# Identification of a plasmid-coded protein required for initiation of ColE2 DNA replication

Mitsuru Kido+, Hisashi Yasueda§ and Tateo Itoh*
Laboratory of Genetics, Department of Biology, Faculty of Science, Osaka University, Toyonaka, Osaka 560, Japan

Received March 19, 1991; Revised and Accepted May 13, 1991

## ABSTRACT

**The product of the *rep* gene of ColE2 is required for initiation of ColE2 DNA replication. The *rep* gene was placed under the control of the promoters, $P_L$ and $P_R$, and the heat-labile *cI857* repressor of bacteriophage λ. The Rep protein was identified as a 35 Kd protein by the maxicell method in combination with heat-induced expression. The protein was efficiently expressed from these promoters in unirradiated cells and accumulated up to a few per cent of the total cellular proteins. It was partially purified (about 80% pure) and its properties examined. The amino acid sequence of the amino terminal portion of the partially purified protein agreed well with that predicted from the nucleotide sequence of the *rep* gene. One of the characteristic features of the *rep* gene is frequent usage of rare codons, especially those for arginine. The protein specifically stimulated replication of ColE2 DNA but not that of ColE3 DNA in crude cell extracts of *Escherichia coli*. Specific binding of the protein to plasmid DNA containing the origin region of ColE2 was demonstrated by the filter binding method. Neither endonuclease activity nor topoisomerase activity was detected by using ColE2 DNA.**

## INTRODUCTION

ColE2 and ColE3 plasmids are small multicopy plasmids of about 7 kb (1). These plasmids require host DNA polymerase I for replication (2, 3). The 1.3-kb regions of these plasmids are sufficient for autonomous replication (4) and their nucleotide sequences are quite homologous (5), indicating that these plasmids share common mechanisms for initiation of replication and its control. Each plasmid requires a plasmid-specific *trans*-acting protein (Rep) for autonomous replication. The protein seems to be specified by an open reading frame which can encode a protein of about 300 amino acids. The *cis*-acting region (origin) required for initiation of replication consists of 32 bp for ColE2 or 33 bp for ColE3. These plasmids can replicate in crude cell extracts of *Escherichia coli* (6). Replication in this system depends upon DNA polymerase I and other host proteins along with the plasmid-coded protein which acts in a plasmid-specific manner. Replication starts at a fixed region containing the cloned origin region and proceeds in a fixed direction.

To study mechanisms of initiation of ColE2 DNA replication and its control, we partially purified the Rep protein and examined its properties.

## MATERIALS AND METHODS

### Bacterial strains and plasmids

The *Escherichia coli* strains used are as follows: *E. coli* B, 011′ *thyA deo* (7; obtained from K. Shimizu); *E. coli* K12, N1790 *recA99 uvrA54* (8) and NT525 *endA rna pnp* (9). Other bacterial strains have been described (4, 6). Plasmids pEC22, pEC22X43, pEC32, pET21, pET31 and pTH453 have been described (4, 6, 10). Plasmid pMY12-6 (11; obtained from T. Tsurimoto) carries the promoters, $P_L$ and $P_R$, and the *cI857* gene of bacteriophage λ. Plasmid pBZ201M9 (12; obtained from A. Markovitz) carries a mutated *lon* gene of *E. coli* which is dominant to the wild-type gene. Plasmid pACYC177-4A1 (13; obtained from L. Simon) carries the T4 *pin* gene.

### Construction of plasmids

Plasmid pTC2082 consists of the 4-kb *Bam*HI-*Xba*I segment of pTH453 and the 1.5-kb *Xba*I-*Bam*HI segment (containing the origin of ColE2) of a pEC22 derivative carrying an insertion of the *Xba*I linker (insertion 8; ref.4). Plasmid pBA32 consists of the 2.7-kb *Pvu*II-*Bam*HI segment of pBR322 and the 1.8-kb *Xmn*I-*Bam*HI segment (containing the origin of ColE3) of pEC32. Plasmid pMY12-6A is a derivative of pMY12-6 (11) carrying the ampicillin resistance gene. Plasmid pTI12-6A is a derivative of pMY12-6A, in which the transcription from the promoters, $P_L$ and $P_R$, and that from the promoter of the ampicillin resistance gene proceed in the same direction as the replication from the pBR322 origin. Plasmid pTI12-6A *rep* is a derivative of pTI12-6A with the 1.5-kb *Bam*HI-*Eco*RI fragment of pAO31 (5) containing the entire *rep* gene of ColE2. Plasmid pTI12-6A *rep-lacZ* is a derivative of pTI12-6A with the 4.6-kb *Bam*HI

* To whom correspondence should be addressed

Present addresses: +Institute for Molecular Biology, Osaka University, Suita, Osaka 550 and §Central Research Laboratories, Ajinomoto Co. Inc., Kawasaki, Kanagawa 210, Japan

fragment of pFC716 (5) containing the *rep-lacZ* translational fusion. Plasmid pTI12-6Ad was constructed from pTI12-6A *rep* by removing the *rep* gene.

