*J. Mol. Biol.* (1986) **187**, 157–167

# Structural and Functional Studies of Insertion Element IS*200*

**Stephen Lam† and John R. Roth**

*Biology Department*
*University of Utah, Salt Lake City, UT 84112, U.S.A.*

*(Received 21 September 1983, and in revised form 15 August 1985)*

The nucleotide sequence of the insertion element IS*200* has been determined partially, including the junctions between the element and the host chromosome at the insertion site. At most, two bases (A-A) are found repeated at the junctions and could be duplications of host sequences generated by the insertion of the element. No obvious sequence repeats, either direct or inverted, have been detected between the sequences just within the two ends of the element. The element is an extremely strong block to host transcription across the insertion site. A sequence similar to known transcription termination signals was found just within the element near the right end. Removal of less than 50 base-pairs at the right end of the element abolishes the transcription block. The putative terminator sequence is located within this 50 base-pair region. Genetic studies suggest that the element contains a promoter located more than 93 base-pairs from its left end. The proposed promoter and terminator are in proper orientation to form an internal transcription unit.

## 1. Introduction

Insertion sequences (IS) belong to the class of genetic elements that can move from one location to another within the bacterial genome. These elements can also mediate various chromosomal rearrangements (including inversion, deletion, and fusion of DNA segments) and alter the expression of adjacent genes (for reviews, see Starlinger, 1980; Calos & Miller, 1980; Kleckner, 1981; Shapiro, 1983). Many IS elements have been studied in *Escherichia coli*. The complete nucleotide sequences of IS*1*, IS*2*, IS*4* and IS*5* have been determined (Ohtsubo & Ohtsubo, 1978; Johnsrud, 1979; Ghosal *et al.*, 1979; Klaer *et al.*, 1981; Engler & van Bree, 1981; Schoner & Kahn, 1981). All of the elements have terminal, inverted repeat sequences. The inserted elements are flanked by short duplications of host DNA at the sites of insertion. In all cases, nonsense codons are present in all three translation reading frames, which may partially account for the strong polar effects of these elements. In addition, sequences resembling known transcription termination signals (Rosenberg & Court, 1979) have been identified in IS*2* and IS*4*.

Recently, a new IS element (IS*200*) was identified in *Salmonella typhimurium*. The element is specific to *Salmonella* and is present in almost all *Salmonella* species tested (Lam & Roth, 1983*a*). *S. typhimurium* strain LT2 has six copies of IS*200*, all of which have been mapped genetically (Lam & Roth, 1983*b*). The auxotrophic mutation *hisD984* in the histidine operon was shown to be the result of insertion of IS*200*. Like most other insertion mutations, *hisD984*::IS*200* is extremely polar; unlike other insertion mutations, however, it does not revert to prototrophy (less than $10^{-11}$; see Lam & Roth, 1983*a*). Deletion derivatives that relieve the strong polar effect have been isolated (Hoppe & Roth, unpublished results). Genetic mapping studies show that some of these derivatives appear to have one endpoint at the site of the *hisD984*::IS*200* mutation. These deletion derivatives have endpoints either within the 700 bp‡ of IS*200* or at the junctions between IS*200* and chromosomal material.

We have determined the nucleotide sequences at the junctions between the IS*200* element and *his* sequences in the mutant *hisD984*::IS*200*, as well as the endpoints of several non-polar deletion derivatives. In this paper, we describe these results and those of other structural and functional studies, which demonstrate that IS*200* contains transcription start and stop signals.

## 2. Materials and Methods

### (a) *Bacterial strains*

All bacterial strains used in these studies are listed in Table 1. The nomenclature for Tn*10* insertions has been

† Present address: CIBA-GEIGY, P.O. Box 12257, Research Triangle Park, NC 27709, U.S.A.

‡ Abbreviation used: bp, base-pair(s).

0022-2836/86/020157-11 $03.00/0

**Table 1**
*Bacterial strains*

| Strain | Genotype† | Source |
|---|---|---|
| TR3715 | *his-644 leu-1176*/F′ *hisC2385* | T. Kohno |
| TR5379 | *his-9569* | I. Hoppe |
| TR5384 | *his-9574* | I. Hoppe |
| TR5395 | *his-9585* | I. Hoppe |
| TR5397 | *his-9587* | I. Hoppe |
| TR5398 | *his-9588* | I. Hoppe |
| TR6238 | *hisD984* | P. E. Hartman |
| TR6286 | *his-2236*/F′ *hisC2385* | |
| TR6287 | *his-9585*/F′ *hisC2385* | |
| TR6288 | *his-9587*/F′ *hisC2385* | |
| TR6289 | *his-9588*/F′ *hisC2385* | |
| TT2144 | *zee-2*::Tn*10 his-9538* | I. Hoppe |
| TT2146 | *zee-2*::Tn*10 his-9540* | I. Hoppe |
| TT2148 | *zee-2*::Tn*10 his-9542* | I. Hoppe |
| TT2151 | *zee-2*::Tn*10 his-9545* | I. Hoppe |
| TT2152 | *zee-2*::Tn*10 his-9546* | I. Hoppe |
| TT3228 | *hisG8543*::Tn*10 hisG9647*::Tn*5* | D. Biek |
| TT6131 | *hisG9648*::Tn*5*(*tnf-39*) *recA1* | D. Biek |
| TT6783 | *metA22 trpB2 strA120 xyl-404 metE551 ilv-452 galE550 his-9539 zee-1*::Tn*10* $r_{LT}-m_{LT}+r_S-m_S+$/pWB91 | |
| TT6943 | *hisG9648*::Tn*5*(*tnf-39*) *hisD984*::IS*200* | |
| TT6958 | *metA22 trpB2 strA120 xyl-404 metE551 ilv-452 galE550 his-9539 zee-1*::Tn*10* $r_{LT}-m_{LT}+r_S-m_S+$/pWB91-*hisG9648*::Tn*5*(*tnf-39*) *hisD984*::IS*200* | |
| TT6945 | *his-9538*/F′ *hisC2385* | |
| TT6946 | *his-9542*/F′ *hisC2385* | |
| TT6947 | *hisG8543*::Tn*10 his-9585*/F′ *hisC2385* | |
| TT6948 | *hisG8543*::Tn*10 his-9587*/F′ *hisC2385* | |
| TT6949 | *hisG8543*::Tn*10 his-9588*/F′ *hisC2385* | |
| TT6950 | *hisG8543*::Tn*10 his-9538*/F′ *hisC2385* | |
| TT6951 | *hisG8543*::Tn*10 his-9542*/F′ *hisC2385* | |

† Nomenclature for Tn*10* insertions is as described by Chumley *et al.* (1979). Unless otherwise noted, all strains were constructed for this study.

described (Chumley *et al.*, 1979). Strains TT6943 and TT6958 are derived from Bruce Stocker and C. Colson.

### (b) *Recombinant phages and plasmids*

The recombinant phage M13Ho176 (Barnes, 1979) and plasmid pWB91 (Barnes, 1977) were kindly provided by W. Barnes. The phage M13Ho176 carries a 3300 bp insert in the M13 phage intergenic region; this insert contains the control region (*hisO*) and the first 2 structural genes (*hisG* and *hisD*) of the wild-type *S. typhimurium his* operon. This same portion of the *his* operon is present in the plasmid pWB91, a ColE1 derivative. The mutation *hisD984*::IS*200* was moved into M13Ho176 and pWB91 by homologous recombination. The resulting phage and plasmid are designated M13Ho176-*hisD984*::IS*200* and pWB91-*hisG*::Tn*5* *hisD984*::IS*200*, respectively. The construction of M13Ho176-*hisD984*::IS*200* has been described (Lam & Roth, 1983*a*); the construction of pWB91-*hisG*::Tn*5* *hisD984*::IS*200* is described below.

