Supplemental Information

"The mechanism and structural basis of indirect DNA sequence recognition and its impact on gene targeting by meganucleases"

Abigail R. Lambert, Jazmine P. Hallinan, Betty W. Shen, Jennifer K. Chik,
Jill M. Bolduc, Nadia Kulshina, Lori I. Robins, Brett K. Kaiser, Jordan Jarjour,
Kyle Havens, Andrew M. Scharenberg and Barry L. Stoddard

Supplemental Figures and Tables

Figure S1. Meganuclease thermal denaturation profiles and stabilities (Related to Figure 1)

Figure S2. Protein-DNA contact map for meganuclease N-terminal domains (Related to Figure 2)

Figure S3. Protein-DNA contact map for meganuclease C-terminal domains (Related to Figure 2)

Figure S4. Schematic of flow cytometric assays for (a) meganuclease expression (b) DNA binding and (c) DNA cleavage (Related to Figure 3)

Figure S5. Analysis and Quantification of Flow Cytometric Binding and Cleavage Data (Related to Figure 4)

Figure S6. Tethered flow cleavage assays for I-PanMI against DNA substrates containing all possible 'central 4' basepair sequence combinations (Related to Figure 5)

Figure S7. Overall DNA bending induced by binding of I-SmaMI to wildtype and mutated DNA target sequences (Related to Figure 6)

Figure S8. Central four cleavage specificity of I-SmaMI in the presence of Mg^{2+} vs. Mn^{2+}
(Related to Figure 6)

Table S1. I-OnuI family meganucleases characterized in this study

Supplemental Data File S1 (zipped file). Binding and cleavage data file for the 9 individual meganucleases described in this study. Related to Figures 3 and 4.

Supplemental Experimental Procedures

1. Meganuclease Sequences (coding and translated)
2. Enzyme expression, purification, biochemical analyses, crystallization and X-ray data collection
3. Meganuclease Target Site Prediction
4. Generation of DNA Substrates for Assays of Binding and Cleavage

Figure S1, related to Figure 1. Meganuclease thermal denaturation profiles and stabilities.

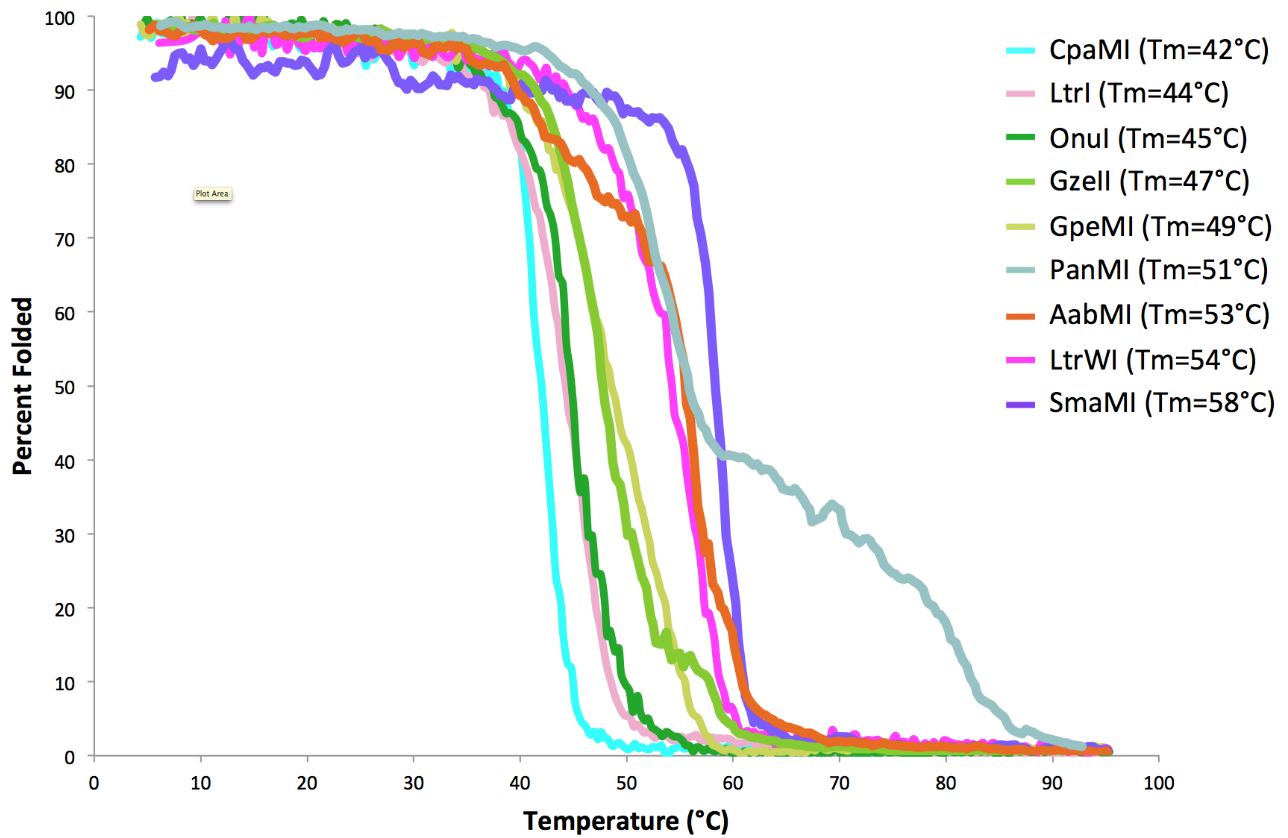


Figure S2, related to Figure 2. Protein-DNA contact map for N-terminal meganuclease domains. Colors match the corresponding protein structures from Figure 1. Blacked out regions indicate the absence of a DNA contact for that particular position while dashes indicate the absence of the amino acid itself at that position.

Domain	Tertiary Structure	Secondary Structure	Position Name	OnuI	GpeMI	CpaMI	LtrI	LtrWI	PanMI	GzeII	SmaMI	AabMI
	LAGLIDADG	Alpha 1	A1_P1	20D	19D	16D	27D	20D	18D	17D	18D	11D
N Terminal			A1_P2	21A	18A	17A	28A	21A	19G	18A	19A	12G
			A1_P3	22E	19E	18E	29E	22E	20E	19E	20E	13E
		LoopA1-B1	A1/B1_P1	23G	20G	19C	30S	23G	21G	20G	21G	14G
	Beta Sheet N Terminal	Beta 1	B1_P1	24S	21S	20S	31S	24S	22S	21S	22S	15C
			B1_P2	25F	19F	21F	32F	25F	23F	22F	23F	16F
			B1_P3	26L	23I	22S	33M	26G	24M	23M	24M	17R
			B1_P4	27L	24L	23I	34L	27V	25L	24I	25V	18I
			B1_P5	28R	25R	24L	35T	28N	26T	25H	26R	19S
			B1_P6	30R	27R	26Q	37S	30V	28I	27E	28R	21T
		LoopB1-B2	B1/B2_P1	32N	29N	29S	39D	32H	30D	29N	30N	22K
			B1/B2_P2	34K	31K	30K	41K	34T	32K	-	32K	-
			B1/B2_P3	35S	32S	31Y	42R	35N	33Y	-	33Y	-
			B1/B2_P4	36S	33S	32S	43N	36K	34K	32K	34K	-
			B1/B2_P5	37V	34A	33T	44T	37T	35L	-	35T	-
		Beta 2	B2_P1	40S	37S	36R	47S	40A	38R	34R	38L	30R
			B2_P2	41T	38T	37I	48V	41V	39V	35V	39V	31V
			B2_P3	42E	39E	38K	49R	42L	40V	36R	40V	32Q
			B2_P4	44G	41G	40V	51R	44Y	42R	38T	42I	34F
			B2_P5	46Q	43Q	42A	53R	46E	44V	40Q	44S	36Q
			B2_P6	48T	45T	44G	55G	48A	46S	42K	46T	38N
			B2_P7	49L	46L	45L	56L	49M	47L	43L	47V	39L
			B2_P8	50H	47H	45H	57H	50N	48H	44D	48D	40H
			B2_P9	53D	50D	49D	60D	53D	51D	47D	51D	43D
		Beta 3	B3_P1	68V	65V	64K	75I	69N	66N	63S	67S	58K
			B3_P2	70A	67A	66H	77T	71Y	68F	65N	69K	60H
			B3_P3	71N	68N	67I	78S	72H	69L	66T	70K	61I
		LoopB3-B4	B3/B4_P1	72S	69S	68H	79D	73N	70M	-	71S	62S
			B3/B4_P2	73G	70G	69G	80I	74P	71T	67S	72G	63G
			B3/B4_P3	75N	-	-	-	76D	-	69K	74S	-
		Beta 4	B4_P1	76A	73A	72T	81D	78T	74S	70E	75T	66L
			B4_P2	78S	75S	74Q	83R	80K	76Q	72V	77S	68Q
			B4_P3	80K	77K	76R	85R	82K	78R	74K	79R	70R
		LoopB4-A3	B4/A3_P1	82T	79T	78D	87E	84S	80E	76R	81E	72Q
			B4/A3_P2	83R	80R	79S	88S	85N	81S	77S	82S	73T
			B4/A3_P3	84F	81F	80P	89L	86I	82L	78L	83S	74F
			B4/A3_P4	85E	82E	81K	90K	87D	83K	79K	84E	75D
		Alpha 4	A4_P1	103K	100K	99K	108K	106K	118K	97K	103K	93K
		LoopA4-A5	A4/A5_P1	117M	114M	113I	122I	120I	115I	111L	117	
			A4/A5_P2	120K	117K	116K	125K	123K	118K	114Q	120K	110N
			A4/A5_P3	122H	119H	118H	127H	125H	120H	116H	122H	112H
			A4/A5_P4	123L	120L	119L	128L	126L	121L	117L	123L	113L
			A4/A5_P5	-	-	-	-	-	-	118S	-	-
		Alpha 5	A5_P1	135K	132K	131K	140K	138K	133K	133R	135K	125K
		LoopA5-A6	A5/A6_P1	138L	135L	134L	143L	141M	136I	136M	138I	128L
			A5/A6_P2	139N	136N	135N	144N	142N	137N	137N	139N	129N
			A5/A6_P3	140W	137W	136L	145L	143L	138N	138L	140K	130L
			A5/A6_P4	141G	138G	137G	146G	144G	139G	139G	139G	131G
			A5/A6_P5	143T	140T	139N	148S	146S	141N	141S	143S	133S

	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11
I-AabMI	A	G	G	T	A	C	C	C	T	T	T	A	A	A	C	C	T	A	C	T	A	A
I-CpaMI	T	A	G	C	C	C	A	C	A	A	T	A	T	T	A	A	G	G	C	C	A	T
I-GpeMI	T	T	T	C	C	G	C	T	T	A	T	T	C	A	A	C	C	C	T	T	T	A
I-GzeII	A	T	G	G	G	T	A	C	C	A	T	A	T	T	G	G	T	A	C	A	A	A
I-LtrI	A	A	T	G	C	T	C	C	T	A	T	A	C	G	A	C	G	T	T	T	A	G
I-LtrWI	A	G	T	A	G	T	G	A	A	G	T	A	T	G	T	T	A	T	T	T	A	A
I-OnuI	T	T	T	C	C	A	C	T	T	A	T	T	C	A	A	C	C	T	T	T	T	A
I-PanMI	G	C	T	C	C	T	C	A	T	A	A	T	C	C	T	T	A	T	C	A	A	G
I-SmaMI	T	A	T	C	C	T	C	C	A	T	T	A	T	C	A	G	G	T	G	T	A	C

Figure S3, related to Figure 2. Protein-DNA contact map for C-terminal meganuclease domains. Colors match the corresponding protein structures from Figure 1. Blacked out regions indicate the absence of a DNA contact for that particular position while dashes indicate the absence of the amino acid itself at that position.

