Supporting Information

Marcaida *et al.* 10.1073/pnas.0804795105

SI Results

Metal role in DNA binding. Divalent metal ions play an essential role in the catalysis of LAGLIDADG meganucleases. The conserved acidic residues at the active sites coordinate these divalent metal ions. The general mechanism of cleavage of the phosphodiester bonds of DNA requires a nucleophile to attack the electron deficient phosphorus atom, a general base to activate the nucleophile, a general acid to protonate the leaving group, and positively charged groups to stabilize the phosphoanion transition state.

Interestingly, the presence of divalent metal ions is dispensable for DNA binding (Fig. S3). Native electrophoresis and NMR analysis showed that DNA binding could be accomplished in the absence of divalent metal ions, suggesting that the presence of the ion is necessary only for catalysis. We monitored changes in the protein signals upon metal ion and/or duplex DNA addition. ^{1}H–^{15}N HSQC experiments where each signal corresponds to one amide N-H group in the protein were recorded. The HSQC spectrum of I-DmoI depicts fewer signals than expected, suggesting that a number of residues experience different dynamics than the rest of the structure—possibly loops with higher mobility or conformational exchange—in agreement with the differences observed between the apo and DNA bound structures (Fig. S1). Upon addition of a 20-fold molar excess of Mg^{2+} over I-DmoI no change in the signals is observed (Fig. S3*b*) suggesting that divalent metal ions do not bind to the protein in the absence of DNA, although transient binding to the invisible more mobile regions cannot be excluded. However, in the presence of 1 equivalent of DNA, changes do occur in a number of signals, and new ones appear, indicating that there is complex formation and that binding stabilizes certain regions of the proteins rendering the intervening residues observable (Fig. S3*c*). Essentially the same changes are observed in the presence of Mg^{2+} (Fig. S3*d*).

Pseudopalindromic versus nonpalindromic DNA recognition. A structure-based sequence alignment of I-DmoI with I-CreI, H-DreI and I-SceI was performed to compare the molecular basis of I-DmoI DNA recognition with other members of the LAGLIDADG family. I-CreI and I-SceI are 2 well-characterized meganucleases representing the homodimeric and monomeric members of the LAGLIDADG family that bind pseudo- and nonpalindromic targets respectively (Fig. S7). H-DreI is a chimerical meganuclease composed of domain A from I-DmoI fused to a I-CreI monomer that recognizes a nonpalindromic target. The alignment illustrates the differences in the primary and secondary structures among these enzymes regarding the location of residues involved in DNA binding (blue residues, Fig. S7*a*). A structural comparison of the DNA binding residues of these homing endonucleases shows that despite the similar structural scaffold, the residues responsible for DNA binding are not topologically conserved (Fig. S7 *a* and *b*). The schematic comparison of the base-protein contacts (including hydrogen bonds and van der Waals interactions) in the different meganuclease-DNA complexes (Fig. S7*c*) illustrates how the homodimeric meganuclease accomplish DNA recognition generating a similar network of protein-DNA contacts on both sides of the pseudopalindromic DNA, whereas the monomeric ones display a tendency to maximize the interactions in one-half of the target DNA.

I-DmoI specificity *in silico* analysis. The differences in target recognition lead us to assess the specificity of I-DmoI and compare it with I-SceI, I-CreI and the chimerical H–DreI. Although it has been shown that the presence of different cations can influence the cleavage and specificity properties of I-DmoI (1, 2), we have analyzed the length of the binding site and the number of specific positions for the target DNA sequences of each meganuclease using our experimental data—the only one available about I-DmoI-DNA complex—with FoldX2.8 (3). Each base was mutated to the other 3 possibilities and the resulting interaction energies were converted to a probability to predict the preference for each base at a determined position (Fig. S8). These predictions were compared with the experimentally determined cleavage preferences (specificity logos) of I-SceI and I-CreI (4, 5). The different bases were considered to be specific when their probability was $>2/3$, and the recognition site was defined as the sequence between the first and last specific nucleotide. We defined a specificity index as the sum of probabilities of the specific bases normalized by the recognition site length (see Materials and Methods). Higher values of this index indicate a high degree of meganuclease specificity. As mentioned in the main text and shown is Fig. S8 and Tables S1 and S2, DmoI is the most specific of the meganucleases studied *in silico*. Although its recognition site is the shortest one, it does not have many unspecific bases within it. The opposite case would be that of I-CreI, where the binding site spans over twenty-two bases, but contains 9 variable positions. I-DmoI also shows a clear preference for cytosine and guanine, reflecting the rich GC content of the thermophilic organism where it is coded. No clear correlation could be found between the interaction energy and the recognition site length or the number of specific bases. To evaluate the specificity of these enzymes in real genomes, we *searched in silico* for putative binding sites in 2 model organisms. We scanned the energy matrices coming from the analysis above along the *S. cerevisiae* and *D. melanogaster* genomes, using 2 different energy thresholds with respect to the wild-type interaction energy. The results did not yield a hit in yeast with the lower energy threshold (I-SceI site in the yeast strain sequenced is disrupted by the insertion of an intron that contains the meganuclease), and very few sites were found in *D. melanogaster* (Table S2). When the energy threshold was increased, weaker hits became apparent (Table S2). This could be important in the context of a highly expressed enzyme or one with enhanced activity.

SI Materials and Methods

Protein Expression, Purification, Protein–DNA Formation and Crystallization. The protein was expressed, purified and crystallized in complex with its target DNA as described in ref. 6. Selenomethionine-labeled I-DmoI was produced as in ref. 7 and purified as described in ref. 6.

Data collection, structure solution, model building and refinement. All data were collected at cryogenic temperatures using synchrotron radiation at 100K. I-DmoI crystals were mounted and cryoprotected. The datasets were collected using synchrotron radiation at the ID29 beamline at the ESRF (Grenoble), and at the PX beamline at the SLS (Villigen). Diffraction data were recorded on an ADSC-Q4, Mar345dtb, Mar225 CCD or Pilatus detectors depending on the beamline. Processing and scaling were accomplished with HKL2000 (8) and XDS (9). Statistics for the crystallographic data and structure solution are summarized

in Table 1. Reduced intensities were used to search for the Se substructure. A resolution cut-off of 4.0 Å was applied during the substructure solution. All of the 9 Se positions in the asymmetric unit were found with SnB (10) and SHELXD (11). These positions were fed into SHARP (12). After solvent flattening with SOLOMON (13) the initial 2.7 Å map, showing all of the Se sites, was used for automatic model building using MAID (14). The models of I-DmoI/DNA complexes were rebuilt with COOT and refined to 2.0 and 2.1 Å (PDB entries 2VS7 and 2VS8) using REFMAC5 (15) (Table S1). Protein-DNA interactions were analyzed with NUCPLOT (16). The DNA conformation and helical parameters were examined with CURVES (17).

Construction of Target Clones. The nonpalindromic twenty-four base pairs long target sequence 5′-GCCTTGCCGGGTAAGTTCCGGCGC-3′ is the natural I-DmoI target (coding strand A). For nomenclature purpose, we divided it in 2 equal parts L and R. The 64 degenerated targets derived from LR sequence were obtained by mutating nucleotides at positions +8, +9, and +10 in the R sequence. Sixty four pairs of oligonucleotides (5′-GCCTTGCCGGGTAAGTTCCNNNGC-3′ and reverse complementary sequences) representing the target library R-10NNN (Right half, positions 10N9N8N) were ordered form Sigma, annealed, and inserted as previously described using the Gateway protocol (Invitrogen) into the yeast pFL39-ADH-LACURAZ containing a I-SceI target site as control (18). *S. cerevisiae* strain FYBL2–7B (MAT a, ura3Δ851, trp1Δ63, leu2Δ1, lys2Δ202) was transformed with the yeast reporter vectors.

Yeast Screening. I-DmoI WT and the 2 I-DmoI mesophilic variants reported, D1 and D2 (19) were screened against the 64 I-DmoI derived targets (R-10NNN, right half positions 10N9N8N) by mating meganuclease expressing yeast clones with yeast strains harboring a reporter system as described in ref. 18. The Meganuclease-induced recombination of the LacZ reporter system restores a functional β-galactosidase gene that can be detected by X-Gal staining.

Interaction Analysis. Specificity logos in Fig. S8 have been calculated with FoldX, by mutating each of the bases of the DNA to all other nucleotides. The height of a given nucleotide is proportional to $\exp(-\Delta\Delta G_{int}/RT)$, where $\Delta\Delta G_{int}$ is the difference in interaction energy between the complex with mutated DNA and the wild type. This function is graphically displayed as information content thanks to the R package seqLogo. Matrices for each of the 4 meganucleases analyzed, containing the difference of interaction energy due to each mutation in the DNA, have been scanned through both the *S. cervisiae* and the *D. melanogaster* genomes to find how many potential binding sites can be found. Two different energy thresholds have been used: 2 kcal/mol, which sets the limit for reasonable binding, and 4 kcal/mol, to show the increase in sites with a more permissive threshold.

