The EMBO Journal vol.13 no.18 pp.4412–4420, 1994

Active site of the replication protein of the rolling circle plasmid pC194

M.F.Noirot-Gros, V.Bidnenko and S.D.Ehrlich

Laboratoire de Génétique Microbienne, Institut National de la Recherche Agronomique, 78352 Jouy en Josas Cedex, France

Communicated by S.D.Ehrlich

Mutation analysis of the rolling circle (RC) replication initiator protein RepA of plasmid pC194 was targeted to tyrosine and acidic amino acids (glutamate and aspartate) which are well conserved among numerous related plasmids. The effect of mutations was examined by an *in vivo* activity test. Mutations of one tyrosine and two glutamate residues were found to greatly impair or abolish activity, without affecting affinity for the origin, as deduced from *in vitro* gel mobility assays. We conclude that all three amino acids have a catalytic role. Tyrosine residues were found previously in active sites of different RC plasmid Rep proteins and topoisomerases, but not in association with acidic residues, which are a hallmark of the active sites of DNA hydrolyzing enzymes, such as the exo- and endonucleases. We propose that the active site of RepA contains two different catalytic centers, corresponding to a tyrosine and a glutamate. The former may be involved in the formation of the covalent DNA–protein intermediate at the initiation step of RC replication, and the latter may catalyze the release of the protein from the intermediate at the termination step.

Key words: catalytic center/DNA replication/DNase/exonuclease/type I topoisomerase

Introduction

Numerous plasmids isolated from Gram-positive bacteria replicate by a rolling circle (RC) mechanism similar to that described for single-stranded (ss) DNA phages in Gram-negatives (te Riele *et al.*, 1986a,b; Gros *et al.*, 1987, 1989; Boe *et al.*, 1989; Sozhamannan *et al.*, 1990). These replicons, called 'rolling circle plasmids' (RCP), have been studied extensively and their features are summarized in several reviews (Gruss and Ehrlich, 1989; Novick, 1989; Del Solar *et al.*, 1993, Jannière *et al.*, 1993). Briefly, the key element of RC replication is the plasmid-encoded initiator protein, termed Rep. Rep initiates replication by introducing a strand-specific nick at the so-called double-stranded origin (DSO). This generates a 3′ hydroxyl end which serves as a primer for DNA synthesis, and a 5′ phosphate end which is displaced by synthesis. After synthesis of a complete DNA strand, Rep also terminates replication by introducing a second cleavage at the same site and ligating the 5′ phosphate end to the newly created 3′ hydroxyl end. A circular ssDNA molecule is thus formed.

The RCP Rep proteins belong to the class of enzymes that can cleave a phosphodiester bond. Such cleavage involves a nucleophilic attack of the phosphate by a hydroxyl group. Two kinds of cleavage mechanism can be distinguished according to the hydroxyl donor, which is either an amino acid or a water molecule. The first mechanism resembles that of type I topoisomerases and related proteins. These enzymes catalyze DNA cleavage and also ligation of the cleaved strands. A covalent protein–DNA intermediate is formed during the reaction. Most often, protein and DNA are linked by a phospho–tyrosyl bond, as shown for the topoisomerase I of *Saccharomyces cerevisiae* or *Escherichia coli* (Lynn *et al.*, 1989). Tyrosine was also identified as the active amino acid in site-specific recombinases, such as the λ Int protein (Pargellis *et al.*, 1988) or the Flp recombinase of the yeast 2μm circle (Evans *et al.*, 1990). Furthermore, it was shown that the active site of the Rep protein of the RC phage φX174 contains two tyrosines, which are alternatively linked to the 5′ end of the cleaved DNA strand (Roth *et al.*, 1984; Van Mansfeld *et al.*, 1986; Hanai and Wang, 1994). However, in the case of the γδ transposon resolvase, the covalent protein–DNA bond is formed via serine (Newman and Grindley, 1984).

The second cleavage mechanism, where the nucleophile is a water molecule, has been described for various exo- and endonucleases. Crystallographic and mutagenesis studies have shown that amino acids carrying γ-carboxylate groups (glutamate and aspartate) are critical in this catalysis. These groups can serve as a general base to promote the formation of an attacking hydroxide ion by displacing a proton from a water molecule, as proposed for RNase H (Nakamura *et al.*, 1991). Alternatively, carboxyl groups can promote activation of the water molecule indirectly, either via a histidyl residue, which acts as a charge relay intermediate in the case of DNase I (Lahm *et al.*, 1991), or by chelating a metal ion, which in turn activates the water molecule as for the 3′–5′ exonuclease activity of *E.coli* DNA polymerase I (Freemont *et al.*, 1988; Beese and Steitz, 1991), or the staphylococcal nuclease (Hale *et al.*, 1993).

In this study we describe a mutation analysis of the Rep protein of the RCP pC194, termed RepA. pC194 belongs to a family containing some 20 plasmids (Seery *et al.*, 1993). Sequence alignment of the 15 most divergent Rep proteins of this family revealed five major conserved domains (Figure 1A). Site-directed mutagenesis was targeted on conserved tyrosine, glutamate and aspartate residues. A tyrosine and two glutamate residues involved in enzyme activity were thus identified. We propose a model of the RepA protein activity based on that of the phage φX174 Rep protein, whereby the active site of

