JOURNAL OF BACTERIOLOGY, Aug. 1975, p. 687-692

Vol. 123, No. 2
Printed in U.S.A.

$\alpha\beta$ Sequence of F is *IS3*[1]

SYLVIA HU,* KAY PTASHNE, STANLEY N. COHEN, AND NORMAN DAVIDSON

Department of Chemistry, California Institute of Technology, Pasadena, California 91125, and Department of Medicine, Stanford University School of Medicine, Stanford, California 94305*

Received for publication 31 March 1975

Previous studies have shown that there is a deoxyribonucleic acid (DNA) segment, of length 1.3 kb and denoted as the $\alpha\beta$ sequence, which occurs twice on the F plasmid at coordinates 93.2 to 94.5/0F kb and 13.7 to 15.0F kb. In the present investigation, heteroduplexes were prepared between a phage DNA carrying the insertion sequence *IS3* and suitable F-prime DNAs. The hybrids formed show that *IS3* is the same as $\alpha\beta$. This result plus previous studies support the view that: (i) the insertion sequences *IS2* and *IS3* occur on F and, in multiple copies, on the main bacterial chromosome of *Escherichia coli* K-12; and (ii) these *IS* sequences on the main bacterial chromosomes are hot spots for Hfr formation by reciprocal recombination with the corresponding sequences of F.

There is a class of spontaneous, strongly polar mutations which occur in *Escherichia coli* and its temperate bacteriophages and are caused by the insertion of a short segment of deoxyribonucleic acid (DNA) into the gene mutated (for a review, see 12). A small number of distinct, specific sequences—*IS1*, *IS2*, *IS3*, and *IS4* with respective lengths of approximately 750, 1,300, 1,300, and 1,400 nucleotide pairs—have been identified as the sequences responsible for many insertion mutations (3, 4).

Hybridization studies have shown that *IS1* and *IS2* are resident on the chromosome of several strains of *E. coli* K-12 in amounts of about eight and five copies, respectively (9). We presume that similar studies would show that *IS1*, *IS2*, *IS3*, and *IS4* all occur at moderate frequencies in most strains of *E. coli*. Recent studies have shown that *IS1*, *IS2*, and *IS3* are present in several R plasmids, and their positions have been mapped by heteroduplex analysis (6, 8). These map positions strongly suggest that *IS1* and *IS3* are active in R plasmid recombination phenomena.

Studies from one of our laboratories have shown that there are certain special segments of the plasmid F which are hot spots for F^- related recombination phenomena (2, 5). Two of these segments, each with a length of 1.3 ± 0.1 kb, are: (i) the $\epsilon\zeta$ sequence with F coordinates of 16.3 to 17.6F; (ii) the $\alpha\beta$ sequence which occurs twice on F, at coordinates 93.2 to 94.5/0F, and 13.7 to 15.0F (for an explanation of our notation and coordinate system see 5). Studies of the structure of the F13 plasmid support the hypothesis that the $\epsilon\zeta$ and $\alpha\beta$ sequences are also present on the main chromosome of *E. coli* and that many cases of Hfr formation from an F^+ cell involve reciprocal recombination between one of the sequences on F and the corresponding sequences on the main chromosome.

It was recently shown that the $\epsilon\zeta$ sequence of F is *IS2* (6). We now report an additional instance where a segment of F active in recombination will hybridize with a genetically characterized *IS* sequence; namely, we show here by heteroduplex analysis that $\alpha\beta$ is *IS3*.

MATERIALS AND METHODS

Bacterial and phage strains. Bacterial strains used in this study are listed in Table 1. λp*lac5S7*-MS505 (abbreviated λp*lac*MS505) carries one copy of *IS3* inserted into the operator region of the *lac* operon (7, 8).

Preparation of plasmid DNA. Bacteria were grown, and closed circular plasmid DNA was extracted and converted to the singly nicked form by X rays as previously described (5, 11). Separated strands of the DNA of phage λp*lac*MS505 were prepared as described previously (8).

