Specificity of DNA recognition in the nucleoprotein complex for site-specific recombination by Tn21 resolvase

Samantha C.Hall and Stephen E.Halford*

Department of Biochemistry, Centre for Molecular Recognition, University of Bristol, Bristol BS8 1TD, UK

Received August 12, 1993; Revised and Accepted October 26, 1993

ABSTRACT

Resolvases from Tn3-like transposons catalyse site-specific recombination at *res* sites. Each *res* site has 3 binding sites for resolvase, I, II, and III. The *res* sites in Tn3 and Tn21 have similar structures at I and II but they differ at III. Mutagenesis of the Tn21 *res* site showed that sub-site III is essential for recombination though the sequences in III that are recognized by Tn21 resolvase are positioned differently from the equivalent sequences in the Tn3 site. The deletion of III caused a 1,000-fold drop in the rate of recombination. But other mutations at III, changing 3 or 4 consecutive base pairs, caused only 1.5- to 4-fold decreases in rate, even when the mutations were in target sequences for this helix-turn-helix protein. The reason why Tn21 resolvase has similar activities at a number of different DNA sequences may be due to the multiplicity of protein-protein and protein-DNA interactions in its recombinogenic complex. This lack of precision may be a general feature of nucleoprotein complexes.

INTRODUCTION

DNA-protein interactions often involve a single protein recognizing an individual DNA sequence. The specificity of these interactions can be extraordinarily high. For example, a 1 bp change in the recognition sequence for a restriction enzyme can reduce DNA cleavage activity by $\geq 10^6$ -fold (1). However, some events on DNA require loci that contain multiple binding sites for the protein, often for 2 or more different proteins (2). In these situations, it is generally thought that the DNA transaction demands such a high level of precision that it can only be met by an array of binding sites, rather than just a single site, in order to create a nucleoprotein assembly of defined structure at that locus on the DNA (3). Site-specific recombination by resolvase is one example of the latter type of interaction.

Tn3-like transposons encode a protein called resolvase which acts at a site within the transposon called *res* (4). During transposition, resolvase catalyses a reciprocal recombination between 2 copies of the transposon, specifically at the *res* sites (5). The recombination can be reproduced *in vitro* with a negatively supercoiled DNA containing 2 *res* sites in head-to-

tail orientation. On the addition of the resolvase from either $\gamma\delta$ (Tn1000) or Tn3 to a DNA with either $\gamma\delta$ or Tn3 *res* sites, the circular substrate is converted to 2 interlinked circles of covalently closed DNA (6, 7). The resolvases from Tn21 or Tn1721 catalyse equivalent rearrangements on substrates containing *res* sites from either Tn21 or Tn1721 though neither Tn21 nor Tn1721 resolvase has any activity on *res* sites from Tn3 (8, 9). The specificities observed *in vitro* concur with genetic studies, which had shown that the resolvases from different Tn3-like transposons belong to distinct complementation groups, with Tn3 and $\gamma\delta$ in one group and Tn21, Tn501 and Tn1721 in another (10, 11). Within each group, the resolvases have >80% amino acid identity and can act at *res* sites from all members of the same group. But between groups, the identity is about 30% and a resolvase from one group has no activity on *res* sites from another group.

In both Tn3 and $\gamma\delta$, *res* contains 3 binding sites for resolvase, named I, II and III (7, 12; Fig. 1). The cross-over is at sub-site I but II and III are also necessary for efficient recombination (13, 14). Each binding site contains 2 copies of a degenerate 12 bp sequence related by dyad symmetry but these are separated by different lengths of DNA at each sub-site (Fig. 1). Each 12 bp sequence is contacted throughout its length by 1 protomer of resolvase, with almost all of the contacts coming from the C-terminal domain of the protein (15). This domain contains a helix-turn-helix motif (16). $\gamma\delta$ resolvase recognizes this DNA from the major groove at the outside edge of each binding site, by its helix-turn-helix, and also from the adjacent minor groove on the same face of the DNA towards the centre of each site (17, 18). These contacts have been confirmed by crystallography on a homologue of resolvase, the C-terminal domain of Hin bound to DNA (R.C. Johnson, personal communication). The binding of the active dimeric form of resolvase to *res* bends the DNA at all 3 sub-sites and produces a compact, highly organized, nucleoprotein structure known as the resolvosome (19, 20).

The *res* sites in Tn21 and Tn1721 also contain 3 binding sites for the cognate resolvases (21; Fig. 1). Sub-sites I and II in Tn21/Tn1721 are organized like those in Tn3/ $\gamma\delta$. The dyad axes at I and II are separated by 53 bp in Tn3/ $\gamma\delta$ and by 53½ bp in Tn21/Tn1721. But in the Tn21 family, sub-site III is ill-defined (21). The right half of III in Tn21 contains a 12 bp sequence that matches the consensus for binding a protomer of Tn21

