have also maximum identity (100%) for conservancy. Moreover, we did immunogenicity prediction of peptides and get the highest pMHC-I immunogenicity scores of 0.02757 (P1), 0.15259 (P2) and 0.14153 (P3). The details are given in Table 1.

3. CD4$^+$ T-Cell Epitope

We identified 162 peptides with $IC50 \leq 100$ but only six peptides (*FVSNGTHWF*, *IRASANLAA*, *FGAISSVLN*, *VKQLSSNFG*, *FAMQMAYRF*, and *FGAGAALQ*) were found to be interacted with most of the HLA-DRB-1 molecules, the details are given in Table 2.

4. B-Cell Linear Epitopes

The B-cell epitopes comprise of peptides which can easily be used to take the place of antigens for immunizations and antibody production. In this study, we used an amino acid scale-based

method in B-cell antibody epitope prediction tool in which we predict linear epitopes from protein sequence. We found six B-cell linear peptides (including *LTPGDSSSGWTAG, YQAGSTPCNGV, YGFQPTNGVGYQ, VITPGTNTSN, QTQTNSPRRARS* and *LPDPSKPSKR*) in the COVID-19 surface glycoprotein which may capable to inducing the desired immune response as B-cell epitopes (Figure 2).

5. Population Coverage Analysis (PCA)

The population coverage analysis calculates the expected response of predicted peptides (including MHC-I & II) in populations from different geographical areas. In this study, PCA for 16 different geographical areas was carried out for the predicted peptides by considering all the MHC class-I & II molecules. The results suggested that the expected response of these peptides is varying for populations residing in different geographical regions, shown in Figure 3. The tool predicted average population coverage of 21.50% and 51.09% for MHC Class-I&II binding peptide fragments respectively. The PCA of MHC class II binding peptide fragments of SARS-CoV-19 surface glycoprotein revealed maximum population coverage e.g., 75.57%, 73.61% and 72.91% for North America, Europe and East Asia populations (The details are given in Supplementary_Figure 2).

6. Autoimmune, Allergic and Toxic Response

In the past, there is many cases has been reported for the development of autoimmune diseases like systemic lupus anemia, mysthemia gravis (hepatitis B), multiple sclerosis (swine flu), diabetes mellitus (Mumps) and Gullian Barr syndrome (influenza and oral polio vaccine)(45),(46). To avoid autoimmunity, it becomes important to discard those peptide agents which are similar to human proteome. In the our study, we found good results, most of the peptides (including MHC-I & II and B-cell peptides) were found to similar with experimentally identified epitopes of SARS_COV (Spike Glycoprotein) except few (*FVSNGTHWF, LTPGDSSSGWTAG and YGFQPTNGVGYQ*) which are resemble with human and other pathogen's proteome. Next, we have screened all the peptides to check their allergic and toxic nature, all the peptides were found non-Allergen as well as non-toxic (Supplementary_Figure 3).

7. Peptide-HLA Interaction Analysis

To ensure the interaction between the $CD^{+}8$ T-cell peptides (*LTDEMIAQY*, *WTAGAAAYY* and *WMESEFRVY*) and HLA molecules (HLA-B*53:01(1A1M), HLA-B*44:03(4JQX) and HLA-B*44:03(4JQX) respectively), we performed molecular docking analysis and found that *LTDEMIAQY* bind with HLA-B*53:01 having binding affinity $-6.70\ kcal/Mol.$ Similarly, *WTAGAAAYY* and *WMESEFRVY* bind with HLA-B*44:03 having binding affinity $-6.50\ kcal/Mol$ and $-7.90\ kcal/Mol$ respectively (Figure 4). Moreover, all the $CD^{+}4$ T-cell peptides were binds into the groove of HLA-DR molecules (HLA-DRB1*01:01(1AQD)) with good binding affinities e.g., $-6.80\ kcal/Mol$ (with *IRASANLAA*), $-8.50\ kcal/Mol$ (with *FGAISSVLN*), $-7.60\ kcal/Mo$l (with *VKQLSSNFG*), $-8.10\ kcal/Mol$ (with *FAMQMAYRF*) and $-8.60\ kcal/Mol$ (with *FGAGAALQ*) and interactions are shown in Figure 5.

Discussion

The development of vaccines refers to one of the most effective and cost-effective medical and public health achievements of all time. This is a very lengthy, complex, and costly process and requires an involvement of public and private combination. In each year, vaccination programs save over three million lives. The peptide as a choice of vaccine agent has made astonishing move towards vaccine design against the viruses, bacteria and cancer. The peptide vaccine is often synthetic and mimics naturally occurring proteins from pathogens and these peptides based vaccine have shown promising successful results in the past for disease like malaria, dengue, multiple sclerosis, influenza. Besides of these diseases, peptide based vaccines have also been developed against several types of cancer like colorectal cancer, myeloid leukaemia and gastric cancer (47),(48),(49). The identification and design of immunogenic peptides (epitopes) is expensive as well as time consuming. So immunoinformatics approach has made easy to identify potent peptides. In the present study, we have identified $CD8^{+}$ T-cell (three peptides), $CD4^{+}$ T-

cell (six peptides) and B-cell linear peptides (six peptides) from SARS-CoV-19 surface glycoprotein using immunoinformatics approaches but we were more emphasized to study T-cell peptides because vaccines against T-cell epitopes are more promising as they evoke a long-lasting immune response, with antigenic drift, an antigen can easily escape the memory response of antibody(50),(51).

For the MHC class-I binding peptides, the immune response for the top five different geographical regions (highest population coverage range 30-34%) including South Africa_33.78% (South Africa), West Indies_33.12% (Cuba, Jamaica, Martinique, Trinidad and Tobago), East Africa_32.25% (Kenya, Uganda, Zambia and Zimbabwe), Central Africa_30.70% (Cameroon, Central African Republic, Congo, Equatorial Guinea, Gabon, Rwanda and Sao Tome & Principe) and East Asia_30% (Japan, South Korea and Mongolia). Similarly, for the MHC class-II binding peptides, the excepted immune response was found to be remarkable (PCA range 60%–76%) for North America (Canada, Mexico and the United States), Europe (Austria, Belarus, Belgium, Bulgaria, Croatia, Czech Republic, Denmark, England, France, Georgia, Germany, Greece, Ireland Northern, Ireland South, Italy, Macedonia, Netherlands, Norway, Poland, Portugal, Romania, Russia, Scotland, Serbia, Slovakia, Spain, Sweden, Switzerland, Turkey, Ukraine, United Kingdom and Wales), East Asia (Japan, South Korea and Mongolia), South Asia (India, Pakistan and Sri Lanka), North Africa (Algeria, Ethiopia, Mali, Morocco, Sudan and Tunisia) and West Indies (Cuba, Jamaica, Martinique, Trinidad and Tobago). Thus, our results suggested that MHC class-II binding peptides may be potent vaccine agents for different populations of the globe.

We know that vaccination is the utmost prevention of epidemiologic infectious diseases but it has low incidence of serious systemic adverse effects (like autoimmune diseases). We have predicted a total of 15 peptides (T-cell and B-cell) but seven peptides are very important including *LTDEMIAQY, IRASANLAA, FGAISSVLN, VKQLSSNFG, FAMQMAYRF, FGAGAALQ* and *VITPGTNTSN* because their peptide fragments are matching with experimentally identified epitopes of glycoprotein of Severe acute respiratory syndrome-related

coronavirus (SARS_CoV)(52–56). So these peptides can be used as vaccine agent against the SARS_CoV-2. Besides, peptide *WTAGAAAYY* and *QTQTNSPRRARS* are similar with epitopes of Human *herpesvirus-4* (with major DNA-binding protein) and Bordetella *pertussis* (with filamentous hemagglutinin) respectively and other three peptides (*WMESEFRVY, YQAGSTPCNG* and *LPDPSKPSKR*) were not found resemble with any organism's proteome. Hence, these peptides (say unique peptides) can also proposed as candidates for further studies in the area of vaccines development, therapeutic antibodies and diagnostic tools against the SARS_CoV-2. Other two peptides *LTPGDSSSGWTAG* and *YGFQPTNGVGYQ* which shown resembles with Mus musculus (mouse) or Homo sapiens (human) proteome were eliminated from the study to avoid autoimmunity risk.

