

# Biochemical and mutagenic analysis of I-Crell reveals distinct but important roles for both the H-N-H and GIY-YIG motifs

Laura E. Corina, Weihua Qiu, Ami Desai and David L. Herrin\*

Section of Molecular Cell and Developmental Biology and Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX 78712, USA

Received January 12, 2009; Revised July 12, 2009; Accepted July 13, 2009

## ABSTRACT

Homing endonucleases typically contain one of four conserved catalytic motifs, and other elements that confer tight DNA binding. I-Crell, which catalyzes homing of the *Cr.psbA4* intron, is unusual in containing two potential catalytic motifs, H-N-H and GIY-YIG. Previously, we showed that cleavage by I-Crell leaves ends (2-nt 3' overhangs) that are characteristic of GIY-YIG endonucleases, yet it has a relaxed metal requirement like H-N-H enzymes. Here we show that I-Crell can bind DNA without an added metal ion, and that it binds as a monomer, akin to GIY-YIG enzymes. Moreover, cleavage of supercoiled DNA, and estimates of strand-specific cleavage rates, suggest that I-Crell uses a sequential cleavage mechanism. Alanine substitution of a number of residues in the GIY-YIG motif, however, did not block cleavage activity, although DNA binding was substantially reduced in several variants. Substitution of conserved histidines in the H-N-H motif resulted in variants that did not promote DNA cleavage, but retained high-affinity DNA binding—thus identifying it as the catalytic motif. Unlike the non-specific H-N-H colicins, however; substitution of the conserved asparagine substantially reduced DNA binding (though not the ability to promote cleavage). These results indicate that, in I-Crell, two catalytic motifs have evolved to play important roles in specific DNA binding. The data also indicate that only the H-N-H motif has retained catalytic ability.

## INTRODUCTION

Intron homing is a unidirectional process in which an intron-minus allele becomes intron-plus (1). Homing of group I introns is catalyzed by an endonuclease, encoded within the invasive intron, which generates a double-strand break (DSB) in the target. Endonucleases similar to the group I intron-encoded proteins are also found as independent genes, a subdomain of group II intron-encoded proteins, and as in-frame insertions in proteins (inteins); some of these elements are also mobile (2).

Homing endonucleases (HEs) have long, asymmetric recognition sequences (14–40 bp), that they can continue to cleave despite multiple substitutions. HEs are usually classified by their catalytic domain, and the vast majority of them have one of the following motifs: LAGLIDADG, GIY-YIG, H-N-H, or His-Cys (2). Looking beyond the primary sequences, however, revealed similarities in the 3D structure of the His-Cys and H-N-H folds, suggesting these enzymes could be related (3). HEs contain additional domains that mediate much of the specific DNA binding.

The largest and best studied family of HEs is LAGLIDADG (2), whereas comparatively fewer His-Cys, H-N-H and GIY-YIG proteins have been examined, and only one member of the latter families has been studied in detail (4–6). The GIY-YIG and H-N-H domains are also found in other types of endonucleases—for example, in certain restriction enzymes (7). Also, the UvrC excinuclease is a GIY-YIG enzyme (8), whereas colicins are non-specific, H-N-H endonucleases (2). The usefulness of H-N-H proteins to biology goes beyond their roles as endonucleases, since an important group of transcription factors in plants has an H-N-H endonuclease in its ancestry (9).

---

\*To whom correspondence should be addressed. Tel/Fax: +1 512 471 3843; Email: herrin@mail.utexas.edu  
Present address:  
Laura E. Corina, Quincy College, Quincy, MA 02169, USA.

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

The GIY-YIG motif is ~85 amino acids (aa). It begins with the consensus GIY and YIG triads, separated by 10–12 aa (2), although an enzyme with only 8 aa separating the triads was recently reported (10). Computational analysis identified four additional, albeit less conserved, sequence elements downstream of the triads, some of which were missing in certain ORFs (11). Pioneering studies with I-TevI showed that it binds double-stranded DNA as a monomer, and cleaves both strands in a sequential fashion, leaving 2-nt 3' overhangs (12). This cleavage pattern is seemingly universal for GIY-YIG enzymes (10,12–15). Mutagenesis and structural analysis helped establish the catalytic ability of the GIY-YIG domain of I-TevI, while also revealing that high-affinity DNA binding is mediated by other modules connected to the GIY-YIG motif by a flexible linker (4,12).

The H-N-H motif is relatively small (35–50 aa) (16), and most of the known H-N-H endonucleases that are not multifunctional proteins bind DNA as a monomer. However, there is evidence for dimerization of colicin E7 in the presence of double-stranded DNA (17), and I-TevIII has recently been shown to be a dimer (18). Cleavage patterns differ significantly among H-N-H endonucleases, e.g. I-HmuI cleaves only one strand (19), whereas I-CmoEI, R.KpnI, and I-TevIII generate a DSB (20–22). There is also variety in the ends left by the latter enzymes: I-CmoEI leaves 4-nt 3' overhangs (20), whereas I-TevIII leaves 5' overhangs (22). X-ray crystal structures of colicins and I-HmuI indicate that the H-N-H motif binds DNA, a divalent metal cation, and encompasses most of the active site (5,23–24).

Holloway *et al.* (25) first suggested that the ORF in the *Cr.psbA4* intron of *Chlamydomonas reinhardtii* contained both H-N-H and GIY-YIG motifs. Subsequently, it was shown that *Cr.psbA4* is efficiently mobile, invading intronless *psbA*, except when the ORF was damaged (26). Kim *et al.* (27) recently established a system for the over-expression and purification of this protein in native form. The enzyme, named I-CreII, generates a DSB in intron-minus, but not intron-plus *psbA* DNA. Also, cleavage by I-CreII leaves 2-nt 3' overhangs similar to GIY-YIG endonucleases, suggesting that it might be the catalytic motif (27).

Holloway *et al.* (25) also suggested that a *psbA* intron in *Chlamydomonas moewusii* was homologous to *Cr.psbA4*, including its ORF, which showed ~58% aa identity to I-CreII. Drouin *et al.* (20) expressed this ORF in *Escherichia coli*, and showed that the protein, I-CmoEI, cleaves intron-minus *psbA* very close to where I-CreII cleaves. However, I-CmoEI leaves 4-nt 3' overhangs, similar to LAGLIDAG HEs. Other functional data, however; especially the relaxed metal requirement, together with the neutralizing effect of substituting a conserved histidine, suggested that I-CmoEI is an H-N-H endonuclease (20).

Thus, the published data suggests that I-CreII and I-CmoEI, despite being homologous enzymes, might be using different motifs for catalysis. This would normally be extraordinary for a HE, but is possible in this case, because of the proposed dual-motif (H-N-H/GIY-YIG) structure (25). To see if this was in fact the case,

we created and characterized a series of variants with an aa substitution in one of the catalytic motifs of I-CreII. As a necessary prelude, we first carried out additional biochemical studies of wild-type (wt) I-CreII.

## MATERIALS AND METHODS

### Site-directed mutagenesis

The alanine-substitution mutants were generated by PCR using the oligonucleotides listed in Supplementary Table 1, and plasmid pI-CreII (27). *Escherichia coli* DH5 $\alpha$  was the host, and the new plasmids were re-sequenced before use. The G220A, I221A, Y222A, G235A and K245A variants were created using megaprimer PCR (28). The megaprimer was synthesized in 50  $\mu$ l-reactions containing 15 ng pI-CreII, 100 pmol of mutagenic and reverse primers, and the *Pfx* polymerase (Invitrogen). Ten microliters of this mixture (~100 ng of megaprimer) was added to the second PCR (50  $\mu$ l), which was similar to the first, except only 3 ng of pI-CreII was used. After asymmetric amplification for five cycles, 100 pmol of oligo 152 was added, and the reaction was cycled 20 $\times$  without an annealing step. The 1-kb product was cut with BspHI and XhoI, and cloned into pET16b as before (27,29). For H147A, an XbaI-AgeI fragment was replaced with a XbaI+AgeI-digested PCR product synthesized with *Vent* polymerase (NEB) and oligos 393 and 396. For N161A, an AgeI-XhoI fragment was replaced with a similarly-digested PCR product synthesized with oligos 394 and 153. Variants H146A, H170A, H174A, E272A, E292A and D299A were created with the QuikChange II mutagenesis kit (Stratagene), and the indicated primers (Supplementary Table 1).

### Expression and purification of I-CreII and variants

I-CreII (including all variants) was produced in an *E. coli* strain that also overexpresses the GroEL/GroES chaperonin as described previously (27). We initially reported that only ~50% of I-CreII was soluble (27), but since then we discovered that with electrophoresis of crude fractions I-CreII comigrates on SDS gels with an abundant membrane protein of *E. coli*. Using longer separating gels, it thus appears that >90% of I-CreII is typically soluble. Purification of the wt protein (for DNA binding and kinetic studies) was as described by Kim *et al.* (27) with some minor modifications detailed below. For purifying the many alanine-substituted mutants, the protocol was shortened by omitting the second chromatography step (on heparin-Sepharose); this change did not significantly affect the activity of the wt enzyme, as it typically yields a preparation that is at least 85% I-CreII, as reported previously (27).

