<table>
<thead>
<tr>
<th>Title</th>
<th>Insertion and Deletion Mismatches Distant from the Target Position Improve Gene Correction with a Tailed Duplex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Kamiya, Hiroyuki; Nishigaki, Natsuki; Ikeda, Akihiro; Yukawa, Seiya; Morita, Yukiko; Nakatsu, Yoshimichi; Tsuzuki, Teruhisa; Harashima, Hideyoshi</td>
</tr>
<tr>
<td>Citation</td>
<td>Nucleosides, nucleotides &amp; nucleic acids, 35(7): 379-388</td>
</tr>
<tr>
<td>Issue Date</td>
<td>2017-07-03</td>
</tr>
<tr>
<td>Doc URL</td>
<td><a href="http://hdl.handle.net/2115/66349">http://hdl.handle.net/2115/66349</a></td>
</tr>
<tr>
<td>Rights</td>
<td>This is an Accepted Manuscript of an article published by Taylor &amp; Francis in Nucleosides, Nucleotides and Nucleic Acids on 2016, available online: <a href="http://www.tandfonline.com/10.1080/15257770.2016.1163384">http://www.tandfonline.com/10.1080/15257770.2016.1163384</a></td>
</tr>
<tr>
<td>Type</td>
<td>article (author version)</td>
</tr>
<tr>
<td>File Information</td>
<td>manuscript.pdf</td>
</tr>
</tbody>
</table>
Insertion and deletion mismatches distant from the target position improve gene correction with a tailed duplex

Hiroyuki Kamiya,a,b,c,* Natsuki Nishigaki,a,b Akihiro Ikeda,a Seiya Yukawa,a Yukiko Morita,c Yoshimichi Nakatsu,d Teruhisa Tsuzuki,d and Hideyoshi Harashima,c

a Graduate School of Science and Engineering, Ehime University, 2-5 Bunkyo-cho, Matsuyama 790-8577, Japan, b Graduate School of Biomedical and Health Sciences, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8553, Japan, c Faculty of Pharmaceutical Sciences, Hokkaido University, Kita-12, Nishi-6, Sapporo 060-0812, Japan, and d Graduate School of Medical Sciences, Kyushu University, Maidashi 3-1-1, Higashi-ku, Fukuoka 812-8582, Japan.

This work was supported in part by Grants-in-Aid from the Japan Society for the Promotion of Science (21300174, 21659015, and 25282144).

*Address correspondence to Hiroyuki Kamiya, Ph. D., Graduate School of Biomedical and Health Sciences, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8553, Japan.
E-mail: hirokam@hiroshima-u.ac.jp

Running title: Indel mismatches improve gene correction
Abstract

A 5'-tailed duplex (TD) DNA corrects a base-substitution mutation. In this study, the effects of insertion and deletion (indel) mismatches distant from the target position on the gene correction were examined. Three target plasmid DNAs with and without indel mismatches ~330 bases distant from the correction target position were prepared, and introduced into HeLa cells together with the TD. The indel mismatches improved the gene correction efficiency and specificity without sequence conversions at the indel mismatch site. These results suggested that the gene correction efficiency and specificity are increased when an appropriate second mismatch is introduced into the TD fragment.

Keywords  Gene correction; Nucleotide sequence conversion; Tailed duplex; Indel mismatch; *rpsL* gene
INTRODUCTION

Many mutations that cause the loss of important functions have been found in disease-responsible genes. Gene correction, the nucleotide sequence conversion of a disease-causing mutation to the normal sequence, is attractive as therapy for such kinds of diseases. [1-7] For example, correction of TAG (termination) to TCG (Ser) etc. would regain the gene product function(s). In addition, the disruption (knockout) of a gene essential for the maintenance of a disease could be achieved with the same technology. This method could also be applied to prepare disease-model animals and cells, as well as useful cells for biotechnology.

Gene correction has advantages over transgene expression as a gene therapy method; however, the correction efficiency must be increased for its clinical use. [8] Heat denatured, several hundred bp polymerase chain reaction (PCR) products containing the normal sequences have been examined in the small fragment homologous replacement method. [1-3] We designed a single-stranded (ss) DNA fragment containing the sense sequence, and it corrected the targeted DNA with one order of magnitude higher efficiency than the denatured PCR product. [9] Moreover, the 5'-tailed duplex (TD) DNA, prepared by annealing an oligonucleotide to the ss DNA fragment (Fig. 1), achieved gene correction more efficiently than the ss DNA fragment. [10] The naked TD fragment altered the sequence of the rpsL chromosomal transgene in mouse livers when it was delivered by hydrodynamic tail vein injection. [11-13] However, the gene correction efficiency was ~0.1%, and thus should be improved.

