

Error-free and mutagenic processing of topoisomerase 1-provoked damage at genomic ribonucleotides

Justin L Sparks[†] & Peter M Burgers^{*}

Abstract

Genomic ribonucleotides incorporated during DNA replication are commonly repaired by RNase H2-dependent ribonucleotide excision repair (RER). When RNase H2 is compromised, such as in Aicardi-Goutières patients, genomic ribonucleotides either persist or are processed by DNA topoisomerase 1 (Top1) by either error-free or mutagenic repair. Here, we present a biochemical analysis of these pathways. Top1 cleavage at genomic ribonucleotides can produce ribonucleoside-2',3'-cyclic phosphate-terminated nicks. Remarkably, this nick is rapidly reverted by Top1, thereby providing another opportunity for repair by RER. However, the 2',3'-cyclic phosphate-terminated nick is also processed by Top1 incision, generally 2 nucleotides upstream of the nick, which produces a covalent Top1–DNA complex with a 2-nucleotide gap. We show that these covalent complexes can be processed by proteolysis, followed by removal of the phospho-peptide by Tdp1 and the 3'-phosphate by Tpp1 to mediate error-free repair. However, when the 2-nucleotide gap is associated with a dinucleotide repeat sequence, sequence slippage re-alignment followed by Top1-mediated religation can occur which results in 2-nucleotide deletion. The efficiency of deletion formation shows strong sequence-context dependence.

Keywords DNA repair; genomic ribonucleotides; mutagenesis; topoisomerase 1

Subject Categories DNA Replication, Repair & Recombination

DOI 10.15252/embj.201490868 | Received 19 December 2014 | Revised 18 February 2015 | Accepted 25 February 2015

Introduction

The incorporation of ribonucleotides during DNA replication and its consequence for genome instability have been the focus of several recent studies [reviewed in Williams & Kunkel (2014)]. Ribonucleotide incorporation is more frequent than previously surmised because the cellular rNTP pools are 10- to 100-fold higher than the dNTP pools (Nick McElhinny *et al.*, 2010b). At dNTP and rNTP

concentrations similar to those found in the yeast cell, replicative DNA polymerases typically incorporate 1–2 ribonucleotides per kilobase (Nick McElhinny *et al.*, 2010b; Sparks *et al.*, 2012). These incorporated ribonucleotides (rNMPs) constitute the most abundant replication errors in cells. Their removal from the genome is important to maintain genomic stability. RNase H2-initiated ribonucleotide excision repair (RER) is the main pathway that removes ribonucleotides from the genome (Rydberg & Game, 2002; Kim *et al.*, 2011; Sparks *et al.*, 2012). In humans, mutations in each of the genes for this three-subunit enzyme are responsible for a genetic syndrome known as Aicardi-Goutières syndrome (AGS) (Crow *et al.*, 2006). The mutations are associated with a reduction in RNase H2 activity, compromising the efficiency of RER (Chon *et al.*, 2013). A complete defect of RNase H2 in mouse causes embryonic lethality (Reijns *et al.*, 2012). And similar to yeast RNase H2 deletion strains (Nick McElhinny *et al.*, 2010a), mouse RNase H2-null cell lines also accumulate ribonucleotides in their genomic DNA (Reijns *et al.*, 2012).

In yeast strains lacking a functional RNase H2, the accumulation of genomic ribonucleotides is partially suppressed by the action of DNA topoisomerase 1 (Top1) (Kim *et al.*, 2011; Williams *et al.*, 2013). However, these RNase H2-deficient yeast strains show a distinct mutational pattern that is characterized by an increase in 2- to 5-nt deletions that occur specifically at repeat sequences, and are dependent on Top1 (Nick McElhinny *et al.*, 2010a; Kim *et al.*, 2011). Top1 is a type IB topoisomerase whose action proceeds through the reversible formation of a DNA nick bounded by a DNA-3'-phosphate-Top1 covalent complex through its active site tyrosine, and a 5'-hydroxyl group. Failure to reverse this step, through the action of drugs such as camptothecin that inhibit this re-ligation step, or because of loss of the adjacent 5'-hydroxyl group, can result in the persistence of the covalent complex (Champoux, 2001). This covalent complex is termed the Top1-cleavage complex (Top1-cc).

Based on genetic studies, a model has been proposed in which a persistent Top1-cc can be resolved using several different pathways. Generally, repair of a Top1-cc is initiated by partial proteolytic degradation of the DNA-bound Top1 (Desai *et al.*, 1997), followed by removal of the remaining peptide by Tdp1 (tyrosyl-DNA

Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO, USA

^{*}Corresponding author. Tel: +1 314 362 3872; E-mail: burgers@biochem.wustl.edu

[†]Present address: Department of Biological Chemistry, Harvard Medical School, Boston, MA, USA

phosphodiesterase) (Pouliot *et al*, 1999). Tdp1 activity is specific for small tyrosyl peptides and does not hydrolyze native Top1 attached to the 3'-phosphate (Debethune *et al*, 2002). Subsequently, the resulting 3'-phosphate can be redundantly removed either by Tpp1 (three prime phosphatase) or by the abasic endonucleases Apn1 or Apn2, which should produce a small gap (Vance & Wilson, 2001). Subsequent gap repair synthesis should proceed in an error-free manner. Alternatively, the structure-specific endonuclease Rad1–Rad10 can excise the Top1-cc as an initiating step in recombinational repair (Vance & Wilson, 2002). A third pathway for resolving Top1-cc has recently been proposed, and it depends on proteolysis of Top1 by Wss1 protease, followed by DNA polymerase ζ -dependent translesion synthesis past the lesion (Stinge *et al*, 2014). To what extent each of these pathways is utilized is currently not obvious, but the importance of the Tdp1 pathway is apparent from the observation that while neither deletion of *TDP1*, *RAD1*, or *WSS1* is particularly sensitive to camptothecin, either the *tdp1 Δ rad1 Δ* double mutant or the *tdp1 Δ wss1 Δ* double mutant is exquisitely sensitive to camptothecin, indicating that the repair of Top1-cc has been severely compromised when both pathways are inactivated (Vance & Wilson, 2002; Stinge *et al*, 2014).

Human and vaccinia viral Top1 proteins show endoribonuclease activity toward DNA-embedded ribonucleotides (Sekiguchi & Shuman, 1997; Kim *et al*, 2011). When Top1 is transiently linked to the 3'-phosphate of the ribonucleotide moiety, nucleophilic attack by the neighboring 2'-hydroxyl group can release Top1 with the formation of a 2',3'-cyclic phosphate nick (Fig 1A). Currently, it is not known whether the cyclic phosphate intermediate can be resolved by DNA repair enzymes. However, DNA with long-lived nicks are substrates for Top1, which cleaves most prominently a few nucleotides upstream of the nick (Christiansen & Westergaard, 1999). When this occurs, the resulting small oligonucleotide that is generated can readily dissociate, leaving a gap bounded by the Top1-cc and a 5'-hydroxyl group (Fig 1A). The gapped Top1-cc can be repaired by the Tdp1 or Rad1-10 pathway. Alternatively, if the small gap occurs in a repetitive DNA sequence, a slippage re-alignment by extrusion of the non-cleaved DNA strand would place the 5'-hydroxyl in proximity of the 3'-phosphate-Top1 and therefore may permit religation of the DNA with release of Top1 (Cho *et al*, 2013). If the bulge is not processed by mismatch repair, the following round of replication would lead to a 2- to 5-nt deletion in one of the daughter cells. The Top1-dependent rNMP-initiated deletion mutagenesis phenomenon shows strong similarity with transcription-associated mutagenesis (TAM) in yeast, which also occurs specifically at repetitive sequences in regions of highly transcribed sequences, and is also dependent on Top1 (Lippert *et al*, 2010; Takahashi *et al*, 2010). TAM is thought to occur by collision between RNA polymerase with a Top1-cc that must be resolved.

The resolution of Top1-cc has been an intense focus of study for some time, as several classes of therapeutic cancer drugs target Top1 and stabilize the cleavage complex leading to DNA damage (Pommier *et al*, 2010). However, the mechanistic details of how the activity of Top1 at genomic ribonucleotides can lead to deletion formation remain to be determined. The model proposed by Cho *et al* (2013) requires that Top1 acts independently at two steps, first at the site of ribonucleotide incision and, when this incision results in the formation of a 2',3'-cyclic phosphate nick, in a second step

upstream of the cyclic phosphate nick. Therefore, genetic studies with Top1 deletion mutants are uninformative since these eliminate both steps. However, a biochemical analysis should allow a probing of each step, thereby providing an opportunity to test the proposed model.

In this paper, we have investigated the activities of Top1 at genomic ribonucleotides and its subsequent activity at the intermediates that have been formed as a result of Top1 action at rNMPs. Our studies highlight the catalytic activity of Top1 at various types of RNA–DNA and DNA–DNA nicks. Surprisingly, the 2',3'-cyclic phosphate-terminated nick is the best substrate for Top1, leading both to a reversal process through nucleophilic attack of Top1 on the cyclic phosphate and to a forward reaction that produces Top1-cc. We have reconstituted the error-free pathway for the repair of Top1-cc using purified enzymes. We also describe the sequence requirements for the formation of 2-nt deletions in a dinucleotide repeat sequence containing a single ribonucleotide, which is the most common type of Top1-dependent deletion mutations produced in the cell (Lippert *et al*, 2010; Nick McElhinny *et al*, 2010a).

Results

Saccharomyces cerevisiae Top1 possesses endoribonuclease activity that produces a 2',3'-cyclic phosphate

We began our study using an oligonucleotide with a (GA)₃ dinucleotide repeat sequence, also named the (TC)₃ hotspot, derived from the yeast *CAN1* gene, which was identified as a Top1-dependent 2-nt deletion hotspot (Fig 1B) (Lippert *et al*, 2010; Nick McElhinny *et al*, 2010a). In order to separate error-free repair from mutagenic repair involving deletions, we screened several oligonucleotides with the rNMP situated at different positions and selected one substrate that yielded no detectable deletions for our studies of error-free repair. The same 32-mer oligonucleotide, with an rUMP at position 16 on the top strand (Fig 1B), was shown by others to produce a uridyl-2',3'-cyclic phosphate intermediate when treated with human Top1 (Kim *et al*, 2011).

The addition of *S. cerevisiae* Top1 to the substrate, 5'-labeled on the rUMP-containing strand, yielded two products: the first is consistent with Top1 cleaving on the 3'-side of the rUMP, and the second is a high molecular weight product that barely migrated into the gel (Fig 1B). These products are dependent on the presence of the rNMP in the substrate (Supplementary Fig S1A, compare lane 7 with 13). Additional control experiments with 3'-labeled substrate, which are detailed in Supplementary Fig S1A, show that the downstream product produced by Top1 action has a 5'-hydroxyl group, as expected from the known mechanism of Top1 (Champoux, 2001).

