TALEN-Induced Double-Strand Break Repair of CTG Trinucleotide Repeats

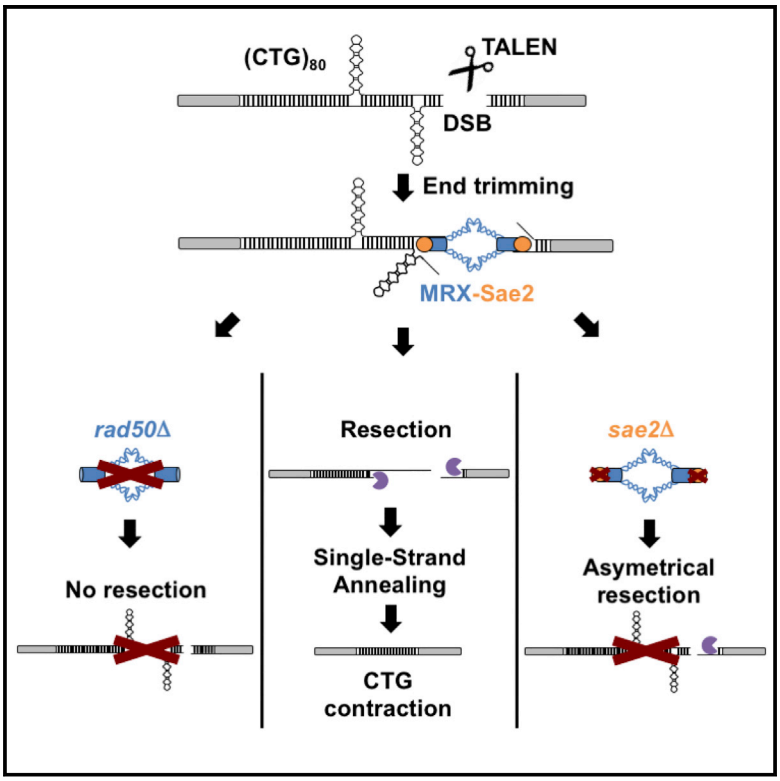
Valentine Mosbach, Lucie Poggi, David Viterbo, Marine Charpentier, Guy-Franck Richard

Article

Cell Reports

TALEN-Induced Double-Strand Break Repair of CTG Trinucleotide Repeats

Graphical Abstract



Authors

Valentine Mosbach, Lucie Poggi, David Viterbo, Marine Charpentier, Guy-Franck Richard

Correspondence

gfrichar@pasteur.fr

In Brief

Mosbach et al. show that a TALEN could successfully contract an expanded CTG repeat tract below pathological length. The TALEN double-strand break needs Rad50, Rad52, and Sae2 to be repaired. A double-strand break end containing a long trinucleotide repeat needs both Rad50 and Sae2 to be efficiently resected.

Highlights

- *RAD50*, *SAE2*, and *RAD52* are involved in repairing a DSB within a CTG repeat
- *POL32*, *DNL4*, and *RAD51* do not play a role in repairing a DSB within a CTG repeat
- Resection of a DSB within a CTG repeat needs the Mre11-Rad50 complex, as well as Sae2
- The double-strand break is repaired by an iterative single-strand annealing process

Mosbach et al., 2018, Cell Reports *22*, 2146–2159
February 20, 2018
https://doi.org/10.1016/j.celrep.2018.01.083

CellPress

OPEN ACCESS
CellPress

Cell Reports
Article

TALEN-Induced Double-Strand Break Repair of CTG Trinucleotide Repeats

Valentine Mosbach,[1,2,3] Lucie Poggi,[1,2,3] David Viterbo,[1,3] Marine Charpentier,[4] and Guy-Franck Richard[1,3,5,*]
[1]Institut Pasteur, Département Génomes & Génétique, 25 rue du Dr Roux, 75015 Paris, France
[2]Sorbonne Universités, UPMC Univ Paris 06, IFD, 4 Place Jussieu, 75252 Paris Cedex 05, France
[3]CNRS, UMR3525, 75015 Paris, France
[4]INSERM U1154, CNRS UMR7196, Museum National d'Histoire Naturelle, 75005 Paris, France
[5]Lead Contact
*Correspondence: gfrichar@pasteur.fr
https://doi.org/10.1016/j.celrep.2018.01.083

SUMMARY

Trinucleotide repeat expansions involving CTG/CAG triplets are responsible for several neurodegenerative disorders, including myotonic dystrophy and Huntington's disease. Because expansions trigger the disease, contracting repeat length could be a possible approach to gene therapy for these disorders. Here, we show that a TALEN-induced double-strand break was very efficient at contracting expanded CTG repeats in yeast. We show that *RAD51*, *POL32*, and *DNL4* are dispensable for double-strand break repair within CTG repeats, the only required genes being *RAD50*, *SAE2*, and *RAD52*. Resection was totally abolished in the absence of *RAD50* on both sides of the break, whereas it was reduced in a *sae2*Δ mutant on the side of the break containing the longest repeat tract, suggesting that secondary structures at double-strand break ends must be removed by the Mre11-Rad50 complex and Sae2. Following the TALEN double-strand break, single-strand annealing occurred between both sides of the repeat tract, leading to repeat contraction.

INTRODUCTION

Microsatellite expansions are responsible for more than two dozen neurological or developmental disorders in humans. Among the most common sequences involved are CAG/CTG trinucleotide repeat tracts, whose expansions are the cause of Huntington's disease, myotonic dystrophy type 1 (DM1 or Steinert disease), and several spinocerebellar ataxias (Orr and Zoghbi, 2007). Despite having been under investigation for more than two decades, the molecular mechanism(s) leading to large expansions is not completely understood, although it is generally accepted that secondary structures formed by these microsatellites may be triggering or amplifying the expansion process (McMurray, 1999). It was shown that CAG and CTG trinucleotide repeats form imperfect hairpins *in vitro* (Gacy et al., 1995; Yu et al., 1995a, 1995b). In addition, there is biochemical and genetical evidence that CAG and CTG hairpins interfere with the mismatch repair machinery, an important player of the expansion process, although its precise role is not totally clear (Foiry et al., 2006; Manley et al., 1999; Owen et al., 2005; Pearson et al., 1997; Pinto et al., 2013; Savouret et al., 2004; Slean et al., 2016; Tian et al., 2009; Tomé et al., 2009, 2013; Viterbo et al., 2016; Williams and Surtees, 2015). Most trinucleotide repeat transmissions from parents to children lead to repeat tract expansion. However, it seldom happens that a large allele contracts to a shorter one. Indeed, in a family affected by DM 1, it was reported that a daughter inherited a contracted allele from her father by a mechanism likely to be gene conversion (O'Hoy et al., 1993). The daughter was followed until the age of 17 and did not develop any of the symptoms of the pathology, showing that a large repeat contraction prevented this kind of disease.

Recent attempts were made to cure trinucleotide repeat disorders by gene therapy. In Huntington's disease pluripotent stem cells, an expanded CAG repeat in the Huntington's disease (HD) gene was replaced by a smaller allele by homologous recombination. In corrected cells, HD disease phenotypes were reversed (An et al., 2012). Two independent groups used SpCas9 either to induce one single double-strand break upstream of the *FMR1* CGG repeat (Park et al., 2015) or two double-strand breaks (DSBs) upstream and downstream of the repeat tract (Xie et al., 2016). In both cases, *FMR1* reactivation was observed in edited cells. More recently, SpCas9 was used to delete the expanded CTG triplet repeat at the DM1 locus by making a DSB upstream and/or downstream of the repeat tract. Again, disease phenotypes were partially suppressed in DM1 myoblasts (van Agtmaal et al., 2017). In all of these cases, DSBs were always induced outside and never inside the trinucleotide repeat tract.

DSB repair is one of the molecular processes leading to trinucleotide repeat contractions and expansions. It was formerly shown that a DSB made by the I-*Sce* I meganuclease within a short CTG repeat tract often led to the loss of the nuclease recognition site and contraction of the repeat tract (Richard et al., 1999). In less frequent cases, it led to both expansions and contractions of the repeat tract by gene conversion during mitosis (Richard et al., 1999, 2000) or meiosis (Richard et al., 2003). Following these early experiments, zinc-finger nucleases (ZFNs) were used to direct a DSB within a CAG or CTG trinucleotide repeat tract. In two separate studies from the same lab, induction of a ZFN in Chinese hamster ovary (CHO) cells led to a 15-fold increase in repeat contractions. However, deletions in

one or both flanking regions were observed in 20% of the cases, whereas insertions of exogenous DNA at the DSB site were found in another 24% of the cases (Mittelman et al., 2009; Santillan et al., 2014). Different authors used another ZFN expressed in HeLa cells containing CAG/CTG trinucleotide repeats integrated in the two possible orientations compared with replication fork progression. They observed contractions as well as expansions of the repeat tract when both ZFN arms were expressed, but only contractions were recovered when one single arm was expressed (Liu et al., 2010). This suggested that one arm of the ZFN was able to homodimerize and induce a DSB by itself. Using a *Saccharomyces cerevisiae* strain in which a large CTG triplet repeat from a DM1 patient was integrated in a yeast chromosome, we were recently able to show that induction of a transcription activator-like effector (TALE) nuclease (TALEN) induced contractions of a CTG triplet repeat tract at a high frequency. Pulse-field gel electrophoresis and genome-wide deep sequencing showed that no other mutation, duplication, or chromosomal rearrangement was induced by the TALEN outside of the repeat tract (Richard et al., 2014). These experiments demonstrated that this new family of nucleases was efficient and specific enough to envision their possible use as a future gene therapy tool in human cells (Richard, 2015). Using a different approach, Cinesi et al. (2016) recently showed that inducing single-strand breaks within a CTG repeat tract using the Cas9D10A mutant nickase also promoted contractions of the repeat tract in model human cells.

Mechanisms of DSB repair have been studied in yeast for several decades, and the main proteins involved in this process have been identified (Krogh and Symington, 2004). A large part of these advances was made possible by the use of highly specific meganucleases such as HO or I-*Sce* I (Fairhead and Dujon, 1993; Haber, 1995; Plessis et al., 1992). However, the fate of a single DSB made within a long repeated and structured DNA sequence was never addressed before.

One of the goals of the present work was to study the role of several recombination genes (namely, *RAD50*, *RAD51*, *RAD52*, *DNL4*, *SAE2*, and *POL32*) in the repair of a single DSB made within a long CTG trinucleotide repeat. *RAD52* encodes a mediator multimeric protein controlling homologous recombination pathways (gene conversion, single-strand annealing [SSA], and break-induced replication [BIR]) (Davis and Symington, 2004; Krogh and Symington, 2004; Sugawara and Haber, 1992). *RAD51* is a RecA homolog responsible for nucleofilament formation and subsequent strand exchange and gene conversion (Shinohara et al., 1992; Sung, 1994). *RAD50* belongs to the multifunctional Mre11-Rad50-Xrs2 complex involved, along with Sae2, in DSB end clipping and resection during meiosis as well as mitosis (Borde et al., 2004; Lee et al., 1998; Mimitou and Symington, 2008; Zhu et al., 2008). *DNL4* encodes ligase IV, the protein responsible for the ligation step during non-homologous end joining (Wilson et al., 1997), and *POL32* is part of the polymerase δ complex and was shown to be essential for BIR (Lydeard et al., 2007) as well as to be an important player in microhomology-mediated repair (Villarreal et al., 2012).