Heteroduplex preparation and other aspects of our electron microscope technique have been described previously (6).

TABLE 1. *E. coli K-12 strains*

Strain	Plasmid	Chromosomal genotype	Source and reference
JE3513	F8-33	*thr leu gal12 lac52 pil fla str*^{r}	E. Ohtsubo (11)
JE513	F13-4	*lac purB str*^{r}	Y. Hirota (5)

[1] Contribution no. 5083 from the Department of Chemistry of the California Institute of Technology.

RESULTS

To test whether *IS3* occurs in F, we have searched for heteroduplex structures between suitable F plasmids and a DNA molecule known to contain *IS3*. For the latter we used λp*lac*MS505 DNA. In a mixture containing the two complementary strands of the phage DNA and the two complementary strands of the plasmid DNA, the rate of formation of a heteroduplex between the short *IS3* sequence on a phage strand and its complement on a plasmid strand is expected to be slow compared to the competing phage-phage and plasmid-plasmid reactions leading to reformation of homoduplexes. As noted previously, satisfactory rates of heteroduplex formation for *IS* segments can be achieved in mixtures containing only one of the strands of the phage DNA and a mixture of the two complementary strands of the plasmid DNA (6, 8). In the present study, a preparation of the light (L) strand of λp*lac*MS505 was used.

One plasmid DNA that is particularly useful in searching for heteroduplexes involving $\alpha\beta$ is F8-33. Its structure is shown in Fig. 1 (11). It contains only one copy of $\alpha\beta$, i.e., $\alpha_1\beta_1$ at coordinates 93.2 to 94.5/0F. (For an explanation of our notation and coordinate system see 5.) F8-33 also contains two copies of *IS2* in an inverted order, so that a single strand of F8-33 frequently shows the inverted repeat duplex structure depicted in Fig. 1. This duplex region provides a reference marker for mapping other features in heteroduplex structures with F8-33. In heteroduplex preparations containing λp*lac*MS505 and F8-33 DNA, approximately 1 to 2% of the F8-33 strands were heteroduplexes with λp*lac*MS505 DNA of the type shown in Fig. 1b. There is a duplex segment of length 1.3 ± 0.3 kb, bounded by single-strand forks on both sides, at a distance of 29.9 ± 1.3 kb from the inverted repeat reference structure. A typical electron micrograph is shown in Fig. 2a.

It should be noted that the λp*lac*MS505 L strand DNA that was used was fragmented, with very few intact strands. Therefore, it was not possible to identify the sequence on the phage DNA which formed a duplex with F8-33 DNA as the *IS3* sequence by observing the position of the duplex region on a complete λp*lac*MS505 strand. However, previous hybridization studies have shown that there are no other λ or *lac* sequences of this (or any other) length complementary to F sequences (6, 8, 10).

