Plasmid RSF1010 DNA replication *in vitro* promoted by purified RSF1010 RepA, RepB and RepC proteins

Eberhard Scherzinger, Volker Haring, Rudi Lurz and Sabine Otto Max-Planck-Institut für Molekulare Genetik, Abteilung Schuster, Ihnestraße 73, D-1000 Berlin 33, FRG

Received January 16, 1991; Revised and Accepted February 19, 1991

ABSTRACT

We have constructed and analyzed an in vitro system that will efficiently replicate plasmid RSF1010 and its derivatives. The system contains a partially purified extract from E.coli cells and three purified RSF1010-encoded proteins, the products of genes repA, repB (or mobA/repB), and repC. Replication in this system mimics the in vivo mechanism in that it (i) is initiated at oriV, the origin of vegetative DNA replication, (ii) proceeds in a population of plasmid molecules in both directions from this 396-base-pair origin region, and (iii) is absolutely dependent on the presence of each of the three rep gene products. In addition, we find that E.coli DNA gyrase, DnaZ protein (γ subunit of pollII holoenzyme) and SSB are required for in vitro plasmid synthesis. The bacterial RNA polymerase, the initiation protein DnaA, and the primosomal proteins DnaB, DnaC, DnaG and DnaT are not required. Furthermore, the replicative intermediates seen in the electron microscope suggest that replication in vitro begins with the simultaneous or nonsimultaneous formation of two displacement loops that expand for a short stretch of DNA toward each other, and form a theta-type structure when the two displacing strands pass each other.

INTRODUCTION

RSF1010 is an 8684-base-pair (bp) multicopy plasmid conferring resistance to streptomycin and sulfonamides. It belongs to the incompatibility group Q and has the remarkable ability to replicate in almost all Gram-negative bacteria (1, 2). In E. coli, DNA replication is initiated within a unique origin on the plasmid genome and may proceed unidirectionally in either direction, or simultaneously in both directions (3). The same origin is also used in P. aeruginosa (4). DNA sequence analysis of this region, referred to as oriV, revealed three major characteristics (1; Fig. 1): (i) three perfect direct repeats of 20 bp, part of which is repeated a fourth time on their left, (ii) two adjacent regions of 28 and 37 bp to the right of the repeats that are extremely rich in G+C and A+T, respectively; and (iii) a large (64-bp stem, 28-bp loop) inverted repeat separated from the AT-rich region by 90 bp of DNA containing a small palindrome. For R1162, a plasmid that is essentially identical to RSF1010, the sequences required for origin function in vivo have been mapped to within two adjacent HpaII fragments of 370 and 210 bp, corresponding

to nucleotide (nt) positions 2181-2550 and 2561-2772 on the RSF1010 map. Although these fragments are both required, they may be separated or inverted with respect to each other, indicating the existence of a separate essential domain on each fragment (5).

In addition to requiring oriV DNA, RSF1010 replication requires at least three plasmid-encoded proteins, the products of genes repA, B and C (6). These have been isolated from overproducing strains and partially characterized in this laboratory. Purified RepC protein, a dimer of 31 kDa subunits, binds specifically to the direct repeats in the oriV region (7, 8). Purified RepA protein, a hexamer of 30 kDa subunits, was found to contain two enzymatic activities: it is a single-stranded (ss) DNA-dependent ATP(dATP)ase and also a DNA helicase (8), which catalyzes the unwinding of duplex DNA with an apparent 5'- to -3' polarity (E.S., unpublished data). The repB gene actually encodes two polypeptides of 36 and 78 kDa which arise from the use of two alternative start codons and the same stop codon in the *repB* sequence (1). DNA sequence comparisons and protein analyses revealed that RepB-78kDa is identical to MobA, a protein known from genetic studies to be required for RSF1010 mobilization in the presence of conjugative plasmids (9). We therefore proposed to call this two-domain protein, which appears to be dispensable for plasmid replication/maintenance in E. coli (2, 10), MobA/RepB (1). Our initial studies of the RepB and MobA/RepB proteins showed both of them to be RSF1010 DNAdependent DNA primases (8). Through the use of a series of M13/RSF1010 hybrid ssDNA templates in a RepB-dependent in vitro replication system, we were able to show that RepB protein synthesizes primers at two specific sites on RSF1010 DNA. These single strand initiation sequences, herein referred to as ssiA and ssiB (formerly designated as oriL and oriR, respectively) (1, 8), are located on opposite strands of the DNA within the large oriV palindrome and have a potential to form stable stem-loop structures in ssDNA. Recently, Honda et al. (11) have shown that the same sequences do also function as RepB-dependent priming signals in vivo when inserted in ori_c-defective M13 phage genomes. Such plasmid-encoded functions are thought to make RSF1010 independent of many host replication proteins (e.g., DnaB, C and G in *E.coli*) (4), although the host protein requirements in vivo have not been explored extensively.

In vitro systems that allow replication of RSF1010 plasmid DNA have been previously described by Diaz and Staudenbauer (12). These systems require cell-free extracts of RSF1010-harboring *E.coli* or *P.aeruginosa* strains and are insensitive to inhibitors of the bacterial RNA polymerase. The activity of the RSF1010-containing *E.coli* extracts was subsequently shown to be due to the plasmid-encoded RepA, B and C proteins (6). It was not known, however, whether replication *in vitro* was initiated at the *in vivo* origin (*ori*V) and whether it proceeded bi- or unidirectionally as it apparently does in the cell.

In this paper, we report that any supercoiled DNA that carries the *ori*V region of RSF1010 can be efficiently replicated in the presence of extract from plasmid-free *E.coli* cells and purified RSF1010 RepA, RepB (or MobA/RepB), and RepC proteins. Using this modified *in vitro* system, an analysis of the DNA sequences and host proteins required to support the replication of RSF1010 DNA has been initiated. An EM analysis of early replicative intermediates generated in the *in vitro* system is also presented. The results suggest that, unlike most other bacterial plasmids, RSF1010 replicates by a host-primosome-independent displacement mechanism, in which both daughter strands are synthesized continuously in the 5' to 3' direction, at least initially.