RAD50 was found to be essential to resect both DSB ends, whereas *SAE2* was needed to resect only the DSB end that contains most of the triplet repeat tract. This observation supports the presence of secondary structures that need a functional Sae2 activity to be removed. *RAD52* was also required to repair the DSB but not *RAD51*, *POL32*, or *DNL4*, suggesting an iterative SSA process that progressively leads to repeat shortening.

RESULTS

A DSB Induced within CTG Repeats Requires the Mre11-Rad50 Complex to be Processed

In the present work, two TALENs were used. The $TALEN_{CTG}$ was the same as the one used in our former publication, designed to induce a DSB within a modified *SUP4* allele containing expanded CTG triplets from the dystrophia myotonica protein kinase (DMPK) human locus (Richard et al., 2014). The $TALEN_{noCTG}$ was designed to induce a DSB within a modified *SUP4* allele containing an I-*Sce* I recognition site (Richard et al., 1999). The trinucleotide repeat tract lengths used here ranged from 20–50 triplets for short alleles to 70–90 triplets for long alleles. In a first series of experiments, the $TALEN_{CTG}$ was expressed in wild-type yeast cells and in isogenic strains mutated for DSB repair genes. Both $TALEN_{CTG}$ arms were carried on centromeric vectors, and their expression was under control of an inducible TetOFF promoter. DSB formation was followed during a time course by Southern blot analysis. When the $TALEN_{CTG}$ was repressed, uncut chromosomes containing CTG repeats of two different lengths were visible. When the $TALEN_{CTG}$ was expressed, signals corresponding to DSB formation were detected (Figure 1A). By using probes specific to each side of the repeat tract, it was possible to distinguish between signals corresponding to 5′ or 3′ ends of the DSB (Figure S1). The 5′ end of the break, containing the long CTG tract appears like a smear. This smear corresponds to different repeat lengths because of progressive CTG repeat contraction over time. The 3′ end of the DSB appears as a sharper band because it contains only a few triplets. The DSB was not visible before 14 hr after $TALEN_{CTG}$ induction (time point labeled 0, Figure 1A). Quantification showed that the maximum of broken molecules was reached 4–6 hr after T0 for all strains except *rad50*Δ and *sae2*Δ (Figure 1B).

To determine how long it would take for cells to completely repair the DSB, a longer time course was run in wild-type cells over 72 hr after TALEN induction (58 hr after DSB formation; Figure 2). This experiment was set up in haploid cells to discriminate between the parental (uncontracted) allele and the contracted allele recovered after DSB repair. Cells were collected at several time points, with particular attention to the 34–46 hr time range. Total genomic DNA was extracted, and Southern blot was run as described previously (Figure 2A). Signal quantification showed that, during the first 40 hr in which ~12% of chromosomes were broken, the DSB signal stayed stable. After that time, it increased to ~20% of broken molecules and stayed at the same level until the end of the time course (Figure 2B). This result may be interpreted in two ways: (1) the nuclease was not active in all cells at the same time. Therefore, only a subfraction of repeat tracts was cleaved in the first 40 hr and another, larger, subfraction was cleaved later on. (2) A first burst of DSBs partially contracted repeat tracts in all cells. A second round of DSBs cleaved shortened repeats more efficiently because they

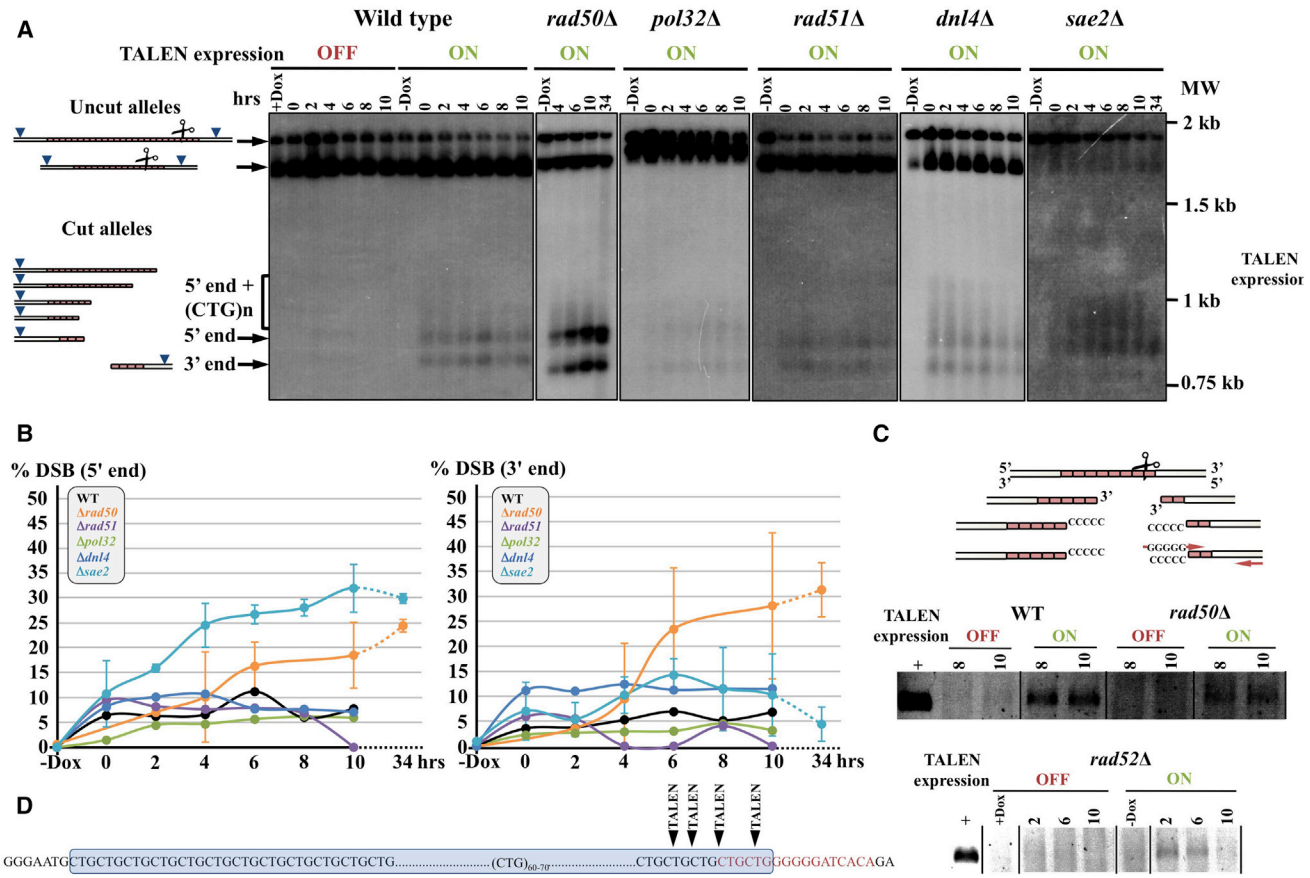


Figure 1. DSB Induction by a TALEN$_{CTG}$ within CTG Repeats

(A) Southern blots of yeast strains during DSB induction. For each wild-type and mutant strain, cells were collected at different time points after induction (+Dox or −Dox). The time point labeled "0" represents the first time point which the DSB was detectable. For all experiments, it corresponds to 14 hr after induction (+Dox or −Dox). When the TALEN$_{CTG}$ was repressed (OFF), no band corresponding to the DSB was visible. When the TALEN$_{CTG}$ was induced (ON), several signals were detected: a smear corresponding to successive contractions of the large trinucleotide repeat tract located 5′ of the DSB, a band corresponding to the 5′ end of the DSB with only a few triplets left, and another band corresponding to the 3′ end of the DSB containing 1–4 triplets. The cartoons at the left describe the different molecules detected. Blue triangles indicate the location of EcoRV sites used for restriction digestion before Southern blotting. Scissors indicate the DSB location.

(B) Quantification of 5′ and 3′ DSB signals. Note that the time course was run during 34 hr only in the *rad50*Δ and *sae2*Δ strains.

(C) Terminal transferase-mediated PCR. After DSB induction, dCTP was added to both 3′ strands by terminal transferase. The 3′ end of the break was subsequently amplified with a poly-dG oligonucleotide and another primer specific of the 3′ end of the DSB. Note that additional time points and controls were present on the same gels, but only significant time points are shown here.

(D) Results of terminal transferase-mediated PCR sequences. The repeat tract is shown in the blue box, and the right TALEN$_{CTG}$ binding site is indicated by red letters. The locations of the four TALEN$_{CTG}$ DSBs sequenced are indicated by black arrowheads.

One time course was performed for wild-type, *lig4*Δ, *rad51*Δ, and *pol32*Δ strains. For the *rad50*Δ and *sae2*Δ strains, values are the average of two or three independent experiments, depending on the time points considered. Error bars correspond to one SD.

were more accessible to the nuclease. This experiment also showed that smear length progressively decreased over time, although the 5′ smear intensity was too low to be reliably quantified (Figure 2A). Given the rate of decrease of the parental band, it is expected to see its complete disappearance after 6 days of induction (Figure 2C).

Time courses for *dnl4*Δ, *pol32*Δ, and *rad51*Δ mutants were similar to the wild-type strain, with a maximum of 12.4% of broken molecules in the *dnl4*Δ mutant (Figures 1A and 1B). We concluded that none of these mutants showed a detectable effect on DSB repair kinetics. On the contrary, in the *rad50*Δ mutant, an accumulation of DSBs was observed (Figure 1A). On both sides of the break, a signal increase was clearly detected (Figure 1B). This suggested that a DSB induced in CTG repeats was not correctly processed in this mutant, leading to an accumulation of unrepaired broken molecules. In the *sae2*Δ mutant, the 5′ and 3′ ends of the DSB showed an asymmetric increase compared with *rad50*Δ cells. The DSB end containing most of the trinucleotide repeat tract (5′ end) shows the same accumulation rate as in *rad50*Δ, whereas the other DSB end, containing only a few triplets, accumulates more slowly (Figures 1A and 1B).

