Viral Integration and Excision: Structure of the Lambda *att* Sites

DNA sequences have been determined for regions involved in lambda site-specific recombination.

Arthur Landy and Wilma Ross

Site-specific recombination is an important genetic pathway in both prokaryotes and eukaryotes. The range of phenomena associated with this genomic transaction in prokaryotes includes the generation and integration of episomal factors, the highly mobile antibiotic resistance elements, the influence on gene expression by small DNA insertion sequences (IS-elements), and the transition between cytoplasmic vegetative phage genomes and host-integrated prophage. Although less is known about eukaryotic systems, recent biochemical experiments and a renewed appreciation of older genetic experiments are beginning to suggest that site-specific recombination, in addition to being important in the integration of viral genomes, may also be relevant in cellular regulation or developmental pathways (or both) (*1*). An up-to-date collection of recent work in this area is presented in (*2*). One of the most thoroughly characterized examples of site-specific recombination, and also the subject of this article, is the integration and excision reaction of bacteriophage lambda (λ) (*3*).

According to the now well-documented model proposed by Campbell for λ integration, the infecting phage genome is first circularized by means of its cohesive termini and then undergoes a single reciprocal recombination event at specific loci (*att* sites) on the bacterial and viral chromosomes (*4*) (Fig. 1). The product of the phage gene *int*, which maps immediately to the right of *att*, is required for both integration and excision of the λ prophage (*5*). The excision reaction, in addition, requires the product of a second phage gene, *xis*, which maps immediately to the right of the *int* gene (*6*). Although no other phage-specific proteins are required (*7*), there are at least one or more host-specified proteins that are also involved in the integration-excision pathway (*8*). The enzymology of the integration-excision reaction has only recently yielded to direct biochemical analysis (*9, 10*), and information crucial to the formulation of a complete model will undoubtedly be forthcoming in the near future. Not included within the 1400 base-pair *att-int-xis* cluster are the regulatory genes *c*II and *c*III, whose products greatly enhance the expression of the *int* gene, but not of *xis*, by mechanisms which remain to be established (*11, 12*).

In current nomenclature the phage (*att*P) and bacterial (*att*B) *att* sites are designated POP′ and BOB′, respectively, and the prophage *att* sites are designated BOP′ (*att*L) and POB′ (*att*R) (Fig. 1). Transducing phage carrying *att*L (λ*gal*) or *att*R (λ*bio*) are generated from a prophage which has excised from the host chromosome by a rare *int*-independent recombination which deletes phage DNA from one end of the prophage and adds bacterial DNA to the other (*4*). A phage carrying the bacterial *att* site, BOB′, is obtained as a product of *int*-promoted recombination between a *gal* (BOP′) and a *bio* (POB′) transducing phage which is capable of transducing *gal* and *bio* together (*13*).

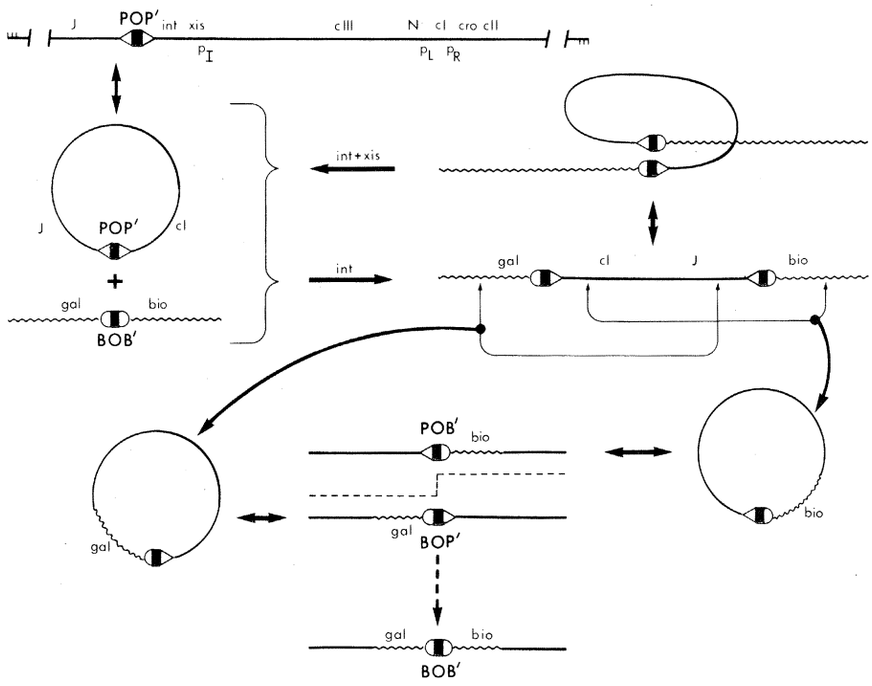
The analysis of *int*-dependent recombination between phage bearing different *att* sites lead to the proposal by Guerrini (*14*) that the arms of the *att* sites, P, P′, B, and B′, differ from each other in nucleotide sequence, and that the extent of *int*-dependent recombination between different *att* sites does not depend on the degree of homology between them. Furthermore, all four arms of a given pair of *att* sites are somehow important together in determining the results of a cross (*13, 15*), as are the temperature and the activity of *xis* protein (*16*). Electron microscopy of appropriate heteroduplex molecules has yielded estimates for the maximum size of the postulated recognition elements, for example, less than 20 base pairs for element B (*17*); however, such conclusions are necessarily qualified because of the number of possible models for *att* site structure (*18*).

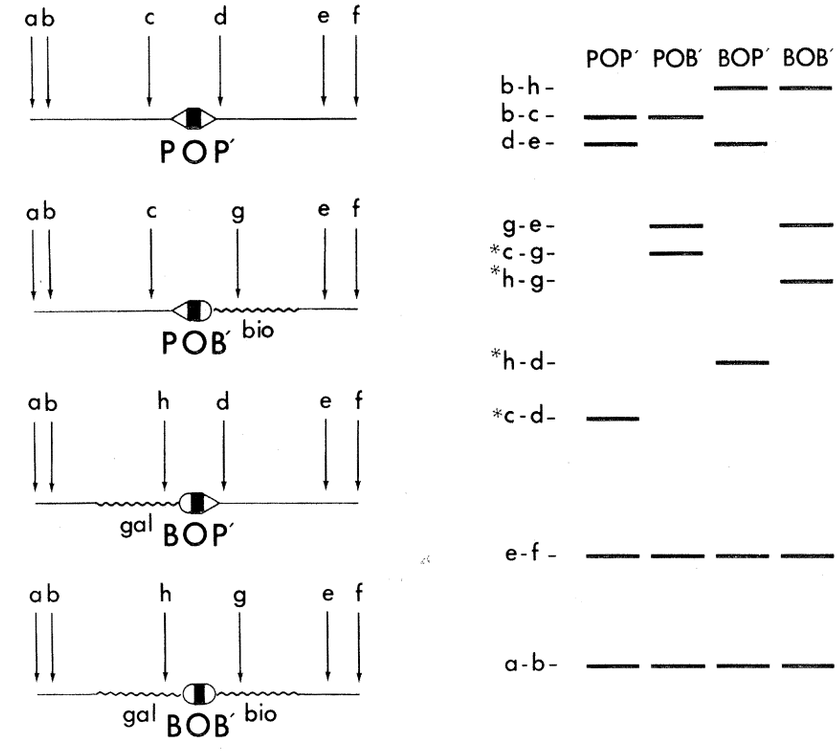
In addition to the unique character of each *att* site, the genetic results of Shulman and Gottesman (*19*) suggested that there is a common region, O, which is shared by all of the λ *att* sites. Mutants defective in *att* function can be crossed from one *att* site into another; and, since some of the progeny of a mutant by wild-type cross are heterozygous, they have postulated that the actual crossover might be made by staggered nicks in the two strands of the common core region (*19, 20*). Failure to find evidence of this type of homology by electron microscope heteroduplex analysis yields an upper limit of 20 to 50 base pairs for a common core region (*17, 21*).

When *Escherichia coli* carries a deletion of the *gal-bio* region, that is, of the primary bacterial *att* site BOB′, *int*-dependent integration of λ can be detected at numerous loci (secondary bacterial *att* sites) on the *E. coli* chromosome (*22*). This integration always involves the phage *att* site (POP′) (*23*) and is thus very similar to the behavior of IS-elements (*24, 25*). The secondary prophage *att* sites are given the general designation ΔOP′ and POΔ′, and they differ from BOP′ and POB′ and from each other in their biological properties as determined by *int*- and *xis*-dependent recombination frequencies with various *att* sites (*26*).

We have undertaken a study of the structures involved in site-specific integrative and excisive recombination in bacteriophage λ. In this article, we report the DNA sequences for each of the four primary *att* sites, POP′, POB′, BOP′, and BOB′. Our results show that the phage *att* site and bacterial *att* site have in common a sequence 15 nucleotide pairs in length. The crossover event of integrative recombination must be within the limits (or at the boundaries) of this "common core" sequence since it occurs unaltered in both the leftward and rightward prophage *att* sites. The four sets of sequences, adjacent to the common core region, are different from one another and hereafter will be referred to as "arms" (P arm, B′ arm, and so forth) without any implication as to their functional significance. Several features of

Dr. Landy is an associate professor of medical science and Wilma Ross is a research associate in the section of Microbiology and Molecular Biology, Division of Biology and Medicine, Brown University, Providence, Rhode Island 02912.





the *att* site structures are discussed in terms of possible models for the site-specific recombination.

Mapping of Primary and Secondary Restriction Fragments

The first step toward direct sequence analysis of the four primary λ *att* sites involved the generation and identification of appropriate DNA restriction fragments with the use of the large collection of site-specific endonucleases that are now available [for recent reviews see (*27*, *28*)]. In order to obtain fragments of an appropriate size and sufficient purity, it was necessary to use a two-step isolation procedure. The logic used in identifying the *att*-containing primary fragments from four phage, each carrying a different *att* site (Fig. 1), is outlined in Fig. 2.

When DNA from each of these four phage is digested with Hind II+III, one unique fragment is found in each gel profile (Fig. 3, legend). This work has been described (*29*). From the sizes of these fragments, and the placement of the Hind III cut site, which forms the left end of the POP′ and BOB′ fragments at approximately 56.8 on the λ map (*30*), it was concluded that the entire attachment region of each phage (centered at 57.3) was probably contained in these primary fragments.

The primary *att*-containing fragments were isolated from preparative-scale gels

Fig. 1 (top). The λ integration-excision pathway and the origin of transducing phage carrying different *att* sites. Integration of the circularized phage genome requires the phage-encoded protein *int* and proceeds *via* site-specific recombination between the phage *att* site POP′ and the primary bacterial *att* site BOB′. The reaction yields a prophage bounded by a left prophage *att* site, BOP′ (*att*L), and a right prophage *att* site, POB′ (*att*R). The reverse reaction, excision, involves recombination between BOP′ and POB′ to yield BOB′ and POP′ (excised phage) and requires *int* and a second phage product *xis*, both of which map immediately to the right of *att*. The *gal* transducing phage carrying the left prophage *att* site and *bio* transducing phage carrying the right prophage *att* site are created by rare excision events in which *int*-independent recombination occurs at sites other than the prophage *att* sites. A *gal-bio* transducing phage carrying the bacterial *att* site BOB′ is generated by *int*-dependent recombination between the prophage *att* sites of coinfecting *gal* and *bio* transducing phage. Fig. 2 (bottom). Logic used to identify restriction fragments carrying the four *att* sites POP′, POB′, BOP′, and BOB′. The DNA's from four phage, each carrying a different *att* site, are cleaved with a restriction endonuclease at the specific sites marked *a* to *h*. The DNA restriction fragments, resolved by electrophoresis in adjacent gel lanes, are predicted to fall into three categories: (i) those common to all four profiles, and deriving from phage DNA outside the boundaries of the bacterial DNA substitutions, for example, fragments a–b and e–f; (ii) those appearing in two of the four profiles, and deriving from within bacterial DNA, from the non-*att* phage-bacterial DNA juncture, or from within phage DNA deleted during formation of the transducing phages, for example, fragments b–c, d–e, b–h, and g–e; and (iii) those occurring in only one profile. These "unique" fragments, of which there should be one per profile, contain the crossover region, for example, fragments c–d, c–g, h–d and h–g (designated with an asterisk in the figure). Preparative quantities of the primary *att*-containing fragments (see text) are subjected to secondary digestions with restriction enzymes having shorter, and therefore more frequent, recognition sites. Secondary *att*-containing fragments can be identified by a similar logic; however, the first category of fragments, those appearing in all four profiles, is not expected in these experiments since the ends of the primary fragments fall within the bacterial DNA substitutions. An exception to the above predictions would be found if a primary or secondary restriction site occurred within a region of homology that was common to all four *att* sites; in this case, no unique, *att*-containing fragments would be found. This situation did not occur in our experiments.

of Hind II+III digests and used as source material for the secondary digests (*29*). The method used to identify the *att*-containing secondary fragments generated by each of five restriction enzymes was similar to that used for primary fragments and is outlined in Fig. 2. A secondary fragment that is not unique can be assigned to a particular arm based upon the two profiles in which it appears. By labeling the 5′ termini of the primary fragment with ^{32}P prior to the secondary digestion, the end fragments can be determined. If the number of fragments assigned to an arm is only one or two, the order of these and of the adjacent unique is readily defined. Simultaneous digestion with two restriction enzymes confirms the placement of cut sites determined from single digests (especially important in the case of sites which are close to one another) and in some instances yields additional information about fragment order. The gel profiles of Hinf I digestion products of the primary fragments are shown in Fig. 3 as an example of these experiments.