* To whom correspondence should be addressed

resolvase while the left half contains two 12 bp blocks that might act as recognition sequences (marked **a** and **b** in Fig. 1). Sequence **a** matches the consensus at the 3' end (as shown in Fig. 2) but, at its 5' end, it is poorly conserved across the Tn21 family and it lacks the GTCA sequence found in almost all other binding sites for Tn21 and Tn/721 resolvases. The converse applies to sequence **b** (Fig. 2). Sequence **a** in the left half of III (Fig. 1) is separated by 8 bp from the right hand sequence, in contrast to the 1 bp spacer in Tn3/ $\gamma\delta$, but this creates a dyad axis 89 bp downstream of the cross-over point in I, similar to the 87½ bp in Tn3/ $\gamma\delta$. Hence, if **a** is the target for Tn21 resolvase, sub-site III is longer in Tn21 than in Tn3/ $\gamma\delta$, but the overall structure of the Tn21 resolvosome is likely to be similar to that in Tn3/ $\gamma\delta$. On the other hand, sequence **b** overlaps the right hand element by 2 bp, thus creating a binding site of similar size to III in Tn3/ $\gamma\delta$, but it produces a dyad axis 5 bp further along the DNA compared to **a**. A 5 bp translocation corresponds to a 170 rotation around the axis of the DNA helix so, if Tn21 resolvase were to recognize **b** rather than **a**, the resultant resolvosome cannot be organized like that in Tn3/ $\gamma\delta$.

The initial objective of this study was to determine whether sequence **a** or **b** has any function in recombination by Tn21 resolvase. However, mutations in the Tn21 *res* site had less effect on the activity of Tn21 resolvase than might have been expected. These observations lead us to propose that a general function of specialized nucleoprotein complexes is to relax specificity for DNA sequence.

MATERIALS AND METHODS

DNA and protein preparations

Plasmids were used to transform either *Escherichia coli* GM2929 (*dam*, *recF*: M.G. Marinus, unpublished) or the *Dam*⁺ strain HB101 (22). The transformants were grown in M9 minimal media with 1 mCi/L [*methyl*-³H]-thymidine and the covalently closed form of the DNA was purified on CsCl/ethidium bromide gradients (23). Unless noted otherwise, all reactions were done on DNA from *E. coli* GM2929. DNA preparations were checked with *Mbo*I, which cleaves unmethylated *dam* sites, and with *Dpn*I, which only cleaves methylated *dam* sites (24): DNA from GM2929 had no *Dam* methylation, DNA from HB101 was fully methylated. The purification of Tn21 resolvase and the procedures for determining both protein and DNA concentrations were as noted previously (25). Molarities of resolvase are given as the dimeric protein of *M_r* 43,000. *Eco*RV was purified as before (26). All other enzymes were from commercial suppliers.

DNA constructs

The plasmid pCP10 is a derivative of pUC19 carrying 2 *res* sites from Tn21, both on 800 bp restriction fragments (27). Standard procedures (22) were used to manipulate 1 of the *res* sites from pCP10 (28). The 800 bp fragment was first digested with *Bal*31 in the presence of Tn21 resolvase. The length of DNA that resolvase protected from *Bal*31 (28) was the same as in DNase I foot-prints (21, 29). The DNA was then ligated *via* linkers to a phagemid vector (30) and subjected to oligonucleotide-directed mutagenesis (30) to change 2 bp within the Tn21 *res* sequence. one to convert CCATGT in the 21 bp spacer between sub-sites I and II to CCATGG, thus creating an *Nco*I site (Fig. 1); another to convert ATATCT between II and III (the 2 bp underlined are those between the end of II and the start of sequence **a** in III) to AGATCT, thus creating a *Bgl*III site (Fig. 1). The DNA,

together with flanking restriction sites from the polylinker (Fig. 1), was subsequently used to replace 1 of the 800 bp fragments on pCP10 to yield a plasmid, pSH1, with 2 Tn21 *res* sites in direct repeat: one with the above changes and the other straight from Tn21. Further plasmids were made by cutting pSH1 with the appropriate pair of restriction enzymes, then blunting the ends and re-ligating the termini, in order to delete sub-site(s) (pSH1ΔIII and pSH1ΔII,III). Alternatively, the DNA between the *Bgl*III and *Pst*I sites was replaced with pairs of synthetic oligonucleotides, in order to mutate sub-site III (Fig. 3: pSH1b–f). Oligonucleotides were synthesized on a Dupont Coder 300. The *res* sites in pSH1 and derivatives were sequenced on a Dupont Genesis.

Recombination reactions

Stock solutions of resolvase were diluted in 50 mM Tris.HCl (pH 8.0), 1.0 M NaCl, 0.1 mM EDTA, 100 µg/ml bovine serum albumin, and added immediately to the DNA to initiate the reaction. The reactions were carried out at 25°C in 200 µl volumes containing 10 nM ³H-labelled plasmid and ≥ 60 nM resolvase in 50 mM Tris.HCl (pH 8.0), 200 mM NaCl, 10 mM MgCl₂, 100 µg/ml albumin. At timed intervals after the addition of resolvase, samples (20 µl) were withdrawn and mixed immediately with 0.1 mM ethidium bromide (2 µl) to stop the recombination reaction (23). The DNA was digested with *Eco*RV and analysed by electrophoresis through agarose. The circular form of pSH1 is cleaved by *Eco*RV to fragments of 3776, 290 and 168 bp, while the catenane produced by Tn21 recombination gives fragments of 3089, 855 and 290 bp. The other plasmids were also converted to catenanes that gave similarly distinct *Eco*RV fragments. The amounts of the *Eco*RV fragments from the substrate and the recombinant products were measured by scintillation counting and % recombination determined as before (23, 31). Values of % recombination refer to the % of the supercoiled substrate converted to product and include a correction factor for nicked DNA in the plasmid preparations (usually < 10%). Each data point in Figs 4–8 is the mean from three separate reactions and the error bars on all of the data points show the range of values from the repeat experiments: in some instances, the symbol illustrating the data point obscures the error bars. The time-course of recombination by resolvase is biphasic (31) and follows the equation

