Webology, Volume 18, Number 1, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I1/202

In-silico Designing and Evaluation of gRNA for the CRISPR/Cas9 system against the Genes associated with Tuberculosis

[1]Shifa Ali Zaidi, [2]Devvret Verma, [1]Anshul Kamboj, [1]Saurav Chandra Bhatt, 2Kumud Pant*, [3]Shreesh Gupta

[1]Department of Life Sciences, Graphic Era Deemed to be University, Clement Town, Dehradun, Uttarakhand, India 248002
[2]Department of Biotechnology, Graphic Era Deemed to be University, Clement Town, Dehradun, Uttarakhand, India 248002
[3]Wecapsulate Pharma, Lane no.8, Kandoli, Rajpur Road, Dehradun, Uttarakhand, India 248001

ABSTRACT

Tuberculosis caused by the bacterium Mycobacterium tuberculosis has been seen to show resistance against the prescribed drugs which are of uttermost importance and act as a first line of defense. The emergence of drug resistance has become a serious concern for the society and need for new and effective armamentarium persists. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR-associated nuclease 9 (Crispr Cas9) is like a boon for the genetic engineering and have a wide application in gene editing. In the Mycobacterium tuberculosis, the gene Rv3378c is responsible for the production of a protein "1-TbAd" which helps the bacterium to survive inside the host for a period. The alteration in the gene Rv3378c can be potential therapeutics for the treatment of Tuberculosis. In the present study, a guide RNA (gRNA) was designed using CRISPOR tool to perform the necessary gene editing in the bacterium. The most suitable gRNA was selected with zero off targets. Zero off targets are necessary to decrease the chances of mistake in cutting the target sequence. This technique could serve as an important armamentarium to serve and save the people suffering from this disease.

KEYWORDS- CRISPR, Tuberculosis, PAM sequences, Guide RNA, Zero-off target, NCBI, Mycobacterium tuberculosis, Rifampicin, Isoniazid, Ethambutol, Pyrazinamide.

INTRODUCTION

Mycobacterium tuberculosis is a species of a pathogenic strain in bacterial family which is a causative agent of tuberculosis (TB). It is a communicable and deadly disease which affects a healthy individual when a person suffering from tuberculosis ejects the pathogen either by sneezing or coughing or sometimes by just even talking. It has existed for epochs and remains a major delinquent to date. It affects 10 million people each year and is a major cause of demise around the sphere, ranking above HIV/AIDS making it even worse. It can affect people in two ways-Either by

Webology, Volume 18, Number 1, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I1/202

infecting the lungs (pulmonary infection) or by infecting any other body part (ex-pulmonary infection) example Spine.

