SCIENTIFIC REP♦RTS

OPEN

Understanding the indirect DNA read-out specificity of I-CreI Meganuclease

Jesús Prieto[1], Pilar Redondo [2], Blanca López-Méndez[3], Marco D'Abramo [4], Nekane Merino [5], Francisco J. Blanco [5,6], Phillipe Duchateau[7], Guillermo Montoya[3] & Rafael Molina[3]

The high DNA specificity of homing endonucleases makes them a powerful protein scaffold to engineer enzymes for genome manipulation. Understanding their molecular recognition of DNA is an important prerequisite to generate engineered enzymes able to cleave DNA in specific desired genome sites. Protein-DNA recognition studies have been mostly focused on specific direct contacts between amino acid side chains and bases to redesign the binding interface. However, the important role of indirect readout in the central region of the target DNA of the homing endonuclease I-CreI suggested that indirect readout may play a key role in the redesign of protein-DNA interactions. The sequences of the I-CreI central substrate region, 2NN, along with the adjacent 5NNN, are key for substrate cleavage. Here, we analyse the mechanism of target discrimination at the 5NNN region by the I-CreI protein, revealing its critical role in the location and occupancy of the catalytic metal ions, which is crucial for cleavage. Our data highlight the importance of indirect readout for target DNA cleavage, thus aiding I-CreI engineering when targeting new DNA sequences.

Indirect readout in protein-DNA recognition is the mechanism by which the protein achieves partial sequence specificity by detecting structural features on the DNA. Therefore, indirect readout has been proposed to involve contacts mediated by water or other small molecules, as well as distortions of the DNA double helix, so that the protein can distinguish different sequences energetically[1]. Structural analysis of LAGLIDADG homing endo-nucleases (LHEs) bound to their targets revealed that the central region of the DNA target is kinked, result-ing in base twisting and unstacking near the scissile phosphate groups, thus allowing binding and position-ing in the active site[2–4]. I-CreI is a homodimeric member of LHEs family, which recognizes and cleaves a 22 bp pseudo-palindromic target (5′- CAAAACGTCGTGAGACAGTTTG -3′). Since different subsets of protein-DNA target contacts may be sufficient to maintain a high degree of sequence-specific homing site recognition and cleavage, some I-CreI-DNA target interactions may be altered and additional changes may be accommodated[5]. Thus, a 24 bp palindromic DNA target is recognized and cleaved by a I-CreI variant (I-CreI D75N) with similar affinity and activity than in the wild-type case[6,7]. Each I-CreI monomer binds its own DNA target region generat-ing the catalytic centre at the dimer interface. This region contains two catalytic aspartic acids (D20, one per each monomer). The aspartic side chains participate in the cleavage of the DNA strands along the minor groove, result-ing in the hydrolysis of specific phosphodiester bonds upon the coordination of three divalent metal ions[8]. The structure of I-CreI in complex with its target DNA shows that each monomer establishes direct interactions with the bases[2], grouped in three boxes called 5NNN located at positions ±3, ±4, ±5; the 7NN located at positions ±6, ±7[9] and 10NNN located at positions ±8, ±9, ±10[6,10] (Fig. 1a). The four base pairs (±1 and ±2), called 2NN,

[1]Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre (CNIO), Macromolecular Crystallography Group, c/Melchor Fdez. Almagro 3, 28029, Madrid, Spain. [2]Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre (CNIO), Cell Signalling and Adhesion Group, c/Melchor Fdez. Almagro 3, 28029, Madrid, Spain. [3]Protein Structure & Function Programme, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, 2200, Copenhagen, Denmark. [4]Department of Chemistry, University of Rome "La Sapienza", Piazzale Aldo Moro 5, 00185, Rome, Italy. [5]CIC bioGUNE, Parque Tecnológico de Bizkaia Edificio 800, 48160, Derio, Spain. [6]IKERBASQUE, Basque Foundation for Science, María Díaz de Haro 3, 48013, Bilbao, Spain. [7]CELLECTIS S.A., 8 rue de la Croix Jarry, 75013, Paris, France. Correspondence and requests for materials should be addressed to R.M. (email: rafael.molina@cpr.ku.dk)
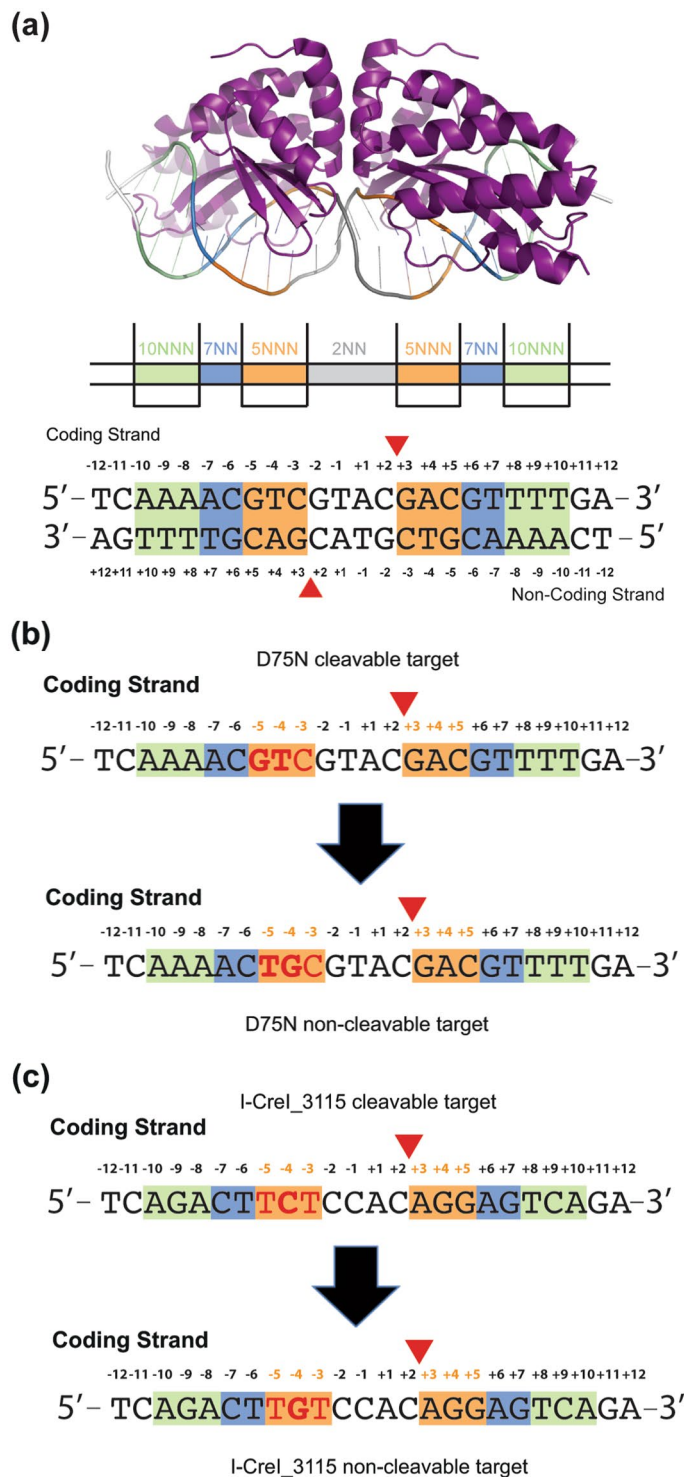
Figure 1. DNA target regions involved in I-CreI_D75N and I-CreI_3115 binding or catalysis. (**a**) The protein–DNA contacts are clustered in three nucleotide regions: 10NNN (± 8, ± 9, ± 10 boxed in green), 7NN (± 6, ± 7 boxed in blue) and 5NNN (± 3, ± 4, ± 5 boxed in orange). The 2NN region (± 2, ± 1) is located at the cleavage site (red triangles indicate the scissile P-O bonds). (**b**) Scheme of the design of I-CreI D75N and (**c**) I-CreI_3115 non-cleavable DNA targets based on their corresponding cleavable DNA targets.

containing the scissile phosphodiester bonds, show only one backbone contact between the nucleotide at position -1 (both strands) and K139 (of each I-CreI monomer)[11]. Changes in 2NN significantly affect substrate binding and cleavage. The influence of the central sequence was explained by its topology, showing a mechanism governing target discrimination not based on specific protein–DNA contacts[12]. Thus, the 2NN region affects the active site rearrangement, the proper protein-DNA complex binding and catalytic ion positioning to lead the cleavage.

