Published online 4 April 2013 *Nucleic Acids Research, 2013, Vol. 41, No. 10* ***5413–5427***
doi:10.1093/nar/gkt186

The monomeric GIY-YIG homing endonuclease I-BmoI uses a molecular anchor and a flexible tether to sequentially nick DNA

Benjamin P. Kleinstiver, Jason M. Wolfs and David R. Edgell*

Department of Biochemistry, Schulich School of Medicine and Dentistry, Western University, London, Ontario N6A 5C1, Canada

Received January 18, 2013; Revised February 25, 2013; Accepted February 27, 2013

ABSTRACT

The GIY-YIG nuclease domain is found within protein scaffolds that participate in diverse cellular pathways and contains a single active site that hydrolyzes DNA by a one-metal ion mechanism. GIY-YIG homing endonucleases (GIY-HEs) are two-domain proteins with N-terminal GIY-YIG nuclease domains connected to C-terminal DNA-binding and they are thought to function as monomers. Using I-BmoI as a model GIY-HE, we test mechanisms by which the single active site is used to generate a double-strand break. We show that I-BmoI is partially disordered in the absence of substrate, and that the GIY-YIG domain alone has weak affinity for DNA. Significantly, we show that I-BmoI functions as a monomer at all steps of the reaction pathway and does not transiently dimerize or use sequential transesterification reactions to cleave substrate. Our results are consistent with the I-BmoI DNA-binding domain acting as a molecular anchor to tether the GIY-YIG domain to substrate, permitting rotation of the GIY-YIG domain to sequentially nick each DNA strand. These data highlight the mechanistic differences between monomeric GIY-HEs and dimeric or tetrameric GIY-YIG restriction enzymes, and they have implications for the use of the GIY-YIG domain in genome-editing applications.

INTRODUCTION

Site-specific DNA nucleases are central to many cellular processes, including DNA recombination and repair, restriction defense systems and the mobility of selfish genetic elements (1,2). A number of nuclease active sites have evolved that are characterized by conserved amino acid motifs and operate via distinct catalytic mechanisms to generate nicks or double-strand breaks (DSBs) in DNA (1–3). There is often little correlation between the identity of the nuclease active site motif and cellular function, as conserved catalytic cores from individual nuclease families are found within diverse protein scaffolds (1,4). As such, a fundamental understanding of the mechanisms by which each nuclease family generates a DSB is required, as several families have been exploited for genome engineering purposes (5–9).

One such nuclease family contains the GIY-YIG catalytic motif, consisting of a compact domain of ~100 amino acids that is characterized by at least two α-helices that surround a three-stranded β-sheet (10–14). The catalytic core includes a number of conserved amino acids that are essential for DNA-hydrolysis via a one-metal ion mechanism (14). The minimal GIY-YIG nuclease domain is often found associated with protein scaffolds that function to anchor the GIY-YIG domain in close proximity to DNA, implying that the GIY-YIG domain itself has weak DNA-binding activity (15–20). Included in this category is the nucleotide excision repair protein UvrC that has acquired distinct binding domains and uses a single GIY-YIG domain to nick 3′ to DNA lesions (11). Alternatively, the restriction enzymes Eco29kI and Hpy188I have acquired supplementary units of structure that are intertwined within the conserved GIY-YIG domain to promote oligomerization and assembly on DNA, with each GIY-YIG monomer nicking one strand (13,14,19,21). The GIY-YIG nuclease domain is also found within homing endonucleases (GIY-HEs) as an N-terminal fusion to sequence-tolerant DNA-binding domains that target extended asymmetric binding sites (22,23).

GIY-HEs function as mobile genetic elements by introducing DSBs at a defined site within naïve genomes to promote mobility of their own genes (2,24,25). I-TevI and I-BmoI, encoded within introns interrupting the *td* and *thyA* thymidylate synthase (TS) genes of bacteriophage T4 and *Bacillus mojavensis*, remain the best-characterized GIY-HEs to date; however, sequencing

*To whom correspondence should be addressed. Tel: +1 519 661 3133; Fax: +1 519 661 3175; Email: dedgell@uwo.ca

efforts have identified additional putative GIY-HEs (26–28). I-BmoI and I-TevI bind a homologous target site in their respective TS genes and consist of N-terminal GIY-YIG domains connected by linkers to C-terminal domains that contact 30–35 bp of substrate (Figure 1A) (15,23,29). Studies of I-TevI indicate that the inter-domain linker is flexible (29,30), supported by its ability to position the catalytic domain on substrate. The modularity of GIY-HEs is also evidenced in their cognate target sites, as the asymmetric sequences contain distinct motifs for DNA binding and cleavage (15). The I-BmoI and I-TevI target sites share 48% identity and previous *in vitro* substrate selections identified a conserved G–C base pair that flanks the I-BmoI and I-TevI cleavage sites as critical for cleavage (Figure 1A) (31,32). Contacts to the G–C base pair are essential for both GIY-HEs to perform the sequential and independent nicking reactions to generate a DSB with a two-nucleotide 3′-overhang (23,33,34). Previous studies with I-TevI showed that the two-domain enzyme bound DNA as a monomer and induced significant substrate bending near the cleavage site (35), whereas chimeric enzymes with the I-TevI nuclease domain fused to the *ryA* zinc-finger generated a DSB in a non-cooperative manner (9). Kinetic analyses of the two-step nicking reaction using I-BmoI showed that the first-strand nicking reaction proceeds ~2-fold faster than the second nicking reaction (28). Moreover, divalent metal ion and DNA sequence at the cleavage site influence I-BmoI activity, likely by modulating local DNA structure and promoting sequential protein interactions with bases in the cleavage site region (36). However, the central question of how the single active site of GIY-HEs is used to successively nick each DNA strand has not been rigorously addressed. Based largely on studies with I-TevI, a conformational change model was proposed, whereby the GIY-YIG nuclease domain and substrate undergo substantial re-arrangements between nicking reactions to reposition the active site (Figure 1B) (35). However, the extreme cytotoxicity of I-TevI precluded detailed kinetic studies that could rule out alternative mechanisms of DSB formation by monomeric nucleases, including transient dimerization between catalytic domains, synapsis of two enzymes bound to separate target sites or cleavage through sequential transesterification reactions (Figure 1B) (37–39).

Here, we capitalize on the ability to purify wild-type I-BmoI to explicitly test models of DSB formation. Significantly, we show that I-BmoI functions as a monomer at all stages of the reaction pathway and does not use a hairpin mechanism for DNA cleavage. Our data are consistent with a model where the monomeric GIY-YIG domain rotates between nicking sites to generate a DSB, with the I-BmoI DNA-binding domain functioning to anchor the linker and nuclease domain to substrate. We envision that a better understanding of the catalytic mechanism will provide insight into the ability of GIY-HEs to invade a diverse range of biological alleles, as well as enhance ongoing efforts to use the GIY-YIG nuclease domain as a genome-editing tool.

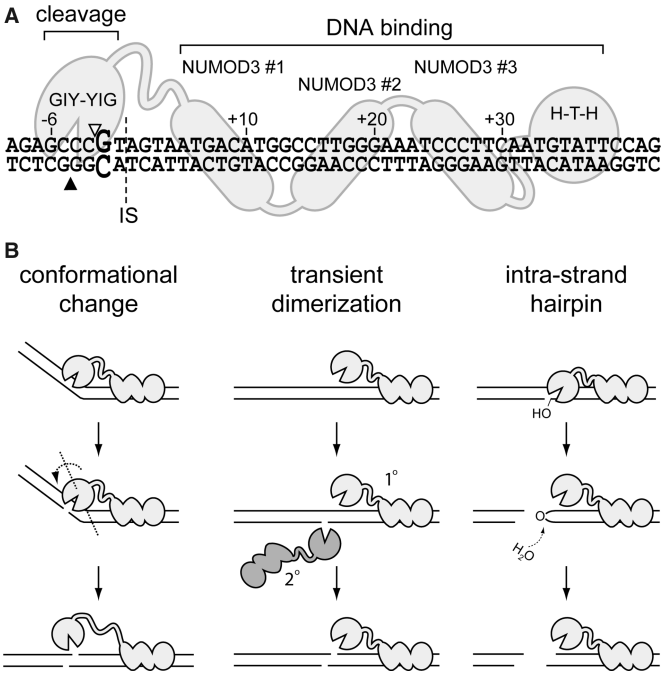

Figure 1. I-BmoI interactions with substrate and monomeric cleavage models. (**A**) Schematic of the modular structure of I-BmoI interacting with intronless *thyA* substrate. Bottom- and top-strand nicking sites are indicated by filled and open triangles, respectively. IS, intron-insertion site. NUMOD, nuclease-associated modular DNA-binding domains. (**B**) Models for DSB formation by monomeric nucleases with single active sites. For transient dimerization, the secondary enzyme molecule can dimerize with the primary DNA-bound molecule from solution or via synapsis as a substrate-bound molecule at an additional target site.

MATERIALS AND METHODS

Strain and plasmid construction

A complete description of strains and plasmids used in this study are found in Supplementary Table S1. *Escherichia coli* strains DH5α, ER2566 and BW25141 (λDE3) were used for plasmid manipulations, expression studies and genetic selections, respectively. Sequences and details of oligonucleotides are in Supplementary Table S2.

Full-length I-BmoI, R27A I-BmoI and N- and C-terminal truncations were expressed and purified as previously described (28). Codon optimized N- and C-terminal I-BmoI domain truncations were cloned by PCR using Phusion (N.E.B.) into the NdeI and KpnI sites of pTYB1. To construct plasmid pLBmo, the PvuII fragment of pBmoHS was subcloned into the PvuII site of LITMUS28i (N.E.B.). A duplex 49mer oligonucleotide containing the I-BmoI *thyA* target site (DE-906/907) was ligated into the SwaI site of pLBmo in the opposite or same orientation as the primary I-BmoI target site to generate the two-site plasmids pLBmo2a and pLBmo2b, respectively. Quikchange (Stratagene) reactions were performed on pLBmo to generate pLBmoNa and pLBmoNb by introducing Nt.BbvCI sites 1082 or 60 bp from the I-BmoI top-strand nick site, respectively. Both Nt.BbvCI sites were introduced into pLBmo to generate pLBmo2N.

To create the pTox plasmid, Quikchange was performed (DE-1171/1172) to remove the EcoRI site from p11-lacY-wtx1 (40), and an oligo cassette containing an EcoRI site (DE-1173/1174) was ligated into XbaI/SphI. The kanamycin-resistant pKox plasmid was generated by cloning the kanamycin resistance gene into the ScaI site of pTox. The vectors pKoxBmoHS and pKoxBmoIn+ were similarly generated from their respective pTox precursors (28). pTox and LITMUS28i plasmids containing variant I-BmoI homing sites with G + 7A, G + 7C or G + 7T mutations were generated by ligating the appropriate duplex oligos into XbaI/EcoRI of either vector. Plasmid constructs with nucleotide insertions in the I-BmoI homing site were generated by ligating the appropriate oligos into XbaI/EcoRI of LITMUS28i. All plasmid constructs were verified by sequencing.

