Related homing endonucleases I-*Bmo*I and I-*Tev*I use different strategies to cleave homologous recognition sites

David R. Edgell*† and David A. Shub†

Department of Biological Sciences and Center for Molecular Genetics, State University of New York, Albany, NY 12222

Communicated by Marlene Belfort, New York State Department of Health, Albany, NY, May 4, 2001 (received for review February 19, 2001)

A typical homing endonuclease initiates mobility of its group I intron by recognizing DNA both upstream and downstream of the intron insertion site of intronless alleles, preventing the endonuclease from binding and cleaving its own intron-containing allele. Here, we describe a GIY-YIG family homing endonuclease, I-*Bmo*I, that possesses an unusual recognition sequence, encompassing 1 base pair upstream but 38 base pairs downstream of the intron insertion site. I-*Bmo*I binds intron-containing and intronless substrates with equal affinity but can nevertheless discriminate between the two for cleavage. I-*Bmo*I is encoded by a group I intron that interrupts the thymidylate synthase (TS) gene (*thy*A) of *Bacillus mojavensis* s87-18. This intron resembles one inserted 21 nucleotides further downstream in a homologous TS gene (*td*) of *Escherichia coli* phage T4. I-*Tev*I, the T4 *td* intron-encoded GIY-YIG endonuclease, is very similar to I-*Bmo*I, but each endonuclease gene is inserted within a different position of its respective intron. Remarkably, I-*Tev*I and I-*Bmo*I bind a homologous stretch of TS-encoding DNA and cleave their intronless substrates in very similar positions. Our results suggest that each endonuclease has independently evolved the ability to distinguish intron-containing from intronless alleles while maintaining the same conserved recognition sequence centered on DNA-encoding active site residues of TS.

Group I introns are intervening sequences that are autocatalytically removed from the RNA transcript of the gene in which the intron is inserted (1). They are also highly efficient mobile genetic elements capable of unidirectional movement at the DNA level between intron-containing and intronless alleles. This process, termed homing (2), is initiated by site-specific DNA endonucleases (homing endonucleases) encoded within the introns themselves (3). Homing endonucleases possess lengthy recognition sequences (14–40 bp) that are generally centered on the intron insertion site (IS) of intronless alleles, and usually introduce a double-strand break within 2–5 nt upstream or downstream of the intron IS (4). By recognizing sequences both upstream (exon1, E1) and downstream (exon2, E2) of the intron IS, homing endonucleases are prevented from cleaving their own intron-containing alleles, as the recognition sequence is interrupted by the intron. Because the cleaved intronless allele is repaired by the double-strand break repair pathway, which uses the intron-containing allele as a template, the site of intron insertion and the endonuclease gene within the intron are present in the repaired allele (5–8).

I-*Tev*I is the well characterized homing endonuclease of the GIY-YIG family, which is encoded within a group I intron interrupting the *td* gene of *Escherichia coli* phage T4. I-*Tev*I is unusual with respect to other well characterized homing endonucleases because it cleaves intronless *td* DNA substrate 23 and 25 nt upstream of the intron IS (8, 9). Biochemical and structural studies have concluded that I-*Tev*I is composed of two functional domains (10). The N-terminal catalytic domain makes a few base-specific contacts near the cleavage sites of intronless DNA substrate (11, 12). The DNA-binding domain is located in the C-terminal half of the protein and makes predominantly minor groove contacts with its DNA substrate (11).

Another group I intron very similar to the phage T4 *td* intron was discovered in the TS gene (*thy*) of phage β22 that infects the Gram-positive bacterium *Bacillus subtilis* (13). Interestingly, the β22 *thy* intron shares 55% identity with the phage T4 *td* intron but is inserted 21 nt upstream from the T4 *td* intron IS in its *thy* gene. In contrast, the TS genes are very divergent, sharing only 30% identity. The I-*Tev*I coding sequence is inserted in structural element P6 of the phage T4 *td* intron; the β22 intron lacks a complete ORF and instead possesses a fragmentary ORF located in structural element P8. This fragmentary ORF shares amino acid similarity with helix-turn-helix (H-T-H) motifs of bacterial transcriptional regulators and with the C-terminal DNA-binding domain of I-*Tev*I.

We have recently discovered another TS (*thy*A) group I intron (C. Eifert and D.A.S., unpublished observations) in a *Bacillus* soil isolate, s87–18, which was later included in a newly described species, *Bacillus mojavensis* (14, 15). The *B. mojavensis thy*A intron is very similar to the β22 intron and is inserted in exactly the same position as its TS gene. However, the *B. mojavensis thy*A intron possesses, in structural element P8 (Fig. 1*A*), the coding sequence for a complete GIY-YIG endonuclease, I-*Bmo*I, that is very similar to I-*Tev*I and the fragmentary ORF of phage β22 (Fig. 1*B*).

Together, these data imply that the thymidylate synthase genes, introns, and intron-encoded ORFs of phage T4 and *B. mojavensis*, while homologs, have independent evolutionary histories, and that acquisition of either intron cannot be explained by an endonuclease-mediated homing event between the two TS genes (13). This suggested to us that I-*Bmo*I and I-*Tev*I might possess fundamentally different modes of recognition of intron-containing and intronless substrates, as each endonuclease independently adapted to the change of intron position.

Materials and Methods

Strains. *E. coli* JM109 or XL-1B were used for plasmid manipulations and grown in liquid or solid LB media supplemented with appropriate antibiotics. *E. coli* ER2566 (New England Biolabs) or BL21(DE3) pLysS (16) were used for protein overexpression.

Plasmid Construction. PCR fragments corresponding to full-length I-*Bmo*I or the 129C domain were cloned into the IMPACT expression vector pTYB1 (17) to yield pDE12 and pBmo129C.

Abbreviations: IS, insertion site; TS, thymidylate synthase; wt, wild type.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. AF321518).

