HAL
open science

Replication stalling and heteroduplex formation within CAG/CTG trinucleotide repeats by mismatch repair

David Viterbo, Grégoire Michoud, Valentine Mosbach, Bernard Dujon, Guy-Franck Richard

To cite this version:

David Viterbo, Grégoire Michoud, Valentine Mosbach, Bernard Dujon, Guy-Franck Richard. Replication stalling and heteroduplex formation within CAG/CTG trinucleotide repeats by mismatch repair. DNA Repair, 2016, 42, pp.94-106. 10.1016/j.dnarep.2016.03.002 . hal-01299733

HAL Id: hal-01299733

https://hal.sorbonne-universite.fr/hal-01299733v1

Submitted on 8 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ACCEPTED MANUSCRIPT

Replication Stalling and Heteroduplex Formation within CAG/CTG Trinucleotide Repeats by Mismatch Repair

David Viterbo[1,2,3], Grégoire Michoud[4], Valentine Mosbach[1,2,3], Bernard Dujon[1,2,3] and Guy-Franck Richard[1,2,3]

[1] Institut Pasteur, Unité de Génétique Moléculaire des Levures, Département Génomes & Génétique, 25 rue du Dr Roux, F-75015 Paris, France

[2] Sorbonne Universités, UPMC Univ Paris 06, IFD, 4 Place Jussieu, 75252 Paris Cedex 05, France

[3] CNRS, UMR3525, F-75015 Paris, France

[4] Biological and Environmental Sciences and Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal 23955, Saudi Arabia

Corresponding author: Guy-Franck Richard.

Tel: +33-1-40-61-34-54; Fax: +33-1-40-61-34-56; e-Mail: gfrichar@pasteur.fr

ACCEPTED MANUSCRIPT

Highlights

- CAG/CTG triplet repeats block replication in both orientations on a yeast chromosome
- Replication fork stalling depends on mismatch repair integrity
- Msh2p is enriched at CAG/CTG triplet repeats in both orientations
- *MSH2* overexpression favors or stabilizes the formation of heteroduplex molecules

ACCEPTED MANUSCRIPT

Abstract

Trinucleotide repeat expansions are responsible for at least two dozen neurological disorders. Mechanisms leading to these large expansions of repeated DNA are still poorly understood. It was proposed that transient stalling of the replication fork by the repeat tract might trigger slippage of the newly-synthesized strand over its template, leading to expansions or contractions of the triplet repeat. However, such mechanism was never formally proven. Here we show that replication fork pausing and CAG/CTG trinucleotide repeat instability are not linked, stable and unstable repeats exhibiting the same propensity to stall replication forks when integrated in a yeast natural chromosome. We found that replication fork stalling was dependent on the integrity of the mismatch-repair system, especially the Msh2p-Msh6p complex, suggesting that direct interaction of MMR proteins with secondary structures formed by trinucleotide repeats *in vivo,* triggers replication fork pauses. We also show by chromatin immunoprecipitation that Msh2p is enriched at trinucleotide repeat tracts, in both stable and unstable orientations, this enrichment being dependent on *MSH3* and *MSH6*. Finally, we show that overexpressing *MSH2* favors the formation of heteroduplex regions, leading to an increase in contractions and expansions of CAG/CTG repeat tracts during replication, these heteroduplexes being dependent on both *MSH3* and *MSH6*. These heteroduplex regions were not detected when a mutant *msh2-E768A* gene in which the ATPase domain was mutated was overexpressed. Our results unravel two new roles for mismatch-repair proteins: stabilization of heteroduplex regions and transient blocking of replication forks passing through such repeats. Both roles may involve direct interactions between MMR proteins and secondary structures formed by trinucleotide repeat tracts, although indirect interactions may not be formally excluded.

Keywords: Replication pause, Trinucleotide repeats, Mismatch repair, Heteroduplex

ACCEPTED MANUSCRIPT

1. Introduction

Trinucleotide repeat expansions are involved in more than two dozen human neurological disorders [1,2]. Expansion mechanisms in human cells have, so far, eluded our understanding, but experiments in model organisms led to conclusions that S-phase replication DNA synthesis, double-strand break repair or gap repair may lead to triplet repeat expansions [2–6]. Replication slippage was historically the first mechanism proposed to induce expansions [7–9]. Slippage was supposed to occur following replication fork stalling and restart, after dissociation/reassociation of the newly-synthesized strand from its template strand [10–12], due to transient formation of CAG/CTG hairpin structures [13–15]. Replication fork pauses were shown to occur within plasmid-borne CCG/CGG, CAG/CTG and GAA/TTC triplet repeats, in bacteria, yeast and human cells [16–20], whereas GAA/TTC repeats were shown to transiently stall replication forks when cloned in a yeast natural chromosome [21]. No proof of replication fork stalling due to chromosomal-borne CAG/CTG repeats was ever shown. Actually, when a $(CTG)_{55}$ triplet repeat was integrated in a yeast natural chromosome, no evidence for a pausing signal at this locus was detected by 2D gel electrophoresis [22]. In addition, although replication slippage models postulate that transient fork stalling is a prerequisite to strand slippage, no formal demonstration of such correlation was ever published.

It was formerly shown that human Msh2 binds *in vitro* to both CAG and CTG synthetic hairpins, binding a CAG oligonucleotide being slightly more efficient than binding a CTG oligonucleotide [23]. It was subsequently suggested that the Msh2-Msh3 complex binds CAG hairpins, inhibiting Msh3 ATPase activity and slowing down release of the complex from the hairpin [24], although this last point is still controversial [25]. It was subsequently shown that the human MSH2-MSH3 complex discriminates between repair-competent loops and repair-resistant loops, such as those formed by long CAG repeats [26]. More recently, it was shown

ACCEPTED MANUSCRIPT

that short CAG/CTG slip-out structures were efficiently repaired by human MutLα and MutSβ complexes [27,28], and to a lesser extent by MutSα [27]. In mice in which Msh2, Msh3, Pms2, Mlh1 or Mlh3 were reduced or abolished, CAG/CTG repeat expansions were decreased [29–37]. This is also true when Msh2 was partially or totally depleted in a mouse model for fragile X premutation [38]. In addition, the Msh2 ATPase domain (abolishing ATPase activity but not DNA-binding ability) was shown to be directly involved in trinucleotide repeat expansions, strongly suggesting that a functional MutS complex was indeed required for such expansions [35]. Finally, it was recently demonstrated that two or three extrahelical CAG or CTG triplets were sufficient to trigger MutLα activation, and depended on MutSβ function [39]. All these data point to a model in which MMR complexes recognize and bind secondary structures associated to CAG/CTG repeats, maybe stabilizing these structures, and that such binding leads to expansions, by a process requiring a functional MMR activity. In support of this model, it was shown that purified yeast Msh2p-Msh3p inhibits Rad27p cleavage at the base of repeat-containing 5' flaps, suggesting that this inhibition eventually leads to repeat expansions *in vivo* [40].

We reasoned that mismatch repair complexes could directly or indirectly interfer with replication by slowing it down or transiently stalling it at trinucleotide repeat tracts. To investigate the precise role of MMR proteins in replication fork stalling, we used an experimental system previously designed to study trinucleotide repeat replication in *Saccharomyces cerevisiae* [22]. We found that replication fork stalling within trinucleotide repeat tracts was dependent on the presence of *MSH2*, *MSH6* and *MLH1*. Chromatin immunoprecipitation of Msh2p showed an enrichment of this protein at the triplet repeat locus, as compared to flanking regions. This enrichment was decreased in *msh6*Δ and in *msh3*Δ cells. Finally, we present molecular evidence for the presence of heteroduplex regions

ACCEPTED MANUSCRIPT

at the triplet repeat locus. These heteroduplexes depend on *MSH2* overexpression and on the presence of both *MSH3* and *MSH6* genes. These data bring direct evidence for the existence of transient DNA-protein intermediates, *in vivo*, containing a slipped-strand DNA structure, suggesting that these transient intermediates might be triggering replication fork stalling, rather than secondary structures themselves.

2. Materials and Methods

2.1 Yeast strains

ARG2 was engineered by integration, in a *MAT*a haploid yeast strain, of a cassette containing trinucleotide repeats in one of two orientations and a selection marker (*TRP1*) [22]. Mutant alleles were completely deleted and replaced by the KANMX4 marker, using the classical gene "ends-out" replacement method [41]. Shortly, 70 nt long oligonucleotides were synthesized (Integrated DNA Technologies). The 5' part of each oligonucleotide (50 nt) was designed to be homologous to the 50 bp right upstream or downstream of the targeted gene. The 3' part of each oligonucleotide (20 nt) was homologous to regions upstream and downstream the KANMX4 cassette. The PCR product amplified on a KANMX4 template using these two oligonucleotides was transformed into yeast using the Lithium Acetate-PEG procedure, as previously described [42]. Gene replacement was verified by Southern blot and trinucleotide repeat size was also checked by Southern blot in correctly deleted transformants. All strains used in the present study are shown in Table 1.

ACCEPTED MANUSCRIPT

Table 1: Strains used in this study

Strain	Genotype	p*MSH2*(1)	Reference
GFY117	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *SUP4*-opal *arg2*Δ::(CTG)98-*TRP1*	-	(2)
GFY170	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *SUP4*-opal *arg2*Δ::(CAG)60-*TRP1* (3)	-	This work
GFY167	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *sup4*-opal *arg2*Δ::(CAG)120-*TRP1*	-	(2)
GFY176	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *lys2*Δ202 *ade2*-opal *SUP4*-opal *arg2*Δ::(CTG)Δ-*TRP1*	-	This work
GFY179	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *SUP4*-opal *arg2*Δ::(CTG)98-*TRP1 msh2*Δ::KANMX4	-	This work
GFY180	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *sup4*-opal *arg2*Δ::(CAG)98-*TRP1 msh2*Δ::KANMX4	-	This work
GFY181	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *SUP4*-opal *arg2*Δ::(CTG)98-*TRP1 msh2*Δ::KANMX4	+	This work
GFY182	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *sup4*-opal *arg2*Δ::(CAG)98-*TRP1 msh2*Δ::KANMX4	+	This work
GFY183	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *SUP4*-opal *arg2*Δ::(CTG)70-*TRP1 msh3*Δ::KANMX4	-	This work
GFY184	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *SUP4*-opal *arg2*Δ::(CTG)70-*TRP1 msh3*Δ::KANMX4	+	This work
GFY185	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *SUP4*-opal *arg2*Δ::(CTG)98-*TRP1 msh6*Δ::KANMX4	-	This work
GFY186	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *SUP4*-opal *arg2*Δ::(CTG)98-*TRP1 msh6*Δ::KANMX4	+	This work
GFY187	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *SUP4*-opal *arg2*Δ::(CTG)98-*TRP1 mlh1*Δ::KANMX4	-	This work
GFY188	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *SUP4*-opal *arg2*Δ::(CTG)98-*TRP1 MSH2*-TAP	-	This work
GFY189	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *SUP4*-opal *arg2*Δ::(CTG)70-*TRP1 msh3*Δ::KANMX4 *MSH2*-TAP	-	This work
GFY190	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *SUP4*-opal *arg2*Δ::(CTG)98-*TRP1 msh6*Δ::KANMX4 *MSH2*-TAP	-	This work
GFY191	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *sup4*-opal *arg2*Δ::(CAG)98-*TRP1 MSH2*-TAP	-	This work
GFY192	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *lys2*Δ202 *ade2*-opal *SUP4*-opal *arg2*Δ::(CTG)Δ-*TRP1 MSH2*-TAP	-	This work
GFY193	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *sup4*-opal *arg2*Δ::(CAG)98-*TRP1 msh3*Δ::KANMX4 *MSH2*-TAP	-	This work
GFY194	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *sup4*-opal *arg2*Δ::(CAG)98-*TRP1 msh6*Δ::KANMX4 *MSH2*-TAP	-	This work
GFY195	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *sup4*-opal *arg2*Δ::(CAG)98-*TRP1 msh3*Δ::KANMX4	-	This work
GFY196	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *sup4*-opal *arg2*Δ::(CAG)98-*TRP1 msh6*Δ::KANMX4	-	This work
GFY197	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *SUP4*-opal *arg2*Δ::(CTG)98-*TRP1 msh2*Δ::KANMX4	+ (4)	This work
GFY198	*MAT*a *ura3*Δ851 *leu2*Δ1 *his3*Δ200 *trp1*Δ63 *ade2*-opal *sup4*-opal *arg2*Δ::(CAG)98-*TRP1 msh2*Δ::KANMX4	+ (4)	This work

