JOURNAL OF BACTERIOLOGY, July 1991, p. 3935–3942
0021-9193/91/133935-08$02.00/0

Vol. 173, No. 13

Critical Sequences in the Core of the P1 Plasmid Replication Origin

THERESE BRENDLER,* ANN ABELES, AND STUART AUSTIN

Laboratory of Chromosome Biology, ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, Frederick, Maryland 21702

Received 8 January 1991/Accepted 22 April 1991

The core of the P1 plasmid replication origin consists of a series of 7-bp repeats and a G+C-rich stretch. Methylation of the GATC sequences in the repeats is essential. Forty different single-base mutations in the region were isolated and assayed for origin function. A single-base change within any 7-bp repeat could block the origin, irrespective of whether GATC bases were affected. The repeats themselves were critical, but the short intervals between them were not. Mutations in the G+C-rich region showed it to be a spacer whose exact length is important but whose sequence can vary considerably. It maintains a precise distance between the 7-bp repeats and binding sites for the P1 RepA initiator protein. It may also serve as a clamp to limit strand separation during initiation.

The prophage of bacteriophage P1 is an autonomous plasmid that can be maintained in its *Escherichia coli* host at a copy number as low as one to two per cell (28). In this respect, it mimics the stringent behavior of the host chromosome. The P1 plasmid origin has some structural similarities to *oriC*, the origin of the host chromosome. The study of these features should be informative as to how replication control can achieve accurate maintenance at such a low copy number. The P1 replicon consists of three contiguous elements. These are the origin of replication, *oriR*; the gene for the essential replication protein, *repA*; and the *incA* locus, which controls plasmid copy number (2). The copy control region is not essential to replication, but when it is deleted, the copy number goes up to about 8 to 10 per cell instead of the 1 to 2 found with the wild-type plasmid (27). This *incA* locus consists of a series of 19-bp repeats of a sequence that is a specific binding determinant for the plasmid RepA protein (1).

The origin, *oriR*, is a sequence of approximately 250 bp (Fig. 1) which is sufficient for replication when the P1 RepA protein is supplied in *trans* (11). It contains two DnaA boxes that are binding sites for the DnaA protein essential for P1 origin function (16). Adjacent to these are five tandem repeats of 7 bp. Four of these repeats contain GATC sites that are substrates for the host DNA adenine methylase (17). There is a fifth GATC sequence outside and to the right of the last repeat. Methylation of one or more of the origin GATC sequences is important for function. The origin requires adenine methylation both in vivo and in vitro (3). Adjacent to the *dam* methylation sites is a 39-bp G+C-rich region. Following this are five 19-bp repeats similar to those in *incA*. P1 RepA protein binds to these origin repeats and is essential for initiation at the origin (1). Following these, and outside the boundary of the origin as originally defined, are three more potential DnaA boxes situated within the leader region of the *repA* gene. In at least a superficial sense, the roles of the DnaA boxes and the RepA-binding repeats are clear; they bind their respective proteins, which then become available for some important activities during initiation. However, the function of the central core region that encompasses the 7-bp repeats, GATC sequences, and G+C-rich region is not clear. To probe the significance of these sequences, we isolated a large number of mutants of the region and tested their phenotypes.

MATERIALS AND METHODS

Buffers, media, enzymes, and reagents. Conditions for cell growth and buffers used for electrophoresis and DNA preparation were previously described (3, 5, 20). Sequenase modified T7 DNA polymerase was supplied by U.S. Biochemical Corp. (Cleveland, Ohio) (30). All other enzymes were obtained from Bethesda Research Laboratories, Inc. (Gaithersburg, Md.), New England BioLabs, Inc. (Beverly, Mass.), Boehringer Mannheim Biochemicals (Indianapolis, Ind.), and Pharmacia LKB Biotechnology, Inc. (Piscataway, N.J.). Details of other reagents and materials were as previously described (1, 3, 5).

Bacterial strains. *E. coli* DH5 (15) was used for DNA cloning and plasmid preparation. P1 *oriR* function was tested in CC708, a *polA12*(Ts) strain (MM383; 24) transformed with plasmid pALA619 that produces the P1 RepA protein.

DNA preparation and sequencing. Plasmid DNA was prepared as described previously (20). DNA sequencing templates were prepared from supercoiled DNA by the method of Chen and Seeburg (12). DNA was sequenced by the dideoxy method, using Sequenase modified T7 DNA polymerase (30). Deoxyoligonucleotides were synthesized on an Applied Biosystems DNA synthesizer (model 380B), using the phosphite triester method (9, 21). The oligonucleotides were purified by Sephadex G-25 chromatography and kinase treated as described by Abeles et al. (5). Complementary oligonucleotides were annealed under the conditions described previously (37).

Nomenclature of mutations. Mutations are named for the change involved and the coordinate of the change on the standard P1 plasmid maintenance region map (2). Thus, mutation A→T,420 has the A at coordinate 420 changed to a T, and ΔT,450 has the T at coordinate 450 deleted. Larger deletions and insertions follow a similar scheme. Thus, Δ480-489 has the indicated 9 bp (inclusive) deleted and +GCTCCATGGC,484 has these 10 bp inserted between base 484 and the preceding base.

