In silico evaluation of surface-exposed proteins of severe acute respiratory syndrome coronavirus 2 to propose a multi-epitope vaccine candidate

Article · January 2021

CITATIONS
0

READS
92

2 authors:

Sepideh Fereshteh
Pasteur Institute of Iran
21 PUBLICATIONS **43** CITATIONS
SEE PROFILE

Farzad Badmasti
Pasteur Institute of Iran
65 PUBLICATIONS **521** CITATIONS
SEE PROFILE

Health Biotechnology and Biopharma (2021), 4(4): 51-72

Original Research Article

In silico evaluation of surface-exposed proteins of severe acute respiratory syndrome coronavirus 2 to propose a multi-epitope vaccine candidate

Sepideh Fereshteh [1, †], Amin Sepehr [1,†], Nazanin Rahimirad [2], Rahimeh Sanikhani [1], Farzad Badmasti [1,*]

[1]Department of Bacteriology, Pasteur Institute of Iran, Tehran, Iran

[2]Department of Molecular Genetics, Tarbiat Modares University, Tehran, Iran

***Corresponding author:** Farzad Badmasti, Department of Bacteriology, Pasteur Institute of Iran, Tehran, Iran*

E-mail: f_badmasti@pasteur.ac.ir

[†] Both authors contributed equally to this manuscript.

DOI: 10.22034/HBB.2020.06

Received: January 18, 2021; **Accepted:** January 25, 2021

ABSTRACT

The novel Coronavirus disease 2019 caused a global outbreak therefore a promising vaccine is needed to combat it. In this study, the Spike (S) glycoprotein, Membrane (M) protein, and Envelope (E) protein have been tested as putative vaccine candidates. It was demonstrated that the second protrusion part of M protein contains several immunogenic epitopes. On the other hand, the S2 domain of S protein was found to be highly conserved. Therefore, we proposed a fusion protein as a multiepitope-based subunit vaccine including the second protrusion part of M protein and S2 domain of S protein with a flexible linker. Finally, the 3D structure and physicochemical properties of this fusion protein were evaluated to select a stable structure.

***Keywords*:** SARS-CoV-2, Spike glycoprotein, membrane protein, multiepitope-based subunit vaccine, *in silico*

INTRODUCTION

In December 2019, unusual pneumonia with an unknown etiology spread to Wuhan, China. In December 2019, studies showed that the causative agent of this pneumonia was a new coronavirus called Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and had developed to be a public health emergency

of international concern [1]. This family of viruses is called the coronaviruses because of the characteristic appearance by which they are identified in the electron microscope. Glycoproteins on the surface of the virus enable them to survive in the gastrointestinal tract and spread through the fecal-oral route [2,3]. In humans, coronaviruses cause respiratory infections such as common colds, which are usually mild. However, rarer forms such as Severe Acute Respiratory Syndrome (SARS), Middle East Respiratory Syndrome (MERS), and SARS-CoV-2 may be fatal [4]. The clinical symptoms of the Coronavirus disease 2019 (COVID-19) are non-specific, and in most patients, it is accompanied by fever and dry cough, and in one-third of cases by shortness of breath. Moreover, some patients show other symptoms such as lethargy, headache, sore throat, and diarrhea [5]. The incubation period of the virus is reported 5.2 up to 14 days and human-to-human transmission is confirmed [6].

The size of the coronavirus genome varies from 26,000 to 32,000 bp consisting of 6 to 11 Open Reading Frames (ORFs). The four major structural proteins include the Spike protein (S), the Envelope protein (E), the Membrane protein (M), and the nucleocapsid protein (N) [7]. A Receptor-Binding Domain (RBD) in SARS-CoV-2 the Spike protein mediates the interaction with the host cell receptor. After attachment to the cell, the host protease in the vicinity of the S protein releases the fusion peptide and facilitates virus entry [7].

At present, there is no effective drug or specific treatment available for COVID-19 [8] so there is an urgent need to develop a safe and effective vaccine to prevent this highly infectious disease. Vaccine development is time-consuming, and considerable collaborative efforts are required for preclinical and clinical assessment of vaccine candidates [8]. The full-length genome phylogenetic analysis suggests that the genetic structure of SARS-CoV-2 is almost 80 % similar to that of SARS-CoV [9,10]. Therefore, previously available related studies and existing data about vaccine designing attempts against the coronavirus (SARS/MERS) disease may be helpful to design a vaccine against COVID-19.

Three different strategies have been used to develop the vaccine for COVID-19 so far. The first generation of vaccine development is the whole virus vaccines include inactivated or live attenuated vaccines. This strategy is of great

importance because it provides the human body with a rapid immunogenic response against the COVID-19 infection due to the presence of natural antigenic materials. Numerous studies have reported developing live attenuated vaccines against coronaviruses in the past [11-13]. The second strategy is designing a protein subunit vaccine, which involves the use of synthetic or isolated parts of highly antigenic protein subunits with the short antigen segment which offers a safer approach [14]. Numerous protein subunit vaccines have been successfully formulated for various pathogens. Similarly, various kinds of full-length or segmented proteins are reported in literature against coronaviruses, which involves the receptor-binding domain, membrane proteins, nucleo-capaside proteins, spike proteins, or envelop protein [15-19]; and in the case of SARS-CoV-2, such researches are focused on the S protein. Another strategy is producing nucleic acid vaccines (genetically engineered DNA or RNA vaccines) with stable antigenic expression delivered by plasmids, which stimulates relatively lower but constant immune responses [20]. Studies reported promising results for nucleic acid vaccines against coronaviruses [21]. The safety and efficacy of these vaccines produced by each method should be experimentally tested and evaluated [22]. Due to the importance of COVID-19 infections, designing a protective vaccine seems necessary. Therefore, the aim of this study was *in silico* evaluation of surface-exposed proteins of severe acute respiratory syndrome coronavirus 2 as putative vaccine candidates.

