HHS Public Access

Author manuscript

DNA Repair (Amst). Author manuscript; available in PMC 2023 September 11.

Published in final edited form as:

DNA Repair (Amst). 2023 June ; 126: 103489. doi:10.1016/j.dnarep.2023.103489.

Spontaneous deamination of cytosine to uracil is biased to the non-transcribed DNA strand in yeast

Jonathan D Williams[a,1], **Demi Zhu**[a], **María García-Rubio**[b], **Samantha Shaltz**[a], **Andrés Aguilera**[b], **Sue Jinks-Robertson**[a,*]

[a]Department of Molecular Genetics and Microbiology, 213 Research Dr., Duke University Medical Center, Durham, NC 27710

[b]Centro Andaluz de Biología Molecular y Medicina Regenerativa (CABIMER), Universidad de Sevilla-CSIC, Seville, Spain

Abstract

Transcription in *Saccharomyces cerevisiae* is associated with elevated mutation and this partially reflects enhanced damage of the corresponding DNA. Spontaneous deamination of cytosine to uracil leads to CG>TA mutations that provide a strand-specific read-out of damage in strains that lack the ability to remove uracil from DNA. Using the *CAN1* forward mutation reporter, we found that C>T and G>A mutations, which reflect deamination of the non-transcribed and transcribed DNA strands, respectively, occurred at similar rates under low-transcription conditions. By contrast, the rate of C>T mutations was 3-fold higher than G>A mutations under high-transcription conditions, demonstrating biased deamination of the non-transcribed strand (NTS). The NTS is transiently single-stranded within the ~15 bp transcription bubble, or a more extensive region of the NTS can be exposed as part of an R-loop that can form behind RNA polymerase. Neither the deletion of genes whose products restrain R-loop formation nor the over-expression of RNase H1, which degrades R-loops, reduced the biased deamination of the NTS, and no transcription-associated R-loop formation at *CAN1* was detected. These results suggest that the NTS within the transcription bubble is a target for spontaneous deamination and likely other types of DNA damage.

Keywords

transcription; mutagenesis; RNA polymerase; R-loops; transcription bubble; deamination

1. INTRODUCTION

Mutations generally have negative consequences but are required for adaptation under stress conditions and for evolutionary change. Most mutations arise during genome duplication and reflect either rare errors of the high-fidelity replicative DNA polymerases or the bypass of DNA damage by error-prone translesion-synthesis DNA polymerases. In addition to

*Communicating author: Department of Molecular Genetics and Microbiology, 213 Research Dr., Duke University Medical Center, Durham, NC 27710, Phone: 919-681-7273, sue.robertson@duke.edu.

[1]Current address: Labcorp, Burlington, NC

replication-associated changes in DNA sequence, mutations also arise in the context of transcription. This may be a particularly important source of genetic alterations in the post-mitotic cells of metazoans, such as those in neural tissue [1]. In microorganisms, transcription-associated mutagenesis (TAM) is operationally defined as an increase in the mutation rate of a target gene when it is transcribed at a high versus low level (reviewed in [2]). Early studies in *Saccharomyces cerevisiae* demonstrated that TAM increases when excision-repair or error-free damage-bypass pathways are inactivated, and decreases when translesion synthesis is disrupted [3]. There also is synergism between the stimulatory effects of exogenous mutagens and transcription on recombination rates [4]. Collectively, these genetic data indicate that transcription increases the exposure of the underlying DNA template to endogenous and exogenous damage. Mutagenic damage can also be inflicted by topoisomerase I (Top1), which resolves the torsional stress that accumulates when DNA strands are separated during transcription [5, 6]. Top1-dependent mutagenesis specifically generates small deletions and reflects sequential cleavage of one DNA strand, with the initial incision occurring at an embedded ribonucleotide [7, 8].

Both replication and transcription impart an asymmetry between the two strands of duplex DNA. During yeast replication, complementary strands are copied by distinct DNA polymerases and the efficiencies of replication-error removal differ, which can lead to mutation asymmetries on the leading and lagging strands (reviewed in [9]). During transcription, only one strand typically serves as the template for RNA synthesis. The template strand is referred to as the transcribed strand or TS, while the complementary strand is the non-transcribed strand or NTS. Strand-related biases are readily evident when mutations are reported with reference to only one DNA strand and both replication- and transcription-associated mutation biases have been reported in cancer genomes [10].

Strand-related biases associated with transcription can reflect more robust repair of the TS than NTS and/or more damage to the NTS than TS. A lesion on the TS that blocks/stalls RNA polymerase, for example, triggers more efficient removal by the nucleotide excision repair pathway than does the same lesion on the NTS (transcription-coupled repair; reviewed in [11]). In addition, single-stranded DNA is more susceptible to chemical and enzymatic damage than is double-stranded DNA (reviewed in [12, 13]). More damage of the NTS relative to the TS presumably reflects its enhanced single-stranded character. During transcription, 10–15 nt of the NTS are exposed within the RNA polymerase (RNAP) transcription bubble, whereas the TS is protected through its pairing with the nascent transcript. A much larger segment of the NTS can be exposed when the RNA threads back after exiting RNAP and pairs with the template strand [14]. The extended RNA-DNA hybrid, together with the single-stranded NTS, is referred to as an R-loop. R-loops are important for transcription termination, but conflicts with the replisome can generate DSBs that are potent initiators of homologous recombination and genetic instability (reviewed in [15]). The contribution of R-loops to the accumulation of spontaneous DNA damage is less clear.

While transcription-associated genetic instability is generally problematic, it is physiologically important in the maturation of the vertebrate immune system. Antigen-stimulated somatic hypermutation (SHM) improves antibody affinity, while class-switch

recombination (CSR) diversifies antibody effector functions (reviewed in [16]). Both processes are limited to B cells and require enzymatic deamination of cytosine to uracil by the single-strand specific enzyme AID (activation-induced cytosine deaminase). *In vitro*, AID travels with phage RNAP [17] and primarily deaminates the NTS, leading to the suggestion that it also targets the transcription bubble. A more recent study with eukaryotic RNAPII, however, demonstrated equal deamination of the two DNA strands by AID and it was suggested that both become accessible in a backtracked transcription bubble [18]. By contrast, other reports have shown that deamination during CSR involves R-loop formation [19]. AID expression in yeast was reported to have an NTS bias in an RNAPII-transcribed gene, but only in a mutant background where R-loops were expected to accumulate at a high level [20].

The metazoan-specific APOBEC (apolipoprotein B mRNA editing enzyme, catalytic peptide) family of cytosine deaminases target viral genomes and are an important component of innate immunity [21]. APOBECs also robustly deaminate single-stranded genomic DNA, however, and are potent drivers of tumorigenesis [22]. In yeast, APOBEC3G primarily targets the NTS of tRNA genes [23], which are strong sites of R-loop accumulation [24]. The NTS bias in tRNA genes was exacerbated by loss of RNase H, which degrades the RNA component of RNA-DNA hybrids, consistent with a primary targeting of single-stranded DNA in the context of R-loops. An APOBEC from lamprey also targeted the NTS in a genome-wide analysis in yeast, but with preferential deamination of transcription start sites that contain stable pre-initiation complexes [25].

While the single-stranded DNA of R-loops can be clearly targeted by cytosine deaminases, a role for R-loops in spontaneous deamination has not been reported. A primary role for the transcription bubble in NTS-biased cytosine deamination, however, was previously inferred in bacterial cells [26], where co-transcriptional translation is assumed to limit R-loop formation. The rapid assembly of yeast transcripts into ribonucleoprotein particles similarly sequesters transcripts to limit R-loop formation [27]. Although most work to date has suggested greater susceptibility of the NTS to various types of DNA damage, the reverse bias was reported at a bacterial promoter when the transcription and replication forks were in a head-on orientation [28]. It was suggested that this reflected a protective interaction of sigma factor with the NTS in a stabilized RNAP open complex.

