Supporting Information

Lin et al. 10.1073/pnas.0909740107

SI Text

SI Materials and Methods. ***Plasmids and bacterial strains.*** pRW3245 and pRW3246 each contain a tract of $(CTG \cdot CAG)_{98}$ repeats, but differ from one another in the orientation of the repeat relative to the origin of replication and to the lacZ promoter. The pGEM series of plasmids was generated by cleaving the $(CTG \cdot CAG)_{98}$ tract from pRW3245 using HindIII and EcoRI restriction enzymes, and cloning the fragment into pGEMZf (+) (Promega). Clones containing the full-length 98 repeat insert (pGEM_CTG98) as well as two deletion derivatives harboring 27 and 52 CTG · CAG repeats (Table S1) were isolated and sequenced. The CTG · CAG tract is located 85 and 93 bp from the T7 and SP6 transcription start sites, respectively. Plasmids pDr_CTA68 and pDr_TAG harboring the $(CTA \cdot TAG)_{68}$ tract were obtained by a complete bisulfite conversion of the CTG · CAG inset followed by PCR and cloning into pDrive (Qiagen). Selected clones were analyzed by DNA sequencing. All plasmids were cloned and maintained in *E. coli* HB101 [*mcrB, mmr, hsdS20* (r_B^-, m_B^-), *recA1, supE44, ara14, galK2, lacY1, proA2, rplS20* (Sm^R), *xyl5, λ*$^-$, *leuB6, mtl-1*].

Oligonucleotides and siRNAs. Oligonucleotides were purchased from Sigma-Genosys, and Stealth siRNAs were synthesized by Invitrogen. Oligonucleotides: CMVD-For TACCGAGCTCGG AT CCACTAGTC, BGHB-Rev GGTTTAAACGGGCCCTCTA GA, and CMV-F-conv TACCAAACTCAAATCCACTAATC; BGH-R-conv GGTTTAAATGG GTTTTTTAGA were used to amplify CTG · CAG repeat region of the bisulfite treated DNA in nondenaturing and denaturing conditions, respectively. Oligonucleotides: HPRT-For CGGCTACAAGGACGACTCT AG; HPRT-Rev TTGATGTAATCCAGCAGGTCAGC; β-actin-For AGAGAGGCATCCTCACCCTG; β-actin-Rev CATGAGG TAGTCAGTCA GGT; RNaseH1-For CACAGAGGATGAGGC CTG; RNaseH1-Rev CAGTGGCTCAC GGAGTC; RNaseH2A-For CTGGGCGTCGATGAGG; RNaseH2B-Rev CCGCTCGC TCTCCAATAG were used in the quantitative RT-PCR reactions to determine the effect of siRNA treatment on RNase H and HPRT genes expression. The primer T70 TAATACGACTCACT ATAGGG was used for DNA sequencing of the cloned PCR products.

The following siRNAs were used (only sense strands are indicated): RNaseH1 siRNA1 GACCGGUUUCCUGCUGCCA GAUUUA; RNaseH1 siRNA2 CCGGAAGUUUCAGAAGGG CAUGAAA; RNaseH2 siRNA1 GGGCGGGUCAAAUACAA CCUGA ACU; RNaseH2 siRNA2 CCACUGGGCUUAUACA GUAU GCAUU; Vimentin-siRNA GAAUGGUACAAAUCCA AGU.

In vitro detection of R loops. In vitro transcription mixture was divided into three 10-μl fractions. One aliquot was incubated for 30 min at 37 °C with RNase A/I (1 unit of RNase A, U.S. Biochemicals; 2 units of the RNase I, Ambion); one was incubated for 30 min at 37 °C with RNase A/I and RNase H (1 unit, Invitrogen); and the remaining aliquot served as a "no RNase" transcriptional control. Reactions were stopped by phenol/chloroform extraction and ethanol precipitation. Samples were loaded onto 0.8% Tris-acetate/EDTA agarose gels. Electrophoresis was conducted at 5 V/cm in the absence of ethidium bromide. Gels were directly exposed to the PhosphorImager screen (GE Healthcare) or x-ray film (Kodak). Agarose gels used for analysis of no-RNase control reactions were stained with ethidium bromide followed by PhosphorImager analyses. A gel purification kit (Qiagen) was utilized to extract RNA · DNA hybrids from the agarose gels followed by a 15 min incubation at 37 °C with DNase I (Ambion). For quantitative analyses of the RNase-A/I resistant and RNase-H sensitive RNA · DNA hybrids, radioactive RNA was extracted from agarose gels, melted in 10 mM Tris·HCl pH 8.0, 1 mM EDTA (TE) buffer, and measured by scintillation counting (Beckman Coulter).

