Structural and Functional Analyses of the Genes on the Invertible DNA Segments in Intact and Defective Bacteriophages

1991 9

AKIRA TOMINAGA

The Graduate School of Natural
Science and Technology
(Doctor Course)

OKAYAMA UNIVERSITY

(1)

Structural and Functional Analyses of the Genes on the Invertible DNA Segments in Intact and Defective Bacteriophages

1991 September

AKIRA TOMINAGA

The Graduate School of Natural
Science and Technology
(Doctor Course)

OKAYAMA UNIVERSITY

PREFACE

Phase variation may be defined as alternative expression between the genes that specify a functionally identical phenotype. For a cell to express a new phenotype, a gene or a set of genes is inactivated and instead another gene or another set is to be activated; sometimes, a DNA segment is chosen among a set of fragments to form another complete gene constituent (Swanson et al. 1986). In procaryotes the phase variation defined as above has been shown to result from DNA rearrangements that require recombination at specific sites (Simon and Silverman 1983). One form of DNA rearrangements that is reversible and controls gene expression is inversion of some DNA segment.

Inversional DNA systems are composed of an invertible DNA segment and a site-specific recombinase gene that is adjacent to, or contained in, the segment and mediates recombination of the segment to evoke a sleeping gene(s). To date, four systems have been characterized in detail: Salmonella typhimurium H segment-hin gene (H-hin) (Silverman and Simon 1983, Johnson and Simon 1985), phage Mu G-gin (van de Putte et al. 1980, Kamp 1981, Giphart-Gassler et al. 1982), phage P1 C-cin (Hiestand-Nauer and Iida 1983, Iida 1984), and Escherichia coli P-pin systems (Enomoto et al. 1983, Plasterk et al. 1983, Plasterk and van de Putte 1985). In S. typhimurium, H inversion switches on-off expression of both phase 2 flagellin and repressor genes that compose an operon, causing repression and derepression of a phase 1 flagellin gene. Inversion of the G segment controls alternation of a host range of Mu by expressing one of the two sets of genes on the G segment. Inversion of the C segment also controls host specificity of P1, but it has not been clear if the C segment contains two sets of genes. For inversion of the P segment, any change in a phenotype has not been detected in the cells that carry the P-pin system (Tominaga, unpublished data). The Hin, Gin, Cin and Pin recombinases are almost alike in amino acid sequences and functionally substitutable for one another in mediating inversion of the H, G, C and P segments (Kutsukake

and Iino 1980, Kamp and Kahmann 1981, Iida et al. 1982, Enomoto et al. 1983, van de Putte et al. 1984, Momota and Enomoto 1986). However, the invertible segments are different not only in function of the gene(s) on the segments but in their structure.

There is an idea to explain the functional and structural similarity among the recombinase genes and the differences among the invertible segments (Simon et al. 1980): an ancestor type of the invertible segment-recombinase gene system would be a kind of transposon, which carries a site-specific recombinase gene and inverted repeats at its both ends that are recognizable by the recombinase. Such the transposon would insert at a site adjacent to some gene and undergo some kinds of rearrangement including nearby DNA, such as replicative transposition, deletion and inversion, to evolve to the system we can see at the present time. If this is the case, one of the experimental methods to approach this idea will be a survey of different kinds of bacteria for the presence of recombinase genes whose products can mediate inversion of the known invertible segments and examination of the putative invertible segments that might be located closely to the recombinase genes isolated. Such the approach in our laboratory has led us to the finding of *pin* in *E. coli* K-12 and recently to the detection of recombinase genes in *Shigella boydii* and *Sh. dysenteriae*. I have studied mainly about functions of the genes on the C segment of phage P1 and on the P segment of *E. coli* e14, since the molecular structures of these segments have previously been reported by other research groups (Iida et al. 1982, Plasterk and van de Putte 1985), but biological functions of the genes on the segments have remained unknown. I have also been studying the structure and function of the genes on the invertible segments newly detected in *Sh. boydii* and *Sh. dysenteriae* with the expectation that they might carry genes encoding completely different functions from those reported previously for the other segments. However, it was found, so far as examined, that the function of the genes on the invertible segments including

the C, P and B segments is concerned with tail fiber proteins of phage or its defective type.

This thesis describes three kinds of experiments that have revealed or suggested the function of the genes on the three invertible segments from phage P1, defective phage e14 of E. coli K-12 and an unknown defective phage of Sh. boydii. For the B-pinB system of Sh. boydii, the structure at a nucleotide level is described in detail. Therefore, the paper is divided into three parts: section I deals with identification of the gene products and determination of their function for the C segment of P1, section II describes identification and a putative role of the gene products encoded on the P segment of e14, and section III deals with characterization of the B-pinB system cloned from a defective phage of Sh. boydii. Each section consists of Abstract, Materials and Methods, Results and Discussion, in which the description about Materials and Methods is limited to only the subjects concerned with special experiments in each section to avoid overlapping. References in each section are gathered and shown in the last section.

CONTENTS

I. Identification and characterization of the gene products encoded on the C segment of bacteriophage P1 ---------- 7

ABSTRACT ---------- 7

INTRODUCTION ---------- 8

MATERIALS AND METHODS ---------- 10

RESULTS ---------- 15

1) Isolation and characterization of P1 mutants carrying a fixed C segment ---------- 15

2) Immunological analysis of P1 mutants with anti-P1 sera ----- 18

3) Identification of polypeptides characteristic of the orientation of the C segment ---------- 19

4) Cloning of the C segment ---------- 20

5) Detection of gene products encoded on the C segment -------- 20

DISCUSSION ---------- 22

TABLES AND FIGURES ---------- 25

II. Identification of the gene products encoded on the invertible P segment of the e14 element in Escherichia coli K-12 -- 36

ABSTRACT ---------- 36

INTRODUCTION ---------- 37

MATERIALS AND METHODS ---------- 38

RESULTS ---------- 40

1) Cloning of the P segment from E. coli K-12 ---------- 40

2) Construction of the plasmids that direct overproduction of polypeptides encoded on the P segment ---------- 40

3) Identification of the plasmid-coded polypeptides ---------- 43

4) Amino acid sequence analysis of four open reading frames --- 43

5) Western immunoblot analysis ---------- 44

DISCUSSION ---------- 46

TABLES AND FIGURES ---------- 48

III. Cloning and characterization of a cryptic invertible B segment from *Shigella boydii* ---------- 62

ABSTRACT ---------- 62

INTRODUCTION ---------- 63

MATERIALS AND METHODS ---------- 65

RESULTS ---------- 68

1) Construction of the assay plasmid for recombinase gene activity ---------- 68

2) Detection of recombinase gene activity from *Shigella* spp. -- 69

3) Inversion activity for H, P, and C segments in *Sh. boydii* and *Sh. dysenteriae* ---------- 70

4) Cloning of the *pinB* gene and its adjacent invertible B segment from *Sh. boydii* ---------- 72

5) Nucleotide sequences of the *pinB* gene and the B segment----- 73

6) Southern blot analysis ---------- 75

7) Western immunoblot analysis ---------- 76

DISCUSSION ---------- 78

TABLES AND FIGURES ---------- 82

REFERENCES ---------- 100

I. Identification and characterization of the gene products encoded on the C segment of bacteriophage P1

ABSTRACT

To examine the gene products of the C segment, P1cin mutants carrying fixed C segments were isolated. P1*cin*C(+) mutants were isolated from a P1C*mclr100* lysogen based on the criterion that the *cin* mutation cannot restore flagellar phase variation of the *hin pin* strain, and a P1*cin*C(-) mutant was isolated from a P1*cin*C(+) lysogen of the *hin*$^+$ *pin*$^+$ strain by utilizing the difference in colonial types at 41°C. The *hin*$^+$ *pin*$^+$/ P1*cin*C(-) strain yielded P1*cin*C(+) lysogen at a rate of 10^{-2} per bacterium while the *hin pin* / P1*cin*C(-) strain was considerably stable. Anti-P1*cin*C(+) serum neutralized both the P1*cin*C(-) and MuG(+) particles to 16% the rate of the homologous combination whereas anti-P1*cin*C(-) serum neutralized the P1*cin*C(+) particles to 45% the rate of the homologous combination. This indicates that P1*cin*C(+) and P1*cin*C(-) bear antigenically distinct polypeptides concerned with host specificity and that P1*cin*C(+) shares some antigen(s) involved in infectivity with MuG(+). The analysis of whole phage polypeptides revealed that P1*cin*C(+) had a 105 kDa polypeptide, and P1*cin*C(-) had 101 kDa instead of 105kDa. Two polypeptide, of 16.5 and 24 kDa, were inferred to be the gene products encoded in the C segment by the maxicell analysis. Phage P1*cin*C(-) was able to form plaques on a wild-type strain of *Escherichia coli* C and on *Shigella sonnei* in the presence of Mg^{2+}.

INTRODUCTION

Bacteriophage P1 has an invertible C segment, and *cin* gene adjacently linked to the C segment mediates its inversion (Iida et al. 1982, Hiestand-Nauer and Iida 1983). C inversion is concerned with host specificity; P1C(+) which has one of two possible orientations of the C segment is infectious to *Escherichia coli* K-12, whereas P1C(-) having opposite orientation cannot grow on it, but becomes infectious to some mutant of *E. coli* C (Iida 1984). It has been inferred by the analysis of structural proteins from infectious P1 particles (Walker and Walker 1981) and from the electron-microscopic observation of some P1 mutants (Walker and Walker 1983) that a 105 kilodalton (kDa) polypeptide, a major component of a tail fiber, is encoded by gene *19* in the C segment region and concerned with infectivity. It has been well studied with phage Mu that the G segment whose inversion is mediated by *gin* contains two sets of genes (Giphart-Gassler et al. 1982); *U* and *U'*, and *S* and *S'* whose common coding region is outside the G segment. MuG(+) expressing *U* and *S* genes shows host specificity to *E. coli* K-12, and MuG(-) expressing *U'* and *S'* genes is infectious to the bacteria other than *E. coli* K-12 (van de Putte et al. 1980). These genes were found to be involved in the biosynthesis of tail fibers and/or their attachment to the Mu particles (Grundy and Howe 1984). So far as the invertible segment is concerned, P1 and Mu have a great similarity; physical homology is shown by heteroduplex analysis (Chow and Bukhari 1976), *S* and *U* mutations of Mu can be rescued by the corresponding wild type alleles of P1 (Toussaint et al. 1978), and P1 can acquire the Mu host range by recombination between the two phages (Toussaint et al. 1978). These results postulate that the C segment of P1 would carry the set of genes which determine the host range like the G segment of Mu.

The present study deals with isolation and characterization of P1*cin*C(+) and P1*cin*C(-) mutants and detection of the polypeptides encoded on the C segment, which involved in host specificity. Salmonella *hin* and *E. coli pin*

genes are known to mediate inversion of the C segment as well as flagellar phase variation (Silverman and Simon 1980, Enomoto et al. 1983) as dose the cin gene of P1 (Kutsukake and Iino 1980a). To isolate cin mutants and to harvest pure P1 particles with fixed C orientation in large quantity, an E. coli K-12 derivative was used that has two Salmonella flagellin genes, but does not manifest phase variation due to hin and pin mutations.

MATERIALS AND METHODS

Bacterial strains and phages.

Strains used were *E. coli* K-12 derivatives (Table I-1), most of which had *Salmonella fliC-i* and *fljB-e,n,x fljA* regions (Enomoto and Stocker 1975, Emonoto 1983). Strain EJ1076 is *pin*$^+$ capable of suppressing a *hin* mutation and manifests phase variation between the two flagellar antigens, *i* and *e,n,x*. From a *zcg3*::Tn*10* derivative of EJ1076, in which *pin*$^+$ and Tn*10* are closely linked (Enomoto et al. 1983), tetracycline-sensitive (Tet^S) colonies were isolated by the positive selection method (Maloy and Nunn 1981), and they were tested for their flagellar expression, and one of them (EJ1420) that stably expressed *i* without showing phase variation was isolated as a Δ*pin* mutant. Wild type *E. coli* C was used as an indicator for P1*cin*C(-). *Shigella sonnei* (IID969, Institute of Medical Science), *Citrobacter freundii* (NIH10018-68, National Institutes of Health), *Enterobacter aerogenes* (NCTC10006, National Collection of Type Cultures), *Klebsiella pneumoniae* (NCTC9632) and *Proteus rettgeri* (NIH96) were examined for the sensitivity to P1*cin*C(-) phage and were kindly supplied by the Medical School, Okayama University. Phage P1Cm*clr100* (Rosner 1972), a temperature sensitive *c1* mutant, was used for isolating *cin* mutants, and P1*vir* was used for phenotypic mixing. Mu*cts* (Howe 1973) was used as control for neutralization tests with anti-P1 sera. Mu*cts* and P1Cm*clr100* were supplied by Y. Komeda and K. Kutsukake, respectively, University of Tokyo.

Media and antisera.

Components of TLY broth, nutrient agar (NA) and nutrient semisolid agar (NSS) were described previously (Enomoto et al. 1983). Chloramphenicol (Cm, Sigma Chemical Co.), tetracycline (Tet, Sigma) and ampicillin (Amp, Sigma) were used at concentrations of 12.5, 25 and 50 mg per liter, respectively. Buffer solution for P1 dilution contained 86 mM NaCl and 10 mM $MgCl_2$ in 10 mM

Tris-HCl (pH 7.5), and the same solution excluded NaCl was used for Mu dilution. Anti-*i* flagellum serum was prepared against a monophasic *S. typhimurium* strain (SL4206, Enomoto and Stocker 1974). Anti-P1*cin*C(+) and anti-P1*cin*C(-) sera were prepared by serial injection of purified P1 particles (about 10^{11} particles/ml, total 7.5 ml) into rabbits and used after treatment for 30 min at 56°C.

Isolation procedures of P1*cin* mutants and phenotypic mixing.

First, P1*cin*$^+$C(+) was treated with 50 mM $NaNO_2$ in 100 mM acetate-acetic acid buffer (pH4.6) (Miller 1972). The treatment was terminated after 20 min by adding 1N NaOH (45 µl per 1 ml reaction mixture), and the fraction of phage surviving was around 10^{-2}. The mutagenized phage suspension (about 10^5 plaque-forming unit=PFU/ml) was diluted ten times with the broth culture of the monophasic *hin pin* strain (EJ1420). An aliquote of the mixture was taken into a plate that contained cooled but not solidified NSS medium containing Cm and anti-*i* serum and mixed quickly. The plates were incubated overnight at 30°C. Nonmotile Cm^r colonies were picked up as candidates for P1*cin*C(+) lysogens and tested for temperature sensitivity at 41°C and for phase variation, and phages induced from these candidates were also tested for their infectivity.

To isolate P1*cin*C(-) lysogens, the *hin*$^+$ *pin*$^+$ strain (EJ1346) was lysogenized with P1*cin*C(+) and then, colonies of the P1*cin*C(+) lysogens were duplicated by replica plating on NA plates containing Cm, and master NA plates and replica plates were incubated overnight at 41°C and 30°C, respectively. Colonies of P1*cin*C(-) lysogens, which resulted from C-segment inversion mediated by *hin*$^+$ and/or *pin*$^+$, were distinguishable from those of P1*cin*C(+) lysogens since in the case of P1*cin*C(-) lysogens, nonlysogenic survivors occurring spontaneously in a colony at 41°C are not infected with liberated P1*cin*C(-) and can form small colonies whereas nonlysogenic survivors from

P1cinC(+) lysogens are almost completely killed by induced phages. Candidates were isolated from replica plates and tested for the same characters as described for P1cinC(+) lysogens. To lysogenized new strains with noninfectious P1cinC(-), phenotypic mixing (Novick and Szilard 1951) was utilized. The broth culture of the strain, P1cinC(-)/hin$^+$ pin$^+$, was treated for 20 min at 41°C immediately after addition of P1vir at a multiplicity of infection of around three and incubated at 37°C until lysis occurred. The lysate evoked 20 times as many Cm^r colonies as did the lysate made without P1vir. Cm^r colonies were stabbed in NA plates on which indicator bacteria (EJ1420) had been previously layered with soft agar and incubated for 2 h at 41°C, then overnight at 37°C. The clone which makes no lysis area around the stab site was presumed to be a P1cinC(-) lysogen.

Purification of phages.

Lysates of P1cin$^+$, P1cinC(+) and P1cinC(-) were prepared by thermal induction (Rosner 1972) of the respective lysogens of the hin pin strain (EJ1420). MuG(+) lysate was prepared by the confluent lysis method (Rosner 1972) on NA plates. One liter of the induced lysate was precipitated by polyethylene glycol 6,000 after removal of bacterial debris, suspended in the buffer solution and purified by CsCl step gradient centrifugation (Yamamoto et al. 1970, Mural et al. 1979). The suspension of the purified particles was dialyzed and finally adjusted to 4.0 ml with the buffer solution. The suspensions of P1cin$^+$ and P1cinC(+) contained 10^{11}-10^{12} PFU/ml, corresponding to around 30 µg protein/ml, and the P1cinC(-) suspension showed the similar protein content. These suspensions were used for the isolation of DNA, immunization of rabbits and analysis of structural proteins.

Cloning of BamHI-5 fragment of P1.

Isolation of plasmid DNA and cloning procedures followed the method

described (Mural et al. 1979). P1 DNA (2μg) and pBR322 DNA (1μg) were mixed, digested with BamHI (Takara Shuzo Co.), ligated with T4 ligase (Takara) and transformed to the hin pin strain according to the method (Dagert and Ehrlich 1979). Since the BamHI-5 fragment of P1 is known to contain the C segment and cin (Mural et al. 1979, Kutsukake and Iino 1980b), Amp^r Tet^s transformants showing phase variation were isolated in transformation with P1cin$^+$ DNA. In transformation with P1cin DNA, Amp^r Tet^s transformants were first selected, and those which harbored hybrid plasmids carrying the BamHI-5 fragment were kept.

Neutralization tests.

Anti-P1cinC(+) and anti-P1cinC(-) sera were diluted with dilution buffer to the concentration at which a phage titer in the homologous combination was reduced to approximately 10^{-4} after the incubation for 60 min at 37°C. Phage lysates were diluted to 5×10^7 PFU/ml. An equal volume of antiserum and phage suspension were mixed, incubated at 37°C and aliquotes were titrated at an interval of five or ten min.

Sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE).

Four ml of the purified phage suspension was treated for two h at 0°C after addition of one ml of 25% trichloroacetic acid and precipitated by ultracentrifugation (200,000 x g, 30 min). To estimate the protein content, the precipitate was neutralized with 0.1 N NaOH, dissolved in distilled water and measured by the Lowry method (Lowry et al. 1951). For PAGE the precipitate was dissolved in approximately 100 μl of Laemmli buffer (1 mg protein/ml) and boiled for three min, and 20 μl was loaded on the gel and electrophoresed according to the method (Laemmli 1970) slightly modified by the use of a 4.75% acrylamide stacking gel and a 12% resolving gel.

Maxicell analysis of plasmid-coded polypeptides.

For detection of plasmid-encoded polypeptides, maxicells were produced by the method of Sancer et al. (1979) and labeled with [^{35}S]methionine. Samples were resolved by SDS-PAGE. Gels were then dried under vacuum and heat and exposed to Fuji X-RAY film for 12 to 24 h.

RESULTS

1. Isolation and characterization of P1 mutants carrying a fixed C segment.

Isolation of P1*cin*C(+) mutant.