As seen from the map shown in Fig. 4, restriction enzyme sites are found at appropriate intervals for sequence analysis of both DNA strands in the P, B, and B′ arms. The P′ arm presented a stretch of approximately 170 base pairs of sequence without a restriction site (between Alu I and Mbo II sites) in an area known to contain the crossover region. Therefore, to supplement the restriction enzymes, a less specific nuclease, endonuclease IV (*31*), was used to generate smaller fragments for sequence analysis in this region (*32*).

DNA Sequencing in the Crossover Region

We have used the dimethyl sulfate–hydrazine chemical modification method developed by Maxam and Gilbert (*33*) for our sequence work. A DNA restriction fragment uniquely labeled with ^{32}P at one

Table 1. Analysis of sequence from restriction fragments containing the P and P′ arms of the phage *att* site. Section 1 includes the P arm sequence and sections 2 and 3 the P′ arm sequence. The fragment numbers (first column), for example, PP′1, have been arbitrarily assigned to facilitate reference between this table, Fig. 6, and the text. A fragment is named (second column) by the primary fragment from which it was derived (see Fig. 3, legend), the secondary restriction enzyme or enzymes used to generate it, its approximate length in base pairs, and, where applicable, the purified strand, *l* or *r* (*l* is the template strand for leftward transcription). The double-stranded fragments used in these experiments are 5′-labeled at only one end, the end created by the first enzyme in the fragment name; the second enzyme named was used after labeling to remove the other 5′ end. Single-stranded fragments were obtained by gel purification of strands after 5′ labeling of a double-stranded fragment generated by either one or two enzymes (Fig. 5 legend). Primary fragment 1290 (from which PΔ′1 was obtained) was derived from a phage carrying a secondary prophage *att* site, λY905 (*29*). (Fragment 1290–Alu I + Hha I–155 was provided by Karen Bidwell.) Numbering of the sequence is as described in Fig. 7, with zero being the center base of the 15-base common core region. The boundaries of the common core are indicated by brackets. The ^{32}P-labeled 5′-end is designated *p. . . . Sequence from the *r* strand is written in reverse chemical polarity (that is, written 5′ to 3′ right to left), and is underlined with dots. Large letters represent sequence read confidently in that particular experiment. Sequence in small letters was either run off the gel or was not determined with certainty in that experiment, but was confirmed by data from other experiments. The 5′ ends of the endonuclease IV fragments are not known precisely; they were estimated on the basis of (i) migration of bands on sequencing gels relative to a bromophenol-blue dye marker and (ii) the preferred recognition sequences of this enzyme (*31*).

Fragment Number	Fragment Name	Sequence
		SECTION 1 P REGION (−114 to −3)
		−110 −100 −90 −80 −70 −60 −50 −40 −30 −20 −10
PP′1	Cl-HinfI-320, l strand EndoIV-54	*p...TCATAGTGACTGCATATGTTGTGTTTTACAGTATTATGTAGTCTGTTTTTTATGOH
PP′2	Cl-HinfI-320, l strand	*p...TCATAGTGACTGCATATGTTGTGTTTTACAGTATTATGTAGTCTGTTTTTTATGCAAAATCTAATTTAATATATTGATATTTATATCATTTTACGTTTCTCGTTCAGCTTT...OH
PP′4	Cl-HinfI-320, r strand EndoIV-43	HOAAGTCGAAA...p*
PP′5	Cl-HinfI-320, r strand EndoIV-55	HOAAGTCGAAA...p*
PΔ′1	1290-AluI+HhaI-155	HO...CAACACAAAATGTCATAATACATCAGACAAAAAATACGTTTTAGATTAAATTATATAACTATAAATATAGTAAAATGCAAAGAGCAAGTCp*
PB′1	D3a-AluI+HinfI-109, l strand	*p...TCATAGTGACTGCATATGTTGTGTTTTACAGTATTATGTAGTCTGTTTTTTATGCAAAATCTAATTTAATATATTGATATTTATATCATTT...OH
PB′2	D3a-AluI+HinfI-109, r strand	HO...TACGTTTTAGATTAAATTATATAACTATAAATATAGTAAAATGCAAAGAGCAAGTCp*
PB′3	D3a-AluI+HinfI-109	HO...GACGTATACAACACAAAATGTCATAATACATCAGACAAAAAATACGTTTTAGATTAAATTATATAACTATAAATATAGTAAAATGCAAAGAGCAAGTCp*
		SECTION 2 P′ REGION (+3 to +105)
		+10 +20 +30 +40 +50 +60 +70 +80 +90 +100
PP′3	Cl-AluI+HinfI-211	*p...ACTAAGTTGGCATTATAAAAAAGCATTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTGATTTCAATTTTGTCCCAC...OH
PP′4	Cl-HinfI-320, r strand EndoIV-43	HO...TGATTCAACCGTAATATTTTTTCGTAACGp*
PP′5	Cl-HinfI-320, r strand EndoIV-55	HO...TGATTCAACCGTAATATTTTTTCGTAACGAATAGTTAAACAACp*
PP′6	Cl-HinfI-320, r strand EndoIV-61	HOTTGTCCAGTGATAGTCAGTTTTATTTTAGTAATAAACTAAAGTTAAAACAGGGTG...p*
PP′8	Cl-HinfI-320, r strand	HO...TTTAGTAATAAACTAAAGTTAAAACAGGGTG...p*
BP′2	B7a-HinfI-227, l strand	*p...ACTAAGTTGGCATTATAAAAAAGCATTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTGATTTCAATTTTG...OH
BP′3	B7a-HinfI-227, r strand	HO...AACTAAAGTTAAAACAGGGTG...p*
		SECTION 3 P′ REGION (+100 to +203)
		+100 +110 +120 +130 +140 +150 +160 +170 +180 +190 +200
PP′6	Cl-HinfI-320, r strand EndoIV-61	HO...AGGGTGAGGGACp*
PP′7	Cl-HinfI-320, l strand EndoIV-41	*pCTGCCTCTGTCATCACGATACTGTGATGCCATGGTGTCCGAOH
PP′8	Cl-HinfI-320, r strand	HO...AGGGTGAGGGACGGAGACAGTAGTGCTATGACACTACGGTACCACAGGCTGAATACGGGCTCTTCTACAACTCGTTTGAATAGCGAATAGACGAAGAGTATCTC...p*
PP′9	Cl-HinfI-320, r strand EndoIV-46	HOAGGGTGAGGGACGGAGACAGTAGTGCTATGACACTACGGTACCACp*
BP′3	B7a-HinfI-227, r strand	HO...AGGGTGAGGGACGGAGACAGTAGTGCTATGACACTACGGTACCACAGGCTGAATACGGGCTCTTCTACAACTCGTTTGAATAGCGAATAGAGGAAGAGTATCTC...p*

5′ end is obtained in one of two ways. A restriction fragment that has been labeled at both 5′ ends with polynucleotide kinase and [γ-^{32}P]-ATP is cut with an appropriate restriction enzyme, and the two singly labeled product fragments are reisolated by gel electrophoresis. Alternatively, the denatured strands can be separated by agarose (*34*) or polyacrylamide gel electrophoresis (*33*). The singly labeled molecule is then chemically modified in four parallel base-specific reactions that render the molecules labile to alkali cleavage at the modified bases. Reaction with dimethyl sulfate is specific for purines, and reaction with hydrazine is specific for pyrimidines. Adenine (A) is distinguished from guanine (G) by differential release of the two bases after the modification step. Thymine (T) is distinguished from cytosine (C) by the fact that a high concentration of salt depresses the modification (and therefore the cleavage) of T relative to C. By limiting the extent of these chemical reactions, the size ranges of cleavage products can be controlled. The cleavage products are subjected to electrophoresis in adjacent gel lanes under conditions that permit resolution of fragments differing in length by a single nucleotide. As can be seen in Fig. 5, the DNA sequence can be read directly from an autoradiograph of the gel by ascertaining in which of the four lanes a band appears for each size increment. Depending on the quality of gel resolution, an average of 80 to 110 bases of sequence may be read from one labeled 5′ end in a series of samples that have undergone electrophoresis for differing times.

Table 2. Analysis of sequence from restriction fragments containing the common core region of the phage and bacterial *att* sites. The boxed area indictes the 15-base common core sequence. See also the legend to Table 1.

Fragment Number	Fragment Name	Sequence (−10 0 +10)
PP′2	C1-HinfI-320, l strand	*p...TTCTCGTTCAGCTTTTTTATACTAAGTTGGCATTA...OH
PP′3	C1-AluI+HinfI-211	*pCTTTTTTATACTAAGTTGGCATTA...OH
PP′4	C1-HinfI-320, r strand EndoIV-43	HOAAGTCGAAAAAATATGATTCAACCGTAAT...p*
PP′5	C1-HinfI-320, r strand EndoIV-55	HOAAGTCGAAAAAATATGATTCAACCGTAAT...p*
PB′4	D3a-AluI+HpaII-275	*pCTTTTTTATACTAACTTGAGCGAA...OH
BP′2	B7a-HinfI-227, l strand	*p...GTTGAAGCCTGCTTTTTTATACTAAGTTGGCATTA...OH
BB′4	C8a-HhaI-163, l strand	*p...GTTGAAGCCTGCTTTTTTATACTAACTTGAGCGAA...OH
BB′5	C8a-HhaI-163, r strand	HO...CAACTTCGGACGAAAAAATATGATTGAACTCGCTT...p*

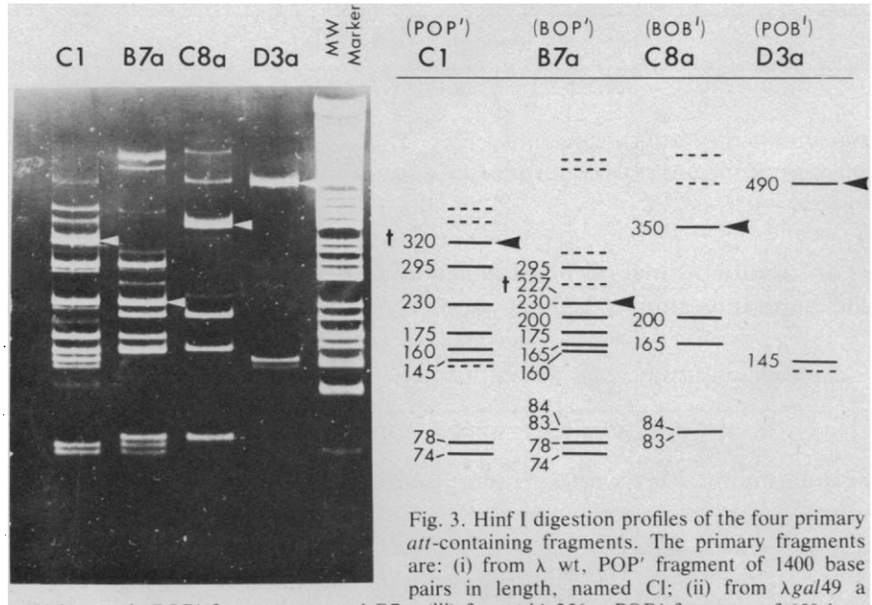


Fig. 3. Hinf I digestion profiles of the four primary *att*-containing fragments. The primary fragments are: (i) from λ wt, POP′ fragment of 1400 base pairs in length, named Cl; (ii) from λ*gal*49 a 1600-base-pair BOP′ fragment, named B7a; (iii) from λ*bio*256 a POB′ fragment of 650 base pairs, named D3a; and (iv) from λ*gal*49*bio*256 a BOB′ fragment of 850 base pairs, named C8a (*29*). A portion (1.0 to 1.5 μg) of each purified primary fragment was incubated with Hinf I restriction endonuclease (*29*). Samples were subjected to electrophoresis on a 1-mm thick, 8 percent acrylamide–0.26 percent methylene-bis-acrylamide gel in tris-borate buffer, *p*H 8.3, for 4 hours at 150 volts. The gel was stained in ethidium bromide (2 μg/ml) and photographed with Polaroid 55 P/N film using a long-wave ultraviolet light source. Arrows mark the unique *att*-containing fragments in both gel and schematic profiles. (The B7a unique fragment, 227, comigrates with fragment 230, which is also found in Cl. In double digests with Hpa II, the 230 fragment is cut, leaving 227 readily apparent as the unique fragment; data not shown.) The molecular weight marker is a Hind II+III digest of ϕ80psu$^{+}_{III}$ DNA (*62*). (Calibration of the smaller fragments in this marker digest will be described elsewhere.) Molecular weight estimates in base pairs are given besides each fragment in the schematic profiles. (———) Hinf I products of primary fragments; (- - - - -) products of other fragments contaminating the primary fragment preparation (present in less than molar yield); (†) fragments whose entire sequences have been determined.