Identification of I-BmoI domains

Limited proteolysis trials were conducted in reactions containing 1× binding buffer (50 mM Tris–HCl, pH 8.0, 50 mM NaCl and 1 mM DTT), 5 μM I-BmoI and 5.5 μM intronless *thyA* (DE-116/117), intron-containing (DE-444/445), non-specific (DE-144/144GC) or mutant (see later in the text) 74mer duplex oligonucleotide substrates. Reactions were incubated at room temperature for 5 min, and an aliquot ('0' time-point) was extracted before the addition of 50 ng trypsin, 100 ng chymotrypsin, 5 ng substilisin A or 1 μg elastase in 1× protease dilution buffer (20 mM Tris–HCl, pH 7.9, 50 mM KCl and 10 mM $MgSO_4$). Aliquots were removed from the digest, added to stop buffer (100 mM EDTA, pH 8.0, 150 mM Tris–HCl, pH 6.8, 4.8% SDS and 24% glycerol) and boiled at 95°C before electrophoresis on 15% SDS–PAGE gels. Coomassie-stained bands of stable domains were excised before in-gel trypsin digestion at the London Regional Proteomics Facility. Recovered peptides were subjected to matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry, and the results analyzed using the UCSF MS-Bridge webpage. To delineate the boundary of the trypsin- and elastase-stable catalytic domain, liquid chromatography (LC) mass spectrometry was performed on a 3Q mass spectrometer (Micromass) equipped with a Z-spray source and run in positive ion nanospray mode. Results were analyzed using MassLynx 4.0.

Biophysical characterization of I-BmoI

Differential scanning calorimetry was performed using a VP-DSC MicroCalorimeter (MicroCal) equilibrated with buffer (20 mM Tris–HCl, pH 8.0, 500 mM NaCl and 5% glycerol) in a 500-μl cell at the Biomolecular Interactions and Conformations Facility at the University of Western Ontario. Trials consisted of 5.2 μM I-BmoI in the presence or absence of 5.9 μM *thyA* 74mer duplex oligonucleotide substrate, with a heating rate of 1°C/min from 10°C to 100°C. Data were analyzed with MicroCal Origin 7 software (including corrections for buffer baseline and protein concentration) using a two-state model that assumes a single independent unfolding transition (41).

DNA-binding affinity of I-BmoI domains

Binding reactions contained 250 mM NaCl, 50 mM Tris, pH 8.0, 2.5 mM EDTA, 1 mM DTT, 5 ng/μl of poly dIdC, 5% glycerol and 0.25 nM of intronless *thyA*, intron-containing or randomized 74-mer substrates radiolabeled on the top-strand with γ-^{32}P. Serial dilutions of the I-BmoI domain truncations were added to the reaction mixture before incubation at room temperature for 15 min and the addition of loading buffer. Reaction was loaded on 10% native polyacrylamide (19:1) and visualized using a Phosphorimager (GE Healthcare). Data were fit to the equation $\frac{[AB]}{[A]_{total}} = \frac{[B]}{[B]+K_D}$.

Gel filtration chromatography

I-BmoI and I-BmoI:*thyA* complexes (DE-130/131) were incubated at room temperature for 10 min, spun for 10 min at 9300*g*, applied to a Superose 12 10/300 GL column (Amersham Biosciences) pre-equilibrated with dialysis buffer (25 mM Tris–HCl, pH 8.0, 500 mM NaCl and 5% glycerol) and run at 4°C at a flow rate of 0.5 ml/min. Fractions of 0.25 ml were collected and run on a 12% SDS–PAGE gel [29:1 (w/w) acrylamide to bis-acrylamide] and a 1% agarose gel. Protein standards used to calibrate the column: carbonic anhydrase, 29 kDa; albumin, 66 kDa; β-amylase, 200 kDa; and apoferritin, 443 kDa. The calibration curve was used to calculate the molecular masses of I-BmoI, *thyA* substrate and the I-BmoI:*thyA* complex by interpolating elution volumes onto the curve.

Cleavage assays with plasmid substrate

The dependence of the initial reaction velocity on protein concentration was determined by cleavage reactions containing 20 mM Tris–HCl, pH 8.0, 250 mM NaCl, 2 mM $MgCl_2$, 2.5% glycerol, 10 nM pBmoHS and I-BmoI (5.5 nM to 175 nM). Aliquots were removed at time points into stop dye (100 mM EDTA, 25% glycerol and 0.2% bromophenol blue), heated for 5 min at 95°C, electrophoresed on a 1% agarose gel and stained in ethidium bromide (Caledon) solution before analysis on an AlphaImager™3400 (Alpha Innotech). Rate constants were calculated as best fit value of at least three experiments, as previously described (9). Initial reaction velocities were determined by calculating the slope of the line for the early time points of product formation. All subsequent plasmid cleavage assays were conducted as described earlier in the text with exceptions indicated later in the text.

Two-site plasmid cleavage assays were performed on the single-site pLBmo plasmid and two-site pLBmo2a and pLBmo2b plasmids. Reactions were supplemented with either 2 or 10 mM $MgCl_2$, and I-BmoI was added to a final concentration of 22.1, 54.3 or 87.5 nM. Domain addition assays were performed on pBmoHS with 10 mM $MgCl_2$. I-BmoI was added to 11 or 22 nM in reactions containing 11 μM N111 I-BmoI, N130 I-BmoI, or bovine serum albumin (BSA) (1000- or 500-fold excess domain, respectively). Assays on plasmids with nucleotide insertions or G + 7 substitutions were performed on pLBmo,

pLBmo + 5T, pLBmo + 5C pLBmo + 3Ta, pLBmo + 3Tb, pLBmo + 3Ca, pLBmo + 3Cb, pLBmoG + 7A, pLBmoG + 7C and pLBmoG + 7T. Reactions were supplemented with 0.5, 2 or 10 mM $MgCl_2$ and 87.5 nM I-BmoI.

Gel-mobility shift assays with pre-nicked substrates

Binding reactions were assembled in the presence of 10 mM EDTA or 10 mM $MgCl_2$ and run as previously described (34). To generate the three substrates required, the top-strand 74mer (DE-116) was 5′-labeled with γ-^{32}P before annealing with the complement full-length bottom-strand oligo (DE-117) to generate the *thyA* substrate (WT). The substrate with the 3′-OH −5/−4 bottom-strand nick was generated with a 3′-23mer (DE-125) and 5′-51mer (DE-123) (−4 nick). To create the −5/−4 bottom-strand nick substrate with a 3′-H (−4 ddGTP), the 3′-23mer (DE-125) and 5′-50mer (DE-124) that had been extended with ddGTP by terminal transferase (N.E.B.) were used. To ensure addition of a single-ddGTP nucleotide to DE-124, we analyzed the product of the terminal transferase reaction by denaturing electrophoresis that revealed only a single nucleotide addition as compared with an unreacted DE-124 oligonucleotide.

thyA cleavage site and spacer selection

The 4-bp upstream and 9-bp downstream of G − 2 in the *thyA* substrate were randomized to generate the pKoxRCS library (DE-1228, GCGGAATTCGANNNN GNNNNNNNNNCATGGCCTTGGGAAATCCCTTC AATGTATTCCGGCATG). A 5′-phosphorylated primer (DE-1229) complimentary to the underlined sequence was annealed to leave a four-nucleotide 3′-SphI overhang. The library was made double stranded by extension with 3′→5′ exo$^-$ Klenow Fragment (N.E.B.), digested with EcoRI and ligated into EcoRI/SphI-cut pKox. The potential complexity of the library was estimated to be $\sim 1.6 \times 10^4$ based on the number of independent transformants.

To screen pKoxRCS for cleavable target sites, 60 µl of BW25141 (λDE3) cells harboring pACYCIBmoI were transformed with 200 ng of pKoxRCS. Transformations were allowed to recover in 500 µl of SOC at 37°C before plating on non-selective media (LB plus 25 µg/ml of chloramphenicol, 50 µg/ml of kanamycin and 0.2% glucose). A total of 2400 transformant colonies, representing ∼15% of the library, were replica-gridded on non-selective and selective media (LB plus 25 µg/ml of chloramphenicol and 10 mM L-(+)-arabinose). Colonies containing cleavable substrates survived on both media, whereas those with non-cleavable substrates survived only on non-selective media. Eighty-six colonies were verified as cleavable by replica-gridding in triplicate, and 96 clones were verified as non-survivors. The survivors and non-survivors were grown in separate 96-well plates in 1 ml of LB + 50 µg/ml of kanamycin and 0.2% glucose, and 200 µl from each well was pooled before DNA extraction to generate survivor and non-survivor plasmid libraries.

Samples were prepared for Ion Torrent sequencing at the London Regional Genomics Centre by PCR amplifying the survivor, non-survivor and pKoxRCS libraries with PWO (Roche) for 25 cycles with barcoded PCR primers. Custom Perl scripts were used to interpret the sequencing results, revealing 9982 unique sequences in the input library with an approximately equal distribution of all four nucleotides at each position. Sequences that occurred >10 times in the survivor and non-survivor pools were retained for further analyses only if the sequences were also present in the input library. Nucleotide proportions at each position in the input and survivor data sets were used to calculate a position-specific scoring matrix (PSSM), and they were used to display relative enrichment by calculating the differences in proportions for each nucleotide at each position between the data sets.

in vivo survival assays

Two plasmid *in vivo* substrate activity assays were performed as previously described (28), with toxic (reporter) plasmids containing the wild-type *thyA* I-BmoI target site or variants with G + 7 mutations. Briefly, the *in vivo* I-BmoI cleavage efficiency for each substrate was calculated by dividing the number of colonies observed on selective plates (LB plus 25 µg/ml of chloramphenicol and 10 mM L-(+)-arabinose) by those on non-selective plates (LB plus 25 µg/ml of chloramphenicol).

OP-Cu in-gel footprinting

The 1,10-phenanthroline copper (OP-Cu) footprinting experiments were performed using 10 pmol of the catalytically inactive R27A variant of I-BmoI, as previously described (34). Duplex 74mer substrates (0.1 pmol) used include *thyA* or *thyA* with G + 7A (DE-1356/1357), G + 7C (DE-1358/1359) or G + 7T (DE-1360/1361) substitutions. Hypo- and hyper-sensitivity to OP-Cu was calculated as the pixel density ratio between bands in the protein:DNA complex (UC) and unbound substrate lanes. Traces were normalized using bands outside of the footprinting region (C-12 and A-11).

RESULTS

I-BmoI is partially disordered in the absence of substrate

Previous bioinformatic and experimental studies have suggested that I-BmoI is a modular protein composed of distinct N- and C-terminal domains connected by a linker (23,42,43). To investigate domain structure changes in the presence of substrate, we performed limited proteolysis experiments with I-BmoI or I-BmoI in complex with its cognate *thyA* substrate, derived from the intronless version of the *B. mojavensis thyA* gene (22). Trypsin, elastase and chymotrypsin digests of I-BmoI revealed a pattern of proteolysis converging on a single-stable domain of ∼10 kDa (Figure 2A and Supplementary Figure S1A). Solution mass-spectrometry, as well as in-gel trypsin digest, followed by mass spectrometry identified the stable domain as the N-terminal GIY-YIG domain consisting of residues 1–92. Strikingly, protease digests of I-BmoI in complex with its cognate 74mer *thyA* substrate under identical conditions revealed a different pattern of protease sensitivity. For all three proteases,

