MOLECULAR AND CELLULAR BIOLOGY, Apr. 1996, p. 1805–1812 Vol. 16, No. 4
0270-7306/96/$04.00+0

Topoisomerase I Involvement in Illegitimate Recombination in *Saccharomyces cerevisiae*

JIE ZHU AND ROBERT H. SCHIESTL*

Department of Molecular and Cellular Toxicology, Harvard School of Public Health, Boston, Massachusetts 02115

Received 3 October 1995/Returned for modification 10 November 1995/Accepted 17 January 1996

Chromosome aberrations may cause cancer and many heritable diseases. Topoisomerase I has been suspected of causing chromosome aberrations by mediating illegitimate recombination. The effects of deletion and of overexpression of the topoisomerase I gene on illegitimate recombination in the yeast *Saccharomyces cerevisiae* have been studied. Yeast transformations were carried out with DNA fragments that did not have any homology to the genomic DNA. The frequency of illegitimate integration was 6- to 12-fold increased in a strain overexpressing topoisomerase I compared with that in isogenic control strains. Hot spot sequences [(G/C)(A/T)T] for illegitimate integration target sites accounted for the majority of the additional events after overexpression of topoisomerase I. These hot spot sequences correspond to sequences previously identified in vitro as topoisomerase I preferred cleavage sequences in other organisms. Furthermore, such hot spot sequences were found in 44% of the integration events present in the *TOP1* wild-type strain and at a significantly lower frequency in the *top1*Δ strain. Our results provide in vivo evidence that a general eukaryotic topoisomerase I enzyme nicks DNA and ligates nonhomologous ends, leading to illegitimate recombination.

Illegitimate recombination (IR) joins two DNA molecules that do not have extensive homology. In most eukaryotic organisms other than *Saccharomyces cerevisiae*, IR events occur more frequently than homologous recombination (21). IR events are associated with chromosome aberrations, which have been implicated in carcinogenesis and many heritable diseases. Despite the importance of IR, however, very little is known about the mechanism and the genetic control of this type of recombination.

Previous analysis of IR in *S. cerevisiae* has been carried out by transforming DNA fragments that contain the *URA3* gene into a *ura3* deletion strain that lacks any homology to the transforming DNA (25). Since the transforming DNA fragment contains no yeast replication origin, the obtained Ura^+ colonies arise after integration of the *URA3*-containing fragment into the genome or after incorporation of mitochondrial DNA that functions as an origin of replication in the nucleus (23, 25). The transforming DNA integrates into different positions in the genome, and several classes of illegitimate integration events have been found (25). About 40% of integration sites contain (G/C)TT sequences, suggesting that topoisomerase I (Topo I) may be involved in IR (23).

By cutting and religating reactions, DNA topoisomerases change the superhelical state of DNA, which is important for transcription and replication of DNA. There are two types of DNA topoisomerases in eukaryotic cells. Type I topoisomerases, including Topo I and Topo III, generate nicks on one DNA strand, while type II topoisomerases generate double-strand breaks (16, 34). Topo I covalently binds to the 3′ end of the nick before religation, and the nicking and religation steps are normally associated. Specialized type I topoisomerases, such as lambda phage integrase, Tn*3* resolvase, yeast FLP recombinase, and P1 phage Cre protein, catalyze DNA strand transfer during site-specific recombination (14, 18). For these recombination events, the site specificity correlates with the sequence specificity of the purified recombinase enzyme. The topoisomerase activity measured by relaxation of supercoiled DNA is also dependent on these specific sequences.

General eukaryotic Topo I enzymes from different species have similar recognition and cleavage sites. Topo I usually generates a nick immediately next to the 3′ end of 5′ (G/C)(A/T)T 3′ sequences (6). These Topo I recognition sites are highly conserved in species ranging from monkeys and rats (3, 20) to wheat to *Tetrahymena pyriformis* (2, 30). The preferred cleavage sequences for Topo III, the second type I topoisomerase in eukaryotes (32), seem more random and different from these Topo I recognition sites (16).

Several lines of evidence suggest that Topo I activities might promote IR between nonhomologous DNA molecules. First, sequences at the crossover points for excision of simian virus 40 DNA from chromosomes are the same as the simian virus 40 DNA sites cleaved by Topo I enzyme in vitro (5). Second, analysis of the hepadnavirus integration sites has identified the presence of Topo I preferred sites, and Topo I can promote the linkage of virus DNA to a 5′ OH end of a heterologous DNA or to a linear double-stranded DNA in vitro (33). Finally, vaccinia virus Topo I, which is packaged within the virus particles and might have a specialized role in recombination and the life cycle of the virus, has been shown to promote IR in *Escherichia coli* (28, 29).

In this study, we have used the yeast transformation system mentioned above to determine whether the general yeast Topo I is involved in IR in *S. cerevisiae*. We determined illegitimate integration frequencies in isogenic strains carrying either the *TOP1* wild-type gene, *top1*Δ, or a *GAL1-TOP1* overexpression construct. We cloned and determined the DNA sequences flanking the integrated fragments of 54 independent Ura^+ integrants from both *TOP1* overexpression and *top1*Δ strains. We found that overexpression of the *TOP1* gene increased the frequency of illegitimate integration 6- to 12-fold and that it also increased the proportion of events of integration into sites that are the same as the Topo I preferred sites in other organisms. The results provide in vivo evidence for an activity of a general eukaryotic Topo I enzyme that nicks DNA and ligates

* Corresponding author. Mailing address: Department of Molecular and Cellular Toxicology, Harvard School of Public Health, 665 Huntington Ave., Boston, MA 02115. Phone: (617) 432-4410. Fax: (617) 432-1780. Electronic mail address: schiestl@mbcrr.harvard.edu.

nonhomologous DNA ends, leading to IR, which is otherwise the hallmark of specialized recombinases.

MATERIALS AND METHODS

Strains and media. *S. cerevisiae* RSY12 (*MAT***a** *leu2-3,112 his3-11,15 ura3*Δ*::HIS3*) (25) containing a deletion of the *URA3* coding sequence was used. All transformation experiments were performed with RSY12 and isogenic derivatives. A derivative of RSY12 lacking the *TOP1* gene (*top1*Δ) was constructed by using plasmid CB25 (kindly provided by R. Sternglanz [31]). In this plasmid, 849 bp of the *TOP1* coding sequence is replaced with the *LEU2* gene. A *Hin*dIII fragment containing this disruption was transformed into RSY12 to create a deletion of the genomic *TOP1* gene.

To construct the *GAL1-TOP1* overexpression strain, plasmid pJZ104 was used (see below for construction). A 4-kb *Hin*dIII-*Pst*I DNA fragment that contained *LEU2* flanked by *TOP1* upstream sequences and by *GAL1-TOP1* was excised from pJZ104. This fragment was transformed into strain RSY12. Leu^+ transformants in which the genomic copy of the *TOP1* gene was replaced with *GAL1-TOP1* were selected. Both the *top1*Δ and the *GAL1-TOP1* constructs were confirmed by Southern blotting and by Topo I assays. Growth conditions and medium preparation were standard (27).