For each variant (and wt) clone, frozen stocks were plated on Luria agar containing 100  $\mu$ g/ml ampicillin, 34  $\mu$ g/ml kanamycin, 34  $\mu$ g/ml chloramphenicol, and after incubation at 37°C overnight, single colonies were inoculated into 2 $\times$  YT medium (10 ml) with the same antibiotics. After incubating for ~3 h (37°C), the cells were harvested by centrifugation (3000  $\times$  g, 15 min), resuspended in 1 l of fresh medium (+ antibiotics), and

split into 4 × 250 ml aliquots for growth (37°C) to mid-log phase ( $A_{600} = 0.7\text{--}1.0$ ). IPTG was added (1 mM) to induce I-CreII, and after 2 h at 30°C, the cells were harvested by centrifugation and stored at  $-75^{\circ}\text{C}$ . I-CreII was purified from the frozen pellets after the cells were disrupted by two passes through a Mini-Bomb (Kontes) at 1500 psi  $\text{N}_2$ . The fractions from the SP-Sephrose and heparin-Sephrose columns were analyzed for activity and by SDS-PAGE. Selected fractions were pooled, dialyzed against 20 mM Tris-HCl pH 7.5, 10% glycerol and aliquotted for storage at  $-75^{\circ}\text{C}$ . Protein concentrations were determined by UV absorbance in 6 M guanidine (for purified I-CreII) (27); with the Bradford assay (which uses coomassie blue binding); and by quantifying dried, coomassie blue-stained SDS gels (30). As with some other proteins, the raw coomassie blue-binding data must be corrected to get accurate estimates of I-CreII, which it underestimates by 40% (27).

### Endonuclease assays and analysis

The standard cleavage assay used plasmid pE4-E5, isolated on CsCl gradients (29) and linearized with ScaI, as substrate (27). The reaction (20  $\mu\text{l}$ ) contained 1 nM I-CreII, 2 nM substrate DNA, 20 mM Tris-HCl pH 8, 10 mM  $\text{MgCl}_2$ , and was incubated at 37°C for 45 min, or as indicated in the text. The reactions were stopped with 0.1 volume of 10× Standard Stop Solution (0.1 M Tris-HCl pH 9, 250 mM EDTA), and analyzed on agarose gels. I-CreII variants were tested for thermal stability as described before (27).

Quantitative kinetic assays were in siliconized tubes (PGC Scientific) using 50  $\mu\text{g}/\text{ml}$  BSA, 20 mM Tris-HCl pH 8, 10 mM  $\text{MgCl}_2$ , 0.5 mM EDTA as buffer; the reactions were stopped with 0.1 volume of 10× Stop Solution 2 (500  $\mu\text{g}/\text{ml}$  proteinase K, 5% SDS, 250 mM EDTA pH 8, 30% glycerol, 0.125% bromphenol blue). To assay cleavage with pre-bound substrate, the reactions were pre-incubated without  $\text{MgCl}_2$  for 15 min at 37°C (or 23°C), and then started by adding  $\text{MgCl}_2$  to 10 mM. Electrophoresis was in 1% agarose gels, which were stained with ethidium bromide, and imaged with a digital scientific camera (DC120, Kodak). The relative amounts of DNAs on the gels were quantified with 1D Image Analysis (Kodak, version 3.6) provided the signals were within the linear range of the DC120. To minimize loading differences, the values for each DNA band were expressed as a fraction of the total DNA in each lane. The data was plotted using curve fitting software (Kaleidagraph, version 3), and  $k_{\text{obs}}$  was determined from substrate-decay curves using the first-order rate equation,  $A = A_0 e^{-kt}$  (where  $k = k_{\text{obs}}$ ), except where indicated.

To determine strand-specific cleavage rates, a small (134-bp) substrate DNA (E4-E5<sub>134</sub>) was synthesized by PCR using plasmid pE4-E5 (27) and oligos 99 and 100 (either one or both was 5' end-labeled with [ $\gamma$   $^{32}\text{P}$ ]-ATP). The product was purified on silica (Qiagen) and eluted in 10 mM Tris-HCl pH 8. Cleavage reactions (in siliconized tubes at 37°C) contained 30 nM end-labeled E4-E5<sub>134</sub> DNA, 230 nM I-CreII, 10 mM  $\text{MgCl}_2$ , 20 mM Tris-HCl pH 8, and aliquots were removed at selected intervals and

stopped with 1.2 vol of 95% formamide, 20 mM EDTA pH 8, 0.05% xylene cyanol, 0.05% bromphenol blue. Denaturing polyacrylamide (6%) gels (29) were run at 50°C, and exposed to X-ray film (Biomax MS, Kodak) without a screen. Multiple exposures were made, and the films quantified using 1D Image Analysis as above; log-linear plots were also made with Kaleidagraph.

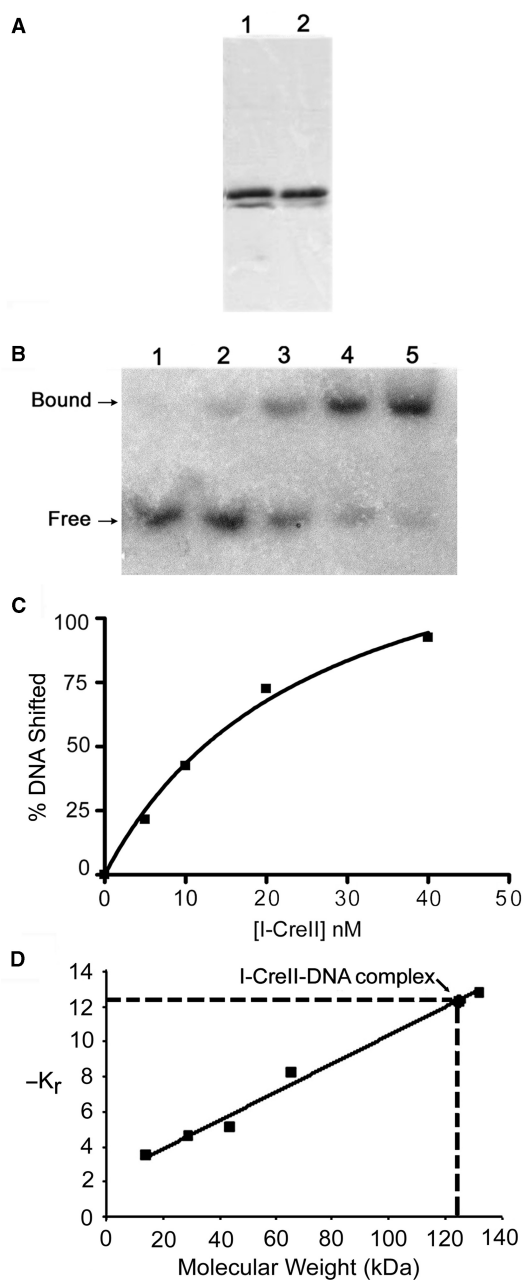
### Electrophoretic mobility shift assay

Binding was performed at 23°C (50  $\mu\text{l}$  volume) using a range of I-CreII concentrations and either 10 or 20 nM  $^{32}\text{P}$ -labeled E4-E5<sub>134</sub> DNA in 20 mM Tris-HCl pH 8, 1 mM DTT, 0.1 mM EDTA, 30  $\mu\text{g}/\text{ml}$  polydeoxyinosinic-deoxycytidylic acid (polydIdC). After 20 min, 0.2 volume of 30% glycerol, 0.125% bromphenol blue was added, and the samples separated on native polyacrylamide gels (room temperature) buffered with 0.5× TBE (29). To estimate  $K_d$ , which is approximately equal to the concentration of I-CreII that shifts 50% of the target DNA, the gels were exposed to X-ray film (Kodak BioMax, MS) and quantified as described above. The data was fit to a one-site saturation-binding curve (GraphPad Prism, version 5).

The size of the I-CreII-DNA complex was determined using a modified Ferguson analysis (31). Electrophoretic mobility shift array (EMSA) reactions that shifted ~50% of the DNA were separated on native gels of 6, 7, 8, 9 and 10% polyacrylamide along with native protein standards (Sigma). The mobility ( $R_f$ ) values of the standards and the I-CreII-DNA complex were plotted against the polyacrylamide concentration, and the slopes of these lines ( $K_f$ ) were plotted against the molecular weights of the standards (20).