MATERIALS AND METHODS

Upstream in silico analysis

Sequence retrieval

The protein sequences of S protein (accession number: YP_009724390.1), M protein (accession number: YP_009724393), and E protein (accession number: YP_009724392) of SARS-CoV-2 were obtained from the NCBI protein database [23] and saved as FASTA format for further investigations.

Prediction of physicochemical properties

The ProtParam tool, an online tool for the assessment of various physicochemical properties of a given protein, was used to predict the physicochemical properties of putative vaccine candidates [24]. The allergenicity of selected proteins was measured using the Algpred online tool.

Antigenicity measurement

The whole protein antigenicity was investigated using the VaxiJen tool with a threshold of 0.4 and immunomedicine Group (predicting antigenic peptides) [25].

Adhesion probability and similarity to the host proteins

The Vaxign 2 platform, a vaccine target prediction and analysis system based on the reverse vaccinology principle, was used to identify the adhesion probability and similarity to host (human, mouse, and pig) proteins [26].

Predicting the subcellular localization of viral proteins within the host cell

The Virus-mPLoc platform was used to predict the subcellular localization of selected proteins within the host and virus-infected cells [27].

Finding the conserved domains

The conserved domain database, Pfam and InterPro were used to identify the conserved domains of studied proteins [28-30].

2D, 3D structures and topology prediction

2D structure and membrane topology predictions of these viral proteins were gained from Phyre2 [31]. In the next step, we used structure prediction and structure-based function annotation of the SARS-CoV-2 genome conducted by I-TASSER [32].

The HMMTOP and CCTOP platforms were used to estimate the transmembrane helix and N-terminus location of these proteins [33,34].

Downstream in silico analyses

Linear and conformational B-cell epitopes prediction

The investigation of linear B-cell epitope prediction was performed by the IEBD with threshold ≥ 1. ElliPro and DiscoTope 2.0 Server was used to conformational epitopes prediction with threshold ≥ 0.7 and -3.7 respectively [35-37].

T-cell epitopes prediction

Prediction of the Major Histocompatibility Complex (MHC) I and II allele binding performed by the IEDB analysis resource (TepiTool) [38]. For binding predictions of MHC I alleles the inclusion criteria was as follows: 1. The host species was human, 2. Alleles were selected from the list of representative alleles from different HLA supertypes, 3. The default was applied based on settings for a moderate number of peptides and duplicate peptides were

removed, 4. The prediction method was according to IEBD recommendation.

The inclusion criteria for the MHC II alleles binding predictions were as follows: 1. The host species was human, 2. The prediction was conducted based on the custom allele set, 3. The default settings were applied based on settings for a moderate number of peptides and duplicate peptides were removed, 4. The prediction method was according to IEBD recommendation and peptides were selected with a predicted consensus percentile rank ≤ 0.5. To compare the immunogenicity of proteins a ratio was defined. For this purpose, the number of obtaining alleles was divided into the number of amino acids of each protein.

S protein conservation among different SARS-CoV-2 strains

The NCBI virus database was used to capture the S protein sequences of SARS-CoV-2 strains (published from 2020-01-12 to 2020-04-30) [39]. In the next step, the redundancies of these sequences were omitted using Jalview software [40] and multiple sequences alignment was done using MEGA software [41] and the conservation of S protein amino acids was investigated using the Consurf online tool [42].

Designing a multi-epitope vaccine

For designing a multi-epitope vaccine construct, the immunogenic epitopes of M protein were attached to the S2 domain of S protein by appropriate peptide linkers such as Pro-Gly-Pro-Gly. The 3D structure prediction of this multi-epitope vaccine construct was predicted using PHYRE2 and the physicochemical parameters for this construct were predicted using ProtParam.

RESULTS

Physicochemical properties, 2D structure, and other characteristics prediction

The results of antigenicity measurements, physicochemical properties, prediction of 2D structure, and other characteristics of S protein, M protein and E protein of SARS-CoV-2 were summarized in Table 1.

Conserved domain search

A conserved domain search for 3 selected proteins showed that E protein with only 1 functional domain plays a central role in the virus morphogenesis and assembly.

The M protein and its interactions with other structural proteins are necessary for the production and release of virus-like particles. This protein showed only 1

functional domain (residue: 5-222) and 3 transmembrane helices.

On the other hand, the S protein showed 3 functional domains. The first domain is the N-terminal domain of the S1 subunit (residue: 13-304) which contributes to the spike trimer interface. The second domain is the receptor-binding domain of the S1 subunit (residue: 319-541) and mediates attachment to the host cell. The last domain is the S2 subunit (residue: 662-1270) and mediates membrane fusion.

Linear and conformational B-cell epitopes prediction

Among the 3 selected proteins, E protein did not show linear and conformational B-cell epitopes according to defined criteria, so we exclude it from our investigation and continued the study with two other proteins.

The data from *in silico* investigations revealed that M protein consists of only 1 surface-exposed linear B-cell epitope (Met1, Ala2, Asp3) according to defined inclusion criteria. Also, this protein consists of 2 surface-exposed conformational B-cell epitopes. The sequences and locations of these epitopes were as follows: 1. Arg72, Leu73, Asn74, Leu76, Thr77, and 2. Met1, Ala2, Asp3, Ser4, Asn5, Gly6, Thr7, Leu8, Thr9, Val10, Glu11, Glu12, Lys15, Gln19, Trp20. These 2 surface-exposed conformational B-cell epitopes are illustrated in Figure 1 in red and yellow colors.

Linear B-cell epitope prediction of S protein was shown in Figure 2 (A). This protein has 10 surface-exposed linear B-cell epitopes with a threshold higher than 1. According to the IEBD results, the S protein consists of two conformational epitopes with thresholds higher than 0.7. These two epitopes and the location are depicted in Figure 2 (B). Based on the DiscoTope results, the S protein showed 73 surface-exposed conformational B-cell epitopes. Data is shown in Figure 2 (C).

T-cell epitopes prediction

MHC I and II binding allele's prediction was done for M protein. Among different introduced MHC I binding epitopes, we considered 5 of them because they were located on the second protrusion part of this protein and simultaneously were surface-exposed conformational B-cell epitopes. There were not any MHC II binding alleles located on the second protrusion part of the M protein according to our defined criteria.

