Application of in-silico reverse vaccinology for designing multi-epitope vaccine against Coronavirus

Kumar Sharp[1], Dr. Shubhangi Dange[2]*

[1]2nd MBBS undergraduate student, Government Medical College and Hospital, Jalgaon.

[2]Associate professor, Department of Microbiology, Government Medical College and Hospital, Jalgaon

Corresponding author: -

[2]* Dr. Shubhangi Dange,

Associate Professor,

Department of Microbiology,

Government Medical College and Hospital, Jalgaon.

Email: dangeshubhangi@gmail.com

Application of in-silico reverse vaccinology for designing multi-epitope vaccine against Coronavirus

Abstract

In absence of any specific medication or vaccine till now, experimentation has reached new heights. With lockdown imposed in almost every country and huge economic losses the search for a suitable vaccine has still been unsuccessful. In this study we have approached through in-silico method or reverse vaccinology taking advantage of the genome sequence of the novel coronavirus. We created a multi-epitope model vaccine which can elicit both humoral as well as cell-mediated immune response. It is also docked with toll-like receptor 8 TLR-8. The sequence obtained is antigenic, non-allergenic and 86.3% residues are in favourable region of Ramachandran plot. This sequence might have good hope of emerging as the vaccine of the current pandemic if studied more in depth.

Keywords: reverse vaccinology; coronavirus; multi-epitope; vaccine

Introduction

As of writing this article, over 34 lakh people all over the world are infected with SARS-CoV2 with over 2 lakh people dead because of it [1]. In absence of any specific medication or vaccine till now, experimentation has reached new heights. With lockdown imposed in almost every country and huge economic losses the search for a suitable vaccine has still been unsuccessful.

In this study we have approached through in-silico method or reverse vaccinology taking advantage of the genome sequence of the novel coronavirus. It serves its benefit of identifying antigens seen by both conventional as well as discovering any novel antigen [2]. With technological advancement in the field of immunology these studies have become easier and more accurate [3]. We created a multi-epitope model vaccine which can elicit both humoral as well as cell-mediated immune response. It is also docked with toll-like receptor 8 TLR-8. These findings can help find a suitable vaccine candidate against coronavirus.

Methodology

The steps followed for reverse vaccinology can be introduced as the following sequence (figure1) and as summarised by Khan S. et al [4]:

1. Selection of proteins
2. Prediction/selection of CTL and HTL epitopes
3. Calculating allergenicity and antigenicity of vaccine sequence
4. Reviewing physio-chemical properties of vaccine sequence
5. Prediction of secondary and tertiary structure of vaccine
6. B-cell epitope prediction
7. Molecular docking with toll-like receptor.
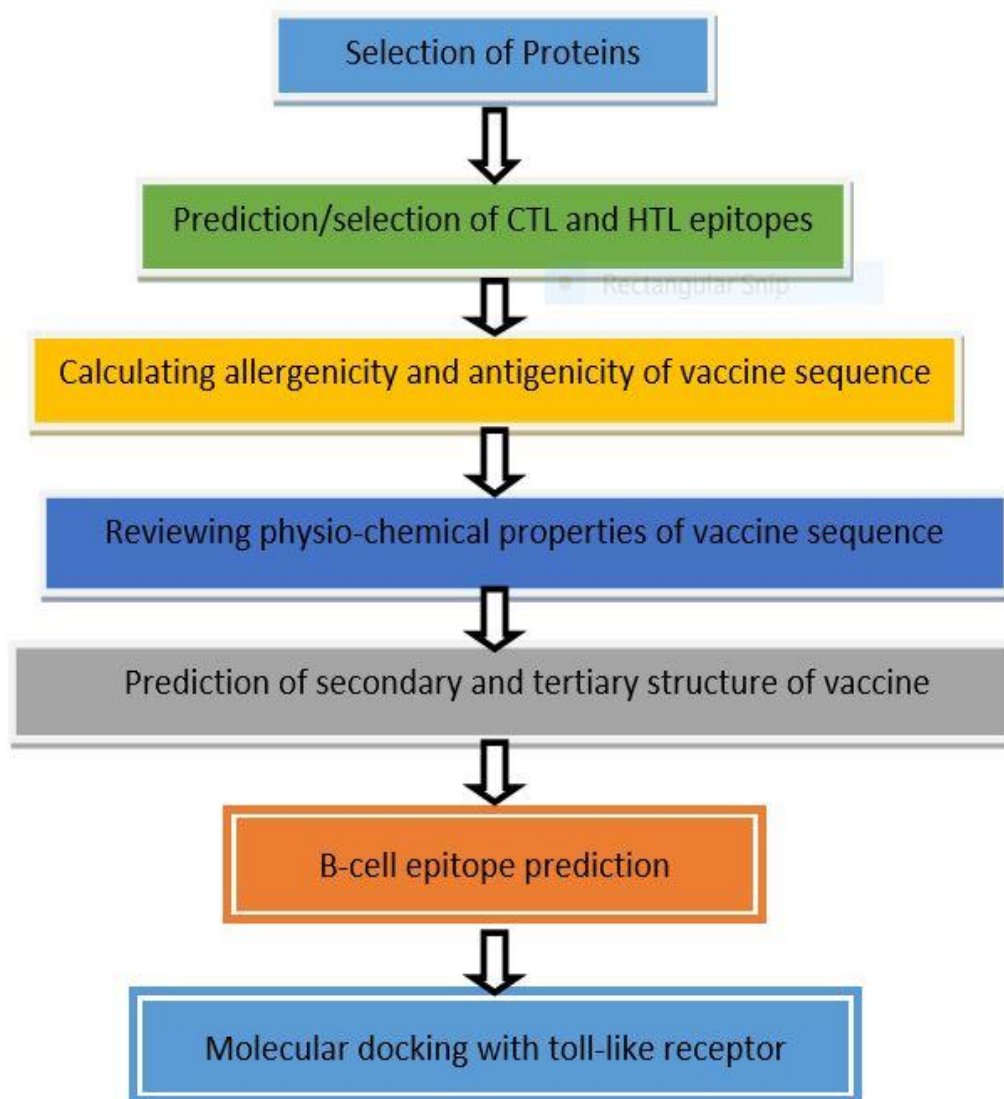
Figure 1: *Methodology summary.*

These steps are further described in detail:

1. *Selection of proteins*

   The complete amino acid sequence for coronavirus genome sequence MT259228 was taken from GenBank [5], NCBI in standard FASTA format. This genome sequence was isolated from SARS-CoV2 Wuhan, China.
   Structural protein sequences like spike/surface glycoprotein, envelope, membrane, nucleocapsid and ORF frames were selected. These structures of the virus are the first to come in contact with human cells and hence facilitate in immunity.

2. *Prediction/selection of CTL and HTL epitopes*

   All proteins sequences were analysed in the same manner as follows.

   For prediction of CTL epitopes, the following sequence was followed:
   a. Vaxijen [6] server was used to determine whether the protein sequence is antigenic or non-antigenic. Only antigenic sequences were chosen.
   b. Net CTL 1.2 [7] server was used to predict CTL epitope sequences and results were sorted as per combined score. Only those sequences were chosen which were predicted to be epitopes(E).
   c. These epitope sequences were again checked individually on Vaxijen server for antigenicity. Only antigenic epitope sequences were selected.
   d. Antigenic CTL epitopes were predicted for toxicity using ToxinPred [8] server. Only non-toxic sequences were selected.
   e. Non-toxic antigenic CTL epitopes were then analysed for immunogenicity by Immunogenicity IEDB database [9]. 2 to 3 higher score sequences were selected.

   For prediction of HTL epitopes, the following sequence was followed:
   f. Protein sequence was analysed by IEDB MHC-2 server [10] and epitope candidates were chosen as per percentile score. The percentile score is inversely proportional to the binding affinity. Only those sequences which had a score less than or equal to 2.0 and classified as strong binders were proceeded with.
   g. The selected epitope sequences were analysed individually by Algpred server [11] for allergenicity. Only non-allergenic sequences were selected since allergenic sequences could induce hypersensitivity reaction in the body.
   h. Non-allergenic HTL epitopes were analysed for antigenicity by Vaxijen server. Only antigenic sequences among them were selected.
   i. Antigenic non-allergenic HTL epitopes were analysed for toxicity using ToxinPred server. Only non-toxic sequences were selected.
   j. Non-toxic Antigenic non-allergenic HTL epitopes were then predicted for interferon IFN inducing property using IFN epitope server [12]. Only positive or "no data available" results were chosen.

