
Electronic Thesis and Dissertation Repository

4-18-2013 12:00 AM

DNA hydrolysis and genome editing applications of GIY-YIG family homing endonucleases

Benjamin P. Kleinstiver
The University of Western Ontario

Supervisor
David Edgell
The University of Western Ontario

Graduate Program in Biochemistry
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of
Philosophy
© Benjamin P. Kleinstiver 2013

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>

 Part of the [Biochemistry Commons](#), [Biotechnology Commons](#), and the [Molecular Biology Commons](#)

Recommended Citation

Kleinstiver, Benjamin P., "DNA hydrolysis and genome editing applications of GIY-YIG family homing endonucleases" (2013). *Electronic Thesis and Dissertation Repository*. 1191.
<https://ir.lib.uwo.ca/etd/1191>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.

DNA HYDROLYSIS AND GENOME EDITING APPLICATIONS OF GIY-YIG FAMILY
HOMING ENDONUCLEASES

(Thesis format: Integrated Article)

by

Benjamin P. Kleinstiver

Graduate Program in Biochemistry

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

© Benjamin P. Kleinstiver 2013

ABSTRACT

The ability to manipulate complex genomes in a precise manner is essential for studying biological processes in model systems, engineering plant strains for agriculture, or advancing human cellular therapies to treat diseases. Genomic alterations are most efficient when a double-strand DNA break is introduced at the loci where the modification is desired. Different classes of naturally occurring DNA endonucleases, including homing endonucleases, have therefore been explored as candidates for genome modification studies as they target long stretches of DNA. Homing endonucleases are mobile genetic elements whose biological role is to introduce site-specific double-strand breaks into naïve genomes, ultimately resulting in the selfish propagation of their own genes. Consequently, homing endonucleases are an ideal enzymatic system whose natural properties can be exploited to manipulate genes.

In the present studies, I examine the cleavage mechanism of GIY-YIG family homing endonucleases, as until now the method by which they hydrolyze DNA has remained poorly understood. Using the GIY-YIG homing endonuclease I-BmoI as a model system, I investigate the amino acid, nucleotide, and divalent metal ion requirements of the GIY-YIG nuclease domain to generate a double-strand break in DNA. I specifically test models by which enzymes with a single active site could nick both strands of DNA, and determine that I-BmoI functions as a monomer throughout the reaction pathway. Furthermore, I demonstrate that the nuclease domain itself has weak binding affinity, is tethered to DNA by a high affinity binding domain, and must reposition across each strand through a series of protein and substrate conformational changes to facilitate DNA hydrolysis.

To explore the relevance of GIY-YIG homing endonucleases as genome editing reagents, I fused the nuclease domain of I-TevI to three different re-targetable DNA-binding platforms utilized in the field. The engineered nucleases developed within the present studies are mechanistically distinct from established technologies, as they function as monomers and cleave DNA at a preferred sequence motif. I therefore envision that the engineered GIY-YIG nucleases may circumvent complications associated with established technologies, and provide an alternative and potentially safer set of genome editing reagents.

KEYWORDS

GIY-YIG homing endonucleases, I-BmoI, I-TevI, monomer, double-strand break, DNA-hydrolysis mechanism, genome editing, engineered nucleases, zinc-finger, transcription activator-like (TAL) effector

CO-AUTHORSHIP

For the peer-reviewed publications presented in Chapters 2, 3, 4, and 5, Ben Kleinstiver performed the research, with exceptions noted below. Ben Kleinstiver and David Edgell conceived and designed the experiments, analyzed the data, and wrote the manuscripts.

Chapter 2 - Andrew Fernandes developed the EoS metric, Greg Gloor generated the structure-based alignment and performed co-variation analyses.

Chapter 3 - Wesley Bérubé-Janzen helped perform plasmid cleavage assays on mutant substrates, Andrew Fernandes calculated rate constants.

Chapter 4 - Jason Wolfs helped perform gel-filtration studies, two-site plasmid cleavage assays, *in vivo* substrate activity assays, OP-Cu footprinting reactions, and protease mapping experiments on mutant substrates.

Chapter 5 - Jason Wolfs helped design and test I-TevI/I-OnuI chimeras, Alanna Roberts, Sherry Hu, and Tomasz Kolaczyk helped perform endless *in vivo* bacterial activity assays. I-TevI/PthXo1 chimeras were designed in collaboration with Dr. Adam Bogdanove and his research technician Li Wang.

DEDICATION

This thesis is dedicated to my mentors and colleagues who have guided me down the rabbit hole of science.

“... science is whatever we want it to be.”

- *Dr. Leo Spaceman*

ACKNOWLEDGMENTS

The primary acknowledgment for this thesis is owed to my supervisor, Dr. Dave Edgell, for his guidance and advice over the course of my Doctoral work. In Dave's lab, one learns quickly to always have the proper scientific controls and to never forget *anything*. I am sincerely grateful for the scientific training and intellectual freedom that Dave has provided me with over the past 5 years.

I would also like to acknowledge my advisory committee, Dr. David Haniford and the late Dr. Megan Davey, for their helpful discussions and technical insight throughout my graduate research. Megan's unwavering dedication to her research throughout her battle with cancer was truly inspirational.

To the past and present members of the Edgell lab, you have helped me remain motivated to come into the lab early or stay late at night, and 'forced' me to enjoy many beers on the Grad Club patio. Particular acknowledgements to Dr. Ewan Gibb for his mentorship and inexhaustible cynicism, Jason Wolfs for wrapping up experiments from the endless list that I'll never complete, Tom McMurrough for his sometimes productive diversions, and Stephanie Dorman for achieving an honorary Edgell Lab membership.

I would also like to extend acknowledgements to my friends and family. Without your love, support, and endless distractions, I may not have survived my PhD. Thank you also to Laura for allowing me to commandeer her apartment while I wrote my thesis.

Thank you to everyone for contributing to some of my happiest and most rewarding years; I've had endless amounts of fun constantly discussing science and life with you.

TABLE OF CONTENTS

ABSTRACT	ii
KEYWORDS	iii
CO-AUTHORSHIP	iv
DEDICATION	v
ACKNOWLEDGMENTS	vi
TABLE OF CONTENTS	vii
LIST OF TABLES	xii
LIST OF FIGURES	xiii
LIST OF APPENDICES	xv
LIST OF ABBREVIATIONS AND SYMBOLS.....	xvi
CHAPTER 1: INTRODUCTION	1
1.1 DNA endonucleases	2
1.1.1 Restriction defense systems	2
1.1.2 Mobile genetic elements	3
1.1.3 Strategies to hydrolyze DNA	5
1.2 Homing and Type IIS restriction endonucleases	8
1.2.1 GIY-YIG homing endonucleases.....	10
1.2.2 Properties of other GIY-YIG family endonucleases.....	12
1.2.3 LAGLIDADG homing endonucleases.....	15
1.2.4 $\beta\beta\alpha$ -Me homing endonucleases	16
1.2.5 Type IIS restriction endonucleases	17
1.3 Genome editing.....	18
1.4 Genome editing with engineered nucleases.....	19
1.4.1 Zinc-finger nucleases	21
1.4.2 Transcription activator-like effector nucleases.....	21

1.4.3	Engineered LAGLIDADG nucleases	23
1.5	Scope of the thesis: can the cleavage mechanism of the GIY-YIG domain detoxify engineered nucleases?.....	25
1.6	References	27
CHAPTER 2: A UNIFIED GENETIC, COMPUTATIONAL AND EXPERIMENTAL FRAMEWORK IDENTIFIES FUNCTIONALLY RELEVANT RESIDUES OF I-BMOI		
50		
2.1	Introduction	50
2.2	Materials and methods.....	53
2.2.1	Strain and plasmid construction.....	53
2.2.2	Genetic selection.....	54
2.2.3	Construction of mutagenized I-BmoI libraries	55
2.2.4	I-BmoI unigenic evolution and selection of variants.....	55
2.2.5	Construction and purification of site-directed mutants.....	56
2.2.6	Characterization of I-BmoI variants	56
2.2.7	Structure based alignment.....	57
2.2.8	Co-variation analyses.....	58
2.3	Results	58
2.3.1	Improved alignment of the GIY-YIG domain	58
2.3.2	Identification of co-evolving positions in the I-BmoI catalytic domain... 60	
2.3.3	Unigenic evolution identifies mutable positions in I-BmoI.....	61
2.3.4	Statistical analysis of unigenic evolution data reveals residues of potential functional importance	68
2.3.5	Genetic analysis of site-directed mutants	70
2.3.6	Cleavage assays with key site-directed mutants	74
2.4	Discussion.....	76
2.4.1	Application of MUSE to GIY-YIG homing endonucleases	76

2.4.2	Residues within the GIY-YIG domain of I-BmoI that are important for function.....	79
2.4.3	S20 and I71	79
2.4.4	H31	82
2.4.5	N42.....	83
2.4.6	I67	83
2.4.7	Residues outside of the catalytic domain predicted to be critical for function.....	83
2.4.8	Conclusion	85
2.5	References	86

CHAPTER 3: DIVALENT METAL ION DIFFERENTIALLY REGULATES THE SEQUENTIAL NICKING REACTIONS OF THE GIY-YIG HOMING ENDONUCLEASE I-BMOI 91

3.1	Introduction	91
3.2	Materials and methods.....	94
3.2.1	Bacterial strains and plasmids.....	94
3.2.2	Oligonucleotides	94
3.2.3	Cleavage assays on plasmid substrates	94
3.2.4	OP-Cu in-gel footprinting.....	96
3.2.5	Molecular modeling.....	97
3.3	Results	97
3.3.1	Magnesium is the preferred divalent metal ion for efficient and specific cleavage.....	97
3.3.2	Limiting divalent metal ion has a more pronounced regulation of second strand nicking.....	99
3.3.3	Assays with mutant substrates reveal three distinct cleavage phenotypes	100
3.3.4	In-gel footprinting reveals multiple minor groove distortions dependent on GC-2	106
3.3.5	Modeling of an I-BmoI-substrate complex.....	107

3.4	Discussion.....	109
3.5	References	113

CHAPTER 4: THE MONOMERIC GIY-YIG HOMING ENDONUCLEASE I-BMOI USES A MOLECULAR ANCHOR AND A FLEXIBLE TETHER TO SEQUENTIALLY NICK DNA 119

4.1	Introduction	119
4.2	Materials and methods.....	122
4.2.1	Strain and plasmid construction.....	122
4.2.2	Identification of I-BmoI domains	123
4.2.3	Biophysical characterization of I-BmoI.....	124
4.2.4	DNA-binding affinity of I-BmoI domains.....	124
4.2.5	Gel filtration chromatography.....	124
4.2.6	Cleavage assays with plasmid substrate	125
4.2.7	Gel-mobility shift assays with pre-nicked substrates.....	125
4.2.8	<i>thyA</i> cleavage site and spacer selection	126
4.2.9	<i>in vivo</i> survival assays.....	127
4.2.10	OP-Cu in-gel footprinting.....	127
4.3	Results	128
4.3.1	I-BmoI is partially disordered in the absence of substrate.....	128
4.3.2	N-terminal domains have weak but specific DNA-binding affinity.....	130
4.3.3	I-BmoI binds DNA as a monomer	132
4.3.4	The oligomeric status of I-BmoI does not change during cleavage.....	134
4.3.5	I-BmoI does not function via a substrate hairpin mechanism.....	136
4.3.6	The I-BmoI target site is modular.....	137
4.3.7	Contacts to G+7 facilitate cleavage by I-BmoI	142
4.4	Discussion.....	146
4.5	References	151

CHAPTER 5: MONOMERIC SITE-SPECIFIC NUCLEASES FOR GENOME EDITING	157
5.1 Introduction	157
5.2 Materials and methods.....	159
5.3 Results	159
5.3.1 Construction and validation of GIY-zinc finger endonucleases.....	159
5.3.2 Tev-ZFEs function as monomers to cleave at a specific sequence.....	162
5.3.3 Tev-ZFEs function in a yeast-based recombination assay.....	164
5.3.4 The I-TevI nuclease domain is portable to the LAGLIDADG architecture.....	166
5.3.5 I-TevI fusions to the TAL effector PthXo1 are functional	169
5.3.6 Tev-TALs tolerate nucleotide substitutions in the DNA spacer.....	169
5.3.7 A 5'-CNNNG-3' cleavage motif is not limiting for targeting	171
5.4 Discussion.....	171
5.5 References	175
CHAPTER 6: DISCUSSION	179
6.1 GIY-HE cleavage mechanism	179
6.2 Considerations for engineered nucleases.....	182
6.3 Implications of GIY-HEs for GIY-ENs.....	187
6.4 Potential of GIY-YIG engineered nucleases	188
6.5 References	191
APPENDICES	196
CURRICULUM VITAE	251

LIST OF TABLES

Table 1.1: Properties of site-specific nucleases	4
Table 2.1: Co-evolving residues in the I-BmoI GIY-YIG domain.....	62
Table 2.2 Survival ratios of I-BmoI variants with mutations in the catalytic domain.....	72
Table 2.3: DNA cleavage by WT I-BmoI and variants identified by MUSE.....	77
Table 3.1: Rate constants for first- and second-strand nicking reactions in different MgCl ₂ concentrations.....	102
Table 3.2: Summary of cleavage data for intronless and mutant substrates.....	103
Table 4.1: Rate constants for first- and second-strand nicking reactions on G+7 mutant substrates	144

LIST OF FIGURES

Figure 1.1: Mobility pathway of selfish intron-encoded endonucleases	6
Figure 1.2: Mechanisms of DNA hydrolysis by site-specific endonucleases.....	7
Figure 1.3: Structures of homing and restriction endonucleases	9
Figure 1.4: GIY-YIG homing endonucleases I-TevI and I-BmoI	11
Figure 1.5: GIY-YIG domain structures.....	14
Figure 1.6: Genome editing outcomes from targeted DNA breaks	20
Figure 1.7: FokI nuclease derived ZFNs and TALENs	22
Figure 1.8: Methods to engineer LAGLIDADG endonuclease specificity	24
Figure 2.1: I-BmoI is a modular GIY-YIG homing endonuclease	52
Figure 2.2: Alignment of the GIY-YIG domain	59
Figure 2.3: Co-evolving residues of the I-BmoI catalytic domain	63
Figure 2.4: The two-plasmid genetic selection	64
Figure 2.5: Unigenic evolution analysis of I-BmoI	66
Figure 2.6: Mapping positions that tolerate non-synonymous substitutions and EoS data onto the I-BmoI homology model	67
Figure 2.7: Cleavage activity of WT I-BmoI and site-directed mutants.....	73
Figure 2.8: Nicking assays with WT and mutant proteins.....	75
Figure 2.9: Summary of functionally relevant residues identified by MUSE	81
Figure 3.1: Model of I-BmoI interactions with intronless and intron-containing substrates based on DNA footprinting experiments	93
Figure 3.2: Magnesium is the preferred divalent metal ion for efficient and specific cleavage by I-BmoI.....	98
Figure 3.3: Limiting divalent metal ion has a greater effect on second strand than first strand nicking	101
Figure 3.4: Magnesium concentrations reveal three distinct classes of substitutions.....	104
Figure 3.5: The GC-2 base pair is required for minor-groove distortions near the bottom-strand nick site.....	108
Figure 3.6: Model of I-BmoI GIY-YIG domain interactions with substrate	110

Figure 4.1: I-BmoI interactions with substrate and monomeric cleavage models.....	121
Figure 4.2: Identification of stable I-BmoI domains	129
Figure 4.3: I-BmoI domains have distinct affinities for DNA.....	131
Figure 4.4: I-BmoI is a monomer in solution and in complex with thyA substrate	133
Figure 4.5: I-BmoI functions as a monomer.....	135
Figure 4.6: I-BmoI does not function via an intra-strand hairpin mechanism.....	138
Figure 4.7: Insertions in the I-BmoI substrate spacer reduce or abolish cleavage	139
Figure 4.8: An <i>in vivo</i> screen identifies I-BmoI nucleotide preference at G+7.....	141
Figure 4.9: Mutations at G+7 affect I-BmoI activity.....	143
Figure 4.10: The domains of I-BmoI form distinct DNA-contacts to distort DNA prior to cleavage.....	147
Figure 5.1: Design and functionality of Tev-ZFEs.....	160
Figure 5.2: TevN201-ZFE is a monomer with a preferred cleavage site.....	163
Figure 5.3: Tev-ZFEs can induce recombination in an eukaryotic system.....	165
Figure 5.4: Design and functionality of Tev-LHEs	167
Figure 5.5: Design and functionality of Tev-TALs	170
Figure 5.6: Cleavage requirements do not limit GIY-EN applicability.....	172
Figure 6.1: Conformational change DSB mechanism for I-BmoI.....	181
Figure 6.2: FokI requirements can lead to off-target DSBs.....	183
Figure 6.3: Design of engineered GIY-YIG endonucleases	186
Figure 6.4: Considerations for engineered GIY-YIG nucleases.....	189

LIST OF APPENDICES

Appendix S1: Copyright permissions for Chapters 2-5.....	196
S1.1 Permission for Chapter 2	196
S1.2 Permission for Chapter 3	197
S1.3 Permission for Chapter 4	198
S1.4 Permission for Chapter 5	199
Appendix S2: Supplementary information for Chapter 2.....	200
S2 Supplementary materials and methods	200
S2 Supplementary figures	201
S2 Supplementary tables.....	203
Appendix S3: Supplementary information for Chapter 3.....	209
S3 Supplementary tables.....	209
Appendix S4: Supplementary information for Chapter 4.....	212
S4 Supplementary materials and methods	212
S4 Supplementary figures	214
S4 Supplementary tables.....	223
S4 Supplementary references.....	230
Appendix S5: Supplementary information for Chapter 5.....	231
S5 Supplementary materials and methods	231
S5 Supplementary figures	235
S5 Supplementary tables.....	239
S5 Supplementary references.....	250

LIST OF ABBREVIATIONS AND SYMBOLS

ATP	adenosine triphosphate
BLAST	basic local alignment search tool
Bmo-ZFE	I-BmoI/zinc-finger fusion endonuclease
bp	base pair
Cas	CRISPR-associated
CRISPR	clustered regularly interspaced short palindromic repeats
DNA	deoxyribonucleic acid
DSB	double-strand break
DSC	differential scanning calorimetry
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EoS	evidence of selection
GIY-EN	GIY-YIG engineered nuclease
GIY-HE	GIY-YIG homing endonuclease
GIY-LHE	GIY-YIG/LAGLIDADG fusion endonuclease
GIY-RE	GIY-YIG restriction endonuclease
GIY-TAL	GIY-YIG/TAL effector fusion endonuclease
GIY-ZFE	GIY-YIG/zinc-finger fusion endonuclease
HE	homing endonuclease
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HR	homologous recombination
H-T-H (or HTH)	helix-turn-helix
IPTG	isopropyl β -D-1-thiogalactopyranoside
K_D	dissociation constant
$k_{\#}$	rate constant for reaction #
kDa	kilodalton
LB	Luria-Bertani
LC	lower complex
LC-MS	liquid chromatography-mass spectrometry
LEM	lamina-associated polypeptide, emerlin, MAN1 domain
LHE	LAGLIDADG homing endonuclease
LTR	long terminal repeat
MALDI-TOF	matrix-assisted laser desorption/ionization time-of-flight
MgCl ₂	magnesium chloride
MGE	mobile genetic element
min	minute
MUSE	<u>m</u> utual information, <u>u</u> nigenic evolution, and <u>s</u> tructure guided <u>e</u> lucidation
NHEJ	nonhomologous end joining

NMR	nuclear magnetic resonance
nt	nucleotide
NUMOD	nuclease associated modular DNA-binding α -helix
ONPG	<i>ortho</i> -nitrophenyl- β -galactoside
OP-Cu	1,10-copper phenanthroline
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB ID	protein data bank identifier
pK_a	acid dissociation constant
PSSM	position-specific scoring matrix
RE	restriction endonuclease
RM	restriction/modification
RNA	ribonucleic acid
rRNA	ribosomal RNA
RVD	repeat-variable diresidue
s	second
SDS	sodium dodecyl sulfate
SNP	single-nucleotide polymorphism
SOC	super optimal broth with catabolite repression
TAE	tris base, acetic acid, and EDTA buffer
TAL	transcription activator-like
TALEN	TAL effector nuclease
<i>td</i>	thymidylate synthase allele of bacteriophage T4
Tev-LHE	I-TevI/LAGLIDADG fusion endonuclease
Tev-TAL	I-TevI/TAL effector fusion endonuclease
Tev-ZFE	I-TevI/zinc-finger fusion endonuclease
<i>thyA</i>	thymidylate synthase allele of <i>Bacillus mojavensis</i>
T_m	melting temperature
tris-HCl	tris(hydroxymethyl)aminomethane hydrochloride
UC	upper complex
UNB	unbound
WT	wild-type
YPD	yeast extract peptone dextrose
ZFN	zinc-finger nuclease
ω	omega factor

“I am against religion because it teaches us to be
satisfied with not understanding the world.”

- *Richard Dawkins*

Chapter 1

1 Introduction

Enzymes that hydrolyze nucleic acids, termed nucleases, are ubiquitous and indispensable for life. Nucleases are critical components in a variety of cellular processes, including nucleotide metabolism and scavenging, DNA repair and recombination, restriction defense systems that protect host genomes, and the mobility of genetic elements (1-3). By lowering the transition state energy required for phosphodiester bond hydrolysis, nucleases act as catalysts to hydrolyze DNA or RNA polynucleotide chains. The focus of this thesis is to characterize the mechanism by which one class of nuclease hydrolyzes DNA to initiate its own genetic propagation between distinct biological alleles. Additionally, the derived molecular understanding provided the necessary insight to engineer this nuclease family for targeted genome editing applications.

Using the GIY-YIG homing endonuclease (GIY-HE) I-BmoI as a model system, I sought to investigate the mechanism by which GIY-HEs generate double-strand breaks (DSBs) in DNA. To better understand the structure and function of I-BmoI, we developed an experimental approach in Chapter 2 that combines bioinformatic, genetic, and structural techniques to identify non-conserved residues within an enzyme that affect catalysis. In Chapter 3, I present work that investigates the divalent metal ion and base pair requirements of the I-BmoI nuclease domain to efficiently cleave its cognate target site. After gaining insight into the biology and function of the nuclease domain, I specifically test models for cleavage by I-BmoI to elucidate the mechanism of hydrolysis in Chapter 4. Finally, in Chapter 5 I present data that demonstrates the GIY-YIG nuclease domain can be fused to the three common DNA-binding platforms used in the field of genome editing to create highly active and targetable chimeric nucleases. By understanding the cleavage mechanism of GIY-HEs prior to designing the chimeric nucleases, we hope to avoid engineering complications related to other genome editing technologies due to the inherent biology of the associated nuclease domain (4).

1.1 DNA endonucleases

Nuclease nomenclature differs depending on where hydrolysis along the DNA strand occurs, what type of nucleic acid is being modified, and whether or not both strands are hydrolyzed. Endonucleases cleave within polynucleotide chains, exonucleases catalyze the removal of nucleotides from either the 5' or 3' end of nucleic acids, and nickases and cleavases hydrolyze one or both strands, respectively. Additionally, nucleases can hydrolyze polynucleotides chains in non-specific, structure-specific, or sequence-specific manners. Non-specific nucleases, such as the *Serratia* and *Anabaena* nucleases, act as extracellular nucleotide scavengers for their prokaryotic hosts (5,6). The *Serratia* nuclease has a single active site and hydrolyzes single- and double-stranded DNA and RNA unless the nucleic acid substrate contains certain polynucleotide tracts (7-9). Structure-specific nucleases typically recognize abnormal DNA conformations and are common within DNA-repair and recombination pathways. AP endonucleases nick adjacent to sites of mismatched or damaged bases to initiate base-excision repair (10-12), and T7 endonuclease I recognizes branched DNA to resolve holiday junctions (13-15). Finally, examples of sequence-specific (or site-specific) nucleases include endonucleases that are involved in host restriction/modification (RM) defense systems (16-18), or enzymes that initiate the mobility of selfish genetic elements (19,20).

1.1.1 Restriction defense systems

First discovered in the mid 1950's (21-23), RM systems have revolutionized molecular biology as purified recombinant restriction endonucleases (REs) are indispensable for many scientific disciplines (18,24). REs constitute half of the RM systems employed by prokaryotes to prevent invasion by foreign DNA (18,25). In addition to REs, RM systems also encode DNA methyltransferases that utilize S-adenosyl-methionine as a co-factor to methylate adenine or cytosine residues, protecting host DNA from self-degradation (26,27). Both endonuclease and methyltransferase components of the RM system bind the same 3-8 base-pair sequence, with some organisms encoding upwards of 20 RM systems that target different sites (28). Since the initial characterization of type II REs (29,30), thousands of distinct REs have been identified that bind and cleave unique sites

by a vast number of mechanisms, necessitating a unified naming convention with distinct subtypes (24,31-33). Type II REs act independently of their associated methyltransferases and can be further divided into over 14 subtypes, including Type IIP and IIS classes to specify enzymes that recognize palindromic sequences or those that cleave outside of their primary recognition site, respectively (24,32). The Type IIP subclass constitutes the majority of REs utilized for biochemistry, while the nuclease domain of the Type IIS RE FokI has been used extensively for genome editing applications due to the challenges associated with reengineering of the natural binding specificities of Type IIP REs (see also Chapter 1.2.5) (34-37). Detailed studies of RM systems from each subclass have provided insight into the distinct properties and molecular mechanisms by which REs hydrolyze DNA (Table 1.1) (38). Interestingly, the conserved motifs that constitute RE active sites have also been identified in other classes of site-specific DNA endonucleases that act as selfish mobile genetic elements (39-42), with REs themselves also implicated to play a role in genome evolution (43,44).

1.1.2 Mobile genetic elements

On an evolutionary timescale, the status of most genomes can be considered fluid and under selective pressure, a result of normal or aberrant cellular processes, spontaneous mutations, or the presence of mobile genetic elements (MGEs) (45-47). Initially thought to be mechanistically selfish and phenotypically neutral with respect to host biology (48,49), certain classes of MGEs can contribute to essential cellular functions including mating type switching in *Saccharomyces cerevisiae* or the maintenance of proper transcript splicing in bacteriophage T4 (50,51). MGEs often encode relatively small genes that act to initiate ‘self-mobility’ between host and recipient alleles (49,52), yet the cumulative presence of multiple MGEs is thought to comprise a significant proportion of many genomes (53-56). During the mobility process, MGEs have been shown to influence recipient genome dynamics by inducing genomic rearrangements (57-59), promoting the spread of antibiotic resistance genes (60-62), or fragmenting formerly intact genes (63-65).

Table 1.1: Properties of site-specific nucleases

Endonuclease family	Oligomeric status	Active site	Divalent metal	DNA binding	Cleavage mechanism
Type IIP RE	Monomeric, dimeric, or tetrameric assembly on DNA	PD(E/D)xK (most common), $\beta\beta\alpha$ -Me, or GIY-YIG motifs in individual enzyme units	1, 2 or 3 divalent metal ions for catalysis; metal identity varies by catalytic family	3-8 bp symmetric binding site, cleave within binding site	Sequential or simultaneous nicking via multiple active sites
Type IIS RE	Bind DNA as monomers, can dimerize for cleavage	Single active site in catalytic domain, can form a composite site at dimer interface	1 or 2 divalent metal ions for catalysis; metal identity varies	4-7 bp asymmetric binding site, cleave 0-21 bp away from site	Sequential cleavage or transient dimerization through catalytic domains
Phospholipase D RE	Monomeric (2-domain) or homodimeric assembly on DNA	HxK catalytic motif in each half-unit at dimer interface to form a composite active site	Metal-independent for catalysis	6 bp asymmetric binding site, cleave 4-7 bp away from site	Transesterification via covalent protein-DNA intermediate
H-N-H ($\beta\beta\alpha$ -Me) HE	Bind DNA as monomers or dimers	Single H-N-H active site in the catalytic domain	Cleavage with Mg^{2+} ; nicking with Ni^{2+} ; Zn^{2+} for protein stability	24-35 bp binding site, nick or cleave asymmetrically within site	Nickase, or stable dimerization for DSB
His-Cys ($\beta\beta\alpha$ -Me) HE	Dimeric assembly on DNA	Composite His-Cys active sites at dimer interface	Cleavage with Mg^{2+} ; Zn^{2+} for protein stability	14-20 bp binding site, cleave within site	Sequential or simultaneous nicking at dimer interface
LAGLIDADG HE	Single-chain monomeric or homodimeric assembly on DNA	LAGLIDADG motif in each half-unit to form a composite active site	1, 2 or 3 divalent metal ions (can be shared at interface), Mg^{2+} optimal for catalysis	16-26 bp pseudo-symmetric or asymmetric binding site, cleave within binding site	Sequential nicking via two active sites from either half-unit
GIY-YIG HE	Bind DNA as monomers	Single GIY-YIG active site in the catalytic domain	Cleavage with single Mg^{2+} ; Mn^{2+} leads to promiscuous cleavage	30-36 bp asymmetric binding site, cleave adjacent to primary binding sequence	Unknown; conformational change proposed

Homing endonucleases (HEs) are one class of MGE that are often found encoded within introns, an arrangement that has been suggested as a mechanism to reduce the impact of the insertion of MGEs on host gene function (66-68). Within these introns, genes of HEs (or other classes of DNA endonucleases) can be found that initiate the mobility process (67,69-71) (Figure 1.1). The first step of intron propagation occurs when the intron-encoded endonuclease is translated and generates a double-strand break within a naïve allele, followed by DNA repair processes within the recipient organism that utilize the MGE-containing donor allele as a template for repair (20,68). Other than HEs, a number of classes of endonuclease are known to initiate similar non-mendelian gene conversion events, including transposons, retrotransposons, and REs (44,71-74). Accordingly, the mechanisms by which various MGEs mobilize their encoding genes have been exploited for targeted genome editing applications (75-79).

1.1.3 Strategies to hydrolyze DNA

Site-specific endonucleases vary considerably in their oligomeric assembly on DNA, target site recognition, and method of DNA hydrolysis (Table 1.1) (18,80). Extensive studies have revealed an assortment of catalytic mechanisms utilized by REs and HEs to generate DSBs (Table 1.1, Figure 1.2). Many Type IIP REs, such as EcoRI, stably oligomerize on DNA to provide two active sites to hydrolyze each DNA strand (81). Alternatively, the Type IIP REs BcnI and MvaI bind DNA as monomers and ‘hop’ across the phosphate backbone between nicking reactions (82,83). Some Type IIS REs have also been shown to transiently dimerize or sequentially nick both strands via a conformational change mechanism (Figure 1.2, see also Chapter 1.2.5) (84,85). Yet another DNA hydrolysis mechanism has been proposed for the phospholipase D RE BfiI, where a single composite active site is formed at a dimer interface prior to hydrolysis through a covalent protein:DNA intermediate (86). A similar catalytic mechanism has been observed for the single active site Tn10 transposase, where successive phosphoryl transfer reactions contribute to DNA hydrolysis via an intra-strand DNA hairpin intermediate (87). The mechanisms by which HEs generate double-strand breaks to initiate their mobility have also been studied, and are discussed below.

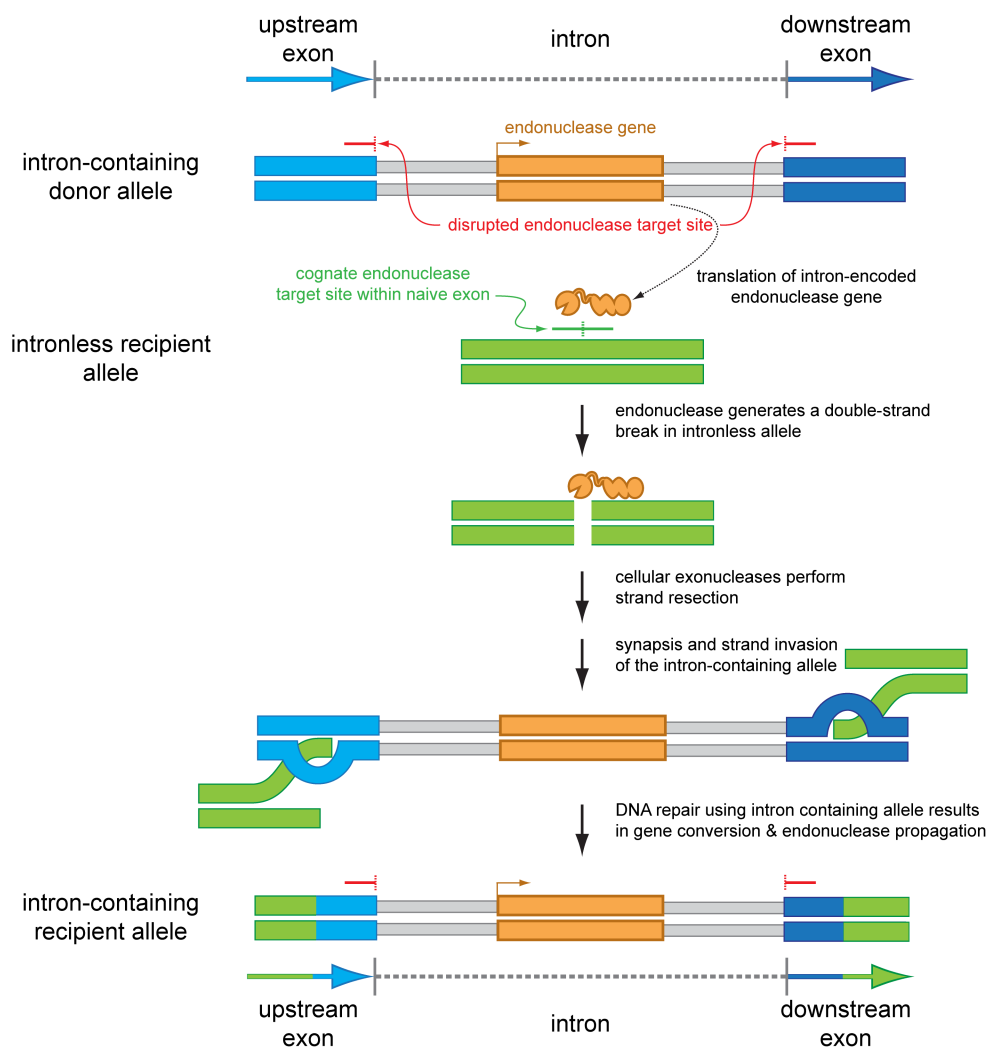


Figure 1.1: Mobility pathway of selfish intron-encoded endonucleases

The expression of an endonuclease gene embedded within an intron results in the lateral transfer of the intron sequence from the donor allele to an intronless recipient allele.

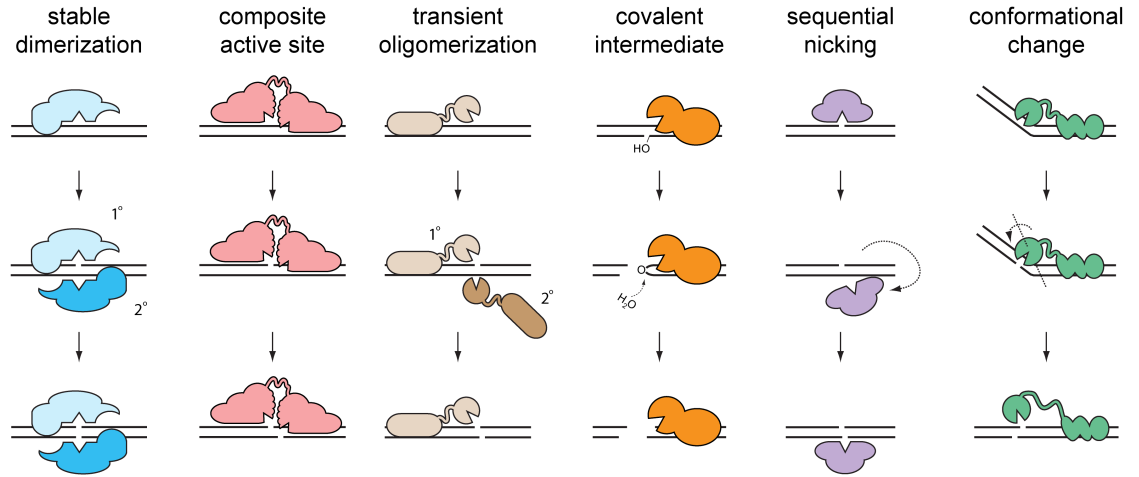


Figure 1.2: Mechanisms of DNA hydrolysis by site-specific endonucleases.

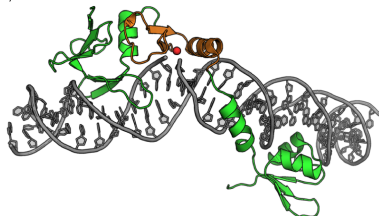
1.2 Homing and Type IIS restriction endonucleases

Like other MGEs, HEs catalyze the ‘selfish’ propagation of their genes to naïve chromosomal sites through a process termed homing (Figure 1.1) (3,66). HEs were discovered when non-mendelian inheritance was observed for an intron in the mitochondrial large rRNA gene of *Saccharomyces cerevisiae* (88,89). It was later shown that the unidirectional gene conversion event was dependent on the intron-encoded open reading frame that expressed the ω factor, now known as the LAGLIDADG HE I-SceI (90,91). HEs share mechanistic and structural similarities with REs (Table 1.1), yet they differ by targeting extended recognition sites (14-36 base pairs) in a sequence tolerant manner (REs are generally intolerant to substitutions within their relatively short 3-8 base pair binding sites) (18,80). While Type IIS REs also bind short sequences, they are mechanistically similar to certain types of HEs by cleaving at a distance from their primary binding site, unlike canonical REs (Table 1.1) (92,93).

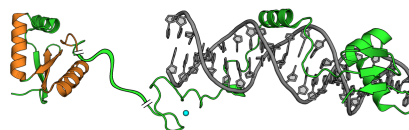
Homing endonuclease genes have been identified in organisms within all kingdoms of life and constitute six different classes based on conserved structural and catalytic motifs (the LAGLIDADG, H-N-H, His-Cys, GIY-YIG, PD-(D/E)xK, and EDxHD families) (Figure 1.3) (94). The H-N-H and His-Cys box families contain highly similar active site structures ($\beta\beta\alpha$ -Me family), while the PD-(D/E)xK and EDxHD classes contain catalytic motifs analogous to those found within orthodox REs (79,95). The relatively small I-Ssp6803I (150 amino acids) was the first identified HE gene that contains the PD-(D/E)xK motif and is distinct from its RE counterparts as it binds a long 23 base pair target site in a tetrameric configuration (41,96). HEs are generally small proteins (<50 kDa) that have mostly adopted two distinct strategies to target extended allelic sites: acquiring multiple DNA-binding elements to bind elongated asymmetric sites in a sequence tolerant manner, or dimeric assembly of comparatively globular folds to double-up their binding capacity on pseudo-palindromic sequences (Figure 1.3) (79,80). The structures and DNA hydrolysis mechanisms utilized by the three major families of HEs and Type IIS REs are further discussed below.

I-HmuI

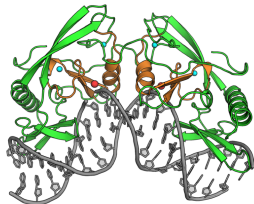
H-N-H ($\beta\beta\alpha$ -Me) HE family
monomer, nickase

**I-TevI**

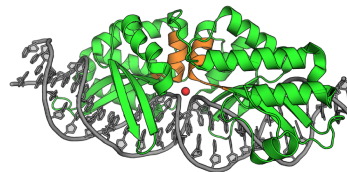
GIY-YIG HE family
monomer, cleavase

**I-PpoI**

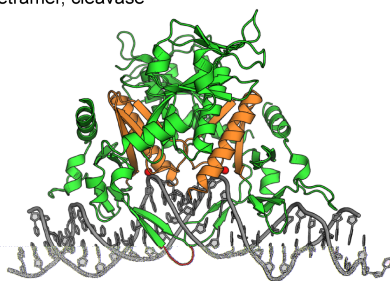
His-Cys ($\beta\beta\alpha$ -Me) HE family
stable dimer, cleavase

**I-OnuI**

LAGLIDADG HE family
single chain monomer, cleavase

**I-Ssp6803I**

PD-(D/E)xK HE family
stable tetramer, cleavase

**FokI**

Type IIS RE family
transient dimer, cleavase

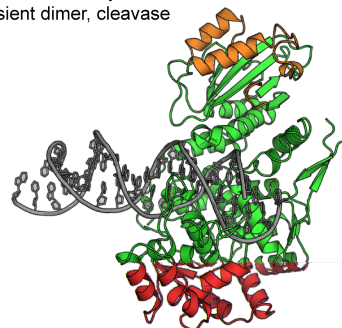


Figure 1.3: Structures of homing and restriction endonucleases

Active site secondary structure elements are shown in orange, magnesium ions in red, and zinc ions in cyan. For simplicity, only the active site structures of the units directly involved in DNA hydrolysis for I-Ssp6803I are highlighted. HE, homing endonuclease; RE, restriction endonuclease; I-HmuI PDB:1U3E; I-TevI PDB:1MK0 & 1I3J; I-PpoI PDB:1A73; I-OnuI PDB:3QQY; I-Ssp6803I PDB:2OST; FokI PDB:1FOK.

1.2.1 GIY-YIG homing endonucleases

Characterization of GIY-YIG HEs began when it was demonstrated that the mobility of the intron associated with the thymidylate synthase (*td*) gene of bacteriophage T4 was dependent on an encoded endonuclease (97). It was later discovered that the endonuclease, named I-TevI, comprised a two-domain structure with an N-terminal domain that contains a GIY-YIG nuclease motif and a C-terminal DNA-binding domain (Figure 1.4A) (3,98). The structural modularity of the endonuclease is also reflected in its target site, as cleavage and DNA-binding regions are present within I-TevI's cognate substrate (99). Furthermore, mutagenesis and interference assays of the intronless *td* substrate revealed that I-TevI's minor groove interactions with substrate are highly tolerant to nucleotide substitutions (100,101). Results from randomized cleavage motif selections and hydrolysis site mapping experiments revealed that I-TevI nicks the bottom- and top-strands of substrate within a CnnnG motif 25 and 23 base pairs upstream of the intron insertion site, respectively (Figure 1.4A) (101,102). Moreover, biochemical assays demonstrated that I-TevI binds DNA as a monomer and induced significant bends in its substrate near the cleavage site (103).

Independent NMR and X-ray crystallography studies of the nuclease and DNA-binding domains of I-TevI provided high-resolution images of the enzyme's modular structure (Figure 1.4C) (104-107). The catalytic domain structure revealed a tightly packed ~90 amino acid core with a $\beta\beta\alpha\alpha\beta\alpha$ fold, where the GIY and YIG motifs are encoded by the first two β -strands within a single active site (104,106,108). Additionally, the C-terminal domain/substrate co-crystal confirmed the sequence-tolerant minor groove interactions by I-TevI, and also identified NUMOD3 α -helix (nuclease associated modular) and helix-turn-helix DNA-binding elements similar to those identified in $\beta\beta\alpha$ -Me HEs (Figure 1.4A, see also Chapter 1.2.4) (105,109-111). Rather unexpectedly, the C-terminal structure of I-TevI revealed a non-canonical zinc-finger module that does not specifically contact DNA (105). It has been speculated that the I-TevI zinc-finger, along with the mostly-unstructured inter-domain linker, acts as a flexible distance determinant to position the GIY-YIG domain on substrate for cleavage (104,107,112,113). Unfortunately, the extraordinarily high specific activity of I-TevI leads to cytotoxicity

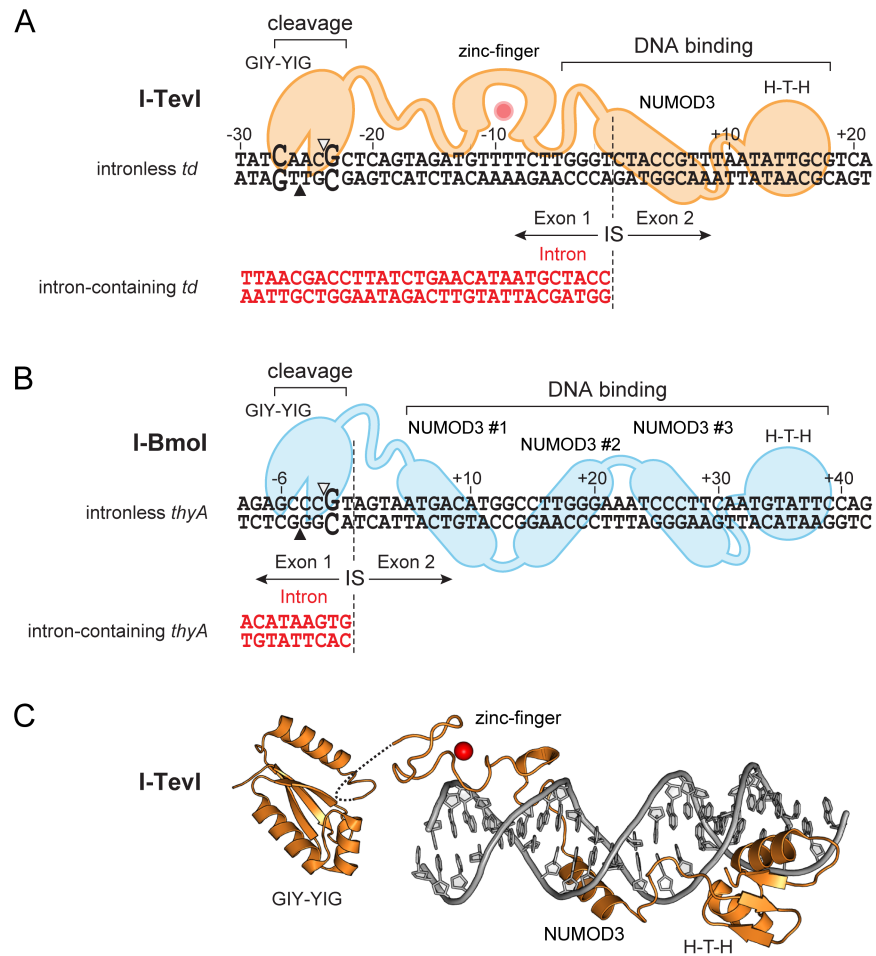


Figure 1.4: GIY-YIG homing endonucleases I-TevI and I-BmoI

Schematic representations highlighting the protein and substrate modularity of (A) I-TevI and (B) I-BmoI. Both enzymes bind asymmetric sequences that span the intron-insertion site (IS) of their respective intronless thymidylate synthase alleles. Intron insertion renders the intron-containing alleles resistant to cleavage by altering critical base pairs at the cleavage site (shown in bold uppercase font). Bottom- and top-strand nicking sites are shown by filled and open triangles, respectively. (C) Composite structural representation of I-TevI displaying the independent N-terminal GIY-YIG domain and C-terminal DNA-binding domain crystal structures.

when expressed in bacteria, precluding purification of the wild-type enzyme (114). Detailed biochemical studies to investigate the cleavage mechanism of I-TevI were therefore prohibited, though a conformational change model of cleavage was proposed where the monomeric nuclease domain sequentially nicks both DNA strands (Figure 1.2) (103).

Fortuitously, an additional GIY-HE similar to I-TevI, named I-BmoI, is encoded within an intron that interrupts the thymidylate synthase gene (*thyA*) of *Bacillus mojavensis* (115). The *thyA* sequence that I-BmoI hydrolyzes is homologous to the *td* allele targeted by I-TevI, with I-BmoI preferring only a guanine nucleotide at its cleavage site in an analogous position to the G of the CnnnG motif observed for I-TevI (Figure 1.4B) (102,116). This preference renders the post-homing intron-containing allele immune to cleavage, as intron insertion disrupts the cognate I-BmoI cleavage site without affecting binding along the asymmetric recognition site (115). I-BmoI is also a modular enzyme where a flexible linker connects the N-terminal GIY-YIG domain to a DNA-binding domain that contains three predicted NUMOD3 α -helices and a helix-turn-helix domain (Figure 1.4B) (115,116). To generate a DSB, I-BmoI nicks the bottom-strand prior to the top strand (non-coding before coding) with multiple distinct DNA-distortions observed surrounding the cleavage site (117). As such, I-BmoI appears to operate via a comparable catalytic mechanism to I-TevI, where substantial protein and substrate conformational rearrangements facilitate cleavage by the GIY-YIG domain. Unlike I-TevI, however, I-BmoI is easily overexpressed and purified from bacteria due to a ~750-fold reduced specific activity (116), allowing us to perform detailed biochemical examinations of protein:DNA interactions and specifically test DNA-hydrolysis models (Chapters 2-4). Additionally, because a major consideration of this thesis is the mechanism by which GIY-HEs hydrolyze DNA, the distribution, modularity, and mechanisms of other enzymes that contain the GIY-YIG motif are described below.

1.2.2 Properties of other GIY-YIG family endonucleases

First discovered within intron-encoded ORFs in bacteriophage and filamentous fungi (118), the GIY-YIG motif is distributed amongst nucleases with diverse structures and functions (104,119,120). Variations of the GIY-YIG consensus are found in HEs and REs

(GIY-REs) (40,97,115,121-126), non-LTR retrotransposons (127-129), non-specific nucleases (130), structure-specific nucleases involved in DNA-repair and recombination (131-136), and the recently described human and *C. elegans* LEM associated nucleases (137,138). Independent crystallographic studies revealed that the GIY-YIG domain is composed of a structurally conserved and tightly packed core that contains three β -strands surrounded by at least two α -helices (Figure 1.5A) (106,139-143). The central GIY-YIG domain has a single active site and has acquired distinct context specific adaptations either within or surrounding the conserved fold to expand its functional diversity.

Prior to studies that solved two separate GIY-RE:substrate co-crystals (142,143), functional roles of conserved tyrosine, histidine, arginine, and asparagine residues remained speculative, other than a metal binding role established for a conserved glutamate residue (E75 in α -helix 3 of I-TevI, residue numbering hereafter for I-TevI). The crystal structures subsequently suggested a concerted single-step cleavage reaction that utilizes a single magnesium ion for hydrolysis (143). The tyrosine residues of the GIY and YIG motifs (Y6 and Y17), found within β -strands 1 and 2, respectively, potentially act interchangeably as the primary general base to activate a water molecule (though Y6 is most likely). Additionally, the imidazole ring of H31 in α -helix 2 is within hydrogen bonding distance of both Y6 and Y17 and may assist in proton abstraction to initiate hydrolysis. The active site divalent metal ion, coordinated primarily by E75 in α -helix 3 and to some extent by N90, is thought to simultaneously stabilize the reaction intermediate and 3' leaving group. Interestingly, the role of R27 in α -helix 1 remains somewhat ambiguous, as it is found in different rotameric conformations between the dimer subunits in the structure of the GIY-RE Eco29kI structure (142). Roles have been proposed for the conserved arginine to contact the phosphate backbone, interact with the neighboring ribose, or to directly participate in catalysis (142,143). Furthermore, it is highly likely that additional contextually distinct functional residues within separate GIY-YIG architectures are unidentifiably by conservation alone, and I present data in Chapter 2 that identifies non-conserved yet functionally relevant residues of the I-BmoI catalytic domain (120).

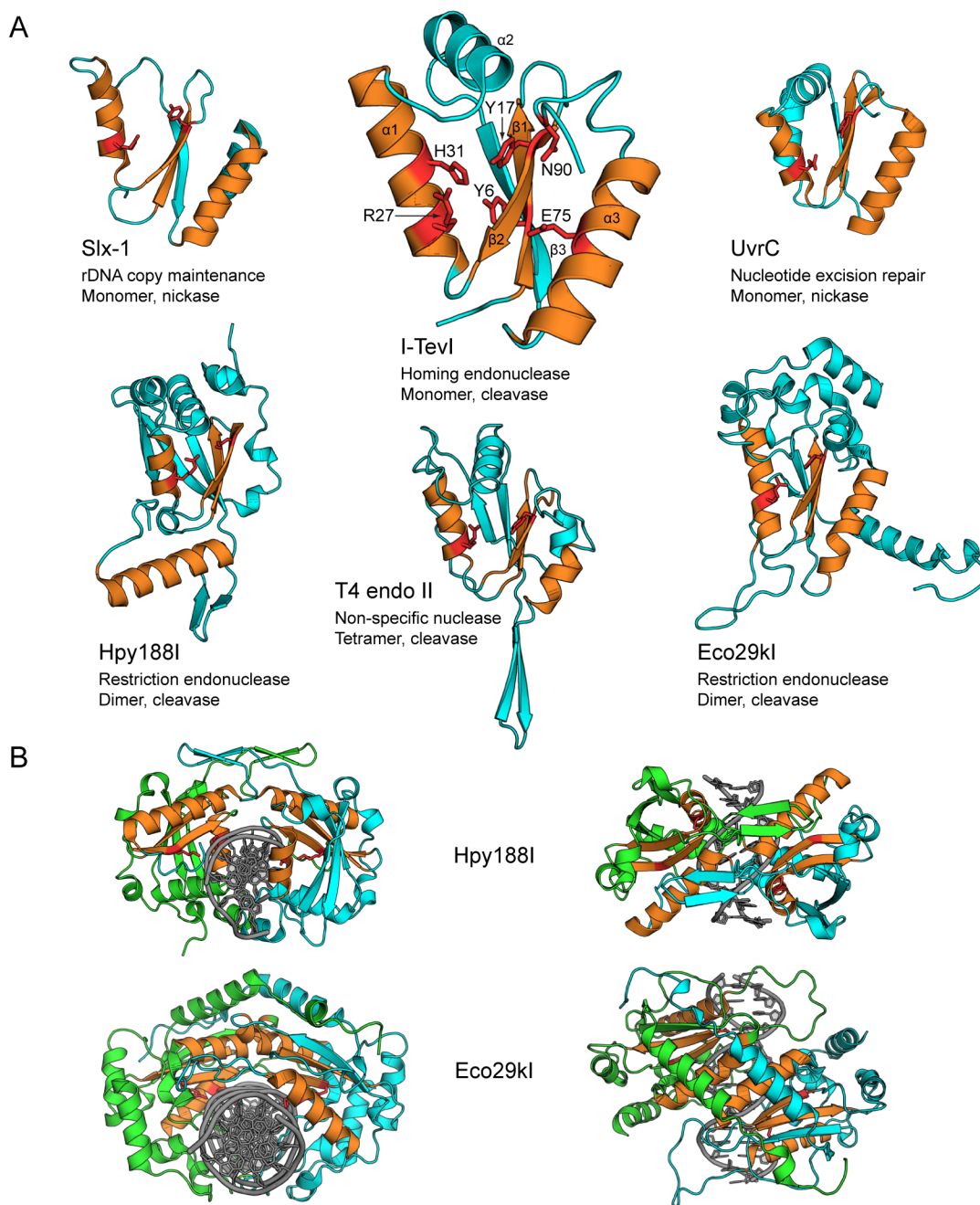


Figure 1.5: GIY-YIG domain structures

(A) Structures of six GIY-YIG nuclease domains. Secondary structure elements that contain the active site residues are shown in orange, and catalytic residues are shown in red for I-TevI. For the remaining five structures, only residues homologous to Y17 and R27 are shown. (B) DNA-bound structures of dimeric GIY-YIG REs Hpy188I and Eco29kI, with dimer half-units shown in cyan and green, and active site structures shown in orange. Views along the DNA axis or perpendicular to the plane of DNA are shown (*left and right panels*, respectively). Slx1 PDB:1ZG2; I-TevI PDB:1MK0; UvrC PDB:1YCZ; Hpy188I PDB:3OQG; T4 endo II PDB:2WSH; Eco29kI PDB:3NIC.

The conserved amino acids of GIY-YIG endonucleases are found within a structurally similar active site (Figure 1.5A), yet individual family members employ distinct architectures and function by a number of mechanisms (reviewed in reference 142). For example, GIY-REs have acquired supplementary protein folds that supply base-specific and backbone contacts necessary to bind 5 or 6 base pair target sites in a specific manner (142,143). The additional folds also assist in oligomerization as a mechanism for DSB formation, where Eco29kI and Hpy188I assemble as dimers (Figure 1.5B) and Cfr42I forms a tetramer upon DNA binding (126,143,144). Similarly, bacteriophage T4 endonuclease II, which degrades host DNA during infection, forms a tetrameric structure that nicks both strands of a single substrate or synapses between multiple targets to nick one strand of each substrate (145). Alternatively, the bacterial nucleotide excision repair proteins UvrC and Cho employ single GIY-YIG domains as nickases to hydrolyze one DNA strand 3' to DNA lesions of different sizes (132,133). These endonuclease also typically encode a secondary nuclease module to nick upstream of the damaged DNA base to fully excise the lesion (119). While it has been speculated that GIY-HEs hydrolyze DNA via a sequential nicking and conformational change mechanism (103,117,146), cleavage by I-TevI or I-BmoI has remained poorly understood. Both characterized GIY-HEs lack the supplementary catalytic domain folds necessary for dimerization and no additional putative nuclease motifs or active sites have been predicted. Therefore, I specifically test models for DNA hydrolysis by single active site enzymes in Chapters 4 and 5 (for I-BmoI and engineered I-TevI nucleases, respectively) towards understanding the GIY-HE cleavage mechanism.

1.2.3 LAGLIDADG homing endonucleases

LAGLIDADG homing endonucleases (LHEs) have a broad phylogenetic distribution and target sequences of moderate length from 16 to 26 base pairs (80,147). As described above for I-SceI, the discovery of LHEs provided a foundation from which we began to understand the mobility of selfish MGEs (52). With over 10 solved crystal structures to date, LHEs are well-characterized structurally and hydrolyze DNA within two active sites located at the base of the pseudo-dimeric enzyme interface (Figure 1.3) (80,148,149). Also termed dodecapeptide-nucleases, LHEs encode α -helices within each subunit that

contain the moderately conserved 10-residue motif and form the major interactions at the subunit interface. LHEs with a single copy of the motif function as homodimers and target near-palindromic sites (150,151), while single peptide chain LHEs with two copies of the motif and are monomers that can target asymmetric sequences (152,153). LHEs utilize four extended antiparallel β -sheets to bind DNA through the major groove where interactions from each half-unit span the cleavage site and induce a directed bend in the substrate (Figure 1.3) (150-154). Each half-unit of either the homodimeric or single chain enzymes hydrolyzes one strand of DNA, while for some LHEs it is suspected that a shared divalent metal ion across both active sites can contribute to DNA hydrolysis (80). As such, the DNA-binding mechanism and unique subunit interface has made LHEs intriguing candidates for genome editing applications (148,149,154,155). These prospects will be further discussed in Chapter 1.4.3, and we investigate the utility of a catalytically inactive LHE as a DNA-binding platform for engineered nucleases in Chapter 5.

1.2.4 $\beta\beta\alpha$ -Me homing endonucleases

His-Cys box and H-N-H family enzymes form structurally similar $\beta\beta\alpha$ -Me active site architectures and are found within a selection of endonucleases that perform various biological functions (80,156). Aside from HEs (109,157), the $\beta\beta\alpha$ -Me fold is found within REs (39,42,158), non-specific bacterial colicins and fungal nucleases (including the *Serratia* nuclease) (159-162), structure-specific holiday junction resolvases (163), and other DNA processing enzymes (164,165). The His-Cys motif is characterized by histidine and cysteine residues and is found within many HE genes (166-169). I-PpoI, the most well characterized His-Cys HE family member, dimerizes to bind and cleave a relatively short pseudo-palindromic 14 base pair recognition sequence (Figure 1.3) (111,170,171). Crystallographic studies revealed a tertiary structure rich in metal ions where multiple zinc ions are coordinated by histidine and cysteine residues to stabilize the protein core, and magnesium ions are coordinated within the $\beta\beta\alpha$ -Me active site to assist in hydrolysis by interacting with the scissile phosphate (157,172). A significant bend in DNA across the base pairs between the two active sites is observed, facilitated by NUMOD4 DNA-binding elements within each monomer (110,111). Interestingly, each subunit positions a hydrophobic leucine side-chain within the minor groove to widen the

distance between the phosphate backbones prior to hydrolysis, a mechanism that has been suggested to increase accessibility to the cleavage site prior to hydrolysis (173).

While they share a $\beta\beta\alpha$ -Me fold, the H-N-H HEs are structurally dissimilar to I-PpoI as they target longer recognition sequences (24-40 base pairs) and do not bind DNA as symmetric dimers (52,79). The crystal structure of the H-N-H HE I-HmuI substrate-bound complex revealed an extended monomeric configuration with N- and C-terminal DNA-binding domains flanking either side of the H-N-H active site (Figure 1.3) (109). Within the context of H-N-H HEs the catalytic motif inherently functions as a nickase, as I-HmuI and I-BasI bind DNA as monomers to nick only a single strand (174-176), and I-TevIII forms an asymmetric dimer to induce a DSB by nicking both strands (177). It is possible that I-HmuI and related H-N-H HEs nick only a single strand of DNA as a result of their monomeric assembly on DNA combined with rotational restriction of the active site by surrounding DNA-binding elements. The I-HmuI N-terminal DNA binding domain displays high similarity to the NUMOD4 domains observed within I-PpoI, while the C-terminal domain contains NUMOD3 α -helices and a helix-turn-helix domain (similar to those observed for I-TevI), sandwiching the active site between elements with substantial affinity for DNA (109,110,175). Furthermore, the DNA-binding modules of I-HmuI span the phosphate backbone twice to induce a 40° bend near the cleavage site, and like I-PpoI widen the minor groove (80,109).

1.2.5 Type IIS restriction endonucleases

Type IIS REs were the first identified Type II subclass due to their capacity to nick at least one strand of DNA at a precise distance anywhere from 0 to 21 base pairs from the primary 4-7 base pair binding site (32,92,93). Like GIY-HEs, many Type IIS REs are two-domain proteins connected by a flexible linker (178-180), with FokI being the most highly characterized of any Type IIS RE. A number of catalytic mechanisms have been described for the four subclasses of Type IIS REs, and FokI binds DNA as a globular monomeric protein to initiate hydrolysis (180,181). Upon DNA binding, the single active site nuclease domain of FokI is released from the globular structure allowing extension to the cleavage site where hydrolysis occurs non-specifically (84,182). Hydrolysis is most

efficient when dimerization with another substrate bound FokI molecule at a proximal secondary site occurs, another hallmark of many Type IIS REs (84,180,183). FokI nicks both DNA strands at 9/13 base pairs away from its primary binding site to generate a DSB, yet the composition of nucleotides surrounding the cleavage site does not affect hydrolysis (184,185). Thus, most Type IIS REs are said to non-specifically cleave DNA while retaining a distance preference for the location of hydrolysis (92,93). The modularity and non-specific nuclease activity of the FokI nuclease domain has garnered significant interest in the field of genome editing and has led to its adaptation within multiple engineered platforms (35,186). The architectures of FokI chimeric nucleases are discussed in Chapters 1.4.1 and 1.4.2, while the genome editing successes, limitations, and prospects are discussed in Chapters 1.5 and 6.

1.3 Genome editing

The ability to stably modify genomes has led to the elucidation of the functions of many genes in model systems and initiated the field of nuclease mediated gene therapy. In the late 1970's, the first site-directed homologous gene targeting method demonstrated that synthetic genes could be stably integrated into *Saccharomyces cerevisiae* (187). The seminal study revealed that yeast could heritably acquire a metabolic gene from an exogenous source through homologous recombination (HR), when a stable *leu2⁻* strain was converted to a *LEU2⁺* strain upon transformation with a bacterial plasmid encoding the corresponding leucine biosynthesis gene. This technique was subsequently adopted for many gene knock-in or knock-out studies in yeast (188-191), yet similar efficient gene conversion was not observed in mice or human cells (192-195). The finding that on average less than 0.01% of human cells were genetically modified was due to the inherent biology of mammalian cells, where nonhomologous end joining (NHEJ) is the predominant DNA repair pathway at most cell stages versus HR in yeast, precluding the methodology from relevant human cell therapies (194,196,197).

A shift in the paradigm of genome editing in mammalian cells occurred when traditional homologous gene replacement methods were coupled with a site-directed DSB (198,199). When a vector expressing the LHE I-SceI was cotransfected with a donor template for DNA repair via HR, greater than 100-fold increases in gene conversion events were

observed versus transfection of the HR template alone (200-202). By generating a sequence-specific break to initiate DNA repair in the presence of a donor template, the natural DSB-repair preference in mammalian cells was shifted from NHEJ to HR. The realization that genomes could be modified at defined locations was encouraging, as early efforts with high efficiency recombination based approaches using transposons or viral integrases remained relatively non-specific (203,204). Consequently, much effort has been devoted to developing site-specific endonuclease architectures for targeted genome modification, a subset of which are further discussed below.

1.4 Genome editing with engineered nucleases

At the beginning of this century, multiple clinical trials were initiated to treat two forms of severe combined immunodeficiency (X-SCID and ADA-SCID) using retroviral integrases (205-207). SCID is a genetic disorder of the adaptive immune system where the development of both T- and B-lymphocytes is impaired, making it an ideal disease model for therapy as corrected cells will selectively proliferate over a patient's intrinsic population of deficient cells (208,209). Retroviral treatments were conducted on extracted hematopoietic stem cells to restore a functional immune system to recipients by stably integrating wild-type copies of their deficient genes (205,206,210). At early stages in their treatment regimens, patients appeared to recover from their SCID symptoms, however it was later discovered that in many cases their therapies had promoted the development of leukemia as a result of non-directed transgene integration adjacent to proto-oncogenes (209,211,212). Though these and other clinical trials prolonged the life of most patients, they reinforced the necessity for the development of highly targeted therapeutic technologies that could function in a site-specific manner.

Around the same time, high frequency gene replacement in mammalian cells using site-specific nucleases renewed optimism within the field of gene therapy (78,213-215). By mimicking the natural propagation mechanisms of 'selfish' genetic elements, engineered endonucleases were employed to disrupt or delete genes, correct deleterious single-nucleotide polymorphisms (SNPs), or introduce exogenous corrective genetic cassettes (Figure 1.6) (216-218). Upon further refinement of engineered nuclease technologies, it has become possible to modulate the desired gene correction mechanism following a

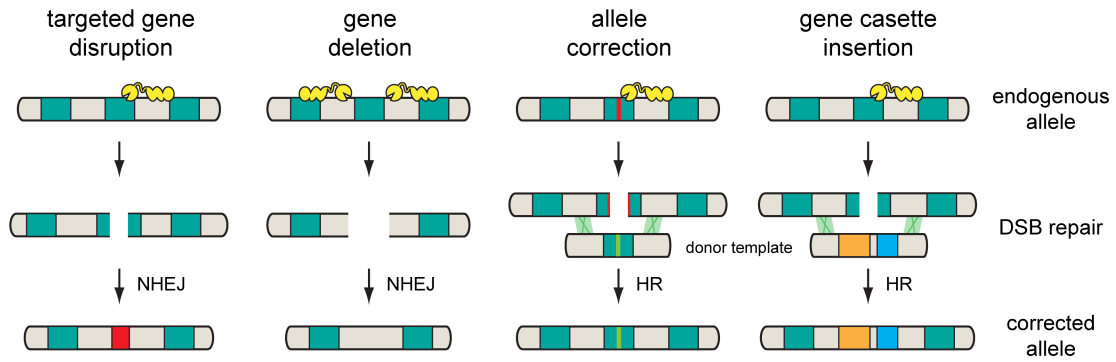


Figure 1.6: Genome editing outcomes from targeted DNA breaks

Site-specific endonucleases initiate DNA repair pathways at endogenous alleles leading to gene conversion events that differ depending on whether or not an exogenous donor template is provided.

DSB (NHEJ vs. HR, elaborated in Chapter 6) (219,220). Though some limitations associated with current platforms have led to the exploration of alternative nuclease technologies (221-223), the majority of engineered site-specific endonucleases utilize one of two architectures: customizable DNA-binding modules fused to the non-specific FokI nuclease domain, or LHEs whose binding specificities have been reengineered to target non-native sequences.

1.4.1 Zinc-finger nucleases

As described in Chapter 1.2.5, the Type IIS RE FokI is a bi-partite enzyme that transiently dimerizes through its catalytic domain to hydrolyze DNA. Accordingly, the modularity and catalytic mechanism of FokI has been exploited to generate chimeric REs by fusing the non-specific nuclease domain to targetable zinc-finger DNA-binding proteins, creating zinc-finger nucleases (ZFNs, Figure 1.7A, B) (34,35,224). The oligomeric nature of FokI requires a “head-to-head” design of ZFNs and necessitated optimization of the dimer interface and for efficient cleavage (Figure 1.7B) (225-227). To expand the genomic space targetable by ZFNs, the re-programmability of zinc-finger proteins as DNA-binding modules has been extensively explored (228-231). Each zinc-finger unit contributes 3-4 base pairs of DNA-binding specificity, where engineered arrays of six zinc-fingers can theoretically bind an 18 base pair sequence (Figure 1.7B) (232). The combined optimization efforts have led to ZFN architecture refinements and the development of high-throughput assembly methods to construct active dimeric ZFN pairs (217,233-235). ZFNs have been notably successful tools for gene editing studies in model organisms and human cell lines (214,236-239), yet the limited targeting range, frequently ambiguous nucleotide recognition, and imperfect assembly of zinc-finger arrays has reduced the widespread technological potential of ZFNs (240,241).

1.4.2 Transcription activator-like effector nucleases

An alternative highly customizable DNA-binding platform has begun to supplant zinc-finger arrays as the preferred targeting module in FokI chimeric nucleases (218,242). While characterizing the virulence of *Xanthomonas* strains as plant pathogens, it was

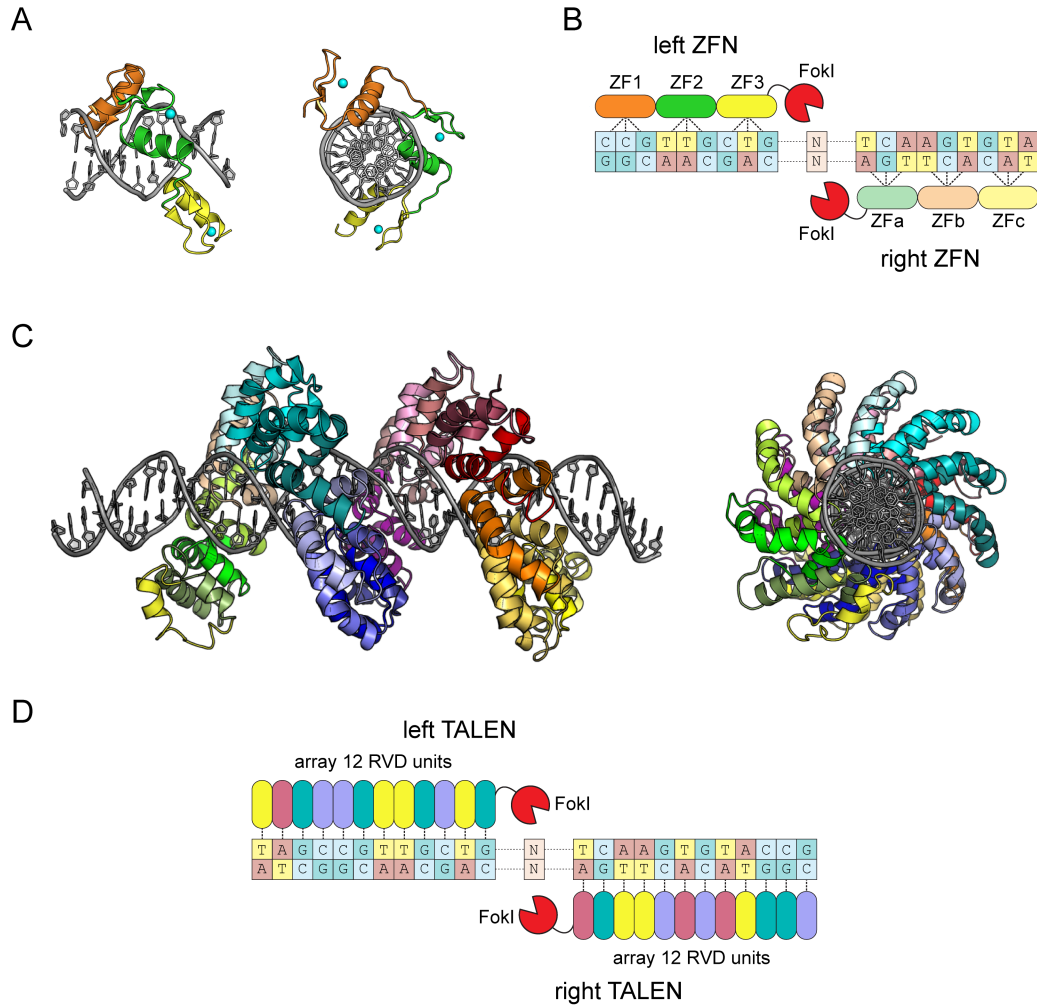


Figure 1.7: FokI nuclease derived ZFNs and TALENs

(A) Structural representations of a zinc-finger array. Separate zinc-finger units that each target 3-4 base pairs are coloured in orange, green, and yellow, with zinc ions coloured in cyan. (B) Schematic of two 3-member ZFNs in a head-to-head orientation, each of which interacts with a 9 base pair target. (C) Structural representations of a 25-repeat TAL array, with repeats that each contact a single base pair coloured separately. (D) Schematic of two 12-repeat TALENs in a head-to-head orientation, each of which interacts with 12 base pairs.

discovered that transcription activator-like (TAL) effector proteins recognize DNA through a highly repetitive structure (Figure 1.7C) (243). Uniquely, 2 amino acids within a repeated and otherwise almost absolutely conserved 33-35 amino acid structure dictate a single-base pair contact (termed the repeat-variable diresidue, or RVD) (244,245). Recent crystallographic studies confirmed the predicted 1:1 cipher that governs DNA recognition, where an array of sequential RVDs within the repeated TAL backbone can be assembled to bind defined sequences (Figure 1.7D) (246,247). It wasn't long before studies emerged demonstrating that TALENs (TAL effector nucleases, where the FokI nuclease domain is fused to the TAL domain) efficiently target desired loci in yeast, plants, zebrafish, and mammalian systems (248-254). The early success of this architecture led to a much more rapid refinement of TALEN development and assembly compared to ZFNs, due in part to the pre-established knowledge of FokI biology. Within three years, numerous high-throughput assembly methodologies have been developed with considerable promise for widespread application of TALENs (249,255-258). Similarly to ZFNs, however, TALENs are subject to the limitations of the inherent biology and mechanism of the FokI nuclease domain and must therefore be constructed in a head-to-head design (4).

1.4.3 Engineered LAGLIDADG nucleases

As the initiators of modern genome engineering, LHEs have continued to drive innovation within the field. Their relatively small size and robust structural characterization have made them ideal candidates for application-oriented genome editing studies (149). Additionally, their natural pseudo-symmetric modularity, modifiable DNA-binding specificity, and growing number of identified family members have begun to translate to an almost unlimited targetable sequence space by engineered LHEs (148,149,154,259,260). Unlike the head-to-head design of FokI derived ZFNs and TALENs, engineered LHEs are single module enzymes that cleave within their respective binding sites. Initial studies to engineer the binding specificity of LHEs involved the generation of I-CreI/I-DmoI chimeras that target novel hybrid DNA sequences. The engineered 'E-DreI' chimeric LHE contained one I-CreI homodimeric subunit that functionally replaced one half of the single chain LHE I-DmoI (Figure 1.8A) (261,262).

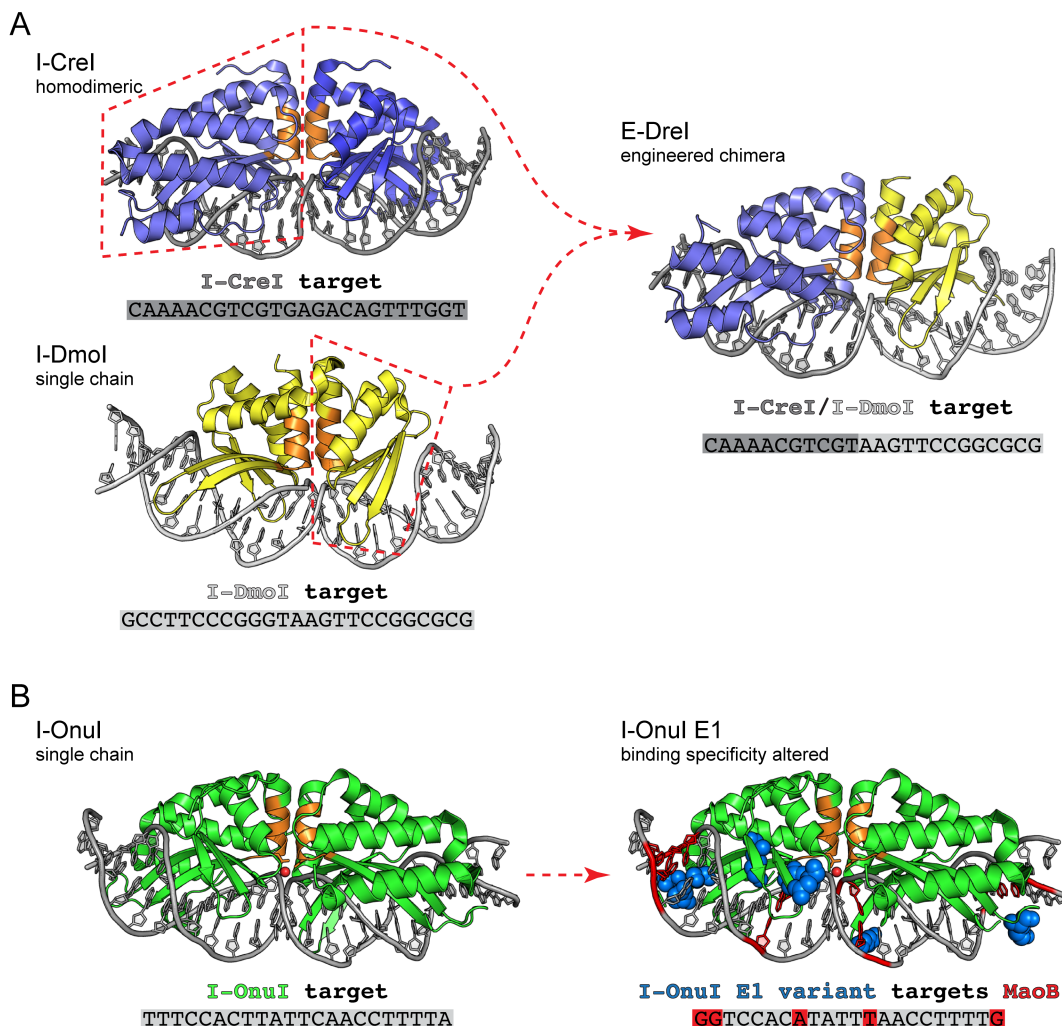


Figure 1.8: Methods to engineer LAGLIDADG endonuclease specificity

(A) Chimeras are created from left and right units derived from distinct parent enzymes and cleave a corresponding hybrid target site. (B) Directed evolution of amino acids that specify DNA interactions. Altered amino acid side chains are space filled in blue, while mutated base pairs are shown in red. Active site LAGLIDADG motifs are coloured in orange.

More recently, a study generated 30 different chimeric LHEs from the I-OnuI subfamily, 14 of which displayed appreciable *in vitro* activity (155). Yet, the most powerful and successful directed evolution approaches have involved the semi-rational high-throughput re-engineering of the LHE DNA-binding interface (Figure 1.8B) (259,263-265). Using this method, LHEs have been re-targeted to engineer maize genes (266), generate populations of malaria resistant mosquitoes (267), and target numerous human alleles (154,268-270). Despite the successes associated with engineered LHEs, each directed evolution method must be coupled with concurrent (and sometimes challenging) re-optimization of catalytic activity and protein stability, as LHEs are highly dynamic single module enzymes for which most intra-molecular interactions are not well understood (271-273).

