2502

Comprehensive Analysis of the Editing Window of TALE Base Editors

Maria Feola[1], Sylvain Pulicani[2], Diane Tkach[1], Robert Hong[1], Alex Boyne[1], J Lomibao[1], Louisa Mayer[1], Aymeric Duclert[2], Philippe Duchateau[2], Alexandre Juillerat[1]

[1] Cellectis Inc, New York, NY; [2] Cellectis SA, Paris, France

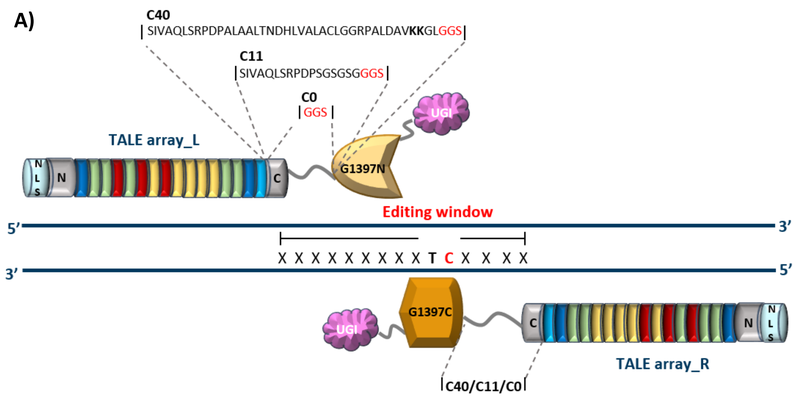
Abstract

TALE base editors (TALE-BE), one of the most recent advances in the toolbox of genome editors, are fusions of a transcription activator-like effector domain (TALE), split-DddA deaminase halves, and a uracil glycosylase inhibitor (UGI). The C-to-T class of TALE-BE can edit double strand DNA by conversion of a cytosine (C) to a thymine (T) via the formation of a uracil intermediate. In this study, we aimed to systematically investigate the influence of the sequence context surrounding the targeted C on TALE-BE C-to-T conversion efficiency. We recently developed a strategy that allowed the comprehensive characterization of editing efficiencies on the function of the TC position within the TALE-BE editing window. This experimental set-up allowed for a high throughput screening of editing efficiency in a precisely defined genomic context *in cellulo*. In particular, the approach minimizes or excludes biases that would arise from microenvironmental and/or epigenetic differences, such as chromosome relaxation, that would be present among different genomic loci. This method also takes advantage of a highly precise and efficient TALEN®-mediated ssODN knock-in in primary T cells to assess how target composition and spacer variations affect TALE-BE activity/efficiency. This robust and versatile approach allowed insight into the editing rules of TALE-BE and further highlighted that the composition of bases surrounding the target TC is strongly correlated with editing efficiencies. Therefore, educated choice of the TALE-BE architecture and its positioning on target DNA could either prevent target sequence limitations (increasing the targetable sequence) or be used to decrease, or eliminate, bystander editing with the editing window (increasing specificity), allowing TALE-BE to be fine-tuned for a desired gene editing outcome. Altogether, the knowledge obtained in this study will enable the design of efficient TALE-BE, while improving the specificity profiles of this innovative gene editing tool for nuclear and mitochondrial therapeutic cell engineering.

#1 Background

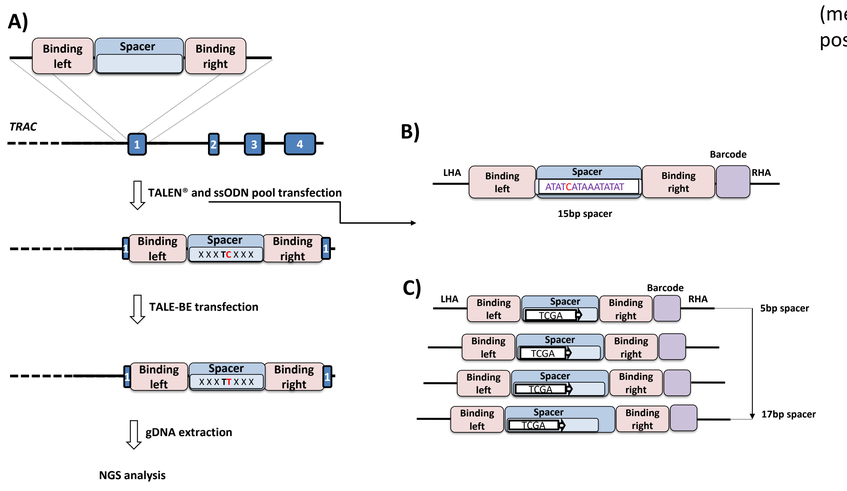
Base editing is a technology that leads to the introduction of point mutations in defined loci of a targeted DNA sequence. Base editors create mutations by deamination of the targeted bases (C or A), which are then converted into T or G, respectively, during the DNA repair process. The first development of TALE-BE rely on the use of a TALEN® scaffold. It has been shown that, in a TALEN® context, the linker could impact the overall efficiency of such molecular tools. To extend our understanding of key determining factors allowing efficient TALE-BE editing, we compared the following scaffolds:

- **C40/C40** scaffold - the linker domain is composed of the native first 40 amino acids from the C-terminal domain of TALE with a flexible GGS linker
- **C11/C11** scaffold – a truncated C-terminal end of the TALE domain with a GGS linker
- **C0/C0** scaffold - maintaining only the GGS linker.



