Molec. gen. Genet. 172, 151–159 (1979)

MGG
© by Springer-Verlag 1979

# Nucleotide Sequence of Small ColE1 Derivatives: Structure of the Regions Essential for Autonomous Replication and Colicin E1 Immunity

Atsuhiro Oka, Nobuo Nomura, Masayuki Morita, Hiroyuki Sugisaki, Kazunori Sugimoto, and Mituru Takanami

Institute for Chemical Research, Kyoto University, Uji-shi, Kyoto-fu 611, Japan

**Summary.** A small ColE1 derivative, pAO2, which replicates like the original ColE1 and confers immunity to colicin E1 on its host cell has been constructed from a quarter region of ColE1 DNA (Oka, 1978). The entire nucleotide sequence of pAO2 (1,613 base pairs) was determined based on its fine cleavage map. The sequence of a similar plasmid, pAO3, carrying additional 70 base pairs was also deduced.

The sequence in the region covering the replication initiation site on these plasmids was consistent with those reported for ColE1 by Tomizawa et al. (1977) and by Bastia (1977). DNA sequences indispensable for autonomous replication were examined by constructing plasmids from various restriction fragments of pAO2 DNA. As a result, a region of 436 base pairs was found to contain sufficient information to permit replication. The occurrence of initiation and termination codons and of the ribosome-binding sequence on pAO2 DNA suggests that a polypeptide chain consisting of 113 amino acid residues may be encoded by the region in which the colicin E1 immunity gene has been mapped.

## Introduction

The colicin E1 plasmid (ColE1) is a closed circular DNA molecule with a molecular weight of 4.2 megadaltons that determines the production of an inducible antibiotic protein, colicin E1, in bacteria and confers immunity to colicin E1 upon its host (Bazaral and Helinski, 1970). Since the ColE1 DNA replication continues in the presence of chloramphenicol till more than 1,000 copies per cell accumulate (Clewell, 1972), ColE1 and its derivatives have been used as a molecular vehicle for cloning DNA (Hershfield et al., 1974; Itakura et al., 1977). The cleavage maps of ColE1 and its deletion derivatives have been constructed with several restriction endonucleases (Oka and Takanami, 1976; Tomizawa et al., 1977; Oka, 1978), and some genetic markers have been localized on the map by analysis of transposon-insertion mutants (So et al., 1975; Dougan and Sherratt, 1977; Inselburg, 1977a) and of deletion mutants (Inselburg, 1977b; Oka, 1977). Because of its small size and relaxed mode of replication, the mechanism involved in the ColE1 DNA replication has extensively been studied (Bazaral and Helinski, 1970; Sakakibara and Tomizawa, 1974; Oka and Inselburg, 1975; Tomizawa et al., 1975). The nucleotide sequence in the region covering the initiation site of replication has also been determined (Tomizawa et al., 1977; Bastia, 1977). Nevertheless, the DNA region essential for the replication of ColE1 DNA is not yet defined. Nor is the structure of any other genes elucidated at the nucleotide sequence level.

We previously demonstrated that the fragment derived from 0.75 to 1.00 map unit[1] of ColE1 DNA can replicate as a plasmid after in vitro circularization (Oka, 1978). This plasmid, pAO2 (Fig. 1), still confers immunity to colicin E1 on the host cell, indicating that genes responsible for replication (*ori*[2]) and colicin E1 immunity (*imm*) are contained within the region covered by pAO2. In this paper, the entire nucleo-

*For offprints contact:* Dr. Atsuhiro Oka

*Abbreviations.* ColE1, colicin E1 plasmid; Tris, tris-(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetate; dNTP, deoxyribonucleoside triphosphates; ATP, adenosine 5′-triphosphate.

[1] Map unit is the fractional length from the 5′ end of the H-strand (Tomizawa, 1975) generated by cleavage of ColE1 DNA by *Eco*RI, and therefore, the *Eco*RI site is at 0.00 (1.00) map unit

[2] *ori* was defined as the smallest DNA region capable of replicating autonomously. It may contain not only the DNA region required for the replication initiation but also that for the replication termination

0026-8925/79/0172/0151/$01.80

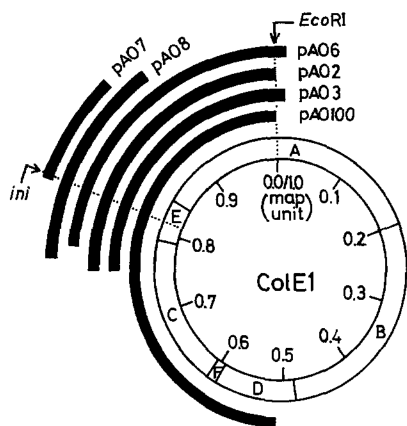


**Fig. 1.** Deletion map of small ColE1 derivatives. Heavy lines show DNA regions carried by each deletion mutant. pAO100, pAO7 and pAO8 carry non-ColE1 DNA in addition to the indicated regions (see text and Table 1). ColE1 DNA replication initiates at *ini* and proceeds counterclockwise (Tomizawa et al., 1977). *Hae*II cleavage sites and resulting fragments A to F (Oka and Takanami, 1976) are indicated as reference in addition to map unit

tide sequence of pAO2 DNA was determined based on its fine cleavage map previously constructed (Oka, 1978). pAO3 (Fig. 1) which is similar to pAO2 but carries additional 70 base pairs derived from 0.00 to 0.01 map unit of ColE1 was also sequenced. The loci of *ori* and *imm* were assigned on the sequence, based on analyses of DNA regions indispensable for autonomous replication and of initiation and termination codons.

