The R46 Site-specific Recombination System is a Homologue of the Tn3 and γδ (Tn1000) Cointegrate Resolution System

By HELEN M. DODD and PETER M. BENNETT*
Department of Microbiology, University of Bristol, University Walk, Bristol BS8 1TD, UK

(Received 29 December 1986; revised 12 March 1987)

The nucleotide sequence of the R46 site-specific recombination system has been determined. The organization of the recombination gene (perRa6) and the site at which it acts (per site), together with the extensive sequence homology displayed with the tnpR genes and res sites of the transposons Tn3 and γδ (Tn1000), suggests that they have been derived from a Tn3-like element. These site-specific recombination functions of R46 play a role in plasmid maintenance.

INTRODUCTION

The N group plasmid R46 encodes a site-specific recombination system comprising a per (plasmid-encoded recombinase) gene and an adjacent site at which the gene product acts, the per site (Dodd & Bennett, 1983, 1986). As has been demonstrated for a number of other plasmids (Summers & Sherratt, 1984; Hakkaart et al., 1984; Austin et al., 1981), a site-specific recombination system ensures that the plasmid is maintained predominantly in the monomeric form. The carriage of the per functions on R46 appears to play a similar role contributing to stable plasmid maintenance.

The per site and the per gene product of R46 were identified as functions related to the site-specific recombination system of the ampicillin-resistance transposon Tn3 (Dodd & Bennett, 1983), and the gene products of the Tn3 tnpR gene and perRa6 have been shown to be interchangeable; both can mediate recombination between the res site of Tn3 and the per site of R46. As a consequence, R46::Tn3 recombinant molecules carry either site-specific deletions or inversions.

The R46 site-specific recombination system has been cloned into pACYC184, where it noticeably increased the stability of the pACYC184 derivatives. Furthermore, insertion of Tn3 into this recombinant molecule generated the same types of site-specific rearrangements seen with R46 (as shown in Fig. 1 and described previously: Dodd & Bennett, 1986).

Here we present the DNA sequence of the perRa6 and the per site. In addition, the nucleotide sequence across the end-points of Tn3-induced deletions and inversions, in different recombinant molecules, have been determined. The data are compared with the related site-specific recombination systems of Tn3 (Heffron et al., 1979) and γδ (Tn1000) (Reed et al., 1982). The possible origin of the R46 per functions is discussed.

METHODS

Plasmids and strains. Escherichia coli UB5201 (pro met recA56 gyrA) (Sanchez et al., 1982) served as bacterial host for all plasmids except those pUC9 derivatives used for nucleotide sequence determination, for which MC1022 (araD139 ara-leu)7697 lacZDM15 galU galK strA) (Casadaban & Cohen, 1980) was the host.

Plasmid pUB2951 is a pACYC184 derivative carrying the cloned R46 per functions (Dodd & Bennett, 1986). The insertion of Tn3. E5 into this molecule and subsequent site-specific recombination between the res site of Tn3. E5 and the R46 per site generated two classes of pUB2951::Tn3. E5 recombinant plasmids comprising...
Fig. 1. Generation of pUB2951::Tn3.E5 recombinant plasmids. Tn3.E5 sequences are represented by blocked-in lines. The left and right inverted repeats of the transposon are labelled L and R respectively. Arrows indicate the orientations of the Tn3.E5 res site (solid box) and the R46 per site (open box). (a) Recombination between the directly repeated recombination sites results in deletion of sequences lying between the two sites. (b) Recombination between inversely repeated recombination sites leads to inversion of the intervening sequences. Hybrid recombination sites, (i) and (ii), are generated as a result of either deletion or inversion.

deletion derivatives and inversion derivatives respectively. The first class, represented by pUB2965 and pUB2966, are deleted for both plasmid and transposon sequences as the result of site-specific deletion (Fig. 1a). The sizes of the deletions are different in the two derivatives, but the deletion junctions generated [marked (i) in Fig. 1a] are the same. The second class is represented by pUB2964, in which Tn3.E5 is inserted in the opposite orientation to that in pUB2965 and pUB2966. This orientation permits inversion of the sequences between the R46 per site and the res site of the transposon, creating two new transposon-plasmid junctions (Fig. 1b), one of which is expected to be in distinguishable from that generated by site-specific deletion [marked (i) in Fig. 1b]. The construction of these plasmids has been described previously (Dodd & Bennett, 1986). Transposon Tn3.E5 was employed in the construction of the recombinant plasmids because it carries an EcoRI linker at the start of the tnpR gene (Tn3 coordinate 3246; Heffron et al., 1978; Fig. 2d) which proved to be very useful in the sequence determination across hybrid recombination sites (see below). The cloning vector pUC9 (Vieira & Messing, 1982) was used to subclone DNA fragments for sequence analysis.

