Immunoinformatic design of a COVID-19 subunit vaccine using entire structural immunogenic epitopes of SARS-CoV-2

Esmaeil Behmard
Bijan Soleymani (bijansolimani@gmail.com)
Ali Najafi (najafi74@bmsu.ac.ir)
Ebrahim Barzegari (e.barzegari@kums.ac.ir)

Method Article

Keywords: SARS-CoV-2, Multi-epitope Vaccine, Molecular Dynamics Simulations, Toll-like Receptor 3, Immunoinformatics

Posted Date: December 2nd, 2020

DOI: https://doi.org/10.21203/rs.3.pex-923/v1

Abstract

One of the strategies against the lack of data and short onset time of the ongoing COVID-19 disease is the use of computational methods in the field of drug and vaccine design. Though demanding experimental validations, these methods can greatly reduce the time and cost of the immunogenic development projects. In this protocol, we describe the use of various immunoinformatics tools to design a multi-epitope vaccine polypeptide with the highest potential for activating the human immune system against SARS-CoV-2. The initial epitope set was extracted from the whole set of viral structural proteins. Their potential non-toxic and non-allergenic T-cell and B-cell binding and cytokine inducing epitopes were then identified through a priori immunoinformatic prediction. Selected epitopes were bonded to each other with appropriate links. A suitable adjuvant was added to the N-terminus of the vaccine polypeptide sequence to increase its immunogenicity. Molecular modelling of the 3D structure of the vaccine polypeptide, docking, molecular dynamics simulations and free energy calculations were used to confirm stability and receptor affinity. Time taken to complete: 43 days.

Introduction

Reagents

Equipment

Procedure

1. SARS-CoV-2 structural protein sequences. The amino acid sequences of SARS-CoV-2 structural proteins, including the spike glycoprotein (S), envelope protein (E), membrane protein (M), and nucleocapsid phosphoprotein (N), were retrieved using the NCBI reference genome of the virus (accession number NC_045512.2).

2. Identifying cytotoxic T lymphocyte (CTL) epitopes. Predicting peptides that are capable of inducing CTL responses is a crucial step in the design of epitope-based vaccine. The MHC-I binding tool of Immune Epitope Database and Analysis Resource (IEDB; http://tools.iedb.org/mhci) was used to predict the $CD8^+$ T-cell epitopes borne in S, E, M and N proteins[1-3]. In this step, the ANN 4.0 method was set as the prediction method. The human was selected as the source species. The maximum IC_{50} value was set to 50 nM, and percentile rank < 1 was considered since lower score indicates high affinity[4,5].

3. Identifying helper T lymphocyte (HTL) epitopes. IEDB (http://www.iedb.org) was used to predict MHC-II binding of 15-mer epitopes for viral structural proteins against human HLAs such as HLA-DRB1*15:01, HLA-DRB4*01:01, HLA-DRB3*01:01, HLA-DRB5*01:01, HLA-DRB1*03:01, HLA-DRB3*02:02 and HLA-DRB1*07:01 (Supplementary Table S8 online), using NN-align 2.3 method[6-9]. Epitopes with the IC_{50} values < 50 nM have high affinity. The maximum value of 500 nM and 5000 nM indicate intermediate and low affinity of epitopes, respectively[10]. The 15-mer epitopes with IC_{50} values < 50 nM were considered for next analysis[10-12].

4. B-cell epitopes prediction. The ABCpreds server was used to predict 16-mer linear B-lymphocyte (LBL) epitopes with a threshold of 0.51[13]. In addition, the ElliPro tool of IEBD was utilized to predict conformational and linear B-cell epitopes of the vaccine polypeptide.

