**ORIGINAL ARTICLE**

# En route to Peptide Therapeutics for COVID 19: Harnessing Potential Antigenic Mimicry Between Viral and Human Proteins

**Maya Madhavan[1] · Sabeena Mustafa[2]**

## Abstract

Molecular mimicry is a general strategy used by pathogens to infect the host cells. The emergence of SARS-CoV-2 virus has resulted in more than 6,700,000 infections and 390,000 deaths worldwide. Coronavirus disease (COVID-19) is an infectious disease caused by this virus. In this project concept, we aim to focus on the peptide-protein interaction analysis using two important drug targets in SARS-CoV-2 such as spike (S) protein and nucleocapsid (N) protein. These proteins play an important role in the virus entry and encapsidation of the viral particles. Motifs or functional regions in these two proteins must be sharing sequence homology with human protein (ACE2) which may be involved in the binding mechanism. The results will show a set of motif regions which can disrupt the viral infection. Once we identify these sets of antigenic determinant regions, antibody binding activity studies can be performed by in vitro methods. Our results from this study may suggest the existence of antigenic mimicry between SARS-CoV-2 and host proteins. The hit peptide components will have therapeutic applications to be developed into a wide variety of medicinal formulations against SARS-CoV-2 such as vaccine, intranasal and inhalation formulations. Also, the choice of conserved regions will lead to development of cross protective therapeutics against wide range of coronaviruses.

**Keywords** SARS-CoV-2 · COVID-19 · Spike · Nucleocapsid · Immunogenicity · Peptide-protein interaction

## Background

There is no approved vaccine or therapeutic option developed yet for COVID-19. So, the availability of various therapeutic options against COVID-19 will provide benefits for the control of the present pandemic situation as well as a control for reemergence of the virus in future. The effort of vaccine development is an important research area. Several strategies such as small molecules, repurposing of FDA approved compounds, and different classes of vaccines candidates are being developed in the labs and some are even under trials.

Peptide-based vaccines are designed to elicit immunity against particular pathogens by selectively stimulating antigens specific for B and T cells. Peptide vaccines have a whole lot of advantages though they suffer from certain challenges such as low immunogenicity, susceptibility to enzymatic digestion etc. (Skwarczynski and Toth 2016). Advanced bioinformatics tools and databases help to speed up and design peptide-based vaccines for SARS-CoV-2 where peptides could be used as ligands (Abdelmageed et al. 2020). The viable approach in designing peptide vaccines should be to identify peptide epitopes on the immunogens and to use synthetic versions of these peptides. Since these vaccines are totally synthetic, they would not carry the risk of reversion or of complete inactivation, and in principle; epitopes could be selected to avoid components that give rise to unwanted side effects (https://www.who.int/biologicals/publications/trs/areas/vaccines/peptide/WHO_TRS_889_A1.pdf?ua=1). Also, peptide-based vaccines do not need in vitro culture making them biologically safe, and their selectivity allows accurate activation of immune responses (Dudek et al. 2010). Studies shows that effective vaccines can be developed using T cell and B cell

✉ Sabeena Mustafa
msabeena@gmail.com

1 Department of Biochemistry, Govt Arts and Science College, Kulathoor, Neyyattinkara, Thiruvananthapuram, India

2 Department of Biostatistics and Bioinformatics, King Abdullah International Medical Research Center (KAIMRC), Ministry of National Guard Health Affairs (MNGHA), Riyadh, Kingdom of Saudi Arabia

epitopes comprising the immunome, which constitutes the immune system (Groot et al. 2008; Tahir ul Qamar et al. 2019). Epitopes are short amino acid sequences of a protein which can induce an immune response and epitope finding is quite expensive and laborious. The use of in silico methods are an alternative and effective strategy due to the importance in biomedical applications such as vaccine design, disease diagnostics, biomarker studies etc. (Bahrami et al. 2019; Backert and Kohlbacher 2015). There are many peptide vaccines in development, such as vaccine for AIDS by the virus human immunodeficiency virus (HIV) (Liu et al. 2007), malaria (Epstein et al. 2007), anthrax (Oscherwitz et al. 2010), human papilloma virus (HPV) (Solares et al. 2011). Our hypothesis is based on the fact that, epitope region of the drug targets which share the sequence identity and similarity to human proteins may be involved in their binding and fusion to human cells (Hwa et al. 2008).