In the current study, we examined the spontaneous deamination of cytosine to uracil in yeast as a function of low- *versus* high-transcription of the *CAN1* forward-mutation reporter. Mutations attributable to cytosine deamination occurred at similar rates on the NTS and TS strands under low-transcription conditions but were strongly biased to the NTS under high-transcription conditions. No correlation between the transcription-associated mutation bias and R-loop accumulation was detected, however, suggesting that the NTS exposed in the transcription bubble may be the primary target of DNA damage in some contexts.

2. MATERIALS AND METHODS

2.1. Strain construction

All strains used for mutational analyses (Table S1) were derived from SJR282 (*MATα ade2–101$_{OC}$ his3Δ200 ura3ΔNco gal80Δ::HIS3*) by transformation. As previously described [5], high transcription of *CAN1* was achieved by replacing the endogenous promoter with the *GAL1* promoter (*pGAL*) linked to the bacterial *kan* marker [29]. *FCY1, MSH6*, *MFT1*, *RNH1*, *RNH201*, *TOP1* or *UNG1* was deleted by one-step allele replacement using PCR-generated cassettes containing an appropriate selectable marker. For experiments involving RNase H1 overexpression, the high-txn *ung1*Δ strain was transformed with the *URA3-CEN* vector pRS316 [30] or a derivative containing a *pGAL1-RNH1* fusion [31].

2.2. Mutation rates and spectra

A minimum of 24 cultures of each strain was used for determining canavanine-resistance (Can-R) rates. Each culture was inoculated with a single colony and was grown to saturation (three days) in YEP-GE (1% yeast extract, 2% Bacto-peptone, 2% glycerol and 2% ethanol) on a roller drum at 30°C. In a *gal80* background, *pGAL* is constitutively active in YEP-GE medium [32]. Cultures were washed with water and the number of Can-R mutants in each was determined by plating an appropriate dilution on synthetic complete medium (SC; 0.17% yeast nitrogen base, 0.5% ammonium sulfate, 2% agar and 0.13% Hartwell's complete amino acid mix) lacking arginine and supplemented with 60 μg/ml canavanine (Sigma). Total cell number was estimated by plating an appropriate dilution on non-selective YPD medium (1% yeast extract, 2% Bacto-peptone, 2% dextrose and 2% agar). Mutation rates were calculated using method of the median [33], and 95% confidence intervals (CIs) were determined as previously described [34].

Independent Can-R mutants for sequencing were obtained using a prong-plating technique that generates ~100 mini-cultures/plate. Briefly, cells were diluted in water and a 100-count, flat-tipped custom pinning device was used to transfer ~10^3 cells/pin onto non-selective YEP-GE plates. Cells were also spotted onto SC-Arg+Can plates to ensure there were no pre-existing Can-R mutants. After three days of growth, cells were replica-plated onto SC-Arg+Can plates and incubated for two days to allow the appearance of Can-R colonies. A single Can-R colony was picked from each spot and following genomic DNA isolation the *CAN1* locus was amplified using MyTaq (Bioline) polymerase.

For PacBio analysis, 16-nt barcode (a list available is at https://github.com/PacificBiosciences/Bioinformatics-Training/blob/master/barcoding/pacbio_384_barcodes.fasta) was attached to the 5′ end of each *CAN1*-specific primer and DNA of each mutant was amplified using a unique pair of barcodes. Barcodes for Can-R mutants isolated from the low-transcription, *pCAN-CAN1* strains or high-transcription *pGAL-CAN1* strains were attached to forward primers 5′-CAGTTTTTAATCTGTCGTCAATCGAAAG or 5′-GCCGAGCGGGTGACAG, respectively. Barcodes were attached to single *CAN1* reverse primer (5′-GGGAGCAAGATTGTTGTGGT). Following product quantification using ImageJ software (https://imagej.nih.gov/ij/), barcoded amplicons were pooled in equal amounts for

subsequent SMRT library preparation. SMRT libraries were constructed and sequenced at Duke Center for Genomic and Computational Biology using the PacBio RSII and Sequel systems. Circular consensus reads were sorted and analyzed using an in-house pipeline [35]. The rates of C>T and G>A mutations were calculated as the product of the total Can-R rate and the proportion of the mutation type in the corresponding spectrum. 95% CIs were assigned to mutation proportions using the Vassarstats web site (vassarstats.net). To assign upper and lower confidence intervals to a C>T or G>A mutation rate, the 95% CIs of both the corresponding Can-R rate and the mutation proportion were combined using the root of the square of the sums [36].

For experiments involving the over-expression of RNase H1, the high-transcription *ung1*Δ mutant was transformed with empty vector or with a plasmid in which *RNH1* was regulated by *pGAL1*. Mutation rates were calculated and Can-R mutants were isolated as described above, except that all pre-growth was in SC medium containing 2% glycerol and 2% EtOH as carbon sources and missing uracil in order to maintain plasmids. Following DNA isolation and PCR amplification of *CAN1* (forward and reverse primers 5′-TTGGTATGATTGCCCTTGGT and 5′-ACAGCGTTGAAGATATGTGGC, respectively), nonsense mutations between positions 412 and 1056 were captured by Sanger sequencing (primer 5′-TATCCACACCTCTGACCAAC). This region contains six CAG/CAA codons and nine TGG codons to monitor C>T and G>A nonsense mutations, respectively.

2.3. DNA-RNA Immunoprecipitation (DRIP)

Cultures (100 ml) grown to mid-log phase in YEP containing 2% galactose were washed with chilled H_2O, resuspended in 1.4 ml spheroplasting buffer (1 M sorbitol, 10 mM EDTA pH 8, 0.1% β-mercaptoethanol, 2 mg/ml Zymolase 20T) and incubated at 30° for 30 min. Spheroplasts were pelleted (5 min at 4000 rpm), rinsed with H_2O and resuspended in 1.65 ml of buffer G2 (800 mM guanidine HCl, 30 mM Tris-Cl pH 8, 30 mM EDTA pH 8, 5% Tween-20, 0.5% Triton X-100). Samples were treated with 10 μl of 10 mg/ml RNase A for 30 min at 37° and then with 75 μl of 20 mg/ml proteinase K (Roche) for 1 h at 50°. Genomic DNA was extracted gently with chloroform:isoamyl alcohol (24:1).

DRIP was performed as previously described [37] with the following modifications. Precipitated genomic DNA was spooled onto a glass rod, washed twice with 70% EtOH, resuspended gently in 250 μl TE (10 mM Tris-HCl pH 8, 1 mM EDTA pH 8) containing 2 mM spermidine and 2.5 μl of 10 mg/ml BSA, and digested overnight with 50U each of *Hind*III, *EcoR*I, *BsrG*I, *Xba*I and *Ssp*I. Half of the DNA was treated with 3 μl RNase H1 (Invitrogen Ambion; 1U/μl) overnight at 37° for an RNase H1-treated control. RNase H1-treated and untreated samples were incubated with 10 μl of Dynabead-coupled S9.6 monoclonal antibody (1 mg/ml) in 500 μl binding buffer (10 mM $NaPO_4$ pH 7.0, 140 mM NaCl, 0.05% Triton X-100) overnight at 4°C and then washed three times with binding buffer. The S9.6 antibody was coupled to Dynabeads Protein A (Invitrogen) for 2 h at 4°. DNA was eluted in 100 μl elution buffer (50 mM Tris pH 8.0, 10 mM EDTA, 0.5% SDS), treated 45 min with 7 μl proteinase K (20 mg/ml) at 55° and purified with the Macherey-Nagel DNA purification kit. Real-time quantitative PCR was performed using iTaq Universal SYBR Green (Biorad) and a 7500 Real-Time PCR machine (Applied Biosystems).

Forward and reverse primers for *PDC1* were 5′-GAAGGTATGAGATGGGCTGGTAA and 5’ CCTTGATACGAGCGTAACCATCA, respectively. *CAN1*-specific forward and reverse primers were 5′ GGTTTTCTTGGGCAATCACTT and 5′ AATTGAATGACTTGGCCAACTACA, respectively. Three biological replicates were performed for each strain; mean and standard deviations are reported. As a positive control for R-loop detection in DRIP experiments, genomic DNA extracted from W303 and an *spt16–11* derivative was used.