*Present address: Molecular Genetics Program, Wadsworth Center, New York State Department of Health and School of Public Health, State University of New York, Albany, NY 12201-2002.

†To whom reprint requests should be addressed. E-mail: shub@csc.albany.edu or edgell@wadsworth.org.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

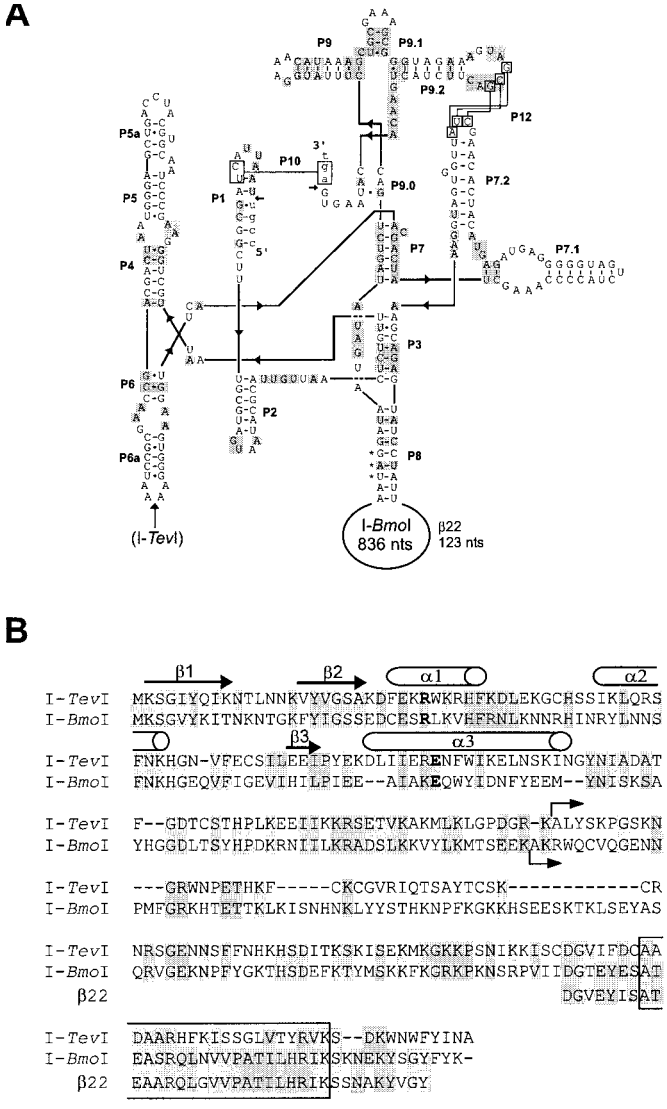

Fig. 1. (*A*) Secondary structure of the *B. mojavensis thy*A intron (29). Identical nt between the *B. mojavensis thy*A and phage T4 *td* introns are shaded. Positions of the I-*Bmo*I, I-*Tev*I, and phage β22 coding sequences are indicated. The I-*Bmo*I stop codon in P8 is indicated by asterisks. (*B*) Amino acid alignment of I-*Tev*I, I-*Bmo*I, and the fragmentary ORF from phage β22. Identical or conserved amino acids are shaded. The H-T-H motif is indicated by a box (13). The catalytic Arg-27 and Glu-75 (30) are in bold type. Alpha helices (cylinders) and beta-sheets (black arrows), representing secondary structure elements of I-*Tev*I, are drawn above amino acid residues to which they correspond (30). Amino acid positions corresponding to cloned C-terminal DNA-binding domains of I-*Tev*I and I-*Bmo*I are indicated by right-facing arrows.

The R27A mutation was introduced into pDE12 by inverse PCR (18) to give pDE13. All inserts were sequenced on both strands to confirm the correct sequence.

To map I-*Bmo*I cleavage sites on intronless substrate, p87–18ΔI was created from pCE14, which carries a 500-bp insert of the *B. mojavensis* s87–18 *thy*A gene with an exact deletion of the intron. Primers 87–18E1.*Xba*I and 87–18E2.*Eco*RI were used to generate a 198-bp PCR product, which was ligated into *Xba*I/*Eco*RI cut pBS.

Protein Purification. I-*Bmo*I wild type (wt), R27A, and 129C were overexpressed by using the IMPACT system (17), with the following changes. Column buffer consisted of 20 mM Tris, pH 8.0/500 mM NaCl/5 mM EDTA/0.1% Triton X-100. A high salt wash (50 mM Tris, pH 8.0/1 M NaCl/1 mM EDTA) was used to remove nonspecifically bound protein. Following elution, protein-containing fractions were pooled and dialyzed twice against 1 liter of 20 mM Tris, pH 8.0/500 mM NaCl/1 mM EDTA at 4°C. For further purification, eluted fractions were loaded directly onto a Hi-Trap Heparin column (Amersham Pharmacia) equilibrated with 20 mM Tris, pH 8.0/500 mM NaCl. Protein was eluted with a linear NaCl gradient of 0.5 M to 1.2 M. I-*Bmo*I eluted at 0.8–1.0 M NaCl. Fractions containing protein were flash frozen and stored at −80°C.

Mapping of I-*Bmo*I Cleavage Sites. Twenty picomoles of reverse and −20 primers, end-labeled with 16.8 pmol (50 μCi) of γ[^{32}P]ATP by T4 polynucleotide kinase, were used in a PCR reaction with p87–18ΔI to generate strand-specific substrates. Cleavage reactions were performed in 50-μl volumes of 50 mM Tris, pH 8.0/10 mM $MgCl_2$/100 mM NaCl that contained 10^5 cpm of end-labeled substrate, and 100 pM I-*Bmo*I at 37°C for 5 min. Cleavage reactions were run alongside sequencing ladders of the same strand on 6% denaturing polyacrylamide gels.