The five panels in Fig. 5 have been chosen as examples of the sequencing gels because they are from the region of the crossover in three of the *att* sites. (A description of the nomenclature for the secondary restriction fragments appears in the legend to Table 1.) The Alu I cut in the P arm was recognized, from mapping studies, to be the closest one to the crossover region in the phage *att* site (see Fig. 4). Panel B of Fig. 5 shows the sequence proceeding rightward from the Alu I cut in the POP′ fragment. In this gel, 30 of the first 33 bases of sequence are identified, and with two additional gels that have undergone electrophoresis for longer times (not shown) the total amount of sequence obtained ran from −6 to +105 (Fig. 7). Figure 6 shows a schematic representation of this sequence (labeled PP′3) as it maps with respect to sequences derived in other experiments. The actual sequence information obtained from this gel series is shown in Tables 1 and 2.

The same Alu I cut labeled in the above series of experiments also occurs in the right prophage *att* site, POB′. Panel A of Fig. 5 shows 31 of the first 34 bases of sequence proceeding rightward from this cut site into the B′ arm, and Fig. 6 shows the total length of sequence determined (labeled PB′4). The two sequences are expected to be identical for some distance and then to diverge. The predicted identity is found in the first 14 bases, 11 of which are shown bracketed in panels A and B of Fig. 5. Although the divergence starts at position +8 (with a G in the POP′ site and a C in the POB′ site; see Fig. 7), positions 9, 10 and 11

Table 3. Analysis of sequence from restriction fragments containing the B (sections 1 and 2) and B′ (section 3) arms of the bacterial *att* site. See also the legend to Table 1.

Fragment Number	Fragment Name	Sequence
		SECTION 1 B ARM (-171 to -80)
		-170 -160 -150 -140 -130 -120 -110 -100 -90 -80
BP′1	B7a-HinfI+AluI-99	HO...GTTACGGTCGCGG...p*
BB′1	C8a-AluI-370, r strand	HO...cTCCATGGTCGCGCCAAACTAGTCTTCCTGCAACTAGCCCGCCCCAACTCp*
BB′2	C8a-HhaI-78, l strand	*pCGGTTTGATCAGAAGGACGTTGATCGGGCGGGGTTGAGCTACAGGCGGTCAGCGTCACGCcAAAAGCCAATGCCAGCGOH
BB′3	C8a-HhaI-78, r strand	HOGCGCCAAACTAGTCTTCCTGCAACTAGCCCGCCCCAACTCGATGTCCGCCAGTCGCAGTGCGGTTTTCGGTTACGGTCp*
		SECTION 2 B ARM (-85 to -3)
		-80 -70 -60 -50 -40 -30 -20 -10
BP′1	B7a-HinfI+AluI-99	HO...TCGCGGTCTGCCCTTTGACTTTTACACAAGTGTCCAACGAGGCCCGATACTTTATCTTTTTACTTAp*
BP′2	B7a-HinfI-227, l strand	*pAATCCGTTGAAGCCTGCTTT...OH
BB′4	C8a-HhaI-163, l strand	*pCCAGACGGGAAACTGAAAATGTGTTCACAGGTTGCTCCGGGCTATGAAATAGAAAAATGAATCCGTTGAAGCCTGCTTT...OH
BB′5	C8a-HhaI-163, r strand	HO...cCCGATACTTTATCTTTTTACTAGGCAACTTCGGACGAAA...p*
		SECTION 3 B′ ARM (+3 to +88)
		+10 +20 +30 +40 +50 +60 +70 +80
PB′4	D3a-AluI+HpaII-275	*p...ACTAACTTGAGCGAAACGGGAAGGTAAAAAGACAAAAAGTTGTTTTTAATACCTTTAAGTGATACCAGATGGCATTGCGCCATCTG...OH
BB′4	C8a-HhaI-163, l strand	*p...ACTAACTTGAGCGAAACGGGAAGGTAAAAAGACAAAAAGTTGTTTTT...OH
BB′5	C8a-HhaI-163, r strand	HO...TGATTGAACTCGCTTTGCCCTTCCATTTTTCTGTTTTTCAACAAAAATTATGGAAATTCACTATGGTCTACCGTAACGCp*

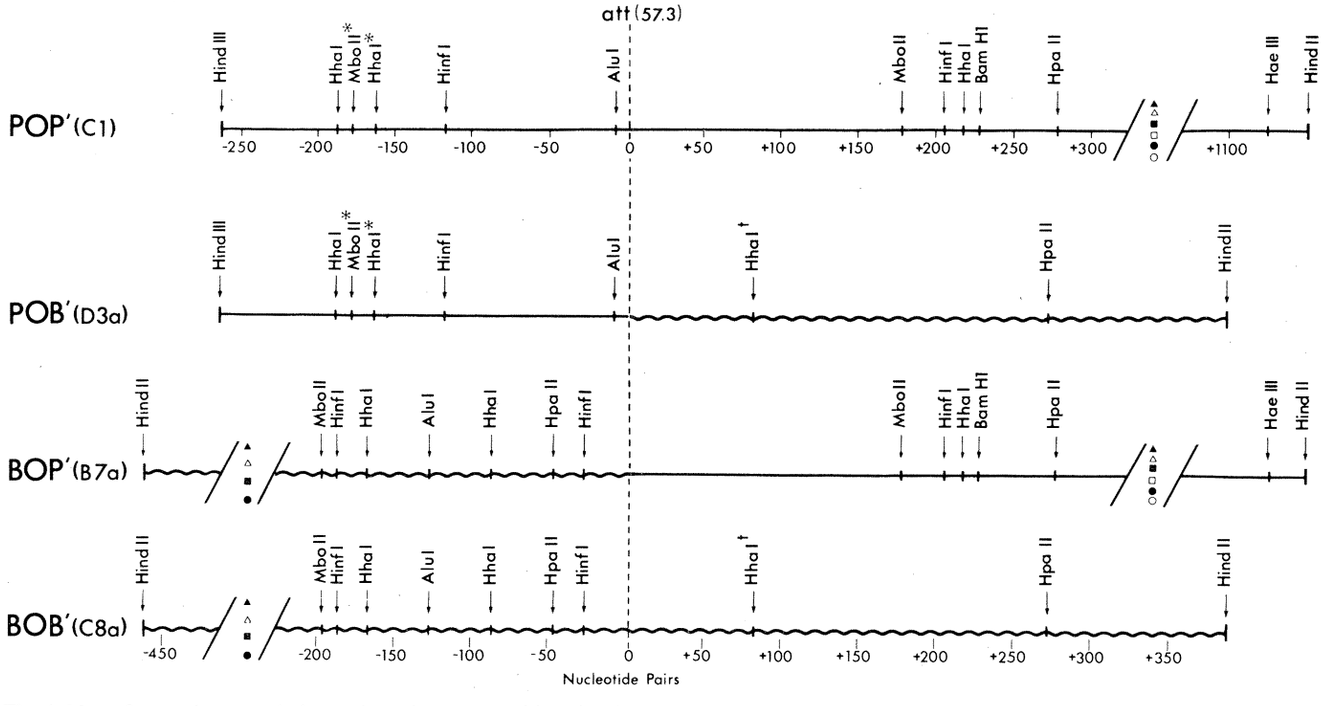
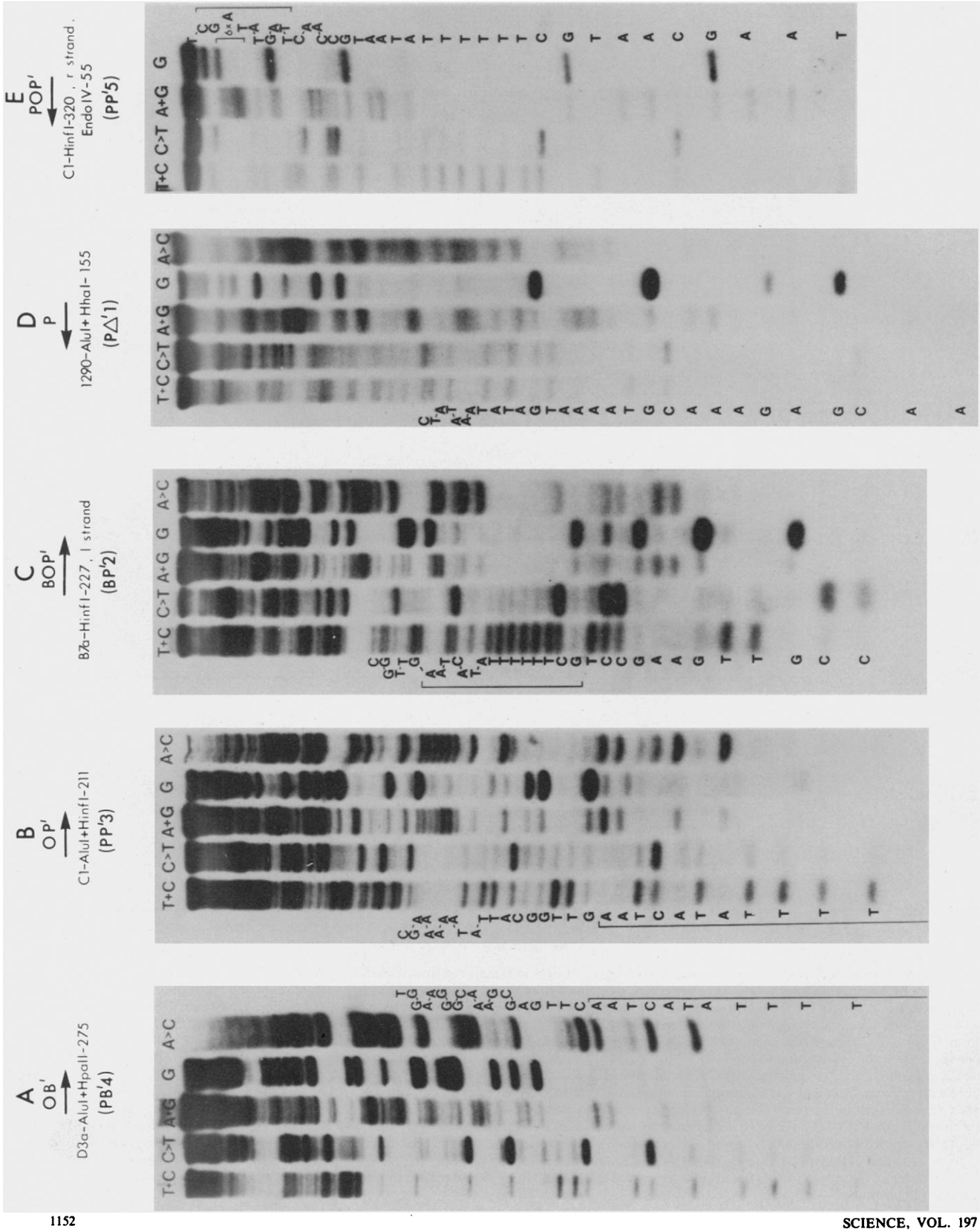