P1*cin* mutants result in fixed orientations of the C segment, C(+) or C(-) (Iida et al. 1982), and lose the ability to mediate flagellar-phase variation (Kutsukake and Iino 1980b). Making use of this character, P1*cin* mutants were isolated. The monophasic *hin pin* strain (EJ1420) expressing flagellar antigen *i* was infected with mutagenized P1Cm*clr100*, and Cm^r colonies were isolated in NSS medium containing Cm and anti-*i* serum. Of more than 200 candidates tested, three were presumed to be P1*cin*C(+) lysogens since (i) they grew at 30°C, but not at 41°C on NA plates containing Cm, (ii) they did not manifest phase variation, and (iii) induced phages formed clear-lysis spots on *E. coli* K-12. Using one of the lysates, strains EJ1420 and EJ1346 (hin^+ pin^+) were lysogenized and designated EJ1452 and EJ1456, respectively.

Isolation of P1*cin*C(-) mutant.

Since C(+) orientation of prophage is expected to be inverted by hin^+ and/or pin^+ in EJ1456, P1*cin*C(-) lysogens were isolated in EJ1456 using difference in colonial types. Colonies of EJ1456 were transferred from 30°C to 41°C after they were duplicated by replica plating and incubated overnight. Two colonial types were observed. The majority were flat trasparent colonies and the few (four of 504) dense irregular ones which seemed to be formed by well-grown secondary colonies resulting from survivors. One of the colonies of the latter type was saved from the replica plate. It neither grew at 41°C nor showed phase variation, and induced phages formed few plaques on *E. coli* K-12. The clone was presumed to be a P1*cin*C(-) lysogen and designated EJ1457. Finally, the *hin pin* strain was lysogenized with the P1*cin*C(-) mutant. The lysate that contained P1*cin*C(-) particles bearing C(+) adsorption properties

was prepared from EJ1457 by the method of phenotypic mixing and applied to EJ1420. Cm^r colonies obtained were tested as to whether they yielded plaque-forming phages upon thermal induction. Of 24 colonies tested, five were judged as P1cinC(-) lysogens, and one of them was designated EJ1534.

Characterization of P1cin lysogens.

Four kinds of P1cin lysogens obtained so far were compared with respect to their characters. First, a proportion of lysogenic bacteria carrying prophage with altered infectivity was measured. Using five ml of broth cultures (1-5 x 10^8/ml) from single colonies, the number of colony-forming units (CFU) and the number of bacteria capable of being infective centers (IC) at s nonpermissive temperature were counted. For IC count the culture was spotted on the lawn of indicator bacteria (EJ1420) and incubated at 41°C. Lysates were also made from these cultures by thermal induction and titrated (Table I-2). The lysogens, hin pin/P1cinC(+) (EJ1452) and hin^+ pin^+/P1cinC(+) (EJ1456), showed an IC/CFU ratio of around one and their induced lysates contained 10^8 to 10^9 PFU/ml. It was expected that the orientation of the C segment in EJ1456 was inverted by hin^+ and/or pin^+ genes, and that the IC/CFU ratio would be lower than that of the hin pin lysogen; however, as described below, the inversion rate was so low that the effect of these genes seemed to be included in the experimental error. hin^+ pin^+/P1cinC(-) (EJ1457) showed the ratio of 1.5 x 10^{-2}, suggesting that one to two % of bacteria in the culture harbored P1cinC(+) whose C segment was inverted by the host genes. This observation postulates that the culture of EJ1456 also contains P1cinC(-) lysogens with a similar rate. With the lysogen hin pin/P1cinC(-) (EJ1534), P1cinC(-) prophage seemed to be considerably stable since the IC/CFU ratio was at the order of 10^{-6} and the induced lysate contained only 10^2 to 10^3 PFU/ml. It is presumed that the induced lysate of EJ1534 contains P1cinC(-) particles comparable to the number of P1cinC(+) ones (10^8-10^9 PFU/ml) induced from EJ1452 since both phage

suspensions had the similar protein content when purified by the same condition. It is, therefore, inferred that both the lysates from EJ1452 and EJ1534 contain phages with the opposite orientation of the C segment at the rate of approximately 10^{-6}. With the four kinds of lysogens, formation of nonlysogenic bacteria occurring spontaneously in the culture was examined Table 2). The ratio of CFU on NA plates at 41°C to that at 30°C was 10^{-3} to 10^{-4} with all the four lysogens, and almost all the colonies grown at 41°C were sensitive to Cm. It was, therefore, demonstrated that the appearance of two colonial types at 41°C was not attributed to the difference in the number of nonlysogens yielded in the colony, but due to the character of induced P1cin of whether it could kill the nonlysogens.

Infectivity of P1cinC(-) mutant.

It was found that the P1cinC(-) mutant can form plaques on the wild-type strain of E. coli C in the presence of Mg^{2+} although it cannot plate on E. coli C in the absence of Mg^{2+} or on E. coli K-12 in the presence or absence of Mg^{2+} (Table I-3). At Mg^{2+} concentrations ranging from 10 to 50 mM, the P1cinC(-) lysate prepared by thermal induction (Rosner 1972) of EJ1420 lysogens gave clear plaques at 41°C on a lawn of E. coli C with high efficiency, showing a titer of the order of 10^8 PFU/ml, but at concentrations below 7.5 mM Mg^{2+} it gave small obscure plaques with low efficiency. Both $MgSO_4$ and $MgCl_2$ had the same effect. Other divalent metal ions, Mn^{2+}, Ba^{2+}, Cu^{2+}, Zn^{2+}, Ni^{2+}, and Fe^{2+}, did not have any effect on plaque formation. On the other hand, it was found that P1cinC(+) required Ca^{2+} for plaque formation as has been reported (Bertani and Nice 1954) and moreover, Ca^{2+} could be substituted for Mg^{2+}. Although P1cinC(+) did not form plaques on a lawn of E. coli C or EJ1420 on an ordinary nutrient agar plate containing 5 mM sodium citrate, it did form plaques when 5 mM $CaCl_2$ or 10 mM $MgSO_4$ was added to this plate. To find other bacterial species sensitive to P1cinC(-) in the presence

of Mg^{2+}, seven strains from six genera of Enterobacteriaceae were examined. To avoid the effect of host-specified restriction systems, first P1Cm lysogens were isolated from each strain, and their induced lysates were examined for plaque formation on homologous strains. Citrobacter freundii, Enterobacter aerogenes and a S. typhimurium galE mutant were not lysed by but were lysogenized with P1cinC(-), whereas Klebsiella pneumoniae, Proteus rettgeri and S. typhimurium LT2 were not susceptible to either P1cinC(-) or P1cinC(+). P1cinC(+) and P1cin^{+} were used as controls (Table I-3). Among the four P1cinC(-) lysates, only the lysate from Sh. sonnei produced clear plaques on the homologous bacteria in the presence of Mg^{2+} with the high efficiency similar to that shown by P1cinC(+). Thus Sh. sonnei was found to be a good indicator for P1cinC(-) in the presence of Mg^{2+}.

2. Immunological analysis of P1 mutants with anti-P1 sera.

To examine the difference between the P1cinC(+) and P1cinC(-) particles, neutralization tests were performed by use of antiserum prepared against purified particles of P1cinC(+) or P1cinC(-). MuG(+) and MuG(-) were also tested since it has been reported (Toussaint et al. 1978) that anti-P1 and anti-Mu sera partially neutralize Mu and P1, respectively. Anti-P1cinC(+) serum neutralized homologous P1cinC(+) to the extent of 3.9×10^{-3} and heterologous P1cinC(-) and MuG(+) to approximately 0.4 by the treatment for 30 min at 37°C when the value at 0 min was taken as 1.0, but not MuG(-) (Fig. I-1A). Anti-P1cinC(-) serum neutralized P1cinC(-) to the extent of 1.8×10^{-3} for 30 min, P1cinC(+) to 0.15 for 20 min, and MuG(-) to 2.5×10^{-2} for 20 min, but not MuG(+) (Fig. I-1B). Titers of both antisera were calculated using the linear slope in Fig. I-1 from the equation, $K=2.3D/t \log P_0/P_t$ (Dulbecco et al. 1956). The titer of anti-P1cinC(+) serum to P1cinC(-) and MuG(+) was approximately 30 per min, corresponding to 16% the titer of the homologous combination (184.5/min), but its titer to MuG(-) was less than 0.1 per min.

This indicates that both P1cinC(-) and MuG(+) particles share common antigenicity in part with the P1cinC(+) particle. The anti-P1cinC(+) and anti-MuG(-) titers of anti-P1cinC(-) serum were 75.2 and 147.4 per min, corresponding to approximately 45% and 88% of the titer in the homologous combination (168.3/min), respectively, and its anti-MuG(+) titer was less than 0.1 per min. This indicates that P1cinC(-), P1cinC(+) and MuG(-) particles contain in common the considerable degree of antigenicity, which is not contained in MuG(+). The neutralization tests suggest that P1cinC(+) and P1cinC(-) bear the antigenically distinct polypeptide(s) as component(s) of the tail structure involved in host specificity though they still have common antigen(s) concerned with infectivity. MuG(+) was partially neutralized with anti-P1cinC(+) serum as reported previously (Toussaint et al. 1978), but not with anti-P1cinC(-) serum, moreover, MuG(-) was neutralized with anti-P1cinC(-) serum to the same extent of homologous combination, but not with anti-P1cinC(+) serum. This suggests that the C and G segment might express infectious specificity similar to each other only when they are in the same orientation.

3. Identification of polypeptides characteristic of the orientation of the C segment.

To detect the difference between the two mutants at the polypeptide level, whole proteins of purified particles of P1cinC(+), P1cinC(-) and P1cin$^+$ were analyzed by SDS-PAGE. MuG(+) proteins were also examined since the partial neutralization was detected by anti-P1cinC(+) serum. A number of polypeptide bands were observed on the gradient polyacrylamide gel (10-20%) (Fig. I-2A) as reported by Walker and Walker (1981). One band of approximately 100 kDa seemed to be different between the two mutants, and any other difference was not detected among the bands with the molecular weight higher than 44 kDa. The bands of the molecular weight lower than 44 kDa were too thin to be compared

precisely between the two mutants. The number of polypeptides composing MuG(+) was much fewer than that of P1, and there seemed to be no band that corresponded to the band of P1cinC(+) in mobility. To detect clearly the difference in 100 kDa bands between P1cinC(+) and P1cinC(-) particles, the samples were run on the 7.5% gel (Fig. I-2B). It was demonstrated that P1cinC(+) particles had a 105 kDa polypeptide, and P1cinC(-) had 101 kDa instead of 105 kDa, and that P1cin$^+$ particles had both polypeptides. When polypeptides were prepared from P1cin$^+$ particles that were propagated with the hin pin strain by infection, not by induction, only 105 kDa polypeptide was detected. The 105 kDa polypeptide has been inferred to be a main component of the tail fiber (Walker and Walker 1983). The two P1cin mutants used differ only in orientation of the C segment. Taken these together, the results suggest that the two polypeptides identified are concerned with host specificity of each mutant and encoded by the genes on the C segment.

4. Cloning of the C segment.

To make sure that the C segments of P1cinC(+) and P1cinC(-) mutants are opposite in orientations, the BamHI-5 fragments of P1cin$^+$, P1cinC(+) and P1cinC(-) were cloned on pBR322, and resulting three hybrid plasmids were designated pAT101, 102 and 104, respectively. They were purified from the hin pin strain, digested with HincII and BamHI, and electrophoresed (Fig. I-3). The digests of pAT102 and 104 were different from each other in the two bands (the second and third bands from the top) whose length became equal when added up while the digests of pAT101 showed the four bands, two of which corresponded to those of either pAT102 or 104. These indicate that P1cin$^+$ has the invertible C segment and that P1cinC(+) and P1cinC(-) have the C segments fixed in the opposite orientations.

5. Detection of gene products encoded on the C segment.

Subcloning of the BamHI-5 fragment.

To identify the gene products encoded on the C segment, 7.6 kb BamHI-5 fragment of pAT101 was subcloned into the BamHI site of expression vector pDR540. The BamHI site of pDR540 is located downstream of a strong tac promoter. Two kinds of recombinant plasmids which contain the BamHI-5 fragments in the opposite orientations were isolated, and designated pHK101 and pHK104. The direction of the inserted fragments were determined by HindIII and BglII digestion followed by agarose electrophoresis. In pHK101 the transcriptional direction of the tac promoter and the cin gene face in the opposite direction, in pHK104 they are in the same direction.

Maxicell analysis.

To analyze the product of tail fiber genes inferred in the C segment, plasmid-coded polypeptides in maxicell were separated on a 12% polyacrylamide gel (Fig. I-4). Four polypeptide bands, approximately 100 kDa, 24 kDa, 21 kDa and 16.5 kDa, could be reproducibly observed through three experiments in pHK101. pHK104 showed only one prominent polypeptide band, 21 kDa, except for weak bands, 16.5 kDa and 24 kDa. The 21 kDa polypeptide was inferred to be Cin protein, since the transcriptional direction of the tac promoter and the cin gene are the same. Approximately 100 kDa polypeptide may correspond to both 101 kDa and 105 kDa polypeptides identified above. The other two bands, of 16.5 kDa and 24 kDa, may correspond to U or U' gene products of phage Mu by analogy between the C segment and the G segment. Since the plasmid pHK101 contains cin^+ gene the polypeptides detected in maxicell were inferred to be total products of the C segment in the both orientation. Thus, it was not determined which polypeptide band correspond to the product specific to the orientation of the C segment.

DISCUSSION

In the present experiments, the *cin* mutants of phage P1 were isolated as lysogens of the *hin pin* strain by utilizing the finding that *cin*$^+$, *hin*$^+$ and *pin*$^+$ genes can be substituted for one another with respect to phase variation (Kutsukake and Iino 1980a, Enomoto et al. 1983). To lysogenize the *hin pin* strain with non-infectious P1*cin*C(-), P1*cin* particles that retain the C(-) orientation, but are infectious to *E. coli* K-12 were prepared by phenotypic mixing and infected to the *hin pin* strain. Of the Cm^r colonies obtained, approximately 20% were the P1*cin*C(-) lysogens. It was found that the rate could increase nearly to 100 % when the *hin pin*/P1*cin*C(-) lysogen was again used for phenotypic mixing. Such lysate would be very useful to isolate P1*cin*C(-) lysogens from strains with various genetic backgrounds. P1*cin*C(-) prophages yielded P1*cin*C(+) ones at the order of 10^{-2} per bacterium in the *hin*$^+$ *pin*$^+$ strain, suggesting that *hin*$^+$ and/or *pin*$^+$ mediate inversion of the C segment at low frequency. Although fixed orientation of the C segment in *hin pin*/P1*cin*C(-) (EJ1534) seemed to be stable enough for biochemical analysis, this strain yielded the lysogens with C(+) orientation at the low rate of approximately 10^{-6} per bacterium. This suggests that P1*cin*C(-) isolated has a leaky *cin* mutation able to revert to *cin*$^+$ at low frequency.

In the course of the experiment, P1*cin*C(-) was found to be able to form plaques on the wild strain of *E. coli* C and *Sh. sonnei* in the presence of Mg^{2+}. The host range of phage Mu is known to somewhat resemble that of P1 (Iida et al. 1982, van de Putte et al. 1980), and MuG(-) can form plaques on *E. coli* C, *Sh. sonnei*, *C. freundii* (van de Putte et al. 1980), *Erwinia* spp., *Enterobacter cloacae*, and *Serratia marcescens* (Sandulache et al. 1985), some of which were also found to be hosts for P1*cin*C(-) as described above.

In the neutralization tests, it raised the question that anti-P1*cin* sera might contain antiserum against the particles with the C segment in opposite orientation to the extent that affects the cross-neutralization rate. However,

this possibility was excluded by the fact that anti-P1cinC(-) serum did not inactivate MuG(+) at all; otherwise the serum will neutralize MuG(+) at a detectable rate.

It has been reported that the genes determining infectious properties characteristic of the C(+) and the C(-) orientations are located on the right and the left half of the segment, respectively when the segment is in the (+) orientation (Iida et al. 1982, Iida 1984). Each gene coding for 105 or 101 kDa polypeptide was estimated to be approximately 3 kb long in this study, and the C segment has 3 kb DNA flanked by 0.6 kb inverted repeats (Chow and Bukhari 1976); therefore, both genes will have a common coding region extending outside the C segment, and one of genes will correspond to gene 19 as has been already suggested (Iida 1984, Walker and Walker 1983). That the two products have the common coding region might explain, as one possibility, the cross neutralization between anti-P1 serum and heterologous P1 phages. Although MuG(+) was partially inactivated by anti-P1cinC(+) serum, MuG(-) was considerably inactivated by anti-P1cinC(-) serum. This indicates that P1cinC(+) and P1cinC(-) phages are antigenically homologous to MuG(+) and MuG(-), respectively. When the analogy with the G segment of Mu was taken into account (Chow and Bukhari 1976, Toussaint et al. 1978), it was expected that at least two sets of polypeptides could be detected in SDS-PAGE as the difference between the two P1cin mutants; difference in only one set of polypeptides, 105 and 101 kDa, was detected between P1cinC(+) and P1cinC(-) phages. Another one set of polypeptides, 16.5 and 24 kDa, was detected in maxicell. It was not determined which polypeptide correspond to the product specific to either the (+) orientation or the (-) orientation of the C segment. In SDS-PAGE, however, the bands with lower molecular weights were always too thin to make the comparison between the two mutants. Recently, it was reported that a polypeptide of 22.5 kDa is encoded within the C segment, which is presumed to be the analogue of the Mu U or U' gene product (Huber et

al. 1985). The 24 kDa polypeptide detected in maxicell may correspond to this 22.5 kDa polypeptide.

Table I-1. Bacterial strains used in this study

Strain	Relevant markers and phages	Source
EJ1076	*fliC-i hin fljB-e,n,x fljA*$^+$ *pin*$^+$	Enomoto et al. (1983)
EJ1346	*fliC-i hin*$^+$ *fljB-e,n,x fljA*$^+$ *pin*$^+$	Enomoto et al. (1983)
EJ1420	*fliC-i hin fljB-e,n,x fljA*$^+$ Δ*pin*	This work
EJ1451	EJ1420/P1*cin*$^+$	This work
EJ1452	EJ1420/P1*cinC*(+)	This work
EJ1456	EJ1346/P1*cinC*(+)	This work
EJ1457	EJ1346/P1*cinC*(-)	This work
EJ1534	EJ1420/P1*cinC*(-)	This work
E. coli C	wild type	Natl. Inst. Genetics

Table I-2. Characters of P1*cin* lysogens

Strain	*hin* *pin*/P1*cin*	IC/CFU[a]	Titer of lysate/ml[b]	Rate of nonlysogens[c]
EJ1452	- - / C(+)	1.0	10^8-10^9	2.6 x10^{-3}
EJ1456	+ + / C(+)	1.1	10^8-10^9	4.1 x10^{-4}
EJ1457	+ + / C(-)	1.5 x 10^{-2}	10^5-10^6	1.3 x10^{-3}
EJ1534	- - / C(-)	3.9 x 10^{-6}	10^2-10^3	1.8 x10^{-4}

a, The number of bacteria capable of being infective centers (IC) at 41°C was divided by the number of colony-forming units (CFU) at 30°C.

b, Aliquotes of the thermally induced lysate were spotted on the lawn of indicator bacteria (EJ1420) and incubated at 37°C.

c, The number of CFU at 41°C was divided by the number of CFU at 30°C.