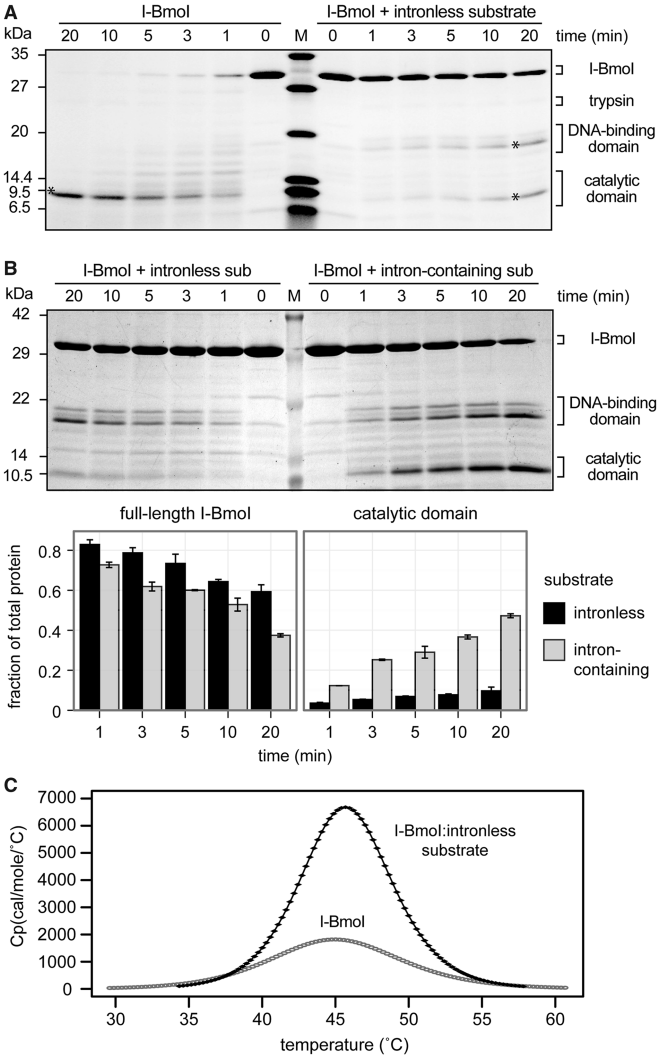

Figure 2. Identification of stable I-BmoI domains. (**A**) Images of Coomassie-stained SDS-gels of trypsin-limited proteolysis time-course experiments performed on I-BmoI and I-BmoI pre-incubated with intronless *thyA* substrate. Aliquots were removed at the indicated time points. In-gel trypsin digest followed by mass spectrometry was performed to identify the peptide products indicated with an asterisk. M, protein marker (sizes in kDa indicated to the left). (**B**) (*top panel*) Coomassie-stained SDS-gel of trypsin-limited proteolysis time-course experiments of I-BmoI in complex with intronless or intron-containing substrates. (*bottom panel*) Plot of the fraction of full-length I-BmoI remaining or increase of the catalytic domain peptide over time. Fraction of total protein is calculated after normalization to the untreated lane, with error bars representing the standard deviation of two replicates. (**C**) Enthalpy of transition curves for differential scanning calorimetry of I-BmoI or I-BmoI pre-incubated with substrate.

substrate-bound I-BmoI was far more resistant to proteolysis (Figure 2A and Supplementary Figure S1A). Additional stable domains of ~18–20 kDa were also observed and subsequently identified by in-gel trypsin digestion followed by mass spectrometry as peptides that mapped C-terminal to residue 130. Western blot analyses using an antibody raised against residues 1–111 of I-BmoI (N111) confirmed the identities of both stable domains (Supplementary Figure S1B). Digests of I-BmoI with all three proteases in the presence of a non-cognate oligonucleotide substrate revealed a pattern of protease sensitivity similar to that observed for I-BmoI without substrate (data not shown). No change in protease sensitivity was observed under cleavage conditions with 10 mM $MgCl_2$. The protease digestion patterns of the I-BmoI:*thyA* complex revealed a stable two-domain structure resulting from specific DNA binding where the inter-domain linker remains moderately accessible to proteases.

An exception was observed when digests were performed with intron-containing *thyA* substrate. I-BmoI binds the intron-containing substrate with similar affinity to the cognate intronless allele; however, nucleotide differences upstream of the intron insertion site within the intron-containing target abolish the I-BmoI cleavage site (23). Trypsin digests of the I-BmoI:intron-containing substrate complex showed a loss of protection of full-length protein, revealing an ~25% greater degradation after 20 min into N- and C-terminal domains (Figure 2B). A ~4-fold increase in the presence of the catalytic domain fragment was also observed, suggesting that nucleotide changes at the cleavage site disengage the catalytic domain from substrate resulting in increased protease sensitivity of the I-BmoI linker. These results are consistent with previous results where I-BmoI was unable to distort the intron-containing substrate as a prerequisite to bottom-strand nicking, suggesting a loss of catalytic domain contacts at the cleavage site (36).

Additional insight into the stability of I-BmoI was gained through differential scanning calorimetry analyses that revealed a broad denaturation profile with an observed T_m of 45.1°C. In the presence of *thyA* substrate, we observed an ~5-fold increase in the enthalpy of transition versus that observed for I-BmoI in solution, consistent with an increase in protein stability (Figure 2C) (44). Collectively, these data suggest that only the N-terminal catalytic domain of I-BmoI is stable in the absence of substrate, and that the C-terminal domain becomes structured on DNA binding.

N-terminal domains have weak but specific DNA-binding affinity

To investigate the solubility and DNA-binding activity of the N- and C-terminal domains of I-BmoI, we expressed and purified a number of truncations. N-terminal constructs that contained the GIY-YIG nuclease domain were soluble and expressed well (N92, N111, N130 and N154), whereas C-terminal constructs were generally insoluble, with the exception of 130 C (Figure 3A). Binding assays with full-length I-BmoI and soluble domain truncations showed stable complex formation with *thyA* substrate (Supplementary Figure S2A), and quantitative DNA-binding assays were performed to determine dissociation constants of each of the N-terminal domain truncations for *thyA* intronless substrate. Binding reactions with intron-containing substrate were also performed to determine whether the N-terminal truncations possessed affinity for the cleavage site region of *thyA* substrate. The dissociation constants for the N92, N111 and N130 truncations were in the 10 μM range, whereas the N154

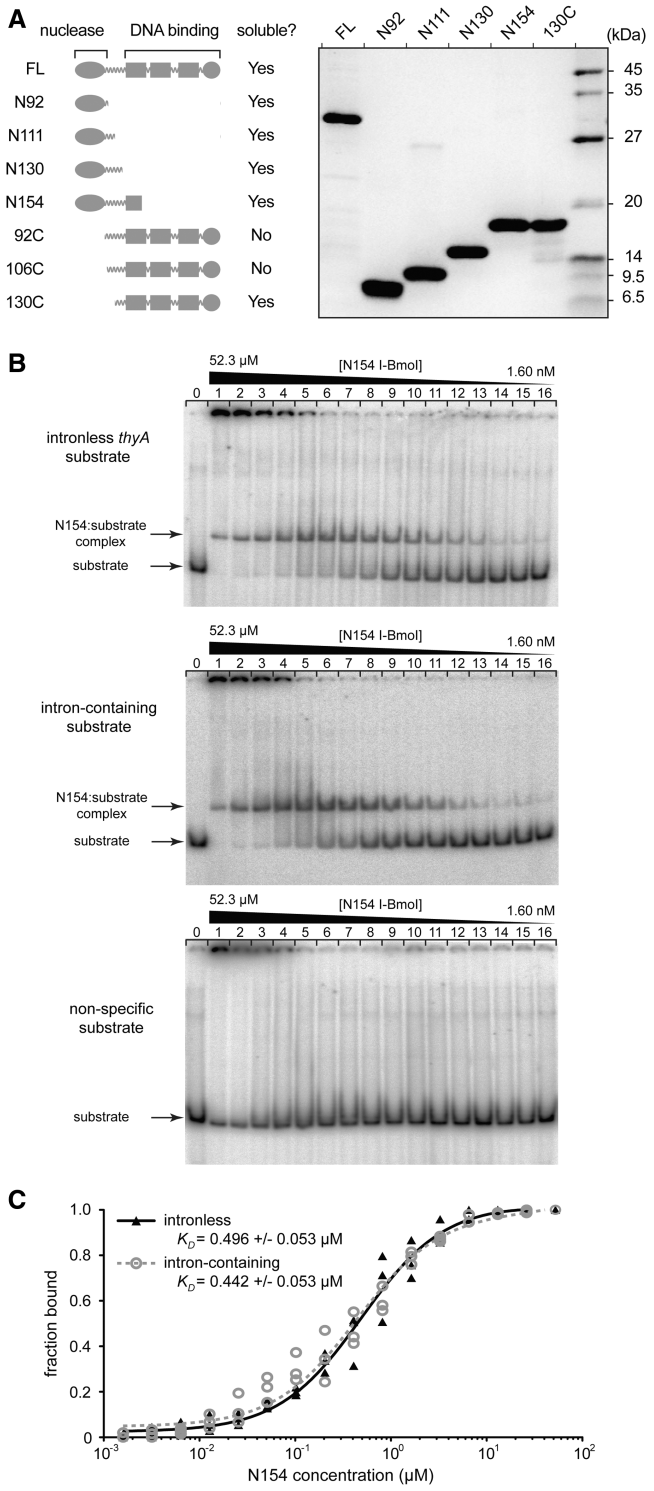

Figure 3. I-BmoI domains have distinct affinities for DNA. (**A**) Schematic of I-BmoI truncations and image of a Coomassie-stained SDS-page gel containing 1.5 µg of each soluble construct (*left* and *right* panels, respectively). FL, full-length I-BmoI; N92, residues 1–92 of I-BmoI; N111, 1–111; N130, 1–130; N154, 1–154; 92C, 92–266; 106C, 1–266; and 130C, 130–266. (**B**). Images of native gel-shift experiments for N154-binding reactions performed on labeled intronless, intron-containing and non-specific substrate. The left lane contains substrate only, and lanes 1 through 16 contain serial dilutions of N154 from 52.3 µM to 1.60 nM. (**C**) Graph of the binding curves of N154 I-BmoI on intronless *thyA* or intron-containing substrate, with data from three replicates plotted (see also Supplementary Table S3).

truncation had a K_D of ~500 nM for both intronless and intron-containing substrates, indicating that the cleavage site region is not a major determinant for DNA binding (Figure 3B, Supplementary Figure S2B and Supplementary Table S3). Discrete complex formation with the N-terminal truncations in the presence of a non-specific substrate was not observed under our binding conditions, apart from N154, which aggregated at a concentration that increased the K_D by >100-fold (Supplementary Table S3 and Supplementary Figure S2B). The dissociation constants for the full-length protein and 130C DNA-binding domain were ~10 nM for both intronless and intron-containing *thyA* substrates, consistent with previous data (23,31). Collectively, these data show that I-BmoI possesses two functional domains connected by a protease-sensitive linker, where the N-terminal catalytic domain has weak affinity for DNA and the C-terminal domain contributes the majority of the DNA-binding energy.

I-BmoI binds DNA as a monomer

One model that has been proposed for DNA cleavage by GIY-YIG homing endonucleases requires the formation of a transient or stable dimer, with each monomer nicking one DNA strand (36). The ability to purify wild-type I-BmoI allowed us to determine the oligomeric status of I-BmoI and I-BmoI:substrate complexes in solution using gel filtration analyses (Figure 4). The elution profile of I-BmoI alone (35 kDa) matched well with the calculated mass of an I-BmoI monomer (32 kDa), indicating that in solution, I-BmoI is monomeric. The 46mer *thyA* DNA substrate eluted at an observed mass of 99 kDa, well above that of its calculated mass of 29 kDa because of aberrant migration of cylindrical DNA through the gel-filtration column. The observed I-BmoI:*thyA* complex, assembled at a 5:1

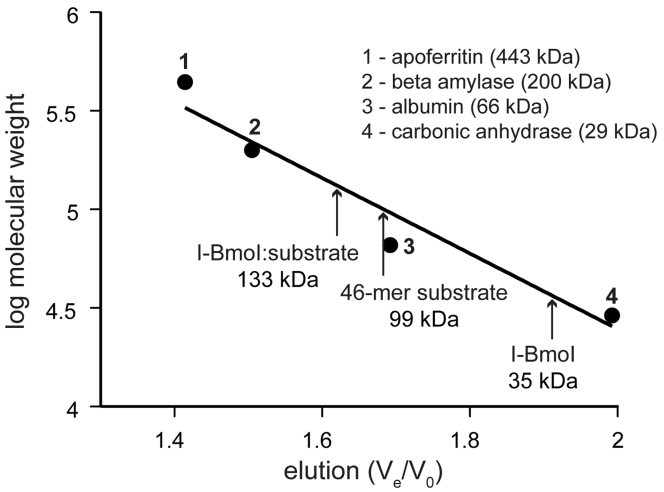

Figure 4. I-BmoI is a monomer in solution and in complex with *thyA* substrate. Graph of gel filtration elution profile of I-BmoI, The 46mer *thyA* substrate, or I-BmoI:substrate complex, with observed molecular weights indicated. Standards used to generate the elution volume standard curve are shown. Gel filtration analyses were performed in duplicate. V_e, elution volume; V_o, void volume.

protein:DNA ratio, eluted with a size (133 kDa) consistent with an monomeric I-BmoI:DNA substrate complex. Changing the protein to DNA ratio to 1:2 did not change the elution profile of the I-BmoI:*thyA* complexes, suggesting that I-BmoI does not synapse between two DNA molecules (data not shown). The gel-filtration results were confirmed by protein cross-linking experiments using two different reagents with I-BmoI and I-BmoI:*thyA* complexes (Supplementary Figure S3). No cross-linked products were observed under binding or cleavage conditions (without exogenous metal or with 10 mM $MgCl_2$, respectively) by gel electrophoresis or western blot analyses, even at micromolar concentrations of protein. Thus, I-BmoI is monomeric and does not form stable higher order complexes in solution or when in complex with substrate.

