The EMBO Journal vol.12 no.5 pp.2141-2149, 1993

The *td* intron endonuclease I-*Tev*I makes extensive sequence-tolerant contacts across the minor groove of its DNA target

Mary Bryk[1], Susan M.Quirk, John E.Mueller, Nick Loizos[1], Charles Lawrence[2] and Marlene Belfort[2,3]

Molecular Genetics Program, Wadsworth Center for Laboratories and Research, New York State Department of Health, Empire State Plaza, Albany, NY 12201-0509, [1]Department of Microbiology, Immunology and Molecular Genetics, Albany Medical College, Albany, NY 12208 and [2]School of Public Health, State University of New York at Albany, Empire State Plaza, Albany, NY 12201-0509, USA
[3]Corresponding author

Communicated by C.Baglioni

I-*Tev*I, a double-strand DNA endonuclease encoded by the mobile *td* intron of phage T4, has specificity for the intronless *td* allele. Genetic and physical studies indicate that the enzyme makes extensive contacts with its DNA substrate over at least three helical turns and around the circumference of the helix. Remarkably, no single nucleotide within a 48 bp region encompassing this interaction domain is essential for cleavage. Although two subdomains (DI and DII) contain preferred sequences, a third domain (DIII), a primary region of contact with the enzyme, displays much lower sequence preference. While DII and DIII suffice for recognition and binding of I-*Tev*I, all three domains are important for formation of a cleavage-competent complex. Mutational, footprinting and interference studies indicate predominant interactions of I-*Tev*I across the minor groove and phosphate backbone of the DNA. Contacts appear not to be at the single nucleotide level; rather, redundant interactions and/or structural recognition are implied. These unusual properties provide a basis for understanding how I-*Tev*I recognizes T-even phage DNA, which is heavily modified in the major groove. These recognition characteristics may increase the range of natural substrates available to the endonuclease, thereby extending the invasive potential of the mobile intron.

Key words: DNA-protein interaction/intron homing site/mobile intron/modified DNA/phage T4

Introduction

Intron-encoded double-strand endonucleases involved in intron mobility have been described in prokaryotes and lower eukaryotes (reviewed by Dujon, 1989; Lambowitz, 1989; Perlman and Butow, 1989; Belfort, 1991; Clyman *et al.*, 1993; Lambowitz and Belfort, 1993). These endonucleases share the ability to recognize and cleave their respective intronless alleles, thereby initiating recombination events that result in intron inheritance. This process is referred to as intron homing, with the homing site comprising the intron insertion site and the endonuclease binding and cleavage sites (Dujon *et al.*, 1989). A distinctive feature of the prokaryotic endonucleases is that they cleave at a distance from the intron insertion site (Bell-Pedersen *et al.*, 1990; Chu *et al.*, 1990; Eddy and Gold, 1991). I-*Tev*I, the 29 kDa endonuclease encoded by the bacteriophage T4 *td* intron, binds an intronless *td* allele and cleaves 23-26 bp upstream of the intron insertion site (Bell-Pedersen *et al.*, 1990, 1991; Chu *et al.*, 1990). A further point of interest of the phage endonucleases is their ability to recognize and cleave both unmodified DNA and T-even phage DNA containing glucosylated hydroxymethylcytosine residues projecting into the major groove (Quirk *et al.*, 1989; Bell-Pedersen, 1991).

DNase I footprint analysis of the *td* homing site has shown that I-*Tev*I makes high-affinity contacts with sequences flanking the intron insertion site, whereas protection extends to the cleavage site at high I-*Tev*I concentrations (Bell-Pedersen *et al.*, 1991). Although disruption of the intron insertion site by insertion or deletion abolishes I-*Tev*I recognition, many missense mutations are tolerated throughout the homing site (Quirk *et al.*, 1989; Bell-Pedersen *et al.*, 1991; Chu *et al.*, 1991; M.Bryk, unpublished). In contrast, I-*Sce*I from the mitochondrial LSU rRNA intron of *Saccharomyces cerevisiae* tolerates only minimal degeneracy over an 18 bp region (Colleaux *et al.*, 1988). I-*Sce*II, from the aI4α intron of the *COXI* gene of *S.cerevisiae*, is less fastidious in recognition of its homing site than I-*Sce*I, but also has distinct sequence preferences at a number of sites (Sargueil *et al.*, 1990; Wernette *et al.*, 1992).

To examine further the interaction of I-*Tev*I with the *td* homing site, we are using a combination of genetic and physical approaches. With degenerate substrates and an enrichment scheme for isolating enzyme-substrate complexes, we have selected variants proficient in both binding and cleavage by I-*Tev*I. No single nucleotide within a 48 bp stretch containing the homing site is absolutely required for I-*Tev*I activity. This sequence tolerance and the ability of I-*Tev*I to act on modified phage DNA support the findings of protection and modification interference studies indicating that I-*Tev*I binds predominantly across the minor groove in close proximity to the phosphate backbone. I-*Tev*I thus joins a small, growing list of proteins that interact largely across the minor groove of DNA and that may recognize structural properties of DNA rather than strictly sequence determinants. The evolutionary implications of such interactions are considered in terms of the ability of an endonuclease to impart and extend the invasive potential of a mobile intron.

Results

The sequence-tolerant td homing site contains hypomutable regions

Previous studies indicated that a high degree of polymorphism is tolerated in the *td* homing site (Quirk *et al.*, 1989; Bell-Pedersen *et al.*, 1991; Chu *et al.*, 1991; M.Bryk, unpublished). To determine the sequence specificity of the

