Enhanced gene repair mediated by methyl-CpG-modified single-stranded oligonucleotides

Carmen Bertoni1,2,*, Arjun Rustagi1 and Thomas A. Rando1,3

1Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, CA 94305, 2Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, CA 90095 and 3Neurology Service, VA Palo Alto Health Care Systems, Palo Alto, CA 94304, USA

Received January 27, 2009; Revised August 21, 2009; Accepted August 28, 2009

ABSTRACT
Gene editing mediated by oligonucleotides has been shown to induce stable single base alterations in genomic DNA in both prokaryotic and eukaryotic organisms. However, the low frequencies of gene repair have limited its applicability for both basic manipulation of genomic sequences and for the development of therapeutic approaches for genetic disorders. Here, we show that single-stranded oligodeoxynucleotides (ssODNs) containing a methyl-CpG modification and capable of binding to the methyl-CpG binding domain protein 4 (MBD4) are able to induce >10-fold higher levels of gene correction than ssODNs lacking the specific modification. Correction was stably inherited through cell division and was confirmed at the protein, transcript and genomic levels. Downregulation of MBD4 expression using RNAi prevented the enhancement of gene correction efficacy obtained using the methyl-CpG-modified ssODN, demonstrating the specificity of the repair mechanism being recruited. Our data demonstrate that efficient manipulation of genomic targets can be achieved and controlled by the type of ssODN used and by modulation of the repair mechanism involved in the correction process. This new generation of ssODNs represents an important technological advance that is likely to have an impact on multiple applications, especially for gene therapy where permanent correction of the genetic defect has clear advantages over viral and other nonviral approaches currently being tested.

INTRODUCTION
Manipulation of genomic sequences through gene trapping or gene targeting technologies has a wide spectrum of applications ranging from site-directed mutagenesis of bacterial vectors to development of animal models through DNA recombination technologies. Among those, gene editing can be used to target specific gene defects and restore protein expression for therapeutic applications (1–5). To date, however, the use of gene editing has been limited and is still at an early stage of development.

Single-stranded oligodeoxynucleotides (ssODNs) have been shown to be able to alter single nucleotides and induce stable alterations at the genomic level (3,6). Gene repair mediated by ssODNs takes advantages of specific repair mechanisms present in the cells that are able to recognize the presence of mismatches in genomic DNA. ssODNs complementary to the target sequence but containing a mismatch at the base targeted for modification are used as templates for the correction process. Once introduced into the cell, they have been shown to anneal to the genomic DNA sequence targeted for repair and initiate the repair process (7) leading to a single nucleotide exchange that is stably inherited throughout cell division. The technology has been successfully applied in different cell types including bacterial, yeast and mammalian cells (1,3,8–12).

Oligonucleotide-mediated gene correction has been investigated in several eukaryotic cell type as well as different models of genetic disorders. Correction has been successfully demonstrated in hepatocytes (13–16), retinal cells (17,18), bone marrow-derived cells (19) and muscle cells (5,9,10,20–22). The level of correction varies depending on the cell type being targeted suggesting that the repair process involved is differentially regulated in different cell types.

In skeletal muscle, the major focus of ssODN-mediated gene editing has been the treatment of genetic disease, in particular Duchenne muscular dystrophy (DMD). This disease is characterized by mutations in the dystrophin gene that lead to complete absence of dystrophin protein expression, progressive muscle degeneration and weakness, and in most of the cases death by the age of 30 years. Targeted single base alterations of the dystrophin gene...
has been successfully achieved both in vitro and in vivo, using mouse and dog models of DMD (5,9,10,20,21). Correction can be obtained in mature myofibers (5,10) as well as muscle progenitor cells (10), suggesting that oligonucleotides can target both postmitotic cells as well as replicating cells and can achieve frequencies of up to 10% depending on the strand targeted for repair and the type of mismatch created by the oligonucleotide (5,10,20,21).

The past few years have seen a considerable expansion of the application of gene editing to neuromuscular disorders. Oligonucleotides can be used not only to correct dystrophin gene defects due to single point mutations, but also to treat frameshift deletions by disrupting intron/exon consensus sequences to redirect mRNA assembly and splicing and induce the expression of novel in-frame transcripts (5,21). Along the same line, ssODNs have been shown to target and correct a splicing regulatory element in the SMN2 gene and to restore SMN mRNA assembly and splicing and induce the expression of novel in-frame transcripts (5,21).

The mechanisms that lead to the repair process remain unclear, but appear to involve multiple steps. The first step requires the annealing of the ssODN to the targeted genomic sequence through the interaction of proteins such as Rad51, Rad54 and XRCC2 (23–25). The second step, the correction of the single base pair mismatch, is less well characterized. Initial data suggested that correction is mediated by mismatch repair (MMR) mechanisms that recognize the mismatch created by the annealing of the ssODN to the genomic DNA (26). Proteins involved in MMR appear to be crucial for the ssODN-mediated base exchange at least in yeast (7). In some mammalian cells, however, gene correction can occur even in the absence of MMR (27,28), suggesting that different mechanisms may operate to facilitate ssODN-mediated gene repair under different cellular contexts. Homologous recombination, nucleotide exchange repair and double-strand break repair have all been implicated as potential mechanistic pathways (29–31), and it is not clear if the different frequencies of gene correction detected in different cell types are in part due to different mechanisms being recruited. This hypothesis led us to investigate whether ssODN-mediated gene correction frequencies could be improved by designing ssODNs capable of activating specific repair mechanisms in the cells targeted for repair.