The oligomeric status of I-BmoI does not change during cleavage

To specifically test whether the oligomeric status of I-BmoI changes during cleavage, we performed *in vitro* kinetic assays to (i) examine the dependence of initial reaction velocity on protein concentration; (ii) determine whether synapsis by two endonuclease monomers in a two target site plasmid can enhance cleavage rate; and (iii) investigate whether adding excess catalytic domain can stimulate cleavage (37,38). Before performing the *in vitro* kinetic assays, we sought to examine whether DNA topology influenced the rate of substrate cleavage by I-BmoI. To accomplish this, we used a plasmid containing a *thyA* target site into which we introduced restriction sites for the Nt.BbvCI nickase at two positions. The same plasmid could, therefore, be used as a supercoiled, relaxed or linear substrate depending on whether the plasmid was relaxed with Nt.BbvCI or linearized with SwaI. As shown, no significant difference was observed for the rate of product formation on supercoiled, relaxed plasmid or linear substrates at 0.5, 2 or 10 mM $MgCl_2$ (Supplementary Figure S4), and no difference in rate was observed on relaxed plasmids that were nicked at either or both Nt.BbvCI sites (data not shown). Subsequent *in vitro* cleavage assays were performed on supercoiled plasmid substrate unless otherwise indicated, as this allowed us to observe and determine rate constants for all steps of the reaction.

Insight into the oligomeric status of I-BmoI during cleavage was obtained by establishing the relationship between enzyme concentration and initial reaction velocity. Single-turnover cleavage reactions with seven concentrations of I-BmoI were performed on plasmid substrates containing a single *thyA* target site where product formation was measured by the appearance of linear DNA (Figure 5A). The initial rate of product appearance for each I-BmoI concentration was calculated and plotted against I-BmoI concentration, yielding a linear dependence (Figure 5B). These data suggest that I-BmoI–catalyzed DNA hydrolysis is first order with respect to protein concentration, indicative of non-cooperative cleavage.

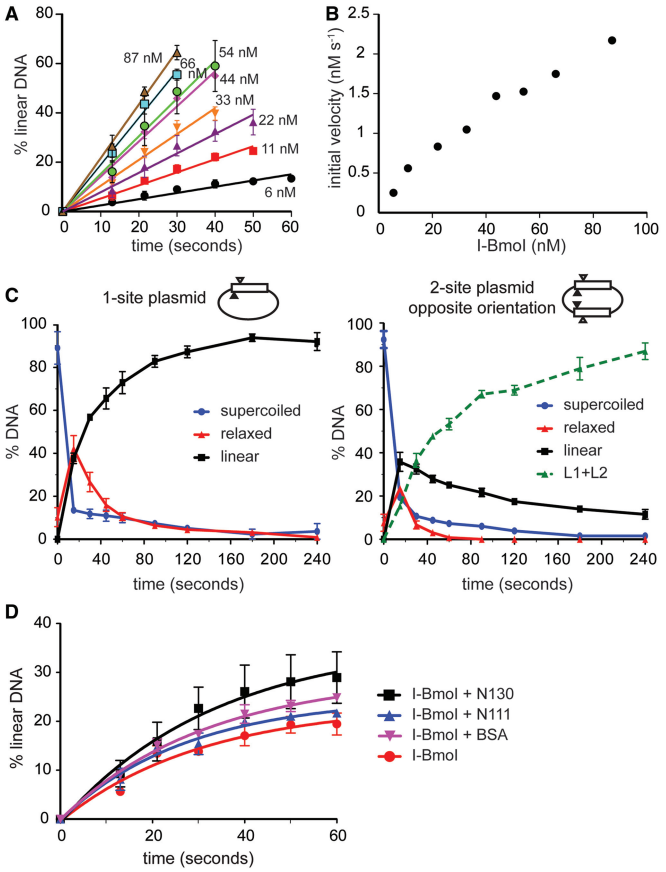

Figure 5. I-BmoI functions as a monomer. (**A**) Graph of initial reaction progresses for cleavage assays with eight I-BmoI concentrations expressed as per cent linear product. (**B**) Plot of initial reaction velocity versus I-BmoI concentration. (**C**) Graph of reaction progress for cleavage assays with I-BmoI and one- or two-site substrate plasmids (*left* and *right panels*, respectively). L1 + L2, linear products from two-site plasmid cleavage. (**D**) Domain addition experiments. Graph of linear product formation for cleavage assays with 22 nM I-BmoI and 10 nM plasmid. Reactions were supplemented with 500-fold molar excess of N111, N130 or BSA. All cleavage assays were performed in triplicate.

Additional cleavage assays were performed on substrates containing two target sites to address whether the cleavage rate could be enhanced by synapsis of two I-BmoI monomers bound to distinct target sites within a single plasmid. Reaction progress was examined on plasmids that contained a primary target site alone, or those that contained an additional secondary target site in the opposite or same orientation relative to the primary site (Figure 5C and Supplementary Figure S5). No rate enhancement was observed for either nicking reaction by I-BmoI on plasmids containing secondary target sites relative to plasmids with a single site, suggesting that synapsis is not required for cleavage. To further address a cooperative or synaptic cleavage mechanism, biotin-tagged/radiolabeled oligonucleotides were immobilized on streptavidin-coated magnetic beads before the addition of I-BmoI (Supplementary Figure S6A). After the beads were stringently washed to remove unbound I-BmoI, cleavage was induced by the addition of

magnesium, and product liberation was measured by quantification of radiolabel released into solution. Cleavage levels were similar regardless of whether excess I-BmoI was added to the pre-formed I-BmoI:substrate complexes during the magnesium cleavage step (Supplementary Figure S6B).

We tested whether transient interactions between catalytic domains could stimulate cleavage by performing cleavage assays with I-BmoI in the presence of excess N-terminal domain truncations. Cleavage assays with I-BmoI and 500-fold molar excess N111 or N130 did not stimulate cleavage rate relative to the addition of BSA (Figure 5D). Assays with a 500-fold molar excess of N154 led to non-specific cleavage, as the concentration of the domain added approached the K_D of non-specific binding for N154 (Supplementary Table S3). A similar pattern of non-specific substrate degradation was observed for reactions containing the same concentration of N154 alone (data not shown). Given that I-BmoI exists as a stable monomer in solution and in complex with substrate, these data demonstrate that I-BmoI acts as a monomer at all stages of the cleavage pathway and rules out transient dimerization through the catalytic domain as a potential cleavage mechanism.

I-BmoI does not function via a substrate hairpin mechanism

As a monomeric nuclease, I-BmoI could use a substrate 'hairpin' mechanism to hydrolyze DNA, where the liberated 3′-OH from the primary nick in DNA acts as the nucleophile to attack the scissile phosphate bond of the opposite strand. The intra-strand transesterification mechanism has been demonstrated for the Tn10 transposase, the retroviral HIV-1 DNA intergrase and V(D)J recombinases (45–47). Alternatively, I-BmoI could form a covalent protein–DNA intermediate, as is the case with the phospholipase D superfamily enzyme BfiI (48). We tested this hypothesis by generating substrates that contained pre-nicked bottom-strands with either a 3′-OH or 3′-H at the bottom-strand nick site. If I-BmoI uses an intra-strand transesterification reaction, the second nicking reaction of the top-strand should be greatly reduced with the 3′-H substrate. Both pre-nicked substrates were bound and cleaved equivalently to the intact *thyA* substrate, demonstrating that the primary 3′-OH product is not required for resolution of the DSB, and that I-BmoI does not function via an intra-strand substrate hairpin mechanism (Figure 6).

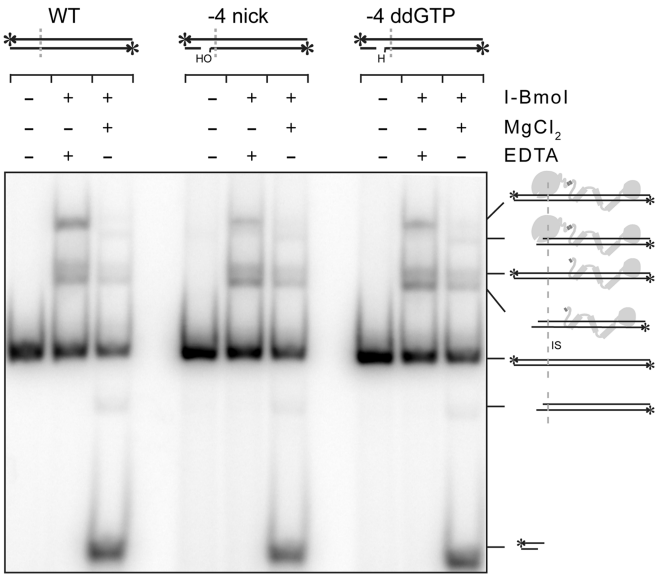

Figure 6. I-BmoI does not function via an intra-strand hairpin mechanism. Image of I-BmoI-binding reactions performed with dual-labeled *thyA* substrate (WT), bottom-strand pre-nicked substrate between −5/−4 (−4 nick) and bottom-strand pre-nicked substrate lacking the 3′-OH at the nick site (−4 ddGTP). Reactions were performed with 10 mM EDTA or 10 mM $MgCl_2$. Schematic and naming of complexes are as previously described (34). Reactions were performed in duplicate showing similar results.

The I-BmoI target site is modular

As I-BmoI acts as a monomer throughout all steps of the cleavage pathway, a potential mechanism for DSB formation involves dynamic conformational changes within the enzyme:substrate complex leading to the reorientation of the single I-BmoI active site. This model implies that the *thyA* substrate should also be modular and contributes to the DNA distortions observed during cleavage (34). Previous results indicated that the G − 2 base is critical for cleavage, and that other bases surrounding the cleavage site can modulate cleavage efficiency (31,36). Additionally, nucleotide insertions downstream from the intron insertion site between positions A + 8/C + 9 alter cleavage by I-BmoI (42).

