JOURNAL OF
EMERGING INVESTIGATORS

Article

Designing gRNAs to reduce the expression of the *DMPK* gene in patients with classic myotonic dystrophy

Saketh Madhusudhan[1], Madhu Rajagopalan[1], Sharifa Sahai[2]

[1] Shrewsbury High School, Shrewsbury, Massachusetts

[2] Harvard University, Cambridge, Massachusetts

SUMMARY

Myotonic dystrophy is a condition that affects thousands globally. It presents as a wide variety of life-threatening conditions, ranging from cardiac conduction defects and myotonia, to diabetes mellitus and hypogonadism. Myotonic dystrophy is mainly caused by the presence of trinucleotide repeats in the myotonic dystrophy protein kinase (*DMPK*) gene, which controls the function of muscle cells, and it currently has no cure. As this condition has a genetic cause, we hypothesized that we could design guide RNA (gRNA) sequences—short RNAs that direct DNA editing—specific to the trinucleotide repeats of the *DMPK* gene that cause myotonic dystrophy. This was extremely important, as these sequences could be used to target these repeats and ultimately suppress the effects of myotonic dystrophy for patients. To accomplish this, we located the sequence of the *DMPK* gene and used it to develop gRNA sequences. To analyze these gRNAs, we generated scores for binding efficiency with CRISPRedict. The sequences demonstrated overall middle to high scores for overall efficiency category, an on-target category, and an off-target category. All of the gRNAs we generated were predicted to bind to the *DMPK* gene with a stronger affinity than a random sequence, but a lower affinity than a sequence known to target *DMPK*. The highest overall efficiency score of any of the generated gRNAs was 0.631, which was similar to the score of a sequence known to effectively reduce *DMPK* expression. These gRNAs have the potential to effectively target the repeats in *DMPK*. They may be promising therapeutics to reduce the symptoms caused by myotonic dystrophy, potentially improving the quality of life of millions with the disease.

INTRODUCTION

Myotonic dystrophy is a condition that has affected and continues to affect thousands of people worldwide. It impacts 1 out of every 8,000 people globally, making myotonic dystrophy the most common type of muscular dystrophy, especially in adults (1). This disease has several detrimental effects, including progressive weakness of many muscles, such as in the arms and legs, as well as hypogonadism and diabetes (1). To diagnose myotonic dystrophy, many doctors look for common symptoms. Some alarming signals include eye issues, such as cataracts, and heart problems, including conduction defects (1,4). Furthermore, many doctors will run laboratory studies such as electrodiagnostic testing and muscle biopsy, to further confirm a diagnosis of myotonic dystrophy (14).

The cause of this disease is found in the genome, specifically a CTG repeat in the untranslated region of the myotonic dystrophy protein kinase (*DMPK*) gene, a gene that controls muscle contraction and relaxation (2). The *DMPK* gene, located on Chromosome 19 is crucial in muscle, heart, and brain cells because of its ability to communicate with other cells (2). Myotonic dystrophy type 1 is caused by a large region of repeating trinucleotide repeats in the introns of the *DMPK* gene (4). The introns of genes have a large impact on post-transcriptional regulation of gene expression, including mRNA localization, stability, and translation (3). In healthy humans, there are about 5–34 CTG repeats in the *DMPK* introns, but patients with myotonic dystrophy usually have more than 50, with some patients even having more than 1,000 repeats (4). These repeats create an unstable region in the genome, leading to defects in the function of the protein and cells, such as muscle cells (5).

Luckily, scientists can combat this condition through the power of genetic editing technologies, specifically the CRISPR-Cas9 system (6). Generally, the CRISPR-Cas9 system is made up of two parts. The first part of the CRISPR-Cas9 system is the Cas9 enzyme, which cuts the DNA at the target location (6). The next part of this genetic editing system is the guide RNA (gRNA), an RNA sequence that binds to the target DNA sequence to guide the Cas9 enzyme (6). Naturally, genome editing is currently of great interest in the prevention and treatment of human diseases, and it has been applied to address many conditions, ranging from Alzheimer's to cystic fibrosis (7).

gRNA plays a major role in the CRISPR-Cas9 system, as it shows the Cas9 enzyme where to cut in the genome of an organism (7). When designing a gRNA, some variables need to be considered, including the target region/gene, the version of Cas-9 enzyme, and the protospacer adjacent motif (PAM) sequences that need to be recognized (8). PAMs are 3 to 4 nucleotide DNA sequences that are found near the target sequence, and they are essential in this process, as they let the Cas9 nuclease know where to cut the DNA (4). Additionally, there are other factors to consider when approaching the process of designing gRNAs. gRNA sequences need to target a region that is specific to the gene of interest (8). The gRNA also needs to be able to target a region that is found in the gene of interest, such as exons, introns, and untranslated regions (UTRs) (8). In addition, researchers need to be able to suppress off-target effects while also considering the likelihood of creating other mutations in the gene of interest, such as frameshift mutations (8).

