Proc. Natl. Acad. Sci. USA
Vol. 75, No. 2, pp. 615–619, February 1978
Biochemistry

Nucleotide sequence of an insertion element, IS1

(*Escherichia coli* plasmids/deletion events/illegitimate recombination/inverted repeat sequences)

HISAKO OHTSUBO AND EIICHI OHTSUBO

Department of Microbiology, Health Sciences Center, State University of New York at Stony Brook, Stony Brook, New York 11794

Communicated by Norman Davidson, October 19, 1977

ABSTRACT **PSM2, PSM1, and PSM15 are small plasmids derived from R100 by spontaneous deletions at either end of the insertion sequence IS1. These plasmids were used to identify regions neighboring IS1 as well as the IS1 DNA itself, by cleavage with *Eco*R1, *Hin*dIII, *Hae* II, *Hae* III, *Hpa* II, *Hha* I, *Hin*f, and *Alu* I. The nucleotide sequencing results demonstrate that IS1 contains 768 bases. About 30 bases at the ends of IS1 were found to be repeated in an inverted order. The deletions occurring at the ends of IS1 were found to be due to illegitimate recombination. The hypothesis that RNA polymerase could play an important role in such recombination phenomena is discussed based on the nucleotide sequences surrounding the recombinational hot spots.**

Insertion sequences (IS) are segments of DNA that occur in several size classes (700–1400 bases) and appear as insertion mutations in various operons (1, 2). These elements are present as repeated sequences in the *Escherichia coli* chromosome and bacterial plasmids including the fertility factor F and the resistance factor R (3–6). Insertion of an IS element into an operon causes strong polarity due to interference at the level of transcription (7, 8). Excision of this element restores normal function of the operon (9). In this way, IS elements can control gene expression.

IS elements have the ability to move into genes or DNA sequences as a single unit, indicating that the ends of the IS elements act as hot spots for recombination. Recent electron microscope heteroduplex studies on R plasmid DNA molecules showed that the ends of IS elements also act as hot spots in the formation of deletions and translocations of large DNA segments adjacent to the IS elements (10–15). These recombination events are illegitimate in nature but are site-specific in terms of involvement of the ends of IS elements. The frequent IS-mediated rearrangements of DNA segments strongly imply that these elements must have played an important role in the evolution of the organisms carrying them.

In this paper, we report on the restriction endonuclease cleavage map and nucleotide sequencing analysis of an IS element, IS1, about 700–800 bases long. We describe the entire nucleotide sequence of IS1, which greatly extends our previous partial sequencing (16). Some of the structural features of interest in the nucleotide sequences of IS1 and its neighboring regions are described and discussed as they relate to the control of gene expression and to site-specific recombination mediated by IS1 DNA.

MATERIALS AND METHODS

Plasmid Strains. The plasmid strains used were pSM1, pSM2, and pSM15, each of which carries one copy of the IS1 sequence (13). Isolation of closed circular duplex DNA of these plasmids has been described (17).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