## Materials and Methods

*Escherichia coli Strains and Plasmids*

An *E. coli* K-12 strain C600 (*thr leu thi tonA lac supE*) was used throughout all experiments unless otherwise noted. GM31 (Marinus, 1973) and A19 (Gesteland, 1966) are *dcm* (DNA cytosine methylation) and *rna* (ribonuclease I) mutants, respectively, Their plasmid-carrying derivatives were constructed by transformation. Plasmids used were ColE1 derived from P678–54(ColE1) (Oka and Inselburg, 1975) and its derivatives pAO100, pAO2, pAO3, pAO43 and pAO48. pAO100 (Oka, 1978) has been constructed by circularization of the mini-ColE1 segment produced by digestion of pML21 (Hershfield et al., 1976) DNA with *Eco*RI and is composed of the region from 0.51 to 1.00 map unit of ColE1 and a non-ColE1 DNA of about 300 base pairs. This is therefore identical with pVH51 (Hershfield et al., 1976). pAO2 (Oka, 1978) has been constructed from the largest *Hph*I-fragment of pAO100 which covers the region from 0.75 to 1.00 map unit of ColE1 after circularization by repair ligation (defined below). pAO3 was prepared from the largest *Hph*I-fragment of ColE1 by the same procedure. This covered the region from 0.75 to 0.01 map unit of ColE1 across the 0.00 (1.00) map unit site and thus contained the *Eco*RI site (Fig. 1). In contrast, pAO2 does not carry the *Eco*RI site, because an *Hph*I recognition sequence was present near the *Eco*RI site-proximal end of the non-ColE1 DNA of pAO100 and the *Eco*RI recognition sequence on pAO100 was disrupted by *Hph*I digestion.

pAO43 and pAO48 are chimeric plasmids produced by transposition of a kanamycin transposon Tn*903* from R6-5 to pAO2. The Tn*903* insertion sites in these plasmids are at 0.99 and 0.91 map unit on the pAO2 DNA molecule, respectively (Oka et al., 1978). pYT10 (Yasuda and Hirota, 1977) is a composite plasmid of pVH51 and an *Eco*RI-fragment of 4.5 megadaltons (Amp-fragment) carrying *bla* (ampicillin resistance). This Amp-fragment cannot replicate by itself.

*Enzymes*

The source and method for preparation of restriction endonucleases *Alu*I, *Eco*RI, *Hae*II, *Hae*III, *Hap*II, *Hga*I, *Hha*I, *Hin*fI and *Hph*I were described previously (Takanami, 1973; Oka, 1978). *Bpa*I and *Bpa*II were isolated from *Bordetella parapertussis* by the method essentially identical with that for *Hap*II (Takanami, 1973). *Ava*II and *Eco*RII were purchased from Bethesda Research Lab. DNA was digested in a reaction buffer containing 10 mM Tris-HCl (pH 7.6)/7 mM $MgCl_2$/7 mM 2-mercaptoethanol. For digestion with *Ava*II, *Bpa*I, *Bpa*II and *Eco*RI, 30 mM NaCl, 50 mM KCl, 50 mM KCl and 150 mM NaCl were added to the reaction buffer, respectively. DNA polymerase, DNA ligase and polynucleotide kinase were prepared from A19 cells infected with T4 amN82 according to the methods reported by Lehman (1974), Weiss (1971) and Richardson (1965), respectively.

*Preparation of Plasmid DNA and Its Restriction Fragments*

*E. coli* cells carrying a plasmid were grown to $3 \times 10^8$ cells per ml in E-broth (10 g polypeptone/1 g yeast extract/10 g glucose/1 g $NH_4Cl$/3 g NaCl/0.1 g $Na_2SO_4$/0.08 g $MgCl_2 \cdot 6H_2O$/15.2 g $Na_2HPO_4 \cdot 12H_2O$/3 g $KH_2PO_4$ in one liter), and chloramphenicol (Sankyo) was added to 80 µg per ml. After incubation at 37° C for 15 h, cells were harvested and the fraction containing covalently closed circular DNA molecules were prepared as described (Oka, 1978). Analysis of plasmid DNA was carried out with 1% agarose (Dohjin) gel columns (0.6 cm by 12 cm) in a running buffer containing 40 mM Tris-acetic acid (pH 7.2)/20 mM sodium acetate/2 mM EDTA (pH 7.2). DNA bands were visualized by staining with ethidium bromide. For preparation of restriction fragments, plasmid DNA was digested with restriction endonucleases and resolved by electrophoresis on 5% polyacrylamide gel columns with a running buffer composed of 90 mM Tris/90 mM boric acid/2.5 mM EDTA-2Na, and DNA fragments were extracted from band regions as described (Oka, 1978).

*Ligation*

Joining of restriction fragments generated by different restriction endonucleases and circularization of a fragment carrying different ragged ends were performed by the following procedure termed repair ligation. The ragged ends of restriction fragments were converted to blunt ends by $3' \rightarrow 5'$ exonucleolysis with T4 DNA polymerase in the absence of dNTP, followed by $5' \rightarrow 3'$ polymerization with the same enzyme in the presence of four dNTP's. The resulting fragments were then joined by T4 DNA ligase, since this enzyme can join DNA fragments with blunt ends as well as those with matched cohesive ends (Sgaramella et al., 1970).

For preparation of fragments with blunt ends, about 10 pmol of a restriction fragment carrying ragged ends was mixed with 10 µg of T4 DNA polymerase in 0.5 ml of a reaction buffer containing 7 mM $MgCl_2$/100 mM KCl/7 mM dithiothreitol/50 mM Tris-HCl (pH 8.0). After incubation at 12° C for 10 min, four dNTP's were added to 0.4 mM each and incubation was continued at 12° C