DNase I Footprinting. Footprinting reactions were performed with the catalytic mutant I-*Bmo*I R27A or I-*Bmo*I 129C in 30-μl volumes with 0.05 or 0.1 pmol of annealed oligonucleotide substrate (FP1/FP2 or INT-E2Top/INT-E2Bot). Reactions were incubated at 24°C for 10 min; 1 unit of DNase I was added and then stopped after 30 s by the addition of 3-μl stop buffer (660 mM Tris, pH 9.5/66 mM EDTA/3.3% SDS) and freezing on dry ice. Aliquots were run alongside the appropriate Maxam–Gilbert A + G sequencing ladders (19) on either 8% or 10% denaturing polyacrylamide gels (19:1 acrylamide: bisacrylamide).

Gel-Shift Assays. Substrates for gel shifts were the same oligonucleotides used in footprinting reactions. Binding reactions were performed at 24°C in 30 μl of 50 mM Tris, pH 8.0/50 mM NaCl with 0.01–0.05 pM substrate. Loading dye (5 μl of 50 mM Tris, pH 8.0/50 mM NaCl/50% glycerol) was added before loading on a 8% native gel (29:1 acrylamide:bisacrylamide), prerun in 0.5 × TBE at 4°C. For competition gel shifts, unlabeled competitor DNA was added to binding reactions before addition of I-*Bmo*I.

Cleavage Assays. Substrates for cleavage assays were singly end-labeled PCR products or annealed complementary oligonucleotides corresponding to E1-E2 or INT-E2 junctions. Reactions were performed in 30-μl volumes consisting of 50 mM Tris, pH 8.0/100 mM NaCl/10 mM $MgCl_2$ and incubated at 37°C for 10 min. Reactions were run alongside appropriate Maxam–Gilbert A + G sequencing reactions on 10% denaturing polyacrylamide gels for oligonucleotide substrates or 6% denaturing gels for PCR-generated substrates.

Oligonucleotides. FP1, CTAGATCCACCTTGAGGTAAGAGCCCGTAGTAATGACATGGCCTTGGGAAATCCCTTCAATGTATTCCAGTACAATGT; FP2, CTAGATTGTACTGGAATACATTGAAGGGATTTCCCAAGGCCATGTCATTACTACGGGCTCTTACCTCAAGGTGGAGTT; INT-E2Top, CTAGAGCGACTTCTACTGAACATAAGTGAGTAATGACATGGCCTTGGGAAATCCCTTCAATGTATTCCAGTACAATGT; INT-E2Bot, CTAGAACATTGTACTGGAATACATTGAAGGGATTTCCCAAGGCCATGTCATTACTCACTTATGTTCAGTAGAAGTCC.

Results

I-*Bmo*I Cleaves Intronless DNA Substrate Close to the Intron Insertion Site. To facilitate study of I-*Bmo*I, we overexpressed the protein by using the IMPACT system (17). Copurifying with full-length

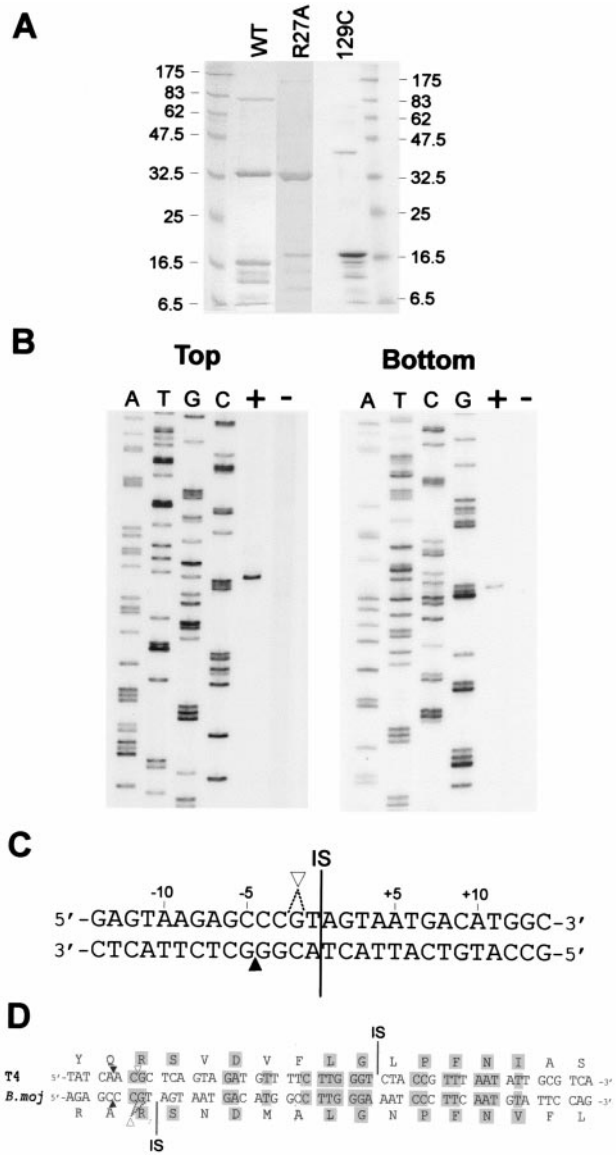

Fig. 2. I-*Bmo*I cleaves intronless substrate close to the intron insertion site. (*A*) SDS/PAGE gel of IMPACT column fractions enriched for I-*Bmo*I wt (31.2 kDa), R27A (31.1 kDa), or 129C (16.1 kDa) proteins. (*B*) Mapping assay to determine I-*Bmo*I cleavage sites on intronless substrate. Intronless substrate, singly end-labeled on top or bottom strands, was incubated with (+) or without (−) I-*Bmo*I and run beside sequencing ladders of the same strand (denoted by the ddNTP used in each reaction). (*C*) DNA sequence surrounding the intron IS of the *B. mojavensis thy*A gene showing positions of I-*Bmo*I cleavage sites. ▲, bottom strand; ▽, top strand. Dashed lines indicate possible positions of the single top strand cleavage site. (*D*) I-*Bmo*I and I-*Tev*I cleave their respective intronless substrates in the same positions. Shown is a protein-based DNA alignment of the coding strand of the *E. coli* phage T4 and *B. mojavensis* s87–18 thymidylate synthase genes. Thick lines indicate intron IS. Identical nt and amino acids are shaded.