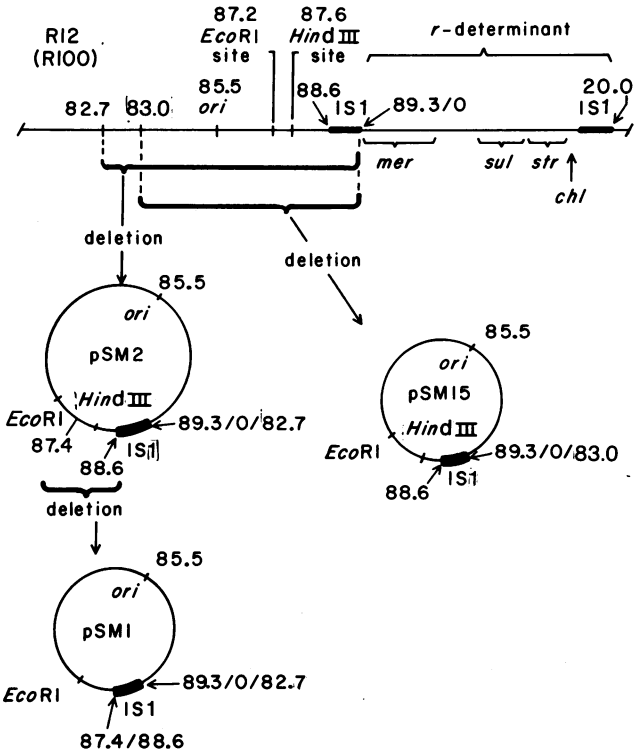

FIG. 1. Pedigree of the small circular plasmids (pSM1, pSM2, and pSM15), showing sequence relationships with their parent R factor, R12 (or R100). A critical part of R12 of circular DNA is shown. All numbers are distances in kilobases from an end point of an IS1 sequence on R12. The positions of the replication origin (*ori*), the *Eco*R1 and *Hin*dIII cutting sites, the sites for deletions to generate pSM plasmids, and *r*-determinant region responsible for resistance to mercuric ion (*mer*), sulfonamide (*sul*), streptomycin (*str*), and chloramphenicol (*chl*) have been mapped (13, 4, 19).

Enzymes. *Eco*R1 and *Hin*dIII were purchased from Miles Laboratory and were assayed as recommended by Miles Laboratory. Other restriction endonucleases (*Hae* II, *Hae* III, *Hha* I, *Hin*fI, *Hpa* II, and *Alu* I) were kindly supplied by H. Ohmori. The reaction mixture for these endonucleases contained 6 mM Tris·HCl (pH 7.9), 6 mM $MgCl_2$, 6 mM 2-mercaptoethanol, and bovine serum albumin (500 μg/ml). Bacterial alkaline phosphatase (36 units/ml) from Worthington Biochemical, in 0.1 M Tris·HCl, pH 8.0/10 mM $MgCl_2$, was used. Polynucleotide kinase (P-L Biochemicals, Inc.) was used for phosphorylation of the 5′ end of the DNA with [γ-^{32}P]ATP (specific activity, ~2000 Ci/mmol, ICN 35001-HH). The reaction mixture for this enzyme contained 50 mM Tris·HCl (pH 7.5), 10 mM $MgCl_2$, and 5 mM dithiothreitol.

Gel Electrophoresis. To separate native DNA fragments, electrophoresis was carried out on a 13 × 15 × 0.2 cm 4% or 10%

Abbreviation: IS, insertion sequence.

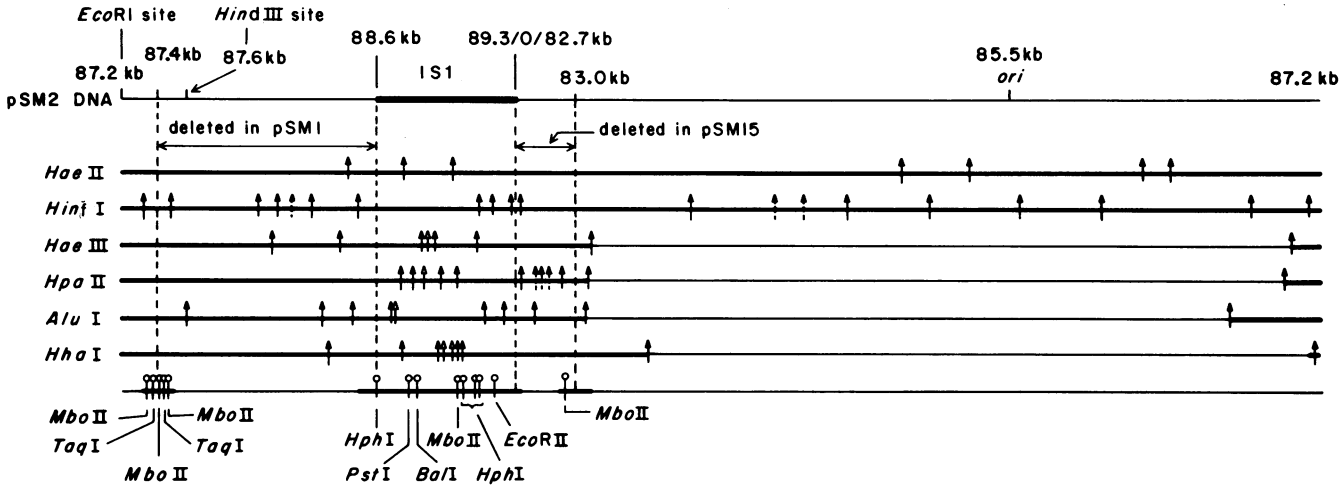

FIG. 2. Cleavage maps of pSM2 DNA with various restriction enzymes. pSM2 is displayed in a linear representation by cutting the circle at the *Eco*R1 cleavage site. Thick lines represent the regions analyzed by endonuclease cleavage. Arrows indicate cleavage sites with various restriction enzymes. Arrows with open heads are the cleavage sites that could only be fixed after nucleotide sequence analysis (see Fig. 4). The bottom line summarizes cleavage sites of restriction endonucleases not used in the experiment; these sites were found after nucleotide sequencing analysis of the thick lined region.

polyacrylamide slab gel (acrylamide/bisacrylamide, 20:1) with E buffer (50 mM Tris·borate, pH 8.3/1 mM EDTA). DNA bands were visualized with ethidium bromide (4 μg/ml in E buffer) under a short-wavelength UV light.