### (c) *Media, phage growth and transductional methods*

All media, phage growth and transductional methods have been described (Lam & Roth, 1983*b*).

### (d) *Preparation of DNA*

Single-stranded and double-stranded DNA of various phages and plasmids were prepared according to Barnes (1978*a*,*b*). Restriction fragments were isolated and recovered from gels by the method described by Maxam & Gilbert (1980).

### (e) *DNA sequencing*

DNA sequencing was carried out according to the methods of Sanger *et al.* (1977), Barnes (1978*b*) and Maxam & Gilbert (1980).

### (f) *Southern hybridization methods*

All materials and methods for Southern (1975) hybridization experiments have been described (Lam & Roth, 1983*a*,*b*).

### (g) *Construction of* pWB91-hisG::*Tn5* hisD984::*IS200*

The mutation *hisG9648*::Tn*5*(*tnf-39*) is a transposition-defective derivative of *hisG9648*::Tn*5* (Biek & Roth, 1981). Strains carrying this mutant insertion element in the *hisG* gene retain *hisD* gene expression, and the resulting strain is phenotypically $HisD^+$ (can grow on histidinol as the source of histidine). The double mutant *hisG9648*::Tn*5*(*tnf-39*) *hisD984*::IS*200* was constructed by transduction. P22 phage were grown on a strain (TT6131) containing the mutation *hisG9648*::Tn*5*(*tnf-39*) and used as donor to transduce a strain containing the mutation *hisD984*::IS*200* (TR6238) to kanamycin resistance. Kanamycin-resistant transductants that are $HisD^-$ contain both mutations. Such a transductant (TT6943) was then used as donor to transduce to kanamycin resistance a strain (TT6783) carrying a non-transducible *his* deletion in the chromosome and the plasmid pWB91, which includes the promoter-proximal portion of the wild-type *his* operon (including the *his* control region,

*hisG*, *hisD* and part of *hisC*). Since *hisG9648*::Tn*5*(*tnf-39*) is defective in transposition, kanamycin-resistant transductants arise only when the Tn*5* mutation recombines into the plasmid pWB91. Kanamycin-resistant transductants were grown in rich medium containing kanamycin to allow for segregation of non-transduced plasmids, and then scored for their HisD$^-$ phenotype. Transductants that were phenotypically HisD$^-$ had inherited IS*200* with the donor Tn*5* insertion.

(h) *Selection of non-polar derivatives of* hisD984::*IS200*

Since *hisD984*::IS*200* is absolutely polar, it is phenotypically HisD$^-$ (by direct damage) and HisC$^-$ because it blocks transcription. Genes located promoter-distal to *hisC*$^-$ function because of a promoter located between the *hisC* and *hisB* genes (Ciampi *et al.*, 1982). When *hisD984*::IS*200* acquires an F′*hisD*$^+$ *hisC*$^-$ plasmid, it becomes phenotypically HisD$^+$ (can grow on histidinol as a histidine source). The diploid remains *hisC*$^-$ and will not grow on minimal medium. Therefore, selection can be made for HisC$^+$ revertants. Among these are the chromosomal deletion mutants that affect IS*200* and are described in this paper.

(i) *Construction of strains containing* hisG::*Tn*10 *and non-polar deletion derivatives of* hisD984::*IS200*

Non-polar deletion derivatives of *hisD984*::IS*200* have restored transcription of the adjacent downstream *hisC* gene by removing the transcription stop signal at the right end of *hisD984*::IS*200*. Since the deletions were selected for *hisC* gene expression, all must retain the *hisC* gene. Two exceptions are deletions *his-9543* and *9545* (see Fig. 4), which express *hisC* despite removal of a small, apparently dispensable, part of the *hisC* gene. To examine whether the transcript expressing *hisC* starts at the *his* promoter or within *hisD984*::IS*200*, a polar *hisG*::Tn*10* mutation was introduced into the upstream *hisG* gene of strains containing such deletions. The resulting strains were examined for their HisC phenotype. If transcription was solely from the *his* promoter, the strains should become HisC$^-$ when they receive the polar *hisG* insert. The HisC phenotype was scored as the ability to complement an F′ plasmid containing the *E. coli his* operon with a *hisC*$^-$ mutation. To allow for scoring of HisC phenotype, the non-polar deletion derivatives of *hisD984*::IS*200* were first moved into strains carrying *hisC*$^-$ F′ plasmids, the *hisG*::Tn*10* mutation was then introduced into the resulting strains. Details are outlined below.

The strains TR3715 and TR6286 contain the F′*hisC2385* plasmid and the deletions *his-644* and *his-2236*, respectively, in the chromosome. Both deletions *his-644* and *his-2236* are *hisD*$^-$ *hisC*$^-$. As a result, the strains TR3715 and TR6286 are phenotypically His$^-$. These strains were used as recipients in transductional crosses, with P22 phages grown on strains containing non-polar (HisC$^+$) deletion derivatives of *hisD984*::IS*200* as donors. Transductants that were phenotypically His$^+$ had inherited the non-polar deletion derivatives in the chromosome. Lack of homology prevents recombination between fragments of the *Salmonella* chromosome and the *E. coli his* region present on the plasmid.

Strain TT3228 carries a Tn*10* and a Tn*5* mutation in the *hisG* gene. The Tn*10* mutation is located closer to the *his* promoter. This strain was used as donor to transduce the strains described in the last paragraph to tetracycline-resistance. Tetracycline-resistant transductants that are kanamycin sensitive have incorporated the Tn*10* mutation but not the Tn*5* mutation, nor any donor genetic material promoter-distal to the Tn*5* mutation and are presumably the double mutants desired. Their genotypes were verified by standard transductional mapping crosses.

## 3. Results

(a) *Partial sequence determination of* hisD984::*IS200*

The mutation *hisD984*::IS*200* was moved, by homologous recombination, into two small DNA molecules amenable to sequence analysis: the single-stranded phage M13Ho176 (Barnes, 1979; see Lam & Roth, 1983*a*) and the plasmid pWB91 (Barnes, 1977; see Materials and Methods). M13Ho176 and pWB91 contain the same portion of wild-type *S. typhimurium his* operon, which includes the *his* control region (*hisO*) and the first two structural genes, *hisG* and *hisD*. The resulting phage and plasmid are designated M13Ho176-*hisD984*::IS*200* and pWB91-*hisG*::Tn*5* *hisD984*::IS*200*, respectively.

The approximate location of the mutation *hisD984*::IS*200* within the *hisD* gene was determined from the following observations. Genetic mapping studies (Hoppe & Roth, personal communication) indicated that *hisD984* is located between the mutations *hisD6404* and *hisD2578*. The mutation *hisD6404* has been shown by Bossi & Roth (1980) to be a C to T base substitution at the 49th nucleotide from the beginning of the *hisD* gene. The mutation *hisD2578* has been shown by Isono & Yourno (1974) to affect a run of C residues, which we now know starts at the 149th nucleotide from the beginning of *hisD*. Thus, the site of insertion of the mutation *hisD984*::IS*200* is located between the 49th and 149th nucleotides from the beginning of *hisD*.