5. Assessment of identified epitopes for antigenicity, allergenicity, and toxicity. The antigenic potential of each of the T and B cells epitopes was predicted by VaxiJen v2.0 with a threshold of 0.4[14]. The predicted T and B cells epitopes were then further evaluated to check their toxicity and allergenicity, with ToxinPred and AllergenFP v1.0 server, respectively[15,16]. Then, the ability of each of the HTL and B cell epitopes ($CD4^+$) to induce the secretion of cytokines, such as interferon-gamma (IFN-γ), interleukin-4 (IL-4) and interleukin-10 (IL-10), to overcome the inflammatory response and prevent tissue damage was predicted by IFNepitope, IL4pred and IL10pred server tools, respectively[17-19]. The IL4pred and IL10pred operations were carried out based on SVM method and hybrid method respectively, with other parameters kept as default[1,8].

6. Designing the multi-epitope vaccine polypeptide construct. To develop a multi-epitope vaccine, we selected those predicted epitopes with high antigenic potential, which were not identified as allergic or toxic, and had good solubility when highly expressed. To construct the vaccine polypeptide, the selected CTL, HTL and LBL epitopes were fused together using appropriate linkers[20,21]. An adjuvant was also considered to increase the immunogenic capacity of the multi-epitope vaccine[1,10]. Accordingly, a 45-amino acid sequence prepared from β-defensin-2 protein was added to the N terminus of the vaccine sequence using the EAAAK linker[21].

7. Immunogenic, allergenic and physiochemical evaluation of vaccine construct. The antigenicity of the multi-epitope vaccine polypeptide was predicted utilizing the VaxiJen v2.0 tool, with the threshold value 0.4[14]. The allergenicity of the vaccine was analysed using AllerTOP v.2.0 and AllergenFP v.1.0

servers[15,22]. The ProtParam server was employed to evaluate the physical chemistry properties of the vaccine construct, such as amino acid composition, molecular weight, theoretical isoelectric point (pI), grand average of hydropathicity (GRAVY), aliphatic and instability index, and *in vitro* and *in vivo* half-life[23].

8. Vaccine polypeptide structure modelling, refinement and validation. The SOPMA server was applied for analysing the secondary structural properties of the multi-epitope vaccine polypeptide[24]. The GalaxyWEB server was employed for modelling and refinement of the 3D structure model[25]. The server relaxes the model structure using repacking and molecular dynamics (MD) simulation. Next, RAMPAGE server and ProSA-web were used to validate the refined 3D model[26,27]. All of these tools give us the overall quality of the 3D structure of the peptide vaccine.

9. Molecular docking of the vaccine polypeptide to TLR-3. The binding of antigenic agents with the target immune cell protein is crucial for the creation of a suitable immune system response. For analysing the binding pattern of the multi-epitope vaccine polypeptide with TLR-3 (PDB ID: 2A0Z)[28], molecular docking analysis was performed by Hdock, Zdock, Cluspro and Hawkdock[29-32]. Among the molecular species docked by each of these applications, the best outputs were extracted, followed by the four docked molecules uploaded into the Hawkdock program, and the free energy of binding of each complex calculated. Based on this screening, it was determined that the model selected from the Cluspro output had the best binding free energy. Therefore, this molecular species was chosen as the primary structure for initiating the molecular dynamics (MD) simulation.

10. MD simulation of the vaccine-receptor complex. To determine the structural stability and to study the molecular details of the interactions between TLR-3 and the multi-epitope vaccine polypeptide in the docked conformation, MD simulation was performed. Briefly, the system including vaccine polypeptide-TLR-3 was simulated by the Gromacs-2020 package applying OPLS-AA force field[33]. The complex system was solvated using TIP3P water model. Then, the genion module was utilized to neutralize the whole system. Next, the conjugate gradient algorithm was applied to minimize the energy of the system. In the NVT ensemble, the temperature of the system gradually increased from 0 to 310 K during 400 ps. Subsequently, in the NPT ensemble, 500 ps simulation was carried out at a pressure of 1 atm and a temperature of 310 K. Production simulation for 24000 ps was then implemented. The particle-mesh Ewald (PME) and the LINCS algorithms were applied to assess all electrostatic connections and to restrain all bond lengths in the protein, respectively. Moreover, periodic boundary condition was utilized during the simulation. The final coordinates obtained for the complex system were analysed with classic

MD analyses, plus the MMPBSA method for calculating the free energy of intermolecular interactions[34]. The results were visualized by Pymol (Schrodinger L.L.C.).

