

Electronic Thesis and Dissertation Repository

8-13-2015 12:00 AM

Examining the Nucleotide Preference of the Linker Domain in Engineered Tev-mTALENs

Brendon C. McDowell
The University of Western Ontario

Supervisor
Dr David Edgell
The University of Western Ontario

Graduate Program in Biochemistry
A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science
© Brendon C. McDowell 2015

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

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

Recommended Citation

McDowell, Brendon C., "Examining the Nucleotide Preference of the Linker Domain in Engineered Tev-mTALENs" (2015). *Electronic Thesis and Dissertation Repository*. 3195.
<https://ir.lib.uwo.ca/etd/3195>

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.

EXAMINING THE NUCLEOTIDE PREFERENCE OF THE LINKER DOMAIN IN
ENGINEERED TEV-MTALENS

(Thesis format: Monograph)

by

Brendon C McDowell

Graduate Program in Biochemistry

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

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

© Brendon C McDowell 2015

Abstract

Tev-mTALENs are genome-editing nucleases which combine the nuclease and linker domains of I-TevI with the DNA-binding domain of a TAL effector. The linker domain interacts with a portion of the Tev-mTALEN target site called the DNA Spacer, facilitating DNA cleavage. Linker-DNA Spacer interactions are poorly understood but necessary for Tev-mTALEN activity. I examined the DNA Spacer sequence requirements of the linker by assaying Tev-mTALEN activity on targets with mutated DNA Spacer sequences. I also performed activity assays using Tev-mTALENs with mutations to the I-TevI linker domain. My results indicate that the linker DNA Spacer sequence requirements are highly cryptic. No single nucleotide requirements exist at any position in the DNA Spacer. However, assays with mutant Tev-mTALENs have shown that small amino acid mutations to the linker domain can alter or relax the sequence requirements of Tev-mTALENs, increasing their targeting potential.

Keywords

Engineered Nucleases, Genome Editing, GIY-YIG Homing Endonucleases, I-TevI, Tev-mTALENs, Transcription Activator-like (TAL) Effector

Acknowledgments

I would like to acknowledge my supervisor, Dr David Edgell, for the guidance and direction he provided over the course of my Masters work. I would also like to acknowledge my advisory committee, Dr Greg Gloor and Dr Caroline Schild-Poulter, for their helpful critiques and suggestions pertaining to my research.

Chapters 2, 3: Ben Kleinstiver designed and cloned all wildtype Tev-mTALEN expression constructs for both yeast and bacteria. The 96 well plate adaptation of the β -Galactosidase reporter assay was designed by Tomasz Kolaczyk.

Chapter 3: Tomasz Kolaczyk performed yeast reporter assays with the N15 library substrates and the subsequent isolation of clones for sequencing.

Table of Contents

Abstract	ii
Acknowledgments.....	iv
Table of Contents	v
List of Tables	viii
List of Figures	ix
List of Appendices	xi
List of Abbreviations	xii
Chapter 1	1
1 Introduction	1
1.1 Site-Specific Nucleases for Precise Genome-Editing.....	2
1.2 I-TevI as an Alternative to the FokI Nuclease	7
1.3 I-TevI-based Engineered Nucleases	9
1.4 Hypothesis.....	11
Chapter 2.....	12
2 The I-TevI Nuclease Domain Retains its Activity in a Modular Fashion when Fused to a TAL Effector Domain.....	12
2.1 Tev-mTALENs Retain the Cleavage Motif Requirements of Native I-TevI Enzyme	12
2.2 Tev-mTALENs Nick the Target DNA at the Same Top and Bottom Strand Sites as the Native I-TevI Enzyme.....	15
2.3 Summary	16
Chapter 3.....	22
3 Tev-mTALENs are Sensitive to the Identity of Several Nucleotides in the DNA Spacer.....	22
3.1 Effects of Single Nucleotide Substitutions in the DNA Spacer on Tev-mTALEN Activity	22

3.2 In Vivo Screen of Tev-mTALEN Target Sites with a Randomized DNA Spacer Sequence	24
3.3 Summary	29
Chapter 4.....	31
4 Novel I-TevI Linker Domains with Altered Specificity can Broaden Tev-mTALEN Targeting Potential.....	31
4.1 Spacer Single Nucleotide Substitution Assays with Tev-mTALEN Linker Variants	31
4.2 Summary	35
Chapter 5.....	45
5 Discussion and Conclusion	45
5.1 Summary	45
5.2 Limitations and Future Directions	46
5.2.1 Limitations of Yeast Reporter Assays	46
5.2.2 Limitations of the N169T120 Tev-mTALEN Architecture.....	48
5.2.3 Limitations of Single Nucleotide Preference Analysis.....	50
5.3 Advantages of Tev-mTALENs.....	52
5.4 Conclusion	54
6 Materials and Methods.....	56
6.1 Bacterial and Yeast Strains	56
6.2 Target-Site Plasmid Construction.....	56
6.2.1 <i>In Vitro</i> Cleavage Assays (Chapter 2.2)	56
6.2.2 Control Targets and Cleavage Motif Mutants (Chapter 2.1)	56
6.2.3 DNA Spacer Single Substitution Targets (Chapters 3.1 and 4).....	57
6.2.4 N15 DNA Spacer Library (Chapter 3.2).....	57
6.3 Expression Plasmid Construction	57

6.3.1 Bacterial Expression Plasmid Construction.....	57
6.3.2 Yeast Expression Plasmids	57
6.4 Purification of the 6xHis-tagged Tev-mTALEN	58
6.5 <i>In vitro</i> Cleavage Assays with the Purified Tev-mTALEN.....	59
6.6 Yeast β -Galactosidase Reporter Assays	60
References.....	62
Appendices.....	70
Curriculum Vitae	91

List of Tables

Supplementary Table S1: Figure 2.1A Data	77
Supplementary Table S2: Figure 2.1B Data	77
Supplementary Table S3: Figure 2.2B Data	77
Supplementary Table S4: Figure 2.2C Data	78
Supplementary Table S5: Figure 2.2D Data	78
Supplementary Table S6: Figure 3.1 Data	80
Supplementary Table S7: Figure 3.2 Data	82
Supplementary Table S8: Figure 3.3A Data	83
Supplementary Table S9: Figure 3.3B Data	84
Supplementary Table S10: Figure 4.1-4.4 Data	86
Supplementary Table S11: Figure 4.5A Data	88
Supplementary Table S12: Figure 4.5B Data	90

List of Figures

Figure 1.1: Nuclease-Mediated Genome Editing.....	3
Figure 1.2: FokI-based Engineered Nucleases.....	5
Figure 1.3: I-TevI and Artificial Tev-mTALEN Nucleases	10
Figure 2.1: <i>In Vivo</i> Yeast Reporter Assays with the N169-T120 Tev-mTALEN Construct	13
Figure 2.2: Purification of the Tev-mTALEN Construct.....	17
Figure 2.3: Cleavage Assays with Varying Protein:DNA Ratios	18
Figure 2.4: Cleavage Assays with Varying Buffer Salt Concentrations.....	19
Figure 2.5: Cleavage Assays with Poly dI/dC as a Non-Specific Competitor.....	20
Figure 2.6: Mapping of Tev-mTALEN Top and Bottom Strand Nick Sites	21
Figure 3.1: DNA Spacer Single Nucleotide Substitution Assays	23
Figure 3.2: <i>In Vivo</i> Screen of Tev-mTALEN Randomized DNA Spacer Targets.....	25
Figure 3.3: Analysis of Sequenced N15 DNA Spacer Clones	26
Figure 3.4: Sequence LOGOs of Active and Inactive N15 DNA Spacer Clones.....	27
Figure 4.1: Spacer Single Nucleotide Substitution Assays with the S134G Tev-mTALEN Variant.....	37
Figure 4.2: Spacer Single Nucleotide Substitution Assays with the S134G/N140S Tev-mTALEN Variant	38
Figure 4.3: Spacer Single Nucleotide Substitution Assays with the K135R/N140S/Q158R Tev-mTALEN Variant	39

Figure 4.4: Spacer Single Nucleotide Substitution Assays with the V117F/D127G Tev-mTALEN Variant 40

Figure 4.5: Comparison of Wildtype and S134G Tev-mTALEN Activity 41

Figure 4.6: Comparison of Wildtype and S134G/N140S Tev-mTALEN Activity 42

Figure 4.7: Comparison of Wildtype and K135R/N140S/Q158R Tev-mTALEN Activity 43

Figure 4.8: Comparison of Wildtype and V117F/D127G Tev-mTALEN Activity..... 44

List of Appendices

Appendix A: Strains and Plasmids	70
Appendix B: Oligonucleotides.....	73
Appendix C: Supplementary Data Tables.....	77

List of Abbreviations

bp	base pair
CAP	catabolite activator protein
Cas	CRISPR-associated
CRISPR	clustered regularly interspaced short palindromic repeat
DNA	deoxyribonucleic acid
DTT	dithiothreitol
GPD	glyceraldehyde-3-phosphate dehydrogenase
gRNA	guide ribonucleic acid
KCl	potassium chloride
LB	Luria-Bertani (bacterial medium)
Mega-Tev	I-TevI meganuclease fusion
NaCl	sodium chloride
NHEJ	non-homologous end joining
NUMOD	nuclease associated modular domain
ONPG	<i>ortho</i> -nitrophenyl- β -galactoside
PAGE	polyacrylamide gel electrophoresis
PAM	protospacer-adjacent motif
PCR	polymerase chain reaction
PNK	polynucleotide kinase
Poly dI/dC	poly(deoxyinosinic-deoxycytidylic) acid
RNA	ribonucleic acid
rpm	rotations per minute
RVD	repeat variable di-residue
SDS	sodium dodecyl sulfate
TAL	transcription activator-like (effector)
TALEN	TAL effector nuclease
Tev-mTALEN	I-TevI/TAL effector nuclease
Tev-ZFE	I-TevI/zinc finger fusion nuclease
td	thymidylate synthase (gene)
Tris-HCl	tris(hydroxymethyl)aminomethane hydrochloride

YPD	yeast extract peptone dextrose (yeast medium)
ZF	zinc finger
ZFN	zinc finger nuclease

Chapter 1

1 Introduction

Genome engineering is the process of making specific, heritable modifications to the DNA of an organism (1, 2). Genome engineering techniques can be used to introduce novel DNA sequences, generate deletions, or make corrections to the DNA of an organism (1). Genome engineering has a broad range of academic, clinical, and industrial applications, but has been limited by the tools and methods available. The earliest genome engineering technique involved transformation of the target cell with a donor DNA template, relying on homologous recombination between the donor DNA and the target chromosomal DNA (3, 4, 5). This technique was first demonstrated in yeast, when a functional LEU2 gene was restored to a leu2- strain through transformation and integration of a LEU2-containing plasmid (3). Though efficient in yeast, this technique is extremely inefficient in mammalian cells (4, 6, 7, 8), which do not readily perform homology-directed repair outside of cell division. The low success rate of conventional recombination techniques is a limiting factor for genome engineering in mammalian cells. Engineered viruses (9, 10, 11, 12, 13, 14) and transposons (15, 16, 17) can also be used to insert novel DNA into a target genome; however, these methods are limited by the precision of their insertion and the ability to re-engineer their targeting specificity (18, 19). In 1997, two papers examining the effects of over-expressed Translation Elongation Factor 1 α on fruit fly longevity had to be retracted after it was shown that the transposon-delivered gene was not being expressed in the modified flies (18). Viral delivery has been successfully used to treat X-linked Severe Combined Immunodeficiency by restoring a functional IL2RG gene to patients with the disease (19); however, several of the patients subsequently developed leukemia as a result of integration of the virus near proto-oncogenes (20, 21, 22). Genome editing tools must therefore be both efficient, simple to engineer, and minimize the occurrence of potentially toxic off-target mutagenesis.

1.1 Site-Specific Nucleases for Precise Genome-Editing

Over the past two decades, site-specific nucleases have become a tool of choice for targeted genome editing. By directing nuclease-induced DNA damage to a specific target site, cellular DNA repair pathways can be stimulated, increasing the frequency of recombination or deletion events at that site (6, 23) (Figure 1.1). Sequence-specific DNA-binding allows nuclease activity to be directed to the desired target site while limiting off-target binding and cleavage. Minimizing off-target DNA cleavage is critical as it could potentially lead to undesired mutagenesis. The potential utility of site-specific nucleases was first demonstrated in the 1980s with the enzymes HO and I-SceI. These enzymes, members of the meganuclease family, were shown to efficiently stimulate recombination in several yeast genes when their respective recognition sites were integrated into the target genes (23, 24). These experiments provided the proof-of-principle for nuclease-mediated genome-editing by showing that targeted DNA damage could stimulate mutagenic repair several orders of magnitude more efficiently than unassisted recombination techniques (6, 23, 24). However, the complex nature of the overlapping DNA binding and cleavage activity of meganucleases makes them laborious to re-engineer for non-wildtype targets. The experiments performed with HO and I-SceI required prior integration of their wildtype target sites into the target genes in yeast. In order for nuclease-mediated genome editing to be feasible, enzymes must be easily re-engineered to target a wide variety of non-engineered, non-integrated target sequences. Efforts are being made to improve the breadth of meganuclease targeting through directed mutagenesis (25, 26), but difficulty in re-engineering target site specificity is still a limiting factor in their application.

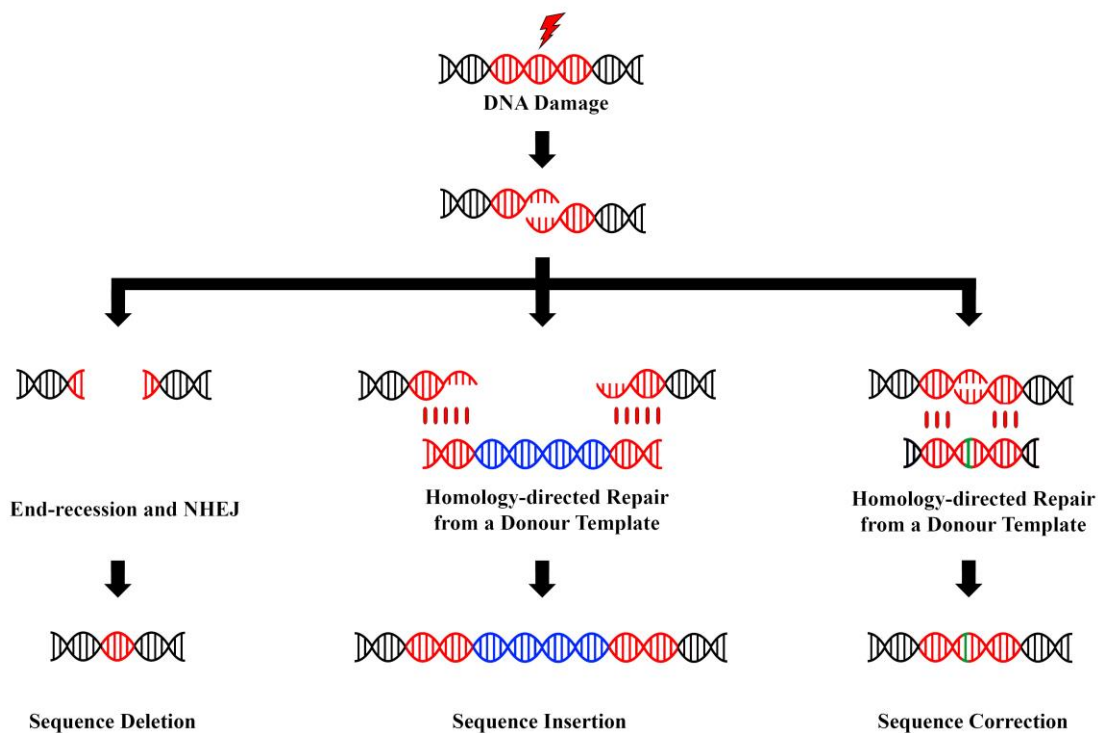


Figure 1.1: Nuclease-Mediated Genome Editing

Nuclease-mediated genome editing involves precisely targeting DNA damage in order to stimulate cellular repair pathways. End-resection followed by NHEJ (non-homologous end joining) will result in a loss of sequence information, ideal for targeted gene knockouts. If a donour template is provided homology-directed repair can lead to insertion of new or corrected sequence information.

The practical barrier to nuclease targeting potential was overcome with the development of engineered, site-specific nucleases (27). Engineered nucleases are proteins that combine the DNA-binding domain and nuclease domain of two different proteins, allowing nuclease activity to be directed to target sites specified by the DNA-binding domain. Engineered nucleases require components that are modular – able to retain their individual functions when fused to non-native domains. If a nuclease or DNA-binding domain is heavily influenced by neighboring domains, fusing it to a non-native domain will likely impair or alter its function in ways that are difficult to predict. The first family of broadly-targetable engineered nucleases were the Zinc Finger Nucleases (ZFNs). Described in 1996 (27), ZFNs combine the nuclease domain of the type II restriction enzyme, FokI, with a DNA-binding domain consisting of several tandem zinc finger (ZF) subunits (Figure 1.2A). Unlike the complex DNA-binding mechanics of the meganucleases, each zinc finger interacts with 3 base pairs (28). A DNA-binding domain with multiple zinc fingers will target a sequence that corresponds to the combined 3 base specificity of each subunit (Figure 1.2B and 1.2F)(29, 30). Because the activity of the FokI nuclease domain is modular, it remains active when fused to non-native DNA binding domains such as a zinc finger array (27, 31, 32). The FokI nuclease domain alone possesses no specific base requirements for DNA cleavage (33), allowing ZFNs to cleave any DNA substrate to which the ZF domain binds. FokI functions most efficiently as a dimer (34), necessitating the design of two ZFNs that position their respective FokI domains at the target (Figure 1.2F)(35), leading to dimerization and subsequent cleavage. Despite the simplicity of their design, ZFN targeting is complicated by the fact that the base-specificity of individual zinc finger subunits is not entirely modular. Interactions between adjacent zinc finger subunits frequently leads individual subunits to take on new base-specificities, resulting in the failure of many ZFN pairs to effectively target their intended substrate (36, 37). Of further concern is the strictness of zinc finger specificity – zinc finger arrays are tolerant of mismatches in their target sites (1, 38). Efforts are underway to predict context-dependencies (39) and alternate methods of assembling zinc finger arrays have been developed in order to minimize the occurrence of off-target breaks (40). In spite of the complications in ZFN targeting, clinical trials

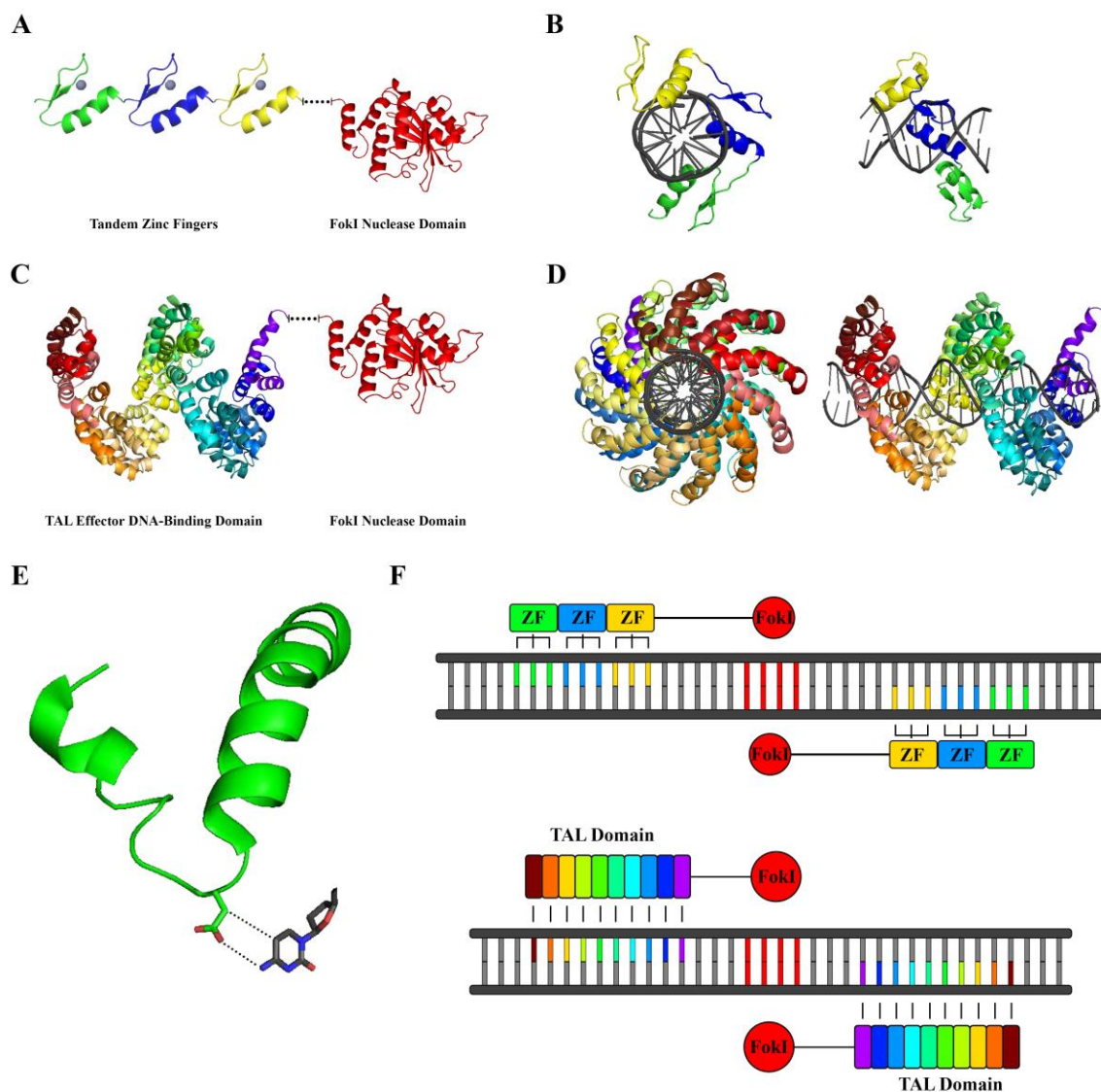


Figure 1.2: FokI-based Engineered Nucleases

(A) Structural layout of a ZFN. (B) Crystal structure of three adjacent ZF domains interacting with a DNA target. (C) Structural layout of a TALEN. (D) Crystal structure of the central DNA-binding domain of TAL effector PthXo1 interacting with its DNA target, each RVD-containing repeat is coloured separately. (E) Example of a cytosine-binding TAL effector repeat with RVD sequence HD. Residue D13 is shown engaging in a hydrogen bond and a van der waals interaction with the cytosine base. (F) Basic schematics of dimeric ZFN and TALEN targeting. (A-E) Crystal structures were adapted from Mak et al (41), Elrod-Erickson et al (42), and Wah et al (43).

are already underway to test the efficacy of ZFNs in generating HIV-resistant stem cells (44).

Another family of FokI-based engineered nucleases known as TAL effector nucleases (TALENs) have gained attention as genome-editing reagents (45). TALENs replace the zinc finger domain of ZFNs with a DNA-binding domain from virulence proteins known as Transcription Activator-Like (TAL) effectors (Figure 1.2C) (45). Originally discovered in the bacterial plant pathogen, *Xanthomonas*, TAL effectors are secreted into the host cell and localized to the nucleus, where they bind specific plant promoters, activating expression of genes that increase susceptibility to infection (46, 47). A TAL effector DNA-binding domain consists of a series of 33-35 amino acid DNA-binding helix-loop-helix repeats, each of which interacts with a single target DNA base (Figure 1.2D)(41, 47, 48). The repeats are largely identical, with the exception of the 12th and 13th amino acids located in the inter-helical loop. These variable amino acids, termed the repeat variable di-residue (RVD), target one DNA base to determine the specificity of each TAL repeat (41, 48, 49). Although many RVDs occur naturally, artificial TAL domains are typically assembled using RVDs with the amino acid sequence HD to target cytosine, NG to target thymine, NI to target adenine, and NN for targeting guanine (41, 45, 48, 49). Figure 1.2E shows the interaction of an HD RVD repeat with a target cytosine. This one-to-one correspondence of TAL effector RVD repeats to DNA base makes the theoretical basis of TALEN targeting extremely simple, and functional TALENs can be designed to target sequences as long as 30 nucleotides (45, 51). Conventional TAL effectors typically require a thymidine immediately 5' to the target sequence (49, 52); however, the discovery of non-*xanthomonas* TAL effectors as well as mutation of certain amino acids in the N-terminal region of the protein have led to novel TAL domains which do not have this requirement (52, 53, 54). Like ZFNs, TALENs utilize the FokI nuclease domain to catalyze DNA hydrolysis, necessitating the design of two TALEN monomers that will align their active sites over the desired target (Figure 1.2F)(45).

The most recently developed family of site-specific nucleases to be used for genome editing are the clustered regularly interspaced short palindromic (CRISPR)-associated

(Cas) nucleases (55). The CRISPR-Cas system is a form of adaptive bacterial immune system that protects against invasive foreign DNA by recognizing and cleaving specific sequences (56). Unlike ZFNs and TALENS, which achieve sequence-specificity through protein domains, the CRISPR-Cas system uses a guide RNA (gRNA) with 20 base pairs of complementarity to the target sequence (56). The gRNA binds to the complementary target sequence, stimulating cleavage by the Cas9 nuclease. One significant constraint on CRISPR-Cas targeting is the requirement for a short motif (NGG), known as a protospacer adjacent motif (PAM), located immediately 3' to the target sequence (55, 57, 58). The PAM sequence must be located adjacent to the target sequence in order to facilitate Cas9 nuclease activity. A further constraint on CRISPR-Cas targeting is the need for a G or GG at the 5' end of the target site (58). This is not a requirement of CRISPR-Cas biology but instead the promoters used to transcribe gRNAs (58). The CRISPR-Cas system has generated great interest as a genome-editing system because of the incredibly facile nature of gRNA targeting. As long as the 3' PAM and 5' G/GG are present, nearly any sequence can be targeted if a complementary gRNA is designed. As promising as the CRISPR system appears, there are concerns regarding the systems proneness to off-target cleavage (59, 60, 61, 62, 63). Studies in human cells have shown that the CRISPR-Cas system can generate a substantial degree of off-target mutagenesis, with many sites differing by up to 5 bases from the gRNA being readily bound and cleaved (59). In spite of this, promising efforts have been made to reduce off-target mutagenesis by the CRISPR-Cas system, including the use of paired CRISPR nickases (64), or truncation of the complementary portion of the gRNA (65). The unmatched simplicity of the CRISPR-Cas system has quickly made it a popular genome editing tool.