$$R_t = R_\infty - Ae^{-k_1 t} - Be^{-k_2 t}$$

where *R_t* and *R_∞* are % recombination at time *t* and time ∞ and *A* and *B* are the amplitudes of the fast and slow phases. The lines drawn through each set of data points are non-linear regression fits to this equation and comparisons of reaction rates are from the values of *k*₁, the rate constant for the fast phase.

RESULTS

To facilitate the analysis of the *res* site from Tn21, a DNA cassette was made by altering the sequence in the spacer regions between sub-sites I and II and between II and III, in order to create new restriction sites at these positions (Fig. 1). Additional restriction sites were placed in the flanking DNA either side of the *res* sequence. Though the single bp mutation to create the *Bgl*III site (AGATCT) between II and III is in the spacer separating these sub-sites, the mutation simultaneously creates a new *dam* site, GATC (24). Since the adenines that would be methylated by the *dam* methyltransferase may be part of III, the DNA substrates

Figure 1. Organization of *res* sites. The *res* site in Tn3 (and $\gamma\delta$) has 3 sub-sites; I, II and III. Each contains two 12 bp recognition sequences for Tn3 resolvase in inverted orientation, marked by filled arrows. The numbers of bp between the two sequences in each sub-site and in the spacers between each sub-site are shown. The recognition sequences for Tn3/ $\gamma\delta$ resolvase were initially considered to be 9 bp and those for Tn21/Tn1721 resolvase 10 bp (12, 21), but it is now known that 1 protomer of $\gamma\delta$ resolvase contacts 12 bp (15). The *res* site in Tn21 is represented as above with 12 bp recognition sequences except that the left half of sub-site III lacks a clearly defined copy of the target sequence: 2 potential sites are marked with open arrows labelled a and b. Tn21 *res* DNA was manipulated in this study to create new restriction sites at the locations indicated.

Figure 2. Potential targets for Tn21 resolvase. The consensus for a 12 bp recognition sequence for Tn21 resolvase is given: Y = C or T, R = A or G, W = A or T and lower case marks poorly conserved bases. The sections of the DNA that, by analogy with $\gamma\delta$ resolvase (17, 18), are likely to be recognized from the major and minor grooves are marked above. The consensus was derived from both halves of sub-sites I and II and the right half of sub-site III in Tn21, Tn501 and Tn1721. The left half of sub-site III, in all 3 members of the Tn21 family, contain 2 potential target sequences, as listed, marked a and b in Fig. 1. Bases matching the consensus are highlighted.

for all experiments reported here (except for those noted otherwise) were from a Dam⁻ strain of *E. coli*.

The cassette was used to construct a plasmid, pSH1, that can act as a substrate for Tn21 resolvase: pSH1 contains both the cassette and a wild-type copy of *res* in direct repeat. The rate of recombination on pSH1 (Fig. 4) was identical to that observed previously on other plasmids carrying 2 wild-type Tn21 sites (27, 31). Moreover, as with the wild-type substrates (23), the maximal rate on pSH1 was achieved with a 6-fold molar ratio of protein dimers over DNA molecules: further increases in the concentration of resolvase produced no further increase in rate (data not shown). The protein concentration that gives a 1:1 ratio of resolvase dimers to resolvase binding sites on the DNA is sufficient to saturate all of the binding sites on pSH1. The changes

Figure 3. Mutations at sub-site III. Part of the DNA sequence from the Tn21 *res* cassette in pSH1 is given in (a). The part shown includes the last bp in sub-site II and all of sub-site III: a *Xho*I site (CTCGAG) was incorporated immediately after the end of sub-site III. To mutate sub-site III, pairs of complementary oligonucleotides were synthesized, as shown in (b)–(f), and annealed together. The resultant duplexes were used to replace the *Bgl*II–*Pst*I fragment in pSH1 to yield the derivatives pSH1b (with duplex b) through to pSH1f (with duplex f). In all 6 duplexes, the bp that differ from the wild-type Tn21 sequence are in bold type and underlined.

in and around the Tn21 site that were needed to make the cassette have no detectable effect on the reaction.

The new restriction sites in and around the cassette occur only once in pSH1. These sites were used to construct several derivatives of pSH1, either by deleting 1 or more of the resolvase binding sites from the cassette, or by replacing the DNA between 2 restriction sites with synthetic oligonucleotides (Fig. 3). No alterations were made to the second (wild-type) copy of the Tn21 *res* site on pSH1, so each derivative involves the pairing of 1 wild-type and 1 mutant site.

