Current Topics in Biochemical Research
Vol. 16, 2014

Original Communication

Evolutionary adaptation of the GIY-YIG domain for specific DNA binding in I-CreII

Weihua Qiu[#], Youzhong Guo[#] and David L. Herrin*

Department of Molecular Biosciences and Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX 78712, USA.

ABSTRACT

Homing endonucleases promote gene invasion by cleaving specific DNA targets, and they typically contain one of several conserved catalytic domains, plus additional motifs that provide specific DNA binding. However, at least two enzymes from green algal chloroplasts, I-CmoeI and I-CreII, have two apparent catalytic domains (H-N-H and GIY-YIG, respectively), and lack most conserved DNA-binding motifs. Mutagenesis of I-CreII indicated that the H-N-H motif is catalytic, and, based on DNA-affinity measurements, that the GIY-YIG motif could be involved in DNA-binding. Here we have used footprinting techniques to investigate target DNA binding by native and substituted forms of I-CreII. Hydroxyl radical footprinting with wild-type I-CreII indicated that the tightest binding occurs downstream of the cleavage site, though binding at the cleavage site was apparent on the top strand, which is the first one cleaved by I-CreII. DNase I footprinting was possible with the variants that have reduced DNA affinity, as well as wild-type. The GIY-YIG variant, G231E/K245A, showed a strong differential effect within the footprint, especially with the hyper-cleavage of several nt just downstream of the cleavage site on the top strand. Structural modeling of the GIY-YIG domain using the solved I-TevI structure as a template indicated that I-CreII has a β-hairpin insertion that interrupts an α-helix normally required for catalysis. These results indicate that the GIY-YIG domain in I-CreII plays an important, if not dominant, role in specific DNA binding. Moreover, they suggest that the evolution of the GIY-YIG domain into this role may have begun with the β-hairpin insertion, which inactivated its catalytic ability.

KEYWORDS: *Chlamydomonas reinhardtii*, *psbA*, endonuclease, H-N-H/GIY-YIG, intron homing

INTRODUCTION

Homing endonucleases (HEs) are a distinct group of proteins encoded mostly within mobile introns, but also within inteins, and as free-standing mobile HEs (see Hafez & Hausner [1] for a recent review). HEs cleave DNA specifically, and close to the site of DNA insertion; most of them generate a double-strand (ds) break, but some cleave only one strand. Cleavage stimulates recombinational repair of the target DNA using the donor DNA − which contains the mobile element − as template [2]. HEs have long (14-40 bp) asymmetric recognition sequences, and can usually tolerate multiple base changes in the sequence, and still effect cleavage [3]. HEs have a catalytic domain, and one or more DNA-binding domains, and are usually classified according to the former. Seven families of HEs are now recognized, with the

*Corresponding author: herrin@utexas.edu

[#]Present addresses: W. Qiu, Department of Pharmacology, College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA;
Y. Guo, Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA.

LAGLIDADG, H-N-H and GIY-YIG being the largest; the others are His-Cys, PD(D/E)×K, ED×HD and HJ Resolvase-like [1, 3-4]. The most relevant to this study are in the GIY-YIG and H-N-H classes.

The GIY-YIG domain is ~90 amino acids long, although its name is based on the conserved GIY-(X_{10-11})-YIG region near the beginning of the domain. Studies of I-TevI, I-TevII, and I-BmoI revealed them to be small, monomeric enzymes with the catalytic GIY-YIG domain at the N-terminus, and multiple DNA-binding regions towards the C-terminus [5]. It was also shown that they recognize long DNA sequences (35-40 bp), and cleave strands sequentially [6-8]. A crystal structure of the GIY-YIG domain of I-TevI revealed a mixed α/β topology, ββααβα, with the three β strands forming a β-sheet, and the conserved triads being part of β1 and β2, respectively [9]. Mutational and structural data confirmed the importance of the triads, and identified key R and E residues in α1 and α3, respectively. The GIY-YIG domains in I-TevI and I-BmoI do not bind DNA with high affinity, as that job falls to three downstream DNA-binding motifs connected to the catalytic domain by a flexible linker [10-11]. The DNA-binding regions act as an anchor so that the catalytic domain can be free to move and cleave both strands [11].

The GIY-YIG motif has also been found at the active site for other types of endonucleases besides HEs, including restriction endonucleases (Eco29kI and Hpy188I), the UvrC nuclease for nucleotide excision repair, endonuclease II from phage T4, and non-specific nucleases (Ankle1 and glutaredoxin AtGRX16a from plant chloroplasts) [12-17]. Unlike the HEs (I-TevI, I-TevII, and I-BmoI), which act as monomers, the restriction endonucleases Eco29kI and Hpy188I function as dimers, and have several interesting modifications to the GIY-YIG active site [18-19].

The classic H-N-H active site motif is only ~30 amino acids, and contains a conserved N flanked by multiple (usually) conserved histidines [20]; however, the motif is often extended at one or both ends to include one or two zinc binding sites [21-22]. The biochemical characteristics of these HEs are more diverse than the other families, as some cleave only one strand (eg. I-HmuI and I-PfoP3I) while others generate double-strand breaks (eg. I-TevIII and I-CmoeI), and while I-TevIII leaves 5' overhangs I-CmoeI leaves 3' overhangs [4, 23-24]. Also, while most H-N-H endonucleases seem to act as monomers, I-TevIII functions as a dimer [25]. DNA-binding domains are also important for the specificity of H-N-H HEs; I-HmuI contains a C-terminal DNA-binding region that has two continuous α-helices and a helix-turn-helix motif [26], and I-TevIII contains two Zn-fingers that interact with DNA [25]. Finally, the H-N-H motif is also found in proteins that appear not to be nucleases, including the AP2 family of plant transcription factors and an annealing helicase AH2 [27-28].

Holloway *et al.* [29] first suggested that the protein encoded by the *Chlamydomonas reinhardtii psbA4* intron contained two of the signature catalytic motifs, GIY-YIG and H-N-H, and was closely related (58% identity) to a *psbA* intron-encoded protein from *Chlamydomonas moewusii*. Biochemical studies of the *C. moewusii* protein, I-CmoeI, showed that it cleaves both DNA strands, and has a flexible metal requirement like H-N-H endonucleases [24]. Subsequently, we showed that the *C. reinhardtii* protein, I-CreII, was required for homing by the *psbA4* intron [30], and that it cleaves the intron-minus (but not intron-plus) gene near the site of intron insertion [31]. The recognition sequence derived from digesting sequence ladders was ~29 bp and included the cleavage site (CS) on each strand, and the intron-insertion site (IIS). I-CreII also exhibited a flexible metal requirement for cleavage like I-CmoeI, but unlike that enzyme it could bind target DNA in the absence of added metals [32], a fact that facilitated further DNA-binding studies. Sizing of the protein-DNA complex and kinetic analyses both indicated that I-CreII cleaves DNA as a monomer, sequentially cleaving the two strands beginning with the top strand [33].