2.4. Whole genome analysis of spontaneous deamination sites

The data for 162 C>T mutations detected from whole-genome sequencing of an *ung1* strain passaged for 81 days was downloaded from Chen *et al.* [38]. The reported genome location and flanking sequences were used to map the exact location of mutations in the Saccharomyces Genome Database (version 6CA_00146045.2). The deaminated strand (NTS or TS) was determined for mutations in RNAPII-transcribed genes using release 91 of the Ensembl genome browser. The transcript level of each gene was downloaded [39, 40] and genes were ranked based on expression level.

2.5. Additional statistical analyses

The numbers of C>T versus G>A mutations observed in different strain backgrounds were compared using either a 2×2 contingency chi-square test or a one-tailed Fisher exact test, as appropriate (vassarstats.net). These tests were also used to compare the relationship between transcript level and NTS/TS deamination in the whole-genome analyses. Chi-square goodness of fit was used to compare observed to expected numbers of events. Rates were significantly differently if the rate in each individual strain did not overlap the corresponding rate-associated CI of the other strain.

3. RESULTS

Relative levels of spontaneous DNA damage associated with transcription of a target gene can be inferred by examining the mutagenic consequences. An inherent complication with most base damages is that they or their repair intermediates slow or stall subsequent DNA replication and transcription. This can trigger specialized bypass mechanisms during replication and/or template-specific engagement of nucleotide excision repair during transcription, which complicates subsequent analyses and inferences. The hydrolytic deamination of cytosine to uracil is an exception, however, because the base-pairing properties of uracil are identical to those of thymine [41]. In wild-type (WT) yeast cells, uracil is efficiently removed by the Ung1 uracil-DNA N-glycosylase and is only rarely mutagenic. In the absence of Ung1 deaminated cytosine persists as part of a U:G mismatch and the uracil is replicated as a thymine, leading to CG>TA transition mutations [42].

3.1. Transcription increases cytosine deamination in *CAN1*

As a reporter for spontaneous cytosine deamination we used the yeast *CAN1* gene, which encodes an arginine permease that transports the toxic analog canavanine and confers canavanine sensitivity. Any forward mutation that inactivates the encoded protein results in canavanine resistance (Can-R) on drug-containing medium. To examine effects of

transcription on spontaneous deamination, strains were used in which the *CAN1* gene was either under control of its native promoter or fused to the highly active *GAL1* promoter (low- and high-transcription conditions, respectively; [5]). Under low-transcription (low-txn) conditions, 31% (69/220) of Can-R mutants isolated in an *UNG1* background contained a CG>TA base substitution (Tables S2-S3). We note that this is higher than the CG>TA proportion among Can-R mutants reported previously (60/277; p=0.018; [43]). This difference could reflect either strain background differences or the use of ten times the standard amount of canavanine in the previous study. Deletion of *UNG1* was accompanied by a 1.6-fold increase in the Can-R rate (Table S4) as well as a proportional increase in CG>TA mutations from 31% to 88% (217/246; p<0.0001). This translates into a 4.6-fold increase in the CG>TA rate in the absence of Ung1 (Figure 1) and confirms that most CG>TA mutations arising in an *ung1*Δ background reflect cytosine deamination.

Highly elevated transcription resulted in an ~12-fold increase in Can-R rate the WT background and deletion of *UNG1* was associated with an additional 1.9-fold rate increase. As under low-txn conditions, there was a proportional increase in CG>TA mutations in the *ung1*Δ background under high-txn conditions: from 6% (41/663) to 45% (249/554; p<0.0001). The rate of CG>TA mutations was 7-fold higher under high-txn than under low-txn conditions in the *ung1*Δ background, demonstrating that transcription increases deamination of the corresponding DNA (Figure 1). The lower proportion of CG>TA mutations under high- than low-txn conditions (45% and 88%, respectively; p<0.0001) reflects the large contribution of Top1-dependent 2–5 bp deletions to the spectrum under high high-txn conditions [5, 6].

3.2. Transcription elevates nonsense mutations more than missense mutations

An inherent problem with forward-mutation assays is that missense mutations, which alter a single amino acid, are often of no functional consequence. Changes that convert a sense to a nonsense/stop codon truncate the corresponding protein product, however, and almost always disable function. At the *CAN1* locus it was previously estimated that only ~5% of missense mutations confer canavanine resistance [43]. The missense-detection issue potentially can be exacerbated under high-txn conditions where there may be enough of a highly dysfunctional protein to confer a canavanine-sensitive phenotype. Among Can-R mutants analyzed, we indeed observed proportionally fewer missense mutations among CG>TA changes detected under high- than under low-txn conditions (83/249 and 114/217, respectively; p<0.0001; Tables S2 and S3). Because of the inability to reliably detect missense mutations under high-txn conditions, only CG>TA mutations that created a TAA, TAG or TGA stop codon were considered in subsequent analyses.

3.3. Transcription-associated deamination is biased to the NTS

By convention, mutations are reported in relation to the coding/sense strand of a gene, which has the same sequence as the mRNA and corresponds to the non-transcribed strand (NTS). Although reporting mutations in this way allows a strand bias to be easily discerned, it often provides no information as to the source of the bias. This can be inferred, however, using DNA polymerases with specific misincorporation patterns [44], disabling a repair pathway that is highly specific for damage to only one of the complementary bases [45] or creating a

single-stranded DNA target [46]. Thus, deamination of cytosine on the NTS leads to a C>T change while deamination of a cytosine located on the transcribed strand (TS) results in a G>A mutational read-out (Figure 2A).

C>T mutations in sense codons generate all three stop codons: CAA>TAA, CAG>TAG and CGA>TGA. By contrast, G>A mutations create only TGA and TAG stop codons, both of which are derived from the TGG sense codon (TGG>TGA and TGG>TAG); a TAA stop codon arises only *via* a G>A change in a preexisting TAG or TGA stop codon. Within the *CAN1* ORF, there are 12 CAA/CAG/CGA codons plus 16 TGG codons (Figure S1) and most were frequent sites of C>T or G>A mutations, respectively. Exceptions were the four TGG codons nearest the 3′ end of the gene, mutation of which presumably generates a truncated but functional protein, and a TGG located 5′ to other positions where nonsense mutations were frequent. Mutations at these five codons were also absent from the spectra of Lang and Murray [43]. Excluding these five TGG codons, there were 12 and 22 positions (22 guanines within the remaining 11 TGG codon targets) in *CAN1* where C>T and G>A mutations, respectively, were monitored.

Even though proportionally fewer C>T than G>A mutations that created stop codons were expected (12/34 and 22/34, respectively), similar numbers of C>T and G>A mutations were observed in the low-txn *ung1*Δ background: 55 and 48, respectively (Table S4; p=0.02). On a per site basis, there were 4.6 C>T changes at cytosines where deamination was monitored (55 mutations at 12 sites) and 2.2 G>A mutations at monitored guanines (48/22). There is thus an approximately 2-fold C>T bias even under low-txn conditions. The corresponding rates of C>T and G>A events, however, were very similar and are shown in Figure 2B. In contrast to similar numbers of C>T and G>A mutations under low-txn conditions, there was a 3.4-fold excess of C>T relative to G>A mutations in the high-txn *ung1*Δ background: 128 and 38, respectively (p<0.0001). On a per site basis, there were 10.7 C>T changes at cytosines where deamination was monitored (128 mutations at 12 sites) and 1.7 G>A mutations at monitored guanines (38/22). In terms of rates, C>T mutations were elevated 14.5-fold under high-txn conditions while G>A mutations were elevated only 4.8-fold (Figure 2B).

3.4. The transcription-associated C>T bias reflects neither enzymatic deamination nor biased U:G repair

Cytosine in single-stranded DNA spontaneously deaminates more rapidly than cytosine in duplex DNA [47], suggesting that the strand-specific deamination bias observed under high-txn conditions reflects the enhanced single-stranded character of the NTS. An alternative possibility, however, is strand-biased deamination of cytosine by the yeast Fcy1 protein, which has been reported to deaminate DNA as well as RNA [48]. If Fcy1 were the primary source of cytosine deamination, then the rate of C>T mutations should be reduced in its absence. There was no change in the rates (Figure 2B; Table S4) or distribution (p=0.12; Tables S2-S3) of C>T and G>A mutations in the *ung1*Δ *fcy1*Δ double mutant relative to the *ung1*Δ single mutant, consistent with strand-biased hydrolytic deamination of cytosine.

