2366–2373 *Nucleic Acids Research, 1998, Vol. 26, No. 10*

Sequence specificity of illegitimate plasmid recombination in *Bacillus subtilis*: possible recognition sites for DNA topoisomerase I

Rob Meima, Gert-Jan Haan, Gerard Venema, Sierd Bron* and Steven de Jong

Department of Genetics, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands

Received January 28, 1998; Revised and Accepted March 25, 1998

ABSTRACT

Previous work in our group indicated that structural plasmid instability in *Bacillus subtilis* is often caused by illegitimate recombination between non-repeated sequences, characterized by a relatively high AT content. Recently we developed a positive selection vector for analysis of plasmid recombination events in *B.subtilis* which enables measurement of recombination frequencies without interference of selective growth differences of cells carrying wild-type or deleted plasmids. Here we have used this system to further analyse the sequence specificity of illegitimate plasmid recombination events and to assess the role of the host-encoded DNA topoisomerase I enzyme in this process. Several lines of evidence suggest that single-strand DNA nicks introduced by DNA topoisomerase I are a major source of plasmid deletions in pGP100. First, strains overproducing DNA topoisomerase I showed increased levels of plasmid deletion. Second, these deletions occurred predominantly (>90% of the recombinants) between non-repeated DNA sequences, the majority of which resemble potential DNA topoisomerase I target sites. Sequence alignment of 66 deletion end-points confirmed the previously reported high AT content and, most importantly, revealed a highly conserved C residue at position –4 relative to the site of cleavage at both deletion termini. Based on these genetic data we propose the following putative consensus cleavage site for DNA topoisomerase I of *B.subtilis*: 5′-$^{A}/_{T}$CAT$^{A}/_{T}$TA$^{A}/_{T}$$^{A}/_{T}$A-3′.

INTRODUCTION

For several reasons the Gram-positive soil bacterium *Bacillus subtilis* is of great fundamental and industrial importance. Because of its genetic amenability, well-studied biochemistry and physiology, high protein secretion capacity and non-pathogenic nature, *B.subtilis* is widely used for a variety of biotechnological applications. Especially since the entire genome sequence was established (1), this bacterium has become the paradigm of *Bacillus* spp. and, in fact, Gram-positive bacteria in general. However, a major problem in the development of suitable cloning and expression vectors for bacilli to exploit their full potential is the frequently observed genetic instability, in particular of foreign DNA.

Illegitimate recombination, which is a major source of genome rearrangements in nature (for reviews see 2,3), was shown to underlie structural plasmid instability in numerous cases. These *recA*-independent events can occur in several different ways. Recombination between short directly repeated sequences (DRs) of 3–20 bp can result from either copy choice recombination or DNA breakage and reunion mechanisms (for reviews see 3,4). Breakage and reunion events also underlie recombination between sequences that share no homology; these processes are thought to involve certain DNA handling enzymes, such as topoisomerases, origin nicking enzymes, transposases, invertases and site-specific recombinases.

DNA topoisomerases are pivotal enzymes in the control of topological transitions of DNA. A role of these enzymes in illegitimate recombination has been demonstrated by several authors (5–11). These events involve uncoupling of the nicking and closing reactions catalysed by DNA topoisomerases and result in joining of unrelated ends, producing recombinant molecules. In addition, specialized type I topoisomerases, like phage λ integrase (12) and P1 Cre recombinase (13), mediate DNA strand transfer in site-specific recombination reactions. Recently Zhu and Schiestl (14) demonstrated that overproduction of topoisomerase I (Topo I) in *Saccharomyces cerevisiae* increased the level of illegitimate integration in transformation studies with non-homologous DNA.

Based on our previous work we suggested that DNA Topo I also underlies plasmid instability in *B.subtilis* (15,16). Based on our results, we suggested that Topo I-dependent single-strand (ss)DNA nicking may be the primary source of type I deletions, which typically occur between non-repeated DNA sequences. A similar role for Topo I in recombination between non-homologous sequences was proposed by Bierne *et al.* (17), who studied deletion as a result of replication arrest in *Escherichia coli*. In addition to type I recombination events, deletions between short DRs have also been reported, in particular in certain strains carrying mutations in genes encoding enzymes involved in homologous recombination and DNA repair (16–19). We conceived that these deletions, which we have denoted type II, resulted from error-prone double-strand break repair involving exonucleolytic processing of linear double-stranded (ds)DNA molecules prior to closure (19,20). Similarly to deletions between short DRs upon replication blockage (17), Topo I is probably not involved in this class of deletions in *B.subtilis*.

*To whom correspondence should be addressed. Tel: +31 50 3632105; Fax: +31 50 3632348; Email: s.bron@biol.rug.nl

Table 1. Bacterial strains and plasmids

Strain or plasmid	Relevant genotype or plasmid marker[a]	Source or reference
Bacillus subtilis		
8G5	*trpC2 tyr-1 his ade met rib ura nic*	24
8G5::pIN5	8G5 carrying the *Streptococcus agalactiae* Tc^r marker in the *amyE* locus	Laboratory collection
8G5::pTOP	8G5 carrying an additional copy of *topA* gene, Em^r	de Jong *et al*., unpublished
8G5::pTOPC1	8G5 carrying multiple copies of the *topA* gene, Em^r, Cli^r	de Jong *et al*., unpublished
Escherichia coli		
MC1061	F^- *araD139* Δ(*ara-leu*) *7696 galE15 galK16* Δ(*lac*)*X74 rpsL* (Str^r) *hsdR2* (r_K^- m_K^+) *mcrA mcrB1*	42
Plasmids		
pBR322	Ap^r, Tc^r, 4.3 kb	43
pWSK29	pSC101 derivative containing extended multiple cloning site; 5.4 kb, Ap^r	44
pTOP	pWSK29 containing a 2.9 kb *Xba*I–*Stu*I fragment carrying the *topA* gene and the upstream *smf* gene; 8.3 kb	de Jong *et al*., unpublished
pTOPEm	pTOP carrying the *B.subtilis* topA gene and the pE194 Em^r marker from pSPT1988E; 9.6 kb	de Jong *et al*., unpublished
pGP100	pGKV2 derivative carrying *penIP* fragment from pGP1 and promoterless *cat-86* gene; Tc^r, Km^r, 9.1 kb	16
pSPT1988E	pSPT1988 carrying the Em^r marker of pE194	32
pIN588	pWSK29 derivative carrying the *addAB* genes, Ap^r Tc^r, 17.5 kb	45

[a]Tc, tetracycline; Em, erythromycin; Cli, clindamycin; Str, streptomycin; Ap, ampicillin; Km, kanamycin.

