DEVELOPMENT OF METHODS TO FACILITATE THE APPLICATION OF PRIME EDITING

PhD thesis book

Dorottya Anna Simon

Semmelweis University Doctoral College

Molecular Medicine Division

Supervisor: Ervin Welker, D.Sc

Official reviewers: Tamás Arányi, Ph.D
Sándor Spisák, Ph.D

Head of the Complex Examination Committee:
Attila Tordai, D.Sc

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

Budapest
2026

1. Introduction

Prime editing is a versatile CRISPR tool that enables the precise installation of substitutions and smaller insertions or deletions. It uses a Cas9 nickase, a variant that cleaves only one strand of the DNA. This Cas9 is fused to a reverse transcriptase domain, which can synthetise DNA from an RNA template. The third component of the prime editor (PE) complex is the prime editing guide RNA (pegRNA). The pegRNA has four main parts: (i) the spacer sequence, located at the 5′ end of the pegRNA, which targets the PE to a specific DNA sequence, by base-pair complementarity; (ii) the scaffold sequence, which forms a secondary structure that binds and activates Cas9; (iii) the reverse transcription template (RTT) contains the desired modification and serves as a template for the reverse transcriptase. (iv) The primer binding site (PBS), located at the 3′ end, is necessary for initiating reverse transcription.

The PE binds to its designated target on the DNA, as specified by the spacer sequence. Then, nCas9 separates the two strands of DNA and cleaves the non-complementary strand. The 3’ end of the cleaved DNA strand then hybridises with the complementary PBS sequence. The reverse transcriptase then extends the DNA strand using the RTT, which contains the desired modification, as a template. This edited sequence can

then be incorporated into the genome by cellular repair mechanisms. This can be promoted by adding a single guide RNA (sgRNA) to guide PE to nick the non-edited strand. The second nick biases DNA repair mechanisms to correct this strand using the edited strand as a template.

The main advantage of prime editing is that the sequence of the RTT can be freely altered within certain limits, enabling a wide range of edits to be introduced. However, its widespread use was initially hindered by its often low efficiency and the fact that the length of the PBS and the RTT, as well as the position of the second nick needed to be optimised for each modification. The main goal of my PhD work was to develop methods that can help to overcome these problems and thus facilitate the application of prime editing.

2. Objectives

In my PhD work my aim was to:

- Develop a reporter system for an easy, fluorescence-based detection of prime editing outcomes that allows maximum flexibility for the target sequences and pegRNA designs to be tested on it.
- To verify whether the same factors influence prime editing efficiency on the PEAR plasmid as in genomic context.
- Find out whether prime edited mammalian cells could be identified and enriched by using a plasmid-based surrogate marker for chromosomal DNA modifications.
- To demonstrate that PEAR can be applied to facilitate the development of new prime editing systems.
- To compare the efficiency of proPE and PE on target-distal edits, using targets identified in the clinically relevant *CYP* gene family.

3. Methods

3.1 Molecular cloning

PEAR target plasmids and plasmids for second nicking sgRNAs and pegRNAs were constructed by one or two rounds of one-pot cloning with the appropriate linker oligonucleotides. In each case, one-pot cloning mixtures were incubated at 37°C for 30 min before being transformed into NEB5-alpha competent cells. The sequence of the cloned plasmids was verified by Sanger sequencing.

3.2 Cell culturing and transfection

HEK293T, U2OS, HEK-BEAR-GFP, HEK-BEAR-mScarlet and K562 cells were grown in their appropriate media supplemented with 10% FBS and 100 units/mL Penicillin-Streptomycin. Cells were cultured at 37°C in a humidified atmosphere of 5% CO_2. HEK cell lines were transfected with TurboFect, K562 and U2OS cells were nucleofected using an Amaxa 4D-Nucleofector, according to the manufacturers' protocols. Each transfection was performed in three replicates. In plasmid-based PEAR experiments cells were analysed by flow cytometry on day 3 from transfection. In enrichment experiments, where cell sorting was used, cells were sorted three days after transfection. *The cell culturing and transfection of U2OS and K562 cells was carried out by Péter István Kulcsár.*

3.3 Flow cytometry and cell sorting

Flow cytometry analysis was carried out using an Attune NxT Acoustic Focusing Cytometer. In all experiments, a minimum of 10,000 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 laser for excitation, and the 440/50 (BFP), 530/30 (GFP), 620/15 (mCherry), and 585/16 nm (mScarlet) filter for emission. Attune Cytometric Software v.4.2 was used for data analysis.

