Immunoinformatics approaches to design: A novel multi-epitope vaccine candidate against SARS-CoV-2 and it's in silico expression

Ravi Deval, Ayushi Saxena, Zeba Mueed, Dibyabhaba Pradhan, Pankaj Kumar Rai

Submitted to: JMIR Bioinformatics and Biotechnology
on: October 19, 2021

Table of Contents

Immunoinformatics approaches to design: A novel multi-epitope vaccine candidate against SARS-CoV-2 and it's in silico expression

Ravi Deval[1] BSc, MSc, PhD; Ayushi Saxena[2] BSc, MSc; Zeba Mueed[2] BSc, MSc, PhD; Dibyabhaba Pradhan[3] BSc, MSc, PhD; Pankaj Kumar Rai[2] BSc, MSc, PhD

[1]Rohilkhand Laboratory and Research Centre Bareilly IN
[2]Department of Biotechnology Invertis University Bareilly IN
[3]ICMR Computational Genomics Centre Informatics, Systems & Research Management (ISRM) Division New Delhi IN

Corresponding Author:
Ravi Deval BSc, MSc, PhD
Rohilkhand Laboratory and Research Centre
B-4 Parsakhera
Bareilly
IN

Abstract

Background: SARS-CoV-2, belonging to the Coronaviridae family, is a novel RNA virus, known for causing fatal disease in humans called COVID-19. Researchers all around the world are keen on developing a precise treatment or vaccine against this deadly disease.

Objective: The main objective of this paper is to design a novel multi-epitope vaccine candidate against SARS-CoV-2 using immunoinformatics tools.

Methods: A consensus sequence was generated from various genomes of SARS-Cov-2 available from various countries of the outbreak at the ViPR database using JalView software. T cell and B cell epitopes were predicted by restricting them to certain HLA alleles using various servers (nHLApred, NetMHCIIpan v.3.1, ABCpred) and were validated using IEDB tools. Using these epitopes and adjuvant, a multi-epitope vaccine was constructed in-silicoand was later subjected to allergenicity, antigenicity and physicochemical properties profiling along with identification of conformational B-cell epitopes. The designed vaccine was evaluated via codon optimization by the Jcat server and finally, it's in-silicoexpression was done in pET-28a(+) vector using snap-gene software.

Results: A total of 18 epitopes (both T and B cell) were predicted that constituted vaccine construct along with adjuvant and end tag. Vaccine construct was validated and its best structure model was successfully docked with human Toll-like receptors. In-silico expression of the designed vector was also seen in pET-28a(+) plasmid.

Conclusions: The designed novel vaccine candidate has been validated in-silico to elicit robust immune responses hence; it can be used as a potential model for further development of multi-epitope vaccines in the laboratory.

(JMIR Preprints 19/10/2021:34349)
DOI: https://doi.org/10.2196/preprints.34349

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Immunoinformatics approaches to design: A novel multi-epitope vaccine candidate against SARS-CoV-2 and it's *in silico* expression

Short Title: *In silico* Vaccine against SARS-CoV-2 and it's expression.

Abstract

Backgroung: SARS-CoV-2, belonging to the Coronaviridae family, is a novel RNA virus, known for causing fatal disease in humans called COVID-19. Researchers all around the world are keen on developing a precise treatment or vaccine against this deadly disease.

Objective: The main objective of this paper is to design a novel multi-epitope vaccine candidate against SARS-CoV-2 using immunoinformatics tools.

Methodology: A consensus sequence was generated from various genomes of SARS-Cov-2 available from various countries of the outbreak at the ViPR database using JalView software. T cell and B cell epitopes were predicted by restricting them to certain HLA alleles using various servers (nHLApred, NetMHCIIpan v.3.1, ABCpred) and were validated using IEDB tools. Using these epitopes and adjuvant, a multi-epitope vaccine was constructed *in-silico*and was later subjected to allergenicity, antigenicity and physicochemical properties profiling along with identification of conformational B-cell epitopes. The designed vaccine was evaluated via codon optimization by the Jcat server and finally, it's *in-silico*expression was done in pET-28a(+) vector using snap-gene software.

Result: A total of 18 epitopes (both T and B cell) were predicted that constituted vaccine construct along with adjuvant and end tag. Vaccine construct was validated and its best structure model was successfully docked with human Toll-like receptors. In-silico expression of the designed vector was also seen in pET-28a(+) plasmid.

Prospects: The designed novel vaccine candidate has been validated *in-silico* to elicit robust immune responses hence; it can be used as a potential model for further development of multi-epitope vaccines in the laboratory.

Keywords: Docking, *In-silico*, Multi-epitope vaccine, SARS-CoV-2, Toll-like receptors

1. Introduction

Coronaviruses (CoVs) belonging to a family of Coronaviridae (subfamily Coronavirinae) are a positive-sense virus with a single-stranded RNA genome (1). These CoVs have a zoonotic origin, however, some of them have been seen to cross this animal barrier and infect humans. Recently, a coronavirus variant, designated as SARS-CoV-2 (Severe Acute Respiratory Syndrome-Coronavirus-2) is reported to be thecausative agent of COVID 19 (2). This causative agent was originally named as 2019-nCoV, and was later changed to SARS-CoV-2 by the Coronavirus Study Group (CSG) of International Committee on Taxonomy of Viruses (ICTV) based on similarity with SARS-CoV (3). This contagious disease COVID-19 is particularly characterized by fever, cough, difficulty in breathing, and may be followed by respiratory tract infections, diarrhea, headache, hemoptysis, etc. (4,5). The origin of this virus has been mapped to the Wuhan province of China from where it spread to other parts of the world (6,7). The pandemic of COVID 19 around the world compelledthe WHO to declareit as the sixth public health emergency (2). Samples from patients have revealed almost 79.6% sequence similarity of SARS-CoV-2 with SARS-CoV and 96% with bat coronavirus (8).

The global health concern has put pressure on the scientific community to develop an effective vaccine. However, in absence of any effective vaccine and precise treatment, all efforts are being taken as curative measures such asthe use of antiviral drugs, protease inhibitors, phosphodiesterase inhibitors, convalescent plasma transfusions, natural herbs, zinc supplements, etc. (9-14). But these are just aimed at providing symptomatic relief. Therefore, the only way to preventthe spread of

COVID 19 is the development of an effective vaccine.Vaccination will help to induce both humoral as well as cell-mediated immunity and thus safeguards the individual from getting infectedby this virus in the future. A little information is available about the pathogenesis of the virus; therefore, an immunoinformatics-based approach to investigate the immunogenic epitopes and vaccine design using data from proteins sequencing of the SARS-CoV 2 is required. The concept of a multi-epitope vaccine is to efficiently identify and assemble B and T-cell epitopes that are more capable of stimulating the immune system to induce more potent and effective immune responses from both arms of Immune system. Peptides and epitopes have shown to be desirable candidates for vaccine development due to their relatively easy production, chemical stability, and lack of infectious potential (7). This study attempts to use immunoinformatic tools to identify specific immunodominant epitopes for generating an effective vaccine against SARS-CoV-2.

