Investigation of the effects of altered and split pegRNAs on intramolecular complementarity and prime editing outcomes

Ph.D. thesis

Zsuzsanna Biczók

Semmelweis University Doctoral College
Molecular Medicine Division

Supervisor: Ervin Welker, D.Sc
Official reviewers: Bernadett Szikriszt, Ph.D
Tamás Mészáros, Ph.D

Head of the Complex Examination Committee:
Attila Tordai, D.Sc
Members of the Final Examination Committee:
Máté Varga, Ph.D
András Váradi, D.Sc
Hedvig Tordai, Ph.D

Budapest, 2026

1. Introduction

The discovery of CRISPR/Cas systems, often described as bacterial immune systems due to their role in defense against viruses and bacteriophages, has enabled the development of revolutionary gene-editing technologies. Owing to their practicality and adaptability, these systems and their variants are now widely employed in targeted genome editing and biological research. The most extensively utilized system derives from *Streptococcus pyogenes* and comprises a Cas9 nuclease (SpCas9) complexed with a single guide RNA (sgRNA). The 20-nucleotide spacer sequence of the sgRNA directs the complex to a complementary DNA target, requiring the presence of a downstream NGG protospacer-adjacent motif (PAM) for successful binding and activity. Upon target recognition, SpCas9 induces a double-strand break three base pairs upstream of the PAM, which is subsequently repaired by cellular mechanisms such as non-homologous end joining or homology-directed repair, enabling the utilization of these for genome modification.

Although SpCas9 exhibits high activity at intended target sites, it can also cleave sequences with partial similarity, resulting in off-target modifications. To mitigate this, engineered variants generated through rational design or directed evolution have been developed, but as these often reduce on-target efficiency, screening is required to identify variants that balance high specificity with sufficient activity.

A new generation of SpCas9-based tools, base editors (BE) and prime editors (PE), decrease the occurrence of unwanted modifications by using an inactive (dead) or nickase form of the protein (dSpCas9 and nSpCas9, respectively), which either possess no cleavage activity or cleave only one strand of the DNA. These methods are also better suited for making small, base-pair-level modifications to the genome. The key advantage of prime editing lies in its versatility, allowing not only substitutions but also insertions and deletions via a prime editing guide RNA (pegRNA), whose 3′ extension provides a template for reverse transcription by an engineered M-MLV RT fused to SpCas9 nickase. Following target recognition and nicking, the edited DNA strand is synthesized using the pegRNA template and incorporated into the genome through cellular repair processes.

Prime editing, however, is a fairly complex, multi-step process, and as a result, a number of limiting factors can reduce its efficiency. The pegRNA can limit prime editing efficiency due to intramolecular complementarity between its spacer and PBS sequences, which may interfere with the editing efficiency. This effect is often exacerbated by longer PBS lengths, making the identification of an optimal PBS length critical for efficient editing. Although various strategies, such as adjusting melting temperature, introducing spacer mismatches, screening large pegRNA libraries, or applying cold shock have been explored, pegRNA optimization remains a significant challenge.

Also, work by Anzalone and colleagues demonstrated that prime editing can introduce edits up to 33 nucleotides

downstream of the cleavage site, however most subsequent studies have focused on optimizing edits near the target sequence. In contrast, editing at more distal positions (>10 nucleotides) generally shows low efficiency, highlighting the need for improved strategies to expand the applicability of prime editing, particularly for therapeutic use.

The proPE method is based on the separated pegRNA functions, but by using two separate guide RNAs. Cleaving of the non-target DNA strand is performed by the essential nicking guide RNA (engRNA), which is a canonical sgRNA and utilizes a standard 20-nucleotide spacer sequence. The template for the newly synthesized DNA strand is provided by the template providing guide RNA (tpgRNA), which possesses the 3' extension of the pegRNA (the PBS and RTT sequences), but uses a short (11–15 nucleotide long) spacer sequence to avoid cleavage of the tpgRNA target site, located within close proximity to the engRNA target site. By using these two guide RNAs in appropriate ratios, several inhibitory factors, affecting the process of prime editing, can be eliminated simultaneously, like the position of the edit or the length of the PBS sequence.

2. Objectives

We aim to investigate how the reduction in the intramolecular complementarity of the pegRNA affects the efficiency and specificity of prime editing, by:

- introducing mismatches into the spacer or PBS sequences,
- introducing 1–4-nucleotide deletions within the long, 20-nucleotide PBS sequence,
- combining the PBS deletions with spacer mismatches,
- and by examining whether this method alters the off-target activity of prime editing.

We also aim to evaluate whether the proPE approach is also effective in a hPSC cell line and on clinically relevant target sites.