All structural protein sequences chosen above were subjected to all steps a-j. If at any step, the determined conditions were not fulfilled, the particular protein was removed from choice.

The selected CTL and HTL epitopes were then joined by AAY and GPGPG linkers [13] and adjuvant was added [14] at the beginning. This finalized the creation of a multi-epitope vaccine sequence.

3. *Calculating allergenicity and antigenicity of vaccine sequence*

- The final sequence formed was predicted for antigenicity using Vaxijen server. If found to be non-antigenic new epitope sequences were added unless it resulted as antigenic.
- Antigenic sequence was then subjected to allergenicity prediction by using Algpred server. Only non-allergenic sequence is considered. If found to be allergenic, new epitope sequences were added unless it resulted as non-allergenic.

4. *Reviewing physio-chemical properties of vaccine sequence*

The physio-chemical properties of the vaccine sequence were analysed using ProtParam server [15]. These properties were molecular weight, theoretical pI, half-life in Escherichia coli, instability index, aliphatic index and hydropathicity.

5. *Prediction of secondary and tertiary structure of vaccine*

The secondary and tertiary structure for the vaccine were constructed by PHYRE2 server [16]. The model was subjected to refinement using Galaxy Refine [17] web-tool. The 3D structure was validated on ProSA-web [18] and RAMPAGE [19]. ProSA-web is used to assign a quality Z-score for the structure. RAMPAGE is used for Ramachandran plot analysis.

6. *B-cell epitope prediction*

B-cell epitopes for the final multi-epitope vaccine sequence were predicted and visualized using Ellipro suite [20].

7. *Molecular docking with toll-like receptor/s.*

The final multi-epitope vaccine was docked with toll-like receptor TLR8 by PatchDock server [21]. PDB id: 3W3G was the protein structure for TLR8 protein. The structure was taken from Protein Data Bank [22]. The results thus obtained were refined using FireDock server [23]. The best ranked-result was chosen as the final docking result.

Result

The final multi-epitope sequence formed after performing step 1 and 2 of methodology comprised of 206 amino acids:

GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRKEAAAKQLTPTWRVYAAYSTDT
GVEHVAAYYIDIGNYTVAAYLSPRWYFYYAAYLLEQWNLVIGPGPGVLSFELLHAPATVCGGPG
PGKRWQLALSKGVHFVCGPGPGPIHFYSKWYIRVGARGPGPGKAYNVTQAFGRRGPEGPGPG
VGLMWLSYFIASFRL

The above sequence distribution is as follows:
GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRK EAAAK
(The above sequence is beta-defensin adjuvant sequence with EAAAK linker)

QLTPTWRVY AAY
(Spike/surface glycoprotein CTL epitope with highest immunogenicity with AAY linker)

STDTGVEHV AAY
(ORF 3A CTL epitope with highest immunogenicity with AAY linker)

YIDIGNYTV AAY
(ORF 8 CTL epitope with highest immunogenicity with AAY linker)

LSPRWYFYY AAY
(Nucleocapsid CTL epitope with highest immunogenicity with AAY linker)

LLEQWNLVI GPGPG
(Membrane protein CTL epitope with highest immunogenicity with GPGPG linker to HTL epitopes)

VLSFELLHAPATVCG GPGPG
(Spike/surface glycoprotein HTL epitope with GPGPG linker)

KRWQLALSKGVHFVC GPGPG
(ORF 3A HTL epitope with GPGPG linker)

PIHFYSKWYIRVGAR GPGPG
(ORF 8 HTL epitope with GPGPG linker)

KAYNVTQAFGRRGPE GPGPG
(Nucleocapsid HTL epitope with GPGPG linker)

VGLMWLSYFIASFRL
(Membrane HTL epitope)

The antigenicity prediction as in step 3 by Vaxijen server predicted it to be a probable antigen with score 0.5433. The allergenicity prediction as in step 3 by Algpred server predicted it to be a non-allergen with score of -0.5025712 (positive predictive value is 18.21% and negative predictive value is 71.24%).