2. Methodology

2.1 Retrieval of Sequence

Polyprotein sequences of 1,389 SARS-CoV-2 genomes were retrieved from virus pathogen database and analysis resource (ViPR) (15). To generate a consensus sequence, Jalview (16), a multiple sequence alignment editor was used. This consensus sequence was then used to find T and B cell epitopes. The process flow diagram of the methodology is depicted in Figure 1.

2.2 Prediction of T-cell epitopes

For lymphocyte-mediated cellular immunity, antigenic peptides must bind to major histocompatibility complexes. In this study we used HLA-A*24:02, HLA-A*24, HLA-B*51:01, HLA-B*54:01 and HLA-C*06:02 alleles for binding of cytotoxic T cell epitopes and HLA-DPA1*02:01/HLA-DPB1*05:01, HLA-DQA1*05:01/HLA-DQB1*03:01 and HLA-DRB4*01:01 alleles for binding of T-helper cell epitopes as suggested by (17–20).

2.2.1 Cytotoxic T Lymphcytes (CTL) epitope prediction

The nHLAPred is a peptide prediction server for MHC class I binding based on the neural network approach. ComPred, a cytotoxic T-cell epitope prediction comprehensive method of nHLAPred, was used to predict MHC binding peptides by using proteasomalmatrices with a cutoff score of 0.5 (21). The results from nHLAPred were validated using IEDB MHC-I binding predictions (22). Percentile rank threshold of 50 was applied for consensus method and peptides having percentile rank above 50 were eliminated.

2.2.2 Helper T Lymphocytes (HTL) epitope prediction

The consensus sequence was screened for binding propensity to MHC class II alleles using NetMHCIIpan 3.1 server (23). This server quantitatively predicts binding affinity and identifies the peptide-binding core register (24). Predicted epitopes were then validated by MHC-II binding prediction of IEDB analysis resource using consensus 2.22 method (25).

2.3 Prediction of B cell epitopes

B cell epitopes are responsible for long term immunity. They secrete antibodies which on binding to antigen causes the B lymphocyte to divide rapidly into effector and memory B cell. We employed an online server ABCpred (26) to predict B cell epitopes. This server is based on a recurrent neural network (RNN) using fixed-length patterns (27). A threshold of 0.51 was applied and window length was set to 20.

2.4 Construction of multiepitope vaccine sequence

After the selection of various CTL, T-helper and B cell epitopes, they were linked together using

appropriate linkers such as **AAY, GPGPG** and **KK** to facilitate the immune processing and epitope presentation (28–31)**.** Adjuvant enhances the immunogenicity of antigens, hence an adjuvant β-defensin 1 was linked at N-terminal by **EAAAK** inker (30). An end tag of 6x-His was linked at the C-terminal end of the vaccine sequence for the purpose of purification and identification (31).

2.5 Allergenicity, Immunogenicity and Physiochemical properties evaluation

We used two servers, AllerTOP v. 2.0 (32) and AllergenFP v.1.0 (33) to test the allergenicity of our vaccine construct as it is important to check if our designed vaccine elicits an allergic reaction inside the body. VaxiJen v.2.0 is an alignment-independent server for prediction of tumor, bacterial and viral origin protective antigens with prediction accuracy of 70-89% (34), hence we employed VaxiJen v.2.0 (35) to check the antigenicity of our vaccine construct at 0.4 threshold. Web server ProtParam (36) was used to assess the physicochemical properties of the construct. Grand average of hydrophobicity (GRAVY), aliphatic index, molecular weight, estimated half-life, etc. are the parameters computed by ProtParam (37).

2.6 Secondary and Tertiary structure prediction and refinement

The secondary structure of our vaccine construct was computed via PSIPRED (38) and GOR (39). Both the servers calculate the number of α-helix, coils, etc. To predict the tertiary structure of the vaccine we used the SCRATCH Protein Predictor suite (40). 3Dpro server of this suite combines the use of predicted structural features, a fragment library and energy terms derived from the PDB statistics to predict the tertiary structure of a protein (41).

The output structure from 3Dpro was refined through the GalaxyRefine web server (42). This server applies template-based modeling to predict structure and *ab initio* modeling to refine loop or terminus regions (43). To refine the query 3D structure, it employs the CASP10 technique (30). The process involves initial reformation of side chains followed by the execution of side-chain repacking and overall structural relaxation by molecular dynamics simulations (31).

2.7 Validation of 3D structure

For validation of refined structure we employed three different servers with various attributes; ProSA-web (44), PROCHECK (45) and ERRAT (46). The basis of ProSA diagnostic tool is statistical analysis and it calculates Z-score which indicates overall model quality. Positive z-score indicates the erroneous nature of the model (47). PROCHECK server was used to generate the Ramachandran plot of the tertiary structure of the designed vaccine. ERRAT server is used to differentiate between correctly and incorrectly determined regions based on characteristic atomic interactions (48).

2.8 Defining discontinuous B-cell epitopes of designed vaccine

To eliminate the infection an antigen epitopes need to interact with antibodies. The most prevalent epitopes recognized by antibodies are conformational epitopes (7). Hence we predicted Conformational B-cell epitopes in our designed vaccine using 3D structure generated by GalaxyRefine via the Ellipro server (49). This server is an attribute of IEDB web tool. The threshold forthe minimum score was set 0.5 with a maximum distance of 6 angstroms.

2.9 Molecular docking between vaccine construct and TLR receptors

Human Toll-like receptors 3 and 4 were utilized for molecular docking to predict interaction patterns and binding affinity of vaccine construct as they have been found prominent inSARS-CoV and MERS-CoV studies TLR 3 has been found to recognize RNA viruses and induce immune response by recruiting TRIF adaptor protein which inturn activates transcription factors such as INF3 and NK-kB.Poly(I:C) andlipopolysaccharide (LPS), agonists of TLR3 and TLR4 respectively have been

found protective against SARS-CoV in mice (50). For this evaluation, PatchDock was used (51). This molecular docking algorithm is geometry-based and screens outtransformations havinga goodyield of molecular shape complementarity. To discard redundant solutions, and RMSD (root mean square deviation) clustering is applied tocandidate solutions (52). For refinement of docking results from PatchDock, another server FireDock (53) was employed. FireDock is a method for flexible refinement and scoring of protein-protein docking. The basis of refined complexes includes partial electrostatics, atomic contact energy, etc. (54). These results can be viewed using the RasMol program which provides molecular graphic visualizations of PDB files (55).

2.10 Validation

2.10.1 Validation by codon optimization

Java Codon Adaptation Tool (Jcat) was used to test high-level expression of a vaccine in K12 strain of *E.coli* for codon optimization with the addition of Xhol and Ndel restriction sites (56). Jcat calculates CAI score which provides the basis for codon optimization along with GC content of improved sequence (57).

2.10.2 *In-silico* vaccine expression

The vaccine was finally cloned into pET-28a (+) using snap-gene software. Restriction cloning tool of snap-gene software was used to insert our multiepitope vaccine sequence at Xhol and Ndel restriction sites on pET-28a (+) vector.

