THE JOURNAL OF GENE MEDICINE
J Gene Med 2011; **13**: 522–537.
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/jgm.1611

RESEARCH ARTICLE

Endonucleases: tools to correct the dystrophin gene

Joel Rousseau[1]
Pierre Chapdelaine[1]
Sébastien Boisvert[2,3]
Luciana P. Almeida[4]
Jacques Corbeil[3]
Alexandre Montpetit[5]
Jacques P. Tremblay[1,3]*

[1]*Unité de recherche de recherche en Génétique Humaine, Centre de recherche de CHUL, CHUQ, Faculté de médecine, Université Laval, Québec, Canada*

[2]*Centre de recherche en infectiologie, Centre Hospitalier Universitaire de Québec (CHUQ), Pavillon CHUL, Québec, Canada*

[3]*Faculté de médecine, Pavillon Ferdinand-Vandry, 1050, avenue de la Médecine, Université Laval, Québec, Canada*

[4]*Department of Biochemistry and Immunology, School of Medicine of Ribeirão Preto, University of São Paulo, SP, Brazil*

[5]*McGill University and Genome Quebec Innovation Centre, Montreal, Canada*

*Correspondence to: J. P. Tremblay, Unité de recherche de recherche en Génétique Humaine, Centre de recherche de CHUL, CHUQ, Faculté de médecine, Université Laval, Québec, Canada.
E-mail: jacques-p.tremblay@crchul.ulaval.ca

Received: 4 July 2011
Revised: 23 August 2011
Accepted: 15 September 2011

Abstract

Background Various endonucleases can be engineered to induce double-strand breaks (DSBs) in chosen DNA sequences. These DSBs are spontaneously repaired by nonhomologous-end-joining, resulting in micro-insertions or micro-deletions (INDELs). We detected, characterized and quantified the frequency of INDELs produced by one meganuclease (MGN) targeting the RAG1 gene, six MGNs targeting three introns of the human dystrophin gene and one pair of zinc finger nucleases (ZFNs) targeting exon 50 of the human dystrophin gene. The experiments were performed in human cells (i.e. 293 T cells, myoblasts and myotubes).

Methods To analyse the INDELs produced by the endonucleases the targeted region was polymerase chain reaction amplified and the amplicons were digested with the Surveyor enzyme, cloned in bacteria or deep sequenced.

Results Endonucleases targeting the dystrophin gene produced INDELs of different sizes but there were clear peaks in the size distributions. The positions of these peaks were similar for MGNs but not for ZFNs in 293 T cells and in myoblasts. The size of the INDELs produced by these endonucleases in the dystrophin gene would have permitted a change in the reading frame. In a subsequent experiment, we observed that the frequency of INDELs was increased by re-exposition of the cells to the same endonuclease.

Conclusions Endonucleases are able to: (i) restore the normal reading of a gene with a frame shift mutation; (ii) delete a nonsense codon; and (iii) knock-out a gene. Endonucleases could thus be used to treat Duchenne muscular dystrophy and other hereditary diseases that are the result of a nonsense codon or a frame shift mutation. Copyright © 2011 John Wiley & Sons, Ltd.

Keywords Duchenne muscular dystrophy; dystrophin; endonuclease; meganuclease; micro-deletion; micro-insertion; reading frame correction; zinc finger nuclease

Introduction

Many hereditary diseases, including Duchenne muscular dystrophy (DMD) are caused by mutations, which lead to a premature termination of protein translation as a result of the presence of a nonsense mutation or a frame shift mutation, which results in a premature stop codon. Internally deleted dystrophin proteins have been shown to have a normal or quasi-normal role because they can integrate into the dystrophin complex and give rise either to asymptomatic subjects or to Becker muscular dystrophy with a milder phenotype [1–3].

Gene targeting is the ultimate tool for genetic modifications for eventually treating a variety of genetic diseases, but its use is often limited by its low

efficiency. Meganucleases (MGNs), also called homing endonucleases, are sequence-specific endonucleases, which recognize unique large (> 12 bp) target sites in living cells [4]. Zinc finger nucleases (ZFNs) are another type of endonucleases, which can also be engineered to target specific DNA sequences [5–12]. More recently, a third type of endonucleases, the Tal effector nucleases (TALENs) has also been engineered to mutate specific DNA sequences [13–18]. All three types of endonucleases induce site-specific double-strand breaks (DSBs), which stimulate the rate of homologous recombination (HR) up to 10000-fold in cultured cells [19,20]. However, repair by HR requires the insertion in the cells of a donor DNA molecule sharing homologies with the targeted gene. Endonuclease induced DSB can also be repaired by nonhomologous end-joining (NHEJ), an error prone process, which frequently results in micro-insertions or micro-deletions (INDELs) at the site of the break [21–23]. However, NHEJ has the great advantage that it does not require any donor DNA and, thus, when introduced in the cells *in vitro* or *in vivo*, only the endonuclease protein could produce NHEJ. It is now possible to engineer highly specific redesigned endonucleases recognizing chosen sequences from almost any chromosomal locus [7,8,10,11,13,15,24–29]. Thus, endonucleases represent universal tools to mutate the genome at specific target sequences. Moreover, it is possible to permanently mutate the genome with only the transduction of an endonuclease protein, avoiding the requirement to introduce in the cells the DNA coding for the endonuclease and a donor DNA.

The present study describes three procedures for characterizing and quantifying INDELs induced by endonucleases at a specific site within a gene. All three procedures include a first step consisting in the polymerase chain reaction (PCR) amplification of a short DNA sequence surrounding the chromosomal target site, although they differ in the techniques used to characterize the resulting amplicons. The first procedure, the Surveyor assay, has already been described [11,28,30]. The second procedure, called subtractive colony hybridization (SCH), is the only one to include the cloning of an individual amplicon. PCR products were cloned in bacterial plasmids, and products containing INDELs were detected by two rounds of negative *in situ* hybridization in bacteria, using a specific oligonucleotide radioactive probe corresponding to the MGN target. This approach permits a semi-quantitative evaluation of INDELs produced by a MGN inside cells. The third procedure is based on based on deep barcoded sequencing (DBS). The DBS method has previously been used by Meng *et al*. [31]. It is both quantitative and extremely sensitive and can thus be used to compare the efficacy of various procedures for delivering the endonuclease genes or proteins not only *in vitro*, but also *in vivo*. Moreover, this procedure permits an exact determination of which base pairs have been added or deleted, and thus determine whether the reading frame of the resulting mRNA has been modified. Our results also demonstrate that the INDELs induced by endonucleases modify the reading frame of the dystrophin and could thus be used to restore the normal reading frame of a dystrophin gene with an out-of-frame deletion.

Material and methods

Materials

Meganuclease-coding plasmids were provided by Cellectis SA Inc. (Romainville, France). The plasmid coding for *RAG1*, corresponds to the scV3^{-}-V2^{+}(G19S) meganuclease described in a previous study [32]. Six MGNs targeting three dystrophin introns (MGN2874, MGN3387, MGN3631, MGN3633, MGN3326 and MGN3330) were produced based on the same scaffold as the *RAG1* meganuclease. The *RAG1* MGN is under the EF1α promoter, whereas the MGNs targeting dystrophin are under the cytomegalovirus (CMV) promoter. The ZFNs targeting the dystrophin gene were purchased from Compo ZR Inc. (Oakville, ON, Canada). The ZFNs were under the control of a CMV promoter in a 4129 bp expression vector called pZFN. Targeted sequences are indicated in Table 1. The plasmid vector containing the dog μ-dystrophin fused to a V5-tag and containing an insertion including the *RAG1* MGN target sequence has been previously described [33]. Surveyor nuclease enzyme was purchased from Transgenomic Inc. (Omaha, NE, USA) and used as in accordance with the manufacturer's instructions to detect INDELs. Phusion High-Fidelity DNA Polymerase (Finnzymes Qy Inc., Vantaa, Finland) and TAQ DNA polymerase used for PCR were purchased from New England Biolabs Ltd (Pickering, ON, Canada). Qiaquick gel extraction kit and Qiagen PCR cloning kit were from Qiagen Inc. (Valencia, CA, USA). The vector pLenti6/V5-D-TOPO and Lipofectamine 2000™ were from Invitrogen Canada Inc. (Burlington, ON, Canada).

Transfection of 293 T cells with endonuclease plasmids

Half a million 293 T cells were plated in a six-well plate. The following day, these cells were transfected with 4 μg of each of the seven MGN or Flag-ZFN plasmids using Lipofectamine 2000™. Three days later, the genomic DNA was extracted from the cells [34] to detect INDELs with the Surveyor enzyme, SCH or DBS procedure. The analysis of INDELs by SCH and DBS was also carried out for 293 T cells transfected either one or four times consecutively with MGN *RAG1* at a 3-day interval using Lipofectamine 2000™. In some experiments, the proteins were also extracted from the cells and analyzed by western blot with a specific polyclonal antibody developed against *I-CreI* MGN that is able to detect the expression of the different MGN [33]. The expression of the Flag-ZFN was confirmed by western blotting using an anti-Flag monoclonal antibody (Sigma-Aldrich Canada Inc., Oakville, Canada).

Table 1. Nucleotide sequences targeted by the different endonucleases

MGN name	Dystrophin intron/exon	Targetted sequence
Rag1		TTGTTCTCAGGTACCTCAGCCAG
I-SceI		CACGCTAGGGATAACAGGGTAATAT
MGN2874	Intron 38	GAAACCTCAAGTACCAAATGTAAA
MGN3387	Intron 38	GAAACCTCAAGTACCAAATGTAAA
MGN3631	Intron 44	AATGTCTGATGTTCAATGTGTTGA
MGN3633	Intron 44	AATGTCTGATGTTCAATGTGTTGA
MGN3326	Intron 42	CAAATCCTGCCTTAAAGTATCTCA
MGN3330	Intron 42	CAAATCCTGCCTTAAAGTATCTCA
ZFNs	Exon 50	CTAGCTCCTGGACTGACCactattGGAGCCTGTAAGTATACTG

Note that there are two different MGNs targeting the same dystrophin sequence.

Lentivirus constructs coding for MGNs

All the plasmids coding for different MGNs obtained from Cellectis Inc. (Romainville, France) have a similar structure. A simple digestion of MGN vector with restriction enzymes *Nde*1 and *Age*1 liberated a DNA fragment of approximately 1.5 kb containing a portion of CMV promoter and a complete MGN coding sequence, which was subsequently cloned directly in the same restriction enzyme sites of a lentiviral vector named pLenti6/V5-D-TOPO in which the blasticidin resistant gene had been replaced by a puromycin resistant gene. The resulting final constructs contained a MGN coding sequence driven by a CMV promoter already present in the lentiviral vector. The lentiviruses coding for different MGNs were produced in the 293 T cells as described below. The supernatant containing the virus were then used to infect human myoblasts. The cells were selected with 2 mg/ml puromycin (2 days) and propagated. Western blot analysis with a specific polyclonal antibody to *I-CreI* MGN, reacting with an epitope common to all these MGNs, was carried out to confirm the expression of each MGN.