(1) Indicates whether strain contains *MSH2*-overexpressing plasmid (+) or not (-)

ACCEPTED MANUSCRIPT

(2) Kerrest *et al.* (2009) *SRS2* and *SGS1* prevent chromosomal breaks and stabilize triplet repeats by restraining recombination. Nature Structural and Molecular Biology 16: 159-167
(3) This strain is a subclone of GFY167 in which the CAG trinucleotide repeat tract was partially contracted
(4) Both strains were transformed with p*msh2-E768A* mutant *MSH2* gene

ACCEPTED MANUSCRIPT

2.2 MSH2 and msh2-E768A overexpression

The wild-type *MSH2* gene was PCR amplified and cloned into pCMha190, a derivative of the multicopy pCM190 plasmid [43], carrying the *URA3* selectable marker. The gene was placed under the control of a doxycyclin-regulated $(TetO)_7$ promoter and tagged in its N-terminus region with the HA tag. Expression of the Msh2 protein was checked by Western blot, using an anti-HA antibody ([12CA5], Abcam). The *msh2-E768A* mutation was ordered as a 1380 bp DNA sequence (Integrated DNA Technologies) containing the point mutation, and subcloned into the pCMha*MSH2* plasmid to replace the wild-type sequence. Wild-type *MSH2* and mutant *msh2-E768A* genes were resequenced after cloning in pCMha190. Plasmid transformation was performed on SC-URA plates supplemented with 10 μg/ml doxycyclin. Trinucleotide repeat size was verified by Southern blot after transformation, on several subcloned transformants. Subsequent cultures for instability experiments were performed in the absence of doxycyclin. Each verified subclone was grown overnight and diluted to an appropriate concentration before plating on SC-URA supplemented with 10 μg/ml doxycyclin. The level of Msh2 protein was quantified on Western blots. Liquid cultures were grown to exponential phase in the presence or absence of doxycyclin. Proteins were extracted on 3 x 10^8 cells in 200 μl Laemmli solution with 100 μl glass beads. Proteins were separated on a 10% acrylamide gel in standard conditions and blotted to a nitrocellulose membrane (Optitran BA-S 83 reinforced NC, Schleicher & Schuell). Protein load was checked after blotting by red Ponceau staining. 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). A secondary goat anti-rabbit antibody conjugated to horseradish peroxidase was used for detection (Thermo Scientific, 0.16 μg/ml final concentration). Quantification was performed using a ChemiDoc MP Imager (Bio-Rad) with the dedicated Image Lab

ACCEPTED MANUSCRIPT

software. The molecular weight marker used was the Precision Plus Protein Standards All Blue (Bio-Rad).

2.3 Analysis of CAG/CTG repeat instability

Total genomic DNA was extracted from individual yeast colonies grown overnight in YPD medium. *Eco*RI and *Pst*I were used to digest genomic DNA of strains containing repeats in the CTG orientation, *Eco*RI and *Xba*I were used in the CAG orientation (Figure 1). Hybridization was performed with a probe covering *ARG2*, labeled by random priming [44]. Southern blots are exposed for a fixed amount of time (2-3 days) on a photostimulated phosphor plate (Fuji) before analysis. The image was treated with Adobe Photoshop (CS3) so that signal over background levels allow to visualize the most intense bands. When this was done, one or more than one band corresponding to triplet repeats was (were) visible. Normal sizes, contractions and expansions were then scored and recorded. If signal over background levels were modified to be more sensitive (increasing background level), more bands were detected and sometimes a smear corresponding to multiple repeat tract lengths within the same colony was visible. These extra bands correspond to instabilities occuring after the very first generation of colony growth and were not recorded. Heteroduplexes corresponded to double bands that have comparable intensities and were therefore infered to result from double-stranded DNA carrying two different informations that segregated at the very first division of the future colony. Fisher exact tests were performed with the 'R' package [45].

2.4 Analysis of replication fork pauses by 2D gels

*MAT*a yeast cells were arrested in G1 with alpha factor, washed and released in fresh medium at 23°C. Progression of S phase was followed by microscope observation and confirmed by FACS analysis on a MACSQuant Analyser (not shown). Time points were collected at 30',

ACCEPTED MANUSCRIPT

40', 60' and 90' after release for wild-type cells, or 40', 50', 60' and 90' for mutant strains and DNA was extracted by the CTAB method [46], with slight modifications to increase DNA yield. Shortly, cells were frozen overnight at -80°C, thawed on ice, resuspended in 2 ml water and 2.5 ml solution I (2% CTAB, 1.4 M NaCl, 100 mM Tris HCl pH 7.6, 25 mM EDTA pH 8.0). Subsequently, 167 µl Zymolyase 100T (Seikagaku, 30 mg/ml) and 5 µl DTT (2M) were added and tubes were left 1 hour at 30°C. After that step, extraction and gel set up were performed as described in [46]. Approximately 20 µg DNA were digested by *Cla*I and hybridized as described before [22]. Hybridization was performed with a probe covering the left part of the *ARG2* gene, labeled by random priming [44]. Replication pausing signals and Y arc signals were quantified on a Fujifilm FLA-9000, as in [16] with slight modifications. Only 2D gels whose Y arc signal was above background signals by 20% or more were kept for further calculation. Pausing index (P) were calculated as the ratio of pausing signal over Y arc signal, compared to their respective areas (Figure 3, right). Pausing indexes above 1 indicate a more intense signal at the trinucleotide repeat locus than on the average of the Y arc. Wilcoxon tests were performed with the 'R' package [45].

2.5 Chromatin immunoprecipitation

Msh2p was tagged in its C-terminal part using the TAP tag [47], amplified by PCR from the Thermo Scientific yeast TAP-fusion library. Correct integration was verified by PCR and protein expression was checked by Western blot with anti-protein A antibody (P3775, Sigma-Aldrich). Correct function of the *MSH2*-TAP construct was verified as follows. The yeast genome contains an (ATT)35 natural microsatellite, which was amplified using primers ATT35f and ATT35r (Supplementary Table 1) in wild-type, *msh2*Δ, *msh2*Δ +pMSH2 and TAP-tagged strains and PCR products were sequenced. Microsatellite instability was detected in the *msh2*Δ strain but not in the three others, confiming the proper function of *MSH2*-TAP.

ACCEPTED MANUSCRIPT

Chromatin immunoprecipitation was performed as previously described [48]. Cells were grown overnight at 23°C in YPD and diluted the following morning to 4-5 10^6 cells/ml. After 4-6 hours of growth in YPD at 23°C, cells were fixed with formaldehyde (1% final, 20 minutes). After two washes in PBS, cells were broken with acid-washed glass beads in lysis buffer (50 mM HEPES pH 7.5, 140 mM NaCl, 5 mM EDTA, 0.1% NP40 and 0.1% Na Deoxycholate). Lysate was sonicated in a Covaris S220 (LGC Genomics) in tubes AFA (12x12 mm) using the following setup: Peak Incident Power: 140 Watts, Duty Factor: 5%, 200 cycles, 300 seconds. The average size of chromatin fragments obtained after sonication was 250 bp, and the distance separating "Up" and "Down" regions amplified by qPCR was larger (350 bp). The lysate was centrifuged for 5 minutes at 4°C and the supernatant was incubated with Dynabeads Protein G (Life Technologies) coated with anti-protein A antibody (P3775, Sigma-Aldrich), overnight at 4°C. After washes, crosslinking was reversed at 65°C overnight and chromatin was digested with proteinase K (Sigma-Aldrich). DNA was purified on Qiaquick columns (Qiagen) and resuspended in 50 µl water. Quantitative real-time PCR was performed on 4 µl of DNA, in triplicate for each couple of primers, using ABsolute SYBR Fluorescein Mix (Thermo Scientific), in a realplex Mastercycler (Eppendorf).

Table 2: Primers used in this study.

Name	Sequence	PCR product
UpCTGf	CTCGGTAGCCAAGTTGGTTTAAG#	70 bp
UpCTGr	GCATTCCCGGCTACAAGGA	
DownCTGf	GGGGGATCACAGACCATTTCTT	77 bp
DownCTGr	GTCGAACGCCCGATCTCAA	
UpCAGf	CCAAGTTGGTTTAAGGCGCA	79 bp
UpCAGr	GGGGGATCACAGACCATTTCT	
DownCAGf	CATTCCCGGCTACAAGGACG	114 bp
DownCAGr	TGCAAGTCTGGGAAGTGAATGG#	
JEM1f*	TGTGATTTGGCTGAGTTACAACG	88 bp
JEM1r*	AACTGCCCAGCGATCCATT	
ARG2f	AGACCCCGAACACTGTGATG	91 bp
ARG2r	TCTTGTCACGTAGTGCCCTC	
JEM1bisf*	TTTTTCCAAGATAACGGCCT	241 bp
JEM1bisr*	GAACTAGCACTAGGCGATAC	
ATT35f	GCTTTTCACGTCGAAAGGAAAA	250 bp
ATT35r	TGACCTTTACAATTTGATATTT	

ACCEPTED MANUSCRIPT

* Two couples of primers were tested to qPCR amplify the *JEM1* locus. Results obtained with both couples of primers were comparable.
# UpCTGf and DownCAGr were used as qPCR primers to amplify a 85 bp DNA fragment at the locus at which triplet repeats were completely deleted (strain GFY192, Table 1 and Figure 5).

