Gene targeting by single-stranded DNA oligonucleotides in mouse embryonic stem cells

Aarts, M.

2010

document version
Publisher's PDF, also known as Version of record

Chapter

5

Gene targeting by incorporation of unmodified single-stranded oligonucleotides in mouse ES cells does not induce DNA damage

Marieke Aarts
Hein te Riele

Submitted for publication.

Division of Molecular Biology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.

Gene targeting by single-stranded oligodeoxyribonucleotides (ssODNs) is a promising tool for introducing site-specific sequence alterations into the genome of mouse embryonic stem (ES) cells. We have developed an ES cell line that carries a mutant *EGFP* reporter gene to provide information about the stability of the introduced sequence alteration, possible toxicity of the ssODN and the viability of targeted cells. We found that gene targeting by unmodified ssODNs did not induce a substantial amount of DNA damage and only mildly affected the viability of targeted cells, which is in sharp contrast to targeting by ssODNs that were end-protected by phosphorothioate linkages. Short and long term analyses reveal that sequence alteration occurs through physical incorporation of the ssODN into the genome during DNA synthesis. We show that semi-conservative replication and cell division lead to a dilution of targeted cells. The use of unmodified ssODNs rather than end-protected ssODNs allows stable outgrowth of targeted cells and may facilitate routine application of this technique.

INTRODUCTION

Gene targeting by single-stranded oligodeoxyribonucleotides (ssODNs) is a promising tool for introducing site-specific sequence alterations into the genome of mammalian cells. The procedure only requires short synthetic ssODNs that are homologous to the target locus, except for a single or a few centrally located bases that comprise the desired genetic alteration.

Although the exact mechanism underlying ssODN-mediated gene targeting remains elusive, several reports have demonstrated the involvement of DNA replication. Synchronizing cells in the S phase of the cell cycle or reducing the rate of replication fork progression improved the targeting frequency in various cell types [1-4]. It has therefore been postulated that the ssODN anneals to its complementary genomic target sequence within the context of a replication fork. Extension of the ssODN by the replication machinery would result in incorporation of the ssODN into the newly synthesized DNA strand [5,6]. Following annealing, DNA mismatches are formed between the ssODN and its chromosomal complement. We have previously shown that these mismatches are recognized and removed by the DNA mismatch repair (MMR) system, thereby preventing introduction of the desired genetic alteration. Deletion or down-regulation of the central MMR gene *Msh2* dramatically improved the targeting frequency in mouse embryonic stem (ES) cells [7,8]. Recently, this inhibitory effect of the MMR system has been confirmed in human hepatocytes [9], CHO-K1 cells [10] and HEK293T Lα cells [11]. The involvement of the MMR system, which functions to correct replication errors that escape proofreading, provides additional proof that ssODN-mediated gene targeting takes place during replication.

Mutant versions of *EGFP* or *neo* reporter genes have been used to quantify the efficacy of ssODN-mediated gene correction. In many mutant *EGFP* reporter cell lines, anti-sense ssODNs complementary to the non-transcribed strand were most effective [1,3,12,13]. Conversely, in mutant *neo* reporter cell lines, sense ssODNs complementary to the transcribed strand showed the highest targeting frequencies [14,15]. In *Escherichia coli* it was shown that ssODNs

corresponding in sequence to the lagging strand performed better than ssODNs corresponding to the leading strand [5,16,17]. Thus, the strand bias may be determined by the direction of the replication fork through the target locus and therefore vary between different target loci.

Besides the polarity of the ssODN, the composition of the ssODN may also influence the efficacy of ssODN-mediated gene targeting. Protection of ssODNs against nucleolytic degradation by phosphorothioate (PTO) linkages [10,13] or locked nucleic acid (LNA) bases [18] seemed to enhance the targeting frequency in various *EGFP* reporter cell lines. However, we observed that PTO modifications decreased the targeting frequency in mouse ES cells carrying a mutant *neo* reporter gene [7,15]. Several groups have shown that correction by PTO-modified ssODNs triggered a DNA damage response, including activation of the ATM/ATR pathway and phosphorylation of histone H2AX, due to the formation of genomic double-stranded DNA breaks (DSBs) [11,12,19]. As a result, the majority of the cells that were initially corrected by PTO-modified ssODNs, arrested in the S [19] or G_2 phase [2,10,11] of the cell cycle and were not able to expand. It is unclear whether this type of toxicity is restricted to chemically-modified ssODNs and whether it also occurs in mouse ES cells. These issues are important as ES cells modified by ssODN-mediated gene targeting, need to retain their pluripotency and germ line competence to allow the generation of mutant mice [8,20].

In the present study, we have addressed the discrepancies between the mutant *EGFP* and mutant *neo* reporter cell lines with respect to the optimal polarity and the chemical composition of the ssODN. We have generated a novel *Msh2*-deficient mouse ES cell line that carries a single copy of a mutant *EGFP* reporter gene at the same chromosomal position as the previously described mutant *neo* reporter gene [7,8,15]. In contrast to the *neo* reporter cell line, this *EGFP* reporter cell line enabled us to monitor gene correction events already 24 hrs after ssODN exposure and to study the fate of corrected cells upon further culturing. By this approach, we have obtained better insight into the mechanism of ssODN-mediated gene targeting and demonstrate that sequence alteration occurs through incorporation of the ssODN into the genome during DNA replication. Importantly, gene targeting by unmodified ssODNs only mildly affected the viability of the targeted cells, which may facilitate future therapeutic applications.

RESULTS

Reporter cell lines

To monitor the frequency of ssODN-mediated gene targeting, we used our previously described neomycin (*neo*) reporter system in which the *neo* gene was mutated by a single base substitution in the start codon (ATG to AAG; Figure 1A). In addition, we constructed a mutant *EGFP* (*mEGFP*) reporter gene of which the translational start site was replaced by the mutant *neo* start sequence (Figure 1A). Single copies of the mutant *neo* and the *mEGFP* genes were stably integrated into the *Rosa26* locus of *Msh2*-deficient mouse embryonic stem (ES) cells [7]. ssODNs of 35-37 nt containing one (1m) or four (4N) centrally-located nucleotide substitutions were used to correct the mutant start sequence or to generate a novel ATG start codon, respectively. Correction of the *neo* reporter gene was monitored by the appearance of G418-resistant colonies after 8-10 days, while correction of the *mEGFP* reporter resulted in EGFP fluorescence which could be detected 24 hrs after ssODN transfection using flow cytometry (Figure

1B). These reporter systems allowed us to compare the frequency of ssODN-mediated gene targeting in mouse ES cells immediately and several days after exposure to ssODNs, providing information about the stability of the introduced sequence alteration, possible toxicity of the ssODNs and the viability of corrected cells.