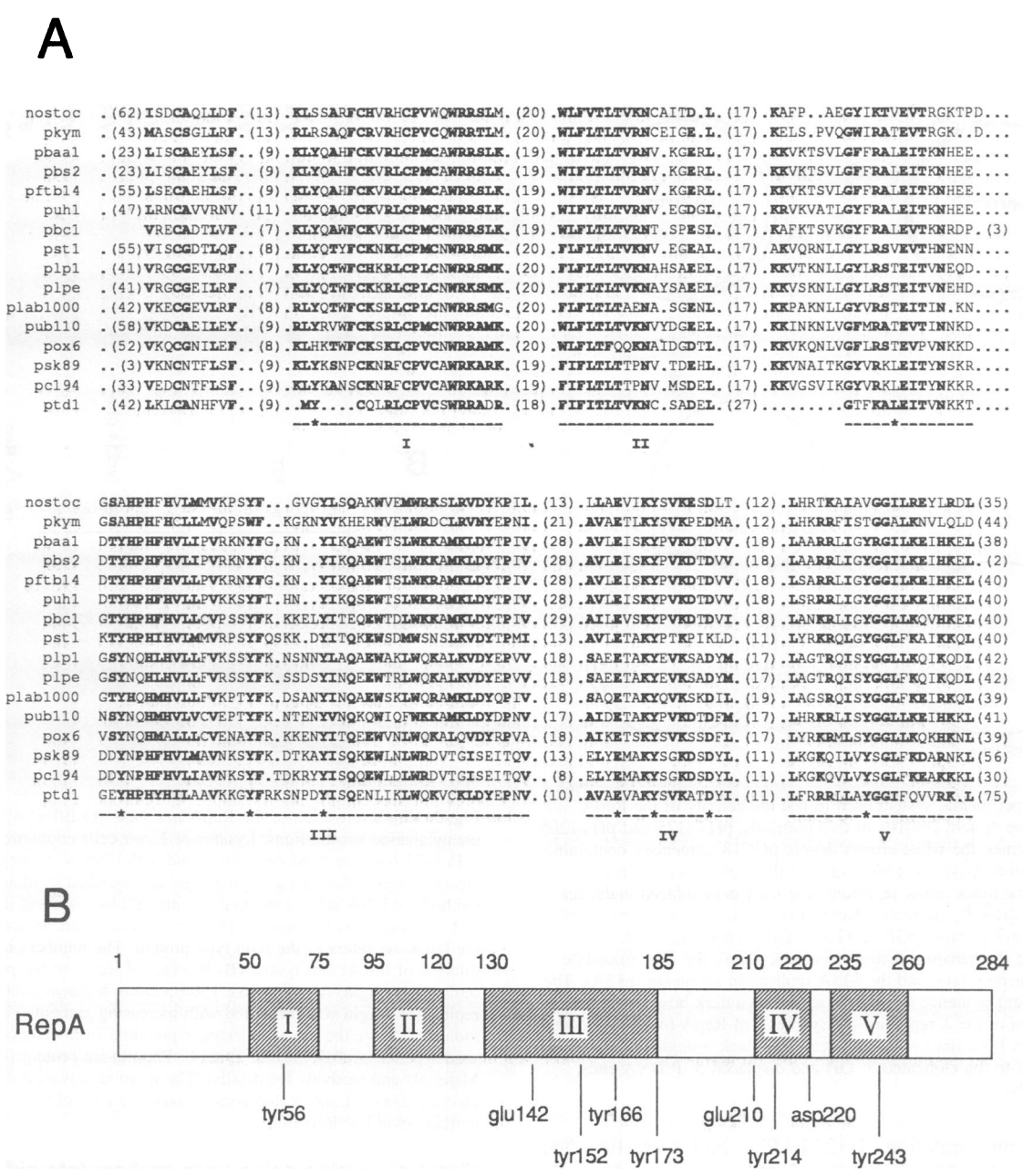

Fig. 1. (**A**) Homology between the Rep proteins of the pC194 plasmid family. Roman numbers identify the conserved regions, which are underlined. Amino acids, conserved or belonging to the same exchange group in at least eight of the 15 Rep proteins, are shown in bold. A star indicates the highly conserved tyrosine and carboxylic residues. (**B**) Schematic representation of the pC194 Rep protein. The conserved regions are represented as hatched boxes. The positions of tyrosines and carboxyl amino acids targeted for mutagenesis are indicated.

RepA contains two non-equivalent catalytic centers which cleave phosphodiester bonds by two different mechanisms, one resembling that of topoisomerases and the other that of nucleases.

Results

Experimental strategy

RepA activity was tested by the capacity of the protein to promote RC replication *in vivo*. Experiments were carried out in *E.coli* using the system represented in Figure 2, which is derived from a system developed in a previous work (Gros *et al.*, 1989). The system comprises two plasmids: (i) a RepA donor (pHV1175), which carries *repA* under the control of the λ P_L promoter, and (ii) a RepA target (pHV1205), which contains the pC194 DSO. The latter plasmid also carries the thermolabile λ cI_{857} repressor gene which controls the λ P_L promoter. The two plasmids are compatible and can be maintained together in *E.coli*. Synthesis of the RepA protein is induced at high temperature and its activity is estimated from the amount of target plasmid ssDNA.

Identification of a catalytic tyrosine

The six conserved tyrosine residues, Tyr56, Tyr152, Tyr166, Tyr173, Tyr214 and Tyr243 (Figure 1), were changed to phenylalanine. The activity of the mutant proteins is shown in Figure 3A and summarized in Table

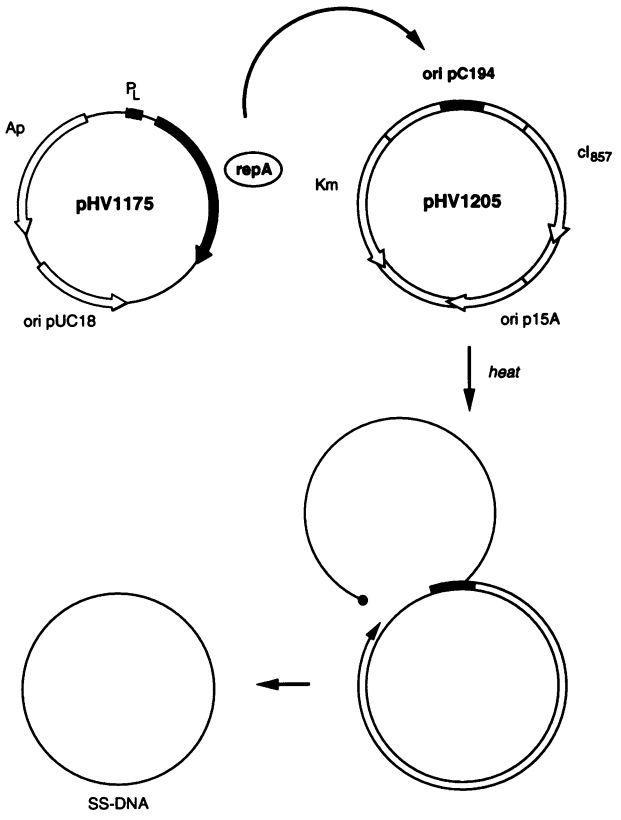

Fig. 2. Experimental system used to test the activity of the RepA protein. The system consists of two plasmids, pHV1175 and pHV1205. For the former, the white arrows denote pUC18 sequences, containing the ampicillin resistance gene (Ap) and the replication origin (ori pUC18); the black arrow represents the *repA* gene, placed under the control of the λ P_L promoter (hatched box). For the latter, the double circle denotes plasmid pGP1-2 (Tabor and Richarson, 1985), comprising the kanamycin resistance gene (Km), the λ thermolabile repressor gene (cI_{857}) and the p15A replication origin (ori p15A). The pC194 origin sequence (ori pC) is shown as a dark box. Upon heat inactivation of the λ repressor the synthesis of RepA (oval) is induced and ssDNA (thin line) is synthesized. The black triangle and dot correspond to the elongated 3′ OH and displaced 5′ P DNA ends, respectively.