Determination of Nucleotide Sequences. 5′-^{32}P-Labeled fragments were obtained by strand separation (18) or by cleavage of the 5′-^{32}P-labeled double strand with another restriction enzyme. The nucleotide sequence of these fragments was determined as described by Maxam and Gilbert (18).

RESULTS

Cleavage Maps of the Region Containing IS1. The physical structures of the three small plasmids used for the mapping experiments are shown in Fig. 1. The sequence relationships among these plasmids and their parental R factor, R12 (same as R100), have been determined with a coordinate system in kilobases by the electron microscope heteroduplex method (13). pSM2 and pSM15 were independently derived from the parent R12 by deletion. pSM1 was derived from pSM2 spontaneously by deletion. It should be noted that all of the deletion events occurred at the ends of the IS1 sequence (13). An alternative map showing sequence relationships among these plasmids is shown in a linear representation at the top of Fig. 2. pSM1 and pSM15 are missing different portions of pSM2, located at either end of the IS1 sequence of pSM2.

These three plasmids were cleaved by *Hae* II, *Hae* III, *Hin*f, *Hha* I, and *Alu* I and were subjected to gel electrophoresis. The deletion in pSM1 or pSM15 resulted in a gel band pattern different from that of pSM2. Comparison of gel band patterns of pSM1 or pSM15 with pSM2 enabled us to identify the fragments containing the neighboring regions of IS1 as well as those containing the ends of IS1 DNA. Elution of the DNA fragments generated by a restriction endonuclease and further digestion of these DNA fragments with another restriction endonuclease allowed us to order the fragments containing the IS1 sequence and its neighboring regions. Fig. 3 shows the cleavage maps of these three small plasmids. All of the fragments were oriented from the *Eco*R1 or *Hin*dIII cutting site, the location of which