Detection of RNA · DNA hybrids in genomic DNA. Approximately 2×10^6 HEK293_5150 cells containing $(CTG \cdot CAG)_{67}$ repeat tract (rCAG transcript) were washed with PBS and harvested by trypsinization. The cell pellet was washed twice with PBS and centrifuged for 3 min at $1,000 \times g$. DNA was extracted in nondenaturing conditions by lysing the cells in TE buffer, pH 8.0 plus 0.5% sodium dodecyl sulfate, followed by proteinase K digestion (0.2 mg/mL, Sigma) overnight at 37 °C. Genomic DNA was subsequently extracted with phenol/chloroform, precipitated in ethanol, and resuspended in TE buffer. Next, the genomic DNA was incubated for 6 h at 37 °C in the presence of 4 units of RNase H (Invitrogen). In parallel, a control reaction without the enzyme was performed under identical conditions. Two micrograms of DNA (treated with RNase H or not treated) were subjected to bisulfite modification using an EpiTect Bisulfite Modification Kit (Qiagen). Bisulfite reactions were conducted in nondenaturing conditions for 12 h at 37 °C. As a positive control for bisulfite modification, a conversion reaction was conducted under denaturing conditions (3 cycles of 99 °C for 5 min followed by 60 °C for 25 min in the first cycle, 85 min in the second cycle, and 175 min in the third cycle). Desulfonation and purification of modified genomic DNA was carried according to the manufacturer's recommendations. Bisulfite modified genomic DNA (100 ng) was used in the PCR amplification using CMVD-For and BGHB-Rev primers and the FailSafe PCR system (Epicentre Biotechnologies). In the case of DNA modified in denaturing conditions, a set of conversion specific PCR primers, CMV-F-conv and BGH-R-conv, was used. All PCR analyses were carried out under the identical conditions for 35 cycles of 95 °C for 20 s, 55 °C for 30 s, and 72 °C for 1 min. PCR products were separated on 1% agarose gels, purified and cloned into pDrive PCR cloning vector (Qiagen) according to the manufacturer's recommendations. Plasmids isolated from individual colonies were sequenced using a T70 sequencing primer.

Assay for genetic instability in E. coli. The efficiency of the pRW3245 and pRW3246 plasmids transformation was determined in every experiment by plating an aliquot of the transformation mixture onto an agar plate containing ampicillin (100 μg/mL). Cultures were initiated with 1,000–2,000 colony forming units. Cells were subcultured into fresh medium during logarithmic growth phase (A_{600} of 0.75–0.9). Cells were recultivated three times in 20-generation intervals. Each subculture was started with 10^3–10^4 cells as an inoculum. An aliquot of cells from each subculture was harvested every 20 generations and stored at −20 °C, so that plasmids from all cultures could be isolated simultaneously, using a Wizard Miniprep System (Promega). Changes in repeat number were determined by restriction digestion to release the repeat-containing insert, which was then end-labeled with [α^{32}P]dATP using the Klenow fragment of DNA polymerase I. Subsequently, inserts containing CTG · CAG repeats were separated by electrophoresis on 5% polyacrylamide gels. Repeat instability, which was measured as the loss of the full-length $(CTG \cdot CAG)_{98}$ repeat tract, was quantitatively analyzed using a PhosphorImager. In

the experiments with inhibited transcription, the plasmid pI^Q-kan was first transformed into the appropriate *E. coli* strain, and the pI^Q-kan-containing competent cells were prepared in medium containing kanamycin (50 μg/mL). Transcription from the lacZ promoter was induced by adding 1 mM IPTG into the LB medium.

Frequency of CTG · CAG contractions in human cells. The FLAH25 cells, which were derived from $HPRT^-$ HT1080 cells, contain an *HPRT* minigene that is inactivated by an intronic insertion of a $(CTG \cdot CAG)_{95}$ repeat oriented to give an rCAG transcript. Contraction of the repeat tract to fewer than 39 repeats leads to the $HRPT^+$ phenotype, which can be selected on hypoxanthine/aminopterin/thymidine supplemented medium (0.1 mM hypoxanthine, 0.4 μM aminopterin, and 16 μM thymine; Sigma). Transcription of the *HPRT* minigene is driven from the Tet-ON promoter, which can be induced by addition of doxycycline (2 μg/mL). Cells were incubated at 37 °C with 5% CO_2 in DMEM-F-12 medium supplemented with 10% dialyzed serum and 1% nonessential amino acids (Invitrogen). Contraction frequencies, which were calculated as the number of $HRPT^+$ colonies divided by the number of viable cells, are the averages of results from at least five experiments. To analyze the effects of RNase H1 and RNase H2 on the frequency of repeat contractions, siRNAs targeting expression of both enzymes were used. siRNAs (200 nM; Invitrogen) were transfected into cells by using Oligofectamine (Invitrogen); vimentin specific siRNA (200 nM), which does not affect repeat contraction, served as a control. Knockdown of target gene expression was evaluated by real-time reverse transcription-PCR (RT-PCR).