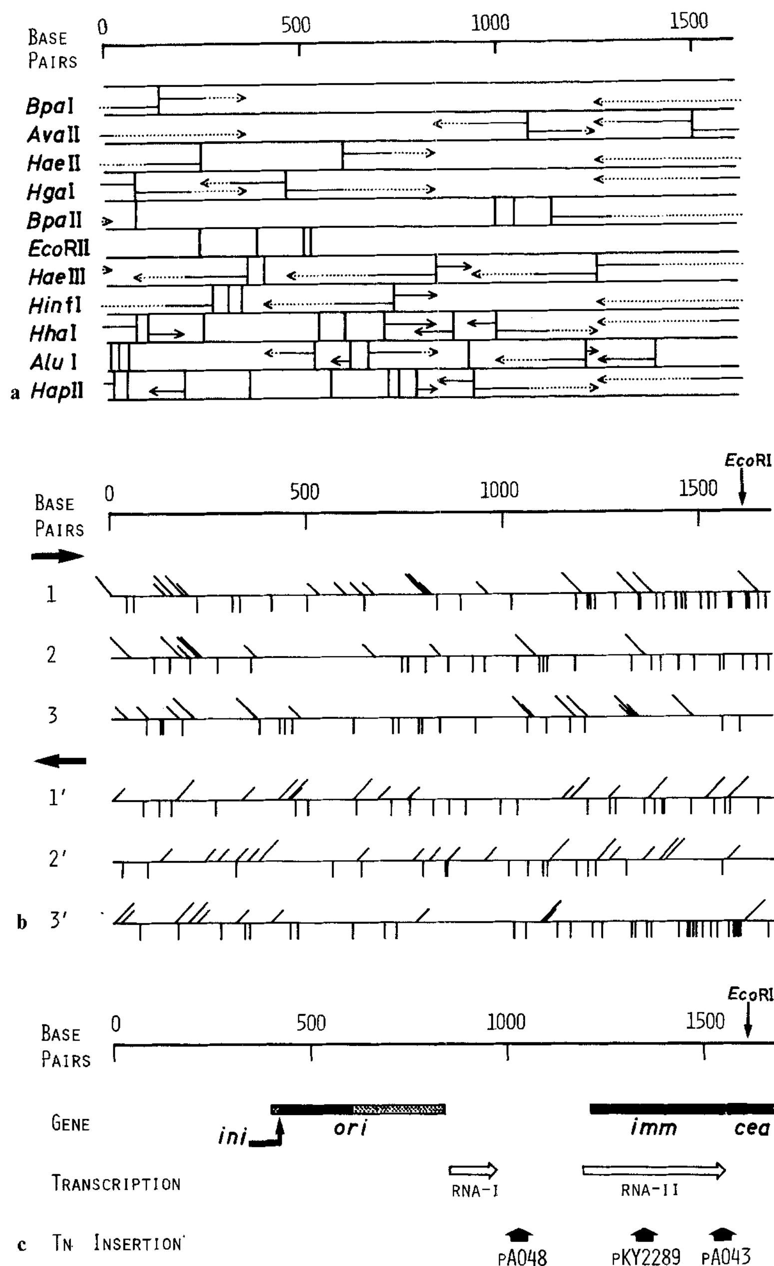


**Fig. 2. a** Fine cleavage maps of pAO2 and labeled strands used for sequence analysis. The maps were oriented with respect to the ligated point of pAO2, and the 5′ end of the H-strand is assigned to the left end. Therefore, the left and right ends correspond to 0.75 and 1.00 map unit of ColE1 DNA, respectively. The number in base pairs corresponds to nucleotide positions on the sequence in Fig. 3. Cleavage sites by eleven restriction endonucleases are illustrated by vertical lines. Arrows indicate labeled strands aligned in the 5′ to 3′ direction. All the strands were obtained by secondary cleavage of labeled double-stranded DNA. Solid line in each arrow denotes the region in which the sequence was deduced. **b** Codon analysis of the pAO3 sequence for protein synthesis. Initiation and termination codons for protein synthesis were plotted on each of three plausible frames on the sequence. Long and short angled-lines represent initiation codons ATG and GTG, respectively, and vertical lines represent termination codons TAG, TAA and TGA. **c** Loci of *ini*, *ori*, *imm*, *cea*, two transcripts in vitro and Tn insertion on the pAO3 sequence. *ini*, initiation site of ColE1 DNA replication (position 418). *ori*, DNA regions essential for autonomous replication of ColE1. Since *ini* is presumably indispensable for the *ori* function, and no plasmid has been constructed with *Alu*I-A and *Hae*II-B of pAO2 (see Table 1), the left end of *ori* is assumed to be ended within positions 403–418 and the right end of *ori* within positions 603–838, respectively. These regions were indicated by shading. *imm*, the gene conferring colicin E1 immunity. *cea*, the structural gene for colicin E1 protein. RNA-I (positions 863–972) and RNA-II, in vitro transcript from pAO2 and ColE1 DNA (Morita and Oka, 1979). Open arrows represent the direction of transcription. Vertical arrows, Tn insertion sites in pAO43, pAO48 (Oka et al., 1978), and pKY2289 (Maeda et al., 1978; Takeya et al., 1978)

for 60 min. The reaction was terminated by treatment with 80% phenol and DNA fragment was isolated by passage through a Sephadex G100 column (0.6 cm by 20 cm).

The following conditions were used for the construction of the *ori*-containing recombinants since the formation of linear or circular products of ligation is a function of the concentration of DNA fragments (Dugaiczyk et al., 1975). A restriction fragment containing *ori* (about 5 pmol) and Amp-fragment purified from pYT10 (about 0.5 pmol), both of which carried blunt ends, were mixed in 0.1 ml of a ligation buffer (7 mM $MgCl_2$/0.1 mM ATP/10 mM dithiothreitol/67 mM Tris-HCl; pH 7.6) containing 2 units of ligase, and incubation was carried out at 12° C for 3 h (1st reaction). The reaction mixture was then diluted ten-folds by the ligation buffer containing 15 units of ligase per ml, and incubation was continued at 12° C for additional 12 h (2nd reaction). When an *ori*-containing fragment was circularized by itself, the conditions used for the 2nd reaction were applied. The reaction mixture was deproteinized by shaking with 80% phenol and ligation products were isolated by passage through a Sephadex G100 column (0.6 cm by 20 cm), which has been equilibrated with 1 mM Tris-HCl (pH 7.5). Sephadex used was sterilized in advance by immersing in 0.1 N NaOH for 1 h.

### *Transformation*

Bacteria were made competent for transformation as follows (Lederberg and Cohen, 1974). Recipient cells grown to $2 \times 10^8$ cells per ml in L-broth (10 g polypeptone/5 g yeast extract/5 g NaCl/1 g glucose in one liter; pH 7.3) were washed once with cold 0.1 M $MgCl_2$, suspended in 0.1 M $CaCl_2$ and held at 0° C for 20 min. Cells were collected by centrifugation and resuspended in cold 0.1 M $CaCl_2$ of one-twentieth of the original culture volume (competent cells).

For transformation, 0.1 ml of plasmid DNA in 1 mM Tris-HCl (pH 7.5) was added to 0.2 ml of the competent cells in an ice bath. After holding at 0° C for 30 min, the mixture was subjected to a heat pulse at 42° C for 2 min and then chilled. Before spreading on selective agar plates, 5 ml of L-broth was added and incubation was carried out with aeration at 37° C for 1 h to allow phenotypic expression. Drug resistance was assayed on L-agar containing 20 µg of ampicillin (Takeda) per ml. Cells immune to colicin E1 were selected on a lawn of C600(ColE1) cells which had been grown overnight and killed with chloroform vapor.

### *Nucleotide-Sequence Determination*

The 5′ ends of restriction fragments were phosphorylated with [$\gamma^{32}$P]ATP (about 4 Ci per µmol; New England Nuclear Co.) and polynucleotide kinase. After separation of the two labeled ends by secondary cleavage with a different restriction endonuclease, the sequence of the 5′$^{32}$P-labeled DNA strand was determined by the method of Maxam and Gilbert (1977).