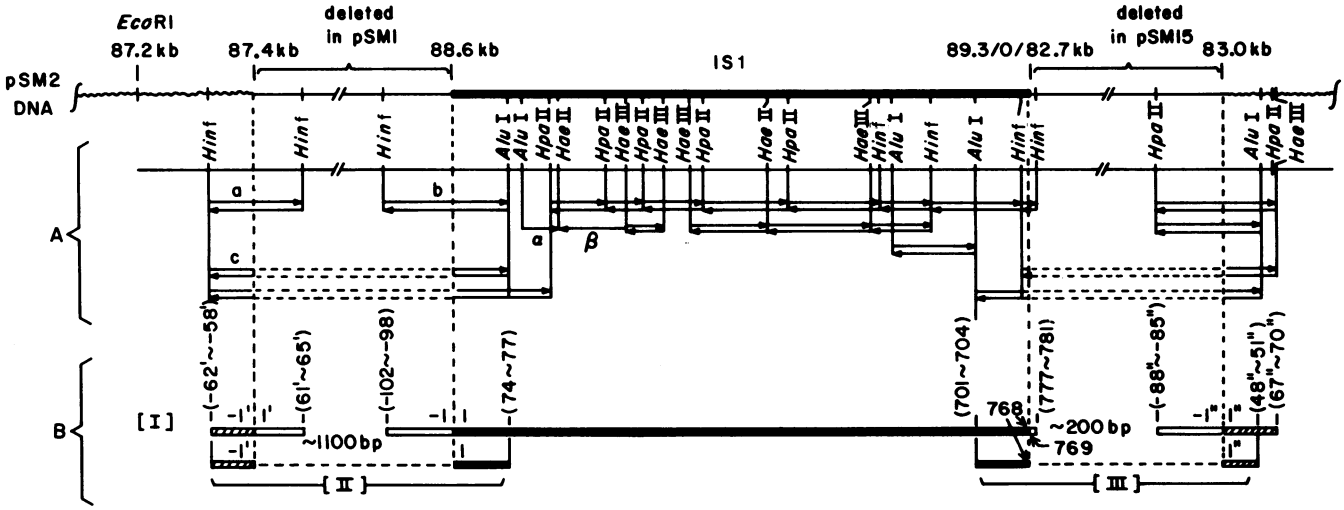

FIG. 3. Labeled strands used for determination of the nucleotide sequence of the regions (I, II, and III) are shown at the bottom of the figure. Arrows indicate 5′-^{32}P-labeled strand aligned in the 5′ → 3′ direction. Most of the strands were obtained by the strand separation technique, except for strands indicated by α and β which were obtained by secondary cleavage of 5′-^{32}P-labeled double-stranded DNA. The numbers in the region I, II, and III indicate the positions of nucleotides from selected origins that were based on nucleotide sequencing results of those regions (see Fig. 4).

