Anatomy of the Replication Origin of Plasmid ColE2-P9

Masaru Yagura,¹ Shin-ya Nishio,¹ Hideki Kurozumi,¹ Cheng-fu Wang,^{1†} and Tateo Itoh^{1,2*}

Department of Biology, Faculty of Science, Shinshu University, Matsumoto, Nagano 390-8621,¹ and Department of Biology, Faculty of Science, Osaka University, Toyonaka, Osaka 560-0043,² Japan

Received 6 July 2005/Accepted 3 November 2005

The plasmid ColE2-P9 origin is a 32-bp region which is specifically recognized by the plasmid-specified Rep protein to initiate DNA replication. We analyzed the structural and functional organization of the ColE2 origin by using various derivatives carrying deletions and single-base-pair substitutions. The origin may be divided into three subregions: subregion I, which is important for stable binding of the Rep protein; subregion II, which is important for binding of the Rep protein and for initiation of DNA replication; and subregion III, which is important for DNA replication but apparently not for binding of the Rep protein. The Rep protein might recognize three specific DNA elements in subregions I and II. The relative transformation frequency of the autonomously replicating plasmids carrying deletions in subregion I is lower, and nevertheless the copy numbers of these plasmids in host bacteria are higher than those of the wild-type plasmid. Efficient and stable binding of the Rep protein to the origin might be important for the replication efficiency to be at the normal (low) level. Subregion II might be essential for interaction with the catalytic domain of the Rep protein for primer RNA synthesis. The 8-bp sequence across the border of subregions II and III, including the primer sequence, is conserved in the (putative) origins of many plasmids, the putative Rep proteins of which are related to the ColE2-P9 Rep protein. Subregion III might be required for a step that is necessary after Rep protein binding has taken place.

DNA replication initiates at a specific chromosomal region called the replication origin. The DNA sequence elements required for replication origin activity have been extensively defined and characterized (30). Typically, they consist of a recognition element for the initiator protein and an AT-rich region. In the process of initiation of replication, the initiator protein binds to the origin and induces local unwinding of the duplex DNA. Subsequent events are the loading of a DNA helicase onto the unwound region followed by recruitment of additional factors to initiate DNA synthesis. In the *Escherichia coli* replication system, the ATP-bound DnaA proteins bind to the 9-mer DnaA boxes in *oriC*, and then the DnaB helicase is loaded onto unwound 13-mer AT-rich sequences, followed by loading of proteins required for initiation of DNA synthesis, including the DnaG primase (30).

The plasmid ColE2-P9 (ColE2) is a circular duplex DNA molecule of about 7 kb (18) and is present at about 10 to 15 copies per host chromosome (4, 21). The plasmid specifies an initiator protein (Rep protein, 35 kDa), which is the only plasmid-specified *trans*-acting factor essential for its replication (21, 26, 60). In addition, plasmid replication requires the host DNA polymerase I (28, 50) along with other host factors. The minimal *cis*-acting region (origin) (Fig. 1) required for plasmid replication consists of only 32 bp, where replication initiates (23, 60). The ColE2 Rep protein specifically binds to the ColE2 origin as revealed by a filter binding assay (26). The cloned origin inhibits replication of autonomously replicating plasmids (IncB incompatibility activity) due to competition for the

Rep proteins (51). In vitro ColE2 DNA replication starts at a unique site in the origin region and proceeds unidirectionally (23). The ColE2 Rep protein synthesizes a unique primer RNA (ppApGpA) for initiation of leading-strand DNA synthesis by DNA polymerase I at the origin (53, 54). This results in formation of a D-loop structure, where various host factors are loaded to continue DNA synthesis by DNA polymerase III holoenzyme. The plasmid ColE2-P9 belongs to a large family of very closely related plasmids (ColE2-related plasmids), of which 17 members are currently known (8, 58, 59). The nucleotide sequences of the replicon regions of 11 representative plasmids have been determined (19, 60). All these plasmids share the identical priming mechanism mediated by the plasmid-specified Rep protein. Among them there are four different specificity groups for interaction of the Rep proteins with the origins (19). Based on comparisons of the amino acid sequences of the Rep proteins and the nucleotide sequences of the origins, we have proposed that the specificity of interaction of the Rep proteins with the origins among these four groups of plasmids might be determined by combinations of the amino acid sequences in the three regions in the C-terminal portions of the Rep proteins and the nucleotide sequences at the three sites in the origins. By using chimeric *rep* genes and chimeric origins along with site-directed mutagenesis, we showed that the specificity between plasmids ColE2 and ColE3-CA38 (ColE3) is determined by combinations of the two sites in the origins with an insertion or deletion of a single base pair (Fig. 1) and corresponding combinations of amino acid sequences in the two regions in the C-terminal portions of the Rep proteins (42). For one of the regions in the Rep protein, the deletion/insertion of a nine-amino-acid sequence is the key factor for determination of the specificity. For the other region in the Rep protein, the specificity seems to be determined by a few amino acids.

* Corresponding author. Mailing address: Department of Biology, Faculty of Science, Shinshu University, Matsumoto, Nagano 390-8621, Japan. Phone: 81-263-37-2489. Fax: 81-263-37-2560. E-mail: tateito@gipac.shinshu-u.ac.jp.

† Present address: Department of Biochemistry, Suzhou Medical College, Suzhou, Jiangsu 215007, People's Republic of China.

FIG. 1. Origin of replication of plasmid ColE2-P9. The positions of nucleotides are numbered from left to right. Sites α and β are the specificity determinants in interaction of the origins with the Rep proteins in plasmids ColE2 and ColE3 (42). Repeats L and R contain directly repeated sequences 5'-CAGATAA-3'. The position of the primer RNA is marked by a square (53).

By using various derivatives of the ColE2 origin with mutations including deletions and a complete set of single-base-pair substitutions, we analyzed the structural and functional organizations of the ColE2 origin and showed that the region may be divided into at least three subregions. We discuss the functional roles of the three subregions.

MATERIALS AND METHODS

Bacterial strains and plasmids. The *Escherichia coli* K-12 strains used have been described elsewhere (19, 21, 23, 26). BL21(DE3) (Novagen) was used for production of the Rep protein. The plasmids used have been described elsewhere (21, 26, 42, 55) except for those described below.

Construction of derivatives of plasmid ColE2 origins with deletions and single-base-pair substitutions. Plasmids carrying derivatives of the ColE2 origin (Fig. 1) with deletions from the left (designated dl) and the right (designated dr) were constructed as described at <http://science.shinshu-u.ac.jp/~bios/PhysDev/tatei/supple/jb0993-05.pdf>. Briefly, deletion mutants were generated by digestion with an exonuclease from either side, and single-base-pair mutants were made by using mutagenic oligonucleotides.

Construction of pET21a-E2Rep. pET21a(+) (Novagen) was digested with BglII, filled with T4 DNA polymerase, and self-ligated to construct pET21a(+) Δ BglII. Plasmid pBE243RI-II (a gift from M. Shinohara), containing an EcoRI site just ahead of the *rep* gene, was constructed from pBE243 (42) by using oligonucleotide EcoRI-ATG. The 5.4-kb EcoRI-XhoI (filled with T4 DNA polymerase) fragment of pET21a(+) Δ BglII was ligated with the 0.9-kb EcoRI-BglII (filled with T4 DNA polymerase) fragment of pBE243RI-II to construct pET21a-E2Rep.

Oligonucleotides used. Oligonucleotides E2-UP and E2-DN have been described previously (60). Other oligonucleotides used in this study are listed at <http://science.shinshu-u.ac.jp/~bios/PhysDev/tatei/supple/jb0993-05.pdf>.

Other materials. The media used have been described previously (19, 21, 23). Chemicals, antibiotics, and enzymes, including some kits, were from commercial sources. Oligonucleotides used for site-directed mutagenesis were supplied by Pharmacia.

Measurement of the replication activity of mutant origins. Plasmid pMB9 was introduced into AG1 *recA1* cells carrying the test plasmids (pTI20dr, pTI21dl, and pTI20 point mutant ori series), and then the bacteria were cultured without selection for the test plasmids for 60 generations. One hundred independent colonies were tested for the presence or absence of the test plasmids to determine the retention frequency. The test plasmids were introduced into the bacteria with or without pMB9 to determine the relative transformation frequency. The relative transformation frequency of the test plasmids was estimated by using plasmid-free bacteria.

Incompatibility test and measurement of plasmid copy number. The test plasmids (pBR322dl, pBR322dr, pTI51dl, pTI51dr, and pBlue point mutant ori series) were introduced into AG1 *recA1* cells carrying pEC22. The bacteria were cultured without selection for pEC22 for 60 generations. One hundred independent colonies were tested for the presence or absence of pEC22 to determine the retention frequency. The plasmid pEC22 or pTI184 (21), a derivative of pACYC184 (7), was introduced into the bacteria with or without each of the test plasmids. The relative transformation frequency of pEC22 was normalized by using plasmid-free bacteria and pTI184. Measurement of plasmid copy number was performed as described previously (43).