Nucleofection of human myoblasts with an endonuclease plasmid

Human myoblasts were nucleofected with 5 μg of an endonuclease plasmid. Nucleofection solution adult (NHDF-adult, Lonza, no. CC-2511; ESBE Scientific, Markham, ON, Canada) was used and the nucleofection apparatus (Amaxa Inc., Amaxa Nucleofector System, Lonza Walkersville Inc., Walkersville, MD, USA) was set to program P-022. After the nucleofection of plasmid, the cells were grown during 3 days in MB1 medium (Hyclone Inc., Logan, UT, USA) culture medium followed by DNA extraction. The endonuclease-targeted region was amplified by PCR. The resulting amplicons were treated with the Surveyor enzyme, as described below.

Western blot analysis of different endonucleases in 293 T cells

Proteins were also extracted from 293 T cells nucleofected with 5 μg of an endonuclease plasmid and analyzed by western blotting with a specific polyclonal antibody developed against the *I-CreI* MGN able to detect the expression of the different MGNs [33] or with an antibody against the Flag (Sigma-Aldrich Canada Inc.) for the ZFNs. An aliquot of 20 μg of protein was loaded in each lane, resolved in 12% sodium dodecyl sulphate (SDS)-polyacrylamide gel electrophoresis (PAGE) and electrotransferred onto 0.45 μm nitrocellulose membrane. The membrane was blocked in 5% (w/v) nonfat dry milk resuspended in phosphate-buffered saline (PBS) containing 0.05% Tween-20 for 1 h and incubated overnight at 4 °C with a rabbit polyclonal antibody (dilution 1:5000 for MGN) or monoclonal anti-FLAG (dilution 1:1500 for ZFN) serum diluted at 1:5000 in PBS-Tween. The membrane was washed for 10 min in PBS with 0.05% Tween three times and incubated with goat anti-rabbit (dilution 1:10000) Ab or rabbit anti-mouse (dilution 1:1500) for 1 h at room temperature. After three washes of 10 min each, the membranes were incubated with an enhanced chemiluminescent reagent (Renaissance reagent, New England Nuclear Corporation, Perkin-Elmer, Waltham, MA, USA) substrate and exposed to autoradiography film.

Surveyor enzyme procedure

The surveyor enzyme procedure is explained in Figure 1A. Briefly, 100 ng of genomic DNA were PCR amplified with the 2X Phusion mix (Finnzymes QY Inc.) in a total volume of 50 μl, using the primers for RAG1 or for each of the dystrophin target (Primers 1–14 in Table 2). PCR parameters were: one step at 98 °C for 1 min followed by 30 cycles of 98 °C for 30 s, 60 °C for 30 s and 72 °C for 30 s. Amplicons were purified with the Qiaquick gel extraction kit (Qiagen Inc.) in accordance with the manufacturers instructions. Purified PCR product (10 μl) was submitted to the Surveyor enzyme protocol in accordance with the manufacturer's instructions.

SCH protocol to identify INDELs caused by endonucleases

A new five-steps procedure, termed SCH (Figure 2A), was developed to characterize INDELs produced in the genome of cells treated with different endonucleases.

DOI: 10.1002/jgm

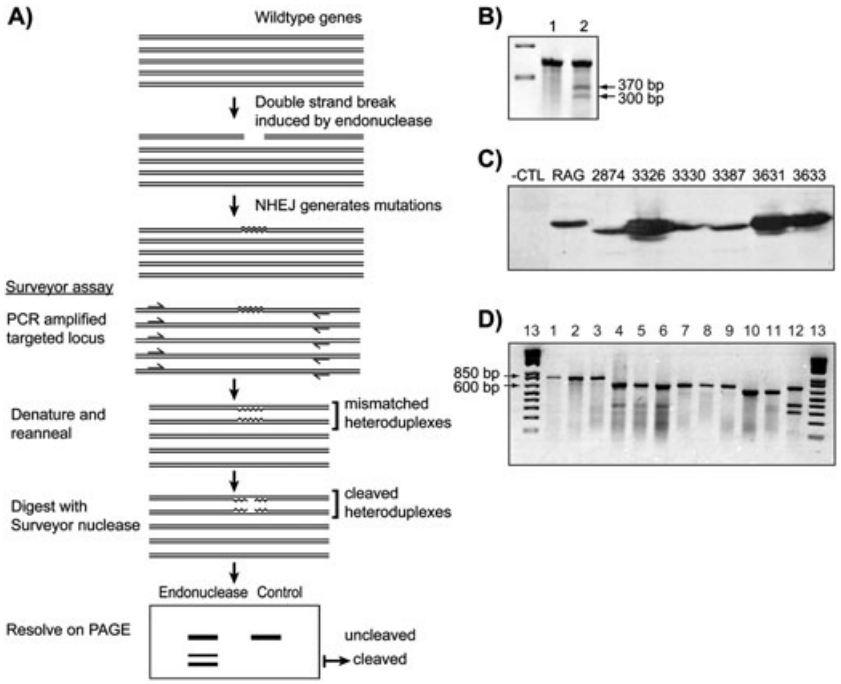

Figure 1. Surveyor enzyme procedure. 1 (A) Schema of the Surveyor enzyme procedure: a DSB is induced by an endonuclease in some of the target genes. These DSBs are spontaneously repaired by NHEJ resulting in INDELs in the target genes, which have been cut. The gene region targeted by an endonuclease can be amplified by PCR using primers located approximately 300 bp from the targeted sequence. This produces amplicons of approximately 600 bp. Only a fraction of these amplicons contain INDELs. The amplicon mixture is then heated at 95 °C to denature the amplicons and re-annealed by slowly cooling at room temperature. Some mismatched hetero-duplexes are then formed. These hetero-duplexes are cut by the Surveyor enzyme producing shorter double-strand DNA fragments. These fragments are separated from the uncut homo-duplexes by separation on agarose gel or by PAGE. (B) Surveyor enzyme procedure on human myoblasts infected with the lentivirus μ-dyst*RAG1*/V5 and nucleofected with the *RAG1* MGN. Human myoblasts were infected with a lentiviral vector coding for μ-dyst*RAG1*/V5, a resistance gene to puromycin, and for μ-dyst*RAG1*/V5. They were selected with the puromycin and proliferated during 2 weeks. Cells were then nucleofected or not with the *RAG1* MGN and their genomic DNA was extracted. The μ-dyst*RAG1*/V5 gene region targeted by the *RAG1* MGN was amplified by PCR. The amplicons were digested with the Surveyor enzyme. Line 1 represents not treated with the *RAG1* MGN. There is only one band at 657 bp, which is a result of amplicons not cut by the Surveyor enzyme. In lane 2, cells were treated with the *RAG1* MGN. Two additional bands are present at 370 and 300 bp because the MGN has induced INDELs and therefore there were hetero-duplexes present in the amplicons that were cut by the Surveyor enzyme. (C) Confirmation of the expression of the MGNs targeting dystrophin and RAG1 genes: Plasmids coding for *RAG1* or for the various MGNs (2874, 3387, 3631, 3633, 3326 and 3330) targeting the dystrophin gene were transfected in 293 T cells. The proteins were extracted 3 days later and positive western blots were obtained using an antibody against an epitope present in all these MGNs. (D) Surveyor enzyme results for MGNs targeting dystrophin: the genomic DNA was extracted from the 293 T cells used in Figure 1C. Cells in lanes 1, 4, 7 and 10 were not transfected with a MGN plasmid. Cells in lanes 2, 3, 5, 6, 8 and 9 were transfected, respectively, with MGN2874, MGN3387, MGN3631, MGN3633, MGN3326 and MGN3330. Specific dystrophin genomic regions targeted by these MGNs were amplified by PCR. Lane 11 transfected with the *RAG1* MGN served as a negative control. PCR products were heated, re-annealed and digested with the Surveyor enzyme. Lane 12 is a positive control for the Surveyor enzyme containing a 1:1 mixture of amplicons of μ-dyst*RAG1*/V5 and μ-dyst*I-SceI*/V5 gene. Lane 13 contains molecular weight markers. The Surveyor enzyme produced additional bands in lane 12 but not in the MGN-treated cells compared to their respective negative controls (lanes 2 and 3 versus lane 1; lanes 5 and 6 versus lane 4; lanes 8 and 9 versus lane 7; lane 11 versus lane 10). Thus, the MGNs targeting the endogenous dystrophin gene did not induce INDELs that were sufficiently abundant to be detected by the Surveyor enzyme test.

The first step: 100 ng of genomic DNA was amplified with the Phusion High-Fidelity DNA Polymerase (New England Biolabs Inc., Ipswich, MA, USA) (one cycle at 98 °C for 1 min; 25 cycles at 98 °C for 10 s, 60 °C for 20 s and 72 °C for 30 s and finally one cycle at 72 °C for 10 min) using primers 1–14 specified in Table 2. At the end of the PCR reaction, 0.5 μl of Taq DNA polymerase (NEB) was added to the PCR medium and brought at 72 °C for 10 min and then loaded and electrophoresed on a 1.4% agarose gel containing ethidium bromide for DNA visualization under ultraviolet (UV) light. A fluorescent 600-pb band was excised from the gel (Qiagen gel extraction kit) and this amplicon was cloned in the TA cloning vector pDrive (Qiagen PCR cloning kit). The ligation medium (10 μl: 0.5 μl pDrive (50 ng/μl), 4.5 μl PCR product, 5 μl ligation 2X) was incubated overnight at 13 °C. The 10 μl of the ligation final volume was mixed directly with 100 μl of DH5a bacteria. The transformed bacteria were spread on 3 LB-agar plates containing ampicillin (100 μg/ml) and kept at 37 °C during 18 h, giving more than 100 colonies per plate. Second step: the next day, the bacterial colonies on the agar plate were transferred to a nylon membrane 0.45 μm (GE Health Bio-Sciences Inc, Baie-D'Urfé, QC, Canada). The membrane (with the colonies on the upper side) was treated successively for 7 min on a membrane Whatman 3 MM CHR chromatography paper (Fisher Scientific Inc., Cooksville, ON, Canada) dipped in a solution containing 0.5 M NaOH and 1.5 M NaCl and twice in 1 M