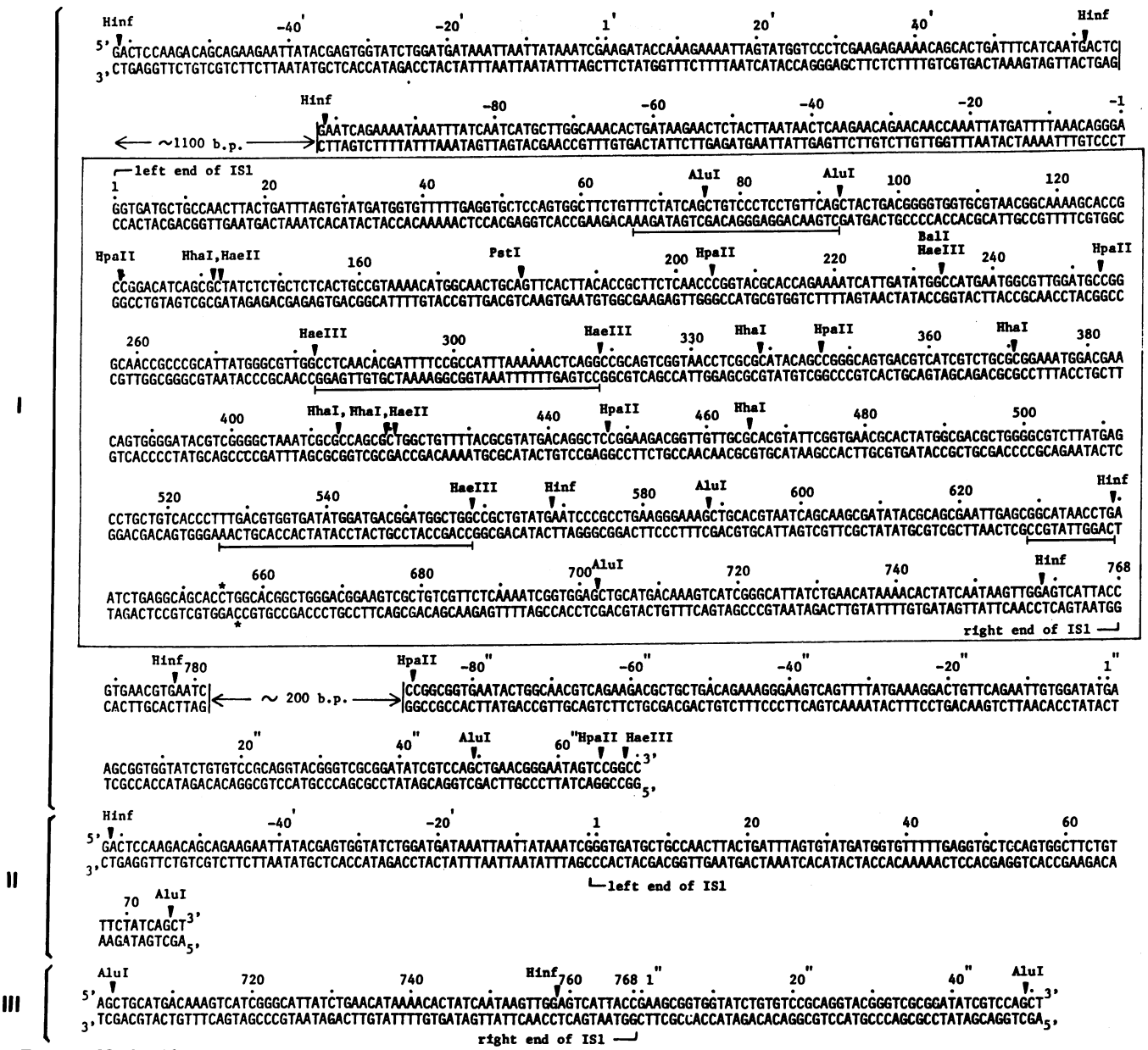

FIG. 4. Nucleotide sequence in the region I, II, and III shown in Fig. 3. Determined nucleotide sequences of IS1 are numbered from 1 to 768. Other numbers outside of IS1 are named as indicated in this figure to facilitate the sequence relationships among regions I, II, and III (see Figs. 5 and 6). Wedges indicate the cleavage sites of each enzyme on one strand which is aligned with 5′ → 3′ direction in this sequence. C at 655 and 657 shows modified dCMP, probably 5-methyl dCMP. The nucleotide sequence for approximately 100 residues from the 5′ end of a strand could be determined with reasonable accuracy. Most of the DNA sequences of IS1 and its neighboring regions were determined by sequencing both strands, except that the regions 67–92, 282–318, 527–558, and 629–639 were determined by sequencing of only one strand as indicated by underlines in the figure.

had been determined (13, 19). As will be discussed below, nucleotide sequencing analysis of these fragments confirmed the cleavage maps that we constructed.

Nucleotide Sequence Analysis in the Regions Containing IS1. Part A of Fig. 3 summarizes the DNA fragments, sequenced by the Maxam and Gilbert method (18), for determination of nucleotide sequences of the regions in pSM1, pSM2, and pSM15 that are indicated in B of Fig. 3. The final nucleotide sequences determined for these regions are shown in Fig. 4. Fig. 5 shows the radioautographs used in the determination of the sequences of two large DNA fragments, b and c, shown in Fig. 3.

Fig. 6 presents the sequences relevant to the identification of the left and right end points of IS1. As shown in B of Fig. 6, sequence c of pSM1 is derived by illegitimate recombination between a and b of pSM2. We define the left end of IS1 as the sequence present in both b of pSM2 and c of pSM1 or as the sequence in c of pSM1 but absent in a of pSM2. This enabled us to determine the left end of IS1 to be the G residue labeled 1. pSM15 and pSM2 were independently derived from R12 by excision events. We define the right end of IS1 as the sequence common to f of pSM3 and e of pSM15 or as the sequence in a of pSM15 but not common to d of pSM2. This enabled us to assign the right end of IS1 to be the G residue labeled 769 or the C residue labeled 768.

DISCUSSION

We have described the nucleotide sequence of IS1 and its neighboring regions that were present in small plasmids derived