1.5 Scope of the thesis: can the cleavage mechanism of the GIY-YIG domain detoxify engineered nucleases?

This thesis continues the story of GIY-HE characterization that began over 24 years ago when I-TevI was discovered (97). To date, studies with I-TevI have provided biochemical and structural insight into the roles of the nuclease, linker, and DNA-binding domains of GIY-HEs, yet were unable to specifically test models for DSB formation. Thus, the DNA hydrolysis mechanism utilized by GIY-HEs to mobilize their own genes has remained unknown. Using the related endonuclease I-BmoI, I have revisited this story to investigate catalytic domain interactions with substrate and specifically test DSB mechanisms. By understanding the biology of GIY-HE's, I have been able to exploit the inherent properties of GIY-YIG nuclease domains to create chimeric enzymes fused to the three targetable DNA-binding platforms used in the field of genome editing.

As described within this thesis, I propose a monomeric cleavage mechanism for GIY-HEs that involves the repositioning of the single active site nuclease domain to perform sequential nicking reactions. The conformational change mechanism proceeds through multiple distinct DNA distortions, the first of which is dependent on an identified protein:DNA interaction that acts as a molecular tether to anchor the nuclease domain on substrate. Given that I-BmoI requires a GC base pair at its cleavage site for efficient

hydrolysis (Figure 1.4B), this added preference would imply that a di-nucleotide targeting rule exists for I-BmoI. Accordingly, an analogous protein:DNA tether for I-TevI would lead to an expanded targeting preference from a CnnnG di-nucleotide to a tri-nucleotide. This inference is plausible, as the functional similarity of the GIY-HE linkers has previously been demonstrated when chimeric TevBmo and BmoTev nucleases were shown to cleave corresponding hybrid target sites (113). This study also emphasized that the I-TevI and I-BmoI nuclease domains retained function in the context of a non-native DNA-binding platform, a significant result when considering either domain for genome editing applications.

I envision the monomeric and site-specific cleavage mechanism of GIY-HEs as an advantageous property to ‘detoxify’ and improve existing genome editing reagents. Studies have revealed that the inherent non-specific nuclease activity and dimeric requirement of FokI-derived ZFNs and TALENs has led to high off-target cleavage rates by established ZFNs that are in clinical trials (274,275). While TALENs have performed slightly better than ZFNs in terms of genome modification activity at similar loci (276), the specificity of either platform has not been thoroughly examined and they remain susceptible to the same non-specific and toxic mechanisms inherent to the FokI nuclease domain (4). Conversely, the directed evolution of LHE DNA-binding specificity has not proceeded as readily as had been hoped. While the confinement of DNA-binding and cleavage properties within a single small architecture has advantages, subtle changes to the engineered LHE structure can have unforeseen consequences and often necessitates laborious re-optimization of catalytic activity. Consequently, the difficulties associated with these technologies have led to the exploration of alternative nuclease platforms that include chimeras of Type IIP REs to zinc-fingers and inactive LAGLIDADG HEs (221,277), and synthetic ‘guide’ RNAs that stimulate DNA hydrolysis by the CRISPR/Cas system (223,278,279). Unfortunately, the efficacy and specificity of either platform have not been robustly tested.

The work presented within this thesis has directed the development of the first set of monomeric reprogrammable nucleases. By understanding the fundamental biology and cleavage mechanism of GIY-HE’s, I have been able to rationally design and test GIY-

YIG engineered nucleases fused to the zinc-finger, LHE, and TAL DNA-binding platforms (GIY-ENs). It is possible that the inherent properties of GIY-ENs may circumvent the drawbacks associated with other technologies, as the requirements to cleave at a preferred motif and function as a monomer may abolish the promiscuous activity observed with FokI-derived nucleases. These potentially advantageous properties are elaborated in Chapter 6, along with future considerations and prospects for GIY-ENs. Overall, it has been important is to understand the inherent cleavage mechanism, nucleotide preference, and modularity of the GIY-HE family prior to generating chimeric and targetable enzymes.

1.6 References

1. Yang,W. (2011) Nucleases: Diversity of structure, function and mechanism. *Q. Rev. Biophys.*, **44**, 1-93.
2. Nishino,T. and Morikawa,K. (2002) Structure and function of nucleases in DNA repair: Shape, grip and blade of the DNA scissors. *Oncogene*, **21**, 9022-9032.
3. Belfort,M. and Roberts,R.J. (1997) Homing endonucleases: Keeping the house in order. *Nucleic Acids Res*, **25**, 3379-88.
4. Halford,S.E., Catto,L.E., Pernstich,C., Rusling,D.A. and Sanders,K.L. (2011) The reaction mechanism of fokI excludes the possibility of targeting zinc finger nucleases to unique DNA sites. *Biochem. Soc. Trans.*, **39**, 584-588.
5. Nestle,M. and Roberts,W.K. (1969) An extracellular nuclease from serratia marcescens. I. purification and some properties of enzyme. *J. Biol. Chem.*, **244**, 5213-8.
6. Muropastor,A.M., Flores,E., Herrero,A. and Wolk,C.P. (1992) Identification, genetic-analysis and characterization of a sugar-nonspecific nuclease from the cyanobacterium anabaena sp pcc-7120. *Mol. Microbiol.*, **6**, 3021-3030.
7. Miller,M.D., Tanner,J., Alpaugh,M., Benedik,M.J. and Krause,K.L. (1994) 2.1-angstrom structure of serratia endonuclease suggests a mechanism for binding to double-stranded dna. *Nat. Struct. Biol.*, **1**, 461-468.
8. Yonemura,K., Matsumoto,K. and Maeda,H. (1983) Isolation and characterization of nucleases from a clinical isolate of serratia-marcescens-kums 3958. *J. Biochem.*, **93**, 1287-1295.

9. Meiss,G., Friedhoff,P., Hahn,M., Gimadutdinow,O. and Pingoud,A. (1995) Sequence preferences in cleavage of dsdna and ssdna by the extracellular serratia-marcescens endonuclease. *Biochemistry (N. Y.)*, **34**, 11979-11988.
10. Wilson,D.M., Takeshita,M., Grollman,A.P. and Demple,B. (1995) Incision activity of human apurinic endonuclease (ape) at abasic site analogs in dna. *J. Biol. Chem.*, **270**, 16002-16007.
11. Lindahl,T. and Wood,R.D. (1999) Quality control by DNA repair. *Science*, **286**, 1897-1905.
12. Mol,C.D., Izumi,T., Mitra,S. and Tainer,J.A. (2000) DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination (vol 403, pg 451, 2000). *Nature*, **404**, 525-525.
13. Holliday,R. (1964) Mechanism for gene conversion in fungi. *Genet. Res.*, **5**, 282-304.
14. Lilley,D.M.J. and White,M.F. (2001) The junction-resolving enzymes. *Nature Reviews Molecular Cell Biology*, **2**, 433-443.
15. Hadden,J.M., Declais,A., Carr,S.B., Lilley,D.M.J. and Phillips,S.E.V. (2007) The structural basis of holliday junction resolution by T7 endonuclease. *Nature*, **449**, 621-U15.
16. Bickle,T.A. and Kruger,D.H. (1993) Biology of dna restriction. *Microbiol. Rev.*, **57**, 434-450.
17. Raleigh,E.A. and Brooks,J.E. (1998) Restriction Modification Systems: Where they are and what they do.
18. Pingoud,A., Fuxreiter,M., Pingoud,V. and Wende,W. (2005) Type II restriction endonucleases: Structure and mechanism. *Cellular and Molecular Life Sciences*, **62**, 685-707.
19. Shapiro,J.A. (1983) Mobile Genetic Elements.
20. Gimble,F.S. (2000) Invasion of a multitude of genetic niches by mobile endonuclease genes. *FEMS Microbiol Lett*, **185**, 99-107.
21. Luria,S.E. and Human,M.L. (1952) A nonhereditary, host-induced variation of bacterial viruses. *J. Bacteriol.*, **64**, 557-569.
22. Bertani,G. and Weigle,J.J. (1953) Host controlled variation in bacterial viruses. *J. Bacteriol.*, **65**, 113-121.
23. Meselson,M. and Yuan,R. (1968) Dna restriction enzyme from E coli. *Nature*, **217**, 1110-4.

24. Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2010) REBASE-a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Res.*, **38**, D234-D236.
25. Yuan,R. (1981) Structure and mechanism of multifunctional restriction endonucleases. *Annu. Rev. Biochem.*, **50**, 285-315.
26. Wilson,G.G. (1991) Organization of restriction-modification systems. *Nucleic Acids Res.*, **19**, 2539-2566.
27. Anderson,J.E. (1993) Restriction endonucleases and modification methylases. *Curr. Opin. Struct. Biol.*, **3**, 24-30.
28. Lin,L.F., Posfai,J., Roberts,R.J. and Kong,H.M. (2001) Comparative genomics of the restriction-modification systems in helicobacter pylori. *Proc. Natl. Acad. Sci. U. S. A.*, **98**, 2740-2745.
29. Kelly,T.J. and Smith,H.O. (1970) A restriction enzyme from hemophilus-influenzae. 2. base sequence of recognition site. *J. Mol. Biol.*, **51**, 393-409.
30. Smith,H.O. and Wilcox,K.W. (1970) A restriction enzyme from hemophilus-influenzae. 1. purification and general properties. *J. Mol. Biol.*, **51**, 379-91.
31. Smith,H.O. and Nathans,D. (1973) Suggested nomenclature for bacterial host modification and restriction systems and their enzymes. *J. Mol. Biol.*, **81**, 419-423.
32. Roberts,R.J., Belfort,M., Bestor,T., Bhagwat,A.S., Bickle,T.A., Bitinaite,J., Blumenthal,R.M., Degtyarev,S., Dryden,D.T., Dybvig,K., et al. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res*, **31**, 1805-12.
33. Orłowski,J. and Bujnicki,J.M. (2008) Structural and evolutionary classification of type II restriction enzymes based on theoretical and experimental analyses. *Nucleic Acids Res.*, **36**, 3552-3569.
34. Kim,Y.G. and Chandrasegaran,S. (1994) Chimeric restriction endonuclease. *Proc Natl Acad Sci U S A*, **91**, 883-7.
35. Kim,Y.G., Cha,J. and Chandrasegaran,S. (1996) Hybrid restriction enzymes: Zinc finger fusions to fok I cleavage domain. *Proc. Natl. Acad. Sci. U. S. A.*, **93**, 1156-60.
36. Lanio,T., Jeltsch,A. and Pingoud,A. (1998) Towards the design of rare cutting restriction endonucleases: Using directed evolution to generate variants of EcoRV differing in their substrate specificity by two orders of magnitude. *J Mol Biol*, **283**, 59-69.

37. Lanio,T., Jeltsch,A. and Pingoud,A. (2000) On the possibilities and limitations of rational protein design to expand the specificity of restriction enzymes: A case study employing EcoRV as the target. *Protein Eng*, **13**, 275-81.
38. Kovall,R.A. and Matthews,B.W. (1999) Type II restriction endonucleases: Structural, functional and evolutionary relationships. *Curr. Opin. Chem. Biol.*, **3**, 578-583.
39. Saravanan,M., Bujnicki,J.M., Cymerman,I.A., Rao,D.N. and Nagaraja,V. (2004) Type II restriction endonuclease R.KpnI is a member of the HNH nuclease superfamily. *Nucleic Acids Res.*, **32**, 6129-6135.
40. Ibryashkina,E.M., Zakharova,M.V., Baskunov,V.B., Bogdanova,E.S., Nagornyykh,M.O., Den'mukhamedov,M.M., Melnik,B.S., Kolinski,A., Gront,D., Feder,M., et al. (2007) Type II restriction endonuclease R.Eco29kI is a member of the GIY-YIG nuclease superfamily. *BMC Structural Biology*, **7**, 48.
41. Zhao,L., Bonocora,R.P., Shub,D.A. and Stoddard,B.L. (2007) The restriction fold turns to the dark side: A bacterial homing endonuclease with a PD-(D/E)-XK motif. *EMBO J*, **26**, 2432-42.
42. Shen,B.W., Heiter,D.F., Chan,S., Wang,H., Xu,S., Morgan,R.D., Wilson,G.G. and Stoddard,B.L. (2010) Unusual target site disruption by the rare-cutting HNH restriction endonuclease pacl. *Structure*, **18**, 734-743.
43. Naito,T., Kusano,K. and Kobayashi,I. (1995) Selfish behavior of restriction-modification systems. *Science*, **267**, 897-899.
44. Kobayashi,I. (2001) Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res.*, **29**, 3742-3756.
45. Kolodkin,A.L., Klar,A.J.S. and Stahl,F.W. (1986) Double-strand breaks can initiate meiotic recombination in *saccharomyces-cerevisiae*. *Cell*, **46**, 733-740.
46. Dawkins,R. (1976) *The Selfish Gene*. Oxford University Press.
47. Dover,G. and Doolittle,W.F. (1980) Modes of genome evolution. *Nature*, **288**, 646-647.
48. Doolittle,W.F. and Sapienza,C. (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature*, **284**, 601-3.
49. Orgel,L.E. and Crick,F.H. (1980) Selfish DNA: The ultimate parasite. *Nature*, **284**, 604-7.
50. Gibb,E.A. and Edgell,D.R. (2010) Better late than early: Delayed translation of intron-encoded endonuclease I-TevI is required for efficient splicing of its host group I intron. *Mol Microbiol*, **78**, 35-46.

51. Jensen,R., Sprague,G.F. and Herskowitz,I. (1983) Regulation of yeast mating-type interconversion - feedback-control of ho gene-expression by the mating-type locus. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, **80**, 3035-3039.
52. Chevalier,B.S. and Stoddard,B.L. (2001) Homing endonucleases: Structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res*, **29**, 3757-74.
53. Sanmiguel,P. and Bennetzen,J.L. (1998) Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Annals of Botany*, **82**, 37-44.
54. Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W., et al. (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921.
55. Li,W.L., Zhang,P., Fellers,J.P., Friebe,B. and Gill,B.S. (2004) Sequence composition, organization, and evolution of the core triticeae genome. *Plant Journal*, **40**, 500-511.
56. Aziz,R.K., Breitbart,M. and Edwards,R.A. (2010) Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Res.*, **38**, 4207-4217.
57. Hughes,J.F. and Coffin,J.M. (2001) Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution. *Nat Genet*, **29**, 487-9.
58. Gilbert,N., Lutz-Prigge,S. and Moran,J.V. (2002) Genomic deletions created upon LINE-1 retrotransposition. *Cell*, **110**, 315-25.
59. Xing,J., Wang,H., Belancio,V.P., Cordaux,R., Deininger,P.L. and Batzer,M.A. (2006) Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 17608-17613.
60. Roberts,M.C. (1996) Tetracycline resistance determinants: Mechanisms of action, regulation of expression, genetic mobility, and distribution. *FEMS Microbiol Rev*, **19**, 1-24.
61. Whittle,G., Shoemaker,N.B. and Salyers,A.A. (2002) The role of bacteroides conjugative transposons in the dissemination of antibiotic resistance genes. *Cell Mol Life Sci*, **59**, 2044-54.
62. Alekshun,M.N. and Levy,S.B. (2007) Molecular mechanisms of antibacterial multidrug resistance. *Cell*, **128**, 1037-1050.
63. Friedrich,N.C., Torrents,E., Gibb,E.A., Sahlin,M., Sjöberg,B.M. and Edgell,D.R. (2007) Insertion of a homing endonuclease creates a genes-in-pieces ribonucleotide reductase that retains function. *Proc Natl Acad Sci U S A*, **104**, 6176-6181.

64. Moran,J.V., Zimmerly,S., Eskes,R., Kennell,J.C., Lambowitz,A.M., Butow,R.A. and Perlman,P.S. (1995) Mobile group II introns of yeast mitochondrial DNA are novel site-specific retroelements. *Mol Cell Biol*, **15**, 2828-38.
65. Moran,J.V., DeBerardinis,R.J. and H.,K.H.,Jr. (1999) Exon shuffling by L1 retrotransposition. *Science*, **283**, 1530-4.
66. Dujon,B. (1989) Group-I introns as mobile genetic elements - facts and mechanistic speculations - a review. *Gene*, **82**, 91-114.
67. Lambowitz,A.M. and Belfort,M. (1993) Introns as mobile genetic elements. *Annu Rev Biochem*, **62**, 587-622.
68. Belfort,M. and Perlman,P.S. (1995) Mechanisms of intron mobility. *J Biol Chem*, **270**, 30237-40.
69. Berg,D.E. and Howe,M.M. (1989) Mobile Dna.
70. Deininger,P.L., Moran,J.V., Batzer,M.A. and H.,K.H.,Jr. (2003) Mobile elements and mammalian genome evolution. *Curr Opin Genet Dev*, **13**, 651-8.
71. Hedges,D.J. and Deininger,P.L. (2007) Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutat. Res. - Fundam. Mol. Mech. Mutag.*, **616**, 46-59.
72. Singer,M.F. (1982) Sines and lines - highly repeated short and long interspersed sequences in mammalian genomes. *Cell*, **28**, 433-434.
73. Burt,A. and Koufopanou,V. (2004) Homing endonuclease genes: The rise and fall and rise again of a selfish element. *Curr. Opin. Genet. Dev.*, **14**, 609-615.
74. Siefert,J.L. (2009) Defining the mobilome. *Methods in Molecular Biology (Clifton, N. J)*, **532**, 13-27.
75. Berg,C.M., Berg,D.E. and Groisman,E.A. (1989) Transposable Elements and the Genetic Engineering of Bacteria.
76. Voigt,K., Izsvak,Z. and Ivics,Z. (2008) Targeted gene insertion for molecular medicine. *Journal of Molecular Medicine-Jmm*, **86**, 1205-1219.
77. Luft,F.C. (2010) Sleeping beauty jumps to new heights. *Journal of Molecular Medicine-Jmm*, **88**, 641-643.
78. Porteus,M.H. and Baltimore,D. (2003) Chimeric nucleases stimulate gene targeting in human cells. *Science*, **300**, 763.
79. Stoddard,B.L. (2011) Homing endonucleases: From microbial genetic invaders to reagents for targeted DNA modification. *Structure*, **19**, 7-15.

80. Stoddard,B.L. (2005) Homing endonuclease structure and function. *Q Rev Biophys*, **38**, 49-95.
81. Pingoud,A. and Jeltsch,A. (2001) Structure and function of type II restriction endonucleases. *Nucleic Acids Res*, **29**, 3705-27.
82. Kaus-Drobek,M., Czapinska,H., Sokolowska,M., Tamulaitis,G., Szczepanowski,R.H., Urbanke,C., Siksnys,V. and Bochtler,M. (2007) Restriction endonuclease MvaI is a monomer that recognizes its target sequence asymmetrically. *Nucleic Acids Res*, **35**, 2035-46.
83. Sasnauskas,G., Kostiuk,G., Tamulaitis,G. and Siksnys,V. (2011) Target site cleavage by the monomeric restriction enzyme BcnI requires translocation to a random DNA sequence and a switch in enzyme orientation. *Nucleic Acids Res.*, **39**, 8844-56.
84. Bitinaite,J., Wah,D.A., Aggarwal,A.K. and Schildkraut,I. (1998) FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci U S A*, **95**, 10570-5.
85. Jakubauskas,A., Sasnauskas,G., Giedriene,J. and Janulaitis,A. (2008) Domain organization and functional analysis of type IIS restriction endonuclease Eco31I. *Biochemistry (N. Y.)*, **47**, 8546-56.
86. Sasnauskas,G., Zakrys,L., Zaremba,M., Cosstick,R., Gaynor,J.W., Halford,S.E. and Siksnys,V. (2010) A novel mechanism for the scission of double-stranded DNA: BfiI cuts both 3'-5' and 5'-3' strands by rotating a single active site. *Nucleic Acids Res.*, **38**, 2399-410.
87. Kennedy,A.K., Guhathakurta,A., Kleckner,N. and Haniford,D.B. (1998) Tn10 transposition via a DNA hairpin intermediate. *Cell*, **95**, 125-34.
88. Coen,D., Deutsch,J., Netter,P., Petrochilo,E. and Slonimski,P.P. (1970) Mitochondrial genetics. I-methodology and phenomenology. *Symp. Soc. Exp. Biol.*, **24**, 449-496.
89. Dujon,B. (1980) Sequence of the intron and flanking exons of the mitochondrial 21s ribosomal-rna gene of yeast strains having different alleles at the omega-loci and rib-1-loci. *Cell*, **20**, 185-197.
90. Macreadie,I.G., Scott,R.M., Zinn,A.R. and Butow,R.A. (1985) Transposition of an intron in yeast mitochondria requires a protein encoded by that intron. *Cell*, **41**, 395-402.
91. Jacquier,A. and Dujon,B. (1985) An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell*, **41**, 383-94.
92. Szybalski,W., Kim,S.C., Hasan,N. and Podhajaska,A.J. (1991) Class-iis restriction enzymes - a review. *Gene*, **100**, 13-26.

93. Mucke,M., Kruger,D.H. and Reuter,M. (2003) Diversity of type II restriction endonucleases that require two DNA recognition sites. *Nucleic Acids Res.*, **31**, 6079-6084.
94. Taylor,G.K. and Stoddard,B.L. (2012) Structural, functional and evolutionary relationships between homing endonucleases and proteins from their host organisms. *Nucleic Acids Res.*, **40**, 5189-200.
95. Taylor,G.K., Heiter,D.F., Pietrokovski,S. and Stoddard,B.L. (2011) Activity, specificity and structure of I-Bth0305I: A representative of a new homing endonuclease family. *Nucleic Acids Res.*, **39**, 9705-9719.
96. Bonocora,R.P. and Shub,D.A. (2001) A novel group I intron-encoded endonuclease specific for the anticodon region of tRNA(fMet) genes. *Mol Microbiol*, **39**, 1299-306.
97. Bell-Pedersen,D., Quirk,S.M., Aubrey,M. and Belfort,M. (1989) A site-specific endonuclease and co-conversion of flanking exons associated with the mobile td intron of phage T4. *Gene*, **82**, 119-26.
98. Derbyshire,V., Kowalski,J.C., Dansereau,J.T., Hauer,C.R. and Belfort,M. (1997) Two-domain structure of the *td* intron-encoded endonuclease I-*TevI* correlates with the two-domain configuration of the homing site. *J Mol Biol*, **265**, 494-506.
99. Bell-Pedersen,D., Quirk,S.M., Bryk,M. and Belfort,M. (1991) I-*TevI*, the endonuclease encoded by the mobile *td* intron, recognizes binding and cleavage domains on its DNA target. *Proc Natl Acad Sci U S A*, **88**, 7719-23.
100. Bryk,M., Quirk,S.M., Mueller,J.E., Loizos,N., Lawrence,C. and Belfort,M. (1993) The *td* intron endonuclease I-*TevI* makes extensive sequence-tolerant contacts across the minor groove of its DNA target. *EMBO J*, **12**, 2141-9.
101. Bryk,M., Belisle,M., Mueller,J.E. and Belfort,M. (1995) Selection of a remote cleavage site by I-*TevI*, the *td* intron-encoded endonuclease. *J Mol Biol*, **247**, 197-210.
102. Edgell,D.R., Stanger,M.J. and Belfort,M. (2004) Coincidence of cleavage sites of intron endonuclease I-*TevI* and critical sequences of the host thymidylate synthase gene. *J Mol Biol*, **343**, 1231-1241.
103. Mueller,J.E., Smith,D., Bryk,M. and Belfort,M. (1995) Intron-encoded endonuclease I-*TevI* binds as a monomer to effect sequential cleavage via conformational changes in the *td* homing site. *EMBO J*, **14**, 5724-35.
104. Kowalski,J.C., Belfort,M., Stapleton,M.A., Holpert,M., Dansereau,J.T., Pietrokovski,S., Baxter,S.M. and Derbyshire,V. (1999) Configuration of the catalytic GIY-YIG domain of intron endonuclease I-*TevI*: Coincidence of computational and molecular findings. *Nucleic Acids Res*, **27**, 2115-25.

105. Van Roey,P., Waddling,C.A., Fox,K.M., Belfort,M. and Derbyshire,V. (2001) Intertwined structure of the DNA-binding domain of intron endonuclease I-TevI with its substrate. *EMBO J*, **20**, 3631-7.
106. Van Roey,P., Meehan,L., Kowalski,J.C., Belfort,M. and Derbyshire,V. (2002) Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI. *Nat Struct Biol*, **9**, 806-11.
107. Liu,Q., Dansereau,J.T., Puttamadappa,S.S., Shekhtman,A., Derbyshire,V. and Belfort,M. (2008) Role of the interdomain linker in distance determination for remote cleavage by homing endonuclease I-TevI. *J Mol Biol*, **379**, 1094-106.
108. Bujnicki,J.M., Rotkiewicz,P., Kolinski,A. and Rychlewski,L. (2001) Three-dimensional modeling of the I-TevI homing endonuclease catalytic domain, a GIY-YIG superfamily member, using NMR restraints and monte carlo dynamics. *Protein Eng.*, **14**, 717-721.
109. Shen,B.W., Landthaler,M., Shub,D.A. and Stoddard,B.L. (2004) DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J Mol Biol*, **342**, 43-56.
110. Sitbon,E. and Pietrokovski,S. (2003) New types of conserved sequence domains in DNA-binding regions of homing endonucleases. *Trends Biochem Sci*, **28**, 473-7.
111. Flick,K.E., Jurica,M.S., J.,M.R.,Jr and Stoddard,B.L. (1998) DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature*, **394**, 96-101.
112. Dean,A.B., Stanger,M.J., Dansereau,J.T., Van Roey,P., Derbyshire,V. and Belfort,M. (2002) Zinc finger as distance determinant in the flexible linker of intron endonuclease I-TevI. *Proc Natl Acad Sci U S A*, **99**, 8554-61.
113. Liu,Q., Derbyshire,V., Belfort,M. and Edgell,D.R. (2006) Distance determination by GIY-YIG intron endonucleases: Discrimination between repression and cleavage functions. *Nucleic Acids Res*, **34**, 1755-64.
114. Wu,W., Wood,D.W., Belfort,G., Derbyshire,V. and Belfort,M. (2002) Intein-mediated purification of cytotoxic endonuclease I-TevI by insertional inactivation and pH-controllable splicing. *Nucleic Acids Res*, **30**, 4864-71.
115. Edgell,D.R. and Shub,D.A. (2001) Related homing endonucleases I-BmoI and I-TevI use different strategies to cleave homologous recognition sites. *Proc Natl Acad Sci U S A*, **98**, 7898-903.
116. Edgell,D.R., Stanger,M.J. and Belfort,M. (2003) Importance of a single base pair for discrimination between intron-containing and intronless alleles by endonuclease I-BmoI. *Curr Biol*, **13**, 973-8.

117. Carter,J.M., Friedrich,N.C., Kleinstiver,B. and Edgell,D.R. (2007) Strand-specific contacts and divalent metal ion regulate double-strand break formation by the GIY-YIG homing endonuclease I-BmoI. *J Mol Biol*, **374**, 306-21.
118. Michel,F. and Dujon,B. (1986) Genetic exchanges between bacteriophage T4 and filamentous fungi? *Cell*, **46**, 323.
119. Dunin-Horkawicz,S., Feder,M. and Bujnicki,J.M. (2006) Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics*, **7**, 98.
120. Kleinstiver,B.P., Fernandes,A.D., Gloor,G.B. and Edgell,D.R. (2010) A unified genetic, computational and experimental framework identifies functionally relevant residues of the homing endonuclease I-BmoI. *Nucleic Acids Research*, **38**, 2411-27.
121. Sandegren,L. and Sjöberg,B.-. (2004) Distribution, sequence homology, and homing of group I introns among T-even-like bacteriophages: Evidence for recent transfer of old introns. *J Biol Chem*, **279**, 22218-27.
122. Nord,D. and Sjöberg,B.-. (2008) Unconventional GIY-YIG homing endonuclease encoded in group I introns in closely related strains of the bacillus cereus group. *Nucleic Acids Res*, **36**, 300-10.
123. Bae,H., Chung,I., Sim,N. and Cho,Y. (2012) Complete genome sequence of pseudomonas aeruginosa siphophage MP1412. *J. Virol.*, **86**, 9537-9537.
124. Bujnicki,J.M., Radlinska,M. and Rychlewski,L. (2001) Polyphyletic evolution of type II restriction enzymes revisited: Two independent sources of second-hand folds revealed. *Trends Biochem Sci*, **26**, 9-11.
125. Kaminska,K.H., Kawai,M., Boniecki,M., Kobayashi,I. and Bujnicki,J.M. (2008) Type II restriction endonuclease R.Hpy188I belongs to the GIY-YIG nuclease superfamily, but exhibits an unusual active site. *BMC Structural Biology*, **8**, 48.
126. Gasiunas,G., Sasnauskas,G., Tamulaitis,G., Urbanke,C., Razaniene,D. and Siksnys,V. (2008) Tetrameric restriction enzymes: Expansion to the GIY-YIG nuclease family. *Nucleic Acids Research*, **36**, 938-49.
127. Lyozin,G.T., Makarova,K.S., Velikodvorskaja,V.V., Zelentsova,H.S., Khechumian,R.R., Kidwell,M.G., Koonin,E.V. and Evgen'ev,M.B. (2001) The structure and evolution of penelope in the virilis species group of drosophila: An ancient lineage of retroelements. *Journal of Molecular Evolution*, **52**, 445-56.
128. Pyatkov,K.I., Arkhipova,I.R., Malkova,N.V., Finnegan,D.J. and Evgen'ev,M.B. (2004) Reverse transcriptase and endonuclease activities encoded by penelope-like retroelements. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 14719-24.

129. Koyama,T., Kondo,H., Aoki,T. and Hirono,I. (2013) Identification of two penelope-like elements with different structures and chromosome localization in kuruma shrimp genome. *Marine Biotechnology*, **15**, 115-123.
130. Carlson,K. and Wiberg,J.S. (1983) In vivo cleavage of cytosine-containing bacteriophage T4 DNA to genetically distinct, discretely sized fragments. *J Virol*, **48**, 18-30.
131. Aravind,L., Walker,D.R. and Koonin,E.V. (1999) Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res*, **27**, 1223-42.
132. Verhoeven,E.E., van Kesteren,M., Moolenaar,G.F., Visse,R. and Goosen,N. (2000) Catalytic sites for 3' and 5' incision of escherichia coli nucleotide excision repair are both located in UvrC. *J Biol Chem*, **275**, 5120-3.
133. Moolenaar,G.F., van Rossum-Fikkert,S., van Kesteren,M. and Goosen,N. (2002) Cho, a second endonuclease involved in escherichia coli nucleotide excision repair. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 1467-1472.
134. Van Houten,B., Eisen,J.A. and Hanawalt,P.C. (2002) A cut above: Discovery of an alternative excision repair pathway in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 2581-3.
135. Fricke,W.M. and Brill,S.J. (2003) Slx1-Slx4 is a second structure-specific endonuclease functionally redundant with Sgs1-Top3. *Genes Dev*, **17**, 1768-78.
136. Aravind,L. and Koonin,E.V. (2001) Prokaryotic homologs of the eukaryotic DNA-end-binding protein ku, novel domains in the ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res*, **11**, 1365-74.
137. Brachner,A., Braun,J., Ghodgaonkar,M., Castor,D., Zlopasa,L., Ehrlich,V., Jiricny,J., Gotzmann,J., Knasmueller,S. and Foisner,R. (2012) The endonuclease Ankle1 requires its LEM and GIY-YIG motifs for DNA cleavage in vivo. *J. Cell. Sci.*, **125**, 1048-1057.
138. Dittrich,C.M., Kratz,K., Sendoel,A., Gruenbaum,Y., Jiricny,J. and Hengartner,M.O. (2012) LEM-3-A LEM domain containing nuclease involved in the DNA damage response in *C. elegans*. *Plos One*, **7**, e24555.
139. Truglio,J.J., Rhau,B., Croteau,D.L., Wang,L., Skorvaga,M., Karakas,E., Dellavecchia,M.J., Wang,H., Van Houten,B. and Kisker,C. (2005) Structural insights into the first incision reaction during nucleotide excision repair. *EMBO J*, **24**, 885-94.
140. Karakas,E., Truglio,J.J., Croteau,D., Rhau,B., Wang,L., Van Houten,B. and Kisker,C. (2007) Structure of the C-terminal half of UvrC reveals an RNase H endonuclease domain with an argonaute-like catalytic triad. *The EMBO Journal*, **26**, 613-22.

141. Andersson,C.E., Lagerback,P. and Carlson,K. (2010) Structure of bacteriophage T4 endonuclease II mutant E118A, a tetrameric GIY-YIG enzyme. *Journal of Molecular Biology*, **397**, 1003-16.
142. Mak,A.N., Lambert,A.R. and Stoddard,B.L. (2010) Folding, DNA recognition, and function of GIY-YIG endonucleases: Crystal structures of R.Eco29kI. *Structure*, **18**, 1321-31.
143. Sokolowska,M., Czapinska,H. and Bochtler,M. (2010) Hpy188I-DNA pre- and post-cleavage complexes--snapshots of the GIY-YIG nuclease mediated catalysis. *Nucleic Acids Research*, **39**, 1554-64.
144. Ibryashkina,E.M., Sasnauskas,G., Solonin,A.S., Zakharova,M.V. and Siksnys,V. (2009) Oligomeric structure diversity within the GIY-YIG nuclease family. *Journal of Molecular Biology*, **387**, 10-6.
145. Lagerback,P., Andersson,E., Malmberg,C. and Carlson,K. (2009) Bacteriophage T4 endonuclease II, a promiscuous GIY-YIG nuclease, binds as a tetramer to two DNA substrates. *Nucleic Acids Res.*, **37**, 6174-83.
146. Kleinstiver,B.P., Berube-Janzen,W., Fernandes,A.D. and Edgell,D.R. (2011) Divalent metal ion differentially regulates the sequential nicking reactions of the GIY-YIG homing endonuclease I-BmoI. *Plos One*, **6**, e23804.
147. Dalgaard,J.Z., Moser,M.J., Hughey,R. and Mian,I.S. (1997) Statistical modeling, phylogenetic analysis and structure prediction of a protein splicing domain common to inteins and hedgehog proteins. *J Comput Biol*, **4**, 193-214.
148. Taylor,G.K., Petrucci,L.H., Lambert,A.R., Baxter,S.K., Jarjour,J. and Stoddard,B.L. (2012) LAHEDES: The LAGLIDADG homing endonuclease database and engineering server. *Nucleic Acids Res.*, **40**, W110-W116.
149. Jacoby,K., Metzger,M., Shen,B.W., Certo,M.T., Jarjour,J., Stoddard,B.L. and Scharenberg,A.M. (2012) Expanding LAGLIDADG endonuclease scaffold diversity by rapidly surveying evolutionary sequence space. *Nucleic Acids Res.*, **40**, 4954-4964.
150. Jurica,M.S., J.,M.R.,Jr and Stoddard,B.L. (1998) DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-CreI. *Mol Cell*, **2**, 469-76.
151. Silva,G.H., Dalgaard,J.Z., Belfort,M. and Van Roey,P. (1999) Crystal structure of the thermostable archaeal intron-encoded endonuclease I-DmoI. *J Mol Biol*, **286**, 1123-36.
152. Moure,C.M., Gimble,F.S. and Quijcho,F.A. (2003) The crystal structure of the gene targeting homing endonuclease I-SceI reveals the origins of its target site specificity. *J Mol Biol*, **334**, 685-95.

153. Bolduc, J.M., Spiegel, P.C., Chatterjee, P., Brady, K.L., Downing, M.E., Caprara, M.G., Waring, R.B. and Stoddard, B.L. (2003) Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing factor. *Genes Dev*, **17**, 2875-88.
154. Takeuchi, R., Lambert, A.R., Mak, A.N., Jacoby, K., Dickson, R.J., Gloor, G.B., Scharenberg, A.M., Edgell, D.R. and Stoddard, B.L. (2011) Tapping natural reservoirs of homing endonucleases for targeted gene modification. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 13077-13082.
155. Baxter, S., Lambert, A.R., Kuhar, R., Jarjour, J., Kulshina, N., Parmeggiani, F., Danaher, P., Gano, J., Baker, D., Stoddard, B.L., et al. (2012) Engineering domain fusion chimeras from I-OnuI family LAGLIDADG homing endonucleases. *Nucleic Acids Res.*, **40**, 7985-8000.
156. Kuhlmann, U.C., Moore, G.R., James, R., Kleanthous, C. and Hemmings, A.M. (1999) Structural parsimony in endonuclease active sites: Should the number of homing endonuclease families be redefined? *FEBS Lett*, **463**, 1-2.
157. Flick, K.E., McHugh, D., Heath, J.D., Stephens, K.M., J., M.R., Jr and Stoddard, B.L. (1997) Crystallization and preliminary X-ray studies of I-PpoI: A nuclear, intron-encoded homing endonuclease from *Physarum polycephalum*. *Protein Sci*, **6**, 2677-80.
158. Sokolowska, M., Czapinska, H. and Bochtler, M. (2009) Crystal structure of the -me type II restriction endonuclease Hpy99I with target DNA. *Nucleic Acids Res.*, **37**, 3799-3810.
159. Kleanthous, C., Kuhlmann, U.C., Pommer, A.J., Ferguson, N., Radford, S.E., Moore, G.R., James, R. and Hemmings, A.M. (1999) Structural and mechanistic basis of immunity toward endonuclease colicins. *Nat. Struct. Biol.*, **6**, 243-252.
160. Pommer, A.J., Cal, S., Keeble, A.H., Walker, D., Evans, S.J., Kuhlmann, U.C., Cooper, A., Connolly, B.A., Hemmings, A.M., Moore, G.R., et al. (2001) Mechanism and cleavage specificity of the H-N-H endonuclease colicin E9. *J Mol Biol*, **314**, 735-49.
161. Cheng, Y.S., Hsia, K.C., Doudeva, L.G., Chak, K.F. and Yuan, H.S. (2002) The crystal structure of the nuclease domain of colicin E7 suggests a mechanism for binding to double-stranded DNA by the H-N-H endonucleases. *J Mol Biol*, **324**, 227-36.
162. Mate, M.J. and Kleanthous, C. (2004) Structure-based analysis of the metal-dependent mechanism of H-N-H endonucleases. *J Biol Chem*, **279**, 34763-9.
163. Biertuempfel, C., Yang, W. and Suck, D. (2007) Crystal structure of T4 endonuclease VII resolving a holliday junction. *Nature*, **449**, 616-U14.

164. Dalgaard,J.Z., Klar,A.J., Moser,M.J., Holley,W.R., Chatterjee,A. and Mian,I.S. (1997) Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. *Nucleic Acids Res.*, **25**, 4626-4638.
165. Mehta,P., Katta,K. and Krishnaswamy,S. (2004) HNH family subclassification leads to identification of commonality in the his-me endonuclease superfamily. *Protein Science*, **13**, 295-300.
166. Johansen,S., Embley,T.M. and Willassen,N.P. (1993) A family of nuclear homing endonucleases. *Nucleic Acids Res*, **21**, 4405.
167. Johansen,S., Elde,M., Vader,A., Haugen,P., Haugli,K. and Haugli,F. (1997) In vivo mobility of a group I twintron in nuclear ribosomal DNA of the myxomycete *didymium iridis*. *Mol Microbiol*, **24**, 737-45.
168. Elde,M., Willassen,N.P. and Johansen,S. (2000) Functional characterization of isoschizomeric his-cys box homing endonucleases from *naegleria*. *Eur J Biochem*, **267**, 7257-66.
169. Galburt,E.A. and Jurica,M.S. (2005) His-cys box homing endonucleases. *Nucleic Acids and Molecular Biology*, **16**, 85-102.
170. Muscarella,D.E., Ellison,E.L., Ruoff,B.M. and Vogt,V.M. (1990) Characterization of I-ppo, an intron-encoded endonuclease that mediates homing of a group I intron in the ribosomal DNA of *physarum polycephalum*. *Mol Cell Biol*, **10**, 3386-96.
171. Wittmayer,P.K. and Raines,R.T. (1996) Substrate binding and turnover by the highly specific I-PpoI endonuclease. *Biochemistry (N. Y.)*, **35**, 1076-1083.
172. Galburt,E.A., Chevalier,B., Tang,W., Jurica,M.S., Flick,K.E., J.,M.R.,Jr and Stoddard,B.L. (1999) A novel endonuclease mechanism directly visualized for I-PpoI. *Nat Struct Biol*, **6**, 1096-9.
173. Galburt,E.A., Chadsey,M.S., Jurica,M.S., Chevalier,B.S., Erho,D., Tang,W., J.,M.R.,Jr and Stoddard,B.L. (2000) Conformational changes and cleavage by the homing endonuclease I-PpoI: A critical role for a leucine residue in the active site. *J Mol Biol*, **300**, 877-87.
174. Goodrich-Blair,H. and Shub,D.A. (1996) Beyond homing: Competition between intron endonucleases confers a selective advantage on flanking genetic markers. *Cell*, **84**, 211-21.
175. Landthaler,M. and Shub,D.A. (2003) The nicking homing endonuclease I-BasI is encoded by a group I intron in the DNA polymerase gene of the *Bacillus thuringiensis* phage *bastille*. *Nucleic Acids Res*, **31**, 3071-7.

176. Landthaler,M., Shen,B.W., Stoddard,B.L. and Shub,D.A. (2006) I-BasI and I-HmuI: Two phage intron-encoded endonucleases with homologous DNA recognition sequences but distinct DNA specificities. *J Mol Biol*, **358**, 1137-51.
177. Robbins,J.B., Stapleton,M., Stanger,M.J., Smith,D., Dansereau,J.T., Derbyshire,V. and Belfort,M. (2007) Homing endonuclease I-TevIII: Dimerization as a means to a double-strand break. *Nucleic Acids Res*, **35**, 1589-600.
178. Li,L., Wu,L.P. and Chandrasegaran,S. (1992) Functional domains in fok I restriction endonuclease. *Proc Natl Acad Sci U S A*, **89**, 4275-9.
179. Wah,D.A., Hirsch,J.A., Dorner,L.F., Schildkraut,I. and Aggarwal,A.K. (1997) Structure of the multimodular endonuclease FokI bound to DNA. *Nature*, **388**, 97-100.
180. Bath,A.J., Milsom,S.E., Gormley,N.A. and Halford,S.E. (2002) Many type IIs restriction endonucleases interact with two recognition sites before cleaving DNA. *J. Biol. Chem.*, **277**, 4024-4033.
181. Sugisaki,H. and Kanazawa,S. (1981) New restriction endonucleases from flavobacterium okeanokoites (FokI) and micrococcus luteus (MluI). *Gene*, **16**, 73-8.
182. Wah,D.A., Bitinaite,J., Schildkraut,I. and Aggarwal,A.K. (1998) Structure of FokI has implications for DNA cleavage. *Proc Natl Acad Sci U S A*, **95**, 10564-9.
183. Vanamee,E.S., Santagata,S. and Aggarwal,A.K. (2001) FokI requires two specific DNA sites for cleavage. *J. Mol. Biol.*, **309**, 69-78.
184. Kim,Y.G., Li,L. and Chandrasegaran,S. (1994) Insertion and deletion mutants of fokI restriction-endonuclease. *J. Biol. Chem.*, **269**, 31978-31982.
185. Perona,J.J. (2002) Type II restriction endonucleases. *Methods (San Diego, Calif)*, **28**, 353-64.
186. Christian,M., Cermak,T., Doyle,E.L., Schmidt,C., Zhang,F., Hummel,A., Bogdanove,A.J. and Voytas,D.F. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, **186**, 757-U476.
187. Hinnen,A., Hicks,J.B. and Fink,G.R. (1978) Transformation of yeast. *Proc. Natl. Acad. Sci. U. S. A.*, **75**, 1929-1933.
188. Orrweaver,T.L., Szostak,J.W. and Rothstein,R.J. (1981) Yeast transformation - a model system for the study of recombination. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, **78**, 6354-6358.
189. Rothstein,R.J. (1983) One-step gene disruption in yeast. *Meth. Enzymol.*, **101**, 202-211.

190. Orrweaver,T.L., Szostak,J.W. and Rothstein,R.J. (1983) Genetic applications of yeast transformation with linear and gapped plasmids. *Meth. Enzymol.*, **101**, 228-245.
191. Szostak,J.W., Orr-Weaver,T.L., Rothstein,R.J. and Stahl,F.W. (1983) The double-strand-break repair model for recombination. *Cell*, **33**, 25-35.
192. Thomas,K.R. and Capecchi,M.R. (1986) Introduction of homologous dna-sequences into mammalian-cells induces mutations in the cognate gene. *Nature*, **324**, 34-38.
193. Thomas,K.R., Folger,K.R. and Capecchi,M.R. (1986) High-frequency targeting of genes to specific sites in the mammalian genome. *Cell*, **44**, 419-428.
194. Capecchi,M. (1990) Gene targeting - how efficient can you get. *Nature*, **348**, 109.
195. Capecchi,M. (1990) Gene targeting - tapping the cellular telephone. *Nature*, **344**, 105.
196. Thomas,K.R. and Capecchi,M.R. (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem-cells. *Cell*, **51**, 503-512.
197. Haber,J.E. (2000) Partners and pathways - repairing a double-strand break. *Trends in Genetics*, **16**, 259-264.
198. Rouet,P., Smih,F. and Jasin,M. (1994) Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc Natl Acad Sci U S A*, **91**, 6064-8.
199. Carroll,D. (2004) Using nucleases to stimulate homologous recombination. *Methods in Molecular Biology (Clifton, N. J)*, **262**, 195-207.
200. Rouet,P., Smih,F. and Jasin,M. (1994) Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol*, **14**, 8096-106.
201. Smih,F., Rouet,P., Romanienko,P.J. and Jasin,M. (1995) Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Res*, **23**, 5012-9.
202. Choulika,A., Perrin,A., Dujon,B. and Nicolas,J.F. (1995) Induction of homologous recombination in mammalian chromosomes by using the I-scei system of *saccharomyces-cerevisiae*. *Mol. Cell. Biol.*, **15**, 1968-1973.
203. Chalberg,T.W., Portlock,J.L., Olivares,E.C., Thyagarajan,B., Kirby,P.J., Hillman,R.T., Hoelters,J. and Calos,M.P. (2006) Integration specificity of phage phi C31 integrase in the human genome. *J. Mol. Biol.*, **357**, 28-48.

204. Silva,G., Poirot,L., Galetto,R., Smith,J., Montoya,G., Duchateau,P. and Paques,F. (2011) Meganucleases and other tools for targeted genome engineering: Perspectives and challenges for gene therapy. *Current Gene Therapy*, **11**, 11-27.
205. Aiuti,A., Slavin,S., Aker,M., Ficara,F., Deola,S., Mortellaro,A., Morecki,S., Andolfi,G., Tabucchi,A., Carlucci,F., et al. (2002) Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. *Science*, **296**, 2410-2413.
206. Hacein-Bey-Abina,S., Von Kalle,C., Schmidt,M., McCormack,M.P., Wulffraat,N., Leboulch,P., Lim,A., Osborne,C.S., Pawliuk,R., Morillon,E., et al. (2003) LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science*, **302**, 415-419.
207. Gaspar,H.B., Parsley,K.L., Howe,S., King,D., Gilmour,K.C., Sinclair,J., Brouns,G., Schmidt,M., Von Kalle,C., Barington,T., et al. (2004) Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. *Lancet*, **364**, 2181-2187.
208. Fischer,A., Le Deist,F., Hacein-Bey-Abina,S., Andre-Schmutz,I., Basile,G.D., de Villartay,J.P. and Cavazzana-Calvo,M. (2005) Severe combined immunodeficiency. A model disease for molecular immunology and therapy. *Immunol. Rev.*, **203**, 98-109.
209. Cavazzana-Calvo,M., Lagresle,C., Hacein-Bey-Abina,S. and Fischer,A. (2005) Gene therapy for severe combined immunodeficiency. *Annu. Rev. Med.*, **56**, 585-602.
210. Logan,A.C., Lutzko,C. and Kohn,D.B. (2002) Advances in lentiviral vector design for gene- modification of hematopoietic stem cells. *Curr. Opin. Biotechnol.*, **13**, 429-436.
211. Nienhuis,A.W., Dunbar,C.E. and Sorrentino,B.P. (2006) Genotoxicity of retroviral integration in hematopoietic cells. *Molecular Therapy*, **13**, 1031-1049.
212. Ellis,J. (2005) Silencing and variegation of gammaretrovirus and lentivirus vectors. *Hum. Gene Ther.*, **16**, 1241-1246.
213. Urnov,F.D., Miller,J.C., Lee,Y.L., Beausejour,C.M., Rock,J.M., Augustus,S., Jamieson,A.C., Porteus,M.H., Gregory,P.D. and Holmes,M.C. (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, **435**, 646-651.
214. Porteus,M.H. and Carroll,D. (2005) Gene targeting using zinc finger nucleases. *Nature Biotechnology*, **23**, 967-73.

215. Alwin,S., Gere,M.B., Guhl,E., Effertz,K., Barbas,C.F., Segal,D.J., Weitzman,M.D. and Cathomen,T. (2005) Custom zinc-finger nucleases for use in human cells. *Molecular Therapy*, **12**, 610-617.
216. Pingoud,A. and Silva,G.H. (2007) Precision genome surgery. *Nat. Biotechnol.*, **25**, 743-744.
217. Cathomen,T. and Joung,J.K. (2008) Zinc-finger nucleases: The next generation emerges. *Molecular Therapy*, **16**, 1200-1207.
218. Joung,J.K. and Sander,J.D. (2013) INNOVATION TALENs: A widely applicable technology for targeted genome editing. *Nature Reviews Molecular Cell Biology*, **14**, 49-55.
219. Certo,M.T., Ryu,B.Y., Annis,J.E., Garibov,M., Jarjour,J., Rawlings,D.J. and Scharenberg,A.M. (2011) Tracking genome engineering outcome at individual DNA breakpoints. *Nature Methods*, **8**, 671-U102.
220. Certo,M.T., Gwiazda,K.S., Kuhar,R., Sather,B., Curinga,G., Mandt,T., Brault,M., Lambert,A.R., Baxter,S.K., Jacoby,K., et al. (2012) Coupling endonucleases with DNA end-processing enzymes to drive gene disruption. *Nature Methods*, **9**, 973-5.
221. Fonfara,I., Curth,U., Pingoud,A. and Wende,W. (2012) Creating highly specific nucleases by fusion of active restriction endonucleases and catalytically inactive homing endonucleases. *Nucleic Acids Res.*, **40**, 847-60.
222. Kleinstiver,B.P., Wolfs,J.M., Kolaczyk,T., Roberts,A.K., Hu,S.X. and Edgell,D.R. (2012) Monomeric site-specific nucleases for genome editing. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 8061-6.
223. Hwang,W.Y., Fu,Y., Reyon,D., Maeder,M.L., Tsai,S.Q., Sander,J.D., Peterson,R.T., Yeh,J.R. and Joung,J.K. (2013) Efficient genome editing in zebrafish using a CRISPR-cas system. *Nat. Biotechnol.*, **31**, 227-9.
224. Chandrasegaran,S. and Smith,J. (1999) Chimeric restriction enzymes: What is next? *Biol. Chem.*, **380**, 841-848.
225. Smith,J., Bibikova,M., Whitby,F.G., Reddy,A.R., Chandrasegaran,S. and Carroll,D. (2000) Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Research*, **28**, 3361-9.
226. Soellue,C., Pars,K., Cornu,T.I., Thibodeau-Beganny,S., Maeder,M.L., Joung,J.K., Heilbronn,R. and Cathomen,T. (2010) Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. *Nucleic Acids Res.*, **38**, 8269-8276.
227. Guo,J., Gaj,T. and Barbas,C.F.,III. (2010) Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. *J. Mol. Biol.*, **400**, 96-107.

228. Choo, Y. and Klug, A. (1994) Toward a code for the interactions of zinc fingers with dna - selection of randomized fingers displayed on phage. *Proc. Natl. Acad. Sci. U. S. A.*, **91**, 11163-11167.
229. Wolfe, S.A., Nekludova, L. and Pabo, C.O. (2000) DNA recognition by cys(2)his(2) zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 183-212.
230. Pabo, C.O., Peisach, E. and Grant, R.A. (2001) Design and selection of novel cys(2)his(2) zinc finger proteins. *Annu. Rev. Biochem.*, **70**, 313-340.
231. Hurt, J.A., Thibodeau, S.A., Hirsh, A.S., Pabo, C.O. and Joung, J.K. (2003) Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 12271-12276.
232. Moore, M., Klug, A. and Choo, Y. (2001) Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. *Proc. Natl. Acad. Sci. U. S. A.*, **98**, 1437-1441.
233. Miller, J.C., Holmes, M.C., Wang, J., Guschin, D.Y., Lee, Y., Rupniewski, I., Beausejour, C.M., Waite, A.J., Wang, N.S., Kim, K.A., et al. (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.*, **25**, 778-785.
234. Maeder, M.L., Thibodeau-Beganny, S., Osiak, A., Wright, D.A., Anthony, R.M., Eichinger, M., Jiang, T., Foley, J.E., Winfrey, R.J., Townsend, J.A., et al. (2008) Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell*, **31**, 294-301.
235. Sander, J.D., Dahlborg, E.J., Goodwin, M.J., Cade, L., Zhang, F., Cifuentes, D., Curtin, S.J., Blackburn, J.S., Thibodeau-Beganny, S., Qi, Y., et al. (2011) Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nature Methods*, **8**, 67-U94.
236. Bibikova, M., Beumer, K., Trautman, J.K. and Carroll, D. (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science*, **300**, 764.
237. Durai, S., Mani, M., Kandavelou, K., Wu, J., Porteus, M.H. and Chandrasegaran, S. (2005) Zinc finger nucleases: Custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Res.*, **33**, 5978-5990.
238. Carroll, D., Beumer, K.J., Morton, J.J., Bozas, A. and Trautman, J.K. (2008) Gene targeting in drosophila and caenorhabditis elegans with zinc-finger nucleases. *Methods in Molecular Biology (Clifton, N. J)*, **435**, 63-77.
239. Carroll, D. (2008) Progress and prospects: Zinc-finger nucleases as gene therapy agents. *Gene Therapy*, **15**, 1463-8.

240. Ramirez,C.L., Foley,J.E., Wright,D.A., Muller-Lerch,F., Rahman,S.H., Cornu,T.I., Winfrey,R.J., Sander,J.D., Fu,F., Townsend,J.A., et al. (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. *Nature Methods*, **5**, 374-375.
241. Jantz,D. and Berg,J.M. (2010) Probing the DNA-binding affinity and specificity of designed zinc finger proteins. *Biophys. J.*, **98**, 852-860.
242. Bogdanove,A.J. and Voytas,D.F. (2011) TAL effectors: Customizable proteins for DNA targeting. *Science*, **333**, 1843-1846.
243. Bogdanove,A.J., Schornack,S. and Lahaye,T. (2010) TAL effectors: Finding plant genes for disease and defense. *Curr. Opin. Plant Biol.*, **13**, 394-401.
244. Moscou,M.J. and Bogdanove,A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 1501-1501.
245. Boch,J., Scholze,H., Schornack,S., Landgraf,A., Hahn,S., Kay,S., Lahaye,T., Nickstadt,A. and Bonas,U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509-1512.
246. Mak,A.N., Bradley,P., Cernadas,R.A., Bogdanove,A.J. and Stoddard,B.L. (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science*, **335**, 716-9.
247. Deng,D., Yan,C., Pan,X., Mahfouz,M., Wang,J., Zhu,J., Shi,Y. and Yan,N. (2012) Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science*, **335**, 720-723.
248. Christian,M., Cermak,T., Doyle,E.L., Schmidt,C., Zhang,F., Hummel,A., Bogdanove,A.J. and Voytas,D.F. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, **186**, 757-U476.
249. Cermak,T., Doyle,E.L., Christian,M., Wang,L., Zhang,Y., Schmidt,C., Baller,J.A., Somia,N.V., Bogdanove,A.J. and Voytas,D.F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.*, **39**, e82.
250. Cade,L., Reyon,D., Hwang,W.Y., Tsai,S.Q., Patel,S., Khayter,C., Joung,J.K., Sander,J.D., Peterson,R.T. and Yeh,J.J. (2012) Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. *Nucleic Acids Res.*, **40**, 8001-8010.
251. Sung,Y.H., Baek,I., Kim,D.H., Jeon,J., Lee,J., Lee,K., Jeong,D., Kim,J. and Lee,H. (2013) Knockout mice created by TALEN-mediated gene targeting. *Nat. Biotechnol.*, **31**, 23-24.

252. Zhang,Y., Zhang,F., Li,X., Baller,J.A., Qi,Y., Starker,C.G., Bogdanove,A.J. and Voytas,D.F. (2013) Transcription activator-like effector nucleases enable efficient plant genome engineering. *Plant Physiol.*, **161**, 20-27.
253. Bedell,V.M., Wang,Y., Campbell,J.M., Poshusta,T.L., Starker,C.G., Krug,R.G.,II, Tan,W., Penheiter,S.G., Ma,A.C., Leung,A.Y.H., et al. (2012) In vivo genome editing using a high-efficiency TALEN system. *Nature*, **491**, 114-U133.
254. Carlson,D.F., Tan,W., Lillico,S.G., Stverakova,D., Proudfoot,C., Christian,M., Voytas,D.F., Long,C.R., Whitelaw,C.B.A. and Fahrenkrug,S.C. (2012) Efficient TALEN-mediated gene knockout in livestock. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 17382-17387.
255. Morbitzer,R., Elsaesser,J., Hausner,J. and Lahaye,T. (2011) Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res.*, **39**, 5790-5799.
256. Mussolino,C., Morbitzer,R., Luetge,F., Dannemann,N., Lahaye,T. and Cathomen,T. (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res.*, **39**, 9283-9293.
257. Reyon,D., Tsai,S.Q., Khayter,C., Foden,J.A., Sander,J.D. and Joung,J.K. (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotechnol.*, **30**, 460-5.
258. Schmid-Burgk,J.L., Schmidt,T., Kaiser,V., Hoening,K. and Hornung,V. (2013) A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat. Biotechnol.*, **31**, 76-81.
259. Arnould,S., Chames,P., Perez,C., Lacroix,E., Duclert,A., Epinat,J.C., Stricher,F., Petit,A.S., Patin,A., Guillier,S., et al. (2006) Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J Mol Biol*, **355**, 443-58.
260. Li,H., Ulge,U.Y., Hovde,B.T., Doyle,L.A. and Monnat,R.J., Jr. (2012) Comprehensive homing endonuclease target site specificity profiling reveals evolutionary constraints and enables genome engineering applications. *Nucleic Acids Res.*, **40**, 2587-2598.
261. Chevalier,B.S., Kortemme,T., Chadsey,M.S., Baker,D., Monnat,R.J. and Stoddard,B.L. (2002) Design, activity, and structure of a highly specific artificial endonuclease. *Mol Cell*, **10**, 895-905.
262. Epinat,J.C., Arnould,S., Chames,P., Rochaix,P., Desfontaines,D., Puzin,C., Patin,A., Zanghellini,A., Paques,F. and Lacroix,E. (2003) A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res*, **31**, 2952-62.

263. Smith,J., Grizot,S., Arnould,S., Duclert,A., Epinat,J.C., Chames,P., Prieto,J., Redondo,P., Blanco,F.J., Bravo,J., et al. (2006) A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res*, **34**, e149.
264. Volna,P., Jarjour,J., Baxter,S., Roffler,S.R., J.,M.R.,Jr, Stoddard,B.L. and Scharenberg,A.M. (2007) Flow cytometric analysis of DNA binding and cleavage by cell surface-displayed homing endonucleases. *Nucleic Acids Res*, **35**, 2748-58.
265. Ashworth,J., Taylor,G.K., Havranek,J.J., Quadri,S.A., Stoddard,B.L. and Baker,D. (2010) Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. *Nucleic Acids Res.*, **38**, 5601-5608.
266. Gao,H., Smith,J., Yang,M., Jones,S., Djukanovic,V., Nicholson,M.G., West,A., Bidney,D., Falco,S.C., Jantz,D., et al. (2010) Heritable targeted mutagenesis in maize using a designed endonuclease. *Plant Journal*, **61**, 176-187.
267. Windbichler,N., Menichelli,M., Papathanos,P.A., Thyme,S.B., Li,H., Ulge,U.Y., Hovde,B.T., Baker,D., Monnat,R.J.,Jr., Burt,A., et al. (2011) A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature*, **473**, 212-5.
268. Arnould,S., Perez,C., Cabaniols,J.P., Smith,J., Gouble,A., Grizot,S., Epinat,J.C., Duclert,A., Duchateau,P. and Paques,F. (2007) Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *J Mol Biol*, **371**, 49-65.
269. Grizot,S., Smith,J., Daboussi,F., Prieto,J., Redondo,P., Merino,N., Villate,M., Thomas,S., Lemaire,L., Montoya,G., et al. (2009) Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. *Nucleic Acids Research*, **37**, 5405-19.
270. Munoz,I.G., Prieto,J., Subramanian,S., Coloma,J., Redondo,P., Villate,M., Merino,N., Marenchino,M., D'Abramo,M., Gervasio,F.L., et al. (2011) Molecular basis of engineered meganuclease targeting of the endogenous human RAG1 locus. *Nucleic Acids Res.*, **39**, 729-743.
271. Prieto,J., Redondo,P., Padro,D., Arnould,S., Epinat,J., Paques,F., Blanco,F.J. and Montoya,G. (2007) 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.
272. Eastberg,J.H., McConnell Smith,A., Zhao,L., Ashworth,J., Shen,B.W. and Stoddard,B.L. (2007) Thermodynamics of DNA target site recognition by homing endonucleases. *Nucleic Acids Res*, **35**, 7209-21.

273. Jarjour,J., West-Foyle,H., Certo,M.T., Hubert,C.G., Doyle,L., Getz,M.M., Stoddard,B.L. and Scharenberg,A.M. (2009) High-resolution profiling of homing endonuclease binding and catalytic specificity using yeast surface display. *Nucleic Acids Res.*, **37**, 6871-6880.
274. Pattanayak,V., Ramirez,C.L., Joung,J.K. and Liu,D.R. (2011) Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nature Methods*, **8**, 765-U115.
275. Gabriel,R., Lombardo,A., Arens,A., Miller,J.C., Genovese,P., Kaepfel,C., Nowrouzi,A., Bartholomae,C.C., Wang,J., Friedman,G., et al. (2011) An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat. Biotechnol.*, **29**, 816-U72.
276. Chen,S., Oikonomou,G., Chiu,C.N., Niles,B.J., Liu,J., Lee,D.A., Antoshechkin,I. and Prober,D.A. (2013) A large-scale in vivo analysis reveals that TALENs are significantly more mutagenic than ZFNs generated using context-dependent assembly. *Nucleic Acids Res.*, **41**, 2769-78.
277. Schierling,B., Dannemann,N., Gabsalilow,L., Wende,W., Cathomen,T. and Pingoud,A. (2012) A novel zinc-finger nuclease platform with a sequence-specific cleavage module. *Nucleic Acids Res.*, **40**, 2623-38.
278. Mali,P., Yang,L., Esvelt,K.M., Aach,J., Guell,M., DiCarlo,J.E., Norville,J.E. and Church,G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823-826.
279. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A., et al. (2013) Multiplex genome engineering using CRISPR/cas systems. *Science*, **339**, 819-823.

Chapter 2

2 A unified genetic, computational and experimental framework identifies functionally relevant residues of the homing endonuclease I-BmoI

The work presented in this chapter is reproduced (with permission, Appendix S1) from:

Kleinstiver, B.P., Fernandes, A.D., Gloor, G.B., Edgell, D.R. (2010) A unified genetic, computational and experimental framework identifies functionally relevant residues of the homing endonuclease I-BmoI. *Nucleic Acids Research* 38: 2411-2427

2.1 Introduction

The explosion of sequence and structural data has rapidly accelerated the pace of protein structure and function studies. Bioinformatic approaches that predict function based on amino acid conservation (1,2), homology modelling studies (3), and identification of co-evolving residues (4,5) are among methods commonly used to address structure and function questions. There are, however, many protein families for which mechanistic insight is lacking. The GIY-YIG homing endonuclease family is one such example. Homing endonucleases are site-specific yet sequence-tolerant DNA endonucleases that are distinguishable from other DNA endonucleases in their ability to bind long target sequences and tolerate multiple substitutions within their binding site (6). They function primarily as mobile genetic elements, initiating the movement of their coding sequence and surrounding DNA by binding and cleaving a target site (the homing site) in genomes that lack the endonuclease (7). Homing endonucleases are phylogenetically widespread, and have traditionally been categorized into one of four large families based on conserved amino acid motifs, the LAGLIDADG, HNH, His-Cys box, and GIY-YIG families (6). The PE-(DE)-XK and Vsr-like enzymes are only recently described and have fewer family members (8,9). Much effort has been devoted towards re-engineering naturally occurring LAGLIDADG endonucleases to cleave novel target sequences with clinical relevance in the human genome (10-13). Similar studies could in principle be performed on any endonuclease family, necessitating a detailed understanding of mechanism.

Within the four largest endonuclease families, the GIY-YIG endonucleases are the least understood in terms of mechanism. The prototypical GIY-YIG family endonuclease is I-TevI, encoded with the genome of *Escherichia coli* phage T4 (14). Studies on I-TevI revealed that the enzyme has a two-domain structure, composed of a N-terminal catalytic domain containing the class-defining GIY-YIG motif that is connected to a C-terminal DNA-binding domain by a flexible linker (15). Substantial experimental evidence suggests that the DNA-binding domain tethers the catalytic domain on its substrate to perform two sequential nicking reactions that generate a staggered 2-nt 3' overhang (16). The catalytic domain has no measurable DNA-binding activity when expressed independently, and a critical function of the linker region in I-TevI, and its isoschizomer I-BmoI, is to correctly position the domain on substrate (17-19). Early bioinformatic studies revealed that the GIY-YIG domain is not exclusive to homing endonucleases (20), illustrated by the presence of the domain in the UvrC nucleotide excision repair protein (21), restriction enzymes (22,23), and retrotransposable elements (24). Structural, biochemical and bioinformatic studies have shown that the GIY-YIG domain is ~90 amino acids with an α/β -fold composed of a central three-stranded antiparallel β -sheet flanked by three helices (21,25) (Figure 2.1). Four highly conserved residues in the GIY-YIG domain, Y17, R27, E74, and N87 (numbered according to the I-BmoI sequence, Figure 1B) comprise a putative active site cleft (Figure 1C), with a single divalent metal ion coordinated by the glutamic acid residue in both the I-TevI and UvrC structures. Mutation of any of these residues abolishes DNA cleavage activity in a number of GIY-YIG enzymes (20-22,26,27).

In spite of a wealth of bioinformatic, biochemical, and structural data, the mechanism by which GIY-YIG homing endonucleases introduce a double-strand break in substrate is unknown (28). The mechanism must involve repositioning of a (presumably) single active site within the catalytic domain on substrate to perform two sequential nicking reactions, with the bottom (non-coding) strand nicked before the top (coding) strand (27,29). This mechanism is likely to be distinct from other enzymes that contain the GIY-YIG domain, including the restriction enzyme Cfr42I that functions as a tetramer (30), Eco29kI that functions as a dimer (31), or the UvrC proteins that nick only a single-strand

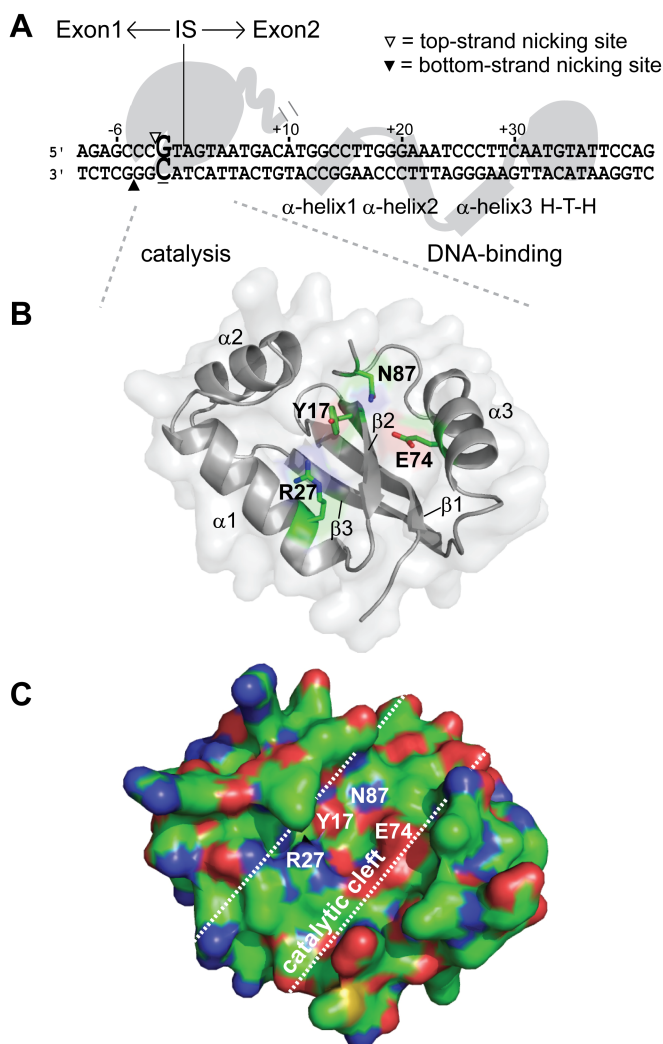


Figure 2.1: I-BmoI is a modular GIY-YIG homing endonuclease

(A) Schematic representation of I-BmoI interactions with intronless *thyA* substrate based on biochemical data (27,32). Top- and bottom-strand nicking sites are shown as open and filled triangles, respectively, and the critical -2 GC base pair is shown in enlarged, bold-type font. The intron insertion site is indicated by a vertical line, with exon 1 sequence upstream (-) and exon 2 sequence downstream (+). (B) Homology model of the I-BmoI catalytic domain (residues 1-88). Highlighted are four highly conserved residues in GIY-YIG alignments that are critical for function, and secondary structure elements of the domain. Subsequent illustrations of the catalytic domain will be shown from this view (front) or a 180-degree rotation (back). (C) Surface representation of a front view of the I-BmoI homology model highlighting the putative catalytic cleft. The sidechains of Y17, R27, E74, and N87 are surface exposed, lie along the base of the cleft, and are situated in close proximity to one another. Patches of charge are shown in color, blue being positive, red being negative, and green being hydrophobic.

adjacent to a damaged base (21). In an effort to gain insight into the mechanism by which GIY-YIG homing endonucleases introduce a double-strand break, we have been studying I-BmoI (Figure 2.1), an isoschizomer of I-TevI (32). Like I-TevI, I-BmoI is a two-domain endonuclease with an extended recognition sequence. Both enzymes cleave at the same positions within their respective intronless substrates, but I-BmoI requires only a critical G-C base pair at position -2 of intronless substrate for cleavage (33). As a model GIY-YIG homing endonuclease, I-BmoI has a number of advantages over I-TevI, including the fact that the wild-type (WT) enzyme can be overexpressed and purified in quantities that are difficult to obtain with I-TevI. Moreover, I-BmoI is ~750-fold less active than I-TevI, suggesting that early steps in the reaction pathway are more amenable to *in vitro* analysis (27,33).

Here, we present a unified experimental framework that will provide a platform on which to base future structure and function studies of GIY-YIG homing endonucleases, and other GIY-YIG-containing enzymes. Our framework, which we term MUSE, synthesizes data from three distinct experimental approaches; mutual information analyses that identify co-evolving residues in the GIY-YIG domain, a unigenic evolution strategy that uses a functional genetic selection to identify hypo- and hyper-mutable residues, and interpretation of the data using paralog-specific sequence alignments and structural models of the GIY-YIG domain. While none of the approaches used in our study are individually novel, the synthesis of data from all three methods facilitated the identification of residues that are unlikely to have been identified as important for function using any one of the approaches in isolation. Mutational analyses of the positions revealed phenotypic differences relative to WT I-BmoI in functional assays, validating that MUSE can successfully identify previously unrecognized residues with the GIY-YIG domain as relevant for function.

2.2 Materials and methods

2.2.1 Strain and plasmid construction

Strains and plasmids used in this study are listed in Supplementary Table S2.1, and oligonucleotides are listed in Supplementary Table S2.2. To construct strain

BW25141(λ DE3) for use in unigenic evolution experiments, *E.coli* BW25141 was lysogenized using the λ DE3 lysogenization kit (Novagen). The toxic plasmid backbone, p11-lacY-wtx1 (38), was used to construct pToxBmoHS and pToxBmoIn⁺ by inserting the corresponding intronless homing site (HS) and intron-containing target site (In⁺), respectively. To construct pToxBmoHS, oligonucleotides DE-395 and DE-396 were annealed and ligated into the XbaI and SphI sites of p11-lacY-wtx1. The 51bp intronless homing site corresponds to positions -10 to +41 of the exon1-exon2 junction relative to the intron insertion site (32). To construct pToxBmoIn⁺, oligonucleotides DE-429 and DE-430 were annealed and ligated into the XbaI and SphI sites of p11-lacY-wtx1. The 51 bp intron-containing site corresponds to the final 10 bp of the 3' end of the intron plus the first 41 bp of the 5' end of exon 2 (32). pIBmoI^E, which is a pUC57 derivative containing a codon optimized I-BmoI gene (IDT DNA), was used as a template for cloning the I-BmoI gene (optimized I-BmoI sequence is presented in Supplementary Figure S2.1). Primers DE-331 and DE-384 were used to amplify and clone the codon optimized I-BmoI into the NdeI and XhoI sites of pACYCDuet-1 to generate pACYCIBmoI. This plasmid was subsequently used as a template to generate pACYCR27A using the Quikchange XL site-directed mutagenesis kit (Stratagene) with primers DE-419 and its reverse complement, DE-420. All plasmid constructs were verified by sequencing.

2.2.2 Genetic selection

To generate strains harbouring the toxic plasmid for unigenic evolution, BW25141(λ DE3) was transformed with one of the 3 toxic (reporter) plasmids (p11-lacY-wtx1, pToxBmoHS, or pToxBmoIn⁺) and plated on LB plates containing 100 μ g/ml ampicillin and 0.2% glucose. For each strain, a single colony was picked to inoculate 500 ml LB plus 100 μ g/ml ampicillin and 0.2% glucose to generate electrocompetent cells. Typically, 50 μ l of electrocompetent cells were transformed with 100 ng of the expression plasmid (pACYCIBmoI or pACYCR27A). The transformations were allowed to recover in 500 μ l of SOC media at 37°C for 5 minutes, then diluted into 2 ml 37°C SOC and shaken at 37°C for 75 minutes. We found that addition of IPTG was not necessary to induce I-BmoI expression, as an IPTG concentration of 0.1 mM led to toxic effects. After incubation, transformations were diluted 1000-fold in SOC, and 100- μ l

aliquots were spread on plates containing LB plus 25 µg/ml chloramphenicol to estimate number of transformants, or plates containing LB plus 25 µg/ml chloramphenicol and 10 mM arabinose to observe the number of colonies surviving the selection. Survival rate was calculated by dividing the number of colonies observed on chloramphenicol plus arabinose plates by colonies observed on chloramphenicol only plates.

2.2.3 Construction of mutagenized I-BmoI libraries

I-BmoI mutant libraries were generated by error-prone PCR from pACYCIBmoI using primers DE-490 and DE-491. The forward primer (DE-490) was designed such that only the ATG start codon was included in the primer, and the reverse primer (DE-491) was designed such that no part of the I-BmoI gene was present in the primer. Three mutagenic libraries were generated using identical PCR conditions in parallel 50-µl reactions containing 80 ng of pACYCIBmoI as template, 20 pmol of each primer (DE-490, DE-491), 0.2 mM dATP, 0.2 mM dGTP, 1 mM dCTP, 1 mM dTTP, 0.5 mM MgCl₂, 0.5 mM MnCl₂, and 2.5 units of Taq polymerase (NEB) in the presence of 1x PCR buffer (10 mM KCl, 10 mM (NH₄)₂SO₄, 20 mM Tris-HCl pH 8.8, 2 mM MgSO₄, 0.1% Triton X-100). A total of 30 PCR cycles were run as follows: 94°C for 60s, 46.5°C for 60s, and 72°C for 60s. Mutagenic PCR products were digested with NdeI and XhoI and ligated into pACYCR27A (used as a ligation target due to the fact that re-ligated singly cut R27A I-BmoI would be non-functional in the selection). The ligated pools were independently transformed into DH5α, grown in 3 ml LB plus 25 µg/ml chloramphenicol for 16 hours at 37°C, and minipreped (QIAGEN) to generate the mutant I-BmoI libraries.

2.2.4 I-BmoI unigenic evolution and selection of variants

The three mutagenic libraries were subjected to unigenic evolution to determine survival percentage and to obtain clones required for sequence analysis. We sequenced a total of 167 selected clones picked from LB plus chloramphenicol and arabinose plates, of which 87 independent clones (36, 34, and 17 clones from pools 1, 2, and 3 respectively) were identified (the rest discarded due to redundancy of DNA or amino acid sequence). These clones contained a total of 460 nucleotide substitutions corresponding to 271 amino acid substitutions. We also sequenced 62 unselected clones from LB plus chloramphenicol

only plates. The unselected clones harboured a total of 760 nucleotide substitutions, 577 amino acid substitutions, and were used to establish baseline mutation frequencies. The EoS value was calculated as described (40).

2.2.5 Construction and purification of site-directed mutants

We created a library of site-directed mutants using the Quikchange® XL Site-Directed Mutagenesis Kit (Stratagene) to generate point mutants in the pACYCIBmoI backbone. For purification purposes, a subset of these mutants were sub-cloned into the pTYB1vector and were expressed and purified as previously described with one change to the protocol (27). Once the clarified lysate has been loaded, the five column volume wash with buffer A now contains a final concentration of 1mM ATP (Bioshop Canada Inc.) to help remove bound chaperones. The concentrations of purified WT I-BmoI and I-BmoI mutants were determined by a standard Bradford assay in duplicate using an Ultrospec 2100 pro (Biochrom Ltd).

2.2.6 Characterization of I-BmoI variants

A set of I-BmoI variants identified from the unigenic evolution study and the library of site-directed mutants were run through the genetic selection (as described above) to determine their survival versus WT. Cleavage assays were subsequently performed with I-BmoI mutants that were amenable to purification. The cleavage activities of WT and I-BmoI mutants were determined using titrations with 10 nM pBmoHS and 2-fold serial dilutions of I-BmoI from 700 nM to 1.37 nM in 10- μ l volumes for 5 min at 37°C in 50 mM Tris pH 7.9, 50 mM NaCl, 2 mM MgCl₂, and 1 mM DTT. Reactions were stopped by addition of 4 μ l stop dye (100 mM EDTA pH 8.0, 25% glycerol, and 0.2% bromophenol blue) and heated at 90°C for 5 minutes. Stopped reaction were run on 1% agarose gels, stained with ethidium bromide (OmniPur) and analyzed on an AlphaImager™3400 (Alpha Innotech). Percent linear DNA is defined as the percentage of total DNA (circular, nicked, and linear) converted to linear product. Nicking assays were conducted as above, with a standard protein concentration of 175 nM and 10- μ l aliquots of a reaction pool were stopped at 15 s, 30 s, 45 s, 1 min, 1min 30 s, 2 min, 3

min, and 5 min. The rate constants for the first strand (circular to nicked) and second strand (nicked to linear) steps were calculated as described in Supplementary Figure S2.3.