Deletions of sub-sites

Recombination by either $\gamma\delta$ or Tn3 resolvase is severely depressed by the deletion of III from 1 of the 2 *res* sites on the substrate, and even more so by the deletion of both II and III (13, 14). However, sub-site III in Tn21 appears to be different from all other resolvase binding sites (21). To examine the role of III in the Tn21 *res* site, 2 derivatives of pSH1 were made, one deleting III and the other both II and III (pSH1 Δ III and pSH1 Δ II,III). The deletions replace sub-site III (or II and III) with a sequence from pUC19, the *Hind*III site from the polylinker and the part of *lacZ* that follows the polylinker: the number of positions where the replacement sequence and the *res* site had identical bp was close to the statistical expectation for random DNA. The new plasmids were tested as substrates for Tn21 resolvase (Fig. 4). Both underwent site-specific recombination between the mutant and wild-type sites, but with much lower efficiencies than 2 wild-typesites. After 21 hr reactions, about 40% of pSH1 Δ III had been converted to the catenane product while 20% of pSH1 Δ II,III had given the product (insert to Fig.

Figure 4. Deletion mutations. The main panel shows the reactions of 60 nM Tn21 resolvase on either 10 nM pSH1 (open-circles) or 10 nM pSH1ΔIII (filled circles). Also given in the main panel is the reaction of 500 nM resolvase on 10 nM pSH1ΔIII (open squares). The insert shows the reactions of 100 nM Tn21 resolvase on either 10 nM pSH1ΔIII (filled circles) or 10 nM pSH1ΔII,III (open triangles) over an extended time-base.

4). Under the same reaction conditions, 40% recombination on pSH1 is achieved in < 1 min.

From the initial rates of these reactions at comparable concentrations of resolvase, the deletion of sub-site III reduced the rate of recombination by a factor of about 1000 relative to wild-type, while the deletion of both II and III decreased the rate by a factor of 2000. Sub-site III is therefore essential for efficient recombination in the Tn21 system while the additional effect caused by deleting both II and III is comparatively minor. However, the amount of recombination on pSH1ΔIII increased with increasing concentrations of Tn21 resolvase (Fig. 4). This suggests that the small amount of recombination seen on this plasmid involves weak nonspecific binding of resolvase to the DNA, perhaps adjacent to sub-site II (14), rather than to just the 5 specific binding sites for resolvase that remain on pSH1ΔIII.

Right half of sub-site III

To identify the sequences in III that are needed for recombination, derivatives of pSH1 were constructed in which sub-site III in the *res* cassette was replaced with altered sequences (Fig. 3). With the $\gamma\delta$ system, single bp alterations to a DNA fragment carrying an isolated binding site for resolvase often had large effects on the binding of the protein to that DNA but, when the altered binding sites were incorporated into a *res* site, they generally had no effect on recombination (18). Hence, in an attempt to ensure detectable effects, most of our mutations in Tn21 were substitutions at 4 consecutive bp. The right half of sub-site III contains a sequence that closely matches the consensus for binding Tn21 resolvase. It includes at its outer edge the 5'-GTCA-3' motif

Figure 5. Right half of sub-site III. The reactions contained 10 nM pSH1b and Tn21 resolvase at the following concentrations: 60 nM (filled squares); 100 nM (filled circles); 200 nM (filled triangles). The error bars on the data from the latter 2 experiments have been omitted. Also shown is the control reaction of 10 nM pSH1 with 60 nM Tn21 resolvase (open circles).

found in almost all other binding sites for Tn21 resolvase and a derivative, pSH1b, was constructed in which this was changed to CCTG. When pSH1b was tested with the minimal amount of resolvase needed for recombination, 6 dimers per mol. DNA, the rate for the recombination of mutant and wild-type sites was 4 times lower than the parental plasmid (Fig. 5). But, in contrast to the parental DNA, increasing concentrations of resolvase enhanced the rate of reaction on pSH1b and, at high protein concentrations, it approached that observed on pSH1 (Fig. 5).

The GTCA motif located on the 'bottom' strand in the right half of sub-site III (Fig. 3) is therefore part of a recognition sequence for Tn21 resolvase. The kinetic data in Fig. 5 demonstrate that the mutation of all 4 bp in this motif weakens the binding of the protein. At a 1:1 ratio of resolvase to resolvase binding sites on pSH1b, the concentration of resolvase is insufficient to saturate the mutant site and saturation is achieved only at higher protein concentrations. But once bound to this mutant site, Tn21 resolvase appears to fulfil all of its functions with unaltered activity. However, the 4 bp mutation in the right half of sub-site III had much less effect on resolvase activity than the deletion of sub-site III (Fig. 4), so resolvase must also recognize other sequences in this binding site.