Recovery of gene correction events varies with the time of readout

We used the *mEGFP* reporter gene to accurately compare the targeting efficacy of unmodified and chemically-modified ssODN 1m (Figure 2A) and ssODN 4N (Figure 2B). As previously shown for other mutant *EGFP* reporter cell lines [1,3,12,13], we found

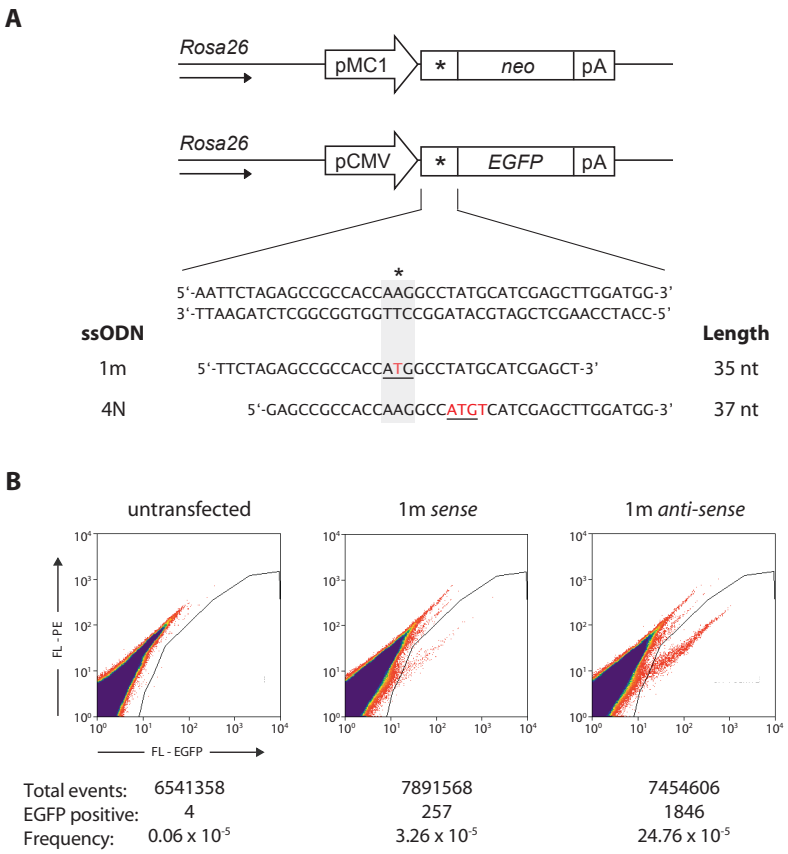


Figure 1 Correction of a mutant EGFP reporter gene by ssODNs.
*(**A**) Sequence of the mutant neomycin (neo) and EGFP reporter genes in which the start codon (ATG) is replaced for AAG (*). Activity of the reporter genes can be restored by single-stranded oligodeoxyribonucleotides (ssODNs) that substitute one (1m) or four nucleotides (4N) (indicated in red) to create a new ATG start codon (underlined). Sequences of sense ssODNs complementary to the transcribed strand are shown. Arrow indicates the direction of the Rosa26 promoter. (**B**) Representative flow cytometry dot plots of Msh2$^{-/-}$ mEGFP cells either untransfected (left panel) or transfected with ssODN 1m sense (middle panel) or 1m anti-sense (right panel) analyzed 24 hrs after ssODN transfection.*

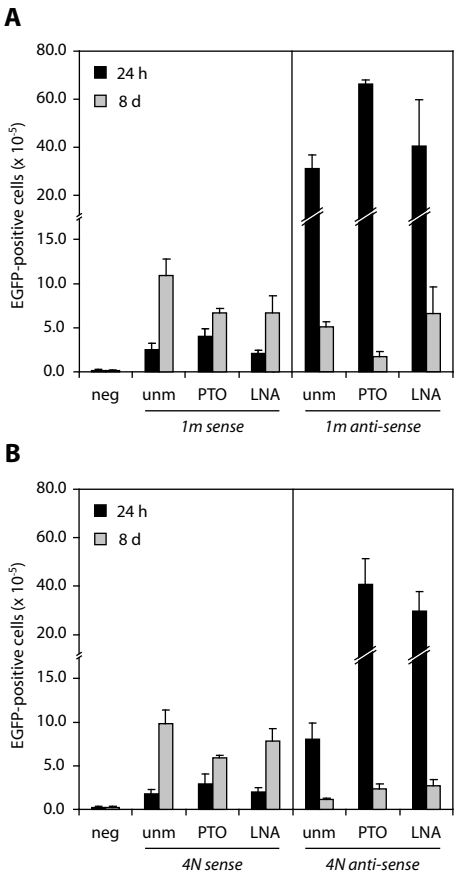


Figure 2 Gene correction frequencies by different types of ssODNs in Msh2-deficient mEGFP ES cells.

*(**A**) Unmodified (unm), PTO-modified and LNA-modified sense and anti-sense ssODN 1m were used to correct the mEGFP gene in Msh2-deficient ES cells. The number of EGFP-positive cells was determined at 24 hrs (black bars) and 8 days (grey bars) after ssODN transfection. (**B**) Correction of the mEGFP gene by unmodified (unm), PTO-modified and LNA-modified sense and anti-sense ssODN 4N. Error bars represent the standard deviation (s.d.) of three independent experiments.*

that anti-sense ssODNs appeared to be 5- to 20-fold more effective than sense ssODNs, when the number of EGFP-positive cells was determined 24 hrs after ssODN exposure (Figure 2A and 2B, black bars). Addition of three flanking phosphorothioate (PTO) linkages or a single locked nucleic acid (LNA) base at each end to protect the ssODN against nucleolytic degradation improved the frequency, in particular of anti-sense ssODNs. Remarkably however, when EGFP expression was analyzed 8 days after ssODN exposure, the targeting frequency of anti-sense ssODNs had dramatically decreased, whereas the targeting frequency of sense ssODNs had increased (Figure 2A and 2B, grey bars). The strand bias was now in favour of the sense ssODNs, which is in agreement with our previous results using the *neo* reporter system [15]. Particularly, PTO-modified anti-sense ssODNs, although being the most effective at 24 hrs after ssODN exposure, showed the strongest reduction in the number of EGFP-positive cells after 8 days (28-fold *versus* 8-fold reduction for unmodified anti-sense ssODNs).

Physical incorporation of ssODNs during DNA replication

Different models have been proposed for the mechanism of ssODN-mediated gene targeting (reviewed in [21]). In one model (Figure 7A, model I), the ssODN anneals to its chromosomal complement and serves as a template for repair of this chromosomal DNA strand. If this were the case, anti-sense ssODNs would stimulate substitution of nucleotides at the non-transcribed strand, whereas sense ssODNs would lead to alteration of the transcribed strand. In an alternative model (Figure 7A, model II), the ssODN becomes physically integrated into the genome within the context of either a D-loop or a replication fork. In this case, the reverse would take place: anti-sense ssODNs would lead to sequence alteration of the transcribed strand and sense ssODNs to alteration of the non-transcribed strand. Copying the genetic alteration to the opposite strand may take more time, possibly even a next round of DNA replication. If

so, correction of the transcribed strand will lead to immediate EGFP expression, whereas correction of the non-transcribed strand may require an extra round of DNA replication to copy the base alteration(s) to the transcribed strand before EGFP expression can be detected.