Plasmids. Two plasmids, YIp*lac*211 and pM20, were used as DNA substrates in yeast transformation to select for illegitimate integration events. Plasmid YIp*lac*211 (13) contains the *URA3* gene at the *Eco*O109 site of pUC19. Plasmid pM20 contains the *URA3* gene placed between the *Bam*HI sites of pUC7 (25). YEp*lac*195 (13) contains the *URA3* gene in a 2μm plasmid. Plasmid pJZ102 contains the *URA3* gene flanked by *LYS2* sequences in pUC8 (26) and was used to determine the frequency of homologous recombination events within the cell. Plasmid pWE3 is a CEN plasmid and contains the *URA3* gene and a *GAL1-TOP1* construct (10) and was used as a control plasmid for Topo I activity determination.

Plasmid pJZ104, which contained the *LEU2* marker flanked by *TOP1* upstream sequences and *GAL1-TOP1* sequences, was constructed as follows. The 500-bp *Mlu*I-*Hin*dIII *TOP1* upstream fragment was obtained from plasmid pNL8 (19), and the *Mlu*I end was converted to a blunt end with Klenow fragment. Plasmid pRDG317 (kindly provided by R. Daniel Gietz, University of Manitoba, Winnipeg, Manitoba, Canada) carries the *Hpa*I-*Acc*I fragment of the *LEU2* gene in the *Sma*I site of pUC9. pRDG317 was cut with *Bam*HI (located on the 5′ end of the *LEU2* gene next to the *Hin*dIII site), the *Bam*HI 5′ overhang was then filled in with Klenow fragment, and the plasmid was digested with *Hin*dIII. The *Mlu*I (blunt)-*Hin*dIII *TOP1* upstream fragment was ligated with the *Hin*dIII-*Bam*HI (blunt-end) fragment of pRDG317 to form plasmid pJZ317. The 1.8-kb *Sal*I-*Eco*RI fragment of *GAL1-TOP1* was derived from pNL46 (kindly provided by Nikki Levin [19]). It contains 800 bp of the *GAL1* inducible promoter and 1 kb of *TOP1* coding sequence truncated at the *Eco*RV site. This *Sal*I-*Eco*RI fragment of *GAL1-TOP1* was converted into an *Eco*RI-*Eco*RI fragment by being cloned into the *Eco*RI and *Sal*I sites of pUC7 and then excised by *Eco*RI digestion, since pUC7 has its multicloning site flanked by two *Eco*RI sites. The resulting 1.8-kb fragment was finally inserted into the *Eco*RI sites of pJZ317 at the 3′ end of *LEU2* to form plasmid pJZ104. The constructed fragment contains *TOP1* upstream sequence-*LEU2-GAL1-TOP1* and can be excised by *Hin*dIII and *Pst*I digestion. Integration of this fragment into the genomic *TOP1* gene reconstructs *TOP1* coding sequences.

Genetic and molecular techniques. The CsCl method was used for large-scale preparation of plasmid DNA (22), and Magic Column (Promega) was used for small-scale preparation of DNA.

For yeast transformation, the lithium acetate method as described previously (12, 24) was used. Plasmid YIp*lac*211 was digested to completion with *Bam*HI. Calf intestine phosphatase (Bethesda Research Laboratories) was added to obtain a dephosphorylated DNA substrate, and the DNA was then phenol extracted and ethanol precipitated. A 10-μg sample of DNA was used per transformation mixture of 200 μl of competent cells, and 30 to 50 μg of DNA was used in each transformation experiment.

To obtain the target sequences of the illegitimate integration events, total yeast DNA was isolated from Ura^+ transformants as described by Denis and Young (8). The genomic DNA was digested with *Cla*I or *Bgl*II, and T4 DNA ligase was added to promote self-ligation. The ligation mixture was then used to transform *E. coli* DH5α to recover plasmid YIp*lac*211 and the flanking junction sequences. The junction sequences were determined by dideoxy sequencing of double-stranded DNA templates (17) with primers 1233 and 1212 (New England Biolabs), which flank the multicloning site of YIp*lac*211. The target sites of integration were determined by comparing the junction sequence with GenBank sequences. All data shown are based on actual target site sequences obtained this way.

Total RNA was isolated by using TRI-REAGENT (Molecular Center Research). RNA was electrophoresed on denaturing agarose-formaldehyde gels according to the protocol of Sambrook et al. (22). The blotted membrane was hybridized with a random-labeled ^{32}P probe containing a *TOP1* gene fragment according to the Quik-Hyb protocol (Stratagene).

Topo I activity assay. Cells from an overnight culture were diluted into fresh YPAD medium and grown to log phase for 5 h. Cell extracts were prepared as described previously (9) with the following modifications. Log-phase cells were harvested by centrifugation, washed with cold double-distilled water, and recentrifuged. The cells were resuspended in 0.5 ml of yeast lysis buffer (20 mM Tris-HCl [pH 7.5]–1 mM Na_2EDTA–1 mM dithiothreitol–1 mM phenylmethylsulfonyl fluoride–500 mM KCl–10% glycerol). A 0.4-ml volume of glass beads (type V, 425- to 600-μm diameter; Sigma) was added to the cells, and the cells were processed by 10 cycles of 10 s of vortexing followed by 10 s of cooling on ice. The lysates were spun for 10 min at 4°C in a tabletop centrifuge. Extracts were serially diluted with cell lysis buffer containing 100 mg of bovine serum albumin (BSA) per ml. A 1-μl sample of each dilution was assayed for Topo I activity in a 10-μl reaction volume containing 0.2 μg of mostly supercoiled YEp*lac*181 (13), 20 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 1 mM Na_2EDTA, 1 mM dithiothreitol, and 30 mg of BSA per ml. Relaxation of DNA was monitored by electrophoresis in a 0.7% agarose gel.

Determination of the frequency of IR. The frequency of IR in isogenic strains RSY12, RSY12 *top1*Δ, and RSY12 *GAL1-TOP1* was determined. RSY12 *GAL1-TOP1* cells were first grown overnight in raffinose medium and then diluted to a concentration of 2×10^6/ml in either glucose- or galactose-containing medium. After 6 h of growth, the cells were transformed with *URA3*-containing plasmid YIp*lac*211 or pM20 after each had been completely digested with *Bam*HI and dephosphorylated. After transformation, cells were plated on uracil-deficient synthetic medium containing glucose. At the same time, cells were transformed with plasmid YEp*lac*195, a 2 μm plasmid, to determine transformation efficiencies. IR frequencies were expressed as events per 10^4 2μm plasmid transformants.

RESULTS

Overexpression of the *TOP1* gene increased the frequency of illegitimate integration. To study the effect of Topo I on IR, a *GAL1-TOP1* strain was constructed as described in Materials and Methods. This strain contained its only genomic copy of the *TOP1* gene under the control of a *GAL1* inducible promoter. *GAL1* is induced in the presence of galactose and is suppressed in the presence of glucose. The levels of *TOP1* mRNA in wild-type *TOP1* cells and *GAL1-TOP1* cells in glucose- or galactose-containing medium were determined by Northern (RNA) analysis. Densitometer scanning of the autoradiographs showed that *GAL1-TOP1* cells grown in galactose increased the expression of the *TOP1* gene 18-fold compared with that of *TOP1* wild-type cells and 180-fold compared with that of *GAL1-TOP1* cells in glucose medium. No expression signal was found for the *top1*Δ strain (data not shown).