Deamination of 5-methylcytosine is among the most prominent genomic alterations in mammalian cells with rates up to 300 events per genome per day (32,33). Loss of the amino group leads the spontaneous conversion of 5-methylcytosine (m5C) into thymine which, if left unrepaired, would lead to the accumulations of C:G-to-T:A mutations (34,35). The presence of 5-methylcytosines and the frequency of mutation at such bases in the context of m5CpG sites contribute to genetic variation and are implicated in tumorigenesis (36,37). Repair of the mismatch created by the deamination of 5-methylcytosine requires the presence of specific repair mechanisms capable of recognizing the mismatch that arises as a result of the deamination. Recognition of deamination events in the context of m5CpG sites is mediated by the base excision repair (BER) mechanism through the specific activation of the methyl-CpG binding domain protein 4 (MBD4). MBD4 (also known as MED1) plays a crucial role in maintaining genome integrity by recognizing G:T or G:U mismatches at m5CpG sites on double-stranded DNA (38,39). MBD4 was originally identified in a yeast two-hybrid screen due to its interaction with MLH1, an important component of the MMR mechanism, which led to the hypothesis that the two systems could be linked (38,40). To date, however, the role of the interaction between MBD4 and MLH1 has yet to be clarified (41). In vitro studies have shown that MBD4 can efficiently recognize and hydrolyze G:T and G:U mismatches at hemi-methylated m5CpG sites. Furthermore, a m5CpG context, although preferred, is not absolutely necessary as G:T and G:U mismatches in nonmethylated CpG sequences can also be recognized and processed, although at a reduced rate (39). Binding of MBD4 to the mismatched T or U leads to glycosylation and removal of the base without altering the sugar phosphate backbone of the DNA. The apurinic site that is generated by MBD4 is then processed by specific endonucleases and ligases which are required to direct the addition of a new cytosine on this site and to complete the repair process (42,43).

We have designed a new generation of ssODNs containing a methyl-CpG modification and tested their ability to mimic, when annealed to the genomic sequence targeted for correction, the G:T mismatch that would occur upon deamination of 5-methylcytosines (to become thymine) in the genomic DNA targeted for correction. Due to the high frequencies at which this process occurs in nature, we hypothesized that the use of these ssODNs could be much more efficient in directing stable single base alterations at the genomic level. Their ability to specifically activate the BER repair by recruiting MBD4 and to induce single base alterations was tested in muscle cells using a GFP reporter system. This system allowed us to determine the efficacy of gene correction of an m5CpG-containing ssODN and to compare its efficacy to targeting ssODNs either lacking any modified cytosines or containing a 5-methylcytosine but not in the context of a CpG dinucleotide, neither of which would be expected to recruit MBD4. We demonstrate that gene correction frequencies were consistently higher when an ssODN containing a methyl-CpG and capable of recruiting MBD4 was used. These studies expand the potential applications of ssODNs in inducing single base alterations at the genomic level and further advance this technology into therapeutic applications for the treatment of many disorders and in particular muscle diseases.

MATERIALS AND METHODS
Oligonucleotide synthesis
ssODNs were purchased from MWG Biotech Inc. (High Point, NC, USA). All ssODNs were labeled at the 5′-end
with fluorescent CY3 and at the 3’-end with a tag (-cgccg) of phosphorothioate bases to increase stability of ssODNs from endonucleases. Oligonucleotides were HPLC purified and exhibited a single peak of the expected molecular weight as determined by MALDI-TOF mass spectroscopy analysis.

Plasmid constructs

The pGFPmut vector was generated using the QuickChange® II Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA, USA), following the manufacturer’s instructions. The pEGFP-N3 vector (BD Bioscience, Clontech, CA, USA) was annealed with a forward primer (GFPmut forward: 5’-GACCTACGGC GTGCGATGATTACGCCTACCCCCCGA-3’) and a reverse primer (GFPmut rev: 5’-GTCGGGTAGCCGCTGAACTCATC-3’) designed to induce the C-to-A transversion in position +213 of the GFP coding sequence. Initial denaturation was carried out at 95°C for 3 min, followed by a 20-cycle step of denaturation at 95°C for 3 s, annealing at 58°C for 1 min and extension at 68°C for 6 min. The PCR product (1 µl) was digested overnight at 37°C with DpnI and transformed in DH5α competent cells (Invitrogen, Carlsbad, CA, USA). Positive clones were selected by restriction endonuclease analysis and the C-to-A mutation was confirmed by direct sequencing using an Applied Biosystems ABI377 automated sequencer.

The pGEX3.MBD4 plasmid was generated by inserting the MBD4 coding sequence into the BamHI and XhoI sites of the pGEX3 vector (Amersham GE Healthcare Bio-Sciences Corp., Piscataway, NJ, USA). The mouse cDNA encoding MBD4 was obtained by direct cloning of PCR products encoding the MBD4 cDNA (GenBank accession number NM_010774). The MBD4 gene encoding full-length cDNA was obtained by direct cloning of PCR products encoding the MBD4 cDNA (GenBank accession number NM_010774). The MBD4 gene encoding full-length cDNA was obtained by direct amplification of a cDNA library using a forward primer (MBD4-forw-BamHI: 5’-GCAGCGAGGATCCCGAGA GCCC-3’) and a reverse primer (MBD4-rev-XhoI: 5’-CT CGGATAGACTTAATTTTTCATG-3’). Amplification was carried out for 30 cycles in the presence of Pfu DNA polymerase (Stratagene) at an annealing temperature of 65°C for 1 min and an extension temperature of 68°C for 2 min. Amplification was terminated by an additional extension of 30 min at 72°C. Amplicons were digested with BamHI and XhoI. PCR products were run on a 1.2% agarose gel and purified using the gel extraction kit (Qiagen, Valencia, CA, USA) as previously described (5). Vector sequences were confirmed using an ABI377 automated sequencer.

Cell culture and transfection

Myoblasts were plated in wells of 6-well dishes (10^4 cells/well) 12 h prior to transfection. ssODNs (10 µg) were complexed with 1 µl of Lipofectamine 2000 (Invitrogen Corp.) for 30 min at room temperature in a total volume of 0.5 ml of Ham’s F10 nutrient mixture. The complex was then added to the wells containing 1.5 ml of GM. Transfections were stopped by replacing the transfection solution with fresh GM.