Vaccinia virus topoisomerase 1 cleaves DNA at an embedded rNMP with the formation of a 2',3'-cyclic phosphate (Sekiguchi & Shuman, 1997). A logical model for this reaction is that during the course of Top1 action, the transient covalent RNA-3'-phosphoryl-Top1 intermediate is attacked by the vicinal 2'-hydroxyl to release Top1 and produce a 2',3'-cyclic phosphate product (Fig 1A). We show that the product produced by yeast Top1 also has a 2',3'-cyclic phosphate terminus. First, it is resistant to alkaline

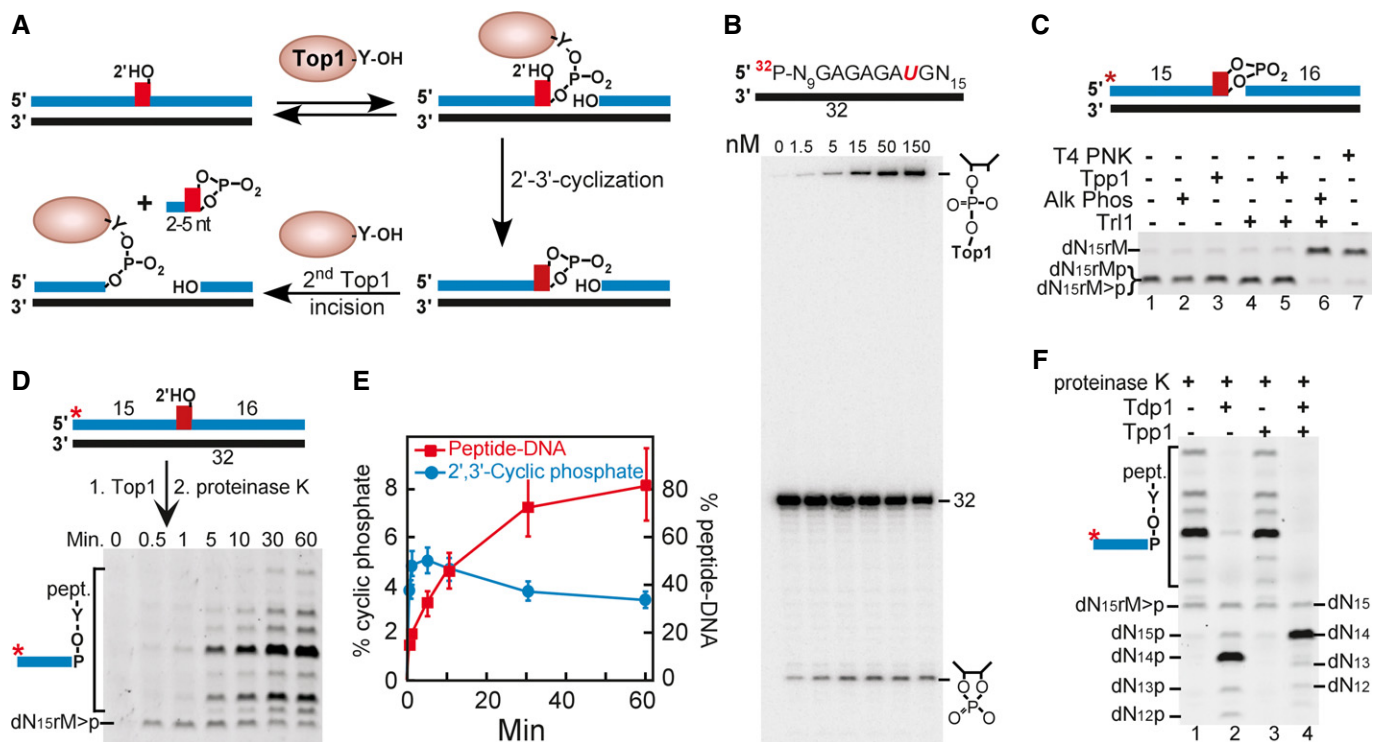


Figure 1. Endoribonuclease activity of Top1 and subsequent cleavage complex formation.

A Model for the formation of a 2',3'-cyclic phosphate-terminated nick and subsequent formation of a Top1-cc.

B Activity of Top1 on dsDNA with a single rUMP at position 16. Standard assay mixtures used 10 nM 5'-³²P-ACTCGTCACGAGAGAUGCCACGGTATTTCAAA hybridized to complementary DNA. Assays with increasing Top1 were carried out at 30°C for 10 min and analyzed on a 7 M urea, 17% polyacrylamide gel.

C The DNA containing the 2',3'-cyclic phosphate-terminated nick produced as in (B), was isolated, and incubated in a second reaction at 25 nM substrate with alkaline phosphatase, Tpp1 (25 nM), Trl1 (25 nM), or in combination as indicated, for 10 min at 30°C, or with T4 polynucleotide kinase (PNK) for 60 min at 37°C.

D Time course of formation of 2',3'-cyclic phosphate and Top1-cc. Standard assays used 25 nM substrate as in (B) and 100 nM Top1. After the indicated times, the reactions were stopped with 10 mM EDTA and 0.5% SDS and treated with proteinase K to degrade Top1 in the Top1-cc, leaving residual DNA-tyrosyl peptides, and analyzed by 7 M urea, 17% polyacrylamide gel.

E Quantification of gel in (D); left y-axis represents percentage cyclic phosphate and right y-axis represents percentage DNA-tyrosyl peptides derived from Top1-cc. Background 2',3'-cyclic phosphate present at t = 0 was subtracted in all graphs. The error bars are the errors in quantification of a representative experiment (see Material and Methods).

F Isolated DNA-tyrosyl peptide products (25 nM) as in (D) were incubated with 25 nM Tdp1 or Tpp1 or both for 10 min at 30°C and analyzed on a 7 M urea, 17% polyacrylamide gel.

Data information: In (C, D, F), * designates 5'-³²P.

phosphatase treatment, which can remove a 2'- or 3'-phosphate, but not a 2',3'-cyclic phosphate (Fig 1C, lane 2); it is also resistant to yeast Tpp1 (three-prime phosphatase), an enzyme involved in the repair of DNA strand breaks (Vance & Wilson, 2001) (lane 3). Third, the cyclic phosphate is resolved by the combined action of yeast Trl1 and alkaline phosphatase (lane 6). Trl1 is a tRNA ligase involved in tRNA splicing (Sawaya *et al.*, 2003). Trl1 is the only yeast enzyme known to process a ribonucleotide-2',3'-cyclic phosphate, but it produces a 2'-phosphate. The activity of Trl1 on the cyclic phosphate-terminated DNA substrate could not be directly demonstrated (lane 4), because both substrate and product migrate similarly on the gel. However, the hydrolysis of the cyclic phosphate by Trl1 was revealed when alkaline phosphatase was also included (lane 6). Inclusion of Tpp1 with Trl1 was ineffective (lane 5), because Tpp1 is specific for 3'-phosphates and does not hydrolyze the 2'-phosphate produced by Trl1. Finally, T4 polynucleotide kinase, which has both 2',3'-cyclic phosphatase and 3'-phosphatase activity (Das & Shuman, 2012), also removed the cyclic phosphate moiety (lane 7).

Top1 processes the rNMP-2',3'-cyclic phosphate nick into a Top1-cc

A very slowly migrating product was also observed, its abundance increasing with increasing Top1 concentrations (Fig 1B). When analyzed on a SDS-polyacrylamide gel, this product migrated as a single species (Supplementary Fig S1B). We hypothesized that this product was the Top1-cc, in which Top1 was covalently linked to the DNA through a 3'-phospho-tyrosyl bond. Proteinase K treatment of the Top1-cc, produced by incubation of the rNMP-containing DNA substrate with a 2-fold molar excess of Top1 for increasing times, showed a series of products reflecting covalent complexes of tyrosyl-oligopeptides with the DNA, with either the peptide or DNA or both varying in size (Fig 1D). These products, but not the ratio between the various products, increased with time. Importantly, the 2',3'-cyclic phosphate product was most abundant at early times, 5% after 1 min, but then slowly decreased with time, while the formation of DNA-peptide products, derived from Top1-cc, proceeded more slowly and accumulated over time, up to 81% after 60 min

(Fig 1D and E). These data are consistent with the model that the formation of the uridyl-2',3'-cyclic phosphate-terminated nick precedes the formation of the Top1-cc (Fig 1A).

Repair of Top1-cc by proteolysis, Tdp1 and Tpp1, and repair synthesis

One pathway for the removal of Top1-cc is through the Tdp1-dependent pathway, which hydrolyzes DNA-3'-tyrosyl peptides to DNA-3'-phosphate (Pouliot *et al*, 1999). We tested the ability of purified *S. cerevisiae* Tdp1 to remove the covalently linked Top1 from our DNA substrate. As previously reported for the human enzyme (Debethune *et al*, 2002), the Top1-cc containing yeast Top1 is also resistant to Tdp1 treatment. However, the tyrosyl-peptide-DNA products isolated after proteinase K digestion of Top1-cc (Fig 1D and F lane 1) were readily processed by Tdp1 to give predominantly a 14-mer-3'-phosphate (lane 2), which was converted into the 3'-hydroxyl 14-mer upon additional treatment with Tpp1 (lane 4). In addition, small amounts of 15-mer, 13-mer, and 12-mer products were also formed. These data indicate that the second incision by Top1 had occurred predominantly at a position two nucleotides upstream of the nick, releasing a dinucleotide-2',3'-cyclic phosphate and a 14-mer Top1-cc. Incision at other positions than the -2 position from the nick accounted for 11% of the different size products formed.

To determine whether proteolytic degradation of Top1-cc and hydrolysis by Tdp1 and Tpp1 are both necessary and sufficient steps to produce a substrate suitable for gap repair, the Top1-cc was subjected to proteinase K digestion and the 5'-labeled DNA-tyrosyl peptides were isolated on a denaturing gel and hybridized back to a 64-mer template (Fig 2). This family of DNA substrates failed to be extended by DNA polymerase δ (Pol δ), with its processivity factor PCNA present (lane 2). Neither did treatment of the DNA with Tdp1 or Tpp1 alone provide a suitable primer terminus for extension by Pol δ (lanes 4 and 6). Pol δ was only capable of extending the primer terminus after treatment with both Tdp1 and Tpp1 (lane 8). The extended product is resistant to treatment with sodium hydroxide (lane 9), consistent with the model in Fig 1A showing removal of the ribonucleotide by two subsequent Top1-mediated events. Our data demonstrate that a Top1-cc produced by Top1 action on misincorporated ribonucleotides during DNA replication, or by other pathways such as treatment of cells with camptothecin, can be efficiently repaired by partial proteolysis, followed by Tdp1- and Tpp1-dependent restoration of a proper 3'-hydroxyl terminus for repair synthesis by Pol δ .