I-*Bmo*I were several smaller proteins (Fig. 2*A*), which we believe to be proteolytic fragments of I-*Bmo*I (see below).

To map I-*Bmo*I cleavage sites, we digested intronless substrate, end-labeled on the top or bottom strands, with I-*Bmo*I. Surprisingly, I-*Bmo*I cleaves intronless substrate very close to the intron IS, 4 nt upstream on the bottom strand, and at a single site 1 or 2 nt upstream on the top strand (Fig. 2 *B* and *C*). Despite the fact that their introns are inserted 21 bp apart, I-*Bmo*I and I-*Tev*I cleave their respective intronless substrates in similar positions (Fig. 2*D*).

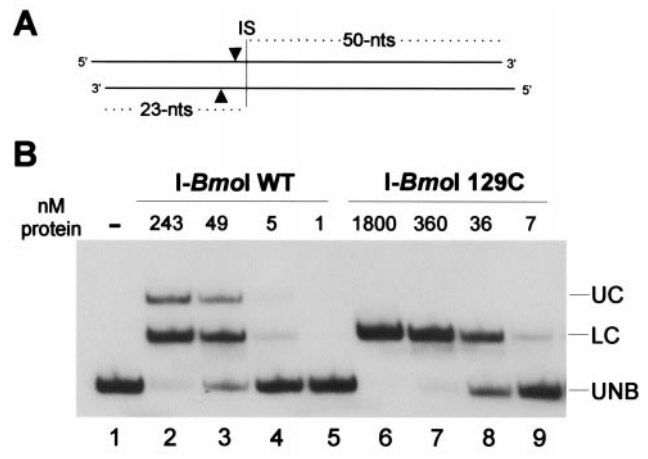

Fig. 3. A C-terminal fragment of I-*Bmo*I retains DNA-binding activity. (*A*) Schematic of intronless (E1–E2) oligonucleotide substrate encompassing 23 bp upstream (E1) and 50 bp downstream (E2) of the intron IS, plus 5 nt corresponding to *Xba*I overhangs. Cleavage sites (black arrows) are indicated for top or bottom strands. (*B*) Two distinct complexes are observed upon I-*Bmo*I binding of intronless substrate. Top strand labeled substrate was incubated with increasing concentrations of I-*Bmo*I wt or 129C proteins and resolved on a nondenaturing gel. UC, upper complex; LC, lower complex; UNB, unbound substrate.

A C-Terminal Fragment of I-*Bmo*I Retains DNA-Binding Activity. To characterize further interactions of I-*Bmo*I with its intronless DNA substrate, we designed complementary 78-mer oligonucleotides (FP1, FP2) encompassing positions −23 to +50 relative to the intron IS of intronless *B. mojavensis thy*A (Fig. 3*A*). When annealed complementary oligonucleotides, end-labeled on top or bottom (not shown) strands, were incubated with increasing concentrations of I-*Bmo*I in the absence of metal ions and resolved on a nondenaturing gel, two distinct complexes were observed: an upper complex and a lower complex (Fig. 3*B*, lanes 2–5). Multiple protein–DNA complexes were also observed during gel-shift studies of I-*Tev*I and its intronless substrate (10, 11). An analogous lower complex was shown to result from interaction of a proteolyzed C-terminal fragment of I-*Tev*I that retained DNA-binding activity, but not catalytic activity, with intronless substrate (11).

Based on an amino acid alignment of I-*Bmo*I and I-*Tev*I, we cloned and overexpressed a C-terminal fragment of I-*Bmo*I (I-*Bmo*I 129C) starting at residue 129 with a predicted molecular mass of 16 kDa (Figs. 1*B* and 2*A*). This amino acid position is similar to that used for cloning and overexpression of the 130C domain of I-*TevI*, which corresponds to a trypsin cleavage site in full-length I-*Tev*I (10). Purified I-*Bmo*I 129C migrates to a position similar to putative proteolyzed fragments of full-length protein (Fig. 2*A*). When we incubated intronless substrate with I-*Bmo*I 129C, we observed only a single complex (Fig. 3, lanes 6–9) similar in mobility to the lower complex observed with full-length preparations of I-*Bmo*I (Fig. 3, lanes 2–5). Thus, it is likely that the lower complex we observe during gel-shift analyses with full-length I-*Bmo*I results from proteolyzed C-terminal fragments that copurify and retain DNA-binding activity.

I-*Bmo*I Has an Unusually Asymmetric Footprint with Respect to the Intron Insertion Site. I-*Tev*I cleaves its intronless substrate 23 and 25 nt upstream from the intron IS (8, 9), yet the DNase I footprint of the protein is still centered on the intron IS (10). I-*Bmo*I cleaves intronless substrate in similar positions as I-*Tev*I, even though the *B. mojavensis thy*A intron IS is 21 nt upstream from the phage T4 *td* intron IS (Fig. 2*D*). Given these contrasting positions of cleavage sites relative to the intron IS, we were