Domain	Tertiary Structure	Secondary Structure	Position Name	OnuI	GpeMI	CpaMI	LtrI	LtrWI	PanMI	GzeII	SmaMI	AabMI
C Terminal	LAGLIDADG	Alpha 7	A7_P1	176S	181S	175S	182A	179E	185D	175S	177S	167S
			A7_P2	177G	182G	176G	183G	180G	176A	176G	178G	168A
			A7_P3	178E	183E	177D	184E	181E	177E	177E	179D	169E
	Beta Sheet C Terminal	Beta 5	B5_P1	179G	184G	178S	185G	182A	178G	176G	180G	170G
			B5_P2	180C	185C	179S	186S	183C	179C	179S	181S	171C
			B5_P3	181F	186F	180F	187F	184F	180F	180F	182F	172F
			B5_P4	182F	187F	181N	188Y	185F	181S	181S	183K	173M
			B5_P5	183V	188V	182V	189I	186V	182V	182V	184S	174V
			B5_P6	184N	189S	183K	190R	187S	183V	183Y	185I	
			B5_P7	185L	190L	184I	191I	188I	184F	184T	186L	176I
			B5_P8	186I	191I	185S	192A	189Y	185F	185T	187K	177A
			B5_P9	187K	192K	186N	193K	190N	186K	186S	188K	178K
		LoopB5-B6	B5/B6_P1	188S	193S	187S	194N	191S	187S	187D	189S	179S
			B5/B6_P2	189K	194K	189T	196T	193K	189T	-	191S	181A
			B5/B6_P3	190S	195S	190S	197L	194S	190S	-	192I	182S
			B5/B6_P4	191K	196K	191L	198K	195K	191K	-	193K	183S
		Beta 6	B6_P1	195Q	200Q	195R	202Q	199A	195A	190Y	197Q	187Q
			B6_P2	196V	201V	196V	203V	200V	196V	191V	198S	188V
			B6_P3	197Q	202Q	197Q	204Q	201Q	197K	192S	199I	189Y
			B6_P4	199V	204V	199R	206V	203V	199S	194S	201V	191T
			B6_P5	201S	206S	201G	208Q	205K	201I	196R	203Q	193I
			B6_P6	203T	208T	203G	210T	207T	203T	198S	205T	195T
			B6_P7	204Q	209Q	204L	211Q	208Q	204Q	199Q	206Q	196Q
		LoopB6-A8	B6/A8_P1	205H	210H	205N	212D	209H	205S	200H	207H	197H
			B6/A8_P2	207K	212R	207R	214R	211R	207R	202K	209R	199R
		Beta 7	B7_P1	-	-	226K	-	-	-	-	-	-
			B7_P2	222Y	228Y	227N	230N	227R	223N	217G	225F	215R
			B7_P3	223I	229I	228I	231I	228V	224T	218F	226I	216
			B7_P4	225K	230K	229Y	232R	229E	225S	220N	227E	217A
			B7_P5	227K	232K	-	234R	231R	227D	222H	229D	219K
			B7_P6	228N	233K	-	235K	232K	228P	223N	230S	-
			B7_P7	229K	234K	-	-	-	-	224K	231R	220R
		Beta 8	B8_P1	232F	237F	-	-	-	-	-	-	-
			B8_P2	233S	238S	-	245T	235E	230G	227K	233P	221N
			B8_P3	234W	239W	234S	246C	236A	231T	228A	234W	222V
			B8_P4	236D	241E	236R	248D	238D	233D	230I	236Y	224E
			B8_P5	238V	243V	238E	250V	242T	235K	232V	238T	226Q
		LoopB8-A9	B8/A9_P1	240T	245T	240V	252T	241V	237T	234R	240T	228S
			B8/A9_P2	241K	246K	241K	253N	242S	238N	235K	241N	229K
			B8/A9_P3	242F	247F	242F	254L	244I	239F	236F	242F	230F
			B8/A9_P4	243S	248S	243S	255D	245K	240S	237E	243S	231S
		Alpha 10	A10_P1	262K	267K	262K	274K	264K	259K	256K	262K	249K
			A10_P2	265D	284K	265D	277D	267N	262D	259D	265D	252D
		LoopA10-A11	A10/A11_P1	281H	286H	281H	293H	283H	278H	275H	281H	268H
			A10/A11_P2	282L	287L	282L	294L	284L	279L	276L	282L	269L
		Alpha 11	A11_P1	294K	299K	294K	306K	296K	291R	288K	294K	281K
			A11_P2	297M	302M	297M	309M	299M	294M	291M	297M	284M
			A11_P3	298N	303N	298N	310N	300N	295N	292N	298N	285N
			A11_P4	299K	304K	-	311R	301T	296T	293S	299K	286R
			A11_P5	-	-	-	-	-	-	294Y	-	-

I-AabMI A G G T A C C C T T T A A A C C T A C T A A

I-CpaMI T A G C C C A C A A T A T T A A G G C C A T

I-GpeMI T T T C C G C T T A T T C A A C C C T T T A

I-GzeII A T G G G T A C C A T A T T G G T A C A A A

I-LtrI A A T G C T C C T A T A C G A C G T T T A G

I-LtrWI A G T A G T G A A G T A T G T T A T T T A A

I-OnuI T T T C C A C T T A T T C A A C C T T T T A

I-PanMI G C T C C T C A T A A T C C T T A T C A A G

I-SmaMI T A T C C T C C A T T A T C A G G T G T A C

Figure S4, related to Figure 3. Schematic of flow cytometric assays for *(a)* meganuclease expression *(b)* DNA binding and *(c)* DNA cleavage. ***Panel a:*** A plot of PE (N-terminal stain) vs. FITC (C-terminal stain) was examined for each enzyme to verify expression of full-length, stable protein on the surface of the yeast. ***Panel b:*** Yeast with surface-expressed meganuclease are stained with a FITC antibody to the C-terminal Myc epitope tag and incubated with fluorescently labeled DNA duplexes (A647 fluorophore) containing a sequence corresponding either to the wild-type target or to a target containing a single basepair substitution. By assaying the relative signal for bound DNA (A647) at concentrations corresponding to the approximate wild-type K_D, basepair substitutions that significantly reduce affinity can be identified. ***Panel c:*** Yeast with surface-expressed meganuclease are incubated with an anti-HA-biotin antibody, a streptavidin-PE stain, and fluorescently labeled DNA substrate (A647 fluorophore) containing a sequence corresponding to either to the wild-type target or a target containing a single basepair substitution. The biotin and streptavidin molecules create a physical tether of the DNA target substrate to the N-terminus of the protein. A plot of PE (N-term protein) vs. A647 (DNA) signals is collected for each DNA construct in both the presence of calcium (which facilitates binding but prevents cleavage) and in the presence of magnesium (which supports cleavage). By superimposing the calcium and magnesium plots, one can determine the relative cleavage activity against each DNA target by comparing loss of the A647 signal (representing cleavage of the tethered DNA target substrate, indicated by arrow.)

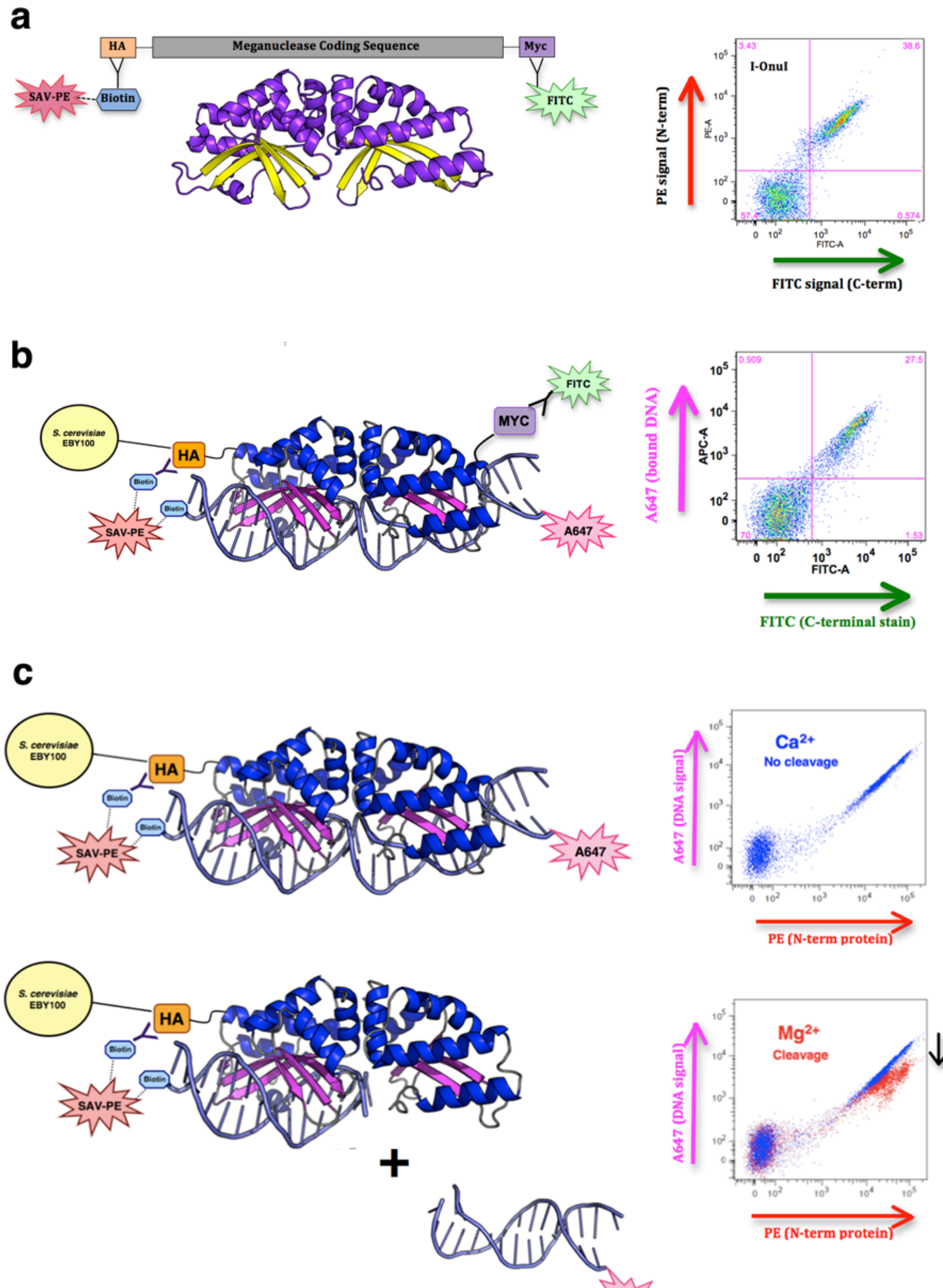
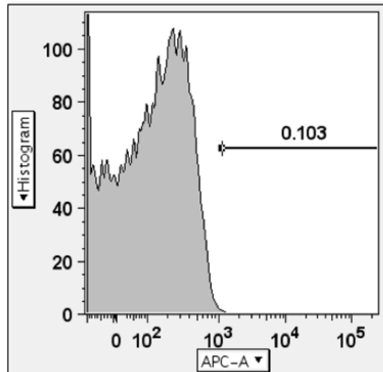
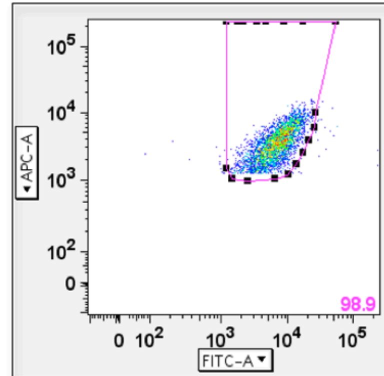