Left half of sub-site III

The construction of the DNA cassette from the Tn21 *res* site (Fig. 1) created a *dam* site that overlaps sequence a, but not b, in the left half of sub-site III. The *dam* methyltransferase methylates the N6 group of adenine in both of the AT bp in its target sequence, GATC (24), and this should create a severe steric barrier against any protein trying to recognize this DNA from the major groove (32). Hence, if resolvase recognizes sequence a rather than b, it might be less active on Dam-methylated pSH1 than on the unmethylated form.

Figure 6. *Dam* methylation. All reactions contained 60 nM Tn21 resolvase and 10 nM plasmid substrate, either pSH1 (circles) or pSH1b (squares). In both cases, the unfilled symbols mark the extent of recombination on DNA lacking *dam* methylation (from *E. coli* GM2929) and the filled symbols the DNA with *dam* methylation (from *E. coli* HB101). The reactions are shown over a shorter time-base shown compared to Figs. 4 and 5.

Samples of pSH1 and pSH1b were isolated from *Dam*⁻ and *Dam*⁺ strains of *E. coli*. Since the rate of recombination by resolvase varies with DNA superhelicity (31), both samples were analysed on chloroquine gels: the 2 strains yielded DNA with similar levels of supercoiling (not shown). When tested for recombination by Tn21 resolvase (Fig. 6), the reaction on the *Dam*⁺ form of pSH1 was identical to that on the *Dam*⁻ form described above, while the *Dam* methylated DNA from pSH1b gave a reaction rate that was 5 times slower than the unmethylated DNA; *i.e.*, 20 times slower than the pSH1 parent. Another plasmid was constructed that differed from pSH1 only by cloning the *res* cassette in reverse orientation. The latter was used to analyse the stability of the synaptic complex between 2 Tn21 *res* sites in inverted orientation as described previously (27): the complex was less stable with *Dam*-methylated DNA than with unmethylated DNA (data not shown).

These observations suggest that sequence **a** in the left half of sub-site III is recognized by Tn21 resolvase. On pSH1b, the interaction of resolvase with sub-site III is already weakened by a 4 bp mutation in the right half (Fig. 5) so that a further disruption elsewhere in the DNA, such as methylation of sequence **a**, has a detectable effect on the rate of recombination, even though the second disruption is by itself insufficient to produce a detectable effect on pSH1 (Fig. 6). Similarly, the synaptic complex between inverted *res* sites is less stable than that between directly repeated sites (27), so only the former yields a measurable effect from *Dam* methylation. However, the DNA fragment used to create the mutation in pSH1b carries an additional *dam* site 5 bp upstream of the Tn21 sequences (Fig. 3). We cannot exclude the possibility that this second site mediates the *dam* effect on pSH1b, perhaps indirectly since *dam* methylation alters the dynamic flexibility of DNA (33).

Figure 7. Left half of sub-site III. The reactions contained 60 nM Tn21 resolvase and either 10 nM pSH1c (filled circles) or 10 nM pSH1d (open circles). The former differs from pSH1 in sequence **a**, the latter in sequence **b**.

Two further derivatives of pSH1 were constructed in the same manner as pSH1b. For one, pSH1c, sequence **a** was altered so that its first 8 bp contained only 1 match to the consensus for Tn21 resolvase (Figs. 2 and 3c). For the other, pSH1d, all 4 bp in the GTCA motif in sequence **b** were altered, again creating a segment with only 1 match to the Tn21 consensus in its first 8 bp (Fig. 3d). Both constructs were tested for recombination with the minimal amount of Tn21 resolvase needed for the reaction, 6 dimers of the protein per mol. plasmid, conditions that produced a slow reaction on pSH1b (Fig. 5). The plasmid altered in sequence **b** (pSH1d) gave a reaction that was similar to the parental plasmid, pSH1 (Fig. 7). The DNA altered in sequence **a** (pSH1c) showed a reduced rate, though the reduction was less than 2-fold (Fig. 7). These experiments have thus failed to identify any function for sequence **b** in sub-site III but they support the view that sequence **a** plays at least some role in recombination by Tn21 resolvase.

Minor groove interactions

Each subunit of $\gamma\delta$ resolvase contacts its 12 bp target sequence in the major groove, at the outside edge of each binding site, and across the adjacent minor groove towards the centre of the site (17, 18): the equivalent regions extrapolated to the Tn21 consensus are shown in Fig. 2. The above mutations, to either left (both **a** and **b**) or right halves of sub-site III, were all in regions where resolvase contacts the DNA in the major groove. Perhaps the interactions across the minor groove play the key roles. In the regions of sub-site III that may be contacted from the minor groove, both **a** in the left half and the target for resolvase in the right half contain the sequence 5'-GAAG-3', in 'top' and 'bottom' strands respectively. Two more plasmids were constructed from pSH1, in which these GAAG blocks were altered to TCCT: pSH1e for **a** in the left side, pSH1f for the right (Fig. 3e and 3f). When tested for recombination with the minimal amount of Tn21 resolvase as above, both gave reactions

Figure 8. Minor groove interactions. The reactions contained 60 nM Tn21 resolvase and 10 nM DNA from 1 of the following plasmids: pSH1e (filled triangles); pSH1f (filled circles); pSH1 (open circles). Both pSH1e and pSH1f differ from the control, pSH1, by 4 bp mutations in the section of the DNA that is likely to be recognized by resolvase from the minor groove: in the left half of sub-site III (sequence **a**) for pSH1e and in the right half of III for pSH1f.

that were slightly slower than pSH1 (Fig. 8). The reduction following the left hand change (pSH1e) was 2-fold, which is slightly larger than the decrease obtained by mutating the part of sequence **a** that is recognized from the major groove (pSH1c: Fig. 7). However, the decrease from mutating all 4 bp in the right-hand GAAG block, on pSH1f, was only a factor of 1.5, less than the 4-fold decrease from changing all 4 bp in the GTCA block on its outside edge (pSH1b: Fig. 5).