Isolation of plasmid DNA, cloning and transformation. Plasmid DNA was isolated by dye buoyant density centrifugation of cleared cell lysates (Cornelis et al., 1976). Restriction enzymes were obtained from various commercial sources and used according to the suppliers' recommendations. DNA fragments for sequence analysis were ligated into pUC9 using T4 ligase and recombinant plasmids were recovered by transformation of E. coli MC1022 by the method of Cohen et al. (1972) with the modification of Humphreys et al. (1979). Transformants were selected on nutrient agar containing ampicillin (100 μg ml⁻¹), the chromogenic β-galactosidase substrate 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-Gal, 30 μl of a 24 mg ml⁻¹ solution in dimethyl formamide per 20 ml agar), and the inducer of β-galactosidase, isopropyl β-D-thiogalactopyranoside (IPTG, 30 μl of the 100 mM solution per 20 ml agar). White colonies contained putative pUC9 derivatives with inserted fragments and were purified for further analysis; blue colonies were discarded.

DNA sequence analysis. The 1.2 kb PstI fragment of pUB2951 (shown in Fig. 2a) was ligated into pUC9. To sequence this region of R46, smaller fragments were then subcloned into pUC9, using restriction enzymes EcoRI, PvuII and TaqI (Fig. 2a). Fragments containing the deletion and inversion end-points generated in pUB2964, pUB2965 and pUB2966 [indicated as (i) and (ii) in Fig. 1] were also ligated into pUC9 and the nucleotide sequences across the junctions were determined from the sites and in the directions shown (Fig. 2b, c). DNA sequence analysis was performed by the chemical method of Maxam & Gilbert (1977).
**RESULTS AND DISCUSSION**

**DNA sequence of per<sub>R46</sub>.**

The sequence of the 1.2 kb PstI fragment which contains the site-specific recombination functions of R46 is shown in Fig. 3. Superficially the overall homology between these R46 sequences and the equivalent regions of Tn3 and γδ was not particularly strong due to major differences at both ends of the sequence (see below). However, an open reading frame of 573 bp (Fig. 3), when aligned with the sequences of Tn3 and γδ resolvase genes, showed striking homology (81% homology with Tn3 and 72% with γδ). There are two other possible initiation codons within the sequence which start open reading frames of more than 500 bp (at coordinates 321 and 324), but both of these lie in the region of R46 corresponding to the per site (see below) and for this reason are thought unlikely to represent the start of the R46 per gene.

It would seem most likely that the 573 bp sequence starting at coordinate 405 (Fig. 3) is the per gene of R46. Consistent with this conclusion is the previous observation that deletion of a 400 bp EcoR1 fragment, which accounts for about half the coding sequence (Fig. 3), abolished R46 resolvase activity (Dodd & Bennett, 1986). Further, the predicted amino acid sequence of the R46 resolvase is very similar to those of the resolvases of Tn3 and γδ. The putative per gene codes for a protein of 190 amino acids (Fig. 4), slightly larger than those predicted for the resolvases of Tn3 and γδ (185 and 183 amino acids respectively). This protein would display 85% homology with the resolvase of Tn3 and 81% homology with that of γδ. Of the mismatched residues, in both cases approximately half represent conservative changes (Fig. 4).
H. M. DODD AND P. M. BENNETT

TGATGCCTATTTGGACCGATACCGGCTACGGCCCACAGAATGATGCACGCTGAMATGCCG
GCCTTTGAAATGGGTTCACTGGCACGTCCACATGCAAAAAGGGCAGTAAATGTTATACCA
CGGACTATTTGCCAGAGGCCCTCATACAGATGAAAATAGCCCGACGTCAGGTCATTCG
GGCTTTGAATGGGTTCATGTGCAGCTCCATCAGCAMAGGGGATGATMGTTTATCACCA
CCG ACTATTTGCAACAGTGCCC