3. Results

3.1 Replication fork stalling occurs in both stable and unstable orientations

In order to address the question of correlation between replication fork stalling and trinucleotide repeat instability, we took advantage of genetically engineered yeast strains previously used in a former study [22]. Triplet repeats of different lengths were integrated into yeast chromosome X, at the *ARG2* locus, in such a way that the nearest replication origin was located ca. 7 kb away, telomere-proximal to *ARG2*. CAG/CTG repeats were cloned in two opposite orientations, the CTG orientation (in which the lagging-strand template carries the CTG sequence), or the CAG orientation (in which Okazaki fragments carry the CTG sequence) (Figure 1). As previously shown [8,22], CAG/CTG repeats of comparable lengths show radically different stabilities in opposite orientations, since 42.9% (82/191) alleles in the CTG orientation exhibit a repeat contraction (Figure 2A), whereas only 15.2% (20/131) harbor a contraction in the CAG orientation (Figure 2B), this difference being very significant (Fisher exact test, p-value= 1.18×10^{-7}). In order to look for a possible correlation of repeat contractions with replication fork pausing, yeast strains containing short (around 60 triplets) or long (around 98 triplets) CAG/CTG repeats, cloned in both orientations, were arrested using alpha-factor, then released into S phase. DNA was collected at different time points after release, and analyzed by 2D gels [22]. If replication forks are paused at -or near- trinucleotide repeats, a bulge on the descending Y arc, where repeats are located, should be detected. Unexpectedly, both orientations stall replication forks with similar efficacies and pauses are more pronunced in S phase as compared to G2 phase, which could be expected (Figure 3). Note that a strain in which the trinucleotide repeat tract was precisely deleted

ACCEPTED MANUSCRIPT

without modifying the rest of the locus showed no detectable replication fork pause (CTGΔ strain, Figure 3). The conical shape visible above the Y arc, presumably corresponding to reversed replication forks was also quantified, as previously [22]. A weak inverse correlation was found between the amount of cone and the pausing signal (Pearson's correlation p-value: 0.097). This suggests that fork reversing may occur more frequently when forks are not stalled by repeats.

These experiments unambiguously show that chromosome-borne CAG/CTG trinucleotide repeats stall replication forks in both orientations, and concluded that there is no direct correlation between replication fork stalling and triplet repeat contractions. Therefore, repeat contractions must arise by a mechanism different from replication fork stalling. Since no expansion was detected in the CTG orientation, and only one in the CAG orientation, it was not possible to address the effect of replication pausing on triplet repeat expansions. This low level of expansions is similar to what was observed by other authors in recent publications [40,49].

3.2 Replication fork stalling depends on the integrity of the mismatch-repair machinery

Then, we reasoned that transient binding of mismatch-repair proteins to CAG and CTG slipped-strand structures *in vivo*, rather than structures alone, could promote replication fork stalling. *MSH2* was therefore deleted in yeast strains harboring long CAG/CTG repeats, and replicating DNA was analyzed by 2D gels, as previously. Results show that replication fork pausing is decreased in the CTG orientation (Figure 4, Wilcoxon p-value= 0.06), but surprisingly not in the CAG orientation (Wilcoxon p-value= 0.6). In order to find out which subcomplex of the mismatch-repair machinery was responsible for replication pausing, *MSH3*, *MSH6*, and *MLH1* were individually deleted in the CTG orientation. In *msh6*Δ and *mlh1*Δ strains, replication pausing is reduced, although more drastically in *msh6*Δ (Wilcoxon

ACCEPTED MANUSCRIPT

p-value= 0.005) than in *mlh1*Δ (p-value= 0.01). In *msh3*Δ, pausing is also reduced but not statistically different from wild type (p-value= 0.13). We therefore concluded that MutSα and the Mlh1-Pms1 complexes contribute to replication fork stalling at CTG trinucleotide repeats.

3.3 MutS complexes promote CAG/CTG heteroduplex formation or stabilization

In order to try to increase replication pausing, *MSH2* was cloned on a multicopy plasmid, under the control of the doxycyclin-regulated $(TetO)_7$ promoter [43]. *msh2*Δ strains containing CAG/CTG repeats were transformed with this construct (subsequently called p*MSH2*), and replicating DNA was analyzed by 2D gels. In both orientations, pausing signals are not increased as compared to *msh2*Δ strains. It must be pointed out however, that strains harboring p*MSH2* grew slower, were harder to synchronize than other strains, and DNA extracted from such strains often showed artefactual signals on 2D gels. In these strains, cell cycle was delayed in such a way that 2.5 hours after alpha-factor release, S phase was still not completed, whereas in the normal situation it was completed 2 hours after release. Therefore, it is possible that this apparent lack of effect on pausing is due to some cell cycle dysregulation by *MSH2* overexpression.

However, subsequent analyses of repeat tract lengths in yeast colonies containing p*MSH2*, showed an unexpected phenotype. In both orientations, contractions dramatically increased, going from 42.9% to 74.1% of repeat tract lengths in the CTG orientation (Figure 2A, Fisher exact test, p-value= 6.50×10^{-11}), and from 15.2% to 65.7% of repeat tract lengths in the CAG orientation (Figure 2B, Fisher exact test, p-value< 2.20×10^{-16}). By comparison, in *msh2*Δ cells there is no difference in stability as compared to wild-type cells, in the CTG orientation (Fisher exact test, p-value= 0.57), but a significant reduction in contraction rate was detected in the CAG orientation (Figure 2B, p-value= 7.59×10^{-4}). Careful examination of DNA from colonies overexpressing *MSH2* showed that most of these contractions were found within

ACCEPTED MANUSCRIPT

colonies exhibiting at least two alleles (sometimes three or four) of different lengths and of comparable intensities (Figure 1). Sectored colonies, containing two different alleles, are classically attributed to non-repaired heteroduplexes occuring during homologous recombination (reviewed in [50]). They are frequently detected within trinucleotide repeat tracts, both during meiotic and mitotic recombination [5,15]. Here, sectored colonies dramatically increase when *MSH2* is overexpressed, going from 4.4% to 76.7% of colonies in the CTG orientation (Fisher exact test, p-value< 2.20×10^{-16}), and from 1.6% to 35.7% of colonies in the CAG orientation (p-value= 1.56×10^{-10}). Sectored colonies also significantly increased when *MSH2* was overexpressed in a wild-type strain (4.4% to 18.1%, Fisher exact test, p-value=7.95×10^{-5}), although not quite to the level of the *msh2*Δ strain overexpressing the same plasmid. It is interesting to note that contractions are increased in both orientations, whereas expansions are only increased in the CAG orientation (Fisher exact test, p-value= 1.07×10^{-3}). This result suggests that overexpressing *MSH2* tends to deregulate trinucleotide repeat stability perhaps by helping formation and/or stabilization of heteroduplex regions containing slipped-stranded structures, thus leading to unrepaired molecules that may ultimately be resolved as two DNA molecules of different lengths. However, it is peculiar that this effect is observed without co-overexpression of *MSH3* or *MSH6*. It is therefore possible that *MSH2* overexpression by itself may deregulate the balance of MutS subcomplexes assembly within cells or titrate some essential replication factor like PCNA, indirectly leading to the observed effect.

It is striking that in some sectored colonies, none of the two bands corresponds in size to the trinucleotide repeat parental size (Figure 1, lanes labeled with an asterisk). These colonies do not correspond to classical heteroduplex molecules in which one strand exhibits the parental allele and the other strand a contracted (or expanded) allele. This observation is reminiscent of the unconventional "rearranged heteroduplexes" described during meiotic recombination of

ACCEPTED MANUSCRIPT

a human minisatellite integrated in the yeast genome [51], and suggests that similar complex rearrangements occur at the trinucleotide repeat tract when the mismatch-repair machinery is perturbed. Contracted allele lengths detected in colonies into which one allele was parental (classical heteroduplex) and contracted allele lengths in colonies into which none of the alleles was parental (complex heteroduplex) were compared. In the CTG orientation no significant difference was observed, most contractions detected in both types of colonies being rather small (Figure 2C). However, in the CAG orientation contractions were larger, particularly in classical heteroduplexes (Figure 2C, one parental band) and both distributions were different. This confirms that classical heteroduplex and complex heteroduplex probably occur by different mechanisms, at least in the CAG orientation.

In order to identify whether a MutS complex was specifically involved in heteroduplex formation, we transformed p*MSH2* in *msh3*Δ and *msh6*Δ strains in the CTG orientation, and analyzed trinucleotide repeat lengths. In the *msh6*Δ strain, levels of both contractions and sectored colonies dramatically dropped, and were not statistically different from those detected in the wild-type strain (Fisher exact test respective p-values= 0.42 and 0.37). In the *msh3*Δ strain, the level of contractions is significantly reduced below wild-type levels (17.2%, Fisher exact test, p-value= 1.39×10^{-3}) and sectored colonies drop to 3.1% of analyzed colonies, a proportion similar to wild type (Figure 2A). Note that both *msh3*Δ and *msh3*Δ +p*MSH2* strains carry a small contraction of the repeat tract (70 instead of 98 triplets) in the CTG orientation. However, there is no statistical difference in sectored colonies between these two strains.

This proves that both MutSα and MutSβ are involved in formation and/or stabilization of heteroduplex regions. As a control, deletions of *MSH3* and *MSH6* by themselves did not increase the proportion of sectored colonies in any orientation (Figure 2). However, in the

ACCEPTED MANUSCRIPT

CAG orientation, contractions were significantly decreased in *msh3*Δ and *msh6*Δ strains (Fisher exact test respective p-values= 0.005 and 0.015), to levels comparable to what was observed in the *msh2*Δ mutant.

In order to determine if *MSH2* ATPase activity was required for heteroduplex formation, we introduced a point mutation in the *MSH2* Walker B motif (Glu768->Ala768). This mutation does not affect heterodimerization nor the ability to bind mismatched DNA but dramatically reduces ATPase activity and results in mutation rates similar to the null allele [52,53]. The *msh2-E768A* gene was placed under the control of the same doxycyclin-regulated $(TetO)_7$ promoter and expressed as previously. In both orientations, the frequency of sectored colonies observed was similar to the frequency observed in wild-type strains (Figure 2). Interestingly, Tome and colleagues also showed that the *MSH2* ATPase domain was required for expansions in mice, suggesting that they occur through a process that is not solely dependent on DNA binding but requires a functional MMR system [35], an observation reminiscent of what was observed here for heteroduplex formation.

3.4 MutS complexes are enriched at the CAG/CTG trinucleotide repeat locus

In order to detect a possible direct interaction of MutS complexes with trinucleotide repeats *in vivo*, chromatin immunoprecipitation was used. The Msh2 protein was tagged at its C-terminal end with the TAP tag [47]. After checking that the tagged protein was properly expressed and functional (see Materials & Methods), ChIP experiments were performed in wild-type strains as well as in *msh3*Δ and *msh6*Δ mutants containing triplet repeats in both orientations. Designing reliable primers for real-time qPCR within long trinucleotide repeat tracts was impossible. Therefore, it was chosen to design two sets of primers, one upstream and one downstream of the trinucleotide repeat tract (Figure 5, "Up" and "Down" regions). In wild-type strains in both orientations, Msh2p binding was increased upstream and

ACCEPTED MANUSCRIPT

downstream the trinucleotide repeat tract (Figure 5). This binding was totally abolished in a *msh3*Δ mutant, in both orientations. In the *msh6*Δ mutant, Msh2 enrichment was lost in the CAG orientation but not in the CTG orientation, showing that MutSα accumulates only in the CAG orientation whereas MutSβ is present at repeat tracts in both orientations. This suggests that different secondary structures may be formed at CAG/CTG triplet repeats and are recognized by different MutS complexes, depending on the orientation of the repeat tract.