To predict the *in silico* binding pattern of I-DmoI and the D1 and D2 variants (Fig. 5*B*), we used the wild type I-DmoI structure and calculated the differences in the interaction energy with the WT DNA sequence. The energies are calculated by adding up the change in interaction energy due to each individual mutation in the DNA.

Band Shift Assay. Band shift assay was performed in 10 mM Tris·HCl pH 8, 50 mM NaCl, 10 mM $CaCl_2$ or $MgCl_2$, 1 mM DTT incubated 1h at 20 or 37°C, using 5 μM 6-FAM DNA duplex (see Fig. 1*B*) and 5 μM I-DmoI protein. After incubation the samples were subject to electrophoresis using a 15% Acrylamide-Tris-Glycine pH 11 gel.

Cleavage assay for double stranded and nicked double stranded DNA duplexes. The DNA substrates were generated after mixture of equimolar amounts of 2 or 3 complementary oligonucleotides in 10 mM Tris·HCl pH 8 (Fig. S6). The sample was heated at 99°C for 5 min and slowly cooled down to 25°C. The corresponding analyzed strand was labeled with 6-FAM (6-carboxyfluorescein) at the 5′.

The standard I-DmoI DNA target oligonucleotide sequences were: For the coding strand labeled duplex, 5′-(6FAM)-GCCTTGCCGGGTAAGTTCCGGCGCG and 5′- CGCGC-CGGAACTTACCCGGCAAGGC. For the noncoding strand labeled duplex, 5′-(6FAM)-CGCGCCGGAACTTACCCG-GCAAGGC and 5′-GCCTTGCCGGGTAAGTTCCG-GCGCG.

The nicked DNA targets were: for the coding strand labeled 5′-(6FAM)-GCCTTGCCGGGTAAGTTCCGGCGCG, 5′-CGCGCCGGAACTTAC and 5′-CCGGCAAGGC. For the noncoding strand labeled 5′-(6FAM)-CGCGCCGGAACT-TACCCGGCAAGGC, 5′-GCCTTGCCGGGTAA and 5′-GTTCCGGCGCG.

Cleavage assays were performed at 65°C in 10 mM Tris·HCl (pH 8.0), 50 mM NaCl, 10 mM $MgCl_2$. The cleavage reactions contained 5 pmol of target DNA and 2.5 pmol of I-DmoI in 10 μl final volume per reaction. To perform the time-course measurements the cleavage reactions were stopped at different times after addition of 10 μl of formamide loading dye on ice. The reactions mixtures were analyzed using 10% polyacrylamide gels with 8 M urea and TBE buffer. The gel was exposed to UV light and the intensity of the bands was quantified using the ImageJ software (http://rsb.info.nih.gov/ij). Three different measurements were performed for each condition and the error bars represent the s.e. of the measured values.

NMR Spectroscopy. Uniformly ^{15}N isotope enriched I-DmoI protein was obtained growing *E. coli* cells in minimal medium containing 1 g/L of ^{15}N-NH_4Cl. NMR experiments were recorded on Bruker AVANCE 600 (with cryoprobe) at 35°C in with 500 μM protein or protein/DNA complex in 300 μl of 20 mM Tris·HCl buffer pH 7.4 with 500 mM NaCl containing 5% (vol/vol) ^2H_2O and 0.03% (wt/vol) NaN_3 and in the presence or absence of 10 mM $MgCl_2$. The high salt concentration was necessary to increase the solubility of the protein and stabilize it at 500 μM (which precipitated over time at lower concentrations of salt). The samples were prepared from protein and DNA duplex stocks with a concentration of ≈1 mM (determined by UV absorbance). Protein purification and DNA duplex preparation was done with buffers that did not contain divalent metal ions, as evaluated by NMR spectroscopy of buffer samples with added EDTA, whose signals change in the presence of divalent cations (data not shown). Still, both the protein and DNA stock solutions were extensively dialyzed before the NMR sample preparation against buffer (20 mM Tris pH 7.4, 500 mM NaCl) containing solid resin Chelex-100 (Bio-Rad) to remove all traces of divalent cations, and the NMR tubes were rinsed in Chelex treated water and dried before use. The concentration of a 1M $MgCl_2$ stock solution was determined directly by gravimetry. Chemical shifts were measured relative to internal DSS for ^1H and calculated for ^{15}N (20). Spectra were processed with NMRPipe (21) and analyzed using NMRView (22).

1. Lykke-Andersen J, Garrett RA, Kjems J (1997) Mapping metal ions at the catalytic centres of 2 intron-encoded endonucleases. *EMBO J* 16:3272–3281.
2. Aagaard C, Awayez MJ, Garrett RA (1997) Profile of the DNA recognition site of the archaeal homing endonuclease I-DmoI. *Nucleic Acids Res* 25:1523–1530.
3. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L (2005) The FoldX web server: An online force field. *Nucleic Acids Res* 33:W382–W388.
4. Doyon JB, Pattanayak V, Meyer CB, Liu DR (2006) Directed evolution and substrate specificity profile of homing endonuclease I-SceI. *J Am Chem Soc* 128:2477–2484.

5. Argast GM, Stephens KM, Emond MJ, Monnat RJ, Jr. (1998) I-PpoI and I-CreI homing site sequence degeneracy determined by random mutagenesis and sequential in vitro enrichment. *J Mol Biol* 280:345–353.
6. Redondo P, Prieto J, Ramos E, Blanco FJ, Montoya G Crystallization and preliminary X-ray diffraction analysis on the homing endonuclease I-Dmo-I in complex with its target DNA *Acta Crystallogr Sect F Struct Biol Cryst Commun* 63:1017–20, 2007.
7. Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol* 229:105–124.
8. Otwinowski Z, Minor W (1997) *Processing of X-ray Diffraction Data Collected in Oscillation Mode* (Academic, New York).
9. Kabsch W (1988) Automatic indexing of rotation diffraction patterns. *J Appl Cryst* 21:67–71.
10. Weeks CM, Miller R (1999) The design and implementation of SnB v20. *J Appl Cryst* 32:120–124.
11. Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. *Acta Crystallogr D* 58:1772–1779.
12. de la Fortelle E, Bricogne G (1997) *Macromolecular Crystallography* (Academic, New York).
13. Abrahams JP, Leslie AG (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr D* 52:30–42.
14. Levitt DG (2001) A new software routine that automates the fitting of protein X-ray crystallographic electron-density maps. *Acta Crystallogr D* 57:1013–1019.
15. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* 53:240–255.
16. Luscombe NM, Laskowski RA, Thornton JM (1997) NUCPLOT: A program to generate schematic diagrams of protein-nucleic acid interactions. *Nucleic Acids Res* 25:4940–4945.
17. Laverly R, Sklenar H (1988) The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J Biomol Struct Dyn* 6:63–91.
18. Arnould S, *et al.* (2006) Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J Mol Biol* 355:443–458.
19. Prieto J, *et al.* (2008) Generation and Analysis of Mesophilic Variants of the Thermostable Archaeal I-DmoI Homing Endonuclease. *J Biol Chem* 283:4364–4374.
20. Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD (1995) 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. *Investigations of nearest-neighbor effects* J Biomol NMR 5:67–81.
21. Delaglio F, *et al.* (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277–293.
22. Johnson BA, Blevins RA (1994) NMRView: A computer program for the visualization and analysis of NMR data. *J Biomol NMR* 4:603–614.