Deletion formation at a dinucleotide repeat requires base pairing at the ligation junction

Having reconstituted the error-free pathway of the repair of rNMP-provoked Top1 activity, we next investigated the determinants that would induce deletion formation. Genetic studies in yeast have indicated that 2-nt deletions in the context of a dinucleotide repeat sequence are the most frequent events of Top1-dependent deletions produced in RER-defective RNase H2 mutants (Kim *et al*, 2011). Deletions are proposed to occur by two sequential Top1 cleavages of the rNMP-containing strand to eventually form the Top1-cc, which is followed by an extrusion of a dinucleotide in the repeat sequence on the complementary strand that re-aligns the 5'-hydroxyl of the

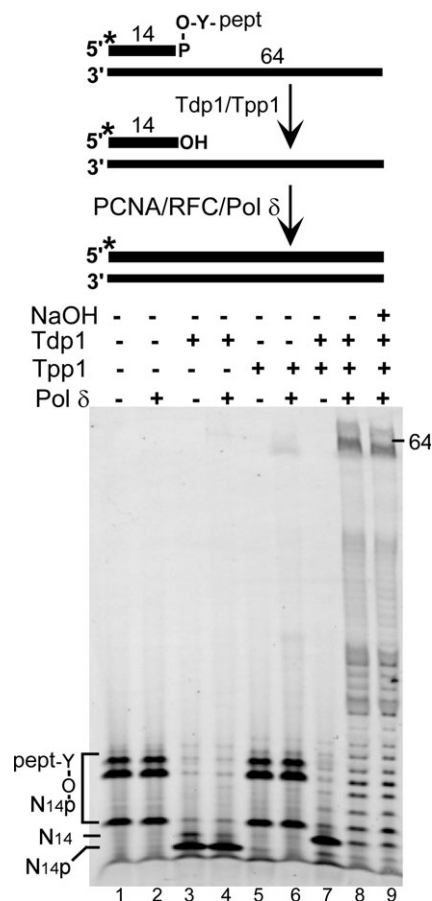


Figure 2. Reconstitution of the Tdp1-dependent repair pathway of Top1-cc.

The 5'-Cy3-labeled 32-mer substrate shown in Fig 1B was treated with 100 nM Top1 under standard assay conditions for 1 h at 30°C. The product was then treated with proteinase K followed by isolation of the main tyrosyl peptide-oligonucleotide products on a 7 M urea, 17% polyacrylamide gel. The isolated DNA was hybridized to a 64-nt template containing both 5' and 3' biotin-streptavidin blocks. The biotin-streptavidin blocks retain PCNA, loaded by its clamp loader RFC onto the DNA substrate allowing for more efficient replication by Pol δ (Ayyagari *et al*, 2003). This DNA (25 nM) was incubated with PCNA (50 nM), RFC (100 nM), RPA (50 nM), 100 μ M ATP as well as Tdp1 (25 nM), Tpp1 (25 nM), or in combination as indicated for 10 min at 30°C either in the presence or in the absence of 50 nM Pol δ . The sample in lane 9 was treated with 0.3 M NaOH at 55°C for 30 min to cleave at potential internal ribose residues. Analysis was on a 7 M urea, 17% polyacrylamide gel. * designates 5'-Cy3.

downstream strand with the Top1-3'-phosphate of the upstream strand, thereby allowing for religation to occur (Fig 3A). The substrate we used for our initial studies in Fig 1, with uridine at position 16, did not yield deletions, presumably because a dinucleotide extrusion in the TC repeat on the complementary strand did not give proper base pairing at the junction [Fig 3B, substrate (a)]. However, when the ribonucleotide was moved to position 14, a properly base-paired junction was generated by TC extrusion on the bottom strand, and a 2-nt deletion product could be observed albeit inefficiently [substrate (b)]. Reasoning that increased base pairing near the junction might stabilize the extrusion and thereby stimulate deletion formation, we increased the repeat from (GA)₃ to (GA)₄ and observed a marked increase in deletion formation [substrate (c)]. Changing the rG to rC in the terminal GA repeat eliminated the possibility for proper junction

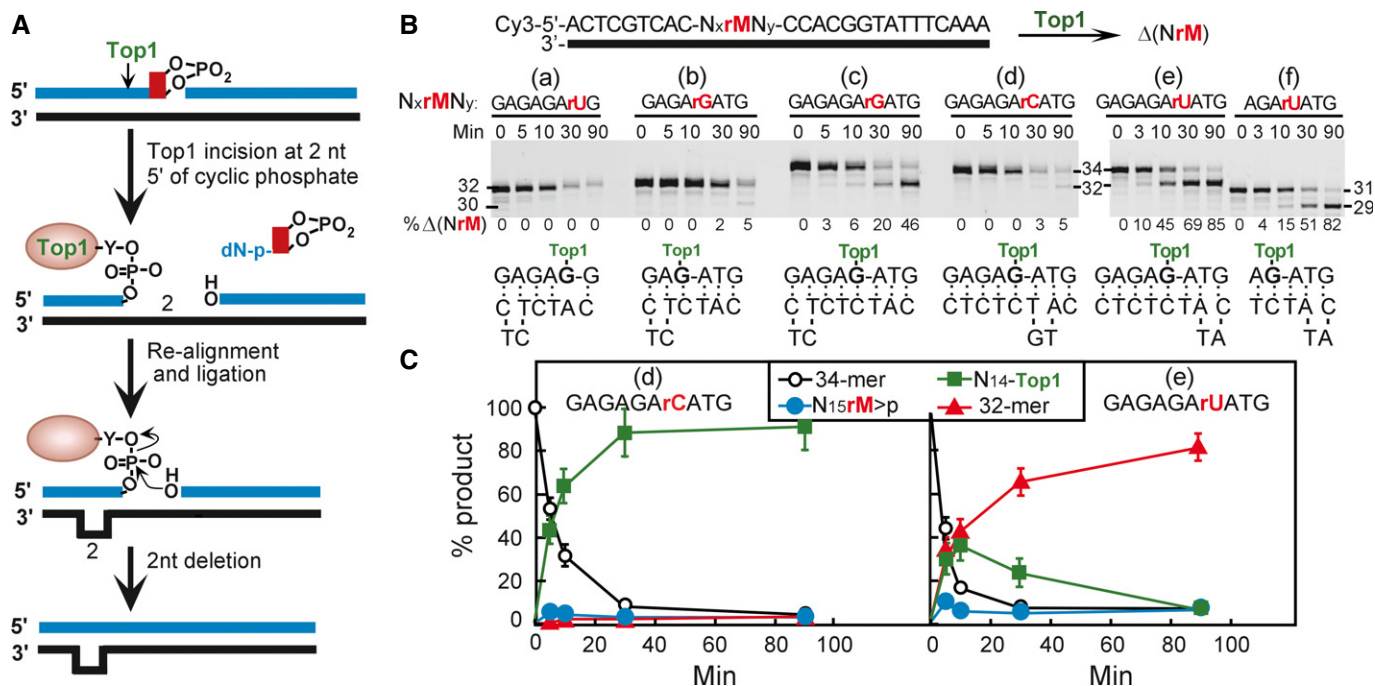


Figure 3. Top1 activity generates deletions when rNMP is in the context of repetitive DNA sequences.

A Model for Top1-dependent dinucleotide deletions.

B Top1-dependent deletions of dinucleotides. Standard assay mixtures contained 25 nM double-stranded oligonucleotide with the rNMP-containing strand 5'-Cy3-labeled. The NxrMNY sequence is shown above each gel with the rNMP in red. The outside sequences are identical for each experiment. Assays were carried out with 300 nM Top1 at 30°C for the time indicated and analyzed on a 7 M urea, 17% polyacrylamide gel.

C Graphic representation of the quantification for oligonucleotide (d) left and (e) right, of the substrate (34-mer) and the three products, cyclic phosphate, Top1-cc, and 32-mer deletion product, as indicated. See Materials and Methods for error analysis.

base pairing after extrusion of a TC dinucleotide on the bottom strand [substrate (d)]. However, some deletion product was observed, which we suggest could originate from dinucleotide extrusion to the right of the junction, which also leaves a base-paired junction. The stability of this ligation junction is probably poor because the extrusion is only one nucleotide from the junction. Therefore, we changed the rG to rU, which increases the base pairing to the right of the junction, and this substrate was highly active for deletion formation [substrate (e)]. The proposed slippage re-alignment for substrate (e) indicates that deletion formation should be independent of the (GA)₃ repeat sequence. To test this, we changed the -GAGAGArUAT- sequence to -AGArUAT-, which eliminates the GA repeats but still retains a strong Top1 recognition site (see below). This substrate (f) is very proficient for deletion formation, indicating that both GA and TA repeats can induce 2-nt deletion formation as determined in the original mapping studies (Lipfert *et al*, 2010; Nick McElhinny *et al*, 2010a). To our knowledge, this is the first biochemical demonstration of Top1-dependent deletion formation at dinucleotide repeats.

We carried out a kinetic analysis of the rates of accumulation of Top1-dependent products, comparing substrate (d) with (e). Within a factor of 2, the consumption of the 34-mer substrate proceeded at comparable rates (Fig 3C). Likewise, the rate of appearance of the cyclic phosphate product and the initial rate of the formation of Top1-cc also proceeded with comparable rates. However, while with substrate (d) the Top1-cc kept on accumulating with time, the Top1-cc of substrate (e) was converted into the 32-mer deletion product. An estimate of the rates shows that deletion formation proceeded

more than 100-fold slower for (d) than for (e). Several control experiments described in detail in Supplementary Fig S2 were carried out to ascertain that the deletion product has lost the dinucleotide, as predicted by the model. First, while the substrate is susceptible to cleavage by NaOH due to the presence of the rNMP, the Top1-dependent religation product is resistant to NaOH (Supplementary Fig S2A). Secondly, we sequenced the substrate and the product of the reaction and showed that the dArG dinucleotide had been deleted as predicted (Supplementary Fig S2B).

The nature of the nick structure determines processing by Top1

The kinetic analysis in Fig 1E shows that the cyclic phosphate intermediate is formed very rapidly, whereas Top1-cc formation was relatively slow. Yet, the cyclic phosphate intermediate accumulated to only a few percent. This suggested to us that the formation of the cyclic phosphate might be a reversible reaction. This proposed interconversion of substrate and product is shown schematically in Fig 4D. In order to test this hypothesis, we generated the series of substrates shown in Fig 4A (boxed substrate), with the same sequence and ribonucleotide position as substrate (e) in Fig 3B, but with all possible nick configurations with regard to phosphate status. The oligonucleotide with the uridine-2'-3'-cyclic phosphate terminus (U>p) was generated by partial RNase A digestion of a longer oligonucleotide and was contaminated with ~25–30% of the uridine-3'-phosphate form (Up, Supplementary Fig S3A). Remarkably, the substrate containing the U>p nick was very rapidly converted