MATERIALS AND METHODS

Bacterial strains, plasmids, and phage

The E. coli K-12 strains used in this study are C600 (14), WM433 (dnaA204) (15), PC2 (dnaC2) (16), and AX727 (dnaZ2016) (17). Plasmids used are RSF1010 (18), ColE1 (19), and the R1 miniderivative pKN402 (20). The recombinant M13 phages mPS8 and mPS9 and their deletion derivatives, previously constructed for sequencing the oriT and oriV regions of RSF1010(1), were obtained from P.Scholz. mPS8 contains a 2.5-kb PvuII/PstI fragment of pKT228 (RSF1010::Tn3), carrying the RSF1010 sequences from nt position 1951 to 3982, cloned between the HincII and PstI restriction sites of the vector M13mp8. mPS9 contains the same fragment cloned between the HincII and PstI sites of M13mp9. Unidirectional deletions extending into the region of oriV containing the 20-bp direct repeats or into the region containing the 152-bp palindromic sequence were created by BAL-31 nuclease digestion after cleavage by BamHI of mPS9 or after cleavage by PstI of mPS8, respectively. The endpoints of the mPS9 deletions used in the study are at nt positions 2130, 2347, and 2369 on the RSF1010 map, and the endpoints of the mP8 deletions are at nt positions 2614, 2662, 2710, 2742, and 2791. A schematic diagram of these constructions is presented in Fig. 1.

Plasmid DNAs, maintained in *E. coli* C600, and M13 phage replicative form (RFI) DNAs, present in infected *E. coli* JM101 (21), were purified by an adaption of the method of Holmes and Quigley (22). The DNA was carried through two cycles of CsCl/ethidium bromide density gradient centrifugation, followed by extraction with 2-propanol, and extensive dialysis against 10 mM Tris HCl, pH 7.6/10 mM NaCl/1 mM EDTA.

Reagents, proteins, and antibodies

Hepes, rifampicin, unlabled nucleoside triphosphates, creatine phosphate, creatine kinase, and restriction endonucleases were obtained from Boehringer Mannheim. Aphidicolin, novobiocin, oxolinic acid and streptolydigin were from Sigma and [methyl-³H]dTTP (30 Ci/mmol) was from Amersham. *E. coli* SSB was purified by a modification of the procedure of Chase et al. (23), using an AMP-agarose fractionation step in place of the DNA-cellulose step (E.Lanka, personal communication). Homogeneous preparations of *E. coli* DnaC protein (fraction VI of ref. 24) and DnaZ protein (prepared by a modification of a

published procedure [25]), as well as antisera directed against DnaB protein, DnaG protein or SSB (26, 27) were obtained from E.Lanka of this department. A partially purified sample of *E.coli* DnaA protein was the gift of W.Messer (this institute), and antiprotein i gamma globulin was the gift of H.Masai (DNAX Research Institute, Palo Alto, USA).

Plasmid-encoded replication proteins

The RepA, MobA/RepB, RepB, and RepC proteins of RSF1010 were purified from the following overproducing *E. coli* strains (1, 7): HB101 [pVH4, pVH1] (RepA and C), HB101 [pVH3, pVH1] (RepB), and HB101 [pVH9, pVH1] (MobA/RepB). The isolation and characterization of each Rep protein will be described elsewhere (8; and unpublished work). For all four preparations used in this study, greater than 90% of the product migrated as a single band during NaDodSO₄/polyacrylamide gel electrophoresis (Fig.2). Purified R1 RepA protein was the gift of R.Díaz (Instituto de Biología Celular, Madrid).

Preparation of cell-free extracts

Fraction I was prepared by the freeze/thaw lysis method of Staudenbauer (28) from cells grown at 30° C. From it, fraction II, proteins insoluble in 73% saturated ammonium sulfate, was prepared as described (6).



Figure 1. RSF1010 map and extent of plasmid DNA carried on recombinant M13 phage DNAs used to define the minimum RSF1010 segment with origin function. The map shown at the top is based on the known at sequence of RSF1010 (1). Base pair coordinates originate from the HpaI cleavage site and read from left to right. Open boxes indicate the regions of the plasmid known to encode functional proteins. Thick lines between adjacent genes indicate overlap of at least one nucleotide. Gene mobB is contained within mobA/repB, and protein RepB (also called RepB'; 1, 8) arises from a fresh translation start in the middle of mobA/repB (marked by a dotted line). The large and small black boxes indicate the approximate extent of the genetic origins of replication (oriV) and DNA transfer (oriT), respectively (5, 13). The triangle labeled 228 marks the site of insertion of Tn3 in pKT228 (left insertion point at 3982). The second line shows the structure of a pKT228-derived, 2.5-kb PvuII/PstI fragment cloned in M13 mp 8/9 vectors. The viral DNAs of mPS8 and mPS9 carry the RSF1010 l strand and r strand sequences, respectively. The position of the major direct and inverted repeats in the oriV region are indicated (black and white arrowheaded boxes) as are a GC-rich region of 28 bp, an AT-rich region of 37 bp, and the 3'end of mobC (open box). The broken line indicates Tn3 sequences. The remaining horizontal lines indicate the extent of pKT228 DNA carried by various deletion derivatives of mPS8/9. The number at the end of each line is the nt position in the coordinates at the deletion end points. The area between the vertical dotted lines is the oriV sequence defined by the present in vitro studies (see Table 2 and the text).

Assay of DNA replication

The standard reaction mixture (25 μ l) contained the following components: Hepes KOH (pH 8.0), 40 mM; KCl, 100 mM; Mg (CH₂COO)₂, 10 mM; dithiothreitol, 0.4 mM; polyethylene glycol 6,000, 2.5%; Brij 58, 0.01%; ATP, 2 mM; GTP, CTP and UTP, each 0.4 mM; dATP, dGTP, dCTP and dTTP, each 50 μ M with [³H] dTTP at 180 cpm/pmol of total deoxynucleotide; NAD, 50 μ M; creatine phosphate, 15 mM; creatine kinase, 0.1 mg/ml; fraction II proteins from E. coli C600, 7-10 mg/ml; RSF1010 RepA protein, 3 µg/ml; RSF1010 RepB protein, 1 µg/ml; RSF1010 RepC protein, 6 µg/ml; and supercoiled DNA template, 13 µg/ml. Components were assembled on ice, and reactions were initiated by incubation at 30°C. The incubation time was 30 min unless otherwise indicated. DNA synthesis was measured by determining incorporation of labeled deoxynucleotide into acid-insoluble material as described (29).