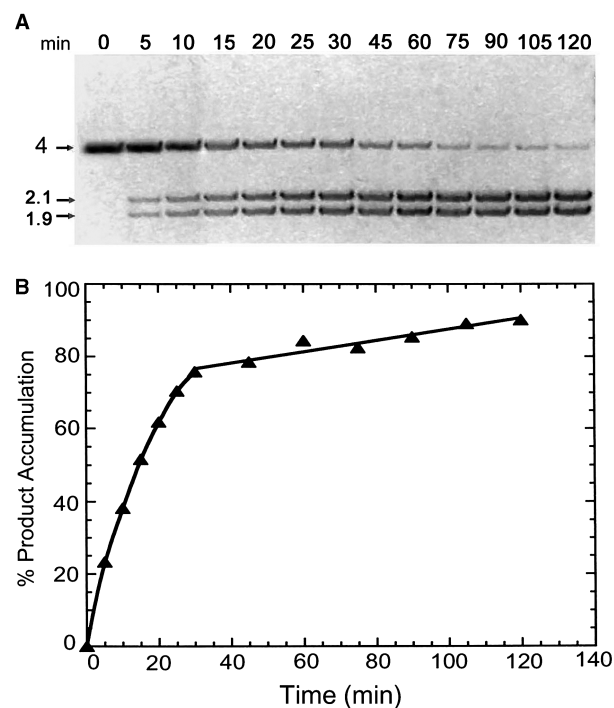
## RESULTS

I-CreII was produced in *E. coli* in native form (27)—a system we used previously to avoid possible artifactual effects of non-native amino acids on protein folding. The host cells also over-produce the chaperonin, GroEL/GroES, which is homologous to the chloroplast protein and was necessary for proper folding of I-CreII in *E. coli* (27). Purification takes advantage of the fact that few *E. coli* proteins are as basic as I-CreII ( $pI = 10.0$ ), and of its affinity for nucleic acids. Figure 1A shows the protein after the ion-exchange chromatography step (lane 1), and after an additional separation on a heparin-affinity column (lane 2). The enzyme, even when highly purified, frequently appeared as a doublet on SDS gels, with the ratio of the two bands being somewhat variable (the upper band is much more prominent in Figure 1A). N-terminal sequencing of both bands, sliced from an SDS gel, showed they have the same N-terminus—that of I-CreII lacking the N-terminal methionine (not shown). Mass spectrometry of a mixture of the two proteins gave two peaks that differed by 80 Daltons (not shown), suggesting the enzyme was mono-phosphorylated—although the presumptive site is unknown. As we did not see evidence of heterogeneity in the kinetic or binding data, we judged the apparent modification to be insignificant. Also, since the biochemical parameters, determined below (Figures 1–4),



**Figure 1.** EMSA and size determination of the I-CreII-DNA complex. (A) SDS gel of I-CreII purified from *E. coli*. Lane 1 was after the SP-sepharose (cation-exchange) chromatography step, and lane 2 was after an additional step through heparin-sepharose. (B) Autoradiograph of a representative EMSA run on an 8% gel. E4-E5<sub>134</sub> DNA (20 nM) was incubated with increasing concentrations (0, 5, 10, 20, 40 nM) of I-CreII, and then separated on a native polyacrylamide gel. B, bound DNA (I-CreII-DNA complex); F, free DNA. (C) Plot of the fraction (in %) of shifted (E4-E5<sub>134</sub>) DNA versus I-CreII concentration for the gel in (B). (D) Ferguson analysis of the I-CreII-DNA complex. The retardation coefficient for each standard protein ( $-K_r$ ) was plotted against its molecular weight, and the  $K_r$  for the I-CreII-DNA complex corresponded to a mass of 124 kDa. This analysis was repeated more than three times with similar results.

were indistinguishable for the enzyme purified by either protocol, the shorter one was used for analysis of the numerous mutants (Figures 6–9). First, however, the characterization of wt I-CreII is presented using data obtained



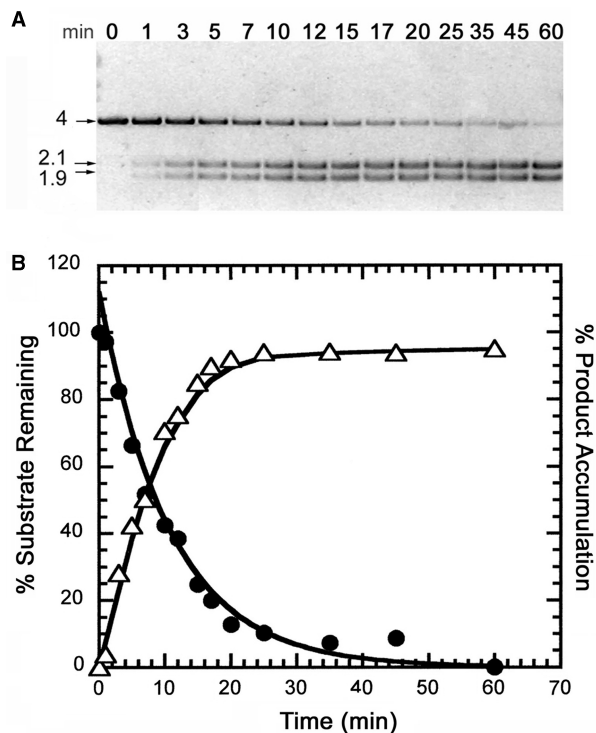
**Figure 2.** Kinetics of DNA cleavage with excess substrate. (A) Agarose gel of a time-course reaction with excess substrate (35 nM pE4-E5 DNA, 25 nM I-CreII). The ethidium-stained gel was digitally imaged, and then inverted. The sizes of the substrate (4) and cleavage products (2.1 and 1.9) are indicated to the left in kb. (B) Plot of product accumulation. The 2.1-kb product in (A) was quantified and plotted against reaction time using the curve-fitting function in Kaleidagraph (version 3.0). The line for 0–30 min is from an exponential curve-fit, and the line for 30–120 min is from a linear curve-fit. The experiment was repeated two more times with similar results.

with the enzyme purified through both chromatographic steps (Figures 1–4).

#### DNA binding by wt I-CreII

Attempts to determine the native size of I-CreII by gel filtration chromatography, low angle X-ray scattering and native gel electrophoresis were unsuccessful due to aggregation of the protein. However, as shown here, I-CreII can bind target DNA in the absence of a divalent cation, and by determining the size of the DNA-protein complex, we could infer the stoichiometry of DNA binding. Figure 1B shows a representative EMSA with the E4-E5<sub>134</sub> target DNA: there is a discrete shift of the DNA to a slower mobility (Figure 1B), and the fraction shifted is proportional to the I-CreII concentration (Figure 1C). Lane 4 (Figure 1B) is additionally informative, because it contained equimolar amounts of I-CreII and target DNA; the fact that nearly all the DNA is shifted indicates that the vast majority of the protein is active. The plot in Figure 1C gave an apparent  $K_d$  of  $\sim 10$  nM.

Ferguson analysis (31) of the DNA-protein complex yielded a mass of 124 kDa (Figure 1D). Since the DNA is 84 kDa, the protein component is inferred to be 40 kDa, which is close to the size of I-CreII predicted from the

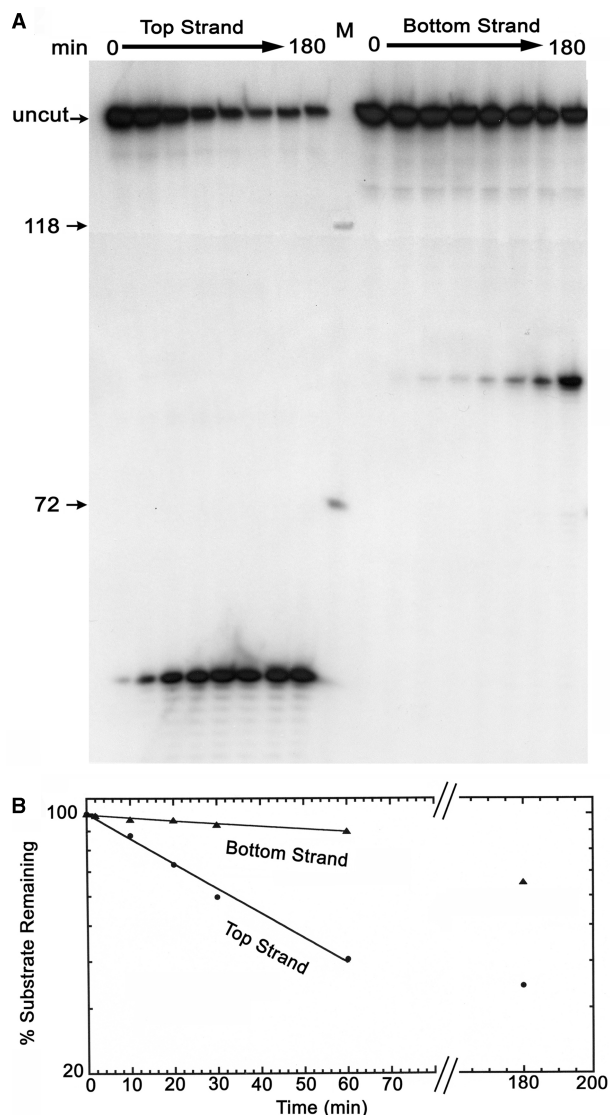


**Figure 3.** Kinetics of single-turnover cleavage with pre-bound substrate. (A) Agarose gel of a reaction with pre-bound substrate. The substrate DNA (13.5 nM) was pre-incubated with a 7-fold molar excess of I-CreII. The reaction was started by adding  $MgCl_2$ , and aliquots were removed at the indicated times. The gel image was inverted. (B) Plot of the time-course reaction in (A). The substrate (4 kb) and 2.1-kb product DNAs were quantified, and used to plot substrate-remaining (filled circles) and product-accumulation (open triangles) curves, respectively. The substrate decay line has an  $R^2$  value of 0.996.