Table 2. Primer sequences

Primer #	Forward/ reverse	Gene	Nucleotide positions	Sequence	Amplicon length
1	Forward/	m-dyst$_{Rag1}$	1027–1051	GACAGTTATCAAACAGCTTTGGAAG	657
2	Reverse	m-dyst$_{Rag1}$	1684–1673	GTAATCTGTGGGTGTCTTGTAAAAGA	
3	Forward/	m-dyst$_{I\text{-}SceI}$	1027–1051	GACAGTTATCAAACAGCTTTGGAAG	657
4	Reverse	m-dyst$_{I\text{-}SceI}$	1684–1673	GTAATCTGTGGGTGTCTTGTAAAAGA	
5	Forward/	RAG1a	9969–9988	GGGAGGCAAAGATGAATCAA	574
6	Reverse	RAG1a	10542–10523	GGGCTTTTAACAATGGCTGA	
7	Forward/	dystrophin intron 38	997822–997843	TCTTGCAGCCTAAAGGAACAAA	797
8	Reverse	dystrophin intron 38	998618–998597	TCCTCTCGCTTTCTCTCATCTG	
9	Forward/	dystrophin intron 42	1036576–1036599	GCAGAGCTAGAGAAGAATGAGAAA	668
10	Reverse	dystrophin intron 42	1037243–1037220	TTTGTTATTGGTTGAGGTTTGCTG	
11	Forward/	dystrophin intron 44	1130033–1130054	GAACAGGTGGTATTACTAGCCA	669
12	Reverse	dystrophin intron 44	1949881–1949860	GGTTGCAGTGAGCTGAGATCAT	
13	Forward/	dystrophin exon 50	1524446–1524469	TTCACCAAATGGATTAAGATGTTC	406
14	Reverse	dystrophin exon 50	1524845–1524826	ACTCCCCATATCCCGTTGTC	
15	Deep Forward/	dystrophin intron 38	998314–998336	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCCTACACGACGCTCTTCCGA TCTNNNGTGTTCTTTACAACTAGATGGGA	219
16	Deep Reverse	dystrophin intron 38	998414–998390	CAAGCAGAAGACGGCATACGAGATCGGT CTCGGCATTCCTGCTGAACCGCTCTTCCGAT CTNNNCATGGAATGCTATTATAAGATACTG	
17	Deep Forward/	dystrophin intron 42	1036800–1036823	AATGATACGGCGACCACCGAGATCTACACT CTTTCCCTACACGACGCTCTTCCGATCTNNNA TTTTCCAAAATATAATTACTGTG	222
18	Deep Reverse	dystrophin intron 42	1036896–1036874	AAGCAGAAGACGGCATACGAGATCGGTCT CGGCATTCCTGCTGAACCGCTCTTCCGATCT NNNAACTTACGATGCTAAGAAATTCT	
19	Deep Forward/	dystrophin intron 44	1130279–1130302	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACGCTCTTCCGATCTNN NTAACAAAAGCCTCCTGGCAAGGAT	249
20	Deep Reverse	dystrophin intron 44	1130409–1130386	AAGCAGAAGACGGCATACGAGATCGGTCT CGGCATTCCTGCTGAACCGCTCTTCCGATCT NNNCTGCCTGAAACACCTGAAACACTT	
21	Deep Forward/	RAG1a	10234–10258	AATGATACGGCGACCACCGAGATCTACACT CTTTCCCTACACGACGCTCTTCCGATCTNNNG ATATTGATATTGGTCTTAATATGA	218
22	Deep Reverse	RAG1a	10326–10301	CAAGCAGAAGACGGCATACGAGATCGGTCT CGGCATTCCTGCTGAACCGCTCTTCCGATCTN NNTGAGTCCCAAGGTGGGTGGGAAAGA	
23	Deep Forward/	dystrophin exon 50	1524579–1524601	AATGATACGGCGACCACCGAGATCTACACT CTTTCCCTACACGACGCTCTTCCGATCTNN NAAGAGCTGAGGGCAAAGCAGCCT	214
24	Deep Reverse	dystrophin exon 50	1524674–1524652	CAAGCAGAAGACGGCATACGAGATCGGTCT CGGCATTCCTGCTGAACCGCTCTTCCGATCT NNNATAGCTAGAGCCAAAGAGAATGG	

Reference rag sequence: NG_007528.1. Reference dystrophin sequence: NG_012232. Primers 1–12 were used to produce long amplicons for the Surveyor enzyme and the SCH procedures. These primers were also used for the first five cycles of the DBS procedure. Primers 13–20 were used for re-amplification of the long amplicons for the DBS procedure. The nucleotide sequences in red are the forward adapters for deep sequencing on the Illumina instrument; thus, they are the same for primers (13, 15, 17 and 19). The three nucleotides indicated as NNN are the barcodes permitting sequencing in the same Illumina lane as the same target gene. The nucleotide positions are given relative to our μ-dystrophin gene, the RAG1 gene (Gene bank NG_007528.1) and the dystrophin gene (GenBank NG_012232).

Tris-HCl (pH 8.0) with 1.5 M NaCl each time for 7 min. After a light air-drying at room temperature, the membrane was exposed to a UV light during 3 min and hybridized with a radioactive ^{32}P-oligonucleotide in hybridization solution (6 × SSC (1 × SSC being 0.15 M NaCl/0.015 M sodium citrate, pH 7.0), 2.5 × Denhardt's, salmon sperm DNA (300 mg/ml) and SDS 1%) during at least 4–18 h at 60 °C in a water bath. The oligonucleotide probe was 24 nucleotides long and complementary to the sequence targeted by a given MGN. It was radioactively labelled with T4 polynucleotide kinase (Invitrogen Inc.) in presence of ^{32}P-γ ATP (6000 Ci/mmol) at 37 °C and then purified. After hybridization, the membrane was washed several times in 1 × SSC solution at 60 °C, air-dried and exposed to X-ray film at −80 °C overnight. The next day, the developed film was superposed to the original agar plate and bacteria colonies presenting no radioactive hybridization signal were picked up and seeded in order on a second agar plate divided into squares and also at the same time in the LB medium to be grown overnight at 37 °C. Third step: a second hybridization was carried out (i.e. the colonies on the second agar plate were transferred to a nylon membrane, hybridized with the same radioactive probe and exposed to a film to confirm that the colonies picked in step 2 were really negative). Fourth step: the plasmids were prepared using the bacteria grown in the LB medium but only for the colonies, which were still negative in step 3. These plasmids were digested with *Eco*R1 and the digestion products were analyzed on a 1.4% agarose gel to confirm the presence inside the pDrive vector of an insert of a size similar to the amplicon obtained in the first step. Fifth step: these plasmids with an insert were then

DOI: 10.1002/jgm

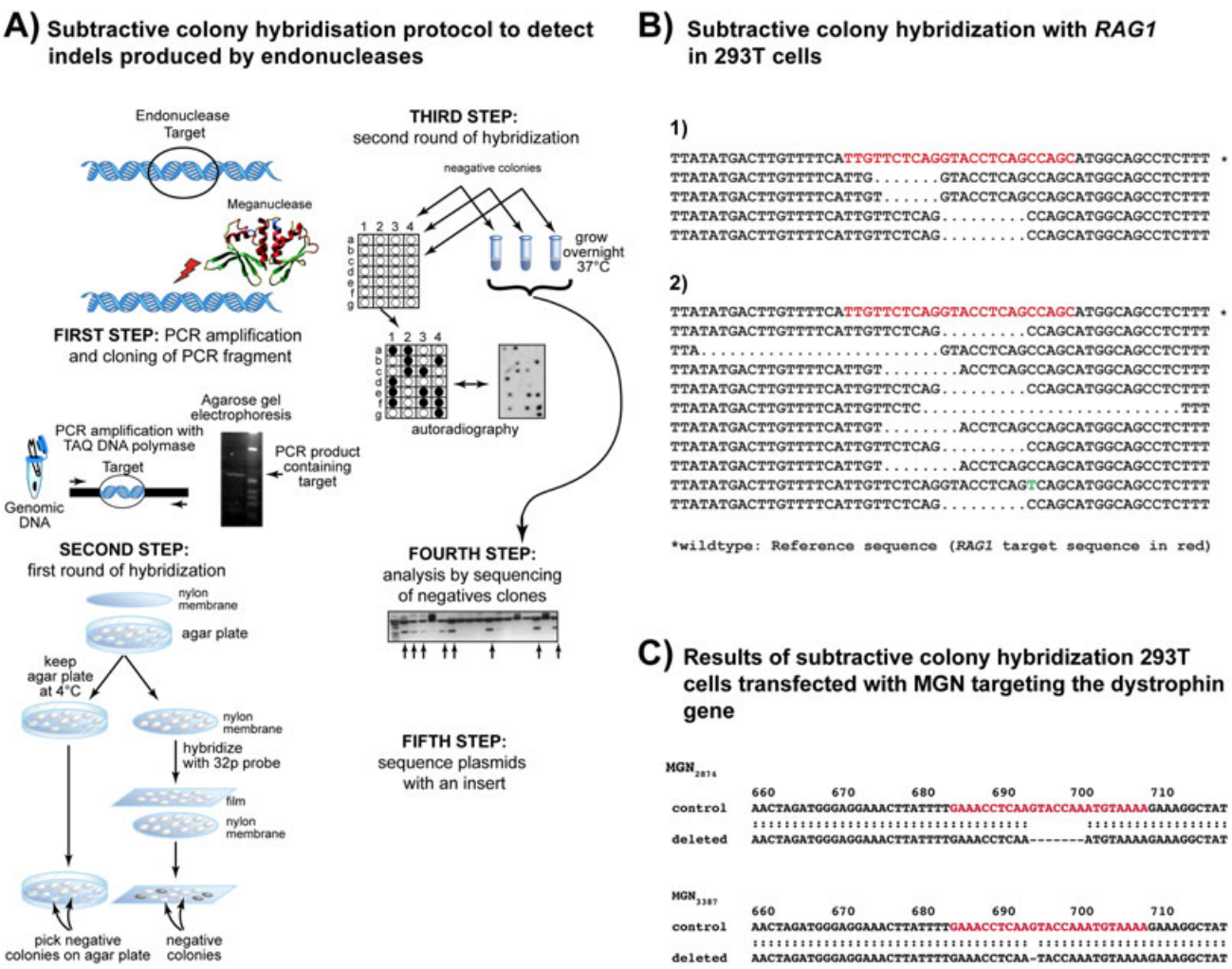

Figure 2. The subtractive colony hybridization (SCH) procedure. 2 (A) Schema of the procedure: This procedure has five steps. Step 1: amplicons of the region targeted by a given MGN were cloned in a TA vector. Step 2: bacterial clones were grown on an agar plate and transferred to a nylon membrane. This membrane was hybridized with a radiolabelled probe specific for the sequence targeted by the MGN. This permitted selection of the negative colonies (unlabelled) in the original agar plate. Step 3: a second round of hybridization eliminated false negative clones. Step 4: the DNA of negative clones was digested with *Eco*R1 to select clones containing insert of the expected sizes. Step 5: the inserts were sequenced and analyzed for the presence of INDELs. (B) Results of SCH in 293 T cells with *RAG1*: the SCH procedure was used to characterize the INDELs produced in 293 T cells transfected one (B1) or four (B2) times with the *RAG1* MGN. The sequence of colonies containing a micro-deletion in the RAG1 gene are illustrated. (C) Results of SCH in 293 T cells with MGN targeting the dystrophin gene: the SCH was also used to identify INDELs produced by the MGN2874 and MGN3387 targeting intron 38 of the endogenous dystrophin gene in 293 T cells. The technique permitted identification of only one micro-deletion in that gene produced by each MGN.

sequenced with the dideoxy method (Sanger) to detect the presence of INDELs produced by the MGN. The sequencing was done using fluorescent dye sequencing [i.e. the sample (recombinant plasmid (mini-prep)] was electrophoresed on capillaries with a sequencing apparatus of type ABI 3730/XL.