3. Results

3.1 Sequence Retrieval

A consensus sequence was generated by aligning 1,389 full-length polyprotein sequences belonging to an outbreak from around the world. The consensus sequence was identical to “Orf1ab” polyprotein of SARS-CoV-2 as per BLAST results.

3.2 T-cell epitope Prediction

For the elimination of respiratory virus infections, a strong virus-specific T-cell response is necessary. Strategies to enhance T-cell response for robust long term memory should be considered by vaccine interventions (7).

3.2.1 Predicted CTL epitopes

CTL release cytotoxic proteins like granzymes, perforins, etc. to neutralize virus-infected cells and damaged cells (7). This occurs after the successful induction of MHC I. Top 4 antigenic sequences were generated for each allele (HLA-A*24:02, HLA-A*24, HLA-B*51:01, HLA-B*54:01 and HLA-C*06:02) of MHC class I. After validation with IEDB consensus method, a total of 12 CTL epitopes were found.

3.2.2 Predicted HTL epitopes

HTL activates B-cells and CTL for antibody production and killing of infected cells, respectively thus, in adaptive immunity they are the key players (28). The consensus sequence was subjected to NetMHCIIpan v.3.1 (23) for HTL epitope prediction. This online server predicts scores and binders. A strong binder is that which has a score of IC50 ≤ 50nM towards the MHCII allele. Three peptide sequences were selected based on their binding scores to all 3 MHCII alleles (HLA-DPA1*02:01/HLA-DPB1*05:01, HLA-DQA1*05:01/HLA-DQB1*03:01 and HLA-DRB4*01:01). These 3 epitopes were then assessed using IEDB tool with a threshold of 50 percentile rank. Since all three peptides fell under the threshold, they were proposed as HTL

epitopes.

3.3 Predicted B-cell epitopes

B cells produce antibodies and recognize antigens hence,they are the main component during adaptive immune response of humoral immunity. Thus, it was necessary to predict B-cell epitopes as well for vaccine construct (28). Top 3 epitopes were selected based on their scores from ABCpred.

3.4 Multiepitope vaccine construct

The main goal of the study was to design a multiepitope vaccine *in silico* from epitopes of COVID-19 that would elicit immune responses. To enhance the immunogenicity of the vaccine construct, human β-defensin 1 (68 amino acid residues) was used as an adjuvant (28). A total of 18 epitopes were determined in which 12 are CTL epitopes, 3 HTL epitopes and 3 B-cell epitopes these epitopes and adjuvants were linked by certain linkers viz. EAAAK, KK, GPGPG and AAY (28, 30) as shown in Figure 2. An end tag of 6x Histone was added (31).

3.5 Vaccine Features

The capability of vaccines to elicit a robust immune response by binding to B and T cell receptors can be predicted in numeric values as its antigenicity. Vaxien v.2.0 calculated antigenicity of our vaccine constructs to be 0.5004 which falls under the criteria of being antigenic. Likewise, the allergenicity was also calculated to ensure no allergic reaction is induced by our vaccine construct (Table I).

Various physiochemical properties of the designed vaccine were calculated through various servers available online (Table I). The designed vaccine (347 aa long) has a molecular weight of 37812.56 Da. Theoretical pI was found to be 9.32 indicating that the vaccine is significantly basic. With the computed instability index as 25.16, the vaccine construct was marked stable. Aliphatic index of 99.8 suggests it as thermostable and Grand average of hydropathicity (GRAVY) as 0.370 which is less positive hence more hydrophilic.

3.6 Secondary and Tertiary structure prediction

The secondary structure of the vaccine forecasted by PSIPRED (38) and GOR (39) server had 44.38 % α-helix, 17.58% extended strand and 38.04 % random coils as shown in Figure 3.

The tertiary structure of our vaccine was predicted by the 3Dpro function of the SCRATCH suite (40) and was refined using the GalaxyRefine web server (43). This refinement server provided 5 model structures with different RMSD (Root mean square deviation) and Rama favored values (Table II). The model with RMSD 0.604 and Rama favored value 97.4 was selected as shown in Figure 4.

3.7 3D structure validations

Ramachandran plot by PROCHECK(45) confirms 93.8% residues were in most favored regions, 5.3% in additional allowed regions, 0.3% in generously allowed regions and 0.7% in disallowed regions as shown in Figure 5. ProSA-web (44) further assessed Z-score to be -2.15 which is within the vicinity of scores of experimental structures (Figure 6). 86.4662 was the quality factor computed by ERRAT (46).

3.8 Conformational B-cell epitopes

Six discontinuous B-cell epitopes with a threshold of score 0.5 were predicted using Ellipro server (49). Amino acid residues, number of residues and their scores are listed in Table IV. The graphical representation of these epitopes in the vaccine construct is shown in Figure 7.

3.9 Vaccine-Receptor Docking

TLR3 and TLR4 were used for docking purposes as suggested by (50) using PatchDock (51) server and the results were refined by the Firedock server (53). FireDock predicted structures based on global energy, Van der waal interactions and hydrogen bonding (Table IV and V). Out of the top ten structures,the graphical representation of the best structure based on global energy is shown in Figure 8.

3.10 Validation

3.10.1 Codon optimization

The quantification of the vaccine's expression level in host *Escherichia Coli* (K12 strain) was done by Jcat (56). Codon adaptation index was calculated to be 1.0 with GC content 51.58% which indicates better expression. For better expression, GC content between 35% - 70% is required. A 1041 bp improved sequence was also generated which was then used as an insert in pET-28a(+) plasmid.

3.10.2 *In-silico* vaccine expression

The designed vaccine was inserted in pET-28a(+) plasmid by restriction cloning tools of snap-gene software. The overall construct of the vector with inserted vaccine sequence is shown in Figure 9. the clone generated was of 6328bp with an insert at number 159 (158-159 : MCS) to 1199.

4. Discussion

With the ongoing outbreak of COVID-19, the world is not only facing a health crisis but also an economic and social catastrophe.Since there is no definite treatment yet, to ensure safety against SARS-CoV-2, researchers around the world are striving to construct vaccines against this virus. An innocuous and effectual vaccine can reduce the rate of this infection. Since long, vaccination has been used effectively by humans against viral diseases. Epitope based vaccines are more precise and safe in comparison to traditionally used live and killed or attenuated vaccines (7, 31). To design an epitope vaccine, tools of immunoinformatics are convenient as they are time and cost-effective. In the present study, we designed a multiepitope vaccine as both T and B cell epitopes are key components of cellular and humoral immunity. The idea behind its multiepitope design is to design a vaccine that can elicit a maximum immune response. We used a consensus sequence of polyproteins from various genomes of the COVID-19 outbreak that resembles "Orf1ab" polyprotein of SARS-Cov-2. This consensus sequence was then used to predict all three types of epitopes i.e. HTL, CTL and B-cell epitopes.T cell epitopes were restricted to few HLA alleles which were found to be associated with SARS-CoV-2 (17,18). Adjuvant and end tag were added with suitable linkers to design vaccine which was validated by different servers for its antigenicity, allergenicity and other physiochemical properties. The vaccine construct was found to be antigenic, non-allergen and stable. Furthermore, secondary and tertiary structures were predicted and validated. With the tertiary structure, we were able to dock the vaccine construct to human toll-like receptors 3 and 4. Finally, the designed vaccine was validated by codon optimization and in-silico cloning. Although there is still a need to validate it in a wet lab.By Analysing the results, it can be said that this multiepitope vaccine construct could be used as a vaccine candidate against COVID-19 after further processing in wet labs.