### Media, enzymes, antibiotics and chemicals

L-broth (14) and M9 medium (15) were as described. Enzymes, antibiotics and chemicals were from commercial sources except for *Eco*RI, *Bam*HI and T4 DNA ligase which were prepared in our laboratory.

### Other materials

[$^3$H]-labelled plasmid DNA was prepared as described (16) by using *E. coli* strain 011′ cells carrying a derivative of pBR322.

Heparin-Sepharose 2B was prepared as described (17). Two oligonucleotides containing the ideal S/D sequence (18) and its complementary sequence, 5′GATCTAAGTAAGGAGGTTTA-AG3′ and 5′GATCCTTAAACCTCCTTACTTA3′, were synthesized using an automatic DNA synthesizer (Applied Biosystems, Model 381A).

### Detection of plasmid-coded proteins

The maxicell method (19) was performed as described with some modifications. N1790 *recA99 uvrA54* cells carrying plasmids to be tested were used. Incubation with [$^{35}$S] methionine (10 $\mu$Ci/ml) was carried out at 30°C or 42°C.

### Purification of the ColE2 Rep protein

W3110 cells newly transformed with pTI12-6A *rep* were grown to $2\times10^8$/ml at 30°C in 7-liter L-broth. Then 4-liter L-broth heated at 60°C was quickly added and mixed to bring the temperature up to 42°C. Incubation was continued at 42°C for 90 min. Approximately 20 g of cells were obtained and stored at −80°C. All the following operations were carried out at 0 to 4°C and centrifugations were at 14,000 rpm with a Hitachi RPR20 rotor, unless otherwise stated. Frozen cells (7.3 g) were suspended in 20 ml of Buffer A (50 mM Tris-HCl pH 7.5/1 mM EDTA/7 mM 2-mercaptoethanol). KCl, DTT, EDTA and lysozyme were added at final concentrations of 1.0 M, 5 mM, 10 mM and 200 $\mu$g/ml, respectively. After 5 min stirring, one twentieth volume of 10% polyethyleneimine (Polymin P ; Miles) pH 7.0 was added and the suspension was mixed gently and stood without stirring on ice for 45 min. The mixture was warmed quickly up to 37°C and after 2 min it was chilled and centrifuged for 60 min. The supernatant was diluted with an equal volume of Buffer A and passed through a DEAE cellulose column (2.0×10 cm) equilibrated with Buffer A containing 0.5 M KCl. Ammonium sulfate was added to the pooled flow through fractions to 55% saturation and the precipitate was dissolved in 10 ml of Buffer A containing 1.0 M KCl. The solution was dialyzed against Buffer A containing ammonium sulfate (35% saturation) and the precipitate was dissolved in 10 ml of Buffer A containing 1.0 M KCl. The solution was diluted with Buffer A (about 200 ml) to a final salt concentration corresponding to 0.3 M KCl and applied to a Heparin-Sepharose 2B column (1.6×8.0 cm) equilibrated with Buffer A containing 0.3 M KCl. The column was washed with the same buffer and the adsorbed proteins were eluted with a linear gradient of KCl (0.3 to 1.0 M; 150 ml). The Rep protein was eluted between 0.4 and 0.6 M KCl. The pooled fractions (54 ml) were diluted 5-fold with Buffer A, applied again on a 5-ml heparin-Sepharose 2B column and then eluted with Buffer A containing 1.0 M KCl. Pooled fractions (about 10 ml) were dialyzed against Buffer A containing ammonium sulfate (70% saturation). The precipitate was dissolved in 0.4 ml of Buffer A containing 1.0 M KCl and applied on a HPLC column (TSK gel G3000SW, Toyo Soda; 0.75×60 cm) equilibrated with Buffer A containing 1.0 M KCl. The column was washed with the same buffer and the Rep protein was eluted between 23 and 28 ml. The protein fraction was dialyzed against Buffer A containing ammonium sulfate (70% saturation). The precipitate was dissolved in 0.26 ml of Buffer A containing 1.0 M KCl, dialyzed against the same buffer and stored at −80°C. No loss of activity was observed after at least 3 months.

### In vitro DNA synthesis

DNA synthesis was measured essentially as described (6). Appropriate amounts of the ColE2 Rep protein fractions to be assayed were added to the standard reaction mixtures. Polyvinyl alcohol (5%) was used in place of polyethylene glycol. Plasmids pET21 and pET31 were used as the template DNAs.