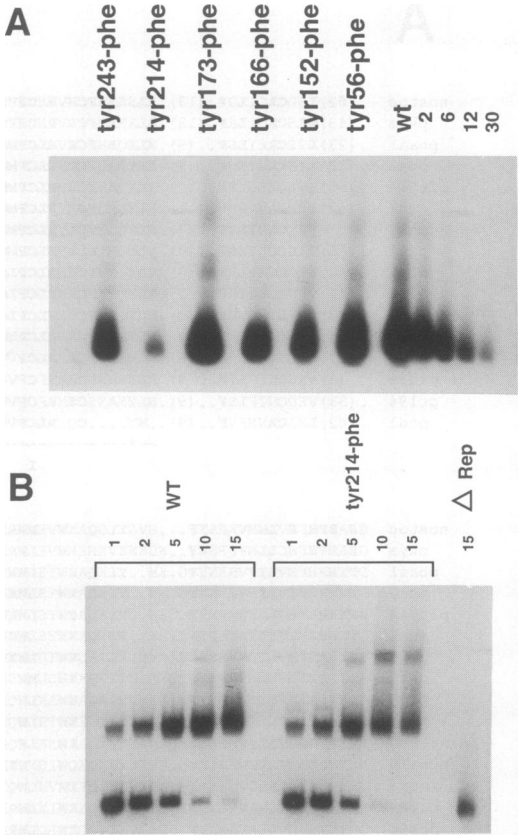

Fig. 3. Characterization of the RepA mutants. (**A**) Effect of tyrosine to phenylalanine substitutions. Lysates of *E.coli* cells containing plasmids pHV1175 (or a mutated derivative) and pHV1205 were prepared and electrophoresed through an agarose gel as described in Materials and methods. ssDNA was transferred to nitrocellulose without denaturation and hybridized with labeled pHV1205. Mutations are indicated above the lanes. wt refers to the wild type protein. The numbers indicate the dilution of the wt cell lysate. (**B**) Binding of RepA to the pC194 replication origin. A radioactive 150 bp DNA segment containing the replication origin was incubated with increasing amounts (1–15 μl, as indicated above the lanes) of extracts prepared from cells containing wild type RepA (WT) or the Tyr214–Phe mutant protein (see Materials and methods for details). The mixture was assayed by gel electrophoresis. Lane Δ Rep corresponds to a control assay with cell extract devoid of RepA.

I. Only one mutation, Tyr214–Phe, had a drastic effect, leaving only ~5% of the wild type RepA activity. Changing Tyr214 to serine also affected protein activity, decreasing the amount of ssDNA ~50-fold. Therefore, both the hydroxyl group and the phenyl radical of Tyr214 appear to be required for activity.

To rule out a possible loss of affinity of the Tyr214–Phe RepA for the origin, we compared the binding properties of the mutant and wild type proteins using a gel retardation assay. For this purpose, a radioactively labeled DNA fragment containing the pC194 origin sequence was incubated with varying amounts of cell extract, prepared after induction of Rep protein synthesis and analyzed by gel electrophoresis (Figure 3B). The binding capacities of the mutant and wild type proteins were similar, as judged by the amount of retarded fragment. Binding was specific, since a fragment which did not contain the origin sequence was not retarded (not shown). We conclude that Tyr214 is involved in the RepA catalytic reaction, since the mutation abolishes ssDNA synthesis without affecting origin binding.

Catalytic amino acids with carboxylate side chains

To test whether carboxyl amino acids are required for RepA activity, we targeted the mutagenesis to the three conserved residues. Two of these, Glu210 and Asp220, are in the vicinity of the catalytic Tyr214 residue, and the third, Glu142, is located within the conserved region III (Figure 1). Each of these amino acids was replaced by alanine and the activity of the mutant protein was determined (Table I). Activity of the mutant Asp220–Ala was not affected significantly, that of Glu142–Ala was barely detectable and that of Glu210–Ala was abolished. The *in vitro* binding of the mutated Glu142–Ala and Glu210–Ala Rep proteins to the origin sequence was not affected, as judged by the gel retardation assay (not shown). Therefore, Glu142 and Glu210 are required for the catalytic activity of the pC194 Rep protein.

Mutagenesis of the conserved region IV of RepA

The conserved region IV, comprising amino acids 209–220 (Tyr209–Glu–Met–Ala–Lys–Tyr–Ser–Gly–Lys–

Asp−Ser−Asp220), is rather rich in residues bearing a reactive side chain. It contains, besides the catalytic amino acids (Glu210 and Tyr214) and the non-essential Asp220, three hydroxyl amino acids (Tyr209, Ser215 and Ser219) and one carboxyl amino acid (Asp218). Furthermore, Tyr214 and Asp218 compose the motif $Tyr(X)_3Asp$, which is found in the catalytic site of the 3′−5′ exonuclease of DNA polymerases (Table II; Freemont *et al.*, 1988; Bernad *et al.*, 1989; Soengas *et al.*, 1992). To test whether these amino acids, which are not conserved in the majority of the proteins of the family, are nevertheless required for RepA activity, we changed them to either phenylalanine (Tyr209) or alanine (Ser215, Ser219 and Asp218). All mutants had an activity comparable with that of the wild type RepA (Table I). This indicates that among the carboxyl and hydroxyl amino acids of this region, only Tyr214 and Glu210 are required for RepA activity.

Table I. Activity of the RepA mutants[a]

Amino acid changes	Activity (%)
Tyr56−Phe	100
Tyr152−Phe	50
Tyr166−Phe	50
Tyr173−Phe	100
Tyr214−Phe	5
Tyr214−Ser	2
Tyr243−Phe	50
Glu142−Ala	<1
Glu210−Ala	2
Asp220−Ala	50
Tyr209−Phe	75
Ser215−Ala	50
Ser219−Ala	75
Asp218−Ala	100
Tyr214−Phe and Tyr209−Phe	5
Tyr214−Phe and Ser215−Ala	2
Tyr214−Phe and Ser219−Ala	5
Tyr214−Phe and Tyr209−Phe and Ser219−Ala	2
Tyr214−Phe and Tyr209−Phe and Ser215−Ala and Ser219−Ala	5
Tyr214−Phe and Asp218−Ala	5
Tyr214−Phe and Asp220−Ala	5
Tyr214−Phe and Glu210−Ala	2

[a]The activity was estimated from the amount of ssDNA produced in the standard assay. It is expressed relative to the activity of the wild type protein.

Characterization of the residual activity of the Tyr214−Phe RepA

The Tyr214−Phe mutant protein displays ~5% of the wild type Rep activity (Figure 3A and Table I). To test whether the residual and wild type activities are qualitatively similar, we checked circularity and strand polarity of the ssDNA generated by this protein. The specificity of the protein for the origin sequence was also investigated.