3. Methods

3.1. Plasmid construction

The PEAR fluorescent reporter target plasmids were developed by Simon et al. (2022) in our group.

PegRNA expressing plasmids contain either an mCherry or a TagBFP to monitor transfection efficiency.

For the construction of pegRNAs or tpgRNAs, a two-step cloning procedure was used: first, the spacer coding linkers were cloned between BpiI sites using 3 units of the BpiI enzyme, 2 units of T4 DNA ligase, 500 μM ATP, 1× Green buffer, 50 ng vector, and 0.25 μM of each of the corresponding oligonucleotides. . In the second cloning step, linkers containing the RTT-PBS were cloned between BsmbI sites with the above-described method but using a different buffer condition (1 mM DTT and 1× Tango buffer).

The following plasmids were obtained from the non-profit plasmid distribution service Addgene: pCMV-PE2 (#132775), X330-Flag-dSpCas9 (Addgene #92113), pX330-Flag-wtSpCas9-H840A (Addgene #80453), pX330-Flag-wtSpCas9 (Addgene #92353).

3.2. Cell culturing and transfection

The HEK293T (CRL-3216) cells used were obtained from ATCC. The HuES9 human embryonic stem cell line was a gift from Dr. Douglas Melton, Harvard University. (HuES9 NIH Approval number: NIHhESC-09-0022 and Health Care Research Council, Human Reproduction Committee in

Hungary, Approval number: 6681/2012-EHR.) The cell lines were authenticated by their respective suppliers, and they tested negative for mycoplasma during our experiments.
In general, transfections of either of the cell types were performed in triplicates and transfected cells were analysed by flow cytometry three days post-transfection. Transfection efficiency was considered in both PEAR and genomic experiments.
For genomic experiments, samples were always analysed by flow cytometry first, followed by genomic DNA purification. When evaluating sequencing data, samples were normalized for transfection efficiency.

3.3. Flow cytometry

Flow cytometry analysis was carried out using an Attune NxT Acoustic Focusing Cytometer (Applied Biosystems by Life Technologies). As a rule, signals from a set target minimum of 20,000 viable single cells were acquired by gating based on the side and forward light-scatter parameters. BFP, GFP, mCherry and mScarlet signals were detected using the 405 (for BFP), 488 (for GFP), and 561 nm (for mCherry and mScarlet) diode lasers for excitation, and the 440/50 (BFP), 530/30 (*GFP*), 620/15 (mCherry) and 585/16 nm (mScarlet) filters for emission.

3.4 Genomic DNA purification and PCR

After flow cytometry genomic DNA was extracted using the Puregene DNA Purification protocol (GentraSystems Inc). Amplicons for next-generation sequencing were generated from

the genomic DNA samples using two rounds of PCR to attach Illumina handles. PCR was done in a S1000 Thermal Cycler (Bio-Rad) or PCRmax AlphaAC2 Thermal Cycler using Q5 high-fidelity polymerase supplemented with Q5 buffer, and 150 ng of genomic DNA in a total volume of 50 µl.

3.5 Next generation sequencing

Sequencing on an Illumina NextSeq instrument was performed by Delta Bio 2000 Ltd. Reads were aligned to the reference sequence using BBMap. The aligned reads were considered as the total reads for each sample. Indels at the pegRNA and 2nd nicking sgRNA target sites were computationally identified from the aligned reads. Indels were searched for at ± 2 bp around the nick/cut sites. Frequency of precise edits generated by prime editing was determined as the percentage of (sequencing reads with the desired modification without indels)/(number of total reads). Reads with the intended modifications were identified by searching for a sequence stretch containing the desired edit, flanked by 5-5 matching nucleotides.

3.6 Analysing data from Yu et al.

We analysed Dataset 3, 4, 7 and 8 from the Table S2 in the publication of Yu and Kim et al. (41) to compare the average PE efficiencies of pegRNAs grouped according to their PBS length between Library-Small and Library-epegRNA, when used with PE2max and PE4max systems. We included editing data in the analyses only from editing via the conventional 'NGG' PAM.

3.7 Statistics

Statistical significance was assessed by two-tailed unpaired or paired t-test for comparing two groups. For data sets with normal distributions statistical significance was assessed using one-way ANOVA with Tukey's post hoc test (when comparing each group to every other group) or with Dunnett's test (when comparing each group to a control group). In cases where data did not pass normality but fulfilled the assumptions of Box-Cox transformation the transformed data were analysed as above. If not, Kruskal-Wallis test with Dunn's test was applied. Statistical tests were performed using GraphPad Prism 9.1.2.