Figure S5, related to Figure 4. Analysis and Quantification of Flow Cytometric Binding and Cleavage Data. ***Panel a:*** Use the "no stain" control to draw a range gate that will include all cells with signal from bound DNA (A647/APC). This gate excludes the negative population. ***Panel b:*** Using the wild-type sample, draw a tight gate around the cells with signal for the bound DNA. Let the top of the gate extend upwards to include any samples that might have higher signal than wild-type. Activate the number/statistic for COUNT (number of cells) inside this gate. Use the numbers of cells inside the gate for quantification of binding relative to the wild-type sample. ***Panel c:*** Use a "no stain" control to draw a range gate for "Expressed" cells (cells with FITC signal). Apply to all samples. From this point forward, you will only be looking at cells with a FITC signal (cells with full-length expressed protein). ***Panel d:*** Draw a rectangular gate to capture a "slice" of the expressed cells (can be a thin or wide rectangle, but positioned at a FITC level that is uniform across all samples). Apply this gate to all samples. ***Panel e:*** Draw a second rectangular gate (thin) to create a "PE slice". Apply to all samples. Activate the statistical value/number for MEDIAN APC signal inside this gate. For each sample, record values for the matched $\mathbf{Ca^{2+}}$ and the $\mathbf{Mg^{2+}}$ samples. The shift observed in this assay can be quantified by calculating the $\mathbf{Ca^{2+}/Mg^{2+}}$ ratio.

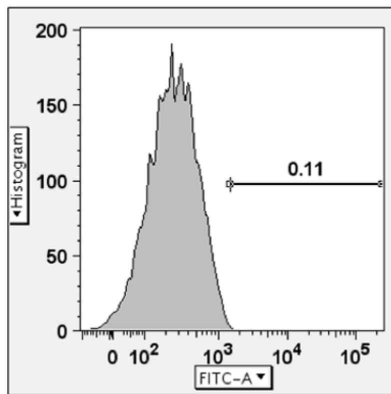
a



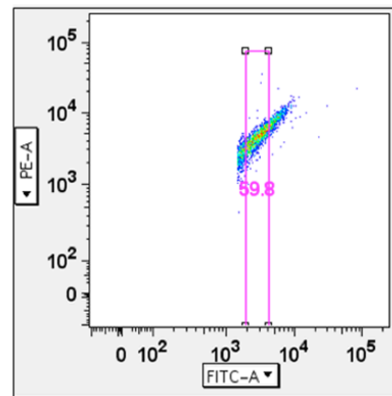
b



c



d



e

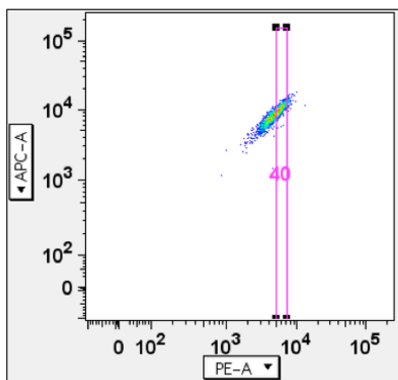
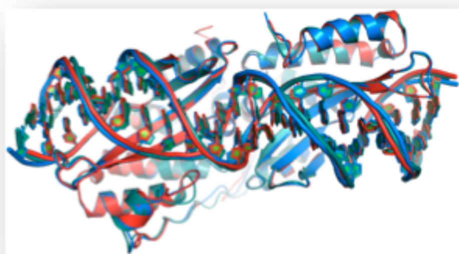


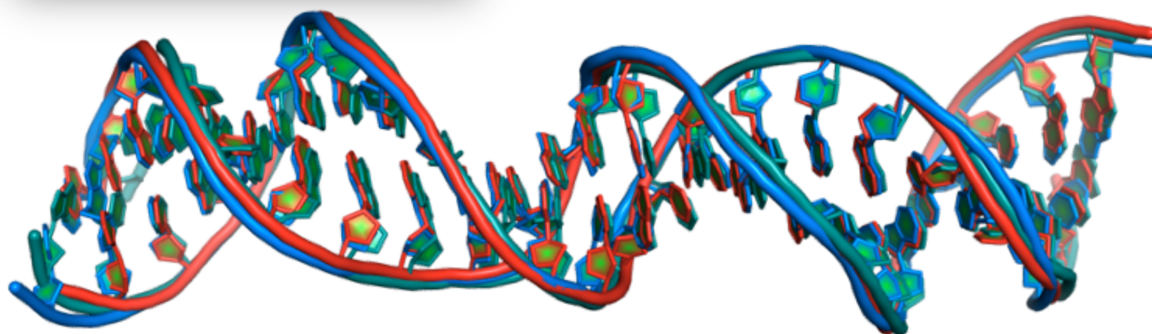
Figure S6, related to Figure 5. Raw data from the flow cytometric tethered cleavage assay for I-PanMI against DNA substrates containing all possible 'central 4' basepair sequences.

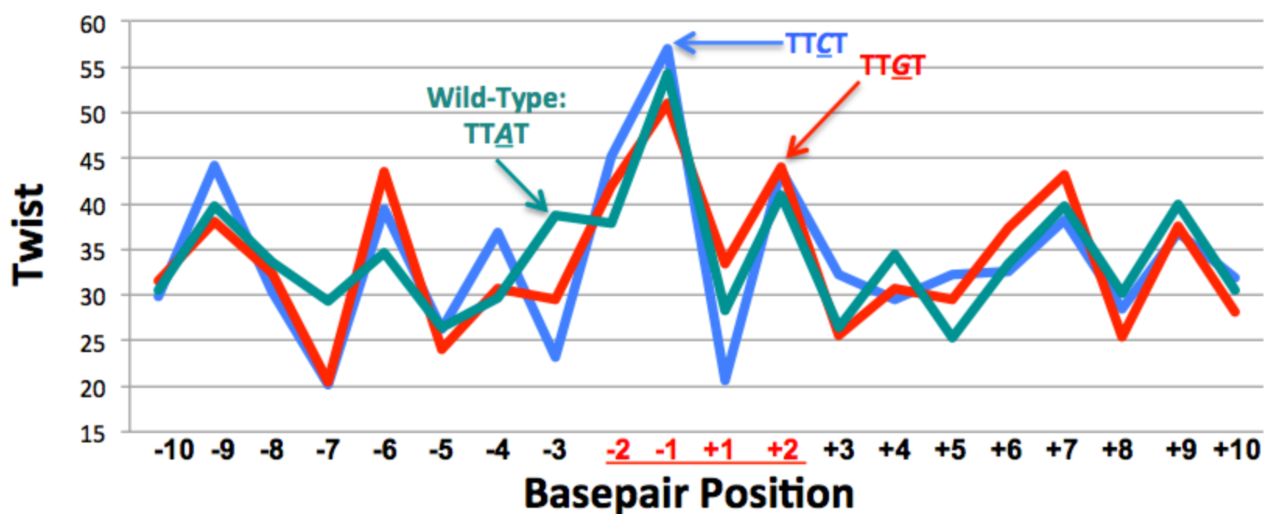
/2015
ll 256 c4's plate 2

01/13/2015
PanMI c4 all
plate 1 repeat
cleavage 7.9

Figure S7, related to Figure 6. 3DNA structural analysis of overall DNA bending induced by binding of I-SmaMI to wildtype and mutated DNA target sequences.







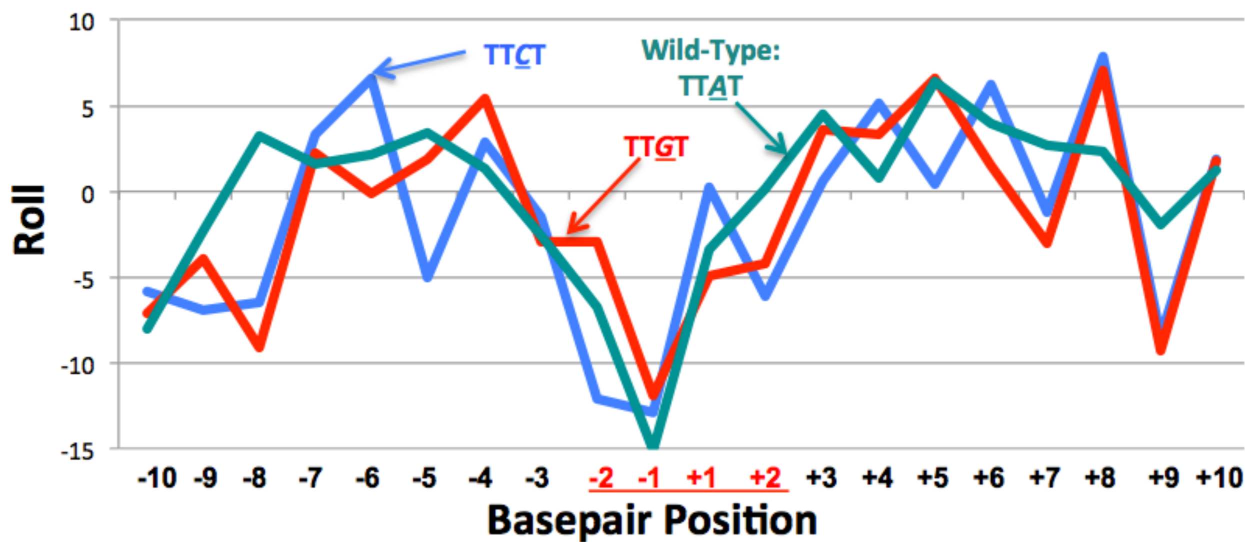
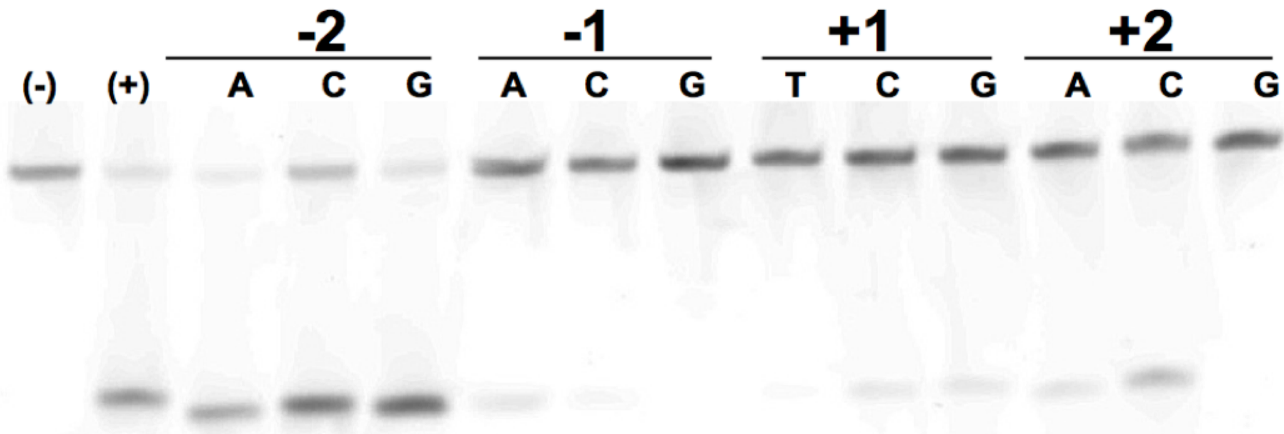


Figure S8, related to Figures 4 and 6. Central four cleavage specificity of I-SmaMI in the presence of Mg^{2+} vs. Mn^{2+}.