Circularity. To determine whether the mutant protein can carry out both the initiation and termination reactions, the circularity of the generated ssDNA was investigated using PCR. For this purpose total DNA was prepared, electrophoresed on agarose gel (see Materials and methods) and the gel regions containing ss and ds (CCC) DNA were cut in slices. The DNA was extracted from the gel, cleaved by *Bam*HI to eliminate any contamination of ss by ds DNA and amplified using two sets of primers (see Figure 4). One set (primers 1 and 2) flanked the origin nick site, whereas the other was located on one side of the nick site (primers 1 and 3). Circular molecules should be amplified by both sets, whereas linear molecules should be ampli-

Table II. Active sites of enzymes that cleave phosphodiester bonds

Rolling circle proteins[a]	210				214				218		
pC194	**E**	**M**	**A**	**K**	**Y**	S	G	**K**	D	**S**	**D**
ϕX174	**Y**	**V**	**A**	**K**	**Y**	V	N	**K**	K	**S**	**D**
	*				*						
Exonucleases[b]											
T5	I	**M**	W	P	**Y**	A	A	**K**	**D**	**T**	**D**
SpoI	I	**L**	K	V	**Y**	L	A	D	**D**	C	**D**
T7	**E**	**M**	M	D	**Y**	N	V	Q	**D**	V	V
phi29	**E**	E	Y	A	**Y**	I	K	N	**D**	I	Q
AcMNPV	V	**I**	**A**	**K**	**Y**	N	V	Q	**D**	C	M
Vaccinia	**D**	**M**	**A**	**R**	**Y**	C	I	**H**	**D**	**A**	C
HCMV	**Q**	**V**	**G**	**R**	**Y**	C	L	Q	**D**	**A**	V
Ec.PolI	**E**	A	**G**	**R**	**Y**	A	A	E	**D**	**A**	**D**
					*				*		
Topoisomerases[c]											
Ectopa	G	Y	I	T	**Y**	M	R	T	**D**	**S**	T
Ssptopa	G	F	I	T	**Y**	M	R	T	**D**	**S**	V
Ysctop3	G	F	I	S	**Y**	P	R	T	**E**	**T**	**D**
Bstop3	K	**L**	V	T	**Y**	P	R	T	**D**	**S**	N
Ssorevgyr	G	**L**	I	T	**Y**	H	R	T	**D**	**S**	N
Topamβ1	K	**L**	L	S	**Y**	P	R	T	**D**	**T**	P
Ecotopb	K	**L**	I	T	**Y**	P	R	S	**D**	C	R

[a]Conserved residues are shown in bold; stars indicate the positions of the two catalytic centers.
[b]Stars indicate the positions of the catalytic tyrosine and aspartate residues. Relevant amino acid similarity with the pC194 Rep protein is indicated in bold. The following abbreviations are used: SpoI, *Bacillus subtilis Spo*I bacteriophage; AcMNPV, *Autographa californica* mononuclear polyhedrosis virus; HCMV, human cytomegalovirus; Ec.PolI, *E.coli* DNA polymerase I. The data are from Soengas *et al.* (1992).
[c]Abbreviations are: Ectopa, *E.coli* topoisomerase I; Ssptopa, *Synechococus* sp. topoisomerase I; Bstop3, *B.subtilis* topoisomerase III (P.Noirot, personal communication); Ecotopb, *E.coli* topoisomerase III; Topamb1, plasmid pAMb1 topoisomerase (Swinfield *et al.*, 1991); Ysctop3, *S.cerevisiae* topoisomerase III; Ssorevgyr, *Sulfolobus acidocaldarius* reverse gyrase (Confalonieri *et al.*, 1993). The alignment was provided by P.Noirot (personal communication).

fied by primers 1 and 3 only. Amplification took place with both sets of primers on ssDNA generated by both proteins (Figure 4A and B), but was less efficient with the control primers (1 and 3), possibly because primer 3 was complementary to a palindromic region of the origin (the amplification products could nevertheless be easily seen on the original gel). DNA prepared from non-induced cells containing the Tyr214−Phe mutant gene was used as control. Amplification products were detected only at the position of dsDNA (Figure 4C), probably because *Bam*HI restriction was incomplete. We conclude that the mutant protein can carry out initiation and termination reactions.

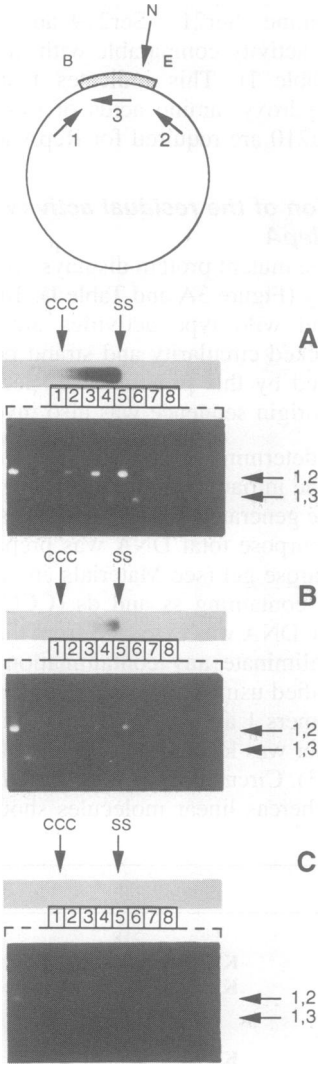

Fig. 4. Wild type RepA and Tyr214−Phe mutant proteins generate circular ssDNA. (**Top**) Schematic representation of the plasmid pHV1205. The pC194 origin is represented by a stippled box, N indicates the position of the nick site, B and E indicate the positions of the *Bam*HI and *Eco*RI restriction sites, respectively. The position and orientation of the primers are indicated by arrows, numbered 1–3. ssDNA synthesis was induced with the wild type RepA protein (**A**) the Tyr214−Phe mutant (**B**) or was not induced (**C**). DNA was extracted and duplicate samples were analyzed by gel electrophoresis. The position of ds covalently closed circular DNA (CCC) was determined by ethidium bromide staining; the position of ssDNA was determined by transferring one of the duplicate samples to nitrocellulose and hybridizing with the labeled pHV1205 probe. Slices of agarose were cut out of the lanes containing the other sample (boxes labeled 1–8), DNA was recovered, restricted with *Bam*HI and used as template for PCR amplification. Primers 1 and 2 (which amplify circular plasmid molecules) and primers 1 and 3 (which amplify both circular and linear molecules) were used. The two samples from each slice were loaded in alternate lanes and analyzed by gel electrophoresis. The figure is mounted such that the slices containing ssDNA (boxes 4 and 5) are above the lanes containing the corresponding amplified samples. The arrows labeled 1,2 and 1,3 indicate the size of the segments expected after amplification with the corresponding primers.

Strand polarity. The polarity of the ssDNA was examined by hybridization with a strand-specific DNA probe. The molecules generated by the wild type and mutant RepA corresponded to the same strand (data not shown).

Origin specificity. Wild type RepA protein introduces a nick into the sequence TTATCTTGATAA, between the G and the A (Michel and Ehrlich, 1986). To test the origin sequence specificity of the mutant RepA, a change was introduced at the nick site, replacing G with A. The wild type protein activity was greatly affected by this mutation, whereas the amount of ssDNA generated by the Tyr214−Phe protein was unchanged (Figure 5). Neither protein induced ssDNA synthesis from a target plasmid lacking the cognate origin sequence (not shown). This suggests that the Tyr214−Phe Rep protein is less specific than the wild type protein.