Fig. 4. Map of secondary restriction endonuclease recognition sites in the *att* regions of the four primary *att*-containing fragments, Cl, D3a, B7a, and C8a. The primary fragments (*29*) were generated by a mixture of the two nucleases Hind II+III, recognizing the sequences GTPyPuAC (*63*) and AAGCTT, respectively (*64*). The secondary enzymes used in studying all four primary fragments were: Hinf I, purified from *Haemophilus influenzae* strain R_f, and recognizing the sequence GANTC (*28*); Hha I, purified from *H. haemolyticus*, recognizing GCGC (*65*); Hpa II, purified from *H. parainfluenzae*, recognizing CCGG (*66*); Mbo II, purified from *Moraxella bovis*, recognizing GAAGA and cutting eight base pairs downstream of the recognition site (*28*); and Alu I purified from *Arthrobacter luteus*, and recognizing AGCT (*67*). Hae III, purified from *Haemophilus aegyptius*, and recognizing GGCC (*28*) and BamH I, purified from *Bacillus amyloliquefaciens H*, recognizing GGATCC (*28*) were studied only with primary fragment C1. Digestion conditions used for all enzymes were as reported (*29*). The methods used in mapping these recognition sites are discussed in the text. The numbering scale, in base pairs, is the same as that used for the sequence in Fig. 7. The precise positions of some restriction cut sites have been determined by sequence analysis (Figs. 6 and 7). Additional sites are known to occur in the B and P′ arms, but their map positions are undetermined. The presence of unmapped sites for particular enzymes is designated by the symbols between slashes in these arms. (▲) Hha I; (△) Mbo II; (■) Hinf I; (□) Alu I; (●) Hpa II; (○) Hae III; (———) phage DNA; (~~~~~) bacterial DNA; (*) the relative order of these two sites is not certain; (+) three additional Hha I sites occur to the right of the one mapped in B′, but their locations have not been determined.

are the same in both sequences. The difference at position +8 has been confirmed in several different experiments (Table 2). Thus the A at position +7 must either be at the crossover point, or it must constitute the right boundary of a common core region (or both). The left boundary of a common region could be no further to the left than the Alu I site, as this occurs only in the P arm. This establishes a maximum of 15 bases for a possible common core region, and the final resolution of its left boundary depends on sequence obtained from the right prophage *att* site, BOP′.

The Hinf I cut forming the left end of a BOP′-unique fragment had been mapped in the B arm a short distance to the left of



the Alu I cut in the P arm (Fig. 4). Therefore, this BOP′-unique fragment should contain a short length of B arm sequence which is not found in either of the fragments discussed above. This sequence would be followed either by a common core region, or there would be no region of identity with POB′; and, at position +8 in BOP′, the sequence would become identical to that determined for POP′. Panel C of Fig. 5 shows 33 of the first 36 bases proceeding rightward from the left Hinf I cut of the BOP′-unique (see also BP′2 in Fig. 6 and Tables 1 to 3). As was expected, the sequence initially differs from the POP′ and POB′ unique fragments. At position −7 (16 bases from the labeled Hinf I terminus) the sequence becomes identical to that found in both POP′ and POB′, and from position +8 rightward it diverges from POB′ and remains identical to POP′. The left boundary of a 15-base-pair common region is thus established at −7 (the G of the AGCT Alu I site found in the P arm).

Panel D of Fig. 5 shows the sequence extending leftward from the Alu I site in the P arm, confirming that this sequence does differ from that presumed to be the B arm sequence (panel C of Fig. 5). Panel E of Fig. 5 shows the sequence derived from an endonuclease IV product of the purified *l* strand of the Hinf I unique fragment from the POP′ *att* site (PP′5 in Fig. 6). This sequence is complementary to a portion of PP′3 (+34 to −6) and provides a "read through" of the Alu I site, thus confirming and connecting the sequences shown in panels B and D of Fig. 5. This shows that there is no small undetected intervening Alu I fragment, a function also served by PP′2 (Fig. 6).

An interesting feature of the O and P′ regions, which can be seen in panels B and E of Fig. 5, is an inverted repeat (see below) 11 base pairs in length and involving positions −7 to +3 and +16 to +26.

Fig. 5. Sequencing gels illustrating the central portion of the phage and bacterial *att* regions. Each gel shows a portion of the sequence that was obtained from the fragment named above the panel, for example, panel A sequence is from PB′4 (Fig. 6), fragment D3a–Alu I + Hpa II–275 (see Tables 2 and 3). Also indicated above each panel are the region and direction of the sequence; for example, the sequence in panel A begins in O, the common core region, and extends into the B′ arm. Brackets indicate those bases that are part of a common core sequence. (In panels A and B, the bands corresponding to the first three nucleotides of the common core have been run off the gels; these bases were seen on other gels, for example, panel E; see also Table 2.) In preparation for sequencing, the fragments used in panels A to D were generated by digestion with the first enzyme in the name, then labeled at the 5′ end with polynucleotide kinase and [γ-^{32}P]ATP as described (*29*). [γ-^{32}P]ATP was synthesized according to modifications of methods described (*33, 68*). The two 5′ ends of each labeled fragment were separated either by (i) digestion with the second named enzyme, followed by gel purification of the products (panels A, B and D), or (ii) separation of the two DNA strands of the fragment by gel electrophoresis (panel C) (*33, 34*). In panel E, purified *r* strand of C1–Hinf I–320 was digested with endonuclease IV (*31*), and the products were labeled at the 5′ end and purified by gel electrophoresis (*32*). Labeled fragments were eluted from gels by diffusion and sometimes purified from soluble gel material by BND cellulose or hydroxylapatite column chromatography. The sequencing method has been described by Maxam and Gilbert (*33*). We chose to use five of the reactions specified by Maxam and Gilbert, including two "alternative" cleavages. For the T+C lane, labeled fragment was incubated with 12 to 15*M* hydrazine hydrate. For the C>T lane, incubation was with 14 to 17*M* hydrazine hydrate, plus 1.5*M* NaCl. For the G+A and G lanes, the fragment was incubated with dimethyl sulfate. Half of the sample, for the G+A lane, was then subjected to the "strong adenine/weak guanine" cleavage method. The remaining half was cleaved by the "alternative guanine" method to specifically cleave methylated G residues. A fifth reaction, the "alternative strong adenine/weak cytosine" cleavage, was used for the A>C lane. The gels illustrated are 16 percent acrylamide–7*M* urea, with electrophoresis at 600 volts for approximately 9 hours, after a 6-hour prerun of the gel. Autoradiography was done at −20°C.

Extent of Sequence Determined

Considerable additional sequence information was determined both in the region of the common core discussed above and in the arms extending leftward and rightward from the common region (Fig. 7). Several secondary fragments from each of the four primary *att*-containing fragments were used to generate the sequences shown in Fig. 7, and the relative positions and extents of sequences determined from these fragments are illustrated in Fig. 6. The derivation of the fragments, labeled PP′1, PP′2, and so forth, can be determined from Tables 1 to 3. Complementary sequences were obtained for all regions reported, with the exception of the distal ends of the arms and a short (18 base-pair) segment of the P′ arm between +35 and +52 (Fig. 6 and Table 1). Often, one strand was sequenced from one primary fragment, the other strand from the other primary fragment carrying the same arm. Obtaining complementary information served as a check on the accuracy of the method, the same base-pair information being obtained from two different base specific reactions. No contradictory information was found. In all cases, the sequences determined from opposite strands did complement each other; and sequences from the two primary fragments carrying each arm were in agreement.

In addition, Fig. 6 illustrates that wherever a restriction site within the reported sequence was used as an end from which to determine sequence (Alu I in the P arm; Alu I, Hha I and Hinf I in the B arm; and Hha I in the B′ arm) the restriction site and adjacent bases to either side were "read through" in sequence determined from another secondary fragment. This confirmed that no very small restriction fragments, undetected by our mapping procedures, occur between the sequenced fragments.

A summary of the data obtained from all fragments sequenced is presented in Tables 1 to 3. All bases reported were unambiguously observed at least once, and almost all in two or more experiments. The sequence information from all four primary λ *att* sites totals approximately 1000 base pairs. For the phage *att* site, the sequence includes 107 base pairs to the left (P arm) and 196 base pairs to the right (P′ arm) of the common core boundaries. Similarly, for the bacterial site it includes 166 base pairs to the left and 81 base pairs to the right of the core boundaries. Preliminary sequence information corresponding to one strand of the phage *att* site has been obtained by Davies using a different approach from that described here; however, identification of the common core was not possible (*35*).

A Common Core

The most striking feature of the results shown in Fig. 7 is the 15-nucleotide-pair sequence that is common to all four *att* sites. The crossover event of integrative recombination must occur within the limits (or at the boundaries) of this "common core" sequence, since this sequence is unaltered in both the leftward and rightward prophage *att* sites. This observation eliminates the possibility, previously untestable by genetic means, that phage insertions result in some sequence alterations that are then reversed in the course of excisive recombination. The common core region could also be viewed as support for the suggestion from genetic experiments that the integration-excision reactions proceed by way of staggered cuts in the two DNA strands of each recombining *att* site (*19, 20*). However, there are other equally attractive functions that can be ascribed to

this region on the basis of sequence features discussed below.

In their detailed analysis of λ secondary insertion sites, Shimada *et al.* (*23*) calculated that, if the *E. coli* genome is treated as a random DNA sequence, the observed frequency of secondary *att* sites would be indicative of a 5- to 6-base-pair recognition site. This discrepancy between the frequency of secondary *att* sites and the observed length of the common core region is suggestive of two possibilities.

According to the first, it is likely that most, or many, secondary bacterial *att* sites do not possess the exact 15-nucleotide-pair sequence of the common core region. This would be consistent with the wide range of efficiencies found among different secondary bacterial *att* sites (*23*). The excision of λ from a secondary *att* site within a bacterial gene results in the restoration of full wild-type function both to the bacterial gene and to the λ *att* site (*23*). Therefore, if excision proceeds via staggered cuts in the two strands of each recombining *att* site, it is not likely that there would be mismatched base pairs within any resulting heteroduplex region. This would suggest that secondary *att* sites are likely to have a shorter but uninterrupted region of homology with the common core region of the phage *att* site. According to the second possibility, which is not mutually exclusive of the first, the λ insertion sites might consist of sequences that occur more frequently in the *E. coli* chromosome because of some particular functional significance.