Expression of recombinant MBD4 protein

Expression constructs of wild-type MBD4 protein were propagated in Escherichia coli strain XL-1 Blue, as previously described (38,45). BL21(DE3)(pLysS) cells transformed with the expression vectors were grown to OD_{600} 0.4 and induced with 1 mM IPTG at 37°C for 2 h. Cells were collected by centrifugation and lysed in 10 mM Tris–HCl (pH 8.0), 500 mM NaCl, 0.1% Nonidet P-40 (NP-40), 10% glycerol and ‘Complete’ protease inhibitors (Boehringer-Mannheim Biochemicals, Indianapolis, IN, USA). After clarification by centrifugation at 12 000g, the soluble protein fraction was diluted in 150 mM NaCl and β-mercaptoethanol (7 mM) was added. Purification of recombinant proteins was conducted as previously earlier (46). Briefly, the sample was applied to a 5-ml Q Sepharose anion exchange column (Amersham) connected in series to a 5-ml SP Sepharose cation exchange column (Amersham). After disconnecting the Q Sepharose column, the SP Sepharose column was washed and elution was performed with buffer containing 1.5 M NaCl. The SP Sepharose eluate was applied to a 1.5-ml nickel-chelating Sepharose column (Amersham). Recombinant MBD4 protein was eluted with 5 ml of 10 mM Tris–HCl (pH 8.0), 300 mM NaCl, 10% glycerol, 2 mM β-mercaptoethanol and 150 mM imidazole. MBD4 protein was further purified by size-exclusion chromatography using a Superdex 200 PC 3.2/30 gel-filtration column with a SMART chromatography system (Pharmacia Laboratories, Piscataway, NJ, USA), as previously described (45,46). The final MBD4 preparation was estimated to be 95% pure by SDS-PAGE. Upon the addition of 10% glycerol, purified MBD4 fractions were frozen and stored in liquid nitrogen. Protein concentration was determined with the Bio-Rad protein assay (Bio-Rad, Hercules, CA, USA), using bovine gamma globulin as a standard.

Flow cytometry

Fluorescence was measured using a Beckton Dickinson FACScalibur flow cytometry (Becton Dickinson, Franklin Lakes, NJ, USA). Cells were trypsinized and harvested at different time after transfection, resuspended in 0.5% BSA, 2 mM EDTA in PBS and processed immediately. For each analysis, a total of 2 x 10^4 cells were used. Experiments were repeated in triplicate for each time point analyzed and a total of three independent experiments were performed.

For DNA content analysis, cells were fixed in 70% ice-cold ethanol, washed in PBS and resuspended in 0.5 ml of PBS containing 50 µg/ml RNAse A, 1% FBS and 2.5 µg/
ml propidium iodide. Analyses were performed using 2 × 10^6 cells.

Western blot analysis

Cells were lysed in RIPA buffer [50 mM Tris–HCl (pH 7.4), 150 mM NaCl, 0.5% deoxycholate and 1% Nonidet P-40] containing aprotinin (20 μg/ml), leupeptin (20 μg/ml), phenylmethylsulfonyl fluoride (10 μg/ml) and sodium orthovanadate (1 mM). Total protein in the extract was determined by the Bio-Rad protein assay. Immunoblot analysis was performed as previously described (10,47). Total protein from each sample (200 μg) was separated by electrophoresis (20 mA for 1.5 h) using 12% SDS-polyacrylamide gels, and then transferred (250 mA for 1 h) onto nitrocellulose membranes. Membranes were probed overnight at 4°C with an antibody directed toward the C-terminal region of the GFP protein (1GFP63; Santa Cruz Biotechnologies, Santa Cruz, CA, USA) or an antibody directed toward the N-terminal domain of the MBD4 protein (Abcam Inc., Cambridge, MA, USA) (21,47). The membranes were blocked with 5% milk in PBS for 1 h at room temperature. Blots were washed with 0.05% Tween-20 in PBS and then incubated with a horseradish peroxidase-coupled anti-mouse secondary antibody (Amersham). Specific antibody binding was detected using an enhanced chemiluminescent system (Amersham).

Genomic DNA and restriction endonuclease analysis

Cultured myoblasts were rinsed twice with PBS and genomic DNA was extracted using the Wizard Genomic DNA extraction kit (Promega, Madison, WI, USA) according to the manufacturer’s instructions. Total genomic DNA (1 μg) was digested overnight at 37°C with 5 U of HinfI or BstI, purified using Amicon Microcon® PCR Centrifugal Filter Devices (Millipore Corporation, Bedford, MA, USA), and resuspended in 20 μl of H2O. For each amplification reaction, 5 μl of digested genomic DNA was subjected to amplification using the forward primer (GFP-forw: 5'-GGTACGGCTACCGGTCCGA C-3') specific to region +17 to −2 of the GFP open reading frame (ORF), and the reverse primer (GFP-rev: 5'-CCTTGAAGTCCATGCCCTTC-3') complementary to the region +397 to +377 of the ORF.

Real-time PCR was performed using a My iQ single-color detection system (Bio-Rad) at the following thermal cycler conditions: 95°C for 10 min followed by 45 cycles of a 3-step reaction consisting of denaturation at 95°C for 30 s, annealing at 60°C for 1 min and extension and data collection at 72°C for 30 s. Amplicons were separated on 1.5% agarose gels and PCR products were purified using the Qiagen gel extraction kit (Qiagen). DNA sequencing was carried out using an Applied Biosystems ABI377 automated sequencer.