Alanine substitution mutagenesis of both motifs indicated that the H-N-H domain was responsible for cleaving both strands [33], thus leaving open the question of the GIY-YIG motif's role. Since I-CreII seemed to lack any conserved or obvious DNA-binding motifs, one possibility was that the GIY-YIG domain played a role, perhaps an important one, in binding DNA. DNA affinity

estimates of the alanine-substitution variants suggested that both motifs were needed for high-affinity DNA binding but that the GIY-YIG motif might be more important, based on the fact that only one H-N-H variant (N161A) showed strongly reduced binding, but most of the GIY-YIG variants exhibited strong reductions in DNA affinity [33]. Since this would constitute a new role for a GIY-YIG motif, we have explored further the interaction of I-CreII with DNA using hydroxyl radical and DNase I footprinting of wild-type (WT) and substitution mutants. The results point to an important, if not dominant, role for the GIY-YIG domain in specific DNA binding. Moreover, 3-D homology modeling indicates that the GIY-YIG domain of I-CreII has a β-hairpin insertion in an α-helix usually required for catalysis. This insertion can explain the loss of catalytic activity, and may have been the key event that shifted the evolutionary path of this domain toward DNA binding.

MATERIALS AND METHODS

I-CreII variants

All of the I-CreII variants were created previously [33] except K245A, which was created using the QuikChange II kit (Stratagene) and oligos 653 (cacgaaaaagatttagcacaaagtacaaaaatg) and 654 (catttttgtactttgtgctaaatctttttcgtg). The parental plasmid was pI-CreII [31] and the new K245A variant was confirmed by sequencing the I-CreII portion on both strands.

Protein expression and purification

Native I-CreII and the substituted variants were produced without any non-native tags as described previously [31], except the second chromatographic step on heparin-sepharose was omitted as before [33]. The protein preparations were quantified using the Bradford assay (BioRad), and analyzed by SDS polyacrylamide (12% acrylamide) gel electrophoresis along with protein standards (Bio-Rad). The gels were stained with coomassie blue and scanned, in order to gauge purity and determine the fraction that was specifically I-CreII, which ranged from 60% - 85% for the variants. Since the footprinting analyses were performed with excess I-CreII (compared to the DNA target), a higher level of purity was unnecessary. The proteins were stored at -70 °C in aliquots that were thawed only once.

Hydroxyl radical protection analysis

The target DNA (141 bp) containing the *psbA* exon 4-exon 5 junction (E4-E5$_{141}$) was synthesized by PCR from plasmid pE4-E5 with 5'-labeled primers as described before [33]. The hydroxyl radical protection analysis was adapted from previous protocols [34-35]. A large excess of I-CreII (~150 pmol) was incubated with 5 pmol of ^{32}P-labeled (top or bottom strand) E4-E5$_{141}$ DNA in 20 μL of 20 mM Tris-HCl pH 8.0, 1 mM DTT, 0.1 mM EDTA, 30 μg/mL polydeoxyinosinic-deoxycytidylic acid (polydIdC) for 20 min at 23 °C to allow equilibrium binding; a reaction without I-CreII served as a control. To effect hydroxyl radical cleavage, the mixtures were adjusted to 0.1 mM $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$, 0.2 mM EDTA, 10 mM sodium ascorbate, 0.15% H_2O_2 in a total vol. of 23 μL, and incubated for 5 min at 23 °C. The reactions were quenched with 1.2 μL of 0.5 M thiourea, and separated on 8% native polyacrylamide gels (23 °C) in 0.5 × TBE. The I-CreII-DNA complex was excised and eluted, extracted with phenol-chloroform, and precipitated with ethanol. The DNA was dissolved in formamide loading buffer (94 °C for 3 min), and analyzed on denaturing polyacrylamide (8%) gels (7 M urea/1 × TBE at 50 °C) together with an A+G ladder of the target DNA. The dried gels were quantified with a phosphorimager and 1D Image Analysis software (Kodak, v 3.6).

DNase I protection analysis

The target DNA was prepared as described in the preceding section, and the DNase I footprinting followed a previous method [36-37]. For binding, varying concentrations (0 to 50 nM) of I-CreII were incubated with ^{32}P-labeled E4-E5$_{141}$ DNA (2.5 nM) in 20 mM Tris-HCl pH 8, 1 mM DTT, 0.1 mM EDTA, 30 μg/mL polydIdC (20 μL) for 20 min at 23 °C. Then, 1 U of DNase I (New England Biolabs) was added, after adjusting the reactions to 2 mM $CaCl_2$, and cleavage proceeded for 10 min at 23 °C. The reactions were quenched with excess EDTA, extracted with phenol/chloroform, and separated on 8% or 6% denaturing polyacrylamide gels, along with an A+G ladder as described in the preceding section.

Structural modeling

Sequence analysis indicated that the H-N-H motif of I-CreII is similar to that of the colicin E9 DNase, whose crystal structure (PDB:1V15) was available [38-39]. The colicin structure was used as a template to model the H-N-H domain of I-CreII with SWISS-MODEL [38], and with TFmodeller [40]. For the GIY-YIG domain of I-CreII, the model was built with SWISS-MODEL using the structure of the GIY-YIG domain of I-TevI (PDB:1LN0) [9] as the template. The RMSD values of the best models were calculated with PyMol v. 1 [41]. The modeled structures and templates were also analyzed with the SSM module (Secondary Structure Matching) of WinCoot.

RESULTS

Hydroxyl radical protection analysis of native I-CreII bound to target DNA

Since hydroxyl radical protection analysis relies on solvent-based OH· to cleave DNA, it reveals sites of tight binding by a protein along the sugar-phosphate backbone [34-35]. Native or wild-type (WT) I-CreII and a segment of the *psbA* gene containing the exon 4-exon 5 junction (E4-E5$_{141}$) [33] were used for this analysis. Binding was performed in the absence of Mg^{2+} with enough excess I-CreII, based on a mobility shift assay [33], to bind all of the target DNA. Since the E4-E5$_{141}$ DNA was end-labeled on each strand separately, the protection patterns are strand-specific.