Table I-3. Efficiencies of plating of P1cin mutants

Phage	Host	Efficiency of plating on E. coli C: Mg^{2+}	Efficiency of plating on E. coli C: Without Mg^{2+}	Host: Mg^{2+}
P1cin^{+}	E. coli K-12 (EJ1420)	1.0	5.6×10^{-1}	5.4×10^{-1}
P1cinC(+)	E. coli K-12 (EJ1420)	1.0	0.9	1.0
P1cinC(-)	E. coli K-12 (EJ1420)	1.0	5.3×10^{-6}	5.8×10^{-6}
P1cinC(+)	Sh. sonnei	1.0	1.0	1.1
P1cinC(-)	Sh. sonnei	1.0	7.5×10^{-5}	1.0
P1cinC(+)	C. freundii	1.0	9.6×10^{-1}	$<10^{-7}$
P1cinC(-)	C. freundii	1.0	4.2×10^{-6}	$<10^{-7}$
P1cinC(+)	E. aerogenes	1.0	1.0	$<10^{-6}$
P1cinC(-)	E. aerogenes	1.0	1.8×10^{-6}	$<10^{-6}$

The efficiency of plating on E. coli C in the presence of Mg^{2+} was set at 1.0. Phage lysates were titrated by the drop-on-lawn method (Ornellas and Stocker 1974) on NA, and at least 0.05 ml of lysate or its dilutions were spotted for each sample. $MgSO_4$ was used at concentration of 20 mM.

Fig. I-1. Neutralization of P1 and Mu by anti-P1cinC(+) serum (A) and by anti-P1cinC(-) serum (B). Phage lysate was diluted in buffer to 5×10^{-7} pfu/ml and mixed with antiserum [(○), P1cinC(+); (□), P1cinC(-); (△), MuG(+); (▽), MuG(-)]. At appropriate times samples were removed, diluted in buffer, and titrated on E. coli K-12 strain EJ1420 to assay P1cinC(+) and MuG(+) phages and on E. coli C to detect P1cinC(-) and MuG(-) phages.

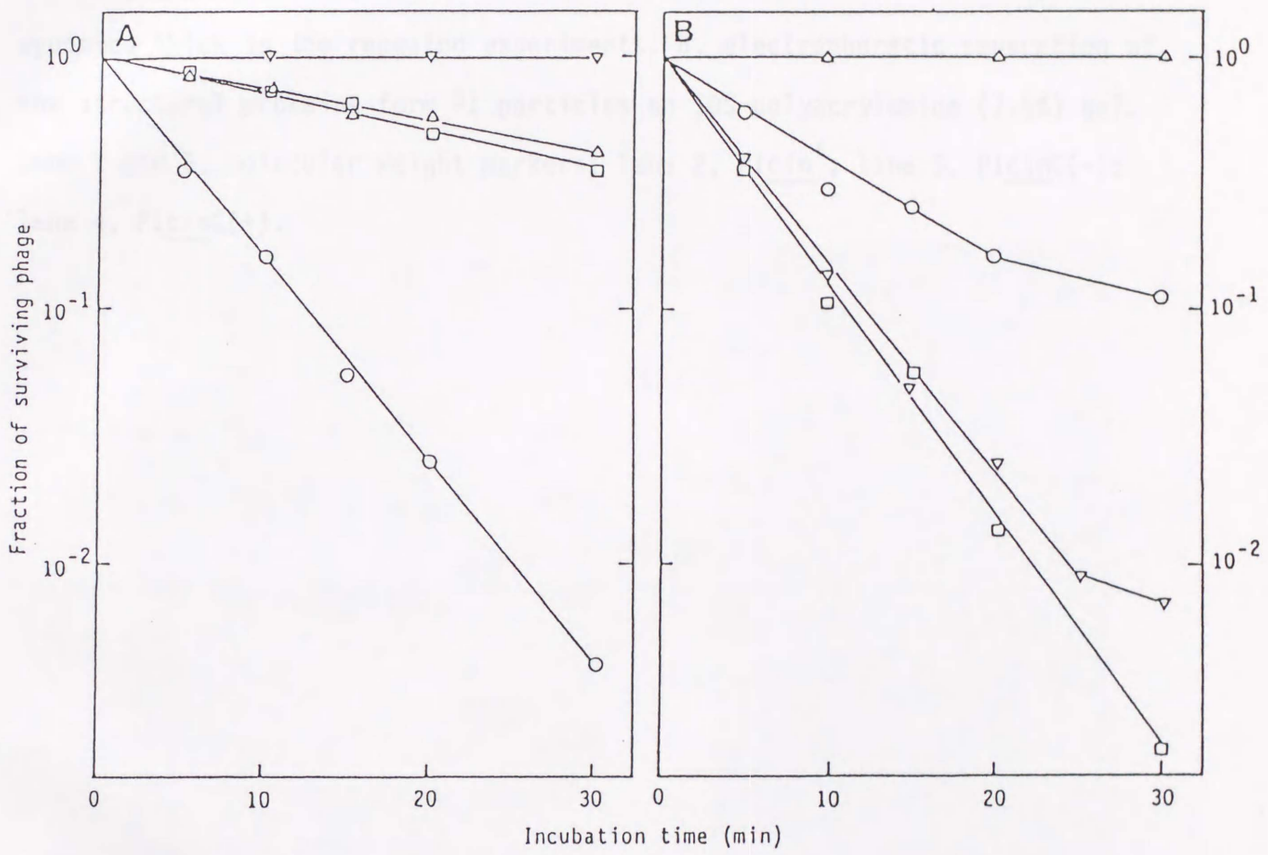
A
B
10^0
10^{-1}
10^{-2}
Fraction of surviving phage
0
10
20
30
Incubation time (min)

Fig. I-1.

Fig. I-2. A, electrophoretic separation of the structural proteins from P1 and Mu particles on SDS-polyacrylamide gradient (10-20%) gel. Lane 1 and 6, molecular weight markers; lane 2, P1*cin*$^{+}$; lane 3, P1*cin*C(-); lane 4, P1*cin*C(+); lane 5, MuG(+). The band corresponding to 12 kDa is thin in lane 3; this has been identified as a head component (Walker and Walker 1981) and appeared thick in the repeated experiments. B, electrophoretic separation of the structural proteins form P1 particles on SDS-polyacrylamide (7.5%) gel. Lane 1 and 5, molecular weight markers; lane 2, P1*cin*$^{+}$; lane 3, P1*cin*C(-); lane 4, P1*cin*C(+).

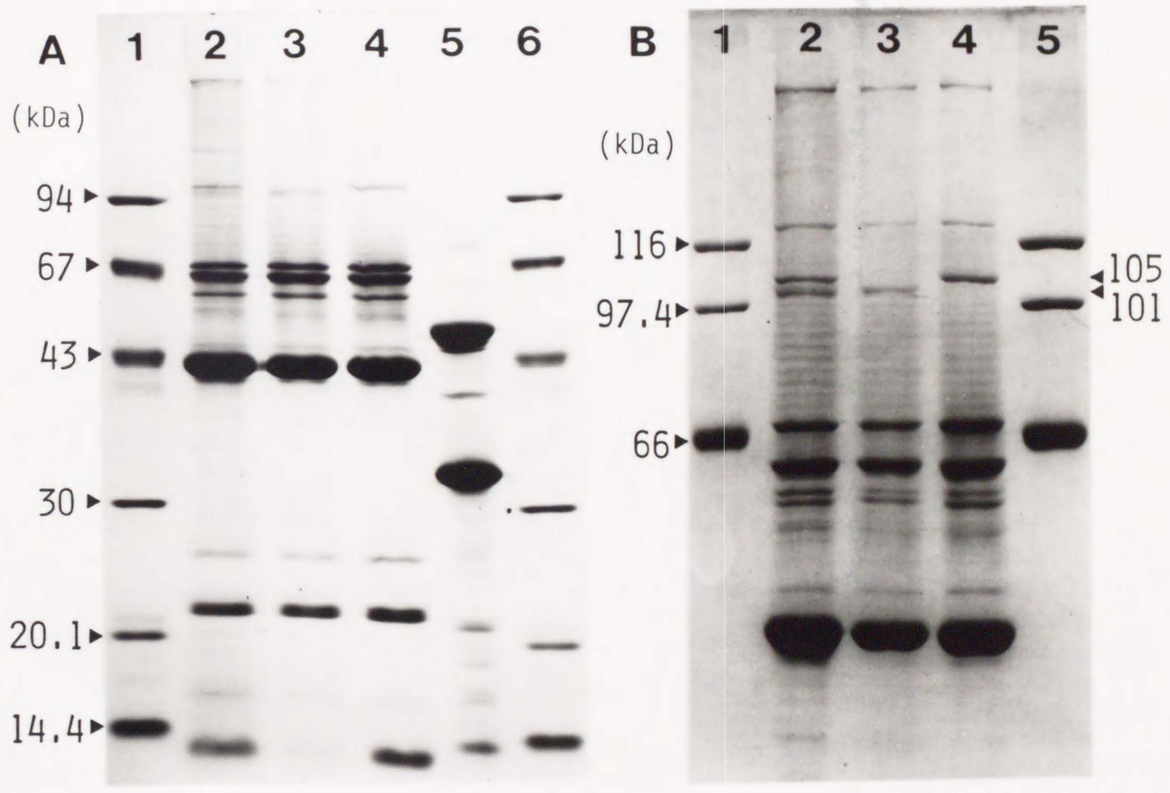

Fig. I-2.

Fig. I-3. A, electrophoretic patterns of HincII-BamHI digests of hybrid plasmids. Lane 1, pAT101; lane 2, pAT102; Lane 3, pAT104. B, the HincII-restriction map of the BamHI-5 fragment. Bold lines are 0.6 kb inverted repeats. Numbers represent kb of the fragments in both directions. B; BamHI, H; HincII.

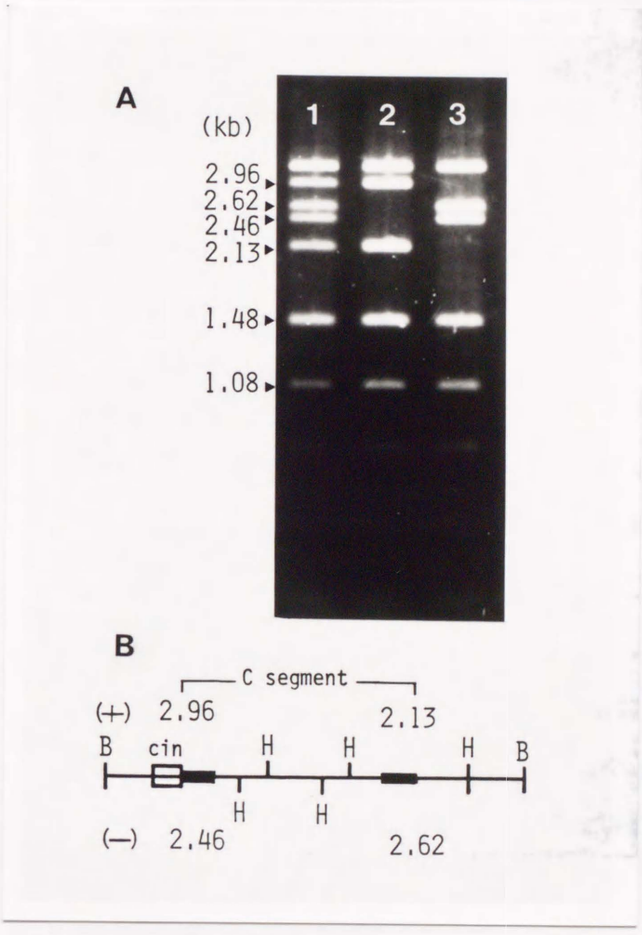
A
(kb)
1
2
3
2.96
2.62
2.46
2.13
1.48
1.08
B
C segment
(+)
2.96
2.13
B
cin
H
H
H
B
H
H
(—)
2.46
2.62

Fig. I-3.

Fig. I-4. [^{35}S]methionine-labeled polypeptides produced in maxicells by plasmids pHK101 and pHK104. Maxicell extracts were prepared from plasmid containing HB101 (recA) after UV irradiation. Polypeptides were separated on a 12% polyacrylamide gel and visualized by fluorography. Arrows indicate specific bands. Lanes: 1, pDR540; 2, pHK101; 3, pHK104.

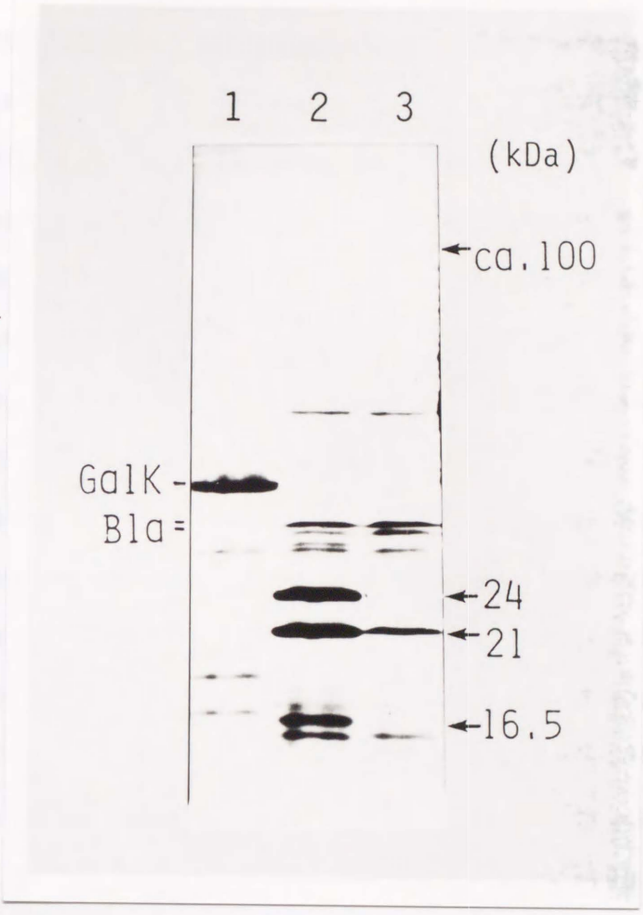

Fig. I-4.

II. Identification of the gene products encoded on the invertible P segment of the e14 element in *Escherichia coli* K-12

ABSTRACT

Many strains of *Escherichia coli* K-12 have the excisable element e14, which contains an invertible DNA segment. This invertible P segment and the *pin* gene responsible for inversion were cloned, together with other e14 sequence. The e14 element contains a gene which represses the growth of phage ø80. This repression was observed when the repressor gene is in the multicopy state. This repressor gene was mapped in a region outside the P segment. Depending on the orientation of the P segment, Two alternate sets of gene products, 16 and 22 kDa in P(+) and 28 and 32 kDa in P(-), were expressed. Two polypeptides of 16 and 32 kDa were inferred to be the products of internal two ORFs in the P segment, P145 and P293, respectively. Western blot analysis suggested that the constant amino-terminal region (Pcon) and the two carboxy-terminal regions, Pplus and Pmin, produce the other two polypeptides of 22 and 28 kDa, respectively, and that the Pcon region encodes a 11 kDa polypeptide. The amino acid sequences of these polypeptides, which are predicted from the nucleotide sequences, were homologous to those of tail fiber and its related gene products.

INTRODUCTION

Flagellar phase variation results from the alternate expression of two kinds of antigenically distinct flagella called phase 1 and phase 2 at a definite rate per bacterium per division (Stocker 1949). Phase 1 and phase 2 flagella are specified by the respective structure genes, *fliC* and *fljB*. H segment is a regulator region linked upstream of the *fljB* gene and considered to be inverted by the function of *hin* gene coded within the H segment (Zieg et al. 1977 and 1978, Kutsukake and Iino 1980). The *hin* mutation which prevents phase variation in *Salmonella*, was found to be phenotypically suppressed in the genetical background of *Escherichia coli* K-12. It was inferred that some gene function in *E. coli* K-12 would inverse the H segment at a definite rate instead of *hin*, and this gene, named *pin*, was located at around 26 min on the chromosome of *E. coli* K-12 (Enomoto et al. 1983). Similarly, Plasterk et al. (1983) also showed that *E. coli* K-12 strain HB101 and CSH520 had a site-specific recombinase gene, identical with *pin*. They reported the restriction map of the *pin* region containing its target region which was the 1,800 bp invertible segment linked to *pin* and named P segment. Furthermore, they mapped the *pin* locus between *purB* and *fabD* on the *E. coli* chromosome. In this region e14, a cryptic, UV-excisable, genetic element is situated (van de Putte et al. 1984, Greener and Hill 1980). Although, the DNA sequence of the *pin* gene and the P segment was determined and the existence of four open reading frames (ORFs) in the P segment was revealed, gene products of the four ORFs were not detected in minicell (Plasterk et al. 1983 and Plasterk and van de Putte 1985).

In this work, I describe the identification and characterization of products of four ORFs encoded on the P segment to disclose the function of the invertible P segment.

MATERIALS AND METHODS

Bacterial strains and plasmid vectors.

All bacterial strains used in this study were derivatives of *E. coli* K-12 and are listed in Table II-1. Strain EJ350 was a motile derivative of PA309 that was found to be *pin*$^+$, capable undergoing phase variation. EJ1076 was constructed from EJ350 by P1 transduction and have a *Salmonella* H1 and H2 region. Plasmid vectors and recombinant plasmids used in this study are listed in Table II-1. pBR322 was used as a vector in cloning of recombinase gene by the shotgun procedure and the restriction mapping of the inserted fragments. pMLmp10 was a vector derived from pBR322 and had a single *Bgl*II site and the *Eco*RI-*Hind*III multicloning site from bacteriophage M13mp10. pYEJ001 is a expression vector derived from pBR327 (Soberon et al. 1980) and contains a synthetic consensus *E. coli* RNA polymerase promoter, tandem synthetic lactose operators and a *Hind*III chloramphenicol acetyltransferase gene cartridge. pYEJ001 was used to investigate the gene products encoding in the P segment.

Purification of plasmid-coded polypeptides and preparation of antisera.

Small scale experiments revealed that the polypeptides, expressed and produced in large quantities from plasmid pPS140 and pPS240, can be recovered from sonicated cell lysate by centrifugation. Strain EJ2371 carrying plasmid was grown in 500 ml of TLY broth medium. Cells were washed once with buffer (50 mM Tris-HCl, pH 8), resuspended in 10 ml of this buffer and broken by sonication. Cell debris was removed by centrifugation for 20 min at 20,000 g. Polypeptides in buffer were precipitated with ammonium sulfate (50% saturation), resuspended in buffer, and dialyzed against the same buffer to exclude ammonium sulfate. Antisera, anti-P(+) serum and anti-P(-) serum, against the polypeptides produced from pPS240 and pPS140, respectively, were prepared as follows; approximately 0.5 to 1.0 mg/ml of purified polypeptide samples, which were recovered from a acrylamide gel, were emulsified with

Freund's complete adjuvant (Sigma) and injected subcutaneously to New Zealand White rabbits four times at 1- or 2-week intervals. Immune sera were obtained after 11 weeks. Contaminating reactivities to bacterial antigens were removed from sera by adsorption against boiled bacterial cells following the method of Helfman et al. (1983).

Western blotting.

