

## Research Article

# Design of CRISPR-based Targets for the Development of a Diagnostic Method for SARS-CoV-2: An *in Silico* Approach

 Abraham Peele Karlapudi,<sup>1</sup>  Chinna Venkateswarulu Thirupati,<sup>1</sup>  Krupanidhi Srirrama,<sup>1</sup>  
 Divyashree Coimbatore Nageswaran,<sup>2</sup>  Indira Mikkili,<sup>1</sup>  Vijaya Sai Ayyagari,<sup>1</sup>  Ranganadha Reddy Aluri,<sup>1</sup>  
 Yasha Nazir Butt,<sup>3</sup>  Aishwarya Gangadhar<sup>1</sup>

<sup>1</sup>Department of Bio-Technology, Vignan's Foundation for Science, Technology and Research, India

<sup>2</sup>Tango Group International Ltd., United Kingdom

<sup>3</sup>Department of Chemistry, Wayne State University, Michigan, USA

### Abstract

**Objectives:** To computationally design CRISPR target-based Single-guided RNA (sg RNA) for the development of a diagnostic method for SARS-CoV-2.

**Methods:** In the present study, Cas12 nuclease-specific protospacer adjacent motif (PAM) and downstream target sequences were identified from conserved regions of the SARS-CoV-2 spike glycoprotein S-gene receptor-binding domain (RBD) using CHOPCHOP and NEB computational tools to design sgRNA. Further, the *in silico* expression vector was constructed using SnapGene software.

**Results:** The 20-nucleotide sequence 5'-ACCATACAGAGTAGTAGTAC-3' showed 100% similarity with SARS-CoV-2 was selected. Target-specific oligonucleotide sgRNA template was generated with the addition of an extra nucleotide 'G' to the 5' end using NEBicalculator.

**Conclusion:** The findings of this study will help designing a CRISPR-based diagnostic kit for the detection of SARS-CoV-2. However, computationally designed sgRNA requires to be tested *in vitro* to demonstrate its potential and efficiency in detecting this novel coronavirus.

**Keywords:** CHOPCHOP, CRISPR-Cas12, SARS-CoV-2, sgRNA

**Cite This Article:** Karlapudi AP, Venkateswarulu Thirupati C, Srirrama K, Nageswaran DC, Mikkili I, Ayyagari VS, Reddy Aluri R, et al. Design of CRISPR-based Targets for the Development of a Diagnostic Method for SARS-CoV-2: an *in Silico* Approach. EJMO 2020;4(4):304–308.

CRISPR technology originated from repetitive nucleic acid sequences discovered in bacterial DNA. Among these sequences viral genome particles were found. These particles are used by CRISPR associated proteins to target the invading viral genome.<sup>[1, 2]</sup> Cas proteins have the ability of cleaving viral nucleic acids when activated by complementary base pairing of CRISPR RNA with the viral genome. Viruses are a threat to humanity as 65% of them are infectious and fatal. Thanks to CRISPR Cas technology, scien-

tists can detect and eliminate viruses with high precision and accuracy. Among CRISPR Cas systems, scientists have taken the advantage of a natural tool called CRISPR Cas9, a powerful gene editing technology with an immense potential to treat human diseases and has wider applications.<sup>[3]</sup> Cas 12 and Cas 13 proteins belonging to the Cas family, use their guide RNAs to search for the viral RNA sequence to introduce collateral cleavage of any RNA encountered.<sup>[4]</sup> The technology is used as a diagnostic tool to detect vi-

**Address for correspondence:** Chinna Venkateswarulu Thirupati, Vignan's Foundation for Science, Technology and Research, Vadlamudi-522213, Guntur, Andhra Pradesh, India

**Phone:** +9000390204 **E-mail:** venki\_biotech327@yahoo.com

**Submitted Date:** July 27, 2020 **Accepted Date:** August 30, 2020 **Available Online Date:** November 30, 2020

©Copyright 2020 by Eurasian Journal of Medicine and Oncology - Available online at [www.ejmo.org](http://www.ejmo.org)