Fig. 1 shows a representative gel of the hydroxyl-radical cleavage patterns of the sense (Top) and antisense (Bottom) strands of E4-E5$_{141}$ in the presence (+) and absence (−) of excess I-CreII (Fig. 1A). After quantification and expression of the data as the relative protection per nt (which is the reduction in signal in the + lanes relative to the – lanes), the results show that I-CreII protects, at least partially, a region of 24 nt on the top strand (−9 to +15 relative to the IIS) and 26 nt on the bottom strand (−8 to +18), thus giving it an overall footprint of 27 bp (−9 to +18) (Fig. 1B). However, most of the strong protection occurred downstream of the CS on each strand, especially on the bottom strand. There was protection around the CS on the top strand, but little around the CS on the bottom strand, which is congruent with the finding that I-CreII cleaves the top strand first [33]. It should also be said that increasing the I-CreII to DNA ratio by a further 2- to 4-fold did not change the hydroxyl radical footprints, confirming that the protein was saturating. In conclusion, the hydroxyl radical protection results indicate that I-CreII binds asymmetrically to its target site (with respect to the CS), with the most and strongest binding occurring downstream of the CS.

DNase I protection analysis of DNA binding by native I-CreII and H-N-H variants

Protection from DNase I [36-37] was also used to examine native I-CreII, and because of its stronger signal-to-noise ratio, it could be used to analyze variants with reduced DNA affinity (the alanine-substituted amino acids are indicated in Fig. 2). Fig. 3 shows the protection patterns of the top (Fig. 3A) and bottom (Fig. 3B) strands for WT and the H146A, N161A, and H174A variants; the results with H170A were indistinguishable from H174A, so only the former is shown.

The protection by the WT enzyme spanned a region from −25 to +18 of the top strand (Fig. 3A) and from −16 to +16 on the bottom strand (Fig. 3B). The increased signals with WT enzyme at −5 on the top strand and −6 on the bottom strand are due to the enzyme cleaving its own CS (both sites are labeled I-CreII in the figure). This was made possible by the divalent cation that had to be included with the DNase I, and is confirmed by the reduced signals with H146A and H174A, which have greatly reduced cleavage activity [33]. Overall, I-CreII protects regions both upstream and downstream of the CS (the most protected regions are indicated with the smaller brackets); however, protection of several nt around the CS (−8, −7, −4 on the top strand and −7, −9, −10 on the bottom) is weaker than the protection of nt in the flanking regions.

Also shown in Fig. 3 are the results with the H-N-H variants; relative to WT, the DNA affinity of H174A and H146A were reduced 2- and 4-fold, respectively, whereas N161A was reduced ~20-fold [33]. Consistent with the EMSA data, higher concentrations of N161A (> 5 nM) were needed to see a footprint, and at 25 nM there is a

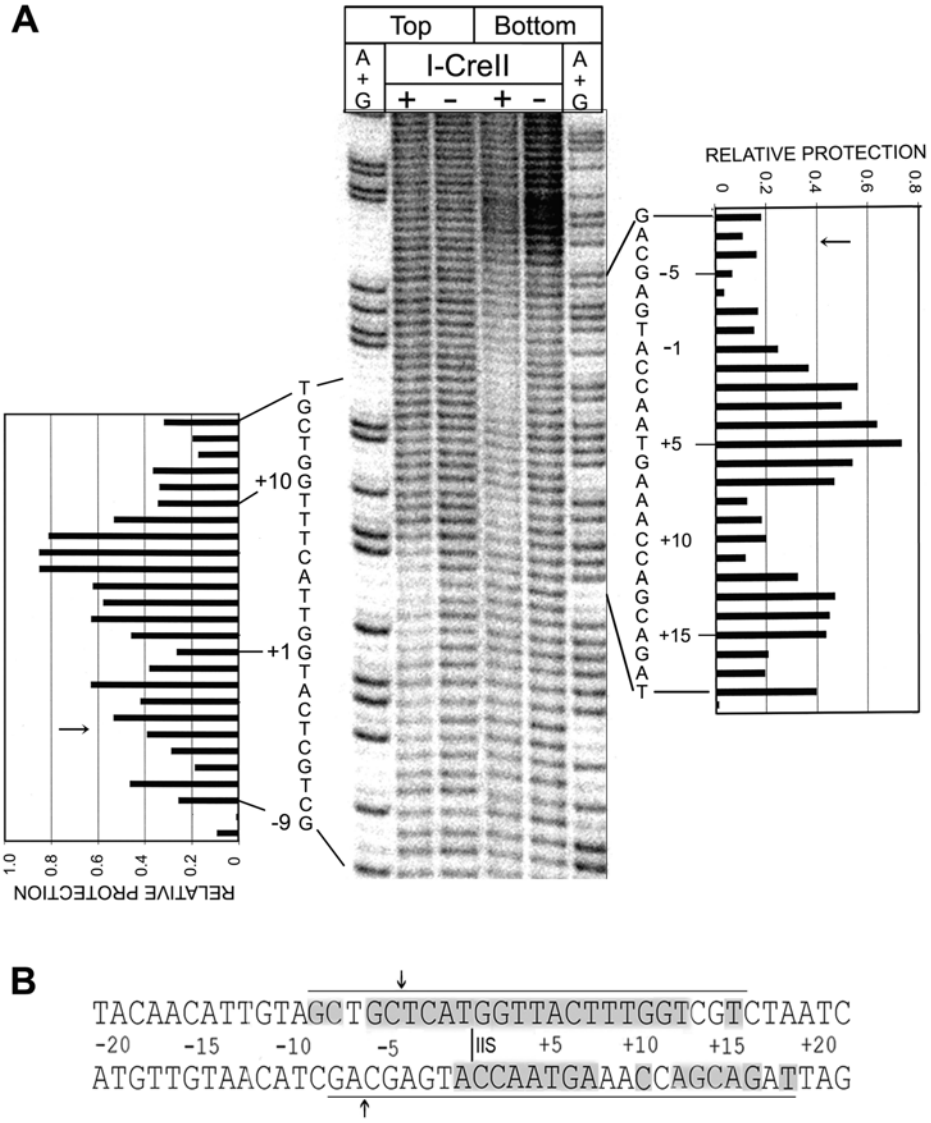

Fig. 1. Hydroxyl radical protection analysis of I-CreII binding to target DNA. **(A)** A phosphorimage of a denaturing polyacrylamide (8%) gel that shows the relative protection of nt in the homing site. The specific strand of E4-E5$_{141}$ DNA (Top or Bottom) that was end-labeled is indicated above the gel, and the symbols indicate the presence (+) or absence (−) of I-CreII. The A+G ladders provided markers for each strand, and the homing site nt are numbered relative to the IIS. The plots of relative protection (for the indicated nt) were determined from the phosphorimage using the equation, 1 – volume with (+) I-CreII /volume without (−) I-CreII; the CS on each strand is indicated by the arrow. To minimize any effects of uneven loading between lanes, the values were normalized by using reference nt that were adjacent to the protected regions. **(B)** The I-CreII homing site and region protected from hydroxyl radicals. The lines delineate the range of the protected region, whereas shading indicates nt with a protection score ≥ 0.2. The nt are numbered relative to the IIS, which is indicated with a vertical line; the CS on each strand is marked with a vertical arrow.