### In vitro binding of the ColE2 Rep protein to DNA

Binding of the Rep protein to DNA was measured by the filter binding assay (20). Standard reaction mixtures (35 $\mu$l) contained Tris-HCl pH 8 (10 mM), KCl (100 mM), $MgCl_2$ (10 mM), EDTA (1 mM), ethylene glycol (10%), 2-mercaptoethanol (7 mM), [$^3$H]-labelled plasmid DNA (20 ng) and the ColE2 Rep protein (7.5 to 450 ng). After incubation at room temperature for 1 hr, reaction mixtures were filtered through nitrocellulose membranes (Millipore). The trapped radioactivity was counted with a liquid scintillation counter.

### Detection of endonuclease and topoisomerase activities

Reaction mixtures (50 $\mu$l) contained Tris-HCl pH 8 (10 mM), KCl (100 mM), $MgCl_2$ (10 mM), EDTA (1 mM), ethylene glycol (10%), 2-mercaptoethanol (7 mM), ATP (1 mM), supercoiled plasmid DNA (1 $\mu$g), and the ColE2 Rep protein (7.5 to 460 ng). After incubation at 37°C for 1 hr the mixtures were electrophoresed on agarose gel (1%). Plasmid DNA molecules were visualized by staining with ethidium bromide.

### Other methods

Protein determination was by the methods described (21; 22). SDS-PAGE was carried out essentially as described (23) with a modification. Analysis of proteins was performed at pH 8.8 or at pH 8.0. DNA sequence analysis was performed as described (24) using plasmid DNA. The $NH_2$-terminal amino acid sequence of the ColE2 Rep protein was determined by Edman degradation method by using an automatic gas-phase protein sequencer (Applied Biosystems, Model 470A).

## RESULTS

### Identification of the ColE2 Rep protein

Replication of ColE2 DNA requires a plasmid-coded *trans*-acting factor(s), one of which has been suggested, on genetic evidence, to be the product of the *rep* gene (4, 5, 6). Attempts to detect the product of the intact *rep* gene by the maxicell method, however, gave only ambiguous results (5), probably because transcription and/or translation of the *rep* gene is poor. To achieve high expression of the *rep* gene, we constructed a derivative of pBR322 which carried the *rep* gene of ColE2 downstream of the promoters, $P_L$ and $P_R$, of bacteriophage λ under the control of

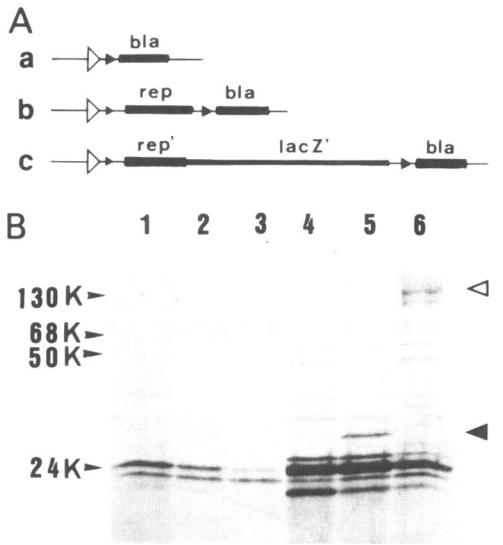


**Fig. 1.** (A) Structures of plasmids used for identification of the ColE2 Rep protein. (a) Plasmid pTI12-6Ad is a derivative of pBR322 carrying the *cI857* repressor gene and the $P_L$ and $P_R$ promoters (open triangle) of bacteriophage λ. (b) Plasmid pTI12-6A *rep* is a derivative of pTI12-6A carrying the entire *rep* gene of ColE2 with its own promoters (filled triangle) downstream of the λ promoters. (c) Plasmid pTI12-6A *rep-lacZ* carries the translational *rep-lacZ* fusion gene in place of the *rep* gene. Only relevant portions are shown. (B) Analysis of the plasmid-coded proteins synthesized in maxicells. Proteins were labelled by [$^{35}$S] methionine and analyzed by SDS-PAGE. Labelling was performed at 30°C (lanes 1−3) or at 42°C (lanes 4−6) using bacteria with pTI12-6Ad (lanes 1 and 4), pTI12-6A *rep* (lanes 2 and 5) or pTI12-6A *rep-lacZ* (lanes 3 and 6). Positions of the Rep protein (filled triangle) and the Rep-LacZ fusion protein (open triangle) are indicated. Mobilities of proteins of known molecular weights are shown on the left.

