Nucleic Acids Research, 1994, Vol. 22, No. 11 2071–2078

# Characterisation of specific and secondary recombination sites recognised by the integron DNA integrase

Gavin D.Recchia[1,2], H.W.Stokes[2] and Ruth M.Hall[1,*]
[1]CSIRO Division of Biomolecular Engineering, Sydney Laboratory, PO Box 184, North Ryde, NSW 2113 and [2]School of Biological Sciences, Macquarie University, NSW 2109, Australia

Received February 16, 1994; Revised and Accepted April 29, 1994

## ABSTRACT

**Integrons determine a site-specific recombination system which is responsible for the acquisition of genes, particularly antibiotic resistance genes. The integrase encoded by integrons recognises two distinct classes of recombination sites. The first is the family of imperfect inverted repeats, known as 59-base elements, which are associated with the mobile gene cassettes. The second consists of a single site into which the cassettes are inserted. This site, here designated *attI*, is located adjacent to the *int* gene in the recipient integron structure. The *attI* site has none of the recognisable features of members of the 59-base element family except for a seven-base core site, GTTRRRY, at the recombination crossover point. Using a conduction assay to quantitate site activity, the sequence required for maximal *attI* site activity was confined to a region of >39 and ≤70 bases. Both integrative and excisive site-specific recombination events involving *attI* and a 59-base element site were demonstrated, but no evidence for events involving two *attI* sites was obtained. Integrase-mediated recombination between a 59-base element and several secondary sites in pACYC184 with the consensus GNT occurred at low frequency, and such events could potentially lead to insertion of gene cassettes at many non-specific sites.**

## INTRODUCTION

Integrons are DNA elements that determine a site-specific recombination system which is able to capture genes (1, 2). The recombination system consists of two determinants, a gene (*int*) encoding a DNA integrase (Int) and an adjacent site recognised by the integrase, here designated the *attI* site, at which the incoming genes are incorporated. To be inserted at this site, incoming genes (usually antibiotic resistance genes) must be associated with a recombination site recognised by the integrase. The most extensively studied integron family is associated with the sulphonamide resistance determinant *sulI* (1). These integrons consist of a conserved sequence (Fig. 1), which is separated by the inserted gene(s) into two regions, the 5′-conserved segment (5′-CS) encoding the integrase and the 3′-conserved segment (3′-CS) encoding *sulI* (1, 3). A large number of distinct antibiotic resistance genes and several unidentified reading frames have been found in the insert region of integrons (see ref. 4). Also, there appears to be no limit to the number of insertion events that can occur as integrons containing several inserted genes have been found in nature. When inserted genes are present they are always in the same orientation with respect to the conserved segments and are transcribed from a common promoter(s) located in the 5′-CS.

The inserted genes are found in discrete units, known as gene cassettes, which include the gene coding region and a recombination site located at the 3′ end of the gene (2, 4–6). The gene cassettes have been shown to be independently mobilisable units which can be deleted or rearranged by Int (5). Cassettes can be excised as covalently closed circular molecules (6) and insertion of circular cassettes into integrons has also been demonstrated (2). The recombination sites associated with the inserted genes were first identified as a family of related imperfect inverted repeat sequences found at the 3′ end of genes and were named 59-base elements (7). Subsequently the 59-base element family was extended to include longer and more diverged elements (4), all of which contain imperfect inverted repeats and are related at their outer ends to a consensus sequence (4, 6). Several different 59-base elements have been shown to function as recombination sites recognised by Int and can participate in recombination events involving either a second 59-base element or *attI* (4–6, 8).

In *sulI*-associated integrons, the region in the conserved recipient structure required for activity of the site into which cassettes are inserted (*attI*) is located at the inner boundary of the 5′-CS (see Fig. 1), and evidence that this site participates in site-specific recombination events has been reported (2, 5, 8). The only obvious similarity between sequences near the inner end of the the 5′-CS and the recognisable features of 59-base elements is a seven-base core site with the consensus GTTAGAC or GTTRRRY (R = purine, Y = pyrimidine) found at the 3′ end of 59-base elements and at the junction of the 5′-CS with the 3′-CS or the first inserted gene (1, 4). The recombination crossover point has been localised to within or on either side of the conserved GTT of this core site (2, 4–6), and the

*To whom correspondence should be addressed

recombination event leading to insertion of a gene cassette into an integron containing no inserted genes thus separates the core sites from the remainder of the original extended recombination sites (59-base element or *attI*), leading to the formation of two composite sites. For simplicity, the complete core site has been assigned to either the 3′-CS or 5′ end of the inserted cassettes (4).

The *attI* site appears to be distinct from the 59-base element sites and it seems unlikely that the core site sequence would be sufficient to confer all of the features required for an active, specific recombination site. It has been proposed that a 14 base sequence, designated hs2, is sufficient for the activity of this site (9). However, these 14 bases were identified only by comparisons of the 5′-CS sequence with a limited number of 59-base element sites in the region of a proposed recombination crossover point which was incorrectly located, and no evidence that this 14 base sequence represents the full extent of the recombination site has been reported.