2.2.7 Structure based alignment

Sequences similar to the endonuclease domain of I-TevI (GI: 29345254) were identified by BLAST in the nonredundant protein database and in the metagenomic protein database at NCBI. All proteins with an E value less than 0.1 were aligned to the I-TevI catalytic domain structure (PDB ID: 1LN0) using the block align feature of Cn3D (<http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml>). The defaults of this feature were changed to not perform global alignments and to allow changes to the block structure of existing alignments. Long gaps were not allowed. One hundred and seventy-four full-length sequences were identified. Partial sequences, sequences that required large gaps or that were missing the presumed catalytic residues (R27, E75, and N90, I-TevI numbering) were excluded from the alignment. The alignment included the I-BmoI sequence (GI: 12958590). The alignment contained 4 bacteriophage sequences, 97 sequences from the marine metagenome, 23 bacterial sequences and 50 eukaryotic sequences. The sequences were culled to remove the shorter member of a redundant pair with a threshold of 90% using JalView (49), leaving 146 sequences. To test the quality of the alignment and for contamination by paralogous GIY-YIG sequences, the UvrC sequence from *Thermotoga maritima* (GI: 8134799) was included in the initial alignment. Using the Cn3D block align method, the sequences were sorted by matches to the position sensitive scoring matrix. The *T. maritima* UvrC sequence was the third worst scoring protein by this measure, suggesting that UvrC proteins were effectively excluded from the alignment. Second, we examined the alignment for an excess of local covariation that can detect as little as 10% contamination of the alignment (35). This often results when paralogous gene families are incorporated into one alignment. No excess local covariation was identified, suggesting that contamination by UvrC or other paralogous GIY-YIG type endonuclease domains was rare. Based on the alignment, we built a homology model of the I-BmoI GIY-YIG domain (residues 1-88) using MODELLER (50) and SWISS-MODEL (51) with the I-TevI structures (1MK0 and

1LN0) as templates. There was no significant difference between the two structural models.

2.2.8 Co-variation analyses

Three semi-independent methods, *Mlp* (34), ΔZp and ΔZpx (35), were applied to the multiple sequence alignment to identify pairs or small groups of positions that showed non-independent evolution in the GIY-YIG domain, using scoring cutoffs of 4.5 (*Mlp*), 3.5 (ΔZpx) and 1.5 (ΔZp), respectively. Mutual information values were assigned to residues using the I-BmoI reference sequence. Distances were calculated using the closest non-hydrogen atom of co-evolving residues in the I-BmoI homology model.

2.3 Results

2.3.1 Improved alignment of the GIY-YIG domain

Previous multiple sequence alignments of the GIY-YIG domain have included sequences of proteins with diverse functions, diluting the phylogenetic signal from GIY-YIG domains specific to homing endonucleases (20,23). To gain better insight into residues that are conserved amongst potential GIY-YIG homing endonucleases, we collected sequences by BLAST and aligned them to the I-TevI catalytic domain structure with Cn3D. BLAST searches were dominated by matches to the nucleotide excision repair protein UvrC, all of which were subsequently discarded. In addition, sequences that contained obvious insertions and deletions, or sequences that lacked residues equivalent to the functionally critical R27, E74 and N87 (I-BmoI numbering), were removed resulting in a final alignment of 146 sequences.

The GIY-YIG domain has previously been separated into 5 conserved regions that are characterized by highly conserved residues, termed motifs A through E (20). In our new alignment (Figure 2.2A), the information content associated with the highly conserved residues has not changed significantly, with the exceptions of increases associated with the tyrosine residues of motif A (Y6 and Y17) and the phenylalanine residue of motif C (F57). There are, however, differences outside of the highly conserved residues that are apparent in our alignment. In particular, there are obvious increases in the information

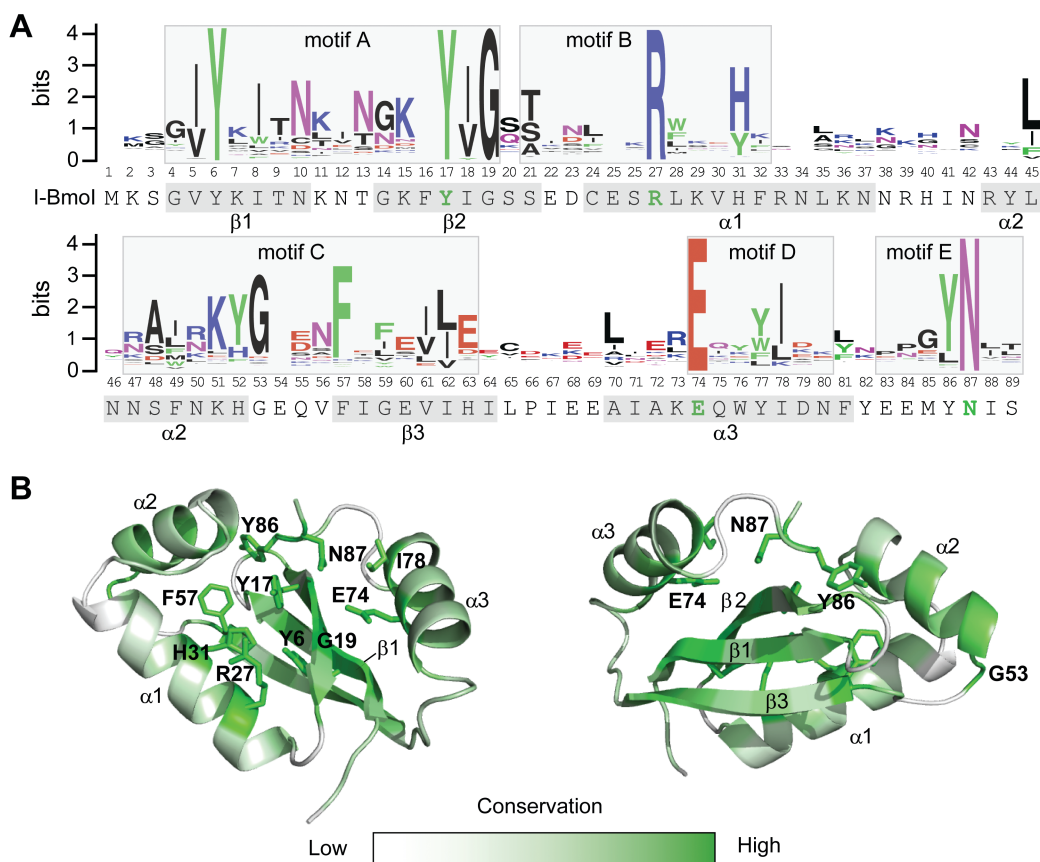


Figure 2.2: Alignment of the GIY-YIG domain

(A) Multiple sequence alignment of 146 sequences represented as a sequence logo (52). Positions are numbered according to the I-BmoI sequence that is shown below the alignment (conserved functionally critical residues are shown in green). Predicted secondary structure elements of the I-BmoI GIY-YIG domain are indicated on the sequence by shading, and motif assignments are identified on the alignment with shaded boxes. (B) Front (left) and rear (right) views of the homology model with the degree of conservation mapped onto the structure. Conserved positions are shown in dark green, with the side chains of highly conserved residues indicated. Variable positions are shown in white.

content of positions that intervene the GIY and YIG elements of motif A (I8, N10, and K15), and the stretch of residues (S48, K51, H52, G53) that precede the phenylalanine (F57) of motif C. In contrast, there is a decrease in the information content of the proline (P84) that precedes the conserved tyrosine (Y86) and asparagine (N87) residues of motif E (Figure 2.2A), attributable to the removal of UvrC sequences from the alignment that predominantly possess a proline at this position.

To display sequence conservation within the GIY-YIG domain, the entropy for each position in the alignment was calculated and subsequently displayed on the I-BmoI homology model, with conserved positions (low entropy, high information content) coloured green and non-conserved positions (high entropy, little information content) coloured white (Figure 2.2B). As expected, the highly conserved residues that characterize the GIY-YIG domain (Y6, Y17, R27, E74, and N87) all surround the proposed catalytic surface. Of particular interest is H31 that displays high information content, and lies in close proximity to the essential R27 in α -helix1, with both residues stacked against each other and orientated towards the presumed catalytic surface (Figure 2.2B). In general, the majority of conserved residues are similarly positioned, with the exception of a block of moderately conserved residues spanning positions 47-53 (located in α -helix 2, orientated away from the catalytic surface). Strikingly, many of the non-conserved residues occupy positions distinct from the proposed catalytic surface, including residues in α -helix 1 and the loop connecting α -helix 1 and α -helix 2. Collectively, these data highlight the conservation of residues that surround the proposed catalytic surface of the GIY-YIG domain, whereas non-conserved residues tend to be on the periphery of the domain.

2.3.2 Identification of co-evolving positions in the I-Bmol catalytic domain

Additional insight into functionally important residues of the I-BmoI catalytic domain was obtained by applying mutual information analyses to detect amino acid positions that co-evolve, or co-vary, with each other. Such analyses can detect non-independent evolution of residues that co-vary with other residues because of functional constraints. Indeed, most residues that co-vary tend to be within contact distance of each other (4,34).

We applied three different methods to analyze co-variation in the GIY-YIG domain alignment, Mip , ΔZpx , and ΔZp (35), and found that the highest scoring amino acid pair by all methods was S20-I71 (Table 2.1). While S20 and I71 are not within predicted contact distance (6.3Å) in the homology model of the I-BmoI GIY-YIG domain, they occupy a surface of the domain that can be envisioned as a gateway to the catalytic cleft formed by the C-terminal end of β -sheet 2 and the N-terminal end of α -helix 3 (Figure 2.3). In addition, we identified a set of four residues, L35-H40-N46-F49, that co-evolve with each other (Table 2.1). This set of residues was of interest because the H40Y mutation in I-TevI reduces catalytic activity relative to the WT protein (36), and the interaction between these residues may be required to position the H40 residue appropriately within the active site (Figure 2.3). Another intriguing residue identified by mutual information analyses was K7 that coevolves with three other residues, T9, F16 and E60 (Table 2.1). All four residues lie in adjacently positioned β -sheets (Figure 2.3), suggesting that interaction between these residues is critical for folding of the GIY-YIG domain. Similarly, other sets of coevolving residues such as K51-H52, and H63-K73-W76 (Table 2.1), are all within contact distance of each other and likely have roles in folding or stability of the domain.

2.3.3 Unigenic evolution identifies mutable positions in I-BmoI

The above results provided a phylogenetic and structural framework for the identification and analysis of potential functionally critical residues in the catalytic domain of I-BmoI. To gain experimental insight into residues that are functionally relevant across the full length of I-BmoI, we adapted the unigenic evolution method (37). In this method, a genetic selection is used to isolate functional variants of I-BmoI after random mutagenesis, facilitating identification of amino acid positions that were either tolerant (hypermutable) or intolerant (hypomutable) of substitutions. The selection utilizes a two-plasmid system that relies on I-BmoI expression from one plasmid (pExp) to cleave a second, toxic plasmid (pTox) that contains the cognate I-BmoI homing site (38) (Figure 2.4). Cells survive plating on selective media only if the toxic plasmid has been cleaved by a functional I-BmoI. As shown in Figure 2.4, we observed a survival ratio of 0.95 when WT I-BmoI was expressed from pExp, and pTox contained the I-BmoI homing site

Table 2.1: Co-evolving residues in the I-BmoI GIY-YIG domain

Pair	Mlp	ΔZ_{px}	ΔZ_p	Distance (Å)
S20-I71	8	5.7	5.5	6.3
H63-K73	4.8	3.8	3.5	3.1
K51-H52	4.7	4.3	2	1.3
K7-E60	4.1	3.6	2	3
L35-N46	5.2	3.4	1.8	3.7
K7-F16	3.9	3.7	1.8	4.2
K73-W76	6	4.6	1.8	3.3
K7-T9	3.2	3	1.8	3.8
Q75-D79	3.8	4.1	1.8	3.1
L35-F49	3.5	3.2	1.7	3.6
N10-S48	4.6	3.9	1.6	3
L35-H40	4.7	3.4	1.5	3.5

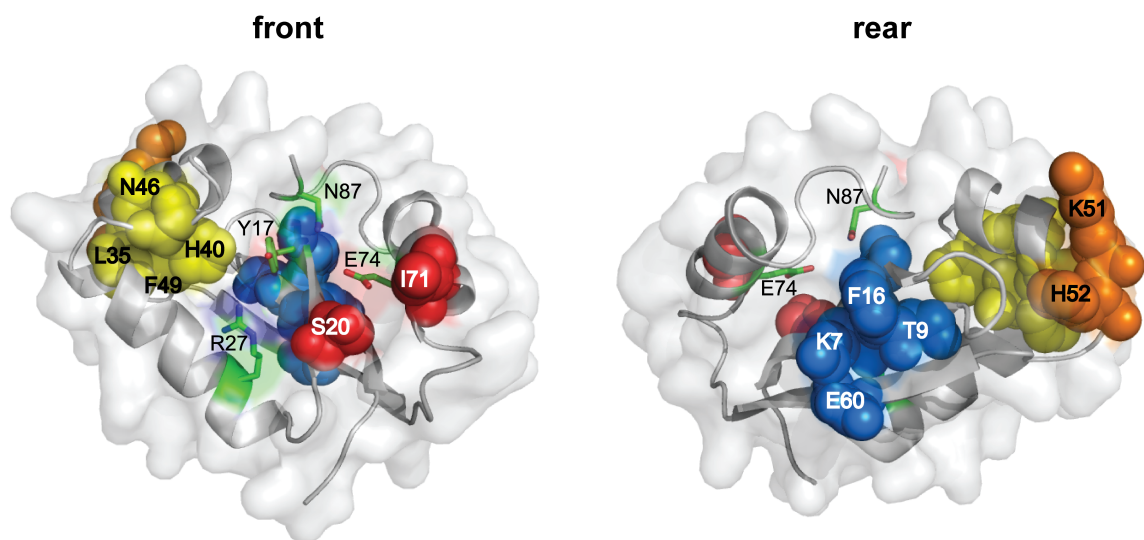


Figure 2.3: Co-evolving residues of the I-Bmol catalytic domain

Position of four sets of co-evolving residues mapped onto the I-Bmol homology model, color-coded by co-evolving residues (yellow, L35, H40, N46, F49; red S20, I71; blue K7, T9, F16, E60; orange K51, H52). Front (left) and rear (right) views are shown, with functionally critical residues highlighted in light green.

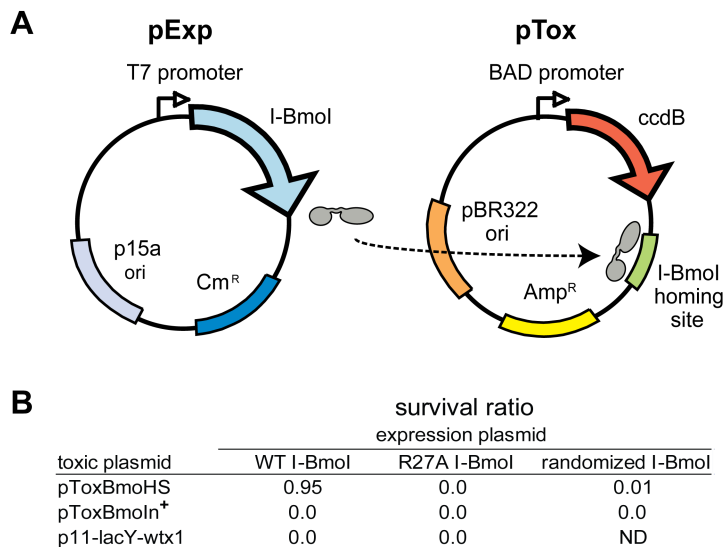


Figure 2.4: The two-plasmid genetic selection

(A) Schematic of the expression plasmid (pExp) and toxic (reporter) plasmid (pTox). (B) Verification of the genetic selection using variants of pExp and pTox. Survival rates are expressed as the ratio of colonies on chloramphenicol + arabinose plates to colonies on chloramphenicol only plates. WT I-Bmol, pExp expressing WT I-Bmol; R27A I-Bmol, pExp expressing an inactive R27A I-Bmol; randomized I-Bmol, library of I-Bmol variants; p11-lacY-wtx1, parental pTox without I-Bmol target site; ND, not determined.

(pToxBmoHS). In contrast, survival ratios of 0 were observed when WT I-BmoI was used with the pTox backbone (p11-lacY-wtx1) containing no homing site, and when a catalytically inactive I-BmoI variant, R27A, was used in combination with pToxBmoHS. Furthermore, the survival ratio was <0.0001 when pTox contained the intron-containing I-BmoI substrate (pToxBmoIn⁺), which I-BmoI cleaves poorly (33). These results validate the genetic selection system, showing that cells survive only when I-BmoI can cleave the toxic plasmid that contains the I-BmoI homing site.

We used error-prone PCR to generate three independent libraries of I-BmoI variants in pExp, mutagenized over the entire I-BmoI coding region. The libraries were transformed into the selection strain carrying pToxBmoHS, and survivors identified by plating on selective media with an average survival ratio of ~ 0.01 for the three pools. We identified and sequenced 87 independent I-BmoI variants that survived the selection, containing an average of 5.28 nucleotide substitutions (or 3.11 amino acid substitutions) per clone (Supplementary Figure S2.2). To determine the mutation frequencies inherent to the error-prone PCR, we sequenced 62 independent clones plated on non-selective media, which contained an average of 12.26 nucleotide substitutions per clone (Supplementary Figure S2.2). As expected, the average number of substitutions in the pool of unselected clones was much greater than the number in the selected pool, and the distribution of the number of changes between the selected and unselected clones differed.

By mapping the mutable positions from the selected clones onto the I-BmoI sequence, we found that amino acid positions tolerant to substitutions were distributed throughout the length of the coding region (Figure 2.5). To gain further insight into tolerated mutations within the GIY-YIG catalytic domain, the mutable residues from positions 1-88 were mapped onto a homology-based 3D model of the I-BmoI catalytic domain (Figure 2.6). Interestingly, the majority of the mutations mapped to the periphery or the backside of the catalytic domain (opposite the proposed catalytic surface). Also, few mutations were observed near the proposed catalytic surface and no substitutions were observed in the four functionally critical residues (Y17, R27, E74, and N87). Some positions tolerated the same substitution more than once, as evidenced by the appearance of the same mutation in independent clones that contained an alternate set of accompanying mutations (Table

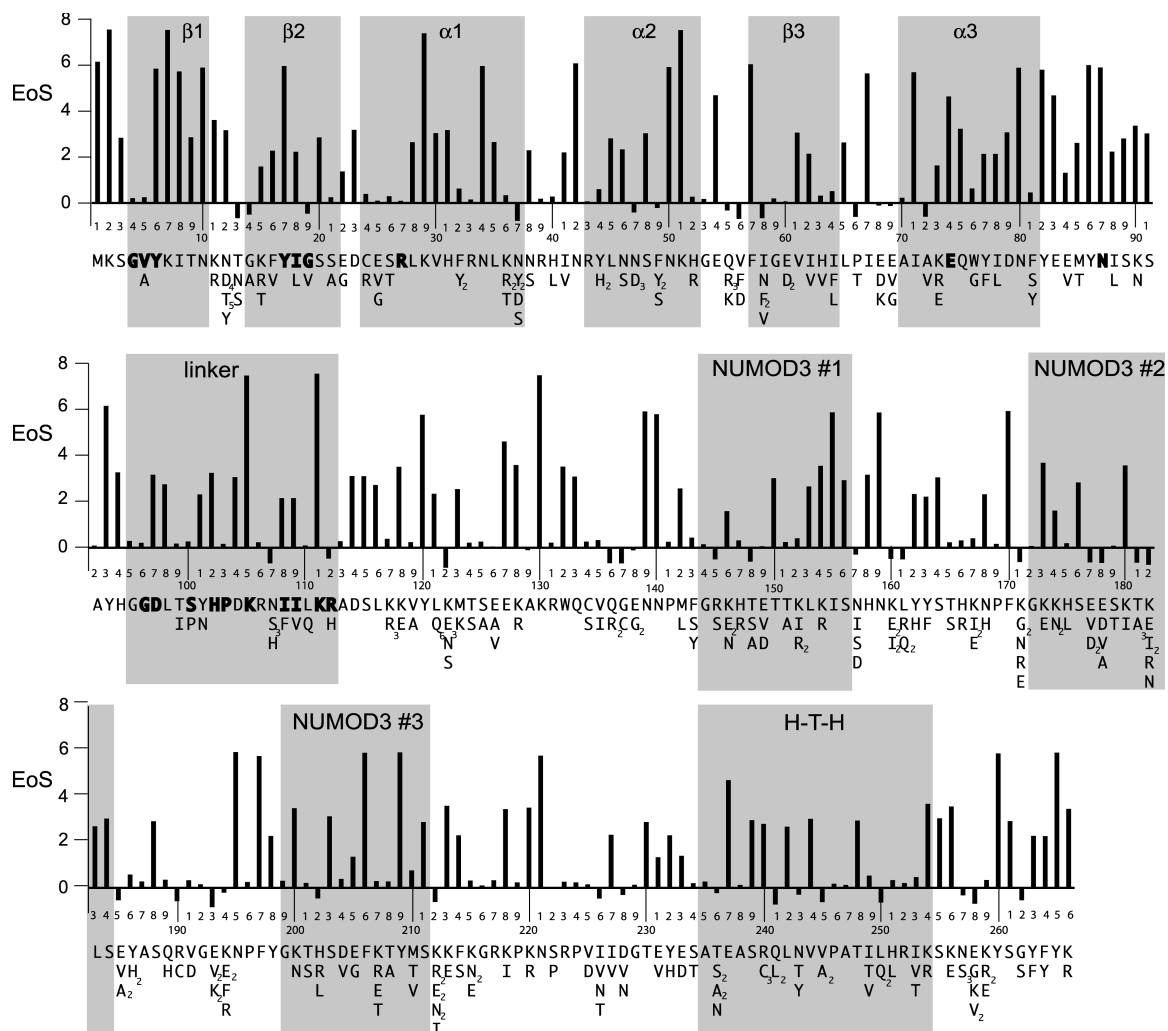


Figure 2.5: Unigenic evolution analysis of I-BmoI

Shown is a summary of the mutations found in the 87 selected clones and the EoS value for each position. Non-synonymous substitutions found at each position are indicated beneath the corresponding I-BmoI sequence over the entire length of the protein (multiple independent occurrences of the same mutation are shown in subscript). Shown above each line of sequence is a graph of the evidence of selection (EoS) at each amino acid position of I-BmoI. Regions of modelled or predicted secondary structure are indicated by grey rectangles, and bold residues indicate the GIY and YIG motifs, functionally critical residues, or residues that are identical between I-BmoI and I-TevI in the linker domain.

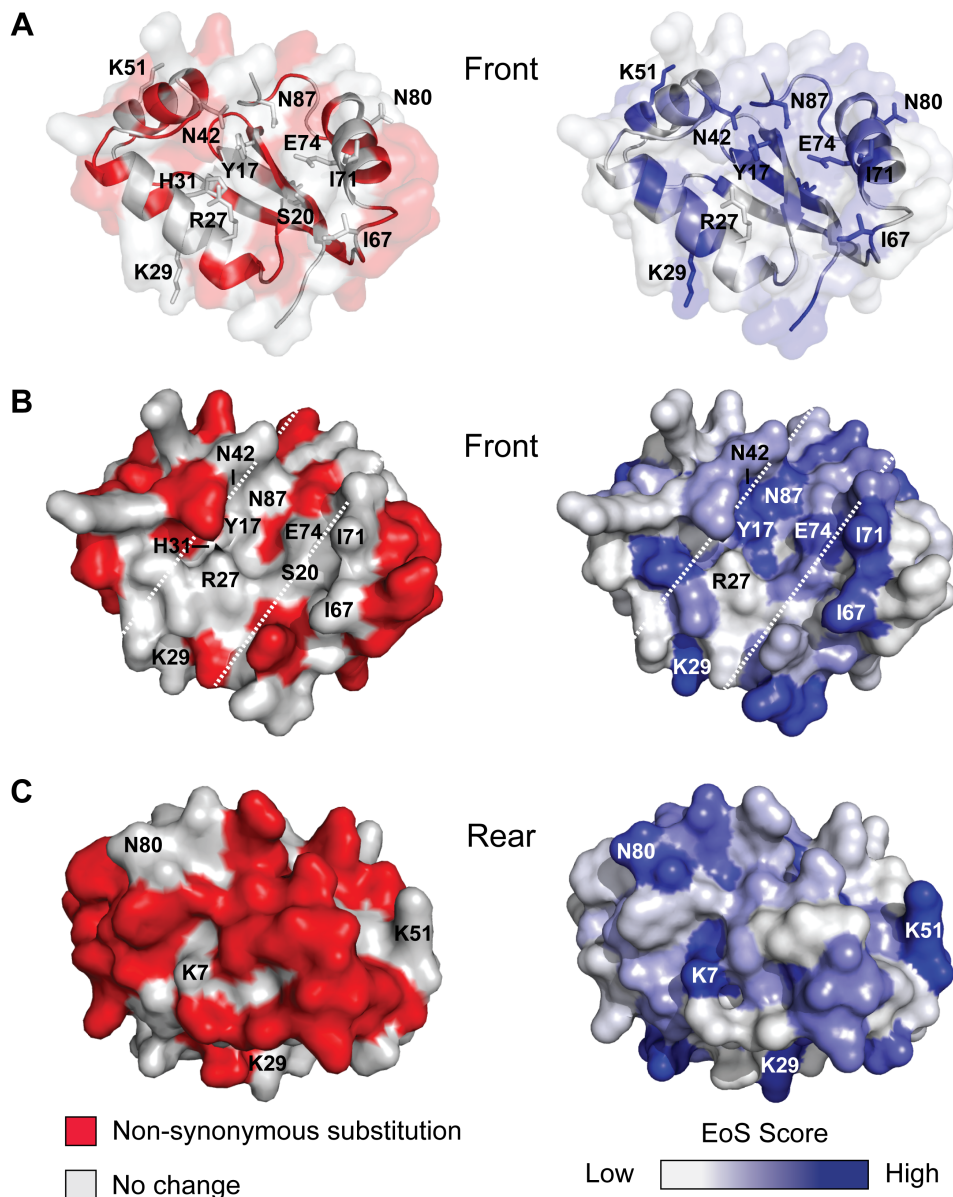


Figure 2.6: Mapping positions that tolerate non-synonymous substitutions and EoS data onto the I-Bmol homology model

For the *left* of each panel, amino acid positions that tolerate non-synonymous substitutions are shown in red and positions where no change was found are in white. On the *right* of each panel, the EoS score is shown as gradient. Positions with high EoS values are blue, and low EoS values are white. (A) Ribbon representations of a front view of the catalytic domain, with functionally critical residues identified by the present and previous studies indicated. (B) Surface representation of a front view of the I-Bmol catalytic domain, with the putative catalytic cleft bounded by white dashed lines. (C) Surface representation of a rear view of the catalytic domain.

S2.3). For example, residue N12 tolerated 10 amino acid substitutions (five to aspartate, four to threonine, and one to tyrosine) in different clones, indicating that the position is highly mutable. The significance of mutations in positions outside of the I-BmoI catalytic domain were difficult to interpret due to the absence of similarity between the DNA-binding domain of I-BmoI and that of I-TevI, for which there is an available structure. However, the I-BmoI and I-TevI linker regions, encompassing residues 95-112 of I-BmoI, show some degree of similarity (18). In particular, alanine-scanning mutants in this region of I-TevI functionally divided the linker into deletion intolerant and deletion tolerant regions (17,20). The equivalent deletion intolerant region of the I-BmoI linker appeared to contain few mutations among the selected clones (Figure 2.5), with some amino acid positions exhibiting high statistical confidence that the lack of mutation was not due to random chance (see below).

2.3.4 Statistical analysis of unigenic evolution data reveals residues of potential functional importance

More rigorous interpretation of the unigenic evolution data to identify hypo- and hypermutable residues requires statistical analyses to determine the probability that any codon will undergo a synonymous or non-synonymous change, taking into account the expected substitution frequencies based on the bias of the mutagenic PCR and the base composition of individual codons. Previous methods utilized a statistic, H , varying between -1 and 1, to determine the mutability of an individual amino acid position (37,39). We found that H both over- and under-estimates the mutability of positions because it lacks the power to determine if a significant difference exists between the selected and unselected clones at each position (40). For instance, the number of hypomutable residues in a given protein was vastly overestimated by assigning a value of -1 for any residue that had no non-synonymous changes, whereas there will be many positions that do not contain sequence changes simply because the mutagenic method is not saturating.

To take this, and other issues into consideration, we developed a new method for analyzing unigenic evolution data called EoS (Evidence of Selection) (40). The EoS value assesses whether the observed frequency of substitutions for any given codon in the

selected clones is statistically different from the expected frequency of mutations based on data for the same codon from the unselected pool of clones. Importantly, the EoS value explicitly represents both selection and the power to detect selection for each residue, and is plotted as the \log_2 ratio of the probability of non-random changes per codon versus the probability of random changes per codon. Thus, an EoS value of 8 represents a 1 in 256 probability that the observed spectrum of mutations occurred at random. Moreover, the fact that some residues possess low EoS scores does not necessarily indicate that these residues are not of potential functional significance, only that the sample size is insufficient to determine their significance.

For example, A/T rich codons are more likely to be mutated due to the bias of Taq polymerase, whereas G/C rich codons are less likely to be mutated. Under the EoS method, a lysine codon (AAA) for which there were no non-synonymous changes in the selected clones would be flagged as significant only if substitutions were observed at the same codon in the unselected clones. As an example, consider position K130. Six non-synonymous substitutions were present in the unselected clones at this position, whereas no substitutions were observed in the selected clones at this position (Supplementary Table S2.3). Because the spectrum of observed substitutions in the selected clones was significantly different from the observed and expected spectrum of mutations in the unselected clones, K130 is assigned a high EoS value of 8.03, thus providing strong evidence that the position is intolerant to substitution. In contrast, K171 has a low EoS value of -0.65 because a similar spectrum of mutations was observed in the both the selected and unselected clones at this position (Supplementary Table S2.3).

Using the EoS value as a guide, many positions throughout the length of I-BmoI appear to have potential functional importance (Figure 2.5). Of particular interest were residues in the catalytic domain that have a high EoS value. For example, K29, N42, N50, K51, F57, I67, I71, N80, and Y82 all have EoS values greater than 6. Apart from F57, none of these positions have been implicated by previous studies to be of potential functional importance. We mapped the EoS values of residues onto the I-BmoI homology model, and compared this representation to that of the mutable amino acid residues mapped onto the model (Figure 2.6). Residues that line the catalytic surface were refractory to

mutation in the selected clones (Figure 2.6A and 2.6B, and coloured white), whereas the EoS values indicate that the majority of these residues are predicted to be functionally critical (Figure 2.6A and 2.6B, and coloured blue). Many residues with high EoS scores were located away from the active site surface (Figure 2.6B and 2.6C), consistent with structural roles (for instance, K51, which strongly co-evolves H52; Table 2.1). Again, it is important to note that residues with a low EoS value (coloured white in Figure 2.6) do not imply that this position is not of potential significance, only that we lack the statistical power to draw such a conclusion given the observed spectrum of mutations (for example, R27).

Inspection of EoS values for the remainder of the I-BmoI sequence revealed that many residues in the linker and C-terminal region appear to be of potential functional significance (Figure 2.5). For instance, two residues in the linker region that are identical between I-TevI and I-BmoI, K105 and K111, possesses EoS values >8 . These residues in I-BmoI are excellent candidates for future mutagenic studies because they correspond to the deletion intolerant region of the I-TevI linker, where mutations drastically reduced or abolished cleavage activity (17,20).

2.3.5 Genetic analysis of site-directed mutants

The data generated by mutual information and unigenic evolution facilitated the identification of amino acids of potential functional importance that could be tested through mutational analyses. We focused on amino acids positions within the N-terminal catalytic domain, because these positions could be interpreted within the context of known and modelled GIY-YIG domain structures and previous mutagenic studies, and because mutations in the catalytic domain would not affect the DNA-binding activity of I-BmoI (15,32). Thus, we could be confident that any phenotype we observed for site-directed mutants was due to a defect related to the function of the catalytic domain. We selected amino acid positions for mutagenesis based on one of the following criteria: (1) the position had a high EoS value, (2) the residue(s) was identified as co-evolving with another residue, or (3) the residue possessed high information content in the alignment and had not been previously analyzed by mutational studies. For instance, H31 was chosen for mutagenesis because it possessed high information content, had a moderate

EoS value of 3.52, and has not previously been targeted by mutagenesis studies of GIY-YIG homing endonucleases. In contrast, position 71, an isoleucine in I-BmoI, displayed very low information content, yet had a high EoS value of 6.37. Furthermore, I71 co-evolves with S20, thus both residues were chosen for further analyses. The rationale for choosing other amino acid positions for mutagenesis is provided in Supplementary Table S2.4. In addition to generating site-specific mutants, we chose two clones identified in the unigenic evolution study that contained single mutations within the catalytic domain (N12D and H52R) for further analyses. These clones are representative of amino acid positions within the catalytic domain that are tolerant to change.

To determine if the identified residues were indeed critical for I-BmoI function, we individually analyzed the mutants using the genetic selection, allowing us to calculate a survival ratio that could be directly compared to that for WT I-BmoI. As shown in Table 2.2, the mutants could be divided into four classes. Class I mutants, Y17F, Y17H, S20A, H31F, H31Y, N42D, and N42A, showed the most dramatic effect as none survived the selection. Class II mutants were severely compromised in their survival, with ratios <0.07 , and included H31A, and I67N. Class III mutants, S20Q, H52R, I71N, and I71A, exhibited moderate survival ratios of between 0.14 and 0.43. The class IV mutants, N12D, S48A, K51L, H52Y, and Y86F all showed survival ratios essentially equivalent to WT I-BmoI, and thus were considered to have little effect on I-BmoI activity. We considered the possibility that a lack of survival in the selection could be due to the generation of a hyperactive mutant or an altered specificity mutant, resulting in a toxic endonuclease that would cause cell death. For all mutants, however, we observed no reduction in viable cells when cultures were plated on non-selective media during the selection protocol, suggesting that none of the mutants were toxic. Furthermore, we purified the H31A and N42D mutants and found by *in vitro* cleavage assays that both mutants were severely compromised for cleavage activity (see below and Figure 2.7), suggesting that the lack of survival in the genetic selection was due to loss of endonuclease activity.

Table 2.2 Survival ratios of I-BmoI variants with mutations in the catalytic domain

Mutant		Survival ratio relative to WT ¹
Class I	Y17F	0
	Y17H	0
	S20A	0
	H31F	0
	H31Y	0
	N42A	0
	N42D	0
Class II	H31A	0.003 +/- 0.002
	I67N	0.06 +/- 0.04
Class III	S20Q	0.12 +/- 0.02
	H52R ²	0.43 +/- 0.04
	I71A	0.22 +/- 0.03
	I71N	0.20 +/- 0.03
Class IV	N12D ²	1.06 +/- 0.03
	S48A	1.02 +/- 0.03
	K51L	0.86 +/- 0.10
	H52Y	0.96 +/- 0.05
	Y86F	0.99 +/- 0.04

¹ Values shown represent averages and standard deviation of at least three independent trials, normalized against the survival ratio for a WT I-BmoI selection performed in parallel.

² The H52R and N12D mutants were isolated from the pool of selected clones.

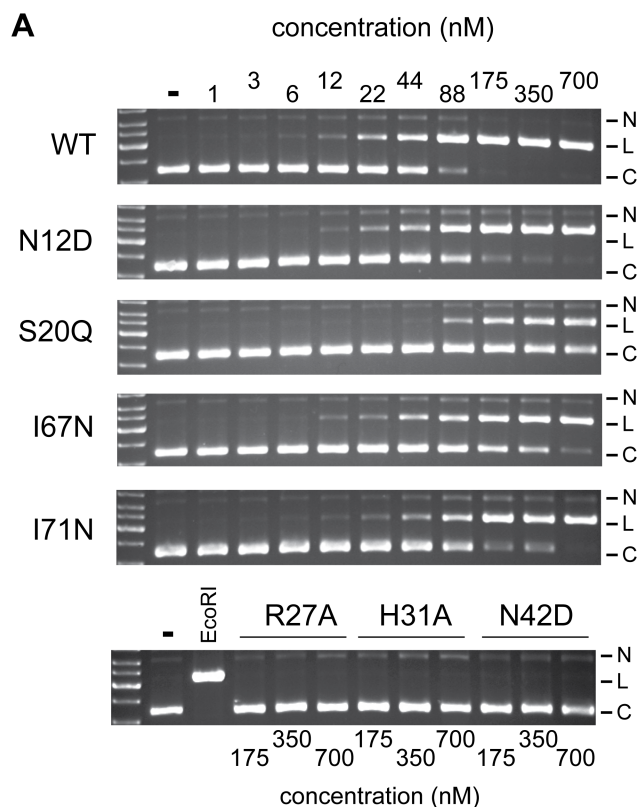


Figure 2.7: Cleavage activity of WT I-BmoI and site-directed mutants

(A) Shown are representative cleavage assays using 2-fold serial dilutions of the WT and mutant I-BmoI proteins, from 700 nM on the right to 1 nM on the left. The first lane of each gel from the left is unreacted substrate. For the R27A, H31A and N42D mutants, only the three highest protein concentrations were tested. In the bottom panel, substrate linearized by EcoRI is shown in the leftmost panel. Circular (C), linear (L), and nicked (N) plasmid forms are indicated to the right of each gel.

2.3.6 Cleavage assays with key site-directed mutants

To better understand the effect of individual mutations on endonuclease activity, we next purified key site-directed mutants (N12D, S20Q, H31A, N42D, I67N, and I71N) for use in *in vitro* cleavage assays (Figure 2.7 and Figure 2.8). The H31Y and I71A mutants were insoluble and not studied further, and apart from the N12D mutant, mutants with similar survival ratios to WT I-BmoI in the genetic assay were not purified for cleavage assays (Table 2.2). We first analyzed the activity of the mutant proteins relative to WT I-BmoI over a wide range of protein concentrations in cleavage assays using a circular substrate that contained a single I-BmoI site. As shown in Figure 2.7, the N12D, S20Q, I67N and I71N mutants exhibited a range of activities relative to WT I-BmoI. The N12D mutant was as active as WT I-BmoI, whereas I61N and I71N displayed intermediate cleavage activities. The S20Q mutant was the least active, and we estimate that this mutant retained ~30% cleavage activity of WT I-BmoI. In contrast, the H31A and N42D mutants were severely compromised for cleavage activity, with no linear product visible at the highest protein concentrations tested (Figure 2.7).

To gain additional insight into the observed phenotypes, we performed time-course cleavage assays with WT and mutant proteins to detect the appearance of a nicked intermediate as well as the linear product (Figure 2.8). These assays were performed with limiting metal ion, as we have previously shown that this reaction condition can effectively distinguish the two sequential nicking reactions that generate a double-strand break (27). With circular plasmid substrates used in this assay, the first nicking reaction generates a nicked intermediate with slow mobility, and the subsequent nicking reaction generates a linear product. The profile of the nicking assays suggested that the I-BmoI cleavage reaction followed the reaction scheme shown in Equation (1):



where C, N, and L are intact plasmid substrate, nicked intermediate, and linear product resulting from a double-strand break. The rate constant for the first nicking reaction is k_1 , and k_2 is the rate constant for the second nicking reaction.

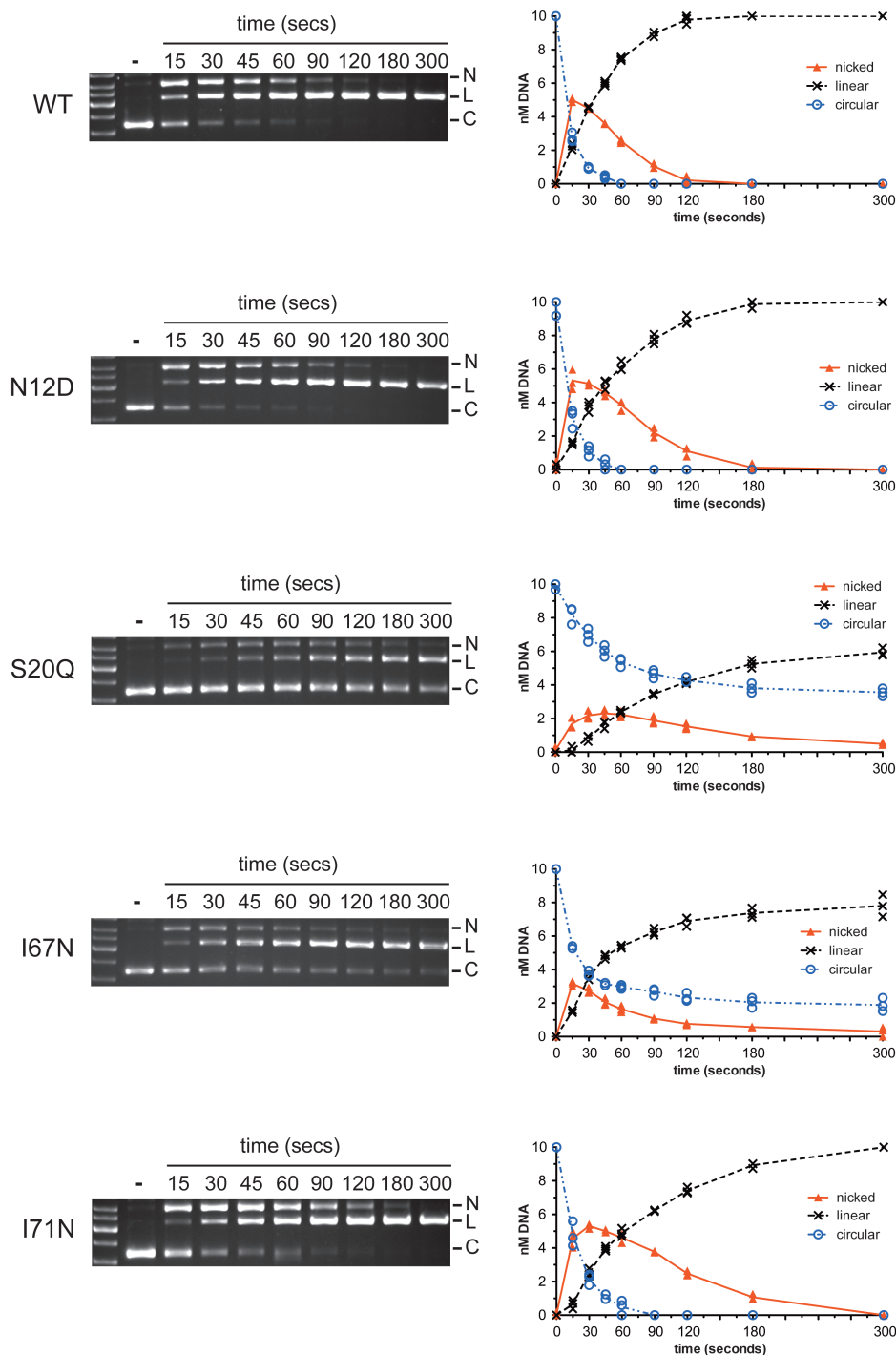


Figure 2.8: Nicking assays with WT and mutant proteins

Shown are representative agarose gels of time-course assays, with time points in seconds indicated above the gels. The first lane of each gel contains unreacted substrate (-). Circular (C), linear (L), and nicked (N) plasmid forms are indicated to the right of each gel. Beside each gel image is a graphical representation of the disappearance of substrate and appearance of products formed over time. Data points from three independent experiments are plotted, with a continuous line drawn through the average.

As shown in Figure 2.8 and Table 2.3, the reaction profiles and rate constants of the WT protein and N12D mutant were very similar, with conversion of the nicked intermediate to linear product complete by ~120 seconds. The reaction profiles of the S20Q and I67N mutants were very different from WT and N12D, with k_1 constants indicative of a slower first nicking reaction. Under conditions used in these assays, neither reaction was complete at the end of the time course, making it difficult to determine k_2 for the S20Q and I67N mutants using Equation 1 (Supplementary methods S2). In contrast, Figure 8 illustrates that the I71N mutant displayed a similar first nicking step to the WT and N12D proteins, but that the conversion to linear product was slower as indicated by kinetic analysis (Table 2.3), consistent with a nicked intermediate accumulating over an extended period of time relative to WT protein. Consequently, conversion to linear product was delayed, but still complete by the end of the time course. Collectively, these results clearly implicate residues identified by MUSE as functionally important, as mutation of these residues generates phenotypes that are distinct from the WT protein.

2.4 Discussion

Detailed insight into protein structure and function can be obtained by synthesizing data from multiple experimental, computational and structural approaches. Here, we elaborate an experimental framework to identify previously unknown functionally relevant residues of the GIY-YIG endonuclease I-BmoI that takes into account evidence other than strict conservation of residues in a multiple sequence alignment. Our goal was to use MUSE to identify key amino acids in I-BmoI for proof-of-principle experiments, highlighting the utility of the MUSE framework. We anticipate that our data will form a platform on which to pursue future structure and function studies of GIY-YIG homing endonucleases and other proteins containing the GIY-YIG domain, and that the MUSE approach will be generally applicable to a broad range of proteins.

2.4.1 Application of MUSE to GIY-YIG homing endonucleases

Past studies on GIY-YIG enzymes have utilized alignments and structural data to identify a set of absolutely conserved residues, subsequent mutation of which abolished cleavage activity. The residues were chosen for mutational analyses on the assumption that

Table 2.3: DNA cleavage by WT I-BmoI and variants identified by MUSE

protein	k_1 (s ⁻¹) ^a	k_2 (s ⁻¹) ^b
WT	0.080 ± 0.0033	0.034 ± 0.00069
N12D	0.073 ± 0.0045	0.024 ± 0.00081
I71N	0.045 ± 0.0028	0.015 ± 0.00051
S20Q	0.012 ± 0.003	N.D.
I67N	0.036 ± 0.0005	N.D.

^a k_1 is the rate constant for the first nicking step that generates nicked intermediate from circular substrate. For S20Q and I67N, k_1 is valid only at the initial time point (Supplementary Figure S2.4).

^b k_2 is the rate constant for the second nicking step that generates a linear product

conserved residues are important for function, and likely are components of the enzyme's active site. Such approaches, however, provided limited mechanistic insight and would miss residues that are not universally conserved amongst GIY-YIG enzymes, but nonetheless may be functionally critical. Furthermore, alignments of GIY-YIG containing proteins are dominated by the nucleotide excision repair protein, UvrC, which has a different set of functional constraints than GIY-YIG homing endonucleases. Thus, one aspect of the MUSE approach was to assemble an alignment of GIY-YIG proteins that closely resembled known homing endonucleases, with the goal of enhancing the information specific to GIY-YIG homing endonucleases. For instance, it is known that ~140 well-aligned sequences will provide sufficient information for covariation analyses to detect co-evolving residues (4). As discussed below, the strongest co-evolving amino acid pair detected by covariation analyses was S20-I71. One of these residues, I71, has not previously been identified for mutational analyses because this position is highly variable in multiple sequence alignments, yet our data indicate that I71 is a functionally significant residue.

That I-BmoI is a site-specific endonuclease was another critical factor in this study, facilitating the use of a genetic selection where survival in the assay was dependent on endonuclease function. Survival could also be influenced by the solubility or stability of I-BmoI variants. We found, however, that very low levels of I-BmoI expression were required for survival in the selection, suggesting that solubility was not likely a factor. Furthermore, use of a functional genetic selection allowed us to screen through a large population of I-BmoI variants mutagenized over the entire coding region, thus avoiding biases introduced by localized mutagenesis of specific residues or regions of the protein. Using the EoS method for analysis of unigenic evolution data, we obtained sufficient power to identify residues of potential functional significance for a protein the length of I-BmoI (266 residues) by sequencing a relatively small number of selected (87) and unselected (62) clones. In addition to identifying residues of potential importance, MUSE is also expected to identify positions that are tolerant to mutation (hypermutable). For instance, N12 was mutated 10 times in the selected clones as opposed to 7 times in the unselected clones, clearly implying a tolerance to mutation that is not expected to drastically affect I-BmoI function. Indeed, our *in vitro* analyses of the N12D mutant (the

most common mutation at that position) indicate essentially WT levels of activity, validating that the MUSE framework has the power to distinguish between residues in I-BmoI the catalytic domain that are functionally important and those that are not.

2.4.2 Residues within the GIY-YIG domain of I-BmoI that are important for function

The proposed catalytic mechanism for the GIY-YIG domain is based on the predicted function of a set of conserved residues (Y17, R27, E74 N87), each with distinct roles (21,25). Y17 is thought to act as a general base to activate a nucleophilic water; R27 may stabilize the 5' phosphate of the cleavage intermediate, or contact substrate; E74 coordinates a divalent metal ion that likely functions as the Lewis acid; and N87 is thought mainly to have a structural role in maintaining the active site architecture. Mutation of any of these residues to alanine abolishes cleavage activity, an uninformative phenotype as limited mechanistic insight is gained from catalytically inactive mutants. Hence, many unanswered questions remain regarding the catalytic mechanism of the GIY-YIG domain. For instance, the path of substrate DNA has only been inferred from the position of the conserved residues in the catalytic domain. In this regard, it is worth noting that an I-BmoI R27A mutant displayed a loss of hypersensitivity in footprinting experiments compared to WT protein (27), suggesting that the R27A mutant possesses a DNA contact defect that would not be expected if R27 functioned as a catalytic residue. Thus, additional types of evidence other than structural data are needed to definitively assign functional roles to the presumed set of catalytic residues. Moreover, the significance of residues that lie very close to the proposed active site of I-TevI and UvrC are unknown, and have largely been ignored because they are not conserved in multiple sequence alignments. In the following sections, we discuss the potential structural and functional significance of residues identified by MUSE for which subsequent mutagenesis revealed effects on I-BmoI cleavage activity.

2.4.3 S20 and I71

The S20-I71 pair was chosen for further study because this pair was the highest scoring set of co-evolving residues (Table 2.1 and Supplementary Table S2.4). In the GIY-YIG

domain alignment, S20 and I71 are both poorly conserved positions (but are conserved between I-BmoI and I-TevI), and are most commonly varied to glutamine and asparagine residues, respectively. In the I-BmoI homology model and the I-TevI structure, the residues are located in the proposed catalytic surface (Figure 2.9). In the I-TevI structure, S20 lies within hydrogen-bonding distance of the metal-coordinating residue E75, and could potentially position E75 or stabilize the interaction of E75 with divalent metal ion. Intriguingly, I71 can be structurally aligned with L116 of the His-Cys box homing endonuclease I-PpoI, where the residue is inserted into the minor groove of its homing site substrate (41,42). Further evidence for significant roles of S20 and I71 stemmed from the unigenic evolution data, as S20 possessed an EoS score of 3.25 and I71 possessed a score of 6.37. Neither position was mutated in the selected clones, whereas both positions were mutated in the unselected clones (Supplementary Table S2.3). We made S20A and S20Q mutations and found that the S20A mutant did not survive the genetic selection, whereas S20Q had a survival ratio of 0.063. Cleavage assays with S20Q revealed that at high protein concentrations the enzyme retained ~30% activity of WT protein, and a similar reduction in cleavage activity was observed when the equivalent residue in UvrC (K32) was mutated (21).

Both I71 mutants (I71A and I71N) survived the genetic selection, but with reduced ratios relative to WT I-BmoI. We purified the I71N mutant for further analysis, and found that it had slightly reduced activity relative to WT protein. Importantly, time-course cleavage assays revealed that the nicked plasmid intermediate persisted longer than for WT protein (Figure 2.8). Because I-BmoI generates a double-strand break by two independent and sequential nicking reactions (27), the phenotype of the I71N mutant implies a defect in the second nicking reaction. While the mechanistic basis of this defect will require future study, it is possible that I71 functions in substrate recognition, and that the I71N mutation greatly reduces substrate interactions that effect the second nicking reaction.

One intriguing question regarding the S20 and I71 residues is why do they co-evolve? It is worth noting that the predicted distance separating the residues (~6Å) could reflect a limitation of the homology model of the I-BmoI catalytic domain, as these residues are ~3.5Å apart in the I-TevI structure. In the 146 sequences used for mutual information

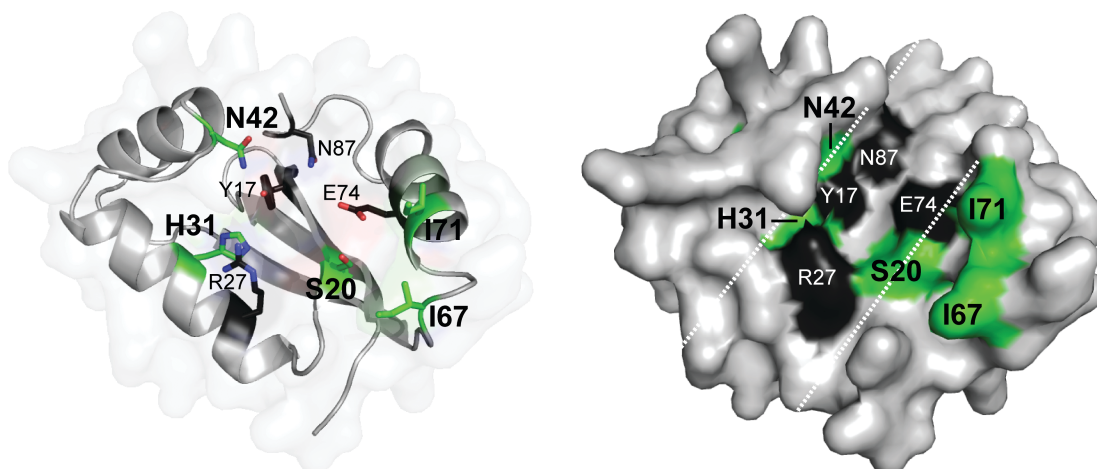


Figure 2.9: Summary of functionally relevant residues identified by MUSE

Shown are ribbon (left) and surface (right) representations of the I-BmoI catalytic domain with the residues identified by MUSE (S20, H31, N42, I67, and I71) highlighted in green, and previously identified functionally critical residues (Y17, R27, E74 and N87) shown in black. The catalytic cleft is highlighted as in Figure 2.1C.

analyses, position 20 is most commonly a serine, and position 71 is most commonly an isoleucine. However, there are almost equal occurrences of Q20-N71, Q20-D71, and S20-L71 pairs among the sequences. We attempted to rescue the S20Q mutant by making a second site mutation in position I71, generating a S20Q/I71N double mutant, reasoning that this double mutant represented a tolerated amino acid pairing at positions 20 and 71. However, the S20Q/I71N double mutant did not survive the genetic selection, suggesting that this combination of residues is not tolerated in the I-BmoI background, and that changes at other positions would be required to restore activity. Both S20 and I71 are in close proximity to the proposed metal binding residue, E74 (E75 in I-TevI), implying that certain combinations of residues may be favoured for correct positioning of the metal-binding residue within the active site (Figure 2.9).

2.4.4 H31

H31 was chosen for further analysis due to its high information content in the alignment, the observation that this position has a relatively high EoS score of 3.52, and a lack of mutations in the selected clones (but multiple mutations were present in the unselected clones; Supplementary Table S2.3). H31 is well conserved amongst GIY-YIG endonucleases, but is replaced by a tyrosine in the equivalent position in UvrC (Y43). In the homology model and I-TevI structure (Figure 2.9), H31 packs against the side chain of R27 and is within hydrogen bonding distance of both H40 and Y6, but no potential functional role was assigned to H31 based on the I-TevI structure. In the UvrC structure, Y43 was proposed to act as a general base to deprotonate a nucleophilic water molecule (21). To test the importance of H31 in I-BmoI, we made three mutations, H31A, H31F and H31Y, reasoning that the H31F would be structurally similar but chemically inert, while H31Y might retain limited function. Intriguingly, the H31F and H31Y mutants did not survive the genetic selection, while the H31A mutation had a very low survival ratio (Table 2.3). Unfortunately, the H31Y mutant proved to be insoluble and could not be studied *in vitro*, whereas no cleavage activity could be detected *in vitro* with the H31A mutant using plasmid-based cleavage assays. The H31A mutant may retain an extremely low level of activity that is difficult to detect using non-radioactive substrates, but that may be sufficient to permit very a low level of survival in the genetic selection. These

data provide the first evidence that H31 may play an important role in the active site of I-BmoI and other GIY-YIG endonucleases, either functioning as a base, by stabilizing the active site architecture by forming a hydrogen bond network with Y6 and H40, or by contacting substrate.

2.4.5 N42

N42 was chosen for mutational studies because it possessed a high EoS score of 6.49 (Figure 2.5 and Supplementary Table S2.4). In the GIY-YIG domain alignment, position 42 has little information content, although there is a tendency for this position to be a polar residue. N42 is located in a loop connecting α -helix1 and α -helix2, and is orientated so that the side chain points towards the predicted active site surface (Figure 2.9). Interestingly, the amino group of N42 is within hydrogen bonding distance of the hydroxyl group of the functionally critical Y17. We made two mutations that would disrupt this interaction, N42A and N42D, and found that both mutants did not survive the genetic selection, and that the N42D mutant displayed no cleavage activity *in vitro*. N42 may be critical because it functions to correctly position Y17 within the active site of the enzyme.

2.4.6 I67

Like I71, I67 is not conserved amongst GIY-YIG endonucleases, with this position displaying no information content (Figure 2.2, Supplementary Table S2.4). In the homology model of the I-BmoI catalytic domain, I67 is positioned in a loop connecting α -helix 3 and β -sheet 3, with its side chain pointed towards the active site surface (Figure 2.9). Mutation to asparagine reduced *in vitro* cleavage activity to approximately half that of WT I-BmoI, with defects in both the first and second nicking reactions (Figure 2.8). We envision that I67 may be involved in substrate interactions, similar to a role for I71.

2.4.7 Residues outside of the catalytic domain predicted to be critical for function

GIY-YIG homing endonucleases are modular proteins, with the N- and C-terminal domains connected by a flexible linker. Past studies on the I-TevI linker have revealed

that the linker is required to correctly position the N-terminal GIY-YIG domain on substrate for efficient cleavage (17,19,43). Remarkably, the linker can extend or retract to correctly position the catalytic domain on substrates that contain insertions or deletions that move the preferred cleavage sites from their WT position. This property of the linker has led to the proposal that I-TevI, and perhaps other GIY-YIG endonucleases, generates a DSB by a conformational change mechanism, whereby the linker is a critical component in repositioning the catalytic domain between the bottom- and top-strand nicking reactions. Interestingly, I-BmoI can also reposition the catalytic domain to cleave substrates with +5 and +10 insertions (18), suggesting that the I-BmoI linker functions analogously to the I-TevI linker in spite of limited amino acid similarity between the two proteins in the linker region. Intriguingly, our unigenic evolution data revealed that many residues within the I-BmoI linker region are predicted to be functionally significant. Notably, K105 and K111, conserved between I-TevI and I-BmoI, have EoS scores greater than 8. These residues correspond to the deletion intolerant region of the I-TevI linker, where 2- or 3-amino acid deletions abolish cleavage activity, although the functional basis for this phenotype is unknown (17,19). Furthermore, two additional sets of residues in the I-BmoI linker, centered on Y120 and K130, also have high EoS scores. Our data clearly implicate the I-BmoI linker as important for function, and identify a set of residues for future mutagenesis and functional studies.

Apart from a repeated nuclease-associated modular DNA-binding domain motif (NUMOD3) (44), I-BmoI shares little amino acid similarity with I-TevI in the C-terminal DNA binding domain, even though the enzymes are isoschizomers and bind the same stretch of thymidylate synthase sequence in *Bacillus mojavensis* and phage T4, respectively (32). The NUMOD3 motif corresponds structurally to a minor-groove binding α -helix that was first identified in the co-crystal of the I-TevI DNA-binding domain with its homing site substrate (45), and later in the structure of an unrelated HNH endonuclease, I-HmuI (46). In the I-TevI structure, the α -helix is positioned along the minor groove of its DNA substrate, and only one residue in the α -helix (S191) makes a single hydrogen bond contact to the phosphate backbone. Immediately preceding the α -helix, however, is H182 that makes two base-specific hydrogen bonds. Based on

sequence predictions, I-BmoI possesses three NUMOD3 motifs, with H147, H175 and H202 corresponding to the critical H182 of I-TevI. The EoS scores for the three histidine residues in I-BmoI are extremely low (Figure 2.5), and each position had multiple substitutions in the selected clones (Supplementary Table S2.3), suggesting that these residues perform different functions than the equivalent H182 of I-TevI. Similarly, residues that comprise the predicted I-BmoI helix-turn-helix (HTH) motif at the C-terminal end of the DNA-binding all have low EoS scores (Figure 2.5). In I-TevI, the analogous HTH motif makes extensive hydrophobic contacts with thymine residues in the substrate, providing specificity to the I-TevI substrate interaction. The tolerance of the I-BmoI HTH to mutation implies that it may function differently than the analogous HTH motif of I-TevI, and a detailed study of the sequence requirements for DNA binding by both HTH motifs would provide an intriguing comparative study.

2.4.8 Conclusion

Using a unified experimental approach that synthesizes three distinct types of data, we have identified previously unknown functionally relevant residues in the GIY-YIG homing endonuclease I-BmoI. Our results will form a platform for future studies of I-BmoI and other GIY-YIG-domain containing proteins because residues identified by MUSE, when mutated, generate distinct phenotypes that will provide mechanistic insight. Many of the positions identified by MUSE are non-conserved, and have escaped detection by traditional analyses such as strict conservation in multiple sequence alignments, providing a cautionary tale against using only a single methodology for structure and function studies. We anticipate that the MUSE framework will be generally applicable to a wide range of protein families, requiring ~140 well-aligned paralogous sequences, an enzymatic activity that forms the basis for a genetic selection and, although not essential, a structural model on which to interpret the MUSE data. For instance, the MUSE framework can be applied to any homing endonuclease or DNA endonuclease without difficulty, and would greatly aid in the re-design of endonucleases against novel target sequences. Unigenic evolution screens have been applied to eukaryotic proteins, including human Pin1 (a prolyl isomerase) and yeast TFIIB (47,48), and each of these proteins could be also analyzed within the MUSE framework.

2.5 References

1. Capra, J.A. and Singh, M. (2007) Predicting functionally important residues from sequence conservation. *Bioinformatics*, **23**, 1875-1882.
2. Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T. and Ben-Tal, N. (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res*, **33**, W299-302.
3. Watson, J.D., Laskowski, R.A. and Thornton, J.M. (2005) Predicting protein function from sequence and structural data. *Current opinion in structural biology*, **15**, 275-284.
4. Martin, L.C., Gloor, G.B., Dunn, S.D. and Wahl, L.M. (2005) Using information theory to search for co-evolving residues in proteins. *Bioinformatics*, **21**, 4116-4124.
5. Tillier, E.R. and Lui, T.W. (2003) Using multiple interdependency to separate functional from phylogenetic correlations in protein alignments. *Bioinformatics*, **19**, 750-755.
6. Stoddard, B.L. (2005) Homing endonuclease structure and function. *Q Rev Biophys*, **38**, 49-95.
7. Belfort, M., Derbyshire, V., Cousineau, B. and Lambowitz, A. (2002) In Craig, N., Craigie, R., Gellert, M. and Lambowitz, A. (eds.), *Mobile DNA II*. ASM Press, New York, pp. 761-783.
8. Zhao, L., Bonocora, R.P., Shub, D.A. and Stoddard, B.L. (2007) The restriction fold turns to the dark side: a bacterial homing endonuclease with a PD-(D/E)-XK motif. *EMBO J*, **26**, 2432-2442.
9. Dassa, B., London, N., Stoddard, B.L., Schueler-Furman, O. and Pietrokovski, S. (2009) Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family. *Nucleic Acids Res*, **37**, 2560-2573.
10. Grizot, S., Smith, J., Daboussi, F., Prieto, J., Redondo, P., Merino, N., Villate, M., Thomas, S., Lemaire, L., Montoya, G. *et al.* (2009) Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. *Nucleic Acids Res*.
11. Rosen, L.E., Morrison, H.A., Masri, S., Brown, M.J., Springstubb, B., Sussman, D., Stoddard, B.L. and Seligman, L.M. (2006) Homing endonuclease I-CreI derivatives with novel DNA target specificities. *Nucleic Acids Res*, **34**, 4791-4800.

12. Ashworth, J., Havranek, J.J., Duarte, C.M., Sussman, D., Monnat, R.J., Jr., Stoddard, B.L. and Baker, D. (2006) Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature*, **441**, 656-659.
13. Arnould, S., Chames, P., Perez, C., Lacroix, E., Duclert, A., Epinat, J.C., Stricher, F., Petit, A.S., Patin, A., Guillier, 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.
14. Quirk, S.M., Bell-Pedersen, D. and Belfort, M. (1989) Intron mobility in the T-even phages: high frequency inheritance of group I introns promoted by intron open reading frames. *Cell*, **56**, 455-465.
15. Derbyshire, V., Kowalski, J.C., Dansereau, J.T., Hauer, C.R. and Belfort, M. (1997) Two-domain structure of the *td* intron-encoded endonuclease I-TevI correlates with the two-domain configuration of the homing site. *J Mol Biol*, **265**, 494-506.
16. Bell-Pedersen, D., Quirk, S.M., Bryk, M. and Belfort, M. (1991) I-TevI, the endonuclease encoded by the mobile *td* intron, recognizes binding and cleavage domains on its DNA target. *Proc Natl Acad Sci U S A*, **88**, 7719-7723.
17. Liu, Q., Dansereau, J.T., Puttamadappa, S.S., Shekhtman, A., Derbyshire, V. and Belfort, M. (2008) Role of the interdomain linker in distance determination for remote cleavage by homing endonuclease I-TevI. *J Mol Biol*, **379**, 1094-1106.
18. Liu, Q., Derbyshire, V., Belfort, M. and Edgell, D.R. (2006) Distance determination by GIY-YIG intron endonucleases: discrimination between repression and cleavage functions. *Nucleic Acids Res*, **34**, 1755-1764.
19. Dean, A.B., Stanger, M.J., Dansereau, J.T., Van Roey, P., Derbyshire, V. and Belfort, M. (2002) Zinc finger as distance determinant in the flexible linker of intron endonuclease I-TevI. *Proc Natl Acad Sci U S A*, **99**, 8554-8561.
20. Kowalski, J.C., Belfort, M., Stapleton, M.A., Holpert, M., Dansereau, J.T., Piotrowski, S., Baxter, S.M. and Derbyshire, V. (1999) Configuration of the catalytic GIY-YIG domain of intron endonuclease I-TevI: coincidence of computational and molecular findings. *Nucleic Acids Res*, **27**, 2115-2125.
21. Truglio, J.J., Rhau, B., Croteau, D.L., Wang, L., Skorvaga, M., Karakas, E., Dellavecchia, M.J., Wang, H., Van Houten, B. and Kisker, C. (2005) Structural insights into the first incision reaction during nucleotide excision repair. *EMBO J*, **24**, 885-894.

22. Lagerback, P. and Carlson, K. (2008) Amino acid residues in the GIY-YIG endonuclease II of phage T4 affecting sequence recognition and binding as well as catalysis. *J Bacteriol*, **190**, 5533-5544.
23. Dunin-Horkawicz, S., Feder, M. and Bujnicki, J.M. (2006) Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics*, **7**, 98.
24. Pyatkov, K.I., Arkhipova, I.R., Malkova, N.V., Finnegan, D.J. and Evgen'ev, M.B. (2004) Reverse transcriptase and endonuclease activities encoded by Penelope-like retroelements. *Proc Natl Acad Sci U S A*, **101**, 14719-14724.
25. Van Roey, P., Meehan, L., Kowalski, J.C., Belfort, M. and Derbyshire, V. (2002) Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI. *Nat Struct Biol*, **9**, 806-811.
26. Ibryashkina, E.M., Zakharova, M.V., Baskunov, V.B., Bogdanova, E.S., Nagornykh, M.O., Den'mukhamedov, M.M., Melnik, B.S., Kolinski, A., Gront, D., Feder, M. *et al.* (2007) Type II restriction endonuclease R.Eco29kI is a member of the GIY-YIG nuclease superfamily. *BMC structural biology*, **7**, 48.
27. Carter, J.M., Friedrich, N.C., Kleinstiver, B. and Edgell, D.R. (2007) Strand-specific contacts and divalent metal ion regulate double-strand break formation by the GIY-YIG homing endonuclease I-BmoI. *J Mol Biol*, **374**, 306-321.
28. Van Roey, P. and Derbyshire, V. (2005) In Belfort M, S. B., Wood DW, Derbyshire V (ed.), *Homing Endonucleases and Inteins*. Springer-Verlag, Berlin, pp. 67-83.
29. Mueller, J.E., Smith, D., Bryk, M. and Belfort, M. (1995) Intron-encoded endonuclease I-TevI binds as a monomer to effect sequential cleavage via conformational changes in the *td* homing site. *EMBO J*, **14**, 5724-5735.
30. Gasiunas, G., Sasnauskas, G., Tamulaitis, G., Urbanke, C., Razaniene, D. and Siksnys, V. (2008) Tetrameric restriction enzymes: expansion to the GIY-YIG nuclease family. *Nucleic Acids Res*, **36**, 938-949.
31. Ibryashkina, E.M., Sasnauskas, G., Solonin, A.S., Zakharova, M.V. and Siksnys, V. (2009) Oligomeric structure diversity within the GIY-YIG nuclease family. *J Mol Biol*, **387**, 10-16.
32. Edgell, D.R. and Shub, D.A. (2001) Related homing endonucleases I-BmoI and I-TevI use different strategies to cleave homologous recognition sites. *Proc Natl Acad Sci U S A*, **98**, 7898-7903.

33. Edgell, D.R., Stanger, M.J. and Belfort, M. (2003) Importance of a single base pair for discrimination between intron-containing and intronless alleles by endonuclease I-BmoI. *Curr Biol*, **13**, 973-978.
34. Dunn, S.D., Wahl, L.M. and Gloor, G.B. (2008) Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics*, **24**, 333-340.
35. Dickson, R.J., Wahl, L.M. and Gloor, G.B. (2009) Identifying and seeing beyond systematic alignment errors using molecular covariation. *submitted*.
36. Wu, W., Wood, D.W., Belfort, G., Derbyshire, V. and Belfort, M. (2002) Intein-mediated purification of cytotoxic endonuclease I-TevI by insertional inactivation and pH-controllable splicing. *Nucleic Acids Res*, **30**, 4864-4871.
37. Deminoff, S.J., Tornow, J. and Santangelo, G.M. (1995) Unigenic evolution: a novel genetic method localizes a putative leucine zipper that mediates dimerization of the *Saccharomyces cerevisiae* regulator Gcr1p. *Genetics*, **141**, 1263-1274.
38. Chen, Z. and Zhao, H. (2005) A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic Acids Res*, **33**, e154.
39. Behrsin, C.D., Brandl, C.J., Litchfield, D.W., Shilton, B.H. and Wahl, L.M. (2006) Development of an unbiased statistical method for the analysis of unigenic evolution. *BMC Bioinformatics*, **7**, 150.
40. Fernandes, A., Kleinstiver, B.P., Edgell, D.R., Wahl, L.M. and Gloor, G.B. (2009) Estimating the evidence of selection and the reliability of inference in unigenic evolution. *Algorithms of Molecular Biology*. **5**:35.
41. Flick, K.E., Jurica, M.S., Monnat, R.J., Jr. and Stoddard, B.L. (1998) DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature*, **394**, 96-101.
42. Galburt, E.A., Chadsey, M.S., Jurica, M.S., Chevalier, B.S., Erho, D., Tang, W., Monnat, R.J., Jr. and Stoddard, B.L. (2000) Conformational changes and cleavage by the homing endonuclease I-PpoI: a critical role for a leucine residue in the active site. *J Mol Biol*, **300**, 877-887.
43. Bryk, M., Belisle, M., Mueller, J.E. and Belfort, M. (1995) Selection of a remote cleavage site by I-TevI, the *td* intron-encoded endonuclease. *J Mol Biol*, **247**, 197-210.

44. Sitbon, E. and Pietrokovski, S. (2003) New types of conserved sequence domains in DNA-binding regions of homing endonucleases. *Trends Biochem Sci*, **28**, 473-477.
45. Van Roey, P., Waddling, C.A., Fox, K.M., Belfort, M. and Derbyshire, V. (2001) Intertwined structure of the DNA-binding domain of intron endonuclease I-TevI with its substrate. *EMBO J*, **20**, 3631-3637.
46. Shen, B.W., Landthaler, M., Shub, D.A. and Stoddard, B.L. (2004) DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J Mol Biol*, **342**, 43-56.
47. Behrsin, C.D., Bailey, M.L., Bateman, K.S., Hamilton, K.S., Wahl, L.M., Brandl, C.J., Shilton, B.H. and Litchfield, D.W. (2007) Functionally important residues in the peptidyl-prolyl isomerase Pin1 revealed by unigenic evolution. *J Mol Biol*, **365**, 1143-1162.
48. Zeng, X., Zhang, D., Dorsey, M. and Ma, J. (2003) Hypomutable regions of yeast TFIIB in a unigenic evolution test represent structural domains. *Gene*, **309**, 49-56.
49. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. and Barton, G.J. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189-1191.
50. Eswar, N., Marti-Renom, A., Webb, B., Madhusudhan, M.S., Eramian, D., Shen, M., Pieper, U. and Sali, A. (2006) Comparative Protein Structure Modeling With MODELLER. *Current Protocols in Bioinformatics*, **15**, 5.6.1-5.6.30.
51. Arnold, K., Bordoli, L., Kopp, J. and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, **22**, 195-201.
52. Schneider, T.D. and Stephens, R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res*, **18**, 6097-6100.

Chapter 3

3 Divalent metal ion differentially regulates the sequential nicking reactions of the GIY-YIG homing endonuclease I-BmoI

The work presented in this chapter is reproduced (with permission, Appendix S1) from:

Kleinstiver, B.P., Bérubé-Janzen, W., Fernandes, A.D., Edgell, D.R. (2011) Divalent metal ion differentially regulates the sequential nicking reactions of the GIY-YIG homing endonuclease I-BmoI. *PLoS One* 6(8):e23804.

3.1 Introduction

Homing endonucleases are DNA endonucleases that primarily function as mobile genetic elements by introducing double-strand breaks or nicks at specific homing sites in naïve genomes (1,2). DNA repair, recombination and replication pathways repair the double-strand break (or nick) using the endonuclease-containing genome as a template, resulting in the mobilization of the homing endonuclease gene and surrounding DNA to the recipient genome (1). In model laboratory systems, the efficiency of endonuclease-mediated homing is extraordinary, reaching 80-100% in some cases (3-5), implying that homing endonucleases can quickly spread through populations of naïve genomes. The rapid accumulation of genome sequence data has revealed an abundance of homing endonuclease genes in bacterial, archaeal, viral, and organellar genomes, including the mitochondrial and nuclear genomes of eukaryotes (6,7). Five distinct homing endonuclease families, defined by conserved amino acids, have been characterized to date. They are the LAGLIDADG, HNH, His-Cys box, GIY-YIG, and PD-(D/E)xK families (8,9).

Homing endonucleases have attracted much interest as potential reagents for manipulating complex genomes, and substantial effort has been devoted to reprogramming the specificity of a select few LAGLIDADG endonucleases to cleave clinically relevant targets (10-14). Understanding the mechanism of DNA hydrolysis by homing endonucleases is critical for their utilization as reagents for targeted manipulation of complex genomes. From a mechanistic perspective, the least understood of the homing

endonuclease families are the GIY-YIG enzymes. GIY-YIG homing endonucleases are modular in nature and contain an N-terminal cleavage domain of ~100 amino acids that includes the class-defining GIY-YIG motif, and a C-terminal DNA-binding domain that is composed of distinct modules (Figure 3.1) (15,16). The N- and C-terminal domains are connected by a flexible linker that functions as a molecular ruler to position the catalytic domain at the correct distance on substrate from the DNA-binding domain (17-20). Two well-studied GIY-YIG endonucleases are the isocaudomers I-TevI and I-BmoI, encoded within group I introns interrupting the thymidylate synthase genes of bacteriophage T4 and *Bacillus mojavensis*, respectively (21-23). I-TevI and I-BmoI introduce a staggered double-strand break by two independent and sequential nicking reactions at the same positions of their substrates, with the bottom (non-coding) strand of substrate nicked before the top (coding) strand (24,25). DNA bending assays and in-gel footprinting demonstrated that significant DNA distortions on the bottom strand occur independently of the first nicking reaction (24,25). Interestingly, studies with mutant DNA substrates revealed distinct sequence requirements for efficient double-strand break formation by each endonuclease. In particular, I-BmoI absolutely requires a GC base pair at position -2 immediately 3' to the top-strand nicking site (26), yet displays no sequence preference for bases flanking the bottom-strand nicking site, while I-TevI requires an additional CG base pair 5' to the bottom-strand nicking site for cleavage activity (27).

The GIY-YIG nuclease motif is not unique to homing endonucleases, as the domain is found in a variety of protein scaffolds (28), including DNA repair proteins (29), retrotransposable elements (30), and restriction endonucleases (31-33). Structural studies of diverse GIY-YIG domains have revealed a compact α/β -fold composed of a central three-stranded antiparallel β -sheet flanked by three α -helices (29,34-38). A set of conserved residues comprise the single active site that uses a one-metal ion mechanism to catalyze DNA hydrolysis (36,37). The function of the GIY-YIG nuclease domain varies in a scaffold-dependent manner, as the domain has acquired additional structural units that are interspersed within the GIY-YIG domain to direct oligomerization or DNA binding. For instance, the GIY-YIG restriction enzyme Eco29kI utilizes additional folds to function as a dimer (39), with each monomer nicking one strand of substrate. In the

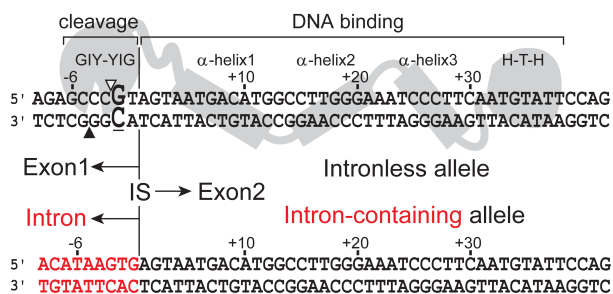


Figure 3.1: Model of I-BmoI interactions with intronless and intron-containing substrates based on DNA footprinting experiments

Shown is a schematic of the modular interaction of I-BmoI with the intronless *thyA* allele (upper), and the resultant changes to the target site upon intron insertion that generates the intron-containing allele (lower). Top- and bottom-strand nicking sites are indicated by open and filled triangles, respectively. The critical GC-2 base-pair is shown in enlarged bold-type font, and the intron insertion site (IS) is indicated by a vertical line with exon 2 downstream and either exon 1 or the intron (red) upstream.

case of the UvrC nucleotide excision repair protein, the single GIY-YIG domain nicks 3' to a lesion site to initiate a repair event (40). The homing endonucleases are distinct from other GIY-YIG family members, as they have an extended structure and bind DNA as monomers (25). They are believed to successively use the single active site to generate a double-strand break (38), but how the active site is reorganized to accommodate different DNA strands for the sequential nicking reactions is unknown.

Here, we investigate the contributions of divalent metal ion and DNA sequence to the sequential nicking reactions of I-BmoI. We show that cleavage by I-BmoI is sensitive to the identity and concentration of divalent metal ion, and that the second-strand nicking rate is affected to a greater extent by low metal ion concentrations than the rate of the first nicking reaction. DNA substrate mutations at key positions within the cleavage site of I-BmoI can differentially attenuate or completely abolish first- and second-strand nicking. Furthermore, we show that the GC-2 base pair is crucial for generating bottom strand minor-groove distortions that are a necessary prerequisite to cleavage.

3.2 Materials and methods

3.2.1 Bacterial strains and plasmids

E.coli strains DH5 α and ER2566 (New England Biolabs) were used for plasmid manipulations and protein expression, respectively. pTYBmoI was used to over-express a wild-type, codon-optimized version of I-BmoI for purification as previously described (41). pBmoHS is a pBS derivative that contains an insert corresponding to 49 bp of the intronless *thyA* substrate (26). A complete description of all plasmids used in this study is found in Supplementary Table S3.1.

3.2.2 Oligonucleotides

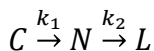
Oligonucleotides used in this study can be found in Supplementary Table S3.2.