The mismatch repair (MMR) system removes errors that arise during DNA replication by initiating excision and re-synthesis of the newly synthesized strand (reviewed in

[49]). Because U:G mismatches due to cytosine deamination occur outside the context of replication, no preference for removing the U- *versus* the G-containing strand would be expected [50]. It is possible, however, that transcription could confer a strand-specific asymmetry in the repair of U:G mismatches. We thus examined the rates of C>T and G>A mutations in the absence of Msh6, a component of the MutSα heterodimer that detects and initiates repair of most base-base mismatches. As expected, deletion of *MSH6* was associated with a stronger mutator phenotype than was *UNG1* deletion in the high-txn background (9.9- and 1.9-fold increases in the Can-R rate, respectively; Table S4). For CG>TA mutations, however, the rate increases relative to WT were comparable in the *msh6*Δ and *ung1*Δ single mutants (16- and 14-fold increases, respectively; Figure 2B). In contrast to the greater number of C>T than G>A mutations in the high-txn *ung1*Δ background (128 and 38, respectively), C>T mutations were much less frequent than G>A mutations in the *msh6*Δ background (1 and 20, respectively; $p<0.0001$). The large increase in C>T relative to G>A mutations rate observed when *UNG1* was deleted in the WT background, however, was recapitulated when *UNG1* was deleted in the *msh6*Δ background (13 C>T and 12 G>A mutations; $p=0.0009$). We conclude that the C>T bias observed under high-txn conditions directly reflects spontaneous cytosine deamination rather than an MMR-related bias during repair of the resulting U:G mismatches.

3.5. The C>T bias under high-txn conditions does not correlate with R-loop accumulation

Pairing between an RNA and its DNA template occurs in the small RNAP-associated transcription bubble and in more extensive R-loops. To investigate whether R-loops are the primary driver of the elevated NTS deamination at *CAN1*, we examined the transcription-associated C>T bias in genetic backgrounds where R-loop levels are expected to increase. To potentially promote more R-loop formation we deleted the *MFT1* gene, the product of which is part of the THO complex and is important for the normal sequestration of transcripts into ribonucleoprotein particles [31]. We also deleted *TOP1*, which discourages R-loop formation by removing the negative supercoils associated with DNA underwinding behind the transcription machinery [51]. Finally, to stabilize R-loops once formed, the enzymatic degradation of the RNA component of RNA-DNA hybrids was blocked by deleting *RNH1* and *RNH201*, which encode RNase H1 and the catalytic subunit of RNase H2, respectively (reviewed in [52]). If R-loops are the primary source of the NTS deamination bias, we expected an even stronger C>T bias in backgrounds that favor R-loop formation/stability. Relative to the bias in the *ung1*Δ single-mutant background (128 C>T and 38 G>A mutations), however, the NTS bias was not enhanced in the *ung1*Δ *top1*Δ (98 C>T and 41 G>A; $p=0.24$), *ung1*Δ *rnh1*Δ *rnh201*Δ (51 C>T and 21 G>A; $p= 0.39$) or *ung1*Δ *mft1*Δ (72 C>T and 35 G>A; $p=0.10$) background. The rates of C>T and G>A changes in these backgrounds are shown in Figure 3A (see Table S4).

As further genetic confirmation that R-loops are not the primary source of the strand-biased deamination at *CAN1*, we examined the effect of over-expressing RNase H1 on the accumulation of C>T *versus* G>A mutations. This approach is frequently used to eliminate R-loops [31] and is the complement to genetically increasing R-loop levels (see above). The high-txn *ung1*Δ strain was transformed with empty vector (EV) or with a plasmid containing a *GAL-RNH1* fusion and cells were grown in minimal medium to maintain the plasmids.

For this analysis, only the central portion of the *CAN1* gene was sequenced; this region contains six CAG/CAA codons and nine TGG codons, which corresponds to six possible C>T changes and 18 G>A changes that can produce a stop codon. There was no effect of *RNH1* over-expression on the overall Can-R rate or on relative numbers of C>T and G>A events in the region monitored (Tables S2-S4). The rates of C>T and G>A events are presented in Figure 3B.

R-loops do not form uniformly in all regions of the yeast genome; sequences prone to R-loop formation are generally GC-rich and have a high rate of transcription [15]. To complement genetic analyses, R-loop accumulation at *CAN1* was examined by quantitative PCR following DNA-RNA immunoprecipitation (DRIP) with the S9.6 antibody [53]. The *PDC1* locus, where R-loops have been detected in other studies [54] was examined in parallel. At the *CAN1* locus there was neither an increase in R-loop accumulation as a function of transcription in WT cells, nor an increase in in R-loops in genetic backgrounds where an increase might have been expected (Figure 4). There similarly was no increase in R-loops at *PDC1* in the mutant backgrounds examined. As a positive control for R-loop detection, *PDC1* was examined in an *spt16–11* derivative of W303; RNA:DNA hybrids were elevated, as expected [55]. Based on the cumulative analyses, there is no indication that R-loops are the primary driver of the strand-biased cytosine deamination associated with transcription of the *CAN1* reporter.

3.6. Genome-wide analysis of transcription-associated cytosine deamination

The relationship between cytosine deamination and transcription was examined genome-wide using data from *ung1*Δ mutation-accumulation lines [38] and transcript profiles obtained by single-molecule sequencing [39]. Among the 162 CG>TA mutations identified in mutation-accumulation lines, 88 were in open-reading frames (ORFs). Each of these mutations was assigned as originating from deamination of the TS or NTS and there was expression data for 87 of the 88 ORFs (Table S5). Among the transcribed genes, there were 54 C>T mutations and 33 G>A mutations, which is significantly different from the equivalent numbers expected if there is unbiased deamination of the NTS and TS (p=0.032). We also separately ranked the genes with a C>T *versus* G>A mutation by relative transcript level. The median transcript level was 83 and 51 for genes containing a C>T and G>A mutation, respectively, and the corresponding transcript distributions were significantly different (p=0.009 by Mann-Whitney U-test). These data suggest that the NTS bias for cytosine deamination observed locally at *CAN1* is a general feature of RNAPII-transcribed genes that correlates globally with transcription level.

4. DISCUSSION

The goal of the current study was to determine whether there is a transcription-associated strand bias in the hydrolytic deamination of cytosine to uracil. The *CAN1* gene was used as a forward-mutation reporter, with C>T and G>A mutations reflecting NTS and TS deamination, respectively. Under high-txn conditions we observed proportionally fewer missense than nonsense mutations in *CAN1*, suggesting that increased expression can compensate for reduced protein function. To eliminate expression-related effects on

mutation detection, only C>T and G>A changes that created nonsense mutations were considered.

Whereas C>T and G>A mutations occurred at similar rates under low-txn conditions, there was a strong bias for C>T mutations under high-txn conditions. A similar NTS bias for enzymatic cytosine deamination by AID or APOBEC proteins has been associated with co-transcriptional R-loops [20, 23], which leave the NTS in an unpaired and vulnerable state (Figure 5). Despite the general correlation between R-loop accumulation and transcription in genome-wide yeast studies [24], there are numerous factors such as base composition, propensity for secondary structure formation on the NTS, the direction of replication and chromatin modifications that affect when and where R-loops form (reviewed in [15]). Wth regard to the spontaneous deamination bias, we were unable to detect a transcription-associated increase in RNA-DNA hybrids at the *CAN1* locus. In addition, the use of genetic backgrounds where R-loops are expected to form more readily (*top1*Δ or *mft1*Δ) or to be more stable once formed (*rnh1*Δ *rnh201*Δ) neither exacerbated the C>T bias nor led to detecteable R-loops. The apparent lack on a role for R-loops was further confirmed by over-production of RNase H1, which did not alter the C>T bias.