To delineate functional regions of the *thyA* substrate and examine the effect of nucleotide insertions on the first- and second-strand nicking rates, we generated a series of plasmid substrates that contained target sites with three-nucleotide insertions between A + 4/A + 5 or A + 8/C + 9 and substrates with five-nucleotide insertions between A + 8/C + 9 (Figure 7A). Substrates containing TTT and CCC insertions between A + 4/A + 5 attenuated cleavage by ∼5-fold under standard conditions with 10 mM $MgCl_2$, leading to second-strand rate constants (k_2) of $0.0084 \pm 0.0006\,s^{-1}$ and $0.0051 \pm 0.0003\,s^{-1}$, respectively (Figure 7B). Also evident from the reaction progress curves is the lack of accumulation of a nicked intermediate, consistent with first-strand nicking being rate limiting ($k_1 < k_2$). Assays performed on the same substrates with 2 or 0.5 mM $MgCl_2$ showed low levels of cleavage (Supplementary Figure S7). Cleavage assays on the four substrates with target sites containing three- or five-nucleotide insertions between A + 8/C + 9 showed only minimal nicking activity after prolonged incubations under standard reaction conditions with 10 mM $MgCl_2$ (data not shown). These substrates allowed us to identify the sequence upstream of A + 5 as tolerant to insertions and the sequence downstream of A + 8 as intolerant to insertions. One interpretation of this data is that specific bases in the A + 5/A + 8 region may act as anchor points to position the I-BmoI linker or DNA-binding domain on substrate.

To determine whether individual bases in the G − 6 to A + 8 region of *thyA* substrate were important for function

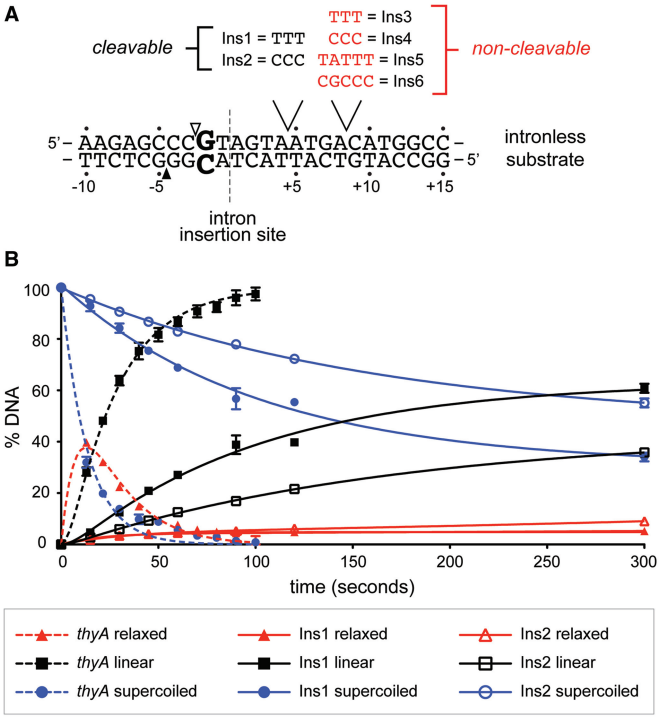

Figure 7. Insertions in the I-BmoI substrate spacer reduce or abolish cleavage activity. (**A**) Schematic of the wild-type *thyA* I-BmoI substrate and the substrates generated containing three- or five-nucleotide insertions. Filled and open triangles represent the bottom- and top-strand nicking sites, respectively. (**B**) Progress curves for cleavage reactions on *thyA*, Ins1 and Ins2 substrates in 10 mM $MgCl_2$. Curves are generated from reactions performed in triplicate.

in vivo, we adapted an *in vitro* substrate selection that previously identified G − 2 to be critical for cleavage by I-BmoI (31). We generated a target site with randomized nucleotides between G − 6 and A + 8, while maintaining G − 2 fixed, to create a library of target site toxic reporter plasmids (pKoxRCS) to be screened *in vivo* (Figure 8A). The pKoxRCS library was transformed into cells harboring the I-BmoI expression plasmid (pACYCIBmoI), and transformants were replica-gridded in triplicate onto selective and non-selective plates (Figure 8B), where control toxic plasmids containing the *thyA* intronless target site survived on selective media and plasmids encoding the intron-containing target did not. We identified 86 robustly cleaved reporter plasmids and sequenced their target site regions, as well as the input library of randomized sites using barcoded PCR products on an Ion Torrent platform. A plot of the relative change in nucleotide proportions at each position between the survivor sequences and input pool, as well as a position-specific scoring matrix (PSSM), identified the wild-type guanine nucleotide as highly preferred at position +7 (Figure 8C and Supplementary Figure S8A). Consistent with previous results, positions within the G − 6 to T − 1 cleavage motif did not show strong nucleotide preference (31,36). Positions that displayed slight nucleotide preference identified the wild-type nucleotide as preferred at that position. A plot of the relative change in nucleotide proportion between the survivor sequences and a pool of non-cleavable ('dead') clones showed similar results, identifying +7 as the only position with a strong preference for any nucleotide (Supplementary Figure S8B).

Contacts to G + 7 facilitate cleavage by I-BmoI

To better understand the preference for guanine at position +7, we designed plasmid substrates that contained G + 7A, G + 7C and G + 7T target site substitutions for *in vitro* and *in vivo* assays. Cleavage assays on the G + 7T substrate with 0.5 mM $MgCl_2$ showed an ~2.5-fold decrease in both k_1 and k_2 versus the wild-type G + 7 substrate (Figure 9A and Table 1). The G + 7A substrate displayed kinetics similar to the wild-type target site, whereas G + 7C showed a slight first-strand nicking defect (Table 1). Metal-dependent rescue of cleavage defects on mutant substrates *in vitro* was observed in reactions with 2 or 10 mM $MgCl_2$, consistent with previous results (36).

The observed cleavage defects on the G + 7 substitution substrates were exaggerated when the same target sites were examined by the *in vivo* two-plasmid survival assay, possibly because of low levels of *in vivo* protein expression versus excess protein to DNA ratios *in vitro* (28,40). Briefly, the cleavage efficiency of I-BmoI on non-cognate substrates can be assessed by the degree to which a target site embedded within a toxic reporter plasmid is cleaved, and survival percentage is reported as the ratio of colonies on selective versus non-selective plates (28). We observed ~90% survival by I-BmoI on the wild-type *thyA* target site and decreasing survival on G + 7A > G + 7C > G + 7T substrates (Figure 9B). The correlation of reduced *in vitro* cleavage and poor *in vivo* survival for transversion substitutions at G + 7 suggests that a purine at position +7 acts as a critical anchor point for I-BmoI on *thyA* substrate.

To establish a physical basis for the observed defect associated with substitutions at G + 7, we performed in-gel footprinting with the minor groove specific reagent 1, 10-phenanthroline copper (OP-Cu) on oligonucleotide substrates containing G + 7 mutations. OP-Cu was selected as a footprinting reagent because minor groove distortions of substrate are identified as hypersensitivity to OP-Cu (49), and because it has previously been used to identify DNA distortions surrounding the I-BmoI cleavage site that are dependent on contacts to G − 2 (34,36). Reactions were performed using the catalytically inactive R27A variant of I-BmoI, as bottom-strand nicked products would otherwise occlude distortions surrounding the bottom-strand nicking site (34). We found no difference in the binding affinity of I-BmoI for the three mutant substrates relative to the wild-type G + 7 substrate. Footprinting reactions revealed drastic reductions in hypersensitivity to OP-Cu versus the wild-type substrate at positions surrounding the bottom-strand nicking site on the G + 7C and G + 7T substrates (G − 5/G − 4/G − 3), whereas modest reductions were observed on the G + 7A substrate (Figure 9C). All three mutant substrates did not exhibit sensitivity to OP-Cu on the bottom-strand at position T + 4 that was previously observed for the R27A footprint on wild-type *thyA*

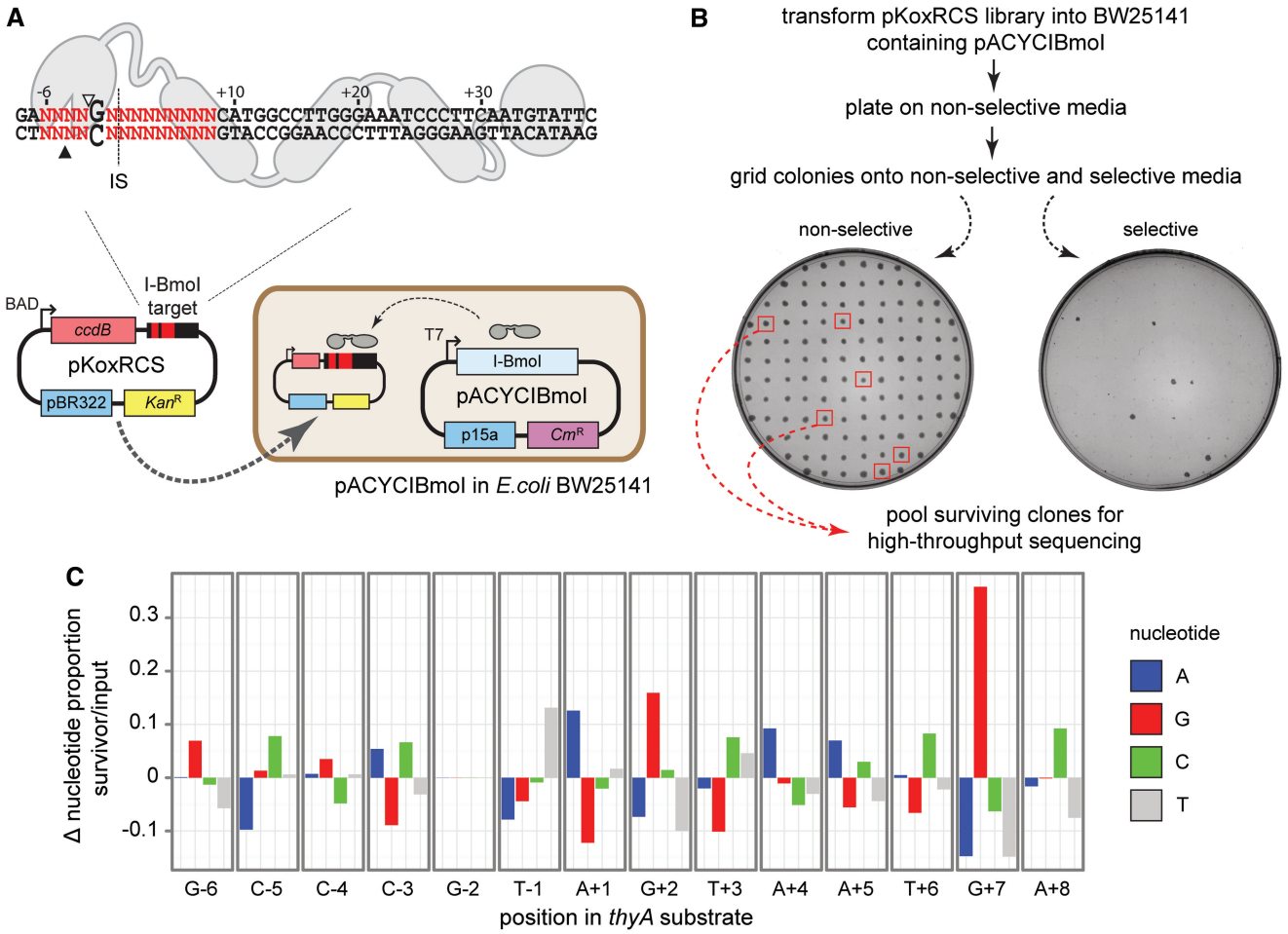

Figure 8. An *in vivo* screen identifies I-BmoI nucleotide preference at G+7. (**A**) (*top*) Schematic representation of the randomized cleavage site from −6 to +8 with G−2 fixed (in bold enlarged font). (*bottom*) The pKoxRCS library and I-BmoI expression plasmids. Filled and open triangles represent the bottom- and top-strand nicking sites, respectively. (**B**) Work flow to select for cleavable target sites. (**C**) Sequencing results from cleavable targets represented as the difference in nucleotide proportion from the clones in the survivor pool versus the input library.

substrate (34). Additionally, a dramatic increase in sensitivity to OP-Cu was observed at A+11 on the bottom-strand on all three mutant substrates. A similar increase in OP-Cu sensitivity at A+11 was previously observed within the lower-complex footprint of I-BmoI (34). The lower-complex consists of I-BmoI:*thyA* complexes that are unable to form DNA-contacts via the catalytic domain, or proteolyzed I-BmoI fragments that retain DNA-binding activity but presumably lack the catalytic domain entirely (23). The footprinting data in Figure 9C suggest that mutations at G+7 reduce I-BmoI catalytic domain and linker contacts that are essential for inducing distortions adjacent to the bottom-strand nicking site, with hypersensitivity at +11 resulting from the catalytic domain and linker of I-BmoI being disengaged from substrate.