Conclusion

The vaccine candidate against SARS-CoV-2 was designed by *in- silico* methods. The consensus sequence of polyproteins from various genomes of the COVID-19 outbreak that resembles "Orf1ab" polyprotein of SARS-Cov-2 was used to predict all three types of epitopes i.e. HTL, CTL and B-cell

epitopes. Adjuvant and end tag were added with suitable linkers to design vaccine which was validated by different servers for its antigenicity, allergenicity and other physiochemical properties. The vaccine construct was found to be antigenic, non-allergen and stable. Furthermore, secondary and tertiary structures were predicted and validated and the designed vaccine was validated by codon optimization and *in-silico* cloning. Although there is still a need to validate it in a wet lab. The designed novel vaccine candidate has been validated *in-silico* to elicit robust immune responses hence; it can be used as a potential model for further development of multi-epitope vaccines in the laboratory.

Conflict of interest

The authors declare that they have no conflict of interest in the publication.

References

1. Dhama Kuldeep a, Khan Sharun b, Tiwari Ruchi c, Dadard Maryam, Singh Yashpal, Malike, *et al*. COVID-19, an emerging coronavirus infection: advances and prospects in designing and developing vaccines, immunotherapeutics, and therapeutics *HUMAN VACCINES & IMMUNOTHERAPEUTICS* https://doi.org/10.1080/21645515.2020.1735227 2020
2. WHO. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). 2020.
3. Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, *et al*. Severe acute respiratory syndrome-related coronavirus: the species and its viruses–a statement of the coronavirus study group. *BioRxiv*. 2020. 2020.02.07.937862. doi:10.1101/ 2020.02.07.937862.
4. Wu YC, Chen CS, Chan YJ. The outbreak of COVID-19: An overview. *J Chin Med Assoc* 2020;83(3):217-220.
5. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, *et al*. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020;395(10223):497-506.
6. Wang, C., Horby, P. W., Hayden, F. G., & Gao, G. F., (2020). A novel coronavirus outbreak of global health concern. *The Lancet*, 395(10223), 470–473. https://doi.org/10.1016/S0140-6736(20)30185-9
7. Maryam Enayatkhani, Mehdi Hasaniazad, Sobhan Faezi, Hamed Guklani, Parivash Davoodian, Nahid Ahmadi *et al*. Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study.*Journal of Biomolecular Structure and Dynamics*, ISSN: 0739-1102 (Print) 1538-0254 (Online) Journal homepage: https://www.tandfonline.com/loi/tbsd20. 2020
8. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, *et al*. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020;579(7798):270-273.
9. Li H, Wang YM, Xu JY, Cao B. Potential antiviral therapeutics for 2019 novel coronavirus. *Chinese journal of tuberculosis and respiratory diseases*, 2020b;43:E002. doi:10.3760/cma.j..1001-0939.2020.0002.
10. Duan K, Liu B, Li C, Zhang H. Effectiveness of convalescent plasma therapy in severe COVID-19 patients. *PNAS*, 2020;117(17):9490-9496.
11. Luo H, Tang QL, Shang YX, Liang SB, Yang M, Robinson N, *et al*. Can Chinese Medicine Be Used for Prevention of Corona Virus Disease 2019 (COVID-19)? A Review of Historical Classics, Research Evidence and Current Prevention Programs. *Chin J Integr Med* 2020;26(4):243-250.
12. Yang Y, Islam MS, Wang J, Li Y, Chen X. Traditional Chinese Medicine in the Treatment of Patients Infected with 2019-New Coronavirus (SARS-CoV-2): A Review and Perspective. *Int J Biol Sci* 2020;16(10):1708-1717.
13. Runfeng L, Yunlong H, Jicheng H, Weiqi P, Qinhai M, Yongxia S, et al. Lianhuaqingwen

exerts anti-viral and anti-inflammatory activity against novel coronavirus (SARS-CoV-2). *Pharmacol Res* 2020;156:104761.
14. Gautret P, Lagier JC, Parola P, Hoang VT, Meddeb L, Mailhe M, et al. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *Int J Antimicrob Agents* 2020:105949.
15. Virus Pathogen Database and Analysis Resource (ViPR) - Genome database with visualization and analysis tools. Available from: *https://www.viprbrc.org/brc/home.spg?decorator=vipr*
16. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009 May 1;25(9):1189–91. Available from: *https://academic.oup.com/bioinformatics/article/25/9/1189/203460*
17. Warren RL, Birol I. HLA predictions from the bronchoalveolar lavage fluid samples of five patients at the early stage of the Wuhan seafood market COVID-19 outbreak. *ArXiv*, 200407108 Q-Bio . 2020 Apr 27; Available from: *http://arxiv.org/abs/2004.07108*
18. Sanchez-Mazas A. HLA studies in the context of coronavirus outbreaks. Swiss Med Wkly . 2020 Apr 16;150(1516). Available from: *https://smw.ch/article/doi/smw.2020.20248*
19. Lin M, Tseng H-K, Trejaut JA, Lee H-L, Loo J-H, Chu C-C, *et al*. Association of HLA class I with severe acute respiratory syndrome coronavirus infection. *BMC Med Genet*. 2003 Sep 12;4(1):9. Available from: *https://doi.org/10.1186/1471-2350-4-9*
20. Tseng HK, Trejaut J, Lee H-L, Loo J-H, Chu C-C, Chen P-J, *et al*. Association of HLA class I with severe acute respiratory syndrome coronavirus infection. BMC Med Genet. 2003 Oct 1;4:9.
21. NHLApred: A neural network method for predicting promiscuous MHC Class I binders. Available from: *http://crdd.osdd.net/raghava/nhlapred/*
22. MHC-I Binding . Available from: *http://tools.iedb.org/mhci/*
23. NetMHCIIpan 3.1 Server. Available from: *http://www.cbs.dtu.dk/services/NetMHCIIpan-3.1/*
24. Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. PubMed-NCBI. Available from: *https://www.ncbi.nlm.nih.gov/pubmed/26416257*
25. MHC-II Binding . Available from: http://tools.iedb.org/mhcii/
26. ABCpred submission page Available from: *https://webs.iiitd.edu.in/raghava/abcpred/ABC_submission.html*
27. Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins*. 2006 Oct 1;65(1):40–8.
28. Kalita P, Padhi AdityaK, Zhang KYJ, Tripathi T. Design of a peptide-based subunit vaccine against novel coronavirus SARS-CoV-2. *Microb Pathog* . 2020 May 4; Available from: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7196559/*
29. Ceraolo C, Giorgi FM. Genomic variance of the 2019-nCoV coronavirus. J Med Virol 2020;92(5):522–8. Available from: *https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.25700*
30. Khan M, Khan S, Ali A, Akbar H, Sayaf AM, Khan A, et al. Immunoinformatics approaches to explore Helicobacter Pylori proteome (Virulence Factors) to design B and T cell multi-epitope subunit vaccine. Sci Rep. 2019 Dec;9(1):13321. Available from: *http://www.nature.com/articles/s41598-019-49354-z*
31. Saha R, Prasad BVLS. In silico approach for designing of a multi-epitope based vaccine against novel Coronavirus (SARS-COV-2). *bioRxiv* 2020 Apr 3;2020.03.31.017459. Available from: https://www.biorxiv.org/content/10.1101/2020.03.31.017459v1
32. Bioinformatics tool for allergenicity prediction. Available from: *https://www.ddg-pharmfac.net/AllerTOP/*