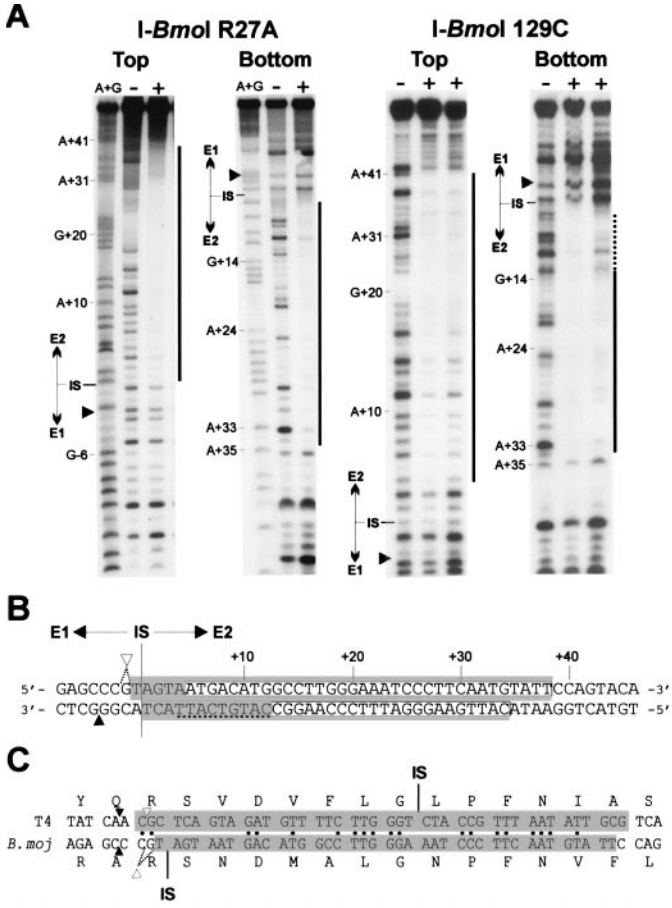

Fig. 4. I-*Bmo*I has an unusually asymmetric footprint with respect to the intron IS. (*A*) Shown is DNase I protection assay with no protein (−), I-*Bmo*I R27A (18 μM), or I-*Bmo*I 129C (33 and 11 μM) on top or bottom strands of intronless substrate. Black bars indicate region protected by each protein. Weak protection was observed on the bottom strand with I-*Bmo*I 129C, indicated by a dashed line. Black triangles indicate position of I-*Bmo*I cleavage sites. (*B*) Extent of I-*Bmo*I R27A footprint is indicated by a shaded box and by a white box for I-*Bmo*I 129C. Cleavage sites are labeled as in Fig. 2. (*C*) I-*Bmo*I and I-*Tev*I protect the same homologous stretch of DNA of their respective intronless substrates. Shaded regions indicate extent of DNase I protection on the top strand of intronless substrates for I-*Bmo*I R27A and I-*Tev*I R27A (10). Identical nt between the two TS genes are indicated by black dots. Intron IS and cleavage sites for I-*Bmo*I and I-*Tev*I are indicated.

curious to determine whether the DNase I footprint of I-*Bmo*I was centered on its intron IS, as would be expected for a typical homing endonuclease.

To this end, we constructed and overexpressed a catalytic mutant (R27A) of I-*Bmo*I (Figs. 1*B* and 2*A*) for use in DNase I footprinting experiments with complementary oligonucleotides (FP1, FP2) corresponding to the E1–E2 junction as substrate (Fig. 3*A*). We found that a region of 39 bp was protected from DNase I digestion by the protein (Fig. 4*A*). Surprisingly, the protected region was not centered on the intron IS, as 38 bp of E2 and 1 bp of E1 were protected from DNase I digestion by the protein (Fig. 4*B*). We also performed DNase I protection assays with I-*Bmo*I 129C and found that the protected region included only the exon 2 sequence (Fig. 4*B*).

The DNase I footprints of full-length I-*Bmo*I R27A and I-*Tev*I R27A are very similar; both proteins protect a longer stretch of DNA on the top versus the bottom strand, and the protection extends up to, but does not include, the cleavage sites of both proteins (10, 11). Similarly, the DNase I footprints of the I-*Tev*I and I-*Bmo*I DNA-binding domains both encompass a subset of the full-length footprint (Fig. 4*B*, ref. 10). Remarkably, I-*Bmo*I and I-*Tev*I protect the same homologous stretch of DNA on their respective intronless substrates, which are 43% identical over the protected region (Fig. 4*C*).

I-*Bmo*I Binds with the Same Relative Affinity to Intron-Containing and Intronless Substrates. The asymmetric footprint of I-*Bmo*I on its intronless substrate led us to consider the possibility that I-*Bmo*I might also bind intron-containing substrate. We designed complementary oligonucleotides (INT-E2Top and INT-E2Bot) of the same length as those used for footprinting and gel-shift assays of the E1–E2 junction, but with 23 bp of exon1 sequence replaced by intron sequence (Fig. 5*A*). We performed a series of binding reactions in which either I-*Bmo*I wt (Fig. 5*B*) or I-*Bmo*I R27A (not shown) was added in increasing equivalent molar excess to either E1–E2 or INT–E2 oligonucleotide duplex substrates labeled on the top strand. Unexpectedly, we observed that equivalent amounts of substrate were shifted at approximately the same concentrations of I-*Bmo*I, irrespective of whether the substrate corresponded to the E1–E2 or INT–E2 junction.

We also performed a competition gel-shift assay in which we added increasing molar excess of cold E1–E2 or INT–E2 competitor substrate to binding reactions containing labeled E1–E2 substrate. For a typical homing endonuclease, addition of E1–E2 substrate should act as a specific competitor, whereas addition of INT–E2 substrate should have a lesser affect on binding of E1–E2 DNA. We found that addition of increasing amounts of either unlabeled E1–E2 or INT–E2 competitor reduced the amount of E1–E2 substrate bound by I-*Bmo*I (Fig. 5*C*). We also performed the complementary experiment, adding increasing amounts of unlabeled E1–E2 or INT–E2 competitor to binding reactions containing labeled INT–E2 substrate, and found that either competitor could effectively decrease the amount of INT–E2 substrate bound by I-*Bmo*I (Fig. 5*C*).