1.2 I-TevI as an Alternative to the FokI Nuclease

Each family of genome-editing nucleases possesses advantageous and disadvantageous properties in accordance with their biology. One potential drawback of conventional ZFNs and TALENs is the non-specific activity of the dimeric FokI nuclease domain. While the lack of sequence-specificity means the nuclease domain does not impose targeting constraints, it also presents the risk of nuclease activity at off-target sites. ZFNs and TALENs are designed to cleave when two monomers bind their separate target sites

and align their FokI domains to form a heterodimeric complex (Figure 1.2F); however, a single DNA-bound FokI monomer can recruit a second monomer from solution and form a functioning homodimer or heterodimer (66, 67, 68, 69). These can lead to potentially toxic cleavage at off-target sites where only a single protein is bound to the DNA.

Mutations can be made to the FokI dimerization interface to reduce the occurrence of homodimers (66, 67, 68, 69), but these may reduce the effectiveness of the enzyme (70) and do not eliminate the occurrence of heterodimers. Recent work by the Edgell lab has shown that the common, FokI-based nucleases prevalent in genome-editing literature can be replaced by monomeric, cleavage site-specific alternatives (71, 78). These enzymes replace the commonly used FokI with the nuclease domain of the protein I-TevI. I-TevI, found in phage T4, is a group I intron-encoded GIY-YIG homing endonuclease (72, 73). The GIY-YIG homing endonucleases are named for their characteristic ~100 amino acid nuclease domain containing the “GIY” and “YIG” consensus motifs (72). In phage T4, I-TevI binds to and cleaves a specific sequence in the thymidylate synthase (*td*) gene (74). Through this process, I-TevI mediates the invasion of its encoding intron into the *td* gene (74, 75). I-TevI consists of three domains – an N-terminal GIY-YIG nuclease domain, a central linker domain containing an atypical zinc finger motif, and a C-terminal DNA-binding domain (76) (Figure 1.3A). The I-TevI nuclease domain cuts the DNA substrate at a defined CA₇AC[↓]G motif upstream of the DNA-binding site (arrows indicate the sites of lower and upper strand cleavage)(74, 77). Analyses with mutant DNA substrates have shown that both the 5' C and 3' G residues of the cleavage motif are essential for efficient I-TevI activity (77, 78, 79). Substitutions within the central 3 nucleotides of the cleavage motif are generally tolerated, with a few exceptions (78, 79). The zinc finger-containing linker domain wraps around the minor groove of the DNA spacer sequence, located between the CNNNG cleavage motif and the binding site (80). Through this interaction, the linker positions the nuclease domain over the cleavage motif, facilitating I-TevI activity. The zinc finger of the linker domain acts as a molecular ruler, allowing I-TevI to discriminate between CNNNG motifs based on their distance from the binding site (76, 81, 82, 83, 84). The C-terminal portion of I-TevI includes an α -helix and a helix-turn-helix motif and acts as the primary DNA-binding domain (80, 81). Similar to FokI, the DNA-binding and nuclease activity of I-TevI are physically separate (71, 78,

79, 85), allowing the nuclease domain to be fused to novel DNA-binding domains and retain its activity (71, 78, 79). However, I-TevI possesses two potentially advantageous properties compared to FokI. First, I-TevI functions as a monomer, meaning that only a single artificial nuclease has to be designed for each targeted sequence (71, 78, 79, 86). Second, is the requirement for an appropriately positioned CNNNG cleavage motif (71, 77, 78, 79). The requirement for the short cleavage motif offers a compromise between the lengthy recognition sites of the meganucleases and the completely non-specific activity of the FokI nuclease domain. A small degree of nuclease sequence-specificity should not be a major targeting constraint, but will reduce the likelihood of off-target cleavage when compared to the non-specific FokI domain.

1.3 I-TevI-based Engineered Nucleases

The genome-editing potential of the I-TevI nuclease domain was first demonstrated with the development of Tev-zinc finger endonucleases (Tev-ZFEs)(71) and Tev-meganuclease fusions (Mega-Tevs)(71, 79). These enzymes consist of the nuclease and linker domains of I-TevI fused to a C-terminal zinc finger domain or a catalytically inactive meganuclease, respectively (71, 79). Both families of enzymes have target site-specific activity comparable to the conventional FokI-based nucleases (71, 79). Activity assays with mutant DNA substrates show that these I-TevI-based nucleases require both a compatible DNA-binding sequence and an appropriately positioned CNNNG motif for efficient cleavage – indicating that I-TevI confers an additional level of target site-discrimination which FokI does not. Following the development of Tev-ZFEs and Mega-Tevs, it was shown that the I-TevI nuclease was also compatible with TAL effector DNA-binding domains (78, 87). These enzymes (referred to as Tev-mTALENs in this report) combine the N-terminal nuclease and linker domains of I-TevI with a C-terminal TAL effector DNA-binding domain (Figure 1.3B). Tev-mTALENs combine the monomeric, site-specific activity of I-TevI with the simple, versatile targeting of TAL effector domains.

Tev-mTALEN targeting involves three distinct interactions between the modular domains of the enzyme and their corresponding DNA targets – the interaction of the I-TevI nuclease domain with the CNNNG cleavage motif, the interaction of the central I-

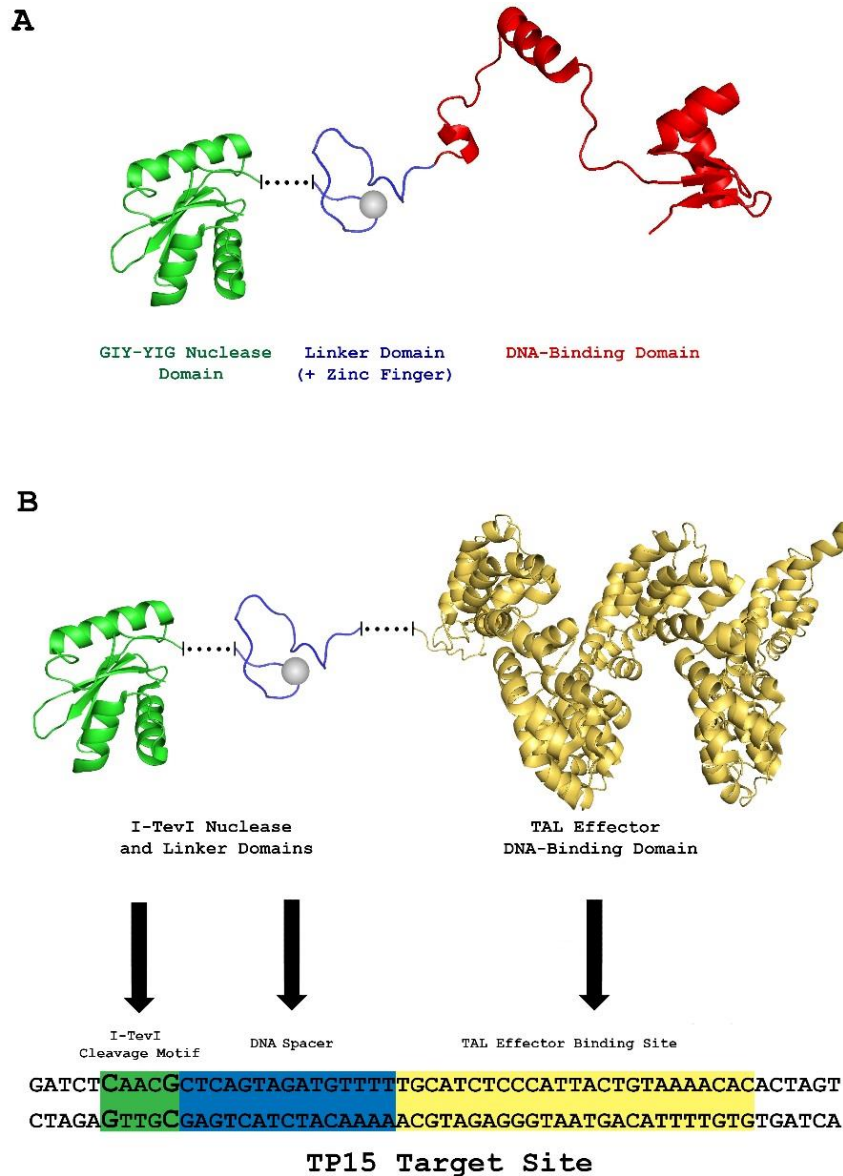


Figure 1.3: I-TevI and Artificial Tev-mTALEN Nucleases

(A) Structure of the GIY-YIG homing endonuclease, I-TevI. Two separate crystal structures of I-TevI (80, 88) have been combined to show the basic structural layout of I-TevI when it is bound to its substrate. (B) Tev(N169)-PthXo1(T120), an example of a Tev-mTALEN nuclease. Partial crystal structures of I-TevI (80, 88) and the TAL effector PthXo1 (41) have been combined to show the modular structure of a Tev-mTALEN nuclease. Shown below is the wildtype TP15 target site. Each of the three domains in the Tev-mTALEN interacts with a corresponding portion of the target site.

TevI linker domain with the DNA spacer, and the interaction of the C-terminal TAL effector domain with the TAL binding site (Figure 1.3B). The interactions of the TAL domain and the I-TevI nuclease domains with their targets are well characterized, but the interaction of the linker domain with the DNA spacer is poorly understood. Previous work with native I-TevI demonstrated that many target sites with non-wildtype DNA spacer sequences are not cleaved efficiently; however, no single DNA spacer nucleotide was identified as critical for activity (89). Jason Wolfs of the Edgell Lab has generated similar results with Mega-Tevs. Based on these data, it is likely that the DNA spacers of potential Tev-mTALEN target sites will not support efficient cleavage.

1.4 Hypothesis

Without a predictive model of DNA spacer compatibility, selecting robust Tev-mTALEN target sites will be imprecise. I hypothesized that the I-TevI linker domain has preferences for certain nucleotides in the DNA spacer, and that these preferences are important for Tev-mTALEN activity. To identify any Tev-mTALEN preferences for specific nucleotides in the DNA spacer of the target site, I performed assays using a Tev-mTALEN nuclease referred to as the N169-T120 construct. *In vitro* and *in vivo* assays were performed to confirm the modular function of the I-TevI nuclease, probe the sequence requirements of the I-TevI linker domain for its spacer DNA target, and explore mutations to the I-TevI linker as a way of broadening Tev-mTALEN targeting potential.

Chapter 2

2 The I-TevI Nuclease Domain Retains its Activity in a Modular Fashion when Fused to a TAL Effector Domain

In order for the I-TevI nuclease domain to be broadly applicable for genome-editing applications, it must retain activity when fused to non-native DNA-binding domains. This section describes experiments that assess whether the I-TevI nuclease domain retains its **CAACG** site-specific activity when fused to the non-native TAL binding domain.

2.1 Tev-mTALENs Retain the Cleavage Motif Requirements of Native I-TevI Enzyme

To determine if the I-TevI nuclease domain retains its function when fused to a TAL effector domain, *in vivo* yeast reporter assays were performed to measure the activity of the N169-T120 Tev-mTALEN construct on several control substrates. The N169-T120 construct (Figure 1.3B) is named for its two components, the N-terminal 169 amino acids of I-TevI (ending in asparagine 169), and the TAL effector PthXo1 lacking residues 1-119 of the N-terminus and residues 1319-1373 of the C-terminus, such that the TAL domain begins at threonine 120 and ends at proline 1318. The N169 fragment of I-TevI was selected because it comprises a minimal functional portion of I-TevI that excludes all known amino acids that make base-discriminant contacts to DNA, while still including the important distance-determining zinc finger of the linker domain. The T120/P1318 truncation of PthXo1 contains the essential RVD-containing DNA-binding domain and nuclear localization signals, while excluding the N-terminal type III secretion signal and C-terminal activation domain of the native protein. The N169-T120 construct activity was tested on a DNA target referred to as the TP15 (Figure 1.3B). This target site consists of, in the 5' to 3' direction, - the wildtype **CAACG** I-TevI cleavage motif, 15 nucleotides of the *td* DNA spacer sequence, and the binding sequence of the PthXo1 TAL effector. Work performed previously by Ben Kleinstiver showed that 15 nucleotides is the optimal DNA spacer length for N169-T120 construct activity (78). The yeast β -

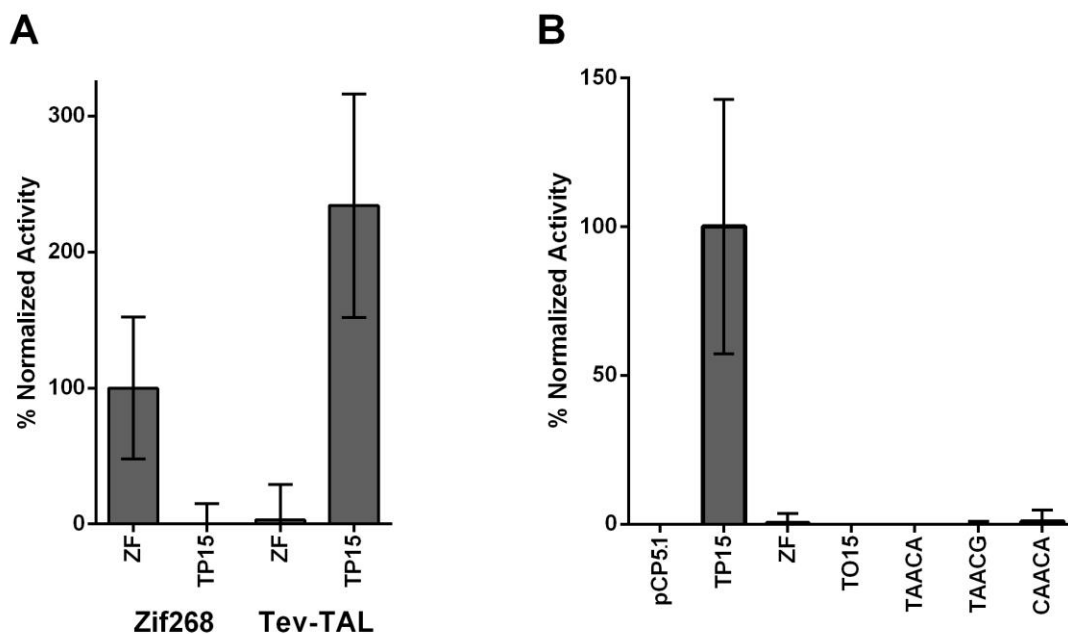


Figure 2.1: *In Vivo* Yeast Reporter Assays with the N169-T120 Tev-mTALEN Construct

Yeast β -Galactosidase reporter assays were performed using the N169-T120 Tev-TAL construct and zinc finger nuclease, Zif268. Activity values are measured in Miller units normalized to the activity of the Tev-TAL construct on the TP15 substrate. **(A)** Control assays with the N169-T120 and Zif268 constructs. Each nuclease was screened against the TP15 and ZF target sites. **(B)** Tev-TAL target site discrimination assays. Tev-TAL activity was measured against the wildtype TP15 target and three TP15 variants with one or both of the critical cleavage motif nucleotides mutated (**TAACA/TAACG/CAACA**). Activity was also measured against the empty target vector (pCP5.1), Zif268 target (ZF), and mega-Tev target site (TO15). **(A + B)** All assays were performed using three biological replicates, each with 3 technical replicates. Error bars indicate standard deviations.

Galactosidase assay reports on nuclease activity through repair of a *LacZ* gene interrupted by the nuclease target site. The *LacZ* gene is also partially duplicated, so cleavage of the target site by the nuclease stimulates single-strand annealing DNA repair, restoring a functional *LacZ* gene that is measured by β -Galactosidase degradation of the colorimetric substrate, ONPG. Initial β -Galactosidase reporter assays were performed with the N169-T120 construct and control Zif268. The Zif268 enzyme is a dimeric FokI-based ZFN. Activity measurements were normalized to either the activity of the Zif268 construct on its ZF substrate (Figure 2.1A) or the Tev-mTALEN construct on the TP15 substrate (figure 2.1B). Zif268 and N169-T120 activity were each measured against the ZF and the TP15 target sites (Figure 2.1A). Against the TP15 substrate, the Tev-mTALEN construct cleaved with an efficiency greater than that of the ZFN control on its respective target site. This confirms that the I-TevI nuclease domain remains functional in the presence of the non-native TAL effector domain and, importantly, with activity comparable to that of the commonly used FokI nuclease. To determine if the Tev-mTALEN retains the requirement for the CAACG cleavage motif, activity was measured on three TP15 target sites with mutated cleavage motifs – TAACA, CAACA, and TAACG (Figure 2.1B). Tev-mTALEN activity on all three cleavage site mutants was at background levels. These results confirm that in addition to being active *in vivo*, Tev-mTALENs possess the same strict cleavage site requirements of the native I-TevI enzyme.

Targeting of Tev-mTALENs should be determined by the TAL DNA binding domain. To determine if the I-TevI nuclease and linker domains affect DNA targeting, Tev-mTALEN activity was measured on the TO15 substrate – a target site that replaces the PthXo1 TAL binding site with that of I-OnuI, a meganuclease. The PthXo1 TAL site and the I-OnuI site are different lengths (25bp versus 22bp) and share 47.06% identity. The TO15 substrate retains the CAACG cleavage motif and spacer DNA of the TP15 substrate. Activity on the TO15 substrate was at background levels, indicating that the I-TevI nuclease and linker domains do not influence targeting by the TAL domain (Figure 2.1B).

2.2 Tev-mTALENs Nick the Target DNA at the Same Top and Bottom Strand Sites as the Native I-TevI Enzyme

In vitro cleavage site-mapping was performed to determine if the Tev-mTALEN construct nicks the CAACG cleavage site at the same bottom and top strand positions as native I-TevI. The N169-T120 was His-tagged at the C-terminal end, over-expressed in *E. coli*, and purified using Ni²⁺ affinity and size exclusion chromatography. The identity of the most prominent polypeptide in the eluted sample was confirmed by MALDI analysis (Figure 2.2). Cleavage assays were performed to determine if the construct is active *in vitro*. Purified N169-T120 protein was incubated with plasmid pSP72 containing the TP15 target site, as well as empty pSP72. The reactions were resolved on agarose gel and the extent of plasmid linearization was compared between the two samples. Varying reaction conditions were tested in order to maximize cleavage of the TP15 target plasmid while minimizing cleavage of the empty plasmid. Reactions were initially performed in standard Tev-mTALEN reaction buffer (Chapter 6.3, Materials and Methods) with protein-substrate ratios of 1:1 and 2:1 (Figure 2.3). Reactions were incubated at 37⁰C for 20 minutes and samples were taken at 3, 7, 15, and 20 minutes (Figure 2.3). With 2-fold excess protein, the majority of the TP15 was linearized (88%) after 20 minutes, however, the empty pSP72 vector was also linearized to 17%. Using a 2-fold excess of protein and a 20 minute reaction time, cleavage assays were then performed with salt levels ranging from 25mM-150mM and with KCl or the standard NaCl of NEBuffer2 (Figure 2.4). Plasmid linearization was greatest at lower salt concentrations, with a sharp decrease in activity occurring from 100mM to 150mM. Compared to NaCl buffers, KCl buffers increased TP15 linearization to 100% for all but the 150mM buffer, however pSP72 linearization was also increased substantially (nearly 75% with 25mM KCl buffer). None of the salt variant buffers improved on the activity of standard NEBuffer 2 (50mM NaCl). In order to reduce promiscuous cleavage, cleavage reactions were performed with the non-specific DNA competitor substrate, poly dI/dC (Figure 2.5). 20 minute reactions were performed with 2-fold molar excess of protein with or without 20ng/μl poly dI/dC. Addition of poly dI/dC reduced TP15 linearization (77% vs 93%) but also reduced activity on the empty pSP72 vector (1% vs 25%). Cleavage reactions were performed using poly dI/dC as shown in Figure 2.5 and

the linearized TP15 product was isolated via gel extraction and sent for run-off sequencing using bottom and top strand mapping primers (Figure 2.6A). The ABI traces from the sequencing reaction of the linearized TP15 product (Figure 2.6B) show that the N169-T120 Tev-mTALEN nicks each strand of the cleavage motif at the same positions (CA \uparrow AC \downarrow G) as the wild-type I-TevI on its cognate DNA substrate.

2.3 Summary

In vitro and *in vivo* assay data show that the I-TevI nuclease domain retains its function when fused to a TAL DNA-binding domain. In the yeast-based assay, Tev-mTALEN activity was comparable to that of the dimeric FokI-ZFN, suggesting that Tev-mTALENs can achieve cleavage efficiencies comparable to the more commonly used FokI-based engineered nucleases. Yeast reporter assays and cleavage mapping have also shown that the I-TevI nuclease domain retains the strict CAACG site-specificity of native I-TevI. These data indicate that the I-TevI nuclease domain (along with the ancillary linker) is modular in function - making Tev-mTALENs a viable alternative to FokI-based TALENs.

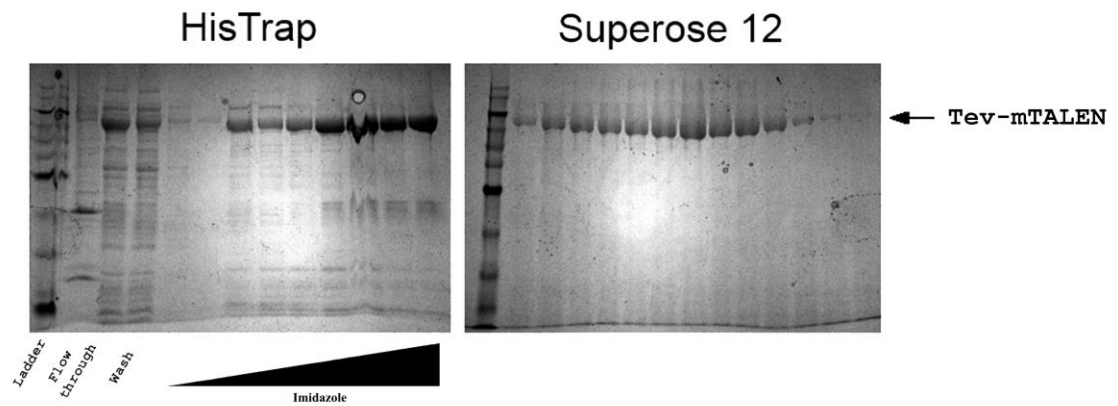


Figure 2.2: Purification of the Tev-mTALEN Construct

2-step purification of the Tev-mTALEN construct. Cell lysate was first run over a Ni^{2+} column. The purest samples were then eluted over a Superose 12 size exclusion gel column. Identity of the Tev-mTALEN construct was confirmed by MALDI.

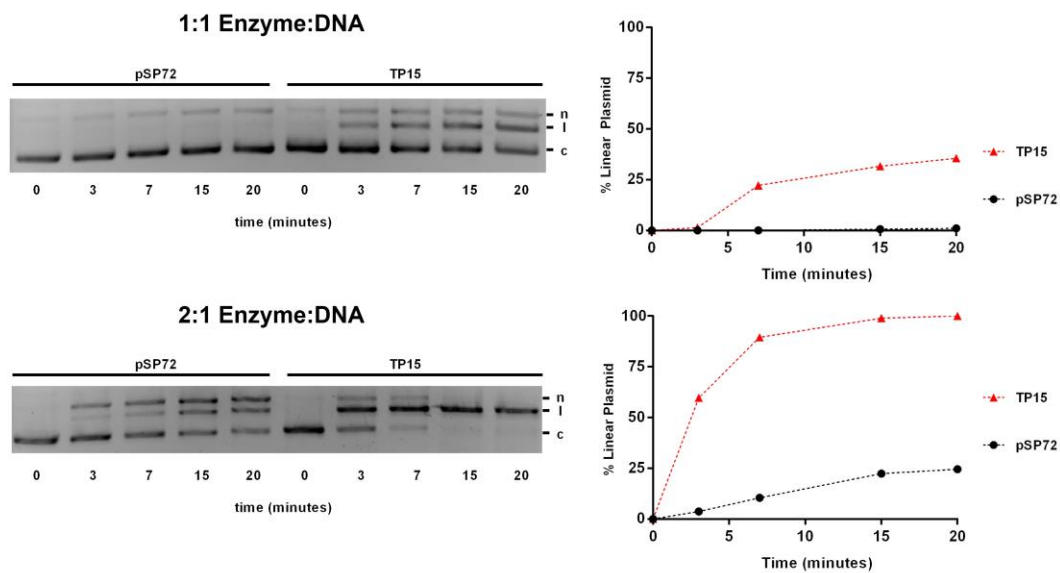


Figure 2.3: Cleavage Assays with Varying Protein:DNA Ratios

Timecourse cleavage assays were performed using equimolar or 2-fold molar excess of enzyme to DNA. Samples were taken and stopped at the time points indicated along the bottom. Graphs show the percentage of plasmid linearized, as measured by gel imaging software.

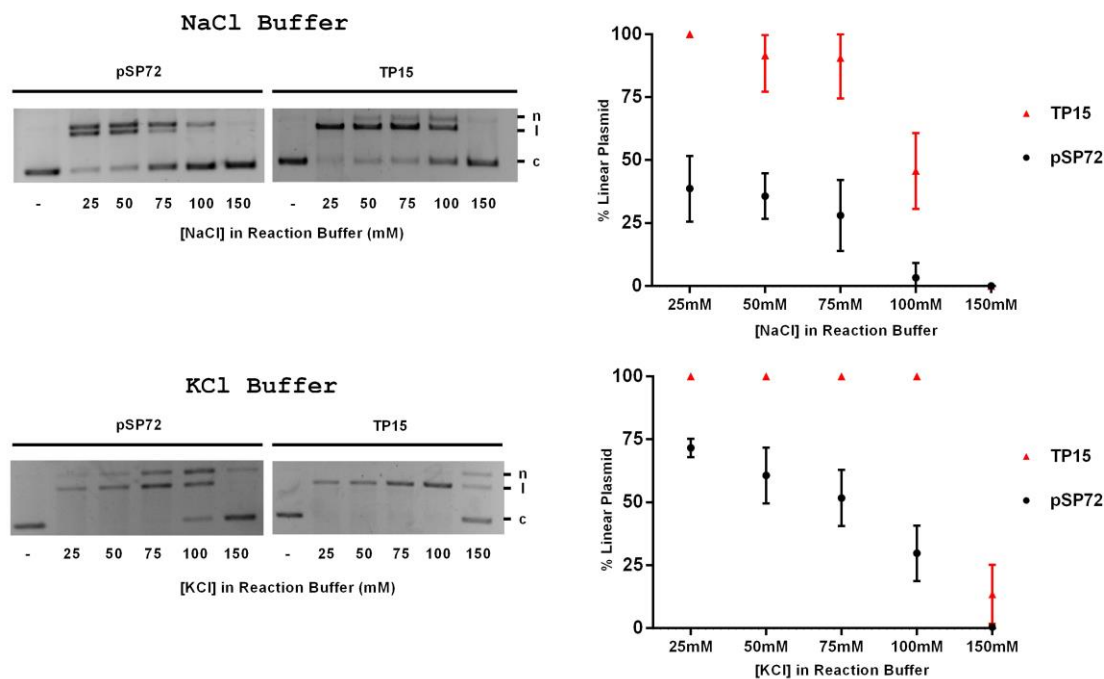


Figure 2.4: Cleavage Assays with Varying Buffer Salt Concentrations

20 minute endpoint cleavage assays were performed using varying concentrations of either NaCl or KCl in the reaction buffer. Graphs show the average and standard deviation of 3 replicate reactions. Graphs show the percentage of plasmid linearized, as measured by gel imaging software.