We compared the Topo I activities in extracts of *TOP1* wild-type, *top1*Δ, and *GAL1-TOP1* cells by assaying the relaxation of supercoiled DNA (Fig. 1). When grown on galactose, *GAL1-TOP1* cells and those carrying pWE3 (CEN, *GAL1-TOP1*) (10) gave an estimated ninefold increase in Topo I activity over the wild-type level on the basis of the appearance of relaxed plasmid DNA (Fig. 1). Similar induction levels were previously shown in a different strain transformed with pWE3 (10). We found no activity in *top1*Δ cells and very little activity in the *GAL1-TOP1* strain grown on glucose (Fig. 1).

Overexpression of *TOP1* in galactose medium increased the IR frequency 6- to 12-fold compared with that of the *GAL1-TOP1* cells in glucose medium, in which the *TOP1* gene is repressed (Table 1). The frequencies of homologous integration in the cells grown in different media were comparable. These results indicate that overexpression of Topo I increased the frequency of IR but not of homologous recombination. The *top1*Δ strain showed an IR frequency similar to that of the *TOP1* wild-type strain. About 40% of integration sites in the *TOP1* wild type contain (G/C)TT sequences, suggesting that Topo I may be involved in IR (23). In the *top1*Δ strain, these Topo I-mediated events may be lacking. However, the variation between different experiments was too large to allow detection of an expected 40% difference in IR frequency.

Topo I mediated illegitimate integration into genomic DNA next to preferred target sites. The target sites of YIp*lac*211 illegitimate integration events in cells carrying either the *top1*Δ allele or the *GAL1-TOP1* allele, in which Topo I was either

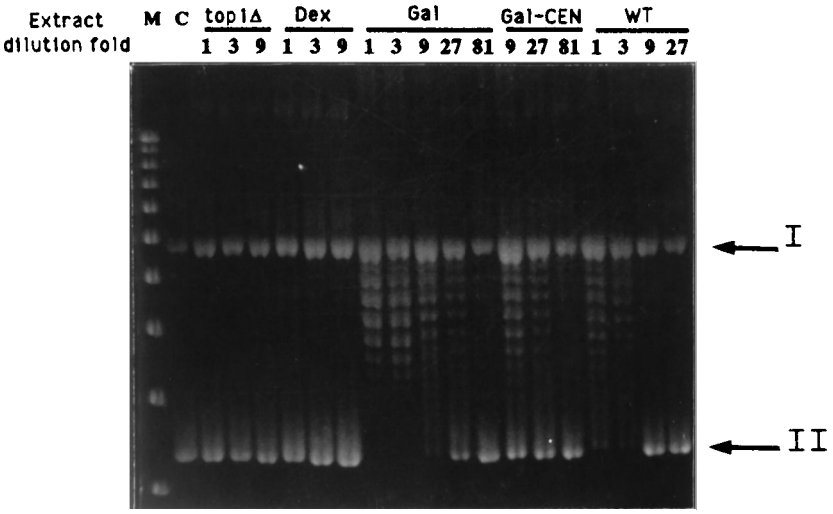

FIG. 1. Assay for DNA Topo I activity. Cell extracts of RSY12 *TOP1* wild-type (WT), *top1*Δ and *GAL1-TOP1* strains on glucose (Dex) and on galactose (Gal) and strain RSY12 *top1*Δ with CEN plasmid pWE3 containing *TOP1* under the control of *GAL1* (Gal-CEN) were prepared as described in Materials and Methods. Lane M, size marker; lane C, control (no extract). The strains from which the extracts were made and the dilution factor of each extract that was incubated with the plasmid substrate are indicated above the lanes. All of the extracts were first adjusted to the same protein concentration (0.6 mg/ml). The positions of nicked DNA (I) and supercoiled DNA (II) are indicated. A characteristic ladder of partially relaxed DNA appears upon Topo I activity.

repressed or overexpressed, respectively, were determined. A total of 54 independent Ura$^+$ transformants were analyzed to determine integration junction sequences. Genomic DNA from each Ura$^+$ colony was isolated and was digested with enzymes that do not cut within plasmid YIp*lac*211. After self-ligation of the resulting fragments, the DNA was transformed into *E. coli* to obtain ampicillin-resistant colonies. The recovered plasmids contained the junction sequences of integration events, and these junctions were sequenced with primers flanking the pUC19 multicloning site. To determine the target sequences, the junction sequences were compared with sequences in the GenBank database. Figure 2 illustrates all integration events found in *GAL1-TOP1* cells on galactose (Fig. 2A) and on glucose (Fig. 2B) and in the *top1*Δ strain (Fig. 2C) that gave matches to both junctions (38 of 54) and shows the DNA substrate (YIp*lac*211) and the genomic target sequences. Figure 2 also shows possible base pairings (underlined) between the target site and the 5′ single-stranded overhang GATC of the plasmid as well as the sites of cleavage that are necessary to explain the sequences of the junctions. It was assumed that deletions of base pairs occurred from the plasmid if they could result from deletions either from the target site or from the plasmid.

TABLE 1. Frequency of illegitimate integration after overexpression or deletion of the *TOP1* gene[a]

Genotype	Medium type	Relative frequency of integration[b]		
		IR		Homologous integration with pJZ102[c]
		URA3 fragment[d]	YIp*lac*211[e]	
TOP1$^+$	Glucose	0.50 ± 0.16	0.27 ± 0.02	41 ± 0.9
	Galactose	0.19 ± 0.04	0.24 ± 0.04	46 ± 7.7
*top1*Δ	Glucose	0.12 ± 0.05	0.23 ± 0.07	25 ± 5.7
GAL1-TOP1	Glucose	1.3 ± 0.20	0.60 ± 0.14	61 ± 15
	Galactose	7.1 ± 1.8	7.1 ± 0.09	42 ± 7.1

[a] The frequencies of IR in isogenic wild-type, *top1*Δ and *GAL1-TOP1* strains were determined. The number of integration events per microgram of DNA per 10^4 2μm transformants is shown, which we call the relative frequency of integration. This number is controlled for the transformation efficiency and represents the average of two experiments ± the standard deviation for each experiment and each plasmid. The data for the *GAL1-TOP1* strain in different media were obtained with the same plasmid preparations and at the same time. Similarly, the data for the *TOP1* wild type and the *top1*Δ strain were obtained at the same time with the same plasmid preparations. However, the *GAL1-TOP1* experiments were done with plasmid preparations different from those used in the experiments with the *TOP1* or the *top1*Δ strain. Thus, the lower frequency of IR in the *TOP1* wild type and the *top1*Δ strain than in the *GAL1-TOP1* strain in glucose may be due to the different plasmid preparations and hence may not be significant.

[b] Cultures were separately transformed with plasmid YEp*lac*195 (2μm *URA3*). To control for differences in transformation efficiency, the relative frequency of integration is the frequency of integration events per 10^4 transformants with YEp*lac*195.