3.2.3 Cleavage assays on plasmid substrates

Single time-point cleavage assays to examine metal dependence were performed in 10- μ l volumes containing 20 mM Tris-HCl pH 8.0, 250 mM NaCl, 2.5% glycerol, 10 nM wild-

type or G-2T pBmoHS, and various concentrations of MgCl₂, CaCl₂, CuCl₂, MnCl₂, NiCl₂, and ZnCl₂. Reactions were started by the addition of I-BmoI to final a concentration of 175 nM, allowed to proceed for 90 seconds at 37°C, and stopped by the addition of 4 µl stop dye (100 mM EDTA, 25% glycerol, 0.2% bromophenol blue). Stopped reactions were heated for 5 minutes at 95°C, cooled on ice for 5 minutes, electrophoresed on a 1% agarose gel, and stained in a 1xTAE solution containing ethidium bromide (Caledon) prior to analysis on an AlphaImager™3400 (Alpha Innotech). Time course cleavage reactions to determine Mg²⁺ dependence were performed on supercoiled pBmoHS in 115-µl volumes with conditions similar to those listed above, with either 0.5 mM, 1 mM, 2 mM, 5 mM, 7 mM, or 10 mM MgCl₂. A 10-µl aliquot was removed prior to the addition of wild-type I-BmoI to final a concentration of 175 nM, and subsequent 10-µl aliquots were removed at ten time points into 4 µl stop dye. Stopped reactions were visualized as indicated above. Reaction progress was determined by the relative amount of circular substrate, nicked intermediate, and linear product. At least 3 independent trials were conducted for each MgCl₂ concentration. Time course cleavage reactions with mutant pBmoHS substrates were performed in 65-µl volumes with conditions similar to those listed above, with either 2 mM or 10 mM MgCl₂. A 10-µl aliquot was removed prior to the addition of wild-type I-BmoI to final a concentration of 175 nM, and subsequent 10-µl aliquots were removed at 15, 30, 60, 120, and 180 seconds into 4 µl stop dye. Mutant substrates were classified as follows: class I substrates displayed a reaction progress similar to wild-type intronless substrate, class II substrates displayed defects for both bottom and top-strand nicking reactions at 2 mM MgCl₂ and ‘rescue’ of cleavage to near WT levels after 180 seconds in 10 mM MgCl₂, and class III substrates showed severe to complete defects in 2 mM MgCl₂ and no appreciable rescue of activity in 10 mM MgCl₂ after 180 seconds. Cleavage assays for class I, II, and a selection of class III substrates (G-2T, C-3G/C-5A, T-1G/C-3G/C-5A, C-3G/C-4A/C-5A, and In+) were conducted as described above. For the remainder of class III substrates, only the accumulation of nicked intermediate or linear product above background levels was measured after a 180 second incubation with I-BmoI in 2 mM or 10 mM MgCl₂. Three independent reactions were performed for substrates that gave

measurable rate constants, and two independent trials were performed for class III substrates that displayed minimal or no cleavage. Rate constants for the reaction



were estimated for the two-constant irreversible kinetic model using Prism5 (GraphPad Software) or via the Bayesian bootstrap (61). The time-course data were fit to the three following equations

$$C = [C_0] \exp(-k_1 t) \quad (\text{equation 1})$$

$$N = [C_0] \frac{k_1}{k_1 - k_2} [-\exp(-k_1 t) + \exp(-k_2 t)] \quad (\text{equation 2})$$

$$L = [C_0] \left[1 + \frac{1}{k_1 - k_2} (k_2 \exp(-k_1 t) - k_1 \exp(-k_2 t)) \right] \quad (\text{equation 3})$$

where C_0 is the initial concentration of circular substrate (in nM), N is the concentration of nicked DNA (in nM), L is the concentration of linear product (in nM), k_1 is the first nicking rate constant (in s^{-1}), k_2 is the second nicking constant (in s^{-1}), and t is time (in seconds). For each bootstrap replicate, parameters were optimized for minimal discrepancy with the data under the half-taxi metric (62). This procedure was found to be more robust than a standard least-squares estimation due to the constraint that the total amount of circular, nicked, and linear DNA is constant. Posterior parameter medians and 95% confidence intervals are reported in Supplementary Table S3.3. The ratio of k_1/k_2 at various $MgCl_2$ concentrations was reported as the value of the $\log_{10} k_1 - \log_{10} k_2$ to minimize over-weighting the end points if plotted on a linear scale.

3.2.4 OP-Cu in-gel footprinting

1,10-phenanthroline copper (OP-Cu) footprinting experiments were conducted as previously described (24). Substrates used included a 74-mer duplex oligonucleotide substrate corresponding to the intronless *thyA* target site (DE-116/DE-117), the intronless 74-mer with G-2T (DE-446/DE-447), G-2A (DE-459/DE-460), or G-2C (DE-461/DE-462) substitutions, the intron-containing 74-mer (DE-444/DE-445), and the intron-containing 74-mer with A-2G substitution (DE-463/DE-464) (see Supplementary Table S3.2). Gel images were analyzed using ImageQuant 5.2 (GE Healthcare Life Sciences). To quantify hypersensitive sites observed in the OP-Cu footprint, two bands outside of the I-BmoI protection region were selected to normalize phosphorimager units (positions

+9 and +10 for intronless substrates, and +11 and +12 for intron-containing substrates, relative to the intron insertion site). Hypersensitivity to OP-Cu at positions -1 and -2 was calculated by expressing the ratio of normalized phosphorimager units at sites in the shifted footprint (UC, upper complex) to the units in the unbound substrate reaction (UNB, unbound) (24).

3.2.5 Molecular modeling

The I-BmoI GIY-YIG domain homology model was built as previously described (41), and alignments to annotated structures were performed using MacPyMOL v1.2r. Residues 6-10 of the homology model were aligned to residues 47-51 of the solution structure of E142Q Eco29kI (3MX1) (Figure 6A) (36). A subsequent alignment was performed with the I-BmoI homology model and a single subunit of the Y49F/L69K Eco29kI structure in complex with its 18 base pair substrate (3NIC) (36). For illustration purposes, only nucleotides (-4)-CCCGCGGGC-(+5) of the Eco29kI substrate were shown. Additional alignments were performed using structures of the I-TevI GIY-YIG domain (1KM0 and 1LN0) in place of the I-BmoI homology model (38), and of Hpy188I in complex with substrate (3OQG) in place of the Eco29kI structure (37), yielding nearly identical results.

3.3 Results

3.3.1 Magnesium is the preferred divalent metal ion for efficient and specific cleavage

To determine the divalent metal ion preference for cleavage by I-BmoI, we tested various metals across a range of concentrations in assays with a supercoiled plasmid (pBmoHS) containing the I-BmoI intronless *thyA* target site (Figure 3.2). I-BmoI reaction progress can be visualized using supercoiled substrate, as the first nicking reaction generates a nicked plasmid intermediate and the second nicking reaction converts the nicked intermediate to linear product (41). Cleavage assays were performed under single turnover conditions (protein excess) with a range of MgCl₂, CaCl₂, CuCl₂, MnCl₂, NiCl₂, and ZnCl₂ concentrations (Figure 3.2). Reactions with CoCl₂ did not yield any products.

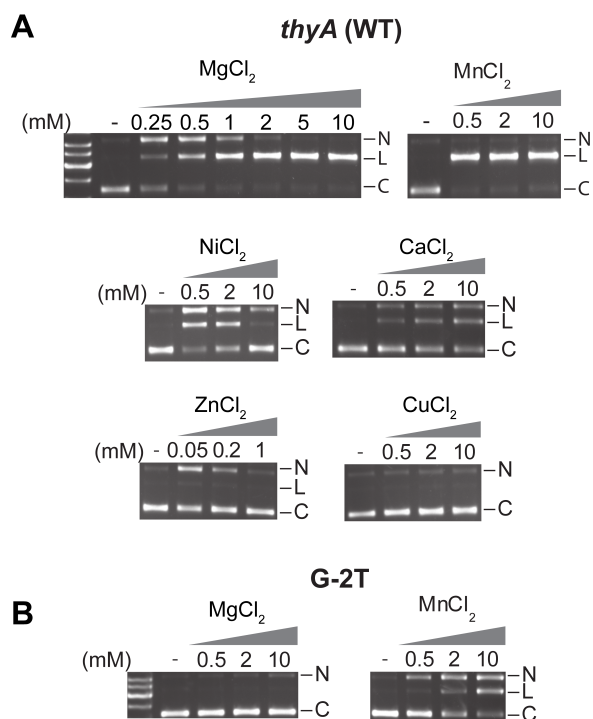


Figure 3.2: Magnesium is the preferred divalent metal ion for efficient and specific cleavage by I-BmoI

(A) Representative gel images of time-point cleavage assays with I-BmoI performed on supercoiled substrate containing the intronless *thyA* target site. Reactions contained increasing concentrations of MgCl₂, CaCl₂, CuCl₂, MnCl₂, NiCl₂, and ZnCl₂. Lanes that lack I-BmoI (-) have 10 mM metal (1 mM for ZnCl₂), and nicked (N), linear (L), and circular (C) plasmid forms are indicated to the right of each gel image. (B) Representative gel images of I-BmoI cleavage assays performed on supercoiled substrate containing a mutation at the critical GC-2 basepair (G-2T). Reactions were performed in the presence of MgCl₂ or MnCl₂.

The overall divalent metal ion preference of I-BmoI was $\text{Mn}^{2+} > \text{Mg}^{2+} > \text{Ni}^{2+} > \text{Ca}^{2+} > \text{Zn}^{2+} \gg \text{Cu}^{2+} = \text{Co}^{2+}$. In particular, cleavage was extremely efficient in the presence of MnCl_2 for all concentrations tested, while reactions in MgCl_2 displayed a second strand defect at 0.25 mM MgCl_2 , near complete conversion of nicked intermediate to linear product at 2 mM MgCl_2 , and complete conversion to product at 10 mM MgCl_2 (Figure 3.2A). NiCl_2 appeared to be inhibitory at 10 mM, yet similar levels of cleavage were observed at 0.5 mM NiCl_2 and MgCl_2 . Cleavage was observed in the presence of CaCl_2 (albeit greatly reduced versus MgCl_2), whereas ZnCl_2 and CuCl_2 were inhibitory at most concentrations tested, with the exception that 0.05 mM ZnCl_2 supported nicking only. These findings correlate well with the typical roles of divalent metal observed for the majority of site-specific DNA endonucleases, including the tetrameric GIY-YIG restriction enzyme Cfr42I, apart from the observation that cleavage by Cfr42I was most efficient in CoCl_2 -containing buffers (31,42-44).

Site-specific endonucleases have been shown to exhibit increased activity in the presence of manganese, but at the cost of fidelity (45,46). To determine if I-BmoI displayed a similar loss of fidelity in the presence of MnCl_2 relative to MgCl_2 , we performed cleavage assays on a plasmid substrate with a G-2T mutation, against which I-BmoI is known to retain only limited activity (Figure 3.2B) (26). We observed minimal nicking of the pBmoHS G-2T substrate by I-BmoI in the presence of 10 mM MgCl_2 , while 2 mM and 10 mM MnCl_2 were conducive for cleavage (but reduced versus intronless *thyA* substrate). Collectively, these data suggested that the observed increase in efficiency in the presence of MnCl_2 is partly due to decreased fidelity, allowing I-BmoI to cleave non-cognate sites. Furthermore, based on these results we determined that of the divalent metal ions tested, a selection of magnesium concentrations would provide the optimal level of efficiency and specificity.

3.3.2 Limiting divalent metal ion has a more pronounced regulation of second strand nicking

To gain insight into the substrate conversion process by I-BmoI, we selected a range of MgCl_2 concentrations to dissect reaction progress as previous data indicated that the sequential reactions have distinct metal requirements (24). We performed time-course

cleavage assays with the *thyA* supercoiled substrate and determined rate constants based

on the following reaction scheme: $C \xrightarrow{k_1} N \xrightarrow{k_2} L$

where k_1 is the rate constant for conversion of supercoiled plasmid (C) to nicked intermediate (N), and k_2 is the rate constant for conversion of nicked intermediate to linear product (L). Representative time course experiments in 0.5 mM and 10 mM MgCl₂ are shown in Figure 3.3, with k_1 and k_2 rate constants for all MgCl₂ concentrations tested summarized in Table 3.1. We observed a 5.2-fold reduction in k_1 when MgCl₂ was reduced from 10 mM to 0.5 mM MgCl₂, while a 12.2-fold decrease was observed for k_2 . The distinct metal requirements for each nicking reaction is represented in the log₁₀ k_1 - log₁₀ k_2 versus MgCl₂ concentration plot, which shows a more pronounced decrease in k_2 as the concentration of MgCl₂ is reduced (Figure 3.3C). These data suggest that divalent metal ion regulates the rate of both nicking reactions, but that the second strand nicking reaction has a more stringent requirement for divalent metal than the first nicking reaction.

3.3.3 Assays with mutant substrates reveal three distinct cleavage phenotypes

To further probe the contribution of divalent metal ion and DNA sequence to the sequential nicking reactions by I-BmoI, we performed time-course cleavage assays with various supercoiled mutant substrates at 2 mM and 10 mM MgCl₂. We took advantage of a series of previously constructed substrates (26), whereby positions -6 through -1 of intronless *thyA* substrate were individually and in combination changed to the corresponding intron-containing sequence (Figure 3.1 and Table 3.2), which I-BmoI does not cleave. End-point assays with these substrates highlighted the importance of the GC-2 base pair in the generation of a DSB by I-BmoI (26). The previous assays, however, were incapable of distinguishing defects in each independent nicking reaction because they were performed on linearized plasmid substrates in 10 mM MgCl₂. Time-course assays using mutant cleavage site plasmids revealed that the substrates segregated into three distinct classes (Figure 3.4 and Table 3.2). Class I mutants were defined as substrates that behaved essentially as intronless *thyA* substrate, showing slightly reduced nicking and slower conversion of nicked intermediate to linear product at 2 mM versus 10 mM MgCl₂

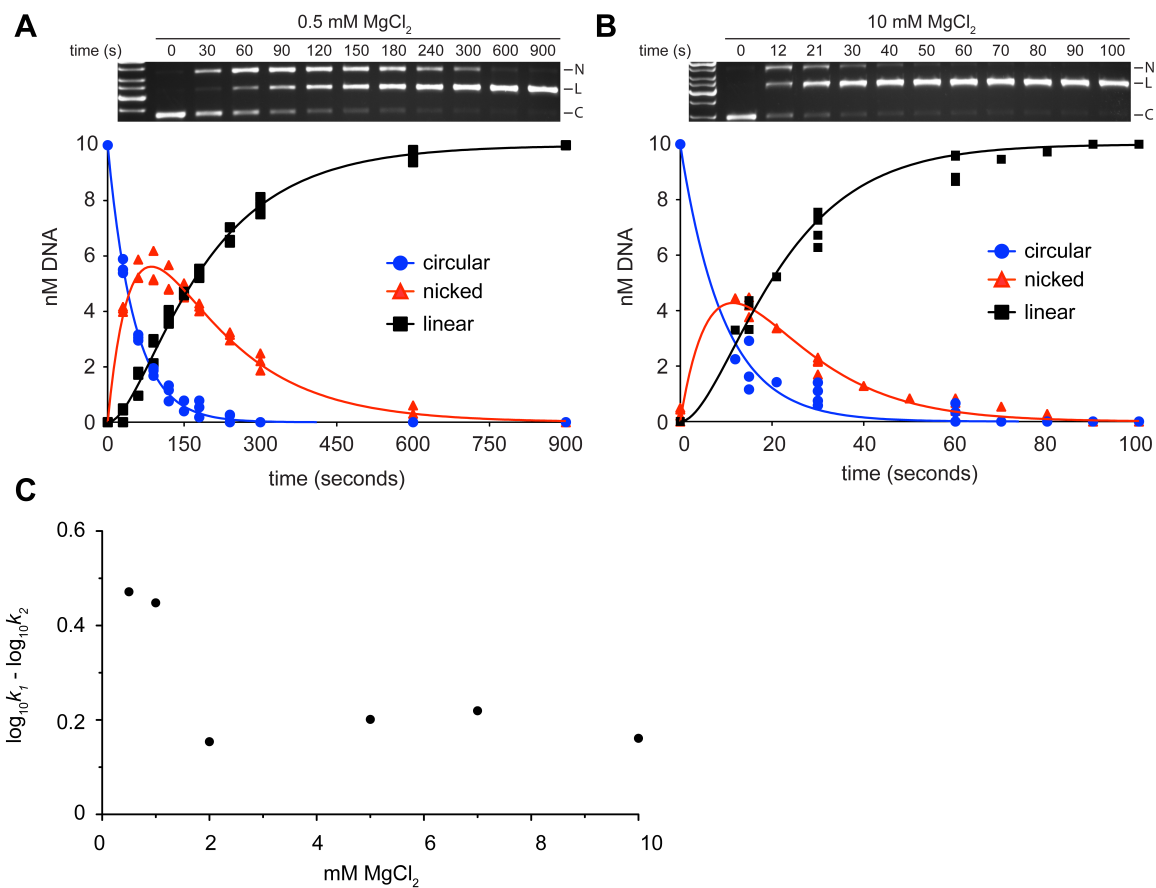


Figure 3.3: Limiting divalent metal ion has a greater effect on second strand than first strand nicking

Shown are representative images of time-course cleavage assays with supercoiled substrate containing the intronless *thyA* target site and I-BmoI in (A) 0.5 mM MgCl₂ and (B) 10 mM MgCl₂, as well as progress curves for each condition. Circular substrate (C), nicked intermediate (N) and linear product (L) are indicated on the gel images. Individual data points from three independent replicates are shown in the progress curves, and the solid continuous lines are the best fit of the data to equations 1 and 2. (C) Plot of the $\log_{10}k_1 - \log_{10}k_2$ value for MgCl₂ concentrations tested.

Table 3.1: Rate constants for first- and second-strand nicking reactions in different MgCl₂ concentrations

[MgCl ₂] (mM)	k_1 (s ⁻¹) ^a	k_2 (s ⁻¹)
10	0.10 ± 0.007	0.078 ± 0.008
7	0.10 ± 0.005	0.060 ± 0.003
5	0.079 ± 0.004	0.050 ± 0.002
2	0.056 ± 0.001	0.036 ± 0.001
1	0.037 ± 0.0006	0.013 ± 0.0004
0.5	0.019 ± 0.0002	0.0064 ± 0.0002

^a k_1 and k_2 rate constants are reported as the best fit value of three independent experiments with standard error.

Table 3.2: Summary of cleavage data for intronless and mutant substrates

	Substrate						Rate constants (s ⁻¹) ^a			
							2mM MgCl ₂		10mM MgCl ₂	
	-6	-5	-4	-3	-2	-1	<i>k</i> ₁ ^b	<i>k</i> ₂ ^c	<i>k</i> ₁	<i>k</i> ₂
<i>(intronless)</i>	G	C	C	C	G	T	0.057	0.026	0.11	0.074
Class I			A				0.05	0.028	0.1	0.057
(like wild-type)	T						0.057	0.024	0.087	0.058
Class II						G	0.031	0.0095	0.065	0.025
(rescue)				G			0.01	0.0084	0.025	0.025
		A					0.0095	0.016	0.031	0.042
			A	G			0.0082	0.0091	0.024	0.022
		A	A				0.0044	0.015	0.014	0.043
	T	A					0.011	0.017	0.028	0.048
	T	A	A				0.0045	0.015	0.014	0.047
Class III					T		0.00027	<i>n.d.</i> ^d	0.00061	<i>n.d.</i>
(no rescue)					T	G	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>
				G	T		<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>
			A		T		<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>
		A			T		<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>
		A		G			0.00037	0.00053	0.0013	0.014
				G	T	G	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>
		A		G		G	0.00016	<i>n.d.</i>	0.00035	0.0026
			A	G	T		<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>
	T		A		T		<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>
		A	A	G			0.0002	0.0017	0.00052	0.0091
		A	A	G	T		<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>
<i>(intron-containing)</i>	T	A	A	G	T	G	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>

^a Rate constants were determined from three independent experimental trials for all substrates.

^b *k*₁, the rate constant for the first nicking reaction that generates nicked intermediate from circular substrate (expressed as the median of a 95% confidence interval, see Supplementary Table S3.3)

^c *k*₂, the rate constant for the second nicking reaction that generates the linear product (expressed as the median of a 95% confidence interval, see Supplementary Table S3.3)

^d *n.d.* not determined.

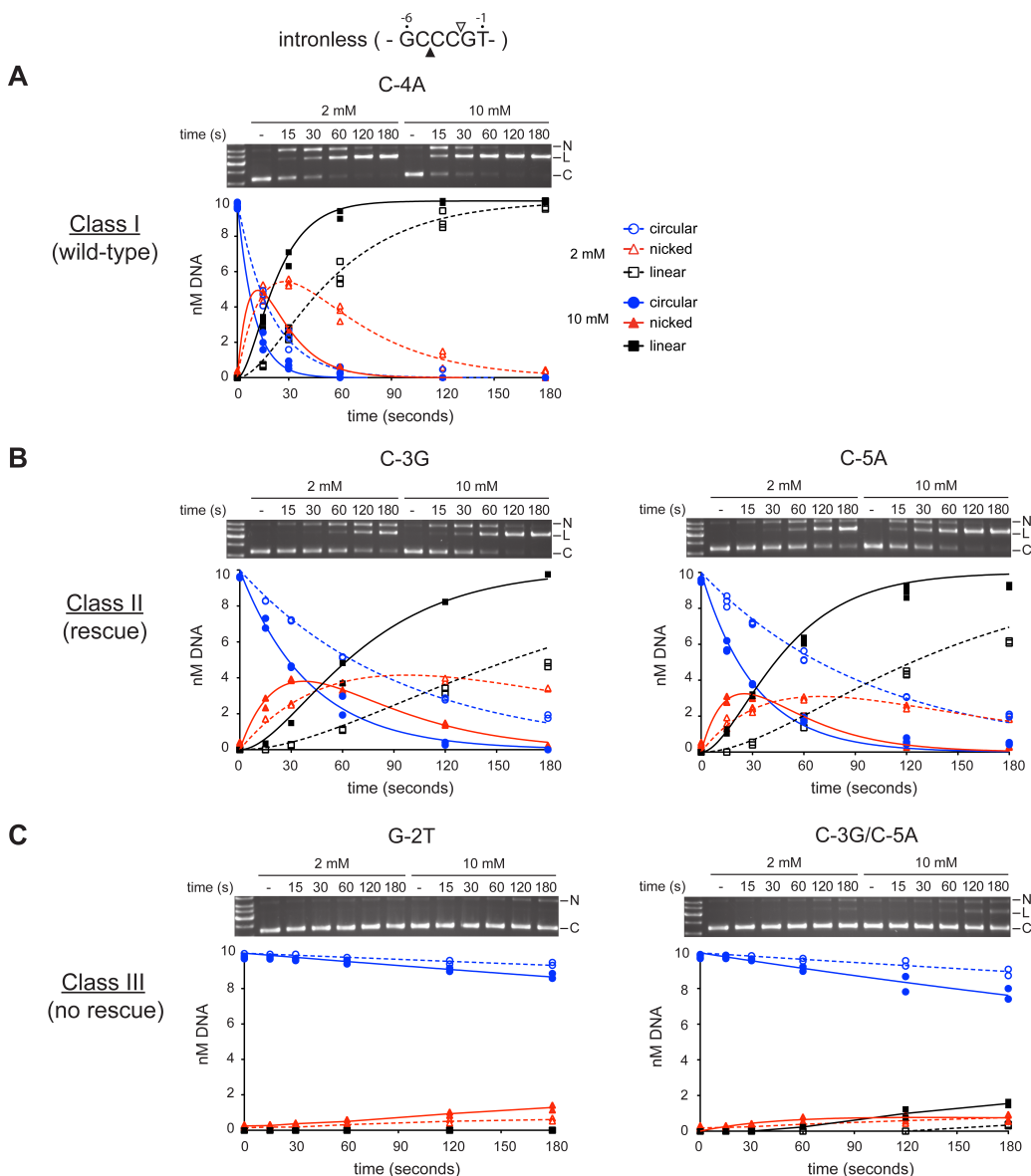


Figure 3.4: Magnesium concentrations reveal three distinct classes of substitutions

Cleavage assays with I-BmoI were conducted on supercoiled plasmid substrates containing substitutions at positions -6 to -1 in the presence of 2 mM or 10 mM MgCl₂ (see also Table 2). Mutant substrates were arranged into three classes. (A) Substrates that showed a phenotype similar to wild-type intronless *thyA* substrate (class I); (B) substrates that demonstrated poor cleavage in 2 mM MgCl₂ and rescued cleavage in 10 mM MgCl₂ (class II); (C) substrates with significantly reduced or no cleavage (class III). Shown are representative gel images of time-course cleavage assays, where the second lane from the left contains unreacted plasmid substrate (-). Nicked (N), linear (L), and circular (C) plasmid forms indicated to the right. Beneath each gel image is a graphical representation of reaction progress over time in 2 mM and 10 mM MgCl₂ using dashed and solid lines, respectively. Data points representing three independent experiments for class I and II, and two experiments for class III are shown.

(Figure 3.4A). The substrates were those with substitutions C-4A or G-6T, and the rate constants for the first and second strand nicking reactions in low and high MgCl_2 were very similar to those obtained for intronless *thyA* substrate (Table 3.2 and Supplementary Table S3.3).

Class II mutants were defined as substrates that exhibited a greater impairment of first and second strand nicking in low versus high MgCl_2 conditions. This class included the singly mutated substrates T-1G, C-3G, and C-5A, as well as the multiply substituted C-3G/C-4A, C-4A/C-5A, C-5A/G-6T and C-4A/C-5A/G-6T substrates (Table 3.2). The general characteristics of substrates in this class were the accumulation and persistence of the nicked intermediate in 2 mM MgCl_2 (which resulted in impaired double-strand break formation), and the rescue of cleavage in 10 mM MgCl_2 resulting in the near complete conversion to linear product by 180 seconds (Figure 3.4B and Table 3.2). Rate constant analysis indicated a further segregation of substrates within class II by differential base-specific effects on first and second strand nicking reactions (Table 3.2). The T-1G substitution decreased the first (k_1) and second (k_2) strand rates by roughly 2- and 3-fold versus intronless *thyA* substrate in the presence of 2 mM and 10 mM MgCl_2 , respectively, suggesting a more profound modulation of second-strand nicking. The C-3G substitution, which displayed similar defects in the presence or absence of a C-4A substitution, generated approximately 5- and 3- fold reductions in k_1 and k_2 versus intronless substrate, regardless of MgCl_2 concentration. Interestingly, the C-5A substitution, alone or in the context of a G-6T substitution, decreased k_1 by roughly 6- and 3-fold in the presence of 2 mM and 10 mM MgCl_2 , respectively. The addition of a C-4A substitution to the C-5A template exacerbated the k_1 defect to 12- and 9-fold in the presence of 2mM and 10 mM MgCl_2 . Surprisingly, the C-5A substitution showed less of an influence on the second-strand nicking rate (k_2), evidenced by a 1.5-fold rate reduction in both MgCl_2 concentrations. This result was observed with the C-5A substitution alone or in the context of the double mutation with C-4A.

Class III mutants were defined as substrates with drastically reduced reaction rates that displayed little or no cleavage in either low- or high-metal ion conditions (Figure 3.4C and Table 3.2). All substrates with mutations at the -2 position fell into this category, in

addition to substrates with mutations at both the -5 and -3 positions. Interestingly, we observed that the G-2T substitution alone, and in combination with substitutions at positions -1 or -4 and -6, demonstrated limited accumulation of nicked intermediate at 180s in high metal. First strand nicking rates (k_1) for class III substrates were 84- to 314-fold reduced as compared to intronless *thyA* substrate, indicating a substantial first strand defect. Additionally, we observed that the substrates containing a combination of C-3G and C-5A mutations retained low levels of double-strand break formation after 180s in 10 mM MgCl₂, and that the addition of a T-1G or C-4A mutation did not abolish the low amount of cleavage. Importantly, the intron-containing substrate showed no evidence of nicking or cleavage after 180 seconds under either condition tested (Table 3.2).

Collectively, the cleavage assays indicated that substrates with individual mutations at positions -4 or -6 had little effect on first- or second-strand nicking in either low or high metal ion conditions. Additionally, substrates with single mutations at positions -1, -3 or -5 exhibited defects in both first- and second-strand nicking in low metal conditions and an observable rescue of cleavage by high divalent metal ion. Substrates with mutations at both the -3 and -5 positions were not significantly rescued by high metal ion, and those with mutations at position -2 were defective for cleavage under all conditions tested.

3.3.4 In-gel footprinting reveals multiple minor groove distortions dependent on GC-2

The cleavage data suggested that divalent metal ion and the GC-2 base pair are required for efficient double-strand break formation. To gain further insight into the role of the GC-2 base pair in the cleavage pathway, we performed in-gel footprinting with the minor groove-specific reagent 1,10-copper phenanthroline (OP-Cu). I previously used in-gel footprinting to show that significant OP-Cu hypersensitive sites were localized to positions -2 and -1 of the bottom strand of the I-BmoI-*thyA* complex (24), consistent with protein-induced DNA distortions that make the minor groove more accessible to OP-Cu. The distortions were also present in I-BmoI-substrate complexes formed with the catalytic mutants R27A and E74A, implying that the distortions precede first-strand nicking and do not require the presence of a metal ion bound by E74 (24). Previous

footprinting studies, however, did not explore the requirement of the GC-2 base pair in formation of OP-Cu hypersensitive sites.

We performed in-gel footprinting with a 74-mer duplex oligonucleotide substrate corresponding to the intron-containing substrate (In+), which contains an alternative sequence upstream of the insertion site relative to the intronless substrate and notably has a G-2T substitution (Figure 3.5). As shown in Figures 3.5B and 3.5C, the OP-Cu hypersensitive sites at positions -2 and -1 on the bottom strand were significantly reduced in the UC consisting of I-BmoI-In+ substrate as compared to UC formed with intronless *thyA* substrate. We next mutated the GC-2 base pair within the context of the intronless substrate to TA-2, AT-2 or CG-2, and measured the extent of the OP-Cu hypersensitive sites at positions -2 and -1 on the bottom strand. These three mutations reduced the OP-Cu hypersensitive sites to the same extent as was observed with intron-containing substrate (Figure 3.5B and 3.5C), indicating the GC-2 base pair is critical for inducing DNA distortions that result in enhanced OP-Cu sensitivity. To provide further evidence for this model, we tested a T-2G substitution within the context of the intron-containing substrate (Figure 3.5A), reasoning that this mutation should restore the OP-Cu hypersensitivity if contacts to the GC-2 base pair are required for minor-groove distortions. Indeed, footprinting reactions on the T-2G intron-containing substrate revealed a restoration of OP-Cu hypersensitivity at positions -2 and -1 to approximately half of those observed on *thyA* intronless substrate. Products seen at the -4 position are indicative of bottom-strand nicks observed when I-BmoI-*thyA* complexes were formed in-gel in the absence of exogenously added metal (24), as Cu²⁺ is not a productive divalent metal ion for cleavage (Figure 3.2A).

3.3.5 Modeling of an I-BmoI-substrate complex

To gain insight into the role of divalent metal ion and substrate contacts in the I-BmoI reaction pathway, we used the recently solved co-crystals of the GIY-YIG restriction enzymes Eco29kI and Hpy188I with their respective substrates to model an I-BmoI-substrate complex (36,37). Eco29kI and Hpy188I function as dimers to cleave palindromic sites, with each GIY-YIG monomer nicking one strand of the substrate(37,39). We first generated a homology model of residues 1-92 of I-BmoI based

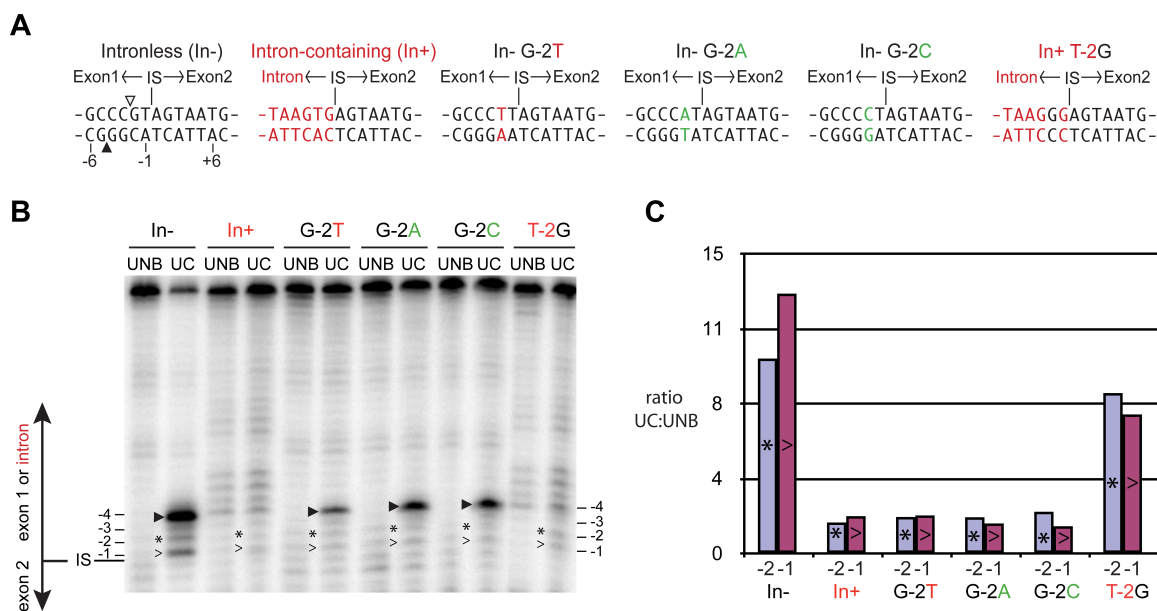


Figure 3.5: The GC-2 base pair is required for minor-groove distortions near the bottom-strand nick site

(A) Illustration of base pairs -6 to +7 of the intronless (In-), intron-containing (In+), and mutant 74-mer substrates used for in-gel 1,10-phenanthroline copper (OP-Cu) footprinting reactions. First- and second-strand nicking sites are indicated by open and filled triangles, respectively. (B) Representative denaturing gel image of OP-Cu footprinting reactions on bottom-strand labeled In-, In+, and mutant substrates. Nicked products at -4 and minor groove distortions at positions -2 and -1 are indicated using filled triangles, asterisks, and greater-than symbols, respectively. Sequence upstream of the intron insertion site (IS) varies whether the intronless or intron-containing substrate was used as a template. (C) Graphical representation of the minor groove sensitivity to OP-Cu at positions -2 (asterisk, blue) and -1 (greater than symbol, purple) for In-, In+, and mutant substrates, expressed as the ratio of normalized phosphorimager units of the upper complex (UC) and unbound substrate (UNB).

on the closely related I-TevI GIY-YIG domain structure (38). Next, we superimposed the I-BmoI model on a monomer of the dimeric Eco29kI and Hpy188I structures by aligning residues 6-10 of I-BmoI with the structurally analogous amino acids of Eco29kI (residues 47-51) and Hpy188I (residues 61-65) (which included key structural residues of the β -sheet 1). We observed that the position of the I-BmoI active site residues (Y6, Y17, R27, H31, E74 and N87) aligned well with the homologous residues in both restriction enzymes (Figure 3.6A and data not shown), as was observed in two other structure superposition studies (36,37). To model the position of the substrate in the I-BmoI active site, we included the substrate DNA from the Eco29kI structure, and similar results were obtained using the DNA from the Hpy188I structure. As shown in Figure 3.6, the path of the DNA follows a previously hypothesized catalytic cleft that is lined by the active site residues of I-BmoI (41), with the metal ion coordinated by E74 positioned in close proximity to the bottom-strand scissile phosphate at position -4. Of particular interest are the OP-Cu hypersensitive sites at positions -2 and -1 of the bottom strand that result from a widening of the minor groove. These distortions could place the base edge of the bottom strand C of the critical GC-2 base pair within hydrogen bonding distance of a number of side chains, highlighting the importance of this base pair in the cleavage pathway. The model also reinforces the notion that the role of divalent metal ion in promoting second-strand nicking must be in repositioning of the substrate-DNA complex rather than by directly participating in DNA hydrolysis.

3.4 Discussion

Enzymes that function as DNA endonucleases almost always require the presence of active site divalent metal ions (8,42,44). In general, the preferred ion is magnesium, which acts to catalyze reaction progress by stabilizing negative phosphoanion transition states, or by acting as Lewis acids to modulate the pKa of coordinated water molecules (42,44). Metal-activated water molecules can function as catalytic agents by adopting the roles of nucleophiles and general bases, or by protonating leaving groups (44,47). Families of site-specific DNA endonucleases have distinct metal ion requirements, and generally function by either a one- or two-metal ion mechanism (44,47). The metal ion requirements for homing endonucleases are best understood for the LAGLIDADG family

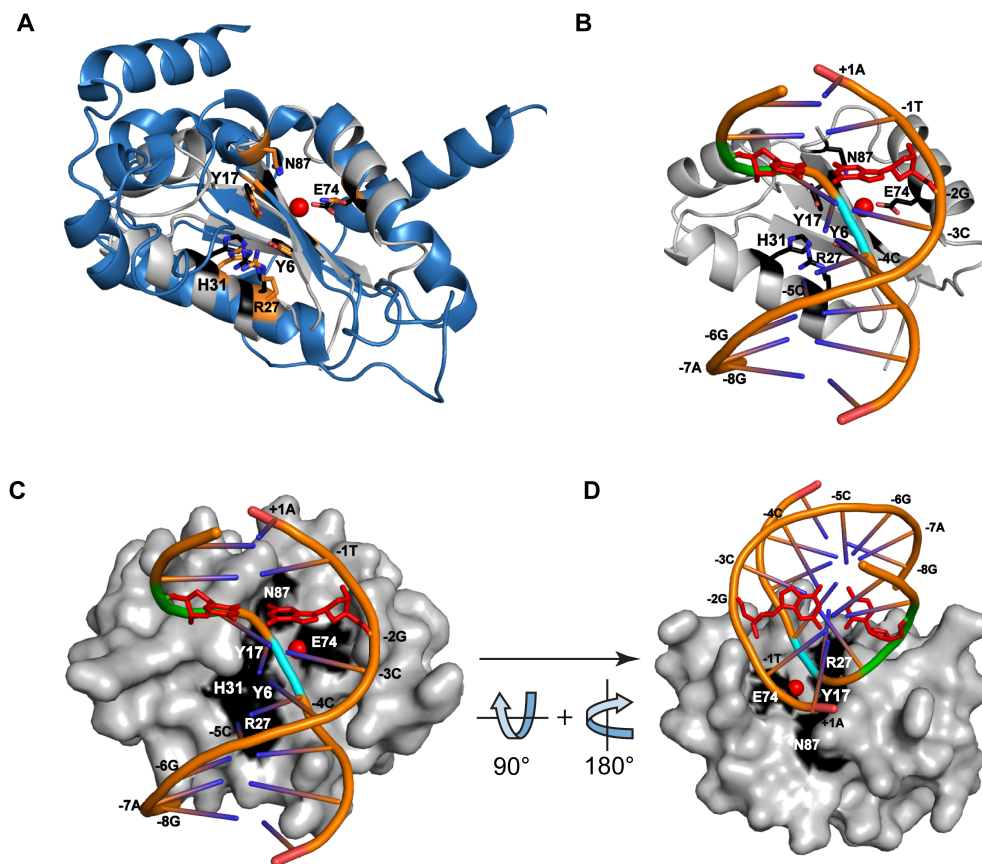


Figure 3.6: Model of I-BmoI GIY-YIG domain interactions with substrate

(A) Cartoon representation of the I-BmoI GIY-YIG domain homology model (gray) (41) aligned with the solution structure of related GIY-YIG restriction enzyme Eco29kI (3MX1, blue) (36). The side chains of the conserved and catalytically relevant residues Y6, Y17, R27, H31, E74, and N87 of I-BmoI are shown in black, and equivalent Eco29kI residues in orange (Y49, Y76, R104, H108, E142, and N154). The position of divalent metal coordinated by Eco29kI is represented by the red sphere. (B) Cartoon representation of the I-BmoI homology model aligned with a segment of the DNA from the substrate-bound Eco29kI structure (3N1C, -4C to +5C). The top strand of the Eco29kI substrate is labeled according to the nucleotides of the I-BmoI intronless allele (-8G to +1A). The phosphate of the bottom strand nick site is highlighted in blue, the bottom strand distortions at positions -2 and -1 are shown in green, and the -2GC base pair is shown as a stick model in red. (C) Surface representation of the GIY-YIG domain for the model shown in panel (B). (D) Similar to panel (C), with a 90° rotation of the model on the horizontal axis, and a 180° rotation around the vertical axis.

that have an active-site preference for magnesium, and can uniquely function by a two-metal mechanism where a third divalent metal ion is shared between two active sites (8,48,49). The H-N-H homing endonuclease I-HmuI, which generates single-stranded nicks in DNA substrates, functions optimally at 1 mM MgCl₂ or MnCl₂ and nicks via a one-metal ion mechanism (43). Studies of the related H-N-H colicin E9 revealed different divalent metal ion requirements for cleavage of distinct nucleic acid substrates, with magnesium promoting cleavage of dsDNA and zinc being more effective for nicking of ssDNA (50). I-PpoI, a His-Cys box homing endonuclease, functions by a one-metal ion mechanism and requires zinc and magnesium for folding and cleavage, respectively (8,51). The focus of this paper, the GIY-YIG homing endonucleases, have traditionally been studied in the presence of magnesium that structural studies indicate is coordinated by a conserved active site glutamate (E74 in I-BmoI) (38). GIY-YIG enzymes likely use a single-metal ion to promote DNA hydrolysis (36,37), and our data agrees with studies of the tetrameric GIY-YIG restriction enzyme Cfr42I that revealed a broad tolerance to divalent-metal ion (31).

Among the GIY-YIG family of enzymes, there is tremendous diversity in how the GIY-YIG nuclease domain is utilized to hydrolyze DNA. For instance, the restriction enzymes Eco29kI and Hpy188I function as dimers (37,39), with each GIY-YIG monomer nicking one strand, whereas the UvrC nucleotide excision repair protein uses a single GIY-YIG domain to nick 3' to a mutagenic lesion (40). In this respect, GIY-YIG homing endonucleases differ from characterized GIY-YIG enzymes that make a DSB in that a single catalytic domain is used in two ordered and sequential nicking reactions to generate a DSB (24,25). How the single active site of the GIY-YIG nuclease domain is reorganized on substrate between nicking reactions is unknown, but previous studies suggested an involvement of divalent metal ion in this process (24). Our current study more accurately defines the roles of divalent metal ion and DNA substrate bases in this process as evidenced by reaction conditions that differentially affect k_1 or k_2 . Not surprisingly, substrates that contained a mutation in the critical GC-2 base pair displayed drastically impaired rate constants, indicating a severe defect in first-strand nicking and an inability to form a DSB. Interestingly, bases surrounding the bottom-strand nicking site at positions -4 and -5 also resulted in reduced nicking rates when mutated. This

observation was somewhat surprising as previous studies had not revealed a critical role of these positions for cleavage by I-BmoI (26), likely because past studies examined only linear product formation under high metal conditions rather than employing reactions conditions that could temporally dissect base-specific nicking effects. The I-BmoI-substrate model predicts that multiple amino acid side chains within the catalytic cleft project into the major groove near positions -4 and -5, implying that base pair substitutions result in altered contacts that consequently reduce k_1 . The observation that k_1 defects associated with these substrates were significantly rescued in the presence of 10 mM $MgCl_2$ suggests that magnesium can compensate for lost I-BmoI-substrate interactions by promoting alternative conformations of protein or substrate within the catalytic cleft that facilitate cleavage. Additionally, substitutions near the bottom strand-nicking site had a comparatively insignificant effect on the top strand-nicking rate (k_2), which is in agreement with the bottom strand contacts being necessary only for k_1 .

A crucial aspect of the I-BmoI cleavage pathway is the introduction of a significant minor groove distortion in *thyA* substrate (24); similar DNA-bending was also observed in I-TevI-*td* substrate complexes (25). These protein-dependent DNA distortions are observed in the absence of exogenous magnesium, and the precise role of these distortions in the cleavage pathway is unknown. Parallels can be drawn with other site-specific endonucleases, as large DNA distortions have been shown to bring opposite strand scissile phosphates into close proximity (52). Conversely, the natural curvature of nucleic acids is highly sequence dependent and can be influenced by the presence of divalent metal ions (53-55). Interestingly, magnesium ions have been shown to induce much greater changes in the curvature of GC- versus AT-rich DNA (53). In this respect, it is noteworthy that the DNA distortions introduced by I-BmoI and I-TevI occur in stretches of GC-rich DNA (24,25), suggestive of distinct metal dependent and protein induced distortions that promote substrate conformational changes between nicking reactions. It is possible that aside from directly participating in catalysis, magnesium may induce a secondary substrate conformation that is distinct from the distortion resulting from contacts to the GC-2 base pair. Reduced rate constants observed for I-BmoI on mutant substrates can therefore be rationalized by lack of base-specific contacts to the GC-2 pair and to other positions, as well as the sequence-dependence of DNA distortions induced

by magnesium. Decreased fidelity of I-BmoI on the G-2T substrate in the presence of manganese can alternatively be explained by different protein-DNA conformations promoted by manganese ions relative to magnesium, as they have differential effects on natural DNA curvature (56).

Similar roles for metal ions in promoting protein-DNA interactions and conformational changes required for efficient catalysis have been dissected in other enzyme systems where a single active site is used sequentially to perform multiple reactions (57). Notably, sub-optimal magnesium concentrations can uncouple excision and strand transfer events, and manganese can reduce target site specificity of the Tn10 transposase (57). However, the modular structure of I-BmoI and other GIY-YIG homing endonucleases resembles that of Type IIs restriction enzymes, including the well-characterized FokI that is a monomer in solution but transiently dimerizes to generate a DSB (58,59). The FokI cleavage mechanism has been suggested as a solution to how the single active site of GIY-YIG homing endonucleases is used to affect a double-strand break (38). Interestingly, divalent metal ion has been shown to stabilize the FokI dimer/substrate complex (59), and it is possible that a reduction in the k_2 rate constant we observe in low magnesium conditions reflects a defect in transient dimerization between I-BmoI monomers. No obvious dimer interface was evident in the structure of the I-TevI catalytic domain (38), which shares high identity with that of I-BmoI, but it is possible that dimerization determinants lie outside of the catalytic domain. Interestingly, the type IIs enzyme Eco31I also has a modular structure but binds and cleaves DNA as a monomer (60), suggestive of mechanistic similarities to I-BmoI and I-TevI.

3.5 References

1. Belfort M, Derbyshire V, Cousineau B, Lambowitz A (2002) Mobile introns: pathways and proteins. In: Craig N, Craigie R, Gellert M, Lambowitz A, editors. *Mobile DNA II*. New York: ASM Press. pp. 761-783.
2. Dujon B, Belfort M, Butow RA, Jacq C, Lemieux C, et al. (1989) Mobile introns: definition of terms and recommended nomenclature. *Gene* 82: 115-118.
3. Bell-Pedersen D, Quirk S, Clyman J, Belfort M (1990) Intron mobility in phage T4 is dependent upon a distinctive class of endonucleases and independent of

- DNA sequences encoding the intron core: mechanistic and evolutionary implications. *Nucleic Acids Res* 18: 3763-3770.
4. Bell-Pedersen D, Quirk SM, Aubrey M, Belfort M (1989) A site-specific endonuclease and co-conversion of flanking exons associated with the mobile *td* intron of phage T4. *Gene* 82: 119-126.
 5. Durrenberger F, Thompson AJ, Herrin DL, Rochaix JD (1996) Double strand break-induced recombination in *Chlamydomonas reinhardtii* chloroplasts. *Nucleic Acids Res* 24: 3323-3331.
 6. Gimble FS (2000) Invasion of a multitude of genetic niches by mobile endonuclease genes. *FEMS Microbiol Lett* 185: 99-107.
 7. Haugen P, Simon DM, Bhattacharya D (2005) The natural history of group I introns. *Trends Genet* 21: 111-119.
 8. Stoddard BL (2005) Homing endonuclease structure and function. *Q Rev Biophys* 38: 49-95.
 9. Zhao L, Bonocora RP, Shub DA, Stoddard BL (2007) The restriction fold turns to the dark side: a bacterial homing endonuclease with a PD-(D/E)-XK motif. *EMBO J* 26: 2432-2442.
 10. Ashworth J, Havranek JJ, Duarte CM, Sussman D, Monnat RJ, Jr., et al. (2006) Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature* 441: 656-659.
 11. Chan SH, Stoddard BL, Xu SY (2011) Natural and engineered nicking endonucleases--from cleavage mechanism to engineering of strand-specificity. *Nucleic Acids Res* 39: 1-18.
 12. Hilario E, Gogarten JP (1993) Horizontal transfer of ATPase genes--the tree of life becomes a net of life. *Biosystems* 31: 111-119.
 13. Rosen LE, Morrison HA, Masri S, Brown MJ, Springstubb B, et al. (2006) Homing endonuclease I-CreI derivatives with novel DNA target specificities. *Nucleic Acids Res* 34: 4791-4800.
 14. Stoddard BL (2011) Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. *Structure* 19: 7-15.
 15. Derbyshire V, Kowalski JC, Dansereau JT, Hauer CR, Belfort M (1997) Two-domain structure of the *td* intron-encoded endonuclease I-*TevI* correlates with the two-domain configuration of the homing site. *J Mol Biol* 265: 494-506.
 16. Kowalski JC, Belfort M, Stapleton MA, Holpert M, Dansereau JT, et al. (1999) Configuration of the catalytic GIY-YIG domain of intron endonuclease I-*TevI*:

- coincidence of computational and molecular findings. *Nucleic Acids Res* 27: 2115-2125.
17. Bryk M, Belisle M, Mueller JE, Belfort M (1995) Selection of a remote cleavage site by I-TevI, the *td* intron-encoded endonuclease. *J Mol Biol* 247: 197-210.
 18. Dean AB, Stanger MJ, Dansereau JT, Van Roey P, Derbyshire V, et al. (2002) Zinc finger as distance determinant in the flexible linker of intron endonuclease I-TevI. *Proc Natl Acad Sci U S A* 99: 8554-8561.
 19. Liu Q, Dansereau JT, Puttamadappa SS, Shekhtman A, Derbyshire V, et al. (2008) Role of the interdomain linker in distance determination for remote cleavage by homing endonuclease I-TevI. *J Mol Biol* 379: 1094-1106.
 20. Liu Q, Derbyshire V, Belfort M, Edgell DR (2006) Distance determination by GIY-YIG intron endonucleases: discrimination between repression and cleavage functions. *Nucleic Acids Res* 34: 1755-1764.
 21. Bell-Pedersen D, Quirk SM, Bryk M, Belfort M (1991) I-TevI, the endonuclease encoded by the mobile *td* intron, recognizes binding and cleavage domains on its DNA target. *Proc Natl Acad Sci U S A* 88: 7719-7723.
 22. Edgell DR, Shub DA (2001) Related homing endonucleases I-BmoI and I-TevI use different strategies to cleave homologous recognition sites. *Proc Natl Acad Sci U S A* 98: 7898-7903.
 23. Quirk SM, Bell-Pedersen D, Belfort M (1989) Intron mobility in the T-even phages: high frequency inheritance of group I introns promoted by intron open reading frames. *Cell* 56: 455-465.
 24. Carter JM, Friedrich NC, Kleinstiver B, Edgell DR (2007) Strand-specific contacts and divalent metal ion regulate double-strand break formation by the GIY-YIG homing endonuclease I-BmoI. *J Mol Biol* 374: 306-321.
 25. Mueller JE, Smith D, Bryk M, Belfort M (1995) Intron-encoded endonuclease I-TevI binds as a monomer to effect sequential cleavage via conformational changes in the *td* homing site. *EMBO J* 14: 5724-5735.
 26. Edgell DR, Stanger MJ, Belfort M (2003) Importance of a single base pair for discrimination between intron-containing and intronless alleles by endonuclease I-BmoI. *Curr Biol* 13: 973-978.
 27. Edgell DR, Stanger MJ, Belfort M (2004) Coincidence of cleavage sites of intron endonuclease I-TevI and critical sequences of the host thymidylate synthase gene. *J Mol Biol* 343: 1231-1241.
 28. Dunin-Horkawicz S, Feder M, Bujnicki JM (2006) Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics* 7: 98.

29. Truglio JJ, Rhau B, Croteau DL, Wang L, Skovvaga M, et al. (2005) Structural insights into the first incision reaction during nucleotide excision repair. *EMBO J* 24: 885-894.
30. Pyatkov KI, Arkhipova IR, Malkova NV, Finnegan DJ, Evgen'ev MB (2004) Reverse transcriptase and endonuclease activities encoded by Penelope-like retroelements. *Proc Natl Acad Sci U S A* 101: 14719-14724.
31. Gasiunas G, Sasnauskas G, Tamulaitis G, Urbanke C, Razaniene D, et al. (2008) Tetrameric restriction enzymes: expansion to the GIY-YIG nuclease family. *Nucleic Acids Res* 36: 938-949.
32. Ibryashkina EM, Zakharova MV, Baskunov VB, Bogdanova ES, Nagornykh MO, et al. (2007) Type II restriction endonuclease R.Eco29kI is a member of the GIY-YIG nuclease superfamily. *BMC Struct Biol* 7: 48.
33. Kaminska KH, Kawai M, Boniecki M, Kobayashi I, Bujnicki JM (2008) Type II restriction endonuclease R.Hpy188I belongs to the GIY-YIG nuclease superfamily, but exhibits an unusual active site. *BMC Struct Biol* 8: 48.
34. Andersson CE, Lagerback P, Carlson K (2010) Structure of bacteriophage T4 endonuclease II mutant E118A, a tetrameric GIY-YIG enzyme. *J Mol Biol* 397: 1003-1016.
35. Karakas E, Truglio JJ, Croteau D, Rhau B, Wang L, et al. (2007) Structure of the C-terminal half of UvrC reveals an RNase H endonuclease domain with an Argonaute-like catalytic triad. *Embo J* 26: 613-622.
36. Mak AN, Lambert AR, Stoddard BL (2010) Folding, DNA recognition, and function of GIY-YIG endonucleases: crystal structures of R.Eco29kI. *Structure* 18: 1321-1331.
37. Sokolowska M, Czapinska H, Bochtler M (2010) Hpy188I-DNA pre- and post-cleavage complexes--snapshots of the GIY-YIG nuclease mediated catalysis. *Nucleic Acids Res* 39: 1554-1564.
38. Van Roey P, Meehan L, Kowalski JC, Belfort M, Derbyshire V (2002) Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-*TevI*. *Nat Struct Biol* 9: 806-811.
39. Ibryashkina EM, Sasnauskas G, Solonin AS, Zakharova MV, Siksnys V (2009) Oligomeric structure diversity within the GIY-YIG nuclease family. *J Mol Biol* 387: 10-16.
40. Verhoeven EE, van Kesteren M, Moolenaar GF, Visse R, Goosen N (2000) Catalytic sites for 3' and 5' incision of *Escherichia coli* nucleotide excision repair are both located in UvrC. *J Biol Chem* 275: 5120-5123.

41. Kleinstiver BP, Fernandes AD, Gloor GB, Edgell DR (2010) A unified genetic, computational and experimental framework identifies functionally relevant residues of the homing endonuclease I-BmoI. *Nucleic Acids Res* 38: 2411-2427.
42. Perona JJ (2002) Type II restriction endonucleases. *Methods* 28: 353-364.
43. Shen BW, Landthaler M, Shub DA, Stoddard BL (2004) DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J Mol Biol* 342: 43-56.
44. Yang W, Lee JY, Nowotny M (2006) Making and breaking nucleic acids: two-Mg²⁺-ion catalysis and substrate specificity. *Mol Cell* 22: 5-13.
45. Vermote CL, Halford SE (1992) EcoRV restriction endonuclease: communication between catalytic metal ions and DNA recognition. *Biochemistry* 31: 6082-6089.
46. Allingham JS, Haniford DB (2002) Mechanisms of metal ion action in Tn10 transposition. *J Mol Biol* 319: 53-65.
47. Galburt EA, Stoddard BL (2002) Catalytic mechanisms of restriction and homing endonucleases. *Biochemistry* 41: 13851-13860.
48. Chevalier B, Sussman D, Otis C, Noel AJ, Turmel M, et al. (2004) Metal-dependent DNA cleavage mechanism of the I-CreI LAGLIDADG homing endonuclease. *Biochemistry* 43: 14015-14026.
49. Chevalier BS, Monnat RJ, Jr., Stoddard BL (2001) The homing endonuclease I-CreI uses three metals, one of which is shared between the two active sites. *Nat Struct Biol* 8: 312-316.
50. Pommer AJ, Cal S, Keeble AH, Walker D, Evans SJ, et al. (2001) Mechanism and cleavage specificity of the H-N-H endonuclease colicin E9. *J Mol Biol* 314: 735-749.
51. Galburt EA, Chevalier B, Tang W, Jurica MS, Flick KE, et al. (1999) A novel endonuclease mechanism directly visualized for I-PpoI. *Nat Struct Biol* 6: 1096-1099.
52. Jurica MS, Monnat RJ, Jr., Stoddard BL (1998) DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-CreI. *Mol Cell* 2: 469-476.
53. Brukner I, Susic S, Dlakic M, Savic A, Pongor S (1994) Physiological concentration of magnesium ions induces a strong macroscopic curvature in GGGCCC-containing DNA. *J Mol Biol* 236: 26-32.
54. Rouzina I, Bloomfield VA (1998) DNA bending by small, mobile multivalent cations. *Biophys J* 74: 3152-3164.

55. Sreedhara A, Cowan JA (2002) Structural and catalytic roles for divalent magnesium in nucleic acid biochemistry. *Biometals* 15: 211-223.
56. Duguid J, Bloomfield VA, Benevides J, Thomas GJ, Jr. (1993) Raman spectroscopy of DNA-metal complexes. I. Interactions and conformational effects of the divalent cations: Mg, Ca, Sr, Ba, Mn, Co, Ni, Cu, Pd, and Cd. *Biophys J* 65: 1916-1928.
57. Junop MS, Haniford DB (1996) Multiple roles for divalent metal ions in DNA transposition: distinct stages of Tn10 transposition have different Mg²⁺ requirements. *EMBO J* 15: 2547-2555.
58. Wah DA, Bitinaite J, Schildkraut I, Aggarwal AK (1998) Structure of FokI has implications for DNA cleavage. *Proc Natl Acad Sci U S A* 95: 10564-10569.
59. Vanamee ES, Santagata S, Aggarwal AK (2001) FokI requires two specific DNA sites for cleavage. *J Mol Biol* 309: 69-78.
60. Jakubauskas A, Sasnauskas G, Giedriene J, Janulaitis A (2008) Domain organization and functional analysis of type IIS restriction endonuclease Eco31I. *Biochemistry* 47: 8546-8556.
61. Rubin DB (1981) The Bayesian Bootstrap. *The Annals of Statistics* 9: 130-134.
62. Miller WE (2002) Revisiting the Geometry of a Ternary Diagram With the Half-Taxi Metric. *Mathematical Geology* 34: 275-290.

Chapter 4

4 The monomeric GIY-YIG homing endonuclease I-BmoI uses a molecular anchor and a flexible tether to sequentially nick DNA

The work presented in this chapter is reproduced from (with permission, Appendix 1) the accepted version of:

Kleinstiver, B.P., Wolfs, J. W., Edgell, D.R. (2013) The monomeric GIY-YIG homing endonuclease I-BmoI uses a molecular anchor and a flexible tether to sequentially nick DNA. *Nucleic Acids Research*, doi: 10.1093/nar/gkt186

4.1 Introduction

Site-specific DNA nucleases are central to many cellular processes, including DNA recombination and repair, restriction defense systems, and the mobility of selfish genetic elements (1,2). A number of nuclease active sites have evolved that are characterized by conserved amino acid motifs and operate via distinct catalytic mechanisms to generate nicks or double-strand breaks (DSBs) in DNA (1-3). There is often little correlation between the identity of the nuclease active site motif and cellular function, as conserved catalytic cores from individual nuclease families are found within diverse protein scaffolds (1,4). As such, a fundamental understanding of the mechanisms by which each nuclease family generates a DSB is required, as several families have been exploited for genome engineering purposes (5-9).

One such nuclease family contains the GIY-YIG catalytic motif, consisting of a compact domain of ~100 amino acids that is characterized by at least two α -helices that surround a three-stranded β -sheet (10-14). The catalytic core includes a number of conserved amino acids that are essential for DNA-hydrolysis via a one-metal ion mechanism (14). The minimal GIY-YIG nuclease domain is often found associated with protein scaffolds that function to anchor the GIY-YIG domain in close proximity to DNA, implying that the GIY-YIG domain itself has weak DNA-binding activity (15-20). Included in this category is the nucleotide excision repair protein UvrC that has acquired distinct binding domains and uses a single GIY-YIG domain to nick 3' to DNA lesions (11).

Alternatively, the restriction enzymes Eco29kI and Hpy188I have acquired supplementary units of structure that are intertwined within the conserved GIY-YIG domain to promote oligomerization and assembly on DNA, with each GIY-YIG monomer nicking one strand (13,14,19,21). The GIY-YIG nuclease domain is also found within homing endonucleases (GIY-HEs) as an N-terminal fusion to sequence-tolerant DNA-binding domains that target extended asymmetric binding sites (22,23).

GIY-HEs function as mobile genetic elements by introducing DSBs at a defined site within naïve genomes to promote mobility of their own genes (2,24,25). I-TevI and I-BmoI, encoded within introns interrupting the *td* and *thyA* thymidylate synthase (TS) genes of bacteriophage T4 and *Bacillus mojavensis*, remain the best-characterized GIY-HEs to date, however sequencing efforts have identified additional putative GIY-HEs (26-28). I-BmoI and I-TevI bind a homologous target site in their respective TS genes and consist of N-terminal GIY-YIG domains connected by linkers to C-terminal domains that contact 30-35 base pairs of substrate (Figure 4.1A) (15,23,29). Studies of I-TevI indicate that the inter-domain linker is flexible (29,30), supported by its ability to position the catalytic domain on substrate. The modularity of GIY-HEs is also evidenced in their cognate target sites, as the asymmetric target sequences contain distinct motifs for DNA binding and cleavage (15). The I-BmoI and I-TevI target sites share 48% identity and previous *in vitro* substrate selections identified a conserved G-C base pair that flanks the I-BmoI and I-TevI cleavage sites as critical for cleavage (Figure 4.1A) (31,32). Contacts to the G-C base pair are essential for both GIY-HEs to perform the sequential and independent nicking reactions to generate a DSB with a 2-nt 3' overhang (23,33,34). Previous studies with I-TevI showed that the two-domain enzyme bound DNA as a monomer and induced significant substrate bending near the cleavage site (35), while chimeric enzymes with the I-TevI nuclease domain fused to the *ryA* zinc-finger generated a DSB in a non-cooperative manner (9). Kinetic analyses of the two-step nicking reaction using I-BmoI showed that the first-strand nicking reaction proceeds approximately 2-fold faster than the second nicking reaction (28). Moreover, divalent metal ion and DNA sequence at the cleavage site influences I-BmoI activity, likely by modulating local DNA structure and promoting sequential protein interactions with bases in the cleavage site region (36). However, the central question of how the single active

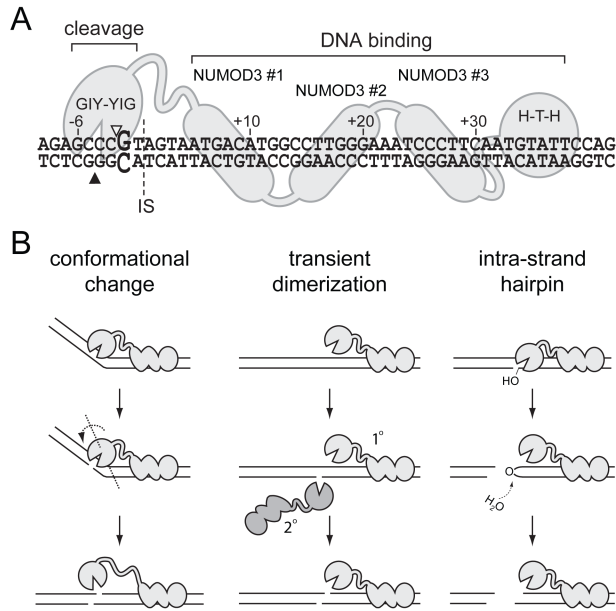


Figure 4.1: I-BmoI interactions with substrate and monomeric cleavage models

(A) Schematic of the modular structure of I-BmoI interacting with intronless *thyA* substrate. Bottom- and top-strand nicking sites are indicated by filled and open triangles, respectively. IS, intron-insertion site. NUMOD; nuclease-associated modular DNA-binding domains. (B) Models for double-strand break formation by monomeric nucleases with single active sites. For transient dimerization, the secondary enzyme molecule can dimerize with the primary DNA-bound molecule from solution or via synapsis as a substrate-bound molecule at an additional target site.

site of GIY-HEs is utilized to successively nick each DNA strand has not been rigorously addressed. Based largely on studies with I-TevI, a conformational change model was proposed whereby the GIY-YIG nuclease domain and substrate undergo substantial rearrangements between nicking reactions to reposition the active site (Figure 4.1B) (35). However, the extreme cytotoxicity of I-TevI precluded detailed kinetic studies that could rule out alternative mechanisms of DSB formation by monomeric nucleases, including transient dimerization between catalytic domains, synapsis of two enzymes bound to separate target sites, or cleavage through sequential transesterification reactions (Figure 4.1B) (37-39).

Here I capitalize on the ability to purify wild-type I-BmoI to explicitly test models of DSB formation. Significantly, we show that I-BmoI functions as a monomer at all stages of the reaction pathway and does not use a hairpin mechanism for DNA cleavage. The data are consistent with a model where the monomeric GIY-YIG domain rotates between nicking sites to generate a DSB, with the I-BmoI DNA-binding domain functioning to anchor the linker and nuclease domain to substrate. I envision that a better understanding of the catalytic mechanism will provide insight into the ability of GIY-HEs to invade a diverse range of biological alleles, as well as enhance ongoing efforts to utilize the GIY-YIG nuclease domain as a genome-editing tool.

4.2 Materials and methods

4.2.1 Strain and plasmid construction

A complete description of strains and plasmids used in this study are found in Supplementary Table S4.1. *E.coli* strains DH5 α , ER2566, and BW25141(λ DE3) were used for plasmid manipulations, expression studies, and genetic selections, respectively. Sequences and details of oligonucleotides are in Supplementary Table S4.2. Full-length I-BmoI, R27A I-BmoI, and N- and C-terminal truncations were expressed and purified as previously described (28). Codon optimized N- and C-terminal I-BmoI domain truncations were cloned by PCR using Phusion (N.E.B.) into the NdeI and KpnI sites of pTYB1. To construct plasmid pLBmo, the PvuII fragment of pBmoHS was subcloned into the PvuII site of LITMUS28i (N.E.B.). A duplex 49mer oligonucleotide containing

the I-BmoI *thyA* target site (DE-906/907) was ligated into the SmaI site of pLBmo in the opposite or same orientation as the primary I-BmoI target site to generate the 2-site plasmids pLBmo2a and pLBmo2b, respectively. Quikchange (Stratagene) reactions were performed on pLBmo to generate pLBmoNa and pLBmoNb by introducing Nt.BbvCI sites 1082 or 60 base pairs from the I-BmoI top-strand nick site, respectively. Both Nt.BbvCI sites were introduced into pLBmo to generate pLBmo2N.

To create the pTox plasmid, Quikchange was performed (DE-1171/1172) to remove the EcoRI site from p11-lacY-wtx1 (40) and an oligo cassette containing an EcoRI site (DE-1173/1174) was ligated into XbaI/SphI. The kanamycin-resistant pKox plasmid was generated by cloning the kanamycin resistance gene into the ScaI site of pTox. The vectors pKoxBmoHS and pKoxBmoIn+ were similarly generated from their respective pTox precursors (28). pTox and LITMUS28i plasmids containing variant I-BmoI homing sites with G+7A, G+7C, or G+7T mutations were generated by ligating the appropriate duplex oligos into XbaI/EcoRI of either vector. Plasmid constructs with nucleotide insertions in the I-BmoI homing site were generated by ligating the appropriate oligos into XbaI/EcoRI of LITMUS28i. All plasmid constructs were verified by sequencing.

4.2.2 Identification of I-BmoI domains

Limited proteolysis trials were conducted in reactions containing 1X binding buffer (50 mM Tris-HCl pH 8.0, 50 mM NaCl, 1mM DTT), 5 μ M I-BmoI, and 5.5 μ M intronless *thyA* (DE-116/117), intron-containing (DE-444/445), non-specific (DE-144/144GC), or mutant (see below) 74-mer duplex oligonucleotide substrates. Reactions were incubated at room temperature for 5 minutes and an aliquot ('0' time-point) was extracted prior to the addition of either 50 ng trypsin, 100 ng chymotrypsin, 5 ng subtilisin A, or 1 μ g elastase in 1X protease dilution buffer (20 mM Tris-HCl pH 7.9, 50 mM KCl, 10 mM MgSO₄). Aliquots were removed from the digest, added to stop buffer (100 mM EDTA pH 8.0, 150 mM Tris-HCl pH 6.8, 4.8 % SDS, 24% glycerol), and boiled at 95°C prior to electrophoresis on 15% SDS-PAGE gels. Coomassie stained bands of stable domains were excised prior to in-gel trypsin digestion at the London Regional Proteomics Facility. Recovered peptides were subjected to MALDI-TOF mass spectrometry and the results

analyzed using the UCSF MS-Bridge webpage. To delineate the boundary of the trypsin- and elastase-stable catalytic domain, LC mass spectrometry was performed on a 3Q mass spectrometer (Micromass) equipped with a Z-spray source and run in positive ion nanospray mode. Results were analyzed using MassLynx 4.0.

4.2.3 Biophysical characterization of I-BmoI

Differential scanning calorimetry was performed using a VP-DSC MicroCalorimeter (MicroCal) equilibrated with buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5% glycerol) in a 500 μ l cell at the Biomolecular Interactions and Conformations Facility at the University of Western Ontario. Trials consisted of 5.2 μ M I-BmoI in the presence or absence of 5.9 μ M *thyA* 74-mer duplex oligonucleotide substrate, with a heating rate of 1°C/minute from 10-100°C. Data were analyzed with MicroCal Origin 7 software (including corrections for buffer baseline and protein concentration) using a two-state model that assumes a single independent unfolding transition (41).

4.2.4 DNA-binding affinity of I-BmoI domains

Binding reactions contained 250 mM NaCl, 50 mM tris pH 8.0, 2.5 mM EDTA, 1 mM DTT, 5 ng/ μ l poly dIdC, 5% glycerol, and 0.25 nM of intronless *thyA*, intron-containing, or randomized 74-mer substrates radio labeled on the top strand with γ -³²P. Serial dilutions of the I-BmoI domain truncations were added to the reaction mixture prior to incubation at room temperature for 15 minutes and the addition of loading buffer. Reaction were loaded on 10% native polyacrylamide (19:1) and visualized using a Phosphorimager (GE Healthcare). Data were fit to the equation $\frac{[AB]}{[A]_{total}} = \frac{[B]}{[B]+K_D}$.

4.2.5 Gel filtration chromatography

I-BmoI and I-BmoI:*thyA* complexes (DE-130/131) were incubated at room temperature for 10 minutes, spun for 10 minutes at 9300x g, applied to a Superose 12 10/300 GL column (Amersham Biosciences) pre-equilibrated with dialysis buffer (25mM Tris-HCl pH 8.0, 500mM NaCl and 5% glycerol), run at 4°C at a flow rate of 0.5 ml/min. Fractions of 0.25 ml were collected and run on a 12% SDS-PAGE gel (29:1 (w/w) acrylamide to bis-acrylamide) and a 1% agarose gel. Protein standards used to calibrate the column:

carbonic anhydrase, 29 kDa; albumin, 66 kDa; β -amylase, 200 kDa; and apoferritin, 443 kDa. The calibration curve was used to calculate the molecular masses of I-BmoI, *thyA* substrate, and the I-BmoI:*thyA* complex by interpolating elution volumes onto the curve.

4.2.6 Cleavage assays with plasmid substrate

The dependence of the initial reaction velocity on protein concentration was determined by cleavage reactions containing 20 mM Tris-HCl pH 8.0, 250 mM NaCl, 2 mM MgCl₂, 2.5% glycerol, 10 nM pBmoHS, and I-BmoI (5.5 nM to 175 nM). Aliquots were removed at time points into stop dye (100 mM EDTA, 25% glycerol, 0.2% bromophenol blue), heated for 5 minutes at 95°C, electrophoresed on a 1% agarose gel, and stained in ethidium bromide (Caledon) solution prior to analysis on an AlphaImagerTM3400 (Alpha Innotech). Rate constants were calculated as best fit value of at least three experiments, as previously described (9). Initial reaction velocities were determined by calculating the slope of the line for the early time points of product formation. All subsequent plasmid cleavage assays were conducted as described above with exceptions indicated below.

Two-site plasmid cleavage assays were performed on the single-site pLBmo plasmid and two-site pLBmo2a and pLBmo2b plasmids. Reactions were supplemented with either 2 mM or 10 mM MgCl₂, and I-BmoI was added to a final concentration of 22.1 nM, 54.3 nM, or 87.5 nM. Domain addition assays were performed on pBmoHS with 10 mM MgCl₂. I-BmoI was added to 11 nM or 22 nM in reactions containing 11 μ M N111 I-BmoI, N130 I-BmoI, or BSA (1000- or 500-fold excess domain, respectively). Assays on plasmids with nucleotide insertions or G+7 substitutions were performed on pLBmo, pLBmo+5T, pLBmo+5C, pLBmo+3Ta, pLBmo+3Tb, pLBmo+3Ca, pLBmo+3Cb, pLBmoG+7A, pLBmoG+7C, and pLBmoG+7T. Reactions were supplemented with 0.5 mM, 2 mM, or 10 mM MgCl₂, and 87.5 nM I-BmoI.

4.2.7 Gel-mobility shift assays with pre-nicked substrates

Binding reactions were assembled in the presence of 10 mM EDTA or 10 mM MgCl₂ and run as previously described (34). To generate the 3 substrates required, the top strand 74mer (DE-116) was 5' labeled with γ -³²P prior to annealing with the complement full-

length bottom-strand oligo (DE-117) to generate the *thyA* substrate (WT). The substrate with the 3'-OH -5/-4 bottom strand nick was generated with a 3'-23mer (DE-125) and 5'-51mer (DE-123) (-4 nick). To create the -5/-4 bottom strand nick substrate with a 3'-H (-4 ddGTP), the 3'-23mer (DE-125) and 5'-50mer (DE-124) that had been extended with ddGTP by terminal transferase (N.E.B.) were used. To ensure addition of a single ddGTP nucleotide to DE-124, we analyzed the product of the terminal transferase reaction by denaturing electrophoresis that revealed only a single nucleotide addition as compared to an unreacted DE-124 oligonucleotide.

4.2.8 *thyA* cleavage site and spacer selection

The 4 bp upstream and 9 bp downstream of G-2 in the *thyA* substrate were randomized to generate the pKoxRCS library (DE-1228, GCGGAATTTCGANNNGNNGNNNNNNNNNNCATGGCCTTGGG-AAATCCCTTCAATGTATTCCGGCATG). A 5'-phosphorylated primer (DE-1229) complimentary to the underlined sequence was annealed to leave a 4-nt 3' SphI overhang. The library was made double stranded by extension with 3'→5' exo⁻ Klenow Fragment (N.E.B.), digested with EcoRI, and ligated into EcoRI/SphI cut pKox. The potential complexity of the library was estimated to be approximately 1.6×10^4 based on the number of independent transformants.

To screen pKoxRCS for cleavable target sites, 60 μ l of BW25141(λ DE3) cells harboring pACYCIBmoI were transformed with 200 ng of pKoxRCS. Transformations were allowed to recover in 500 μ l SOC at 37°C prior to plating on non-selective media (LB plus 25 μ g/ml chloramphenicol, 50 μ g/ml kanamycin, and 0.2% glucose). 2400 transformant colonies, representing ~15% of the library, were replica-gridded on non-selective and selective media (LB plus 25 μ g/ml chloramphenicol and 10 mM L-(+)-arabinose). Colonies containing cleavable substrates survived on both media, whereas those with non-cleavable substrates survived only on non-selective media. 86 colonies were verified as cleavable by replica-gridding in triplicate, and 96 clones were verified as non-survivors. The survivors and non-survivors were grown in separate 96-well plates in 1 ml LB + 50 μ g/ml kanamycin and 0.2% glucose, and 200 μ l from each well was pooled prior to DNA extraction to generate survivor and non-survivor plasmid libraries.

Samples were prepared for Ion Torrent sequencing at the London Regional Genomics Centre by PCR amplifying the survivor, non-survivor, and pKoxRCS libraries with PWO (Roche) for 25 cycles with barcoded PCR primers. Custom Perl scripts were used to interpret the sequencing results, revealing 9982 unique sequences in the input library with an approximately equal distribution of all four nucleotides at each position. Sequences that occurred greater than 10 times in the survivor and non-survivor pools were retained for further analyses only if the sequences were also present in the input library. Nucleotide proportions at each position in the input and survivor datasets were used to calculate a Position-Specific Scoring Matrix (PSSM), and used to display relative enrichment by calculating the differences in proportions for each nucleotide at each position between the datasets.

4.2.9 *in vivo* survival assays

Two plasmid *in vivo* substrate activity assays were performed as previously described (28), with toxic (reporter) plasmids containing the wild-type *thyA* I-BmoI target site or variants with G+7 mutations. Briefly, the *in vivo* I-BmoI cleavage efficiency for each substrate was calculated by dividing the number of colonies observed on selective plates (LB plus 25 μ g/ml chloramphenicol and 10 mM L-(+)-arabinose) by those on non-selective plates (LB plus 25 μ g/ml chloramphenicol).

4.2.10 OP-Cu in-gel footprinting

1,10-phenanthroline copper (OP-Cu) footprinting experiments were performed using 10 pmol of the catalytically inactive R27A variant of I-BmoI, as previously described (34). Duplex 74-mer substrates (0.1 pmol) used include: *thyA*, or *thyA* with G+7A (DE-1356/1357), G+7C (DE-1358/1359), or G+7T (DE-1360/1361) substitutions. Hypo- and hyper-sensitivity to OP-Cu was calculated as the pixel density ratio between bands in the protein:DNA complex (UC) and unbound substrate lanes. Traces were normalized using bands outside of the footprinting region (C-12 and A-11).

4.3 Results

4.3.1 I-BmoI is partially disordered in the absence of substrate

Previous bioinformatic and experimental studies have suggested that I-BmoI is a modular protein composed of distinct N- and C-terminal domains connected by a linker (23,42,43). To investigate domain structure changes in the presence of substrate, we performed limited proteolysis experiments with I-BmoI or I-BmoI in complex with its cognate *thyA* substrate, derived from the intronless version of the *B. mojavensis thyA* gene (22). Trypsin, elastase, and chymotrypsin digests of I-BmoI revealed a pattern of proteolysis converging on a single stable domain of ~10 kDa (Figure 4.2A, Supplementary Figure S4.1A). Solution mass-spectrometry, as well as in-gel trypsin digest followed by mass spectrometry identified the stable domain as the N-terminal GIY-YIG domain consisting of residues 1-92. Strikingly, protease digests of I-BmoI in complex with its cognate 74-mer *thyA* substrate under identical conditions revealed a different pattern of protease sensitivity. For all three proteases, substrate-bound I-BmoI was far more resistant to proteolysis (Figure 4.2A, Supplementary Figure S4.1A). Additional stable domains of ~18-20 kDa were also observed and subsequently identified by in-gel trypsin digestion followed by mass spectrometry as peptides that mapped C-terminal to residue 130. Western blot analyses using an antibody raised against residues 1-111 of I-BmoI (N111) confirmed the identities of both stable domains (Supplementary Figure S4.1B). Digests of I-BmoI with all 3 proteases in the presence of a non-cognate oligonucleotide substrate revealed a pattern of protease sensitivity similar to that observed for I-BmoI without substrate (data not shown). No change in protease sensitivity was observed under cleavage conditions with 10 mM MgCl₂. The protease digestion patterns of the I-BmoI:*thyA* complex revealed a stable two-domain structure resulting from specific DNA binding where the inter-domain linker remains moderately accessible to proteases.