A. Schematic representation of a TALE-BE in the **C40**, **C11** and **C0** scaffolds.

#2 Definition of the optimal TALE-BE editing window: high throughput testing

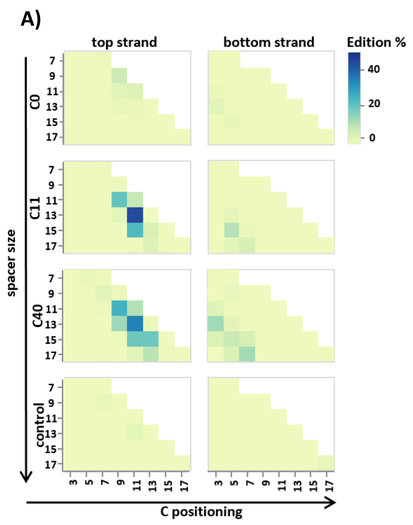


A. The strategy used to generate artificial base editor target sites.

B. Schematic representation of the 15 ssODN pool collection to determine the impact of the three linkers (C40, C11 and C0) on editing of targets containing a unique TC within a spacer of 15 bp, known to be optimal for the C40 TALE-BE scaffold.

C. Schematic representation of the 37 unique ssODN pools characterized by spacers with varying lengths spanning from 5 to 17 bp (tested in denominations of two; e.g. 5 bp, 7 bp, 11 bp, etc.), to determine the impact of shortening the linker on editing efficiency as a function of the spacer length. A TCGA quadruplex target sequence was incorporated into the spacer at varying positions from 5' to 3'.

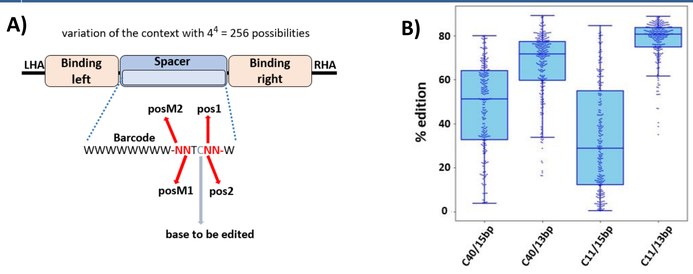
#3 Analysis of the linker nature and spacer length influence on TALE-BE C-to-T conversion



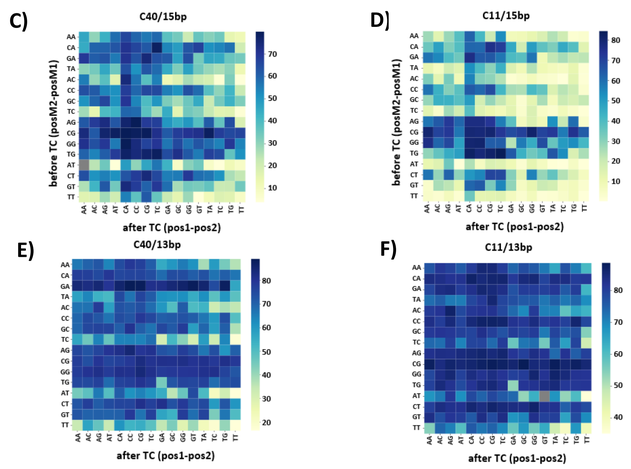
A. The C-to-T conversion of TALE-BE measured on the target collections with spacer lengths spanning 5 to 17 bp.

- In all three linker combinations, absent or very low editing on both the top and bottom strand was observed for spacer length ≤ 9 bp (max editing values: C40: C_4 1.3%; C11: C_2 3.3%; C0: C_6 1%).
- With a spacer length of 11bp, the C40 and C11 TALE-BE architecture combinations showed similar editing levels almost exclusively on the top strand (max editing: C40: C_8 25%; C11: C_8 20%).
- Highest editing rates were obtained on 13 bp and 15 bp spacers (max editing: C40: C_{10} 34%; C11: C_{10} 44% on 13 bp; C40: C_{10} 19%; C11: C_{10} 22% on 15 bp).
- Shortening the spacer from 15 to 13 bp increased editing, but did not allow rescue of activity for the C0 architecture.