differential effect on the top-strand pattern (Fig. 3A); the reduction in protection of nt upstream of the CS (from −7 to −12) is greater than the reduction in protection of several nt located downstream (from −1 to +5). A similar effect can be seen for the bottom strand with N161A, where the reduction in protection of nt just upstream of the CS is greater than the reduction in protection of

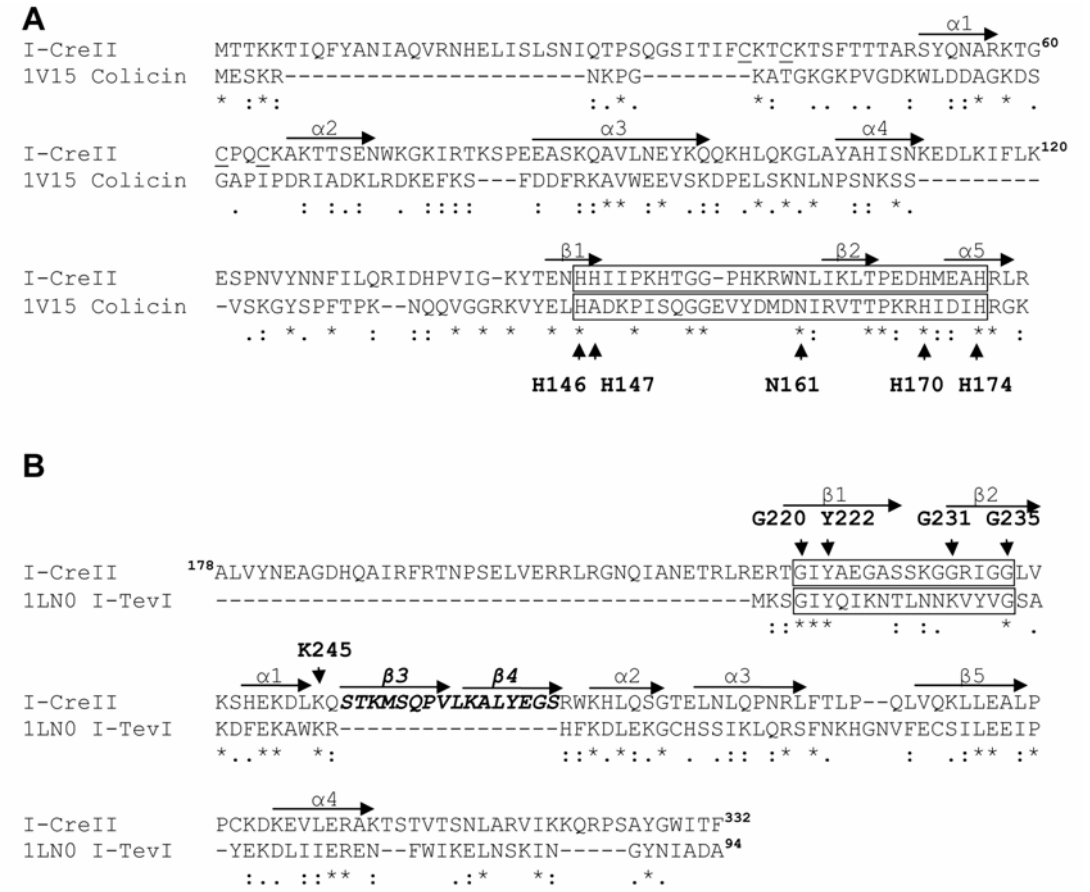

Fig. 2. Sequence alignments of the I-CreII domains with reference proteins of known 3D structure: The secondary structures predicted for I-CreII and the alanine-substituted amino acids are also indicated. Clustal X generated the alignments. The symbols are: dash = gap, asterisk (*) = identity, colon (:) = conservation of a strong group in the Gonnet Pam250 matrix, and period (.) = conservation of a weaker group in the matrix. **(A)** Alignment of the N-terminal region of I-CreII (aa 1-178) with the same region of colicin E9 (1V15 Colicin). The conserved core of the H-N-H motif is boxed, and the residues that were changed to alanine are indicated below the alignment. The secondary structures above the alignment are for I-CreII, but except for α1 and α2 they are also found in the colicin. **(B)** Alignment of the C-terminal region of I-CreII (aa 178-332) with the GIY-YIG domain of I-TevI (1LN0 I-Tev). The conserved core of the GIY-YIG motif is boxed, and the substituted residues are indicated above the alignment. The secondary structures above the alignment are for I-CreII, but except for the β-hairpin insertion (β3 and β4), whose seqeunce is in bold italics, they are also found in I-TevI.

those downstream (−5 to +1) (Fig. 3B). It should be noted that the apparent lack of protection by N161A at −5 on the top strand (labeled as I-CreII) is probably due to this variant's own cleavage activity [33]; this interpretation is supported by the increased signal at the higher protein concentrations.

DNase I protection analysis of DNA binding by the GIY-YIG variants of I-CreII

DNase I protection analysis was also performed with the GIY-YIG variants created previously [33]. And as part of this effort we discovered that the K245A variant had another mutation that changed G231 to E231 (see Fig. 2B), thus making it a double mutant, which we renamed G231E/K245A. A new K245A variant was created that had only the single substitution. It should be noted that the cleavage efficiency of all the GIY-YIG variants is close to WT, but their affinity for DNA is reduced, especially G231E/K245A [33].