Physio-chemical properties as estimated by step 4 using ProtParam server gave the following results:

Molecular weight:22677.26 Daltons

Theoretical pI=9.67

Estimated half-life in E. coli: >10 hours (in vivo)

Instability index: 36.46(stable)

Aliphatic index: 76.26

Grand average of hydropathicity (GRAVY): -0.129

The secondary structure of the final multi-epitope vaccine sequence was computed using PHYRE 2 server: 32% comprised of alpha-helix, 37% of beta-strand and 19% was disordered (Figure 2).


Figure 2: *Secondary structure of vaccine sequence.*

The tertiary structure obtained from PHYRE2 server was subjected to refinement by Galaxy Refine tool which generated 5 models as follows (Table 1):

| Model	GDT-HA	RMSD	MolProbity	Clash Score	Poor rotamers	Rama favoured
Initial	1.0000	0.000	4.327	143.8	11.9	63.7
Model 1	0.9078	0.536	2.166	9.8	0.6	85.3
Model 2	0.9005	0.555	2.234	10.7	0.0	83.3
Model 3	0.8908	0.588	2.226	10.7	0.0	83.8
Model 4	0.8871	0.586	2.192	9.8	0.6	83.8
Model 5	0.9053	0.563	2.095	7.9	0.6	84.8

Table 1: *Galaxy Refine structure models*

Model 1 was chosen as the best tertiary structure of the sequence for further analysis. It was visualized using UCSF Chimera software [24] (Figure 3).
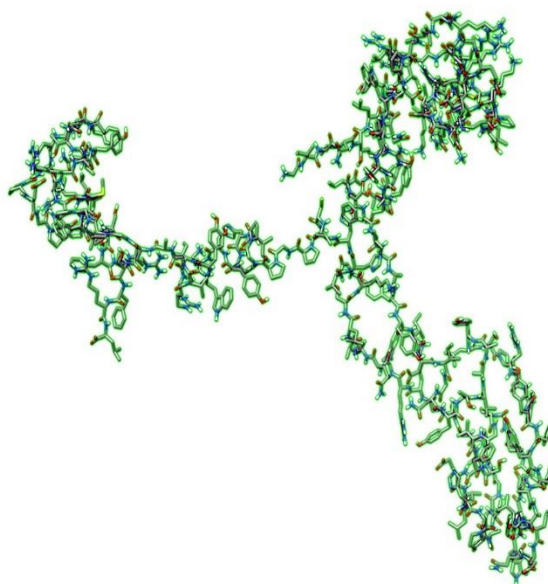
Figure 3: *Model 1 tertiary structure.*