DNA preparation for DBS

Genomic DNA was extracted from 293 T or human myoblasts. For DBS, 100 ng of genomic DNA were PCR amplified for only five cycles using the same PCR primers and amplification parameters as for the Surveyor enzyme or SCH procedures. Five microlitres of PCR medium (first step) were then used for a second PCR amplification with a new set of forward and reverse primers containing nucleotides (23–25 nts) complementary to the target gene (primers 15–24 in Table 2). These primers contained in their 5′ end additional nucleotides (58 nts for Deep forward and 61 nts for Deep reverse primers (adapters required for the Illumina® sequencing technology) [35] followed by a three nucleotides barcode. It should be noted that our barcoding procedure differs from the one previously described by Illumina Inc. (San Diego, CA, USA) [36]. For the PCR amplification (50 µl final volume), 100 ng of the DBS primers reverse and forward mixed with 5 µl of the first PCR reaction were used as follows: one cycle at 98 °C for 1 min followed by five cycles at 98 °C for 5 s, 55 °C for 10 s and 72 °C for 10 s and 28 cycles at 98 °C for 5 s, 65 °C for 10 s and 72 °C for 10 s. The amplicons resulting from the second PCR were electrophoresed

on agarose gel stained with ethidium bromide and then purified with QIAquick gel extraction kit (Qiagen Inc.) in accordance with the manufacturer's instructions. The resulting size of the amplicon issue from the second PCR with DBS primers was smaller than the amplicons obtained in the first PCR used for Surveyor enzyme or SCH procedures (Table 2); this second PCR being a nested PCR. Because only 76 nucleotide sequences were obtained by DBS, the primer sequences complementary to the target gene were positioned at approximately the same distance from the sequence targeted by a given endonuclease (Table 1). The concentrations of the purified amplicons were evaluated with PicoGreen® dsDNA Quantification Reagent (Molecular Probes Inc., Eugene, OR, USA). The resulting fluorescence was read (Ex 502 nm/Em 523 nm) on Victor^3 multi-label reader (Perkin-Elmer). An equal amount of the different amplicons were pooled together to obtain a final DNA concentration of 10 ng/μl and sent for DBS at the McGill Genome Center. Pools containing approximately 0.02 pmol of each sample and phiX control DNA were used for cluster generation on the cBot and paired-end sequencing on one lane of the Illumina Genome Analyser IIx (Illumina Inc.) for 2×76 cycles.

Analysis of the sequences obtained with the Illumina® instrument

The Illumina Genome Analyser IIx permits approximately 25 million sequences of 32–200 bp per lane to be obtained. A flow cell contains eight lanes producing a total of 200 million sequences. The number of sequences obtained per lane was much more than needed for the present study. We modified the Illumina sequencing adapters to permit multiplexing to analyse simultaneously, in the same lane, the activity of different endonucleases under different experimental conditions. Because the number of sequences needed for our experiments was low compared to the sequencing capacity of the Illumina instrument, our samples were usually mixed with other samples of yeast genome sequencing. Our samples represented only 1% of the total DNA loaded in one lane. The individual amplicon sequences were retrieved from the pooled reads using the 3-bp barcode and the first 8 bp of the specific primers used from each amplicon. Between 15,000–65 000 paired reads of 76 bases were obtained for each of our specimens. These sequences included the 3-bp barcode at each end, the forward or reverse primer sequences, which varied between 23 and 25 pb (Table 2). Thus there were between 47 and 53 bp in the target gene where INDELs could be detected. Therefore, 72 or 73 bp (excluding the barcode, but including the forward primer or the reverse primer plus the 47–53 bp of the target gene) were aligned onto the wild-type human reference gene (RAG1 or dystrophin) using BLAST software [37]. This allowed the determination of how many sequences had mutations and their separation from the wild-type sequences for further analyses. Each of the mutated sequence was then compared with the wild-type sequence (NG_007528.1 for the RAG1 gene and NG_012232 for the dystrophin gene) using software that developed for the present study. This was carried out using the 3 bp of multiplex tag sequence and the first 8 bp of the specific primers used form each amplicon. The BLASTALL software was then used to align selected sequences with the reference gene permitting to detect the exact position and composition of each INDEL. This software permitted the determination of how many base pairs were inserted or deleted, as well as the exact position of the event. For the final result compilation, the primer sequences were masked. We calculated that the overall sequencing error rate of the Illumina instrument was approximately 0.6% by aligning phiX to its genome using the GERALD tool from Illumina. To eliminate these sequencing errors, we used forward and reverse sequences, which were overlapping, and only the matching sequences were retained for further analysis. This pairing of the forward and reverse sequences eliminated the sequencing errors made by the Illumina instrument. However, this pairing of the forward and the reverse sequences could not be performed for MGN3631 and MGN3633 because a long stretch of thymidine near the targeted sequence precludes the design a reverse primer that was sufficiently close to the forward primer to obtain an amplicon of approximately 76 bp, permitting the overlap of the forward and the reverse sequences obtained with the Illumina instrument. Thus, for these two MGNs some of the 1-bp mutations may be sequencing errors of the Illumina instrument, which produces an approximately 0.6% bp identification error but never produces INDELs.

ZFNs

A pair of ZFNs targeting a sequence in exon 50 of human dystrophin (Table 1) was purchased from Compo ZR Inc.

Results

Detection of INDELs with the Surveyor enzyme

Figure 1A shows a schema on the Surveyor enzyme methodology used to detect the activity of an endonuclease targeting a specific sequence in the genome of a cell. The DNA break caused by an endonuclease is subsequently repaired within the cell by NHEJ generating a mutation of the target gene. The initial experiments to establish the Surveyor enzyme method in our laboratory were performed with human myoblasts infected with a lentivirus coding for a modified micro-dystrophin gene containing the target sequence for the *RAG1* MGN and a V5 flag at the 3′ end (μ-dyst$_{RAG1/V5}$) [33]. These cells were then nucleofected with a plasmid coding for the *RAG1* MGN. The sequence within the μ-dyst$_{RAG1/V5}$ gene targeted by this MGN was then amplified by PCR (using

DOI: 10.1002/jgm

primers 1 and 2 in Table 2). The resulting amplicons were then digested with the Surveyor enzyme and the digested DNA was separated by gel electrophoresis. The presence of two additional bands confirmed that the RAG1 MGN had mutated the target gene (Figures 1A and 1B). It should be noted that, although the Surveyor enzyme procedure permitted confirmation that the *RAG1* MGN mutated the target sequence, it did not permit an exact quantification of the frequency of the induced mutations, nor the indentification of which base pairs were added or deleted.

Six different MGNs targeting three sequences located in introns 38, 42 and 44 of the human dystrophin gene were then produced (Table 1). The sequences targeted by these MGNs are unique in the human genome as confirmed by a BLAST with GRCh37/hg19. The plasmids coding for each of these MGNs were transfected in a human 293 T cells. The expression of all six MGNs was confirmed by western blot analysis (Figure 1C). The regions targeted by the six MGNs were then amplified by PCR (Primers 7–12 in Table 2). These amplicons were then digested with the Surveyor enzyme. No additional bands as a result of mutations of the human dystrophin gene by these MGNs were observed (Figure 1D). This result indicates that either these MGNs were inactive or that the frequency of mutations induced by these MGNs was low. Indeed, to detect additional bands with the Surveyor enzyme method, the mutation frequency should be at least 1–3% [38–40]. Thus, a more sensitive method has to be developed to verify whether these new MGNs targeting human dystrophin had some activity.

SCH

The Surveyor enzyme procedure can indicate that INDELs are present in the target gene but it does not provide any information about the exact frequency and size of the INDELs. It also does not permit identification of the deleted or inserted base pairs. Therefore, a second procedure, namely SCH, was developed to obtain this information (Figure 2A). Because our previous experiments had confirmed that the *RAG1* MGN was active, the procedure was initially tested using this MGN in 293 T cells, transfected one (Figure 2B1) or four consecutive times (Figure 2B2). The RAG1 gene region targeted by the *RAG1* MGN was first amplified by PCR using the same primers as for the Surveyor enzyme experiment (Primers #11–12 in Table 2). These amplicons were cloned in the pDrive plasmid. These plasmids were then transduced in bacteria. A total of 395 and 188 bacterial colonies were present on the agar plate, respectively, for the analysis of 293 T cells transfected one or four times. The bacteria were hybridized with a radioactive oligonucleotide probe, which was complementary to the target sequence; respectively, 28 and 40 of them were negative after the first hybridization, suggesting that the RAG1 gene sequence cloned in these bacteria had been mutated by the *RAG1* MGN. However, only 18 and 32 colonies, respectively, were still negative after the second hybridization. The DNA of these negative colonies were digested with *Eco*R1 to verify that there was an insert in the pDrive plasmid and only six of them contained an insert in the pDrive plasmid for 293 T cells transfected 1 time with *RAG1* MGN, whereas 12 clones contained an insert for 293 T cells transfected four times. These six and 12 colonies were sent for sequencing. Among them, two for each experiment had a wild-type insert in the pDrive plasmid (i.e. identical to the RAG1 MGN target). However, four colonies in the first experiment contained a pDrive plasmid with an INDEL identified by sequencing (Figure 2B1), whereas ten colonies in the second experiment presented nine deletions and one insertion (Figure 2B2). The percentage of INDELs was calculated as the number of bacterial colonies with an INDEL divided the total number of colonies on the agar plate minus the colonies, which did not contained an insert in the pDrive plasmid: $(4/[395-(18–6)]) \times 100 = 1.0\%$) or $(10/[188-(32–12)]) \times 100 = 5.9\%$. Thus, the SCH permitted determination of the frequency of INDELs and identification of the bp, which were inserted or deleted.

The SCH procedure was also used to identify the INDELs produced by the MGN2874 and MGN3387 targeting intron 38 of the endogenous dystrophin gene in 293 T cells. However, this procedure permitted identification of only one micro-deletion in the dystrophin gene produced by each MGN (Figure 2C). Thus, this procedure permitted identification of INDELs within the cell genome produced at low frequency by a MGN. However, this procedure was too labour intensive and was abandoned because in, the mean time, we had developed the DBS procedure, which was both less labour intensive and more sensitive.