DISCUSSION

Organization of *res* sites

In the left half of sub-site III in the Tn21 *res* site, sequence **a** appears to be the target for resolvase. This is consistent with foot-printing data on the Tn21 *res* site (21, 29). The binding of Tn21 resolvase to this DNA protects most of the phosphodiester bonds in both **a** and **b** from digestion by DNase I. But in both halves of sub-sites I and II, and in the right half of III, the GTCA component of the recognition sequence contains 2 phosphodiester bonds that are not protected from DNase I by the binding of resolvase. Re-analysis of the DNase I foot-prints on the Tn21 *res* site (21, 29) shows that the equivalent bonds in **a**, but not **b**, remain susceptible to DNase I. Previously (21), these phosphodiester bonds had been assigned to the spacer between sub-sites II and III. In both $\gamma\delta$ and Tn3, the spacer between II and III is 5 bp long and is accessible to DNase I in the presence of the cognate resolvase (7, 12). In contrast, sequence **a** in the Tn21 *res* site is separated by only 2 bp from sub-site II so Tn21 resolvase might prevent DNase I from reaching the II-III spacer.

The overall organization of the resolvosome, formed by the binding of Tn21 resolvase to all 3 of its binding sites in *res*, is likely to be similar to that in Tn3/ $\gamma\delta$. In the *res* sites from both

Tn3/ $\gamma\delta$ and Tn21, the centre-to-centre distance between sub-sites I and II corresponds to about 5 helical turns of B-DNA (4). This phasing is necessary for the reaction since the insertion of 6 or 17 bp between I and II in $\gamma\delta$ reduced recombination, while the addition of 10 or 21 bp permitted full activity (19). Presumably, this distance is required for the correct spatial orientation of the proteins bound to these 2 sites. Since Tn21 resolvase recognizes sequence **a** in sub-site III, rather than **b**, the distance between the centres of II and III corresponds to about 3.3 helical turns in both Tn21 and in Tn3/ $\gamma\delta$. The organization of the resolvosome might require that the proteins bound to II and III have a fixed orientation relative to each other, determined by the twist of the DNA as is the case between I and II. Alternatively, it may demand a precise length of DNA between the centres of sub-sites II and III.

Resolvase needs to be able to distinguish sub-site I from the accessory sub-sites, II and III, since it carries out the strand transfer reaction at I. Studies on the Tn3/ $\gamma\delta$ systems had led to the proposal that this was achieved by the lengths of II and III being set at 3 and 2 helical turns respectively while I was 2½ turns long. Consequently, the binding of the dimeric form of resolvase to I results in the bending and the unwinding of the DNA at the centre of the site, in order to bring the 2 symmetry-related recognition sequences in each half of the site onto the same face of the DNA (34). However, *res* in Tn21 fails to conform to this pattern. In Tn21, sub-sites I and II are approximately 2½ and 3 helical turns long, but the length of III is midway between the lengths of I and II.

Multiple binding sites

Site-specific recombination by resolvase is one of many DNA transactions that occur at loci with multiple binding sites for protein(s). Other examples include the initiation of DNA replication (2), transposition (35, 36) and transcription (37). The binding of protein(s) to an array of sites on DNA often generates a specialized nucleoprotein complex, held together by a network of protein-DNA and protein-protein interactions. These nucleoprotein complexes are clearly essential to organize and control complicated transactions on DNA (3). In the case of resolvase, the organization of the synaptic complex determines its topological specificity for *res* sites in head-to-tail orientation (20, 27, 38). It has also been suggested that nucleoprotein complexes are essential for DNA transactions that demands a higher level of precision than can be achieved by a single protein interacting with an isolated site on the DNA (3). However, the degree of specificity needed to target a reaction to just 1 locus in a genome can be achieved by a single protein acting at an individual site. Type II restriction enzymes such as *EcoRV* and *EcoRI* (1) cleave their recognition sequences very much faster than any other sequence: sites that differ from the recognition site by just 1 bp are cut between 10^6 and 10^9 times more slowly (26, 39). Moreover, these endonucleases usually make double strand breaks at their cognate sites but they cleave noncognate sites by means of 2 consecutive single strand breaks. The nicked DNA can be repaired by DNA ligase before the endonuclease cuts the second strand (26, 39) so, in the presence of ligase as must be the case *in vivo*, the discrimination factors shown by restriction enzymes exceed 10^9 -fold.