The reduced DNA binding of these variants is also apparent from the DNase I protection results

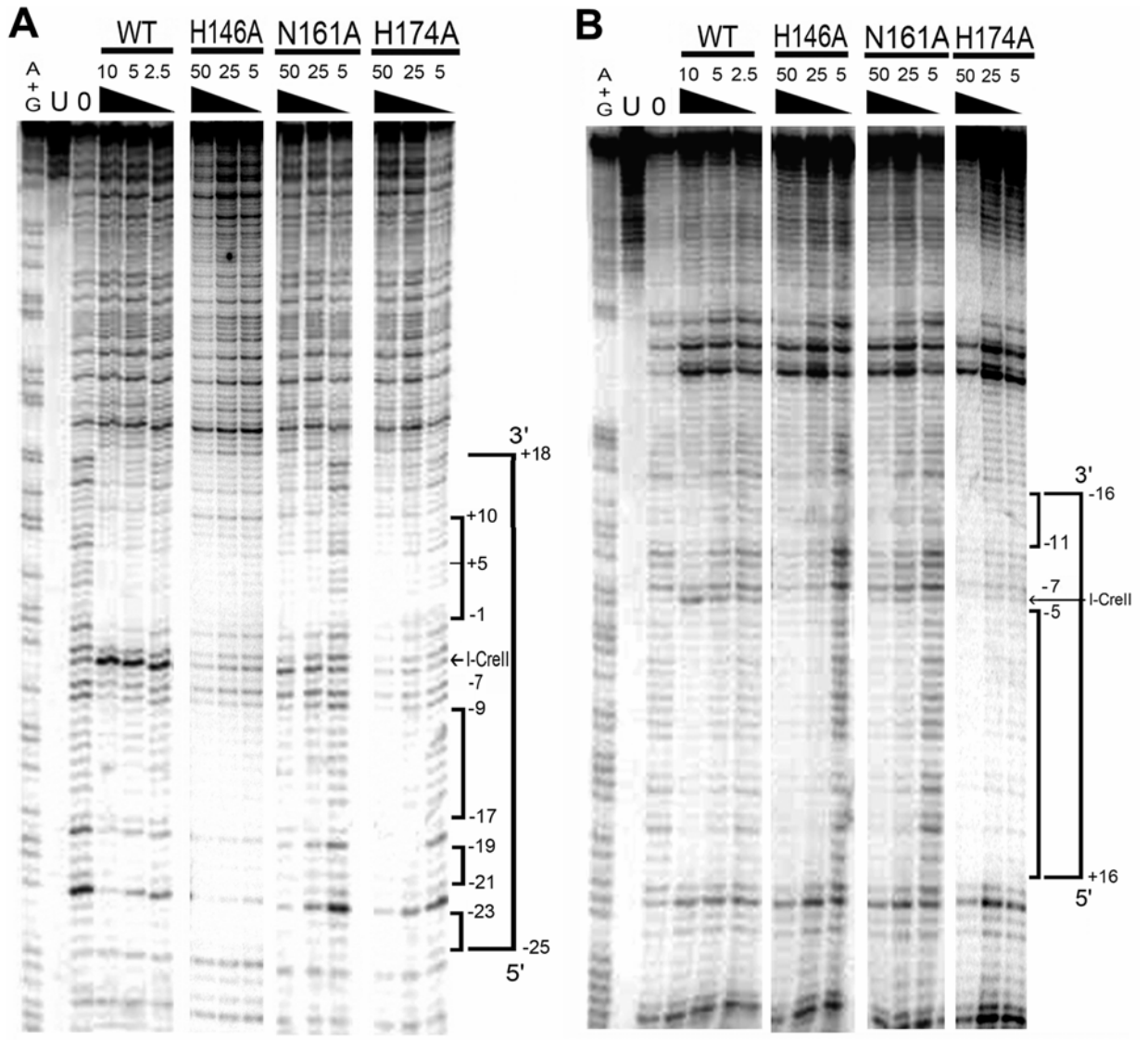

Fig. 3. DNase I footprinting of WT I-CreII and variants with an alanine substitution in the H-N-H motif. **(A)** Analysis of the top strand. Shown are phosphorimages of denaturing 8% polyacrylamide gels. The target DNA was 5'-labeled on the top strand, and the nt numbering scheme is similar to Fig. 1. The A+G reaction (A+G) served as a marker; untreated DNA (U), and DNA cleaved with DNase I in the absence of I-CreII (0), were controls. The concentrations in nM of the I-CreII variants are indicated above the triangles. The outer bracket indicates the range of the binding region, whereas the inner brackets indicate regions of strongest protection. The product of cleavage by I-CreII itself is identified (I-CreII). Note: DNase I products lack a 3' phosphate and migrate about 1.5 bands slower than their counterparts in the A+G ladder. **(B)** Analysis of the bottom strand. The analysis was similar to (A), except the gels were 6% polyacrylamide, and the target DNA was labeled on the bottom strand.

for the top strand (Fig. 4A), where protein concentrations > 5 nM were required to obtain good footprints. The most dramatic effect on the protection pattern was obtained with the G231E/K245A variant, which induced hyper-cleavage of the 4 nt downstream of the CS, especially at −4 and −2 (Fig. 4A). There was also reduced protection farther downstream from the CS (from −1 to +9) with G231E/K245A, and yet its protection at the other end of the footprint (from −25 to −17) was still robust. On the bottom strand (Fig. 4B), there was reduced protection