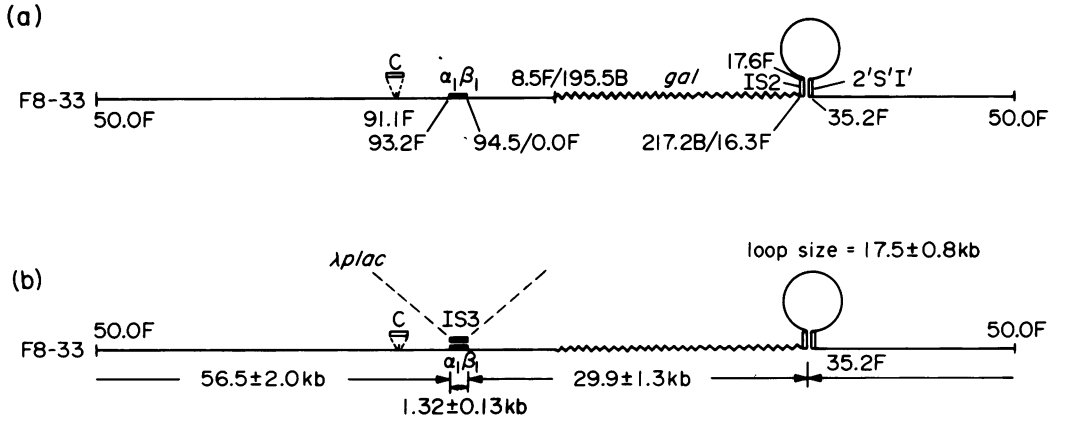

FIG. 1. *Schematic representations of the structure of F8-33 and of its heteroduplex with λplacMS505. To conserve space, circular maps are represented in a linear form. Our system for coordinate assignments on F and F-prime plasmids has been described elsewhere (5, 11). The essential points are that coordinates in kilobase units are given relative to certain defined origins. F sequences are shown as a straight line with coordinates denoted by the suffix F. Bacterial chromosomal sequences are shown as a sawtooth line with coordinates denoted by the suffix B. Special sequences are shown as a heavier or a double line. (a) The structure of a single strand of F8-33. This plasmid contains only one copy of $\alpha\beta$, denoted as $\alpha_1\beta_1$. It contains two copies of IS2 in an inverted order. In a spreading of a single strand, the two complementary IS2 sequences often pair to give the inverted repeat structure shown in the drawing. This provides a reference point for mapping other features. (b) The structure of a heteroduplex of F8-33 with fragmented strands of λplacMS505 which contains IS3. There is a duplex region bounded by single-strand forks on each side due to hybridization of the IS3 sequence on the bacteriophage DNA with its homologous sequence on F8-33. The distance of this duplex region from the inverted repeat shows that the duplex region of the heteroduplex is within experimental error at the position of $\alpha_1\beta_1$, with an alternative assignment eliminated by the study of λplacMS505/F13-4 heteroduplexes depicted in Fig. 3.*