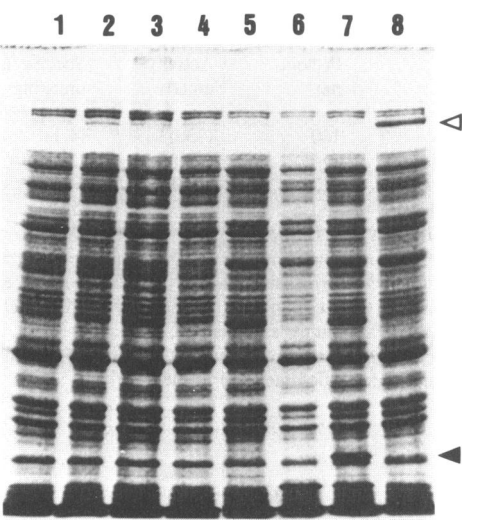


**Fig. 2.** Accumulation of the ColE2 Rep protein. W3110 cells without a plasmid (lanes 1 and 5), with pTI12-6Ad (lanes 2 and 6), pTI12-6A *rep* (lanes 3 and 7) or pTI12-6A *rep-lacZ* (lanes 4 and 8) were grown at 30°C to $2\times10^8$/ml and then each culture was divided into two portions. One was incubated at 30°C (lanes 1−4) and the other at 42°C (lanes 5−8). After incubation for 90 min total cellular proteins were analyzed by SDS-PAGE (pH 8.0) and visualized by staining with Coomassie Brilliant Blue. Positions of the Rep protein (filled triangle) and the Rep-LacZ fusion protein (open triangle) are indicated.

the heat-labile *cI857* repressor (Fig. 1A). We also constructed two other plasmids; one which did not carry the *rep* gene and the other carried the *rep* gene fused with the *lacZ* gene in frame. When the λ promoters were repressed at 30°C, all three plasmids gave similar results (Fig. 1B, lanes 1 to 3). The plasmid-coded proteins identified were the product of the β-lactamase gene and the *cI857* repressor. The product of the *rep* gene was hardly visible. On the other hand, when the promoters were activated at 42°C, some additional proteins were made, indicated by the appearance of several new bands with different mobilities characteristic for each plasmid (Fig. 1B, lanes 4 to 6). A protein with a molecular weight of about 35 K was specifically seen for the plasmid carrying the intact *rep* gene. The protein was not produced by the plasmid with the *rep-lacZ* fusion gene, but instead proteins with molecular weights of about 150 K and 130 K were specifically produced. These results indicate that the 35 Kd protein is the Rep protein. The estimated molecular weight of the Rep protein is consistent with that predicted from the nucleotide sequence of the *rep* gene (5). The 130 Kd protein is probably a degradation product of the 150 Kd protein which is the product of the *rep-lacZ* fusion gene. Two additional proteins in samples prepared at 42°C were commonly seen for all three plasmids and, therefore, are not related to the *rep* gene. The products of the β-lactamase gene appear as three proteins of different molecular weights under certain conditions (25).

### Accumulation of the ColE2 Rep protein under various conditions

We then estimated accumulation of the protein in unirradiated W3110 cells carrying pTI12-6A *rep* (Fig. 1A, plasmid b) after heat-induction at 42°C. Staining the proteins by Coomassie Brilliant Blue after SDS-PAGE revealed that a significant amount of the Rep protein was in fact accumulated in the cells (Fig. 2). Electrophoresis at pH 8.0 instead of pH 8.8 gave better separation of the Rep protein from *E. coli* proteins with similar mobility. The amount of the Rep protein accumulated in the cells reached its maximum of a few per cent of the total cellular proteins in 2 to 4 hr after temperature shift to 42°C and leveled off (data not shown).

Accumulation of the Rep protein (Figs. 1B and 2) was less than that of β-lactamase, although the *rep* gene and the ampicillin resistance gene were mainly transcribed together from the λ promoters and the *rep* gene was promoter-proximal. The results suggest that the region of RNA coding for the Rep protein is translated less efficiently than that coding for the β-lactamase and/or that the Rep protein is very unstable. We therefore constructed several derivatives of pTI12-6A *rep* carrying the synthetic ideal S/D sequence upstream of the first ATG codon around position 560 or the GTG codon around position 535 of the *rep* gene. The size of the accumulated protein with the former plasmids was identical to that with the original plasmid and clearly smaller than that with the latter plasmids (data not shown), suggesting that the first ATG codon is the real initiation codon. The amount of the Rep protein accumulated by using these plasmids, however, was no more than that by using the original plasmid. We also tested plasmid pBZ201M9 which produces a dominant mutant Lon protease of *E. coli* (12) or pACYC177-4A1 which produces the Pin protease inhibitor of bacteriophage T4 (13). Coexistence of these plasmids in the host cells did not improve the extent of accumulation of the Rep protein (data not shown), although these plasmids have been known to improve accumulation of some proteins by protecting them from cellular protease activities.