To discriminate between these two models, we quantified the number of EGFP-positive cells every 24 hrs over a period of 8 days. As shown in Figure 3A (black bars), the targeting frequency of sense ssODN 1m was only 3.0 ± 1.6 at the 24 hrs time point (day 1), but sharply increased to a maximum of 14.5 ± 1.9 x 10^{-5} on day 2. During the next two days, the number of EGFP-positive cells slightly decreased and stabilized at a frequency of 11.0 ± 1.5 x 10^{-5} on day 4. Anti-sense ssODN 1m showed the maximum targeting frequency (35.0 ± 6.2 x 10^{-5}) on day 1 (Figure 3B, black bars), after which the number of EGFP-positive cells gradually decreased until day 4-5 and stabilized at a frequency of 4.5 ± 0.4 x 10^{-5}. These observations indicate that the genetic modification induced by sense ssODNs required an extra round of DNA replication to become expressed, supporting the model that the sense ssODNs became physically integrated into the non-transcribed strand. The immediate EGFP-positivity observed with the anti-sense ssODNs indicates that these ssODNs integrated into the transcribed strand.

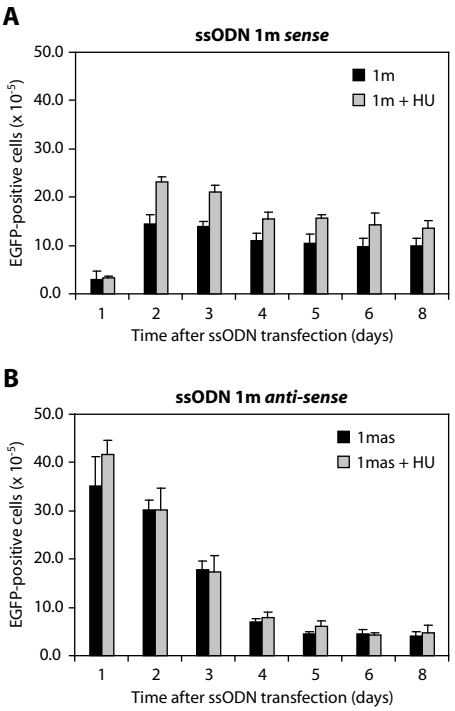


Figure 3 Time course of recovery of EGFP-positive cells obtained by unmodified ssODN 1m.
*$Msh2^{-/-}$ mEGFP ES cells were exposed to ssODN 1m without treatment (black bars), or in the presence of 50 μM hydroxyurea (HU; grey bars). The number of EGFP-positive cells was quantified by flow cytometry at the indicated time points after transfection. Error bars represent the s.d. of four independent experiments. (**A**) Correction by unmodified sense ssODN 1m (1m). (**B**) Correction by unmodified anti-sense ssODN 1m (1mas).*

We have previously postulated that ssODN-mediated gene targeting in mouse ES cells occurs within the context of a replication fork [15]. The different accessibility of the leading and the lagging strand during DNA replication could explain the preferential incorporation of sense over anti-sense ssODNs after 8 days (Figure 3). Moreover, this model provides a simple explanation for the decline of EGFP-positive cells obtained following exposure to anti-sense ssODNs (Figure 3B). Incorporation of anti-sense ssODNs into the transcribed strand immediately provides a template for EGFP production, probably already in the G_2 phase of the cell cycle (Figure 7C). However, in these cells only one of the four single DNA strands contains the corrected start codon. Subsequent semi-conservative DNA replication and cell division will gradually dilute the number of EGFP-positive cells, theoretically 4-fold. In contrast, cells corrected

by sense ssODNs, need to replicate their DNA once to transmit the genetic modification to the transcribed strand (Figure 7B), which is reflected by the sharp increase in EGFP-positive cells on day 2 (Figure 3A). The modest decline of EGFP-positive cells, which we reproducibly observed at day 4 (Figure 3A), indeed indicates that EGFP was expressed in G_2 cells immediately after the corrected start codon was transmitted to the transcribed strand. After mitosis, this led to green cells that did not contain the corrected *EGFP* gene and were therefore no longer detectable by flow cytometry upon further culturing (Figure 7B).

We have previously shown that exposure of cells to hydroxyurea, which slows down replication fork progression by depleting cellular dNTP pools through inhibition of ribonucleotide reductase [22], enhanced the targeting frequencies in *Msh2*$^{-/-}$ *neo* ES cells. Reducing the rate of DNA synthesis by hydroxyurea increases the exposure of the single-stranded DNA regions that are present during replication, particularly in the lagging strand. Here, we investigated the effect of hydroxyurea on the frequency of sense and anti-sense ssODN-mediated correction of the *mEGFP* reporter over a time course of 8 days. *Msh2*$^{-/-}$ *mEGFP* ES cells were incubated with 50 µM hydroxyurea between 6 hrs before and 22 hrs after ssODN exposure. For sense ssODN 1m, hydroxyurea treatment resulted in a 1.5-fold increase in targeting frequency (Figure 3A, grey bars), confirming our results obtained with the *neo* reporter [15]. However, the targeting frequencies of anti-sense ssODN 1m were not significantly altered upon hydroxyurea treatment (Figure 3B, grey bars). The observation that hydroxyurea augmented the accessibility of the target locus to sense ssODNs, strongly suggests that these ssODNs were incorporated into the lagging strand.

Unmodified ssODNs only mildly reduced the survival of corrected cells

The gradual loss of EGFP-positive cells upon gene correction by anti-sense ssODNs as shown in Figure 3B, has also been observed by others [2,11,19]. In these experiments, cells corrected by PTO-modified ssODNs showed a reduced viability due to high levels of DNA damage. Indeed, we observed a much stronger reduction of EGFP-positive cells after 8 days when gene correction was obtained with PTO-modified ssODNs than with unmodified ssODNs (Figure 2). This strong reduction can certainly not be explained by semi-conservative replication alone. Also, the loss of EGFP-positive cells obtained with unmodified anti-sense ssODNs seemed to be somewhat larger than the 4-fold reduction predicted by semi-conservative DNA replication. We therefore investigated whether ssODN-mediated gene targeting *per se* affected the viability of cells.

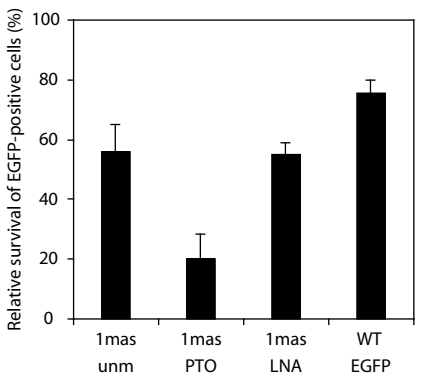


Figure 4 Relative colony survival of EGFP-positive cells.