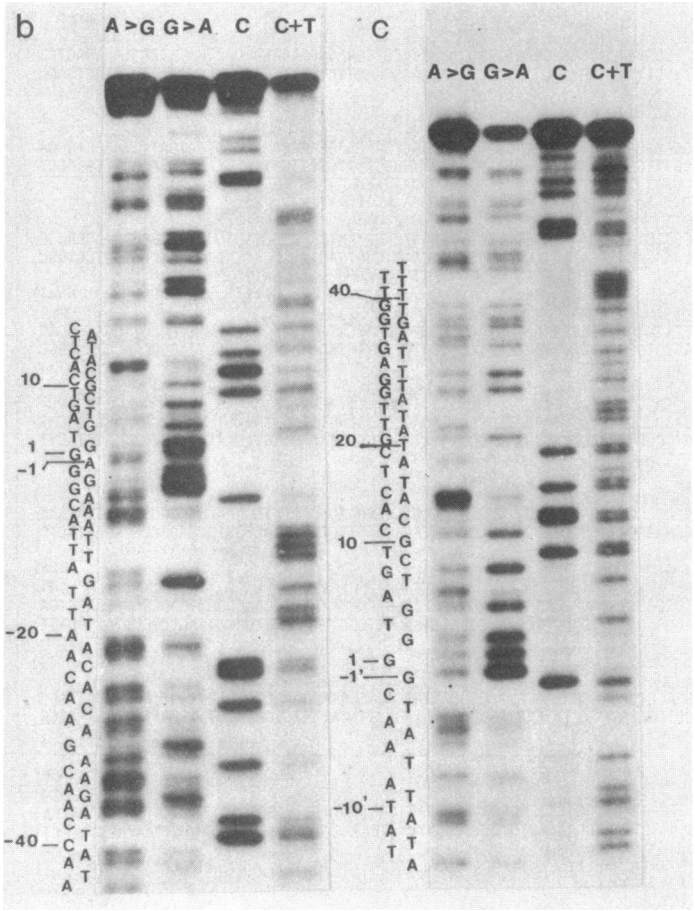

FIG. 5. Radioautograph used for determination of the sequences of strands b and c shown in Fig. 3. These sands contained sequences b and c which are shown in B of Fig. 6 for the determination of the left end of IS1. Each strand, labeled at the 5′ end with ^{32}P, was subjected to the four sets of chemical reactions. Aliquots of the four samples of base-specific cleavage products of each strand were analyzed by electrophoresis for 36 hr in a 30 × 40 × 0.15 cm slab gel of 20% polyacrylamide/7 M urea, according to the method developed by Maxam and Gilbert (18). Critical parts of the sequence read from gel patterns are indicated with numbers.

from an R factor, R12. IS1 DNA was determined to contain 768 or 769 bases. As explained above, the end points of the IS1 sequence were defined as the points where deletions appear to have occurred specifically. This raises the possibility that hot spots for deletion are not exactly the ends of IS1. At present, the nucleotide sequence results that are available for determination of the ends of IS1 are the preliminary ones obtained by Grindley (20) on the end region of an IS1 that occurred in the *gal* operon. Our results agree with his when we take the G residue labeled 1 as the left end of IS1 and the C residue at 768 as the right end of IS1 as shown in Fig. 6. IS1 DNA was therefore determined to contain 768 bases. This may also indicate that both deletions and insertions are site-specific events.

It is interesting to note that we have found inverted repeat sequences at the very ends of IS1, although there are some mismatched regions in those repeats. Fig. 7 shows this inverted repetition as an inversion loop structure. These repeated sequences at the ends of IS1 are reminiscent of the fact that most transposable DNA segments carrying antibiotic resistance genes are flanked by inverted repeated sequences (10–12, 21–23). *Tn9*, a transposable element for the chloramphenicol resistance gene, however, is flanked by direct repeats of IS1 (14, 15). In this case, however, *Tn9* also has inverted repeats, because IS1

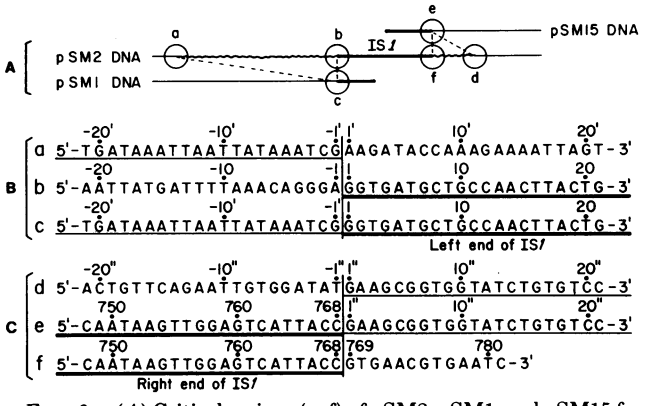

FIG. 6. (*A*) Critical regions (a–f) of pSM2, pSM1, and pSM15 for the determination of nucleotide sequence at the ends of IS1. (*B*) Comparison of the sequences a, b, and c fixes the left end of IS1. (*C*) Comparison of the sequences d, e, and f fixes the right end of IS1.

itself was found to have inverted sequences at its ends. These facts suggest that the presence of inverted repeats is vital for insertion of IS1 and transpositions.

As shown in B of Fig. 6, two different sequences, a and b, recombine to give sequence c. This recombination takes place at a nonhomologous region, although these regions contain (AT)-rich clusters that always occur at almost the same positions relative to the site of recombination (see fragments a and b in B of Fig. 6). This illegitimate but site-specific recombination may thus be due to the presence of some specific sequences around the recombination sites. We notice that these regions contain nucleotide sequences that are similar to or the same as those seen in promoter and operator regions of various operons that are recognized and bound by DNA-dependent RNA polymerase. The sequences

$$\overset{30}{\dot{T}}\text{-A-T-G-A-T-}\overset{36}{\dot{G}} \text{ and } \overset{730}{\dot{G}}\text{-A-T-A-A-T-}\overset{724}{\dot{G}}$$

which are found near the left and right ends of IS1, are seen in the t-RNA^{Tyr} promoter region (24) and in the rightward early promoter region (*pr*) of phage λ (25, 26). These sequences are similar to the sequence, T-A-T-A-A-T-R, which was proposed by Pribnow (27) to be an important sequence involved in RNA polymerase binding. It should be noted that many other sequences similar to Pribnow's are also found in the region within 50 bases from the recombination junction sites (see Fig. 6).