DBS

Experiments using the RAG1 MGN targeting the RAG1 gene in 293 T

The DBS procedure is schematized in Figure 3A. It was first tested for the detection of INDELs using the *RAG1* MGN because the Surveyor enzyme procedure and the SCH had already confirmed its activity. Examples of INDELs induced by MGN *RAG1* in the 293 T cell genome and detected by DBS are presented in Figure 3B. As summarized in Table 3, 15501 RAG1 amplicons from 293 T control cells (not transfected) were sequenced and none were mutated. However, 1031 micro-deletions and 15 micro-insertions corresponding to 6.69% of INDELs were detected using 15844 forward and reverse sequences of amplicons of 293 T cells transfected with the *RAG1* MGN (Table 3; see also Supporting information, Table S1). Moreover, this also permitted identification of exactly which bp were deleted or inserted (examples in Figure 3B) and the exact frequency of various INDELs was easily determined (Figure 3C). The size distribution of the micro-deletions was not uniform as initially hypothesized; indeed, a peak was observed for the 9-bp micro-deletions for the *RAG1* MGN. The RAG1 amplicons were also sequenced again in a different sequencing lane of the Illumina flowcell and a very similar distribution with a

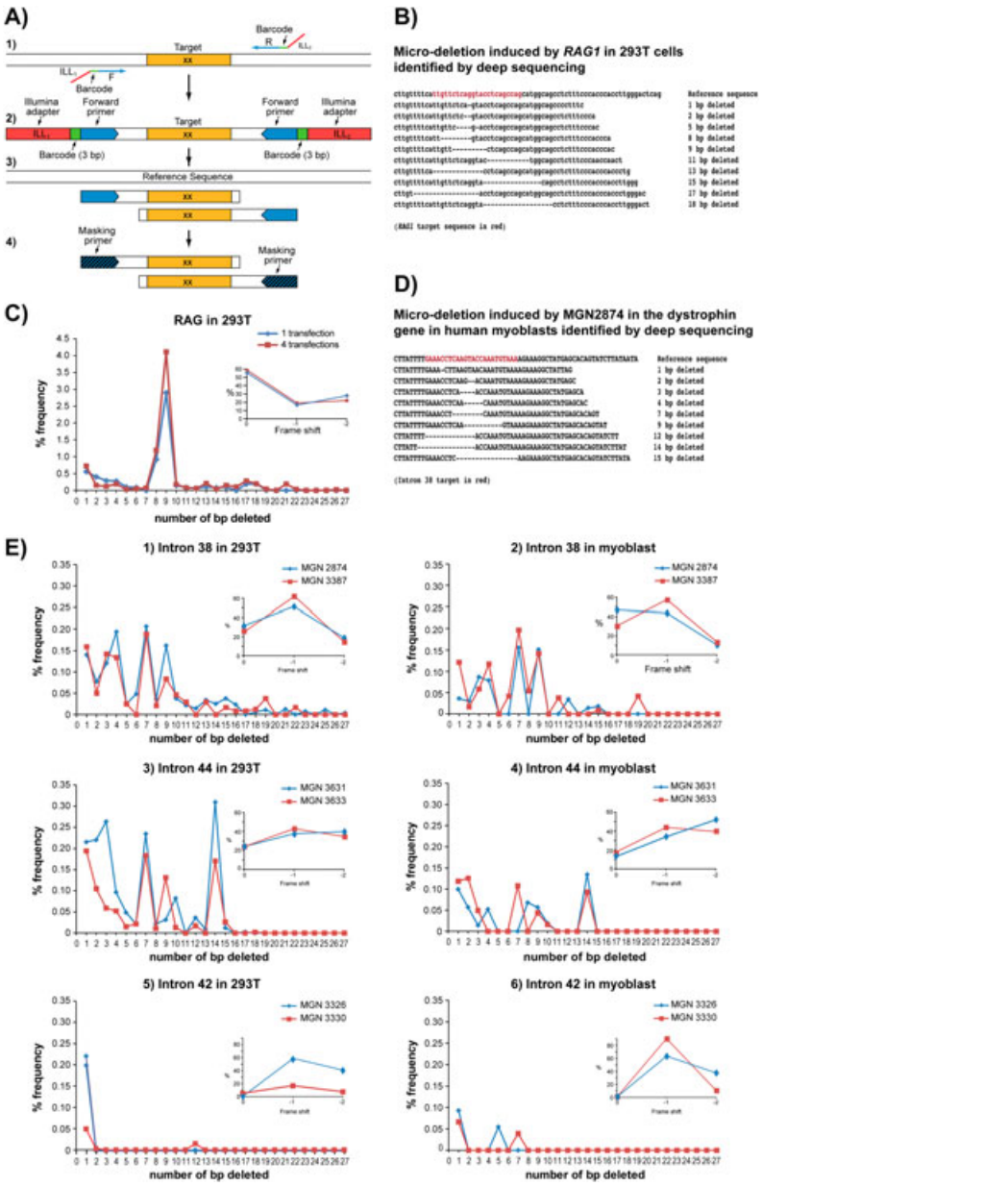

Figure 3. DBS procedure. (A) Schema of the DBS procedure: the region targeted by the MGN was amplified by PCR with primers containing the Illumina adapters for deep sequencing: a 3-bp barcode and a sequence that was complementary to the target gene. The Illumina sequences were separated according to the barcodes. Only the sequences in which the primer sequences were perfect were used to identify the INDELs. Only the sequences that were identical in the forward and in the reverse direction were used for the final identification of INDELs. (B) Identification by DBS of micro-deletions induced by MGNs targeting the RAG1 gene in 293 T cells: a few examples of the micro-deletions of various sizes in the RAG1 gene that were identified following the transfection of the *RAG1* MGN in 293 T cells. (C) Distribution of the sizes of micro-deletions induced by the *RAG1* MGN in 293 T cells: the *RAG1* MGN was transfected four times at 3-day intervals in 293 T cells. The DNA was extracted from samples of cells 3 days after the first and the fourth transfection. 15844 and 30948 RAG1 amplicons obtained after the first and the fourth transfection were analyzed by DBS for the presence of INDELs. The figure illustrates the percentage of amplicons in function of the size of the deletion. There was a clear peak in this distribution at 9 bp in both cases. The peak was, however, higher after four transfections (4.11%) than after one transfection (2.90%), indicating that repetitive administration of the same MGN progressively increases the frequency of INDELs. The insert shows that, as a result mostly of this peak at 9 bp, 55% of the micro-deletions would not have shifted reading frame of the resulting mRNA if MGN targeted a coding sequence, whereas 17% would have changed it by 1 bp and 28% by 2 bp. Similar results were obtained after four transfections. (D) Micro-deletions induced by MGNs targeting the dystrophin gene in human myoblasts identified by DBS: a few examples of the micro-deletions of various sizes in the dystrophin gene that were identified following the infection of human myoblasts with a lentivirus coding for the MGN2874. (E) Distribution of the sizes of INDELs identified in the dystrophin gene: 293 T cells were transfected ounce with one of the six different MGNs targeting the dystrophin gene. Human myoblasts were also infected once with the same MGNs. The regions of the dystrophin gene targeted by these MGNs were amplified and thousands of amplicons were sequenced using the Illumina platform. The figure illustrates the frequencies of the deletions of various sizes for each of these MGNs. There are peaks in these distributions, which are at the same positions for the two MGNs targeting the same dystrophin region. These peaks are in the same positions in 293 T cells and in myoblasts. However, MGNs targeting different dystrophin regions have peaks of different base pairs. The inserts indicate by how many base pairs the reading frame of an mRNA would be shifted if the MGNs targeted a coding region.

DOI: 10.1002/jgm

Table 3. Micro-deletions induced by MGNs targeting the dystrophin and the RAG1 genes in 293 T cells identified by DBS

	293 T			Myoblasts		Myotubes
	Control	RAG				
Deletions	0.00%	6.60%				
Insertions	0.00%	0.09%				
		DMD21 (Intron 38)		DMD21 (Intron 38)		
	Control	MGN 2874	MGN 3387	MGN 2874	MGN 3387	
Deletions	0.03%	1.25%	1.01%	0.61%	0.32%	
Insertions	0.00%	0.10%	0.02%	0.05%	0.00%	
		DMD31 (Intron 44)		DMD31 (Intron 44)		
	Control	MGN 3631	MGN 3633	MGN 3631	MGN 3633	
Deletions	0.00%	1.61%	1%	0.50%	0.55%	
Insertions	0%	0.16%	0.10%	0.00%	0.00%	
		DMD33 (Intron 42)		DMD33 (Intron 42)		
	Control	MGN 3326	MGN 3330	MGN 3326	MGN 3330	
Deletions	0.02%	0.20%	0.1%	0.15%	0.11%	
Insertions	0.02%	0.13%	0.04%	0.00%	0.01%	
	Control	ZFNs		control	ZFNs	ZFNs
Deletions	0.00%	6.27%		0.03%	0.69%	0.86%
Insertions	0.00%	3.10%		0.00%	0.22%	0.26%

The cells and the DNA extracts are from the same cells used in Figure 3. The endogenous dystrophin gene and the RAG1 gene in 293 T cells were mutated by MGNs. The regions of the dystrophin and the RAG1 genes targeted by the various MGNs were amplified by PCR and analyzed by DBS. Numbers of amplicons were analyzed: MGN2874: 28125 amplicons; MGN3387: 13850 amplicons; MGN3631: 21036 amplicons; MGN3633: 27514 amplicons; MGN3326: 17065 amplicons; MGN3330: 19261 amplicons; *RAG1* MGN; 10437 amplicons. As a negative control for the *RAG1* MGN, the targeted region of the RAG1 gene was amplified from cells transfected with the MGN2874. As negative controls for the MGNs targeting the dystrophin gene, the same dystrophin gene region was amplified from cells transfected with the *RAG1* MGN. No INDEL or a very low percentage of INDELs was detected in these negative controls. All the MGNs produced micro-deletions and micro-insertions that were more abundant than in the control cells. Thus, all the MGNs were active; however, the activity of MGN3326 and MGN3330 is very low.

peak at 9 bp was obtained. The DBS experiment was also repeated with 293 T cells that were transfected four times at 3-day intervals with the plasmid coding for the *RAG1* MGN. The percentage of INDELs was increased to 8.4%, indicating that repetitive administrations of the same MGN progressively increased the frequency of INDELs. The distribution of the size of the micro-deletions were similar, with a deletion peak at 9 bp (Figure 3C), which was higher after foru transfections (4.11%) than after one transfection (2.90%) These last results demonstrate that the sensitivity and precision of the DBS analysis of INDELs is better than those of the SCH. It is important to note that the large majority of INDELs produced by *RAG1* are micro-deletions and that micro-insertions were only 1.2–1.4% of the INDELs.