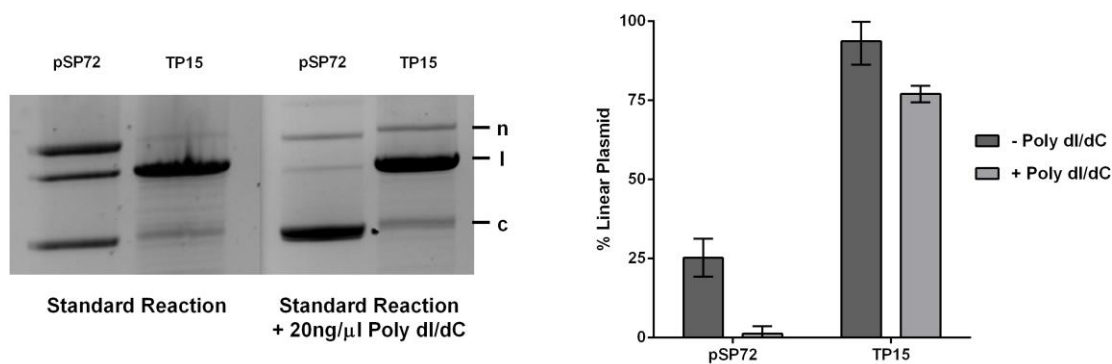


Figure 2.5: Cleavage Assays with Poly dI/dC as a Non-Specific Competitor

20 minute endpoint cleavage assays with 2-fold molar excess of protein to DNA and 50mM NaCl buffer in the absence or presence of the non-specific DNA substrate, polydI/dC. Graphs show the average and standard deviation of 3 replicate reactions. Graphs show the percentage of plasmid linearized, as measured by gel imaging software.

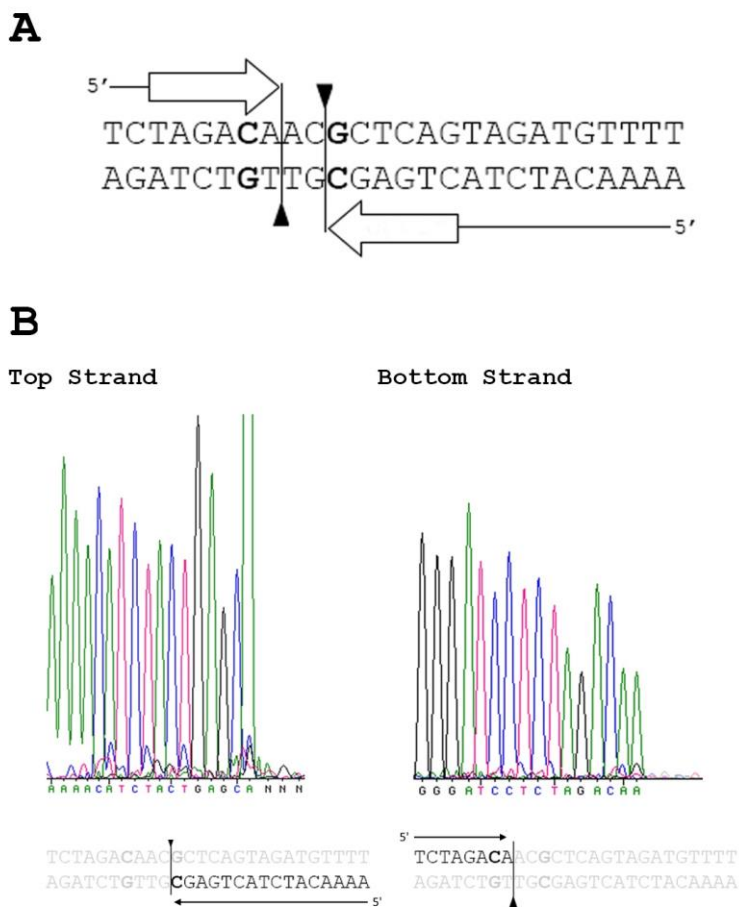


Figure 2.6: Mapping of T_{ev}-mTALEN Top and Bottom Strand Nick Sites

(A) Outline of the cleavage mapping process. Bottom strand cleavage is mapped by synthesizing the top strand, which terminates at the nucleotide opposite the bottom strand nick site. Top strand cleavage is mapped by synthesizing the bottom strand, up to the nucleotide opposite the site of top strand nicking. (B) Sequencing traces from cleavage mapping. ABI traces indicate that the sites of lower and upper strand nicking match the pattern of the wildtype I-TevI enzyme ($CA_{\uparrow}AC_{\downarrow}G$). Note that the additional A residue at the 3' end of each strand is a product of Taq polymerase extension.

Chapter 3

3 Tev-mTALENs are Sensitive to the Identity of Several Nucleotides in the DNA Spacer

Interaction of the I-TevI linker with the target DNA spacer is critical for Tev-mTALEN activity, and many DNA spacer sequences do not promote efficient cleavage (78). An understanding of the nucleotide preferences of the I-TevI linker for its DNA spacer will allow for more accurate prediction of whether or not a putative target site can be cleaved efficiently. To identify nucleotide positions in the DNA spacer that are important for Tev-mTALEN target site cleavage, I examined the effects of single base substitutions in the wildtype DNA spacer on Tev-mTALEN activity. I also screened a library of Tev-mTALEN target sites with a fully randomized DNA spacer sequence in order to enrich well-cleaved targets and identify trends in their nucleotide sequences.

3.1 Effects of Single Nucleotide Substitutions in the DNA Spacer on Tev-mTALEN Activity

To determine what positions in the DNA spacer sequence are important for cleavage activity, yeast reporter assays were performed to measure the activity of the Tev-mTALEN against each of the 45 possible single-nucleotide DNA spacer mutations of the TP15 target site (Figure 3.1). Nuclease activity on each of the mutant substrates was compared to that of the TP15 target in order to determine the mutations that impaired Tev-mTALEN activity. Assay results show that the Tev-mTALEN is sensitive to mutations at several positions in the DNA spacer. At position C1, substitution of A or T reduced activity to 14% and 33% respectively. Substitution of a G increased activity to an average of 258%. At position T2, substitution of any other base reduced activity, though an A was tolerated to a greater degree than a C or G. Activity on the T2A substrate was 19%, while the T2C and T2G were cleaved at 6% and <1% respectively. At position C3, substitution of T consistently reduced activity to ~5%. The C3A and C3G mutants were cleaved with 57% and 71% efficiency respectively. Substitution of G5 to an A reduced activity by 50%. At position T6 substitution of a C or G reduced activity to 12% and substitution of an A reduced activity to 7%. At position G8, substitution of A,

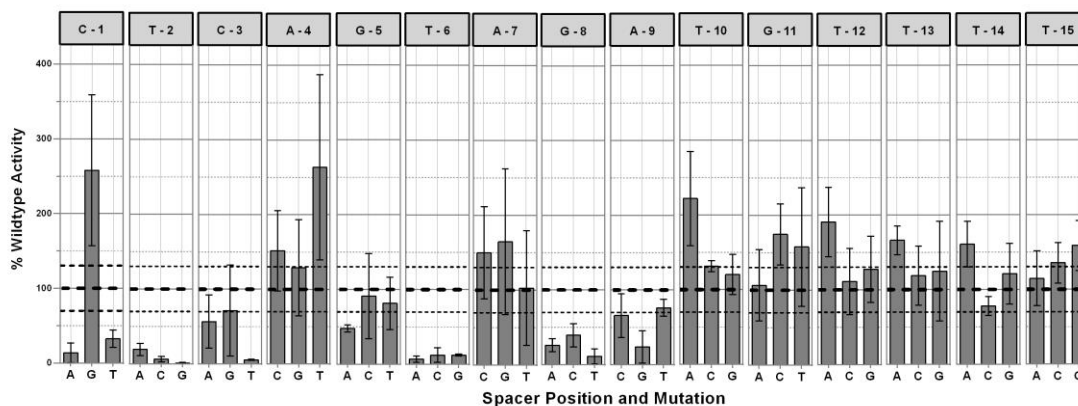


Figure 3.1: DNA Spacer Single Nucleotide Substitution Assays

Yeast B-Galactosidase reporter assays were performed to measure Tev-TAL activity on each of the 45 possible single nucleotide TP15 spacer mutants. Activity measurements are normalized to the activity of the N169-T120 construct on the TP15 substrate. The wildtype nucleotide and spacer positions (with 1 being directly adjacent to the cleavage motif) are shown along the top, with nucleotide substitutions indicated along the bottom axis. Average activity and error bars are based on three biological replicates each with 2 technical replicates. The thick dashed line indicates the average activity of the Tev-mTALEN on the TP15, with small dashes indicating 1 standard deviation.

C, and T reduced activity to 25%, 39%, and 11% respectively. At position A9, substitution of G reduced activity to 24%. The A9C and A9T substrates were cleaved with 66% and 76% efficiency respectively. Activity on the T14C substrate was 77%. All other single base substitutions were tolerated. Several mutant substrates were cleaved more efficiently than the TP15 – most noticeably the C1G and A4T, which were cleaved at 258% and 263% respectively. These data show that the identity of several nucleotides, primarily within the first 9 positions of the DNA spacer sequence, are important for Tev-mTALEN activity.

3.2 In Vivo Screen of Tev-mTALEN Target Sites with a Randomized DNA Spacer Sequence

While the assays performed on the DNA spacer single mutants provide insight into the nucleotide preferences of the linker, the effects of individual substitutions were examined in the context of a DNA spacer that had an otherwise identical sequence to that of the TP15 target. To determine if nucleotide context in the DNA spacer influences cleavage activity, I constructed a library of TP15 target site variants in which the DNA spacer sequence was completely randomized. The DNA spacer library, referred to as the N15, was transformed into yeast. Individual yeast clones were grown in 96 well plate format to isolate a single DNA spacer sequence per well. Clones harbouring the target site plasmid were mated with the yeast strain expressing the Tev-mTALEN, and β -Galactosidase reporter assays were performed in triplicate on a total of 753 clones from the random library (Figure 3.2A and B). Each 96 well plate included a well with the TP15 target and a well with the Zif268 target to act as positive and negative controls respectively. In each plate replicate, activity of the Tev-mTALEN on each of the N15 clones was normalized to the positive control for the plate. PCR was performed to amplify the target sequence from 62 non-active and 50 active clones, and the nucleotide content in the randomized DNA spacer region was analyzed by DNA sequencing (Figure 3.2C). N15 clones were considered active if the average activity of 3 replicates was not less than 2 standard deviations below the activity of the TP15 target. Figure 3.3A shows the relative nucleotide frequency at each position of the DNA spacer for the sequenced

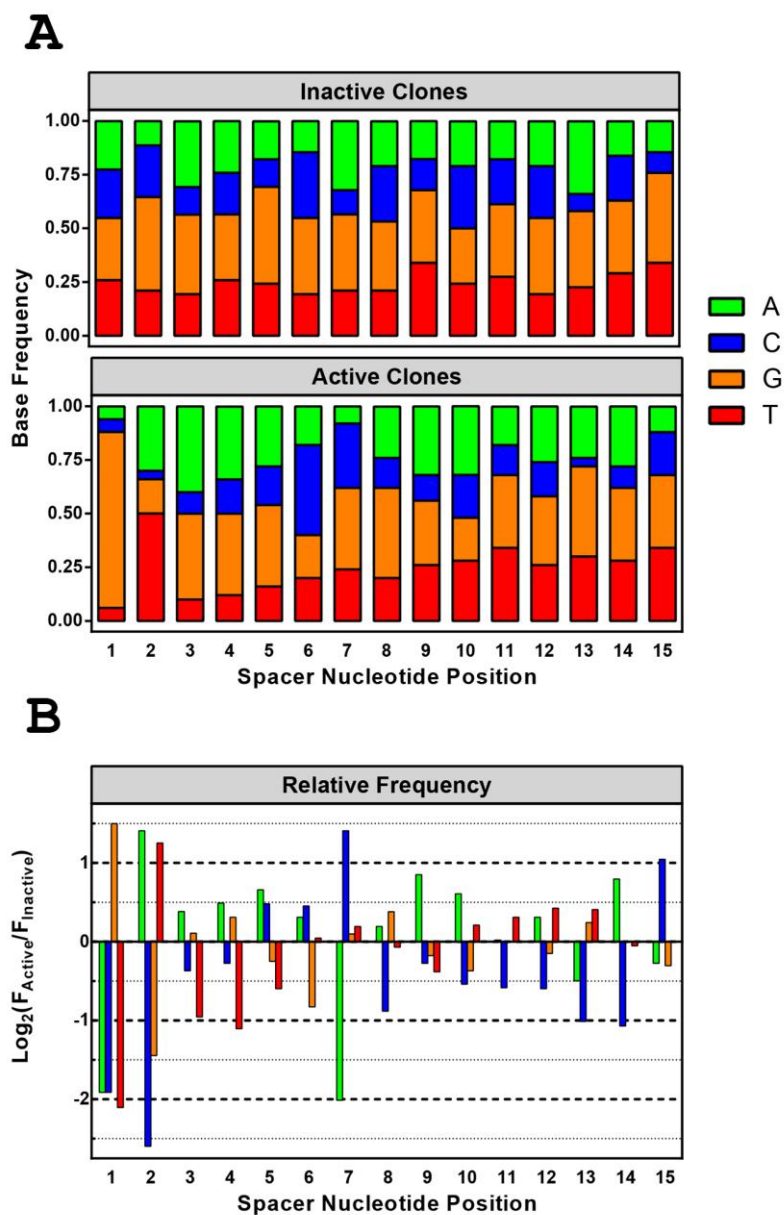


Figure 3.3: Analysis of Sequenced N15 DNA Spacer Clones

(A) Nucleotide frequencies at each DNA spacer position in the sequenced inactive and active N15 clones. (B) Comparison of the nucleotide frequencies for active and inactive N15 clones. Differences in nucleotide frequency are measured by as the ratio of the nucleotide frequency in the active clones over the nucleotide frequency in the inactive clones, converted to \log_2 scale.

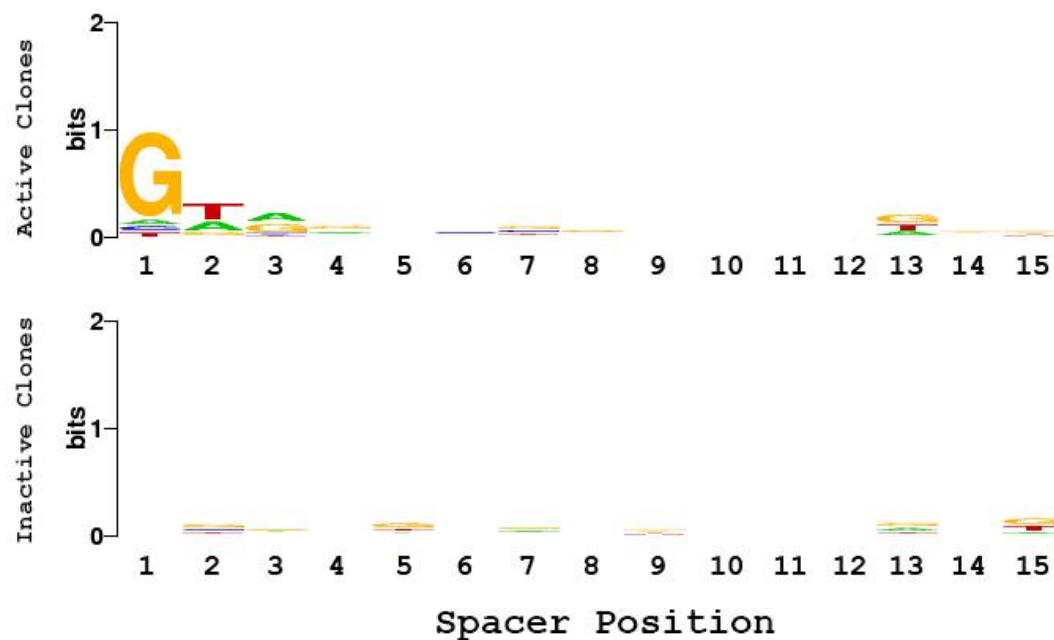


Figure 3.4: Sequence LOGOs of Active and Inactive N15 DNA Spacer Clones

Sequence LOGOs were generated for the sets of active and inactive clones identified in the random library screens. LOGOs were generated using the WebLogo tool provided by the Computational Genomics Research Group at the University of California, Berkeley (90). Values along the Y axis measure information content of the DNA spacer sequences.

active and inactive clones. Figure 3.3B shows a comparison of the nucleotide frequencies at each position between the active and inactive clones. Differences in nucleotide frequency are expressed as the \log_2 of the ratio of the frequency in the active clones over the frequency in the inactive clones. A value of 0 indicates that the nucleotide occurs at a specific position with equal frequency in the active and inactive pools. Positive values indicate that the nucleotide occurs more frequently in the active clones, while negative values indicate that a nucleotide occurs less frequently in the active clones. Comparison of the nucleotide frequencies between the two groups showed preferences for and against certain nucleotides at several positions along the DNA spacer. Several of the nucleotide preferences observed agree with those identified in the DNA spacer single nucleotide substitution assays (Figure 3.1). At position 1, active clones showed a 3-fold enrichment of G and 4-5 fold less A, T, and C. The observed preference for a G nucleotide is consistent with the spacer single substitution assays, in which the C1G was consistently cleaved more efficiently than the TP15 target. The reduced frequencies of A and T are also consistent with single substitution data - substitution of an A or T substantially reduced Tev-mTALEN activity. The wildtype C nucleotide is also under-represented in active clones. At position 2, active clones showed a 2.5-fold enrichment of A and T, a 3-fold reduction in G and a 6-fold reduction in C. The reduction in C and G nucleotide frequencies is consistent with the results of single substitution assays, in which the T2C and T2G substrates were cleaved with 6% and <1% normalized activity, respectively. The enrichment of A at position 2, by comparison, is not consistent with single substitution data as the T2A mutant was also cleaved much less efficiently than the TP15 substrate. At position 3, active clones showed a 2-fold reduction in T, consistent with the results of spacer single substitution assays. A similar reduction in T was also observed at position 4. This result is in contrast to the spacer single substitution data, where the A4T substrate was cleaved with an average of 263% normalized activity. No significant nucleotide preference was observed at position 5. At position 6, where all single nucleotide substitutions of the wildtype T substantially reduced Tev-mTALEN activity (Figure 3.1), only a minor preference against G was observed (less than a 2-fold reduction in active clones). At position 7 active clones showed a 3-fold increase in C and a 4-fold reduction in the wildtype A. At position 8

only a minor (< 2 -fold) preference against C was observed, in spite of the spacer single substitution data showing a strong preference for the wildtype G. At position 9 a 2-fold increase in the wildtype A was observed. No significant preferences were observed from positions 10-12. Positions 13 and 14 showed a 2-fold reduction in C, while position 15 showed a 2-fold increase in C. A < 2 -fold increase in A was observed at position 14.

Figure 3.4 shows sequence LOGOs of the active and inactive clones. Although the sample sizes (50 active and 62 inactive) are too small to identify a consensus sequence, the LOGOs show that the linker domain is tolerant of non-wildtype nucleotides across the DNA spacer. The preference for G at position 1 and T at position 2 can be seen, but even these are clearly not essential for activity. In agreement with data from native I-TevI and Mega-Tevs, sequencing of the active clones reveals that no single nucleotide at any position in the DNA spacer is absolutely required for Tev-mTALEN activity - nor does a single nucleotide at any position universally inhibit activity. While there are nucleotide preferences at several positions along the DNA spacer, the sequence requirements are context-dependent. This means that Tev-mTALEN cleavage of a putative target site cannot be predicted simply by examining the identity of individual nucleotides in the DNA spacer independent of each other.

3.3 Summary

Analyses of Tev-mTALEN DNA spacer requirements have identified nucleotide preferences at several positions in the DNA spacer. The DNA spacer single nucleotide substitution assays have identified preferences for and against certain nucleotides at DNA spacer positions 1, 2, 3, 5, 6, 8, 9, and 14. Screens of fully randomized DNA spacers from the N15 library have also identified nucleotide preferences in a broader context. Sequencing of Tev-mTALEN targets with cleavage-promoting DNA spacers has shown enrichment and reduction of certain nucleotides at positions 1, 2, 3, 6, 7, 8, 9, 13, 14, and 15. Many of the preferences identified in the N15 library screen agree with the results of the single substitution assays, such as the enrichment of G at position 1 and the position 2 preference for a T. However, many preferences were observed in the single substitution assays which were not observed in the N15 library screen. Two examples of this are the preferences at positions 6 and 8 – DNA spacer single nucleotide substitution data showed

that mutation of either of these positions to non-wildtype nucleotides substantially reduced activity. However, no substantial enrichment of either wildtype nucleotide was observed in the 50 active clones isolated from the N15 library. Conversely, nucleotide preferences were observed in the N15 library screen which were not apparent in the single substitution assays. Active clones from the N15 library had nearly a 3-fold greater abundance of A at position 2 than the inactive clones, even though the single substitution data showed that the T2A mutant target was cleaved 5-fold less efficiently than the TP15. The contrast between the results of the single nucleotide substitution assays and the N15 library screen indicate that nucleotide context is important in determining which DNA spacer sequences promote efficient Tev-mTALEN activity. Predicting which DNA spacer sequences will support robust cleavage by Tev-mTALENs is not possible by examining the identity of nucleotides at specific positions independently.

Interestingly, the results of these assays agree qualitatively with studies of the I-TevI linker preference in the context of Mega-Tev nucleases. Jason Wolfs performed *in vitro* assays designed to enrich Mega-Tev targets with cleavage-promoting DNA spacer sequences from a randomized library. These assays have identified base preferences at the same positions for Mega-Tevs as those observed for Tev-mTALENs. Furthermore, preferences for specific bases at these positions are largely the same between the two families of enzymes. This suggests that like the nuclease domain, the I-TevI linker also functions in a largely modular fashion, retaining its nucleotide preferences in the presence of non-native DNA-binding domains.

Chapter 4

4 Novel I-TevI Linker Domains with Altered Specificity can Broaden Tev-mTALEN Targeting Potential

Single nucleotide substitution assays and screening of the N15 DNA spacer library have shown a preference for certain nucleotides at several positions. These preferences mean that many potential Tev-mTALEN targets will not be cleaved efficiently due to the DNA spacer sequence. Using Mega-Tev nucleases, Jason Wolfs has isolated I-TevI variants from a partially randomized library of the linker domain that facilitate activity on DNA spacer substrates that are poorly cleaved by the wildtype enzyme. The similarities between the DNA spacer preferences in Tev-mTALENs and Mega-Tevs suggested that these mutations might also confer altered sequence preferences to Tev-mTALENs. Thus, Tev-mTALENs bearing these same linker mutations were generated and screened against the DNA spacer single substitution target sites to determine if the nucleotide preferences of the linker could be similarly altered.

4.1 Spacer Single Nucleotide Substitution Assays with Tev-mTALEN Linker Variants

Four Tev-mTALEN constructs with mutations in the N169 I-TevI domain were generated, S134G, S134G/N140S, V117F/D127G, and K135R/N140S/Q158R. The spacer single nucleotide substitution assay described in chapter 3 was repeated using each of the mutant Tev-mTALENs. Activity measurements were normalized to that of the mutant Tev-mTALEN on the TP15 substrate and plotted alongside the wildtype enzyme measurements from chapter 2 for comparison.

The S134G mutant linker was isolated from screen of the partially randomized I-TevI linker against the T6G DNA spacer substrate that is poorly cleaved by the wild-type I-TevI linker. Figure 4.1 shows the activity of the Tev-mTALEN S134G variant on each of the 45 single nucleotide substitution TP15 substrates, normalized to the activity of the S134G on the TP15 substrate (with wildtype data from Chapter 3 for comparison). On most mutant substrates, including the T6G, relative activity of the S134G was comparable or slightly lower than the wildtype enzyme. However, the S134G showed

improved relative activity on several mutant substrates which the wildtype enzyme cleaved poorly. Activity on the C1T substrate was 65% TP15 normalized, twice the 32.5% of the wildtype enzyme. Relative activity on all three T2 mutant substrates was improved, most notably the T2C mutant which the S134G cleaved at 50% average normalized activity, compared with the 6% of the wildtype enzyme. Relative activity on the T6C substrate was improved from 12% to 42%. At position 8, the S134G cleaved the G8T substrate with an average of 24% normalized, double that of the wildtype enzyme.

Like Tev-mTALENs, Mega-Tevs are sensitive to substitution of spacer nucleotide G8 to any other base, although to a much greater degree. The S134G/N140S and K135R/N140S/Q158R mutants (Figure 4.2 and 4.3 respectively) were both isolated in screens against the G8A mutant spacer substrate. Both sets of mutations relieved sensitivity to the G8A substitution, with the Tev-mTALEN variants cleaving the mutant substrate more efficiently than the TP15 (211% activity for the S134G/N140S and 437% activity for the K135R/N140S/Q158R). Additionally, both variants were not sensitive to several position 1, 2, 3, 6, 7, 8, and 9 substitutions which reduced wildtype Tev-mTALEN activity.

The V117F/D127G linker mutant (Figure 4.4) was isolated in a screen against a substrate with multiple non-wildtype spacer nucleotides. Based on the DNA spacer sequence, this substrate was predicted to be cleaved efficiently by the wildtype Mega-Tev nuclease, but when assayed was cleaved poorly. Against the majority of mutant substrates, V117F/D127G activity was equal to or greater than activity on the TP15, including position 1, 2, 3, 6, 8, and 9 mutants to which the wildtype enzyme is particularly sensitive. This relaxation of nucleotide preferences was most apparent at positions 2 and 6. The wildtype Tev-mTALEN is sensitive to all single base substitutions at positions T2 and T6, but the V117F/D127G variant cleaves all substrates with equal or greater efficiency compared to the TP15.