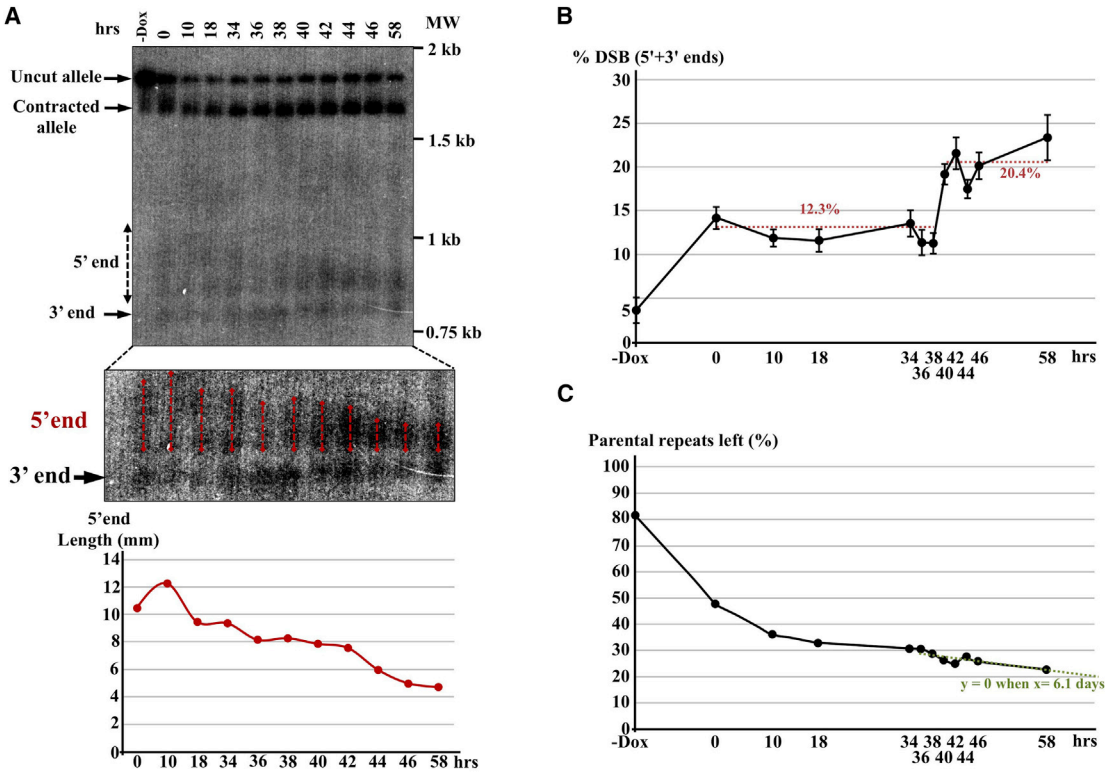


Figure 2. Expression of a TALEN$_{CTG}$ during a 72-hr Time Course

(A) Southern blot of a time course in a haploid strain containing only one CTG repeat allele, for 3 days.

(B) Quantification of the DSB level during the time course. For this experiment, 5′ and 3′ end signals were added and compared with the total signal in each lane. The two horizontal red dotted lines correspond to average DSB levels in each time frame.

(C) Quantification of the amount of parental repeat length left during TALEN$_{CTG}$ induction. The ratio of the parental allele over total signal is represented. The amount of parental allele decreased rapidly during the first day, corresponding to cells that repaired the DSB by SSA and, therefore, contracted the repeat tract. Then parental allele reduction occurs more slowly, when new repeat tracts are cut and contracted over time. Linear regression of the slower part of the graph (green dotted line) shows that the parental allele would have completely disappeared 6 days after the beginning of the induction. Note that the graphs in both (B) and (C) show the average of two independent experiments. Error bars correspond to one SD.

In the *rad52*Δ mutant, no signal corresponding to DSB formation could be detected by Southern blot (Figure S2A). TALEN expression was verified by western blot analysis (Figure S3). In the presence of doxycycline, no signal was detected in any of the strains. In the absence of doxycycline, the hemagglutinin (HA)-tagged TALEN was clearly detected in all strains. Its relative level was similar in *rad50*Δ, *rad51*Δ, and *rad52*Δ strains but ~10-fold higher in wild-type cells. We concluded that the absence of visible DSB in the *rad52*Δ strain was not due to a lack of expression of the nuclease because it was present in similar amounts in the two other mutant strains, in which the DSB was clearly detected. To check whether the absence of detectable DSB was due to some mutation unrelated to the *RAD52* deletion itself, we performed the same experiment in a *rad52*Δ/*RAD52* heterozygote. In this strain, the DSB was clearly visible, showing that complementing the *rad52*Δ deletion with a *RAD52* gene restored the wild-type phenotype (Figure S2B).

Subsequently, a terminal deoxynucleotidyl transferase-mediated PCR approach was used to amplify the DSB (Förstemann et al., 2000). PCR products were visible in the wild-type and *rad50*Δ strains used as positive controls at 8 hr and 10 hr, but no product was detected when the TALEN$_{CTG}$ was repressed (Figure 1C). PCR products corresponding to DSB amplification were also visible in the *rad52*Δ mutant when the TALEN$_{CTG}$ was expressed, but very faintly. We concluded that DSBs occurred as expected within the repeat tract in the *rad52*Δ strain but that their level was too low to be detected by Southern blot. Sanger sequencing of the terminal end of the PCR product generated by terminal deoxynucleotidyl transferase-mediated PCR products showed that the DSB occurred in the very last 1–4 CTG triplets of the repeat tract (Figure 1D).

DSB Accumulation in *rad50*Δ Depends on CTG Repeats

To determine whether DSB accumulation was only dependent on the presence of CTG repeats at the end of the DSB, a second TALEN was designed to recognize a *SUP4* allele that did not contain a repeat tract. This TALEN was called TALEN$_{noCTG}$ to distinguish it from the TALEN$_{CTG}$. In the wild-type strain, the DSB signal was weaker because only one of the two chromosomes could be cut by the TALEN$_{noCTG}$ (Figures 3A and 3B).

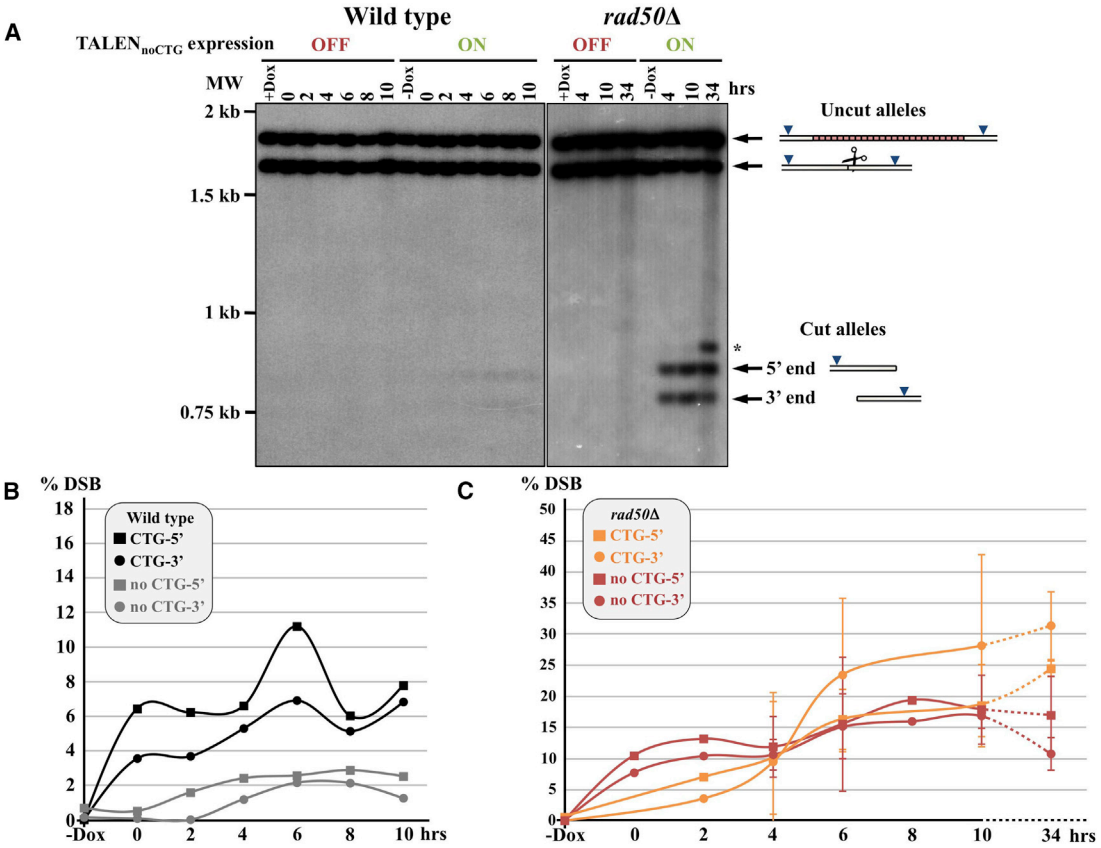


Figure 3. DSB Induction by a TALEN$_{noCTG}$ within a Non-repeated Region

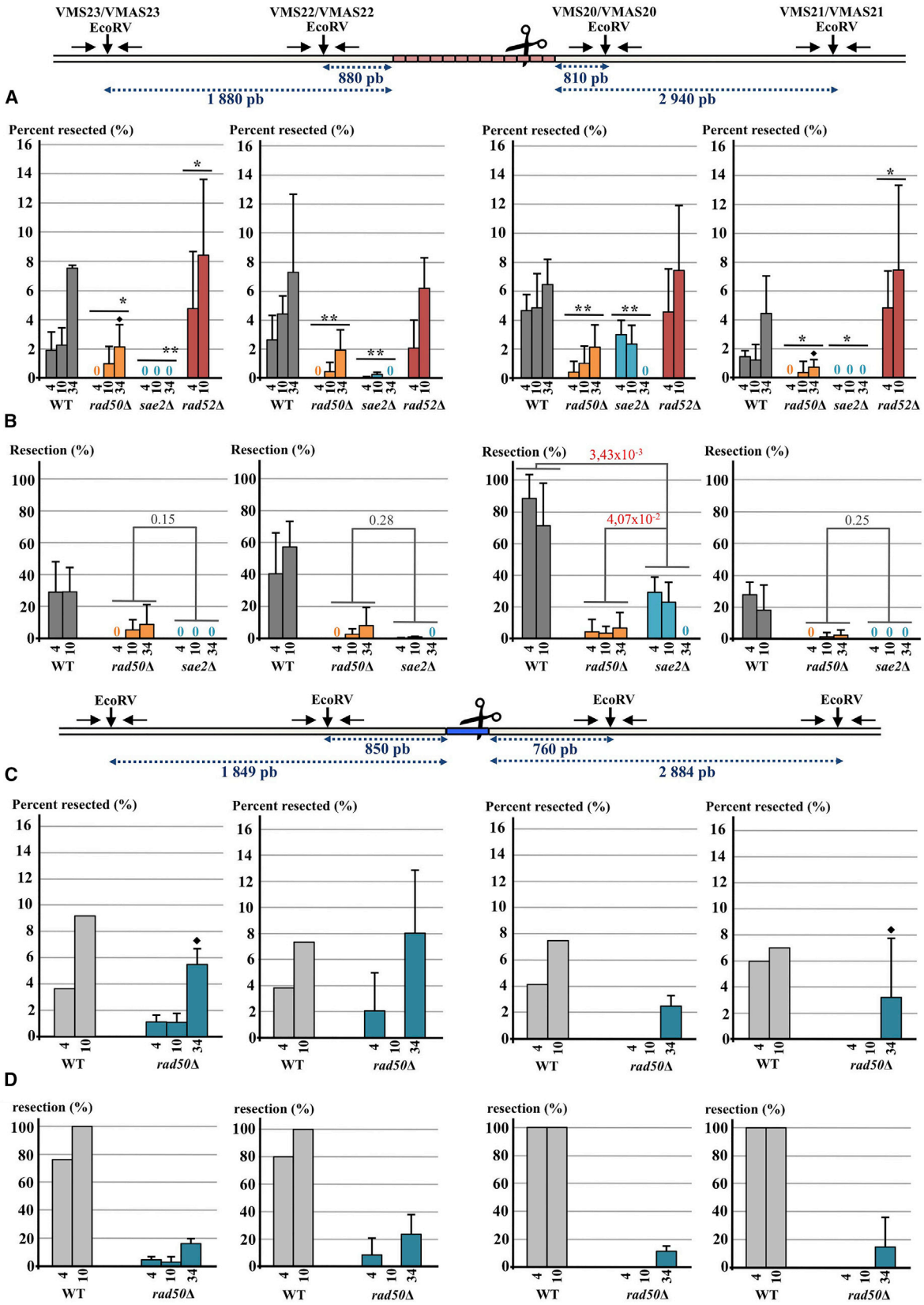
(A) Southern blots of time courses during DSB induction of the TALEN$_{noCTG}$ in wild-type and *rad50*Δ strains. The asterisk indicates an extra band only visible in the *rad50*Δ strain, probably corresponding to some chromosomal rearrangement specific to this mutant.