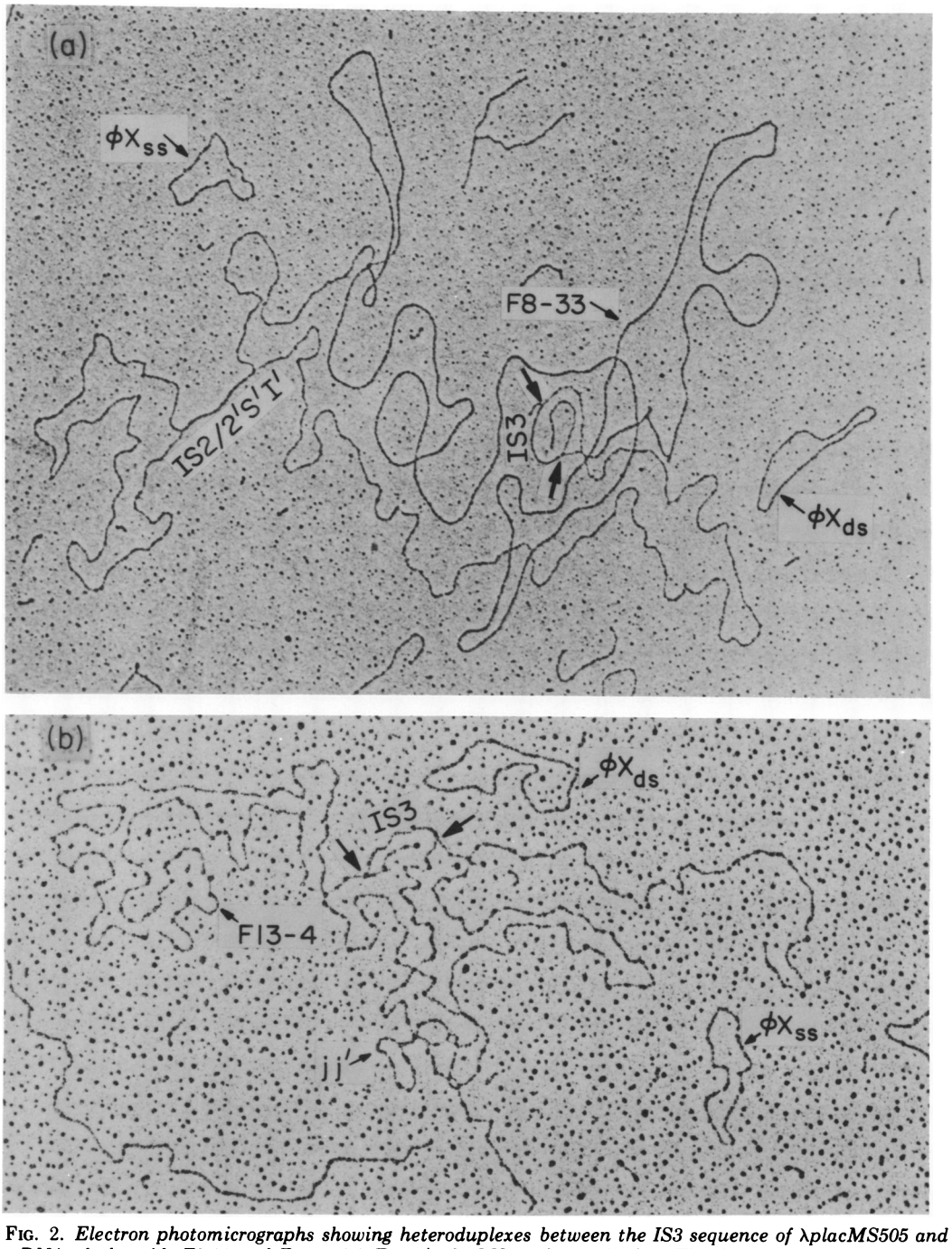

FIG. 2. *Electron photomicrographs showing heteroduplexes between the IS3 sequence of λplacMS505 and the DNA of plasmids F8-33 and F13-4. (a) F8-33/λplacMS505 heteroduplex. The duplex region bounded by single-strand forks corresponding to the IS3/$\alpha_1\beta_1$ hybrid is marked by two arrows. The inverted repeat structure due to pairing of the two IS2 sequences on F8-33 is also shown. φX single-strand and φX double-strand DNA are present as internal standards for length measurements. Part of the F8-33 DNA is duplex as a result of association with a fragment of complementary F8-33 DNA, but this does not affect the interpretation of the micrograph. (b) Electron photomicrograph of a structure of a broken strand of F13-4 with a fragment of λplacMS505 DNA. There is a short duplex region bounded by single-strand forks due to base pairing of IS3 of the phage DNA with $\alpha_2\beta_2$ of F13-4. The reference feature jj' of F13-4 is identified.*

Therefore, we conclude that there is an *IS3* sequence on F8-33.

The structure in Fig. 1a therefore shows that *IS3* on F8-33 is either 29.9 (±1.3) kb clockwise or counterclockwise from the inverted repeat sequence, corresponding either to F coordinates 65.1 to 66.4 kb or (within experimental error) to the coordinates of $\alpha_1\beta_1$ at 93.2 to 94.5/0F.

This twofold ambiguity was resolved in favor of the latter choice by analyzing the heteroduplexes of λp*lac*MS505 with F13-4 DNA. The structure of F13-4 (5) is shown in Fig. 3a. F13-4 contains both of the copies on F of $\alpha\beta$, i.e., $\alpha_1\beta_1$ at 93.2 to 94.5/0F and $\alpha_2\beta_2$ at 13.7 to 15.0F. There is a small inverted repeat structure denoted *jj'* within the chromosomal sequences of F13-4; this feature provides a reference point on a single strand of F13-4 DNA. F13-4 also contains *lac* sequences within the chromosomal sequences.