While the results of chapter 3 have shown that data from DNA spacer single substitution assays (Figure 3.1) are not completely indicative of the nucleotide preferences in a broader context (Figure 3.3B), many nucleotide preferences observed for the wildtype

Tev-mTALEN in single substitution assays were also observed in a screen of fully randomized N15 DNA spacer substrates. Based on this, it is likely that the altered nucleotide preferences of the four mutant linker variants examined here are similarly reflective of their DNA spacer compatibility on a broader scale. This means that the targeting constraints imposed by the wildtype enzyme's DNA spacer requirements can be overcome by the development of mutants which effectively cleave those targets which the wildtype does not. However, the activity measurements for the mutant constructs and the wildtype Tev-mTALEN in figures 4.1-4.4 are each normalized to their own activity on the TP15. This allows for a comparison on nucleotide preferences, relative to each enzymes activity on the TP15 target, but it does not provide a direct comparison of actual activity. Many of the linker variants examined cleaved the TP15 target poorly, so their relaxed nucleotide preferences do not necessarily indicate an actual improvement in activity compared to the wildtype Tev-mTALEN.

Figures 4.5-4.8 show a direct comparison of mutant and wildtype Tev-mTALEN activity on each of the DNA spacer targets. Mutant and wildtype enzyme activity was normalized to the activity of the wildtype enzyme on the TP15, to allow for a direct comparison of activity on each substrate (Figure 4.5A/4.6A/4.7A/4.8A). Figures 4.5B/4.6B/4.7B/4.8B show the Log_2 of the ratio of mutant construct activity over wildtype Tev-mTALEN activity for each of the DNA spacer substrates. A value of 0 indicates that the mutant Tev-mTALEN cleaved the substrate with the same activity as the wildtype enzyme. Values above 0 indicate that the linker variant cleaved the substrate with greater activity than the wildtype, while negative values indicate that the mutant construct was less efficient than the wildtype.

The K135R/N140S/Q158R variant was 11-fold less active on the TP15 substrate than the wildtype enzyme and 2- to 16-fold less active on the majority of mutant substrates (Figure 4.7B). The S134G/N140S variant was also generally less active than the wildtype enzyme. S134G/N140S activity on the TP15 was ~40% of the wildtype (Figure 4.6A) and activity on most mutant substrates was 2- to 4-fold less than the wildtype; however, the mutations slightly improved activity on several position 2, 6, and 8 single substitution substrates which the wildtype enzyme cleaved poorly (Figure 4.6B). The

V117F/D127G mutations tended to reduce activity compared to the wildtype enzyme. Activity on the TP15 was 4-fold less than wildtype and, similar to the S134G/N140S variant, activity on most other substrates was reduced 2- to 4-fold (Figure 4.8B). The V117F/D127G mutations improved activity on all the position 2 and 6 single substitution substrates as well as the C3T substrate. The most noticeable improvements in activity were on the T2C and T2G substrates. The wildtype enzyme cleaved these substrates extremely poorly, with 6% and <1% TP15-normalized activity respectively; however, the V117F/D127G variant cleaved both of these substrates with ~40% of the activity of the wildtype enzyme on the TP15 (Figure 4.8A).

Unlike the other three Tev-mTALEN variants, the S134G was generally more active than the wildtype enzyme. Activity on the TP15 was twice that of the wildtype, and on the majority of mutant substrates the S134G had greater activity than the wildtype (Figure 4.5A and B). The S134G mutation significantly improved Tev-mTALEN activity on several substrates which the wildtype enzyme cleaved poorly. Mutation of DNA spacer nucleotide C1 to an A or T reduced wildtype Tev-mTALEN activity to 14% and 33% normalized, respectively. S134G activity on the C1A substrate was more than 3-fold greater at 47% normalized, and activity on the C1T improved 4-fold from 33% to 129%. Activity on all of the position 2 mutants was improved compared to the wildtype: activity on the T2A substrate was increased 4-fold from 19% to 82%; activity on the T2C substrate increased 16-fold from 6% to 95%; and activity on the T2G was 23%, compared with the background levels of activity seen for the wildtype enzyme. Activity on the C3A substrate increased from 56% to 144%, and activity on the C3T substrate increased 3-fold from 6% to 19%. S134G activity on the T6C substrate was 92%, more than a 7-fold increase compared to the wildtype enzyme which cleaved with ~12% normalized activity. The S134G mutation also improved activity on the T6G target from 12% to 28%. The S134G cleaved all of the DNA spacer G8 mutants more efficiently than the wildtype enzyme: activity on the G8A substrate increased 2-fold from 25% to 48%; activity on the G8C increased more than 3-fold from 39% to 129%; and activity on the G8T increased 5-fold from 11% to 54%. Interestingly, the S134G was not significantly less active on any of the DNA spacer single substitution substrates than the wildtype, with the exception of the T10A. Average wildtype activity on the T10A was

222%, while the S134G cleaved it with ~100% normalized activity. Overall, mutation of serine 134 in the I-TevI linker to a glycine broadly increased average Tev-mTALEN activity against most of the substrates assayed, without rendering the enzyme ineffective against any substrates which the wildtype was not. These results agree with assay data from Jason Wolfs, which showed that the S134G linker mutant also improved Mega-Tev activity.

Even though three of the linker variants generally reduced Tev-mTALEN activity, the assays with the mutant enzymes still demonstrate that mutations to amino acids in the I-TevI linker domain can alter the DNA spacer nucleotide preferences of the enzyme. Each of the mutant Tev-mTALEN enzymes examined here have altered DNA spacer nucleotide preferences compared to the wildtype enzyme. In spite of their reduced average activity, both the V117F/D127G and S134G/N140S mutations improved Tev-mTALEN activity on several position 2, 6, and 8 DNA spacer substitution targets which the wildtype enzyme cleaved poorly. The most striking assay result were those of the S134G variant. Compared to the wildtype Tev-mTALEN, the S134G linker variant had improved activity on the TP15 target and most mutant substrates, including many poorly cleaved position 1, 2, 3, 6, 8, and 9 single substitution substrates. The mutant I-TevI linkers examined here represent only a few of the variants isolated in the original Mega-Tev screen performed by Jason Wolfs. Furthermore, the initial screen was performed only on a small selection of poorly-cleaved mutant substrates. It is likely that further selection on other wildtype-poor substrates will yield new I-TevI linker mutations which can modulate altered or relaxed Tev-mTALEN DNA spacer preferences.

4.2 Summary

Examination of several mutant Tev-mTALENs has shown that as little as 1 amino acid mutation in the I-TevI linker domain can relax or alter the DNA spacer nucleotide preference of Tev-mTALENs. While these data do not elucidate the mechanism of Tev-mTALEN targeting specificity, they do demonstrate that mutations to the I-TevI linker domain can produce enzymes with altered or relaxed DNA spacer sequence-compatibility. A sufficient repertoire of functional linker variants with altered DNA

spacer nucleotide preferences will improve the targeting potential of Tev-mTALENs beyond that of the wildtype enzyme.

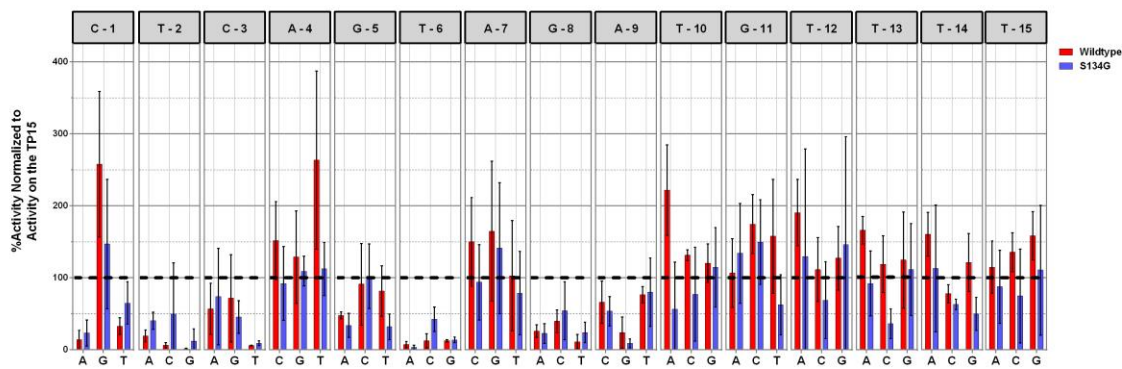


Figure 4.1: Spacer Single Nucleotide Substitution Assays with the S134G Tev-mTALEN Variant

Spacer single nucleotide substitution assays (Chapter 3.1) were performed with the S134G Tev-mTALEN linker variant. Activity measurements are normalized to the activity of the S134G on the TP15 target. Wildtype Tev-mTALEN measurements from chapter 3 are shown in red alongside the corresponding S134G measurements for comparison. Errors bars indicate the standard deviation of three replicates.

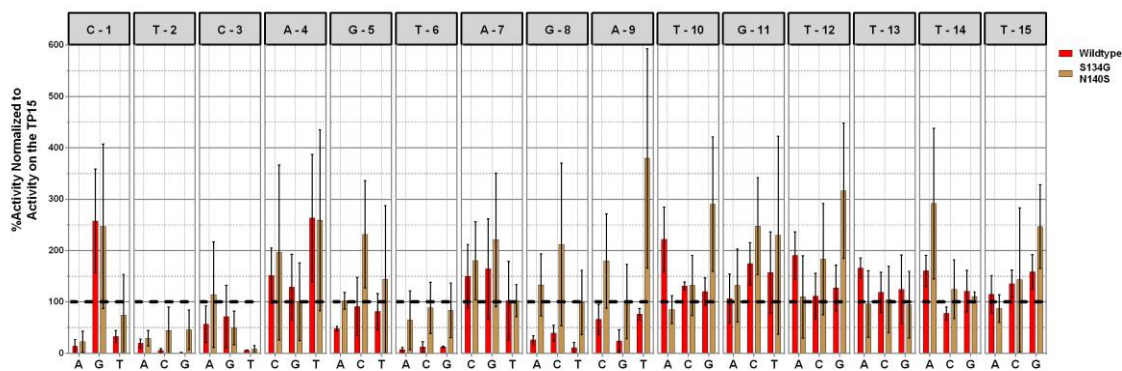


Figure 4.2: Spacer Single Nucleotide Substitution Assays with the S134G/N140S Tev-mTALEN Variant

Spacer single nucleotide substitution assays (Chapter 3.1) were performed with the S134G/N140S Tev-mTALEN linker variant. Activity measurements are normalized to the activity of the S134G/N140S on the TP15 target. Wildtype Tev-mTALEN measurements from chapter 3 are shown in red alongside the corresponding S134G/N140S measurements for comparison. Error bars indicate the standard deviation of three replicates.

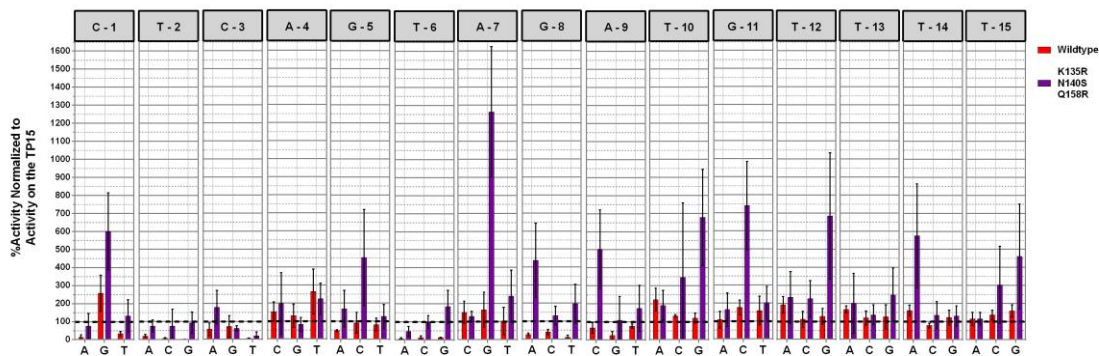


Figure 4.3: Spacer Single Nucleotide Substitution Assays with the K135R/N140S/Q158R Tev-mTALEN Variant

Spacer single nucleotide substitution assays (Chapter 3.1) were performed with the K135R/N140S/Q158R Tev-mTALEN linker variant. Activity measurements are normalized to the activity of the K135R/N140S/Q158R on the TP15 target. Measurements from chapter 3 are shown in red alongside the corresponding K135R/N140S/Q158R measurements for comparison. Errors bars indicate the standard deviation of three replicates.

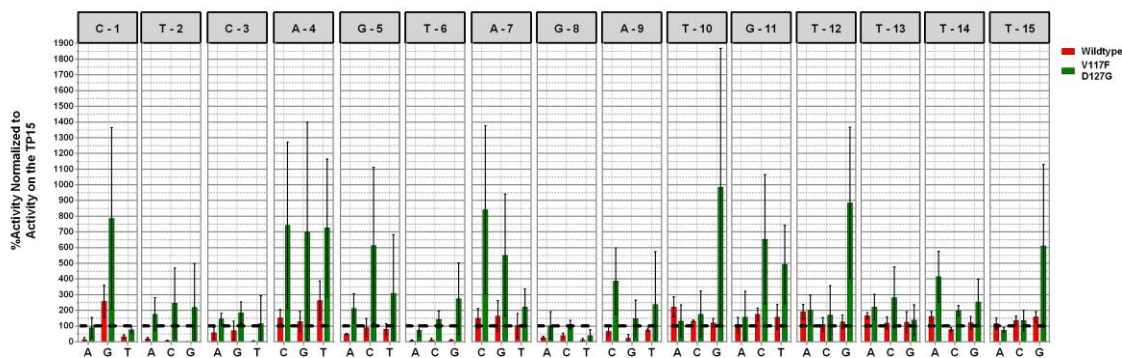


Figure 4.4: Spacer Single Nucleotide Substitution Assays with the V117F/D127G Tev-mTALEN Variant

Spacer single nucleotide substitution assays (Chapter 3.1) were performed with the V117F/D127G Tev-mTALEN linker variant. Activity measurements are normalized to the activity of the V117F/D127G on the TP15 target. Measurements from chapter 3 are shown in red alongside the corresponding V117F/D127G measurements for comparison. Errors bars indicate the standard deviation of three replicates.

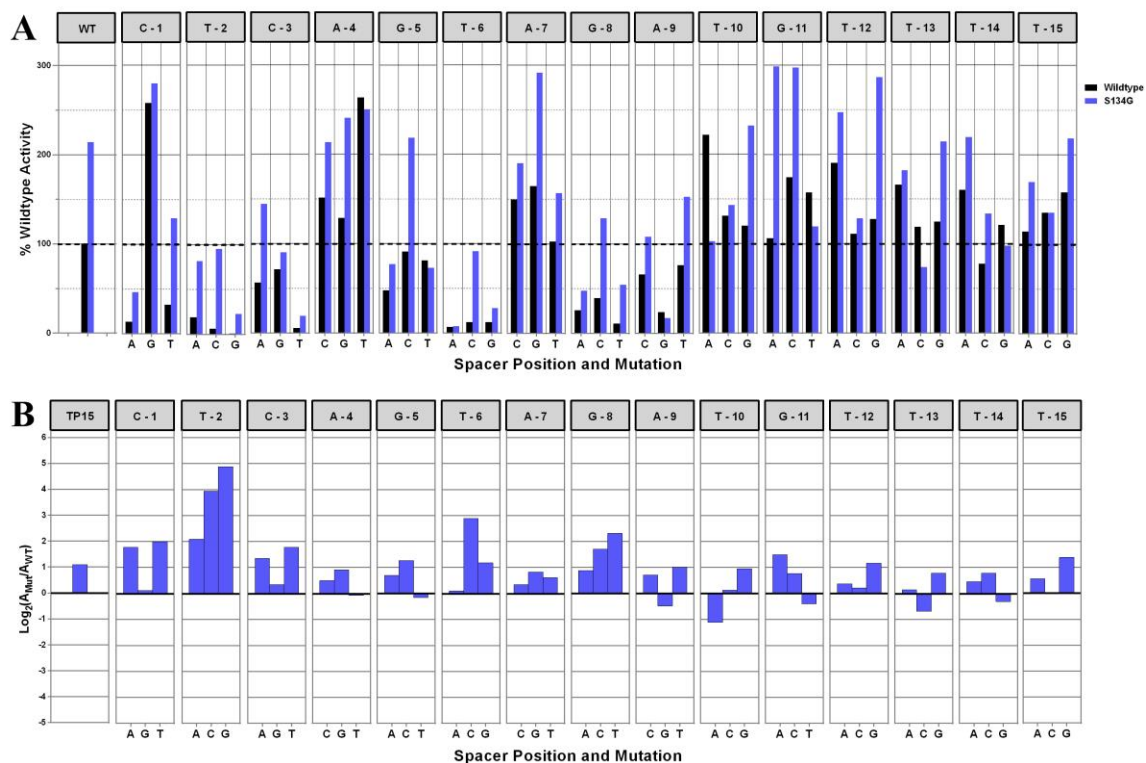


Figure 4.5: Comparison of Wildtype and S134G Tev-mTALEN Activity

(A) Comparison of spacer single nucleotide substitution data for the S134G construct (blue bars) and the wildtype enzyme (black bars). Wildtype Tev-mTALEN activity measurements are taken from Figure 3.1. S134G Tev-mTALEN activity measurements are taken from the data set used for figure 4.1 and adjusted so that activity is normalized to the wildtype Tev-mTALEN on the TP15. (B) Relative activity of the S134G on the DNA spacer single substitution targets compared to the wildtype Tev-mTALEN. Relative activities are expressed as the ratio of mutant construct activity over wildtype activity on each substrate, in Log₂ scale.

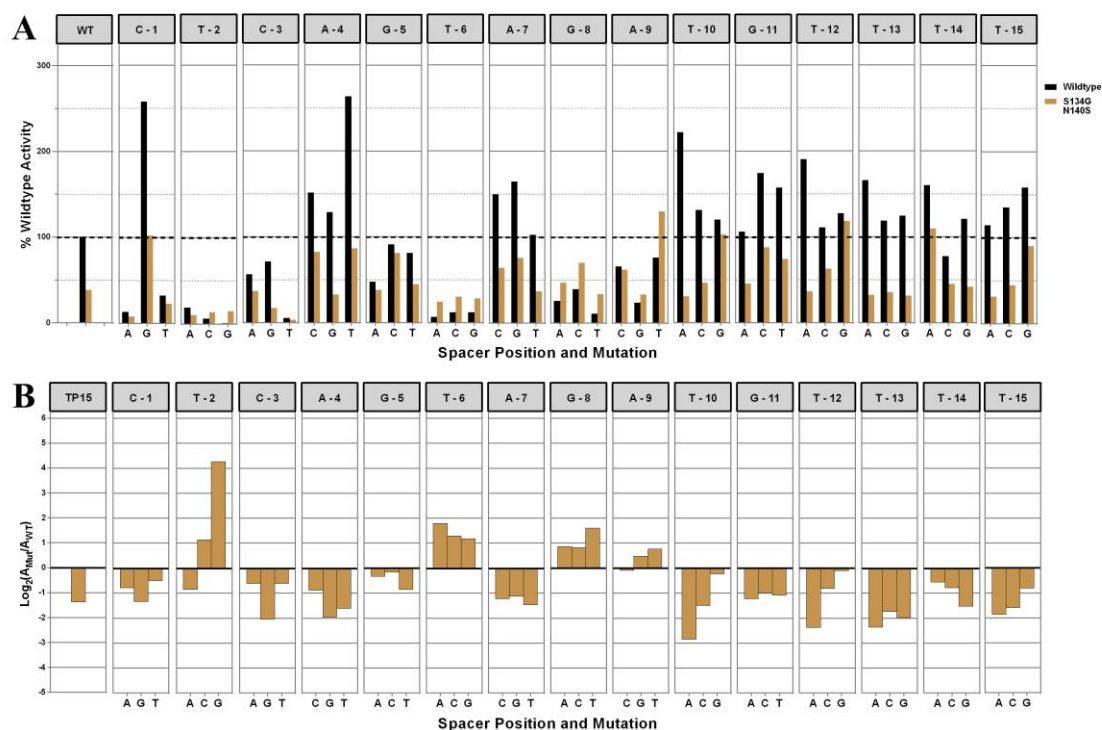


Figure 4.6: Comparison of Wildtype and S134G/N140S Tev-mTALEN Activity

(A) Comparison of spacer single nucleotide substitution data for the S134G/N140S construct (golden bars) and the wildtype enzyme (black bars). Wildtype Tev-mTALEN activity measurements are taken from Figure 3.1. S134G/N140S Tev-mTALEN activity measurements are taken from the data set used for figure 4.2 and adjusted so that activity is normalized to the wildtype Tev-mTALEN on the TP15. (B) Relative activity of the S134G/N140S on the DNA spacer single substitution targets compared to the wildtype Tev-mTALEN. Relative activities are expressed as the ratio of mutant construct activity over wildtype activity on each substrate, in Log₂ scale.

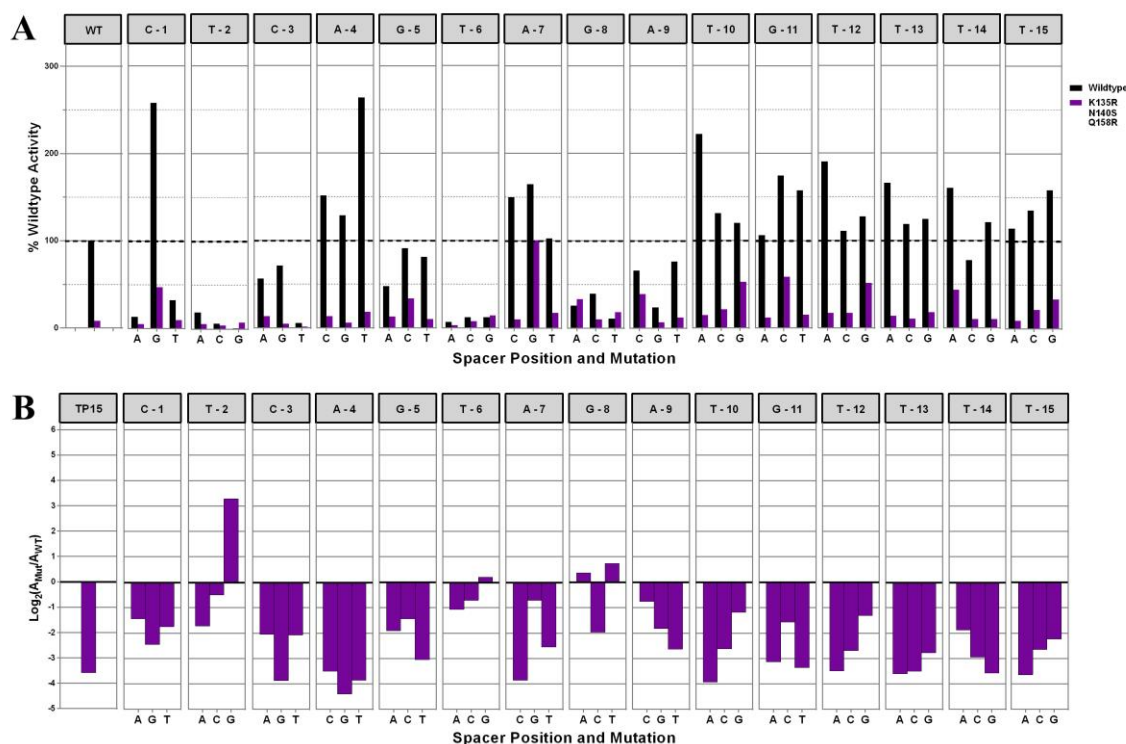


Figure 4.7: Comparison of Wildtype and K135R/N140S/Q158R Tev-mTALEN Activity

(A) Comparison of spacer single nucleotide substitution data for the K135R/N140S/Q158R construct (purple bars) and the wildtype enzyme (black bars). Wildtype Tev-mTALEN activity measurements are taken from Figure 3.1. K135R/N140S/Q158R Tev-mTALEN activity measurements are taken from the data set used for figure 4.3 and adjusted so that activity is normalized to the wildtype Tev-mTALEN on the TP15. (B) Relative activity of the K135R/N140S/Q158R on the DNA spacer single substitution targets compared to the wildtype Tev-mTALEN. Relative activities are expressed as the ratio of mutant construct activity over wildtype activity on each substrate, in Log₂ scale.

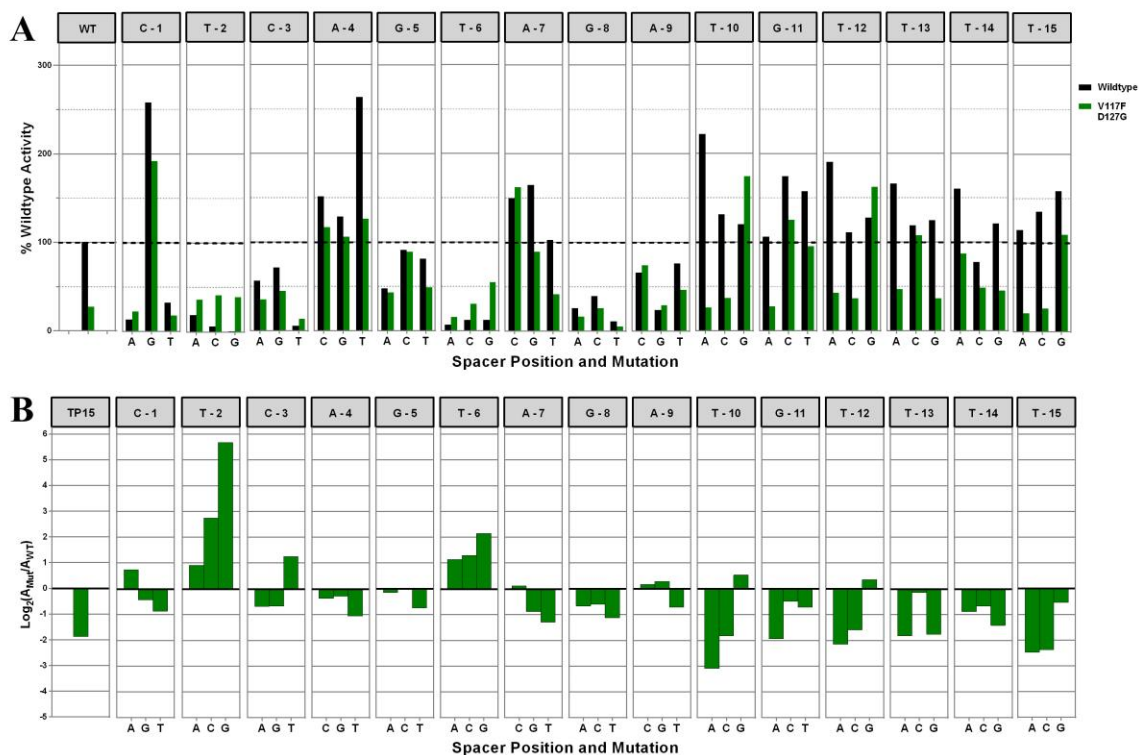


Figure 4.8: Comparison of Wildtype and V117F/D127G Tev-mTALEN Activity

(A) Comparison of spacer single nucleotide substitution data for the V117F/D127G construct (green bars) and the wildtype enzyme (black bars). Wildtype Tev-mTALEN activity measurements are taken from Figure 3.1. V117F/D127G Tev-mTALEN activity measurements are taken from the data set used for figure 4.4 and adjusted so that activity is normalized to the wildtype Tev-mTALEN on the TP15. (B) Relative activity of the V117F/D127G on the DNA spacer single substitution targets compared to the wildtype Tev-mTALEN. Relative activities are expressed as the ratio of mutant construct activity over wildtype activity on each substrate, in Log₂ scale.