[c] pJZ102 was digested with *Eco*RI and *Hin*dIII to release a *URA3* fragment flanked by the 5′ and 3′ ends of *LYS2* and used for homologous integration into the *LYS2* locus.

[d] Plasmid pM20 was digested with *Bam*HI to release a 1.1-kb *URA3* fragment which was used for transformation.

[e] Plasmid YIp*lac*211 (integrating, *URA3*) was digested with *Bam*HI and used for transformation.

Of the 14 target sequences from the galactose-grown *GAL1-TOP1* strain, 12 had 5′ (G/C)(A/T)T 3′ sequences at the 5′ end of either one end or both ends of their junctions. In other organisms, these sequences are the preferred (consensus) recognition and nicking sequences for eukaryotic Topo I (6). Table 2 presents the distribution of Topo I preferred triplet sequences at the 5′ ends of junctions of illegitimate integration in *TOP1* wild-type, *top1*Δ, and *GAL1-TOP1* cells on either glucose or galactose. Sequences of 23 previously published target sites in *TOP1* wild-type cells (23, 25, 26) are also included for comparison purposes. The probability of random integration of a DNA fragment next to a given triplet sequence is $(1/4)^3 \times 2 = 1/64 \times 2 = 1/32$, since each transformation event has two junctions. Hence, the four possible sequences of the Topo I consensus sequence (CTT, GTT, CAT, and GAT) would occur in a random distribution at a frequency of 4/32, or 12.5%. In the *top1*Δ strain or the *GAL1-TOP1* strain in glucose, the percentage of Topo I preferred triplet sequences found next to integration junctions is similar to that expected if integration is random (7 and 11%, respectively). A total of 10

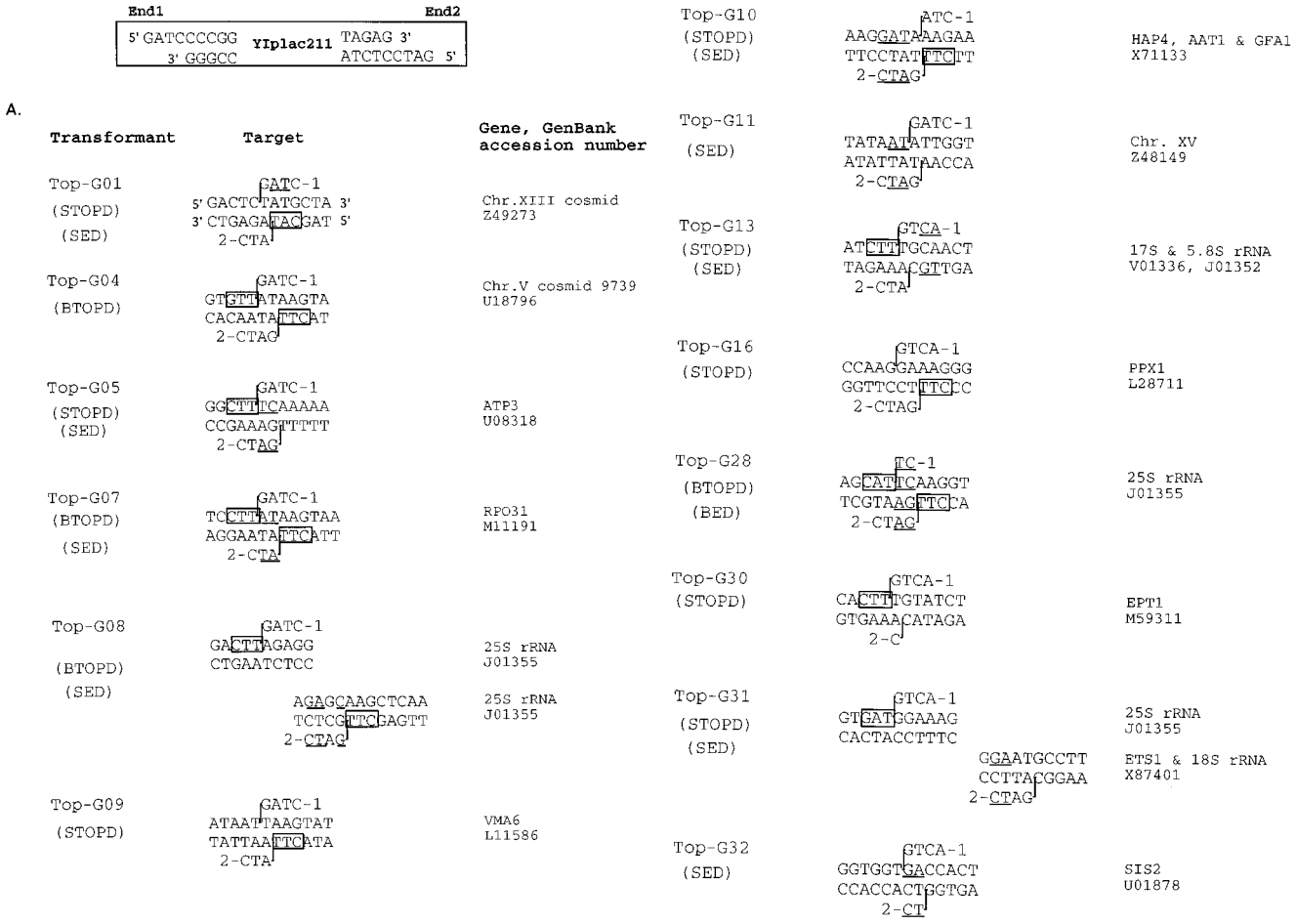

FIG. 2. Target sequences of illegitimate integration events. *Bam*HI-treated YIp*lac*211 DNA was transformed into strain RSY12 *GAL1-TOP1* on galactose (A) and on glucose (B) as well as into strain *top1*Δ (C). The transformed YIp*lac*211 DNA and flanking sequences were rescued, and both ends were sequenced with primers flanking the multicloning site of plasmid YIp*lac*211. The identities of the target sequences were obtained by homology searches of the GenBank database. Shown are the recombination substrate YIp*lac*211, the target sequences and the genes into which the plasmid integrated, and their GenBank accession numbers. Sometimes, one to four bases were deleted during integration from the 5′ ends of plasmid YIp*lac*211. For instance, for transformant Top-G01, one base was lost from end 2 but no base was lost from end 1. The GATC bases from ends 1 and 2 that are shown were present during integration and were recovered as double-stranded regions at the junctions after integration. Underlined sequences represent sequence homology between the 5′ ends of the plasmid and the target. For instance, in transformant Top-G01, the AT sequence of end 1 of the plasmid is homologous to the TA sequence at the target. We have shown how the target sequences were cleaved to allow integration of the ends of the plasmid to give rise to the observed sequences at the junctions. For example, for transformant Top-G01, the C at the target was ligated to the G of end 1 of the plasmid. Topo I preferred sites (boxed) [5′ (G/C)(A/T)T 3′] at the target are indicated. Transformants are classified according to the number of ends of the plasmid that are homologous to the target: SED, single end directed; NED, not end directed; BTOPD and STOPD, both ends or single end Topo I directed (indicating the number of ends at which Topo I might have acted). The direction of the DNA is 5′ to 3′. Chr., chromosome; mitoch., mitochondrial; rDNA, genes coding for rRNA; NTS, nontranslated sequence.