Here we conduct a study of the impact of some 5NNN target bases composition in catalysis using biochemical, computational and structural analysis. The results suggest a mechanism controlling target discrimination not only based on specific protein-DNA contacts, but on the proper positioning of the catalytic ion. These findings, in line with previous reports[12] allow us to further rationalize the search for new target sequences in the development of new-engineered homing endonucleases for therapeutic and biotechnological applications.

Results

The DNA bases located at the 5NNN region are key for cleavage.

Previous studies have pointed out that I-CreI meganuclease has a preference for certain nucleotides at the 5NNN of its target sequence, and that a G in position −4 strongly hampers cleavage[6,9]. To decipher why a G at this position restricts cleavage, we made DNA substitutions in the cleavable DNA target to render it non-cleavable (Fig. 1b,c). For this purpose, we exchanged GT in positions −5 and −4 of the wild type coding strand target (I-CreI_D75N_target, Fig. 1b upper sequence) by TG, generating the non-cleavable target I-CreI_D75N_target-null (Fig. 1b bottom sequence). To broaden our analysis, we also studied the I-CreI heterodimeric variant I-CreI_3115 (Y33G/Q38K/Q44K/R68Y/R70S/ D75N/I77Y/I132V-G19S/Y33V/Q38R/S40Q/Q44D/R68A/R70S/D75K/I77R). This mutant was generated using the methodology previously described[6,13,14], and is able to recognize and cleave the human HBB (Haemoglobin beta subunit) gene (I-CreI_3115_target, Fig. 1c upper sequence) whose mutations cause sickle cell anaemia. C in position −4 of the coding strand target sequence was exchanged to G, generating the non-cleavable target I-CreI_3115_target-null (Fig. 1c bottom sequence). In vitro plasmid cleavage assays (Fig. 2a) revealed that both I-CreI_D75N and I-CreI_3115 proteins cleaved their targets with the same efficiency, despite the differences in base composition, amino acid sequences and contacts (see below). Then, to check how a G in position −4 affects target cleavage we performed an in vitro cleavage assay using labelled targets, resulting in an inhibition of cleavage when the G is in position −4 position in both DNA targets (Fig. 2b, Supplementary Fig. S1).

The DNA bases located at the 5NNN region are critical in substrate binding.

To analyse whether cleavage differences may arise from changes in binding between the wild type and null DNA targets, we performed MicroScale Thermophoresis (MST) experiments to assess the K_D between proteins and target DNAs. Exploratory binding assays using fluorescently labelled proteins or fluorescently labelled DNA targets yielded similar results (Supplementary Fig. S2, Supplementary Table S1). The MST measurements were then performed keeping constant the fluorescently labelled proteins and titrating the corresponding target DNAs in the presence of the non-catalytic cation Ca^{2+} (10 mM). This metal ion allows binding but hinders catalysis[12]. Our results indicate that there are significant differences in the K_D between cleavable and non-cleavable targets with both proteins, I-Cre_D75N and I-CreI_3115 (Fig. 2c,d, Supplementary Table S2), providing a possible explanation about cleavage differences at the concentration sampled.

The 5NNN non-cleavable sequences prevent the central metal positioning in the catalytic site.

To understand the molecular mechanism that inhibits cleavage when a G is at −4 position flanked by a pyrimidine at position −5, we solved the crystal structures of the cleavable and the corresponding non-cleavable variant targets in the presence of catalytic (Mg^{2+}) and non-catalytic (Ca^{2+}) cations (Supplementary Table S2, Supplementary Table S4). As observed in our previous studies concerning the central 2NN target region[12], the absence of cleavage may arise not only from binding differences but also from changes of the DNA structure at the 5NNN region. To check this point, we analysed a superimposition of the crystal structures of both I-CreI variants bound to their target DNA sequences in the non-cleaved state (Supplementary Fig. S3). In both cases subtle protein-DNA interaction changes occurred at the 5NNN region modified, but they did not alter the conformation of the active site, suggesting that cleavage differences were not due to alterations of the catalytic centre configuration. After discarding DNA conformational changes at 5NNN region that could have affected the configuration of the active site, we analysed the crystal structures of both I-CreI_D75N and I-CreI_3115 proteins in complex with cleavable and non-cleavable DNA target sequences and in the presence of the catalytic ion Mg^{2+} in order to decipher the molecular basis of this different behaviour. As expected from in vitro cleavage experiments, the comparison of the crystal structures of the I-CreI variants in complex with its cleavable and non-cleavable target DNAs in favourable catalytic conditions (2 mM Mg^{2+}), showed the hydrolysed phosphodiesters in the cleavable targets (Fig. 3a,b left panels) while the non-cleavable targets displayed the intact phosphodiester bonds (Fig. 3a,b right panels). Noteworthy, the comparison of the isomorphous signal of the electron density maps of the I-CreI variants, suggested a different number of catalytic ions in the active site for the cleavable and non-cleavable targets, even though the configuration of the active sites was similar. We observed that the structures of the variants in complex with the cleavable targets showed the presence of the 3 reported catalytic ions[15], while the structures in complex with the non-cleavable targets indicated the presence of just 2 catalytic ions, with the central metal position occupied by a water molecule. The absence of the central metal ion in the non-cleaved structures[16], would explain the different catalytic behaviour. To assess unambiguously the number of catalytic ions present in the active site, we solved the crystal structure of the two proteins in complex with both target DNAs in the presence of Mn^{2+}, which allows catalysis and unambiguous detection and location through its anomalous diffraction signal. In addition, Mn^{2+} is less restrictive than Mg^{2+}[17] allowing phosphoryl transfer reactions in non-ideal substrates[18], including the non-cleavable targets studied here. In-depth analysis showed differences in the occupancy of the central metal at the active site between cleavable (Fig. 3c,d left panels) and non-cleavable targets (Fig. 3c,d right panels). A high occupancy level of the central metal ion is essential for the cleavage[16], and the lower occupancy level of the central Mn^{2+} ion found in the non-cleavable targets (46% and 74% for I-CreI_D75N_target-null and I-CreI_3115_target-null, respectively) compared with the cleavable ones (100% in both cases; see Supplementary Table S2) revealed that the non-cleavable targets interfere with the positioning of the central metal ion affecting catalysis.
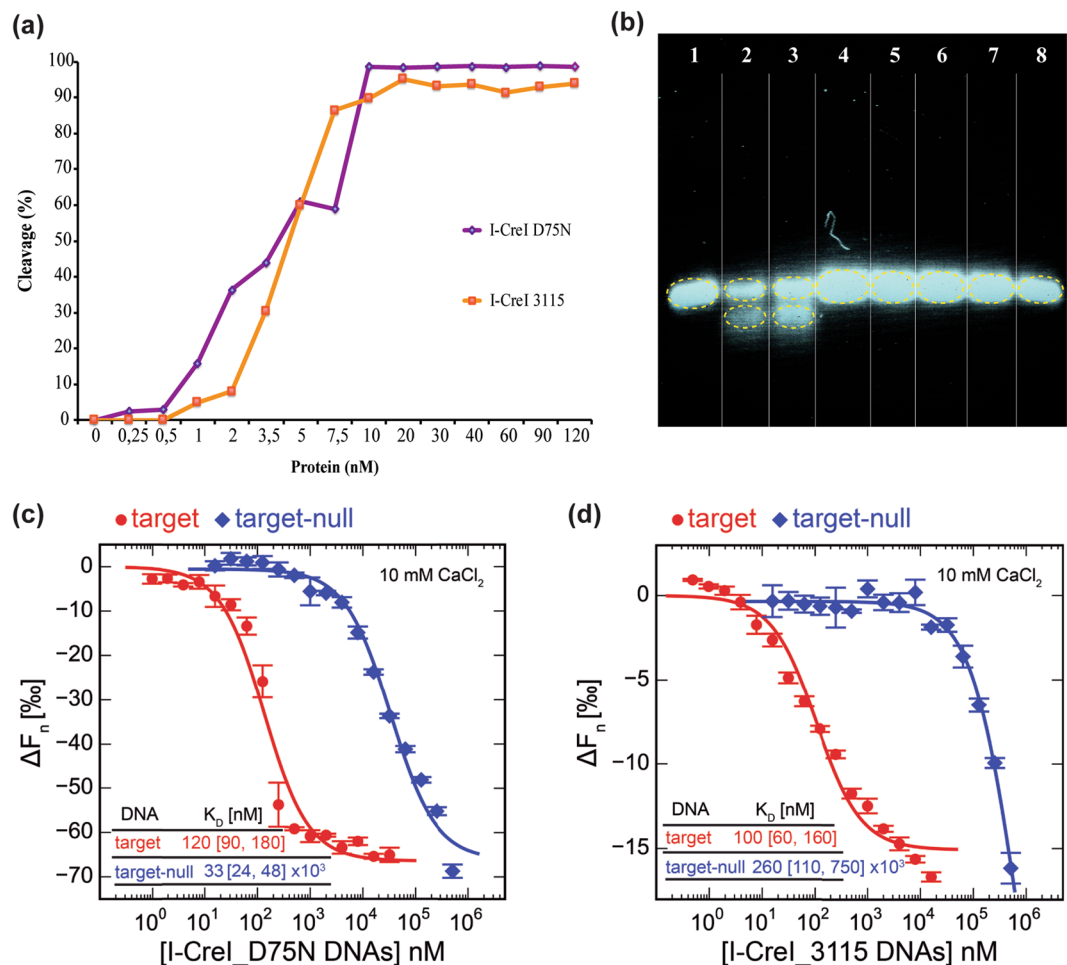
Figure 2. I-CreI *in vitro* cleavage and binding experiments. (**a**) *In vitro* cleavage titration analysis for I-CreI_D75N_target and I-CreI_3115_target using linearized plasmids containing the cleavable target sequences. (**b**) Validating 5NNN non-cleavable target patterns using labelled duplex DNA targets. Lanes: **1**, I-CreI_D75N_target; **2**, I-CreI_D75N_target in the presence of I-CreI_D75N protein; **3**, I-CreI_3115_target in the presence of I-CreI_3115 protein; **4**, I-CreI_3115_target; **5**, I-CreI_D75N_target-null (non-cleavable derivative); **6**, I-CreI_D75N_target-null in the presence of I-CreI_D75N protein; **7**, I-CreI_3115_target-null (non-cleavable derivative) in the presence of I-CreI_3115 protein; **8**, I-CreI_3115_target-null (non-cleavable derivative). (**c**) I-CreI_D75N/I-CreI_D75N_target or target-null binding by microscale thermophoresis (in the presence of 10 mM CaCl$_2$). (**d**) I-CreI_3115/I-CreI_3115_target or target-null binding by microscale thermophoresis (in the presence of 10 mM CaCl$_2$). The change in thermophoresis is plotted against the concentration of DNA in a logarithmic scale and fitted to a 1:1 binding model to yield the corresponding K_Ds. Numbers in brackets represent the 68.3% confidence interval calculated by ESP (error-surface projection).