**OPEN ACCESS** This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.



ral and bacterial pathogens. The CRISPR Cas gene-editing technique now plays a prominent role in diagnostic biology.<sup>[5]</sup> There are four sub-types of CRISPR RNA combined with Cas 13; namely, Cas 13 (VI-A, VI-B, VI-C, and VI-D) and Cas 12 consisted of two subtypes; namely (V-A and V-B).<sup>[6]</sup> Nucleic acid samples go through amplification, which is an essential requirement for virus detection. The amplified fragment is made accessible by CRISPR-Cas proteins specially designed to get output from collateral cleavage, and its signal is detected by a flow sensing device. DETECTR (DNA Endonuclease Targeted CRISPR Trans Reporter) and SHERLOCK (Specific High Sensitivity Enzymatic Reporter UNLOCKing) are Cas12 and Cas 13-based detection techniques respectively. The natural RNase activity of Cas 13 and Cas 12 proteins is employed to target both RNA and non-specific ssDNA (single-stranded DNA) degradation respectively. The two systems, DETECTR and SHERLOCK generate a messenger signal to detect the presence of the target virus, SARS-CoV-2. The method follows the amplification of target DNA or RNA using RPA (recombinase polymerase amplification) or RT-RPA (reverse transcription recombinase polymerase amplification), respectively.<sup>[7]</sup> Cas 12a recognizes a 5-TTN-3 PAM site, whereas Cas 13 recognizes a 3 A/G/U PAM site. DETECTR and SHERLOCK systems can work in a similar way by targeting nucleotide-specific sequences and a fluorescent quencher reports the output after the nucleic acids have been degraded. The two detection systems were developed for rapid diagnosis, low cost and high-sensitivity that clearly distinguishing the close members of the coronavirus family, including SARS-CoV, MERS-CoV and others at atomolar levels within an hour.<sup>[8]</sup> The detection of this novel coronavirus, SARS-CoV-2 involves an incubation process in which the extracted nucleic acid is isothermally amplified with the addition of engineering the Cas protein.<sup>[9]</sup> Designing the sgRNA requires accurate target sites with the lowest off-target potential.<sup>[10]</sup> In this study, we suggest the *in silico* designed potential sgRNAs to detect the Spike protein gene of the COVID-19 virus from human nucleic acid samples.

## Methods

### Phylogenetic Tree

The S-gene sequences of SARS-CoV-2 from different geographical locations were selected from the NCBI server (<https://www.ncbi.nlm.nih.gov/>), and a phylogenetic tree was created according to the Maximum-Likelihood method using the open-source MEGA 7.0 version.<sup>[11]</sup> The conserved sequence of the S-gene (Gene ID: 43740568) was selected from the Wuhan reference genome (NC\_045512.2) and compared to other closely related species encoding the spike glycoprotein.

### Designing of CRISPR-Cas Targets

CHOPCHOP is an easy and straightforward tool to design sgRNA against any gene of interest and is available in both CRISPR and TALEN modes. CHOPCHOP algorithms search for potential off-targets and predict any frame-shift mutation frequencies. The selected conserved sequence of the S-gene was submitted to CHOPCHOP (<https://chopchop.cbu.uib.no/>). The genomic targets of 20 nucleotide gene sequences preceding 5-TTN-3 PAM were generated.<sup>[12]</sup> The output sequences were selected according to sequencing by minimum off-targets and were submitted to NCBI-Blast to further filter out the sequence showing the lowest sequence similarity to other viral genomes. We designed target-specific DNA oligos from the selected 20 nucleotide sequence using NEBioCalculator (<https://nebiocalculator.neb.com/>) and thus synthesized sgRNA. Hence, primer oligos were mainly designed for isothermal amplification of sgRNA followed by CRISPR-Cas12 based detection.