DNA sequence, 37.9 kDa. Thus, the EMSA data indicates that I-CreII binds to target DNA as a monomer.

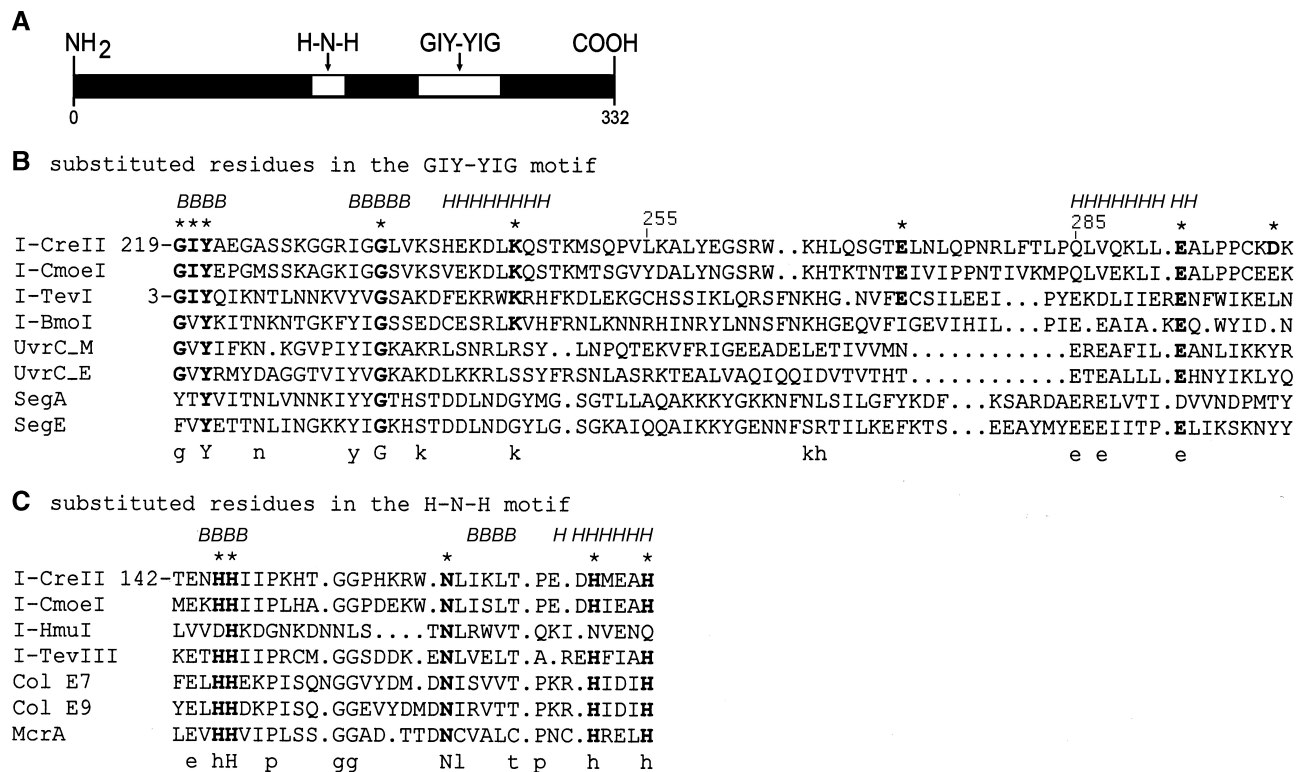
### Kinetics of DNA cleavage

The addition of  $Mg^{2+}$  to the EMSA assay gave cleavage products of the expected size (46 and 88 bp) and ratio (not shown), indicating the protein does not hold onto either of its products as tightly as it does the substrate. Thus, it seemed likely that I-CreII, unlike I-CreI (32), could perform turnover cleavage, a finding that could also have implications for why this intron homes more efficiently than *psbA* intron 2 (J.Lee, N.Deshpande, and D.L.Herrin, manuscript in preparation). To verify this feature, and determine if product release limits turnover cleavage, we returned to the linearized plasmid-based assay described previously (27). The plasmid, pE4-E5, contains the same 134-bp fragment of the intronless *psbA* gene that was used for the EMSA (above), cloned into the pCR2.1 vector. A representative time-course reaction with a 40% excess of plasmid substrate is shown in Figure 2A. The progress plot (Figure 2B) shows two distinct phases: the first, or fast, phase ( $k_{obs} = 0.09 \text{ min}^{-1}$ ) lasted for ~30 min and represents the first cleavage event, whereas the second phase lasted until the end of the experiment (another 90 min), and was



**Figure 4.** Strand-specific cleavage kinetics. (A) Autoradiographs of cleavage reactions for each strand of the E4-E5<sub>134</sub> DNA. Single-end-labeled DNAs (30 nM) were cleaved with excess I-CreII (230 nM); aliquots were removed at selected intervals and separated on a 6% polyacrylamide/urea gel. HaeIII-digested  $\phi$ X174 DNA was used as size markers (lane M), and the sizes of two fragments are indicated (in nt). Undigested E4-E5<sub>134</sub> DNA is also indicated (uncut). (B) Log-linear plots of the time-course reactions in (A); the  $R^2$  values are 0.997 for the top strand, and 0.983 for the bottom strand.

~60-fold slower ( $k_{obs} = 0.0015 \text{ min}^{-1}$ ). Although the [substrate] exceeded the [enzyme] by only 40% in this experiment, the slow phase of the plot is 3-fold longer than the fast phase; also, longer incubation times, which could have allowed for more substrate to be used, were not useful due to increased enzyme inactivation (not shown). The biphasic rate profile suggests that the slow step occurs after cleavage chemistry, most likely product release. It is also noteworthy that the fraction of DNA cleaved during the fast phase (~75%) is similar to the molar ratio of I-CreII to substrate (71%), thus supporting the assertion that I-CreII cleaves as a monomer.



**Figure 5.** Sequence alignments and aa residues selected for substitution. (A) Pictorial diagram of I-CreII showing the relative locations of the H-N-H and GIY-YIG motifs. (B) Alignment of GIY-YIG motifs: the substituted residues are in bold letters with overlying asterisks. Residues in I-CreII that are in predicted secondary structures, consistent with the I-TevI structure, are indicated in italics—*B* is for beta strand and *H* for alpha helix. The sequences include intron-encoded endonucleases [I-CreII (gi:21675095), I-CmoE (gi:12486), I-TevI (gi:3033368), I-BmoI (gi:12958590)], free-standing HEs [SegA (gi:5354447), SegE (gi:9632827)], and excision endonucleases [UvrC\_M (gi:8134799), UvrC\_E (gi:8134799)]. (C) Alignment of H-N-H motifs: the substituted residues and predicted secondary structures are marked as for the GIY-YIG motif (B). The selected sequences include intron-encoded endonucleases [I-HmuI (gi:465641), I-TevIII (gi:579158)], colicins [E7(gi:510385), E9 (gi:1418695)], and a restriction enzyme, McrA (gi:146794). Below the alignments are consensus residues that are in lower case if present in 50% or more of the proteins, but in upper case if 100% conserved. Initial alignments were with Clustal X, but final alignments were adjusted manually; gaps (dots) were used to maximize similarity. The number of the aa that precedes each motif is indicated for I-CreII and I-TevI; selected residues in the GIY-YIG of I-CreII are numbered above the aa.

To estimate the rate of cleavage chemistry, time-course reactions were performed with all of the substrate pre-bound to the enzyme (to remove any effects of DNA binding on the observed reaction rate). This was accomplished by pre-incubating the substrate with a molar excess of I-CreII, and then initiating the reaction with Mg<sup>2+</sup>; a representative experiment (and plot) is shown in Figure 3. The decrease in substrate with time is a first-order reaction, which reproducibly gave a rate constant (*k*<sub>obs</sub>) of 0.1 ± 0.01 min<sup>-1</sup>. This rate is quite close to the fast phase in the experiment with excess substrate (Figure 2), indicating that DNA binding under those conditions is rapid and probably not limiting.

**Strand-specific cleavage kinetics**

Time-course cleavage reactions with supercoiled DNA indicated that most of it is not converted directly to linear DNA, suggestive of a sequential cleavage mechanism (33, and see below). Thus, we examined the rate of cleavage of each strand using single-end-labeled E4-E5<sub>134</sub> DNA and single-turnover conditions (i.e. excess enzyme, but not pre-bound to substrate). The data in Figure 4

show that the top strand is cleaved 10-fold faster than the bottom strand (*k*<sub>obs</sub> = 0.031 min<sup>-1</sup> for the top strand, and 0.003 min<sup>-1</sup> for the bottom strand). Similar results were obtained with DNA labeled at both ends (not shown). One implication of this data is that the rate of DSB formation (e.g. Figure 3) reflects the rate of cleavage of the antisense (bottom) strand, because it is the slower cleavage event. The results are also consistent with a sequential cleavage mechanism for I-CreII that begins with top-strand cleavage. Finally, we note that the observed rate for cleaving E4-E5<sub>134</sub> is considerably (~25-fold) slower than the rate for cleaving the pE4-E5 plasmid. The reason for this is not clear at present, but it may not simply be an effect of the shorter length of the radio-labeled PCR product, since binding to this DNA is efficient—based on the EMSA—and binding would be the step most affected by DNA length.