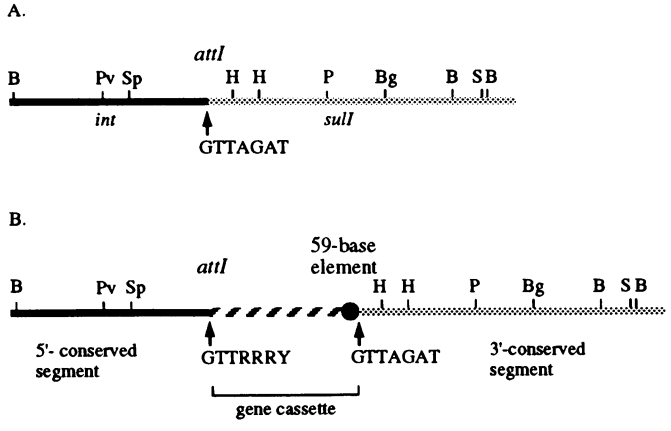


**Figure 1.** Structure of *sulI* integrons. (**A**) An integron with no inserted genes and (**B**) an integron with one inserted gene cassette. The 5′-conserved segment (5′-CS) is shown as a thick black line and the 3′-conserved segment (3′-CS) as a stippled line. The 5′-CS includes the DNA integrase gene, *int*, and the bulk of the *attI* site, and the 3′-CS includes a sulphonamide resistance gene, *sulI*. A 59-base element associated with the inserted cassette is represented by a solid black circle. The locations of the seven-base core recombination sites are indicated by vertical arrows and their sequences are indicated. Restriction sites shown are: B, *Bam*HI; Bg, *Bgl*II; H, *Hin*dIII; P, *Pst*I; Pv, *Pvu*II; S, *Sal*I; Sp, *Sph*I.

In this study, we have determined experimentally the extent of the region required for activity of the *attI* site. Site activity was assayed using the conduction assay developed by Martinez and de la Cruz (8, 10), which measures the formation of cointegrates between the IncW plasmid R388 ($Tra^+$) containing the integron In3 and a pACYC184-based plasmid ($Tra^-$, $Mob^-$) containing a cloned fragment which includes a recombination site. The fragments tested included various lengths of the 5′-CS together with adjacent sequences from the 3′-CS or an inserted cassette which provide the seven-base core site. A region of greater than 39 bases and no more than 70 bases was shown to specify the features required for the activity of *attI*. The *attI* site was also shown to participate in resolution events, but no evidence for recombination between two *attI* sites was obtained. Recombination catalysed by Int between the orfA 59-base element in R388 and non-specific sites in pACYC184 was also observed to occur at a very low frequency. A trinucleotide with the consensus GNT was found in the pACYC184 sequence at the crossover point. This type of event could lead to the insertion of gene cassettes into non-specific sites, and may be important in the dissemination of genes, particularly antibiotic resistance genes.

## MATERIALS AND METHODS

### Bacterial strains and plasmids

*Escherichia coli* strain JM101 *thi*, Δ(*lac-proAB*), [F′, *traD36*, *proAB*, $lacI^q$, *lacZΔM15*] (11) was used as a host for plasmid constructions. Strains UB1637 (*his*, *lys*, *trp*, *recA56*, *rpsL*) and UB5201 ($F^-$, *pro*, *met*, *recA56*, *gyrA*) (12) were used in cointegration and resolution experiments. Plasmids are listed in Table I. Appropriate *Hin*cII−*Hin*dIII and *Rsa*I−*Hin*dIII fragments from pRMH71 or pRMH232 were isolated and cloned into *Eco*RV−*Hin*dIII digested pACYC184 to generate pRMH95 and pRMH248 respectively. A *Taq*I fragment from pRMH232 extending from a *Taq*I site in the 5′-CS to a *Taq*I site in the vector 5 bases beyond the *Hin*dIII site was end-filled using the Klenow fragment of DNA polymerase (Pharmacia) and cloned into the *Eco*RV site of pACYC184. Recombinants were sequenced using the primer RH62 (pACYC184, complement of bases 1717−1733 in ref. 13) to determine that pRMH251 contains the *Taq*I fragment in the same orientation as the inserts in other plasmid constructs. pRMH500 was constructed by digestion of pRMH262 with *Hin*dIII and re-ligation to remove two *Hin*dIII fragments. R388

**Table I.** Plasmids

| Plasmid | Description | Relevant phenotype[a] | Reference |
|---|---|---|---|
| R388 | | $Tp^R$, $Su^R$, $Tra^+$ | 25 |
| pACYC184 | | $Cm^R$, $Tc^R$ | 26 |
| pSU2056 | 1176 bp *Rsa*I−*Bam*HI fragment of Tn*21* in pUC9 | $Ap^R$, $Int^+$ | 8 |
| pMAQ28 | 400 bp *Hin*dIII−*Sau*3A fragment of pDGO100 in pACYC184 | $Cm^R$ | 4 |
| pMAQ76 | Bal 31 generated deletion of pRMH251 | $Cm^R$ | This study |
| pMAQ79 | Bal 31 generated deletion of pMAQ76 | $Cm^R$ | This study |
| pRMH95 | 448bp *Hin*cII−*Hin*dIII fragment from pRMH71 in pACYC184 | $Cm^R$ | This study |
| pRMH232 | 770bp *Sph*I−*Hin*dIII fragment from pRMH71 in pACYC184 | $Cm^R$ | 2 |
| pRMH248 | 310bp *Rsa*I−*Hin*dIII fragment from pRMH232 in pACYC184 | $Cm^R$ | This study |
| pRMH251 | 266bp *Taq*I fragment from pRMH232 in pACYC184 | $Cm^R$ | This study |
| pRMH262 | Int-mediated insertion of *aadA2* into pRMH232 | $Cm^R$, $Sp^R$, $Sm^R$ | 2 |
| pRMH500 | 750bp *Sph*I−*Hin*dIII fragment from pRMH262 in pACYC184 | $Cm^R$ | This study |

[a]Ap, ampicillin; Cm, chloramphenicol; Sm, streptomycin; Sp, spectinomycin; Tc, tetracycline; Tp, trimethoprim; Su, sulphonamide.

was introduced into UB1637 by conjugation and all other plasmids were introduced into the appropriate strain by transformation, either using the calcium chloride procedure (14) or by electroporation using a Gene Pulsar (Bio-Rad). When pSU2056 was present, it was transformed into the strain last, and stocks were frozen in 50% glycerol at −70°C as soon as possible after purification of transformants to reduce the occurrence of deleterious Int-mediated recombination events.