Purification of Rep protein. BL21(DE3) cells harboring pET21a-E2Rep were grown in 100 ml of L broth containing ampicillin (200 μ g/ml) at 37°C to $\sim 2 \times 10^8$ cells/ml. Rep protein production was induced by treatment with 400 μ M isopropylthio- β -galactoside for 2 h. Cells were harvested and resuspended in 1 ml of buffer

Name	Sequence
ColE2 WT	5' -AAAATGAGACCAGATAAGCCTTATCAGATAACAGCGCCCT-3'
dl55	AAAATGAGACCAGATAAGCCTTATCAGATAACAGCGCCCT
dl4	tcgATGAGACCAGATAAGCCTTATCAGATAACAGCGCCCT
dl3	ggTcTGAGACCAGATAAGCCTTATCAGATAACAGCGCCCT
dl1	ctcTaGAGACCAGATAAGCCTTATCAGATAACAGCGCCCT
dl2	ccctcTAgACCAGATAAGCCTTATCAGATAACAGCGCCCT
dl53	AAgcTtgACCAGATAAGCCTTATCAGATAACAGCGCCCT
dl5	gAgTcagTACCAGATAAGCCTTATCAGATAACAGCGCCCT
dr27	AAAATGAGACCAGATAAGCCTTATCAGATAACAGCGCCga
dr55	AAAATGAGACCAGATAAGCCTTATCAGATAACAGCGCtag
dr61	AAAATGAGACCAGATAAGCCTTATCAGATAACAGCGgtCg
dr62	AAAATGAGACCAGATAAGCCTTATCAGATAACAGCAtCCc
dr56	AAAATGAGACCAGATAAGCCTTATCAGATAACAtCGatga
dr4	AAAATGAGACCAGATAAGCCTTATCAGAgctttatcgatg
2-dA	AAAATGAGACCAGATAAGCCTTATCAGTAacagcgCccg
dr8	AAAATGAGACCAGATAAGCCgaAgCttATcgatGataagc
dr31	AAAATGAGACCgaAgctttatcgATgAtAgctgtCaaaCa

FIG. 2. Mutant origins with deletions. Nucleotides of mutant origins identical to those of the wild-type origin at the corresponding positions are shown by uppercase letters. dl55 carries three base pairs derived from plasmid ColE2 to the left of position 1.

A (50 mM NaH_2PO_4 , 1 M NaCl, 0.5% polyoxyethylene sorbitan monolaurate, pH 8.0) containing 20 mM imidazole and 1 mM phenylmethylsulfonyl fluoride and were stored frozen at -80°C until use. All manipulations were done on ice unless otherwise stated. Frozen cells were thawed, incubated for 30 min with egg white lysozyme (1 mg/ml), sonicated, and clarified by centrifugation at $12,000 \times g$ for 30 min at 4°C . The supernatant was mixed with 200 μ l Ni^{2+} -nitrilotriacetic acid resin (50% slurry in buffer A). The resin was washed with buffer A containing 60 mM imidazole and eluted with buffer A containing 250 mM imidazole. The eluate was adjusted to final concentrations of 1 M Tris-HCl, pH 8.0, 0.01% bovine serum albumin, 1 mM dithiothreitol, 0.85 M NaCl, and 50% glycerol and stored at -20°C . The purity of the Rep protein was about 90%, as analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and staining with Coomassie brilliant blue.

Gel mobility shift assay. Binding reactions were performed in a final volume of 20 μ l containing 400 mM Tris-HCl, pH 7.5, 200 mM KCl, 20 mM MgSO_4 , 0.4 mM EDTA, 20% ethylene glycol, 5 nM ^{32}P -end-labeled DNA fragment, and an appropriate amount of the Rep protein. Salmon sperm DNA (0.5 μ g) was added as a nonspecific competitor. After incubation at 25°C for 30 min, the reaction mixtures were separated on 6% native polyacrylamide gels (10 by 10 cm) at 300 V (constant voltage) for 30 min in $0.5 \times$ Tris-borate-EDTA buffer at 4°C . Visualization of the DNA fragments was performed with a Fuji BAS 1500 phosphor-imager.

RESULTS AND DISCUSSION

Fine mapping of the boundaries of a functional origin. We have previously shown that the 32-bp region of plasmid ColE2 (positions 7 to 38; Fig. 1) is capable of functioning as the specific replication origin in the presence of the ColE2 Rep protein (60). To define the 5' (left) and 3' (right) ends of the origin more precisely, we constructed various deletion derivatives of the ColE2 origin (Fig. 2) with deletions from the left (dl55, dl4, dl3, dl1, dl2, dl53, and dl5) and from the right (dr27, dr55, dr61, dr62, dr56, dr4, 2-dA, dr8, and dr31) as described at <http://science.shinshu-u.ac.jp/~bios/PhysDev/tatei/supple/jb0993-05.pdf>. The replication activity of mutant origins was measured by a method using plasmid pMB9 as described previously (60). Plasmid ColE1 or its derivatives are readily lost from growing bacteria when the host bacteria also carry pMB9 (17, 56). When plasmid ColE1 carries another unrelated replicon (the ColE2 replicon in this study), it can be maintained by

TABLE 1. In vivo replication activity and Rep binding activity of deletion mutant origins of plasmid ColE2

Mutant name	Replication activity				IncB incompatibility activity (relative transformation frequency of pEC22 ^c with resident plasmid)	
	Retention frequency of test plasmids ^a			Relative transformation frequency of test plasmids ^b	pBR322 series	pTI51 series
	High	Low	None			
ColE2 (wild type)	1	0	0	1	<0.002	<0.002
dl55	1	0	0	0.5	<0.002	<0.002
dl4	ND ^d	ND	ND	ND	0.05	ND
dl3	1	0	0	0.7	0.1	<0.002
dl1	1	0	0	0.2	ND	0.2
dl2	1	0	0	0.06	ND	1
dl53	1	0	0	0.06	1	1
dl5	0	0	1	<0.002	ND	1
dr27	1	0	0	0.6	<0.002	<0.002
dr55	1	0	0	0.6	<0.002	<0.002
dr61	1	0	0	1	ND	ND
dr62	1	0	0	0.2	ND	ND
dr56	0	0	1	<0.002	0.03	<0.002
dr4	0	0	1	<0.002	0.13	0.01
2-dA	0	0	1	<0.001	0.15	0.013
dr8	0	0	1	<0.001	ND	1
dr31	0	0	1	<0.001	1	1

^a Retention frequency of the test plasmids (pTI20dl ori series and pTI21dr ori series; ColE1 vector plasmids carrying these mutant origins and the *rep* gene) without selection for 60 generations after introduction of pMB9. At least 100 colonies were tested. High, the bacteria grew well; Low, the bacteria grew poorly; None, no bacteria grew. The ratio of the colony number to the total colony number for each of the three criteria is shown.

^b Relative transformation frequency of the bacteria carrying pMB9 with the test plasmids. The transformation frequency of pTI20 or pTI21 carrying the wild-type ColE2 origin sequence is taken as 1.

^c Relative transformation frequency of the bacteria carrying the test plasmids (derivatives of pBR322 and pTI51 carrying these mutant origins) with pEC22. The transformation frequency of bacteria carrying the vector plasmid is taken as 1.

^d ND, not determined.

bacteria with coexisting pMB9, as it is capable of replication by use of the second replicon. When the ColE1-ColE2 composite plasmid carries a partially defective ColE2 origin, establishment of the ColE1-ColE2 composite plasmid introduced into the cells already loaded with pMB9 is more difficult than its retention by the cells carrying the composite plasmid into which pMB9 is introduced. Therefore, the relative transformation frequency is the better measure for detecting a relatively mild effect of mutations on the replication activity of the ColE2 origin, whereas the retention frequency is the better measure for detecting a relatively severe effect. When pMB9 was introduced into bacteria carrying derivatives of plasmid ColE1 with cloned mutant ColE2 origins and the ColE2 *rep* gene (Table 1), those with dl55, dl3, dl1, dl2, dl53, dr27, dr55, dr61, and dr62 were maintained with pMB9, indicating that these mutant origins are active as the functional origin of replication. Those carrying the rest of the mutant origins were all readily lost, indicating that they are inactive. On the other hand, when plasmids with mutant ColE2 origins were introduced into bacteria carrying pMB9 (Table 1), the relative transformation frequencies for dl55, dl3, dr27, dr55, and dr61 were as high as that for the intact ColE2 origin, whereas the frequencies were a little lower for dl1 and dr62 and considerably lower for dl2 and dl53. These results showed that the 5' end of the minimum origin required for replication is at position 8 and that the sequence up to position 5 is required for the full activity. These results also showed that the 3' end of the minimum origin required for replication is located at position 34 or 35 and that the region up to position 36 is required for the full activity.