The conservation of amino acids of S protein was investigated using the MEGA software and the Consurf online tool. The

results demonstrated that this protein, especially the S2 domain, is highly conserved among different strains of the SARS-CoV-2.

Designing a Multiepitope-based Subunit Vaccine

The sequence of the Multiepitope-based Subunit Vaccine was deposited in the GenBank database (accession number: MT585376). The ProtParam analyses showed that this construct includes 636 amino acids, with molecular weight 69538.58 (Da), and theoretical pI 5.51. The estimated half-life was >10 h in *Escherichia coli* (in vivo), >20 h in yeast (*in vivo*), and 7.2 h in mammalian reticulocytes (in vitro). This construct was stable according to the instability index (37.66). Based on the PHYRE2 results, this construct had disordered (21 %), alpha-helix (49 %), beta-strand (15 %), and TM helix (10 %). The 3D structure prediction was shown in Figure 3.

Table 1. Antigenicity measurement, physicochemical properties, 2D structure prediction, and other characteristics of 3 surface-exposed proteins of SARS-CoV-2

Viral Proteins / Characteristics	S protein	M protein	E protein
Number of amino acids	1273	222	75
Molecular weight (Da)	141178.47	25146.62	8365.04
Theoretical pI	6.24	9.51	8.57
Asp + Glu	110	13	3
Arg + Lys	103	21	5
The estimated half-life	>10 h (*Escherichia coli, in vivo*	>10 h (*Escherichia coli, in vivo)*	>10 h (*Escherichia coli, in vivo)*
Aliphatic index	84.67	120.86	144.00
Instability index	Stable	Stable	Stable
Average antigenic propensity	1.0416	1.0532	1.1202
Overall prediction for the protective antigen	0.46 (Probable ANTIGEN)	0.51 (Probable ANTIGEN)	0.60 (Probable ANTIGEN)
N-terminus location	OUT	OUT	OUT
Number of transmembrane helices	1	3	1
Disordered	16 %	15 %	21 %
Alpha helix	26 %	41 %	80 %

Beta strand	37 %	29 %	3 %
Transmembrane helices	7 %	27 %	56 %
Allergenicity	Potential ALLERGEN	NON-ALLERGEN	NON-ALLERGEN
Adhesion probability score	0.635	0.282	0.234
The similarity to host protein (human, mouse, pig)	No	No	No
Subcellular localization of viral proteins within the host cell	Host endoplasmic reticulum	The host cell membrane and host endoplasmic reticulum.	Host cell membrane
Surface exposed linear B-cell epitope	10	1	0
Surface exposed conformational B-cell epitope	2	2	0
MHC class I binding alleles ratio[a]	0.2	0.25	Not detected
MHC class II binding alleles ratio[b]	0.01	0.03	Not detected

[a,b] The number of obtaining alleles was divided into the number of amino acids of each protein, and the results were reported as a ratio.

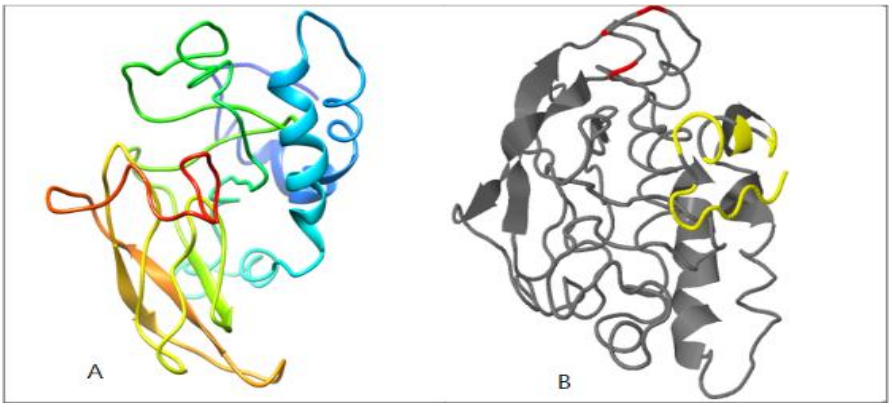

Figure 1. The 3D structure of the M protein of SARS-CoV-2 (A). Two surface-exposed conformational B-cell epitopes of M protein of SARS-CoV-2 were shown in yellow and red colors (B).

DISCUSSION

SARS-CoV-2 is a new human-infecting Betacoronavirus divergent from SARS-CoV. The phylogenetic analysis suggests that bats might act as the original hosts for this virus. However, the intermediate hosts are unknown and an animal sold at the seafood market in Wuhan should be the intermediate host. SARS-CoV-2 use Angiotensin-Converting Enzyme 2 (ACE2) as a receptor to enter the target cells. This receptor is widely expressed in the heart, liver, testis, kidney, intestine cells, and other tissues. The RBD domain of S protein interacts with the ACE2 receptor and mediates cell entry [44]. This virus causes severe lung pathology by inducing an inflammatory form of programmed cell death called pyroptosis. The exact pathogenic mechanism of this virus is still unclear. However, it is suggested that a viral protein directly interacts with inflammasome NLRP3 (nucleotide-binding domain leucine-rich repeat (NLR) and pyrin domain-containing receptor 3) and subsequently induces pro-inflammatory cytokines (IL-1β and IL-18) production [45]. According to the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, in 2021, 96, 185, 360 confirmed cases of Covid-19 and 2,057, 743 deaths are reported worldwide [46]. Given the rapid spread of the disease as well as the high morbidity and mortality rate, design and manufacture a vaccine to prevent the disease seems necessary.