The chosen tertiary structure was subjected to validation by Pro-SA web sever which computed a Z-score of -1.07 (Figure 4).
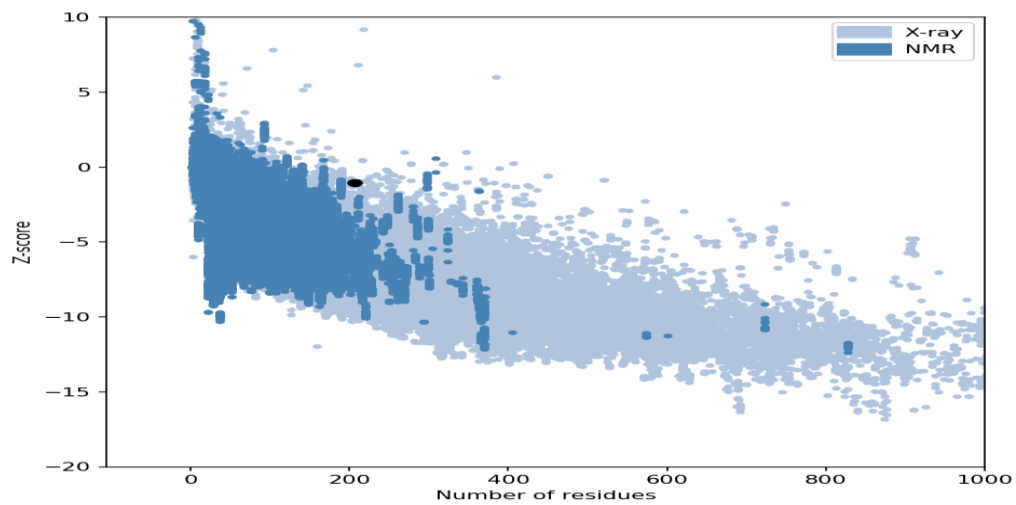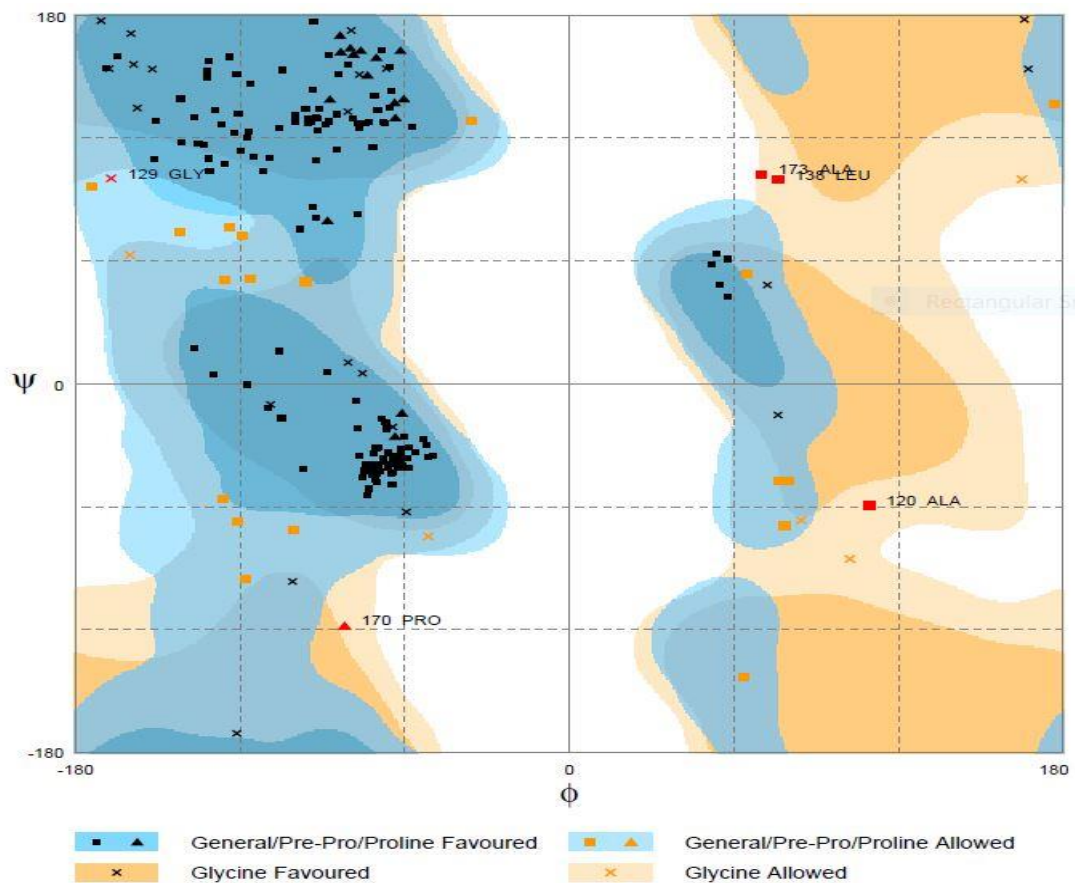

Figure 4: *Pro-SA web validation depicting position by black dot on the graph.*

Ramachandran plot analysis by RAMPAGE server (Figure 5) gave the following result:

Number of residues in favoured region (~98.0% expected): 176 (86.3%)
Number of residues in allowed region (~2.0% expected): 23 (11.3%)
Number of residues in outlier region: 5 (2.5%)

Figure 5: *Ramachandran Plot analysis of the final vaccine tertiary structure.*

The predicted B-cell linear epitopes were calculated using Ellipro suite (Figure 6):

| No.	Chain	Start	End	Peptide	Number of residues	Score
1	_	41	54	CRRKEAAAKQLTPT	14	0.777
2	_	166	206	RGPGPGKAYNVTQAFGRRGPEGPGPGVGLMWLSYFIASFRL	41	0.752
3	_	91	109	YFYYAAYLLEQWNLVIGPG	19	0.738
4	_	117	129	LLHAPATVCGGPG	13	0.729
5	_	12	28	RVRGGRCAVLSCLPKEE	17	0.605
6	_	32	38	KCSTRGR	7	0.537

Figure 6: B-cell linear epitopes predicted by Ellipro suit.