Molecular dynamics simulations reveal a link between 5NNN region and key Adenine +1.

To shed light on possible dynamic-conformational effects involved in the DNA cleavage process, we performed MD simulations of the I-CreI_3115 and I-CreI_D75N bound to the different DNA target sequences (Supplementary Table S3). As reported in our previous works[12,16,19], the cleavage activity of this class of enzymes can be rationalized in terms of both the proper arrangement of the active site (including water, ions and DNA) and the correct perturbation provided by the protein. Therefore, we compared here the DNA conformational dynamics near the cleavage site. Our results clearly point out that in the cases of I-CreI_D75N_target-null and I-CreI_3115_target-null sequences bound to the I-CreI_D75N and I-CreI_3115 proteins, A at position +1 largely deviates with respect to the usual B-DNA conformation (Fig. 4). In particular, we found that in ~ 30% (I-CreI_D75N_target-null, Fig. 4a) and ~ 60% (I-CreI_3115_target-null, Fig. 4b) of the simulation time A +1 is not paired and is found stacked to the corresponding nucleobases. On the other hand, when the two enzyme variants were simulated bound to the I-CreI_D75N_target and I-CreI_3115_target DNA sequences, the helical parameters of A +1 were similar to those found in the ideal B-DNA conformation, with minor fluctuations typical of atomic thermal motions (Fig. 4a,b). These results suggest that the DNA sequences in I-CreI_D75N_target-null and I-CreI_3115_target-null allow for a larger conformational flexibility to A +1, making the proper positioning of the P-O bond less frequent as compared with the DNA sequence in I-CreI_D75N_target and I-CreI-3115_target. Hence, although in all cases the DNA deviates from the ideal B-DNA conformation, their sequences in the
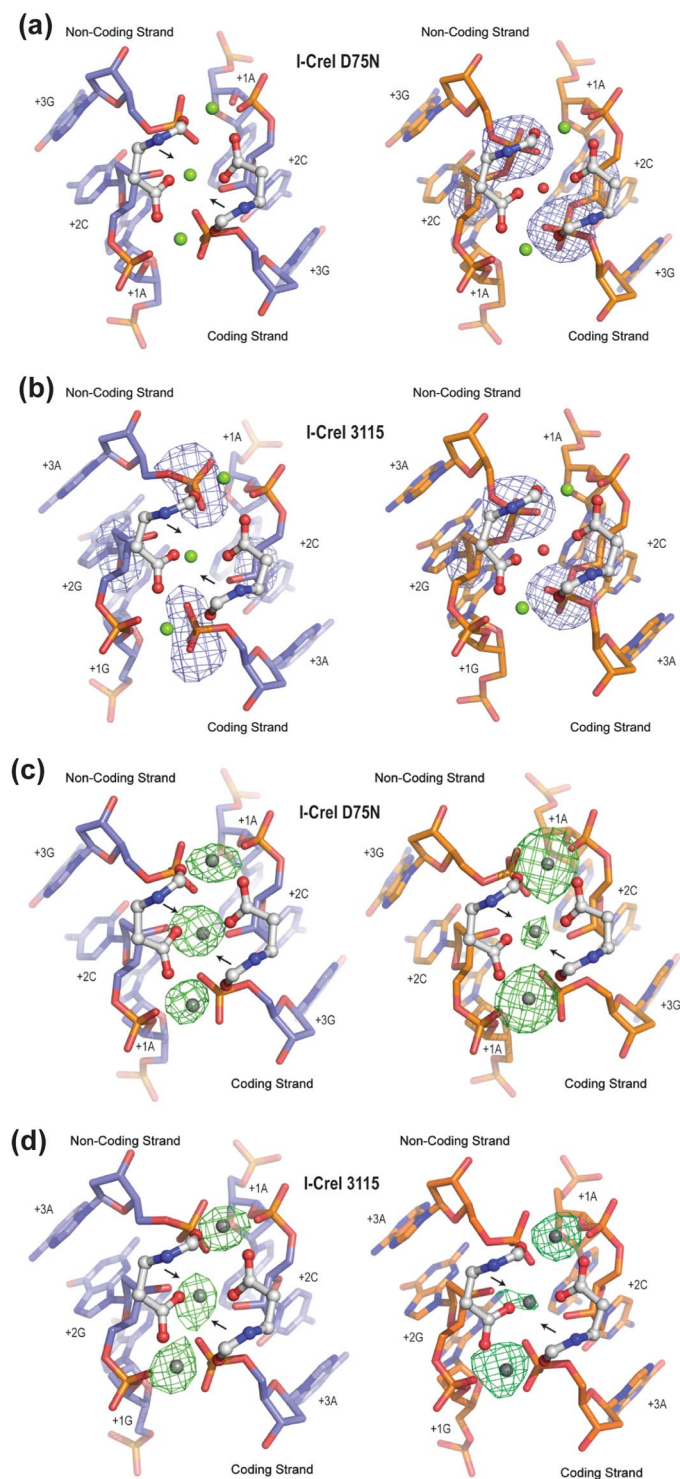
Figure 3. Active site detailed view of both I-CreI variants in complex with cleavable and non-cleavable targets. (**a**) Cleavage analysis of I-CreI_D75N_target (left panel) vs I-CreI_D75N_target-null (right panel) and (**b**) I-CreI_3115_target (left panel) vs I-CreI_3115_target-null (right panel). Omit map ($\sigma = 6.0$) around the cleavable phosphodiester revealed no cleavage in the corresponding target-null and phosphodiester cleavage with their original targets. The green spheres represent the Mg^{2+} ions and the red sphere a water molecule. (**c**) Metal positioning analysis of I-CreI_D75N_target (left panel) vs I-CreI_D75N_target-null (right panel) and (**d**) I-CreI_3115_target (left panel) vs I-CreI_3115_target-null (right panel). Anomalous maps ($\sigma = 6.0$) around the active site revealed the presence of Mn^{2+} (grey spheres) as a catalytic ion with different occupancies depending on the target to cleave. Black arrows indicate the phosphodiester bond cleaved.
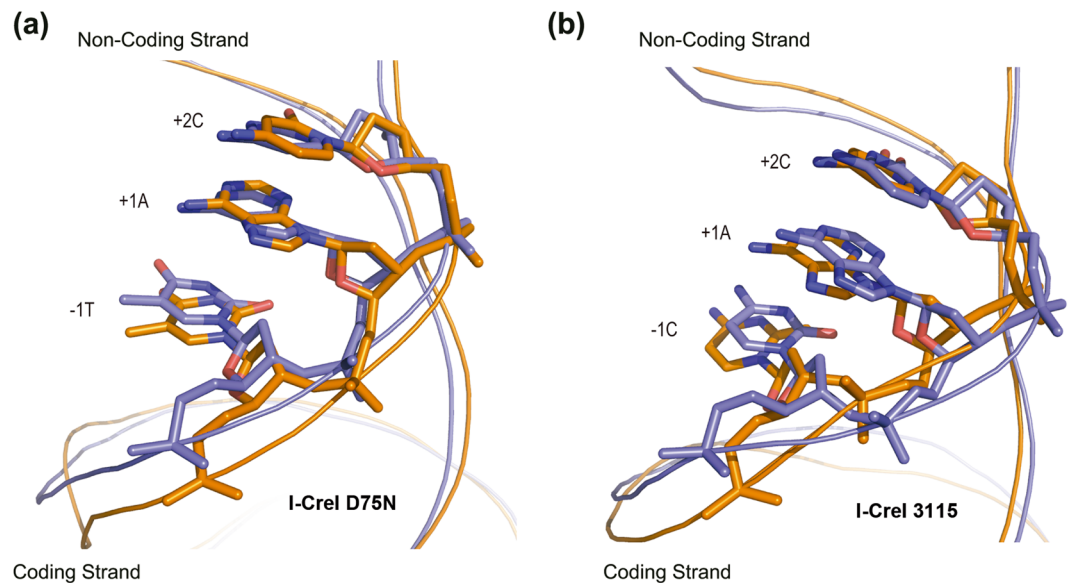
Figure 4. Zoom view at the Adenine +1 remarking the stacking changes between cleavable and non-cleavable targets. (**a**) Structure comparison between I-CreI_D75N_target (blue), where Adenine follows the ideal B-DNA geometry, vs I-CreI_D75N_target-null (orange), where Adenine +1 largely is deviated with respect to the ideal B-DNA geometry. (**b**) Structure comparison between I-CreI_3115_target (blue), where Adenine follows the ideal B-DNA geometry, vs I-CreI_3115_target-null (orange), where Adenine +1 largely is deviated with respect to the ideal B-DNA geometry.