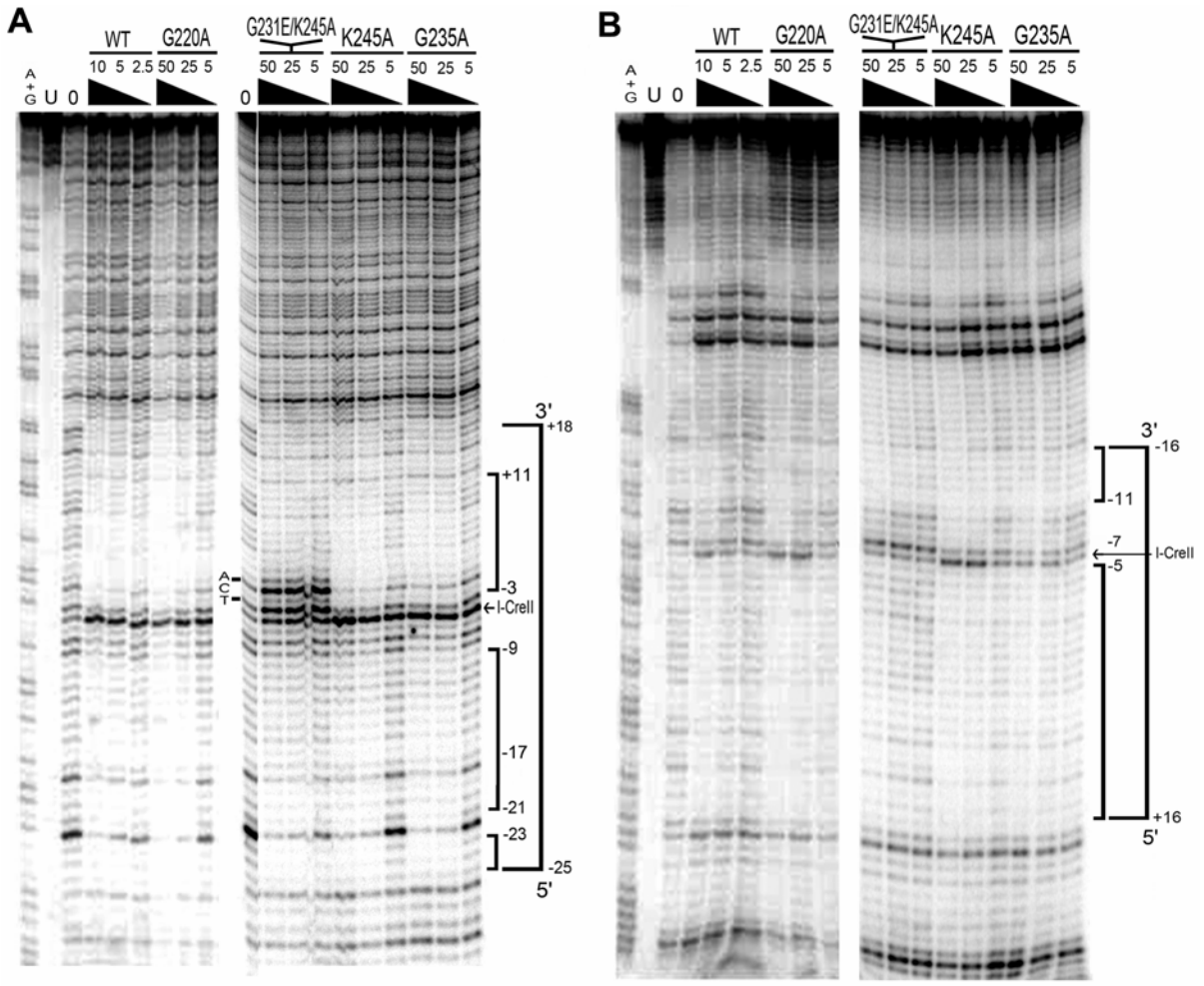

Fig. 4. DNase I protection analysis of GIY-YIG motif variants. **(A)** Analysis of the top strand. Phosphorimages of denaturing 8% polyacrylamide gels are shown. The target DNA was 5'-labeled on the top strand, and the nt numbering is similar to Fig. 1. The A+G reaction (A+G) served as a marker. Untreated DNA (U), and DNA cleaved with DNase I in the absence of I-CreII (0), were controls. The concentrations in nM of the I-CreII variants are indicated above the triangles. The outer bracket indicates the range of the binding region, whereas the inner brackets indicate regions of strongest protection. The corresponding nt identities of the 3 products above the I-CreII product are indicated between the gels. **(B)** Analysis of the bottom strand. The analysis was similar to (A), except the gels were 6% polyacrylamide, and the target DNA was end-labeled on the bottom strand.

with G231E/K245A on both sides of the CS, but the footprint (at 25 nM protein) was weakened throughout the downstream nt (from −5 to +16). On the other hand, the G231E/K245A footprint was still evident farther upstream from the CS (from −11 to −16). In summary, the effect of the G231E/K245A substitutions on the DNAse I footprint is stronger for the top strand than for the bottom and is centered just downstream of the CS, suggesting that is where the GIY-YIG motif binds to the DNA.

Summary of the DNA footprinting analyses

A summary of the target residues protected from hydroxyl radical and DNase I cleavage is presented in Fig. 5. The hydroxyl radical footprint is highly asymmetric (with respect to the CS), as most of it is located downstream of the CS, especially on the bottom strand. The hydroxyl radical protection around the CS on the top strand is consistent with the known preference of the enzyme for cleaving this strand first [33]. The DNase I footprint aligns fairly closely with the

hydroxyl radical footprint, except the former extends farther upstream and was weaker around the CS for both strands (indicated by the dashes in Fig. 5). Two nt that became hyper-sensitive to DNase I with the G231A/K245A variant are indicated with asterisks.

Homology modeling of I-CreII

Homology-based structural modeling provided additional insight into I-CreII structure and evolution. The H-N-H domain was modeled using the structure of colicin E9 [38, 39], whose sequence was used in the alignment in Fig. 2A, as the template. The superimposed models [40] were quite similar overall, but were especially congruent for the core H-N-H motif (not shown). A notable difference between the two structures involved the N-terminal regions, where I-CreII was predicted to contain two alpha helices (α1 and α2 in Fig. 2A) not found in the colicin. The GIY-YIG domain of I-CreII (from R218-T331) was modeled using the structure of the GIY-YIG domain of I-TevI [9] as the template (Fig. 6). Like I-TevI, the I-CreII structure is a three-stranded

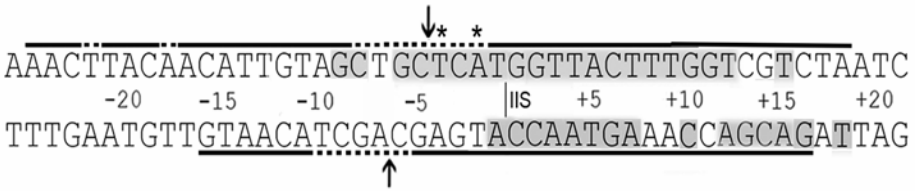

Fig. 5. Summary of the DNase I and hydroxyl radical protection analysis. The horizontal lines delimit the DNase I footprint; the solid portions indicate strong protection and the dashed portions, weak or minimal protection. The shaded letters indicate nt protected (≥ 0.2 relative protection) from hydroxyl radical cleavage. The asterisks (*) indicate nt that were hypersensitive to DNase I in the G231E/K245A variant. The I-CreII CS on each strand is indicated with an arrow, and the IIS with a vertical line; nt numbering is relative to the IIS.