Figure 1 shows a restriction map of the region, including the insertion. The nucleotide sequences of the left and right junctions between the insertion mutation and adjacent *his* material were determined by the chain-termination method of Sanger *et al.* (1977) and Barnes (1978*b*). The template used was single-stranded DNA isolated from M13Ho176-*hisD984*::IS*200*. The primer used for obtaining the left junction sequences was pWB91 DNA digested with the enzyme *Pvu*II. The *Pvu*II site in *hisD* (see Fig. 1) is the only *Pvu*II site on pWB91 (W. Barnes, personal communication). The primer used for obtaining the right junction sequences was the *Eco*RI-*Hin*dIII fragment internal to IS*200* (see Fig. 1), isolated from replicative form DNA of M13Ho176-*hisD984*::IS*200*. Autoradiograms showing the junction sequences are presented in Figure 2.

Additional IS*200* sequence near the right junction was determined by the chemical cleavage method of Maxam & Gilbert (1980), using the *Hin*dIII fragment that contains the right junction.

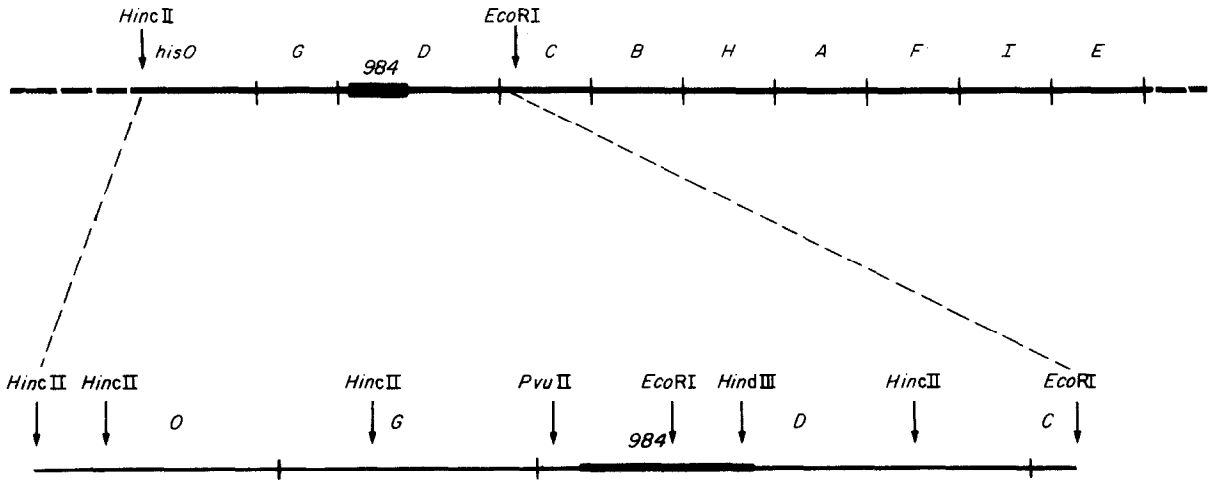


**Figure 1.** Restriction map of the cloned region of the *S. typhimurium his* operon. The top line shows a genetic map of the entire *his* operon. The direction of transcription is from left to right. The bottom line shows the region of the operon that is present in M13Hol6-*hisD984*::IS*200*, with the locations of relevant restriction sites indicated above the line.

The fragment was isolated from pWB91-*hisG*::Tn*5* *hisD984*::IS*200* plasmid DNA. Combined sequence data of the two junctions, together with the corresponding wild-type *his* sequences (W. Barnes, personal communication), are shown in Figure 3.

The following structural features are apparent: (1) no *his* sequence is missing; (2) there is a possible duplication of, at most a 2 bp sequence (A-A) at the junctions; (3) no obvious inverted terminal repeats are observed within the ends of the element; and (4)

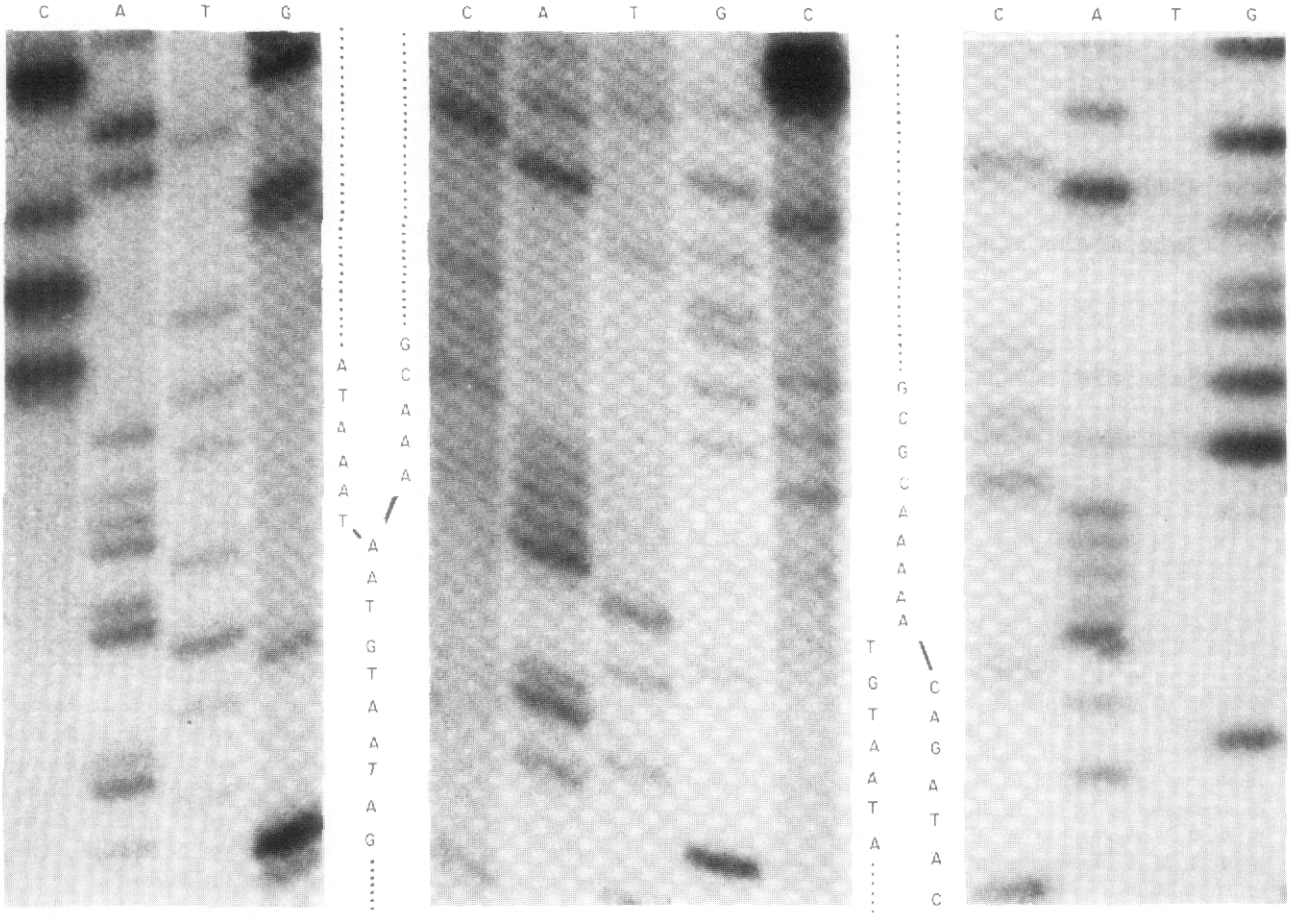


**Figure 2.** Autoradiograms of sequencing gels showing the junctions between *hisD984*::IS*200* and *his* sequences. The left panel shows the left junction sequence, the middle panel shows wild-type *his* sequence, and the right panel shows the right junction sequence. Sequencing was done according to the methods of Sanger *et al.* (1977) and Barnes (1979).