## Results and Discussion

### *The Nucleotide Sequences of pAO2 and pAO3*

The cleavage map of pAO2 DNA with seven restriction endonucleases, *Alu*I, *Hae*II, *Hae*III, *Hap*II, *Hga*I, *Hha*I and *Hin*fI, has previously been reported (Oka, 1978). In addition, the restriction sites of *Ava*II, *Bpa*I and *Bpa*II were determined by the procedure described earlier (Oka and Takanami, 1976). The cleavage sites for *Eco*RII were identified by using pAO2 DNA prepared from a *dcm* strain, GM31(pAO2). The cleavage sites with all these enzymes are summarized in Fig. 2a. Based on this map, the sequences of various restriction fragments were determined to construct the entire nucleotide sequence of pAO2 by the method of Maxam and Gilbert (1977). DNA strands analysed are indicated by arrows in Fig. 2a.

As shown in Fig. 1, pAO3 is composed of the whole pAO2 DNA sequence and an additional 70 base pair segment derived from 0.00 to 0.01 map unit of ColE1 DNA. This was confirmed by restriction analysis: pAO3 contained only two cleavage sites, the *Eco*RI site at 0.00 map unit and an *Alu*I site, in addition to those on pAO2. Thus the sequence of pAO3 was examined by applying the cleavage map of pAO2.

The sequence was verified by duplicate analyses of the same fragment or by analysis of overlapped fragments until unambiguous results were obtained. The nucleotide sequences of pAO2 (1,613 base pairs) and pAO3 (1,683 base pairs) thus deduced are shown in Fig. 3. Nucleotides were numbered in the 5′ to 3′ direction of the H-strand (Tomizawa, 1975) from the ligated point of pAO2. The nucleotide sequence in positions 1480–1683 was also confirmed by analysis of the original ColE1 DNA. The sequence surrounding the replication initiation site (*ini*)(position 418; see below) was consistent with that reported for ColE1 by Tomizawa et al. (1977) (positions 244–606) except for the nucleotide in position 248, and that by Bastia (1977)(positions 198–473) except for the nucleotides in positions 243, 364 and 386. Since the nucleotides in position 243 and 386 are located within the *Eco*RII recognition site and their methylated cyto-

**Fig. 3.** Nucleotide sequence of pAO2 and pAO3. pAO2 and pAO3 were composed of 1,613 and 1,683 base pairs, respectively. The nucleotides are contiguous in positions 1 and 1,613 in pAO2, and in positions 1 and 1,683 in pAO3. The sequence was shown in the same orientation as in Fig. 2. The upper and lower sequences are H- and L-strands, respectively. The recognition sites of twelve restriction endonucleases, expressed by nucleotide number at the 5′ end on the H-strand, are as follows: *Bpa*I (GT$^{AT}_{CG}$AC), 140; *Ava*II (GG$^{T}_{A}$CC), 1077, 1498; *Hae*II (RGCGCY), 243, 602; *Hga*I (GACGC, GCGTC), 75, 460; *Bpa*II (CGCG), 74, 987, 1038, 1134; *Eco*RII (CC$^{T}_{A}$GG), 240, 385, 506, 519; *Hae*III (GGCC), 358, 401, 837, 1250; *Hin*fI (GANTC), 269, 310, 344, 729; *Hha*I GCGC), 73, 103, 244, 536, 603, 703, 879, 988; *Alu*I (AGCT), 10, 29, 54, 525, 615, 661, 920, 1217, 1393, 1665; *Hap*II (CCGG), 15, 49, 195, 360, 565, 712, 738, 782, 930; and *Eco*RI (GAATTC), 1611

```
                                                                                                 100
GACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTGAGCGGATGCCGGGAGCTGACAAGCCCGTCAGGGCGCGTCAGCAGGTTTTAGCGGGTGTCG
CTGTGTACGTCGAGGGCCTCTGCCAGTGTCGAACAGACACTCGCCTACGGCCCTCGACTGTTCGGGCAGTCCCGCGCAGTCGTCCAAAATCGCCCACAGC

                                                                                                 200
GGGCGCAGCCCTGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTTAACCATGCGGCATCAGTGCGGATTGTATGAAAAGTACGCCATGCCGGGT
CCCGCGTCGGGACTGGGTCAGTGCATCGCTATCGCCTCACATATGACCGAATTGGTACGCCGTAGTCACGCCTAACATACTTTTCATGCGGTACGGCCCA

                                                                                                 300
GTGAAATGCCGCACAGATGCGTAAGGAGAAAATGCACGTCCAGGCGCTTTTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGCG
CACTTTACGGCGTGTCTACGCATTCCTCTTTTACGTGCAGGTCCGCGAAAAGGCGAAGGAGCGAGTGACTGAGCGATGCGAGCCAGCAAGCTGACGCCGC

                                                                                                 400
AGCGGTACTGACTCACACAAAAACGGTAACACAGTTATCCACAGAATCAGGGGATAAGGCCGGAAAGAACATGTGAGCAAAAGACCAGGAACAGGAAGAA
TCGCCATGACTGAGTGTGTTTTTGCCATTGTGTCAATAGGTGTCTTAGTCCCCTATTCCGGCCTTTCTTGTACACTCGTTTTCTGGTCCTTGTCCTTCTT

                                                                                                 500
GGCCACGTAGCAGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATAGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAA
CCGGTGCATCGTCCGCAAAAAGGTATCCGAGGCGGGGGGACTGCTCGTAGTGTTTTTATCTGCGAGTTCAGTCTCCACCGCTTTGGGCTGTCCTGATATT

                                                                                                 600
AGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCG
TCTATGGTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCCTATGGACAGGCGGAAAGAGGGAAGCCCTTCGC

                                                                                                 700
TGGCGCTTTCTCATAGCTCACGCTGTTGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCG
ACCGCGAAAGAGTATCGAGTGCGACAACCATAGAGTCAAGCCACATCCAGCAAGCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGC

                                                                                                 800
CTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGGCACGCCTTAACGCCACTGGCAGCAGCCACTGGTAACCGGATTAGCAGAGCGATG
GACGCGGAATAGGCCATTGATAGCAGAACTCAGGTTGGGCCATTCCGTGCGGAATTGCGGTGACCGTCGTCGGTGACCATTGGCCTAATCGTCTCGCTAC

                                                                                                 900
ATGGCACAAACGGTGCTACAGAGTTCTTGAAGTAGTGGCCCGACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTAC
TACCGTGTTTGCCACGATGTCTCAAGAACTTCATCACCGGGCTGATGCCGATGTGATCTTCCTGTCATAAACCATAGACGCGAGACGACTTCGGTCAATG

                                                                                                1000
CTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGTTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAA
GAAGCCTTTTTCTCAACCATCGAGAACTAGGCCGTTTGTTTGGTGGCAACCATCGCCACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTT

                                                                                                1100
GGATCTCAAGAAGATCCTTTAATCTTTTCTACTGAACCGCGATCCCCGTCAGTTTAGAAGAGGAGGATGGTGCGATGGTCCCTCCCTGAACATCACGTAT
CCTAGAGTTCTTCTAGGAAATTAGAAAAGATGACTTGGCGCTAGGGGCAGTCAAATCTTCTCCTCCTACCACGCTACCAGGGAGGGACTTGTAGTGCATA

                                                                                                1200
ATAGTTAGCCTGACATCCAACAAGGAGGTTTATCGCGAATATTCCCACAAAAAATCTTTTCCTCATAACTCGATCCTTATAAAATGAAAAGAATATATGG
TATCAATCGGACTGTAGGTTGTTCCTCCAAATAGCGCTTATAAGGGTGTTTTTTAGAAAAGGAGTATTGAGCTAGGAATATTTTACTTTTCTTATATACC

                                                                                                1300
CGAGGTTTAATTTATGAGCTTAAGATACTACATAAAAAATATTTATTTGGCCTGTACTGCACACTTATATATATATACCTTATAACAAAAAACAGCGAA
GCTCCAAATTAAATACTCGAATTCTATGATGTATTTTTTATAAAATAAACCGGACATGACGTGTGAATATATATATATGGAATATTGTTTTTTGTCGCTT

                                                                                                1400
GGGTATTATTTCCTTGTGTCAGATAAGATGCTATATGCAATAGTGATAAGCACTATTCTATGTCCATATTCAAAATATGCTATTGAATACATAGCTTTTA
CCCATAATAAAGGAACACAGTCTATTCTACGATATACGTTATCACTATTCGTGATAAGATACAGGTATAAGTTTTATACGATAACTTATGTATCGAAAAT

                                                                                                1500
ACTTCATAAAGAAAGATTTTTTCGAAAGAAGAAAAAACCTAAATAACGCCCCCGTAGCAAAATTAAACCTATTTATGCTATATAATCTACTTTGTTTGGT
TGAAGTATTTCTTTCTAAAAAAGCTTTCTTCTTTTTTGGATTTATTGCGGGGGCATCGTTTTAATTTGGATAAATACGATATATTAGATGAAACAAACCA

                                                                                                1600
CCTAGCAATCCCATTTGGATTGCTAGGACTTTTTATATCAATAAAGAATAATTAAATCCCTAACACCTCATTTATAGTATTAAGTTTATTCTTATCAATA
GGATCGTTAGGGTAAACCTAACGATCCTGAAAAATATAGTTATTTCTTATTAATTTAGGGATTGTGGAGTAAATATCATAATTCAAATAAGAATAGTTAT

TAGGAGCATAGAATTCCTGTAACAATAGCAATACCCCAAATACCTAATGTAGTTCCAGCAAGCAAGCTAAAAAGTAAAGCAAC  3' H
ATCCTCGTATCTTAAGGACATTGTTATCGTTATGGGGTTTATGGATTACATCAAGGTCGTTCGTTCGATTTTTCATTTCGTTG  5' L
```