**Table 1.** Purification of the ColE2 Rep protein.

| Step | Protein mg | Activity units[a] | Specific activity units/mg |
|---|---|---|---|
| 1. Extraction | 180 | (730)[b] | (4.1) |
| 2. DEAE-cellulose | 176 | (910)[b] | (5.2) |
| 3. Ammonium sulfate | 22 | 1110 | 50 |
| 4. Heparin-Sepharose | 1.6 | 384 | 240 |
| 5. HPLC-gel filtration | 0.5 | 124 | 248 |

[a]One unit of ColE2 DNA synthesizing activity is defined as the amount of the protein which leads to the incorporation of 1 nmol of dCMP in 60 min at 32°C in the standard reaction mixture as described.
[a]Activity was not estimated accurately due to the existence of inhibitory effects by Polymin P and unidentified factor(s) in the fractions.

**Table 2.** *In vitro* ColE2 DNA synthesis with the purified ColE2 Rep protein.

| Template DNA | Addition or omission | dCMP incorporation pmol |
|---|---|---|
| ColE2 | none | 13 |
| ColE2 | −Rep protein | 0.9 |
| ColE2 | −cell extract | 0.4 |
| ColE2 | +novobiocin | 1.1 |
| ColE3 | none | 1.1 |

The standard reaction mixtures (25 μl), with addition or omission as indicated, were incubated for 60 min at 32°C. The Rep protein, cell extract and novobiocin added were 50 ng, 5 μl and 0.1 μg, respectively.

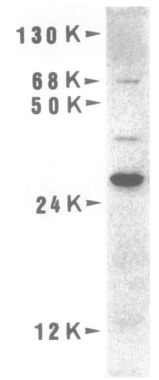


**Fig. 3.** Partially purified Rep protein. Positions of proteins of known molecular weights are indicated.

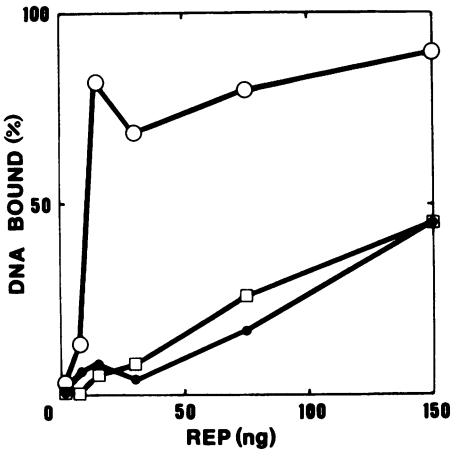


**Fig. 4.** Specific binding of the ColE2 Rep protein to ColE2 origin. The partially purified ColE2 Rep protein was incubated with [$^3$H] DNA of a derivative of pBR322 carrying the ColE2 (open circles), ColE3 (open squares) or neither (dots) origin. All the plasmid DNA used contained the pBR322 origin of replication. The binding was measured by counting the radioactivity trapped on nitrocellulose membranes as described in Materials and Methods.

## Purification of the ColE2 Rep protein

The Rep protein of ColE2 was partially purified and a summary of the purification steps is shown in Table 1. HPLC-gel filtration was performed for removal of a 100 Kd protein (data not shown). The final purity of the Rep protein was about 80% as judged by SDS-PAGE (Fig. 3).

## Properties of the ColE2 Rep protein

The amino acid sequence at the amino terminal portion of the partially purified ColE2 Rep protein was Ser-Ala-Val-Leu-Gln-Arg-Phe (data not shown). This corresponds exactly to the predicted first 7 amino acid residues of the protein translated from the first ATG codon of the open reading frame for the ColE2 *rep* gene, the first methionine residue of which is removed. The ColE2 Rep protein consists of 296 amino acid residues and the calculated molecular weight is 33,630, which roughly corresponds to that determined by SDS-PAGE.

During the course of purification we noticed that the ColE2 Rep protein formed insoluble aggregates at a salt concentration below 0.3 M KCl and became irreversibly inactivated. It was soluble as a monomer at 1.0 M KCl as revealed by HPLC-gel filtration (data not shown). Most of the purification procedures were, therefore, carried out at a salt concentration between 0.3 to 1.0 M KCl.

Endonuclease activity and topoisomerase activity in the partially purified ColE2 Rep protein were examined with negatively supercoiled ColE2 DNA. No changes in the structure of the plasmid DNA were detected (data not shown), indicating that the ColE2 Rep protein possesses neither endonuclease activity nor topoisomerase activity under the conditions used.

The effect of the partially purified ColE2 Rep protein on *in vitro* DNA synthesis was examined (Table 2). ColE2 DNA synthesis required the Rep protein and an *E. coli* cell extract. The synthesis was roughly proportional to the amount of the Rep protein added and leveled off above a protein/template ratio of about 30 (data not shown). Inclusion of rifampicin and chloramphenicol in the system indicated that the reaction is independent of DNA-dependent RNA polymerase and *de novo* protein synthesis. DNA synthesis was sensitive to novobiocin, indicating involvement of DNA gyrase. The synthesis was specific for ColE2 DNA, while ColE3 DNA was completely inactive as template. These results are consistent with those obtained by using crude cell extracts of *E. coli* carrying ColE2 plasmid in place of a crude cell extract of plasmid-free bacteria and the partially purified Rep protein (6).