Although we focused on a strand-specific bias in cytosine deamination, it should be noted that G>A as well as C>T mutations were elevated under high-txn conditions. This indicates increased hydrolytic damage of both DNA strands, which could reflect the enhanced single-strand character of underwound, negatively supercoiled DNA that forms in the wake of the transcription machinery. There was no further increase in C>T or G>A mutations, however, in the absence of Top1. An alternative possibility inferred from *in vitro* studies is that the backtracking of RNA polymerase might expose both DNA strands to damage [18].

There are not yet estimates of how frequently R-loops form during transcription, but they are expected to be rare. By contrast, a small transcription bubble forms with the passage of each RNAP. The melting of duplex DNA when it enters RNAP potentially renders both strands more susceptible to damage, while subsequent pairing between of the TS with ~8 nt of nascent RNA affords some protection (Figure 5). I*n vitro* studies with bacterial RNAP have shown that the NTS in the back half of an arrested elongation complex can be cleaved by micrococcal nuclease, indicating enhanced enzymatic accessibility [56]. Furthermore, the crystal structure of yeast RNAPII with a nucleic acid scaffold that mimics a transcription bubble revealed that part of the NTS is on the surface of the enzyme [57]. A least in the case of *CAN1*, we suggest that the biased deamination of the NTS occurs primarily in the context of transcription bubbles. An interesting possibility is that frequent pausing of RNAPII at defined positions *in vivo* [58] could further modify the landscape of spontaneous DNA damage. Although R-loops do not appear to be responsible for the strand-specific bias in spontaneous deamination within the *CAN1* gene, they may nevertheless be the major driver in other genetic contexts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank Tom Petes and members of the SJR lab for discussions throughout the course of this work and for helpful comments on the manuscript. Work in the SJR was supported by National Institutes of Health grant R35GM118077 and JDW was supported by a Tri-Institutional Molecular Mycology and Pathogenesis Training Program (5T32AI052080) postdoctoral fellowship. Work in the AA lab was supported by the grant PID2019-104270GB-I00/BMC from the Agencia Estatal de Investigación of the Spanish Ministry of Science and Innovation MCIN/AEI/10.13039/501100011033.

REFERENCES

[1]. Lodato MA, Woodworth MB, Lee S, Evrony GD, Mehta BK, Karger A, Lee S, Chittenden TW, D'Gama AM, Cai X, Luquette LJ, Lee E, Park PJ, Walsh CA, Somatic mutation in single human neurons tracks developmental and transcriptional history. Science 350 (2015) 94–98. [PubMed: 26430121]

[2]. Jinks-Robertson S, Bhagwat AS, Transcription-associated mutagenesis. Annu. Rev. Genet 48 (2014) 341–359. [PubMed: 25251854]

[3]. Datta A, Jinks-Robertson S, Association of increased spontaneous mutation rates with high levels of transcription in yeast. Science 268 (1995) 1616–1619. [PubMed: 7777859]

[4]. Garcia-Rubio M, Huertas P, Gonzalez-Barrera S, Aguilera A, Recombinogenic effects of DNA-damaging agents are synergistically increased by transcription in Saccharomyces cerevisiae: New insights into transcription-associated recombination. Genetics 165 (2003) 457–466. [PubMed: 14573461]

[5]. Lippert MJ, Kim N, Cho JE, Larson RP, Schoenly NE, O'Shea SH, Jinks-Robertson S, Role for topoisomerase 1 in transcription-associated mutagenesis in yeast. Proc. Natl. Acad. Sci. USA 108 (2011) 698–703. [PubMed: 21177427]

[6]. Takahashi T, Burguiere-Slezak G, Van der Kemp PA, Boiteux S, Topoisomerase 1 provokes the formation of short deletions in repeated sequences upon high transcription in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA 108 (2011) 692–697. [PubMed: 21177431]

[7]. Sparks JL, Burgers PM, Error-free and mutagenic processing of topoisomerase 1-provoked damage at genomic ribonucleotides. EMBO J 34 (2015) 1259–1269. [PubMed: 25777529]

[8]. Huang SY, Ghosh S, Pommier Y, Topoisomerase I alone is sufficient to produce short DNA deletions and can also reverse nicks at ribonucleotide sites. J. Biol. Chem 290 (2015) 14068–14076. [PubMed: 25887397]

[9]. Lujan SA, Williams JS, Kunkel TA, Eukaryotic genome instability in light of asymmetric DNA replication. Crit. Rev. Biochem. Mol. Biol 51 (2016) 43–52. [PubMed: 26822554]

[10]. Haradhvala NJ, Polak P, Stojanov P, Covington KR, Shinbrot E, Hess JM, Rheinbay E, Kim J, Maruvka YE, Braunstein LZ, Kamburov A, Hanawalt PC, Wheeler DA, Koren A, Lawrence MS, Getz G, Mutational strand asymmetries in cancer genomes reveal mechanisms of DNA damage and repair. Cell 164 (2016) 538–549. [PubMed: 26806129]

[11]. Pani B, Nudler E, Mechanistic insights into transcription coupled DNA repair. DNA Repair 56 (2017) 42–50. [PubMed: 28629777]

[12]. Lindahl T, Instability and decay of the primary structure of DNA. Nature 362 (1993) 709–715. [PubMed: 8469282]

[13]. Chan K, Gordenin DA, Clusters of multiple mutations: incidence and molecular mechanisms. Annu. Rev.Genet 49 (2015) 243–267. [PubMed: 26631512]

[14]. Roy D, Yu K, Lieber MR, Mechanism of R-loop formation at immunoglobulin class switch sequences. Mol. Cell. Biol 28 (2008) 50–60. [PubMed: 17954560]

[15]. García-Muse T, Aguilera A, R Loops: from physiological to pathological roles. Cell 179 (2019) 604–618. [PubMed: 31607512]

[16]. Maizels N, Immunoglobulin gene diversity. Annu. Rev. Genet 39 (2005) 23–46. [PubMed: 16285851]

[17]. Senavirathne G, Bertram JG, Jaszczur M, Chaurasiya KR, Pham P, Mak CH, Goodman MF, Rueda D, Activation-induced deoxycytidine deaminase (AID) co-transcriptional scanning at single-molecule resolution. Nat. Commun 6 (2015) 10209. [PubMed: 26681117]

[18]. Pham P, Malik S, Mak C, Calabrese PC, Roeder RG, Goodman MF, AID-RNA polymerase II transcription-dependent deamination of IgV DNA. Nucleic Acids Res 47 (2019) 10815–10829. [PubMed: 31566237]

[19]. Zhang ZZ, Pannunzio NR, Han L, Hsieh CL, Yu K, Lieber MR, The strength of an Ig switch region is determined by its ability to drive R loop formation and its number of WGCW sites. Cell Rep 8 (2014) 557–569. [PubMed: 25017067]

[20]. Gomez-Gonzalez B, Aguilera A, Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. Proc. Natl. Acad. Sci. USA 104 (2007) 8409–8414. [PubMed: 17488823]

[21]. Harris RS, Dudley JP, APOBECs and virus restriction. Virology 479–480 (2015) 131–145.

[22]. Mertz TM, Collins CD, Dennis M, Coxon M, Roberts SA, APOBEC-induced mutagenesis in cancer. Annu. Rev. Genet (2022).