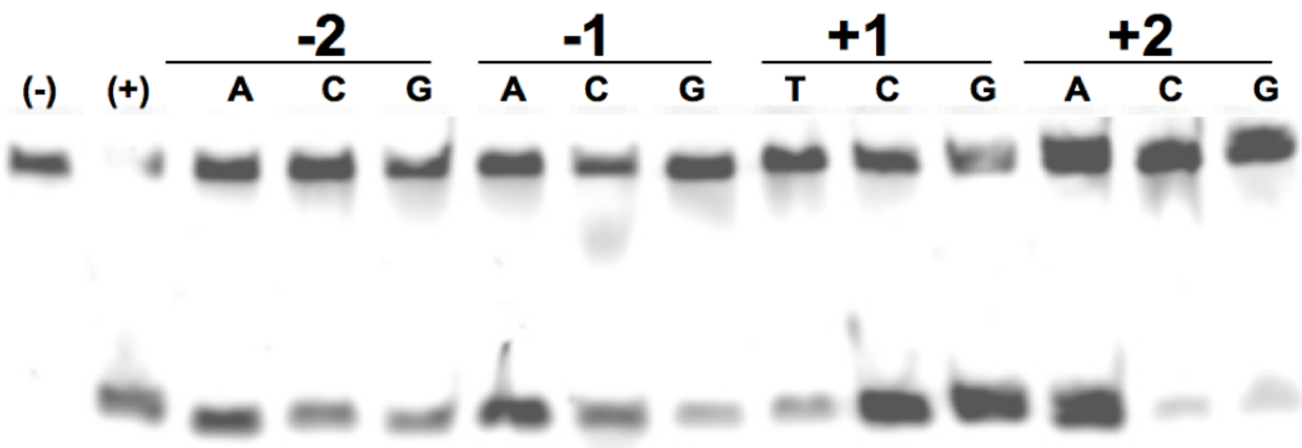


Table S1, related to Figure 1. I-OnuI family meganucleases characterized in this study.

Nuclease	Host organism	Host Gene	Surface Expression	Cleavage Activity	Crystal structure
I-AabMI	Ascocalyx abietina	Small subunit ribosomal RNA gene	✓	✓	✓
I-AaeMIP	Agrocybe aegerita	Small subunit ribosomal RNA gene			
I-ApaMIP	Amoebidium parasiticum	Large subunit ribosomal RNA gene			
I-CkaMI	Cordyceps kanzashiana	Small subunit ribosomal RNA gene	✓	✓	
I-CpaMI	Cryphonectria parasitica	Small subunit ribosomal RNA gene	✓	✓	✓
I-CpaMIIP	Cryphonectria parasitica	S5 ribosomal protein	✓		
I-CpaMIIIP	Cryphonectria parasitica	Small subunit ribosomal RNA gene	✓		
I-CraMI	Cordyceps ramosopulvinata	Small subunit ribosomal RNA gene	✓	✓	
I-GpeMI	Grosmannia penicillata	Large subunit ribosomal RNA gene	✓	✓	✓
I-GpiI	Grosmannia piceiperda	Large subunit ribosomal RNA gene	✓	✓	
I-GzeI	Gibberella zeae	NADH:ubiquinone oxidoreductase subunit 3	✓	✓	
I-GzeII	Gibberella zeae	NADH:ubiquinone oxidoreductase subunit 2	✓	✓	✓
I-GzeMIIIP	Gibberella zeae	NADH:ubiquinone oxidoreductase subunit 2			
I-HjeMII	Hypocrea jecorina	NADH dehydrogenase subunit 2	✓	✓	
I-LtrI	Leptographium truncatum	Large subunit ribosomal RNA gene	✓	✓	✓
I-LtrWI	Leptographium truncatum	Small subunit ribosomal RNA gene	✓	✓	✓
I-MpeMI	Moniliophthora perniciosa	NADH dehydrogenase subunit 4	✓	✓	
I-MveMIP	Mortierella verticillata	Small subunit ribosomal RNA gene	✓		
I-NcrMIP	Neurospora crassa	ATPase subunit 8			
I-OnuI	Ophiostoma novo-ulmi	Large subunit ribosomal RNA gene	✓	✓	✓
I-OsoMI	Ophiocordyceps sobolifera	Small subunit ribosomal RNA gene	✓	✓	
I-OsoMII	Ophiocordyceps sobolifera	Small subunit ribosomal RNA gene	✓	✓	
I-OsoMIIIP	Ophiocordyceps sobolifera	Small subunit ribosomal RNA gene			
I-OsoMIVP	Ophiocordyceps sobolifera	Small subunit ribosomal RNA gene			
I-PanMI	Podospora anserina	NADH dehydrogenase subunit 2	✓	✓	✓
I-PanMIIP	Podospora anserina	cytochrome c oxidase cox1			
I-PanMIIIP	Podospora anserina	Dod ND4L i2 grp IC protein			
I-PnoMI	Phaeosphaeria nodorum	NADH dehydrogenase subunit 5	✓		
I-ScuMIP	Smittium culisetae	Large subunit ribosomal RNA gene			
I-ScuMIIP	Smittium culisetae	cytochrome c oxidase cox1			
I-ScuMIIIP	Smittium culisetae	cytochrome c oxidase cox1			
I-ScuMIVP	Smittium culisetae	cytochrome c oxidase cox1			
I-SmaMI	Sordaria macrospora	cytochrome c oxidase cox1	✓	✓	✓
I-SscMI	Sclerotinia sclerotiorum	Small subunit ribosomal RNA gene	✓	✓	

Supplemental Experimental Procedures

1. Meganuclease Sequences (coding and translated)

Note: Highlighted residues are mutations designed to decrease surface hydrophobicity

I-AabMI (AF284853)

**STSDNNGNIKINPWFLTGFIDGEGCFRISVTKINRAIDWRVQLFFQINLHEKDRALLESIKDYLGVGKIHISGKN
LVQYRIQTFDELTILIKHLKEYPLVSKKRADFELFNTAHKLIKNNEHLNKEGINKLVSLKASLNLGLSESLKLA
FPNVISATRLTDFTVNIPDPHWLSGFASAEGCFMVGIAKSSASSTGYQVYLTFILTQHVRDENLMKCLVDYFN
WGRLARKRNVYEYQVSKFSDVEKLLSFFDKYPILGEKAKDLQDFCSVSDLMKSKTHLTEEGVAKIRKIKEGM
NRGR**

TCTACTTCTGACAACAACGGTAACATCAAGATCAACCCATGGTTCCTGACTGGTTTCATCGACGGTGAAGGTTG
TTTCAGAATCTCTGTCACTAAGATCAACAGAGCTATCGACTGGAGAGTCCAGCTGTTCTTCCAGATCAACCTGC
ACGAAAAGGACAGAGCTCTGCTGGAATCTATCAAGGACTATCTGGGTGTCGGTAAGATCCACATCTCTGGTAA
GAACCTGGTCCAGTATAGAATCCAGACTTTCGACGAACTGACTATCCTGATCAAGCACCTGAAGGAATATCCA
CTGGTCTCTAAGAAGAGAGCTGACTTCGAACTGTTCAACACTGCTCACAAGCTGATCAAGAACAACGAACACC
TGAACAAGGAAGGTATCAACAAGCTGGTCTCTCTGAAGGCTTCTCTGAACCTGGGTCTGTCTGAATCTCTGAAG
CTGGCTTTCCCAAACGTCATCTCTGCTACTAGACTGACTGACTTCACTGTCAACATCCCAGACCCACACTGGCT
GTCTGGTTTCGCTTCTGCTGAAGGTTGTTTCATGGTCGGTATCGCTAAGTCTTCTGCTTCTTCTACTGGTTATCA
GGTCTATCTGACTTTCATCCTGACTCAGCACGTCAGAGACGAAAACCTGATGAAGTGTCTGGTCGACTATTTCA
ACTGGGGTAGACTGGCTAGAAAGAGAAACGTCTATGAATATCAGGTCTCTAAGTTCTCTGACGTCGAAAAGCT
GCTGTCTTTCTTCGACAAGTATCCAATCCTGGGTGAAAAGGCTAAGGACCTGCAGGACTTCTGTTCTGTCTCTG
ACCTGATGAAGTCTAAGACTCACCTGACTGAAGAAGGTGTCGCTAAGATCAGAAAGATCAAGGAAGGTATGA
ACAGAGGTAGA

I-CpaMI (AAB84212.1)

**NTSSSFNPWFLTGFSDAECSFSILIQANSKYSTGWRIKPVFAIGLHKKDNELLKRIQSYLGVGKIHIHGKDSIQF
RIDSPKELEVIINHFENYPLVTAKQADYTLFKKALDVIKNKEHLSQKGLLKLVGIKASLNLGLNGSLKEAFPN
WEELQIDRPSYVNKGIPDPNWISGFASGDSSFNVKISNSPTSLLNKRVQLRFGIGLNIREKALIQYLVAYFDLSD
NLKNIYFDLNSARFEVVKFSDITDKIIPFFDKYSIQGKKSQDYQNFKEVADIIKSKNHLTSEGFQEILDIKASMN
K**

AACACTTCTTCTTCTTTCAACCCATGGTTCCTGACTGGTTTCTCTGACGCTGAATGTTCTTTCTCTATCCTGATCC
AGGCTAACTCTAAGTATTCTACTGGTTGGAGAATCAAGCCAGTCTTCGCTATCGGTCTGCACAAGAAGGACAA
CGAACTGCTGAAGAGAATCCAGTCTTATCTGGGTGTCGGTAAGATCCACATCCACGGTAAGGACTCTATCCAG
TTCAGAATCGACTCTCCAAAGGAACTGGAAGTCATCATCAACCACTTCGAAAACTATCCACTGGTCACTGCTAA
GCAGGCTGACTATACTCTGTTCAAGAAGGCTCTGGACGTCATCAAGAACAAGGAACACCTGTCTCAGAAGGGT
CTGCTGAAGCTGGTCGGTATCAAGGCTTCTCTGAACCTGGGTCTGAACGGTTCTCTGAAGGAAGCTTTCCCAAA
CTGGGAAGAACTGCAGATCGACAGACCATCTTATGTCAACAAGGGTATCCCAGACCCAAACTGGATCTCTGGT
TTCGCTTCTGGTGACTCTTCTTTCAACGTCAAGATCTCTAACTCTCCAACTTCTCTGCTGAACAAGAGAGTCCAG
CTGAGATTCGGTATCGGTCTGAACATCAGAGAAAAGGCTCTGATCCAGTATCTGGTCGCTTATTTCGACCTGTC
TGACAACCTGAAGAACATCTATTTCGACCTGAACTCTGCTAGATTCGAAGTCGTCAAGTTCTCTGACATCACTG
ACAAGATCATCCCATTCTTCGACAAGTATTCTATCCAGGGTAAGAAGTCTCAGGACTATCAGAACTTCAAGGA
AGTCGCTGACATCATCAAGTCTAAGAACCACCTGACTTCTGAAGGTTTCCAGGAAATCCTGGACATCAAGGCTT
CTATGAACAAG

I-GpeMI (ACV41146.1)