Plasmids and cloning vectors. Plasmids pUC18 (39), pUC19 (39), and pALA619 (6), which contains the gene for the P1 RepA protein in a pSC101-based vector, were previously described. The P1 minimal-origin clones were constructed as

* Corresponding author.

follows. Plasmid pALA621 was derived from the hydroxylamine-induced mutation *rep-30*, which changes C to T at P1 base 457 (8). For the purpose of uniformity, this mutation is referred to as C→T,457 in this report. A *Hin*dIII-to-*Eco*RI fragment containing the mutant origin was excised from M13-P1*ori*-#88 (3) and inserted between the corresponding sites of pUC18. All other plasmids were derived from the P1 minimal-origin clone pALA618, a pUC19 derivative (6). Plasmid pALA626 was constructed from a specific double-stranded oligonucleotide spanning a 66-bp region between two adjacent *Ava*II sites of pALA618 (pBR322 bp 3729; P1 bp 424). This created a *Bgl*II site by changing P1 bp 414 from a C to a T in the first 7-bp tandem repeat. Plasmids pALA630 and pALA631 were constructed by inserting the P1 *oriR* sequence from bp 386 to 502 between the *Hin*dIII (pUC19) site and *Nru*I (P1 bp 502) site of pALA618, using a double-stranded oligonucleotide with a *Hin*dIII site at the P1 bp 386 end. Plasmid pALA631 differs from pALA630 in having a *Ssp*I site due to an A-to-T mutation at bp 419 between the first and second 7-bp tandem repeats. Plasmid pALA632 was constructed by replacing the 117-bp P1 *oriR* region of pALA630 between *Hin*dIII (P1 bp 386) and *Nru*I (P1 bp 502) with a 40-bp oligonucleotide containing both *dnaA* boxes and the first 7-bp tandem repeat mutated to a *Bgl*II site (C→T,414). The purpose of this construct was to facilitate mutagenesis with oligonucleotides containing random base changes (see below). The first 7-bp tandem repeat could be regenerated by ligation of the *Bgl*II 5′ half site of the vector to an oligonucleotide containing a *Bam*HI 3′ half site. Plasmid pALA1204 was generated from a double-stranded oligonucleotide encompassing P1 *oriR* sequences 410 to 502, which replaced the sequences between the *Bgl*II and *Nru*I site of pALA632. Plasmid pALA1204 has a *Cla*I site due to a T-to-A substitution at bp 466. The mutant behaves essentially as the wild-type pALA630 in all assays (see Results). Spacing mutations pALA1209 to pALA1221 were derived from the insertion of synthetic oligonucleotides between the new *Cla*I (P1 bp 464) and *Nru*I (P1 bp 502) sites of pALA1204 and revert the silent mutation at bp 466 to wild type. Saturation mutagenesis of the first 7-bp tandem repeats was accomplished by insertion of synthetic oligonucleotides between the *Hin*dIII (P1 bp 386) and *Cla*I (P1 bp 464) sites of pALA1204. All of the resulting clones retain the silent mutation at bp 466.

Random mutagenesis of the P1 minimal origin with synthetic oligonucleotides. The region between P1 bp 414 and 500 was targeted for random mutagenesis. This region spans four of the five 7-bp tandem repeats, four of the five *dam* sites, and a G+C-rich region (Fig. 1). Random base changes throughout the region of interest were accomplished through the chemical synthesis of a mixed population of oligonucleotides which contain a low level of contamination by the other three nucleotides at each position (18, 26). The level of contaminating nucleotides was calculated from the binomial distribution derived by McNeil and Smith (22) and chosen to be one hit per clone. The resulting oligonucleotides of 116 bases contained the targeted region of 85 bp flanked by a 5′ *Bam*HI site and a 3′ *Nru*I site. Distal to the *Nru*I site was a sequence complementary to the universal M13 17-bp primer (23). Priming, synthesis of the complementary strand, and cloning of the resultant *Bam*HI-*Nru*I fragments were done as described previously (18, 26, 29). The fragments were inserted between the *Bgl*II and *Nru*I ends of pALA632.

Screening of clones from mutagenesis with synthetic oligonucleotides. Candidate transformants were replica plated onto nitrocellulose filters and hybridized to P1 *ori* region probes as described previously (31, 32). Colonies with inserts were initially identified by using an oligonucleotide probe spanning the entire P1 *ori* region targeted for mutagenesis. Using a series of seven overlapping oligonucleotide probes of 16 to 19 bases in length, clones likely to contain inserts with single-base changes were identified by using stringent hybridization conditions to detect single-base-pair mismatches (33, 34). This also allowed us to determine in which short interval the putative mutation was located. Hybridizations were performed 5°C below the theoretical T_m (33) in 6× SSPE (1× SSPE is 0.18 M NaCl, 0.01 M sodium phosphate [pH 7.7], and 0.001 M EDTA). Washes were done at 2 to 5°C below the T_m in 1× to 2× SSPE.

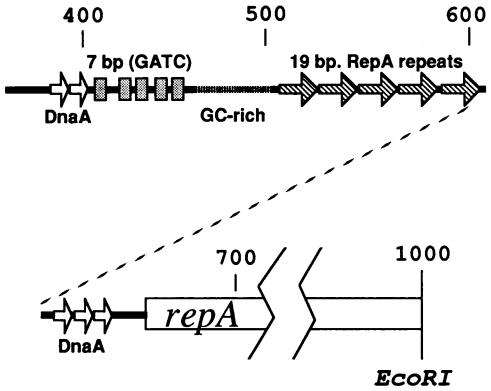

FIG. 1. P1 plasmid replication origin and associated sequences. Stippled boxes are the five 7-bp imperfect repeats. All but the second box contain GATC adenine methylation sites. Open arrows are potential DnaA boxes. Shaded arrows are 19-bp RepA-binding repeats. Vertical shading distinguishes the 39-bp G+C-rich sequence. The upper part of the map represents the minimal origin as originally defined (11); the lower part shows associated sequences including the rightward set of DnaA boxes and the start of the *repA* gene. Coordinates are in standard base pair units (2).

In vivo assay for P1 *oriR* function. The in vivo replication assay was essentially performed as described by Abeles et al. (6). The pUC19-derived clones were introduced into strain CC708 by transformation using the calcium chloride method (20). Strain CC708 is *polA*(Ts) and supplies the P1 RepA protein in *trans* from a compatible plasmid, pALA619 (6). Relative transformation frequencies were determined by selection of ampicillin-resistant transformants at 30 or 42°C, counting the resulting colonies after 24 h.