4. Discussion

4.1 Replication pauses occur within plasmidic and chromosomal CAG/CTG trinucleotide repeats

It was previously published that CAG/CTG trinucleotide repeats stall replication forks when they are carried by a bacterial [19] or yeast plasmid [18]. In the present work, we show for the first time that replication fork stalling occurs in both orientations at a chromosome-borne CAG/CTG trinucleotide repeat. Stalling is not linked to contractions, more stable repeats stalling forks as efficiently as unstable ones. In a former work, using a similar experimental set up but much shorter triplet repeats (~50-55 CTG triplets), it was not possible to detect a transient pausing signal by 2D gel [22]. This demonstrates that a minimal repeat size is required in order to stall replication forks in such a way that they may be detected by 2D gel.

Replication fork stallings happening at the Ter sequence in *Escherichia coli* and *Bacillus subtilis*, in yeast rDNA repeats, at the *Schizosaccharomyces pombe mat1* locus, or in the Epstein-Barr virus, are all due to DNA-protein complexes transiently arresting forks (reviewed in [54]) and sometimes leading to formation of a double-strand break (reviewed in [55]). Here, we propose that replication fork pausing at CAG/CTG trinucleotide repeats is due to transient association of such repeats with mismatch-repair proteins, possibly through interactions with DNA secondary structure.

ACCEPTED MANUSCRIPT

4.2 Mismatch-repair dependent trinucleotide repeat instability

We also show that contractions (and expansions in the CAG orientation) are increased when *MSH2* is overexpressed. In the *msh2*Δ mutant, the only significant effect detected was a decrease of contractions in the CAG orientation. Former reports showed small but detectable effects of mismatch-repair mutants on CAG/CTG instability in yeast. Schweitzer and Livingston [56] showed a clear effect of MMR mutants on repeat tract length, corresponding to additions/deletions of one repeat unit only, similar to the well-described microsatellite instability phenotype associated to mismatch-repair mutations and colorectal cancers [57,58]. Schmidt and colleagues [59] made a similar observation in bacteria. However, these authors did not score sectored colonies. Other authors found a 2-3 fold increase in contractions [60,61], or a 10-fold decrease in expansions [62] in a *msh2*Δ mutant, but sectored colonies were not reported. Therefore, the large increase in sectored colonies observed in the present work, when *MSH2* is overexpressed, could not be observed before and cannot be compared to former published results. In a recent report, small incremental CAG/CTG expansions were detected by individual tracking of repeat length in independent yeast colonies during 14 days time course experiments. The authors showed that these expansions were *MSH3* dependent [49]. It must also be noted that previous authors worked with shorter repeats than those used in the present study (25 triplets in [62], 30 triplets in [60], and 50 triplets in [61]), and it is most likely that triplet repeats need to reach a threshold length in order to form secondary structures bound by the MMR machinery. However, given that this sectoring phenotype was observed when *MSH2* alone was overexpressed (without co-expression of *MSH3* or *MSH6*) it is possible that the observed phenotype resulted from an indirect effect due to perturbations in MutS subcomplexes assembly or titration of some essential replication factor like PCNA.

ACCEPTED MANUSCRIPT

It would have been interesting to address the effect of *msh2-E768A* overexpression on replication fork pausing and binding to repeat tract. We predict that binding by itself should not be altered, therefore replication fork pausing should still be detected. However, strains overexpressing *MSH2* or its mutant allele are hard to synchronize and exhibit dysregulation of the cell cycle that make 2D gels hard to interpret. Therefore, other experiments need to be designed in order to address these specific questions.

4.3 Similarities and differences between mammalian and yeast models

Mapping of replication origins by nascent strand analysis in a mice model of myotonic dystrophy, showed that the expanded triplet repeat was replicated in the CAG orientation [63]. When the level of Msh2 or Msh3 was reduced in these same mice, expansions were reduced too. Here, when *MSH2* is deleted in the CAG orientation, there is no difference in the level of expansions observed (Figure 2B). However, when it is overexpressed, expansions are increased (Fisher exact test p-value= 1.074×10^{-3}), suggesting the existence of subtle differences in CAG/CTG repeat replication - or in MMR protein levels - between yeast and mammalian cells.

In human cells, MutSα is normally more abundant than MutSβ [64,65], unless *MSH3* is overexpressed in a cancer cell line, leading to an increase in MutSβ complexes [65]. In mice it is the opposite, since MSH3 is more abundant than MSH6 in many tissues, suggesting that MutSβ is more abundant than MutSα [36]. In yeast, Msh6p is about 7 times more abundant than Msh3p [66], suggesting that MutSα is more abundant than MutSβ. Given that Msh2p is twice as much abundant than needed to assemble the two complexes, it might be possible that some free uncomplexed Msh2p or Msh2p homodimers may exist within the cell [66]. It is worth noting that *MLH1* overexpression in yeast strongly increases spontaneous mutagenesis [67] whereas *MSH2* overexpression does not [68]. Furthermore, hMLH1 or hMSH2

ACCEPTED MANUSCRIPT

overexpression increases apoptosis in human cells [69] whereas the overexpression of yeast *MSH2* was not reported to affect viability [68], pointing to subtle differences between yeast and human MMR.

It is known that CAG/CTG trinucleotide repeats mainly contract in yeast cells in both orientations (although some expansions are detected in the CAG orientation), whereas there is a clear trend towards expansions in human and mice cells (reviewed in [2]). Chromatin structures of DNA in these two eukaryotic cell types share similarities as well as differences, and could certainly be one of the factors involved in this discrepancy. In human cells, the HDAC3 and HDAC5 histone deacetylase complexes are involved in CAG/CTG repeat expansions, and in yeast cells the HDAC3 homologue has the same effect [30,70]. However, it was shown that chromatin organization at trinucleotide repeats embedded in a yeast natural chromosome exhibit a noncanonical structure, involving binding of Hmo1p, a high mobility group protein without obvious orthologue in the human genome [71].

4.4 Evidence for secondary stuctures at CAG/CTG trinucleotide repeats in vivo

In a former work, Liu and colleagues [72] showed that induction of a CAG/CTG repeat-targeted nuclease in mammalian cells, induced the formation of molecules that were interpreted to be products of hairpin digestion by the nuclease, supporting the existence of such secondary structures *in vivo*. Very recently, Axford and colleagues [73] used an antibody directed against cruciform structures [74] to immunoprecipitate DNA extracted from myotonic dystrophy patient tissues. They showed that immunoprecipitated DNA was enriched in secondary structures, suggesting that such structures existed *in vivo*, in non-replicating DNA.

We propose a new model (Figure 6), in which slipped-stranded secondary structure formation may occur at CAG/CTG trinucleotide repeat tracts, probably happening more often on the

ACCEPTED MANUSCRIPT

lagging- or leading-strand templates than on newly-synthesized strands. Secondary structures may be recognized by MutSα following the replication fork during S phase. We postulate that the complex may not be released from the structure until the following S phase, leading to replication fork stalling in a few instances (6-9% of replication forks are stalled, based on pausing indexes). Stalling may occur through direct interaction of MutSα with the approaching replication machinery, or indirectly by signaling unrepaired damage to the cell. These rare events may occasionally lead to fork breakage, although double-strand break frequency is too low to be detected in our experimental set up, despite numerous attempts.

If *MSH2* is overexpressed, secondary structures may bind MutSα and/or MutSβ, stabilizing heteroduplex regions and leading to formation of transient hDNA resolved during the following S phase as two DNA molecules of different lengths, one of them giving rise to a contracted allele. In the rare cases in which hairpin formation occurs on the newly-synthesized strand, resolution of the heteroduplex region will give rise to an expansion, as seen in the CAG orientation. As an alternative hypothesis, given the large increase of contractions in both orientations when *MSH2* is overexpressed, it is possible that overexpression of Msh2p induces more slipped-stranded structures, thus generating more contractions. One may wonder how imperfect hairpins may survive another cell cycle without being processed by structure-specific nucleases, such as the Mre11-Rad50 complex [75]. It is however possible that MutSα/β binding to such hairpins protects DNA from degradation by these nucleases.

Finally, the effect of *MSH2* overexpression may not be due to a direct binding of MMR complexes to secondary structures but to partial titration of PCNA, reducing replication fork processivity, therefore increasing replication slippage, leading to contractions. The fact that S phase duration is increased when *MSH2* is overexpressed is in favor of this hypothesis, whereas Msh2p enrichment at trinucleotide repeats, as well as the necessary presence of

ACCEPTED MANUSCRIPT

Msh3p and Msh6p to observe sectored colonies do not support this hypothesis, although the two models (secondary structure interaction and PCNA interaction) are not formally exclusive from each other.

5. Conclusions

In the present work we show that there is no correlation between contractions and replication fork pausing within CAG/CTG trinucleotide repeats in yeast cells, suggesting that another mechanism is responsible for repeat contractions. In addition, we also show that the MMR complex, involved in detecting and repairing replication errors, is also involved in replication fork pausing within trinucleotide repeats. This complex contains the Msh2 protein and we show that this protein is found enriched at the trinucleotide repeat locus, this enrichment depending on two other proteins of the same complex, Msh3 and Msh6. Finally, trinucleotide repeat tract length analysis in several mutants of the MMR complex suggests that it triggers or stabilizes the formation of transient heteroduplex DNA regions containing different repeat lengths on each DNA strand. These discoveries will be helpful to understand processes underlying trinucleotide repeat instabilities in human cells.

Acknowledgements

G.-F. R. thanks C. Saveanu and F. Feuerbach for the use of the TAP-tag library and advices on ChIP.

Conflict of interest

The authors declare no conflict of interest regarding any material discussed in the present manuscript.

ACCEPTED MANUSCRIPT

References

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

[2] C.E. Pearson, K.N. Edamura, J.D. Cleary, Repeat instability: mechanisms of dynamic mutations, Nat. Rev. Genet. 6 (2005) 729–742.

[3] B.A. Lenzmeier, C.H. Freudenreich, Trinucleotide repeat instability: a hairpin curve at the crossroads of replication, recombination, and repair, Cytogenet. Genome Res. 100 (2003) 7–24.

[4] C.T. McMurray, Mechanisms of trinucleotide repeat instability during human development, Nat Rev Genet. 11 (2010) 786–99. doi:10.1038/nrg2828.

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

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

[7] J.M. Darlow, D.R.F. Leach, The effects of trinucleotide repeats found in human inherited disorders on palindrome inviability in Escherichia coli suggest hairpin folding preferences in vivo, Genetics. 141 (1995) 825–832.

[8] C.H. Freudenreich, J.B. Stavenhagen, V.A. Zakian, Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome, Mol Cell Biol. 17 (1997) 2090–2098.

[9] S. Kang, A. Jaworski, K. Ohshima, R.D. Wells, Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in E. coli, Nat. Genet. 10 (1995) 213–217.

[10] S.M. Mirkin, Toward a unified theory for repeat expansions, Nat Struct Mol Biol. 12 (2005) 635–7.

[11] S.M. Mirkin, DNA structures, repeat expansions and human hereditary disorders, Curr Opin Struct Biol. 16 (2006) 351–8.

[12] G.F. Richard, A. Kerrest, B. Dujon, Comparative genomics and molecular dynamics of DNA repeats in eukaryotes, Microbiol Mol Biol Rev. 72 (2008) 686–727.