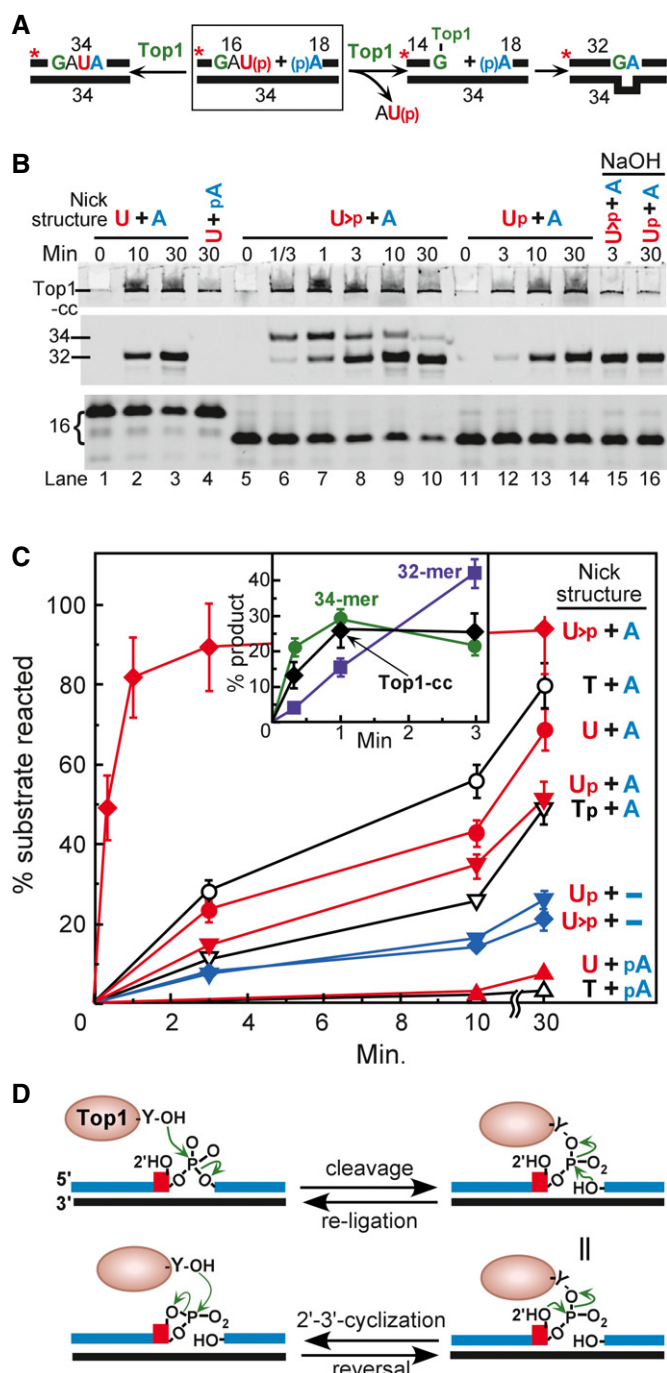


Figure 4. Structure of the RNA-DNA nick determines processing by Top1.

- A** Schematic for the Top1-dependent activity at various rNMP-terminated DNA nick structures. The possible activities of the whole family of nicks are shown. A 5'-Cy3-labeled 16-mer with various 3'-terminated structures and an 18 mer with either 5'-phosphate or 5'-hydroxyl were hybridized to a template 34-mer leaving a nick (the starting substrate is boxed). Top1 can either catalyze religation, but only when the nick contains a 2',3'-cyclic phosphate, or catalyze a second cleavage and potentially catalyze subsequent religation after DNA slippage re-alignment, giving a 2-nt deletion product. The upstream top strand is 5'-Cy3-labeled (indicated with *). Construction of the substrates is described in the text and in Supplementary Fig S3.
- B** The different RNA-DNA nick substrates were treated with 100 nM Top1 under standard assay conditions for the indicated times at 30°C and analyzed on a 7 M urea, 17% polyacrylamide gel. The labeled 16-mer upstream oligos were terminated with a uridine with either no phosphate (U, lanes 1–4), a 2',3'-phosphate (U > p, lanes 5–10, 15), or a 3'-phosphate (Up, lanes 11–14, 16), and the 18-mer was 5'-terminated with either a 5'-hydroxyl (A) or a 5'-phosphate (pA). Only the Top1-cc, the 32 and 34 region, and the 16-mer regions of the gel are shown. The complete gel, containing additional DNA-DNA nick structures and controls, is shown in Supplementary Fig S3C.
- C** Quantification of the products from (B) and Supplementary Fig S3C; RNA-DNA nicks are red-blue and DNA-DNA nicks are black-blue. U > p + – and U > p + – represent the absence of the downstream 18-mer (Supplementary Fig S3C). The inset shows quantification of the initial rates of formation of various products by reaction of U > p + A with Top1, as indicated. See Materials and Methods for error analysis.
- D** Model for Top1 activity on cyclic phosphate intermediates.

by Top1 into a ligated 34-mer product (Fig 4B, lanes 5–10). Conversely, formation of Top1-cc by Top1 cutting 2 nt upstream of the U > p nick occurred at a lower rate, and so did subsequent formation of the 32-mer 2-nt deletion product (Fig 4C, inset). The 34-mer but not the 32-mer contained the ribonucleotide, as it was sensitive to alkali treatment (lane 15). The substrate remaining after 10 min is predominantly the contaminating uridine-3'-phosphate nicked form (Supplementary Fig S3C, lanes 24 and 25). The uridine-3'-phosphate nick did not form a ligated 34-mer product, but rather, it slowly formed the 32-mer deletion product (lanes 11–14), which was resistant to alkali (lane 16). The phosphate-less nick was also converted to the 32-mer deletion product (lanes 1–3). However, the

normal 5'-phosphate nick was converted very slowly into a Top1-cc, and this Top1-cc failed to convert into the 32-mer deletion product, because it lacked the 5'-hydroxyl group that is essential for religation (lane 4). We also tested three all-DNA substrates and determined that each of their reactivities with Top1 was comparable to that of the analogous ribo-containing substrate (Supplementary Fig S3C). The rates of conversion of the seven nick substrates are plotted in Fig 4C, along with two additional control substrates that lacked the downstream oligonucleotide (Supplementary Fig S3C, U > p + – and Up + –). Their conversion to the Top1-cc product occurred much slower than that of the analogous nicked substrates. Two substrates stand out in this comparative study. First, a regular DNA nick (T + pA), which is formed as an intermediate during virtually all DNA transactions, is basically inactive for modification by Top1. Second, the U > p + A nick is converted an order of magnitude faster than other nicked substrates, indicating that the U > p + A nick, formed by Top1 activity on genomic ribonucleotides, is an unstable high-energy intermediate, that is rapidly processed.

Influence of ribonucleotide position in Top1-cc and deletion formation

In order to understand the rNMP positional determinants for the formation of Top1-dependent products, we tested the DNA substrate (c) from Fig 3B, but with the single rNMP placed at eight different positions on the top strand within and near the (GA)₄ repeat, as well as on the complementary positions on the bottom strand (Fig 5, Supplementary Fig S4). We designate the position prior to the (GA)₄ repeat as the –1 position, and the first nucleotide of the repeat as the 1 position, etc. The opposite positions on the (TC)₄ strand are designated as –1', 1', etc. Figure 5A shows the distribution of products after 90-min incubation of the oligonucleotide substrates with a

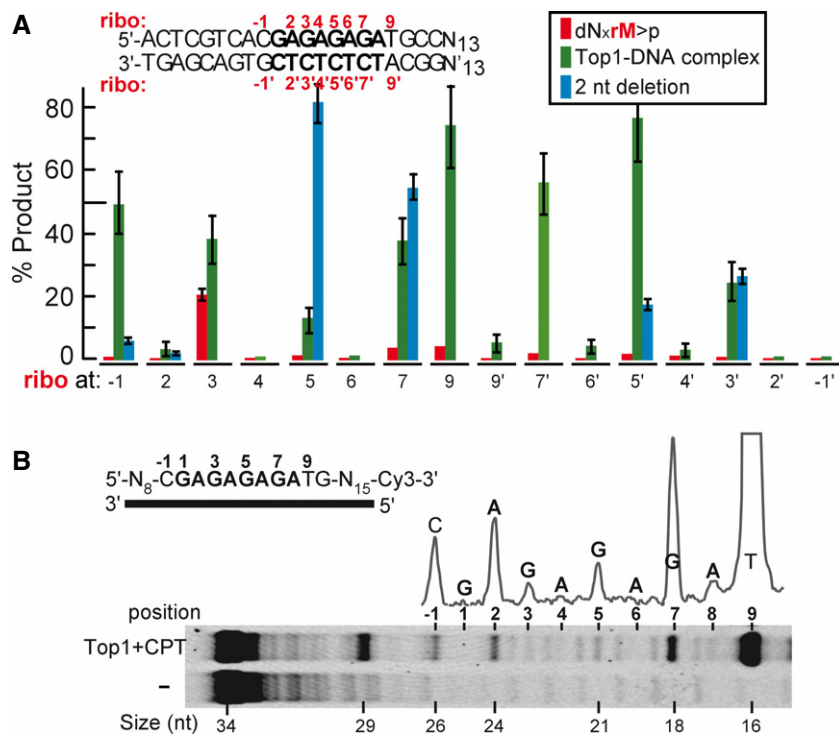


Figure 5. Position of the rNMP within a repeat sequence determines the efficiency of Top1-mediated deletion formation.

A A series of dsDNA substrates was made with the rNMP either on the top strand (positions -1 to 9 , in which position 1 is the first G in the $(GA)_4$ repeat) or on the bottom strand (positions $-1'$ to $9'$, as indicated). In each substrate, the rNMP-containing strand was 5'-labeled with a Cy3 fluorophore. The series of substrates were treated with 100 nM Top1 under standard assay conditions for 90 min at 30°C and analyzed on a 7 M urea, 17% polyacrylamide gel. Additional time courses (10 , 30 , 90 min) are given in Supplementary Fig S4A and B. The three products [cyclic phosphate (red), Top1-cc (green), and 32 -mer (blue)] were quantified and shown. See Materials and Methods for error analysis.

B The sequence preference for the initial Top1 cleavage on the $(GA)_4$ -containing strand was determined using the same oligonucleotide as in (A) but lacking a rNMP, and labeled with a $3'$ -Cy3-label. The positions of the $(GA)_4$ repeat are labeled. The DNA was either not treated or treated with 100 nM Top1 together with 5 μM camptothecin for 1 min at 30°C and analyzed on a 7 M urea, 17% polyacrylamide gel. The band intensity of the untreated DNA was subtracted from that of the treated DNA, and this is shown in the blue trace.

2 -fold molar excess of Top1. The analysis of the 16 substrates shows a large variation in the formation of Top1-cc and of 2 -nt deletions (Fig 5A, Supplementary Fig S4).

Topoisomerase 1 shows strong sequence preference for cleavage. The consensus sequence for mammalian Top1 is $5'-(A/T)(G/C)(A/T)T$ with the covalent linkage occurring at the $3'$ -T residue (Champoux, 2001). We determined the DNA cleavage site preference for yeast Top1 in the presence of camptothecin, which freezes Top1 in its cleavage complex, on the top strand of the main oligonucleotides used in this study (Fig 5B, Supplementary Fig S4C). Analogous to mammalian Top1, yeast Top1 also showed a strong cleavage preference for the T residue on position 9 of the sequence in Fig 5B, which is a perfect consensus site (AGAT). Therefore, it is not surprising that placing a rU at that position resulted in the rapid formation of the Top1-cc (Fig 5A). Of equal interest, however, are the positions on the DNA that did not show cleavage by Top1, for example, A-4 and A-6. Since the Top1 recognition sequence is about four nucleotides long, both the A-4 and A-6 position would present the same sequence to Top1, that is, GAGA. When an rA was placed in each of these positions, the rNMP-containing DNA was resistant to Top1 treatment. Therefore, if a ribonucleotide were inserted at that position during replication, and if RER were defective, as in an RNase H2-deletion strain, this ribonucleotide would be expected to remain in the genome.

Consistent with the known sequence preference of Top1, the GA-containing top strand showed higher activity than the CT-containing bottom strand. This is not that clearly seen in Fig 5, but the kinetic analysis in Supplementary Fig S4A and B provides strong support for this conclusion. After 10 min of reaction with Top1, the top-strand positions reacted (to cyclic phosphate, Top1-cc, and 2 -nt deletions) at an average of 24% (0 – 59%), whereas the bottom-strand positions reacted to an average of 7% (0 – 27%).