Experiments using six MGNs targeting the human dystrophin gene

Experiment in 293 T cells

Following the establishment of the DBS procedure with the *RAG1* MGN, we tested it with the six new MGNs targeting three dystrophin introns (Table 1) for which no INDELs were detected with the Surveyor procedure and only one INDEL was detected by the SCH procedure for MGN2874 and MGN3384. The detection of more INDELs by the SCH procedure would have required much more work. The initial experiments were carried out in 293 T cells that were transfected separately with different plasmids each coding for one of the MGNs targeting introns of the human dystrophin gene. Amplicons were obtained as described in the Materials and methods (for sequences of primers, see Table 2). Note that, for the two different MGNs targeting the same dystrophin introns, different barcodes were used to permit simultaneous sequencing in the same lane of the Illumina flowcell. This permitted sequencing of 49 bp of the dystrophin gene, which were between the primer sequences. INDELs were detected in the dystrophin gene of 293 T cells transfected with each one of the six MGN plasmids. When the 293 T cells were transfected with the *RAG1* MGN plasmid (i.e. an irrelevant MGN for the dystrophin gene), only very low percentages (0.00% to 0.04%) of INDELs were detected in the dystrophin introns 38, 42 and 44 (Table 3). The MGN3631 was the most active with 1.77% INDELs, while MGN3330 was the least active, with only 0.14% INDELs. It is important, however, to note that the frequency of INDELs measured for MGN3631 and MGN3633 targeting intron 44 is based only on a forward sequence of the amplicons. This is because there was a long series of thymidines after the 3′ end of the intron sequence targeted by these MGNs. This did not permit the design of primers to obtain the amplified sequence between 43 and 52 bp as for the other MGNs; the amplified sequences were 83 bp long and thus longer sequences would have been necessary to obtain overlapping forward and reverse sequences. Thus, our results for these two MGNs may contain some erroneous INDELs, especially those of only 1 bp. We also calculated the distribution of INDELs of various numbers of base pairs for these six MGNs targeting dystrophin in 293 T cells (Figures 3E1, E3 and E5). The distributions of the size of INDELs were not uniform, with peaks in different positions for these six MGNs. If the MGN3387 had targeting a coding sequence, 62% of the INDELs that it produced would have shifted their reading by −1 (Figure 3E1, insert). It is important to note that, as for the *RAG1* MGN, the

DOI: 10.1002/jgm

percentages of INDELs that were micro-insertions were lower than micro-deletions for MGNs targeting the dystrophin gene, being 7.4% for MGN2874, 2.0% for MGN3387, 9.3% for MGN3631, 8.7% for MGN3633, 41.7% for MGN3326 and 22.9% for MGN3330. The last two MGNs, however, had a very low total mutagenic activity.

Experiments in human myoblasts

Human myoblasts are difficult to transfect; therefore, six lentivirus vectors were thus constructed, each one containing the gene coding for one of the six MGNs targeting the dystrophin gene and a puromycin resistance gene. Myoblast cultures were infected with each of these lentiviruses, selected with puromycin during 48 h and propagated for 2 weeks. The expression of the MGN proteins by the infected myoblasts was confirmed by western blotting. Myoblasts infected with a lentiviral vector coding for EGFP were used as negative control. As for the experiment carried out with 293 T cells, the dystrophin regions targeted by the MGNs were analyzed by DBS. Low frequency INDELs were detected in myoblasts demonstrating the high sensitivity of this technique and indicating that these MGNs were able to reach and mutate the endogenous dystrophin gene in these primary human myoblast cultures. Examples of micro-deletions produced by MGN 2874 in human myoblasts are presented in Figure 3D and the results are summarized in Table 3. The MGN2874 was the most active in myoblasts. These results indicate that the Surveyor enzyme and the SCH procedures were not sufficiently sensitive to detect these INDELs. The sizes of the INDELs produced by the MGNs targeting dystrophin in myoblasts (Figures 3E2, E4 and E6) were also not uniformly distributed. The peaks in those distributions were frequently at the same positions for the two MGNs targeting the same dystrophin region. It is important to note that the peaks are at the same number of base pairs in both 293 T cells and in myoblasts. If the target sites would had been located in coding sequences, reading frame shifts would have been produced (Figure 3E, inserts). Again, the percentage of micro-insertions produced by the MGNs targeting the dystrophin gene were a low percentage of the total INDELs, being 2.9% for MGN2874, 0% for MGN3387, 0% for MGN3631, 0% for MGN3633, 0% for MGN3326 and 10.0% for MGN3330.

Mutations of the dystrophin gene with ZFNs targeting exon 50

The previous experiments were made with MGNs targeting dystrophin introns. However, to restore the reading frame of the dystrophin gene in DMD patients that have out-of-frame deletions by inducing INDELs, it is exons that have to be targeted. We thus obtained (from Compo ZR) a pair of ZFNs targeting the end of exon 50. Each of the ZFNs contained six fingers. The sequences targeted by this pair of ZFNs are: CTAGCTCCTGGACTGACC and GGAGCCTGTAAGTATACTG. We initially tested these ZFNs by transfecting the plasmids coding for them in 293 T cells. The DNA and the proteins were extracted from the cells 3 days later. The expression of the ZFNs was confirmed by western blotting (Figure 4A-1). The results of this experiment were first analyzed with the Surveyor enzyme method. Two additional bands were observed indicating that the ZFNs had mutated the exon 50 (Figure 4A-2). The results were then further analyzed with the DBS procedure and 9.4% of dystrophin exon 50 had an INDEL, whereas there were only 0.03% INDELs in the control cells not transfected with the ZFNs. Distributions of the sizes of these micro-deletions and micro-insertions are illustrated in Figures 4B-1 and 4B–2. It is important to note that the micro-insertions represent 43% of all INDELs.

The experiment was then repeated with human myoblasts. In this experiment, the control myoblasts were nucleofected with a plasmid coding for enhanced green fluorescent protein, whereas the experimental cells were co-nucleofected with the plasmids coding for both ZFNs. Again the DNA and the proteins were extracted 3 days later. The expression of the ZFNs was confirmed by western blotting (Figure 4A-3). However, no additional bands were detected with the Surveyor enzyme procedure, indicating that there was either no mutation induced by the ZFNs in human myoblasts or that the rate of mutation was low. However, with the DBS, we observed only 0.03% INDELs in the control myoblasts, whereas there were 0.92% INDELs in the cells that received the ZFNs. The distributions of the sizes of these INDELs are illustrated in Figures 4B-3 and 4B-4. To verify whether the expression of the dystrophin gene could increase the accessibility of the ZFNs to that gene to mutate it, a subpopulation of the myoblasts nucleofected with the ZFN plasmids in the previous experiments was placed at high density in a low serum culture medium to favour their fusion. Myotubes started to form after 5 days. The proteins and the DNA were extracted 7 days after the nucleofection. We observed that 1.11% of the exons 50 of dystrophin contained INDELs. The size distributions of these INDELs are illustrated in Figures 4B-5 and 4B-6. It is important to note that 23% of the INDELs were micro-insertions. It should be noted that 831 INDELs out of the 1916 detected in myotubes (i.e. 43% of INDELs) changed the reading frame by $1 + 3n$ bp, whereas 732 INDELs (38%) changed the reading frame by $2 + 3n$ bp.

Discussion

Comparison of the three techniques to detect INDELs

We have tested three different procedures to detect INDELs produced by endonucleases. The first procedure, the Surveyor nuclease treatment, uses the capacity of this enzyme to recognize and cleave all insertions, deletions and base substitutions that may be present in DNA

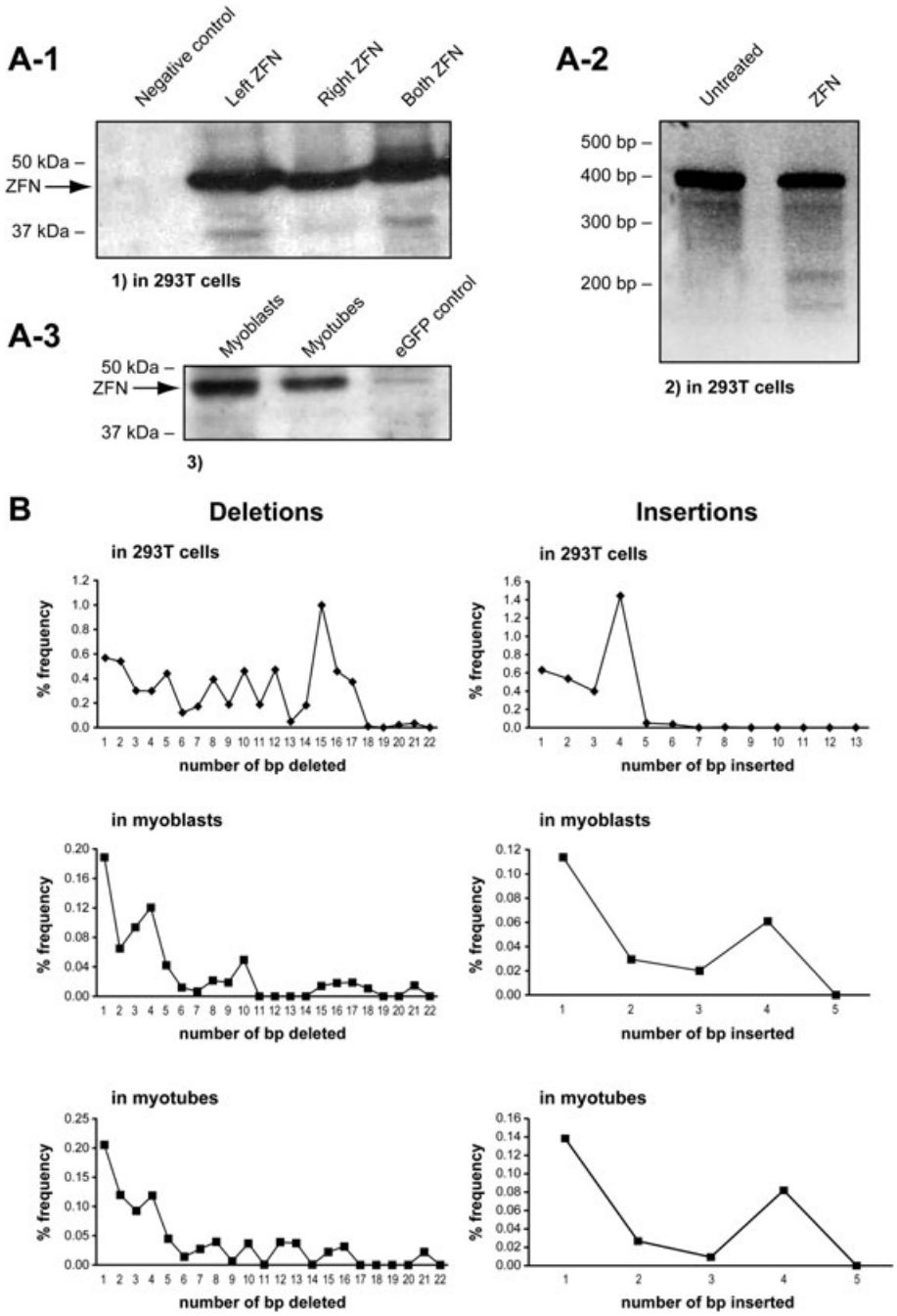

Figure 4. Results with ZFNs targeting exon 50 of dystrophin. (A-1) The ZFN proteins targeting dystrophin exon 50 were detected by western blotting in homogenates of 293 T cells transfected either with plasmids coding for the right and left ZFN alone or together. The band near 50 kDa was not present for proteins extracted from 293 T cells not transfected with a ZFN plasmid. (A-2) Two additional bands were detected following digestion with the Surveyor enzyme of the DNA extracted from 293 T cell transfected with both ZFNs. (A-3) Human myoblasts were nucleofected with the plasmids coding for the right or left ZFN or with an enhanced green fluorescent protein plasmid. The ZFN proteins were detected in human myoblasts and in the myotubes resulting from their fusion. (B) The size of the micro-deletions and the micro-insertions produced by the ZFNs in 293 T cells in myoblasts and in myotubes were not uniformly distributed.

hetero-duplexes. This technique is rapid but is not sensitive enough to detect rare INDELs. Indeed, previous studies have reported a range of 1% to 3% of mutations have to be present in the target sequence to be detected by the Surveyor enzyme [38–40]. These results are supported by our observations that we did not observe additional bands with the Surveyor enzyme in human myoblasts nucleofected with MGNs or ZFNs, whereas approximately 1% INDELs were detected by DBS. Moreover, the Surveyor enzyme procedure does not provide information on the types and the frequencies of various INDELs produced by endonucleases.