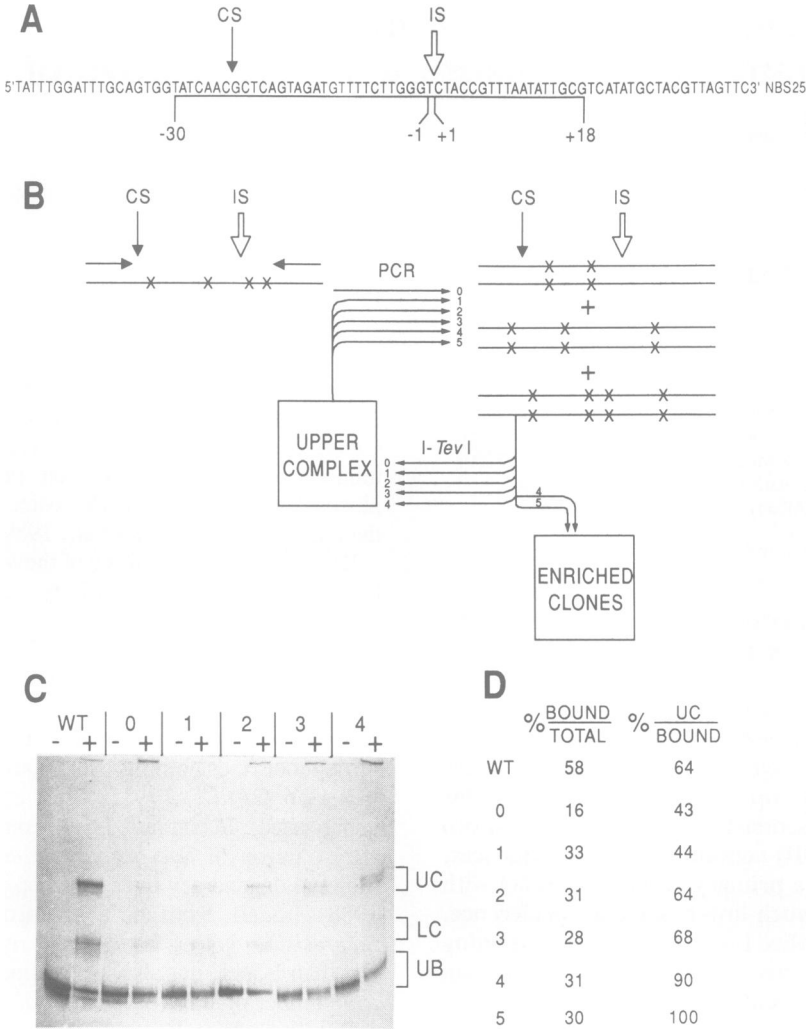

Fig. 1. Binding enrichment scheme. (**A**) The homing-site oligonucleotide (NBS25). The degenerate region is underlined. IS = intron insertion site; CS = I-*Tev*I cleavage site. Bases are numbered from −1 and +1 extending upstream and downstream of the IS, respectively. (**B**) Schematic of binding-enrichment drill. NBS25 was made double-stranded by PCR and incubated with I-*Tev*I; bound DNA (UC, panel C) was then isolated for use in the subsequent cycle. (**C**) Gel-shift analysis. PCR-amplified *td* homing sites from successive cycles of enrichment (0−4; cycle 5 not shown) or the wild type (WT) were incubated in the presence (+) or absence (−) of I-*Tev*I and electrophoresed. UB, unbound; LC, lower complex; UC, upper complex. (**D**) Quantification of binding enrichment. The percent of the homing site pool bound by I-*Tev*I [(UC+LC)×100/(UC+LC+UB)] and the percent in the upper complex relative to the total amount bound [(UC)×100/(UC+LC)] are shown for the WT and successive cycles (0−5).

I-*Tev*I−target interaction directly, an enrichment scheme was developed to isolate variant *td* homing sites that maintain the ability to bind I-*Tev*I despite extensive sequence alterations (Figure 1B). Mutant homing sites were prepared by synthesizing oligonucleotides that were contaminated to 15% with the other three nucleotides (Figure 1A). Degeneracy of the mutant oligonucleotides was confirmed to be 15% by cloning and sequencing (Table IA, line 1, 14.9%). The degenerate duplex pool was incubated with I-*Tev*I and separated by electrophoresis on native gels into three major species (Figure 1C): unbound DNA, a lower complex and an upper complex. For binding enrichment, the catalytically active upper complex was purified away from the inactive lower complex, which is an intermediate or by-product of the binding reaction (J.E.Mueller and M.Belfort, in preparation). Purified DNA from the upper complex was amplified by PCR and then used as the target for the subsequent round of binding. A doubling in the fraction of bound DNA, from 16% to 30%, was observed after five enrichment cycles (Figure 1D). Although the fraction bound did not reach wild-type levels (30% versus 58%), the relative amount of the active upper complex increased over five cycles of binding enrichment from 43% to 100%, exceeding the fraction of upper complex for the wild-type homing site (64%).

DNA from the bound fraction from cycles 4 and 5 was cloned and mutants were sequenced and classified according to their ability to be cleaved by I-*Tev*I (Figure 2; Table I). Comparison of the compiled sequences (Figure 2B) of cloned variants with that of the wild-type *td* homing site revealed that I-*Tev*I is, in general, extremely tolerant of base changes. In 62 cleavage-proficient variants (including two that were unselected), no single base within the 48 bp degenerate region was absolutely required for recognition by I-*Tev*I (Figure 2C). From Table I, which contrasts mutation frequency in noncleavable (−) and cleavable (+/− to +++) variants, it is apparent that the average mutation frequency among 60 enriched cleavage-proficient homing

Table I. Analysis of *td* homing site variants

	Selected	Cleavage proficiency[a]	No. sequenced[b]	Mutation average[c]	Mutation frequency (%)[d] Entire (48)	DI (8)	DII (7)	DIII (16)	Other (17)
A.1.	No	N.t.	19	7.2	14.9	15.8	13.6	14.1	15.8
2.	Yes	−	20	9.9	20.5	16.9	9.3	25.9	18.5
3.	Yes	+/− to +++	60[e]	4.7	9.8	4.4	3.3	10.8	13.9
B.1.	Yes	+/−	14	6.9	14.3	8.0	5.1	14.3	21.0
2.	Yes	+[f]	7	4.1	8.6	7.1	4.1	8.9	13.4
3.	Yes	++	33	4.0	8.3	2.7	2.6	9.5	12.3
4.	Yes	+++	6	3.8	8.0	2.1	2.4	12.5	8.8

[a]Cleavage proficiency was assayed as described in Materials and methods. N.t., not tested; '−', 0% cleaved in 5 min; '+/−', 5–50% cleaved in 5 min; '+', 40–50% cleaved in 1 min; '++', >50% cleaved in 1 min; '+++', >90% cleaved in 1 min. (In three cases, where assignment to + or ++ classes was ambiguous, cleavage of ≥90% at 5 min was used to assign variants to the ++ class.)
[b]Number of variants sequenced.
[c]Average number of mutations per 48 bp.
[d]Mutation frequencies presented for: 'Entire' = 48-base degenerate region (bp −30 to +18); 'DI' = hypomutable domain in exon I (bp −21 to −14); 'DII' = hypomutable domain in exon II (bp +9 to +15); 'DIII' = (bp −8 to +8), defined on the basis of physical studies (Bell-Pedersen *et al.*, 1991); 'Other' = remaining 17 bases in degenerate region, collectively. The number of base pairs in each region is indicated in parentheses.
[e]This set was analyzed statistically (Materials and methods). The breakdown of the cleavage-proficient variants from line 3 of part A is presented in part B.
[f]Wild-type *td* homing site is a member of the (+) class.

sites was reduced from 14.9% to 9.8% (Table IA, lines 1 and 3). In contrast, for the cleavage-minus variants, the mutation frequency over the 48 bp region was 20.5% (Table IA, line 2). These results suggest that, although binding is tolerant of sequence alterations, constraints exist for the formation of cleavage-competent complexes.

To determine whether there are specific regions of the cleavage-proficient variants that are more sensitive to mutation than others, a two-stage statistical analysis was performed (Materials and methods). In the first, exploratory phase with 24 variants, two regions appeared conserved. These domains are DI in exon I (8 bp, −21 to −14) and DII in exon II (7 bp, +9 to +15) (Figure 2B and C; mutation frequencies of 3.6% and 2.4%, respectively). In the second, confirmatory phase with 36 variants, the hypomutability of DI and DII was confirmed: the observed mutation frequencies, 4.9% and 4.0% respectively, are significantly lower than the input level of 15% ($P < 0.05$). The mutation frequencies of DI and DII for all 60 cleavage-competent variants are 4.4% and 3.3%, respectively (Table IA, line 3). Of the 20 binding-proficient mutants that are not cleaved (Table IA, line 2), DII was conserved (mutation frequency of 9.3%; $P < 0.05$ for the significance of the difference from 15%), suggesting a requirement for DII in I-*Tev*I binding, while DI was not (mutation frequency of 16.9%), indicating a possible role for DI in the formation of the catalytically active complex.