We then tested whether the ColE2 Rep protein specifically recognizes and binds to the origin of ColE2 DNA replication by the filter-binding assay (Fig. 4). As the mechanism of initiation of replication in ColE2 and ColE3 is quite different from that in pBR322 (4, 5), we used pBR322 DNA for examination of possible non-specific binding of the ColE2 Rep protein to DNA. DNA molecules with the origin of ColE2 were efficiently trapped

**Table 3.** Codon usage of the Rep proteins of ColE2 and ColE3

| Amino acid | Codon | Total codons ColE2 | ColE3 | Amino acid | Codon | Total codons ColE2 | ColE3 | Amino acid | Codon | Total codons ColE2 | ColE3 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Phe | **UUU** | 2 | 1 | Pro | CCU* | 4 | 5 | Asn | AAU* | 5 | 5 |
| | UUC | 6 | 7 | | CCC* | 4 | 4 | | AAC | 7 | 8 |
| | | | | | CCA | 3 | 2 | | | | |
| Leu | UUA* | 3 | 2 | | CCG | 10 | 10 | Lys | AAA | 14 | 13 |
| | UUG* | 2 | 2 | | | | | | AAG* | 5 | 5 |
| | CUU* | 7 | 6 | Thr | ACU | 3 | 3 | | | | |
| | CUC* | 6 | 7 | | **ACC** | 1 | 1 | Asp | GAU | 9 | 8 |
| | CUA* | 0 | 0 | | ACA* | 2 | 3 | | GAC | 5 | 5 |
| | CUG | 9 | 10 | | ACG | 3 | 7 | | | | |
| | | | | | | | | Glu | GAA | 10 | 12 |
| Ile | AUU | 7 | 9 | Ala | GCU | 8 | 10 | | GAG* | 4 | 4 |
| | AUC | 6 | 5 | | GCC | 10 | 8 | | | | |
| | AUA* | 0 | 0 | | GCA | 9 | 7 | Cys | UGU | 5 | 4 |
| | | | | | GCG | 11 | 10 | | UGC | 2 | 3 |
| Met | AUG | 2 | 2 | | | | | | | | |
| | | | | Tyr | **UAU** | 2 | 4 | Trp | UGG | 8 | 9 |
| Val | GUU | 3 | 2 | | UAC | 13 | 12 | | | | |
| | GUC | 3 | 2 | | | | | Arg | CGU | 8 | 7 |
| | GUA | 4 | 4 | Ter | UAA | 0 | 0 | | CGC | 6 | 9 |
| | GUG | 5 | 6 | | UAG | 0 | 1 | | CGA* | 4 | 2 |
| | | | | | UGA | 1 | 0 | | CGG* | 3 | 4 |
| Ser | UCU | 2 | 2 | | | | | | AGA* | 3 | 2 |
| | UCC | 1 | 0 | His | CAU | 3 | 2 | | AGG* | 2 | 2 |
| | UCA* | 5 | 5 | | CAC | 4 | 5 | | | | |
| | UCG* | 4 | 3 | | | | | Gly | **GGU** | 1 | 3 |
| | AGU* | 1 | 2 | Gln | CAA* | 2 | 3 | | GGC | 6 | 8 |
| | AGC | 6 | 5 | | CAG | 7 | 8 | | GGA* | 2 | 0 |
| | | | | | | | | | GGG* | 5 | 5 |

The ColE3 Rep protein was assumed to start at the first AUG codon. Rare codons are marked with asterisks. Codons which are used frequently for most *E. coli* proteins but infrequently for the Rep proteins of ColE2 and ColE3 are shown in bold face letters.

on the nitrocellulose membranes by the presence of the partially purified Rep protein. Efficient filter binding required the ColE2 Rep protein with a protein/DNA ratio of about 50. With greater excess of the Rep protein DNA molecules with the origin of ColE3 or even those without the cloned origin were trapped. So far we do not know whether the non-specific binding was due to the ColE2 Rep protein or to other proteins in the sample.