Eight out of the sixteen rNMP-containing positions examined were resistant to Top1; more than 90% of substrate remained after 90 min of reaction (Fig 5A). The other eight substrates showed significant product formation. Only one substrate (with the rNMP at position 3) accumulated as much as 21% of the cyclic phosphate intermediate over time. For the other substrates, a steady-state level of up to 5% cyclic phosphate was maintained, while other products (Top1-cc and 2 -nt deletions) accumulated over time (Supplementary Fig S4A and B). Only four of the substrates that accumulated substantial Top1-cc over time, yielded significant ($>5\%$) 2 -nt deletions. One possible reason why deletion formation would not occur could be because Top1 cutting upstream of the cyclic phosphate nick produced a different size gap than the 2 -nt gap we observed so far. Therefore, we investigated at which position upstream of the cyclic phosphate nick the second Top1 cleavage occurred, as

described in Fig 1F (Supplementary Fig S4D). For the seven substrates investigated, the predominant second cleavage site was 2 nt upstream of the cyclic phosphate nick, and second cleavages at 3, 4, and 5 nt upstream were very infrequent. No second cleavage at 1 nt upstream of the first cleavage was observed.

Discussion

The action of Top1 on genomic ribonucleotides is unique because of the potential for release of Top1 through the formation of a 2',3'-cyclic phosphate intermediate. With one single exception (when the ribonucleotide was at position 3), these cyclic phosphate intermediates did not accumulate to a substantial percentage (Fig 5A). Rather, Top1 has evolved the ability to recognize the 2',3'-cyclic phosphate and mediate fast religation to the starting ribonucleotide-containing DNA (Fig 4A and D). The initial kinetics of the reaction between the 2',3'-cyclic phosphate-containing nick and Top1 indicate that reversal is favored over additional cleavage by Top1 (Fig 4B and C inset). The reversal reaction is particularly important in situations where RER is compromised because of defects in RNase H2. A subset of Aicardi-Goutières syndrome patients show a reduction in RNase H2 activity, thereby compromising the efficiency of RER (Chon *et al*, 2013). The Top1-catalyzed conversion of the 2',3'-cyclic phosphate nick into rNMP-containing DNA allows for another opportunity for processing of genomic ribonucleotides by error-free RER.

In addition to reversal, the 2',3'-cyclic phosphate nick can also be processed by a second Top1 cleavage, which occurs predominantly 2 nt upstream of the nick (Supplementary Fig S4D). This reaction is irreversible because the small oligonucleotide generated by cleavage readily dissociates, as indicated in the model in Fig 6. DNA nicks have previously been shown to promote formation of covalent Top1 complexes on the 5'-strand (Henningfeld & Hecht, 1995; Lebedeva *et al*, 2008). Our kinetic analysis of different nick architectures suggests that natural DNA nicks are actually very poor substrates for Top1 cleavage, while the 2',3'-cyclic phosphate nick is 100–1,000 times more reactive (Fig 4B and C). Phosphate-less nicks or 3'-phosphate nicks show intermediate reactivity. Therefore, Top1 is relatively inactive to normal intermediates of DNA metabolism, but has evolved to specifically process the cyclic phosphate intermediates that it generates at genomic ribonucleotides. At a low rate, other unusual nick architectures are also processed by Top1.

The very small gaps that are formed as a result of tandem Top1 action at genomic ribonucleotides are most often the subject of error-free DNA repair. This repair may be initiated by ubiquitin-mediated proteolysis of the covalently bound Top1 (Desai *et al*, 1997); however, processing by a specialized protease Wss1 has also been reported (Stingele *et al*, 2014). Hydrolysis of the DNA-tyrosyl peptide by Tdp1 (Pouliot *et al*, 1999; Debethune *et al*, 2002), and of the 3'-phosphate by Tpp1 or other cellular 3'-phosphatases such as Apn1 or Apn2 (Vance & Wilson, 2001), produces a small phosphate-less gap that should be an apt substrate for gap repair. Here, we have reconstituted the error-free repair pathway with purified proteins, starting with the tyrosyl peptide intermediate that was generated by proteinase K rather than by ubiquitin-mediated proteolysis or Wss1 (Fig 2).

Two alternative modes of processing of the small 2- to 5-nt gaps with covalently attached Top1 can be envisioned. First, when

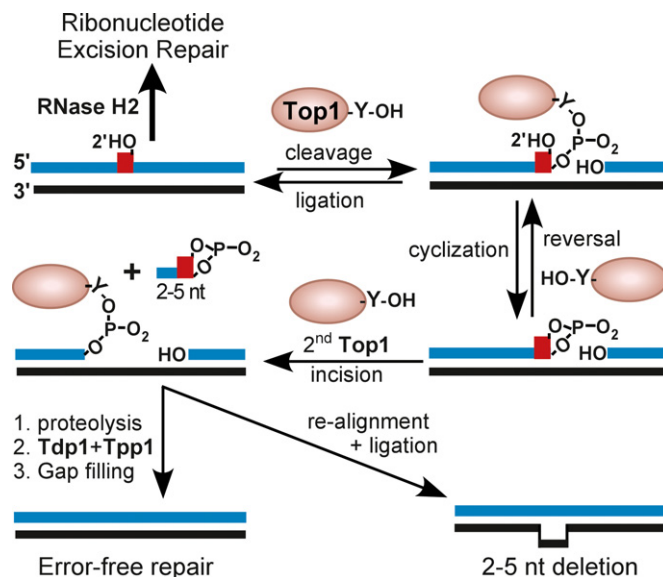


Figure 6. Model for Top1-dependent activities at genomic ribonucleotides.

See text for details.

sequence slippage re-alignment in a repeat sequence is favorable, the 5'-hydroxyl of the downstream DNA strand can be brought in proximity to the Top1-3'-phosphate of the upstream strand allowing DNA ligation with release of Top1 (Fig 6). In this paper, we focused on the most common gap of 2 nt that can produce a 2-nt deletion. This is an extremely rare event in the cell, but it is very sensitive to genetic detection because of the frame-shift mutations that can result because of it (Kim *et al*, 2011; Cho *et al*, 2013). Second, these gaps can be further enlarged by 5'-nucleases such as Exo1. Gap enlargement would protect the Top1-cc containing DNA from the slippage re-alignment and ligation pathway that yields deletions. In fact, a recent study showed that yeast Exo1 and the helicase Srs2 may collaborate in this gap-enlarging process, as mutations in either gene enhance the rate of Top1-dependent frame-shift mutagenesis (Potenski *et al*, 2014). In that study, the proposal was made that Exo1 and Srs2 may act at the 2',3'-cyclic phosphate-terminated nicks that are generated after the first Top1 cleavage. However, since these nicks are very transient and are rapidly processed by a second Top1 cleavage, we propose that the Top1-cc with the small gap is a more appropriate substrate for Exo1 and Srs2. In principle, these alternatives can be distinguished by using separation of function mutations in *TOP1*. We think that generating such separation of function mutations should be feasible because of a significant difference in DNA structures used by Top1 for each of the two steps. In the first step, Top1 recognizes upstream dsDNA and cuts 3' of its binding site, while in the second step, Top1 recognizes a specialized nick and cuts a few nucleotides upstream of that nick. If such mutants were available, one would expect that a *TOP1* mutant, which cuts the first but not the second time, would accumulate 2',3'-cyclic phosphate products and therefore would be predicted to have a greater reliance on Exo1/Srs2 if these enzymes act on the cyclic phosphate nick, as proposed (Potenski *et al*, 2014).

Repeat sequences are generally favorable sequences for slippage re-alignment, which is a prerequisite for Top1-mediated deletion

formation. Our analysis of the influence of different sequence environments on deletion formation has given us an insight into the rules that govern this. Seven out of sixteen rNMP-containing sequence environments examined produced a substantial fraction of Top1-cc product with predominantly a 2-nt gap (Fig 5A and Supplementary Fig S4D). For four of those, the Top1-cc reacted further by slippage re-alignment and ligation to give a 2-nt deletion products. These four DNAs are characterized by their ability to realign either the upstream or downstream top strand to form a nick in which perfect base pairing occurs, similar to shown in Fig 3B. In addition, one of the two base pairs at the nick junction is a strong G-C base pair. The substrate with the rNMP at position 9 did not produce a deletion because perfect base pairing at the nick cannot be accomplished. Interestingly, the substrate with the rNMP at position 7' also did not yield a deletion, even though perfect base pairing was possible. However, in that case, the nick was anchored by two A-T base pairs which may not have been able to generate a slippage re-alignment product with enough stability for re-ligation. Together with the data in Fig 3, we can conclude that deletion formation requires a re-aligned sequence with perfect base pairing at the nick, which is anchored with at least one G-C base pair.

Our studies provide evidence for the proposal that repetitive DNA sequences may allow for slippage re-alignment of the DNA ends to achieve perfect base pairing with the bulging out of several nucleotides upstream or downstream of the site of religation (Cho *et al*, 2013). We have also reconstituted the predominant repair pathway of Top1-initiated processing of genomic ribonucleotides, in which the Top1-cc is processed by the Tdp1-dependent pathway (Fig 2). Based on genetic data (Kim *et al*, 2011), we can make a rough estimate of the frequency of these three pathways in a wild-type cell. RER repairs the vast majority (~0.95–0.99) of ribonucleotides, while error-free repair via Top1-cc occurs at a frequency of 0.01–0.05 and repeat-induced deletion formation at a frequency of $\sim 10^{-8}$ – 10^{-9} . In *rnh201Δ* cells, processing by Top1 is dramatically increased, and so are the rates of deletions. These numbers are based on studies of the *CAN1* locus (Kim *et al*, 2011) and may not accurately reflect rates in the entire genome. The increased burden of Top1-mediated processing of ribonucleotides in an RER-defective strain also results in genomic stress, in particular if the frequency of rNMP incorporation is increased, as in a *pol2-M644G* mutant of Pol ϵ . The *rnh201Δ pol2-M644G* double mutant shows constitutive checkpoint activation and an increased sensitivity to replication stress that is dependent on Top1 (Williams *et al*, 2013).

While this manuscript was under review, several papers appeared describing genome-wide mapping data of genomic ribonucleotides in yeast RNase H2 mutants (Clausen *et al*, 2015; Koh *et al*, 2015; Reijns *et al*, 2015). These data should provide a basis for determining which genomic ribonucleotides are subject to attack by Top1 and which may result in the formation of 2- to 5-nt deletions.

Materials and Methods

Proteins and oligonucleotides

RPA (Henricksen *et al*, 1994), PCNA (Eissenberg *et al*, 1997), RFC (Gomes *et al*, 2000), and FEN1 (Gomes & Burgers, 2000) were

purified from *E. coli* overexpression systems, while Pol δ (Fortune *et al*, 2006) was purified from a yeast overexpression system. Tdp1, Tpp1, and Trl1 proteins were overexpressed from the Open Biosystems *S. cerevisiae* open reading frame library in the vector backbone BG1805 and purified as described (Gelperin *et al*, 2005).