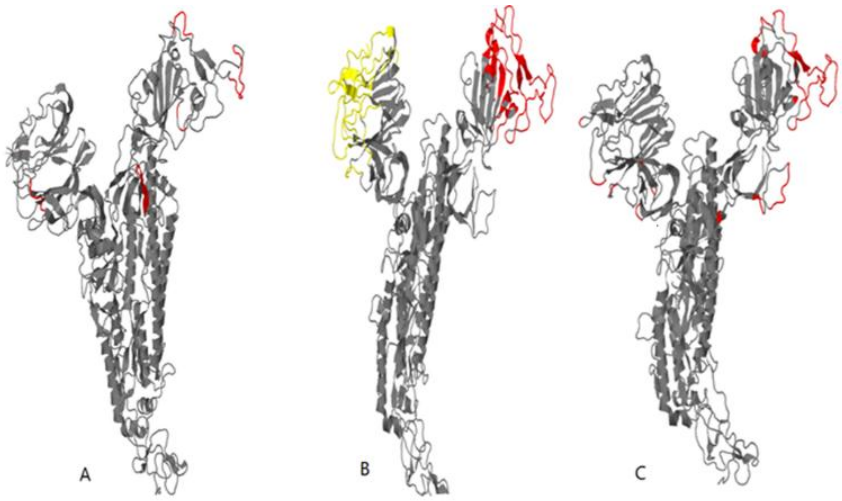

Figure 2. The 3D structure of the S protein of SARS-CoV-2 (A-C). Predicted linear B-cell epitopes were shown in red color (A). Two conformational B-cell epitopes according to IEBD results were shown in yellow and red colors (B). Surface-exposed conformational B-cell epitopes based on DiscoTope results were shown in red color (C).

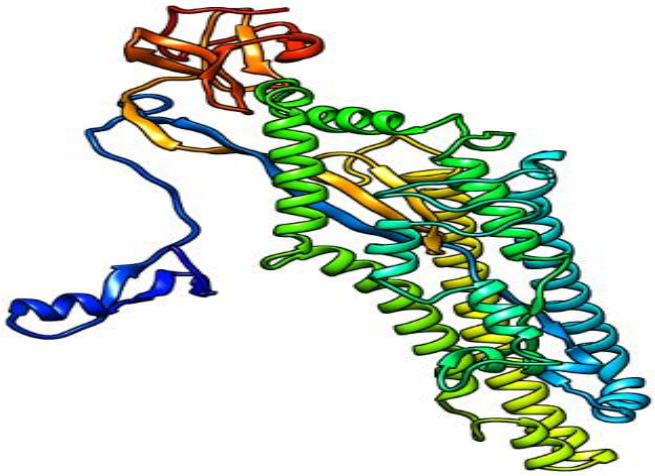

Figure 3. The 3D structure prediction of multi-epitope vaccine construct. This construct included the S2 domain of S protein of SARS-COV2 and protrusion part of M protein.

Act as the original hosts for this virus. However, the intermediate hosts are unknown and an animal sold at the seafood market in Wuhan should be the intermediate host [43].

Nowadays, numerous researchers have tried to design and develop an effective vaccine against coronavirus, and researchers are testing 41 vaccines in phase I, 22 vaccines in phase II, and 23 vaccines in phase III clinical trials. Moreover, 8 vaccines are approved for limited and 2 vaccines were approved for full use [47]. Several efforts have been attempted to design the inactivated vaccine formulations to get effective protection from SARS-COV 2. Among the developed vaccines in phase III clinical trials, the Wuhan Institute of Biological Products produced an inactivated virus vaccine, BBIBP-CorV, which showed promising results in phase III trials and is currently approved for limited use in U.A.E [48]. Another inactivated virus vaccine launched Phase III trials in the U.A.E and Argentina named Sinopharm developed by the Beijing Institute of Biological Products [49]. Moreover, the CoronaVac by Sinovac Biotech and the Covaxin by Bharat Biotech has also inactivated virus vaccines in phase III trials.

Although the inactivated vaccines are considered safe due to the absence of living pathogens and their inability to possible re-infection, they might revert into virulent phases [50-52]. Also, some studies reported inflammation and lung lesion with eosinophil infiltration in the case of inactivated coronaviruses vaccines [52,53]. A few articles reported weaker immune response or delayed immune response in inactivated vaccines [54] which requires multiple dosages [55]. Therefore, weaker and unexpected immune response might be the major limitation associated with the use of inactivated vaccines.

Another vaccine designing strategy is called the third generation vaccines or nucleic acid vaccines which use fragments of genetic material such as the S (Spike) protein as RNA vaccines. Although RNA vaccines can be quickly designed and easily manufactured, they induce a poor immune response. DNA vaccines could potentially integrate into the host genome, while there are concerns about the stability of RNA vaccines [56]. There are currently no approved DNA or RNA vaccines for medical use in humans [57].

Moderna in partnership with the National Institutes of Health (NIH) developed an approved mRNA vaccine to produce viral proteins in the body [58]. The mRNA-1273, an mRNA vaccine that encodes a SARS-CoV-2 spike protein stabilized in the prefusion conformation, could protect against SARS-CoV-2 infection in the lungs and noses of mice without evidence of

immunopathology [59]. In this approach, the S protein residues 986 and 987 were mutated to produce prefusion-stabilized SARS-CoV-2 S (2P) protein. The efficacy of this vaccine is 94.5 % and needs 2 doses, 4 weeks apart. Moderna vaccine can be stored with refrigeration for 30 days and at –4 °F (–20 °C) for 6 months.

Another approved mRNA vaccine is developed by The German company BioNTech with collaborations of Pfizer, and the Fosun Pharma company named BNT162b1. This vaccine is a lipid-nanoparticle-formulated, nucleoside-modified mRNA encoding the trimerized Receptor-Binding Domain (RBD) of the spike glycoprotein of SARS-CoV-2 [60]. Nowadays, the injection of Pfizer-BioNTech vaccine is approved in Canada, Saudi Arabia, the U.S, and other countries. The efficacy of this vaccine is 95 % and needs 2 doses 3 weeks apart. This vaccine must be stored only at –94 °F (–70 °C) so the distribution of it around the world may be challengeable.