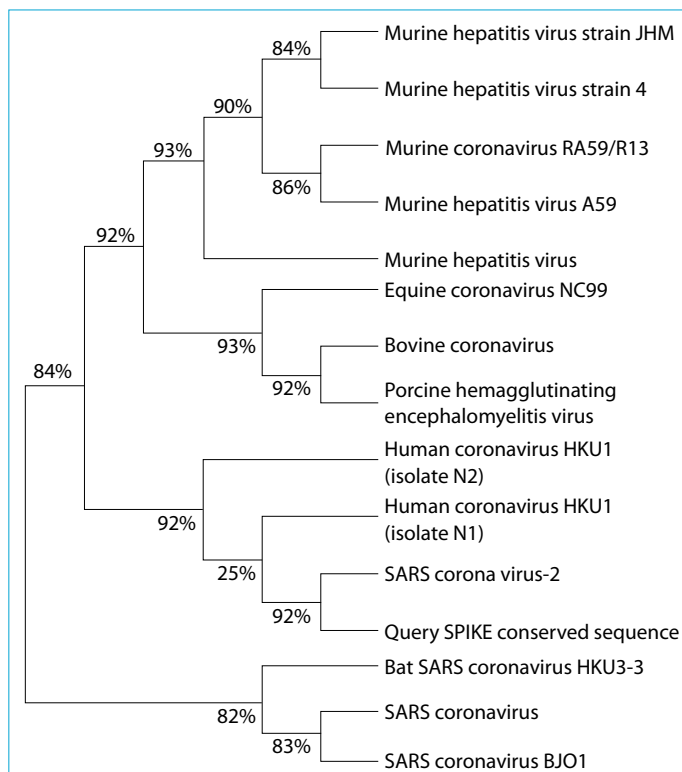
### In Silico Vector Design

To design an expression vector construct for its introduction into the mammalian cells, we downloaded the pMBP-LbCas12a H759A vector from Addgene ([www.addgene.org](http://www.addgene.org)) which possesses a T7 promoter and other viral promoters encoding both the sgRNA and the Cas12 protein. We used SnapGene (<https://www.snapgene.com>) software to construct the expression vector inserted with the selected sgRNA sequence.<sup>[13]</sup> We used SnapGene software to introduce the 20 nucleotide S1-RBD target sequence into the exon region was generated by CHOPCHOP. Using the "Edit sequence" option, the selected 20-mer was inserted adjacent to the T7 promoter site.

## Results

### Phylogenetic Tree Analysis

The conserved sequence of the S-gene was identified using the Multiple sequence alignment (MSA) tool, where the sequences of the S1-RBD coding region (22553 to 23093) of SARS-CoV-2 from different geographical locations were compared with other viruses. The phylogenetic tree cluster was the best fit to showing maximum percent similarity under a common human clade SARS-CoV-2 and distantly related to other groups. The S protein sequence of 15 different corona virus isolates such as SARS-COV, Bat coronavirus, murine corona virus, murine hepatitis virus, and bovine coronavirus spike proteins were aligned using the muscle program in MEGA software and the sequence identities obtained for the full-length shown on the S protein diagram. A Maximum-Likelihood tree was generated based on the alignment (Fig. 1). In addition, the S-gene sequence was used to perform CHOPCHOP analysis.



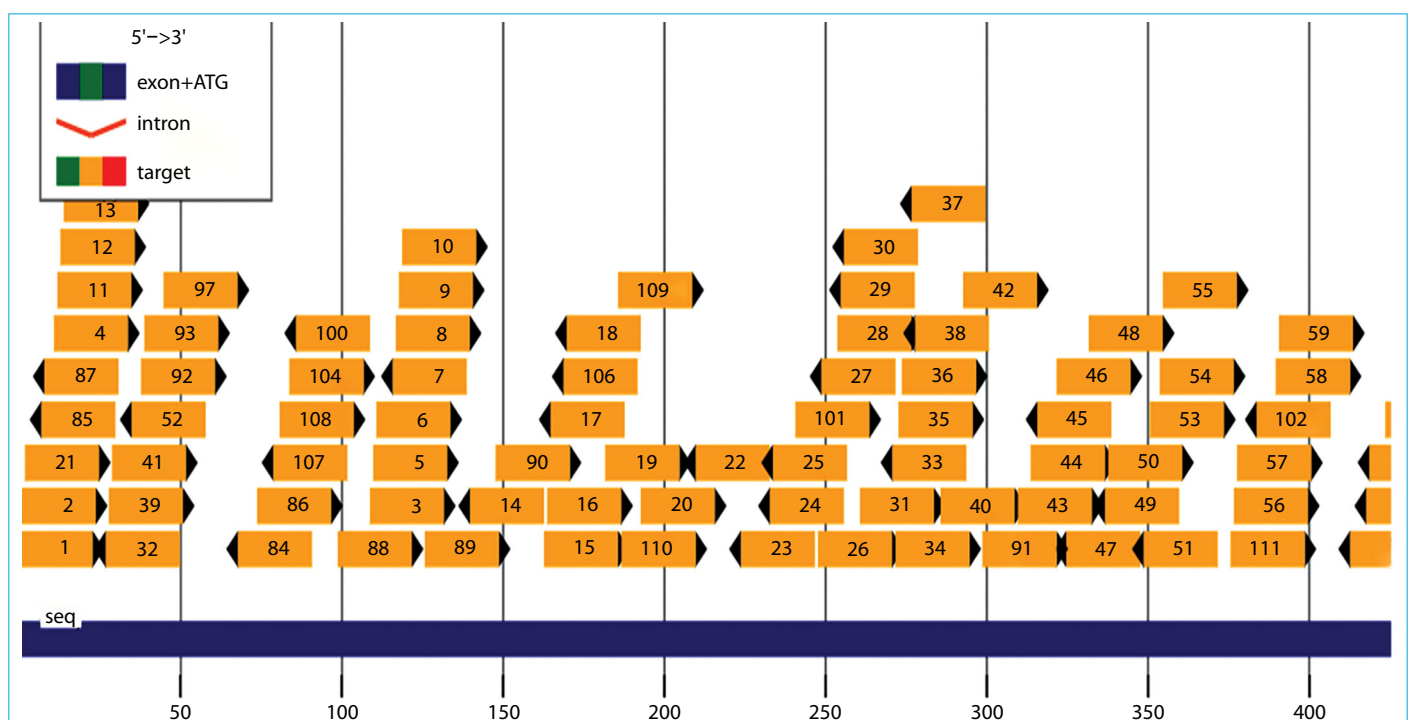
**Figure 1.** Phylogenetic tree construct showing the relation among the selective spike glycoprotein gene sequences of different corona virus species. Maximum-likelihood phylogenetic tree of spike protein of SARS-CoV-2. Percentage next to the branching points indicate the relative support from bootstrap (500 times).