Surprisingly, the region flanking the intron insertion site, previously shown to constitute a portion of the high-affinity I-*Tev*I binding site (Bell-Pedersen *et al.*, 1991), which we term DIII (16 bp, −8 to +8) (Figure 2C), was significantly more tolerant of mutation than DI and DII ($P < 0.05$) although, at a mutation frequency of 10.8%, it was conserved relative to the input degeneracy of 15% ($P < 0.05$; Table IA, line 3). The importance of this domain in generation of the cleavage-competent complex is further underscored by comparing the mutation frequency of DIII for the cleavage-proficient variants (Table IA, line 3; 10.8%) with the mutants unable to be cleaved (Table IA, line 2; 25.9%). In contrast with DI, DII and DIII, the other 17 positions in the 48 bp homing site ('Other', Table I), which collectively had a mutation frequency of 13.9%, were unconserved ($P > 0.05$).

By grouping variants into cleavage classes, it became clear that the mutation frequency over 48 bases decreased as cleavage efficiency increased (Figure 2B, Table IB). Although sample size within each class precludes a rigorous statistical analysis, it is interesting that the reduced tolerance for mutation is reflected in a steady decrease in mutation frequency for both DI and DII as cleavage ability increased (Table IB). In contrast, no steady decline was observed for DIII. Instead, the mutation frequency first dropped and then increased with increasing cleavage efficiency (Table IB, lines 1–4). That is, for those variants that are cleaved more efficiently by I-*Tev*I than the wild-type *td* homing site (++ and +++), sequence constraints on DI and DII are heightened, whereas this is not the case for DIII. It remains to be determined whether the increased mutation frequency in DIII in the +++ class reflects a statistical fluctuation, or whether these mutations in DIII (Figure 2B and C) are responsible for the enhancement of function.

Physical evidence for I-TevI interactions at discrete domains

To evaluate physical contacts made by I-*Tev*I at the homing site, hydroxyl radical protection experiments were performed with the active upper complex and inactive lower complex using Fe(II)EDTA reagent (Figure 3). The strongest region of hydroxyl radical protection flanks the intron insertion site. In a typical Fe(II)EDTA cleavage experiment, a second region of protection is evident in the upper complex, downstream from the I-*Tev*I cleavage site (Figure 3A and B). Interestingly, I-*Tev*I does not protect the nucleotides that immediately flank the cleavage site on the top strand. Protection of the bottom strand in this region is unclear,

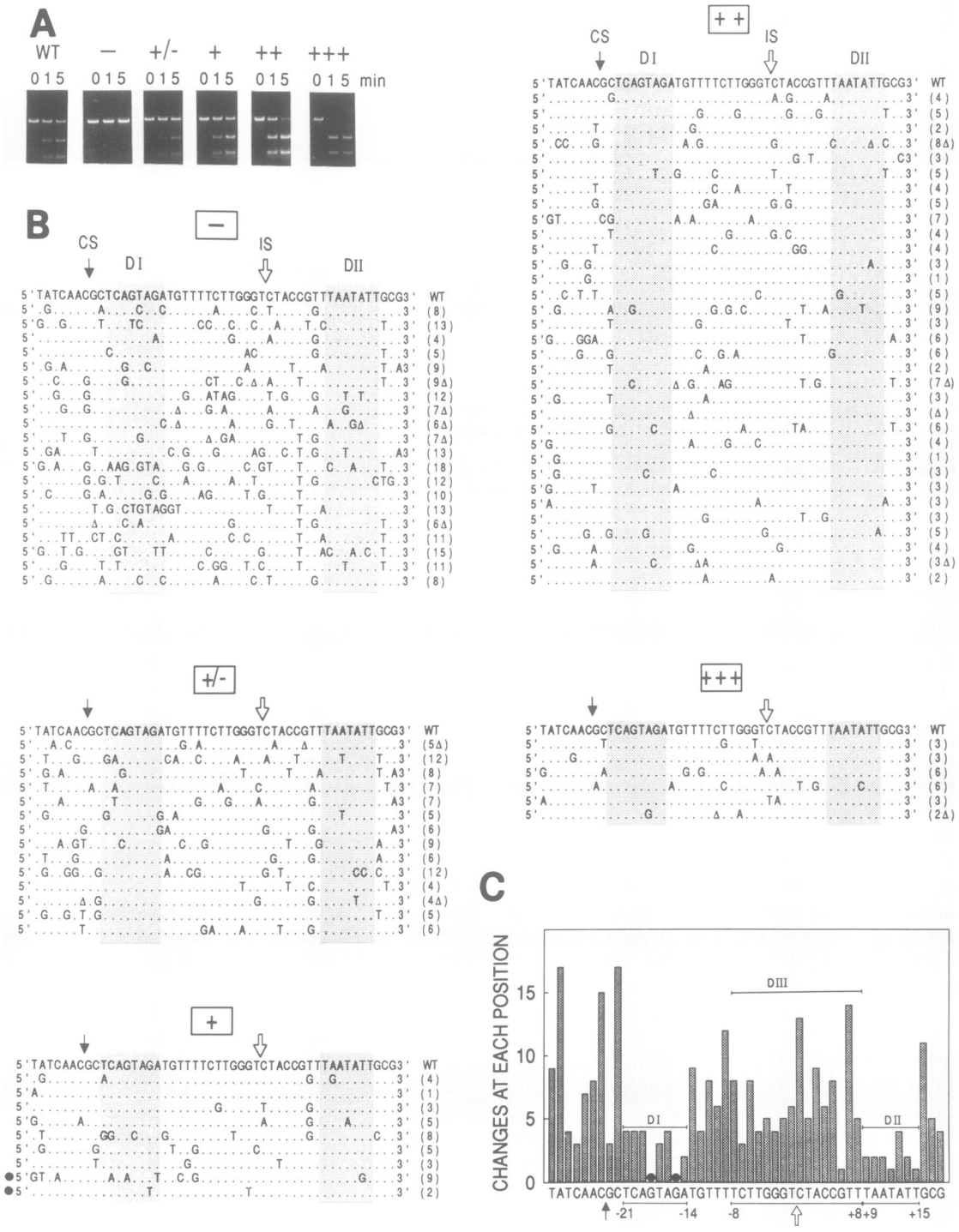

Fig. 2. Compilation of variant *td* homing sites. **(A)** Cleavage assays (0, 1 or 5 min). Representatives of each cleavage class, as defined in the legend to Table I, are shown. **(B)** Cleavage classification. The wild-type sequence (WT) is shown at the top of each class. The number of changes and the presence of deletions (Δ) is indicated in parentheses. The sequences marked ● in the (+) class were not generated by the enrichment scheme, but show alterations at −15G and −18G (see panel C). DI and DII (shaded) = hypomutable domains. IS, CS as in Figure 1. Average number of mutations in each class is presented in Table I. **(C)** Histogram of the number of mutations at each position for 60 cleavage-proficient variants. DIII = interaction domain defined by previous physical studies (see Bell-Pedersen *et al.*, 1991). DI, DII and ●, as in B.

because it is obscured by cleavage of this strand by I-*Tev*I in the presence of Fe(II)EDTA reagents (Figure 3).

In contrast to the upper complex, I-*Tev*I protects only the region near the insertion site in the lower complex (Figure 3B), with no consistent protection apparent in the region extending towards the cleavage site. These results corroborate the genetic experiments in which DI, near the cleavage site, was hypomutable only in the cleavage-competent variants (Table I), implying that contacts near the cleavage site are important for the cleavage event.