Chapter 5

5 Discussion and Conclusion

For the most advanced applications, nuclease-mediated genome editing requires tools that have highly precise sequence specificity and are simple to engineer. An enzyme may cleave its intended target with high efficiency, but also cleave similar sequences with nearly equal efficiency, making it unsuitable for sensitive applications. Conversely, an enzyme family may possess exquisite sequence discrimination but be extremely laborious to re-engineer for non-wildtype target sequences, making its use impractical. Tev-mTALENs have the potential to be both simple to engineer and minimally toxic; however, the cryptic DNA spacer sequence requirements of the I-TevI linker domain complicate the otherwise simple targeting process. Work presented here has shown that, unlike the relatively simple sequence requirements of the I-TevI nuclease domain and the TAL effector DNA-binding domain, the sequence requirements of the I-TevI linker are more complex.

5.1 Summary

Assays on the DNA spacer single substitution substrates have shown that Tev-mTALENs are sensitive to the identity of several nucleotides in the DNA spacer. Substitution of one or more non-wildtype nucleotides at DNA spacer positions 1, 2, 3, 5, 6, 8, and 9 was shown to reduce average activity relative to the TP15 target. Screening of the randomized N15 DNA spacer library has confirmed many of these preferences. However, many preferences observed in the single substitution assays were not observed in the library screen. Similarly, many preferences seen in the N15 screen were not identified in the single substitution assays. These results indicate that the DNA spacer sequence requirements are context-dependent, and no single nucleotide requirements exist at any position in the DNA spacer. However, sequencing data has identified clear nucleotide preferences at several positions in the DNA spacer. Understanding these will improve the success of Tev-mTALEN target site selection.

Analysis of mutant Tev-mTALEN variants has identified several amino acids in the I-TevI linker domain which can alter Tev-mTALEN DNA spacer nucleotide preferences. These mutants represent only a handful of potential linker variants, but have shown that small mutations to the I-TevI linker domain can significantly alter or relax the nucleotide preferences of Tev-mTALENs. Of the linker variants examined, the S134G stands out for its broadly increased activity and relaxed DNA spacer nucleotide preferences. Because many DNA spacer sequences fail to promote efficient cleavage by the wildtype Tev-mTALEN, mutant I-TevI linker domains with a broader tolerance for DNA spacer sequences such as the S134G will be essential for broadening Tev-mTALEN targeting potential in the future.

5.2 Limitations and Future Directions

5.2.1 Limitations of Yeast Reporter Assays

A major limitation of the data presented here is the assay used to measure Tev-mTALEN activity. Activity measurements obtained from the β -galactosidase reporter assay tended to vary significantly. In a single 96 well plate replicate, the standard error of wildtype Tev-mTALEN activity on the TP15 target was anywhere from 20% to 40% of the value of the average measurement. In some technical replicates, activity on the TP15 was at background levels. Similar variation was observed for the majority of the DNA spacer single substitution targets, with the standard error of some well-cleaved substrates exceeding 100% of average TP15 activity. The high degree of variation in measurements makes it difficult to determine if minor differences in average activity on different substrates are a genuine result of linker nucleotide preferences or just inherent assay variation. Average activity on many of the DNA spacer single substitutions was greater than the TP15, however, reliably identifying which of these are true preferred substrates and which are simply a result of assay error is not possible. Substrates which were poorly cleaved compared to the TP15 (<50% normalized activity) tended have more consistent measurements, so the assay in its current form can only reliably identify substrates which are noticeably less efficient than the TP15. There are likely more subtle nucleotide preferences at other positions in the DNA spacer which were not determined in this report, due to the poor precision of the reporter assay. A similar degree of

variation was observed in the assays performed with the Tev-mTALEN linker mutants. Many average activity and error measurements were skewed by a single replicate which was unusually high or low compared to the others. Furthermore, comparison of activity measurements between linker variants is dependent on the precision of the positive control measurements in each plate. Each of the four Tev-mTALEN linker variants was assayed against the TP15 target and 45 DNA spacer mutants in separate batches of plate replicates. In each plate, the wildtype Tev-mTALEN was tested against the TP15 target, acting as a positive control. Linker variant activity measurements against each substrate were normalized to the wildtype-TP15 positive control in their respective plates. The four linker variants were then compared to one another and the wildtype enzyme based on how efficiently they cleaved the DNA spacer substrates, relative to the positive control (Figure 4.5). This means that plate-to-plate variations in wildtype activity on the TP15 target will alter how efficient each Tev-mTALEN linker variant appears compared to the others. Low positive control activity in a plate will exaggerate the efficiency of the Tev-mTALEN linker variant on each of the DNA spacer targets, while an unusually high positive control measurement will make the variant appear generally less active than it actually is. The linker variant experiments were performed in sets of 3 plate replicates, with 2 technical replicates per plate in order to mitigate the effects of plate-to-plate variation; however, variation inherent to the assay is still a concern.

One factor that may contribute to the high degree of variation in activity measurements is the amount of nuclease expressed in each culture. The reporter assay measures cleavage of a target site indirectly through β -Galactosidase cleavage of ONPG. Yeast diploids harbouring the nuclease expression plasmid and the target site plasmid are generated through mating and then allowed to grow for 18 hours. During this time, the nuclease is expressed under the control of the strong, constitutive glyceraldehyde-3-phosphate dehydrogenase (GPD) promoter. Cleavage of the target site stimulates repair of a functional LacZ gene. A well-cleaved target site will result in more repair events in the yeast culture and more β -Galactosidase expression. When the yeast are lysed during the assay phase, cultures with well-cleaved targets will release more β -Galactosidase enzyme, resulting in more ONPG degradation and an increase in the OD_{405nm} reading of the culture. The basic assumption of this assay is that differences in the OD_{405nm} reading

of each well are a direct result of varying cleavage efficiency. OD_{595nm} readings are taken for each culture and factored into the activity calculation, in order to account for the effect of cell density on the frequency of repair events. Both the nuclease and the repaired LacZ gene are under the control of the GPD promoter. By accounting for cell density and using a strong, constitutive promoter for both the nuclease and the reporter enzyme, the protocol attempts to minimize the impact of variable protein expression on target site cleavage and β -Galactosidase expression - making target site cleavage the primary determinant of OD_{405nm} readings. However, no direct measurement of nuclease expression is performed. It is possible that expression of the nuclease may vary significantly from culture to culture without being accounted for, increasing the standard error of activity measurements. A possible improvement to the reporter assay may be the use of a reporter tag, such as a fluorescent protein fusion, to directly measure enzyme levels. This would allow for correction of OD_{405nm} readings by normalizing to the levels of nuclease expression in each culture.

5.2.2 Limitations of the N169T120 Tev-mTALEN Architecture

The construct used in this report consists of the N169 truncation of I-TevI and the T120 N-terminal truncation of PthXo1; however, other functional truncations for both proteins exist. Alternate I-TevI truncations, such as the D184 and S206, can be used to create functional Tev-mTALENs, both of which tend to have greater activity on their optimal substrates than the N169 fragment (78). As well, multiple functional truncations of the TAL domain N-terminus exist – Tev-mTALENs and FokI TALENs constructed with the V152 TAL effector truncation are among the most active (78, 91). Furthermore, the assays described here were only performed on target sites with a DNA spacer length of 15bp. This length was chosen because it is the optimal length for N169-T120 activity; however, the enzyme can also cleave targets with DNA spacers of varying lengths (78). Because many TAL binding sites will not have a CNNNG motif exactly 15 bp upstream, being able to target Tev-mTALEN activity to motifs within a broader window of distance from a TAL site would increase the number of potential Tev-mTALEN targets significantly. Previous work by Ben Kleinstiver (78) has shown that, in addition to the N169-T120, each of the three Tev-mTALEN truncation variants: S206-V152, N169-

V152, and D184-V152, function optimally on substrates with different spacer lengths. Between these four constructs, a CNNNG motif located anywhere from 13-31bp upstream of a TAL binding site could be cleaved with efficiency comparable to or greater than the N169-T120 on the TP15 target. For this reason, future work should involve examining the I-TevI linker preferences in the context of alternate nuclease architectures and with varying spacer lengths. Although 15bp is the optimal length for the N169-T120 construct, the enzyme can also target substrates with spacers ranging in length from 13-19bp and 25-29bp (78). Examining the nucleotide preferences of the enzyme for shorter and longer spacers would be a suitable follow-up to the experiments presented here. As well, all three of the alternative Tev-mTALEN truncation variants mentioned are active on the TP15 target (78) – based on this, a natural follow-up experiment would be to repeat the TP15 DNA spacer single substitution assays described in Chapter 3 using the S206-V152, N169-V152, and D184-V152 Tev-mTALEN constructs, for a direct comparison of nucleotide substitution sensitivity between the four constructs.

Another follow-up experiment to consider would be to repeat the N15 DNA spacer library screen in Chapter 3 using the mutant linker variants examined in Chapter 4 – particularly the S134G variant. Results from Chapter 4 seem to indicate that the S134G mutation relaxes several of the nucleotide preferences of the wildtype enzyme, and increases the activity of the enzyme in general. Screening the S134G construct against the same N15 clones assayed in Chapter 3 (or at least the sequenced clones) would determine whether or not the relaxed nucleotide preferences observed in the single substitution screen are reflective of relaxed preferences in general. In Chapter 3, N15 DNA spacer clones were considered active if the activity of the Tev-mTALEN on the target site was no less than 2 standard deviations below average activity on the TP15. Based on this cut-off, roughly 22% of the random DNA spacer targets screened can be considered active. If the S134G linker mutation can improve this success rate, it would demonstrate the potential of linker mutations to broaden Tev-mTALEN targeting.

An alternative to mutagenic screening is identifying new GIY-YIG domains similar to I-TevI. I-TevI is the most well-characterized of the GIY-YIG family of homing endonucleases, however, numerous other GIY-YIG homing endonucleases exist. A

search of the Interpro database for “group I intron endonuclease” (IPR006350) yields hundreds of confirmed or putatively identified GIY-YIG homing endonucleases similar to I-TevI. Several of these enzymes, such as I-BmoI(92), I-TevII(93), I-BanI(94), and I-BthII(95) have been characterized to varying degrees. These enzymes have the same basic structure of I-TevI: an N-terminal GIY-YIG nuclease domain; a central domain containing 1-3 beta-turn-loop-helix motifs (NUMOD3 – nuclease associated modular domain 3); and a C-terminal helix-turn-helix motif (NUMOD1) (96). Each of these enzymes binds and cleaves a sequence that differs from the I-TevI homing site (92, 93, 94, 95). It is likely that the nuclease and linker domains from many of these proteins can be substituted for I-TevI, resulting in mTALENs with new cleavage motifs and DNA spacer requirements.

5.2.3 Limitations of Single Nucleotide Preference Analysis

Activity assays on non-wildtype DNA spacer substrates have identified certain nucleotide preferences for the DNA spacer sequence, but no absolute requirements at any position. These results are consistent with previous studies of the native I-TevI enzyme. While I-TevI does make a small number base-discriminant DNA contacts (80), no single nucleotide at any position of the wildtype *td* target site is essential for binding and cleavage (89). It is possible that identifying a defined Tev-mTALEN DNA spacer motif is impossible because linker-DNA spacer compatibility is not a result of direct readout of individual base identities. Crystal structure data and footprinting analyses have shown that the linker and DNA-binding domains of native I-TevI wrap primarily around the minor groove of the target site (80, 89), with the majority of contacts occurring at the minor groove and phosphate backbone (89). Major groove-binding proteins, such as zinc finger-based transcription factors, tend to achieve sequence recognition primarily through base-specific hydrogen bonding (97). Minor groove-binding proteins, in contrast, often recognize target sites through differences in sequence-dependent DNA structural properties, such as minor groove compression, asymmetric charge neutralization in the phosphate backbone, and bending stiffness (97, 98, 99). It is possible that Tev-mTALEN DNA spacer preferences are a result of variable target site bending stiffness.

Introduction of a bend at the cleavage site is an important step in I-TevI activity (86). I-TevI catalysis involves two sequential nicks to the target DNA, performed by the single active site within the GIY-YIG domain. The first nick occurs on the bottom strand of the CAACG cleavage site (CA_↑ACG). Following the first nick, a ~38° bend is introduced at the cleavage site and the second nick (CAAC_↓G) occurs. This distortion is believed to assist I-TevI catalysis by making the top strand nick site more accessible to the active site (86). The bend was mapped to the cleavage motif and the first 9 nucleotides of the DNA spacer, centered on DNA spacer nucleotide 2 (89). Interestingly, this includes a region of the native I-TevI *td* target site identified as DI (89, 100). DI consists of DNA spacer nucleotides 2-9, and represents a region of the target site in which native I-TevI is particularly sensitive to mutations (89, 101). DI also corresponds to a region of the target site in which I-TevI makes extensive minor-groove and phosphate backbone contacts (89). These enzyme-substrate contacts are directly associated with formation of the bend during I-TevI catalysis (86, 89). Based on these data, the authors proposed a model in which the I-TevI linker assists I-TevI catalysis through primarily minor groove and backbone interactions which facilitate bending at the cleavage site following the bottom strand nick (86). This distortion makes the top strand nick site accessible to the nuclease domain (86). Substitution of the critical C or G nucleotides of the CAACG motif with any other nucleotide (mutations which compromise I-TevI activity) resulted in both a significant reduction in bend formation and a reduction in the angle of the bend (86). In contrast, mutation of the A nucleotide in the central triplet (CAACG) to a C had no apparent effect on bend formation or bend angle (86). These data highlight the importance of target site bending in I-TevI catalysis, and suggests that the linker domain interacts with the DNA spacer sequence to facilitate this distortion. This means that the differences in Tev-mTALEN activity on the N15 DNA spacer substrates may be a result of varying target site flexibility. A reduction in target site bend formation and the maximum angle of the bend will likely impair the ability of the I-TevI nuclease to perform its two-step nicking activity. Interestingly, the majority of Tev-mTALEN nucleotide preferences identified in this report occur within the first 9 positions of the DNA spacer, corresponding to DI – the region in which substrate bending, minor groove and backbone contacts, and mutation-sensitivity were observed for the native I-TevI

enzyme. The connection between nucleotide sequence and target site distortion has been observed for other proteins. Mutations in the binding site of *E. coli* Catabolite Activator Protein (CAP) were shown to alter the flexibility of the binding site DNA, and a strong correlation between target site bending and CAP binding affinity was observed (101). A similar correlation between sequence-dependent DNA distortion and binding affinity was observed for the eukaryotic TATA binding protein (102). Structural readout of the cleavage motif and DNA spacer by the linker domain would explain why identifying preferences according to individual nucleotide identities is so difficult. DNA bending stiffness is primarily a product of interactions between adjacent base pairs (103), so nucleotide identities must be examined in the context of adjacent nucleotides.

Future experiments should examine the possible relationship between target site distortion and Tev-mTALEN activity. The assays performed with native I-TevI (86, 89) should be repeated with Tev-mTALENs to determine if there is a relationship between how efficiently a substrate is cleaved and how readily the substrate is distorted at the target site. Control assays on a small number of well-cleaved and poorly-cleaved targets would be a suitable starting point for confirming any possible connection. If it is shown that poorly-cleaved DNA spacer mutants are also less prone to cleavage site distortion, this would provide strong evidence that Tev-mTALEN DNA spacer preferences are not primarily a result of direct base readout. This would change the approach taken to predicting Tev-mTALEN target site cleavage, shifting the focus to prediction of cleavage motif and DNA spacer bending stiffness.

5.3 Advantages of Tev-mTALENs

Each of the existing genome-editing nuclease families has specific advantages and disadvantages based on their biology. CRISPRs and FokI-based TALENs have quickly become the two most commercially popular systems due to the simple nature of their sequence-specificity; however, at this point no single tool can be regarded as universally superior to others. There are a number of factors to consider, including the method of delivery, cell type, and type of modification desired, so tool selection must be done on a case-by-case basis. In addition to their monomeric and site-specific activity, a major advantage of Tev-mTALENs is the high degree of precision with which TAL effector

binding can be targeted. Early guidelines for TAL effector targeting outlined 5 rules for selecting robust binding sites: a T nucleotide at position 0 of the target site, position 1 V (not T), position 2 B (not A), a T nucleotide at the 3'-most position, and an overall nucleotide composition that does not differ by more than two standard deviations from that of naturally occurring TAL sites (104, 105). These guidelines were an early estimate based on examination of identified *Xanthomonas* TAL effector binding sites. However, a large-scale screen of TALEN activity against 48 target sites in the eGFP sequence and 96 human gene targets subsequently showed that the latter 4 guidelines are unnecessary (105). Only the presence of a position 0 T showed any correlation to TALEN activity (105). The discovery of TAL effector N-terminal domains with altered position 0 nucleotide preferences (52, 53) or none at all (54) means that this does not have to be a targeting constraint either. The lack of any theoretical sequence constraints makes TAL effectors the most precise DNA-binding domains available. This level of precision may not be necessary for gene knockout, since frameshift mutations can be equally effective across a broad window of DNA targets; however, when sequence insertions or corrections are needed, precision is critical. Homology-directed repair is naturally inefficient in mammalian cells (4, 6, 7, 8), and tends to be less efficient the further away the DSB is from the site of the intended mutation (106, 107, 108). Given that the average GC content of the human genome is ~41% (109), the CRISPR motifs GN₁₉NGG and GGN₁₈NGG can be expected to occur once every ~63 and ~313 bp, respectively. If paired CRISPR nickases are used, then two of these sites have to be located typically within 100bp and on opposite strands of each other (64), closer if efficient homology-directed repair is necessary. In the case of Tev-mTALENs, the only major targeting constraints are the CNNNG cleavage motif and DNA spacer requirements. A CNNNG motif will occur on average every ~13 bases. If ~22% of DNA spacers sequences are permissive of wildtype Tev-mTALEN activity (based on the results of Chapter 3), then this means a compatible site will occur every ~57bp. These numbers are only estimates, as the targeting requirements of both families are not completely understood. Some TALENs and CRISPRs will fail to efficiently bind and cleave perfect-match target sites, indicating that the sequence requirements of both families are more nuanced than current targeting models suggest (105, 110). Furthermore, the full targeting potential of both

families has yet to be explored. Recent work with the CRISPR/Cas system has shown that mutated Cas9 nuclease variants can be engineered to target new PAM sequences, expanding the number of potential target sites significantly (111). Similarly, the use of new/mutated GIY-YIG nuclease and linker domains may allow for targeting of new cleavage motifs and DNA spacer sequences, further expanding the targeting breadth and precision of the mTALEN family.

5.4 Conclusion

Tev-mTALEN activity assays with non-wildtype spacers have shown that examining base preferences alone cannot fully account for the complex linker-DNA spacer requirements. Regardless of the observed preferences, a putative target site may have a combination of spacer bases that make it seem like a viable target, yet still be cleaved poorly. Conversely, a target site may appear to be a poor substrate according to individual base preferences, yet be cleaved more efficiently than the TP15 target. With over a billion potential 15bp DNA spacer sequences (4^{15}) and a lack of any absolute base requirements at any position, exhaustive assaying of positional base preferences alone is not a feasible way of developing an accurate targeting model. Future work should include investigating what role, if any, target site distortion plays in determining Tev-mTALEN DNA spacer requirements. Published work on I-TevI has established a strong connection between target site distortion and I-TevI catalysis - the modular activity of the I-TevI nuclease and linker domains suggests that they may behave in the same manner when fused to the non-native TAL domain. Furthermore, additional Tev-mTALEN linker mutants with relaxed sequence requirements must be examined, in order to maximize the potential number of robust Tev-mTALEN targets. An accurate model of DNA spacer compatibility, combined with a sufficient repertoire of linker mutants with relaxed DNA spacer preferences, would eliminate the last practical barrier in Tev-mTALEN targeting prediction and, as a result, the need for control screening of each putative target. With all the modular sequence requirements understood, a program could be designed to scan a target for compatible CNNNG motifs and adjacent DNA spacer sequences, and a Tev-mTALEN enzyme with an appropriate TAL domain and I-TevI linker variant could be assembled for the target site. Monomeric, cleavage site-specific

activity, in combination with a predictable targeting system would make mTALEN enzymes a promising tool for sensitive genome editing applications, even capable of competing with the popular CRISPR nucleases.

Chapter 6

6 Materials and Methods

6.1 Bacterial and Yeast Strains

Bacterial strain *E. coli* DH5 α (Invitrogen) was used for all plasmid propagation. Strain *E. coli* ER2566 (New England Biolabs) was used for purification of the Tev-mTALEN construct. All β -galactosidase reporter assays were performed with strains *S. cerevisiae* YPH499 and YPH500 (104). All *in vivo* nuclease target sites were ligated into plasmid pCP5.1 and transformed into YPH499. All nuclease constructs were expressed *in vivo* from plasmid pGPD, which was transformed into YPH500. See Appendix A for detailed strain information.

6.2 Target-Site Plasmid Construction

6.2.1 *In Vitro* Cleavage Assays (Chapter 2.2)

The *in vitro* Tev-mTALEN substrate, pSP72-TP15, was generated by annealing of oligonucleotides DE1811/DE1812, phosphorylation with T4 PNK (New England Biolabs), and ligation into BglII/XbaI digested pSP72.

6.2.2 Control Targets and Cleavage Motif Mutants (Chapter 2.1)

The Zif268 target site plasmid, pCP5.1-ZF, was received from the lab of Dr Adam Bogdanove. The Mega-Tev target site plasmid, pCP5.1-TO15, was received from Jason Wolfs. Other substrates were generated by annealing of complementary oligonucleotides, phosphorylation with T4 PNK, and ligation into BglII/SpeI digested pCP5.1. Oligonucleotides DE1811/DE1812 were used to generate target vector pCP5.1-TP15. Oligonucleotides DE1612/DE1613, DE1734/DE1735, and DE1736/DE1737 were used to generate mutant cleavage motif vectors pCP5.1-TP15(TAACA), pCP5.1-TP15(TAACG), and pCP5.1-TP15(CAACA), respectively.

6.2.3 DNA Spacer Single Substitution Targets (Chapters 3.1 and 4)

pCP5.1-TP15 with positions 1 and 7-15 DNA spacer single substitutions were generated by annealing of consecutive oligonucleotide pairs from DE1546-DE1551 and DE1734-DE1791, with even numbers corresponding to top strand oligos and odd numbers corresponding to bottom strand oligos (full oligonucleotide pairings can be found in Appendix B). Oligos were phosphorylated with T4 PNK and ligated into BglII/SpeI digested pCP5.1. pCP5.1-TP15 with positions 2-6 DNA spacer single substitutions were generated by annealing of DE1496 to each of DE1715-DE1729, extension via Klenow Fragment exo- (New England Biolabs), digestion of extended product with BglII/SpeI, and ligation into BglII/SpeI digested pCP5.1.

6.2.4 N15 DNA Spacer Library (Chapter 3.2)

The N15 DNA spacer library was generated by annealing of DE1496 to DE1333, extension with Klenow fragment exo-, digestion with BglII/SpeI, and ligation into BglII/SpeI digested pCP5.1.

6.3 Expression Plasmid Construction

6.3.1 Bacterial Expression Plasmid Construction

Bacterial expression vector pACYC.Pci-N169T120(12RVD) was cloned by Ben Kleinstiver. The 12RVD variant of the N169T120 construct is otherwise identical to the full protein, but with 12 RVDs instead of the wildtype 23.

6.3.2 Yeast Expression Plasmids

Expression vectors pGPD, pGPD-N169T120, and pGPD-Zif268 were all received from the lab of Dr Adam Bogdanove. The Tev-mTALEN linker variants were generated using mutated I-TevI linkers from plasmids pACYC.Pci-N169ONU(S134G), pACYC.Pci-N169ONU(S134G/N140S), pACYC.Pci-N169ONU(K135R/N140S/Q158R), and pACYC.Pci-N169ONU(V117F/D127G), provided by Jason Wolfs. Site-directed mutagenesis was performed on each of the obtained plasmids using primers DE1175 and DE1176, in order to eliminate a PciI site located in the I-TevI coding sequence – this step is necessary in order to use an upstream 5' PciI site in subsequent cloning steps, but does

not alter the amino acid sequence of the enzyme. PciI restriction digests were performed to confirm the elimination of the PciI site from the plasmids (the number of sites in the plasmids is reduced from 2 to 1, if successful). After confirming successful elimination of the internal PciI site, the mutant I-TevI N169 domains were PCR-amplified using primers DE1013 and DE1213, then digested with enzymes PciI and BamHI. pACYC.Pci-N169T120 was digested with PciI and BamHI, to remove the wildtype I-TevI N169 domain, and the corresponding mutant N169 domains were each ligated in its place, generating plasmids pACYC.Pci-N169T120(S134G), pACYC.Pci-N169T120(S134G/N140S), pACYC.Pci-N169T120(K135R/N140S/Q158R), and pACYC.Pci-N169T120(V117F/D127G). Each of the pACYC.Pci mutant plasmids was digested with PciI/XhoI to isolate the complete mutant Tev-mTALEN sequences. The mutant constructs were subsequently ligated into PciI/XhoI digested pGPD, substituting the wildtype Tev-mTALEN for the mutant constructs. The resulting yeast expression plasmids were named pGPD-N169T120(S134G), pGPD-N169T120(S134G/N140S), pGPD-N169T120(K135R/N140S/Q158R), and pGPD-N169T120(V117F/D127G).