The chances of being infected by TB is much higher in the people suffering from HIV or people who are affected by high-risk factors such as diabetes, smoking, alcohol consumption. To cure this ailment the effective medication started in the 1940s. According to the report of WHO the currently endorsed treatment for drug-susceptible cases of Tuberculosis, a 6-month administration of four major drugs is available as first line of defense. The drugs are namely-isoniazid, rifampicin, ethambutol, and pyrazinamide. Total 194 member states of WHO regularly report to the organization about the success rate of the treatments regarding the drug-susceptible cases of TB, reporting to about at least 85%. Medication pertaining to different types of resistant TB namely-Rifampicin Resistant (RR-TB) and multi-drug resistant (MDR-TB) tuberculosis is highly expensive and prolonged . In accordance with WHO, a recent report suggested that at least 123 countries have reported the XDR-TB case which implies that the TB shows resistance to the four main anti - tuberculosis drug, involving resistance to two main tuberculosis drugs isoniazid and rifampicin [1].
The bacterium *Mycobacterium tuberculosis* chiefly affects the lungs i.e. it is majorly a mammalian respiratory system pathogen. Its genome was first sequenced in 1998 [2][3] and is understood as a complex containing 9 members, namely-: *M tuberculosis sensustricto*, *M. africanum*, *M. canetti*, *M. bovis*, *M. caprae*, *M. microti*, *M. pinnipedii*, *M. mungi*, and *M. orygis* [4]. The unusual cell wall of the bacterium *Mycobacterium tuberculosis* is a lipid rich cell wall basically mycolic acid and is an active virulence factor which is responsible for conferring resistance to the bacteria against desiccation. The bacterium carries Ure C gene responsible for the prevention of acidification of the phagosome [5]. The bacterium also produces a diterpene molecule is known as "isotuberculosinol" also called as nosyberkol or edaxadiene, which aids in its pathogenesis. It functions by preventing the maturation of the host cell phagosome in which the bacterium resides [6]. The bacterium also avoids the assassination of the macrophages by a neutralization reaction in which the intermediates of reactive nitrogen are neutralized [7]. A new and recent study suggests that the bacterium *Mycobacterium tuberculosis* covers itself in an unusual nucleoside which acts as an antacid and permits the bacterium to induce swelling in the lysosome and neutralizes the p^H, called as1-Tb-Ad (1-tuberculosiyladenosine). 1-TbAd is coded by the gene "Rv3378c" [8][9]. A screen for the 1-TbAd mutant, gene transfer, and complementation studies suggest that 1-TbAd is synthesized by the gene Rv3378c. The role of this gene was earlier believed to purpose as a phosphate, but later studies suggest its part as tuberculosinyl transferase and put forward the amended biosynthetic pathway for sequential action of the gene. In accordance with the model the recombinant rv3378c protein produced 1-TbAd. The crystal assembly of this protein also exposed a fold namely cis- prenyl transferase fold with hydrophobic residues for isoprenoid binding and a second binding pocket is also present which is suitable for the nucleoside substrate. This dual substrate pocket differentiates the standard cis-prenyl transferase from Rv3378c. This provides a unique model for the prenylation of diverse metabolites. Since the Terpene nucleosides are very rare in nature, similarly 1-TbAd is only known in Tuberculosis bacterium. Therefore, the juncture of these two (nucleosides and terpene) pathways arose late in the development of MTB complex. 1-TbAd now serves as an enough marker of MTB and the extracellular export of this molecule likely accounts for the virulence promoting effects of Rv3378c gene [10]. This gene plays a major role in shielding the bacteria and hence giving it property to subsist in the host. In the current study, the CRISPR-Cas9 system was

Webology, Volume 18, Number 1, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I1/202

used to design the guide RNA for the gene Rv3378c and hence empowering gene editing. This system is a very exciting development in molecular biology. CRISPR stands for "Clustered Regularly Interspaced Short Palindromic Repeats", which serves as a tool of the adaptive immune system along with the associate endonucleases (e.g. Cas9) in the archaea and bacteria providing them protection against the invading virus. It contains sequences of the genetic code which are repeated and are interrupted by "spacer sequences". These spacer sequences are thought of as the remains of the genetic code of the past invaders. It works by cutting at the DNA of the invader at specific sequences [11]. It is a simpler technique as compared to the previously present techniques namely, Zinc Finger Nucleases (ZFNs) and Transcription Activator Like Effector Nucleases (TALENs). Both above-mentioned techniques require more efforts and are time consuming since they require custom proteins for each DNA target. All this is not required in the RNA programming for CRISPR- Cas9 system [12]. The role of Cas9 nuclease is to encourage the double strand breaks on the target position of the genome which is guided by the single-guide RNAs (sg-RNAs). Different studies have been put forward suggesting that this nuclease-based genome editing may introduce off-target effects that can result in significant level of non-specific editing at another untargeted genome locus [13]. However, recent studies depict that the off-target effects of this technology are not as vital as mentioned earlier, but still the effectiveness and the uniqueness of this Cas9-sgRNA should be carefully studied before using it *in vivo* [14]. The basic designing of a CRISPR Cas9 system contains two necessary RNA components namely, CRISPR RNA (cr-RNA) and Trans Activating cr-RNA (tr-RNA). Both complement each other to activate and guide the Cas9 nuclease [15]. Scientists have taken efforts and it was suggested that the production of sgRNA is functionally equivalent to the work done by the complex of cr-RNA and tr-RNA [16]. Designing a guide RNA in silico is a very crucial step in CRISPR Cas9 system and continuous efforts are being made to design a sg-RNA which has both high on-target efficiency and least off-target effects [17].