**PTRNESINPWVLTGFADAEGSFILRIRNNNKSSAGYSTELGFQITLHKKDKSILENIQSTWKVGVIANSGDNAVS
LKVTRFEDLRVVLNHFEKYPLITQKLGDYLLFKQAFSVMENKEHLKIEGIKRLVGIKANLNWGLTDELKEAF
VASGGENIFVASGGERSLINKNIPNSGWLAGFTSGEGCFFVSLIKSKSKLGVQVQLVFSITQHARDRELMDNLV
TYLGCGYIKEKKKSEFSWLEFVVTKFSDIKDKIIPVFQENNIIGVKLEDFEDWCKVAKLIEEKKHLTESGLEEI
RNIKLNMNKGRVL**

CCAACTAGAAACGAATCTATCAACCCATGGGTCCTGACTGGTTTCGCTGACGCTGAAGGTTCTTTCATCCTGAG
AATCAGAAACAACAACAAGTCTTCTGCTGGTTATTCTACTGAACTGGGTTTCCAGATCACTCTGCACAAGAAGG

ACAAGTCTATCCTGGAAAACATCCAGTCTACTTGGAAGGTCGGTGTCATCGCTAACTCTGGTGACAACGCTGTC
TCTCTGAAGGTCACTAGATTCGAAGACCTGAGAGTCGTCCTGAACCACTTCGAAAAGTATCCACTGATCACTCA
GAAGCTGGGTGACTATCTGCTGTTCAAGCAGGCTTTCTCTGTCATGGAAAACAAGGAACACCTGAAGATCGAA
GGTATCAAGAGACTGGTCGGTATCAAGGCTAACCTGAACTGGGGTCTGACTGACGAACTGAAGGAAGCTTTCG
TCGCTTCTGGTGGTGAAAACATCTTCGTCGCTTCTGGTGGTGAAAGATCTCTGATCAACAAGAACATCCCAAAC
TCTGGTTGGCTGGCTGGTTTCACTTCTGGTGAAGGTTGTTTCTTCGTCTCTCTGATCAAGTCTAAGTCTAAGCTG
GGTGTCCAGGTCCAGCTGGTCTTCTCTATCACTCAGCACGCTAGAGACAGAGAACTGATGGACAACCTGGTCA
CTTATCTGGGTTGTGGTTATATCAAGGAAAAGAAGAAGTCTGAATTCTCTTGGCTGGAATTCGTCGTCACTAAG
TTCTCTGACATCAAGGACAAGATCATCCCAGTCTTCCAGGAAAACAACATCATCGGTGTCAAGCTGGAAGACT
TCGAAGACTGGTGTAAGGTCGCTAAGCTGATCGAAGAAAAGAAGCACCTGACTGAATCTGGTCTGGAAGAAAT
CAGAAACATCAAGCTGAACATGAACAAGGGTAGAGTCCTG

I-GzeII (ABC86622.1)

**DLSTSINPWFVTGFTDAEGSFMIHLEKNKDKWRVRPTFQIKLDIRDKSLLEEIKNYFNNTGSINTSNKECVYKV
RSLKDISIIISHFDKYNLITQKKADFELFKKIINKLNSQEHLSYEVGATVLQEIISIRASMNLGLSSSVKEDFPHIIP
SNRPLIENMNIPHPEWMAGFVSGEGSFSVYTTSDDKYVSLSFRVSQHNKDKQLLKSFVDFFGCGGFNYHNKG
NKAVIFVTRKFEDINDKIIPLFNEYKIKGVKYKDFKDWSKVAKMIESKSHLTTNGYKEICKIKENMNSYRKSSV
N**

GACCTGTCTACTTCTATCAACCCATGGTTCGTCACTGGTTTCACTGACGCTGAAGGTTCTTTCATGATCCACCTG
GAAAAGAACAAGGACAAGTGGAGAGTCAGACCAACTTTCCAGATCAAGCTGGACATCAGAGACAAGTCTCTG
CTGGAAGAAATCAAGAACTATTTCAACAACACTGGTTCTATCAACACTTCTAACAAGGAATGTGTCTATAAGG
TCAGATCTCTGAAGGACATCTCTATCATCATCTCTCACTTCGACAAGTATAACCTGATCACTCAGAAGAAGGCT
GACTTCGAACTGTTCAAGAAGATCATCAACAAGCTGAACTCTCAGGAACACCTGTCTTATGAAGTCGGTGCTA
CTGTCCTGCAGGAAATCATCTCTATCAGAGCTTCTATGAACCTGGGTCTGTCTTCTTCTGTCAAGGAAGACTTCC
CACACATCATCCCATCTAACAGACCACTGATCGAAAACATGAACATCCCACACCCAGAATGGATGGCTGGTTT
CGTCTCTGGTGAAGGTTCTTTCTCTGTCTATACTACTTCTGACGACAAGTATGTCTCTCTGTCTTTCAGAGTCTCT
CAGCACAACAAGGACAAGCAGCTGCTGAAGTCTTTCGTCGACTTCTTCGGTTGTGGTGGTTTCAACTATCACAA
CAAGGGTAACAAGGCTGTCATCTTCGTCACTAGAAAGTTCGAAGACATCAACGACAAGATCATCCCACTGTTC
AACGAATATAAGATCAAGGGTGTCAAGTATAAGGACTTCAAGGACTGGTCTAAGGTCGCTAAGATGATCGAAT
CTAAGTCTCACCTGACTACTAACGGTTATAAGGAAATCTGTAAGATCAAGGAAAACATGAACTCTTATAGAAA
GTCTTCTGTCAAC

I-LtrWI (ADN05145.1)

**MINLKNNIEYLNWYICGLVDAEGSFGVNVVKHATNKTGYAVLTYFELAMNSKDKQLLELIKKTFDLECNIYH
NPSDDTLKFKVSNIEQIVNKIIPFFEKYTLFSQKRGDFILFCKVVELIKNKEHLTLNGLNKILSIKAAMNLGLSE
NLKKEFPGCLSVKRPEFGLSNLNKRWLAGFIEGEACFFVSIYNSPKSKLGKAVQLVFKITQHIRDKILIESIVEL
LNCGRVEVRKSNEACDFTVTSIKEIENYIIPFFNEYPLIGQKLKNYEDFKKIFDMMKTKDHLTEEGLSKIIEIKN
KMNTNR**

ATGATCAACCTGAAGAACAACATCGAATATCTGAACTGGTATATCTGTGGTCTGGTCGACGCTGAAGGTTCTTT
CGGTGTCAACGTCGTCAAGCACGCTACTAACAAGACTGGTTATGCTGTCCTGACTTATTTCGAACTGGCTATGA
ACTCTAAGGACAAGCAGCTGCTGGAACTGATCAAGAAGACTTTCGACCTGGAATGTAACATCTATCACAACCC
ATCTGACGACACTCTGAAGTTCAAGGTCTCTAACATCGAACAGATCGTCAACAAGATCATCCCATTCTTCGAAA
AGTATACTCTGTTCTCTCAGAAGAGAGGTGACTTCATCCTGTTCTGTAAGGTCGTCGAACTGATCAAGAACAAG
GAACACCTGACTCTGAACGGTCTGAACAAGATCCTGTCTATCAAGGCTGCTATGAACCTGGGTCTGTCTGAAA
ACCTGAAGAAGGAATTCCCAGGTTGTCTGTCTGTCAAGAGACCAGAATTCGGTCTGTCTAACCTGAACAAGAG
ATGGCTGGCTGGTTTCATCGAAGGTGAAGCTTGTTTCTTCGTCTCTATCTATAACTCTCCAAAGTCTAAGCTGGG
TAAGGCTGTCCAGCTGGTCTTCAAGATCACTCAGCACATCAGAGACAAGATCCTGATCGAATCTATCGTCGAA
CTGCTGAACTGTGGTAGAGTCGAAGTCAGAAAGTCTAACGAAGCTTGTGACTTCACTGTCACTTCTATCAAGGA
AATCGAAAACTATATCATCCCATTCTTCAACGAATATCCACTGATCGGTCAGAAGCTGAAGAACTATGAAGAC
TTCAAGAAGATCTTCGACATGATGAAGACTAAGGACCACCTGACTGAAGAAGGTCTGTCTAAGATCATCGAAA
TCAAGAACAAGATGAACACTAACAGA

I-PanMI (NP_074914.1)

**STLESKLNPSYISGFVDGEGSFMLTIIKDNKYKLGWRVVCRFVISLHKKDLSLLNKIKEFFDVGNVFLMTKDSA
QYRVESLKGLDLIINHFDKYPLITKKQADYKLFKMAHNLIKNKSHLTKEGLLELVAIKAVINNGLNNDLSIAFP
GINTILRPDTSLPQILNPFWLSGFVDAEGCFSVVVFKSKTSKLGEAVKLSFILTQSNRDEYLIKSLIEYLGCGNT**

**SLDPRGTIDFKVTNFSSIKDIIVPFFIKYPLKGNKNLDFTDFCEVVRLMENKSHLTKEGLDQIKKIRNRMNTNR
K**

TCTACTTTGGAATCTAAGTTGAACCCATCTTACATCTCTGGTTTCGTCGACGGTGAAGGTTCTTTCATGTTGACT
ATCATCAAGGACAACAAGTACAAGTTGGGTTGGAGAGTTGTTTGTAGATTCGTTATCTCTTTGCACAAGAAGG
ACTTGTCTTTGTTGAACAAGATCAAGGAATTTTTCGACGTCGGTAACGTTTTCTTGATGACTAAGGACTCTGCTC
AATACAGAGTTGAATCTTTGAAGGGTTTGGACTTGATCATCAACCACTTCGACAAGTACCCATTGATCACTAAG
AAGCAAGCTGACTACAAGTTGTTCAAGATGGCTCACAACTTAATTAAGAACAAGTCTCACTTGACTAAGGAAG
GTTTGTTGGAATTGGTTGCTATCAAGGCTGTTATCAACAACGGTTTGAACAACGACTTGTCTATCGCTTTCCCA
GGTATCAACACTATCTTGAGGCCTGACACTTCTTTGCCACAAATCTTGAACCCATTCTGGTTGTCTGGTTTCGTT
GACGCTGAAGGTTGTTTCTCTGTTGTTGTTTTCAAGTCTAAGACTTCTAAGTTGGGTGAAGCTGTTAAGTTGTCT
TTCATCTTGACTCAATCTAACAGAGACGAATACTTGATCAAGTCTTTGATCGAATACCTAGGTTGTGGTAACAC
TTCTTTGGACCCAAGAGGTACTATCGACTTCAAGGTTACTAACTTCTCTTCTATCAAGGACATCATCGTTCCATT
CTTCATCAAGTACCCATTGAAGGGTAACAAGAACTTGGACTTCACTGACTTCTGTGAAGTTGTTAGATTGATGG
AAAACAAGTCTCACTTGACTAAGGAAGGTTTGGACCAAATCAAGAAGATCAGAAACAGAATGAACACTAACA
GAAAG

I-SmaMI (XP_003342391.1)

**SKGENSKLNPWAVVGFIDAEGSFMVRVRKNSKYKTGWLVVAIFSVTVDKKDLFLLESLKTFFGGLGSIKKSG
NSTFSYRIESSEQLTKIILPFFDKYSLITEKLGDYLLFKKVLELMGTKEHLTQRGLEKIVSLKASINKGLSEELQ
AAFPQCVPTPRPEINNKNIPDPFWLAGFVSGDGSFKSILKKSESIKVGFQSILVFQITQHARDVKLMESLISYLG
CGFIEKDSRGPWLYYTVTNFSDIQGKIIPFFHQYKIIGSKYGDYQDWCKIALIMQNKNHLTPEGLNEIRALKG
GMNKGRL**