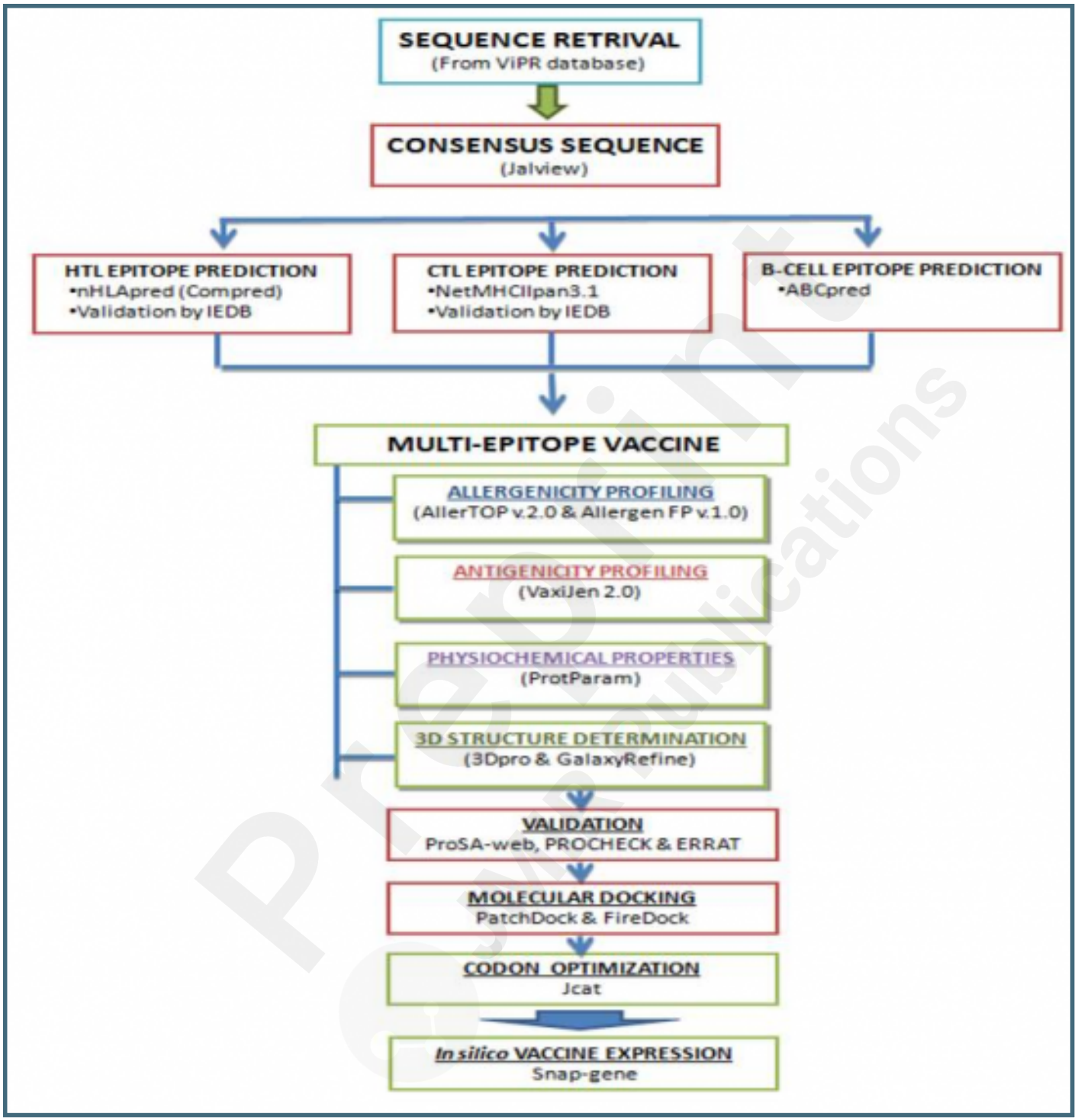
33. Bioinformatics tool for allergenicity prediction. Available from: http://ddg-pharmfac.net/AllergenFP/
34. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*. 2007 Jan 5;8(1):4. Available from: https://doi.org/10.1186/1471-2105-8-4
35. VaxiJen. Available from: http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html
36. ExPASy - ProtParam tool. *Available from:* https://web.expasy.org/protparam/
37. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, *et al*. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, editor. *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press; 2005.p. 571–607. Available from: http://link.springer.com/10.1385/1-59259-890-0:571
38. PSIPRED Workbench. Available from: http://bioinf.cs.ucl.ac.uk/psipred/result/caf171d2-3a1f-11e5-84d8-00163e110593
39. NPS@ : GOR4 secondary structure prediction. Available from: https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html
40. Scratch Protein Predictor. Available from: http://scratch.proteomics.ics.uci.edu/
41. Cheng J, Randall AZ, Sweredoski MJ, Baldi P. SCRATCH: a protein structure and structural feature prediction server. *Nucleic Acids Res*. 2005 Jul 1;33(Web Server issue):W72–6. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1160157/
42. GalaxyWEB. Available from: http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE
43. GalaxyWEB server for protein structure prediction and refinement, *Nucleic Acids Research Oxford Academic*. Available from: https://academic.oup.com/nar/article/40/W1/W294/1078340
44. ProSA-web - Protein Structure Analysis. Available from: https://prosa.services.came.sbg.ac.at/prosa.php
45. PROCHECK - DOE-MBI Structure Lab UCLA. Available from: https://servicesn.mbi.ucla.edu/PROCHECK/
46. ERRAT - DOE-MBI Structure Lab UCLA. Available from: https://servicesn.mbi.ucla.edu/ERRAT/
47. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, *Nucleic Acids Research*, Oxford Academic. Available from: https://academic.oup.com/nar/article/35/suppl_2/W407/2920938
48. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci Publ Protein Soc*. 1993 Sep;2(9):1511–9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2142462/
49. ElliPro: Antibody Epitope Prediction . Available from: http://tools.iedb.org/ellipro/
50. Totura A, Whitmore A, Agnihothram S, Schäfer A, Katze M, Heise M, et al. Toll-Like Receptor 3 Signaling via TRIF Contributes to a Protective Innate Immune Response to Severe Acute Respiratory Syndrome Coronavirus Infection. *mBio*. 2015 May 26;6.
51. PatchDock Server: An Automatic Server for Molecular Docking. Available from: https://bioinfo3d.cs.tau.ac.il/PatchDock/php.php
52. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res*. 2005 Jul 1;33(Web Server issue):W363–7. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1160241/
53. FireDock Server. Available from: http://bioinfo3d.cs.tau.ac.il/FireDock/php.php
54. FireDock: a web server for fast interaction refinement in molecular docking. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2447790/
55. RasMol and OpenRasMol. Available from: http://www.openrasmol.org/
56. JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. Andreas Grote, andreas.grote@tu-bs.de; 2004. Available from: http://www.jcat.de/

57. Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res*. 2005 Jul 1;33(Web Server issue):W526–31. Available from: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1160137/*

Supplementary Files

Figures

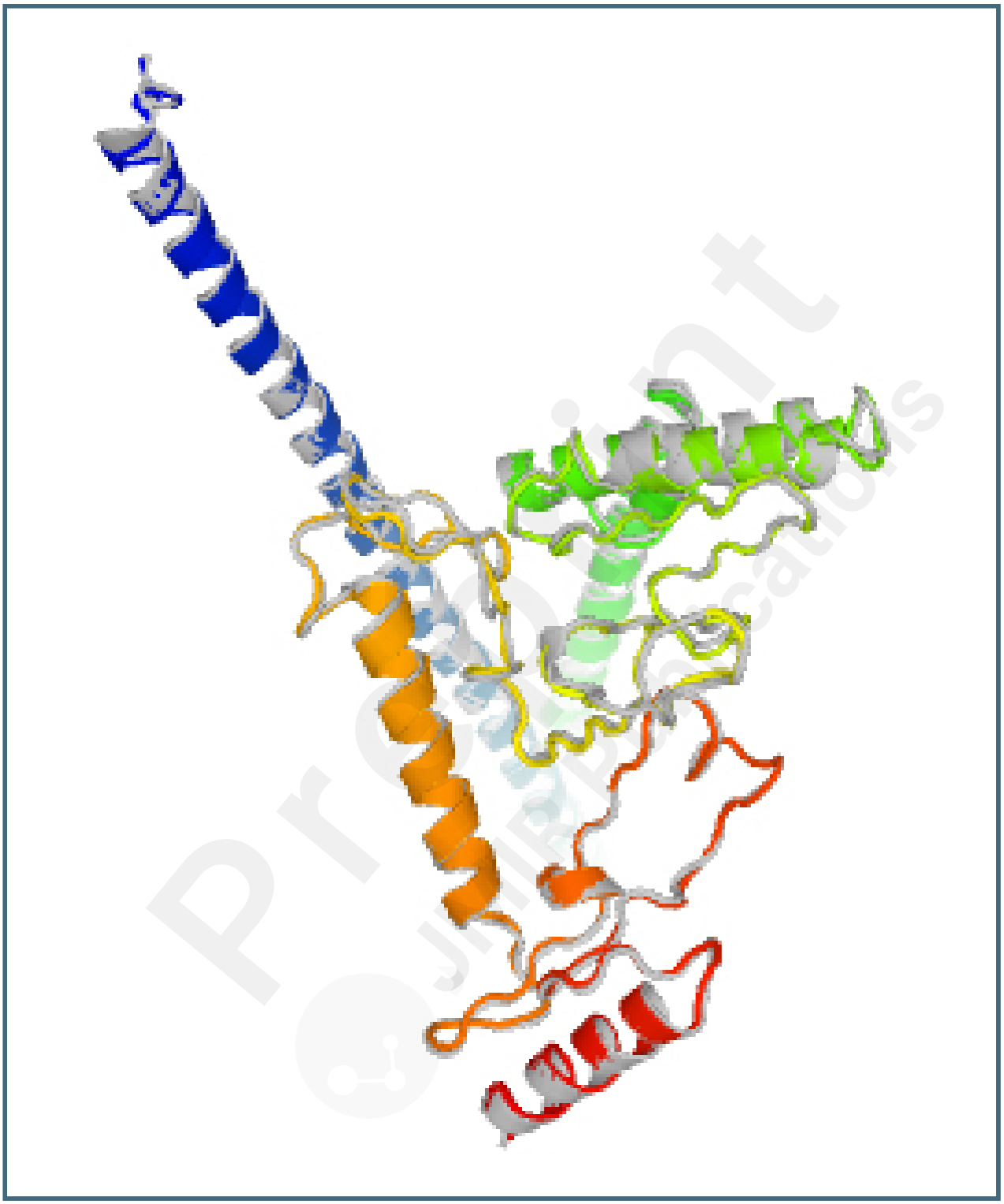
Process flow of multiepitope vaccine design and its expression in silico.