Although the molecular mechanisms of gene correction by the ss and TD DNA fragments have not been elucidated, homologous recombination (HR) proteins are considered to be involved.
The proteins would search for the homology between the gene correction DNA fragments and their target DNAs in the gene correction process. Single base-base mismatches are present between the fragments and the targets in the correction of base-substitution mutations: when the normal and mutated bases are G and T, respectively (in the case of G→T substitution mutations), an A:G mismatch would be formed. The second mismatch at another position was expected to reduce the correction efficiency, due to the decreased homology. In this study, the effects of insertion and deletion (indel) mismatches ~330 bases distant from the correction target position were examined. We found that the indel mismatches increased the gene correction efficiency and specificity. In addition, the sequence alteration was induced at the target position, but not at the indel mismatch positions. The results obtained in this study suggested that both the gene correction efficiency and specificity are improved by the introduction of an appropriate second mismatch.

EXPERIMENTAL

General

Oligodeoxyribonucleotides (ODNs) were obtained from Sigma Genosys Japan (Ishikari), Hokkaido System Science (Sapporo), and Eurofins Genomics (Tokyo) in purified forms. The pSSW plasmid, containing the normal Escherichia coli rpsL gene and the TTTT sequence in the region 92-95 bp upstream from the start codon, is referred to as pSSW(T4) in this study (Fig. 1). [15] The pSSW(T3) plasmid containing the TTT sequence (~1 deletion mutation), instead of the TTTT sequence, was obtained from a spontaneous mutant colony. The pSSW(T5) plasmid was constructed
by the insertion of ODNs containing the TTTTTT sequence (+1 insertion mutation) into the BstX I and Hin1 I sites of pSSW(T3). The plasmid DNAs were amplified in *E. coli* DH10B cells, and were purified with a Pure Yield Plasmid Miniprep System kit (Promega, Madison, WI, U.S.A.).

**Preparation of the TD DNA fragment**

The *rpsL* gene fragment, containing the T→G mutation at position 239 (the 2nd position of codon 80), was amplified by PCR and introduced into the *Eco*R I site of M13mp2 DNA. The ss phage DNA was obtained by the standard method. [16] The 791-base ss fragment was prepared from the phage DNA, by annealing with its respective scaffold ODNs (5'-dAAACCGGGAATTCGTAATCA-3' and 5'-dGGCCAGTGAATTCAGATCGA-3') followed by *Eco*R I digestion. The digested DNA was loaded onto low-melting point agarose gels, and the 791-base DNA fragment was extracted from the gels. After purification by phenol, phenol-chloroform, and chloroform extractions, followed by ethanol precipitation, the fragment was further purified by gel filtration chromatography. Its UV spectrum was measured to check its purity and to calculate the yield. The concentration was determined by the molar absorption coefficient of DNA: 1.0 OD$_{260}$ equals 40 µg of ss DNA. [16] The ss DNA fragment was mixed with a 10-fold molar excess of a 35mer ODN (5'-dCAGATCGAAGTACTCAGTTTAAGCTGCCTCCCATG-3'). The mixtures were heat-denatured at 98°C for 5 min, and were immediately chilled on ice for at least 5 min. They were then heated at 85°C for 5 min and cooled slowly to room temperature.
Cell culture and transfection

HeLa cells were seeded at 2 \( \times 10^5 \) cells/well in 6-well plates and cultured in Dulbecco’s modified Eagle’s medium, supplemented with 10% fetal bovine serum, at 37°C under a 5% CO\(_2\) atmosphere for 24 h. The TD fragment (5 pmol) was mixed with 12.5 fmol of the target plasmid (pSSW). Transfection into HeLa cells was performed with the Lipofectamine Reagent (Life Technologies, Carlsbad, CA, U.S.A.), according to the supplier’s instructions. After 48 h, the cells were harvested and the plasmid DNA was recovered, as described previously. [9]

Determination of gene correction efficiency

Electro-competent *E. coli* DH10B cells were prepared, essentially according to the method described in the literature. [17] The DNA recovered from the HeLa cells was electroporated at 1.8 kV, 25 \( \mu \)F and 200 \( \Omega \) with a Gene Pulser II (Bio-Rad, Hercules, CA, U.S.A.) into the electro-competent cells. The bacteria were seeded onto LB agar plates containing 200 \( \mu \)g/ml of streptomycin and 50 \( \mu \)g/ml of kanamycin (selection plates) and LB agar plates containing 50 \( \mu \)g/ml of kanamycin (titer plates), as described previously. [18] Gene correction efficiencies were calculated by dividing the number of streptomycin-resistant colonies on the selection plates by the number of colonies on the titer plates.