Electron microscopy

Spreading of DNA was carried out by a modification of the method of Morris et al. (30). DNA molecules were spread at room temperature from a hyperphase of 50% (v/v) formamide, 0.01% cytochrome C, 20 mM Na₂CO₃ and 2 mM EDTA (pH 10.0) on a hypophase of double-distilled water. The spreading film was picked up with Parlodion-coated copper grids (60×300 mesh), dehydrated in 96% ethanol and rotary shadowed at an angle of $\approx 6^{\circ}$ with platinum-iridium (80:20). The grids were examined in a Philips EM400 electron microscope, and DNA molecules were photographed at a primary magnification of 5,200. Replicated and unreplicated regions of the molecules were measured on 16-fold enlarged negatives using an LM4 digitizer board (Brühl, Nürnberg). The total length of the DNA molecules was taken as 100% and the variance in size did not exceed 5%.

RESULTS

RepA, B, C- dependence of RSF1010 replication in vitro

A partially purified extract from *E. coli* C600 cells, fraction II prepared as described in ref. 6, supported the replication of plasmid ColE1 (6; see also Table 2). This *E. coli* enzyme fraction,



Figure 2. Polyacrylamide gel analysis of purified RSF1010 Rep proteins. Proteins ($\approx 4 \ \mu g$ each) were denatured, reduced, and electrophoresed in a 15% polyacrylamide gel containing 0.1% NaDodSO₄. Protein bands were stained with Coomassie blue R-250. *Lane* 1, MobA/RepB (predicted $M_{\rm T}$, 77,945); *lane* 2, RepB (35,889); *lane* 3, RepA (29,896); *lane* 4, RepC (31,123); *lane* S, protein size standards; in order of decreasing molecular weight, they are: DNA polymerase I, BSA, ovalbumin, chymotrypsinogen A, and lysozyme.

on the other hand, was incapable of catalyzing any DNA synthesis with plasmid RSF1010 as template, even when supplemented with a mixture of either the RSF1010 RepA and RepB proteins, the RepA and RepC proteins, or the RepB and RepC proteins (Table 1, lines 1-4). However, RSF1010-directed DNA synthesis did occur upon supplementing all three of these purified RSF1010 proteins (line 5), in agreement with the results of our earlier in vitro studies using partially purified Rep protein fractions (6). Kinetic analysis showed that the reaction started after a lag of 4-7 min and continued for about 60 min, at which time the extent of incorporation was typically 0.7-0.8 nmol of total deoxynucleotide per 25-µl standard reaction mixture (data not shown); this is equivalent to 70-80% of the DNA added to the assay mixture. As shown in Fig. 3, the *in vitro* reaction required relatively low concentrations of each purified Rep protein. In the presence of 13 μ g (2.3 pmol) of RSF1010 DNA per ml, half maximal activity was obtained with approximately 480 ng/ml of RepA, 70 ng/ml of RepB and 750 ng/ml of RepC; this corresponds to Rep monomer: DNA molar ratios of 7:1, 0.9:1 and 11:1, respectively. The optimal concentration of the C600 fraction II proteins was around 8 mg per ml; concentrations above 10 mg/ml were inhibitory.

Table	1.	Requirements	for	RSF1010	DNA	synthesis	in	vitro
					~	0,		

Omission	RepA	RepB	RepC	DNA synthesis, pmol
1. None	_	_	_	3
2. None	+	+	_	3
3. None	+	-	+	2
4. None	_	+	+	3
5. None	+	+	+	318
6. RSF1010 DNA	+	+	+	2
7. None*	+	+	+	3
8. C600 proteins	+	+	+	<1
9. dATP, dCTP, dGTP	+	+	+	4
10. CTP, GTP, UTP	+	+	+	221
11. ATP	+	+	+	3
12. CK, CP	+	+	+	42
13. PEG 6,000	+	+	+	64
14. None**	+	+	+	342

Reaction mixtures were incubated under standard conditions with RSF1010 DNA as template, except that individual components were omitted as indicated. Purified RSF1010 RepA, B and C proteins were added as indicated. CK, creatine kinase; CP, creatine phosphate.

* Linear RSF1010 DNA (by HpaI digestion) was used as template.

** The 36-kDa RepB protein was replaced by an equimolar amount of the 78-kDa MobA/RepB protein.



Figure 3. Titration of the amount of each RSF1010 Rep protein required for maximal replication activity. Reaction mixtures were incubated under standard conditions with RSF1010 DNA as template, except that the concentration of the specified Rep protein was varied as indicated, while the concentration of the other two proteins was fixed at 4 μ g/ml each.



Figure 4. Alkaline CsCl gradient centrifugation of DNA synthesized *in vitro*. Standard reaction mixtures (25 μ l) with RSF1010 DNA as template were incubated for 60 min at 30°C in the presence of [³H] dTTP (180 cpm/pmol), 75 ng of RepA, 150 ng of RepC, and either 25 ng of RepB (A) or 54 ng of MobA/RepB (B). The mixtures were then adjusted to 25 mM EDTA and 0.2 M NaOH and layered onto preformed 4.2 ml, linear CsCl gradients (density, 1.2-1.4 g/ml) containing 0.2 M NaOH. Centrifugation was for 1 h at 50,000 rpm in an SW60 rotor at 5°C. Fractions were collected from the bottom of the tubes and assayed for acid-insoluble radioactivity. *Arrows* marked *cc* and *oc* indicate, respectively, the positions where supercoiled and linear RSF1010 DNA sedimented in a parallel gradient.

Table 1 summerizes other features of the in vitro RSF1010 replication reaction. Activity was entirely dependent upon the addition of closed-circular DNA to serve as a template. Linear RSF1010 DNA was inert in this system (line 7). In addition to a dependence on each of the three RSF1010 Rep proteins, replication was found to be absolutely dependent upon the presence of the E. coli enzyme fraction, ATP, dNTPs, and Mg²⁺ions. The presence of an ATP-regenerating system and polyethylene glycol 6,000 (optimal concentration, 2,5%) greatly increased the activity, as previously observed for other in vitro RSF1010 replication systems (6, 12). The 78-kDa MobA/RepB protein, which is not essential for RSF1010 replication in vivo, at least in E. coli (2, 10), was an effectual substitute for the 36-kDa RepB protein in the *in vitro* system (*line* 14). As seen with RepB, the MobA/RepB-promoted replication was completely dependent upon the presence of the other two RSF1010 Rep proteins, A and C (data not shown).