4. Results

4.1 Mismatches in the pegRNA can improve editing efficiency

Prime editing efficiency is strongly influenced by the length of the pegRNA 3′ extension, particularly the PBS which has a rather broad effective range (typically 8–15 nucleotides) and which must be optimized in a target-dependent manner. Notably, longer PBS lengths (15–17 nucleotides) are often associated with reduced efficiency, potentially arising from spacer–PBS interactions, a hypothesis supported by improved outcomes following pegRNA splitting or modified refolding conditions.

4.1.1 Mismatches in the pegRNA spacer

We systematically introduced mismatches into both the spacer and PBS while using varying PBS lengths to assess their impact on spacer–PBS complementarity and prime editing efficiency.

First, we analysed the mismatch tolerance of the SpCas9 protein, which is known to tolerate PAM-distal mismatches well. Using three targets with spacer mismatches starting from the 10th position, NGS results showed efficient cleavage for the two targets, while for the third all mismatches were detrimental likely because its 21-nucleotide length, which has been connected with increased fidelity previously. Next, we examined pegRNAs whose spacer contain one, two, or three nucleotide differences, mainly at PAM-distal positions and using PBSs of

varying lengths (10, 13, 17). The mismatches were able to enhance the efficiency even for the more efficient short PBS pegRNAs, although the biggest enhancements could be observed with PBS17 (up to 7.2-fold), where the compatible section is expected to be the longest. Mismatches placed in or before position 10, were mostly detrimental, as expected, while the ones affecting the sequence after the complementary region had mostly no effect. The use of multiple mismatches was also tested, and these were also able to enhance the efficiency of prime editing, although there seemed to be a higher chance for a detrimental effect, than with single mismatches.

We also evaluated the efficiency of mismatches in pegRNAs with a 20-nucleotide long PBS, which we hypothesize could potentially enhance transcription initiation at some targets, but there was no improvement in enhancing editing efficiency compared to the mismatched PBS17 pegRNAs.

4.1.2 Mismatches in the pegRNA PBS

Incorporating mismatches into the PBS sequence rather than in the spacer, may offer the advantage of lowering PBS complementarity, even in cases where SpaCas9 shows limited tolerance to mismatches in the spacer. We hypothesize that mutations starting from position 5 of the PBS can effectively reduce intramolecular complementarity while not interfering with the process of RT binding and the initiation of reverse transcription. We re-examined the three targets using pegRNAs with PBS of varying lengths (10, 13, and 20 nucleotides) that contained one or multiple mismatches in the PBS. The

mismatches only showed an effect with PBS20 and increased the prime editing efficiency up to 6.7-fold.

4.2 Deletions in the PBS sequence and the combination of spacer and PBS modifications

4.2.1 *Deletions in the pegRNA PBS*

We then investigated whether deletions introduced into the long (20-nucleotide) PBS sequence exhibit the same effect as mismatches. We introduced deletions of 1 to 4 nucleotides, starting at position 6 of the PBS sequence, into the pegRNAs. As controls, we also tested pegRNAs containing a shorter PBS region corresponding to the intact sequence located upstream of the implanted deletions. Deletions of different sizes and of various positions were also capable of increasing prime editing efficiency relative to the pegRNA with a long PBS, sometimes even surpassing the efficiency of the pegRNA optimized for the target site. Deletions starting at position 13 resulted in the highest efficiency, thus we selected these variants for subsequent experiments.

We also examined the intact sequence of the PBS, that remains after the deletion, by scrambling this section, however the experiment shoved conflicting results, indicating that the 3′ PBS region influences efficiency beyond simple PBS–target strand binding.

4.2.2 Combining spacer and PBS modifications

We next examined whether the effects of spacer mismatches and PBS deletions are cumulative. We selected a

one- and a two-base deletion starting at position 13 (located within a 20-nt PBS) and paired them with mismatches that are not directly opposite these deletions and are located at PAM-distal positions.

Compared to the optimal pegRNA, we observed no reduced efficiency, and the combinations resulted in increased editing efficiency on one target.

We then selected 10 genomic targets previously described in the literature and compared the spacer- and PBS-modified pegRNAs with the ones used in the original study. All newly designed pegRNAs contained a single-nucleotide deletion at position 13 of the 20 nt PBS, as well as an additional single mismatch (transition or transversion) at positions 13, 15, or 18 of the spacer. Of the 10 edits, 2 outperformed the optimal pegRNAs reported in the literature.

4.3 The impact of a single nucleotide deletion in the PBS on editing efficiency and off-target activity

When we compared pegRNAs containing combined modifications with pegRNAs containing only a deletion, in which the deletion occurred only at position 13 of the 20-nt PBS, the results showed that additional spacer variations are not necessarily required. PegRNAs containing only the deletion were already able to reach editing efficiencies reported in the literature, suggesting that this approach may reduce the burden of optimizing combined modifications; therefore, we named this solution SPELL (Streamlined Prime Editing with fixed-Length

PBS Leverage). These results were validated on multiple endogenous targets and in a hPSC cell line.