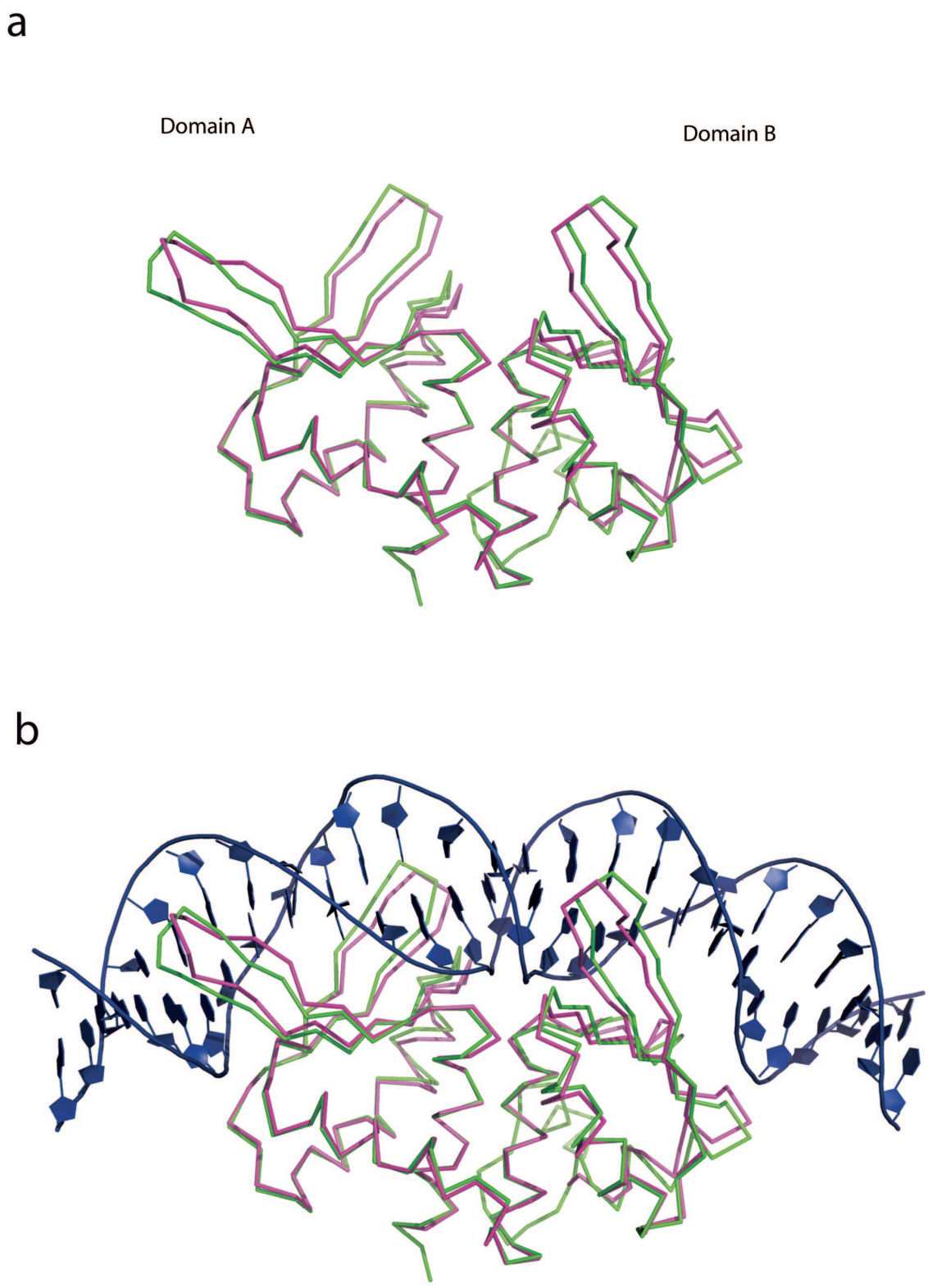

Fig. S1. Superposition of the I-DmoI apo (PDB entry 1B24) (magenta) structure with the DNA bound enzyme (green). (*a*) The main differences in the protein moiety can be observed in the β-sheets and the loops that join the β-strands to accommodate the DNA molecule. (*b*) Same view including the DNA molecule (blue).

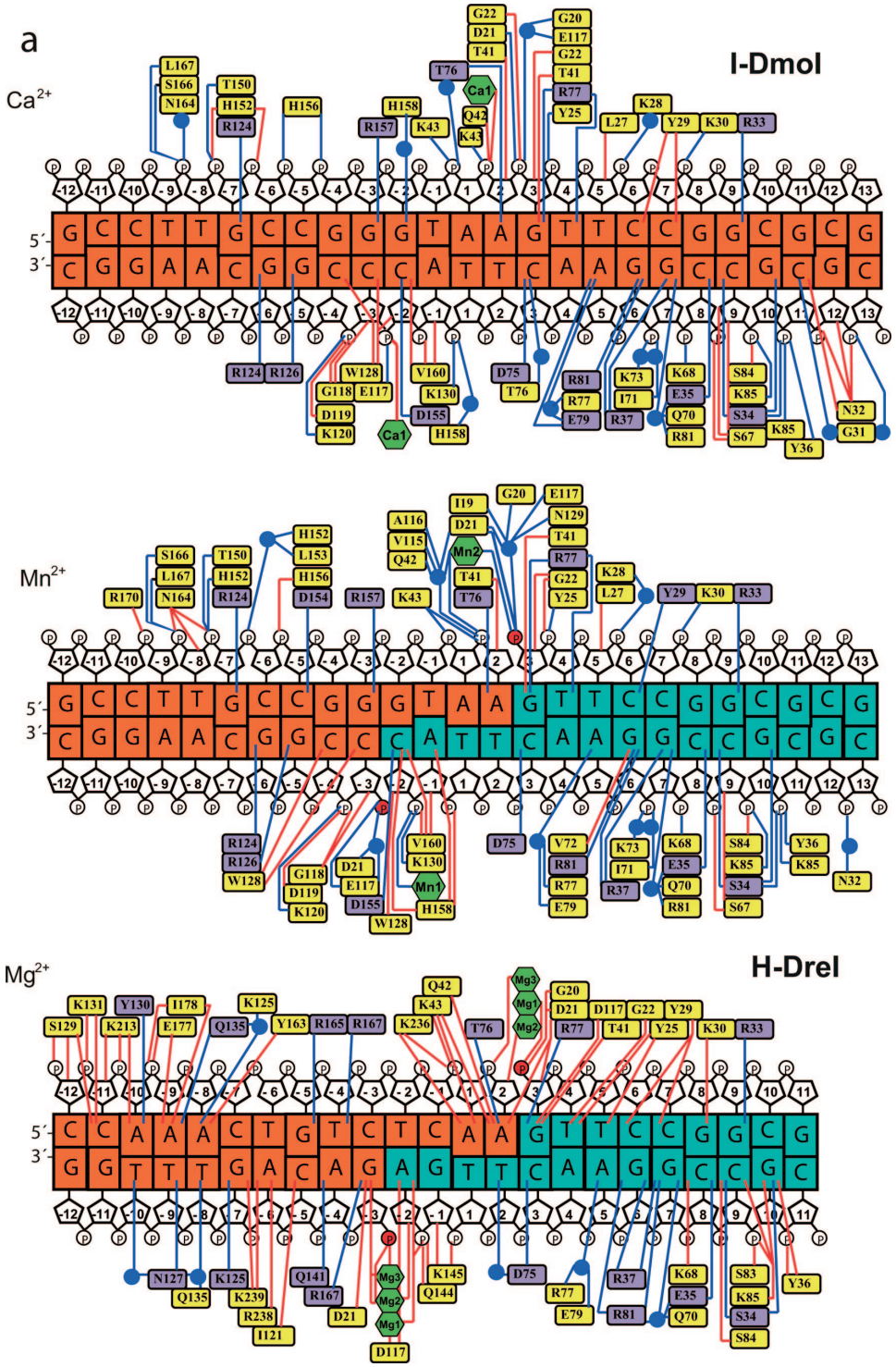

Fig. S2. (*a*) Scheme of the Protein-DNA contacts in the Ca^{2+} and Mn^{2+} bound I-DmoI structures compared with H-DreI. The DNA is in orange in the substrate bound structure whereas in the Mn^{2+} structure, each DNA cleavage product is in blue (domain A) and orange (domain B) according to the protein domain they bind to. Phosphates at which cleavage occurs are in red in the Mn^{2+} and Mg^{2+} panels. Lines in blue and red indicate hydrogen bonds and van der Waals interactions, respectively. Blue dots represent water molecules involved in the interaction. Amino acids depicted on violet boxes represent hydrogen bond interactions with the bases and the residues on yellow boxes represent van der Waals interactions with the DNA (bases, riboses or phosphates). (*b*) Scheme of the protein-base contacts in the I-DmoI and H-DreI structures. (*c*) Schematic representation of the I-DmoI and H-DreI DNA targets using CURVES (17) showing their curved global axis in red.