### Bacterial culture conditions

Bacteria were cultured in L-broth (15) supplemented with the appropriate antibiotic(s). Solid medium contained 1.5% bacto-agar. Antibiotic concentrations used were; 25 μg/ml or 100 μg/ml of ampicillin (Ap), 25 μg/ml of chloramphenicol (Cm), 25 μg/ml of nalidixic acid (Nx), 25 μg/ml of spectinomycin (Sp), and 25 μg/ml of trimethoprim (Tp). Antibiotics were purchased from Sigma Chemical Co. (St Louis, MO).

### DNA procedures

Plasmid pRMH251 was purified for use as a substrate for Bal 31 nuclease digestion using a Qiagen-100 column (Diagen GmbH) according to manufacturer's instructions. For restriction enzyme digestion, plasmid DNA was isolated by the alkaline lysis method (14). Restriction fragments were analysed by electrophoresis in 0.8% agarose gels using *Eco*RI digested SPP-1 phage DNA (Bresatec, Adelaide, SA) and *Bam*HI digested R388 as size standards.

For DNA sequencing, plasmid DNA was isolated using the Magic Minipreps DNA Purification System (Promega) according to manufacturer's instructions. Sequencing was performed using synthetic primers, annealed to the plasmid template using the protocol of Jones and Schofield (16), and a Sequenase 2.0 DNA sequencing kit (United States Biochemical Corporation). Primers used to sequence cointegrate junctions were HS38 (3′-CS, complement of bases 380−395 in ref. 1) and HS94 (orfA, bases 302−317 in Fig. 4 of ref. 17).

### Bal 31 deletion

Plasmid pRMH251 was linearised with *Bam*HI and digested with 1 unit of Bal 31 nuclease (US Biolabs) at 25°C in a final volume of 50 μl of Bal 31 reaction buffer supplied by the manufacturer. Samples were removed at 1 min intervals at times between 10 and 20 min, and the reaction stopped by addition of EDTA to a final concentration of 50 mM and heating to 65°C for 15 min. Digestion products were extracted with an equal volume of phenol/chloroform (1:1) and ethanol precipitated. Fragments were end-filled with T4 DNA polymerase (Pharmacia) and ligated with T4 DNA ligase (Promega) at 4°C overnight. $Cm^R$ transformants were recovered and screened by restriction digestion. Plasmids which retained the *Sal*I site of pACYC184 were sequenced using the primer RH20 (3′-CS, complement of bases 116−131 in ref. 1) to determine the extent of deletion. A single plasmid, pMAQ76, which retains 32 bases of the 5′-CS and pACYC184 sequences from the *Hin*dIII site clockwise to base 1979 (13) was identified. A unique *Ban*II site is located in the pACYC184 sequence 6 bases beyond this junction, and pMAQ79 was generated by linearisation of pMAQ76 with *Ban*II and digestion with Bal 31 for 20 s. pMAQ79 retains 30 bases of the 5′-CS adjacent to base 1986 of pACYC184 (13).

### Conduction assays

The ability of particular DNA segments to act as sites for Int-mediated site-specific recombination was determined by assaying conduction of the test plasmid by the conjugative plasmid R388 from the donor strain UB1637 (*recA56*, $Sm^R$) to the recipient strain UB5201 (*recA56*, $Nx^R$) as described previously (4, 8, 10). The plasmids tested are all derivatives of pACYC184 and are $Cm^R$. R388 ($Tp^R$, $Su^R$) is a $Tra^+$ plasmid containing the

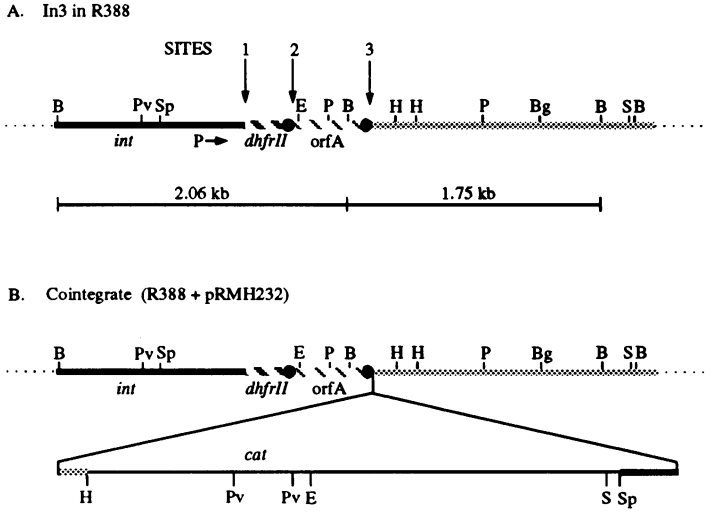


**Figure 2.** (**A**) Structure of the integron In3 in R388. Three integrase recombination sites indicated by vertical arrows are; (1) *attI*, (2) the *dhfrII* 59-base element, and (3) the orfA 59-base element. The lengths of the two diagnostic *Bam*HI fragments used to determine the site in R388 of integration of plasmids tested are also shown. (**B**) A typical cointegrate formed by recombination between the orfA 59-base element of R388 and *attI* in pRMH232. pRMH232 is shown below R388. The 5′-CS and 3′-CS are depicted as in Fig. 1 and pACYC184 derived sequence is represented by a thin black line. Dotted lines represent the remainder of R388. Symbols and restriction sites are as for Fig. 1 with the addition of E, *Eco*RI.

integron In3 (Fig. 2). The integron DNA integrase was supplied *in trans* by the plasmid pSU2056 ($Ap^R$) (8). For each plasmid tested two independent transformants containing pSU2056 were assayed. Transconjugants were selected on plates containing Tp and Nx and containing Cm and Nx and the conduction frequency was calculated as the ratio of $Cm^R$ to $Tp^R$ transconjugants.