Troubleshooting

Time Taken

SARS-CoV-2 structural protein sequences. 1 days

Identifying cytotoxic T lymphocyte (CTL) epitopes. 2 days

Identifying helper T lymphocyte (HTL) epitopes. 2 days

B-cell epitopes prediction. 2 days

Assessment of identified epitopes for antigenicity, allergenicity, and toxicity. 2 days

Designing the multi-epitope vaccine polypeptide construct. 1 days

Immunogenic, allergenic and physiochemical evaluation of vaccine construct. 3 days

Vaccine polypeptide structure modelling, refinement and validation. 5 days

Molecular docking of the vaccine polypeptide to TLR-3. 5 days

MD simulation of the vaccine-receptor complex. 20 days

Whole procedure: 43 days

Anticipated Results

References

1 Abdulla, F., Adhikari, U. K. & Uddin, M. K. Exploring T & B-cell epitopes and designing multi-epitope subunit vaccine targeting integration step of HIV-1 lifecycle using immunoinformatics approach. *Microb Pathog* **137**, 103791, doi:10.1016/j.micpath.2019.103791 (2019).

2 Larsen, M. V. *et al.* Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC bioinformatics* **8**, 424, doi:10.1186/1471-2105-8-424 (2007).

3 Yadav, G., Rao, R., Raj, U. & Varadwaj, P. K. Computational modeling and analysis of prominent T-cell epitopes for assisting in designing vaccine of ZIKA virus. *Journal of Applied Pharmaceutical Science* **7**, 116-122, doi:10.7324/japs.2017.70816 (2017).

4 Andreatta, M. & Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* **32**, 511-517, doi:10.1093/bioinformatics/btv639 (2016).

5 Nielsen, M. & Lund, O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC bioinformatics* **10**, 296, doi:10.1186/1471-2105-10-296 (2009).

6 Greenbaum, J. *et al.* Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. *Immunogenetics* **63**, 325-335, doi:10.1007/s00251-011-0513-0 (2011).

7 Weiskopf, D. *et al.* Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. *Proceedings of the National Academy of Sciences of the United States of America* **110**, E2046-2053, doi:10.1073/pnas.1305227110 (2013).

8 Khatoon, N., Pandey, R. K. & Prajapati, V. K. Exploring Leishmania secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach. *Sci Rep* **7**, 8285, doi:10.1038/s41598-017-08842-w (2017).

9 Narula, A., Pandey, R. K., Khatoon, N., Mishra, A. & Prajapati, V. K. Excavating chikungunya genome to design B and T cell multi-epitope subunit vaccine using comprehensive immunoinformatics approach to control chikungunya infection. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **61**, 4-15, doi:10.1016/j.meegid.2018.03.007 (2018).

10 Duvvuri, V. R. *et al.* Preexisting CD4+ T-cell immunity in human population to avian influenza H7N9 virus: whole proteome-wide immunoinformatics analyses. *PloS one* **9**, e91273, doi:10.1371/journal.pone.0091273 (2014).

11 Shey, R. A. *et al.* In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Sci Rep* **9**, 4409, doi:10.1038/s41598-019-40833-x (2019).

12 Zhou, W. Y. *et al.* Therapeutic efficacy of a multi-epitope vaccine against Helicobacter pylori infection in BALB/c mice model. *Vaccine* **27**, 5013-5019, doi:10.1016/j.vaccine.2009.05.009 (2009).

13 Saha, S. & Raghava, G. P. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* **65**, 40-48, doi:10.1002/prot.21078 (2006).

14 Doytchinova, I. A. & Flower, D. R. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC bioinformatics* **8**, 4, doi:10.1186/1471-2105-8-4 (2007).