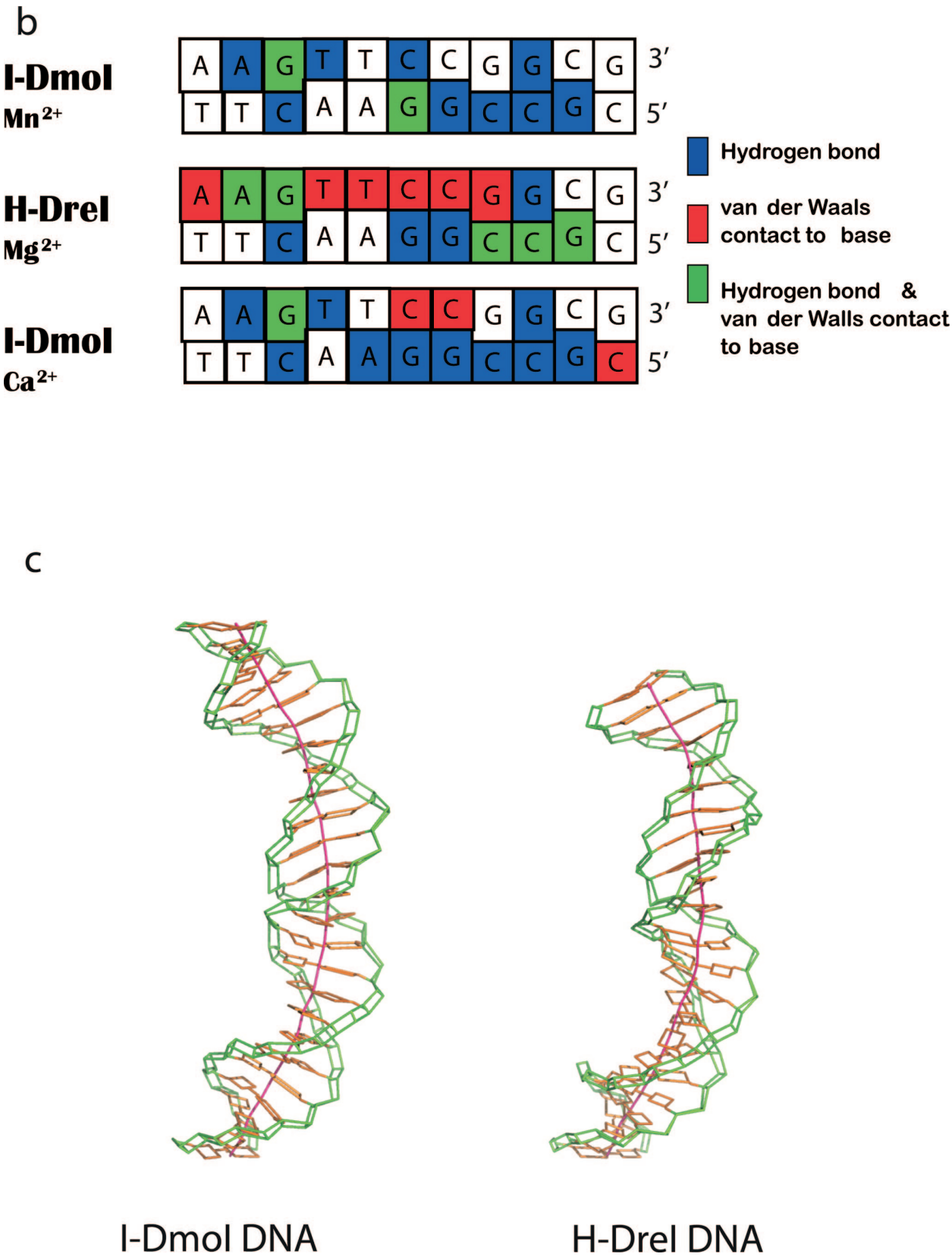

Fig. S2 (continued).

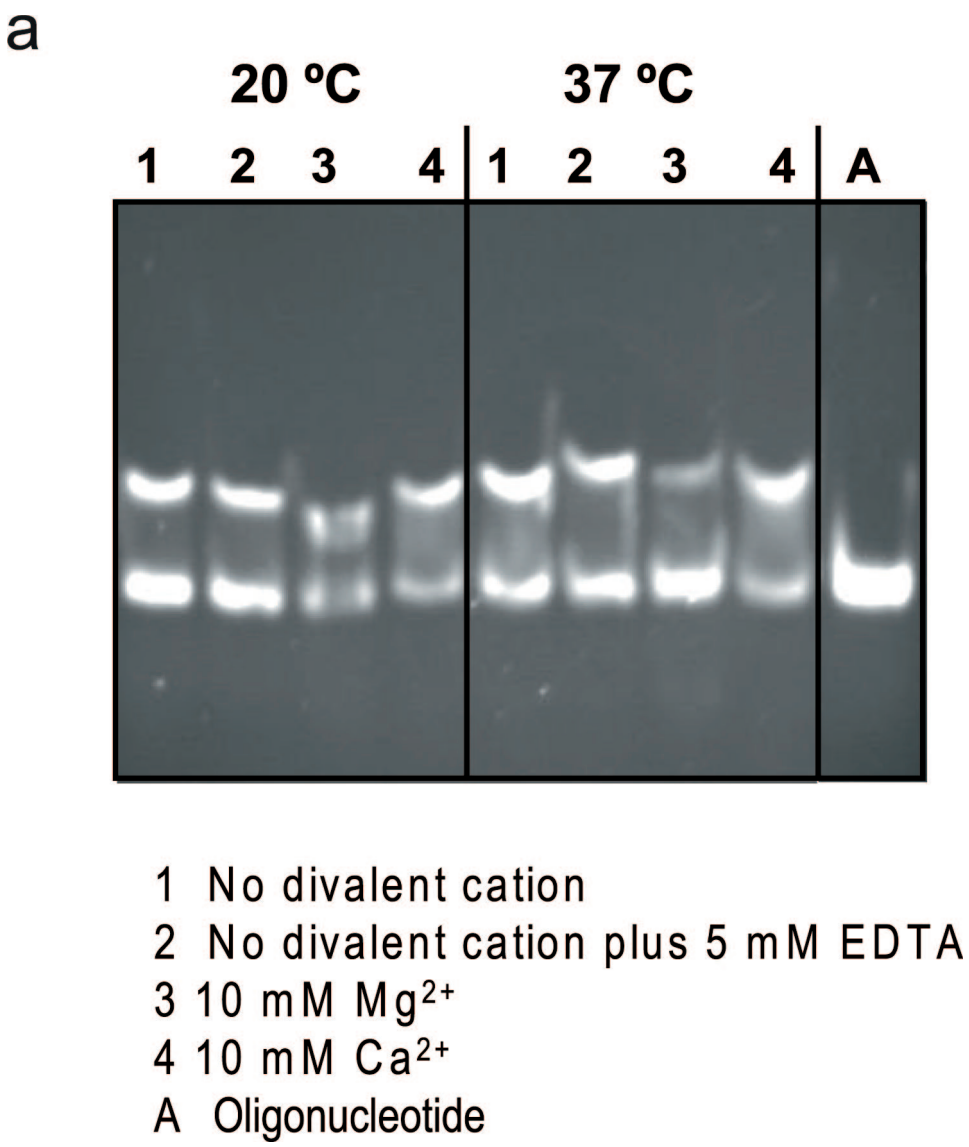

Fig. S3. (*a*) Band shift analysis of the I-DmoI-DNA complex in the presence of cations and EDTA. (*b*) NMR ^{1}H-^{15}N HSQC spectra of I-DmoI. The spectrum of 500 μM I-DmoI alone is shown in black, with overlays, in red color, of the spectra of I-DmoI in the presence of 10 mM $MgCl_2$. (*c*) DNA duplex (500 μM). (*d*) DNA duplex (500 μM) and 10 mM $MgCl_2$. Conditions: 600 MHz, 35°C, 20 mM Tris pH 7.4 and 500 mM NaCl.

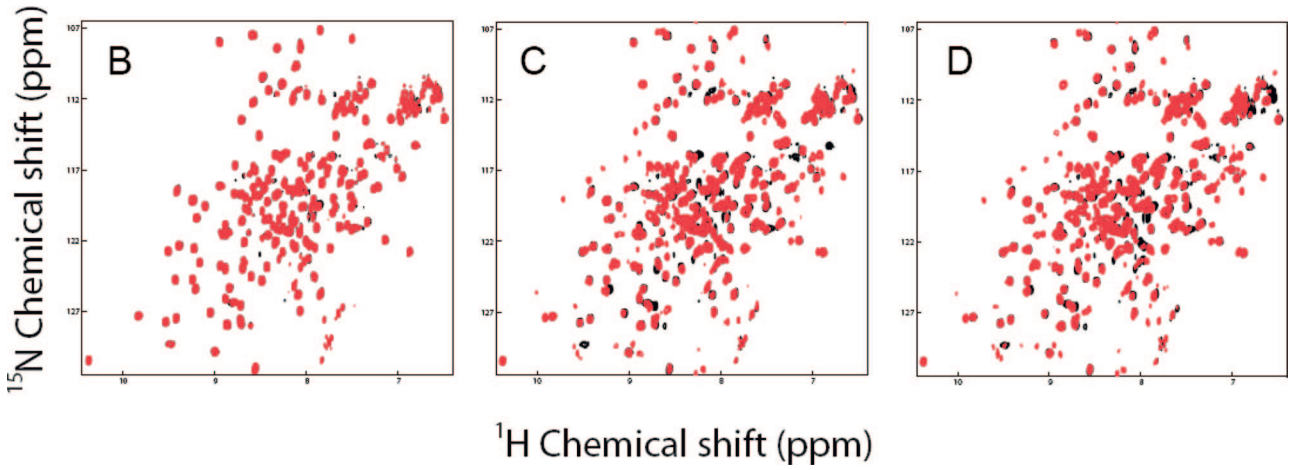

Fig. S3 (continued).