**Alanine substitution in the GIY-YIG and H-N-H motifs of I-CreII**

To assess the roles of the GIY-YIG and H-N-H motifs in I-CreII (Figure 5A), alanine substitution was used to

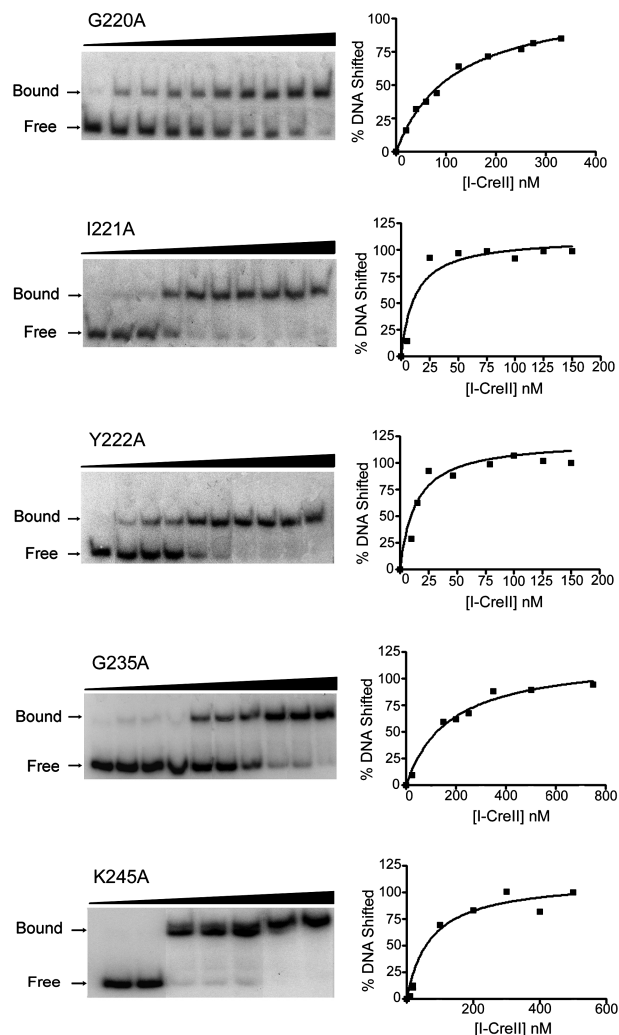
minimize effects on protein structure. The poor conservation of the GIY-YIG motif, and the paucity of structures made structural modeling of this region (by using I-TevI for instance) difficult, so we relied primarily on sequence alignments, and to a lesser extent, secondary structure predictions. The secondary structures indicated above the alignment in Figure 5B were predicted for I-CreII and generally agreed with I-TevI; predictions of the region between the second beta strand and the alpha helix did not agree with the I-TevI structure and were omitted for simplicity's sake. The aa in I-CreII that were substituted are indicated above the alignment with asterisks. Initially, the most conserved aa of the triads, G220, Y222 and G235, were substituted; the tyrosine of the second triad is already changed in wt I-CreII. K245 was substituted, because we thought it might be equivalent to R27 in I-TevI (11); and I221 was targeted to provide a non-conserved change. We also substituted residues E272, E292 and D299 in an attempt to obtain a catalytic mutant analogous to the E75 mutant of I-TevI (4,11).

Although the smaller H-N-H motif in I-CreII could be modeled using colicins, we also relied heavily on sequence alignments, and the previous mutagenesis of colicin E9 (34) to select residues for substitution. Figure 5C shows an alignment of H-N-H motifs from selected endonucleases, and the aa that were changed to alanine (H146, H147, H170, H174, N161) are indicated by asterisks. We note that I-Hmul is an example of the H-N-N subset of H-N-H proteins, whereas the H-N-H motif of I-CreII fits best into subset 1 of Mehta *et al.* (16).

### DNA binding by the I-CreII variants

The alanine-substituted variants (and wt) were purified through the ion-exchange chromatography step, which yields a preparation that is at least 85% I-CreII (27) and whose activity is indistinguishable from the homogeneous preparation used for Figures 1–4; an SDS gel of representative variants was included in the Supplementary Data (Supplementary Figure 1). The DNA-binding ability of the variants was estimated with the EMSA. Figure 6 shows results for the GIY-YIG mutants: the I221A and Y222A variants were similar to wt ( $K_d$ , 5–10 nM), but for the G220A, G235A and K245A variants, DNA binding was reduced 10- to 25-fold ( $K_d$ , ~100, ~150 and ~65 nM, respectively). These data suggest that the GIY-YIG motif contributes significantly to specific DNA binding by I-CreII.

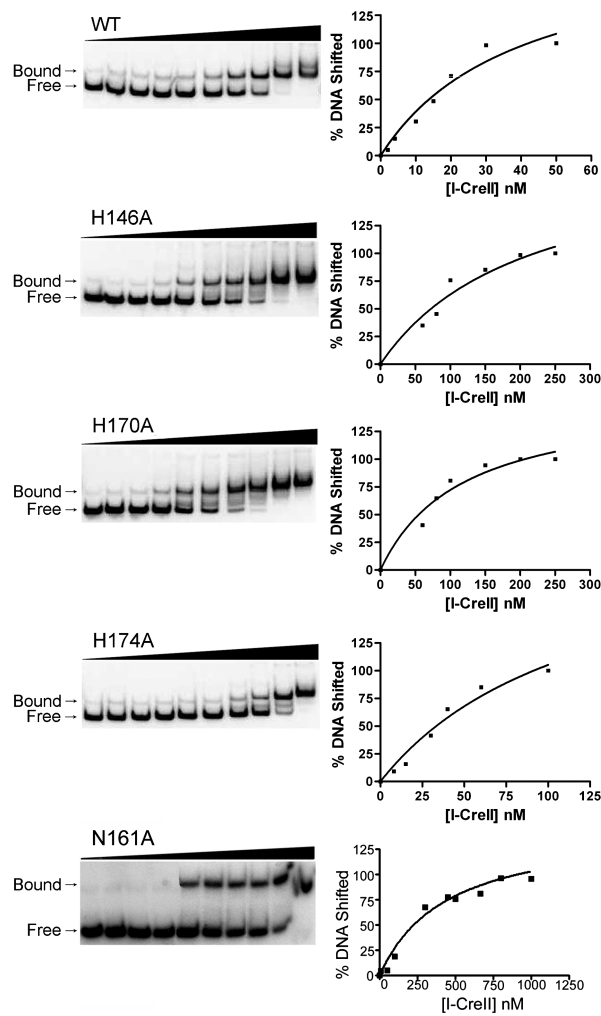
EMSA of the H-N-H mutants (Figure 7) show that substituting the conserved histidines H146, H170 and H174 produced a modest (~3- to 4-fold) decline in DNA binding that was similar for all three variants. Substituting the asparagine (N161A variant), however, reduced DNA binding at least 25-fold. In fact, this variant had the worst apparent  $K_d$  (~250 nM) among the proteins that were quantified (the H147A variant could not shift DNA at all, but its structure may have been compromised, based on the fact that it was less soluble and bound poorly to the cation-exchange column). From this data, we conclude that the H-N-H motif also contributes substantially to specific DNA binding by I-CreII.



**Figure 6.** EMSAs of the GIY-YIG variants.  $^{32}\text{P}$ -labeled E4-E5<sub>1,34</sub> (20 nM) was incubated with increasing concentrations of the indicated variants, and then separated on native polyacrylamide gels. On the left are autoradiographs of native gels, and on the right are the corresponding plots. Bound, protein–DNA complex; Free, free DNA.