Adenovirus vector-based vaccines use the antigenic protein component of the pathogenic virus within a non-virulent vector and mimic the natural infection process with subsequent cellular and humoral immunogenicity [61]. It is difficult to rapidly design a vector-based vaccine for newly emerging viruses like SARS-CoV-2 since it is necessary to know the exact epidemiology, genotoxicity, and virology of both pathogenic and vector viruses [19]. Also, the major disadvantage of this strategy is delayed immune response against the pathogenic virus, because the initial immune response is mainly due to the vector and not the pathogen [61]. Moreover, the risk of mutation and unexpected virulence ability of the engineered vector should not be overlooked [62]. An Adenovirus-based vaccine developed by Chinese company CanSino Biologics called Ad5, expressing the spike glycoprotein of a SARS-CoV-2 strain showed promising results and is currently approved for limited use in China [63]. Also, The Gamaleya Research Institute, part of Russia's Ministry of Health, launched 3 phase clinical trials for a vaccine called Gam-Covid-Vac consisting of two adenoviruses, Ad5 and Ad26. A recombinant adenovirus type 26 (rAd26) vector and a recombinant adenovirus type 5 (rAd5) vector, both carry the spike glycoprotein (rAd26-S and rAd5-S) gene for SARS-CoV-2. This is now approved for early use in Russia and negotiated to supply the vaccine to other countries as well [64].

The AstraZeneca company and the University of Oxford developed a vaccine based on a chimpanzee adenovirus called (ChAdOx1 nCoV-19) expressing the SARS-CoV-2 full-length spike surface glycoprotein compared with a meningococcal conjugate vaccine (MenACWY) as control (19, 65). Another recombinant nanoparticle vaccine rSARS-CoV-2 called NVX-CoV2373 composed of trimeric full-length SARS-CoV-2 Spike glycoproteins and Matrix-M1 adjuvant launched phase III clinical trials [66].

Recent advancements in vaccine development for SARS-COV suggested that the protein subunit vaccine designing is an ideal and safer option to design a vaccine against the SARS-CoV-2 virus. These vaccines include only essential antigenic protein subunits for stimulating the immunogenic responses. They are less likely to cause complications and adverse effects in vaccinated individuals because the antigenic components employed in protein subunit vaccines are purified and do not involve the use of pathogenic viruses [14]. Some studies reported the induction of immunogenic response, for instance neutralizing antibodies, IgA, IgG, Th-1, and Th-2 (67) [18,68,69]. The most important feature of these types of vaccines is their higher safety and a lower rate of adverse effects compared to the first generation of vaccines since actual naturally occurring viral components are not available [14]. Moreover, this approach provides an opportunity to design vaccines against multiple epitopes, for similar or different kinds of pathogen strains [16]. The lack of other viral components may lead to a poor or delayed immunogenic response, and this could be addressed by the use of appropriate adjuvants [17,68,70]. An outstanding advantage of subunit vaccines is their availability, rapid production, and potential enhancements by the use of effective adjuvants, particularly in the pandemic situation due to definite immunogenic components of protein subunits. Currently, there are many vaccines under pre-clinical studies using this platform. A notable example of a vaccine candidate designed based on this platform is developed by Novavax, which is currently in phase III clinical trial.

In this study, we evaluated 3 structural proteins of SARS-CoV-2 to find the best putative vaccine candidate. *In silico* analysis demonstrated that among the 3 studied proteins, E protein is not considered as the desired vaccine candidate because it lacks immunogenic epitopes. This result is consistent with the study of Nieto-Torres. They declared that E protein is consists of

76–109 amino acids in different types of coronaviruses and its immunogenicity is limited [71].

According to *in silico* results derived from the literature, the S protein is considered the most desirable vaccine candidate [72]. Not only the full-length protein, but also different parts of this protein such as NTD, S1 domain, and S2 domain have the potential to be considered as putative vaccine candidates.

In this study, we evaluated the full-length S protein and discovered this protein is a promising vaccine candidate because the data showed the highest adhesion probability score and 1 transmembrane helix. Adhesions are important vaccine targets as they are critical for viruses to invade host cells. Moreover, the recombination and purification process of proteins containing more than 1 transmembrane helix often leads to failure [73]. Although full-length protein can maintain the correct conformation of the protein and is capable of providing more epitopes and exhibiting higher immunogenicity, the S2 domain is more desirable because it is highly conserved among different strains of SARS-Cov2 and contains multiple immunogen epitopes. Therefore, we selected this part of the S protein to construct a Multiepitope-based Subunit Vaccine. Moreover, M protein is the most abundant protein in the virion and has a high antigenic propensity [74]. It was reported that immunization with the full length of M protein can induce neutralizing antibody production in SARS infected patients [75]. Due to this protein has 3 transmembrane helices, we focused on protrusion parts and predicted conformational B-cell and MHC I binding epitopes and attached these epitopes to the S2 domain of S protein to design a multi-epitope construct as a putative vaccine with the highest protection. According to the study of Jafari *et al*, we selected a Pro-Gly-Pro-Gly linker, which demonstrated that this linker could be a good choice for fusion proteins especially subunit vaccine [76]. For designing a Multiepitope-based Subunit Vaccine, the instability index, and the half-life of the construct should be considered. For this purpose, we made two different constructs first including the M protein with the S2 domain and second including the S2 domain with the M protein. The ProtParam analyses showed that the first construct was stable and had more half-life and worthy to clone and express in a suitable host to evaluate its efficacy. Moreover, in the original form of S protein, the S2 domain is located on the

C terminal so the construct of M protein with the S2 domain was closer to the original conformation.

In Silico Study conducted by Avinash and Bhuvnesh demonstrated that the spike protein of the SARS-CoV-2 has 65 antigenic sites so can be considered as a promising vaccine candidate [77]. In another study designed by Singh *et al.,* B-cell, T-helper, and cytotoxic T-cell epitopes of Spike protein were selected and linked with GPGPG and AAY linkers as a vaccine candidate. Consistent with our study they focused on spike protein to introduce a promising vaccine candidate. However, besides the spike protein, we considered the importance of membrane protein for vaccine design [78]. In the study by Bhattacharya and colleagues, they collected 13 MHC-I and 3 MHC-II 9 mer epitopes of spike protein to construct a peptide-based vaccine against SARS-CoV-2 [79]. Moreover, Normalina and colleagues supplied data regarding mutations in the S glycoprotein and proposed an epitope-based vaccine candidate against Indonesian SARS-CoV-2 [80]. Then, Enayatkhani *et al.,* have constructed a multi-epitope vaccine candidate based on three known antigenic proteins (Nucleocapsid, ORF3a, and Membrane protein) and showed that this chimeric protein could elicit humoral and cellular immune responses [81].