Inhibition of the I-TevI–DNA interaction by phosphate ethylation

Given the extensive interaction between I-*Tev*I and the *td* homing site and considering the sequence tolerance of the site, it became of interest to probe the involvement of the

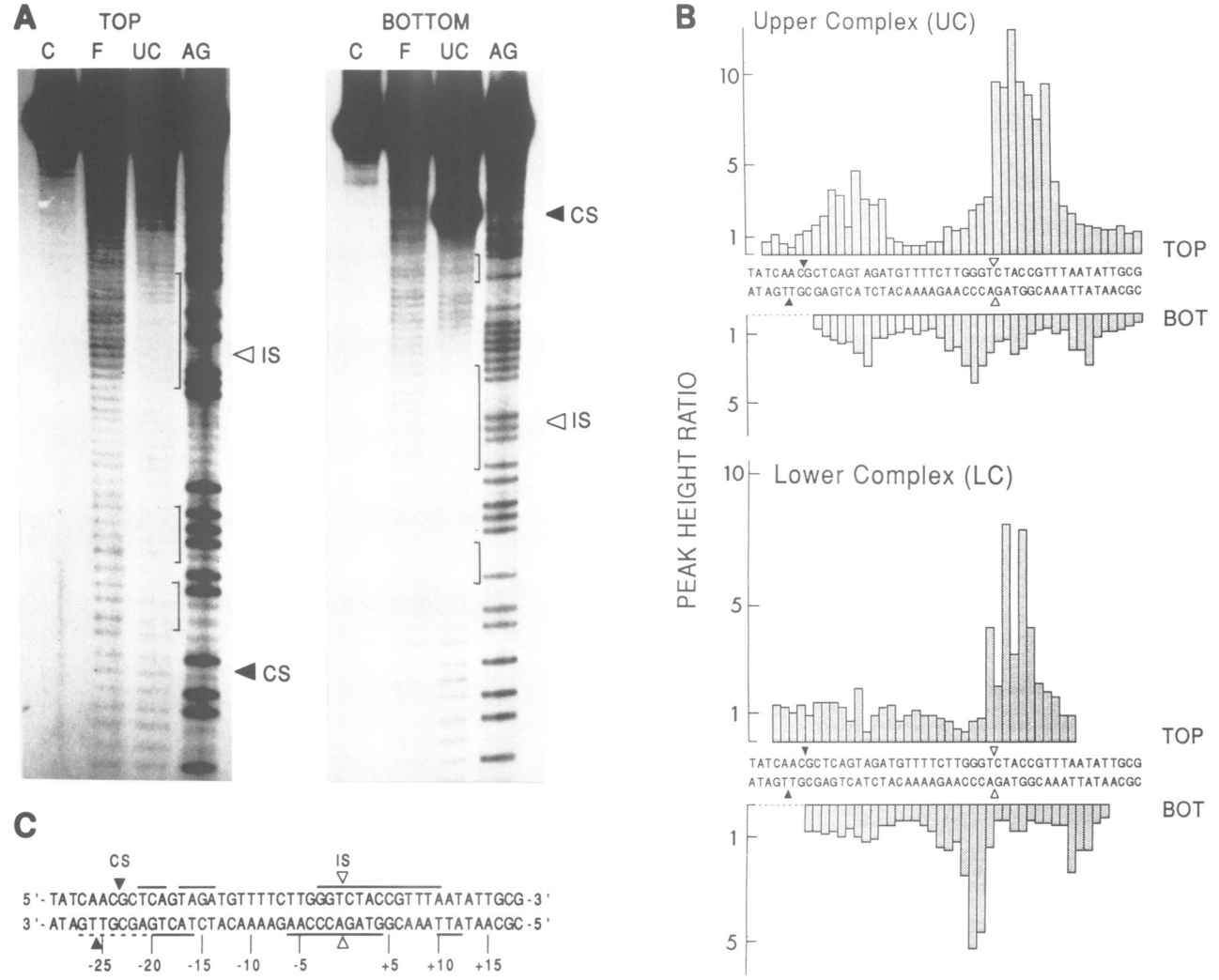

Fig. 3. Hydroxyl radical protection analysis. (**A**) Footprint of I-*Tev*I-homing site. Top and bottom refer to individually labeled strands (OP46 and OP47, see Materials and methods). Lanes: C, untreated duplex; F, free duplex treated with Fe(II)EDTA in the absence of I-*Tev*I; UC, upper complex treated with Fe(II)EDTA; AG, products from a modified Maxam–Gilbert A+G reaction (Bencini *et al.*, 1984). IS, CS as in Figure 1. Brackets represent protected areas. (**B**) Protection by I-*Tev*I for the upper and lower complexes. Peak heights for individual bases were determined by densitometry and expressed as ratios of F to UC or LC (autoradiogram not shown). Dashed lines represent region obscured by I-*Tev*I cleavage (also in C, below). Ratios ≥1.75 for the top strand and ≥1.4 for the bottom strand indicate I-*Tev*I protection. (**C**) Schematic of hydroxyl radical footprint (solid lines) for the upper complex. These data are representative of the results from at least three different experiments, which were reproducible except for small variations at the borders of the protected regions.

phosphate backbone. The results of a typical ethylation interference experiment are shown in Figure 4. Ethylation of phosphates flanking the insertion site interferes with upper and lower complex formation (Figure 4A and B). In contrast, ethylation of phosphates extending towards the cleavage site interfere only with upper complex formation, paralleling the hydroxyl radical protection results (cf. Figure 3B and 4B).

Chemical probing of major and minor groove interactions

Dimethylsulfate (DMS) modification of purine residues was used to examine groove-specific interactions of I-*Tev*I with the *td* homing site (Figure 5). Interference patterns indicate that methylation of purines flanking the insertion site interferes with both upper and lower complex formation (Figure 5A). Again, purine modifications approaching the cleavage site interfere only with upper complex formation (Figure 5A 'TOP', cf. UC and LC lanes), consistent with the foregoing genetic and physical studies (Figures 2–4). As with ethylation, no modification completely eliminated binding. Results of numerous methylation experiments are summarized in Figure 5B to indicate the purines that interfere with I-*Tev*I binding relative to free DNA (F lane) (●), and those that only interfere relative to the unbound DNA fraction (UB lane) (○). This latter fraction contains enhanced bands due to modified DNA molecules that are excluded from the bound fractions. Of the seven more strongly interfering nucleotides, six are A residues, with three As and one G in sequences flanking the insertion site, and three As in the region approaching the cleavage site (Figure 5B). Considering that DMS methylates adenine residues at N3 in the minor groove and guanine residues at N7 in the major groove (Siebenlist and Gilbert, 1980), the methylation interference data suggest predominantly minor groove interactions around the intron insertion site and approaching the cleavage site of the upper complex. These inferences are confirmed by the other physical data, which are summarized in Figure 6 and will be considered further in the Discussion.