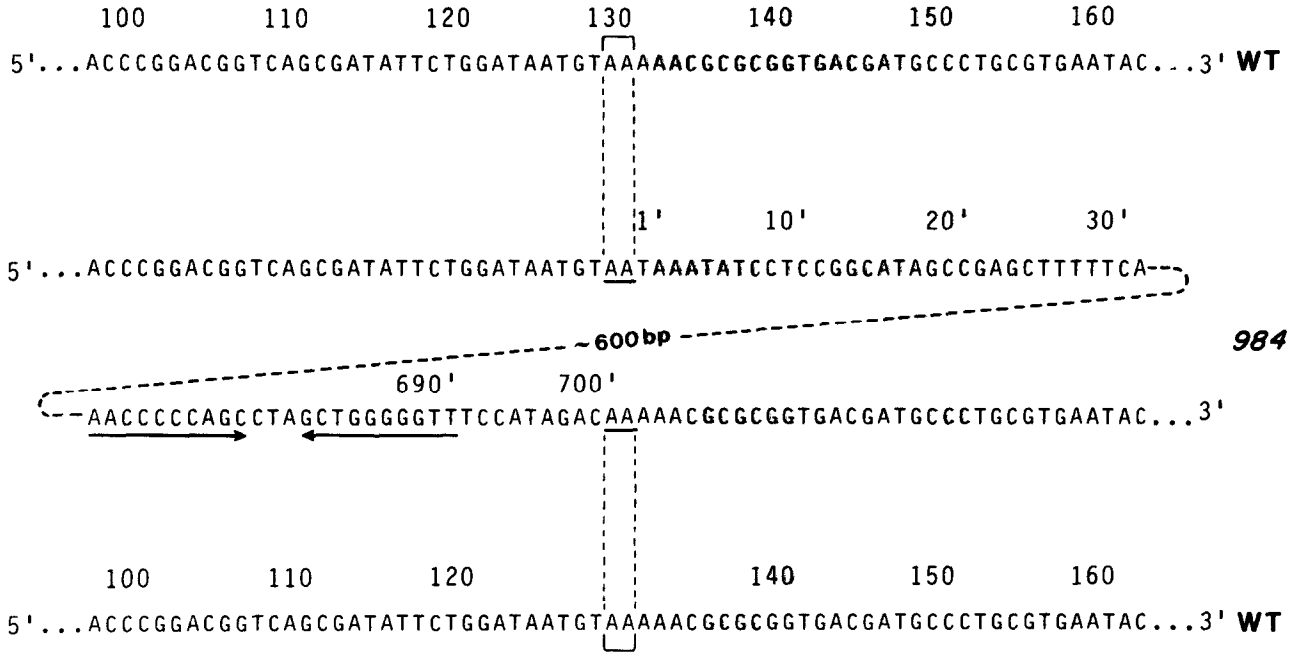


**Figure 3.** Nucleotide sequences at the junctions of the insertion mutation *hisD984*::IS*200*. Only one strand of the DNA is shown. The top and bottom lines show wild-type sequences. The middle 2 lines show the sequences at the 2 junctions. Numbering of *his* sequence starts at the beginning of the *hisD* gene. Numbering of IS*200* sequence starts at the left end of *hisD984*::IS*200*, and uses numbers followed by primes (1′, 10′, etc.). For the purpose of numbering, the length of the inserted sequence is arbitrarily taken to be 700 bp. The arrows indicate sequences that can potentially form a structure resembling known transcription termination signals (see Fig. 7). WT, wild-type. Sequence hyphens are omitted for clarity.

a sequence near the right end of the insertion can potentially form a structure resembling *rho*-independent transcription terminators (Rosenberg & Court, 1979). This hypothetical structure is discussed below.

### (b) *A possible terminator near the right end of* hisD984::*IS*200

Insertion mutant *hisD984*::IS*200* blocks transcription of the adjacent promoter-distal *hisC* gene. Therefore, IS*200* must include a transcription termination signal. Several sponteneous deletion derivatives of IS*200* have been selected that have regained *hisC* expression (see Fig. 4). This selection method is described in Materials and Methods. Each of these deletions must have removed the region(s) of IS*200* that is responsible for the polar effect. By analyzing the extent of such deletions, it was possible to infer the approximate location of the transcription termination site. The site removed by all non-polar deletion mutants lies within 50 bp of the right end of IS*200*. Examination of the DNA sequence in this region revealed a likely candidate for the termination signal.

The IS*200* element contains a *Hin*dIII site near its right end. Restriction fragments ending at this site had been used to obtain DNA sequences at the right end of the element. Although the exact DNA sequence near the site was not determined, the approximate distance between the site and the right end of the element was estimated, from the spacing on the gels to be less than 50 bp. Several non-polar deletion derivatives entering IS*200* from the right (see Fig. 4) were examined by Southern hybridization experiments, to see if they have removed this *Hin*dIII site. The results are shown in Figure 5. In these experiments, the DNA was digested with *Hin*dIII, and the probe used was M13Hol6. Lane 1 contains DNA from the wild-type strain LT2, lanes 2 to 8 contain DNA from the deletion derivatives, and lane 9 contains DNA from the parent strain carrying the mutation *hisD984*::IS*200* (TR6238). *Hin*dIII does not cut within the cloned portion of

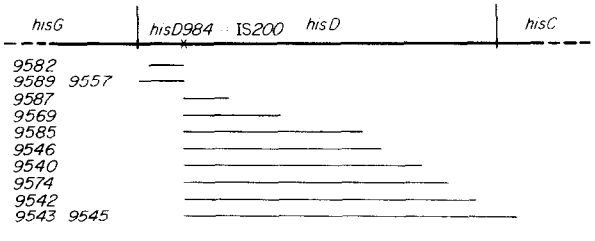


**Figure 4.** Deletion derivatives of the mutation *hisD984*::IS*200*. The top line shows the first 3 structural genes of the *S. typhimurium his* operon. These deletions were selected for loss of the polar effect of *hisD984*::IS*200* on the *hisC* gene. All deletions, even those entering the *hisC* gene, retain *hisC* function. Deletions presented appear to have 1 endpoint either within or adjacent to IS*200*. Other deletions (not shown) removed the entire insertion element.