[23]. Saini N, Roberts SA, Sterling JF, Malc EP, Mieczkowski PA, Gordenin DA, APOBEC3B cytidine deaminase targets the non-transcribed strand of tRNA genes in yeast. DNA Repair 53 (2017) 4–14. [PubMed: 28351647]

[24]. Wahba L, Costantino L, Tan FJ, Zimmer A, Koshland D, S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. Genes Dev 30 (2016) 1327–1338. [PubMed: 27298336]

[25]. Lada AG, Kliver SF, Dhar A, Polev DE, Masharsky AE, Rogozin IB, Pavlov YI, Disruption of transcriptional coactivator Sub1 leads to genome-wide re-distribution of clustered mutations induced by APOBEC in active yeast genes. PLoS Genet 11 (2015) e1005217. [PubMed: 25941824]

[26]. Mokkapati SK, Bhagwat AS, Lack of dependance of transcription-induced cytosine deaminations on protein synthesis. Mutat. Res 508 (2002) 131–136. [PubMed: 12379468]

[27]. Gómez-González B, García-Rubio M, Bermejo R, Gaillard H, Shirahige K, Marín A, Foiani M, Aguilera A, Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. EMBO J 30 (2011) 3106–3119. [PubMed: 21701562]

[28]. Sankar TS, Wastuwidyaningtyas BD, Dong Y, Lewis SA, Wang JD, The nature of mutations induced by replication-transcription collisions. Nature 535 (2016) 178–181. [PubMed: 27362223]

[29]. Longtine MS, McKenzie III A, Demarini DJ, Shah NG, Wach A, Brachat A, Philippsen P, Pringle JR, Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast 14 (1998) 953–961. [PubMed: 9717241]

[30]. Sikorski RS, Hieter P, A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics 122 (1989) 19–27. [PubMed: 2659436]

[31]. Huertas P, Aguilera A, Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol. Cell 12 (2003) 711–721. [PubMed: 14527416]

[32]. Torchia TE, Hamilton RW, Cano CL, Hopper JE, Disruption of regulatory gene GAL80 in Saccharomyces cerevisiae: Effects on carbon-controlled regulation of the galactose/melibiose pathway genes. Mol. Cell. Biol 4 (1984) 1521–1527. [PubMed: 6092916]

[33]. Lea DE, Coulson CA, The distribution of the numbers of mutants in bacterial populations. J. Genet 49 (1949) 264–285. [PubMed: 24536673]

[34]. Spell RM, Jinks-Robertson S, Determination of mitotic recombination rates by fluctuation analysis in Saccharomyces cerevisiae, in Genetic Recombination: Reviews and Protocols, Waldman AS, Editor. 2004, Humana Press: Totowa, NJ. p. 3–12.

[35]. Guo X, Lehner K, O'Connell K, Zhang J, Dave SS, Jinks-Robertson S, SMRT sequencing for parallel analysis of multiple targets and accurate SNP phasing. G3 (Bethesda) 5 (2015) 2801–2808. [PubMed: 26497143]

[36]. Moore A, Dominska M, Greenwell P, Aksenova AY, Mirkin S, Petes T, Genetic control of genomic alterations induced in yeast by interstitial telomeric sequences. Genetics 209 (2018) 425–438. [PubMed: 29610215]

[37]. Ginno PA, Lott PL, Christensen HC, Korf I, Cheden F, R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol. Cell 45 (2012) 814–825. [PubMed: 22387027]

[38]. Chen X, Chen Z, Chen H, Su Z, Yang J, Lin F, Shi S, He X, Nucleosomes suppress spontaneous mutations base-specifically in eukaryotes. Science 335 (2012) 1235–1238. [PubMed: 22403392]

[39]. Lipson D, Raz T, Kieu A, Jones DR, Giladi E, Thayer E, Thompson JF, Letovsky S, Milos P, Causey M, Quantification of the yeast transcriptome by single-molecule sequencing. Nat. Biotechnol 27 (2009) 652–658. [PubMed: 19581875]

[40]. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M, The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320 (2008) 1344–1349. [PubMed: 18451266]

[41]. Wardle J, Burgers PM, Cann IK, Darley K, Heslop P, Johansson E, Lin LJ, McGlynn P, Sanvoisin J, Stith CM, Connolly BA, Uracil recognition by replicative DNA polymerases is limited to the archaea, not occurring with bacteria and eukarya. Nucleic Acids Res 36 (2008) 705–711. [PubMed: 18032433]

[42]. Impellizzeri KJ, Anderson B, Burgers PM, The spectrum of spontaneous mutations in a Saccharomyces cerevisiae uracil-DNA-glycosylase mutant limits the function of this enzyme to cytosine deamination repair. J. Bacteriol 173 (1991) 6807–6810. [PubMed: 1938887]

[43]. Lang GI, Murray AW, Estimating the per-base-pair mutation rate in the yeast Saccharomyces cerevisiae. Genetics 178 (2008) 67–82. [PubMed: 18202359]

[44]. Nick McElhinny SA, Gordenin DA, Stith CM, Burgers PM, Kunkel TA, Division of labor at the eukaryotic replication fork. Molecular Cell 30 (2008) 137–144. [PubMed: 18439893]

[45]. Pavlov Y, Newlon CS, Kunkel TA, Yeast origins establish a strand bias for replicational mutagenesis. Mol. Cell 10 (2002) 207–213. [PubMed: 12150920]

[46]. Yang Y, Sterling J, Storici F, Resnick MA, Gordenin DA, Hypermutability of damaged single-strand DNA formed at double-strand breaks and uncapped telomeres in yeast Saccharomyces cerevisiae. PLoS Genet 4 (2008) e1000264. [PubMed: 19023402]

[47]. Frederico LA, Kunkel TA, Shaw BR, A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. Biochemistry 29 (1990) 2532–2537. [PubMed: 2185829]

[48]. Su XA, Freudenreich CH, Cytosine deamination and base excision repair cause R-loop-induced CAG repeat fragility and instability in Saccharomyces cerevisiae. Proc. Natl .Acad. Sci. USA 114 (2017) E8392–E8401. [PubMed: 28923949]

[49]. Kunkel TA, Erie DA, DNA mismatch repair. Annu. Rev. Biochem 74 (2005) 681–710. [PubMed: 15952900]

[50]. Rodriguez GP, Romanova NV, Bao G, Rouf NC, Kow YW, Crouse GF, Mismatch repair-dependent mutagenesis in nondividing cells. Proc. Natl. Acad. Sci. USA 109 (2012) 6153–6158. [PubMed: 22474380]

[51]. Pommier Y, Sun Y, Huang SN, Nitiss JL, Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. Nat. Rev. Mol. Cell Biol 17 (2016) 703–721. [PubMed: 27649880]

[52]. Cerritelli SM, Crouch RJ, Ribonuclease H: the enzymes in eukaryotes. FEBS J 276 (2009) 1494–1505. [PubMed: 19228196]

[53]. Boguslawski SJ, Smith DE, Michalak MA, Mickelson KE, Yehle CO, Patterson WL, Carrico RJ, Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. J Immunol Methods 89 (1986) 123–130. [PubMed: 2422282]

[54]. García-Benítez F, Gaillard H, Aguilera A, Physical proximity of chromatin to nuclear pores prevents harmful R loop accumulation contributing to maintain genome stability. Proc. Natl. Acad. Sci. USA 114 (2017) 10942–10947. [PubMed: 28973905]

[55]. Herrera-Moyano E, Mergui X, Garcia-Rubio ML, Barroso S, Aguilera A, The yeast and human FACT chromatin-reorganizing complexes solve R-loop-mediated transcription-replication conflicts. Genes Dev 28 (2014) 735–748. [PubMed: 24636987]

[56]. Wang D, Landick R, Nuclease cleavage of the upstream half of the nontemplate strand DNA in an Escherichia coli transcription elongation complex causes upstream translocation and transcriptional arrest. J. Biol. Chem 272 (1997) 5989–5994. [PubMed: 9038220]

[57]. Barnes CO, Calero M, Malik I, Graham BW, Spahr H, Lin G, Cohen AE, Brown IS, Zhang Q, Pullara F, Trakselis MA, Kaplan CD, Calero G, Crystal structure of a transcribing RNA polymerase II complex reveals a complete transcription bubble. Mol. Cell 59 (2015) 258–269. [PubMed: 26186291]

[58]. Churchman LS, Weissman JS, Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 469 (2011) 368–373. [PubMed: 21248844]