To provide further evidence that I-BmoI catalytic domain contacts are disturbed by G+7 substitutions, we performed protease-mapping experiments on I-BmoI:G+7A, G+7C and G+7T substrate complexes. Loss of catalytic domain or linker contacts to substrate would be observed as an increase in protease sensitivity within the I-BmoI linker, and time course reactions with trypsin revealed a reduction in the protection of full-length I-BmoI from protease digestion (Figure 9D and E). An increase in abundance of the catalytic and DNA-binding domain peptides was observed, indicating greater protease sensitivity within the I-BmoI linker in the presence of G+7 mutant substrates versus I-BmoI:*thyA* complexes. An increase in protease sensitivity was also observed for the 130C I-BmoI DNA-binding domain construct in complex with the G+7T substrate. Trypsin digests revealed a significantly greater rate of proteolytic degradation versus 130C I-BmoI:*thyA* complexes, indicating that I-BmoI contacts to the G+7 base pair are C-terminal to I-BmoI residue 130 (Supplementary Figure S9).

The results from the physical assays probing G+7 interactions are consistent with a model whereby G+7 mutations led to (i) a reduction in I-BmoI catalytic domain or linker contacts to DNA surrounding the cleavage site that are necessary to induce bottom-strand distortions near the bottom-strand nicking site; (ii) loss in sensitivity at +4 that is indicative of changes in protein–DNA interactions downstream of the cleavage site; and (iii) a disengagement of the catalytic domain from

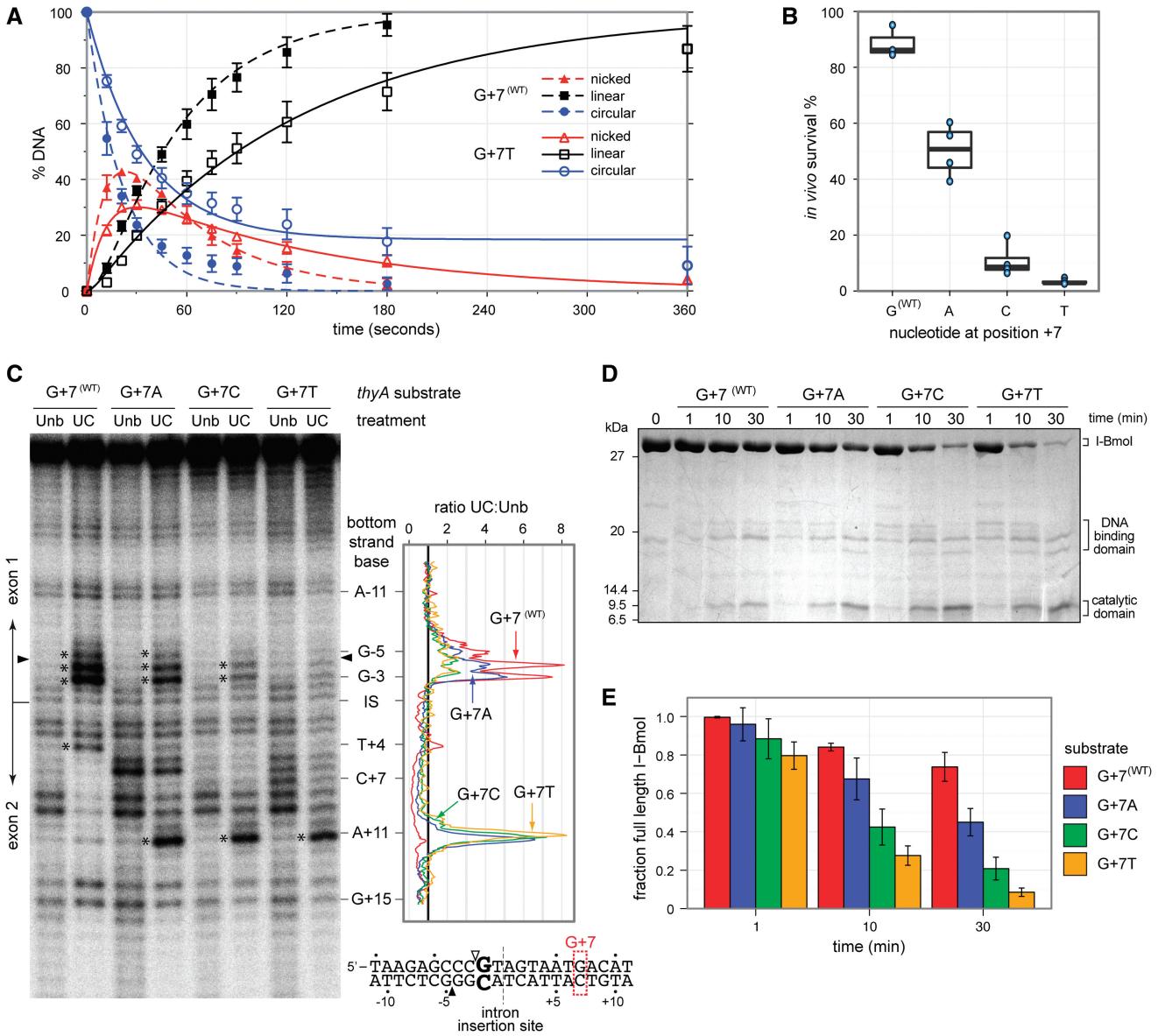

Figure 9. Mutations at G+7 affect I-BmoI activity. (**A**) Plot of reaction progress with I-BmoI and G+7 (WT) or G+7T plasmid substrates in 0.5 mM $MgCl_2$. Curves are plotted as a fit to the mean of three replicates with standard deviation shown. (**B**) *In vivo* survival of I-BmoI on *thyA*, G+7A, G+7C and G+T toxic reporter plasmid substrates, with four replicates plotted. (**C**) Denaturing gel image of in-gel OP-Cu footprinting reactions with R27A I-BmoI on bottom-strand–labeled G+7 (WT), G+7A, G+7C and G+7T substrates (Unb, unbound substrate; UC, full-length I-BmoI bound to substrate). Sites that are hypersensitive to the footprinting reagent are highlighted with an asterisk, and filled triangles indicate the bottom-strand nick site. To the right of the gel image is a graph of the normalized pixel density ratio for bands in the UC lane versus the Unb lane. (**D**) Coomassie-stained SDS-gel of trypsin-limited proteolysis time-course experiments of I-BmoI in the presence of G+7 (WT), G+7A, G+7C and G+7T substrates. (**E**) Plot of the fraction of full-length I-BmoI remaining over time for reactions shown in panel (D), calculated as the fraction remaining after normalization to the untreated lane (error bars represent the standard deviation of two replicates).

Table 1. Rate constants for first- and second-strand nicking reactions on G+7 mutant substrates

Substrate	k_1 (s^{-1})	k_2 (s^{-1})
G+7 (WT)	0.0441 ± 0.0021	0.0198 ± 0.0015
G+7A	0.0478 ± 0.0010	0.0303 ± 0.0007
G+7C	0.0313 ± 0.0014	0.0320 ± 0.0009
G+7T	0.0172 ± 0.0009	0.0082 ± 0.0005

substrate that leads to new DNA distortions at +11, as well as increased protease sensitivity within the linker and across the DNA-binding domain (Figure 10).

DISCUSSION

Early efforts with I-TevI to address the mechanistic question of how the single active site of GIY-HEs is used to make a DSB revealed that the two-domain protein bound DNA as a monomer, that the linker connecting the functional domains was flexible and that the

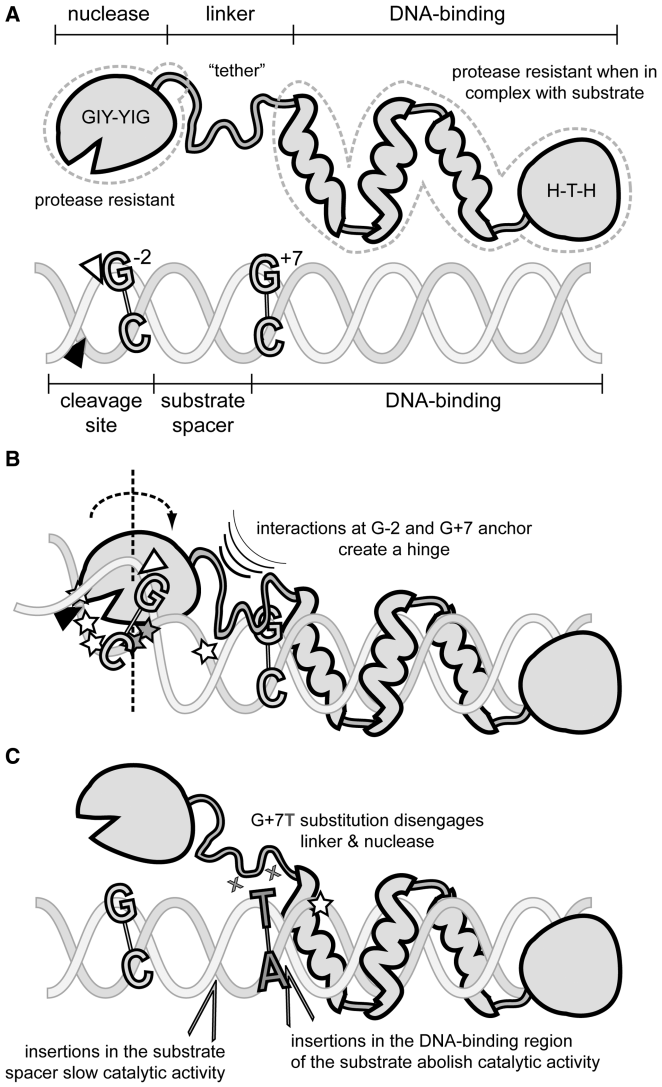

Figure 10. The domains of I-BmoI form distinct DNA-contacts to distort DNA before cleavage. (**A**) Schematic of I-BmoI domains and substrate modularity. The protease-resistant domains of I-BmoI and substrate bases required for efficient cleavage are highlighted. Bottom- and top-strand nicking sites are indicated by filled and open triangles, respectively. (**B**) Model of I-BmoI-bending DNA to induce substrate distortions near the bottom-strand nick site (indicated by white and grey stars that represent OP-Cu hypersensitivity for R27A and WT I-BmoI footprinting reactions, respectively). Rotation of the enzyme:substrate complex after bottom-strand nicking to reposition the nuclease domain for top-strand nicking is indicated by a dashed arrow around a vertical axis. Interactions at G − 2 and G + 7 may form a hinge necessary to distort DNA. (**C**) Schematic of nucleotide substitutions and insertions that affect I-BmoI catalytic activity.