Two kinds of heteroduplexes between F13-4 and λp*lac*MS505 were seen. In one kind, there was a duplex region of length 3.9 (±0.3) kb with an insertion loop of length 1.3 kb due to heteroduplexes between the *lac* sequences of F13-4 and of λp*lac*MS505, as indicated in Fig. 3b. More importantly, heteroduplexes with a duplex segment of length 1.3 kb due to the *IS3* sequence were also seen. These duplex regions were mapped relative to the *jj'* feature. This analysis shows that *IS3* occurs at $\alpha_1\beta_1$ and $\alpha_2\beta_2$, but never at the alternative position of 65.1 to 66.4F allowed by the F8-33 study. The frequency observed for structures that we attribute to heteroduplexes of the *IS3* sequence on the phage DNA with an $\alpha\beta$ sequence on an F13-4 strand with an identifiable *jj'* feature was of the order of 1 to 2%. In control experiments with λp*lac*MS505 DNA omitted from the hybridization mixture the frequency of similar structures (due to out-of-register renaturation, i.e., for example a broken fragment of F13-4 containing $\alpha_2\beta_2$ hybridizing with $\alpha_1\beta_1$ on a large F13-4 strand) was unobservably small. Micrographs with typical heteroduplexes are shown in Fig. 2b.

We conclude that the $\alpha\beta$ sequence of F is identical, by the heteroduplex criterion, with *IS3*.

DISCUSSION

It has now been shown that the $\alpha\beta$ and $\epsilon\zeta$ sequences of F form duplex hybrids with *IS3* and *IS2*, respectively. As pointed out previously (6), two segments that associate to form a duplex under electron microscope spreading conditions may not be perfectly complementary—there can be some partial sequence mis-

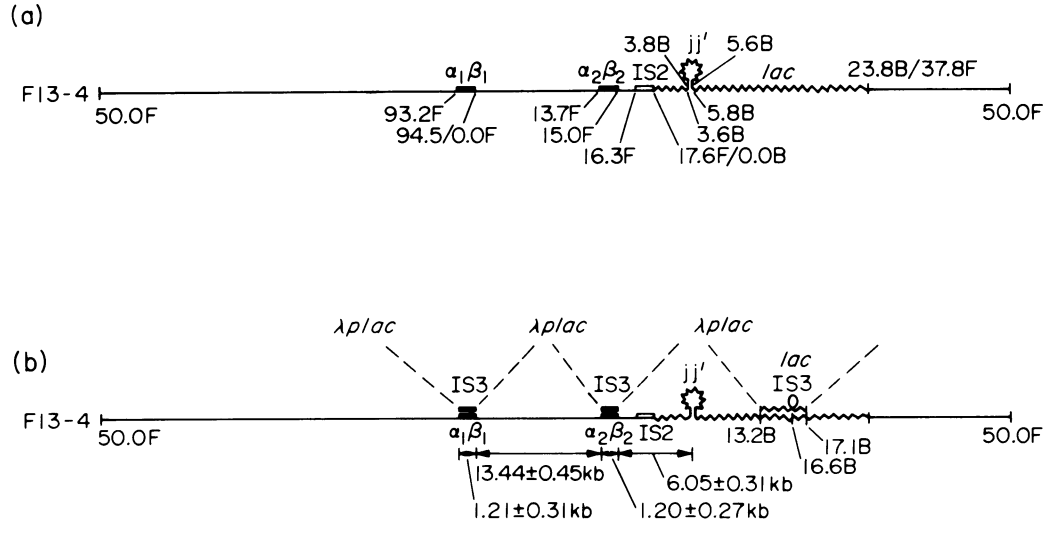

FIG. 3. *Schematic representation of the structure of F13-4 and of the several possible heteroduplexes with λplacMS505. (a) Structure of a single strand of F13-4. The notation is basically the same as that used in Fig. 1. There is a small inverted repeat feature labeled jj' in the chromosomal sequences of F13-4 which is useful as a reference marker for mapping purposes. Note that both $\alpha_1\beta_1$ and $\alpha_2\beta_2$ occur on the F sequences of F13-4, but that the F sequences from 17.6F to 37.8F have been substituted by chromosomal sequences, including the jj' feature and lac genes. (b) Structures of possible heteroduplexes between F13-4 and λplacMS505. Heteroduplexes were observed with the lac sequences of the phage DNA hybridized with those of F13-4, and of the IS3 sequence of the phage hybridized with $\alpha_1\beta_1$ and with $\alpha_2\beta_2$ of F13-4.*