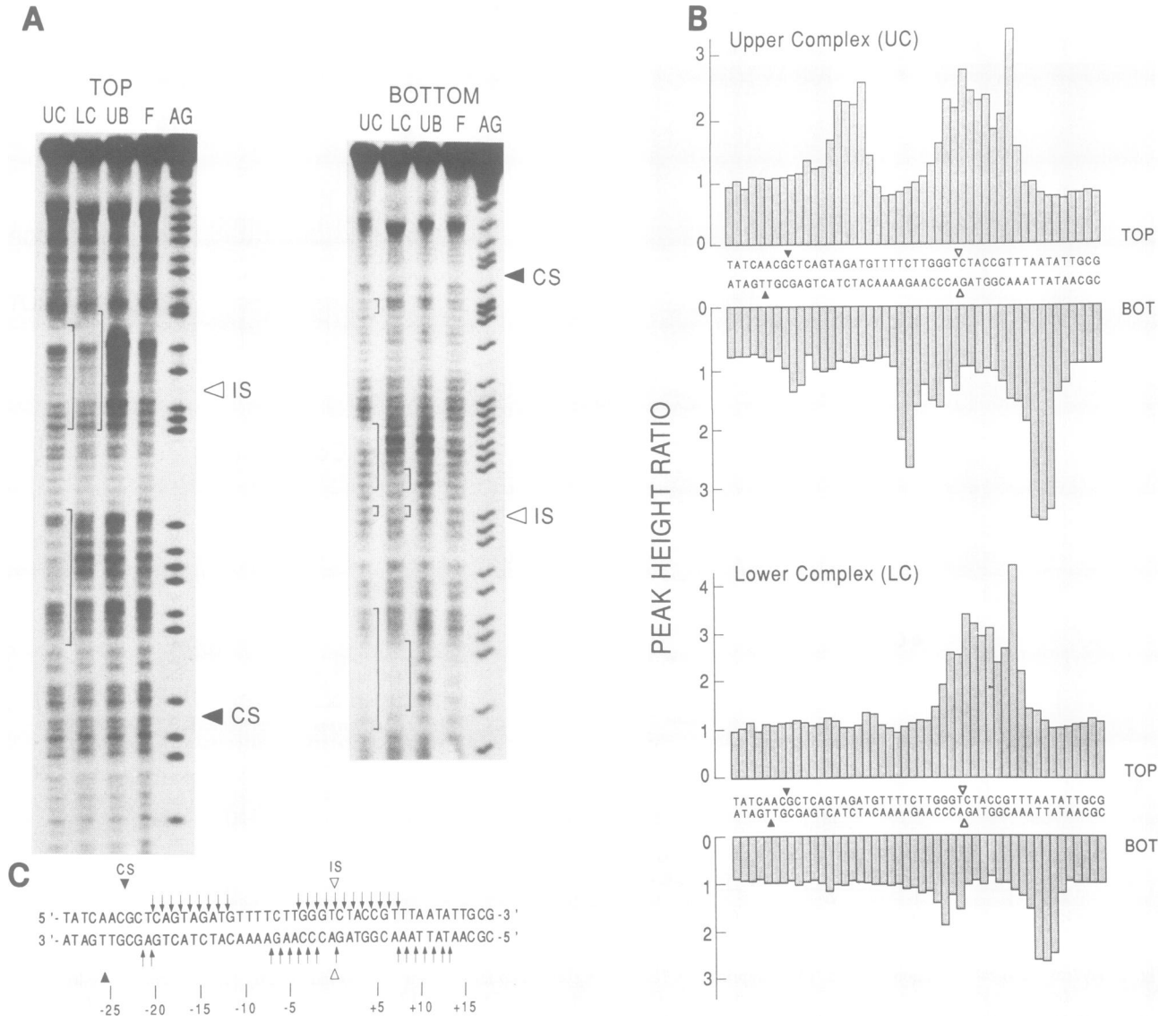

Fig. 4. Ethylation interference analysis. (**A**) Interference of I-*Tev*I binding by DNA ethylation. Lanes: UC, upper complex; LC, lower complex; UB, unbound DNA; F, free ethylated DNA; AG, G>A Maxam−Gilbert sequencing reaction. Other labels as for Figure 3A. (**B**) Graphical representation of interference patterns. Bar height represents the effect of ethylation of the 5′ phosphate of the corresponding base. Analysis as for Figure 3B except that peak-height ratios of UB to UC or LC are shown. Results were qualitatively similar when peak-height ratios were determined against the F lane. (**C**) Schematic representation of the upper complex. Arrows indicate ethylation interference (peak-height ratio ≥ 1.3). These results were reproducible over three different experiments, except for minor variations in intensity at the interference boundaries.

Discussion

The genetic and physical analyses combine to underscore several distinctive features of the I-*Tev*I−target interaction. Remarkably, the *td* homing site variants revealed that no individual base within 48 bp is absolutely required for the interaction (Figure 2). Furthermore, the interaction between I-*Tev*I and the homing site is extensive, spanning at least three helical turns and occupying the entire circumference of the helix (Figure 6). Additionally, the homing site can be divided into three domains, DI, DII and DIII, based on genetic and physical criteria. DI (in exon I) and DII (in exon II) are domains in which sequence preferences are more clearly manifest than in DIII, flanking the insertion site. By physical criteria, DII and DIII form a contiguous region important in formation of both inactive and cleavage-competent I-*Tev*I−DNA complexes. By contrast, DI has been implicated specifically in formation of the active complex. Finally, there appear to be extensive interactions involving both the minor groove and the phosphate backbone.

A broad body of evidence is accumulating in favor of interaction of I-*Tev*I predominantly across the minor groove. First, hydroxyl radical protection and ethylation interference are apparent across the minor groove both around the insertion site (DIII) and towards the cleavage site (DI) (summarized and most clearly apparent in Figure 6). Second, complementary strands at the insertion site (DIII) and the downstream rim of the cleavage site region (DI) exhibit peaks of hydroxyl radical protection (Figures 3B and 6) as well as ethylation interference (Figures 4B and 6) that are shifted by several bases in the 3′ direction, supportive of minor groove interactions (Dervan, 1986). Third, direct evidence against major groove interactions upstream of the intron

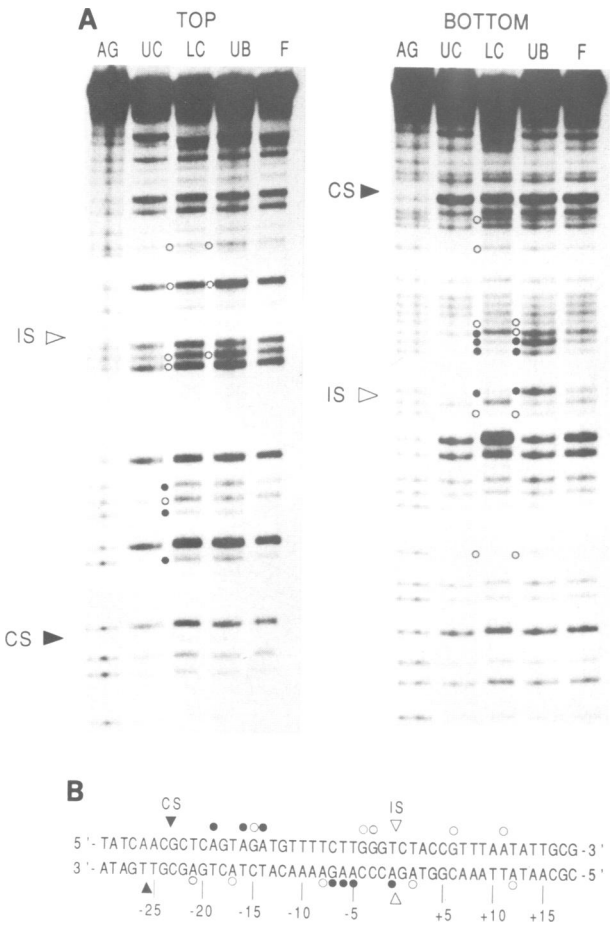

Fig. 5. Methylation interference analysis. (**A**) Interference of I-*Tev*I interaction by DNA methylation. Labeling is as in Figure 4 except that F = free methylated DNA. Modified purine residues that interfered with I-*Tev*I binding over three experiments are marked: ●, interference evident against F and UB; ○, interference evident against UB only. (**B**) Schematic representation of interference of the UC.

insertion site was provided by creating, by site-specific mutagenesis, a *dam* methylation site (GATC) within 2 bp of the intron insertion site (G−2A) (Figure 6A). Methylation at this site did not reduce cleavage efficiency, suggesting that I-*Tev*I is insensitive to methylation of adenine in the major groove at −2 (M.Bryk, unpublished). Fourth, although the physical data do not support minor groove interactions exclusively, the minor groove seems dominant, based on the observation that there are more methylated As (six) than Gs (one) which interfere strongly with binding in both DI and DIII (Figures 5B and 6). It remains unclear whether guanine interference is due to isolated major groove contacts near regions of minor groove interaction, or whether this results from indirect inhibitory effects on complex formation (Siebenlist and Gilbert, 1980; Yang and Nash, 1989; Derbyshire and Grindley, 1992).