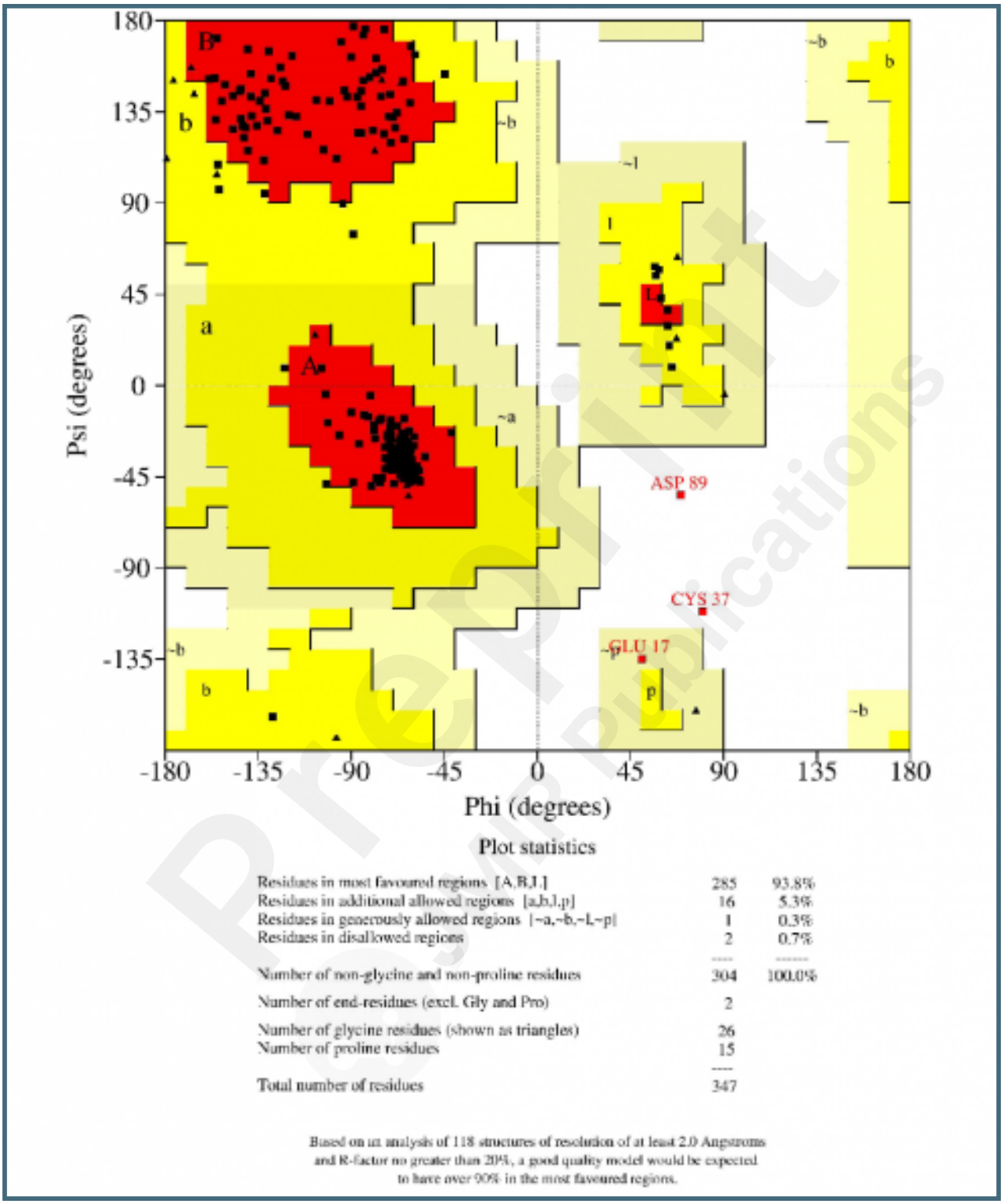
Linear Multiepitope vaccine design with end linked adjuvant.

PSIPRED web tool predicted secondary structure of vaccine construct showing 44.38% ?-helix, 17.58% extended strand and 38.04 % random coils.

3D structure of the final vaccine refined by the GalaxyRefine web server.



Ramachandran plot of vaccine construct showing most favored (93.8%), additional allowed (5.3%), generously allowed (0.3%) and disallowed (0.7%) regions.