The *topA* gene of *B.subtilis*, specifying Topo I, was recently cloned and sequenced by our group (S.de Jong *et al.*, submitted for publication). It was shown to be highly homologous to Topo I from *E.coli* and *Thermotoga maritima*. As in *E.coli*, the enzyme contains three C-terminally located putative Zn binding domains; successive deletion of these domains indicated that at least one of them is required for complementation of a temperature-sensitive *topA* mutant of *E.coli*.

In the present study we assessed the sequence specificity of type I deletion events, as well as the possible role of Topo I therein, by analysis of plasmid deletion derivatives produced in a TopoI overproducing host. For this purpose we made use of a recently developed system for assaying illegitimate plasmid recombination, pGP100 (16). Comparison of the deletion end-points obtained here and those analysed in our previous studies (15,16,19) revealed a strongly conserved C nucleotide at position –4 relative to both deletion termini. The presence of a C residue at this position is characteristic of TopoI-mediated cleavage observed in other organisms (21). Furthermore, plasmid deletion frequencies were increased in TopoI overproducing strains. Together these data indicate that Topo I-dependent nicking of supercoiled plasmid DNA is a likely source of type I plasmid rearrangements.

MATERIALS AND METHODS

Bacterial strains and plasmids

Bacterial strains and plasmids used throughout this study are listed in Table 1.

Chemicals and enzymes

Chemicals used were of analytical grade and were obtained from Merck (Darmstadt, Germany). Restriction endonucleases were used as indicated by the manufacturer (Boehringer, Mannheim, Germany); T4 DNA ligase and T4 DNA polymerase (Boehringer) were used according to Sambrook *et al.* (22). Ampicillin, clindamycin and chloramphenicol were purchased from Sigma (St Louis, MO); erythromycin, kanamycin and tetracycline were from Boehringer.

Media and growth conditions

Bacillus subtilis minimal medium consisted of Spizizen's minimal salts (23) supplemented with glucose (0.5%) and casein hydrolysate (0.02%; Difco, Detroit, MI). Amino acids and nucleotides, if required, were added at 20 μg/ml each; vitamins were added at 0.4 μg/ml. Minimal agar consisted of minimal salts supplemented with 0.5% glucose, 0.02% casein hydrolysate, required growth factors and 1.5% agar. TY broth contained (per litre) 10 g trypton (Difco), 5 g yeast extract (Difco), 10 g NaCl, pH 7.4. Ampicillin (Ap) was used at 100 μg/ml for *E.coli*. Unless indicated otherwise, chloramphenicol (Cm) was used at 5 μg/ml for *B.subtilis* and *E.coli*. Erythromycin (Em) was used at a final concentration of 150 μg/ml for *E.coli* (or 50 μg/ml when used in combination with Cm) and 2 μg/ml for *B.subtilis*. Tetracycline (Tc) and kanamycin (Km) were added at 12.5 and 50 μg/ml respectively for both *E.coli* and *B.subtilis*. Optical densities of cultures were measured at 600 nm using a Vitatron microprocessor controlled Photometer (Vital Scientific, Dieren, The Netherlands). Growth in microtitre dishes (Corning Glass Works, New York, NY) was performed in 100 μl TY broth at 37°C without shaking or additional aeration; growth rates (μ_{max}) were determined from viable cell counts (c.f.u./ml), measured by plating culture samples collected at several time points during growth.

DNA manipulations

Chromosomal DNA from *B.subtilis* was isolated according to Bron and Venema (24). Large scale or mini preparations of plasmid DNA from both *B.subtilis* and *E.coli* were obtained by the alkaline lysis method (22). pGP100 DNA for *in vitro* relaxation assays was purified from CsCl gradients as described previously (22). Cloning procedures were performed as described by Sambrook *et al.* (22). DNA restriction fragments were recovered from agarose gels using a Qiaex purification kit (Qiagen, Hilden, Germany). Southern blot analyses were performed using the non-radioactive ECL direct nucleic acid labelling and

detection system (Amersham International, Amersham, UK). DNA sequencing was according to Sanger *et al*. (25) using Pharmacia reagents (Uppsala, Sweden).

Topo I activity in cell-free extracts

Topo I-dependent relaxation of supercoiled DNA was quantified as follows. Cells of *B.subtilis* strains producing different levels of Topo I were grown in TY broth containing the appropriate antibiotics. At the onset of the stationary growth phase cells were harvested by centrifugation and washed once with buffer 1 [50 mM Tris–HCl, pH 7.5, 1 mM EDTA, 1 mM dithiothreitol (DTT), 1 mM phenylmethylsulfonyl fluoride (PMSF)]. Cell-free extracts were prepared by rupturing the cells in a French press as described by Kooistra *et al*. (26), using a J5-598A Laboratory Pressure Cell Press (Aminco, Silver Spring, MA). Subsequently the lysates were centrifuged at 4°C (13 000 *g*) for 15 min to remove intact cells and debris. Protein concentrations were determined as described by Bradford (27), with bovine serum albumin (BSA) (Sigma) as the standard. Approximately 1 µg supercoiled pBR322 DNA was incubated in a total volume of 30 µl for 15 min at 37°C with varying amounts of lysate in buffer 1 supplemented with 150 mM KCl, 5 mM $MgCl_2$ and 100 µg/ml BSA. DNA gyrase activity in the lysates was inhibited by addition of nalidixic acid at a final concentration of 1 mM. The relaxation reactions were stopped by addition of 0.2% SDS and the samples were loaded onto 0.8% agarose gels containing 0.5 µg/ml ethidium bromide (Merck). Electrophoresis (2 V/cm, 10 mA) was performed at room temperature for ~16 h, to ensure optimal separation of supercoiled and relaxed forms of the plasmid.

Analysis of plasmid supercoiling

pGP100 plasmid DNA to be analysed by gel electrophoresis was isolated from late exponential cultures using the alkaline lysis method (see above). Approximately 1 µg plasmid DNA was electrophoresed in 0.8% agarose slab gels in TAE buffer (22) at 2.5 V/cm for 16–20 h at room temperature. The gels contained 6 µg/ml chloroquine to resolve the topoisomers present in the mixture of supercoiled plasmid DNA molecules. After electrophoresis the gels were washed with 1 mM $MgSO_4$ to remove the chloroquine (28) and the DNA was subsequently transferred to Qiabrane Nylon Plus membranes (Qiagen) for Southern hybridization as described above.