As we know that HLA-alleles are polymorphic in nature; so peptide can interact with it. In our study, MHC class-I binding peptides (8-9 residues long) are binds in the wide groove (~1.25À) of HLA molecules (HLA-B*53:01 and HLA-B*44:03). Peptide (*LTDEMIAQY)* bind with HLA-B*53:01 by *TRP-147, TYR-84, ARG-62* residues and surrounding residues which gives it polymorphic nature lie at the interface of this helix and the bottom of the peptide-binding site, namely at positions *ASN-63, PHE- 67,* and *TYR-74*. In docking with HLA-B*44:03, the *WTAGAAAYY* and W*MESEFRVY* peptide got expanded conformation within the same trench and both peptides were surrounded by polymorphic positions (*TYR*-9, *THR-24* and *LEU-163*). Further, we check the interaction of MHC class-II binding peptides with HLA-DR molecules. All the five peptides are buried by interactions with the *GLU-28, SER-53, TYR-60, ASN-62, ASP-66, ASN-69, ARG-71, HIS-81 and ASN-8* and are preferred bind in the groove of HLA-DR molecules. The docking studies suggested that these peptides are preferably bound into the groove of respective alleles and can induce the $CD^{+}8$ and $CD^{+}4$ T-cell immunity.

Since, there is currently no vaccine or specific treatment for COVID-19 and vaccine development is a long way from being translated into practical applications. So at this crunch situation, we have suggested that the predicted peptides would be capable to prompt an effective immune response as a peptide vaccine against the COVID-19.

Conclusion

In our study, integrated computational approaches have identified a total of 15 peptides (T and B cells) from SARS-CoV-19 surface glycoprotein, of which nine peptides resemble with experimentally identified epitopes of SARS-CoV and other pathogen. Consequently, these peptides may be used for synthetic vaccine design to combat with emerging COVID-19 at this deadly situation. However, in vivo and in vitro, experimental validation is needed for their efficient use as a vaccine candidate.

Author Contributions

RI and AA conceived the study design instructed on data analysis. AK and AA curated data and performed statistical analyses. N.I and MFS curated data and draw figures. All the authors read, edited and approved of the manuscript.

Funding

This work did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

The R.I and AK are grateful to the Centre for Interdisciplinary Research in Basic Sciences (CIRBSc), Jamia Millia Islamia-110025 for providing the research infrastructure. A.Alam acknowledges the Department of Health Research (DHR) New Delhi, India for the award of "Young Scientist" fellowship (R.12014/06/2019-HR). N.Imam thankful to Indian Council of Medical Research (ICMR), New Delhi, India. Md.F.Siddiqui is grateful to Prof. Muratov Zhanybek Kudaibakovich and Dr. Syed Ali Abbas from Osh State University, Kyrgyzstan for their guidance and support

References

1. Guo Y-R, Cao Q-D, Hong Z-S, Tan Y-Y, Chen S-D, Jin H-J, Tan K-S, Wang D-Y, Yan Y. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. *Mil Med Res* (2020) **7**:11. doi:10.1186/s40779-020-00240-0

2. Pharmaceutical Tech. Zydus Cadila looks to expedite Covid-19 vaccine development. (2020) Available at: https://www.pharmaceutical-technology.com/news/zydus-cadila-covid-19-vaccine/

3. Stephanie Soucheray. HHS partners with drug makers on COVID-19 vaccine, drugs. (2020) Available at: http://www.cidrap.umn.edu/news-perspective/2020/02/hhs-partners-drug-makers-covid-19-vaccine-drugs

4. SARS treatments could help with the new coronavirus. Why were they shelved? *fortune* (2020) Available at: https://fortune.com/2020/02/18/coronavirus-sars-vaccine-development/

5. Taiwan firm teams up with US to develop Covid-19 vaccine. *The Star* (2020) Available at: https://www.thestar.com.my/news/regional/2020/02/17/taiwan-firm-teams-up-with-us-to-develop-covid-19-vaccine

6. University of Oxford. Oxford university begins enrolling over 500 volunteers for coronavirus vaccine trial Read more at: https://economictimes.indiatimes.com/news/international/world-news/oxford-university-begins-enrolling-over-500-volunteers-for-coronavirus-vaccine-trial/articleshow/74864754.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst. *Econ Times* (2020) Available at: https://economictimes.indiatimes.com/news/international/world-news/oxford-university-begins-enrolling-over-500-volunteers-for-coronavirus-vaccine-trial/articleshow/74864754.cms?from=mdr

7. Dong L, Hu S, Gao J. Discovering drugs to treat coronavirus disease 2019 (COVID-19). *Drug Discov Ther* (2020) **14**:58–60. doi:10.5582/ddt.2020.01012

8. Tambunan. n Silico Analysis of Envelope Dengue Virus-2 and Envelope Dengue Virus-3 Protein as the Backbone of Dengue Virus Tetravalent Vaccine by Using Homology Modeling Method. *OnLine J Biol Sci* (2009) **9**:6–16. doi:10.3844/ojbsci.2009.6.16

9. López JA, Weilenman C, Audran R, Roggero MA, Bonelo A, Tiercy JM, Spertini F, Corradin G. A synthetic malaria vaccine elicits a potent CD8(+) and CD4(+) T lymphocyte immune response in humans. Implications for vaccination strategies. *Eur J Immunol* (2001) **31**:1989–1998. doi:10.1002/1521-4141(200107)31:7<1989::aid-immu1989>3.0.co;2-m

10. Shahsavandi S, Ebrahimi MM, Sadeghi K, Mahravani H. Design of a heterosubtypic epitope-based peptide vaccine fused with hemokinin-1 against influenza viruses. *Virol Sin* (2015) **30**:200–207. doi:10.1007/s12250-014-3504-0

11. Bourdette DN, Edmonds E, Smith C, Bowen JD, Guttmann CRG, Nagy ZP, Simon J, Whitham R, Lovera J, Yadav V, et al. A highly immunogenic trivalent T cell receptor peptide vaccine for multiple sclerosis. *Mult Scler Houndmills Basingstoke Engl* (2005) **11**:552–561. doi:10.1191/1352458505ms1225oa

12. Knutson KL, Schiffman K, Disis ML. Immunization with a HER-2/neu helper peptide vaccine generates HER-2/neu CD8 T-cell immunity in cancer patients. *J Clin Invest* (2001) **107**:477–484. doi:10.1172/JCI11752

13. Petrovsky N, Brusic V. Computational immunology: The coming of age. *Immunol Cell Biol* (2002) **80**:248–254. doi:10.1046/j.1440-1711.2002.01093.x

14. Brusic V, Bajic VB, Petrovsky N. Computational methods for prediction of T-cell epitopes--a framework for modelling, testing, and applications. *Methods San Diego Calif* (2004) **34**:436–443. doi:10.1016/j.ymeth.2004.06.006

15. Nielsen M, Lundegaard C, Lund O, Keşmir C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* (2005) **57**:33–41. doi:10.1007/s00251-005-0781-7

16. Buchan DWA, Jones DT. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res* (2019) **47**:W402–W407. doi:10.1093/nar/gkz297

17. Hasan MA, Hossain M, Alam MJ. A computational assay to design an epitope-based Peptide vaccine against saint louis encephalitis virus. *Bioinforma Biol Insights* (2013) **7**:347–355. doi:10.4137/BBI.S13402

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

19. Lundegaard C, Lund O, Nielsen M. Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers. *Bioinforma Oxf Engl* (2008) **24**:1397–1398. doi:10.1093/bioinformatics/btn128

20. Giguère S, Drouin A, Lacoste A, Marchand M, Corbeil J, Laviolette F. MHC-NP: predicting peptides naturally processed by the MHC. *J Immunol Methods* (2013) **400–401**:30–36. doi:10.1016/j.jim.2013.10.003

21. Fraser CC, H. Altreuter D, Ilyinskii P, Pittet L, LaMothe RA, Keegan M, Johnston L, Kishimoto TK. Generation of a universal CD4 memory T cell recall peptide effective in humans, mice and non-human primates. *Vaccine* (2014) **32**:2896–2903. doi:10.1016/j.vaccine.2014.02.024

22. Oyarzún P, Ellis JJ, Bodén M, Kobe B. PREDIVAC: CD4+ T-cell epitope prediction for vaccine design that covers 95% of HLA class II DR protein diversity. *BMC Bioinformatics* (2013) **14**:52. doi:10.1186/1471-2105-14-52

23. Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, Peters B. Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics* (2010) **11**:568. doi:10.1186/1471-2105-11-568