6.4 Purification of the 6xHis-tagged Tev-mTALEN

Expression vector pACYC.Pci-N169T120(12RVD) was transformed into chemically-competent *E. coli* ER2566 using the standard heat-shock protocol. Cultures were grown in 1L LB broth (+100µg/ml ampicillin) to $OD_{600nm} \sim 0.5$, at which point expression was induced by addition of IPTG to a final concentration of ~1mM. Induction was allowed to proceed at 16°C for 16 hours, at 200rpm in a baffled flask. Induced cultures were pelleted and resuspended in Buffer A (200mM NaCl, 20mM Tris-HCl pH 7.6, 1mM DTT, 1mM EDTA, 5% glycerol) at 4°C, then lysed in an EmulsiFlex cell homogenizer (Avestin). Cell lysate was spun in a pre-chilled JA25.50 rotor (Beckman Coulter) at 13000rpm for 25 minutes to pellet cellular debris. Clarified supernatant was run through a 1ml HisTrap-FF Ni²⁺ column (GE Healthcare) at a rate of 0.3ml/min. The column was washed with 10ml of Buffer A. Elutions were then performed with 2ml of Buffer A with increasing concentrations of imidazole - in order, elutions were performed with 2ml of 30mM buffer, 2ml of 50mM buffer, 5ml of 60mM buffer, and 5ml of 70mM buffer. Elution samples were taken in 1ml fractions. Small samples of each elution fraction were

run on SDS PAGE to examine fraction purity. The cleanest fractions were pooled and then spun down to 1ml using Vivaspin sample concentrators (GE Healthcare) in a JS5.3 rotor (Beckman Coulter) at 5300rpm. Concentrated samples were spun at 4°C and maximum speed in a tabletop centrifuge to pellet any precipitate. Clarified protein sample was loaded onto a Superose 12 10/300 GL size exclusion column (GE Healthcare) pre-equilibrated with Buffer A using an AKTA FPLC (GE Healthcare). 1 column volume (~24ml) of Buffer A was run over the column at a rate of 0.3ml/min with fractions taken in 0.25ml fractions. Using the AKTA chromatogram, peak fractions were selected and run on SDS PAGE. The purest samples were pooled and split into 20µl aliquots, then stored at -80°C. Samples of the primary polypeptide were sent for MALDI analysis at the UWO MALDI MS Facility, and confirmed to be the N169T120(12RVD) construct.

6.5 *In vitro* Cleavage Assays with the Purified Tev-mTALEN

In vitro cleavage assays were performed in 20µl reactions using a standard Tev-mTALEN reaction buffer based on NEBuffer 2 (New England Biolabs). The standard buffer consists of 50mM NaCl, 20mM Tris-HCl pH 8.0, 10mM MgCl₂, and 1mM DTT. Several variations of this buffer with altered salt or addition of the competitive DNA substrate, poly dI/dC, are outlined in Chapter 2.2. The DNA substrate (either pSP72 or pSP72-TP15) was maintained at a final concentration of 10nm. The substrate and reaction buffer were mixed on ice prior to addition of the protein. N169T120(12RVD) Tev-mTALEN protein samples were thawed on ice, diluted in reaction buffer, and added to the mixture. Reaction mixtures were then incubated at 37°C for 20 minutes and quenched with stop solution (100mM EDTA, 0.1% SDS). Stopped reactions were resolved on 1% agarose gel. Cleavage mapping was performed using a 5X larger reaction (100µl) and the excised linear band was purified using the Biobasic EZ-10 spin column protocol. Purified linear DNA was sequenced at the London Regional Genomics Center using oligos DE1114 and DE1452 and the resulting ABI traces were used for cleavage site mapping.

6.6 Yeast β -Galactosidase Reporter Assays

In vivo activity of the Tev-mTALEN constructs against target site substrates was measured using a modified version of the yeast reporter assay described by Christian et al (45), designed for 96-well microtitre plates. Yeast strain YPH499 was transformed with vector pCP5.1 containing the indicated target site and plated on minimal medium agar lacking tryptophan and uracil (Trp-/Ura-)(0.75% yeast nitrogenous base, 2% glucose, 0.6% casein hydrolysate, 0.01% adenine). Yeast strain YPH500 was transformed with vector pGPD containing the indicated Tev-mTALEN construct (wildtype or linker variant) or Zif268 and plated on minimal medium agar lacking histidine (His-)(0.75% yeast nitrogenous base, 2% glucose, 0.01% adenine, 0.01% leucine, 0.01% lysine, 0.0025% uracil, 0.005% tryptophan). Single clones of each YPH499 target site transformant and YPH500 expression transformant were grown overnight in Trp-/Ura- and His- liquid media, respectively. In each well of a 96 well plate, 50 μ l of expression strain culture and 50 μ l of target site culture were added to 1ml of YPD medium and allowed to mate at 30⁰C for 4 hours. Cells were then pelleted and resuspended in medium lacking both tryptophan and histidine (His-/Trp-)(0.75% yeast nitrogenous base, 2% glucose, 0.01% adenine, 0.01% leucine, 0.01% lysine, 0.0025% uracil) and grown for 18 hours to select for diploids harbouring both the expression and target plasmids. Following growth of diploid cultures, cells were resuspended in 1ml LacZ reaction buffer (60mM Na₂HPO₄, 40mM NaH₂PO₄, 10mM KCl, 1mM MgSO₄, 35mM β -mercaptoethanol). OD_{595nm} readings were taken by plate reader to determine cell density. Chloroform and SDS were added to 0.06% and 0.01% respectively to lyse the cells. Plates were incubated at 30⁰C for 1 hour after addition of chloroform and SDS. After cells were lysed, ONPG solution was added to a final concentration of 0.3mg/ml and reactions were incubated for a duration of 30 minutes at 30⁰C. Reactions were then stopped by the addition of NaCO₃ to a concentration of 0.2M. Plates were spun at 2000xG for 5 minutes to pellet cellular debris. OD_{405nm} readings were taken on the clarified reaction solutions. All assays were performed in biological triplicates, with 2 technical replicates each.

Activity in each well was measured according to the following equation:

$$\text{Activity} = 2500 * [\text{OD}_{405\text{nm}} - \text{Neg}] / (\text{T} * (\text{D} / 2.5) * \text{V})$$

Where: $[\text{OD}_{405\text{nm}} - \text{Neg}]$ indicates the $\text{OD}_{405\text{nm}}$ reading of the well, minus the $\text{OD}_{405\text{nm}}$ reading of the negative control well for the plate; T indicates the reaction duration (30 minutes), D indicates the cell density of the well, as measured by $\text{OD}_{595\text{nm}}$ readings; and V indicates the volume of the reaction (1ml). Activity measurements were then normalized to the positive control value for the plate.

References

- 1 - Carroll, D. (2011) Genome Engineering With Zinc-Finger Nucleases. *Genetics* **188**, 773-782.
- 2 – Carr, P.A., and Church, G.M. (2009) Genome Engineering. *Nature Biotechnology*, **27**(12), 1151-1162.
- 3 – Hinnen, A., Hicks, J.B., and Fink, G.R. (1978) Transformation of Yeast. *PNAS*, **75**(4), 1929-1933.
- 4 – Capecchi, M., R. (1989) Altering the Genome by Homologous Recombination. *Science* **244**(4910), 1288-1292.
- 5 – Scherer, S., and Davis, R.W. (1979) Replacement of chromosome segments with altered DNA sequences constructed *in vitro*. *PNAS* **76**(10), 4951-4955.
- 6 – Taghian, D.G., and Nickoloff, J.A. (1997) Chromosomal Double-Strand Breaks Induce Gene Conversion at High Frequency in Mammalian Cells. *Molecular and Cellular Biology* **17**(11), 6386-6393.
- 7 – Perez et al (2005) Factors affecting double-strand break-induced homologous recombination in mammalian cells. *Biotechniques* **39**, 109-115.
- 8 – Chang, X., and Wilson, J.H. (1987) Modification of DNA ends can decrease end joining relative to homologous recombination in mammalian cells. *PNAS* **84**, 4959-4963.
- 9 – Xiao et al (1997) Gene Transfer by Adeno-Associated Virus Vectors into the Central Nervous System. *Experimental Neurology* **144**, 113-124.
- 10 – Snyder et al (1997) Persistent and therapeutic concentrations of human factor IX in mice after hepatic gene transfer or recombinant AAV vectors.
- 11 – Flotte et al (1993) Stable *in vivo* expression of the cystic fibrosis transmembrane conductance regulator with an adeno-associated virus vector. *PNAS* **90**, 10613-10617.
- 12 – Acland et al (2001) Gene therapy restores vision in a canine model of childhood blindness. *Nature Genetics* **28**, 92-95.
- 13 – Auricchio, A. (2003) Pseudotyped AAV vectors for constitutive and regulated gene expression in the eye. *Vision Research* **43**, 913-918.
- 14 – Rabinowitz et al (2002) Cross-Packaging of a Single Adeno-Associated Virus (AAV) Type Vector Genome into Multiple AAV Serotypes Enables Transduction with Broad Specificity. *Journal of Virology* **76**(2), 791-801.

- 15 – Yusa, K. (2013) Seamless genome editing in human pluripotent stem cells using custom endonuclease-based gene targeting and the piggyback transposon. *Nature Protocols* **8**, 2061-2078.
- 16 – Geurts et al (2003) Gene Transfer into genomes of human cells by the sleeping beauty transposon system. *Molecular Therapy* **8**(1), 108-117.
- 17 – Martinez-Garcia and de Lorenzo (2012) Transposon-Based and Plasmid-Based Genetic Tools for Editing Genomes of Gram-Negative Bacteria. *Methods in Molecular Biology* **813**, 267-283.
- 18 – Kaiser et al (1997) P-element inserts in transgenic flies: a cautionary tale. *Heredity* **78**, 1-11.
- 19 – Cavazzana-Calvo et al (2000) Gene Therapy of Human Severe Combined Immunodeficiency (SCID)-X1 Disease. *Science* **288**, 669-672.
- 20 – Hacein-Bey-Albine et al (2003) A Serious Adverse Event after Successful Gene Therapy for X-Linker Severe Combined Immunodeficiency. *New England Journal of Medicine* **348**(3), 255-266.
- 21 – Hacein-Bey-Albina et al (2003) LMO2-Associated Clonal T Cell Proliferation in Two Patients after Gene Therapy for SCID-X1. *Science* **302**, 415-419.
- 22 – Davé, U.P., Jenkins, N.A., Copeland, N.G. (2004) Gene Therapy Insertional Mutagenesis Insights. *Science* **303**, 333.
- 23 – Kolodkin et al (1986) Double-Strand Breaks Can Initiate Meiotic Recombination in *S. cerevisiae*. *Cell* **46**, 733-740.
- 24 – Plessis et al (1992) Site-Specific Recombination Determined by I-SceI, a Mitochondrial Group I Intron-Encoded Endonuclease Expressed in the Yeast Nucleus. *Genetics* **130**, 451-460.
- 25 - Rosen et al (2006) Homing endonuclease I-CreI derivatives with novel DNA target specificities. *Nucleic Acids Research* **34**(17), 4791-4800.
- 26 - Baxter et al (2012) Engineering domain fusion chimeras from I-OnuI family LAGLIDAG homing endonucleases. *Nucleic Acids Research* **40**(16), 7985-8000.
- 27 - Kim et al (1996) Hybrid Restriction Enzymes: Zinc Finger Fusions to Fok I Cleavage Domain. *PNAS* **93**, 1156-1160.
- 28 - Pavletich, N.P., and Pabo, C.O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å resolution. *Science* **252**, 809–817.
- 29 – Carroll et al (2006) Design, construction and *in vitro* testing of zinc finger nucleases. *Nature Protocols* **1**, 1329-1341.

- 30 – Bibikova et al (2002) Targeted Chromosomal Cleavage and Mutagenesis in *Drosophila* Using Zinc-Finger Nucleases. *Genetics* **161**, 1169-1175.
- 31 – Kim, Y.G., and Chandresagaren, S. (1994) Chimeric Restriction Endonuclease. *PNAS* **91**, 883-887.
- 32 – Kim et al (1998) Chimeric restriction enzyme: Gal4 fusion to *FokI* cleavage domain. *Biological Chemistry* **379**, 489-405.
- 33 – Sugisaki, H., and Kanazawa, S. (1981) New restriction endonucleases from *Flavobacterium okeanokoites* (FokI) and *Micrococcus luteus* (MluI). *Gene* **16**, 73-78.
- 34 – Bittinaite et al (1998) FokI dimerization is required for DNA cleavage. *PNAS* **95**, 10570-10575.
- 35 - Smith et al (2000) Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nature Biotechnology* **25**(7), 786-793.
- 36 - Ramirez et al (2008) Unexpected failure rates for modular assembly of engineered zinc finger nucleases. *Nature Methods* **5**, 374-375.
- 37 – Kim, J.S., Lee, H.J., and Carroll, D. (2010) Genome editing with modularly assembled zinc-finger nucleases. *Nature Methods* **7**, 91.
- 38 – Beumer et al (2006) Efficient Gene Targeting in *Drosophila* With Zinc-Finger Nucleases. *Genetics* **172**, 2391-2403.
- 39 - Meng et al (2007) Profiling the DNA-binding specificities of engineered Cys2His2 zinc finger domains using a rapid cell-based method. *Nucleic Acids Research* **35**(81), 773-782.
- 40 - Sander et al (2011) Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nature Methods* **8**, 67-79.
- 41 - Mak et al (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* **335**(6069), 716-719.
- 42 - Elrod-Erickson, M., Benson, T.E., and Pabo, C.O. (1998) High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure* **6**, 451-464.
- 43 – Wah et al (1997) Structure of the multimodular endonuclease FokI bound to DNA. *Nature* **388**, 97-100.
- 44 – Tebas et al (2014) Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *New England Journal of Medicine* **370**(10), 901-910.

- 45 - Christian et al (2010) Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics* **186**, 757-761.
- 46 – Boch, J., and Bonas, U. (2010) Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annual Review of Phytopathology* **48**, 419-436.
- 47 – Gurlebeck, D., Thieme, F., and Bonas, U. (2006) Type III effector proteins from the plant pathogen Xanthomonas and their role in the interaction with the host plant. *Journal of Plant Physiology* **163**, 233-255.
- 48 – Deng et al (2012) Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* **335**(6069), 720-723.
- 49 – Boch et al (2009) Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* **326**, 1509-1512.
- 50 – Stella et al (2013) Structure of the AvrBs3-DNA complex provides new insights into the initial thymine-recognition mechanism. *Acta Crystallographica* **69**, 1707-1716.
- 51 – Doyle et al (2013) TAL effectors: highly adaptable phyto-bacterial virulence factors and readily engineered DNA-targeting proteins. *Trends in Cell Biology* **23**(8), 390-398.
- 52 – Doyle et al (2013) Tal Effector Specificity for base 0 of the DNA Target is Altered in a Complex, Effector- and Assay-Dependent Manner by Substitutions for the Tryptophan in Cryptic Repeat -1. *PLOS ONE* **8**(12), e82120.
- 53 – Lamb, B.M., Mercer, A.C., Barbas, C.F. (2013) Directed evolution of the TALE N-terminal domain for recognition of all 5' bases. *Nucleic Acids Research* **41**(21), 9779-9785.
- 54 – Juillerat et al (2014) BurrH: a new modular DNA binding protein for genome engineering. *Scientific Reports* **4**, 3831.
- 55 – Ran et al (2013) Genome Engineering using the CRISPR-Cas9 system. *Nature Protocols* **8**(11), 2281-2308.
- 56 – Horath, P., and Barrangou, R. (2010) CRISPR/Cas, the Immune System of Bacteria and Archae. *Science* **327**, 167-170.
- 57 – Hwang et al (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotech* **31**, 227-229.
- 58 – Nakayama et al (2013) Simple and Efficient CRISPR/Cas9-Mediated Targeted Mutagenesis in *Xenopus tropicalis*. *Genesis* **00**, 1-9.
- 59 – Fu et al (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology* **31**(9), 822-826.

- 60 – Hsu et al (2013) DNA targeting specificity of RNA-guided Cas9 Nucleases. *Nature Biotechnology* **31**, 827-832.
- 61 – Lin et al (2014) CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Research* **42**, 7473-7485.
- 62 – Cradick et al (2013) CRISPR/Cas9 systems targeting β -globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Research* **41**, 9584-9592.
- 63 – Pattanayak et al (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnology* **31**, 839-843.
- 64 – Ran et al (2013) Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154**, 1380-1389.
- 65 – Fu et al (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnology* **32**(2), 279-284.
- 66 – Creating Designed Zinc Finger Nucleases with Minimal Cytotoxicity. *Journal of Molecular Biology* **405**(3), 630-641.
- 67 – Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nature Methods* **8**(1), 74-81.
- 68 – Cade et al (2012) Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. *Nucleic Acids Research* **40**(16), 8001-8010.
- 69 – Szczepek et al (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nature Biotechnology* **25**(7), 786-793.
- 70 – Miller et al (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. *Nature Biotechnology* **25**, 778-785.
- 71 – Kleinstiver et al (2012) Monomeric site-specific nucleases for genome editing. *PNAS* **109**(21), 8061-8066.
- 72 – Stoddard, B. (2006) Homing endonuclease structure and function. *Quarterly Reviews of Biophysics* **38**(1), 49–95.
- 73 – Michel, F., and Dujon, B. (1986) Genetic exchanges between bacteriophage T4 and filamentous fungi? *Cell* **46**(3), 323.
- 74 – Bell-Pedersen et al (1991) I-TevI, the endonuclease encoded by the mobile td intron, recognizes binding and cleavage domains on its DNA target. *PNAS* **88**, 7719-7723.
- 75 – Bell-Pedersen et al (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 Research* **18**(13), 3763-3770.

76 – Van Roey, P., Belfort, M., and Derbyshire, V. (2000) *Homing Endonuclease I-TevI: An Atypical Zinc Finger with a Novel Function*. Madame Curie Bioscience Database, Landes Biosciences, Austin, Texas.

77 – Edgell et al (2004) Coincidence of Cleavage Sites of Intron Endonuclease I-TevI and Critical Sequences of the Host Thymidylate Synthase Gene. *Journal of Molecular Biology* **343**, 1231-1241.

78 – Kleinstiver et al (2014) The I-TevI nuclease and linker domains contribute to the specificity of monomeric TALENs. *Genes/Genomes/Genetics* **4**(6), 1155-1165.

79 – Wolfs et al (2014) MegaTevs: single-chain dual nucleases for efficient gene disruption. *Nucleic Acids Research* **42**(13), 8816-8829.

80 – Van Roey et al (2001) Intertwined structure of the DNA-binding domain of intron endonuclease I-TevI with its substrate. *EMBO* **20**, 3631-3637.

81 – Dean et al (2002) Zinc finger as distance determinant in the flexible linker of intron endonuclease I-TevI. *PNAS* **99**(13), 8554-8561. *Nature Structural and Molecular Biology* **11**(10), 936-944.

82 – Robbins et al (2011) Redox-Responsive Zinc Finger Fidelity Switch in Homing Endonuclease and Intron Promiscuity in Oxidative Stress. *Current Biology* **21**, 243-248.

83 – Liu et al (2008) Role of the Interdomain Linker in Distance Determination for Remote Cleavage by Homing Endonuclease I-TevI. *Journal of Molecular Biology* **379**, 1094-1106.

84 – Liu et al (2006) Distance determination by GIY-YIG intron endonucleases: discrimination between repression and cleavage functions. *Nucleic Acids Research* **34**(6), 1755-1764.

85 – Derbyshire et al (1997) Two-domain Structure of the td Intron-encoded Endonuclease I-TevI Correlates with the Two-domain Configuration of the Homing Site. *Journal of Molecular Biology* **265**, 494-506.

86 - Mueller et al (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.

87 – Beurdeley et al (2013) Compact designer TALENs for efficient genome engineering. *Nature Communications* **4**(1762).

88 – Van Roey et al (2002) Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI. *Nature Structural Biology*. **9**, 806-811.

- 89 – Bryk et al (1993) The td intron endonuclease I-TevI makes extensive sequence-tolerant contacts across the minor groove of its DNA target. *The EMBO Journal* **12**(5), 2141-2149.
- 90 – Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E. (2004) WebLogo: A sequence logo generator. *Genome Research* **14**, 1188-1190.
- 91 – Miller et al (2011) A TALE nuclease architecture for efficient genome editing. *Nature Biotechnology* **29**(2), 143-150.
- 92 – 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. *Current Biology* **13**, 973-978.
- 93 - Lozios et al (1996) Intron-encoded Endonuclease I-TevII Binds Across the Minor Groove and Induces Two Distinct Conformational Changes in its DNA Substrate. *Journal of Molecular Biology* **255**, 412-424.
- 94 – Nord et al (A Functional Homing Endonuclease in the Bacillus anthracis nrdE Group I Intron. *Journal of Bacteriology* **189**(14), 5293-5301.
- 95 – Unconventional GIY-YIG homing endonuclease encoded in group I introns in closely related strains of the Bacillus cereus group. *Nucleic Acids Research* **36**(1), 300-310.
- 96 – Sitbon, E., and Pietrokovski, S. (2003) New types of conserved sequence domains in DNA-binding regions of homing endonucleases. *Trends in Biochemical Science* **28**(9), 473-477.
- 97 – Seeman, N.C., Rosenberg, J.M., and Rich, A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins. *PNAS* **73**, 804-808.
- 98 – Bewley, C.A, Gronenborn, A.M., and Clore, G.M (1998) MINOR GROOVE-BINDING ARCHITECTURAL PROTEINS: Structure, Function, and DNA Recognition. *Annual Review of Biophysical and Biomolecular Structure* **27**, 105-131.
- 99 – Strauss, J.K., and Maher III, L.J. (1994) DNA Bending by Asymmetric Phosphate Neutralization. *Science* **266**, 1829-1834.
- 100 – Bryk et al (1995) Selection of a Remote Cleavage Site by I-TevI, the td Intron-encoded Endonuclease. *Journal of Molecular Biology* **247**, 197-210.
- 101 – Gartenberg, M.R., and Crothers, D.M. (1988) DNA sequence determinants of CAP-induced bending and protein binding affinity. *Nature* **330**, 824-829.
- 102 – Starr, D.B, Hoopes, B.C., and Hawley, D.K. (1995) DNA Bending is an Important Component of Site-specific Recognition by the TATA Binding Protein. *Journal of Molecular Biology* **250**, 434-446.

- 103 – Sarai et al (1988) Origin of DNA Helical Structure and Its Sequence Dependence. *Biochemistry* **27**, 8498-8502.
- 104 – Cermak et al (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research* **39**(12), e82.
- 105 – Reyon et al (2012) FLASH Assembly of TALENs Enables High-Throughput Genome Editing. *Nature Biotechnology* **30**(5), 460-465.
- 106 – Elliot et al (1998) Gene conversion tracts from double-strand break repair in mammalian cells. *Molecular Cell Biology* **18**, 93-101.
- 107 – Yang et al (2013) Optimization of scarless human stem cell genome editing. *Nucleic Acids Research* **41**, 9049-9061.
- 108 – Findlay et al (2014) Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* **513**, 120-123.
- 109 – Antonarakis, S.E. (2010) *Human Genome Sequence and Variation*. Vogel and Motulsky's Human Genetics: Problems and Approaches, Springer-Verlag, Berlin, Germany.
- 110 – Cho et al (2014) Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Research* **24**, 132-141.
- 111 – Kleinstiver et al (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, 14592.