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

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

[15] H. Moore, P.W. Greenwell, C.-P. Liu, N. Arnheim, T.D. Petes, Triplet repeats form secondary structures that escape DNA repair in yeast, Proc Natl Acad Sci USA. 96 (1999) 1504–1509.

[16] R.P. Anand, K.A. Shah, H. Niu, P. Sung, S.M. Mirkin, C.H. Freudenreich, Overcoming natural replication barriers: differential helicase requirements, Nucleic Acids Res. 40 (2012) 1091–105. doi:10.1093/nar/gkr836.

[17] H.M. Kim, V. Narayanan, P.A. Mieczkowski, T.D. Petes, M.M. Krasilnikova, S.M. Mirkin, et al., Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair, EMBO J. 27 (2008) 2896–906. doi:10.1038/emboj.2008.205.

[18] R. Pelletier, M.M. Krasilnikova, G.M. Samadashwily, R. Lahue, S.M. Mirkin, Replication and expansion of trinucleotide repeats in yeast, Mol Cell Biol. 23 (2003) 1349–57.

ACCEPTED MANUSCRIPT

[19] G. Samadashwily, G. Raca, S.M. Mirkin, Trinucleotide repeats affect DNA replication in vivo, Nat. Genet. 17 (1997) 298–304.
[20] I. Voineagu, C.F. Surka, A.A. Shishkin, M.M. Krasilnikova, S.M. Mirkin, Replisome stalling and stabilization at CGG repeats, which are responsible for chromosomal fragility, Nat Struct Mol Biol. 16 (2009) 226–8. doi:10.1038/nsmb.1527.
[21] A.A. Shishkin, I. Voineagu, R. Matera, N. Cherng, B.T. Chernet, M.M. Krasilnikova, et al., Large-scale expansions of Friedreich's ataxia GAA repeats in yeast, Mol Cell. 35 (2009) 82–92. doi:10.1016/j.molcel.2009.06.017.
[22] A. Kerrest, R. Anand, R. Sundararajan, R. Bermejo, G. Liberi, B. Dujon, et al., SRS2 and SGS1 prevent chromosomal breaks and stabilize triplet repeats by restraining recombination, Nat. Struct. Mol. Biol. 16 (2009) 159–167.
[23] C.E. Pearson, A. Ewel, S. Acharya, R.A. Fishel, R.R. Sinden, Human MSH2 binds to trinucleotide repeat DNA structures associated with neurodegenerative diseases, Hum Mol Genet. 6 (1997) 1117–1123.
[24] B.A. Owen, Z. Yang, M. Lai, M. Gajek, J.D. Badger 2nd, J.J. Hayes, et al., (CAG)(n)-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition, Nat Struct Mol Biol. 12 (2005) 663–70.
[25] L. Tian, C. Hou, K. Tian, N.C. Holcomb, L. Gu, G.M. Li, Mismatch recognition protein MutSbeta does not hijack (CAG)n hairpin repair in vitro, J Biol Chem. 284 (2009) 20452–6. doi:10.1074/jbc.C109.014977.
[26] W.H. Lang, J.E. Coats, J. Majka, G.L. Hura, Y. Lin, I. Rasnik, et al., Conformational trapping of mismatch recognition complex MSH2/MSH3 on repair-resistant DNA loops, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) E837–844. doi:10.1073/pnas.1105461108.
[27] G.B. Panigrahi, M.M. Slean, J.P. Simard, O. Gileadi, C.E. Pearson, Isolated short CTG/CAG DNA slip-outs are repaired efficiently by hMutSbeta, but clustered slip-outs are poorly repaired, Proc Natl Acad Sci U A. 107 (2010) 12593–8. doi:10.1073/pnas.0909087107.
[28] G.B. Panigrahi, M.M. Slean, J.P. Simard, C.E. Pearson, Human mismatch repair protein hMutLalpha is required to repair short slipped-DNAs of trinucleotide repeats, J Biol Chem. 287 (2012) 41844–50. doi:10.1074/jbc.M112.420398.
[29] L. Foiry, L. Dong, C. Savouret, L. Hubert, H. te Riele, C. Junien, et al., Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice, Hum. Genet. 119 (2006) 520–6. doi:10.1007/s00439-006-0164-7.
[30] A.M. Gannon, A. Frizzell, E. Healy, R.S. Lahue, MutSbeta and histone deacetylase complexes promote expansions of trinucleotide repeats in human cells, Nucleic Acids Res. 40 (2012) 10324–33. doi:10.1093/nar/gks810.
[31] M. Gomes-Pereira, M.T. Fortune, L. Ingram, J.P. McAbney, D.G. Monckton, Pms2 is a genetic enhancer of trinucleotide CAG.CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion, Hum Mol Genet. 13 (2004) 1815–25.
[32] K. Manley, T.L. Shirley, L. Flaherty, A. Messer, Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice, Nat Genet. 23 (1999) 471–473.
[33] C. Savouret, E. Brisson, J. Essers, R. Kanaar, A. Pastink, H. te Riele, et al., CTG repeat instability and size variation timing in DNA repair-deficient mice, Embo J. 22 (2003) 2264–73.
[34] C. Savouret, C. Garcia-Cordier, J. Megret, H. te Riele, C. Junien, G. Gourdon, MSH2-dependent germinal CTG repeat expansions are produced continuously in spermatogonia from DM1 transgenic mice, Mol. Cell. Biol. 24 (2004) 629–637.

ACCEPTED MANUSCRIPT

[35] S. Tome, I. Holt, W. Edelmann, G.E. Morris, A. Munnich, C.E. Pearson, et al., MSH2 ATPase domain mutation affects CTG*CAG repeat instability in transgenic mice, PLoS Genet. 5 (2009) e1000482. doi:10.1371/journal.pgen.1000482.
[36] S. Tome, K. Manley, J.P. Simard, G.W. Clark, M.M. Slean, M. Swami, et al., MSH3 Polymorphisms and Protein Levels Affect CAG Repeat Instability in Huntington's Disease Mice, PLoS Genet. 9 (2013) e1003280. doi:10.1371/journal.pgen.1003280.
[37] R.M. Pinto, E. Dragileva, A. Kirby, A. Lloret, E. Lopez, J. St Claire, et al., Mismatch Repair Genes Mlh1 and Mlh3 Modify CAG Instability in Huntington's Disease Mice: Genome-Wide and Candidate Approaches, PLoS Genet. 9 (2013) e1003930. doi:10.1371/journal.pgen.1003930.
[38] R.A. Lokanga, X.N. Zhao, K. Usdin, The mismatch repair protein MSH2 is rate limiting for repeat expansion in a fragile X premutation mouse model, Hum Mutat. 35 (2014) 129–36. doi:10.1002/humu.22464.
[39] A. Pluciennik, V. Burdett, C. Baitinger, R.R. Iyer, K. Shi, P. Modrich, Extrahelical (CAG)/(CTG) triplet repeat elements support proliferating cell nuclear antigen loading and MutLalpha endonuclease activation, Proc Natl Acad Sci U A. 110 (2013) 12277–82. doi:10.1073/pnas.1311325110.
[40] A. Kantartzis, G.M. Williams, L. Balakrishnan, R.L. Roberts, J.A. Surtees, R.A. Bambara, Msh2-Msh3 interferes with Okazaki fragment processing to promote trinucleotide repeat expansions, Cell Rep. 2 (2012) 216–222. doi:10.1016/j.celrep.2012.06.020.
[41] A. Baudin, O. Ozier-Kalogeropoulos, A. Denouel, F. Lacroute, C. Cullin, A simple and efficient method for direct gene deletion in Saccharomyces cerevisiae., Nucleic Acids Res. 21 (1993) 3329–3330.
[42] R.D. Gietz, R.H. Schiestl, A.R. Willems, R.A. Woods, Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure, Yeast. 11 (1995) 355–60.
[43] E. Gari, L. Piedrafita, M. Aldea, E. Herrero, A set of vectors with a tetracycline-regulatable promoter system for modulating gene expression in Saccharomyces cerevisiae, Yeast. 13 (1997) 837–848.
[44] A.P. Feinberg, B. Vogelstein, A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity, Anal Biochem. 132 (1983) 6–13.
[45] G. Millot, Comprendre et réaliser les tests statistiques à l'aide de R, 2nd ed., de boeck, Brussels, 2011.
[46] G. Liberi, C. Cotta-Ramusino, M. Lopes, J. Sogo, C. Conti, A. Bensimon, et al., Methods to study replication fork collapse in budding yeast, Methods Enzym. 409 (2006) 442–62.
[47] G. Rigaut, A. Shevchenko, B. Rutz, M. Wilm, M. Mann, B. Seraphin, A generic protein purification method for protein complex characterization and proteome exploration, Nat Biotechnol. 17 (1999) 1030–2. doi:10.1038/13732.
[48] V. Borde, W. Lin, E. Novikov, J.H. Petrini, M. Lichten, A. Nicolas, Association of Mre11p with double-strand break sites during yeast meiosis, Mol Cell. 13 (2004) 389–401. doi:S1097276504000346 [pii].
[49] G.M. Williams, J.A. Surtees, MSH3 Promotes Dynamic Behavior of Trinucleotide Repeat Tracts In Vivo, Genetics. 200 (2015) 737–54. doi:10.1534/genetics.115.177303.
[50] F. Pâques, J.E. Haber, Multiple Pathways of Recombination Induced by Double-Strand Breaks in Saccharomyces cerevisiae, Microbiol Mol Biol Rev. 63 (1999) 349–404.
[51] H. Debrauwère, J. Buard, J. Tessier, D. Aubert, G. Vergnaud, A. Nicolas, Meiotic instability of human minisatellite CEB1 in yeast requires double-strand breaks, Nat Genet. 23 (1999) 367–371.