https://doi.org/10.59720/24-129

There are many rules and guidelines that can aid in increasing the efficiency of the gRNAs while also avoiding negative outcomes (8). For example, when designing gRNA sequences, it is important to select target sites that are crucial for protein function (8). There are also some stability rules that can aid in making gRNAs more effective, including the guanine and cytosine content of the gRNA being between 40 to 80 percent, which makes the sequence very stable (8). Additionally, the efficiency can be increased by utilizing multiple gRNAs to target the same location, as employing multiple gRNAs can increase the likelihood of editing the gene of interest (8).

Currently, there is not much research that has been done on CRISPR-Cas9 and myotonic dystrophy, as these research areas have been studied independently. Therefore, our goal was to link these areas of research and design gRNAs that will specifically target the trinucleotide repeats that cause myotonic dystrophy. We hypothesized that we would be able to design gRNA sequences that could target the trinucleotide repeats that cause myotonic dystrophy. We first designed these sequences by utilizing guidelines such as nucleotide composition and selection of specific target sites. Next, we tested these sequences by measuring their overall efficiency, on-target, and off-target scores to determine how effective these sequences were in deleting the trinucleotide repeats using the CRISPRedict tool (12). In this case, higher scores would correspond to better guides, and so, our end goal was to design a sequence that would produce high scores.

To summarize, we designed four gRNA sequences to specifically target CTG repeats in the *DMPK* gene, and we also tested their efficiency. We then used algorithms such as CRISPRedict to obtain scores for these sequences, which determined how successful our designs were. After designing and testing our sequences, we found that gRNA sequence 1 had the highest overall efficiency score of 0.631, similar to the efficiency score of the positive control, a sequence previously used in live cells. This design sequence could be taken further and tested in live cells, which would aid in truly determining the efficiency of this sequence via *in vitro* methods.

RESULTS

Our goal was to create gRNA sequences that would be able to cut out sections of trinucleotide repeats in the *DMPK* gene, ultimately decreasing the number of repeats in patients with myotonic dystrophy (**Figure 1**). To design efficient sequences, we manually created four gRNAs off the *DMPK* gene and tested them using two different algorithms. We began by locating sections of repeats in the *DMPK* gene, as well as a PAM sequence so that the gRNAs would know where to cut. Furthermore, by using guidelines such as manipulating specific nucleotides to increase stability, we were able to design four gRNA sequences. After this, we input the sequences into CRISPRedict and the IDT CRISPR-Cas9 Guide Design RNA Checker to obtain an overall efficiency score, an on-target score, and an off-target score that would predict how sensitive and specific the sequences were (11,12). The overall efficiency score (scored from 0–1) represents the overall performance of the sequence, while the on-target and off-target scores (scored from 0–100) focus more on the ability of guides to target specific sequences. For all the scores, the higher the score, the better the sequence performed in that category. In the off-target score category, a higher score would mean that there is a lower risk of off-targets. As such, an ideal sequence would have a high-efficiency score, a high on-target score, and a high off-target score. Finally, we compared these design sequences to a negative control (a randomly generated sequence) and a positive control (a sequence that is known to target CTG repeats in *DMPK*) to assess the efficiency of the gRNA designs.

We found that all four gRNA sequences could bind to Chromosome 19 (**Table 1**). For the overall efficiency score, the negative control had a score of 0.310, while the positive control had a score of 0.666. The design gRNAs had scores that were in between this range. This was also similar for the on-target score, as the negative control had a score of 13, while the positive control had a score of 72. Once again, the design sequences had scores in between the negative and positive controls, and they mostly hovered around a score of 40. However, this changed for the off-target category. The negative control had a score of 0, and the positive control had a score of 42, but gRNA 1 had a score of 90. Furthermore, the rest of the gRNAs were mostly between the range of 30–60, while gRNA 4 had a score of 8. This was important, as it showed that gRNA 1 outperformed the positive control in this category. Finally, when we calculated the *p*-values of each sequence via an ANOVA test, we found that of the four gRNA sequences we designed, gRNA 2, gRNA 3, and gRNA 4 all showed significantly lower efficiency levels compared to the positive control ($p < 0.05$), while gRNA 1 showed similar efficiency levels compared to the positive control ($p < 0.05$). This trend was generally applicable to all three scoring categories, and when compared to both the positive and negative controls in the on-target and off-target categories, the *p*-values were less than 0.05. This indicated that there was a statistically significant difference between the groups.