15 Dimitrov, I., Naneva, L., Doytchinova, I. & Bangov, I. AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics* **30**, 846-851, doi:10.1093/bioinformatics/btt619 (2014).

16 Gupta, S. *et al.* In silico approach for predicting toxicity of peptides and proteins. *PloS one* **8**, e73957, doi:10.1371/journal.pone.0073957 (2013).

17 Dhanda, S. K., Gupta, S., Vir, P. & Raghava, G. P. Prediction of IL4 inducing peptides. *Clinical & developmental immunology* **2013**, 263952, doi:10.1155/2013/263952 (2013).

18 Dhanda, S. K., Vir, P. & Raghava, G. P. Designing of interferon-gamma inducing MHC class-II binders. *Biol Direct* **8**, 30, doi:10.1186/1745-6150-8-30 (2013).

19 Nagpal, G. *et al.* Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Sci Rep* **7**, 42851, doi:10.1038/srep42851 (2017).

20 Chen, X., Zaro, J. L. & Shen, W. C. Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev* **65**, 1357-1369, doi:10.1016/j.addr.2012.09.039 (2013).

21 Mohan, T., Sharma, C., Bhat, A. A. & Rao, D. N. Modulation of HIV peptide antigen specific cellular immune response by synthetic alpha- and beta-defensin peptides. *Vaccine* **31**, 1707-1716, doi:10.1016/j.vaccine.2013.01.041 (2013).

22 Dimitrov, I., Bangov, I., Flower, D. R. & Doytchinova, I. AllerTOP v.2–a server for in silico prediction of allergens. *J Mol Model* **20**, 2278, doi:10.1007/s00894-014-2278-5 (2014).

23 Gasteiger, E. *et al.* in *The Proteomics Protocols Handbook Springer Protocols Handbooks* (ed J.M. Walker) 571-607 (Humana Press, 2005).

24 Geourjon, C. & Deleage, G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Computer applications in the biosciences : CABIOS* **11**, 681-684, doi:10.1093/bioinformatics/11.6.681 (1995).

25 Ko, J., Park, H., Heo, L. & Seok, C. GalaxyWEB server for protein structure prediction and refinement. *Nucleic acids research* **40**, W294-297, doi:10.1093/nar/gks493 (2012).

26 Wang, W. *et al.* Data set for phylogenetic tree and RAMPAGE Ramachandran plot analysis of SODs in Gossypium raimondii and G. arboreum. *Data in brief* **9**, 345-348, doi:10.1016/j.dib.2016.05.025 (2016).

27 Wiederstein, M. & Sippl, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research* **35**, W407-410, doi:10.1093/nar/gkm290 (2007).

28 Bell, J. K. *et al.* The molecular structure of the Toll-like receptor 3 ligand-binding domain. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 10976-10980,

doi:10.1073/pnas.0505077102 (2005).

29 Kozakov, D. *et al.* The ClusPro web server for protein-protein docking. *Nature protocols* **12**, 255-278, doi:10.1038/nprot.2016.169 (2017).

30 Pierce, B. G., Hourai, Y. & Weng, Z. Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PloS one* **6**, e24657, doi:10.1371/journal.pone.0024657 (2011).

31 Weng, G. *et al.* HawkDock: a web server to predict and analyze the protein-protein complex based on computational docking and MM/GBSA. *Nucleic acids research* **47**, W322-W330, doi:10.1093/nar/gkz397 (2019).

32 Yan, Y., Zhang, D., Zhou, P., Li, B. & Huang, S. Y. HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic acids research* **45**, W365-W373, doi:10.1093/nar/gkx407 (2017).

33 Van Der Spoel, D. *et al.* GROMACS: fast, flexible, and free. *J Comput Chem* **26**, 1701-1718, doi:10.1002/jcc.20291 (2005).

34 Kumari, R., Kumar, R. & Lynn, A. g_mmpbsa--a GROMACS tool for high-throughput MM-PBSA calculations. *Journal of chemical information and modeling* **54**, 1951-1962, doi:10.1021/ci500020m (2014).

Acknowledgements