TCP_T_C_A_A_G
APG_
&&A-ATA_T_AG
_Cc
C_c_p_G_c_T_G C
A_cGT_c_A_T_TG_G_
GTXCTSSfi&C&TAS C_G_~C_C_C_ATAA_T_T_C&~~T-C~~~~T~~~~G~~~~_G_~A~~~C~ A CJ;_G_G

CA_-
TACCGGCCTCCCTGCCTGATGGTCATGAATTAACAATTTCGCAACCCGTCGAAAATATAT
AAATTCTCGGACACACTAAAATGGTTGTTCTAGGTGTCTATTAATATCGATTTTTTTGTA
TAACAGACACACCACGTCCGTATTTTAGGATACATATTATGCGACTTTTTGGTT
ACGCACGGGTATCAACAGCCGACAGTCTCCTGATATTCGATACAGATCAAGGGGCTTAAAGG
CGGGGAGTCGCGGTCCGATTCATAGACGACGGGATCAGCACCGACGGTGAAATGGGGAAAA
TGGTTGTGACCATTCTTTTCGATAGCTGATCGAACCTATGCAAAAGGAATATCAGCAGGTC
CGAAGCTGGATCTGCTGCGAATGAAGGTGGAGGAAGGTGACGTCACCTTGGTGAAGAAAC
EwRI
TCGACCGGCTTGGCCGCGACACTGCCGATATGATCCAACTGACAAAAGAATTCGACGCTC
AGGGAGTCGGCTTGGCACACTGCCGATATGATCCAACTGACAAAAGAATTCGACGCTC

EcoRI

TCGACCGGCTTGGCCGCGACACTGCCGATATGATCCAACTGACAAAAGAATTCGACGCTC
AGGGAGTCGGCTTGGCACACTGCCGATATGATCCAACTGACAAAAGAATTCGACGCTC

Fig. 3. Single-strand DNA sequence of the R46 PstI fragment containing the R46 site-specific recombination functions. The open reading frame, corresponding to perR46, is underlined (continuous line), as is the sequence that shows homology to the sequences at the starts of the tnpA genes of Tn3 and γδ (dashed line). EcoRI sites are indicated. The IS46 sequence which displays homology with IS15 is in bold letters with the 14 bp inverted repeat indicated by the horizontal arrow above the sequence.

It has been demonstrated that γδ resolvase undergoes chymotryptic cleavage at a specific site between residues 140 and 141 (Abdel-Meguid et al., 1984). The 140-residue amino-terminal domain is thought to contain the catalytic site for the recombination reaction. γδ and Tn3 sequences in this region are conserved to a higher degree (91% identical) than is found in the smaller carboxyl-terminal domain (see below). This conservation is also seen with the perR46.
Fig. 4. Comparison of the predicted amino acid sequences of the site-specific resolvases of perR46, Tn3 and γδ. The residues which differ from those of R46 are indicated for Tn3, above, and for γδ, below the R46 sequence. Amino acid residues are represented by the standard one-letter abbreviations. The vertical line between residues 140 and 141 indicates the site of chymotryptic cleavage (Abdel-Meguid et al., 1984). The open boxes represent the regions at which conserved residues are clustered in a number of site-specific recombination enzymes (Newman & Grindley, 1984; Diver et al., 1983). Asterisks indicate residues that are required for resolvase activity in γδ (Newman & Grindley, 1984).

Site I

Site II

Site III

Fig. 5. Comparison of the R46 per site with the Tn3 and γδ res sites. Nucleotide substitutions seen in the Tn3 and γδ sequences are indicated, respectively above and below the R46 sequence. The open boxes indicate the three resolvase-binding sites in the Tn3 and γδ res sites, each bounded by a 9 bp inverted sequence (horizontal arrow; from Grindley et al., 1982; Wells & Grindley, 1984). The start of perR46 is underlined. The coordinates run from the centre of the cleavage point, within site I. The dashes in the Tn3 sequence, at positions −35 and −36 represent 'padding' required to maximize homology. The ATG underlined at R46 coordinates −59 to −61 corresponds to the start codon of the tnpA genes of Tn3 and γδ (see text for details).
resolvase sequence (Fig. 4). The first 140 residues share 89% homology with \( \gamma \delta \) and 92% homology with Tn3.