N-terminal GIY-YIG nuclease domain possessed a single active site (10,15,30,35). These data all pointed towards GIY-HEs functioning as monomers, with significant domain rearrangements postulated to reposition the nuclease domain between strand-specific nicking reactions to generate a DSB. However, the extreme cytotoxicity of I-TevI precluded necessary kinetic studies to determine whether the enzyme functioned as a monomer in all stages of the reaction pathway, and it prohibited the exclusion of alternative cleavage mechanisms. A number of mechanisms have been described for monomeric nucleases that make DSBs, and we discuss the relevance of these mechanisms in light of our current data supporting a monomeric mode of action by I-BmoI whereby the C-terminal domain acts as a molecular anchor to tether the N-terminal nuclease domain to sequentially nick both DNA strands.

As structurally bipartite enzymes, GIY-HEs closely resemble the type IIS restriction enzymes FokI and MboII that bind DNA through N-terminal domains and cleave at a distance via C-terminal single-active site catalytic domains (50,51). FokI can form a DSB by a cooperative transient dimerization mechanism where a substrate bound FokI can dimerize with additional solution monomers, or synapse with the catalytic domain of another FokI molecule bound to a secondary target site (37,38). We did not find evidence for I-BmoI functioning via a transient dimerization mechanism, as the rate of cleavage is linear with respect to protein concentration, addition of excess catalytic domain cleavage does not stimulate product formation and two-site plasmids are cleaved at the same rate as a single site plasmid. Furthermore, the structure of the I-TevI catalytic domain and homology model of the I-BmoI catalytic domain did not reveal an obvious dimer interface (10,28,36). Interestingly, the type IIS enzyme Eco31I is also a two-domain protein with a single HNH active site that cleaves close to its binding site (52,53). Unlike FokI, Eco31I does not transiently dimerize to generate a DSB, meaning that the single HNH active site sequentially nicks each strand. The reaction pathway of Eco31I proceeds with roughly equal rate constants for both nicking reactions and, therefore, parallels that of I-BmoI, where the single GIY-YIG active site sequentially nicks the bottom- and top-strands with only an ~2-fold difference in rates.

The two-domain structure of I-BmoI is also similar to I-HmuI, an HNH family homing endonuclease that binds an extended target site (54). I-HmuI consists of an N-terminal HNH catalytic domain, and a C-terminal DNA-binding domain composed of minor-groove binding α-helix (NUMOD3) and helix-turn-helix modules that are similar to those found within the I-TevI DNA-binding domain (43,54,55). I-HmuI, however, nicks only one strand of its substrate, presumably because DNA-binding elements found on either side of the HNH nuclease motif prevent release of the catalytic domain after nicking. Our data suggest that the I-BmoI catalytic domain is released from substrate between nicking reactions, as the minimal I-BmoI N-terminal domain (N92) possesses weak DNA affinity ($K_D > 23$ μM), and limited-protease digestions of full-length I-BmoI on intron-containing and G + 7 mutant substrates revealed that the catalytic domain was disengaged from substrate. Moreover, the homology model of the I-BmoI catalytic domain does not possess additional units of protein structure that bind DNA, consistent with the evolutionary acquisition of distinct DNA-binding modules by GIY-YIG family enzymes because of the inherently weak DNA affinity of the nuclease domain.

DNA cleavage mechanisms that require conformational changes and rearrangements of single active sites have been proposed for the phospholipase D superfamily restriction enzyme BfiI and the type IIP restriction enzymes BcnI and MvaI (56–58). BfiI exists as a dimer; yet, it forms only a single active site at the dimer interface that rotates and switches orientation between sequential nicking reactions (56). I-BmoI is unlike BfiI, as there remains no evidence for the formation of an I-BmoI dimer, and the proposed catalytic mechanism for BfiI involves a covalent DNA intermediate. Based on active site similarity to other GIY-YIG nucleases (10–14,36), it is unlikely that GIY-HEs form covalent protein:DNA intermediates, and we have shown that I-BmoI does not function via an intra-strand transesterification mechanism. Conversely, a mechanism has been demonstrated for BcnI where the enzyme acts as a monomer to nick one strand before 'hopping' to the opposite strand where it interacts in the reverse orientation to generate the second nick (57). The affinity of the I-BmoI DNA-binding domain for substrate and the fact that I-BmoI remains bound to the cleavage product would preclude the full-length enzyme from hopping to the opposite strand and nicking in an alternate orientation. It is possible, however, that the I-BmoI catalytic domain dissociates and rearranges between nicking reactions because of its inherently weak DNA affinity and the flexible nature of the linker.

We envision a monomeric cleavage mechanism by I-BmoI where the DNA-binding domain acts as a molecular tether to allow the GIY-YIG catalytic domain to diffuse and reposition to effect the sequential nicking reactions. Conformational changes in the I-BmoI:substrate complex that precede bottom-strand nicking are dependent on contacts to the G − 2 and G + 7 base pairs, as mutation of either position abolishes DNA distortions by I-BmoI that precede nicking (36). The function of the G + 7 anchor point to position the I-BmoI catalytic domain and linker is supported by +4/+5 and +8/+9 insertions in *thyA* substrate that drastically reduce k_1 or abolish activity, respectively. Insertions at +4/+5 extend the spacing that intervenes the cleavage site and G + 7 anchor to a distance where bottom-strand nicking is constrained, whereas insertions at +8/+9 destroy I-BmoI interactions with G + 7 leading to undetectable levels of cleavage. I-BmoI contains three predicted NUMOD3 minor-groove DNA-binding repeats similar to those that are found in I-TevI and I-HmuI, with the most N-terminal NUMOD3 of the I-BmoI DNA-binding domain encompasses residues G144–S156 (28). We found an ~100-fold difference in the K_D values of N130 and N154 I-BmoI for *thyA* substrate, and protease mapping of 130C I-BmoI performed on the G + 7T mutant substrate led to a higher level of proteolytic digestion than was observed on the wild-type *thyA* substrate. It is, therefore, possible that the N-terminal I-BmoI NUMOD3 repeat interacts with *thyA* substrate at G + 7 to anchor the linker to substrate to facilitate positioning of the nuclease domain.

Given that I-BmoI generates a DSB by sequentially nicking the bottom- and top-strands, the GIY-YIG domain must reposition to the opposite DNA strand after first-strand nicking to hydrolyze the phosphodiester bond that lies in the reverse orientation. The observation that the two nicking reactions differ by only an ~2-fold rate difference suggests that conformational changes during cleavage are not limiting for second-strand nicking. After first-strand hydrolysis, the nicked DNA intermediate could rotate around the top-strand backbone, reducing the spatial reorientation required to move the top-strand scissile phosphate bond into the active site of the catalytic domain. This model is consistent with the fact that I-BmoI can cleave substrates where the bottom- or top-strands have been resected to leave a single-stranded target on the non-resected strand (34).

Homing endonucleases often target functionally critical regions of conserved genes as a strategy to ensure that their target site will be present in naïve genomes (59). For instance, the critical G–C base pair at position −2 of the I-BmoI target corresponds to the second nucleotide position of a conserved arginine codon that is present in all TS genes (31,32). Likewise, the G–C base pair at position +7 that acts as an I-BmoI anchor point corresponds to the first nucleotide position of an aspartate codon that is also highly conserved. The same G–C base pair and aspartate codon is also present in the I-TevI *td* target site, falling within a hypomutable region of the *td* DNA substrate (22). Additionally, ethylation and methylation interference assays of I-TevI:*td* complexes suggested that I-TevI contacts the G of the G–C base pair and phosphate backbone at this position (22). It is, therefore, possible that mutations at this position would also impact I-TevI cleavage efficiency; however, this nucleotide position has not previously been the target of mutagenesis studies.

In summary, we show that I-BmoI generates a DSB via a mechanism that is distinct from other GIY-YIG domain-containing enzymes that oligomerize. The I-BmoI GIY-YIG domain is fused to a C-terminal DNA-binding domain to overcome the inherently weak affinity of the nuclease domain for DNA. The linker that connects the I-BmoI N- and C-terminal domains functions as both an anchor to position the catalytic domain and a flexible tether to permit conformational rotation between the sequential nicking reactions. It is interesting to note that the I-BmoI and I-TevI linkers share significant amino acid similarity, suggesting that the enzymes function via similar cleavage mechanisms. Moreover, the fact that both the I-BmoI and I-TevI substrates share the G + 7 nucleotide in their native substrates becomes an important consideration when targeting chimeric GIY-YIG nucleases for genome-editing applications. In particular, investigating whether I-TevI cleavage efficiency is also influenced by the analogous G + 7 position will help determine whether the proximity of the anchor point to the cleavage site can influence catalytic activity of chimeric GIY-YIG nucleases.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–3, Supplementary Figures 1–9,

Supplementary Methods and Supplementary Reference [60].

ACKNOWLEDGEMENTS

The authors thank David Haniford for discussions regarding the manuscript.

FUNDING

Natural Sciences and Engineering Research Council of Canada Discovery Grant [311610-2010 to D.R.E.]; Alexander Graham Bell Canada Graduate Scholarship from the Natural Sciences and Engineering Research Council of Canada (to B.P.K.); Queen Elizabeth II Graduate Scholarship in Science and Technology from the government of Ontario (to J.M.W.). Funding for open access charge: Natural Sciences and Engineering Research Council of Canada.

Conflict of interest statement. None declared.