The products of *in vitro* DNA synthesis in the presence of RepA, RepC, and either RepB or MobA/RepB were analyzed by alkaline CsCl gradient centrifugation (Fig. 4). In both cases, approximately 80% of the incorporated label was found in a single peak of fast-sedimenting DNA corresponding to closed-circular (cc) monomers of RSF1010. A small fraction of the label sedimented ahead of the monomeric cc form and may represent catenated (interlocked) plasmid forms. Analysis of ³²P-labeled reaction products by neutral agarose gel electrophoresis further showed that the newly formed ccDNA possessed the same superhelical density as the RSF1010 DNA added to the reaction mixture (data not shown).

Specificity of replication for RSF1010 oriV sequences

DNA synthesis in the presence of *E. coli* extract and purified Rep A, B and C proteins was specific for RSF1010 and its derivatives (Table 2, *Exp*.A). Plasmid pKN402, a mini-derivative of R1, and the RFI DNAs of phages ϕ X174 and M13 (which code for their own initiators) did not replicate in this system. On the other

 Table 2. Template activity of various supercoiled DNAs

		DNA synthesis, pm	ol	
-	T 1.	Without	With	
Exp.	Template	керА,В,С	керА,В,С	
Α	RSF1010	2	332	
	ColE1	314	295	
	pKN402	3	2	
	φX174	3	4	
	M13mp8	4	3	
	mPS8	4	462	
	mPS9	5	498	
В	mPS9-2	3	474	
	mPS9-15	5	482	
	mPS9-22	3	5	
	mPS8-54	4	486	
	mPS8-85	4	490	
	mPS8-78	5	264	
	mPS8-55	3	278	
	m PS8-8 1	4	4	

Supercoiled plasmid or phage RFI DNAs $(0,32 \ \mu g)$ were added to standard reaction mixtures. Replication activity was measured in the presence or absence of the RSF1010 Rep proteins as indicated.

hand, mPS8 and mPS9, two *in vitro* recombinants consisting of a pKT228 (RSF1010::Tn3)-derived, 2.5-kb *PvuII/PstI* fragment cloned between the *HincII* and *PstI* multilinker sites of M13 mp8 and M13 mp9, respectively, were as effective as templates as the intact RSF1010 plasmid. The insert of mPS8/9 comprises the RSF1010 sequences from nt position 1951 to 3982 (Fig. 1), and it has been shown *in vivo* to carry both a functional origin of replication (*oriV*) and a functional origin of transfer (*oriT*) (8; and unpublished data). Plasmid ColE1 was also an effective template. However, unlike RSF1010 and mPS8/9, ColE1 was replicated in the presence of the C600 protein fraction alone, and the addition of the RSF1010 Rep proteins had no discernible effect on the ColE1-dependent reaction.

To investigate the outer limits of RSF1010 DNA required for RepA, B, C-dependent in vitro replication initiation, we used a series of mPS8/9 mutant templates with BAL-31-generated deletions extending toward the oriV region from either the left or the right. Maps of 8 such derivatives are presented in Fig. 1, and the results of the replication assays are shown in Table 2 (Exp. B). From the center of the PvuII site at nt position 1948, up to 396 bp of DNA could be deleted rightward (left end point at 2347) with no detectable effect on template activity. Thus, the leftmost of the four 20-bp direct repeats (the presumptive RepC protein binding sites; 7, 8) and all of the RSF1010 DNA to the left of it is dispensable. Deletion of an additional 22 bp of DNA, eliminating 19 bp of the second 20-bp repeat (left end point at 2369), resulted in the loss of activity. From the nearest PstI site within the Tn3 part of pKT228, deletions of up to 1705 bp, eliminating all of the Tn3 sequences (465 bp) and all of the oriT region of RSF1010 (right end point at 2742), had no effect on template activity. Deletion of an additional 32 or 80 bp of DNA, eliminating part or all of the right arm of the 152-bp palindrome (right end points at 2710 and 2662), resulted in a 1.6-to-2-fold reduction in activity, and it was abolished when two thirds of the left arm were deleted as well (right end point at 2614). The latter data are in close quantitative agreement with in vitro replication assays performed on a series of single-stranded M13/RSF1010 hybrid DNAs (8; and unpublished data), showing that each arm of the oriV palindrome carries, on opposite strands of the DNA, a signal sequence for RepB (or MobA/RepB)-

 Table 3. Effect of inhibitors of DNA gyrase, RNA polymerase and DNA polymerase II on RSF1010 DNA synthesis in vitro

Inhibitor	Concentration	DNA synthesis, pmol	
None		328	
Novobiocin	5 μ g/ml	6	
Oxolinic acid	$100 \ \mu g/ml$	11	
Rifampicin	$25 \ \mu g/ml$	309	
Streptolydigin	$50 \ \mu g/ml$	278	
Aphidicolin	50 µM	307	
-	250 µM	314	

Inhibitors were added just prior to the start of the reaction.



Figure 5. Effect of antibodies against DnaB protein (A), DnaG protein (B) and protein i (C) on DNA synthesis. Replication of RSF1010 DNA (*closed circles*) was performed under standard conditions, except that the C600 protein fraction had been preincubated with the indicated amounts of anti-DnaB serum, anti-DnaG serum, or anti-protein i IgG for 10 min at 0°C; the level of serum protein in the various assay mixtures was kept constant by mixing the antibody preparations with appropriate amounts of the corresponding preimmune preparations. Control reactions using ColE1 DNA as template (*open circles*) were performed in an identical fashion, but for these reactions the RSF1010 Rep proteins were omitted. The 100% values were 235–280 and 210–275 pmol of dNMP incorporated for RSF1010 and ColE1, respectively.

dependent initiation of complementary strand synthesis. We conclude that the minimal RSF1010 segment competent for RepA, B, C-dependent initiation of bidirectional replication is greater than 342 bp and not more than 396 bp, lying between nt positions 2347 and 2742 and including the three perfectly conserved 20-bp repeats and all of the 152-bp palindromic sequence (Fig. 1). This segment of RSF1010 DNA is not known to overlap a protein-coding sequence (1), and no sequences analogous to the 9-bp *E.coli Dna*A box (31), the uX174-type primosome assembly site (32), or the G4-type DnaG recognition site (32) are found within or in the vicinity of it.