Effect of hydroxyl and carboxyl amino acids of region IV on the Tyr214−Phe mutant activity

The residual activity of the protein carrying the Tyr214−Phe substitution qualitatively resembles that of the wild type protein (see above). To test whether a

Fig. 5. Origin specificity of RepA. ssDNA was produced by the wild type (+) or the Tyr214−Phe (−) RepA protein (Rep), acting on the wild type (+) or the mutated (−) origin (Ori). The mutation changed a G adjacent to the nick site into an A.

proximal hydroxyl group might be involved in this activity, additional mutations were introduced in the Tyr214−Phe protein. RepA variants containing double, triple and quadruple substitutions were thus constructed by replacing Tyr209 with Phe and Ser215, and Ser219 with Ala. The different mutants generated slightly different amounts of ssDNA, but the quadruple mutant was as active as the Tyr214−Phe protein (Table I). Therefore, none of the hydroxyl amino acids in the vicinity of Tyr214 are involved in the residual activity of the RepA Tyr214−Phe mutant.

Similarly, double mutants were constructed containing the Tyr214−Phe mutation associated with substitutions of the acidic residues in this region (Glu210, Asp218 or Asp220). The residual activity persisted in all mutants (Table I), indicating that it does not require any of the carboxyl groups present in the vicinity of Tyr214.

Discussion

Mechanism of action of the pC194 Rep protein

Mutation analysis of the pC194 Rep protein shows that its active site contains at least three amino acids, Tyr214, Glu142 and Glu210. The presence of Tyr214 is not surprising, since active sites of enzymes which 'cut and join' DNA often contain a tyrosine residue (Pargellis *et al.*, 1988; Lynn *et al.*, 1989; Evans *et al.*, 1990). Furthermore, tyrosine appears to be required for the activity of two replication proteins which have homology with the pC194 Rep active site, that of *E.coli* phage ϕX174 (Roth *et al.*, 1984; Van Mansfeld *et al.*, 1986; Gros *et al.*, 1987; see Table II) and of *Shigella sonnei* plasmid pKYM (Yasukawa *et al.*, 1991). In contrast, the presence of the other two amino acids was not expected since, to our knowledge, carboxyl groups have not been found previously to be involved in the catalytic mechanism of enzymes which form covalent adducts with DNA via tyrosine. However, carboxyl residues are invariably present in active sites of enzymes which cleave DNA, such as the exo- and endonucleases (see King *et al.*, 1989; Lin and Sancar, 1992; Lin *et al.*, 1992; Hale *et al.*, 1993). Interestingly, recent crystallographic studies of the 67 kDa N-terminal fragment of the *E.coli* DNA topoisomerase I revealed similarities between the active site of this enzyme and that of the 3′−5′ exonuclease of DNA polymerase I, including the presence of acidic residues (Lima *et al.*, 1994).

The model presented below could account for our findings. It is based on that proposed previously for the ϕX174 Rep protein, but assigns to carboxyl amino acids the role they may play in different nucleases. The mechanism of action of the ϕX174 Rep protein is represented in Figure 6 (Roth *et al.*, 1984; Van Mansfeld *et al.*, 1986; Hanai and Wang, 1994). The critical amino acids are two tyrosine residues localized on the same side of an α-helical portion of the protein. DNA replication is initiated by a nucleophilic attack of the phosphodiester bond between the two nucleosides at the replication origin (A and G) by the hydroxyl group of a tyrosine (step 1). As a result, the protein is covalently bound to the 5′ end of the DNA molecule via a tyrosyl−phosphodiester bond, and a 3′ hydroxyl end is formed that can be used as a primer for DNA replication (step 2). After one round of replication the free tyrosine initiates a nucleophilic attack

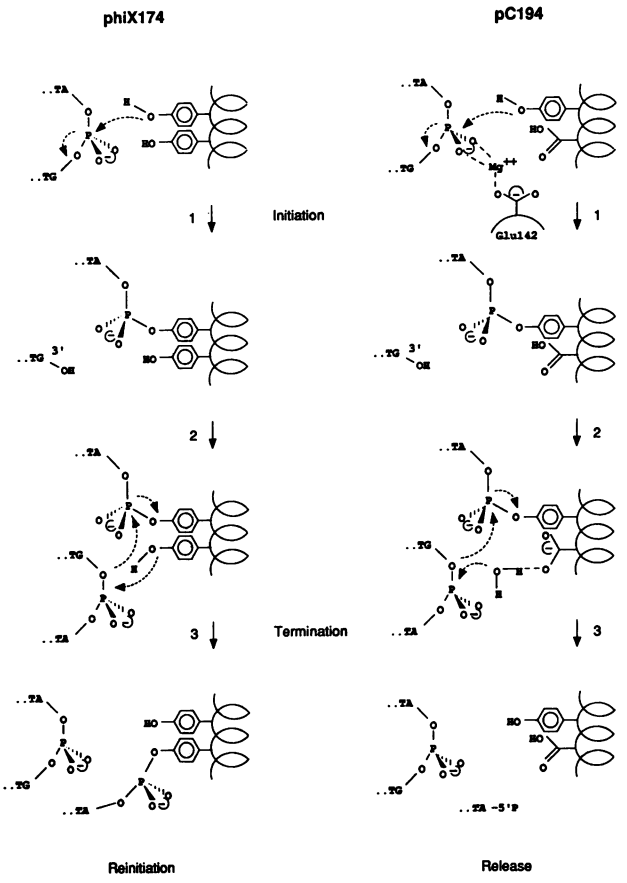

Fig. 6. Models of phage ϕX174 and plasmid pC194 Rep protein action. (**Left**) ϕX174 (after Van Mansfeld *et al.*, 1986). The active site contains two catalytic tyrosines separated by one turn of the α-helix. At initiation a hydroxyl group of one of the tyrosines carries out a nucleophilic attack (symbolized by dashed arrows) of the phosphodiester bond at the origin (step 1). A covalent linkage between the protein and the 5′ P end is thus formed and the resulting 3′ OH end is used as primer for DNA synthesis. After one round of replication, the newly synthesized origin is presented to the active site (step 2). At termination a novel nucleophilic attack by the hydroxyl group of the free tyrosine induces a double transesterification reaction. A circular ssDNA is generated and the protein is linked to the new 5′ P end by another tyrosyl−phosphoester bond. Replication can thus be reinitiated. (**Right**) pC194. The active site contains one catalytic tyrosine and one glutamate separated by one turn of α-helix. At initiation the tyrosine hydroxyl group carries out a nucleophilic attack, which may be facilitated by phosphate polarization, possibly mediated by electrostatic interaction with a metal ion (Mg^{2+}). Glu142 may be involved in chelating the metal ion. At termination a nucleophilic attack of the newly synthesized origin by a water molecule takes place. Glu210 may serve as a general base catalyst to promote formation of the hydroxide ion. A transesterification reaction is thus initiated, resulting in the formation of a circular ssDNA molecule and the release of the protein. Replication therefore cannot be reinitiated.

of the phosphodiester bond between the same bases (A and G) of the newly synthesized strand (step 3). A series of transesterification reactions follows, generating a circular ssDNA molecule and a molecule carrying the Rep protein at the 5′ end. Another round of replication can thus be reinitiated at the 3′ end of the same DNA molecule (not represented).