I-CreI_D75N_target-null and I-CreI_3115_target-null are much more flexible, thus probably sampling wider regions of the conformational space not compatible with the phosphorous-oxygen bond break.

Discussion

The specificity of I-CreI, a widely-redesigned enzyme with therapeutic and biotechnological purposes, strongly depends on indirect readout to recognize and cleave its target sequence. So far, the redesign of the protein-DNA binding properties in this scaffold has been performed taking into account the specific protein-DNA contacts. In this sense, understanding of the 2NN region role in target recognition and cleavage had a strong impact in meganuclease engineering targeting new DNA sequences that avoid the presence of the non-preferred bases in the central region, thus optimizing meganuclease tailoring. Here we provide evidence that the location of a guanine at position −4 (Fig. 1) of the target DNA, in the 5NNN region, strongly affects the conformational dynamics of the adenine at position +1 in the active site independently of the target sequence or protein variant. The larger flexibility of the A +1 in the active centre explains the low occupancy of the central metal ion within the active site, disturbing the positioning of the P-O bond for catalysis, thus inhibiting cleavage. This correlation between bases at −4 and +1 positions of the target depends mainly on DNA sequence composition, thus adding new restrictions to those imposed by the 2NN[12] (Fig. 1).

Hence, the molecular assembly of the protein-DNA complex is strongly dependent on the DNA sequence in the 5NNN region, regardless of the specific contacts arising from the interaction between the protein and other DNA regions. This suggests the need of a restrictive sequence induced conformation around the active site to allow the formation of a ternary protein-DNA-metal complex competent for cleavage. Our results highlight the importance of the indirect readout of the DNA target in the I-CreI scaffold. The combination of these restrictions in the 2NN and 5NNN regions provide a new scenario for the selection of adequate sequences to be recognized for a given I-CreI variant in genome editing.

Methods

Protein expression and purification. The homodimeric I-CreI_D75N variant was cloned and expressed as reported[20]. The co-expression, purification and storage of the heterodimeric I-CreI_3115 variant (Y33G/Q38K/Q44K/R68Y/R70S/D75N/I77Y/I132V-G19S/Y33V/Q38R/S40Q/Q44D/R68A/R70S/D75K/I77R) was carried out as described[13]. Both proteins are folded and share biophysical properties (oligomerization status) with the respective wild type proteins as assessed by circular dichroism and SEC-MALS.

***In vitro* cleavage assay conditions.** Plasmid cleavage assays were performed at 37 °C in 10 mM Tris-HCl (pH 8), 50 mM NaCl, 10 mM MgCl$_2$ and 1 mM DTT. 100 nM of the linearized plasmid target (3-Kb) and 120-0.25 nM dilutions of I-CreI_D75N and I-CreI_3115 proteins were used reaching 25 μl of final reaction volume. Reactions were stopped after 1 hour by addition of 5 μl of 45% Glycerol, 95 mM EDTA (pH 8), 1.5% (w/v) SDS, 1.5 mg/ml Proteinase K and 0.048% (w/v) Bromophenol blue (6x Buffer Stop), incubated at 37 °C for 30 minutes and electrophoresed in a 1% agarose gel[20]. The gels were stained with SYBR Safe DNA gel staining kits (Invitrogen) and the intensity of the bands observed upon illumination with UV light was determined with the

ImageJ software (http://rsb.info.nih.gov/ij/). The linearized target plasmid was 3 kb in size and yielded two smaller bands, of 2 kb and 1 kb, upon cleavage with the meganuclease. The percentage of cleavage was calculated with the following equation: % cleavage $= 100 \times (I_{2kb} + I_{1kb})/(I_{3kb} + I_{2kb} + I_{1kb})$, where I_{1kb}, I_{2kb} and I_{3kb} are the intensities of the 1, 2 and 3 kb bands, respectively.

Double-stranded oligonucleotide cleavage assays were performed at 37 °C in 10 mM Tris-HCl (pH 8), 50 mM NaCl, 10 mM MgCl$_2$ and 1 mM DTT, in 10 ml of final reaction volume. Concentrations were: 1 mM for the 6 FAM-labelled DNA duplexes (Supplementary Table S3), and 2 μM for I-CreI_D75N and I-CreI_3115 proteins. Reactions were stopped after 30 minutes by addition of 2 μl of 45% Glycerol, 95 mM EDTA (pH 8), 1.5% (w/v) SDS, 1.5 mg/ml Proteinase K and 0.048% (w/v) Bromophenol blue, incubated at 37 °C for 30 minutes and electrophoresed in a 3.5% agarose gel.

The D75N mutation of the I-CreI scaffold does not affect protein structure and facilitates the enzyme purification. As it has been reported previously, I-CreI and its D75N variant display similar *in vitro* activities and levels of specificity[6].