Involvement of host proteins in RSF1010 replication

We tested the effects of specific inhibitors (antibiotics, antibodies) or assayed the activity of protein fractions from dna^{ts} mutants of *E. coli*, in an attempt to identify some of the host proteins participating in the *in vitro* RSF1010 replication reaction. The results of these experiments, in which plasmids ColE1 and pKN402 served as controls of specifity, are summarized below, in Tables 3 and 4 and in Figs. 5 and 6.

DNA gyrase, RNA polymerase and DNA polymerase II: The sensitivity of RSF1010 replication to novobiocin and oxolinic acid at concentrations known to block the activity of DNA gyrase (33) clearly implicates this bacterial enzyme as a participant in the reaction. Rifampicin and streptolydigin, which act on the bacterial RNA polymerase, and aphidicolin, which selectively inhibits



Figure 6. Effect of antibody against SSB on RSF1010 DNA synthesis. Assays were as described in the legend to Fig. 5 except that antiserum against SSB and purified SSB were added as indicated. (A) Inhibition of the reaction by the antiserum. (B) Recovery of activity by addition of SSB to assay mixtures containing 2 μ l of anti-SSB serum.

Table 4. Replication activity of extracts from E. coli dna mutants

Source	<i>E. coli</i> protein added	DNA synthesis, pmol			
extract		RSF1010	ColE1	pKN402	
C600	None	323	289	_	
WM433	None	292	-	4	
	DnaA	251	_	198	
PC2	None	256	9	_	
	DnaC	238	84	_	
AX727	None	2	11	_	
	DnaZ	62	106	_	

Replication of RSF1010 DNA was performed under standard conditions, with fraction II protein as follows: C600, dna^+ (200 μ g); WM433, $dnaA^{15}$ (250 μ g); PC2, $dnaC^{15}$ (220 μ g); and AX727, $dnaZ^{15}$ (190 μ g). The mutant fractions were prepared from cells grown at 30°C and were used without prior heat treatment. Partially purified DnaA protein (6 μ g), purified DnaC protein (0.18 μ g), and purified DnaZ protein (0.1 μ g) were added as indicated. ColE1 replication was assayed under identical conditions, except that the RSF1010 Rep proteins were omitted. Assay mixtures with pKN402 template DNA contained purified R1 RepA protein (2 μ g) in place of the RSF1010 Rep proteins and were incubated at 30°C for 60 min.

DNA polymerase II of *E. coli* (Ki = 50 μ M) (34), failed to inhibit the reaction (Table 3).

DnaB, DnaG, and DnaT: Antibodies prepared against the *E.coli* primosomal proteins DnaB (helicase), DnaG (primase), or i (*dna*T gene product) also failed to inhibit RSF1010 DNA synthesis in this *in vitro* system (Fig. 5). By contrast, DNA synthesis in control incubations with ColE1 DNA as template was inhibited by each of these preparations, although to a different extent. Thus, the anti-DnaB and anti-protein i antibodies almost completely eliminated replication (95 and 86% inhibition, respectively), whereas the anti-DnaG antibody caused maximally only a 60% inhibition. This pattern of response is in accord with earlier *in vitro* studies on ColE1 (or pBR322) replication showing that the DnaB helicase and protein i are involved in the synthesis of both DNA strands and that the DnaG primase is required for synthesis of the H strand only (35, 36).

SSB: Involvement of this host protein in the replication of RSF1010 DNA is indicated by inhibition of the *in vitro* reaction by antiserum directed against SSB; addition of purified SSB neutralized the effect of the antibody (Fig. 6).

DnaA: Fraction II prepared from E. coli WM433 ($dnaA^{ts}$) was inactive in pKN402 replication when supplemented with purified R1 RepA protein; the addition of a partially purified DnaA protein fraction readily stimulated pKN402 replication, as was found by Díaz et al. (27). The same protein fraction (supplemented with

1208 Nucleic Acids Research, Vol. 19, No. 6



Figure 7. Electron micrographs of mPS9 molecules replicating in vitro. DNA synthesis was carried out with mPS9 RFI (13 µg/ml) as template in a 125-µl reaction mixture containing ddTTP (25 μ M) and rifampicin (20 μ g/ml) in addition to the standard components. Incubation was at 30°C for 20 min after which the reaction was stopped by addition of EDTA to 25 mM and NaDodSO4 to 0.5%. The DNA was purified essentially as described by Kaguni et al. (39). The sample was heated at 65 °C for 3 min and then incubated for 30 min at 37°C with proteinase K (0.5 mg/ml). The mixture was extracted twice with phenol/chloroform/isoamylalcohol (25:24:1), twice more with diethyl ether, layered onto 75 μ l of a CsCl shelf (density, 1.42 g/ml), and centrifuged for 1 h at room temperature in a Beckman airfuge. The CsCl shelf was recovered, diluted 2-fold with TE (10 mM Tris·HCl, pH 8.0/1 mM EDTA), and the DNA was precipitated with 2.5 volumes of 95% ethanol. After washing with 70% ethanol, the pellet was redissolved in 30 μ l of TE and incubated with pancreatic RNase (5 μ g/ml) at 37°C for 30 min to digest contaminating RNA. Ten-microliter samples ($\approx 0.4 \ \mu g \ DNA$) were then treated with 4 units of EcoRI or PstI endonuclease for 30 min at 37°C in reaction mixtures (25 µl) containing 100 mM Tris HCl (pH 7.6), 50 mM NaCl and 5 mM MgCl₂ or 10 mM Tris·HCl (pH 7.6), 100 mM NaCl and 10 mM MgCl₂, respectively. Restriction reactions were terminated by additon of EDTA to 20 mM, and the DNA was prepared for electron microscopy as described under Materials and Methods . Representative types of replicating molecules obtained after cleavage with EcoRI (a-d) or PstI (e, f) are shown. Arrows point to parental single strands displaced by replication.

the RSF1010 RepA, B, and C proteins), on the other hand, supported a normal level of RSF1010 DNA synthesis and, unlike the case of the R1 system, supplementing wild-type DnaA protein had no stimulatory effect (Table 4).

DnaC: Fraction II prepared from *E.coli* PC2 ($dnaC^{ts}$) was inefficient in ColE1 replication but showed RSF1010 replication activity similar to that of the C600 (dna^+) extract. The small amount of incorporation observed with ColE1 was stimulated 8-fold by addition of purified DnaC protein, indicating that the PC2 protein fraction was deficient in DnaC activity (Table 4).