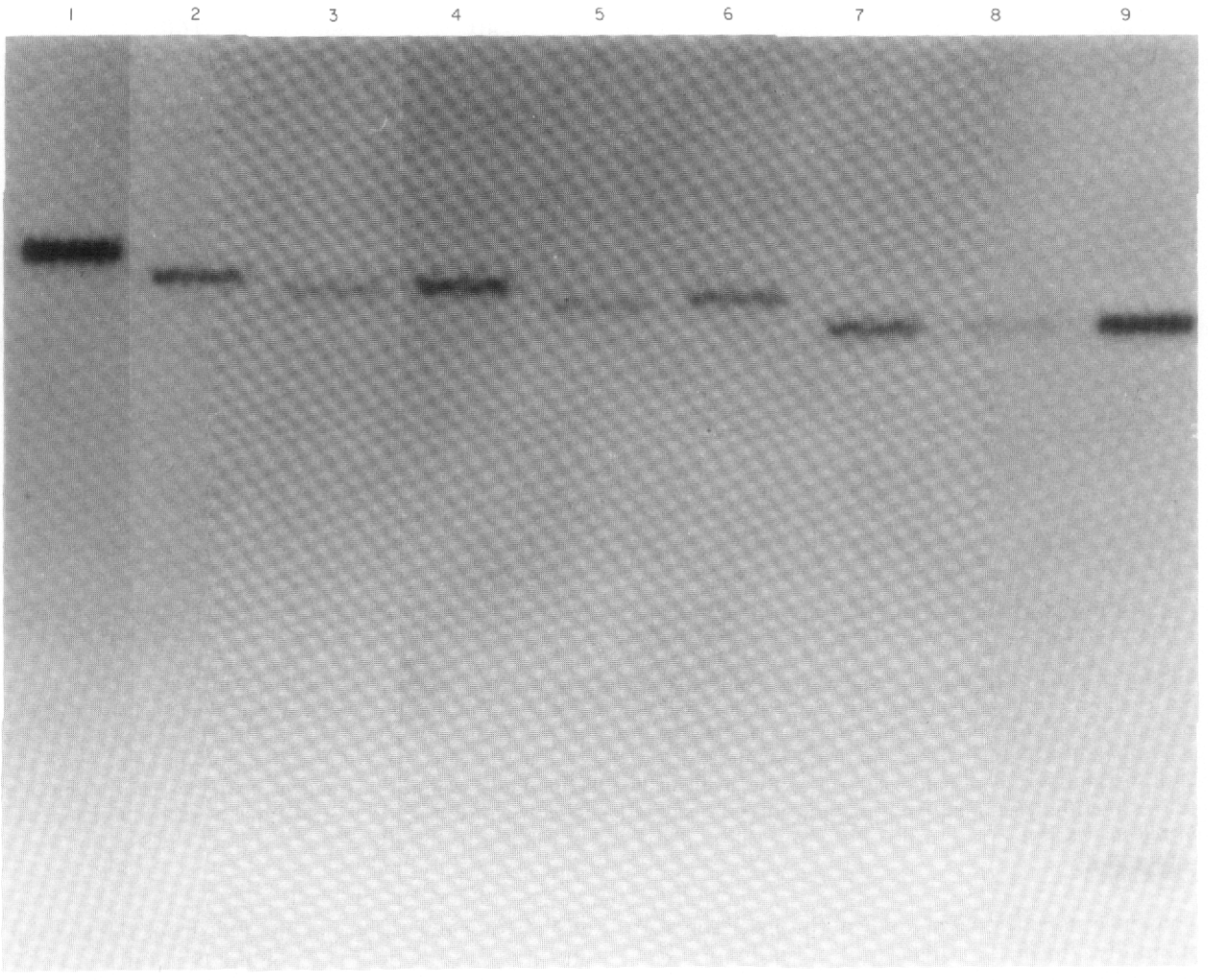


**Figure 5.** Southern blot analysis of deletion derivatives of *hisD984*::IS*200*. DNA from each strain was digested with *Hin*dIII; the probe used was M13Hol6. Details are described in the text. Lane 1 contains DNA from the wild-type strain LT2, lane 9 contains DNA from the parent strain carrying the mutation *hisD984*::IS*200* (TR6238), and lanes 2 to 8 contain DNA from strains carrying the following non-polar deletion derivatives of *hisD984*::IS*200*; (2) *his-9569* (TR5379); (3) *his-9585* (TR5395); (4) *his-9546* (TT2152); (5) *his-9540* (TT2146); (6) *his-9574* (TR5384); (7) *his-9542* (TT2148); and (8) *his-9545* (TT2151).

the wild-type *his* operon present in M13Hol76. The wild-type strain LT2 therefore yielded only one hybridized fragment. The strain containing *hisD984*::IS*200* (TR6238) yielded two hybridized fragments, because of the *Hin*dIII site within IS*200*. The larger fragment contains the *his* control region *hisO*, *hisG* and part of *hisD*; the smaller fragment (which runs much faster and is barely visible in Fig. 5) contains part of *hisD* and part of *hisC* (see Fig. 5). Each of the strains containing, respectively, the deletions *his-9569* (lane 2), *his-9585* (lane 3), *his-9546* (lane 4), *his-9540* (lane 5), and *his-9574* (lane 6) yielded one hybridized fragment with a molecular weight differing from that of the larger fragment in the parent strain carrying *hisD984*::IS*200* (TR6238; lane 9), suggesting that these deletions have removed the *Hin*dIII site in IS*200* and have fused the two fragments in the parent strain (TR6238). The deletions *his-9542* (lane 7) and *his-9545* (lane 8) yielded one hybridized fragment with the same molecular weight as the large fragment in the parent strain (TR6238), suggesting that these two deletions, which enter the element from the right have not removed the *Hin*dIII site but have removed essentially all of the second fragment. An alternative possibility is that these deletions have removed the *Hin*dIII site, but the fusion fragment is fortuitously of the same size as the larger fragment in the parent strain (TR6238). This possibility is eliminated by the Southern blot experiments presented below.

Further Southern hybridization analyses of these deletion derivatives indicate that the deletion *his-9542* does not remove the *Hin*dIII site. The results of these analyses are shown in Figure 6. The top half of Figure 6 shows an autoradiogram from Southern hybridization experiments designed to define the left endpoints of the two deletions *his-9542* and *his-9545*. The bottom half shows the interpretation of the data. Four different strains were examined in two sets of experiments: the