### Resolution of cointegrates

To test for Int-dependent resolution, plasmid pSU2056 (*int*$^+$) was introduced into the *recA*$^-$ strain UB5201 containing a cointegrate. Cells were grown overnight, plasmid DNA extracted and re-transformed into UB5201 using a low DNA concentration as described by Collis and Hall (5). $Cm^R$ transformants were isolated and screened for resistance to Tp. Plasmid DNA from transformants was digested with *Hin*dIII and with *Bam*HI and *Eco*RI if required.

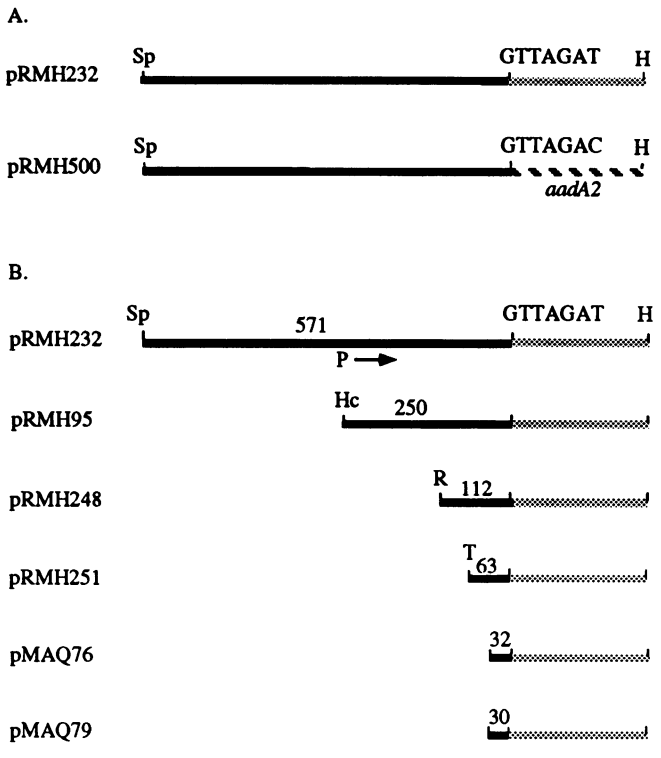


**Figure 3.** Structures of cloned integron fragments. (**A**) The fragments contained in the plasmids pRMH232 and pRMH500 are derived from deletion derivatives of the integron In6 from pSa. All fragments were cloned into pACYC184 in the same orientation. (**B**) Integron fragments with progressive deletions of the 5′-CS. Numbers indicate the length of 5′-CS sequence present (bp). The location of the promoter $P_{ant}$ in the 5′-CS is shown below pRMH232. Restriction sites shown are: H, *Hin*dIII; Hc, *Hinc*II; R, *Rsa*I; Sp, *Sph*I; T, *Taq*I.

## RESULTS

### Activity of the *attI* site

Site activity was assayed *in vivo* by testing the ability of plasmids containing integron fragments to form cointegrates with the conjugative plasmid R388 (8, 10). The integron In3, found in R388, contains two inserted cassettes, *dhfrII* and orfA, and therefore three potential targets for site-specific recombination. These are the *attI* site and the *dhfrII* and orfA 59-base elements (Fig. 2A). Activity was tested in a *recA*$^-$ background and the integron DNA integrase was supplied *in trans* by including the plasmid pSU2056 which expresses a high level of the integrase (Int). The activity of *attI* was first tested using pRMH232, which contains a fragment from an integron with no inserted genes, consisting of 571 bases of the 5′-CS and 199 bases of the 3′-CS (Fig. 3A). This plasmid has previously been shown to contain an active site for the insertion of circular gene cassettes (2). The frequency of conduction of pRMH232 was substantially higher in the presence of pSU2056 ($1.1\times10^{-2}$) than in its absence ($1\times10^{-7}$), and the frequency of Int-dependent conduction of pRMH232 was more than three orders of magnitude higher than that of the vector pACYC184 (Table II). This confirms that an active recombination site recognised by Int is present in pRMH232, and the activity of this site is somewhat higher than that of the *aadB* 59-base element present in plasmid pMAQ28 (4). To determine the site of insertion of pRMH232 into R388, plasmid DNA from four transconjugants from each of six independent crosses was analysed by digestion with *Bam*HI and with *Hin*dIII. Cointegrate formation by recombination with the orfA 59-base element should result in the loss of the 1.76 kb *Bam*HI fragment of R388 whereas insertion into either *attI* or the *dhfrII* 59-base element should result in the loss of the 2.06 kb *Bam*HI fragment (see Fig. 2A). All cointegrates appeared to have been formed by recombination between *attI* in pRMH232 and the orfA 59-base element in R388, and a typical cointegrate of R388 and pRMH232 is shown in Fig. 2B. The 3′ end of the orfA 59-base element was also the preferred site in R388 for cointegrate formation when various 59-base elements were tested (4).