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

25. Larsen J, Lund O, Nielsen M. [No title found]. *Immunome Res* (2006) **2**:2. doi:10.1186/1745-7580-2-2

26. Chou PY, Fasman GD. Empirical Predictions of Protein Conformation. *Annu Rev Biochem* (1978) **47**:251–276. doi:10.1146/annurev.bi.47.070178.001343

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

28. Karplus PA, Schulz GE. Prediction of chain flexibility in proteins: A tool for the selection of peptide antigens. *Naturwissenschaften* (1985) **72**:212–213. doi:10.1007/BF01195768

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

30. Parker JMR, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and x-ray-derived accessible sites. *Biochemistry* (1986) **25**:5425–5432. doi:10.1021/bi00367a013

31. Bui H-H, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics* (2007) **8**:361. doi:10.1186/1471-2105-8-361

32. Calis JJA, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, Keşmir C, Peters B. Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput Biol* (2013) **9**:e1003266. doi:10.1371/journal.pcbi.1003266

33. Thévenet P, Shen Y, Maupetit J, Guyon F, Derreumaux P, Tufféry P. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Res* (2012) **40**:W288-293. doi:10.1093/nar/gks419

34. Smith KJ, Reid SW, Harlos K, McMichael AJ, Stuart DI, Bell JI, Jones EY. Bound water structure and polymorphic amino acids act together to allow the binding of different peptides to MHC class I HLA-B53. *Immunity* (1996) **4**:215–228. doi:10.1016/s1074-7613(00)80430-6

35. Rist MJ, Theodossis A, Croft NP, Neller MA, Welland A, Chen Z, Sullivan LC, Burrows JM, Miles JJ, Brennan RM, et al. HLA Peptide Length Preferences Control CD8 $^{+}$ T Cell Responses. *J Immunol* (2013) **191**:561–571. doi:10.4049/jimmunol.1300292

36. Murthy VL, Stern LJ. The class II MHC protein HLA-DR1 in complex with an endogenous peptide: implications for the structural basis of the specificity of peptide binding. *Struct Lond Engl 1993* (1997) **5**:1385–1396. doi:10.1016/s0969-2126(97)00288-8

37. Berman HM. The Protein Data Bank. *Nucleic Acids Res* (2000) **28**:235–242. doi:10.1093/nar/28.1.235

38. Bui H-H, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics* (2006) **7**:153. doi:10.1186/1471-2105-7-153

39. Fleri W, Vaughan K, Salimi N, Vita R, Peters B, Sette A. The Immune Epitope Database: How Data Are Entered and Retrieved. *J Immunol Res* (2017) **2017**:5974574. doi:10.1155/2017/5974574

40. Dhanda SK, Mahajan S, Paul S, Yan Z, Kim H, Jespersen MC, Jurtz V, Andreatta M, Greenbaum JA, Marcatili P, et al. IEDB-AR: immune epitope database-analysis resource in 2019. *Nucleic Acids Res* (2019) **47**:W502–W506. doi:10.1093/nar/gkz452

41. Saha S, Raghava GPS. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res* (2006) **34**:W202-209. doi:10.1093/nar/gkl343

42. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Open Source Drug Discovery Consortium, Raghava GPS. In silico approach for predicting toxicity of peptides and proteins. *PloS One* (2013) **8**:e73957. doi:10.1371/journal.pone.0073957

43. Antunes DA, Moll M, Devaurs D, Jackson KR, Lizée G, Kavraki LE. DINC 2.0: A New Protein–Peptide Docking Webserver Using an Incremental Approach. *Cancer Res* (2017) **77**:e55–e57. doi:10.1158/0008-5472.CAN-17-0511

44. Wang S, Li W, Liu S, Xu J. RaptorX-Property: a web server for protein structure property prediction. *Nucleic Acids Res* (2016) **44**:W430–W435. doi:10.1093/nar/gkw306

45. Cohen AD, Shoenfeld Y. Vaccine-induced autoimmunity. *J Autoimmun* (1996) **9**:699–703. doi:10.1006/jaut.1996.0091

46. Jain S, Baranwal M. Computational analysis in designing T cell epitopes enriched peptides of Ebola glycoprotein exhibiting strong binding interaction with HLA molecules. *J Theor Biol* (2019) **465**:34–44. doi:10.1016/j.jtbi.2019.01.016

47. Maslak PG, Dao T, Bernal Y, Chanel SM, Zhang R, Frattini M, Rosenblat T, Jurcic JG, Brentjens RJ, Arcila ME, et al. Phase 2 trial of a multivalent WT1 peptide vaccine (galinpepimut-S) in acute myeloid leukemia. *Blood Adv* (2018) **2**:224–234. doi:10.1182/bloodadvances.2017014175

48. Sundar R, Rha SY, Yamaue H, Katsuda M, Kono K, Kim HS, Kim C, Mimura K, Kua L-F, Yong WP. A phase I/Ib study of OTSGC-A24 combined peptide vaccine in advanced gastric cancer. *BMC Cancer* (2018) **18**:332. doi:10.1186/s12885-018-4234-8

49. Melief CJM, van der Burg SH. Immunotherapy of established (pre)malignant disease by synthetic long peptide vaccines. *Nat Rev Cancer* (2008) **8**:351–360. doi:10.1038/nrc2373

50. Chiou S-S, Fan Y-C, Crill WD, Chang R-Y, Chang G-JJ. Mutation analysis of the cross-reactive epitopes of Japanese encephalitis virus envelope glycoprotein. *J Gen Virol* (2012) **93**:1185–1192. doi:10.1099/vir.0.040238-0

51. Alam A, Ali S, Ahamad S, Malik MdZ, Ishrat R. From ZikV genome to vaccine: *in silico* approach for the epitope-based peptide vaccine against Zika virus envelope glycoprotein. *Immunology* (2016) **149**:386–399. doi:10.1111/imm.12656

52. He Y, Zhou Y, Wu H, Luo B, Chen J, Li W, Jiang S. Identification of immunodominant sites on the spike protein of severe acute respiratory syndrome (SARS) coronavirus: implication for developing SARS diagnostics and vaccines. *J Immunol Baltim Md 1950* (2004) **173**:4050–4057. doi:10.4049/jimmunol.173.6.4050

53. Yang J, James E, Roti M, Huston L, Gebe JA, Kwok WW. Searching immunodominant epitopes prior to epidemic: HLA class II-restricted SARS-CoV spike protein epitopes in unexposed individuals. *Int Immunol* (2009) **21**:63–71. doi:10.1093/intimm/dxn124

54. ohn Sidney; Jason Botten; Benjamin Neuman; Michael Buchmeier; Alessandro Sette. HLA DRB1*01:01 binding capacity of selected SARS-derived peptides. *IEDB_SUBMISSION* (2006) Available at: https://www.iedb.org/reference/1000399

55. Wang B, Chen H, Jiang X, Zhang M, Wan T, Li N, Zhou X, Wu Y, Yang F, Yu Y, et al. Identification of an HLA-A*0201–restricted CD8+ T-cell epitope SSp-1 of SARS-CoV spike protein. *Blood* (2004) **104**:200–206. doi:10.1182/blood-2003-11-4072

56. Guo J-P, Petric M, Campbell W, McGeer PL. SARS corona virus peptides recognized by antibodies in the sera of convalescent cases. *Virology* (2004) **324**:251–256. doi:10.1016/j.virol.2004.04.017

Abbreviations

WHO	World Health Organization	CDC	Center for Disease Control and Prevention
COVID-19	Coronavirus Disease-2019	HCoVs	Human Coronaviruses
HBV	Hepatitis B Virus	MERS-CoV	Middle East Respiratory Syndrome Coronavirus
SARS-CoV	Severe Acute Respiratory Syndrome Coronavirus	MHC Class-I & II	Major Histocompatibility Complex Class-I & II
HLA	Human Leukocyte Antigen	CD8 and CD4	Cluster Of Differentiation 8 and 4
T and B Cell	Thymus **Cells and B**one Marrow- or Bursa-Derived **Cells**	PCA	Population Coverage Analysis
IEDB	Immune Epitope Database		