#4 Influence of the TC context: 15 and 13 bp spacer length



A. Schematic representation of the ssODN pool collection containing the targeted 5'TC at the optimal positioning (position C_5), within optimal spacer lengths (13 and 15 bp). **B.** % C-to-T conversion of the 256 possible targets showed overall editing on the 13 bp collection (median: 72% for C40, 81% for C11) was higher when compared to the 15 bp collection (median: 51% for C40, 29% for C11), which was expected from the slightly more favorable positioning of the TC (position C_5) within the former spacer length.



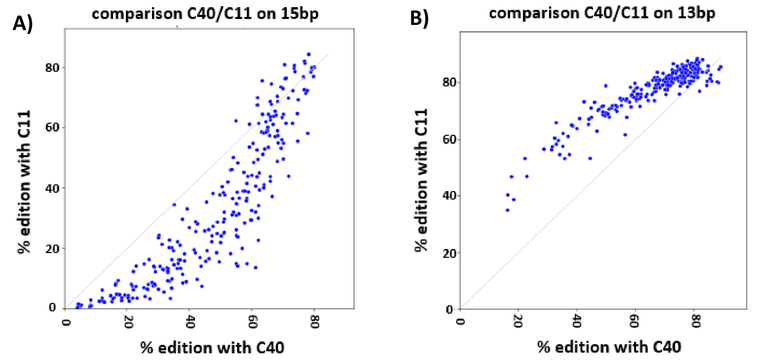
C-D. C11/C11 editing is more stringent against a suboptimal context, compared to C40/C40. The 15 bp spacer collection in the C40 and C11 architectures show the following nucleotide preferences:

posM2: A = T << G < C posM1: T = C < A << G
pos1: T < G < A << C pos2: T < C < G < A

E-F. The 13 bp context in the C40 and C11 architectures seem to have less importance for editing efficiency, which shows higher overall editing, compared to (C-D). However, the 13 bp spacer collection show a similar context preference to the above:

posM2: T < A < G < C posM1: T < C < A < G
pos1: T < G < A < C pos2: T < A = C < G

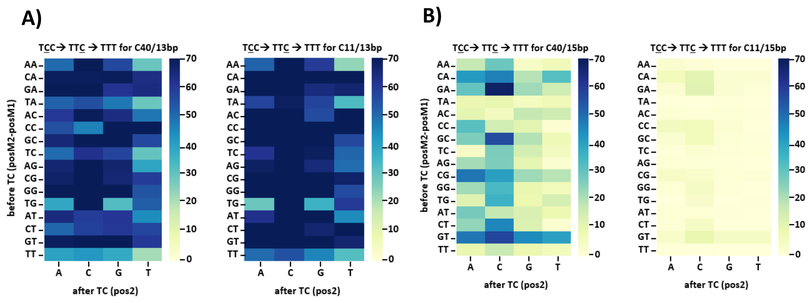
#5 Editing efficiency comparison between scaffolds on 15 bp and 13 bp spacer length



A. With the 15 bp spacer, editing percentages between the two scaffolds (C40/C40 v. C11/C11) show a non-linear correlation, suggesting that for contexts unfavorable to editing, activity with C11/C11 is lower than for the C40/C40 scaffold; for contexts that are the most favorable to editing, the activity is similar in C11/C11 and C40/C40, and both scaffolds can reach ~80% C-to-T conversion.

B. In the 13bp spacer context, comparison of editing percentage between C40/C40 and C11/C11 also show a non-linear correlation, with C40/C40 scaffold being more tolerant in unfavorable contexts (reverse trend to above); however, both scaffolds reach higher editing values (range: 88%-100%, mean = 96%), compared to the 15 bp spacer.

#6 Choice of TALE-BE design minimizes editing of Cs



A. % C-to-T conversion of the pos1 C was measured in a 13 bp spacer context. Both the C11 and C40 architecture showed a permissive profile with high rates of editing (C11: 70.96 +/- 14.53, C40: 60.78 +/- 13.74; median +/- stddev). Comparison of the editing results showed small differences in mutation rates on the pos1 C, almost independent of context requirements.

B. In the 15 bp spacer context, editing results showed a clear difference in mutation rates on the pos1 C, with the C40 linker combination showing more permissive activity overall, with less dependence on context requirements than the C11 scaffold. The 15bp spacer in both architectures showed more restrictive editing (C11: 1.09 +/- 2.68, C40: 17.21 +/- 15.41, median +/- stddev). The use of the C11 architecture with the 15 bp spacer nearly abolished editing in most contexts, revealing the possibility to prevent edits on stretches of multiple cytosines.

#7 Conclusions

This data demonstrates that:

- It is possible to narrow editing on the top strand using a C11/C11 combination
- The optimal spacer length (13/15 bp) for highly efficient TALE-BE for both C40/C40 and C11/C11 scaffolds
- The optimal common sequence context for high editing rates
- The editing efficiency of the C11/C11 combination is highly dependent on cytosine position, resulting in more stringent activity in a context of 15 bp spacer length and decreased bystander editing

Altogether, a combination of architecture, sequence composition, and spacer length enables us to identify the best TALE-BE candidate for a given locus