Cell sorting was carried out on a FACSAria III cell sorter (BD Biosciences). The live single-cell fraction was acquired by gating based on side and forward light-scatter parameters. BFP or GFP signals were detected using the 405 or 488 nm diode laser for excitation and the 450/50 or 530/30 nm filter for emission, respectively. A minimum of 50,000 cells were sorted in all experiments. *In every experiment, cell sorting was carried out by György Várady.*

3.4 Genomic DNA purification and genomic PCR

Genomic DNA was extracted according to the Puregene DNA Purification protocol. The purified genomic DNA was subjected to PCR analysis conducted with Q5 polymerase and

locus-specific primers. PCR products were gel purified and subjected to next-generation sequencing.

3.5 Next generation sequencing analysis

3.5.1 PEAR enrichment experiments

Reads were aligned to the reference sequence using BBMap. Indels were counted computationally among the aligned reads that matched at least 75% to the first 20 bp of the reference amplicon. Indels without mismatches were searched at ± 2 bp around the cut site with allowing indels of any size. For each sample, indel frequency was determined as (number of reads with an indel)/(number of total reads). Frequency of substitution without indels generated by prime editing was determined as the percentage of (sequencing reads with the intended modification, without indels)/(number of total reads). By contrast, frequency of intended insertions or deletions generated by prime editing was determined as the percentage of (all sequencing reads with the intended modification)/(number of total reads). For these samples, the indel background was calculated from reads containing different types of indels, than the aimed edit. To avoid falsely high specificity ratios during calculations indels lower than 0.05% were assumed to be 0.05%. *In all PEAR NGS experiments bioinformatic calculations were performed by Sarah Laura Krausz.*

3.5.2 Experiments on the *CYP* gene targets

During analysis, reads derived from non gene-specific primer-annealing and from mixed PCR products due to template switching were excluded by exploiting two gene-specific motifs located at different positions of the amplicon. Editing and indel values were calculated as described above (Section 3.5.1).

3.6 Statistics

Unless stated otherwise, differences between samples were tested using one-way ANOVA with Tukey's post hoc test for homoscedastic samples. Homogeneity of variances was tested by Brown-Forsythe test and normality of residuals was tested by D'Agostino-Pearson omnibus (K2) test. 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.

Figure 12: For each normalised dataset, a straight-line model was fitted using non-linear regression with the least squares fitting method. The null hypothesis that the best-fit slope is the same for all datasets was tested with an extra sum-of-squares F-test. Error bars include error propagation.

Statistical tests were performed using GraphPad Prism 9.2.

4. Results

4.1 Development of the prime editor activity reporter (PEAR)

PEAR is based on a GFP protein whose coding sequence is split by an intron that has a modified 5' splice site to abolish splicing and therefore the fluorescent signal. This can be restored by PEs, so the appearance of a fluorescent signal indicates successful prime editing. Experiments have confirmed that PEAR is sensitive enough to distinguish between the editing efficiencies of different prime editing setups. By design, PEAR provides unrestricted flexibility in the target sequence and with the tolerance of the splice donor site for substitutions, even the region of the edit is readily adjustable.

4.2 PEAR in genomic context

To test whether the same factors influence prime editing on plasmids as on genomic targets, editing efficiencies were compared on two cell lines containing an intron-disrupted GFP or mScarlet with an inactive splice site and plasmids coding the same disrupted proteins. For all three targets tested the measured efficiencies on plasmids and genomic targets correlated strongly. Implying that our system reflects accurately the main features of prime editing and factors affecting its efficiency.

4.3 PEAR as a selection marker for prime edited cells

Previous work has shown that Cas9 edited cells can be successfully enriched with a plasmid-based surrogate marker that is subjected to the same type of genetic modification as the intended genomic edit. In this section, I tested the applicability of PEAR as an enrichment marker in HEK293T, K562, and U2OS cells. To facilitate transfection, the PEAR-GFP-2in1 plasmid was constructed, which encodes both the PEAR reporter and the pegRNA targeting the reporter. In these experiments the pegRNA and sgRNA targeting the genome also coded BFP as a transfection marker. Three days after transfection, cells were sorted into three fractions: (i) single living cells, i.e. no enrichment, (ii) cells expressing BFP - transfection marker enrichment and (iii) cells expressing GFP - PEAR enrichment. From the sorted populations, editing and indel formation at the genomic target sites was quantified by NGS.