Also, after the analysis of two targets with known off-target sites (*HEK4* and *HBB*), no substantial increase in off-target indel activity or off-target prime editing activity was observed.

4.4 SPELL pegRNAs do not act by restoring nicking activity

At the beginning of our work, Éva Bakos conducted an experiment in which we compared the cleavage activity of the SpCas9 nuclease using conventional sgRNAs and pegRNAs containing a long (15–17 nucleotide) PBS sequence. The results showed that even at targets where prime editing has high activity with these pegRNAs, albeit with varying efficiency, cleavage activity is reduced compared to sgRNA, leading to the conclusion that the PBS-spacer complementarity in the pegRNAs may affect the nicking function

When we compared the cleavage activity of SpCas9 using pegRNAs containing PBSs of different lengths, one of which was the SPELL pegRNA containing a deletion, we found that although these pegRNAs yielded prime editing results of varying efficiency at multiple targets, their cleavage activity did not differ significantly. This surprising result suggests that this self-complementary effect does not necessarily influence nicking activity, but rather affects another step of prime editing, for example efficient RT priming.

4.5 ProPE enhances editing efficiency on low performing and target-distal edits

ProPE is a tool based on prime editing technology and its mechanism relies on the separation of the pegRNA functions, namely the cleavage and the templating of new DNA, which have been split into two separate pegRNAs (engRNA and tpgRNA). ProPE demonstrated increased efficiency even at editing positions distant from the nick site (edits located >10 positions downstream), where prime editing typically exhibits low editing efficiency (below 5% in HEK293T cells), thereby successfully broadening the editing window of prime editing.

10 endogenous targets were tested in a hPSC cell line, where prime editing is less active, to investigate the broader applicability of the system. Improvements in editing efficiency and specificity (indel/edit ratio) were observed in more than half of the tested targets at at least one of the three tested engRNA ratios (the optimal ratio was found to be 1:40). ProPE exhibited increased activity on all clinically relevant targets as well, with at least one of the two tested tpgRNAs, and in most cases, editing specificity was also higher with proPE.

5. Conclusions

Mismatches in the pegRNA spacer while using PBS sequences of different lengths (10, 13, 17, and 20 nucleotides) revealed that when located in the spacer-PBS complementary region, they can improve prime editing efficiency. The greatest improvement was observed for 17-nucleotide-long PBSs.

Mismatches within the PBS sequence were mostly detrimental when using shorter (10- or 13-nucleotide) sequences, whereas for 20-nucleotide-long PBSs, we observed improvements, albeit to varying degrees.

When using 1–4-nt deletions starting at position 6 in a 20-nucleotide-long PBS sequence, we observed varying effects depending on the length and location of the deletions; however, in some cases, the pegRNAs containing the deletion were able to match or even exceed the efficiency of the pegRNA with the PBS length optimal for the given targets.

After combining spacer mismatches and PBS deletions, we observed that although in certain cases the mismatches could help enhance the effect of PBS deletions, a single deletion at position 13 of the 20-nt-long PBS achieved optimal pegRNA efficiency for the majority of targets. We named this method SPELL pegRNA.

We also observed no notable differences in off-target editing and indel activities compared to previously characterized pegRNAs. Overall, SPELL pegRNAs are capable of achieving effective prime editing even without extensive optimization.

Finally, we evaluated proPE on target-distal edits and found that it enables more efficient editing compared to conventional PE, thereby even allowing the editing of hard-to-reach pathogenic SNPs.

6. Bibliography of the candidate's publications

Publications related to the thesis:

Biczók Z, Krausz SL, Simon DA, Tóth E, Varga É, Annus T, et al. Disrupting pegRNA intramolecular complementarity via PBS and spacer sequence alterations can enhance prime editing efficiency. Nucleic Acids Res. 2026 Apr 24;54(7):gkag292. doi:10.1093/nar/gkag292

IF: 13,1

Krausz SL, Simon DA, Bartos Z, Biczók Z, Varga É, Huszár K, et al. ProPE expands the prime editing window and enhances gene editing efficiency where prime editing is inefficient. Nat Catal. 2025 Oct;8(10):1100–16. doi:10.1038/s41929-025-01406-6

IF: 44,6

Publications not related to the thesis:

Simon DA, Tálas A, Kulcsár PI, Biczók Z, Krausz SL, Várady G, et al. PEAR, a flexible fluorescent reporter for the identification and enrichment of successfully prime edited cells. Lapinaite A, Stainier DY, Hamilton JR, editors. eLife. 2022 Feb 23;11:e69504. doi:10.7554/eLife.69504

IF: 7,7