sine is not identified by the Maxam and Gilbert method (Ohmori et al., 1978), our data in these positions obtained using the *dcm* mutant would be correct. The differences in other positions are not elucidated.

### *DNA Regions Indispensable for Autonomous Replication (ori)*

The initiation site of ColE1 DNA replication as the transition point from RNA to DNA (*ini*) has been determined by Tomizawa et al. (1977). The site is position 418 on the sequence in Fig. 3. However, DNA regions essential for autonomous replication (*ori*) have not yet been determined. In order to approach to this question, we looked for the smallest segment of ColE1 DNA that can replicate autonomously. Ligation products of restriction fragments derived from ColE1 and pAO2 were prepared under the conditions in which circular forms are predominantly produced. In addition, recombinants between these fragments and the non-replicating Amp-fragment which provides a selective marker were prepared by repair ligation. The products were then used to transform C600 cells, and each transformant colony produced was inspected to carry an expected plasmid. The restriction fragments used and plasmids constructed were listed in Table 1. Electrophoretic patterns of these plasmids on agarose gel are presented in Fig. 4. Among these plasmids, the one containing the smallest ColE1 segment was pAO7 which was constructed from *Hae*III-C (436 base pairs) of pAO2 and Amp-fragment. Digestion of pAO7 DNA by *Eco*RI yielded two fragments with sizes corresponding to *Hae*III-C and Amp-fragment, as expected[3] (Fig. 5c). Plasmids containing dimer and trimer of *Hae*III-C were also isolated (pAO7-2 and pAO7-3, respectively) (Fig. 4g, h, Fig. 5d, e). Thus it was concluded that the 436 base pair segment carried the information enough for autonomous replication, though pAO7 was not so stable as ColE1 and pAO2 in a bacterial cell. This segment contains only 16 base pairs downstream, but 420 base pairs upstream, from *ini*. So far we could not construct plasmids with *Alu*I-A and *Hae*II-B of pAO2 which respectively contain 108 and 184 base pairs upstream from *ini* (Table 1). The result suggests that a relatively long sequence in the upstream region is involved in *ori* (see the legend to Fig. 2c). This view is consistent with the result reported by Bolivar et al. (1977). They compared the nucleotide sequences covering *ini* of pBR345 and ColE1, which are considered to have been evolved from a common ancestor, and found that the sequence was well conserved in the upstream region, but not in the downstream region, of *ini*.