Bias in Base Composition

It is not surprising that a common core region, although it had been looked for carefully, was never detectable in electron microscope heteroduplex analyses (*17, 21*). In addition to its small size, the core is 80 percent A+T and hence would tend to form relatively unstable du-

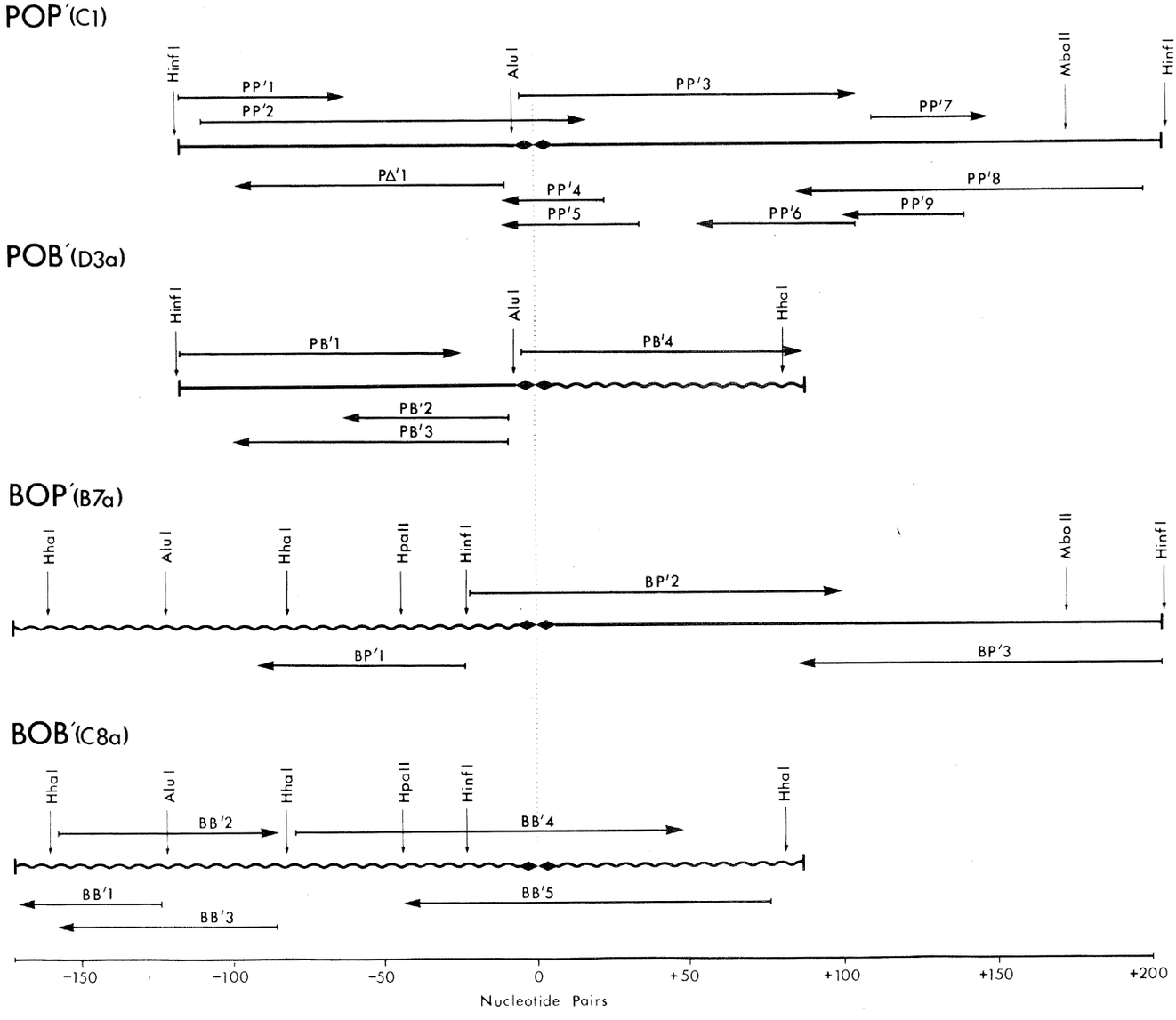


Fig. 6. Schematic summary of results from Tables 1 to 3, showing the extent of complementary and overlapping sequence information. Only those portions of the four primary fragments from which sequence was determined are represented. Restriction sites in the sequenced region are indicated. Arrows represent lengths of sequence determined in particular experiments; the source of fragments, PP′1, and so forth, and actual sequence determined from them can be found in Tables 1 to 3. Arrows indicate sequence in a 5′ to 3′ direction. Those pointing rightward represent sequence from the *l* strand, those pointing leftward represent sequence from the *r* strand. The scale, in nucleotide pairs, is as designated for Fig. 7. (——) Phage DNA; (~~~~~) bacterial DNA; (◆◆) common core region.

plexes. As is shown in Fig. 8, the bias in base composition is not confined to the common core region, but extends from approximately −40 to +60 (average 70 percent A+T) in the bacterial *att* site and from +100 to at least −100 (average 75 percent A+T) in the phage *att* site. In the P arm, our sequence does not extend far enough to delimit the region of high A+T, and these sequences are probably contiguous with, or a part of, the readily denatured region that extends approximately from the *att* site leftward into the *b*2 region (*36*).

The significance of this extreme bias in base composition is most likely related to the recent findings that negatively supercoiled DNA is required as a substrate for integrative recombination (*37*). Since negatively supercoiled DNA will tend to partially denature in regions of high A+T (*38*, *39*), there is the strong suggestion of a preference for single-stranded DNA by one or more of the components in the recombination pathway. Our results so far with secondary bacterial *att* sites, however, suggest that an extended region of high A+T may be a required feature only in the phage *att* site.

Molecular Palindromes

The rotationally symmetrical character of a molecular palindrome (a DNA sequence whose complementary strands read the same in the 5′ to 3′ direction) permits its recognition by a protein composed of two or more identical subunits, each of which can interact with identical features on the complementary DNA strands. In addition, this symmetry would facilitate recognition from either direction by a protein capable of one-dimensional diffusion along the DNA molecule. The significance of molecular palindromes as protein recognition elements is most convincingly substantiated by their persistent recurrence in a large number of different sequences comprising recognition sites for a variety of restriction endonucleases.

A particularly rich cluster of overlapping palindromes is found in the P arm in the region from −25 to −55 (Fig. 7). It is interesting that the phage *att* site is far richer than the bacterial *att* site in palindromic sequences, and this, as well as other sequence features (see below), suggests that the phage elements may play a more critical role than the bacterial elements in integrative and excisive recombination. There are no molecular palindromes that are unique to either of the prophage *att* sites.

Identities and Direct Repeats

As is seen in Fig. 7, there are a few places where the P and B arms (or P′ and B′ arms) have similar sequences (identities). Since there is no a priori reason to expect any such identities, the sequences may possibly be of functional significance. It may also be of interest that one of the largest identities (at positions +20 and +30) includes one element of the largest inverted repeat found in the entire sequence (this inverted repeat, which crosses the OP′ juncture, is of particular interest and is discussed below). Whether or not the sequence similarities are functionally related to site-specific recombination, they may reflect some evolutionary link in the generation of these two genetic regions.

There are two primary reasons for noting sequences that are repeated on the same DNA strand in the neighborhood of the *att* sites. The first concerns those schemes for the *att* site recognition elements which suggest that the same sequence might be used on both sides of a common core region—for example, POP in the phage *att* site and BOB in the bacterial *att* site (*40*). However, the *att* site sequences do not contain those features which would be predicted by this class of models (Fig. 7). The second reason for noting repeated sequences concerns the possible existence of multiple binding sites for one or more of the proteins which interact with this region. This type of interaction, for example, is found in the sequential binding of multiple repressor molecules to the λ operator regions (*41*).

Inverted Repeats

An inverted repeat consists of two DNA sequences which read (with the same chemical polarity) the same on opposite strands, and it has the property of rotational symmetry. Therefore, when the distance between the two elements of an inverted repeat is not too large, these sequences are, like molecular palindromes, potential sites for interaction with dimeric or tetrameric proteins. These sequences also have the capacity for intrastrand H-bonding, and the formation of double hairpin structures (*42*). Although there is at present no evidence to support the natural existence of double hairpins in native DNA (*43*), there have been several models advanced for both generalized and site-specific recombination in which these structures play a prominent role (*44*). Finally, inverted repeats and direct repeats have a special significance in reactions, such as site-specific recombination, which consist of symmetrical operations. (RNA polymerase binding and initiation of transcription is an example of a nonsymmetrical operation.) If there is a polarity to the action of a protein that binds on both sides of the crossover region, the recognition sequences would constitute an inverted repeat. Proteins that have a nonpolar mode of action could just as well be associated with direct repeats centered around the crossover region (see above.)

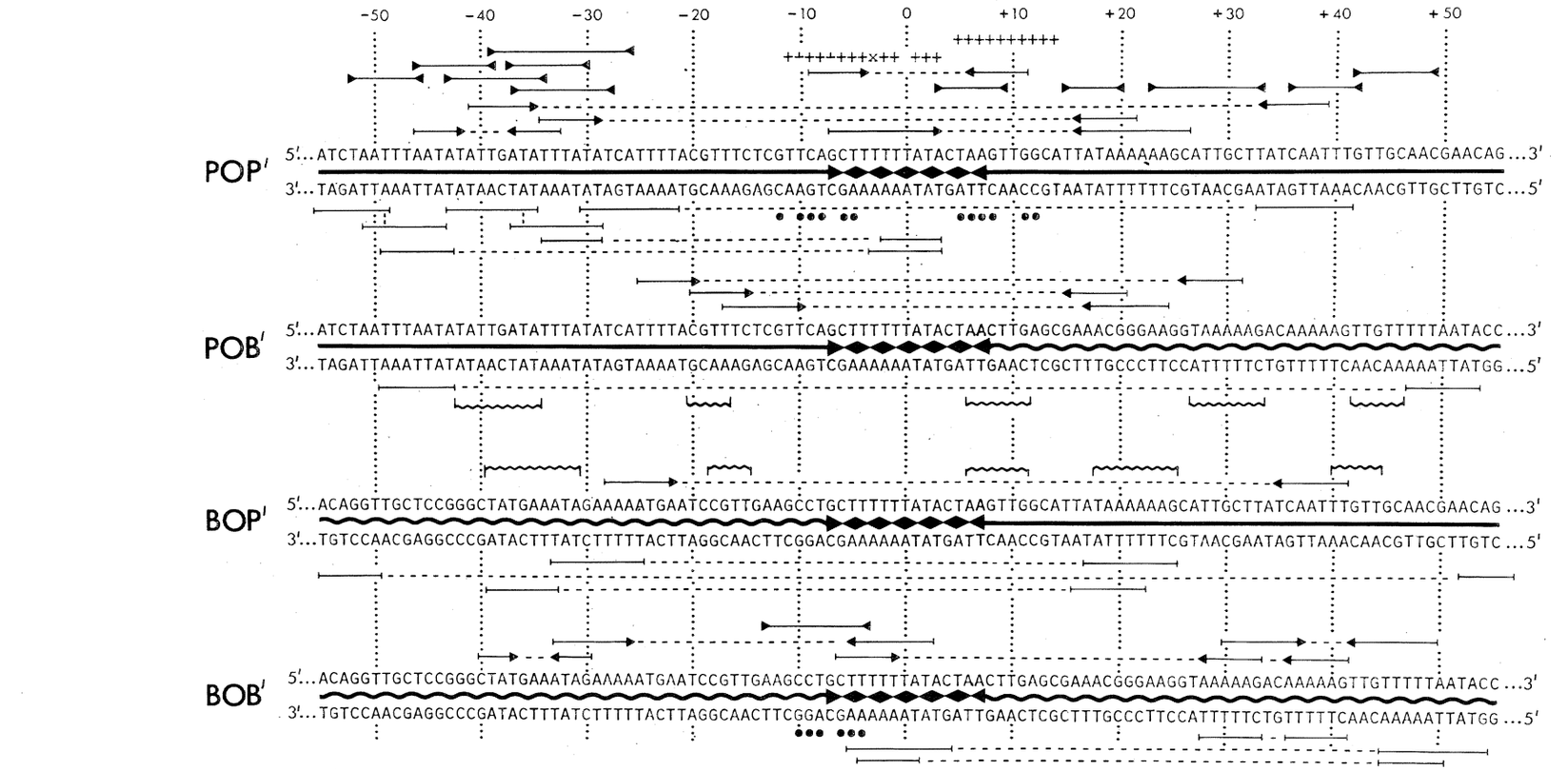
Three of the four *att* sites have inverted repeats that are centered to varying degrees around the common core region, with the bacterial *att* site, BOB′, lacking such a configuration. The most symmetrically disposed inverted repeat occurs in the phage *att* site with element centers at −38 and +36. It may be interesting that both of the symmetrical inverted repeats in the phage *att* site have one element within the region of the P arm which also has a high concentration of palindromic sequences. Among the inverted repeats that are not centered with respect to the center core region, there is at least one in each of the four *att* sites which involves a portion of the core region (see Fig. 7).

Features of the Core-Arm Junctures

The existence of the 15-nucleotide-pair common core region could be taken as evidence in support of the genetic prediction that integration and excision proceed by way of staggered cuts in the two DNA strands of each recombining molecule (*19*). If, in fact, this is the primary significance of the common core region, then we would expect that one cut is made in each strand at the two opposite ends of the common core sequence.

Examination of the *att* sequences at the core-arm junctures for features that might be suggestive of cut sites at these loci reveals a very short inverted repeat that also contains a very short (and "degenerate") molecular palindrome (*45*). However, both of these sequence features are too short (or too degenerate) to even meet the criteria for inclusion in Fig. 7 (see legend); it is only their placement at the core-arm junctures that warrants attention.