Colony-forming ability of EGFP-positive Msh2$^{-/-}$ *mEGFP ES cells after correction by unmodified (unm), PTO-modified and LNA-modified anti-sense ssODN 1m (1mas). Msh2*$^{-/-}$ *ES cells carrying a wild-type (WT) EGFP gene were used as control. The colony survival of the corrected EGFP-positive cells was normalized to the colony survival of uncorrected EGFP-negative cells that had been exposed to the same ssODNs. Error bars represent the s.d. of two independent experiments.*

Twenty-four hours after exposure of *Msh2*$^{-/-}$ *mEGFP* ES cells to various types of anti-sense ssODN 1m, EGFP-positive and -negative cells were separated by flow cytometry and plated to determine their colony-forming ability. For all types of ssODNs, the viability of EGFP-positive cells was reduced compared to EGFP-negative cells (Figure 4). However, *Msh2*$^{-/-}$ ES cells expressing a wild-type (WT) *EGFP* gene showed 75% survival compared to EGFP-negative *Msh2*$^{-/-}$ *mEGFP* reporter cells, indicating that EGFP expression itself already reduced cell viability. Twenty-four hours after exposure to unmodified or LNA-modified ssODNs, 55% of the corrected cells were able to form colonies. In contrast, only 20% of the EGFP-positive cells obtained after correction by PTO-modified ssODNs survived. Thus, the apparent loss of cells corrected by unmodified anti-sense ssODNs observed in Figure 3B can largely be explained by dilution of the corrected DNA strand due to semi-conservative DNA replication and to a minor extent by a reduced viability of the targeted cells. In contrast, the loss of cells corrected by PTO-modified ssODNs seems to be mainly caused by a strong reduction in cell viability.

ssODN uptake did not affect survival

The reduced viability of corrected cells may be due to toxicity of high levels of ssODNs in cells. To investigate whether unmodified ssODNs conferred toxicity, we determined the cell viability as a function of the level of ssODN uptake. *Msh2*$^{-/-}$ *neo* cells were exposed to 5'-Cy5-labelled sense ssODN 4N and sorted into four equal bins based on their Cy5 fluorescence signal 24 hrs after transfection. After sorting, part of the cells was plated to assess their colony-forming ability, while the remainder of the cells was subjected to G418 selection to determine the targeting frequency. Although the sorting procedure reduced the colony-forming ability of ES cells, there was no correlation between cell viability and the level of Cy5-ssODN uptake (Figure 5A). Thus, cells with a high Cy5-ssODN uptake (bin 4) were as

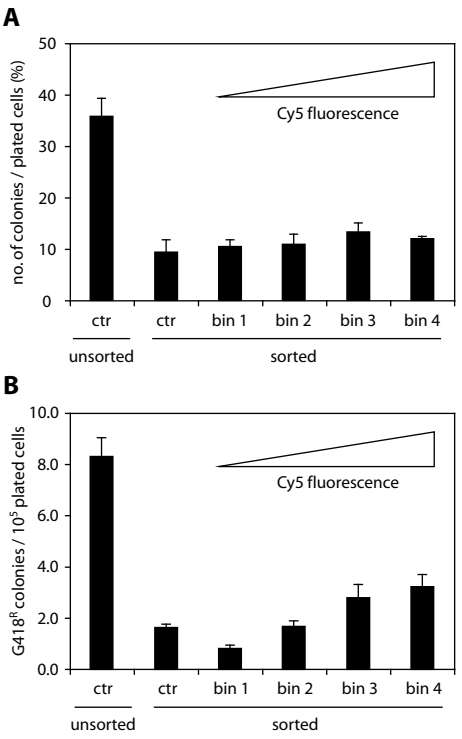


Figure 5 Effect of the level of ssODN uptake on colony survival and targeting frequency.
Msh2$^{-/-}$ *neo ES cells were transfected with Cy5-labelled sense ssODN 4N and sorted into four equal bins based on their Cy5 fluorescence signal 24 hrs after ssODN transfection (bin 1, 25% of cells with the lowest Cy5-ssODN uptake; bin 4, 25% of cells with the highest Cy5-ssODN uptake). Msh2*$^{-/-}$ *neo ES cells transfected with Cy5-labelled sense ssODN 4N, either unsorted or sorted but not binned, were included as controls (ctr). (**A**) Colony survival of cells that were binned by their Cy5-ssODN uptake. (**B**) Targeting frequency of cells that were binned by their Cy5-ssODN uptake. Targeting frequency is the number of G418-resistant colonies per 10*^{5} *cells that were plated after sorting. Note that the calculated targeting frequencies after sorting are lower due to the reduced survival of the plated cells. Error bars represent the s.d. of two independent sorting experiments consisting of two replicate runs per sort.*

viable as cells with a low uptake (bin 1), indicating that the presence of unmodified ssODNs in cells did not affect cell viability. On the other hand, increased Cy5-ssODN uptake did result in an increased targeting frequency (Figure 5B), as reported previously [14]. Cells with the highest Cy5-ssODN uptake (bin 4) showed a 4-fold higher targeting frequency than cells with the lowest Cy5-ssODN uptake (bin 1).

Unmodified ssODNs did not induce DNA damage in ES cells

It has been reported that correction by PTO-modified ssODNs resulted in an increased level of genomic double-stranded DNA breaks (DSBs) [11,12,19]. We have performed the neutral comet assay to determine whether exposure to or incorporation of unmodified ssODNs induced DSB formation. *Msh2*$^{-/-}$ *mEGFP* ES cells were exposed to unmodified or PTO-modified anti-sense ssODN 1m for 24 hrs. Subsequently, the EGFP-positive and -negative cells were separated by flow cytometry and subjected to the neutral comet assay [23]. The amount of DNA damage was quantified by the tail moment (TM), which represents the product of the tail length and the percentage of DNA in the comet tail (Figure 6A). After exposure to unmodified anti-sense ssODN 1m, there was no significant difference in the median TM value of the corrected EGFP-positive cells and the uncorrected EGFP-negative cells (8.4 and 11.4, respectively) (Figure 6B). Moreover, mock-transfected *Msh2*$^{-/-}$ *mEGFP* cells and *Msh2*$^{-/-}$ *WT EGFP* control cells showed comparable TM values (9.4 and 10.6, respectively), indicating that exposure to or incorporation of unmodified anti-sense ssODN 1m had not induced a substantial amount of DNA damage. In contrast, the EGFP-positive cells that were corrected by PTO-modified anti-sense ssODN 1m showed high levels of DSBs (median TM = 39.4). A similarly high level of DSBs was found in control *Msh2*$^{-/-}$ *mEGFP* cells that had been exposed to 25 Gy of γ-irradiation (median TM = 39.2). After exposure to PTO-modified ssODNs, even the uncorrected EGFP-negative cells showed elevated TM values (median TM = 25.0), suggesting that both exposure to and incorporation of PTO-modified ssODNs led to DNA damage.

In summary, we have shown that correction by unmodified ssODNs did not induce a substantial amount of DNA damage and only marginally reduced the viability of mouse ES cells, in sharp contrast to correction by PTO-modified ssODNs. Our results clearly demonstrate that gene correction was mediated by incorporation of the ssODN into the genome during DNA replication. Therefore, the decline of EGFP-positive cells corrected by unmodified anti-sense ssODNs was largely the consequence of semi-conservative replication rather than deleterious effects of ssODN exposure or incorporation.