[3] An *Eco*RI susceptible site should have been created at the junction between a *Hae*III-fragment and a repaired *Eco*RI-fragment

**Table 1.** Construction of small ColE1 derivatives

| Original restriction fragment (Nucleotide position[a]) | Plasmid | (Selection[b]) |
|---|---|---|
| *Hph*I–A*[c] of pAO100 (1–1613) | pAO2 | (Imm$^+$) |
| *Hph*I–A* of ColE1 (1–1683) | pAO3 | (Imm$^+$) |
| *Hph*I–A·*Bpa*I–1* of ColE1 (142–1683) | pAO6 | (Imm$^+$) |
| *Bpa*II–A of pAO2 (76–988) and Amp-fragment* | pAO8 | (Amp$^r$) |
| *Hae*III–C[d] of pAO2 (403–838) and Amp-fragment* | pAO7 | (Amp$^r$) |
| *Hae*II–B* of pAO2 (248–602) and Amp-fragment* | N.I.[e] | (Amp$^r$) |
| *Alu*I–A of pAO2 (56–526) and Amp-fragment* | N.I. | (Amp$^r$) |

[a] Nucleotide position corresponds to that of Fig. 3
[b] Selection indicates selective marker used in transformation
[c] A fragment carrying repaired blunt ends was indicated by an asterisk
[d] Though *Hae*III-C of pAO2 migrates faster than *Hae*III-B of pAO2 on polyacrylamide gel electrophoresis (Oka, 1978), *Hae*III-C was longer than *Hae*III-B by 23 base pairs
[e] N.I. means that no plasmid has been isolated

It has been shown that the ColE1 DNA replication occurs in the absence of ColE1-encoded proteins (Tomizawa et al., 1975; Donoghue and Sharp, 1978; Kahn and Helinski, 1978). In accordance with this, the *Hae*III-C sequence has no capacity coding for a long polypeptide chain[4], as judged from the occurrence of initiation (ATG) and termination codons (TGA, TAA, TAG) in all possible frames (Fig. 2b).

### *The Gene Conferring Colicin E1 Immunity (imm)*

The map position of *imm* has been allocated around 0.96 map unit of ColE1 DNA (Inselburg, 1977b; Oka, 1977). A more striking evidence on the location of *imm* was provided by Maeda et al. (1978), who isolated a Tn*3*-insertion derivative (pKY2289) of ColE1 lacking the ability to produce colicin E1 immunity. Its insertion site was located in positions 1344–1348[5] (Fig. 2c) (Takeya et al., 1978). However, no informa-

[4] *Hae*III-C can code for a polypeptide of 39 amino acid residues (frame-1′). Besides, if GTG were taken as the initiation codon, *Hae*-III-C can code for several polypeptide chains which do not exceed 52 amino acid residues. However, all these coding sequences do not seem to be preceded by potential ribosome-binding sequences

[5] The Tn insertion site was expressed by the nucleotide positions of direct repeats generated by the insertion; Tn*3* and Tn*903* generate five and nine base pair repeats, respectively (Takeya et al., 1978; Oka et al., 1978)

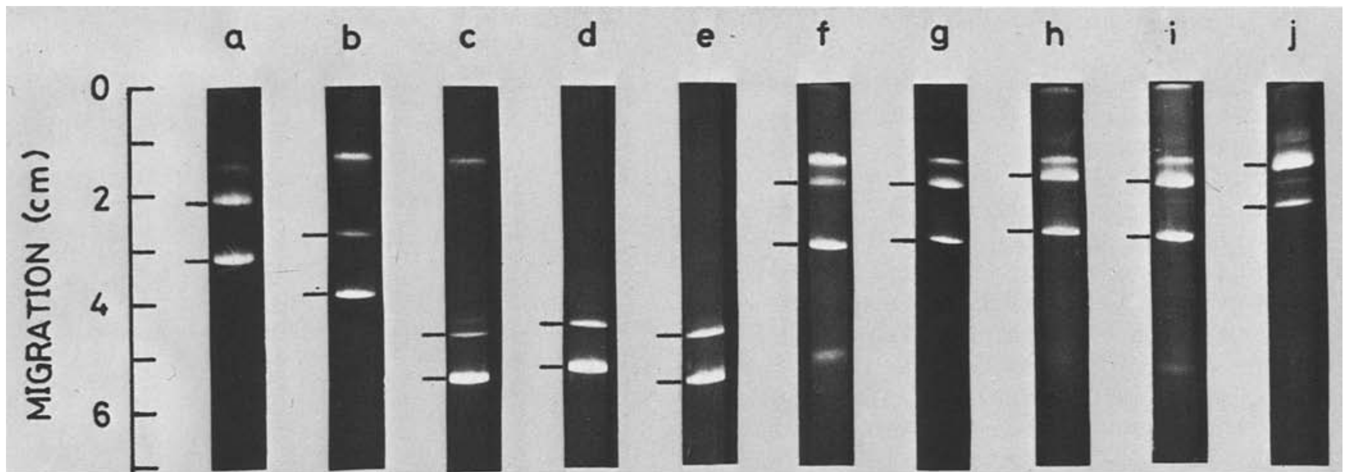


**Fig. 4.** 1% agarose gel electrophoretic patterns of ColE1 (a), pAO100 (b), pAO2 (c), pAO3 (d), pAO6 (e), pAO7 (f), pAO7-2 (g), pAO7-3 (h), pAO8 (i) and pYT10 (j). The faster and slower bands shown by bars correspond to covalently closed circular and open circular DNA molecules, respectively. The slowly moving components are bacterial DNA fragments

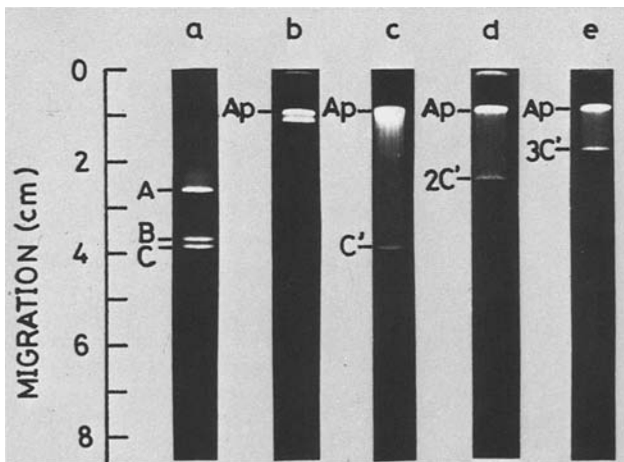


**Fig. 5.** 5% polyacrylamide gel electophoretic patterns of a *Hae*III digest of pAO2 (a), and *Eco*RI digests of pYT10 (b), pAO7 (c), pAO7-2 (d) and pAO7-3 (e). Ap represents the Amp-fragment and A, B and C represent *Hae*III-A, -B and -C of pAO2, respectively. C′, 2C′ and 3C′ indicate fragments having the sizes of monomer, dimer and trimer of *Hae*III-C of pAO2, respectively. These fragments should carry ragged ends produced by *Eco*RI, compared with *Hae*III-C having blunt ends. Digestion of the fragments 2C′ and 3C′ by *Hae*III yielded fragments with a size corresponding to that of *Hae*III-C of pAO2 (data not shown)

tion is available as to the size of the *imm* gene product. We have found that the insertion of Tn*903* into either positions 1023–1031 (pAO48) or positions 1537–1545 (pAO43) on the pAO2 sequence did not afect the level of colicin E1 immunity (Fig. 2c) (Oka et al., 1978). It is therefore likely that *imm* is flanked by these two insertion sites.