RT–PCR

Total RNA was extracted from cultured myotubes using TRI-REAGENT (Sigma, St. Louis, MO, USA). For each reaction, 2 μg of RNA was treated with 1 U of DNase I (GIBCO/BRL) at room temperature and reverse transcribed using the first strand cDNA synthesis kit (Invitrogen) in the presence of SuperscriptIII (Invitrogen) to a final volume of 20 μl. Of those, 10 μl were digested overnight with 5 U HinfI or BstI in a final volume of 20 μl. The remaining 10 μl of undigested cDNA obtained after reverse transcription were brought to a final volume of 20 μl containing isomolar concentrations of salt and glycerol and were used to normalize transcript levels. PCR reactions were carried out using 4 μl of digested product in the presence of the GFP primers described earlier. Each amplification mixture contained 25 pmols of appropriate primer, 10% DMSO, 0.5 U Master Taq DNA polymerase (Takara, Panvera Corp., Madison, WI, USA), and 5 mM of each deoxyribonucleotide triphosphate. After an initial step of denaturation at 95°C for 5 min, amplification was performed for 35 cycles at 95°C for 1 min followed by annealing at 55°C for 3 min and extension at 72°C for 2 min. Amplification reactions were terminated by an additional extension step at 72°C for 10 min and products were fractionated on 1.5% agarose gels.

Electrophoretic mobility shift assay

An oligonucleotide homologous to the region of the GFP gene targeted for correction and ssODNs were synthesized by MWG and purified by HPLC. Targeting and control oligonucleotides were identical to the one previously described (Figure 1), but lacked the CY3 modification at their 5'-end. The probe was end-labeled with [γ-32P]ATP using T4 polynucleotide kinase (Amersham). The ability of MBD4 to form a complex with the m5CpG ssODN alone was investigated using the m5CpGGFP labeled at its 5'-end with [γ-32P]ATP. Unincorporated label was removed with G25 spin columns (Amersham). Control and target- ing ssODNs were annealed to equal amounts (~100,000 c.p.m.) of radiolabeled probe using a thermal cycler at the following setting conditions: 95°C for 5 min for the initial denaturation step, followed by 80°C for 5 min, 75°C for 5 min, 70°C for 15 min, 55°C for 15 min, 37°C for 15 min and 20°C for 20 min. One microgram of recombinant MBD4 protein was incubated in binding buffer [50 mM Tris–HCl (pH 7.5), 2.5 mM dithiothreitol, 2.5 mM EDTA, 250 mM NaCl, 5 mM MgCl2, and 20% glycerol] in the presence of 1 μg of Poly(dI-dC):Poly(dI-dC) competitor (Amersham) for 15 min at room temperature. Double-stranded oligonucleotides were then added to the mixture and incubated for an additional 15 min. Samples were loaded on 5% polyacrylamide gels and resolved at 200 V for 2 h. Gels were dried for 1 h at 80°C prior to autoradiographic exposure.

siRNA treatment

The target sites were predicted in silico using on-line tools from Invitrogen Corp (http://ribiserv.tch.ub.edu/mfold/). In brief, the sequences targeted by the Stealth™ siRNA were listed as residues 589–614 of the ORF with the following sequence: 5'-GCAGCCAAA UGAAACUGACGUUCA-3' and 3'-CGUCGGUUUA CUUUGACUGCAAGU-5'. The negative control
Stealth™ siRNA used were: 5′-GCAACAAGUAACAG UCAUGUCGUCA-3′ and 3′-CGUUGUUCAUUGUCA GUACAGCAGU-5′. No homologous sequences were found in mouse genome. The double-stranded siRNAs were synthesized chemically and modified into stealth siRNA (Invitrogen Corp) to enhance the stability in vitro. Transfection of the siRNAs was performed using Lipofectamine 2000 (Invitrogen) following the manufacturer’s instructions. Efficient silencing was demonstrated by real-time PCR in cells transfected with the siRNAs and was performed on RNA isolated at different time after siRNA transfection (Figure 9). First strand cDNA was prepared and reverse transcribed into cDNA using SuperScript III Reverse Transcriptase (Invitrogen) according to the manufacturer’s protocol. Quantitative real-time RT–PCR was performed using a My iQ single-color detection system (Bio-Rad) using the following thermal cycler conditions: 95°C for 10 min followed by 45 cycles of a 3-step reaction consisting of denaturation at 95°C for 30 s, annealing at 60°C for 1 min and extension and data collection at 72°C for 30 s. The primers RTMBD4-forw (5′-GATGTATACCTTTATCGGCCAC AAG-3′) and RTMBD4-rev (5′-CTGGCCGACTCTGAA GTACCAC-3′) were used to detect MB4 mRNA. Transcripts levels were normalized to those of GAPDH using the following primers: GAPDH-forw: 5′-TACCGAC TTCACAGCAACTC-3′; GAPDH-rev: 5′-ATGTAGG CCATGAGGTCCAC-3′ as described earlier(48).

Statistical analysis
Data are presented as means and standard deviations. Comparisons between groups were done using Student’s t-test assuming two-tailed distribution and unequal variances.