The Rep protein binding activities of these mutant origins, including dl4 (Table 1), were estimated by their IncB incompatibility activities, namely, inhibition of ColE2 replication due to titration of the Rep protein by binding to the cloned origin (51). For that purpose, the mutant origins were cloned on pBR322 and pNT51, a derivative of ColE1 with a copy number of 200 per host chromosome (56). Establishment of the ColE2 replicon in the cells carrying pNT51 derivatives with the partially defective mutant ColE2 origin is more difficult than that in the cells carrying pBR322 derivatives. Therefore, the relative transformation frequency of the cells carrying pNT51 derivatives is the better measure for detecting a relatively severe effect of mutations on the Rep protein binding activity of the mutant ColE2 origin, whereas that of the cells carrying pBR322 derivatives is the better measure for detecting a relatively mild effect. When an autonomously replicating plasmid ColE2 (pEC22) was introduced into bacteria carrying derivatives of pNT51 with the cloned mutant origins (Table 1), no transformants were obtained for dl55, dl3, dr27, dr55, and dr56, while the relative transformation frequencies were very low for dr4 and 2-dA, low for dl1, and very high for dl2, dl53, dl5, dr8, and dr31, compared with bacteria carrying the pNT51 vector plasmid. On the other hand, when pEC22 was introduced into bacteria carrying pBR322 derivatives (Table 1), no transformants were obtained for dl55, dr27, and dr55, while the relative transformation frequencies were low for dl4, dl3, dr56, dr4, and 2-dA. These results showed that the mutant origins dl55, dr27, and dr55 are fully active, whereas the mutant origin dr56 is a little less active, the mutant origins dl4 and dl3 are less

TABLE 2. In vivo replication activity and Rep binding activity of single point mutant origins of plasmid ColE2

Position no.	Base pair		Replication activity				IncB incompatibility activity			
			Retention frequency of test plasmids ^a			Relative transformation frequency of test plasmids ^b	Retention frequency of pEC22 ^c			Relative transformation frequency of pEC22 ^d
	Wild type	Mutant	High	Low	None		High	Low	None	
1	A-T	G-C	1	0	0	1.9	0	0.01	0.99	<0.001
1	A-T	T-A	1	0	0	2.1	0	0	1	<0.001
1	A-T	C-G	1	0	0	1.1	0.02	0.02	0.96	<0.001
2	A-T	G-C	1	0	0	2.5	0	0	1	<0.002
2	A-T	T-A	1	0	0	3	0	0.01	0.99	<0.001
2	A-T	C-G	1	0	0	2.2	0	0	1	<0.002
3	A-T	G-C	1	0	0	1	0.01	0.01	0.98	<0.001
3	A-T	T-A	1	0	0	1.4	0	0.02	0.98	<0.001
3	A-T	C-G	1	0	0	1.3	0	0.1	0.9	<0.002
4	A-T	G-C	1	0	0	2.4	0	0	1	<0.001
4	A-T	T-A	1	0	0	1.2	0.01	0.02	0.97	<0.002
4	A-T	C-G	1	0	0	1.8	0	0.04	0.96	<0.001
5	T-A	A-T	1	0	0	0.47	0.94	0.06	0	0.47
5	T-A	G-C	1	0	0	0.35	1	0	0	0.056
5	T-A	C-G	1	0	0	0.17	0.99	0.01	0	<0.001
6	G-C	A-T	1	0	0	0.39	0.99	0.01	0	0.065
6	G-C	T-A	1	0	0	0.22	1	0	0	0.31
6	G-C	C-G	1	0	0	0.043	1	0	0	1.6
7	A-T	G-C	1	0	0	0.72	0.99	0.01	0	0.33
7	A-T	T-A	1	0	0	0.28	1	0	0	0.69
7	A-T	C-G	1	0	0	0.58	0.97	0.03	0	0.007
8	G-C	A-T	1	0	0	0.55	1	0	0	1
8	G-C	T-A	1	0	0	2.4	0.98	0.02	0	0.018
8	G-C	C-G	1	0	0	0.32	1	0	0	0.61
9	A-T	G-C	1	0	0	0.5	1	0	0	1.5
9	A-T	T-A	1	0	0	0.52	1	0	0	2
9	A-T	C-G	1	0	0	0.4	0.98	0.02	0	0.29
10	C-G	A-T	1	0	0	0.027	1	0	0	0.58
10	C-G	G-C	1	0	0	0.52	1	0	0	0.1
10	C-G	T-A	1	0	0	0.2	1	0	0	0.28
11	C-G	A-T	1	0	0	0.85	1	0	0	0.74
11	C-G	G-C	0.13	0.87	0	0.004	1	0	0	3.2
11	C-G	T-A	1	0	0	2.3	1	0	0	0.3
12	A-T	G-C	1	0	0	0.21	0	0.92	0.08	<0.001
12	A-T	T-A	1	0	0	0.23	1	0	0	0.73
12	A-T	C-G	1	0	0	0.037	1	0	0	1.4
13	G-C	A-T	1	0	0	0.22	0.01	0	0.99	<0.001
13	G-C	T-A	1	0	0	0.56	0.04	0.69	0.27	<0.001
13	G-C	C-G	1	0	0	0.52	0.76	0.2	0.04	0.001
14	A-T	G-C	1	0	0	0.22	0.99	0.01	0	0.16
14	A-T	T-A	1	0	0	0.4	0	0.89	0.11	<0.001
14	A-T	C-G	1	0	0	0.4	1	0	0	0.074
15	T-A	A-T	1	0	0	0.73	0.95	0.05	0	0.003
15	T-A	G-C	1	0	0	0.003	1	0	0	1.5
15	T-A	C-G	1	0	0	0.19	1	0	0	0.18
16	A-T	G-C	0.01	0.94	0.05	0.008	1	0	0	0.77
16	A-T	T-A	1	0	0	0.65	0.93	0.07	0	0.014
16	A-T	C-G	0	0	1	<0.001	1	0	0	0.44
17	A-T	G-C	1	0	0	0.39	0.97	0.03	0	0.079
17	A-T	T-A	1	0	0	0.032	0.98	0.02	0	0.23
17	A-T	C-G	0	0	1	<0.001	1	0	0	2.4
18	G-C	A-T	1	0	0	0.16	0.95	0.05	0	<0.002
18	G-C	T-A	1	0	0	0.26	1	0	0	0.32
18	G-C	C-G	1	0	0	0.71	0.95	0.05	0	<0.002
19	C-G	A-T	0	0	1	<0.001	0.98	0.02	0	2.7
19	C-G	G-C	0	0.32	0.68	<0.001	1	0	0	0.76
19	C-G	T-A	0	0.27	0.73	<0.001	1	0	0	0.8
20	C-G	A-T	0	0.12	0.88	0.001	0.15	0.85	0	1.3
20	C-G	G-C	0	0.9	0.1	0.044	1	0	0	0.47
20	C-G	T-A	0	0.93	0.07	0.034	1	0	0	0.51
21	T-A	A-T	1	0	0	0.31	1	0	0	<0.001
21	T-A	G-C	1	0	0	1.1	0.73	0.19	0.08	<0.002
21	T-A	C-G	1	0	0	1.9	0.93	0.07	0	<0.001

Continued on following page

TABLE 2—*Continued*

Position no.	Base pair		Replication activity				IncB incompatibility activity			
			Retention frequency of test plasmids ^a			Relative transformation frequency of test plasmids ^b	Retention frequency of pEC22 ^c			Relative transformation frequency of pEC22 ^d
	Wild type	Mutant	High	Low	None		High	Low	None	
22	T-A	A-T	0	0	1	<0.001	0.97	0.03	0	1.4
22	T-A	G-C	0	0	1	<0.001	1	0	0	1.5
22	T-A	C-G	0	0	1	<0.001	1	0	0	1.6
23	A-T	G-C	1	0	0	0.45	0.98	0.02	0	0.038
23	A-T	T-A	1	0	0	0.47	1	0	0	0.06
23	A-T	C-G	1	0	0	0.17	1	0	0	0.1
24	T-A	A-T	0	0	1	<0.001	1	0	0	2.2
24	T-A	G-C	0	0	1	<0.001	1	0	0	2.5
24	T-A	C-G	0	0	1	<0.001	1	0	0	0.6
25	C-G	A-T	0	0	1	<0.001	1	0	0	0.069
25	C-G	G-C	0	0	1	<0.001	1	0	0	0.045
25	C-G	T-A	0	0	1	<0.001	1	0	0	0.09
26	A-T	G-C	0	0	1	<0.001	0.99	0.01	0	0.034
26	A-T	T-A	0	0	1	<0.001	0.98	0.02	0	0.008
26	A-T	C-G	0	0.09	0.91	<0.001	0.92	0.08	0	0.002
27	G-C	A-T	0	0.05	0.95	<0.001	0.91	0.08	0.01	0.002
27	G-C	T-A	0	0.01	0.99	<0.001	0	0.82	0.18	<0.002
27	G-C	C-G	0	0.05	0.95	<0.001	0.06	0.93	0.01	<0.002
28	A-T	G-C	1	0	0	0.22	0	0.01	0.99	<0.002
28	A-T	T-A	0	0.1	0.9	<0.001	0	0.98	0.02	<0.001
28	A-T	C-G	0	0.17	0.83	0.069	0.05	0.93	0.02	<0.001
29	T-A	A-T	0	0.01	0.99	<0.001	0	0.7	0.3	<0.001
29	T-A	G-C	0	0	1	<0.001	0	0.73	0.27	<0.001
29	T-A	C-G	0	0	1	<0.001	0	0.75	0.25	<0.001
30	A-T	G-C	0	0.38	0.62	0.03	0	0.68	0.32	<0.001
30	A-T	T-A	0	0.61	0.39	0.057	0	0.51	0.49	<0.001
30	A-T	C-G	0	0.02	0.98	0.057	0	0.57	0.43	<0.001
31	A-T	G-C	0	0.76	0.24	0.26	0	0.21	0.79	<0.001
31	A-T	T-A	0	0.61	0.39	0.23	0	0.03	0.97	<0.001
31	A-T	C-G	0.18	0.65	0.17	0.68	0	0.06	0.94	<0.001
32	C-G	A-T	0	0.96	0.04	<0.002	0	0.18	0.82	0.001
32	C-G	G-C	0	0.55	0.45	<0.002	0	0.13	0.87	<0.001
32	C-G	T-A	0.13	0.87	0	0.028	0	0.16	0.84	0.001
33	A-T	G-C	0	0	1	<0.002	0	0.02	0.98	<0.001
33	A-T	T-A	0	0.03	0.97	<0.002	0	0.06	0.94	0.003
33	A-T	C-G	0	0.14	0.86	<0.002	0	0.29	0.71	0.001
34	G-C	A-T	0	0	1	<0.002	0	0.35	0.65	0.001
34	G-C	T-A	0	0.04	0.96	<0.002	0	0.41	0.59	<0.001
34	G-C	C-G	0	0.02	0.98	0.002	0	0.1	0.9	0.001
35	C-G	A-T	0	0.05	0.95	0.002	0	0.05	0.95	<0.001
35	C-G	G-C	0	0.08	0.92	<0.002	0	0.18	0.82	0.001
35	C-G	T-A	0	0.17	0.83	<0.002	0	0	1	<0.001
36	G-C	A-T	1	0	0	0.35	0	0	1	<0.001
36	G-C	T-A	1	0	0	0.54	0	0	1	<0.001
36	G-C	C-G	1	0	0	0.21	0	0	1	<0.001
37	C-G	A-T	1	0	0	0.38	0	0	1	<0.001
37	C-G	G-C	1	0	0	0.45	0	0	1	<0.001
37	C-G	T-A	1	0	0	0.42	0	0	1	<0.001