Binding Assays by MicroScale Thermophoresis (MST). MST experiments were performed on a NanoTemper Monolith NT.115 instrument with blue/red channels. Most of the experiments were performed using fluorescently labelled proteins and titrating with the corresponding DNAs, except for the I-CreI_3115/I-CreI_3115_target interaction which was also performed titrating the fluorescently labelled DNA with increasing concentrations of protein. Both labelling approaches gave similar results (Supplementary Table S1). Cy5-labeled I-CreI_3115_target oligo-nucleotide (coding strand) was purchased from TAG Copenhagen A/S (Copenhagen, Denmark) and annealed with its complementary oligonucleotide prior to MST experiments. I-CreI_3115 and I-CreI_D75N were labelled using the Monolith NT Protein Labelling Kit RED-NHS (NanoTemper Technologies GmbH) according to the supplied protocol. The fluorophore-conjugated protein and the extent of labelling were determined by spectrophotometry. MST experiments were performed at 25 °C in 50 mM Tris pH 8.0, 300 mM NaCl, 2 or 10 mM CaCl$_2$, 0.5 mM TCEP, 0.05% Tween-20 (in the case of I-CreI_3115) or in 50 mM Hepes pH 7.5, 300 mM NaCl, 10 mM CaCl$_2$, 0.5 mM TCEP, 0.05% Tween-20 for the I-CreI_D75N protein. In all cases, premium coated capillaries were used and measurements were performed at 40% MST power. Laser off/on times were 5 s and 30 s, respectively. The fluorescently labelled DNA or proteins were used at 25 nM concentration. The fluorescently labelled partner was mixed with equivalent volumes of a two-fold serial dilution of the different ligands starting at concentrations of 16 μM or 512 μM (for the non-cleavable DNA variants). The fraction of bound protein was derived from the ratio of the averaged normalized fluorescence at later (2.0 s prior to switching off the IR-laser) and early times (0.5 s after switching on the IR-laser) of the thermophoretic time traces at different ligand concentrations. The K_Ds were obtained by fitting the fraction of bound protein to the quadratic solution of the binding reaction equilibrium derived from the law of mass action. The number of independent repeats was 3 for all measurements.

Crystallization. The 24-mer DNA oligonucleotides were purchased from Proligo and sequences are described at Supplementary Table S3. The DNA substrates were generated after mixture of equimolar amounts of oligonucleotides in 10 mM Tris-HCl pH 8. All of them form a 24-bp blunt-ended duplexes after incubation for 5 minutes at 99 °C and slowly cooling down to 25 °C, with the reverse complementary oligonucleotide. I-Cre_D75N_target is palindromic and thus self-annealing. The protein–DNA complexes were: I-CreI_D75N:I-CreI_D75N_target:Ca^{2+}, I-CreI_D75N:I-CreI_D75N_target:Mn^{2+}, I-CreI_D75N:I-CreI_D75N_target:Mg^{2+}, I-CreI_D75N:I-CreI_D75N_target-null:Mn^{2+}, I-CreI_D75N:I-CreI_D75N_target-null:Mg^{2+}, I-CreI_3115:I-CreI_3115_target:Ca^{2+}, I-CreI_3115:I-CreI_3115_target:Mn^{2+}, I-CreI_3115:I-CreI_3115_target-3115:Mg^{2+}, I-CreI_3115:I-CreI_3115_target-null:Mn^{2+}, I-CreI_3115:I-CreI_3115_target-null:Mg^{2+} obtained in the presence of 2 mM metal ion. Complexes were formed by pre-warming the meganuclease and the oligonucleotide samples at 37 °C and mixing them in a 0.75:1 molar ratio (DNA:protein). The mixture was incubated for 50 minutes at this temperature, and then spun down for 5 minutes to remove insoluble material. The final concentration of protein in the DNA–protein complex solution was 4 mg/ml.

All protein-DNA complex crystals were grown in the presence of 2 mM Ca^{2+}, Mg^{2+} or Mn^{2+} using the hanging-drop method at 290 K, in 2 μl droplets formed by 1 μl of the DNA–protein complex and 1 μl of precipitant solution consisting of 0.05–0.1 M calcium acetate, 0.1 M sodium acetate pH 4.6–5.4, 33–40% (v/v) 1,2-propanediol. The complexes were cryo-protected by adding 20% (v/v) glycerol to the mother liquor before flash-frozen in liquid nitrogen.

Data collection, structure solution, model building and refinement. All data were collected from frozen crystals at 100 K with PILATUS detectors at beamlines PXI (SLS, Villigen, Switzerland) and XALOC (ALBA Synchrotron, Barcelona, Spain). Data processing and scaling were accomplished with XDS[21] and Scala from the CCP4 package[22]. Statistics for the crystallographic data and structure solution are summarized in Supplementary Table S2. The structures were solved by molecular replacement, as implemented in the program PHASER[23]. The search models were based on the PDB entries 1G9Y (I-CreI:DNA:Ca^{2+}) and 1G9Z (I-CreI:DNA:Mg^{2+}). The models were then subjected to iterative cycles of model building and refinement with Coot[24] and PHENIX[25]. The identification and analysis of the protein–DNA hydrogen bonds and van der Waals contacts was done with the Protein Interfaces, Surfaces and Assemblies service PISA at the European Bioinformatics Institute (http://www.ebi.ac.uk/msdsrv/prot_int/pistart.html). DNA structures were analysed using 3DNA[26].

Molecular Dynamics Simulations. The molecular dynamics (MD) simulations lasting 100 ns each were performed using the Gromacs software v 5.0, starting from the experimental crystal structures using the

non-cleaved constructs and properly substituting the DNA sequence. We utilized the amber99sb force field[27] for the simulation of the four systems (I-CreI_D75N and I-CreI_3115 variants bound to their DNA targets and to non-cleavable DNA sequences (null): I-CreI_D75N:I-CreI_D75N_target, I-CreI_D75N:I-CreI_D75N_target-null, I-CreI_3115:I-CreI_3115_target, I-CreI_3115:I-CreI_3115_target-null). The systems were solvated using the SPC water model. After usual minimization, thermalization and equilibration steps, the systems were simulated using an integration step of 2 fs with periodic boundary conditions and Particle Mesh Ewald method for the calculation of the electrostatics. The pressure and temperature were kept constant by means of the Parrinello-Rahman barostat[28]. (P = 1 bar) and the isokinetic temperature coupling algorithm (T = 310 K, the same temperature of the experimental cleavage assays).

Data availability.

References

1. Garvie, C. W. & Wolberger, C. Recognition of specific DNA sequences. *Mol Cell* **8**, 937–946 (2001).
2. Jurica, M. S., Monnat, R. J. Jr. & Stoddard, B. L. DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-CreI. *Mol Cell* **2**, 469–476 (1998).
3. Marcaida, M. J., Munoz, I. G., Blanco, F. J., Prieto, J. & Montoya, G. Homing endonucleases: from basics to therapeutic applications. *Cell Mol Life Sci* **67**, 727–748, https://doi.org/10.1007/s00018-009-0188-y (2010).
4. Stoddard, B. L. Homing endonuclease structure and function. *Q Rev Biophys* **38**, 49–95, https://doi.org/10.1017/S0033583505004063 (2005).
5. Argast, G. M., Stephens, K. M., Emond, M. J. & Monnat, R. J. Jr. I-PpoI and I-CreI homing site sequence degeneracy determined by random mutagenesis and sequential *in vitro* enrichment. *Journal of molecular biology* **280**, 345–353, https://doi.org/10.1006/jmbi.1998.1886 (1998).
6. Arnould, S. *et al.* Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *Journal of molecular biology* **355**, 443–458, https://doi.org/10.1016/j.jmb.2005.10.065 (2006).
7. Chevalier, B., Turmel, M., Lemieux, C., Monnat, R. J. Jr. & Stoddard, B. L. Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-CreI and I-MsoI. *Journal of molecular biology* **329**, 253–269 (2003).
8. Chevalier, B. *et al.* Metal-dependent DNA cleavage mechanism of the I-CreI LAGLIDADG homing endonuclease. *Biochemistry* **43**, 14015–14026 (2004).
9. Grizot, S., Duclert, A., Thomas, S., Duchateau, P. & Paques, F. Context dependence between subdomains in the DNA binding interface of the I-CreI homing endonuclease. *Nucleic Acids Res* **39**, 6124–6136, https://doi.org/10.1093/nar/gkr186 (2011).
10. Arnould, S. *et al.* Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *Journal of molecular biology* **371**, 49–65, https://doi.org/10.1016/j.jmb.2007.04.079 (2007).
11. Ulge, U. Y., Baker, D. A. & Monnat, R. J. Jr. Comprehensive computational design of mCreI homing endonuclease cleavage specificity for genome engineering. *Nucleic Acids Res* **39**, 4330–4339, https://doi.org/10.1093/nar/gkr022 (2011).
12. Molina, R. *et al.* Non-specific protein-DNA interactions control I-CreI target binding and cleavage. *Nucleic Acids Res* **40**, 6936–6945, https://doi.org/10.1093/nar/gks320 (2012).
13. Redondo, P. *et al.* Molecular basis of xeroderma pigmentosum group C DNA recognition by engineered meganucleases. *Nature* **456**, 107–111, https://doi.org/10.1038/nature07343 (2008).
14. Smith, J. *et al.* A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res* **34**, e149, https://doi.org/10.1093/nar/gkl720 (2006).
15. Chevalier, B. S., Monnat, R. J. Jr. & Stoddard, B. L. The homing endonuclease I-CreI uses three metals, one of which is shared between the two active sites. *Nat Struct Biol* **8**, 312–316, https://doi.org/10.1038/86181 (2001).
16. Molina, R. *et al.* Visualizing phosphodiester-bond hydrolysis by an endonuclease. *Nat Struct Mol Biol* **22**, 65–72, https://doi.org/10.1038/nsmb.2932 (2015).
17. Yang, W., Lee, J. Y. & Nowotny, M. Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity. *Mol Cell* **22**, 5–13, https://doi.org/10.1016/j.molcel.2006.03.013 (2006).
18. Vaisman, A., Ling, H., Woodgate, R. & Yang, W. Fidelity of Dpo4: effect of metal ions, nucleotide selection and pyrophosphorolysis. *EMBO J* **24**, 2957–2967, https://doi.org/10.1038/sj.emboj.7600786 (2005).
19. Molina, R. *et al.* Engineering a Nickase on the Homing Endonuclease I-DmoI Scaffold. *J Biol Chem* **290**, 18534–18544, https://doi.org/10.1074/jbc.M115.658666 (2015).
20. Prieto, J. *et al.* The C-terminal loop of the homing endonuclease I-CreI is essential for site recognition, DNA binding and cleavage. *Nucleic Acids Res* **35**, 3262–3271, https://doi.org/10.1093/nar/gkm183 (2007).
21. Kabsch, W. X. *Acta Crystallogr D Biol Crystallogr* **66**, 125–132, https://doi.org/10.1107/S0907444909047337 (2010).
22. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**, 72–82, https://doi.org/10.1107/S0907444905036693 (2006).
23. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674, https://doi.org/10.1107/S0021889807021206 (2007).
24. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501, https://doi.org/10.1107/S0907444910007493 (2010).
25. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221, https://doi.org/10.1107/S0907444909052925 (2010).
26. Lu, X. J. & Olson, W. K. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat Protoc* **3**, 1213–1227, https://doi.org/10.1038/nprot.2008.104 (2008).
27. Perez, A. *et al.* Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J* **92**, 3817–3829, https://doi.org/10.1529/biophysj.106.097782 (2007).
28. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys* **52**, 9 (1981).