Protein samples extracted from cells were electrophoretically transferred to the nitrocellulose membrane after fractionating by SDS-PAGE as described earlier. The electrophoretic blot was air-dried and soaked in the blocking solution [5% nonfat dry milk, 75 mM NaCl, 30 mM Tris-HCl (pH7.4)] for 30 min at a room temperature and soaked again for 3 hr in the blocking solution containing anti-P(+) serum or anti-P(-) serum at a concentration of 1/500. The sheet was washed by three changes of the washing buffer [20 mM Tris-HCl (pH7.4), 150 mM NaCl, 0.1% SDS] during total 15 min, subjected for 2 hr to the second antibody [perooxidase-conjugated rabbit anti-IgG (Amersham)] that was diluted to 1/500 with the blocking solution, and washed again with the buffer as described above. For the color reaction, the blot was soaked in the solution containing 4-Chloro-1-naphtol of (480 μg per ml), 0.01% H_2O_2, 10 mM Tris-HCl (pH7.4), and 100 mM NaCl. This solution was freshly prepared from the stock solution of 3 mg 4-Chloro-1-naphtol dissolved in 1 ml methyl alcohol and of 30% H_2O_2. The reaction was terminated after 20-30 min by washing with distilled water. The blot was air-dried and stored in dark.

SDS polyacrylamide gel electrophoresis containing 8 M urea (Urea-SDS-PAGE).

Urea-SDS-PAGE was the method originally developed by Storti et al. (1976), and the practical method followed the description (Schleif and Wensink 1981) except that the sample buffer contained 8 M urea, and samples were not boiled before electrophoresis.

RESULTS

1. Cloning of the P segment from *E. coli* K-12.

Since the function of P segment is unknown, no phenotypic marker can be used for the selection of the P segment. Thus, the *pin* gene and its adjacent P segment was cloned from the chromosomal DNA of W1485 into the *Hin*dIII site of the vector pBR322 by the shotgun procedure. The recombinant plasmids were transformed into EJ1079 (Δ*pin hin* fixed-*i*^{on}) and Tet^{r} *e,n,x*^{on} swarms were selected on NSS containing tetracycline and anti-*i* serum. One of the recombinant plasmids, pSI701, contained 11.2 kb *pin*$^{+}$ fragment. The restriction map are shown in Fig. II-1A and was almost identical with that reported by Plasterk et al. (1983). This fragment was judged to have the invertible region from the electrophoretic pattern (data not shown). When the plasmid was digested with *Pst*I, 7 bands (1.2, 1.3, 1.6, 1.9, 4.5, 4.9 and 6.2 kb) appeared, the total length of which was 17.3 kb, exclusive of a 4.3 kb vector fragment, this is longer than the cloned fragment by 6.1 kb. It was inferred that the extra 6.1 kb fragment (1.2 + 4.9 or 1.6 + 4.5) resulted from a invertible segment, which could have alternative orientation. The sites of inverted repeats were inferred from the length of *Pst*I fragments and found to be identical with those already reported (Plasterk et al. 1983).

2. Construction of plasmids that direct overproduction of polypeptides encoded on the P segment.

Construction of plasmids carrying the subcloned fragment of e14.

From the sequence of the P region it was revealed that the P segment contains four *orfs* (Plasterk and van de Putte 1985) ; two internal *orfs* (P293, 293 amino acids and P145, 145 amino acids) and two incomplete *orfs* (Pmin, 156 amino acids and Pplus, 107 amino acids) read into the invertible segment (in the same reading frame with in the inverted repeat) are present (Fig. II-1B).

Therefore, it is thought that orf Pmin and orf Pplus represent variable regions of a gene which starts with a constant segment (Pcon) in the non-inverting DNA, and that orf Pcon and orf Pmin (or orf Pplus) are couple in the (-) [or (+)] orientation to create orf Pc-min (or orf Pc-plus). To identify the gene products encoded on the P segment two recombinant plasmids, which have a fixed P segment in one orientation, were constructed. These plasmids were prepared from pSI701 by deleting the pin gene. The 11.2 kb HindIII fragment which contained the complete pin gene and invertible P segment was isolated from pSI701 and cloned into the HindIII site of the vector pMLmp10 derived from pBR322. The plasmid isolated was named pSI703. pMLmp10 has a single BglII site and the 11.2 kb HindIII fragment also has one BglII site in the pin gene. Therefore, pSI703 was digested by BglII and ligated to isolate the plasmid deleting most part of the pin gene and a part of pMLmp10. The ligated DNA was transformed into EJ1709 (Δpin hin fixed-i^{on}) and nonmotile transformants on NSS containing anti-i serum were screened. Plasmid DNA was isolated from the nonmotile transformant and analyzed by PstI digestion followed by agarose gel electrophoresis. As a result of analysis two types of plasmids, one containing the P segment frozen in (+) orientation, and the other containing the P segment frozen in (-) orientation, were isolated and named pSI704 and pSI705, respectively (Fig. II-1A). First, pSI704 and pSI705 were digested by BamHI and ligated to isolate the plasmids deleting two (1.6 kb and 5.7 kb) BamHI fragments, since three BamHI sites are present in these plasmids. The resultant plasmids were named pPS130 and pPS230. Then, 2.8 kb EcoRI-BglII fragments of pPS130 and pPS230 were subcloned into the expression vector pYEJ001. pPS130 and pPS230 were codigested with EcoRI and BglII, and fragments were separated by agarose gel electrophoresis. 2.8 kb EcoRI-BglII fragments were recovered from the gel and ligated in the EcoRI and the BamHI sites of pYEJ001, since the cleavage of both BamHI and BglII generate compatible cohesive termini, to transcribe genes encoded in the P segment from

the synthetic consensus promoter of pYEJ001. The resulting plasmids derived from pPS130 and pPS230 were designated pPS140 and pPS240, respectively (Fig. II-1A).

Characterization of strains carrying the subcloned fragments of e14.

It has been noted previously that the e14 element is a nonessential constituent of the E. coli K-12 chromosome and the curing causes no change in growth rate or growth requirements (Greener and Hill 1980). To detect the gene dosage effect of parts of e14 element, the growth rate, growth requirements and sensitivity for bacteriophages were examined in E. coli C (recA) background. Plasmid pSI701 and its derivatives were transformed into E. coli C (recA) and Amp^r transformants were isolated. The growth rate and growth requirements of these transformants were not different from their parental strain (data not shown). The sensitivity for many phages was tested by spot method. Phage lysates of T2, T3, T4, T5, T6, T7, P1vir, P2vir, ø186, λ and ø80 were appropriately diluted and spotted on the lawn of the strain containing one of these plasmid. One exception of ø80, other phage lysates produced many plaques on every strains examined. The efficiency of plating of lambdoid phages were examined and the result is shown in Table II-2. Three strains containing the plasmids, of pSI701, pSI704 and pSI705, restricted the plaque formation of ø80 and BF1 (hλi80) which has λ host range and ø80 immunity (Coste and Bernardi 1987) but two strains containing the plasmids, of pPS140 and pPS240 did not restrict the growth of all phages used, except phage P1 which shows poor growth on the recA background. Since phage ø80 and BF1 have the ø80 immunity region which is negatively regulated by the ø80 repressor, the restriction against these phages may reside in the repression by some protein from e14. There was no difference in EOP between the strains carrying pSI705 or pPS140 and those carrying pSI704 or pPS240, respectively. Therefore, the inversion of the P segment may have no effect on the restriction and the

growth of phages examined.

3. Identification of the plasmid-coded polypeptides.

The polypeptides produced from the plasmids constructed were examined. The polypeptide pattern of strains carrying the plasmid pPS140[P(-)] or pPS240[P(+)] was analyzed by SDS-PAGE. Plasmids pPS140 and pPS240 were transformed into EJ2371 (Δ*lac* *recA*) to avoid the partial repression of *lac* operators by *lacI* gene. Samples of whole cell lysate were separated on an Urea-SDS-PAGE (Fig. II-2). Four polypeptide bands were specifically overproduced; 28 kDa and 32 kDa polypeptides are specific in pPS140[P(-)], and 16 kDa and 22 kDa are in pPS240[P(+)]. Since two internal ORFs P293 (a molecular weight of 32,831) and P145 (16,726) are predicted in the P(-) and P(+) orientation, respectively, 32 kDa and 16 kDa polypeptides were inferred to correspond to products of these two *orfs* *P293* and *P145*, respectively. Since the molecular weight of ORF Pmin and ORF Pplus are calculated at 16,734 and 11,041, the other two polypeptides 22 and 28 kDa may correspond to products of two *orfs* *Pc-plus* and *Pc-min*, respectively, assuming that the *orf* *Pcon* region encodes approximately 11 kDa polypeptide. The ORF Pcon region coding for 11 kDa polypeptide is estimated to be approximately 0.3 kb long in this study, therefore the transcriptional start of *orf* *Pcon* may be present in the subcloned fragments of pPS140 and pPS240.

4. Amino acid sequence analysis of four ORFs.

To obtain a clue to four uncharacterized polypeptides the predicted amino acid sequences of four ORFs were compared with the National Biochemical Research Foundation protein sequence database using DNASIS sequence analysis software. Partial homology was found with the tail fiber and related genes of bacteriophages in three ORFs (Pmin, P293 and P145). Their amino acid sequences are compared with ORFs in the P segment (Fig. II-3). It was found that the

predicted amino acid sequence of ORF Pmin shows more than 60% homology with the amino acid sequences of the tip of phage T4 tail fiber (gene 37) (Oliver and Crowther 1981) and an ORF B314 in the phage λ genome (Sanger et al. 1982) (Fig. II-3, A and C). Moreover, the predicted amino acid sequence of ORF P293 shows more than 60% homology with the amino acid sequences of the tail assembly polypeptide of T4 (gene 38) (Oliver and Crowther 1981) and an ORF A194 (Sanger et al. 1982) in the phage λ genome immediately downstream of ORF B314 (Fig. II-3, B and D). The predicted amino acid sequence of ORF P145 also shows 62% homology with the amino acid sequence of the ORF A194. The homology found is summarized in Fig. II-3, F. Although, most of the homologous region between the ORFs predicted in the P segment and other ORFs is small, these three sets of ORFs are similar in organization. ORF Pplus had no significant homology with proteins examined, except for weak homology with products of bacteriophage genes (including K3-gp36, IKe-geneIV and λ-B314).

5. Western immunoblot analysis.

The amino acid homology found between the ORFs predicted in the P segment and two ORFs from phage T4 and λ suggest some relationship between the e14 element and these phages. In order to examine a relation among these ORFs, immunological analysis of these phage polypeptides were carried out. The polypeptides over produced in the strain carrying pPS140 or pPS240 were purified by stepwise addition of solid ammonium sulfate (Fig. II-4). Anti-P(-) and anti-P(+) sera were prepared against purified one set of polypeptides 28 kDa and 32 kDa, and another set of polypeptides 16 kDa and 22 kDa, respectively.

To examine the specificity of antisera prepared against the polypeptides encoded in the P segment the purified polypeptides, which are specific to the orientation of the P segment, were analyzed by Western blotting. The result is shown in Fig. II-5. Anti-P(+) serum reacted strongly with two polypeptides, of

16 and 22 kDa, and weakly with one polypeptide of 28 kDa. On the other hand, anti-P(-) serum reacted strongly with two polypeptides, of 28 and 32 kDa, and weakly with one polypeptide of 22 kDa. These results indicate the cross reaction between anti-P(+) serum and 28 kDa polypeptide and between anti-P(-) serum and 22 kDa polypeptide, because two sets of polypeptides, 16 and 22 kDa and 28 and 32 kDa, are specific to the orientation of the P segment, P(+) and P(-), respectively. This cross reaction between two polypeptides, of 22 and 28 kDa, and antisera is consistent with the idea that 22 kDa and 28 kDa polypeptides correspond to the products of two ORFs, of Pc-plus and Pc-min, respectively.

To detect the immunological relation between the ORFs encoded in the P segment and the tail fiber polypeptides of phages that showed high homology with these ORFs, the polypeptides of phage particles, of λ and T4, were analyzed on Western blotting with anti-P(+) and anti-P(-) sera. None of the polypeptides of phages examined reacted with these antisera (data not shown). Since two ORFs of phage λ is encoded b region, which is not essential for lytic growth, these ORFs may not be contained in phage particle or not be produced. To examine these possibility the crude induction lysates of a λ lysogen were analyzed by Western blot. Although, the component polypeptides of phage λ particle were detected with anti-λ serum, no polypeptide was detected with anti-P(+) and anti-P(-) sera. Since the products of gene 38 (gp 38) of phage T4 catalyzes assembly of tail fiber tip (gp 37), the gp 38 cannot be detected in this experiment. The polypeptides of phage ϕ80, which cannot grow on the strain carrying a plasmid containing a part of e14 element, was also examined with anti-P(+) and anti-P(-) sera, but there was no positive reaction (data not shown).

DISCUSSION

The invertible P segment of *E. coli* K-12 was cloned and gene products encoded in the P segment and its adjacent region were identified. In the (+) orientation of the P segment 16 and 22 kDa polypeptides were detected, and 28 and 32 kDa polypeptides were in the (-) orientation of the P segment. The genetic organization of the P segment has been predicted by Plasterk and van de Putte (1985), and the P segment contains two complete (*P145* and *P293*) and two incomplete (*Pplus* and *Pmin*) *orfs*. Based on the molecular weight of ORFs, which is calculated from the predicted amino acid sequences, and the orientation of the P segment, two polypeptides, 16 and 32 kDa, were inferred to be the products of two *orfs*, *P145* and *P293*, respectively. Another two incomplete ORFs, Pplus and Pmin, were inferred to be part of the two polypeptides, 22 and 28 kDa, respectively, based on their molecular weight, assuming that the constant region (Pcon) of both ORFs encodes 11 kDa polypeptide. Thus, the products of four *orfs* were identified in this experiment, although previous report (Plasterk and van de Putte 1985) did not observe any polypeptide band in minicell, except for a weak band corresponding to 14 kDa polypeptide in the P(-) plasmid. In this experiment, the band corresponding to 14 kDa polypeptide could not be detected. The reason that no expression of polypeptides coded in the P region was observed in minicell is that sequences further away than 1 kb from the P segment has no promoter sequence, since the *tac* promoter located 1 kb upstream of the P segment resulted in overproduction of polypeptides in this experiment.

The e14 element has the characters of a prophage. This element is excised as a 14.4 kb DNA circle on induction of SOS system (Greener and Hill 1980) and has an attachment site in the *E. coli* chromosome (Brody and Hill 1988). This element is nonreplicative and has no plaque forming activity (Brody and Hill 1988). As shown here a repressor gene for ø80 is presumed on this element. It was implied that the e14 element is a defective prophage. Although, it

contains inversion system homologous to phage Mu and P1, there is no similarity to the G or C segment either in the DNA sequence or in the amino acid sequences. In the comparison of amino acid sequences, however, two amino acid sequences of ORF Pmin and ORF P293, which are expressed on the (-) orientation of the P segment, are homologous to those of the tail fiber related gene products. The T4 gp37 and λ ORF B314 are homologous to ORF Pmin, and the T4 gp38 and λ ORF A194 are homologous to ORF P293. Recently, it was reported that the T4 gp38 can be functionally replaced by the product of orf A194 (Montag and Henning 1987) and that part of ORF B314 plus ORF A194 can functionally replace a C-terminal area of T4 gp37 and all of gp38, respectively (Montag et al. 1989). orf B314 and gene 37 are located immediately upstream of orf A194 and gene 38, respectively. Since the organization of two orfs, Pmin and P293, in the P segment is similar to that of T4 genes and λ orfs in their genomes, the function of ORF Pmin and ORF P293 may correspond to that of gp37 and gp38, respectively. In immunological analysis, however, anti-P(-) serum, which was prepared against 28 and 32 kDa polypeptides, did not react to the components of phage T4 and λ. Although the comparison of amino acid sequences showed partial homology between the ORFs in the P segment and the tail fiber related proteins, the function of the invertible P segment is still unclear. It does not seem a degeneration of the G or C segment, but it may belong to another group of invertible segment. (i) Although the product of genes encoded in the G and C segment are components of the tail fibers, one of the ORFs in the P segment may be a tail assembly polypeptide. (ii) The e14 element has a repressor gene for phage ϕ80, which is a member of lambdoid phage.

Table II-1. Bacterial strains and plasmids used in this study

Strain or plasmid	Relevant characteristics
E. coli strains	
EJ1076	*fliC-i hin fljB-e,n,x fljA*$^+$ *pin*$^+$
EJ1709	*fliC-i hin fljB-e,n,x fljA*$^+$ Δ*pin recA56*
EJ2371	*hsdR* Δ(*lacI-A*)*X74 recA56* e14°
W1485	e14$^+$
E. coli C	e14° *recA56*
Plasmids	
pBR322	Tet^r Amp^r ColE1-derived replicon
pMLmp10	Amp^r pBR322-derived replicon
pYEJ001	Tet^r Amp^r Cm^r synthetic promoter vector

Table II-2. Efficiencies of plating of lambdoid phages

Phage	Efficiency of plating on *E. coli* C					
	None	pSI701	pSI704	pSI705	pPS140	pPS240
λ	1.0	1.0	1.0	1.0	1.0	1.0
ϕ80	1.0	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$	1.0	1.0
λi434	1.0	1.0	1.0	1.0	1.0	1.0
BF1(hλi80)	1.0	$<10^{-8}$	$<10^{-8}$	$<10^{-8}$	1.0	1.0
F82(h80iλ)	1.0	1.0	1.0	1.0	1.0	1.0

The efficiency of plating on *E. coli* C containing no plasmid was set at 1.0. The titer of all the phages on *E. coli* C was between 1.2×10^{9} and 3.9×10^{11}/ml.

Fig. II-1. (A) Restriction maps of the fragments containing the P segment. pSI701 carries the 11.2 kb HindIII fragment with P-pin from W1485. pSI704 and pSI705 carry the 10.1 kb HindIII-BglII fragment with a fixed p segment. pPS140 and pPS240 carry the 2.8 kb EcoRI-BglII fragment with a fixed fragment. Boxed arrowheads show inverted repeat sequences. Solid arrow represents the pin gene. Restriction cleavage site: B, BamHI; Bg, BglII; E, EcoRI; H, HindIII; and P, PstI. Restriction sites within the P segment are shown with two orientations. (B) Diagrammatic representation of ORFs. ORFs are indicated by open arrows and the number of amino acids in each ORF, except Pcon, is shown in parentheses.