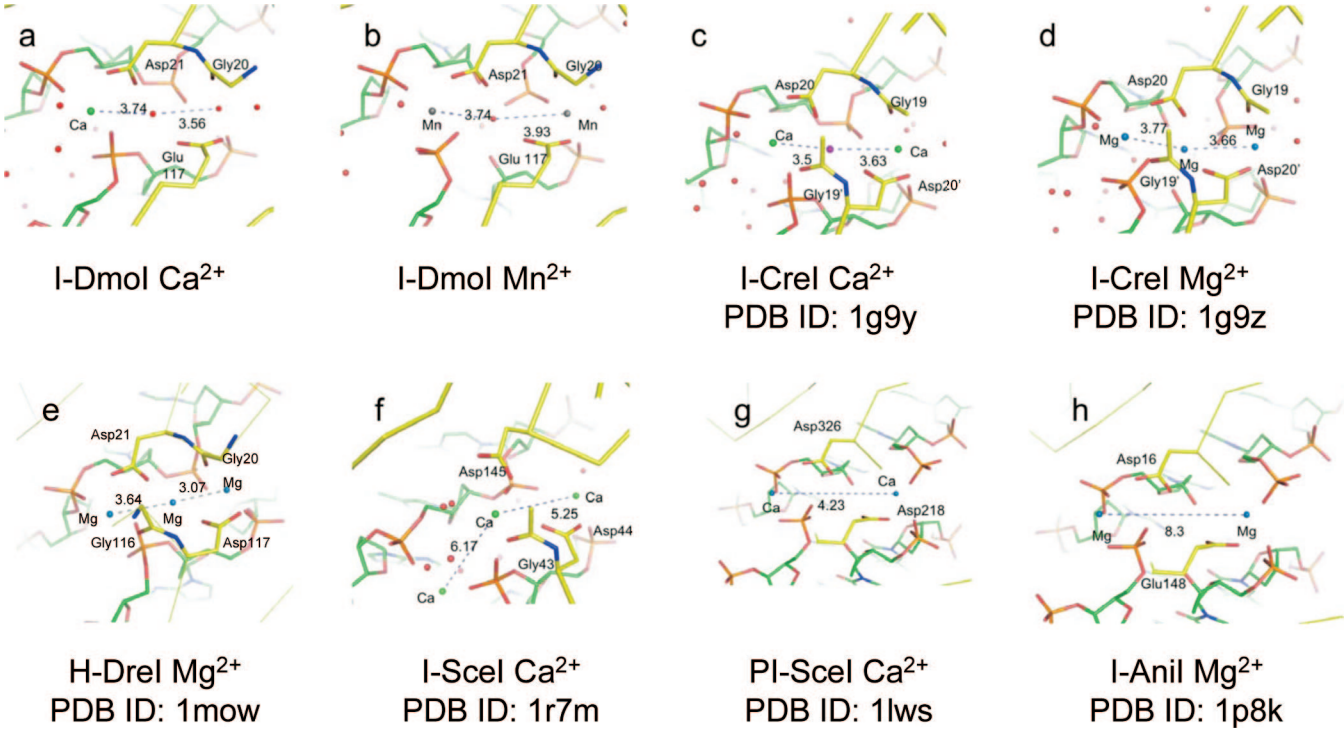

Fig. S4. Comparison of the active centres of I-DmoI, I-CreI, PI-SceI, I-AniI, H-DreI and I-SceI. The active centres were aligned and figures taken from the same viewpoint after their superposition. DNA is in green, the protein in yellow with ordered water molecules in red. Calcium ions are shown as green spheres, magnesium ions as blue spheres, manganese ions as gray spheres and sodium ions as purple spheres. Distances between the active site cations are shown in angstroms.

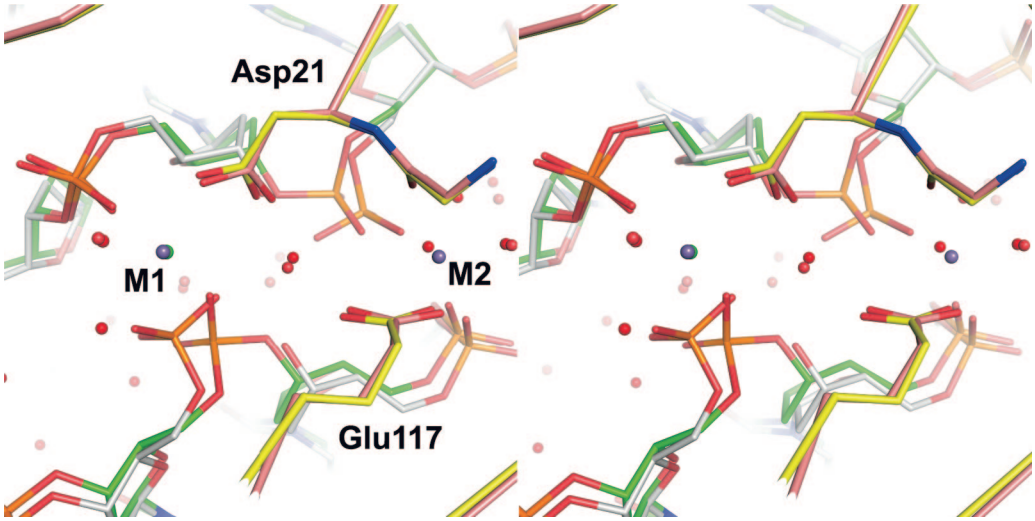

Fig. S5. Stereo comparison of the enzyme active site in the substrate (protein in yellow and DNA in green) and product (protein in pink and DNA in white) bound structures. The manganese atoms are in blue and the calcium atom in green. The metal sites are labeled (M1) for the shared position between calcium and manganese and (M2) for the second manganese atom.