Acknowledgements

Author Contributions

J.P. performed the *in vitro* cleavage assays. P.R. carried out the crystallization experiments. B.L.M. accomplished the binding assays. M.D. run the molecular dynamics simulations. N.M. did the protein expression and purification. R.M. performed the X-ray data collection crystal data processing, model building and refinement. J.P. and R.M. designed and implemented the research, evaluated the results of the work and wrote the manuscript. F.J.B. and G.M. discussed the data and revised the manuscript. P.D. provided the clones for protein expression and purification.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-28599-0.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

SUPPLEMENTARY INFORMATION

Understanding the indirect DNA read-out specificity of I-CreI Meganuclease

Jesús Prieto[1], Pilar Redondo[2], Blanca López-Méndez[3], Marco D'Abramo[4], Nekane Merino[5], Francisco J Blanco[5,6], Phillipe Duchateau[7], Guillermo Montoya[3], Rafael Molina[3,] *

[1] Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre (CNIO), Macromolecular Crystallography Group, c/Melchor Fdez. Almagro 3, 28029 Madrid, Spain

[2] Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre (CNIO), Cell Signalling and Adhesion Group, c/Melchor Fdez. Almagro 3, 28029 Madrid, Spain

[3] Protein Structure & Function Programme, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, 2200 Copenhagen, Denmark.

[4] Department of Chemistry, University of Rome "La Sapienza," Piazzale Aldo Moro 5, 00185, Rome, Italy.

[5] CIC bioGUNE, Parque Tecnológico de Bizkaia Edificio 800, 48160 Derio, Spain.

[6] IKERBASQUE, Basque Foundation for Science, María Díaz de Haro 3, 48013 Bilbao, Spain.