Determination of plasmid recombination frequencies

The frequency of recombination events resulting in formation of pGP100 derivatives expressing chloramphenicol resistance was measured by the following methods. (i) *Direct plating assay.* Plating of cultures directly on Cm-containing agar medium was performed as described previously (16); recombination frequencies were determined by measuring the percentage of Cm^r clones as a function of the total number of viable cells. (ii) *Fluctuation assay.* Deletion frequencies were measured as described previously (16,29) using the principles of the Luria–Delbrück fluctuation test (30). A large number of identical but independent cultures were inoculated with a limited number of cells and grown in 96-well microtitre plates. Based on the estimates of recombination frequencies from the direct plating assay, cells were diluted to a final titre that was low enough to ensure that the cultures did not contain recombinants at the onset of the experiment (i.e. 1000× lower than the expected recombination frequencies, as determined by the direct plating assay). The cultures were then grown until each, on average, contained one recombinant cell. Selection was then applied by the addition of Cm, which blocks growth of all the cultures that did not contain recombinants but allows continued growth of cells containing a recombinant plasmid(s). Frequencies of plasmid recombination were expressed per cell generation. Growth in microtitre dishes of the various strains used in our assays was analysed prior to the fluctuation experiments. This was achieved by plating samples collected at several stages during growth. Based on the growth curves obtained and given the inocculum size, the cultivation time required to obtain a situation in which ~50% of the cultures contained a recombinant plasmid (P_0 = 0.5) could be estimated. These conditions were tested for each of the strains analysed and used throughout the experiments, to ensure accurate measurements of recombination frequencies.

RESULTS

Topo I overproduction does not affect pGP100 copy number

To investigate the possible role of Topo I in the initial stages of deletion plasmid pGP100 was introduced into *B.subtilis* strains expressing increased levels of Topo I activity. These derivatives were constructed by integration and subsequent amplification of the *topA* gene in the chromosome of our laboratory strain, 8G5. For this purpose we made use of an integrational *E.coli* plasmid vector, pTOPEm, which carries a wild-type copy of the *topA* gene and a selectable marker for integration and amplification in *B.subtilis* (de Jong *et al*., unpublished results). Multiple copy amplification of the *topA* gene was achieved by plating *B.subtilis* cells carrying this plasmid on medium containing increasing concentrations of clindamycin (Cli), as described previously (16,31). Colonies grown at various concentrations of Cli were characterized as follows. First, Topo I-dependent relaxation of supercoiled plasmid DNA was measured *in vitro* in cell-free lysates (Fig. 1). As can be seen in lane 7, ~50% of the supercoiled pBR322 plasmid DNA was relaxed upon incubation with a 16-fold diluted lysate of the strain carrying multiple copies of the *topA* gene, whereas no significant relaxation was observed under similar conditions in lysates of wild-type cells. The fact that the molecules with higher mobility are indeed relaxed circles was demonstrated by the observations that: (i) the mobility shift could be reversed by addition of ATP in a reaction typically catalysed by DNA gyrase and (ii) when electrophoresed in the absence of ethidium bromide the newly formed product had a much lower mobility, as expected for relaxed circular molecules (results not shown). The results of these assays indicated that the Topo I activity was increased ~8-fold in strains in which amplification of the *topA* gene had occurred. Furthermore, the Topo I activity in the overproducing strains was comparable in clones isolated at either 1 or 20 µg/ml Cli, suggesting a limit to the level of overexpression. One clone obtained at 1 µg/ml Cli, designated 8G5::pTOPC1, was selected for further analysis. Southern hybridization showed that in this strain the *topA* gene was indeed present in multiple copies at the expected site in the chromosome; densitometric scanning of the resulting blots revealed that ~10 additional copies of the *topA* gene were present in 8G5::pTOPC1 (not shown).

To address the question whether possible differences in plasmid copy number could be responsible for differences in recombination frequencies (see below) the relative copy numbers of pGP100 in the parental strain and the strains carrying one or more additional

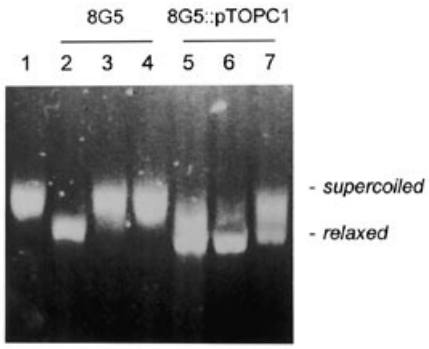

Figure 1. *In vitro* relaxation of supercoiled pBR322 DNA (lane 1) in cell-free lysates of strains 8G5 and 8G5::pTOPC1. pBR322 DNA was incubated either with undiluted lysate (± 2 µg total protein; lanes 2 and 5), 4-fold (lanes 3 and 6) or 16-fold diluted lysate (lanes 4 and 7).

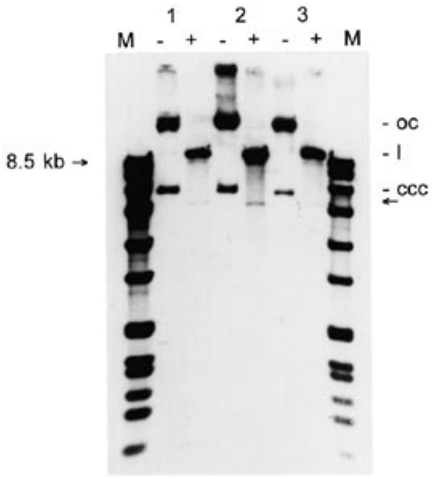

Figure 2. Southern blot analysis of total DNA isolated from strains carrying plasmid pGP100. *Bst*YI-digested pGP1 and *Asp*700-digested pIN588 were used as probes. The latter was used as an internal standard (*addAB*) of chromosomal DNA to correct for differences in the total amount of DNA applied to the gel. M, molecular weight marker (*Eco*RI-digested phage SPP1 DNA); lane 1, 8G5; lane 2, 8G5::pTOP; lane 3, 8G5::pTOPC1. –, untreated; +, digested with *Pst*I. *oc*, open circular; *l*, linear; *ccc*, covalently closed circular; the arrow on the right indicates the position of a 4.9 kb chromosomal DNA fragment hybridizing with the *addAB*-specific probe (32).

copies of the *topA* gene were analysed (Fig. 2). Plasmid copy numbers in the various strains were analysed by Southern hybridization of total chromosomal and plasmid DNA extracted from exponentially growing cells of these strains and comparison of the hybridization signals. As an internal standard for the amount of chromosomal DNA a probe directed against the *addAB* gene (32) was used. Densitometric scanning of the hybridization signal of *Pst*I-linearized pGP100 and that from the chromosomal standard, which runs just below the linear pGP100 fragment, showed that no significant differences in copy number (expressed per chromosome equivalent) of pGP100 existed between the strains analysed. In all three strains tested the ratio of plasmid and chromosomal DNA was between 1:75 and 1:100.