## Selection of S and N Protein Drug Targets of SARS-CoV-2

### Key Properties of Two Drug Targets, S and N Proteins

SARS-Co-V-2 possesses many druggable targets. Coronavirus entry into host cell is mediated by the transmembrane Spike (S) glycoprotein that forms homotrimers protruding from the viral surface (Tortorici and Veesler 2019). S comprises two functional subunits responsible for binding to the host cell receptor ($S_1$ subunit) and fusion of the viral and cellular membranes ($S_2$ subunit). SARS-CoV-2 and several SARS-related coronaviruses (SARS-CoV) interact directly with angiotensin-converting enzyme 2 (ACE2) via S to enter target cells (Song et al. 2018; Kirchdoerfer et al. 2018; Ge et al. 2013). As the coronavirus S glycoprotein is surface-exposed and mediates entry into host cells, it is the main target of neutralizing antibodies (Abs) upon infection and the focus of therapeutic and vaccine design. After binding of RBD in S1 subunit of S protein on the virion to the ACE2 receptor on the target cell, the heptad repeat 1 (HR1) and 2 (HR2) domains in its S2 subunit interact with each other to form a six-helix bundle (6-HB) fusion core, bringing viral and cellular membranes into close proximity for fusion and infection (Bosch et al. 2004).

The nucleocapsid (N) protein is one of the major structural proteins of the virus and is involved in the transcription and replication of viral RNA, packaging of the encapsidated genome into virions (Chang et al. 2006; Hurst et al. 2009) and interference with cell cycle processes of host cells. Moreover, in many coronaviruses, including SARS-CoV-2, the N protein has high immunogenic activity and is abundantly expressed during infection. Thus, both S and N proteins may be potential targets for design of peptide vaccines against COVID-19.

## SARS-CoV and SARS-CoV-2 Sequence Homology

SARS-CoV and SARS-CoV-2 have 89.8% sequence identity in their S protein S2 subunits, which mediate the membrane fusion process, and both of their S1 subunits utilize human angiotensin-converting enzyme 2 (hACE2) as the receptor to infect human cells (Zhou et al. 2020). Also, the ACE2-binding affinity of the receptor-binding domain (RBD) in S1 subunit of S protein of SARS-CoV-2 is 10- to 20-fold higher than that of SARS-CoV which may contribute to the higher infectivity and transmissibility of SARS-CoV-2 compared to SARS-CoV (Wrapp et al. 2020). Sequence alignment studies also show that S2 subunits of SARS-CoV-2 and SARS-CoV share consensus regions with 92.6% and 100% in HR1 and HR2 domains, respectively (Xia et al. 2020). The presence of large amount of sequence identity between SARS-CoV and SARS-CoV-2 enables designing of vaccines which are not only cross-protective across a wide spectrum of betacoronaviruses but also relatively resistant to the ongoing virus evolution.

## Methodology

### Identification and Retrieval of S And N Protein Drug Targets from Repositories

We searched the online repositories namely Genbank, NCBI protein and UniProtKb to retrieve the suitable sequence entries of the drug targets. The search resulted in many sequence entries and we selected two ids for further analysis as shown in Table 1.

As the first step in our analysis, S and N protein sequences will be analyzed to identify their immunogenic regions. Algorithms which can predict immunogenicity, secondary structure analysis, protein topology features and hydrophobicity will be studied to design peptides for further testing. Immunogenic viral peptides could be identified based on the algorithm developed by Kolaskar and Tongaonkar. This method makes use of the physicochemical properties of amino acid residues and its frequencies of occurrence in experimentally identified epitopes which are developed to identify antigenic determinants in proteins (Kolaskar and Tongaonkar 1990). These epitopes are the short residue set of peptides with antigenicity. IEDB (https://tools.iedb.org) is one of the tools using this algorithm, which is mainly used to determine the antigenic sites with a default threshold value of 1.024. The prediction algorithm works as follows based on the paper published by Kolaskar and Tongaonkar in 1990. The algorithm is as follows.