For maintenance assays, transformants were isolated on selective medium at 30°C, at which the pUC19 origin is functional (7, 19). After the cultures were grown in selective medium at 30°C to an optical density at 600 nm of 0.2, the cultures were diluted into nonselective medium and grown at 42°C to block pUC19 origin function (19). Periodically, aliquots were withdrawn during growth at 42°C and plasmid retention was determined as the ratio of colonies formed at 30°C on agar with and without selective antibiotic.

Graphical treatment of data. Data were tabulated and plotted by using the Cricket Graph program (Cricket Software, Malvern, Pa.). Using the same software, the slopes of the curves $\log_2$ [plasmid retention] versus number of generations were calculated by regression. The zero generation point was ignored for this purpose to avoid the initial shoulder present on most of these curves.

TABLE 1. Single-base mutations[a]

Plasmid	Mutation	42/30°C[b]	Slope[c]	Class
pALA630	Wild type	0.7	−0.16	1
pUC19	Vector	0.0	−0.87	3
pALA1228	A→T,409	0.8[d]	−0.46	2
pALA1229	A→G,409	0.0	−0.70	3
pALA1230	G→A,410	0.0	−0.68	3
pALA1231	G→C,410	0.0	−0.91	3
pALA1232	A→T,411	0.6[d]	−0.45	2
pALA1233	A→C,411	0.0	−0.85	3
pALA1234	T→A,412	0.4[d]	−0.49	2
pALA1235	T→G,412	0.0	−0.78	3
pALA1236	T→C,412	0.0	−0.71	3
pALA1237	C→A,413	0.0	−1.03	3
pALA1238	C→G,414	0.0	−0.99	3
pALA626	C→T,414	0.0	−0.68	3
pALA1239	A→G,415	1.0	−0.26	1
pALA631	A→T,419	0.9	−0.13	1
pALA1243	C→T,427	0.2[d]	−0.47	2
pALA1244	C→A,427	0.6	−0.26	1
pALA1245	A→T,427	1.0	−0.07	1
pALA1246	C→G,436	0.0	−0.78	3
pALA1248	T→G,441	0.9	−0.10	1
pALA1249	A→G,445	0.6[d]	−0.43	2
pALA1250	A→C,453	0.0	−0.73	3
pALA1251	G→T,454	0.0	−0.82	3
pALA621	C→T,457	0.0	−0.90	3
pALA1252	T→A,463	0.8	−0.26	1
pALA1253	C→A,464	0.9	−0.31	1
pALA1204	T→A,466	0.9	−0.14	1
pALA1254	C→A,474	1.0	−0.28	1
pALA1255	G→C,486	1.0	−0.22	1
pALA1256	T→G,489	0.8	−0.13	1
pALA1257	G→C,492	0.7	−0.12	1
pALA1258	A→G,493	1.0	−0.26	1
pALA1259	G→A,495	1.0	−0.28	1
pALA1260	ΔA,416	0.9	−0.25	1
pALA1262	ΔC,421	1.0	−0.04	1
pALA1263	ΔT,435	0.6[d]	−0.38	2
pALA1264	ΔT,450	0.8[d]	−0.33	2
pALA1265	ΔT,466	0.0	−0.53	3
pALA1266	ΔC,474	0.0	−0.71	3

[a] Mutations are in map order, with the deletion mutations listed separately.
[b] Ratio of transformation frequencies 42 and 30°C.
[c] Values of the slopes of the graphs of $\log_2$ [plasmid retention] versus number of generations (see Materials and Methods). The classes are derived from these data (see text and Fig. 3). The phenotype of each mutant was confirmed in duplicate experiments (data not shown).
[d] Minute colonies at 42°C.

RESULTS

Assay of origin function. We have developed an assay for the P1 origin that uses pUC19 derivatives that contain it. An isolated P1 origin promotes replication when the P1 RepA protein is supplied in *trans* (11). We introduce the pUC19 derivatives into a strain in which DNA polymerase I is temperature sensitive and the RepA protein is supplied [CC708 *polA12*(Ts)]. At 30°C, pUC19 and all of its derivatives transform the strain to ampicillin resistance, as the pUC19 origin is active (Table 1). At 42°C, pUC19 gives no measurable transformation. However, derivatives such as pALA630, which contains an active P1 origin in addition to that of pUC19, do transform efficiently at 42°C, reflecting that the P1 origin does not require DNA polymerase I (2). The ability to transform this strain at 42°C can be used as a qualitative test for P1 origin function, indicating whether the origin is competent for plasmid establishment. A refinement of the *polA*(Ts) test allows quantitation by assaying plasmid maintenance. Transformants are isolated and purified at 30°C, at which the origin of pUC19 is functional. After a period of growth at 30°C in medium containing ampicillin that selects for the retention of the test plasmid, the culture is diluted into antibiotic-free medium at 42°C. The proportion of cells retaining the plasmid is then determined at various points during exponential growth (see Materials and Methods).

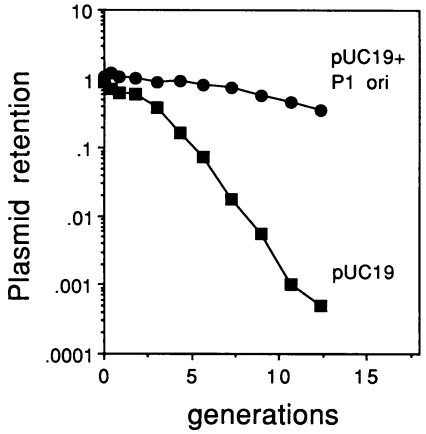

FIG. 2. Quantitative assay for origin function. The curves represent the proportion of cells retaining the plasmid in nonselective culture after a shift to 42°C as a function of the number of generations of growth without selection. Results with the wild-type P1 origin plasmid pALA630 and the vector pUC19 are shown. In the latter case, the ultimate rate of loss is as predicted for a plasmid that does not replicate at all (the slope of $\log_2$ [retention] versus number of generations = −1). The equivalent slope of the curve for pALA630 is −0.17.

