Western University

Scholarship@Western

9-18-2015 12:00 AM

Isolation and Characterisation of Relaxed Specificity I-TevI Nuclease Domain Mutants

Alexander C. Roy
The University of Western Ontario

Supervisor
Prof. David R. Edgell
The University of Western Ontario

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

Part of the Biochemistry Commons

**ISOLATION AND CHARACTERISATION OF RELAXED SPECIFICITY
I-TevI NUCLEASE DOMAIN MUTANTS**

by

Alexander C. Roy

Submitted in partial fulfillment of the requirements
for the degree of Master of Science

at

Western University
London, Ontario
August 2013

ABSTRACT

Engineering nucleases is important to the advancement of genetic engineering and gene therapy approaches. Engineering requires a knowledge of which residues are contributing to each function of the nuclease. The residues which contribute to cleavage specificity of the I-TevI nuclease domain (ND) are unknown. I suspect that some of these contributions derive from the ND, thus my null hypothesis is that mutation of the ND will not alter the substrates this enzyme can cut. I have mutagenised the I-TevI nuclease domain and using directed evolution I have isolated mutations which were characterised *in vivo* and *in vitro*. These mutations permit cleavage of otherwise cleavage resistant substrates, indicating that the ND does contribute to cleavage specificity. Mutations which provided the greatest increase in activity against cleavage resistant substrates (K26R, T95S, and Q158R) were combined into a single relaxed specificity nuclease domain which exhibits 1.2-5-fold improved cleavage of resistant substrates.

KEYWORDS

DEDICATION

This thesis is dedicated to my darling Cynthia,

whose patience, confidence, and love have been my constant muses,

and without whom this thesis could not have happened.

FOREWORD & ACKNOWLEDGEMENTS

I have always been enchanted by the fundamental concept of an enzyme; the image they conjure in my mind of an incredibly-efficient, atom-scale, precision machine, operating at hundreds of actions per second, seems more in line with science-fiction than reality. The reality of enzymes is even more fantastic when you consider that these machines continue to out-perform synthetic catalysts in terms of the versatility, specificity, and efficiency of the reactions they catalyse; their nitrogen-fixing, carbon-fixing, and oxygen-producing activities are apparently capable of Terra-forming an entire planet; and that, in a human-being, a single dysfunctional enzyme can be fatal! The potential I see in enzymes – for scientific and technological advancement – drew me to the field of enzymology; a study of the mechanisms by which enzymes function.

The studies and insights that I convey in this thesis are the embodiment of my interest; however, I acknowledge that a proper thesis requires more than academic curiosity, or time spent at a lab-bench. I may never have pursued my curiosities without the support and guidance of those people in my life who wanted me to succeed, and who provided me the tools to do so. I would like to thank my mother, Beverly Joyce Clark Roy, and my father, Robert Gordon Roy, for filling my childhood with the gratification that comes from scientific discovery and elegant design. I would like to thank my grand-father, Ralph Alexander Clark, from whom I inherited – whether by nature, or by nurture – an insatiable desire to tinker with, to re-imagine, and to restore. Most of all, I would like to convey my gratitude for Professor David Edgell; he pulled me from the flames of a career path gone awry, and set me on track with his firm, but fair guidance. I would be a lesser scientist were it not for his encouragement, his insight, and his mentorship over the past two years. Finally, I confess I would need to write a second thesis to adequately thank my girlfriend Cynthia Mancinelli, who not only makes my workload bearable, but enjoyable, rewarding and fulfilling.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations and Symbols Used

ADA-SCID	adeonsine deaminase - SCID
A	adenosine
A_{280}	absorbance at 280 nm
b	cuvette pathlength
bp	basepair
c	concentration by UV absorbance
C	cytosine
cas	CRISPR-associated
cfu	colony forming units
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
Da	Dalton
DBD	DNA binding domain
ddiH$_2$O	distilled deionised water
DNA	deoxyribonucleic acid
DSB	double-strand break
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
ES	enzyme-substrate complex

f_S	fraction of substrate remaining
FRET	Förster resonance energy transfer
G	guanosine
h	hours
HE	homing endonuclease
HDR	homology directed repair
HIV	human immunodeficiency virus
HTH	helix-turn-helix
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
k_{app}	apparent kinetic constant describing substrate decay
k_{cat}	enzymatic turnover number
K_M	Michaelis constant
LB	lysogeny broth
LHE	LAGLIDADG HE
MegaTev	fusion of I-TevI ND with meganuclease
MGE	mobile genetic element
min	minute
MMEJ	microhomology-mediated end joining
MWCO	molecular weight cutoff
m_1	correction for baseline
m_2	correction for initial condition

m_3	see: k_{app}
ND	nuclease domain
NHEJ	nonhomologous end joining
NMR	nuclear magnetic resonance
NNN	positions 2, 3, and 4 of the I-TevI cleavage motif
nt	nucleotides
OD_{600}	optical density at 600 nm
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB ID	protein databank identification code
RE	restriction endonuclease
RNA	ribonucleic acid
RPM	revolutions per minute
RVD	repeat variable diresidue
R1	round 1 of selection
R2	round 2 of selection
R3	round 3 of selection
R4	round 4 of selection
R^2	coefficient of determination
SCID	severe combined immunodeficiency
SDS	sodium dodecyl sulphate

sgRNA	single guide RNA
SOC	super optimal broth with catabolite repression
SOEing	splicing by overlap extension
SPR	surface plasmon resonance
ssRNA	single-stranded RNA
STD-NMR	saturation transfer difference NMR
t	time
T	thymidine
TBE	tris-borate-EDTA
TAL	transcription-activator like
TALE	TAL effector
TALEN	TALE nuclease
TALER	TALE recombinase
T3	K26R T95S Q158R triple mutant
WT or wt	wild-type
ZF	zinc finger
ZFA	ZF array
ZFN	ZF nuclease
ZFR	ZF recombinase
ε_{280}	extinction coefficient at 280 nm
ω	omega-factor

Chapter 1   INTRODUCTION

Engineering nucleases is important for the advancement of genome editing, a core component of genetic engineering and gene therapy approaches[1-3]. Optimally, a system would be developed such that a nuclease could be immediately identified and produced to edit any gene. Such an enzyme would have to specifically target an extended DNA sequence to ensure editing at a single unique site in a genomic context. This level of specificity requires extensive protein-DNA contacts over a long stretch of DNA. Homing endonucleases (HEs) are a class of nucleases that recognise DNA sequences that are 12-40 bp in length, a characteristic that allows them to potentially effect a genetic change at a single position in a host organism's genome[4]. HEs effect contacts over these lengthy DNA sequences through the combination of multiple DNA binding 'modules' that each contribute to the relative degree of cleavage that a nuclease can effect on each of a related set of substrates, hereafter called its cleavage profile. These modules can be recombined to generate libraries of engineered nucleases, each with a unique cleavage profile. Such libraries could represent a source of versatile genome editing tools.

I-TevI is a HE comprised of three modules: an N-Terminal nuclease domain (ND), a C-terminal DNA binding domain (DBD), and a linker region that connects the two domains[5]. The I-TevI ND has been recombined with other DBDs to generate chimaeric nucleases that combine the cleavage profile of the chosen DBD with that of the I-TevI ND[6]. The range of cleavage profiles possible from chimaeric nucleases like these could be further extended by rationally designing a library of engineered I-TevI NDs with distinct cleavage profiles. Rational design of a nuclease requires that the

amino acids which contribute to DNA sequence readout are known; this is not the case for the I-TevI ND.

In this thesis I describe how I used directed evolution to screen for mutant I-TevI NDs that were able to cleave otherwise poorly cleaved DNA substrates. Further, I describe the process by which I isolated and characterised the impact of mutations that were identified from the screens on the I-TevI ND cleavage profile *in vivo* and *in vitro*. Subsequently, I relate the identified mutations and their impacts on the I-TevI ND cleavage profile to previously identified mutations in the I-TevI ND and to other attempts to develop nucleases with altered cleavage profiles. Finally, I give an insight into how these mutant NDs will be used to better understand the source of its cleavage profile towards the ultimate goal of rationally designing a library of I-TevI NDs with unique cleavage profiles for use in genome editing.

1.1   Genome Editing

Genome editing is a technique in which a specific genetic locus of an organism's genome is targeted for removal, replacement, or insertion of new genetic material. The editing process is facilitated by nucleases that recognise the chosen target locus, and create a nick or double-strand break (DSB) within the locus. The breakage of a DNA strand then elicits the DNA repair pathways to mend the break, either by nonhomologous end joining (NHEJ), microhomology-mediated end joining (MMEJ), or homology-directed repair (HDR), which can be utilised to effect genetic changes. The variety and versatility of genome editing tools has seen a tremendous amount of growth in the past decade. These genome editing tools are typically exploited by one of two

related fields: genetic engineering (§ 1.1.2) or gene therapy (§ 1.1.3).

1.1.1   Mechanisms of Genome Editing with Nucleases

Genome editing with nucleases is achieved by eliciting DNA repair pathways, each of which offer unique opportunities for editing (Figure 1.1)[7]. NHEJ, and MMEJ can remove nucleotides surrounding the site of damage and thus can be used to create a deletion or gene knock-out; HDR uses a template to effect reconstruction of the broken strand, and thus if a donor template is provided, the owriginal sequence surrounding the break can be replaced with a DNA sequence from the donor template.

1.1.2   Genome Editing in Genetic Engineering

Genetic engineering is the artificial genetic modification of an organism in order to make it more useful in a particular context. Genetic engineering is used extensively in the field of biomedical research to create genetically modified organisms that provide model systems for human diseases[8]. Other genetically modified organisms created using genetic engineering include knock-out and knock-in mice, which express an aberrant phenotype that facilitates an understanding of gene function[9]. Genetic engineering can also be used to create gene-fusions that encode a protein of interest joined with a reporting element such as green fluorescent protein which allow for tracking and localisation of said protein in a single-cell or whole-organism context[10,11].

Genetic engineering is also used in the agricultural industry, which has benefitted tremendously from the development of transgenic plants. Thus there is a great interest in engineering genome editing tools to aid in further developments. To this end the HE
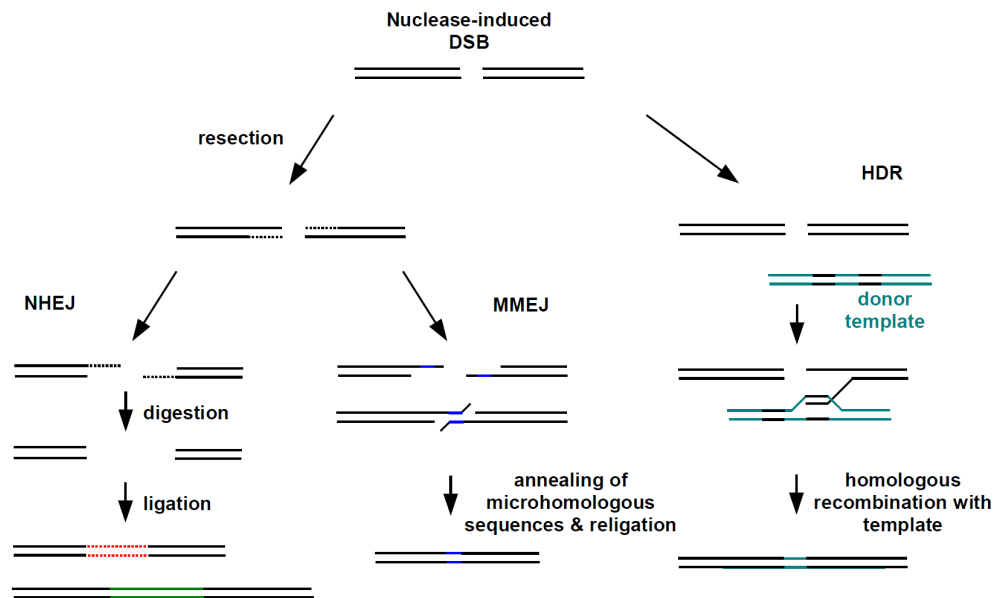
Figure 1.1. Mechanisms of DNA DSB repair provide opportunities for genome editing. The three cannonical modes of DNA DSB repair are shown above. NHEJ and MMEJ both begin with resection of the 5' ends created by the DSB, but their mechanisms diverge thereafter. In NHEJ, digestion of the 3' overhangs created by resection creates blunt ends, which are ligated together. In MMEJ, regions of microhomology between the 3' overhangs are brought together, excess 3' ends are removed, and missing nucleotides are filled in adjacent to the microhomologous region. HDR uses a donor template with homologous regions on both sides of the DSB. The region adjacent to the DSB is replaced by replicating the donor template.

I-CreI has been re-engineered to generate transgenic varieties of *maize* by effecting insertions and deletions at a defined site within immature embryos[12]. Further, multigene plant transformation vectors have been developed that employ a cloning system composed of zinc-finger (ZF) nucleases (ZFNs) and HEs[13], and plant viral vectors have been used to deliver endonuclease genes[14].

The medical and pharmaceutical industries have benefited from genetically engineered organisms that produce therapeutic proteins, such as human insulin and human growth hormone[15]. Genetic engineering is also a promising avenue of research for the field of renewable energy. Genetically engineered photosynthetic algae present the unique opportunity to harness solar energy and convert it directly into biofuels[16], or value-added products[17]. One interesting offshoot of genetic engineering is its impact on computer technology. Bacterial chromosomes were harnessed to generate a DNA-based retrievable data storage unit[18,19] and to effect digital control of gene expression, which are first steps towards a biological computer[20].

Genetic engineering is also being investigated for its potentially transformative effect on population genetics and allelic frequencies. In one application, potentially disease carrying mosquitoes of the species *Anopheles gambia* were engineered for reduced fertility by targeting their genome with a synthetic genetic element containing the HE I-SceI[21,22]. Similar results were observed using the I-PpoI HE gene in male mosquitoes[23]. Related studies of the propagation of malaria by mosquitoes predict that transmission of a HE gene in this manner would reduce the incidence and spread of malaria[24].

1.1.3   Genome Editing in Gene Therapy

Gene therapy is *in situ* or *ex vivo* DNA correction or manipulation to restore a healthy state to a diseased organism. Such corrections could ameliorate countless hereditary disorders, or eliminate core components of a provirus. The promise of much needed medical advancements has spurred countless attempts to apply genome editing techniques to human disease. The more promising applications involve monogenic diseases that can be treated *ex vivo,* such as blood disorders, skin ailments, and immunodeficiencies[25]. Notably, a successful proof of concept treatment of the monogenic immunodeficiency, ADA-SCID was demonstrated using using viral vectors[26]. However, off-target gene integration led to leukemia in five patients, and one patient's death. Thus, engineered genome editing tools with more precise targeting are needed[27].

Some preliminary success in using genome editing tools for gene therapy have been achieved using engineered recombinases. Recombinases are a class of enzymes that are capable of translocating genetic material between a DNA vector and a genome (see §1.2.1). This ability to move genetic material has been exploited in mice to develop a number of potential treatments for diseases such as hemophilia[28], muscular dystrophy[29], Junctional epidermolysis bullosa[30], peripheral vascular disease[31], and rheumatoid arthritis[32]. Some success has also been had with recombinases in human cells, such as the genetic correction of dystrophic epidermolysis bullosa in primary patient cells[33], and the reprogramming of somatic cells to pluripotency[34, 35].

A lot of work has gone into developing gene therapy approaches for treating HIV infection. In one such study designer endonucleases were coupled *in trans* with DNA

end-processing enzymes to bias DNA repair towards using HDR[36]. In one such application a 3′ repair exonuclease, Trex2, was overexpressed along with ZFNs, transcription activator like effector nucleases (TALENs), or an engineered I-CreI LAGLIDADG HE (LHE). The result was an improved yield of targeted gene disruption in several different cell lines. In particular, the I-CreI based approach effected a seven-fold increase in gene disruption of the endogenous HIV coreceptor CCR5[36]. Furthermore, the ZFN-based approach is now in clinical trials[37]. Other approaches have sought to eliminate the virus in its proviral phase. The LHE I-AniI has been shown to cure cells of latent HIV infection by mutagenising key proviral sequences[38].

Progress has also been made in developing therapeutics for other human diseases. An engineered I-CreI LHE was developed that could target and correct a defect in the XPC1 gene of patients with Xeroderma pigmentosum[39,40]. In a similar effort, another engineered I-CreI LHE was developed to target the RAG1 gene in SCID[41,42]. In another example, an engineered I-CreI LHE was developed to correct a dystrophin gene defect underlying Duchenne muscular dystrophy[43].

Therapies built around genome editing tools are still in their infancy[25,44], and there are a few main obstacles to their application. For example, many of these therapies operate by genetically modifying a subset of the patient's (or a compatible donor's) cells in culture, and transplanting them into the relevant tissue, where they multiply to supplant the diseased cells; however, such a strategy is impossible in non-dividing cells. Furthermore, genome editing with nucleases relies upon HDR for the insertion or replacement of genetic material; however, HDR is downregulated in many terminally differentiated cells, such as cardiomyocytes[45], or neurons[46]. Another complication arises

from off-target DNA cleavage, which can cause unwanted genetic modifications. Take for example the observation of Cre and φC31-mediated recombination of off-target pseudo-recombination sites[47,48], which has been shown to lead to deletions and chromosomal re-arrangements in cultured cells[49-51] and mice[52]. Additionally, DNA damage has been observed at sites of ZF recombinase (ZFR)-mediated recombination[53]. Some of these modifications may convert proto-oncogenes to oncogenes, and lead to cancer, as was the case with one notable attempt to treat ADA-SCID with gene therapy using a viral vector and recombinase[27]. However, protein engineering approaches provide an opportunity to develop new tools for genome editing that are more selective, robust, and reliable.

1.2   Tools for Genome Editing

As a result of the many promises of the field of genetic engineering and gene therapy presented above, a great deal of interest has been poured into developing tools to advance the field of genome editing[1,2,54-58]. These tools are typically engineered variants of enzymes that effect site-specific cleavage, such as recombinases, integrases, HEs, or restriction endonucleases (REs). These engineered enzymes may also be coupled to proteins whose native function requires sequence-specific recognition of DNA, such as ZFs or TALEs. Each of these engineered enzymes has its own strengths and failings, leading to the great diversity of genome editing tools – potential or proven (*e.g.* TALENs, and CRISPRs). However, due to the caveats described above, none of these tools have proven themselves sufficiently reliable in a clinical or therapeutic context. Efforts to develop a robust therapeutic genome editing tool benefit from

concurrent development of a variety of tools, each with unique characteristics. The following sections describe those tools that are currently under development, and which may yet provide a robust therapeutic genome editing tool.

1.2.1   Recombinases/Integrases

Recombinases (also known as integrases) are a class of enzymes responsible for integration and excision of viral genomes, activation of developmental genes and transposition of mobile genetic elements (MGEs)[59]. Identification of recombinase minimal nucleotide target sequences has permitted their use in metabolic and genetic engineering, and synthetic biology. In this capacity, recombinases have been used to create gene knock-outs[60-62]. Their high site-specificity also makes them useful for targeted integration and excision of transgenic elements and selectable markers[63-66].

1.2.2   Restriction Endonucleases

Recombinant type II REs are instrumental to molecular biology; both for their usefulness in molecular cloning, and because of their high-fidelity and straightforward reaction conditions, requiring only Mg^{2+} as a cofactor[67]. REs were initially discovered during investigations of viral restriction, when it was observed that viral DNA was eliminated from a bacterial cell[68]. Further investigation revealed that REs were targeting specific 4-8 bp palindromic DNA sequences and methylation states to eliminate non-self DNA. Indeed, REs are highly sequence-specific, as evidenced by intolerance to substitutions in their target. For example, plasmid pAT153 has 12 EcoRV sites that differ from the cognate target sequence by a single nucleotide, and the best of these was

cleaved 6 orders of magnitude less efficiently (k_{cat}/K_m) than the cognate target[69]. Further, a typical type II RE binds its cognate target with nano- to picomolar affinity, but against a non-cognate target, affinity is only in the µM range[70].

One consequence of such high sequence specificity is the difficulty of making altered specificity type II RE mutants. Attempts to generate novel specificities in type II REs by substituting amino acids used in base-specific interactions proved futile, and tended to lead to reduced activity without significantly changing specificity[71-73]. This outcome was explained by the role to which REs have evolved. REs must retain high specificity against a single target, and so have redundant means of recognition, conferred by an extensive network of intramolecular contacts and bound waters[74-78]. Furthermore, crystal structures provide only ground-state depictions of the enzyme-substrate (ES) complex, and gross amino acid substitutions often leave functional groups at the wrong distance or orientation. However, some facile specificity changes have been identified, but they required at least a pair of amino acid substitutions, one for each nucleobase in the basepair[79].

Despite being highly sequence-specific, most REs are fundamentally unsuitable for genome editing. A given RE would be expected to cleave every 4^h bp, where h is the length of the cognate site. Consequently, widespread cleavage would be expected in a genomic context. For example, the human genome has *ca.* 50,000-13,000,000 sites of 8-4 bp, respectively. However, the type IIS RE FokI has proven invaluable to genome editing efforts in that it has a non-specific ND that has been conjugated to DBDs to generate chimeric nucleases, described in greater detail below (see §1.2.3 and 1.2.4). Additionally, type V REs are a more recently discovered family of REs, and are

uniquely suited to genome editing. Type V REs are components of the CRISPR/Cas restriction system, which uses sequence complementarity to guide RNAs to target DNA cleavage, and is the topic of the next section.

1.2.2.1 CRISPR/Cas9

First – and unknowingly – observed in 1987 during sequencing of the *E. coli iap* gene[80], it wasn't until 2005 that CRISPRs were recognised for what they were; a bacterial defence system[81-83]. With further investigation, the full process was eludicated[84, 85]. Foreign DNA sequences are incorporated into a CRISPR locus for later recognition. These sequences are then transcribed as crRNAs, or "guide RNAs", which are ssRNAs that are bound by proteins expressed from CRISPR associated (*cas*) genes to cleave foreign DNA complementary to the crRNA. Further study of CRISPR function revealed that in type II CRISPR immune systems cleavage of DNA targeted by the crRNA could be effected by a single gene product, Cas9[86]. Cas9 binds another RNA, tracrRNA, which itself binds a complementary region of the crRNA in order to recruit crRNAs to Cas9. By merging the crRNA and tracrRNA into a single chimeric guide RNA, or sgRNA, cleavage could be accomplished by a single RNA-enzyme pair[87]. The direct method of targeting through complementary basepairing and the simplicity of a two component sgRNA-Cas9 system lended itself to genome editing approaches. Since the year 2013, papers have been published demonstrating genome editing with the CRISPR/Cas9 in human cells, zebrafish embryos, and bacteria[88].

1.2.2.2 FokI Nuclease Domain is Useful for Genome Editing

FokI is a modular type IIS RE, comprised of a site-specific N-terminal DBD and non-specific C-terminal ND that is functional as a dimer[89,90]. Although REs are generally unsuitable for genome editing for the reasons discussed above, the FokI ND alone provides an alternate strategy for targeting a specific genetic locus. FokI can be attached to a pair of DBDs, each flanking a chosen target site to confer endonuclease function against that site. DBDs that have been successfully used include ZFs and TAL effectors, which are described in further detail below.

1.2.3 Zinc-finger Nucleases

Motivated by failed attempts at re-engineering REs, Chandrasegaran *et al.* took the nonspecific ND from FokI and combined it with a ubiquitous DNA recognition domain, ZF protein, as the DBD[91]. Fusions of these domains functioned as endonuclease with a target sequence defined by the ZF-DBD. The variety of ZFs, each recognising a distinct trinucleotide gave promise to the approach[92]. Indeed, several ZFs could be appended to the FokI ND to make ZF arrays (ZFAs)[93]. Each ZFA can be designed to target a sequence of 9-12 bps, meaning that a complete FokI dimer recruited by a pair of ZFAs can target a site defined by 18-24 bp. This extent of sequence recognition provides a potential means to target a single site in a genomic context. However, the trinucleotide recognition by the ZF proteins proved less stringent than their successor, TALE proteins.

1.2.4 TAL Effector Nucleases

The modular approach to defining a target site used for ZFEs was the inspiration for another class of engineered nucleases, based upon TAL effectors[94]. Identified from

Xanthomonas bacteria[95], TALEs are proteins comprised of repeating peptides that differ by only two residues, known as the repeat variable diresidue (RVD)[96]. These RVDs each define a single nucleobase in the TALE recognition site, based upon the identity of the two amino acids in the variable di-residue. Joined with the FokI ND, these TALE nucleases provide a straightforward method of targeting any DNA sequence.