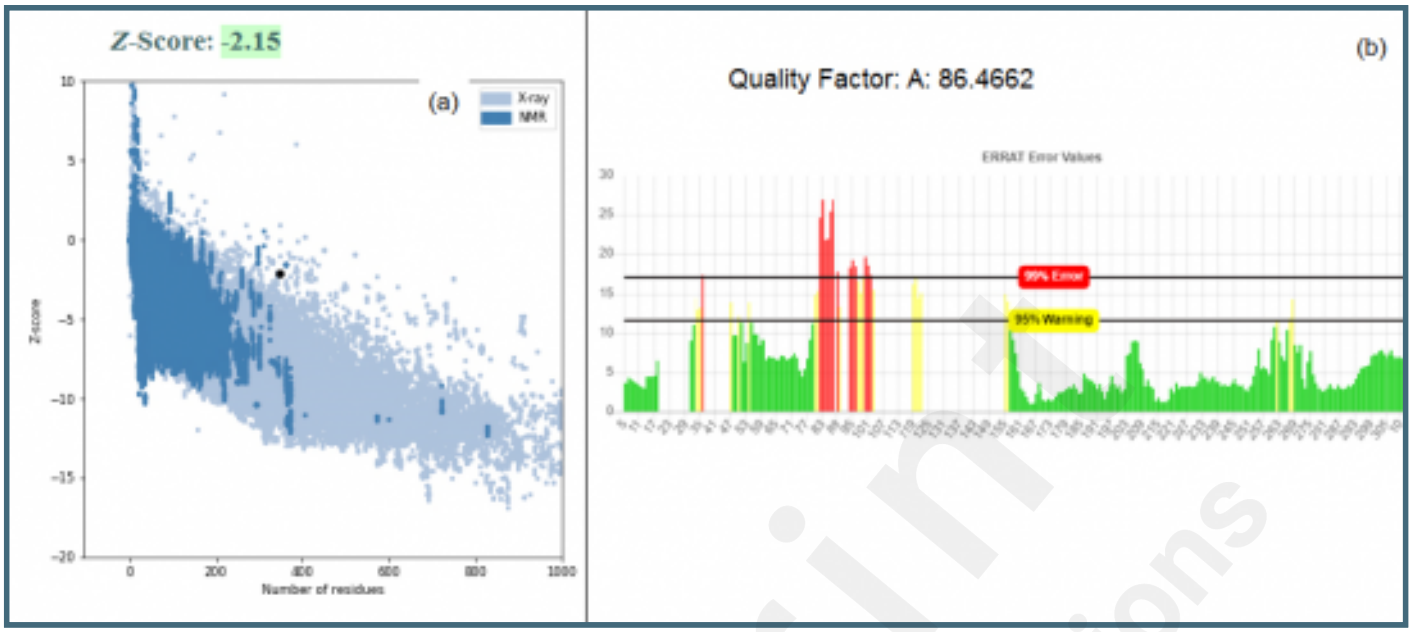


Plot statistics

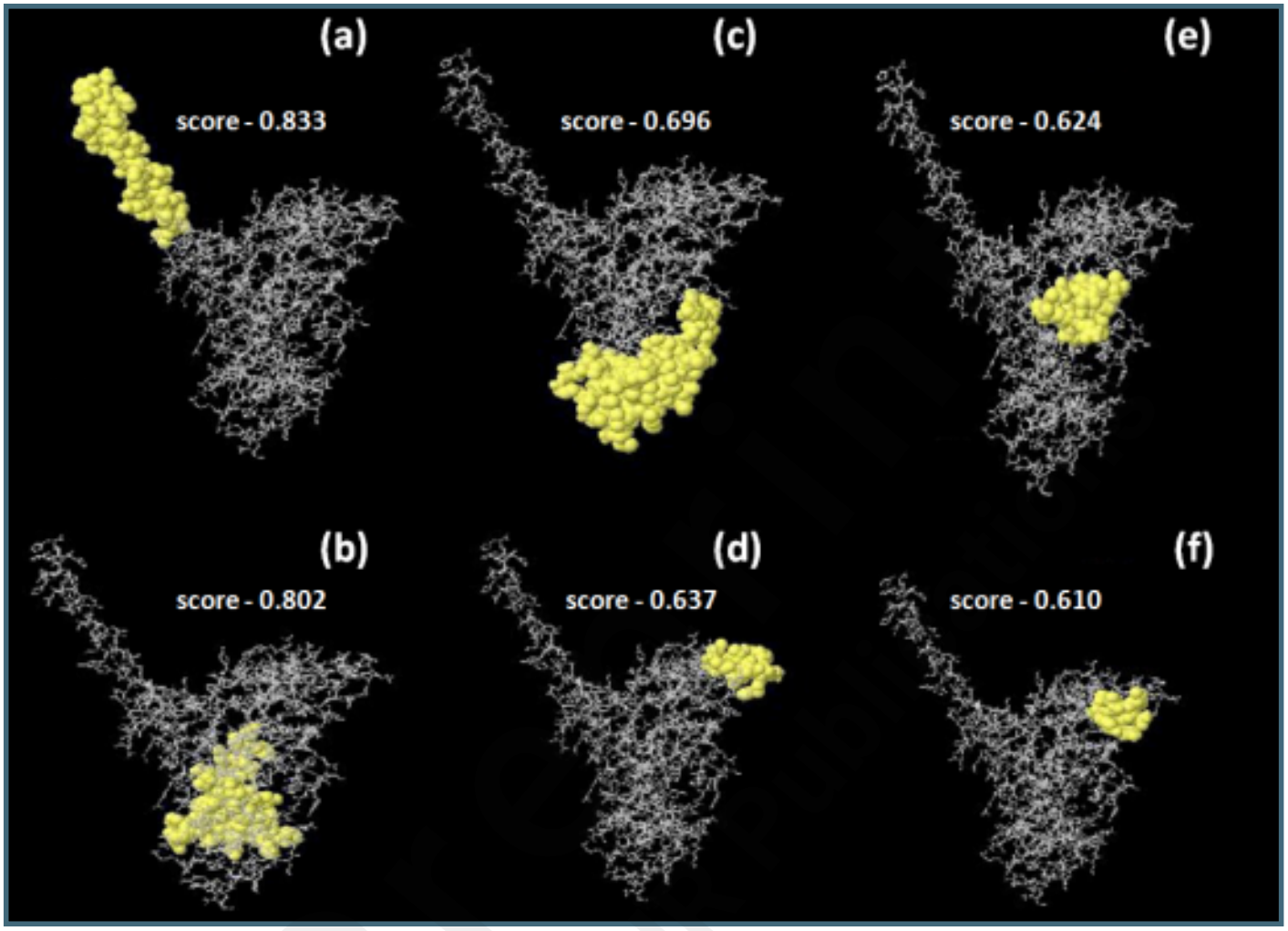
Residues in most favoured regions [A,B,L]	285	93.8%
Residues in additional allowed regions [a,b,l,p]	16	5.3%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.3%
Residues in disallowed regions	2	0.7%
Number of non-glycine and non-proline residues	304	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	26	
Number of proline residues	15	
Total number of residues	347	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

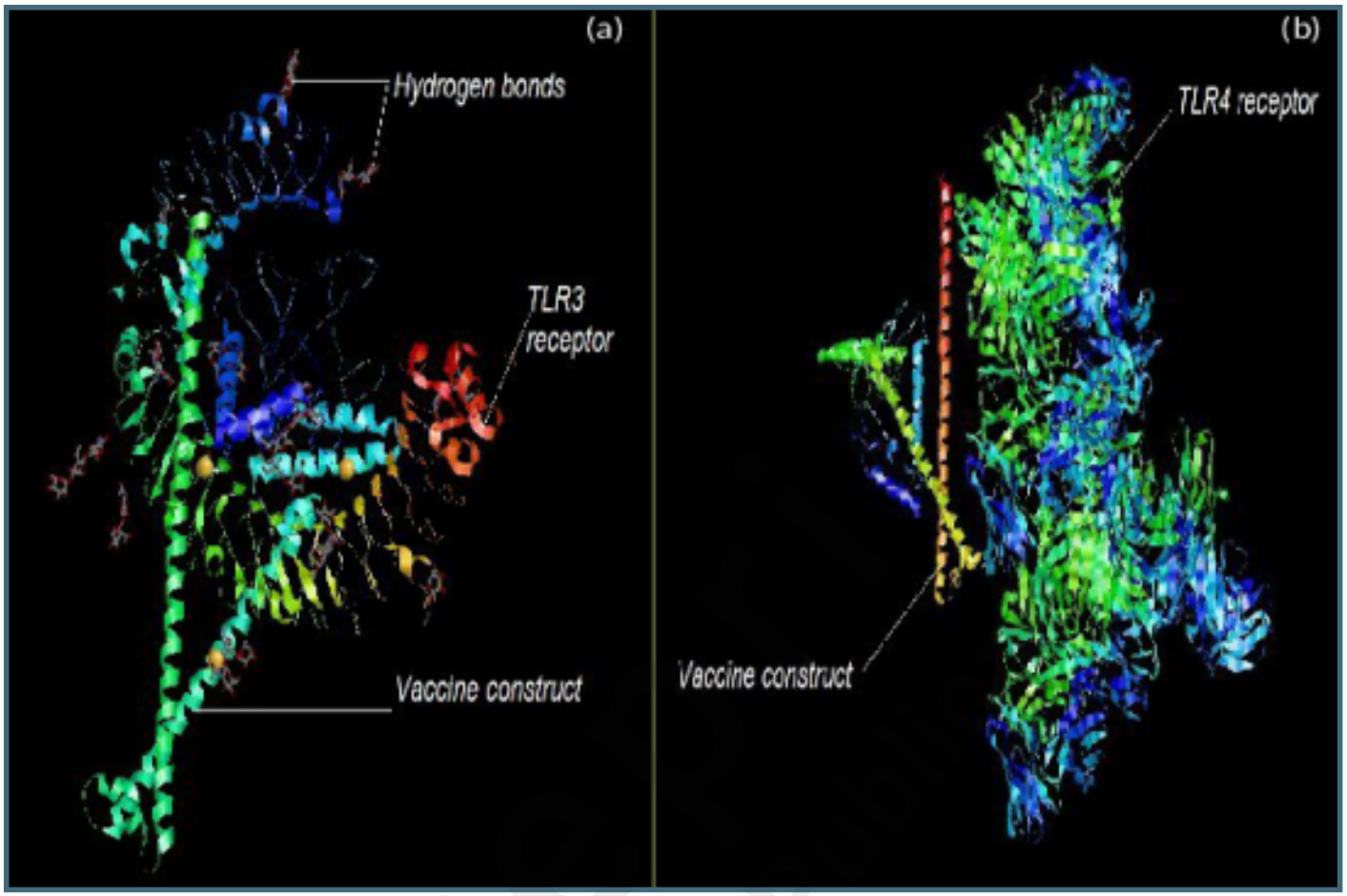
Validation of designed vaccine by (a) Z-score plot and (b) ERRAT plot depicting quality factor.



3D representation of conformational B-cell epitopes from designed vaccine. Epitopes are shown in yellow color and grey sticks represent bulk of polyprotein. The numbers correspond to values in Table III.



Graphical representation of vaccine-receptor docking using RasMol; (a): Vaccine TLR3 docking complex; (b): Vaccine TLR4 docking complex.



Structure of cloned pET-28a (+) vector with inserted vaccine sequence (highlighted in magenta) along with various restriction sites.