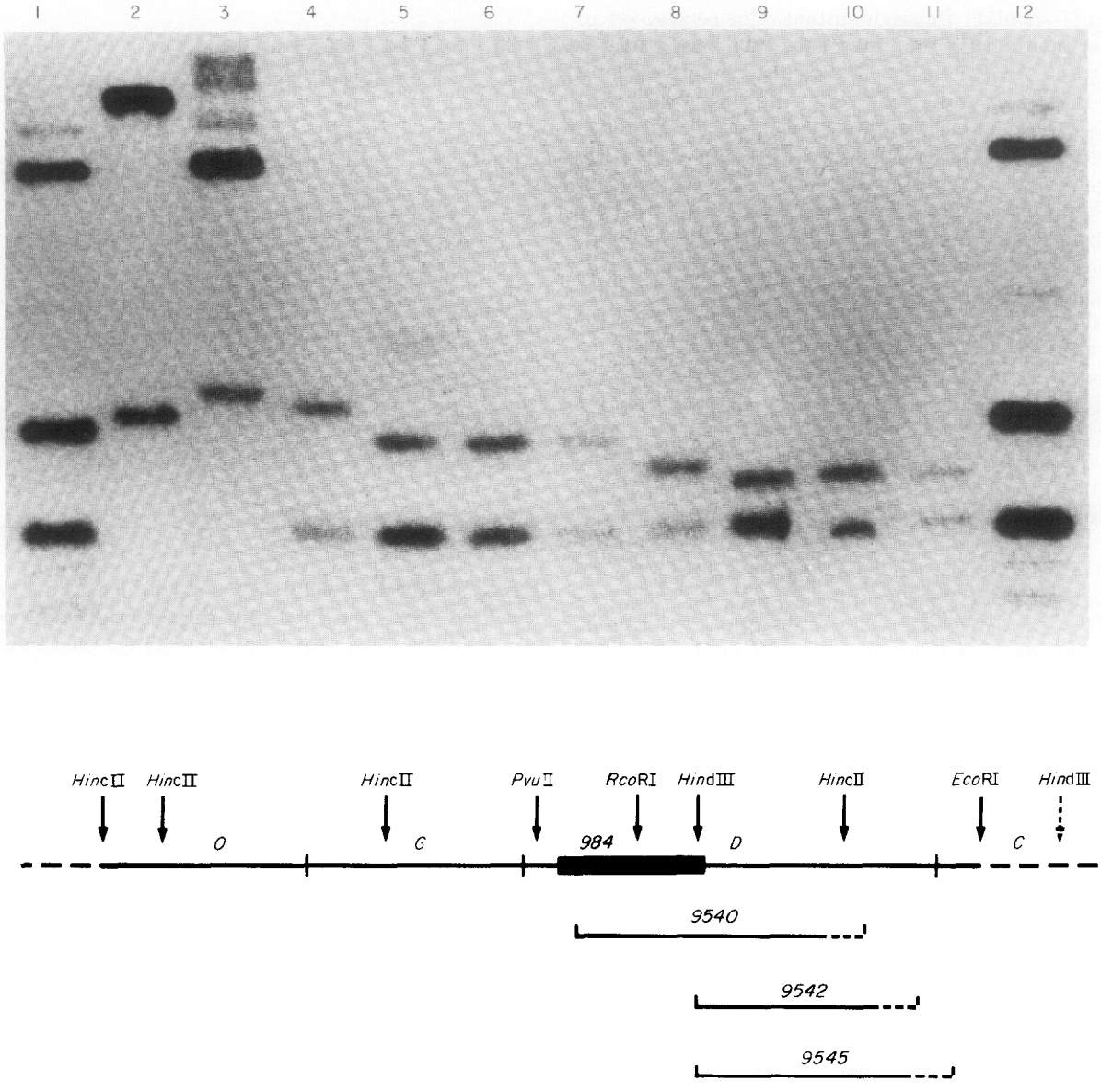


**Figure 6.** Southern blot analysis of deletion derivatives of *hisD984*::IS*200*. Top: autoradiogram of Southern blot. The probe used was M13Hol6, which includes material indicated by the unbroken horizontal line in the map, but without IS*200*. Two sets of double digest experiments are shown: in the first set (lanes 4 to 7), DNA was digested with *Hin*cII and *Hin*dIII; in the second set (lanes 8 to 11), DNA was digested with *Hin*cII and *Eco*RI. Four strains were analyzed in each set of experiments: the parent strain containing *hisD984*::IS*200* (TR6238; lanes 5 and 9) and 3 deletion derivatives; *his-9540* (TT2146; lanes 4 and 8), *his-9542* (TT2148; lanes 6 and 10) and *his-9545* (TT2151; lanes 7 and 11). Lanes 2 and 3 contain DNA from the parent strain (TR6238) digested with *Eco*RI and *Hin*dIII, respectively. Lanes 1 and 12 contain the molecular weight standards: DNA from the plasmid pWB91, digested with *Hin*cII. This plasmid contains the promoter-proximal portion of the *his* operon. Bottom: restriction map of the region and interpretation of the hybridization data.

parent strain (containing *hisD984*::IS*200*) and three deletion derivatives, *his-9540*, *his-9542* and *his-9545*. The deletion *his-9540* has been sequenced partially (discussed below). Although the exact junction has not been determined, it is known that the deletion enters IS*200* from the right leaving approximately 70 bp at the left end of *hisD984*::IS*200* and has thus removed both the *Eco*RI and the *Hin*dIII sites internal to the IS*200* element.

In the first set of experiments (Fig. 6, lanes 4 to 7), DNA from each strain was digested by the two enzymes *Hin*cII and *Hin*dIII. The larger fragment in each lane (4 to 7) corresponds to the *Hin*cII-*Hin*dIII fragment, including part of *hisG* and most of the IS element. It can be seen that for *his-9540* (lane 4), this fragment size is different from that of the other strains (lanes 5, 6 and 7). This is expected, since the deletion has removed the *Hin*dIII site, thus joining the fragments on either side of it. The other two deletion strains (lanes 6 and 7) yielded fragments with the same size as the parent (lane 5). The simple interpretation is that the deletions end to the right of the *Hin*dIII site, thus leaving the

*Hinc*II-*Hin*dIII fragment intact. The second set of experiments (Fig. 6, lanes 8 to 11) demonstrates that the *Eco*RI site of IS*200* also is removed by deletion *9540* and retained by the other deletions. This set of experiments is identical to the first, except that the DNA was digested with the enzymes *Hinc*II and *Eco*RI. The larger fragment in each lane (8 to 11) corresponds to the *Hinc*II-*Eco*RI fragment spanning *hisG* and part of the IS element. Deletion mutant *his-9540* (lane 8) yielded a large fragment, different in size from that of the other strains, while the other two deletion strains (lanes 10 and 11) yielded fragments of identical size to those of the parent (lane 9). Thus deletions *his-9542* and *his-9545* leave both the *Eco*RI and the *Hin*dIII sites of IS*200*. In both of the above digests, extra bands are predicted that are not apparent in the Figure. Several of these fragments comigrate as the smaller band. (Note the differences in intensity of the band in various deletions.) The other fragments are much smaller and blot so poorly they are not apparent in the photograph.

The above conclusion is strengthened because one can predict that a larger *Hinc*II-*Hin*dIII fragment would have been detectable had these deletions removed the *Hin*dIII site. Genetic mapping data (Hoppe & Roth, personal communication) indicate that the right endpoint of *his-9542* is within *hisD* and, therefore, retains the *Eco*RI site in *hisC*. Deletion *his-9545*, though entering the *hisC* gene, expresses *hisC* function. It must have removed only a small, apparently dispensable, portion of the *hisC* gene and is also likely to have retained the *Eco*RI site in *hisC*. The distance from *Eco*RI to *Hin*dIII in *hisC* is detectably larger than the distance from *Eco*RI to *Hin*dIII in IS*200*. (This is demonstrated by the difference in the smallest bands in lanes 2 and 3 in Fig. 6.) Knowing the above facts, one can predict that if deletions *9542* or *9545* ended between the *Eco*RI and the *Hin*dIII sites of IS*200*, the *Hinc*II-*Hin*dIII fragment generated would have been considerably larger than the parent fragment. This was not the case (lanes 5 to 7 in Fig. 6).

Thus, the deletion *his-9542*, which abolished the strong polar effect of the mutation *hisD984*::IS*200*, has removed less than 50 bp of the right end of the element. The conclusion is, therefore, that a strong polar signal is encoded within 50 bp of the right end of *hisD984*::IS*200*. This sequence was scanned for potential termination sequences. The most promising candidate is presented in Figure 7. All deletions that destroy the transcription blocking ability of IS*200* remove this sequence. The sequence has a strong resemblance to known transcription termination sites (Rosenberg & Court, 1979; Platt & Bear, 1983). Therefore, we feel that the sequence is likely to serve as a transcription block within IS*200*.

The DNA sequences of the deletion derivatives *his-9540*, *his-9557*, *his9582* and *his-9587* have been determined. The sequence data obtained support the above conclusion. The deletions were crossed into M13Ho176 by homologous recombination

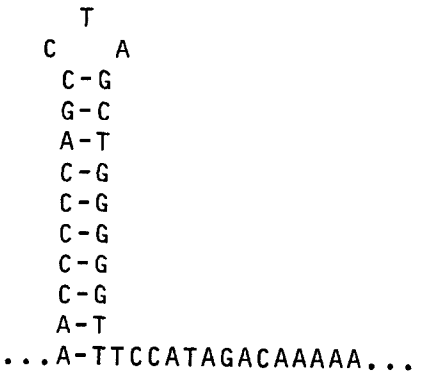


**Figure 7.** Potential transcription terminator structure near right end of *hisD984*::IS*200*. This structure was removed by all deletions analyzed that removed the transcription blocking ability of IS*200*. Sequence hyphens have been omitted but base-pairing is indicated.