REFERENCES

1. Yang,W. (2011) Nucleases: diversity of structure, function and mechanism. *Q. Rev. Biophys.*, **44**, 1–93.
2. Stoddard,B.L. (2005) Homing endonuclease structure and function. *Q. Rev. Biophys.*, **38**, 49–95.
3. Galburt,E.A. and Stoddard,B.L. (2002) Catalytic mechanisms of restriction and homing endonucleases. *Biochemistry*, **41**, 13851–13860.
4. Taylor,G.K. and Stoddard,B.L. (2012) Structural, functional and evolutionary relationships between homing endonucleases and proteins from their host organisms. *Nucleic Acids Res.*, **40**, 5189–5200.
5. Kim,Y.G., Cha,J. and Chandrasegaran,S. (1996) Hybrid restriction enzymes: zinc finger fusions to fok I cleavage domain. *Proc. Natl Acad. Sci. USA*, **93**, 1156–1160.
6. Ashworth,J., Havranek,J.J., Duarte,C.M., Sussman,D., Monnat,R.J. Jr, Stoddard,B.L. and Baker,D. (2006) Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature*, **441**, 656–659.
7. Chan,S.H., Stoddard,B.L. and Xu,S.Y. (2011) Natural and engineered nicking endonucleases—from cleavage mechanism to engineering of strand-specificity. *Nucleic Acids Res.*, **39**, 1–18.
8. Fonfara,I., Curth,U., Pingoud,A. and Wende,W. (2012) Creating highly specific nucleases by fusion of active restriction endonucleases and catalytically inactive homing endonucleases. *Nucleic Acids Res.*, **40**, 847–860.
9. Kleinstiver,B.P., Wolfs,J.M., Kolaczyk,T., Roberts,A.K., Hu,S.X. and Edgell,D.R. (2012) Monomeric site-specific nucleases for genome editing. *Proc. Natl Acad. Sci. USA*, **109**, 8061–8066.
10. Van Roey,P., Meehan,L., Kowalski,J.C., Belfort,M. and Derbyshire,V. (2002) Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-*Tev*I. *Nat. Struct. Biol.*, **9**, 806–811.
11. Truglio,J.J., Rhau,B., Croteau,D.L., Wang,L., Skorvaga,M., Karakas,E., Dellavecchia,M.J., Wang,H., Van Houten,B. and Kisker,C. (2005) Structural insights into the first incision reaction during nucleotide excision repair. *EMBO J.*, **24**, 885–894.
12. Andersson,C.E., Lagerback,P. and Carlson,K. (2010) Structure of bacteriophage T4 endonuclease II mutant E118A, a tetrameric GIY-YIG enzyme. *J. Mol. Biol.*, **397**, 1003–1016.
13. Mak,A.N., Lambert,A.R. and Stoddard,B.L. (2010) Folding, DNA recognition, and function of GIY-YIG endonucleases: crystal structures of R.Eco29kI. *Structure*, **18**, 1321–1331.
14. Sokolowska,M., Czapinska,H. and Bochtler,M. (2010) Hpy188I-DNA pre- and post-cleavage complexes—snapshots of the GIY-YIG nuclease mediated catalysis. *Nucleic Acids Res.*, **39**, 1554–1564.
15. Derbyshire,V., Kowalski,J.C., Dansereau,J.T., Hauer,C.R. and Belfort,M. (1997) Two-domain structure of the *td* intron-encoded endonuclease I-*Tev*I correlates with the two-domain configuration of the homing site. *J. Mol. Biol.*, **265**, 494–506.
16. Pyatkov,K.I., Arkhipova,I.R., Malkova,N.V., Finnegan,D.J. and Evgen'ev,M.B. (2004) Reverse transcriptase and endonuclease activities encoded by penelope-like retroelements. *Proc. Natl Acad. Sci. USA*, **101**, 14719–14724.
17. Dunin-Horkawicz,S., Feder,M. and Bujnicki,J.M. (2006) Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics*, **7**, 98.
18. Lagerback,P., Andersson,E., Malmberg,C. and Carlson,K. (2009) Bacteriophage T4 endonuclease II, a promiscuous GIY-YIG nuclease, binds as a tetramer to two DNA substrates. *Nucleic Acids Res.*, **37**, 6174–6183.
19. Ibryashkina,E.M., Sasnauskas,G., Solonin,A.S., Zakharova,M.V. and Siksnys,V. (2009) Oligomeric structure diversity within the GIY-YIG nuclease family. *J. Mol. Biol.*, **387**, 10–16.
20. Brachner,A., Braun,J., Ghodgaonkar,M., Castor,D., Zlopasa,L., Ehrlich,V., Jiricny,J., Gotzmann,J., Knasmueller,S. and Foisner,R. (2012) The endonuclease Ankle1 requires its LEM and GIY-YIG motifs for DNA cleavage *in vivo*. *J. Cell. Sci.*, **125**, 1048–1057.
21. Kaminska,K.H., Kawai,M., Boniecki,M., Kobayashi,I. and Bujnicki,J.M. (2008) Type II restriction endonuclease R.Hpy188I belongs to the GIY-YIG nuclease superfamily, but exhibits an unusual active site. *BMC Struct. Biol.*, **8**, 48.
22. Bryk,M., Quirk,S.M., Mueller,J.E., Loizos,N., Lawrence,C. and Belfort,M. (1993) The *td* intron endonuclease I-*Tev*I makes extensive sequence-tolerant contacts across the minor groove of its DNA target. *EMBO J.*, **12**, 2141–2149.
23. Edgell,D.R. and Shub,D.A. (2001) Related homing endonucleases I-*Bmo*I and I-*Tev*I use different strategies to cleave homologous recognition sites. *Proc. Natl Acad. Sci. USA*, **98**, 7898–7903.
24. Quirk,S.M., Bell-Pedersen,D. and Belfort,M. (1989) Intron mobility in the T-even phages: high frequency inheritance of group I introns promoted by intron open reading frames. *Cell*, **56**, 455–465.
25. Edgell,D.R. (2009) Selfish DNA: homing endonucleases find a home. *Curr. Biol.*, **19**, 115–117.
26. Sandegren,L. and Sjoberg,B.M. (2004) Distribution, sequence homology, and homing of group I introns among T-even-like bacteriophages—evidence for recent transfer of old introns. *J. Biol. Chem.*, **279**, 22218–22227.
27. Nord,D. and Sjöberg,B. (2008) Unconventional GIY-YIG homing endonuclease encoded in group I introns in closely related strains of the bacillus cereus group. *Nucleic Acids Res.*, **36**, 300–310.
28. Kleinstiver,B.P., Fernandes,A.D., Gloor,G.B. and Edgell,D.R. (2010) A unified genetic, computational and experimental framework identifies functionally relevant residues of the homing endonuclease I-BmoI. *Nucleic Acids Res.*, **38**, 2411–2427.
29. Liu,Q., Dansereau,J.T., Puttamadappa,S.S., Shekhtman,A., Derbyshire,V. and Belfort,M. (2008) Role of the interdomain linker in distance determination for remote cleavage by homing endonuclease I-TevI. *J. Mol. Biol.*, **379**, 1094–1106.
30. Dean,A.B., Stanger,M.J., Dansereau,J.T., Van Roey,P., Derbyshire,V. and Belfort,M. (2002) Zinc finger as distance determinant in the flexible linker of intron endonuclease I-TevI. *Proc. Natl Acad. Sci. USA*, **99**, 8554–8561.
31. Edgell,D.R., Stanger,M.J. and Belfort,M. (2003) Importance of a single base pair for discrimination between intron-containing and intronless alleles by endonuclease I-*Bmo*I. *Curr. Biol.*, **13**, 973–978.
32. Edgell,D.R., Stanger,M.J. and Belfort,M. (2004) Coincidence of cleavage sites of intron endonuclease I-TevI and critical sequences of the host thymidylate synthase gene. *J. Mol. Biol.*, **343**, 1231–1241.
33. Bell-Pedersen,D., Quirk,S.M., Bryk,M. and Belfort,M. (1991) I-*Tev*I, the endonuclease encoded by the mobile *td* intron, recognizes binding and cleavage domains on its DNA target. *Proc. Natl Acad. Sci. USA*, **88**, 7719–7723.
34. Carter,J.M., Friedrich,N.C., Kleinstiver,B. and Edgell,D.R. (2007) Strand-specific contacts and divalent metal ion regulate

double-strand break formation by the GIY-YIG homing endonuclease I-BmoI. *J. Mol. Biol.*, **374**, 306–321.
35. Mueller,J.E., Smith,D., Bryk,M. and Belfort,M. (1995) Intron-encoded endonuclease I-*Tev*I binds as a monomer to effect sequential cleavage via conformational changes in the*td*homing site. *EMBO J.*, **14**, 5724–5735.
36. Kleinstiver,B.P., Berube-Janzen,W., Fernandes,A.D. and Edgell,D.R. (2011) Divalent metal ion differentially regulates the sequential nicking reactions of the GIY-YIG homing endonuclease I-BmoI. *PLoS One*, **6**, e23804.
37. Bitinaite,J., Wah,D.A., Aggarwal,A.K. and Schildkraut,I. (1998) FokI dimerization is required for DNA cleavage. *Proc. Natl Acad. Sci. USA*, **95**, 10570–10575.
38. Vanamee,E.S., Santagata,S. and Aggarwal,A.K. (2001) FokI requires two specific DNA sites for cleavage. *J. Mol. Biol.*, **309**, 69–78.
39. Sasnauskas,G., Connolly,B.A., Halford,S.E. and Siksnys,V. (2007) Site-specific DNA transesterification catalyzed by a restriction enzyme. *Proc. Natl Acad. Sci. USA*, **104**, 2115–2120.
40. Chen,Z. and Zhao,H. (2005) A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic Acids Res.*, **33**, e154.
41. Spink,C.H. (2008) Differential scanning calorimetry. *Methods Cell Biol.*, **84**, 115–141.
42. Liu,Q., Derbyshire,V., Belfort,M. and Edgell,D.R. (2006) Distance determination by GIY-YIG intron endonucleases: discrimination between repression and cleavage functions. *Nucleic Acids Res.*, **34**, 1755–1764.
43. Sitbon,E. and Pietrokovski,S. (2003) New types of conserved sequence domains in DNA-binding regions of homing endonucleases. *Trends Biochem. Sci.*, **28**, 473–477.
44. Johnson,C.M. (2013) Differential scanning calorimetry as a tool for protein folding and stability. *Arch. Biochem. Biophys.*, **531**, 100–109.
45. Kennedy,A.K., Guhathakurta,A., Kleckner,N. and Haniford,D.B. (1998) Tn10 transposition via a DNA hairpin intermediate. *Cell*, **95**, 125–134.
46. Engelman,A., Mizuuchi,K. and Craigie,R. (1991) HIV-1 DNA integration—mechanism of viral-DNA cleavage and DNA strand transfer. *Cell*, **67**, 1211–1221.
47. Mcblane,J.F., Vangent,D.C., Ramsden,D.A., Romeo,C., Cuomo,C.A., Gellert,M. and Oettinger,M.A. (1995) Cleavage at a V(d)j recombination signal requires only Rag1 and Rag2 proteins and occurs in 2 steps. *Cell*, **83**, 387–395.
48. Sasnauskas,G., Halford,S.E. and Siksnys,V. (2003) How the BfiI restriction enzyme uses one active site to cut two DNA strands. *Proc. Natl Acad. Sci. USA*, **100**, 6410–6415.
49. Spassky,A. and Sigman,D.S. (1985) Nuclease activity of 1,10-phenanthroline-copper ion. Conformational analysis and footprinting of the lac operon. *Biochemistry*, **24**, 8050–8056.
50. Li,L., Wu,L.P. and Chandrasegaran,S. (1992) Functional domains in fok I restriction endonuclease. *Proc. Natl Acad. Sci. USA*, **89**, 4275–4279.
51. Soundararajan,M., Chang,Z.Y., Morgan,R.D., Heslop,P. and Connolly,B.A. (2002) DNA binding and recognition by the IIs restriction endonuclease MboII. *J. Biol. Chem.*, **277**, 887–895.
52. Jakubauskas,A., Giedriene,J., Bujnicki,J.M. and Janulaitis,A. (2007) Identification of a single HNH active site in type IIS restriction endonuclease Eco31I. *J. Mol. Biol.*, **370**, 157–169.
53. Jakubauskas,A., Sasnauskas,G., Giedriene,J. and Janulaitis,A. (2008) Domain organization and functional analysis of type IIS restriction endonuclease Eco31I. *Biochemistry*, **47**, 8546–8556.
54. Shen,B.W., Landthaler,M., Shub,D.A. and Stoddard,B.L. (2004) DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J. Mol. Biol.*, **342**, 43–56.
55. Van Roey,P., Waddling,C.A., Fox,K.M., Belfort,M. and Derbyshire,V. (2001) Intertwined structure of the DNA-binding domain of intron endonuclease I-*Tev*I with its substrate. *EMBO J.*, **20**, 3631–3637.
56. Sasnauskas,G., Zakrys,L., Zaremba,M., Cosstick,R., Gaynor,J.W., Halford,S.E. and Siksnys,V. (2010) A novel mechanism for the scission of double-stranded DNA: BfiI cuts both 3′-5′ and 5′-3′ strands by rotating a single active site. *Nucleic Acids Res.*, **38**, 2399–2410.
57. Sasnauskas,G., Kostiuk,G., Tamulaitis,G. and Siksnys,V. (2011) Target site cleavage by the monomeric restriction enzyme BcnI requires translocation to a random DNA sequence and a switch in enzyme orientation. *Nucleic Acids Res.*, **39**, 8844–8856.
58. Kaus-Drobek,M., Czapinska,H., Sokolowska,M., Tamulaitis,G., Szczepanowski,R.H., Urbanke,C., Siksnys,V. and Bochtler,M. (2007) Restriction endonuclease MvaI is a monomer that recognizes its target sequence asymmetrically. *Nucleic Acids Res.*, **35**, 2035–2046.
59. Edgell,D.R., Belfort,M. and Shub,D.A. (2000) Barriers to intron promiscuity in bacteria. *J. Bacteriol.*, **182**, 5281–5289.
60. Crona,M., Moffatt,C., Friedrich,N.C., Hofer,A., Sjoberg,B. and Edgell,D.R. (2011) Assembly of a fragmented ribonucleotide reductase by protein interaction domains derived from a mobile genetic element. *Nucleic Acids Res.*, **39**, 1381–1389.