### Catalytic activity of the GIY-YIG motif variants

Information gleaned from the EMSAs allowed us to examine the kinetics of DNA cleavage with pre-bound substrate for most variants. Time-course reactions for the GIY-YIG motif variants, G220A, I221A, Y222A, G235A and K245A are presented in Figure 8. Surprisingly, the cleavage efficiency of all of these proteins was nearly indistinguishable from wt, except for Y222A, whose rate was 30% of wt ( $k_{\text{obs}} = 0.03 \pm 0.01 \text{ min}^{-1}$  compared to  $0.1 \pm 0.01 \text{ min}^{-1}$  for wt). It should be said that the stabilities of these proteins under cleavage conditions (37°C for 1 h) were similar to wt, except G235A, which lost activity in 30 min (Supplementary Figure 2). G235A is stable at 23°C, however (Supplementary Figure 2); so it was assayed at that temperature. Although the endonuclease activity of the E272A, E292A and D299A variants was assessed using a cruder fraction [after the ammonium sulfate fractionation

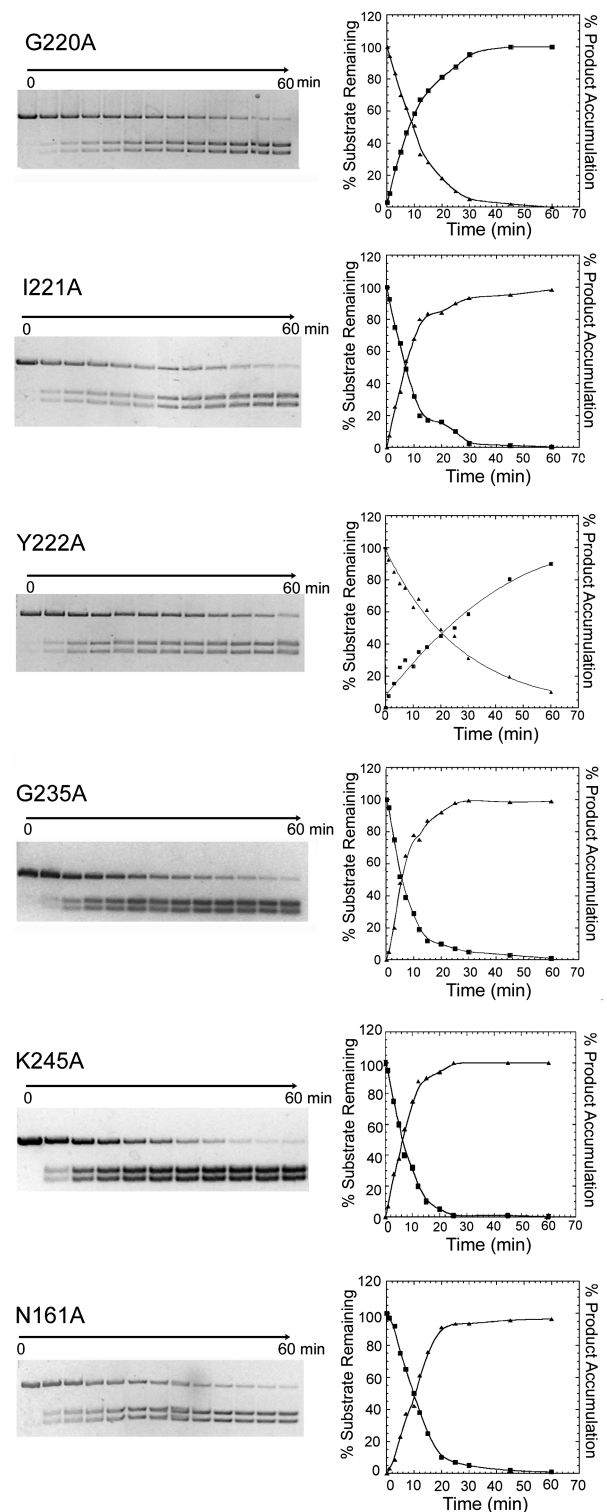


**Figure 7.** EMSAs of the H-N-H mutants. <sup>32</sup>P-labeled E4-E5<sub>134</sub> DNA (20 nM) was incubated with increasing concentrations (0–1000 nM) of the indicated I-CreII variant, and then separated on native polyacrylamide gels. On the left are autoradiographs of representative gels, and their corresponding plots are on the right.

step (27)], these proteins were clearly quite competent in DNA cleavage (not shown), and were not investigated further. Together, these data suggest that the GIY-YIG motif in I-CreII is not catalytic, although it may have been in the evolutionary past.

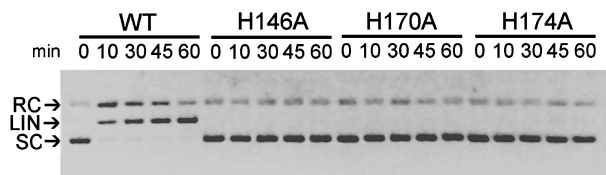
**H-N-H motif variants with and without catalytic activity**

Among the H-N-H variants, only N161A exhibited significant catalytic activity, and its cleavage kinetics were indistinguishable from wt (Figure 8). In contrast, substitution of any of the four conserved histidines severely affected DNA cleavage. As shown by the digestions with supercoiled DNA (Figure 9), the H146A, H170A and H174A variants were also defective in nicking (cleaving only one strand) as well as DSB formation. The H146A variant did exhibit a small amount of cleavage (~1% of wt), whereas no cleavage was detected for the H170A and H174A variants. Finally, the H147A variant was inactive in DNA cleavage, but also in DNA binding (see above).



**Figure 8.** Mutant cleavage kinetics with pre-bound substrate. The proteins were pre-incubated with substrate DNA (pE4-E5) as follows: for those with a  $K_d$  similar to wt, an ~7-fold molar excess of enzyme was added. For the variants with a higher  $K_d$ , this was impractical, so I-CreII was added to  $\sim 2.5 \times K_d$  to pre-bind all the substrate. The reactions were performed as in Figure 3, except for G235A, which had to be incubated at 23°C. Agarose gels of time-course reactions with the indicated variants are to the left, and the corresponding substrate-remaining (filled circles) and product-accumulation (triangles) plots are on the right. The plot lines have  $R^2$  values of 0.969–0.995.





**Figure 9.** Cleavage assays with H-N-H variants and supercoiled substrate. Standard cleavage conditions were used, except the plasmid (pE4-E5) was not linearized and >90% was in supercoiled form. The DNA concentration was 2nM, and the I-CreII concentrations were: 1nM for WT, 4nM for H146A, 4nM for H170A, and 2.5nM for H174A. The ethidium-stained gel was imaged, and the image inverted (to black-on-white). RC, relaxed circles; LIN, linear; SC, supercoiled.

The time-course digestion of supercoiled DNA with the wt enzyme (WT lanes in Figure 9) indicates that supercoiled (SC) DNA is converted to relaxed circles (RC), and then to linear DNA (LIN), consistent with a sequential mechanism.

## DISCUSSION

### I-CreII cleavage mechanism

The EMSA demonstrated that tight DNA binding by I-CreII ( $K_d$ , 5–10 nM) does not require free divalent cations, an ability that is shared by many other endonucleases (e.g. 11,32,35–37), but interestingly, not by I-CmoI (20). This could be an underestimate of the *in vivo* affinity of I-CreII for DNA, however; based on the fact that DNA binding by I-CreI increases with added  $Ca^{2+}$  (38). A similar analysis could not be done with wt I-CreII, because it can use  $Ca^{2+}$  for DNA cleavage (27). The I-CreII–DNA complex has a 1:1 stoichiometry of enzyme and DNA, indicating that I-CreII binds its target as a monomer. The kinetics of DNA cleavage with excess substrate, and the lack of symmetry in the native target sequence (27) are also consistent with I-CreII functioning as a monomer. Monomeric enzymes that generate DSBs are relatively uncommon, but include I-CmoI and several phage GIY-YIG endonucleases (12,20,36). There is also evidence that the H-N-H colicin, E7, may act as a monomer or a dimer, depending on the substrate (17).

Many HEs are poor at cleaving multiple substrate molecules (20,32,39–41), usually because of slow product release. The biphasic cleavage-rate profile for I-CreII suggests that its turnover cleavage is also limited by product release. Nonetheless, the ability of I-CreII to carry out more than one cycle of cleavage could help explain why homing of *Cr.psbA4* is so efficient (26), and especially why it is more efficient at homing than *Cr.psbA2* (J. Lee, N. Deshpande and D.L. Herrin, manuscript in preparation). Also, it has not escaped our notice that multiple-turnover HEs could have a distinct advantage over one-shot HEs (e.g. I-CreI) in environments like chloroplasts or mitochondria, where there can be 50–100 genome copies in one organelle. This hypothesis could be tested if a variant of I-CreII was available that was

incapable of performing turnover cleavage, but was otherwise similar to the wt enzyme.

The strand-specific cleavage rates estimated with the PCR substrate indicate that I-CreII cleaves the top strand 10-fold faster than it cleaves the bottom strand, suggestive of a sequential cleaving mechanism. The kinetic profile for cleaving supercoiled plasmid DNA is also consistent with a sequential mechanism. It is not yet known if this is an obligatory order of events, or if it is mainly a kinetic phenomenon, but these models should be testable (e.g. by using phosphorothioate substitution in the top strand to block the first cleavage). Sequential cleavage is an uncommon mechanism for enzymes that cleave both DNA strands, but it is also likely used by I-TevI and I-TevII, which cleave the bottom strand first (12,36), and by the restriction enzyme, BfiI (42).

### A role for the GIY-YIG motif in specific DNA binding

Alanine substitution of several residues in the GIY-YIG motif had substantial effects on DNA affinity, but little effect on catalysis. The Y222A variant did show a modest (3-fold) decrease in cleavage rate, but not in DNA binding, suggesting that this residue might play a role in catalysis. However, the corresponding mutation in I-TevI (Y6A) abolished DNA cleavage, though there was some evidence that the protein might have been structurally compromised (11). Substitution of the conserved glycines of the triads (G220 and G235, respectively) substantially reduced DNA binding by I-CreII, without affecting catalysis. In contrast, a mutation in I-TevI (G19A) similar to G235A resulted in a loss of cleavage activity, but not DNA binding (11). It could be said that this highly conserved glycine has reversed roles in I-CreII.