## Target Sequence Identification

We used CHOPCHOP (<https://chopchop.rc.fas.harvard.edu>) to identify the protospacer adjacent motif (PAM) sequence from the query RBD gene sequence of SARS-CoV-2 in order to design sgRNA. CHOPCHOP generated 110 target sequences based on PAM provided within the exon region of the S-gene that was obtained from the reference genome (Wuhan coronavirus) and selected the 20 nucleotide target sequence (Fig. 2). The CHOPCHOP tool ranked the sequences by many in-built parameters, including GC content and off-targets. Besides, the CHOPCHOP tool provided target-specific primers for S-gene sequence amplification and restriction site analysis. The twenty nucleotide long sequence 5'ACCATACAGAGTAGTAGTAC3' showed 100% similarity with SARS-CoV-2 and demonstrated <93% similarity with SARS-CoV and other closely related species (data not shown). The target sequence is (highlighted in green; Figure 3), transcription of T7 promoter site sequence (shown in red) and the overlapping sequence was added as the remaining sequence to design guide RNA shown in blue (Fig. 3).

## In silico Expression Vector Design

Target-specific oligonucleotide sgRNA template was generated with an extra nucleotide 'G' added to the 5' end using NEBicalculator EnGen sgRNA Template Oligo design-



**Figure 2.** CHOPCHOP analysis for RBD gene coding sequences of SARS-CoV-2. The workflow of CHOPCHOP in Cas12 nickase mode which allows DNA sequence and genomic coordinates. Cas 12a recognizes a 5'-TTN-3' PAM site. The target color indicates the quality of each sgRNA starting from best (green) to worst (red) and sgRNA represented with primers and restriction sites.

Target-specific DNA oligo template (5'→3'):

TTCTAATACGACTCACTATA GACCATACAGAGTAGTAGTACGTTTTAGAGCTAGA

Transcribed sgRNA sequence (5'→3'):