On average, in all enrichment experiments, PEAR enrichment increased the number of edited cells by 2.8-fold compared to samples without enrichment. With this, it was twice as efficient as transfection marker enrichment, which only resulted in a 1.4-fold increase. Using the HEK293T samples, I also examined whether PEAR enrichment influenced off-target prime editing. Out of the fifteen known Cas9 off-target sites

tested, there was only one site with detectable editing and indel formation, which was also enriched by PEAR enrichment. To mitigate the off-target effect of the pegRNA targeting the PEAR plasmid, the PEAR-GFP-2in1-2.0 plasmid was constructed, on which the pegRNA has no detectable off-target sites in the human genome as determined by GUIDE-seq. This plasmid could enrich various types of edits with the same efficiency as the original PEAR-GFP-2in1 plasmid.

4.4 The application of PEAR in the development of a novel PE tool

ProPE (prime editing with prolonged editing window), a novel PE tool was recently developed in our research group. During the development process, we used the PEAR system to determine the effective parameter ranges of proPE, and then to demonstrate some of its efficiency enhancing effects compared to PE. One of these was the experiment presented in this section, which I designed to assess the efficiency-decreasing effect of degraded RNAs on PE and proPE. The results showed that: (i) PEAR is sensitive enough to detect the difference in the efficiency-decreasing effect of degraded RNAs; and (ii) proPE is less susceptible to this effect.

4.5 Comparing the efficiency of proPE and PE on target-distal edits

In early experiments on genomic targets with proPE it was found that the efficiency of PE usually declined when the distance of the edit was 10 or more nucleotides from the nick site. To evaluate proPE's ability to enhance the efficiency of such edits, I designed edits lying outside the target sequences to install naturally occurring single nucleotide polymorphisms in the *CYP1A1*, *CYP1A2* and *CYP2B6* genes. The results showed that, proPE could significantly increase the median editing efficiency from 1.2% to 5.2%, and the median specificity 4.1-fold.

5. Conclusions

- PEAR, the reporter system developed here, exclusively reports on prime editing and is not sensitive to potentially caused indels. However, it is sufficiently sensitive to detect differences in the editing efficiencies of different pegRNA – second nicking sgRNA combinations.
- PEAR can effectively restore various active splice sequences and can be used with different target sequences.
- Experiments with the BEAR-GFP and BEAR-mScarlet plasmids and cell lines have shown that the efficiency of prime editing on the PEAR plasmid is affected by the same factors as in the genomic context.
- PEAR can be applied as a surrogate marker to enrich prime edited cells in multiple cell lines. The PEAR-GFP-2in1 plasmid could enrich the edited population in HEK293T, K562 and U2OS cells, resulting in up to 76% editing, without compromising specificity. Off-target analysis of the HEK293T samples showed that PEAR enrichment only minimally affects the levels of off-target prime editing.
- To minimise the genome-wide off-target effect of the PEAR-GFP-2in1 plasmid, PEAR-GFP-2in1-2.0 was constructed.

This plasmid could enrich prime edited HEK293T cells similarly to PEAR-GFP-2in1.

- The inhibitory effect of degraded RNAs on PE and proPE was assessed using the PEAR system. Based on the results, proPE is less affected by the presence of PBS-less and (RTT-PBS)-less RNAs.
- The efficiency-enhancing effect of proPE on target-distal edits was demonstrated by installing naturally occurring SNPs in the *CYP1A1*, *CYP1A2* and *CYP2B6* genes.

6. Bibliography of the candidate's publications

6.1 Publications related to the thesis

- S. L. Krausz, D. A. Simon, Z. Bartos, Z. Biczók, É. Varga, K. Huszár, P. I. Kulcsár, A. Tálas, Z. Ligeti, and E. Welker, "ProPE expands the prime editing window and enhances gene editing efficiency where prime editing is inefficient," *NATURE CATALYSIS*, vol. 8, pp. 1100–1116, 2025.
- D. A. Simon, A. Tálas, P. I. Kulcsár, Z. Biczók, S. L. Krausz, G. Várady, and E. Welker, "PEAR, a flexible fluorescent reporter for the identification and enrichment of successfully prime edited cells," *ELIFE*, vol. 11, 2022.

6.2 Publications not related to the thesis

- Tálas, D. A. Simon, P. I. Kulcsár, É. Varga, S. L. Krausz, and E. Welker, "BEAR reveals that increased fidelity variants can successfully reduce the mismatch tolerance of adenine but not cytosine base editors," *NATURE COMMUNICATIONS*, vol. 12, no. 1, 2021.