When pUC19 was used, the plasmid was rapidly lost from the population at 42°C (Fig. 2). The starting population already contained cured cells. This reflects the low copy number of plasmids driven by the pUC origin in the *polA12*(Ts) strain at 30°C coupled with the random distribution of the copies to daughter cells (7). After an initial plateau, the loss rate curve of growth versus $\log_2$ ampicillin-resistant cells/total cells is linear, with a slope of about −1.0 (see Materials and Methods). Thus, after an initial increase, the total number of plasmid-containing cells in the culture is constant, showing that pUC19 replication is blocked. Presumably, the early kinetics of loss reflect an initial segregation of the plasmid to the point where no cell contains more than one plasmid. From this point, a fixed number of plasmid-containing cells are simply diluted by growth.

When the pUC19 clone containing the wild-type P1 *ori* was used (pALA630), the starting population had no measurable plasmid-free cells and continued to maintain the plasmid for some considerable time at 42°C without selection (Fig. 2). The same plasmid behaved like its pUC19 parent if the RepA protein was not supplied (data not shown), showing that the improved maintenance is due to P1 origin function. The gradual loss of the P1-driven plasmid without selection is presumably due to the random distribution of the

plasmid at cell division, as the clones have no plasmid partition system (4). After an initial dilution of any excess copies present at zero time, the rate of loss ($\log_2$ slope = −0.17) should be a function of the average number of copies per dividing cell produced by the P1 origin and the distribution of copy number in individual cells. Random partitioning would give the observed rate of loss if each dividing cell had about eight copies (25). In practice, the real average copy number must be higher than this since some variation of copy number must exist from cell to cell. Mutant origins with impaired function are expected to give lower average copy numbers and should segregate plasmid-free cells more rapidly. In extreme cases, origin-defective mutants should behave like the pUC19 control, whereas silent mutants should behave like pALA630.

Mutagenesis of the origin core. The 85-bp origin core (Fig. 1) was subjected to mutagenesis by the introduction of random bases in a synthetic oligonucleotide. The synthesis was designed to give an average of one base change per molecule. After synthesis of a complementary DNA strand, the double-stranded oligonucleotides were ligated into a pUC19-based plasmid in such a way as to restore a complete P1 plasmid origin sequence with the random mutations in place (see Materials and Methods). The DNA content of cells transformed with the ligated products was screened by filter hybridization using labeled synthetic oligonucleotides corresponding to short intervals of the wild-type region as hybridization probes. Conditions were established such that any mismatches to the probe inhibit hybridization (see Materials and Methods). Using the resulting data, we judged that 90% of the recovered clones contained an oligonucleotide insert. Of these, 28% appeared to be wild type in sequence, 36% showed evidence of mutations in more than one interval, and 36% appeared to have changes in only one of the short intervals. Cells corresponding to 50 of this latter class were cultured, plasmid DNA was prepared from them, and the DNA sequence of the origin region was determined for each.

DNA sequences of the mutant candidates. Although the hybridization screen was designed to eliminate mutants with complex changes, it was not completely effective, as some of the candidates showed multiple changes, including evidence of imprecise ligation of the oligonucleotide ends resulting in deletions. However, 16 mutants had the desired single-base changes in an otherwise wild-type origin sequence. In addition, six mutants had unique single-base deletions. Mutants of this latter class were unexpected but were nevertheless useful. We assume that they were the result of errors by the oligonucleotide synthesizer. The 22 mutants with single-base changes or deletions were used for further study, along with an additional three single-base-change mutants made by base-specific oligonucleotide-directed mutagenesis and a previously isolated hydroxylamine-induced single-base-change mutant (pALA621 C→T,457) (see Materials and Methods). The properties of the resulting 26 mutants are discussed below.

Mutations in the first 7-bp repeat. The strategy used to isolate random point mutants precluded isolation of mutations in all but the last base of the first 7-bp repeat (A,415). None of the random mutations isolated affected this base. One mutant (pALA626, C→T,414) was isolated by insertion of a specific 66-bp region constructed from oligonucleotides. It differs from other origin constructs in having 20 additional P1 bases to the left of the DnaA boxes at bp 386. To expand the number of mutations affecting this repeat, we created 12 additional mutants by replacing the relevant region of pALA1204 with specific oligonucleotides containing the mutant sequence. Unlike the other origin constructs, these 12 all retain the silent mutation T→A,466 of pALA1204 which created the *Cla*I site used in their construction. The DNA sequence of the mutants was confirmed.

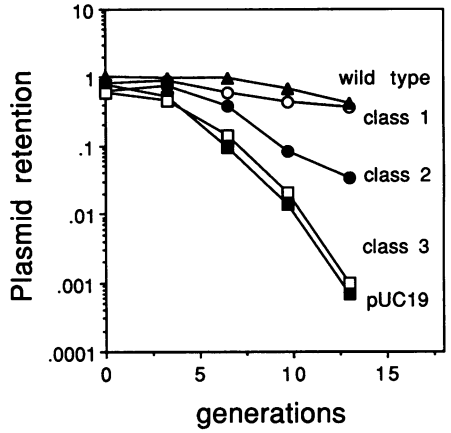

FIG. 3. Maintenance stability of typical mutants. The maintenance curves are typical for class 1, 2, and 3 origin mutations. The assay is the same as used for Fig. 2. The control curves for pALA630 (wild type) and the pUC19 vector were obtained in the same experiment and are comparable to those obtained in Fig. 2. The examples given are mutations A→T,419; C→T,427; and C→T,457, respectively.