1.2.5   Homing Endonucleases

HEs are derived from a class of MGE that defy the laws of mendelian inheritance. Discovered in 1970, HEs are unidirectionally inherited, owing to their unique method of transmission[97]. HEs are responsible for catalysing a DSB in a naïve allele, and co-opting DNA DSB repair pathways to integrate its host intron in a process termed 'homing' (Figure 1.2)[98]. Since their discovery, HEs have been found in all three domains of life[99], and in different genetic contexts, within group I or group II introns[100], as self-splicing inteins[101-103], and as free-standing genes[104].

Evolution of HEs has been guided by two somewhat contradictory forces. On the one hand, the process of homing requires that HEs be highly site-specific such that they insert reliably into the naïve allele, which is often a functionally critical gene[105], without causing deleterious mutations. On the other hand, their continued propagation requires that they be able to target homologous genes in (*i.e.* "jump" to) other species[98]. Indeed, experimental determinations of HE target specificity have found that they bind long targets (12-40 nts), and are thus highly site specific, yet tend to tolerate individual substitutions[106,107]. The ability to bind long target sites in a site-specific (if not fully sequence specific manner), and use DNA repair to alter a single genetic locus makes

expression

**binding &
 cleavage**

HDR & insertion
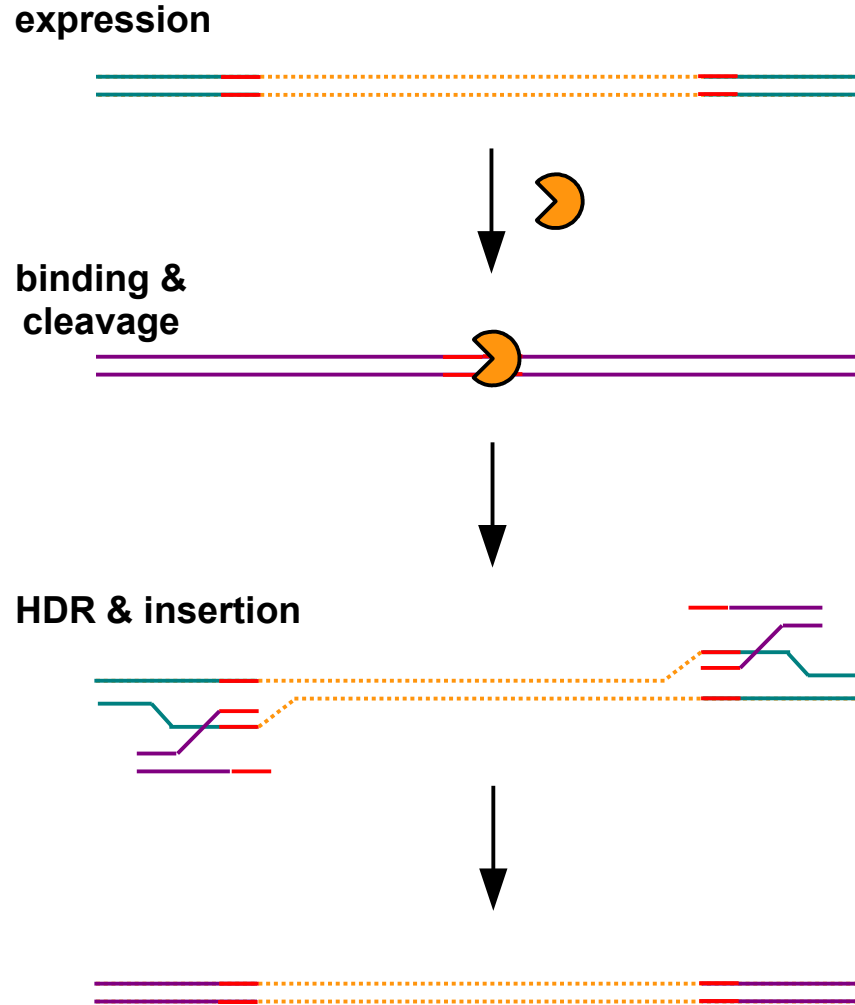
Figure 1.2. Endonuclease homing is a process which relies upon precise cleavage and repair, and is responsible for propagation of its MGE. Homing is the process by which lateral transfer of a MGE (orange) from one allele (turqoise) to another (magenta) is facilitated by the endonuclease which it encodes. Once the endonuclease is expressed, it binds specifically to its homing site (red) a sequence near the intended MGE insertion-site, and typically induces a DSB in a naïve allele. When the MGE harbouring allele is used as a template during HDR of the nuclease induced DSB, the MGE becomes incorporated into the repaired strand, separating the homing site across the MGE and preventing further DSBs.

HEs a promising platform for the development of tools for genome editing.

All HEs currently identified fall into one of six families that are named after the consensus amino acid sequence that defines the family. The six families are HNH, His-Cys, PD-(D/E)xK, EDxHD, LAGLIDADG, and GIY-YIG[1]. Although each of these families share the above-mentioned characteristics of HEs, only two of these families have been exploited for genome editing applications, LHEs and GIY-YIG HEs.

1.2.6 LAGLIDADG Family Homing Endonucleases

LHEs (also known as meganucleases) were the first family of HEs to be identified, and since their discovery in 1970 they have provided a system to understand MGEs[108]. This long history of investigation has produced a wealth of information about LHE structure and mechanism, including 37 crystal structures to date. LHEs consist of two LAGLIDADG domains that are either subunits of a dimer, or domains of a monomer, that bind the enzyme's 16-26 bp cognate target (Figure 1.3)[109,110]. The single active site of LHEs is formed at the interface between pseudo-symmetric LAGLIDADG domains, and is responsible for cleavage of both DNA strands. It is still unclear if LHEs as a family require two or more divalent metal ions to effect catalysis, as examples of both exist in the literature[111,112].

1.2.7 Engineered LHEs for Genome Editing

LHEs provide a promising platform for genome editing. Their relatively small size facilitates expression in the host organism, and extensive structural characterisation facilitates rational design and targeted mutagenesis.
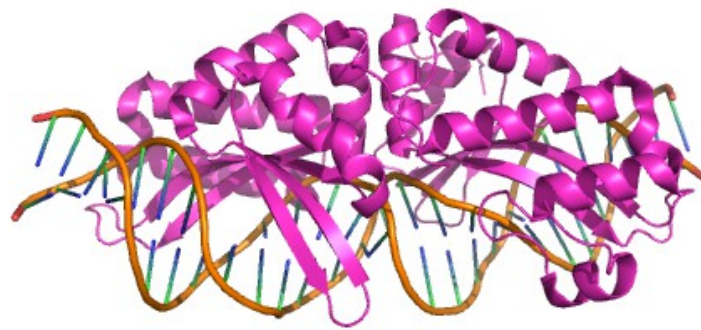
Figure 1.3. I-OnuI LHE has a pseudosymmetric dimer of two structurally similar domains. The structure of I-OnuI, like other LHEs, has a striking 2-fold rotational axis. These domains are nonidentical, and each imposes its own distinct substrate sequence preference.

Furthermore, their two-domain pseudo-symmetric structure and growing number of identified family members supports an intuitive mode of developing LHEs with altered target site preferences *via* domain swapping. These characteristics combined have allowed a repertoire of engineered LHEs with an extensive array of target site preferences to emerge.

Efforts to engineer LHEs with altered target site preference fall into two broad categories: mutagenesis-based approaches, and domain swapping approaches. Mutagenesis-based approaches use directed evolution, rational design or computer assisted design to identify amino acids of importance for binding or catalysis and substitute one or more of these amino acids to effect a change in specificity. Domain swapping approaches seek to first define independent functional domains of a multidomain enzyme. These domains can then be recombined with functional domains of other enzymes to produce a chimaera with a new, combined function.

In one example of the mutagenesis-based approach, the LHE I-OnuI was engineered using directed evolution to preferentially target the human MAO B gene[4]. The MAO B gene has been implicated in the development of Parkinson's disease, and MAO B itself is a therapeutic target. Within the MAO B gene there is a sequence that differs from the native I-OnuI target sequence at only 5 base-pairs. The engineered LHEs were generated by saturating mutagenesis of amino acid positions identified from crystal structures to be in contact with those nucleobases that differed between the native and MAO B target site, and selected over several rounds of directed evolution. Ultimately, an engineered I-OnuI variant named I-OnuI E2 was developed that showed an ~2.5-fold preference for the MAO B target over the native target site. This engineered

LHE was later incorporated into a chimaeric fusion with the ND of I-TevI to create a dual-cutting endonuclease that presented a high frequency of site-specific gene disruption in mammalian cell culture, without the need for end processing enzymes such as Trex2[113].

In an example of the domain swapping approach, the pseudosymmetry of LHEs was capitalised on to generate an engineered HE named E-DreI[114,115]. The "left" and "right" domains of E-DreI were derived from two distinct members of the LHE family, I-DmoI, and I-CreI. The N-terminal domain of I-DmoI and a monomer of I-CreI were computationally combined, and amino acid substitutions were identified that optimised the interdomain interface. In this way, the authors generated a chimaeric HE that targeted a combined target site with an enzymatic efficiency on par with the parent enzymes. Building on the success of E-DreI, 30 chimaeric LHEs were generated, of which 14 displayed catalytic activity[116].

1.2.8   GIY-YIG Homing Endonucleases: I-TevI

GIY-YIG HEs possess traits consistent with the prototypical GIY-YIG HE I-TevI. I-TevI has an N-terminal GIY-YIG ND and a C-terminal helix-turn-helix (HTH) DBD tethered by a flexible linker region (Figure 1.4A)[5], making I-TevI inherently modular. A crystal structure of the C-terminal DBD with substrate, coupled with affinity assays reveals that the DBD is responsible for most of the enzyme's binding affinity and sequence recognition[117,118]. Functional characterisations have identified the N-terminal ND as being responsible for cleavage[118], where it acts as a monomer to target sites based upon both distance from the DBD and its own limited sequence specificity[119]. Cleavage
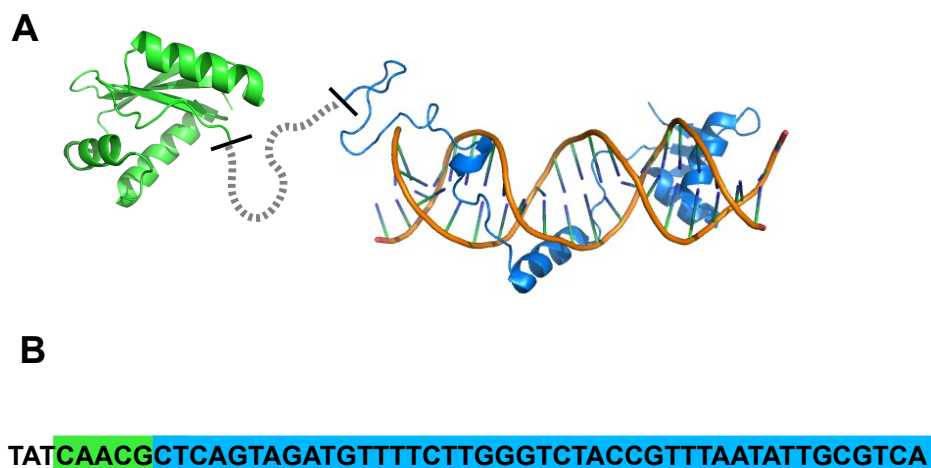
Figure 1.4. Structure of I-TevI HE and its cognate target site convey corresponding modularity. The structure of I-TevI (panel A) is composed of an N-terminal ND (green, from PDB 1MK0), connected *via* a flexible linker to a C-terminal DBD (both in blue, from PDB 1IJ3). A portion of the I-TevI linker region did not form a single ordered structure in the cocrystal, and is thus shown diagramatically as a dotted grey line. Similarly, the cognate homing site (panel B) can be divided into the cleavage motif (green), which is connected by a spacer to the I-TevI binding site (both in blue).

is effected in two sequential nicking reactions, where the bottom strand is nicked first in a metal independent reaction, prior to the Mg^{2+}-dependent nicking of the top strand[119]. The I-TevI ND recognises a five-basepair cognate cleavage motif, 5' – CAACG – 3'. Further investigation of the I-TevI ND cleavage profile revealed that it recognises the more general cleavage motif, 5' – CNNNG – 3'[120], equivalent to ~6 bits of information. There is no clear pattern for cleavage of the 64 possible triplets (NNNs). Although some NNNs are not cleaved, promiscuity abounds, and cleavage efficiency spans three orders of magnitude[113]. The lack of direct correlation between NNN sequence and cleavage efficiency indicates that indirect readout is likely playing a role in cleavage motif recognition.

Investigation of I-TevI is hampered by the inherent toxicity of this enzyme to *E. coli* cells[121], which precludes traditional overexpression and purification techniques. For this reason, studies of the I-TevI ND have been carried out using fusions of the ND with other DBDs[6] – analogous to FokI-based nucleases – or by extrapolating from experiments done using a close relative of I-TevI, I-BmoI[122].

The structural and functional modularity of I-TevI is evident in its modular target site (Figure 1.4B)[123], which is consistent with the view that the ND alone is responsible for catalysis and is a contributor to cleavage motif sequence recognition. Although the low binding affinity and dynamic nature of its mechanism[118] make the I-TevI ND less accessible to characterisation by techniques that rely upon stable interactions (*e.g.* crosslinking, STD-NMR, FRET, SPR, x-ray crystallography, or ITC), important residues have been identified by mutagenesis. R27[118], H40[123], and E75[124] have all been identified as important catalytic residues, as I-TevI R27A or E75A were unable to effect DSBs,

and H40Y had reduced activity. The function of these and other residues have also been predicted by homology, using GIY-YIG NDs that exist in a variety of enzymes, including I-BmoI, UvrC, Eco29kI, and Hpy188I (see below, Figure 3.5). Although indirect methods have fostered a greater understanding of the I-TevI ND, the knowledge required for re-engineering is currently incomplete.

As alluded to above, the I-TevI ND domain and its associated specificity can be ported to other DBDs. This ND has been successfully paired with Zinc fingers[6], TALEs[125, 126], and LHEs[6, 113]. The value of such portability has already been demonstrated by the extensive use of the FokI ND, described above. In the case of a combined I-TevI and LHE chimaera (MegaTev), the combined cleavage activities and specificities of these two enzymes together has been demonstrated to efficiently effect target gene disruption[113]. Furthermore, the chimaeric MegaTev is not as toxic to *E. coli* as I-TevI, and can be overexpressed and purified[113]. For these reasons, I used a fusion of the I-TevI ND to a catalytically inactive variant of the LHE I-OnuI – where it functions as a DBD – to generate the results described in this thesis.

The I-TevI ND possesses a number of characteristics that make it a potentially useful component of genome editing tools. As mentioned above, the ND is active as a monomer, which simplifies engineering constraints, and it has its own sequence specificity, which reduces off-target cleavage. However, the use of this ND in genome editing is restricted by its limited specificity, driving the need for I-TevI ND variants with altered sequence specificity. This is made challenging by the lack of information about the exact source of its cleavage specificity. Further, rational design is made impossible by the lack of a co-crystal of the holo-enzyme, complete with substrate, thus

directed evolution approaches are indicated.

1.3   Towards Engineering and Understanding the I-TevI ND

1.3.1   Hypothesis & Objectives

As described above, the I-TevI ND is highly sequence tolerant, likely stemming from the use of indirect readout to recognise and cleave the 5' – CNNNG – 3' cleavage motif. Understanding the mechanism by which readout is conveyed by I-TevI will facilitate engineering of this portable ND. *Thus, my null hypothesis is that readout of the cleavage motif is not conveyed by residues of the ND, and that altering these residues will not alter the cleavage profile of the I-TevI ND.*

To test this hypothesis I pursued several research objectives:

Objective 1)   Create a library of I-TevI ND mutants using random mutagenesis.

Objective 2)   Use a directed evolution approach to selectively identify mutants that are active on cleavage motifs that the wild-type I-TevI ND is not.

Objective 3)   Identify which individual mutations or combinations thereof are responsible for conferring said cleavage activity.

Objective 4)   Overexpress, purify, and kinetically characterise a mutant MegaTev with a new cleavage activity to determine its cleavage profile.

1.3.2   Scope & Relevance

My goal in this thesis was to identify mutations, and thus amino acid substitutions of the I-TevI ND that alter its cleavage profile. I expect that the positions of

these substitutions correspond to those that convey indirect readout of the cleavage motif. By identifying amino acids that are putatively involved in conveying indirect readout I hope to provide a means to better understand and potentially engineer I-TevI NDs with a cleavage profiles that are orthogonal to that of the wild-type ND. Such NDs could contribute to the development of a therapeutic genome editing tool.

The following chapter (Chapter 2) details the methods I used to carry out my experimental objectives. Chapter 3 recounts the results of my selections, the mutants I identified and their characterisation *in vivo*. Further, it described the *in vitro* characterisation of the I-TevI ND triple mutant T3, which was revealed to have a significantly relaxed cleavage specificity. Chapter 4 discusses the implications of the mutations that I have identified with respect to previously identified mutations, and presents the experiments that are now possible, and that I intend to carry out in pursuance of my PhD. Finally, supplementary figures and tables present the results of individual *in vivo* and *in vitro* assays, and specify the bacterial strains, plasmids, and primers used to complete this work.

Chapter 2   METHODS

All bacterial strains, plasmids, and oligonucleotide primers are listed in supplementary Tables S.1, S.2, and S.3, respectively. All restriction enzymes were acquired from NEB. Unless stated otherwise, small molecule reagents were acquired from EMD.

2.1   Construction of Mutagenised I-TevI Libraries

I-TevI ND mutant libraries were generated using Mutazyme II (Agilent), a mix of DNA polymerases for error prone PCR. Primers DE-840 and DE-1912 were used to select the region to be mutagenised. 0.2 ng of the I-TevI ND was then mutagenised throughout amino acids 10 – 95 under manufacturer-defined conditions for 30 PCR cycles. Mutagenesis was repeated as before for another 30 cycles in a fresh reaction to further increase the extend of mutagenesis before end-point PCR with *Taq* DNA polymerase (NEB) was used to amplify the mutant ND sequences. A truncated I-TevI linker region (residues 96-169) was amplified using end-point PCR with *Taq* and primers DE-1424 and DE-1045, and then combined with the I-TevI ND mutant library using splicing by overlap extension (SOEing) PCR with *Phusion* DNA polymerase (Thermo Scientific) and primers DE-840 and DE-1045. The ND mutant library with wild-type linker was digested with NcoI-HF and BamHI-HF and ligated using T4 DNA ligase (NEB) into the PciI and BamHI sites of an I-OnuI E1 E22Q with hexahistidine tag encoding plasmid, pACYCOnuE1E22Q(+H). Negative ligation controls were conducted by omitting insert in a parallel ligation set up. Complexity of library was determined based upon difference between colony count on ligation plate, and colony

count on the negative ligation control. In the final construct, the library of chimaeric MegaTevs with mutant NDs, was downstream of a T7 promoter/*lac* operator and a ribosome binding site, and upstream-adjacent to a sequence encoding residues 4-307 of I-OnuI E1 E22Q with a C-terminal hexahistidine tag.

2.2   Directed Evolution and Selection of Variants

Electrocompetent cells for directed evolution were prepared from *E. coli* strain BW25141(λDE3) transformed with a pTox plasmid as described previously[6]. Batches of electrocompetent cells were tested for lack of retention of the toxic plasmid by transforming them with 10 ng pACYCDuet-1; batches of cells that displayed survival greater than 0.1% under selective conditions (expression of the toxic protein, Ccdb) were discarded. Typically, 50 μL of electrocompetent cells were transformed with 10 ng of plasmid harbouring the I-TevI ND mutant library, and immediately diluted with 500 μL of SOC media for incubation at 37°C with shaking (280 RPM) for an amount of time that depended on the round of selection underway. For the first round of selection, cultures were incubated at 37°C for 6 h, while subsequent rounds were incubated for 1 h. 100 μL was diluted and plated as described below for *in vivo* survival assays. Another 200 μL was removed and diluted into two separate 5 mL aliquots of lysogeny broth (LB) media: a "non-selective" media with chloramphenicol (25 μg/mL) alone, and a selective media that also contained arabinose (10 mM). The diluted cultures were incubated at 30°C with shaking (280 RPM) for 18 h before being harvested by centrifugation and their plasmids isolated using a plasmid miniprep kit (Bio Basic) for subsequent rounds of selection.

After two rounds of selection, those populations of mutant NDs that showed a measureable increase in survival were PCR amplified with primers DE-840 and DE-1045. The amplified DNA was treated with DpnI (NEB) to destroy any remaining round 2 plasmids, digested with NcoI-HF and BamHI-HF, and religated back into pACYCDuet-1(PciI).

Round 4 survivors were sampled by picking five colonies from the corresponding selective plates, and incubating them overnight in 5 mL LB media with chloramphenicol (25 μg/mL) at 37°C, for subsequent plasmid isolation with a plasmid miniprep kit.

2.3   *in vivo* Survival Assays

Electrocompetent cells harbouring pTox plasmids were generated as described in the previous section, and were typically transformed with 50 ng of plasmid harbouring a MegaTev with a mutant ND, and immediately diluted with 500 μL of SOC media for incubation at 37°C for 1 h. Cultures were diluted 1/1, 1/10, 1/100, 1/1000 and 1/10000, and 100 μL of diluted culture was plated on selective (chloramphenicol [25 μg/ml] and arabinose [10 mM]) and non-selective (chloramphenicol [25 μg/ml]) LB media, and incubated for 20 h at 37°C, and colonies counted. Data quality was improved by discarding plates that did not meet the following criteria: colonies were only counted on those plates that had >10 colonies (preferrably hundreds), or >0.1 % survival, whichever was greater.

2.4   Construction of I-TevI Nuclease Domain Mutants

Construction of I-TevI NDs that were not identified by sampling round 4 survivors was typically achieved by SOEing together fragments of previously identified I-TevI ND mutants using *Phusion* DNA polymerase. To generate K26R, primers DE-840 and DE-1912 were used to amplify the ND of K26R Q158R from amino acids 1 to 95, the wild-type linker region of I-TevI was amplified using primers DE-1045 and DE-1424, and these two regions were joined using SOEing PCR as above for library construction. Similarly, K26R T95S Q158R was made by combining the ND from K26R T95S and the linker from Q158R. To generate K26S mutants, a pair of complementary primers with single basepair mismatches to the wild-type I-TevI ND sequence installed a K26S mutation. These primers were used in combination with primers DE-840 and DE-1045 to construct each of the K26S mutants. A similar strategy was used to restore T95S to T95 by amplifying the I-TevI ND with primers DE-840 and DE-2167, and combining it with the linker sequence as above in order to generate the single mutants C39R, and I86V.

2.5   Purification of Chimaeric MegaTevs

2.5.1   *Overexpression of Chimeric MegaTevs in* E. coli

Plasmids harbouring MegaTevs comprised of either a wild-type I-TevI ND or the T3 ND were transformed into ER2566 *E. coli* cells (NEB), plated on LB media with chloramphenicol (25 µg/mL), and incubated for 18 h at 37°C. A single colony per plate was picked and used to inoculate a 20-mL LB culture (with 25 µg/mL chloramphenicol), which was incubated at 37°C for 4-6 h before being diluted into 1 L LB culture (with 25 µg/mL chloramphenicol) and grown to $OD_{600} = 0.8$. The culture was then chilled on ice

for 30 min, and 1 mL of 1 M isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce enzyme expression, before being incubated for a further 13 h at 15°C. The cells were then harvested from the culture (now at OD_{600} of 1.1-1.4) by centrifugation (4000 ×g, 10 min), and the pellet collected and stored at -80°C for 16-24 h.

2.5.2 *Chromatographic Purification of Chimaeric MegsTevs*

The cell pellet was resuspended into 35 mL of binding buffer (20 mM Tris·HCl [pH 8.0], 500 mM NaCl, 5% (*v/v*) glycerol, 10 mM imidazole) containing a protease inhibitor mix (1/20th of a cOmplete™ protease inhibitor pellet [Roche, added as a suspension of pellet in ddiH$_2$O], per gram of cell pellet). Cells were then sonicated (power 5, 50% duty cycle, pulsed mode, 5 × 20 pulses), and subjected to centrifugation (20,000×g, 15 min) to separate the cell pellet from the soluble fraction, which was removed and applied to a His-Bind column (Amersham). The column was then loaded with a procession of buffers: ~45 mL of binding buffer, 15 mL of wash buffer (binding buffer with 50 mM imidazole), and 5 mL of elution buffer (binding buffer with 300 mM imidazole). The final 5 mL of eluate was dialysed (10,000 Da molecular weight cutoff (MWCO) [Spectra/Por]) against 500 mL storage buffer (20 mM Tris·HCl [pH 8.0], 500 mM NaCl, 5% (*v/v*) glycerol, 1 mM dithiothreitol [DTT]) for 6 h at 4°C, before the buffer was replaced, and dialysis continued for a further 12 h. In the case of the MegaTev with a wild-type ND, aliquots and frozen at -80°C; these aliquots were typically active for over a month when stored in this fashion. In the case of MegaTev T3, the dialysed stock was kept at 4°C and used within a week.