In all three efficiency score categories, the negative control had the lower scores, and the scores were usually the highest in the positive control. This was expected, as the positive control was designed to be the most efficient at targeting the correct sequence. Therefore, it would be the most likely to target the correct sequence, especially since it had been tested in cells. However, in the off-target category,

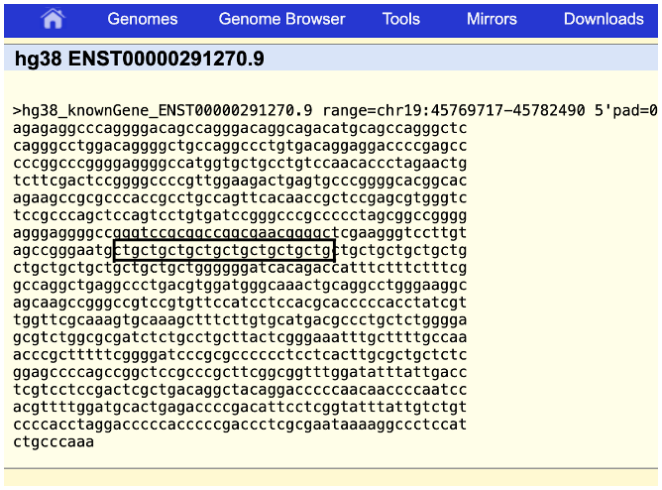

Figure 1: Image of *DMPK* untranslated region in UCSC Browser. Sequence of *DMPK* gene used to design gRNA sequences. Box indicates sample section of CTG repeats inside genome. Data obtained from the UCSC Genome Browser (15).

https://doi.org/10.59720/24-129

gRNA 1 had a better off-target score than the positive control, which meant that it was the best at avoiding other sequences in the genome.

DISCUSSION

In this study, we designed gRNAs targeting the human *DMPK* gene, specifically in the CTG repeat region, and tested their efficiency. We then compared the scores to negative and positive controls and finally determined what potential gRNA would be the best to test in future experiments. Based on the results provided by the experiment, our hypothesis that, if the necessary guidelines and rules are abided by when designing gRNA sequences, then they will be able to successfully target the *DMPK* gene and specifically the trinucleotide repeats that cause classic myotonic dystrophy, was supported. Through tools such as the UCSC Genome Browser and CRISPRedict, we designed four successful gRNA sequences to possibly bind the *DMPK* gene (12). To analyze the efficiency of each sequence, we generated an overall efficiency score, the on-target score, and the off-target score. The negative control had the lowest score out of the sequences in every category, and the positive control had the highest scores in every category except the off-target category. The experimental sequences were mostly between the negative and positive controls in all score categories, which showed that they were able to outperform the negative control. However, the positive control had a very high on-target score, while also having a lower off-target score. This may have occurred due to the positive control having high affinity for the target sequence and other sequences in the genome, which would then negatively impact the off-target score. Overall, out of the designed sequences, gRNA sequence 1 was the most efficient as it had the highest score out of all designed sequences in both the overall efficiency category and the off-target category and even outperformed the positive control with its off-target score with a high score of 90. This indicates that gRNA1 would have a better chance of performing well than the other sequences.

However, some follow up experiments could be done to reinforce the results. We could design more sequences that would target different locations, as the sequences in this study mostly targeted larger sections of CTG repeats. This could allow for a larger experimental group, which would identify more designs for future applications. If this could not be done, then baseline criteria could be set to pick specific locations in the first place. This would be very beneficial, as it would narrow down the specific target locations, making the design of the gRNAs more specific to those target locations when starting the experiment. To further test the sequences in this experiment, *in vitro* testing should be carried out to examine the effectiveness of these sequences in skeletal muscle cells due to the presence of the *DMPK* gene.

CRISPR-Cas9 is becoming much more well-known in the scientific world, so it is crucial to study more of its uses and how it can be used for various diseases. This study explored how gRNA sequences can be designed to target specific gene sequences to cut harmful parts of the genome, which can aid many people with genetic diseases. In fact, some patients have even been cured of sickle cell disease with the help of CRISPR, showing that CRISPR can play a part in curing other diseases as well (8). By using delivery pathways such as plasmids, we can test these sequences in live cells to truly analyze the efficiency of the experimental design sequences. This would also shine light on how well the scores obtained in this experiment correlate to *in vivo* results. There are many reasons why our efficiency scores may not correlate with *in vivo* results, such as chemical modifications, such as phosphorothioate, which are added to gRNA sequences to improve their stability (10). Thus, we could then ask whether high efficiency scores in these categories correlate with their ability to cut target sequences at the right location (12).