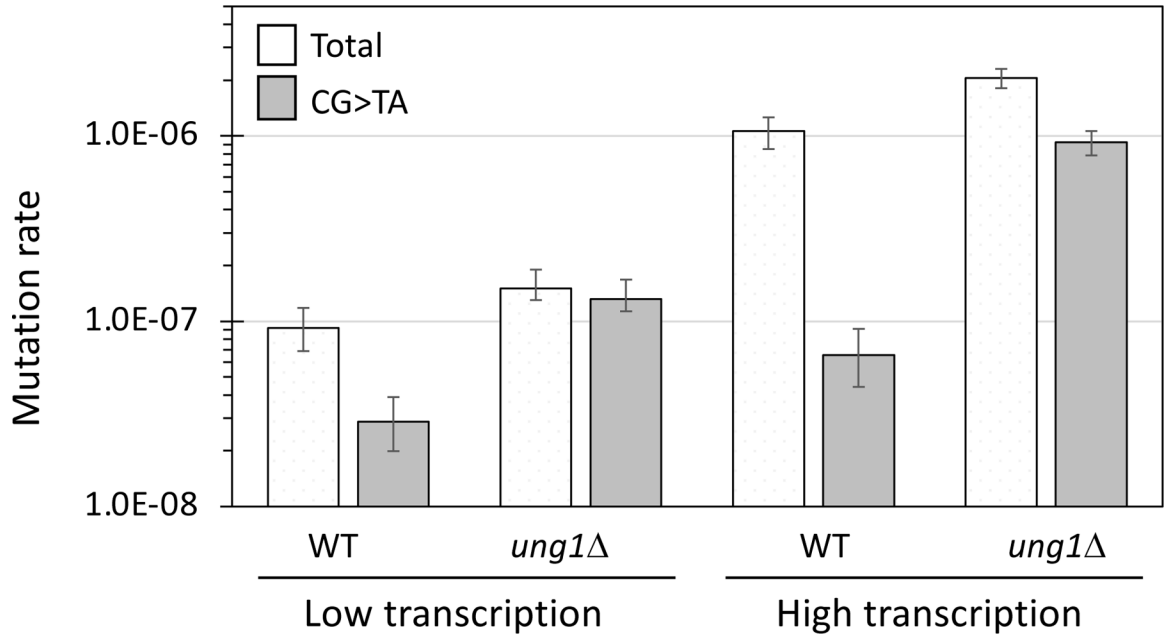

Figure 1.

Loss of Ung1 preferentially elevates CG>TA mutations in *CAN1* under low- and high-txn conditions. White bars correspond to the total mutation rate and gray bars to the rate of CG>TA mutations. Error bars are the 95% confidence interval (CI).

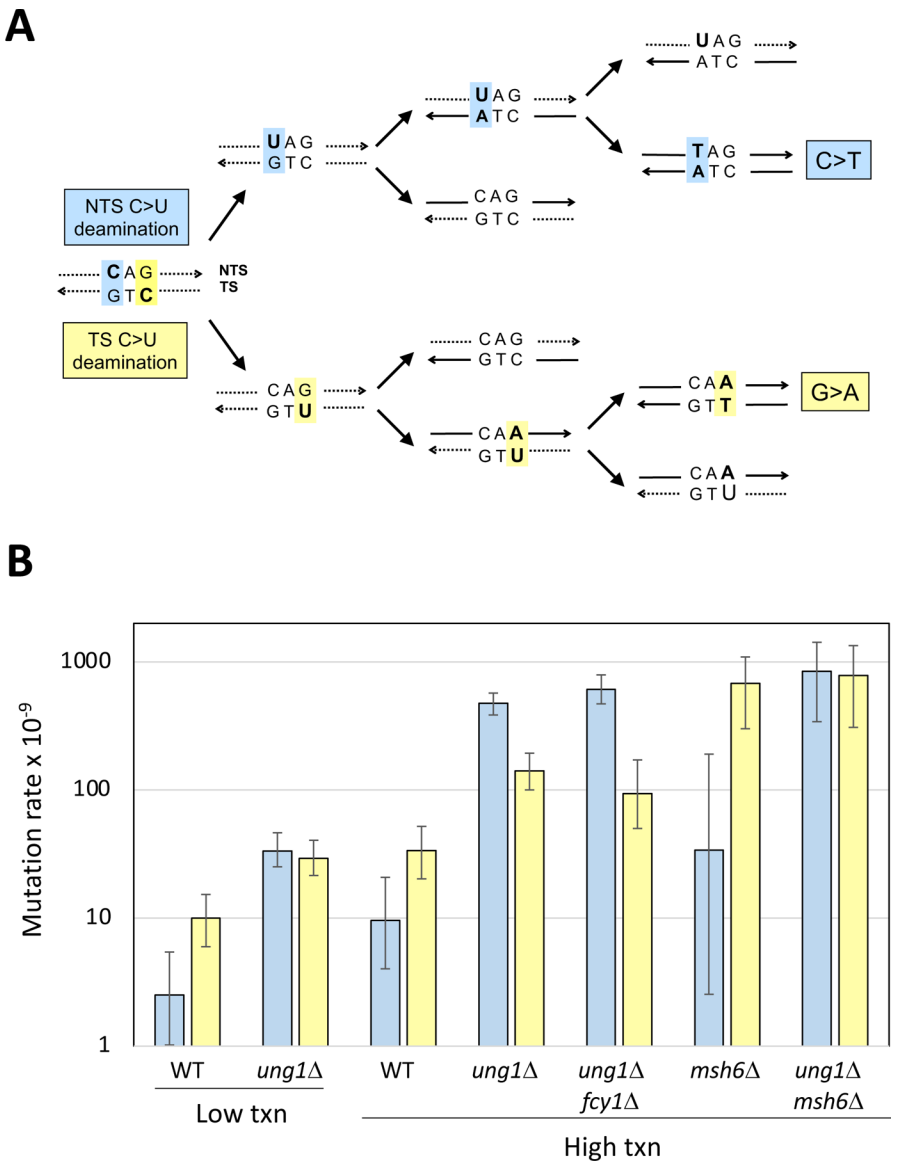

Figure 2.

Strand-specific mutation accumulation. (**A**) Cytosine deamination on complementary DNA strands yields C>T and G>A mutations. In an *ung1*Δ background, a uracil that results from cytosine deamination persists and templates dAMP incorporation during subsequent DNA synthesis. After two rounds of replication, a deaminated cytosine on the NTS (blue) or TS (yellow) gives rise to a C>T or G>A mutation, respectively, when the NTS sequence is reported. Dotted lines represent the starting DNA strands and arrowheads correspond to 3′ ends. (**B**) In the absence of Ung1, high levels of transcription elevate C>T more that G>A mutations (blue and yellow bars, respectively) demonstrating a bias for NTS deamination. This bias reflects neither Fcy1 nor MMR activity. Error bars are 95% CIs.