2.5.3   Determining MegaTev Quality and Quantity

All MegaTev purifications were followed by electrophoretic separation to determine the purity of the aliquots before the concentration of MegaTev is quantified. The concentration of MegaTevs in solution was determined by measuring the UV absorbance of the solution at 280 nm (A_{280}) and comparing it to the predicted extinction coefficient (ε_{280}) of the chimaeric MegaTev (67380 $M^{-1}{\cdot}cm^{-1}$). Predicted ε_{280} values were calculated using the "ProtParam" tool on the ExPASy website (http://web.expasy.org/protparam/) assuming no disulfide bonds. The precise concentration was determined using the Beer-Lambert-Bouguer law[127-129] (eqn 2.1; c is the molarity, b is the pathlength [in cm]). It was assumed that the only significant protein component in solution was MegaTev, on the basis of the SDS–PAGE results (Figure 3.7, below).

$$c = A_{280} / (b{\cdot}\varepsilon_{280}) \tag{2.1}$$

2.6   Barcode Assays and Kinetic Characterisation of Chimaeric MegaTevs

Barcode assay substrates were prepared by using pTox as template with a pair of flanking primers equidistant from the cleavage motif (see supplementary Table S.3), in end-point PCR. Substrates of 2200, 1900, 1600, or 1320 bp were made, and combined into a single reaction. Substrates contained a 42 bp MegaTev target site comprised of a 5 bp cleavage motif, a 15 bp spacer from the I-TevI native target, and a 22 bp I-OnuI E1 target sequence from the human MAO B gene. The cleavage motif was placed such that substrates would be cleaved in half to create two equally-sized products. Unreactive pre-

mixtures were prepared on ice, and were comprised of 5 nM of each substrate, 250 nM enzyme, and cleavage buffer (50 mM Tris·HCl pH 8.0, 100 mM NaCl, 1 mM DTT, 5% glycerol). An aliquot of pre-mixture was removed immediately prior to starting the reaction by adding 2 mM MgCl$_2$, and incubating at 5°C for 30 min. Aliquots were removed from the reaction mixture at 1, 2, 4, 10, and 30 min time-points (although for practical purposes some aliquots were removed at 11 or 13 min instead of 10 min), and quenched by the introduction of ethylenediaminetetraacetic acid (EDTA) and sodium dodecylsulphate (SDS) (final concentrations of 83 mM, and 8.3%, respectively). Time-points were resolved using agarose gel electrophoresis in TBE (100 mM Tris base, 100 mM boric acid, 2 mM EDTA [pH 8.0]) and spot densitometry was used to measure the quantity of substrate remaining in the reaction, and the quantity of product formed. The intensity of the corresponding substrate and product bands at each time-point are summed, and normalised to the to the intensity of the substrate band at $t = 0$ (forcing mass balance). The fraction of substrate remaining (f_S) is then simply the ratio of the normalised substrate band intensity to the initial intensity. Triplicate values were plotted as fractions of substrate remaining at each time-point, and fit by non-linear regression to a first-order decay curve (eqn 2.2, where f_S is the fraction of remaining substrate, m_1 and m_2 correct for a non-zero baseline or non-unity starting condition, respectively, m_3 is the k_{app} in reciprocal minutes, and t is the amount of time passed, in minutes). The apparent first-order rate constant of decay (k_{app}) was normalised to k_{app} for the native cleavage motif decay curve, and reported as relative k_{app}.

$$f_S = m_1 + m_2^{-m_3 t} \qquad\qquad (2.2)$$

2.7   Sequencing of MegaTev T3 Cleavage Products

Barcode assay substrates harbouring position 1 substitutions or the native cleavage motif were digested with MegaTev T3 for 1 h at 37°C, substrates were isolated from enzyme using a PCR cleanup kit (Bio Basic), and submitted for Sanger sequencing with one of two flanking primers to obtain the sequence of the top strand (DE-410) and bottom strand (DE-411).

Chapter 3   RESULTS

3.1   Mutagenesis, Genetic Selection, and Isolation of I-TevI Nuclease Domain Mutants

My first research objective was to create a library of I-TevI ND mutants, with the goal of isolating mutants with activity on cleavage motifs that are poor substrates for the wild-type enzyme. To accomplish this objective, I first generated a library of I-TevI ND mutants using an end point PCR technique that makes use of a mix of engineered DNA polymerases that ensure an equal proportion of each mutation (*e.g.* A→C, G, or T).

I wanted to ensure that the library was of sufficient complexity to contain all possible single amino acid substitutions at every position of the ND (20 amino acid possibilities for each of 86 positions from 10-95, or 1720 single amino acid substitutions in total). The complexity of the initial library was assessed in two ways: by the number of successful transformants made with the library, and by the number and variety of mutations found therein. *E. coli* BW25141(λDE3) were transformed with the MegaTev ND mutant library and a subset were plated on LB media with chloramphenicol (25 µg/mL). After an 18 h incubation at 37°C, colonies arising from this subset of the full culture were counted, and their number extrapolated to the full culture volume. Based upon the number of colony forming units, the library was estimated to contain approximately 70,000 cfu. Six of these colonies (LIB-1-LIB-6) were chosen at random, grown overnight, and harvested to isolate their pENDO plasmids, which were sequenced (Figure 3.1). One of the sequenced plasmids appeared to have undergone an insertion/deletion reaction (LIB-4), and as a result, ~90% of its sequence had been frameshifted; this sequence was excluded from further analyses. The other sequences

Figure 3.1

Figure 3.1. PCR with Mutazyme II Consistently Mutates the I-TevI Nuclease Domain. Sequences of six clones from the the mutagenised nuclease domain library (LIB-1 – LIB-6) are shown above, amino acid sequences on the left, and nucleotide sequences on the right. In each of the six NDs there was a single, double, or triple amino acid substitution. An indel in LIB-4 has resulted in a frameshift mutation that has affected almost all of the amino acid sequence.

were each revealed to have 1-4 amino acid substitutions, either transversions or transitions, as summarised in Table 3.1 and 3.2.

A ratio of transversions and transitions as close to 2:1, but certainly greater than 1:1 was considered important because there are twice as many codons available through transversions, and thus a greater set of possible mutations at a particular amino acid position. The extent of mutagenesis observed in this sampling of the library yielded a ratio of transversions to transitions of 1.5:1.

With a sufficiently complex library of I-TevI ND mutants in hand, my next objective was to identify mutants that were active on cleavage motifs that the wild-type ND is not. This objective was accomplished using a directed-evolution approach, in the context of the MegaTev chimaeric nuclease. The library of randomly mutagenised I-TevI NDs was fused *via* a partial I-TevI linker to a catalytically inactive I-OnuI E1 E22Q and subjected to multiple rounds of selection and enrichment using a bacterial 2-plasmid assay, delineated in Figure 3.2. This assay facilitated rapid phenotypic screening of a library in a stringent, selective system with an easily controlled selective pressure in the form of a double stranded plasmid DNA-substrate (pTox). pTox harbours a toxic gene (*ccdb*, encoding the topoisomerase-inhibiting peptide, Ccdb), which is under arabinose-mediated metabolic control (using the araBAD promoter). In this system, cleavage of the target site linearises pTox, which is then degraded by the *E. coli* RecBCD complex, allowing growth of cells with an active endonuclease. This selection is bacteriostatic, not bacteriocidal, because the CcdB toxin inhibits DNA gyrase. Thus, even very limited cleavage of pTox was sufficient to overcome the selective challenge. Since the selections were done with a library of ND mutants under direct competition,

Table 3.1. Survey of Observed
Nucleotide and Amino Acid Substitutions

Nucleotide Substitutions	# (%)	Amino Acid Substitutions	# (%)
Transitions	6 (0.4)	Missense	10 (2.35)
Transversions	9 (0.9)	Nonsense	2 (0.47)
Total	15 (1.5)	Total	12 (2.82)
Indels	1 (N/A*)		

*Percent calculations did not include the sequence with an indel

Table 3.2. Survey of mutation rates in sample sequences.

		Final					
		T	C	A	G		
Initial	T	-	3	4	1	-	Identity
	C		-				Transition
	A	2		-	2		Transversion
	G	1		2	-		
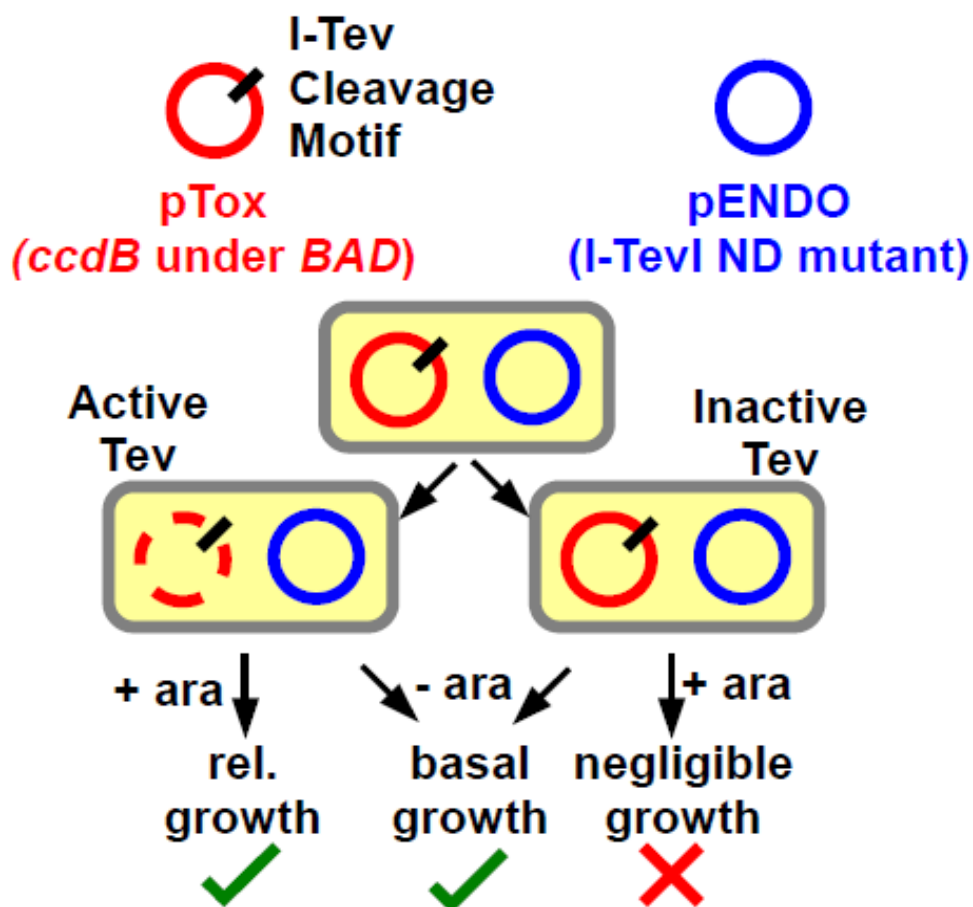
Figure 3.2. Bacterial two-plasmid selection discriminates between sufficiently or insufficiently active I-TevI cleavage domains. Plasmids encoding the mutant enzymes are transformed into *E. coli* harbouring a second plasmid encoding the toxic gene *ccdB* under the arabinose-inducible *BAD* promoter, and a putative endonuclease target site. If the target site is cleaved, the plasmid encoding the toxic gene is rapidly degraded, and similar growth is observed in the presence and absence of arabinose; if, however, the target site is intact, negligible growth is observed in the presence of arabinose. By comparing the relative growth under selective (+ara) *vs.* non-selective (-ara) conditions, % survival was determined.

alleviating the bacteriostatic effect granted a critical selective advantage over non-replicating or slowly replicating competitors. This approach was chosen because the strongest survivors would be enriched and would dominate the population.

I expected that the mutations present in my library would be more likely to broaden the cleavage profile of the I-TevI ND, rather than fully eliminate activity against the native 5' – CAACG – 3' cleavage motif. Thus, if the library did not retain any residual activity against the native cleavage motif, then it would have indicated that the level of mutagenesis was too high, and had led to complete attenuation of ND activity. Survival assays testing the library against the native cleavage motif revealed that the library was able to survive (Table 3.3), albeit at a reduced level compared to the wild-type ND, which has previously conferred 100% survival.

Confident that I had a library which contained active I-TevI ND mutants, I chose 16 substrates that have been shown previously to be highly cleavage resistant to cleavage by the wild-type I-TevI ND[113] from the set of all 64 NNNs as the first priority for extending the versatility of the I-TevI ND through a broadened cleavage profile. I anticipated that mutations which resulted in cleavage of a poor substrate were likely to directly influence nuclease activity, rather than result from indirect effects such as increased protein stability or expression. The initial library was screened against all 16 poor substrates one by one (R1, Figure 3.3). Each screen required an independent transformation of the library into competent cells harbouring an individual substrate. For R1 only, freshly transformed cells were incubated in SOC media at 37°C for 6 h, before selection proceeded for 18 h at 30°C. This generated 16 populations of enriched I-TevI ND mutants.

Table 3.3. Summary of initial library and R4 population survival rates against selected substrates and specific mutations derived from R4.

Population	Selected Against	wild-type Survival Rate (n=3, %)	R4 Survival Rate (%)	R4 Survival Rate vs CAACG (n=2, %)	Mutant	# of Observations
A	AAG	1.6 ± 0.8	16	143, 64	Q158R	4
					K26R Q158R	1
B	CCC	nil*	36	113, 109	T95S	5
C	GAA	13.5 ± 6.1	61	86, 87	Q158R	5
					I86V T95S	1
D	GCC	nil*	20	84, 81	C39R T95S	1
					C39R I86V T95S	3
E	GGA	nil*	29	107, 133	T95S	3
					K26R T95S	2
F	TGG	nil*	37	55, 87	T95S	2
					K26R T95S	1

Initial Library Survival Rate vs CAACG (n=2, %): 17, 6

*values of survival rate did not exceed background survival of negative controls (0.1%)
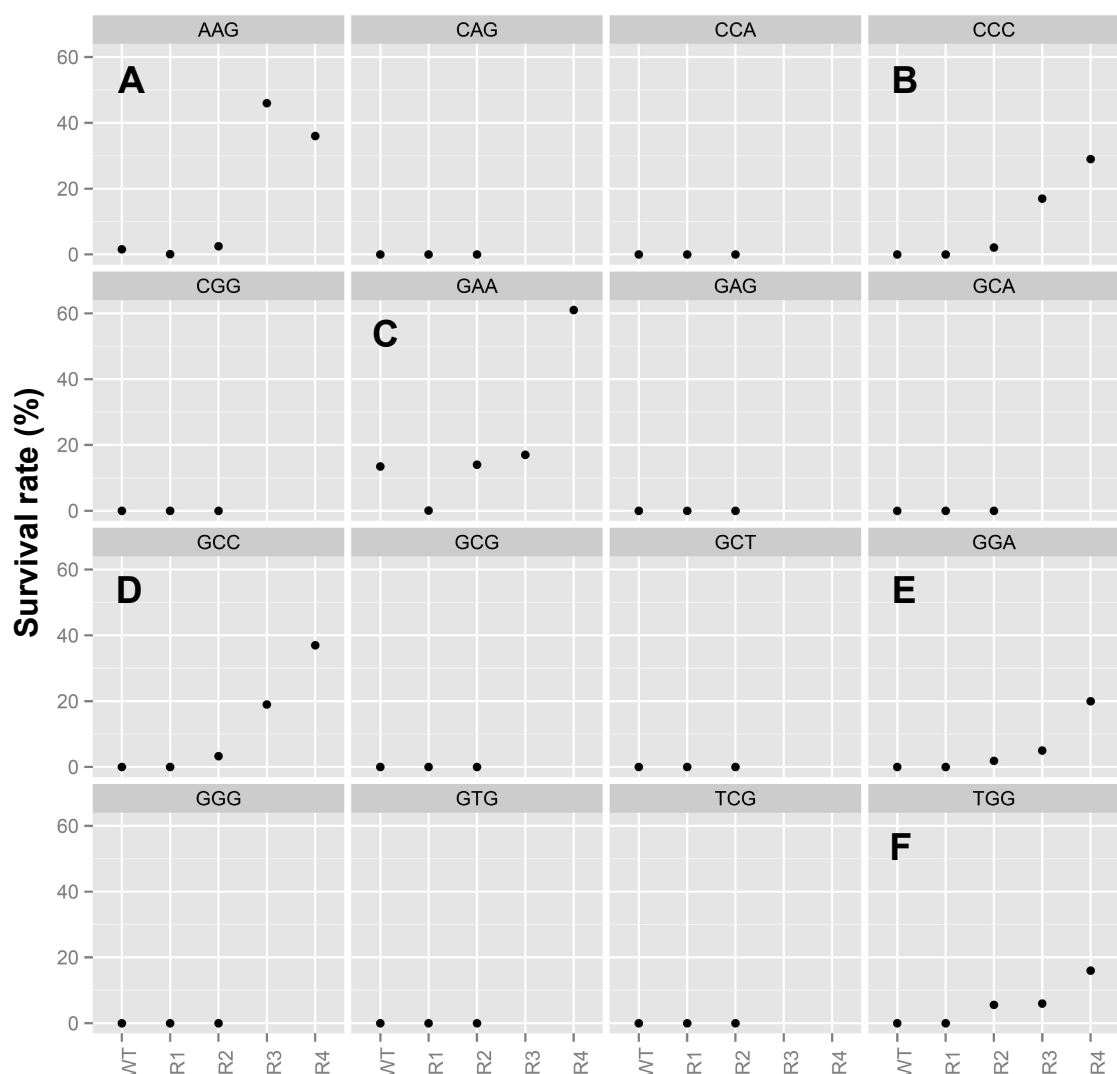
Figure 3.3. Mutant populations of I-TevI cleavage domains confer survival against toxic plasmids harbouring CNNNG cleavage motifs with cleavage resistant triplets. After a round of selection on the library (R1), the population of survivors against each substrate was isolated, and subjected to a second round of selection (R2). Those populations (A-F) that confer a measurable improvement in survival over wild-type (WT), were recloned and subjected to two additional rounds of selection (R3 & R4).

In order to further enrich the mutants that cleaved poor substrates, the resulting 16 populations were each subjected to another round of selection against their respective substrates (R2, Figure 3.3). For this and all further rounds of selection, selection was made more stringent by incubating freshly transformed cells for only 1 h, rather than 6 h. The survival rate of each population after round 2 was compared to survival of the wild-type I-TevI ND against the same substrate (*e.g.* survival rate of population A on 5' – CAAGG – 3', compared to survival conferred by the wild-type enzyme against 5' – CAAGG – 3'). Only those populations that showed any observable improvement in survival over the wild-type I-TevI ND (wt, Figure 3.3) in R2 were pursued further. Improvements in survival compared to the wild-type enzyme were often very clear, since the wild-type did not survive to any extent. In these cases, survival greater than 0.1% (*i.e.* greater than background survival observed with an inactive ND) was sufficient to merit further rounds of selection. In those situations where the wild-type enzyme did confer survival to some extent, survival equal to, or greater than the wild-type was deemed sufficient. Such a lenient margin of success was chosen because mutation is expected to reduce activity in general. Thus populations that were indistinguishable from the wild-type enzyme in terms of activity would be expected to also contain individual mutants that were more active than wild-type.

The populations I obtained in R2 could have been the result of mutations outside of the I-TevI ND (*e.g.* chance mutations to promoter leading to increased expression of endonuclease). I wanted to ensure that only mutations to I-TevI were maintained, and so I recloned the open reading frames (ORFs) containing the I-TevI ND and partial linker from each population into fresh background vector prior to further rounds of genetic

selection.

After four rounds of selection (R4), six populations (A-F) were identified that showed a marked improvement in survival, ranging from 4-fold for population C, to >370-fold for population F, as summarized in Table 3.3. I wanted to assess to what extent these populations had diverged from the initial library, and the wild-type I-TevI ND. Namely, I wanted to know whether or not my directed evolution approach had selected enzymes that preferred the cleavage motif they were selected against (*e.g.* 5' – CAAGG – 3' for population A) over the native cleavage motif (5' – CAACG – 3'). Unfortunately, bacterial 2-plasmid assays of R4 populations against the native cleavage motif revealed that the native cleavage motif was still preferred by each of these populations.

Although the R4 populations preferred the native cleavage motif, they displayed substantial improvements in survival over the wild-type I-TevI ND against poor substrates. Thus I wanted to know what mutations were present in these populations that might confer said survival. Five colonies were chosen from plates of survivors on selective media from each of the six populations (A-F), and their MegaTev ORFs sequenced to identify their mutations. The number of each mutant genotype observed in each population are tabulated in Table 3.3. One surprising mutation I observed (Q158R) was outside of the mutagenised ND region of I-TevI, and instead was found in the partial I-TevI linker. Otherwise, all mutations were observed within the I-TevI ND region. Importantly, none of the MegaTev ORFs sequenced contained the wild-type I-TevI ND and partial linker.

3.2   *in vivo* Characterisation of I-TevI Nuclease Domain Mutants

Having isolated several mutant I-TevI NDs, I was interested in understanding how – and critically if – each amino acid substitution is affecting the ND cleavage profile. To determine the effect of individual mutations, mutant I-TevI NDs were made containing single or double mutants from amino acid substitutions identified from the genetic selections. The ability of each of these substitutions to confer survival in a survival assay was determined in triplicate, and is summarised as a heatmap of average survival values in Figure 3.4, and a table of values in supplementary Table S.4.

Because survival is an indirect measurement of cleavage activity, I was concerned that the *in vivo* survival I had observed might be caused by a mechanism independent of cleavage. If substrate pTox plasmids did not obviate survival in the presence of a catalytically inactive I-TevI ND, then some cleavage independent mechanism of survival could be providing the results I observed. To confirm that the results I was observing required a catalytically active ND, triplicate negative control survival assays were conducted using each substrate and a chimaeric MegaTev with a catalytically inactive R27A ND mutant; no survival greater than 0.1 % was observed.