In contrast to these weak patterns at the four core-arm junctures, we would like to point out that three very distinctive features of the *att* site sequences involve the OP′ juncture—thus suggest-


POP′
POB′
BOP′
BOB′
5′...ATCTAATTTAATATATTGATATTTATATCATTTTACGTTTCTCGTTCAGCTTTTTTATACTAAGTTGGCATTATAAAAAAGCATTGCTTATCAATTTGTTGCAACGAACAG...3′
3′...TAGATTAAATTATATAACTATAAATATAGTAAAATGCAAAGAGCAAGTCGAAAAAATATGATTCAACCGTAATATTTTTTCGTAACGAATAGTTAAACAACGTTGCTTGTC...5′
5′...ATCTAATTTAATATATTGATATTTATATCATTTTACGTTTCTCGTTCAGCTTTTTTATACTAACTTGAGCGAAACGGGAAGGTAAAAAGACAAAAAGTTGTTTTTTAATACC...3′
3′...TAGATTAAATTATATAACTATAAATATAGTAAAATGCAAAGAGCAAGTCGAAAAAATATGATTGAACTCGCTTTGCCCTTCCATTTTTCTGTTTTTCAACAAAAATTATGG...5′
5′...ACAGGTTGCTCCGGGCTATGAAATAGAAAAATGAATCCGTTGAAGCCTGCTTTTTTATACTAAGTTGGCATTATAAAAAAGCATTGCTTATCAATTTGTTGCAACGAACAG...3′
3′...TGTCCAACGAGGCCCGATACTTTATCTTTTTACTTAGGCAACTTCGGACGAAAAAATATGATTCAACCGTAATATTTTTTCGTAACGAATAGTTAAACAACGTTGCTTGTC...5′
5′...ACAGGTTGCTCCGGGCTATGAAATAGAAAAATGAATCCGTTGAAGCCTGCTTTTTTATACTAACTTGAGCGAAACGGGAAGGTAAAAAGACAAAAAGTTGTTTTTTAATACC...3′
3′...TGTCCAACGAGGCCCGATACTTTATCTTTTTACTTAGGCAACTTCGGACGAAAAAATATGATTGAACTCGCTTTGCCCTTCCATTTTTCTGTTTTTCAACAAAAATTATGG...5′

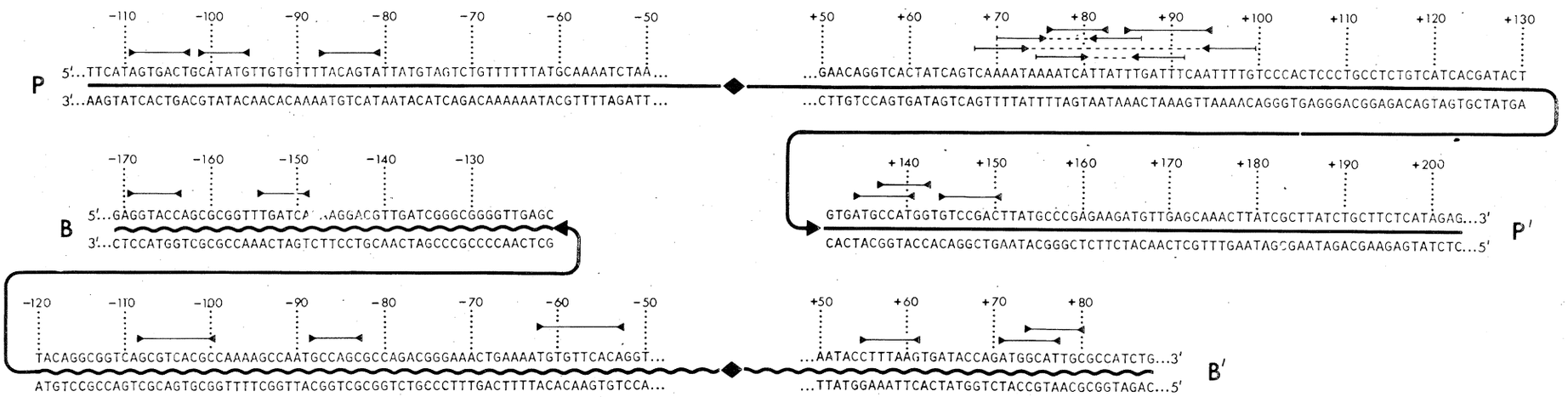

P
5′...TTCATAGTGACTGCATATGTTGTGTTTTACAGTATTATGTAGTCTGTTTTTTTATGCAAAATCTAA...
3′...AAGTATCACTGACGTATACAACACAAAATGTCATAATACATCAGACAAAAAATACGTTTTAGATT...
...GAACAGGTCACTATCAGTCAAAATAAAATCATTATTTGATTTCAATTTTGTCCCACTCCCTGCCTCTGTCATCACGATACT
...CTTGTCCAGTGATAGTCAGTTTTATTTTAGTAATAAACTAAAGTTAAAACAGGGTGAGGGACGGAGACAGTAGTGCTATGA
B
5′...GAGGTACCAGCGCGGTTTGATCA...AGGACGTTGATCGGGCGGGGTTGAGC
3′...CTCCATGGTCGCGCCAAACTAGTCTTCCTGCAACTAGCCCGCCCCAACTCG
GTGATGCCATGGTGTCCGACTTATGCCCGAGAAGATGTTGAGCAAACTTATCGCTTATCTGCTTCTCATAGAG...3′
CACTACGGTACCACAGGCTGAATACGGGCTCTTCTACAACTCGTTTGAATAGCGAATAGACGAAGAGTATCTC...5′
P′
TACAGGCGGTCAGCGTCACGCCAAAAGCCAATGCCAGCGCCAGACGGGAAACTGAAAATGTGTTCACAGGT...
ATGTCCGCCAGTCGCAGTGCGGTTTTCGGTTACGGTCGCGGTCTGCCCTTTGACTTTTACACAAGTGTCCA...
...AATACCTTTAAGTGATACCAGATGGCATTGCGCCATCTG...3′
...TTATGGAAATTCACTATGGTCTACCGTAACGCGGTAGAC...5′
B′

Fig. 7. Sequences determined in the phage, prophage, and bacterial *att* site regions. The sequences are numbered with zero as the central base of the common region. Sequence extending rightward from zero into the P′ and B′ arms is assigned positive numbers, and leftward from zero into the P and B arms negative numbers. (——) Phage DNA; (⌒⌒⌒) bacterial DNA; (◆◆) the common core region. The top section of the figure includes the central 111 bases of each of the four *att* sites, 55 bases to each side of zero. The lower section includes the distal portions of the four arms, P and B from −50 toward the left, and P′ and B′ from +50 toward the right. (There is a 5-base overlap of each arm with the top section.) In the lower section, the central 101 bases are represented by the appropriate symbols. Three types of structural features are identified in the sequence. Molecular palindromes are indicated by (▶—◀) above the corresponding sequences in both the top and lower sections. There are no palindromes meeting the set criteria (see below) which are unique to the prophage *att* sites. Direct repeats are only marked in the top section and are indicated below the corresponding sequences by (⟵⟶). Dashed lines connect the pairs of repeats. In the prophage sequences only those direct repeats (see below for criteria) which are unique to these sequences, that is, those with elements in P and B′ or B and P′, are shown. Inverted repeats are only marked in the top section and are indicated by (⟶▶) above the corresponding sequence. The two elements of each inverted repeat are connected by dashed lines. The only inverted repeats included in the lower section are the three that occur in the proposed *int* gene termination region (see text). The criteria used in marking each of the above sequence features were a minimum of six base pairs with no mismatches, a single central mismatch with at least three base pairs to each side, or two central mismatches with at least four base pairs to each side. Three sequences, of special interest because of their location, have been marked as inverted repeats even though they do not meet the stated criteria. Two of these occur from −46 to −33 and from −40 to −30 in the P and B arms, respectively. The third sequence involves the core-arm junctures in the phage *att* site (see text). Direct and inverted repeats, with elements often separated by considerable distance, also occur in the distal portions of the arms (lower section) but have not been included. In the central 111 base pairs, similar sequences ("identities") occurring in approximately the same positions in the P and B arms, and in the P′ and B′ arms, are marked between the two prophage sequences by (⌒⌒⌒). A sequence of 10 bases (+5 to +14) in the phage *l* strand (marked +++) is identical to part of an inverted repeat structure in one arm of IS*l* (*47*). A similar but less perfect homology exists between the *r* strand (+3 to −11) and the other element of the same inverted repeat structure that occurs in the other IS*l* arm (+, exact match; ⊥, purine or pyrimidine match; and ×, A-T base pair). Homology between the core-arm junctures and the 3′ end of 16S ribosomal RNA are indicated by (•) under the matching bases (see *69*).

ing the possibility of a special or important role for this particular region.

The largest inverted repeat seen in all of the sequences that we have determined has one 11-base-pair element comprised of a large part of the common core region (−7 to +3); the other 11-base-pair element, separated by an interval of 12 nucleotide pairs, occurs in the P′ arm (+26 to +16). Both elements of this 11-base-pair inverted repeat contain the sequence T_6A, which is associated with termination of transcription at those sites which do not require the protein factor ρ (*46*). The *E. coli* promoter sequence, TATPuATG (*47*), also overlaps the boundaries of both repeat elements. Although in vitro experiments indicate that RNA synthesis per se is not required for integration or excision (*9*, *10*), it is possible that RNA polymerase or some related element plays a role which does not actually require transcription. This unusual concentration of potential RNA polymerase binding sites in the core and P′ arm could also be the reflection of an evolutionary relation between RNA polymerase subunits and some of the proteins now involved in site-specific recombination. Other functions that might be assigned to these sequences involve the regulation of transcription into and out of the prophage when it is integrated into the *E. coli* chromosome. This function provides an interesting analogy with the effect of IS-elements on the expression of nearby chromosomal genes [see (*2*, *25*)].

The second distinctive feature is a 7-base-pair molecular palindrome that overlaps the OP′ juncture and extends from +3 in the core region to +9 in the P′ arm.

Finally, the third feature of the OP′ juncture is a significant region of homology with sequences in the arms of the IS*l* insertion sequence. Grindley (*48*) has found that the first 23 nucleotides at each end of the IS*l* sequence constitute an imperfect inverted repeat (18 matches out of 23 base pairs in each element of the repeat). In Fig. 7 we have indicated the extent and nature of the homology between the IS*l* arms and the core-arm juncture. There is a perfect 10-base-pair match between one arm of IS*l* and the OP′ juncture (+5 to +14). The match between the second IS*l* arm and the PO juncture is not as extensive (see Fig. 7); however, by deleting three contiguous bases from the second IS*l* sequence, one obtains a perfect 13-base-pair match with the OP′ juncture (not shown).

This homology raises interesting speculations about a functional or evolutionary relationship between *int* protein and the proteins which act on IS*l* or the involvement of these sequences in recognition by host proteins. It will also be interesting to see whether this sequence homology is related to the earlier observation that at least one secondary insertion site for λ on the *E. coli* chromosome (in segment six of the *gal*T gene) is genetically inseparable from a site frequently occupied by IS sequences (*49*).

These three sequence features, which occur at the OP′ juncture, lead to the suggestion that site-specific recombination may not be a reaction between two partners of equal importance or specificity; in particular, the phage *att* site, POP′, would have a special role in integrative recombination, and the left prophage *att* site, BOP′, would have a similar special role in excisive recombination.

Would a bias in the reactivity of the recombining sites be consistent with other available data? Parkinson (*50*) has pointed out, on the basis of suggestive (although not conclusive) genetic experiments, that efficient *int*-promoted recombination seems to require that at least one of the recombining *att* sites contains P′. There are, however, some unpublished experiments that are not in agreement with this conclusion (*51*). A special role for the core-P′ region is also consistent with the segregation patterns obtained by Shulman *et al.* (*19*, *20*) for a number of independently isolated *att* site mutants. In each case examined, the type of segregation pattern observed depends on whether the P′ arm is associated with the mutant or the wild-type parent.

If the *int* protein interacts more strongly with the OP′ region, one might expect to find this reflected in binding studies in vitro with phage DNA's carrying different *att* sites. Kotewicz, Chung, Takeda, and Echols (*52*) have found that *int* protein does indeed bind most efficiently to the phage *att* site, POP′, and the left prophage *att* site, BOP′. The right prophage *att* site POB′, binds very poorly. The only finding from this study which is not consistent with the above model, that BOB′ also binds to *int* protein, is not corroborated by the results of either Nash or Kamp (*53*). They find that *int* protein binds BOB′ much less well than it binds either *att*P or *att*L, which is consistent with our suggestion.

Thus, while there is information from other experiments that would be consistent with a special role for OP′, resolution of this question is to be found in the results of further experiments.