Phenotypes of the mutant origin plasmids. Table 1 shows the relative transformation frequencies of the mutant origins at 30 and 42°C. All mutants transformed normally at 30°C, showing that they are not incapable of transformation and do not, for example, kill the cell by runaway replication. At 42°C, the mutants fall into three classes: those transforming well, like the wild type; those giving no transformants; and an intermediate class. The intermediate class gave normal or somewhat reduced frequencies of transformation at 42°C, but the colonies obtained were only pinpoint in size after overnight incubation, as opposed to the wild-type colonies that are approximately 2 mm in diameter. Each mutant was subjected to the maintenance test described above. Loss of each plasmid was determined at five different time points during unselected growth at 42°C (Fig. 3). Inspection of the total data showed a good correlation between the rates at which the plasmids were lost during growth and the transformation phenotypes described above. As reproduction of the graphical data for all of the mutants would be burdensome, we have tabulated estimates of the slopes of these loss curves determined as described in Materials and Methods (Table 1). We note that these values are determined from a minimal number of time points for each mutant (see Materials and Methods) and therefore are only approximations of the steady-state loss rates. Nevertheless, the values obtained are completely rational with respect to the transformation phenotypes, with the normally transforming mutants all giving slope values of greater than −0.29 and nontransforming mutants giving lower values approaching that of pUC19 (<−0.52). The seven mutants that transform, but give small colonies each have intermediate values (−0.49 to −0.33). We conclude that the ability to transform at 42°C and the maintenance fidelity of the mutants show a good correlation. Thus, both phenotypes appear to reflect origin function, and classification into the three basic groups given by

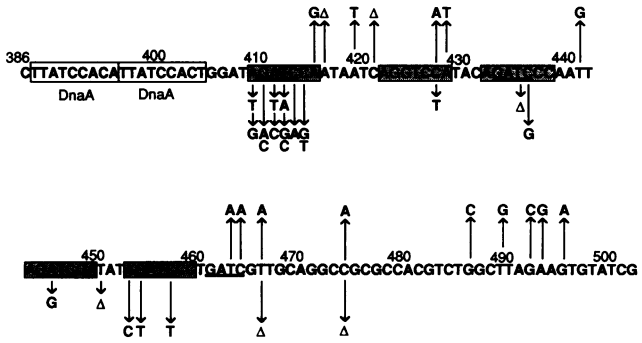

FIG. 4. Mutations of the 7-bp repeat and G+C-rich regions. The DnaA boxes (open boxed regions) and 7-bp repeats (stippled boxes) are marked. The GATC methylation sites are underlined. Arrows show the single-base-change mutations, their size, and their direction, indicating which of the three phenotypic classes they fall into as determined from the data in Table 1. Upward arrows indicate class 1 mutations (little or no effect on origin function; see Fig. 3, for example). Full down arrows indicate mutations that block origin function (class 3 mutations; Fig. 3). Half down arrows mark mutations that have partial function (class 2 mutations; Fig. 3). The base altered in each mutant is shown. Δ indicates the missing base in single-base-deletion mutations.

the transformation assays seems justified. Figure 4 shows the base change and classification of each mutant according to this scheme.

Each 7-bp repeat is important for origin function. Inspection of Fig. 4 shows that the 7-bp repeats are sensitive to mutation. For each repeat we found at least one single-base change that had a deleterious effect on replication. Thus, all of the repeats appear to play a role, including the atypical one that lacks the GATC methylation site. The phenotypes of the comprehensive set of mutations in the first 7-bp repeat show that 6 of the 7 bp are important for origin function. This may also be true for all of the repeats, as seen in the pattern of mutations in other boxes; mutations within the first 6 bp are, with a single exception, deleterious, whereas changes in the last position have little or no effect (Fig. 4). Deleterious mutations include, but are not limited to, the GATC sequences themselves. We conclude that although the GATC sequences themselves are important, it is the integrity of the short repeat motif consisting of at least 6 of the 7 bp that is the critical determinant of functionality of this region. The fifth GATC sequence that is not associated with a short repeat appears to be gratuitous. The second 7-bp repeat that lacks a GATC sequence tolerates a change at the sixth base (C→A,427). However, the sequence of this repeat is important because another change at the same position (C→T,427) is deleterious.

Spacing between the 7-bp repeats. A number of mutations in the 3- to 6-bp spacers between the 7-bp repeats give no phenotype. These include single-base deletions between boxes 1 and 2. The exception is mutant ΔT,450 (pALA1264), which reduces the short spacer region between boxes 4 and 5 to just 2 bp. This suggests that a minimal spacer length may be needed but that otherwise neither the sequence nor length of these short spacers is critical.

The G+C-rich region to the right of the short repeats. Of the nine base change mutations that lie in the G+C-rich region, none have a readily measurable effect on origin function. Several of these change a C or G to an A, showing that not all of the G and C bases are critical. In contrast, the two single-base deletions (ΔT,466 [pALA1265] and ΔC,474 [pALA1266]) both have drastic effects, blocking origin function. This suggests that the G+C-rich region is a spacer whose length is more critical than its base sequence.