of 23 target sites in the *TOP1* wild-type strain (44% [23, 25, 26]) and 12 of 14 target sites (86%) in the *GAL1-TOP1* strain on galactose contained the Topo I preferred sequences on at least one of the two junctions (Table 2). The differences between the *top1*Δ strain and the strain overexpressing Topo I or between the strain repressing Topo I and the strain overexpressing Topo I are both statistically highly significant ($P < 0.0005$) as determined by using chi-square distribution values. This indicates that most, if not all, additional integrations in cells overexpressing Topo I were into Topo I preferred sites. The difference between the *TOP1* wild type and the *top1*Δ strain is also significant ($P < 0.025$), indicating that Topo I mediates IR in the wild type. These results were obtained by pooling data for the target sites of plasmid pM20 and plasmid YIp*lac*211 for the *TOP1* wild type. However, if just the YIp*lac*211 target sites (23) for the *TOP1* wild type were considered (4 of 10 targets contained Topo I preferred sites) in comparison with the YIp*lac*211 target sites for the *top1*Δ strain, a significant difference ($P < 0.05$) still was obtained.

Four of 12 Topo I preferred sequences found at IR target sites (33%) from the galactose-grown *GAL1-TOP1* strain, were present at both junctions, suggesting that two different Topo I molecules, one at each end of the fragment, might have been involved during integration (Top-G04, Top-G07, Top-G08, and Top-G28; Fig. 2A). Similar flanking Topo I sequences at both junctions of plasmid YIp*lac*211 were found in only 1 of 10 cases (10%) in the *TOP1* wild-type strain, and no such events were found for the *GAL1-TOP1* strain grown in glucose or in the *top1*Δ strain. Larger amounts of Topo I enzyme within the cell might have increased the likelihood that two Topo I molecules were involved in one integration event. In most cases, the two Topo I preferred sequences were close to each other on the target, resulting in 2-bp deletions (Top-G04 and Top-G07, Fig. 2A) or in no loss of sequences (Top-G28, Fig. 2B;

B.

Transformant	Target	Gene, GenBank accession number
Top-d02 (BED)	ATC-1 5′ CATGCCATCTTT 3′ 3′ GTACGGTAGAAA 5′ 2-CTAG	GDH2 X72015
Top-d03 (BED)	ATC-1 GTTTTATCCGGT CAAAATAGGCCA 2-CTAG	Chr.XII cosmid U21094
Top-d05 (SED)	GATC-1 AGCTGGATTCCA TCGACCTAAGGT 2-CTA	MALp (MALR) X15241
Top-d06 (SED)	ATC-1 TATGCATTATC ATACGTAATAG 2-CTA	MSD1(mitoch.) M26020
Top-d07 (SED)	TC-1 AATCTTATGT TTAGAATACA	tRNA-Met(mitoch.) L36888
	ATTATATTTA TAATATAAAT 2-CTA	CYT C2 (mitoch.) V00706
Top-d10 (SED) (STOPD)	ATC-1 CTTTGTATCTT GAAACATAGAA 2-CTA	STM1/TIF3 X71996
Top-d14 (BED)	ATC-1 GTCATTATCGC CAGTAATAGCG 2-CTA	rDNA NTS X00486
Top-d24 (BED)	ATC-1 TCAATCACATTGA AGTTAGTGTAACT 2-CTAG	BSA1 M58057
Top-d35 (BED)	TC-1 CCTGGTTCGTTA GGACCAAGCAAT 2-CTAG	Cosmid 9986 U00027

C.

Transformant	Target	Gene, GenBank accession number
Top1-5 (SED)	GATC-1 5′ GCGAGGAATT 3′ 3′ CGCTCCTTAA 5′ 2-CTAG	Y′ element M58718
Top1-22 (BED)	ATC-1 ATTGCTATTAGA TAACGATAATCT 2-CTA	rDNA X00486
Top1-30 (SED)	TC-1 GCGGTAATTCCA CGCCATTAAGGT 2-CTA	rDNA V01335
Top2-7 (BED)	ATC-1 TAAAATCGCCGC ATTTTTAGCGGCG 2-CTAG	SLT2 X59262
Top2-10 (NED)	TC-1 AAACTGTGAA TTTGACACTT 2-CTAG	PEM1 M16987
Top2-13 (BED)	GATC-1 TTGTGATCCGTC AACACTAGGCAG 2-CTAG	COQ3 M73270
Top2-30 (SED)	1 TCATATTCTTTC AGTATAAGAAAG 2-CTAG	tRNA-Ser V01329
Top203 (BED)	GATC-1 GAATCCGATGT CTTAGGGCTACA 2-CTAG	YBL0421,0438 X77291
Top204 (BED)	GATC-1 GAAGGATTGTGG CTTCCTAACACC 2-CTA	2μm circle J01347
Top206 (SED) (STOPD)	ATC-1 TGCTTTATTAC ACGAATAATG 2-CTA	2μm circle J01347
Top207 (BED)	GATC-1 AGAGAAGATAAT TCTCTTCTATTA 2-CT	Chr.IV cosmid 9410 U33050
Top208 (BED)	GATC-1 GAAAGAGCTCTCAAATCTGT CTTTCTCGAGAGTTTAGACA 2-CTAG	rDNA V0133
Top391 (BED)	GATC-1 ATACCGATCTGATTC TATGGCTAGACTAAG 2-CTAG	Chr.VI cosmid 9993 D44603
Top412 (BED)	ATC-1 CCAACAATCACATGAG GGTTGTTAGTGTACTC 2-CTAG	Chr.XI ORF YKL106W Z28106
Top415 (BED)	GATC-1 GACTAGATCGGGTGGTG CTAGTCTAGCCCACCAC 2-CTAG	18S rDNA J01353

WT2-25 in Fig. 1 of reference 23). However, in one case (Top-G08), the two Topo I sites were separated, leading to the rearrangement of two genes coding for the 25S rRNA.

It has been previously reported that a considerable fraction of the IR target sites contain sequence homology of two, three, or four bases to the 5′ overhangs of the plasmid, namely, the sequence GATC after digestion with *Bam*HI (23, 25, 26) or the sequence AATT after digestion with *Eco*RI (23). Classification of the target sites according to such microhomology-mediated recombination events reveals another difference between the target sites among different strains. In the *GAL1-TOP1* strain grown on galactose, the fraction of target sites containing 3 or 4 bp of homology to the GATC sequence at both ends of the plasmid (both ends directed [BED]) was completely lacking (0 of 14). In contrast, this class was present in the *GAL1-TOP1* strain grown on glucose in 2 of 9 cases (22%; Top-d02 and Top-d03 in Fig. 2B), indicating a trend; in the *top1*Δ strain, this class of BED events was present in 8 of 15 cases (53%; Top2-7, Top2-13, Top203, Top204, Top208, Top391, Top412, and Top415 in Fig. 2C), representing a significant difference ($P < 0.01$); and in the *TOP1* wild-type strain, this class of BED

TABLE 2. Frequency of Topo I preferred triplet sequences next to illegitimate integration target sites