Helix-turn-helix proteins such as resolvase (16) are generally less efficient than restriction enzymes at rejecting noncognate sequences. Even so, a 1 bp change in the target sequence for a helix-turn-helix protein can cause a 1000-fold decrease in

binding constant and the $\Delta\Delta G^\circ$ values for changing 2 or more bp are generally additive (40). Hence, changing 4 bp in a binding site for a helix-turn-helix protein ought to make that site equivalent to nonspecific DNA. But all of our plasmids carrying 4 bp mutations in sub-site III were much better substrates for Tn21 resolvase than the plasmid where III had been deleted. The low activity on pSH1 Δ III was not due to the particular DNA from pUC19 used to replace sub-site III: other plasmids where III had been replaced with different DNA sequences gave similarly low activities (data not shown). The 4 bp mutations caused only marginal reductions to the rate of recombination, by factors varying from 1.5 to 4, in contrast to the 1000-fold drop with the deletion. A 4 bp mutation at one end of the sequence and Dam methylation at the other end produced only a 20-fold decrease in rate (Fig. 6). It seems that the retention of any part of sub-site III allows resolvase to function better than when III is replaced by random DNA. This may be due to flexibility in the resolvase protein, given that it can bind specifically to DNA sites of varying lengths (4; Fig. 1). However, many of the amino acid substitutions that have been made in the DNA recognition domain of resolvase have little effect on recombination (4, 25). Even the replacement of the entire recognition helix in the helix-turn-helix motif of Tn21 resolvase, with the equivalent amino acid sequence from Tn3 resolvase, only weakened the binding of the protein to Tn21 DNA: it still allowed for recombination of Tn21 *res* sites to the exclusion of Tn3 sites (29).

Site-specific recombination by resolvase is thus remarkably insensitive to mutations in either the protein or the DNA target. Other DNA loci that form nucleoprotein complexes show similar insensitivities to mutation. For example, single bp changes at the recombinogenic sites for λ integrase often have no effect on recombination though the integrase reaction can be blocked by altering 3 bp in a single binding site (41). The insensitivity of these systems is likely to be due to the multiplicity of interactions in the nucleoprotein complex, so that the ΔG° for any one interaction is small compared to the overall ΔG° for the formation of the complex. Previously, it had been found that single bp changes to a DNA fragment carrying an isolated binding site for $\gamma\delta$ resolvase could abolish any detectable binding of the protein to that DNA but, when the single bp changes were made to an intact *res* site, they had almost no effect on the efficiency of recombination (18). But it is difficult to extrapolate from a binding constant for a DNA with a single site to the equilibrium constant for the formation of a nucleoprotein complex. In this study, the dependence of the rate of recombination on the concentration of Tn21 resolvase (Fig. 5) demonstrates that a 4 bp mutation at one sub-site in *res* makes the ΔG° for the formation of the recombinogenic complex less negative than before but a slight increase in the protein concentration is sufficient to result in a negative ΔG for the formation of this complex.

Given the partitioning of the free energy for the formation of a nucleoprotein complex into numerous protein-DNA and protein-protein interactions, an inescapable feature of multiprotein complexes is that they will discriminate against alternative DNA sequences less well than individual proteins interacting with isolated sites. This lack of sensitivity may be advantageous for the reactions on DNA where the survival of the DNA depends upon the completion of that reaction, since it minimizes the consequences of ill-matching protein to DNA. The broad specificity of Tn21 resolvase for a wide range of related DNA sequences confers an additional evolutionary advantage. The Tn21 family of Tn3-like transposons (11) now contains over 30

members that vary in size from < 6 kb (TnI722) to > 60 kb (Tn4653). Part of this diversity is due to Tn21 carrying an integrase that enables it to incorporate new antibiotic resistance functions (42). But other members of the Tn21 family, such as Tn501 and TnI721, are clearly related to each other by recombination events that can be localized to the cross-over point in *res* (11, 43). Such events were presumably mediated by a resolvase with a broad specificity for DNA sequence.

ACKNOWLEDGEMENTS

We thank Christian Parker, Panos Soultanas and Karen Burman for contributing to this work and John Grinstead, Tony Maxwell, Mark Oram, Marshall Stark and Panos for discussions and comments on the manuscript. This work was supported by the Wellcome Trust, an SERC studentship (S.C.H.) and a Royal Society Leverhulme Trust Research Fellowship (S.E.H.)