In conclusion, this experiment designed gRNAs intended for use with the CRISPR-Cas9 system to suppress the effects of detrimental mutations that caused classic myotonic dystrophy. gRNA sequence 1 was the most efficient of the designs, which means that it may have a higher chance of

Guide RNA Sequence	Nucleotide Sequence	Overall Efficiency Score (0 - 1)	On-Target Score (0 - 100)	Off-Target Score (0 - 100)
Negative Control	ACGACGACGACGACGACCCC	0.310	13	0
gRNA Sequence 1	CUGCUGCUGCUGCUGCUGGG	0.631	33	90
gRNA Sequence 2	ACCUUCCUGACUCACGGGCCC	0.451	41	37
gRNA Sequence 3	GCGAUCUCUGCCUGCUUACUCGGG	0.559	40	63
gRNA Sequence 4	GGGCAAACUGCAGGCCUGGGAAGG	0.373	40	8
Positive Control	CUGGUCAUGGAGUAUUACGU	0.666	72	42

Figure 2: Overall efficiency, on-target score, and off-target score for each gRNA sequence. Overall efficiency, on-target, and off-target scores for four gRNAs designed to target the *DMPK* gene as well as a random gRNA sequence (negative control) and a gRNA sequence known to target *DMPK* in vitro (positive control). Overall efficiency scores were obtained from the CRISPRedict algorithm (12). On-target scores and off-target scores were provided by the IDT CRISPR-Cas9 guide RNA design checker (11).

https://doi.org/10.59720/24-129

success when tested in live cells. Research involving gene editing is extremely crucial today, as scientists are now turning to gene therapy. Therefore, these new technologies need to be studied further to truly utilize them to benefit mankind.

MATERIALS AND METHODS

To obtain the DNA sequences, we used the UCSC Genome Browser (15). This tool provides the entire human genome, and we utilized it to find sections of CTG repeats in Chromosome 19. To achieve the maximum impact, we needed to target the correct region in the *DMPK* gene (Ensemble Gene ID: ENSG00000104936.13). Then, we manually located the introns where the repeats were present in large groups, which helped us determine our target sites.

Using the reference target sequence, we designed guide sequences. We did this by utilizing guidelines of stability. This can be manipulated by the nucleotide make of the gRNAs (through the number of guanines and cytosines), which is why this was crucial in increasing the efficiency of gRNAs (8). We decided on around 10–12 guanines and/or cytosines in order to improve stability. However, when doing this, we also had to consider that the sequences would still be able to target the right location. To confirm this, we repeatedly tested these sequences with different numbers of cytosines and guanines, as well as different numbers of adenines and thymines, to implement stability and efficiency. Furthermore, we increased and decreased the lengths of the design sequences to improve efficiency. This was done by checking how efficient the sequences were through testing of different lengths. By following these steps, we were able to design sequences that will be able to bind to the target sequence, which was around 10–15 nucleotides.

The presence of a PAM sequence is also required for CRISPR-Cas9. We located the PAM manually, and then we designed the gRNA sequences by converting the DNA into RNA using base pair complementarity. We repeated these steps to design four gRNA sequences targeting introns with repeats in the UTR. Each gRNA design was slightly different, as each targeted different regions in the *DMPK* gene. All of the sequences targeted regions of around 20–30 nucleotides that mostly contained CTG repeats. After we manually designed the experimental sequences, we obtained a negative control through a random gRNA sequence generator (12). We acquired the positive control through the Synthego Knockout Guide Designer Tool (16). We chose this sequence because it was tested in live cells and proven to be efficient, making it a great positive control. These controls served as sequences that we could compare to our designs to determine their efficiency.

Once we designed the sequences, we tested them using predictive algorithms to assess their efficiency. First, we used the CRISPRedict algorithm, where the gRNAs were inputted. Using rules such as nucleic acid composition, the software then determined the overall efficiency score (12). This would ultimately help the algorithm, as guidelines such as nucleic acid composition indicate the binding stability of the gRNA. Additionally, other guidelines would also be taken in account to determine this score. Using this information, the algorithm could then take in the number of guanines (represented by G) in the design sequence and analyze this factor, which would help in providing a score for these sequences. Next, we input the sequences into the IDT CRISPR-Cas9 Guide Design RNA Checker, which generated two scores: an on-target score that determined how likely the sequences would target the correct location, and an off-target score that determined how likely the sequences were to target a location somewhere else in the human genome (11). Through these algorithms, we predicted the efficiency of the gRNA sequences.

