COMPUTATIONAL DETERMINATION OF POTENTIAL TALEN TARGETS IN HERPES SIMPLEX VIRUS 2 GENOME

W.G.B.P. Dharmasena[1] and D.H.H. Munasinghe[2]
[1, 2] Department of Botany, University of Sri Jayewardenepura, Sri Lanka

buddhikapolgahawela@gmail.com, helani@sci.sjp.ac.lk

ABSTRACT: *Human alphaherpesvirus 2, also known as Herpes simplex virus 2 (HSV-2), causes genital herpes and is one of the viruses that lead to chronic infections in human. Several complications occur due to this viral infection and proper medication is still not available to cure the disease. However, the modern gene editing technique, Transcription Activator Like Effector Nuclease (TALEN), has a potential to use as an alternative strategy to eradicate the disease. In this study, the potential TALEN target sites present in the genome of HSV-2 were identified using bioinformatic tools. The viral gene UL21, which is essential for the propagation of the virus, and the gene UL30 (DNA polymerase) which is essential for the viral multiplication, were selected to find targets for TALENs so that these genes can be altered to diminish the viral persistence and multiplication. We used the tool 'TAL effector nucleotide targeter 2.0' to identify the possible TALEN targets and some targets were selected based on binding efficiency and validated for the absence of off-target sites in human and murine genome and other locations of the same viral genome using the tools 'Paired target finder' and 'PROGNOS'. In addition, a rough prediction of off-targets in genomes of other organisms was performed by searching for local alignments of the TALEN target sequence using the tool 'nBLAST'. The TALEN sites with better binding affinity and null off-target effect were selected and the putative functions of the mutated protein were predicted so as to avoid the sites that lead to mutated proteins having an undesired function. Considering all these criteria, best scoring TALEN target sites were selected. The target sites selected for UL21 gene are, 5'T CACGGGCACCCGCGCCCCCA gaccgatggccgggaccg GGGCCGGGGGCGCGGGGGCC A3', 5'T GGACGCCCGCGACCGGCGC acggatgtcgtgatcacgg GCACCCGCGCCCCCAGACCG A3', 5'T CCGCCAGCGCGGCCTGCGG gacgtgcggcccgtggg GGAGGACGAGGTGTTCCTGG A3' and 5'T CGGGATTCTCAGCCGGGGG aaattgccaagtttg GCCTGGTGGTCCGGGGGAC A3'. The target sites selected for UL30 gene are 5'T CCACGACGGCCGCCTCCGGC gcgcccctaaggtgtact GCGGGGGGGACGAGCGCG A3', 5'T CGCCCCGCGTTCGCTGGACG aggacgcccccgcggagcag CGCACCGGGGTCCACG A3', 5'T GCGCGCCGCCCAGCTCCACG agcgatttatggacgccatc ACGCCCGCCGGGACCGTC A3', 5'T GGACGCCATCACGCCCGCCG ggaccgtcatcacgc TTCTGGGTCTGACCCCCGA A3' and 5'T GGTACCGCCTCAAGCCCG gccgcgggaacgcgc CGGCCCAACCGCGCCCCCCG A3'.These potential target sites could be used to construct site specific TALENs or to construct vectors possessing the gene for TALENs which can be used as therapeutic agents for the treatment of HSV-2 infection, but the on-target and off-target effects should be further assessed by in vitro and in vivo studies. These target sites may not be universally applicable to all the strains of the virus and the off-targets may present in the genomes that are not available in the GenBank database.*

Keywords: TALEN, *Human alphaherpesvirus 2*, UL21, UL30, Bioinformatics

1. INTRODUCTION

Mutation by Transcription Activator Like Effector Nuclease (TALEN) is a modern method of gene editing which uses site specific nuclease to induce double strand breaks and hence to induce mutations by non-homologous end joining repair. Therefore, TALENs have a potential to be used as an alternative method to cease the activity of human pathogens for which medications are not available to date. In this study, we computationally identified and validated the TALEN target sites for Herpes Simplex Virus 2 (*Human alphaherpesvirus 2*) hence the

ISBN 978-955-627-120-1

TALEN proteins specifically designed to bind with these targets can be used to alter the genome of Simplex Virus 2. To identify TALEN sites, we selected UL21 gene of the virus which is essential for the viral propagation as described by Le Sage *et al.* (2013) and moreover the gene UL30 or the DNA polymerase, which is essential for the multiplication of the virus (Knafels *et al.,* 2006). Mutations of these genes may alter the function of them and may diminish the viral activity and may reduce the viral population.