To confirm that sequences from the 3′-CS other than the seven-base core site are not required for *attI* site activity, the activity of the plasmid pRMH500 was also determined. The integron

**Table II.** Conduction frequencies

| Plasmid[a] | Junction | pSU2056 ($Int^+$) | Conduction frequency[b] | | |
|---|---|---|---|---|---|
| | | | No. of determinations | Range | Average |
| pRMH232 | 5′-CS–3′-CS | + | 11 | $2.9\times10^{-3}$–$2.7\times10^{-2}$ | $1.1\times10^{-2}$ |
| pRMH232 | 5′-CS–3′-CS | – | 5 | $1.7\times10^{-8}$–$1.6\times10^{-7}$ | $1.0\times10^{-7}$ |
| pACYC184 | | + | 22 | $8.0\times10^{-7}$–$1.7\times10^{-5}$ | $4.2\times10^{-6}$ |
| pACYC184 | | – | 5 | | $<10^{-9}$ |
| pMAQ28 | *aadB*–3′-CS | + | 26 | $7.4\times10^{-4}$–$8.0\times10^{-3}$ | $3.7\times10^{-3}$ |
| pRMH500 | 5′-CS–*aadA2* | + | 8 | $4.1\times10^{-3}$–$2.9\times10^{-2}$ | $1.3\times10^{-2}$ |

[a]In all cases the donor strain also contained R388.
[b]Conduction frequency = chloramphenicol resistant ($Cm^R$) transconjugants/trimethoprim resistant ($Tp^R$) transconjugants.

fragment in pRMH500 (Fig. 3A) contains the same region of 5′-CS as pRMH232, but the 3′-CS region is replaced by a segment of the *aadA2* cassette, and the sequence of the core site is GTTAGAT in pRMH232 and GTTAGAC in pRMH500 (Fig. 3A). The conduction frequencies of pRMH232 and pRMH500 were essentially identical (Table II) indicating that only 5′-CS sequences together with a seven-base core site from either the 3′-CS or first inserted gene cassette (*aadA2*) are required for site activity. DNA from four cointegrates of R388 and pRMH500 from each of three independent crosses was analysed and again all cointegrates had been generated by recombination between *attI* in pRMH500 and the orfA 59-base element in R388. That recombination was site-specific was demonstrated by sequencing one junction from three independent cointegrates of R388 and pRMH500 and from the sequences (Fig. 4) it was possible to confine the recombination crossover point to within or on either side of the GTTAGA of the core sites at the 5′ end of the *aadA2* cassette and at the 3′ end of the orfA 59-base element. This is consistent with the results of previous studies which localised the crossover point to within or on either side of the GTT triplet in the core site (2, 4–6).

In this study conduction of pRMH232 occurred at a low frequency in the absence of pSU2056 (Table II), and this is likely to result from the activity of the integrase encoded by R388. In previous studies (4, 8, 10), clear evidence for activity of the R388 integrase was not obtained. However as this could have been due to either a weaker promoter for *int* in R388 or the lower copy number of R388 or both, it was not possible to conclude that the In3-encoded Int is inactive. To confirm that the cointegrates observed here arose by Int-mediated recombination 15 transconjugants were analysed. All contained cointegrates which appeared to have been formed by recombination with *attI* in pRMH232. For eight cointegrates the In3 recombination site was the *dhfrII* 59-base element, and for seven the site was the orfA 59-base element. Thus the In3-encoded Int is active, though the site preferences appear to be somewhat different as recombination at the *dhfrII* 59-base element is very rarely observed in cointegrates formed when the In2-encoded Int is supplied by pSU2056.

The vector pACYC184 does not contain a known integrase recognition site yet shows a low but significant conduction frequency ($4.2 \times 10^{-6}$) in the presence of pSU2056, and this is at least 5,000-fold higher than the conduction frequency of pACYC184 when pSU2056 is not present. This suggests that the integrase is able to recognise, albeit with low efficiency, sites other than 59-base elements or the *attI* site. Sequencing of the boundaries of selected pACYC184 + R388 cointegrates indicated that several sites in pACYC184 are recognised (see below).

| Plasmid | Junction | Sequence |
|---|---|---|
| R388 | orfA - 3′-CS | GCTCAACTTCAAC **GTTAGAT** GCACTAAGCACAT |
| R388+pRMH500 | orfA - *aadA2* | GCTCAACTTCAAC **GTTAGAC** ATCATGAGGGTAG |
| pRMH500 | 5′-CS - *aadA2* | GCCCTAAAACAAA **GTTAGAC** ATCATGAGGGTAG |

**Figure 4.** Recombinant junctions. Sequences of one recombinant junction from cointegrates formed between *attI* in pRMH500 and the orfA 59-base element in R388, and of the original boundaries in R388 and pRMH500 are shown. The seven-base core site is shown in bold type. Colons indicate bases identical in the original plasmids and the cointegrate, and the region to which the recombination crossover event can be confined is underlined. The junction sequence was determined from three independent cointegrates.

## Extent of the *attI* site

To determine the extent of 5′-CS sequence required for site activity, plasmids containing progressively shorter fragments of the 5′-CS were constructed using restriction sites in the 5′-CS (Fig. 3B). The plasmid with the shortest 5′-CS fragment (pRMH251) retained full activity (Table III), indicating that the presence of 63 bases of the 5′-CS is sufficient for full *attI* activity. Bal 31 nuclease-generated deletions were used to further define the sequences required. Two plasmids, pMAQ76 and pMAQ79, containing 32 and 30 bases of the 5′-CS respectively (Figs 3B and 5) were isolated (Table III). The Int-dependent conduction frequency of pMAQ76 was 40-fold lower than that of pRMH251 but still approximately 40-fold higher than that of pACYC184 (see Tables II and III) indicating that more than 32 bases of the 5′-CS are required for full *attI* activity. Four cointegrates from each of 14 independent crosses were analysed, and the majority had been formed by recombination between *attI* in pMAQ76 and the orfA 59-base element in R388. The sequences of both recombinant junctions of two such cointegrates confirmed that recombination was site-specific (data not shown).

The conduction frequency of pMAQ79 which retains only 30 bases of the 5′-CS was approximately 600-fold lower than that of pRMH251 and this is only 3-fold above the activity of the vector pACYC184. However, analysis of cointegrates revealed that the majority were formed by site-specific events. Of 63 cointegrates (four cointegrates from each of 12 crosses and 15 from one further cross), 59 were formed by recombination between the orfA 59-base element in R388 and *attI* in pMAQ79.