DISCUSSION

Many reports have demonstrated the feasibility of ssODN-mediated gene targeting using mutant *EGFP* reporter systems in a variety of cell lines [2,10-13,19]. Using these *EGFP* reporters some general conclusions were drawn: (1) anti-sense ssODNs complementary to the non-transcribed strand, were most effective; (2) PTO modifications improved the targeting frequency; (3) most corrected cells arrested in the G_2 phase of the cell cycle and were not able to form stable colonies. Remarkably, these conclusions did not match the results we have obtained previously using a mutant *neo* reporter ES cell line, in which the formation of G418-resistant colonies served as a readout for the targeting frequency [15]. In the present study, we have generated a novel *Msh2*-deficient mouse ES cell line that carries a

single copy integration of a mutant *EGFP* reporter gene at exactly the same position in the *Rosa26* locus as the previously described mutant *neo* reporter gene [15]. This *EGFP* reporter cell line enabled us to investigate the discrepancies between the *neo* and *EGFP* reporter systems and to gain more insight into the mechanism of ssODN-mediated gene targeting in ES cells. We demonstrate that (1) the apparent superiority of anti-sense ssODNs over sense ssODNs was a consequence of early readout; (2) PTO-modified ssODNs were deleterious to cells, while unmodified ssODNs only mildly affected cell viability; (3) unmodified ssODNs did not induce DNA damage.

Superiority of anti-sense ssODNs over sense ssODNs is a consequence of early readout

Using the *neo* reporter cell line, we have shown that the targeting frequency of sense ssODN 1m was $7.0 \pm 2.8 \times 10^{-5}$, while the targeting frequency of anti-sense ssODN 1m was $3.7 \pm 3.0 \times 10^{-5}$ [15]. Here, using the novel *mEGFP* reporter ES cell line, similar targeting frequencies were obtained when the number of EGFP-positive cells was determined 8 days after ssODN exposure

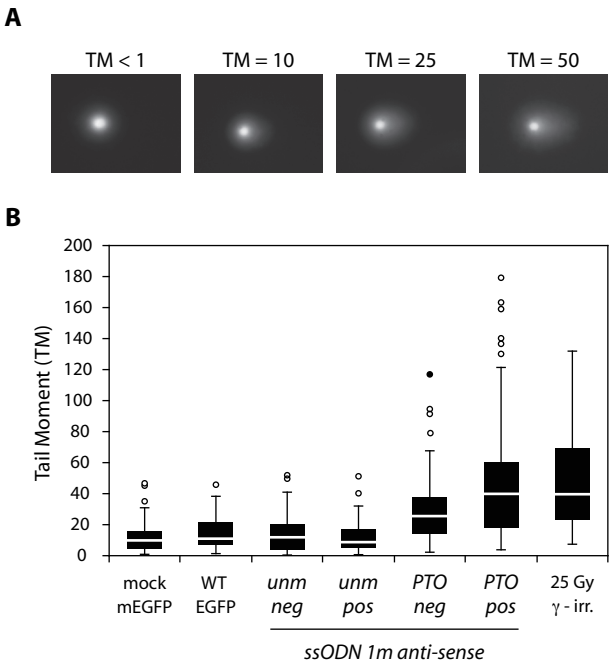


Figure 6 Quantification of DNA damage after ssODN-mediated gene targeting using the neutral comet assay.

*(**A**) Representative images from the neutral comet assay performed on ssODN-transfected Msh2$^{-/-}$ mEGFP ES cells with various levels of DNA damage as quantified by the tail moment (TM; %DNA in tail x tail length). (**B**) Box plot representing the distribution of the comet tail moments after exposure to unmodified (unm) and PTO-modified anti-sense ssODN 1m. Corrected EGFP-positive (pos) and uncorrected EGFP-negative (neg) cells were separated by flow cytometry 24 hrs after ssODN exposure and the level of DNA damage was quantified using the neutral comet assay. Mock-transfected Msh2$^{-/-}$ mEGFP cells (without ssODN), Msh2$^{-/-}$ WT EGFP cells, and Msh2$^{-/-}$ mEGFP cells exposed to 25 Gy of γ-irradiation were included as controls.*

(Figures 2A and 3), indicating that both reporter systems provided an accurate readout of the targeting frequency. Strikingly, in the *mEGFP* reporter cell line the recovery of gene correction events was strongly influenced by the time of readout: initially anti-sense ssODNs performed best, while later sense ssODNs were most effective. We postulate that the initial strand bias indicates that ssODNs become incorporated into the genome: by incorporating into the transcribed strand and replacing the original genetic information, anti-sense ssODNs allow immediate detection of the corrected *EGFP* gene. In contrast, incorporation of the sense ssODNs into the non-transcribed strand requires the genetic alteration to be transferred to the transcribed strand before *EGFP* can be expressed (Figure 7B). Since the delay in EGFP expression was approximately one day (Figure 3A), it is most likely that this process occurred during DNA replication (Figure 7A, model II). As we have hypothesized before [15], the ultimate superiority of sense ssODNs over anti-sense ssODNs may be the result of the differential accessibility of the leading and lagging strand during replication. In our reporter system, sense ssODNs seem to be incorporated into the lagging strand.

Unmodified ssODNs only mildly reduced the viability of corrected cells and did not induce DNA damage