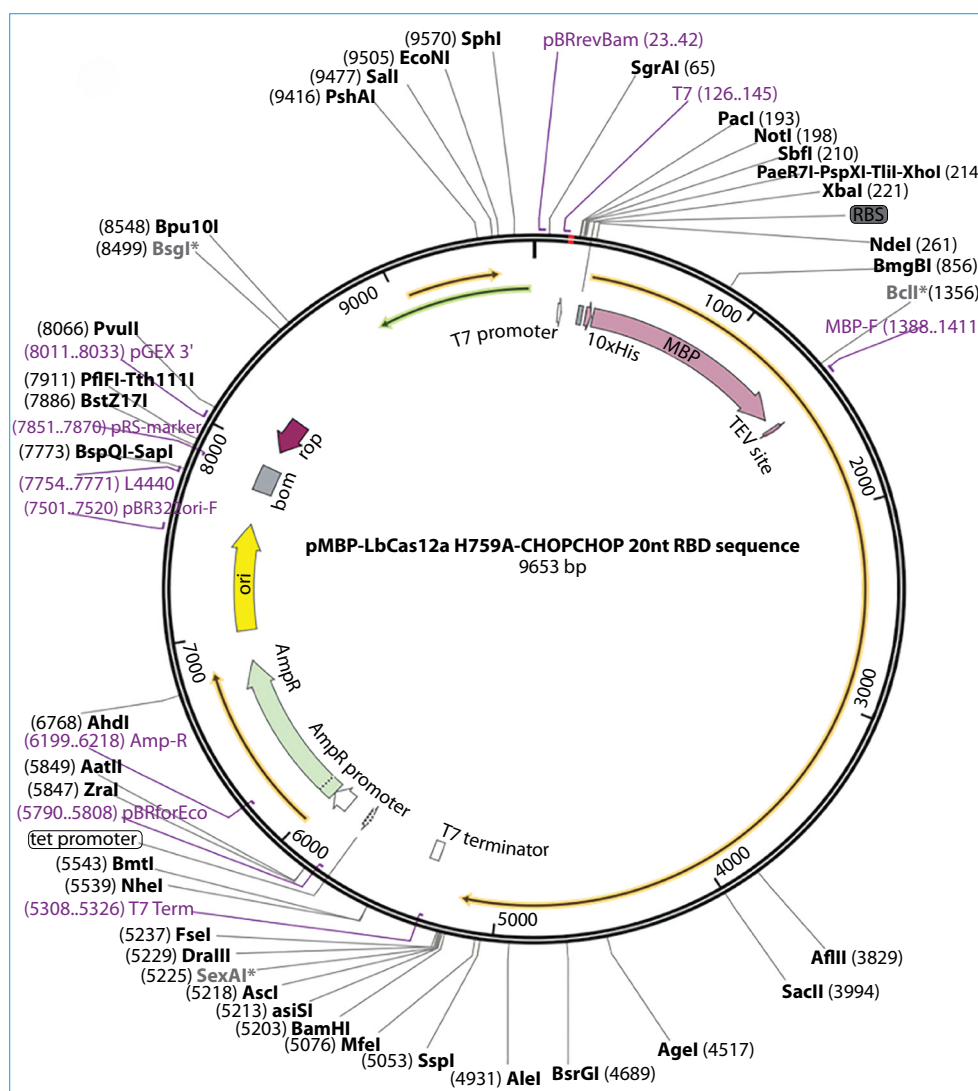
GACCAUACAGAGUAGUAGUACGUUUUAGAGCUAGAAAUAGCAAGUUAAAAU

**Figure 3.** Design of sgRNA using NEBioCalculator. The target-specific oligo contains the T7 promoter sequence, ~20 nucleotides of target-specific sequence and a 14 nucleotide overlap sequence complementary to the Cas12 Scaffold. Two oligos anneal at the 14 nucleotide overlap region of complementarity. DNA polymerase makes a double-stranded DNA template and RNA polymerase T7 promoter (shown in red) and initiates transcription. The sgRNA contains the target-specific sequence (shown in green) as well as the tracrRNA. In order for transcription to occur, a promoter sequence is required and overlapping sequence shown in blue color.

er (<https://nebiocalculator.neb.com/#/sgrna>). The sgRNA template was attached to the T7 promoter site for transcription. The NEBioCalculator also provided the overlapping sequence attached to the designed guide RNA (Fig. 3). The generated sequence *in silico* has been used to synthesize the ssDNA oligonucleotide primers to develop the *in vitro* transcription kit (Fig. 4).