**Table 1** Spike protein and nucleocapsid protein sequences information from repositories

| Sl. no | Sequence | Sequence repository and corresponding id | Length of residues (aa) |
|---|---|---|---|
| 1 | Spike | Genbank-NC_045512.2, NCBI protein: QHR63290.2, QJS57231.1 UniProtKB-P0DTC2 | 1273 |
| 2 | Nucleocapsid | Genbank-NC_004718.3, NCBI protein: BCF79921.1, BCF79911.1 UniProtKB- P0DTC9 | 419 |

1. Calculate the average propensity values (Ap) for each peptides from N terminus to C terminus of the protein and assign the result to the fourth $(i+3)$ residue in the segment.
2. The average antigenic propensity of the protein will be determined
3. If the Ap of the protein is above 1.0 then all residues are potentially antigenic.
4. If the average for the protein is below 1.0 then all residues having above the average for the protein are potentially antigenic.
5. To select the antigenic determinant as a condition, six consecutive residues must satisfy step 3

The above step will help to identify the region of viral proteins (S and N) that are homologous to human protein which will be shortlisted based on the score. The peptides which display high immunogenicity, location on the protein surface and low hydrophobicity are to be selected, following which these peptides will be synthesized and tested by ELISA for binding to antibodies present in the convalescent plasma of COVID 19 patients. This is in line with the theory that the simplest and most direct approach to combating SARS-CoV-2 during the outbreak would be to use plasma from the convalescent patients (Kruse 2020). Patients who have been diagnosed with the disease, ascertained by RT-PCR and who are in their convalescent stage (> 20 days of incubation) should be selected for the study. The peptides which show good amount of binding and with negligible auto immune response can be selected for the design of peptide vaccines. Table 2 shows the timeline, major resources and the milestones of the project on a quarterly basis. The detailed methodology is represented in the flowchart (Fig. 1).

## Summary and Conclusion

Our results may suggest the molecular mimicry between SARS-CoV-2 drug target protein sequences (spike and nucleocapsid) and human proteins. Our study will use immunoinformatics tools to identify the possible epitopes of spike and nucleocapsid proteins of SARS-CoV-2. These epitopes can be the ideal candidates to formulate the initial component to develop a therapy against COVID-19 either in the form of peptide and sera based vaccine, based on the selected antigenic determinants. However, our analysis will require several in vitro and in vivo validations before formulating the therapeutic products to resist COVID-19. Antibody binding activity of these motifs will be performed in vitro to evaluate the neutralization activity against SARS-CoV-2 infection or participate in the immunopathogenesis induced by SARS-CoV-2. Our studies on the mimicry motifs of S and N protein, may provide approaches to inhibit or disrupt the infection of SARS-CoV-2. The peptides selected from the study after testing for toxicity in vivo, can have wide range of applications- be made into vaccines or intranasal and inhalation formulations. Also, since these peptides are selected on the basis of their binding specificity to regions

**Table 2** Timeline, major resources and the milestones of the project on a quarterly basis

| | Q1 | Q2 | Q3 |
|---|---|---|---|
| Timeline | Literature search, retrieval of S and N protein sequences from repositories. Query sequences will be analyzed to find immunogenic regions using peptide-protein interaction studies | Identify the regions of S and N which are homologous to human protein. A list of peptides with high immunogenicity, location on the protein surface and low hydrophobicity are to be selected | Peptide synthesis and wetlab validation using sera from patients. Evaluate the antibody formation in response to SARS-CoV-2 infection ELISA test and evaluation |
| Resources | PUBMED, GENBANK, UniProtKb, PDB, Piper (Schrodinger) | NCBI BLAST, MSA, IEDB analysis resource database, NetMHC server, EpiJen, MHCpred, EpiSearch, PepSurf, BIMAS, Pep3D | Peptide Synthesizer, RT-PCR, ELISA system |