An exception was observed when digests were performed with intron-containing *thyA* substrate. I-BmoI binds the intron-containing substrate with similar affinity to the cognate intronless allele, however nucleotide differences upstream of the intron insertion site within the intron-containing target abolish the I-BmoI cleavage site (23). Trypsin digests

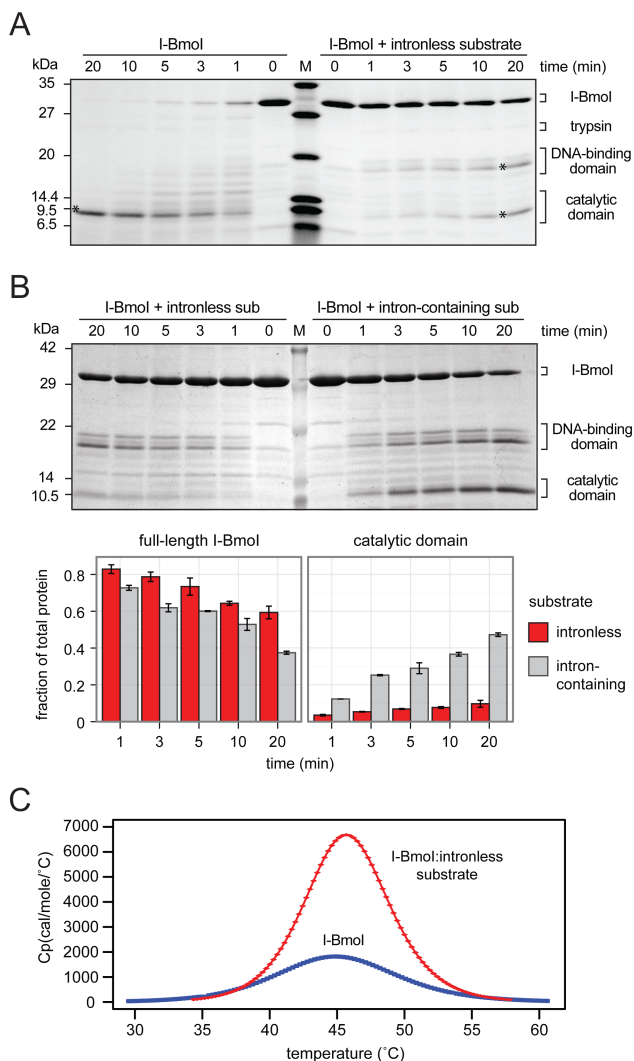


Figure 4.2: Identification of stable I-Bmol domains

(A) Images of coomassie stained SDS-gels of trypsin limited proteolysis time-course experiments performed on I-Bmol and I-Bmol pre-incubated with intronless *thyA* substrate. Aliquots were removed at the indicated time points. In-gel trypsin digest followed by mass spectrometry was performed to identify the peptide products indicated with an asterisk (*). M, protein marker (sizes in kDa indicated to the left). (B) (*top panel*) Coomassie stained SDS-gel of trypsin limited proteolysis time-course experiments of I-Bmol in complex with intronless or intron-containing substrates. (*bottom panel*) Plot of the fraction of full length I-Bmol remaining or increase of the catalytic domain peptide over time. Fraction of total protein is calculated after normalization to the untreated lane, with error bars representing the standard deviation of two replicates. (C) Enthalpy of transition curves for differential scanning calorimetry of I-Bmol or I-Bmol pre-incubated with substrate.

of the I-BmoI:intron-containing substrate complex showed a loss of protection of full-length protein, revealing a ~25% greater degradation after 20 minutes into N- and C-terminal domains (Figure 4.2B). A ~4 fold increase in the presence of the catalytic domain fragment was also observed, suggesting that nucleotide changes at the cleavage site disengage the catalytic domain from substrate resulting in increased protease sensitivity of the I-BmoI linker. These results are consistent with previous results where I-BmoI was unable to distort the intron-containing substrate as a prerequisite to bottom strand nicking, suggesting a loss of catalytic domain contacts at the cleavage site (36).

Additional insight into the stability of I-BmoI was gained through differential scanning calorimetry analyses that revealed a broad denaturation profile with an observed T_m of 45.1 °C. In the presence of *thyA* substrate, we observed a ~5-fold increase in the enthalpy of transition versus that observed for I-BmoI in solution, consistent with an increase in protein stability (Figure 4.2C) (44). Collectively, these data suggest that only the N-terminal catalytic domain of I-BmoI is stable in the absence of substrate and that the C-terminal domain becomes structured upon DNA binding.

4.3.2 N-terminal domains have weak but specific DNA-binding affinity

To investigate the solubility and DNA-binding activity of the N- and C-terminal domains of I-BmoI, we expressed and purified a number of truncations. N-terminal constructs that contained the GIY-YIG nuclease domain were soluble and expressed very well (N92, N111, N130, and N154), while C-terminal constructs were generally insoluble, with the exception of 130C (Figure 4.3A). Binding assays with full-length I-BmoI and soluble domain truncations showed stable complex formation with *thyA* substrate (Supplementary Figure S4.2A), and quantitative DNA-binding assays were performed to determine dissociation constants of each of the N-terminal domain truncations for *thyA* intronless substrate. Binding reactions with intron-containing substrate were also performed to determine whether the N-terminal truncations possessed affinity for the cleavage site region of *thyA* substrate. The dissociation constants for the N92, N111, and N130 truncations were in the 10 μ M range, while the N154 truncation had a K_D of ~500 nM for both intronless and intron-containing substrates indicating that the cleavage site region is not a major determinant for DNA binding (Figure 4.3B, Supplementary Figure

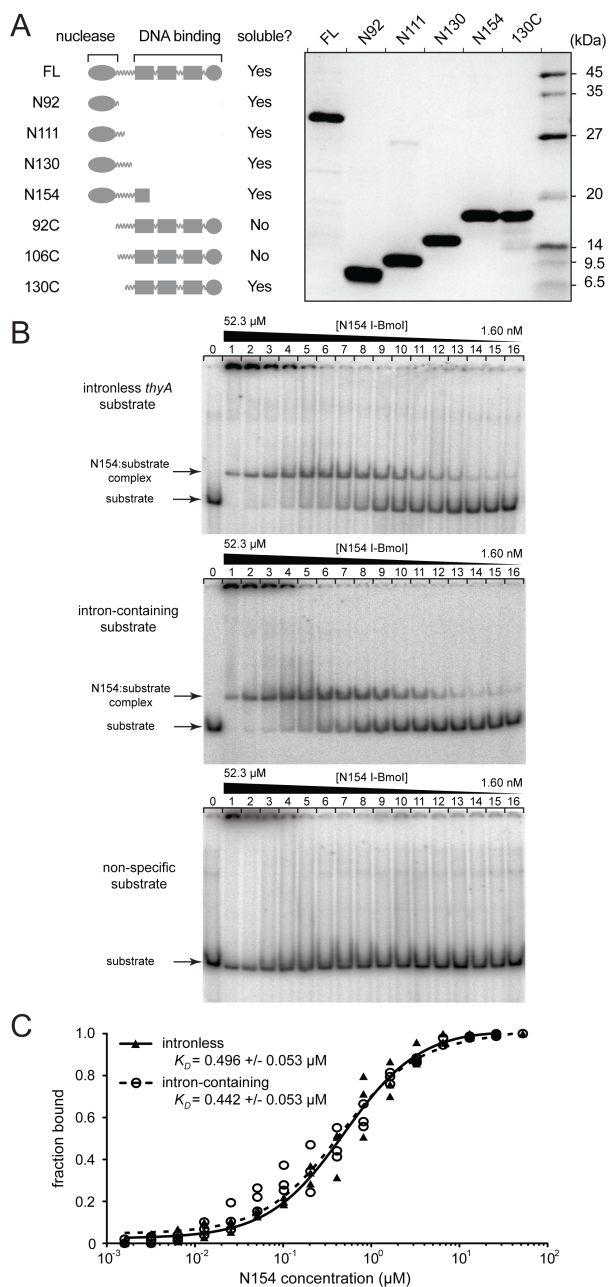


Figure 4.3: I-BmoI domains have distinct affinities for DNA

(A) Schematic of I-BmoI truncations and image of a coomassie stained SDS-page gel containing 1.5 μg of each soluble construct (*left and right* panels, respectively). FL, full length I-BmoI; N92, residues 1-92 of I-BmoI; N111, 1-111; N130, 1-130; N154, 1-154; 92C, 92-266; 106C, 1-266; 130C, 130-266. (B) Images of native gel-shift experiments for N154 binding reactions performed on labeled intronless, intron-containing, and non-specific substrate. The left lane contains substrate only, and lanes 1 through 16 contain serial dilutions of N154 from 52.3 μM to 1.60 nM. (C) Graph of the binding curves of N154 I-BmoI on intronless *thyA* or intron-containing substrate, with data from three replicates plotted (see also Supplementary Table S4.3).

S4.2B, and Supplementary Table S4.3). Discrete complex formation with the N-terminal truncations in the presence of a non-specific substrate was not observed under our binding conditions, apart from N154 which aggregated at a concentration that increased the K_D by over 100-fold (Supplementary Table S4.3, Supplementary Figure S4.2B). The dissociation constants for the full-length protein and 130C DNA-binding domain were ~ 10 nM for both intronless and intron-containing *thyA* substrates, consistent with previous data (23,31). Collectively these data show that I-BmoI possesses two functional domains connected by a protease-sensitive linker, where the N-terminal catalytic domain has weak affinity for DNA and the C-terminal domain contributes the majority of the DNA-binding energy.

4.3.3 I-BmoI binds DNA as a monomer

One model that has been proposed for DNA cleavage by GIY-YIG homing endonucleases requires the formation of a transient or stable dimer, with each monomer nicking one DNA strand (36). The ability to purify wild-type I-BmoI allowed us to determine the oligomeric status of I-BmoI and I-BmoI:substrate complexes in solution using gel filtration analyses (Figure 4.4). The elution profile of I-BmoI alone (35 kDa) matched very well with the calculated mass of an I-BmoI monomer (32 kDa), indicating that in solution I-BmoI is monomeric. The 46-mer *thyA* DNA substrate eluted at an observed mass of 99 kDa, well above that of its calculated mass of 29 kDa, due to aberrant migration of cylindrical DNA through the gel-filtration column. The observed I-BmoI:*thyA* complex, assembled at a 5:1 protein:DNA ratio, eluted with a size (133 kDa) consistent with an monomeric I-BmoI:DNA substrate complex. Changing the protein to DNA ratio to 1:2 did not change the elution profile of the I-BmoI:*thyA* complexes, suggesting that I-BmoI does not synapse between two DNA molecules. The gel-filtration results were confirmed by protein cross-linking experiments using two different reagents with I-BmoI and I-BmoI:*thyA* complexes (Supplementary Figure S4.3). No cross-linked products were observed under binding or cleavage conditions (without exogenous metal or with 10 mM $MgCl_2$, respectively) by gel electrophoresis or Western blot analyses, even at μM concentrations of protein. Thus, I-BmoI is monomeric and does not form stable higher order complexes in solution or when in complex with substrate.

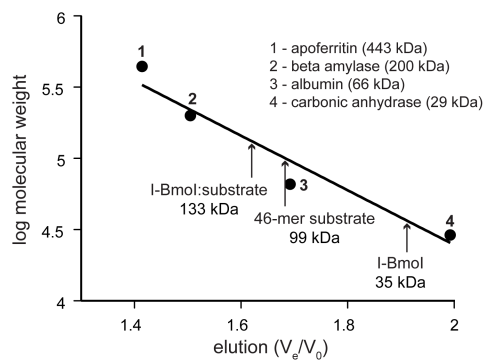


Figure 4.4: I-BmolI is a monomer in solution and in complex with *thyA* substrate

Graph of gel filtration elution profile of I-BmolI, 46-mer *thyA* substrate, or I-BmolI:substrate complex, with observed molecular weights indicated. Standards used to generate the elution volume standard curve are shown. Gel filtration analyses were performed in duplicate. V_e , elution volume; V_o , void volume.

4.3.4 The oligomeric status of I-BmoI does not change during cleavage

To specifically test whether the oligomeric status of I-BmoI changes during cleavage, we performed *in vitro* kinetic assays to 1) examine the dependence of initial reaction velocity on protein concentration, 2) determine whether synapsis by two endonuclease monomers in a two target site plasmid can enhance cleavage rate, and 3) investigate whether adding excess catalytic domain can stimulate cleavage (37,38). Prior to performing the *in vitro* kinetic assays, we sought to examine whether DNA topology influenced the rate of substrate cleavage by I-BmoI. To accomplish this, we used a plasmid containing a *thyA* target site into which we introduced restriction sites for the Nt.BbvCI nickase at two positions. The same plasmid could therefore be used as a supercoiled, relaxed, or linear substrate depending on whether the plasmid was relaxed with Nt.BbvCI or linearized with SmaI. As shown, no significant difference was observed for the rate of product formation on supercoiled, relaxed plasmid, or linear substrates at 0.5 mM, 2 mM, or 10 mM MgCl₂ (Supplementary Figure S4.4), and no difference in rate was observed on relaxed plasmids that were nicked at either or both Nt.BbvCI sites (data not shown). Subsequent *in vitro* cleavage assays were performed on supercoiled plasmid substrate unless otherwise indicated, as this allowed us to observe and determine rate constants for all steps of the reaction.

Insight into the oligomeric status of I-BmoI during cleavage was obtained by establishing the relationship between enzyme concentration and initial reaction velocity. Single-turnover cleavage reactions with eight concentrations of I-BmoI were performed on plasmid substrates containing a single *thyA* target-site where product formation was measured by the appearance of linear DNA (Figure 4.5A). The initial rate of product appearance for each I-BmoI concentration was calculated and plotted against I-BmoI concentration, yielding a linear dependence (Figure 4.5B). These data suggest that I-BmoI-catalyzed DNA hydrolysis is first order with respect to protein concentration, indicative of non-cooperative cleavage.

Additional cleavage assays were performed on substrates containing two target sites to address whether the cleavage rate could be enhanced by synapsis of two I-BmoI

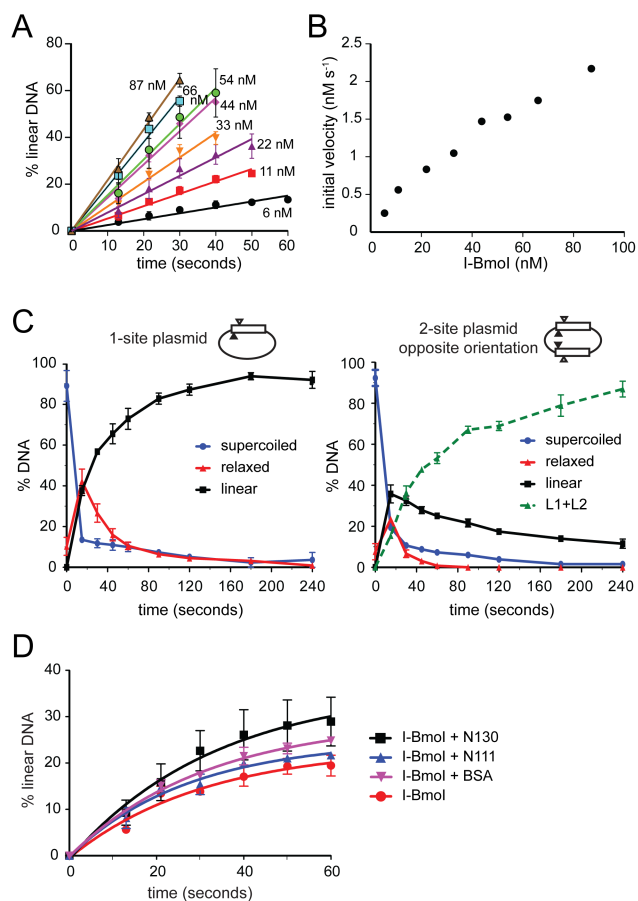


Figure 4.5: I-Bmol functions as a monomer

(A) Graph of initial reaction progress for cleavage assays with eight I-Bmol concentrations expressed as percent linear product. (B) Plot of initial reaction velocity versus I-Bmol concentration. (C) Graph of reaction progress for cleavage assays with I-Bmol and 1- or 2-site substrate plasmids (*left and right panels*, respectively). L1+L2, linear products from 2-site plasmid cleavage. (D) Domain addition experiments. Graph of linear product formation for cleavage assays with 22 nM I-Bmol and 10 nM plasmid. Reactions were supplemented with 500-fold molar excess of N111, N130, or BSA. All cleavage assays were performed in triplicate. BSA, bovine serum albumin.

monomers bound to distinct target sites within a single plasmid. Reaction progress was examined on plasmids that contained a primary target site alone, or those that contained an additional secondary target site in the opposite or same orientation relative to the primary site (Figure 4.5C and Supplementary Figure S4.5). No rate enhancement was observed for either nicking reaction by I-BmoI on plasmids containing secondary target sites relative to plasmids with a single site, suggesting that synapsis is not required for cleavage. To further address a cooperative or synaptic cleavage mechanism, biotin-tagged/radiolabeled oligonucleotides were immobilized on streptavidin-coated magnetic beads prior to the addition of I-BmoI (Supplementary Figure S4.6A). After the beads were stringently washed to remove unbound I-BmoI, cleavage was induced by the addition of magnesium and product liberation was measured by quantification of radiolabel released into solution. Cleavage levels were similar regardless of whether or not excess I-BmoI was added to the pre-formed I-BmoI:substrate complexes during the magnesium cleavage step (Supplementary Figure S4.6B).

Next, we tested whether transient interactions between catalytic domains could stimulate cleavage by performing cleavage assays with I-BmoI in the presence of excess N-terminal domain truncations. Cleavage assays with I-BmoI and 500-fold molar excess N111 or N130 did not stimulate cleavage rate relative to the addition of BSA (Figure 4.5D). Assays with a 500-fold molar excess of N154 led to non-specific cleavage, as the concentration of the domain added approached the K_D of non-specific binding for N154 (Supplementary Table S4.3). A similar pattern of non-specific substrate degradation was observed for reactions containing the same concentration of N154 alone (data not shown). Given that I-BmoI exists as a stable monomer in solution and in complex with substrate, these data demonstrate that I-BmoI acts as a monomer at all stages of the cleavage pathway and rules out transient dimerization through the catalytic domain as a potential cleavage mechanism.

4.3.5 I-BmoI does not function via a substrate hairpin mechanism

As a monomeric nuclease, I-BmoI could employ a substrate ‘hairpin’ mechanism to hydrolyze DNA, where the liberated 3’-OH from the primary nick in DNA acts as the nucleophile to attack the scissile phosphate bond of the opposite strand. The intra-strand

transesterification mechanism has been demonstrated for the Tn10 transposase, the retroviral HIV-1 DNA integrase, and V(D)J recombinases(45-47). Alternatively, I-BmoI could form a covalent protein-DNA intermediate, as is the case with the phospholipase D superfamily enzyme BfII (48). We tested this hypothesis by generating substrates that contained pre-nicked bottom strands with either a 3'-OH or 3'-H at the bottom strand nick site. If I-BmoI uses an intra-strand transesterification reaction, the second nicking reaction of the top strand should be greatly reduced with the 3'-H substrate. Both pre-nicked substrates were bound and cleaved equivalently to the intact *thyA* substrate, demonstrating that the primary 3'-OH product is not required for resolution of the DSB and that I-BmoI does not function via an intra-strand substrate hairpin mechanism (Figure 4.6).

4.3.6 The I-BmoI target site is modular

Since I-BmoI acts as a monomer throughout all steps of the cleavage pathway, a potential mechanism for DSB formation involves dynamic conformational changes within the enzyme:substrate complex leading to the reorientation of the single I-BmoI active site. This model implies that the *thyA* substrate should also be modular and contributes to the DNA distortions observed during cleavage (34). Previous results indicated that the G-2 base is critical for cleavage and that other bases surrounding the cleavage site can modulate cleavage efficiency (31,36). Additionally, nucleotide insertions downstream from the intron insertion site between positions A+8/C+9 alter cleavage by I-BmoI (42).

To delineate functional regions of the *thyA* substrate and examine the effect of nucleotide insertions on the first- and second-strand nicking rates, we generated a series of plasmid substrates that contained target sites with 3-nucleotide insertions between A+4/A+5 or A+8/C+9, and substrates with 5-nucleotide insertions between A+8/C+9 (Figure 4.7A). Substrates containing TTT and CCC insertions between A+4/A+5 attenuated cleavage by ~5-fold under standard conditions with 10 mM MgCl₂, leading to second-strand rate constants (k_2) of $0.0084 \pm 0.0006 \text{ s}^{-1}$ and $0.0051 \pm 0.0003 \text{ s}^{-1}$, respectively (Figure 4.7B). Also evident from the reaction progress curves is the lack of accumulation of a nicked intermediate, consistent with first-strand nicking being rate limiting ($k_1 < k_2$). Assays performed on the same substrates with 2 mM or 0.5 mM MgCl₂ showed low levels of

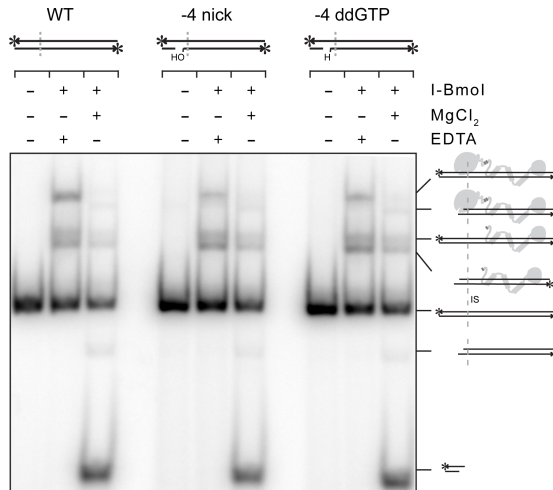


Figure 4.6: I-BmoI does not function via an intra-strand hairpin mechanism

Image of I-BmoI binding reactions performed with dual labeled *thyA* substrate (WT), bottom-strand pre-nicked substrate between -5/-4 (-4 nick), and bottom strand pre-nicked substrate lacking the 3'-OH at the nick site (-4 ddGTP). Reactions were performed with 10 mM EDTA or 10 mM MgCl₂. Schematic and naming of complexes are as previously described (34). Reactions were performed in duplicate showing similar results.

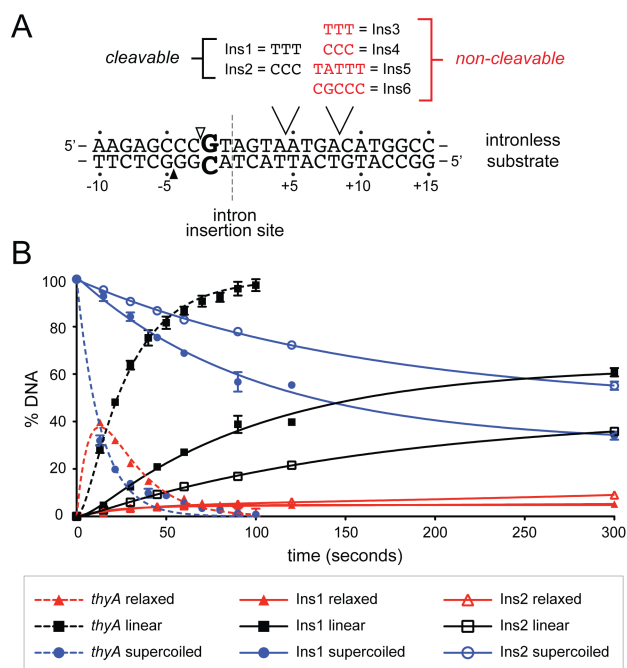


Figure 4.7: Insertions in the I-BmoI substrate spacer reduce or abolish cleavage

(A) Schematic of the wild-type *thyA* I-BmoI substrate and the substrates generated containing 3- or 5-nucleotide insertions. Filled and open triangles represent the bottom- and top-strand nicking sites, respectively. (B) Progress curves for cleavage reactions on *thyA*, Ins1, and Ins2 substrates in 10 mM MgCl₂. Curves are generated from reactions performed in triplicate. The x-axis has been manipulated so that reaction progress for all substrates can be visualized.

cleavage (Supplementary Figure S4.7). Cleavage assays on the four substrates with target sites containing 3 or 5 nucleotide insertions between A+8/C+9 showed only minimal nicking activity after prolonged incubations under standard reaction conditions with 10 mM MgCl₂ (data not shown). These substrates allowed us to identify the sequence upstream of A+5 as tolerant to insertions and the sequence downstream of A+8 as intolerant to insertions. One interpretation of this data is that specific bases in the A+5/A+8 region may act as anchor points to position the I-BmoI linker or DNA-binding domain on substrate.

To determine if individual bases in the G-6 to A+8 region of *thyA* substrate were important for function *in vivo*, we adapted an *in vitro* substrate selection that previously identified G-2 to be critical for cleavage by I-BmoI (31). We generated a target site with randomized nucleotides between G-6 and A+8, while maintaining G-2 fixed, to create a library of target site toxic reporter plasmids (pKoxRCS) to be screened *in vivo* (Figure 4.8A). The pKoxRCS library was transformed into cells harboring the I-BmoI expression plasmid (pACYCIBmoI) and transformants were replica-gridded in triplicate onto selective and non-selective plates (Figure 4.8B), where control toxic plasmids containing the *thyA* intronless target site survived on selective media and plasmids encoding the intron-containing target did not. We identified 86 robustly cleaved reporter plasmids and sequenced their target site regions, as well as the input library of randomized sites using barcoded PCR products on an Ion Torrent platform. A plot of the relative change in nucleotide proportions at each position between the survivor sequences and input pool, as well as a position-specific scoring matrix (PSSM), identified the wild-type guanine nucleotide as highly preferred at position +7 (Figure 4.8C, Supplementary Figure S4.8A). Consistent with previous results, positions within the G-6 to T-1 cleavage motif did not show strong nucleotide preference (31,36). Positions that displayed slight nucleotide preference identified the wild-type nucleotide as preferred at that position. A plot of the relative change in nucleotide proportion between the survivor sequences and a pool of non-cleavable ('dead') clones showed similar results, identifying +7 as the only position with a strong preference for any nucleotide (Supplementary Figure S4.8B).

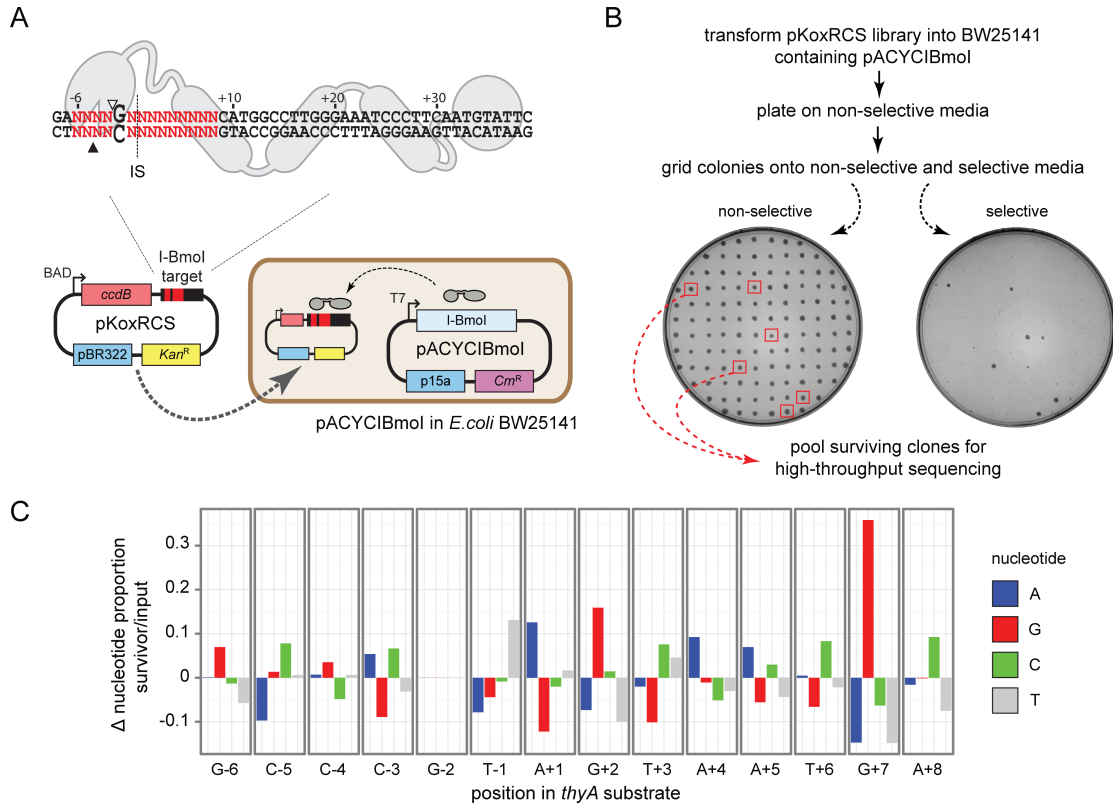


Figure 4.8: An *in vivo* screen identifies I-Bmol nucleotide preference at G+7

(A) (*top*) Schematic representation of the randomized cleavage site from -6 to +8 with G-2 fixed (in bold enlarged font). (*bottom*) The pKoxRCS library and I-Bmol expression plasmids. Filled and open triangles represent the bottom- and top-strand nicking sites, respectively. IS, intron-insertion site. (B) Work flow to select for cleavable target sites. (C) Sequencing results from cleavable targets represented as the difference in nucleotide proportion from the clones in the survivor pool versus the input library.

4.3.7 Contacts to G+7 facilitate cleavage by I-BmoI

To better understand the preference for guanine at position +7, we designed plasmid substrates that contained G+7A, G+7C, and G+7T target site substitutions for *in vitro* and *in vivo* assays. Cleavage assays on the G+7T substrate with 0.5 mM MgCl₂ showed a ~2.5 fold decrease in both k_1 and k_2 versus the wild-type G+7 substrate (Figure 4.9A, Table 4.1). The G+7A substrate displayed kinetics similar to the wild-type target site, while G+7C showed a slight first-strand nicking defect (Table 4.1). Metal-dependent rescue of cleavage defects on mutant substrates *in vitro* was observed in reactions with 2 mM or 10 mM MgCl₂, consistent with previous results (36).

The observed cleavage defects on the G+7 substitution substrates were exaggerated when the same target sites were examined by the *in vivo* two-plasmid survival assay, possibly due to low levels of *in vivo* protein expression versus excess protein to DNA ratios *in vitro* (28,40). Briefly, the cleavage efficiency of I-BmoI on non-cognate substrates can be assessed by the degree to which a target site embedded within a toxic reporter plasmid is cleaved and survival percentage is reported as the ratio of colonies on selective versus non-selective plates (28). We observed ~90% survival by I-BmoI on the wild-type *thyA* target site and decreasing survival on G+7A > G+7C > G+7T substrates (Figure 4.9B). The correlation of reduced *in vitro* cleavage and poor *in vivo* survival for transversion substitutions at G+7 suggests that a purine at position +7 acts as a critical anchor point for I-BmoI on *thyA* substrate.

To establish a physical basis for the observed defect associated with substitutions at G+7, we performed in-gel footprinting with the minor groove specific reagent 1,10-phenanthroline copper (OP-Cu) on oligonucleotide substrates containing G+7 mutations. OP-Cu was selected as a footprinting reagent because minor groove distortions of substrate are identified as hypersensitivity to OP-Cu (49), and because it has previously been used to identify DNA distortions surrounding the I-BmoI cleavage site that are dependent on contacts to G-2 (34,36). Reactions were performed using the catalytically inactive R27A variant of I-BmoI, as bottom-strand nicked products would otherwise occlude distortions surrounding the bottom-strand nicking site (34). We found no difference in the binding affinity of I-BmoI for the three mutant substrates relative to the

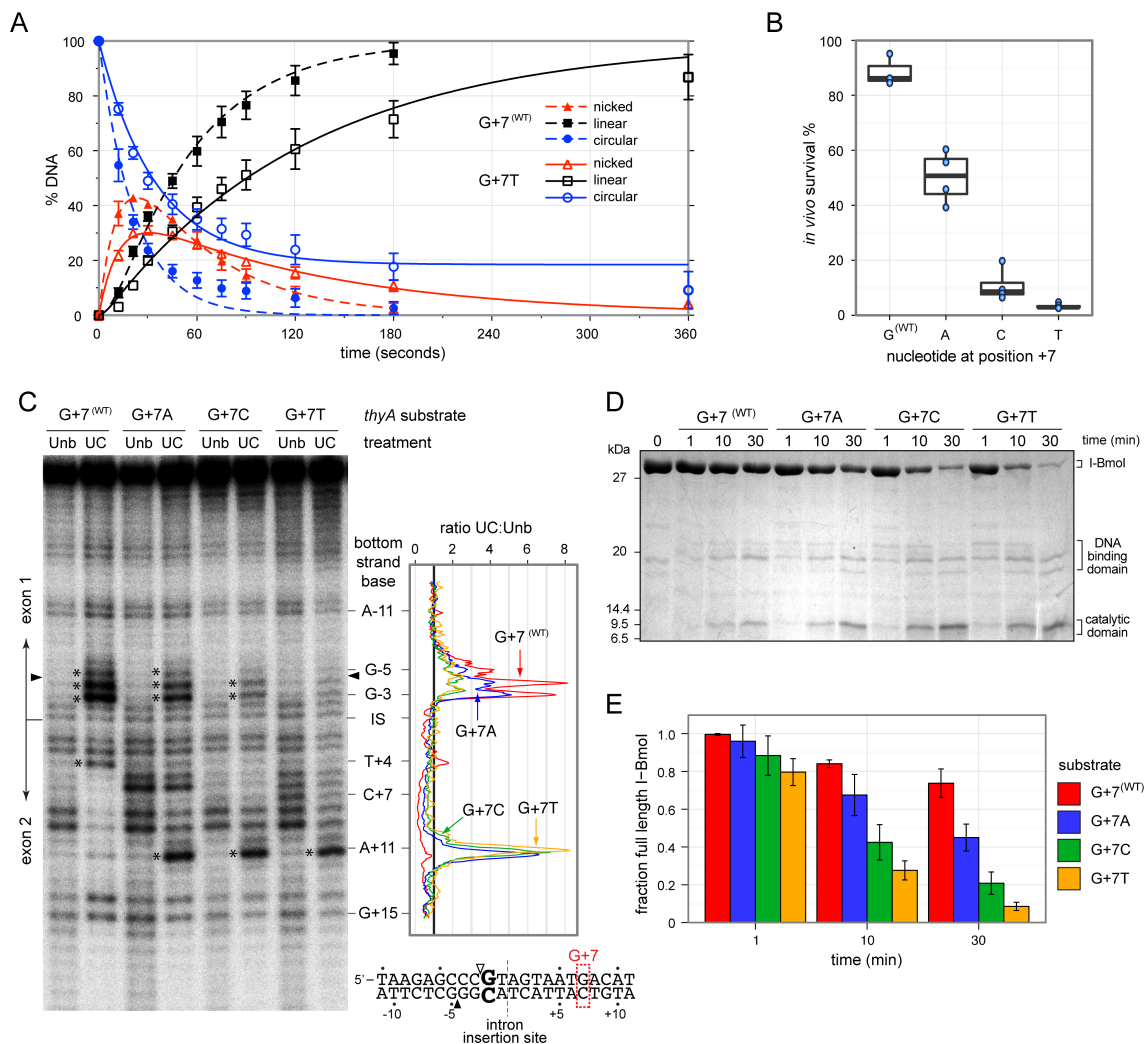


Figure 4.9: Mutations at G+7 affect I-BmoI activity

(A) Plot of reaction progress with I-BmoI and G+7 (WT) or G+7T plasmid substrates in 0.5 mM MgCl₂. Curves are plotted as a fit to the mean of three replicates with standard deviation shown. (B) *In vivo* survival of I-BmoI on *thyA*, G+7A, G+7C, and G+T toxic reporter plasmid substrates, with four replicates plotted (C) Denaturing gel image of in-gel OP-Cu footprinting reactions with R27A I-BmoI on bottom-strand labeled G+7 (WT), G+7A, G+7C, and G+7T substrates (Unb, unbound substrate; UC, full-length I-BmoI bound to substrate). Sites that are hypersensitive to the footprinting reagent are highlighted with an asterisk, and filled triangles indicate the bottom strand nick site. To the right of the gel image is a graph of the normalized pixel density ratio for bands in the UC lane vs. the Unb lane. (D) Coomassie stained SDS-gel of trypsin limited proteolysis time-course experiments of I-BmoI in the presence of G+7 (WT), G+7A, G+7C, and G+7T substrates. (E) Plot of the fraction of full length I-BmoI remaining over time for reactions shown in panel (D), calculated as the fraction remaining after normalization to the untreated lane (error bars represent the standard deviation of two replicates).

Table 4.1: Rate constants for first- and second-strand nicking reactions on G+7 mutant substrates

substrate	k_1 (s ⁻¹)	k_2 (s ⁻¹)
G+7 (WT)	0.0441 ± 0.0021	0.0198 ± 0.0015
G+7A	0.0478 ± 0.0010	0.0303 ± 0.0007
G+7C	0.0313 ± 0.0014	0.0320 ± 0.0009
G+7T	0.0172 ± 0.0009	0.0082 ± 0.0005

wild-type G+7 substrate. Footprinting reactions revealed drastic reductions in hypersensitivity to OP-Cu versus the wild-type substrate at positions surrounding the bottom-strand nicking site on the G+7C and G+7T substrates (G-5/G-4/G-3), while modest reductions were observed on the G+7A substrate (Figure 4.9C). All three mutant substrates did not exhibit sensitivity to OP-Cu on the bottom strand at position T+4 that was previously observed for the R27A footprint on wild-type *thyA* substrate (34). Additionally, a dramatic increase in sensitivity to OP-Cu was observed at A+11 on the bottom strand on all 3 mutant substrates. A similar increase in OP-Cu sensitivity at A+11 was previously observed within the lower-complex footprint of I-BmoI (34). The lower-complex consists of I-BmoI:*thyA* complexes that are unable to form DNA-contacts via the catalytic domain, or proteolyzed I-BmoI fragments that retain DNA-binding activity but presumably lack the catalytic domain entirely (23). The footprinting data in Figure 4.9C suggests that mutations at G+7 reduce I-BmoI catalytic domain and linker contacts that are essential for inducing distortions adjacent to the bottom-strand nicking site, with hypersensitivity at +11 resulting from the catalytic domain and linker of I-BmoI being disengaged from substrate.

To provide further evidence that I-BmoI catalytic domain contacts are disturbed by G+7 substitutions, we performed protease-mapping experiments on I-BmoI:G+7A, G+7C, and G+7T substrate complexes. Loss of catalytic domain or linker contacts to substrate would be observed as an increase in protease sensitivity within the I-BmoI linker, and time course reactions with trypsin revealed a reduction in the protection of full-length I-BmoI from protease digestion (Figure 4.9D, 4.9E). An increase in abundance of the catalytic and DNA-binding domain peptides was observed, indicating greater protease sensitivity within the I-BmoI linker in the presence of G+7 mutant substrates versus I-BmoI:*thyA* complexes. An increase in protease sensitivity was also observed for the 130C I-BmoI DNA-binding domain construct in complex with the G+7T substrate. Trypsin digests revealed a significantly greater rate of proteolytic degradation versus 130C I-BmoI:*thyA* complexes, indicating that I-BmoI contacts to the G+7 base pair are C-terminal to I-BmoI residue 130 (Supplementary Figure S4.9).

The results from the physical assays probing G+7 interactions are consistent with a model whereby G+7 mutations lead to 1) a reduction in I-BmoI catalytic domain or linker contacts to DNA surrounding the cleavage site that are necessary to induce bottom strand distortions near the bottom strand nicking site, 2) loss in sensitivity at +4 that is indicative of changes in protein-DNA interactions downstream of the cleavage site, and 3) a disengagement of the catalytic domain from substrate that leads to new DNA distortions at +11 as well as increased protease sensitivity within the linker and across the DNA binding domain (Figure 4.10).

4.4 Discussion

Early efforts with I-TevI to address the mechanistic question of how the single active site of GIY-HEs is utilized to make a DSB revealed that the two-domain protein bound DNA as a monomer, that the linker connecting the functional domains was flexible, and that the N-terminal GIY-YIG nuclease domain possessed a single active site (10,15,30,35). These data all pointed towards GIY-HEs functioning as monomers, with significant domain rearrangements postulated to reposition the nuclease domain between strand-specific nicking reactions to generate a DSB. However, the extreme cytotoxicity of I-TevI precluded necessary kinetic studies to determine whether the enzyme functioned as a monomer in all stages of the reaction pathway, and prohibited the exclusion of alternative cleavage mechanisms. A number of mechanisms have been described for monomeric nucleases that make DSBs, and we discuss the relevance of these mechanisms in light of our current data supporting a monomeric mode of action by I-BmoI whereby the C-terminal domain acts as a molecular anchor to tether the N-terminal nuclease domain to sequentially nick both DNA strands.

As structurally bipartite enzymes, GIY-HEs closely resemble the type IIS restriction enzymes FokI and MboII that bind DNA through N-terminal domains and cleave at a distance via C-terminal single-active site catalytic domains (50,51). FokI can form a DSB by a cooperative transient dimerization mechanism where a substrate bound FokI can dimerize with additional solution monomers, or synapse with the catalytic domain of another FokI molecule bound to a secondary target site (37,38). We did not find evidence for I-BmoI functioning via a transient dimerization mechanism, as the rate of cleavage is

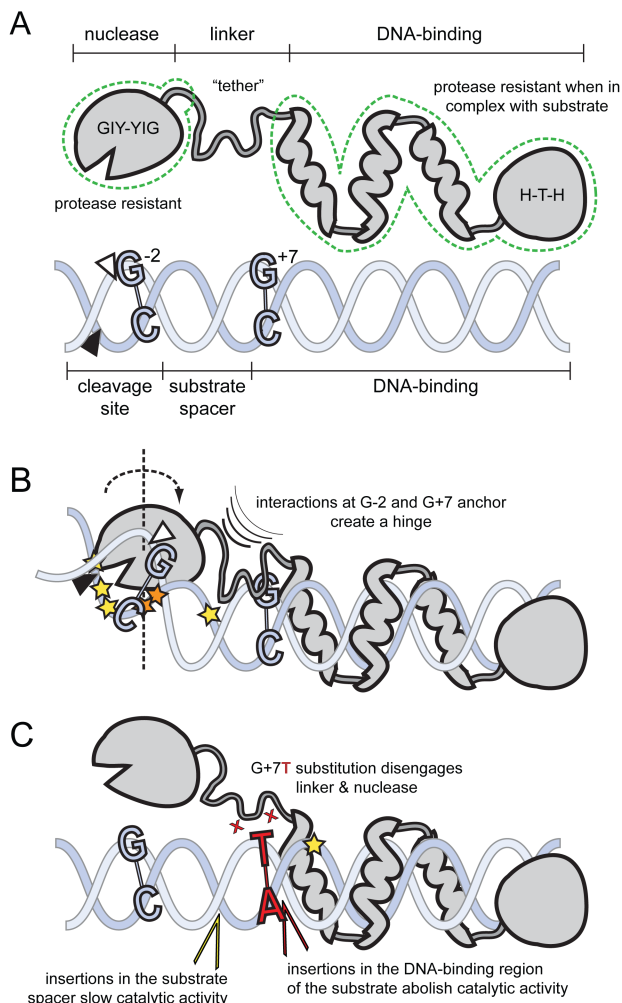


Figure 4.10: The domains of I-BmoI form distinct DNA-contacts to distort DNA prior to cleavage

(A) Schematic of I-BmoI domains and substrate modularity. The protease resistant domains of I-BmoI and substrate bases required for efficient cleavage are highlighted. Bottom- and top-strand nicking sites are indicated by filled and open triangles, respectively. (B) Model of I-BmoI bending DNA to induce substrate distortions near the bottom-strand nick site (indicated by yellow and orange filled stars that represent OP-Cu hypersensitivity for R27A and WT I-BmoI footprinting reactions, respectively). Rotation of the enzyme:substrate complex after bottom-strand nicking to reposition the nuclease domain for top-strand nicking is indicated by a dashed arrow around a vertical axis. Interactions at G-2 and G+7 may form a hinge necessary to distort DNA. (C) Schematic of nucleotide substitutions and insertions that affect I-BmoI catalytic activity.

linear with respect to protein concentration, addition of excess catalytic domain cleavage does not stimulate product formation, and two-site plasmids are cleaved at the same rate as a single site plasmid. Furthermore, the structure of the I-TevI catalytic domain and homology model of the I-BmoI catalytic domain did not reveal an obvious dimer interface (10,28,36). Interestingly, the type IIS enzyme Eco31I is also a two-domain protein with a single HNH active site that cleaves close to its binding site (52,53). Unlike FokI, Eco31I does not transiently dimerize to generate a DSB, meaning that the single HNH active site sequentially nicks each strand. The reaction pathway of Eco31I proceeds with roughly equal rate constants for both nicking reactions and therefore parallels that of I-BmoI, where the single GIY-YIG active site sequentially nicks the bottom- and top-strands with only a ~2-fold difference in rates.

The two-domain structure of I-BmoI is also similar to I-HmuI, an HNH family homing endonuclease that binds an extended target site (54). I-HmuI consists of an N-terminal HNH catalytic domain, and a C-terminal DNA-binding domain composed of minor-groove binding α -helix (NUMOD3) and helix-turn-helix modules that are very similar to those found within the I-TevI DNA-binding domain (43,54,55). I-HmuI, however, nicks only one strand of its substrate, presumably because DNA-binding elements found on either side of the HNH nuclease motif prevent release of the catalytic domain after nicking. Our data suggest that the I-BmoI catalytic domain is released from substrate between nicking reactions, as the minimal I-BmoI N-terminal domain (N92) possesses weak DNA affinity ($K_D > 23 \mu\text{M}$), and limited-protease digestions of full-length I-BmoI on intron-containing and G+7 mutant substrates revealed that the catalytic domain was disengaged from substrate. Moreover, the homology model of the I-BmoI catalytic domain does not possess additional units of protein structure that bind DNA, consistent with the evolutionary acquisition of distinct DNA-binding modules by GIY-YIG family enzymes due to the inherently weak DNA affinity of the nuclease domain.

DNA cleavage mechanisms that require conformational changes and rearrangements of single active sites have been proposed for the phospholipase D superfamily restriction enzyme BfiI and the type IIP restriction enzymes BcnI and MvaI (56-58). BfiI exists as a dimer yet forms only a single active site at the dimer interface that rotates and switches

orientation between sequential nicking reactions (56). I-BmoI is unlike BfiI as there remains no evidence for the formation of an I-BmoI dimer, and the proposed catalytic mechanism for BfiI involves a covalent DNA intermediate. Based on active site similarity to other GIY-YIG nucleases (10-14,36), it is unlikely that GIY-HEs form covalent protein:DNA intermediates, and we have shown that I-BmoI does not function via an intra-strand transesterification mechanism. Conversely, a mechanism has been demonstrated for BcnI where the enzyme acts as a monomer to nick one strand prior to ‘hopping’ to the opposite strand where it interacts in the reverse orientation to generate the second nick (57). The affinity of the I-BmoI DNA-binding domain for substrate and the fact that I-BmoI remains bound to the cleavage product would preclude the full-length enzyme from hopping to the opposite strand and nicking in an alternate orientation. It is possible, however, that the I-BmoI catalytic domain dissociates and rearranges between nicking reactions due to its inherently weak DNA affinity and the flexible nature of the linker.

We envision a monomeric cleavage mechanism by I-BmoI where the DNA-binding domain acts as a molecular tether to allow the GIY-YIG catalytic domain to diffuse and reposition to effect the sequential nicking reactions. Conformational changes in the I-BmoI:substrate complex that precede bottom-strand nicking are dependent on contacts to the G-2 and G+7 base pairs, as mutation of either position abolishes DNA distortions by I-BmoI that precede nicking (36). The function of the G+7 anchor point to position the I-BmoI catalytic domain and linker is supported by +4/+5 and +8/+9 insertions in *thyA* substrate that drastically reduce k_1 or abolish activity, respectively. Insertions at +4/+5 extend the spacing that intervenes the cleavage site and G+7 anchor to a distance where bottom-strand nicking is constrained, while insertions at +8/+9 destroy I-BmoI interactions with G+7 leading to undetectable levels of cleavage. I-BmoI contains three predicted NUMOD3 minor-groove DNA-binding repeats similar to those that are found in I-TevI and I-HmuI, with the most N-terminal NUMOD3 of the I-BmoI DNA-binding domain encompasses residues G144-S156 (28). We found a ~100-fold difference in the K_D values of N130 and N154 I-BmoI for *thyA* substrate, and protease mapping of 130C I-BmoI performed on the G+7T mutant substrate led to a higher level of proteolytic digestion than was observed on the wild-type *thyA* substrate. It is therefore possible that

the N-terminal I-BmoI NUMOD3 repeat interacts with *thyA* substrate at G+7 to anchor the linker to substrate to facilitate positioning of the nuclease domain.

Given that I-BmoI generates a DSB by sequentially nicking the bottom- and top-strands, the GIY-YIG domain must reposition to the opposite DNA strand after first-strand nicking to hydrolyze the phosphodiester bond that lies in the reverse orientation. The observation that the two nicking reactions differ by only an ~2-fold rate difference suggests that conformational changes during cleavage are not limiting for second-strand nicking. Following first-strand hydrolysis, the nicked DNA intermediate could rotate around the top-strand backbone, reducing the spatial reorientation required to move the top-strand scissile phosphate bond into the active site of the catalytic domain. This model is consistent with the fact that I-BmoI can cleave substrates where the bottom- or top-strands have been resected to leave a single-stranded target on the non-resected strand (34).

Homing endonucleases often target functionally critical regions of conserved genes as a strategy to ensure that their target site will be present in naïve genomes (59). For instance, the critical G-C base pair at position -2 of the I-BmoI target corresponds to the second nucleotide position of a conserved arginine codon that is present in all TS genes (31,32). Likewise, the G-C base pair at position +7 that acts as an I-BmoI anchor point corresponds to the first nucleotide position of an aspartate codon that is also highly conserved. The same G-C base pair and aspartate codon is also present in the I-TevI *td* target site, falling within a hypomutable region of the *td* DNA substrate (22). Additionally, ethylation and methylation interference assays of I-TevI:*td* complexes suggested that I-TevI contacts the G of the G-C base pair and phosphate backbone at this position (22). It is therefore possible that mutations at this position would also impact I-TevI cleavage efficiency; however this nucleotide position has not previously been the target of mutagenesis studies.

In summary, we show that I-BmoI generates a DSB via a mechanism that is distinct from other GIY-YIG domain containing enzymes that oligomerize. The I-BmoI GIY-YIG domain is fused to a C-terminal DNA-binding domain to overcome the inherently weak

affinity of the nuclease domain for DNA. The linker that connects the I-BmoI N- and C-terminal domains functions as both an anchor to position the catalytic domain and a flexible tether to permit conformational rotation between the sequential nicking reactions. It is interesting to note that the I-BmoI and I-TevI linkers share significant amino acid similarity, suggesting that the enzymes function via similar cleavage mechanisms. Moreover, the fact that both the I-BmoI and I-TevI substrates share the G+7 nucleotide in their native substrates becomes an important consideration when targeting chimeric GIY-YIG nucleases for genome-editing applications.. In particular, investigating whether I-TevI cleavage efficiency is also influenced by the analogous G+7 position will help determine if the proximity of the anchor point to the cleavage site can influence catalytic activity of chimeric GIY-YIG nucleases.

4.5 References

1. Yang,W. (2011) Nucleases: Diversity of structure, function and mechanism. *Q. Rev. Biophys.*, **44**, 1-93.
2. Stoddard,B.L. (2005) Homing endonuclease structure and function. *Q Rev Biophys*, **38**, 49-95.
3. Galburt,E.A. and Stoddard,B.L. (2002) Catalytic mechanisms of restriction and homing endonucleases. *Biochemistry*, **41**, 13851-60.
4. Taylor,G.K. and Stoddard,B.L. (2012) Structural, functional and evolutionary relationships between homing endonucleases and proteins from their host organisms. *Nucleic Acids Res.*, **40**, 5189-5200.
5. Kim,Y.G., Cha,J. and Chandrasegaran,S. (1996) Hybrid restriction enzymes: Zinc finger fusions to fok I cleavage domain. *Proc. Natl. Acad. Sci. U. S. A.*, **93**, 1156-1160.
6. Ashworth,J., Havranek,J.J., Duarte,C.M., Sussman,D., J.,M.R.,Jr, Stoddard,B.L. and Baker,D. (2006) Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature*, **441**, 656-9.
7. Chan,S.H., Stoddard,B.L. and Xu,S.Y. (2011) Natural and engineered nicking endonucleases--from cleavage mechanism to engineering of strand-specificity. *Nucleic Acids Research*, **39**, 1-18.
8. Fonfara,I., Curth,U., Pingoud,A. and Wende,W. (2012) Creating highly specific nucleases by fusion of active restriction endonucleases and catalytically inactive homing endonucleases. *Nucleic Acids Res.*, **40**, 847-860.

9. Kleinstiver,B.P., Wolfs,J.M., Kolaczyk,T., Roberts,A.K., Hu,S.X. and Edgell,D.R. (2012) Monomeric site-specific nucleases for genome editing. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 8061-8066.
10. Van Roey,P., Meehan,L., Kowalski,J.C., Belfort,M. and Derbyshire,V. (2002) Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI. *Nat Struct Biol*, **9**, 806-11.
11. Truglio,J.J., Rhau,B., Croteau,D.L., Wang,L., Skorvaga,M., Karakas,E., Dellavecchia,M.J., Wang,H., Van Houten,B. and Kisker,C. (2005) Structural insights into the first incision reaction during nucleotide excision repair. *EMBO J*, **24**, 885-94.
12. Andersson,C.E., Lagerback,P. and Carlson,K. (2010) Structure of bacteriophage T4 endonuclease II mutant E118A, a tetrameric GIY-YIG enzyme. *Journal of Molecular Biology*, **397**, 1003-16.
13. Mak,A.N., Lambert,A.R. and Stoddard,B.L. (2010) Folding, DNA recognition, and function of GIY-YIG endonucleases: Crystal structures of R.Eco29kI. *Structure*, **18**, 1321-31.
14. Sokolowska,M., Czapinska,H. and Bochtler,M. (2010) Hpy188I-DNA pre- and post-cleavage complexes--snapshots of the GIY-YIG nuclease mediated catalysis. *Nucleic Acids Research*, **39**, 1554-64.
15. Derbyshire,V., Kowalski,J.C., Dansereau,J.T., Hauer,C.R. and Belfort,M. (1997) Two-domain structure of the *td* intron-encoded endonuclease I-TevI correlates with the two-domain configuration of the homing site. *J Mol Biol*, **265**, 494-506.
16. Pyatkov,K.I., Arkhipova,I.R., Malkova,N.V., Finnegan,D.J. and Evgen'ev,M.B. (2004) Reverse transcriptase and endonuclease activities encoded by penelope-like retroelements. *Proc Natl Acad Sci U S A*, **101**, 14719-24.
17. Dunin-Horkawicz,S., Feder,M. and Bujnicki,J.M. (2006) Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics*, **7**:98.
18. Lagerback,P., Andersson,E., Malmberg,C. and Carlson,K. (2009) Bacteriophage T4 endonuclease II, a promiscuous GIY-YIG nuclease, binds as a tetramer to two DNA substrates. *Nucleic Acids Research*, **37**, 6174-83.
19. Ibryashkina,E.M., Sasnauskas,G., Solonin,A.S., Zakharova,M.V. and Siksnyus,V. (2009) Oligomeric structure diversity within the GIY-YIG nuclease family. *Journal of Molecular Biology*, **387**, 10-16.
20. Brachner,A., Braun,J., Ghodgaonkar,M., Castor,D., Zlopasa,L., Ehrlich,V., Jiricny,J., Gotzmann,J., Knasmueller,S. and Foisner,R. (2012) The endonuclease Ankle1 requires its LEM and GIY-YIG motifs for DNA cleavage in vivo. *J. Cell. Sci.*, **125**, 1048-57.

21. Kaminska,K.H., Kawai,M., Boniecki,M., Kobayashi,I. and Bujnicki,J.M. (2008) Type II restriction endonuclease R.Hpy188I belongs to the GIY-YIG nuclease superfamily, but exhibits an unusual active site. *BMC Structural Biology*, **8**:48.
22. Bryk,M., Quirk,S.M., Mueller,J.E., Loizos,N., Lawrence,C. and Belfort,M. (1993) The *td* intron endonuclease I-TevI makes extensive sequence-tolerant contacts across the minor groove of its DNA target. *EMBO J*, **12**, 2141-9.
23. Edgell,D.R. and Shub,D.A. (2001) Related homing endonucleases I-BmoI and I-TevI use different strategies to cleave homologous recognition sites. *Proc Natl Acad Sci U S A*, **98**, 7898-903.
24. Quirk,S.M., Bell-Pedersen,D. and Belfort,M. (1989) Intron mobility in the T-even phages: High frequency inheritance of group I introns promoted by intron open reading frames. *Cell*, **56**, 455-65.
25. Edgell,D.R. (2009) Selfish DNA: Homing endonucleases find a home. *Current Biology*, **19**, 115-17.
26. Sandegren,L. and Sjoberg,B.M. (2004) Distribution, sequence homology, and homing of group I introns among T-even-like bacteriophages - evidence for recent transfer of old introns. *J. Biol. Chem.*, **279**, 22218-27.
27. Nord,D. and Sjöberg,B.-. (2008) Unconventional GIY-YIG homing endonuclease encoded in group I introns in closely related strains of the bacillus cereus group. *Nucleic Acids Res*, **36**, 300-10.
28. Kleinstiver,B.P., Fernandes,A.D., Gloor,G.B. and Edgell,D.R. (2010) A unified genetic, computational and experimental framework identifies functionally relevant residues of the homing endonuclease I-BmoI. *Nucleic Acids Research*, **38**, 2411-27.
29. Liu,Q., Dansereau,J.T., Puttamadappa,S.S., Shekhtman,A., Derbyshire,V. and Belfort,M. (2008) Role of the interdomain linker in distance determination for remote cleavage by homing endonuclease I-TevI. *J Mol Biol*, **379**, 1094-106.
30. Dean,A.B., Stanger,M.J., Dansereau,J.T., Van Roey,P., Derbyshire,V. and Belfort,M. (2002) Zinc finger as distance determinant in the flexible linker of intron endonuclease I-TevI. *Proc Natl Acad Sci U S A*, **99**, 8554-61.
31. Edgell,D.R., Stanger,M.J. and Belfort,M. (2003) Importance of a single base pair for discrimination between intron-containing and intronless alleles by endonuclease I-BmoI. *Curr Biol*, **13**, 973-8.
32. Edgell,D.R., Stanger,M.J. and Belfort,M. (2004) Coincidence of cleavage sites of intron endonuclease I-TevI and critical sequences of the host thymidylate synthase gene. *J Mol Biol*, **343**, 1231-1241.

33. Bell-Pedersen,D., Quirk,S.M., Bryk,M. and Belfort,M. (1991) I-*TevI*, the endonuclease encoded by the mobile *td* intron, recognizes binding and cleavage domains on its DNA target. *Proc Natl Acad Sci U S A*, **88**, 7719-23.
34. Carter,J.M., Friedrich,N.C., Kleinstiver,B. and Edgell,D.R. (2007) Strand-specific contacts and divalent metal ion regulate double-strand break formation by the GIY-YIG homing endonuclease I-BmoI. *J Mol Biol*, **374**, 306-21.
35. Mueller,J.E., Smith,D., Bryk,M. and Belfort,M. (1995) Intron-encoded endonuclease I-*TevI* binds as a monomer to effect sequential cleavage via conformational changes in the *td* homing site. *EMBO J*, **14**, 5724-35.
36. Kleinstiver,B.P., Berube-Janzen,W., Fernandes,A.D. and Edgell,D.R. (2011) Divalent metal ion differentially regulates the sequential nicking reactions of the GIY-YIG homing endonuclease I-BmoI. *Plos One*, **6**, e23804.
37. Bitinaite,J., Wah,D.A., Aggarwal,A.K. and Schildkraut,I. (1998) FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci U S A*, **95**, 10570-5.
38. Vanamee,E.S., Santagata,S. and Aggarwal,A.K. (2001) FokI requires two specific DNA sites for cleavage. *J. Mol. Biol.*, **309**, 69-78.
39. Sasnauskas,G., Connolly,B.A., Halford,S.E. and Siksnys,V. (2007) Site-specific DNA transesterification catalyzed by a restriction enzyme. *Proc Natl Acad Sci U S A*, **104**, 2115-20.
40. Chen,Z. and Zhao,H. (2005) A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic Acids Res*, **33**, e154.
41. Spink,C.H. (2008) Differential scanning calorimetry. *Biophysical Tools for Biologists: Vol 1 in Vitro Techniques*, **84**, 115-41.
42. Liu,Q., Derbyshire,V., Belfort,M. and Edgell,D.R. (2006) Distance determination by GIY-YIG intron endonucleases: Discrimination between repression and cleavage functions. *Nucleic Acids Res*, **34**, 1755-64.
43. Sitbon,E. and Pietrokovski,S. (2003) New types of conserved sequence domains in DNA-binding regions of homing endonucleases. *Trends Biochem Sci*, **28**, 473-7.
44. Johnson,C.M. (2012) Differential scanning calorimetry as a tool for protein folding and stability. *Arch. Biochem. Biophys.*, doi: 10.1016/j.abb.2012.09.008. [Epub ahead of print]
45. Kennedy,A.K., Guhathakurta,A., Kleckner,N. and Haniford,D.B. (1998) Tn10 transposition via a DNA hairpin intermediate. *Cell*, **95**, 125-34.

46. Engelman,A., Mizuuchi,K. and Craigie,R. (1991) Hiv-1 dna integration - mechanism of viral-dna cleavage and dna strand transfer. *Cell*, **67**, 1211-21.
47. Mcblane,J.F., Vangent,D.C., Ramsden,D.A., Romeo,C., Cuomo,C.A., Gellert,M. and Oettinger,M.A. (1995) Cleavage at a V(d)j recombination signal requires only Rag1 and Rag2 proteins and occurs in 2 steps. *Cell*, **83**, 387-95.
48. Sasnauskas,G., Halford,S.E. and Siksnys,V. (2003) How the BfiI restriction enzyme uses one active site to cut two DNA strands. *Proc Natl Acad Sci U S A*, **100**, 6410-5.
49. Spassky,A. and Sigman,D.S. (1985) Nuclease activity of 1,10-phenanthroline-copper ion. conformational analysis and footprinting of the lac operon. *Biochemistry*, **24**, 8050-6.
50. Li,L., Wu,L.P. and Chandrasegaran,S. (1992) Functional domains in fok I restriction endonuclease. *Proc. Natl. Acad. Sci. U. S. A.*, **89**, 4275-4279.
51. Soundararajan,M., Chang,Z.Y., Morgan,R.D., Heslop,P. and Connolly,B.A. (2002) DNA binding and recognition by the IIs restriction endonuclease MboII. *J. Biol. Chem.*, **277**, 887-95.
52. Jakubauskas,A., Giedriene,J., Bujnicki,J.M. and Janulaitis,A. (2007) Identification of a single HNH active site in type IIS restriction endonuclease Eco31I. *J. Mol. Biol.*, **370**, 157-69.
53. Jakubauskas,A., Sasnauskas,G., Giedriene,J. and Janulaitis,A. (2008) Domain organization and functional analysis of type IIS restriction endonuclease Eco31I. *Biochemistry (N. Y.)*, **47**, 8546-56.
54. Shen,B.W., Landthaler,M., Shub,D.A. and Stoddard,B.L. (2004) DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J Mol Biol*, **342**, 43-56.
55. Van Roey,P., Waddling,C.A., Fox,K.M., Belfort,M. and Derbyshire,V. (2001) Intertwined structure of the DNA-binding domain of intron endonuclease I-TevI with its substrate. *EMBO J*, **20**, 3631-7.
56. Sasnauskas,G., Zakrys,L., Zaremba,M., Cosstick,R., Gaynor,J.W., Halford,S.E. and Siksnys,V. (2010) A novel mechanism for the scission of double-stranded DNA: BfiI cuts both 3'-5' and 5'-3' strands by rotating a single active site. *Nucleic Acids Res.*, **38**, 2399-410.
57. Sasnauskas,G., Kostiuk,G., Tamulaitis,G. and Siksnys,V. (2011) Target site cleavage by the monomeric restriction enzyme BcnI requires translocation to a random DNA sequence and a switch in enzyme orientation. *Nucleic Acids Res.*, **39**, 8844-56.

58. Kaus-Drobek,M., Czapinska,H., Sokolowska,M., Tamulaitis,G., Szczepanowski,R.H., Urbanke,C., Siksnys,V. and Bochtler,M. (2007) Restriction endonuclease MvaI is a monomer that recognizes its target sequence asymmetrically. *Nucleic Acids Res*, **35**, 2035-46.
59. Edgell,D.R., Belfort,M. and Shub,D.A. (2000) Barriers to intron promiscuity in bacteria. *J Bacteriol*, **182**, 5281-9.
60. Crona,M., Moffatt,C., Friedrich,N.C., Hofer,A., Sjoberg,B. and Edgell,D.R. (2011) Assembly of a fragmented ribonucleotide reductase by protein interaction domains derived from a mobile genetic element. *Nucleic Acids Res.*, **39**, 1381-89.

Chapter 5

5 Monomeric site-specific nucleases for genome editing

The work presented in this chapter is reproduced (with permission, Appendix S1) from:

Kleinstiver, B.P., Wolfs, J.M., Kolaczyk, T., Roberts, A.K., Hu, S.X., Edgell, D.R. (2012) Monomeric site-specific nucleases for genome editing. *Proceedings of the National Academy of Sciences USA* 109(21):8061-6

The reproduced text has been modified from the published version by the addition of sections 5.3.5 and 5.3.6 to include data that was not available at the time of publication. Other sections have also been modified accordingly.

5.1 Introduction

Precise genome editing often requires the introduction of a double-strand break (DSB) at defined positions (1-3), and two distinct site-specific DNA endonuclease architectures have been developed towards this goal. One of these architectures relies on reprogramming the DNA-binding specificity of naturally occurring LAGLIDADG homing endonucleases (LHEs) to target desired sequences (4,5). The other architecture utilizes the reprogrammable DNA-binding specificity of zinc-finger proteins or TAL-effector domains that are fused to the non-specific nuclease domain of the type IIS restriction enzyme FokI to create chimeric zinc-finger nucleases (ZFNs) or TAL effector nucleases (TALENs) (6-8). Regardless of the architecture, the underlying biology of the component proteins imposes design challenges and the relative merits of the LHE and the ZFN/TALEN architectures are the subject of much debate in the literature (6,9). One notable constraint imposed by the FokI nuclease domain is the requirement to function as a dimer to efficiently cleave DNA (10,11). For any given DNA target, this necessitates the design of two distinct ZFNs (or two TALENs), such that each pair of zinc finger or TAL effector domains is oriented for FokI dimerization and DNA cleavage (12).

Expanding the repertoire of DNA nuclease domains with distinctive properties is necessary to facilitate the development of new genome editing reagents. Indeed, a number of recent studies have explored the potential of the PvuII restriction enzyme as an

alternative site-specific nuclease domain for genome editing applications (13,14). The PvuII chimeras, however, share similar design constraints as ZFNs and TALENs, requiring two nuclease fusions for precise targeting. In considering alternative nuclease domains for genome editing, we were intrigued by the properties of the GIY-YIG nuclease domain that is associated with a variety of proteins of diverse cellular functions (15). The small (~100 aa) globular GIY-YIG domain is characterized by a structurally conserved central three-stranded antiparallel β sheet, with catalytic residues positioned to utilize a single metal ion to promote DNA hydrolysis (16-18). Intriguingly, the GIY-YIG homing endonucleases, typified by the isoschizomers I-TevI and I-BmoI (19), bind DNA as monomers (20), and generate a DSB with 2-nt, 3' overhangs. It is unknown, however, if GIY-YIG homing endonucleases function as monomers in all steps of the reaction, as the oligomeric status during cleavage has yet to be studied. Notably, GIY-YIG homing endonucleases prefer a specific DNA sequence to generate a DSB (21,22). For I-TevI, the bottom (\uparrow) and top (\downarrow) strand nicking sites lie within a 5'-CN \uparrow NN \downarrow G-3' motif (CNNNG), with the critical G optimally positioned ~28 bp from where the helix-turn-helix (H-T-H) module of the I-TevI DNA-binding domain interacts with substrate (21,22). From an engineering perspective, the modularity and sequence specificity of the GIY-YIG nuclease domain is appealing to create new chimeric endonucleases. Indeed, swapping of the I-BmoI and I-TevI catalytic and DNA-binding domains suggested that the GIY-YIG nuclease domain could be fused to unrelated DNA-targeting platforms (23).

To highlight the genome engineering potential of the GIY-YIG nuclease domain, we fused the domain to 3-member zinc fingers to construct GIY-YIG zinc finger endonucleases (GIY-ZFEs). The GIY-ZFEs are active in bacterial and yeast cells, and *in vitro* data show that they function catalytically as monomers and retain the cleavage specificity associated with the parental GIY-YIG nuclease domain. The GIY-YIG nuclease domain is also portable to the LHE platform, as we constructed monomeric GIY-LHEs that are active *in vivo* and possess ~18-bp binding specificity. We selected LHEs as a DNA targeting domain because of the greater sequence specificity compared to 3-member zinc fingers, the ability to reprogram LHE DNA-binding specificity (24-26), and recent success in generating PuvII-LHE fusions (13). Finally, we demonstrate the flexibility and applicability of the I-TevI nuclease domain by fusing it the TAL platform

(2), revealing that GIY-TALs are functional in yeast and can cleave substrates that contain nucleotide mutations near the cleavage site. Collectively, our data highlight the unique biochemical properties of the GIY-YIG nuclease domain as an alternative to the FokI nuclease domain for genome editing applications.

5.2 Materials and methods

See detailed *S5 Supplementary materials and methods*. Briefly, Tev-ZFE, Tev-LHE, and Tev-TAL fusions and hybrid target sites were modeled in PyMOL using the I-TevI 130C (PDB 1I3J), Zif268 (PDB 1AAY), I-OnuI (PDB 3QQY), and PthXo1 (PDB 3UGM) co-crystal structures (25,27,28,38). The *in vivo* activity of fusions was determined using a two-plasmid bacterial selection (31) or yeast-based reporter assay (that was used to calibrate activity of Tev-ZFEs and Tev-TALs) against a characterized ZFN (35). TevN201-ryA was purified using nickel affinity chromatography to determine the *in vitro* biochemical properties of Tev-ZFEs. Cleavage assays were performed as described (43). A custom Perl script was created to determine CNNNG occurrences relative to 8,829 predicted ZFN sites on zebrafish chromosome 1 (40).

5.3 Results

5.3.1 Construction and validation of GIY-zinc finger endonucleases

To create novel chimeric enzymes, we modeled GIY-zinc finger endonucleases (GIY-ZFEs) using existing crystal structures of the I-TevI 130C DNA binding domain and the Zif268 zinc finger (27,28). One notable feature of our constructs is the polarity, as the I-TevI nuclease domain is fused to the N-terminal end of the 3-finger ryA zinc-finger protein to mimic its native orientation, unlike FokI constructs that are fused to the C-terminal end of zinc-finger proteins. We modeled the Zif268 zinc finger in place of the H-T-H module at the C-terminus of I-TevI, providing the rationale to subsequently fuse various lengths of the I-TevI N-terminal region to the ryA zinc finger that targets a sequence in the *Drosophila rosy* gene to create Tev-ryA zinc finger endonucleases (Tev-ZFEs, Figure 5.1A) (29). The Tev-zinc finger DNA substrates (TZ) consisted of 30 to 38 bps of the I-TevI *td* homing site joined to the 9-bp ryA target site. The TZ substrates differ in the distance of the CNNNG cleavage motif relative to the ryA-binding site

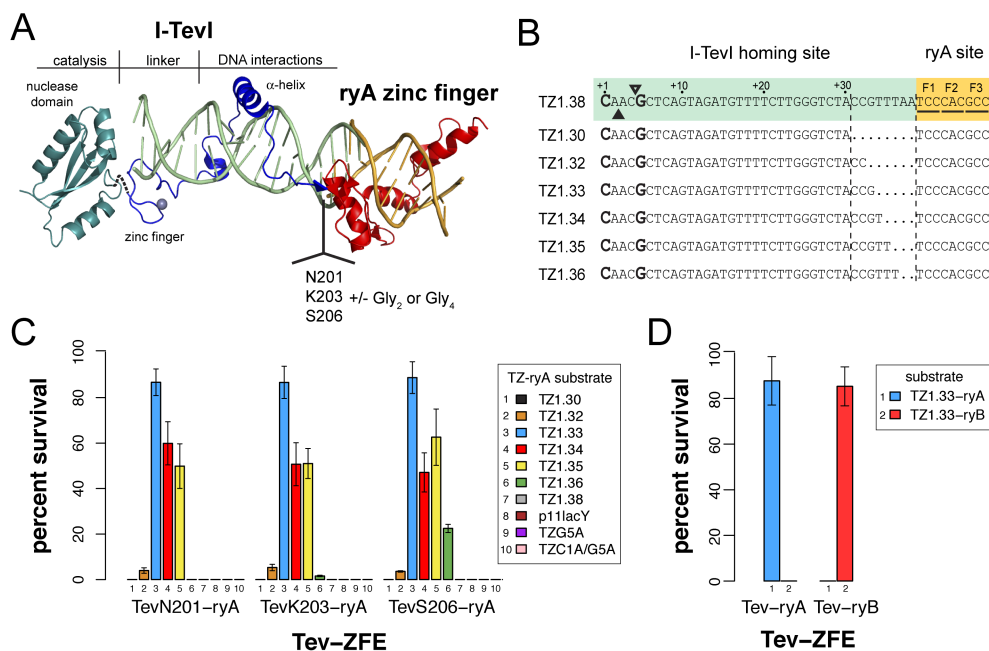


Figure 5.1: Design and functionality of Tev-ZFEs

(A) Modeling of a Tev-zinc finger fusion with DNA substrate (light green) using structures of the I-TevI catalytic domain in green (PDB 1MK0), the I-TevI DNA-binding domain co-crystal in blue (PDB 1I3J), and the Zif268 co-crystal in red (PDB 1AA5) (B) The TZ-ryA substrate is colored according to the structural model. Shown is the top strand of the I-TevI *td* homing site substrate fused to the 5' end of the ryA-binding site for all wild-type substrates tested. The substrate is numbered from the first base of the *td* homing site sequence (the numbering scheme is reverse of that used for the native *td* homing site). The substrates tested differ by insertion or deletion of *td* sequence at the junction of the *td*/ryA sites. (C) Percent survival of three representative Tev-ryA ZFEs in the bacterial two-plasmid selection. All Tev-ryA ZFEs were tested against plasmids containing various length substrates (TZ1.30-1.38), plasmids lacking a target site (p11lacY), and TZ1.33 plasmids with single or double mutations in the CNNNG motif (G5A and C1A/G5A) (Supplementary Table S5.1). (D) Percent survival of TevN201-ryA and TevN201-ryB ZFEs on their cognate and reciprocal target sites. Data are plotted with standard deviation for $n \geq 3$.

(Figure 5.1B). Each TZ substrate possesses a single zinc finger targeting sequence, rather than two head-to-head zinc finger sites necessary for efficient ZFN cleavage. A similar set of I-BmoI-ryA fusions (Bmo-ZFEs) and substrates (BZ) were constructed (Supplementary Figure S5.1). We tested the activity of the GIY-ZFEs using a well-described two-plasmid bacterial selection system, where survival is dependent on the endonuclease cleaving a target plasmid (30,31). Eight Tev-ZFEs were tested on seven TZ substrates cloned into the reporter plasmid (Figure 5.1B and C, Supplementary Table S5.1). In general, the survival of all Tev-ZFEs was highest against TZ substrates where the preferred CNNNG motif was positioned between 33 and 35-bp from the ryA binding site. Low survival (~4-6%) was observed for all Tev-ZFEs against the TZ1.32 substrate, while none survived on the TZ1.30 substrate. Likewise, there was no survival against the longer substrates, with the exception that the longest fusion (TevS206-ryA) exhibited ~22% survival against the TZ1.36 substrate. No survival was observed when the Tev-ZFEs were tested against the target plasmid without a target site (p11lacYwtx1). Mutation of the catalytic arginine 27 of the I-TevI nuclease domain to alanine to create TevR27A-ryAs showed that survival is dependent on GIY-YIG nuclease activity as none of the Tev-R27A constructs survived (Supplementary Table S5.1).

We also constructed and tested a fusion of the TevN201 domain to a different 3-member zinc finger, the ryB zinc finger, creating TevN201-ryB. The TevN201-ryB showed survival in the bacterial selection assay against a corresponding TZ-ryB target, indicating that the I-TevI nuclease domain can function in the context of two different 3-member zinc fingers, but did not survive when tested against the TZ-ryA substrate (Figure 5.1D). Likewise, the TevN201-ryA fusion did not survive against the TZ-ryB substrate, indicating that the zinc finger alone directs DNA-binding. We also tested the Bmo-ZFEs in the genetic selection, but did not observe significant survival for any of the fusions, consistent with the ~750-fold reduced activity of wild-type I-BmoI relative to I-TevI (32). However, as described below, enzymatic activity was detected *in vitro* using purified Bmo-ZFEs. Collectively, these data show that two different GIY-YIG nuclease domains could be fused to zinc finger DNA binding domains to create active site-specific chimeric nucleases.

5.3.2 Tev-ZFEs function as monomers to cleave at a specific sequence

To study the GIY-ZFE biochemical characteristics in more detail, we purified TevN201-ryA for cleavage assays and *in vitro* mapping. We first performed cleavage assays to determine the relationship between TevN201-ryA enzyme concentration and initial reaction velocity using a plasmid substrate with a single TZ-ryA target site. The reaction progress curves indicated an initial burst of cleavage followed by a slower rate of product accumulation (Figure 5.2A), consistent with product release being the rate-limiting step. The initial burst phase was used to estimate initial velocity, and plotting against protein concentration yielded a linear relationship (Figure 5.2A), suggesting that DNA hydrolysis catalyzed by TevN201-ryA is first order with respect to protein concentration.

The model TZ-ryA substrates were designed as a single ryA zinc finger site fused to the I-TevI target sequence. To determine if cleavage by TevN201-ryA was influenced by additional Tev-ryA target sites, we constructed two-site plasmids that differed in whether the target sites were in the same or opposite orientations relative to each other. The single- or two-site plasmids were used in time-course cleavage assays under single-turnover conditions (~10-fold molar excess of protein to substrate) to determine reaction rates. As shown in Figure 5.2B, cleavage of the one-site plasmid yielded $k_{\text{obs}(1\text{-site})} = 0.099 \pm 0.001 \text{ s}^{-1}$, and cleavage of the two-site plasmids with target sites in the opposite or same (Supplementary Figure S5.2B) orientations generated very similar rate constants, $k_{\text{obs}(2\text{-site})} = 0.088 \pm 0.001 \text{ s}^{-1}$ and $0.089 \pm 0.001 \text{ s}^{-1}$, respectively, to the one-site plasmid. In contrast, similar experiments with FokI showed a significant rate enhancement for two-site plasmids relative to one-site plasmids, consistent with FokI functioning as a dimer (33). We conclude that cleavage by TevN201-ryA is non-cooperative and that efficient DNA hydrolysis does not require two sites, consistent with TevN201-ryA functioning catalytically as a monomer.