Toll-like receptor 8 was docked with the final model by PatchDock server and results were refined using FireDock server. The results were tabulated below (Table 2):

Rank	Solution number	Global energy	Attractive VdW	Repulsive VdW	ACE	HB
1	7	-2.31	-2.88	0.00	0.23	0.00
2	1	4.73	-3.47	0.95	1.86	0.00
3	4	5.57	-6.72	3.21	0.70	-0.47
4	5	12.05	-6.31	2.02	3.52	-0.87
5	3	147.38	-53.03	243.75	16.99	-7.49
6	10	1251.90	-37.08	1640.80	-3.79	-4.21
7	6	1971.10	-71.11	2501.80	23.67	-7.27
8	9	3102.50	-61.08	3966.91	7.23	-6.80
9	2	8424.86	-77.31	10692.53	7.72	-10.68
10	8	9662.56	-125.16	12275.23	21.48	-21.45

Table 2: FireDock ligand binding score table.

Solution number 7 was the most favourable binding conformation with global energy at -2.31 and 0.00 repulsive Vander Waal forces. The docked model was visualized using UCSF Chimera (Figure 7).
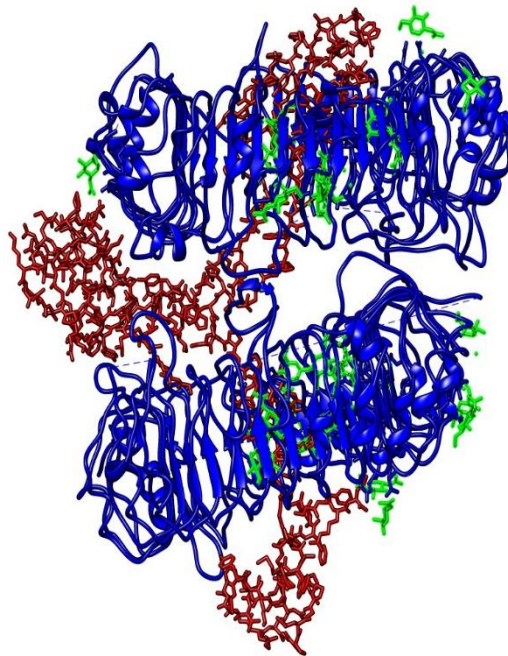
Figure 7: *Vaccine model(red) docked with Toll-like receptor 8(blue). Non-standard ligands of TLR-8 are highlighted in green.*

Discussion

The multi-epitope vaccine sequence obtained selected epitopes from those proteins which first come into contact with immune response according to viral replication cycle, like envelope, surface/spike glycoprotein and membrane protein. The protein sequence is predicted to be antigenic as well as non-allergic, hence proving its advantage of not producing any harmful hypersensitivity reaction in the body. It is basic in nature and has low molecular weight hence suitable for any rote of administration except oral. It's half-life in E. coli is >10 hours, hence can easily be cultured and extracted. It is thermally stable as indicated by instability index. It has various B-cell epitope stimulating site and molecular docking with toll like receptor TLR-8 shows that it binds easily it without any repulsive Van der Waal forces. Toll-like receptor 8 which induces immune response against ss-RNA organisms will elicit an immune response against this sequence considering it be an active coronavirus and thus fulfilling its purpose as a vaccine. 86.3% of the residues were in the favoured region of Ramachandran plot. These points make it favourable for in-vitro trials and further refinement. This sequence might have good hope of emerging as the vaccine of the current pandemic if studied more in depth. All these studies were on web-tool prediction servers designed for such type of studies. However, since they work on growing databases, these cannot give a complete surety for success in future stages. Along with the advantage of the study, there are some limitations. Larger peptide sequences like of ORF 1ab polyprotein was not performed because of limited computational resources. 19% of the predicted secondary structure is disordered. Instead of 98% proteins being in the favourable region of Ramachandran plot only 86.3% of them is present. Advanced molecular dynamic simulations were not performed like RMSD (root mean square deviation). Only proteins were considered for reverse vaccinology in this study [2]. These disadvantages need to be overcome with better resources but these early results do serve as a guiding path to build future work upon it.