Appendices

Appendix A: Strains and Plasmids

Strains	Description	Source
<i>E. coli</i> - ER2566	F- λ - fhuA2 [lon] ompT lacZ::T7 gene 1 gal sulA11 Δ (mcrC-mrr)114::IS10 R(mcr-73::miniTn10-TetS)2 R(zgb-210::Tn10)(TetS) endA1 [dcm]	New England Biolabs
<i>E. coli</i> – DH5 α	F' endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ 80dlacZ Δ M15 Δ (lacZYA-argF)U169, hsdR17(r_K m_K^+), λ -	Invitrogen
<i>S. cerevisiae</i> – YPH499	MATa ura3-52 lys2-801_amber ade2-101_ochre trp1- Δ 63 his3- Δ 200 leu2- Δ 1	Dr Adam Bogdanove (90)
<i>S. cerevisiae</i> – YPH500	MATa ura3-52 lys2-801_amber ade2-101_ochre trp1- Δ 63 his3- Δ 200 leu2- Δ 1	Dr Adam Bogdanove (90)
Target-site Plasmids	Description	Source
pSP72	Amp ^R , pBR322 origin	Promega
pSP72-TP15	pSP72 with the TP15 target site ligated in via BglII/XbaI	This report
pCP5.1	Derivative of vector pCP5, Amp ^R , ColE1 origin, target-site plasmid used in yeast reporter assays, nuclease target sites are cloned in between a fragmented LacZ gene via restriction sites BglII/SpeI, contains the selectable markers URA3 and TRP1	Dr Adam Bogdanove (90)
pCP5.1-TP15	pCP5.1 containing the TP15 target site	This report
pCP5.1-ZF	pCP5.1 containing the Zif268 target site	Dr Adam Bogdanove (90)
pCP5.1-TO15	pCP5.1 containing the TO15 (Mega-Tev) target site	Jason Wolfs
pCP5.1-TP15(TAACA)	pCP5.1 containing the TP15 target site with the mutated cleavage motif, TAACA	This report
pCP5.1-TP15(TAACG)	pCP5.1 containing the TP15 target site with the mutated cleavage motif, TAACG	This report
pCP5.1-TP15(CAACA)	pCP5.1 containing the TP15 target site with the mutated cleavage motif, CAACA	This report
pCP5.1-TP15(C1A)	pCP5.1 containing the TP15 target site with the C1A DNA spacer mutation	This report
pCP5.1-TP15(C1G)	pCP5.1 containing the TP15 target site with the C1G DNA spacer mutation	This report
pCP5.1-TP15(C1T)	pCP5.1 containing the TP15 target site with the C1T DNA spacer mutation	This report
pCP5.1-TP15(T2G)	pCP5.1 containing the TP15 target site with the T2G DNA spacer mutation	This report
pCP5.1-TP15(T2C)	pCP5.1 containing the TP15 target site with the T2C DNA spacer mutation	This report
pCP5.1-TP15(T2A)	pCP5.1 containing the TP15 target site with the T2A DNA spacer mutation	This report
pCP5.1-TP15(C3G)	pCP5.1 containing the TP15 target site with the C3G DNA spacer mutation	This report
pCP5.1-TP15(C3A)	pCP5.1 containing the TP15 target site with the C3A DNA spacer mutation	This report
pCP5.1-TP15(C3T)	pCP5.1 containing the TP15 target site with the C3T DNA spacer mutation	This report
pCP5.1-TP15(A4G)	pCP5.1 containing the TP15 target site with the A4G DNA spacer mutation	This report

pCP5.1-TP15(A4C)	pCP5.1 containing the TP15 target site with the A4C DNA spacer mutation	This report
pCP5.1-TP15(A4T)	pCP5.1 containing the TP15 target site with the A4T DNA spacer mutation	This report
pCP5.1-TP15(G5C)	pCP5.1 containing the TP15 target site with the G5C DNA spacer mutation	This report
pCP5.1-TP15(G5A)	pCP5.1 containing the TP15 target site with the G5A DNA spacer mutation	This report
pCP5.1-TP15(G5T)	pCP5.1 containing the TP15 target site with the G5T DNA spacer mutation	This report
pCP5.1-TP15(T6G)	pCP5.1 containing the TP15 target site with the T6G DNA spacer mutation	This report
pCP5.1-TP15(T6C)	pCP5.1 containing the TP15 target site with the T6C DNA spacer mutation	This report
pCP5.1-TP15(T6A)	pCP5.1 containing the TP15 target site with the T6A DNA spacer mutation	This report
pCP5.1-TP15(A7G)	pCP5.1 containing the TP15 target site with the A7G DNA spacer mutation	This report
pCP5.1-TP15(A7C)	pCP5.1 containing the TP15 target site with the A7C DNA spacer mutation	This report
pCP5.1-TP15(A7T)	pCP5.1 containing the TP15 target site with the A7T DNA spacer mutation	This report
pCP5.1-TP15(G8C)	pCP5.1 containing the TP15 target site with the G8C DNA spacer mutation	This report
pCP5.1-TP15(G8A)	pCP5.1 containing the TP15 target site with the G8A DNA spacer mutation	This report
pCP5.1-TP15(G8T)	pCP5.1 containing the TP15 target site with the G8T DNA spacer mutation	This report
pCP5.1-TP15(A9G)	pCP5.1 containing the TP15 target site with the A9G DNA spacer mutation	This report
pCP5.1-TP15(A9C)	pCP5.1 containing the TP15 target site with the A9C DNA spacer mutation	This report
pCP5.1-TP15(A9T)	pCP5.1 containing the TP15 target site with the A9T DNA spacer mutation	This report
pCP5.1-TP15(T10G)	pCP5.1 containing the TP15 target site with the T10G DNA spacer mutation	This report
pCP5.1-TP15(T10C)	pCP5.1 containing the TP15 target site with the T10C DNA spacer mutation	This report
pCP5.1-TP15(T10A)	pCP5.1 containing the TP15 target site with the T10A DNA spacer mutation	This report
pCP5.1-TP15(G11C)	pCP5.1 containing the TP15 target site with the G11C DNA spacer mutation	This report
pCP5.1-TP15(G11A)	pCP5.1 containing the TP15 target site with the G11A DNA spacer mutation	This report
pCP5.1-TP15(G11T)	pCP5.1 containing the TP15 target site with the G11T DNA spacer mutation	This report
pCP5.1-TP15(T12G)	pCP5.1 containing the TP15 target site with the T12G DNA spacer mutation	This report
pCP5.1-TP15(T12C)	pCP5.1 containing the TP15 target site with the T12C DNA spacer mutation	This report
pCP5.1-TP15(T12A)	pCP5.1 containing the TP15 target site with the T12A DNA spacer mutation	This report
pCP5.1-TP15(T13G)	pCP5.1 containing the TP15 target site with the T13G DNA spacer mutation	This report
pCP5.1-TP15(T13C)	pCP5.1 containing the TP15 target site with the T13C DNA spacer mutation	This report
pCP5.1-TP15(T13A)	pCP5.1 containing the TP15 target site with the T13A DNA spacer mutation	This report
pCP5.1-TP15(T14G)	pCP5.1 containing the TP15 target site with the T14G DNA spacer mutation	This report

pCP5.1-TP15(T14C)	pCP5.1 containing the TP15 target site with the T14C DNA spacer mutation	This report
pCP5.1-TP15(T14A)	pCP5.1 containing the TP15 target site with the T14A DNA spacer mutation	This report
pCP5.1-TP15(T15G)	pCP5.1 containing the TP15 target site with the T15G DNA spacer mutation	This report
pCP5.1-TP15(T15C)	pCP5.1 containing the TP15 target site with the T15C DNA spacer mutation	This report
pCP5.1-TP15(T15A)	pCP5.1 containing the TP15 target site with the T15A DNA spacer mutation	This report
pCP5.1-N15	pCP5.1 with the library of N15 DNA spacer sequences (oligo DE1333)	This report
Nuclease Expression Plasmids	Description	Source
pACYC.PciI	Modified version of pACYC-Duet1 (Novagen) with the NcoI site in MCS1 replaced with a PciI site, constructs are cloned in and expressed via restriction sites PciI/XhoI, Cm ^R	Ben Kleinstiver
pACYC.Pci-N169T120	pACYC-PciI, containing the N169-T120 Tev-mTALEN, cloned in via PciI and XhoI	Ben Kleinstiver
pACYC.Pci-N169T120(12RVD)	pACYC-PciI, containing the 6x his-tagged N169-T120 Tev-mTALEN with 12RVDs, cloned in via PciI and XhoI	Ben Kleinstiver
pACYC.Pci-N169T120(S134G)	pACYC-PciI expressing the S134G Tev-mTALEN mutant	This report
pACYC.Pci-N169T120(S134G/N140S)	pACYC-PciI expressing the S134G/N140S Tev-mTALEN mutant	This report
pACYC.Pci-N169T120(K135R/N140S/Q158R)	pACYC-PciI expressing the K135R/N140S/Q158R Tev-mTALEN mutant	This report
pACYC.Pci-N169T120(V117F/D127)	pACYC-PciI expressing the V117F/D127G Tev-mTALEN mutant	This report
pACYC.Pci-N169ONU(S134G)	pACYC-PciI expressing the S134G Mega-Tev mutant	Jason Wolfs
pACYC.Pci-N169ONU(S134G/N140S)	pACYC-PciI expressing the S134G/N140S Mega-Tev mutant	Jason Wolfs
pACYC.Pci-N169ONU(K135R/N140S/Q158R)	pACYC-PciI expressing the K135R/N140S/Q158R Mega-Tev mutant	Jason Wolfs
pACYC.Pci-N169ONU(V117F/D127G)	pACYC-PciI expressing the V117F/D127G Mega-Tev mutant	Jason Wolfs
pGPD	Modified version of yeast expression vector p416GPD, Amp ^R , nuclease constructs are cloned in via PciI/XhoI sites, contains the selectable HIS3 marker	Dr Adam Bogdanove (90)
pGPD-N169T120 (pLWN37)	pGPD expressing the N169-T120 Tev-mTALEN	Dr Adam Bogdanove
pGPD-Zif268	pGPD expressing the dimeric Zif268 ZFN	Dr Adam Bogdanove (90)
pGPD-N169T120(S134G)	pGPD expressing the S134G Tev-mTALEN	This report
pGPD-N169T120(S134G/N140S)	pGPD expressing the S134G/N140S Tev-mTALEN	This report
pGPD-K135R/N140S/Q158R	pGPD expressing the K135R/N140S/Q158R Tev-mTALEN	This report
pGPD-V117F/D127G	pGPD expressing the V117F/D127G Tev-mTALEN	This report

Appendix B: Oligonucleotides

Oligonucleotides	Sequence (5'-3' direction)	Description
DE1013	GCGACATGTCTAAAAGCGGAATTTATCA GATT	Forward PCR primer for I-TevI (1-169) in pACYC-Pci, with 5'PciI overhang
DE1114	GTCGTTAGAACGCGGC	Reverse sequencing/PCR primer for pSP72
DE1175	GCAACGTTTGGTGATACGTGTTCTACGCA TCCATTA AAAAG	Site-directed mutagenesis primer to eliminate the PciI site in I-TevI
DE1176	CTTTTAATGGATGCGTAGAACACGTATCA CCAAACGTTGC	Site-directed mutagenesis primer to eliminate the PciI site in I-TevI
DE1213	CGCGGATCCACCAGAACACCATTCTG CATTTACTACAAG	Reverse PCR primer for I-TevI (1-169) in pACYC-Pci, with 3' BamHI
DE1333	GCAATGAGATCTCAACGNNNNNNNNNN NNNNNTGCATCTCCCATTACTGTAAAACA C	Upper oligo for the N15 DNA spacer library
DE1452	CGTATTACCGCCTTTGAG	Forward sequencing/PCR primer for pSP72
DE1496	GCTATGACTAGTGTGTTTTACAGTAATGG GAGA	Lower strand extension primer for oligos DE1333 and DE1715-DE1729 (SpeI)
DE1546	GATCTGCCAACGATCAGTAGATGTTTTG CATCTCCCATTACTGTAAAACACA	Top strand oligo for the TP15(C1A) target (BglII/SpeI)
DE1547	CTAGTGTGTTTTACAGTAATGGGAGATGC AAAAACATCTACTGATCGTTGGCA	Bottom strand oligo for the TP15(C1A) target (BglII/SpeI)
DE1548	GATCTGCCAACGGTCAGTAGATGTTTTG CATCTCCCATTACTGTAAAACACA	Top strand oligo for the TP15(C1G) target (BglII/SpeI)
DE1549	CTAGTGTGTTTTACAGTAATGGGAGATGC AAAAACATCTACTGACCGTTGGCA	Bottom strand oligo for the TP15(C1G) target (BglII/SpeI)
DE1550	GATCTGCCAACGTTCCAGTAGATGTTTTG CATCTCCCATTACTGTAAAACACA	Top strand oligo for the TP15(C1T) target (BglII/SpeI)
DE1551	CTAGTGTGTTTTACAGTAATGGGAGATGC AAAAACATCTACTGAACGTTGGCA	Bottom strand oligo for the TP15(C1T) target (BglII/SpeI)
DE1612	GATCTTAACACTCAGTAGATGTTTTGCA TCTCCCATTACTGTAAAACACA	Top strand oligo for the TP15(TAACA) target (BglII/SpeI)
DE1613	CTAGTGTGTTTTACAGTAATGGGAGATGC AAAAACATCTACTGAGTGTTAA	Bottom strand oligo for the TP15(TAACA) target (BglII/SpeI)
DE1736	GATCTCAACACTCAGTAGATGTTTTGCA TCTCCCATTACTGTAAAACACA	Top strand oligo for the TP15(CAACA) target (BglII/SpeI)
DE1737	CTAGTGTGTTTTACAGTAATGGGAGATGC AAAAACATCTACTGAGTGTTGA	Bottom strand oligo for the TP15(CAACA) target (BglII/SpeI)
DE1715	GCAATGAGATCTCAACGCGCAGTAGATG TTTTTGCATCTCCCATTACTGTAAAACAC	Top strand extension oligo for the TP15(T2G) target BglII
DE1716	GCAATGAGATCTCAACGCCAGTAGATG TTTTTGCATCTCCCATTACTGTAAAACAC	Top strand extension oligo for the TP15(T2C) target BglII
DE1717	GCAATGAGATCTCAACGCACAGTAGATG TTTTTGCATCTCCCATTACTGTAAAACAC	Top strand extension oligo for the TP15(T2A) target BglII
DE1718	GCAATGAGATCTCAACGCTGAGTAGATG TTTTTGCATCTCCCATTACTGTAAAACAC	Top strand extension oligo for the TP15(C3G) target BglII
DE1719	GCAATGAGATCTCAACGCTAAGTAGATG TTTTTGCATCTCCCATTACTGTAAAACAC	Top strand extension oligo for the TP15(C3A) target BglII
DE1720	GCAATGAGATCTCAACGCTTAGTAGATG TTTTTGCATCTCCCATTACTGTAAAACAC	Top strand extension oligo for the TP15(C3T) target BglII
DE1721	GCAATGAGATCTCAACGCTCGGTAGATG TTTTTGCATCTCCCATTACTGTAAAACAC	Top strand extension oligo for the TP15(A4G) target BglII
DE1722	GCAATGAGATCTCAACGCTCCGTAGATG TTTTTGCATCTCCCATTACTGTAAAACAC	Top strand extension oligo for the TP15(A4C) target BglII
DE1723	GCAATGAGATCTCAACGCTCTGTAGATG TTTTTGCATCTCCCATTACTGTAAAACAC	Top strand extension oligo for the TP15(A4T) target BglII

DE1724	GCAATGAGATCTCAACGCTCACTAGATG TTTTTGCATCTCCCATTACTGTAAAACAC	Top strand extension oligo for the TP15(G5C) target BglII
DE1725	GCAATGAGATCTCAACGCTCAATAGATG TTTTTGCATCTCCCATTACTGTAAAACAC	Top strand extension oligo for the TP15(G5A) target BglII
DE1726	GCAATGAGATCTCAACGCTCATTAGATG TTTTTGCATCTCCCATTACTGTAAAACAC	Top strand extension oligo for the TP15(G5T) target BglII
DE1727	GCAATGAGATCTCAACGCTCAGGAGATG TTTTTGCATCTCCCATTACTGTAAAACAC	Top strand extension oligo for the TP15(T6G) target BglII
DE1728	GCAATGAGATCTCAACGCTCAGCAGATG TTTTTGCATCTCCCATTACTGTAAAACAC	Top strand extension oligo for the TP15(T6C) target BglII
DE1729	GCAATGAGATCTCAACGCTCAGAAAGATG TTTTTGCATCTCCCATTACTGTAAAACAC	Top strand extension oligo for the TP15(T6A) target BglII
DE1734	GATCTTAACGCTCAGTAGATGTTTTTGCA TCTCCCATTACTGTAAAACACA	Top strand oligo for the TP15(TAACG) target (BglII/SpeI)
DE1735	CTAGTGTGTTTTACAGTAATGGGAGATGC AAAACATCTACTGAGCGTTAA	Bottom strand oligo for the TP15(TAACG) target (BglII/SpeI)
DE1738	GATCTCAACGCTCAGTGGATGTTTTTGCA TCTCCCATTACTGTAAAACACA	Upper oligo for the TP15(A7G) target BglII/SpeI
DE1739	CTAGTGTGTTTTACAGTAATGGGAGATGC AAAACATCCACTGAGCGTTGA	Lower oligo for the TP15(A7G) target BglII/SpeI
DE1740	GATCTCAACGCTCAGTCGATGTTTTTGCA TCTCCCATTACTGTAAAACACA	Upper oligo for the TP15(A7C) target BglII/SpeI
DE1741	CTAGTGTGTTTTACAGTAATGGGAGATGC AAAACATCGACTGAGCGTTGA	Lower oligo for the TP15(A7C) target BglII/SpeI
DE1742	GATCTCAACGCTCAGTTGATGTTTTTGCA TCTCCCATTACTGTAAAACACA	Upper oligo for the TP15(A7T) target BglII/SpeI
DE1743	CTAGTGTGTTTTACAGTAATGGGAGATGC AAAACATCAACTGAGCGTTGA	Lower oligo for the TP15(A7T) target BglII/SpeI
DE1744	GATCTCAACGCTCAGTACATGTTTTTGCA TCTCCCATTACTGTAAAACACA	Upper oligo for the TP15(G8C) target BglII/SpeI
DE1745	CTAGTGTGTTTTACAGTAATGGGAGATGC AAAACATGTACTGAGCGTTGA	Lower oligo for the TP15(G8C) target BglII/SpeI
DE1746	GATCTCAACGCTCAGTAAATGTTTTTGCA TCTCCCATTACTGTAAAACACA	Upper oligo for the TP15(G8A) target BglII/SpeI
DE1747	CTAGTGTGTTTTACAGTAATGGGAGATGC AAAACATTTACTGAGCGTTGA	Lower oligo for the TP15(G8A) target BglII/SpeI
DE1748	GATCTCAACGCTCAGTATATGTTTTTGCA TCTCCCATTACTGTAAAACACA	Upper oligo for the TP15(G8T) target BglII/SpeI
DE1749	CTAGTGTGTTTTACAGTAATGGGAGATGC AAAACATATACTGAGCGTTGA	Lower oligo for the TP15(G8T) target BglII/SpeI
DE1750	GATCTCAACGCTCAGTAGGTGTTTTTGCA TCTCCCATTACTGTAAAACACA	Upper oligo for the TP15(A9G) target BglII/SpeI
DE1751	CTAGTGTGTTTTACAGTAATGGGAGATGC AAAACACCTACTGAGCGTTGA	Lower oligo for the TP15(A9G) target BglII/SpeI
DE1752	GATCTCAACGCTCAGTAGCTGTTTTTGCA TCTCCCATTACTGTAAAACACA	Upper oligo for the TP15(A9C) target BglII/SpeI
DE1753	CTAGTGTGTTTTACAGTAATGGGAGATGC AAAACAGCTACTGAGCGTTGA	Lower oligo for the TP15(A9C) target BglII/SpeI
DE1754	GATCTCAACGCTCAGTAGTTGTTTTTGCA TCTCCCATTACTGTAAAACACA	Upper oligo for the TP15(A9T) target BglII/SpeI
DE1755	CTAGTGTGTTTTACAGTAATGGGAGATGC AAAACAACCTACTGAGCGTTGA	Lower oligo for the TP15(A9T) target BglII/SpeI
DE1756	GATCTCAACGCTCAGTAGAGGTTTTTGCA ATCTCCCATTACTGTAAAACACA	Upper oligo for the TP15(T10G) target BglII/SpeI
DE1757	CTAGTGTGTTTTACAGTAATGGGAGATGC AAAACCTCTACTGAGCGTTGA	Lower oligo for the TP15(T10G) target BglII/SpeI
DE1758	GATCTCAACGCTCAGTAGACGTTTTTGCA TCTCCCATTACTGTAAAACACA	Upper oligo for the TP15(T10C) target BglII/SpeI

DE1759	<u>CTAGTGTGTTTTACAGTAATGGGAGATGC</u> AAAAACGTCTACTGAGCGTTGA	Lower oligo for the TP15(T10C) target BglII/SpeI
DE1760	GATCTCAACGCTCAGTAGAAGTTTTTGC ATCTCCCACTACTGTA AACACA	Upper oligo for the TP15(T10A) target BglII/SpeI
DE1761	<u>CTAGTGTGTTTTACAGTAATGGGAGATGC</u> AAAAACTTCTACTGAGCGTTGA	Lower oligo for the TP15(T10A) target BglII/SpeI
DE1762	GATCTCAACGCTCAGTAGATCTTTTTGCA TCTCCCACTACTGTA AACACA	Upper oligo for the TP15(G11C) target BglII/SpeI
DE1763	<u>CTAGTGTGTTTTACAGTAATGGGAGATGC</u> AAAAAGATCTACTGAGCGTTGA	Lower oligo for the TP15(G11C) target BglII/SpeI
DE1764	GATCTCAACGCTCAGTAGATTTTTTGC TCTCCCACTACTGTA AACACA	Upper oligo for the TP15(G11A) target BglII/SpeI
DE1765	<u>CTAGTGTGTTTTACAGTAATGGGAGATGC</u> AAAAATATCTACTGAGCGTTGA	Lower oligo for the TP15(G11A) target BglII/SpeI
DE1766	GATCTCAACGCTCAGTAGATTTTTTGC TCTCCCACTACTGTA AACACA	Upper oligo for the TP15(G11T) target BglII/SpeI
DE1767	<u>CTAGTGTGTTTTACAGTAATGGGAGATGC</u> AAAAAAATCTACTGAGCGTTGA	Lower oligo for the TP15(G11T) target BglII/SpeI
DE1768	GATCTCAACGCTCAGTAGATGGTTTTGC ATCTCCCACTACTGTA AACACA	Upper oligo for the TP15(T12G) target BglII/SpeI
DE1769	<u>CTAGTGTGTTTTACAGTAATGGGAGATGC</u> AAAACCATCTACTGAGCGTTGA	Lower oligo for the TP15(T12G) target BglII/SpeI
DE1770	GATCTCAACGCTCAGTAGATGCTTTTTGCA TCTCCCACTACTGTA AACACA	Upper oligo for the TP15(T12C) target BglII/SpeI
DE1771	<u>CTAGTGTGTTTTACAGTAATGGGAGATGC</u> AAAAGCATCTACTGAGCGTTGA	Lower oligo for the TP15(T12C) target BglII/SpeI
DE1772	GATCTCAACGCTCAGTAGATGATTTTTGC ATCTCCCACTACTGTA AACACA	Upper oligo for the TP15(T12A) target BglII/SpeI
DE1773	<u>CTAGTGTGTTTTACAGTAATGGGAGATGC</u> AAAATCATCTACTGAGCGTTGA	Lower oligo for the TP15(T12A) target BglII/SpeI
DE1774	GATCTCAACGCTCAGTAGATGTGTTTTGC ATCTCCCACTACTGTA AACACA	Upper oligo for the TP15(T13G) target BglII/SpeI
DE1775	<u>CTAGTGTGTTTTACAGTAATGGGAGATGC</u> AAACACATCTACTGAGCGTTGA	Lower oligo for the TP15(T13G) target BglII/SpeI
DE1776	GATCTCAACGCTCAGTAGATGTCTTTGCA TCTCCCACTACTGTA AACACA	Upper oligo for the TP15(T13C) target BglII/SpeI
DE1777	<u>CTAGTGTGTTTTACAGTAATGGGAGATGC</u> AAAGACATCTACTGAGCGTTGA	Lower oligo for the TP15(T13C) target BglII/SpeI
DE1778	GATCTCAACGCTCAGTAGATGTATTTGC ATCTCCCACTACTGTA AACACA	Upper oligo for the TP15(T13A) target BglII/SpeI
DE1779	<u>CTAGTGTGTTTTACAGTAATGGGAGATGC</u> AAATACATCTACTGAGCGTTGA	Lower oligo for the TP15(T13A) target BglII/SpeI
DE1780	GATCTCAACGCTCAGTAGATGTTGTTGC ATCTCCCACTACTGTA AACACA	Upper oligo for the TP15(T14G) target BglII/SpeI
DE1781	<u>CTAGTGTGTTTTACAGTAATGGGAGATGC</u> ACAACATCTACTGAGCGTTGA	Lower oligo for the TP15(T14G) target BglII/SpeI
DE1782	GATCTCAACGCTCAGTAGATGTTCTTGCA TCTCCCACTACTGTA AACACA	Upper oligo for the TP15(T14C) target BglII/SpeI
DE1783	<u>CTAGTGTGTTTTACAGTAATGGGAGATGC</u> AGAACATCTACTGAGCGTTGA	Lower oligo for the TP15(T14C) target BglII/SpeI
DE1784	GATCTCAACGCTCAGTAGATGTTATTGC ATCTCCCACTACTGTA AACACA	Upper oligo for the TP15(T14A) target BglII/SpeI
DE1785	<u>CTAGTGTGTTTTACAGTAATGGGAGATGC</u> AATAACATCTACTGAGCGTTGA	Lower oligo for the TP15(T14A) target BglII/SpeI
DE1786	GATCTCAACGCTCAGTAGATGTTGTGC ATCTCCCACTACTGTA AACACA	Upper oligo for the TP15(T15G) target BglII/SpeI
DE1787	<u>CTAGTGTGTTTTACAGTAATGGGAGATGC</u> ACAACATCTACTGAGCGTTGA	Lower oligo for the TP15(T15G) target BglII/SpeI

DE1788	GATCTCAACGCTCAGTAGATGTTTCTGCA TCTCCCACTACTGAAAAACA <u>A</u>	Upper oligo for the TP15(T15C) target BglII/SpeI
DE1789	<u>CTAGTGTGTTTTACAGTAATGGGAGATGC</u> AGAAACATCTACTGAGCGTTGA	Lower oligo for the TP15(T15C) target BglII/SpeI
DE1790	GATCTCAACGCTCAGTAGATGTTTATGC ATCTCCCACTACTGAAAAACA <u>A</u>	Upper oligo for the TP15(T15A) target BglII/SpeI
DE1791	<u>CTAGTGTGTTTTACAGTAATGGGAGATGC</u> ATAAACATCTACTGAGCGTTGA	Lower oligo for the TP15(T15A) target BglII/SpeI
DE1811	GATCTCAACGCTCAGTAGATGTTTTGCA TCTCCCACTACTGAAAAACA <u>A</u>	Top strand oligo for the TP15 target (BglII/SpeI)
DE1812	<u>CTAGTGTGTTTTACAGTAATGGGAGATGC</u> AAAAACATCTACTGAGCGTTGA	Bottom strand oligo for the TP15 target (BglII/SpeI)

Appendix C: Supplementary Data Tables

Enzyme	Target	Average Normalized Activity	Standard Error
Zif268	ZF	100.00%	52.23%
Zif268	TP15	0%	31.06%
Tev-mTALEN	ZF	234.17%	82.01%
Tev-mTALEN	TP15	3.22%	26.18%

Supplementary Table S1: Figure 2.1A Data

Target	Average Normalized Activity	Standard Error
TP15	100.00%	42.74%
pCP5a	0.00%	2.85%
Zif	0.39%	3.30%
TO15	0.00%	2.23%
TNNNA	0.00%	0.44%
TNNNG	0.00%	1.29%
CNNNA	0.98%	3.81%

Supplementary Table S2: Figure 2.1B Data

Time (min)	1:1 Ratio		2:1 Ratio	
	pSP72	pSP72-TP15	pSP72	pSP72-TP15
0	0%	0%	0%	0%
3	0%	1.41%	2.8%	51.9%
7	0%	22.17%	7.3%	77.1%
15	0.6%	31.65%	15.6%	87.1%
20	1.08%	35.53%	17.1%	88.4%