We propose that the pC194 protein initiates replication in a way similar to that of ϕX174, but that its action differs at the termination step (Figure 6). As in ϕX174, two of the active site amino acids of the pC194 Rep

protein, Tyr214 and Glu210, are placed within an α-helical region of the protein, as deduced from the computer analysis of the sequence, and their functional groups should protrude on the same side of the helix. Replication of pC194 is initiated by a nucleophilic attack on the phosphodiester bond involving the hydroxyl group of the Tyr214 (step 1). The active site Glu142 may facilitate the attack by binding and positioning a metal ion (Mg^{2+}), which would help to neutralize negative charges of the phosphodiester substrate, as described for the 3′−5′ exonuclease activity of *E.coli* DNA polymerase I (Freemont *et al.*, 1988; Beese and Steitz, 1991). The protein is thus covalently bound to the 5′ end of the DNA molecule and replication can start at the 3′ hydroxyl end (step 2). At the end of the round of replication, a water molecule activated by Glu210 mediates a nucleophilic attack of a phosphodiester bond on the newly synthesized strand (step 3, another possibility is that Glu210 positions and orients a water molecule, while the formation of the attacking hydroxide ion is promoted by a divalent metal ion). A concomitant transesterification reaction generates a circular ssDNA molecule and detaches the protein from the 5′ end of the DNA. Removal of the protein prevents reinitiation of a new round of DNA replication. It has been shown already that reinitiation does not take place during pC194 replication (Gros *et al.*, 1989), which is a prerequisite for regulating the plasmid copy number by controlling the level of plasmid Rep protein in the cell. In contrast to the plasmid life cycle, that of phage ϕX174 does not require copy number regulation, and the structure of its replication protein appears to favor reinitiation.

According to the proposed model, the active site of the pC194 protein is composed of two catalytic centers which are not equivalent and which cleave phosphodiester bonds by two different mechanisms, transesterification and hydrolysis. These mechanisms have been found so far in two different classes of enzyme, topoisomerases and nucleases, respectively. The pC194 replication protein could therefore be a link between these two classes.

It was shown recently that the replication protein of another RCP, pT181, is covalently bound to an oligonucleotide after a round of replication and is thus no longer active (Rasooly and Novick, 1993). Within the context of the model proposed here (Figure 6), similar inactivation of RepA would require that the cleavage at the termination step takes place at a phosphodiester bond downstream from the one cleaved at the initiation step. This would generate a circular DNA molecule and the protein attached to a short oligonucleotide. Further experiments would be necessary to test whether the RepA protein is released free or bound to an oligonucleotide.

Residual activity of the pC194 Rep Tyr214−Phe mutant

Somewhat unexpectedly, removal of the hydroxyl group crucial for the activity of the pC194 Rep protein does not completely inactivate the protein. However, several precedents of such functional flexibility exist for enzymes which 'cut and join' or which cleave DNA. For instance, removal of the hydroxyl group involved in the covalent binding of the yeast FLP recombinase by substituting Phe for Tyr343 does not abolish the protein activity, since the mutant can mediate a low level of strand transfer *in vitro* if one of the substrates has a 5′ hydroxyl end (Serre *et al.*, 1992), or in the presence of tyramine a tyrosine analog (Lee and Jayaram, 1993). Another example of functional flexibility is the variety of nucleophiles used by the retroviral integrase (IN) proteins, which catalyze cleavage of the dinucleotides from the ds viral DNA ends (Sherman and Fyfe, 1990). The natural nucleophile is most probably water, but the human immunodeficiency virus (HIV) IN protein can use alcohols, amino acids or even the 3′ hydroxyl group of the viral DNA (Vink *et al.*, 1991; Van Gent *et al.*, 1993). It is possible that in the pC194 Rep protein the natural nucleophile, which appears to be the hydroxyl group of the Tyr214, can be replaced by a hydroxyl group of another amino acid or even by a water molecule, albeit at a low efficiency. The mutation analysis has shown that neither the hydroxyl nor carboxyl group carried by a neighboring amino acid is involved in this reaction.

Interestingly, a point mutation at the putative nick site of the pC194 replication origin, which interferes drastically with the wild type Rep protein activity, does not affect the residual activity of the Tyr214−Phe mutant. The capacity to act upon a different sequence could be an intrinsic property of the mutant protein. Alternatively, the wild type protein may promote unspecific cleavages at a low rate. A precedent for such a phenomenon is the so-called star activity of restriction endonucleases. In the case of *Eco*RI endonuclease, the star activity is enhanced by a substitution of its Tyr193 residue, involved in DNA recognition, by a phenylalanine (Heitman and Model, 1990). It was proposed that the mutation could promote a change of the enzyme conformation triggering DNA cleavage at a non-canonical sequence. The hypothesis that RepA protein could promote unspecific cleavage is also supported by the finding that the Rep protein of plasmid pKYM, which is homologous to RepA, cleaves an origin sequence poorly related to that of other pC194-like plasmids (Yasukawa *et al.*, 1991, 1993).

Residues in the vicinity of the active site tyrosine

The three catalytic amino acids required for the activity of the pC194 Rep protein are conserved in all homologous plasmids, which indicates that the proposed mechanism may be extended to all members of this plasmid family. Two lysines in the vicinity of the active site tyrosine (Lys213 and Lys217) are also strictly conserved in all plasmids and in phage ϕX174, which suggests that they may be part of the active site. Their role has not been analyzed in this work, but basic residues were proposed previously to be involved in (i) DNA binding, as *Eco*RI endonuclease (Heitman and Model, 1990), (ii) neutralizing the negative charges of the phosphodiester substrate and thus facilitating the cleavage reaction mediated by the staphylococcal nuclease or FLP recombinase (Parsons *et al.*, 1990; Pourmotabbed *et al.*, 1990), and (iii) orientation and polarization of the Glu residue which catalyzes the ATPase reaction of the DNA gyrase B of *E.coli* and related proteins (Jackson and Maxwell, 1993). Remarkably, Tyr214 is part of the motif $Tyr(X)_3Asp$, which is conserved in >50% of the plasmid replication proteins (Figure 1) and is found in the active site of the 3′−5′ exonuclease domain of DNA polymerases and of type I DNA topoisomerases (Table II). Although the aspartate is

not involved in the activity of either the wild type or Tyr214–Phe mutant pC194 protein, as judged by our mutation analysis, the conservation of the motif may indicate an evolutionary relationship between proteins which 'cut and join' and proteins which cleave DNA.