^a Retention frequency of the test plasmids (pTI20 point mutant ori series; ColE1 vector plasmids carrying these mutant origins and the *rep* gene) without selection for 60 generations after introduction of pMB9. At least 100 colonies were tested. High, the bacteria grew well; Low, the bacteria grew poorly; None, no bacteria grew. The ratio of the colony number to the total colony number for each of the three criteria is shown.

^b Relative transformation frequency of the bacteria carrying pMB9 with the test plasmids. The transformation frequency of pTI20 carrying the wild-type ColE2 origin sequence is taken as 1.

^c Retention frequency of pEC22 without selection for 60 generations after introduction of the test plasmids [pBlue point mutant ori series; a pBluescriptII KS(+) derivative with these mutant origins]. At least 100 colonies were tested. High, the bacteria grew well; Low, the bacteria grew poorly; None, no bacteria grew. The ratio of the colony number to the total colony number for each of the three criteria is shown.

^d Relative transformation frequency of the bacteria carrying the test plasmids with pEC22. The transformation frequency of bacteria carrying the vector plasmid is taken as 1.

active, and the mutant origins dl1, dr4, and 2-dA are much less active. The IncB incompatibility activities of dl2, dl53, dl5, dr8, and dr31 were undetectable by this method. Although the IncB incompatibility activities of dl2 and dl53 were undetectable,

they are active as the functional origin to support plasmid replication, as described above, indicating that they must be capable of interacting with the Rep protein. In contrast, the mutant origins dr56, dr4, and 2-dA retain a significant level of

The replication activities of these mutant origins carrying single-base-pair substitutions were measured as described in the previous section (Table 2 and Fig. 3). All the mutants with mutations in the region from position 1 to 14 retained the replication activity. From position 1 to 4, all the mutations had no effect on the replication activity. From position 5 to 10, some of the mutations slightly decreased the replication activity. These results together with the results of deletion analyses (Table 1) indicated involvement of positions 5 to 8 in the replication activity. One of the mutations at position 11 showed significantly lower activity, while the other two showed no

In the region from position 14 to 26 (subregion II; Fig. 3), many mutations severely decreased the IncB incompatibility

```

      AAAATGAGACCAGAT-AAGCCTTATCAGATAACAGCGC ColE2 origin
      AWTGCCYGATGCGRCGYNNNNMCGSCTTATCMGGCCTACR E. coli PU consensus sequence
      *** **      * *****      *

      AAAATGAGACCAGATAAGCCTTATCAGA-TAACAGCGC ColE2 origin
      YGTAGGCGCKGATAAGSCGKNNNNRCGYCGCATCRGGCAWT E. coli PU consensus sequence(c)
      ** ** ***** *      * * ** *

```

FIG. 5. *E. coli* sequences showing high homology with the ColE2 origin. *E. coli* PU constituting a family of repetitive DNA sequences present in large numbers in *E. coli* (15). (c) denotes the complementary sequence of the PU consensus sequence. Gaps (-) are inserted when appropriate to facilitate the alignment. Positions with identical residues in the two sequences are indicated by asterisks.

activity and the replication activity. Another DNA binding domain of the Rep protein might recognize the region from position 14 to 20 (site b; Fig. 3). The region from position 22 to 26 might be the third recognition element (site c; Fig. 3) for Rep protein binding. Some mutations in subregion II decreased the IncB incompatibility activity nearly below the detectable level and still retained the replication activity. By gel mobility shift assay, we detected weak but significant Rep protein binding activity of these mutant origins, and the result with ori 20G/C is shown in Fig. 4.

The A/T base pair in the ColE3 origin inserted at the corresponding position between positions 21 and 22 at the α site of the ColE2 origin (Fig. 1) is the specificity determinant in binding of the Rep proteins to the origins in the plasmids ColE2 and ColE3, and it was proposed to be a spacer (42). Mutations at position 21 hardly affected the IncB incompatibility activity and replication activity. Therefore, the T/A base pair at position 21 in the ColE2 origin and the corresponding base pair in the ColE3 origin (along with the inserted A/T base pair) might be a spacer between the two recognition elements (sites b and c). Even the region from position 21 to 23 could be a spacer. Mutations at position 22, however, severely affected the IncB incompatibility activity and the replication activity. These mutations might cause a conformational change of the origin DNA structure, and the proper interaction of these recognition elements with the Rep protein might be very important for the following step, probably primer RNA synthesis.

In the region from position 27 to 35 (subregion III; Fig. 3) mutations had only little effect on the IncB incompatibility activity. At positions 36 and 37, mutations had no effect on the activity. The results with mutations at positions 28 to 30 are apparently a little inconsistent with those of deletion analyses

with dr4 and 2-dA (Table 2). Multiple substitutions in the sequence to the right of positions 28 and 29 in dr4 and 2-dA, respectively, might have affected the IncB incompatibility activity. Even though the region from position 32 to 35 is not covered with the Rep protein (34), this region is required for the replication and might interact with a host factor required for replication. Initiation of plasmid ColE2 replication requires DNA polymerase I, which could be a candidate. DNA polymerase I is known to bind to highly repetitive bacterial DNA sequences called the palindromic unit (PU) in a sequence-specific manner (15). There are two regions in the ColE2 origin that show high homology with a part of the PU (Fig. 5), although both are outside subregion III. These sequences might still act to recruit DNA polymerase I for efficient utilization of the primer RNA. We are currently trying to detect specific interaction of DNA polymerase I with the ColE2 origin.

In the ColE2 origin there are two perfect direct repeats of 5'-CAGATAA-3' (Fig. 1). Repeated sequences are quite common in the origin sequences of various DNA replication systems (10). In every case the sequence is the recognition ele-

TABLE 4. Copy numbers of autonomously replicating plasmids carrying mutant origins with single-base-pair substitutions

Replication activity	IncB incompatibility activity ^a			
	$\pm$	+	++	+++
++++-++	8AT, 1.6; 8CG, 1.6; 9GC, 1.9; 9TA, 2.7	6TA, 2.0; 9CG, 2.5; 10TA, 2.4; 11TA, 1.3; 15CG, 1.1; 18TA, 0.7	4GC, 0.9; 5GC, 1.7; 6AT, 1.3; 8TA, 1.0; 10GC, 1.1; 12GC, 0.7; 13AT, 0.7; 15AT, 1.1; 16TA, 1.1; 17GC, 1.0; 23CG, 0.7	
++-+	6CG, 1.8; 10AT, 2.1; 11GC, 1.3; 15GC, 1.6; 16GC, 1.0	17TA, 1.0		ND ^b

^a Numbers are relative copy numbers of plasmids with mutant origins (pEC point mutant ori series), and the copy number of the wild-type plasmid is taken as 1.0. These numbers are averages from at least three independent assays. The replication activity (relative transformation frequency [TF] and retention frequency of test plasmids) and IncB incompatibility activity (relative transformation frequency and retention frequency of pEC22) of these mutant origins are from Table 2. Replication activity: +++, $0.5 < TF \leq 1$; ++, $0.1 < TF \leq 0.5$; +, $0.01 < TF \leq 0.1$; $\pm$, $0.002 < TF \leq 0.01$; $\pm$, $TF \leq 0.002$ and all the cells retained the test plasmids. IncB incompatibility activity: +++, $TF \leq 0.002$ and none of the cells retained pEC22; ++, $TF \leq 0.002$ and some of the cells did not retain pEC22; +, $0.002 < TF \leq 0.01$; $\pm$, $0.01 < TF \leq 0.1$; $\pm$, $0.1 < TF \leq 0.5$; $\pm$, $0.5 < TF$ and all the cells retained pEC22.

^b ND, not determined.