A comparison of amino acid sequences of several site-specific recombination enzymes has shown that in the amino-terminal domain, identical residues are clustered in four groups (Newman & Grindley, 1984; Diver et al., 1983; indicated in Fig. 4). The functional importance of two of these conserved regions has been demonstrated by the isolation of independent mutations of \( \gamma \delta \) resolvase that are defective in resolution but retain repressor activity (Newman & Grindley, 1984). These mutations each contain an alteration in a single amino acid (shown as asterisks in Fig. 4). These essential residues are retained in the \( \text{per}_{\text{R46}} \) sequence. In general, mismatches are found outside the conserved regions (Fig. 4). This finding is consistent with the proposal that these sequences have a specific role in the activity of these enzymes (see Newman & Grindley, 1984).

The smaller carboxyl-terminal domain of \( \gamma \delta \), which recognizes specific binding sequences in the \( \text{res} \) site (Abdel-Meguid et al., 1984), has significantly lower homology with the equivalent regions of Tn3 despite the fact that the enzymes are interchangeable. A similar amino acid degeneracy in this region is found when the \( \text{per}_{\text{R46}} \) gene product is compared with the resolvases of Tn3 and \( \gamma \delta \) (Fig. 4). It is interesting to note that, despite the lower homology between \( \text{per}_{\text{R46}} \) and \( \gamma \delta \) sequences in this domain (53% identical), there is a region in the last 14 amino acids that contains a stretch of eight identical and two conservatively changed residues. Homology between the gene products of \( \text{per}_{\text{R46}} \) and Tn3 in the same region is even more pronounced: there are 15 identical residues (Fig. 4). These sequences lie within those domains of the individual proteins that show some homology to the DNA-binding domains of the catabolite gene activator protein (Weber et al., 1982) and cro protein (Ohlendorf et al., 1982; see Abdel-Meguid et al., 1984).

### \( \text{R46 per site} \)

The \( \text{res} \) sites of both Tn3 and \( \gamma \delta \) are similarly located, namely upstream of and adjacent to \( \text{tnpR} \), with the transposase gene \( \text{tpmA} \) on the other side (Heffron et al., 1979; Reed et al., 1982). Unlike these transposons, there is no evidence that R46 encodes a \( \text{tpmA} \) gene product, but sequences upstream of the proposed start of \( \text{per}_{\text{R46}} \) show strong homology with the analogous regions of Tn3 and \( \gamma \delta \). Further, the homology extends past the intercistronic sequences into those which comprise the promoter-proximal regions of the \( \text{tpmA} \) genes of Tn3 and \( \gamma \delta \) (Fig. 5), indicating that R46 originally possessed a \( \text{tpmA} \) analogue and that the arrangement of its site-specific recombination proteins and the transposase gene was the same as that seen on Tn3 and \( \gamma \delta \).

It has been demonstrated that the resolvase proteins of \( \gamma \delta \) and Tn3 bind specifically to three distinct sites within a 115 bp DNA segment of the \( \text{tnpA-tnpR} \) intercistronic region (Grindley et al., 1982; Kitts et al., 1983). These sequences are sufficient to constitute a fully functional \( \text{res} \) site (Wells & Grindley, 1984). The sequence similarity between R46 and \( \gamma \delta \) in this region is striking, particularly within the three binding sites, where, apart from two base pair changes within site I, the sequences are identical (Fig. 5). It has been postulated that a 9 bp sequence, with a consensus of 5'-TGTCYNTA, constitutes the binding sequence recognized by resolvase (Grindley et al., 1982; Abdel-Meguid et al., 1984; Wells & Grindley, 1984). This sequence, in a slightly degenerate form, flanks all three binding sites of the \( \text{res} \) site of \( \gamma \delta \). In the R46 sequence, the right-hand side of site I differs from that of \( \gamma \delta \) in two positions (Fig. 5). Nonetheless, resolvase activity is retained. One change involves the substitution of a C for a G at position 11 (Fig. 5) and generates a better match with the consensus sequence. Wells & Grindley (1984) have suggested that one or both of the last two nucleotides of the 9 bp consensus sequence (TA) are specifically recognized by resolvase because a mutant \( \gamma \delta \) \( \text{res} \) site, in which the right-hand resolvase-binding site has GG at these positions, is inactive. In the R46 sequence a G has been substituted for the T (Fig. 5) without obvious effect on recombination ability, indicating that this change is neutral. This finding suggests that the last nucleotide in the consensus sequence, A, which is conserved in each of the binding sites of \( \gamma \delta \), Tn3 and R46, plays a critical role in the activity of the \( \text{res} \) site.