Using the same ratio of I-*Bmo*I R27A to substrate as in DNase I protection assays of the E1–E2 substrate, we performed DNase I protection assays with the INT–E2 substrate (Fig. 5*D*) and found the same protection pattern on INT–E2 substrate as on E1–E2 substrate (compare Fig. 4*B* with Fig. 5*E*). All of these results are consistent with I-*Bmo*I being unable to distinguish intron-containing from intronless substrate for binding.

I-*Bmo*I Can Effectively Distinguish Intronless from Intron-Containing Substrates for Cleavage. The finding that I-*Bmo*I cannot distinguish intron-containing from intronless substrates for binding raised the obvious question of whether I-*Bmo*I can distinguish between the two for cleavage, as would be expected of a typical homing endonuclease. We therefore incubated the same molar concentration of intronless or intron-containing oligonucleotide substrates, labeled on top or bottom strands, with increasing concentrations of I-*Bmo*I and resolved the reaction products on denaturing gels. Under conditions where intronless substrate was extensively cleaved on the top strand, intron-containing substrate appeared not to be cleaved (Fig. 6*A*). Identical results were obtained for intronless and intron-containing substrates labeled on the bottom strand (not shown).

We also tested the cleavage preference of I-*Bmo*I on longer intron-containing and intronless substrates, generated by the PCR and differentially end-labeled on top or bottom strands (Fig. 6*B*). When incubated with the same increasing concentrations of I-*Bmo*I, intron-containing substrate was not cleaved, whereas intronless substrate was extensively cleaved (Fig. 6*B*, compare lanes 2–4 with lanes 6–8). As with cleavage of intron-containing oligonucleotide substrates (Fig. 6*A*), longer exposures revealed a weak cleavage product for intron-containing substrate (Fig. 6*B*, arrowheads). When both substrates were combined in a mixed cleavage assay, we observed a reduction in

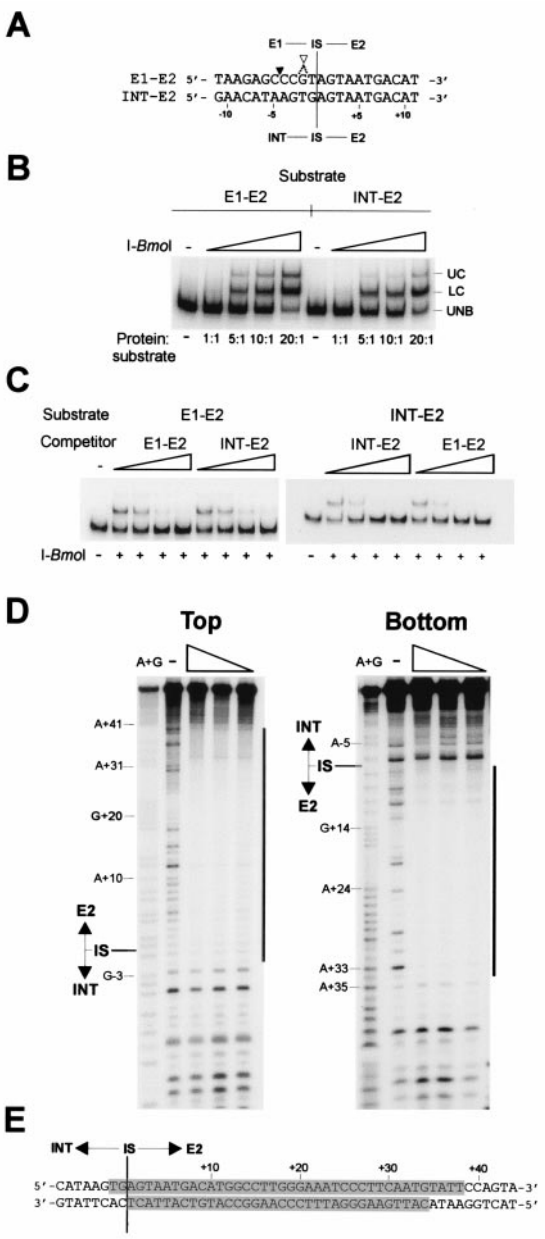

Fig. 5. I-*Bmo*I binds with the same relative affinity to intron-containing or intronless substrates. (*A*) Alignment of coding strand sequence from positions −11 to +11 surrounding the intron IS of intron-containing (INT–E2) or intronless (E1–E2) substrates. I-*Bmo*I cleavage sites are labeled as in Fig. 2. Intron-containing and intronless substrates are the same length (78 nt). (*B*) Gel-shift assay with increasing concentrations of I-*Bmo*I and equal concentrations of E1–E2 or INT–E2 oligonucleotide substrate end-labeled on the top strand. (*C*) Intron-containing and intronless substrate can effectively compete for binding of intronless substrate by I-*Bmo*I. Equal concentrations of intronless or intron-containing substrate end-labeled on the top strand were incubated with I-*Bmo*I wt protein and increasing concentrations of unlabeled intron-containing or intronless substrate. (*D*) I-*Bmo*I R27A protects the same region of intron-containing as intronless substrate. Shown is a DNase I protection assay with decreasing concentrations of I-*Bmo*I R27A on top or bottom strands of INT–E2 substrate, labeled as in *A*. (*E*) Regions protected (shaded) by I-*Bmo*I R27A from DNase I digestion on an INT–E2 substrate.