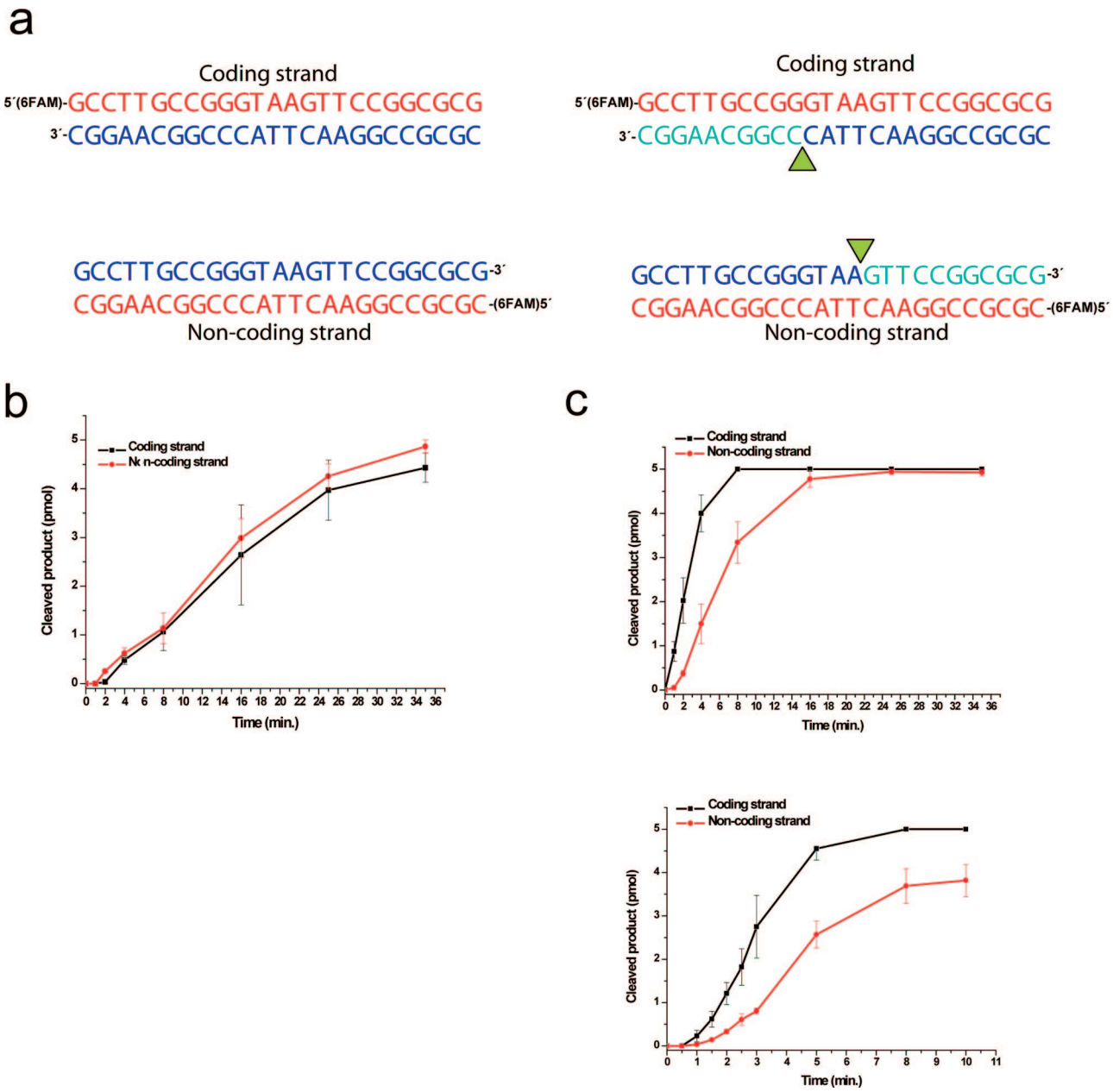

Fig. S6. Cleavage assay for double stranded and nicked double stranded DNA duplexes. (*a*) Sequences of the DNA duplexes used as substrates for the cleavage reaction. In each case the analyzed strand was labeled with 6-FAM. The DNA targets displayed on the left side were used to analyze the cleavage rate of the coding versus noncoding stand. The right side DNA targets contained a nick in the position indicated with a green arrow that coincides with the corresponding cleavage site on each strand. (*b*) Time-course cleavage of coding versus noncoding strand. (*c*) Time-course cleavage of coding versus noncoding strand using the nicked substrates. The lower graph displays another set of experiments with additional measurements at shorter times to calculate the rates of cleavage accurately. The cleavage of the coding strand was approximately 2 times faster than the noncoding one (Cleavage rate of 0.0165 pmol of substrate per second versus 0.0081 pmol of substrate per second). Three experiments were performed for each graph and the error bars represent the s.e. of the measured values.

a

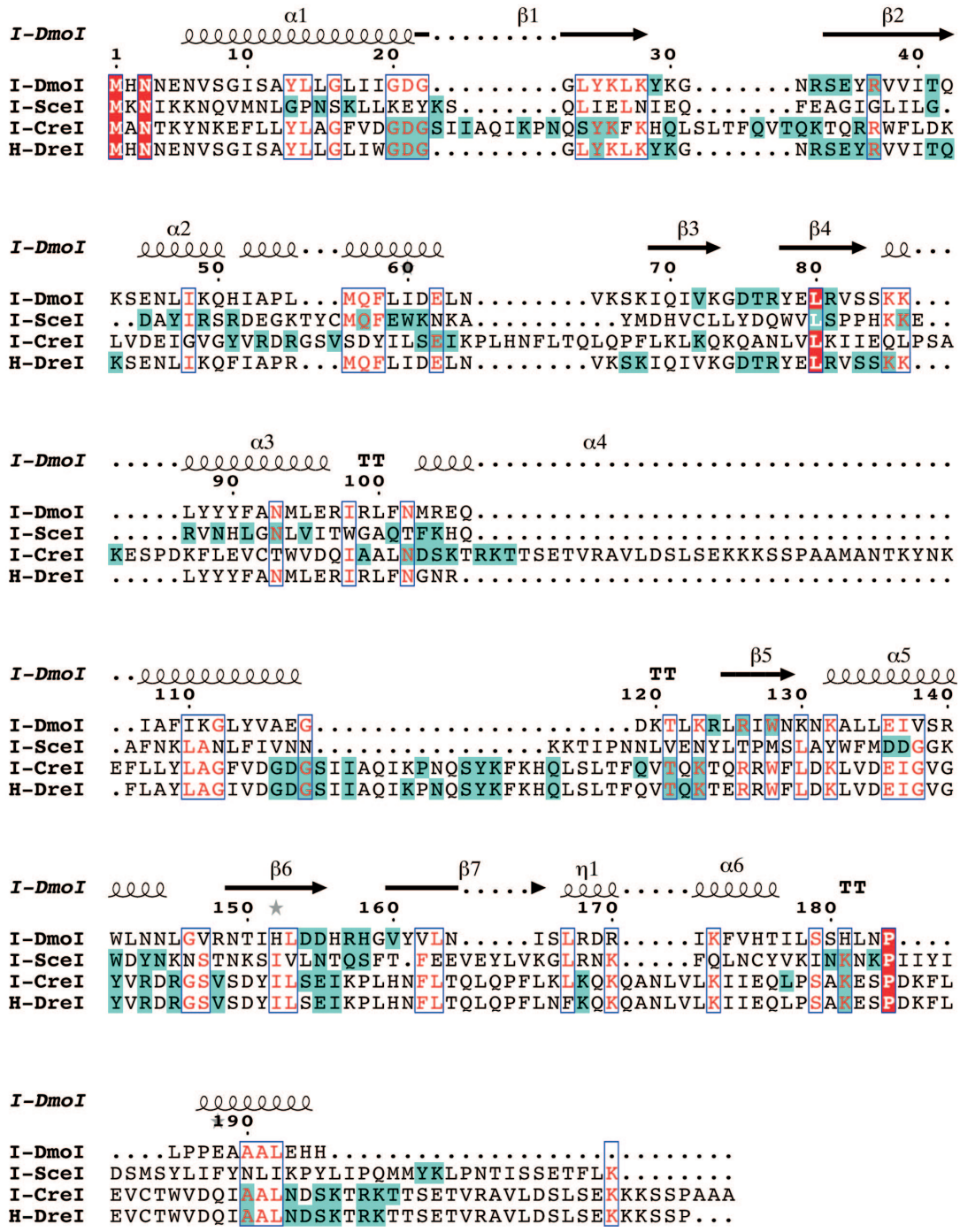

Fig. S7. Structural basis of DNA recognition. (*a*) Structure-guided sequence alignment between the archaeal I-DmoI, H-DreI chimera, eukaryotic I-SceI and I-CreI homing endonucleases. Secondary structure elements of the homing endonuclease I-DmoI are shown above the alignment. Identical residues are boxed with a red background or shown in red when they are the same in all of the sequences. Residues with a blue background are those involved in hydrogen bonds and/or van der Waals interactions with the bases in the complex crystal structures. Sequence alignment was carried out with Clustal (1) and the structural alignment with ESPript (2). (*b*) Comparison of the location of the residues involved in protein-base interactions (regions in blue) in the I-DmoI, H-DreI, I-SceI and I-CreI protein-DNA structures. (*c*) Schematic view of the DNA, showing protein-base interactions in the substrate bound structures.

1. Larkin MA, *et al.* (2007) Clustal W and Clustal X version 20. *Bioinformatics* 23:2947–2948.
2. Gouet P, Courcelle E, Stuart DI, Metoz F (1999) ESPript: Analysis of multiple sequence alignments in PostScript. *Bioinformatics* 15:305–308.