Statistical analysis

Statistical significance was examined by the Student’s \( t \)-test.
RESULTS

Increased correction efficiency with indel mismatches

We previously reported gene corrections with ss and 5'-TD DNA fragments, using the *rpsL* gene as a model target gene. [10,11,18] In this study, we compared the gene correction (sequence alteration) efficiencies when indel mismatches were present ~330 bases distant from the target position. The target base was T at position 239 (the 2nd position of codon 80, ATC) in the pSSW plasmid DNAs, and the corresponding base was G in the TD fragment (AGC) (Fig. 1B and Supplementary Fig. 1). The conversion of the T base to the G base changes the 80th amino acid of the *E. coli* ribosomal protein S12, and this mutation confers streptomycin resistance (*strA*) to *E. coli*. The TTTT sequence is present in the region 92-95 bp upstream from the start codon, and this sequence was the target of the indel mismatches (Fig. 1A and Supplementary Fig. 1). The TD DNA also possesses the TTTT sequence, and the corresponding sequences in the pSSW(T3) and pSSW(T5) plasmids are TTT and TTTTT, respectively. Thus, the TD fragment has +1 (+T) and −1 (∆T) mutations for the pSSW(T3) and pSSW(T5) plasmids, respectively (Fig. 1B).

The TD fragment was introduced into HeLa cells, together with a target plasmid (Fig. 1C). The TD fragment was used in a 400-fold molar excess over the target plasmid, and the co-transfection was performed with cationic lipids. After 48 h, the plasmid DNA was recovered from the transfected HeLa cells and electroporated into the *E. coli* DH10B (*recA* and *rpsL*) strain. The transformed DH10B cells were seeded onto agar plates containing streptomycin. The number of
streptomycin-resistant colonies was used for the calculation of the correction efficiency, as described in the Materials and Methods section.

As shown in Fig. 2, the co-introduction of the TD fragment and the pSSW(T4) plasmid, which both contain the TTTT sequence, resulted in 0.21% “gene correction efficiency”. This value was the frequency of streptomycin-resistant colonies, but not the actual gene correction efficiency, as described below. Unexpectedly, the correction efficiency was increased to 2.7% when the pSSW(T3) plasmid DNA was co-transfected with the TD fragment. This value was ~13-fold higher than that of pSSW(T4), indicating that the presence of the +1 mutation on the TD fragment dramatically improved the gene correction efficiency. Likewise, the correction efficiency was increased when the pSSW(T5) plasmid DNA was co-transfected with the TD fragment. In this case, the TD fragment contained the –1 mutation in the upstream region of the target site, and the enhancement was ~6.1-fold. These results indicated that the indel mismatches enhanced the gene correction efficiency.

**Sequence analysis of plasmids recovered from transfected cells**

We sequenced the plasmid DNAs in the streptomycin-resistant colonies, to examine whether the expected sequence alteration was induced. In total 41, 36, and 33 colonies were analyzed in the pSSW(T3), pSSW(T4), and pSSW(T5) experiments, respectively. As shown in Fig. 3, the successful correction at the targeted position was induced in the plasmids. Table 1 summarizes the sequence changes observed in the analyzed plasmids. In the pSSW(T3) experiment, all colonies contained the on-target T→G correction, and no nucleotide alterations in the flanking sequences.
were detected. Likewise, the correct sequence alteration was confirmed in almost all of the plasmids in the pSSW(T5) experiment. Interestingly, the TTT and TTTTTT sequences were maintained, although the target position was converted to that of the TD fragment (Fig. 3 and Table 1). An untargeted mutation was found in one colony in the pSSW(T5) experiment. In contrast, a significant proportion of the plasmids contained untargeted mutations for pSSW(T4). Nearly half of the corrected plasmids contained additional large insertion mutations. These results indicated that the indel mismatches improve both the specificity and efficiency of the gene correction.

DISCUSSION

In this study, we found that the gene correction efficiency was promoted when the indel mismatches (−1 and +1), in addition to the mismatch at the target position, were present between the TD and target DNAs. Moreover, the sequence was unchanged at the indel mismatch position, in contrast to the expected correction at the target position. In addition, the indel mismatches suppressed the untargeted mutations, indicating that they improved the gene correction specificity. These findings would be useful for the future design of improved gene correction systems.