match. However, the simpler interpretation of the results is that *IS3* and *IS2* are identical to $\alpha\beta$ and $\epsilon\zeta$ on F, respectively.

The F-prime plasmids F42, F100, and F152 all have structures consistent with the hypothesis that the parental Hfr cells (P804 for F42, P3 for F100 and F152) were formed by reciprocal recombination between $\alpha_1\beta_1$ (*IS3*) on F and *IS3* resident on the chromosome of the F^+ parent of the Hfr. The structure of F13 is consistent with the hypothesis that Hfr 13 was formed by a reciprocal recombination involving *IS2* on F and on the bacterial chromosome (5). The structure of a group of F-primes derived from AB312 suggest that this Hfr was formed by reciprocal recombination involving *IS2* (M. Guyer, personal communication). The points of origin and directions of transfer of many other Hfr's suggest that they were formed by reciprocal recombination between *IS2* or *IS3* on the main *E. coli* chromosome with *IS2* or *IS3* on F (2, 5). This view has received further support from direct observation of the structure of integrated F in the DNAs from several different Hfr strains (R. C. Deonier, personal communication). (However, other F-primes, F14 and F210, have structures indicating that their Hfr parents were formed by a recombination at some other position on F.)

Translocation of an *IS* sequence is *recA* independent and is probably a site-specific recombination involving the ends of the *IS* segment (12). Correspondingly, the reciprocal recombination involved in Hfr formation is probably not due to *recA*$^+$-directed homologous exchange at any point along the *IS* segment, but rather a *recA* independent site-specific recombination involving the ends.

There are approximately five copies of *IS2* on the chromosome of several strains of *E. coli* K-12 (9); we presume there are roughly comparable numbers of copies of *IS2* and of *IS3* on all strains of *E. coli* K-12. The frequency of *IS* insertions into a single gene has been estimated as about one per 10^7 cells (12). Consider a mutation for which there is a probability of P per cell division. Assume that $P \ll 1$ and that the back mutation rate is negligible. Starting with one bacterium, the numbers of mutated and wild-type bacteria after one cell division are $2P$ and $2(1 - P)$, respectively. In the second cycle of cell divisions, the $2P$ mutant cells give $4P$ mutant cells; whereas the $2(1 - P)$ wild-type cells give $4P(1 - P)$ mutants and $4(1 - P)^2$ wild-type cells. After n cycles of cell division, the number of wild-type bacteria is $2^n(1 - P)^n$; therefore the number of mutants is $2^n[1 - (1 - P)^n] \approx 2^n(nP)$. To generate 10^7 bacteria, $n = 23$. If one of these is mutant, we have $1 = 2^{23}(23P)$, or $P \approx 4 \times 10^{-9}$. Thus, we estimate the probability of appearance of an insertion as 4×10^{-9} per gene per generation.

If there is an approximately constant steady-state number of *IS* sequences in the bacterial genome, *IS* sequences must leave loci at approximately the same rate as they enter new loci. Let g be the number of genes, f the number of insertion sequences, r the probability per generation that an *IS* sequence leaves a particular position, and P the probability per generation that any *IS* sequence appears within a particular gene. Then, at the steady state, $P = rf/g$. We take the approximate values, $f = 10$, $g = 10^3$, and $P = 4 \times 10^{-9}$. The above equation then gives $r = 4 \times 10^{-7}$ per generation. Thus, on the average, an *IS* sequence has a residence time at any one locus in the genome of 2.5×10^6 generations.