MATERIAL AND METHODS

Data procurement and analysis- The coding sequence (CDS) for the molecule 1-TbAd of the studied MTB strains were extracted from the nucleotide database available at the National Centre for Biotechnological Information (NCBI) (https://www.ncbi.nlm.nih.gov/) in the fasta format [18]. The accession numbers with the information of the gene and the CDS region coding 1-TbAd are shown in table 1.

TABLE-1: Data obtained from NCBI for the Target gene.

S.NO	GENE	ACCESSION NUMBER	CODING REGION
1.	Rv3378c	NC_000962.3	3792358-3793248

Multiple Sequence Alignment- The sequence alignment was done with the help of BLAST tool of NCBI ((https://blast.ncbi.nlm.nih.gov/Blast.cgi). Using nucleotide BLAST, the sequence was checked for its specificity in the genome and then Global Alignment was done at the same site comparing *Mycobacterium tuberculosis* genome with that of the human (human genome taxid:9605) and results were checked and compared [19].

Designing guide RNA- Now when the gene of interest has been identified designing a guide RNA is the major task that was done using the CRISPOR tool which can be accessed using the link http://crispor.tefor.net/. CRISPOR uses the BWA sequence search algorithm [20]. This website helped to select and express the CRISPR guide sequences. DNA sequence were used as input and

Webology, Volume 18, Number 1, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I1/202

genome were selected. The tool lists the guide sequences with more information found in the database and algorithms which includes genome variants, off-target and on-target scores. Primers are also designed for each guide sequence so that it can be amplified if required and produce guide

10	**GUIDE SEQUENCE**	**MIT SPECIFICITY SCORE**	**CFD SPECIFICITY SCORE**	**PREDICTED EFFICIENCY**		**OUTCOME**		OFF-TARGETS FOR 0-1-2-3-4 MISMATCHES
				DOENCH'16	MOR.-MATEOS	OUT OF FRAME	LINDEL	
1.29/forward	GGTTAGCGAAAAAGAATTCT *TGG*	100	100	35	11	62	74	0-0-0-0-0
2.72/forward	GCAGAGATTGTACGTTGCCG *TGG*	100	100	61	38	71	65	0-0-0-0-0
3.80/forward	TGTACGTTGCCGTG	100	100	59	45	75	90	0-0-0-0-0

RNA sequences in vitro by transcription after annealing of overlapping primers or for cloning into plasmids. The tool also helps to find out the PAM sequences [21].
The CRISPOR tool directs to select the desired PAM sequence upstream and itself gives us the possible PAM sequences
Since the off-target and on targets are also identified by the tool the best guide RNA will be the one which has off-target score "0".

RESULT AND DISCUSSION

Tuberculosis has been giving mankind a tough competition whether how to contain it or whether regarding drugs. The different known drugs for this ailment have started to show resistant which is a major cause of concern. The objective of the study was to find an alternative regarding the resistant Tuberculosis. The study was conducted on the gene Rv3378c which codes for the protein 1-TbAd in *Mycobacterium tuberculosis.* The exceptionality of Rv3378c in the *Mycobacterium tuberculosis* makes it a potential target for gene editing. The finding also suggests that the gene codes for the protein (1-TbAd) only in *Mycobacterium tuberculosis* and helps it to remain untargeted by the immune system of the

Webology, Volume 18, Number 1, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I1/202

	GACCGA *AGG*							
4.102/forward	GTATCCGTCTTTCC CTTCGA *TGG*	100	100	51	39	75	79	0-0-0-0-0
5.115/forward	CCTTCGATGGCAC ACGCCGC *TGG*	100	100	36	42	79	43	0-0-0-0-0

host and thus making it more detrimental for the host. The result of the BLAST confirmed its existence only in the bacteria.