DnaZ: The *dna*Z gene product has been characterized as the 52-kDa γ subunit of DNA polymerase III holoenzyme (25) but

has also been implicated in DNA polymerase II-catalyzed DNA synthesis reactions (37). An essential role of this host protein in RSF1010 replication is indicated by (i) the complete inability of an enzyme fraction to support the *in vitro* reaction when it was prepared from *E. coli* AX727 (*dnaZ*^{ts}), and (ii) the capacity of purified DnaZ protein to complement the defect and restore DNA synthesis (Table 4). The slight activity observed with ColE1 most likely represents early intermediate synthesis by DNA polymerase I, which is known to occur in the absence of functional DnaZ protein (38).

Electron microscopy of replicative intermediates

To ascertain that the extensive RepA, B, C-dependent DNA synthesis obtained with mPS9 RFI was indeed initiated from the RSF1010 oriV sequence and to obtain information as to the directionality of replication, early replicative intermediates were analyzed by electron microscopy. These were forcibly accumulated by carrying out the reaction in the presence of the chain terminating inhibitor 2',3'-dideoxy TTP. In the experiment to be reported (see legend to Fig. 7), we chose a concentration of ddTTP relative to dTTP that allowed [3H] dTMP incorporation to proceed to 7% of the maximum obtained in the absence of the inhibitor. Possible contribution of the sequence in the M13 viral strand recognized by RNA polymerase for complementary strand synthesis has been eliminated by carrying out the reaction in the presence of rifampicin. Following incubation for 20 min at 30°C, the DNA was purified and a portion of the sample was immediately prepared for electron microscopy using the formamide technique to demonstrate singlestranded DNA (30). The preparation was found to contain up to 25% of the total DNA molecules as theta-shaped replication intermediates. The extent of replication ranged from 2 to 30%. Circular molecules with tails attached to them were not observed in significant numbers (< 1%), although short single-stranded tails were seen at some of the replicative forks in the theta-shaped intermediates. These whiskers presumably arose from branch migration at the replicative fork and have been previously observed in the replicative intermediates from various systems (40).

After treatment of the DNA with the endonucleases *Eco*RI or PstI, each of which cleaves mPS9 such that the oriV region is located near an end of the linear molecule (see Fig. 8), most (87%) replicative intermediates were seen as linear molecules composed of a bubble of replicated DNA bounded by one short and one long unreplicated arm. In about 30-40% of these molecules the replication bubble appeared to have a typical Dloop structure (one duplex and one single-stranded arm) (Fig. 7a, b and f), but the majority of the intermediates contained a stretch of duplex DNA on both arms of the replicative loop. The duplex and single-stranded regions were frequently located at positions that were trans with respect to each other at both replicative forks, as in the examples shown in Fig. 7c, d and e. Replicative intermediates with similar structural features have been previously observed for chloroplast DNA from Chlamydomonas reinhardii in which replication initiates with the formation of two displacement loops (41). The remaining restricted intermediates observed (13%) consisted of a partially single-stranded circle of mPS9 unit length with two branches of replicated DNA attached to it. The structures of these molecules, like the one shown in Fig. 7b, suggested that the restriction site fell into a three-stranded displacement loop. Among a total of 126 EcoRI- or PstI-cleaved intermediates scored, only one double-Y-shaped molecule was observed (not shown).



Figure 8. Line diagrams of partially replicated mPS9 molecules cleaved with *Eco*RI (A) or *Psi*I (B). DNA molecules similar to those shown in Fig. 7 were photographed, and lengths of the replicated and unreplicated regions were measured from the projected micrographs. The molecules are aligned in order of increasing extent of replication with the long unreplicated region of DNA to the right. The *boxes* represent the replication 'bubble'. The corresponding linear maps of mPS9 (10.1 kb) are shown on the top of the figure. The *dark area* represents DNA from RSF1010; the *hatched area*, DNA from Tn3; and the *thin line*, M13 mp9 sequences. The *bracket* indicates the extent of the *ori*V sequence defined in the present work, and the *dotted area* marks the location of the 152-bp palindrome contained in it. Restriction sites are: A, AccI; E, EcoRI; and P, PstI.



Figure 9. Model for the initiation of leading strand synthesis at the two RepB sites within *ori*V. See the text for details. *Big circles* and *small circles* labeled *A* represent RepC and RepA protein molecules, respectively. *Dark boxes* on the l strand and *open boxes* on the r strand represent the *ssi*A and *ssi*B priming signals, respectively. Oligonucleotide primers and newly synthesized DNA chains are indicated by *interrupted* and *arrowheaded lines*, respectively.

In orienting the molecules in Fig. 8 A and B, the 152-bp oriV palindrome (nt position 2589–2740 in RSF1010) is located at a distance of approx. 7.1% from the left end of the *Eco*RI- cut DNA and 17.6% from the left end of the *Pst*I-cut DNA. Inspection of the least replicated molecules reveals that most initiations occurred at or near this site. Of the 126 intermediates analyzed, only 8 were judged not to have been replicated from

the normal origin as the replicated loop did not overlap the oriV palindrome within the error of measurement (2.5% of genome length). Given the location of the preferred initiation site, the pattern of more extensively replicated molecules indicates that the progress of replication from that site was bidirectional in some molecules and undirectional (in either direction), in others. A similar non-homogeneous pattern of replication from oriV has also been observed *in vivo* (3, 4).

We have repeated the EM analysis of replication initiation *in vitro* after carrying out the reaction in the presence of the bifunctional MobA/RepB protein in place of the RepB protein. No difference to the pattern of replication seen in Fig. 8 was noted for the intermediates generated in this alternative *in vitro* RSF1010 replication system (data not shown).

DISCUSSION

A relatively crude *E. coli* protein fraction (AS 0-73) capable of carrying out the complete process of ColE1 DNA replication can also replicate the broad-host-range plasmid RSF1010 if it is supplemented with purified RSF1010 RepA, B and C proteins. The *in vitro* reaction was found to depend on Mg^{2+} , ATP, the four dNTPs, and the host proteins DNA gyrase and DnaZ in addition to each of the three plasmid-encoded Rep proteins. Our studies of RSF1010 replication in the soluble enzyme system failed to detect an obvious role for either the bacterial RNA polymerase, the initiation protein DnaA, or the primosomal proteins DnaB, DnaC, DnaG, and DnaT. These results are consistent with observations *in vivo* (4) and earlier *in vitro* results obtained with less purified RSF1010 replication systems (6, 12).

In addition, our present data suggest the involvement of E. coli SSB in replication of the RSF1010 plasmid.