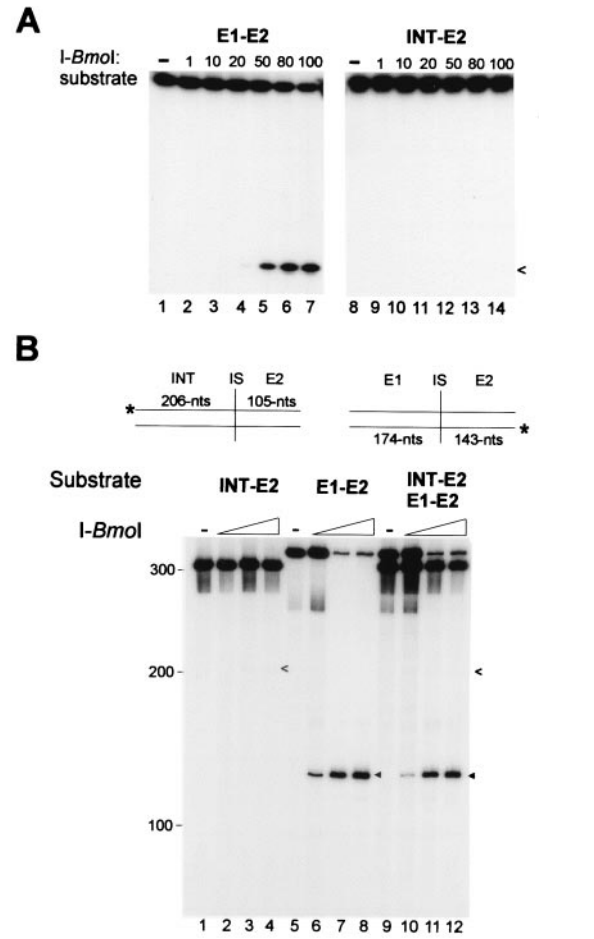

Fig. 6. I-*Bmo*I cleaves intronless, but not intron-containing substrate. (*A*) Cleavage assays with intronless (E1–E2) or intron-containing (INT–E2) oligonucleotide substrates labeled on the top strand with increasing ratio of I-*Bmo*I to substrate. On longer exposure, faint cleavage products could be visualized for INT–E2 top and bottom (not shown) substrates, indicated by open arrowheads. (*B*) Mixed cleavage assays with PCR-generated intron-containing (INT–E2) and intronless (E1–E2) substrates. Black triangles indicate E1–E2 cleavage products, and arrowheads indicate INT–E2 cleavage products. Positions of molecular weight markers are indicated.

the amount of intronless cleavage product (Fig. 6*B*, compare lanes 6–8 with 10–12), as would be expected if intron-containing and intronless substrate both compete for binding of I-*Bmo*I (Fig. 5). These results show that I-*Bmo*I can effectively distinguish between intron-containing and intronless substrates for cleavage specificity.

Discussion

That group I introns and their associated endonucleases can have independent evolutionary histories has been well documented in a number of systems (20–22). However, the situation we describe here is unique in that it involves two very similar group I introns, inserted 21 bp apart in homologous TS genes. Both introns possess homologous GIY-YIG endonucleases, but they are inserted in different positions of each intron. Furthermore, although both the T4 *td* and *B. mojavensis thy*A introns possess functional endonucleases, it is unlikely that either endonuclease mediated movement of the intron by 21 bp in a typical homing event between the two TS genes. Regardless of where the intronless allele was cleaved, double-strand break repair would faithfully use the intron-containing allele as the template, preserving the site of intron insertion in the recipient. Thus, the change in position of the T4 *td* and *B. mojavensis thy*A introns likely arose by an endonuclease-independent mechanism, perhaps mediated by a reverse-splicing event (23).

Irrespective of whether movement of the endonucleases within their introns occurred before or after movement of the TS introns by 21 bp, I-*Bmo*I and I-*Tev*I adapted to a different intron IS by independently evolving the ability to distinguish intron-containing from intronless substrates. By comparing I-*Tev*I substrate specificities with those of I-*Bmo*I presented here, we can begin to address the biochemical basis of substrate recognition by each endonuclease in response to a change of intron position.

I-*Tev*I is a two-domain protein, consisting of an N-terminal catalytic domain and a C-terminal DNA-binding domain (10). The DNase I footprint of the DNA-binding domain is centered over the intron IS of intronless *td* substrate (10). In contrast, the N-terminal catalytic domain is positioned for cleavage 23–25 nt further upstream, makes few base-specific contacts, and is remarkably tolerant of nucleotide substitutions, insertions, and deletions near the cleavage sites (11, 12). Thus, I-*Tev*I discriminates between intron-containing and intronless substrates by sequence-specific contacts made with its DNA-binding domain.

We suggest that I-*Bmo*I does not discriminate between intronless and intron-containing substrates at the DNA recognition step, but does so at the catalytic step for the following reasons. First, the DNase I footprint of I-*Bmo*I 129C encompasses only exon2 sequence, which does not include the intron IS (Fig. 4). However, the DNase I footprint of full-length I-*Bmo*I R27A does extend over the intron IS into exon1 (Fig. 4). Second, as the cleavage sites of I-*Bmo*I in exon1 map very close to the intron IS, the N-terminal catalytic domain of I-*Bmo*I must contact DNA in this region of intronless substrate, facilitating cleavage. Third, I-*Bmo*I binds with approximately equal affinity to intron-containing or intronless substrates (Fig. 5) but efficiently cleaves only intronless substrate (Fig. 6). It is possible that the N-terminal catalytic domain of I-*Bmo*I requires specific DNA sequence for cleavage, which is present only in exon1 of intronless substrate.

The observation that the footprint of I-*Bmo*I on intronless substrate is not centered on the intron IS (Fig. 4) is unexpected for a homing endonuclease, but not without precedent. I-*Sce*I, a LAGLIDAG family endonuclease encoded within the mitochondrial rRNA group I intron of *Saccharomyces cerevisiae*, also possesses an asymmetric recognition sequence and binds intron-containing substrate (24). I-*Hmu*I and I-*Hmu*II, H-N-H family endonucleases found within group I introns interrupting the DNA polymerase genes of *Bacillus* phages SP01 and SP82 (25), each cleaves intron-containing DNA (26).