It has been demonstrated that there are two promoters for expression of \( \text{tnpR} \) of \( \gamma \delta \). The weaker, detected in vitro (Reed et al., 1982), lies within site I of the \( \text{res} \) site, with its -10 region
overlapping the sequence at which site-specific recombination occurs (Fig. 5). In the equivalent region of R46, sequence divergence has resulted in changes to the −35 region of this potential promoter, reducing its similarity to the consensus sequence (Rosenberg & Court, 1979). In addition, the proposed initiation site for the γδ tnpR transcript associated with the promoter has been altered (CAC to GAC, coordinate 11, Fig. 5). An alternative promoter for γδ tnpR, which is about four times stronger, lies within binding site II (Wells & Grindley, 1984). This sequence is fully conserved in R46 (Fig. 5).

Tn3- and γδ-mediated site-specific recombination involves cleavage of sequences at the centre of binding site I (Reed, 1981; Kostriken et al., 1981) within the palindromic sequence TTATAA (Reed & Grindley, 1981). This is conserved in R46 in a 16 bp region, common to all three sequences (Fig. 5).

R46 : : Tn3.E5 deletion and inversion end-points

Fusion of Tn3.E5 and R46 sequences would be expected to occur at the cleavage site within the res site of the former and the per site of the latter. To test this hypothesis the nucleotide sequences across the recombination junctions of the two deletion derivatives pUB2965 and pUB2966 were determined, together with the two alternative R46–Tn3.E5 junction points, generated by inversion, from pUB2964 (see Figs 1 and 2).

Fragments containing deletion end-points carried, as expected, those R46 sequences to the left of the cleavage site within binding site I. These could be determined unambiguously to a CG base pair (at position −11 within site I, Fig. 5). The next 16 base pairs are identical in R46 and Tn3.E5 (Fig. 5) and so cannot be assigned unambiguously. At position 7, to the right of the point of resolvase-mediated cleavage, the CG base pair found in R46 is absent and is replaced by an AT base pair as in Tn3.E5 (Fig. 5). Sequences thereafter are derived from Tn3.E5. The nucleotide sequence of the R46–Tn3.E5 junction in pUB2964, designated inversion end-point 1 (see Fig. 2b), was identical to the R46–Tn3.E5 junctions present in the deletion derivatives pUB2965 and pUB2966. The inversion end-point 2 sequence (Fig. 2c) had, as expected, the opposite arrangement, namely Tn3.E5 sequences to the left of the cleavage site and R46 sequences to the right of the cleavage site (Fig. 5). Again the crossover occurred within the 16 bp region common to both sequences (Fig. 5).

These results confirm that the site-specific rearrangements in these R46 : : Tn3.E5 recombinants were generated by recombination between the related res and per sites creating hybrid recombination sites containing sequences derived from both R46 and Tn3.E5.

Sequences of IS46

R46 carries two copies of a 0.81 kb insertion sequence, IS46 (Brown et al., 1984). One of these elements is adjacent to the R46 per site, with one terminus lying within the PstI fragment which has been sequenced in this work. IS46 does not encode a site-specific recombination system and has been shown to transpose independently of the adjacent R46 per functions. Hybridization studies have shown that IS46 is related to a number of other IS elements (Brown et al., 1984), including IS15, which forms the terminal repeats of the Km' transposon Tn1525 (Labigne-Roussel et al., 1983). Comparison of the available R46 sequence with that of IS15 (Trieu-Cuot & Courvalin, 1984) showed that the first 141 bp of the R46 sequences obtained (Fig. 3) were identical to one end of IS15. At this point, which represents the outermost nucleotide pair of IS15, homology was lost.