Topo I preferred triplet sequence	No. of triplet sequences present[a]			
	RSY12 *top1*Δ	RSY12 *TOP1* wild type	RSY12 *GAL1-TOP1*	
			In glucose[b]	In galactose[c]
5′ −3 −2 −1 3′				
C T T	1/30	6/46	0/18	12/28
G T T	0/30	4/46	0/18	1/28
C A T	0/30	0/46	0/18	2/28
G A T	0/30	1/46	1/18	1/28
(G/C)(A/T)T				
Per junction	1/30	11/46	1/18	16/28
Per event (%)	1/15 (6.7)	10/23 (44)	1/9 (11)	12/14 (86)

[a] Independent Ura$^+$ transformants were analyzed to determine the integration junction sequences. Genomic DNA from each clone was digested with *Cla*I or *Bgl*II, and self-ligation was promoted. The ligated DNA was used to transform *E. coli* DH5α to recover plasmid YIp*lac*211, which contained the integration junction sequences. The junction sequences were determined by DNA sequencing using primers flanking the YIp*lac*211 multicloning sites. Genomic target sequences were obtained in 15 of 19, 9 of 16, and 14 of 19 junction sequences from the *top1*Δ strain, the *GAL1-TOP1* strain in glucose medium, and the *GAL1-TOP1* strain in galactose medium, respectively, resulting in 38 target sites. Data for additional 23 target sites for the *TOP1* wild type were pooled from the target sites of Fig. 5 of Schiestl and Petes (25), Fig. 1 and 4 of Schiestl et al. (23), and Fig. 1A of Schiestl et al. (26).

[b] In glucose, *GAL1-TOP1* was repressed (Table 1).

[c] Topo I overexpressed (Table 1).

events was present in 10 of 23 cases (44%), also representing a significant difference ($P < 0.01$).

It has also been previously shown that a certain fraction of Ura3$^+$ isolates contained plasmid YIp*lac*211 ligated to mitochondrial DNA sequences (23). Only two target sites containing mitochondrial sequences with both ends giving sequence matches (Top-d06 and Top-d07) were obtained from the *GAL1-TOP1* strain grown on glucose (Fig. 2B). However, several partial target sites (with a match obtained for only one junction) containing mitochondrial sequences (not shown in Fig. 2) were found. Plasmids with mitochondrial sequences were obtained in 3 of 19 isolates of the *top1*Δ strain, 4 of 27 isolates of the *TOP1* wild type, 4 of 15 isolates of the *GAL1-TOP1* strain grown on glucose, and 0 of 18 isolates of the *GAL1-TOP1* strain on galactose. Taken together, these data support our hypothesis that Topo I is involved in IR. In the *GAL1-TOP1* strain on galactose, there is a shift in the ratio of the different IR events in favor of Topo I-mediated events so that the occurrence of microhomology-mediated events and events of ligation of plasmid YIp*lac*211 to mitochondrial DNA is underrepresented.

DISCUSSION

In *S. cerevisiae*, overexpression of *TOP1* increased the frequency of nonhomologous integration of transformed DNA into hot spot sequences 6- to 12-fold over levels in strains with repressed or deleted *TOP1*. These hot spot sequences correspond to sequences established as Topo I preferred sites for wheat germ, *T. pyriformis*, rats, and monkeys by multiple criteria (2, 3, 20, 30). Although the recognition sites of yeast Topo I have not yet been determined, the yeast enzyme is highly conserved in its amino acid sequence and sensitivity to camptothecin (4). Since the Topo I recognition sites are highly conserved from plants to mammals, it seems reasonable to assume that these sites would be the same in yeasts.

In the *TOP1* wild-type strain, a variety of classes of IR events occur, such as microhomology-mediated recombination events, integrations into Topo I preferred sites, and ligation of plasmid YIp*lac*211 to mitochondrial DNA. In the *top1*Δ strain, the occurrence of integrations into Topo I preferred sites is significantly reduced. In the *GAL1-TOP1* strain grown on galactose, there is a shift in the ratio of the different classes of IR events in favor of integration into Topo I preferred sites so that microhomology-mediated events and ligation of plasmid YIp*lac*211 to mitochondrial DNA are underrepresented or missing in our sample. These data strongly suggest that Topo I is involved in IR in the wild-type *TOP1* strain and even more so after Topo I is artificially overexpressed.

If illegitimate integration sites coincide with the preferred binding and nicking sites of Topo I, our data may also indicate that we have identified the preferred sequences for yeast Topo I, which have not been studied before. Our IR target sites show a more stringent sequence consensus than the preferred sequences that have been determined for Topo I enzymes of other organisms with other methods. In the *TOP1* wild type, there were six CTT and four GTT sequences at 46 junctions of the IR events, whereas CAT and GAT occurred at lower frequencies similar to those expected if integration is random (Table 2). After the overexpression of Topo I, 12 of 28 junctions contained CTT sequences and 2 of 28 junctions contained CAT sequences, while integration into either of the other two sequences occurred at a frequency similar to that expected if integration is random. There was a significantly higher frequency of IR junctions containing either CTT sequences alone or CTT plus GTT sequences in the *TOP1* overexpression strain than in the *TOP1* wild type. However, we found that the CTT-to-GTT ratio of 12:1 for the *TOP1* overexpression strain was not significantly different from the CTT-to-GTT ratio of 6:4 for the *TOP1* wild-type strain. Taken together, these results suggest that in the *TOP1* wild type as well as in the *GAL1-TOP1* strain grown on galactose, IR target sites show a strong preference for CTT and GTT. A less stringent consensus sequence of 5′ (G/C)(A/T)T 3′ has been found for wheat germ and calf thymus Topo I (3, 30), and an even less stringent consensus sequence of (G/C/A)(A/T)T has been found for Topo I enzymes from monkey cells (20) and rat liver cells (3). Topo I preferred sites in these organisms have been identified by determining Topo I break sites in vitro with the purified enzyme either with or without camptothecin and in vivo with camptothecin. If illegitimate integration sites coincide with the preferred binding and nicking sites of Topo I, our data may suggest that yeast Topo I has a more stringent preference for (G/C)TT.

Changes in the abundance of Topo I may result in a change in the superhelical state of genomic DNA (34), which may directly or indirectly result in a greater abundance of recombinagenic DNA damage (11). In addition, transcription activity of other genes may be affected by changes in Topo I abundance (7). Thus, the IR-promoting effect after overexpression of Topo I may be indirectly caused by changes in the overall topological state of chromosomal DNA or by regulation of genes involved in recombination. An examination of the integration sites, however, suggests that Topo I is directly involved in IR by nicking chromosomal DNA and promoting joining with the nonhomologous DNA molecules transformed into the cells.