Materials and methods

Bacterial strains

The *E.coli* strain CJ236 *dut1 ung1 thi1 relA1* (Kunkel *et al.*, 1987) was used to produce uracyl containing ss template for *in vitro* mutagenesis experiments; JM109 *recA1 endA1 gyrA96 thi hsdR17 supE44 relA1* Δ(*lac-proAB*)/F′ (*traD36 proA*$^+$ *proB*$^+$ *lacI*^{q} *lacZΔM15*) was used as a *dut*$^+$ *ung*$^+$ strain. The *E.coli* strain MC1061 (λ) *hsdR hsdM*$^+$ *hsdS*$^+$ *araD*139Δ (ara-leu) 7697Δ (*lacX74*) *galU galK rpsL* (str^r) (Remaut *et al.*, 1981), which expresses the λ cI repressor constitutively, was used for construction of plasmids containing the inducible λ P_L promoter. Strain HVC45 *thi1 leuB6 lacY1 tonA21 supE44 hsdR* str^r, harboring plasmids pHV1205 and pHV1175 or their mutated derivatives, was used to test the activity of the RepA protein.

Plasmids

Plasmid pHV1175 is a derivative of pHV980 (Gros *et al.*, 1989) which contains the following elements: (i) a sequence from pUC18 encompassing the ampicillin resistance gene and the pBR322 replication origin; (ii) a *PvuI–Bam*HI fragment derived from pLC28 (Remaut *et al.*, 1981), containing the thermoinducible λ P_L promoter; and (iii) the *RsaI–PvuII* segment of pC194 (nucleotides 1533–2592; Dagert *et al.*, 1984) encoding the *repA* gene, placed downstream of the P_L promoter. The pHV1175 derivatives carrying mutant *repA* were built by fragment exchange (see below). Plasmid pHV1205 contains the complete pC194 replication origin sequence, inserted in a 4 kb *Eco*RI–*Bam*HI fragment of vector pGP1-2 (Tabor and Richardson, 1985) which encompasses the thermolabile λ repressor gene cI_{857}, the replication origin of p15A and the kanamycin resistance gene of pACYC177. This plasmid is compatible with pHV1175 in *E.coli*.

In vitro mutagenesis

Mutagenesis experiments were performed according to the method of Kunkel (1987). The pC194 *repA* gene was inserted into M13mp18 and introduced into strain CJ236. ss molecules of the recombinant phage were extracted and used as template for *in vitro* complementary strand DNA synthesis initiated at oligonucleotide primers carrying a mutation to be introduced in the *repA* gene. The presence of the mutation in phage progeny was ascertained by DNA sequencing. The mutation was then transferred into plasmid pHV1175 by fragment exchange.

Induction of RC replication and ssDNA detection

Strain HVC45, containing the Rep protein donor plasmid pHV1175 (or mutated derivatives) and the recipient plasmid pHV1205, was grown to an OD_{650} of ~0.6 at 28°C in LB medium supplemented with ampicillin (100 μg/ml) and kanamycin (30 μg/ml). RC replication was induced by transferring the culture to 40°C. Samples were taken after 60 min and kept on ice. Total cell DNA was prepared from samples containing an equal number of cells, separated by agarose gel electrophoresis and transferred onto nitrocellulose filters without prior denaturation as described previously (te Riele *et al.*, 1986a,b). ssDNA was detected by hybridization with a pHV1205 DNA radioactive probe prepared by nick translation (Amersham kit 5500).

Preparation of extracts containing the RepA protein

HVC45 cells harboring pHV1175 derivatives and pGP1-2 were cultivated at 28°C in 200 ml LB to an OD_{650} of 0.8. RepA synthesis was induced by transferring the culture to 40°C for 60 min and then to 38°C for 2 h. Cells were collected by centrifugation, washed in 20 ml extraction buffer A (50 mM Tris–HCl, 5 mM Na_2EDTA, 1 mM DTT, 1 mM PMSF and 5% sucrose) and resuspended in 300 μl of the same buffer. They were then treated with lysozyme (0.5 mg/ml) for 30 min at 4°C, quickly frozen at −80°C and thawed at 4°C. NaCl was added to 1 M and cells were kept on ice until complete lysis. The lysate was cleared by centrifugation for 30 min at 35 000 r.p.m. in an SW50.1 rotor. It was dialyzed against buffer B (50 mM Tris–HCl, 100 mM KCl, 1 mM Na_2EDTA, 1 mM DTT, 1 mM PMSF), ethylene glycol was added to 10% and the preparation was stored at −80°C. 5 μl of the different extracts were loaded on SDS–acrylamide gel to confirm the presence of the Rep protein.

Protein binding assay

The binding of RepA to the pC194 replication origin was tested by a gel retardation assay. A 150 bp fragment containing the complete 55 bp origin sequence cloned in pUC18 (Gros *et al.*, 1989) was prepared by PCR using forward and reverse primers, labeled at the 5′ end by T4 kinase and used for the assay. Binding mixtures contained 5–10 ng of labeled fragment, 5 μg of carrier poly(dI–dC) and 1–15 μl of cell extract completed to 25 μl with buffer B (see below). They were incubated for 5–10 min at room temperature and then on ice for 1 h. The mixtures were analyzed by electrophoresis on a 5% polyacrylamide gel at 4°C, 10 mA, with a circulation of 0.2×TBE buffer. The gel was then transferred onto Whatman paper, dried under vacuum and autoradiographed.

Analysis of the circularity of ssDNA

Duplicate samples of total cell DNA were electrophoresed through an agarose gel. The position of ds covalently closed circles was identified by ethidium bromide fluorescence and a track containing one of the samples was sliced. DNA from the remainder of the gel was transferred to a nitrocellulose filter without prior denaturation and hybridized with the pHV1205 DNA probe to identify the position of the ssDNA. The DNA embedded in agarose slices was extracted by incubating the slices at −80°C for 30 min, transferring them to 37°C for 20 min and centrifuging the solution through a SPIN-XTM cellulose acetate filter unit (provided by COSTAR). The DNA solution (50 μl) was treated with *Bam*HI (10 U for 2 h at 37°C) and 10 μl were used for PCR amplification. The conditions were 94°C for 30 min, 52°C for 45 min and 72°C for 1 min, 30 cycles. PCR products were electrophoresed on a 2.5% agarose gel.

Acknowledgements

M.F.G. gratefully acknowledges the hospitality of Dr R.Novick during an early phase of this work. We thank P.Noirot for the alignment of topoisomerase active sites, C.Bruand and M.A.Petit for helpful and stimulating discussions during the course of this work and C.Anagnastopoulos, E.Cassuto, B.Michel, P.Noirot and M.Salas for their useful critical comments of the manuscript. This work was supported, in part, by the EEC grant BIOT-CT-91-0268.