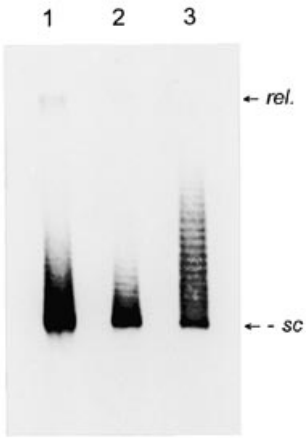

Figure 3. Southern blot analysis of topoisomer distribution of pGP100 *in vivo*. Lane 1, 8G5 (pGP100); lane 2, 8G5::pTOP (pGP100); lane 3, 8G5::pTOPC1 (pGP100). rel., relaxed; – sc, negatively supercoiled pGP100 DNA.

pGP100 linking numbers *in vivo*

To test whether Topo I overproduction changed the overall supercoiling of pGP100 *in vivo* plasmid DNA was isolated from late-exponential cultures and analysed for linking number distribution. For this purpose the plasmid DNA was separated by electrophoresis in chloroquine-containing agarose gels, as indicated in Materials and Methods. The distribution of topoisomers of pGP100 in the various hosts is shown in Figure 3. When compared with the wild-type, the average linking number was shifted from the position of highly negatively supercoiled DNA towards that of more relaxed DNA in the Topo I overproducing background. This showed that *in vivo* the overall supercoiling of pGP100 had markedly changed as a result of Topo I overexpression. In the strain carrying only one additional copy of the *topA* gene no such shift was observed.

pGP100 deletion is increased by Topo I overproduction

To analyse whether Topo I could be involved in the illegitimate recombination events studied here plasmid pGP100 was introduced into strains producing different levels of Topo I and the structural stability of pGP100 was studied. The result of a typical experiment is shown in Table 2. The average recombination frequency, determined from at least four independent assays, was $\sim 5 \times 10^{-8}$ per generation in the control strain (8G5). When one additional copy of the *topA* gene was present, in strain 8G5::pTOP, the instability of pGP100 was not significantly different from the wild-type. However, when multiple copies of the *topA* gene were present (8G5::pTOPC1) the recombination frequency was clearly higher (2- to 4-fold, determined from three individual experiments) than in the parental strain. Since the copy numbers of pGP100 in the strains used were similar (Fig. 2; see above), the differences in plasmid stability reflect differences in molecular recombination frequencies. Strikingly, these frequencies were even considerably higher (~10-fold) in both the parental and the Topo I overproducing strain when tetracycline was included in the medium during the experiments (not shown; see Discussion).

Table 2. pGP100 recombination frequencies in strains producing different levels of Topo I

Strain	Frequency[a]	Relative frequency[b]
8G5	5.4×10^{-8}	1
8G5::pTOP	5.8×10^{-8}	1.07
8G5::pTOPC1[c]	1.2×10^{-7}	2.22

[a]Recombination frequencies are expressed per cell generation and were determined as described in Materials and Methods. Chloramphenicol was added at 15 μg/ml.
[b]Recombination frequencies as compared with the wild-type strain, 8G5.
[c]To avoid loss of copies of the amplified *topA* gene in this strain clindamycin was added to a final concentration of 1 μg/ml in these experiments.

We next tried to analyse pGP100 deletion in *B.subtilis* strains with reduced levels of Topo I activity, since, conceivably, this might improve pGP100 stability. Reduced Topo I activity could, for instance, be achieved by using varying dosages of specific inhibitors of this enzyme. However, no such inhibitors are available at present for prokaryotic type I DNA topoisomerases. So, in order to address this question two mutagenesis-based approaches were followed. First, we attempted to construct *B.subtilis* strains producing lower levels of Topo I by placing the *topA* gene under control of an inducible promoter. Second, we tried to construct strains carrying *topA* gene copies encoding 3′-terminally truncated Topo I enzymes lacking one or several of the Zn(II)-finger domains. The latter mutants were shown previously to complement a temperature-sensitive *topA* mutant of *E.coli* (S.de Jong *et al.*, submitted for publication). However, using the same clones in *B.subtilis* we were unable to construct strains expressing the C-terminally truncated Topo I proteins. Neither were we successful in obtaining strains expressing reduced levels of the enzyme using an inducible promoter upstream of the *topA* gene. Apparently, down-regulation of Topo I-dependent relaxation activity is not tolerated by *B.subtilis* (S.de Jong *et al.*, submitted for publication). As a consequence, we could not determine the effect of reduced cellular levels of Topo I on pGP100 stability.

Sequences involved in deletion in a Topo I overproducing host

Increased Topo I-mediated relaxation of negatively supercoiled DNA evidently causes a change in the level of DNA supercoiling. Since several processes, such as DNA replication and transcription, depend on accurate control of DNA supercoiling (33), it is essential to maintain an equilibrium in the distribution of positively and negatively supercoiled molecules. Upon Topo I overproduction this disturbed equilibrium may be restored by a simultaneous increase in the activity of other topoisomerases, the expression of which is regulated by DNA supercoiling levels. For instance, synthesis of *E.coli* DNA gyrase is highest when the DNA template is relaxed (34), whereas expression of the *topA* gene requires a supercoiled DNA template (35). Furthermore, *E.coli topA* mutants are only viable when compensatory mutations are accumulated elsewhere, e.g. in the DNA gyrase (36) or DNA topoisomerase IV gene (37). Together these observations suggest a homeostatic control mechanism of DNA supercoiling (34). Elevated levels of illegitimate recombination as a consequence of Topo I overproduction are thus not necessarily the result of increased Topo I-related nicking activity, but may also reflect an increased nicking and closing activity by, for instance, DNA gyrase. Since DNA gyrase introduces dsDNA breaks rather than the ssDNA breaks supposed to underlie type I deletion, elevated gyrase levels might alter the distribution between type I and type II deletion events or, alternatively, produce an entirely different class of deletions. If, however, the increased activity of Topo I is directly responsible for the increase in recombination frequency, a significant increase in the relative frequency of type I recombinants is expected. For statistical reasons testing this prediction would require the sequencing of a very large number of deletion fusion points, since in the parental strain 90% of deletions are of the latter class. Nevertheless, we have analysed a number of deletion fusion points isolated from the overproducing strain to gain an impression of the distribution of type I and type II recombinants in this background (Fig. 4). As expected, these deletion derivatives all lacked the T_1T_2 terminator region of pGP100, allowing expression of the *cat-86* gene. In 12 of the 13 recombinants analysed here deletion had occurred between non-repeated sequences. Moreover, we observed a strong preference (9 out of 12) for deletion within the promoter region of the Tc^r gene (16), a feature which is very similar to what was observed for a different promoter in previous studies (15,18,19).