2. METHODOLOGY

The nucleotide sequences of the UL21 and UL30 genes of several strains of the virus were obtained from 'GenBank' database of National Center for Biotechnology Information(NCBI), USA. The accession numbers of the sequences were NC_001798.2, KF781518.1, KP334097.1, KP334096.1, Z86099.2, KU310668.1, KP334094.1, KP192856.1, KX574908.1 and KX574906.2 for UL21 gene and KF588390.1, KY363356.1, KY363355.1, KF588396.1, JX905317.1, JQ352199.1, JQ352188.1, KF588398.1, HQ123166.1, JN561323.2, KP334093.1, KP334093.1, KP334093.1, JX905316.1 and KY363358.1 for UL30 gene. These sequences of both genes were separately aligned using the tool 'UGENE' (Okonechnikov *et al.,* 2012) using the MUSCLE alignment method (Edgar, 2004). Then the conserved regions of the two genes were selected and these regions were used to identify TALEN target sites. This was performed by using the tool 'TALEN targeter' (Doyle *et al.*, 2012). Among the output obtained, several TALEN target sites having at least one unique restriction site at the spacer region and having high percentage of Repeat Variable Diresidues (RVDs) of HD or NH were selected. These selected TALEN target sites were then analyzed for the presence of potential off-targets of the respective TALEN proteins in human genome, murine genome and other loci of the same genome by using the tools 'Paired target finder' (Doyle *et al.,* 2012) and 'PROGNOS' (Fine *et al.*, 2013). Moreover, the off-targets in genomes of other organisms were searched using the tool 'nBLAST' (Altschul *et al.,* 1990). Finally, putative functions of the proteins formed after the mutation in different TALEN target sites were identified using the tools 'The sequence manipulation suite' (Stothard *et al.,* 2000), 'ORF finder' (Rombel *et al.,* 2002) and 'pBLAST' (Altschul *et al.,* 1990) which were used respectively to obtain the possible gene sequence after the mutation, to identify novel ORFs in mutated sequence and to identify the putative functions of the proteins coded by the novel ORFs. Considering all above criteria, we selected several TALEN sites which are free from off-target effects and absent of any novel putative function of the mutated protein.

3. RESULTS AND DISCUSSION

Among the TALEN targets selected from the results of the tool 'TALEN targeter', few targets were identified with the potential of mutating the UL21 and UL30 genes. The TALENs respect to these targets may posses higher binding specificity and higher binding efficiency due to the selection of targets with high percentage of RVD HD or NH in respective TALENs. The selection of RVD 'NH'

Kommentar [h1]: Write the names of the genes

ISBN 978-955-627-120-1

instead of 'NN' to bind with G enhance these effects. The selection of TALEN targets having at least one unique restriction site at the spacer region is essential in the practical identification of the DNA cleavage. The TALEN targets that are located in the initial part of the gene is prioritized hence it makes the mutated protein considerably deviated from the original. But the putative function of the mutated protein should be identified hence to avoid TALENs that may lead to the formation of mutated proteins with undesirable functions. The TALEN target sites selected for UL21 and UL30 genes are shown in table 1 and 2, respectively.

Table 1. Potential TALEN target sites for UL21 gene.

TALEN name	Plus strand target sequence	Cut site	TAL1 RVDs	TAL2 RVDs
HSV2UL21-01	T CACGGGCACCCGCGCCCCCA gaccgatggccgggaccg GGGCCGGGGGCGCGGGGGCC A	667	HD NI HD NH NH NH HD NI HD HD HD NH HD NH HD NH HD HD HD HD HD NI	NH NH HD HD HD HD HD NH HD NH HD HD HD HD HD NH NH HD HD HD
HSV2UL30-06	T GGACGCCCGCGACCGGCGC acggatgtcgtgatcacgg GCACCCGCGCCCCCAGACCG A	633	NH NH NI HD NH HD HD HD NH HD NH NI HD HD NH NH HD NH HD	HD NH NH NG HD NG NH NH NH NH NH HD NH HD NH NH NG NH HD
HSV2UL30-07	T CCGCCAGCGCGGCCTGCGG gacgtgcggcccgtggg GGAGGACGAGGTGTTCCTGG A	242	HD HD NH HD HD NI NH HD NH HD NH NH HD HD NG NH HD NH NH	HD HD NI NH NH NI NI HD NI HD HD NG HD NH NG HD HD NG HD HD
HSV2UL30-08	T CGGGATTCTCAGCCGGGGG aaattgccaagtttg GCCTGGTGGTCCGGGGGAC A	143	HD NH NH NH NI NG NG HD NG HD NI NH HD HD NH NH NH NH NH	NH NG HD HD HD HD HD NH NH NI HD HD HD NI HD HD NI NH NH HD

Table 2. Potential TALEN target sites for UL30 gene.