b

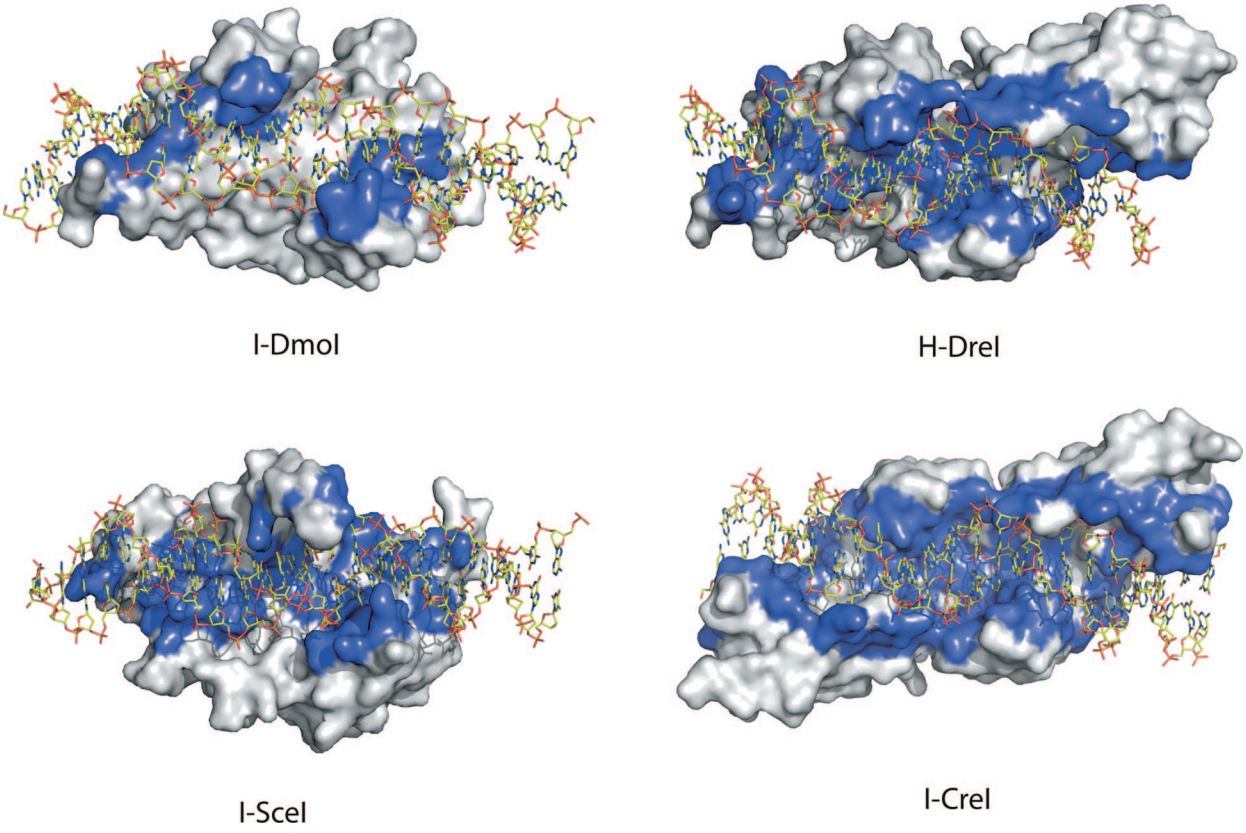

Fig. S7 (continued).

C

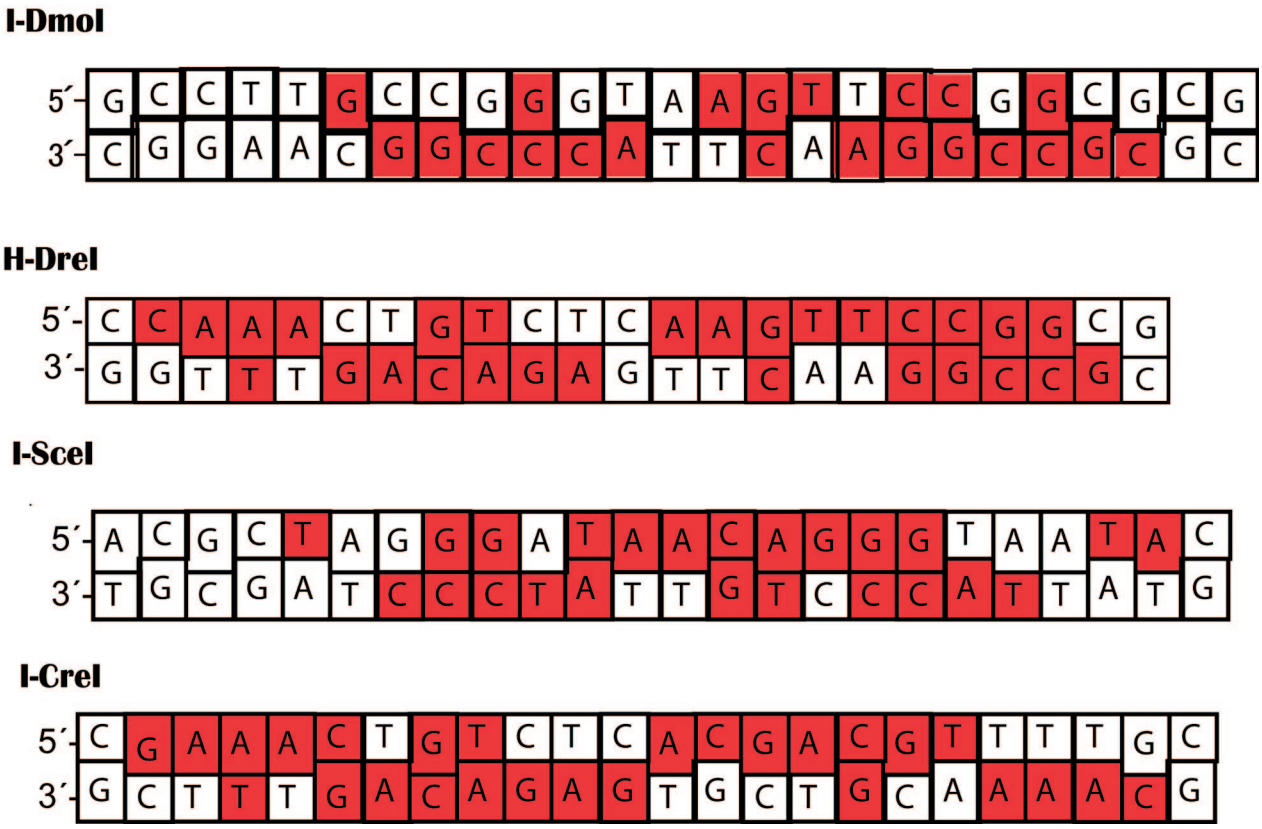

Fig. S7 (continued).

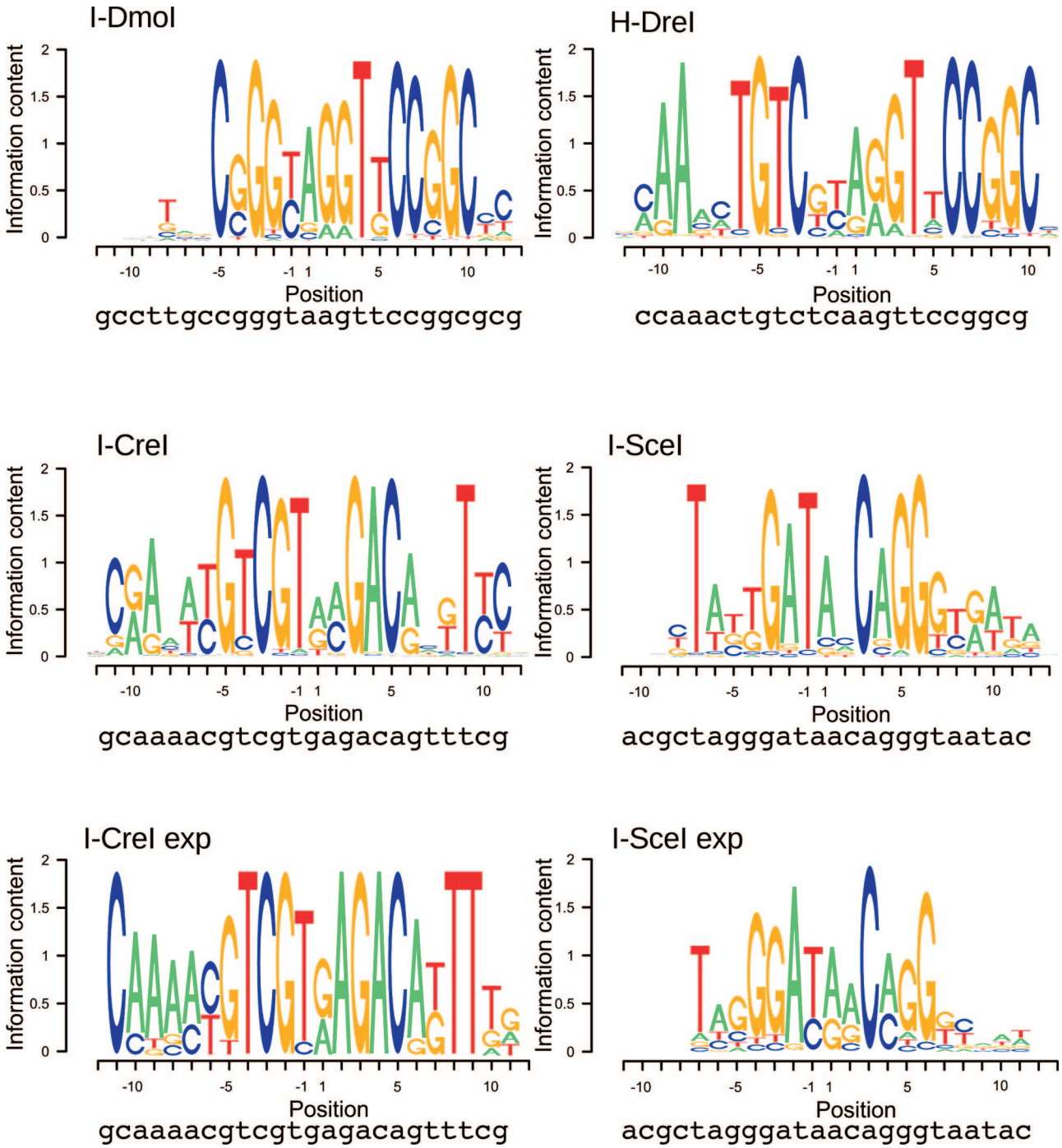

Fig. S8. Specificities for I-DmoI, H-DreI, I-CreI and I-SceI. The energy-based logos display the different binding specificities of the meganucleases for the top 4 logos. I-DmoI presents a short binding site with the highest specificity, whereas I-SceI has a long but quite tolerant binding site. Base discrimination predicted by FoldX compares well with respect to available experimental results for I-SceI (1) and I-CreI (2), shown at the bottom. Under each logo, the top strand of the crystallized DNA sequence is shown.