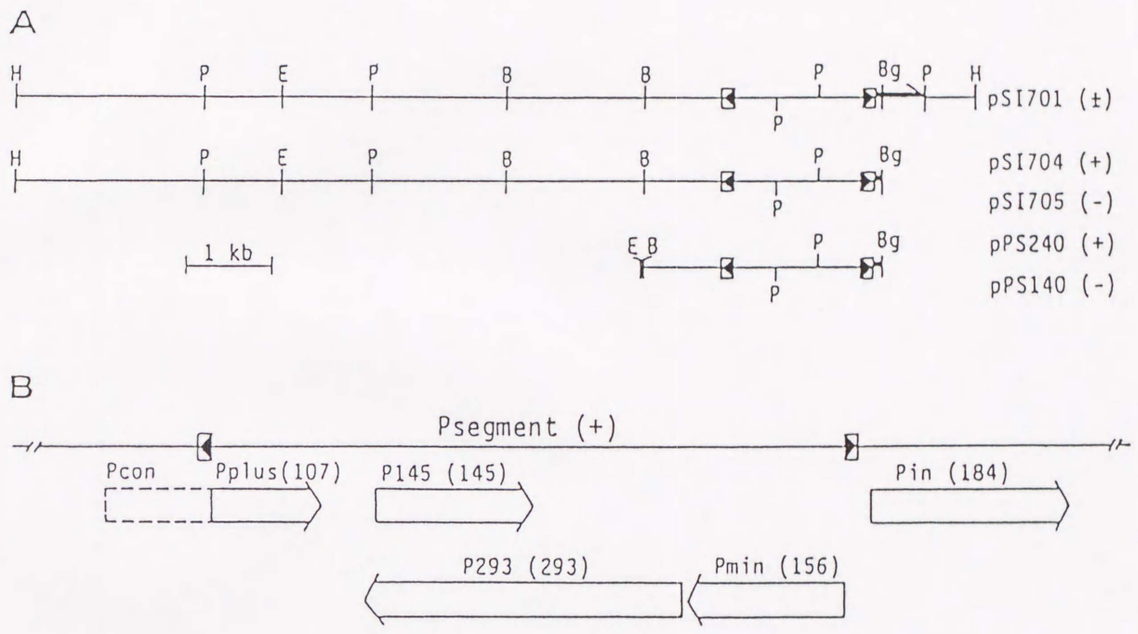
A
H
P
E
P
B
B
P
Bg
P
H
pSI701 (±)
P
pSI704 (+)
pSI705 (-)
1 kb
E B
pPS240 (+)
pPS140 (-)
B
Psegment (+)
Pcon
Pplus(107)
P145 (145)
Pin (184)
P293 (293)
Pmin (156)

Fig. II-1.

Fig. II-2. Urea-SDS-polyacrylamide gel electrophoresis of crude extracts prepared from strain EJ2371 containing pPS140 and pPS240. Lanes: 1, crude extract prepared from EJ2371 harboring pYEJ001; 2, crude extract prepared from EJ2371 containing pPS140; 3, crude extract prepared from EJ2371 containing pPS240; 4, molecular mass standards.

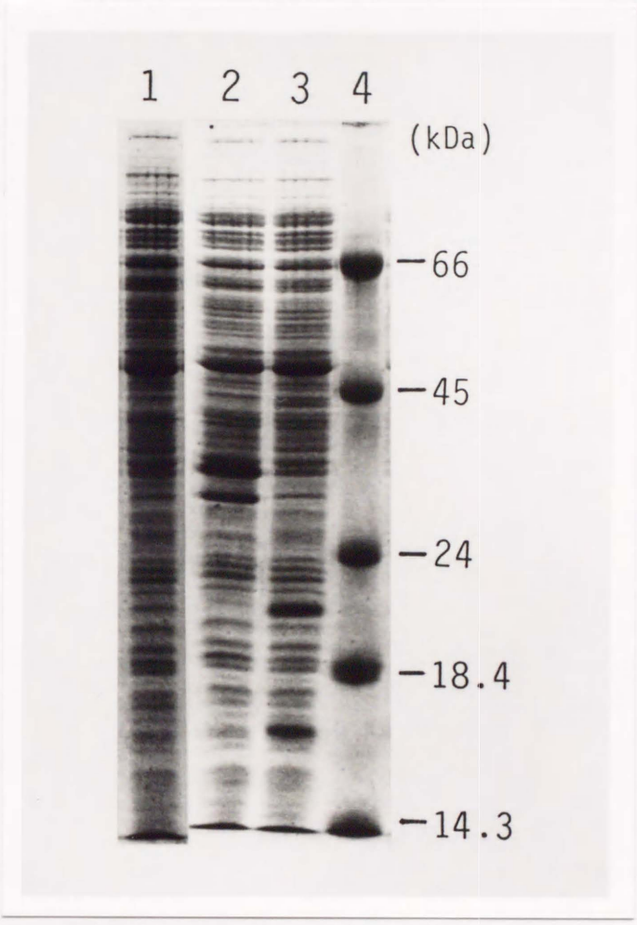

Fig. II-2.

Fig. II-3. Comparison of the putative polypeptides encoded on the P segment with the tail fiber related genes and ORFs. Amino acids are represented by the single letter code. The amino acid sequences of bacteriophage T4 37 and 38 gene products and λ two ORFs, A194 and B314, are indicated as gp37, gp38, A194 and B314, respectively. Panels: A, Pmin versus gp37; B, P293 versus gp38; C, Pmin versus B314; D, P293 versus A194; E, P145 versus A194; F, Diagrammatic presentation of homology comparison of ORFs. ORFs are indicated by open arrows and the number of amino acids in each ORF is shown in parentheses. Numbers indicate the percentage of amino acid homology between ORFs coded in the P segment and tail fiber related ORFs.

A

```
PMIN     1  SALPVGVPVPWPSATPPTGW-LKCNGAAF--SAEEYPELAKAYPTNKLPDLRGEFIRGWD  57
gp37   811  SSYPIGAPIPWPSDSVPAGFALM-EGQTFDKSA--YPKLAVAYPSGVIPDMRGQTIKG-- 865

PMIN    58  DGRGIDTGRSILSIQGYA--TEDHAHGLPSRSTIVTD-ATINFYFDEIWVNS-GTDII-- 111
gp37   866  --KP--SGRAVLSAEADGVKA--HSHSASASSTDLGTKTTSSF--D--YG-TKGTNSTGG 914
```

B

```
P293    18  VTVYNYDGETREYISTSNEYLAV-GVGIPACSCLDAPG-THKAGYAICRSADFN-SWEYV  74
gp38     1  MKIYHYYFDTKEFYKE-ENYKPVKGLGLPAHSTIKKPLEPKEG-YAVVF--D-ERTQDWI  55

P293    75  PD--HRGE-TVYSTKTGESKEIKAPGD-Y-PENTTTIAPLSPYDKWDG-EKW-V-TD    123
gp38    56  YEEDHRGKRA-WTF---N-KE-----EIFISDIGSPVG-IT-FD--EPGE-FDIWTD     97
```

C

```
PMIN     1  SALPVGVPVPWPSATPPTGWLKCNGAAF--SAEEYPELAKAYPTNKLPDLRGEF-IRGWD  57
B314    68  SAFPAGAPIPWPSDIVPSGYVLMQGQAFDKSA--YPKLAVAYPSGVLPDMRG-WTIKGK- 123

PMIN    58  DGRGIDTGRSILSIQGYA--TEDHAHGLPSRSTIVTD-ATINFYFD               100
B314   124  -P-A--SGRAVLSQEQDGIKS--HTHSASASGTDLGTKTTSSFDYG               163
```

Fig. II-3.

D

```
P293    11  IA--TKAGDVTVYNYDGETREYISTSNEYLAVGVG-IPACSCLD-APG--THKAGY-AIC  63
            ..  ...  . .::  . : :.:. .. :.. . . .:: :  : ::.  .  ::. :.
A194     5  MSEQPRT--IKIYNLLAGTNEFIGEGDAYIP-PHTGLPANST-DIAPPDIP--AGFVAVF  58

P293    64  RSADFNS-WEYVPDHRGETVYS-TKTGESKEIKA--PGDYPENTTTIAPLSP-YDKWDGE 118
             : : .. :  : :::: :::  . .:..  : .  ..  ::: : ..: .. :.::.:
A194    59  NS-D-EASWHLVEDHRGKTVYDVA-SGDALFI-SELGPL-PENFTWLSP-GGEYQKWNGT 112

P293   119  KWVTDTEAQHSAAV-DAAEAQRQSLID-AA--MASI--SLIQLKLQAGRKLTQAETTRLN 172
             :: ::::..   . .: :. . ::.. :.  .:..  .  .: . :..   . ::. :.
A194   113  AWVKDTEAEKLFRIREAEETKK-SLMQVASEHIAPLQDAA-DLEI-ATK---E-ETSLLE 165

P293   173  A-----VLDYIDAVTATDTSTAPDVIWPELP                              198
             :     ::  .. :   :::::::. :: .:
A194   166  AWKKYRVL--LNRV---DTSTAPDIEWPAVP                              191
```

E

```
P145    22  VRDANFQ-KQMLISD-ATDFINSRQWQGKAALGRL--KEDE-LKQ-YNLW-LDYLEALEL  74
            .:.:. . :  :. . :.. : . : .. : :  .  ::.  : . .  . .  :  :.
A194   126  IREAE-ETKKSLM-QVASEHIAPLQ-DA-ADLE-IATKEETSLLEAWKKYRV--L--LNR 176

P145    75  VDTSSAPDIEWPTL                                               88
            ::::.::::::::..
A194   177  VDTSTAPDIEWPAV                                               190
```

Fig. II-3. (continued)

F

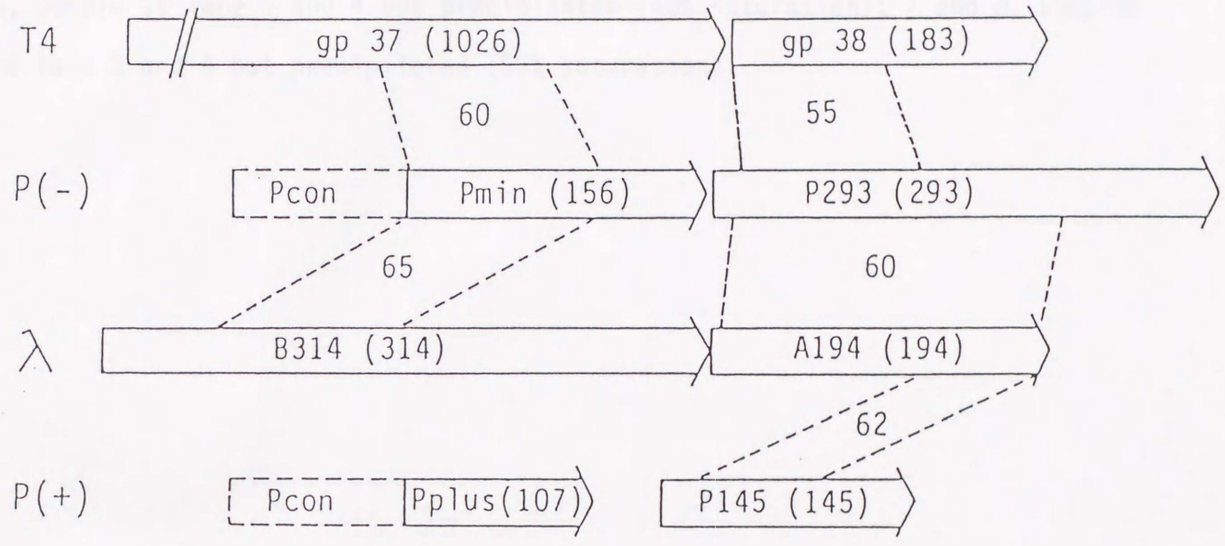

Fig. II-3. (continued)

Fig. II-4. Urea-SDS-polyacrylamide gel electrophoresis of samples from the purification steps of polypeptides. The polypeptides were separated on a 12.5% acrylamide gel. Positions and size (kDa) of molecular weight markers in the same gel are given to the right of the panel. Lanes: 1, 3, 5 and 7, samples prepared from strain EJ2371 containing pPS140; 2, 4, 6 and 8, samples prepared from EJ2371 containing pPS240; 1 and 2, crude extracts; 3 and 4, soluble fraction of samples precipitated with ammonium sulfate (30% saturation; 5 and 6, sample as lane 3 and 4 but precipitated (40% saturation); 7 and 8, samples as lane 3 and 4 but precipitated (50% saturation).

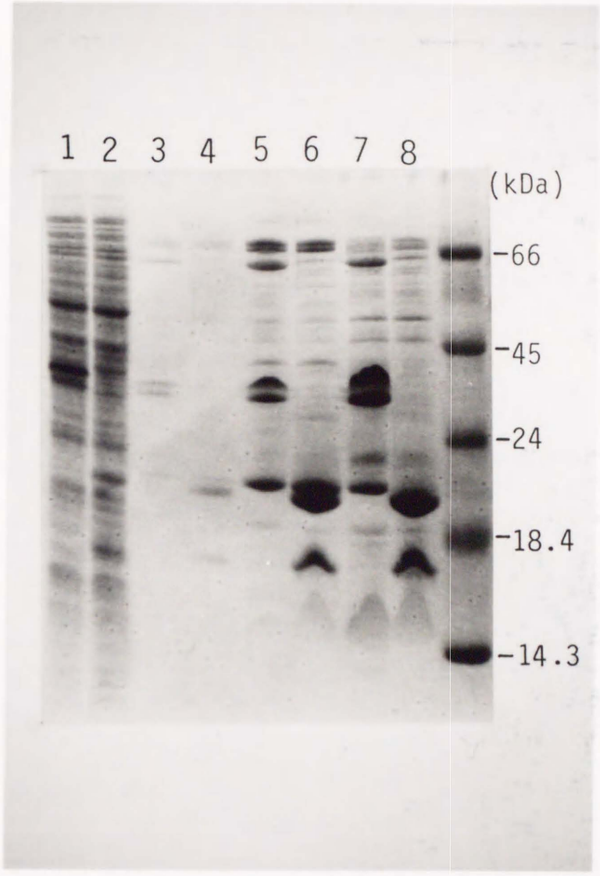

Fig. II-4.

Fig. II-5. Urea-SDS-polyacrylamide gel electrophoresis and Western blotting with antisera. All of the purified samples were prepared from EJ2371 containing the plasmid. Samples (5μg) were separated in a 12.5% acrylamide gel. Lanes: 1 and 2, SDS-polyacrylamide gels stained with Coomassie brilliant blue; 3 to 6, nitrocellulose blot probed with anti-P(+) serum (lane 3 and 4) and anti-P(-) serum (lane 5 and 6); 1, 3 and 5, purified samples prepared from EJ2371 containing pPS140; 2, 4 and 6, purified samples prepared from EJ2371 containing pPS240. Numbers represent kDa of the products produced from the plasmids.

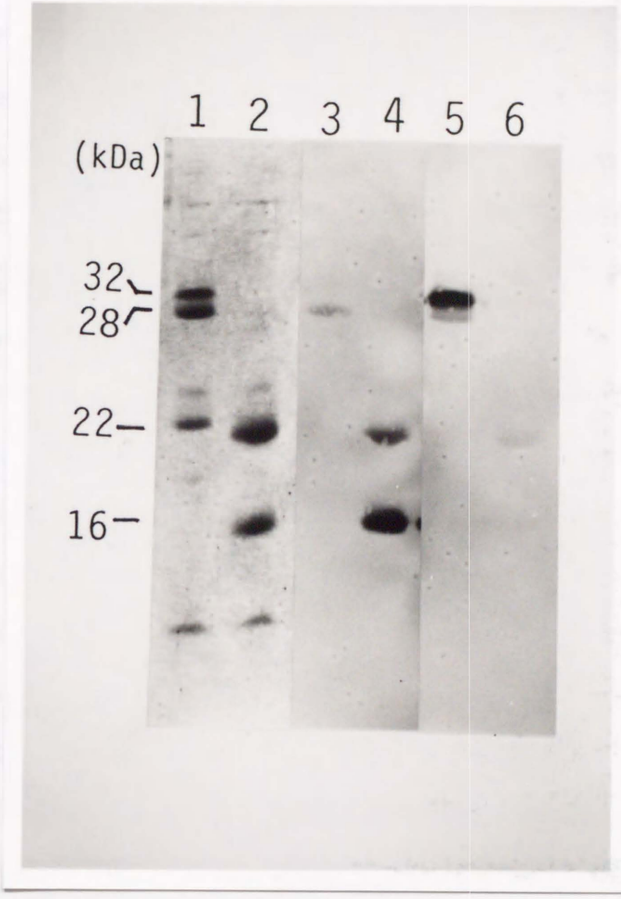

Fig. II-5.

III. Cloning and characterization of a cryptic invertible B segment from *Shigella boydii*

ABSTRACT

Inversional switching systems in procaryotes are composed of an invertible DNA segment and a site-specific recombinase gene adjacent to or contained in the segment. Four related but functionally distinct systems have previously been characterized in detail: the *Salmonella typhimurium* H segment-*hin* gene, H-*hin*, phage Mu G-*gin*, phage P1 C-*cin*, and *Escherichia coli* e14 P-*pin*. In this article we report the isolation and characterization of three new recombinase genes: *pinB*, *pinD*, and defective *pinF* from *Shigella boydii*, *Sh. dysenteriae* and *Sh. flexneri*, respectively. The genes *pinB* and *pinD* were detected by the complementation of a *hin* mutation of *Salmonella* and were able to mediate inversion of the H, P, and C segments. *pinB* mediated H inversion as efficiently as the *hin* gene did and mediated C inversion with a frequency three orders of magnitude lower than that of the *cin* gene. *pinD* mediated inversion of H and P segments with frequencies ten times as high as those for the genes intrinsic to each segment and mediated C inversion with a frequency ten times lower than that for *cin*. Therefore, the *pinB* and *pinD* genes were inferred to be different from each other. The invertible B segment-*pinB* gene cloned from *Sh. boydii* is highly homologous to the G-*gin* in size, organization, and nucleotide sequence of open reading frames, but the 5' constant region outside the segment is quite different in size and predicted amino acid sequence. Two polypeptides, 20 and 22 kDa, which are predicted as the products of internal ORFs in the B segment, were detected and reacted with anti-P1 sera. The B segment underwent inversion in the presence of *hin*, *pin*, or *cin*.

INTRODUCTION

Inversional switching systems in procaryotes, which mediate alternative expression of two sets of genes, are composed of an invertible DNA segment and a site-specific recombinase gene adjacent to or contained in the segment. To date, four systems have been characterized in detail: the *Salmonella typhimurium* H segment-*hin* gene (H-*hin*) (Silverman et al. 1979, Zieg and Simon 1980, Szekely and Simon 1981), phage Mu G-*gin* (Koch et al. 1987), phage P1 C-*cin* (Iida et al. 1982, Hiestand-Nauer and Iida 1983, Iida 1984) and *Escherichia coli* e14 P-*pin* (Plasterk et al. 1983, Plasterk and van de Putte 1985) systems. In *S. typhimurium*, H inversion switches expression of phase 2 flagellin and repressor genes, and in both phages inversion of the segments causes expression of another set of tail fiber genes. Any phenotypic change caused by P inversion is unknown. These systems are different not only in function but in structure. The sizes of the invertible segments are 1, 1.8, 3, and 4.2 kb in the H, P, G, and C segments, respectively. The positions and orientations of recombinase genes are also different from one another with respect to the invertible segment, although the G-*gin* and P-*pin* systems are similar in organization (Glasgow et al. 1989). Despite these differences, nucleotide sequences of *inv* sites where recombination takes place share high homology among the systems (Hatfull and Grindley 1988), and the recombinase genes are substitutable among them (Kutsukake and Iino 1980, Kamp and Kahmann 1981, van de Putte et al. 1984, Momota and Enomoto 1986). These findings together with the relationship of the resolution system in transposon Tn*3* to the switching system (Simon et al. 1980), have led to the idea that these switching systems have evolved from some composite structures in which recombinase genes have a common origin and in which the genes of the associated set(s) have an origin distinct from one another (Kamp and Kahmann 1981, Kamp et al. 1984, Sherratt 1989). One of the approaches to developing the idea would be to detect and analyze more primitive switching systems which

may have imperfect invertible segments or may consist only of a recombinase gene with the inv sites at its ends. Assuming that site-specific recombinases are usually substitutable for one another, it seems possible to detect a new recombinase gene from other organisms when inversion of a known segment can be used as a selection method. The present report deals with the detection and characterization of site-specific recombinase genes from strains of the genus Shigella. Genes from Sh. boydii and Sh. dysenteriae mediate inversion of the known invertible segments with different efficiencies. The recombinase gene and its adjacent invertible segment were cloned from Sh. boydii and sequenced with the result that they highly resemble the G segment-gin system of phage Mu in structure and nucleotide sequence. In Western blot, two polypeptides derived from the B segment reacted with anti-P1 sera.