Each individual amino acid substitution conferred improvement in survival against cleavage-resistant substrates, which was generally enhanced when substitutions were combined. For example the K26R mutant displays an ~31% survival rate against 5' – CAAGG – 3'. Similarly, the Q158R mutant displays an ~53% survival rate against 5' – CAAGG – 3'. Combined, the K26R Q158R mutant displays an ~86% survival rate against 5' – CAAGG – 3', and is an example of a combination of mutations that led to an
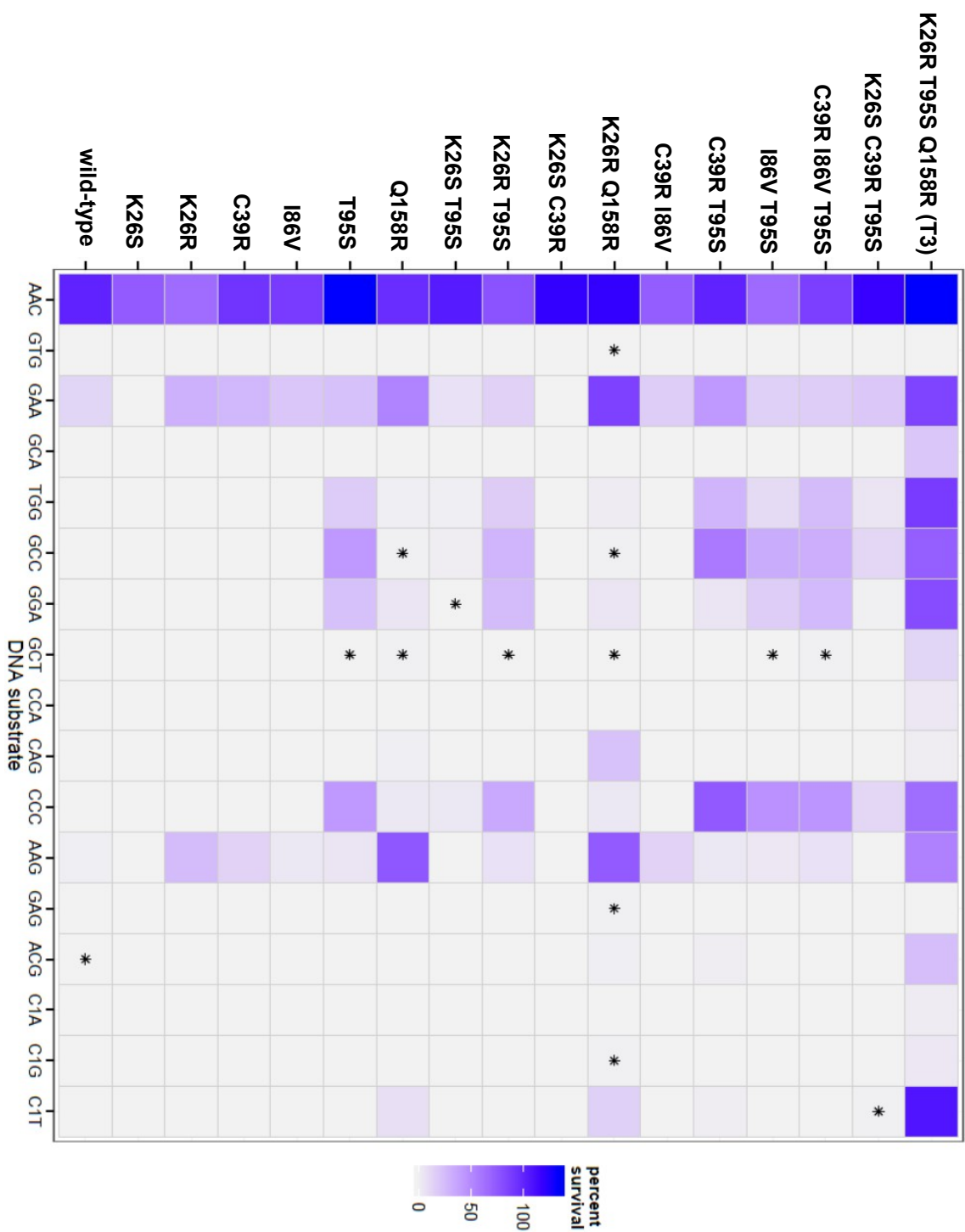
Figure 3.4

Figure 3.4. I-TevI ND T3 cleavage specificity is a combined effect of individual mutations. Individual mutations that were identified by survival assay screening were introduced into the I-TevI ND individually, or in combination, and their ability to confer survival in a 2-plasmid assay was assessed in triplicate. pTox plasmids harbouring the native cleavage motif, or one of 16 cleavage-resistant substrates, differing in their NNN triplet, were used in survival assays as described in the text. Values below 1% are marked with an asterisk.

additive effect on survival. The same combination of mutations resulted in an ~ 23% survival rate against 5' – CCAGG – 3', despite the observation that K26R has no impact on survival against that substrate individually, and Q158R has only an ~1.5% survival rate against 5' – CCAGG – 3'. The observation that Q158R and K26R together have a greater survival rate against 5' – CCAGG – 3' is an example of a combination of mutations that led to a cooperative effect on survival. Through a combination of additive and cooperative effects, the triple mutant K26R T95S Q158R (T3) conferred the highest survival rates against the broadest range of substrates tested.

Perhaps most promising was the ~100% survival conferred by T3 against a C1T substitution in the cleavage motif. C1 of the cleavage motif has previously been shown to be necessary for cleavage by the wild-type I-TevI ND. This result is the first indication that mutants could be developed that cleave targets which differ at this position of the cleavage motif.

Since these amino acid substitutions conferred enhanced survival under the conditions described above, they represented putative functionally important residues. Thus I expected that exchange of these residues with those found at analogous positions within another GIY-YIG ND would bestow some of that ND's substrate preference. Comparison of the I-TevI ORF with the related GIY-YIG HE I-BmoI, revealed that all of the positions identified here were also positions of non-identity with I-BmoI, as depicted in Figure 3.5. Thus, mutants were made that possessed amino acids consistent with I-BmoI at these positions (K26S, C39R, & T95S). Disappointingly, I-TevI ND mutants with these amino acid substitutions fared no better than wild-type against the substrate containing a CCCCG cleavage motif, which contains an NNN triplet identical

to the native I-BmoI cleavage motif GCCCG.

3.3 *in vitro* Barcode Assays

The triple mutant T3 was able to confer survival against a broad range of substrates that the wild-type I-TevI ND could not. However, these results could have been explained by other convoluting variables, such as decreased toxicity, increased *in vivo* stability, increased catalytic activity, other modes of pTox deactivation, or a combination of these effects; thus, *in vitro* characterisation was indicated.

The barcode assay developed by Monnat *et al.* for rapidly determining HE target sites[22] can be used to assess cleavage of four unique substrates in a single, competitive, *in vitro*, kinetic assay, and is described schematically in Figure 3.6. This assay can be used to quantitatively determine kinetic constants for individual substrates relative to the native 5' – CAACG – 3' cleavage motif. For these assays, two I-TevI chimaeras were overexpressed and purified as described above. The I-TevI chimaeras were comprised of the first 169 amino acids of I-TevI, comprising the ND and a partial linker region from either wild-type or T3, and a C-terminal, catalytically inactive I-OnuI E1 E22Q. Purifications resulted in active enzyme of sufficient purity to proceed with *in vitro* assays (Figure 3.7). Although enzyme activity was observed to decline over time (weeks for the wild-type ND and days for the T3 ND), it was assumed that this did not affect the relative cleavage of each substrate.

The I-TevI ND T3 was assayed *in vitro* against assorted substrates that were predicted to be poor substrates of the wild-type I-TevI ND, and those substrates that differed from a poor substrate by a single basepair (Figure 3.8, supplementary Table S.5,

Figure 3.5. Protein sequence alignment of I-TevI and I-BmoI NDs reveals key sequence dispairities between these orthologs that correspond to mutations identified in selections. Sequences of I-TevI and I-BmoI NDs were aligned using Clustal ω. Residues that were identified in selections are marked with red asterisks. Additional GIY-YIG domains from Eco29KI, Hpy188I, and UvrC are also aligned for reference.

Figure 3.6. 'Bar code' *in vitro* **cleavage assay facilitates quantitative assessment of mutant I-TevI cleavage domain activity.** Substrates of varying length (panel A), each bisected by a unique cleavage motif are combined into a single competitive reaction with a Tev-Onu mutant (panel B), started by addition of Mg^{2+}, halted by sequestration of Mg^{2+} by EDTA, and visualised on an agarose gel (represented by panel C). The varied length substrates facilitate measuring relative cleavage of each substrate. An example of an agarose gel is shown (panel D) from which band densities are measured and used to calculate disappearance of substrate over time (measured as $[S_t]/[S_0] = f_S$ or fraction of substrate remaining), which are plotted and fit with a first-order decay curve to determine the rate of decay (panel E).

and supplementary Figure S.1). Altogether, assays were conducted on 34 distinct substrates, using a 50-fold excess of endonuclease to ensure that substrates were saturated with bound endonuclease, and thus only the rate of cleavage was being measured. Furthermore, assays were conducted at 5°C to ensure that initial cleavage rates were slow enough to be measured. Each assay contained a substrate with the native 5' – CAACG – 3' cleavage motif, which was used as an internal standard for the rate of enzyme catalysed substrate decay (represented by k_{app}). All enzyme activities are reported relative to cleavage of the native cleavage motif standard (relative k_{app}). Although there is no consistent ratio between the wild-type and T3 ND rates of cleavage for any particular substrate, the T3 ND is generally more promiscuous (Figure 3.8).

The *in vitro* results were generally consistent with the *in vivo* results; increases in survival conferred by the T3 ND were associated with increased catalytic activity. However, there are cases where a small increase in survival rate was associated with a large increase in cleavage activity. For example, survival on the 5' – CGCTG – 3' cleavage motif by the T3 ND increased to 13% from 0% for the wild-type. This relatively modest increase in survival rate was associated with a nearly 4-fold increase in relative k_{app}, from 0.18 for the wild-type ND to 0.73 for the T3 ND. Conversely, there are also cases where a small increase in cleavage efficiency was associated with a large increase in survival. For example, survival on the 5' – CAAGG – 3' cleavage motif increased from 1.6% for the wild-type ND to 54% for the T3 ND. This pronounced increase in survival rate was associated with an only 1.3-fold increase in relative k_{app}, from 0.38 for the wild-type ND to 0.51 for the T3 ND. Although cases such as these do exist, they represent the minority. In the majority of cases, a large increase in survival
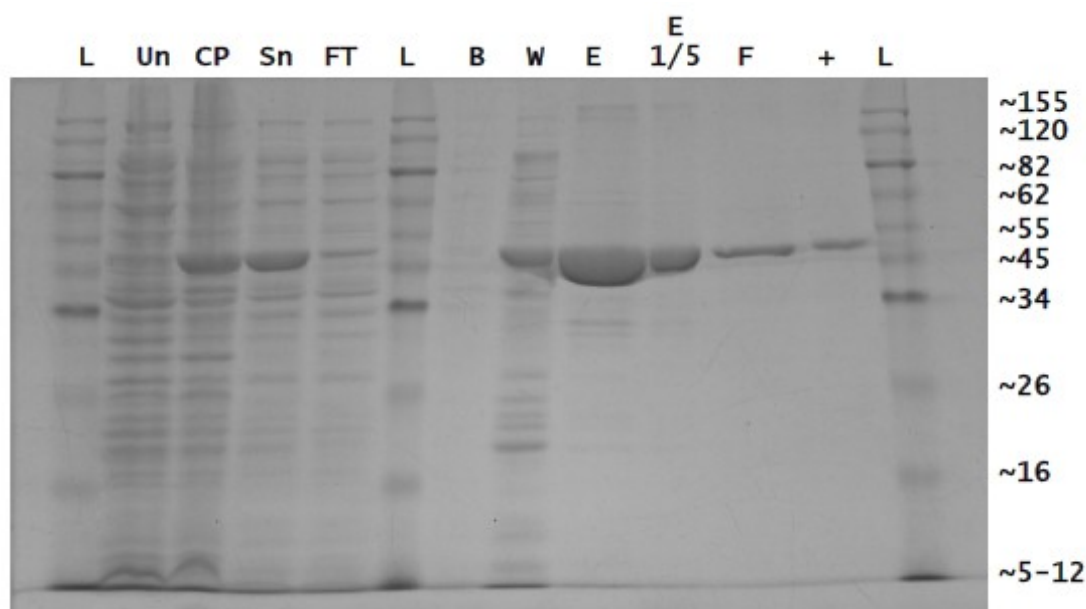
Figure 3.7. I-TevI chimaeras were isolated from *E. coli* ER2566 cells at >95% purity. The results of a typical purification is displayed by the polyacrylamide gel, shown above. Uninduced cells express an undetectable amount of the I-TevI chimaera (Un), but after induction with IPTG, I-TevI chimaera can be observed after lysis and centrifugation of the cells in both the insoluble cell pellet (CP) and supernatant (Sn). The I-TevI chimaera is retained by a His-Bind column, and is not observed in the flow-through (FT) or after the first wash with binding buffer (B). Wash buffer (W) does remove some of the I-TevI chimaera, but renders the elution (E, or diluted 1/5 as E 1/5) almost completely free of non-specifically bound proteins. Complete removal of Ni^{2+} from the column with EDTA reveals that very little I-TevI chimaera remains on the column (F), and that the purification consistently yields a protein of the predicted molecular weight (54.6 kDa), when compared to prior purification (+) and known molecular weight standards (L).
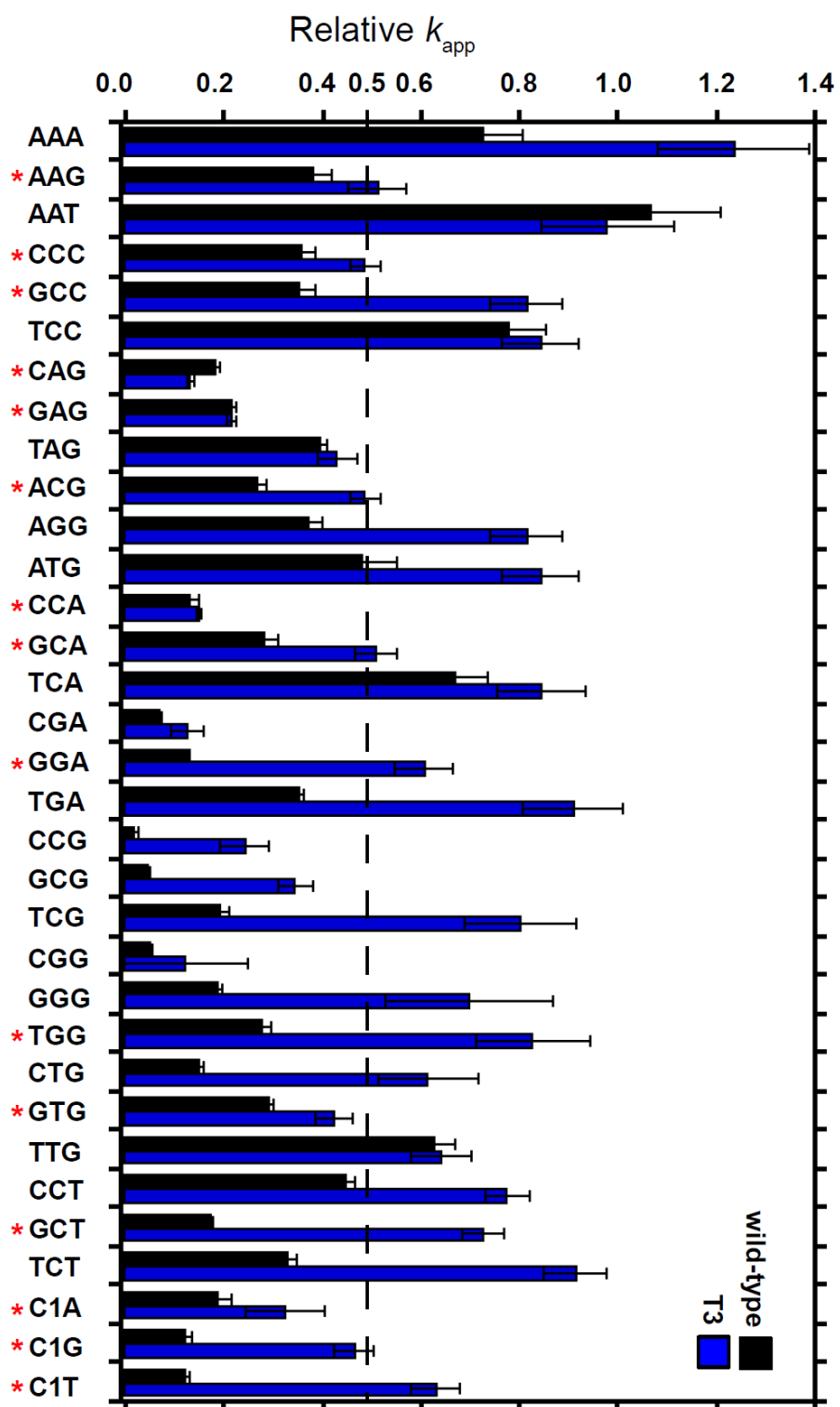
Figure 3.8

Figure 3.8. Kinetic assays reveal that the T3 mutant has a distinct cleavage profile. Incubation of barcode assay substrates with a chimaeric fusion comprised of either the wild-type I-TevI ND or the T3 ND, resulted in a first-order decay of said substrates. The apparent first order kinetic constant for this decay (k_{app}) was determined for each substrate, and normalised to the k_{app}-value for the native target site substrate present in each assay. These relative k_{app}-values are graphed for both the wild-type ND and the T3 mutant. The substrates used differed from the native target site by either the NNN triplet, or at position 1 of the cleavage motif. Further, they are either highly cleavage-resistant substrates used *in vivo* (marked with a red asterisk; *e.g.* the TGG triplet), or else related to such a substrate by a single nucleotide substitution (*e.g.* the triplets AGG, or TCG). Values that exceed the dashed line at 0.5 roughly correspond to those for which survival was observed *in vivo*.

 rate conferred by the T3 ND was accompanied by a correspondingly large increase in relative k_{app} value, and *vice versa*.

Because barcode assays are inherently competitive, any differences between the substrates could be contributing to their differing rates of cleavage. Thus I was concerned that the length of the substrates may have influenced my results. Although the exact mechanism by which the chimaeric MegaTev constructs bind their substrate is unknown, one possible mechanism involves a slow DNA binding step, followed by a rapid sliding of the MegaTev along the DNA helix to find its target sequence, as is the case for the RE EcoRV[130]. In such a situation, a longer substrate would be expected to have an accelerated DNA binding step, ultimately leading to faster cleavage.

 To determine if the length of the substrate had an impact on cleavage efficiency, the native I-TevI cleavage motif (5'–CAACG–3') substrates were synthesized in each of the four possible lengths (2200, 1900, 1600, and 1320 bp). These substrates were mixed, and cleavage monitored, as shown in Figure 3.9A. It was determined that substrate length had a negligible effect on cleavage rate.

Another consideration for any competitive assay must be the effect of residual substrates on the rate of cleavage of their competitors. This may be observed as a cooperative effect, where the cleavage of each substrate is enhanced or attenuated by the presence of its competitors. To determine if there was any impact of cooperative effects between multiple enzyme-substrate pairs on cleavage efficiency, a set of substrates were synthesized that contained a highly cleavage-resistant cleavage-motif 5' – AAACA – 3' (1A5A) in three lengths (2200, 1600, and 1320 bp) and a native cleavage motif substrate 5' – CAACG – 3' (1C5G) of the remaining length (1900 bp). As shown in Figure 3.9B,
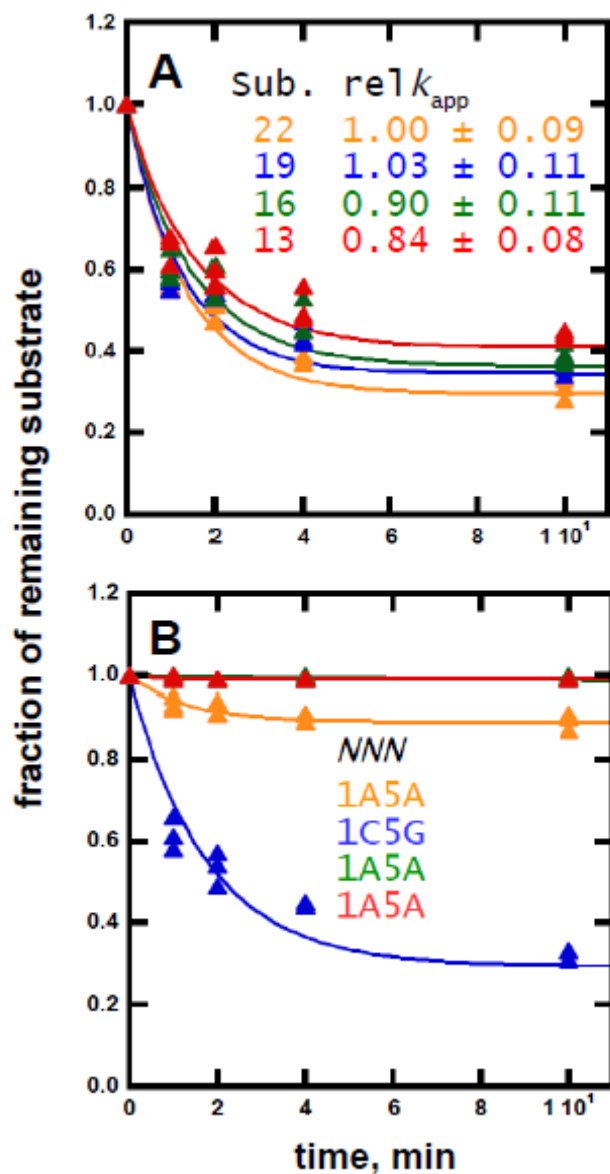
Figure 3.9. Cleavage assays were unaffected by the length of substrate used, or the concentration of other substrates in the reaction. Cleavage assays were conducted under the same conditions as similar bar code assays used previously. Despite being of different lengths 2.2, 1.9, 1.6 or 1.32 kbp, substrates containing the native cleavage motif were not cleaved at rates more disparate than standard error (A). Further, no effect of uncleaved substrate (C1A G5A [1A5A]) on cleavage of the native cleavage motif (1C5G) was observed (B).

only the native cleavage motif showed any significant decay over the assay period, and this decay was aptly fit by a first-order decay curve ($R^2 = 0.98$). Despite a large quantity of residual substrate being present during the full extent of the assay, the decay of the remaining substrate was not observably perturbed. Although some decay was seen, this was likely an artifact, considering that none of the other cleavage resistant substrates were cleaved. The artifact in question is most common when all substrates are present; each substrate causes an increase in background intensity for the substrate immediately above it on the gel image. Thus as the substrate containing the native cleavage motif decays into products, the background intensity of the first cleavage-resistant substrate declines as well, leading to an apparent drop in intensity over time. Regardless, such artifacts were not expected to have any effect on determination of k_{app} because the correction factors m_1 and m_2 (described more fully in § 2.6) compensate for this. Collectively, these experiments show that cleavage is non-cooperative and unaffected by substrate length.

3.4　Cleavage Site Sequencing

Although promiscuity by I-TevI has already been observed for positions 2, 3, and 4 of the cleavage motif, a position 1 C of the 5' – CAACG – 3' cleavage motif was previously determined to be necessary for efficient target site cleavage. Thus, the observation that the T3 triple mutant aptly cleaves C1T *in vivo*, and all position 1 substitutions *in vitro* was surprising. This promiscuity could be explained by cleavage of a secondary target-site that is triggered by the absence of C1. To explore this possibility, the products of cleavage reactions with T3 and each of these substrates were sequenced

using the Sanger method. These sequencing data revealed that the cleavage site has not changed for any of the position 1 substitutions (Figure 3.10). The cleavage motif is nicked on the bottom strand between positions 2 and 3 of the cleavage motif (Figure 3.10 for.), and on the top strand between positions 4 and 5 (Figure 3.10 rev.). It is important to note that the *Taq* DNA polymerase used for Sanger sequencing affixes a single adenosine to the 3' end of a nascent strand which is apparent in the readouts from the upstream primer as an additional 3' adenine, and on the readouts from the downstream primer as a 5' thymine.
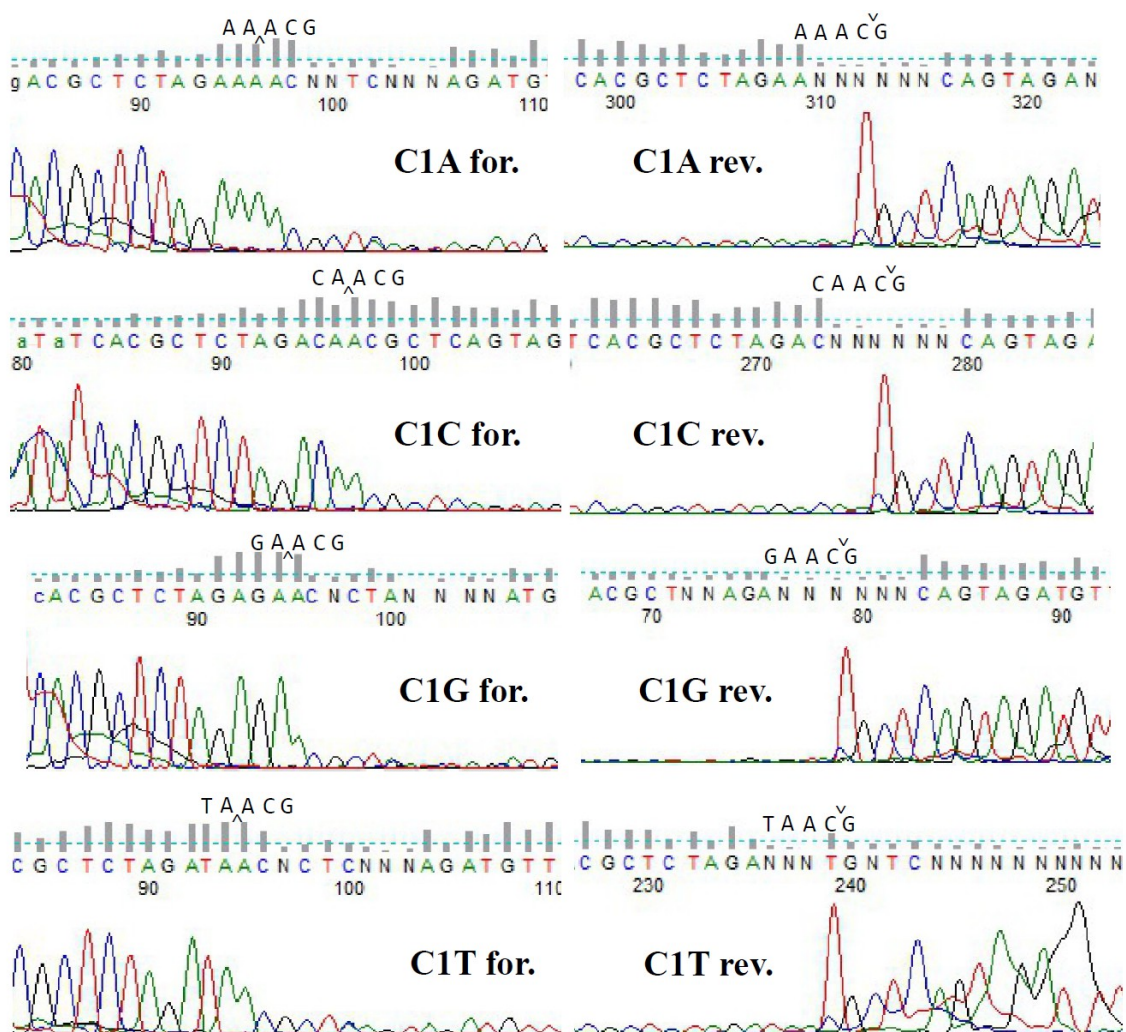
Figure 3.10. The position of substrate nicking reactions of I-TevI T3 are unaffected by subsitutions at position 1 of the cleavage motif. pTox plasmids harbouring native cleavage motif (C1C [C, D]), or one of three position 1 substitutions (C1A [A, B], C1G [E, F], C1T [G, H]) were sequenced using flanking primers: one upstream of the cleavage motif (for. [A, C, E, G]) and one downstream (rev. [B, D, F, H], the reverse complement is shown). Sanger sequencing readouts are shown with traces for adenine (green), cytosine (blue), guanine (black), and thymine (red). The cleavage motif is given above the corresponding region of the readout, with a chevron indicating the predicted nicking position. A drop-off in fluorescence intensity is seen in each sanger readout corresponding to the predicted nicking positions in all eight cases.