Captions

Figures	
Figure 1	(A) The 3D structure of HLA molecules including *HLA-B*53:01*, *HLA-B*44:03* and *HLA-DRB1*01:01*. (B) The peptide structure of three peptides for CD8+ T-cells (*LTDEMIAQY, WTAGAAAYY* and *WMESEFRVY*). (C) The structure of six CD4+ T-cells peptides including *FVSNGTHWF, IRASANLAA, FGAISSVLN, VKQLSSNFG, FAMQMAYRF*, and *FGAGAALQ*.
Figure 2	Six Peptides representing the B cell epitopes which have highest antigenic propensity and surrounded by six different coloured lines, each line indicating different analysis methods (Bepipred linear epitope prediction, Chou & Fasman beta-turn prediction, Emini surface accessibility prediction, Karplus & Schulz flexibility prediction, Kolasker & Tongaonkar antigenicity prediction and Parker hydrophilicity prediction.) with the maximum scores.
Figure 3	Population coverage of the identified peptides in 16 geographical regions of the world: (i) Population coverage represents the fraction of individuals expected to respond to a given epitope set, (ii) Average number of epitope hits / HLA combinations recognized by the population and (iii) Minimum number of epitope hits/HLA combinations recognized by 90% of the population (PC90).
Figure 4	(A) Peptide LTDEMIAQY binds in the groove of the HLA-B*53:01. Residues TRP-147, TYR-84 and ARG-62 form regular hydrogen bonds. (B &C) Two epitopes *WTAGAAAYY* and *WMESEFRVY* bind in the groove of the HLA-B*44:03. Residues which form hydrogen bond are GLN-155, TYR-99, ASN-17, TRP-147 and THR-69, TYR-171, SER-167, TYR-99, TRP-147, GLU-76 respectively.
Figure 5	Five peptides including *IRASANLAA*(green), *FGAISSVLN*(magenta), *VKQLSSNFG*(yellow), *FAMQMAYRF*(blue) and *FGAGAALQI*(red) bind in the groove of HLA-DRB1*01:01. Most of the residues commonly form H-bond with all the peptides, e.g., GLU-28, SER-53, TYR-60, ASN-62, ASP-66, ASN-69, ARG-71, HIS-81 and ASN-8.
Tables	
Table 1	The three potential CD8+ T-cell epitopes along with their interacting MHC class-I alleles and NetCTL combine score, epitopes conservancy hits and pMHC-I immunogenicity score.
Table 2	The selected six most potential CD4+ T-cell epitopes along with their interacting MHC class II alleles with affinity IC50 < 100.
Supplementary Data	
Suppli_1	Secondary Structure composition and Solvent Accessibility of SARS-CoV-2 Surface Glycoprotein.
Suppli_2	Population Coverage of 16 geographical regions including MHC class I and MHC class I.
Suppli_3	Toxicity and Allergy prediction results.

Table 1: The three potential CD8+ T-cell epitopes along with their interacting MHC class-I alleles and NetCTL combine score, epitopes conservancy hits and pMHC-I immunogenicity score.

Peptide	Position	NetCTL Combined Score	MHC-1 IC50 Score <400	MHC-NP Prediction Result	pMHC-I immunogenicity score	Epitope Conservancy Hit (%)

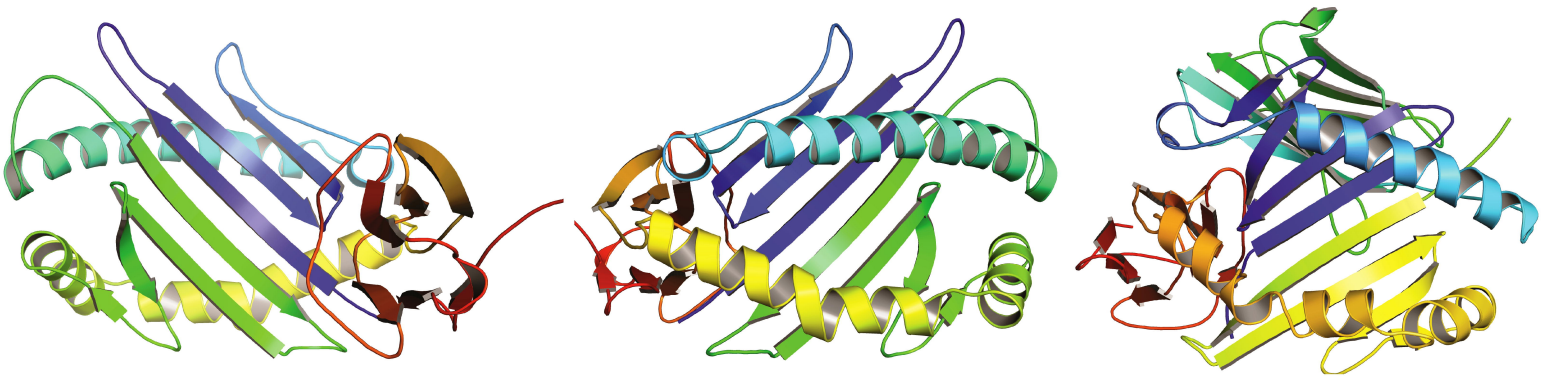
LTDEMIAQY	864-873	A1= 3.66 A2 =1.06	3.37 (0.2) 156 (0.74) 261 (0.54) 330 (0.79) 359(0.81) 376(0.87)	0.9457 HLA-B*57:01 0.8907 HLA-B*57:01 0.8203 HLA-B*53:01 **0.7935 HLA-B*53:01** 0.7806 HLA-B*53:01 0.7793 HLA-B*15:01	0.02757	100
WTAGAAAYY	258-267	A1=3.11 A26=2.00	68.6 (0.18) 48.4 (0.44) 132 (0.66) 93.7 (0.6) 91.8(0.59) 89.9(0.55)	0.9404 HLA-B*44:03 0.873 HLA-B*53:01 0.8287 HLA-B*44:03 0.8185 HLA-B*44:03 0.7519 HLA-B*57:01 0.7412 HLA-A*68:02	0.15259	100
WMESEFRVY	151-160	A1=1.92 B62=1.24	49.7 (0.29) 265 (0.55)	0.8901 HLA-B*57:01 0.8378 HLA-B*44:03 0.8341 HLA-B*57:01 0.7539 HLA-B*44:03 0.734 HLA-B*57:01	0.14153	100

Table 2: The selected six most potential CD4+ T-cell epitopes along with their interacting MHC class II alleles with affinity IC50 < 100

S.NO.	Epitopes	Position	Interacting MHC class II allele	IC50 Values (< 100)
1	FVSNGTHWF	1094-1108	HLA-DRB1*01:01	21.43

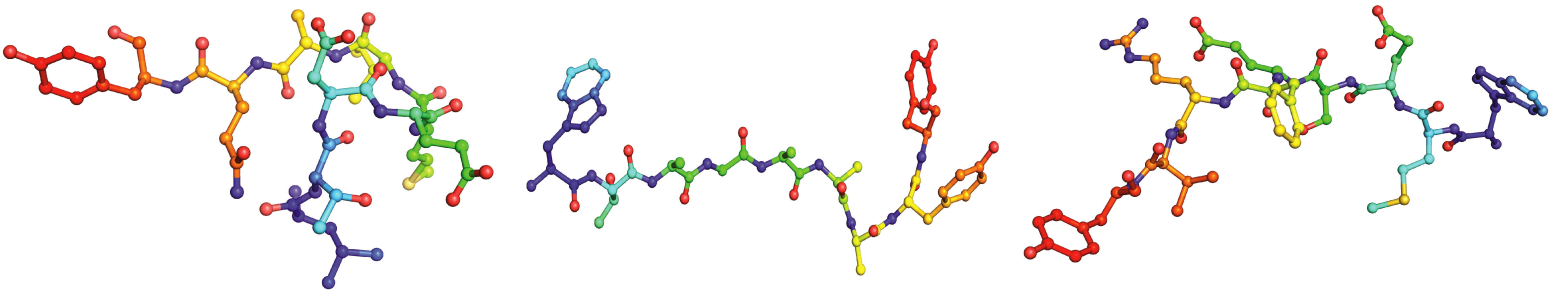
			HLA-DRB1*13:02 HLA-DRB1*07:01 HLA-DRB1*13:02 HLA-DRB1*09:01	93.35 89.95 47.25 98.31
2	IRASANLAA	1017-1031	HLA-DRB1*01:01 HLA-DRB1*04:01 HLA-DRB1*13:02 HLA-DRB1*07:01 HLA-DRB1*09:01	19.9 97.67 42.12 84 88.15
3	FGAISSVLN	969-983	HLA-DRB1*01:01 HLA-DRB1*04:01 HLA-DRB1*04:05 HLA-DRB1*07:01 HLA-DRB1*09:01	20.2 98.03 96.39 86.61 88.46
4	VKQLSSNFG	960-974	HLA-DRB1*01:01 HLA-DRB1*15:01 HLA-DRB1*04:01	22.84 91.6 85.68
5	FAMQMAYRF	897-911	HLA-DRB1*01:01 HLA-DRB1*07:01 HLA-DRB1*09:01 HLA-DRB1*11:01 HLA-DRB1*15:01	12.49 42.09 53.55 48.76 50.31
6	FGAGAALQI	887-901	HLA-DRB1*01:01 HLA-DRB1*09:01 HLA-DRB1*07:01	16.21 24.83 48.18