The BLAST result depicted that the gene sequence showed 100% query cover which means that it was preserved only in the *Mycobacterium tuberculosis.* The BLAST was again performed with excluding the organism human (taxid:9605) and no significant similarity was found in the result. Therefore, the coding region of this gene was taken as a target sequence and further designing of the guide RNA was done. The CRISPOR tool was used for designing the guide RNA. Here the PAM sequences were selected, and the analysis were performed (Table 2).

The PAM sequence which was selected was -20bp-NGG-Sp Cas9, SpCas9-HF-1, eSpCas9 (the results are depicted in the figures 1-7).

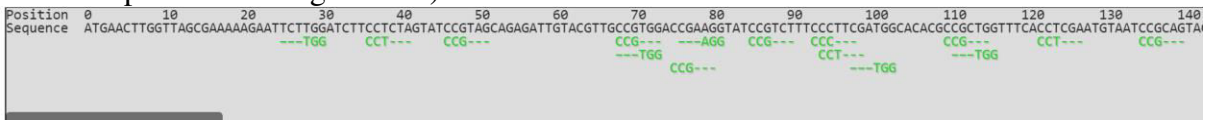

FIGURE-1: The green color depicts the possible PAM sequences from position 0-140.

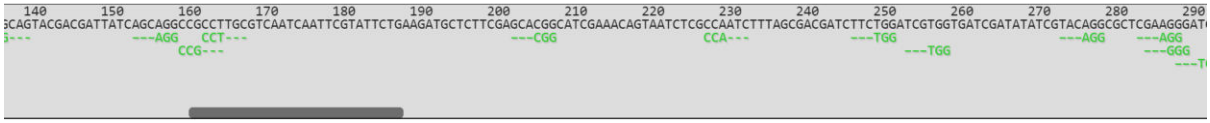

FIGURE-2: The green color depicts the possible PAM sequences from position 140-290.

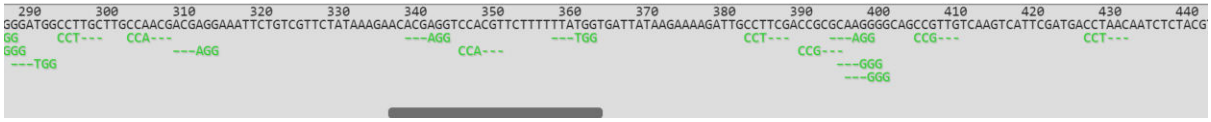

FIGURE-3: The green color depicts the possible PAM sequences from position 290-440.

FIGURE-4: The green color depicts the possible PAM sequences from position 440-590.

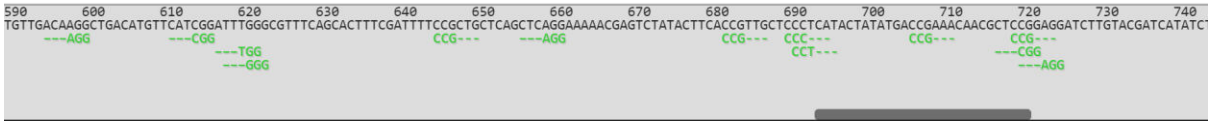

FIGURE-5: The green color depicts the possible PAM sequences from position 590-740.

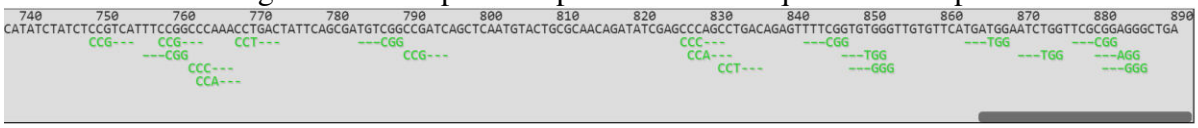

FIGURE-6: The green color depicts the possible PAM sequences from position 740-890.