Topo I may function only in nicking and in attracting the transforming DNA to the nick, while some other enzyme such as DNA ligase may be responsible for the ligation step. On the other hand, Topo I may be involved in both nicking and joining steps. In all cases, the Topo I preferred sites occurred 5′ of the integrating fragment without any deletion of sequences at the target site (Fig. 2A). In contrast, in 11 of 12 cases, at least 1 bp

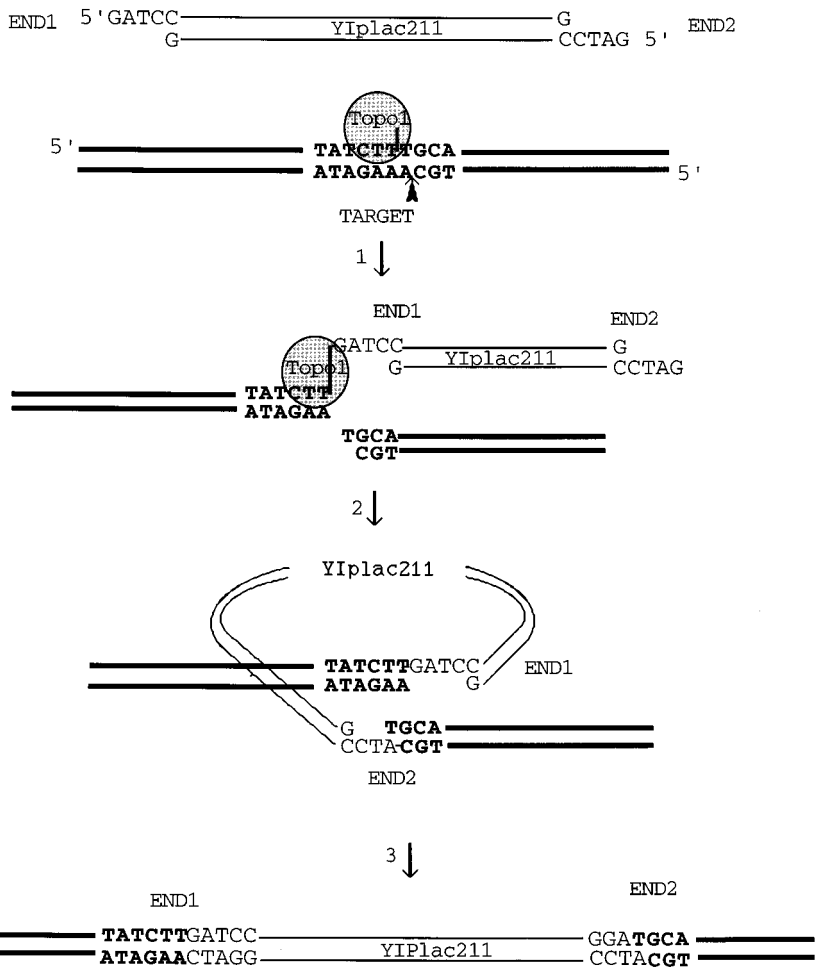

FIG. 3. Possible mechanism for Topo I-mediated integration of plasmid YIp*lac*211, corresponding to event Top-G13 of Fig. 2A. One Topo I molecule binds to the CTT preferred site, makes a nick, and covalently attaches itself to the 3′ end of the T residue. At this point or prior to this point, a nick is made on the opposite strand and a double-strand break develops. In step 1, the Topo I enzyme ligates the 3′ end to end 1 of the YIp*lac*211 recombination substrate. In step 2, end 2 of the recombination substrate is ligated to the other end of the double-strand break. During the integration event, the TA base pair is deleted from the target site. This may have occurred either before the second ligation step and the bases GA were removed from the 5′ single-stranded overhang of end 2 of the substrate, or alternatively, as shown, one T was left at the site of the double-strand break at the target which may have paired with the A of end 2 of the substrate after removal of the terminal G of end 2. In step 3, both single-stranded gaps are filled and the final integration product is obtained.

was deleted from the target site 3′ of the ligation (Fig. 2A). If Topo I were only providing a recombinagenic nick and a subsequent mechanism operated for integration, these deletions may have affected both ends, including the Topo I preferred site. This would have resulted in a more frequent occurrence of (G/C)(A/T) or G/C sequences flanking the integration event, which was not found. This argues that Topo I was involved in both the nicking and the ligation reaction.

Experiments with FLP recombinase, a type I topoisomerase mechanistically similar to eukaryotic Topo I, suggest that the pairing of the FLP target sites is mediated by protein-protein interaction of molecules bound to each site (1). The same is true for gamma-delta transposition (15) and has been suggested for excisional recombination promoted by vaccinia virus Topo I (29). Our experimental design is such that one recombination substrate already provides a double-strand break and a free 5′ OH but no Topo I site at the end. Thus, it is not likely that the two recombination substrates are brought together by protein-protein interaction of two Topo I molecules.

Previous work has shown the stimulatory effect of a vaccinia virus DNA Topo I on IR leading to excision of bacteriophage lambda in *E. coli* (28). Our results are different in several respects from the events catalyzed by vaccinia virus Topo I. The events catalyzed by vaccinia virus Topo I showed that (i) the recognition site was present on both junctions in five of five cases, (ii) the presence of direct repeats at the target sites extended beyond the Topo I recognition sites in three of five cases, and (iii) there were no deletions 3′ of the sites of Topo I cleavage. Thus, vaccinia virus Topo I-promoted recombination might proceed via the model mentioned above by protein-protein interaction of two Topo I molecules. None of these characteristics were found in our yeast experiments.

An alternative model for Topo I-mediated nonhomologous recombination was proposed by Champoux and Bullock (6), in which Topo I recognizes and generates a nick on the 3′ end of a (G/C)TT sequence (Fig. 3). The enzyme covalently binds to the 3′ T residue and can then attach this 3′ end to another DNA molecule to promote recombination in the absence of any protein-protein interaction. On the basis of this model (Fig. 3), the preferred sequences need only appear in one of the two junctions for Topo I to promote recombination. In fact, in the majority of the recombination junctions from mammalian cells analyzed for IR, only one of the two junctions showed a Topo I recognition site (6), which is also true for the

events in our study. We propose that our events occurred by a mechanism similar to the one proposed by Champoux and Bullock (6) and thus provide a good model for IR in mammalian cells.

Analysis of the integrants in cells overexpressing *TOP1* also showed several other interesting features. In two cases, transforming DNA was flanked by two different sequences, either far apart within one gene or between two genes (Fig. 2A). This may have occurred via chromosomal deletion or duplication. Also interesting is a group of four integrants that contain (G/C)TT sequences on both junctions of the integration event. In one case (Top-G08), this led to a rearrangement at the locus of the gene coding for rRNA. Thus, Topo I may mediate the integration of DNA as well as translocations and deletions between chromosomal DNA sequences.

ACKNOWLEDGMENTS

We thank Rolf Sternglanz, Dan Gietz, Nicki Levin (Gerry Fink's laboratory), and Arturo Saavedra for plasmids and strains. Tom Petes, in whose laboratory this project was started, is thanked for his generous support and encouragement. Margaret Dominska of Tom Petes's laboratory is thanked for expert technical help in the sequencing of the sites flanking the IR target sites from the *top1*Δ strain. Tom Petes, Stuwart Shuman, and Wendy Yap are thanked for helpful suggestions during the work. We also thank members of the Schiestl laboratory, Stephanie Kong, and Tom Petes for comments on the manuscript.

This research was supported by grant CN-83 from the American Cancer Society to R.H.S.