TCTAAGGGTGAAAACTCTAAGCTGAACCCATGGGCTGTCGTCGGTTTCATCGACGCTGAAGGTTCTTTCATGGT
CAGAGTCAGAAAGAACTCTAAGTATAAGACTGGTTGGCTGGTCGTCGCTATCTTCTCTGTCACTGTCGACAAGA
AGGACCTGTTCCTGCTGGAATCTCTGAAGACTTTCTTCGGTGGTCTGGGTTCTATCAAGAAGTCTGGTAACTCT
ACTTTCTCTTATAGAATCGAATCTTCTGAACAGCTGACTAAGATCATCCTGCCATTCTTCGACAAGTATTCTCTG
ATCACTGAAAAGCTGGGTGACTATCTGCTGTTCAAGAAGGTCCTGGAACTGATGGGTACTAAGGAACACCTGA
CTCAGAGAGGTCTGGAAAAGATCGTCTCTCTGAAGGCTTCTATCAACAAGGGTCTGTCTGAAGAACTGCAGGC
TGCTTTCCCACAGTGTGTCCCAACTCCAAGACCAGAAATCAACAACAAGAACATCCCAGACCCATTCTGGCTG
GCTGGTTTCGTCTCTGGTGACGGTTCTTTCAAGTCTATCCTGAAGAAGTCTGAATCTATCAAGGTCGGTTTCCAG
TCTATCCTGGTCTTCCAGATCACTCAGCACGCTAGAGACGTCAAGCTGATGGAATCTCTGATCTCTTATCTGGG
TTGTGGTTTCATCGAAAAGGACTCTAGAGGTCCATGGCTGTATTATACTGTCACTAACTTCTCTGACATCCAGG
GTAAGATCATCCCATTCTTCCACCAGTATAAGATCATCGGTTCTAAGTATGGTGACTATCAGGACTGGTGTAAG
ATCGCTCTGATCATGCAGAACAAGAACCACCTGACTCCAGAAGGTCTGAACGAAATCAGAGCTCTGAAGGGTG
GTATGAACAAGGGTAGACTG

2. Enzyme expression, purification, biochemical analyses, crystallization and X-ray data collection

Sequence verified plasmids were transformed into BL21(DE3) RIL *E.coli* cells and plated on LB-Amp plates to grow at 37°C. Colonies were grown in 10 mL overnight cultures of LB + Amp (100 μg/mL) and diluted 1:100 the next day to a final volume of 1 L. Cell cultures were shaken at 37°C until the cells reached an OD_{600} between 0.6-0.8. Cells were then incubated on ice for 30 minutes, induced with 200 mM IPTG, and incubated overnight at 16°C. Induced cells were pelleted and stored at -20°C. Successful protein induction was verified by SDS-PAGE.

Meganuclease purification. Cell pellets were resuspended in a buffer containing 25 mM Tris/HCl pH 7.5, 200 mM NaCl, and 5% glycerol. PMSF and benzonase (Sigma-Aldrich) were added to 1 mM concentrations prior to sonication. Cell debris was pelleted and the supernatant was filtered through a 0.20 μm filter. Untagged protein samples were loaded onto a 5 mL Heparin HP HiTrap column (GE Life Sciences) and eluted with a linear salt gradient (Buffer A: 25 mM Tris/HCl pH 7.5, 200 mM NaCl, 5% glycerol, Buffer B: 25 mM Tris/HCl pH 7.5, 1M NaCl). The meganuclease constructs eluted between 400 mM and 800 mM NaCl.

His-tagged proteins were incubated with pre-equilibrated Nickel-NTA resin (approximately 3 mL bed volume), gravity-loaded onto a column, and washed with buffer containing increasing concentrations of imidazole at pH 8.0. Fractions containing eluted protein were concentrated and exchanged to a thrombin cleavage buffer (0.3 M NaCl, 25 mM Tris pH 7.5, and 5% v/v glycerol). The His-tag was

removed with an overnight incubation at 4°C with biotinylated thrombin. The following morning, streptavidin-conjugated agarose resin was added to the sample and incubated for 30 minutes at room temperature. Thrombin was then removed by gravity filtration over an empty column, and pure, tagless protein was collected.

All untagged constructs were then concentrated to 5 - 20 mg/mL and passed over a size exclusion column (15 mL Superdex 200 10/300 GL, GE Life Sciences) in the presence of 25 mM Tris/HCl pH 7.5, 200 mM NaCl, and 5% glycerol.

Meganuclease thermal stability assays. Purified recombinant meganuclease constructs were diluted to between 10-20 µM concentration and dialyzed overnight into 10 mM potassium phosphate buffer at pH 8.0. Circular dichroism (CD) thermal denaturation experiments were performed on a JASCO J-815 CD spectrometer with a Peltier thermostat. Initial wavelength scans (190-250 nm) were carried out for each construct at 20°C and 80°C; the wavelength where the change in CD signal strength was greatest (210 nm) was used for the variable temperature scan. Thermal denaturation was monitored over a temperature range of 4°C to 95°C (0.1cm pathlength cell), with measurements taken every 2 degrees. Sample temperature was allowed to equilibrate for 30 seconds before each measurement. Thermal denaturation half-points (Tm values) were determined by curve-fitting using the SpectraManager software supplied by the instrument manufacturer.

DNA cleavage assays. To complement our flow cytometric DNA cleavage assay results, we also performed a *non-tethered* cleavage assay using either enzyme released from the surface of yeast or purified recombinant protein (as was used in our structural studies). For enzyme released from the surface of yeast, 5 million induced yeast cells were washed with OCB and mixed in a 45 µL volume with 10 mM DTT (breaks the disulfide bond linking the protein to the yeast cell surface), OCB, 20 nM A647-labeled DNA target substrate, and 5 mM $CaCl_2$ (no cleavage control), $MgCl_2$ (cleavage), or $MnCl_2$ (reduced specificity). For purified recombinant enzyme, the same reactions were set up with a final enzyme concentration of 10 nM and no DTT. The digest reactions were incubated at 37°C for 30 minutes. A colorless Ficoll loading dye was added to each reaction before loading samples onto an 11.5% acrylamide gel (120 volts for 1 hour). Cleavage products were visualized with the red laser (633nM) and a 670 BP 30 filter on a Typhoon Trio imaging system (GE Healthcare), and quantification of bands was performed using Image Studio (LICOR) software.

Meganuclease-DNA crystallization and data collection. Recombinant proteins were incubated with either $CaCl_2$ or $MgCl_2$ and a double-stranded DNA oligonucleotide (IDT) containing the enzyme's 22 bp target site with various lengths of flanking random sequence with either blunt ends, single base 3' overhangs or single base 5' overhangs. The mixtures were set up at a 1 : 1.15 protein:DNA ratio. Crystallization of the protein/DNA complexes was initially screened in 96-well trays using a mosquito robot (TPP Labtech) with three pre-made crystallization grids: PEGs Suite (Qiagen), Index I & II (Hampton Research), and Wizard Classic (Rigaku/Emerald BioStructures). Crystal hits from initial screens were further optimized in larger scale 24-well hanging drop trays. Detailed crystallization and cryoprotection conditions for each meganuclease are provided in the **Supplemental Experimental Procedures**.

Data was collected either on a Rigaku Micromax 007HF rotating anode generator with a RaxisIV++ or Saturn 944+ detector or at the Advanced Light Source (Beamline 5.0.1 or 5.0.2). Data was processed using HKL2000 (Otwinowski and Minor, 1997). Phases were obtained by molecular replacement with Phaser (McCoy et al., 2007) using either the I-OnuI structure (PDB ID 3QQY) or the I-LtrI structure (PDB ID 3R7P) as a search model. Model building and refinement were performed using Coot (Emsley et al., 2010) and Refmac (Murshudov et al., 1997), respectively.

I-AabMI crystallized in 25% (w/v) PEG 550 MME, 0.1M HEPES pH 7.5, and 20mM $CaCl_2$ with a 25bp duplex DNA + a single base 3' overhang:

Top: 5'-CAGGTACCCTTTAAACCTACTAACCC-3'
Bottom: 5'-GGTTAGTAGGTTTAAAGGGTACCTGG-3'

The crystal was cryoprotected in artificial mother liquor with 25% sucrose and flash frozen in liquid nitrogen. Data was collected with a Rigaku Micromax-007HF rotating anode generator on a Saturn944+ detector.

I-CpaMI crystallized in 26% (w/v) PEG 3350, 0.1M Tris/HCl pH 7.5, 5mM DTT, and 5mM $CaCl_2$ with a 27bp duplex DNA + a single base 3' overhang:

Top: 5'-CCTAGCCCACAATATTAAGGCCATCCCA-3'
Bottom: 5'-GGGATGGCCTTAATATTGTGGGCTAGGT-3'

The crystal was cryoprotected and then flash frozen in liquid nitrogen. Data was collected with a Rigaku Micromax-007HF rotating anode generator on a Saturn944+ detector.

I-GpeMI crystallized in 22.5% (w/v) PEG 3000, 0.1M Tris/HCl pH 8.5, and 50mM $CaCl_2$ with a 26bp duplex DNA + a single base 3' overhang:

Top: 5'-CCTTTCCGCTTATTCAACCCTTTACCC-3'
Bottom: 5'-GGTAAAGGGTTGAATAAGCGGAAAGGG-3'

The crystal was cryoprotected in artificial mother liquor with 15% ethylene glycol and flash frozen in liquid nitrogen. Data was collected at the Advanced Light Source (Beamline 5.0.1).

I-GzeII crystallized in 30% (w/v) PEG 6000, 0.1M HEPES pH 7.5, and 2mM $CaCl_2$ with a 26bp duplex DNA + a single base 3' overhang:

Top: 5'-CCTTTGTACCAATATGGTACCCATCCC-3'
Bottom: 5'-GGATGGGTACCATATTGGTACAAAGGG-3'

The crystal was cryoprotected in artificial mother liquor with 20% glycerol and flash frozen in liquid nitogren. Data was collected at the Advanced Light Source (Beamline 5.0.2).

I-LtrWI crystallized in 30% (w/v) PEG-MME 550, 0.1M Bis-Tris pH 6.5, and 50mM $CaCl_2$ with a 25bp duplex DNA + a single base 3' overhang:

Top: 5'-CAGTAGTGAAGTATGTTATTTAACCC-3'
Bottom: 5'-GGTTAAATAACATACTTCACTACTGG-3'

No cryoprotectant was used prior to freezing. Data was collected with a Rigaku Micromax-007HF rotating anode generator on a RaxisIV++ detector.

I-PanMI crystallized in 28% PEG 3350, 0.1M Tris pH 7.0 with a 26bp duplex DNA + a single base 5' overhang:

Top: 5'-CCCGCTCCTCATAATCCTTATCAAG-3'
Bottom: 5'-GGGCTTGAATAAGGATTATGAGGAGC-3'

The crystal was cryoprotected with 15% sucrose and then flash frozen in liquid nitrogen. Data was collected with a Rigaku Micromax-007HF rotating anode generator on a Saturn944+ detector.

I-SmaMI crystallized in 28 to 32% (w/v) PEG-MME 550, 0.1M HEPES pH 7.5 and 10 to 20 mM $CaCl_2$ with a 25bp blunt duplex DNA:

Top: 5'-CGTACACCTGATAATGGAGGATACC-3'
Bottom: 5'-GGTATCCTCCATTATCAGGTGTACG-3'

No cryoprotectant was used prior to freezing. Data was collected with a Rigaku Micromax-007HF rotating anode generator on a RaxisIV++ detector.