We do not know the mechanisms of translocation of *IS* sequences, and the data for the value of P are very approximate. An important assumption in the calculation is that the probability of translocation of an *IS* sequence is uniform throughout the genome. This assumption may not be accurate. Nevertheless, the calculation is useful as an order of magnitude estimate for the value of the rate of migration of *IS* sequences. The result suggests that *IS* sequences migrate from place to place in the *E. coli* chromosome at a finite but very low rate. They therefore can function as relatively stable hot spots for Hfr formation. Curtiss pointed out in 1969 (1) that some of the major sublines of *E. coli* K-12 had not shared a common ancestor for 10,000 to 40,000 generations, and there was a real possibility of evolutionary divergence between such sublines. The translocation of *IS* sequences is possibly an example of such divergence.

ACKNOWLEDGMENTS

We again acknowledge the gift of the phage λp*lac*MS505 from M. Malamy. This research has been supported by Public Health Service grants GM 10991 from the National Institute of General Medical Sciences and AI-08619 (from the National Institute of Allergy and Infectious Diseases), GB-30 581 from the National Science Foundation, and VC-139 from the American Cancer Society. Kay Ptashne is the recipient of a postdoctoral research fellowship (SF02ES54436) from National Institute of Environmental Health Sciences.

LITERATURE CITED

1. Curtiss, R. 1969. Bacterial conjugation. Annu. Rev. Microbiol. **23**:69–136.
2. Davidson, N., R. C. Deonier, S. Hu, and E. Ohtsubo. 1975. The DNA sequence organization of F and F$^-$ primes and the sequences involved in Hfr formation. Microbiology **1**:56–65.

3. Fiandt, M., W. Szybalski, and M. H. Malamy. 1972. Polar mutations in *lac*, *gal* and phage λ consist of a few *IS*-DNA sequences inserted with either orientation. Mol. Gen. Genet. **119:**223-231.
4. Hirsch, H.-J., P. Starlinger, and P. Brachet. 1972. Two kinds of insertions in bacterial genes. Mol. Gen. Genet. **119:**191-206.
5. Hu, S., E. Ohtsubo, and N. Davidson. 1975. Electron microscope heteroduplex studies of sequence relations among plasmids of *E. coli*. XI. The structure of F13 and related F⁻ primes. J. Bacteriol. **122:**749-763.
6. Hu, S., E. Ohtsubo, N. Davidson, and H. Saedler. 1975. Electron microscope heteroduplex studies of sequence relations among bacterial plasmids. XII. Identification and mapping of the insertion sequences *IS1* and *IS2* in F and R plasmids. J. Bacteriol. **122:**764-775.
7. Malamy, M. H., M. Fiandt, and W. Szybalski. 1972. Electron microscopy of polar mutations in the *lac* operon of *Escherichia coli*. Mol. Gen. Genet. **119:**207-222.
8. Ptashne, K., and S. N. Cohen. 1975. Occurrence of *IS* regions on plasmid DNA and inverted nucleotide sequence duplications. J. Bacteriol. **122:**776-781.
9. Saedler, H., and B. Heiss. 1973. Multiple copies of the insertion DNA sequences *IS1* and *IS2* in the chromosome of *E. coli* K-12. Mol. Gen. Genet. **122:**267-277.
10. Saedler, H., H. J. Reif, S. Hu, and N. Davidson. 1974. *IS2*, a genetic element for turn-off and turn-on of gene activity in *E. coli*. Mol. Gen. Genet. **132:**265-289.
11. Sharp, P. A., M.-T. Hsu, E. Ohtsubo, and N. Davidson. 1972. Electron microscope heteroduplex studies of sequence relations among plasmids of *E. coli*. I. Structure of F-prime factors. J. Mol. Biol. **71:**471-497.
12. Starlinger, P., and H. Saedler. 1972. Insertion mutations in microorganisms. Biochimie **54:**177-185.