To further analyse the nature of the target sites for type I deletion we aligned 66 independently obtained deletions of this class (Fig. 5A). These were isolated using plasmids pGP1 (15,18,19) and pGP100 (16,19; this study). This alignment showed a weak consensus target site for type I deletion: (–5) 5′-$T^{A}/_{T}{}^{A}/_{T}TA^{A}/_{T}AT\mathbf{G}^{A}/_{T}$-3′ (+5). The assignment of these nucleotides was based on their occurrence in at least 35% of the deletions. As mentioned above, the (A+T)-rich nature of this target sequence strongly suggests that it is, in fact, a Topo I cleavage site. Assuming that Topo I-dependent cleavage is indeed the first step in type I deletion, cleavage may occur in either strand of the supercoiled plasmid molecule. If cleavage occurs in the strand opposite to that shown in Figures 4 and 5A, this would place a C residue at position –4, as can be seen in Figure 5B: (–5) 5′-$^{A}/_{T}\mathbf{C}AT^{A}/_{T}\mathbf{T}A^{A}/_{T}{}^{A}/_{T}$A-3′ (+5).

Together the data described here suggest that the increased recombination frequencies observed in the Topo I overproducing host are directly related to increased Topo I activity. Apparently ssDNA nicks introduced by Topo I are a major source of type I deletion.

DISCUSSION

In previous work (16,19) we described analyses of plasmid deletion in *B.subtilis*. Deletions were found to occur between either non-repeated sequences (designated type I) or between short DRs (type II). Based on these data and those previously obtained by Peijnenburg *et al.* (15,18) we proposed a role for Topo I in type I deletion. Since the *topA* gene of *B.subtilis* was recently cloned and characterized (S.de Jong *et al.*, submitted for publication) we were now able to further analyse the role of Topo I in plasmid instability. Moreover, the present studies greatly benefited from the availability of a positive selection vector for determination of plasmid recombination frequencies (16), which enabled us to exploit a microtitre-based fluctuation assay, eliminating the possible effects of selective growth differences between cells containing parental plasmids and cells carrying recombinant plasmids (16,29). Using the improved assay we analysed recombination frequencies in a *B.subtilis* strain overexpressing Topo I. In this background, deletion frequencies were ~2- to 3-fold higher than in the parental strain and the strain carrying one additional copy of the *topA* gene. In addition, in the Topo I overproducing strain the distribution of pGP100 topoisomers was shifted from highly negatively supercoiled

Endpoint		*Sequence*		*Features**
pGP100-TΔ3		CCTATTTGGAATTATAATAATAGGTGGTCCCTTTATACCG		*Type I*
Δ=4483 bp	2560	-CCTATTTGG**AATTATAATAAATT**GTTTCGGGTCGGTAATT-	2599	[A/T]$_7$TA▾A▾ATT
	7043	-CCTTTTCCCGTTCCTCATCATAGGTGGTCCCTTTATACCG-	7082	
pGP100-TΔ4		GGTTCAATTAAAAGAGGGAATTCTAGAAGCTTCGACTCTA		*Type II*
Δ=2006 bp	2252	-GGTTCAATTAAAAGAGGGAAGCGTATCATTAACCCTATAA-	2281	10 bp DR
	4258	-TCATCCGCCAAAACAGGGAATTCTAGAAGCTTCGACTCTA-	4297	
pGP100-TΔ5		ATTTGAAACTTAGTTTATATACTGTATGCTCGTTAATTCT		*Type I*
Δ=3902 bp	2160	-ATTTGAAACTTAG**TTTATAT**GTGGTAAAATGTTTTAATCA-	2199	TTTATA▾T▾
	6062	-CGTAAGACATATAAAATTTTACTGTATGCTCGTTAATTCT-	6101	
pGP100-TΔ6		CCTATTTGGAATTATAATAAGCGGATCTTTAAATGGAGTG		*Type I*
Δ=4142 bp	2560	-CCTATTTGG**AATTATAATAAATT**GTTTCGGGTCGGTAATT-	2599	[A/T]$_7$TAA▾ATT
	6702	-TAAAAAATCATACAGCTCGCGCGGATCTTTAAATGGAGTG-	6741	
pGP100-TΔ8		AGAATTTGAAACTTAGTTTAAGGGAATTCTAGAAGCTTCG		*Type I*
Δ=2095 bp	2157	-AGAATTTGAAACTTAG**TTTATAT**GTGGTAAAATGTTTTAA-	2196	TTTA▾TAT
	4252	-CTTCTCTCATCCGCCAAAACAGGGAATTCTAGAAGCTTCG-	4291	
pGP100-TΔ11		TAAAATGTTTTAATCAAGTTGCTGTCTCCCAGGTCGCCGT		*Type I*
Δ=4441 bp	2184	-TAAAATGTTTTAATC**AAGTTTA**GGAGGAATTAATTATGAA-	2223	AA▾G▾T▾T▾TA
	6625	-CCATCTTTCACAAAGATGTTGCTGTCTCCCAGGTCGCCGT-	6664	
pGP100-TΔ1A		TAAAATGTTTTAATCAAGTTGATGTTGCTGTCTCCCAGGT		*Type I*
Δ=4435 bp	2184-	TAAAATGTTTTAATCAAG**TTTA**GGAGGAATTAATTATGAA-	2223	TT▾TA
	6619	-ACTTTGCCATCTTTCACAAAGATGTTGCTGTCTCCCAGGT-	6658	
pGP100-TΔ3A		CAAGTTTAGGAGGAATTAATGCTGTCTCCCAGGTCGCCGT		*Type I*
Δ=4427 bp	2198	-CAAGTTTAGGAGG**AATTAATTAT**GAAGTGTAATTAATGTA-	2237	[A/T]$_5$A▾T▾TAT
	6625	-CCATCTTTCACAAAGATGTTGCTGTCTCCCAGGTCGCCGT-	6664	
pGP100-T7AΔ		TTAGGAGGAATTAATTATGAGCGGATCTTTAAATGGAGTG		*Type I*
Δ=4499 bp	2203-	TTAGGAGG**AATTAATTATGAA**GTGTAATTAATGTAACAGG-	2242	[A/T]$_8$ATGA▾A
	6702-	TAAAAAATCATACAGCTCGCGCGGATCTTTAAATGGAGTG-	6741	
pGP100-TΔ10A		AATGTAACAGGGTTCAATTAGATGTTGCTGTCTCCCAGGT		*Type I*
Δ=4387 bp	2232-	AATGTAACAGGGTTC**AATTAAAA**GAGGGAAGCGTATCATT-	2271	AATTA▾AAA
	6619-	ACTTTGCCATCTTTCACAAAGATGTTGCTGTCTCCCAGGT-	6658	
pGP100-TΔ11A		TAGGAGGAATTAATTATGAATTGTACTTTACTACACTTTA		*Type I*
Δ=3946 bp	2204-	TAGGAGG**AATTAATTATGAA**GTGTAATTAATGTAACAGGG-	2243	[A/T]$_8$ATG▾A▾A▾
	6150-	ACATCTAAGCTCCCGATTAATTGTACTTTACTACACTTTA-	6189	
pGP100-TΔ17A		CAAGTTTAGGAGGAATTAATTTGCCATCTTTCACAAAGAT		*Type I*
Δ=4404 bp	2198	-CAAGTTTAGGAGG**AATTAATTAT**GAAGTGTAATTAATGTA-	2237	[A/T]$_5$A▾T▾T▾AT
	6602	-GATCAATAAAGCCACTTACTTTGCCATCTTTCACAAAGAT-	6641	
pGP100-TΔ18A		TCTCATTACCTGATATTGCATCGGGCTTTTCCGTCTTTAA		*Type I*
Δ=4247 bp	2418	-TCTCATTACCTGATATTGC**AAAT**GATTTTAATAAACCACC-	2457	A▾AAT
	6665	-GGGAAAAGACAAGTTCCTCTTCGGGCTTTTCCGTCTTTAA-	6704	