ACCEPTED MANUSCRIPT

[52] E. Antony, S. Khubchandani, S. Chen, M.M. Hingorani, Contribution of Msh2 and Msh6 subunits to the asymmetric ATPase and DNA mismatch binding activities of Saccharomyces cerevisiae Msh2-Msh6 mismatch repair protein, DNA Repair Amst. 5 (2006) 153–62. doi:10.1016/j.dnarep.2005.08.016.
[53] K. Drotschmann, W. Yang, T.A. Kunkel, Evidence for sequential action of two ATPase active sites in yeast Msh2-Msh6, DNA Repair Amst. 1 (2002) 743–53. doi:S1568786402000812 [pii].
[54] E.V. Mirkin, S.M. Mirkin, Replication fork stalling at natural impediments, Microbiol. Mol. Biol. Rev. 71 (2007) 13–35.
[55] R. Rothstein, B. Michel, S. Gangloff, Replication fork pausing and recombination or "gimme a break," Genes Dev. 14 (2000) 1–10.
[56] J.K. Schweitzer, D.M. Livingston, Destabilization of CAG trinucleotide repeat tracts by mismatch repair mutations in yeast, Hum Mol Genet. 6 (1997) 349–355.
[57] R. Fishel, M.K. Lescoe, M.R.S. Rao, N.G. Copeland, N.A. Jenkins, J. Garber, et al., The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer, Cell. 75 (1993) 1027–1038.
[58] M. Strand, T.A. Prolla, R.M. Liskay, T.D. Petes, Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair, Nature. 365 (1993) 274–276.
[59] K.H. Schmidt, C.M. Abbott, D.R. Leach, Two opposing effects of mismatch repair on CTG repeat instability in Escherichia coli, Mol. Microbiol. 35 (2000) 463–471.
[60] G.-F. Richard, B. Dujon, J.E. Haber, Double-strand break repair can lead to high frequencies of deletions within short CAG/CTG trinucleotide repeats, Mol Gen Genet. 261 (1999) 871–882.
[61] J.J. Miret, L. Pessoa-Brandao, R.S. Lahue, Instability of CAG and CTG trinucleotide repeats in Saccharomyces cerevisiae, Mol Cell Biol. 17 (1997) 3382–3387.
[62] J.J. Miret, P.-B. L., R.S. Lahue, Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in Saccharomyces cerevisiae, Proc Natl Acad Sci USA. 95 (1998) 12438–12443.
[63] J.D. Cleary, S. Tome, A. Lopez Castel, G.B. Panigrahi, L. Foiry, K.A. Hagerman, et al., Tissue- and age-specific DNA replication patterns at the CTG/CAG-expanded human myotonic dystrophy type 1 locus, Nat Struct Mol Biol. 17 (2010) 1079–87. doi:10.1038/nsmb.1876.
[64] J. Genschel, S.J. Littman, J.T. Drummond, P. Modrich, Isolation of MutSβ from Human Cells and Comparison of the Mismatch Repair Specificities of MutSβ and MutSα, J. Biol. Chem. 273 (1998) 19895–19901. doi:10.1074/jbc.273.31.19895.
[65] G. Marra, I. Iaccarino, T. Lettieri, G. Roscilli, P. Delmastro, J. Jiricny, Mismatch repair deficiency associated with overexpression of the MSH3 gene, Proc Natl Acad Sci U A. 95 (1998) 8568–73.
[66] S. Ghaemmaghami, W.K. Huh, K. Bower, R.W. Howson, A. Belle, N. Dephoure, et al., Global analysis of protein expression in yeast, Nature. 425 (2003) 737–41. doi:10.1038/nature02046.
[67] P.V. Shcherbakova, M.C. Hall, M.S. Lewis, S.E. Bennett, K.J. Martin, P.R. Bushel, et al., Inactivation of DNA Mismatch Repair by Increased Expression of Yeast MLH1, Mol. Cell. Biol. 21 (2001) 940–951. doi:10.1128/MCB.21.3.940-951.2001.
[68] K. Drotschmann, A.B. Clark, H.T. Tran, M.A. Resnick, D.A. Gordenin, T.A. Kunkel, Mutator phenotypes of yeast strains heterozygous for mutations in the MSH2 gene, Proc. Natl. Acad. Sci. 96 (1999) 2970–2975. doi:10.1073/pnas.96.6.2970.
[69] H. Zhang, B. Richards, T. Wilson, M. Lloyd, A. Cranston, A. Thorburn, et al., Apoptosis Induced by Overexpression of hMSH2 or hMLH1, Cancer Res. 59 (1999) 3021–3027.

ACCEPTED MANUSCRIPT

[70] K. Debacker, A. Frizzell, O. Gleeson, L. Kirkham-McCarthy, T. Mertz, R.S. Lahue, Histone deacetylase complexes promote trinucleotide repeat expansions, PLoS Biol. 10 (2012) DOI: 10.1371/journal.pbio.1001257. doi:10.1371/journal.pbio.1001257.
[71] H. Kim, D.M. Livingston, A high mobility group protein binds to long CAG repeat tracts and establishes their chromatin organization in Saccharomyces cerevisiae, J Biol Chem. 281 (2006) 15735–40. doi:10.1074/jbc.M512816200.
[72] G. Liu, X. Chen, J.J. Bissler, R.R. Sinden, M. Leffak, Replication-dependent instability at (CTG) x (CAG) repeat hairpins in human cells, Nat Chem Biol. 6 (2010) 652–9. doi:10.1038/nchembio.416.
[73] M.M. Axford, Y.H. Wang, M. Nakamori, M. Zannis-Hadjopoulos, C.A. Thornton, C.E. Pearson, Detection of Slipped-DNAs at the Trinucleotide Repeats of the Myotonic Dystrophy Type I Disease Locus in Patient Tissues, PLoS Genet. 9 (2013) DOI: 10.1371/journal.pgen.1003866. doi:10.1371/journal.pgen.1003866.
[74] M. Tam, S. Erin Montgomery, M. Kekis, B.D. Stollar, G.B. Price, C.E. Pearson, Slipped (CTG).(CAG) repeats of the myotonic dystrophy locus: surface probing with anti-DNA antibodies, J Mol Biol. 332 (2003) 585–600. doi:S0022283603008805 [pii].
[75] K.S. Lobachev, D.A. Gordenin, M.A. Resnick, The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements, Cell. 108 (2002) 183–93.

ACCEPTED MANUSCRIPT

Figure Captions

Figure 1 A: Schematic representation of the *ARG2* locus containing CAG/CTG repeats in the CTG orientation. Arrow on the left (Rep.) indicates replication fork movement coming from ARS 1010 [22]. Restriction sites used for Southern blots are indicated (*Eco*RI and *Pst*I). Hybridization with a probe covering *ARG2* reveals three different signals: band A (1388 bp), band C (692 bp) and a variable band B containing the trinucleotide repeat tract. Below are shown representative Southern blots, in each strain background studied. Each lane contains total genomic DNA extracted from one independent yeast colony, grown for 25 generations, digested by *Eco*RI and *Pst*I. Positions of bands A, B and C, revealed by hybridization with the *ARG2* probe are indicated to the left. (CNG)n size markers show positions of PCR fragments hybridizing to the probe and used on each gel to estimated CAG or CTG trinucleotide repeat sizes. One millimeter on a gel corresponds to 15-23 base pairs (mean= 19 ± 1 bp), corresponding to 5-8 triplets, which is the best possible resolution on such gels. Sectored colonies, detected as lanes containing two or more bands of relatively similar intensities, are frequent when *MSH2* is overexpressed from the p*MSH2* plasmid, unless *MSH3* or *MSH6* are deleted. Lanes labeled with an asterisk (*****) correspond to colonies harboring complex events, i.e. colonies in which none of the two bands observed correspond to the repeat parental size, and which therefore do not correspond to classical heteroduplex molecules (see subsection 3.3). **B**: Western blot showing the level of Msh2p in the presence or absence of 10µg/ml doxycyclin in the culture medium, in the *msh2*Δ +p*MSH2* strain (GFY182) and in the *msh2*Δ +p*msh2-E768A* strain (GFY198), along with the corresponding wild-type strain before its transformation with one of the two plasmids (GFY167). Msh2p is 108.8 kD and migrates right above the 100 kD MW marker. Protein levels normalized to an internal standard and compared to GFY167 strain are indicated below each lane. Note that the anti-Msh2p antibody cross-reacts with another protein present in near-stochiometric amount as compared to Msh2p

ACCEPTED MANUSCRIPT

and migrating slightly faster (shown by bracket). It cannot be Msh3p or Msh6p which are both larger than Msh2p, but may be a degradation product of one of the MutS proteins. **C**: Same as A, in the CAG orientation. Restriction sites used for Southern blots are *Eco*RI and *Xba*I in this orientation. Frequent sectored colonies were also detected in this orientation when *MSH2* was overexpressed. **D**: Same as A, for strains harboring the p*msh2*-E768A overexpression plasmid, in both orientations.

Repeat tracts are around 100 triplets long in each strain, except in the wild-type strain in the CAG orientation in which it is slightly longer (*ca.* 120 triplets), and in the *msh3*Δ strain in which it is slightly shorter (*ca.* 70 triplets long), despite our repeated unsuccessful attempts to build a *msh3*Δ strain containing a longer repeat tract.

ACCEPTED MANUSCRIPT

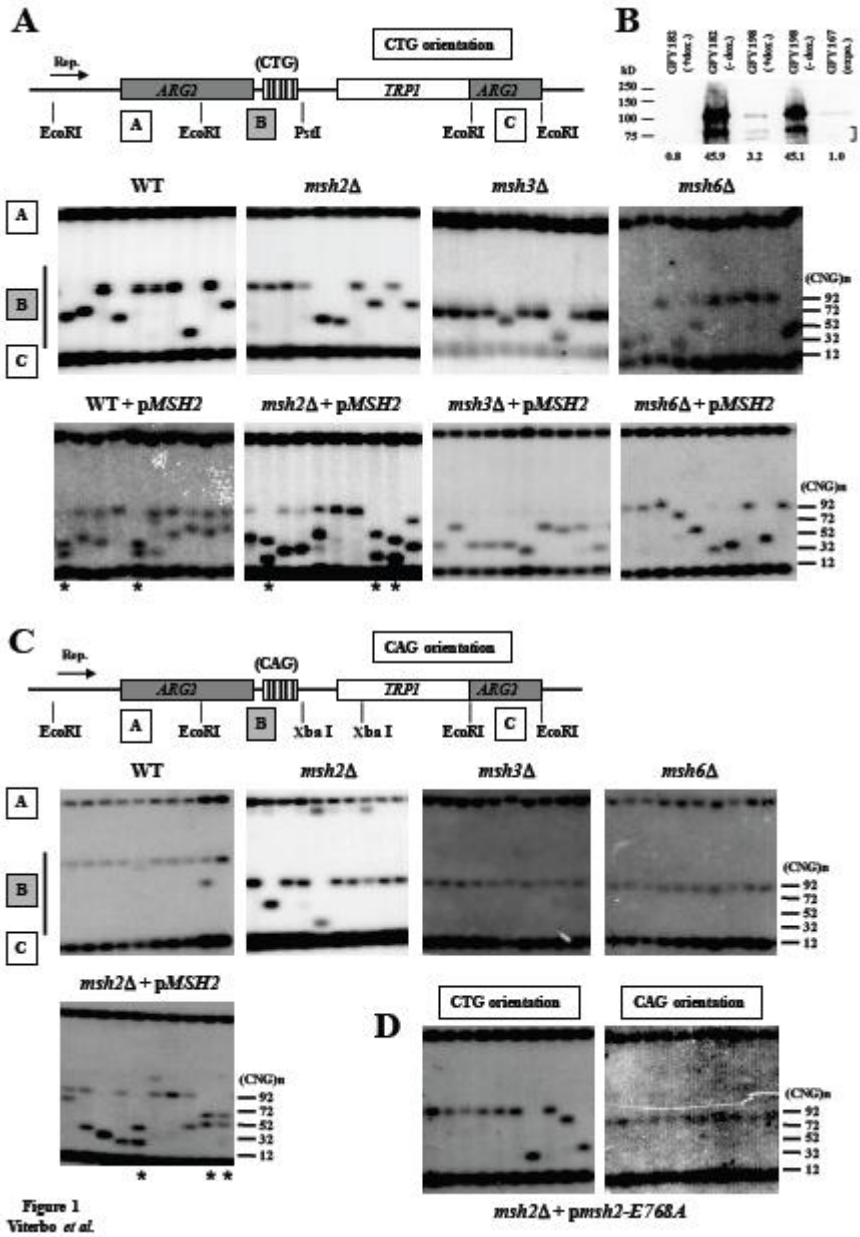
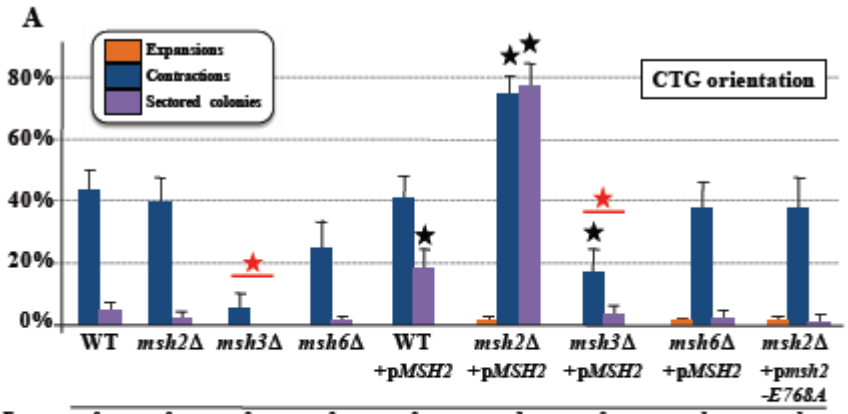