**Fig. 1.** The detailed methodology is represented in the flowchart below
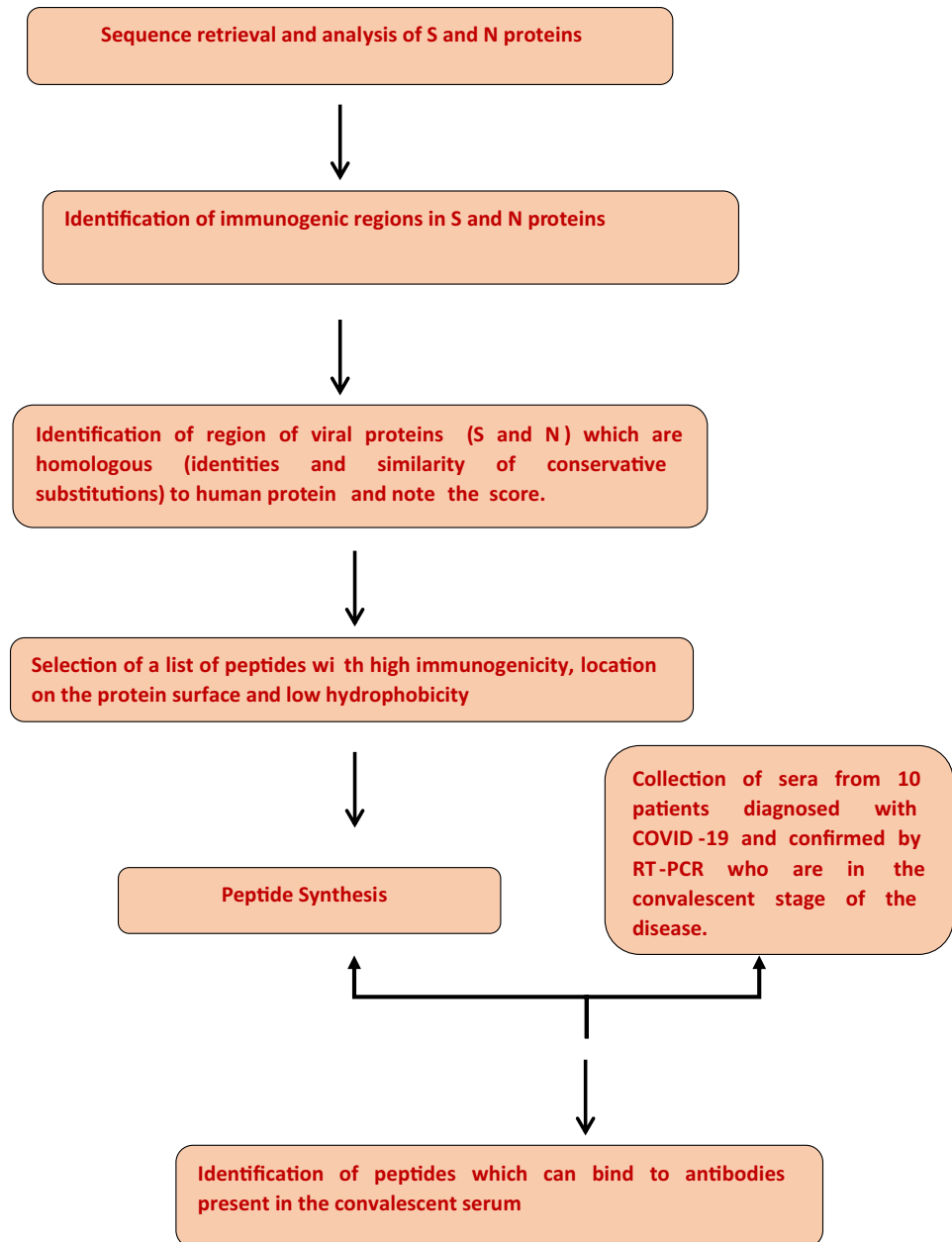


of the viral proteins that are highly conserved across various coronaviruses, the execution of such therapeutic approaches could lead to cross-protective antibody responses against future coronavirus epidemics and pandemics and help us be better prepared for the next novel coronavirus outbreak.

# References

Abdelmageed MI, Abdelmoneim AH, Mustafa MI, et al (2020) Design of multi epitope-based peptide vaccine against E protein of human COVID-19: an immunoinformatics approach. bioRxiv 2020.02.04.934232. https://doi.org/10.1101/2020.02.04.934232

Backert L, Kohlbacher O (2015) Immunoinformatics and epitope prediction in the age of genomic medicine. Genome Med 7:119. https://doi.org/10.1186/s13073-015-0245-0

Bahrami AA, Payandeh Z, Khalili S et al (2019) Immunoinformatics. In silico approaches and computational design of a multi-epitope. Immun Protein Int Rev Immunol 38:307–322. https://doi.org/10.1080/08830185.2019.1657426