References

Beese,L.S. and Steitz,T.A. (1991) *EMBO J.*, **10**, 25–33.
Bernad,A., Blanco,L., Lazaro,J.M., Martin,G. and Salas,M. (1989) *Cell*, **59**, 219–228.
Boe,L., Gros,M.-F., te Riele,H., Ehrlich,S.D. and Gruss,A. (1989) *J. Bacteriol.*, **171**, 3366–3372.
Confalonieri,F., Elie,C., Nadal,M., Bouthier de La Tour,C., Forterre,P. and Duguet,M. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 4753–4757.
Dagert,M., Jones,I., Goze,A., Romac,S., Niaudet,B. and Ehrlich,S.D. (1984) *EMBO J.*, **3**, 81–86.
Del Solar,G., Morosco,M. and Espinosa,M. (1993) *Mol. Microbiol.*, **8**, 789–796.
Evans,B.R., Chen,J.-W., Parsons,R.L., Bauer,T.K., Teplow,D.B. and Jayaram,M. (1990) *J. Biol. Chem.*, **265**, 18504–18510.
Freemont,P.S., Friedman,J.M., Beese,L.S., Sanderson,M.R. and Steitz,T.A. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 8924–8928.
Gros,M.-F., te Riele,H. and Ehrlich,S.D. (1987) *EMBO J.*, **6**, 3663–3669.
Gros,M.-F., te Riele,H. and Ehrlich,S.D. (1989) *EMBO J.*, **8**, 2711–2716.
Gruss,A. and Ehrlich,S.D. (1989) *Microbiol. Rev.*, **53**, 231–241.
Hale,S.P., Leslie,B.P. and Gerlt,J.A. (1993) *Biochemistry*, **32**, 7479–7487.
Hanai,R. and Wang,C.J. (1994) *J. Biol. Chem.*, **268**, 23830–23836.
Heitman,J. and Model,P. (1990) *EMBO J.*, **10**, 3369–3378.
Jackson,A.P. and Maxwell,A. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 11232–11236.
Jannière,L., Gruss,A. and Ehrlich,S.D. (1993) In Sonenshein,A.L., Hoch,J.A. and Losick,R. (eds), *Bacillus subtilis and other Gram-positive Bacteria: Biochemistry, Physiology, and Molecular Genetics.* American Society for Microbiology, Washington, DC, pp. 625–644.
King,K. Benkovic,S.J. and Modrich,P. (1989) *J. Biol. Chem.*, **264**, 11807–11815.

Kunkel,T.A., Roberts,J.D. and Dakour,R.A. (1987) *Methods Enzymol.*, **154**, 367–382.
Lahm,A., Weston,S.A. and Suck,D. (1991) *Nucleic Acids Mol. Biol.*, **5**, 171–186.
Lee,J. and Jayaram,M. (1993) *J. Biol. Chem.*, **268**, 17564–17579.
Lima,C.D., Wang,J. and Mandragon,A. (1994) *Nature*, **367**, 138–145.
Lin,J.-J. and Sancar,A. (1992) *J. Biol. Chem.*, **267**, 17688–17692.
Lin,J.-J., Phillips,M., Hearst,J.E. and Sancar,A. (1992) *J. Biol. Chem.*, **267**, 17693–17700.
Lynn,R.M., Bjornsti,M.-A., Caron,P.R. and Wang,J.C. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 3559–3563.
Michel,B. and Ehrlich,S.D. (1986) *EMBO J.*, **5**, 3691–3696.
Nakamura,H. *et al.* (1991) *Proc. Natl Acad. Sci. USA*, **88**, 11535–11539.
Newman,B.J. and Grindley,N.D.F. (1984) *Cell*, **38**, 463–469.
Novick,R.P. (1989) *Annu. Rev. Microbiol.*, **43**, 537–565.
Pargellis,C.A., Nunes-Duby,S.E., Moitoso de Vargas,L. and Landy,A. (1988) *J. Biol. Chem.*, **263**, 7678–7685.
Parsons,R.L., Evans,B.R., Zheng,L. and Jarayam,M. (1990) *J. Biol. Chem.*, **265**, 4527–4533.
Pourmotabbed,T., Dell'Acqua,M. and Gerlt,J.A. (1990) *Biochemistry*, **29**, 3677–3683.
Rasooly,A. and NovickR.P. (1993) *Science*, **262**, 1048–1050.
Remaut,E., Stansens,P. and Fiers,W. (1981) *Gene*, **15**, 81–93.
Roth,M.J., Brown,D.R. and Hurwitz,J. (1984) *J. Biol. Chem.*, **259**, 10556–10567.
Seery,L.T., Nolan,N.C., Sharp,P.M. and Devine,K.M. (1993) *Plasmid*, **3**, 185–193.
Serre,M.C., Zheng,L. and Jayaram,M. (1992) *J. Biol. Chem.*, **268**, 455–463.
Sherman,P.A. and Fyfe,J.A. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 5119–5123.
Soengas,M.S., Esteban,J.A., Lazaro,J.M., Bernad,A., Blasco,M.A., Salas,M. and Blanco,L. (1992) *EMBO J.*, **11**, 4227–4237.
Sozhamannan,S., Dabert,P., Moretto,V., Ehrlich,S.D. and Gruss,A. (1990) *J. Bacteriol.*, **172**, 4543–4548.
Swinfield,T.-J., Jannière,L., Ehrlich,S.D. and Minton,P.N. (1991) *Plasmid*, **26**, 209–221.
Tabor,S. and Richardson,C.C. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 1074–1078.
te Riele,H., Michel,B. and Ehrlich,S.D. (1986a) *EMBO J.*, **5**, 631–637.
te Riele,H., Michel,B. and Ehrlich,S.D. (1986b) *Proc. Natl Acad. Sci. USA*, **83**, 2541–2545.
Van Gent,D.C., Oude Groeneger,A.A.M. and Plasterk,R.H.A. (1993) *Nucleic Acids Res.*, **21**, 3373–3377.
Van Mansfeld,A.D.M., Van Teeffelen,H.A.A.M., Baas,P.D. and Jansz,H.S. (1986) *Nucleic Acids Res.*, **14**, 4229–4238.
Vink,C., Yeheskiely,E., Van der Marel,G.A., Van Boom,J.H. and Plasterk,R.H.A. (1991) *Nucleic Acids Res.*, **19**, 6691–6698.
Yasukawa,H., Hase,T., Sakai,A. and Masamune,Y. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 10282–10286.
Yasukawa,H., Hase,T. and Masamune,Y. (1993) *J. Gen. Appl. Microbiol.*, **39**, 237–245.

Received on April 5, 1994; revised on July 6, 1994