TABLE 3. Properties of autonomously replicating plasmids carrying deletion mutant origins

Property	Value for ColE2 origin ^a :					
	Wild type	dl55	dl3	dl1	dl2	dl53
Relative transformation frequency	1.0	1.0	1.0	0.3	0.01	0.015
Maintenance by host bacteria ^b	Stable	Stable	ND ^c	ND	Unstable	Unstable
Relative copy no.	1.0	1.0	1.5	2.5	2.5	1.8

^a These origins were carried by pEC22 (wild type), pEC2dl55, pEC2dl3, pEC2dl1, pEC2dl2, and pEC2dl53.

^b Retention frequency without selection after 100 generations: 1.0, stable; 0.90, unstable.

^c ND, not determined.

plasmid name	accession number	ori sequence	position number
pColE2-P9	D30054	AAAATGAGACCAGATAA-GCCT-TATCAGATAACAGCGC	1480
pColE3-CA38	D30055	AGAGT-AGACCAAATAA-GCCTATATCAGATAACAGCGC	1502
pColE5-099	D30060	AGAGT-AGACCAAATAAAACCTATATCAGATAACAGCGC	1235
pColE2-CA42	D30056	TGGAAGCGGAAAAATA-GCCTATATCAGATAACAGCGC	1485
pAsa1	AY301063	TGAGTGATC-AAAAACAAGCCTATATCAGATAACAGCGC	2755
pAsa3	AY301065	TGAGTGATCA-AAACAAAGCCTATATCAGATAACAGCGC	2931
pAsa11	AJ508382	TGAGTGATCA-AACAAA-GCCT-TATCAGATAACAGCGC	26
pP	AY079199	TTCGCGGACATGAACAA-GCCTATATCAGATAACAGCGC	3519
pEI2 ori1	AF244084	AGAGTGCACCAAATAAA-ACCTATATCAGATAACAGCCC	1815
pEI2 ori2	AF244084	AGAGTGCACCAAGATAACAACCTATATCAGATAACAGCCC	4215
pEI2 ori3	AF244084	GAGTGACACCAAGATAACAACCTATATCAGATAACAGCCC	5281
pUb6060 ori1	AJ249644	GGATTGATTGCTTGATAGTATAATAGCAGATAAAAAAGG	4228
pUb6060 ori2	AJ249644	CTGATGTGGCACTAAACAACCTATATCAGATAACAGCCC	5769
pMGD2	AY033498	GGCAGTTGGCACTAAACAACCTATATCAGATAACAGCCC	162
pEMCJH03 ori1	AY167745	*AAAATGACACCAGATAA-GCCTAAATCAGATAACAGCCC	1178
pEMCJH03 ori2	AY167745	*TTTTTGACAGCAGATAA-GCCTAAATCAGATAACAGCCG	1256
pVN84	AB200915.1	*AATGTGTGGCACTAAACAAGCTATATCAGATAACAGCCC	4686
pT3.2I	AJ007958	AGGTGTCTCACGAACCG-GCCTATATCAGATAACAGCCG	4368
pJD1 ori1	M10316	AATAAGAACCCTATCGGGCTGTTATCTGATTATGGTTA	702
pJD1 ori2	M10316	TACAACATATGTCAAATAACCATAATCAGATAACAGCCC	680c
86-028NP	NZ_AAE01000003	*ATTTGTAACAAAATATCATCCCATATCAGATAATTAATAA	2103
pDC3000A	AE016855	*CTTCGGGTGTGCAACATGGTGTACATCAGATACTGCGG	1605
pDC3000B	AE016854	*CTCCAGGTGTGCAATCCGGTGTACATCAGATACTGCGG	1604
pAV505	AJ222648	*CCTCGAGTGTGCAATCCGGTGTACATCAGATACTGCGG	1577
pPT23A	AJ224509	*CTCCAGGTGTGCAATCCGGTGTACATCAGATACTGCGG	1564
pPSR1	AY342395	*TCTTAAGTGTGCAACATGGTGTACATCAGATACTGCGG	1255
pFKN	AF359557	*CCTCGGGTGTGCAATCCGGTGTACATCAGATACTGCGG	1298
CMCP6 chr2	NC_004460	*TTTGTCTATGTTAATGATGCTGTACATCAGATACTGCTC	760874c
pTiK12	L36865	*CCGAGGATGTTAAACATGGTGCACATCAGGTAACAGCCG	2286
pCPP519	AF029751	*CTACGCCATCATTCAGTTTGTACGTACAGATAGCGCCCG	2338
pRSB101	AJ698325	*GTTGAAGTATGAATCATTACATTCATCAGATAATTCAGTT	14788
pPSS4918	AY768807.1	*CCTCGGGTGTGCAATCCGGTGTACATCAGATACTGCGG	1255
pB86-17A	AY768804.1	*CTCCGGGTGTGCAATCCGGTGTACATCAGATACTGCGG	1255
pOK1A	AY768806.1	*CTCCAGGTGTGCAATCCGGTGTACATCAGATACTGCGG	1255
pCG131	AY768805.1	*CCGCGGGTGTGCAATCCGGTGTACATCAGATACTGCGG	1255
pPG2708	AY768799.1	*CTCCAGGTGTGCAATCCGGTGTACATCAGATACTGCGG	1255
pPA0893A	AY768793.1	*CCTCGAGTGTGCAATCCGGTGTACATCAGATACTGCGG	1255
pPA0893C	AY768794.1	*CCTCGAGTGTGCAATATGGTGTACATCAGATACTGCGG	1255
pPSM9032A	AY768801.1	*CTTCGGGTGTGCAACATGGTGTACATCAGATACTGCGG	1255
pPSM9032B	AY768802.1	*CTTCGGGTGTGCAACATGGTGTACATCAGATACTGCGG	1255
pPS0693A	AY768795.1	*CCTCGGGTGTGCAATATGGTGTACATCAGATACTGCGG	1255
pPS0485A	AY768796.1	*ACCAGGGGGTGTATATGCTGTACATCAGATACTGCGG	1255
pPS0485B	AY768797.1	*ACCAGGGGGTGTATATGCTGTACATCAGATACTGCGG	1255
pPS0485C	AY768798.1	*ACCATGGGGTGTATATGCTGTACATCAGATACTGCGG	1255
pPSTA0893A	AY768803.1	*CTTCGGGTGTGCAACATGGTGTACATCAGATACTGCGG	1255
pPSM8810	AY768800.1	*CTTCGGGTGTGCAACATGGTGTAAAGCAGATACTGCGG	1255
pPMA4326A	AY603979.1	*ACCAAGGTGCTCCATATGGTGTACATCAGATACTGCGG	1254
pPMA4326B	AY603980.1	*ACCAAGGCGCTCCATATGGTGTACATCAGATACTGCGG	1243
pTT8 ori1	AB119193	CCCTCGCTAAATCGCAGGACTCTTCCAGATAAACCCCG	3576
pTT8 ori2	AB119193	AGACACCTTGAGATCCGAGAGAAAGTGAGATAACTACA	4990
pXAC33	AE008924	*GCCCGCTTTGGCGAGCGGGCCT-TGACAGATAACAGGG	11345
pXAC64	AE008925.1	*TGCCCTTCTCTCTCTTCCAACCTCTCTCAGATAACGGTGG	9864c
pHE1	AJ132759	*GCATCATCGTCTACATCAGCGAGCGCAACATAACAGCCG	161

FIG. 6. Comparison of the putative origin sequences of various plasmids from gram-negative bacteria with the origin sequences of ColE2-related plasmids (the four plasmids at the top). 86-028NP is one of the whole-genome shotgun sequences of *Haemophilus influenzae* 86-028NP. CMCP6 chr2 is the second chromosome sequence of *Vibrio vulnificus*. The position number in the database sequence of the nucleotide at the right

ment for the initiator protein. The present study clearly showed that these two sequences in the ColE2 origin are functionally distinct. The 3' half of the direct repeat on the right is not covered by the Rep protein as revealed by DNase I footprint analysis (34).

Properties of autonomously replicating plasmids carrying mutant origins. To examine the properties of the mutant origins with deletions from the left (dl1, dl2, dl3, and dl53) in more detail, we constructed autonomously replicating plasmids carrying the mutant origins and *rep* gene (Table 3). The relative transformation frequencies of these plasmids carrying mutant origins dl1, dl2, and dl53 were considerably lower than those of plasmids with the intact origin. Plasmids with mutant origins dl2 and dl53 were lost from about 2% of host bacteria after 40 generations and from 10% after 100 generations without selection for resistance to the antibiotics. Plasmids with the intact origin were stably maintained by host bacteria at least for 100 generations. We then examined the copy numbers of these autonomously replicating plasmids in exponentially growing bacteria (Table 3). Unexpectedly, the plasmids carrying mutant origins dl1, dl2, and dl53, with much lower relative transformation frequency, were maintained by host bacteria with significantly higher copy numbers than those of other plasmids. The plasmid with dl3 also exhibited some increase in the copy number. The properties of the mutant origin dl55 were quite similar to those of the intact origin.

We further examined the copy numbers of autonomously replicating plasmids carrying mutant origins with single-base-pair substitutions (Table 4). For the mutant origins with Rep protein binding activity decreased below the detectable level and with (near-) normal replication activity, the copy numbers of the autonomously replicating plasmids were 1.6 to 2.7 times higher than that of the wild-type origin. For some mutant origins with the binding activity decreased below the detectable level and with the replication activity reduced severely, the relative copy numbers were still 1.0 to 2.1. For the mutant origins in which mutations had a significant effect on the Rep protein binding activity, the relative copy numbers of the plasmids with mutant origins in subregion I were 1.3 to 2.5 and those of plasmids with mutant origins in subregion II were 0.7 to 1.1. For those mutant origins in which mutations had a little or no effect on both the Rep protein binding activity and replication activity, the relative copy numbers were 0.7 to 1.7.