CONCLUSION

Numerous vaccines against the SARS-COV-2 virus were being developed and tested worldwide in early-phase studies and some vaccine candidates are already in phase III clinical trials for assessing their efficacy. Unfortunately, no effective medical interventions like vaccines have been provided for the treatment of this infection. There is an urgent need for a vaccine to control the existing pandemic, and due to the gene sequence homology of SARS-CoV and SARS-CoV-2, the past efforts of vaccine development against SARS and MERS will be of great value in the present scenario. The most commonly used antigen for vaccine development is the RDB S protein segment, and our data showed a highly antigenic nature to induce both humoral and cellular immune responses. In the present study, we used an integrated vaccinology approach to identify the surface-exposed peptides in lesser time and lower cost. Application of immunoinformatic tools and databases helped us to predict physicochemical characteristics, antigenicity, adhesion, subcellular localization, conservation, 3D

and 2D structures of the S protein, and then discover linear and conformational epitopes of B and T-cells. Finally, we designed an effective and specific vaccine candidate.

REFERENCES

[1]. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J Autoimmun*. 2020: 102433.

[2]. Mustafa N , Zahoor H , Majoo F . Pandemic SARS Coronavirus-2 infections in humans-COVID-19. *Bilim. Derg.* 2020; (10): 77-93.

[3]. Chen Y, Liu Q, Guo D. Emerging coronaviruses: genome structure, replication, and pathogenesis. *J Med Virol*. 2020; 92(4): 418-23.

[4]. Peeri NC, Shrestha N, Rahman MS, Zaki R, Tan Z, Bibi S, et al. The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned? *Int. J. Epidemiol.* 2020.

[5]. Del Rio C, Malani PN. 2019 Novel coronavirus important information for clinicians. *Jama*. 2020; 323(11): 1039-40.

[6]. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, *et al.* Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus infected pneumonia in Wuhan, China. *Jama*. 2020; 323(11): 1061-69.

[7]. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet. 2020; 395(10224): 565-74.

[8]. Iyengar K, Bahl S, Vaishya R, Vaish A. Challenges and solutions in meeting up the urgent requirement of ventilators for COVID-19 patients. *Diabetes Metab Syndr*. 2020.

[9]. Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ, et al. COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet*.. 2020; 395(10229): 1033.

[10]. Dhama K, Sharun K, Tiwari R, Dadar M, Malik YS, Singh KP, et al. COVID-19, an emerging coronavirus infection: advances and prospects in designing and developing vaccines, immunotherapeutics, and therapeutics. Hum Vaccin Immunother. 2020:1-7.

[11]. Regla-Nava JA, Nieto-Torres JL, Jimenez-Guardeño JM, Fernandez-Delgado R, Fett C, Castaño-Rodríguez C, *et al.* Severe acute respiratory syndrome coronaviruses with mutations in the E protein are attenuated and promising

vaccine candidates. *J. Virol.* 2015; 89(7): 3870-87.

[12]. Menachery VD, Gralinski LE, Mitchell HD, Dinnon KH, Leist SR, Yount BL, *et al*. Combination attenuation offers strategy for live attenuated coronavirus vaccines. *J. Virol.* 2018; 92(17).

[13]. Jimenez-Guardeño JM, Regla-Nava JA, Nieto-Torres JL, DeDiego ML, Castaño-Rodriguez C, Fernandez-Delgado R, et al. Identification of the mechanisms causing reversion to virulence in an attenuated SARS-CoV for the design of a genetically stable vaccine. *PLoS Pathog.* 2015; 11(10): 1005215.

[14]. Vartak A, Sucheck SJ. Recent advances in subunit vaccine carriers. *Vaccines.* 2016; 4(2): 12.

[15]. Ababneh M, Mu'men Alrwashdeh MK. Recombinant adenoviral vaccine encoding the spike 1 subunit of the Middle East respiratory syndrome Coronavirus elicits strong humoral and cellular immune responses in mice. *Vet World.* 2019; 12(10): 1554.

[16]. Kalita P, Padhi A, Zhang KY, Tripathi T. Design of a peptide-based subunit vaccine against novel coronavirus SARS-CoV-2. *Microb Pathog.* 2020: 104236.

[17]. Kim E, Erdos G, Huang S, Kenniston TW, Balmert SC, Carey CD, *et al.* Microneedle array delivered recombinant coronavirus vaccines: Immunogenicity and rapid translational development. *EBio Med.* 2020: 102743.

[18]. Du L, Tai W, Yang Y, Zhao G, Zhu Q, Sun S, et al. Introduction of neutralizing immunogenicity index to the rational design of MERS coronavirus subunit vaccines. *Nat Commun.* 2016; 7(1): 1-9.

[19]. Folegatti PM, Bittaye M, Flaxman A, Lopez FR, Bellamy D, Kupke A, *et al.* Safety and immunogenicity of a candidate Middle East respiratory syndrome coronavirus viral-vectored vaccine: a dose-escalation, open-label, non-randomised, uncontrolled, phase 1 trial. *Lancet Infect. Dis*. 2020.

[20]. Liu MA. DNA vaccines: an historical perspective and view to the future. *Immunol. Rev.* 2011; 239(1): 62-84.

[21]. Kato T, Takami Y, Deo VK, Park EY. Preparation of virus-like particle mimetic nanovesicles displaying the S protein of Middle East respiratory syndrome coronavirus using insect cells. *J.* Biotechnol. 2019; 306: 177-84.