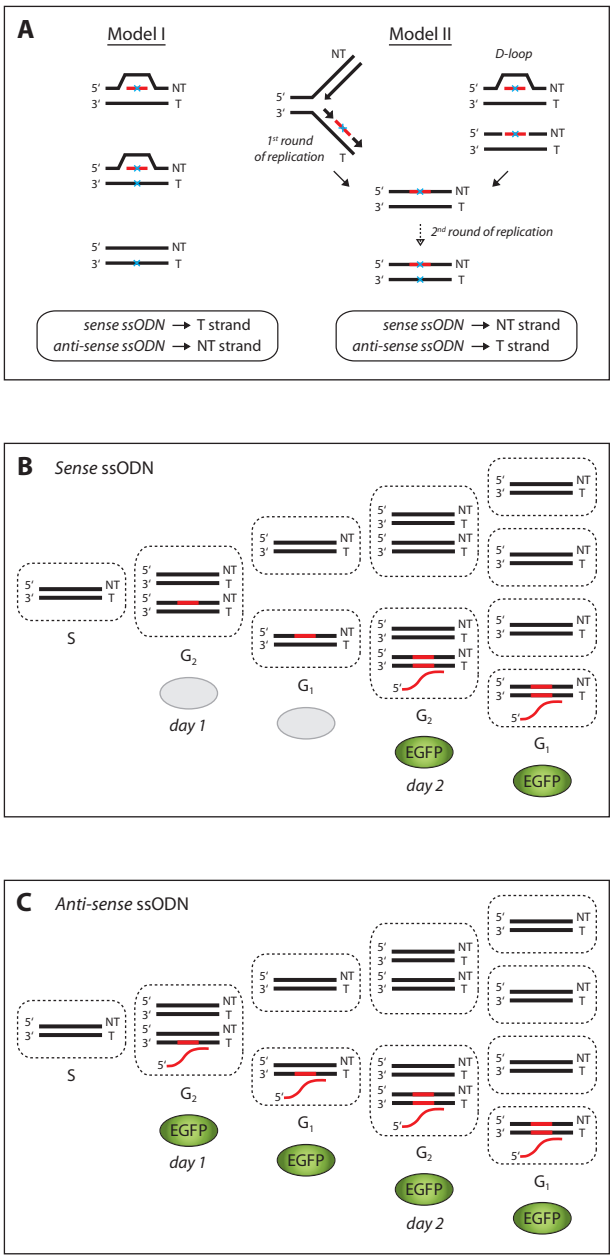
The addition of PTO linkages enhanced the targeting frequency of particularly anti-sense ssODNs at 24 hrs after transfection (Figure 2), probably by protecting the ssODNs against nucleolytic degradation. However, after 8 days, PTO-modified anti-sense ssODNs showed the most prominent decrease in targeting frequency. In agreement with previous results [10,13], cells corrected by PTO-modified ssODNs showed a strongly reduced colony-forming ability, which is in contrast to cells corrected by unmodified ssODNs (Figure 4). This reduced viability has been attributed to an increased level of DSBs resulting in a G_2 cell cycle arrest [11,12]. Due to the low targeting frequencies in mouse ES cells, we could not determine the cell cycle profile of corrected cells. Nevertheless, our data demonstrate that unmodified ssODNs, unlike PTO-modified ssODNs, did not evoke DNA damage in mouse ES cells. First, the survival of cells corrected by unmodified ssODNs was only mildly affected when compared to cells expressing a *WT EGFP* gene (Figure 4). Second, the level of uptake of unmodified ssODNs did not affect the cell viability (Figure 5A). Third, EGFP-positive cells corrected by unmodified ssODNs showed similar low levels of DSBs as uncorrected EGFP-negative cells, as determined by the neutral comet assay (Figure 6). Our results imply that the observed DNA damage response was specific for gene correction by PTO-modified ssODNs, although the mechanism is still largely unknown. On the one hand, PTO-modified ssODNs may titrate repair factors away from chromosomal damage. This could explain the increased level of DSBs in the EGFP-negative cells that had been exposed to PTO-modified ssODNs 1m anti-sense but were not corrected (Figure 6B). On the other hand, integration of PTO-linked nucleotides may provide a barrier to the replication machinery and lead to fork stalling and DSB formation. The deleterious effects of PTO-modified ssODNs do not seem to be caused by the absence of DNA MMR. Previously, we have shown that PTO modifications reduced the targeting frequency to a similar extent in both wild-type and *Msh2*-deficient ES cells carrying the mutant *neo* gene [15]. In addition, MSH2 down-regulation [10] or MLH1 status [11] did not influence the proliferative capacity of cells after correction by end-protected PTO-ssODNs.

Figure 7 Model for ssODN-mediated correction of the mEGFP gene.

*(**A**) Hypothetical models for gene correction by sense ssODNs (indicated in red) in mouse ES cells. Model I: the ssODN anneals to its chromosomal complement and serves as a template for repair of this chromosomal DNA strand. If so, sense ssODNs would stimulate substitution of nucleotides (indicated by blue X) of the transcribed strand allowing immediate expression of EGFP. Model II: the ssODN becomes integrated into the genome within the context of either a replication fork or a D-loop. In this model, sense ssODNs would lead to sequence alteration of the non-transcribed strand. A second round of DNA replication is required to transfer the genetic alteration to the opposite strand to allow EGFP expression. Our findings are most compatible with integration of the ssODN during DNA replication (Model II). NT, non-transcribed strand; T, transcribed strand.*

*(**B**) Incorporation of sense ssODNs (indicated in red) into the non-transcribed strand during S phase requires an extra round of replication to transmit the corrected EGFP sequence to the transcribed strand (day 2). After cell division, both daughter cells appear EGFP-positive although only one contains the corrected EGFP sequence.*

*(**C**) In contrast, incorporation of anti-sense ssODNs (indicated in red) into the transcribed strand immediately provides a template for EGFP production, probably already in the G_2 phase of the cell cycle (day 1). In these cells only one of the four DNA strands contains the corrected EGFP start codon. Subsequent semi-conservative DNA replication and cell division will gradually dilute the number of EGFP-positive cells, theoretically four-fold.*



Semi-conservative replication leads to an apparent loss of corrected cells

Our data strongly indicate that ssODNs become incorporated into the newly synthesized DNA strand during replication, consistent with previous findings [5,6]. After incorporation of the ssODN during replication, the corrected DNA template is transmitted to only one daughter cell. Then, another cell division is required to yield a daughter cell that carries the corrected DNA template in both DNA strands. Hence, after two consecutive cell divisions only one out of four daughter cells is stably corrected. For anti-sense ssODNs, this will lead to a 75% reduction in targeting frequency within two cell divisions after *EGFP* correction (Figure 7C). For sense ssODNs, it takes already one cell division before a corrected cell becomes EGFP-positive and therefore only a 50% reduction in targeting frequency would be expected (Figure 7B). Because EGFP has a half-life of ~24 hrs [24], the fluorescent signal may be transmitted to both daughter cells and slowly fade in one cell, while the other carries the corrected DNA template and will stably express EGFP fluorescence. This long half-life of EGFP may mask and seemingly delay the decrease in targeting frequency. Besides the reduction due to semi-conservative replication, EGFP-positive cells also showed a decreased colony survival compared to EGFP-negative cells (Figure 4). This may have further reduced the long-term recovery of initially corrected EGFP-positive cells by anti-sense ssODNs (day 8 *versus* day 1; Figure 3B). This reduction in long-term recovery appeared to be less pronounced for cells corrected by sense ssODNs (day 8 *versus* day 2; Figure 3A). Possibly, this difference may be attributed to a bias towards viable cells: cells that were not able to divide after correction of the non-transcribed strand by sense ssODNs did not become EGFP-positive and were therefore not observed.

In conclusion, we have developed a unique set of reporter cell lines carrying a single copy of either a mutant *neo* or a mutant *EGFP* gene at the same chromosomal position to study the short- and long-term effects of ssODN-mediated gene targeting. Using these cell lines, we could demonstrate that the strand bias in favour of anti-sense ssODNs was a consequence of early readout. Thus, reporter systems based on EGFP expression allow rapid recovery of gene correction events, but may be more prone to misinterpretation. We provide strong evidence that ssODNs are incorporated during DNA replication. Importantly, we have shown that unmodified ssODNs do not induce a substantial amount of genomic DSBs neither through incorporation into the target locus nor via titration of DNA repair factors. Our data imply that the apparent loss of targeted cells obtained by anti-sense ssODNs was merely the consequence of semi-conservative replication and is therefore inherent to the targeting procedure. The use of unmodified ssODNs for gene targeting ensures stable outgrowth of targeted cells, facilitating application of this technique.