Bosch BJ, Martina BEE, Van Der Zee R et al (2004) Severe acute respiratory syndrome coronavirus (SARS-CoV) infection inhibition using spike protein heptad repeat-derived peptides. Proc Natl Acad Sci USA 101:8455–8460. https://doi.org/10.1073/pnas.0400576101

Chang C, Sue S-C, Yu T et al (2006) Modular organization of SARS coronavirus nucleocapsid protein. J Biomed Sci 13:59–72. https://doi.org/10.1007/s11373-005-9035-9

De Groot AS, Moise L, McMurry JA, Martin W (2008) Epitope-based immunome-derived vaccines: a strategy for improved design and safety. Clin Appl Immun 2:39–69. https://doi.org/10.1007/978-0-387-79208-8_3

Dudek NL, Perlmutter P, Aguilar M-I et al (2010) Epitope discovery and their use in peptide based vaccines. Curr Pharm Des 16:3149–3157. https://doi.org/10.2174/138161210793292447

Epstein JE, Giersing B, Mullen G et al (2007) Malaria vaccines: are we getting closer? Curr Opin Mol Ther 9:12–24

Ge X-Y, Li J-L, Yang X-L et al (2013) Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature 503:535–538. https://doi.org/10.1038/nature12711

Hurst KR, Koetzner CA, Masters PS (2009) Identification of in vivo-interacting domains of the murine coronavirus nucleocapsid protein. J Virol 83:7221–7234. https://doi.org/10.1128/JVI.00440-09

Hwa K-Y, Lin WM, Hou Y-I, Yeh T-M (2008) Peptide mimicrying between SARS coronavirus spike protein and human proteins reacts with SARS patient serum. J Biomed Biotechnol 2008:326464. https://doi.org/10.1155/2008/326464

Kirchdoerfer RN, Wang N, Pallesen J et al (2018) Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. Sci Rep 8:15701. https://doi.org/10.1038/s41598-018-34171-7

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

Kruse RL (2020) Therapeutic strategies in an outbreak scenario to treat the novel coronavirus originating in Wuhan. China. F1000Research 9:72. https://doi.org/10.12688/f1000research.22211.2

Liu Y, McNevin J, Zhao H et al (2007) Evolution of human immunodeficiency virus type 1 cytotoxic T-lymphocyte epitopes: fitness-balanced escape. J Virol 81:12179–12188. https://doi.org/10.1128/JVI.01277-07

Oscherwitz J, Yu F, Cease KB (2010) A synthetic peptide vaccine directed against the 2ß2-2ß3 loop of domain 2 of protective antigen protects rabbits from inhalation anthrax. J Immunol 185:3661–3668. https://doi.org/10.4049/jimmunol.1001749

Skwarczynski M, Toth I (2016) Peptide-based synthetic vaccines. Chem Sci 7:842–854. https://doi.org/10.1039/c5sc03892h

Solares AM, Baladron I, Ramos T et al (2011) Safety and immunogenicity of a human papillomavirus peptide vaccine (CIGB-228) in Women with high-grade cervical intraepithelial neoplasia: first-in-human proof-of-concept trial. ISRN Obstet Gynecol 2011:292951. https://doi.org/10.5402/2011/292951

Song W, Gui M, Wang X, Xiang Y (2018) Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. PLoS Pathog 14:e1007236–e1007236. https://doi.org/10.1371/journal.ppat.1007236

Tahir ul Qamar M, Saleem S, Ashfaq UA et al (2019) Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: an immune-informatics study. J Transl Med 17:362. https://doi.org/10.1186/s12967-019-2116-8

Tortorici MA, Veesler D (2019) Structural insights into coronavirus entry. Adv Virus Res 105:93–116. https://doi.org/10.1016/bs.aivir.2019.08.002

Wrapp D, Wang N, Corbett KS et al (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 367:1260–1263. https://doi.org/10.1126/science.abb2507

Xia S, Liu M, Wang C et al (2020) Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. Cell Res 30:343–355. https://doi.org/10.1038/s41422-020-0305-x

Zhou P, Yang X-L, Wang X-G et al (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270–273. https://doi.org/10.1038/s41586-020-2012-7