(A). HLA Molecules



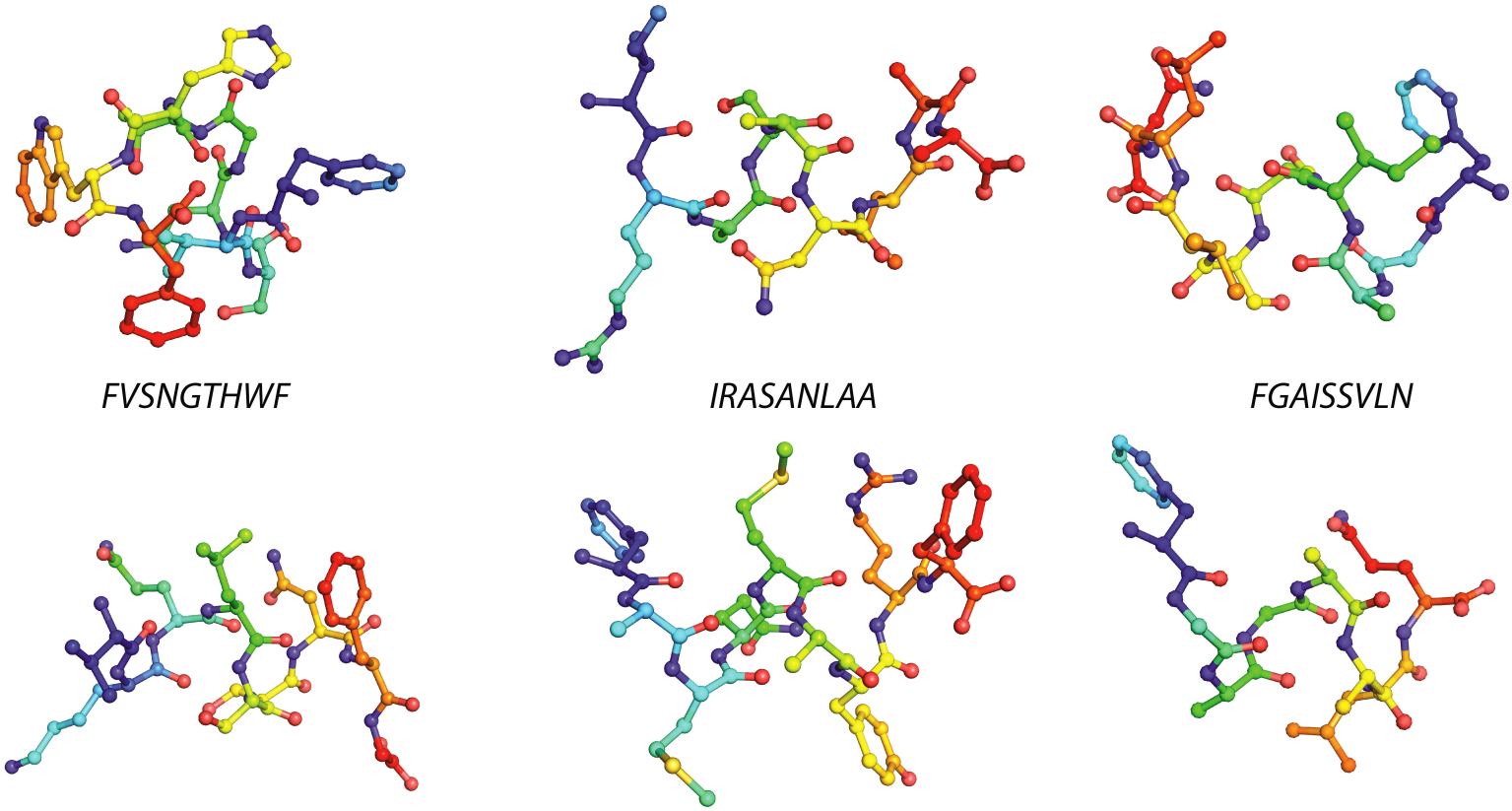
HLA-B*53:01 (PDB ID: 1A1M) HLA-B*44:03(PDB ID: 4JQX) HLA-DRB1*01:01(1AQD)

(B). CD8+ T-Cell Peptide (Epitope)