Oligonucleotides were purchased from IDT (Coralville, IA) and purified by HPLC and/or PAGE chromatography. They are based on the 32-mer yeast *CAN1* sequence ACTCGTCACGAGAGATGC-CACGGTATTTCAAA in which 2-nt deletions have been detected (Kim *et al*, 2011). Sequence alterations and ribonucleotide modifications are as indicated. Oligonucleotides were either 5'- or 3'-labeled by treatment with [γ - 32 P]-ATP and T4 polynucleotide kinase or [α - 32 P]-dATP and terminal transferase, respectively, or purchased with either a 5'- or 3'-Cy3 fluorophore. Most rNMP-containing oligonucleotides are contaminated with the 2',3'-cyclic phosphate product, up to ~0.5%. This contamination increased when they were 5'- 32 P labeled with polynucleotide kinase. For this reason, we preferred using Cy3-labeled oligos, which have lower background levels. Quantification of cyclic phosphate products included subtraction of the background present at $t = 0$. Each labeled rNMP-containing oligonucleotide was hybridized to a 2-fold excess unlabeled complementary DNA oligonucleotide.

Top1 overexpression and purification

Top1 was overexpressed using plasmid pRS425-GAL-GST-*TOP1* containing the *Schistosoma japonicum* glutathione S-transferase (GST) gene fused to the N-terminus of the *TOP1* gene in vector pRS425-GALGST-term (Walker *et al*, 1993; Bylund *et al*, 2006). The GST tag is separated from the N-terminus of the Top1 by a recognition sequence for the human rhinoviral 3C protease (LEV-LFQ/GP). Following cleavage by the protease, the N-terminus of Top1 is extended with the GPEFDIKL sequence. Top1 overproduction was carried out in *S. cerevisiae* strain FM113 (MATa *ura3-52 trp1-289 leu2-3112 prb1-1122 prc1-407 pep4-3*), transformed with plasmid pRS425-GAL-GST-*TOP1*. Growth, galactose induction, extraction preparation, and ammonium sulfate precipitation (0.3 g/ml) were similar to the procedures described previously (Bylund *et al*, 2006). The ammonium sulfate precipitate was resuspended in buffer A₀ (buffer A: 60 mM HEPES-NaOH [pH 7.4], 10% glycerol, 1 mM dithiothreitol, 1 mM EDTA, 0.01% polyoxyethylene (10) lauryl ether, 1 mM sodium bisulfite, 1 μ M pepstatin A, 1 μ M leupeptin; subscript indicates the mM sodium chloride concentration), until the lysate conductivity was equal to that of buffer A₄₀₀. The lysate was then used for batch binding to glutathione-Sepharose 4B beads (GE Healthcare), equilibrated with buffer A₄₀₀, and gently rotated at 4°C for 2 h. The beads were collected at 1,000 rpm in a swinging-bucket rotor, followed by batch washes (3 \times 20 ml of buffer A₄₀₀). The beads were transferred to a 10-ml column and washed at 2.5 ml/min with 100 ml of buffer A₄₀₀. The second washing was with 50 ml buffer A₄₀₀ containing 5 mM Mg-acetate and 1 mM ATP. And the third washing used 50 ml of buffer A₄₀₀ and 30 ml of buffer A₂₀₀. Elution was carried out at a flow rate of 0.2 ml/min with buffer A₁₅₀ containing 30 mM glutathione (pH adjusted to 8.1). Fractions containing Top1 were combined and incubated overnight at 4°C with 30 U of rhinoviral 3C protease. The following day, the Top1 protein was loaded on a heparin column in buffer A₁₅₀ without protease inhibitors. The

column was washed with 10 column volumes of buffer A₃₀₀, and the protein was eluted with buffer A₇₅₀. Fractions containing pure Top1 were collected and dialyzed overnight to A₂₀₀ without protease inhibitors.

Top1 assays

The standard 10 µl assay mixture contains 20 mM Tris-HCl (pH 7.8), 100 µg/ml bovine serum albumin, 1 mM DTT, 5 mM Mg-acetate, 50 mM NaCl, 10 nM of ³²P-end-labeled oligonucleotide substrate or 25 nM of Cy3-end-labeled oligonucleotide, and enzyme. Incubations were carried out at 30°C for the indicated time periods. Deviations for the standard assay conditions are indicated in the legends of the figures. Reactions were stopped with stop buffer containing 10 mM final concentration of EDTA, 0.05% SDS, and 40% formamide. Samples were analyzed on a 7 M urea, 17% polyacrylamide gel. Cy3 gels were directly imaged on a Typhoon fluorescence imager, while ³²P gels were dried and subjected to phosphorimaging. ImageJ was used for quantification. Images were contrast-enhanced for better visualization.

Repair assays were generally carried out in two stages. In the first stage, the rNMP-containing substrate was treated with a twofold excess concentration of Top1 to produce a mixture of full-length substrate, rNMP 2',3'-cyclic phosphate nicked substrate, and Top1-covalently linked substrate. The reactions were stopped with stop buffer as previously described. To purify the 2',3'-cyclic phosphate-containing substrate, the DNA was ethanol-precipitated and used for subsequent reactions. The Top1-covalently linked substrate was further processed by treatment with 0.2 mg/ml proteinase K at 42°C for 30 min, followed by phenol-chloroform extraction and ethanol precipitation. These products were rehybridized to complementary DNA for further studies or treated consecutively with Tdp1 and Tpp1, as described in legends to figures, in order to generate the free 3'-hydroxyl DNA.

All experiments were carried out three independent times or more, with the exception of the experiment in Fig 4, which was carried out twice, but showed high reproducibility. The errors of quantification were small for most products (< 5%). Errors were higher in the quantification of Top1-cc and of complexes isolated by phenol extraction and ethanol precipitation (10–20%). The graphs in Figs 1 and 3 are representative experiments with errors of quantification shown. The graphs in Figs 5 and 6 show averages of two and three experiments, respectively, and standard deviations are shown.

Supplementary information for this article is available online: <http://emboj.embopress.org>

Acknowledgements

The authors thank Carrie Stith for help with protein purification. This work was supported by grant GM032431 from the National Institutes of Health.

Author contributions

JLS and PB designed and carried out the experiments and wrote the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Ayyagari R, Gomes XV, Gordenin DA, Burgers PM (2003) Okazaki fragment maturation in yeast. I. Distribution of functions between FEN1 AND DNA2. *J Biol Chem* 278: 1618–1625
- Bylund GO, Majka J, Burgers PM (2006) Overproduction and purification of RFC-related clamp loaders and PCNA-related clamps from *Saccharomyces cerevisiae*. *Methods Enzymol* 409: 1–11
- Champoux JJ (2001) DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem* 70: 369–413
- Cho JE, Kim N, Li YC, Jinks-Robertson S (2013) Two distinct mechanisms of Topoisomerase 1-dependent mutagenesis in yeast. *DNA Repair (Amst)* 12: 205–211
- Chon H, Sparks JL, Rychlik M, Nowotny M, Burgers PM, Crouch RJ, Cerritelli SM (2013) RNase H2 roles in genome integrity revealed by unlinking its activities. *Nucleic Acids Res* 41: 3130–3143
- Christiansen K, Westergaard O (1999) Mapping of eukaryotic DNA topoisomerase I catalyzed cleavage without concomitant religation in the vicinity of DNA structural anomalies. *Biochim Biophys Acta* 1489: 249–262
- Clausen AR, Lujan SA, Burkholder AB, Orebaugh CD, Williams JS, Clausen MF, Malc EP, Mieczkowski PA, Fargo DC, Smith DJ, Kunkel TA (2015) Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation. *Nat Struct Mol Biol* 22: 185–191
- Crow YJ, Leitch A, Hayward BE, Garner A, Parmar R, Griffith E, Ali M, Semple C, Aicardi J, Babul-Hirji R, Baumann C, Baxter P, Bertini E, Chandler KE, Chitayat D, Cau D, Dery C, Fazzi E, Goizet C, King MD *et al* (2006) Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutieres syndrome and mimic congenital viral brain infection. *Nat Genet* 38: 910–916
- Das U, Shuman S (2012) Mechanism of RNA 2',3'-cyclic phosphate end healing by T4 polynucleotide kinase-phosphatase. *Nucleic Acids Res* 41: 355–365
- Debethune L, Kohlhagen G, Grandas A, Pommier Y (2002) Processing of nucleopeptides mimicking the topoisomerase I-DNA covalent complex by tyrosyl-DNA phosphodiesterase. *Nucleic Acids Res* 30: 1198–1204
- Desai SD, Liu LF, Vazquez-Abad D, D'Arpa P (1997) Ubiquitin-dependent destruction of topoisomerase I is stimulated by the antitumor drug camptothecin. *J Biol Chem* 272: 24159–24164
- Eissenberg JC, Ayyagari R, Gomes XV, Burgers P (1997) Mutations in yeast proliferating cell nuclear antigen define distinct sites for interaction with DNA polymerase delta and DNA polymerase epsilon. *Mol Cell Biol* 17: 6367–6378
- Fortune JM, Stith CM, Kissling GE, Burgers PM, Kunkel TA (2006) RPA and PCNA suppress formation of large deletion errors by yeast DNA polymerase delta. *Nucleic Acids Res* 34: 4335–4341
- Gelperin DM, White MA, Wilkinson ML, Kon Y, Kung LA, Wise KJ, Lopez-Hoyo N, Jiang L, Piccirillo S, Yu H, Gerstein M, Dumont ME, Phizicky EM, Snyder M, Grayhack EJ (2005) Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes Dev* 19: 2816–2826
- Gomes XV, Burgers PMJ (2000) Two modes of FEN1 binding to PCNA regulated by DNA. *EMBO J* 19: 3811–3821
- Gomes XV, Gary SL, Burgers PM (2000) Overproduction in *Escherichia coli* and characterization of yeast replication factor C lacking the ligase homology domain. *J Biol Chem* 275: 14541–14549
- Henningfeld KA, Hecht SM (1995) A model for topoisomerase I-mediated insertions and deletions with duplex DNA substrates containing branches, nicks, and gaps. *Biochemistry* 34: 6120–6129