### Use of rare codons in the *rep* genes of ColE2 and ColE3

Codon usage in the ColE2 and ColE3 *rep* genes (Table 3) is rather unusual, when compared with that for other proteins expressed in *E. coli* (26). Some of the synonymic codons that are used infrequently for synthesis of most *E. coli* proteins are used frequently. For example, there are 12 (ColE2) and 10 (ColE3) rare codons (CGA, CGG, AGA and AGG) among 26 arginine codons and there are 7 (ColE2) and 5 (ColE3) rare codons (GGA and GGG) among 14 (ColE2) and 16 (ColE3) glycine codons. They are scattered throughout the *rep* genes. On the other hand, some of the synonymic codons that are used frequently for efficiently expressed genes of *E. coli* are only rarely used. The most frequently used threonine codon (ACC) is used only once among 9 (ColE2) and 14 (ColE3) threonine codons. For three frequently used serine codons, UCC is used only once in 19 serine codons (ColE2) and not at all in 17 codons (ColE3). For the two frequently used glycine codons, GGU is used only once in 14 glycine codons (ColE2). Similar unusual use of codons has been reported for genes carried by some colicin plasmids (27, 28, 29). Unusually frequent use of rare codons has been known for regulatory genes of *E. coli* which are expressed inefficiently (26).

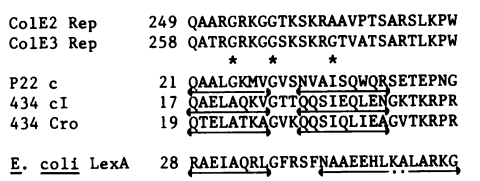


**Fig. 5.** Portions of amino acid sequences of the ColE2 and ColE3 Rep proteins showing some homology to the DNA binding domains of the P22 *c*, 434 *cI*, 434 *cro* (30), and *E. coli lexA* (31) repressors. Asterisks indicate positions 5, 9 and 15 which are most highly conserved (30). α-helix portions of the repressor proteins are indicated by arrows. Numbers on the left side are the position numbers of the amino acids at the beginning of the sequences shown.

## DISCUSSION

We have identified and partially purified the ColE2 Rep protein, a plasmid-coded protein required for initiation of ColE2 DNA replication. Activity of the Rep protein was detected using ColE2 DNA replication in crude cell extracts of *E. coli* (5). The protein was detected by its size in gel electrophoresis, when the *rep* gene was transcribed from the promoters of bacteriophage λ under the control of heat-labile *cI* repressor, but not when expressed from its natural promoter.

The existence of many rare codons might be one of the reasons for inefficient expression of the *rep* gene. Furthermore the protein seems to be very unstable. When host bacteria entered the late logarithmic growth phase, the activity of the Rep protein quickly disappeared, although the expression of the *rep* gene still

continued at a reduced rate for several hours as revealed by measuring the β-galactosidase activity expressed from the *rep-lacZ* translational fusion gene under the same conditions (data not shown).

The ColE2 Rep protein was shown to bind specifically to DNA containing the origin of replication of ColE2 and to stimulate its replication in crude cell extracts of *E. coli*. The ColE2 Rep protein showed no endonuclease activity or topoisomerase activity on ColE2 DNA. The protein might be a plasmid-coded primase by itself or it might help the host primase (and other host proteins) to interact with the origin.

Genetic analyses have shown that the plasmid-specificity of interaction of the Rep proteins of ColE2 and ColE3 with the origins is determined by the absence (ColE2) or the presence (ColE3) of a single base-pair at a fixed position in otherwise identical origin sequences (8). The amino acid sequences of the Rep proteins of ColE2 and ColE3 differ by about 10% in the positions which are located in the carboxy terminal halves of the proteins (5). We are currently trying to identify the amino acid residue(s) responsible for discrimination between the ColE2 and ColE3 origins.

Many sequence-specific DNA binding proteins share sequence homologies in the DNA binding domains and use a helix-turn-helix structure for recognition and binding to specific DNA sequences (30). In the nonhomologous carboxy terminal regions of the Rep proteins of ColE2 and ColE3, there are sequences which seem to share some homology with the DNA binding domains of some sequence-specific DNA binding proteins (Fig. 5). Homology is better in the putative first α-helix portion and the Rep proteins share especially good homology with the repressor of bacteriophage P22. For the putative second α-helix portion, homology is very poor, but they might still share some homology with the LexA repressor. NMR study of the amino terminal DNA binding domain of the LexA repressor revealed that there is a helix-loop-helix structure with the two helices located further apart (Fig. 5), although the secondary structure prediction gave only a low probability for the existence of a helix-turn-helix domain (31). Furthermore there are a number of LexA mutants with a decreased affinity for operator binding, which are located in the region of the LexA protein shown in Fig. 5 (cited in ref. 31). The method for secondary structure prediction (32) showed that the regions of the Rep proteins of ColE2 and ColE3 shown in Fig. 5 have only a low probability of α-helix structure for both of the putative α-helix portions (data not shown). The sequence homology, however, might suggest that the region of each Rep protein conforms the sequence-specific binding domain involved in recognition of the origin sequence. Detailed study of direct physical interaction of the Rep protein with the origin must await establishment of more efficient procedures for purification of the Rep protein.