The sequence

$$\overset{1}{\dot{G}}\text{-G-T-G-A-T-G-C-T-}\overset{10}{\dot{G}}\text{-C-C-A-A-}\overset{15}{\dot{C}}\text{-T-T,}$$

seen at the left end of IS1, is quite similar to the promoter sequence of *pl*

$$\overset{-14}{\dot{G}}\text{-G-T-G-}\overset{-10}{\dot{A}}\text{-T-A-C-T-G-A-G-C-A-}\overset{0}{\dot{C}}\text{-A-T,}$$

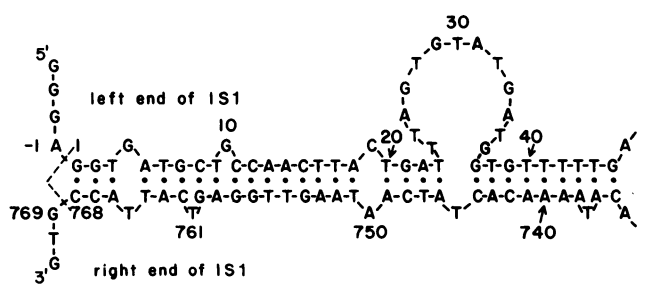

FIG. 7. Inverted repeat sequences at the ends of IS1 shown as an inversion loop structure.

which is protected by RNA polymerase from the attack of DNase *in vitro* (28, 29). The sequence

$$\overset{4''}{\dot{G}}\text{-C-G-G-T-G-}\overset{10''}{\dot{G}}\text{-T-A-T-C-T-G-T-G-T-}\overset{20''}{\dot{C}}\text{-C-G-C-A}$$

is similar to the sequence

$$\overset{-16}{\dot{C}}\text{-C-A-G-}\overset{-20}{\dot{T}}\text{-G-G-T-A-T-T-T-A-T-}\overset{-30}{\dot{G}}\text{-T-C-A-A-C-A}$$

which is found in the promoter region (*pl*) of phage λ. The sequence

$$\overset{-21'}{\dot{T}}\text{-G-A-T-A-A-A-T-T-}\overset{-12'}{\dot{A}}$$

is exactly the same as the sequence seen in *pl* promoter region −56 to about −47, which is thought to be recognized by RNA polymerase.

Although experimental evidence is not yet available, the similarities between the nucleotide sequences in the IS1 end regions and in known promoter regions of various genes suggest that RNA polymerase might play an important role in IS1-mediated recombination. We propose that the RNA polymerase can recognize the DNA sequences at the ends of IS1 and then opens up (or unwinds) these DNA strands. The recent report by Ikeda and Kobayashi (30) on the results of their genetic analysis of RNA polymerase involvement in a *recA*-independent (or illegitimate) pathway of genetic recombination in *E. coli* could be related to our suggestion that RNA polymerase activity may be required for illegitimate recombination.

The polarity effect caused by IS1 could also be explained by involvement of RNA polymerase binding regions that are generated by the insertion of IS1 and block the normal initiation of transcription of structural genes. However, several nonsense codons occur in all three possible reading frames in the IS1 sequences; this fact may be sufficient to explain polarity effects.

Finally, it is interesting to note that a part of our sequence at the right end of IS1,

$$\overset{1}{\dot{G}}\text{-G-T-G-A-T-G-C-T-}\overset{10}{\dot{G}}\text{-C-C-A-A-C-T-T-A-C-}\overset{20}{\dot{T}}\text{-G-A-T}$$

(which is inverted at the left end of IS1 as described above) is the same as the nucleotide sequence seen in the attachment site of phage λ,

$$\overset{20}{\dot{T}}\text{-T-}\overset{15}{\dot{T}}\text{-T-A-T-A-}\overset{10}{\dot{A}}\text{-T-G-C-C-}\overset{5}{\dot{A}}\text{-A-C-T-T-}\dot{A}\text{-G-T-A-T}$$

which was determined by Landy and Ross (31). Although this may be a coincidence, it is possible that these site-specific recombination systems follow similar or identical pathways.