The second procedure, the SCH permitted identification of the sizes and the exact sequences of the INDELs induced by MGNs. The SCH procedure has several pitfalls in comparison to deep sequencing technology. The

DOI: 10.1002/jgm

sensitivity was affected by numerous manipulations required for the five steps. Amplification of genomic DNA by PCR, ligation of the PCR products in TA vector pDrive followed by transformation in the bacteria, colony lift on nylon membranes, labelling of the oligoprobes, hybridization (Tm) conditions of the radioactive probe with membrane, washings of the membrane and autoradiography (time exposure film) are all factors that influenced the sensitivity. However, even if these factors could affect the sensitivity, we were able to evaluate INDELs corresponding to the deletion of one to several base pairs. However, this procedure was very labour intensive and it would had been necessary to sequence much more additional clones to obtain a better representation of the characteristics and frequencies of INDELs induced by the different MGNs, especially those targeting the dystrophin gene. As a result of the amount of work required by this technique, it was given up in favour of the DBS, which was explored in parallel and which produced more rapidly abundant results. Although the SCH technology involves more work than the deep sequencing technology, it has two advantages: (i) it is accessible to laboratories that do not have access to the deep sequencing technology and (ii) it provides a quick turnover of results. The most sensitive procedure and the one that permitted the best characterization of these INDELs was by far the DBS procedure, which allowed us to rapidly quantify and characterize the INDELs. Indeed, simply increasing the total number of amplicons, which are sequenced, can further increase the sensitivity of this procedure. However, the main problem that we had with the deep sequencing is that, because we were buying only one line out of the eight present in the Illumina flowcell, we frequently had to wait several weeks to have the other lines bought by another laboratory requesting the same number of base pairs sequencing and having a spare line or accepting that we spike our samples in their sample. Indeed, the deep sequencing technology is usually used by laboratories performing large-scale genomic studies.

Endonucleases are able to mutate the endogenous dystrophin gene

Our initial experiments with MGNs were carried out in a model system using plasmids carrying the target gene sequence. This made these initial experiments for screening for activity easier. However, the endogenous dystrophin gene has to be targeted with endonucleases to show a potential usefulness of this therapeutic approach.

It was important to establish the frequency of mutations in the target gene, which have not been exposed to an endonuclease. When we did such experiments in the 293 T cell and in human myoblasts, we observed no INDEL in the RAG1 gene but approximately 0.03% INDELs were detected in different introns and in exon 50 of the dystrophin gene. These INDELs are not the result of a sequencing error of the Illumina Genome Analyser IIx because they were identical in the forward and in the reverse sequences. Most of the INDELs in the control cells not treated with a relevant endonuclease were only 1 bp in length. It is possible that some of them are the results of mutation introduced by the PCR amplification. It is however, important the emphasize that four of the MGNs (2874, 3387, 3631 and 3633) targeting the dystrophin gene produced INDELs that were more than ten-fold more abundant than the INDELs detected in the control cells. Moreover, the ZFNs increased the frequency of INDELs by 30-fold in human myoblasts and by more than 300-fold in 293 T cells. Thus, these endonucleases are able to induce mutations in the endogenous dystrophin gene. However, MGNs 3326 and 3330 had, at best, only a low level of activity in both 293 T and in myoblasts.

Reproducibility of the results obtained by DBS

The DBS results are very reproducible because the same frequency of INDELs and a similar distribution of INDEL sizes were obtained when amplicons were sequenced in different sequencing lanes.

Frequency of INDELs produced by endonucleases

Experiments were initially made in 293 T cells because these human cells are easy to transfect with Lipofectamine 2000™, whereas human myoblasts requested nucleofection with plasmids or infection with a lentivirus vector. Subsequent experiments were however, made with human myoblasts because these cells could eventually be genetically corrected with the endonucleases and transplanted back to the same DMD patient to restore the expression of dystrophin [41–43]. The frequencies of INDELs produced by the endonucleases, especially in the dystrophin gene, were always lower in myoblasts than in 293 T cells because the dystrophin gene is not expressed in both types of cells; the difference may be a result only of the more effective transfection of the 293 T cells. However, when a gene is not expressed in a cells, its DNA is in a compact form (i.e. heterochromatin) or methylated and thus less accessible to the endonucleases. The dystrophin DNA may thus be more susceptible to mutations by endonucleases in myotubes or in muscle fibres because the dystrophin gene is expressed in these cells. However, the frequency of INDELs was only slightly increased in the experiments in which we allowed the myoblasts to form myotubes. It should be noted that the myotubes were formed only after the introduction of the endonuclease gene. It may be possible to increase the frequency of INDELs by carrying out experiments at 30 °C, as recently reported for ZFNs [44], or by treating the cells with histone deacetylase inhibitors, which increase the expression of some gene by uncoiling the DNA [45]. Increasing the endonuclease expression level by using a better transfection or transduction procedure may also increase the

INDELs. Moreover, because the genes, that have been mutated by a given endonuclease are no longer susceptible to the actions of that endonuclease, whereas the wild-type genes can still be mutated, it is possible to progressively increase the frequency of INDELs in the target gene by repeated administration of the endonuclease. Indeed, the higher frequency of INDELs in the 293 T cells transfected four times with the *RAG1* MGN sustains this hypothesis (Figure 3C). The DBS procedure is sensitive enough to be used to detect rare INDELs produced *in vivo* by endonucleases and to quantify the increased frequency of these INDELs that would be obtained with repeated endonuclease administration. It is very important to note that, although the frequency of INDELs produced by the endonucleases targeting dystrophin is low less than 2%, this is not an obstacle to an eventual use of specific endonucleases for treating DMD because the frequency of gene correction could be increased by re-administrating the endonucleases several times until a satisfactory level of gene correction has been reached.

Distribution of INDEL sizes and restoration of the normal dystrophin reading frame

The most novel and surprising aspect of our results is the distribution of the size of micro-deletions induced by the endonucleases. This phenomenon has not been previously reported and it can be observed only because large numbers of micro-deletions were identified by DBS. These peaks are not randomly distributed because they are present at the same position in both 293 T cells and myoblasts and at similar positions for MGNs targeting the same dystrophin nucleotide sequence. When we started the present study, the accepted concept was that the INDELs induced by NHEJ is a random process and thus their sizes was expected to be uniformly distributed. It was thus unexpected to observe that 45% of the micro-deletions induced by *RAG1* MGN in 293 T cells were of 9 bp (this increased to 50% in the cells transfected four times with this MGN). Indeed, 55% of the INDELs produced by the *RAG1* MGN if they were in the coding region of the gene would not have changed the reading frame, whereas only 17% would have shifted the reading frame by 1 bp and 28% by 2 bp. We were expecting that there would be an equal distribution of the reading frame shift at approximately 33% each. We were initially disappointed by this result because this was not a confirmation of our hypothesis and thus most of the INDELs produced by the *RAG1* MGN would not have changed the reading frame of a gene with an out-of-frame deletion. On the other hand, an endonuclease that produces most of its micro-deletions without changing the reading frame would be excellent to delete a nonsense codon and thus could be very useful for correcting a genetic disease with this type of mutation. At the present time, we do not have data from a sufficient number of different endonucleases to understand why the *RAG1* MGN preferentially delete 9 bp. This may be a result of the structure of the MGN, of the targeted sequence or of the sequences surrounding it. The same reasons may also explain why the ZFNs are producing much more micro-insertions than MGNs.

The results obtained with the MGNs and the ZFNs targeting the dystrophin gene were quite different than those obtained with the *RAG1* MGN. There were indeed multiple peaks in the distributions of the size of the INDELs. The positions of the peaks were similar for MGNs in the 293 T cells and in the myoblasts but were different for the ZFNs targeting dystrophin exon 50. Thus, the mechanisms controlling the size of INDELs are not well understood.

It is interesting to note that approximately 44% of the micro-deletions produced by the ZFNs in exon 50 in myoblasts or in myotubes changed the reading frame by $1 + 3n$ bp. This type of micro-deletion would have restored to correct dystrophin reading frame in DMD patients with deletions of exons 51, 51–53 or 51–60. However, approximately 36% of the micro-deletions produced by the same ZFNs in the same cells changed the reading frame by $2 + 3n$ bp. This would have permitted restoration of the normal dystrophin reading frame in DMD patients with a deletion of exons 51–56.

The restoration of the dystrophin reading frame in 1% of the nuclei is enough to correct the disease

It is important to note that although many INDELs produced by MGNs or ZFNs would not restore the normal reading frame in the DMD patient with an out-of-frame deletion; this is not a problem. Indeed, before treatment with an endonuclease, the dystrophin gene did not produce an in-frame dystrophin mRNA; thus, for the dystrophin gene in which the reading frame is not corrected by the activity of the endonuclease, the situation is not worse than before. However, the correction of the reading frame of some dystrophin genes permits the production of some dystrophin protein with normal N and C terminals. To correct DMD, it may be sufficient to have only a low percentage of the nuclei inside the muscle fibre that are able to produce dystrophin. Indeed, we have previously show, by our myoblast transplantation experiments, that the presence of only one competent nucleus in a muscle fibre permitted the expression of dystrophin over 400 μm of the muscle fibre, although this portion of the muscle fibres contained more than one hundred nuclei, most of them unable to produce a dystrophin with a C terminal [46].