(Lam & Roth, 1983*a*). Single-stranded DNA was isolated from the resulting phages and sequences were obtained by the chain-termination methods of Sanger *et al.* (1977) and Barnes (1978*b*). The primer used in these experiments was pWB91 DNA digested with the enzyme *Pvu*II. The extents of the deletions are shown in Figure 8. All of the deletions end within the IS*200* element, leaving behind different amounts of the element. The nucleotide sequences at the joint points of the deletions are shown in Figure 9. The DNA sequences of deletion *his-9540* has been determined only partially. The exact sequence of an approximately 10 bp region surrounding the joint point was not determined. This uncertainty is indicated by brackets in Figure 9. It can be seen that the two deletions entering the IS*200* element from the left, *his-9582* and *his-9557*, have removed practically all of the element, leaving only 5 bp and 11 bp, respectively, at the right end, whereas the two deletions entering the element from the right, *his-9587* and *his-9540*, have left 93 bp and approximately 70 bp, respectively, at the left end of the element. (We

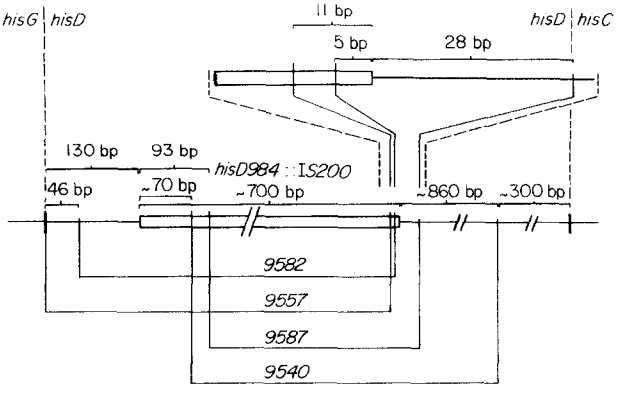


**Figure 8.** Extent and general location of 4 non-polar deletion derivatives of *hisD984*::IS*200*, *his-9582*, *his-9587* and *his-9540*.

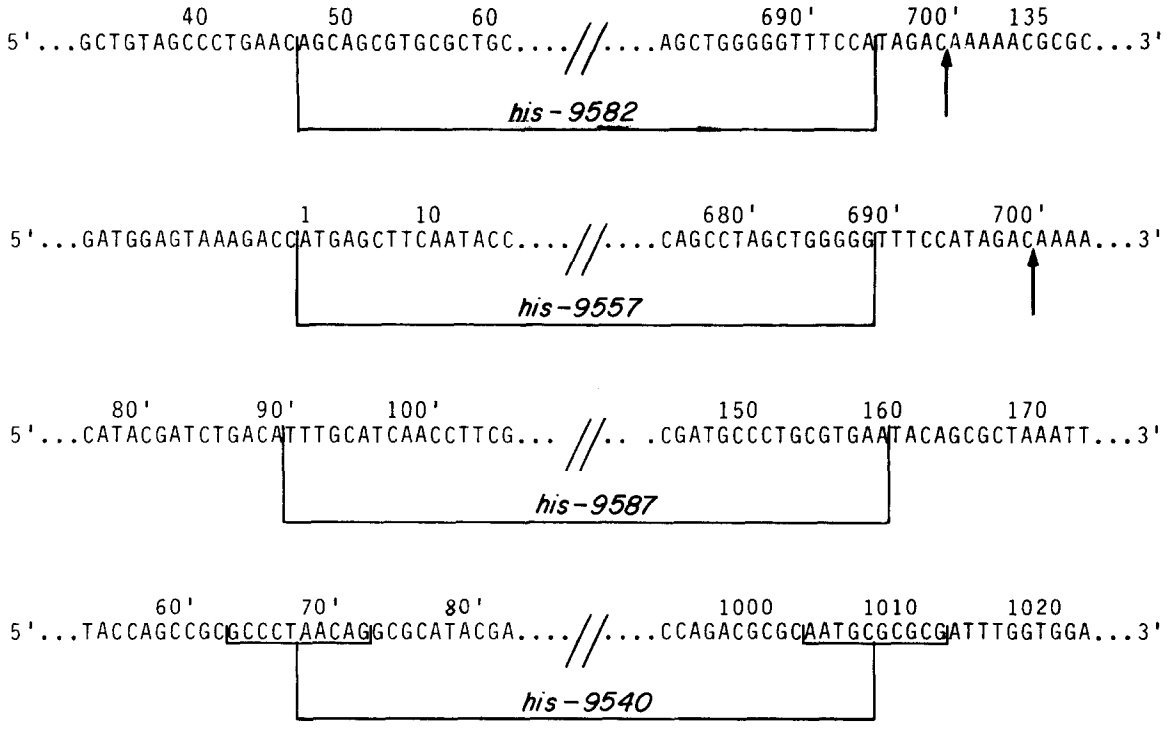


**Figure 9.** Nucleotide sequences at the joint points of 4 non-polar deletion derivatives of *hisD984*::IS*200*: *his-9582*, *his-9557*, *his-9587* and *his-9540*. Only one strand of the DNA is shown. The numbering system is as described in the legend to Fig. 3. The vertical arrows indicate the junction between *his* and IS*200* sequences. In deletion *his-9587*, sequence A-T on both sides of the joint point makes it impossible to identify the actual site of the deletion; only one possibility is shown. The DNA sequence of *his-9540* has been determined only partially. The exact sequence of an approximately 10 bp region surrounding the joint point was not determined. This uncertainty is indicated by brackets in the Figure. Sequence hyphens are omitted for clarity.

have described above two additional non-polar deletions that remove less than 50 bp from the right end of IS*200*.) These results are expected if the termination signal removed by these deletions is located near the right end of the element. All four of the deletions presented in Figure 9 have removed the potential stem-and-loop structure seen in the DNA sequence of the region (see Fig. 7). Thus these four deletions and the two deletions that attracted our attention to this end of the element carry deletions in or near this region of IS*200*.

### (c) *IS*200 *includes a promoter function*

The deletion derivatives of *hisD984*::IS*200* described have apparently restored *hisC* gene expression by removing a transcription stop signal at the right end of *hisD984*::IS*200*. This transcript could have started at the *his* promoter and extended across the remaining portion of *hisD984*::IS*200* and into *hisC*. Alternatively, the transcript could start within the IS*200* element and extend across the right end of the element into *hisC*. To distinguish between the two possibilities, a *hisG*::Tn*10* mutation was introduced into a number of strains carrying different non-polar derivatives of *hisD984*::IS*200* (see Materials and Methods). These Tn*10* insertions in *hisG* are absolutely polar (Ciampi *et al.*, 1982). The *hisG*::Tn*10* mutation prevents any transcription initiated at the *his* promoter from reaching the *hisD* and *hisC* genes. Strains containing both the *hisG*::Tn*10* insert and the non-polar derivative of IS*200* were tested for their ability to express the *hisC* gene. (If the strains are $HisC^+$ they will complement an F′ *hisC*$^-$ mutation.) If transcription originates only at the *his* promoter, the strains should become $HisC^-$; if transcription is initiated within *hisD984*::IS*200*, the strains should remain $HisC^+$.

Five derivatives were tested. The deletions *his-9585* (TR5395), *his-9587* (TR5397), and *his9542* (TT2148) have removed *his* sequences as well as IS*200* sequences (see Fig. 4). For the other two derivatives, *his-9588* (TR5388) and *his-9538* (TT2144), no deletions of *his* material was detected in genetic mapping crosses (Hoppe & Roth, personal communication); presumably, these derivatives harbor changes internal to the IS element that affect the terminator structure. Introduction of the *hisG*::Tn*10* mutation into the strains containing *his-9585* and *his-9587* rendered these strains $HisC^-$, indicating that transcription was initiated at the *his* promoter only; presumably these deletions remove the promoter within IS*200*. The other three strains (*9542*, *9588* and *9538*) remained $HisC^+$, indicating that transcription was initiated within *hisD984*::IS*200* sequences.