We are grateful to A. Maxam for providing the details of the procedure for determining DNA sequences prior to publication. We also thank H. Ohmori and J. I. Tomizawa for invaluable guidance with the sequencing techniques, H. Ohmori for providing various restriction enzymes, J. Feingold for large-scale bacterial cultures, J. Rosen for correction of errors in the original manuscript, and J. Kates for his constant encouragement and his generous criticism during preparation of the manuscript. H.O. received partial support from U.S. Public Health Service Training Grant CA09176. This work was supported by U.S. Public Health Service Grant GM22007 to E.O.

1. Hirsch, H. J., Starlinger, P. & Brachet, P. (1972) *Mol. Gen. Genet.* **119,** 191–206.
2. Fiandt, M., Szybalski, W. & Malamy, M. H. (1972) *Mol. Gen. Genet.* **119,** 223–231.
3. Saedler, H. & Heiss, B. (1973) *Mol. Gen. Genet.* **122,** 267–277.
4. Hu, S., Ohtsubo, E., Davidson, H. & Saedler, H. (1975) *J. Bacteriol.* **122,** 764–775.
5. Hu, S., Ptashne, K., Cohen, S. N. & Davidson, N. (1975) *J. Bacteriol.* **123,** 687–692.
6. Ohtsubo, H. & Ohtsubo, E. (1977) in *DNA Insertion Elements, Plasmids and Episomes,* eds. Bukhari, A. I., Shapiro, J. & Adhya, S. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 591–594.
7. Starlinger, P., Saedler, H., Rak, B., Tillmann, E., Venkow, P. & Waltschewa, L. (1973) *Mol. Gen. Genet.* **122,** 279–286.
8. Adhya, S., Gottesman, M., de Crombrugghe, B. & Court, D. (1976) in *RNA Polymerase,* eds. Losick, R. & Chamberlin, M. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 719–730.
9. Saedler, H. & Starlinger, P. (1967) *Mol. Gen. Genet.* **100,** 178–189.
10. Tye, B. K., Chan, R. K. & Botstein, D. (1974) *J. Mol. Biol.* **85,** 485–500.
11. Kleckner, N., Chan, R. K., Tye, B. K. & Botstein, D. (1975) *J. Mol. Biol.* **97,** 561–575.
12. Berg, D. E., Davis, J., Allet, B. & Rochaix, J. D. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 3628–3632.
13. Mickel, S., Ohtsubo, E. & Bauer, W. (1977) *Gene,* in press.
14. Chow, L. T. & Bukhari, A. I. (1977) in *DNA Insertion Elements, Plasmids and Episomes,* eds. Bukhari, A. I., Shapiro, J. & Adhya, S. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 295–306.
15. MacHattie, L. A. & Jackowski, J. B. (1977) in *DNA Insertion Elements, Plasmids and Episomes,* eds. Bukhari, A. I., Shapiro, J. & Adhya, S. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 219–228.
16. Ohtsubo, H. & Ohtsubo, E. (1977) in *DNA Insertion Elements, Plasmids and Episomes,* eds. Bukhari, A. I., Shapiro, J. & Adhya, A. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 46–63.
17. Clewell, D. B. & Helinski, D. R. (1970) *Biochemistry* **9,** 4428–4440.
18. Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
19. Ohtsubo, E., Feingold, J., Ohtsubo, H., Mickel, S. & Bauer, W. (1977) *Plasmid,* **1,** 8–18.
20. Grindley, N. G. F. (1977) in *DNA Insertion Elements, Plasmids and Episomes,* eds. Bukhari, A. I., Shapiro, J. & Adhya, S. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 591–596.
21. Heffron, F., Rubens, C. & Falkow, S. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 3623–3627.
22. Kopecko, D. J. & Cohen, S. N. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 1373–1377.
23. Cohen, S. N. (1976) *Nature* **263,** 731–738.
24. Sekiya, T., Takeya, T., Coutreras, R., Kupper, H., Khorana, H. G., & Landy, A. (1976) in *RNA Polymerase,* eds. Losick, R., & Chamberlin, M. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 455–472.
25. Pirrotta, V. (1975) *Nature* **254,** 114–117.
26. Walz, A. & Pirrotta, V. (1975) *Nature* **254,** 118–121.
27. Pribnow, D. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 784–788.
28. Maniatis, T., Ptashne, M., Barrell, B. G. & Donelson, J. (1974) *Nature* **250,** 394–397.
29. Maniatis, T., Jeffrey, A. & Kleid, D. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 1184–1188.
30. Ikeda, H. & Kobayashi, I. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 3932–3936.
31. Landy, A. & Ross, W. (1977) *Science* **197,** 1147–1160.