From the occurrence of initiation and termination codons in all the possible frames within the region between the two Tn*903* insertion sites, the sequences of positions 1214–1552 (frame-3), 1328–1552 (frame-3) and 1475–1552 (frame-3) in the left to right direction and those of positions 1166–1032 (frame-1′), 1407–1309 (frame-2′) and 1392–1309 (frame-2′) in the opposite direction can code for polypeptide chains of 113, 75, 26, 45, 33 and 28 amino acid residues, respectively (Fig. 2b). Among these possible coding sequences, two which occurred in positions 1214–1552 (frame-3) and in positions 1166–1032 (frame-1′) are accompanied by a potential ribosome-binding sequence (GAGGT in positions 1202–1206 and TAAGGA in positions 1179–1174, respectively) (Steitz and Jakes, 1975). Since the latter sequence does not cross the Tn*3* insertion site in pKY2289, the former sequence which can code for a polypeptide of 113 amino acid residues appears to be *imm*. This was further supported by analysis of transcriptional units on pAO2 DNA (Morita and Oka, 1979). In an in vitro transcriptional system, two unique sizes of RNA (RNA-I of 110 bases and RNA-II of about 360 bases) were synthesized from the L-strand of pAO2 as indicated in Fig. 2c. RNA-II appears to be just long enough to cover the above candidate sequence for *imm* (positions 1214–1552). If this assignment is correct, the carboxyl-terminal portion of the *imm* product may not be important for its function, because the Tn*903* insertion in pAO43 showing Imm$^+$ phenotype occurred at positions 1537–1545; according to our sequence data on pAO43 (Oka et al., 1978), two amino acid residues present at the carboxyl-terminus of the *imm* product should be replaced by nine amino acid residues coded by the Tn*903* sequence.

*The Structural Gene for Colicin E1 Protein (cea)*

It is known that when a foreign DNA fragment is inserted at the position of the *Eco*RI site on ColE1 DNA, the resulting plasmid loses the ability to produce colicin E1 (Hershfield et al., 1974). Therefore, it is likely that the cleavage with *Eco*RI occurs either in *cea* or in the regulatory region for the expression of *cea*. Genetic analysis of mutants lacking an ability to produce colicin E1 has suggested that *cea* is located between 0.25 map unit and the vicinity of the *Eco*RI site (So et al., 1975; Inselburg, 1977a, 1977b; Dougan and Sherratt, 1977). Furthermore, Yang and Zubay (1978) have provided evidence suggesting that the *Eco*RI site is located within the carboxyl-terminal region of the *cea* gene. If *cea* indeed crosses the *Eco*RI site, no termination codons in the right to left direction should occur on the right side of the *Eco*RI site of the pAO3 sequence shown in Fig. 2b. Only frame-2′ satisfies such a condition and a termination codon TAA appears at positions 1554–1552. It is of interest to note that after a short gap, *cea* is followed by a sequence CAATCCAAAT (positions 1522–1513) homologous to the rho-dependent termination site found in the *supF* gene (CAATCAAAT) (Küpper et al., 1978) and $\lambda_{tR1}$ (CAATCAATT) (Rosenberg et al., 1978).

Since pAO2 and pAO3 replicate in cells like the original ColE1, it is easy to prepare these DNA molecules in a large scale. The determination of the entire nucleotide sequence and assignment of the DNA region essential for autonomous replication now makes it possible to use these plasmids for various purposes; e.g., construction of new vectors, supply of accurate DNA size markers, determination of recognition sequence of new restriction endonucleases and analysis of insertion specificity of translocatable DNA elements.

*Acknowledgements.* We thank Drs. K. Shimada and T. Takeya for the exchange of unpublished information. This work was supported by Grant-in-Aid for Scientific Research No. 311207 to M. T. and No. 311209 to A. O. from the Ministry of Education, Science and Culture, Japan.

## References


Bastia, D.: The nucleotide sequence surrounding the origin of DNA replication of ColE1. Nucleic Acids Res. **4**, 3123–3142 (1977)

Bazaral, M., Helinski, D.R.: Replication of a bacterial plasmid and an episome in *Escherichia coli*. Biochemistry **9**, 399–406 (1970)

Bolivar, F., Betlach, M., Heyneker, H.L., Shine, J., Rodriguez, R.L., Boyer, H.W.: Origin of replication of pBR345 plasmid DNA. Proc. Natl. Acad. Sci. U.S.A. **74**, 5265–5269 (1977)

Clewell, D.B.: Nature of ColE1 plasmid replication in *Escherichia coli* in the presence of chloramphenicol. J. Bacteriol. **110**, 667–676 (1972)

Donoghue, D.J., Sharp, P.A.: Replication of colicin E1 plasmid DNA in vivo requires no plasmid-encoded proteins. J. Bacteriol. **133**, 1287–1294 (1978)

Dougan, G., Sherratt, D.: The transposon Tn1 as a probe for studying ColE1 structure and function. Mol. Gen. Genet. **151**, 151–160 (1977)

Dugaiczyk, A., Boyer, H., Goodman, H.M.: Ligation of *Eco*RI endonuclease-generated DNA fragments into linear and circular structures. J. Mol. Biol. **96**, 171–184 (1975)

Gesteland, R.F.: Isolation and characterization of ribonuclease I mutants of *Escherichia coli*. J. Mol. Biol. **16**, 67–84 (1966)

Hershfield, V., Boyer, H.W., Chow, L., Helinski, D.R.: Characterization of a mini-ColE1 plasmid. J. Bacteriol. **126**, 447–453 (1976)