RESULTS
The GFP reporter system and assay
We have established a reporter system and used it to follow gene repair over time in muscle cells. The pGFPmut vector contains a single point mutation (a C-to-A transversion at position +213 of the coding region), which creates a TGA stop codon and abolishes GFP protein expression in mammalian cells (Figure 2). The plasmid was transfected into murine myoblasts and a stable cell line was selected by G418 using the neomycin resistance gene encoded by the plasmid. Expression of GFP mRNA transcripts was confirmed using RT–PCR. Genomic DNA analysis performed using real-time PCR revealed the presence of 1.4 copies of integrated vector per cell. No GFP protein expression was detected either by fluorescence microscopy in those cultures. No GFP protein expression was detected either by immunofluorescence microscopy by the presence of the CY3 tag, was followed for up to 1 week after transfection (Figure 3). The fluorescent tag was detected in nearly 100% of cells in cultures treated with ssODNs 24 h after transfection (Figure 3A). FACS analysis demonstrated similar pattern of distribution of fluorescence in cells treated with control or targeting oligonucleotides suggesting equal transfection efficiencies among methylated and unmethylated ssODNs (Figure 3B). Fluorescence remained stable for up to 48 h after transfection and then steadily declined (Figure 3C).
Figure 1. Wild-type and mutant GFP sequences and ssODN design. (A) The pGFPmut2 vector contains a single point mutation at base pair +213 of the GFP coding sequence that creates a stop codon in this position. ssODNs were designed to target bases in the stop codon to restore GFP expression. (B) Sequences of targeting (CORGFP, m5CpGGFP, m5CpCGFP) and control (CTLGFP) ssODNs. (C) Annealing of the targeting ssODNs to the genomic sequence creates a mismatch (box), whereas the control ssODN pair perfectly. The CORGFP ssODN is designed to induce an A-to-C conversion that eliminates the stop codon and restores wild-type GFP expression. The m5CpGGFP ssODN creates a G:T mismatch in the context of a m5CpG, thus mimicking the delamination of a m5C. This will induce a T-to-C substitution which will convert the stop codon into an arginine. The m5CpCGFP ssODN targets the same base as CORGFP but contains an m5C modification which will lead to an A-to-G conversion and result in a single amino acid substitution (a tryptophan in place of a cysteine). The CTLGFP ssODN perfectly complements the mutant sequence and therefore should not induce any genomic alterations.

Figure 2. Binding of GFP ssODNs to MBD4. A 45-nt oligonucleotide radiolabeled probe corresponding to the region of the pGFPmut2 vector targeted for repair was incubated with or without ssODNs and with and without recombinant MBD4 at 37°C for 30 min and analyzed by EMSA. The position of the ssODN/probe complexes is indicated by the thin arrow. The thick arrow above shows the position of the supershift achieved by the binding of MBD4 with the ssODN/probe complex. MBD4 was unable to form a complex with the target sequence in the absence of an ssODN bearing a mismatch. No binding of MBD4 to m5CpG [as assessed using a radioactively labeled m5CpG (m5CpG*)] was detected in the absence of the oligonucleotide sequence targeted for repair.
By 1 week, fluorescence had almost disappeared providing clues as to the stability of the oligonucleotides in the cells.

**Effects of ssODNs on cell viability**

We used a cell viability assay to assess the toxicity of ssODNs on myoblast cultures for up to 4 days after transfection (Figure 4). A significant increase in non-viable cells was detected in all cultures treated with ssODNs 24 h after transfection. Oligonucleotides containing a 5-methyl cytosine appeared to be slightly more toxic than COR\(^{\text{GFP}}\), but those differences were not statistically significant. Those differences, however, were only transient and disappeared by 48 h after transfection. By 72 h, the number of apoptotic cells had returned to levels similar to those observed in sham transfected cells and remained stable through the duration of the experiment (Figure 4).

The rate of population growth was determined in cells treated with the control and targeting ssODNs and compared with those of cells transfected with vehicle only (Figure 5). Cell number doubled approximately...
every 18–20 h independently of the type of ssODN used. No significant differences in population growth were detected in cultures treated with targeting compared with control ssODNs.

We next analyzed the effects of ssODNs on cell cycle progression to determine whether ssODNs could affect the cell cycle. We did not observe any significant changes in the number of cells present at each phase of the cell cycle, even when the uptake of ssODNs was at maximal level within the first 48 h after transfection (Table 1). Furthermore, there were no changes related to the presence of methylated ssODNs compared with unmethylated ssODNs.

**Restoration of GFP expression after targeting ssODN treatment**

Flow cytometry was used to follow restoration of GFP expression over time in cultures transfected with ssODNs (Figure 6). No GFP-positive cells were detected at any time in untransfected cultures or in cultures transfected with the control ssODN. GFP-positive cells were detected as early as 1 day after transfection of targeting ssODNs. Two days after transfection, cultures transfected with m5CpGGFP showed up to 6-fold more GFP-positive cells than did cultures transfected with either m5CpCGFP or COR GFP (Figure 6A). GFP-positive cells were clearly detected in cultures treated with targeting oligonucleotides by both FACS and microscopic analyses (Figure 6B and C).

To confirm the stability of the repair process over time, myoblasts treated with ssODNs were analyzed by western blot 1 week after ssODN transfection. Expression of GFP protein was detected in cultures treated with targeting ssODNs but not with the control ssODNs (Figure 7). Myoblasts transfected with the m5CpGGFP ssODN showed 2.5-fold higher levels of GFP protein expression than cells treated with the other targeting ssODNs.

**Assessment of gene correction efficiency**

To further confirm alteration of the GFP sequence targeted by the ssODNs and to have a precise estimate of frequencies of gene correction, we analyzed the genomic sequences by real-time PCR using DNA samples isolated 3 days after ssODN transfection. The GFP cDNA mutation in the pGFPmut vector creates a new restriction site (HinfI, Figure 8A). Using cells transfected with either COR GFP or m5CpCGFP, digestion of total genomic DNA with HinfI allowed us to cleave all copies of the mutant GFP sequence, leaving intact for amplification only that portion of DNA that had undergone single base pair alteration rendering it refractory to HinfI restriction digestion. Amplification of a specific DNA product was observed only in cells treated with the targeting ssODNs but not in untreated cells or cells transfected with the control ssODN (Figure 8B). For analysis of cells treated with m5CpGGFP, we used a similar strategy but a different restriction enzyme because of the different base targeted for correction (Figure 8A). In that case, digestion of total genomic DNA with BstI should eliminate all copies not altered by ssODN-mediated gene modification. Again, amplification of a specific DNA product was observed only in cells treated with the targeting ssODN, but not in untreated cells or cells transfected with the control ssODN (Figure 8B). The amount of DNA amplified was higher in cells treated with m5CpGGFP, consistent with the data obtain at the protein level. Direct sequencing of the products obtained after restriction enzyme digestion and PCR amplification demonstrated the specificity of the PCR reaction and confirmed the presence of the predicted single base pair substitutions at the genomic level (Figure 8C).