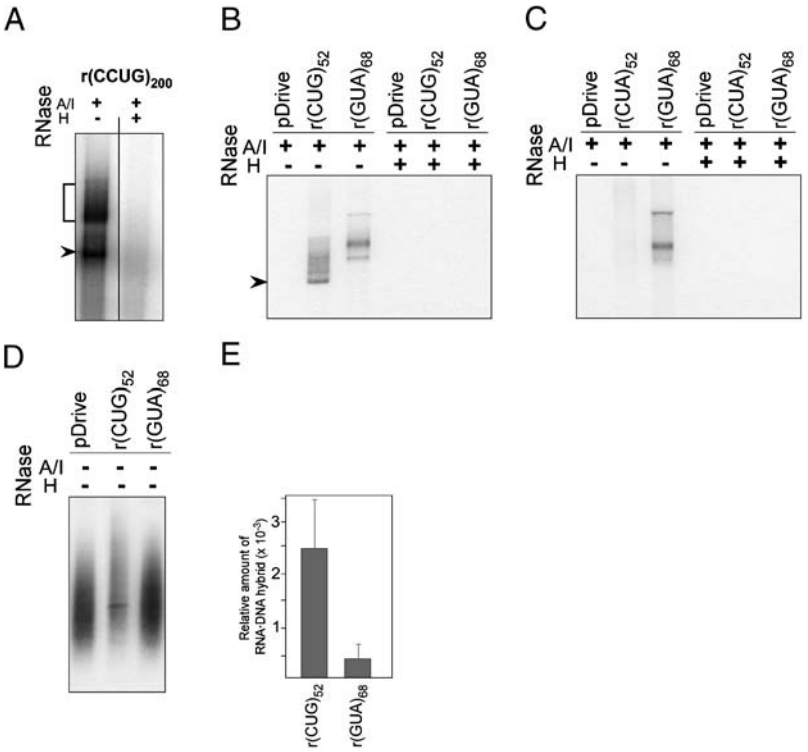

Fig. S1. Formation of RNA · DNA hybrids upon in vitro transcription of $(CCTG \cdot CAGG)_{200}$ and $(CTA \cdot TAG)_{68}$ repeats. In vitro transcriptions of supercoiled plasmids containing 200 CCTG · CAGG repeats (rCCUG transcript) (*A*) and 68 (CTA · TAG) repeats (rCUA and rGUA transcripts) (*B*, *C*) were conducted using T7 RNAP. Transcription reactions were treated with ribonuclease A/I and RNase A/I/H mixtures for 30 min prior to the phenol/chloroform extraction and agarose gel electrophoresis as indicated. Supercoiled pDrive (Qiagen) and the plasmid harboring $(CTG \cdot CAG)_{52}$ repeats were used as negative and positive controls, respectively. The arrowhead indicates migration of the supercoiled DNA, and the position of the relaxed plasmids is marked by a bracket. The efficiencies of R-loop formation by the $r(GUA)_{68}$ and the $r(CUG)_{52}$ tracts were compared relative to the overall efficiencies of the in vitro transcription through these sequences (*D*, *E*).