The I-TevI nuclease domain preferentially cleaves DNA within a 5'-CN↑NN↓G-3' motif, with ↑ and ↓ representing the bottom- and top-strand nicking sites, respectively (22). I-TevI defaults to cleave at the wild-type distance on substrates *in vitro* when this motif is moved closer to, or distant from, the primary binding site, whereas mutants in the I-TevI

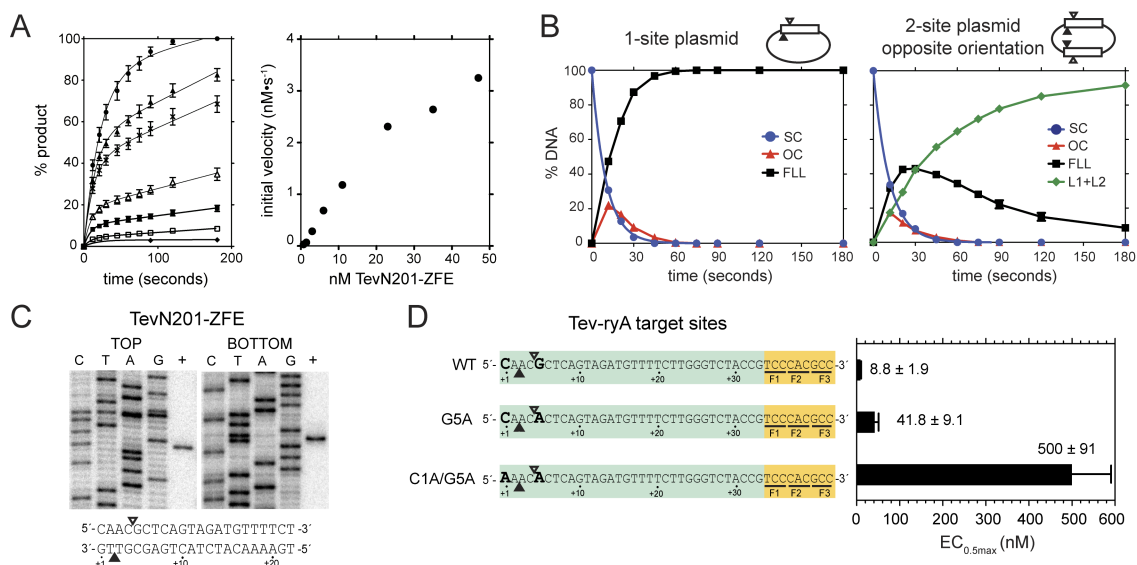


Figure 5.2: TevN201-ZFE is a monomer with a preferred cleavage site

(A) *Left panel*: plot of initial reaction progress for seven TevN201-ZFE concentrations expressed as percent linear product. Protein concentrations from highest to lowest are 47 nM, 32.5 nM, 23 nM, 11nM, 6nM, 3 nM, and 0.7 nM. *Right panel*: graph of initial reaction velocity (nM s^{-1}) versus TevN201-ZFE concentration (nM). (B) Graphical representation of cleavage assays with 90 nM TevN201-ZFE and 10 nM one- or two-site TZ1.33 plasmids (*left and right panels*, respectively). The two-site plasmid had the TZ-ryA sites in the opposite (shown) or same (Fig. S2B) orientation. SC, supercoiled; OC, open-circle (nicked); FLL, full-length linear; L1+L2, linear products. (C) Mapping of TevN201-ZFE cleavage sites on the TZ1.33 substrate, with top and bottom cleavage sites indicated below on the TZ-ryA substrate by open and closed triangles, respectively. (D) Activity of TevN201-ZFE on the wild-type TZ1.33, or the TZ1.33 G5A and TZ1.33 C1A/G5A mutant substrates. A graph of $EC_{0.5\text{max}}$ determinations for each substrate is shown to the right, with $EC_{0.5\text{max}}$ values in nM. Data are plotted as averages of three independent replicates with standard deviations

specific zinc finger cleave at the correct sequence rather than the wild-type distance on mutant substrates (34). To determine the cleavage preference of the TevN201-ryA construct, we mapped the bottom- and top-strand nicking sites using strand-specific end-labeled substrates to the CNNNG motif (Figure 5.2C). Combined with data from the genetic assays showing no survival on substrates that displace the CNNNG motif from an optimal position, our data suggests that in the context of an ryA fusion, the TevN201 domain acts as a molecular ruler with a distance preference.

To further demonstrate TevN201-ryA cleavage preference, we introduced mutations in the CNNNG motif that were previously shown to drastically reduce I-TevI cleavage efficiency (Figure 5.2D) (21,22). Significantly, we observed no survival under selective conditions in the two-plasmid assay on plasmids carrying either the single G5A (CNNNA) or double C1A/G5A (ANNNA) substitutions (Figure 5.1C), equivalent to positions C-27 and G-23 of the I-TevI *td* substrate, respectively. We also performed *in vitro* cleavage assays on wild type and mutant substrates with increasing concentrations of TevN201-ZFE to determine the amount of protein required for half-maximal cleavage ($EC_{0.5max}$). As shown in Figure 5.2D, ~60 fold and ~4.7 fold more protein were required to achieve half-maximal cleavage of the double- and single-mutant substrates relative to the wild-type substrate. The greater substrate discrimination observed in the genetic assay likely reflects lower *in vivo* protein concentrations than those used for *in vitro* cleavage assays. These results show that the TevN201-ryA fusion retains the cleavage specificity of the parental I-TevI enzyme and that double nucleotide substitutions significantly reduce cleavage efficiency. To determine if Bmo-ZFEs also retained substrate specificity, the bottom- and top-strand nicking sites of the BmoN221-ryA fusion were mapped to a 5'-NN↑NN↓G-3' motif, consistent with the cleavage site preference of I-BmoI (Supplementary Figure S5.1D) (19).

5.3.3 Tev-ZFEs function in a yeast-based recombination assay

To extend the *in vivo* relevance of the Tev-ZFE fusions, we utilized a well-described yeast-based recombination assay to test Tev-ZFE function in a eukaryotic system (35). This assay provides a quantitative β -galactosidase readout if the nuclease cleaves its target site that is positioned between a partially duplicated lacZ gene. Furthermore, the

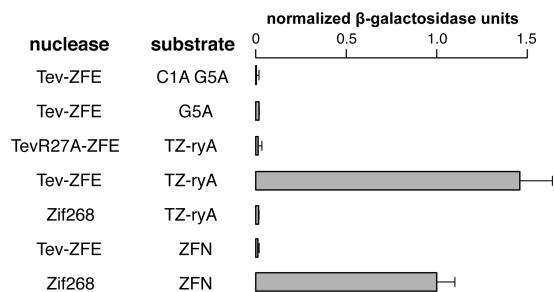


Figure 5.3: Tev-ZFEs can induce recombination in an eukaryotic system

Shown are normalized β -galactosidase units from a yeast-based recombination assay for the indicated nuclease/substrate combinations. Activity was normalized to a homodimeric FokI-Zif268 ZFN positive control. Data are plotted with standard deviation for $n = 4$.

assay allowed us to calibrate TevN201-ryA activity relative to a homodimeric FokI-Zif268 control with previously measured *in vivo* activity sufficient to induce recombination events for genome engineering applications (35). As shown in Figure 5.3, the level of β -galactosidase activity for the TevN201-ryA fusion on its cognate TZ-ryA substrate was ~1.4-fold higher than the Zif268 ZFN control. The TevN201-ryA or Zif268 ZFN constructs displayed no activity on each other's substrates, and activity was dependent on a functional I-TevI nuclease domain, as the TevN201R27A catalytic mutant was unable to induce recombination. Furthermore, TevN201-ryA activity was not observed on mutant substrates where one or both of the critical residues of the CNNNG motif were mutated in the TZ1.33 substrate. Collectively, these assays show that the I-TevI nuclease domain functions in a eukaryotic system with activity on par to a characterized ZFN.

5.3.4 The I-TevI nuclease domain is portable to the LAGLIDADG architecture

To demonstrate that the I-TevI nuclease domain functions in the context of DNA-targeting platforms with greater specificity than 3-member zinc fingers, we constructed fusions of the domain to a catalytically inactive monomeric single-chain LAGLIDADG homing endonuclease (Tev-LHE). As with the Tev-ZFE constructs, we modeled a Tev-LHE chimera using the co-crystal of I-OnuI with its DNA substrate such that the I-TevI nuclease and linker domains were fused to the N-terminus of I-OnuI, which is partially disordered in the structure (Figure 5.4A) (25). Based on this model, we fused TevN201G₄ and TevK203 fragments to a catalytically dead I-OnuI E1 E22Q mutant. A series of model DNA substrates were constructed by fusing the *td* target site to the I-OnuI E1 binding site in the human MAO-B gene, differing in the position of the CNNNG cleavage motif relative to the I-OnuI E1 site (TO1.12 to TO1.30) (Figure 5.4B).

In the bacterial two-plasmid selection, we found that the TevN201G₄-Onu and TevK203-Onu fusions were active against a range of DNA substrates. Notably, the fusions displayed maximal survival on longer targets (TO1.26, TO1.28, and TO1.30), and lower survival against shorter targets (TO1.18 and TO1.20 targets). The two groups of substrate differ by approximately one helical turn of DNA, meaning that the preferred CNNNG

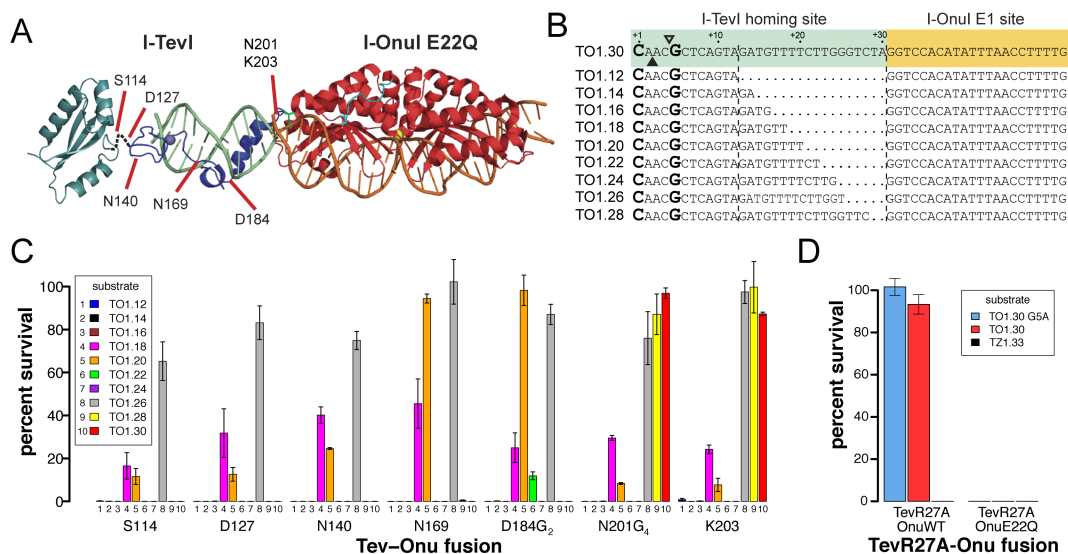


Figure 5.4: Design and functionality of Tev-LHEs

(A) Modeling of a Tev-Onu E1 fusion with DNA substrate (light green) using structures of the I-TevI catalytic domain in green (PDB 1MK0), the I-TevI DNA-binding domain co-crystal in blue (PDB 1I3J), and the I-OnuI co-crystal in red (PDB 3QQY). Shown are fusion points at which the I-TevI fragment has been shortened. (B) The Tev-Onu E1 (TO) substrate is colored according to the structural model. Shown is the top strand of the I-TevI *td* homing site substrate fused to the 5' end of the Onu E1-binding site. The substrates are numbered from the first base of the *td* homing site sequence and differ by the deletion of *td* nucleotides at the junction of the *td*/Onu E1 sites. (C) Percent survival of Tev-LHEs in the bacterial two-plasmid selection with various length target sites (TO1.12-1.30). All Tev-LHEs tested were in the I-OnuI E1 E22Q background. (D) Percent survival of TevR27A(N201G₄)-OnuE1 and TevR27A(N201G₄)-OnuE1(E22Q) on TO1.30, TO1.30G5A, and TZ1.33. Data are plotted with standard deviation for n = 3.

motif would be presented on the same face of the substrate even though the motif is closer to the I-OnuI E1 binding site on the shorter targets. Similar periodic cleavage patterns have been observed *in vitro* with I-TevI on substrates with a displaced CNNNG motif (36). This result also implies that the N-terminus of I-OnuI possesses inherent flexibility to allow the I-TevI nuclease domain to search out the CNNNG motif, in contrast to the ruler-like behaviour of the Tev-ZFE constructs, likely because the zinc finger N-terminus is inflexible. Importantly, the Tev-Onu fusions were not active against the TZ-ryA zinc finger substrates (Supplementary Table S5.2), showing that the LHE, and not the I-TevI linker, directs DNA targeting. Survival was also dependent on an active I-TevI nuclease domain, as TevR27A fusions in the context of the I-OnuI E22Q mutant did not survive (Figure 5.4D). Conversely, the targeting and activity of wild-type I-OnuI E1 was not affected by fusion of the I-TevI domain, as the TevR27A-OnuWT fusions survived against TO substrates (Figure 5.4D).

The apparent flexibility of the N-terminus and the greater specificity of I-OnuI prompted us to test fusions containing shorter fragments of the I-TevI nuclease domain (Figure 5.4A). Based on structural and genetic data, we constructed TevS114-Onu, TevD127-Onu, TevN140-Onu, TevN169-Onu, and TevD184G₂-Onu fusions, progressively removing amino acid residues of I-TevI that make specific base-pair contacts to the *td* substrate (Figure 5.4A) (28). Notably, the TevS114, TevD127, TevN140 and TevN169 removed the α -helix that binds in the minor groove, as well as residues shown by structural data to make base-specific contacts (28). The TevS114 fusion point lies at the boundary of the deletion tolerant region of the I-TevI linker, and represents a functionally minimal GIY-YIG nuclease domain (36,37). We found that the shorter fusions were not active against the longer TO1.28 and TO1.30 substrates, yet displayed the same periodic activity on the shorter substrates (Figure 5.4C and Supplementary Table S5.2). A single exception was the TevD184G₂ fusion that showed low survival against the TO1.22 substrates, against which no other fusion survived. No survival was observed on mutant substrates that contained single (CNNNA) or double (ANNNA) mutations in the CNNNG motif, recapitulating the necessity for an appropriately positioned CNNNG as seen with the Tev-ZFE fusions.

5.3.5 I-TevI fusions to the TAL effector PthXo1 are functional

To expand the binding capacity of chimeric I-TevI nucleases beyond that of engineered ZFEs and LHEs, the logical next step was to model fusions of the I-TevI nuclease domain with the recently solved crystal structure of the modular PthXo1 TAL effector DNA-binding platform (Figure 5.5A) (38). The repetitive structure of TAL effectors can be exploited to assemble highly specific DNA-binding modules whose nucleotide specificity is dictated at a 1:1 ratio via the RVDs (repeat variable diresidues) within the repeated unit (2,7,35). To capitalize on the customizable specificity of TALs, multiple truncation points within the I-TevI linker and the N- and C-terminal ends of the PthXo1 backbone were explored to generate a set of I-TevI-PthXo1 fusions (Tev-TALs, Figure 5.5B). Additionally, model hybrid substrates comprising *td* and PthXo1 target site sequences with varied spacer lengths were cloned into the yeast based recombination assay vector to assess activity.

Several constructs revealed appreciable *in vivo* activity, with the N169-T120 Tev-PthXo1 fusion demonstrating the most robust activity across targets with varied spacer lengths (Figure 5.5C). The periodic activity of the N169 I-TevI truncation is reminiscent of the cleavage pattern observed for the similar Tev-LHE construct, though the magnitude of activity across a larger number of targets appears to be greater for the Tev-TAL platform. This may reflect an increased flexibility of the N169 fusion in the context of the TAL versus LHE platforms, or may result from greater sensitivity inherent to the *in vivo* activity assays (yeast reporter compared to survival in bacteria). Similar to the Tev-ZFEs tested in the yeast-based recombination system, the N169-T120 construct has activity on par with a ZFN pair that induces recombinogenic events at a level sufficient for gene targeting in plant and mammalian cells (35).

5.3.6 Tev-TALs tolerate nucleotide substitutions in the DNA spacer

To determine the tolerance of the Tev-TAL architecture to mutations that intervene the cleavage motif and substrate spacer, hybrid target sites were generated containing sequences from alternative bacteriophage thymidylate synthase alleles. The N169-T120 construct efficiently cleaved most substrates with activity comparable to the cognate *td*

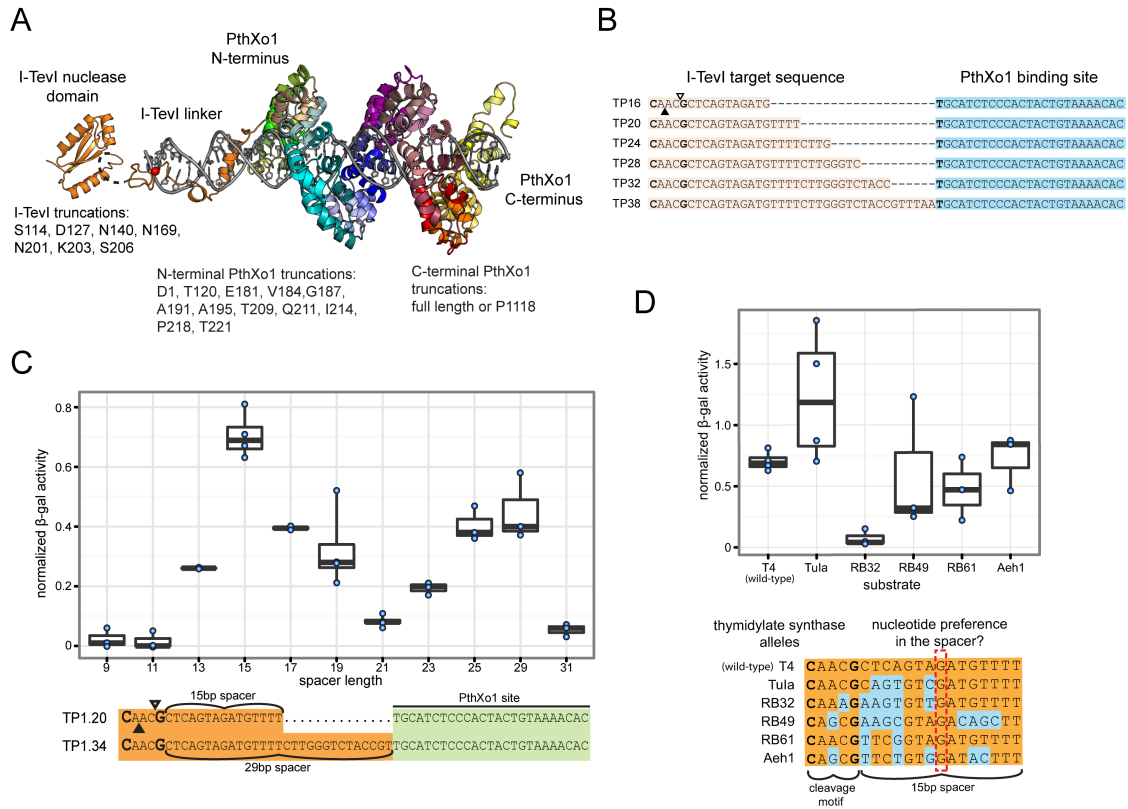


Figure 5.5: Design and functionality of Tev-TALs

(A) Modeling Tev-PthXo1 fusions using structures of the I-TevI nuclease and DNA-binding domains coloured in orange (PDB 1MK0 and 1I3J), aligned with the multicoloured PthXo1 cocystal (PDB 3UGM). Various truncation points of I-TevI and PthXo1 were utilized to identify functional constructs (B) Tev-PthXo1 model substrates with *td* sequence coloured orange and the PthXo1 binding site coloured blue. Substrates are numbered from the first base of the *td* homing site and differ by the number of nucleotides included prior to the junction of the *td*/PthXo1 binding site. (C) Activity of the N169-T120 (I-TevI and N-terminal PthXo1 truncation points, respectively) TevPthXo1 construct on various model substrates, reported as β -galactosidase activity normalized to a Zif268 ZFN control. Substrates are numbered according to the length of the spacer sequence that intervenes the G of the CnnnG cleavage motif and the PthXo1 binding site. (D) Normalized β -galactosidase activity of N169-T120 TevPthXo1 on mutant target sites generated from alternative thymidylate synthase alleles.

allele from bacteriophage T4 (Figure 5.5D). Observed exceptions included the Tu1a allele that displayed activity almost 2-fold greater than the *td* target, and the RB32 allele that was poorly cleaved. These results are consistent with previous *in vitro* results that suggested I-TevI is tolerant to nucleotide substitutions within its homing site (39). Interestingly, the nucleotide analogous to the critical G+7 anchor point observed for the GIY-YIG homing endonuclease I-BmoI was not mutated in any of the tested alleles (Figure 5.5D, see also Chapters 4.3.7 and 6.4).

5.3.7 A 5'-CNNNG-3' cleavage motif is not limiting for targeting

An important consideration in the design of engineered GIY-YIG nucleases for genome-editing applications is the targeting requirements, notably the need for the CNNNG di-nucleotide cleavage motif (Figure 5.6A). In a complex genome of $\sim 3 \times 10^9$ bp, the statistically predicted occurrence of the CNNNG motif is once every 15 bp assuming a 50% GC content. To determine if the frequency of the CNNNG motif would be limiting for targeting applications, we examined 35 bp flanking 8,829 computationally predicted ZFN sites on zebrafish chromosome 1 for the occurrence of the CNNNG motif (40). As shown in Figure 5.6B, the motif is highly represented at all positions within a 35-bp window relative to the ZFN sites. Of the 8,829 sites examined, 88% (7,845) of ZFN sites possessed at least one motif within 35 bp of the predicted binding site (Figure 5.6C). These requirements contrast sharply with those of the recently described PvuII-LHEs and PvuII-ZFNs that require the 6-bp 5'-CAGCTG-3' PvuII site in addition to the LHE or ZF binding site (13,14). Of the 8,829 ZFN sites, 97% lacked a PvuII site within the 35-bp window (Supplementary Figure S5.3). Thus, the requirement for a di-nucleotide cleavage motif in the context of GIY-YIG engineered nucleases will not severely limit potential targeting sites.

5.4 Discussion

Here, we provide evidence that the GIY-YIG nuclease domain is a potential alternative to the currently used FokI nuclease domain for genome editing applications. We show that the I-TevI GIY-YIG nuclease domain is portable to the three reprogrammable DNA-binding scaffolds utilized in the field, the 3-member zinc fingers, catalytically inactive

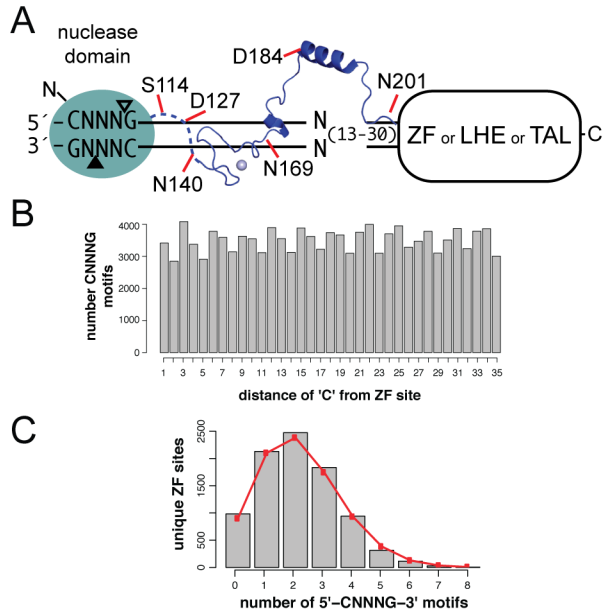


Figure 5.6: Cleavage requirements do not limit GIY-EN applicability

(A) A diverse set of monomeric and sequence specific reagents can be generated by fusing distinct GIY-YIG domain linker lengths to engineered DNA-binding platforms, including zinc-finger arrays, inactive LAGLIDADGs, and TAL effectors (GIY-ENs). (B) Shown is the distribution of the CNNNG motif in a 35-bp window flanking 8,829 predicted ZFN sites on zebrafish chromosome 1. The number of occurrences of the 'C' of the motif at each distance is indicated. (C) Unique ZFN sites were grouped according to the number of occurrences of the CNNNG motif in the 35-bp window. The red line is the expected number of ZFN sites for each group based on a binomial distribution.

LAGLIDADG homing endonucleases, and modular TAL effectors. The Tev-ZFE, Tev-LHE, and Tev-TAL fusions are active *in vitro* and *in vivo*, with the activity of Tev-ZFEs and Tev-TALs in a yeast-based recombination assay on par with that of a characterized ZFN. We foresee the monomeric nature of the I-TevI fusions to each platform as a key advantage over existing ZFNs and TALENs, as a single fusion protein need be designed to target a given sequence, rather than two ZFNs or TALENs required to promote dimerization of the FokI nuclease domain (12). Moreover, the fact that the I-TevI nuclease domain possesses a preferred cleavage motif adds another layer of specificity to targeting requirements, potentially limiting DSBs at off-target sites that do not possess the cleavage motif.

One targeting consideration for chimeric GIY-YIG endonucleases is the DNA sequence requirement of the I-TevI linker. The I-TevI linker is a complex structure, consisting of defined structural elements with distinct roles in I-TevI function (28,34,36). The primary role of the linker is to position the nuclease domain on substrate for cleavage at the CNNNG motif, which is found at a defined distance from the binding site on naturally occurring I-TevI substrates. However, the linker can direct the nuclease domain *in vitro* to search out displaced CNNNG motifs on both native and non-native substrates with insertions or deletions, albeit with reduced cleavage efficiency (39). Our Tev-LHE fusions recapitulate this distance versus sequence behaviour *in vivo*, as the fusions can cleave displaced CNNNG motifs with a periodicity that parallels the helical nature of DNA. We partially attribute this ability of the Tev-Onu fusions to the flexible N-terminus of I-OnuI. The substrate flexibility of different length Tev-Onu fusions is an important consideration for targeting, as CNNNG motifs at various positions relative to the LHE binding site would be accessible by the choice of the appropriate Tev-LHE fusion. In contrast, the apparently inflexible N-terminus of the 3-member zinc fingers constrains cleavage to a distance of 33-36 bp from the ryA-binding site, mimicking the spacing of the CNNNG motif on native *td* substrate. Our longest Tev-ZFE and Tev-LHE fusions encompass all the known elements of the I-TevI linker that make multiple base-specific and non-specific contacts to DNA (28). However, biochemical studies revealed that I-TevI retains significant cleavage activity on substrates with multiple substitutions in the central region of its cognate DNA substrate that is contacted by the linker, equivalent to

positions 6-33 of our longest chimeric substrates (39). The shortest Tev-LHE fusions do not contain any linker elements that are known to make base-specific DNA contacts, and cleave only at the preferred CNNNG motif, implying that the I-TevI linker may contact substrate nucleotides adjacent to the CNNNG motif. The presumption the I-TevI linker can tolerate changes within the substrate spacer is supported by the fact that Tev-TAL fusions can cleave substrates composed of alternative thymidylate synthase alleles. The substrates examined, however, did not provide a saturating mutagenic approach making it difficult to draw conclusions about all nucleotide positions. Interestingly, no changes were made at a position equivalent to the G+7 nucleotide in the GIY-YIG homing endonuclease I-BmoI substrate that corresponds to a critical protein:DNA anchor (Chapter 4.3.7). Should this functional interaction be conserved within the context of I-TevI chimeric nucleases, potential contacts may play a role in the positioning of the nuclease domain rather than being necessary for cleavage (36).

Future work on Tev-ZFEs, Tev-LHEs, and Tev-TALs will require a more thorough dissection of binding affinity and specificity, and characterization of cellular toxicity that results from cleavage at off-target sites. In their current form, the targeting specificity of the Tev-ZFEs is a function of the 3-zinc finger domain, which could be further enhanced by addition of zinc fingers to generate a 4-, 5-, or 6-zinc finger fusion to increase specificity, as has been done with a variety of ZFNs (41). In contrast, the ~18-bp specificity of LHEs is sufficient to direct targeting and cleavage at endogenous loci in human cells. LHEs, however, are tolerant of nucleotide substitutions within their recognition sequence, and I-OnuI E1 cleaves off-target sites that differ by one or two nucleotide substitutions (25). In the context of Tev-LHEs, decoupling of DNA-cleavage and DNA-binding activity by using a catalytically dead LHE scaffold, combined with the requirement for a preferred I-TevI CNNNG cleavage motif, would significantly reduce cleavage at off-target sites (Supplementary Figure S5.4). Another advantage of the decoupled activities of Tev-LHEs is that they would not require re-optimization of catalytic activity that is often necessary in LHEs that have been reprogrammed to bind non-native target sites (25,42). The Tev-TAL platform, however, appears to be able provide the specificity required to target a monomeric nuclease within the human genome. Studies have shown that TALENs can be specifically targeted to sites in a

variety of organisms (7,35), and that assembly of custom DNA-binding units is relatively straightforward (2,35). Similar to the exploration of alternative DNA-binding platforms (2), it is imperative to incorporate nuclease domains with distinct biochemical properties into the genome engineering pipeline to create highly precise tools. With further optimization, the I-TevI nuclease domain may become an alternative to the FokI-derived ZFNs and TALENs.

5.5 References

1. Bibikova M, Beumer K, Trautman JK, & Carroll D (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science* 300(5620):764.
2. Bogdanove AJ & Voytas DF (2011) TAL effectors: customizable proteins for DNA targeting. *Science* 333(6051):1843-1846.
3. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, & Gregory PD (2010) Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* 11(9):636-646.
4. Marcaida MJ, Munoz IG, Blanco FJ, Prieto J, & Montoya G (2010) Homing endonucleases: from basics to therapeutic applications. *Cell Mol Life Sci* 67(5):727-748.
5. Stoddard BL (2011) Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. *Structure* 19(1):7-15.
6. Cathomen T & Joung JK (2008) Zinc-finger nucleases: the next generation emerges. *Mol Ther* 16(7):1200-1207.
7. Christian M, *et al.* (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186(2):757-761.
8. Kim YG, Cha J, & Chandrasegaran S (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A* 93(3):1156-1160.
9. Halford SE, Catto LE, Pernstich C, Rusling DA, & Sanders KL (2011) The reaction mechanism of FokI excludes the possibility of targeting zinc finger nucleases to unique DNA sites. *Biochem Soc Trans* 39(2):584-588.
10. Bitinaite J, Wah DA, Aggarwal AK, & Schildkraut I (1998) FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci U S A* 95(18):10570-10575.
11. Vanamee ES, Santagata S, & Aggarwal AK (2001) FokI requires two specific DNA sites for cleavage. *J Mol Biol* 309(1):69-78.

12. Smith J, *et al.* (2000) Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res* 28(17):3361-3369.
13. Fonfara I, Curth U, Pingoud A, & Wende W (2011) Creating highly specific nucleases by fusion of active restriction endonucleases and catalytically inactive homing endonucleases. *Nucleic Acids Res* 40(2):847-860.
14. Schierling B, *et al.* (2011) A novel zinc-finger nuclease platform with a sequence-specific cleavage module. *Nucleic Acids Res*:doi.10.1093/nar/gkr1112.
15. Dunin-Horkawicz S, Feder M, & Bujnicki JM (2006) Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics* 7:98.
16. Mak AN, Lambert AR, & Stoddard BL (2010) Folding, DNA recognition, and function of GIY-YIG endonucleases: crystal structures of R.Eco29kI. *Structure* 18(10):1321-1331.
17. Sokolowska M, Czapinska H, & Bochtler M (2011) Hpy188I-DNA pre- and post-cleavage complexes--snapshots of the GIY-YIG nuclease mediated catalysis. *Nucleic Acids Res* 39(4):1554-1564.
18. Van Roey P, Meehan L, Kowalski JC, Belfort M, & Derbyshire V (2002) Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI. *Nat Struct Biol* 9(11):806-811.
19. Edgell DR & Shub DA (2001) Related homing endonucleases I-BmoI and I-TevI use different strategies to cleave homologous recognition sites. *Proc Natl Acad Sci U S A* 98(14):7898-7903.
20. Mueller JE, Smith D, Bryk M, & Belfort M (1995) Intron-encoded endonuclease I-TevI binds as a monomer to effect sequential cleavage via conformational changes in the *td* homing site. *EMBO J* 14(22):5724-5735.
21. Bryk M, Belisle M, Mueller JE, & Belfort M (1995) Selection of a remote cleavage site by I-TevI, the *td* intron-encoded endonuclease. *J Mol Biol* 247(2):197-210.
22. Edgell DR, Stanger MJ, & Belfort M (2004) Coincidence of cleavage sites of intron endonuclease I-TevI and critical sequences of the host thymidylate synthase gene. *J Mol Biol* 343:1231-1241.
23. Liu Q, Derbyshire V, Belfort M, & Edgell DR (2006) Distance determination by GIY-YIG intron endonucleases: discrimination between repression and cleavage functions. *Nucleic Acids Res* 34(6):1755-1764.
24. Rosen LE, *et al.* (2006) Homing endonuclease I-CreI derivatives with novel DNA target specificities. *Nucleic Acids Res* 34(17):4791-4800.

25. Takeuchi R, *et al.* (2011) Tapping natural reservoirs of homing endonucleases for targeted gene modification. *Proc Natl Acad Sci U S A* 108(32):13077-13082.
26. Jacoby K, *et al.* (2012) Expanding LAGLIDADG endonuclease scaffold diversity by rapidly surveying evolutionary sequence space. *Nucleic Acids Res*:doi.10.1093/nar/gkr1303.
27. Elrod-Erickson M, Rould MA, Nekludova L, & Pabo CO (1996) Zif268 protein-DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions. *Structure* 4(10):1171-1180.
28. Van Roey P, Waddling CA, Fox KM, Belfort M, & Derbyshire V (2001) Intertwined structure of the DNA-binding domain of intron endonuclease I-TevI with its substrate. *EMBO J* 20(14):3631-3637.
29. Bibikova M, Golic M, Golic KG, & Carroll D (2002) Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics* 161(3):1169-1175.
30. Chen Z & Zhao H (2005) A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic Acids Res* 33(18):e154.
31. Kleinstiver BP, Fernandes A, Gloor GB, & Edgell DR (2010) A unified genetic, computational, and experimental framework identifies non-conserved residues as critical for function of the homing endonuclease I-BmoI. *Nucleic Acids Res* 38:2411-2427.
32. Edgell DR, Stanger MJ, & Belfort M (2003) Importance of a single base pair for discrimination between intron-containing and intronless alleles by endonuclease I-BmoI. *Curr Biol* 13(11):973-978.
33. Catto LE, Ganguly S, Milsom SE, Welsh AJ, & Halford SE (2006) Protein assembly and DNA looping by the FokI restriction endonuclease. *Nucleic Acids Res* 34(6):1711-1720.
34. Dean AB, *et al.* (2002) Zinc finger as distance determinant in the flexible linker of intron endonuclease I-TevI. *Proc Natl Acad Sci U S A* 99(13):8554-8561.
35. Cermak T, *et al.* (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* 39(12):e82.
36. Liu Q, *et al.* (2008) Role of the interdomain linker in distance determination for remote cleavage by homing endonuclease I-TevI. *J Mol Biol* 379(5):1094-1106.
37. Kowalski JC, *et al.* (1999) Configuration of the catalytic GIY-YIG domain of intron endonuclease I-TevI: coincidence of computational and molecular findings. *Nucleic Acids Res* 27(10):2115-2125.

38. Mak AN, *et al.* (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* 335(6069):716-719.
39. Bryk M, *et al.* (1993) The *td* intron endonuclease I-TevI makes extensive sequence-tolerant contacts across the minor groove of its DNA target. *EMBO J*, 12(5):2141-2149.
40. Foley JE, *et al.* (2009) Rapid mutation of endogenous zebrafish genes using zinc finger nucleases made by Oligomerized Pool ENgineering (OPEN). *PLoS One* 4(2):e4348.
41. Shimizu Y, *et al.* (2011) Adding fingers to an engineered zinc finger nuclease can reduce activity. *Biochemistry* 50(22):5033-5041.
42. Takeuchi R, Certo M, Caprara MG, Scharenberg AM, & Stoddard BL (2009) Optimization of in vivo activity of a bifunctional homing endonuclease and maturase reverses evolutionary degradation. *Nucleic Acids Res* 37(3):877-890.
43. Kleinstiver BP, Berube-Janzen W, Fernandes AD, & Edgell DR (2011) Divalent metal ion differentially regulates the sequential nicking reactions of the GIY-YIG homing endonuclease I-BmoI. *PLoS One* 6(8):e23804.

Chapter 6

6 Discussion

Selfish mobile genetic elements (MGEs) that encode DNA endonucleases can initiate the lateral transfer of genetic material between organisms. Consequently, MGEs comprise a significant proportion of any given genome (1-3), subsets of which encode genes of site-specific homing endonucleases (HEs) (4-6). HEs encode six distinct class-defining active-site motifs and differ from most other cellular endonucleases by targeting long recognition sites (14-36 base pairs) (5,7). While each family of HE utilizes a number of different strategies to cleave DNA, the mechanism by which GIY-YIG family homing endonucleases (GIY-HEs) initiate their own genomic mobility has remained poorly understood until now. Within this thesis I demonstrate that the model GIY-HE I-BmoI functions as a monomer and investigate the dynamic protein:DNA interactions that contribute to DNA-hydrolysis. Further, I explicitly test models for double-strand break (DSB) formation by monomeric nucleases to elucidate the mechanism by which single active site GIY-HEs cleave DNA. To apply these findings within the context of engineered nucleases, I demonstrate that the cleavage properties of GIY-HEs are recapitulated in the context of GIY-YIG engineered nucleases (GIY-ENs) when the nuclease domain of I-TevI is fused to three different re-targetable DNA-binding platforms. GIY-ENs are functional *in vivo* and operate via a different mechanism than canonical FokI nuclease domain fusions that constitute the majority of engineered nuclease architectures. Distinctly, GIY-ENs cleave at a preferred nucleotide motif and function as monomers, whereas FokI-derived nucleases cleave nonspecifically and function as dimers. These topics are elaborated further below, where I discuss the advantageous properties of GIY-ENs for genome editing applications.

6.1 GIY-HE cleavage mechanism

Unlike DNA hydrolysis mechanisms that have been established for other HE families (5,7), the inherently high nuclease activity of the GIY-HE I-TevI prohibited expression and purification from bacteria that would enable detailed *in vitro* or *in vivo* analyses. To

investigate the quarter-century old question of how GIY-HEs generate a DSB, I purified the related GIY-HE I-BmoI to study the sequential protein:DNA interactions that mediate DNA cleavage. DSB formation by I-BmoI proceeds through a number of substrate distortions that are dependent on contacts to nucleotides that surround the cleavage site (8,9). In Chapters 2 and 3, I investigated the contributions of amino acid and nucleotide positions to hydrolysis by generating mutations in the catalytic domain or substrate cleavage motif, respectively. In Chapter 4, I discovered additional protein:DNA interactions outside of the cleavage motif that act as an anchor point to position the I-BmoI linker on substrate and initiate DNA distortions. Importantly, I determined that I-BmoI functions as a monomer at all steps of the reaction pathway and does not hydrolyze DNA through protein- or DNA-mediated transesterification reactions. The cumulative evidence from Chapters 2, 3, and 4 suggests a highly dynamic monomeric cleavage mechanism and also validates predictions made previously for I-TevI (10). It is plausible that I-TevI is mechanistically comparable to I-BmoI, as they share high similarity in their nuclease domains and linkers (11-13), bind analogous thymidylate synthase alleles as monomers (10,14), encode single predicted active sites (9,13,15), and induce significant distortions in their respective substrates adjacent to their cleavage site (8,10). I therefore propose a model for DSB formation by GIY-HEs where the DNA-binding domain anchors the low-affinity GIY-YIG nuclease domain on substrate to sequentially nick both DNA strands at a preferred sequence (Figure 6.1).

To further characterize the sequential conformational changes that mediate DSB formation, it may be possible to perform fluorescence resonance energy transfer (FRET) experiments as has been done to track changes in the PI-SceI:substrate complex through the cleavage reaction (16). I-BmoI contains a single cysteine within the nuclease domain (exposed on the surface opposite to the catalytic cleft), and one cysteine within the remainder of the protein located in the linker (which has previously been mutated to serine without affecting activity, Figure 4.5). Using FRET, the effects of amino acid mutations within the nuclease domain, roles of divalent metal and nucleotide substitutions at the cleavage site, and effects of substrate insertions within the spacer detailed in Chapters 2, 3, and 4, respectively, could be further examined and exploited to confirm the mechanistic predictions for GIY-HEs.

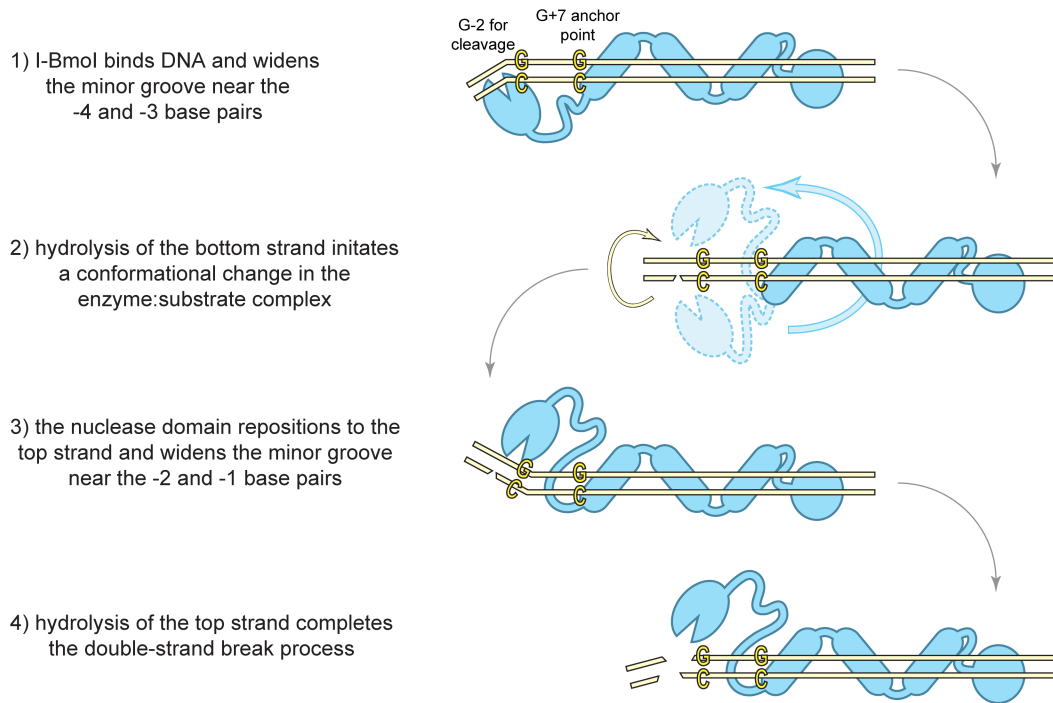


Figure 6.1: Conformational change DSB mechanism for I-BmoI

I-BmoI utilizes multiple DNA-distortions and a critical anchor point within the linker to overcome the inherently weak DNA-binding affinity of the nuclease domain to hydrolyze DNA.

The proposed monomeric DSB mechanism for GIY-HEs contrasts with the models established for other endonucleases of the GIY-YIG family (17,18). As described in Chapter 1.2.2, other enzymes that contain the GIY-YIG motif have acquired context dependent adaptations that promote oligomeric assembly to nick both DNA strands (18-22), or constitute components of larger architectures that allow the GIY-YIG domain to function as a nickase (23,24). The GIY-YIG nuclease domain is therefore highly adaptable and can function via context-dependent single- or double-strand hydrolysis mechanisms. Interestingly, studies that swapped the nuclease and DNA-binding domains of I-TevI and I-BmoI revealed that each domain retained their intrinsic functional properties when removed from their natural context (11). Additionally, the sequence-tolerant DNA-binding domains utilized by I-BmoI and I-TevI display evidence of shared functionality with other enzymes, as similar NUMOD and helix-turn-helix modules are found in the $\beta\beta\alpha$ -Me HE family (5,25-28). These properties, along with the evolutionary flexibility and mechanistic plasticity of the nuclease motif, suggested the GIY-YIG domain as an ideal candidate to consider within the context of novel engineered nucleases.

6.2 Considerations for engineered nucleases

The ability to manipulate genomes for reverse genetics studies in model organisms or alter genes in human cell lines for the treatment of monogenic diseases has led to significant academic, clinical, and biotechnological interest in the development and implementation of engineered nucleases. Their application has been so widespread that in 2011, Nature Methods chose genome editing with engineered nucleases as their Method of the Year (29). The efficacy of a zinc-finger nuclease (ZFN) mediated therapy to treat HIV-1 infection is currently the subject of a clinical trial (30). Unfortunately, the inherent biology of the FokI nuclease component of ZFNs was not rigorously understood prior to the study, with many suggesting inadequate fidelity of engineered ZFNs for human cell therapy (31,32). The requirement of the FokI nuclease domain to form a dimer and its ability to cleave DNA non-specifically has led to undesirable off-target DSBs in model systems. Off-target DSBs can occur when one ZFN module binds an unintended degenerate target site and recruits another ZFN to initiate cleavage (Figure 6.2A). In this

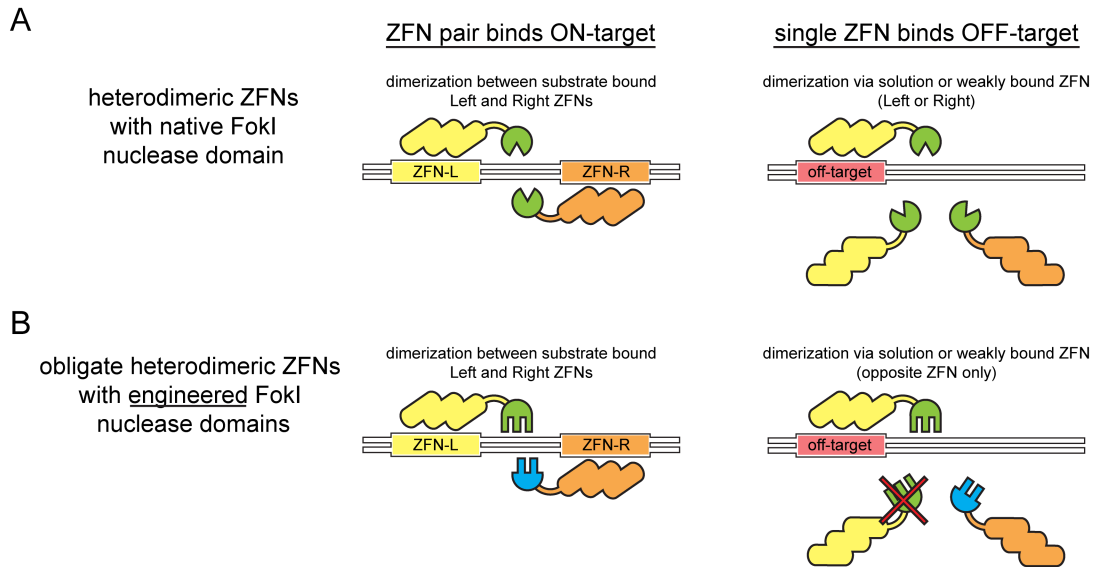


Figure 6.2: FokI requirements can lead to off-target DSBs

(A) The dimerization mechanism of FokI and imperfect binding specificity of zinc-fingers leads to promiscuous cleavage at off-target sites where either or both ZFN units interact with an unintended target. (B) ZFNs that contain an engineered obligate heterodimeric version of the FokI nuclease domain reduce mechanisms of off-target cleavage, but still remain prone to a subset of off-target DSBs.

scenario, the secondary ZFN molecule is not necessarily substrate-bound and can also result from homodimerization between the FokI domains of identical ZFN subunits. To circumvent undesirable homodimerization, studies have been conducted to engineer an obligate heterodimeric FokI dimer interface where a 'left' and 'right' ZFN interaction constitutes the only condition that is productive for cleavage (Figure 6.2B) (33). Unfortunately, dimerization can still occur between a DNA-bound left ZFN unit and a right ZFN that is weakly bound or in solution, leading to apprehension regarding the specificity of FokI-derived ZFNs and TALENs (TAL effector nucleases) (31,34). The frequencies of off-target DSBs at unintended sites are an important consideration for both platforms, as they can have toxic consequences in the form of severe adverse events (discussed in Chapter 1.5).

Another aspect of genome editing procedures to consider is the ability to bias intrinsic cellular DNA-repair pathways to mutagenic nonhomologous end-joining (NHEJ) events in the absence of an exogenous template, or corrective homologous recombination (HR) pathways when an exogenous donor template is provided. NHEJ mediated events can either proceed through a classical mechanism that generally results in precise repair or through an alternative NHEJ pathway where DNA end-processing enzymes generate deletions and frameshifts (35,36). In scenarios where deleterious events are desired, studies have shown that engineered nucleases can be coupled with exonucleases to increase the frequency of mutagenic repair via alternative NHEJ pathways (37). Unfortunately, NHEJ-mediated repair events are often associated with significant genomic instability in the forms of sequence loss or chromosomal translocations. Consequently, there has been interest in the development of engineered site-specific nickases to circumvent these unfavourable outcomes, as they initiate HR at similar frequencies to cleavases, yet do not activate NHEJ (38). Nickase variants of engineered LAGLIDADG HEs (LHEs) and FokI-derived platforms have therefore been explored to more precisely modulate gene repair outcomes (39,40). As described in Chapter 6.3, it may also be possible to expand the potential of GIY-ENs by generating engineered GIY-YIG nickases.

Another important consideration beyond the mutagenic event induced by each nuclease technology is the specificity associated with the three DNA-targeting platforms described within this thesis. The construction of highly precise zinc-finger arrays has not proven to be as trivial as once hoped as many nucleotide triplets remain inaccessible by engineered zinc-fingers (41), and the modular assembly of multiple fingers does not appear to circumvent their inherent promiscuous binding (32). These issues are further compounded by the fact that the specificity provided by an array of 4-5 zinc-fingers (that bind a 12-15 base pair target) within the context of a monomeric nuclease may be insufficient for targeting in human cells. Conversely, many LHEs offer the capacity to bind genomic sites with greater than 18 base pair specificity. Combined with their small size and potential to maintain a second active catalytic motif, engineered LHEs represent an interesting platform to continue to pursue within the context of GIY-YIG fusions. TAL effectors have shown great promise for the design and assembly of modules that can target virtually any sequence with binding sites that exceed 20 base pairs. While their large size and repetitive coding sequence may present challenges for viral based delivery approaches, monomeric GIY-YIG fusions to the TAL platform reduce the size and design complexity of TALENs by a factor of 2. At present it appears that TAL effectors may offer the most customizable and high fidelity option for specific binding (42), though the specificity and promiscuity of all platforms must be rigorously tested to determine the utility and relevance of each architecture.

While the benefits and limitations of existing genome editing technologies are analyzed, I have been motivated to consider the GIY-YIG nuclease domain as an alternative nuclease module. In Chapter 5, I demonstrate that the I-TevI nuclease domain can be fused to three targetable DNA-binding platforms to generate functional I-TevI-zinc-finger (Tev-ZFE), I-TevI-LAGLIDADG (Tev-LHE), or I-TevI-TAL effector (Tev-TAL) engineered nucleases (collectively abbreviated Tev-ENs) (Figure 6.3A). Significantly, the intrinsic properties of I-TevI are recapitulated within the context of Tev-ENs as they function catalytically as monomers and require an appropriately spaced CnnnG cleavage motif to generate a DSB. Fusions of the I-BmoI nuclease domain to the zinc-finger architecture (to create Bmo-ZFEs, Supplementary Figure S5.1) hydrolyzed substrates *in vitro* at the preferred cleavage site, however significant activity *in vivo* was not detected. The failure

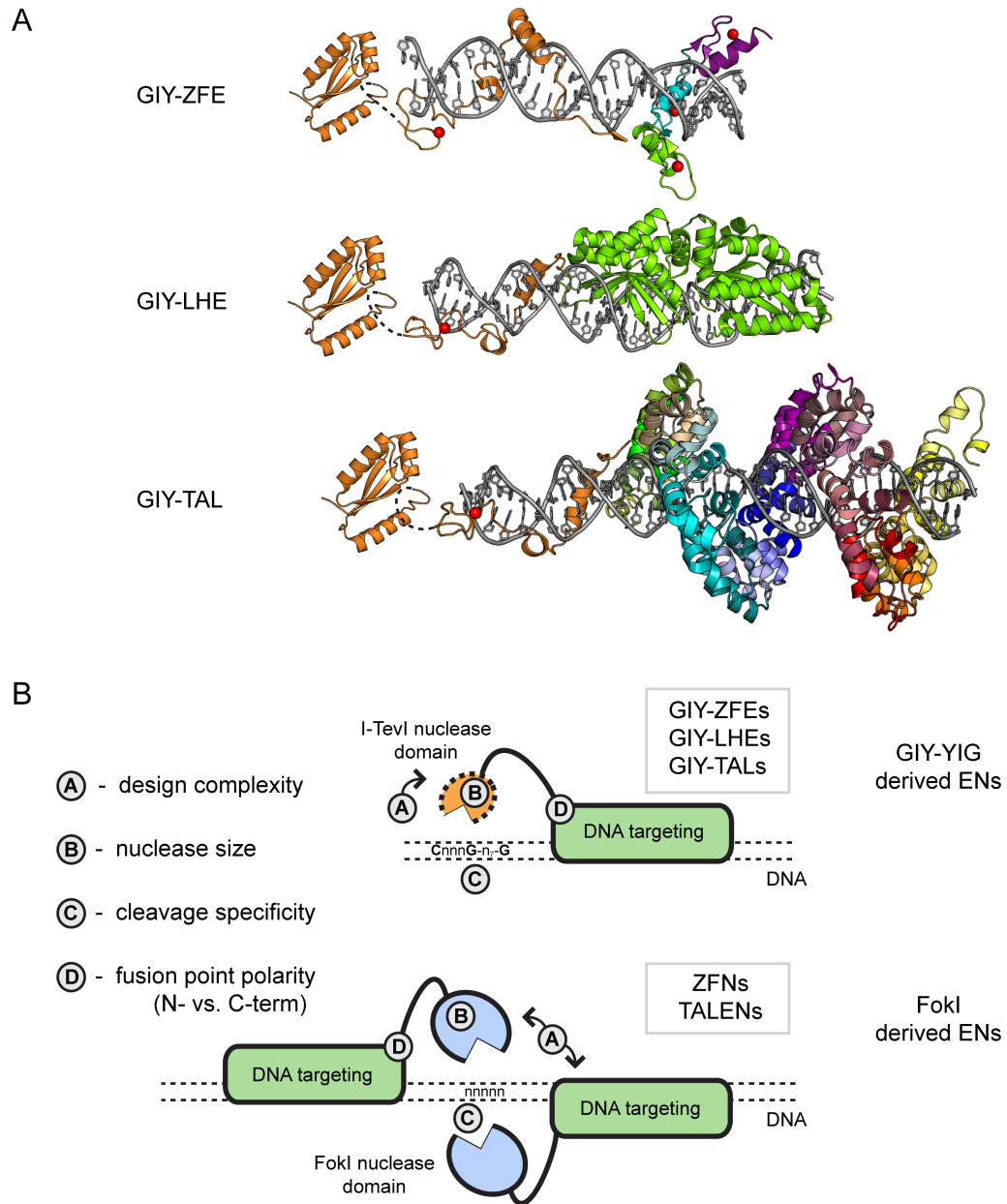


Figure 6.3: Design of engineered GIY-YIG endonucleases

(A) Models of the Tev-ZFE, Tev-LHE, and Tev-TAL platforms. (B) Schematic illustrating the fundamental differences between engineered GIY-YIG and FokI nuclease platforms.

of Bmo-ZFEs *in vivo* can likely be attributed to the lack of structural information required to rationally design functional fusions to the zinc-finger domain, as well as the difference in specific activity versus I-TevI (~750-fold lower) (14). Nevertheless, Tev-ZFEs and Tev-TALs induce gene conversion events in yeast on par with an established ZFN pair (Figures 5.3 & 5.5), and Tev-LHEs have activity in bacteria comparable to the engineered catalytically active LHE alone (Figure 5.4). Overall, the presented results are encouraging as GIY-ENs are functionally robust and mechanistically distinct when compared to analogous FokI-derived nucleases. However, there are properties of the I-TevI GIY-YIG domain and linker that must be considered for GIY-ENs as an alternative genome editing technology.

6.3 Implications of GIY-HEs for GIY-ENs

A cautionary lesson stemming from the implementation of FokI-containing ZFNs and TALENs is that one must understand the inherent biology of a nuclease domain prior its application in engineered architectures (31). Consequently, I envision the monomeric and sequence-tolerant GIY-YIG domain as an alternative to the FokI nuclease domain due to their structural and mechanistic differences (Figure 6.3B). One advantageous property of GIY-ENs is the decrease in design complexity by a factor of 2 versus FokI chimeras. Not only does this simplify the targeting of GIY-ENs to a single site, it also reduces the amount of fusion-encoding DNA that must be transfected into cells. This is an important consideration when utilizing the TALEN architecture, as the TAL domain is significantly larger than other DNA-binding platforms and a pair of nucleases must be transfected. Another advantageous property of GIY-ENs is the requirement to recognize a preferred nucleotide motif for cleavage (43,44). In scenarios where the DNA-binding module of a GIY-EN interacts with an unintended target site, the off-target site may be rendered immune to cleavage unless an appropriate cleavage motif is found within an appropriate linker/spacer distance (Supplementary Figure S5.4). Therefore, the monomeric mechanism and sequence preference by GIY-ENs offer advantages over ZFNs and TALENs by simplifying the assembly and delivery processes and potentially reducing off-target DSB frequencies.

One property observed for GIY-YIG family endonucleases by examining their natural functional diversity is the contextual ability to function as a nickase, as many GIY-YIG enzymes nick only a single strand of DNA (see also Chapter 1.2.2) (23,45,46). As described in Chapter 6.2, nickases represent a potentially safer nuclease mediated technology to induce desired gene conversion outcomes via HR. It may be possible to convert the GIY-YIG nuclease domain to a functional nickase through directed selection approaches or by altering the contextual modularity of the domain itself. For example, high-throughput selections and rational design methods have been utilized to convert both REs and LHEs into functional nickases (38-40,47-49). Conversely, additional DNA-binding elements could be fused upstream of the GIY-YIG domain to constrain repositioning of the nuclease motif after nicking (similar to the modular structure of GIY-YIG DNA-repair enzymes or the H-N-H HE I-HmuI) (28,50), or the nuclease domain could be fused to the C-terminal end of the LHE or TAL DNA-binding platforms to mimic the natural polarity of the LEM associated or Penelope nucleases (51,52).

6.4 Potential of GIY-YIG engineered nucleases

While progress has been made in assembling sequence-specific GIY-ENs that are distinct from FokI-derived chimeras, a number of properties must be further investigated for the technology to become widely applicable (Figure 6.4). Nucleotide requirements within the I-TevI cleavage motif and substrate spacer must be robustly investigated in both the LHE and TAL platforms to be able to predict which genomic sequences will be efficiently cleaved by Tev-ENs and whether additional targeting constraints exist. Preliminary data with Tev-TALs suggest that the substrate spacer preference at G+7 observed for I-BmoI may be recapitulated within the context of Tev-ENs (Figures 4.8 & 5.5). This result must be reexamined in a more robust manner, however, as this screen was limited due to the codon conservation of the thymidylate synthase genes. An added nucleotide preference for I-TevI in the substrate spacer would increase the nuclease domain's specificity from a CnnnG di-nucleotide to a tri-nucleotide motif. While this property may serve to further de-toxify GIY-ENs by conferring a more stringent cleavage preference, it may also limit the targeting capacity to fewer genomic sites. Accordingly, it may be important to identify and modify the residues that form protein:DNA contacts at G+7 for I-BmoI or an

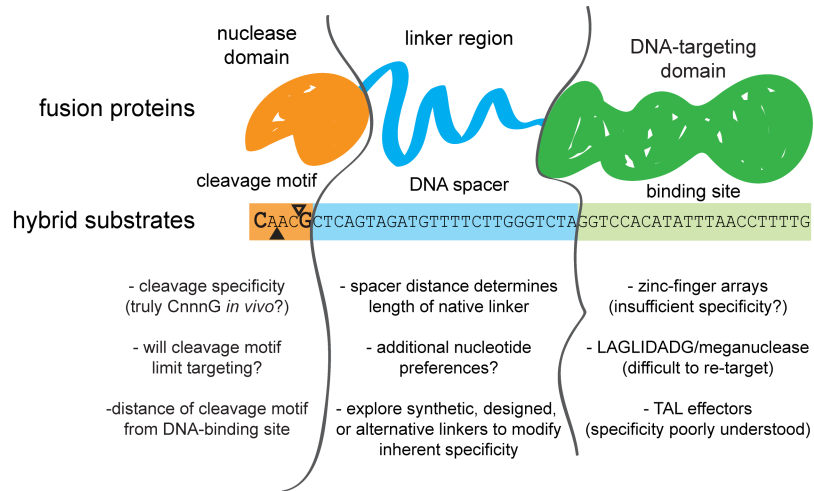


Figure 6.4: Considerations for engineered GIY-YIG nucleases

analogous position for Tev-ENs. If after further study the cleavage requirements of I-TevI limit the targeting potential of Tev-ENs, it would be prudent to investigate the nucleotide preferences of additional GIY-HE nuclease domains. Preliminary results indicate that a nuclease domain from a GIY-HE similar to I-TevI (from the bacteriophage Tu1a) has substantially different nucleotide preferences in the substrate spacer (*J. Wolfs, unpublished*). Over 150 putative GIY-HEs with potentially different targeting constraints have been identified from bacteriophage, bacterial, eukaryotic, and marine metagenomic sequences, many of which could be used to expand the diversity and specificity of GIY-ENs (13,17,53-55).

Another important consideration is functionality in relevant cellular systems. Genome editing interventions using engineered nucleases are most prominently implemented in model organisms to understand biological processes, or within plant and human cells for agriculture or experimental medicine. While it is encouraging that GIY-ENs fused to all three DNA-binding platforms are functional in bacteria and yeast, it will be necessary to ensure that this functionality is preserved within all relevant organisms. At this point it remains unknown if endogenous expression pathways would express GIY-ENs or whether their presence would elicit an innate immune response. If DSB efficiencies to similar FokI-derived architectures can be achieved in the relevant systems, a head-to-head comparison between the GIY-YIG and FokI platforms would be obligatory to compare on-target versus off-target cleavage rates. These studies would also elucidate whether the added specificity provided by GIY-ENs would detoxify existing technologies, or simply limit their targeting potential.

The data presented within this thesis towards the development of a novel monomeric genome-editing platform is encouraging, as GIY-ENs may serve as an alternative genome editing technology to begin to answer interesting biological questions and treat monogenic diseases. Ultimately, however, further refinements of each GIY-EN platform will dictate the utility, safety, and potential of GIY-ENs.

6.5 References

1. Sanmiguel,P. and Bennetzen,J.L. (1998) Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Annals of Botany*, **82**, 37-44.
2. Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W., et al. (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921.
3. Aziz,R.K., Breitbart,M. and Edwards,R.A. (2010) Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Res.*, **38**, 4207-4217.
4. Belfort,M., Derbyshire,V., Cousineau,B. and Lambowitz,A. (2002) Mobile introns: Pathways and proteins. In Craig,N., Craigie,R., Gellert,M. and Lambowitz,A. (eds.), *Mobile DNA II*. ASM Press, New York, pp. 761-783.
5. Stoddard,B.L. (2005) Homing endonuclease structure and function. *Q Rev Biophys*, **38**, 49-95.
6. Stoddard,B.L. (2011) Homing endonucleases: From microbial genetic invaders to reagents for targeted DNA modification. *Structure*, **19**, 7-15.
7. Taylor,G.K. and Stoddard,B.L. (2012) Structural, functional and evolutionary relationships between homing endonucleases and proteins from their host organisms. *Nucleic Acids Res.*, **40**, 5189-200.
8. Carter,J.M., Friedrich,N.C., Kleinstiver,B. and Edgell,D.R. (2007) Strand-specific contacts and divalent metal ion regulate double-strand break formation by the GIY-YIG homing endonuclease I-BmoI. *J Mol Biol*, **374**, 306-21.
9. Kleinstiver,B.P., Berube-Janzen,W., Fernandes,A.D. and Edgell,D.R. (2011) Divalent metal ion differentially regulates the sequential nicking reactions of the GIY-YIG homing endonuclease I-BmoI. *Plos One*, **6**, e23804.
10. Mueller,J.E., Smith,D., Bryk,M. and Belfort,M. (1995) Intron-encoded endonuclease I-TevI binds as a monomer to effect sequential cleavage via conformational changes in the *td* homing site. *EMBO J*, **14**, 5724-35.
11. Liu,Q., Derbyshire,V., Belfort,M. and Edgell,D.R. (2006) Distance determination by GIY-YIG intron endonucleases: Discrimination between repression and cleavage functions. *Nucleic Acids Res*, **34**, 1755-64.
12. Liu,Q., Dansereau,J.T., Puttamadappa,S.S., Shekhtman,A., Derbyshire,V. and Belfort,M. (2008) Role of the interdomain linker in distance determination for remote cleavage by homing endonuclease I-TevI. *J Mol Biol*, **379**, 1094-106.

13. Kleinstiver,B.P., Fernandes,A.D., Gloor,G.B. and Edgell,D.R. (2010) A unified genetic, computational and experimental framework identifies functionally relevant residues of the homing endonuclease I-BmoI. *Nucleic Acids Research*, **38**, 2411-27.
14. Edgell,D.R. and Shub,D.A. (2001) Related homing endonucleases I-BmoI and I-TevI use different strategies to cleave homologous recognition sites. *Proc Natl Acad Sci U S A*, **98**, 7898-903.
15. Van Roey,P., Meehan,L., Kowalski,J.C., Belfort,M. and Derbyshire,V. (2002) Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI. *Nat Struct Biol*, **9**, 806-11.
16. Noel,A.J., Wende,W. and Pingoud,A. (2004) DNA recognition by the homing endonuclease PI-SceI involves a divalent metal ion cofactor-induced conformational change. *J Biol Chem*, **279**, 6794-804.
17. Dunin-Horkawicz,S., Feder,M. and Bujnicki,J.M. (2006) Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics*, **7**, 98.
18. Mak,A.N., Lambert,A.R. and Stoddard,B.L. (2010) Folding, DNA recognition, and function of GIY-YIG endonucleases: Crystal structures of R.Eco29kI. *Structure*, **18**, 1321-31.
19. Gasiunas,G., Sasnauskas,G., Tamulaitis,G., Urbanke,C., Razaniene,D. and Siksnys,V. (2008) Tetrameric restriction enzymes: Expansion to the GIY-YIG nuclease family. *Nucleic Acids Research*, **36**, 938-49.
20. Ibryashkina,E.M., Sasnauskas,G., Solonin,A.S., Zakharova,M.V. and Siksnys,V. (2009) Oligomeric structure diversity within the GIY-YIG nuclease family. *Journal of Molecular Biology*, **387**, 10-6.
21. Lagerback,P., Andersson,E., Malmberg,C. and Carlson,K. (2009) Bacteriophage T4 endonuclease II, a promiscuous GIY-YIG nuclease, binds as a tetramer to two DNA substrates. *Nucleic Acids Research*, **37**, 6174-83.
22. Sokolowska,M., Czapinska,H. and Bochtler,M. (2010) Hpy188I-DNA pre- and post-cleavage complexes--snapshots of the GIY-YIG nuclease mediated catalysis. *Nucleic Acids Research*, **39**, 1554-64.
23. Verhoeven,E.E., van Kesteren,M., Moolenaar,G.F., Visse,R. and Goosen,N. (2000) Catalytic sites for 3' and 5' incision of escherichia coli nucleotide excision repair are both located in UvrC. *J Biol Chem*, **275**, 5120-3.
24. Kaliraman,V. and Brill,S.J. (2002) Role of SGS1 and SLX4 in maintaining rDNA structure in saccharomyces cerevisiae. *Curr Genet*, **41**, 389-400.

25. Flick,K.E., Jurica,M.S., J.,M.R.,Jr and Stoddard,B.L. (1998) DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature*, **394**, 96-101.
26. Van Roey,P., Waddling,C.A., Fox,K.M., Belfort,M. and Derbyshire,V. (2001) Intertwined structure of the DNA-binding domain of intron endonuclease I-TevI with its substrate. *EMBO J*, **20**, 3631-7.
27. Sitbon,E. and Pietrokovski,S. (2003) New types of conserved sequence domains in DNA-binding regions of homing endonucleases. *Trends Biochem Sci*, **28**, 473-7.
28. Shen,B.W., Landthaler,M., Shub,D.A. and Stoddard,B.L. (2004) DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J Mol Biol*, **342**, 43-56.
29. Anonymous. (2012) Method of the year 2011. *Nature Methods*, **9**, 1-1.
30. Holt,N., Wang,J., Kim,K., Friedman,G., Wang,X., Taupin,V., Crooks,G.M., Kohn,D.B., Gregory,P.D., Holmes,M.C., et al. (2010) Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. *Nat. Biotechnol.*, **28**, 839-U120.
31. Halford,S.E., Catto,L.E., Pernstich,C., Rusling,D.A. and Sanders,K.L. (2011) The reaction mechanism of fokI excludes the possibility of targeting zinc finger nucleases to unique DNA sites. *Biochem. Soc. Trans.*, **39**, 584-588.
32. Pattanayak,V., Ramirez,C.L., Joung,J.K. and Liu,D.R. (2011) Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nature Methods*, **8**, 765-U115.
33. Soellue,C., Pars,K., Cornu,T.I., Thibodeau-Beganny,S., Maeder,M.L., Joung,J.K., Heilbronn,R. and Cathomen,T. (2010) Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. *Nucleic Acids Res.*, **38**, 8269-8276.
34. Cathomen,T. and Joung,J.K. (2008) Zinc-finger nucleases: The next generation emerges. *Molecular Therapy*, **16**, 1200-1207.
35. McVey,M. and Lee,S.E. (2008) MMEJ repair of double-strand breaks (director's cut): Deleted sequences and alternative endings. *Trends in Genetics*, **24**, 529-538.
36. Lieber,M.R. (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.*, **79**, 181-211.
37. Certo,M.T., Gwiazda,K.S., Kuhar,R., Sather,B., Curinga,G., Mandt,T., Brault,M., Lambert,A.R., Baxter,S.K., Jacoby,K., et al. (2012) Coupling endonucleases with DNA end-processing enzymes to drive gene disruption. *Nature Methods*, **9**, 973-5.

38. Chan,S.H., Stoddard,B.L. and Xu,S.Y. (2011) Natural and engineered nicking endonucleases--from cleavage mechanism to engineering of strand-specificity. *Nucleic Acids Research*, **39**, 1-18.
39. McConnell Smith,A., Takeuchi,R., Pellenz,S., Davis,L., Maizels,N., Monnat,R.J.J. and Stoddard,B.L. (2009) Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, .
40. Ramirez,C.L., Certo,M.T., Mussolino,C., Goodwin,M.J., Cradick,T.J., McCaffrey,A.P., Cathomen,T., Scharenberg,A.M. and Joung,J.K. (2012) Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Res.*, **40**, 5560-5568.
41. Ramirez,C.L., Foley,J.E., Wright,D.A., Muller-Lerch,F., Rahman,S.H., Cornu,T.I., Winfrey,R.J., Sander,J.D., Fu,F., Townsend,J.A., et al. (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. *Nature Methods*, **5**, 374-375.
42. Joung,J.K. and Sander,J.D. (2013) INNOVATION TALENs: A widely applicable technology for targeted genome editing. *Nature Reviews Molecular Cell Biology*, **14**, 49-55.
43. Edgell,D.R., Stanger,M.J. and Belfort,M. (2003) Importance of a single base pair for discrimination between intron-containing and intronless alleles by endonuclease I-BmoI. *Curr Biol*, **13**, 973-8.
44. Edgell,D.R., Stanger,M.J. and Belfort,M. (2004) Coincidence of cleavage sites of intron endonuclease I-TevI and critical sequences of the host thymidylate synthase gene. *J Mol Biol*, **343**, 1231-1241.
45. Fricke,W.M. and Brill,S.J. (2003) Slx1-Slx4 is a second structure-specific endonuclease functionally redundant with Sgs1-Top3. *Genes Dev*, **17**, 1768-78.
46. Andersson,C.E., Lagerback,P. and Carlson,K. (2010) Structure of bacteriophage T4 endonuclease II mutant E118A, a tetrameric GIY-YIG enzyme. *Journal of Molecular Biology*, **397**, 1003-16.
47. Samuelson,J., Zhu,Z. and Xu,S. (2004) The isolation of strand-specific nicking endonucleases from a randomized sapI expression library. *Nucleic Acids Res.*, **32**, 3661-3671.
48. Heiter,D., Lunnen,K. and Wilson,G. (2005) Site-specific DNA-nicking mutants of the heterodimeric restriction endonuclease R.BbvCI. *J. Mol. Biol.*, **348**, 631-640.
49. Niu,Y., Tenney,K., Li,H. and Gimble,F.S. (2008) Engineering variants of the I-SceI homing endonuclease with strand-specific and site-specific DNA-nicking activity. *J. Mol. Biol.*, **382**, 188-202.

50. Truglio,J.J., Rhau,B., Croteau,D.L., Wang,L., Skorvaga,M., Karakas,E., Dellavecchia,M.J., Wang,H., Van Houten,B. and Kisker,C. (2005) Structural insights into the first incision reaction during nucleotide excision repair. *EMBO J*, **24**, 885-94.
51. Brachner,A., Braun,J., Ghodgaonkar,M., Castor,D., Zlopasa,L., Ehrlich,V., Jiricny,J., Gotzmann,J., Knasmueller,S. and Foisner,R. (2012) The endonuclease Ankle1 requires its LEM and GIY-YIG motifs for DNA cleavage in vivo. *J. Cell. Sci.*, **125**, 1048-1057.
52. Pyatkov,K.I., Arkhipova,I.R., Malkova,N.V., Finnegan,D.J. and Evgen'ev,M.B. (2004) Reverse transcriptase and endonuclease activities encoded by penelope-like retroelements. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 14719-24.
53. Sandegren,L. and Sjoberg,B.M. (2004) Distribution, sequence homology, and homing of group I introns among T-even-like bacteriophages - evidence for recent transfer of old introns. *J. Biol. Chem.*, **279**, 22218-27.
54. Bae,H., Chung,I., Sim,N. and Cho,Y. (2012) Complete genome sequence of pseudomonas aeruginosa siphophage MP1412. *J. Virol.*, **86**, 9537-9537.
55. Nord,D. and Sjöberg,B.M. (2008) Unconventional GIY-YIG homing endonuclease encoded in group I introns in closely related strains of the bacillus cereus group. *Nucleic Acids Res*, **36**, 300-10.

Appendices

Appendix S1: Copyright permissions for Chapters 2-5

S1.1 Permission for Chapter 2

FW: Permission to Use Copyrighted Material in a Doctoral Thesis Hide Details

FROM: [redacted] + Wednesday, January 23, 2013 7:47:40 AM

TO: bkleinst@uwo.ca ★

Dear Mr Kleinstiver,

Re: Benjamin P. Kleinstiver, Andrew D. Fernandes, Gregory B. Gloor, and David R. Edgell. A unified genetic, computational and experimental framework identifies functionally relevant residues of the homing endonuclease I-Bmol *Nucl. Acids Res.* (2010) 38(7): 2411-2427

Thank you for your recent email requesting permission to reuse all or part of your article in a new publication, a thesis or as part of your teaching.

As part of your copyright agreement with Oxford University Press you have retained the right, after publication, to use all or part of the article and abstract, in the preparation of derivative works, extension of the article into book-length or in other works, provided that a full acknowledgement is made to the original publication in the journal. As a result, you should not require direct permission from Oxford University Press to reuse your article.

However, if you are required by your new publisher or employer to obtain full written permission prior to reuse, please let us know and we will draw up a letter as soon as possible.

For full details of our publication and rights policy please see the attached link to our website:

http://www.oxfordjournals.org/access_purchase/publication_rights.html

If you have any other questions or queries, please feel free to contact us.

Kind regards,

[redacted]

[redacted]

Rights Assistant

Academic Rights & Journals

Tel: +44 [redacted]

Email: [redacted]@oup.com

S1.2 Permission for Chapter 3

RE: Permission to Use Copyrighted Material in a Doctoral Thesis Hide Details

FROM: PLoS + Tuesday, January 22, 2013 5:28:46 PM

TO: 'Benjamin Kleinstiver' ★

Dear Dr Kleinstiver,

Thank you for your enquiry and thanks for publishing with PLOS.

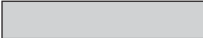
As you know we are an open access publisher and as such everything we publish is freely available online throughout the world, for you to read, download, copy, distribute, and use (with attribution) any way you wish.

No permission required.

For information on the Creative Commons Attribution License please follow this link <http://www.plos.org/about/open-access/license/>

Hope this is helpful.

Kind regards,


PLOS
22 January 2013

S1.3 Permission for Chapter 4

RE: FW: Permission to Use Copyrighted Material in a Doctoral Thesis Hide Details

FROM: +

TO: Benjamin Kleinstiver

Thursday, March 7, 2013 7:21:39 AM ★

Dear Mr Kleinstiver,

Kleinstiver, B.P., Wolfs, J. W., Edgell, D.R. (2013) The monomeric GIY-YIG homing endonuclease I-Bmol uses a molecular anchor and a flexible tether to sequentially nick DNA. *Nucleic Acids Research*, in press.

Thank you for your recent email requesting permission to reuse all or part of your article in a new publication, a thesis or as part of your teaching.

As part of your copyright agreement with Oxford University Press you have retained the right, after publication, to use all or part of the article and abstract, in the preparation of derivative works, extension of the article into book-length or in other works, provided that a full acknowledgement is made to the original publication in the journal. As a result, you should not require direct permission from Oxford University Press to reuse you article.

However, if you are required by your new publisher or employer to obtain full written permission prior to reuse, please let us know and we will draw up a letter as soon as possible.

For full details of our publication and rights policy please see the attached link to our website:
http://www.oxfordjournals.org/access_purchase/publication_rights.html

If you have any other questions or queries, please feel free to contact us.

Kind regards,

Rights Assistant

Academic Rights & Journals

Tel: +44

Email: @oup.com

S1.4 Permission for Chapter 5

RE: Permission to Use Copyrighted Material in a Doctoral Thesis Hide Details

FROM: PNAS Permissions + Thursday, February 7, 2013 12:46:08 PM

TO: 'Benjamin Kleinstiver' ★

Dear Ben Kleinstiver,

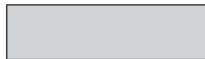
Authors need not obtain permission for the following uses of material they have published in PNAS: (1) to use their original figures or tables in their future works; (2) to make copies of their papers for their classroom teaching; or (3) to include their papers as part of their dissertations (or thesis).

Of course, citation to the original source should be included (full journal references).

Please feel free to contact us with any additional questions you might have.

Thank you!

Best regards,



Executive Editor
PNAS

Appendix S2: Supplementary information for Chapter 2

S2 Supplementary materials and methods

The WT, N12D, and I71N time-course data were fit via regularized nonlinear least-squares to the first-order kinetic model



which describes the stepwise conversion of DNA from circular to nicked to linear form. Regression curves (Figure S4) are shown overlaid with data and empirical monotonic Hermite smoothing-spline interpolants. These figures show no inconsistency between the first order kinetic model and the observed data for WT, N12D and I71N proteins.

In contrast, the first-order model above cannot account for the behavior of the I67N or S20Q mutants. Instead, values of k_1 were estimated in the absence of nicked or linear forms. For S20Q, circular DNA concentrations were fit to the empirical formula

$$C = C_0 + \exp(-\kappa_1 t + \kappa_2) + \kappa_3$$

where the comparable initial reaction velocity is given as $k_1 = \kappa_1 e^{\kappa_2}$. This empirical model could not reproduce the dynamic behavior of the I67N mutant due to the rapid initial conversion of C to N. Instead, the initial reaction velocity was estimated by the slope of the smoothing spline at time zero. The empirical differences in reaction kinetics between the I67N and S20Q mutants suggest that each has distinct mechanistic changes compared to each other, not just WT.