(B) Quantifications of the TALEN$_{noCTG}$ 5′ and 3′ DSB signals and comparisons with the TALEN$_{CTG}$. For each time point, the amounts of 5′ or 3′ signals were quantified and plotted as a ratio of the total signal in the lane. One time course was performed for the wild-type strain. For the *rad50*Δ strain, values are the average of two or three independent experiments, depending on the time points. Error bars correspond to one SD.

The smear detected when the TALEN$_{CTG}$ was induced was not visible with the TALEN$_{noCTG}$, proving that is corresponds to different repeat lengths because of progressive repeat contraction over time. The number of broken molecules increased at a slower rate over time compared with the TALEN$_{CTG}$. Six hours after DSB, non CTG-containing ends are four times less abundant than CTG-containing ends. Repeat-containing broken molecules are more persistent, suggesting that non CTG-containing ends are repaired faster than CTG-containing ends. In the *rad50*Δ mutant, the DSB also accumulates in the TALEN$_{noCTG}$ strain compared with the TALEN$_{CTG}$ strain (Figure 3C). However, the amount of non CTG-containing ends decreases slowly after 10 hr, whereas CTG-containing ends keep on accumulating. We concluded that cut fragments were greatly stabilized in the absence of *RAD50* but that repair of non-repetitive ends eventually occurs at the last time point, whereas it is definitely compromised when CTG-containing ends need to be repaired. This strongly suggests that these repeats form secondary structures at DSB ends that need the Mre11-Rad50-Xrs2 (MRX) complex to be removed for repair to occur.

DSB Resection within CTG Repeats Is Almost Completely Abolished in *rad50*Δ and *sae2*Δ

DSB resection was determined by qPCR of total genomic DNA preliminary digested by a restriction endonuclease (EcoRV in the present case) (Chen et al., 2013). Restriction sites that were resected during the course of the experiment could not be digested by EcoRV because of their single-stranded nature and could therefore be amplified by PCR primers located around the restriction site. Comparisons of cycle threshold (Ct) obtained in the fraction digested by EcoRV with Ct obtained in the undigested fraction was indicative of the amount of resection at this particular restriction site. Four EcoRV sites were studied: two of them located 800–900 base pairs (bp) upstream and downstream of the repeat tract and two located further away, 1.8–2.9 kb upstream and downstream of the CTG repeat (Figure 4). In all experiments, raw resection and relative resection values were calculated. Raw resections were computed as a percentage of PCR product amplified in EcoRV digested compared with non-digested fractions (Experimental Procedures). Relative resection values were calculated as the ratio of raw values to DSB amounts quantified on Southern blots.



(legend on next page)

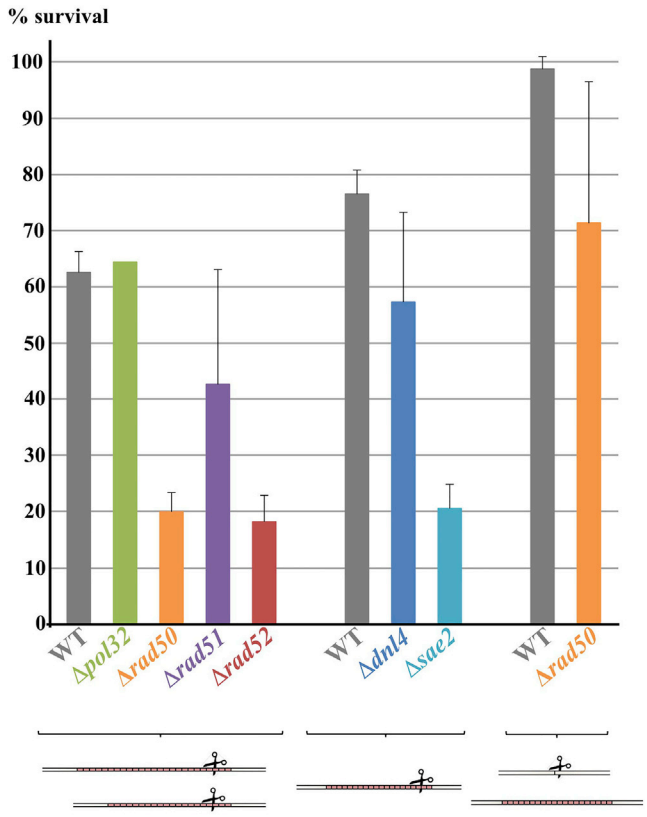


Figure 5. Survival of TALEN-Induced DSBs

Left: survival in diploid cells carrying two alleles of different trinucleotide repeat tract lengths. In these strains, both chromosomes carry trinucleotide repeats and were cut by the nuclease, as shown in the bottom cartoon. Center: survival in haploid cells carrying only one repeat tract. Right: survival in diploid cells carrying only one repeat tract. In these strains, only the chromosome that did not carry trinucleotide repeats was cut by the nuclease. Survival values are the average of two to five experiments, expect for *pol32*Δ. Error bars are equal to one SD.

Because DSB signals in the *rad52*Δ strain were too low to be quantified, relative resection values could not be calculated for this mutant.

In the wild-type strain, raw resection as well as relative resection values were significantly higher at sites proximal to the DSB (VMS20/VMAS20 and VMS22/VMAS22) than at distal sites (VMS21/VMAS21 and VMS23/VMAS23) (Figures 4A and 4B; two-tailed t test p = 0.0085). Resection values dramatically dropped in the *rad50*Δ strain as well as in the *sae2*Δ mutant to become barely detectable at early time points. Note that *SAE2* was found to be more important for relative resection on the DSB end containing the longer repeat tract, whereas *RAD50* was required on both DSB ends (Figure 4B; VMS20/VMAS20, p = 0.04).

In the *rad52*Δ strain, resection was not different from the wild-type at proximal sites but increased at distal sites, suggesting that, in the absence of *RAD52*, resecting enzymes have a better access to DNA ends.

In the wild-type strain expressing the $TALEN_{noCTG}$, raw as well as relative resection values were much higher compared with the $TALEN_{CTG}$ values, particularly at distal EcoRV sites (Figures 4C and 4D). In addition, there was no detectable difference between resection values at proximal versus distal sites. In the *rad50*Δ mutant expressing the $TALEN_{noCTG}$, resection was increased 34 hr after DSB formation compared with the $TALEN_{CTG}$ *rad50*Δ strain, but this increase was found to be significant only at distal restriction sites (Figures 4A and 4C; two-tailed t test, p = 0.0197). This shows that, in the absence of *RAD50*, resection is less efficient when CTG repeats are present at the DSB.

RAD50, *RAD52*, and *SAE2* Are Needed to Repair a TALEN-Induced DSB

Survival of the DSB was determined in each strain by the ratio of colony-forming units (CFUs) when the $TALEN_{CTG}$ was expressed to CFUs when the $TALEN_{CTG}$ was repressed. In wild-type diploid cells, survival was 62.5% ± 3.7%, not significantly different from *pol32*Δ and *rad51*Δ mutant strains (64.4% and 42.6% ± 20.4%, respectively), although a somewhat higher SD was observed for *rad51*Δ (Figure 5). Therefore, the absence of these genes did not significantly affect DSB repair efficacy, consistent with similarities in DSB formation and processing during time courses (Figure 1). On the contrary, the absence of *RAD50* or *RAD52* led to a higher mortality because only 19.9% ± 3.4% of cells survived in the *rad50*Δ strain and 18.2% ± 4.7% survived in *rad52*Δ cells, proving that the product of both genes was required to repair the break. Survival in diploid strains heterozygous for the *rad50*Δ or the *rad52*Δ deletion was not significantly different from the wild-type (Figure S2C). When a chromosome that did not contain CTG repeats was cut by the $TALEN_{noCTG}$ in the *rad50*Δ strain, survival was not significantly different from the wild-type, confirming that this gene product was not essential to repair a DSB in non-repeated DNA (Figure 5).

Figure 4. Quantification of DSB Resection

Resection graphs for each primer pair are plotted under each EcoRV site.

(A) Raw values of resection with the $TALEN_{CTG}$. Statistical comparisons between the wild-type and each mutant strain were determined by two-tailed t tests and are shown above each graph (*p < 0.05, **p < 0.01). Diamonds show significant differences between $TALEN_{CTG}$ and $TALEN_{noCTG}$ in *rad50*Δ strains at both distal sites at 34 hr (t test, p = 0.0197, comparison with C).

(B) Relative values of resection with the $TALEN_{CTG}$. Same as (A), except that resection values were divided by the amount of DSBs detected on Southern blots. Statistical comparisons between the wild-type, *rad50*Δ, and *sae2*Δ were determined by two-tailed t tests and are shown by vertical gray lines along with corresponding p values. Because DSB signals were not detectable by Southern blots in *rad52*Δ, relative resection values could not be calculated.

(C) Raw values of resection with the $TALEN_{noCTG}$. Diamonds show significant differences between $TALEN_{CTG}$ and $TALEN_{noCTG}$ in *rad50*Δ strains at both distal sites at 34 hr (t test, p = 0.0197, comparison with A). Proximal sites do not show any significant difference.

(D) Relative values of resection with the $TALEN_{noCTG}$.

Raw and resection values are the average of two to four independent experiments for each strain and each time point, except in the wild-type strain with the $TALEN_{noCTG}$ (only one experiment), but in this strain, 100% of broken molecules were resected at each EcoRV site. Error bars are equal to one SD.

In haploid wild-type cells that are unable to repair the DSB by homologous recombination, 76.5% ± 4.3% of cells survived. This frequency slightly decreased in the *dnl4*Δ mutant (57.3% ± 16%) but was not significantly different from the wild-type (t test, p = 0.06). A significant decrease in survival was observed in haploid *sae2*Δ cells (20.5% ± %4.3%), proving that this gene was also needed to repair such a break.

DSB Repair within CTG Repeat Tracts Is Mainly an Intramolecular Mechanism

Repeat lengths were analyzed in several surviving colonies after $TALEN_{CTG}$ induction in wild-type and mutant strains by two different techniques. First, Southern blots were run on yeast colonies for each strain to determine the overall range of allele contractions (Figure 6A). Then a subset of these colonies was PCR-amplified at the repeat locus and Sanger-sequenced. When both alleles carried repeat tracts of the same exact length, the sequence was very clear before and after the repeat tract, as previously demonstrated (Richard et al., 2014). On the contrary, a sequence becoming fuzzy after the repeat tract was the signature of two alleles of different lengths. Sequences were therefore classified in two categories: homozygous (when both repeat tract alleles shared the same length) or heterozygous (different lengths). Distribution of repeat lengths in wild-type survivors showed that a large majority of clones (86.1%) carried contracted repeats with less than 20 CTG triplets, most of them (51.6%) exhibiting very large contractions (only 4–10 CTG triplets left; Figure 6B). Among survivors, only 11.1% were homozygous, all of them exhibiting large contractions. These homozygous survivors may correspond to a minority of cells that have repaired the DSB by gene conversion using an already contracted repeat tract as a template or, alternatively, to cells that independently repaired both alleles to the same length by chance.