Next, we performed an ANOVA test to determine the variance of our scores. We did this by comparing the scores of each category for the design sequences with both the positive and negative controls individually, which helped us analyze whether our data was statistically different from both controls. Furthermore, a Tukey HSD test was performed post-hoc with a significant *p*-value of 0.05 to determine what was causing the difference between the gRNAs. This then indicated whether our data was statistically different from both controls.

Received: May 17, 2024
Accepted: August 26, 2024
Published: April 17, 2025

REFERENCES

1. [MH1] “Myotonic Dystrophy: MedlinePlus Genetics.” *Medlineplus.gov*, medlineplus.gov/genetics/condition/myotonic-dystrophy.
2. Lanni, Stella, and Christopher E. Pearson. “Molecular Genetics of Congenital Myotonic Dystrophy.” *Neurobiology of Disease*, vol. 132, Dec. 2019, p. 104533, https://doi.org/10.1016/j.nbd.2019.104533.
3. Llamusí, Beatriz, and Ruben Artero. “Molecular Effects of the CTG Repeats in Mutant Dystrophia Myotonica Protein Kinase Gene.” *National Library of Medicine*, 9 Dec. 2008, https://www.doi.org/10.2174/138920208786847944 . Accessed 31 July 2024.
4. Bird, Thomas D. “Myotonic Dystrophy Type 1.” *GeneReviews*, 17 Sept 1999, https://www.ncbi.nlm.nih.gov/books/NBK1165/
5. ““DMPK Gene: Medlineplus Genetics.” *MedlinePlus*, U.S. National Library of Medicine, medlineplus.gov/genetics/gene/dmpk/.
6. Jinek, Martin, *et al*. “RNA-Programmed Genome Editing in Human Cells.” *eLife*, vol. 2, 29 Jan. 2013, https://doi.org/10.7554/elife.00471.
7. “What Are Genome Editing and CRISPR-Cas9?: Medlineplus Genetics.” MedlinePlus, U.S. National Library of Medicine, 22 Mar. 2022, medlineplus.gov/genetics/understanding/genomicresearch/genomeediting/.
8. Pak, Ekaterina. “CRISPR: A Game-Changing Genetic Engineering Technique.” *Science in the News*, Harvard University, 31 July 2014, sitn.hms.harvard.edu/flash/2014/crispr-a-game-changing-genetic-engineering-technique/ .
9. “CRISPR Guide Design Tools and Algorithm | STEMCELL Technologies.” www.stemcell.com/crispr-guide-design-algorithm#:~:text=An%20on%2Dtarget%20score%20is . Accessed 27 Apr. 2024.
10. Doench, John G, *et al*. “Optimized SgRNA Design to Maximize Activity and Minimize Off-Target Effects of CRISPR-Cas9.” *Nature Biotechnology*, vol. 34, no. 2, 18 Jan. 2016, pp. 184–191, https://doi.org/10.1038/nbt.3437
11. “CRISPR-Cas9 guide RNA design checker.” *Integrated DNA Technologies*, 3 May 2022, www.idtdna.com/site/

https://doi.org/10.59720/24-129

order/designtool/index/CRISPR_SEQUENCE . Accessed 31 July 2024.
12. Konstantakos, Vasileios, *et al*. "CRISPRedict: a CRISPR-Cas9 web tool for interpretable efficiency predictions." *National Library of Medicine*, 5 July 2022, pubmed.ncbi.nlm.nih.gov/35670672 /. Accessed 31 July 2024.
13. Johns Hopkins. "Myotonic Muscular Dystrophy." *Johns Hopkins Medicine*, 2 Feb. 2022, www.hopkinsmedicine.org/neurology-neurosurgery/specialty-areas/muscular-dystrophy/myotonic-muscular-dystrophy#:~:text=The%20diagnosis%20of%20Myotonic%20Dystrophy,(EMG)%20and%20muscle%20biopsy . Accessed 31 July 2024.
14. Hippensteele, Alana. "CRISPR Technology Has Cured Patients of Certain Genetic Diseases, But Not All Patients Can Receive It Due to Cost, Accessibility." *Pharmacy Times*, 10 Mar. 2023, www.pharmacytimes.com/view/crispr-technology-has-cured-patients-of-certain-genetic-diseases-but-not-all-patients-can-receive-it-due-to-cost-accessibility . Accessed 31 July 2024.
15. UCSC Genome Browser, "The UCSC Genome Browser Home," University of California, Santa Cruz, https://genome.ucsc.edu/. Accessed: Oct. 16, 2024.
16. Synthego, "CRISPR Design Tool," https://design.synthego.com/#/, Accessed Oct. 16, 2024.