These results showed that some of the mutant origins of subregion I with Rep protein binding activity below the detectable level may initiate replication more efficiently than the wild-type origin. We suggest that efficient and stable binding of the Rep protein to the origin might be important for the replication efficiency to be at the normal (low) level. This possibility has been previously suggested by the study of the

chimeric origins and Rep proteins between plasmids ColE2 and ColE3 (42).

Models for binding of the Rep protein to the origin. Based upon the results describe above, we discuss possible models. If we assume that the Rep protein binds to the origin as a monomer, the helix-turn-helix motif in the C-terminal region of the Rep protein (19) recognizes and binds to site a (Fig. 3), and the putative second and third DNA binding domains specifically bind to sites b and c separated by the spacer. The N-terminal domain of the Rep protein might recognize and bind to the region around site c to synthesize the primer RNA. The spacer might be important for the proper positioning of the N-terminal domain on the template DNA strand. If we assume that the ColE2 Rep protein binds to the origin as a dimer, there could be two alternative models. In either model, two distinct binding domains in the C-terminal half of one of the Rep proteins in the dimer recognizes sites a and b and binds to the origin DNA, and the spacer might be important for the proper positioning of the second Rep protein. Binding of the second Rep protein to the origin is facilitated by protein-protein interaction. In one model, the third DNA recognition domain of the second Rep protein in the dimer recognizes site c and catalyzes the primer RNA synthesis. In the other model, either of the two recognition domains of the second Rep protein in the dimer interacts with site c, which is only weakly homologous to site a or b. Although the Rep proteins form stable dimers even in the presence of 1 M NaCl (35), so far we do not know whether the Rep protein binds to the origin as a monomer or a dimer (or as an oligomer).

ColE2-related plasmids in gram-negative and gram-positive bacteria. For the past 10 years or so, growing numbers of plasmids whose putative Rep proteins are related to the ColE2-P9 Rep protein have been described (Fig. 6 and 7) (1, 2, 3, 5, 6, 9, 11, 12, 13, 14, 16, 19, 20, 22, 25, 27, 29, 31, 32, 33, 36, 37, 38, 39, 40, 41, 44, 45, 46, 49, 52, 57, 61, 62). For the ColE2-related plasmids, such as plasmids ColE2-P9, ColE3-CA38, ColE5-099, and ColE2-CA42, the origin sequences where the Rep proteins bind in a plasmid-specific manner and initiate plasmid replication have been identified (19, 23, 53, 60). These origin sequences showed good homology with each other, and the sequences around the primer sequence are almost identical (Fig. 6). For only some of the remaining plasmids have the origin sequences been proposed. For the rest of the plasmids, we searched for candidates based on the sequence homology with the origins of the ColE2-related plasmids (Fig. 6 and 7). The putative origin sequences of pEMCJH03 proposed here are very similar to the ColE2-P9 origin sequence. The putative replication origins of pFAJ2600, namely, *ori1*, *ori2*, and *ori3* (11) and *ori4* (32), have been mentioned. We propose here three other sequences (*ori5*, *ori6*, and *ori7*) as better candidates based on the

end of each (putative) origin sequence is shown. c denotes the complementary strand of the database sequence. The (putative) origin sequences of the following plasmids have been presented previously: ColE2-P9, ColE3-CA38, ColE5-099, and ColE2-CA42 (19); pAsa1, pAsa3, and pAsa11 (5); pP (16); pEI2 (13); pUb6060 (3); pMGD2 (61); pT3.2I (2); pJD1 (11, 29, 32); 86-028NP (38); pDC3000A and pDC3000B (6); pAV505 and pPT23A (14); pPSR1 (46); pFKN (39); CMCP6 chr2 (27); pTiK12 (12); pCPP519 (22); pPSRB101 (49); pSS4918, pB86-17A, pOK1A, pCG131, pPG2708, pPA0893A, pPSM9032A, pPSM9032B, pPS0693A, pPS0485A, pPS0485B, pPS0485C, pPSTA0893A, and pPSM8810 (62); pPMA4326A and pPMA4326B (44); pTT8 (52); pXAC33 and pXAC64 (9); and pHE1 (57). The asterisks denote the putative origin sequences proposed in this study. The position of the primer is marked. Regions with partial homology further extend on both sides of the highly conserved sequences marked by a square.

plasmid name	accession number	ori sequence	position number
pColE2-P9	D30054	AAAATGAGACCAGATAAGCCTTATCAGATAACAGCGC	1480
pRBL1	U39878	CAAGGTAGGCATCGAAAGCAATATCAGATGGTTCCTG	713c
pBL-A8	Y11902	CAAAGTGGGTCTCGAAAGCAATATCAGATGGTTCCTG	1204c
pB264 ori1	AY297818	*TCAAAGTTAATCGTGAAGCCAAGTCAGATGTTTCGGT	2975
pB264 ori2	AY297818	*TCAAAGTTAATCGTGAAGCCAAGTCAGATGTTTCGGT	3031
pKA22	AF165152	*TCAAAGTTAATCGTGAAGCCAAGTCAGATGTTTCGGT	1131
pFAJ2600 ori1	AF015088	GGGGTGGGTTTCAGCTCCAAAAATTCGACTACCTGAC	3212c
pFAJ2600 ori2	AF015088	GAAAAATATTCGACTAGGCATCCGAAGCCAAGTCAG	3201
pFAJ2600 ori3	AF015088	CCCTCGCACGCATTTGACGACTATTCGACTAGGCACG	3166
pFAJ2600 ori4	AF015088	CATCCCGAAGCCAAGTCAGGTAGTCGAATTTTGGAG	3234
pFAJ2600 ori5	AF015088	*CGACTAGGCATCCCGAAGCCAAGTCAGGTAGTCGAAT	3226
pFAJ2600 ori6	AF015088	*CATGCAACACGCCGCGTGCCTAGTCGAATAGTCGTCA	3144c
pFAJ2600 ori7	AF015088	*TGACTTGGCTTCGGGATGCCTAGTCGAATATTTTCA	3179c
pDOJH10S	AF538869	*TCGGAATATGTTACGAACCAAGTCAGATATTTTCAGC	3583c
pMB1	X84655	TGAACAAATACGCGGGAACCAAGTCAGATATTTTCAGC	394c
pXZ10142 ori1	NC_002009	AACCCCTGAAACCCCGGAAACATCTGACTTGGTTAC	667
pXZ10142 ori2	NC_002009	AGACGCGCTAAACCCCGGAAATATCTGACTTGGTTTCG	1842c
pXZ10142 ori3	NC_002009	GTTCAAAGTCCATCGTAACCAAGTCAGATGTTTCCTG	645c
pXZ10142 ori4	NC_002009	*TTCATAATAGGTACGAACCAAGTCAGATATTTTCCTG	1943
pRC4	AB040101	*TTATGTAGGACACGCCAACCAAGTCAGATGTTTCCTG	1073c
pAL5000 ori1	M23557	TGGGGTGGCCCCCTCAGCGAAATATCTGACTTGGAGCT	4586
pAL5000 ori2	M23557	TATGGTCCGACACGAGCTCCAAGTCAGATATTTTCCTG	4018c
SS9 ch2 ori1	CR378680	*TTATGGCAGATGCTTCGCAACGGTCAGATGAAAAATG	192361
SS9 ch2 ori2	CR378680	*AGCAGGGGTAAAGTGCCGCATCGTCAGATCGCCAATA	192818
SS9 ch2 ori3	CR378680	*GTTGCGGATATCAACAAAGTCCGTCAGATAATCGAGG	192948
pRGO1	AB007909	*CGGAAGTACGCCGCTAAACCAAAATCAGATAGTTTCCTG	2629c
pN30	AF312210	*CACCGGATTTTTTGGCTCCGCCCATCAGATAATCGCCT	2034c
p545	AF291751	*CACTCAAGGTCGTTTTTGGGGTGTCAGATGAGACCTC	929
pLME106	AJ250233	*TGAACCAGGGGAACAGGGCTAGTCAGATAGGATCGA	4279
pLIM	AY004211	*TGCGGGTGGCTGAGTCGGGGCGTTTCAGATGTTGCCAG	3676c

FIG. 7. Comparison of the putative origin sequences of various plasmids from gram-positive bacteria with the origin sequence of plasmid ColE2-P9. SS9 chr2 is the second chromosome sequence of *Photobacterium profundum*. The (putative) origin sequences of the following plasmids have been presented previously: pRBL1 (1, 11, 32); pBLA8 (32); pB264 (33); pKA22 (31); pFJ2600 (11, 32); pMB1 (11, 32, 40); pXZ10142 (11, 32, 41); pRC4 (20); pAL5000 (11, 32, 45); pRGO1 (25); pLME106 (36); and pLIM (37). Regions with partial homology further extend on both sides of the highly conserved sequences. The position numbers, symbols, and marks are as in Fig. 6.

higher similarity with the origins of the ColE2-related plasmids, which partly overlap *ori1*, *ori3*, and *ori2*, respectively, in the opposite orientation. For pXZ10142, *ori1* and *ori2* (11) and *ori3* (32) have been proposed, and we propose here another candidate (*ori4*). *ori3* and *ori4* are apparently better candidates than *ori1* and *ori2*. Similarly *ori2* of pAL5000 seems to be a better candidate than *ori1*. Both the RepB binding H site and L site are located close to *ori1* and *ori2* (45).