Distribution of repeat lengths observed in *pol32*Δ and *rad51*Δ mutants was not statistically different from the wild-type (homogeneity chi-square test = 5.10 and 5.58, respectively) (Figure 6B). In *rad51*Δ, no homozygous clone was found, consistent with the hypothesis that the few homozygous clones observed in wild-type cells indeed corresponded to gene conversion events. Distributions of repeat lengths in *rad50*Δ and *rad52*Δ were significantly different from the wild-type (chi-square test = 31.33 and 20.08, respectively). In both strains, all survivors were heterozygous, and a large majority of them harbored short repeat lengths (20–25 triplets in *rad50*Δ, 4–10 triplets in *rad52*Δ). The high mortality in these two mutant backgrounds suggests that surviving clones had spontaneously contracted CTG repeat tracts below the minimal length required for TALEN nuclease activity (less than 17 triplets; Richard et al., 2014) and, therefore, did not receive any DSB. Alternatively, survivors may correspond to a subset of cells that were able to repair the break in the absence of Rad50 or Rad52, both hypotheses being not mutually exclusive. The few homozygous events detected in *rad52*Δ (5.7%) probably correspond to cells that independently contracted both alleles to the same length by chance. There is no statistical difference between repeat tract length distribution in *dnl4*Δ or *sae2*Δ strains compared with the wild-type, most survivors being contracted to 4–10 CTG triplets (Figure 6C). However, we must note that only 18 of 48 *sae2*Δ survivors showed a clear sequencing product, suggesting that the length distribution observed might represent only a subset of all repair events obtained in this mutant background.

All of these results showed that repair of a DSB induced in CTG repeats involved *RAD50*, *SAE2*, and *RAD52*. We ruled out that this repair could occur by BIR or gene conversion because neither *POL32* nor *RAD51* was involved. Instead, SSA, which is a non-conservative intrachromosomal recombination mechanism depending on *RAD52*, appeared to be the favored pathway to repair the break. Recombination may occur between CTG repeats present on both sides of the break, eventually resulting in large repeat contractions. Two arguments support iterative cycles of repeat contractions. First, progressive contractions of repeat tracts occurred between time points during TALEN induction (Figure 7A). Second, the reduction of smear length was clearly visible on Southern blots over the duration of a time course (Figures 2 and 7B).

DISCUSSION

Integrity of Sae2 and of the Mre11-Rad50 Complex Is Essential for DSB Processing within CTG Repeats

Former studies showed that Mre11 was not required to process "clean ends" such as those resulting from multiple HO DSBs in yeast (Llorente and Symington, 2004), although resection and single-strand DNA formation were delayed when the Mre11-Rad50 complex was not functional (Lee et al., 1998; Sugawara and Haber, 1992). Here we show that, when a DSB was induced in a *rad50*Δ strain in non-repeated DNA ($TALEN_{noCTG}$), resection and repair were delayed, but survival was not significantly decreased, confirming that the Mre11-Rad50 complex was not essential to resect clean ends (Figure 4C).

On the contrary, resection and repair of a TALEN-induced DSB within a CTG trinucleotide repeat was completely abolished in a *rad50*Δ strain (Figures 4A and 4B). This result is consistent with a former work in which a natural chromosomal break within a long CTG repeat tract in yeast was left unrepaired and accumulated in a *rad50*Δ strain in such proportions that it was possible to detect the broken chromosome by pulse-field gel electrophoresis (Freudenreich et al., 1998). It is also compatible with a recent study using *Xenopus* egg extracts in which Liao et al. (2016) showed that Mre11 was essential for resection of DNA with 3′ damaged nucleotides and 3′ or 5′ bulky adducts.

Sae2 as well as the Mre11-Rad50 complex were previously shown to be required to resolve hairpin-capped natural DSBs in yeast cells (Lobachev et al., 2002). Consistent with this study, the purified Sae2 protein was shown to exhibit endonuclease activity on DNA gaps close to a hairpin structure, and this activity was stimulated by the Mre11-Rad50 complex (Lengsfeld et al., 2007). Later on, Sae2 was shown to be involved in resection at the *MAT* locus following HO DSB only in the absence of the single-strand binding protein Rfa1. In that particular case, Sae2 was required to remove hairpin-like folded back structures at DSB ends (Chen et al., 2013). In our present experiments, a clear absence of resection was observed in the *sae2*Δ mutant on the 5′ DSB end that contained most of the repeat tract

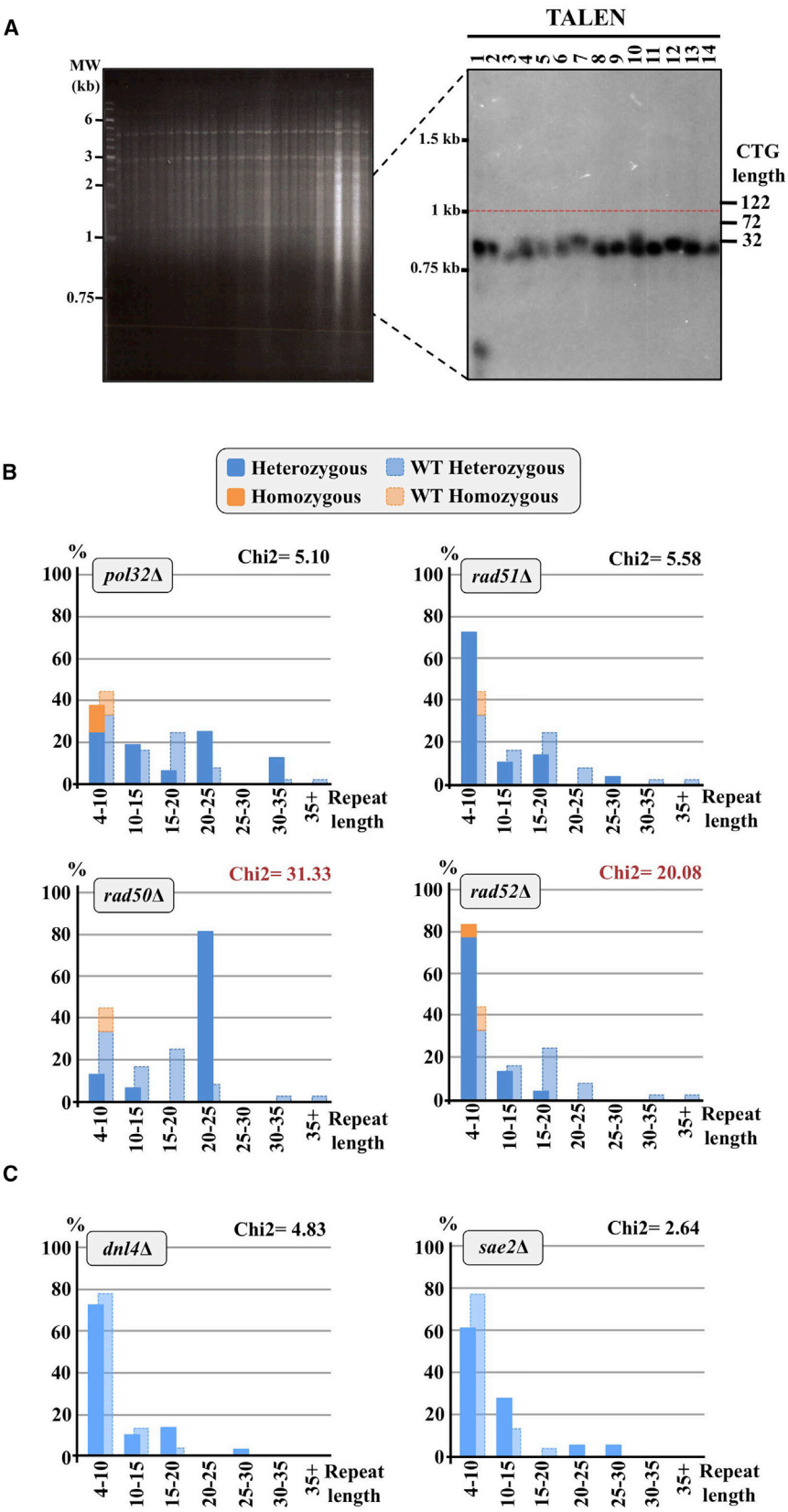


Figure 6. Molecular Analysis of CTG Repeat Length after DSB Repair

(A) Representative Southern blot showing 14 haploid colonies in which the $TALEN_{CTG}$ was induced. In each lane, total genomic DNA was extracted from one single yeast colony and analyzed. Two molecular weight markers were used: the GeneRuler 1-kb ladder (Sigma-Aldrich) on the left and a homemade CTG repeat tract length on the right (Viterbo et al., 2016). The red dotted line shows the parental CTG repeat tract length. In colonies after $TALEN_{CTG}$ induction, one or more bands containing contracted CTG repeat tracts were detected.

(B) Molecular analysis of cells after $TALEN_{CTG}$ induction. DNA was extracted from colonies after DSB induction, and the repeat containing-locus was sequenced. After Sanger sequencing of the PCR product, two outcomes could be obtained. when the two alleles contained the same exact number of triplets, one unique sequence was clearly read (homozygous, in orange); when the two alleles contained different numbers of triplets, the sequence was blurry and unreadable after the shortest of the two repeat tracts (heterozygous, in blue). Repeat lengths are given in number of triplets. The number of colonies sequenced in each strain was as follows: WT, 36; *pol32*Δ, 16; *rad50*Δ, 31; *rad51*Δ, 29; *rad52*Δ, 53. Chi-square test values (degrees of freedom [ddl] = 3) of comparisons between wild-type and mutant distributions are indicated above each graph. Only two distributions (in red) are significantly different from the wild-type: *rad50*Δ and *rad52*Δ.

(C) The same as (B), except that repeat tract lengths were compared between the wild-type strain and *dnl4*Δ or *sae2*Δ haploid strains, so only one trinucleotide repeat allele was present. The number of colonies sequenced in each strain was as follows: WT, 32; *dnl4*Δ, 40; *sae2*Δ, 18. chi-square test values of comparisons between wild-type and mutant distributions are indicated above each graph (ddl = 2).