Since the total concentration of circular, nicked, and linear DNA is presumed constant, a ternary plot showing the relative proportion of each component is presented in Figure S4. Ternary-plot trajectories are displayed without regard to time, and the overall shapes of the reaction regression curves suggest mechanistic differences without regard to differences in reaction rate. Specifically, the plot suggests that the (WT, N12D, I71N), (I67N), and (S20Q) reaction curves are all mechanistically distinct since they group into three observably different classes.

S2 Supplementary figures

Nat ATGAAA TCAGGC GTT TAT AAAA TAACAAA TAAA AATACAGGGAAGTTTATATTGGCAGTTCA GAAGAC
CO ATGAAA TCTGGT GTT TAC AAAA TCACCAAC AAAA AACACCGGTAATTTCTACATCGGTTCTTCT GAAGAC
M K S G V Y K I T N K N T G K F Y I G S S E D

Nat TGT GAAAGTAGACTAAAAGTT CATTTTAGAAATCTA AAAAAC AATAGACATATTAATAGATATTTA AAC
CO TGGGAA TCTCGTCTGAAAGTT CACTTCCGTAACCTG AAAAAC AACCGTCACATCAACCGTTACCTGAAC
C E S R L K V H F R N L K N N R H I N R Y L N

Nat AATTCGTTTAAATAAA CATGGAGAGCAGGTATTTATTGGAGAGGTAATTCATATTTTGCCCATAGAAGAG
CO AACTCTTTCAACAAA CACGGTGAA CAGGTTTTCATCGGTGAAGTTATCCACATCCTGCCGATCGAAGAA
N S F N K H G E Q V F I G E V I H I L P I E E

Nat GCT ATAGCCAAGGAGCAA TGG TATATTGATAAT TTC TAT GAAGAAATGTACAAC ATAAGTAAA TCA GCT
CO GCT ATCGCTAAAGAACAGTGG TACATCGACAAC TTC TAC GAAGAAATGTACAAC ATCTCTAAA TCT GCT
A I A K E Q W Y I D N F Y E E M Y N I S K S A

Nat TAC CATGGCGGAGAC TTAACAAGCTATCAC CCA GACAAA CGA AACATCATCCTCAAAAAGAGCCGACAGT
CO TAC CACGGTGGT GAC CTGACCTCTTAC CAC CCG GACAAA CGT AACATCATCCTGAAA CGTGC T GACTCT
Y H G G D L T S Y H P D K R N I I L K R A D S

Nat TTGAAGAAAGTTTATTTGAAGATGACATCTGAAGAAAAGGCTAAGCGATGGCAATCTGTTCAAGGAGAA
CO CTGAAA AAAGTTTACCTGAAAATGACCTCTGAAGAAAAGGCTAACCGTGGCAGTGC GTTCAGGGTGAA
L K K V Y L K M T S E E K A K R W Q C V Q G E

Nat AATAAT CCGATGTTTGAAGAAA CATACA GAAACGACAAAAGCTAAAGATTTCAAATCATAATAAGCTT
CO AACAAAC CCGATGTTTCGGTCGTAAA CACACC GAAACCCACAAACTGAAAACTCTAACCACAACAAACTG
N N P M F G R K H T E T T K L K I S N H N K L

Nat TAT TACTCAACTCACAAA AATCCATTTAAAGGTAAAAAA CATAGTGAGGAGAGTAAGACTAAGTTGTCA
CO TACTACTCTACC CACAAA AACCCGTTCAAAGGTAAAAAA CACTCTGAAGAACTAAAACCAAACTGTCT
Y Y S T H K N P F K G K K H S E E S K T K L S

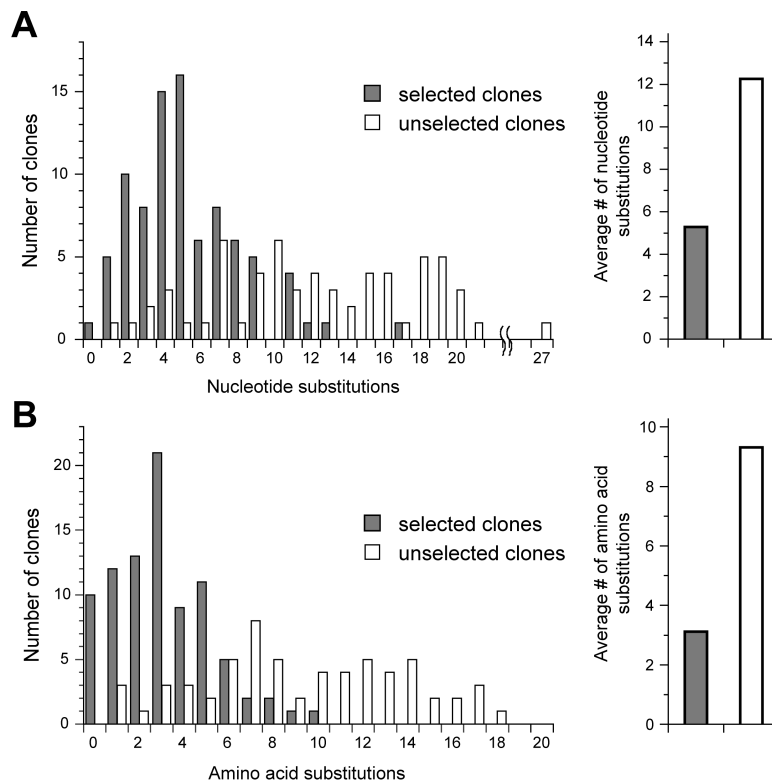
Nat GAA TATGCTTCTCAAAGAGTGGTGAAAAAATCCGTTC TATGGA AAAACA CACAGTGATGAATTTAAG
CO GAA TACGCTTCTCAGCGTGGTGGTGAAAAA AACCCGTTC TACGGTAAAACCCACTCTGACGAATTCAAA
E Y A S Q R V G E K N P F Y G K T H S D E F K

Nat ACTTATATGTCTAAA AAGTTTAAA GGCAGAAAGCCA AAAAATTC AAGACCA GTTATC ATAGATGGAACA
CO ACCTACATGTCTAAA AAATTC AAAAGGTCGTAAACCG AAAA AACTCTCGTCCG GTTATC ATCGACGGTACC
T Y M S K K F K G R K P K N S R P V I I D G T

Nat GAA TATGAAAGTGCTACA GAAGCTTCAAGA CAGTTAAATGTAGTT CCTGCTACTATC CTCCATAGAATC
CO GAA TACGAA TCTGCT ACCGAAGCTTCTCGT CAGCTGAACGTT GTT CCGGCT ACCATC CTGCACCGTATC
E Y E S A T E A S R Q L N V V P A T I L H R I

Nat AAA AGTAAA AATGAAAAATAC AGTGGATAC TTTTACAAA
CO AAA TCTAAA AACGAAAAATAC TCTGGTTAC TCTTACAAA
K S K N E K Y S G Y F Y K

Supplementary Figure S2.1: Comparison of the sequence of codon-optimized (CO) I-BmoI with the native sequence (Nat). Optimized codons for expression in *E. coli* are shown in red.



Supplementary Figure S2.2: Nucleotide mutation frequencies.

(A) Substitution frequencies from 87 selected and 62 unselected clones. (B) Predicted distributions based on simulations of the mutagenic method.

S2 Supplementary tables

Supplementary Table S2.2: Oligonucleotides used in Chapter 2

Name	Sequence (5'-3')	Notes
DE-331	CGCGCGCGCC <u>CATATG</u> AAATCTGGTGTTTACAAA ATC	NdeI site underlined
DE-384	CGCGCGCGCCTCGAGTTATTATTTGTAGAAGTA ACCAGAGTATTTTTTCG	XhoI site underlined
DE-395	<u>CTAGAAAGAGCCC</u> GTAGTAATGACATGGCCTTG <u>GGAAATCCCTTCAATGTATTCAGGCATG</u>	XbaI (5') and SphI (3') overhangs underlined
DE-396	<u>CCTGGAATACATTGAAGGGATTTCCCAAGGCCA</u> <u>TGTCATTACTACGGGCTCTTT</u>	XbaI (5') and SphI (3') overhangs underlined
DE-419	GAAGACTGCGAATCT <u>GCCCTG</u> AAAGTTCACTTC	Alanine codon underlined for R27A
DE-420	GAAGTGAAC <u>TTTCAGGGCAGATTCGCAGTCTTC</u>	Reverse complement of DE-419
DE-429	<u>CTAGAAACATAAGTGAGTAATGACATGGCCTTG</u> <u>GGAAATCCCTTCAATGTATTCAGGCATG</u>	XbaI (5') and SphI (3') overhangs underlined
DE-430	<u>CCTGGAATACATTGAAGGGATTTCCCAAGGCCA</u> <u>TGTCATTACTCACTTATGTTT</u>	XbaI (5') and SphI (3') overhangs underlined
DE-490	GTATAAGAAGGAGATATA <u>CATATG</u>	NdeI site underlined
DE-491	TTCTTTACCAGACT <u>CGAGTTATTA</u>	XhoI site underlined

Supplementary Table S2.3: EoS values and mutations for nuclease & linker residues

Position	Codon	Residue	log2_EoS	Sel. clones # sub.	Sel. clones #nonsynon. sub.	Unsel. clones # sub.	Unsel. clones #nonsynon. sub.
1	ATG	M	6.17	0	0	0	0
2	AAA	K	7.35	0	0	3	2
3	TCT	S	2.93	2	0	3	0
4	GGT	G	0.15	5	0	5	1
5	GTT	V	0.05	6	1	8	5
6	TAC	Y	5.77	0	0	1	1
7	AAA	K	7.21	0	0	6	4
8	ATC	I	5.75	1	0	6	6
9	ACC	T	2.91	0	0	3	2
10	AAC	N	5.84	1	0	2	2
11	AAA	K	3.49	2	1	3	3
12	AAC	N	3.26	10	10	8	7
13	ACC	T	-0.56	4	2	4	4
14	GGT	G	-0.50	2	1	4	1
15	AAA	K	1.52	2	2	4	3
16	TTC	F	2.19	1	1	4	4
17	TAC	Y	5.89	0	0	2	2
18	ATC	I	2.15	2	1	6	5
19	GGT	G	-0.60	2	1	1	0
20	TCT	S	3.00	1	0	4	3
21	TCT	S	0.26	4	1	7	3
22	GAA	E	1.22	1	1	1	1
23	GAC	D	3.22	0	0	3	3
24	TGC	C	0.51	2	1	2	2
25	GAA	E	-0.14	3	2	4	2
26	TCT	S	0.25	6	1	8	4
27	CGT	R	0.08	1	0	4	2
28	CTG	L	2.73	0	0	3	2
29	AAA	K	7.42	0	0	6	6
30	GTT	V	2.96	0	0	2	1
31	CAC	H	3.21	0	0	6	6
32	TTC	F	0.72	2	2	3	3
33	CGT	R	0.11	2	0	4	2
34	AAC	N	5.93	1	0	1	1
35	CTG	L	2.81	0	0	0	0
36	AAA	K	0.30	6	3	5	5
37	AAC	N	-0.75	4	4	4	4
38	AAC	N	2.12	2	1	3	3
39	CGT	R	0.22	5	0	5	2
40	CAC	H	0.33	1	1	3	3
41	ATC	I	2.28	1	1	4	3
42	AAC	N	5.89	0	0	6	6
43	CGT	R	0.31	1	0	3	0
44	TAC	Y	0.61	2	2	1	1
45	CTG	L	2.81	1	0	6	6
46	AAC	N	2.35	1	1	6	6
47	AAC	N	-0.34	3	3	5	5
48	TCT	S	3.12	2	0	3	1

49	TTC	F	-0.29	3	3	6	6
50	AAC	N	5.98	0	0	6	5
51	AAA	K	7.39	2	0	5	4
52	CAC	H	0.33	1	1	2	2
53	GGT	G	0.29	1	0	4	2
54	GAA	E	4.65	1	0	3	3
55	CAG	Q	-0.30	5	4	3	2
56	GTT	V	-0.40	6	2	6	2
57	TTC	F	6.00	0	0	5	5
58	ATC	I	-0.58	5	4	6	6
59	GGT	G	0.11	1	0	4	1
60	GAA	E	0.11	2	2	1	1
61	GTT	V	2.79	3	0	4	4
62	ATC	I	2.01	1	1	4	4
63	CAC	H	0.32	1	1	1	1
64	ATC	I	0.53	3	2	7	6
65	CTG	L	2.74	0	0	3	3
66	CCG	P	-0.47	2	1	1	1
67	ATC	I	5.67	0	0	2	2
68	GAA	E	-0.14	3	2	3	1
69	GAA	E	0.08	2	2	3	2
70	GCT	A	0.06	1	0	3	1
71	ATC	I	5.66	1	0	4	4
72	GCT	A	-0.57	2	1	1	0
73	AAA	K	1.45	5	2	6	5
74	GAA	E	4.69	1	0	2	1
75	CAG	Q	3.23	0	0	4	4
76	TGG	W	0.53	1	1	1	1
77	TAC	Y	2.27	1	1	2	2
78	ATC	I	2.13	1	1	5	5
79	GAC	D	3.18	0	0	3	2
80	AAC	N	5.87	0	0	2	2
81	TTC	F	0.67	2	2	5	5
82	TAC	Y	5.75	0	0	3	2
83	GAA	E	4.57	0	0	1	0
84	GAA	E	1.33	2	1	6	4
85	ATG	M	2.41	1	1	7	7
86	TAC	Y	5.98	0	0	5	5
87	AAC	N	5.93	1	0	6	6
88	ATC	I	2.13	1	1	5	5
89	TCT	S	2.82	0	0	4	4
90	AAA	K	3.52	1	1	2	2
91	TCT	S	2.95	3	0	3	2
92	GCT	A	0.19	0	0	5	1
93	TAC	Y	5.65	1	0	0	0
94	CAC	H	3.30	0	0	5	5
95	GGT	G	0.19	3	0	4	0
96	GGT	G	0.13	2	0	3	1
97	GAC	D	3.01	0	0	5	5
98	CTG	L	2.56	0	0	4	4
99	ACC	T	0.26	2	1	1	1
100	TCT	S	0.24	2	1	0	0
101	TAC	Y	2.18	1	1	6	6

102	CAC	H	3.26	0	0	7	6
103	CCG	P	0.17	1	0	0	0
104	GAC	D	3.23	0	0	2	2
105	AAA	K	7.54	1	0	7	6
106	CGT	R	0.17	0	0	5	1
107	AAC	N	-0.61	5	4	6	6
108	ATC	I	2.08	2	1	5	3
109	ATC	I	2.05	3	1	5	5
110	CTG	L	0.10	1	1	3	3
111	AAA	K	7.54	1	0	12	11
112	CGT	R	-0.48	4	1	3	1
113	GCT	A	0.06	2	0	3	1
114	GAC	D	3.30	0	0	0	0
115	TCT	S	2.88	0	0	3	1
116	CTG	L	2.74	0	0	5	5
117	AAA	K	0.34	4	3	14	11
118	AAA	K	3.47	3	1	5	4
119	GTT	V	0.20	2	1	8	4
120	TAC	Y	6.00	0	0	3	3
121	CTG	L	2.24	6	6	1	1
122	AAA	K	-0.69	6	5	12	11
123	ATG	M	2.54	1	1	5	5
124	ACC	T	0.19	1	1	3	3
125	TCT	S	0.09	2	1	0	0
126	GAA	E	-0.08	3	2	8	3
127	GAA	E	4.53	1	0	1	0
128	AAA	K	3.58	1	1	3	1
129	GCT	A	0.25	0	0	2	1
130	AAA	K	7.36	1	0	8	6
131	CGT	R	0.23	1	0	2	0
132	TGG	W	3.33	0	0	3	3
133	CAG	Q	3.21	0	0	2	2
134	TGC	C	0.34	1	1	4	4
135	GTT	V	0.22	5	1	8	2
136	CAG	Q	-0.45	2	2	5	5
137	GGT	G	-0.38	3	1	2	0
138	GAA	E	-0.01	3	2	3	2
139	AAC	N	5.84	0	0	6	6
140	AAC	N	5.74	0	0	1	1
141	CCG	P	0.00	0	0	2	1
142	ATG	M	2.57	1	1	6	6
143	TTC	F	0.59	2	2	7	6
144	GGT	G	0.33	1	0	1	1
145	CGT	R	-0.59	1	1	5	0
146	AAA	K	1.54	2	2	6	6
147	CAC	H	0.31	1	1	3	3
148	ACC	T	-0.70	3	2	3	2
149	GAA	E	-0.03	3	2	6	5
150	ACC	T	2.91	0	0	4	4
151	ACC	T	0.25	1	1	3	3
152	AAA	K	0.12	4	3	7	6
153	CTG	L	2.65	0	0	2	2
154	AAA	K	3.59	1	1	9	8

Supplementary Table S2.4: Rationale for selecting co-evolving residues to investigate

Information Content (Fig. 2.2A)	Co-evol. partner (Table 2.1)	Mutations in selected clones.	Mutations in unselected clones	EoS Score (Fig. 2.5 & Table S2.2)	I-TevI Residue	UvrC Residue	Mutation made in I-Bmol	Genetic selection Survival vs. WT (Table 2.2)
Y17	-	none	H ₂	5.89	Y17	Y29	F H	0 0
<p>Y17 is extremely well conserved in all GIY-YIG enzymes and lies in β-strand 2. The hydroxyl of I-TevI Y17 is surface exposed and points into the catalytic pocket. When modelled onto I-PpoI, Y17 is in a similar position to H98, which activates H₂O to act as a nucleophile. I-TevI Y17A is marginally soluble, and has 1% of WT activity. Y29 of UvrC forms a hydrogen bond to a metal-bound H₂O, and Y29A and Y29F mutants have ~1-2% activity. In UvrC, the Y29 hydroxyl protrudes from the active site surface and is in close proximity to divalent metal, where it is thought to lower the pK_a to serve as a general base (a proton shuttle). We made a Y17F mutation in I-Bmol as structurally inert mimic, while Y17H was made with the rationale that histidine mutant may retain activity if it can function as a base.</p>								
S20	I71	none	P ₂ , T	3	S20	K32	A Q	0 0.12
<p>S20, which is also found in β-strand 2, is a serine in some GIY-YIG members (including I-TevI) but is a strictly conserved lysine in UvrCs (K32). In the UvrC and I-TevI structures, K32 and S20 are solvent accessible and positioned in close proximity to the coordinated metal ion. In addition, UvrC K32 is near UvrC R39 (analogous to I-TevI R27), and thus may be involved in stabilization of the 5'P of the cleavage intermediate. The I-Bmol S20Q mutation was made since it was the most common substitution found among co-evolving S20-I71 pairs, and the S20A mutation was made to determine the requirement for polar residue at this position. In addition, this residue has not been mutated in any previous study.</p>								
H31	-	none	L ₂ , P, R ₂ , Q	3.21	H31	Y43	A F Y	0.003 0 0
<p>H31 is found in α-helix 1 and is conserved in most GIY-YIG-like endonucleases, and is either a Y or an H in all UvrC members. H31 of I-TevI is hydrogen-bonded to both Y6 and H40, and Y43 of UvrC is hydrogen bonded to Y19. In both I-TevI and UvrC, the residue is surface exposed in the catalytic pocket, and in the I-Bmol homology model, H31 is within hydrogen-bonding distance of Y17. The I-Bmol H31A mutation was made to determine whether H31 was essential, H31F was made as a structurally similar but chemically inert substitution to determine the role of the H31 side chain, and the H31Y mutation was generated in an attempt to mimic the proton shuttle if in fact the H31 side chain does have a role in catalysis.</p>								
N42	-	none	D, I, S ₂ , T, Y	5.89	S42	-	A D	0 0
<p>I-TevI has a serine at this position (S42), whereas no similar residues are present UvrCs at this position. This residue was of interest to mutate due to the fact that it was highly mutable in the unselected population, yet intolerant of mutation in the selected population, and had a high EoS score. In the alignment, slightly polar residues are prevalent at this position, and the I-Bmol homology model predicts that N42 is surface exposed. Interestingly, the amino group of N42 appears to be within hydrogen bonding distance of the hydroxyl of Y17. The N42A mutation was made to determine whether N42 is required for function, while the N42D mutation was made to specifically remove the positive nature of the asparagine side chain and to disrupt contacts to Y17.</p>								
S48	N10	none	A	3.12	S48	E57	A	1.02
<p>S48 lies in a 9 amino acid stretch (from L45 to G53) that shows a high degree of conservation (formerly partly characterized as motif C). Along with K51 and H52, this stretch of amino acids comprises most of α-helix 2. There is a hydrogen bond network between the side chains of N10, S48, and Y86 (with N10 and Y86 being well conserved). S48 coevolves with N10, and while the interaction may be structural, it was our goal to investigate the function of interaction. The S48A mutation was made not only because it is a potentially deleterious substitution, but also because there is a tendency for this position to be alanine.</p>								

Information Content (Fig. 2.2A)	Co-evol. partner (Table 2.1)	Mutations in selected clones.	Mutations in unselected clones	EoS Score (Fig. 2.5 & Table S2.2)	I-TevI Residue	UvrC Residue	Mutation made in I-Bmol	Genetic selection Survival vs. WT (Table 2.2)
K51	H52	none	N, V, X ₂	7.39	K51	-	L	0.86
K51 is a well-conserved residue that lies in α -helix 2, and may form important structural contacts. K51 lies in a stretch of conserved residues that comprise motif C in previous alignments; S48-G53 is conserved between I-TevI and I-Bmol, while UvrC lacks K51-F56 (all of motif C). K51 has a very high EoS score, and K51 co-evolves with H52 (K51-H52 is the third highest scoring co-evolving pair). The K51L mutation removes the charges functional groups from the side chain, but also was the second most common amino acid at the position from the alignment.								
H52	K51	R	I,2	0.33	H52	-	R	0.43
This residue is also a part of α -helix 2, where it is either a histidine or tyrosine. The rationale for mutation of this position was similar to that for K51. Interestingly, we isolated a clone from the unigenic evolution screen that had a single mutation at this position (H52R). The H52Y mutation was made because tyrosine was the second most common residue at that position in the alignment, and tyrosine would be structurally similar to histidine.								
I67	-	none	N, V	5.67	Y66	N68	N	0.06
I67 is found in an unstructured region between β -strand 3 and α -helix 3, and was interesting to investigate from a structural standpoint. It is one of 4 isoleucines that are solvent exposed and flank the catalytic pocket (two on either side of the cleft). Similar to the proposed role for I71, I67 may be involved in contacting substrate. The I67N mutation was generated to substitute the inert nature of an isoleucine residue with the charge of an asparagine residue, while maintaining a similar size.								
I71	S20	none	F ₄	5.66	I72	F73	A	0.22
I71 is found in α -helix 3 and co-evolves with S20, which is the highest-scoring co-evolving pair. I-TevI also contains an isoleucine at this position and UvrC's have a highly conserved phenylalanine. The analogous Leu 116 of I-PpoI is in the metal binding α -helix and is inserted into the minor groove of substrate, forming Van der Waals contacts with adenine. A L116A mutant in I-PpoI has reduced catalytic activity, and a DNA-binding defect. The I71N mutation was generated because asparagine was the second most common amino acid at this position among 146 sequences, and the I71A mutation was generated to determine the effect of the loss of the large hydrophobic sidechain.								
Y86	-	none	C ₂ , D, F ₂	5.98	Y89	Y87	F	0.99
Y86 is highly conserved in all GIY-YIG enzymes and is found next to the functionally critical N87 in a loop region between α -helix 3 and the linker region. This residue was mutated because of its high degree of conservation, and because it forms a hydrogen bond network with N10 and S48. The Y86F mutation was generated to determine the affect of removing the polar hydroxyl group.								

Appendix S3: Supplementary information for Chapter 3

S3 Supplementary tables

Supplementary Table S3.1: Strains and plasmids used in this study

Strains	Description	Source
DH5 α	F ⁻ , ϕ 80 <i>dlacZ</i> Δ M15, Δ (<i>lacZYA-argF</i>)U169, <i>deoR</i> , <i>recA1</i> , <i>endA1</i> , <i>hsdR17</i> (rk ⁻ , mk ⁺), <i>phoA</i> , <i>supE44</i> , λ ⁻ , <i>thi-1</i> , <i>gyrA96</i> , <i>relA1</i>	Invitrogen
ER2566	F- λ - <i>fhuA2</i> [lon] <i>ompT lacZ::T7 gene 1 gal sulA11</i> Δ (<i>mcrC-mrr</i>)114::IS10 R(<i>mcr-73::miniTn10-TetS</i>)2 R(<i>zgb-210::Tn10</i>)(TetS) <i>endA1</i> [dcm]	N.E.B.
Plasmids	Description	Source
pTYBmoI	pTYB1 derivative containing the 266 amino acid codon-optimized I-BmoI gene	Ref 1.
pBmoHS	pBS derivative containing a 48-base pair XbaI/BamHI insert corresponding to the intronless <i>B. mojavensis thyA</i> gene	Ref 2.
pDE212	Similar to pBmoHS, with the intron-containing sequence upstream of the intron insertion site (substitutions relevant to this study are: T-1G, G-2T, C-3G, C-4A, C-5A, and G-6T)	Ref 2.
pDE213	Similar to pBmoHS, with a T-1G substitution	Ref 2.
pDE214	Similar to pBmoHS, with a G-2T substitution	Ref 2.
pDE215	Similar to pBmoHS, with a C-3G substitution	Ref 2.
pDE216	Similar to pBmoHS, with a C-4A substitution	Ref 2.
pDE217	Similar to pBmoHS, with a C-5A substitution	Ref 2.
pDE218	Similar to pBmoHS, with a G-6T substitution	Ref 2.
pDE219	Similar to pBmoHS, with T-1G and G-2T substitutions	Ref 2.
pDE220	Similar to pBmoHS, with G-2T and C-3G substitutions	Ref 2.
pDE221	Similar to pBmoHS, with G-2T and C-4A substitutions	Ref 2.
pDE222	Similar to pBmoHS, with G-2T and C-5A substitutions	Ref 2.
pDE223	Similar to pBmoHS, with C-3G and C-4A substitutions	Ref 2.
pDE224	Similar to pBmoHS, with C-3G and C-5A substitutions	Ref 2.
pDE225	Similar to pBmoHS, with C-4A and C-5A substitutions	Ref 2.
pDE227	Similar to pBmoHS, with T-1G, G-2T, and C-3G substitutions	Ref 2.
pDE228	Similar to pBmoHS, with T-1G, C-3G, and C-5A substitutions	Ref 2.
pDE229	Similar to pBmoHS, with G-2T, C-3G, and C-4A substitutions	Ref 2.
pDE230	Similar to pBmoHS, with G-2T, C-4A, and G-6T substitutions	Ref 2.
pDE231	Similar to pBmoHS, with C-3G, C-4A, and C-5A substitutions	Ref 2.
pDE232	Similar to pBmoHS, with C-4A, C-5A, and G-6T substitutions	Ref 2.
pDE233	Similar to pBmoHS, with G-2T, C-3G, C-4A, and C-5A substitutions	Ref 2.

1. Kleinstiver, B.P., Fernandes, A.D., Gloor, G.B. and Edgell, D.R. (2010) A unified genetic, computational and experimental framework identifies functionally relevant residues of the homing endonuclease I-BmoI. *Nucleic Acids Res*, **38**, 2411-2427.
2. Edgell, D.R., Stanger, M.J. and Belfort, M. (2003) Importance of a single base pair for discrimination between intron-containing and intronless alleles by endonuclease I-BmoI. *Curr Biol*, **13**, 973-978.

Supplementary Table S3.2: Oligonucleotides used in Chapter 3

Name	Sequence (5'-3')	Notes
DE-37	CCACCTTGAGGTAAGAGCCCGTAGTAATGACAT GGCCTTGGGAAATCCCTTCAATGTATTCCAG	Top strand intronless <i>thyA</i> 64mer from +22 to -42 relative to the intron insertion site
DE-38	CTGGAATACATTGAAGGGATTTCCCAAGGCCAT GTCATTACTACGGGCTCTTACCTCAAGGTGG	Bottom strand intronless <i>thyA</i> 64mer from +22 to -42 relative to the intron insertion site
DE-116	AAACTCCACCTTGAGGTAAGAGCCCGTAGTAA TGACATGGCCTTGGGAAATCCCTTCAATGTATT CCAGTACAA	Top strand intronless <i>thyA</i> 74mer from +27 to -47 relative to the intron insertion site
DE-117	TTGTACTGGAATACATTGAAGGGATTTCCCAAG GCCATGTCATTACTACGGGCTCTTACCTCAAGG TGGAGTTT	Bottom strand intronless <i>thyA</i> 74mer from +27 to -47 relative to the intron insertion site
DE-444	AGTAGCGACTTCTACTGAACATAAGTGAGTAAT GACATGGCCTTGGGAAATCCCTTCAATGTATTC CAGTACAA	Top strand intron-containing <i>thyA</i> 74mer from +27 to -47 relative to the intron insertion site
DE-445	TTGTACTGGAATACATTGAAGGGATTTCCCAAG GCCATGTCATTACTCACTTATGTTTCAGTAGAAG TCGCTACT	Bottom strand intron-containing <i>thyA</i> 74mer from +27 to -47 relative to the intron insertion site
DE-446	AAACTCCACCTTGAGGTAAGAGCCCTTAGTAAT GACATGGCCTTGGGAAATCCCTTCAATGTATTC CAGTACAA	Top strand intronless <i>thyA</i> 74mer from +27 to -47 relative to the intron insertion site, with a <u>G</u> : <u>2T</u> substitution
DE-447	TTGTACTGGAATACATTGAAGGGATTTCCCAAG GCCATGTCATTACTA <u>A</u> GGGCTCTTACCTCAAGG TGGAGTTT	Bottom strand intronless <i>thyA</i> 74mer from +27 to -47 relative to the intron insertion site, with a <u>C-2A</u> substitution
DE-459	AAACTCCACCTTGAGGTAAGAGCC <u>A</u> TAGTAAT GACATGGCCTTGGGAAATCCCTTCAATGTATTC CAGTACAA	Top strand intronless <i>thyA</i> 74mer from +27 to -47 relative to the intron insertion site, with a <u>G</u> : <u>2A</u> substitution
DE-460	TTGTACTGGAATACATTGAAGGGATTTCCCAAG GCCATGTCATTACTA <u>T</u> GGGCTCTTACCTCAAGG TGGAGTTT	Bottom strand intronless <i>thyA</i> 74mer from +27 to -47 relative to the intron insertion site, with a <u>C-2T</u> substitution
DE-461	AAACTCCACCTTGAGGTAAGAGCC <u>C</u> TAGTAAT GACATGGCCTTGGGAAATCCCTTCAATGTATTC CAGTACAA	Top strand intronless <i>thyA</i> 74mer from +27 to -47 relative to the intron insertion site, with a <u>G</u> : <u>2C</u> substitution
DE-462	TTGTACTGGAATACATTGAAGGGATTTCCCAAG GCCATGTCATTACTA <u>G</u> GGGCTCTTACCTCAAGG TGGAGTTT	Bottom strand intronless <i>thyA</i> 74mer from +27 to -47 relative to the intron insertion site, with a <u>C-2G</u> substitution
DE-463	AGTAGCGACTTCTACTGAACATAAGGGAGTAAT GACATGGCCTTGGGAAATCCCTTCAATGTATTC CAGTACAA	Top strand intron-containing <i>thyA</i> 74mer from +27 to -47 relative to the insertion site, with a A-2G substitution
DE-464	TTGTACTGGAATACATTGAAGGGATTTCCCAAG GCCATGTCATTACT <u>C</u> CCTTATGTTTCAGTAGAAGT CGCTACT	Bottom strand intron-containing <i>thyA</i> 74mer from +27 to -47 relative to the insertion site, with a T-2C substitution

Supplementary Table S3.3: Confidence intervals for rate constants determined on mutant substrates in low and high MgCl₂ concentrations.

	2mM MgCl ₂										10mM MgCl ₂														
	Substrate					k_1 (s ⁻¹) ^a					k_2 (s ⁻¹) ^b					k_1 (s ⁻¹)					k_2 (s ⁻¹)				
	-6	-5	-4	-3	-2	-1	low	median	high	low	high	low	median	high	low	high	low	median	high	low	high	low	median	high	low
intronless	G	C	C	G	T	0.0531	0.0574	0.0628	0.0236	0.0281	0.0762	0.111	0.131	0.068	0.0738	0.0759									
Class I	A	0.0461	0.0501	0.0645	0.0207	0.0279	0.0561	0.0922	0.1044	0.1178	0.0537	0.0571	0.0591												
(like wild-type)	T	0.0504	0.0572	0.061	0.0239	0.0241	0.0244	0.0763	0.0872	0.0937	0.0556	0.0575	0.0586												
Class II	G	0.0266	0.0308	0.0315	0.0095	0.0095	0.0101	0.0534	0.0651	0.0755	0.0245	0.0245	0.0255												
(rescue)	G	0.01	0.0102	0.0104	0.0083	0.0084	0.0085	0.0243	0.0246	0.0263	0.0247	0.0249	0.0257												
A	0.0095	0.0095	0.0095	0.015	0.0155	0.0156	0.0282	0.0309	0.0312	0.0409	0.0421	0.0428													
A G	0.008	0.0082	0.0086	0.0088	0.0091	0.0094	0.0243	0.0244	0.0245	0.0218	0.0221	0.0225													
A A	0.0042	0.0044	0.005	0.015	0.0153	0.0156	0.0128	0.0136	0.0143	0.0411	0.0434	0.0448													
T A	0.0103	0.0105	0.0106	0.0164	0.017	0.019	0.0269	0.0277	0.0287	0.0467	0.0484	0.0502													
T A A	0.0043	0.0045	0.0046	0.0146	0.0154	0.0155	0.0129	0.0135	0.0143	0.0453	0.0468	0.0486													
Class III	T	0.00025	0.00027	0.00028	n.d.	n.d.	n.d.	0.00059	0.00061	0.00063	n.d.	n.d.	n.d.												
(no rescue)	T G	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.												
G T	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.												
A T	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.												
A T	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.												
A G	0.00035	0.00037	0.0004	0.00026	0.00053	0.0015	0.00085	0.0013	0.00145	0	0.0138	0.0148													
G T G	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.												
A G G	0.00008	0.00016	0.00025	n.d.	n.d.	n.d.	0.00032	0.00035	0.00036	0.0022	0.0026	0.0028													
A G T	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.												
T A	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.												
A A G	0.00018	0.0002	0.00021	0.00008	0.00169	0.00271	0.00039	0.00052	0.0007	0.0087	0.0091	0.0106													
A A G T	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.												
A A G T G	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.												

^a k_1 , the mean of the 95% confidence interval for the rate constant that describes the first nicking reaction, which generates nicked intermediate

^b k_2 , the mean of the 95% confidence interval for the rate constant that describes the second nicking reaction, which generates linear product
n.d., not determined

Appendix S4: Supplementary information for Chapter 4

S4 Supplementary materials and methods

DNA-binding affinity of I-BmoI domains. Binding reactions shown in Supplementary Figure S4.2 contained 200 mM NaCl, 50 mM tris pH 8.0, 5% glycerol, either 10 mM EDTA or 10 mM MgCl₂, 6 μM *thyA* 74-mer substrate, and 12 μM full-length I-BmoI or domain truncation in 15 μl volumes. Reactions were incubated at room temperature for 15 minutes, stopped by the addition of 5 μl loading buffer (10% glycerol, 100mM EDTA), and half of the reaction was loaded on a 10% native polyacrylamide (19:1) gel. Gels were stained in ethidium bromide (Caledon) prior to analysis on an AlphaImagerTM3400 (Alpha Innotech).

Cross-linking reactions. Prior to cross-linking reactions, purified I-BmoI was dialyzed into 50 mM HEPES pH 8.0, 500 mM NaCl, and 5% glycerol. Typically, 20-μl reactions contained 7.6 μM I-BmoI, 15 mM MgCl₂, and 1.125 μM *thyA* 74mer substrate (DE-116/DE-117) when included. Reactions were supplemented with dilutions of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide) (EDC, Fluka Analytical) from 144 μM to 48 mM in conjunction with *N*-Hydroxysuccinimide (NHS, Sigma-Aldrich Inc.) from 92 μM to 31 mM. I-BmoI was the final component to be added to the reactions that were subsequently stopped by adding 12 μl of stop buffer (792 mM Tris-HCl pH 8.0, 4 % SDS, 20 % glycerol) at either 2, 5, or 10 minutes. Stopped reactions were run against a medium molecular weight ladder (PiNK plus, GeneDirex) and proteins were identified after electrophoresis on 15% SDS-PAGE gels either by coomassie stain (86 pmol I-BmoI per lane) or Western blot (3.6 pmol per lane, see below). Additional cross-linking reactions were conducted with a range of I-BmoI concentrations and were conditionally supplemented with 15 mM MgCl₂, *thyA* 74-mer substrate (DE-116/DE-117), and various concentrations of EDC/NHS, bisulfosuccinimidyl suberate (BS3, Thermo Scientific), or glutaraldehyde (EM grade, Polysciences Inc.).

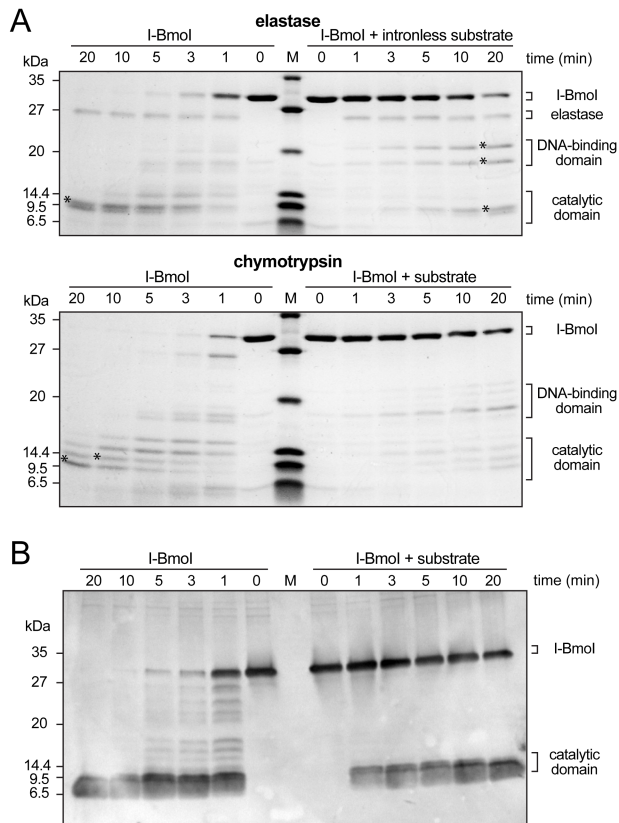
For Western blot analysis, a polyclonal anti-N111 I-BmoI antibody (Pacific Immunology) was raised against the N111 truncation of I-BmoI. Electrophoresis of 3.5 pmol of I-BmoI cross-linking reactions on a 15% SDS-PAGE gel was conducted prior to

transfer to a PVDF membrane (Pall Life Sciences) at 100V. Membranes were washed on a rocker in casein blocking buffer (Sigma) for 60 minutes, washed with a 1:10,000 dilution of the anti-N111 antibody in TBST (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% Tween 20) for 60 minutes, washed 4 x 5 minutes in TBST, exposed to a 1:10,000 dilution of anti-rabbit IgG peroxidase conjugate (Sigma) for 60 minutes in TBST, and washed 4 x 5 minutes in TBST. Luminescent peptides were detected using ECL Plus (GE Healthcare) detection reagent and were visualized at 450 nm on a StormTM860 (Molecular Dynamics).

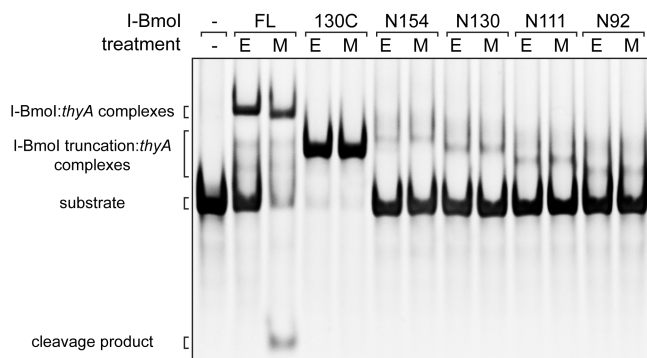
Substrate topology assays. Cleavage assays were performed similar to those described in the Methods section (Chapter 4.2), on supercoiled pLBmo, pLBmo linearized with *Swa*I, and the relaxed plasmids pLBmoNa, pLBmoNb, and pLBmo2N that were pre-nicked by *Nt.Bbv*CI (N.E.B.).

Cleavage assays with immobilized substrates. The immobilized substrate consisted of duplex oligonucleotides, where the 5' end of the top strand contained a biotin moiety (DE-746), and the bottom strand (DE-747) was 5' labeled with γ^{32} -P prior to annealing. Duplex oligonucleotides (0.32 pmol) were incubated with 10 μ g streptavidin coated magnetic beads (binding capacity of 1 nmol/mg Promega) to achieve ~3% occupancy. The beads were subsequently washed in triplicate in 0.5x SSC (75 mM NaCl, 7.5 mM Na₃Citrate) prior to excess I-BmoI (175 nM) binding in 50 mM Tris-HCl pH 8.0, 50 mM NaCl for 5 minutes at 37°C. Protein-substrate complexes were washed in triplicate in wash buffer (20 mM Tris-HCl pH 8.0, 250 mM NaCl, 5% glycerol, 10 mM EDTA) prior to the addition cleavage buffer (20 mM Tris-HCl pH 8.0, 250 mM NaCl, 5% glycerol, 10 mM MgCl₂). Cleavage by I-BmoI was determined by monitoring the release of radiolabel into the supernatant using a scintillation counter (Beckman Coulter CS5801).

S4 Supplementary figures

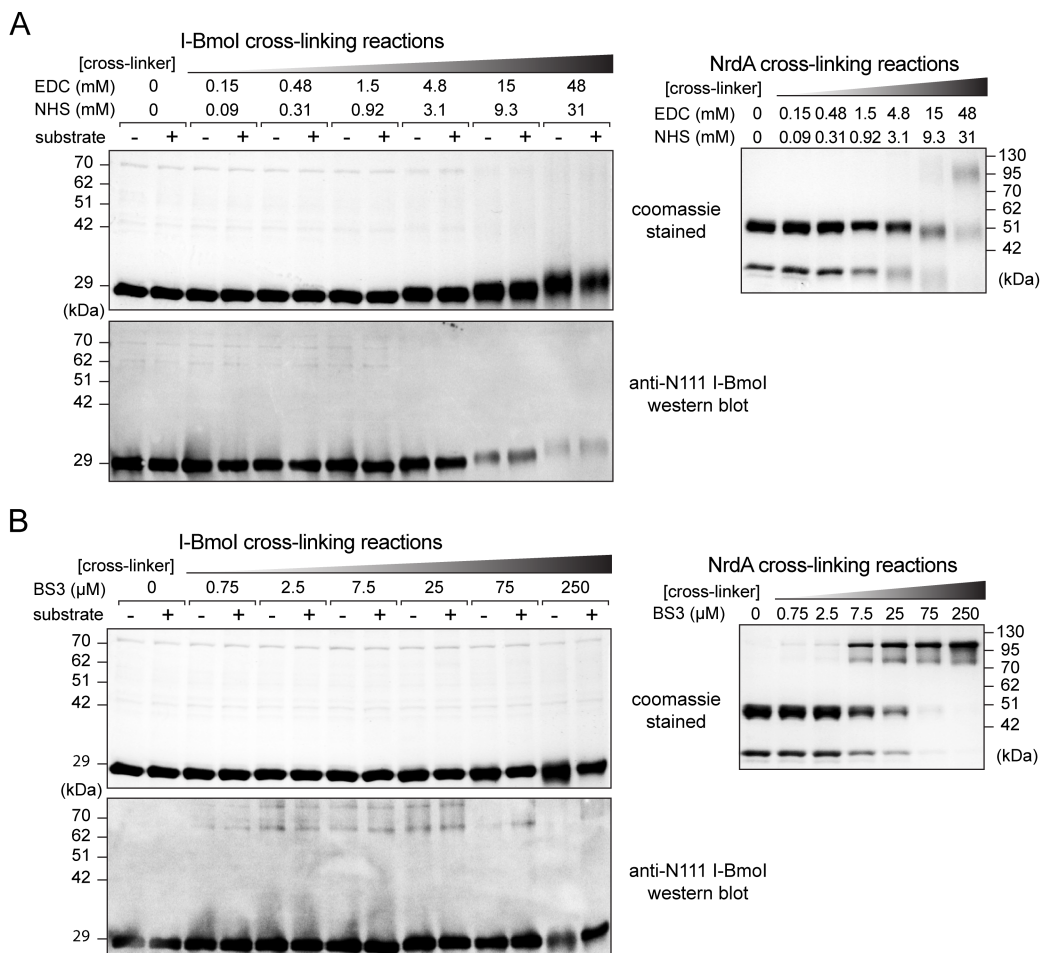
**Supplementary Figure S4.1: Time-course proteolysis of I-Bmol.**

(A) Images of coomassie stained SDS-gels of elastase and chymotrypsin limited proteolysis time-course experiments performed on I-Bmol and I-Bmol pre-incubated with intronless *thyA* substrate. Aliquots were removed at the indicated time points. In-gel trypsin digest followed by mass spectrometry was performed to identify the peptide products indicated with an asterisk (*). M, protein marker (sizes in kDa indicated to the left). (B) Western blot image of trypsin limited proteolysis time-course experiment. The western blot was performed with an antibody against N111 I-Bmol on a gel with aliquots of similar reactions to those shown in Figure 4.1A.



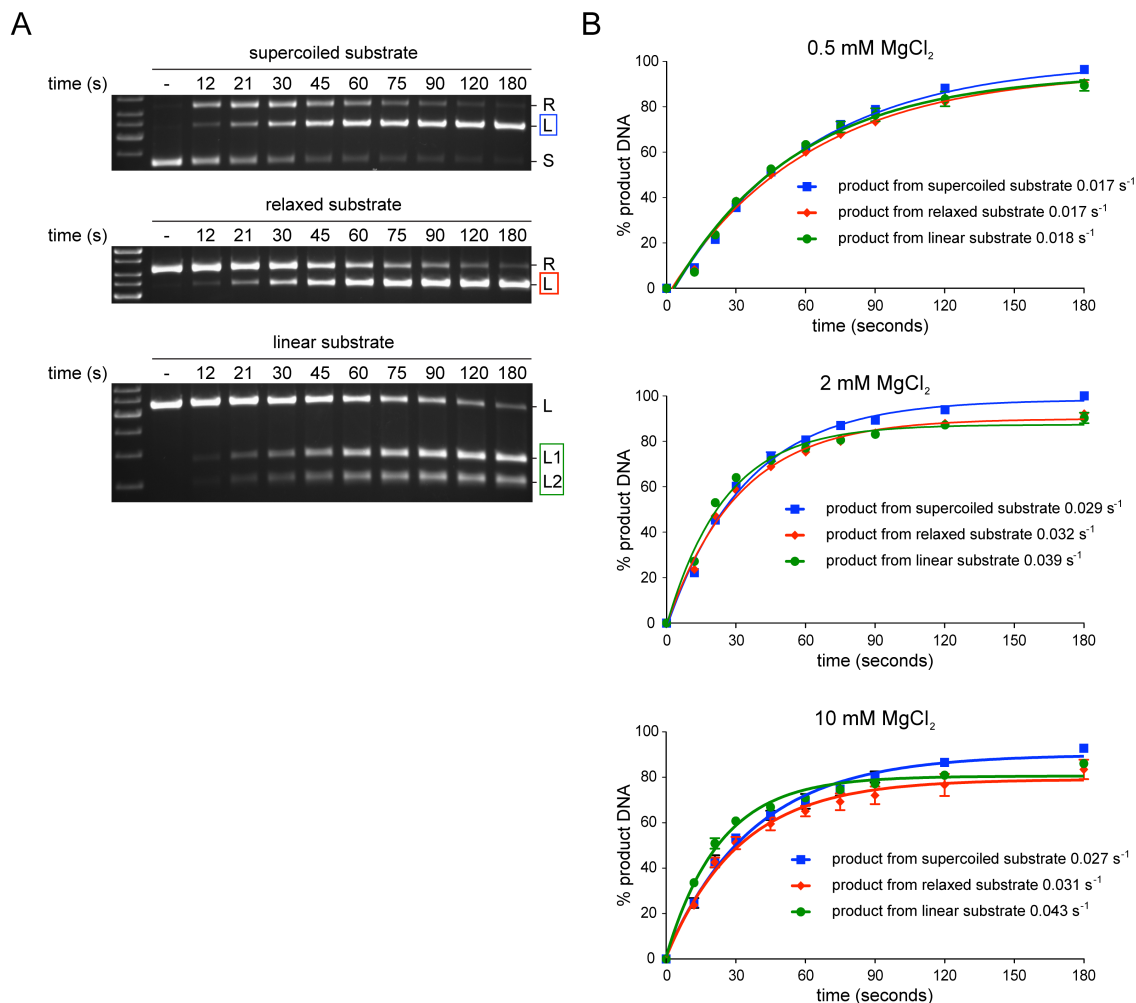
Supplementary Figure S4.2: DNA-binding activity of full-length I-BmolI and purified domains.

Image of a native polyacrylamide gel with binding reactions containing full length and truncated I-BmolI:*thyA* complexes, stained with ethidium-bromide. Binding reactions contained either 10 mM EDTA (E) or MgCl₂ (M).



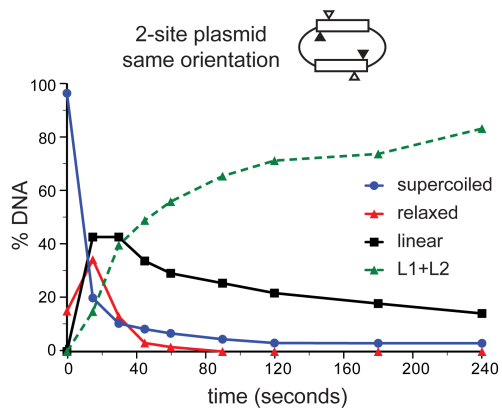
Supplementary Figure S4.3: Crosslinking of I-Bmol in solution and in complex with substrate.

(*top panels*) Representative SDS-gel images of cross-linking reactions of I-Bmol or I-Bmol:substrate complexes using (A) NHS and EDC, or (B) BS³. (*bottom panels*) Western blots using an anti-N111 I-Bmol antibody performed to detect cross-linked complexes below the detection limit of coomassie stain. (*right panels*) Positive control cross-linking reactions performed at equivalent protein:cross-linker ratios with the Aeh1 NrdA subunits that are known to associate (60). Molecular weight markers are shown beside the gels, in kDa.



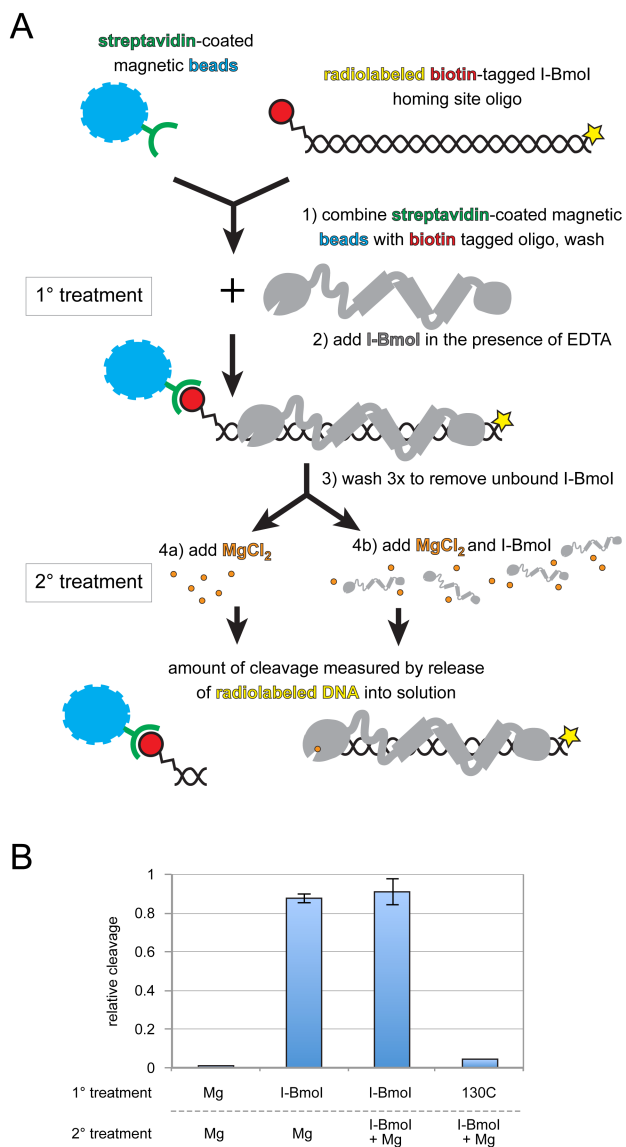
Supplementary Figure S4.4: Substrate topology does not affect I-BmoI cleavage kinetics:

(A) Representative gel images of *in vitro* time-course cleavage assays in 0.5 mM MgCl₂ on supercoiled, relaxed, or linear substrate (*top*, *middle*, and *bottom* panels, respectively). The left-most lane contains a DNA size marker, and the second lane from the left contains unreacted substrate (-). Relaxed (R), linear (L), supercoiled (S), and two linear product (L1, L2) DNA forms are indicated on the right, with product species for each reaction boxed and colored according to reaction progress curves in (B). (B) Plots of product formation for cleavage assays with 0.5 mM, 2 mM, and 10 mM MgCl₂ (*top*, *middle*, and *bottom panels*, respectively) on supercoiled, relaxed, or linear substrate (blue box, red diamond, and green circle, respectively). The average and standard deviation of three replicates is shown.



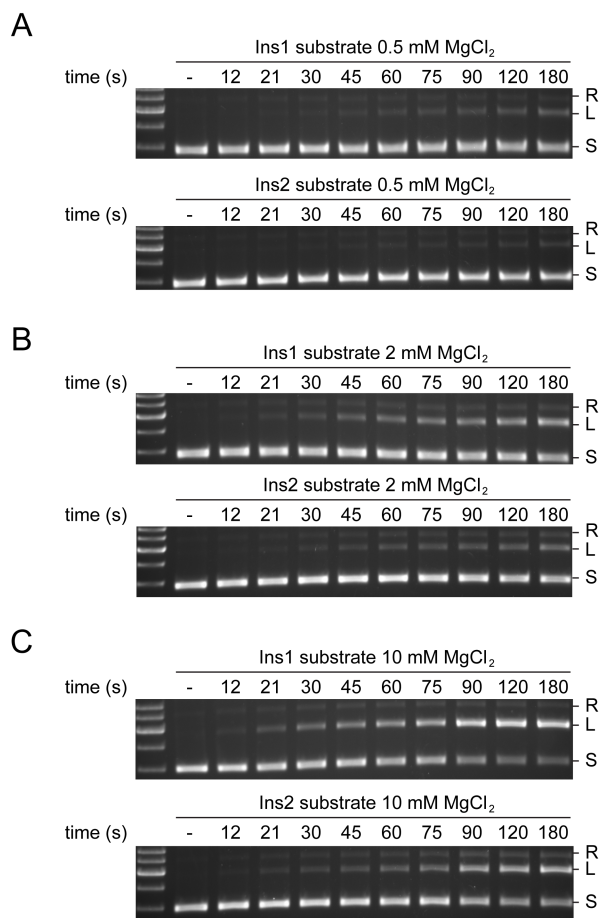
Supplementary Figure S4.5: I-BmoI cleavage of a double target-site plasmid.

Plot of reaction progress for *in vitro* cleavage assays on the two-site substrate plasmid with the target sites in the same orientation. Cleavage assays were performed in triplicate. L1+L2, linear products from 2-site plasmid cleavage.



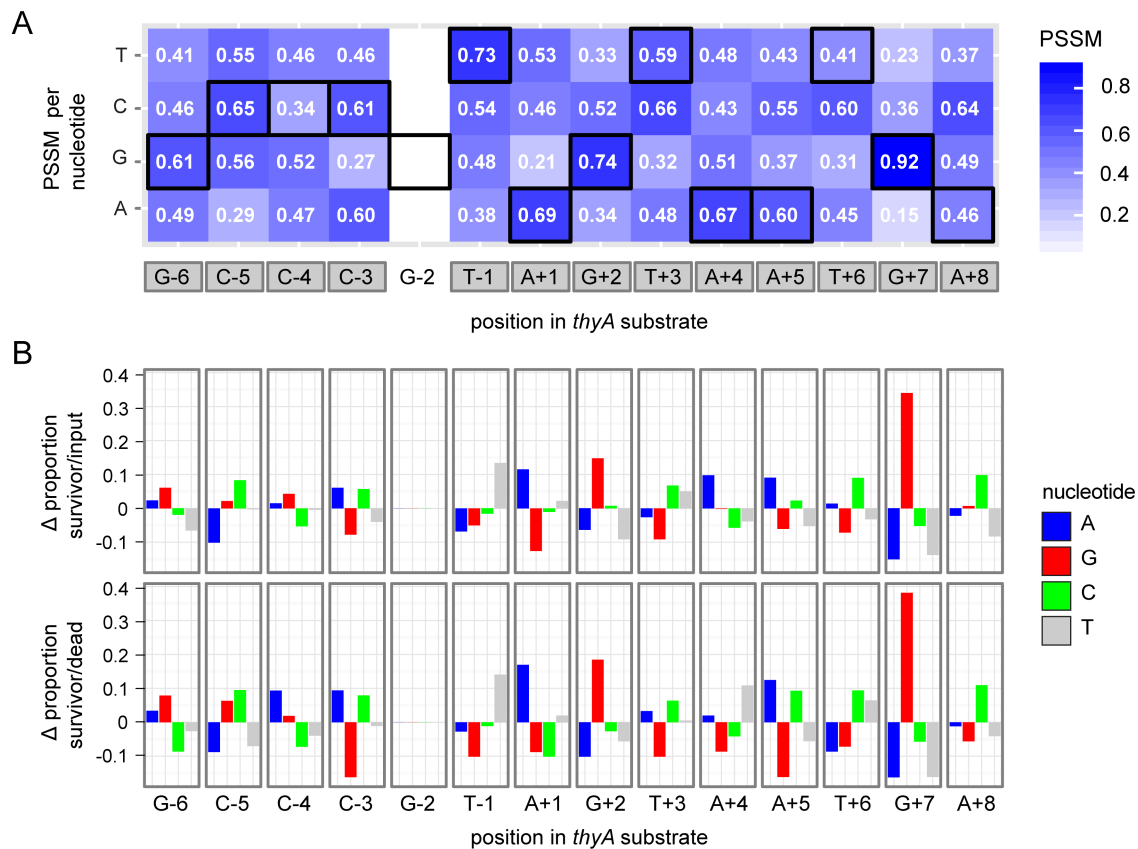
Supplementary Figure S4.6: Immobilized substrates are cleaved by I-BmoI.

Schematic of the assembly of immobilized I-BmoI:*thyA* substrate complexes using streptavidin coated magnetic beads. Cleavage was detected by measuring the release of radiolabel into solution. (B) Graph of cleavage activity between various primary and secondary treatments, reported as cleavage normalized relative to EcoRI (an EcoRI site is found between the biotin tag and the I-BmoI cleavage site). Secondary treatments containing I-BmoI were performed with excess I-BmoI. At least two replicates were performed for the trials where standard error is shown, otherwise only one replicate.



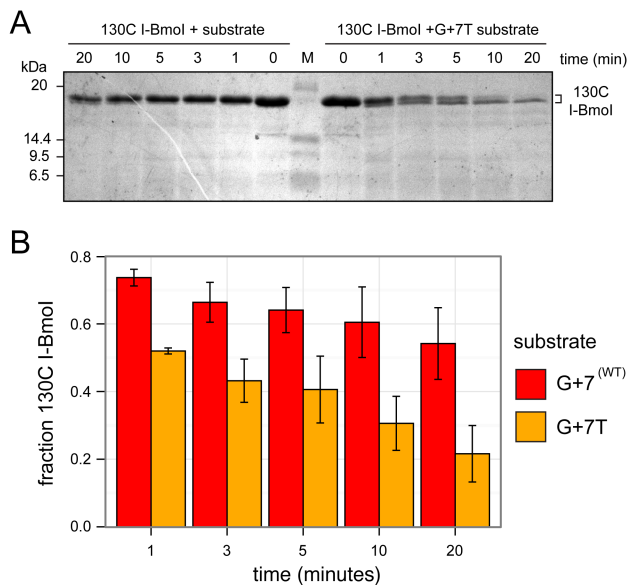
Supplementary Figure S4.7: Gel images of I-BmoI cleavage assays on supercoiled substrates with nucleotide insertions.

Cleavage reactions contained (A) 0.5 mM, (B) 2 mM, or (C) 10 mM MgCl₂, and substrates described in Figure 7A. The left lane contains DNA size marker, and the second lane from the left contains unreacted substrate (-). Relaxed (R), linear (L), and supercoiled (S) DNA forms are shown to the right.



Supplementary Figure S4.8: Frequency of nucleotides in the survivor pool.

(A) Sequencing results from cleavable targets represented as a position-specific scoring matrix. (B) Plots to compare the change in nucleotide proportion in the randomized target site for survivor versus input and survivor versus dead plasmid pools (*top* and *bottom* panels, respectively).



Supplementary Figure S4.9: Limited proteolysis of 130C I-Bmol.

(A) Coomassie stained SDS-gel of trypsin proteolysis time-course experiments performed with 130C I-Bmol:substrate and 130C:G+7T substrate complexes. Aliquots were removed at the indicated time points prior to electrophoresis. M, protein marker (sizes in kDa indicated to the left). (B) Plot of the fraction of 130C I-Bmol remaining over time for reactions shown in panel (A). Disappearance of 130C I-Bmol is calculated as the fraction remaining after normalization to the untreated lane, with error bars and standard deviation of two replicates.

S4 Supplementary tables

SupplementaryTable S4.1: Strains and plasmids used in Chapter 4

Strains	Description	Source
DH5 α	F ⁻ , ϕ 80 <i>lacZ</i> Δ M15, Δ (<i>lacZYA-argF</i>)U169, <i>deoR</i> , <i>recA1</i> , <i>endA1</i> , <i>hsdR17</i> (rk ⁻ , mk ⁺), <i>phoA</i> , <i>supE44</i> , λ^- , <i>thi-1</i> , <i>gyrA96</i> , <i>relA1</i>	Invitrogen
ER2566	F- λ - <i>fhuA2</i> [lon] <i>ompT lacZ::T7</i> gene 1 <i>gal sulA11</i> Δ (<i>mcrC-mrr</i>)114::IS10 R(<i>mcr-73::miniTn10-TetS</i>)2 R(<i>zgb-210::Tn10</i>)(TetS) <i>endA1</i> [dcm]	N.E.B.
BW25141 (λ DE3)	F ⁻ <i>lacI^f</i> <i>rrnB_{T14}</i> <i>DlacZ_{WJ16}</i> <i>DphoBR580</i> <i>hsdR514</i> <i>DaraBAD_{AH33}</i> <i>DrhaBAD_{LD78}</i> <i>galU95</i> <i>endA_{BT333}</i> <i>uidA</i> (<i>Dmlu1</i>): <i>pir+</i> <i>recA1</i> , λ DE3 lysogen	Ref 28.
Plasmids	Description	Source
pTYB1	<i>ori_{M13}</i> , amp	N.E.B.
pTYBmoI	pTYB1 derivative containing the 266 amino acid codon-optimized I-BmoI gene	Ref 28.
pTYBmoIR27A	Similar to pTYBmoI, with an R27A mutation to knockout catalytic activity	Ref 34.
pTYBmoN92	pTYB1 derivative containing residues 1-92 of codon-optimized I-BmoI (cloned NdeI/XhoI)	This study
pTYBmoN111	pTYB1 derivative containing residues 1-111 of codon-optimized I-BmoI (cloned NdeI/XhoI)	This study
pTYBmoN130	pTYB1 derivative containing residues 1-130 of codon-optimized I-BmoI (cloned NdeI/XhoI)	This study
pTYBmoN154	pTYB1 derivative containing residues 1-154 of codon-optimized I-BmoI (cloned NdeI/XhoI)	This study
pTYBmo92C	pTYB1 derivative containing residues 92-266 of codon-optimized I-BmoI (cloned NdeI/XhoI)	This study
pTYBmo106C	pTYB1 derivative containing residues 106-266 of codon-optimized I-BmoI (cloned NdeI/XhoI)	This study
pTYBmo130C	pTYB1 derivative containing residues 130-266 of codon-optimized I-BmoI (cloned NdeI/XhoI)	This study
pBmoHS	pBS derivative containing a 54-base pair XbaI/BamHI insert corresponding to the intronless <i>B. mojavensis thyA</i> gene	Ref 31.
LITMUS28i	<i>ori_{M13}</i> , amp	N.E.B.
pLBmo	LITMUS28i derivative, with the PvuII fragment of pBmoHS (containing the 54 bp I-BmoI <i>thyA</i> target site) subcloned into PvuII	This study
pLBmo2a	pLBmo derivative containing a second I-BmoI target site cloned into SmaI using oligos DE-906/907 (sites in the opposite orientation)	This study
pLBmo2b	Similar to pLBmo2a, except that the I-BmoI target sites are in the same orientation	This study
pLBmoNa	pLBmo derivative that contains an Nt.BbvCI nicking site 1082 base pairs away from the I-BmoI top-strand cleavage site.	This study
pLBmoNb	pLBmo derivative that contains an Nt.BbvCI nicking site 60 base pairs away from the I-BmoI top-strand cleavage site.	This study
pLBmo2N	pLBmo derivative that contains Nt.BbvCI nicking sites 1082 and 60 base pairs away from the I-BmoI top-strand cleavage site.	This study
p11-lacY-wtx1	<i>ori_{pBR322}</i> , amp	Ref 40.
pTox	p11-lacY-wtx1 derivative with EcoRI site destroyed and new EcoRI site between the XbaI and SphI sites using oligos DE-1173/ 1174	This study

pToxBmoHS	p11-lacY-wtx1 plus 52bp intronless <i>thyA</i> homing site insert in XbaI and SphI sites	Ref 28.
pToxBmoIn+	p11-lacY-wtx1 plus 52bp intron-containing homing site insert in XbaI and SphI sites	Ref 28.
pKox	pTox derivative that has the kanamycin promoter and gene blunt cloned into the ScaI site	This study
pKoxBmoHS	pKox derivative that contains a 52bp intronless <i>thyA</i> homing site insert in XbaI and SphI sites	This study
pKoxBmoIn+	pKox derivative that contains a 52bp intron-containing homing site insert in XbaI and SphI sites	This study
pToxBmoG+7A	pTox derivative that has the I-BmoI target site with a G+7A mutation, cloned into the XbaI and EcoRI sites	This study
pToxBmoG+7C	pTox derivative that has the I-BmoI target site with a G+7C mutation, cloned into the XbaI and EcoRI sites	This study
pToxBmoG+7T	pTox derivative that has the I-BmoI target site with a G+7T mutation, cloned into the XbaI and EcoRI sites	This study
pLBmoG+7A	LITMUS28i derivative that has the I-BmoI target site with a G+7A mutation, cloned into the XbaI and EcoRI sites	This study
pLBmoG+7C	LITMUS28i derivative that has the I-BmoI target site with a G+7C mutation, cloned into the XbaI and EcoRI sites	This study
pLBmoG+7T	LITMUS28i derivative that has the I-BmoI target site with a G+7T mutation, cloned into the XbaI and EcoRI sites	This study
pLBmo+5T	LITMUS28i derivative that has the I-BmoI target site with a TATTT insertion between base pairs +8 and +9, cloned into XbaI/EcoRI	This study
pLBmo+5C	LITMUS28i derivative that has the I-BmoI target site with a CGCCC insertion between base pairs +8 and +9, cloned into XbaI/EcoRI	This study
pLBmo+3Ta	LITMUS28i derivative that has the I-BmoI target site with a TTT insertion between base pairs +8 and +9, cloned into XbaI/EcoRI	This study
pLBmo+3Ca	LITMUS28i derivative that has the I-BmoI target site with a CCC insertion between base pairs +8 and +9, cloned into XbaI/EcoRI	This study
pLBmo+3Tb	LITMUS28i derivative that has the I-BmoI target site with a TTT insertion between base pairs +4 and +5, cloned into XbaI/EcoRI	This study
pLBmo+3Cb	LITMUS28i derivative that has the I-BmoI target site with a CCC insertion between base pairs +4 and +5, cloned into XbaI/EcoRI	This study

Supplementary Table S4.2: Oligonucleotides used in Chapter 4

Name	Sequence (5'-3')	Notes
DE-116	AAACTCCACCTTGAGGTAAGAGCCCG TAGTAATGACATGGCCTTGGGAAATC CCTTCAATGTATTCCAGTACAA	Top strand intronless I-BmoI <i>thyA</i> target 74-mer from +27 to -47 relative to the intron insertion site
DE-117	TTGTACTGGAATACATTGAAGGGATT TCCCAAGGCCATGTCATTACTACGGG CTCTTACCTCAAGGTGGAGTTT	Bottom strand intronless I-BmoI <i>thyA</i> target 74-mer from +27 to -47 relative to the intron insertion site
DE-123	TTGTACTGGAATACATTGAAGGGATT TCCCAAGGCCATGTCATTACTACGG	Bottom strand intronless <i>thyA</i> 51-mer from -4 to -47, relative to the intron insertion site
DE-124	TTGTACTGGAATACATTGAAGGGATT TCCCAAGGCCATGTCATTACTACG	Bottom strand intronless <i>thyA</i> 50-mer from -3 to -47, relative to the intron insertion site
DE-125	GCTCTTACCTCAAGGTGGAGTTT	Bottom strand intronless <i>thyA</i> 23-mer from +27 to +5, relative to intron insertion site
DE-130	GAGCCCGTAGTAATGACATGGCCTTG GGAAATCCCTTCAATGTATT	Top strand intronless I-BmoI <i>thyA</i> 46- mer from +8 to -38 relative to the intron insertion site
DE-131	AATACATTGAAGGGATTTCCCAAGGC CATGTCATTACTACGGGCTC	Bottom strand intronless I-BmoI <i>thyA</i> 46-mer from +8 to -38 relative to the intron insertion site
DE-144	AATTGTGAATTTAATTATCCGTGACG AAGCCATCACGGCGTATATGTCGGAT TGTTAGCTCAGGAGATT	Top strand non-specific 69-mer
DE-144 GC	AATCTCCTGAGCTAACAATCCGACAT ATACGCCGTGATGGCTTCGTCACGGA TAATTAATTCACAATT	Bottom strand non-specific 69-mer
DE-331	CGCGCGGCCATATGAAATCTGGTGT TTACAAAATC	<u>NdeI</u> site; for N-terminal cloning of I- BmoI & truncations into pTYB1
DE-444	AGTAGCGACTTCTACTGAACATAAGT GAGTAATGACATGGCCTTGGGAAATC CCTTCAATGTATTCCAGTACAA	Top strand intron-containing I-BmoI target 74-mer from +27 to -47 relative to the intron insertion site
DE-445	TTGTACTGGAATACATTGAAGGGATT TCCCAAGGCCATGTCATTACTCACTT ATGTTCAAGTAGAAGTCGCTACT	Bottom strand intron-containing I- BmoI target 74-mer from +27 to -47 relative to the intron insertion site
DE-498	CGCGGGGTACCCTTGGCAAAGCATT TGTAGAAGTAACCAGAGTATTTTCG	<u>KpnI</u> site; for C-terminal cloning of I- BmoI & truncations into pTYB1
DE-680	CGCGGGTACCCTTGGCAAAGCAAGCA GATTTAGAGATGTTGTACATTTCTTCG	<u>KpnI</u> site; for C-terminal cloning of N92 I-BmoI into pTYB1 to create pTYBmoN92
DE-681	CGCGGGTACCCTTGGCAAAGCATTTC AGGATGATGTTACGTTTGTCCGGG	<u>KpnI</u> site; for C-terminal cloning of N111 I-BmoI into pTYB1 to create pTYBmoN111

DE-682	CGCGGGTACCCTTGGCAAAGCATTTA GCTTTTTCTTCAGAGGTCATTTTCAGG	<u>KpnI</u> site; for C-terminal cloning of N130 I-BmoI into pTYB1 to create pTYBmoN130
DE-683	CGCGGGTACCCTTGGCAAAGCATTTTC AGTTTGGTGGTTTCGGTGTGTTTACGA CCG	<u>KpnI</u> site; for C-terminal cloning of N154 I-BmoI into pTYB1 to create pTYBmoN154
DE-684	CGCGCCATATGGCTTACCACGGTGGT GACCTGACCTC	<u>NdeI</u> site; for N-terminal cloning of 92C I-BmoI into pTYB1 to create pTYBmo92C
DE-685	CGCGCCATATGCGTAACATCATCCTG AAACGTGCTG	<u>NdeI</u> site; for N-terminal cloning of 106C I-BmoI into pTYB1 to create pTYBmo106C
DE-686	CGCGCCATATGAAACGTTGGCAGTGC GTTACAGGGTG	<u>NdeI</u> site; for N-terminal cloning of 130C I-BmoI into pTYB1 to create pTYBmo130C
DE-746	(biotin)GCGAATTCGTAAGAGCCCGTA GTAATGACATGGCCTTGGGAAATCCC TTCAATGTATTCCA	Top strand (5'-biotin) modified intronless <i>thyA</i> I-BmoI homing site from -12 to +41 with <u>EcoRI</u> site
DE-747	TGGAATACATTGAAGGGATTTCCTAA GGCCATGTCATTACTACGGGCTCTTA CGAATTCGC	Bottom strand intronless <i>thyA</i> I-BmoI homing site from -12 to +41 with <u>EcoRI</u> site
DE-906	AGAGCCCCTAGTAATGACATGGCCTT GGGAAATCCCTTCAATGTATTCC	Top strand intronless I-BmoI <i>thyA</i> target 49-mer from +9 to -40
DE-907	GGAATACATTGAAGGGATTTCCTAAAG GCCATGTCATTACTACGGGCTCT	Bottom strand intronless I-BmoI <i>thyA</i> target 49-mer from +9 to -40
DE-945	CCGAAGGTAACCTGGCCTCAGCAGAG CGCAGATAACC	Top strand Quikchange primer to add an <u>Nt.BbvCI</u> site to pLBmo to make pLBmo2a
DE-946	GGTATCTGCGCTCTGCTGAGGCCAGT TACCTTCGG	Bottom strand Quikchange primer to add an <u>Nt.BbvCI</u> site to pLBmo to make pLBmo2a
DE-956	GCTCGGAATTAACCCTCAGCAAAGG GAACAAAAGCTTGC	Top strand Quikchange primer to add an <u>Nt.BbvCI</u> site to pLBmo to make pLBmo2b
DE-957	GCAAGCTTTTGTTCCTTTGCTGAGG GTTAATTCCGAGC	Bottom strand Quikchange primer to add an <u>Nt.BbvCI</u> site to pLBmo to make pLBmo2b
DE-1171	CACTATAGGGAGACCGGATTTCTGG CACGACAGG	Top strand Quikchange primer to remove the <u>EcoRI</u> site from p11-lacY-wtx1
DE-1172	CCTGTCGTGCCAGGAAATCCGGTCTC CCTATAGTG	Bottom strand Quikchange primer to remove the <u>EcoRI</u> site from p11-lacY-wtx1
DE-1173	<u>CTAGACATACGGAATTCCATACGCAT</u> <u>G</u>	Top strand oligo containing an EcoRI site to be cloned into p11-lacY-wtx1 <u>XbaI/SphI</u>
DE-1174	<u>TGTATGCCTTAAGGTATGC</u>	Bottom oligo containing an EcoRI site to be cloned into p11-lacY-wtx1 <u>XbaI/SphI</u>
DE-1228	GCGGAATTCGANNNNGNNNNNNNNN CATGGCCTTGGGAAATCCCTTCAATG TATCCGGCATG	Top strand oligo with randomized base pairs in the I-BmoI <i>thyA</i> target site (G-2 fixed); <u>EcoRI</u> site and <u>SphI</u> overhang

DE-1229	<u>CCGGAATACATTGAAGGG</u>	Bottom strand primer with <u>SphI</u> site used to fill in DE-1228
DE-1230	<u>AATTCGAGCCCGTAGTAATGATATTT</u> CATGGCCTTGGGAAATCCCTTCAATG TATTCGGCATG	Top strand 49-mer <i>thyA</i> target with TATTT insertion between +8/+9, <u>EcoRI/SphI</u>
DE-1231	<u>CCGGAATACATTGAAGGGATTCCCA</u> AGGCCATG AAAT ATCATTACTACGG CTCG	Bottom strand 49-mer <i>thyA</i> target with AAATA insertion between +8/+9, <u>SphI/ EcoRI</u>
DE-1232	<u>AATTCGAGCCCGTAGTAATGACGCC</u> CATGGCCTTGGGAAATCCCTTCAATG TATTCGGCATG	Top strand 49-mer <i>thyA</i> target with CGCCC insertion between +8/+9, <u>EcoRI/SphI</u>
DE-1233	<u>CCGGAATACATTGAAGGGATTCCCA</u> AGGCCATG GGGCG TCATTACTACGG GCTCG	Bottom strand 49-mer <i>thyA</i> target with GGGCG insertion between +8/+9, <u>SphI/ EcoRI</u>
DE-1234	<u>CTAGAGAGCCCGTAGTAATGATTTCA</u> TGGCCTTGGGAAATCCCTTCAATGTA TTCCGG	Top strand 49-mer <i>thyA</i> target with TTT insertion between +8/+9, <u>XbaI/EcoRI</u>
DE-1235	<u>AATTCGGAATACATTGAAGGGATTT</u> CCCAAGGCCATG AAAT CATTACTACG GGCTCT	Bottom strand 49-mer <i>thyA</i> target with AAA insertion between +8/+9, <u>EcoRI/XbaI</u>
DE-1236	<u>CTAGAGAGCCCGTAGTAATGACCCC</u> ATGGCCTTGGGAAATCCCTTCAATGT ATTCCGG	Top strand 49-mer <i>thyA</i> target with CCC insertion between +8/+9, <u>XbaI/EcoRI</u>
DE-1237	<u>AATTCGGAATACATTGAAGGGATTT</u> CCCAAGGCCATG GGG TCATTACTACG GGCTCT	Bottom strand 49-mer <i>thyA</i> target with GGG insertion between +8/+9, <u>EcoRI/XbaI</u>
DE-1238	<u>CTAGAGAGCCCGTAGTATTTATGACA</u> TGGCCTTGGGAAATCCCTTCAATGTA TTCCGG	Top strand 49-mer <i>thyA</i> target with TTT insertion between +4/+5, <u>XbaI/EcoRI</u>
DE-1239	<u>AATTCGGAATACATTGAAGGGATTT</u> CCCAAGGCCATGTCATA AAATA CTACG GGCTCT	Bottom strand 49-mer <i>thyA</i> target with AAA insertion between +4/+5, <u>EcoRI/XbaI</u>
DE-1240	<u>CTAGAGAGCCCGTAGTACCCATGAC</u> ATGGCCTTGGGAAATCCCTTCAATGT ATTCCGG	Top strand 49-mer <i>thyA</i> target with CCC insertion between +4/+5, <u>XbaI/EcoRI</u>
DE-1241	<u>AATTCGGAATACATTGAAGGGATTT</u> CCCAAGGCCATGTCATG GGT ACTACG GGCTCT	Bottom strand 49-mer <i>thyA</i> target with GGG insertion between +4/+5, <u>EcoRI/XbaI</u>
DE-1280	<u>CCATCTCATCCCTGCGTGTCTCCGACT</u> <u>CAGTTGGTGGATAACAGGGTAATAT</u> CACGC	Primer with <u>left adaptor</u> for Ion Torrent sequencing, barcode , and complementarity to pKox (for the substrate library)
DE-1281	<u>CCATCTCATCCCTGCGTGTCTCCGACT</u> <u>CAGTCTTCGGATAACAGGGTAATATC</u> ACGC	Primer with <u>left adaptor</u> for Ion Torrent sequencing, barcode , and complementarity to pKox (for the non-survivor library)
DE-1282	<u>CCATCTCATCCCTGCGTGTCTCCGACT</u> <u>CAGTCAAGGGATAACAGGGTAATAT</u> CACGC	Primer with <u>left adaptor</u> for Ion Torrent sequencing, barcode , and complementarity to pKox (for the survivor library)
DE-1289	<u>CCTCTCTATGGGCAGTCGGTGATGCA</u> <u>TGCCGGAATACATTGAAGGG</u>	Primer with <u>right adaptor</u> for Ion Torrent sequencing and complementarity to pKox

DE-1348	<u>CTAGAGAGCCCGTAGTAATAACATGG</u> CCTTGGGAAATCCCTTCAATGTATTC CGG	Top strand 49 mer <i>thyA</i> target with G+7A mutation, <u>XbaI/EcoRI</u> ends
DE-1349	<u>AATTC</u> CGGAATACATTGAAGGGATTT CCCAAGGCCATGTTATTACTACGGGC TCT	Bottom strand 49 mer <i>thyA</i> target with G+7A mutation, <u>EcoRI/XbaI</u> ends
DE-1350	<u>CTAGAGAGCCCGTAGTAATCACATGG</u> CCTTGGGAAATCCCTTCAATGTATTC CGG	Top strand 49 mer <i>thyA</i> target with G+7C mutation, <u>XbaI/EcoRI</u> ends
DE-1351	<u>AATTC</u> CGGAATACATTGAAGGGATTT CCCAAGGCCATGTGATTACTACGGGC TCT	Bottom strand 49 mer <i>thyA</i> target with G+7C mutation, <u>EcoRI/XbaI</u> ends
DE-1352	<u>CTAGAGAGCCCGTAGTAATTACATGG</u> CCTTGGGAAATCCCTTCAATGTATTC CGG	Top strand 49 mer <i>thyA</i> target with G+7T mutation, <u>XbaI/EcoRI</u> ends
DE-1353	<u>AATTC</u> CGGAATACATTGAAGGGATTT CCCAAGGCCATGTAATTACTACGGGC TCT	Bottom strand 49 mer <i>thyA</i> target with G+7T mutation, <u>EcoRI/XbaI</u> ends
DE-1356	AAACTCCACCTTGAGGTAAGAGCCCG TAGTAAT <u>A</u> ACATGGCCTTGGGAAATC CCTTCAATGTATTCCAGTACAA	Top strand intronless I-BmoI <i>thyA</i> target 74-mer from +27 to -47 relative to the intron insertion site, with <u>G+7A</u> mutation
DE-1357	TTGTACTGGAATACATTGAAGGGATT TCCCAAGGCCATGT <u>T</u> ATTACTACGGG CTCTTACCTCAAGGTGGAGTTT	Bottom strand intronless I-BmoI <i>thyA</i> target 74-mer from +27 to -47 relative to the intron insertion site, with <u>C+7T</u> mutation
DE-1358	AAACTCCACCTTGAGGTAAGAGCCCG TAGTAAT <u>C</u> ACATGGCCTTGGGAAATC CCTTCAATGTATTCCAGTACAA	Top strand intronless I-BmoI <i>thyA</i> target 74-mer from +27 to -47 relative to the intron insertion site, with <u>G+7C</u> mutation
DE-1359	TTGTACTGGAATACATTGAAGGGATT TCCCAAGGCCATGT <u>G</u> ATTACTACGGG CTCTTACCTCAAGGTGGAGTTT	Bottom strand intronless I-BmoI <i>thyA</i> target 74-mer from +27 to -47 relative to the intron insertion site, with <u>C+7G</u> mutation
DE-1360	AAACTCCACCTTGAGGTAAGAGCCCG TAGTAAT <u>T</u> ACATGGCCTTGGGAAATC CCTTCAATGTATTCCAGTACAA	Top strand intronless I-BmoI <i>thyA</i> target 74-mer from +27 to -47 relative to the intron insertion site, with <u>G+7T</u> mutation
DE-1361	TTGTACTGGAATACATTGAAGGGATT TCCCAAGGCCATGT <u>A</u> ATTACTACGGG CTCTTACCTCAAGGTGGAGTTT	Bottom strand intronless I-BmoI <i>thyA</i> target 74-mer from +27 to -47 relative to the intron insertion site, with <u>C+7A</u> mutation

Supplementary Table S4.3: K_D values (μM) for I-BmoI truncations on DNA substrates

substrate	I-BmoI domain			
	N92	N111	N130	N154
intronless <i>thyA</i>	> 23	> 35	> 15	0.50 ± 0.05
intron-containing	> 23	> 35	> 15	0.44 ± 0.05
non-specific	> 23	> 35	> 42	> 15

S4 Supplementary references

28. Kleinstiver,B.P., Fernandes,A.D., Gloor,G.B. and Edgell,D.R. (2010) A unified genetic, computational and experimental framework identifies functionally relevant residues of the homing endonuclease I-BmoI. *Nucleic Acids Res.*, **38**, 2411-27.
31. Edgell,D.R., Stanger,M.J. and Belfort,M. (2003) Importance of a single base pair for discrimination between intron-containing and intronless alleles by endonuclease I-BmoI. *Curr. Biol.*, **13**, 973-8.
34. Carter,J.M., Friedrich,N.C., Kleinstiver,B. and Edgell,D.R. (2007) Strand-specific contacts and divalent metal ion regulate double-strand break formation by the GIY-YIG homing endonuclease I-BmoI. *J. Mol. Biol.*, **374**, 306-21.
40. Chen,Z. and Zhao,H. (2005) A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic Acids Res.*, **33**, e154.
60. Crona,M., Moffatt,C., Friedrich,N.C., Hofer,A., Sjoberg,B. and Edgell,D.R. (2011) Assembly of a fragmented ribonucleotide reductase by protein interaction domains derived from a mobile genetic element. *Nucleic Acids Res.*, **39**, 1381-89.