ACKNOWLEDGEMENTS

We are grateful to Frank van Diepen and Anita Pfauth from the flow cytometry facility for their excellent technical support. We thank Sietske Bakker, Henri van de Vrugt and Eva Wielders for critical reading of the manuscript. This work was supported by grants from the Netherlands Genomics Initiative [Horizon Breakthrough 050-71-007, Horizon 050-71-051].

MATERIALS AND METHODS

Culture conditions

129/Ola-derived E14-IB10 ES cells were cultured on MEF feeders in Glasgow minimal essential medium (GMEM) supplemented with 10% fetal calf serum (FCS), 1mM sodium pyruvate, 1x non-essential amino acids, 1 mM 2-mercaptoethanol and 1000 U per ml of leukemia inhibitory factor. For transfections and antibiotic selections, ES cells were cultured onto gelatin-coated plates in BRL (Buffalo rat liver cells)-conditioned medium.

Neo reporter cell line

We developed a selectable mutant neomycin (*neo*) reporter cell line in which the start codon of the *neo* resistance gene was mutated from ATG to AAG (Figure 1). A single copy of the mutant *neo* reporter gene was stably integrated into the *Rosa26* locus of *Msh2*-deficient ES cells as described previously [7].

EGFP reporter cell lines

Plasmid pEGFP-N1 carries the enhanced green fluorescent protein (EGFP) ORF under the control of the CMV IE promoter/enhancer (Clontech). The coding sequence of *EGFP* without start codon was amplified by PCR from pEGFP-N1 using primers 5'-CGAGGTACCATGTGAGCAAGGGCGAGGA-3' and 5'-CTAGAGTCGCGGCCGCTTTAC-3' to introduce *Kpn*I and *Not*I sites (underlined). To generate a mutant *EGFP* start region, we annealed two synthetic DNA oligonucleotides encoding the start region of the mutant *neo* reporter in order to obtain a double-stranded DNA fragment with *Bgl*II and *Kpn*I overhangs (5'-GATCTGACCCAATTCTAGAGCCGCCACC**AAG**GCCTATGCATCGAGCTTGGATGGATTGCACGCAGGGTAC-3'). The mutant *EGFP* start region was ligated together with the *Kpn*I/*Not*I-digested *EGFP* PCR product (without start region) into the *Bgl*II and *Not*I sites of pEGFP-N1. Similarly, a plasmid with a wild-type start region (**ATG** instead of **AAG**) was constructed for control purposes. Subsequently, the 1.7-kb pCMV-EGFP-SV40pA fragment was amplified using the Expand Long Template PCR system (Roche) and primers 5'-CCGCTCGAGATTAATAGTAATCAATTACGGGGT-3' and 5'-CCGCTCGAGCTTAAGATACATTGATGAGTTTGG-3'and inserted into the *Xho*I site of the *Rosa26-His* targeting vector. A single copy of the mutant and wild-type *EGFP* gene was stably integrated into the *Rosa26* locus of *Msh2*-deficient ES cells as described previously [7]. Single copy integration was confirmed by Southern blot analysis.

Transfection

ssODNs were transfected following the TransFast-mediated transfection method described earlier [8]. Briefly, 7 x 10^5 ES cells were seeded onto a gelatin-coated 6-well in BRL-conditioned medium the day before transfection. For one well, 3 µg of ssODNs and 27 µl of TransFast transfection agent (Promega) were diluted in 1.4 ml serum-free medium and incubated for 15 min at room temperature. After 75 min of exposure to the transfection mixture at 37 °C, 4 ml of BRL-conditioned medium with serum was added to the cells. Cells were harvested for flow cytometry at the indicated time points after transfection. The targeting frequency is the number of EGFP-positive cells per 10^5 cells that were analyzed by flow cytometry. Cells were exposed to 50 µM hydroxyurea (Sigma-Aldrich) for 6 hrs prior to ssODN transfection and for another 22 hrs during ssODN exposure.

Oligonucleotides

ssODNs were designed to correct the mutant start region of the *neo* or the *EGFP* reporter gene by introducing an in-frame ATG start codon (sequences of sense ssODNs are listed in Figure 1). Anti-sense ssODNs introduce identical base changes but have opposite polarity (*i.e.*, are complementary to the non-transcribed strand). When indicated, ssODNs were modified with either three PTO linkages or a single LNA base on each end to protect them against nucleolytic degradation. ssODNs were purchased from Sigma-Aldrich and Eurogentec.

Flow cytometry

Quantification of EGFP-positive cells was performed on a CyAn ADP flow cytometer using Summit V4.3.01 software for analysis (DAKO Cytomation). Cells were harvested 24 hrs after transfection and resuspended in PBS supplemented with 0.1% FCS and DAPI (4', 6'-diamidino-2-phenylindole) in order to exclude dead cells from the analysis. Cells were gated based on cell size (FSC/SSC), doublet discrimination (pulse width/SSC) and viability (DAPI/SSC). For each sample approximately 4 x 10^6 cells were acquired and analyzed.

Cell sorting

Cell sorting was conducted on a BD FACSAria flow cytometer (BD Biosciences). Cells were harvested 24 hrs after ssODN transfection and resuspended in PBS supplemented with 2% FCS. Cells were separated by a FSC/SSC amplitude gate; doublets were excluded by FSC/SSC width gates.

To determine the colony survival of corrected cells, *Msh2*$^{-/-}$ *mEGFP* ES cells were exposed to various types of anti-sense ssODN 1m for 24 hrs and separated into 2,000 EGFP-positive and 2,000 EGFP-negative cells. For control purposes, *Msh2*$^{-/-}$ *WT EGFP* ES cells were mixed with EGFP-negative *Msh2*$^{-/-}$ *mEGFP* ES cells and subsequently sorted. The sorted cells were plated onto three gelatin-coated 100 mm dishes. After 8-10 days, surviving colonies were stained with Leishman's eosin methylene blue solution (Merck) and counted. The colony survival of EGFP-positive cells was normalized to the colony survival of EGFP-negative cells that had been exposed to the same ssODNs.

For the Cy5-ssODN uptake experiments, *Msh2*$^{-/-}$ *neo* ES cells were transfected with 5'-Cy5-labelled sense ssODN 4N mixed with unlabelled sense ssODN 4N (ratio 1:4). The day after transfection, cells were binned into four separate groups based on their Cy5 fluorescence signal. Approximately 2.0-2.5 x 10^6 cells were sorted per bin in two replicate runs per experiment. For each bin, cells were plated onto two gelatin-coated 100 mm dishes at 1,000 cells/dish to assess their colony-forming ability. The remainder of the cells was plated for G418 selection onto three gelatin-coated 100 mm dishes to determine the targeting frequency. G418 selection (800 µg/ml; GIBCO-Invitrogen) was started 24 hrs after sorting. After 8-10 days, surviving colonies were stained and counted. Targeting frequencies were calculated by dividing the number of G418-resistant colonies by the number of cells that were plated after sorting.