[22]. Chen W-H, Strych U, Hotez PJ, Bottazzi ME. The SARS-CoV-2 vaccine pipeline: an overview. *Curr Trop Med Rep.* 2020: 1-4.

[23]. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, *et al.* Database resources of the national center

for biotechnology information. *Nucleic Acids Res*. 2007; 36: 13-21.

[24]. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. *The proteomics protocols handbook: Springer;* 2005. 571-607.

[25]. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC bioinformatics. 2007; 8(1): 4.

[26]. He Y, Xiang Z, Mobley HL. Vaxign: the first web-based vaccine design program for reverse vaccinology and applications for vaccine development. *Biomed Res Int.* 2010;2010.

[27]. Shen H-B, Chou K-C. Virus-mPLoc: a fusion classifier for viral protein subcellular location prediction by incorporating multiple sites. *J Biomol. Struct Dyn.* 2010; 28(2): 175-86.

[28]. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res. 2020;48(D1):D265-D8.

[29]. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* 2019; 47: 427-32.

[30]. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, *et al*.: genome-scale protein function classification. *Bioinformatics.* 2014; 30(9): 1236-40.

[31]. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc.* 2015; 10(6): 845.

[32]. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nature methods.* 2015; 12(1): 7-8.

[33]. Tusnady GE, Simon I. The HMMTOP transmembrane topology prediction server. *Bioinformatics.* 2001; 17(9): 849-50.

[34]. Dobson L, Reményi I, Tusnády GE. CCTOP: a Consensus Constrained TOPology prediction web server. *Nucleic Acids Res.* 2015; 43: 408-12.

[35]. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res*. 2019;47: 339-43.

[36]. Ponomarenko J, Bui H-H, Li W, Fusseder N, Bourne PE, Sette A, *et al.* ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC *Bioinformatics*. 2008; 9(1): 514.

[37]. Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B cell epitope predictions: impacts of method

development and improved benchmarking. *PLoS Comput Biol.* 2012; 8(12).

[38]. Paul S, Sidney J, Sette A, Peters B. TepiTool: a pipeline for computational prediction of T cell epitope candidates. *Curr Protoc Immunol.* 2016; 114(1): 1-.9.

[39]. Hatcher EL, Zhdanov SA, Bao Y, Blinkova O, Nawrocki EP, Ostapchuck Y, *et al*. Virus variation resource improved response to emergent viral outbreaks. *Nucleic Acids Res.* 2017; 45: 482-90.

[40]. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. A multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 2009; 25(9): 1189-91.

[41]. Stecher G, Tamura K, Kumar S. Molecular evolutionary genetics analysis (MEGA) for macOS. *Mol Biol Evo.* 2020; 37(4): 1237-39.

[42]. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, *et al.* An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 2016; 44: 344-50.

[43]. Guo Y-R, Cao Q-D, Hong Z-S, Tan Y-Y, Chen S-D, Jin H-J, *et al*. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak an update on the status. *Mil Med Res.* 2020; 7(1): 1-10.

[44]. Guo, L, Yu, K, Li, D, Yang, H, Liu, L, Fan, J, Sun, N, Yang, X. Potential pathogenesis of multiple organ injury in COVID-19. 2020.

[45]. Merad M, Martin JC. Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. *Nat Rev Immunol.* 2020:1-8.

[46]. WHO. Coronavirus disease 2019 (COVID-19) Situation Report.

[47]. Parikhani Arezoo Beig, et al. The inclusive review on SARS-CoV-2 biology, epidemiology, diagnosis, and potential management Options. *Current microbiology,* (2021): 1-16.

[48]. Xia S, Zhang Y, Wang Y, Wang H, Yang Y, Gao GF, et al. Safety and immunogenicity of an inactivated SARS-CoV-2 vaccine, BBIBP-CorV: a randomised, double-blind, placebo-controlled, phase 1/2 trial. *Lancet Infect. Dis*. 2020.

[49]. Poland GA, Ovsyannikova IG, Kennedy RB. SARS-CoV-2 immunity: review and applications to phase 3 vaccine candidates. *Lancet.* 2020.

[50]. Xiong S, Wang Y-F, Zhang M-Y, Liu X-J, Zhang C-H, Liu S-S, et al. Immunogenicity of SARS inactivated vaccine in BALB/c mice. *Immunol Lett.* 2004; 95(2): 139-43.

[51]. Takasuka N, Fujii H, Takahashi Y, Kasai M, Morikawa S, Itamura S, *et al*. A subcutaneously injected UV-inactivated SARS coronavirus vaccine elicits systemic humoral immunity in mice. *Int Immunol.* 2004; 16(10): 1423-30.

[52]. He Y, Zhou Y, Siddiqui P, Jiang S. Inactivated SARS-CoV vaccine elicits high titers of spike protein-specific antibodies that block receptor binding and virus entry. *Biochem Biophys Res Commun.* 2004; 325(2): 445-52.

[53]. Agrawal AS, Tao X, Algaissi A, Garron T, Narayanan K, Peng BH, *et al.* Immunization with inactivated Middle East Respiratory Syndrome coronavirus vaccine leads to lung immunopathology on challenge with live virus. *Hum Vaccin Immunother.* 2016; 12(9): 2351-56.

[54]. Zhang C-h, Lu J-h, Wang Y-f, Zheng H-y, Xiong S, Zhang My, *et al.* Immune responses in Balb/c mice induced by a candidate SARS-CoV inactivated vaccine prepared from F69 strain. Vaccine. 2005; 23(24): 3196-201.

[55]. Roberts A, Lamirande EW, Vogel L, Baras B, Goossens G, Knott I, *et al.* Immunogenicity and protective efficacy in mice and hamsters of a β-propiolactone inactivated whole virus SARS-CoV vaccine. *Viral Immunol.* 2010; 23(5): 509-19.

[56]. Zhang C, Maruggi G, Shan H, Li J. Advances in mRNA vaccines for infectious diseases. *Front Immunol.* 2019; 10: 594.

[57]. Hobernik D, Bros M. DNA vaccines—how far from clinical use? *Int J Mol Sci.* 2018; 19(11): 3605.