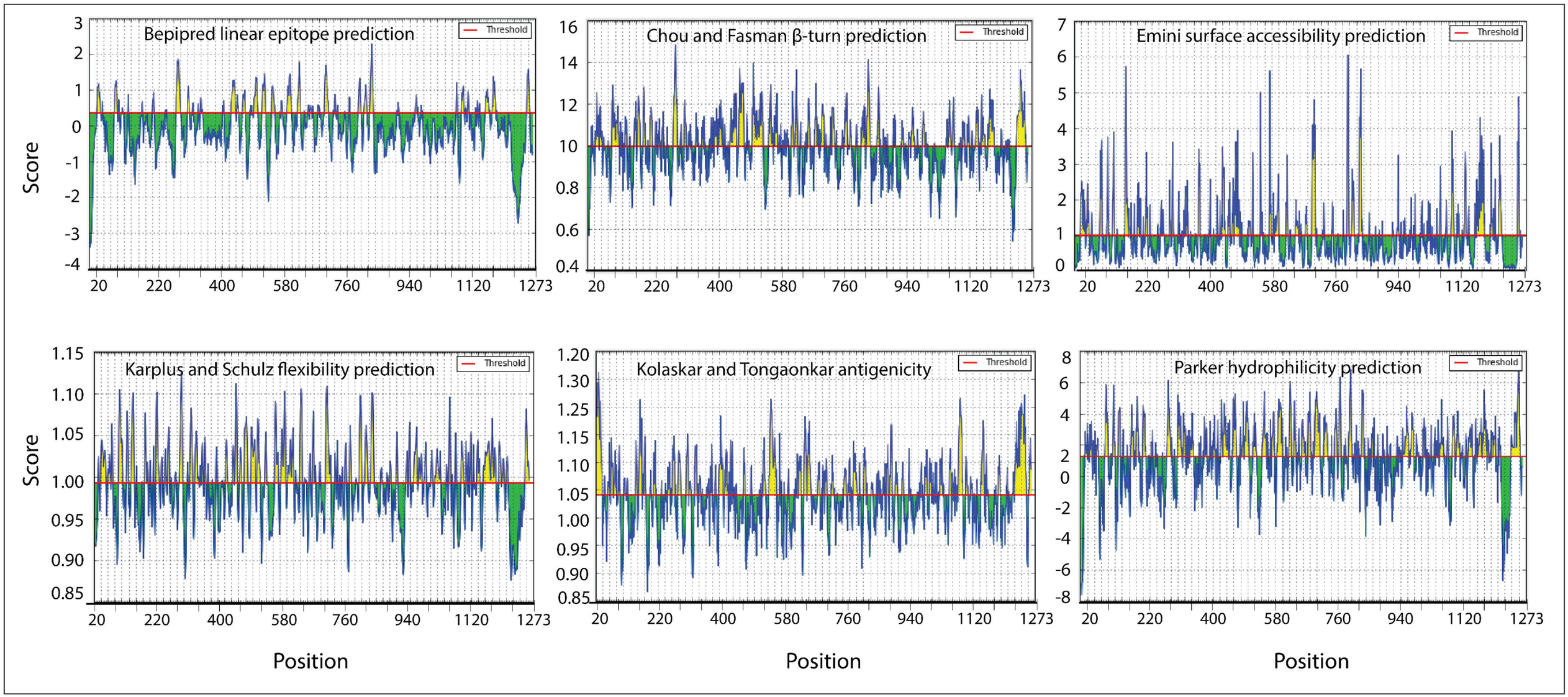
LTDEMIAQY *WTAGAAAYY* *WMESEFRVY*

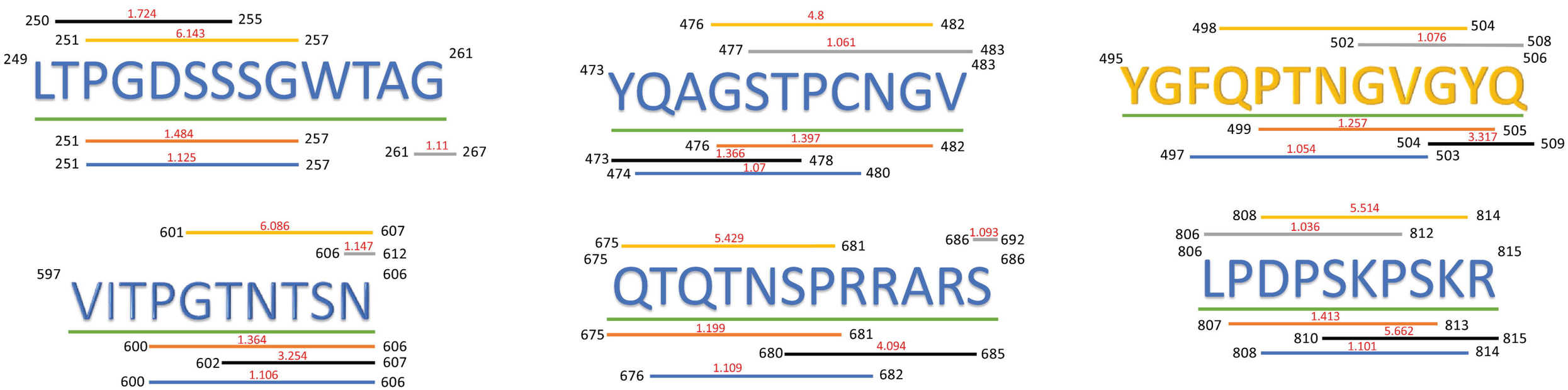
(C). CD4+ T-Cell Peptide (Epitope)