Figure 1
Viterbo *et al.*

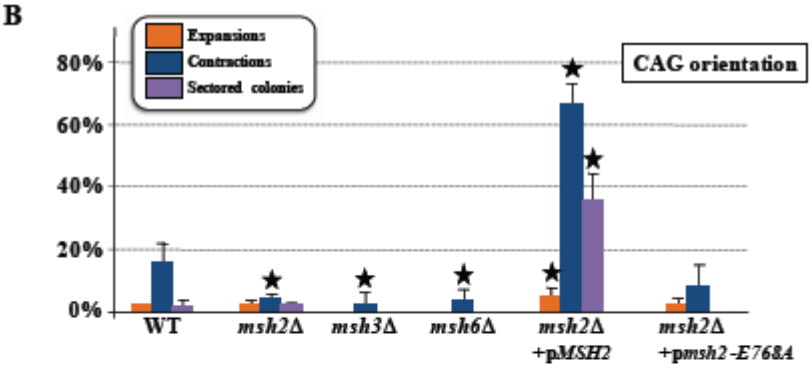
ACCEPTED MANUSCRIPT

Figure 2 Instability of CAG/CTG trinucleotide repeat tracts and sectored colonies in MMR mutant backgrounds. **A**: Instability in the CTG orientation. Contractions and expansions are represented respectively in blue and orange, as the percentage of contracted or expanded alleles (above or below the original size) detected on Southern blots as compared to the total number of alleles detected. Purple bars correspond to the percentage of colonies in which more than one alleles was detected. Error bars correspond to 95% confidence intervals. **B**: Same as above in the CAG orientation. Black stars show mutant distributions significantly different from their respective wild-type distributions (CAG or CTG orientations), as determined by Fisher exact tests (p-value < 0.01). Red stars indicate that contractions and sectored colonies are not statistically different (p-value > 0.99) in *msh3*Δ and *msh3*Δ +p*MSH2* strains in the CTG orientation. Raw numbers are shown under each graph. Expans.: expansions; Contract.: contractions; Sectored: number of colonies showing more than one allele. In parenthesis is shown the number of sectored colonies in which none of the observed alleles is of parental size. Alleles: total number of different alleles detected on Southern blots, including those from sectored colonies; Colonies: total number of colonies analyzed by Southern blots. C: Contracted allele length comparisons of sectored colonies containing one parental band (classical heteroduplex, in blue) with colonies containing no parental band (complex heteroduplex, in orange), in both orientations. Contractions were separated into four different categories, less than 26 triplets, 25-49 triplets, 50-89 triplets and more than 89 triplets (close to starting size, only one allele showed such length).

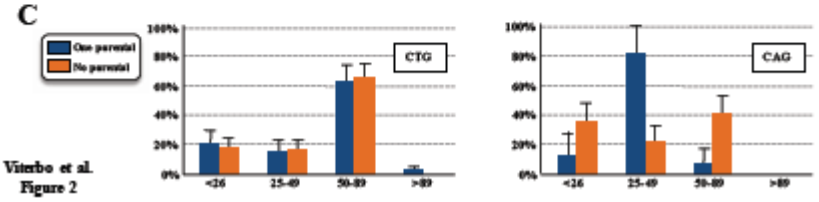
ACCEPTED MANUSCRIPT



	WT	msh2Δ	msh3Δ	msh6Δ	WT +pMSH2	msh2Δ +pMSH2	msh3Δ +pMSH2	msh6Δ +pMSH2	msh2Δ +pmsh2 -E768A
Expans.	0	0	0	0	0	3	0	1	1
Contract.	82	56	5	23	69	152	17	51	36
Sectored	8 (0)	3 (1)	0	1 (1)	26 (8)	89 (41)	3 (0)	3 (0)	1 (0)
Alleles	191	142	91	93	170	205	99	135	96
Colonies	183	139	91	92	144	116	96	132	95



	WT	msh2Δ	msh3Δ	msh6Δ	msh2Δ +pMSH2	msh2Δ +pmsh2 -E768A
Expans.	1	3	0	0	8	1
Contract.	20	5	2	2	115	6
Sectored	2 (0)	2 (2)	0	0	46 (31)	0
Alleles	131	167	81	69	175	73
Colonies	129	165	81	69	129	73

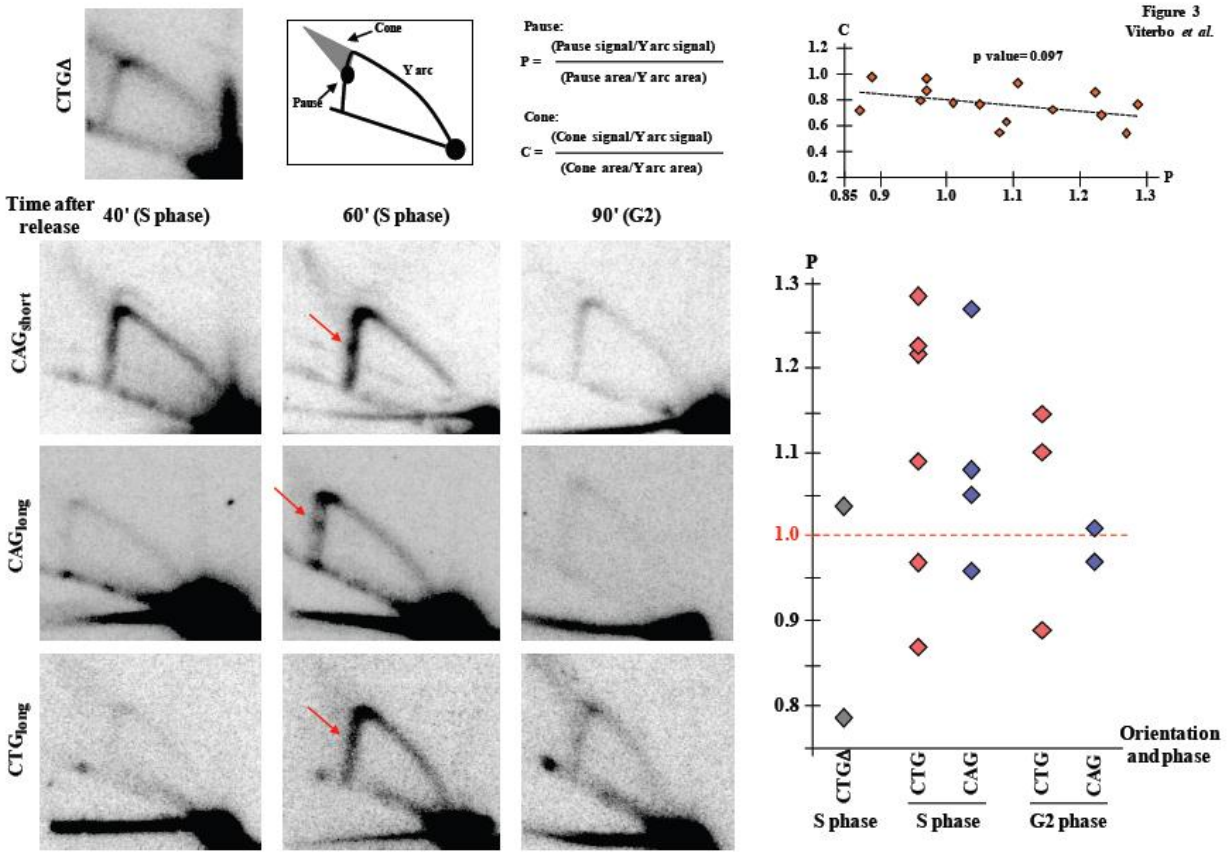


Viterbo et al.
Figure 2

ACCEPTED MANUSCRIPT

Figure 3 Replication fork stalling at CAG/CTG trinucleotide repeats happens in both orientations. Left: 2D gels of DNA collected during S-phase time courses are shown. Two repeat lengths were used for the CAG orientation: short repeat tract corresponding to *ca.* 60 triplets, long repeat tracts corresponding to *ca.* 120 triplets. Red arrows point to the locus at which trinucleotide repeat tracts are integrated, and where thickening of the Y arc corresponding to replication fork stalling is also detected. A 2D gel of the strain in which the trinucleotide repeat tract was perfectly deleted and in which no pausing signal was detected is shown above (CTGΔ). The graph to the upper right shows the weak inverse correlation between replication fork pausing (P) and cone signal (C). Replication pausing indexes (P) are calculated as the ratio of pausing signal over Y arc signal, compared to their respective areas. Right: The graph shows pausing indexes for each time point, in each orientation, grey diamonds corresponding to the CTGΔ strain, blue diamonds corresponding to CAG orientation and red diamonds to CTG orientation. Pausing indexes are shown for S or G2 phases of the cell cycle, as determined by visual inspection and flow cytometry analyses (not shown). For the CAG orientation, short and long repeat tracts were merged in this graph. For the CTGΔ strain, only S phase time points were collected.

ACCEPTED MANUSCRIPT



ACCEPTED MANUSCRIPT

Figure 4 Effect of mismatch-repair mutants on replication fork stalling. Upper part: 2D gels are shown, for CTG orientation and CAG orientation. Repeat tract lengths are as in Figure 1. Red arrows point to replication fork pauses whose pausing indexes are statistically above 1. Dotted red arrows point to pausing signals that are statistically lower than the wild-type signal. Lower part: Graphs showing P in *msh2*Δ or *msh2*Δ +p*MSH2* cells in both orientations (left), or in other MMR mutants in the CTG orientation only (right). Pausing indexes are shown for S or G2 phases of the cell cycle, as determined by visual inspection and flow cytometry analyses (not shown). The p-values of Mann-Whitney-Wilcoxon tests showing a significant difference between a given mutant and the wild-type reference (Figure 3) are indicated. For the tests S and G2 values were merged.