I-SmaMI with the “**TTCT**” one-off DNA substrate crystallized in 25% PEG 3350, 10mM Bis-Tris pH 6.5, and 5mM $CaCl_2$ with a 27bp duplex DNA + a single base 3’ overhang:

Top: 5’-CCTATCCTCCATTCTCAGGTGTACCACC-3’
Bottom: 5’-GTGGTACACCTGAGAATGGAGGATAGGG-3’

The crystal was cryoprotected in artificial mother liquor with 15% glycerol and flash frozen in liquid nitrogen. Data was collected at the Advanced Light Source (Beamline 5.0.2).

I-SmaMI with the “**TTGT**” one-off DNA substrate crystallized in 25% PEG 2000 MME, 10mM Tris pH 6.0, and 5mM $CaCl_2$ with a 25bp duplex DNA + a single base 3’ overhang:

Top: 5’-CTATCCTCCATTGTCAGGTGTACCCC-3’
Bottom: 5’-GGGTACACCTGACAATGGAGGATAGG-3’

The crystal was cryoprotected in artificial mother liquor with 15% ethylene glycol and flash frozen in liquid nitrogen. Data was collected with a Rigaku Micromax-007HF rotating anode generator on a RaxisIV++ detector.

3. Meganuclease Target Site Prediction (Example shown for I-GpeMI)

Run a BLAST search to find a meganuclease of interest in its natural host gene:

Amino acid sequence of GpeMI (trimmed to match the N- and C-termini of I-OnuI):

PTRNESINPWVLTGFADAEGSFILRIRNNNKSSAGYSTELGFQITLHKKDISILENIQSTWKVGVIANSGDNAVSLKVTRFEDLR
VVLNHFEKYPLITQKLGDYLLFKQAFSVMENKEHLKIEGIKRLVGIKANLNWGLTDELKEAFVASGGENIFVASGGERSLINK
NIPNSGWLAGFTSGEGCFFVSLIKSKSKLGVQVQLVFSITQHARDRALMDNLVTYLGCGYIKEKKKSEFSWLEFVVTKFSDIK
DKIIPVFQVNNIIGVKLEDFEDWCKVAKLIEEKKHLTESGLEEIRNIKLNMNKGRVL

Hit from BLAST search for I-GpeMI sequence:

Organism: Grosmannia penicillata (straing WIN(M)27)
Host gene: ribosomal protein 3
Meganuclease/homing endonuclease coding sequence: highlighted in **yellow**

```
>gi|257097851|gb|FJ607136.1| Grosmannia penicillata strain WIN(M)27 large subunit ribosomal RNA (rnl) gene, partial
sequence; and ribosomal protein 3/homing endonuclease-like fusion protein (rps3/HEG fusion) gene, complete cds;
mitochondrial
GGGTTAATTTCAAGAATTACTTATACTATAAGTAAATTTGAAGCGAAGTTTATTTATATTTAATAAATAATAATGTTCAACGACTAAGGGTGAGTTAATAACAATAATCCCTTATTAAT
ACCCAACACTTTTTTTTTGTAATTTAGTTTAATAATATAAAATCAATAAAATGCAAAAATATACCAAATTTTTAAGTAATATATTTGTCAAAAATATTAATAATAAATATAAATCTATA
CCATTGAATACTAGAATTAATTTTGTTGGTGAAACTAGATATTTCCCTTCTGATTTTAAAGAATGAACTAATAGTGTTTATTACTTTAATTCTAATAATATTAAAAATTTTCCCATTTA
TGACTTAAATGTAAGTAAATTATTAAAAGGTTATTTTGATTTATATTTTGATCGTGAAAATATAAAATTGAATTATAAATCTTATGAAAAAAAAGATTTTACGGCGCAAGCTCTAAATA
AAATCTTTGTAAGTAAACCTGAAATAAAACATACTAATTCTAAAGCCATAATAACTATTTATGTTTATAATAGAGAAAGAGTTATTTTGTTAAATAAATTAAATCAATTAAATATAGGA
ATTTTAGGACTAAATAAATTTTTTTATTTATGTAAAAAAATCTCTGGGGATTTATATAGTAAATATATTAAACAAGTTTTATATAAAGAATTCTTATTTTTAAGAAGATCTAAATTAAA
ACTTAATCTTAATGGATTAAAATTCCAAGATAAATTCTTATTTAAATTAAGTAAATTAATAAGTAAATTTTATAAAAAAAAAGTCGAATTTAATATTATTAATTTAAAATCTATAAATT
TGAATCCTAATATATTTACTGAAATAATGGCAAAAAAATTTATGAATAGAAATGCATCTATTATGCAAATAATGAAATTTATCTTAGATAAAAGTATTATTTTAAATGAGGATGGCGGC
ATAGTTTCACTAGCCTCATCCGGTGTAACCTCCTCCGGATCAGGATTAGAAAAAAGTAGAAAAATTAAAAATGTAAATTTAAATTTAATAGAAAATAAATATAAAAATCTGAATATTAA
CTCTATAGTTACTGATGGAGATATAAATGATAATATAAAAGAATTTTATACTAAAAATAGTGAAGATACTATTTTTGATTCTAAATATAAAAATTTAGGAGGTATAAGATTAGAAGCTA
AAGGAAGATTGACTAGACGTTATAGAGCGGATAGAGCAGTATCTAAAGTTAATATAAAAGGAGGATTGAAAAATATAGATTCATCTTACAAAGGTTTATCTTCTATTAATTATATAGGA
AAAATTAATTCAAGTATGGAATATTCAATGGATATATCAAAACGTCGTGTAGGTGCATTTGCTATAAAGGGTTGAATTTCAGGTAGATCTTACAGTACAACTGCCAATCCTACAAGAAA
TGAAAGTATAAATCCTTGAGTTCTTACTGGGTTTGCAGATGCTGAAGGTAGTTTTATACTAAGAATAAGAAATAATAATAAAAGTTCTGCAGGTTATTCTACAGAATTAGGATTTCAAA
TTACTTTGCATAAAAAAGACATATCTATCCTAGAAAATATTCAATCTACTTGAAAAGTAGGAGTTATAGCTAATAGCGGTGATAACGCCGTAAGTTTAAAAGTAACACGGTTTGAGGAT
TTAAGAGTAGTATTAAATCATTTTGAGAAATATCCTTTGATAACTCAAAAATTAGGTGATTATCTATTATTTAAACAAGCCTTTAGTGTTATGGAAAACAAAGAACATTTAAAAATTGA
AGGTATAAAAAGATTAGTTGGAATTAAGGCCAATTTAAATTGAGGTCTTACTGATGAACTAAAGGAGGCTTTTGTCGCCTCCGGCGGTGAAAATATATTTGTCGCCTCCGGCGGTGAAA
GATCTCTAATAAATAAAAAATATACCTAATTCTGGATGATTAGCTGGCTTTACTTCTGGTGAGGGTTGTTTTTTTGTTAGTTTAATAAAATCTAAATCTAAATTAGGGGTTCAGGTACAA
TTGGTATTTTCTATTACTCAACACGCGAGAGATAGAGCATTGATGGATAATTTAGTAACATATCTTGGATGTGGATATATTAAGGAAAAAAAGAAATCTGAGTTTTCATGATTAGAATT
TGTTGTTACAAAGTTTTCAGACATTAAGGATAAAATTATTCCAGTTTTTCAAGTAAATAATATAATAGGTGTAAAATTGGAAGACTTTGAAGATTGATGTAAAGTCGCTAAATTAATTG
AAGAAAAAAAACATTTAACTGAATCAGGCTTGGAAGAAATACGTAATATAAAATTAAACATGAATAAAGGAAGAGTTCTTTAAGTTAACTACATACTAAGTGCAGTATATAATTTCAAT
ATTTAAACTCTAGTATATTTTAATACTACGTTAGCAAGCTCCCATACTTCGTATGAGAATAAGCGGAAAATAATTTTTTTTAGGTTTTAATTACAATAAATAAAAAAAAGTGAAGAAAT
AGTCTGAACCATTTTGTGAAAAA
```

Run a BLAST search on the sequence prior to the meganuclease (the host gene in which the meganuclease resides):

```
>hostgene
GGGTTAATTTCAAGAATTACTTATACTATAAGTAAATTTGAAGCGAAGTTTATTTATATTTAATAAATAATAATGTTCAACGACTAAGGGTGAGTTAATAACAATAATCCCTTATTAAT
ACCCAACACTTTTTTTTTGTAATTTAGTTTAATAATATAAAATCAATAAAATGCAAAAATATACCAAATTTTTAAGTAATATATTTGTCAAAAATATTAATAATAAATATAAATCTATA
CCATTGAATACTAGAATTAATTTTGTTGGTGAAACTAGATATTTCCCTTCTGATTTTAAAGAATGAACTAATAGTGTTTATTACTTTAATTCTAATAATATTAAAAATTTTCCCATTTA
TGACTTAAATGTAAGTAAATTATTAAAAGGTTATTTTGATTTATATTTTGATCGTGAAAATATAAAATTGAATTATAAATCTTATGAAAAAAAAGATTTTACGGCGCAAGCTCTAAATA
AAATCTTTGTAAGTAAACCTGAAATAAAACATACTAATTCTAAAGCCATAATAACTATTTATGTTTATAATAGAGAAAGAGTTATTTTGTTAAATAAATTAAATCAATTAAATATAGGA
ATTTTAGGACTAAATAAATTTTTTTATTTATGTAAAAAAATCTCTGGGGATTTATATAGTAAATATATTAAACAAGTTTTATATAAAGAATTCTTATTTTTAAGAAGATCTAAATTAAA
```

```
ACTTAATCTTAATGGATTAAAATTCCAAGATAAATTCTTATTTAAATTAAGTAAATTAATAAGTAAATTTTATAAAAAAAAAGTCGAATTTAATATTATTAATTTAAAATCTATAAATT
TGAATCCTAATATATTTACTGAAATAATGGCAAAAAAATTTATGAATAGAAATGCATCTATTATGCAAATAATGAAATTTATCTTAGATAAAAGTATTATTTTAAATGAGGATGGCGGC
ATAGTTTCACTAGCCTCATCCGGTGTAACCTCCTCCGGATCAGGATTAGAAAAAAGTAGAAAAATTAAAAATGTAAATTTAAATTTAATAGAAAATAAATATAAAAATCTGAATATTAA
CTCTATAGTTACTGATGGAGATATAAATGATAATATAAAAGAATTTTATACTAAAAATAGTGAAGATACTATTTTTGATTCTAAATATAAAAATTTAGGAGGTATAAGATTAGAAGCTA
AAGGAAGATTGACTAGACGTTATAGAGCGGATAGAGCAGTATCTAAAGTTAATATAAAAGGAGGATTGAAAAATATAGATTCATCTTACAAAGGTTTATCTTCTATTAATTATATAGGA
AAAATTAATTCAAGTATGGAATATTCAATGGATATATCAAAACGTCGTGTAGGTGCATTTGCTATAAAGGGTTGAATTTCAGGTAGATCTTACAGTACAACTGCCAAT
```

In the list of results, look for a match of the same host gene (in this case, large subunit ribosomal RNA gene) with no mention of a homing endonuclease.