- Henricksen LA, Umbricht CB, Wold MS (1994) Recombinant replication protein A: expression, complex formation, and functional characterization. *J Biol Chem* 269: 11121–11132
- Kim N, Huang SN, Williams JS, Li YC, Clark AB, Cho JE, Kunkel TA, Pommier Y, Jinks-Robertson S (2011) Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I. *Science* 332: 1561–1564
- Koh KD, Balachander S, Hesselberth JR, Storici F (2015) Ribose-seq: global mapping of ribonucleotides embedded in genomic DNA. *Nat Methods* 12: 251–257
- Lebedeva N, Rechkunova N, Boiteux S, Lavrik O (2008) Trapping of human DNA topoisomerase I by DNA structures mimicking intermediates of DNA repair. *IUBMB Life* 60: 130–134
- Lippert MJ, Kim N, Cho JE, Larson RP, Schoenly NE, O'Shea SH, Jinks-Robertson S (2010) Role for topoisomerase 1 in transcription-associated mutagenesis in yeast. *Proc Natl Acad Sci USA* 108: 698–703
- Nick McElhinny SA, Kumar D, Clark AB, Watt DL, Watts BE, Lundstrom EB, Johansson E, Chabes A, Kunkel TA (2010a) Genome instability due to ribonucleotide incorporation into DNA. *Nat Chem Biol* 6: 774–781
- Nick McElhinny SA, Watts BE, Kumar D, Watt DL, Lundstrom EB, Burgers PM, Johansson E, Chabes A, Kunkel TA (2010b) Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc Natl Acad Sci USA* 107: 4949–4954
- Pommier Y, Leo E, Zhang H, Marchand C (2010) DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chem Biol* 17: 421–433
- Potenski CJ, Niu H, Sung P, Klein HL (2014) Avoidance of ribonucleotide-induced mutations by RNase H2 and Srs2-Exo1 mechanisms. *Nature* 511: 251–254
- Pouliot JJ, Yao KC, Robertson CA, Nash HA (1999) Yeast gene for a Tyr-DNA phosphodiesterase that repairs topoisomerase I complexes. *Science* 286: 552–555
- Reijns MA, Rabe B, Rigby RE, Mill P, Astell KR, Lettice LA, Boyle S, Leitch A, Keighren M, Kilanowski F, Devenney PS, Sexton D, Grimes G, Holt IJ, Hill RE, Taylor MS, Lawson KA, Dorin JR, Jackson AP (2012) Enzymatic removal of ribonucleotides from DNA is essential for Mammalian genome integrity and development. *Cell* 149: 1008–1022
- Reijns MA, Kemp H, Ding J, de Proce SM, Jackson AP, Taylor MS (2015) Lagging-strand replication shapes the mutational landscape of the genome. *Nature* 518: 502–506
- Rydberg B, Game J (2002) Excision of misincorporated ribonucleotides in DNA by RNase H (type 2) and FEN-1 in cell-free extracts. *Proc Natl Acad Sci USA* 99: 16654–16659
- Sawaya R, Schwer B, Shuman S (2003) Genetic and biochemical analysis of the functional domains of yeast tRNA ligase. *J Biol Chem* 278: 43928–43938
- Sekiguchi J, Shuman S (1997) Site-specific ribonuclease activity of eukaryotic DNA topoisomerase I. *Mol Cell* 1: 89–97
- Sparks JL, Chon H, Cerritelli SM, Kunkel TA, Johansson E, Crouch RJ, Burgers PM (2012) RNase H2-initiated ribonucleotide excision repair. *Mol Cell* 47: 980–986
- Stingle J, Schwarz MS, Bloemeke N, Wolf PG, Jentsch S (2014) A DNA-dependent protease involved in DNA-protein crosslink repair. *Cell* 158: 327–338
- Takahashi T, Burguiere-Slezak G, Van der Kemp PA, Boiteux S (2010) Topoisomerase 1 provokes the formation of short deletions in repeated sequences upon high transcription in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 108: 692–697
- Vance JR, Wilson TE (2001) Repair of DNA strand breaks by the overlapping functions of lesion-specific and non-lesion-specific DNA 3' phosphatases. *Mol Cell Biol* 21: 7191–7198
- Vance JR, Wilson TE (2002) Yeast Tdp1 and Rad1-Rad10 function as redundant pathways for repairing Top1 replicative damage. *Proc Natl Acad Sci USA* 99: 13669–13674
- Walker J, Crowley P, Moreman AD, Barrett J (1993) Biochemical properties of cloned glutathione S-transferases from *Schistosoma mansoni* and *Schistosoma japonicum*. *Mol Biochem Parasitol* 61: 255–264
- Williams JS, Smith DJ, Marjavaara L, Lujan SA, Chabes A, Kunkel TA (2013) Topoisomerase 1-mediated removal of ribonucleotides from nascent leading-strand DNA. *Mol Cell* 49: 1010–1015
- Williams JS, Kunkel TA (2014) Ribonucleotides in DNA: origins, repair and consequences. *DNA Repair (Amst)* 19: 27–37

SUPPLEMENTARY MATERIAL

Error-free and mutagenic processing of Topoisomerase I-provoked damage at genomic ribonucleotides

Justin L. Sparks^{1,2} and Peter M. Burgers^{1*}

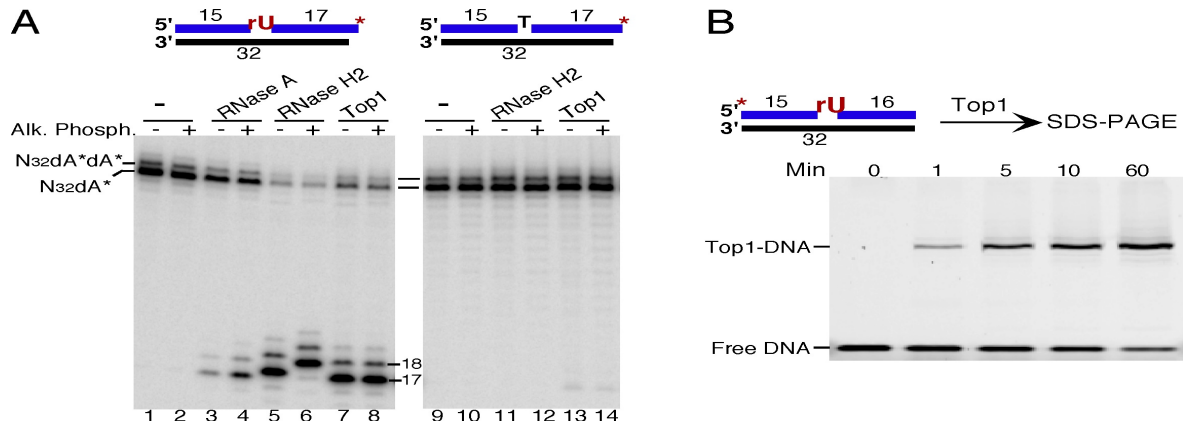


Figure S1 - Top1-cc formation requires embedded ribonucleotides.

A A 32-mer oligonucleotide containing either a single rUMP or dTMP at position 16 was 3'-³²P labeled with [α -³²P]-dATP and terminal transferase. About 10% of the oligo was labeled with 2 dAMPs which accounts for the doublet bands observed in the gel. The labeled oligos were hybridized to the complementary oligo and then incubated with either 1 mg/ml RNase A, 0.5 nM RNase H2, or 100 nM Top1 for 10 minutes at 30 °C under standard assay conditions followed by treatment with alkaline phosphatase at 37 °C for 15 minutes where indicated. The product produced by RNase H2 treatment has a 5'-phosphate that is removed by additional alkaline phosphatase treatment (compare lanes 5 and 6). In contrast, the product produced by Top1 treatment does not have a 5'-phosphate (compare lanes 7 and 8), and is 1 nt shorter.

B Formation of Top1-cc on rNMP-containing dsDNA substrate. Standard assays contained 25 nM 5'-cy3-labeled 32-mer ssDNA oligonucleotide containing a single rNMP, and 100 nM Top1 for the indicated times. The reactions were analyzed by 11% SDS-PAGE.

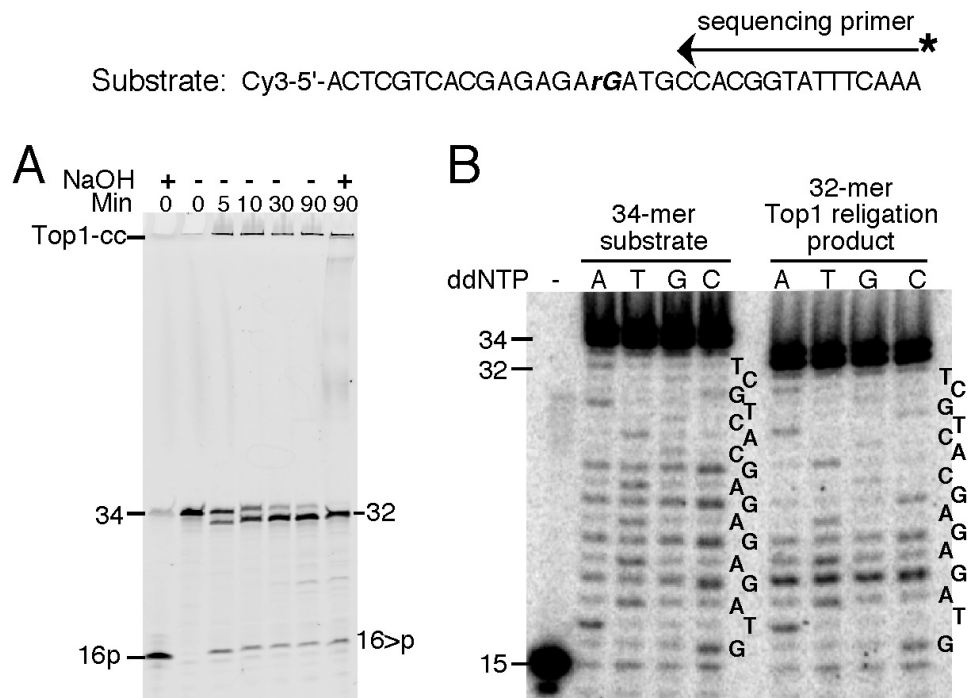


Figure S2 - Top1 religation product leads to removal of di-nucleotide including the rNMP from DNA.

A The rNMP is removed in the Top1 religation products. Standard assay mixtures used 25 nM 5'-cy3-labeled oligonucleotide with a single rGMP, as indicated, in the (GA)₄ repeat hybridized to the complementary DNA strand. The reactions were carried out with 250 nM Top1 at 30 °C for the time indicated. The 0 and 90-min samples were also treated with 0.3 M NaOH at 60 °C for 30 minutes and neutralized with 0.3 M acetic acid. Analysis was on a 7 M urea-17% polyacrylamide gel.

B The predicted dArG dinucleotide is deleted from the Top1 religation product. The 32-nucleotide religation product as obtained in (A) was PAGE-purified on a 7 M urea-17% polyacrylamide gel, followed by hybridization with a ³²P-5'-end-labeled 15-nucleotide primer. Dideoxysequencing reactions were carried out using sequenase at 30°C using the indicated termination mix, and analyzed on a 7 M urea-17% polyacrylamide gel. Sequence positions are indicated to the right of the sequencing ladders.