The inability of linear RSF1010 DNA to serve as template in this system suggests that replication is activated by negative supercoiling of the template DNA. A role for gyrase in RSF1010 replication beyond that of generating superhelical template is predicted from the inhibitory effects of novobiocin and oxolinic acid on the replication of plasmid DNA that is added to the reaction mixture in a supercoiled form. The other two host proteins shown here to be involved in the in vitro reaction. DnaZ and SSB, are known to function in the elongation of RNA- or DNA-primed single-stranded DNA that is catalyzed by DNA polymerases II or III in conjunction with other elongation factors (37). Since aphidicolin, a specific inhibitor of pol II (34), failed to inhibit the reaction, it seems likely that pol III holoenzyme, including the DnaZ protein (γ subunit), is the DNA polymerase responsible for the incorporation of most of the deoxynucleotides in this system.

Our studies have localized the RSF1010 sequences required for initiating replication within an intergenic region of 396 bp, from nt position 2347 to 2741 (Fig. 1). All chimeric M13/RSF1010 RFI DNAs containing this sequence were as efficiently replicated in vitro as authentic RSF1010 plasmid DNA, showing that the leftmost (imperfect) of the four 20-bp direct repeats and the sequences to the right of the 152-bp palindrome are not required. Note that the direct repeats are the site of RepC binding (7, 8) and that the palindromic sequence includes two single strand initiation signals (ssiA and B) that are specifically recognized by RepB protein for primer synthesis (8,11). The two RepB sites, which can form a hairpin structure in ssDNA. are located on opposite DNA strands and are oriented such that the chain elongations initiated from them converge. Deletion of the ssiA site caused a marked decrease in the template utilization efficiency in vitro (mPS8-55 and -78; Table 2), and strand specifity analysis revealed that only 1 strand synthesis occurred with these mutant template DNAs (data not shown). Deletion of both ssiA and ssiB resulted in the complete loss of template activity (mPS8-81; Table 2), even though the RepC binding sites and 200 bp of DNA to the right of them, including the GC-rich and AT-rich regions, remained. These results suggest that ssiA and B, initially identified on ssDNA (8), are the actual replication initiation sites in RSF1010 and that DNA synthesis may be initiated independently at each site. In keeping with this, our EM studies of initiation in vitro indicated the following (Figs. 7 and 8): (i) replication is primarily initiated from two specific sites within the oriV sequence; (ii) initiation at those two sites is not always synchronized; (iii) at each origin, replication begins with the formation of a D-loop resulting from the synthesis of one daughter strand; and (iv) the two displacing daughter strands expand toward each other and initiate the formation of a structure that looks like a theta-type intermediate when they elongate past each other. Thus, RSF1010 replication in our in vitro system appears to involve only a continuous mode of DNA synthesis such as occurs during the replication of the circular duplex DNAs from plant chloroplasts and animal mitochondria or certain viruses with linear duplex genomes (e.g., adenoviruses and B. subtilis phage ϕ 29) (42). Although the pattern of RSF1010 replication seen in vivo (3, 4) is not inconsistent with such a simple mode of replication, the fine structure of the replicative intermediates has not been characterized in the previous in vivo studies, making a distinction between completely continuous and semidiscontinuous replication difficult; thus the physiological significance of the in vitro results remains to be seen.

One of the major unanswered questions regarding initiation of RSF1010 replication is the precise mechanism of how binding of RepC protein to the direct oriV repeats leads to initiation of leading strand synthesis at nearly 200 (ssiB) and 300 bp of DNA (ssiA) to the right. Since RepB can synthesize oligonucleotide primers at ssiA and B on a cloned single-stranded DNA but not on the duplex DNA containing the same sequences (8), it is predicted that specific protein-DNA and protein-protein interactions at oriV somehow destabilize the duplex near the two priming signals to permit their recognition by RepB protein. This could be achieved, subsequent to RepC binding, by unidirectional unwinding of the duplex by the RepA helicase, as depicted schematically in Fig. 9. Studies with the *E. coli ori*C and λdv plasmid replication systems indicated that AT-rich sequences present in the respective origins are very likely the site where the DnaB helicase and other initiation proteins are transferred to the DNA (43, 44). A similar AT-rich region present within oriV (Fig. 1) could serve as an entry site for RepA protein, which then translocates in the 5' to 3' direction along the 1 strand, melting the duplex as it moves. The unwinding may be aided by DNA gyrase and stabilized by SSB. A short but highly GCrich (86%) segment of DNA located between the RepC binding sites and the AT-rich zone (Fig. 1) might initially block a second RepA molecule, positioned on the r strand, from initiating unwinding in the leftward direction. When the rightward-moving RepA molecule arrives at an ssi site, the RepB primase can recognize this sequence on the separated DNA strands and initiate leading strand synthesis, in either direction. Once initiated, both daughter strands will be synthesized continuously, with RepA protein presumably positioned at each growing fork to facilitate displacement of the nonreplicated parental strand. At this time, however, we cannot rule out the possibility that assembly of a nucleoprotein complex at oriV directly induces a superhelixdriven opening of the duplex near the two ssi sites without involving RepA-mediated DNA unwinding.

During analysis of the protein requirements for the *in vitro* RSF1010 replication reaction, we found that extensive DNA synthesis also occurred when the 78-kDa MobA/RepB protein was substituted for the 36-kDa RepB protein. The cofactor and protein requirements for replication in the presence of MobA/RepB were otherwise identical to those for reactions carried out with RepB, and replication was initiated from the normal RSF1010 origin, oriV. This result was not unexpected since our initial studies of the purified MobA/RepB and RepB proteins showed both of them to be DNA primases, specific for the RSF1010 ssiA and B sequences (8). Interestingly, our recent results have shown that protein MobA/RepB also possesses a strand- and sequence-specific DNA breakage-reunion activity that acts at the plasmid's origin of transfer and depends on the simultaneous presence of at least one other RSF1010 protein, MobC (E.S., manuscript in preparation). It is likely that this activity of MobA/RepB is required during the initiation and termination steps in RSF1010 transfer replication.

ACKNOWLEDGEMENT

We thank Heinz Schuster for generous support and helpful discussions. We especially express our gratitude to Drs. Ramon Diaz, Erich Lanka, Hisao Masai and Walter Messer for their generous gift of *E. coli* replication proteins and antibodies. We also thank Renate Spann for preparation of the manuscript.