The input sequences were 891 bp long and contained 83 possible guide sequences. The PAM sites

Webology, Volume 18, Number 1, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I1/202

and the expected cleavage position were located at -3bp 5' of the PAM site. Colors green, yellow and red indicated high, medium, and low specificity of the PAM's guide sequence in the genome. The tool allowed to show the 20bp long guide sequence by clicking the match for the PAM mentioned below, when opened in the browser.
The selected guide sequences had-

- Specificity score in the range of 0-100. This predicted the off-target cleavage by RNA guide sequence elsewhere in the genome. It was calculated using the formula from MIT CRISPR website. The best score chosen was 100.
- Efficiency score in the range of 0-100. This predicted the efficiency of target being cut by RNA guide sequence. The tool used two scores, Doench 2016 score and Moreno Mateos 2016.
- "Zero" off-target score. The range for this was 0-100. The off-target mismatch counts all the off-target possible for each number of mismatches in the genome. The ZERO score is considered as the best.
- Out of frame scores. This predicted the out of frame deletions which are led by the guide RNA.

Protospacer adjacent motifs are abbreviated as PAM sequences. These are 2-6 base-pair long DNA sequences. They immediately follow the DNA sequence targeted by the Cas9 nuclease in the adaptive immune system of the bacteria i.e. CRISPR [22]. It is not a component of the bacterial CRISPR locus because it is not found in the bacterial host genome rather it is a part of the invading virus or plasmid. Until and unless the PAM sequence is present the Cas9 system will not cleave the target [23][24][25].Therefore, the PAM is an important component for determining the self and non-self-DNA and preventing the CRISPR locus from being targeted and being destroyed by the CRISPR associated nuclease [26]. The official PAM sequence is 5'-NGG-3', where "N" is any nucleobase followed by two guanines "G" nucleobase [27]. The guide RNAs can transfer Cas9 to any location for gene editing in the genome but no editing can occur at any site other than that identified by the Cas9 PAM sequences.
While designing guide RNA which targets the DNA region of interest the following were taken care of (i.) The target sequence of the gene. (ii.) The type of Cas9 nuclease which can be used and the PAM sequences to be recognized. (iii) Which promoter is to be used for the in vivo and in vitro expression of the guide RNA sequence and to exclude the terminator sequence from guide-RNA sequences? (iv.) The cloning strategy is also to be determined. (v.) The guide RNA sequence can bind to both strands i.e. positive and negative strands.

Earlier studies show that CRISPR has been used a diagnostic tool to for tuberculosis, which detects the MTB complex and requires less sample input [28]. Not much has been worked on CRISPR regarding Tuberculosis which puts forward more hope for studies and recognition for this paper.

CONCLUSION

Gene editing with CRISPR can be considered as a boon to mankind as it can solve major health problems. It simply needs a guide RNA sequence which can guide the Cas enzyme associated with the target sequence to go and perform the function with great accuracy and effectiveness. In this paper, a constant approach was made to design a guide RNA for the target sequence which codes for 1-TbAd in the bacteria *Mycobacterium tuberculosis.* It can have possible positive effects for the

Webology, Volume 18, Number 1, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I1/202

people suffering from Multi-Drug Resistance Tuberculosis as it has some hidden challenges for the scientist to make and understand as to how the working of *Mycobacterium tuberculosis* is affected by the hindrance of this protein and can be likely a source of new drug formulation regarding tuberculosis. This strategy can be a novel therapeutic at the genomic level to save mankind regarding this disease. But a challenge still awaits, as to how this gene editing system can be introduced into the human body and reach the goals (this technique has not been considered here in this paper).

ACKNOWLEDGEMENT- The authors express their heartfelt gratitude to the Department of Life Sciences, as they provided the best environment to work in with all the facilities that were needed to be accessed. Heartfelt gratitude to the co-workers who shared their work experience and guided us with the work.