Regions of exons to be targeted by endonucleases to correct a frame shift deletion

To restore the reading frame of the dystrophin gene, it is important to target specific region at the end of the exon that precedes the patient deletion or at the beginning of the exon that follows the exon deletion. The sequence

DOI: 10.1002/jgm

that can be targeted in the exon that precedes or follows the patient deletion depends on the type of frame shift (i.e. –1 or a −2 bp) induced by the patient deletion. When targeting the exon that precedes the patient deletion, it is important to target a sequence of this preceding exon, which is such that, if an INDEL restoring the reading frame is produced, this INDEL does not generate a stop codon within this preceding exon. When targeting the exon, which follows the patient deletion, the sequence to be targeted by the endonuclease has to be located before the first stop codon resulting from the frame shift induced by the patient deletion. Thus, depending on the patient deletion, a more or less large fraction of the preceding or following exon can be targeted. Thus, as a general rule, the end of the preceding exon or the beginning of the following exon has to be targeted. The sequences to be targeted for different types of dystrophin deletion are provided in the Supporting information (Table S2).

Deleting nonsense codons with endonucleases

Approximately one-third of DMD patients have a point mutation that results in a nonsense codon and thus in the production of a truncated dystrophin protein, which cannot integrate in the dystrophin complex. Specifically engineered endonucleases could be used to treat such patients. Indeed, endonucleases can induce micro-deletions, which are multiples of 3 bp in size. Thus, by targeting a sequence of base pairs that includes or is near a nonsense codon mutation with endonucleases, it may be possible to delete this nonsense codon but maintain the reading frame of the resulting mRNA. The resulting protein would be internally truncated by a few amino acids, although it would probably be partially functioning in most cases. Up to recently, it would have been difficult to engineer endonucleases (MGNs or ZFNs) that could target such a precise sequence. Recently, a new type of endonucleases called TALEN has been discovered [47,48]. These TALEN can be more easily modified to target specific sequence [13,15]. The only restriction for a sequence to be targeted by a TALEN is that it has to start by a thymidine. Thus, it should be possible to engineer TALENs that target sequences overlapping with every nonsense codon. It is important to note that the nonsense codon to be corrected does not have to be at the beginning or at the end of an exon, and may be anywhere within the exon.

Conclusions and future directions

The results obtained in the present study clearly demonstrate that endonucleases targeting different regions of a gene can induce INDELs. Thus, eventually, by selecting the right endonuclease targeting the right sequence, it could be possible to induce either micro-deletions including a nonsense codon at the same time as maintaining the normal reading frame or INDELs, which restore the reading frame in a gene with a frame shift mutation. The great advantage of correcting a mutated gene with an endonuclease by inducing a NHEJ is that this process does not require the insertion in the cells of a donor sequence, as is the case for the correction of DSBs by homologous recombination. The absence of the requirement for a donor sequence makes it possible to correct a gene by introducing only an endonuclease protein rather than by introducing the endonuclease gene in the cells. It is possible to achieve such a protein transduction by fusing the endonuclease protein with a protein transduction domain [49,50]. Our laboratory is actively pursuing this avenue of research.

Acknowledgements

This work was supported by a grant from Association Française contre les Myopathies. S.B. is recipient of a doctoral award from the Canadian Institutes of Health Research (200910GSD-226209-172830).

References

1. Hoffman EP, Brown RH Jr, Kunkel LM. Dystrophin: the protein product of the Duchenne muscular dystrophy locus. *Cell* 1987; **51**: 919–928.
2. Trimarco A, Torella A, Piluso G, *et al*. Log-PCR: a new tool for immediate and cost-effective diagnosis of up to 85% of dystrophin gene mutations. *Clin Chem* 2008; **54**: 973–981.
3. England SB, Nicholson LV, Johnson MA, *et al*. Very mild muscular dystrophy associated with the deletion of 46% of dystrophin. *Nature* 1990; **343**: 180–182.
4. Stoddard BL. Homing endonuclease structure and function. *Q Rev Biophys* 2005; **38**: 49–95.
5. Durai S, Mani M, Kandavelou K, *et al*. Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Res* 2005; **33**: 5978–5990.
6. Lloyd A, Plaisier CL, Carroll D, *et al*. Targeted mutagenesis using zinc-finger nucleases in *Arabidopsis*. *Proc Natl Acad Sci USA* 2005; **102**: 2232–2237.
7. Mani M, Kandavelou K, Dy FJ, *et al*. Design, engineering, and characterization of zinc finger nucleases. *Biochem Biophys Res Commun* 2005; **335**: 447–457.
8. Porteus MH, Carroll D. Gene targeting using zinc finger nucleases. *Nat Biotechnol* 2005; **23**: 967–973.
9. Lombardo A, Genovese P, Beausejour CM, *et al*. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nature Biotechnol* 2007; **25**: 1298–1306.
10. Cathomen T, Joung JK. Zinc-finger nucleases: the next generation emerges. *Mol Ther* 2008; **16**: 1200–1207.
11. Maeder ML, Thibodeau-Beganny S, Osiak A, *et al*. Rapid 'open-source' engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell* 2008; **31**: 294–301.
12. Hockemeyer D, Soldner F, Beard C, *et al*. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol* 2009; **27**: 851–857.
13. Cermak T, Doyle EL, Christian M, *et al*. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* 2011; **39**: e82.
14. Christian M, Cermak T, Doyle EL, *et al*. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 2010a; **186**: 757–761.
15. Li T, Huang S, Zhao X, *et al*. Modularly assembled designer TAL effector

nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res* 2011; **39**: 6315–6325.
16. Mahfouz MM, Li L, Shamimuzzaman M, *et al*. De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc Natl Acad Sci USA* 2011; **108**: 2623–2628.
17. Miller JC, Tan S, Qiao G, *et al*. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* 2010; **29**: 143–148.
18. Morbitzer R, Elsaesser J, Hausner J, *et al*. Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res* 2011; **39**: 5790–5799.
19. Choulika A, Perrin A, Dujon B, *et al*. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*. *Mol Cell Biol* 1995; **15**: 1968–1973.
20. Rouet P, Smih F, Jasin M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol* 1994; **14**: 8096–8106.
21. Paques F, Haber JE. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* 1999; **63**: 349–404.
22. Beumer K, Bhattacharyya G, Bibikova M, *et al*. Efficient gene targeting in Drosophila with zinc-finger nucleases. *Genetics* 2006; **172**: 2391–2403.
23. Bibikova M, Golic M, Golic KG, *et al*. Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics* 2002; **161**: 1169–1175.
24. Doyon JB, Pattanayak V, Meyer CB, *et al*. Directed evolution and substrate specificity profile of homing endonuclease I-SceI. *J Am Chem Soc* 2006; **128**: 2477–2484.
25. Ashworth J, Havranek JJ, Duarte CM, *et al*. Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature* 2006; **441**: 656–659.
26. Arnould S, Chames P, Perez C, *et al*. Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J Mol Biol* 2006; **355**: 443–458.
27. Smith J, Grizot S, Arnould S, *et al*. A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res* 2006; **34**: e149.
28. Miller JC, Holmes MC, Wang J, *et al*. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol* 2007; **25**: 778–785.
29. Foley JE, Yeh JR, Maeder ML, *et al*. Rapid mutation of endogenous zebrafish genes using zinc finger nucleases made by Oligomerized Pool ENgineering (OPEN). *PLoS One* 2009; **4**: e4348.
30. Perez EE, Wang J, Miller JC, *et al*. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nature Biotechnol* 2008; **26**: 808–816.
31. Meng X, Noyes MB, Zhu LJ, *et al*. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol* 2008; **26**: 695–701.
32. Grizot S, Smith J, Daboussi F, *et al*. Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. *Nucleic Acids Res* 2009; **37**: 5405–5419.
33. Chapdelaine P, Pichavant C, Rousseau J, *et al*. Meganucleases can restore the reading frame of a mutated dystrophin. *Gene Ther* 2010; **17**: 846–858.
34. Chapdelaine P, Delahaye S, Gauthier E, *et al*. A one-hour procedure for the preparation of genomic DNA from frozen tissues. *Biotechniques* 1993; **14**: 163–164.
35. Bentley DR, Balasubramanian S, Swerdlow HP, *et al*. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008; **456**: 53–59.
36. Craig DW, Pearson JV, Szelinger S, *et al*. Identification of genetic variants using bar-coded multiplexed sequencing. *Nat Methods* 2008; **5**: 887–893.
37. Altschul SF, Madden TL, Schaffer AA, *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; **25**: 3389–3402.
38. Guschin DY, Waite AJ, Katibah GE, *et al*. A rapid and general assay for monitoring endogenous gene modification. *Meth Mol Biol* 2010; **649**: 247–256.
39. Jiang Y, Palma JF, Agus DB, *et al*. Detection of androgen receptor mutations in circulating tumor cells in castration-resistant prostate cancer. *Clin Chem* 2010; **56**: 1492–1495.
40. Pilch J, Asman M, Jamroz E, *et al*. Surveyor nuclease detection of mutations and polymorphisms of mtDNA in children. *Pediatr Neurol* 2010; **43**: 325–330.
41. Skuk D, Goulet M, Roy B, *et al*. Dystrophin expression in muscles of duchenne muscular dystrophy patients after high-density injections of normal myogenic cells. *J Neuropathol Exp Neurol* 2006; **65**: 371–386.
42. Quenneville SP, Chapdelaine P, Skuk D, *et al*. Autologous transplantation of muscle precursor cells modified with a lentivirus for muscular dystrophy: human cells and primate models. *Mol Ther* 2007; **15**: 431–438.
43. Skuk D, Goulet M, Roy B, *et al*. First test of a 'high-density injection' protocol for myogenic cell transplantation throughout large volumes of muscles in a Duchenne muscular dystrophy patient: 18 months follow-up. *Neuromuscul Disord* 2007; **17**: 38–46.
44. Doyon Y, Choi VM, Xia DF, *et al*. Transient cold shock enhances zinc-finger nuclease-mediated gene disruption. *Nat Methods* 2010; **7**: 459–460.
45. Thiagalingam S, Cheng KH, Lee HJ, *et al*. Histone deacetylases: unique players in shaping the epigenetic histone code. *Ann NY Acad Sci* 2003; **983**: 84–100.
46. Kinoshita I, Vilquin JT, Asselin I, *et al*. Transplantation of myoblasts from a transgenic mouse overexpressing dystrophin produced only a relatively small increase of dystrophin-positive membrane. *Muscle Nerve* 1998; **21**: 91–103.
47. Boch J, Scholze H, Schornack S, *et al*. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 2009; **326**: 1509–1512.
48. Christian M, Cermak T, Doyle EL, *et al*. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 2010b; **186**: 757–761.
49. Tilstra J, Rehman KK, Hennon T, *et al*. Protein transduction: identification, characterization and optimization. *Biochem Soc Trans* 2007; **35**: 811–815.
50. Chauhan A, Tikoo A, Kapur AK, *et al*. The taming of the cell penetrating domain of the HIV Tat: myths and realities. *J Control Release* 2007; **117**: 148–162.