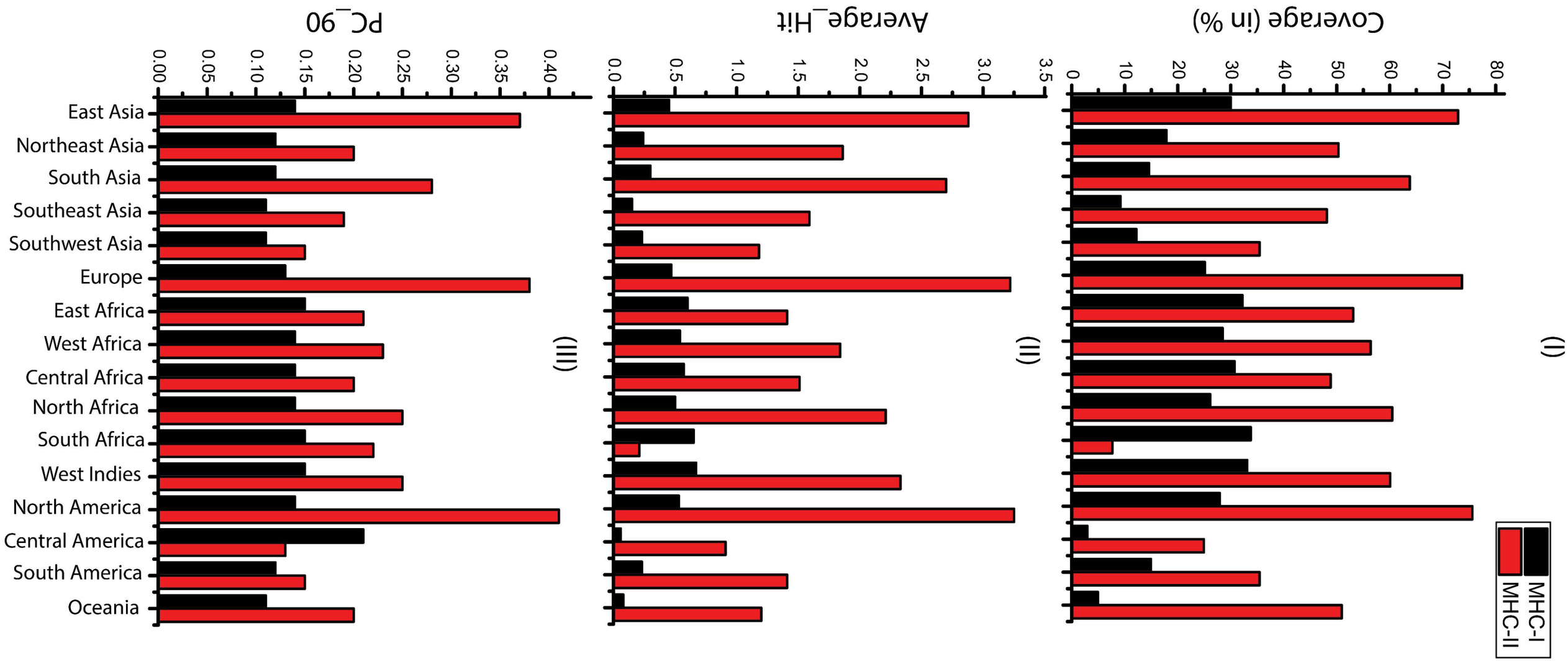
FVSNGTHWF *IRASANLAA* *FGAISSVLN*

VKQLSSNFG *FAMQMAYRF* *FGAGAALQ*



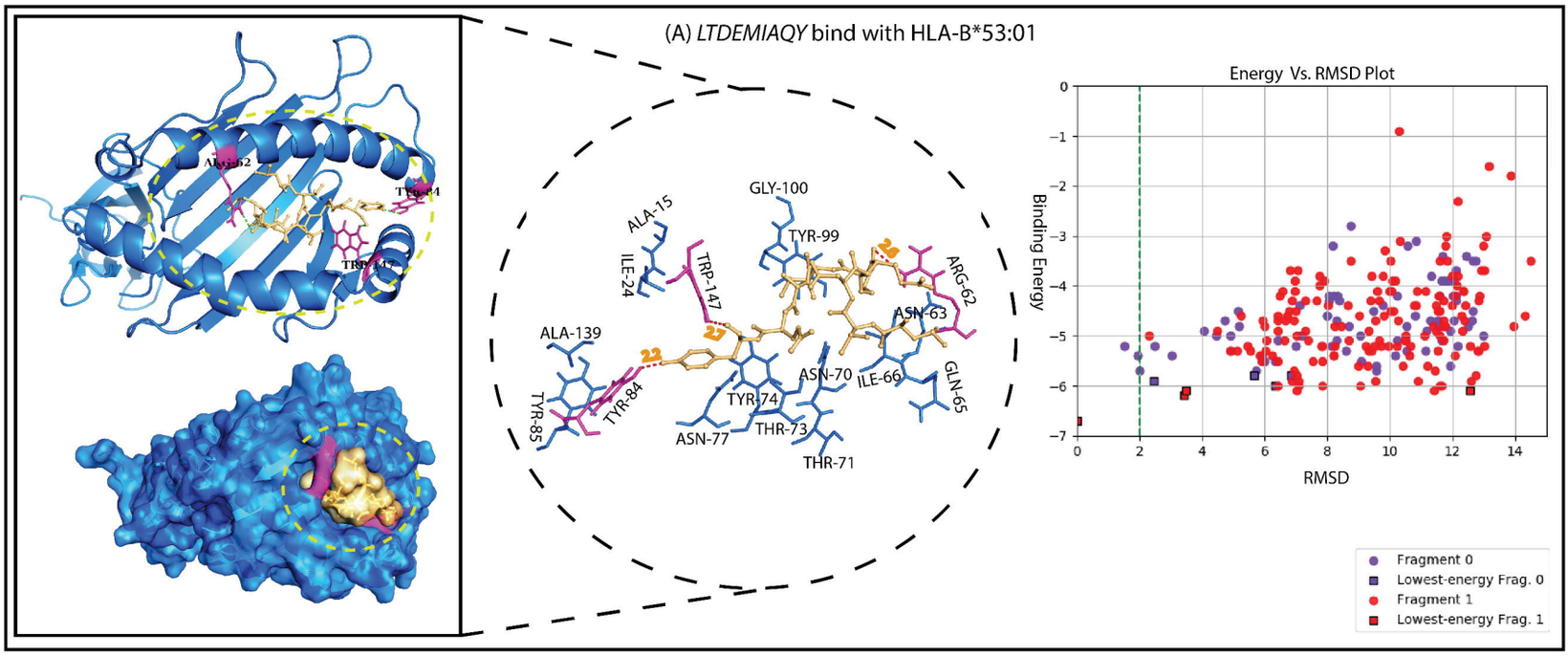


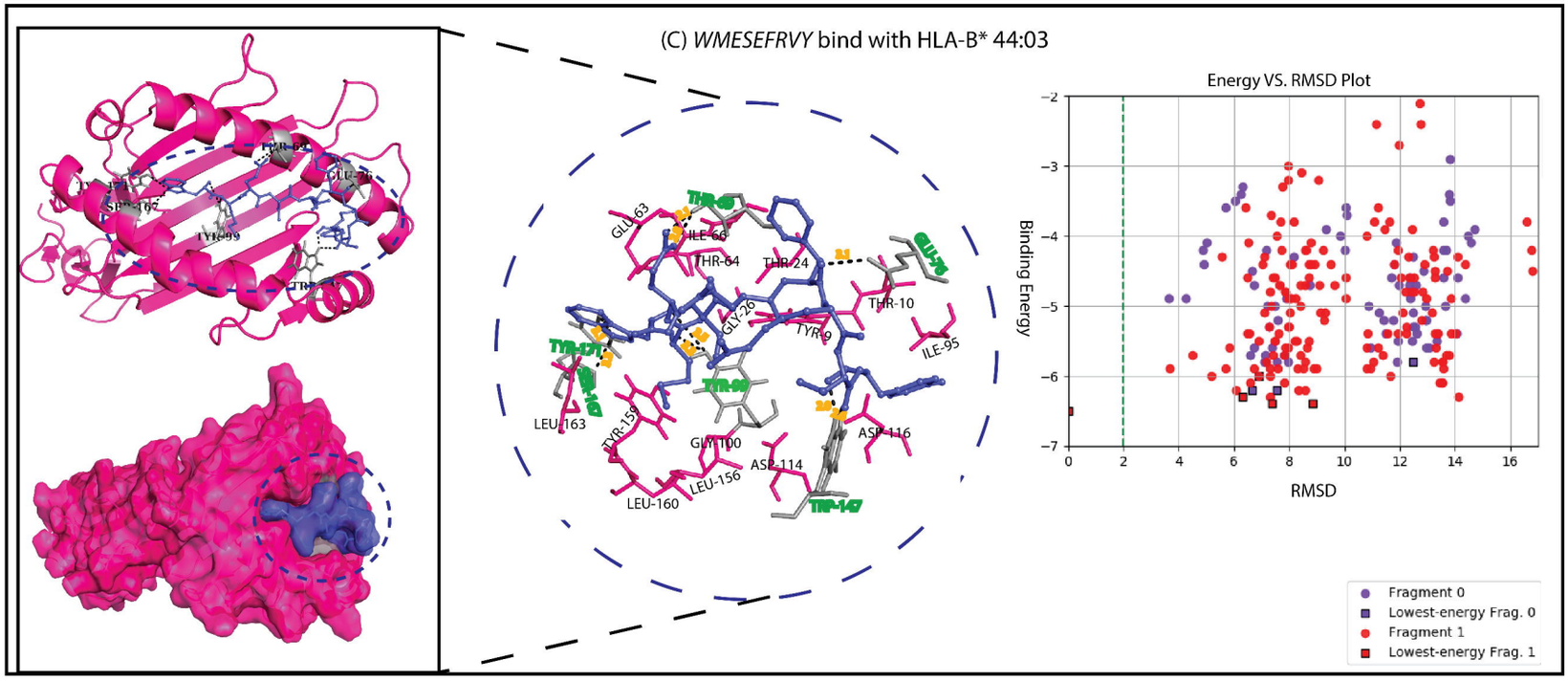
Parker Hydrophilicity prediction
Chau & Fasman beta turn prediction
Bepipred linear epitope prediction
Emini Surface accessibility prediction
Karplus & Schulz Flexibility prediction
Kolaskar & Tongaonkar Antigenicity prediction