Interaction across the minor groove and phosphate backbone, the least discriminatory determinants of DNA sequence (Seeman *et al.*, 1976), is consistent with the relaxed specificity of I-*Tev*I (Figures 1 and 2). It is not yet clear to what degree I-*Tev*I interaction involves discriminatory contacts within the minor groove and to what degree it depends on the structural properties of DNA dictated by nucleotide sequence. Nevertheless, interaction outside the major groove is in accord with the ability of the enzyme to cleave T-even phage DNA, which has bulky glucosyl groups protruding into the major groove from all hydroxymethylcytosine residues.

Both members of the prokaryotic HU/IHF family (Tanaka *et al.*, 1984; Yang and Nash, 1989) and TFIID, a eukaryotic transcription accessory which interacts with the TATA box, bind primarily to the minor groove of their respective targets (Lee *et al.*, 1991; Starr and Hawley, 1991). I-*Tev*I contacts may resemble TFIID− and IHF−DNA interactions, although these latter proteins appear to be more sequence-specific than I-*Tev*I. However, like I-*Tev*I, IHF and TFIID envelop the helix. Interestingly, IHF and TFIID share a common structural motif, the antiparallel β-ribbon, capable of forming arms that can encircle the DNA (Tanaka *et al.*, 1984; Nash and Granston, 1992). Likewise, the N-terminal region of I-*Tev*I, which contains a conserved sequence motif (GIY...YIG; Michel and Dujon, 1986) likely to be part of the enzyme's active site, is predicted by Chou−Fasman and GOR analyses (Devereux *et al.*, 1984) to form a β-ribbon (M.Bryk, unpublished). Experiments in progress will address the functional relevance of the common features in these systems.

The incomplete interference caused by base modification within the homing site (Figures 4 and 5) and the high tolerance of I-*Tev*I for sequence polymorphisms in its substrate (Figure 2) are likely to be related phenomena, which may be explained in one of two ways. One interpretation is that the enzyme makes redundant contacts with its DNA target, consistent with the failure to completely inhibit binding by modification of any particular nucleotide [Figures 4 and 5 and hydroxyl radical interference (missing nucleotide) analysis, J.E.Mueller, unpublished]. Additionally, we have isolated a series of single- and double-mutant pairs (C+4A, A+11C and C+4A:A+11C) where the single mutants have barely detectable phenotypes but the double mutant is highly cleavage-defective (M.Bryk, unpublished). These observations may suggest that I-*Tev*I interacts with both residues +4 and +11, and that either contact will maintain function but eliminating both determinants seriously impairs the enzyme−DNA interaction. The alternative interpretation is that I-*Tev*I recognizes structural features of the DNA. Thus binding interference might be partial if more than one base modification is required to disrupt a structural element. Similarly, the structural component may be preserved in the single mutants (C+4A and A+11C), but disrupted in the corresponding double mutant (C+4A:A+11C).

In the context of DNA structure, it is informative to examine the hydroxyl radical cleavage data. Hydroxyl radicals are sensitive probes for structural distortion of DNA in the minor groove (Burkhoff and Tullius, 1987). The hydroxyl radical cleavage pattern on free *td* homing site DNA reproducibly indicates that the top-strand bases at the intron insertion site are more susceptible to cleavage than neighboring residues, suggesting a structural perturbation in DIII (Figure 3A, lanes F). Also, the A/T-rich regions of the homing site (bp −14 to −5 and bp +7 to +15) are resistant to DNase I cleavage in the absence of I-*Tev*I compared with other regions of the homing site (Bell-Pedersen *et al.*, 1991), further suggesting local structural anomalies, such as differences in the width of the minor groove (Drew and Travers, 1985). The latter DNase I-