## 4. Discussion

The *Salmonella*-specific transposable element, IS*200*, shows several unusual features.

(1) Although insertion of the element did not remove any *his* sequences, the insertion mutation does not revert by precise excision. Most transposable elements studied do excise precisely at detectable frequencies (for a review, see Calos & Miller, 1980). The only other exceptions are Tn*1* (Weinstock & Botstein, 1979; Hernalsteens *et al.*, 1977) and Mu (Bukhari *et al.*, 1977). The failure of the IS*200* insertion to revert may be a consequence of the lack of repeated sequences within the element or due to the minimal duplication of host sequences flanking the insertion (Albertini *et al.*, 1982). In other elements, ability to excise precisely has been attributed to both of these structures (for a review, see Kleckner, 1981).

(2) At most, 2 bp of *his* sequences may have been duplicated at the ends of the element. Since this is the only insertion site studied, the actual ends of the element cannot be determined. All other transposable elements studied have been found to generate more extensive duplication of host sequences at the insertion sites (see Calos & Miller, 1980). The extent of this duplication differs among elements, but the majority generate either a 9 bp or a 5 bp duplication. There does not seem to be any correlation between the length of duplication and excision frequency. The element IS*4*, which generates an 11 bp or 12 bp repeat (Haberman *et al.*, 1979), excises precisely at a frequency of 1 per $10^9$ cells (Pfeifer, cited by Starlinger, 1980), whereas IS*1*, which generates a 9 bp repeat, reverts at a frequency as high as $10^{-6}$ per cell plated (Jordan *et al.*, 1968).

(3) Comparison of the sequences at the ends of IS*200* revealed no obvious relationship between them. There is no hint of an inverted or a direct repeat, which is a common feature among transposable elements. It is interesting to note that phage Mu, which also does not revert spontaneously, is the only other transposable element without obvious inverted repeats.

Southern hybridization studies and sequence data suggest that a transcription termination signal is located at the right end of the IS*200* element. The element is strongly polar. A non-polar deletion derivative that enters the element from the right was shown to have removed less than 50 bp at the right end of the element. The two derivatives tested that enter the element from the left were also both shown to have removed IS*200* material from this 50 bp region. Examination of the DNA sequence in this region revealed a sequence similar to known transcription termination signals (Rosenberg & Court, 1979). It seems likely that this structure is responsible for the transcription block caused by the insertion.

The DNA sequence of four deletion derivatives that end within the IS*200* element have been determined. All have very different endpoints, suggesting that they are spontaneous deletions that were probably not formed by events catalyzed by the transposition activity of the element.

Genetic studies of the derivatives of *hisD984*::IS*200* that express the downstream *hisC* gene indicate that, in some of these derivatives, transcription initiates within IS*200* sequences. Of the three mutants that show this property, *his-9538* and *his-9588* have not been studied physically. Genetic mapping studies indicate that the mutations involved have not removed any *his* sequences adjoining the insertion (Hoppe & Roth, personal communication). The changes in *hisD984*::IS*200* sequences in these mutations may be localized near the region of the terminator structure. The third mutation showing transcription from within IS*200*, *his-9542*, has been shown by Southern hybridization studies to have removed less than 50 bp at the right end of *hisD984*::IS*200*. Thus, most of the element remained intact in the strain carrying this mutation. The mutation *his-9587*, which showed no transcription initiated from within *hisD984*::IS*200* sequences, has removed all but 83 bp of *hisD984*::IS*200* sequences at the left end. These data are consistent with the hypothesis that *hisD984*::IS*200* contains a promoter located more than 93 bp from its left end. The proposed promoter and terminator are in proper orientation to form a transcription unit within IS*200*.

## References


Albertini, A. M. Hofer, M., Calos, M. P. & Miller, J. H. (1982). *Cell*, **29**, 319–328.

Barnes, W. M. (1977). *Science*, **195**, 393–394.

Barnes, W. M. (1978*a*). *J. Mol. Biol.* **119**, 83–99.

Barnes, W. M. (1978*b*). *Proc. Nat. Acad. Sci., U.S.A.* **95**, 4281–4285.

Barnes, W. M. (1979). *Gene*, **5**, 127–139.

Biek, D. & Roth, J. R. (1981). *Cold Spring Harbor Symp. Quant Biol.* **45**, 189–191.

Bossi, L. & Roth, J. R. (1980). *Nature (London)*, **286**, 123–137.

Bukhari, A. I., Ljungquist, E., de Bruijn, F. & Khatoon, H. (1977). In *DNA Insertion Elements, Plasmids and Episomes* (Bukhari, A. I., Shapiro, J. A. & Adhya, S. L., eds), pp. 249–261, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Calos, M. P. & Miller, J. H. (1980). *Cell*, **20**, 579–595.

Chumley, F. G., Menzel, R. & Roth, J. (1979). *Genetics*, **91**, 639–655.

Ciampi M. S., Schmid, M. & Roth, J. R. (1982). *Proc. Nat. Acad. Sci., U.S.A.* **79**, 5016–5020.

Engler, J. A. & van Bree, M. P. (1981). *Gene*, **14**, 155–163.

Ghosal, D., Sommer, H. & Saedler, H. (1979). *Nucl. Acids Res.* **6**, 1111–1112.

Haberman, P., Klaer, K., Kuhn, S. & Starlinger, P. (1979). *Mol. Gen. Genet.* **175**, 369–373.

Hernalsteens, J. P., Villarroel-Mandiola, R., Van Montagu, M. & Schell, J. (1977). In *DNA Insertion Elements, Plasmids and Episomes* (Bukhari, A. I., Shapiro, J. A. & Adhya, S. L., eds), pp. 179–183, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Isono, S. & Yourno, J. (1974). *J. Mol. Biol.* **82**, 355–360.

Johnsrud, L. (1979). *Mol. Gen. Genet.* **169**, 213–218.
Jordan, E., Saedler, H. & Starlinger, P. (1968). *Mol. Gen. Genet.* **102**, 353–363.
Klaer, R., Kuhn, S., Tillman, E., Fritz, H.-J. & Starlinger, P. (1981). *Mol. Gen. Genet.* **181**, 169–175.
Kleckner, N. (1981). *Annu. Rev. Genet.* **15**, 341–404.
Lam, S. & Roth, J. R. (1983*a*). *Cell* **34**, 951–960.
Lam, S. & Roth, J. R. (1983*b*). *Genetics* **105**, 801–811.
Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
Ohtsubo, H. & Ohtsubo, E. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 615–619.
Platt, T. & Bear, D. G. (1983). In *Gene Function in Prokaryotes* (Beckwith, J., Davis, J. & Gallant, J., eds), pp. 123–161, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
Rosenberg, M. & Court, D. (1979). *Annu. Rev. Genet.* **13**, 319–354.
Sanger, F., Nicklen, S. & Coulson, A. R. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 5463–5467.
Schoner, B. & Kahn, M. (1981). *Gene*, **14**, 165–174.
Shapiro, J. A. (1983) Editor of *Mobile Genetic Elements*, Academic Press, New York.
Southern, E. M. (1975). *J. Mol. Biol.* **98**, 503–517.
Starlinger, P. (1980). *Plasmid*, **3**, 241–259.
Weinstock, G. M. & Botstein, D. (1979). *Cold Spring Harbor Symp. Quant. Biol.* **43**, 1209–1215.

*Edited by I. Herskowitz*