## Discussion

CRISPR genome editing consists of various CRISPR-Cas complex systems with a variety of proteins and enzymes. The computational *in silico* method of designing guide RNA serves as one of the most important steps in targeting the novel coronavirus. The results showed that the conserved sequence was within the region of the RBD domain and the selected target sequence has been used to design CRISPR-Cas detection-based kits for the SARS-Cov-2. Once



**Figure 4.** *In silico* expression vector construct (pMBP-LbCas12a H759A plasmid) using SnapGene.

the virus is detected by the designed sgRNA construct, the target region would be subjected to cleavage by the Cas 12 enzyme expressed by the pMBP-LbCas12a H759A plasmid construct and other collateral cleavage activity would be recognized using fluorescent quencher. The specificity of the designed guide RNA distinguishes the novel coronavirus from other closely related species, even with a single nucleotide difference. *In vivo* and *in vitro* validation methods should be used to test the accuracy of the designed novel coronavirus detection kit using *in silico* CRISPR-Cas system tools. The currently developed STOPCovid nucleic acid-based test kit could easily be performed by the user using a simple test strip. Currently, there is great competition and significant progress among several Bio-science companies in the development of home-use viral detection test kits. A simple cartridge microfluidic device-based CRISPR-Cas systems may further simplify the currently employed diagnostic method in the near future.

#### Disclosures

**Acknowledgements:** The authors acknowledge VFSTR (Deemed to be university) and DST-FIST (LSI- 576/2013) for providing the networking facility that was crucial to carry out this work.

**Ethics Committee Approval:** Not applicable.

**Peer-review:** Externally peer-reviewed.

**Conflict of Interest:** There is no conflict of interest.

**Authorship Contributions:** Concept – A.P.K.; Design – A.P.K., V.T.C., V.S.A.; Supervision – V.T.C., K.S.; Data collection &/or processing – A.P.K., V.T.C., V.S.A.; Analysis and/or interpretation – A.P.K., V.T.C., V.S.A., I.M., K.S.; Literature search – A.P.K., R.R.A., A.G.; Writing – A.P.K., V.T.C., K.S., D.C.N., I.M., V.S.A., R.R.A., Y.N.B., A.G.; Critical review – D.C.N., Y.N.B.

#### References

- Ishino Y, Krupovic M, Forterre P. History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology. *J Bacteriol* 2018;200:e00580–17.
- Nasko DJ, Ferrell BD, Moore RM, Bhavsar JD, Polson SW, Womack KE. CRISPR Spacers Indicate Preferential Matching of Specific Viroplankton Genes. *mBio* 2019;10:e02651–18.
- Jiang F, Doudna JA. CRISPR-Cas9 Structures and Mechanisms. *Annu Rev Biophys* 2017;46:505–29.
- Zhu Y, Klompe SE, Vlot M, van der Oost J, Staals RHJ. Shooting the messenger: RNA-targeting CRISPR-Cas systems. *Biosci Rep* 2018;38:BSR20170788.
- Strich JR, Chertow DS. CRISPR-Cas Biology and Its Application to Infectious Diseases. *J Clin Microbiol* 2019;57:e01307–18.
- Tang Y, Fu Y. Class 2 CRISPR/Cas: an expanding biotechnology toolbox for and beyond genome editing. *Cell Biosci* 2018;8:59.
- Kellner MJ, Koob JG, Gootenberg JS, Abudayyeh OO, Zhang F. SHERLOCK: nucleic acid detection with CRISPR nucleases. *Nat Protoc* 2019;14:2986–3012.
- Yuen KS, Ye ZW, Fung SY, Chan CP, Jin DY. SARS-CoV-2 and COVID-19: The most important research questions. *Cell Biosci* 2020;10:40.
- Udugama B, Kadhiresan P, Kozlowski HN, Malekjahani A, Osborne M, Li VYC, et al. Diagnosing COVID-19: The Disease and Tools for Detection. *ACS Nano* 2020;14:3822–35.
- Liu G, Zhang Y, Zhang T. Computational approaches for effective CRISPR guide RNA design and evaluation. *Comput Struct Biotechnol J* 2019;18:35–44.
- Jaimes JA, André NM, Chappie JS, Millet JK, Whittaker GR. Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop. *J Mol Biol* 2020;432:3309–25.
- Karlapudi AP, T C V, Tammineedi J, Srirama K, Kanumuri L, Prabhakar Kodali V. In silico sgRNA tool design for CRISPR control of quorum sensing in *Acinetobacter* species. *Genes Dis* 2018;5:123–9.
- Abraham Peele K, Srihansa T, Krupanidhi S, Ayyagari VS, Venkateswarulu TC. Design of multi-epitope vaccine candidate against SARS-CoV-2: a in-silico study. *J Biomol Struct Dyn*. 2020 Jun 1:1-9. doi: 10.1080/07391102.2020.1770127. [Epub ahead of print].