Figure 4. Nucleotide sequences near fusion points in deletion derivatives of pGP100 isolated from the Topo I-overproducing strain 8G5::pTOPC1. For each deletion the sequence of the fusion is shown in the upper line. The middle and lower lines represent the sequences flanking the left and right deletion termini respectively. Solid triangles indicate possible breakage points within the AT-rich sequences associated with left deletion termini; these sequences are shown in bold in the middle lines. Directly repeated sequences (DRs) associated with the fusion points are underlined. Nucleotide numbers correspond to the linear map of pGP100 (16). *, characteristics of highly clustered deletion end-points in the promoter region of the Tc^r gene (left end-points).

towards more relaxed forms. The plasmid copy numbers were essentially unaffected by Topo I overproduction, indicating that the change in overall DNA supercoiling did not affect the replication efficiency of pGP100.

Sequence analysis of deletion end-points generated in the Topo I overproducer showed that type I deletions were predominant in this host. The distribution of type I plasmid recombinants resembled that observed in the parental strain and no relative increase in type II deletions was observed. Moreover, the majority of end-points flanking one side of the deletion, referred to here as left deletion end-points, were located in the promoter region of the Tc^r gene (nt 2162–2315; see also 16). A similar clustering of deletion end-points was observed previously in the parental strain, 8G5 (16). Since this promoter region is almost equally rich in A and T nucleotides as the coding region of the gene (68.6 versus 64.5% AT), this clustering of end-points seems to indicate that the Tc^r promoter region is a preferred target for deletion, in much the same way as the *penP* promoter in pGP1, which we have studied before (15,18). Our favoured explanation for this observation is that transcription, creating domains of negative

A.

5'	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	3'
G	12	3	9	9	8	12	23	30	**83**	14	
A	21	44	39	23	**61**	43	48	12	11	32	
T	52	38	38	42	27	39	26	50	5	44	
C	15	15	14	26	4	6	3	8	1	10	
	T	A/T	A/T	T	A	A/T	A	T	G	A/T	

B.

5'	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	3'
	A/T	C	A	T	A/T	T	A	A/T	A/T	A	

Figure 5. Sequences associated with type I plasmid deletion in *B.subtilis*. (**A**) Alignment of sequences associated with type I deletion in plasmids pGP1 and pGP100. The incidence of nucleotides around 66 left-hand deletion end-points in the promoter areas of *penP* and Tc^r in pGP1 (15,18,19) and pGP100 (16,19; this study) respectively are compiled in this diagram. (**B**) Putative consensus cleavage site for the *B.subtilis* DNA topoisomerase I enzyme, as deduced from the sequence alignment shown in (A).

hyper-supercoiling (38), rather than nucleotide sequence *per se*, is an important factor in the formation and localization of Topo I-dependent (type I) deletions. Indeed, our recent unpublished results indicated that elevated expression of the Tc^r gene in the presence of tetracycline, as measured by *lacZ* fusions with the Tc^r promoter, resulted in markedly increased deletion frequencies.

About half of the end-points located in the Tc^r promoter region were associated with a predicted stem–loop structure centred around position 2223, suggesting that, in addition to the observed preference for AT-rich sequences, structural features are also involved in determining the target sites for type I deletion. This preference for junctions between ssDNA and dsDNA has also been suggested by Kirkegaard *et al.* (39), who analysed the specificity of *E.coli* Topo I, and, more recently, by Bierne *et al.* (17), who proposed that partly single-stranded molecules (i.e. arrested replication forks) may act as substrates for Topo I-mediated recombination between non-repeated DNA sequences.

In their study on deletion in *E.coli* Bierne *et al.* (17) proposed a Topo I-mediated recombination process, similar to that analysed here. Deletions between non-repeated sequences, normally observed in wild-type cells, were not found in a *topA* mutant. We have also attempted to construct *B.subtilis* strains expressing lower levels of Topo I, but were unable to do so. Nevertheless, the observation that increased Topo I levels stimulate deletion is in good accordance with the decreased recombination between non-repeated sequences in *topA* mutants of *E.coli* reported by Bierne and co-workers (17).