**Table III.** Conduction frequencies of plasmids containing 5′-CS fragments

| Plasmid[a] | Length of 5′-CS[b] (bp) | Conduction frequency[c] | | |
|---|---|---|---|---|
| | | No. of determinations | Range | Average |
| pRMH232 | 571 | 11 | $2.9\times10^{-3}-2.7\times10^{-2}$ | $1.1\times10^{-2}$ |
| pRMH95 | 250 | 8 | $2.0\times10^{-2}-5.3\times10^{-2}$ | $3.4\times10^{-2}$ |
| pRMH248 | 112 | 7 | $1.3\times10^{-3}-8.8\times10^{-3}$ | $5.3\times10^{-3}$ |
| pRMH251 | 63 | 11 | $3.0\times10^{-3}-2.6\times10^{-2}$ | $8.2\times10^{-3}$ |
| pMAQ76 | 32 | 15 | $3.8\times10^{-5}-5.5\times10^{-4}$ | $1.8\times10^{-4}$ |
| pMAQ79 | 30 | 15 | $1.3\times10^{-6}-3.3\times10^{-5}$ | $1.3\times10^{-5}$ |

[a]In all cases the donor strain also contained R388 and pSU2056.
[b]Total required length also includes the further 7 core site bases.
[c]Conduction frequency = $Cm^R$ transconjugants/$Tp^R$ transconjugants.

**Table IV.** Int-dependent resolution of cointegrates

| Plasmids | Phenotypes of $Cm^R$ transformants | |
|---|---|---|
| | $Tp^S$ | $Tp^R$ |
| R388 + pMAQ28, pSU2056 | 33 | 67 |
| R388 + pMAQ28 | 2 | 98 |
| R388 + pRMH251, pSU2056 | 15 | 235 |
| R388 + pRMH251 | 1 | 249 |

The sequences of the recombinant junctions in two independent cointegrates of this type confirmed this conclusion. Three of the remaining four cointegrates arose by recombination between the orfA 59-base element and a site in the vector sequence of pMAQ79. Thus, while the low conduction frequency indicates that the two bases of 5'-CS present in pMAQ76 but not in pMAQ79 are critical for site activity, the ability of the remnant site in pMAQ79 to recombine predominantly with the orfA 59-base element in R388 is retained. This implies that at least some features of the *attI* site responsible for determining site-specificity are still present in the 30 base region of the 5'-CS.

### Can two *attI* sites participate in a single recombination event?

To date site-specific recombination has only been clearly demonstrated to occur either between two 59-base element sites or between a 59-base element and an *attI* site (2, 4–6, 8, 10) and no clear evidence for recombination between two *attI* sites has been reported. However the possibility that recombination may occur between two *attI* sites was not rigorously investigated in these studies. In the present study such an event could potentially be identified by screening for trimethoprim sensitive ($Tp^S$) cointegrates. The plasmids pRMH248, pRMH251, pMAQ76 and pMAQ79 do not contain the promoter $P_{ant}$ which is located in the 5'-CS and is responsible for transcription of inserted genes (Fig. 3B), and the orientation of the integron fragment in these plasmids is such that a vector-derived promoter does not replace $P_{ant}$. As recombination between *attI* in one of these plasmids and *attI* in R388 would separate the *dhfrII* gene from $P_{ant}$ and transcription of the *dhfrII* gene should not occur, such events would result in the formation of $Tp^S$ cointegrates. Between 1 and 2% of cointegrates of R388 and each of the $P_{ant}$-deficient plasmids were found to be $Tp^S$. However, analysis of the plasmid DNA from the 23 $Tp^S$ cointegrates isolated revealed that recombination had not occurred between the *attI* site in the plasmid tested and *attI* in R388. Rather, recombination had occurred between the orfA 59-base element in R388 and the *attI* site of the incoming plasmid, and the $Tp^S$ phenotype was due to loss of the *dhfrII* cassette. Thus no evidence of integrative site-specific recombination between two *attI* sites was obtained. However, due to the observed strong preference for recombination events involving the orfA 59-base element in R388, it is not possible to conclude unambiguously that recombination between two *attI* sites cannot occur.

### Int-dependent resolution of cointegrates

Int-dependent, *recA*-independent resolution of cointegrates, occurring by recombination between two 59-base elements, has previously been demonstrated for cointegrates of R388 and pMAQ28 (4). Whether *attI* can also participate in excision events was addressed by testing for resolution of a cointegrate of R388 and pRMH251. The *recA*$^-$ strain UB5201 was used as the host and the DNA integrase was supplied by introducing the plasmid pSU2056. A strain without pSU2056 was used as an Int$^-$ control. Equivalent strains containing an R388 + pMAQ28 cointegrate were also included for comparison. Plasmid DNA isolated from these strains was re-transformed into UB5201 and transformants resistant to chloramphenicol (Cm) (encoded by pACYC184) were selected and screened for Tp resistance. As shown in Table IV, significant numbers of $Cm^RTp^S$ transformants were obtained using plasmid DNA from strains which contained both an R388 + pRMH251 cointegrate and pSU2056, indicating that Int-dependent resolution had occurred. One of 14 $Tp^S$ transformants analysed contained a cointegrate that had lost the *dhfrII* cassette. However, the remaining 13 $Tp^S$ transformants contained only a small plasmid. Digestion with *Bam*HI and *Eco*RI revealed four classes of resolution products. Seven plasmids were identical to pRMH251. The rest were related to pRMH251 but included one or more cassettes derived from R388. Three contained the orfA cassette, three contained both the *dhfrII* and orfA cassettes and one contained the *dhfrII* cassette and two copies of the orfA cassette in the order orfA–*dhfrII*–orfA. The resolution products containing the *dhfrII* cassette are unable to express Tp resistance as pRMH251 does not include the $P_{ant}$ promoter to direct transcription of the *dhfrII* gene. The single $Tp^S$ transformant obtained with DNA from the control strain which lacked pSU2056 also contained a plasmid identical in size to pRMH251, which presumably arose from a resolution event catalysed by the R388(In3)-encoded integrase.