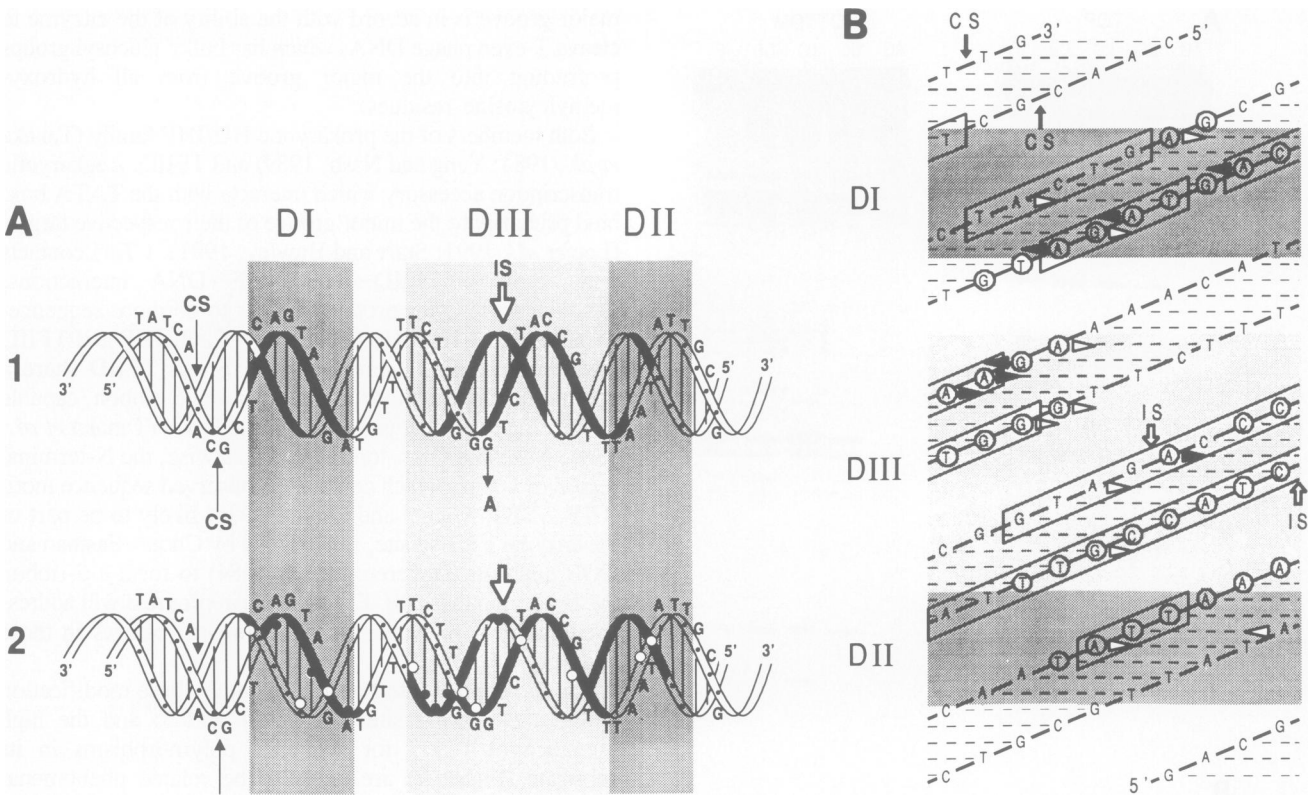

Fig. 6. Summary of physical and genetic analyses of the I-*Tev*I–homing site interaction. (**A**) Helical representation. Heavily shaded areas indicate the hypomutable domains DI and DII and lightly shaded panel represents DIII. (1) Hydroxyl radical protection. Protected regions are filled in black. Arrow under helix shows G–2A mutation creating a *dam* methylation site. (2) Ethylation and methylation interference. Areas filled in black represent regions of ethylation interference. ○ and ● represent methylation inferference as in Figure 5. (**B**) Planar diagram. Shading is as in A. Open boxes indicate hydroxyl radical protection. Circles around bases indicate interference due to ethylation of the corresponding 5′ phosphate. Triangles represent methylation interference with filled and open symbols corresponding to the circles in Figure 5. Arrows pointing to the minor groove or major groove correspond to methylated A and G residues, respectively. IS and CS are as in Figure 1.

resistant region includes the hypomutable DII domain. Methylation, ethylation and hydroxyl radical interference (missing nucleotide) analyses (Figure 6 and data not shown) indicate that DII is an important recognition element, despite only partial contact detected by hydroxyl radical protection analysis. A possible explanation for incomplete protection in the face of modification interference is that DII assumes a recognition structure that is disrupted by modification. There are precedents for non-contacted bases influencing formation of protein–DNA complexes, including the P22 and lambda repressor–operator interactions (Koudelka *et al.*, 1987, 1988; Wu *et al.*, 1992) and SU(Hw) protein interaction with specific sequences of the *Drosophila* gypsy element (Spana and Corces, 1990). It is of further interest that the 7 bp DII is composed entirely of A and T residues and that poly(dA/dT) tracts are structurally distinct from B-DNA, having both deeper and narrower minor grooves (Burkhoff and Tullius, 1987).

Modified DNAs are resistant to restriction endonucleases when the modified nucleotide is within the recognition sequence of the enzyme. This is true for adenine-methylated as well as glucosylated hydroxymethylcytosine DNAs. Unlike restriction enzymes, which make base-specific contacts predominantly in the major groove (for example, *Eco*RI; reviewed in Jen-Jacobson *et al.*, 1991), I-*Tev*I cleaves modified DNAs very efficiently and with equivalent specificity to unmodified substrates (Bell-Pedersen *et al.*, 1991; this work). Clearly, sequence-tolerant recognition based on interactions across the minor groove and phosphate backbone and possibly involving structural features of the DNA extends the range of substrates that are susceptible to cleavage by this intron endonuclease. These recognition properties would increase the potential of the intron-encoded endonuclease to act on a variety of substrates in the wild, thereby extending the invasive properties and potential for dissemination of the mobile intron.

Materials and methods

Oligonucleotides

Mutant *td* homing sites were selected from a degenerate oligonucleotide pool (NBS25) synthesized with phosphoramidite mixtures contaminated to 15% with each of the other three bases at 48 positions (Figure 1A). The PCR primers used to amplify the degenerate *td* homing sites were NBS26 (5′-AAAGGATCCGAACTAACGTAGCATATGA-3′) and NBS27 (5′-AAAGAATTCTATTTGGATTTGCAGTGG-3′), which contain terminal *Bam*HI and *Eco*RI sites, respectively. Physical analysis of the I-*Tev*I–DNA interaction was performed with a synthetic 80 bp *td* homing-site duplex formed with purified oligonucleotide OP46 (5′-AATTCGGATTTGCAGTGGTATCAACGCTCAGTAGATGTTT-TCTTGGGTCTACCGTTTAATATTGCGTCATATGCTACGTTAGT-TCG-3′) and its complement OP47 (5′-GATCCGAACTAACGTAGCA-TATGACGCAATATTAAACGGTAGACCCAAGAAAACATCTACTG-AGCGTTGATACCACTGCAAATCCG-3′).

I-TevI preparations

I-*Tev*I was partially purified from *Escherichia coli* (Bell-Pedersen *et al.*, 1991) or synthesized *in vitro* in wheat germ extracts (Promega, Madison, WI) (Bell-Pedersen *et al.*, 1989). One unit of activity is defined as the amount

of enzyme required to cleave 250 ng (78 fmol) of linearized target (~5 kb) to 50% completion in 1 min at 37°C. Physical analyses and the binding enrichment were performed with more highly purified *in vivo* preparations (0.1–0.25 U/μl; 0.25–0.3 μg/μl total protein). Cleavage assays to classify *td* homing-site mutants utilized the more active *in vitro* I-*Tev*I preparations (5–10 U/μl; 100 μg/μl total protein).

Enrichment for functional td homing site mutants

Homing-site variants enriched for the ability to bind I-*Tev*I were isolated by a series of sequential I-*Tev*I binding and PCR amplification steps. The degenerate oligonucleotide NBS25 (25 pmol) (Figure 1) was amplified using a Perkin Elmer–Cetus (Norwalk, CT) PCR kit with ^{32}P-labeled primers NBS26 and NBS27. Double-stranded DNA was eluted from native 12% gels [29:1, acrylamide:bisacrylamide in 90 mM Tris–borate, 2 mM EDTA pH 8.0 (TBE)] in 0.5 M ammonium acetate, extracted with phenol/chloroform/isoamyl alcohol and ethanol-precipitated. The degenerate pool of duplex homing sites (10–15 pmol) was then incubated with 1.0–1.5 U of I-*Tev*I in binding buffer [50 mM Tris–HCl pH 8.0, 20 μg/ml poly(dI/dC), 10 μg/ml BSA] for 5 min. TBE dye mix (TBE containing 0.01% bromophenol blue, 0.01% xylene cyanol, 50% glycerol) was added to 1/10 vol and I-*Tev*I–DNA complexes were isolated from native gels, as above. The enriched pools from cycles 4 and 5 were cloned into pTZ18U (United States Biochemical). Mutant *td* homing sites were sequenced using the M13 universal primer (5′-GTAAAACGACGGCCAGT-3′).

Statistical methods

A two-stage sampling procedure (Tukey, 1977) was used to identify conserved regions within the functional homing-site variants. In the first stage, the exploratory phase, 24 variants were examined and two hypomutable regions, DI (8 bp, −21 to −14) and DII (7 bp, +9 to +15), were identified on the basis of lower mutation frequency (Figure 2C). In the second stage, the confirmatory phase, 36 variants were examined, where the normal approximation to the binomial was employed to compare the level of mutation in DI and DII with the input frequency of 15%. To assess variation in DIII (16 bp, −8 to +8), a region defined by prior physical analysis, the observed mutation frequency of all 60 cleavage-proficient variants was compared with the 15% input degeneracy using the normal approximation to the binomial. To test the hypothesis that DI and DII were more conserved than DIII, a χ^2 analysis comparing the mutation frequencies in DI and DII versus DIII was used.