MATERIALS AND METHODS

Bacterial strains, phages and tester plasmids.

Four strains representing each of the four *Shigella* subgroups (Rowe and Gross 1981) were used: *Sh. dysenteriae* Sh16 (Scott 1968), *Sh. boydii* C12, *Sh. flexneri* IID642 and *Sh. sonnei* IID969 (Tominaga and Enomoto 1986). All strains except Sh16 were kindly supplied by the Medical School, Okayama University. *S. typhimurium* SL4273 (Enomoto and Stocker 1975) was used as a *hin*$^+$ control. Strains EJ1449 (Momota and Enomoto 1986) and SL4266 (Enomoto and Stocker 1975) were used as DNA donors for cloning a *hin* gene and a fragment containing *fljB* (phase 2 flagellin) and mutant *hin* genes, respectively. The main *E. coli* K-12 strains used were as follows: W3110 and EJ2518, a *pin*$^+$ derivative of MC1061 [*hsdR* Δ(*lacI-A*)*X74*] (Casadaban et al. 1983), were used as *pin*$^+$ strains. K802 (*hsdR*) (Wood 1966) was used as an indicator for P1C(+) and also as a parent of the Δ*pin* Δ*fliC* (flagellin gene) strains. EJ2282 (Δ*fliC purB51* Δ*pin recA56 hsdR*), a host for recombinant plasmids, was constructed from K802 by the successive transduction of *fliC*::Tn*10* (Enomoto et al. 1985), *purB51*-Δ*pin*-*zcg2*::Tn*10* (Enomoto et al. 1983) and *recA56*-*srlA*::Tn*10* (Csonka and Clark 1980), with Tn*10* excision (Maloy and Nunn 1981) after each transduction. Similarly, EJ2517 (Δ*fliC purB51* Δ*pin zcg2*::Tn*10 hsdR*) was made from K802 by the transduction of *fliC*::Tn*10* followed by Tn*10* excision and of *purB51*-Δ*pin*-*zcg2*::Tn*10*.

P1*vir* and P1*cin*C(+) were used for transduction. A P1*cin* mutant (Tominaga et al. 1986) from thermosensitive P1C*mclr100* (Rosner 1972) has a C segment fixed in either the C(+) or C(-) orientation. P1*cin*C(-) was used for the measurement of C inversion. To isolate a P1*cin*C(-) lysogen from *Sh. dysenteriae* that was resistant to P1C(-) (Tominaga and Enomoto 1986), first a P1*cin*C(+) lysogen was made and then a C(-) variant was chosen by testing a number of colonies of P1C(+) lysogens for their inability to lyse indicator bacteria (K802) at 42°C. Mu*cts* (Tominaga et al. 1986) was used as probe for hybridization.

Plasmid pSI710, derived from pACYC184, carries a 3.3 kb HindIII fragment which contains a pin gene from an e14 variant (EJ1076) (Enomoto et al. 1983) in which a new HindIII site had been created in the P segment (unpublished data). pSI730 is a pACYC177 (Chang and Cohen 1978) derivative which carries a 3 kb BamHI-HincII fragment (Tominaga et al. 1986) that contains the cin gene of phage P1. pSI743, a pACYC177 derivative, carries a 4.5 kb PstI fragment (Szekely and Simon 1981) which contains the hin gene from strain EJ1449.

Chemicals and antiserum.

Chloramphenicol (Sigma Chemical Co.), kanamycin (Meiji Seika Co.), tetracycline (Sigma) and ampicillin (Sigma) were used at concentrations of 12.5, 50, 25 and 50 mg/l, respectively. X-gal plates contained 40 mg/l of 5-bromo-4-chloro-3-indolyl-β-D-galactoside (Sigma). Anti-flagellum serum against an e,n,x antigen was as previously described (Momota and Enomoto 1986).

DNA manipulations.

Total host DNA was prepared by the method described previously (Birnboim and Doly 1979), except a solution of 1% sodium dodecyl sulfate (SDS) was used instead of the alkaline SDS solution. Isolation of plasmid DNA and cloning procedures were described previously (Maniatis et al. 1982). Restriction endonucleases, T4 DNA ligase and nuclease Bal 31, were purchased from Takara Shuzo Co. and used as recommended. Restriction fragments were examined by agarose gel electrophoresis with 1% agarose (Takara) in 40 mM Tris-acetate (pH 8.0)-1 mM EDTA buffer as described previously (Maniatis et al. 1982). Nucleotide sequences were determined by the dideoxy chain termination method of Sanger et al. (Sanger et al. 1977) using a sequencing kit (Takara) and [^{35}S]dCTP (Amersham International, 400Ci/mmol). Fragments subcloned from plasmid pTSB916 into pUC118 and pUC119 (Vieira and Messing 1987) were deleted stepwise in opposite direction using exonuclease III and mung-bean nuclease

(Takara), and series of deletion plasmids were used for sequencing. Nucleotide and amino acid sequences were analyzed by using DNASIS sequence analysis software (Hitachi Software Engineering Co.). A restriction fragment, which was used as a probe and had been electroeluted from an agarose gel, and whole Mu DNA, prepared as described previously (Tominaga et al. 1986), were labeled by the random primed method (Feinberg and Vogelstein 1983) with digoxigenin-labeled dUTP and by using the nonradioactive DNA labeling kit (Boehringer Mannheim). For hybridization, (Southern 1975) total bacterial DNA which had been electrophoresed after digestion with several restriction enzymes was transferred to a nylon membrane filter (Amersham Hybond-N) by vacuum blotting (Olszewska and Jones 1988) and was fixed by UV-irradiation. Hybridization was done by the standard method (Maniatis et al. 1982) and detection of hybrid bands was performed by the Dig-ELISA method using the Detection kit (Boehringer Mannheim).

RESULTS

1. Construction of the assay plasmids for recombinase gene activity.

Plasmid pTY109, used for the measurement of H segment inversion, was constructed with a 6.75 kb SalI fragment (*hin* *fljB*^{off}) from strain SL4266 (*fljB*^{on}). First a plasmid carrying a *fljB*^{on} fragment was isolated and transformed into strain EJ2282 with a *cin*$^{+}$ plasmid (pSI730). Then, the *fljB*^{off} plasmid made by H inversion due to the *cin* gene was isolated after transformation into strain EJ2282 (Δ*fliC*). The 6.75 kb SalI fragment was subcloned into mini F plasmid pTN1105 (Nohno et al. 1986) and designated pTY501.

Plasmid pPZ202 (*lacZ*^{off}), used for the measurement of P inversion, was constructed as follows. A 2.8 kb BamHI-BglII fragment (Plasterk and van de Putte 1985) containing the P segment of e14 was subcloned from a P-*pin* plasmid (pHA201 Fig. III-1B) into the BamHI site of pACYC184 (Chang and Cohen 1978). Then, a 3 kb PstI fragment containing the *lacZ* gene lacked p_{lac} was inserted from plasmid pMC1871 (Casadaban et al. 1983) in frame into the PstI site of the major open reading frame (ORF) (Plasterk and van de Putte 1985) of the cloned P fragment. Moreover, the promoter of a tetracycline resistance (Tet^{r}) gene on the vector, which can be used for *lacZ* expression when the P segment undergoes inversion, was deleted by Bal 31 digestion at a HindIII site in the promoter sequence to yield plasmid pPZ102 (*lacZ*^{on}). The P segment of pPZ102 was inverted by using another *cin*$^{+}$ plasmid (pAT101) (Tominaga et al. 1986) derived from pBR322 to yield pPZ202, which was determined to have the P segment fixed in the (+) orientation. The β-galactosidase activity expressed by pPZ202 and pPZ102 in strain MC1061 was approximately 20 and 350 Miller units (Miller 1972), respectively. The two plasmids differed in EcoRI-digestion pattern because of the presence of one EcoRI site in the chloramphenicol acetyltransferase gene on the vector (Chang and Cohen 1978) and two EcoRI sites in the *lacZ* gene (Casadaban et al. 1983).

2. Detection of recombinase gene activity from Shigella spp..

To search Shigella spp. for the presence of new switching systems, transduction by P1cinC(+) was carried out from four Shigella strains representing different subgroups into nonmotile E. coli K-12 strain EJ2282 (ΔfliC Δpin recA) with plasmid pTY109 (hin fljB^{off}). The presence of recombinase genes capable of complementing hin mutation was inferred from restoration of fljB expression (Table III-1). Sh. dysenteriae, Sh. boydii and E. coli W3110, used as a pin^{+} control, produced swarms (motile colonies) on the plain NSS medium, but not on the medium containing anti-e,n,x serum, indicating that silent fljB (antigen e,n,x) that is expressed on pTY109 by H inversion from the off to the on orientation and its inversion is mediated by a hin^{+} equivalent abortively transduced from Sh. dysenteriae or Sh. boydii. Trail production (Table III-1), which suggests the presence of cryptic flagellin genes in Shigella spp., will be described elsewhere.

To examine the possibility that the hin^{+} equivalents of the two strains are located on e14 (Enomoto et al. 1983) or an e14-like element on the chromosome (as in the case of P-pin), transduction was carried out by using the two strains as donors and strain EJ2517 (ΔfliC), which carries purB Δpin zcg2::Tn10 (in that order) as the recipient (Enomoto et al. 1983). Pur^{+} Tet^{S} transductants are expected to restore the ability for H inversion when a gene of interest is on the element. The donor Sh. dysenteriae yielded a number of Pur^{+} clones but none of them (60 clones) were Tet^{S} and the five clones transformed with pTY109 did not cause H inversion either. Sh. boydii yielded only five Pur^{+} clones by repeated transduction and all of them were Tet^{S} Hin. Transduction from Sh. sonnei and Sh. flexneri was also attempted. Sh. sonnei yielded 27 (45%) Pur^{+} Tet^{S} cotransductants among 60 Pur^{+} transductants, and all were Hin. Sh. flexneri yielded Pur^{+} Tet^{S} cotransductants at a frequency of 48% (39 of 81 Pur^{+} transductants); of the ten cotransductants transformed with

pTY109, five became motile due to H inversion, although a *hin*$^{+}$ equivalent had not been detected in the previous transduction (Table III-1). These results suggest that the *hin*$^{+}$ equivalents of *Sh. dysenteriae* and *Sh. boydii* are not on e14 as in *E. coli*, that *Sh. flexneri* probably has defective *pin* gene at the same locus as *E. coli*, and that its defect can be repaired by recombination or another mechanism after transduction to *E. coli* Δ*pin*. Henceforth, the defective gene of *Sh. flexneri* is designated *pinF*, and that of *E. coli* is designated *pinE*. Putative genes in *Sh. boydii* and *Sh. dysenteriae* are designated *pinB* and *pinD*, respectively.

3. Inversion activity for H, P and C segments in *Sh. boydii* and *Sh. dysenteriae*.

The genes *pinB* and *pinD* may be alike or different from each other in structure and function. To distinguish between *pinB* and *pinD* genes, inversion rates of H, P and C segments were measured in *Sh. boydii* and *Sh. dysenteriae* and compared with the rate of the recombinase gene intrinsic to each segment. To examine the rate of H inversion, the two strains transformed with pTY109 were incubated successively and the plasmid isolated after appropriate generations was transformed into strain EJ2282 (Δ*fliC*) to test whether the plasmid confers motility on the recipient as a result of H inversion. The *fljB*^{on} plasmid was detected in all the samples isolated after 50 generations (Table III-2). The inversion rate of the H segment in *Sh. dysenteriae* was estimated to be 1.4×10^{-3}, which is 10 times higher than in *Sh. boydii* and *S. typhimurium* used as control.

P inversion was examined by a change in color on X-gal plates of the transformants with plasmid pPZ202 (*lacZ*^{off}), in which expression of *lacZ* is controlled by the orientation of the P segment. The rate was measured only for *Sh. dysenteriae* and the *E. coli pin*$^{+}$ strain EJ2518 (Δ*lacZ*) (Table III-3), since the two types of *Sh. boydii* transformants obtained with pPZ202 or pPZ102

(*lacZ*^{on}) were blue on the X-gal plate and indistinguishable from each other. The rate of P inversion in *Sh. dysenteriae* was 1.6 to 3.6 x 10^{-4}, which is ten times higher than the rate in the *pin*$^{+}$ strain EJ2518. Plasmid DNA isolated from each blue colony was digested with *Eco*RI and electrophoresed to visualize inverted DNA fragments; the plasmids from all the blue colonies (total, 5) had digestion patterns identical to that of pPZ102 (data not shown). Although P inversion could not be tested in *Sh. boydii*, the activity of the *pinB* gene for P inversion was indicated by the change in the *Eco*RI digestion pattern of pPZ202, which had been propagated in the *Δpin* strain (EJ2282) with a *pinB* plasmid (pTSB916) (see below).

To examine the rate of C inversion, the strains lysogenized with phage P1*cin*C(-) were cultured in broth at 30°C, and the occurrence of P1C(+) lysogens was measured at 42°C (Table III-4). The mean inversion rates of the C segment in *Sh. boydii* and *Sh. dysenteriae* were 2.6 x 10^{-5} and 4.1 x 10^{-3}, respectively. Since the rate from C(-) to C(+) by the *cin* gene has been reported to be 3.4 x 10^{-2} (Momota and Enomoto 1986), it was determined that *Sh. boydii* and *Sh. dysenteriae* can mediate C inversion with frequencies three orders and one order of magnitude lower than that of the *cin* gene, respectively. A similar experiment was performed with *Sh. sonnei*, since earlier study had suggested that this strain can mediate C inversion (Tominaga and Enomoto 1986). It was found that the supernatant from a broth culture of *Sh. sonnei* contains some kind of colicin factor, which is able to make a very small lysis spot which is indistinguishable from a P1*cin*C(+) plaque on the indicator strain K802. When the bacteria lysogenized with P1*cin*C(-) were spotted on the indicator that had previously been made resistant to this colicin and were tested by shifting up to 42°C, no lysis spot or plaque could be detected, showing that the previous finding may be ascribed to the misjudgment of lysis spots by the colicin of *Sh. sonnei*.

These results showed that *Sh. boydii* mediates H inversion with a frequency

similar to that of the hin gene and mediates C inversion with a frequency 1,000 times lower than that of the cin gene, while Sh. dysenteriae mediates inversion of the H and P segments with frequencies 10 times higher than those of the genes intrinsic to each segment and mediates inversion of the C segment with a frequency 10 times lower than that of the cin gene. Assuming that pinB or pinD is the only recombinase gene in each strain, the pinD gene is clearly different in activity from pinB and also different from hin, pinE and cin genes.

4. Cloning of the pinB gene and its adjacent invertible B segment from Sh. boydii.

Cloning of B-pinB system.

BamHI fragments from chromosomal DNA of Sh. boydii were inserted into pBR322 and transformed into nonmotile strain EJ2286 containing the mini F plasmid pTY501 (fljB^{off}). Recombinant plasmid pTSB916, which had conferred motility on the recipient, was found to carry a 5.9 kb BamHI fragment (Fig. III-1A). When the plasmid was digested with BamHI and MluI, six bands (4.3, 3.8, 3.3, 1.5, 1.1 and 1.0 kb) appeared, the total length of which was 10.7 kb, exclusive of a 4.3 kb vector fragment; this is longer than the cloned fragment by 4.8 kb. We inferred that the extra 4.8 kb fragment (3.8 + 1.0 or 3.3 + 1.5) resulted from a putative invertible segment, which was cloned together with the pinB gene and could have alternative orientation. It is reasonable to think that a MluI site is located asymmetrically in the segment since the BamHI sites used are the cloning site. When a 2 kb SalI fragment or a 3.1 kb AvaI fragment (Fig. III-1A) was deleted, the plasmid produced the 1.0 or 3.3 kb fragment by digestion with BamHI and MluI in addition to the large fragment containing the vector. This indicates that the pinB gene, or part of it, is located on these fragments and that its defect results in fixed orientation of the invertible segment. Therefore, the pinB gene was mapped on

the right side of the invertible segment that contained the MluI site (Fig. III-1A). The invertible segment, designated the B segment, seemed to be larger than 2.3 kb, the distance between the MluI sites at B(+) and B(-) orientations, but smaller than 3.3 kb (2.3 + 2 x 0.5 kb) since the distance between the MluI site at B(+) orientation and the AvaI site close to it was 0.5 kb and the AvaI site was not included in the B segment. In fact, the B segment was found to be 3 kb by sequencing this region as described below. The arrangement of the B segment in pinB resembles those of P-pinE, C-cin, and G-gin, unless a promoter site for each recombinase gene is taken into account. However, the size of the B segment was clearly different from those of the P and C segments and similar to that of the G segment (3 kb); nevertheless, the B and G segments were different in restriction sites; the B segment has one EcoRV, while the G segment has no such site but instead has one KpnI and one HincII site (Kahmann and Kamp 1987).

Inversion of the B segment.

The B segment was found to undergo inversion in the presence of the known recombinase genes, pinE, cin and hin. Plasmid pTSB917 was made from pTSB916 by deleting the SalI fragment that contained a part of pinB (Fig. III-1A); this was confirmed later by sequencing pinB. When pTSB917 coexisted with a plasmid carrying one of the three genes, pinE (pSI710), cin (pSI730) and hin (pSI743) in Δpin bacteria (EJ2282) during approximately 30 generations, pTSB917 produced two sets of fragments resulting from two orientations of the B segment after digestion with MluI and SalI (Fig. III-2). This result, together with the finding that the pinB gene can mediate H, P and C inversions, indicates that the B-pinB system operates well and is indeed a member of the family of the inversional switching systems represented by H-hin.