Mutations that vary the length of the G+C-rich spacer. Using defined oligonucleotides to replace the G+C-rich region with altered sequences, we isolated and sequenced a number of additional mutants with variations in the length or composition of this region. Table 2 shows the sequences of these mutants and the results obtained. None of the mutants with altered spacer lengths were able to transform the *polA*(Ts) strain at 42°C, showing that the length can be neither increased nor decreased without deleterious effects. The slopes of the curves from maintenance experiments confirm this finding (Table 2). A mutation that adds a single

TABLE 2. G+C-rich spacer mutations

Plasmid	Spacing change[a]	Mutation	42/30°C[b]	Slope[c]	Class
pALA1209	−1	ΔC,484	0.0	−1.0	3
pALA1211	+1	+G,484	0.0	−0.68	3
pALA1212	+1/−1	+G,484;ΔA,494	1.0	−0.15	1
pALA1213	+5	+TGGAT,481	0.0	−0.78	3
pALA1214	−5	Δ481–484;Δ488	0.0	−1.0	3
pALA1215	+9	+GCTCCATGG,484	0.0	−0.89	3
pALA1216	−9	Δ481–484;Δ488–492	0.0	−0.98	3
pALA1217	+10	+GCTCCATGGC,484	0.0	−1.1	3
pALA1218	−10	Δ481–484;Δ488–493	0.0	−0.96	3
pALA1219	−10	Δ480–489	0.0	−0.86	3
pALA1220	G/C→A/T	—[d]	0.0	−1.0	3
pALA1221	A/T→G/C	—[d]	1.0	−0.35	1 or 2[e]

[a] Number of bases added (+) or removed (−) from the wild-type spacer region. Note that some of the changes involve two mutations.

[b] Ratio of transformation frequencies of strain CC708 at 42 and 30°C.

[c] Values of the slopes of the graphs of $\log_2$ [plasmid retention] versus number of generations (see Materials and Methods). The class designations are derived from these data (see text and Fig. 3).

[d] Has gross changes in the 39-bp spacer region as follows: all G and C bases changed to A and T, respectively (pALA1220); all A and T bases changed to G and C, respectively (pALA1221).

[e] Like class 1 mutants, pALA1221 transforms with high efficiency at 42°C and gives normal-size colonies. However, the slope value (−0.35) falls within the range normally seen for class 2 mutants.

base (+G,484; pALA1211) is defective, but its derivative, which restores the spacer length by deleting a base elsewhere (+G,484;ΔA,494 [pALA1212]), transforms normally and is maintained similarly to the wild type. The data show no evidence that an increase or decrease of the length of the region by 10 bp, which corresponds approximately to one turn of the helix, can be tolerated. We conclude that the region is a spacer of critical length.

Base composition of the spacer region. As the minor changes in base sequence of the spacer region had little effect, it seemed possible that the sequence of the spacer, as opposed to its length, was irrelevant or that only the overall G+C-rich nature of the sequence was important. A mutant was therefore constructed in which all of the spacer GC base pairs were replaced by AT pairs (pALA1220) and another in which all of the AT base pairs were replaced by GC pairs (pALA1221; Table 2). Both mutant regions have the same number of total base pairs as the equivalent wild-type spacer. Table 2 shows that the all-GC spacer, although perhaps not as effective as the wild type, is functional. The all-AT version is not. Thus, much of the base sequence information of the spacer is relatively unimportant. Rather, it seems likely that the G+C-rich composition is the more critical feature.

DISCUSSION

Bramhill and Kornberg (10) have proposed a general model for the events that initiate DNA synthesis from a class of origins typified by the *E. coli* host origin *oriC*. Two principal elements are involved. The first consists of a series of repeats of a binding site for an initiator protein. Binding of the protein causes the DNA to be bent and wrapped around a core of the protein. The second element is adjacent to this region. This region responds to the distortion or the protein bound to the first region by a specific strand-opening event. The single-stranded sequences of this second region are modified by the binding of several host proteins to form a pre-primer complex that will eventually lead to primer synthesis and DNA replication. The proposed regions of strand opening are generally A+T rich (although the P1 region is an exception) and also contain repeat sequences (10). In the P1 origin, this region would be the sequence containing the 7-bp repeats in the origin core (10). The function of these repeated motifs is not yet clear. In the case of *oriC*, the initiator protein is the host DnaA protein (14), which binds to four repeats of the sequence known as the DnaA box (13). The second region contains 13-bp A+T-rich repeats that have been shown to be the sites of strand opening.

For a class of plasmid origins that show similarities to *oriC*, the proposed initiator protein is a plasmid-encoded product that binds to plasmid-specific repeats within the origin. In some cases, however, including that of P1, DnaA boxes are also present, and both the plasmid-specific and DnaA proteins are required for initiation (36). Here it is proposed that the plasmid-specific initiator protein triggers strand opening and the DnaA protein acts subsequently, perhaps during pre-primer formation (10). This concept has received support from recent studies on the P1 origin (6, 35). DnaA boxes were shown to be essential components of the P1 origin. However, a single box was sufficient and could be placed either upstream or downstream of the origin core (6). This virtually rules out the obligate involvement of the P1 DnaA boxes in a complex, coiled structure that initiates a precise strand opening as is proposed for the boxes in *oriC*.

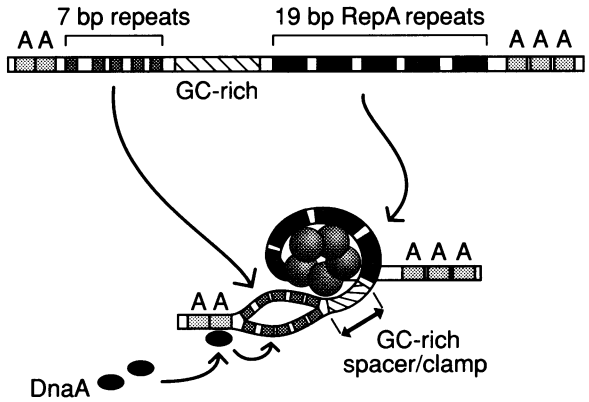

FIG. 5. Model for early steps in initiation at the P1 plasmid origin. This model is an adaption of the general model of Bramhill and Kornberg (10). DnaA boxes (A), the 7-bp repeats, the G+C-rich spacer, and the 19-bp RepA-binding repeats are shown as boxes distinguished by different shadings (top line). Binding of the P1 RepA initiator protein (shaded spheres) to the 19-bp repeat region bends the DNA, wrapping it around a protein core (below). Specific contacts between the wrapped complex and the 7-bp repeat region cause strand melting. This is dependent on the precise relative positioning of the two regions defined by the G+C-rich spacer. The stability of the G+C-rich spacer duplex prevents strand opening from spreading back to destroy the wrapped complex. The DnaA boxes are not an integral part of the complex but serve to capture DnaA protein (shaded spheroids) for delivery to the melted region, where it plays a role in pre-primer formation on the exposed single strands.