A final interesting feature, which we examined in conjunction with prelimi-

nary studies by Nash on the nature of host factors required for integrative recombination in vitro, is a homology between the core-arm junctures and the sequence at the 3′ end of 16S ribosomal RNA. The extent of matching observed is equal to, or better than, that found with the ribosomal binding sites of many mRNA's (see Fig. 7 and legend); however, the significance of this homology remains to be determined.

The *int* Gene

It has been suggested, from the properties of the deletion mutant *att*501 (*20*), that the distal end of the *int* gene is probably within 50 nucleotides of some functional region of the *att* site (*7*). This deletion is too small to be seen in the electron microscope, yet it both destroys *att* site function (*20*) and fails to recombine with several *int* amber mutations (*7*). We have examined this point by noting the reading frame of all chain termination triplets in the DNA strand which would correspond to the messenger RNA (mRNA) of the *int* gene, that is, the *r* strand of the P′ arm. As is seen in Fig. 9, there are two adjacent chain termination triplets at positions +83 and +80, which are not preceded (in the 120 nucleotides sequenced thus far) by any other chain termination signals in the same reading frame. In the downstream direction, within the next 25 nucleotide pairs, this reading frame generates three more chain termination triplets (+65, +62, and +59), which are again in the tandem arrangement frequently observed at the chain termination signals of several other *E. coli* phage genes (*54*, *55*). If the genetic and electron microscope data are taken at face value, neither of the other two reading frames that have chain termination triplets as far upstream as positions +187 and +198, respectively, could code for the *int* protein. Further confirmation of these conclusions should be available as a natural consequence of our present efforts to analyze the sequence of significant portions of the *int* and *xis* genes.

In the region of the postulated *int* termination signal, between +76 and +94, there are only two base pairs that do not participate in a palindromic sequence seven or more base pairs in length. This region is also bracketed by three inverted repeats (Fig. 7). Two of these, if considered together, are capable of forming a stem and loop structure, with 12 complementary bases in the stem and 5 unpaired bases in a loop that includes the first chain termination codon. The significance of this stem and loop may be related to the observed enhancement of termination site function by such structures (*56*). The strongest candidate for the RNA polymerase termination site of the *int* messenger RNA, GCT_6AT (*46*), occurs at the +20 position, 62 bases downstream from the chain termination codon. If this site is functional, it would have to be sufficiently leaky to account for some prophage transcription from *int* across *att*L into adjacent bacterial genes (*57*). The existence of an RNA polymerase termination site not far downstream from the *int* gene is consistent with two observations: (i) that the rate of leftward transcription in the P′ arm decreases beyond the *att* site (*58*) and (ii) that constitutivity of an *int*C mutation does not extend beyond the *int* gene into the *b*2 region (*11*). Both of these results, however, could be affected by mRNA processing (*59*), and further experiments are needed to clarify this question.

It should also be noted that the potential RNA polymerase termination sequence, T_6A, is also found at two places on the *l* strand at positions −65 and −3. If functional, these sites would serve to terminate transcription coming from the *b*2 region toward the *int* gene.

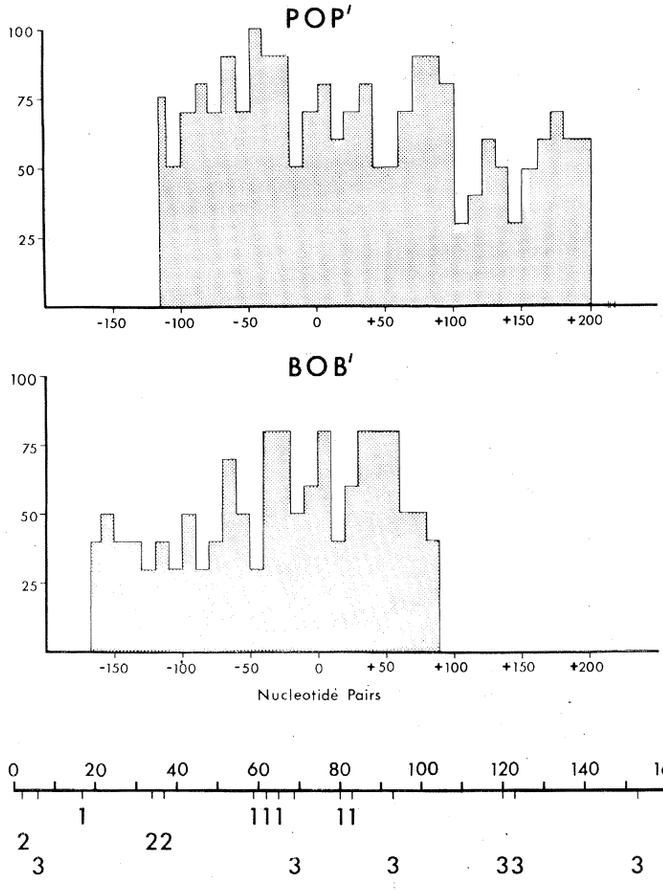


Fig. 8. Base composition of phage and bacterial *att* sites. Average base composition of blocks of ten base pairs, beginning at zero (0 to 10, 11 to 20, and so on) is represented as the percent of A+T.

Fig. 9. Potential chain termination triplets for the *int* gene. The positions of all chain termination codons (UAA, UGA, or UAG) are shown for each of the three reading frames in the *r* strand (corresponding to the mRNA) of the P′ arm. Numbering of the P′ arm is as described for Fig. 7; U, uridine.

Summary

The results shown in Fig. 7 provide a framework for experiments and models pertinent to the mechanism of site-specific recombination. Utilizing the large amount of information that has been accumulated about the integration-excision reaction of bacteriophage λ, as well as those general insights derived from the analysis of numerous other nucleic acid sequences, we have attempted to identify those features of the *att* site sequences that are most likely to be relevant in the functional interactions of this DNA region.

The actual site of the crossover for both integration and excision must take place within, or at the boundaries of, the 15-base-pair sequence that is common to all four *att* sites. The common core region provides exactly the structure suggested by Shulman and Gottesman (*19*)

on the basis of their genetic studies of *att* site mutants and could be viewed as support for the suggestion that integration and excision proceed via staggered cuts in the two DNA strands of each recombining *att* site. However, one interpretation of the *att* site sequence features that is particularly attractive places special emphasis on the OP′ juncture and suggests that the primary significance of the common core region is as a protein recognition site (or part of one) (*60*). According to this view, there may be staggered cuts, but they could lie close to one another. Alternatively, the crossover event need not involve staggered cuts at all, and the genetic results of Shulman and co-workers (*19*, *20*) primarily reflect a particular class of mutant *att* sites rather than the behavior of the wild-type *att* sites.

Several distinctive sequence features overlap the common core and the P′ arm: the largest inverted repeat in the sequence, a 7-base-pair molecular palindrome, and a strong homology with both arms of IS*1* (*48*). These observations lead to the suggestion that possibly the phage *att* site (POP′) in integrative recombination and the left prophage *att* site (BOP′) in excisive recombination have a special role in terms of recognition or initiation of the reactions.

There would also seem to be some special function associated with the P arm, as suggested by its unusual concentration of molecular palindromes and direct repeats, which also coincide with "identities" and the elements of inverted repeats. The bacterial *att* site is by comparison sparse in such sequence features, and a less "active" role seems consistent with the high frequency (and implied lack of specificity) of secondary bacterial *att* sites.

There are two other possibly interesting features of the *att* site sequences. One is the occurrence of potential RNA polymerase recognition sites, both in the form of termination signals and sequences associated with *E. coli* promoters. These sequences may, of course, be fortuitous, or vestiges of evolutionary significance. However, if functionally significant, these sites would probably be enhanced by the extensive bias for A·T base pairs in this region (*39*, *61*). The latter is also a feature that is probably related to the requirement for negatively supercoiled DNA as a substrate for the integration reaction (*37*). The second feature is a homology between sequences at the core-arm juncture and the 3′ end of 16*S* ribosomal RNA (*69*).

Many of these questions should be answered, or at least better defined, by extension of the work reported here. Sequence analysis of mutant and secondary *att* sites, including *int*-dependent phage deletions, will be quite helpful in identifying critical features of the *att* structures. The precise internucleotide bonds which are cut during recombination remain to be determined. Interactive studies with purified *int* protein (*9*, *52*) and relevant host proteins are now possible and will be crucial in putting together a more complete picture at the molecular level of this model system for site-specific recombination.