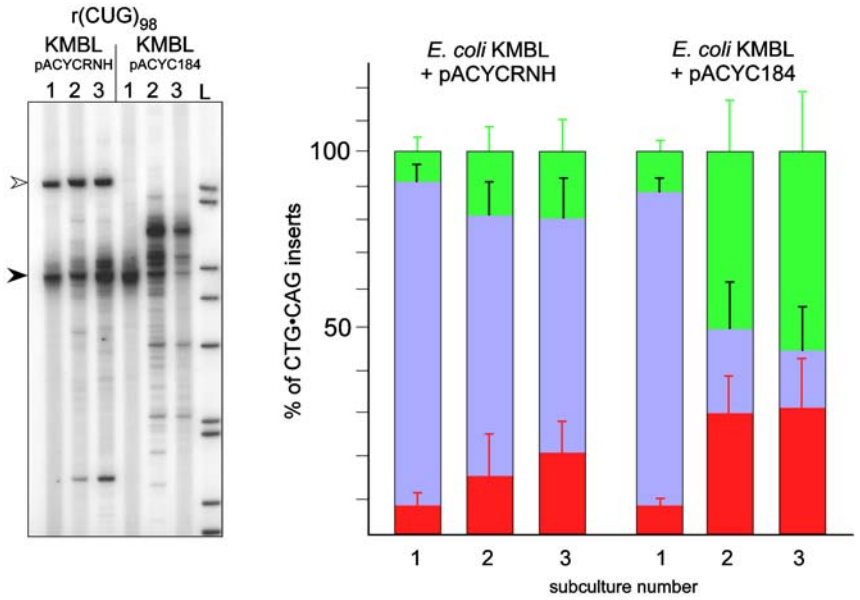

Fig. S2. Overexpression of RNAse H alleviates transcription-induced CTG · CAG repeat instabilities in *E. coli*. KMBL1001 was cotransformed with pRW3245 harboring the $(CTG \cdot CAG)_{98}$ insert together with pACYCRNH (expressing RNase HI) or pACYC184 (control). Transcription through the repeats was stimulated by 1 mM IPTG. In vivo recultivation experiments were carried out throughout ≈60 cell generations, subcultured in 20 generation intervals. The amount of expansions, deletions, and full-length CTG · CAG inserts was quantified after each of three subcultivations using Phorophorimager. The filled arrowhead indicates the electrophoretic migration of the full-length $(CTG \cdot CAG)_{98}$ insert and the open arrowhead designates a restriction fragment of the RNase HI-containing pACYCRNH. Error bars represent standard deviations derived from three independent cotransfection experiments. Repeat expansions are indicated in green, deletions in red, and the remaining fraction of the full-length repeat insert is shown in blue.

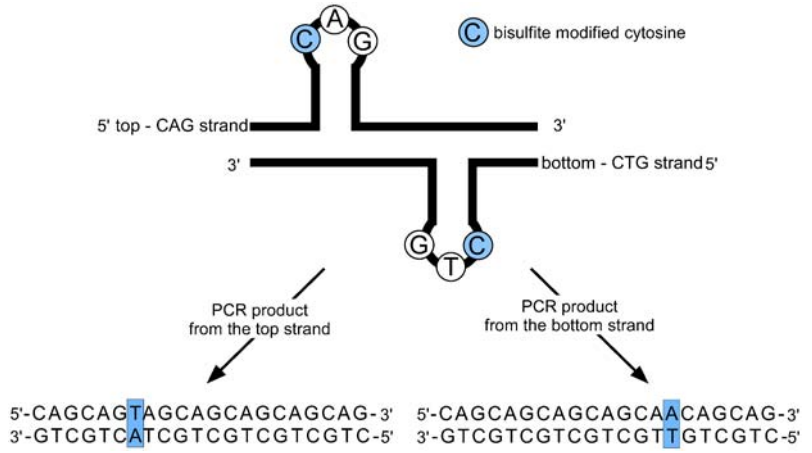

Fig. S3. Strand-specific bisulfite modification. Bisulfite conversion of the cytosine residue in the CTG strands leads to the conversion into the TTG · CAA triplet, whereas conversion of cytosine in the complementary CAG strand results in the TAG · CTA triplet formation.

Table S1. Plasmids used in this study

Plasmid	Backbone	Gene	Repeats	Resistance marker
pRW3245	pUC19	—	$(CAG \cdot CTG)_{98}$	Amp
pRW3246	pUC19	—	$(CTG \cdot CAG)_{98}$	Amp
pGEM_CTG27	pGEM3Zf(+)	—	$(CTG \cdot CAG)_{27}$	Amp
pGEM_CTG52	pGEM3Zf(+)	—	$(CTG \cdot CAG)_{52}$	Amp
pGEM_CTG98	pGEM3Zf(+)	—	$(CTG \cdot CAG)_{98}$	Amp
pRW5150	pcDNA5/FRT/TO	—	$(CTG \cdot CAG)_{67}$	Amp
pRW3142	pGEM3Zf(+)	—	$(TAA)_5TAGTAAGAA(TAA)_{89}$	Amp
pDr_CTA68	pDrive	—	$(CTA \cdot TAG)_{68}$	Amp
pDr_TAG68	pDrive	—	$(TAG \cdot CTA)_{68}$	Amp
pI^Q-kan	—	LacZ repressor	—	Kan
pACYCRNH	pACYC184	RNase HI	—	Cm, Tet
pACYC184	—	—	—	Cm, Tet
pGEM3Zf(+)	—	—	—	Amp
pDrive	—	—	—	Amp