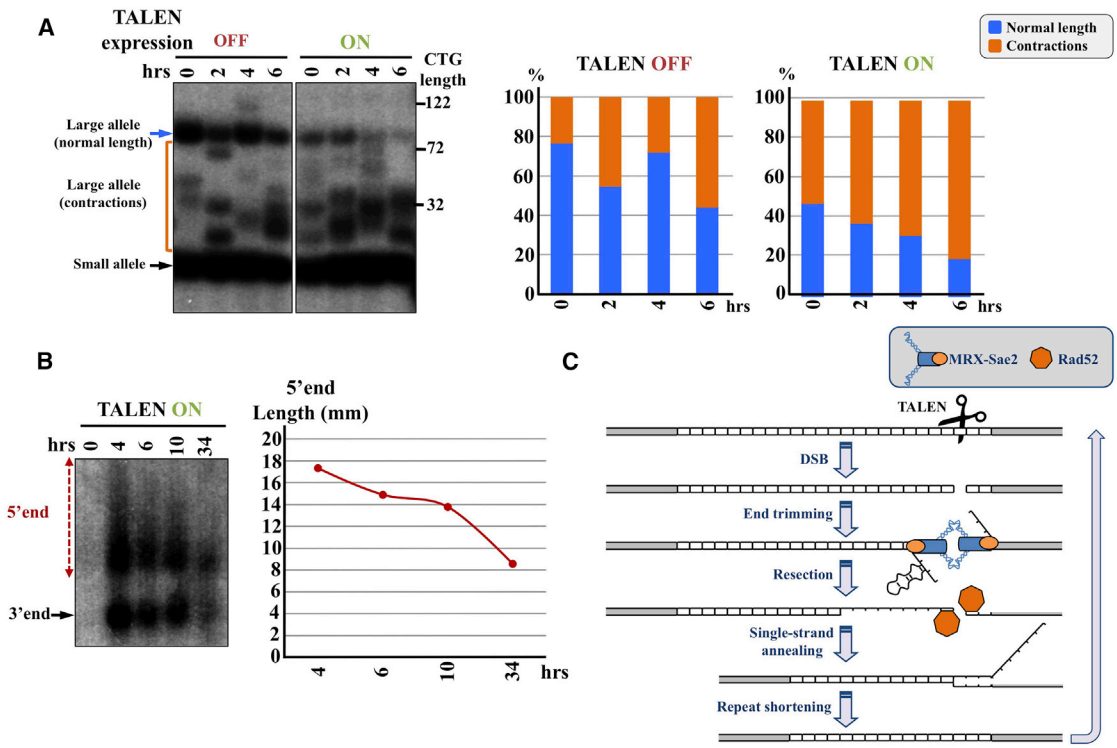


Figure 7. A Model of Progressive CTG Repeat Contractions following Iterative DSB Repair

(A) Diploid yeast cells were collected at several time points during TALEN induction and plated on non-inducible medium so that the nuclease was turned off during colony growth. From 20 to 30 colonies were picked and pooled before DNA extraction and analysis. Repeat tract lengths were analyzed by Southern blot. In each lane, three types of signals were detected: the large uncontracted allele (normal length), large contracted alleles, and the small allele. Note that the small allele is too small to be efficiently cut and contracted by the TALEN in this experiment. Graphs show quantifications of large allele signals, when the TALEN was repressed or expressed during the time course.

(B) Total smear length was measured in a wild-type time course (left), showing length reduction over time (right).

(C) Mechanistic model for CTG repeat contractions following a TALEN-induced DSB. After DSB formation, the ends of the break require the integrity of the MRX complex and Sae2 to be trimmed. Following resection, Rad52 binds to DSB ends and catalyzes SSA between the two ends of the DSB. With the DSB occurring very close to the repeat tract end, only a few triplet repeats may be involved in SSA, leading to a moderate repeat length shortening. The shortened trinucleotide repeats may be the substrate of one or more other round(s) of breakage and SSA until the repeat tract is too short for the TALEN to bind and induce a DSB.

(Figure 4B). On the 3′ end, though, resection was reduced at the proximal site compared with the wild-type but was higher than in the *rad50*Δ mutant. It is possible that the few triplet repeats left at the 3′ end after DSB induction are sufficient to form a small hairpin that decreases or delays resection. This delay might also explain why resection at the distal EcoRV site is decreased on the 3′ end of the DSB. We concluded that DSB ends of CTG trinucleotide repeats most probably harbor some kind of secondary structures, are therefore not clean, and absolutely require a functional Mre11-Rad50 complex as well as Sae2 to be processed for repair to start.

On the opposite, resection was increased at longer distances in the *rad52*Δ mutant (Figure 4A). This suggests that binding of the Rad52 protein on early recombination intermediates interferes with DNA end resection, as already shown in former publications (Van Dyck et al., 1999; Frank-Vaillant and Marcand, 2002; Parsons et al., 2000; Ristic et al., 2003; Sugawara and Haber, 1992; White and Haber, 1990). It is at the present time unclear how competition occurs and is resolved between Rad52p and resection proteins.

Differences between Yeast and Mammalian Systems

Former works showed the role of the Mre11-Rad50 complex on natural CTG trinucleotide repeat expansions in yeast (Sundararajan et al., 2010; Ye et al., 2016). When an I-*Sce* I DSB was repaired using a long CTG trinucleotide repeat as a template, expansions occurring during DSB repair were larger in strains overexpressing *MRE11* or *RAD50* (Richard et al., 2000). In the present experiments, DSB induction in long CTG repeats only led to contractions of the repeat tract, in wild-type as well as in *rad50*Δ strains (Figure 6B), and no expansion was ever observed. Cinesi et al. (2016) reported some expansions when inducing a DSB within a CTG repeat tract in human cells using either wild-type Cas9 or the mutant Cas9D10A nickase. This suggests that DSB repair mechanisms within CTG repeats exhibit subtle differences between yeast and human cells. The chromatin environment in human cells is different and may affect the way a DSB within a CTG repeat tract will be repaired (reviewed in House et al., 2014). In addition, although human cells contain *RAD51* and *RAD52* homologs, two additional genes, *BRCA1* and *BRCA2*, involved in breast cancer, play a central

role in homologous recombination, whereas yeast cells lack these genes (Moynahan et al., 1999, 2001). Comparing results obtained in yeast and in human cells will also hopefully help our understanding of the respective roles of these factors during DSB repair of CTG repeats.

Single-Strand Annealing Is the Main Mechanism of DSB Repair within a CTG Repeat Tract

Former studies of SSA requirements on direct repeats showed that its efficacy relied on three factors: homology length between the two repeated sequences, resection rate, and proximity on the DNA molecule, with closer sequences recombining more easily than distant ones (Lazzaro et al., 2008; Sugawara and Haber, 1992). In addition, *RAD52* was shown to be important for SSA reaction between 15- to 18-bp microhomologous sequences but strongly inhibited SSA between 6- to 13-bp microhomologies (Villarreal et al., 2012). In the present case, the DSB was made close to the 3′ end of the repeat tract, leaving only 1–4 repeat units (3–12 bp) on the 3′ end of the break but a much longer stretch of repeats on the 5′ end (around 70 triplets). SSA between triplet repeats was partially *RAD52*-dependent (survival was 3-fold decreased), suggesting that 15-bp microhomologies were sometimes present and used for SSA between triplet repeats. These results are in good accordance with our former work in which repair occurred in 67% of the cases by annealing between two short repeats flanking an I-*Sce* I restriction site (Richard et al., 1999).

Although no effect of *POL32* and *RAD51* on DSB repair of a CTG repeat tract was detected in the present experiments, it is interesting to mention that both genes were involved in spontaneous expansions of CAG repeats, probably by a BIR-related mechanism. However, expansions rates were low (10^{-5}–10^{-6} per cell per division), and the authors could not test the possible role of these two genes in repeat contractions in their experimental system (Kim et al., 2017).

A Model Supporting Progressive Repeat Contractions Associated with TALEN-Induced DSB Repair

We propose a model involving iterative SSA between short repeat-containing DNA ends after DSB induction (Figure 7C). In this model, the Mre11-Rad50 complex and Sae2 are essential to process DSB ends, after which Rad52 annealing activity catalyzes the SSA reaction. Given that the DSB occurs only a few triplets before the end of the repeat tract, homology available to anneal both ends is very small. This "short SSA" can hardly lead to large repeat contractions in one single step. We thus propose that, following repair of this first DSB, the repeat tract may still be a substrate for the nuclease and could receive a second DSB, leading to further contraction of the repeat, and so on, until it is too small to be efficiently cut by the $TALEN_{CTG}$. Progressive contractions of repeat tracts (Figure 7A) as well as reduction of smear length between time points (Figure 7B) support this model. However, we cannot completely exclude near-complete contraction of repeat tracts in a subpopulation of cells receiving a DSB, given that TALEN efficacy is very low (~10% broken molecules at each time point). Hence, this apparent progressive repeat contraction could indeed represent complete contraction in this subpopulation at each time point. It is unclear at the present time why the TALEN is so inefficient at inducing a DSB compared with known meganucleases such as HO or I-*Sce* I expressed in yeast cells. Further experiments are needed to address this specific question.

EXPERIMENTAL PROCEDURES

Yeast Strains and Plasmids

All mutant strains were built from strains GFY6162-14A and GY6162-3D by classical gene replacement method (Orr-Weaver et al., 1981) using KANMX4 as a marker (Table S1), amplified from the European *Saccharomyces cerevisiae* archive for functional analysis (EUROSCARF) deletion library, using primers located 1 kb upstream and downstream of the cassette (Table S2). The VMY350 strain was used to construct the VMY650 strain by mating-type switching, using the pJH132 vector carrying an inducible HO endonuclease (Holmes and Haber, 1999). Plasmid pCLS9996 carrying the $TALEN_{CTG}$ right arm was digested by NcoI (New England Biolabs) and EagI (Takara). The fragment containing the right arm was cloned in the centromeric pCMha182, digested by BamHI (NEB) and PstI (NEB) using two oligomeric adaptors of 16 bp (BamHI-NcoI) and 19 bp (EagI-PstI). The resulting vector, pCMha182KN9996, was transformed in the haploid strain GFY6162-14A and its mutant derivatives. Plasmid pCLS16715 carrying the $TALEN_{CTG}$ left arm was digested by NcoI (NEB) and EagI (Takara). The fragment containing the left arm was cloned in the centromeric pCMha188, digested by BamHI (NEB) and PstI (NEB) using two oligomeric adaptors of 16 bp (BamHI-NcoI) and 19 bp (EagI-PstI). The resulting vector, pCMha188KN16715, was transformed in the GFY6161-3D haploid strain and its mutant derivatives. Haploid transformants were crossed on rich medium (yeast extract peptone dextrose medium [YPD]) supplemented with doxycycline (10 μg/mL), and diploids containing both $TALEN_{CTG}$ arms were selected on synthetic complete medium lacking uracil and tryptophan (SC-Ura-Trp) with doxycycline (10 μg/mL). For the *dnl4*Δ mutant, both $TALEN_{CTG}$ arms were transformed in haploid strains GFY6162-3D and VMY104 because NHEJ is downregulated in diploid cells (Frank-Vaillant and Marcand, 2001; Valencia et al., 2001).

The $TALEN_{noCTG}$ was designed by the Muséum National d'Histoire Naturelle platform. The target sequence was chosen to be an I-*Sce* I recognition site integrated in the *SUP4* gene (Richard et al., 1999). Plasmid pR1 was used as a template to PCR-amplify the $TALEN_{noCTG}$ right arm using primer pairs VMS25/VMAS25, containing a 50-bp tail homologous to sequences flanking a KpnI site on pCMha182KN. The PCR product and pCMha182KN linearized by KpnI (NEB) were directly cloned in yeast cells. Plasmid pL1 was used as a template to PCR-amplify the $TALEN_{noCTG}$ left arm using primer pairs VMS25/VMAS25, containing a 50-bp tail homologous to sequences flanking a KpnI site on pCMha188KN. The PCR product and pCMha188KN linearized by KpnI (NEB) were also directly cloned in yeast. Centromeric vectors pCMha182KNR1 and pCMha188KNL1, carrying respectively, the $TALEN_{noCTG}$ right arm and the $TALEN_{noCTG}$ left arm, were transformed in diploid strains VMY001 and VMY002.