ACCEPTED MANUSCRIPT

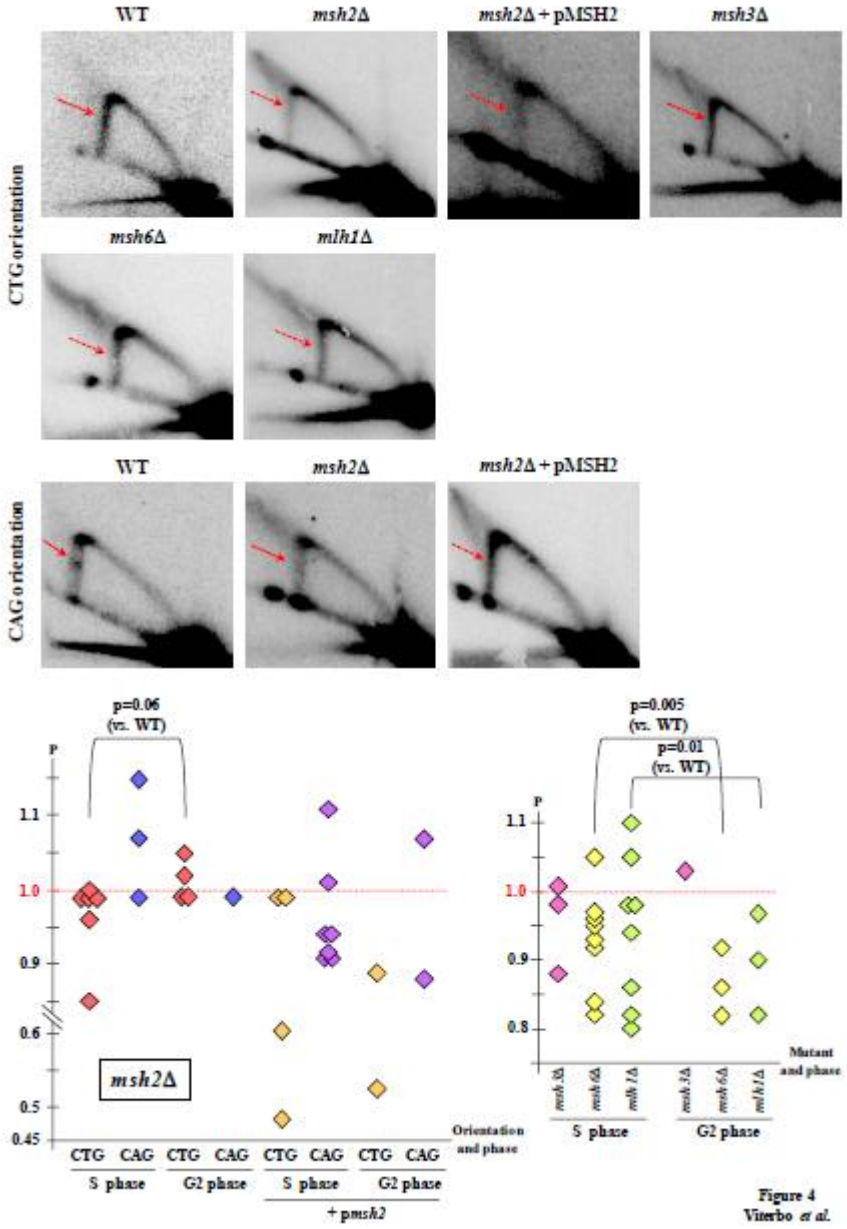
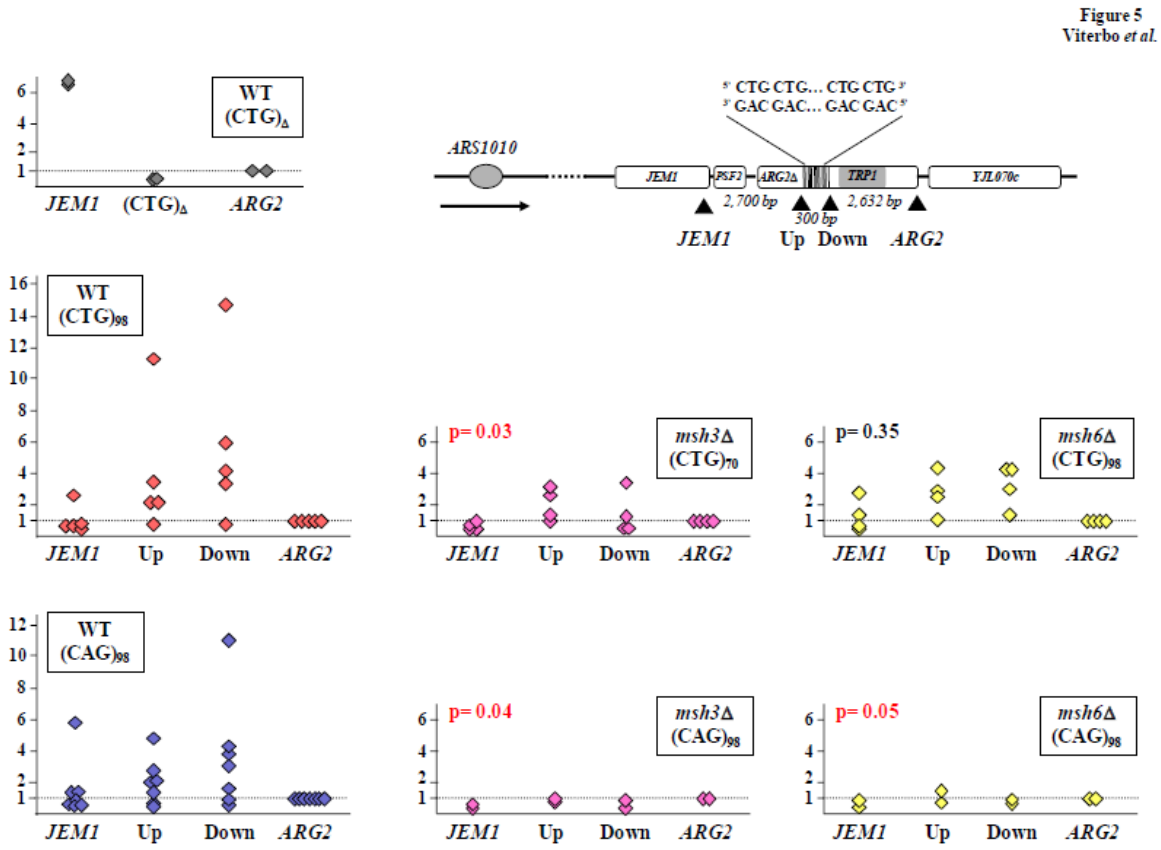


Figure 4
Viterbo *et al.*

ACCEPTED MANUSCRIPT

Figure 5 Real-time qPCR quantification of Msh2p immunoprecipitation. Two couples of primers located directly upstream (Up) and downstream (Down) of the trinucleotide repeat tract were designed, as well as two other couples of primers, located respectively 2.7 kb upstream (*JEM1*) and downstream (*ARG2*) of the repeat tract (Table 2). Enrichment rates were determined as the ratio of immunoprecipitated DNA at each locus on immunoprecipitated DNA at the *ARG2* locus. Each diamond represents one independent experiment. X axis: the four different loci studied. Y axis: ratio of (immunoprecipitated DNA at each locus) divided by (immunoprecipitated DNA at *ARG2*). Statistical significance (Mann-Whitney-Wilcoxon test p-values) of Msh2p loss of enrichment in each mutant strain was compared to the corresponding wild-type strain and shown on each graph. Only the *msh6*Δ strain in the CTG orientation was not statistically different from the wild-type strain in the same orientation, although the overall level of Msh2p enrichment was slightly reduced.



ACCEPTED MANUSCRIPT

Figure 6 Models of interactions between MMR proteins and CAG/CTG trinucleotide repeats. For the sake of clarity, only one orientation was shown here and only one strand (the lagging-strand template) was shown to carry secondary structures, but such structures may form on any strand during replication. Single-stranded DNA may transiently appear on the lagging-strand (or leading-strand) template, allowing formation of DNA hairpin(s) and/or slipped-strand structures. MMR proteins traveling with replication forks may recognize and bind such structures, interfering with the replication fork (represented here by PCNA/Pol δ) and in rare instances promoting replication fork stalling. Alternatively, fork reversion may occur, as suggested by the small increase in conical shapes visible on 2D gels when pausing signals decrease. Some secondary structures may be unwound by fork reversal, although this is highly speculative, as shown by the questionmark.

When *MSH2* is overexpressed, formation of MutSα/β complexes is facilitated and such complexes bind to secondary structures, stabilizing them and protecting them from degradation by structure-specific nucleases. These heteroduplex regions will give rise to one normal-length allele and one contracted allele during the next replication round. Note that if secondary structures occur on newly-synthesized fragments themselves, MMR binding and hDNA stabilization should preferentially lead to expansions, which is seldom observed. Alternatively, *MSH2* overexpression may lead to partial titration of PCNA, reducing replication fork processivity, therefore increasing replication slippage leading to contractions. The fact that S phase duration is increased when *MSH2* is overexpressed favors this hypothesis, whereas Msh2p enrichment at trinucleotide repeats, as well as the necessary presence of Msh3p and Msh6p to observe sectored colonies do not support this hypothesis, although the two models (secondary structure interaction and PCNA interaction) are not formally exclusive from each other.

ACCEPTED MANUSCRIPT

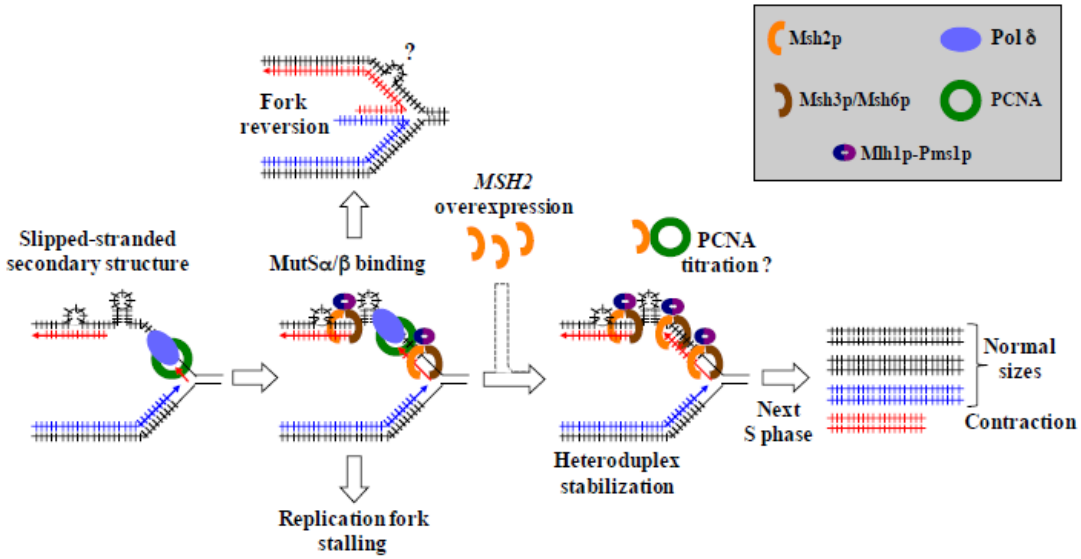


Figure 6
Viterbo *et al.*