References and Notes

1. D. Reanney, *Bacteriol. Rev.* **40**, 552 (1976); N. Battula and H. M. Temin, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 281 (1977).
2. A. Bukhari, J. Shapiro, S. Adhya, Eds., *DNA Insertion Elements, Plasmids, and Episomes* (Cold Spring Harbor Press, Cold Spring Harbor, N.Y., in press).
3. M. E. Gottesman and R. A. Weisberg, in *The Bacteriophage Lambda*, A. D. Hershey, Ed. (Cold Spring Harbor Press, Cold Spring Harbor, N.Y., 1971), p. 113; H. A. Nash, *Current Top. Microbiol. Immunol.*, in press.
4. A. Campbell, *Adv. Genet.* **11**, 101 (1962).
5. J. Zissler, *Virology* **31**, 189 (1967); R. Gingery and H. Echols, *Proc. Natl. Acad. Sci. U.S.A.* **58**, 1507 (1967); M. E. Gottesman and M. B. Yarmolinsky, *J. Mol. Biol.* **31**, 487 (1968).
6. G. Guarneros and H. Echols, *J. Mol. Biol.* **47**, 565 (1970); A. Kaiser and T. Masuda, *ibid.*, p. 557.
7. L. W. Enquist and R. A. Weisberg, *ibid.* **111**, 97 (1977).
8. H. I. Miller and D. I. Friedman, in (*2*); J. G. K. Williams, D. L. Wulff, H. A. Nash, in (*2*); A. Kikuchi and R. A. Weisberg, personal communication.
9. H. A. Nash, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 1072 (1975).
10. S. Gottesman and M. E. Gottesman, *ibid.*, p. 2188.
11. N. Katzir, A. Oppenheim, M. Belfort, A. B. Oppenheim, *Virology* **74**, 324 (1976).
12. S. Chung and H. Echols, *ibid.*, in press; D. Court, S. Adhya, H. A. Nash, L. W. Enquist, in (*2*).
13. H. Echols, *J. Mol. Biol.* **47**, 575 (1970).
14. F. Guerrini, *ibid.* **46**, 523 (1969).
15. H. Echols and D. Court, in *The Bacteriophage Lambda*, A. D. Hershey, Ed. (Cold Spring Harbor Press, Cold Spring Harbor, N.Y., 1971), p. 701.
16. R. A. Weisberg and M. E. Gottesman, in *ibid.*, p. 489; G. Guarneros and H. Echols, *Virology* **52**, 30 (1973); E. R. Signer, J. Weil, P. Kimball, *J. Mol. Biol.* **46**, 543 (1969).
17. R. W. Davis and J. S. Parkinson, *J. Mol. Biol.* **56**, 403 (1971).
18. For example, as was pointed out by Shulman *et al.* (*20*), two recognition elements suffice to create four different *att* sites—POP′, NOP′, PON, and NON, where N is any sequence except P or P′.
19. M. J. Shulman and M. E. Gottesman, *J. Mol. Biol.* **81**, 461 (1973).
20. M. J. Shulman, K. Mizuuchi, M. M. Gottesman, *Virology* **72**, 13 (1976).
21. Z. Hradecna and W. Szybalski, *ibid.* **38**, 473 (1969).
22. K. Shimada, R. A. Weisberg, M. E. Gottesman, *J. Mol. Biol.* **63**, 483 (1972).
23. ———, *ibid.* **93**, 415 (1975).
24. M. Fiandt, Z. Hradecna, H. A. Lozeron, W. Szybalski, in *The Bacteriophage Lambda*, A. D. Hershey, Ed. (Cold Spring Harbor Press, Cold Spring Harbor, N.Y., 1971), p. 329; M. H. Malamy, M. Fiandt, W. Szybalski, *Mol. Gen. Genet.* **119**, 207 (1972); H. J. Hirsch, P. Starlinger, P. Brachet, *ibid.*, p. 191.
25. For a recent review of IS-elements, see P. Starlinger and H. Saedler, *Current Top. Microbiol. Immunol.* **75**, 111 (1976).
26. Similar observations and conclusions also come from the results of Davis and Parkinson (*17*) with *int*-promoted deletions extending rightward (POΔ′) or leftward (ΔOP′) from the phage *att* site, and it has been suggested that these are analogous to secondary *att* sites (*23*).
27. D. Nathans and H. O. Smith, *Annu. Rev. Biochem.* **44**, 273 (1975).
28. R. J. Roberts, *CRC Crit. Rev. Biochem.* **4**, 123 (1976).
29. J. C. Marini and A. Landy, *Virology* **76**, 196 (1977); J. C. Marini, R. A. Weisberg, A. Landy, *ibid.*, in press.
30. L. Robinson and A. Landy, *Gene* **2**, 1 (1977); *ibid.*, p. 33.
31. P. D. Sadowski and J. Hurwitz, *J. Biol. Chem.* **244**, 6192 (1969); A. Bernardi, J. Maat, A. deWaard, G. Bernardi, *Eur. J. Biochem.* **66**, 175 (1976).
32. W. Ross, A. deWaard, A. Landy, unpublished results.
33. A. M. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
34. G. C. Hayward, *Virology* **49**, 342 (1972); P. A. Sharp, D. H. Gallimore, S. J. Flint, *Cold Spring Harbor Symp. Quant. Biol.* **39**, 547 (1974).
35. W. Davies, personal communication.
36. A. Skalka, E. Burgi, A. D. Hershey, *J. Mol. Biol.* **34**, 1 (1968); R. B. Inman and M. Schnos, *ibid.* **49**, 93 (1970).
37. K. Mizuuchi and H. A. Nash. *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3524 (19 → M. Gellert, M. H. O'Dea, T. Itoh, J. Tomizawa, *ibid.*, p. 4474.
38. C. Brack, T. A. Bickle, R. Yuan, *J. Mol. Biol.* **96**, 693 (1975).
39. P. Botchan, *ibid.* **105**, 161 (1976).
40. S. Adhya, personal communication.
41. T. Maniatis and M. Ptashne, *Nature (London)* **246**, 133 (1973).
42. A. Gierer, *ibid.* **212**, 1480 (1966).
43. J. C. Wang, M. D. Barkley, S. Bourgeois, *ibid.* **251**, 247 (1974).
44. H. M. Sobell, *Adv. Genet.* **17**, 411 (1973); *Proc. Natl. Acad. Sci. U.S.A.* **72**, 279 (1975).
45. Three of the four core-arm junctures (−4 to −9 on the left and +6 to +11 on the right) are comprised of inverted repeats that correspond to the generalized sequence 5′ . . . AA$^{G}_{C}$PyTG (where Py is either pyrimidine); the fourth at the BO juncture differs from this by one base pair (Fig. 7). The palindrome within the elements of these inverted repeats is AA$^{G}_{C}$TT (+6 to +10) at the OP′ and OB′ junctures, and the PO and BO junctures (−4 to −8) the related sequences are AAGCT and AAGCA, respectively.
46. Sequence GCT$_6$AT is also found in the φX174 genome at a postulated RNA polymerase termination site (*54*), and the "oop" RNA of phage P22 terminates with the sequence GCT$_{6-7}$ (M. Rosenberg, personal communication). The more general sequences T$_6$AT and T$_6$A have been found at the termini of a number of in vitro and in vivo transcripts [P. Lebowitz, S. M. Weissman, C. M. Radding, *J. Biol. Chem.* **246**, 5120 (19 → G. Pieczenik, B. G. Barrell, M. L. Gefter, *Arch. Biochem. Biophys.* **152**, 152 (1972); J. E. Dahlberg and F. R. Blattner, *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **32**, 664 (1973); T. Ikemura and J. E. Dahlberg, *J. Biol. Chem.* **248**, 5024 (1973); K. Bertrand, L. Korn, F. Lee, T. Platt, C. L. Squires, C. Squires, C. Yanofsky, *Science* **189**, 22 (1975); M. Rosenberg, B. deCrombrugghe, S. M. Weissman, *J. Biol. Chem.* **250**, 4755 (1975)]. Transcription termination at this site does not require, but is enhanced by [M. Rosenberg, B. deCrombrugghe, S. M. Weissman, *ibid.*, p. 4755], the protein termination factor ρ [J. W. Roberts, *Nature (London)* **224**, 1168 (1969)]. These sites also tend to be preceded by a GC-rich region of variable length, but in the case of the *att* sequences this feature is not very pronounced.
47. D. Pribnow [*Proc. Natl. Acad. Sci. U.S.A.* **72**, 784 (1975)] has pointed out that in all *E. coli* promoters there is a homologous sequence approximately ten base pairs upstream of the start point of transcription. Several mutations that increase the efficiency of the lac promoter provide further definition of the "ideal" sequence to yield TATAATPu (where Pu is either purine) [see W. Gilbert, in *RNA Polymerase*, R. Losick and M. Chamberlin, Eds. (Cold Spring Harbor Press, Cold Spring Harbor, N.Y., 1976), p. 193]. The majority of *E. coli* promoters sequenced thus far (including four from λ, three from φX174, three from fd, two from T7, one in SV40, and the trp promoter) have the sequence GTTG (or an analog differing in no more than one position) approximately 22 base pairs upstream from the "Pribnow sequence." (The lac and tRNA^{tyrT} promoters are among those that do not have this sequence.) For the two "Pribnow sequences" in the OP′ inverted repeat, the GTTG feature is weak; on the left, GTTT occurs 18 base pairs

upstream, and on the right GTTG occurs 28 base pairs upstream.
48. N. Grindley, personal communication; in (2); see also E. Ohtsubo, in (2).
49. K. Shimada, R. A. Weisberg, M. E. Gottesman, *J. Mol. Biol.* **80**, 297 (1973).
50. J. S. Parkinson, *ibid.* **56**, 385 (1971).
51. R. A. Weisberg, personal communication.
52. M. Kotewicz, S. Chung, Y. Takeda, H. Echols, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 1511 (1977).
53. H. A. Nash, personal communication; D. Kamp, manuscript in preparation.
54. F. Sanger *et al.*, *Nature (London)* **265**, 687 (1977).
55. W. Min Jou, G. Haegeman, M. Ysebaert, W. Fiers, *ibid.* **237**, 82 (1972).
56. M. Rosenberg, personal communication.
57. K. Shimada and A. Campbell, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 237 (1974).
58. H. J. J. Nijkamp, K. Bovre, W. Szybalski, *Mol. Gen. Genet.* **111**, 22 (1971).
59. H. A. Lozeron, J. E. Dahlberg, W. Szybalski, *Virology* **71**, 262 (1976).
60. Shulman *et al.* (*20*) have pointed out that even when two *att* sites both carry the same mutation they do not undergo efficient *int*-dependent recombination (*19*). This suggests that, if the mutation is in a common core region, perfect homology is not sufficient for *int*-dependent recombination; there must also be a specific sequence, for example, for recognition by a protein such as *int*.
61. J. M. Saucier and J. Wang, *Nature (London) New Biol.* **239**, 167 (1972).
62. A. Landy, E. Ruedisueli, L. Robinson, C. Foeller, W. Ross, *Biochemistry* **13**, 2134 (1974).
63. T. J. Kelly and H. O. Smith, *J. Mol. Biol.* **51**, 393 (1970).
64. R. Old, K. Murray, G. Roizes, *ibid.* **92**, 331 (1975).
65. R. J. Roberts. P. A. Myers, A. Morrison, K. Murray, *ibid.* **103**, 199 (1976).
66. D. E. Garfin and H. M. Goodman, *Biochem. Biophys. Res. Commun.* **59**, 108 (1974).
67. R. J. Roberts, P. A. Myers, A. Morrison, K. Murray, *J. Mol. Biol.* **102**, 157 (1976).
68. I. M. Glynn and J. B. Chappell, *Biochem. J.* **90**, 147 (1964).
69. Portions of the sequence at the 3′ end of 16*S* ribosomal RNA, GAUCACCUCCUUA$_{OH}$, are complementary to mRNA sequences associated with ribosomal binding and the initiation of protein synthesis. The extent of complementarity ranges from three to nine bases, depending upon the particular mRNA [J. Shine and L. Dalgarno, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1342 (1974); K. U. Sprague and J. A. Steitz, *Nucl. Acids Res.* **2**, 787 (1975)]. A number of *E. coli* proteins are involved in this site-specific recognition. For a recent review see J. A. Steitz *et al.*, in *Nucleic Acid-Protein Recognition*, H. J. Vogel, Ed. (Academic Press, New York, 1977), p. 491.
70. We thank A. M. Maxam and W. Gilbert for communicating their sequencing procedure prior to its publication, R. J. Roberts for information on the purification of restriction enzymes, A. deWaard for gifts of endonuclease IV, C. Foeller and M. McNamara for technical assistance, and C. Chute for computer programming. We also thank M. J. Shulman, H. A. Nash, R. A. Weisberg, L. W. Enquist, M. E. Gottesman, M. Rosenberg, N. Grindley, H. Echols, and D. Kamp for communicating their results prior to publication and for stimulating conversations. Supported by NIH grant CA11208 and by grant 1-543 from The National Foundation–March of Dimes. A.L. is a Faculty Research Associate of the American Cancer Society, Inc.

Disasters as a Necessary Part of Benefit-Cost Analyses

Water-project costs should include the possibility of events such as dam failures.

R. K. Mark and D. E. Stuart-Alexander

Although it is well known that some low-probability events can be very costly, benefit-cost analyses for water projects generally have not included the probable value of these costs. The significance of expected costs is reflected in the growing difficulty of obtaining liability insurance against dam failure, even at highly inflated prices (*1*). This article is intended to stimulate discussion of the need to include events such as dam failures, impoundment-induced earthquakes, and landslides in the benefit-cost analyses of reservoir projects. It is not concerned with actual methods of estimation (*2*).

Dam Failures

Dam failures are not uncommon. Gruner (*3*) reported that 33 dams failed in the United States between 1918 and 1958; five of these were major disasters involving the loss of 1680 lives. He stated that these 33 were part of a list of 1764 dams built prior to 1959. Other references list about 1000 more dams completed by 1959 (*4*, *5*), so that we are assuming an average of 1600 dams over this 40-year period (*5*). These data suggest a failure rate of approximately 5×10^{-4} per dam-year and a major disaster rate of approximately 0.8×10^{-4} per dam-year. Kiersch (*6*) reported that from 1959 to 1965 "nine major dams of the world have failed in some manner." In 1962 there were 7833 major dams (*7*), indicating a worldwide failure rate of about 2×10^{-4} per dam-year for that period. In 1976 there were six dam failures, four of which are considered major disasters, resulting in significant property damage and a total of more than 700 deaths. Dam failures have generally resulted from design, construction, or site inadequacies, or from natural phenomena, primarily storms or earthquakes.

Generalized estimates of dam failure probabilities can be based on historical frequency observations, either aggregated (as above) or, if sample size permits, disaggregated into categories and time periods (*5*). Historical trends may result from the balance between improving technology and the need to use more difficult dam sites. For instance, calculations based on data compiled by the Committee on Failures and Accidents to Large Dams (*5*) indicate that the U.S. major failure rate did decline by about an order of magnitude during the first four decades of this century, but has since fluctuated between about 1×10^{-4} and 2×10^{-4} per dam-year.

The number of failures in small populations of dams are generally insufficient to give precise estimates of failure rates, nevertheless it can be instructive to consider restricted populations. For example, the U.S. Bureau of Reclamation had accumulated approximately 4500 dam-years' experience on earth-filled dams for reservoirs in excess of 1000 acre-feet storage capacity (*8*), when its Teton Dam in Idaho failed on 5 June 1976. Terminating the sample at this first disastrous failure (*9*) (that is, an inverse binomial sample, with ~4500 dam-years to first failure), we obtain a median-unbiased estimate of the failure rate (*10*) as approximately

$$\frac{2}{3} \times \frac{1}{4500} = 1.5 \times 10^{-4} \text{ per dam-year}$$

Such an estimate, based on a single failure, has a very wide confidence interval; however, it is generally consistent with the other worldwide estimates.

Project-specific probabilities of failure as well as generalized probabilities of failure can be estimated by "fault tree" analysis of the probabilities of casual

R. K. Mark is a physical scientist and D. E. Stuart-Alexander is a geologist in the Branch of Western Environmental Geology at the U.S. Geological Survey, Menlo Park, California 94025.