Chapter 4   DISCUSSION & CONCLUSIONS

Altering the specificity of existing HEs is central to any effort to extend the versatility of this genome editing platform; however, altering the specificity of HEs has proven difficult. Typically, success is achieved by separating the HE into 'modules' that retain their function when recombined with modules from other proteins. In the case of I-TevI, the ND has been shown to be such a module, and has been ported to numerous other DBDs; however, the native cleavage specificity of the I-TevI ND is limiting, and thus installing new specificities are desired. In previous cases, new specificities were developed through a relaxed specificity intermediate. Further, installing new specificities is facilitated by a knowledge of which amino acids are responsible for conveying substrate specificity and defining the cleavage profile. Thus identification of amino acids in the I-TevI ND that result in relaxed specificity is a twofold success; in addition to generating a relaxed specificity mutant, it provides indirect evidence of the amino acids that convey specificity in the wild-type ND. My goal in this thesis has been to determine if the I-TevI ND is responsible for controlling the cleavage motif cleavage profile. I did this by testing the null hypothesis that mutagenesis of the I-TevI ND would not alter the cleavage profile.

4.1   Directed Evolution of I-TevI Nuclease Domains

My first and second research objectives were the creation of a library of mutagenised I-TevI ND from which I would try to identify I-TevI ND mutants that could cleave substrates that the wild-type could not. In the preceding chapters, I described how mutagenic PCR was used to generate a library of I-TevI ND mutants, from which NDs

with altered specificities were selected over multiple rounds of directed evolution. Six populations of mutant I-TevI NDs survived 4 rounds of selection on substrates AAG, CCC, GAA, GCC, GGA, and TGG (population A-F, respectively). It was interesting that in these 6 populations, all 12 individual mutants identified possessed either a T95S or a Q158R mutation. This was the first indication that these mutations would prove important for altering the I-TevI ND cleavage profile. Another notable mutation was the K26R mutation, which is immediately adjacent to the catalytically critical R27, and was the only mutant identified in combination with both Q158R and T95S, indicating a potentially significant impact on catalysis when combined with both of these mutations. However, the absence of K26R, C39R, or I86V mutations in isolation indicates that these mutations may be less important for catalysis. Ultimately, the small sample size of isolated mutants (5 per population), and the potential for a founder effect in the PCR mutagenesis cast doubt on these assertions, and a more detailed study of these mutations was needed; regardless, my first two objectives were complete.

4.2   Individual Mutations: Potential Impacts on Catalysis and Structure

My third objective was to identify individual I-TevI ND mutations or combinations thereof that would confer survival against substrates that the wild-type ND could not cleave. My intention was to identify the ND mutations that had the strongest impact on survival, and thus were most likely to be of direct catalytic relevance. The exact mechanism by which the mutations I identified alter specificity is unknown, and was not directly attended to in this thesis; however, most of the mutations described above are oriented towards the putative active site of the I-TevI ND (Figure 4.1),

presenting a possibility for them to have a direct role in catalysis. Furthermore, some of these mutations occur adjacent to residues with an established catalytic role.

4.2.1   *Individual Mutations: K26R is Adjacent to Catalytically Critical R27*

An R27A mutation in the I-TevI ND abolishes the second, top-strand, nicking reaction and thus R27 likely has a direct role in cleavage of the top-strand phosphodiester bond. Although the precise mechanism of I-TevI is not known, the mechanism of other GIY-YIG NDs have been elucidated in greater detail. Hpy188I is a RE from *Helicobacter pylori*, and contains a GIY-YIG ND. Crystallographic studies have solved the structure of this enzyme with its substrate bound, in which R84 (analogous to R27 in I-TevI, Figure 3.5) is observed in a crystal structure orienting the water molecule that makes a nucleophilic attack on the phosphodiester bond[131]. UvrC, a component of the DNA damage repair pathway also contains a GIY-YIG ND, in which R39 (analogous to R27 in I-TevI, Figure 3.5) appears to be involved in charge balancing of the pentavalent phosphate intermediate that forms following nucleophilic attack by water[132]. Thus K26R may simply assist R27 by positioning the scissile phosphate accordingly, or it may stabilise the pentavalent phosphate intermediate that accompanies phosphodiester bond cleavage as in the mechanism of the HE I-PpoI[133], Eco29kI[134], or UvrC[132]. One intriguing possibility is that K26R is acting as a redundant catalytic residue, that steps in to catalyse cleavage of substrates for which the orientation of R27 is sub-optimal due to perturbations of DNA backbone structure that accompany alterations of DNA sequence (in this case, the cleavage motif). Such a possibility could be tested by generating a K26R R27A mutant, and testing for *in vitro* cleavage or *in vivo*
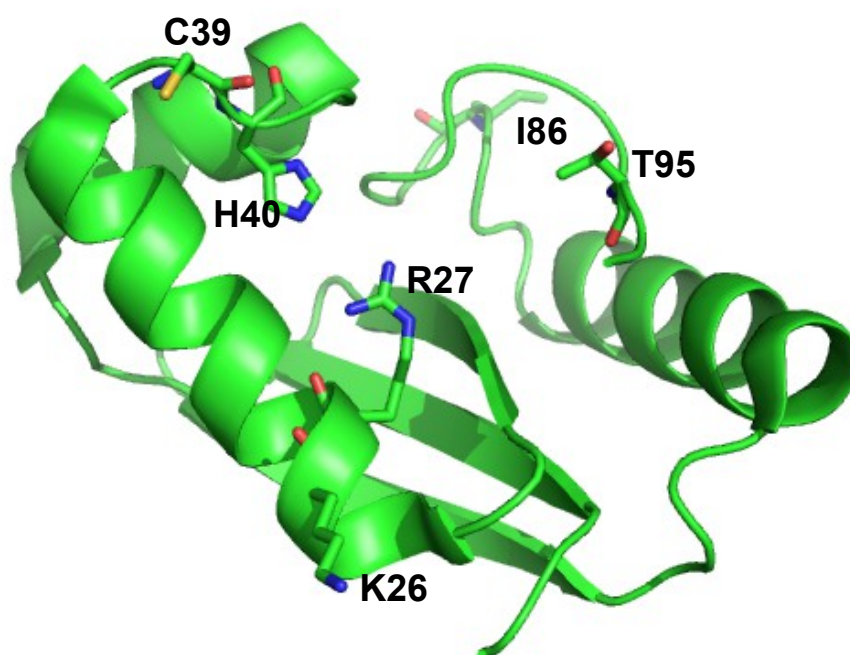
Figure 4.1. Mutations K26R, C39R, and T95S affect the I-TevI ND active site.
Key catalytic residues H40 and R27 are oriented towards the active site groove of
the I-TevI ND. K26, C39, and T95 are also oriented towards this groove, or could
adopt a conformation to do so.

survival as described in this thesis.

4.2.2   Individual Mutations: C39R is Adjacent to Catalytically Important H40

H40 has been identified as being catalytically important: H40Y is structurally stable, as evidenced by substrate bending studies, but is less able to effect catalysis. Histidine residues typically contribute to catalysis by acting as a general base, stabilising anionic intermediates *via* charge balancing or H-bonding, or by chelating metal atoms. In the proposed mechanism of the GIY-YIG ND in Eco29kI, a histidine residue (which does not coincide with H40 in sequence alignments with I-TevI) is responsible for deprotonating one of the conserved tyrosine residues of the GIY-YIG sequence, which in turn deprotonates a water molecule such that it can nucleophilically attack the scissile phosphate to effect phosphodiester bond cleavage[134]. In the proposed mechanism of Hpy188I, a histidine residue aids in coordinating the divalent metal ion responsible for orienting the phosphodiester group such that nucleophilic attack by water ejects the 3'-hydroxyl group of the downstream nucleotide[131]. Each of these roles could be modulated, enhanced, or abolished by a nearby guanidinium group, as in the C39R mutation.

It is worth noting that cysteine can act in a similar fashion to the histidine residues in the examples above; however, given that this residue is not conserved across GIY-YIG domains, and no C39A or similar mutations exist, it is currently difficult to speculate about what catalytic role – if any – this residue could have. There also exists the possibility that C39R has no direct catalytic impact at all. Consider that, as can be seen in Figure 3.4, the C39R mutation has only a weak ability to confer survival. This

weak influence on survival is in contrast to the strong influence on survival that a substitution directly involved in catalysis would be expected to have (*e.g.* catalytically dead R27A NDs confers no selective advantage over empty backbone vector, as evidenced by the <0.1% survival observed for both). These observations may point to another possible explanation. C39 is notable for being present as a cystine in the I-TevI R27A ND crystal structure (PDB ID: 1LN0)[135], forming a disulfide bond to C39 of an adjacent I-TevI ND. Although C39R has a relatively small impact on changing specificity, it may reduce the sensitivity of the ND to oxidative stress or post-translational modification, thus its role may be connected to *in vivo* stability moreso than catalysis. It may be illuminating to measure or follow the expression of the MegaTev chimaera with or without the C39R mutation using S^{35} pulse-chase to determine levels of expression in the cell and turnover.

4.2.3   *T95S: Implications for the I-TevI Nuclease Domain C-Terminal Region*

T95S contributed significantly to survival against a number of poor substrates with NNN substitutions (GAA, TGG, GCC, GGA, and CCC) but not against any of the poor substrates with position 1 substitutions (C1A, C1G, or C1T). This could indicate the role of this amino acid in I-TevI. T95 is located at the border of the ND and the linker region, a region of poorly defined structure and function, however its impact on NNN triplet recognition suggests that this region is important for defining how this triplet is recognised. Because the triplets seem to be read – in part – through indirect readout (*i.e.* through the response of the DNA to structural perturbations rather than by specific H-bonding patterns of the major-groove surface) the C-terminal region of the

ND in which T95S is found may be responsible for positioning the substrate for readout, or straining the substrate DNA to bring the catalytic residues and substrate into the proper orientation (activated state) for catalysis. Such a difference might be borne out by a thermodynamic study of the cleavage of substrates with varying NNN triplets by the I-TevI wild-type and T3 ND to explore differences in ground-state substrate binding and transition-state binding between these two NDs, and between cleavage resistant and cleavage facile substrates.

4.2.4   *Significance of Similarities of Mutations to I-BmoI Sequence*

Some of the amino acid substitutions identified by the selections described herein are already present in another GIY-YIG HE, I-BmoI, which has the native cleavage motif 5' – GCCCG – 3'. Thus a set of mutations were installed in I-TevI to emulate the sequence of I-BmoI at these positions, namely K26S, C39R, and T95S, to see if such substitutions led to marked improvement of cleavage of substrates similar to the I-BmoI cleavage motif, namely C1G (5' – GAACG – 3') or CCC (5' – CCCCG – 3'). Disappointingly, K26S and C39R did not impart any significant advantage, either alone or with other mutations, against substrates similar to the I-BmoI native cleavage motif. However, T95S was singularly responsible for improved survival against CCC, and another K26 substitution, K26R, when combined with Q158R and T95S, conferred survival against all position 1 substitutions, including C1G.

4.2.5   *The Curious Case of Q158R*

One of the mutations identified from selections was particularly surprising,

namely Q158R. This mutation was not identified in any of the regions that were intentionally mutagenised, and likely arose spontaneously during PCR amplification of the linker region during library construction. Furthermore, Q158 is within a ZF DBD that has not been thus far linked to the specificity or DNA-binding affinity of the enzyme[136], but rather to distance determination. The significant increase in survival conferred by Q158R against both cleavage resistant NNN triplets and position 1 cleavage motif substitutions cannot easily be explained by altered cleavage distance determination, especially in light of the cleavage motif sequencing results, which indicate that in fact the cleavage site has not changed.

One possible reason for the impact of the Q158R mutation on survival may be that it is acting as a suppressing mutation of sorts, and counteracting disturbances created in the linker due to the fusion of residues 1-169 of I-TevI to I-OnuI E1 E22Q. This mechanism of expanding the cleavage profile of I-TevI could be investigated by including Q158R in chimaeric MegaTevs that have more of the native linker region between the ND and I-OnuI E1 E22Q. If a MegaTev with a longer linker does not have a broader cleavage specificity after inclusion of Q158R, then it would be unlikely that this substitution suppresses the effect of fusions made only 11 aa downstream to the same extent as fusions made as many as 37 aa downstream (as would be the case for fusions with residues 1-201 of I-TevI)[6]. A more likely cause for this result in this case would be that the substitution effects its change upstream instead, perhaps by influencing the orientation of the linker, and through mechanical coupling, the orientation of the ND.

4.2.6   The Triple Mutant: K26R T95S Q158R (T3)

The most promising individual mutations identified in survival assays were K26R, T95S, and Q158R. Unsurprisingly, when these mutations were combined into a single triple mutant, the result was an I-TevI ND and linker with a significantly expanded survivability in survival assays. The creation of a triple ND mutant with expanded survivability is not consistent with the null hypothesis that the I-TevI ND is not responsible for the cleavage motif cleavage profile. However, since the survival assays do not measure cleavage *per se*, it was not known whether or not these substitutions were the result of a change in the catalytic activity and cleavage profile of the ND. Thus these results were not necessarily *inconsistent* with the null hypothesis either, and direct measurement of cleavage using *in vitro* assays was indicated.

4.3   Information Gleaned From *in vitro* Assays

My fourth research objective was to overexpress, purify and kinetically characterise an I-TevI ND mutant to determine its cleavage profile. If the cleavage profile has clearly changed in a manner consistent with the ND mutations assayed *in vivo*, then indeed the ND is at least partly responsible for defining the cleavage motif cleavage profile. As alluded to in the previous section, *in vivo* data obtained from survival assays cannot be directly correlated with activity because of numerous convoluting variables in a biological system. Indeed, tight regulation of HEs is required to prevent detrimental effects to the host organism[137,138], which are not necessarily replicated in the bacterial 2-plasmid system. Thus survival may be conveyed through reduced toxicity (*e.g.* reduced affinity/activity for some as-yet unknown site in a critical component of the *E. coli* genome), increased *in vivo* stability leading to higher steady-

state enzyme concentrations, or even through unanticipated downregulation of the topoisomerase inhibitor used as selective pressure in these assays, similar to the autoregulation of wild-type I-TevI, in which I-TevI binds and obscures a regulatory site, but cannot cut it because of an absent cleavage motif[137]. Even if these results are due to a catalytic effect, the data presented in sections 3.1 and 3.2 does not discriminate between increased activity against all substrates (since there is an upper bound of ~100% survival), or altered specificity through increased promiscuity. Furthermore, the *in vivo* data provides only a coarse estimate of activity since the standard error for such an experiment can be quite large. Thus, *in vitro* assays are indispensable for distilling the enzymatic consequences from a whole-cell system; this kind of experiment has the sensitivity to discriminate between increased activity against all substrates.

In general, it can be seen that the T3 ND cleaves the cleavage-resistant substrates more efficiently than the wild-type ND, as measured by the first order rate constant that describes the decay of each substrate over time. For some substrates assayed (*e.g.* GCA, GTG, and GCT) there was a large disparity between the *in vitro* cleavage rates between wild-type and T3 NDs, but relatively a relatively small enhancement of *in vivo* survival rate by the T3 ND. Conversely, assay of some substrates (*e.g.* CCC, and AAG) revealed a comparatively small disparity between *in vitro* cleavage rates, despite a large increase in survival. Although this initially seemed to point to a conflict between the two data sets, further examination revealed that the data is quite consistent with a threshold effect: survival *in vivo* corresponded to a relative k_{app} value greater than or approximately equal to 0.5.

While this correlation is somewhat crude and exceptions exist, it could be still be

used as a rough measure of the extent of convoluting variables affecting *in vivo* data: if the *in vitro* data suggest that there clearly should or should not be survival for a chosen enzyme-substrate pairing, then conflicting *in vivo* data would indicate that variables other than cleavage rate may need to be considered. That being said, further characterisation to ensure this threshold effect is observed with other substrates *in vitro* and *in vivo* would be necessary to give confidence to such a metric.

The NNN triplets presented in Figure 3.8 represent only half of all possible NNN triplets. Thus one obvious step forward will be to determine the exact kinetic impact of each of these mutations on cleavage of all 64 possible NNNs. Another consideration that was not attended to in this thesis is the impact that these mutations may have on cleavage of position 5 substitutions. Although the position 5 G of the I-TevI cleavage motif, like the position 1 C has been previously demonstrated to be required for cleavage, the mutations I've described above may provide the ability to relax this requirement.

4.4   Future Directions

Although the results of the work described in this thesis are inherently informative, they also form the foundation of the project that I will be undertaking towards completion of my PhD thesis: a thermodynamic and kinetic investigation of the underlying cause of the I-TevI cleavage profile. As mentioned in the introduction to this thesis, there is no obvious pattern to the NNN triplets that I-TevI will cleave, and indirect readout through the biophysical characteristics of the triplets is likely responsible for the cryptic cleavage profile. By identifying I-TevI ND mutants that

present an altered cleavage profile, I can begin identifying correlations between the specific thermodynamic and kinetic characteristics of the mutants, and the ability to cleave a particular NNN triplet. In the same spirit as Prof. Richard Feynman's famous words, "If I can build it, then I understand it" my hope is that these correlations can be used to reverse engineer I-TevI NDs with practically any desired cleavage profile. Below are briefly described some of the projects I will be pursuing in my PhD, which are derived from the results in this thesis.

4.4.1   *Additional Directed Evolution of I-TevI Nuclease Domains*

Although there was extensive redundancy in the mutations identified from my genetic selections, a more thorough exploration of the surviving I-TevI ND mutants may identify additional mutations that expand the I-TevI ND cleavage profile, diminish the number of cleavage resistant substrates, and complement experiments designed to elucidate the mechanism by which ND mutations expand the I-TevI cleavage profile.

In engineering an enzyme with a new substrate specificity, it can be useful to first develop a more promiscuous enzyme capable of acting on both its original substrate and the new substrate. This promiscuous enzyme can then be refined to act selectively on the new substrate. I propose that the same approach could be applied to the I-TevI ND. The T3 mutant has a more relaxed cleavage profile compared to wild-type, for the set of substrates assayed thus far. An I-TevI ND mutant with an orthogonal cleavage profile might be developed by further mutagensing the T3 mutant, and then conducting rounds of selection in which cleavage of a chosen substrate is selected against, by including the corresponding target site into the pEndo vector, which harbours the I-TevI

ND mutant itself. Thus only mutants that cleave the target site in pTox, but *not* in pEndo will be maintained in the population. Mutations identified by such a bi-functional selection could not only inform on those amino acids that facilitate cleavage of cleavage resistant substrates, but also reveal those amino acids that must be maintained in order to facilitate cleavage of the target site included on pEndo.

4.4.2 *Kinetic Investigations of the I-TevI Nuclease Domain Cleavage Profile*

I-TevI has a two step mechanism in which the bottom strand is nicked prior to cleavage. Studies with I-BmoI have indicated that the rate of each nicking reaction varies depending on the chemical environment of the ND active site, namely the divalent metal ion present[139]. I posit that the resistance of each NNN triplet to cleavage is mediated by perturbing the active site chemical environment, and that this will manifest as a reduction of the rate of one or both nicking reactions. Moreover, I propose that the extent to which each nicking reaction is compromised may be correlated to the nucleobases on the same strand as, and directly adjacent to, the scissile phosphate. This information is hidden in measurements of overall cleavage (such as the barcode assay) because overall cleavage would proceed at a rate defined only by the nicking reaction that becomes the rate-limiting step. I further propose that the ND mutants that I have identified will display an increase in the rate of one or both nicking reactions. Collectively, I expect that the substrates that a particular mutant cleaves better than wild-type will be those substrates that compromise the same nicking reaction that the mutation accelerates. In other words: the cleavage-resistant substrates impose a rate-limiting step that the mutations counterbalance, such that on the whole, the reaction

proceeds efficiently.

4.4.3   Thermodynamic Investigations of the I-TevI Nuclease Domain Cleavage Profile

Even in the absence of direct observation, there is a great deal about the mechanism of an enzyme that can be elucidated by examining the thermodynamics of its function. The extent to which a particular kinetic constant varies with temperature can be used to determine thermodynamic constants for that reaction, such as the individual contributions of enthalpy and entropy changes to transition state stabilisation and ground state destabilisation[140,141]. An enzyme that contorts its substrate might be expected to strongly reduce the entropy of bound substrate, as it is gripped tightly and forced into an unfavourable conformation. Such contortions would have to be compensated for by a similarly strong reduction of enthalpy, typically effected using extensive hydrogen bonding, close packing of hydrophobic surfaces, and geometrically optimal salt bridges. As discussed in the introduction, indirect readout seemingly plays a part in the I-TevI ND cleavage profile. Since indirect readout is in essence recognising the response of a segment of DNA to strain, I propose that cleavage of the native I-TevI cleavage motif will be accompanied by a large decrease in entropy. It stands to reason that indirect readout cannot function properly if the required strain cannot be generated. Thus, I propose that cleavage resistant substrates are as such because they are also resistant to the contortions required for indirect readout, either because they are too stiff to be contorted to a significant degree, or they are flexible and are able to contort their structure without induction of significant strain. In either case, I expect that cleavage resistance will manifest as a complete or partial mitigation of the large reduction of

entropy anticipated for binding of the native target. Consequently, I propose that the I-TevI ND mutants that effect cleavage of cleavage resistant substrates will display an enhanced reduction of entropy upon substrate binding regardless of the substrate being cleaved.

4.5   Conclusions

Efforts to re-engineer nucleases have made significant progress over the past decade. However, these efforts have also proven challenging, and the goal of a fully customiseable nuclease is still incomplete. This work represents the first time the cleavage specificity towards the I-TevI cleavage motif has been altered. Furthermore, since the I-TevI ND and partial linker are portable to other DBDs, this result is a step towards improving the versatility of a genome editing system in which a DBD and an I-TevI ND mutant are combined on the basis of their specificity to effect genome editing at any chosen locus. Additional work will be required to further alter and hone the specificity of these mutants, using both positive selection as described above, and an additional negative selection, which eliminates ND mutants that cleave a particular cleavage motif by – for instance – placing that cleavage motif in the ND mutant expression plasmid. Perhaps more importantly, the mechanism by which the mutants identified by genetic selection change the I-TevI cleavage profile is poorly understood and requires further study.