We conclude that IS46 is very closely related to IS15 and, at least at one terminus, the two elements are indistinguishable. It would seem likely that IS46 and IS15 have arisen from a common ancestral insertion sequence which has since spread throughout bacterial populations and is now found as a component of a number of plasmids and transposons (see Brown et al., 1984).

Evolution from a Tn3-like element

Sequences lying to the left of the R46 per site display homology with the start of the Tn3 and γδ tnpA genes. These R46 sequences align precisely with the equivalent sequences of γδ. There are
several mismatches in this region, but an ATG is found in the same position (coordinate -59, Fig. 5) as the translational start of the \( \gamma 6 \) \textit{tnpA} gene. When the equivalent R46 and Tn3 sequences are aligned to give maximum homology, the first 102 bp of the \textit{tnpA} gene of Tn3 are 80\% homologous to R46 sequences. The amino acids which would be encoded from these sequences are 71\% homologous. This homology continues to the start of IS46 (shown in Fig. 3).

The \textit{tnpA} gene of Tn3 is approximately 3 kb long (Heffron \textit{et al.}, 1979). The insertion of IS46 near the beginning of such a gene on an ancestor of R46 would result in gene disruption with the majority of the transposase gene lying on the opposite side of the element. This sequence has not been determined, but no homology between R46 and either Tn3 or \( \gamma 6 \) sequences was detected in this region (data not shown). IS46, as with other transposable elements, can generate adjacent deletions (Brown \textit{et al.}, 1984) and such an event may have removed the residual transposase sequence. Further sequencing at the far terminus of the IS46 element will be required to determine the possible pathway of events leading to the present arrangement of sequences in R46.

The nucleotide sequence of \( \textit{per} \sb{R46} \) shares a greater degree of similarity with \textit{tnpR} of Tn3 (81\%) than it does with its homologue on \( \gamma 6 \) (72\%). In contrast, sequences within the intercistronic regions of R46 and \( \gamma 6 \) are 88\% similar, compared with 76\% homology between those of R46 and Tn3. These findings might suggest that the ancestral element of the R46 \textit{per} system was a Tn3-\( \gamma 6 \) hybrid generated by the elements' site-specific recombination systems, in which case the formation of a hybrid \textit{res} site would be expected (Reed, 1981; Kostriken \textit{et al.}, 1981). In addition, in such a hybrid sequences to one side of the cross-over will originate from Tn3 while those on the other side will be derived from \( \gamma 6 \). However, within the R46 sequence the switch from sequences more closely related to \( \gamma 6 \) than to Tn3, to those more closely resembling Tn3, occurs, not at the normal recombination cross-over point within site I (Fig. 5), but at a site within what are the \textit{tnpR} genes of the two transposons. There is a 102 bp region in the R46 sequence, starting at coordinate 36 (Fig. 5) and proceeding into the \textit{per} gene, that is identical to the equivalent \( \gamma 6 \) sequence, apart from four mismatches between sites II and III (coordinates 72 to 75, Fig. 5). Accordingly, the element from which the R46 site-specific recombination system originated may have been a Tn3-\( \gamma 6 \) hybrid transposon generated by recombination, not at the \textit{res} sites as might have been expected, but within the \textit{tnpR} genes. Alternatively, of course, the R46 \textit{per} system may have evolved independently from either Tn3 or \( \gamma 6 \) or from an ancestral element that was common to all three systems. The data permit no informed choice between these possibilities.

It is concluded that the nucleotide sequence homology displayed between the site-specific recombination system of R46 and those of Tn3 and \( \gamma 6 \) indicates that the \textit{per} functions of R46 originated on an element from the Tn3 family of transposons. The ability of the element to retranspose has been lost due, at least in part, to the insertion of IS46 within the transposase gene. In contrast, the resolvase function (\textit{per} \sb{R46}) and recombination site (\textit{per} site) are fully functional.

The R46 site-specific recombination system would appear to be an example of the positive selection of sequences, initially foreign, which are beneficial to the plasmid, in this case improving its stability (Dodd & Bennett, 1986). This finding suggests that the carriage of transposons encoding site-specific recombination functions, in addition to increasing the plasticity of the surrounding genome, may also play a role in the maintenance of many plasmids throughout nature.

This work was supported by a grant from the Science and Engineering Research Council (UK) to P. M. B.

REFERENCES


The R46 site-specific recombination system