[7] CELLECTIS S.A., 8 rue de la Croix Jarry, 75013 Paris, France

* To whom correspondence should be addressed. Email: rafael.molina@cpr.ku.dk
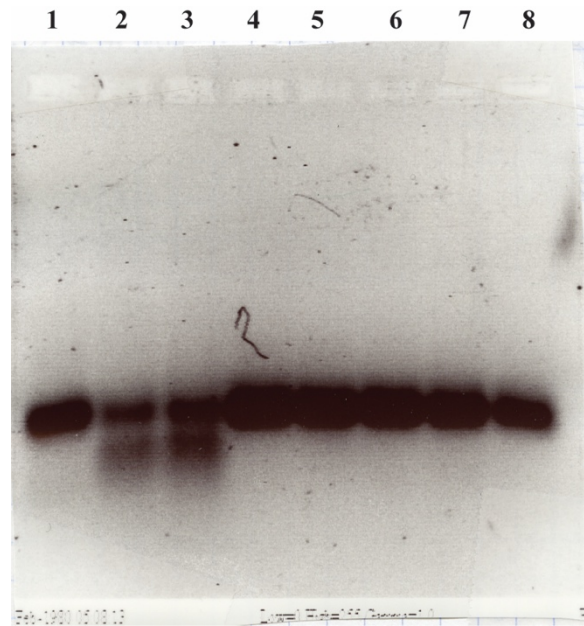
Supplementary Figure S1. Full-length gel from the I-CreI *in vitro* cleavage experiments to validate 5NNN non-cleavable target patterns using labelled duplex DNA targets. Lanes: **1**, I-CreI_D75N_target; **2**, I-CreI_D75N_target in the presence of I-CreI_D75N protein; **3**, I-CreI_3115_target in the presence of I-CreI_3115 protein; **4**, I-CreI_3115_target; **5**, I-CreI_D75N_target-null (non-cleavable derivative); **6**, I-CreI_D75N_target-null in the presence of I-CreI_D75N protein; **7**, I-CreI_3115_target-null (non-cleavable derivative) in the presence of I-CreI_3115 protein; **8**, I-CreI_3115_target-null (non-cleavable derivative).
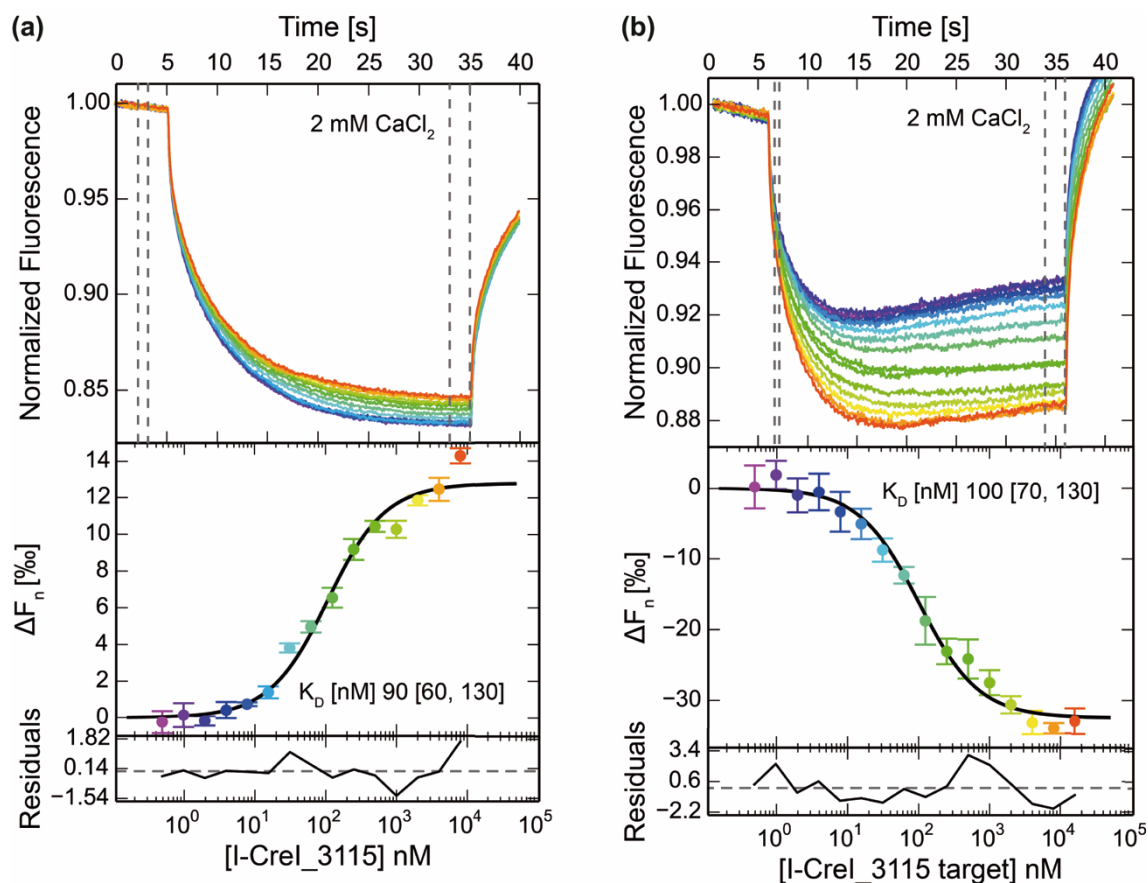
Supplementary Figure S2. Thermophoresis data for the I-CreI_3115/DNAs interactions in the presence of 2 mM CaCl₂. I-CreI_3115 and DNA target interaction using complementary labelling schemes, i.e., fluorescently labeled DNA duplex in **(a)** or fluorescently labeled protein in **(b)**. The upper panel shows the thermophoresis time traces with the discontinuous lines boundary regions used to calculate the observed thermophoresis related to the free- and bound-biomolecule populations. The fluorescence is normalized to a starting value of 1.0. The binding curve data is shown in the middle panel, each point is the average of three experiments and the error bars show the standard deviation. The line is the best fit to a 1:1 binding model. The bottom panel shows residuals between the fit and experimental data.
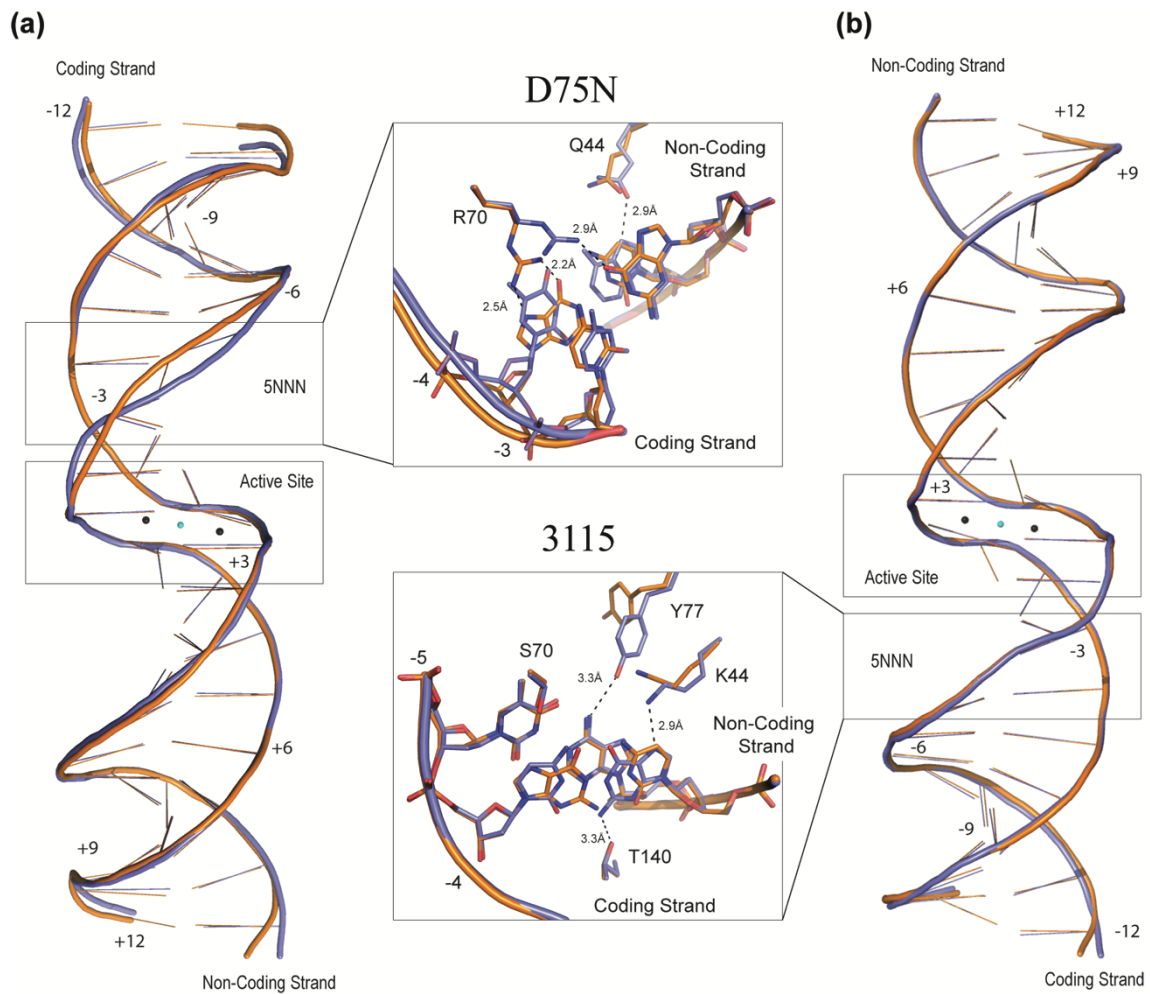
Supplementary Figure S3. DNA structure superimposition of cleavable (blue) and their corresponding non-cleavable (orange) DNA targets in their non-cleavage states including zoom views at the 5NNN region. **(a)** I-CreI_D75N_target vs I-CreI_D75N_target-null DNA target. **(b)** I-CreI_3115_target vs I-CreI_3115_target-null DNA target. The black spheres represent the metal ions and the blue sphere a water molecule.

Supplementary Table S1. *Meganuclease binding affinity for DNA targets measured by microscale thermophoresis.* [a]Cy5-labelled I-CreI-3115 DNA was used at a 25 nM concentration. [b,c]I-CreI_3115 and I-CreI_D75N were fluorescently labelled and used at a 25 nM concentration. [d]Numbers in brackets represent the 68.3% confidence interval for the fitted K_Ds calculated by ESP (error-surface projection).

Fluorescent labelled partner	Ligand	K_D [nM]	$CaCl_2$ [mM]
I-CreI_3115_target[a]	I-CreI_3115	90 [60, 130][d]	2
I-CreI_3115_target[a]	I-CreI_3115	110 [70, 160][d]	10
I-CreI_3115[b]	I-CreI_3115_target	100 [70, 130][d]	2
I-CreI_3115[b]	I-CreI_3115_target	100 [60, 160][d]	10
I-CreI_3115[b]	I-CreI_3115_target-null	500 [200, 1200][d] x 10^3	2
I-CreI_3115[b]	I-CreI_3115_target-null	260 [110, 750][d] x 10^3	10
I-CreI_D75N[c]	I-CreI_D75N_target	120 [90, 180][d]	10
I-CreI_D75N[c]	I-CreI_D75N_target-null	33 [24, 48][d] x 10^3	10

Supplementary Table S2. *Data collection and refinement statistics.*

	I-CreI3115: I-CreI_ta-3115: Ca^{2+}	I-CreI3115: I-CreI_ta-3115: Mg^{2+}	I-CreI3115: I-CreI_ta-3115: Mn^{2+}
PDB code	6FB0	6FB1	6FB2
Data collection			
Space group	P2$_1$	P2$_1$	P2$_1$
Cell dimensions			
a, b, c (Å)	45.44, 67.67, 83.53	45.10, 66.66, 91.39	44.89, 67.36, 90.87
α, β, γ (°)	90, 96.75, 90	90, 96.80, 90	90, 96.80, 90
Wavelength	0.87	1.00	1.60
Resolution (Å)	37.81-2.15 (2.27-2.15)*	66.66-3.00 (3.15-3.00)*	44.57-2.95 (3.11-2.95)*
R_{merge}	0.07 (0.46)	0.08 (0.47)	0.08 (0.50)
Mean I / σI	10.5 (3.0)	5.4 (1.6)	9.4 (2.3)
Completeness (%)	100 (99.9)	99.7 (98.4)	97.9 (96.6)
Redundancy	4.2 (4.2)	2.4 (2.4)	3.5 (3.4)
Refinement			
A site metal occupancy	100 (Ca^{2+})	100 (Mg^{2+})	100 (Mn^{2+})
B site metal occupancy	100 (Ca^{2+})	100 (Mg^{2+})	100 (Mn^{2+})
C site metal occupancy	0	100 (Mg^{2+})	100 (Mn^{2+})
Resolution (Å)	33.84-2.15	33.98-3.02	41.99-2.95
No. Reflections	26697	18027	11150
R_{work} / R_{free}	0.18/0.23	0.21/0.25	0.19/0.23
No. Molecules in a.u.	1	1	1
DNA strands cleaved?	None	Both	Both
No. Ions at a.s.	2	3	3
No. Atoms			
Protein	2486	2486	2486
DNA	978	980	980
Ions	4	3	3
Water	180	15	11
R.m.s. deviations			
Bond lengths (Å)	0.013	0.002	0.003
Bond angles (°)	1.154	0.696	0.803

*Values in parentheses are for highest-resolution shell. One crystal was used to solve each structure.