Engineering successes and newly opened avenues of research aside, the question remains: were the results of this thesis consistent with the null hypothesis or no? I observed that the ND mutations K26R and T95S were able to expand survivability *in*

vivo, especially when combined with Q158R. Importantly, Q158R was not able to confer survival against some NNNs such as TGG without the assistance of K26R, and especially T95S. However, it could be argued that the inclusion of Q158R undermines the premise of the experiment. Q158R is not a ND mutation, and its influence on the cleavage motif cleavage profile convolutes the influence of the true ND mutations. Yet, the correlation between *in vitro* cleavage and *in vivo* survival strongly suggests that pronounced survival against a substrate was the result of substrate cleavage. *The single mutant T95S was able to confer pronounced survival against an expanded set of substrates, which in light of the correlation between survival and cleavage, indicates that it did indeed relax the cleavage motif cleavage profile, in direct conflict with the null hypothesis.*

I think it is worth noting, however, that whatever contribution the T95S mutation had on relaxing the I-TevI cleavage profile, it was greatly augmented by the addition of Q158R. Further, T95S is at the extreme limit of the canonically defined ND of I-TevI, and those mutations that were clearly within the ND proved far less capable of conferring expanded survivability *in vivo*. Thus I propose a new hypothesis: *The cleavage motif cleavage profile is defined in part by residues within the I-TevI ND, and in part by residues of the linker region, and that these residues work cooperatively.*

REFERENCES

1. Belfort M., Bonocora R. P. (2014) Homing endonucleases: from genetic anomalies to programmable genomic clippers. *Methods Mol. Biol.* **1123,** 1-26. 10.1007/978-1-62703-968-0_1 [doi].

2. Gaj T., Sirk S. J. & Barbas C. F.,3rd. (2014) Expanding the scope of site-specific recombinases for genetic and metabolic engineering. *Biotechnol. Bioeng.* **111,** 1-15. 10.1002/bit.25096 [doi].

3. [Anonymous]. (2012) Method of the Year 2011. *Nat Meth* **9,** 1-1.

4. Takeuchi R., Lambert A. R., Mak A. N., Jacoby K., Dickson R. J., Gloor G. B., Scharenberg A. M., Edgell D. R. & Stoddard B. L. (2011) Tapping natural reservoirs of homing endonucleases for targeted gene modification. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 13077-13082. 10.1073/pnas.1107719108 [doi].

5. Derbyshire V., Kowalski J. C., Dansereau J. T., Hauer C. R. & Belfort M. (1997) Two-domain structure of the td intron-encoded endonuclease I-TevI correlates with the two-domain configuration of the homing site. *J. Mol. Biol.* **265,** 494-506. S0022-2836(96)90754-8 [pii].

6. Kleinstiver B. P., Wolfs J. M., Kolaczyk T., Roberts A. K., Hu S. X. & Edgell D. R. (2012) Monomeric site-specific nucleases for genome editing. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 8061-8066. 10.1073/pnas.1117984109 [doi].

7. Sander J. D., Joung J. K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32,** 347-355. 10.1038/nbt.2842 [doi].

8. Walrath J. C., Hawes J. J., Van Dyke T. & Reilly K. M. (2010) Genetically engineered mouse models in cancer research. *Adv. Cancer Res.* **106,** 113-164. 10.1016/S0065-230X(10)06004-5 [doi].

9. Doyle A., McGarry M. P., Lee N. A. & Lee J. J. (2012) The construction of transgenic and gene knockout/knockin mouse models of human disease.

Transgenic Res. **21,** 327-349. 10.1007/s11248-011-9537-3 [doi].

10. Wood D. W., Camarero J. A. (2014) Intein applications: from protein purification and labeling to metabolic control methods. *J. Biol. Chem.* **289,** 14512-14519. 10.1074/jbc.R114.552653 [doi].

11. Young C. L., Britton Z. T. & Robinson A. S. (2012) Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. *Biotechnol. J.* **7,** 620-634. 10.1002/biot.201100155 [doi].

12. Gao H., Smith J., Yang M., Jones S., Djukanovic V., Nicholson M. G., West A., Bidney D., Falco S. C., Jantz D. & Lyznik L. A. (2010) Heritable targeted mutagenesis in maize using a designed endonuclease. *Plant J.* **61,** 176-187. 10.1111/j.1365-313X.2009.04041.x [doi].

13. Zeevi V., Liang Z., Arieli U. & Tzfira T. (2012) Zinc finger nuclease and homing endonuclease-mediated assembly of multigene plant transformation vectors. *Plant Physiol.* **158,** 132-144. 10.1104/pp.111.184374 [doi].

14. Vainstein A., Marton I., Zuker A., Danziger M. & Tzfira T. (2011) Permanent genome modifications in plant cells by transient viral vectors. *Trends Biotechnol.* **29,** 363-369. 10.1016/j.tibtech.2011.03.007 [doi].

15. Stryjewska A., Kiepura K., Librowski T. & Lochynski S. (2013) Biotechnology and genetic engineering in the new drug development. Part I. DNA technology and recombinant proteins. *Pharmacol. Rep.* **65,** 1075-1085.

16. Stephenson P. G., Moore C. M., Terry M. J., Zubkov M. V. & Bibby T. S. (2011) Improving photosynthesis for algal biofuels: toward a green revolution. *Trends Biotechnol.* **29,** 615-623. 10.1016/j.tibtech.2011.06.005 [doi].

17. Hemaiswarya S., Raja R., Carvalho I. S., Ravikumar R., Zambare V. & Barh D. (2012) An Indian scenario on renewable and sustainable energy sources with emphasis on algae. *Appl. Microbiol. Biotechnol.* **96,** 1125-1135. 10.1007/s00253-012-4487-0 [doi].

18. Bonnet J., Subsoontorn P. & Endy D. (2012) Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 8884-8889. 10.1073/pnas.1202344109 [doi].

19. Siuti P., Yazbek J. & Lu T. K. (2013) Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.* **31,** 448-452. 10.1038/nbt.2510 [doi].

20. Bonnet J., Yin P., Ortiz M. E., Subsoontorn P. & Endy D. (2013) Amplifying genetic logic gates. *Science* **340,** 599-603. 10.1126/science.1232758 [doi].

21. Chan Y. S., Naujoks D. A., Huen D. S. & Russell S. (2011) Insect population control by homing endonuclease-based gene drive: an evaluation in Drosophila melanogaster. *Genetics* **188,** 33-44. 10.1534/genetics.111.127506 [doi].

22. Windbichler N., Menichelli M., Papathanos P. A., Thyme S. B., Li H., Ulge U. Y., Hovde B. T., Baker D., Monnat R. J.,Jr, Burt A. & Crisanti A. (2011) A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature* **473,** 212-215. 10.1038/nature09937 [doi].

23. Klein T. A., Windbichler N., Deredec A., Burt A. & Benedict M. Q. (2012) Infertility resulting from transgenic I-PpoI male Anopheles gambiae in large cage trials. *Pathog. Glob. Health.* **106,** 20-31. 10.1179/2047773212Y.0000000003 [doi].

24. Deredec A., Godfray H. C. & Burt A. (2011) Requirements for effective malaria control with homing endonuclease genes. *Proc. Natl. Acad. Sci. U. S. A.* **108,** E874-80. 10.1073/pnas.1110717108 [doi].

25. Marcaida M. J., Munoz I. G., Blanco F. J., Prieto J. & Montoya G. (2010) Homing endonucleases: from basics to therapeutic applications. *Cell Mol. Life Sci.* **67,** 727-748. 10.1007/s00018-009-0188-y [doi].

26. Kohn D. B., Hershfield M. S., Carbonaro D., Shigeoka A., Brooks J., Smogorzewska E. M., Barsky L. W., Chan R., Burotto F., Annett G., Nolta J. A., Crooks G., Kapoor N., Elder M., Wara D., Bowen T., Madsen E., Snyder F. F., Bastian J., Muul L., Blaese R. M., Weinberg K. & Parkman R. (1998) T lymphocytes with a normal ADA gene accumulate after transplantation of transduced autologous umbilical cord blood CD34+ cells in ADA-deficient SCID neonates. *Nat. Med.* **4,** 775-780.

27. Pessach I. M., Notarangelo L. D. (2011) Gene therapy for primary immunodeficiencies: looking ahead, toward gene correction. *J. Allergy Clin. Immunol.* **127,** 1344-1350. 10.1016/j.jaci.2011.02.027 [doi].

28. Olivares E. C., Hollis R. P., Chalberg T. W., Meuse L., Kay M. A. & Calos M. P. (2002) Site-specific genomic integration produces therapeutic Factor IX levels in mice. *Nat. Biotechnol.* **20,** 1124-1128. 10.1038/nbt753 [doi].

29. Bertoni C., Jarrahian S., Wheeler T. M., Li Y., Olivares E. C., Calos M. P. & Rando T. A. (2006) Enhancement of plasmid-mediated gene therapy for muscular dystrophy by directed plasmid integration. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 419-424. 0504505102 [pii].

30. Ortiz-Urda S., Thyagarajan B., Keene D. R., Lin Q., Calos M. P. & Khavari P. A. (2003) PhiC31 integrase-mediated nonviral genetic correction of junctional epidermolysis bullosa. *Hum. Gene Ther.* **14,** 923-928. 10.1089/104303403765701204 [doi].

31. Portlock J. L., Keravala A., Bertoni C., Lee S., Rando T. A. & Calos M. P. (2006) Long-term increase in mVEGF164 in mouse hindlimb muscle mediated by phage phiC31 integrase after nonviral DNA delivery. *Hum. Gene Ther.* **17,** 871-876. 10.1089/hum.2006.17.871 [doi].

32. Keravala A., Portlock J. L., Nash J. A., Vitrant D. G., Robbins P. D. & Calos M. P. (2006) PhiC31 integrase mediates integration in cultured synovial cells and enhances gene expression in rabbit joints. *J. Gene Med.* **8,** 1008-1017. 10.1002/jgm.928 [doi].

33. Ortiz-Urda S., Thyagarajan B., Keene D. R., Lin Q., Fang M., Calos M. P. & Khavari P. A. (2002) Stable nonviral genetic correction of inherited human skin disease. *Nat. Med.* **8,** 1166-1170. 10.1038/nm766 [doi].

34. Thyagarajan B., Liu Y., Shin S., Lakshmipathy U., Scheyhing K., Xue H., Ellerstrom C., Strehl R., Hyllner J., Rao M. S. & Chesnut J. D. (2008) Creation of engineered human embryonic stem cell lines using phiC31 integrase. *Stem Cells* **26,** 119-126. 2007-0283 [pii].

35. Ye L., Chang J. C., Lin C., Qi Z., Yu J. & Kan Y. W. (2010) Generation of induced pluripotent stem cells using site-specific integration with phage integrase. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 19467-19472. 10.1073/pnas.1012677107 [doi].

36. Certo M. T., Gwiazda K. S., Kuhar R., Sather B., Curinga G., Mandt T., Brault M., Lambert A. R., Baxter S. K., Jacoby K., Ryu B. Y., Kiem H. P., Gouble A., Paques F., Rawlings D. J. & Scharenberg A. M. (2012) Coupling endonucleases with DNA end-processing enzymes to drive gene disruption. *Nat. Methods* **9,** 973-975. 10.1038/nmeth.2177 [doi].

37. Cannon P., June C. (2011) Chemokine receptor 5 knockout strategies. *Curr. Opin. HIV. AIDS.* **6,** 74-79. 10.1097/COH.0b013e32834122d7 [doi].

38. Aubert M., Ryu B. Y., Banks L., Rawlings D. J., Scharenberg A. M. & Jerome K. R. (2011) Successful targeting and disruption of an integrated reporter lentivirus using the engineered homing endonuclease Y2 I-AniI. *PLoS One* **6,** e16825. 10.1371/journal.pone.0016825 [doi].

39. Arnould S., Perez C., Cabaniols J. P., Smith J., Gouble A., Grizot S., Epinat J. C., Duclert A., Duchateau P. & Paques F. (2007) Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *J. Mol. Biol.* **371,** 49-65. S0022-2836(07)00579-7 [pii].

40. Redondo P., Prieto J., Munoz I. G., Alibes A., Stricher F., Serrano L., Cabaniols J. P., Daboussi F., Arnould S., Perez C., Duchateau P., Paques F., Blanco F. J. & Montoya G. (2008) Molecular basis of xeroderma pigmentosum group C DNA recognition by engineered meganucleases. *Nature* **456,** 107-111. 10.1038/nature07343 [doi].

41. Grizot S., Smith J., Daboussi F., Prieto J., Redondo P., Merino N., Villate M., Thomas S., Lemaire L., Montoya G., Blanco F. J., Paques F. & Duchateau P. (2009) Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. *Nucleic Acids Res.* **37,** 5405-5419. 10.1093/nar/gkp548 [doi].

42. Munoz I. G., Prieto J., Subramanian S., Coloma J., Redondo P., Villate M., Merino N., Marenchino M., D'Abramo M., Gervasio F. L., Grizot S., Daboussi F., Smith J., Chion-Sotinel I., Paques F., Duchateau P., Alibes A., Stricher F., Serrano L., Blanco F. J. & Montoya G. (2011) Molecular basis of engineered meganuclease targeting of the endogenous human RAG1 locus. *Nucleic Acids Res.* **39,** 729-743. 10.1093/nar/gkq801 [doi].

43. Chapdelaine P., Pichavant C., Rousseau J., Paques F. & Tremblay J. P. (2010) Meganucleases can restore the reading frame of a mutated dystrophin. *Gene Ther.* **17,** 846-858. 10.1038/gt.2010.26 [doi].

44. Silva G., Poirot L., Galetto R., Smith J., Montoya G., Duchateau P. & Paques F. (2011) Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Curr. Gene Ther.* **11,** 11-27. ABS- 58 [pii].

45. Lovric J., Mano M., Zentilin L., Eulalio A., Zacchigna S. & Giacca M. (2012) Terminal differentiation of cardiac and skeletal myocytes induces permissivity to AAV transduction by relieving inhibition imposed by DNA damage response proteins. *Mol. Ther.* **20,** 2087-2097. 10.1038/mt.2012.144 [doi].

46. Iyama T., Wilson D. M.,3rd. (2013) DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair (Amst)* **12,** 620-636. 10.1016/j.dnarep.2013.04.015 [doi].

47. Chalberg T. W., Portlock J. L., Olivares E. C., Thyagarajan B., Kirby P. J., Hillman R. T., Hoelters J. & Calos M. P. (2006) Integration specificity of phage phiC31 integrase in the human genome. *J. Mol. Biol.* **357,** 28-48. S0022-2836(05)01586-X [pii].

48. Thyagarajan B., Guimaraes M. J., Groth A. C. & Calos M. P. (2000) Mammalian genomes contain active recombinase recognition sites. *Gene* **244,** 47-54. S0378-1119(00)00008-1 [pii].

49. Ehrhardt A., Engler J. A., Xu H., Cherry A. M. & Kay M. A. (2006) Molecular analysis of chromosomal rearrangements in mammalian cells after phiC31-mediated integration. *Hum. Gene Ther.* **17,** 1077-1094. 10.1089/hum.2006.17.1077 [doi].

50. Liu J., Jeppesen I., Nielsen K. & Jensen T. G. (2006) Phi c31 integrase induces chromosomal aberrations in primary human fibroblasts. *Gene Ther.* **13,** 1188-1190. 3302789 [pii].

51. Loonstra A., Vooijs M., Beverloo H. B., Allak B. A., van Drunen E., Kanaar R., Berns A. & Jonkers J. (2001) Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 9209-9214. 10.1073/pnas.161269798 [doi].

52. Schmidt E. E., Taylor D. S., Prigge J. R., Barnett S. & Capecchi M. R. (2000) Illegitimate Cre-dependent chromosome rearrangements in transgenic mouse spermatids. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 13702-13707. 10.1073/pnas.240471297 [doi].

53. Gaj T., Mercer A. C., Sirk S. J., Smith H. L. & Barbas C. F.,3rd. (2013) A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. *Nucleic Acids Res.* **41,** 3937-3946. 10.1093/nar/gkt071 [doi].

54. Christian M., Cermak T., Doyle E. L., Schmidt C., Zhang F., Hummel A., Bogdanove A. J. & Voytas D. F. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186,** 757-761. 10.1534/genetics.110.120717 [doi].

55. Li T., Huang S., Jiang W. Z., Wright D., Spalding M. H., Weeks D. P. & Yang B. (2011) TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.* **39,** 359-372. 10.1093/nar/gkq704 [doi].

56. Marcaida M. J., Munoz I. G., Blanco F. J., Prieto J. & Montoya G. (2010) Homing endonucleases: from basics to therapeutic applications. *Cell Mol. Life Sci.* **67,** 727-748. 10.1007/s00018-009-0188-y [doi].

57. Stoddard B. L. (2011) Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. *Structure* **19,** 7-15. 10.1016/j.str.2010.12.003 [doi].

58. Urnov F. D., Rebar E. J., Holmes M. C., Zhang H. S. & Gregory P. D. (2010) Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* **11,** 636-646. 10.1038/nrg2842 [doi].

59. Grindley N. D., Whiteson K. L. & Rice P. A. (2006) Mechanisms of site-specific recombination. *Annu. Rev. Biochem.* **75,** 567-605. 10.1146/annurev.biochem.73.011303.073908 [doi].

60. Kuhn R., Schwenk F., Aguet M. & Rajewsky K. (1995) Inducible gene targeting in mice. *Science* **269,** 1427-1429.

61. Feil R., Brocard J., Mascrez B., LeMeur M., Metzger D. & Chambon P. (1996) Ligand-activated site-specific recombination in mice. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 10887-10890.

62. Logie C., Stewart A. F. (1995) Ligand-regulated site-specific recombination. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 5940-5944.

63. Fukushige S., Sauer B. (1992) Genomic targeting with a positive-selection lox integration vector allows highly reproducible gene expression in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 7905-7909.

64. O'Gorman S., Fox D. T. & Wahl G. M. (1991) Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science* **251,** 1351-1355.

65. Sauer B., Henderson N. (1990) Targeted insertion of exogenous DNA into the eukaryotic genome by the Cre recombinase. *New Biol.* **2,** 441-449.

66. Dale E. C., Ow D. W. (1991) Gene transfer with subsequent removal of the selection gene from the host genome. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 10558-10562.

67. Roberts R. J. (2005) How restriction enzymes became the workhorses of molecular biology. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 5905-5908. 0500923102 [pii].

68. Pingoud A., Wilson G. G. & Wende W. (2014) Type II restriction endonucleases--a historical perspective and more. *Nucleic Acids Res.* **42,** 7489-7527. 10.1093/nar/gku447 [doi].

69. Taylor J. D., Halford S. E. (1989) Discrimination between DNA sequences by the EcoRV restriction endonuclease. *Biochemistry* **28,** 6198-6207.

70. Alves J., Selent U. & Wolfes H. (1995) Accuracy of the EcoRV restriction endonuclease: binding and cleavage studies with oligodeoxynucleotide substrates containing degenerate recognition sequences. *Biochemistry* **34,** 11191-11197.

71. Alves J., Ruter T., Geiger R., Fliess A., Maass G. & Pingoud A. (1989) Changing the hydrogen-bonding potential in the DNA binding site of EcoRI by site-directed mutagenesis drastically reduces the enzymatic activity, not, however, the preference of this restriction endonuclease for cleavage within the site-GAATTC-. *Biochemistry* **28,** 2678-2684.

72. Heitman J., Model P. (1990) Mutants of the EcoRI endonuclease with promiscuous substrate specificity implicate residues involved in substrate recognition. *Embo j.* **9,** 3369-3378.

73. Wenz C., Selent U., Wende W., Jeltsch A., Wolfes H. & Pingoud A. (1994) Protein engineering of the restriction endonuclease EcoRV: replacement of an amino acid residue in the DNA binding site leads to an altered selectivity towards unmodified and modified substrates. *Biochim. Biophys. Acta* **1219,** 73-80. 0167-4781(94)90248-8 [pii].

74. Thomas M., Davis R. W. (1975) Studies on the cleavage of bacteriophage lambda DNA with EcoRI Restriction endonuclease. *J. Mol. Biol.* **91,** 315-328.

75. Forsblom S., Rigler R., Ehrenberg M. & Philipson L. (1976) Kinetic studies on the cleavage of adenovirus DNA by restriction endonuclease Eco RI. *Nucleic Acids Res.* **3,** 3255-3269.

76. Alves J., Pingoud A., Haupt W., Langowski J., Peters F., Maass G. & Wolff C. (1984) The influence of sequences adjacent to the recognition site on the cleavage of oligodeoxynucleotides by the EcoRI endonuclease. *Eur. J. Biochem.* **140,** 83-92.

77. Van Cleve M. D., Gumport R. I. (1992) Influence of enzyme-substrate contacts located outside the EcoRI recognition site on cleavage of duplex oligodeoxyribonucleotide substrates by EcoRI endonuclease. *Biochemistry* **31,** 334-339.

78. Lukacs C. M., Kucera R., Schildkraut I. & Aggarwal A. K. (2000) Understanding the immutability of restriction enzymes: crystal structure of BglII and its DNA substrate at 1.5 A resolution. *Nat. Struct. Biol.* **7,** 134-140. 10.1038/72405 [doi].

79. Morgan R. D., Luyten Y. A. (2009) Rational engineering of type II restriction endonuclease DNA binding and cleavage specificity. *Nucleic Acids Res.* **37,** 5222-5233. 10.1093/nar/gkp535 [doi].

80. Ishino Y., Shinagawa H., Makino K., Amemura M. & Nakata A. (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J. Bacteriol.* **169,** 5429-5433.

81. Bolotin A., Quinquis B., Sorokin A. & Ehrlich S. D. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151,** 2551-2561. 151/8/2551 [pii].

82. Mojica F. J., Diez-Villasenor C., Garcia-Martinez J. & Soria E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60,** 174-182. 10.1007/s00239-004-0046-3 [doi].

83. Pourcel C., Salvignol G. & Vergnaud G. (2005) CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151,** 653-663. 151/3/653 [pii].

84. Barrangou R., Marraffini L. A. (2014) CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Mol. Cell* **54,** 234-244. 10.1016/j.molcel.2014.03.011 [doi].

85. van der Oost J., Westra E. R., Jackson R. N. & Wiedenheft B. (2014) Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **12,** 479-492. 10.1038/nrmicro3279 [doi].

86. Sapranauskas R., Gasiunas G., Fremaux C., Barrangou R., Horvath P. & Siksnys V. (2011) The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Res.* **39,** 9275-9282. 10.1093/nar/gkr606 [doi].

87. Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J. A. & Charpentier E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816-821. 10.1126/science.1225829 [doi].

88. Sternberg S. H., Doudna J. A. (2015) Expanding the Biologist's Toolkit with CRISPR-Cas9. *Mol. Cell* **58,** 568-574. 10.1016/j.molcel.2015.02.032 [doi].

89. Li L., Wu L. P. & Chandrasegaran S. (1992) Functional domains in Fok I restriction endonuclease. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 4275-4279.

90. Wah D. A., Hirsch J. A., Dorner L. F., Schildkraut I. & Aggarwal A. K. (1997) Structure of the multimodular endonuclease FokI bound to DNA. *Nature* **388,** 97-100. 10.1038/40446 [doi].

91. Kim Y. G., Cha J. & Chandrasegaran S. (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 1156-1160.

92. Wolfe S. A., Nekludova L. & Pabo C. O. (2000) DNA recognition by Cys2His2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.* **29,** 183-212. 29/1/183 [pii].

93. Segal D. J., Beerli R. R., Blancafort P., Dreier B., Effertz K., Huber A., Koksch B., Lund C. V., Magnenat L., Valente D. & Barbas C. F.,3rd. (2003) Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry* **42,** 2137-2148. 10.1021/bi026806o [doi].

94. Joung J. K., Sander J. D. (2013) TALENs: a widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.* **14,** 49-55. 10.1038/nrm3486 [doi].

95. Boch J., Bonas U. (2010) Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu. Rev. Phytopathol.* **48,** 419-436. 10.1146/annurev-phyto-080508-081936 [doi].

96. Streubel J., Blucher C., Landgraf A. & Boch J. (2012) TAL effector RVD specificities and efficiencies. *Nat. Biotechnol.* **30,** 593-595. 10.1038/nbt.2304 [doi].

97. Netter P., Petrochilo E., Slonimski P. P., Bolotin-Fukuhara M., Coen D., Deutsch J. & Dujon B. (1974) Mitochondrial genetics. VII. Allelism and mapping studies of ribosomal mutants resistant to chloramphenicol, erythromycin and spiramycin in S. cerevisiae. *Genetics* **78,** 1063-1100.

98. Lambowitz A. M., Belfort M. (1993) Introns as mobile genetic elements. *Annu. Rev. Biochem.* **62,** 587-622. 10.1146/annurev.bi.62.070193.003103 [doi].

99. Stoddard B. L. (2011) Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. *Structure* **19,** 7-15. 10.1016/j.str.2010.12.003 [doi].

100. Dalgaard J. Z., Garrett R. A. & Belfort M. (1993) A site-specific endonuclease encoded by a typical archaeal intron. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 5414-5417.