The differences in fold changes of PCR products amplified after HinfI or BstI digestion were determined using a standard delta-Ct method (48). The levels of gene correction in cells treated with m5CpGGFP were >10-fold higher than those detected using the COR GFP.
ssODN (Figure 8D) and corresponded to an average of 7% of the total number of copies of the pGFPmut vector present in culture. No significant changes in gene correction levels were detected when comparing the gene repair frequencies of \textsuperscript{5}CpG\textsuperscript{GFP} with those achieved using the COR\textsuperscript{GFP} ssODN (Figure 8D).

Endonuclease digestion analysis was also used to test for GFP mRNA expression as the result of the single base substitution at the genomic level. Total mRNA was isolated from cells 1 week after transfection and reverse transcribed prior to \textit{HinfI} or \textit{BtsI} digestion (see ‘Materials and Methods’ section). The control ssODN failed to produce any amplification of cDNA after restriction enzyme digestion (Figure 8E). The amount of GFP mRNA that was refractory to endonuclease digestion as a result of ssODN-mediated gene editing was consistently higher in cells treated with the MBD4-binding ssODN as compared with the other targeting ssODNs. Direct sequencing of the PCR products was used to confirm expression of GFP transcripts containing the desired single base alteration and to ensure specificity of the amplification product obtained after endonuclease digestion (data not shown).

**Figure 6.** Time course and stability of gene repair. (A) FACS analysis was used to follow GFP protein expression over time. GFP was detected as early as 1 day after ssODN transfection and did not increase significantly after 2 days. Cells transfected with \textsuperscript{5}CpG\textsuperscript{GFP} showed a significantly greater number of GFP-positive cells as compared with cultures treated with targeting ssODNs, but lacking the methyl-CpG modification. \( ^* P \leq 0.008 \). (B) FACS dot-plots of myoblasts treated with ssODNs 2 and 7 days after transfection of oligonucleotides. (C) Analysis of cultures by fluorescence microscopy after treatment with the \textsuperscript{5}CpG\textsuperscript{GFP} ssODN and maintained in culture for 4 days. Bright fluorescence intensity was detected in a subset of replicating cells.

**Figure 7.** GFP protein expression in muscle precursor cells in culture. Western blot analysis of C57 myoblasts stably expressing the GFP-mutant vector 1 week after ssODN transfection. Full-length GFP protein expression was higher in cells treated with \textsuperscript{5}CpG\textsuperscript{GFP} compared with the other targeting ssODNs. No protein was detected in cells after transfection with the control ssODN.
Figure 8. Evidence of ssODN-mediated single base modification at the molecular level. (A) The single base mutation on the GFP reporter vector (base underlined) generates a new restriction site for HinfI endonuclease. Promotion of an A-to-C transversion mediated by the COR GFP ssODN and an A-to-G transition mediated by the m5CpC GFP ssODN eliminate the restriction site and render the sequence refractory to HinfI digestion and cleavage. The T-to-C conversion mediated by m5CpG GFP eliminates the BtsI restriction site. (B) Genomic DNA was isolated from myoblasts stably expressing the mutant GFP 1 week after ssODN transfection and treated with either HinfI (for COR GFP, m5CpC GFP and CTL GFP) or BtsI (for m5CpG GFP and CTL GFP). Amplification was carried using a forward and a reverse primer encompassing the GFP mutation. No amplification was detected in samples treated with CTL GFP using either HinfI or BtsI. A specific PCR product of identical molecular weight to that obtained after amplification of a wild-type GFP plasmid was detected only in cells treated with targeting ssODNs. (C) Direct sequencing of PCR products obtained as in (B) by amplifying total genomic DNA having undergone HinfI or BtsI digestion demonstrated single base alterations at the genomic level and the specificity of the amplification products. (D) Real-time PCR was used to determine the relative amount of genomic DNA having undergone ssODN-mediated single base alteration. Cultures treated with m5CpG GFP showed 10-fold higher levels of gene editing compared with cultures treated with m5CpC GFP. No significant differences in efficiency were detected between cultures treated with the COR GFP and the m5CpC GFP ssODNs. The total amount of plasmid was determined by amplification of the undigested samples using the same primer combinations and was used as internal standard control to normalize DNA levels. *P = 0.001. (E) Total mRNA was isolated 1 week after ssODN transfection and reverse transcribed using an oligo dT primer. The cDNA was then digested overnight with HinfI or BtsI depending of the ssODN transfected (A) and amplified using the same primers used for the genomic DNA analysis. Modification of the mutant GFP sequence using the m5CpG GFP ssODN was more efficient than that observed in cultures treated with targeting ssODNs lacking the 5-methylcytosine (COR GFP) or the ssODN containing the mC modification but not in a CpG context (m5CpC GFP).
Effects of knockdown of MBD4 expression on ssODN-mediated base alteration

The involvement of MBD4 in the gene correction process mediated by methyl-CpG-modified ssODNs was investigated in cells by downregulating MBD4 expression. siRNA was used to knockdown expression of MBD4, leading to a significant reduction of MBD4 transcript levels as soon as 24 h after siRNA transfection and the effect was maintained up to 48 h after transfection. *P ≤ 0.001. (B) Western blot analysis was carried out using total protein isolated 2 days after siRNA transfection. MBD4 protein was markedly reduced in myoblast cultures treated with MBD4 siRNA compared with untreated cells or cells treated with the control siRNA.

![Image](https://example.com/image1.png)

**Figure 9.** Reduction of MBD4 expression by siRNA interference. (A) RNA interference was used to downregulate expression of MBD4 in myoblast cultures. Cells were transfected with a targeting or control siRNA and gene expression was followed over time by quantitative RT–PCR. Efficient downregulation of MBD4 expression was achieved by 24 h after siRNA transfection and the effect was maintained up to 48 h after transfection. *P ≤ 0.001. (B) Western blot analysis was carried out using total protein isolated 2 days after siRNA transfection. MBD4 protein was markedly reduced in myoblast cultures treated with MBD4 siRNA compared with untreated cells or cells treated with the control siRNA.