Thus, as proposed by the general model, the initiator protein is probably the P1 RepA protein that binds to five closely spaced repeats of a 19-bp recognition sequence to the right of the origin core (Fig. 5). The region of strand opening would correspond to the origin core itself, with its 7-bp repeats and adenine methylation sites.

The Bramhill and Kornberg model, as adapted to the P1 case, is illustrated in Fig. 5. Our results allow some refinement of the model. First, elements of the region include a spacer of precise length. Single-base-pair deletions in the G+C-rich region block origin function, whereas two comparable mutations that shorten the distance between 7-bp repeats do not. Thus, it is the length of the interval between the 19-bp RepA repeats on the right and the 7-bp repeats on the left that must be maintained. We propose that this spacer positions the two regions correctly for a physical interaction (Fig. 5). The spacer is G+C rich overall (62%) and has a stretch of nine consecutive GC base pairs. A substitute spacer consisting of 100% GC base pairs is functional. The double-helix structure of the wild type should be relatively stable, and the functional mutant spacer, with its 39 consecutive GC base pairs, should be exceptionally resistant to melting. This suggests that melting of the 7-bp region to the left may not be caused by direct propagation of strand opening from the RepA-binding region on the right. Rather, we favor the idea that contact between the RepA-DNA complex and the 7-bp repeats may cause a local strand opening that leaves the G+C-rich spacer unmelted (Fig. 5). The functions of the spacer could be both to position the strand-melting contact correctly and to limit the propagation of the strand opening to the specific region required for pre-primer formation. This clamp effect might also be responsible for directing the replication fork which proceeds unidirectionally to the left (36).

Whatever the precise role of the 7-bp sequences, it is sensitive to the methylation state of the repeats. The origin requires *dam* methylation both in vivo and in vitro (3). From in vitro experiments, in which the DNA templates contain no other GATC sequences except those present in the origin, a requirement for origin methylation could be demonstrated directly. We show above that all of the mutations affecting the four GATC sites within 7-bp repeats are deleterious. Thus, it is likely that methylated GATC is an important determinant at each of these four positions. Methylation of the fifth GATC that is not associated with a 7-bp repeat does not appear to be important, as the site can be mutated to GAAC or GATA without significant effect. Adenine methylation has been implicated in helix destabilization (38), so its role in the 7-bp repeats may be to facilitate strand opening. However, the 7-bp repeats presumably play a role in addition to that of serving as substrates for methylation, as mutations in the non-GATC bases of the repeats often have drastic effects. In addition, the second repeat, which naturally lacks a GATC sequence, is important for origin function.

The function of the 7-bp repeats themselves remains completely conjectural. However, two possibilities come to mind. Appropriate contact with the RepA protein-DNA complex could conceivably require a repetitive motif in the 7-bp region so that each bound unit of RepA can contact one 7-bp sequence. Alternatively, some other protein might recognize the 7-bp repeats, either before or after strand melting. The repetitive nature of the region could then reflect a need for binding multiple copies of this protein.

ACKNOWLEDGMENTS

We thank Lucretia Reaves for invaluable technical assistance and Julie Ratliff for preparation of the manuscript.

This research was sponsored by the National Cancer Institute under contract NO1-CO-74101 with ABL.