## ACKNOWLEDGEMENT

We are grateful to Drs. H.Ogawa, T.Ogawa and T.Horii for their helpful discussions and continuing encouragement and Dr. J.Tomizawa for critical reading of the manuscript. We thank Drs. H.Kagamiyama, K.Inoue and S.Kuramitsu for their help in determination of the $NH_2$-terminal amino acid sequence of the ColE2 Rep protein. We also thank Drs. Y.Eguchi, A.Markovitz, K.Shigesada, L.Simon, and T.Tsurimoto for materials used in this study. This work was supported in part by a grant-in-aid for Scientific Research from the Ministry of Education, Science and Culture of Japan to T.I.

## REFERENCES

1. Herschman,H.R. and Helinski,D.R. (1967) *J. Bacteriol.*, **94**, 691–699.
2. Kingsbury,D.T. and Helinski,D.R. (1970) *Biochem. Biophys. Res. Commun.*, **41**, 1538–1544.
3. Tacon,W. and Sherratt,D.J. (1976) *Mol. Gen. Genet.*, **147**, 331–335.
4. Horii,T. and Itoh,T. (1988) *Mol. Gen. Genet.*, **212**, 225–231.
5. Yasueda,H., Horii,T. and Itoh,T. (1989) *Mol. Gen. Genet.*, **215**, 209–216.
6. Itoh,T. and Horii,T. (1989) *Mol. Gen. Genet.*, **219**, 249–255.
7. Studier,F.W. (1969) *Virology*, **39**, 562–574.
8. Sancar,A., Hack,A.M. and Rupp,W.D. (1979) *J. Bacteriol.*, **137**, 692–693.
9. Gellert,M., O'Dea,M., Itoh,T. and Tomizawa,J. (1976) *Proc. Natl. Acad. Sci. USA*, **73**, 4474–4478.
10. Tajima,Y., Horii,T. and Itoh,T. (1988) *Mol. Gen. Genet.*, **214**, 209–216.
11. Tsurimoto,T., Hase,T., Matsubara,H. and Matsubara,K. (1982) *Mol. Gen. Genet.*, **187**, 79–86.
12. Zehnbauer,B.A. and Markovitz,A. (1980) *J. Bacteriol.*, **143**, 852–863.
13. Simon,L.D., Randolph,B., Irwin,N. and Binkowski,G. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 2059–2062.
14. Ogawa,T. (1975) *J. Mol. Biol.*, **94**, 327–340.
15. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
16. Sakakibara,Y. and Tomizawa,J. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 802–806.
17. Kameji,T., Murakami,Y., Fujita,K. and Hayashi,S. (1982) *Biochim. Biophys. Acta.*, **717**, 111–117.
18. Shine,J. and Dalgarno,L. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 1342–1346.
19. Brent,R. and Ptashne,M. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 1932–1936.
20. Riggs,A.D., Suzuki,H. and Bourgeois,S. (1970) *J. Mol. Biol.*, **48**, 67–83.
21. Lowry,O.H., Rosebrough,N.J., Farr,A.L. and Randall,R.J. (1951) *J. Biol. Chem.*, **193**, 265–275.
22. Bradford,M.M. (1976) *Anal. Biochem.*, **72**, 248–254.
23. Laemmli,U.K. (1970) *Nature*, **227**, 680–685.
24. Smith,A.J.H. (1980) *Methods in Enzymol.*, **65**, 560–580.
25. Sutcliffe,J.G. (1978) *Cold Spring Harbor Symp. on Quant. Biol.*, **43**, 77–90.
26. Maruyama,T., Gojobori,T., Aota,S. and Ikemura,Y. (1986) *Nucleic Acids Res.*, r151-r197.
27. van den Elzen,P.J.M., Walters,H.H.B., Veltkamp,E. and Nijkamp,H.J.J. (1983) *Nucleic Acids Res.*, **11**, 2465–2477.
28. Morlon,J., Lloube 's,R., Varenne,S. Chartier,M. and Lazdunski,C. (1983) *J. Mol. Biol.*, **170**, 271–285.
29. Cole,S.T., Saint-Joanis,B. and Pugsley,A.P. (1985) *Mol. Gen. Genet.*, **198**, 465–472.
30. Pabo,C.O. and Sauer,R.T. (1984) *Ann. Rev. Biochem.*, **53**, 293–321.
31. Lamerichs,R.M.J.N, Padilla,A., Boelens,R., Kaptein,R., Ottleben,G., Ruterjans,H., Granger-Schnarr,M., Oertel,P. and Schnarr,M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6863–6867.
32. Chou,P.Y. and Fassmann,G.D. (1979) *Adv. Enzymol.*, **47**, 45–148.