The reason(s) for the enhanced correction efficiencies remain unknown. We have two lines of evidence suggesting that HR is involved in the gene correction processes. First, the order of the gene correction efficiencies for the base-substitution mutations (5'-TD > 3'-TD > ss DNA fragment) was identical to that of the in vitro DNA strand exchange efficiencies catalyzed by the RAD51 protein. [10,19] Second, the target plasmid DNA became radioactively labeled when radiolabeled ss
and TD DNA fragments were introduced into mammalian cells (reference 14 and Tsuchiya et al., unpublished results). Given the recognition of the TD and target DNAs by HR proteins, the higher gene correction efficiencies in the presence of the indel mismatches might be due to the higher strand exchange efficiencies by HR protein(s) in cells. Alternatively, the higher efficiencies observed in this study might be explained by more efficient processing after the strand exchange. To reveal the reason(s) for the present findings, further studies including in vitro experiments are necessary.

The corrections of the –1 and +1 indel mutations with the ss and TD fragments were achieved with much lower efficiencies than those of the base-substitution mutations. [20,21] Although the sequence alteration at the indel mismatch site itself does not confer streptomycin resistance to E. coli cells, the fact that the sequence alteration occurred at the target (base-base mismatch) site, but not at the indel mismatch site, might be related to these previous findings. Furthermore, the present findings indicated that the entire 791-base TD strand is not incorporated into the target DNA.

One may think that the indel mutations between the –35 and –10 promoter sequences would affect the rpsL gene expression, and hence the formation of streptomycin-resistant colonies. However, the mismatch site is not included in the core promoter sequences, although it is near the –35 sequence. Moreover, the E. coli DH10B cells possess the mutated rpsL gene and are resistant to streptomycin. The expression of the normal rpsL gene confers streptomycin sensitivity to DH10B cells. Direct transformations of DH10B cells with the pSSW(T3), pSSW(T4), and pSSW(T5) plasmids produced streptomycin-sensitive colonies, indicating the sufficient expression of the rpsL gene from the plasmids. Given that one molecule of the plasmid DNA would be introduced upon the
transformation of the bacterial cells, the putative expression alteration, if present, would not affect the selection efficiency by antibiotic resistance. The fact that all of the streptomycin-resistant colonies contained mutated \( rpsL \) genes supports this interpretation.

In addition to the increased efficiency, the indel mismatches improved the specificity of the gene correction (Table 1). In the case of pSSW(T4), half of the “corrected” molecules contained additional mutations. In contrast, no and only one undesired mutations were detected for pSSW(T3) and pSSW(T5), respectively. Again, additional studies including \textit{in vitro} experiments are necessary to reveal the reason(s) for this observation.

In this study, the \( rpsL \) gene was used as a model target gene. To expand knowledge about the second mismatch effects, comprehensive studies are required. However, this gene is inadequate to evaluate second mismatches located downstream of the target position since introduction of indel mutations into the coding region confers streptomycin resistance. Thus, other genes are expected to be more useful for further studies.

Recently, genome editing using artificial nucleases has been receiving plenty of attention. [22-26] However, off-target cleavage(s) of the genomic DNA by the artificial nucleases might cause severe side effects when they are used in gene therapy. The sequence conversion with the TD fragment could achieve genome editing without the artificial nucleases, and could be used as a safer clinical application.

In conclusions, the indel mismatches located upstream of the target position improved the gene correction with the TD fragment. To design a more efficient gene correction system,
comprehensive studies on the effects of the second mismatches should be conducted, and are currently ongoing in our laboratory.
REFERENCES


Figure legends

Fig. 1. (A) Base sequence of the upstream region of the *rpsL* gene, (B) schematic presentation of the plasmids and the TD DNA fragments used in this study, and (C) procedures of the gene correction assay. (A) The TTTT sequence, where the –1 and +1 mutations were induced in the present study, is shown in bold. The –35 and –10 promoter sequences (–97 to –103 and –75 to –80, respectively) are underlined. The A base of the initiation codon is numbered as +1. The sequence of first 40 bases of the coding region is also shown. (B) The runs of Ts and the targeted regions (ATC in the plasmids and AGC in the TD fragment) are shown. More details on sequence are shown in Supplementary Fig. 1. (C) The TD fragment and the target plasmid (pSSW) were transfected into HeLa cells by lipofection. After 48 h, the plasmid DNA was recovered from the cells and was introduced into *E. coli* DH10B cells. The bacterial colonies on agar plates containing streptomycin were selected as ones containing corrected plasmid molecules.