### Int-dependent recombination with pACYC184

From the conduction frequencies (Table II) it appears that Int-mediated recombination between R388 and the vector pACYC184 can occur at a low frequency. To examine the nature of the recombination events, plasmid DNA from six transconjugants from each of six independent crosses was analysed by restriction enzyme digestion. The *Bam*HI digestion patterns indicated that all except three cointegrates arose from events involving the orfA 59-base element. For the remaining three cointegrates the recombination event involved either the *dhfrII* 59-base element or the *attI* site of R388. Thus one site involved in the formation of all cointegrates was a known integrase recognition site. From *Hin*dIII and *Eco*RI digestions it was possible to determine approximately the location in pACYC184 at which recombination had occurred and to determine if all cointegrates from a single cross resulted from the same recombination event. In two crosses all six cointegrates had identical restriction patterns and are likely to have arisen from a single recombination event. For the remaining four crosses more than one event had occurred, as between two and four distinct restriction patterns were evident. Thus, sites other than 59-base elements or *attI* are able to be recognised at low efficiency by the integrase, and many sites in pACYC184 appeared to have been used. However, some cointegrates from independent crosses showed identical restriction patterns indicating that preferred recombination sites may exist.

The sequences of both recombinant junctions of seven cointegrates and one junction from a further six cointegrates formed by recombination of the orfA 59-base element with pACYC184 confirmed that recombination occurred at several sites in pACYC184 (Fig. 6). The only common feature in the vicinity of the pACYC184 sites was the triplet GNT at the crossover points. In six of 13 cases the second nucleotide of this triplet was a T residue. This triplet is similar to the first three conserved bases of the core site, GTT, present at the recombination crossover point in 59-base elements and in *attI*. From the


**Figure 5.** Extent of the *attI* site. The complete 5′-CS region and part of the 3′-CS region in pRMH251 are shown. Dots represent bases identical to the pRMH251 sequence. Italicised sequences are those from the vector pACYC184 which replace the deleted portion of the 5′-CS and the underlined sequence is that of *aadA2*. The core recombination site (bold type) and the potential secondary core site GTTACGC in the 5′-CS are boxed. Short arrows between these two core sites indicate the short inverted repeats.

sequences of the cointegrate junctions it is possible to localise the recombination crossover point to the bases underlined in Fig. 6. In several cases the crossover must have occurred on either side of the G residue which corresponds to the first base of the core site. Even though recombination in pACYC184 is restricted to regions outside the chloramphenicol acetyl transferase (*cat*) gene and REP region, many GNT sites are available. It is therefore interesting to note that four sites in pACYC184 were each chosen in two independent recombination events, indicating that preferred recombination sites are likely to exist.

| | Position in pACYC184 | No. of independent cointegrates |
|---|---|---|
| TTTTATCTGATTAATAAGATGATCTTCTTGAGATCGTTTT | 1399 - | 1 |
| GAAGATCATCTTATTAATCAGATAAAATATTTCTAGATTT | 1413 + | 2 |
| GTCAGCCCCATACGATATAAGTTGTAATTCTCATGTTTGA | 1488 + | 2 |
| ATATAAGTTCTAATTCTCATGTTTGACAGCTTATCATCGA | 1502 + | 1 |
| CAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGC | 3470 - | 1 |
| ATAAAAACGGTTAGCGCTTCGTTAATACAGATGTAGGTGT | 3497 - | 1 |
| CCCAGAGCCTGATAAAAACGGTTAGCGCTTCGTTAATACA | 3508 - | 2 |
| CTCTGGGAGGCAGAATAAATGATCATATCGTCAATTATTA | 3542 + | 2 |
| TGAATAAGTGATAATAAGCGGATGAATGCAGAAATTCGGA | 3736- | 1 |

**Figure 6.** Comparison of the nucleotide sequences surrounding recombination sites in the vector pACYC184. The region to which recombination crossover event can be confined is underlined. The G and T nucleotides of the conserved GNT triplet are shown in bold type. Numbers indicate the position of the G residue of this triplet in the pACYC184 sequence (13), + indicates the strand shown in the published sequence and − indicates the complementary strand.

## DISCUSSION

The gene cassettes found in integrons are inserted at the unique *attI* site adjacent to the *int* gene and in integrons which contain cassettes this site is associated with the conserved sequences located 5′ to the cassette genes. The extent of the region required for maximal *attI* site activity was determined in this study to be more than 39 bases and a region containing only 37 bases retained $<0.1\%$ of full site activity which is barely detectable above the background. Thus the region of 14 bases predicted by Schmidt *et al.* (9), and designated hs2, is insufficient for this site. Others have claimed to have demonstrated that the 14 base hs2 site is sufficient for site activity (18, 19). However none of the constructs used by Nücken *et al.* (18) contain less than 186 bases of the 5′-CS. In the study of Zühlsdorf and Wiedemann (19) the recombination site was identified by demonstrating that the *aacA1* gene could be deleted by Int and comparing the sequence beyond the 3′ end of the *aacA1* gene with that of the inner boundary of the 5′-CS. However, the fragment used in the deletion studies was longer than that sequenced, and the sequences of the new boundaries in deletants were not determined. Thus the recombination site used in deletion of the *aacA1* cassette could be a 59-base element located in the unsequenced region.