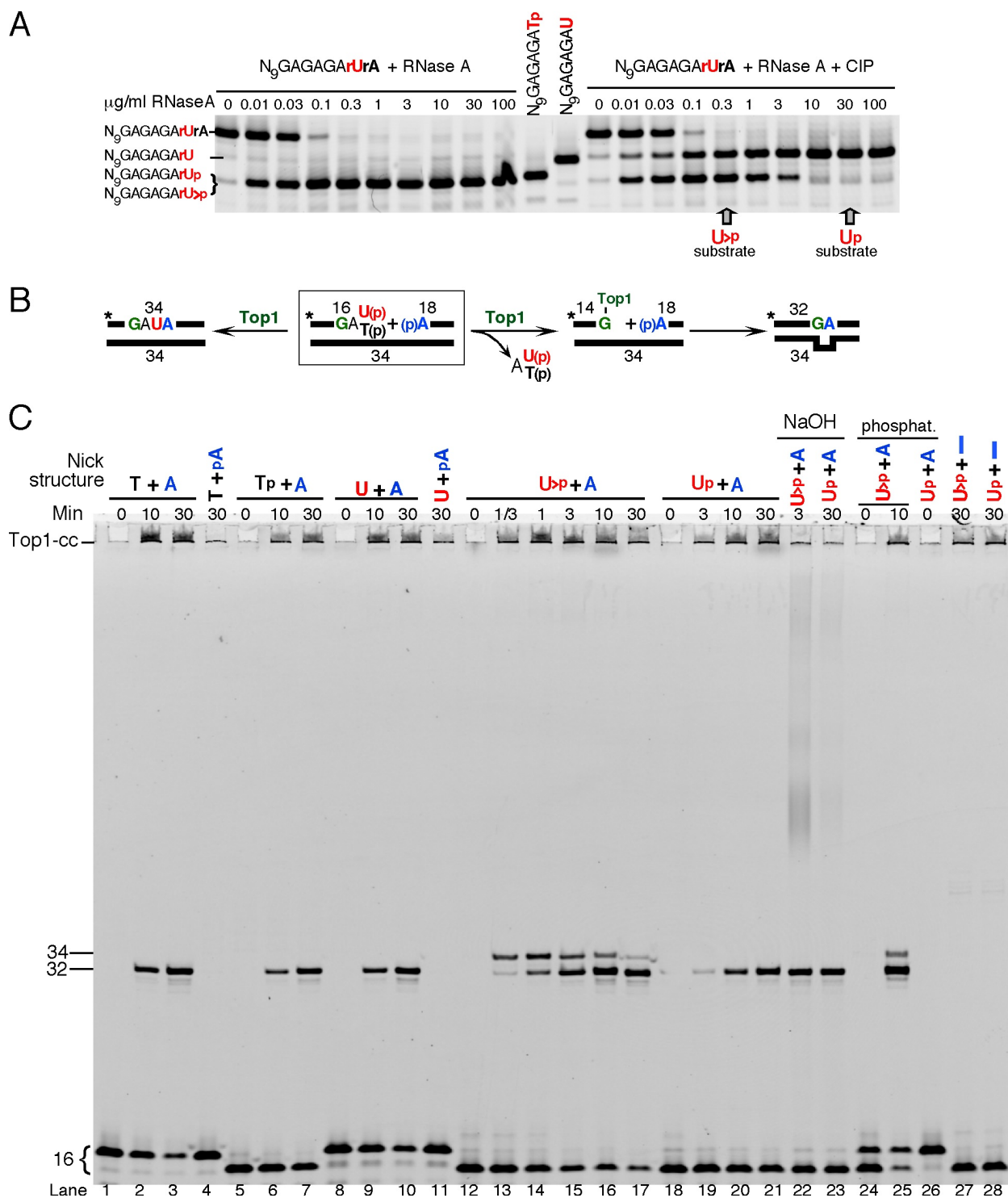


Figure S3 - Structures of RNA-DNA and DNA-DNA nicks determine processing by Top1.

A Synthesis of the oligonucleotides used for making nick substrates. 5'-cy3-
ACTCGTCACGAGAGArUrA (5 μ M) was treated in standard assay buffer with increasing
concentrations of RNaseA as indicated, for 15 min at 37 $^{\circ}$ C. The reactions were terminated by
addition of phenol/chloroform. After a second phenol/chloroform extraction and ethanol
precipitation, the oligonucleotide was re-dissolved in RNase A-free buffer, and an aliquot of

each sample was treated with alkaline phosphatase. The samples were analyzed on a 7 M urea-17% polyacrylamide gel. The sample at 0.3 $\mu\text{g/ml}$ RNase A showed no more starting material and only 30% oligonucleotide 3'-phosphate, which is sensitive to alkaline phosphatase, whereas the remaining 70% was unaffected by alkaline phosphatase, indicating that the oligonucleotide was terminated with a 2',3'-cyclic phosphate. The sample at 30 $\mu\text{g/ml}$ was almost entirely converted to the 3'-phosphate. These two oligonucleotides were used for the experiments.

B Schematic for the Top1-dependent activity at various RNA-DNA and DNA-DNA nick structures. A 5'-cy3-labeled 16mer with various 3'-terminated structures and an 18mer with either 5'-phosphate or 5'-hydroxyl were hybridized to a template 34-mer leaving a nick (the starting substrate is boxed). Top1 can either catalyze religation, but only when the nick contains a 2',3'-cyclic phosphate, or catalyze a second cleavage and potentially catalyze subsequent religation after DNA realignment, giving a 2 nt deletion product. The upstream top strand is 5'-cy3 labeled (indicated with *).

C The different RNA-DNA and DNA-DNA nick substrates were treated with 100 nM Top1 under standard assay conditions for the indicated times at 30 °C, and analyzed on a 7 M urea-17% polyacrylamide gel. The labeled 16-mer upstream oligos were terminated with a dT with either no phosphate (T, lanes 1-4) or a 3'-phosphate (Tp, lanes 5-7), or terminated with an rU with either no phosphate (U, lanes 8-11), a 2',3'-cyclic phosphate (U>p, lanes 12-17, 22,24,25,27), or a 3'-phosphate (Up, lanes 18-21,23,26,28) and the 18mer was 5' terminated with either a 5'-hydroxyl (A) or a 5'-phosphate (pA), or no downstream 18-mer was hybridized to the complementary strand (lanes 27,28). The samples in lanes 24-26 were first treated with Top1 for the indicated time, then incubated at 70 °C for 10 min to inactivate Top1, followed by treatment with alkaline phosphatase. This analysis showed that after 10 min, the U>p + A substrate contained negligible 2',3'-cyclic phosphate oligonucleotide (lane 25), while most of the contaminating 3'-phosphate substrate still remained (compare lane 24 with 25). This gel was processed and select sections are shown in Fig. 4B. Note that the lane numberings of this figure and that in 4B do not correspond.

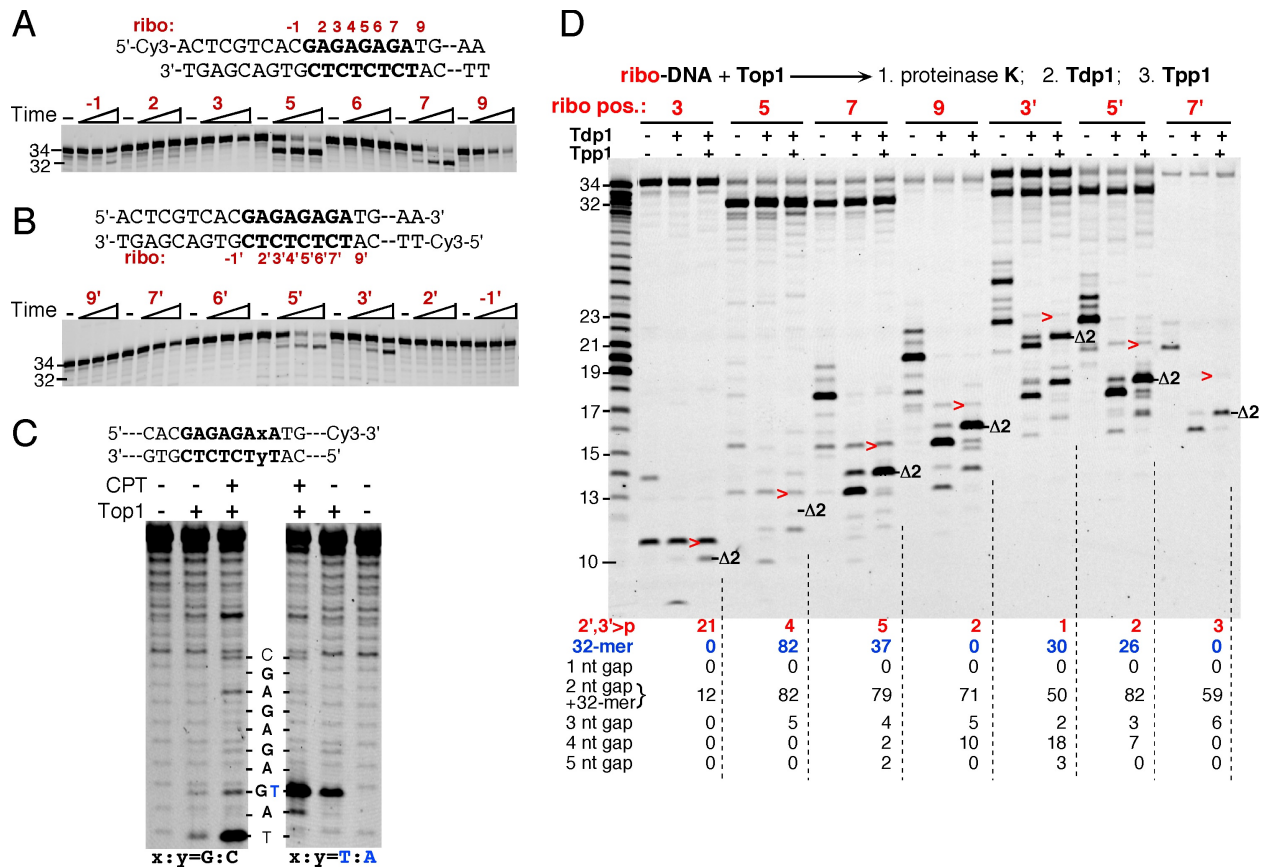


Figure S4. Position of the rNMP within a repeat sequence determines the efficiency of Top1-mediated deletion formation. Associated with Fig. 5. (A,B) A series of dsDNA substrates was made with the rNMP either on the top strand (positions -1 - 9, in which position 1 is the first G in the (GA)₄ repeat), or on the bottom strand (positions -1' - 9', as indicated). In each substrate, the rNMP-containing strand was 5'-labeled with a cy3 fluorophore. The series of substrates were treated with 100 nM Top1 under standard assay conditions at 30 °C for 10, 30, or 90 min, and analyzed on a 7 M urea-17% polyacrylamide gel. The 32-34 regions are shown. (C) The sequence preference for the initial Top1 cleavages on the (GA)₄ and (GA)₄TA containing strands for the substrates shown was determined using the same oligonucleotides as in Fig. 5 and Fig. 4, respectively, but lacking a rNMP, and labeled with a 3'-cy3-label. The DNA was either not treated, or treated with 100 nM Top1 with or without 5 μM camptothecin for 1 min at 30 °C, and analyzed on a 7 M urea-17% polyacrylamide gel. (D) Top1 reactions with the indicated substrates for 90 min, as in (A) and (B), were stopped with 10 mM EDTA and 0.5% SDS and treated with proteinase K to degrade Top1 in the Top1-cc, leaving residual DNA-tyrosyl peptides. Proteinase K was removed by two phenol/chloroform extractions followed by ethanol precipitation. The mixtures of DNA and DNA-tyrosyl peptide products (25 nM total concentration) were incubated, or not, with 25 nM Tdp1 or Tpp1 or both for 10 minutes at 30 °C, and analyzed on a 7 M urea-17% polyacrylamide gel. For each set of three lanes, the > marking indicates the position of the cyclic phosphate, and the Δ2 marking indicates the expected position of the oligonucleotide resulting from Top1 cutting 2 nt upstream of the cyclic phosphate. The

table beneath the figure shows a quantification of each of the products formed after proteinase K, Tdp1 and Tpp1 action. A x-nt gap results from Top1 cutting x nt 5'- of the ribo-2'3'-cyclic phosphate. Since the 32-mer deletion product resulted from realignment and ligation of a 2-nt gap, the % of deletion product and the actual 2-nt gap product were added together to yield the total % of Top1 cutting at a position 2 nt 5'- of the ribo-2'3'-cyclic phosphate.