REFERENCES

1. **Abeles, A.** 1986. P1 plasmid replication: purification and DNA-binding activity of the replication protein RepA. J. Biol. Chem. **261**:3548–3555.
2. **Abeles, A., K. Snyder, and D. Chattoraj.** 1984. P1 plasmid replication: replicon structure. J. Mol. Biol. **173**:307–324.
3. **Abeles, A. L., and S. J. Austin.** 1987. P1 plasmid replication requires methylated DNA. EMBO J. **6**:3185–3189.
4. **Abeles, A. L., S. A. Friedman, and S. J. Austin.** 1985. Partition of unit-copy miniplasmids to daughter cells. III. The DNA sequence and functional organization of the P1 partition region. J. Mol. Biol. **185**:261–272.
5. **Abeles, A. L., L. D. Reaves, and S. J. Austin.** 1989. Protein-DNA interactions in the regulation of P1 plasmid replication. J. Bacteriol. **171**:43–52.
6. **Abeles, A. L., L. D. Reaves, and S. J. Austin.** 1990. A single dnaA box is sufficient for initiation from the P1 plasmid origin. J. Bacteriol. **172**:4386–4391.
7. **Austin, S., S. Friedman, and D. Ludtke.** 1986. Partition functions of unit-copy plasmids can stabilize the maintenance of plasmid pBR322 at low copy number. J. Bacteriol. **168**:1010–1013.
8. **Austin, S. J., R. J. Mural, D. K. Chattoraj, and A. L. Abeles.** 1985. *Trans*- and *cis*-acting elements for the replication of P1 miniplasmids. J. Mol. Biol. **183**:195–202.
9. **Beaucage, S. L., and M. H. Caruthers.** 1981. Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis. Tetrahedron Lett. **22**:1859–1862.
10. **Bramhill, D., and A. Kornberg.** 1988. A model for initiation at origins of DNA replication. Cell **54**:915–918.
11. **Chattoraj, D. K., K. M. Snyder, and A. Abeles.** 1985. P1 plasmid replication: multiple functions of RepA protein at the origin. Proc. Natl. Acad. Sci. USA **82**:2588–2592.
12. **Chen, E. Y., and P. H. Seeburg.** 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA **4**:165–170.
13. **Fuller, R. S., B. Funnell, and A. Kornberg.** 1984. The *dnaA* protein complex with the *E. coli* chromosome replication origin (*oriC*) and other DNA sites. Cell **38**:889–900.
14. **Fuller, R. S., and A. Kornberg.** 1983. Purified *dnaA* protein in initiation of replication at the *Escherichia coli* chromosomal origin of replication. Proc. Natl. Acad. Sci. USA **80**:5817–5821.
15. **Hanahan, D.** 1985. Techniques for transformation of *E. coli*, p. 109–135. *In* D. M. Glover (ed.), DNA cloning: a practical approach. IRL Press Lt., Oxford.
16. **Hansen, E. B., and M. Yarmolinsky.** 1986. Host participation in plasmid maintenance: dependence upon *dnaA* of replicons derived from P1 and F. Proc. Natl. Acad. Sci. USA **83**:4423–4427.
17. **Hattman, S., J. E. Brooks, and M. Masurekar.** 1978. Sequence specificity of the P1 modification methylase (M. *Eco* PI) and the DNA methylase (M. *Eco dam*) controlled by the *Escherichia coli dam* gene. J. Mol. Biol. **126**:367–380.
18. **Hill, D. E., A. R. Oliphant, and K. Struhl.** 1987. Mutagenesis with degenerate oligonucleotides: an efficient method for saturating a defined DNA region with base pair substitutions. Methods Enzymol. **155**:558–568.
19. **Kingsbury, D., and D. Helinski.** 1970. DNA polymerase as a requirement for the maintenance of the bacterial plasmid colicinogenic factor E1. Biochem. Biophys. Res. Commun. **41**: 1538–1544.
20. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
21. **Matteucci, M. D., and M. H. Caruthers.** 1981. Synthesis of deoxyoligonucleotides on a polymer support. J. Am. Chem. Soc. **103**:3185–3193.
22. **McNeil, J. B., and M. Smith.** 1985. *Saccharomyces cerevisiae CYC1* mRNA 5′-end positioning: analysis by in vitro mutagenesis, using synthetic duplexes with random mismatch base pairs. Mol. Cell. Biol. **5**:3545–3551.
23. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
24. **Monk, M., and J. Kinross.** 1972. Conditional lethality of *recA* and *recB* derivatives of a strain of *Escherichia coli* K-12 with a temperature-sensitive deoxyribonucleic acid polymerase I. Int. J. Syst. Bacteriol. **109**:971–978.
25. **Nordström, K., and S. Austin.** 1989. Mechanisms that contribute to the stable segregation of plasmids. Annu. Rev. Genet. **23**:37–69.
26. **Oliphant, A. R., A. L. Nussbaum, and K. Struhl.** 1986. Cloning of random-sequence oligodeoxynucleotides. Gene **44**:177–183.
27. **Pal, S., R. J. Mason, and D. K. Chattoraj.** 1986. Role of initiator titration in copy number control. J. Mol. Biol. **192**:275–285.
28. **Prentki, P., M. Chandler, and L. Caro.** 1977. Replication of the prophage P1 during the cell cycle of *Escherichia coli*. Mol. Gen. Genet. **152**:71–76.
29. **Schneider, T. D., and G. D. Stormo.** 1989. Excess information at bacteriophage T7 genomic promoters detected by a random cloning technique. Nucleic Acids Res. **17**:659–674.
30. **Tabor, S., and C. C. Richardson.** 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84**:4767–4771.
31. **Vogeli, G., and P. S. Kaytes.** 1987. Amplification, storage, and replication of libraries. Methods Enzymol. **152**:407–415.
32. **Wahl, G. M., and S. L. Berger.** 1987. Screening colonies or plaques with radioactive nucleic acid probes. Methods Enzymol. **152**:415–423.
33. **Wallace, B. R., and C. G. Miyada.** 1987. Oligonucleotide probes for the screening of recombinant DNA libraries. Methods Enzymol. **52**:432–442.
34. **Wallace, B. R., J. Shaffer, R. F. Murphy, J. Bonner, T. Hirose, and K. Itakura.** 1979. Hybridization of synthetic oligodeoxyribonucleotides to ϕX174 DNA: the effect of single base pair mismatch. Nucleic Acids Res. **6**:3543–3557.
35. **Wickner, S., J. Hoskins, D. Chattoraj, and K. McKenney.** 1990.

Deletion analysis of the mini-P1 plasmid origin of replication and the role of *Escherichia coli* DnaA protein. J. Biol. Chem. **265**:11622–11627.

36. **Wickner, S. H., and D. K. Chattoraj.** 1987. Replication of mini-P1 plasmid DNA in vitro requires two initiation proteins encoded by the *repA* gene of phage P1 and the *dnaA* gene of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **84**:3668–3672.
37. **Wosnick, M. A., R. W. Barnett, A. M. Vincentini, H. Erfle, R. Elliot, M. Sumner-Smith, N. Mantei, and R. W. Davies.** 1987. Rapid construction of large genes: total chemical synthesis of two different versions of the bovine prochymosin gene. Gene **60**:115–127.
38. **Yamaki, H., E. Ohtsubo, K. Nagai, and Y. Maeda.** 1988. The *oriC* unwinding by *dam* in *Escherichia coli*. Nucleic Acids Res. **16**:5067–5073.
39. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.