5. Nucleotide sequences of the pinB gene and the B segment.

The nucleotide sequence of the 4.8 kb BamHI-MluI fragment at the B(+) orientation (Fig. III-1A) was determined with both strands (Fig. III-3A). Three complete and two incomplete open reading frames (ORFs) were revealed (Fig. III-3B). The location of pinB was determined on the 3' side of the fragment from a predicted amino acid sequence which is highly homologous to those reported for several site-specific recombinases (Zieg and Simon 1980, Hiestand-Nauer and Iida 1983, Plasterk et al. 1983). It starts with GTG instead of ATG as in the gin gene (Plasterk et al. 1983), is preceded by a Shine-Dalgarno sequence and a Pribnow box, consists of 585 bp (position 3796 to 4380), and encodes a 195 amino acid polypeptide with a molecular mass of 21,931 Da. It terminates with TAA followed by inverted repeats (IRs) which were inferred to form a Rho-independent transcriptional terminator, since they are G-C rich and are followed by poly T_5. The ORF can be truncated by SalI or AvaI, which explains why the deletion plasmids produced by these enzymes show the fixed orientation of the invertible segment. The presence of the B segment was inferred by the presence of two 38 bp IRs with only 3 bp differing between them. They contain a 26 bp inv consensus sequence (Hatfull and Grindley 1988), but 3 bp in the right IR sequence (IRR) and 4 bp in the left IR sequence (IRL) differ from the consensus. The occurrence of B inversion by recombination between the two IRs was verified by sequencing about 50 bp of the 3' flanking region of IRL and 50 bp of the 5' flanking region of IRR in the B(-) orientation. The B segment consists of 3,033 bp exclusive of the two IRs and contains the two complete internal ORFs, B175 and B177, which are transcribable in opposite directions: B175, which was transcribed toward pinB in the B(+) orientation, starts with ATG, consists of 525 bp (position 1679 to 2203, 175 amino acids), and terminates with TAA, and B177, which starts with ATG, consists of 531 bp (2770 to 2238, 177 amino acids), and terminates with TGA. These ORFs are followed by a common potential stem-loop structure, which would function as a transcriptional terminator for both ORFs. The other two

ORFs are incomplete, one is on the 5' side of the cloned fragment and contains a 1,674 bp sequence which encodes 558 amino acids. The IRL within the sequence can separate it into a constant region (Bc; 1 to 654) and a variable region (Bv; 655 to 1674), provided that the crossover site is located within the central dinucleotide of the inv site (Plasterk and van de Putte 1984, Johnson and Simon 1985, Iida and Hiestand-Nauer 1986). Another variable region (Bv'; 3727 to 2774), which may take the place of Bv when the B segment undergoes inversion, starts within the IRR present upstream of pinB, but no start codon or SD sequence was detected. The two variable regions, Bv and Bv', encode 340 and 318 amino acids, respectively. Neither promoter sequence nor a start codon was observed in the constant region, and thus they may be present beyond the extreme left end of the fragment.

6. Southern blot analysis.

To confirm that the B segment and the pinB gene are derived from genomic DNA of Sh. boydii and to determine whether the sequence of the B segment is homologous to that of the G segment of Mu, a hybridization assay was carried out. Sh. boydii total DNA and plasmid pTSB916 DNA were digested with BamHI and SalI and were hybridized with the 5.9 kb BamHI fragment containing either the B segment and the pinB gene in pTSB916 or the whole Mu genome. Three inherent plasmids of Sh. boydii were not detected by these probes (data not shown). In the case of the B-pinB probe, the BamHI cut showed only one band (5.9 kb) in the genomic digest (data not shown), and the BamHI and SalI cut showed two bands, of 1.7 and 4.2 kb (Fig. III-4, lane 2). These results are consistent with the existence of the B segment and pinB gene in the genome. With the Mu probe, the BamHI and SalI cut showed only one band, of 4.2 kb, which contained the B segment and the 5'-terminal half of the pinB gene in pTSB916 (Fig. III-4, lane 3) and showed one weak band of approximately 23 kb in addition to the 4.2 kb band in the genomic digest (Fig. III-4, lane 4). These results

demonstrate the homology between the B segment and the G segment and suggest that a fragment of at least 1.7 kb which lies downstream of the B segment is nonhomologous to the Mu genome and that a DNA structure other than B-*pinB* is present in the genome of *Sh. boydii* as the structure homologous to Mu DNA.

7. Western immunoblot analysis.

Construction of plasmids that direct production of polypeptides encoded on the B segment.

To detect the gene products encoded in the B segment, two recombinant plasmids carrying a fixed B segment were constructed. 5.9 kb *BamHI* fragment of plasmid pTSB916 was subcloned into the *BamHI* site of vector plasmid pUC118. One of the recombinant plasmids, isolated was named pPSB100. The direction of the insert fragment of pPSB100 was determined by the *SalI* cut, which showed two bands of 1.7 and 7.4 kb. To isolate the plasmid deleting most *pinB* gene, 7.4 kb *SalI* fragment was ligated and transformed into EJ2282 (Δ*pin recA*) and Amp^r transformants were isolated. Plasmid DNA was isolated from Amp^r transformants and analyzed by *MluI* digestion followed by agarose gel electrophoresis. As a result of analysis two types of plasmids, one containing the B segment frozen in the (-) orientation, and the other containing the B segment frozen in the (+) orientation, were isolated and named pPSB110 and pPSB120, respectively.

Detection of the plasmid-coded polypeptides.

Both plasmids, pPSB110 and pPSB120, have a *lac* promoter in front of the ORFs encoded on the B segment frozen in one orientation. To avoid the repression by *lacI* gene, plasmid pPSB110 and pPSB120 were transformed into EJ2371 (Δ*lac recA*). The polypeptide pattern of strain containing the plasmid pPSB110 or pPSB120 was analyzed by the SDS-polyacrylamide gel electrophoresis.

The crude extracts were separated on a 12% polyacrylamide gel (Fig. III-5, lane 1 and 2). Two polypeptides were overproduced; 20 kDa is specific to pPSB120 and 22 kDa is specific to pPSB110. Since two internal ORFs, B175 (a molecular weight of 20,057) and B177 (20,587), are predicted in the B(+) and B(-) orientation, respectively, 20 kDa and 22 kDa polypeptides were inferred to correspond to the products of B175 and B177, respectively. The polypeptides corresponding to two ORFs, Bcv and Bcv', were not detected in this experiment. Since the Bc region common to these ORFs do not contain an N-terminal region and its promoter, these truncated polypeptides may be unstable or the expression of these polypeptides may be inhibited at translational level.

Western blotting.

It was shown that the ORFs on the B segment is highly homologous to those of G segment in their organization and amino acid sequences and the amino acid sequence from Bv is fairly homologous to that of the C-terminal region of the gene 19 on the C segment of phage P1 (see discussion). To examine the immunological relation between the internal two ORFs of the B segment and those of the C segment, Western blot analysis of these two ORFs, of B175 and B177, with anti-P1 sera were carried out. The result is shown Fig. III-5. Anti-P1C(+) serum reacted slightly with 20 kDa polypeptide, which is specific to the B(+) orientation, and anti-P1(-) serum reacted strongly with 22 kDa polypeptide, which is specific to the B(-) orientation. High molecular polypeptide bands reacted with anti-P1 sera are nonspecific bands, since these polypeptides are not coded in the plasmids used.

DISCUSSION

It was shown that the site-specific recombinase genes pinB and pinD are present in Sh. boydii and Sh. dysenteriae, respectively. No recombinase activity was detected in Sh. flexneri and Sh. sonnei, but in the former the presence of the defective pinF gene was suggested. The pinB gene mediated H inversion as efficiently as the hin gene intrinsic to the H segment did, while pinD showed a frequency for H and P inversion that was ten times higher than that for the recombinase genes intrinsic to each segment. The frequency of C inversion for pinB and pinD was three orders and one order of magnitude lower than that for the cin gene, respectively. The difference in activity among these recombinase genes suggests that pinB and pinD are different kinds of genes and that pinD is also different from any of the known recombinase genes, such as hin, pin and cin. Phenotypes of the pinB and pinD genes are unknown, as is the case for in the pin gene on E. coli e14. The transduction of these two genes into the mutant pin locus of E. coli was negative, suggesting that these genes are not present on e14, which would be homologous to that of E. coli, and that even if they were on e14 or on a similar element, their loci in both strains would differ from those of E. coli.

The 4.8 kb fragment containing the pinB gene and the invertible B segment was cloned from Sh. boydii and sequenced. The pinB gene, which consists of 585 bp (195 amino acids) has its putative promoter (-10 sequence) within the IRR, although a -35 sequence could not be detected and is transcribable toward the outside of the IRR. The B segment, flanked by the two IRs, contains the two complete ORFs, B175 and B177, and the two variable regions, Bv and Bv', which may be transcribed with the 5'-constant region (Bc) outside the B segment in the B(+) and B(-) orientations, respectively, and may result in the ORFs Bcv and Bcv', whose 5' coding region is not contained in the cloned fragment. Such a transcription system has been previously reported for the G, C and P segments (Iida 1984, Plasterk and van de Putte 1985, Koch et al. 1987). The

organization of B-*pinB* is similar to that of G-*gin* in the size of the invertible segment and in the transcriptional direction of the recombinase gene.

When the predicted amino acid sequence of PinB was compared with those of the four recombinases reported previously by several groups (Hatfull and Grindley 1988) (Fig. III-6), the highest homology was observed for Gin (193 amino acids): 78.9% (154 of 195) of the amino acids are identical in all 5 sequences. The homology of PinB to PinE (184 amino acids), Cin (186 amino acids) and Hin (190 amino acids) is 72.8% (142 amino acids identical), 68.7% (134 amino acids identical) and 66.7% (130 amino acids identical), respectively. All these values are more than 80% when equivalent amino acids are also taken into account. The number of amino acids that are conserved as a common sequence in the five recombinases including PinB is 104, corresponding to 53 to 56% of each sequence. The amino acids that are conserved in the four recombinases but not in PinB are at positions 160 (Pro→Asp) and 175 (Leu→Ile) in the C-terminal region. This region has been reported to be important for target recognition (Plasterk and van de Putte 1984) and partial DNA binding (Bruist et al. 1987), and hence these amino acid changes in PinB may cause the rather low frequency of segment inversion in *Sh. boydii*. The enhancer *sis*, which is required for efficient inversion of the invertible segments (Huber et al. 1985, Johnson and Simon 1985, Kahmann et al. 1985), is present in all the recombinase genes reported to date (Hubner and Arber 1989). The *sis* site was also found in the 5'-coding region of *pinB* (Fig. 3) and its 90 bp sequence is completely identical to that of *gin* (Kahmann et al. 1985). When a 50 bp sequence located between the IRR and the *pinB* gene was compared with the corresponding regions of the four recombinases, 80% (40 of 50 nucleotides) and 52% (26 of 50) homology with those regions of *gin* and *pinE*, respectively, were observed, indicating that the three genes are closely related with respect not only to the coding regions but also to their upstream

regions.

The ORFs on the B segment were also found to be highly homologous to those of the G segment in their organization and amino acid sequences. B175 and B177 on the B segment correspond to the two complete genes U (175 amino acids) and U' (177 amino acids) on the G segment (Kahmann and Kamp 1987), respectively. Likewise, Bv and Bv' on the B segment correspond to the two variable regions Sv (327 amino acids) and S'v (311 amino acids), respectively, on the G segment; Sv and S'v and the constant 5' region (Sc) compose the genes S and S' (Kahmann and Kamp 1987). The homology in predicted amino acid sequences of the two corresponding genes or ORFs between the B and G segments is as follows: B175 vs U, 90% (158 identical amino acids of 175); B177 vs U', 98% (173 of 177); Bv vs Sv, 62% (211 of 340) and Bv' vs S'v, 96% (306 of 318). On the other hand, the homology between the two constant regions outside the segments was less than 30% when the perfect 177 amino acid sequence of Sc was compared with several 177 amino acid sequences arbitrarily chosen from Bc (218 amino acids). The amino acid sequence from Bv was found to have fairly high homology with that of the C-terminal region of the gene 19 (328 amino acids) (Kamp et al. 1984) on the C segment of phage P1: 61% (208 of 340) of the amino acids are identical. Each of the four ORFs on the B segment was compared with four ORFs on the P segment of the e14 element (Plasterk and van de Putte 1985). The maximum-amino acid homology between the ORFs of the B and P segments, which were chosen in every possible combination, is 25 to 30%, indicating that there is no homology between the B and P segments at the amino acid sequence level.

The IRL and IRR of the B segment are 38 bp sequences (in which 3 bp are different) and contain the 26 bp inv consensus sequence (Hatfull and Grindley 1988), which is separated by a 5'-AA dinucleotide into two imperfect 12 bp IRs (Hatfull and Grindley 1988). When this sequence of the B segment was compared with those of other invertible segments, 23 (88.5%) of 26 bp for IRL and 24 (92.3%) of 26 bp for IRR were found to be identical to those of the G segment.

Moreover, instead of the AA dinucleotide where recombination occurs (Plasterk and van de Putte 1984, Johnson and Simon 1985, Iida and Hiestand-Nauer 1986), a 5'-GA was observed in both IRs of the B segment, as reported for inv of the G segment (Plasterk et al. 1983). The homology with inv sequences from the H, C and P segments is 53.9 to 80.8% for IRL and 53.9 to 84.6% for IRR (Hatfull and Grindley 1988).

Taken together, these homology tests show that B-pinB of Sh. boydii is similar to G-gin of phage Mu in every way but that the 5'-constant region outside the B segment is quite different in size and predicted amino acid sequence. Phage has not been detected in Sh. boydii, though its culture fluid was tested on the indicator strains E. coli C and K802 (hsdR) before and after its induction by mitomycin C or UV-irradiation. Mu and its related phage D108 are known to be noninducible and inducible, respectively (Dubow 1987). Thus, if an intact prophage like Mu or D108 were present in Sh. boydii, it would be detected by such tests. Although the hybridization assay suggested that the sequence homologous to the Mu genome is present in Sh. boydii, it remains unknown whether the B-pinB structure is part of a defective prophage into which a Mu-like phage has degenerated or whether it is a prototype structure to be integrated into a preprophage to result in an intact phage such as Mu. Two polypeptides of 20 and 22 kDa were inferred to be the product of two ORFs, B175 and B177, respectively, and these polypeptides reacted with anti-P1 sera. This suggests that the amino acid sequences of B175 and B177 is homologous to those of tfs and tfs' (tail fiber specificity) gene products, respectively, which are presumed in the C segment (Kamp et al. 1984). Assuming that B-pinB is part of some prophage, the products of the ORFs on the B segment and its flanking region would be tail fiber proteins concerned with host recognition similar to those on the G segment of Mu and on the C segment of P1.

Table III-1. Occurrence of motile clones from a nonmotile *E. coli* strain with tester plasmid pTY109 in transduction from *Shigella* spp.

Donor	NSS medium containing:			
	no antiserum		anti-*e,n,x* serum	
	swarm	trail	swarm	trail
Sh. dysenteriae	+	-	-	-
Sh. boydii	+	+	-	+
Sh. flexneri	-	+	-	+
Sh. sonnei	-	+	-	+
E. coli W3110	+	+	-	+

P1*cin*C(+) grown on each donor was mixed with 1 ml of broth culture of strain EJ2282 (Δ*fliC* Δ*pin recA*) with pTY109 (Ap^r *hin* $fljB^{off}$) at a multiplicity of infection of around 0.5 and was then incubated at 37°C for 20 min for adsorption. After centrifugation, about half of the mixture was streaked as lines on each of the two NSS plates either containing ampicillin or containing antiflagellum serum and ampicillin and then was incubated at 30°C for 15 to 20 h. +, Detection of swarms or trails; -, No detection. The number of swarms detected was 1 or 2 per line and that of trails was less than 10.

Table III-2. Inversion rates of the plasmid-encoded H segment in *Sh. dysenteriae* and *Sh. boydii*

Strain	No. of swarms/ No. of Amp^r transformants tested			Inversion rate
	(30)	(50)	(100)	
Sh. dysenteriae	16/376	28/388	NT	1.4×10^{-3}
Sh. boydii	0/264	2/400	8/644	1.0×10^{-4}
S. typhimurium SL4273	1/528	3/528	NT	1.1×10^{-4}

A single colony from each strain with pTY109 was cultured in broth for more than 100 generations by successive inoculation, and the plasmid isolated after 30, 50, or 100 generations (rough estimate) was transformed into EJ2282 (Δ*fliC*). The Amp^r transformants obtained were tested for motility by stabbing them into NSS medium containing ampicillin. The inversion rate was expressed as a value (no. of swarms per plasmid per generation) from H(-) to H(+) orientation during 50 generations. NT, Not tested.

Table III-3. Inversion rates of the plasmid-encoded P segment in *Sh. dysenteriae*

Strain	No. of blue colonies/ No. of colonies examined	No. of generations	Inversion rate
Sh. dysenteriae	1/199	30.9	1.6×10^{-4}
	2/182	30.8	3.6×10^{-4}
	1/116	30.1	2.9×10^{-4}
E. coli EJ2518	1/1,090	30.0	3.1×10^{-5}

A single colony from each strain transformed with pPZ202 (Cm^r *lacZ*^{off}) was cultured in broth at 37°C for about 30 generations. The culture was diluted and plated on X-gal plates containing chloramphenicol, and the number of blue colonies was counted after overnight incubation at 37°C. The rate was expressed as a value (no. of blue colonies per cell per generation) in inversion from the P(+) to the P(-) orientation. This will be reduced to at least 1/10 when expressed as a value (no. of blue colonies per plasmid per generation), assuming that 10 to 20 copies of plasmid are maintained in a bacterium and one of them undergoes inversion.

Table III-4. Inversion rates of the C segment in Sh. dysenteriae and Sh. boydii

Strain	IC/ml (10^5)	CFU/ml (10^6)	Ratio IC/CFU	No. of generations	Mean rate
Sh. dysenteriae	290	338	8.5×10^{-2}	28.3	
	290	372	7.8×10^{-2}	28.5	4.1×10^{-3}
	1600	850	1.9×10^{-1}	29.7	
Sh. boydii	127	153	8.7×10^{-4}	30.5	
	163	220	7.4×10^{-4}	31.0	2.6×10^{-5}
	220	287	7.6×10^{-4}	31.4	

A P1cinC(-) lysogen from each strain was cultured in broth at 30°C. After about 30 generations the number of bacteria capable of being infective centers (IC) because of the change to P1cinC(+) was counted by spotting aliquots of the dilutions on the lawn of indicator bacteria (K802) and incubating them overnight at 42°C. The number of viable bacteria (CFU) of the same culture was counted at 30°C. The rate was expressed as a value (IC per CFU per generation) in inversion from C(-) to C(+) orientation.

Fig. III-1. Restriction maps of the fragments containing the invertible segment and recombinase gene. (A) pTSB916 that carries the 5.9 kb BamHI fragment with B-pinB from Sh. boydii. (B) pHA201 that carries the 8.0 kb BamHI fragment with P-pinE from E. coli strain EJ350. Boxed arrow heads show IR sequences. Solid arrows represent recombinase genes. The thick solid line is from pBR322. Restriction cleavage sites; A, AvaI; B, BamHI; Bg, BglII; E, EcoRI; Hc, HincII; Hd, HindIII; Hp, HpaI; M, MluI; P, PstI; S, SalI; V, EcoRV. Restriction sites within the invertible sequences are shown with two orientations.

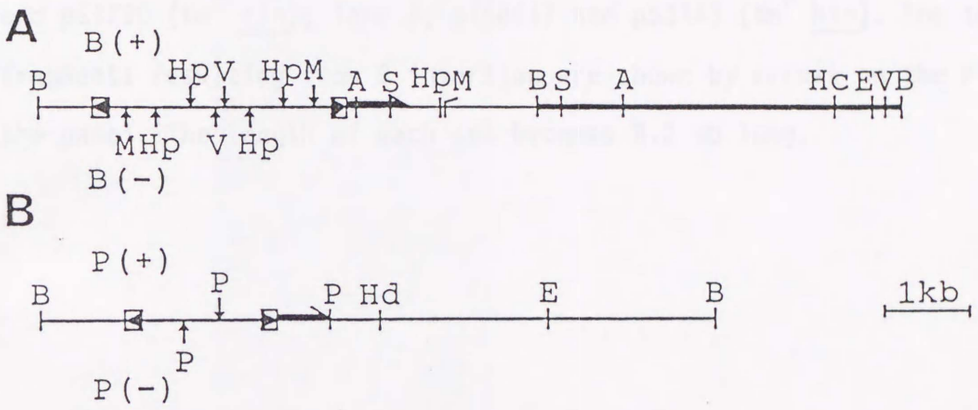

Fig. III-1.