TALEN name	Plus strand target sequence	Cut site	TAL1 RVD	TAL2 RVD
HSV2UL30-02	T CCACGACGGCCGCCTCCGGC gcgcccctaaggtgtact GCGGGGGGGACGAGCGCG A	313	HD HD NI HD NH NI HD NH NH HD HD NH HD HD NG HD HD NH NH HD	HD NH HD NH HD NG HD NH NG HD HD HD HD HD HD HD NH HD
HSV2UL30-03	T CGCCCCGCGTTCGCTGGACG aggacgcccccgcggagcag CGCACCGGGGTCCACG	263	HD NH HD HD HD HD NH HD NH NG NG HD NH HD NG	HD NH NG NH NH NI HD HD HD HD NH NH NG

ISBN 978-955-627-120-1

	A		NH NH NI HD NH	NH HD NH
HSV2UL30-05	T GCGCGCCGCCCAGCTC CACG agcgatttatggacgccatc ACGCCCGCCGGGACCG TC A	518	NH HD NH HD NH HD HD NH HD HD NI NH HD NG HD HD NI HD NH	NH NI HD NH NH NG HD HD HD NH NH HD NH NH NH HD NH NG
HSV2UL30-06	T GGACGCCATCACGCCCG CCG ggaccgtcatcacgc TTCTGGGTCTGACCCCC GA A	545	NH NH NI HD NH HD HD NI NG HD NI HD NH HD HD HD NH HD HD NH	NG HD NH NH NH NH NH NG HD NI NH NI HD HD HD NI NH NI NI
HSV2UL30-07	T GGTACCGCCTCAAGCCC G gccgcgggaacgcgc CGGCCCAACCGCGCCC CCCG A	970	NH NH NG NI HD HD NH HD HD NG HD NI NI NH HD HD HD NH	HD NH NH NH NH NH NH HD NH HD NH NH NG NG NH NH NH HD HD NH

The target sequence is for the plus strand of the gene and the spacer region is represented in lowercase letters. The TAL1 RVDs are responsible for binding with plus strand target which is shown in uppercase letters prior to spacer region and the TAL2 RVDs are for the TALEN protein that bind with the minus strand target which is the complementary sequence of the nucleotides shown after the spacer region. The cut site is the location of the double strand break from the start of the gene.

These TALEN targets were designed based on the conserved residues of the genes. But these conserved residues may not be universally conserved among all the strains of the virus because the study was carried out only for some selected strains of the virus. In identification of potential off-target effects, pre-loaded human and mouse genomes were used. These sequences are consensus sequences and therefore it cannot be concluded that off-targets are absent in each and every human or mouse. To minimize this effect two tools were used as 'Paired target finder' and 'PROGNOS' in which the genome entries are different. Among these two tools PROGNOS tool display descriptive results irrespective of the presence or the absence of off-targets but Paired target finder gives results only if the off-targets are present. The off-target effect identification in murine genome is necessary because then the mouse can be used for *in vivo* testing. The identification of off-targets in genomes of other organisms is necessary to avoid mutations in them in an unexpected environmental release of the constructed TALENs and to avoid mutations in commensals of mouse and human. The nBLAST tool was used to identify the off-targets in genomes of other organisms, but this method was not accurate due to the presence of spacer region in the target that does not participate in specific binding. Also, the

ISBN 978-955-627-120-1

GenBank database does not contain all the genomes of organisms available in the environment which may posses off-targets. The mutation of the protein was considered in the two occasions of one base pair frame shifts and two base pair frame shifts. The identification of the putative function of the protein is based on the data available in the 'Protein' data base of NCBI and cannot exclude the possibility of having undesirable functions of proteins unavailable in the database.

4. CONCLUSION

The selected TALEN targets have a potential to be used in the mutagenesis of Herpes Simplex 2 viral genome and hence the pathogenic effect of the virus could be diminished but should be tested by laboratory experiments.

5. REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403-410.

Doyle, E. L., Booher, N. J., Standage, D. S., Voytas, D. F., Brendel, V. P., VanDyk, J. K., & Bogdanove, A. J. (2012). TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Research*, 40(Web Server issue), W117–W122.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics*, *5*(1), 113.

Fine, E. J., Cradick, T. J., Zhao, C. L., Lin, Y., & Bao, G. (2013). An online bioinformatics tool predicts zinc finger and TALE nuclease off-target cleavage. *Nucleic acids research*, *42*(6), e42-e42.

Le Sage, V., Jung, M., Alter, J. D., Wills, E. G., Johnston, S. M., Kawaguchi, Y., Banfield, B. W. (2013). The Herpes Simplex Virus 2 UL21 Protein Is Essential for Virus Propagation. *Journal of Virology*, *87*(10), 5904–5915.

Okonechnikov, K., Golosova, O., Fursov, M., & UGENE team. (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, *28*(8), 1166-1167.

Rombel, I. T., Sykes, K. F., Rayner, S., & Johnston, S. A. (2002). ORF-FINDER: a vector for high-throughput gene identification. *Gene*, *282*(1), 33-41.

Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques*, *28*(6), 1102-1104.

ISBN 978-955-627-120-1