DISCUSSION

We have studied the ability of a modified-ssODN (mCpG<sub>GFP</sub>), capable of binding MBD4 to direct a site-specific modification at the genomic level and to mediate enhanced gene repair compared with an older generation of correcting ssODN (COR<sub>GFP</sub>) unable to recruit the specific repair mechanism. Binding was demonstrated in vitro using recombinant MBD4 and was specific only to the ssODN containing the methyl-CpG modification (Figure 2). Gene correction was consistently more effective after restriction enzyme digestion of total DNA using BstI. The reduction in gene repair activity in cells in which MBD4 was knocked down was analyzed as a function of time after MBD4 siRNA transfection. A decrease in gene correction efficiency occurred as early as 24 h after co-transfection of mCpG<sub>GFP</sub> and the MBD4 siRNA compared with that achieved after co-transfection of a control siRNA, and this effect was even greater at 48 h when the siRNA-mediated knockdown was most effective (Figure 10B).
when the MBD4-binding ssODN was used as compared with an ssODN containing the 5-methylcytosine, but not in a CpG context or to an ssODN with no 5-methylcytosine modifications (Figure 8). These data were confirmed at the protein, transcript and genomic levels after treatment of cells with targeting ssODNs.

The differences in gene correction detected do not appear to be due to preferential uptake of ssODNs containing the m5CpG sequence nor to an increase in their stability after transfection, as suggested by FACS analysis of myoblasts transfected with fluorescently labeled control and targeting ssODNs and followed for up to 1 week after transfection (Figure 3). Although some toxicity was observed in ssODNs containing m5C, the increase in cell death was only transient and was not specific to MBD4-binding ssODNs as no significant differences were detected between cells transfected with m5CpGGFP and m5CpCGFP oligonucleotides. Finally, no differences in cell cycle regulation (Table 1) or number of cell divisions (Figure 5) were observed in cells treated with m5CpGGFP as compared with the other ssODNs or to sham transfected cells. Together, these data suggest that the differences in gene correction detected in myoblast cultures were not due to the differences in the ability of MBD4-binding ssODNs to influence cell replication, but were likely due to their ability to increase gene repair. The percentages of GFP-positive cells detected over time after ssODN transfection remained stable in all cultures treated with targeting oligonucleotides, but were significantly higher in cells treated with m5CpGGFP ssODN. These results clearly demonstrate that gene editing in myoblasts is stable over time, confirming the results previously obtained in our laboratory using gene editing strategies for the dystrophin gene (5,10,21). Furthermore, these data demonstrate that the use of methyl-CpG-modified ssODNs activating the BER through MBD4 binding can significantly enhance gene repair.

Restriction enzyme digestion analysis in combination with real-time PCR demonstrated that the use of ssODNs containing m5CpG and capable of binding MBD4 were able to induce up to 10-fold higher levels of gene correction than older generation oligonucleotides lacking any specific modification. In assessing the frequencies of gene correction, we have taken into consideration several parameters that are likely to yield different estimates of efficiency at the genomic, transcript and protein levels. Such variable include the number of copies of the pGFPmut plasmid integrated at the genomic level, level of expression of the plasmid after G418 selection, the possibility of silencing of plasmid gene expression in vitro and the sensitivities of the FACS and western blot analyses to detect low levels of protein expression. The use of quantitative PCR analysis allowed us to detect correction events occurring at the genomic level independently of possible silencing effects and to estimate frequencies of gene repair per total number of vectors integrated in the genome. The possibility that the results obtained were due to PCR artifacts can be excluded on several grounds. First, each real-time PCR performed had several internal controls, including DNA isolated from untransfected cells and cells transfected with the control ssODN and then subjected to PCR analysis after restriction enzyme digestion. The absence of amplification in those latter samples ensured us that the endonuclease digestion had reached completion. Second, PCR analyses following restriction enzyme digestion were repeated in triplicate experiments using different DNA samples isolated after transfection. Third, each run was loaded on agarose gels to confirm that no amplification product, other than the one expected in samples treated with the targeting ssODNs, would be present in the reaction. Finally, PCR products obtained after quantitative analysis were excised from the gel and sequenced to confirm the presence of the desired single base pair alteration.

MBD4 appears to be a major component of the mechanism of gene correction mediated by methyl-CpG-modified ssODNs. Downregulation of MBD4 alone was sufficient to prevent the restoration of GFP expression after modified ssODN treatment, as demonstrated by analysis at both the protein and the genomic DNA levels (Figure 10). These observations are supported by studies in MBD4 knockout models. Mice lacking MBD4 expression accumulate C-to-T mutations at CpG sites at a rate of 2- to 3-fold higher than wild-type mice (49,50). Other repair mechanisms, although important for maintaining genomic integrity in mammalian cells, do not appear to play an active role in recognizing deamination of 5-methylcytosine at m5CpG sites, nor can they efficiently compensate for the loss of MBD4 in mice. Therefore, the use of ssODNs containing methyl-CpG modifications and capable of annealing to the genomic sequence creating a mismatched T opposite the G in the methyl-CpG can be used to specifically recruit and activate the MBD4 pathway. The increase in the level of gene correction by ssODNs containing methyl-CpG modifications is likely the result of the efficacy of the repair mechanisms activated to precisely recognize and cleave the mismatched base on the genomic DNA targeted for repair and opposite to the ssODN.