A surprising finding of our characterization of I-*Bmo*I R27A was that the footprint encompassed the homologous region of intronless DNA substrate as is protected by I-*Tev*I R27A on its intronless substrate (Fig. 4*C*). This region of TS coding sequence includes the highly conserved Arg-218, Ser-219, Asp-221, and Asn-229 residues, all of which are involved in binding dUMP in the active site of the enzyme (Fig. 2*D*; ref. 27). For homing endonucleases, one potential strategy to maximize spread to intronless alleles would be to use as recognition sites nucleotide sequences corresponding to important functional domains of proteins, as those sequences are likely to be conserved between related organisms (28).

Movement of the intron IS within a conserved recognition sequence would be tolerated so long as the intron-encoded GIY-YIG endonucleases could adapt to the intron position and discriminate between intron-containing and intronless alleles to promote homing. In this respect, it is noteworthy that although the amino acid similarity between I-*Bmo*I and I-*Tev*I is greatest in the N-terminal catalytic domain of each protein, both proteins possess an H-T-H motif in their C-terminal DNA-binding domain (Fig. 1*B*). By using the H-T-H motif as the primary DNA recognition determinant, each protein could independently evolve additional DNA binding and cleavage specificity within other domains to individually adapt to a particular intron IS without compromising the ability to bind a conserved recognition sequence.

We thank Archana Belle, Richard Bonocora, Marlene Belfort, Vicky Derbyshire, Cheryl Eifert, Markus Landthaler, and Niles Lehman for helpful discussions and Cheryl Eifert for providing primers and genomic DNA. This work was supported by a Postdoctoral Fellowship from the Canadian Institute for Health Research (to D.R.E.) and by National Institutes of Health Grants GM37746 and GM44844.

1. Cech, T. R. (1990) *Annu. Rev. Biochem.* **59,** 543–568.
2. Dujon, B., Belfort, M., Butow, R. A., Jacq, C., Lemieux, C., Perlman, P. S. & Vogt, V. M. (1989) *Gene* **82,** 115–118.
3. Lambowitz, A. M. & Belfort, M. (1993) *Annu. Rev. Biochem.* **62,** 587–622.
4. Jurica, M. S. & Stoddard, B. L. (1999) *Cell. Mol. Life Sci.* **55,** 1304–1326.
5. Jacquier, A. & Dujon, B. (1985) *Cell* **41,** 383–394.
6. Zinn, A. R. & Butow, R. A. (1985) *Cell* **40,** 887–895.
7. Quirk, S. M., Bell-Pedersen, D. & Belfort, M. (1989) *Cell* **56,** 455–465.
8. Bell-Pedersen, D., Quirk, S., Clyman, J. & Belfort, M. (1990) *Nucleic Acids Res.* **18,** 3763–3770.
9. Chu, F. K., Maley, F., Wang, A. M., Pedersen-Lane, J. & Maley, G. (1991) *Nucleic Acids Res.* **19,** 6863–6869.
10. Derbyshire, V., Kowalski, J. C., Dansereau, J. T., Hauer, C. R. & Belfort, M. (1997) *J. Mol. Biol.* **265,** 494–506.
11. Bryk, M., Quirk, S. M., Mueller, J. E., Loizos, N., Lawrence, C. & Belfort, M. (1993) *EMBO J.* **12,** 2141–2149.
12. Bryk, M., Belisle, M., Mueller, J. E. & Belfort, M. (1995) *J. Mol. Biol.* **247,** 197–210.
13. Bechhofer, D. H., Hue, K. K. & Shub, D. A. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 11669–11673.
14. Duncan, K. E., Istock, C. A., Graham, J. B. & Ferguson, N. (1994) *Evolution (Lawrence, Kans.)* **43,** 1585–1609.
15. Roberts, M. S. & Cohan, F. M. (1995) *Evolution (Lawrence, Kans.)* **49,** 1081–1094.
16. Studier, F. W. (1991) *J. Mol. Biol.* **219,** 37–44.
17. Chong, S., Mersha, F. B., Comb, D. G., Scott, M. E., Landry, D., Vence, L. M., Perler, F. B., Benner, J., Kucera, R. B., Hirvonen, C. A., *et al.* (1997) *Gene* **192,** 277–281.
18. Ochman, H., Gerber, A. S. & Hartl, D. L. (1988) *Genetics* **120,** 621–623.
19. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. S. & Struhl, K. (1990) *Current Protocols in Molecular Biology* (Wiley, New York).
20. Mota, E. M. & Collins, R. A. (1989) *Nature (London)* **332,** 654–656.
21. Loizos, N., Tillier, E. R. M. & Belfort, M. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 11983–11987.
22. Saguez, C., Lecellier, G. & Koll, F. (2000) *Nucleic Acids Res.* **28,** 1299–1306.
23. Woodson, S. A. & Cech, T. R. (1989) *Cell* **57,** 335–345.
24. Perrin, A., Buckle, M. & Dujon, B. (1993) *EMBO J.* **12,** 2939–2947.
25. Goodrich-Blair, H. & Shub, D. A. (1994) *Nucleic Acids Res.* **22,** 3715–3721.
26. Goodrich-Blair, H. & Shub, D. A. (1996) *Cell* **84,** 211–221.
27. Carreras, C. W. & Santi, D. V. (1995) *Annu. Rev. Biochem.* **64,** 721–762.
28. Edgell, D. R., Belfort, M. & Shub, D. A. (2000) *J. Bacteriol.* **182,** 5281–5289.
29. Cech, T. R., Damberger, S. H. & Gutell, R. R. (1994) *Nat. Struct. Biol.* **1,** 265–266.
30. Kowalski, J. C., Belfort, M., Stapleton, M. A., Holpert, M., Dansereau, J. T., Pietrokovski, S., Baxter, S. M. & Derbyshire, V. (1999) *Nucleic Acids Res.* **27,** 2115–2125.