From alignment of the deletion end-point sequences shown in Figure 5A we have deduced a putative consensus target site for *B.subtilis* DNA topoisomerase I (Fig. 5B). This sequence shows a strong preference for a T residue at position +1 (61%), while C

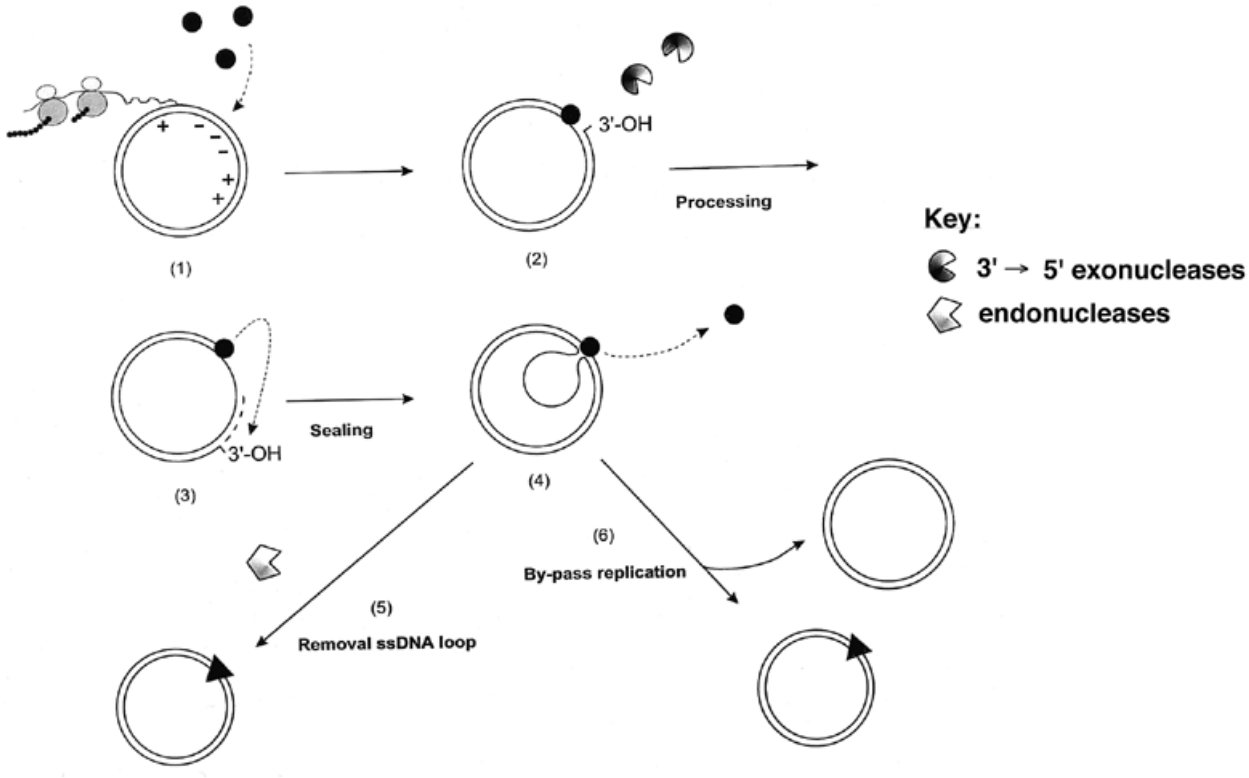

Figure 6. Working model for DNA topoisomerase I-mediated type plasmid deletion in *B.subtilis*. Upon active transcription [1] the DNA template becomes hyper-negatively supercoiled behind the progressing transcription fork. Upon Topo I (●) cleavage and temporal uncoupling of the nicking and closing reaction a free 3′-OH end is generated [2]. Processing of the nick by 3′→5′ exonucleases leads to gap extension [3]; subsequent Topo I-mediated ligation produces a heteroduplex dsDNA molecule containing a ssDNA loop [4]. This loop is then removed by either the activity of endonucleases [5] or, alternatively, through by-pass replication [6] producing a wild-type and deletion plasmid (▲).

is predominant at position –4 (83%). This may indicate that these residues are important for Topo I-mediated DNA cleavage in *B.subtilis*; in this process the T nucleotides may be important for cleavage, whereas the conserved C at position –4 could, for instance, be important for (stabilization of) DNA binding. The occurrence of the C nucleotide at position –4 is in good accord with earlier work by Dean *et al.* (21), who have shown a similar preference (100% in 30 independent cleavage sites) for a C at –4 for both the *E.coli* and *Micrococcus luteus* Topo I enzymes. In their study the authors did not, however, suggest a role for this invariant nucleotide.

In summary, the observations that (i) in the Topo I overproducing strain plasmid recombination frequencies were increased, (ii) pGP100 was present in a more relaxed configuration and (iii) type I deletions prevailed in this host, indicate that Topo I-dependent nicking is a major source of DNA lesions resulting in this type of deletion. The introduction of a transient nick by Topo I followed by exonucleolytic processing of this nick may be sufficient to produce a type I deletion, since, in theory, the covalently attached Topo I molecule may join the enzyme-bound 5′-phosphoryl end to any free end bearing a 3′-hydroxyl group (Fig. 6). Alternatively, Topo I may introduce and seal two separate single-strand nicks, resulting in a recombination event similar to those described previously (5,17). In any case, the partly single-stranded nature of the substrate (e.g. the predicted stem–loop structure between nt 2162 and 2315) may favour uncoupling of the cleavage and ligation reactions required for coupling of unrelated ends (7). When we aligned the end-points flanking the other side of the deletions (right hand end-points) a high incidence of a C residue at position –4 was again observed (76%). This may indicate that, indeed, Topo I introduces two separate nicks in one strand, thereby initiating intrastrand recombination and thus type I deletion. However, the overall occurrence of A or T nucleotides at these ends was much lower; instead, a C was present at positions –3 and –1 (45 and 42% respectively). At present it is not clear whether these sites may act as targets for Topo I and, therefore, we cannot rule out either model. To assess whether simultaneous nicking by DNA gyrase could be involved in specifying the right-hand end-points we have screened the vicinity of these termini for the presence of sequences resembling the consensus cleavage sites proposed for the *E.coli* (40) and *B.subtilis* (41) DNA gyrase enzymes. This revealed the presence of a sequence which resembled (part of) the *B.subtilis* DNA gyrase site (GNATGAT) at position –7 in only two of the 66 end-points analysed. Therefore, our current data do not support models in which DNA gyrase-dependent cleavage would specify the right-hand termini in the deletions analysed in this study.