The crystal structure of the I-TevI domain shows that the two triads are in distinct  $\beta$ -strands that, together with a third-strand downstream, form a  $\beta$ -sheet that could interact with DNA (4). Although the I-TevI structure may help explain the results with the G220A and G235A variants of I-CreII, high-affinity DNA binding by I-TevI is mediated mainly by downstream regions (43). So, one should probably expect to see some structural changes in the GIY-YIG motif of I-CreII, compared to I-TevI. Along this line, secondary structure predictions hint at additional  $\beta$ -strands in the region between L255 and Q285 of I-CreII (Figure 5B) that are not present in I-TevI (unpublished results). Thus, solving the structure of this domain from I-CreII could provide insight into its evolution from a catalytic motif.

Substitution of K245, which we thought might have been analogous to the catalytically important R27 residue of I-TevI (11), decreased DNA binding, but not catalysis. Also, alanine substitution of the downstream acidic residues (E272A, E292A, D299A) had little effect on DNA cleavage. This region was expected to contain a critical, metal-binding residue equivalent to E75 in I-TevI (4,11), but that residue is apparently missing, or is in the wrong structural context, to act catalytically in I-CreII. On the other hand, it seems clear that the GIY-YIG motif of I-CreII has evolved an increased affinity for target DNA compared to the I-TevI domain (4), and that

these evolutionary changes have impacted the roles of the most conserved aa residues.

Like I-TevI, the H-N-H HEs that are highly specific, such as I-HmuI (5), contain additional DNA-binding domains that are conserved among different HEs (e.g. two domains in I-HmuI are similar to two in I-TevI) (44). These domains are not found in I-CreII, which is consistent with the notion that the GIY-YIG motif in I-CreII is performing part of this function. Bioinformatic analysis also supports this view, inasmuch as several programs that predict DNA-binding sites from sequence data (including PredDNA and BindN) identify this region of I-CreII as having DNA-binding potential. Interestingly, they also predict that residues N-terminally adjacent to N161, whose substitution strongly decreased DNA binding (see below), and some located far upstream of the H-N-H motif also bind DNA (unpublished results).

### Catalytic and DNA-binding roles for the H-N-H motif

The H-N-H active site is believed to use a metal cation and a conserved histidine, which acts as a general base, to promote attack at the scissile phosphate by an activated hydroxyl (5,45). In colicin E9, the catalytic metal is coordinated to two of the conserved histidines (34,45), and substituting the equivalent histidines in I-CreII resulted in proteins with very little (H146A) or no (H170A) cleaving activity, just as with the colicin. Substituting H174 in I-CreII nearly eliminated DNA cleavage, although substituting the equivalent residue in colicin E9 (H131) and colicin E7 (H573), where it also coordinates the metal (46), only reduced it (34,46). This might indicate a difference in active-site structure, or bound metal species, or both, in I-CreII (preliminary metal analysis of wt I-CreII suggests that it has both  $Zn^{2+}$  and  $Mg^{2+}$  species; L.C. Corina and D.L. Herrin unpublished results). Residue H147 in I-CreII should act as the general base, so it was unexpected that the H147A variant would be compromised in folding and/or structure; however, this could indicate a dual role for this residue. In summary, the fact that three of the histidine variants (H146A, H170A, H174A) shifted target DNA and had only modest (~3-fold) reductions in DNA binding, would indicate they are catalytic mutants, and as such, provide strong evidence for the H-N-H motif mediating cleavage of both strands.

The aforementioned results were somewhat unexpected, because the ends left by I-CreII cleavage are the same as GIY-YIG endonucleases. However, H-N-H endonucleases leave a variety of ends (45)—although this would be the first report, to our knowledge, of an H-N-H endonuclease leaving 2-nt 3' overhangs.

The conserved-asparagine variant, N161A, exhibited normal catalysis, but strongly reduced DNA binding. This result was surprising, since this residue is purported to be mainly structural (47). In colicin E7, it acts remotely to orient the general base histidine (H545) for efficient DNA cleavage, and substituting it with alanine did not decrease DNA binding (47). Although we did not see evidence of structural destabilization with N161A,

at least not like the thermal instability of G235A, a highly localized perturbation may not have been manifest. An alternative, and more likely explanation in our view, is that evolutionary adaptation of the H-N-H motif in I-CreII has resulted in an important role for this residue in DNA binding. An increase (or enhancement) in aa that interact with the DNA target makes sense in this case, because, unlike colicins, I-CreII is a highly specific endonuclease.

### Occurrence and evolution of dual-motif (H-N-H/GIY-YIG) HEs

One question is whether there are other dual-motif proteins like I-CreII (besides I-CmoeI). Since *Cr.psbA4* is a relatively old intron (25,48), one might expect to find proteins related to I-CreII in other green algal chloroplast genomes. And indeed, there are viable candidates in the chloroplast genomes of *Stigeoclonium helveticum* (gb, DQ630521.1), *Pseudoclonium akinetum* (gb, AY835431.1), and *Oedogonium cardiacum* (gb, EU677193.1) (49). In each case, the ORF is encoded in a group I intron of the *psbA* gene, but only in *Oedogonium* is the intron inserted at the same location as *Cr.psbA4* (unpublished results). Intron-encoded ORFs that resemble degenerate versions of an H-N-H/GIY-YIG endonuclease are also resident in those genomes. Since the number of green-algal chloroplast genomes that have been sequenced is still quite limited, the potential for finding more enzymes like I-CreII is very high. It might also be appropriate to consider the H-N-H/GIY-YIG enzymes as a distinct class of HEs.

The origin of I-CreII is not clear, but a plausible mechanism could have been an invasion of one single-motif HE by another, followed by retention of both motifs in the new protein. There is an example, demonstrated experimentally, of invasion between HE families (50). Thus, I-CreII could have arisen by the invasion of an intron-encoded H-N-H enzyme by a GIY-YIG endonuclease, and then during evolution, the catalytic activity of the GIY-YIG motif was lost, while its DNA-binding ability was enhanced.

### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

### ACKNOWLEDGEMENTS

The authors are grateful to O.W. Odom and Youzhong Guo for expert advice.

### FUNDING

Department of Energy [DE-FG03-02ER15352]; Robert A. Welch Foundation [F-1164]; Texas Advanced Research Program [ARP 003658-0144-2007]; Undergraduate Research Fellowship to A.D. Funding for open access charge: Department of Energy.

*Conflict of interest statement.* None declared.