Hershfield, V., Boyer, H.W., Yanofsky, C., Lovett, M., Helinski, D.R.: Plasmid ColE1 as a molecular vehicle for cloning and amplification of DNA. Proc. Natl. Acad. Sci. U.S.A. **71**, 3455–3459 (1974)

Inselburg, J.: Isolation, mapping and examination of effects of TnA insertions in ColE1 plasmids. J. Bacteriol. **129**, 482–491 (1977a)

Inselburg, J.: Studies of colicin E1 plasmid functions by analysis of deletions and TnA insertions of the plasmid. J. Bacteriol. **132**, 332–340 (1977b)

Itakura, K., Hirose, T., Crea, R., Riggs, A.D., Heyneker, H.L., Bolivar, F., Boyer, H.W.: Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. Science **198**, 1056–1063 (1977)

Kahn, M., Helinski, D.R.: Construction of a novel plasmid-phage hybrid: Use of the hybrid to demonstrate ColE1 DNA replication in vivo in the absence of a ColE1-specific protein. Proc. Natl. Acad. Sci. U.S.A. **75**, 2200–2204 (1978)

Küpper, H., Sekiya, T., Rosenberg, M., Egan, J., Landy, A.: A $\rho$-dependent termination site in the gene coding for tyrosine tRNA su3 of *Escherichia coli*. Nature **272**, 423–428 (1978)

Lederberg, E.M., Cohen, S.N.: Transfromation of *Salmonella typhimurium* by plasmid deoxyribonucleic acid. J. Bacteriol. **119**, 1072–1074 (1974)

Lehman, I.R.: T4 DNA polymerase. In: Methods in enzymology (Grossman, L., Moldave, K., eds.), Vol. 29, pp. 46–53. New York: Academic Press 1974

Maeda, S., Shimada, K., Takagi, Y.: Colicinogenic mutant of ColE1 plasmid that fails to confer immunity to colicin E1. Gene **3**, 1–7 (1978)

Marinus, M.G.: Location of DNA methylation genes on the *Escherichia coli* K-12 genetic map. Mol. Gen. Genet. **127**, 47–55 (1973)

Maxam, A.M., Gilbert, W.: A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**, 560–564 (1977)

Morita, M., Oka, A.: The structure of a transcriptional unit on colicin E1 plasmid. Eur. J. Biochem., in the press (1979)

Ohmori, H., Tomizawa, J., Maxam, A.M.: Deletion of 5-methylcytosine in DNA sequences. Nucleic Acids Res. **5**, 1479–1485 (1978)

Oka, A.: DNA regions essential for replication of colicin E1 plasmid. In: Progress in virology (Institute for Virus Research, Kyoto Univ., ed.), pp. 112–124 (1977)

Oka, A.: Fine cleavage map of a small colicin E1 plasmid carrying genes responsible for replication and colicin E1 immunity. J. Bacteriol. **133**, 916–924 (1978)

Oka, A., Inselburg, J.: Replicative intermediates of colicin E1 plasmid DNA in minicells. Proc. Natl. Acad. Sci. U.S.A. **72**, 829–833 (1975)

Oka, A., Nomura, N., Sugimoto, K., Sugisaki, H., Takanami,

M.: Nucleotide sequence at the insertion sites of a kanamycin transposon. Nature **276**, 845–847 (1978)

Oka, A., Takanami, M.: Cleavage map of colicin E1 plasmid. Nature **264**, 193–196 (1976)

Richardson, C.C.: Phosphorylation of nucleic acid by an enzyme from T4 bacteriophage-infected *Escherichia coli.* Proc. Natl. Acad. Sci. U.S.A. **54**, 158–165 (1965)

Rosenberg, M., Court, D., Shimatake, H., Brady, C., Wulff, D.L.: The relationship between function and DNA sequence in an intercistronic regulatory region in phage $\lambda$. Nature **272**, 414–423 (1978)

Sakakibara, Y., Tomizawa, J.: Replication of colicin E1 plasmid DNA in cell extracts. Proc. Natl. Acad. Sci. U.S.A. **71**, 802–806 (1974)

Sgaramella, V., Sande, J.H. van de, Khorana, H.G.: Studies on polynucleotides, C. A novel joining reaction catalized by the T4-polynucleotide ligase. Proc. Natl. Acad. Sci. U.S.A. **67**, 1468–1475 (1970)

So, M., Gill, R., Falkow, S.: The generation of ColE1-Ap$^r$ cloning vehicle which allows detection of inserted DNA. Mol. Gen. Genet. **142**, 239–249 (1975)

Steitz, J.A., Jakes, K.: How ribosome select initiator regions in mRNA: Base pair formation between the 3′ terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli.* Proc. Natl. Acad. Sci. U.S.A. **72**, 7434–4738 (1975)

Takanami, M.: Specific cleavage of coliphage fd DNA by five different restriction endonucleases from *Haemophilus* genus. FEBS Lett. **34**, 318-322 (1973)

Takeya, T., Shimada, K., Takagi, Y.: Translocation mechanism of Tn*3*. Jpn. J. Genet. (Abstr.) **53**, 466 (1978)

Tomizawa, J.: Two distinct mechanisms of synthesis of DNA fragments on colicin E1 plasmid DNA. Nature **257**, 253–254 (1975)

Tomizawa, J., Ohmori, H., Bird, R.E.: Origin of replication of colicin E1 plasmid DNA. Proc. Natl. Acad. Sci. U.S.A. **74**, 1865–1869 (1977)

Tomizawa, J., Sakakibara, Y., Kakefuda, T.: Replication of colicin E1 plasmid DNA added to cell extracts. Proc. Natl. Acad. Sci. U.S.A. **72**, 1050–1054 (1975)

Weiss, B.: DNA ligase from *Escherichia coli* infected with bacteriophage T4. In: Methods in enzymology (Grossman, L., Moldave, K., eds.), Vol. 21, pp. 319–326. New York: Academic Press 1971

Yang, H., Zubay, G.: Expression of the *cel* gene in ColE1 and certain hybrid plasmids derived from *Eco*RI-treated ColE1. In: Microbiology (Schlessinger, D., ed.), pp. 154–155. Washington, D.C.: American Society for Microbiology 1978

Yasuda, S., Hirota, Y.: Cloning and mapping of the replication origin of *Escherichia coli.* Proc. Natl. Acad. Sci. U.S.A. **74**, 5458–5462 (1977)

Communicated by T. Yura

Received December 26, 1978