TALEN Inductions

Before nuclease induction, Southern blot analyses were conducted on several independent subclones to select cells with different CTG repeat tract lengths on both chromosomes. Yeast cells were grown at 30°C in liquid SC-Ura-Trp medium complemented with 10 μg/ml of doxycycline. Cells were washed with sterile water to eliminate doxycycline and then split in two cultures at 9×10^6 cells/mL, one in SC-Ura-Trp medium complemented with 10 μg/mL of doxycycline (TALEN-repressed) and the other in SC-Ura-Trp (TALEN-induced). For each time point, 2×10^8 cells were collected at time point (T) = 0, 14, 16, 18, 20, 22, 24, or 48 hr afterward, rapidly centrifuged, washed with water, and frozen in dry ice before DNA extraction. To determine viability after DSB induction, cells were plated at 24 hr on SC-Ura-Trp plates supplemented with doxycycline (10 μg/mL) for the TALEN-repressed culture and on SC-Ura-Trp plates for the TALEN-induced culture. CFUs were counted after 3–5 days of growth at 30°C.

DSB Analysis and Quantification

Total genomic DNA (4 μg) of cells collected at each time point was digested for 6 hr by EcoRV (20 U) (NEB) and analyzed by Southern blot as described previously (Viterbo et al., 2016). Alternatively, a terminal transferase-mediated PCR assay (Förstemann et al., 2000) was used to amplify the TALEN-induced DSB. Genomic DNA (100 ng) of cells collected at different time points was heat-denatured and treated with 7 U of terminal deoxynucleotidyl transferase (Takara) in a volume of 10 μL (100 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid [HEPES], pH 7.2; 40 mM $MgCl_2$; 0.5 mM DTT; 0.1% BSA; and 1 mM deoxycytidine triphosphate [dCTP]) for 30 min at 37°C to add polyC tails to 3′OH free ends. The enzyme was inactivated for 10 min at 65°C and 5 min at 94°C. Then 30 μL of PCR mix was added to each reaction to obtain a final volume of 40 μL containing 67 mM Tris HCl (pH 8.8), 16 mM $(NH_4)_2SO_4$, 5% glycerol, 0.01% Tween, 200 μM each deoxynucleotide triphosphate (dNTP), and 40 nM of each primer (G18 and VMS14). The following PCR program was used: 94°C for 2 min (94°C for 20 s, 62°C for 12 s, and 72°C for 20 s) for 45 cycles and then 72°C for 5 min. For each reaction, 20 μL was loaded on a 1% analytical agarose gel, and 20 μL was sent for Sanger sequencing.

Trinucleotide Repeat Length Analysis

Several colonies from each induced or repressed plates were picked, total genomic DNA was extracted, and 4 μg was digested for 6 hr by SspI (20 U) (NEB) and analyzed by Southern blot as described previously (Viterbo et al., 2016). Repeat tracts in some of the survivors were subsequently amplified using primers su3 and su7 and sequenced using a third internal primer, su2 (Table S2). Sanger sequencing was performed by GATC Biotech.

Analysis of DSB End Resection

A real-time PCR assay using primer pairs flanking EcoRV sites 0.81 kb and 2.94 kb away from the 3′ end of the CTG repeat tract (VMS20/VMAS20 and VMS21/VMAS21, respectively) and 0.88 kb and 1.88 kb away from the 5′ end of the CTG repeat tract (VMS22/VMAS22 and VMS23/VMAS23, respectively) was used to quantify end resection. Another pair of primers was used to amplify a region of chromosome X near the *ARG2* gene to serve as an internal control of the DNA amount (JEM1f-JEM1r). Genomic DNA of cells collected at T = 0 hr, T = 18 hr, T = 24 hr, and T = 48 hr was split in two fractions; one was used for EcoRV digestion and the other one for a mock digestion in a final volume of 15 μL. Samples were incubated for 5 hr at 37°C and then the enzyme was inactivated for 20 min at 80°C. DNA was subsequently diluted by adding 55 μL of ice-cold water, and 4 μL was used for each real-time PCR reaction in a final volume of 25 μL. PCRs were performed with Absolute SYBR Green Fluorescein Mix (Thermo Scientific) in a Mastercycler S Realplex (Eppendorf) using the following program: 95°C for 15 min, 95°C for 15 s, 55°C for 30 s, and 72°C for 30 s repeated 40 times, followed by a 20-min melting curve. Reactions were performed in triplicate, and the mean value was used to determine the amount of resected DNA using the following formula: raw resection = $2/(1+2^{\Delta Ct})$ with $\Delta Ct = C_{t,EcoRV} - C_{t,mock}$. Relative resection values were calculated by dividing raw resection values by the percentage of DSB quantified at the corresponding time point. All t tests were performed using the R package (Millot, 2011).

Western Blots

Liquid cultures were grown to exponential phase in the presence or absence of 10 μg/mL doxycycline. Proteins were extracted on 2×10^8 cells in 200 μL Laemmli solution with 100 μL glass beads. Proteins were separated on a 10% acrylamide gel under standard conditions and blotted to a nitrocellulose membrane (Optitran BA-S 83 reinforced NC, Schleicher & Schuell). For TALEN detection, a polyclonal anti-HA antibody was used (ab9110, Abcam, 0.25 μg/mL final concentration). For Msh2 detection, the primary antibody was a polyclonal rabbit antibody directed against an internal part of the yeast Msh2 protein (N3C2, GeneTex, 1 μg/mL final concentration) (Viterbo et al., 2016). A secondary goat anti-rabbit antibody conjugated to horseradish peroxidase was used for detection in both cases (Thermo Scientific, 0.16 μg/mL final concentration). Quantification was performed using a ChemiDoc MP Imager (Bio-Rad) with the dedicated Image Lab software. The molecular weight marker used was the Precision Plus Protein Standards All Blue (Bio-Rad).

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and two tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.01.083.

ACKNOWLEDGMENTS

V.M. was the recipient of a graduate student award from the Fondation pour la Recherche Médicale (PLP20131028794) and was supported by Fondation Guy Nicolas and Fondation Hardy. L.P. is the recipient of a graduate student CIFRE fellowship from SANOFI. This project was supported by the ValoExpress program of the Institut Pasteur (project TRINUCONTRACT).

AUTHOR CONTRIBUTIONS

V.M. designed and performed most of the TALEN experiments and all resection experiments, analyzed the data, and wrote the manuscript. L.P. performed some TALEN experiments, analyzed the data, and contributed to the manuscript. D.V. contributed to building mutant strains. M.C. designed and built the $TALEN_{noCTG}$. G.-F.R. designed the project, analyzed the data, and wrote the manuscript.

DECLARATION OF INTERESTS

There is a patent related to this work, held by the Institut Pasteur (WO 2015/078935 A1).

Received: February 27, 2017
Revised: December 19, 2017
Accepted: January 25, 2018
Published: February 20, 2018

REFERENCES

An, M.C., Zhang, N., Scott, G., Montoro, D., Wittkop, T., Mooney, S., Melov, S., and Ellerby, L.M. (2012). Genetic correction of Huntington's disease phenotypes in induced pluripotent stem cells. Cell Stem Cell *11*, 253–263.

Borde, V., Lin, W., Novikov, E., Petrini, J.H., Lichten, M., and Nicolas, A. (2004). Association of Mre11p with double-strand break sites during yeast meiosis. Mol. Cell *13*, 389–401.

Chen, H., Lisby, M., and Symington, L.S. (2013). RPA coordinates DNA end resection and prevents formation of DNA hairpins. Mol. Cell *50*, 589–600.

Cinesi, C., Aeschbach, L., Yang, B., and Dion, V. (2016). Contracting CAG/CTG repeats using the CRISPR-Cas9 nickase. Nat. Commun. *7*, 13272.

Davis, A.P., and Symington, L.S. (2004). RAD51-dependent break-induced replication in yeast. Mol. Cell. Biol. *24*, 2344–2351.

Fairhead, C., and Dujon, B. (1993). Consequences of unique double-stranded breaks in yeast chromosomes: death or homozygosis. Mol. Gen. Genet. *240*, 170–178.

Foiry, L., Dong, L., Savouret, C., Hubert, L., te Riele, H., Junien, C., and Gourdon, G. (2006). Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice. Hum. Genet. *119*, 520–526.

Förstemann, K., Höss, M., and Lingner, J. (2000). Telomerase-dependent repeat divergence at the 3′ ends of yeast telomeres. Nucleic Acids Res. *28*, 2690–2694.

Frank-Vaillant, M., and Marcand, S. (2001). NHEJ regulation by mating type is exercised through a novel protein, Lif2p, essential to the ligase IV pathway. Genes Dev. *15*, 3005–3012.

Frank-Vaillant, M., and Marcand, S. (2002). Transient stability of DNA ends allows nonhomologous end joining to precede homologous recombination. Mol. Cell *10*, 1189–1199.

Freudenreich, C.H., Kantrow, S.M., and Zakian, V.A. (1998). Expansion and length-dependent fragility of CTG repeats in yeast. Science *279*, 853–856.

Gacy, A.M., Goellner, G., Juranić, N., Macura, S., and McMurray, C.T. (1995). Trinucleotide repeats that expand in human disease form hairpin structures in vitro. Cell *81*, 533–540.

Haber, J.E. (1995). In vivo biochemistry: physical monitoring of recombination induced by site-specific endonucleases. BioEssays *17*, 609–620.

Holmes, A.M., and Haber, J.E. (1999). Double-strand break repair in yeast requires both leading and lagging strand DNA polymerases. Cell *96*, 415–424.

House, N.C.M., Koch, M.R., and Freudenreich, C.H. (2014). Chromatin modifications and DNA repair: beyond double-strand breaks. Front. Genet. *5*, 296.

Kim, J.C., Harris, S.T., Dinter, T., Shah, K.A., and Mirkin, S.M. (2017). The role of break-induced replication in large-scale expansions of (CAG)n/(CTG)n repeats. Nat. Struct. Mol. Biol. *24*, 55–60.

Krogh, B.O., and Symington, L.S. (2004). Recombination proteins in yeast. Annu. Rev. Genet. *38*, 233–271.

Lazzaro, F., Sapountzi, V., Granata, M., Pellicioli, A., Vaze, M., Haber, J.E., Plevani, P., Lydall, D., and Muzi-Falconi, M. (2008). Histone methyltransferase Dot1 and Rad9 inhibit single-stranded DNA accumulation at DSBs and uncapped telomeres. EMBO J. *27*, 1502–1512.

Lee, S.-E., Moore, J.K., Holmes, A., Umezu, K., Kolodner, R.D., and Haber, J.E. (1998). Saccharomyces Ku70, mre11/rad50 and RPA proteins regulate adaptation to G2/M arrest after DNA damage. Cell *94*, 399–409.

Lengsfeld, B.M., Rattray, A.J., Bhaskara, V., Ghirlando, R., and Paull, T.T. (2007). Sae2 is an endonuclease that processes hairpin DNA cooperatively with the Mre11/Rad50/Xrs2 complex. Mol. Cell *28*, 638–651.