## REFERENCES

- Belfort, M., Derbyshire, V., Parker, M.M., Cousineau, B. and Lambowitz, A.M. (2002) Mobile introns: pathways and proteins. In Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds), *Mobile DNA II*. ASM Press, Washington DC, pp. 761–783.
- Belfort, M., Stoddard, B.L., Wood, D.W. and Derbyshire, V. (eds) (2005) *Homing Endonucleases and Inteins. Vol. 16, Nucleic Acids and Molecular Biology*, Gross, H.J. Series (series ed.). Springer-Verlag, Berlin, p. 377
- Kühlmann, U., Moore, G., James, R., Kleanthous, C. and Hemmings, A. (1999) Structural parsimony in endonuclease active sites: should the number of homing endonuclease families be redefined? *FEBS Lett.*, **463**, 1–2.
- Van Roey, P., Meehan, L., Kowalski, J., Belfort, M. and Derbyshire, V. (2002) Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI. *Nat. Struct. Biol.*, **9**, 806–811.
- Shen, B., Lanthaler, M., Shub, D. and Stoddard, B. (2004) DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J. Mol. Biol.*, **342**, 43–56.
- Galburt, E., Chevalier, B., Tang, W., Jurica, M., Flick, K., Monnat, R. and Stoddard, B. (1999) A novel endonuclease mechanism directly visualized for I-PpoI. *Nat. Struct. Biol.*, **6**, 1096–1099.
- Orlowski, J. and Bujnicki, J.M. (2008) Structural and evolutionary classification of Type II restriction enzymes based on theoretical and experimental analyses. *Nucleic Acids Res.*, **36**, 3552–3569.
- Dunin-Horkawicz, S., Feder, M. and Bujnicki, J.M. (2006) Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics*, **7**, 98–117.
- Magnani, E., Sjölander, K. and Hake, S. (2004) From endonucleases to transcription factors: evolution of the AP2 DNA binding domain in plants. *Plant Cell*, **16**, 2265–2277.
- Nord, D. and Sjöberg, B.-M. (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.
- Kowalski, J.C., Belfort, M., Stapleton, M.A., Holpert, M., Dansereau, J.T., Pietrovski, S., Baxter, S.M. and Derbyshire, V. (1999) Configuration of the catalytic GIY-YIG domain of intron endonuclease I-TevI: coincidence of computational and molecular findings. *Nucleic Acid Res.*, **27**, 2115–2125.
- Mueller, J.E., Smith, D., Bryk, M. and Belfort, M. (1995) Intron-encoded endonuclease I-TevI binds as a monomer to effect sequential cleavage via conformational changes in the *td* homing site. *EMBO J.*, **14**, 5724–5735.
- Belle, A., Landthaler, M. and Shub, D.A. (2002) Intronless homing: site-specific endonuclease SegF of bacteriophage T4 mediates localized marker exclusion analogous to homing endonucleases of group I introns. *Genes Dev.*, **16**, 351–362.
- Liu, Q., Belle, A., Shub, D.A., Belfort, M. and Edgell, D.R. (2003) SegG endonuclease promotes marker exclusion and mediates co-conversion from a distant cleavage site. *J. Mol. Biol.*, **334**, 13–23.
- Sandegren, L., Nord, D. and Sjöberg, B.-M. (2005) SegH and Hef: two novel homing endonucleases whose genes replace the *mobC* and *mobE* genes in several T4-related phages. *Nucleic Acids Res.*, **33**, 6203–6213.
- Mehta, P., Katta, K. and Krishnaswamy, S. (2004) HNH family subclassification leads to identification of commonality in the His-Me endonuclease superfamily. *Protein Sci.*, **13**, 295–300.
- Cheng, Y.-S., Kuo-Chiang, H., Doudeva, L.G., Chak, K.-F. and Yuan, H.S. (2002) The crystal structure of the nuclease domain of colicin E7 suggests a mechanism for binding to double-stranded DNA by the H-N-H endonucleases. *J. Mol. Biol.*, **324**, 227–236.
- Robbins, J.B., Stapleton, M., Stanger, M.J., Smith, D., Dansereau, J.T., Derbyshire, V. and Belfort, M. (2007) Homing endonuclease I-TevIII: dimerization as means to a double-strand break. *Nucleic Acids Res.*, **35**, 1589–1600.
- Goodrich-Blair, H. and Shub, D.A. (1996) Beyond homing: competition between intron endonucleases confers a selective advantage on flanking genetic markers. *Cell*, **84**, 211–221.
- Drouin, M., Lucas, P., Otis, C., Lemiueux, C. and Turmel, M. (2000) Biochemical characterization of I-Cmoel reveals that this H-N-H homing endonuclease shares functional similarities with the H-N-H colicins. *Nucleic Acids Res.*, **28**, 4566–4572.
- Saravanan, M., Bujnicki, J., Cymerman, I., Rao, D. and Nagaraja, V. (2004) Type II restriction endonuclease R.KpnI is a member of the HNH nuclease superfamily. *Nucleic Acids Res.*, **32**, 6129–6135.
- Eddy, S.R. and Gold, L. (1991) The phage T4 *nrdb* intron: a deletion mutant of a version found in the wild. *Genes Dev.*, **5**, 1032–1041.
- Pommer, A., Cal, S., Keelbe, A., Walker, D., Evans, S., Kuhlman, U., Cooper, A., Connolly, B., Hemmings, A., Moores, G., et al. (2001) Mechanism and cleavage specificity of the H-N-H endonuclease colicin E9. *J. Mol. Biol.*, **314**, 735–749.
- Mate, J. and Kleanthous, C. (2004) Structure-based analysis of the metal-dependent mechanism of H-N-H endonucleases. *J. Biol. Chem.*, **279**, 34763–34769.
- Holloway, S.P., Deshpande, N.N. and Herrin, D.L. (1999) The catalytic group I introns of the *psbA* gene of *Chlamydomonas reinhardtii*: core structures, ORFs and evolutionary implications. *Curr. Genet.*, **36**, 69–78.
- Odom, O.W., Holloway, S.P., Deshpande, N.N., Lee, J. and Herrin, D.L. (2001) Mobile introns from the *psbA* gene of *Chlamydomonas reinhardtii*: highly efficient homing of an exogenous intron containing its own promoter. *Mol. Cell. Biol.*, **21**, 3472–3481.
- Kim, H.-H., Corina, L.E., Suh, J.-K. and Herrin, D.L. (2005) Expression, purification and biochemical characterization of intron-encoded endonuclease I-CreII. *Protein Expr. Purif.*, **44**, 162–172.
- Lai, R., Bekessy, A., Chen, C.-C., Walsh, T. and Barnard, R. (2003) Megaprimer mutagenesis using very long primers. *Biotechniques*, **34**, 52–56.
- Sambrook, J. and Russell, D. (2001) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685.
- Orchard, K. and May, G.E. (1993) An EMSA-based method for determining the molecular weight of a protein-DNA complex. *Nucleic Acids Res.*, **21**, 3335–3336.
- Wang, J., Kim, H.-H., Yuan, X. and Herrin, D.L. (1997) Purification, biochemical characterization and protein-DNA interactions of the I-CreI endonuclease produced in *E. coli*. *Nucleic Acids Res.*, **25**, 3767–3776.
- Corina, L.E. (2005) Homing endonuclease I-CreII: a novel dual-motif enzyme that catalyzes group I intron homing. *Ph.D. Dissertation*. University of Texas at Austin.
- Walker, D., Georgiou, T., Pommer, A., Walker, D., Moore, G., Kleanthous, C. and James, R. (2002) Mutagenic scan of the H-N-H motif of colicin E9: implications for the mechanistic enzymology of colicins, homing enzymes, and apoptotic endonucleases. *Nucleic Acids Res.*, **30**, 3225–3234.
- Edgell, D.R. and Shub, D.A. (2001) Related homing endonucleases I-BmoI and I-TevI use different strategies to cleave homologous recognition sites. *Proc. Natl Acad. Sci. USA*, **98**, 7898–7903.
- Loizos, N., Silva, G.H. and Belfort, M. (1996) Intron-encoded endonuclease I-TevII binds across the minor groove and induces two distinct conformational changes in its DNA substrate. *J. Mol. Biol.*, **255**, 412–424.
- Truglio, J., Rhau, B., Croteau, D., Wang, L., Skorvaga, M., Karakas, E., DellaVecchia, M., 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.
- Chevalier, B., Sussman, D., Otis, C., Noel, A., Turmel, M., Lemiueux, C., Stephens, K., Monnat, R. and Stoddard, B. (2004) Metal-dependent DNA cleavage mechanism of the I-CreI LAGLIDADG homing endonuclease. *Biochemistry*, **43**, 14015–14026.
- Perrin, A., Buckle, M. and Dujon, B. (1993) Asymmetrical recognition and cleavage activity of the I-SceI endonuclease on its site and on intron-exon junctions. *EMBO J.*, **12**, 2939–2947.
- Wittmayer, P.K. and Raines, R.T. (1996) Substrate binding and turnover by the highly specific I-PpoI endonuclease. *Biochemistry*, **35**, 1076–1083.
- Lykke-Andersen, J., Garrett, R.A. and Kjems, J. (1997) Mapping metal ions at the catalytic centres of two intron-encoded endonucleases. *EMBO J.*, **2**, 3272–3281.

42. 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.
43. 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-TevI correlates with the two-domain configuration of the homing site. *J. Mol. Biol.*, **265**, 494–506.
44. 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.
45. Keeble,A.H., Mate,M.J. and Kleanthous,C. (2005) HNH endonucleases. In Belfort,M., Stoddard,B.L., Wood,D.W. and Derbyshire,V. (eds), *Homing Endonucleases and Inteins*. Springer-Verlag, Berlin, pp. 49–65.
46. Doudeva,L.G., Huang,H., Hsia,K-C., Shi,Z., Li,C-L., Shen,Y., Cheng,Y-S. and Yuan,H.S. (2006) Crystal structure and metal-dependent stability and activity studies of the ColE7 endonuclease domain in complex with DNA/Zn<sup>2+</sup> or inhibitor/Ni<sup>2+</sup>. *Protein Sci.*, **15**, 269–280.
47. Huang,H. and Yuan,H.S. (2007) The conserved asparagine in the HNH motif serves an important structural role in metal finger endonucleases. *J. Mol. Biol.*, **368**, 812–821.
48. Pröschold,T., Marin,B., Schlösser,U.G. and Melkonian,M. (2001) Molecular phylogeny and taxonomic revision of *Chlamydomonas* (Chlorophyta). I. Emendation of *Chlamydomonas* Ehrenberg and *Chloromonas* Gobi, and description of *Oogamochlamys* gen. nov. and *Lobochlamys* gen. nov. *Protist*, **152**, 265–300.
49. Brouard,J-S., Otis,C., Lemieux,C. and Turmel,M. (2008) Chloroplast DNA sequence of the green alga *Oedogonium cardiacum* (Chlorophyceae): unique genome architecture, derived characters shared with the Chaetophorales and novel genes acquired through horizontal transfer. *BMC Genomics*, **9**, 290–306.
50. Saguez,C., Lecelier,G. and Koll,F. (2000) Intronic GIY-YIG endonuclease gene in the mitochondrial genome of *Podospira curvicola*: evidence for mobility. *Nucleic Acids Res.*, **28**, 1299–1306.