REFERENCES

- Scholz, P., Haring, V., Wittmann-Liebold, B., Ashman, K., Bagdasarian, M. and Scherzinger, E. (1989) Gene 75, 271-288.
- Frey, J. and Bagdasarian, M. (1989) In Thomas, C. M. (ed.), Promiscuous Plasmids of Gram Negative Bacteria, Academic Press, London, pp. 79-94.
- de Graaff, J., Crosa, J. H., Heffron, F. and Falkow, S. (1978) J. Bacteriol. 134, 1117-1122.
- Scholz, P., Haring, V., Scherzinger, E., Lurz, R., Bagdasarian, M. M., Schuster, H. and Bagdasarian, M. (1985) In Helinski, D. R., Cohen, S. N., Clewell, D. B., Jackson, D. A. and Hollaender, A. (eds.), Plasmids in Bacteria, Plenum Press, New York, pp. 243-259.
- Meyer, R. J., Lin, L.-S., Kim, K. and Brasch, M. A. (1985) In Helinski, D. R., Cohen, S. N., Clewell, D. B., Jackson, D. A. and Hollaender, A. (eds.), Plasmids in Bacteria, Plenum Press, New York, pp. 173-188.
- Scherzinger, E., Bagdasarian, M. M., Scholz, P., Lurz, R., Rückert, B. and Bagdasarian, M. (1984) Proc. Natl. Acad. Sci. USA 81, 654-658.
- Haring, V., Scholz, P., Scherzinger, E., Frey, J., Derbyshire, K., Hatfull, G., Willetts, N. S. and Bagdasarian, M. (1985) Proc. Natl. Acad. Sci. USA 82, 6090-6094.
- 8. Haring, V. and Scherzinger, E. (1989) In Thomas, C. M. (ed.), Promiscuous Plasmids of Gram Negative Bacteria, Academic Press, London, pp. 95–124.
- Derbyshire, K. M., Hatfull, G. and Willetts, N. S. (1987) Mol. Gen. Genet. 206, 161–168.
- Bagdasarian, M. M., Scholz, P., Frey, J. and Bagdasarian, M. (1987) In Novick, R. and Levy, S. (eds.), Evolution and Environmental Spread of Antibiotic Resistance Genes, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 209-223.
- 11. Honda, Y., Sakai, H., Komano, T. and Bagdasarian, M. (1989) Gene 80, 155-159.
- 12. Diaz, R. and Staudenbauer, W. L. (1982) Nucl. Acids Res. 10, 4687-4702.
- 13. Derbyshire, K. M. and Willetts, N. S. (1987) Mol. Gen. Genet. 206, 154-160.
- 14. Bachman, B. J. (1972) Bacteriol. Rev. 36, 525-557.
- Tippe-Schindler, R., Zahn, G. and Messer, W. (1979) Mol. Gen. Genet. 168, 185-195.
- 16. Carl, P. L. (1970) Mol. Gen. Genet. 109, 107-122.
- Filip, C. C., Allen, J. S., Gustafson, R. A., Allen, R. G. and Walker, J. R. (1974) J. Bacteriol. 119, 443-449.
- Guerry, P., van Embden, J. and Falkow, S. (1974) J. Bacteriol. 117, 619-630.
- 19. Bazaral, M. and Helinski, D. R. (1970) Biochemistry 9, 399-406.
- 20. Uhlin, B. E. and Nordström, K. (1978) Mol. Gen. Genet. 165, 167-179.
- 21. Messing, J. (1983) Methods Enzymol. 101, 20-78.
- Holmes, D. S. and Quigley, M. (1981) Anal. Biochem. 114, 193-197.
 Chase, J. W., Whittier, R. F., Auerbach, J., Sancar, A. and Rupp, W. D. (1980) Nucl. Acids Res. 8, 3215-3227.
- 24. Lanka, E. and Schuster, H. (1983) Nucl. Acids Res. 11, 987–997.
- 25. Hübscher, U. and Kornberg, A. (1980) J. Biol. Chem. 255, 11698–11703.
- 26. Böldicke, T. W., Hillenbrand, G., Lanka, E. and Staudenbauer, W. L. (1981)
- Nucl. Acids Res. 9, 5215-5231.
- Ortega, S., Lanka, E. and Diaz, R. (1986) Nucl. Acids Res. 14, 4865–4879.
 Staudenbauer, W. L. (1976) Mol. Gen. Genet. 145, 273–280.
- Schuster, H., Mikolajczyk, M., Rohrschneider, J. and Geschke, B. (1975) Proc. Natl. Acad. Sci. USA 72, 3907-3911.
- Morris, C. F., Sinha, N. K. and Alberts, B. M. (1975) Proc. Natl. Acad. Sci. USA 72, 4800-4804.
- 31. Fuller, R. S., Funnel, B. E. and Kornberg, A. (1984) Cell 38, 889-900.
- Masai, H., Nomura, N., Kubota, Y. and Arai, K. (1990) J. Biol. Chem. 265, 15124-15133.
- 33. Gellert, M. (1981) Annu. Rev. Biochem. 50, 879-910.
- Chen, H., Lawrence, C. B., Bryan, S. K. and Moses, R. E. (1990) Nucl. Acids Res. 18, 7185-7186.
- Staudenbauer, W. L., Scherzinger, E. and Lanka, E. (1979) Mol. Gen. Genet. 162, 243-249.
- 36. Masai, H. and Arai, K. (1988) J. Biol. Chem. 263, 15016-15023.
- Wickner, S. and Hurwitz, J. (1976) Proc. Natl. Acad. Sci. USA 73, 1053-1057.
- 38. Staudenbauer, W. L. (1977) Mol. Gen. Genet. 156, 27-34.
- 39. Kaguni, J. M., Fuller, R. S. and Kornberg, A. (1982) Nature 296, 623-627.
- Younghusband, H. B. and Inman, R. B. (1974) Annu. Rev. Biochem. 43, 605-618.
- 41. Waddell, J., Wang, X.-M. and Wu, M. (1984) Nucl. Acids Res. 12, 3843-3856.
- Kornberg, A. (1982) 1982 Supplement to DNA Replication, W. H. Freeman and Co., San Francisco.
- 43. Bramhill, D. and Kornberg, A. (1988) Cell 52, 743-755.
- 44. Alfano, C. and McMacken, R. (1988) Nucl. Acids Res. 16, 9611-9630.