Liao, S., Tammaro, M., and Yan, H. (2016). The structure of ends determines the pathway choice and Mre11 nuclease dependency of DNA double-strand break repair. Nucleic Acids Res. *44*, 5689–5701.

Liu, G., Chen, X., Bissler, J.J., Sinden, R.R., and Leffak, M. (2010). Replication-dependent instability at (CTG) x (CAG) repeat hairpins in human cells. Nat. Chem. Biol. *6*, 652–659.

Llorente, B., and Symington, L.S. (2004). The Mre11 nuclease is not required for 5′ to 3′ resection at multiple HO-induced double-strand breaks. Mol. Cell. Biol. *24*, 9682–9694.

Lobachev, K.S., Gordenin, D.A., and Resnick, M.A. (2002). The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. Cell *108*, 183–193.

Lydeard, J.R., Jain, S., Yamaguchi, M., and Haber, J.E. (2007). Break-induced replication and telomerase-independent telomere maintenance require Pol32. Nature *448*, 820–823.

Manley, K., Shirley, T.L., Flaherty, L., and Messer, A. (1999). Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. Nat. Genet. *23*, 471–473.

McMurray, C.T. (1999). DNA secondary structure: a common and causative factor for expansion in human disease. Proc. Natl. Acad. Sci. USA *96*, 1823–1825.

Millot, G. (2011). Comprendre et réaliser les tests statistiques à l'aide de R (de boeck).

Mimitou, E.P., and Symington, L.S. (2008). Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. Nature *455*, 770–774.

Mittelman, D., Moye, C., Morton, J., Sykoudis, K., Lin, Y., Carroll, D., and Wilson, J.H. (2009). Zinc-finger directed double-strand breaks within CAG repeat tracts promote repeat instability in human cells. Proc. Natl. Acad. Sci. USA *106*, 9607–9612.

Moynahan, M.E., Chiu, J.W., Koller, B.H., and Jasin, M. (1999). Brca1 controls homology-directed DNA repair. Mol. Cell *4*, 511–518.

Moynahan, M.E., Pierce, A.J., and Jasin, M. (2001). BRCA2 is required for homology-directed repair of chromosomal breaks. Mol. Cell *7*, 263–272.

O'Hoy, K.L., Tsilfidis, C., Mahadevan, M.S., Neville, C.E., Barceló, J., Hunter, A.G., and Korneluk, R.G. (1993). Reduction in size of the myotonic dystrophy trinucleotide repeat mutation during transmission. Science *259*, 809–812.

Orr, H.T., and Zoghbi, H.Y. (2007). Trinucleotide repeat disorders. Annu. Rev. Neurosci. *30*, 575–621.

Orr-Weaver, T.L., Szostak, J.W., and Rothstein, R.J. (1981). Yeast transformation: a model system for the study of recombination. Proc. Natl. Acad. Sci. USA *78*, 6354–6358.

Owen, B.A., Yang, Z., Lai, M., Gajec, M., Badger, J.D., 2nd, Hayes, J.J., Edelmann, W., Kucherlapati, R., Wilson, T.M., and McMurray, C.T. (2005). (CAG)(n)-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. Nat. Struct. Mol. Biol. *12*, 663–670.

Park, C.-Y., Halevy, T., Lee, D.R., Sung, J.J., Lee, J.S., Yanuka, O., Benvenisty, N., and Kim, D.-W. (2015). Reversion of FMR1 Methylation and Silencing by Editing the Triplet Repeats in Fragile X iPSC-Derived Neurons. Cell Rep. *13*, 234–241.

Parsons, C.A., Baumann, P., Van Dyck, E., and West, S.C. (2000). Precise binding of single-stranded DNA termini by human RAD52 protein. EMBO J. *19*, 4175–4181.

Pearson, C.E., Ewel, A., Acharya, S., Fishel, R.A., and Sinden, R.R. (1997). Human MSH2 binds to trinucleotide repeat DNA structures associated with neurodegenerative diseases. Hum. Mol. Genet. *6*, 1117–1123.

Pinto, R.M., Dragileva, E., Kirby, A., Lloret, A., Lopez, E., St Claire, J., Panigrahi, G.B., Hou, C., Holloway, K., Gillis, T., et al. (2013). Mismatch repair genes Mlh1 and Mlh3 modify CAG instability in Huntington's disease mice: genome-wide and candidate approaches. PLoS Genet. *9*, e1003930.

Plessis, A., Perrin, A., Haber, J.E., and Dujon, B. (1992). Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. Genetics *130*, 451–460.

Richard, G.F. (2015). Shortening trinucleotide repeats using highly specific endonucleases: a possible approach to gene therapy? Trends Genet. *31*, 177–186.

Richard, G.-F., Dujon, B., and Haber, J.E. (1999). Double-strand break repair can lead to high frequencies of deletions within short CAG/CTG trinucleotide repeats. Mol. Gen. Genet. *261*, 871–882.

Richard, G.-F., Goellner, G.M., McMurray, C.T., and Haber, J.E. (2000). Recombination-induced CAG trinucleotide repeat expansions in yeast involve the MRE11-RAD50-XRS2 complex. EMBO J. *19*, 2381–2390.

Richard, G.-F., Cyncynatus, C., and Dujon, B. (2003). Contractions and expansions of CAG/CTG trinucleotide repeats occur during ectopic gene conversion in yeast, by a MUS81-independent mechanism. J. Mol. Biol. *326*, 769–782.

Richard, G.F., Viterbo, D., Khanna, V., Mosbach, V., Castelain, L., and Dujon, B. (2014). Highly specific contractions of a single CAG/CTG trinucleotide repeat by TALEN in yeast. PLoS ONE *9*, e95611.

Ristic, D., Modesti, M., Kanaar, R., and Wyman, C. (2003). Rad52 and Ku bind to different DNA structures produced early in double-strand break repair. Nucleic Acids Res. *31*, 5229–5237.

Santillan, B.A., Moye, C., Mittelman, D., and Wilson, J.H. (2014). GFP-based fluorescence assay for CAG repeat instability in cultured human cells. PLoS ONE *9*, e113952.

Savouret, C., Garcia-Cordier, C., Megret, J., te Riele, H., Junien, C., and Gourdon, G. (2004). MSH2-dependent germinal CTG repeat expansions are produced continuously in spermatogonia from DM1 transgenic mice. Mol. Cell. Biol. *24*, 629–637.

Shinohara, A., Ogawa, H., and Ogawa, T. (1992). Rad51 protein involved in repair and recombination in S. cerevisiae is a RecA-like protein. Cell *69*, 457–470.

Slean, M.M., Panigrahi, G.B., Castel, A.L., Tomkinson, A.E., and Pearson, C.E. (2016). Absence of MutSβ leads to the formation of slipped-DNA for CTG/CAG contractions at primate replication forks. DNA Repair *42*, 107–118.

Sugawara, N., and Haber, J.E. (1992). Characterization of double-strand break-induced recombination: homology requirements and single-stranded DNA formation. Mol. Cell. Biol. *12*, 563–575.

Sundararajan, R., Gellon, L., Zunder, R.M., and Freudenreich, C.H. (2010). Double-strand break repair pathways protect against CAG/CTG repeat expansions, contractions and repeat-mediated chromosomal fragility in Saccharomyces cerevisiae. Genetics *184*, 65–77.

Sung, P. (1994). Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. Science *265*, 1241–1243.

Tian, L., Hou, C., Tian, K., Holcomb, N.C., Gu, L., and Li, G.-M. (2009). Mismatch recognition protein MutSbeta does not hijack (CAG)n hairpin repair in vitro. J. Biol. Chem. *284*, 20452–20456.

Tomé, S., Holt, I., Edelmann, W., Morris, G.E., Munnich, A., Pearson, C.E., and Gourdon, G. (2009). MSH2 ATPase domain mutation affects CTG*CAG repeat instability in transgenic mice. PLoS Genet. *5*, e1000482.

Tomé, S., Manley, K., Simard, J.P., Clark, G.W., Slean, M.M., Swami, M., Shelbourne, P.F., Tillier, E.R., Monckton, D.G., Messer, A., and Pearson, C.E. (2013). MSH3 polymorphisms and protein levels affect CAG repeat instability in Huntington's disease mice. PLoS Genet. *9*, e1003280.

Valencia, M., Bentele, M., Vaze, M.B., Herrmann, G., Kraus, E., Lee, S.-E., Schär, P., and Haber, J.E. (2001). NEJ1 controls non-homologous end joining in Saccharomyces cerevisiae. Nature *414*, 666–669.

van Agtmaal, E.L., André, L.M., Willemse, M., Cumming, S.A., van Kessel, I.D.G., van den Broek, W.J.A.A., Gourdon, G., Furling, D., Mouly, V., Monckton, D.G., et al. (2017). CRISPR/Cas9-Induced (CTG•CAG)n Repeat Instability in the Myotonic Dystrophy Type 1 Locus: Implications for Therapeutic Genome Editing. Mol. Ther. *25*, 24–43.

Van Dyck, E., Stasiak, A.Z., Stasiak, A., and West, S.C. (1999). Binding of double-strand breaks in DNA by human Rad52 protein. Nature *398*, 728–731.

Villarreal, D.D., Lee, K., Deem, A., Shim, E.Y., Malkova, A., and Lee, S.E. (2012). Microhomology directs diverse DNA break repair pathways and chromosomal translocations. PLoS Genet. *8*, e1003026.

Viterbo, D., Michoud, G., Mosbach, V., Dujon, B., and Richard, G.-F. (2016). Replication stalling and heteroduplex formation within CAG/CTG trinucleotide repeats by mismatch repair. DNA Repair (Amst.) *42*, 94–106.

White, C.I., and Haber, J.E. (1990). Intermediates of recombination during mating type switching in Saccharomyces cerevisiae. EMBO J. *9*, 663–673.

Williams, G.M., and Surtees, J.A. (2015). MSH3 Promotes Dynamic Behavior of Trinucleotide Repeat Tracts In Vivo. Genetics *200*, 737–754.

Wilson, T.E., Grawunder, U., and Lieber, M.R. (1997). Yeast DNA ligase IV mediates non-homologous DNA end joining. Nature *388*, 495–498.

Xie, N., Gong, H., Suhl, J.A., Chopra, P., Wang, T., and Warren, S.T. (2016). Reactivation of FMR1 by CRISPR/Cas9-Mediated Deletion of the Expanded CGG-Repeat of the Fragile X Chromosome. PLoS ONE *11*, e0165499.

Ye, Y., Kirkham-McCarthy, L., and Lahue, R.S. (2016). The Saccharomyces cerevisiae Mre11-Rad50-Xrs2 complex promotes trinucleotide repeat expansions independently of homologous recombination. DNA Repair (Amst.) *43*, 1–8.

Yu, A., Dill, J., and Mitas, M. (1995a). The purine-rich trinucleotide repeat sequences d(CAG)15 and d(GAC)15 form hairpins. Nucleic Acids Res. *23*, 4055–4057.

Yu, A., Dill, J., Wirth, S.S., Huang, G., Lee, V.H., Haworth, I.S., and Mitas, M. (1995b). The trinucleotide repeat sequence d(GTC)15 adopts a hairpin conformation. Nucleic Acids Res. *23*, 2706–2714.

Zhu, Z., Chung, W.H., Shim, E.Y., Lee, S.E., and Ira, G. (2008). Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. Cell *134*, 981–994.