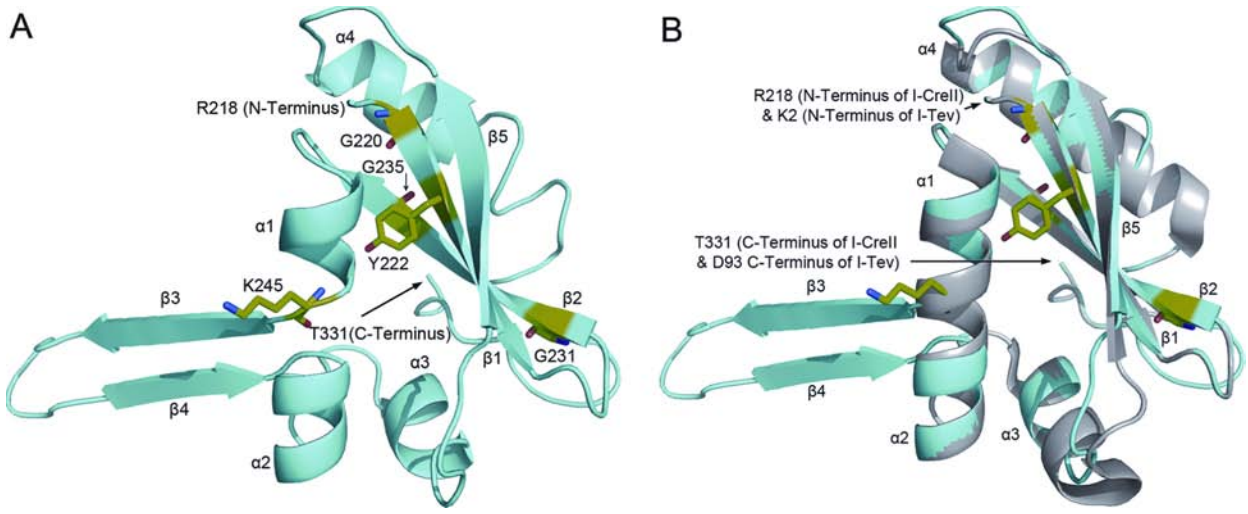

Fig. 6. Structural model of the GIY-YIG domain of I-CreII. **(A)** The model (amino acids 218-331) was built with SWISS-MODEL using the structure of the GIY-YIG domain of I-TevI (PDB: 1LN0) as template. The conserved portion of the GIY-YIG domain that contains the triads, G220-G235, is part of a β-sheet composed of three β-strands (β1, β2, β5) similar to I-TevI; residues that were substituted with alanine (G220, Y222, G231, G235, K245) are highlighted (green) on the ribbon structure (cyan). **(B)** Superimposed structures of the GIY-YIG domain of I-CreII (cyan) with that of I-TevI (gray). The structures are very similar, except for the β-hairpin in I-CreII (β3 and β4), which disrupts the long α-helix of I-TevI located just downstream of β2 (α1 in I-TevI).

β-sheet (β1, β2 and β5) flanked by α-helices, with the conserved triads located in β1 (GIY) and β2 (IGG), respectively (Fig. 6A). However, as the superimposed models show (Fig. 6B), I-CreII has a β-hairpin insertion that interrupts the α1 helix of I-TevI. As this helix is required for catalysis [42], the β-hairpin insertion probably destroyed any catalytic activity that this domain possessed before the insertion. When the β-hairpin was excluded from the comparison, the two structures had an average RMSD value of 0.091, which is indicative of an accurately predicted structure for I-CreII [41].

DISCUSSION

I-CreII is an atypical HE that appeared to contain two catalytic motifs (H-N-H and GIY-YIG) and to lack known DNA-binding motifs [29, 33]. Previously, we employed substitution mutagenesis to demonstrate that the H-N-H motif is the catalytic site [33]. Here, we have used footprinting techniques with the substitution variants to investigate specific DNA binding by I-CreII. Hydroxyl radical footprinting of WT I-CreII on its native DNA target indicated that tight binding is concentrated downstream of the CS on both strands; however, evidence for tight binding on both sides of the CS was obtained for the top strand. This point is significant because I-CreII cleaves the top strand first. DNase I footprinting was effective not only with the WT enzyme but also with the H-N-H and GIY-YIG substitution variants, and the results indicate that the GIY-YIG domain binds downstream of the CS. Homology-based 3D modeling of the GIY-YIG domain of I-CreII using the structure from I-TevI as a template suggests that the I-CreII domain contains a β-hairpin insertion, which interrupts an α-helix that is required for catalysis by GIY-YIG endonucleases [8, 19, 42]. The insertion thus explains the loss of catalytic activity by the GIY-YIG domain in I-CreII, and moreover, raises the possibility that it was the key step in the evolution of this catalytic domain into a DNA-binding domain.

Hydroxyl radicals are small solvent-based agents that attack DNA sugars and thus identify sites of tight backbone binding. However, the signal to noise ratio was better with DNase I, and this was necessary for analyzing the footprints of the variants with reduced DNA affinity. With the WT enzyme, the DNase I footprint aligns fairly well with the hydroxyl radical footprint, except at the upstream border, where the DNase I footprint extends farther upstream, especially on the top strand. The lack of protection from hydroxyl radicals for this region suggests that binding is not as tight and possibly restricted to phosphates, which are the sites of DNase I action. This scenario conjures up the mechanism used by I-BmoI (a GIY-YIG HE that cleaves as a monomer), in which tight DNA binding downstream of the CS anchors the protein to the target, so that a weakly-binding catalytic site can rotate and cleave both strands [11]. It should be noted, however, that the I-CreII footprint is not as asymmetric as the I-BmoI pattern, which protects residues only on the downstream side of the CS [8]. Thus, further work will be necessary to determine whether the I-BmoI model fits I-CreII.

One thing we need to know about I-CreII is the 3-D structure and function of the non-conserved N-terminal region, which is predicted to contain two α-helices (Fig. 2A), but may also contain two Zn-binding sites [21-22]. The potential Zn-binding sites are the CXXC residues at aa 39-42 and 61-64 that flank α1 of the H-N-H domain (Fig. 2A). It is thus reasonable to suggest that this region is involved in binding DNA on the upstream side, but at some distance from the CS, since the H-N-H core must be positioned at the CS to start the reaction cycle.

With the I-CreII variants, a differential effect on the protected region was observed only with the mutants that had strong (> 15-fold) reductions in DNA binding, namely N161A and G231E/K245A. And among these, the effect was dramatic only for the GIY-YIG domain variant, G231E/K245A, which actually exhibited DNase I hyper-cleavage of several nt downstream of the CS on the top strand. This effect indicates that not only was there unfettered access of DNAse I, but it also suggests that the structure of these nt was distorted when the variant was bound.

Evidence that the H-N-H motif has been adapted for purposes other than as a catalytic endonuclease motif has been published [27-28], but evolutionary repurposing of the GIY-YIG motif has not been reported to our knowledge. In this case,

the GIY-YIG motif has been adapted for specific DNA binding, presumably because the catalytic endonuclease function could also be supplied by the H-N-H domain. Although the origins of I-CreII are not yet clear, a plausible scenario that we suggested earlier [33] is still a viable explanation. I-CreII could have been created by the invasion of an H-N-H endonuclease by a GIY-YIG endonuclease (there is an example of a LAGLIDADG endonuclease being invaded by a GIY-YIG endonuclease [43]), and then during evolution the catalytic activity of the GIY-YIG domain was lost and its DNA-binding ability enhanced. These results suggest that the loss of catalytic activity could have occurred in one step, with the insertion of the β-hairpin sequence into the R27 α-helix. Further work will be necessary to pinpoint the changes that increased the ability of this domain to bind DNA specifically.

ACKNOWLEDGEMENTS

We are grateful for helpful discussions with Dr. O. W. Odom, and for financial support from the Robert A. Welch Foundation (F-1164) and the Texas Advanced Research Program (003658-0144-2007).