Supplementary Table S2 (Cont.). *Crystallographic data collection and refinement statistics.*

	I-CreI3115: I-CreI_ta-3115-null: Mg^{2+}	I-CreI3115: I-CreI_ta-3115-null: Mn^{2+}	I-CreI: I-CreI_ta: Ca^{2+}
PDB code	6FB5	6FB6	4AAG
Data collection			
Space group	$P2_1$	$P2_1$	$P22_12_1$
Cell dimensions			
a, b, c (Å)	45.29, 68.00, 83.03	45.18, 67.99, 83.04	47.47, 71.41, 171.21
α, β, γ (°)	90, 96.55, 90	90, 96.45, 90	90, 90, 90
Wavelength	1.00	1.60	1.00
Resolution (Å)	41.24-2.20 (2.32-2.20)*	44.89-2.60 (2.74-2.60)*	47.47-2.80 (2.95-2.80)*
R_{merge}	0.07 (0.25)	0.11 (0.54)	0.09 (0.48)
Mean I / σI	8.9 (3.4)	10.6 (3.3)	5.3 (1.5)
Completeness (%)	99.5 (96.7)	96.1 (95.6)	99.9 (99.7)
Redundancy	3.3 (3.0)	6.8 (6.7)	4.2 (3.9)
Refinement			
A site metal occupancy	100 (Mg^{2+})	100 (Mn^{2+})	100 (Ca^{2+})
B site metal occupancy	100 (Mg^{2+})	100 (Mn^{2+})	100 (Ca^{2+})
C site metal occupancy	0	74 (Mn^{2+})	0
Resolution (Å)	37.73-2.20	44.89-2.60	45.74-2.80
No. Reflections	47710	28700	14934
R_{work} / R_{free}	0.19/0.22	0.19/0.24	0.18/0.25
No. Molecules in a.u.	1	1	1
DNA strands cleaved?	None	Both	None
No. Ions at a.s.	2	3	2
No. Atoms			
Protein	2477	2477	2459
DNA	978	981	977
Ions	4	3	2
Water	127	11	40
R.m.s. deviations			
Bond lengths (Å)	0.021	0.003	0.008
Bond angles (°)	0.821	0.780	1.280

*Values in parentheses are for highest-resolution shell. One crystal was used to solve each structure.

Supplementary Table S2 (Cont.). *Crystallographic data collection and refinement statistics.*

	I-CreI: I-CreI_ta: Mg^{2+}	I-CreI: I-CreI_ta: Mn^{2+}	I-CreI: I-CreI_ta-null: Mg^{2+}
PDB code	4AAB	6FB7	6FB8
Data collection			
Space group	$P22_12_1$	$P22_12_1$	$P4_3$
Cell dimensions			
a, b, c (Å)	46.81, 70.93, 171.51	49.02, 71.50, 169.15	79.48, 79.48, 125.76
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength	1.00	1.60	1.00
Resolution (Å)	44.51-2.50 (2.64-2.50)*	49.02-2.69 (2.83-2.69)*	27.44-2.45 (2.58-2.45)*
R_{merge}	0.07 (0.41)	0.10 (0.55)	0.11 (0.64)
Mean I / σI	9.9 (1.9)	15.1 (3.7)	8.4 (2.1)
Completeness (%)	99.9 (100)	97.1 (93.5)	100.0 (99.9)
Redundancy	4.8 (4.9)	8.8 (8.6)	6.1 (6.0)
Refinement			
A site metal occupancy	100 (Mg^{2+})	100 (Mn^{2+})	100 (Mg^{2+})
B site metal occupancy	100 (Mg^{2+})	100 (Mn^{2+})	100 (Mg^{2+})
C site metal occupancy	100 (Mg^{2+})	100 (Mn^{2+})	0
Resolution (Å)	34.82-2.50	49.02-2.69	27.42-2.45
No. Reflections	20614	30910	56079
R_{work} / R_{free}	0.17/0.22	0.17/0.21	0.19/0.22
No. Molecules in a.u.	1	1	1
DNA strands cleaved?	Both	Both	None
No. Ions at a.s.	3	3	2
No. Atoms			
Protein	2472	2472	2495
DNA	979	979	978
Ions	3	3	4
Water	122	22	87
R.m.s. deviations			
Bond lengths (Å)	0.007	0.009	0.010
Bond angles (°)	1.234	0.992	1.263

*Values in parentheses are for highest-resolution shell. One crystal was used to solve each structure.

Supplementary Table S2 (Cont.). *Crystallographic data collection and refinement statistics.*

	I-CreI: I-CreI_ta-null: Mn^{2+}
PDB code	6FB9
Data collection	
Space group	$P4_3$
Cell dimensions	
a, b, c (Å)	79.73, 79.73, 126.25
α, β, γ (°)	90, 90, 90
Wavelength	1.80
Resolution (Å)	49.49-2.95 (3.11-2.95)*
R_{merge}	0.08 (0.56)
Mean I / σI	12.4 (2.3)
Completeness (%)	98.6 (97.6)
Redundancy	3.8 (3.7)
Refinement	
A site metal occupancy	100 (Mn^{2+})
B site metal occupancy	100 (Mn^{2+})
C site metal occupancy	46 (Mn^{2+})
Resolution (Å)	42.05-2.95
No. Reflections	32279
R_{work} / R_{free}	0.17/0.20
No. Molecules in a.u.	1
DNA strands cleaved?	Both
No. Ions at a.s.	3
No. Atoms	
Protein	2487
DNA	980
Ions	5
Water	22
R.m.s. deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.287

*Values in parentheses are for highest-resolution shell. One crystal was used to solve each structure.

Supplementary Table S3. *Oligonucleotide sequences sampled as I-CreI DNA target* (labelled at 5' for in vitro cleavage assays / MST experiments and not labelled for crystallization experiments). Bold letters represent oligonucleotide changes derived from their corresponding wild-type DNA sequence.

Number	DNA target	Sequence
1	I-CreI D75N wild-type	5'-TCAAAACGTCGTACGACGTTTTGA-3'
2	I-CreI D75N null	5'-TCAAAAC**TG**CGTACGACGTTTTGA-3'
3	I-CreI 3115 wild-type	5'-TCAGACTTCTCCACAGGAGTCAGA-3'
4	I-CreI 3115 null	5'-TCAGACTT**G**TCCACAGGAGTCAGA-3'

Supplementary Table S4. *Summary of crystallographic data.*
In light green are the structures utilized in the molecular dynamics simulations.

Structure	Metal Concentration	Number	Cleaved
I-CreI_D75N: I-CreI_D75N_target	Ca^{2+} [2 mM]	2	No
I-CreI_D75N: I-CreI_D75N_target	Mn^{2+} [2 mM]	3	Yes
I-CreI_D75N: I-CreI_D75N_target	Mg^{2+} [2 mM]	3	Yes
I-CreI_D75N: I-CreI_D75N_target-null	Mn^{2+} [2 mM]	3	Yes
I-CreI_D75N: I-CreI_D75N_target-null	Mg^{2+} [2 mM]	2	No
I-CreI_3115: I-CreI_3115_target	Ca^{2+} [2 mM]	2	No
I-CreI_3115: I-CreI_3115_target	Mn^{2+} [2 mM]	3	Yes
I-CreI_3115: I-CreI_3115_target	Mg^{2+} [2 mM]	3	Yes
I-CreI_3115: I-CreI_3115_target-null	Mn^{2+} [2 mM]	3	Yes
I-CreI_3115: I-CreI_3115_target-null	Mg^{2+} [2 mM]	2	No