101. Kane P. M., Yamashiro C. T., Wolczyk D. F., Neff N., Goebl M. & Stevens T. H.

(1990) Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science* **250,** 651-657.

102. Hirata R., Ohsumk Y., Nakano A., Kawasaki H., Suzuki K. & Anraku Y. (1990) Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of Saccharomyces cerevisiae. *J. Biol. Chem.* **265,** 6726-6733.

103. Dalgaard J. Z., Klar A. J., Moser M. J., Holley W. R., Chatterjee A. & Mian I. S. (1997) Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. *Nucleic Acids Res.* **25,** 4626-4638. gka746 [pii].

104. Sharma M., Ellis R. L. & Hinton D. M. (1992) Identification of a family of bacteriophage T4 genes encoding proteins similar to those present in group I introns of fungi and phage. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 6658-6662.

105. Swithers K. S., Senejani A. G., Fournier G. P. & Gogarten J. P. (2009) Conservation of intron and intein insertion sites: implications for life histories of parasitic genetic elements. *BMC Evol. Biol.* **9,** 303-2148-9-303. 10.1186/1471-2148-9-303 [doi].

106. Colleaux L., D'Auriol L., Galibert F. & Dujon B. (1988) Recognition and cleavage site of the intron-encoded omega transposase. *Proc. Natl. Acad. Sci. U. S. A.* **85,** 6022-6026.

107. Bryk M., Quirk S. M., Mueller J. E., Loizos N., Lawrence C. & Belfort M. (1993) The td intron endonuclease I-TevI makes extensive sequence-tolerant contacts across the minor groove of its DNA target. *Embo j.* **12,** 2141-2149.

108. Jacquier A., Dujon B. (1985) An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell* **41,** 383-394. S0092-8674(85)80011-8 [pii].

109. Jurica M. S., Monnat R. J.,Jr & Stoddard B. L. (1998) DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-CreI. *Mol. Cell* **2,** 469-476. S1097-2765(00)80146-X [pii].

110. Silva G. H., Dalgaard J. Z., Belfort M. & Van Roey P. (1999) Crystal structure of the thermostable archaeal intron-encoded endonuclease I-DmoI. *J. Mol. Biol.* **286,** 1123-1136. S0022-2836(98)92519-0 [pii].

111. Chevalier B. S., Monnat R. J.,Jr & Stoddard B. L. (2001) The homing endonuclease I-CreI uses three metals, one of which is shared between the two active sites. *Nat. Struct. Biol.* **8,** 312-316. 10.1038/86181 [doi].

112. Chevalier B., Sussman D., Otis C., Noel A. J., Turmel M., Lemieux C., Stephens K., Monnat R. J.,Jr & Stoddard B. L. (2004) Metal-dependent DNA cleavage mechanism of the I-CreI LAGLIDADG homing endonuclease. *Biochemistry* **43,** 14015-14026. 10.1021/bi048970c [doi].

113. Wolfs J. M., DaSilva M., Meister S. E., Wang X., Schild-Poulter C. & Edgell D. R. (2014) MegaTevs: single-chain dual nucleases for efficient gene disruption. *Nucleic Acids Res.* **42,** 8816-8829. 10.1093/nar/gku573; 10.1093/nar/gku573.

114. Chevalier B. S., Kortemme T., Chadsey M. S., Baker D., Monnat R. J. & Stoddard B. L. (2002) Design, activity, and structure of a highly specific artificial endonuclease. *Mol. Cell* **10,** 895-905. S1097276502006901 [pii].

115. Epinat J. C., Arnould S., Chames P., Rochaix P., Desfontaines D., Puzin C., Patin A., Zanghellini A., Paques F. & Lacroix E. (2003) A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res.* **31,** 2952-2962.

116. Baxter S., Lambert A. R., Kuhar R., Jarjour J., Kulshina N., Parmeggiani F., Danaher P., Gano J., Baker D., Stoddard B. L. & Scharenberg A. M. (2012) Engineering domain fusion chimeras from I-OnuI family LAGLIDADG homing endonucleases. *Nucleic Acids Res.* **40,** 7985-8000. 10.1093/nar/gks502 [doi].

117. Van Roey P., Waddling C. A., Fox K. M., Belfort M. & Derbyshire V. (2001) Intertwined structure of the DNA-binding domain of intron endonuclease I-TevI with its substrate. *Embo j.* **20,** 3631-3637. 10.1093/emboj/20.14.3631.

118. Derbyshire V., Kowalski J. C., Dansereau J. T., Hauer C. R. & Belfort M. (1997) Two-domain structure of the td intron-encoded endonuclease I-TevI correlates with the two-domain configuration of the homing site. *J. Mol. Biol.* **265,** 494-506. 10.1006/jmbi.1996.0754.

119. Mueller J. E., Smith D., Bryk M. & Belfort M. (1995) Intron-encoded endonuclease I-TevI binds as a monomer to effect sequential cleavage via conformational changes in the td homing site. *Embo j.* **14,** 5724-5735.

120. Edgell D. R., Stanger M. J. & Belfort M. (2004) Coincidence of cleavage sites of intron endonuclease I-TevI and critical sequences of the host thymidylate synthase gene. *J. Mol. Biol.* **343,** 1231-1241. S0022-2836(04)01120-9 [pii].

121. Bell-Pedersen D., Quirk S. M., Aubrey M. & Belfort M. (1989) A site-specific endonuclease and co-conversion of flanking exons associated with the mobile td intron of phage T4. *Gene* **82,** 119-126.

122. Edgell D. R., Shub D. A. (2001) Related homing endonucleases I-BmoI and I-TevI use different strategies to cleave homologous recognition sites. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 7898-7903. 10.1073/pnas.141222498 [doi].

123. Bell-Pedersen D., Quirk S. M., Bryk M. & Belfort M. (1991) I-TevI, the endonuclease encoded by the mobile td intron, recognizes binding and cleavage domains on its DNA target. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 7719-7723.

124. Kowalski J. C., Belfort M., Stapleton M. A., Holpert M., Dansereau J. T., Pietrokovski S., Baxter S. M. & Derbyshire V. (1999) Configuration of the catalytic GIY-YIG domain of intron endonuclease I-TevI: coincidence of computational and molecular findings. *Nucleic Acids Res.* **27,** 2115-2125.

125. Kleinstiver B. P., Wang L., Wolfs J. M., Kolaczyk T., McDowell B., Wang X., Schild-Poulter C., Bogdanove A. J. & Edgell D. R. (2014) The I-TevI nuclease and linker domains contribute to the specificity of monomeric TALENs. *G3 (Bethesda)* **4,** 1155-1165. 10.1534/g3.114.011445 [doi].

126. Beurdeley M., Bietz F., Li J., Thomas S., Stoddard T., Juillerat A., Zhang F., Voytas D. F., Duchateau P. & Silva G. H. (2013) Compact designer TALENs for efficient genome engineering. *Nat. Commun.* **4,** 1762. 10.1038/ncomms2782 [doi].

127. Beer. (1852) Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. *Annalen Der Physik* **162,** 78-88. 10.1002/andp.18521620505.

128. Lambert J. H. (1760) Photometria sive de mensura et gradibus luminis, colorum et umbrae. 391.

129. Bouguer P. (1729) Essai d'Optique, sur la gradation de la lumiere. 16--22.

130. Jeltsch A., Wenz C., Stahl F. & Pingoud A. (1996) Linear diffusion of the restriction endonuclease EcoRV on DNA is essential for the in vivo function of the enzyme. *Embo j.* **15,** 5104-5111.

131. Sokolowska M., Czapinska H. & Bochtler M. (2011) Hpy188I-DNA pre- and post-cleavage complexes--snapshots of the GIY-YIG nuclease mediated catalysis. *Nucleic Acids Res.* **39,** 1554-1564. 10.1093/nar/gkq821 [doi].

132. Truglio J. J., Rhau B., Croteau D. L., Wang L., Skorvaga M., Karakas E., DellaVecchia M. J., Wang H., Van Houten B. & Kisker C. (2005) Structural insights into the first incision reaction during nucleotide excision repair. *Embo j.* **24,** 885-894. 7600568 [pii].

133. Galburt E. A., Chevalier B., Tang W., Jurica M. S., Flick K. E., Monnat R. J.,Jr & Stoddard B. L. (1999) A novel endonuclease mechanism directly visualized for I-PpoI. *Nat. Struct. Biol.* **6,** 1096-1099. 10.1038/70027 [doi].

134. Mak A. N., Lambert A. R. & Stoddard B. L. (2010) Folding, DNA recognition, and function of GIY-YIG endonucleases: crystal structures of R.Eco29kI. *Structure* **18,** 1321-1331. 10.1016/j.str.2010.07.006 [doi].

135. Van Roey P., Meehan L., Kowalski J. C., Belfort M. & Derbyshire V. (2002) Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI. *Nat. Struct. Biol.* **9,** 806-811. 10.1038/nsb853 [doi].

136. Van Roey, P., Belfort, M. & Derbyshire, V. (2005). Homing endonuclease I-TevI: An atypical zinc finger with a novel function. In (Iuchi, S., Kuldell, N., eds.), pp. 35-38, Springer US,.

137. Edgell D. R., Derbyshire V., Van Roey P., LaBonne S., Stanger M. J., Li Z., Boyd T. M., Shub D. A. & Belfort M. (2004) Intron-encoded homing endonuclease I-TevI also functions as a transcriptional autorepressor. *Nat. Struct. Mol. Biol.* **11,** 936-944. 10.1038/nsmb823 [doi].

138. Gibb E. A., Edgell D. R. (2010) Better late than early: delayed translation of intron-encoded endonuclease I-TevI is required for efficient splicing of its host group I intron. *Mol. Microbiol.* **78,** 35-46. 10.1111/j.1365-2958.2010.07216.x [doi].

139. Kleinstiver B. P., Berube-Janzen W., Fernandes A. D. & Edgell D. R. (2011) Divalent metal ion differentially regulates the sequential nicking reactions of the GIY-YIG homing endonuclease I-BmoI. *PLoS One* **6,** e23804. 10.1371/journal.pone.0023804 [doi].

140. Eyring H. (1935) The Activated Complex in Chemical Reactions. *J. Chem. Phys.* **3,** 107-115. 10.1063/1.1749604.

141. Evans M. G., Polanyi M. (1935) Some applications of the transition state method to the calculation of reaction velocities, especially in solution. *Trans.Faraday Soc.* **31,** 875-894. 10.1039/TF9353100875".

Appendix 1: Bacterial strains, plasmids and primers used for the development of this thesis, and raw data underlying the results.

Below are found tables listing the all of the bacterial strains, plasmids, and primers used to develop this thesis (Table S.1, S.2, and S.3, respectively). Also included are tables quantifying the results of the numerous survival assays completed to generate figure 3.4 (Table S.4). Finally, a table of data summarising the k_{app}-values (Table S.5) and the plots with fitting data used to derive them (Figure S.1)

93

Supplemental Table S1: Bacterial strains used in this study

Strains	Description	Source
E.coli – NEB5α	F⁻, φ80d*lacZ*ΔM15, Δ(*lacZYA-argF*)U169, *deoR*, *recA1*, *endA1*, *hsdR*17(rk⁻, mk⁺), *phoA*, *supE*44, λ⁻, *thi*-1, *gyrA*96, *relA1*	N.E.B.
E.coli - ER2566	F– λ– *fhuA2* [*lon*] *ompT lacZ*::T7 gene 1 gal *sulA*11 Δ(*mcrC-mrr*)114::IS10 R(*mcr*-73::mini*Tn10*-TetS)2 R(*zgb*-210::*Tn10*)(TetS) *endA1* [*dcm*]	N.E.B.
E.coli - BW25141(λDE3)	F⁻ *lacI*^{q} *rrnB*$_{T14}$ *ΔlacZ*$_{WJ16}$ *DphoBR580 hsdR514 DaraBAD*$_{AH33}$ *DrhaBAD*$_{LD78}$ *galU95 endA*$_{BT333}$ *uidA*(D*MluI*)::*pir*+ *recA1*, λDE3 lysogen	Ref [6]

Supplemental Table S2: Plasmids used in this study

Plasmids	Description	Source
pACYCDuet-1(PciI)	*ori*$_{p15A}$, cm, pACYCDuet-1 with a PciI site substituted for the NcoI site	Novagen
p11-lacY-wtx1	*ori*$_{pBR322}$, amp	Ref [6]
pSP72	*ori*$_{pBR322}$, amp	Promega
pToxTO1.20	p11-lacY-wtx1, that contains a 42-bp hybrid I-TevI/I-OnuI E1 homing site (*td* bases -27 to -8 fused to the I-OnuI E1 site) cloned into the XbaI and SphI sites(DE1064/1065)	Ref [6]
pToxTO1.20 C1A/G5A	Similar to pToxTO20, with C1A and G5A substitution(DE1156/1157)	Ref [6]
pACYCOnuE1(E22Q)(+H)	pACYCDuet-1(PciI), containing the I-OnuI E1 gene with a E22Q mutation cloned into the BamHI and XhoI sites	Ref [6]
pTevN169-OnuE1(E22Q)(+H)	pACYCOnuE1(E22Q)(+H), with residues 1-N169 of I-TevI (DE) cloned into the PciI and BamHI sites (+6xHis)	Ref [6]

Supplemental Table S3: Oligonucleotides used in this study

Name	Sequence (5'-3')	Notes[1]
DE410	GGAAGAAGTGGCTGATCTCAGC	Forward primer to generate all cycle-seq products for target sites cloned into pTox
DE411	CAGACCGCTTCTGCGTTCTG	Reverse primer to generate all cycle-seq products for target sites cloned into pTox
DE840	GCCGCCATGGGTAAAAGCGGAATTTATCAGATT	Forward primer for I-TevI cloning, NcoI site underlined
DE1045	CGCGGATCCATTTCTGCATTTACTACAAG	Reverse primer for TevN169 cloning, BamHI site underlined
DE1424	CGTTTGGTGATACATGTTCTACG	Reverse primer for I-TevI linker cloning.
DE1912	CGTAGAACATGTATCACCAAACG	Reverse primer for mutagenesis of the I-TevI nuclease domain, PciI site is underlined
DE2183	GGAAGTGCTAAAGATTTTGAATCGAGATGGAAGAGGCATTTTAAAG	Forward primer for installation of K26S into top strand.
DE2184	CTTTAAAATGCCTCTTCCATCTCGATTCAAAATCTTTAGCACTCCC	Reverse primer for installation of K26S into bottom strand.
DE2222	CCCAAACAGGTCGCTGAAATGC	Forward primer for generating the 2200 bp barcode assay substrate from pTox or pKox templates.
DE2223	TGTCACGCTCGTCGTTTGGTATGGC	Reverse primer for generating the 2200 bp barcode assay substrate from pTox templates.
DE2224	ATGACGACCGTAGTGATGAATCTCTCC	Forward primer for generating the 1900 bp barcode assay substrate from pTox or pKox templates.
DE2225	TCATGGTTATGGCAGCACTGC	Reverse primer for generating the 1900 bp barcode assay substrate from pTox templates.
DE2226	AAAAAAATCGAGATAACCGTTGGC	Forward primer for generating the 1600 bp barcode assay substrate from pTox or pKox templates.
DE2227	CCGCGCCACATAGCAGAACTTTAAAAGTGC	Reverse primer for generating the 1600 bp barcode assay substrate from pTox or pKox templates.
DE2228	ATTGTCCATATTGCATCAGACATTGC	Forward primer for generating the 1300 bp barcode assay substrate from pTox or pKox templates.
DE2229	ACTTCACCAGCGTTTCTGG	Forward primer for generating the 1300 bp barcode assay substrate from pTox or pKox templates.
DE2230	AAATTAATAGGTTGTATTGATGTTGGACGAGTCG	Reverse primer for generating the 2200 bp native I-TevI target barcode assay substrate from pKox templates.
DE2231	AAATTGCAGTTTCATTTGATGCTCG	Reverse primer for generating the 1900 bp native I-TevI target barcode assay substrate from pKox templates.
DE2296	TGAGACACAACGTGGCTTTGTTGAATAAATCG	Reverse primer for generating the 1900 bp non-native I-TevI target barcode assay substrate from pKox templates.
DE2297	TCCATGTTGGAATTTAATCGCGGCCTCG	Reverse primer for generating the 2200 bp non-native I-TevI target barcode assay substrate from pKox templates.

[1] underlined nucleotides refer to restriction enzyme sites

Table S.4. (Part 1 of 2) Survival Rates Determined from *in vivo* 2-Plasmid Survival Assay

Substrate	wt	1A(Q158R)	1B(K26R Q158R)	2A(T95S)	2B(I86V T95S)	2C(C39R I86V T95S)	2D(K26R T95S)	2E(C39R T95S)
AAC	89.00±2.65	95.67±13.05	118.00±41.76	131.67±33.65	66.17±29.10	86.43±30.51	77.30±34.13	100.17±37.52
TCG	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
GTG	0.00±0.00	0.00±0.00	0.13±0.06	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
GAA	13.50±6.06	53.33±9.29	86.00±34.04	23.33±4.51	15.77±6.64	17.67±5.13	15.33±2.52	42.67±6.43
GCA	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
TGG	0.00±0.00	1.50±1.22	3.33±1.21	18.23±14.65	10.53±9.04	25.40±14.25	17.87±13.97	29.33±12.70
GCC	0.00±0.00	0.37±0.06	0.90±0.10	43.00±28.05	33.43±22.13	32.67±20.01	29.90±24.06	58.33±28.43
GGA	0.00±0.00	5.90±4.28	5.93±4.27	23.50±18.30	17.37±13.56	26.97±24.11	25.93±24.10	5.73±1.10
GGG	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
GCT	0.00±0.00	0.20±0.00	0.10±0.00	0.10±0.00	0.10±0.00	0.30±0.00	0.10±0.00	0.00±0.00
CCA	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
CAG	0.00±0.00	1.50±0.26	23.00±3.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
CCC	0.00±0.00	4.53±2.27	4.27±2.57	42.33±26.01	46.47±32.10	45.00±22.91	35.20±31.70	74.33±16.56
AAG	1.60±0.82	75.00±36.59	73.67±30.89	6.47±5.20	4.70±3.82	7.47±6.12	7.87±6.47	4.43±3.49
GAG	0.00±0.00	0.00±0.00	0.40±0.17	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
ACG	0.07±0.12	0.00±0.00	2.00±0.35	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	2.43±3.09
C1A	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
C1G	0.00±0.00	0.00±0.00	0.40±0.10	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
C1T	0.00±0.00	8.30±4.12	15.67±8.14	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	2.43±2.57

Table S.4. (Part 2 of 2) Survival Rates Determined from *in vivo* 2-Plasmid Survival Assay

Substrate	2F(K26S T95S)	2G(K26S C39R T95S)	0A(K26R)	0B(C39R)	0C(I86V)	0D(C39R I86V)	0E(K26S)	0F(K26S C39R)	K26R T95S Q158R (T3)
AAC	104.00±30.64	116.33±43.02	64.30±6.55	92.33±11.02	88.60±26.38	72.50±35.57	73.33±22.05	118.33±43.19	139.37±83.39
TCG	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
GTG	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
GAA	7.87±7.06	20.00±18.25	31.33±18.77	27.67±11.24	20.67±10.79	17.33±10.41	0.00±0.00	0.00±0.00	84.67±31.90
GCA	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	21.00±6.24
TGG	1.86±0.22	5.82±3.82	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	88.33±33.95
GCC	2.50±1.83	12.83±10.25	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	71.33±33.84
GGA	0.08±0.13	0.99±1.41	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	80.33±13.80
GGG	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
GCT	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	13.33±0.58
CCA	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	4.97±6.26
CAG	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	2.33±0.55
CCC	4.13±0.68	12.67±20.21	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	64.00±53.23
AAG	0.00±0.00	0.00±0.00	26.33±9.07	14.27±7.75	4.60±2.56	14.27±13.13	0.00±0.00	0.00±0.00	54.00±24.25
GAG	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
ACG	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	24.67±21.94
C1A	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	3.07±2.16
C1G	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	4.93±2.78
C1T	0.00±0.00	0.80±0.44	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	107.33±51.52

Table S.5. Apparent First-Order Rate Constants for Substrate Decay by Chimaeric MegaTevs with wild-type or T3 NDs

	k_{app}			k_{app}	
Substrate	wild-type	T3	Substrate	wild-type	T3
AAA	0.73 + 0.08	0.41 + 0.05	AAC	1.07 + 0.18	0.29 + 0.04
AAC	1.00 + 0.11	0.33 + 0.05	CCG	0.02 + 0.01	0.07 + 0.01
AAG	0.38 + 0.04	0.17 + 0.02	GCG	0.05 + 0.01	0.10 + 0.01
AAT	1.07 + 0.14	0.32 + 0.04	TCG	0.21 + 0.02	0.23 + 0.03
AAC	0.74 + 0.07	0.85 + 0.09	AAC	0.63 + 0.06	0.15 + 0.04
CCC	0.27 + 0.02	0.41 + 0.03	CGG	0.032 + 0.004	0.02 + 0.02
GCC	0.26 + 0.03	0.69 + 0.06	GGG	0.12 + 0.01	0.11 + 0.03
TCC	0.58 + 0.06	0.72 + 0.07	TGG	0.18 + 0.01	0.13 + 0.02
AAC	0.74 + 0.06	0.63 + 0.07	AAC	0.51 + 0.05	0.40 + 0.05
CAG	0.14 + 0.01	0.11 + 0.01	CTG	0.08 + 0.01	0.25 + 0.04
GAG	0.16 + 0.01	0.18 + 0.01	GTG	0.15 + 0.01	0.17 + 0.02
TAG	0.29 + 0.01	0.37 + 0.03	TTG	0.32 + 0.02	0.26 + 0.03
AAC	1.00 + 0.07	0.85 + 0.09	AAC	0.63 + 0.04	0.11 + 0.01
ACG	0.27 + 0.02	0.41 + 0.03	CCT	0.28 + 0.01	0.083 + 0.005
AGG	0.38 + 0.03	0.69 + 0.06	GCT	0.11 + 0.00	0.078 + 0.005
ATG	0.48 + 0.07	0.72 + 0.07	TCT	0.21 + 0.01	0.10 + 0.01
AAC	0.90 + 0.09	0.53 + 0.06	C1A	0.15 + 0.02	0.05 + 0.01
CCA	0.12 + 0.02	0.08 + 0.00	C1C	0.78 + 0.05	0.16 + 0.02
GCA	0.26 + 0.03	0.27 + 0.02	C1G	0.10 + 0.01	0.08 + 0.01
TCA	0.60 + 0.06	0.45 + 0.05	C1T	0.10 + 0.01	0.10 + 0.01
AAC	0.57 + 0.04	0.35 + 0.06	2200	0.75 + 0.07	*n.d.* *
CGA	0.041 + 0.002	0.05 + 0.01	1900	0.77 + 0.08	*n.d.* *
GGA	0.072 + 0.003	0.22 + 0.02	1600	0.67 + 0.08	*n.d.* *
TGA	0.20 + 0.01	0.32 + 0.04	1320	0.63 + 0.06	*n.d.* *

*Not determined

Figure S.1. Barcode assay kinetic data. MegaTevs with a wild-type ND (WT [A, C, E, G, I, K, M, O, Q, S, U, W, X]) and MegaTevs with a triple mutant K26R T95S Q158R (T3 [B, D, F, H, J, L, N, P, R, T, V]) were assayed against four substrates of varying lengths (2200 [yellow ▲], 1900 [blue ▲], 1600 [green ▲], and 1320 bp [red ▲]). Substrates harboured I-TevI cleavage motifs, one of which was the native cleavage motif (5' – CAACG – 3'), and the others were comprised of NNN triplet substitutions (5' – CNNNG – 3' [A-T]), position 1 substitutions (5' – NAACG – 3' [U, V]), or control substrates. The controls involved either all native cleavage motifs, one of each of four lengths (W), or a single native cleavage motif (1C5G [X]) among cleavage resistant motifs (5' – AAACA – 3', 1A5A [X]). All assays were conducted at 5°C, with 250 nM enzyme, and 5 nm of each substrate. The equation of fit is explained in detail in the text (see eqn. 2.2). Note that the equation given below is superficially different; using the identities $y = f_S$, and $x = t$, the equations below become eqn 2.2.