REFERENCES

1. **Amin, A. A., L. G. Beatty, and P. D. Sadowski.** 1990. Synaptic intermediates promoted by the FLP recombinase. J. Mol. Biol. **214:**55–72.
2. **Andersen, H. A., E. Gocke, B. J. Bonven, O. F. Nielsen, and O. Westergaard.** 1985. Topoisomerase I has a strong binding preference for a conserved hexadecameric sequence in the promoter region of the rRNA gene from Tetrahymena pyriformis. Nucleic Acids Res. **13:**1543–1556.
3. **Been, M. D., R. R. Burgess, and J. J. Champoux.** 1984. Nucleotide sequence preference at rat liver and wheat germ type I topoisomerase breakage sites in duplex SV40 DNA. Nucleic Acids Res. **12:**3097–3114.
4. **Bjornsti, M. A., A. M. Knab, and P. Benedetti.** 1994. Yeast *Saccharomyces cerevisiae* as a model to study the cytotoxic activity of the antitumor drug camptothecin. Cancer Chemother. Pharmacol. **34**(Suppl)**:**S1–S5.
5. **Bullock, P., J. J. Champoux, and M. Botchan.** 1985. Association of crossover points with topoisomerase I cleavage sites: a model for nonhomologous recombination. Science **230:**954–958.
6. **Champoux, J. J., and P. A. Bullock.** 1988. Possible role for the eukaryotic type I topoisomerase in illegitimate recombination, p. 655–666. *In* R. Kucherlapati and G. R. Smith (ed.), Genetic recombination. American Society for Microbiology, Washington, D.C.
7. **Choder, M.** 1991. A general topoisomerase I-dependent transcriptional repression in the stationary phase in yeast. Genes Dev. **5:**2315–2326.
8. **Denis, C. L., and E. T. Young.** 1983. Isolation and characterization of the positive regulatory gene *ADR1* from *Saccharomyces cerevisiae*. Mol. Cell. Biol. **3:**360–370.
9. **DiNardo, S., K. Voelkel, and R. Sternglanz.** 1984. DNA topoisomerase II mutant of *Saccharomyces cerevisiae*: topoisomerase II is required for segregation of daughter molecules at the termination of DNA replication. Proc. Natl. Acad. Sci. USA **81:**2616–2620.
10. **Eng, W.-K., L. Faucette, R. K. Johnson, and R. Sternglanz.** 1988. Evidence that DNA topoisomerase I is necessary for the cytotoxic effects of camptothecin. Mol. Pharmacol. **34:**755–760.
11. **Gangloff, S., M. R. Lieber, and R. Rothstein.** 1994. Transcription, topoisomerase and recombination. Experientia **50:**261–269.
12. **Gietz, R. D., A. St. Jean, R. A. Woods, and R. H. Schiestl.** 1992. Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Res. **20:**1425.
13. **Gietz, R. D., and A. Sugino.** 1988. New yeast-Escherichia coli shuttle vectors constructed with in vitro-mutagenized yeast genes lacking six-base-pair restriction sites. Gene **74:**527–534.
14. **Hatfull, G. F., and N. D. F. Grindley.** 1988. Resolvases and DNA-invertases: a family of enzymes active in site-specific recombination, p. 357–396. *In* R. Kucherlapati and G. R. Smith (ed.), Genetic recombination. American Society for Microbiology, Washington, D.C.
15. **Hughes, R. E., G. F. Hatful, P. Rice, T. A. Steitz, and N. D. F. Grindley.** 1990. Cooperativity mutants of the gamma delta resolvase identify an essential interdimer interaction. Cell **63:**1331–1338.
16. **Kim, R. A., and J. C. Wang.** 1992. Identification of the yeast *TOP3* gene product as a single strand-specific DNA topoisomerase. J. Biol. Chem. **267:** 17178–17185.
17. **Kraft, R., J. Tardiff, K. S. Krauter, and L. A. Leinwand.** 1988. Using miniprep plasmid DNA for sequencing double stranded templates with Sequenase. BioTechniques **6:**544–545.
18. **Landy, A.** 1989. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. Annu. Rev. Biochem. **58:**913–949.
19. **Levin, N. A., M. Bjornsti, and G. R. Fink.** 1993. A novel mutation in DNA topoisomerase I of yeast causes DNA damage and RAD9-dependent cell cycle arrest. Genetics **133:**799–814.
20. **Porter, S. E., and J. J. Champoux.** 1989. Mapping in vivo topoisomerase I sites on simian virus 40 DNA: asymmetric distribution of sites on replicating molecules. Mol. Cell. Biol. **9:**541–550.
21. **Roth, D., and J. Wilson.** 1988. Illegitimate recombination in mammalian cells, p. 621–653. *In* R. Kucherlapati and G. R. Smith (ed.), Genetic recombination. American Society for Microbiology, Washington, D.C.
22. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **Schiestl, R. H., M. Dominska, and T. D. Petes.** 1993. Transformation of *Saccharomyces cerevisiae* with nonhomologous DNA: illegitimate integration of transforming DNA into yeast chromosomes and in vivo ligation of transforming DNA to mitochondrial DNA sequences. Mol. Cell. Biol. **13:**2697–2705.
24. **Schiestl, R. H., and R. D. Gietz.** 1989. High efficiency transformation of intact yeast cells by single stranded nucleic acids as carrier. Curr. Genet. **16:**339–346.
25. **Schiestl, R. H., and T. D. Petes.** 1991. Integration of DNA fragments by illegitimate recombination in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **88:**7585–7589.
26. **Schiestl, R. H., J. Zhu, and T. D. Petes.** 1994. Effect of mutations in genes affecting homologous recombination on restriction enzyme-mediated and illegitimate recombination in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **14:** 4493–4500.
27. **Sherman, F., G. R. Fink, and J. B. Hicks.** 1986. Methods in yeast genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
28. **Shuman, S.** 1989. Vaccinia DNA topoisomerase I promotes illegitimate recombination in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **86:**3489–3493.
29. **Shuman, S.** 1991. Recombination mediated by vaccinia virus DNA topoisomerase I in *Escherichia coli* is sequence specific. Proc. Natl. Acad. Sci. USA **88:**10104–10108.
30. **Tanizawa, A., K. W. Kohn, and Y. Pommier.** 1993. Induction of cleavage in topoisomerase I c-DNA by topoisomerase I enzyme from calf thymus and wheat germ in the presence and absence of camptothecin. Nucleic Acids Res. **21:**5157–5166.
31. **Thrash, C., A. T. Bankier, B. G. Barrell, and R. Sternglanz.** 1985. Cloning, characterization, and sequence of the yeast DNA topoisomerase I gene. Proc. Natl. Acad. Sci. USA **82:**4374–4378.
32. **Wallis, J. W., G. Chrebet, G. Brodsky, M. Rolfe, and R. Rothstein.** 1989. A hyper-recombination mutation in *S. cerevisiae* identifies a novel eukaryotic topoisomerase. Cell **58:**409–419.
33. **Wang, H. P., and C. E. Rogler.** 1991. Topoisomerase I-mediated integration of hepadnavirus DNA in vivo. J. Virol. **65:**2381–2392.
34. **Wang, J. C.** 1985. DNA topoisomerases. Annu. Rev. Biochem. **54:**665–697.