Fig. III-2. Analysis of the invertible B segment by agarose gel electrophoresis. Plasmid pTSB917 carrying the B segment was propagated in a Δ*pin* strain EJ2282 with a plasmid carrying one of the three genes, *pinE*, *cin* and *hin*, for about 30 generations in the presence of an antibiotic corresponding to a resistance marker encoded on each plasmid. The plasmids isolated were digested with *Sal*I and *Mlu*I and electrophoresed. Lane 1, pTSB916 (Ap^r) carrying the B-*pinB* fragment; lane 2, pTSB917 (Ap^r) with the B segment fixed in (+) orientation; lane 3, pTSB918 (Ap^r) with the B segment fixed in (-) orientation; lane 4, pTSB917 and coresident pSI710 (Cm^r *pinE*); lane 5, pTSB917 and pSI730 (Km^r *cin*); lane 6, pTSB917 and pSI743 (Km^r *hin*). The two sets of fragments resulting from B inversion are shown by arrows on the right side of the panel. The length of each set becomes 8.2 kb long.

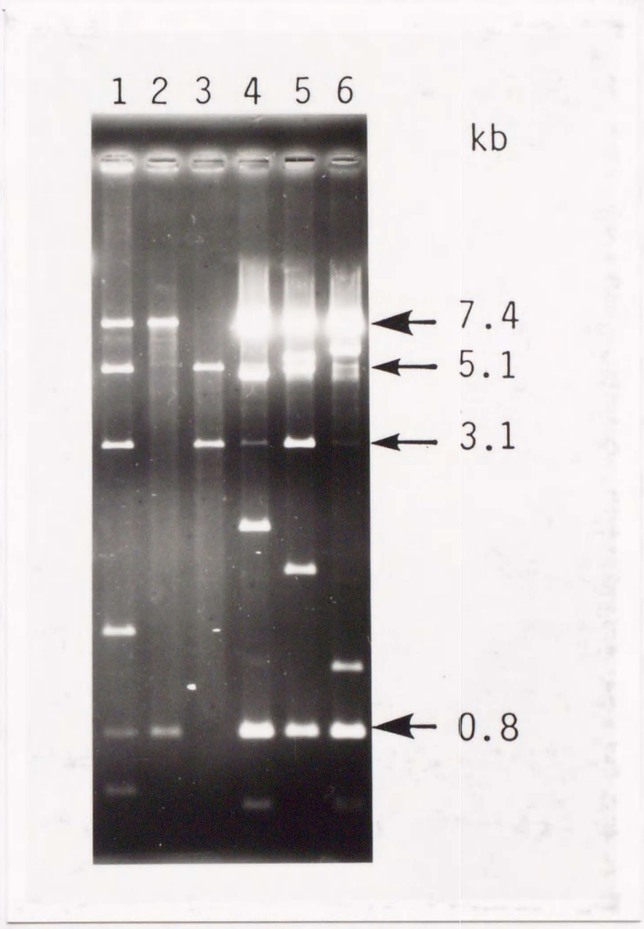

Fig. III-2.

Fig. III-3. Nucleotide sequences and genetic organization of the B segment at (+) orientation and the pinB gene. (A) The sequence of the 4810 bp BamHI-MluI fragment is numbered from the BamHI site. Left and right inverted repeat sequences, IRL and IRR, which flank the B segment, are boxed. The inv sequences are marked by broken lines. The sis site in pinB is indicated by < to >. The deduced amino acid sequences of five potential ORFs are shown: B175, B177 and PinB are complete ORFs, and Bc and Bv' are coupled in (-) orientation to create Bcv'. The translation start and stop sites of the three complete ORFs are indicated by arrows and asterisks, respectively. The Bc start site seems to be located beyond the 5' end of the fragment. Bv and Bv' are considered to start from glycine in IRL and IRR (Glasgow et al. 1989). The Shine-Dalgarno (SD) sequence and the consensus promoter sequence (-10) are underlined. Palindromic sequences which may serve as transcriptional terminators are indicated by converging arrows. The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL and GenBank Nucleotide Sequence Databases under accession number D00660. (B) Diagrammatic representation of ORFs. ORFs are indicated by open arrows and the number of amino acids in each ORF is shown in parentheses.

A

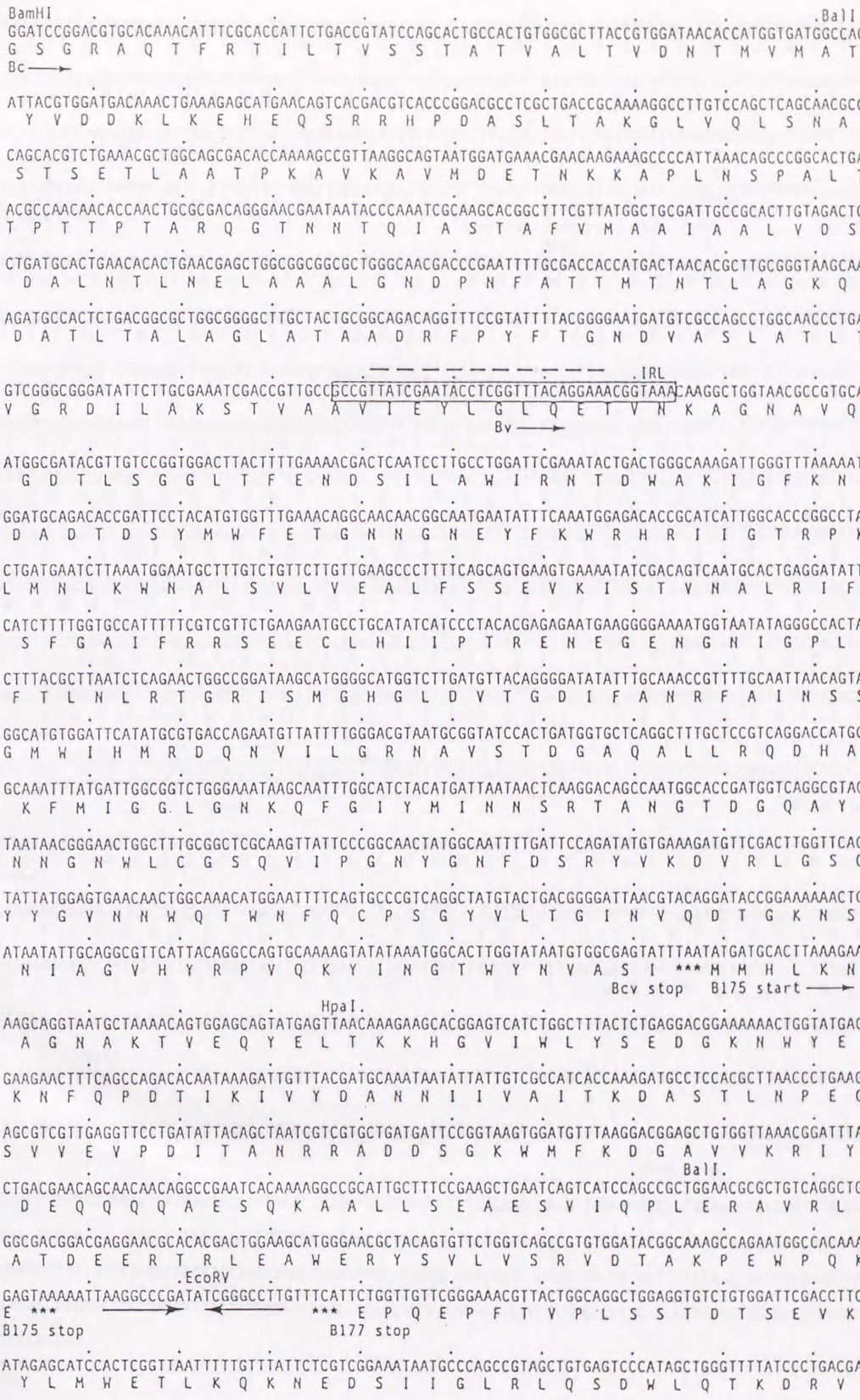
BamHI
.BalI
Bc
.IRL
Bv
HpaI.
Bcv stop
B175 start
BalI.
.EcoRV
B175 stop
B177 stop

Fig. III-3.

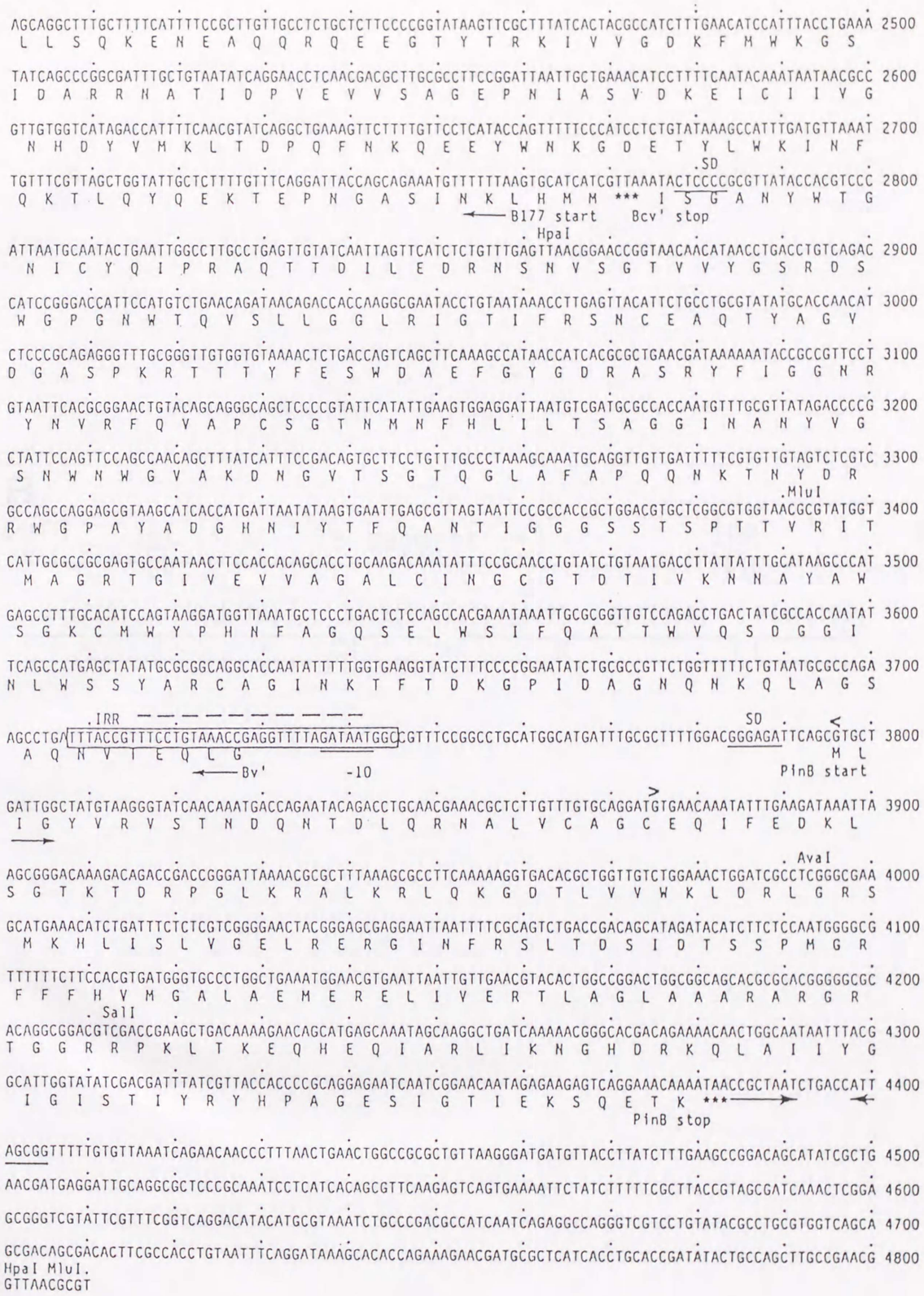

Fig. III-3. (continued)

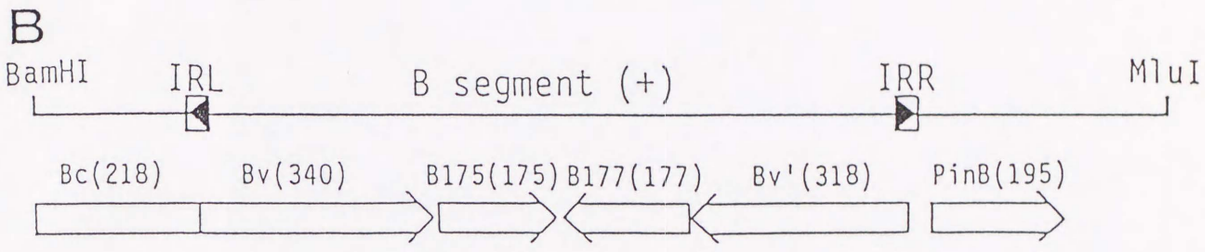

Fig. III-3. (continued)

Fig. III-4. Southern blot analysis of the B segment and pinB gene in Sh. boydii. Genomic or plasmid DNA was digested with BamHI and SalI and fractionated by electrophoresis in a 1% agarose gel. DNA was hybridized with nonradiolabeled probe containing the B segment and pinB gene (lane 1 and 2) or whole Mu genome (lane 3 and 4). Lanes: 1 and 3, pTSB916; 2 and 4, Sh. boydii genomic DNA.

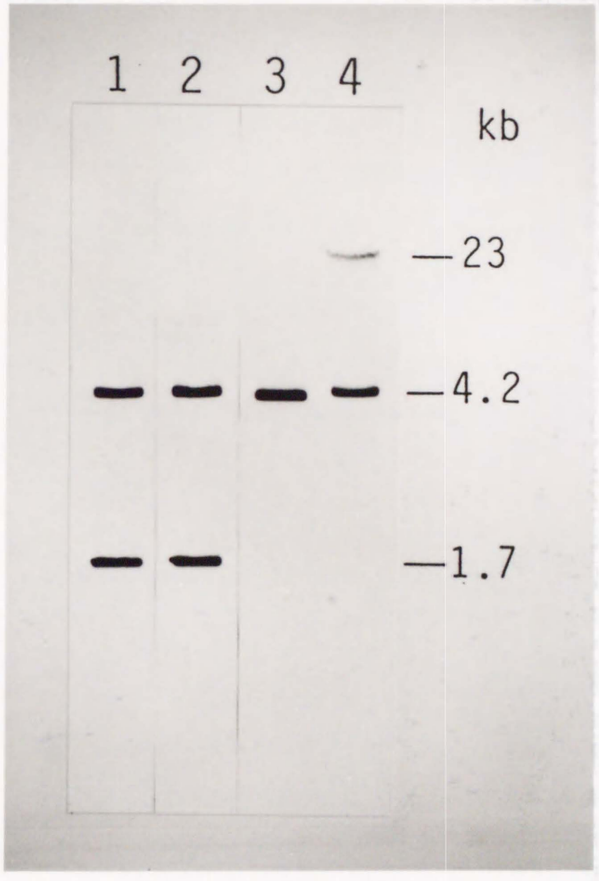

Fig. III-4.

Fig. III-5. SDS-polyacrylamide gel electrophoresis and Western blotting with anti-P1 sera. All of the crude extracts were prepared from EJ2371 containing plasmids. Samples were separated in a 12% acrylamide gel. Lanes: 1 and 2, SDS-polyacrylamide gels stained with Coomassie brilliant blue; 3 to 6, nitrocellulose blot probed with anti-P1C(+) serum (lane 3 and 4) and anti-P1C(-) serum (lane 5 and 6); 1, 3 and 5, crude extracts prepared from EJ2371 containing pPSB120; 2,4 and 6, crude extract prepared from EJ2371 containing pPSB110. Position and size (kDa) of molecular weight markers are given to the left of the panel. Numbers given to the right of the panel represent kDa of the polypeptides reacted with anti-P1 sera.

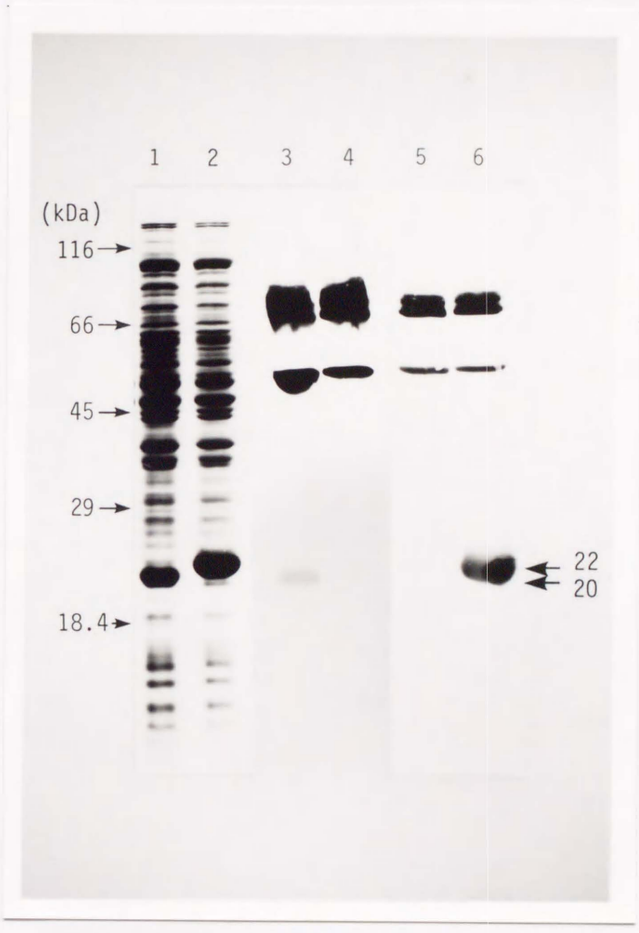

Fig. III-5.

Fig. III-6. Alignment of the predicted amino acid sequence of the recombinase gene products. Dots are shown where the amino acid is identical to that of PinB. Amino acids in the boxed areas are identical in all recombinases.

```
PinB    MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTKTDRPGLKRALKRLQKGDTLVVWKLDRLGRSM
Gin     ......................................R........D......................
PinE    ......................N.....L.....I....SE.....KL.PT.SA...V............
Cin     ...........E...A......ES....L.....A..K.AE.....KV.RM.SR................
Hin  MAT.....I.....I..I.......TS.N.DR....RI..KIAN.........YVN...............V

        KHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFHVMGALAEMERELIVERTLAGLAAARARGRTGGRR
        .................................................I...M.......NK..I...P
        R..VV..E...................T.........................K....ET..Q..I....
        R..VV..E...D...............T.........................R....D...E..I....
        .N.VA.IS..H...AH.H..........A.........S..............................Q..L...P

        PKLTKEQHEQIARLIKNGHDRKQLAIIYGIGISTIYRYHPAGESIGTIEKSQETK
        .....AEW..AG..LAQ.IP...V.L..DVAL..L.KK...KRAHIENDDRIN
        ....P..WA.AG...AA.TP.QKV.....DV.V..L.KRF...DK
        ..YQE.TWQ.MR..LEK.IP...V.....DVAV..L.KKF..SSFQS
        RAIN.HEQ...S..LEK..P.Q.......F...V..L...F..SSIKKRMN
```

Fig. III-6.