Activity assays

*Sca*I-linearized pTZ18U containing either the wild-type or variant *td* homing sites (200 ng; ~3.0 kb) were incubated with 0.5 U I-*Tev*I at 37°C in assay buffer (50 mM Tris–HCl pH 8.0, 100 mM NaCl, 10 mM $MgCl_2$). At 1 or 5 min, the reaction was stopped by adding stop-load buffer (0.05 mM EDTA, 5% SDS, 25% glycerol, 0.25% bromophenol blue) and samples were electrophoresed in 1% agarose gels in TAE (40 mM Tris–acetate, 1 mM EDTA). The extent of cleavage was estimated by densitometry (Hoefer GS-300) with negative films of ethidium bromide-stained gels.

Oligonucleotide labeling and duplex formation for physical analyses

Non-purified oligonucleotides (800 pmol) were end-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase. Full-length oligonucleotides were purified from denaturing (8.3 M urea) 10% TBE gels (19:1, acrylamide: bisacrylamide). Purified ^{32}P-labeled oligonucleotides (~100 pmol) were annealed to 200 pmol of their non-labeled HPLC-purified complements by heating to 90°C for 5 min and slow cooling.

Hydroxyl radical protection analyses

Hydroxyl radical footprinting was done according to Tullius and Dombroski (1985) with the following modifications. I-*Tev*I (30 μl; 3.0–7.5 U) was incubated with 20 pmol duplex in binding buffer for 1 min at 23°C. The mixture was then adjusted to 10 mM L-ascorbic acid, 0.15% H_2O_2, 0.1 mM $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$, 0.2 mM EDTA. After 1 min the reaction was quenched with 16 μl TBE dye mix. Free DNA and bound complexes were isolated from native gels and separated on denaturing gels. Dried gels were exposed to X-ray film, which was analyzed by densitometry.

Ethylation and methylation interference

End-labeled DNA was modified by treatment with ethylnitrosourea or dimethylsulfate as described by Siebenlist and Gilbert (1980). Modified DNA (1 pmol) in binding buffer was incubated with 0.05–0.1 U I-*Tev*I for 2 min at 23°C. I-*Tev*I–DNA complexes, unbound DNA and free DNA were eluted and cleaved at modified positions with NaOH (Hendrickson and Schleif, 1985), separated on 10% denaturing gels and analyzed by densitometry.

Acknowledgements

We thank Maureen Belisle, Justine DeVost and Dorie Monroe for technical assistance, Tom Tullius and Amy Kimball for advice on hydroxyl radical footprinting, Bob Cowan for useful discussions and artwork, Maryellen Carl for expert manuscript preparation and J.Clyman, D.Court and M.Reaban for critical comments on the manuscript. This work was funded by grants from the National Institutes of Health GM44844 and GM39422 to M.B. and GM13582 to S.M.Q.

References

Belfort,M. (1991) *Annu. Rev. Genet.*, **24**, 363–385.
Bell-Pedersen,D. (1991) Dissertation. State University of New York, Albany, NY.
Bell-Pedersen,D., Quirk,S., Clyman,J. and Belfort,M. (1990) *Nucleic Acids Res.*, **18**, 3763–3770.
Bell-Pedersen,D., Quirk,S.M., Bryk,M. and Belfort,M. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 7719–7723.
Bencini,D.A, O'Donovan,G.A. and Wild,J.R. (1984) *BioTechniques*, **2**, 4–5.
Burkhoff,A.M. and Tullius,T.D. (1987) *Cell*, **48**, 935–943.
Chu,F.K., Maley,G., Pedersen-Lane,J., Wang,A.-M. and Maley,F. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 3574–3578.
Chu,F.K., Maley,F., Wang,A.-M., Pedersen-Lane,J. and Maley,G. (1991) *Nucleic Acids Res.*, **19**, 6863–6869.
Clyman,J., Quirk,S. and Belfort,M. (1993) In Karam,J. *et al.* (eds), *Molecular Biology of Bacteriophage T4*. American Society for Microbiology, Washington, DC, in press.
Colleaux,L., D'Auriol,L., Galibert,F. and Dujon,B. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 6022–6026.
Derbyshire,K.M. and Grindley,N.D.F. (1992) *EMBO J.*, **11**, 3449–3456.
Dervan,P.B. (1986) *Science*, **232**, 464–471.
Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
Drew,H.R. and Travers,A.A. (1985) *Nucleic Acids Res.*, **13**, 4445–4467.
Dujon,B. (1989) *Gene*, **82**, 91–114.
Dujon,B., Belfort,M., Butow,R.A., Jacq,C., Lemieux,C., Perlman,P.S. and Vogt,V.M. (1989) *Gene*, **82**, 115–118.
Eddy,S.R. and Gold,L. (1991) *Genes Dev.*, **5**, 1032–1041.
Hendrickson,W. and Schleif,R. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 3129–3133.
Jen-Jacobson,L., Lesser,D.R. and Kurpiewski,M.R. (1991) In Eckstein,F. and Lilley,D.M.J. (eds), *Nucleic Acids and Molecular Biology*. Springer Verlag, Berlin, Vol 5, pp. 141–170.
Kouldelka,G.B., Harrison,S.C. and Ptashne,M. (1987) *Nature*, **326**, 886–888.
Kouldelka,G.B., Harbury,P., Harrison,S.C. and Ptashne,M. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 4633–4637.
Lambowitz,A.M. (1989) *Cell*, **56**, 323–326.
Lambowitz,A.M. and Belfort,M. (1993) *Annu. Rev. Biochem.*, **62**, 87–622.
Lee,D.K., Horikoshi,M. and Roeder,R.G. (1991) *Cell*, **67**, 1241–1250.
Michel,F. and Dujon,B. (1986) *Cell*, **46**, 323.
Nash,H.A. and Granston,A.E. (1991) *Cell*, **67**, 1037–1038.
Perlman,P.S. and Butow,R.A. (1989) *Science*, **246**, 1106–1109.
Quirk,S.M., Bell-Pedersen,D. and Belfort,M. (1989) *Cell*, **56**, 455–465.
Sargueil,B., Hatat,D., Delahodde,A. and Jacq,C. (1990) *Nucleic Acids Res.*, **18**, 5659–5665.
Seeman,N.C., Rosenburg,J.M. and Rich,A. (1976) *Proc. Natl. Acad. Sci. USA*, **73**, 804–808.
Siebenlist,U. and Gilbert,W. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 122–126.
Spana,C. and Corces,V.G. (1990) *Genes Dev.*, **4**, 1505–1515.
Starr,D.B. and Hawley,D.K. (1991) *Cell*, **67**, 1231–1240.
Tanaka,I., Appelt,K., Dijk,J., White,S.W. and Wilson,K.S. (1984) *Nature*, **310**, 376–381.
Tukey,J.W. (1977) *Exploratory Data Analysis*. Addison-Wesley Publishing Co., Reading, MA.
Tullius,T.D. and Dombroski,B.A. (1985) *Science*, **230**, 679–681.
Wernette,C.M., Saldanha,R., Smith,D., Ming,D., Perlman,P.S. and Butow,R.A. (1992) *Mol. Cell. Biol.*, **12**, 716–723.
Wu,L., Vertino,A. and Kouldelka,G.B. (1992) *J. Biol. Chem.*, **267**, 9134–9139.
Yang,C.-C. and Nash,H.A. (1989) *Cell*, **57**, 869–880.

Received on November 25, 1992; revised on February 10, 1993