For this example, the second hit on the list of BLAST results is the SAME host gene in a different strain of Grosmannia penicillata. This is a perfect match to show us what the original host gene looked like before it was invaded by a meganuclease.

```
>gi|257097855|gb|FJ607138.1| Grosmannia penicillata strain WIN(M)136 large subunit ribosomal RNA (rnl) gene, partial sequence;
and ribosomal protein 3 (rps3) gene, complete cds; mitochondrial
CTCATATTTTTTTAATGATAACATATTGGGTTAATTTCAAGAATTACTTATACTATAAGTAAATTTGAAGCGAAGTTTATTTATATTTAATAAATAATAATGTTCAACGACTAAGGGTGAGTTAATAA
CAATAATCCCTTATTAATACCCAACACTTTTTTTTTGTAATTTAGTTTAATAATATAAAATCAATAAAATGCAAAAATATACCAAATTTTTAAGTAATATATTTGTCAAAAATATTAATAATAAATAT
AAATCTATACCATTGAATACTAGAATTAATTTTGTTGGTGAAACTAGATATTTCCCTTCTGATTTTAAAGAATGAACTAATAGTGTTTATTACTTTAATTCTAATAATATTAAAAATTTTCCCATTTA
TGACTTAAATGTAAGTAAATTATTAAAAGGTTATTTTGATTTATATTTTGATCGTGAAAATATAAAATTGAATTATAAATCTTATGAAAAAAAAGATTTTACGGCGCAAGCTCTAAATAAAATCTTTG
TAAGTAAACCTGAAATAAAACATACTAATTCTAAAGCCATAATAACTATTTATGTTTATAATAGAGAAAGAGTTATTTTGTTAAATAAATTAAATCAATTAAATATAGGAATTTTAGGACTAAATAAA
TTTTTTTATTTATGTAAAAAAATCTCTGGGGATTTATATAGTAAATATATTAAACAAGTTTTATATAAAGAATTCTTATTTTTAAGAAGATCTAAATTAAAACTTAATCTTAATGGATTAAAATTCCA
AGATAAATTCTTATTTAAATTAAGTAAATTAATAAGTAAATTTTATAAAAAAAAAGTCGAATTTAATATTATTAATTTAAAATCTATAAATTTGAATCCTAATATATTTACTGAAATAATGGCAAAAA
AATTTATGAATAGAAATGCATCTATTATGCAAATAATGAAATTTATCTTAGATAAAAGTATTATTTTAAATGAGGATGGCGGCATAGTTTCACTAGCCTCATCCGGTGTAACCTCCTCCGGATCAGGA
TTAGAAAAAAGTAGAAAAATTAAAAATGTAAATTTAAATTTAATAGAAAATAAATATAAAAATCTGAATATTAACTCTATAGTTACTGATGGAGATATAAATGATAATATAAAAGAATTTTATACTAA
AAATAGTGAAGATACTATTTTTGATTCTATTAAATATAAAAATTTAGGAGGTATAAGATTAGAAGCTAAAGGAAGATTGACTAGACGTTATAGAGCGGATAGAGCAGTATCTAAAGTTAATATAAAAG
GAGGATTGAAAAATATAGATTCATCTTACAAAGGTTTATCTTCTATTAATTATATAGGAAAAATTAATTCAAGTATGGAATATTCAATGGATATATCAAAACGTCGTGTAGGTGCATTTGCTATAAAG
GGTTGAATAAGCGGAAAATAATTTTTTTTAGGTTTTAATTACAATAAATAAAAAAAAGTGAAGAAATAGTCTGAACCATTTTGTGAAAAAT
```

Use Clustal Omega to align this new hit (with no meganuclease) to the original sequence with the GpeMI:

```
Gpe       ATCTAAAGTTAATATAAAAGGAGGATTGAAAAATATAGATTCATCTTACAAAGGTTTATC
hit       ATCTAAAGTTAATATAAAAGGAGGATTGAAAAATATAGATTCATCTTACAAAGGTTTATC
          ************************************************************

Gpe       TTCTATTAATTATATAGGAAAAATTAATTCAAGTATGGAATATTCAATGGATATATCAAA
hit       TTCTATTAATTATATAGGAAAAATTAATTCAAGTATGGAATATTCAATGGATATATCAAA
          ************************************************************

Gpe       ACGTCGTGTAGGTGCATTTGCTATAAAGGGTTGAATTTCAGGTAGATCTTACAGTACAAC
hit       ACGTCGTGTAGGTGCATTTGCTATAAAGGGTTGAATAAGCGGAAAATAATTTTTTTTAGG
          ************************************    ** * **   *     *  *

Gpe       TGCCAATCCTACAAGAAATGAAAGTATAAATCCTTGAGTTCTTACTGGGTTTGCAGATGC
hit       TTTTAATTACAATAAATAAAAAAAAGTGAAGAAATAGTCTGAAC--CATTTTGTGAAAAA
          *   ***    *  * * *  ***    * **     *     *          ****   *

Gpe       TGAAGGTAGTTTTATACTAAGAATAAGAAATAATAATAAAAGTTCTGCAGGTTATTCTAC
hit       T-----------------------------------------------------------
          *

Gpe       AGAATTAGGATTTCAAATTACTTTGCATAAAAAAGACATATCTATCCTAGAAAATATTCA
hit       ------------------------------------------------------------
```

The two sequences match perfectly up to a very distinct drop-off point. This may indicate where the inserted meganuclease gene begins.

Find the drop-off point in the original GpeMI fusion sequence (green highlight).

Then find the second half of the host gene sequence in the match with no homing endonuclease present (red highlight).

```
>gi|257097851|gb|FJ607136.1| Grosmannia penicillata strain WIN(M)27 large subunit ribosomal RNA (rnl) gene, partial
sequence; and ribosomal protein 3/homing endonuclease-like fusion protein (rps3/HEG fusion) gene, complete cds;
mitochondrial
GGGTTAATTTCAAGAATTACTTATACTATAAGTAAATTTGAAGCGAAGTTTATTTATATTTAATAAATAATAATGTTCAACGACTAAGGGTGAGTTAATAACAATAATCCCTTATTAAT
ACCCAACACTTTTTTTTTGTAATTTAGTTTAATAATATAAAATCAATAAAATGCAAAAATATACCAAATTTTTAAGTAATATATTTGTCAAAAATATTAATAATAAATATAAATCTATA
CCATTGAATACTAGAATTAATTTTGTTGGTGAAACTAGATATTTCCCTTCTGATTTTAAAGAATGAACTAATAGTGTTTATTACTTTAATTCTAATAATATTAAAAATTTTCCCATTTA
TGACTTAAATGTAAGTAAATTATTAAAAGGTTATTTTGATTTATATTTTGATCGTGAAAATATAAAATTGAATTATAAATCTTATGAAAAAAAAGATTTTACGGCGCAAGCTCTAAATA
AAATCTTTGTAAGTAAACCTGAAATAAAACATACTAATTCTAAAGCCATAATAACTATTTATGTTTATAATAGAGAAAGAGTTATTTTGTTAAATAAATTAAATCAATTAAATATAGGA
ATTTTAGGACTAAATAAATTTTTTTATTTATGTAAAAAAATCTCTGGGGATTTATATAGTAAATATATTAAACAAGTTTTATATAAAGAATTCTTATTTTTAAGAAGATCTAAATTAAA
ACTTAATCTTAATGGATTAAAATTCCAAGATAAATTCTTATTTAAATTAAGTAAATTAATAAGTAAATTTTATAAAAAAAAAGTCGAATTTAATATTATTAATTTAAAATCTATAAATT
TGAATCCTAATATATTTACTGAAATAATGGCAAAAAAATTTATGAATAGAAATGCATCTATTATGCAAATAATGAAATTTATCTTAGATAAAAGTATTATTTTAAATGAGGATGGCGGC
ATAGTTTCACTAGCCTCATCCGGTGTAACCTCCTCCGGATCAGGATTAGAAAAAAGTAGAAAAATTAAAAATGTAAATTTAAATTTAATAGAAAATAAATATAAAAATCTGAATATTAA
CTCTATAGTTACTGATGGAGATATAAATGATAATATAAAAGAATTTTATACTAAAAATAGTGAAGATACTATTTTTGATTCTAAATATAAAAATTTAGGAGGTATAAGATTAGAAGCTA
AAGGAAGATTGACTAGACGTTATAGAGCGGATAGAGCAGTATCTAAAGTTAATATAAAAGGAGGATTGAAAAATATAGATTCATCTTACAAAGGTTTATCTTCTATTAATTATATAGGA
AAAATTAATTCAAGTATGGAATATTCAATGGATATATCAAAACGTCGTGTAGGTGCATTTGCTATAAAGGGTTGAATTTCAGGTAGATCTTACAGTACAACTGCCAATCCTACAAGAAA
TGAAAGTATAAATCCTTGAGTTCTTACTGGGTTTGCAGATGCTGAAGGTAGTTTTATACTAAGAATAAGAAATAATAATAAAAGTTCTGCAGGTTATTCTACAGAATTAGGATTTCAAA
TTACTTTGCATAAAAAAGACATATCTATCCTAGAAAATATTCAATCTACTTGAAAAGTAGGAGTTATAGCTAATAGCGGTGATAACGCCGTAAGTTTAAAAGTAACACGGTTTGAGGAT
TTAAGAGTAGTATTAAATCATTTTGAGAAATATCCTTTGATAACTCAAAAATTAGGTGATTATCTATTATTTAAACAAGCCTTTAGTGTTATGGAAAACAAAGAACATTTAAAAATTGA
AGGTATAAAAAGATTAGTTGGAATTAAGGCCAATTTAAATTGAGGTCTTACTGATGAACTAAAGGAGGCTTTTGTCGCCTCCGGCGGTGAAAATATATTTGTCGCCTCCGGCGGTGAAA
GATCTCTAATAAATAAAAATATACCTAATTCTGGATGATTAGCTGGCTTTACTTCTGGTGAGGGTTGTTTTTTTGTTAGTTTAATAAAATCTAAATCTAAATTAGGGGTTCAGGTACAA
TTGGTATTTTCTATTACTCAACACGCGAGAGATAGAGCATTGATGGATAATTTAGTAACATATCTTGGATGTGGATATATTAAGGAAAAAAAGAAATCTGAGTTTTCATGATTAGAATT
```

TGTTGTTACAAAGTTTTCAGACATTAAGGATAAAATTATTCCAGTTTTTCAAGTAAATAATATAATAGGTGTAAAATTGGAAGACTTTGAAGATTGATGTAAAGTCGCTAAATTAATTG
AAGAAAAAAAAACATTTAACTGAATCAGGCTTGGAAGAAATACGTAATATAAAATTAAACATGAATAAAGGAAGAGTTCTTTAAGTTAACTACATACTAAGTGCAGTATATAATTTCAAT
ATTTAAACTCTAGTATATTTTAATACTACGTTAGCAAGCTCCCATACTTCGTATGAGAATAAGCGGAAAATAATTTTTTTTAGGTTTTAATTACAATAAATAAAAAAAAAGTGAAGAAAT
AGTCTGAACCATTTTGTGAAAAA

The gray highlighting is extra sequence carried with the meganuclease when it invades the host gene.

Remove the gray and yellow-highlighted sequences to form the predicted target site of the GpeMI enzyme.

Predicted target sequence for GpeMI TGCTATAAAGGGTTGAATAAGCGGAAAATAATTTTT

The flow-cytometric cleavage assay can then be used to test for cleavage activity against this predicted target sequence.

4. Generation of DNA Substrates for Assays of Binding and Cleavage

Primer #1 with conjugated Biotin:

5’-/5Biosg/TCAGCACAGCACTACG-3’

Primer #2 with conjugated A647 fluorophore:

5’-/5Alex647N/TGGACACGACTTGAGC-3’

Single-stranded template with 22bp target site:

5’-TGGACACGACTTGAGCTTTCCACTTATTCAACCTTTTACGTAGTGCTGTGCTGA-3’

References

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Cryst D *66*, 486 - 501.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. . J. Appl. Crystal. *40*, 658 - 674.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by teh maximum-likelihodd method. Acta Cryst D *53*, 240 - 255.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. In Methods in Enzymology, J. C. W. Carter, and R.M. Sweet, eds. (Academic Press), pp. 307 - 326.