The sequence 5'-ATCAGATA-3' is found in the putative origins of most of the plasmids from the gram-negative bacteria (Fig. 6). Closely related sequences are found in pUB6060 (*ori1*), pTik12, pCPP519, pTT8 (*ori1*), pXAC33, and pHE1. The sequence 5'-ATCAGATA-3' or similar sequences are also found in many of the plasmids from the gram-positive bacteria (Fig. 7). This sequence is located across the boundary of subregions II and III of the ColE2-P9 origin (Fig. 3), contains the primer sequence, and is very important for in vivo replication as described above. The DnaG primase of *E. coli* has been

shown to interact with the DNA sequences upstream and downstream of the primer start position (24). The catalytic domains of the Rep proteins of the ColE2-related plasmids for primer synthesis might interact with the 5'-TATCTGAT-3' sequence on the template strand. It can be speculated that all these plasmids (Fig. 6 and 7) share a common mechanism of initiation of DNA replication, that is, the primer RNA synthesis by the Rep proteins. It is noteworthy that these plasmids are widespread in gram-positive and -negative bacteria and form a large plasmid family.

ACKNOWLEDGMENTS

We are grateful to H. Ogawa and T. Ogawa for helpful discussions and continuing encouragement. We thank T. Horii, A. Shinohara, M. Shinohara, H. Kubo, and Y. Itoh for discussion and A. Ohshima, Desaki, M. Shinohara, K. Matsumoto, J. Itoh, T. Tanigaki, and Y. Suzuki for materials used. We also thank G. Minote for radioisotope experiments in the Research Center for Human and Environmental Sciences of Shinshu University.

This work was supported in part by a grant-in-aid for scientific research from the Ministry of Education, Science and Culture of Japan.

REFERENCES

- Ankri, S., I. Bouvier, O. Reyes, F. Predali, and G. Leblon. 1996. A *Brevibacterium linens* pRBL1 replicon functional in *Corynebacterium glutamicum*. *Plasmid* 36:36–41.
- Aparicio, T., P. Lorenzo, and J. Perera. 2000. pT3.2I, the smallest plasmid of *Thiobacillus* T3.2. *Plasmid* 44:1–11.
- Avison, M. B., T. R. Walsh, and P. M. Bennett. 2001. pUb6060: a broad-host-range, DNA polymerase-I-independent ColE2-like plasmid. *Plasmid* 45:88–100.
- Bazal, M., and D. R. Helinski. 1968. Circular DNA forms of colicinogenic factors E1, E2 and E3 from *Escherichia coli*. *J. Mol. Biol.* 36:185–194.
- Boyd, J., J. Williams, B. Curtis, C. Kozera, R. Singh, and M. Reith. 2003. Three small, cryptic plasmids from *Aeromonas salmonicida* subsp. *salmonicida* A449. *Plasmid* 50:131–144.
- Buell, C. R., V. Joardar, M. Lindeberg, J. Selengut, I. T. Paulsen, M. L. Gwinn, R. J. Dodson, R. T. Deboy, A. S. Durkin, J. F. Kolonay, R. Madupu, S. Daugherty, L. Brinkac, M. J. Beanan, D. H. Haft, W. C. Nelson, T. Davidsen, N. Zafar, L. Zhou, J. Liu, Q. Yuan, H. Khouri, N. Fedorova, B. Tran, D. Russell, K. Berry, T. Utterback, S. E. Van Aken, T. V. Feldblyum, M. D'Ascenzo, W. L. Deng, A. R. Ramos, J. R. Alfano, S. Cartinhour, A. K. Chatterjee, T. P. Delaney, S. G. Lazarowitz, G. B. Martin, D. J. Schneider, X. Tang, C. L. Bender, O. White, C. M. Fraser, and A. Collmer. 2003. The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000. *Proc. Natl. Acad. Sci. USA* 100:10181–10186.
- Chang, A. C., and S. N. Cohen. 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.* 134:1141–1156.
- Cooper, P. C., and R. James. 1984. Two new E colicins, E8 and E9, produced by a strain of *Escherichia coli*. *J. Gen. Microbiol.* 130:209–215.
- da Silva, A. C., J. A. Ferro, F. C. Reinach, C. S. Farah, L. R. Furlan, R. B. Quaggio, C. B. Monteiro-Vitorello, M. A. Van Sluys, N. F. Almeida, L. M. Alves, A. M. do Amaral, M. C. Bertolini, L. E. Camargo, G. Camarotte, F. Cannavan, J. Cardozo, F. Chamberg, L. P. Ciapina, R. M. Ciccarelli, L. L. Coutinho, J. R. Cursino-Santos, H. El-Dorry, J. B. Faria, A. J. Ferreira, R. C. Ferreira, M. I. Ferro, E. F. Formighieri, M. C. Franco, C. C. Greggio, A. Gruber, A. M. Katsuyama, L. T. Kishi, R. P. Leite, E. G. Lemos, M. V. Lemos, E. C. Locali, M. A. Machado, A. M. Madeira, N. M. Martinez-Rossi, E. C. Martins, J. Meidanis, C. F. Menck, C. Y. Miyaki, D. H. Moon, L. M. Moreira, M. T. Novo, V. K. Okura, M. C. Oliveira, V. R. Oliveira, H. A. Pereira, A. Rossi, J. A. Sena, C. Silva, R. F. de Souza, L. A. Spinola, M. A. Takita, R. E. Tamura, E. C. Teixeira, R. I. Tezza, M. Trindade dos Santos, D. Truffi, S. M. Tsai, F. F. White, J. C. Setubal, and J. P. Kitajima. 2002. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* 417:459–463.
- del Solar, G., R. Giraldo, M. J. Ruiz-Echevarria, M. Espinosa, and R. Diaz-Orejas. 1998. Replication and control of circular bacterial plasmids. *Microbiol. Mol. Biol. Rev.* 62:434–464.
- De Mot, R., I. Nagy, A. De Schrijver, P. Pattanapitpaisal, G. Schoofs, and J. Vanderleyden. 1997. Structural analysis of the 6 kb cryptic plasmid pFAJ2600 from *Rhodococcus erythropolis* N186/21 and construction of *Escherichia coli*-*Rhodococcus* shuttle vectors. *Microbiology* 143:3137–3147.
- English, R. S., S. C. Lorbach, K. M. Huffman, and J. M. Shively. 1995. Isolation and characterization of the replicon of a *Thiobacillus intermedius* plasmid. *Plasmid* 33:1–6.
- Fernandez, D. H., L. Pittman-Cooley, and R. L. Thune. 2001. Sequencing and analysis of the *Edwardsiella ictaluri* plasmids. *Plasmid* 45:52–56.
- Gibbon, M. J., A. Sesma, A. Canal, J. R. Wood, E. Hidalgo, J. Brown, A. Vivian, and J. Murillo. 1999. Replication regions from plant-pathogenic *Pseudomonas syringae* plasmids are similar to ColE2-related replicons. *Microbiology* 145:325–334.
- Gilson, E., D. Perrin, and M. Hofnung. 1990. DNA polymerase I and a protein complex bind specifically to *E. coli* palindromic unit highly repetitive DNA: implications for bacterial chromosome organization. *Nucleic Acids Res.* 18:3941–3952.
- Gregorova, D., J. Matiasovicova, A. Sebkova, M. Faldynova, and I. Rychlik. 2004. *Salmonella enterica* subsp. *enterica* serovar Enteritidis harbours ColE1, ColE2, and rolling-circle-like replicating plasmids. *Can. J. Microbiol.* 50:107–112.
- Hashimoto-Gotoh, T., and K. N. Timmis. 1981. Incompatibility properties of Col E1 and pMB1 derivative plasmids: random replication of multicopy replicons. *Cell* 23:229–238.
- Herschman, H. R., and D. R. Helinski. 1967. Comparative study of the events associated with colicin induction. *J. Bacteriol.* 94:691–699.
- Hiraga, S., T. Sugiyama, and T. Itoh. 1994. Comparative analysis of the replicon regions of eleven ColE2-related plasmids. *J. Bacteriol.* 176:7233–7243.
- Hirasawa, K., Y. Ishii, M. Kobayashi, K. Koizumi, and K. Maruhashi. 2001. Improvement of desulfurization activity in *Rhodococcus erythropolis* KA2-5-1 by genetic engineering. *Biosci. Biotechnol. Biochem.* 65:239–246.
- Horii, T., and T. Itoh. 1988. Replication of ColE2 and ColE3 plasmids: the regions sufficient for autonomous replication. *Mol. Gen. Genet.* 212:225–231.
- Huang, T. C., and T. J. Burr. 1999. Characterization of plasmids that encode streptomycin-resistance in bacterial epiphytes of apple. *J. Appl. Microbiol.* 86:741–751.
- Itoh, T., and T. Horii. 1989. Replication of ColE2 and ColE3 plasmids: in vitro replication dependent on plasmid-coded proteins. *Mol. Gen. Genet.* 219:249–255.
- Khopde, S., E. E. Biswas, and S. B. Biswas. 2002. Affinity and sequence specificity of DNA binding and site selection for primer synthesis by *Escherichia coli* primase. *Biochemistry* 41:14820–14830.
- Kiatpapan, P., Y. Hashimoto, H. Nakamura, Y. Z. Piao, H. Ono, M. Yamashita, and Y. Murooka. 2000. Characterization of pRGO1, a plasmid from *Propionibacterium acidipropionici*, and its use for development of a host-vector system in propionibacteria. *Appl. Environ. Microbiol.* 66:4688–4695. (Erratum, 67:1024, 2001.)
- Kido, M., H. Yasueda, and T. Itoh. 1991. Identification of a plasmid-coded protein required for initiation of ColE2 DNA replication. *Nucleic Acids Res.* 19:2875–2880.
- Kim, Y. R., S. E. Lee, C. M. Kim, S. Y. Kim, E. K. Shin, D. H. Shin, S. S. Chung, H. E. Choy, A. Progulske-Fox, J. D. Hillman, M. Handfield, and J. H. Rhee. 2003. Characterization and pathogenic significance of *Vibrio vulnificus* antigens preferentially expressed in septicemic patients. *Infect. Immun.* 71:5461–5471.
- Kingsbury, D. T., and D. R. Helinski. 1970. DNA polymerase as a requirement for the maintenance of the bacterial plasmid colicinogenic factor E1. *Biochem. Biophys. Res. Commun.* 41:1538–1544.
- Korch, C., P. Hagblom, H. Ohman, M. Goransson, and S. Normark. 1985. Cryptic plasmid of *Neisseria gonorrhoeae*: complete nucleotide sequence and genetic organization. *J. Bacteriol.* 163:430–438.
- Kornberg, A., and T. A. Baker. 1992. DNA replication, 2nd ed. Freeman and Company, New York, N.Y.
- Kulakov, L. A., M. J. Larkin, and A. N. Kulakova. 1997. Cryptic plasmid pKA22 isolated from the naphthalene degrading derivative of *Rhodococcus rhodochrous* NCIMB13064. *Plasmid* 38:61–69.
- Leret, V., A. Trautwetter, A. Rince, and C. Blanco. 1998. pBLA8, from *Brevibacterium linens*, belongs to a gram-positive subfamily of ColE2-related plasmids. *Microbiology* 144:2827–2836.
- Lessard, P. A., X. M. O'Brien, D. H. Currie, and A. J. Sinskey. 2004. pB264, a small, mobilizable, temperature sensitive plasmid from *Rhodococcus*. *BMC Microbiol.* 4:15.
- Matsui, H., and T. Itoh. Unpublished data.
- Matsumoto, K., and T. Itoh. Unpublished data.
- Miescher, S., M. P. Stierli, M. Teuber, and L. Meile. 2000. Propionin SM1, a bacteriocin from *Propionibacterium jensenii* DF1: isolation and characterization of the protein and its gene. *Syst. Appl. Microbiol.* 23:174–184.
- Moore, M., C. Svenson, D. Bowling, and D. Glenn. 2003. Complete nucleotide sequence of a native plasmid from *Brevibacterium linens*. *Plasmid* 49:160–168.
- Munson, R. S., Jr., A. Harrison, A. Gillaspay, W. C. Ray, M. Carson, D. Armbruster, J. Gipson, M. Gipson, L. Johnson, L. Lewis, D. W. Dyer, and L. O. Bakaletz. 2004. Partial analysis of the genomes of two nontypeable *Haemophilus influenzae* otitis media isolates. *Infect. Immun.* 72:3002–3010.
- Rohmer, L., S. Kjemtrup, P. Marchesini, and J. L. Dangl. 2003. Nucleotide sequence, functional characterization and evolution of pFKN, a virulence plasmid in *Pseudomonas syringae* pathovar maculicola. *Mol. Microbiol.* 47:1545–1562.
- Rossi, M., P. Brigidi, A. Gonzalez Vara y Rodriguez, and D. Matteuzzi. 1996. Characterization of the plasmid pMB1 from *Bifidobacterium longum* and its use for shuttle vector construction. *Res. Microbiol.* 147:133–143.
- Shen, T., S. Na, W. Xiao, and P. Jia. 1996. Essential region for self-replication of *Corynebacterium* plasmid pXZ10145. *Wei Sheng Wu Xue Bao* 36:168–172.
- Shinohara, M., and T. Itoh. 1996. Specificity determinants in interaction of the initiator (Rep) proteins with the origins in the plasmids ColE2-P9 and ColE3-CA38 identified by chimera analysis. *J. Mol. Biol.* 257:290–300.
- Som, T., and J. Tomizawa. 1982. Origin of replication of *Escherichia coli* plasmid RSF 1030. *Mol. Gen. Genet.* 187:375–383.
- Stavrindes, J., and D. S. Guttman. 2004. Nucleotide sequence and evolution of the five-plasmid complement of the phytopathogen *Pseudomonas syringae* pv. *maculicola* ES4326. *J. Bacteriol.* 186:5101–5115.
- Stolt, P., and N. G. Stoker. 1996. Functional definition of regions necessary for replication and incompatibility in the *Mycobacterium fortuitum* plasmid pAL5000. *Microbiology* 142:2795–2802.
- Sundin, G. W., C. T. Mayfield, Y. Zhao, T. S. Gunasekera, G. L. Foster, and M. S. Ullrich. 2004. Complete nucleotide sequence and analysis of pPSR1 (72,601 bp), a pPT23A-family plasmid from *Pseudomonas syringae* pv. *syringae* A2. *Mol. Genet. Genomics* 270:462–476.