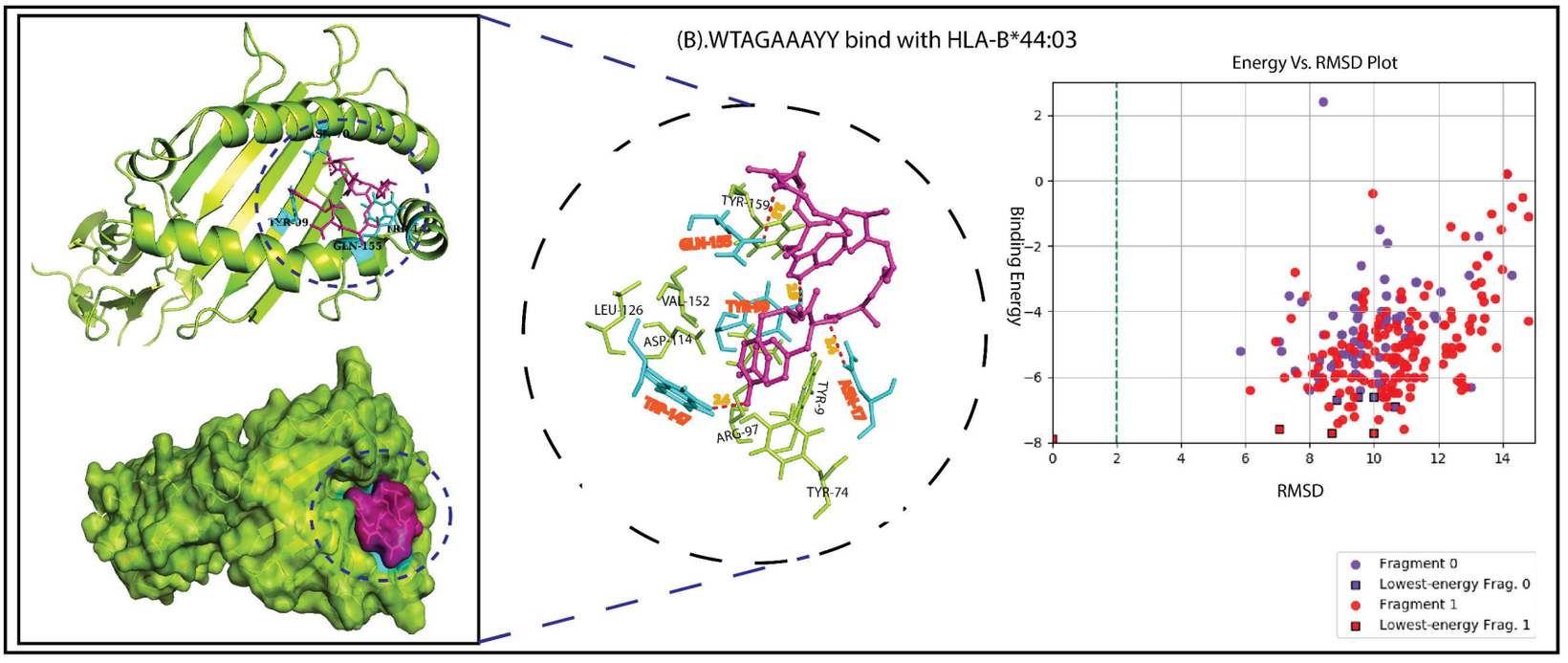


No:te

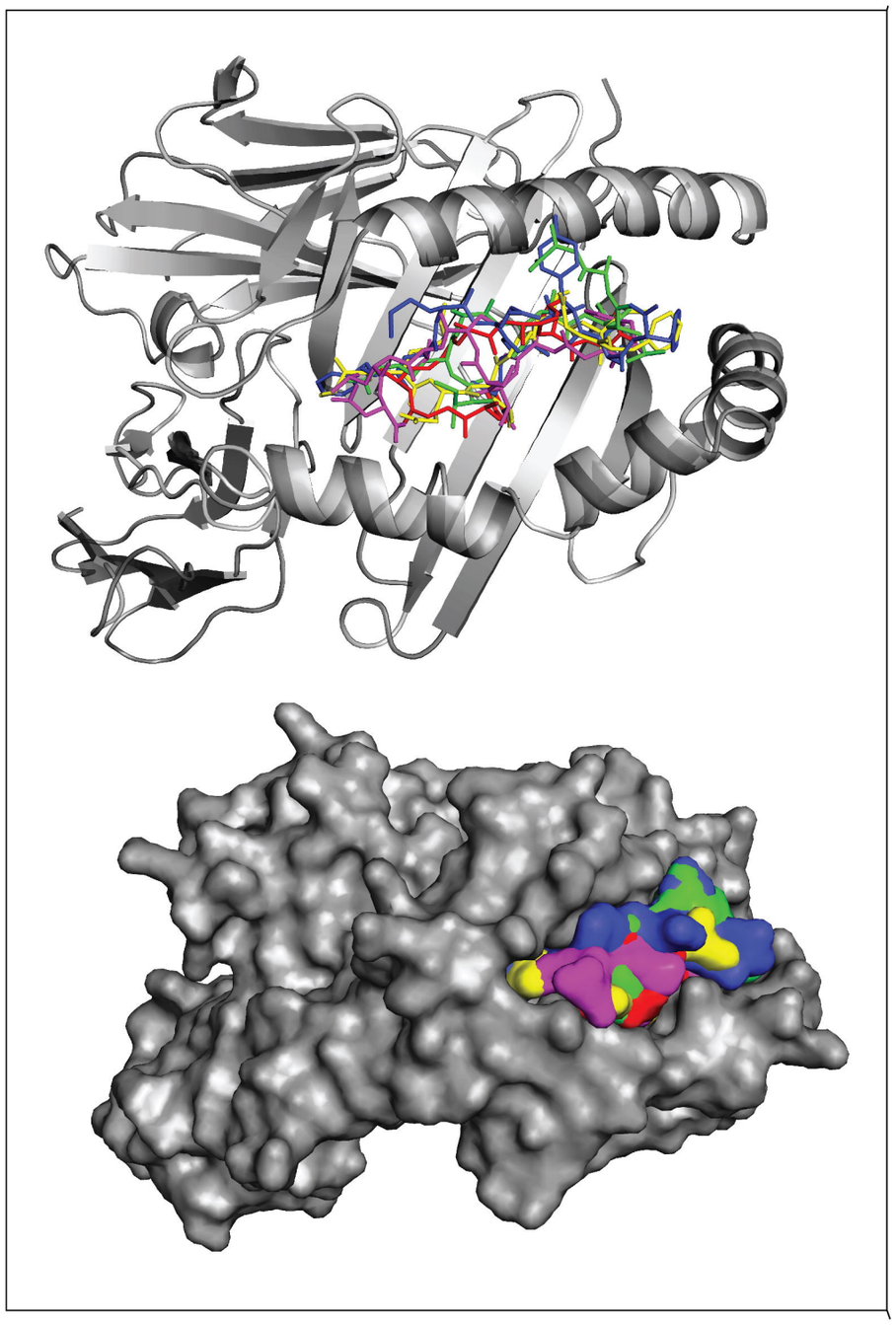
Coverage = Projected Population Coverage
Avergae_Hit= Average Number of Epitope hits / HLA Combinations Recognized by the Population
PC_90 = Minimum Number of Epitope hits / HLA Combinations Recognized by 90% of the Population

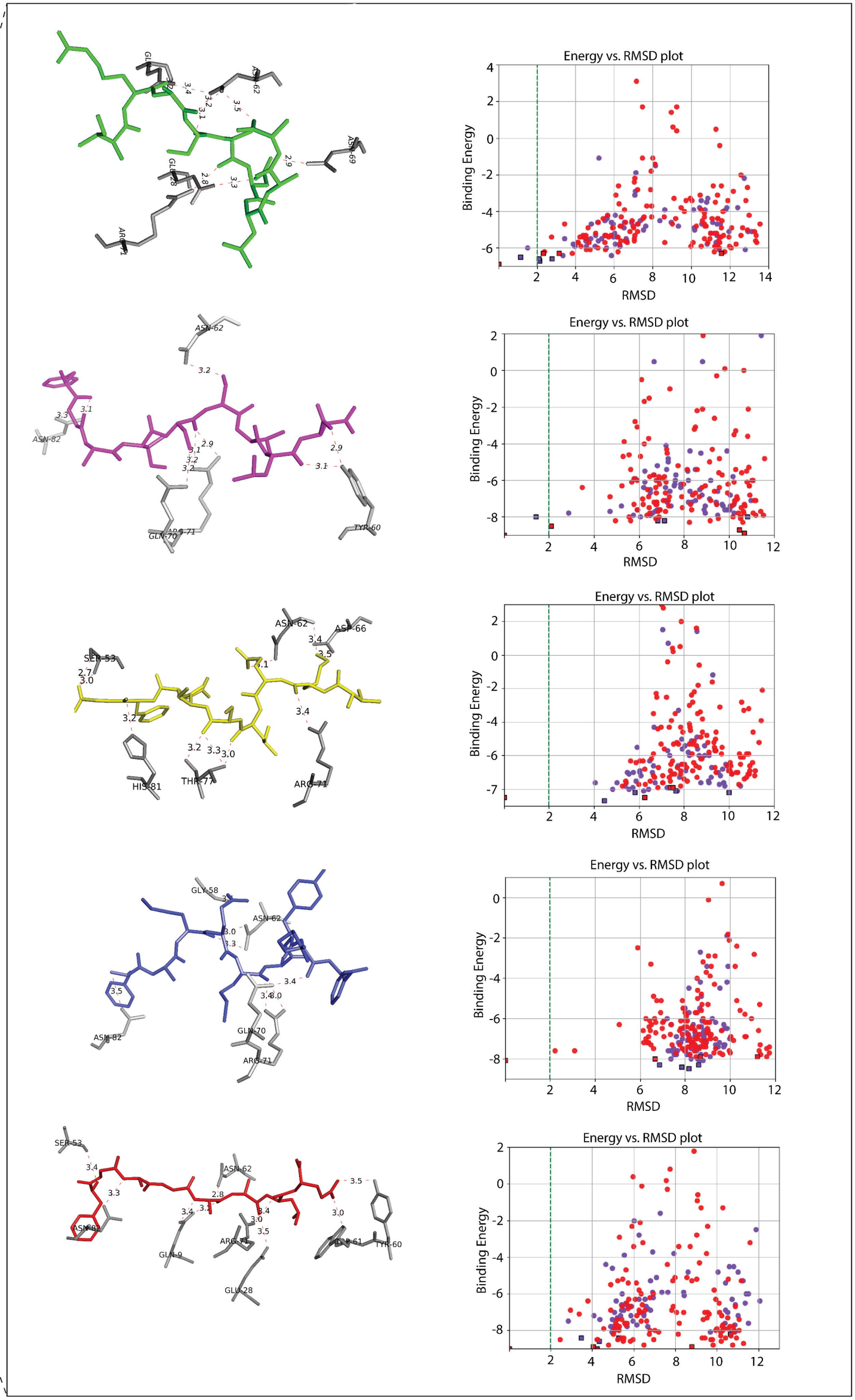

(A) LTDEMIAQY bind with HLA-B*53:01
GLY-100
ALA-15
TYR-99
ILE-24
TRP-147
ARG-62
ASN-63
ALA-139
ASN-70
ILE-66
GLN-65
TYR-85
TYR-84
TYR-74
THR-73
ASN-77
THR-71
Energy Vs. RMSD Plot
Binding Energy
RMSD
Fragment 0
Lowest-energy Frag. 0
Fragment 1
Lowest-energy Frag. 1


(C) WMESEFRVY bind with HLA-B* 44:03
GLU-63
ILE-66
THR-64
THR-24
THR-10
GLU-26
TYR-9
ILE-95
LEU-163
TYR-159
GLY-100
ASP-114
LEU-156
LEU-160
ASP-116
Energy VS. RMSD Plot
Binding Energy
RMSD
Fragment 0
Lowest-energy Frag. 0
Fragment 1
Lowest-energy Frag. 1


(B).WTAGAAAYY bind with HLA-B*44:03
TYR-99
GLN-155
TYR-159
LEU-126
VAL-152
ASP-114
TYR-9
ARG-97
TYR-74
Energy Vs. RMSD Plot
Binding Energy
RMSD
Fragment 0
Lowest-energy Frag. 0
Fragment 1
Lowest-energy Frag. 1

MHC class-II peptides docked with HLA-DRB1*01:01





Fragment 0
Lowest-energy Frag. 0
Fragment 1
Lowest-energy Frag. 1
IRASANLAA
FGAISSVLN
VKQLSSNFG
FAMQMAYRF
FGAGAALQ