Neutral comet assay

Msh2$^{-/-}$ *mEGFP* ES cells were exposed to unmodified or PTO-modified anti-sense ssODN 1m for 24 hrs and separated into EGFP-positive and EGFP-negative cells by flow cytometry. Mock-transfected *Msh2*$^{-/-}$ *mEGFP* cells (without ssODN), *Msh2*$^{-/-}$ *WT EGFP* cells, and *Msh2*$^{-/-}$ *mEGFP* cells exposed to 25 Gy of γ-irradiation were included as controls. The neutral comet assay was performed as previously described [23]. Briefly, approximately 8,000 sorted cells from each population were resuspended in 1.2 ml 0.75% low-gelling-temperature agarose (Sigma-Aldrich) at 40°C and pipetted onto glass slides pre-coated with agarose. After gelling for 2 min, slides were immersed into N1 lysis buffer (2% sarkosyl, 0.5 M Na_2EDTA, 0.5 mg/ml proteinase K) and incubated at 37°C in the dark for 22 hrs. After lysis, slides were rinsed three times in N2 solution (90 mM Tris, 90 mM boric acid, 2 mM Na_2EDTA pH8.5) for 30 min at RT. Then, electrophoresis was conducted in fresh N2 solution for 25 min at 0.6 V/cm. Slides were stained in 2.5 µg/ml propidium iodide for 20 min. Images were captured using a Zeiss AxioObserver Z1 inverted fluorescence microscope equipped with a Hamamatsu ORCA-ER B/W CCD camera using identical exposure settings. Images were processed using AxioVision software (Zeiss). The comets tail moment (TM) was determined using the CASP software, version 1.2.2 (www.casp.of.pl). For each sample, the TM of at least 80 randomly chosen cells was measured.

REFERENCES

1. **Brachman EE, Kmiec EB.** (2005) Gene repair in mammalian cells is stimulated by the elongation of S phase and transient stalling of replication forks. *DNA Repair (Amst)*. 4, 445-57.
2. **Olsen PA, Randol M, Krauss S.** (2005) Implications of cell cycle progression on functional sequence correction by short single-stranded DNA oligonucleotides. *Gene Ther*. 12, 546-51.
3. **Wu XS, Xin L, Yin WX, Shang XY, Lu L, Watt RM, Cheah KS, Huang JD, Liu DP, Liang CC.** (2005) Increased efficiency of oligonucleotide-mediated gene repair through slowing replication fork progression. *Proc Natl Acad Sci U S A*. 102, 2508-13.
4. **Engstrom JU, Kmiec EB.** (2008) DNA replication, cell cycle progression and the targeted gene repair reaction. *Cell Cycle*. 7, 1402-14.
5. **Huen MS, Li XT, Lu LY, Watt RM, Liu DP, Huang JD.** (2006) The involvement of replication in single stranded oligonucleotide-mediated gene repair. *Nucleic Acids Res*. 34, 6183-94.
6. **Radecke S, Radecke F, Peter I, Schwarz K.** (2006) Physical incorporation of a single-stranded oligodeoxynucleotide during targeted repair of a human chromosomal locus. *J Gene Med*. 8, 217-28.
7. **Dekker M, Brouwers C, te Riele H.** (2003) Targeted gene modification in mismatch-repair-deficient embryonic stem cells by single-stranded DNA oligonucleotides. *Nucleic Acids Res*. 31, e27.
8. **Aarts M, Dekker M, de Vries S, van der Wal A, te Riele H.** (2006) Generation of a mouse mutant by oligonucleotide-mediated gene modification in ES cells. *Nucleic Acids Res*. 34, e147.
9. **Igoucheva O, Alexeev V, Anni H, Rubin E.** (2008) Oligonucleotide-mediated gene targeting in human hepatocytes: implications of mismatch repair. *Oligonucleotides*. 18, 111-22.
10. **Papaioannou I, Disterer P, Owen JS.** (2009) Use of internally nuclease-protected single-strand DNA oligonucleotides and silencing of the mismatch repair protein, MSH2, enhances the replication of corrected cells following gene editing. *J Gene Med*. 11, 267-74.
11. **Olsen PA, Solhaug A, Booth JA, Gelazauskaite M, Krauss S.** (2009) Cellular responses to targeted genomic sequence modification using single-stranded oligonucleotides and zinc-finger nucleases. *DNA Repair (Amst)*. 8, 298-308.
12. **Bonner M, Kmiec EB.** (2009) DNA breakage associated with targeted gene alteration directed by DNA oligonucleotides. *Mutat Res*.
13. **Olsen PA, Randol M, Luna L, Brown T, Krauss S.** (2005) Genomic sequence correction by single-stranded DNA oligonucleotides: role of DNA synthesis and chemical modifications of the oligonucleotide ends. *J Gene Med*. 7, 1534-44.
14. **Murphy BR, Moayedpardazi HS, Gewirtz AM, Diamond SL, Pierce EA.** (2007) Delivery and mechanistic considerations for the production of knock-in mice by single-stranded oligonucleotide gene targeting. *Gene Ther*. 14, 304-15.
15. **Aarts M, te Riele H.** (2009) Parameters of oligonucleotide-mediated gene modification in mouse ES cells. *J Cell Mol Med*. In press.
16. **Ellis HM, Yu D, DiTizio T, Court DL.** (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc Natl Acad Sci U S A*. 98, 6742-6.
17. **Li XT, Costantino N, Lu LY, Liu DP, Watt RM, Cheah KS, Court DL, Huang JD.** (2003) Identification of factors influencing strand bias in oligonucleotide-mediated recombination in Escherichia coli. *Nucleic Acids Res*. 31, 6674-87.
18. **Andrieu-Soler C, Casas M, Faussat AM, Gandolphe C, Doat M, Tempe D, Giovannangeli C, Behar-Cohen F, Concordet JP.** (2005) Stable transmission of targeted gene modification using single-stranded oligonucleotides with flanking LNAs. *Nucleic Acids Res*. 33, 3733-42.
19. **Ferrara L, Kmiec EB.** (2006) Targeted gene repair activates Chk1 and Chk2 and stalls replication in corrected cells. *DNA Repair (Amst)*. 5, 422-31.
20. **Dekker M, Brouwers C, Aarts M, van der Torre J, de Vries S, van de Vrugt H, te Riele H.** (2006) Effective oligonucleotide-mediated gene disruption in ES cells lacking the mismatch repair protein MSH3. *Gene Ther*. 13, 686-94.
21. **Engstrom JU, Suzuki T, Kmiec EB.** (2009) Regulation of targeted gene repair by intrinsic cellular processes. *Bioessays*. 31, 159-68.
22. **Rosenkranz HS, Levy JA.** (1965) Hydroxyurea: A Specific Inhibitor of Deoxyribonucleic Acid Synthesis. *Biochim Biophys Acta*. 95, 181-3.
23. **Olive PL, Banath JP.** (2006) The comet assay: a method to measure DNA damage in individual cells. *Nat Protoc*. 1, 23-9.
24. **Li X, Zhao X, Fang Y, Jiang X, Duong T, Fan C, Huang CC, Kain SR.** (1998) Generation of destabilized green fluorescent protein as a transcription reporter. *J Biol Chem*. 273, 34970-5.