These interpretations were also supported by EMSA assays which demonstrated that the binding of MBD4 occurred only when a G:T mismatch was created by the binding of an ssODN containing the m5CpG modification to the target sequence (Figure 2). The presence of a single methyl-CpG modification on the ssODN was sufficient to recruit MBD4 to the target sequence. Furthermore, the ability of MBD4 to recognize and bind the targeted sequence only in the presence of the methyl-CpG-modified ssODN annealed to the DNA and creating a G:T mismatch is a clear evidence that initiation of the process requires the ssODN to first anneal to the target sequence within the genomic DNA. No MBD4/DNA complexes were detected when full-length MBD4 protein was added to the reaction mixture containing the methyl-CpG-modified ssODN alone or when MBD4 was added to the reaction containing ssODNs without a methyl-CpG even in the presence of the target sequence (Figure 2).

For therapeutic applications, the levels of gene correction are likely to require higher efficiencies than those currently being achieved in different systems using ssODNs. The use of methyl-CpG-modified ssODNs is a
technical advance that enhances gene repair efficiency as it
directs the correction event specifically to the genomic
sequence targeted for repair, diminishing the likelihood
that the repair mechanism might instead alter the
ssODN sequence. Understanding additional steps in the
biochemical mechanisms that mediate the repair will also
lead to further increases in gene repair efficiencies.

MBD4-binding ssODNs, although more effective than
unmodified ssODNs, are sequence specific and can only
direct conversion of a thymine into a cytosine.
Furthermore, because MBD4 recognizes G:T mismatches
only in the context of CpG sites, the sequence targeted by
the ssODN requires the presence of a guanine immediately
3′ of the base targeted for repair. This further limits the
number of mutations that can be targeted for repair in the
context of human diseases. However, it should be noted
that ssODN-mediated gene editing can be targeted to both
the coding and noncoding strands (3,5,12,51,52), thus
increasing the possible target sequences. In addition,
methyl-CpG ssODNs may be useful in oligonucleotide-
mediated exon skipping (21). This approach takes
advantage of the ability of ssODNs to target and disrupt
consensus sequences necessary for intron/exon splicing
and assembly of mature mRNA. The alteration at the
genomic level causes the skipping of one or more exons
during the assembly of mRNA transcripts, and ssODNs
designed to disrupt specific splicing regulatory elements
can have therapeutic applications in cases where shorter,
in-frame transcripts allow the production of partially
functional proteins (21). This approach has already
being employed in a mouse model of DMD using older
generations of oligonucleotides with encouraging results
(21). Although intron/exon are not specifically enriched
in TG dinucleotides, that are the specific targets of
methyl-CpG-modified ssODNs, such dinucleotides are
frequently present in consensus sequences and other
splicing control elements and are thus potential targets
for altering splicing to restore functional protein expres-
sion from mutant genes. The ability of methyl-CpG-
modified ssODNs to induce single base pair substitutions
at the intron/exon boundaries of the dystrophin gene is
currently being tested in our laboratory. The development
of a successful approach using methyl-CpG-modified
ssODNs in oligonucleotide-mediated exon skipping of
the dystrophin gene would have a wide range of clinical
applications. Its use could be applied to the majority of
DMD patients (60–70%) in which large deletions or
frameshift mutations could be corrected by restoring the
dystrophin reading frame.

Random integration, mispairing and activation of
homologous recombination at sites different from those
targeted for repair might ultimately preclude this technol-
ogy from entering a clinical scenario. For approaches
aimed at repairing postmitotic muscles cells, those issues
remain less of a concern since the effects are likely to be
confined within a limited number of cells that have under-
gone repair. For those approaches aimed at targeting and
correct muscle stem cells, a detailed analysis of cells under-
gone repair will be critical to demonstrate its safety and
long-term effects. Those studies will have to focus on indi-
vidual cells not just a pool of corrected cells so that the
effects of the correction process can be determined in
detail on a single cell base.

Future studies in our laboratory will be aimed at
determining the safety of gene correction process in
muscle cells and will be focused primarily on assessing
the fate of cells that have undergone repair. The
optimization of culturing conditions capable of maintain-
ing in culture clones isolated after cell sorting will be
critical to characterize those cells and study changes
in gene expression profile in each individual cell that
have undergone gene repair. To date, those studies have
been problematic due to the inability to expand primary
myoblasts isolated using sorting techniques. New
technologies are now being developed that will allow us
to study expression profiles from single cells and might
enable us not only to study the effects of gene repair
within each individual cell, but also to determine the
effects of ssODNs immediately after transfection. Those
technologies might ultimately hold the key in determining
the applicability of ssODN-mediated gene correction for
the treatment of many genetic defects.

Ultimately, safe and effective treatments of genetic
diseases using ssODNs will necessitate synergistic efforts
aimed not only at increasing the level of gene repair but
also methods to insure efficient delivery of ssODNs to
all the cells whose correction would be necessary for
maximum therapeutic effect. For diseases in which the
tissue being targeted undergoes continuous turnover
either under normal physiological conditions or in the
setting of the disease, gene correction of the stem or pro-
genitor cells that are responsible for forming new tissue
would be essential to insure sustained therapeutic effects.
These delivery issues are faced by every form of gene
therapy. However, especially for nonviral forms of gene
therapy, the efficacy of the vector once delivered to the
cell, both in terms of magnitude and duration, will be a
critical determinant of the success of that therapeutic
approach. The use of methyl-CpG-modified ssODNs is
just such a method to enhance the efficacy of ssODN-
mediated gene editing, and may have applications across
a wide range of genetic disorders.

FUNDING
UCLA Human Gene Medicine Seed Grant (to C.B.);
grants from the Muscular Dystrophy Association (USA)
to C.B. and T.A.R.).

Conflict of interest statement. None declared.

REFERENCES
1. Cole-Strauss,A., Yoon,K., Xiang,Y., Byrne,B.C., Rice,M.C.,
the mutation responsible for sickle cell anemia by an RNA-DNA
2. Chan,P.P., Lin,M., Faruqi,A.F., Powell,J., Seidman,M.M. and
Glazer,P.M. (1999) Targeted correction of an episomal gene in
mammalian cells by a short DNA fragment tethered to a triplex-