CONFLICT OF INTEREST STATEMENT

The authors report no conflicts of interest.

REFERENCES

1. Hafez, M. and Hausner, G. 2012, Genome, 55, 553.
2. Belfort, M., Derbyshire, V., Parker, M. M., Cousineau, B. and Lambowitz, A. M. 2002, Mobile DNA II, N. L. Craig, R. Cragigie, M. Gellert and A. M. Lambowitz (Eds.), ASM Press, New York, 761.
3. Stoddard, B. L. 2005, Q. Rev. Biophys., 38, 49.
4. Taylor, G. K. and Stoddard, B. L. 2012, Nucleic Acids Res., 40, 5189.
5. Belfort, M. 2005, Homing Endonucleases and Inteins, M. Belfort, B. L. Stoddard, D. W. Wood and V. Derbyshire (Eds.), Springer-Verlag, New York, 1.
6. Mueller, J. E., Smith, D., Bryk, M. and Belfort, M. 1995, EMBO J., 14, 5724.
7. Derbyshire, V., Kowalski, J. C., Dansereau, J. T., Hauer, C. R. and Belfort, M. 1997, J. Mol. Biol., 265, 494.
8. Edgell, D. R. and Shub, D. A. 2001, Proc. Natl. Acad. Sci. USA, 98, 7898.
9. van Roey, P., Meehan, L., Kowalski, J. C., Belfort, M. and Derbyshire, V. 2002, Nat. Struct. Biol., 9, 806.
10. van Roey, P., Waddling, C. A., Fox, K. M., Belfort, M. and Derbyshire, V. 2001, EMBO J., 20, 3631.
11. Kleinstiver, B. P., Wolfs, J. M. and Edgell, D. R. 2013, Nucleic Acids Res., 41, 5413.
12. 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, EMBO J., 24, 885.
13. Ibryashkina, E. M., Zakharova, M. V., Baskunov, V. B., Bogdanova, E. S., Nagornykh, M. O., Den'mukhamedov, M. M., Melnik, B. S., Kolinski, A., Gront, D., Feder, M., Solonin, A. S. and Bujnicki, J. M. 2007, BMC Struct. Biol., 7, 48.
14. Kaminska, K. H., Kawai, M., Boniecki, M., Kobayashi, I. and Bujnicki, J. M. 2008, BMC Struct. Biol., 8, 48.
15. Lagerback, P., Andersson, E., Malmberg, C. and Carlson, K. 2009, Nucleic Acids Res., 37, 6174.
16. Brachner, A., Braun, J., Ghodgaonkar, M., Castor, D., Zlopasa, L., Ehrlich, V., Jiricny, J., Gotzmann, J., Knasmueller, S. and Foisner, R. 2012, J. Cell. Sci., 125, 1048.
17. Liu, X., Liu, S., Feng, Y., Liu, J-Z., Chen, Y., Pham, K., Deng, H., Hirschi, K. D., Wang, X. and Cheng, N. 2013, Proc. Natl. Acad. Sci. USA, 110, 9565.
18. Mak, A. N., Lambert, A. R. and Stoddard, B. L. 2010, Structure, 18, 1321.
19. Sokolowska, M., Czapinska, H. and Bochtler, M. 2011, Nucleic Acids Res., 39, 1554.
20. Gorbalenya, A. E. 1994, Protein Sci., 3, 1117.
21. Xu, S-Y. and Gupta, Y. 2012, Nucleic Acids Res., 41, 378.
22. Xu, S-Y., Kuzin, A. P., Seetharaman, J., Gutjahr, A., Chan, S-H., Chen, Y., Xiao, R., Acton, T. B., Montelione, G. T. and Tong, L. 2013, PLoS One, 9, e72114.

23. Kong, S., Liu, X., Fu, L., Yu, X. and An, C. 2012, PLoS One, 7, e43738.
24. Drouin, M., Lucas, P., Otis, C., Lemieux, C. and Turmel, M. 2000, Nucleic Acids Res., 28, 4566.
25. Robbins, J. B., Stapleton, M., Stanger, M. J., Smith, D., Dansereau, J. T., Derbyshire, V. and Belfort, M. 2007, Nucleic Acids Res., 35, 1589.
26. Shen, B. W., Landthaler, M., Shub, D. A. and Stoddard, B. L. 2004, J. Mol. Biol., 342, 43.
27. Magnani, E., Sjolander, K. and Hake, S. 2004, Plant Cell, 16, 2265.
28. Yusufzai, T. and Kadonaga, J. T. 2010, Proc. Natl. Acad. Sci. USA, 107, 20970.
29. Holloway, S. P., Deshpande, N. N. and Herrin, D. L. 1999, Curr. Genet., 36, 69.
30. Odom, O. W., Holloway, S. P., Deshpande, N. N., Lee, J. and Herrin, D. L. 2001, Mol. Cell. Biol., 21, 3472.
31. Kim, H. H., Corina, L. E., Suh, J. K. and Herrin, D. L. 2005, Protein Expr. Purif., 44, 162.
32. Corina, L. E. 2005, PhD Dissertation, University of Texas at Austin, Austin, TX.
33. Corina, L. E., Qiu, W., Desai, A. and Herrin, D. L. 2009, Nucleic Acids Res., 37, 5810.
34. Jain, S. S. and Tullius, T. D. 2008, Nat. Protoc., 3, 1092.
35. Tullius, T. D., Dombroski, B. A., Churchill, M. E. and Kam, L. 1987, Methods Enzymol., 155, 537.
36. Brenowitz, M., Senear, D. F., Shea, M. A. and Ackers, G. K. 1986, Methods Enzymol., 130, 132.
37. Landthaler, M., Shen, B. W., Stoddard, B. L. and Shub, D. A. 2006, J. Mol. Biol., 358, 1137.
38. Arnold, K., Bordoli, L., Kopp, J. and Schwede, T. 2006, Bioinformatics, 22, 195.
39. Mate, M. J. and Kleanthous, C. 2004, J. Biol. Chem., 279, 34763.
40. Contreras-Moreira, B., Branger, P. A. and Collado-Vides, J. 2007, Bioinformatics, 23, 1694.
41. deLano, W. L. 2002, The PyMOL Molecular Graphics System, DeLano Scientific, Palo Alto.
42. Kowalski, J. C., Belfort, M., Stapleton, M. A., Holpert, M., Dansereau, J. T., Pietrokovski, S., Baxter, S. M. and Derbyshire, V. 1999, Nucleic Acids Res., 27, 2115.
43. Saguez, C., Lecellier, G. and Koll, F. 2000, Nucleic Acids Res., 28, 1299.