Though 59-base elements differ substantially in both sequence and length, recent compilations of their sequences (4, 6), have revealed several features common to all members of this family. Firstly 59-base elements consist of imperfect inverted repeats flanking a short central loop. They are bounded by a core site at the 3′ end and an inverse core site at the 5′ end, and the sequences of these core sites are the most conserved features of 59-base elements. The sequences of at least 20 bases at the outer ends are related to a consensus sequence (6), while the central region is of variable length and sequence but generally retains an inverted repeat structure. Furthermore, as several different 59-base elements have been shown to be active as integrase recombination sites (2, 4−6, 8) it seems reasonable to assume that at least some of the common features are required for activity. However, when the sequence of the region required for *attI* site activity was examined for these features only a short inverted repeat (see Fig. 5) and the core site were found. Thus *attI* is not obviously a member of the 59-base element family. The finding that deletion of two bases (GT) from pMAQ76 to give rise to pMAQ79 leads to a substantial loss of site activity indicates that these two bases are critical. As they are part of the sequence GTTACGC which has six bases identical to the core site consensus (boxed in Fig. 5), it is tempting to speculate that this potential core site is also an important feature of the *attI* site. However, more detailed analysis of *attI* using directed mutagenesis will be necessary to identify the key features.

The *attI* site also lacks the features common to the recombination sites of other well characterised members of the integrase family. The recombination sites for l Int, the FRT site for FLP and the *loxP* site for Cre all consist of a pair of inverted repeats (7−13 bases) flanking a short overlap region (7−8 bases) (for reviews see refs. 20−23). The inverted repeats include integrase binding domains and strand cleavage occurs at either end of the overlap region. The 59-base elements and the *attI* site of integrons do not at present appear to have the same general

organisation. However, as the recombination crossover point for events catalysed by the integron integrase has been localised to one end of the seven-base core site, it is possible that this core site represents at least part of an Int binding domain.

As well as recombination events involving two specific sites, the integron integrase is also able to catalyse recombination between one specific site and secondary sites conforming to the consensus GNT. Recently, others have also demonstrated recombination at secondary sites (24). The recombination crossover point has previously been localised to within or on either side of the conserved GTT triplet of the seven-base core site (2, 4–6) and from the sequences of the junctions in recombinants arising from events involving a secondary site, the crossover point can be further localised to either side of the conserved G residue. The same conclusion has been reached by Francia *et al* . (24). Finally, it is tempting to speculate that recombination events involving a circular gene cassette and a secondary site could lead to the insertion of cassette-associated antibiotic resistance genes at many different locations. If the cassette is correctly oriented with respect to a promoter in the recipient, the gene should be expressed. Though such insertions would occur only rarely they should be extremely stable as the cassette would be flanked by only one specific recombination site making excision unlikely. This type of event could play an important role in the evolution of the genomes of bacteria and their plasmids and transposons.

## ACKNOWLEDGEMENTS

We thank Tina Collis for the construction of several plasmids used in this study. This work was supported in part by a grant from the Australian National Health and Medical Research Council.

## REFERENCES

1. Stokes, H.W. and Hall, R.M. (1989) *Mol. Microbiol.*, **3**, 1669–1683.
2. Collis, C.M., Grammaticopoulos, G., Briton J., Stokes H. W. and Hall, R. M. (1993) *Mol. Microbiol.*, **9**, 41–52.
3. Hall, R.M. and Vockler, C. (1987) *Nucleic Acids Res.*, **15**, 7491–7501.
4. Hall, R.M., Brookes, D.E. and Stokes, H.W. (1991) *Mol. Microbiol.*, 5, 1941–1959.
5. Collis, C.M. and Hall, R.M. (1992a) *J. Bacteriol.*, 174, 1574–1585.
6. Collis, C.M. and Hall, R.M. (1992b) *Mol. Microbiol.*, 6, 2875–2885.
7. Cameron, F.H., Groot Obbink, D.J., Ackerman, V.P. and Hall, R.M. (1986) *Nucleic Acids Res.*, **14**, 8625–8635.
8. Martinez, E. and de la Cruz, F. (1990) *EMBO J.*, **9**, 1275–1281.
9. Schmidt, F.R.J., Nücken, E.J. and Henschke, R.B. (1989) *Mol. Microbiol.*, **3**, 1545–1555.
10. Martinez, E. and de la Cruz, F. (1988) *Mol. Gen. Genet.*, **211**, 320–325.
11. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) *Gene*, **33**, 103–119.
12. de la Cruz, F. and Grinsted, J. (1982) *J. Bacteriol.*, **151**, 222–228.
13. Rose, R.E. (1988) *Nucleic Acids Res.*, **16**, 355.
14. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
15. Miller, J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
16. Jones, D.S.C. and Schofield, J.P. (1990) *Nucleic Acids Res.*, **18**, 7463–7464.
17. Sundström, L., Radström, P., Swedberg, G. and Sköld, O. (1988) *Mol. Gen. Genet.*, **213**, 191–201.
18. Nücken, E.J., Henschke, R.B. and Schmidt, F.R.J. (1991) *Mol. Gen. Genet.*, **229**, 137–146.
19. Zühlsdorf, M.T. and Wiedemann, B. (1993) *J. Gen. Microbiol.*, **139**, 995–1002.
20. Landy, A. (1989) *Annu. Rev. Biochem.*, **58**, 913–949.
21. Stark, W.M., Boocock, M.R. and Sherratt, D.J. (1992) *Trends Genet.*, **8**, 432–439.
22. Sadowski, P.D. (1986) *J. Bacteriol.*, **165**, 341–347.
23. Sadowski, P.D. (1993) *FASEB J.*, **7**, 760–767.
24. Francia, M.V., de la Cruz, F. and Garcia Lobo, J.M. (1993) *Mol. Microbiol.*, **10**, 823–828.
25. Datta, N. and Hedges, R.W. (1972) *J. Gen. Microbiol.*, **72**, 349–355.
26. Chang, A.C.Y. and Cohen, S.N. (1978) *J. Bacteriol.*, **134**, 1141–1156.