Appendix S5: Supplementary information for Chapter 5

S5 Supplementary materials and methods

Bacterial strains and plasmid construction. *Escherichia coli* strains DH5 α and ER2566 (New England Biolabs) were used for plasmid manipulations and protein expression, respectively. *E. coli* strain BW25141(λ DE3) was used for genetic selection assays (1). A complete description of all plasmids used in this study are listed in Supplementary Table S5.3, and oligonucleotides are listed in Supplementary Table S5.4. The ryA and ryB zinc-finger genes were synthesized by Integrated DNA Technologies with 5'-BamHI and 3'-XhoI sites and a C-terminal 6-histidine tag, and cloned into pACYCDuet-1 to generate pACYCryAZf+H and pACYCryBZF+H, respectively. A stop codon was introduced at the 3' end of the ryAZf gene using Quikchange (Stratagene) to generate pACYCryAZf. To create GIY-ZFEs, the I-TevI and I-BmoI GIY-YIG domains were PCR amplified from bacteriophage T4 gDNA and pACYCIBmoI, respectively, and cloned into the NcoI/BamHI sites of pACYCryAZf+H, pACYCryAZf, and pACYCryBzf+H. TevN201-ryA and TevN201R27A were subcloned into the XbaI and EcoRV sites of pTAL3 to generate the expression plasmids for the yeast reporter assay (pYTZN201 and pYTZN201R27A). To generate Tev-LHEs, the I-OnuI E1 gene was amplified with BamHI and Sall ends to clone into the BamHI and XhoI sites of pACYCDuet-1(PciI) to create pACYCOnuE1(+H). This vector was subsequently Quikchanged to introduce an E22Q mutation in I-OnuI E1 to create pACYCOnuE1E22Q (+H). I-TevI catalytic domains were amplified as above and cloned into PciI/BamHI of pACYCOnuE1E22Q (+H). The R27A mutants of Tev-ZFEs and Tev-LHEs were generated using Quikchange mutagenesis. Hybrid GIY-ZFE and Tev-LHE target sites (Fig. 5.1B and 5.4B, Supplementary Tables S5.1 and S5.2) were cloned into the toxic plasmid p11-lacY-wtx1 to generate reporter plasmids for the bacterial selection. Tev-ryA and Bmo-ryA target sites were cloned into pSP72 for *in vitro* cleavage assays. The Tev-ryA site hybrid homing site was also cloned into LITMUS28i using BamHI and XhoI to generate pTZHS1.35. The two-site Tev-ZF plasmids were created by sub-cloning the PvuII/HpaI fragment from pTZHS1.35 into the SwaI site of pTZHS1.35 to generate pTZHS2.35 and pTZHS3.35 (with the second TZHS in either orientation). The G5A or C1A/G5A

mutations were introduced into pToxTZ and pTZHS plasmids by Quikchange mutagenesis. To generate the target plasmids for the yeast reporter assay, the TZ-ryA target sites from toxic plasmids containing TZ1.33, TZ1.33G5A, and TZ1.33C1A/G5A were amplified and cloned into the BglIII and SpeI sites of pCS753. All constructs were verified by sequencing.

Two-plasmid genetic selection. The two plasmid genetic selection was performed as described (1), with toxic (reporter) plasmids containing hybrid TZ-ryA, TZ-ryB, BZ-ryA, TO target sites (Supplementary Tables S5.1 and S5.2), mutant target sites (with G5A or C1A/G5A substitutions), or plasmids lacking a target site (p11-lacY-wtx1). Survival percentage was calculated by dividing the number of colonies observed on selective by those observed on non-selective plates.

Yeast reporter assay. Transformants of *S.cerevisiae* YPH500(α) with Tev-ZFE or Tev-TAL constructs and YPH499(a) with target constructs were grown overnight (~230 rpm) at 30°C in synthetic complete medium lacking histidine (Tev-ZFEs and Tev-TALs) or lacking tryptophan and uracil (targets). Tev-ZFEs and targets were mated by adding equal densities (~400 μ l) of overnight culture to 1 ml YPD and left stationary for 5-6 hours at 30°C. Cells harvested by centrifugation were washed in 1 ml and resuspended in 4 ml of synthetic medium lacking histidine and tryptophan prior to shaking overnight at 30°C. Cells were harvested by centrifugation, washed in 1 ml Z buffer (60 mM Na₂PO₄, 40 mM NaH₂PO₄, 10 mM KCl, 1 mM MgSO₄, pH 7.0), and suspended in 250 μ l Z buffer. The suspension was diluted 20-fold into 1 ml Z buffer containing 0.27% β -mercaptoethanol, and 75 μ l CHCl₃ and 45 μ l 0.1% SDS were added prior to vortexing. Lysates were pre-incubated at 30°C prior to the addition of 100 μ l 4 mg/ml ONPG. Reactions proceeded until a yellow colour developed whereby progress was stopped by the addition of 300 μ l 1M Na₂CO₃. Stopped reactions were pelleted and the absorbance of the supernatant was analyzed at 420 nm and 550 nm.

Protein purification. Cultures overexpressing either TevN201-ZFE or BmoN221-ZFE were grown at 37°C to an OD₆₀₀ ~0.5 and expression induced by 0.5 mM IPTG (Bio Basic Inc.) overnight at 15°C. Cells were harvested by centrifugation at 8983 x g for 12

minutes, re-suspended in binding buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM imidazole, 5% glycerol, and 1 mM DDT), and lysed by homogenization at 15,000 psi. The cell lysate was clarified by centrifugation at 20400 x g, followed by sonication for 30 seconds, and centrifugation at 20400 x g for 15 minutes. The clarified lysate was loaded onto a 1 mL HisTrap-HP column (GE Healthcare), washed with 15 mL binding buffer and then 10 mL wash buffer (20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 50 mM imidazole, 5% glycerol, and 1 mM DTT). Bound proteins were eluted in 1.5 mL fractions in four 5 mL step elutions with increasing concentrations of imidazole. Fractions containing GIY-ZFEs were dialyzed twice against 1L dialysis buffer (20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 5% glycerol, and 1 mM DDT) prior to storage at -80°C.

Cleavage assays. Single time-point cleavage assays to determine the $EC_{0.5max}$ of TevN201-ryA were performed in buffer containing 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM $MgCl_2$, 5% glycerol, 1 mM DTT and 10 nM pTZHS1.33. Reactions were incubated for 3 minutes at 37°C, stopped with 5 μ l stop solution (100 mM EDTA, 40% glycerol, and bromophenol blue), and electrophoresed on a 1% agarose gel prior to staining with ethidium bromide and analysis on an AlphaImagerTM3400 (Alpha Innotech). The $EC_{0.5max}$ was determined by fitting the data to the equation

$$f_{(endo)} = \frac{f_{max} * [endo]^H}{EC_{0.5max} + [endo]^H}$$

where $f_{(endo)}$ is the fraction of substrate cleaved at concentration of TevN201-ryA [endo], f_{max} is the maximal fraction cleavage, with 1 being the highest value, and H is the Hill constant that was set to 1. The initial reaction velocity was determined using supercoiled plasmid substrate with varying concentrations of TevN201-ryA (0.7 nM to 47 nM) and buffer as above. Aliquots were removed at various times, stopped and analyzed as above.

The data for product appearance was fitted to the equation $P = A(1 - e^{-k_1 t}) + k_2 t$

where P is product (in nM), A is the magnitude of the initial burst, k_1 is the rate constant (s^{-1}) of the initial burst phase and k_2 is the steady state rate constant (s^{-1}). The two-site plasmid cleavage assays were conducted as above, using 10 nM pTZHS2.33 or pTZHS3.33 as substrates, and ~90 nM purified TevN201-ryA. The k_{obs} rate constants were calculated from the decay of supercoiled substrate by fitting to the equation

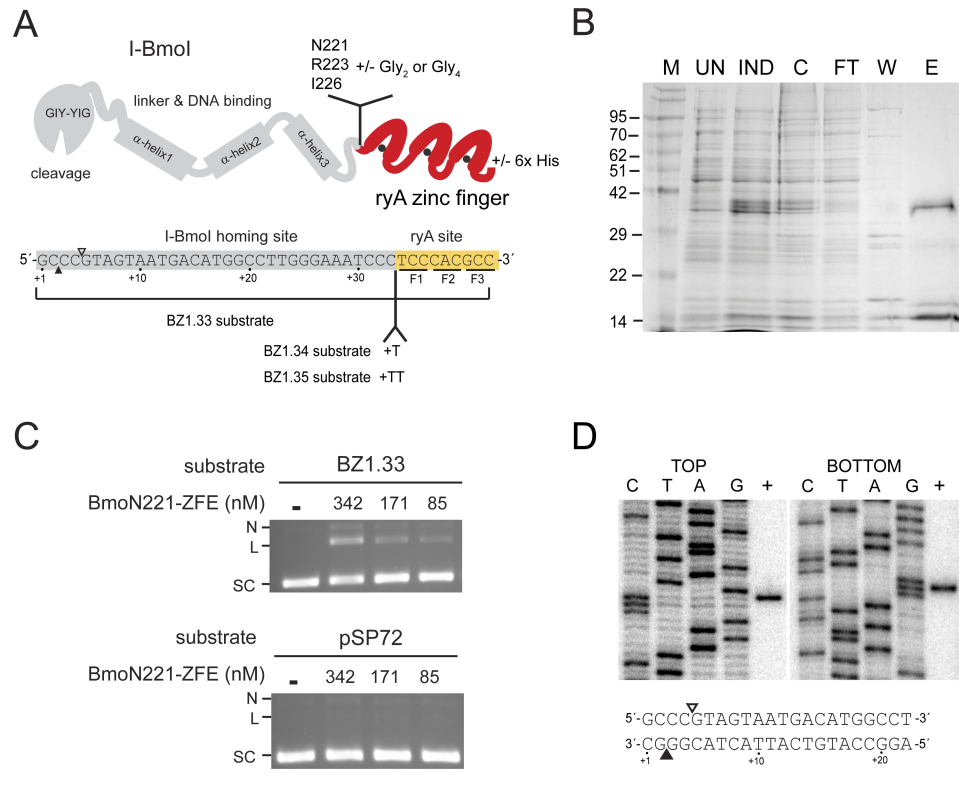
$$[C] = [C_0]\exp(-k_1 t)$$

where $[C]$ is the concentration (nM) of supercoiled plasmid at time t , $[C_0]$ is the initial concentration of supercoiled substrate (nM), and k_1 is the first order rate constant (in s^{-1}). At least 3 independent trials were conducted for each data set.

Cleavage mapping. Mapping of cleavage sites was performed as described (2). Briefly, primers were individually end-labeled with γ - ^{32}P ATP, and used in PCR reactions with pTox or pSP72 plasmids carrying TZ-ryA or BZ-ryA target sites to generate strand-specific substrates. The substrates were incubated with purified protein as above, and electrophoresed in 8% denaturing gels alongside sequencing ladders generated by cycle sequencing with the same end-labeled primers (US Biologicals).

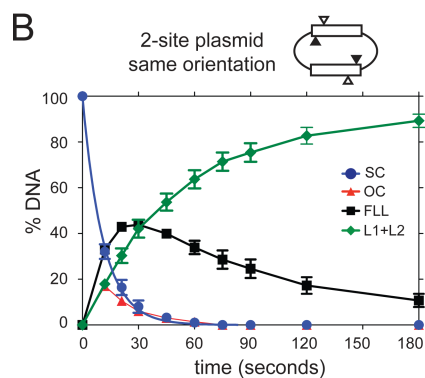
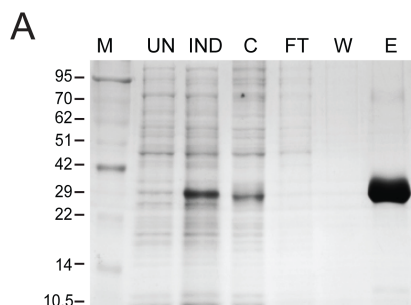
Bioinformatics. The distribution of the CNNNG cleavage motif was examined using a custom Perl script in a 35-bp window of ZFN sites predicted for *Dania rerio* chromosome 1 (Ensembl release 51) (6). Briefly, 40 bp flanking the downstream region of each predicted ZFN was extracted from the corresponding zebrafish chromosome 1 cDNA, and searched for position $i = \text{C}$ and position $i+4 = \text{G}$, with the occurrences of each CNNNG reported at position i (the C of the motif). The number of CNNNG motifs for unique ZFN sites was fit to a binomial distribution, and plotted in R.

S5 Supplementary figures



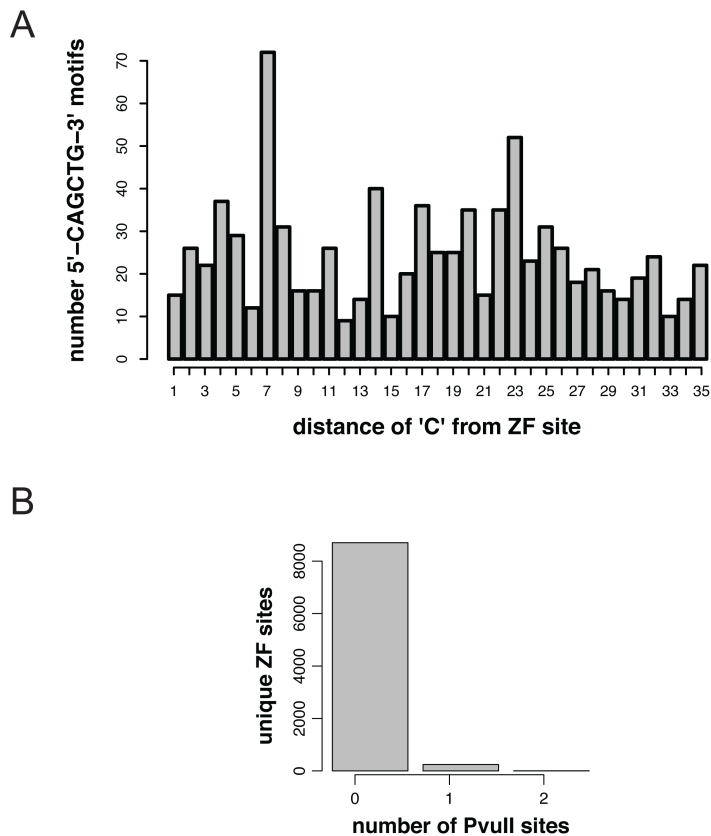
Supplementary Figure S5.1: Design and functionality of Bmo-ZFEs.

(A) Schematic of Bmo-ZFE constructs, with I-BmoI protein and substrate shown in grey, and the ryA protein and binding site shown in red and yellow, respectively. *Top panel*, the fusion points for each of the Bmo-ZFEs are indicated as the last I-BmoI amino acid, with or without a 2xGlycine or 4xGlycine linker. Constructs were made with a 6xHis tag on the C-terminal end. *Bottom panel*, the substrate shown consists of 33-nts of the top strand I-BmoI *thyA* target site (BZ1.33), fused to the 5' end of the ryA binding site. Substrates tested differ by the insertion of one or two T nucleotides at the junction of the *thyA*/ryA sites. (B) Purification of His-tagged BmoN221-ryA. Shown is a representative SDS-PAGE gel; M, marker with molecular weights in kDa indicated on the left; UN, uninduced culture; IND, induced culture; C, crude lysate; FT, flow-through from metal-affinity column; W, wash; E, elution. (C) BmoN221-ryA cleavage specificity. Shown are representative agarose gels of cleavage assays with 10 nM pBZ1.33 or pSP72 (no target site) substrates and the indicated concentrations of BmoN221-ryA under standard assay conditions for 10 minutes. BmoN221-ryA cleaves the BZ-ryA target site plasmid (BZ1.33) but not the control plasmid (pSP72) lacking the target site. N, nicked; L, linear; SC, supercoiled. (D) Mapping of BmoN221-ryA cleavage sites on the BZ1.33 substrate, with top and bottom cleavage sites indicated by open and closed triangles, respectively.



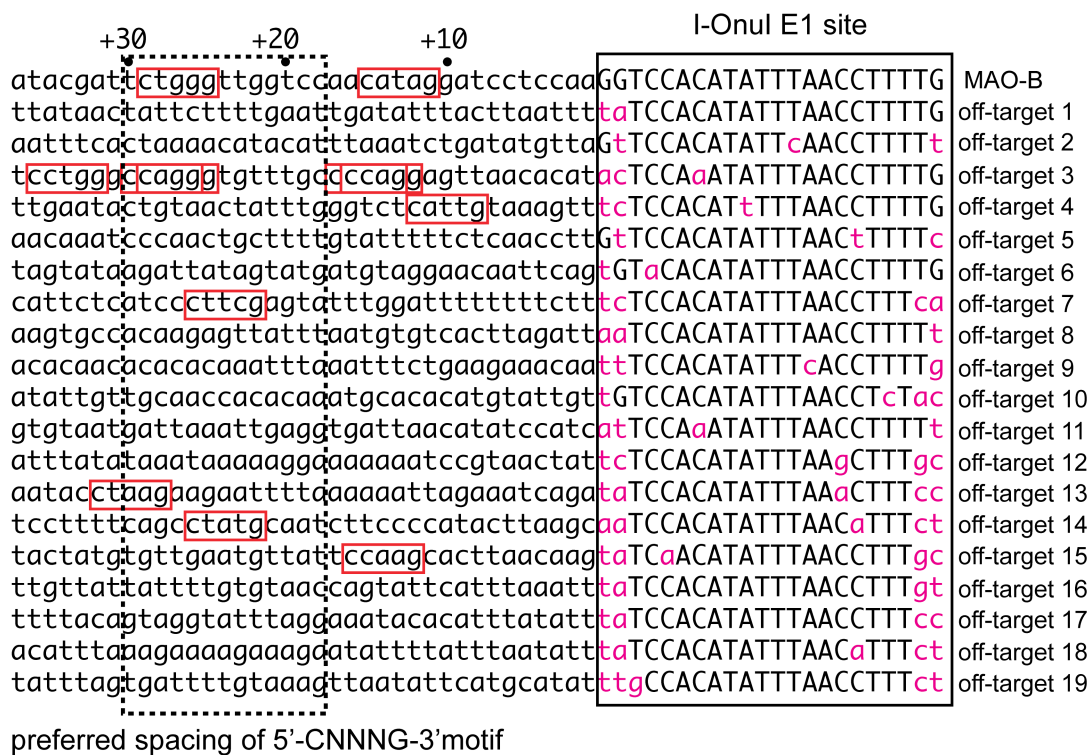
Supplementary Figure S5.2: TevN201-ryA purification for *in vitro* experiments.

(A) Purification of TevN201-ryA. Shown is a representative SDS-PAGE gel; M, marker with molecular weights in kDa indicated on the left; UN, uninduced culture; IND, induced culture; C, crude lysate; FT, flow-through from metal-affinity column; W, wash; E, elution. (B) Graphical representation of cleavage assays with 90 nM TevN201-ZFE and 10 nM two-site pTZ1.33 plasmid with target sites in the same orientation. Data are plotted as averages of three independent replicates with standard deviations; SC, supercoiled; OC, open-circle (nicked); FLL, full-length linear; L1+L2, linear products.



Supplementary Figure S5.3: PvuII site analysis.

(A) Shown is the distribution of the 5'-CAGCTG-3' motif in a 35-bp window flanking 8,829 predicted ZFN sites on zebrafish chromosome 1 (6). The number of occurrences of the 'C' of the motif at each distance is indicated. (B) Unique ZFN sites were grouped according to the number of occurrences of the 5'-CAGCTG-3' motif in the 35-bp window.



Supplementary Figure S5.4: Occurrence of the 5'-CNNNG-3' motif upstream of I-OnuI E1 off-target sites.

Shown is 37-nt of upstream sequence adjacent to the 22-nt I-OnuI E1 MAO-B target site, along with 19 predicted off-target sites (3). CNNNG motifs are highlighted in red, with only 3 of 19 predicted I-OnuI E1 off target sites containing a CNNNG motif at a targetable distance by Tev-LHE fusions. Nucleotide differences of the off-target sites to the I-OnuI E1 site are indicated in magenta lower case font.

S5 Supplementary tables

Supplementary Table S5.1: Tev-ZFE selection data

GIY-ZFE	Toxic plasmid											
	pToxTZ1.30			pToxTZ1.32			pToxTZ1.33			pToxTZ1.34		
	WT	WT	G5A	WT	WT	G5A	C1A/G5A	WT	WT	G5A	C1A/G5A	
TevN201	0	4 ± 1.2 (3)	86.8 ± 5.9 (6)	0	0	0	0	59.9 ± 9.5 (6)	0	0	0.2 ± 0.1 (3)	
TevN201G ₂	ND	ND	72.7 ± 10.7 (6)	0	0	0	0	56.9 ± 11.2 (6)	0	0	0	
TevN201G ₄	ND	ND	83.7 ± 15.2 (4)	0	0	0	0	42.8 ± 12.6 (6)	0	0	0	
TevN201R27A	ND	ND	0	0	0	0	0	0	0	0	0	
TevK203	0	5.4 ± 1.4 (3)	86.8 ± 7.1 (6)	0	0	0	0	50.7 ± 9.5 (6)	0	0	0	
TevK203G ₂	ND	ND	88 ± 13.9 (6)	0	0	0	0	53.7 ± 10.4 (6)	0	0	0	
TevK203G ₄	ND	ND	80.7 ± 7.9 (4)	0.2 ± 0.2 (3)	0.4 ± 0.3 (3)	0	0	43.6 ± 13.0 (6)	0	0	0	
TevK203R27A	ND	ND	0	0	0	0	0	0	0	0	0	
TevS206	0	0	86.6 ± 6.9 (6)	0	0	0	0	47.1 ± 8.6 (6)	0	0	0	
TevS206G ₂	ND	ND	70.7 ± 8.7 (4)	0	0	0	0	27.8 ± 7.4 (6)	0	0	0	
TevS206R27A	ND	ND	0	0	0	0	0	0	0	0	0	
TevN201-ryB	ND	ND	0	ND	ND	ND	ND	ND	ND	ND	ND	

GIY-ZFE	Toxic plasmid											
	pToxTZ1.35			pToxTZ1.36			pToxTZ1.38			pTox ryB		
	WT	G5A	C1A/G5A	WT	WT	G5A	WT	WT	G5A	p11acywtx1	WT	
TevN201	49.8 ± 9.8 (6)	0	0	0	0	0	0	0	0	0	0	
TevN201G ₂	38.6 ± 10.3 (4)	0	0	ND	ND	ND	ND	ND	0	0	ND	
TevN201G ₄	36.3 ± 7.1 (4)	0	0	ND	ND	ND	ND	ND	0	0	ND	
TevN201R27A	0	0	0	ND	ND	ND	ND	ND	0	0	ND	
TevK203	51.0 ± 6.6 (5)	0	0	1.6 ± 0.3 (3)	0	0	0	0	0	0	ND	
TevK203G ₂	46.5 ± 10.9 (5)	0	0	ND	ND	ND	ND	ND	0	0	ND	
TevK203G ₄	48.0 ± 6.1 (4)	0	0	ND	ND	ND	ND	ND	0	0	ND	
TevK203R27A	0	0	0	ND	ND	ND	ND	ND	0	0	ND	
TevS206	62.3 ± 12.4 (4)	0	0	26.1 ± 1.8 (3)	0	0	0	0	0	0	ND	
TevS206G ₂	44.2 ± 16.4 (4)	0	0	ND	ND	ND	ND	ND	0	0	ND	
TevS206R27A	0	0	0	ND	ND	ND	ND	ND	0	0	ND	

*ND, not determined; ±, standard deviation of replicates indicated in brackets

Supplementary Table S5.2: Tev-LHE selection data

Tev-LHE	Toxic plasmid											
	pToxTO1.12		pToxTO1.14		pToxTO1.16		pToxTO1.18		pToxTO1.20		pToxTO1.22	
	WT		WT		WT		WT		WT		WT	
TevS114	0.2 (2)	0	0	0	16.5 ± 6.2 (3)	0	11.6 ± 3.7 (3)	0	0	0	0	0
TevD127	0	0	0.1 (2)	0	31.8 ± 11.3 (3)	0	12.6 ± 3.2 (3)	0	0	0	0	0
TevN140	0	0	0	0	40.2 ± 3.8 (3)	0	24.6 ± 0.4 (3)	0	0	0	0	0
TevN169	0	0	0	0	45.5 ± 11.5 (3)	0	94.5 ± 2.1 (3)	0	0	0	0	0
TevD184G ₂	0	0.1 (2)	0	0	25 ± 6.9 (3)	0	98.3 ± 7.1 (3)	0	0	0	11.9 ± 1.8 (3)	0
TevN201G ₄	0	0	0.1 (2)	0	29.5 ± 1.2 (3)	0	8.3 ± 0.4 (3)	0	0	0	0	0
TevN203	0.8 (2)	0	0.1 (2)	0	24.2 ± 2.0 (3)	0	7.7 ± 3.1 (3)	0	0	0	0	0
Tev-LHE	Toxic plasmid											
	pToxTO1.24		pToxTO1.26		pToxTO1.28		pToxTO1.30		pToxTO1.33-ryA			
	WT		WT		WT		WT		WT		WT	
TevS114	0	65.2 ± 9 (3)	0	0	0	0	ND	ND	ND	ND	ND	ND
TevD127	0	83.2 ± 7.9 (3)	0	0	0	0	ND	ND	ND	ND	ND	ND
TevN140	0	74.9 ± 4.2 (3)	0	0	0	0	ND	ND	ND	ND	ND	ND
TevN169	0	102.2 ± 10.4 (3)	0	0	0.4 ± 0.3 (3)	0	ND	ND	ND	ND	ND	ND
TevD184G ₂	0	87.0 ± 4.7 (3)	0	0	0.0	0	ND	ND	ND	ND	ND	ND
TevN201G ₄	0	76.0 ± 12.3 (3)	0	0	87.1 ± 9.5 (3)	96.9 ± 2.5 (3)	0	0	0	0	0	0
TevN203	0	97.5 ± 5.3 (3)	0	0	99.7 ± 12 (3)	87.2 ± 0.8 (3)	ND	ND	ND	ND	ND	ND
TevR27A(N201G ₄)- OnuE1 (E22Q)	ND	ND	ND	ND	ND	0	0	0	0	0	0	0
TevR27A(N201G ₄)- OnuE1	ND	ND	ND	ND	ND	101.6 ± 4.1 (3)	93.3 ± 4.6 (3)	ND	ND	ND	ND	ND

*ND, not determined; ±, standard deviation of replicates indicated in brackets

Supplementary Table S5.3: Strains and plasmids used in Chapter 5

Strains	Description	Source
<i>E.coli</i> - DH5 α	F ⁻ , ϕ 80d <i>lacZ</i> Δ M15, Δ (<i>lacZYA-argF</i>)U169, <i>deoR</i> , <i>recA1</i> , <i>endA1</i> , <i>hsdR17</i> (rk ⁻ , mk ⁺), <i>phoA</i> , <i>supE44</i> , λ , <i>thi-1</i> , <i>gyrA96</i> , <i>relA1</i>	Invitrogen
<i>E.coli</i> - ER2566	F- λ - <i>fhuA2</i> [lon] <i>ompT lacZ::T7 gene 1 gal sulA11</i> Δ (<i>mcrC-mrr</i>)114::IS10 R(<i>mcr-73::miniTn10-TetS</i>)2 R(<i>zgb-210::Tn10</i>)(TetS) <i>endA1</i> [<i>dcm</i>]	N.E.B.
<i>E.coli</i> - BW25141(λ DE3)	F ⁻ <i>lacI^d rrnB_{T14} DlacZ_{WJ16} DphoBR580 hsdR514 DaraBAD_{AH33} DrhaBAD_{LD78} galU95 endA_{BT333} uidA(DMluI)::pir⁺ recA1</i> , λ DE3 lysogen	Ref (1),(4)
<i>S.cerevisiae</i> - YPH499	MATa <i>ura3-52 lys2-801_amber ade2-101_ochre trp1-Δ63 his3-Δ200 leu2-Δ1</i>	Dr. Bogdanove Ref (5)
<i>S.cerevisiae</i> - YPH500	MAT α <i>ura3-52 lys2-801_amber ade2-101_ochre trp1-Δ63 his3-Δ200 leu2-Δ1</i>	Dr. Bogdanove Ref (5)
Plasmids	Description	Source
pACYCDuet-1	<i>ori_{p15A}</i> , <i>cm</i>	Novagen
pACYCDuet-1(PciI)	<i>ori_{p15A}</i> , <i>cm</i> , pACYCDuet-1 with a PciI site substituted for the NcoI site	Novagen
p11-lacY-wtx1	<i>ori_{pBR322}</i> , <i>amp</i>	Ref (4)
pSP72	<i>ori_{pBR322}</i> , <i>amp</i>	Promega
LITMUS28i	<i>ori_{pMB1}</i> , <i>amp</i>	N.E.B.
pACYCIBmol	pACYCDuet-1, containing the 798bp codon optimized I-Bmol gene in the NdeI and XhoI sites	Ref (1)
pryAzf	<i>ori_{pUC}</i> , <i>kan</i>	I.D.T.
pTAL3	<i>ori_{pBR322}</i> , <i>amp</i>	Dr. Bogdanove
pCP5.1	pCP5 derivative, <i>amp</i> , with ColE1 origin from pBluescript IIKS(-), reporter plasmid for the yeast based recombination assay, see [5]	Dr. Bogdanove
pOnuE1	<i>ori_{pBR322}</i> , <i>amp</i>	Ref (3)
pACYCryAZf+H	pACYCDuet-1, containing the ryA zinc-finger gene with a c-terminal 6-histidine tag cloned into the BamHI and XhoI sites	This study
pACYCryAZf	pACYCDuet-1, containing the ryA zinc-finger gene cloned into the BamHI and XhoI sites	This study
pACYCryBZf+H	pACYCDuet-1, containing the ryB zinc-finger gene with a c-terminal 6-histidine tag cloned into the BamHI and XhoI sites	This study
pTevN201-ZFE (or +H)	pACYCryAZf (or +H), with residues 1-N201 of I-TevI (DE832/840) cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pTevN201-ryB (or +H)	pACYCryBZf (or +H), with residues 1-N201 of I-TevI (DE832/840) cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pTevN201G ₂ -ZFE (or +H)	pACYCryAZf (or +H), with residues 1-N201 of I-TevI (DE833/840) + 2 glycine residues cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pTevN201G ₄ -ZFE (or +H)	pACYCryAZf (or +H), with residues 1-N201 of I-TevI (DE834/840) + 4 glycine residues cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pTevK203-ZFE (or +H)	pACYCryAZf (or +H), with residues 1-K203 of I-TevI (DE835/840) cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pTevK203G ₂ -ZFE (or +H)	pACYCryAZf (or +H), with residues 1-K203 of I-TevI (DE836/840) + 2 glycine residues cloned into the NcoI and BamHI sites (+/- 6xHis)	This study

pTevK203G ₄ -ZFE (or +H)	pACYCryAZf (or +H), with residues 1-K203 of I-TevI (DE837/840) + 4 glycine residues cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pTevS206-ZFE (or +H)	pACYCryAZf (or +H), with residues 1-S206 of I-TevI (DE838/840) cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pTevS206G ₂ -ZFE (or +H)	pACYCryAZf (or +H), with residues 1-S206 of I-TevI (DE839/840) + 2 glycine residues cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pBmoN221-ZFE (or +H)	pACYCryAZf (or +H), with residues 1-N221 of I-BmoI (DE841/849) cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pBmoN221G ₂ -ZFE (or +H)	pACYCryAZf (or +H), with residues 1-N221 of I-BmoI (DE842/849) + 2 glycine residues cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pBmoN221G ₄ -ZFE (or +H)	pACYCryAZf (or +H), with residues 1-N221 of I-BmoI (DE843/849) + 4 glycine residues cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pBmoR223-ZFE (or +H)	pACYCryAZf (or +H), with residues 1-R223 of I-BmoI (DE844/849) cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pBmoR223G ₂ -ZFE (or +H)	pACYCryAZf (or +H), with residues 1-R223 of I-BmoI (DE845/849) + 2 glycine residues cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pBmoR223G ₄ -ZFE (or +H)	pACYCryAZf (or +H), with residues 1-R223 of I-BmoI (DE846/849) + 4 glycine residues cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pBmoI226-ZFE (or +H)	pACYCryAZf (or +H), with residues 1-I226 of I-BmoI (DE847/849) cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pBmoI226G ₂ -ZFE (or +H)	pACYCryAZf (or +H), with residues 1-I226 of I-BmoI (DE848/849) + 2 glycine residues cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pTevN201R27A	Similar to pTevN201-ZFE, with an R27A mutation	This study
pTevN201G ₂ R27A	Similar to pTevN201G ₂ -ZFE, with an R27A mutation	This study
pTevN201G ₄ R27A	Similar to pTevN201G ₄ -ZFE, with an R27A mutation	This study
pTevK203R27A	Similar to pTevK203-ZFE, with an R27A mutation	This study
pTevK203G ₂ R27A	Similar to pTevK203G ₂ -ZFE, with an R27A mutation	This study
pTevK203G ₄ R27A	Similar to pTevK203G ₄ -ZFE, with an R27A mutation	This study
pTevS206R27A	Similar to pTevS206-ZFE, with an R27A mutation	This study
pTevS206G ₂ R27A	Similar to pTevS206G ₂ -ZFE, with an R27A mutation	This study
pToxTZ1.35	p11-lacY-wtx1, that contains a 44-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +8 fused to the 9-bp ryAZf site) cloned XbaI/SphI (DE824/825)	This study
pToxBZ1.35	p11-lacY-wtx1, that contains a 44-bp hybrid I-BmoI/ryA zinc-finger site (<i>thyA</i> bases -6 to +27 fused to the 9-bp ryAZf site) cloned XbaI/SphI (DE826/827)	This study
pSP-TZHS1.35	pSP72, that contains a 44-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +8 fused to the 9-bp ryAZf site) cloned XbaI/SphI (DE824/825)	This study
pTZHS1.35	LITMUS28i, with the 44-bp hybrid I-TevI/ryA zinc-finger site (<i>td</i> bases -27 to +8 fused to the 9-bp ryAZf site) sub-cloned from pSP-TZHS1.35 into BamHI/XhoI	This study
pBZHS1.35	pSP72, that contains a 44-bp hybrid I-BmoI/ryA zinc-finger homing site (<i>thyA</i> bases +6 to -27 fused to the 9-bp ryAZf site) cloned XbaI/SphI (DE826/827)	This study

pTZHS2.35	Similar to pTZHS1.35, with a second Tev-ZFE1.35 target site sub-cloned from pSP-TZHS1.35 (using PvuII/HpaI) into the SwaI site	This study
pTZHS3.35	Similar to pTZHS2.35, with the second Tev-ZFE1.35 target site in the alternate orientation	This study
pToxTZ1.35G5A	Similar to pToxTZ1.35, with a G5A substitution (DE917/918)	This study
pToxTZ1.35C1A/G5A	Similar to pToxTZ1.35, with C1A and G5A substitutions (DE919/920)	This study
pTZHS1.35G5A	Similar to pTZHS1.35, with a G5A substitution	This study
pTZHS1.35C1A/G5A	Similar to pTZHS1.35, with C1A and G5A substitutions	This study
pToxTZ1.34	p11-lacY-wtx1, that contains a 43-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +7 fused to the 9-bp ryAZf site) cloned XbaI/SphI	This study
pToxTZ1.34G5A	Similar to pToxTZ1.34, with G5A substitution	This study
pToxTZ1.34C1A/G5A	Similar to pToxTZ1.34, with C1A and G5A substitutions	This study
pToxTZ1.33	p11-lacY-wtx1, that contains a 42-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +6 fused to the 9-bp ryAZf site) cloned into the XbaI and SphI sites	This study
pToxTZ1.33G5A	Similar to pToxTZ1.33, with G5A substitution	This study
pToxTZ1.33C1A/G5A	Similar to pToxTZ1.33, with C1A and G5A substitutions	This study
pToxTZ1.33-ryB	p11-lacY-wtx1, that contains a 42-bp hybrid I-TevI/ryB zinc-finger homing site (<i>td</i> bases -27 to +6 fused to the 9-bp ryBZf site) cloned into XbaI/SphI sites	This study
pToxBZ1.34	p11-lacY-wtx1, that contains a 43-bp hybrid I-BmoI/ryA zinc-finger site (<i>thyA</i> bases -6 to +26 fused to the 9-bp ryAZf site) cloned XbaI/SphI (DE826/827)	This study
pToxBZ1.33	p11-lacY-wtx1, that contains a 42-bp hybrid I-BmoI/ryA zinc-finger site (<i>thyA</i> bases -6 to +25 fused to the 9-bp ryAZf site) cloned into XbaI/SphI (DE826/827)	This study
pTZHS1.34	Similar to pTZHS1.35, with a 43-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +7 fused to the 9-bp ryAZf site)	This study
pTZHS1.33	Similar to pTZHS1.35, with a 42-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +6 fused to the 9-bp ryAZf site)	This study
pBZHS1.34	Similar to pBZHS1.35, with a 43-bp hybrid I-BmoI/ryA zinc-finger homing site (<i>thyA</i> bases +6 to -26 fused to the 9-bp ryAZf site)	This study
pBZHS1.33	Similar to pBZHS1.35, with a 42-bp hybrid I-BmoI/ryA zinc-finger homing site (<i>thyA</i> bases +6 to -25 fused to the 9-bp ryAZf site)	This study
pTZHS2.34	Similar to pTZHS2.35, with both Tev-ZFE target sites as 43-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +7 fused to the 9-bp ryAZf site)	This study
pTZHS3.34	Similar to pTZHS3.35, with both Tev-ZFE target sites as 43-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +7 fused to the 9-bp ryAZf site)	This study
pTZHS2.33	Similar to pTZHS2.35, with both Tev-ZFE target sites as 42-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +6 fused to the 9-bp ryAZf site)	This study

pTZHS3.33	Similar to pTZHS3.35, with both Tev-ZFE target sites as 42-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +6 fused to the 9-bp ryAZf site)	This study
pTZHS1.34G5A	Similar to pTZHS1.34, with a G5A substitution	This study
pTZHS1.33G5A	Similar to pTZHS1.33, with a G5A substitution	This study
pTZHS1.34C1A/G5A	Similar to pTZHS1.34, with C1A and G5A substitutions	This study
pTZHS1.33C1A/G5A	Similar to pTZHS1.33, with C1A and G5A substitutions	This study
pToxTZ1.30	p11-lacY-wtx1, that contains a 39-bp hybrid I-TevI/ryA zinc-finger site (<i>td</i> bases -27 to +3 fused to the 9-bp ryAZf site) cloned XbaI/SphI (DE1085/1086)	This study
pToxTZ1.32	p11-lacY-wtx1, that contains a 41-bp hybrid I-TevI/ryA zinc-finger site (<i>td</i> bases -27 to +5 fused to the 9-bp ryAZf site) cloned XbaI/SphI (DE1087/1088)	This study
pToxTZ1.36	p11-lacY-wtx1, that contains a 45-bp hybrid I-TevI/ryA zinc-finger site (<i>td</i> bases -27 to +9 fused to the 9-bp ryAZf site) cloned XbaI/SphI (DE1134/1135)	This study
pToxTZ1.38	p11-lacY-wtx1, that contains a 47-bp hybrid I-TevI/ryA zinc-finger site (<i>td</i> bases -27 to +11 fused to the 9-bp ryAZf site) cloned XbaI/SphI (DE1136/1137)	This study
pToxTZ1.33ryB	p11-lacY-wtx1, that contains a 42-bp hybrid I-TevI/ryB zinc-finger homing site (<i>td</i> bases -27 to +6 fused to the 9-bp ryBZf site) cloned XbaI/SphI	This study
pToxTO1.12	p11-lacY-wtx1, that contains a 34-bp hybrid I-TevI/I-OnuI E1 site (<i>td</i> bases -27 to -16 fused to the I-OnuI E1 site) cloned XbaI/SphI (DE1072/1073)	This study
pToxTO1.14	p11-lacY-wtx1, that contains a 36-bp hybrid I-TevI/I-OnuI E1 site (<i>td</i> bases -27 to -14 fused to the I-OnuI E1 site) cloned XbaI/SphI (DE1070/1071)	This study
pToxTO1.16	p11-lacY-wtx1, that contains a 38-bp hybrid I-TevI/I-OnuI E1 site (<i>td</i> bases -27 to -12 fused to the I-OnuI E1 site) cloned XbaI/SphI (DE1068/1069)	This study
pToxTO1.18	p11-lacY-wtx1, that contains a 40-bp hybrid I-TevI/I-OnuI E1 site (<i>td</i> bases -27 to -10 fused to the I-OnuI E1 site) cloned XbaI/SphI (DE1066/1067)	This study
pToxTO1.20	p11-lacY-wtx1, that contains a 42-bp hybrid I-TevI/I-OnuI E1 site (<i>td</i> bases -27 to -8 fused to the I-OnuI E1 site) cloned into the XbaI and SphI (DE1064/1065)	This study
pToxTO1.22	p11-lacY-wtx1, that contains a 44-bp hybrid I-TevI/I-OnuI E1 site (<i>td</i> bases -27 to -6 fused to the I-OnuI E1 site) cloned XbaI/SphI (DE1062/1063)	This study
pToxTO1.24	p11-lacY-wtx1, that contains a 46-bp hybrid I-TevI/I-OnuI E1 homing site (<i>td</i> bases -27 to -4 fused to the I-OnuI E1 site) cloned XbaI/SphI (DE1060/1061)	This study
pToxTO1.26	p11-lacY-wtx1, that contains a 48-bp hybrid I-TevI/I-OnuI E1 site (<i>td</i> bases -27 to -2 fused to the I-OnuI E1 site) cloned into the XbaI and SphI (DE1058/1059)	This study
pToxTO1.28	p11-lacY-wtx1, that contains a 50-bp hybrid I-TevI/I-OnuI E1 site (<i>td</i> bases -27 to +1 fused to the I-OnuI E1 site) cloned into the XbaI and SphI (DE1056/1057)	This study
pToxTO1.30	p11-lacY-wtx1, that contains a 52-bp hybrid I-TevI/I-OnuI E1 site (<i>td</i> bases -27 to +3 fused to the I-OnuI E1 site) cloned into the XbaI and SphI (DE976/977)	This study

pToxTO1.18C1A/G5A	Similar to pToxTO18, with C1A and G5A substitution(DE1154/1155)	This study
pToxTO1.20C1A/G5A	Similar to pToxTO20, with C1A and G5A substitution(DE1156/1157)	This study
pToxTO1.26C1A/G5A	Similar to pToxTO26, with C1A and G5A substitution(DE1158/1159)	This study
pToxTO1.30G5A	Similar to pToxTO30, with a G5A substitution	This study
pToxTO1.30C1A/G5A	Similar to pToxTO30, with C1A and G5A substitution	This study
pACYCOnuE1(+H)	pACYCDuet-1(PciI), containing the I-OnuI E1 gene cloned into the BamHI and XhoI sites	This study
pACYCOnuE1(E22Q)(+H)	pACYCDuet-1(PciI), containing the I-OnuI E1 gene with a E22Q mutation cloned into the BamHI and XhoI sites	This study
pTevK203-OnuE1(E22Q)	pACYCOnuE1(E22Q)(+H), with residues 1-K203 of I-TevI (DE) cloned into the PciI and BamHI sites	This study
pTevN201G ₄ -OnuE1(E22Q)(+H)	pACYCOnuE1(E22Q)(+H), with residues 1-N201 of I-TevI + 4 glycine residues (DE) cloned into the PciI and BamHI sites (+6xHis)	This study
pTevD184G ₂ -OnuE1(E22Q)(+H)	pACYCOnuE1(E22Q)(+H), with residues 1-D184G ₂ of I-TevI + 2 glycine residues (DE) cloned into the PciI and BamHI sites (+6xHis)	This study
pTevN169-OnuE1(E22Q)(+H)	pACYCOnuE1(E22Q)(+H), with residues 1-N169 of I-TevI (DE) cloned into the PciI and BamHI sites (+6xHis)	This study
pTevN140-OnuE1(E22Q)(+H)	pACYCOnuE1(E22Q)(+H), with residues 1-N140 of I-TevI (DE) cloned into the PciI and BamHI sites (+6xHis)	This study
pTevD127-OnuE1(E22Q)(+H)	pACYCOnuE1(E22Q)(+H), with residues 1- D127 of I-TevI (DE) cloned into the PciI and BamHI sites (+6xHis)	This study
pTevS114-OnuE1(E22Q)(+H)	pACYCOnuE1(E22Q)(+H), with residues 1- S114 of I-TevI (DE) cloned into the PciI and BamHI sites (+6xHis)	This study
pTevN201G ₄ R27A-OnuE1(+H)	pACYCOnuE1(+H), with residues 1-N201 of I-TevI + 4 glycine residues (DE) with a R27A mutation in I-TevI (+6xHis)	This study
pTevN201G ₄ R27A-OnuE1(E22Q)(+H)	pACYCOnuE1(E22Q)(+H), similar to pTevN201G ₄ -OnuE1(E22Q)(+H) with an R27A mutation in I-TevI	This study
pZif268	pTAL3 expression vector containing homodimeric FokI-Zif268 ZFN	Dr. Bogdanove Ref (5)
pZif268target	pCS753 reporter plasmid with a homodimeric FokI-Zif268 target site interrupting a partially duplicated <i>lacZ</i> gene	Dr. Bogdanove Ref (5)
pYTZN201	pTAL3, with TevN201-ryA cloned in XbaI/EcoRV using DE1121/1128	This study
pYTZN201R27A	Similar to pYTZN201, with an R27A I-TevI mutation	This study
pTZYHS1.33	pCS753, with the 42-bp hybrid I-TevI/ryA zinc-finger homing site cloned into the BglII/SpeI sites using DE1117/1118	This study
pTZYHS1.33G5A	Similar to PTZYHS1.33, with G5A substitution	This study
pTZYHS1.33C1A/G5A	Similar to PTZYHS1.33, with C1A and G5A substitutions	This study

Supplementary Table S5.4: Oligonucleotides used in Chapter 5

Name	Sequence (5'-3')	Notes
DE410	GGAAGAAGTGGCTGATCTCAGC	Forward primer to generate all cycle-seq products for target sites cloned into pTox
DE411	CAGACCGCTTCTGCGTTCTG	Reverse primer to generate all cycle-seq products for target sites cloned into pTox
DE613	GCTAAAGATTTTAAAAGGCATGG AAGAAGCATTTTAAAG	Forward quikchange primer to create R27A Tev-ZFEs
DE614	CTTTAAAATGCTTCTTCCATGCCTT TTCAAATCTTTAGC	Reverse quikchange primer to create R27A Tev-ZFEs
DE824	<u>CTAGACAACGCTCAGTAGATGTTT</u> <u>TCTTGGGTCTACCGTTTCCCACGCC</u> <u>GCATG</u>	Top-strand oligo to clone the hybrid 35-bp I-TevI/9-bp <i>ryAZf</i> using <u>XbaI</u> and <u>SphI</u>
DE825	<u>CGGCGTGGGAAACGGTAGACCCAA</u> <u>GAAAACATCTACTGAGCGTTGT</u>	Bottom-strand oligo to clone the hybrid 35-bp I-TevI/9-bp <i>ryAZf</i> using <u>XbaI</u> and <u>SphI</u>
DE826	<u>CTAGAGCCCGTAGTAATGACATGG</u> <u>CCTTGGGAAATCCCTTTCCCACGCC</u> <u>GCATG</u>	Top-strand oligo to clone the hybrid 35-bp I-BmoI/9-bp <i>ryAZf</i> using <u>XbaI</u> and <u>SphI</u>
DE827	<u>CGGCGTGGGAAAGGGATTTCCCAA</u> <u>GGCCATGTCATTACTACGGGCT</u>	Bottom-strand oligo to clone the hybrid 35-bp I-BmoI/9-bp <i>ryAZf</i> using <u>XbaI</u> and <u>SphI</u>
DE832	<u>CCGCGGATCCATTACTAGGCTTTT</u> <u>TACC</u>	Reverse primer for TevN201-ZFE cloning, BamHI site underlined
DE833	<u>CCGCGGATCCACCACCATTACTAG</u> <u>GCTTTTACC</u>	Reverse primer for TevN201G ₂ -ZFE cloning, BamHI site underlined
DE834	<u>CCGCGGATCCACCACCACCACCAT</u> <u>TACTAGGCTTTTACC</u>	Reverse primer for TevN201G ₄ -ZFE cloning, BamHI site underlined
DE835	<u>CCGCGGATCCTTTAATATTACTAG</u> <u>GCTTTTACC</u>	Reverse primer for TevK203-ZFE cloning, BamHI site underlined
DE836	<u>CCGCGGATCCACCACCTTTAATAT</u> <u>TACTAGGCTTTTACC</u>	Reverse primer for TevK203G ₂ -ZFE cloning, BamHI site underlined
DE837	<u>CCGCGGATCCACCACCACCACCTT</u> <u>TAATATTACTAGGCTTTTACC</u>	Reverse primer for TevK203G ₄ -ZFE cloning, BamHI site underlined
DE838	<u>CCGCGGATCCTGAAATCTTTTAA</u> <u>TATTACTAGGC</u>	Reverse primer for TevS206-ZFE cloning, BamHI site underlined
DE839	<u>CCGCGGATCCACCACCTGAAATCT</u> <u>TTTTAATATTACTAGGC</u>	Reverse primer for TevS206G ₂ -ZFE cloning, BamHI site underlined
DE840	<u>GCCGCCATGGGTAAAAGCGGAATT</u> <u>TATCAGATT</u>	Forward primer for Tev-ZFE cloning, NcoI site underlined
DE841	<u>CCGCGGATCCGTTTTTTCGGTTTAC</u> <u>GACC</u>	Reverse primer for BmoN221-ZFE cloning, BamHI site underlined
DE842	<u>CCGCGGATCCACCACCGTTTTTTCG</u> <u>GTTTACGACC</u>	Reverse primer for BmoN221G ₂ -ZFE cloning, BamHI site underlined
DE843	<u>CCGCGGATCCACCACCACCACCGT</u> <u>TTTTCGGTTTACGACC</u>	Reverse primer for BmoN221G ₄ -ZFE cloning, BamHI site underlined
DE844	<u>CCGCGGATCCACGAGAGTTTTTTCG</u> <u>GTTTACG</u>	Reverse primer for BmoR223-ZFE cloning, BamHI site underlined
DE845	<u>CCGCGGATCCACCACCACGAGAGT</u> <u>TTTTCGGTTTACG</u>	Reverse primer for BmoR223G ₂ -ZFE cloning, BamHI site underlined
DE846	<u>CCGCGGATCCACCACCACCACCAC</u> <u>GAGAGTTTTTTCGGTTTACG</u>	Reverse primer for BmoR223G ₄ -ZFE cloning, BamHI site underlined
DE847	<u>CCGCGGATCCGATAACCGGACGA</u> <u>GAGTTTTTTCGG</u>	Reverse primer for BmoI226-ZFE cloning, BamHI site underlined

DE848	CCGCGGATCCACCACCGATAACCG GACGAGAGTTTTTCGG	Reverse primer for BmoI226G ₂ -ZFE cloning, BamHI site underlined
DE849	GCCGCCATGGGTAAATCTGGTGTT TACAAAATC	Forward primer for Bmo-ZFE cloning, NcoI site underlined
DE850	CTTGGGTCTACCGT <u>CCCACGCCG</u> ATG	Forward quikchange primer to make the 1.34 I-TevI/ <i>ryA</i> <i>zinc-finger</i> target site
DE851	CATGCGGCGTGGGAACGGTAGACC CAAG	Reverse quikchange primer to make the 1.34 I-TevI/ <i>ryA</i> <i>zinc-finger</i> target site
DE852	CTTGGGTCTACCGT <u>CCCACGCCG</u> ATG	Forward quikchange primer to make the 1.33 I-TevI/ <i>ryA</i> <i>zinc-finger</i> target site
DE853	CATGCGGCGTGGGACGGTAGACCC AAG	Reverse quikchange primer to make the 1.33 I-TevI/ <i>ryA</i> <i>zinc-finger</i> target site
DE854	GCCTTGGGAAATCCCT <u>CCCACGC</u> CGCATG	Forward quikchange primer to make the 1.34 I-BmoI/ <i>ryA</i> <i>zinc-finger</i> target site
DE855	CATGCGGCGTGGGAAGGGATTTCC CAAGGC	Reverse quikchange primer to make the 1.34 I-BmoI/ <i>ryA</i> <i>zinc-finger</i> target site
DE856	GCCTTGGGAAATCCCT <u>CCCACGCC</u> GCATG	Forward quikchange primer to make the 1.33 I-BmoI/ <i>ryA</i> <i>zinc-finger</i> target site
DE857	CATGCGGCGTGGGAGGGATTTCCC AAGGC	Reverse quikchange primer to make the 1.33 I-BmoI/ <i>ryA</i> <i>zinc-finger</i> target site
DE858	CAGAAACAGCTGGTT <u>TAATAACAT</u> CATCACCACtaactcg	Forward quikchange primer to add <u>stops</u> to the 3'-end of the <i>ryA</i> zinc-finger
DE859	cgagttaGTGGTGATGATGTTATTTAAA CCAGCTGTTTCTG	Reverse quikchange primer to add stops to the 3'-end of the <i>ryA</i> zinc-finger
DE917	CTAGACAAC <u>ACTCAGTAGATGTTT</u> TCTTGGGTCTACCGTT <u>CCCACGCC</u> GCATG	Top strand oligo similar to DE824 with <u>G-23A</u> substitution
DE918	CGGCGTGGGAAACGGTAGACCCAA GAAAACATCTACTGAGTGTGT	Bottom strand oligo similar to DE825 with <u>C-23T</u> substitution
DE919	CTAGAAAAC <u>ACTCAGTAGATGTTT</u> TCTTGGGTCTACCGTT <u>CCCACGCC</u> GCATG	Top strand oligo similar to DE824 with <u>G-23A</u> and <u>C-27A</u> substitutions
DE920	CGGCGTGGGAAACGGTAGACCCAA GAAAACATCTACTGAGTGTTT	Bottom strand oligo similar to DE825 with <u>C-23T</u> and <u>G-27T</u> substitutions
DE973	<u>CTAGACAACGCTCAGTAGATGTTT</u> TCTTGGGTCTACCGT <u>TAGCTACTAC</u> GCATG	Top-strand oligo to clone the hybrid 33-bp I-TevI/9-bp <i>ryBZf</i> using <u>XbaI</u> and <u>SphI</u>
DE974	<u>CGTAGTAGCTAACGGTAGACCCAA</u> GAAAACATCTACTGAGCGTTGT	Bottom-strand oligo to clone the hybrid 33-bp I-TevI/9-bp <i>ryBZf</i> using <u>XbaI</u> and <u>SphI</u>
DE976	<u>CTAGACAACGCTCAGTAGATGTTT</u> TCTTGGGTCTAGGTCCACATATTTAA CCTTTTGCATG	Top strand oligo to clone the hybrid 30-bp I-TevI/22-bp <i>I-OnuI E1</i> using <u>XbaI</u> and <u>SphI</u>
DE977	<u>CAAAAGGTTAAATATGTGGACCTAG</u> ACCCAAGAAAACATCTACTGAGCG TTGT	Bottom strand oligo to clone the hybrid 30-bp I-TevI/22-bp <i>I-OnuI E1</i> using <u>XbaI</u> and <u>SphI</u>
DE978	<u>CTAGACAACACTCAGTAGATGTTT</u> TCTTGGGTCTAGGTCCACATATTTAA CCTTTTGCATG	Top strand oligo similar to DE976 with <u>G5A</u> substitution
DE979	<u>CAAAAGGTTAAATATGTGGACCTA</u> GACCCAAGAAAACATCTACTGAGT GTTGT	Bottom strand oligo similar to DE977 with <u>G5A</u> substitution
DE980	<u>CTAGAAAACACTCAGTAGATGTTT</u> TCTTGGGTCTAGGTCCACATATTTAA CCTTTTGCATG	Top strand oligo similar to DE976 with <u>C1A</u> and <u>G5A</u> substitution

DE981	<u>C</u> AAAAAGGTTAAATATGTGGACCTAG ACCCAAGAAAACATCTACTGAGTG TTTT	Bottom strand oligo similar to DE977 with <u>C1A</u> and <u>G5A</u> substitution
DE982	CGGTTTCGCCGACGCGCAAGGCTC CTTTTTGCTGCG	Forward quikchange primer to create E22Q I-OnuI E1 mutant
DE983	CGCAGCAAAAAGGAGCCTTGCGC GTCGGCGAAACCG	Reverse quikchange primer to create E22Q I-OnuI E1 mutant
DE991	CGCGTCGACTTAGAATACTCTGCC CTTGTT	Reverse primer to amplify the I-OnuI E1 gene with a 3' Sall site
DE1017	CGCGGATCCACCACCGTCTGAATG CTTATGATTAAG	Reverse primer for TevD184G ₂ cloning, BamHI site underlined
DE1040	CGCGGATCCAGAACGTTTCTTAAT AATTTC	Reverse primer for TevS114 cloning, BamHI site underlined
DE1042	CGCGGATCCATCAGGTCCAAGTTT AAGC	Reverse primer for TevD127 cloning, BamHI site underlined
DE1044	CGCGGATCCGTTTTTACTTCCGGG TTTAC	Reverse primer for TevN140 cloning, BamHI site underlined
DE1045	CGCGGATCCATTTCTGCATTTACT ACAAG	Reverse primer for TevN169 cloning, BamHI site underlined
DE1056	<u>CTAGACAACGCTCAGTAGATGTTT</u> <u>TCTTGGGTCGGTCCACATATTTAAC</u> <u>TTTTGCATG</u>	Top-strand oligo to clone the hybrid 28-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1057	<u>C</u> AAAAAGGTTAAATATGTGGACCGAC CCAAGAAAACATCTACTGAGCGTT GT	Bottom-strand oligo to clone the hybrid 28-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1058	<u>CTAGACAACGCTCAGTAGATGTTT</u> <u>TCTTGGGGTCCACATATTTAACCTT</u> <u>TTGCATG</u>	Top-strand oligo to clone the hybrid 26-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1059	<u>C</u> AAAAAGGTTAAATATGTGGACCCCA AGAAAACATCTACTGAGCGTTGT	Bottom-strand oligo to clone the hybrid 26-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1060	<u>CTAGACAACGCTCAGTAGATGTTT</u> <u>TCTTGGGTCCACATATTTAACCTTTT</u> <u>GCATG</u>	Top-strand oligo to clone the hybrid 24-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1061	<u>C</u> AAAAAGGTTAAATATGTGGACCCAA GAAAACATCTACTGAGCGTTGT	Bottom-strand oligo to clone the hybrid 24-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1062	<u>CTAGACAACGCTCAGTAGATGTTT</u> <u>TCTGGTCCACATATTTAACCTTTTGC</u> <u>ATG</u>	Top-strand oligo to clone the hybrid 22-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1063	<u>C</u> AAAAAGGTTAAATATGTGGACCAGA AAACATCTACTGAGCGTTGT	Bottom-strand oligo to clone the hybrid 22-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1064	<u>CTAGACAACGCTCAGTAGATGTTT</u> <u>TGGTCCACATATTTAACCTTTTGCAT</u> <u>G</u>	Top-strand oligo to clone the hybrid 20-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1065	<u>C</u> AAAAAGGTTAAATATGTGGACCAAA ACATCTACTGAGCGTTGT	Bottom-strand oligo to clone the hybrid 20-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1066	<u>CTAGACAACGCTCAGTAGATGTTG</u> <u>GTCCACATATTTAACCTTTTGCATG</u>	Top-strand oligo to clone the hybrid 18-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1067	<u>C</u> AAAAAGGTTAAATATGTGGACCAAC ATCTACTGAGCGTTGT	Bottom-strand oligo to clone the hybrid 18-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1068	<u>CTAGACAACGCTCAGTAGATGGGT</u> <u>CCACATATTTAACCTTTTGCATG</u>	Top-strand oligo to clone the hybrid 16-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1069	<u>C</u> AAAAAGGTTAAATATGTGGACCCATC TACTGAGCGTTGT	Bottom-strand oligo to clone the hybrid 16-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>

DE1070	<u>CTAGACAACGCTCAGTAGAGGTCC</u> <u>ACATATTTAACCTTTTGCATG</u>	Top-strand oligo to clone the hybrid 14-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1071	<u>CAAAAGGTTAAATATGTGGACCTCTA</u> <u>CTGAGCGTTGT</u>	Bottom-strand oligo to clone the hybrid 14-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1072	<u>CTAGACAACGCTCAGTAGGTCCAC</u> <u>ATATTTAACCTTTTGCATG</u>	Top-strand oligo to clone the hybrid 12-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1073	<u>CAAAAGGTTAAATATGTGGACCTACT</u> <u>GAGCGTTGT</u>	Bottom-strand oligo to clone the hybrid 12-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1074	<u>CGCGTCGACTTAGTGGTGATGATGG</u> <u>TGATGGAATACTCTGCCCTTGTTTC</u>	Reverse primer to amplify the I-OnuI E1 gene with a 3' <i>his-tag</i> and <u>Sall</u> site
DE1082	<u>GCGAGATCTGGTTCCGCCTATATG</u> <u>TCCCCG</u>	Forward primer to amplify the I-OnuI E1 gene with a 5' BamHI site
DE1085	<u>CTAGACAACGCTCAGTAGATGTTT</u> <u>TCTTGGGTCTACCTCCCACGCCGCA</u> <u>TG</u>	Top-strand oligo to clone the hybrid 32-bp I-TevI/9-bp ryAZf using <u>XbaI</u> and <u>SphI</u>
DE1086	<u>CGGCGTGGGAGGTAGACCCAAGA</u> <u>AAACATCTACTGAGCGTTGT</u>	Bottom-strand oligo to clone the hybrid 32-bp I-TevI/9-bp ryAZf using <u>XbaI</u> and <u>SphI</u>
DE1087	<u>CTAGACAACGCTCAGTAGATGTTT</u> <u>TCTTGGGTCTATCCCACGCCGCATG</u>	Top-strand oligo to clone the hybrid 30-bp I-TevI/9-bp ryAZf using <u>XbaI</u> and <u>SphI</u>
DE1088	<u>CGGCGTGGGATAGACCCAAGAAA</u> <u>ACATCTACTGAGCGTTGT</u>	Bottom-strand oligo to clone the hybrid 30-bp I-TevI/9-bp ryAZf using <u>XbaI</u> and <u>SphI</u>
DE1117	<u>GCAGATCTCTAGCATTACGCTAGG</u> <u>G</u>	Forward primer to amplify target site in pTox with 5' <u>BglII</u> site to clone into pCS753
DE1118	<u>GCACTAGTCTTCTCTCATCCGCC</u>	Reverse primer to amplify target site in pTox with 3' <u>SpeI</u> site to clone into pCS753
DE1121	<u>CGTCTAGAATGAAAAGCGGAATTT</u> <u>ATCAGATT</u>	Forward primer to amplify I-TevI with 5' <u>XbaI</u> site to clone into pTAL3
DE1128	<u>CGGATATCTTATTAACCAGCTGT</u> <u>TTCTGACGCAGG</u>	Reverse primer to amplify ryAZF with 3' <u>EcoRV</u> site to clone into pTAL3
DE1134	<u>CTAGACAACGCTCAGTAGATGTTT</u> <u>TCTTGGGTCTACCGTTTTCCCACGC</u> <u>CGCATG</u>	Top-strand oligo to clone the hybrid 36-bp I-TevI/9-bp ryAZf using <u>XbaI</u> and <u>SphI</u>
DE1135	<u>CGGCGTGGGAAAACGGTAGACCCA</u> <u>AGAAAACATCTACTGAGCGTTGT</u>	Bottom-strand oligo to clone the hybrid 36-bp I-TevI/9-bp ryAZf using <u>XbaI</u> and <u>SphI</u>
DE1136	<u>CTAGACAACGCTCAGTAGATGTTT</u> <u>TCTTGGGTCTACCGTTTTCCCACGCC</u> <u>GCATG</u>	Top-strand oligo to clone the hybrid 38-bp I-TevI/9-bp ryAZf using <u>XbaI</u> and <u>SphI</u>
DE1137	<u>CGGCGTGGGAAAACGGTAGACCCAA</u> <u>GAAAACATCTACTGAGCGTTGT</u>	Bottom-strand oligo to clone the hybrid 38-bp I-TevI/9-bp ryAZf using <u>XbaI</u> and <u>SphI</u>
DE1154	<u>CTAGAAAACACTCAGTAGATGTTT</u> <u>TCTTGGGGGTCCACATATTTAACCTT</u> <u>TTGCATG</u>	Top-strand oligo to clone the hybrid 26-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1155	<u>CAAAAGGTTAAATATGTGGACCCCCA</u> <u>AGAAAACATCTACTGAGTGTTTT</u>	Bottom-strand oligo to clone the hybrid 26-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1156	<u>CTAGAAAACACTCAGTAGATGTTT</u> <u>TGGTCCACATATTTAACCTTTTGCAT</u> <u>G</u>	Top-strand oligo to clone the hybrid 20-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1157	<u>CAAAAGGTTAAATATGTGGACCAAA</u> <u>ACATCTACTGAGTGTTTT</u>	Bottom-strand oligo to clone the hybrid 20-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1158	<u>CTAGAAAACACTCAGTAGATGTTG</u> <u>GTCCACATATTTAACCTTTTGCATG</u>	Top-strand oligo to clone the hybrid 18-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1159	<u>CAAAAGGTTAAATATGTGGACCAAC</u> <u>ATCTACTGAGTGTTTT</u>	Bottom-strand oligo to clone the hybrid 18-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>

S5 Supplementary references

1. Kleinstiver BP, Fernandes A, Gloor GB, Edgell DR (2010) A unified genetic, computational, and experimental framework identifies non-conserved residues as critical for function of the homing endonuclease I-BmoI. *Nucleic Acids Res* 38: 2411-2427.
2. Edgell DR, Shub DA (2001) Related homing endonucleases I-BmoI and I-TevI use different strategies to cleave homologous recognition sites. *Proc Natl Acad Sci U S A* 98: 7898-7903.
3. Takeuchi R, Lambert AR, Mak AN, Jacoby K, Dickson RJ, et al. (2011) Tapping natural reservoirs of homing endonucleases for targeted gene modification. *Proc Natl Acad Sci U S A* 108: 13077-13082.
4. Chen Z, Zhao H (2005) A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic Acids Res* 33: e154.
5. Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, et al. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* 39: e82.
6. Foley JE, et al. (2009) Rapid mutation of endogenous zebrafish genes using zinc finger nucleases made by Oligomerized Pool ENgineering (OPEN). *PLoS One* 4(2):e4348.

BENJAMIN P. KLEINSTIVER

Department of Biochemistry
University of Western Ontario
London, Ontario N6A 5C1

- EDUCATION**
- University of Toronto, Toronto, ON 2003-2007
B.Sc. (Hons with distinct) Biochemistry and Human Biology
- University of Western Ontario, London, ON 2007-2013
Ph.D. Biochemistry
- PUBLICATIONS**
- Kleinstiver, B.P.**, Wang, L., Wang, X., Kolaczyk, T., Schild-Poulter, C., Bogdanove, A.J., Edgell, D.R. (2013) Monomeric TAL effector nucleases. *In preparation*.
- Kleinstiver, B. P.**, Wolfs, J. M., Edgell, D. R. (2013) The monomeric GIY-YIG homing endonuclease I-BmoI uses a molecular anchor and a flexible tether to sequentially nick DNA. *Nucleic Acids Research*, doi: 10.1093/nar/gkt186.
- Kleinstiver, B. P.**, Wolfs, J.M., Kolaczyk, T., Roberts, A.K., Hu, S.X., Edgell, D.R. (2012) Monomeric site-specific nucleases for genome editing. *Proceedings of the National Academy of Sciences U.S.A.* 109(21):8061-8066
- Kleinstiver, B. P.**, Bérubé-Janzen, W., Fernandes, A. D., Edgell, D. R. (2011) Divalent metal ion differentially regulates the sequential nicking reactions of the GIY-YIG homing endonuclease I-BmoI. *PLoS ONE*. 6(8): 323804
- Fernandes, A., **Kleinstiver, B. P.**, Edgell, D. R., Wahl, L. M., Gloor, G.B. (2010) Estimating the evidence and reliability of selection in unigenic evolution. *Algorithms of Molecular Biology*. 5:35
- Kleinstiver, B. P.**, Fernandes, A. D., Gloor, G. B., Edgell, D. R. (2010) A unified genetic, computational and experimental framework identifies functionally relevant residues of the homing endonuclease I-BmoI. *Nucleic Acids Research*. 38(7): 2411-2427
- Carter, J., Friedrich, N., **Kleinstiver, B.P.**, Edgell, D.R. (2007) Strand-specific Contacts and Divalent Metal Ion Regulate Double-strand Break Formation by the GIY-YIG Homing Endonuclease IBmoI. *Journal of Molecular Biology*. 374(2): 306-321

- PATENTS** Edgell, D. R. and **Kleinstiver, B.P.** (2011) Endonuclease. United States Provisional Patent application number 61/628,810.
- ABSTRACTS** **Kleinstiver, B.P.***, Wang, L., Bogdanove, A.J., Edgell, D.R. Genome editing enzymes towards the treatment of HIV. London Health Research Day, London, Ontario. March 19th, 2013 (invited speaker)
- Kleinstiver, B.P.***, Wolfs, J.M., Wang, L., Kolaczyk, T., Bogdanove, A., Edgell, D.R. Monomeric nucleases for genome engineering applications. FASEB Science Research Conference: Genome Engineering – Research & Applications, Lucca, Italy. September 2nd-7th, 2012 (invited speaker)
- Kleinstiver, B.P.***, Wolfs, J.M., Kolaczyk, T., Roberts, A.K., Hu, S.X., Edgell, D.R. Developing precise enzymes for genome editing applications. London Health Research Day, London, Ontario. March 20th, 2012 (abstract/poster)
- Kleinstiver, B.P.***, Wolfs, J.M., Roberts, A.K., Hu, S.X., Edgell, D.R. A monomeric GIY-YIG nuclease domain for genome engineering applications. Workshop on Genome Engineering, University of Washington Medicine, Seattle, Washington). November 8th, 2011 (abstract/poster)
- Kleinstiver, B.P.***, Wolfs, J.M., Edgell, D.R. GIY-YIG homing endonucleases function as monomers. Recombination, replication, repair, and chromosomal segregation conference, University of Western Ontario, London, Ontario. July 2011 (invited speaker)
- Kleinstiver, B. P.***, Edgell, D. R. The GIY-YIG homing endonuclease I-BmoI functions as a monomer to generate a double-strand break. Workshop on Genome Engineering, University of Washington Medicine, Seattle, Washington. November 16th, 2010 (abstract/poster)
- Kleinstiver, B. P.***, Edgell, D. R. Evidence that the GIY-YIG homing endonuclease I-BmoI functions as a monomer to generate a double-strand break. 6th New England Biolabs Meeting on DNA Restriction and Modification, Jacobs University, Bremen, Germany. August 1st-6th, 2010 (abstract/poster)
- Kleinstiver, B. P.***, Fernandes, A. D., Gloor, G. B., Edgell, D. R. MUSE identifies functionally critical residues of I-BmoI. Margaret Moffat Research Day, University of Western Ontario, London, Ontario. April 2010 (abstract/poster)

Kleinstiver, B. P.*, Fernandes, A. D., Gloor, G. B., Edgell, D. R. MUSE identifies functionally critical residues of I-BmoI. Workshop on genome engineering, Seattle Children's Research Institute, Seattle, Washington. November 3rd, 2009 (abstract/poster)

Kleinstiver, B. P.*, Edgell, D. R. An 'aMUSEing' strategy to investigate I-BmoI. Recombination, replication, repair, and chromosomal segregation conference, University of Toronto, Toronto, Ontario. July 2009 (invited speaker)

Kleinstiver, B. P.*, Edgell, D. R. Purification and cleavage mechanism determination of GIY-YIG homing endonuclease I-BmoI. Recombination, replication, repair, and chromosomal segregation conference, University of Western Ontario, London, Ontario. July 2008 (invited speaker)

AWARDS

Alexander Graham Bell Canada Graduate; value \$105,000
May 2010-April 2013, National Science and Engineering Research Council of Canada

Western Graduate Research Scholarship; value \$7000 / year
September 2007-April 2013, Department of Biochemistry, U.W.O.

Schulich Graduate Thesis Research Fund; value \$510
September 2012, Department of Biochemistry, U.W.O

Workshop on Genome Engineering Award; value \$800
November 2011, Northwest Genome Engineering Consortium

New England Biolabs Award; value \$430
August 2010, Jacobs University

Workshop on Genome Engineering Award; value \$1300
November 2010, Northwest Genome Engineering Consortium

Ontario Graduate Scholarship; value \$15,000 (declined)
May 2010-April 2011, Government of Ontario

Schulich Graduate Thesis Research Fund; value \$770
August 2010, Department of Biochemistry, U.W.O.

Workshop on Genome Engineering Award; value \$1200
November 2009, Northwest Genome Engineering Consortium

Ontario Graduate Scholarship; value \$15,000
May 2009-April 2010, Government of Ontario

Ontario Graduate Scholarship in Science and Technology; value
\$15,000

May 2008-April 2009, Government of Ontario

Regents In-Course Scholarship; value \$1000

September 2007-April 2007, University of Toronto

TEACHING

2011 Biochemistry 3380G*, University of Western Ontario

2010 Biochemistry 4445F*, University of Western Ontario

2010 Biochemistry 3380G*, University of Western Ontario

2009 Biochemistry 4445F, University of Western Ontario

*nominated for teaching award

COMMITTEES

University Council on Research Ethics, 2009-2011

Society of Graduate Students, 2009-2012

Biochemistry Department Graduate Studies, 2010-2012

Exam Proctor Committee, 2009-2013

Biochemistry Graduate Student Treasurer, 2010-2012

Area, Safety, and Equipment, 2011-2012

Biochemistry Department Social Committee, 2010-2012