A   **WT**

fraction of remaining substrate

time, min

	2200
	1900
	1600
	1320

NNN	relk_{app}
AAA	0.73
AAC	1
AAG	0.38
AAT	1.07

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.26672	0.022298
m2	0.77003	0.026134
m3	0.72657	0.080771
Chisq	0.10158	NA
R²	0.97315	NA

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.21954	0.019507
m2	0.7785	0.023552
m3	0.99918	0.11075
Chisq	0.090722	NA
R²	0.97811	NA

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.21968	0.025995
m2	0.77533	0.028382
m3	0.38357	0.039951
Chisq	0.088952	NA
R²	0.9719	NA

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.26594	0.021861
m2	0.73233	0.026557
m3	1.0718	0.1471
Chisq	0.11747	NA
R²	0.96875	NA

B   **T3**

fraction of remaining substrate

time, min

	2200
	1900
	1600
	1320

NNN	relk_{app}
AAA	1.24
AAC	1
AAG	0.51
AAT	0.98

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.45429	0.02078
m2	0.53871	0.023282
m3	0.49065	0.050414
Chisq	0.057022	NA
R²	0.95781	NA

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.46689	0.024035
m2	0.52137	0.028393
m3	0.32727	0.046726
Chisq	0.079347	NA
R²	0.94432	NA

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.48946	0.021818
m2	0.49894	0.022716
m3	0.16828	0.020137
Chisq	0.040459	NA
R²	0.99978	NA

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.54266	0.02021
m2	0.44995	0.022156
m3	0.32148	0.044577
Chisq	0.054451	NA
R²	0.94729	NA

Figure S.1 (page 1 of 12)

Figure S.1 (page 2 of 12)

E **WT**

fraction of remaining substrate

time, min

Legend: 2200, 1900, 1600, 1320

NNN	relk_app
AAC	1
CAG	0.19
GAG	0.22
TAG	0.39

(orange) $y = m1 + m2*\exp(-m3*x)$

	Value	Error
m1	0.13896	0.01825
m2	0.85073	0.021503
m3	0.73913	0.060693
Chisq	0.069929	NA
R^2	0.99496	NA

(blue) $y = m1 + m2*\exp(-m3*x)$

	Value	Error
m1	0.14903	0.011958
m2	0.85209	0.012249
m3	0.13732	0.0060803
Chisq	0.009237	NA
R^2	0.99616	NA

(green) $y = m1 + m2*\exp(-m3*x)$

	Value	Error
m1	0.21554	0.013964
m2	0.78538	0.014486
m3	0.1603	0.0077294
Chisq	0.015871	NA
R^2	0.99327	NA

(red) $y = m1 + m2*\exp(-m3*x)$

	Value	Error
m1	0.26658	0.01136
m2	0.73208	0.012339
m3	0.29108	0.013698
Chisq	0.016393	NA
R^2	0.99362	NA

F **T3**

fraction of remaining substrate

time, min

Legend: 2200, 1900, 1600, 1320

NNN	relk_app
AAC	1
CAG	0.18
GAG	0.29
TAG	0.58

(orange) $y = m1 + m2*\exp(-m3*x)$

	Value	Error
m1	0.3278	0.020595
m2	0.66961	0.023961
m3	0.62977	0.07073
Chisq	0.082875	NA
R^2	0.9698	NA

(blue) $y = m1 + m2*\exp(-m3*x)$

	Value	Error
m1	0.36523	0.011082
m2	0.63538	0.011224
m3	0.11471	0.0062668
Chisq	0.00478	NA
R^2	0.99669	NA

(green) $y = m1 + m2*\exp(-m3*x)$

	Value	Error
m1	0.39563	0.0088968
m2	0.60338	0.0093326
m3	0.18478	0.0075888
Chisq	0.0072656	NA
R^2	0.99507	NA

(red) $y = m1 + m2*\exp(-m3*x)$

	Value	Error
m1	0.42001	0.017089
m2	0.57328	0.018964
m3	0.39661	0.034454
Chisq	0.043032	NA
R^2	0.9751	NA

Figure S.1 (page 3 of 12)

G

fraction of remaining substrate

time, min

WT

Legend: 2200, 1900, 1600, 1320

NNN	rel k_{app}
AAC	1
ACG	0.27
AGG	0.38
ATG	0.48

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.19627	0.012901
m2	0.9026	0.014353
m3	1.0016	0.070997
Chisq	0.061903	NA
R²	0.98892	NA

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.808	0.015996
m2	0.27281	0.015218
m3	0.19257	0.015337
Chisq	0.044457	NA
R²	0.9881	NA

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.28648	0.01711
m2	0.71351	0.018145
m3	0.37652	0.028243
Chisq	0.082236	NA
R²	0.97945	NA

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.39348	0.026439
m2	0.60615	0.028395
m3	0.48355	0.069985
Chisq	0.18705	NA
R²	0.93185	NA

H

fraction of remaining substrate

time, min

T3

Legend: 2200, 1900, 1600, 1320

NNN	rel k_{app}
AAC	1
ACG	0.49
AGG	0.83
ATG	0.85

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.21974	0.016022
m2	0.77637	0.024674
m3	0.84635	0.092108
Chisq	0.095567	NA
R²	0.97599	NA

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.21983	0.020706
m2	0.77501	0.017967
m3	0.41404	0.027747
Chisq	0.040707	NA
R²	0.98743	NA

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.28599	0.017012
m2	0.71025	0.019938
m3	0.65279	0.0624
Chisq	0.05903	NA
R²	0.98129	NA

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.3481	0.015798
m2	0.64898	0.018565
m3	0.71638	0.06641
Chisq	0.051673	NA
R²	0.98057	NA

Figure S.1 (page 4 of 12)

Figure S.1 (page 5 of 12)

K   **WT**

fraction of remaining substrate vs time, min

NNN	rel k_{app}
AAC	1
CGA	0.07
GGA	0.13
TGA	0.36

$y = m1 + m2*exp(-m3*x)$

	Value	Error
m1	0.18126	0.017442
m2	0.81434	0.020111
m3	0.5663	0.042612
Chisq	0.056603	NA
R^2	0.98655	NA

$y = m1 + m2*exp(-m3*x)$

	Value	Error
m1	0.17552	0.029355
m2	0.82574	0.02878
m3	0.41089	0.0025193
Chisq	0.0023481	NA
R^2	0.99763	NA

$y = m1 + m2*exp(-m3*x)$

	Value	Error
m1	0.088642	0.015325
m2	0.9165	0.01495
m3	0.072085	0.0026775
Chisq	0.0045219	NA
R^2	0.99779	NA

$y = m1 + m2*exp(-m3*x)$

	Value	Error
m1	0.1249	0.0082741
m2	0.87619	0.0087264
m3	0.20138	0.0053848
Chisq	0.0067098	NA
R^2	0.99791	NA

L   **T3**

fraction of remaining substrate vs time, min

NNN	rel k_{app}
AAC	1
CGA	0.13
GGA	0.61
TGA	0.91

$y = m1 + m2*exp(-m3*x)$

	Value	Error
m1	0.50615	0.027152
m2	0.48439	0.030043
m3	0.35449	0.062448
Chisq	0.10656	NA
R^2	0.9179	NA

$y = m1 + m2*exp(-m3*x)$

	Value	Error
m1	0.47875	0.077525
m2	0.52033	0.075795
m3	0.045878	0.012398
Chisq	0.024926	NA
R^2	0.95258	NA

$y = m1 + m2*exp(-m3*x)$

	Value	Error
m1	0.38209	0.022285
m2	0.54707	0.020287
m3	0.21633	0.021725
Chisq	0.037817	NA
R^2	0.97136	NA

$y = m1 + m2*exp(-m3*x)$

	Value	Error
m1	0.4441	0.019137
m2	0.60859	0.024447
m3	0.32391	0.036664
Chisq	0.06775	NA
R^2	0.96425	NA

Figure S.1 (page 6 of 12)

104

Figure S.1 (page 7 of 12)

O **WT**

	NNN	rel k_app
AAC	1
CGG	0.50
GGG	0.19
TGG	0.28

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.22524	0.020991
m2	0.78983	0.024422
m3	0.63308	0.062861
Chisq	0.086295	NA
R²	0.97629	NA

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.090642	0.072446
m2	0.91389	0.071467
m3	0.031952	0.0039282
Chisq	0.0053181	NA
R²	0.99414	NA

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.35666	0.010953
m2	0.64763	0.011085
m3	0.12029	0.0052305
Chisq	0.0070111	NA
R²	0.995	NA

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.30636	0.015029
m2	0.69109	0.015717
m3	0.17882	0.010765
Chisq	0.020199	NA
R²	0.98948	NA

P **T3**

	NNN	rel k_app
AAC	1
CGG	0.12
GGG	0.70
TGG	0.83
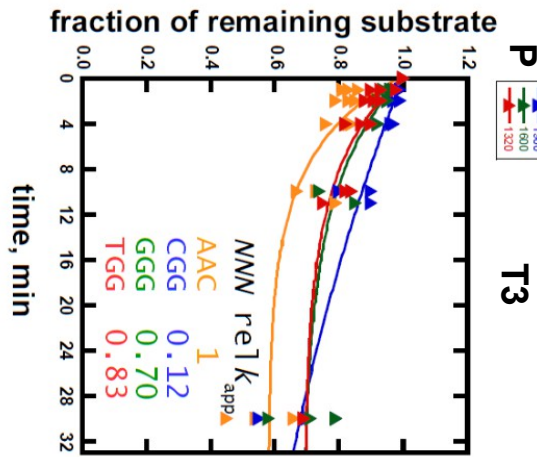
y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.57937	0.034611
m2	0.40329	0.035076
m3	0.15335	0.035812
Chisq	0.094348	NA
R²	0.86636	NA

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.2696	0.029492
m2	0.72887	0.55657
m3	0.01912	0.019202
Chisq	0.040094	NA
R²	0.87602	NA

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.6882	0.029492
m2	0.30783	0.029515
m3	0.10768	0.026089
Chisq	0.042572	NA
R²	0.87592	NA

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.693	0.016682
m2	0.30121	0.016953
m3	0.1271	0.018465
Chisq	0.017607	NA
R²	0.94677	NA

Figure S.1 (page 8 of 12)

Q — WT

NNN	rel k_{app}
AAC	1
CTG	0.15
GTG	0.29
TTG	0.63

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.18822	0.023399
m2	0.80439	0.026763
m3	0.5134	0.051001
Chisq	0.097196	NA
R²	0.97417	

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.48115	0.019658
m2	0.51695	0.019125
m3	0.5134	0.0066471
Chisq	0.0089017	NA
R²	0.98712	

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.1909	0.013289
m2	0.81287	0.013715
m3	0.15016	0.0065781
Chisq	0.013499	NA
R²	0.9945	

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.19212	0.017346
m2	0.8035	0.019033
m3	0.32473	0.02168
Chisq	0.041115	NA
R²	0.98727	

R — T3

NNN	rel k_{app}
AAC	1
CTG	0.62
GTG	0.43
TTG	0.64

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.40069	0.023721
m2	0.59081	0.026542
m3	0.39928	0.051572
Chisq	0.087392	NA
R²	0.95462	

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.61415	0.021556
m2	0.38169	0.023083
m3	0.24675	0.0401
Chisq	0.052668	NA
R²	0.92561	

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.54684	0.015538
m2	0.45133	0.016192
m3	0.17034	0.016085
Chisq	0.020737	NA
R²	0.97455	

y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.56517	0.014344
m2	0.43427	0.015413
m3	0.2572	0.025194
Chisq	0.024003	NA
R²	0.97287	

Figure S.1 (page 9 of 12)

108

Figure S.1 (page 10 of 12)

U **WT**

Legend: 2200, 1900, 1600, 1320

NNN	rel k_{app}
C1A	0.19
C1C	1
C1G	0.12
C1T	0.12

Orange: y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.75697	0.012596
m2	0.29899	0.012923
m3	0.14913	0.02315
Chisq	0.013596	NA
R^2	0.94448	NA

Blue: y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.20296	0.011692
m2	0.79477	0.013745
m3	0.78122	0.045276
Chisq	0.030598	NA
R^2	0.99258	NA

Green: y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.52484	0.019846
m2	0.47218	0.019561
m3	0.096382	0.010414
Chisq	0.015662	NA
R^2	0.97888	NA

Red: y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.51454	0.018956
m2	0.48285	0.01867
m3	0.095133	0.0095497
Chisq	0.013668	NA
R^2	0.98192	NA

V **T3**

Legend: 2200, 1900, 1600, 1320

NNN	rel k_{app}
C1A	0.33
C1C	1
C1G	0.47
C1T	0.63

Orange: y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.74157	0.033182
m2	0.25906	0.0323376
m3	0.053452	0.013315
Chisq	0.0082978	NA
R^2	0.95198	NA

Blue: y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.3451	0.028091
m2	0.63861	0.029131
m3	0.16352	0.019525
Chisq	0.067685	NA
R^2	0.95915	NA

Green: y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.44089	0.021978
m2	0.55548	0.021504
m3	0.076145	0.0068206
Chisq	0.011265	NA
R^2	0.98595	NA

Red: y = m1 + m2*exp(-m3*x)

	Value	Error
m1	0.45628	0.017474
m2	0.53652	0.0174444
m3	0.10363	0.008385
Chisq	0.014474	NA
R^2	0.98412	NA

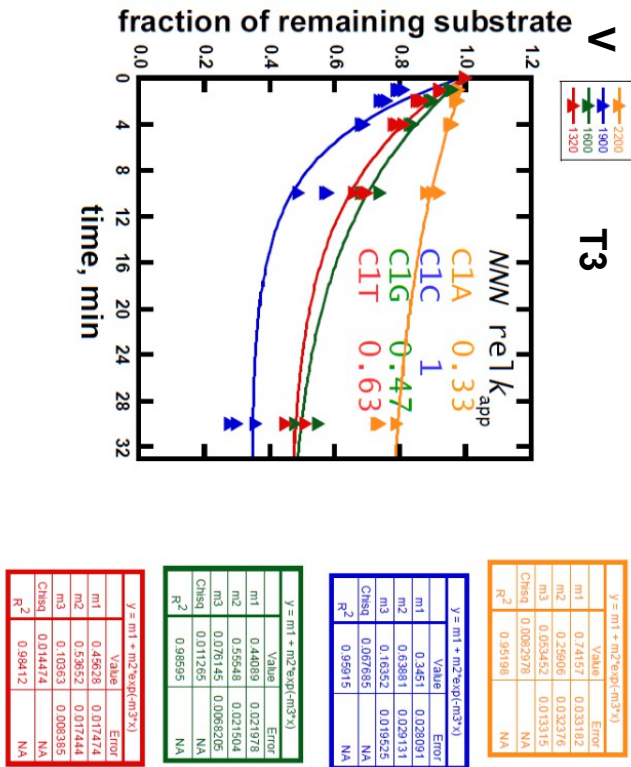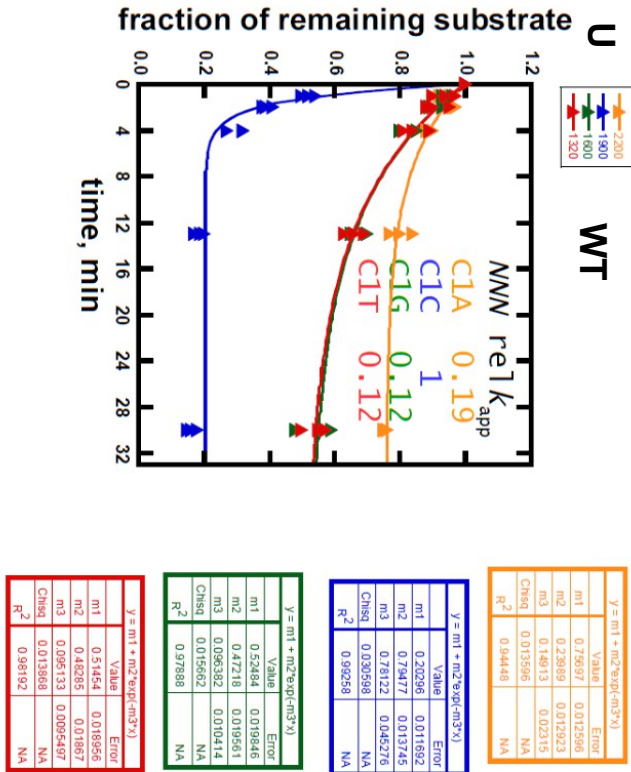**Figure S.1 (page 11 of 12)**

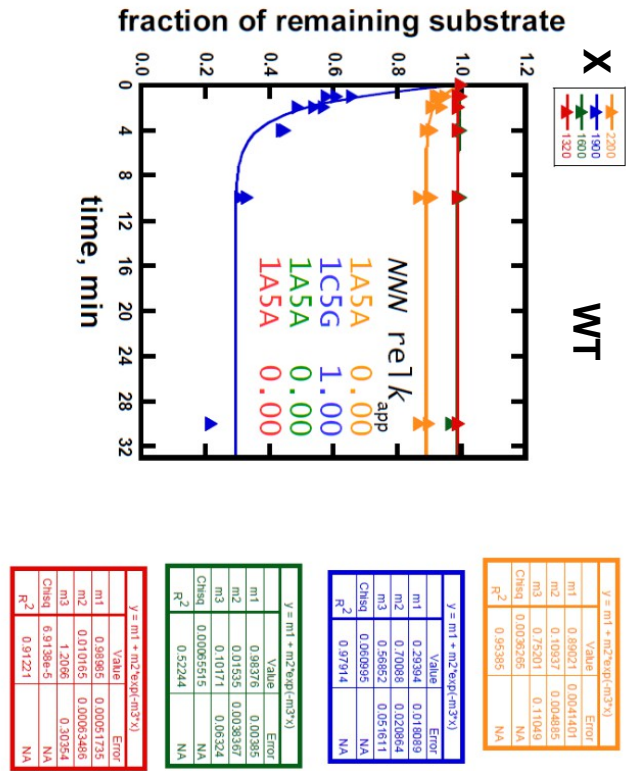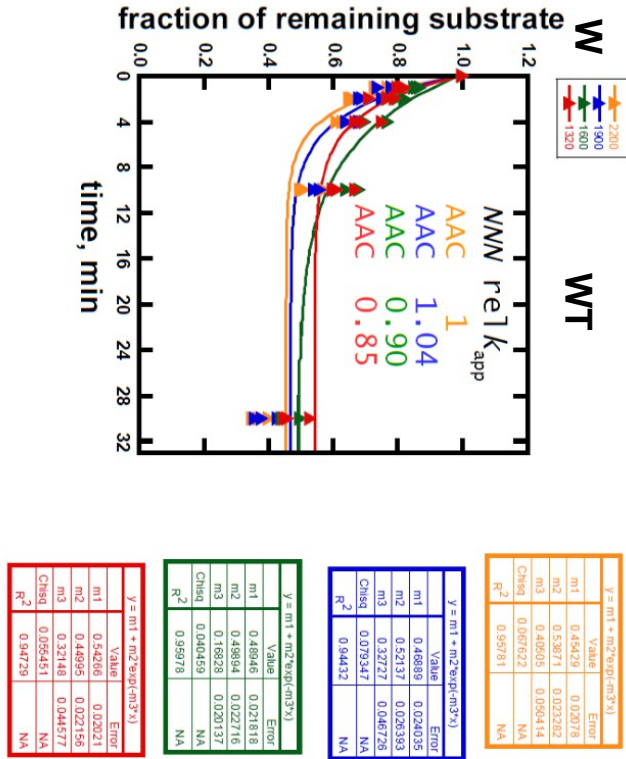**Figure S.1 (page 12 of 12)**

Curriculum Vitae

Alexander Clark Roy

Date of Birth: February 25th, 1987, Halifax, Nova Scotia

EDUCATION

2013   **MSc (Chemistry)**
            Department of Chemistry
            Dalhousie University

            Awards: NSHRF Student Research Award (2011)
                         NSERC A.G. Bell Canadian Graduate Scholarship (2010)

            Thesis Title: *Catalysis by extremozymes: comparing orotidine 5′-monophosphate decarboxylases from psychrophiles, mesophiles, and thermophiles*

            Supervisor: Professor Stephen L. Bearne

2009   **BSc (Honours, Biochemistry & Molecular Biology and Chemistry)**
            Department of Biochemistry & Molecular Biology
            Dalhousie University

            Awards: NSERC Undergraduate Student Research Award (2008)
                         Dalhousie In-Course Scholarship (2007)
                         Norah Stephens Oncology Scholar Award (2007)
                         Dalhousie Entrance Scholarship (2005)

            Thesis Title: *Identification and Characterisation of the GTP-binding Site of CTP Synthase*

            Supervisor: Professor Stephen L. Bearne

PUBLICATIONS

Peer-Reviewed Publications:
(* designates corresponding author, supporting agencies are given in parentheses)

Roy, A.C., Lunn, F.A., and Bearne, S.L.* (2010) "Inhibition of CTP Synthase from *Escherichia coli* by Xanthines and Uric Acids", *Bioorg. Med. Chem. Lett.* **20**, 141-144. (**CIHR, NSERC**)

AWARDS & SCHOLARSHIPS

2012   NSERC Postgraduate Scholarship - Doctoral (National-level)

2011   NSHRF Student Research Award (Provincial-level)

2010   NSERC A.G. Bell Canadian Graduate Scholarship - Master's (National-level)

2008   NSERC Undergraduate Student Research Award (National-level)

2007   Norah Stephen Oncology Scholar Award (Provincial-level)

2007   Dalhousie In-course Scholarship (Institution-level)

2005   Dalhousie Entrance Scholarship (Institution-level)

RESEARCH CONTRIBUTIONS

A. *Development of Inhibitors for CTP Synthase (CTPS)*

During the completion of my undergraduate degree I studied the enzyme CTP synthase (CTPS); a target for chemotherapeutics. To explore this topic, I synthesized azide-based photo-labels to locate the GTP-binding site of CTPS and aid the rational design of GTP-based inhibitors. I also synthesized guanosine-based inhibitors with varying functional groups (amines, alcohols, carboxylic acids) at the 6-position of the guanine heterocycle. Prof. Bearne and I designed these inhibitors to investigate the nature of interactions between CTPS and the 6-position of GTP, which had been shown to be important for activation of CTPS, but was predicted in modeling studies to have no protein contacts. Using spectrophotometric assays, I found that the best inhibitor I developed was actually one of the labels, which possessed a large, non-polar, aromatic benzyl group at the 6-position, implying that this part of the molecule was likely buried in a hydrophobic binding pocket. I was able to use this label to irreversibly inhibit CTPS upon photoactivation of the azide moiety (implying that labeling had occurred); however, my attempts to identify the labeled site using ESI-MS proved impossible.

Before commencing my graduate studies I studied CTPS-inhibition by methylated xanthines (such as caffeine) and uric acids. I identified the urates (conjugate bases of uric acids) as better inhibitors, and performed additional experiments to investigate why. I found that at a higher pH, inhibition by the xanthines increased, implying that deprotonation of the xanthine rings (*i.e.* negative charge) augments inhibition. Further, I observed that methylated xanthines and uric acids are more tightly bound than guanosine, or the guanosine nucleotides. This data was published in Bio-organic and Medicinal Chemistry Letters (see above).

B. *Thermodynamic Comparison of Enzymes from Psychrophiles, Mesophiles, and Thermophiles*

I am currently completing my Chemistry MSc thesis on enzymes from psychrophilic ("cold loving") organisms. In particular, I am using the temperature dependence of k_{cat}, K_m, and k_{non} to study the thermodynamic differences between the mechanism of cold-adapted OMP decarboxylases (ODCases) and their moderate- or high-temperature-adapted homologs. From cloned ORFs encoding *Psychrobacter arcticus* 273-4 and *Colwellia psychrerythraea* 34 H ODCases, I was able to over-express and purify these enzymes using transformed *E. coli* BL21(DE3) cells. By comparing the thermodynamic parameters of cold-adapted ODCases to their homologs, I observed that the cold-adapted variants reduced the enthalpy and entropy of the reaction barrier to a greater extent. This is consistent with the lesser impact of entropy reduction on free energy increase at low temperatures (*i.e.* $\Delta G = \Delta H - T\Delta S$).

TEACHING EXPERIENCE

While completing my undergraduate degree, I volunteered for "Let's Talk Science!", a program that gives university students of science and engineering the chance to present interactive science demonstrations to students in elementary, junior-high, and high-schools. I taught a group of grade six students about how batteries work, and it was a lot of fun. In addition, I was a chemistry and physics tutor to high-school students between my undergraduate classes.

During my graduate studies I had the opportunity to work as a teaching assistant (part-time) in the undergraduate-student organic chemistry laboratory. I was responsible for setting up demonstrations, marking reports and giving feedback, teaching students techniques, and generally ensuring a safe and efficient laboratory experience. Because I had a great deal of experience with the faculties and undergraduate courses of both biochemistry and chemistry, my students also relied on me for advice about course selection, choosing a major, and choosing which supervisor they should do their undergraduate research project with.