REFERENCES

1. Roberts, R.J. and Halford, S.E. (1993) In Roberts, R.J., Linn, S.M. and Lloyd, R.S., (eds.), *Nucleases*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, chapter 2.
2. Kornberg, A. and Baker, T.A. (1992) *DNA Replication*, 2nd ed., W.H. Freeman, New York.
3. Echols, H. (1986) *Science* **233**, 1050–1056.
4. Hatfull, G.F. and Grindley, N.D.F. (1988) In Kucherlapati, R. and Smith, G.M. (eds.), *Genetic Recombination*, American Society for Microbiology, Washington, DC, pp. 357–396.
5. Sherratt, D.J. (1989) In Berg, D.E. and Howe, M.M. (eds.), *Mobile DNA*, American Society for Microbiology, Washington, DC, pp. 163–184.
6. Reed, R.R. (1981) *Cell* **25**, 713–719.
7. Kitts, P.A., Symington, L.S., Dyson, P. and Sherratt, D.J. (1983) *EMBO J.* **2**, 1055–1060.
8. Halford, S.E., Jordan, S.L. and Kirkbride, E.A. (1985) *Mol. Gen. Genet.* **200**, 169–175.
9. Rogowsky, P. and Schmitt, R. (1985) *Mol. Gen. Genet.* **200**, 176–181.
10. Diver, W.P., Grinstead, J., Fritzinger, D.C., Brown, N.L., Altenbuchner, J., Rogowsky, P. and Schmitt, R. (1983) *Mol. Gen. Genet.* **191**, 189–193.
11. Grinstead, J., de la Cruz, F. and Schmitt, R. (1990) *Plasmid* **24**, 163–189.
12. Grindley, N.D.F., Lauth, M.R., Wells, R.G., Wityk, R.J., Salvo, J.J. and Reed, R.R. (1982) *Cell* **30**, 19–27.
13. Wells, R.G. and Grindley, N.D.F. (1984) *J. Mol. Biol.* **179**, 667–687.
14. Bednarz, A.L., Boocock, M.R. and Sherratt, D.J. (1990) *Genes and Devel.* **4**, 2366–2375.
15. Rimphanitchayakit, V., Hatfull, G.F. and Grindley, N.D.F. (1989) *Nucleic Acids Res.* **17**, 1035–1050.
16. Abdel-Meguid, S.S., Grindley, N.D.F., Smyth-Templeton, N. and Steitz, T.A. (1984). *Proc. Natl. Acad. Sci., U.S.A.* **81**, 2001–2005.
17. Graham, K.S. and Dervan, P.B. (1990) *J. Biol. Chem.* **265**, 16534–16540.
18. Rimphanitchayakit, V. and Grindley, N.D.F. (1990) *EMBO J.* **9**, 719–725.
19. Salvo, J.J. and Grindley, N.D.F. (1988) *EMBO J.* **7**, 3609–3616.
20. Benjamin, H.A. and Cozzarelli, N.R. (1988) *EMBO J.* **7**, 1897–1905.
21. Rogowsky, P., Halford, S.E. and Schmitt, R. (1985) *EMBO J.* **4**, 2135–2141.
22. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual* (2nd ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
23. Castell, S.E., Jordan, S.L. and Halford, S.E. (1986) *Nucleic Acids Res.* **14**, 7213–7226.
24. Barras, F. and Marinus, M.G. (1989) *Trends Genet.* **5**, 139–143.
25. Ackroyd, A.J., Avila, P., Parker, C.N. and Halford, S.E. (1990) *J. Mol. Biol.* **216**, 633–643.
26. Vipond, I.B. and Halford, S.E. (1993) *Mol. Microbiol.* **9**, 225–231.
27. Parker, C.N. and Halford, S.E. (1991) *Cell* **66**, 781–791.
28. Parker, C.N. (1990) Ph.D. Thesis, University of Bristol.
29. Avila, P., Ackroyd, A.J. and Halford, S.E. (1990) *J. Mol. Biol.* **216**, 645–655.
30. Stannsens, P., Opsomer, C., McKeown, Y., Kramer, W., Zabeau, M. and Fritz, H. (1989) *Nucleic Acids Res.* **17**, 4441–4454.
31. Castell, S.E. and Halford, S.E. (1989) *Nucleic Acids Res.* **17**, 7045–7058.

32. Frederick, C.A., Quigley, G.J., van der Marel, G.A., van Boom, J.H., Wang, A.H. and Rich, A. (1988) *J. Biol. Chem.* **263**, 17872–17879.
33. Fazakerley, G.V., Quignard, E., Teoule, R., Guy, A. and Guschlbauer, W. (1987) *Eur. J. Biochem.* **167**, 397–404.
34. Hatfull, G.F., Noble, S.M. and Grindley, N.D.F. (1987) *Cell* **49**, 103–110.
35. Mizuuchi, M., Baker, T.A. and Mizuuchi, K. (1992) *Cell* **70**, 303–311.
36. Bainton, R.J., Kubo, K.M., Feng, J. and Craig, N.L. (1993) *Cell* **72**, 931–943.
37. Mitchell, P.J. and Tjian, R. (1989) *Science* **245**, 371–378.
38. Stark, W.M., Boocock, M.R. and Sherratt, D.J. (1989) *Trends Genet.* **5**, 304–309.
39. Heitman, J. (1992) *BioEssays* **14**, 445–454.
40. Sarai, A. and Takeda, Y. (1989) *Proc. Natl. Acad. Sci., U.S.A.* **86**, 6513–6517.
41. Numrych, T.E., Gumport, R.I. and Gardner, J.F. (1990) *Nucleic Acids Res.* **18**, 3953–3959.
42. Martinez, E. and de la Cruz, F. (1990) *EMBO J.* **9**, 1275–1281.
43. Grinstead, J. and Brown, N.L. (1984) *Mol. Gen. Genet.* **197**, 497–502.