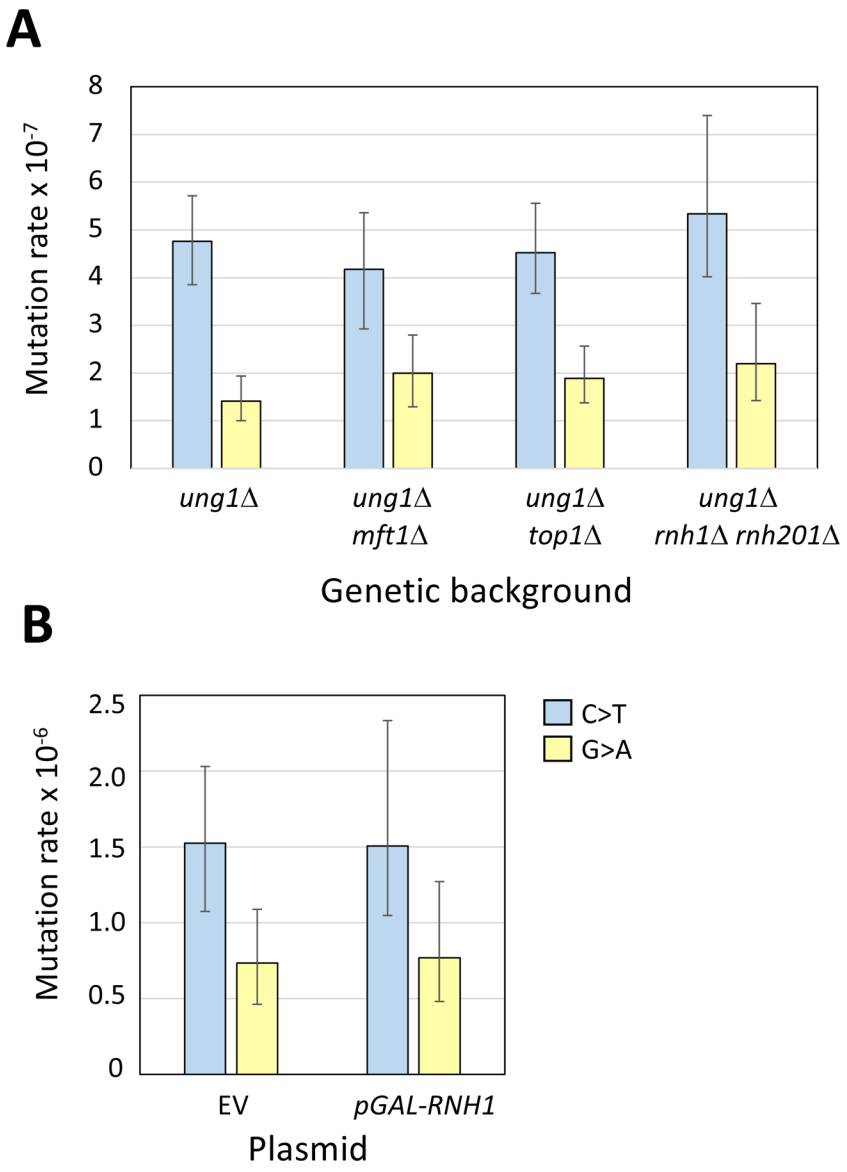

Figure 3.

Biased deamination of the NTS under high-txn conditions is not exacerbated in genetic backgrounds where R-loops are expected to (**A**) increase or (**B**) decrease. Blue and yellow bars reflect the rates of C>T and G>A mutations, respectively, in the indicated genetic backgrounds. Error bars are the 95% CIs.

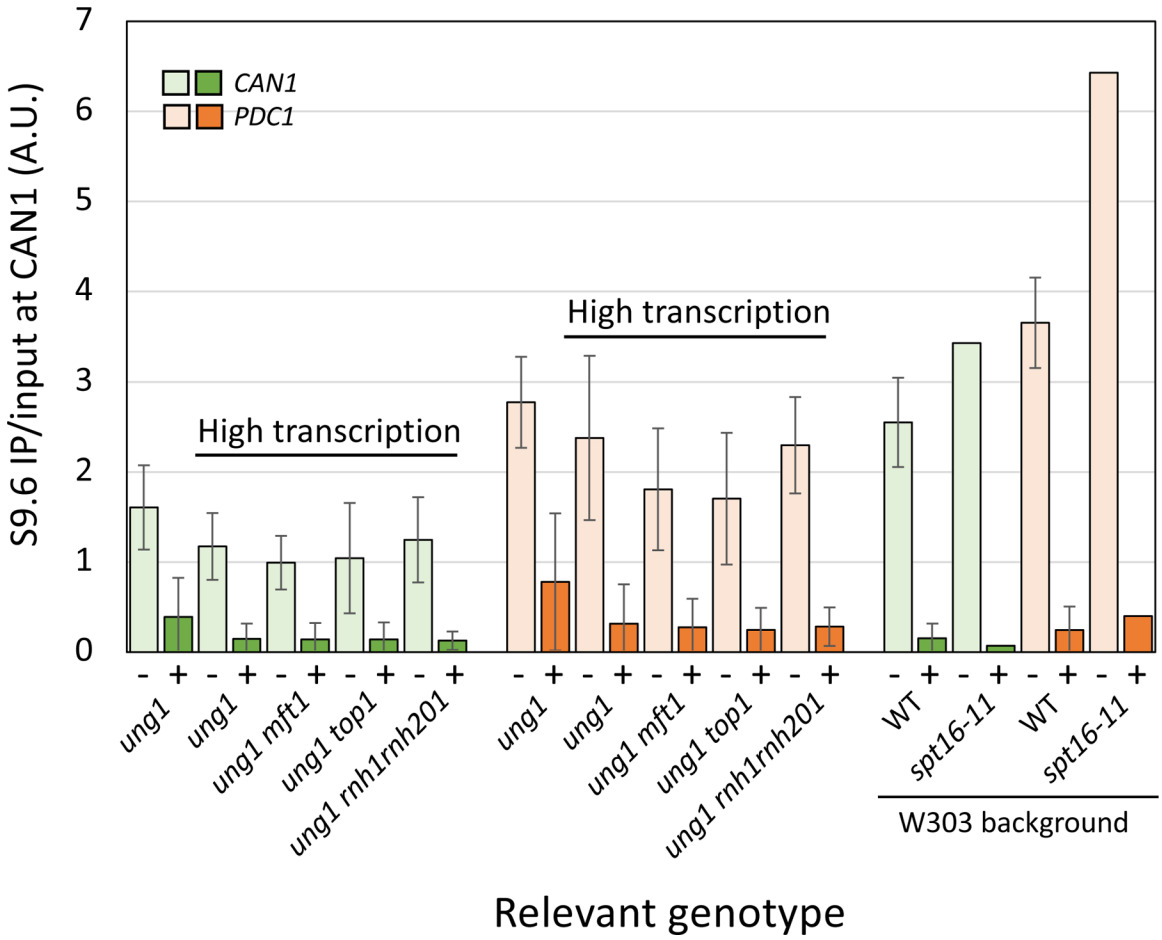

Figure 4.

Relative levels of RNA-DNA hybrids at *CAN1* and *PDC1*. Samples were or were not (+ or −, respectively) treated with RNase H1, which degrades RNA-DNA but not RNA-RNA duplexes, prior to immunoprecipitation with the S9.6 antibody. Values plotted are arbitrary units (A.U.) that represent the mean of three biological replicates after immunoprecipitation (IP) with S9.6 relative to the IP input. Error bars are standard deviations. The *spt16–11* derivative of W303, which accumulates R-loops, was used as positive control for RNA-DNA detection and was analyzed in parallel with experimental strains; duplicate samples were used for WT but only a single sample was analyzed from the *spt16–11* background.

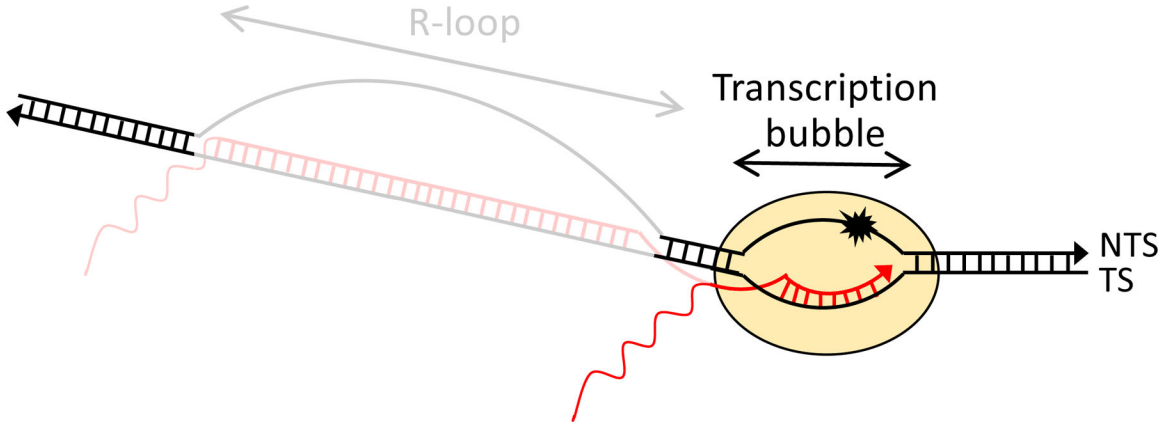

Figure 5.

Vulnerability of the NTS to DNA damage during transcription. A small portion of the NTS is rendered single-stranded as part of each transcription bubble; re-formed duplex DNA and the nascent transcript leave RNAP through separate channels. If the transcript threads back and pairs with the TS to form an R-loop, a larger portion of the NTS is exposed. In the case of *CAN1*, a contribution of R-loops to the biased deamination of the NTS was detected neither genetically or physically. Damage to the NTS that occurs within the transcription bubble is indicated by a star.