ACKNOWLEDGEMENTS

We thank Henk Mulder for preparing the figures. Funding for the project, of which this work is a part, was provided by Stichting Technische Wetenschappen, The Netherlands. Parts of this work were supported by Gist-brocades B.V. (Delft, The Netherlands) and a CEU grant to S.B. and S.de J. (BRIDGE Programme, no. BIOT-CT910268).

REFERENCES

1 Kunst,F. *et al*. (1997) *Nature* **390**, 249–256.
2 Meuth,M. (1989) In Berg,D.E. and Howe,M.M. (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 833–860.
3 Ehrlich,S.D., Bierne,H., d'Alençon,E., Vilette,D., Petranovic,M., Noirot,P. and Michel,B. (1993) *Gene*, **135**, 161–166.
4 Ehrlich,S.D., Noirot,P., Petit,M.A., Jannière,L., Michel,B. and te Riele,H. (1986) In Setlow,J.K. and Hollaender,A. (eds), *Genetic Engineering*. Plenum Press, New York, NY, Vol. 8, pp. 71–83.
5 Bullock,P., Champoux,J.J. and Botchan,M. (1985) *Science*, **230**, 954–958.
6 Champoux,J.J. and Bullock,P.A. (1988) In Kucherlapati,R. and Smith,G.R. (eds), *Genetic Recombination*. American Society for Microbiology, Washington, DC, pp. 655–666.
7 Henningfeld,K.A. and Hecht,S.M. (1995) *Biochemistry*, **34**, 6120–6129.
8 Ikeda,H. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 3724–3728.
9 Ikeda,H., Kawasaki,I. and Gellert,M. (1984) *Mol. Gen. Genet*., **196**, 546–549.
10 Shuman,S. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 10104–10108.
11 Wang,H.P. and Rogler,C.E. (1991) *J. Virol*., **65**, 2381–2392.
12 Craig,N.L. and Nash,H.A. (1983) *Cell*, **35**, 795–803.
13 Abremski,K., Wierzbicki,A., Frommer,B. and Hoess,R.H. (1986) *J. Biol. Chem*., **261**, 391–396.
14 Zhu,J. and Schiestl,R.H. (1996) *Mol. Cell. Biol*., **16**, 1805–1812.
15 Peijnenburg,A.A.C.M., Bron,S. and Venema,G. (1988) *Plasmid*, **20**, 23–32.
16 Meima,R., Venema,G. and Bron,S. (1996) *Plasmid*, **35**, 14–30.
17 Bierne,H., Ehrlich,S.D. and Michel,B. (1997) *EMBO J*., **16**, 3332–3340.
18 Peijnenburg,A.A.C.M., Breed,P.V., Bron,S. and Venema,G. (1989) *Plasmid*, **21**, 205–215.
19 Meima,R., Haijema,B.J., Dijkstra,H., Haan,G.J., Venema,G. and Bron,S. (1997) *J. Bacteriol*., **179**, 1219–1229.
20 Conley,E.C., Saunders,V.A., Jackson,V. and Saunders,J.R. (1986) *Nucleic Acids Res*., **14**, 8919–8932.
21 Dean,F., Krasnow,M.A., Otter,R., Matzuk,M.M., Spengler,S.J. and Cozzarelli,N.R. (1983) *Cold Spring Harbor Symp. Quant. Biol*., **47**, 769–777.
22 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
23 Spizizen,J. (1958) *Proc. Natl. Acad. Sci. USA*, **44**, 1072–1078.
24 Bron,S. and Venema,G. (1972) *Mutat. Res*., **15**, 1–10.
25 Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
26 Kooistra,J. and Venema,G. (1991) *J. Bacteriol*., **173**, 3644–3655.
27 Bradford,M.M. (1976) *Anal. Biochem*., **72**, 248–254.
28 Pruss,G.J. (1985) *J. Mol. Biol*., **185**, 51–63.
29 Chedin,F., Dervyn,E., Dervyn,R., Ehrlich,S.D. and Noirot,P. (1994) *Mol. Microbiol*., **12**, 561–569.
30 Luria,S.E., and Delbrück,M. (1943) *Genetics*, **28**, 491–511.
31 Chopin,M.-C., Chopin,A., Rouault,A. and Galleron,N. (1989) *Appl. Environ. Microbiol*., **55**, 1769–1774.
32 Meima,R., Haijema,B.J., Venema,G. and Bron,S. (1995) *Mol. Gen. Genet*., **248**, 391–398.
33 Pruss,G.J. and Drlica,K. (1989) *Cell*, **56**, 521–523
34 Menzel,R. and Gellert,M. (1983) *Cell*, **34**, 105–113.
35 Tse-Dinh,Y.-C. (1985) *Nucleic Acids Res*., **13**, 4751–4763.
36 Oram,M. and Fisher,L.M. (1992) *J. Bacteriol*., **174**, 4175–4178.
37 Raji,A.J., Zabel,D.J., Laufer,C.S. and Depew,R.E. (1985) *J. Bacteriol*., **162**, 1173–1179.
38 Liu,L.F. and Wang,J.C. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 7024–7027.
39 Kirkegaard,K., Pflugfelder,G. and Wang,J.C. (1984) *Cold Spring Harbor Symp. Quant. Biol*., **49**, 411–419.
40 Morrison,A. and Cozzarelli,N.R. (1979) *Cell*, **17**, 175–184.
41 Bashkirov,V.I. and Zvingila,D.J. (1991) *Genetica*, **85**, 3–12.
42 Wertman,K.F., Wyman,A.R. and Botstein,D. (1986) *Gene*, **49**, 253–262.
43 Bolivar,F., Rodriquez,R.L., Greene,P.J., Betlach,M.V., Heynecker,H.L., Boyer,H.W., Crosa,J.H. and Falkow,S. (1977) *Gene*, **2**, 95–113.
44 Wang,R.F. and Kushner,S.R. (1991) *Gene*, **100**, 195–199.
45 Haijema,B.J., Noback,M., Hesseling,A., Kooistra,J., Venema,G. and Meima,R. (1996) *Mol. Microbiol*., **21**, 989–999.