Conclusion

This study has highlighted a potential vaccine candidate against coronavirus using reverse vaccinology approach. Further in-depth studies and refinement and serve as prove to be successful since it has very good results at such an early stage. It is needed to be validated experimentally. This study will aid infection control by creating immunological memory against coronavirus infection.

References

1. Coronavirus Situation report 105(4th May,2020) by World Health Organization.
2. Rappuoli R. Reverse vaccinology. Current opinion in microbiology. 2000 Oct 1;3(5):445-50.
3. Ali A, Khan A, Kaushik AC, Wang Y, Ali SS, Junaid M, Saleem S, Cho WC, Mao X, Wei DQ. Immunoinformatic and systems biology approaches to predict and validate peptide vaccines against Epstein–Barr virus (EBV). Scientific reports. 2019 Jan 24;9(1):1-2.
4. Khan S, Khan A, Rehman AU, Ahmad I, Ullah S, Khan AA, Ali SS, Afridi SG, Wei DQ. Immunoinformatics and structural vaccinology driven prediction of multi-epitope vaccine against Mayaro virus and validation through in-silico expression. Infection, Genetics and Evolution. 2019 Sep 1; 73:390-400.
5. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic acids research. 2011 Jan;39(Database issue): D32.
6. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC bioinformatics. 2007 Dec 1;8(1):4.
7. 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 Dec 1;8(1):424.
8. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R. Consortium, OSDD; Raghava, GPS In Silico Approach for Predicting Toxicity of Peptides and Proteins. PLoS One. 2013;8: e73957.
9. Calis JJ, 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 computational biology. 2013 Oct;9(10).
10. 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 Dec 1;11(1):568.
11. Saha S, Raghava GP. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic acids research. 2006 Jul 1;34(suppl_2): W202-9.
12. Dhanda SK, Vir P, Raghava GP. Designing of interferon-gamma inducing MHC class-II binders. Biology direct. 2013 Dec;8(1):30.
13. Sabourin M, Tuzon CT, Fisher TS, Zakian VA. A flexible protein linker improves the function of epitope-tagged proteins in Saccharomyces cerevisiae. Yeast. 2007 Jan;24(1):39-45.
14. Lee SJ, Shin SJ, Lee MH, Lee MG, Kang TH, Park WS, Soh BY, Park JH, Shin YK, Kim HW, Yun CH. A potential protein adjuvant derived from Mycobacterium tuberculosis Rv0652 enhances dendritic cells-based tumor immunotherapy. PloS one. 2014;9(8).
15. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. InThe proteomics protocols handbook 2005 (pp. 571-607). Humana press.
16. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nature protocols. 2015 Jun;10(6):845.
17. galaxy.seoklab.org [Internet]. [cited 2020May5]. Available from: http://galaxy.seoklab.org/

18. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic acids research. 2007 Jul 1;35(suppl_2): W407-10.
19. RAMPAGE: Ramachandran Plot Assessment. [cited 2020May5]. Available from: http://mordred.bioc.cam.ac.uk/~rapper/rampage.php
20. Ponomarenko J, Bui HH, Li W, Fusseder N, Bourne PE, Sette A, Peters B. ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC bioinformatics. 2008 Dec 1;9(1):514.
21. Duhovny D, Nussinov R, Wolfson HJ. Efficient unbound docking of rigid molecules. InInternational workshop on algorithms in bioinformatics 2002 Sep 17 (pp. 185-200). Springer, Berlin, Heidelberg.
22. Bank PD. Protein data bank. Nature New Biol. 1971; 233:223.
23. Mashiach E, Schneidman-Duhovny D, Andrusier N, Nussinov R, Wolfson HJ. FireDock: a web server for fast interaction refinement in molecular docking. Nucleic acids research. 2008 Apr 19;36(suppl_2): W229-32.
24. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. Journal of computational chemistry. 2004 Oct;25(13):1605-12.