Fig. 2. Gene correction efficiencies with the TD DNA fragment. The TD DNA fragment was co-transfected with the target plasmids at a 400:1 molar ratio into HeLa cells, and the plasmids recovered from the cells at 48 h post-transfection were used to transform the DH10B cells. The gene correction efficiencies were determined as described in the Materials and Methods section. Transfection experiments were performed six (pSSW(T3) and pSSW(T5)) and eight (pSSW(T4)) times. Data are expressed as the means ± standard errors. ***$P < 0.001$ vs. pSSW(T4).
Fig. 3. Sequence analysis of streptomycin-resistant colonies obtained in the (A, B) pSSW(T3), (C, D) pSSW(T4), and (E, F) pSSW(T5) experiments. (A, C, E) The indel mismatch site and (B, D, F) the target position. Results obtained with a reverse primer are shown in panels B and D.
A)

-120  GGATCGTTGT ATATTTCTTG  ACACCTTTTC  GGCATCGCCC  -81
-80  TAAAATTCGG  CGTCCTCATA  TTGTGTGAGG  ACGTTTTATT  -41
-40  ACGTGTTTAC  GAAGCAAAAAG  CTTAAAACCAG  GAGCTATTTA  -1
+1  ATGGCAACAG  TTAACCAGCT  GGTACGCAAA  CCACGTGCTC  +40

B)

tailed duplex  

\[ \begin{align*}
5' & \quad TTTT & \quad +238 & \quad +240 & \quad +529 & \quad +563 \\
3' & \quad & \quad AGC & \quad & \quad & \\
\end{align*} \]

pSSW(T3)  

\[ \begin{align*}
\cdots & \quad TTT & \quad \cdots & \quad \cdots & \quad \cdots & \quad \cdots \\
\cdots & \quad AAA & \quad \cdots & \quad ATC & \quad \cdots & \quad \cdots \\
\end{align*} \]

pSSW(T4)  

\[ \begin{align*}
\cdots & \quad TTTT & \quad \cdots & \quad \cdots & \quad \cdots & \quad \cdots \\
\cdots & \quad AAA & \quad \cdots & \quad ATC & \quad \cdots & \quad \cdots \\
\end{align*} \]

pSSW(T5)  

\[ \begin{align*}
\cdots & \quad TTTTT & \quad \cdots & \quad \cdots & \quad \cdots & \quad \cdots \\
\cdots & \quad AAAA & \quad \cdots & \quad ATC & \quad \cdots & \quad \cdots \\
\end{align*} \]

Fig. 1
HeLa cells

tailed duplex

target plasmid

lipofection

recovery of plasmid at 48 h post-transfection

transformation of DH10B cells

streptomycin-resistance selection

corrected plasmid

Fig. 1
Fig. 2

Gene correction efficiency (%)

pSSW(T3)

pSSW(T4)

pSSW(T5)

***

**pSSW(T4)***

**pSSW(T5)***
Fig. 3

A) 5'- T T T -3'

B) 3'- CGA -5'

C) 5'- T T T T -3'

D) 3'- CGA -5'
Fig. 3
Table 1. Sequence analysis of plasmids. a)

<table>
<thead>
<tr>
<th></th>
<th>pSSW(T3)</th>
<th>pSSW(T4)</th>
<th>pSSW(T5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>gene correction at codon 80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATC --&gt; AGC</td>
<td>41 (100) b)</td>
<td>31 (86) c, d)</td>
<td>33 (100) e, f)</td>
</tr>
<tr>
<td>other mutations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130A --&gt; T</td>
<td>0 (0)</td>
<td>1 (3)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>344C --&gt; T</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (3) f)</td>
</tr>
<tr>
<td>large insertion</td>
<td>0 (0)</td>
<td>18 (50) d)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>large deletion</td>
<td>0 (0)</td>
<td>1 (3)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>total colonies analyzed</td>
<td>41 (100)</td>
<td>36 (100)</td>
<td>33 (100)</td>
</tr>
</tbody>
</table>

a) All data are represented as cases found (%).
b) All colonies contained the TTT sequence.
c) All colonies contained the TTTT sequence.
d) Fifteen colonies contained a large insertion and the targeted sequence alteration.
e) All colonies contained the TTTTT sequence.
f) One colony contained the 344C --> T mutation and the targeted sequence alteration.