1. Doyon JB, Pattanayak V, Meyer CB, Liu DR (2006) Directed evolution and substrate specificity profile of homing endonuclease I-SceI. *J Am Chem Soc* 128:2477–2484.
2. Argast GM, Stephens KM, Emond MJ, Monnat RJ, Jr. (1998) I-PpoI and I-CreI homing site sequence degeneracy determined by random mutagenesis and sequential in vitro enrichment. *J Mol Biol* 280:345–353.

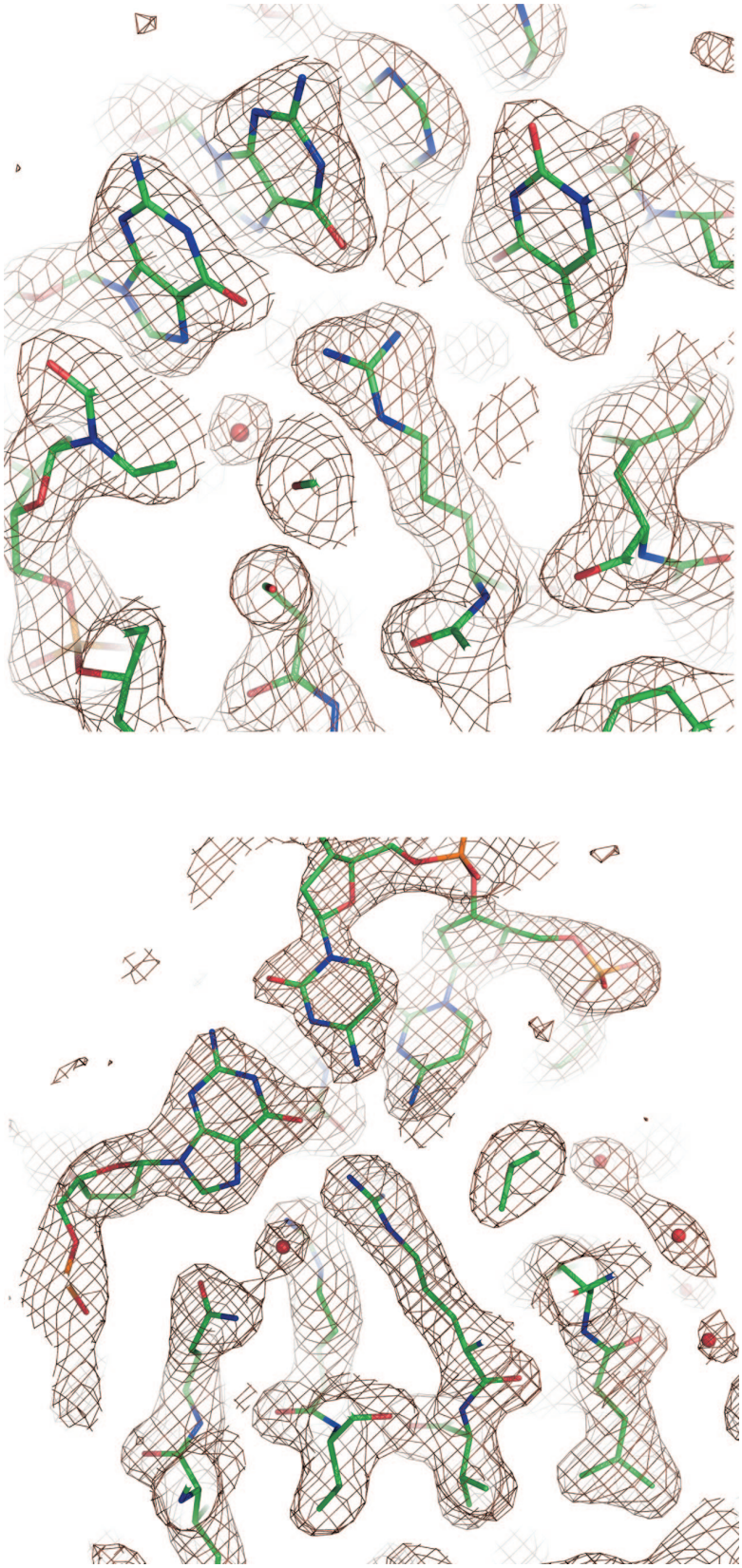

Fig. S9. Two different views of the $2f_o - f_c$ electron density map of the I-Dmo/DNA complex structure contoured at 1.25 σ.

Table S1. Data collection, phasing, and refinement statistics

Data Collection	I-DmoI/DNA Mg^{2+} Se-SAD	I-DmoI/DNA Ca^{2+}	I-DmoI/DNA Ca^{2+} Anomalous data set	I-DmoI/DNA Mn^{2+}	I-DmoI/DNA Mn^{2+} Anomalous data set
Number of crystals	1	1	1	1	1
Temperature, K	100	100	100	100	100
Environment	ADSC-Q315, ESRF, ID-29	MAR225, SLS, PX beamline	MAR345dtb In house	PILATUS, SLS	PILATUS, SLS
Wavelength, Å	0.979	0.9	1.5414	1.00	1.549
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Cell dimensions, Å, °	a = 106.6, b = 70.3, c = 107.14, ag = 90° b = 119.9°	a = 106.9, b = 70.5, c = 106.8, ag = 90° b = 119.7°	a = 106.9, b = 70.5, c = 106.8, ag = 90° b = 119.7°	a = 107.1, b = 70.6, c = 106.9, ag = 90° b = 119.8°	a = 106.7, b = 70.4, c = 106.9, ag = 90° b = 119.8°
No. mol ASU	3	3	3	3	3
Resolution, Å	40.3–2.6	25–2.0	25–2.89	26.82–2.10 (2.21–2,10)	42.64–2.40 (2.53–2.40)
Unique reflections	40,782	87,351	30,936	80,887	51,394
Average multiplicity	2.8(2.4)	3.6(2.9)	5.9	3.0(2.8)	3 (2.4)
Completeness, %	95.9 (77.2)	95.4(72.5)	99.7(99)	99.7(99.5)	95.1 (76.3)
R_{merge}*	0.04 (0.19)	0.06(0.45)	0.08(0.32)	0.047(0.45)	0.036 (0.11)
$\langle I/s(I)\rangle$	14.4(3.8)	14.9(1.0)	8.4(2.3)	12.8(2.3)	15.9 (6.9)
No. Se atoms found	3/3				
FOM (SHARP)	0.56				
RCullis (SHARP)	0.8				
Phasing Power	1.01				
Density modification					
FOM (Solomon)	0.85				
Reflections, no.		79,135		76,811	
Resolution range, Å		25.7–2.0		25.7–2.1	
R factor/R free, %		18.7/22.8		20.1/24.8	
No. protein atoms (Average B, Å²)†		4,668 (41.5)		4,573 (41.4)	
No. DNA atoms (Average B, Å²)†		3,057 (51.5)		3,063 (54.8)	
No. water molecules (Average B, Å²)†		613 (55.7)		392 (50.5)	
Other atoms		3 Ca 1 Acetate		6 Mn 3 Acetate	
rms bond length, Å		0.008		0.012	
rms bond angle, °		1.38		1.69	
Ramachandran plot outliers (number)‡		0		0	

Values in the highest resolution shell are given in parentheses.

*$R_{merge} = S_h S_i |I_{h,i} - \langle I_h\rangle| / S_h S_i |I_{h,i}|$.

†Calculated using MOLEMAN.

‡Calculated using PROCHECK.

Table S2. DNA Interaction analysis for I-DmoI, H-DreI, I-CreI and I-SceI

Enzyme	Recognition site length, bp	No. of specific bases	Specificity index	ΔG, kcal/mol	*Saccharomyces cerevisiae* sites ΔΔG < 2 kcal/mol (< 4 kcal/mol)	*Drosophila melanogaster* sites ΔΔG < 2 kcal/mol (< 4 kcal/mol)
I-DmoI	15	14	0.85	−34	0 (0)	2 (33)
H-DreI	20	14	0.67	−33	0 (0)	1 (2)
I-CreI	22	13	0.54	−35	0 (1)	1 (5)
I-SceI	14	11	0.71	−26	0 (6)	2 (48)