47. Suzuki, M. 1993. Common features in DNA recognition helices of eukaryotic transcription factors. *EMBO J.* **12**:3221–3226.
48. Suzuki, M., and N. Yagi. 1994. DNA recognition code of transcription factors in the helix-turn-helix, probe helix, hormone receptor, and zinc finger families. *Proc. Natl. Acad. Sci. USA* **91**:12357–12361.
49. Szczepanowski, R., I. Krahn, B. Linke, A. Goesmann, A. Puhler, and A. Schluter. 2004. Antibiotic multiresistance plasmid pRSB101 isolated from a wastewater treatment plant is related to plasmids residing in phytopathogenic bacteria and carries eight different resistance determinants including a multidrug transport system. *Microbiology* **150**:3613–3630.
50. Tacon, W., and D. Sherratt. 1976. ColE plasmid replication in DNA polymerase I-deficient strains of *Escherichia coli*. *Mol. Gen. Genet.* **147**:331–335.
51. Tajima, Y., T. Horii, and T. Itoh. 1988. Replication of ColE2 and ColE3 plasmids: two ColE2 incompatibility functions. *Mol. Gen. Genet.* **214**:451–455.
52. Takayama, G., T. Kosuge, H. Maseda, A. Nakamura, and T. Hoshino. 2004. Nucleotide sequence of the cryptic plasmid pTT8 from *Thermus thermophilus* HB8 and isolation and characterization of its high-copy-number mutant. *Plasmid* **51**:227–237.
53. Takechi, S., and T. Itoh. 1995. Initiation of unidirectional ColE2 DNA replication by a unique priming mechanism. *Nucleic Acids Res.* **23**:4196–4201.
54. Takechi, S., H. Matsui, and T. Itoh. 1995. Primer RNA synthesis by plasmid-specified Rep protein for initiation of ColE2 DNA replication. *EMBO J.* **14**:5141–5147.
55. Takechi, S., H. Yasueda, and T. Itoh. 1994. Control of ColE2 plasmid replication: regulation of Rep expression by a plasmid-coded antisense RNA. *Mol. Gen. Genet.* **244**:49–56.
56. Tomizawa, J., and T. Itoh. 1981. Plasmid ColE1 incompatibility determined by interaction of RNA I with primer transcript. *Proc. Natl. Acad. Sci. USA* **78**:6096–6100.
57. Vargas, C., G. Tegos, C. Drainas, A. Ventosa, and J. J. Nieto. 1999. Analysis of the replication region of the cryptic plasmid pHE1 from the moderate halophile *Halomonas elongata*. *Mol. Gen. Genet.* **261**:851–861.
58. Watson, R., W. Rowsome, J. Tsao, and L. P. Visentin. 1981. Identification and characterization of Col plasmids from classical colicin E-producing strains. *J. Bacteriol.* **147**:569–577.
59. Watson, R. J., T. Vernet, and L. P. Visentin. 1985. Relationships of the Col plasmids E2, E3, E4, E5, E6, and E7: restriction mapping and colicin gene fusions. *Plasmid* **13**:205–210. (Erratum, **14**:97.)
60. Yasueda, H., T. Horii, and T. Itoh. 1989. Structural and functional organization of ColE2 and ColE3 replicons. *Mol. Gen. Genet.* **215**:209–216.
61. Yoo, J. S., H. S. Kim, S. Y. Chung, Y. C. Lee, Y. S. Cho, and Y. L. Choi. 2001. Characterization of the small cryptic plasmid, pGD2, of *Klebsiella* sp. KCL2. *J. Biochem. Mol. Biol.* **34**:584–589.
62. Zhao, Y., Z. Ma, and G. W. Sundin. 2005. Comparative genomic analysis of the pPT23A plasmid family of *Pseudomonas syringae*. *J. Bacteriol.* **187**:2113–2126.