Supplementary Table S3: Figure 2.3 Data

	pSP72					pSP72-TP15				
NaCl	R1	R2	R3	Avg	StDev	R1	R2	R3	Avg	StDev
25mM	25%	51%	40%	38.7%	13.1%	100%	100%	100%	100%	100%
50mM	27%	45%	35%	35.7%	9%	100%	100%	75%	91.6%	91.7%
75mM	41%	30%	13%	28%	14.1%	100%	100%	72%	90.6%	90.7%
100mM	10%	0%	0%	3.3%	5.8%	60%	30%	47%	45.6%	45.7%
150mM	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
KCl	R1	R2	R3	Avg	StDev	R1	R2	R3	Avg	StDev
25mM	76%	69%	70%	71.6%	3.7%	100%	100%	100%	100%	100%
50mM	73%	57%	52%	60.7%	11%	100%	100%	100%	100%	100%
75mM	62%	53%	40%	51.7%	11.1%	100%	100%	100%	100%	100%
100mM	36%	36%	17%	29.8%	11.1%	100%	100%	100%	100%	100%
150mM	1%	0%	0%	0.5%	0.8%	19%	21%	0%	13.4%	13.4%

Supplementary Table S4: Figure 2.4 Data

	- Poly dI/dC		+ 20ng/μl Poly dI/dC	
	Avg	StDev	Avg	StDev
pSP72	25.25%	93.75%	1.3%	77%
pSP72-TP15	6.1%	7.5%	2.3%	2.6%

Supplementary Table S5: Figure 2.5 Data

Target	Plate 1	Plate 2	Plate 3	Average	Standard Error
TP15	100.00%	100.00%	100.00%	100.00%	0.00%
pCP5a	0.00%	0.00%	0.00%	0.00%	0.00%
C1A	11.29%	1.52%	28.07%	13.62%	13.43%
C1G	286.50%	145.64%	341.06%	257.73%	100.83%
C1T	44.47%	21.18%	31.94%	32.53%	11.66%
T2G	-2.04%	2.31%	-0.04%	0.08%	2.18%
T2C	10.01%	2.85%	5.73%	6.20%	3.60%
T2A	28.58%	13.56%	15.50%	19.21%	8.17%
C3G	141.19%	31.47%	41.57%	71.41%	60.64%
C3A	92.58%	21.54%	55.54%	56.55%	35.53%
C3T	6.34%	5.17%	5.30%	5.60%	0.64%
A4G	197.73%	70.67%	118.17%	128.86%	64.20%
A4C	92.99%	199.09%	162.74%	151.61%	53.92%
A4T	268.19%	384.97%	137.29%	263.49%	123.91%
G5C	60.81%	55.69%	156.41%	90.97%	56.73%
G5A	50.47%	50.35%	42.30%	47.71%	4.68%
G5T	75.52%	49.48%	119.08%	81.36%	35.17%
T6G	10.91%	13.79%	12.54%	12.41%	1.44%
T6C	1.04%	17.65%	18.51%	12.40%	9.85%
T6A	12.07%	4.17%	4.87%	7.04%	4.38%
A7G	112.05%	104.97%	276.72%	164.58%	97.18%
A7C	116.24%	112.35%	220.99%	149.86%	61.63%
A7T	116.80%	19.99%	171.24%	102.67%	76.61%
G8C	57.19%	32.61%	28.47%	39.42%	15.52%
G8A	34.02%	16.75%	26.66%	25.81%	8.67%
G8T	5.04%	5.26%	22.63%	10.98%	10.09%
A9G	15.65%	7.23%	48.39%	23.76%	21.75%
A9C	97.97%	59.07%	40.56%	65.87%	29.30%
A9T	75.59%	65.04%	87.79%	76.14%	11.39%
T10G	132.93%	89.31%	137.99%	120.08%	26.77%
T10C	125.66%	128.36%	139.67%	131.23%	7.43%
T10A	159.77%	285.75%	219.97%	221.83%	63.01%
G11C	203.69%	127.67%	191.88%	174.41%	40.91%
G11A	62.78%	157.35%	99.26%	106.46%	47.69%
G11T	71.89%	227.89%	172.84%	157.54%	79.12%
T12G	160.23%	77.20%	144.50%	127.31%	44.10%
T12C	161.93%	90.91%	80.70%	111.18%	44.25%
T12A	208.71%	137.99%	224.91%	190.54%	46.22%
T13G	56.61%	127.08%	190.02%	124.57%	66.74%
T13C	74.53%	131.22%	150.22%	118.66%	39.38%
T13A	187.96%	152.07%	157.97%	166.00%	19.25%
T14G	167.48%	93.99%	101.77%	121.08%	40.37%
T14C	65.50%	90.97%	76.70%	77.72%	12.77%

T14A	174.91%	180.99%	125.39%	160.43%	30.50%
T15G	136.39%	196.92%	141.92%	158.41%	33.47%
T15C	163.10%	133.73%	109.22%	135.35%	26.98%
T15A	125.41%	144.44%	73.96%	114.61%	36.46%

Supplementary Table S6: Figure 3.1 Data

Inactive Clones	Active Clones	% Normalized Activity
CATAGGCGCTACATG	GAAAATTGCGTGTCCG	17
ATGACGGGCGTGATT	CGAGATGATGGGATC	18
CGCCACTTTCTAATT	AGGCGCGTCTGTTGG	18
GGGGTTATGTGATGC	GAGAGGGGTGAAGGG	23
CGCGACGCCAGGAGG	TAGAGCTTACGCGTT	27
GTGCCAGTGCGTGAA	GAAGAATGGTTTATA	29
CGCGGCGTGTAAAGT	CAAAGTGGTCGCAGT	30
AGGTGGACGAAAACC	GAAGGCGGGGCGCAT	31
GGGGGCAAGATTACG	GTAGGAGGCGGTTGT	35
TTCAAAGTTCTGTAG	AGGTACGGATCAGTG	35
CCGGGGCCCGGGGGG	GTAAACCCCATAGTA	43
AGACTGTGTCCGACC	GAGTGCGCAATATAT	45
GCTGGAGTGGGCAGC	GTGATACGCCAGATT	46
CGTGAAAATCCGTA	GTAGGGCATTGACGG	47
AGAGCGTGCCGAGGG	GGTCTCTTGACAGCC	48
TGAAGGATTGCGCGG	GTACATGAGTTATAG	49
CGAGAGTGAGCGTTG	GAAGCACAGGAGACA	49
ATAATCGGGGCGCTT	GATCCTGGAAATGGG	49
AGAGGCGGACGGGTT	GAGGAGGGGCGTTAA	51
TGGTTCATCCGCGCG	GTGGCTGAACCTGGT	54
GTTTGAGCTTTCTAG	GTAAATTGCGTGTCCG	57
TTTTGATAATCAAAG	GTGCCTCCGTTGGTC	63
CGACGTAAGACGTCT	GAAAGGTATTTGATG	65
CAGCGGGCTGGTAGG	GATACCTTTTCCTGA	69
GCAGGTGGTGTAGGG	GTGGGATCAAGAAAT	81
TGGTGTGATCTAGGG	GTGGTCCTTGTGGGT	82
TCTTGGTTGGTCGAA	GGCGTCAAGTGCGCC	83
GAGAGGATGTGGGGT	GCGGCTCGAATCATG	84
CCCTTTGATGACTTT	GTGGGCATAAACGGG	87
GGAGGCATTTGCTGA	GTAGAGCAAGAAGAC	89
TATGCCAATCAGAGT	GTACTIONACTAGTT	90
CAATTAACAGGGGGC	GAGCGATAGAGGTAC	92
CTTTTTCTGCTGTTT	GTTGTGCAAATGTTG	94
TGTATGTCCACCAGT	GTCGCCAGGAGGGGC	95
GGTGTGAGTCACGTT	GTCGCCTTTAGTTGT	95
GCGAAGACACGGCTA	GTGGGCCGTGCAGGG	96
GCGGTCAGTAGGGTG	AGGATCGCGTGGTGC	104

ATTTGCTGTACATGG	TGATCCTGATGTGGG	104
AGAAACGGTTTGTCT	GTAAAGGGACTTAAT	105
GCGAAGACACGGCTA	TGATGACCTAACGTG	105
TTTCTCTCTAATCGG	GTGTGCATTTCCGGAT	118
AAGGTTTTGCTGGCT	CTAAGTGGTAGGTTA	118
TCATCGAGCAGATGG	GAGTACCCGATTTTT	122
CCCCGGGGAGGCTTA	GTAAAGGGAcTtgAT	134
AGACGCACCTTTTTT	GTAAAGGGACTTAAT	140
ACACGCGGTGTAACG	GTCATCCTAAAAAAT	170
AAGATGGGGACGAGG	GTGAAGGGGCATAGG	178
CGGTGTCAGCCTAAG	GAGCGATAGAGGTAC	181
TCCAATAGTTCGACT	GTAAGCCAGTTAGAC	248
GCGCCGGCGGTCCGT	GCCGGCGTTTGC GGG	295
TGGTGCGATGACAAT		
GGTGTGTGGCAAGCG		
GCACGCCAGCATGCA		
TGAGGGCAGCGTGAA		
TGAAGTAAAGGTAAT		
ATGACAACGTTTCGAG		
GTAGGCTAATGGGTG		
GCGCTCGCTTGAGGG		
TTGTAGGCAACTACT		
ATGAATCCGTTTATG		
GGGGAAGGGATCGCC		
TGATCTTGTAATTTT		

Supplementary Table S7: Figure 3.2 Data

	Nucleotide Abundance							
	Active Clones				Inactive Clones			
	A	C	G	T	A	C	G	T
1	0.060	0.060	0.820	0.060	0.226	0.226	0.290	0.258
2	0.300	0.040	0.160	0.500	0.113	0.242	0.435	0.210
3	0.400	0.100	0.400	0.100	0.306	0.129	0.371	0.194
4	0.340	0.160	0.380	0.120	0.242	0.194	0.306	0.258
5	0.280	0.180	0.380	0.160	0.177	0.129	0.452	0.242
6	0.180	0.420	0.200	0.200	0.145	0.306	0.355	0.194
7	0.080	0.300	0.380	0.240	0.323	0.113	0.355	0.210
8	0.240	0.140	0.420	0.200	0.210	0.258	0.323	0.210
9	0.320	0.120	0.300	0.260	0.177	0.145	0.339	0.339
10	0.320	0.200	0.200	0.280	0.210	0.290	0.258	0.242
11	0.180	0.140	0.340	0.340	0.177	0.210	0.339	0.274
12	0.260	0.160	0.320	0.260	0.210	0.242	0.355	0.194
13	0.240	0.040	0.420	0.300	0.339	0.081	0.355	0.226
14	0.280	0.100	0.340	0.280	0.161	0.210	0.339	0.290
15	0.120	0.200	0.340	0.340	0.145	0.097	0.419	0.339

Supplementary Table S8: Figure 3.3A Data

	Log₂(F_{active}/F_{inactive})			
	A	C	G	T
1	-1.88291	-2.46787	1.527113	-2.07555
2	1.439022	-2.5674	-1.4154	1.224009
3	0.339486	-0.33859	0.137853	-0.92355
4	0.432596	-0.24548	0.339486	-1.07555
5	0.68741	0.509411	-0.29794	-0.5674
6	0.339486	0.483876	-0.79802	-0.07555
7	-1.98244	1.439022	0.04998	0.224009
8	0.224009	-0.85316	0.339486	-0.03903
9	0.880055	-0.24548	-0.14594	-0.46787
10	0.545937	-0.50851	-0.33859	0.239951
11	0.04998	-0.5536	-0.05283	0.339486
12	0.339486	-0.5674	-0.21305	0.454964
13	-0.46787	-0.98244	0.272372	0.339486
14	0.824913	-1.03903	0.034632	-0.13
15	-0.50851	1.076452	-0.27349	0.034632

Supplementary Table S9: Figure 3.3B Data

	S134G		S134G/N140S		K135R/N140S/Q158R		V117F/D127G	
	Avg	StDev	Avg	StDev	Avg	StDev	Avg	StDev
TP15	100.00%	61.77%	100.00%	60.82%	100.00%	22.30%	100.00%	66.90%
C1A	23.14%	17.92%	22.50%	17.15%	74.72%	81.53%	90.24%	72.69%
C1G	146.72%	89.95%	247.30%	290.96%	600.54%	356.18%	785.71%	748.51%
C1T	64.73%	29.39%	74.53%	158.50%	132.23%	108.85%	76.93%	38.02%
T2A	40.19%	11.90%	29.33%	10.47%	74.62%	35.41%	175.64%	119.91%
T2C	49.31%	71.20%	44.38%	63.20%	73.30%	110.58%	246.18%	274.21%
T2G	11.92%	16.96%	45.72%	44.02%	93.08%	55.96%	219.55%	372.58%
C3A	73.69%	67.10%	114.48%	38.60%	177.91%	109.61%	144.88%	62.21%
C3G	45.42%	22.58%	49.63%	21.96%	60.39%	28.65%	184.06%	101.10%
C3T	9.04%	3.25%	3.62%	5.93%	8.59%	46.28%	114.86%	234.60%
A4C	91.94%	51.28%	196.80%	341.38%	196.42%	191.08%	742.96%	842.62%
A4G	109.05%	20.60%	100.45%	159.49%	80.59%	55.31%	698.67%	1013.00%
A4T	112.06%	37.05%	259.14%	197.32%	222.46%	176.46%	725.41%	556.38%
G5A	33.70%	16.79%	101.93%	17.88%	169.38%	95.21%	212.37%	174.32%
G5C	101.86%	44.82%	231.77%	93.82%	452.75%	340.13%	612.32%	459.94%
G5T	31.74%	17.82%	144.28%	13.53%	125.56%	61.61%	308.81%	356.03%
T6A	3.19%	3.08%	64.49%	40.47%	45.93%	43.30%	73.63%	37.02%
T6C	42.37%	16.98%	88.56%	55.12%	97.43%	41.53%	142.55%	64.75%
T6G	13.55%	3.91%	83.76%	86.95%	183.49%	91.17%	274.94%	320.33%
A7C	93.67%	52.46%	180.25%	49.55%	127.37%	62.65%	840.76%	557.69%
A7G	141.09%	90.87%	221.21%	120.95%	1262.32%	454.01%	551.24%	357.76%
A7T	78.44%	57.91%	102.60%	58.18%	239.46%	156.97%	220.69%	109.51%
G8A	22.49%	13.55%	133.25%	130.33%	437.26%	255.32%	103.09%	79.64%
G8C	53.93%	40.31%	211.96%	251.97%	129.99%	109.46%	109.39%	57.24%
G8T	23.90%	14.12%	99.32%	93.85%	197.37%	164.82%	38.62%	55.36%
A9C	53.42%	20.41%	179.75%	146.22%	503.25%	311.61%	386.11%	191.20%
A9G	8.85%	6.19%	100.88%	30.85%	109.09%	241.34%	148.71%	167.69%
A9T	79.74%	47.40%	379.80%	294.27%	174.10%	214.26%	237.26%	404.47%
T10A	56.16%	65.50%	85.11%	33.87%	190.17%	155.07%	131.91%	96.14%
T10C	76.96%	65.10%	132.20%	87.95%	344.64%	619.94%	175.77%	199.94%
T10G	114.27%	55.15%	290.56%	249.57%	678.70%	403.48%	983.50%	818.67%
G11A	133.94%	69.45%	132.24%	110.02%	162.81%	174.20%	157.66%	231.98%
G11C	149.42%	58.77%	247.65%	85.44%	740.67%	400.83%	653.56%	389.41%
G11T	62.37%	41.68%	229.81%	116.24%	201.80%	121.26%	492.80%	495.08%
T12A	129.52%	149.53%	110.06%	86.34%	232.81%	154.06%	202.23%	130.12%
T12C	68.82%	53.33%	183.10%	191.60%	225.00%	153.64%	169.78%	208.89%
T12G	146.06%	149.82%	316.50%	171.81%	683.98%	329.37%	884.92%	536.18%
T13A	91.88%	45.07%	96.26%	56.51%	198.69%	157.74%	221.09%	119.23%

T13C	36.04%	20.67%	104.88%	102.85%	136.09%	54.42%	281.59%	326.60%
T13G	111.30%	63.84%	94.83%	79.56%	245.50%	140.87%	138.52%	96.65%
T14A	112.73%	88.17%	216.36%	177.93%	585.23%	261.92%	371.81%	259.38%
T14C	63.00%	7.31%	120.91%	49.33%	139.10%	63.04%	150.72%	103.75%
T14G	49.72%	23.11%	110.35%	67.51%	130.98%	76.16%	253.78%	179.09%
T15A	87.45%	50.72%	94.14%	55.36%	108.92%	56.43%	65.41%	43.87%
T15C	74.54%	65.17%	150.83%	51.43%	332.23%	202.76%	129.57%	103.94%
T15G	110.51%	90.19%	260.99%	132.15%	536.82%	220.27%	614.12%	513.97%

Supplementary Table S10: Figure 4.1-4.4 Data

	S134G		S134G/N140S		K135R/N140S/Q158R		V117F/D127G	
	Avg	StDev	Avg	StDev	Avg	StDev	Avg	StDev
TP15	214.26%	168.92%	38.44%	24.02%	8.52%	3.75%	27.20%	25.01%
C1A	46.59%	1.52%	7.87%	8.81%	5.01%	1.95%	22.63%	17.81%
C1G	279.59%	122.34%	102.05%	145.19%	47.07%	4.24%	191.81%	255.96%
C1T	128.82%	60.24%	22.88%	38.34%	9.69%	4.72%	17.87%	13.72%
T2G	22.73%	3.38%	14.75%	9.38%	7.48%	6.74%	39.30%	10.99%
T2C	95.10%	7.83%	13.65%	14.14%	4.36%	1.83%	41.38%	13.85%
T2A	81.56%	31.49%	10.59%	4.37%	5.84%	3.19%	36.24%	21.60%
C3G	90.44%	47.41%	17.29%	8.93%	4.87%	1.93%	44.95%	63.06%
C3A	144.48%	154.29%	37.01%	19.99%	13.63%	8.75%	35.26%	21.98%
C3T	19.13%	13.37%	2.49%	4.04%	0.17%	3.35%	13.30%	36.01%
A4G	240.97%	145.59%	33.02%	40.09%	6.15%	1.03%	105.99%	203.90%
A4C	213.46%	159.56%	82.64%	164.90%	13.37%	8.90%	116.63%	137.05%
A4T	250.24%	32.31%	86.63%	48.37%	18.25%	24.66%	126.32%	51.47%
G5C	218.66%	95.16%	81.39%	11.58%	33.55%	3.80%	89.05%	50.71%
G5A	76.95%	34.91%	38.18%	15.78%	12.78%	3.56%	43.02%	26.51%
G5T	72.85%	24.70%	44.83%	8.47%	9.86%	4.63%	48.78%	23.79%
T6G	28.06%	9.96%	28.21%	19.63%	14.27%	4.75%	54.80%	2.03%
T6C	91.49%	44.99%	30.47%	12.04%	7.58%	1.97%	30.24%	4.95%
T6A	7.28%	4.21%	24.44%	9.40%	3.35%	1.89%	15.50%	7.77%
A7G	291.46%	134.51%	75.68%	15.64%	100.58%	23.33%	89.35%	25.91%
A7C	190.42%	122.37%	64.19%	9.48%	10.30%	4.19%	162.47%	110.42%
A7T	157.02%	101.18%	37.11%	8.75%	17.60%	7.32%	41.75%	20.34%
G8C	128.90%	162.32%	69.94%	62.90%	10.10%	4.32%	26.03%	7.84%
G8A	47.67%	53.67%	46.89%	29.93%	33.25%	3.53%	16.31%	8.22%
G8T	54.43%	86.13%	33.51%	22.89%	18.37%	24.71%	4.31%	8.29%
A9G	17.03%	7.62%	33.33%	2.49%	6.72%	15.98%	28.95%	9.63%
A9C	108.28%	36.00%	62.26%	29.34%	38.96%	15.77%	74.26%	23.48%
A9T	152.86%	68.17%	129.97%	36.87%	12.30%	11.80%	46.36%	77.17%
T10G	232.27%	55.11%	102.88%	56.22%	52.90%	15.71%	174.17%	90.86%
T10C	143.10%	95.71%	46.45%	10.82%	21.27%	44.42%	36.98%	34.22%
T10A	102.67%	107.11%	30.84%	0.94%	14.59%	7.70%	26.14%	15.22%
G11C	297.47%	43.28%	88.09%	4.22%	58.67%	15.79%	125.13%	57.25%
G11A	298.61%	201.46%	45.74%	16.72%	12.17%	6.28%	27.92%	46.82%
G11T	119.37%	96.35%	74.29%	34.91%	15.41%	0.76%	95.62%	31.28%
T12G	286.37%	113.08%	118.53%	37.62%	51.59%	8.44%	162.19%	41.04%
T12C	128.49%	101.92%	63.04%	47.32%	17.16%	0.83%	36.65%	41.49%
T12A	247.17%	61.32%	36.56%	28.66%	17.04%	3.93%	42.78%	43.97%
T13G	214.22%	21.42%	31.76%	16.64%	18.15%	2.00%	36.58%	48.01%

T13C	73.64%	53.48%	35.54%	23.83%	10.54%	1.78%	107.67%	289.11%
T13A	182.19%	88.06%	32.36%	11.36%	13.66%	2.11%	47.03%	28.40%
T14G	97.53%	65.32%	42.07%	14.90%	10.13%	1.78%	45.40%	17.28%
T14C	223.38%	83.41%	43.44%	19.46%	10.52%	2.80%	22.43%	8.97%
T14A	273.96%	296.20%	74.34%	33.09%	44.66%	9.18%	63.56%	30.19%
T15G	328.98%	277.40%	96.88%	67.36%	41.54%	9.80%	110.60%	98.98%
T15C	181.35%	160.05%	48.46%	17.93%	24.60%	5.82%	23.11%	12.07%
T15A	216.30%	176.01%	34.84%	17.48%	8.56%	3.49%	15.60%	22.31%

Supplementary Table S11: Figure 4.5A/4.6A/4.7A/4.8A Data

		S134G	S134G/N140S	V117F/D127G	K135R/N140S/Q158R
	TP15	1.09936254	-1.379319759	-1.878675088	-3.584962501
C - 1	A	1.773381182	-0.781039925	0.732228836	-1.443550187
	G	0.117433624	-1.336555831	-0.4262012	-2.45279506
	T	1.985476538	-0.507473603	-0.864500322	-1.746210027
T - 2	A	2.085753941	-0.859859423	0.915475121	-1.717244339
	C	3.939881072	1.138984332	2.739369376	-0.504959958
	G	4.883383843	4.260038694	5.673404685	3.279464864
C - 3	A	1.353119261	-0.611826869	-0.68130684	-2.05317933
	G	0.340888046	-2.045910139	-0.667591048	-3.874132441
	T	1.770976793	-0.605273355	1.247069022	-2.074859073
A - 4	C	0.493587497	-0.875421036	-0.3783118	-3.503621554
	G	0.903086367	-1.9645008	-0.281884398	-4.38981912
	T	-0.074380072	-1.604717574	-1.060666701	-3.851743338
G - 5	A	0.689792933	-0.321373784	-0.149071581	-1.900679389
	C	1.265226571	-0.160421138	-0.030775242	-1.439078087
	T	-0.159324696	-0.859640653	-0.737931339	-3.044172483
T - 6	A	0.094560937	1.796083427	1.138993847	-1.069296274
	C	2.883221379	1.296891544	1.285958983	-0.708802064
	G	1.176795809	1.184316094	2.142373074	0.201431723
A - 7	C	0.345544403	-1.22326981	0.116587592	-3.862432298
	G	0.824508873	-1.120805034	-0.881303217	-0.710493383
	T	0.612870591	-1.468228459	-1.298145082	-2.544461032
G - 8	A	0.8850506	0.861350138	-0.661878541	0.365424199
	C	1.709130594	0.827136788	-0.598874453	-1.963742512
	T	2.309962025	1.610008159	-1.122944372	0.743173376
A - 9	C	0.717235169	-0.081165726	0.173036774	-0.757431435
	G	-0.480536399	0.488344603	0.285064802	-1.82108384
	T	1.005453	0.77145218	-0.715877739	-2.629605333
T - 10	A	-1.111346121	-2.846579817	-3.085307423	-3.926073128
	C	0.124858892	-1.498243538	-1.8271504	-2.624979554
	G	0.951846994	-0.222946543	0.536512703	-1.182525302
G - 11	A	1.487889733	-1.218723238	-1.930813565	-3.128955569
	C	0.770266178	-0.985405585	-0.479121048	-1.571815437
	T	-0.400277876	-1.084607721	-0.720283578	-3.354091517
T - 12	A	0.37539608	-2.381862222	-2.15494786	-3.482789225
	C	0.208721368	-0.818634134	-1.601012184	-2.695495613

	G	1.169534619	-0.103052922	0.349309476	-1.303368836
T - 13	A	0.134207742	-2.358751127	-1.819530004	-3.602801847
	C	-0.688231628	-1.739412892	-0.140266162	-3.492846388
	G	0.78215896	-1.971522442	-1.767829688	-2.778650281
T - 14	A	0.450698695	-0.548994278	-0.886121598	-1.879874839
	C	0.780888276	-0.782663505	-0.674623511	-2.943516424
	G	-0.312091927	-1.524982552	-1.415302306	-3.57830536
T - 15	A	0.568291566	-1.855915846	-2.474299568	-3.637826283
	C	0.002555894	-1.594373042	-2.367955297	-2.643145765
	G	1.3788	-0.814544845	-0.538497257	-2.22970693

Supplementary Table S12: Figure 4.5B/4.6B/4.7B/4.8B Data

Curriculum Vitae

Name: Brendon McDowell

**Post-secondary
Education and
Degrees:** McMaster University
Hamilton, Ontario, Canada
2008-2012 B.S.

The University of Western Ontario
London, Ontario, Canada
2012-2015 M.S.

**Honours and
Awards:** London Health Research Day, Biochemistry Poster Winner
March 2013

Dean's Honour List, McMaster University
2008-2009, 2011-2012

McMaster Honour Awards Entrance Scholarship, Value \$2000
2008-2010

Publications:

Kleinstiver, B.P. et al (2014) The I-TevI nuclease and linker domains contribute to the specificity of monomeric TALENs. *Genes/Genomes/Genetics* **4**(16), 1155-1165.

Presentations:

McDowell, B., Kleinstiver, B.P., Kolaczyk, T., and Edgell, D.R.
Determining the Nucleotide Preference of the Inter-Domain Linker for its DNA Spacer Target in Engineered Tev-TAL Nucleases. London Health Research Day, London Convention Center, London, Ontario. March 2013. (Poster)

McDowell, B., Kleinstiver, B.P., Kolaczyk, T., and Edgell, D.R.
Determining the Nucleotide Preference of the Inter-Domain Linker for its DNA Spacer Target in Engineered Tev-TAL Nucleases. Harold B Stewart Research Showcase, University of Western Ontario, London, Ontario. January 2013. (Poster)