[58]. Le TT, Andreadakis Z, Kumar A, Roman RG, Tollefsen S, Saville M, *et al.* The COVID-19 vaccine development landscape. *Nat Rev Drug Discov.* 2020; 19(5): 305-6.

[59]. Corbett KS, Edwards D, Leist SR, Abiona OM, Boyoglu-Barnum S, Gillespie RA, *et al*. SARS-CoV-2 mRNA vaccine development enabled by prototype pathogen preparedness. *BioRxiv.* 2020.

[60]. Mulligan MJ, Lyke KE, Kitchin N, Absalon J, Gurtman A, Lockhart S, *et al.* Phase I/II study of COVID-19 RNA vaccine BNT162b1 in adults. *Nature.* 2020: 1-5.

[61]. Bouard D, Alazard-Dany N, Cosset FL. Viral vectors: from virology to transgene expression. *Br J Pharmacol.* 2009; 157(2): 153-65.

[62]. Ura T, Okuda K, Shimada M. Developments in viral vector-based vaccines. *Vaccines.* 2014; 2(3): 624-41.

[63]. Zhu F-C, Li Y-H, Guan X-H, Hou L-H, Wang W-J, Li J-X, *et al.* Safety, tolerability, and immunogenicity of a recombinant adenovirus type-5 vectored

COVID-19 vaccine: a dose-escalation, open-label, non-randomised, first-in-human trial. *The Lancet.* 2020.

[64]. Logunov DY, Dolzhikova IV, Zubkova OV, Tukhvatullin AI, Shcheblyakov DV, Dzharullaeva AS, *et al*. Safety and immunogenicity of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine in two formulations: two open, non-randomised phase 1/2 studies from Russia. *The Lancet.* 2020; 396(10255): 887-97.

[65]. Folegatti PM, Ewer KJ, Aley PK, Angus B, Becker S, Belij-Rammerstorfer S, et al. Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *The Lancet.* 2020; 396(10249): 467-78.

[66]. Keech C, Albert G, Cho I, Robertson A, Reed P, Neal S, *et al.* Phase 1–2 trial of a SARS-CoV-2 recombinant spike protein nanoparticle vaccine. *N Engl J Med.* 2020.

[67]. He Y, Zhou Y, Liu S, Kou Z, Li W, Farzan M, *et al.* Receptor-binding domain of SARS-CoV spike protein induces highly potent neutralizing antibodies: implication for developing subunit vaccine. Biochem. Biophys. *Res. Commun.* 2004; 324(2): 773-81.

[68]. Zakhartchouk AN, Sharon C, Satkunarajah M, Auperin T, Viswanathan S, Mutwiri G, et al. Immunogenicity of a receptor-binding domain of SARS coronavirus spike protein in mice: implications for a subunit vaccine. *Vaccine*. 2007; 25(1): 136-43.

[69]. Tang J, Zhang N, Tao X, Zhao G, Guo Y, Tseng C-TK, *et al.* Optimization of antigen dose for a receptor-binding domain-based subunit vaccine against MERS coronavirus. *Hum Vaccin Immunother*. 2015; 11(5): 1244-50.

[70]. Guo Y, Sun S, Wang K, Zhang S, Zhu W, Chen Z. Elicitation of immunity in mice after immunization with the S2 subunit of the severe acute respiratory syndrome coronavirus. *DNA Cell Biol.* 2005; 24(8): 510-15.

[71]. Nieto-Torres JL, DeDiego ML, Verdia-Baguena C, Jimenez-Guardeno JM, Regla-Nava JA, Fernandez-Delgado R, *et al.* Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogenesis. *PLoS Pathog*. 2014; 10(5).

[72]. Zhang J, Zeng H, Gu J, Li H, Zheng L, Zou Q. Progress and prospects on vaccine development against SARS-CoV-2. *Vaccines.* 2020; 8(2): 153.

[73]. Xiang Z, He Y. Genome-wide prediction of vaccine targets for human

herpes simplex viruses using Vaxign reverse vaccinology. *BMC Bioinformatics*. 2013; 14(4): 2.

[74]. Neuman BW, Kiss G, Kunding AH, Bhella D, Baksh MF, Connelly S, *et al*. A structural analysis of M protein in coronavirus assembly and morphology. *J Struct Biol*. 2011; 174(1): 11-22.

[75]. Pang H, Liu Y, Han X, Xu Y, Jiang F, Wu D, *et al.* Protective humoral responses to severe acute respiratory syndrome-associated coronavirus: implications for the design of an effective protein-based vaccine. *J Gen Virol.* 2004; 85(10): 3109-13.

[76]. Jafari D, Dehghan NF, Honari H, Hoseini R, Jafari R. Bioinformatic analysis of different fusions of ipaD, PA20 and CTxB antigens: a preliminary analysis for vaccine design. 2016.

[77]. Kumar A, Sharma BP. *In silico* study to predict and characterize of SARS CoV 2 surface glycoprotein. *Vaccine Res.* 2020; 7(1): 10-6.

[78]. Singh H, Jakhar R, Sehrawat N. Designing spike protein (S-Protein) based multi-epitope peptide vaccine against SARS COVID-19 by immunoinformatics. *Heliyon.* 2020; 6(11): 5528.

[79]. Bhatnager R, Bhasin M, Arora J, Dang AS. Epitope based peptide vaccine against SARS-COV2: an immune-informatics approach. *J Biomol Struct Dyn.* 2020: 1-16.

[80]. Normalina I, Indrasari S, Nidom RV, Kusala MK, Alamudi MY, Santoso KP, *et al.* Characterization of the spike glycoprotein and construction of an epitope-based vaccine candidate against Indonesian SARS-CoV-2: *In silico* study. *Sys Rev Pharm.* 2020; 11(7): 404-13.

[81]. Enayatkhani M, Hasaniazad M, Faezi S, Guklani H, Davoodian P, Ahmadi N, *et al.* Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study. *J Biomol Struct Dyn*. 2020: 1-16.