REFRENCES

1. World Health Organization. (2019). Global tuberculosis report 2019. World Health Organization
2. Cole, S., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D et.al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 1998;*393*(6685), 537-544.
3. Camus, J. C., Pryor, M. J., Médigue, C., & Cole, S. T. Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv. Microbiology, 2002; *148*(10), 2967-2973.
4. van Ingen, J., Rahim, Z., Mulder, A., Boeree, M. J., Simeone, R., Brosch, et.al. Characterization of Mycobacterium orygis as M. tuberculosis complex subspecies. Emerging infectious diseases, 2012; *18*(4), 653.
5. Bell, E. A souped-up version of BCG. Nature Reviews Immunology, 2005 ;*5*(10), 746-746.
6. Mann, F. M., Xu, M., Chen, X., Fulton, D. B., Russell, D. G., & Peters, R. J. Edaxadiene: a new bioactive diterpene from Mycobacterium tuberculosis. **J Am Chem Soc**, 2009;*131*(48), 17526-17527.
7. Flynn, J. L., & Chan, J. Immune evasion by Mycobacterium tuberculosis: living with the enemy. Current opinion in immunology, 2003;*15*(4), 450-455.
8. Buter, J., Cheng, T. Y., Ghanem, M., Grootemaat, A. E., Raman, S., Feng, X., et.al. Mycobacterium tuberculosis releases an antacid that remodels phagosomes. Nature chemical biology, 2019;*15*(9), 889-899.
9. Brodin, P., & Hoffmann, E. T (oo) bAd. Nature chemical biology, 2019; *15*(9), 849-850.
10. Layre, E., Lee, H. J., Young, D. C., Martinot, A. J., Buter, J., Minnaard et.al. Molecular profiling of Mycobacterium tuberculosis identifies tuberculosinyl nucleoside products of the virulence-associated enzyme Rv3378c. Proceedings of the National Academy of Sciences, 2014; *111*(8), 2978-2983.
11. Fusi, N., Smith, I., Doench, J., & Listgarten, J. In silico predictive modeling of CRISPR/Cas9 guide efficiency. BioRxiv, 2015; 021568.
12. Urnov, F. D., Miller, J. C., Lee, Y. L., Beausejour, C. M., Rock, J. M., Augustus, S., et.al. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature, 2005; *435*(7042), 646-651.

Webology, Volume 18, Number 1, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I1/202

13. Mussolino, C., Morbitzer, R., Lütge, F., Dannemann, N., Lahaye, T., & Cathomen, T. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic acids research, 2011; *39*(21), 9283-9293.
14. Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., et.al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nature biotechnology, 2013; *31*(9), 827.
15. Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., et.al. Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. Nature biotechnology, 2014; *32*(12), 1262.
16. Chuai, G. H., Wang, Q. L., & Liu, Q. In silico meets in vivo: towards computational CRISPR-based sgRNA design. Trends in biotechnology, 2017;*35*(1), 12-21.
17. Jaskula-Ranga, V., & Zack, D. J. grID: A CRISPR-Cas9 guide RNA database and resource for genome-editing. bioRxiv,2016;097352.
18. NCBI Resource Coordinators, Database Resources of the National Center for Biotechnology Information, *Nucleic Acids Research,* Volume 45, Issue D1, Pages D12–D17, January 2017.
19. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. Basic local alignment search tool. Journal of molecular biology, 1990; *215*(3), 403-410.
20. Li, H., & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. bioinformatics, 2009;*25*(14), 1754-1760.
21. Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J. B. et.al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome biology, 2016;*17*(1), 148.
22. Shah, S. A., Erdmann, S., Mojica, F. J., & Garrett, R. A. Protospacer recognition motifs: mixed identities and functional diversity. RNA biology, 2013; *10*(5), 891-899.
23. Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J., & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology, 2009; *155*(3), 733-740.
24. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. science, 2012;*337*(6096), 816-821.
25. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature, 2014;*507*(7490), 62-67.
26. Mali, P., Esvelt, K. M., & Church, G. M. Cas9 as a versatile tool for engineering biology. Nature methods, 2013;*10*(10), 957-963.
27. Anders, C., Niewoehner, O., Duerst, A., & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature, 2014; *513*(7519), 569-573.
28. Ai, J. W., Zhou, X., Xu, T., Yang, M., Chen, Y., He, G. Q., et.al. CRISPR-based rapid and ultra-sensitive diagnostic test for Mycobacterium tuberculosis. Emerging microbes & infections, 2019;*8*(1), 1361-1369.