# GENERATION AND CHARACTERIZATION OF TAL EFFECTOR NUCLEASES WITH NOVEL CATALYTIC DOMAINS

FABIAN BIETZ

Dissertation zur Erlangung des Akademischen Grades des Doktors der Naturwissenschaften (Dr. rer. nat.)

Fachbereich 08
Institut für Biochemie
Justut Liebig Universität
July, 2015

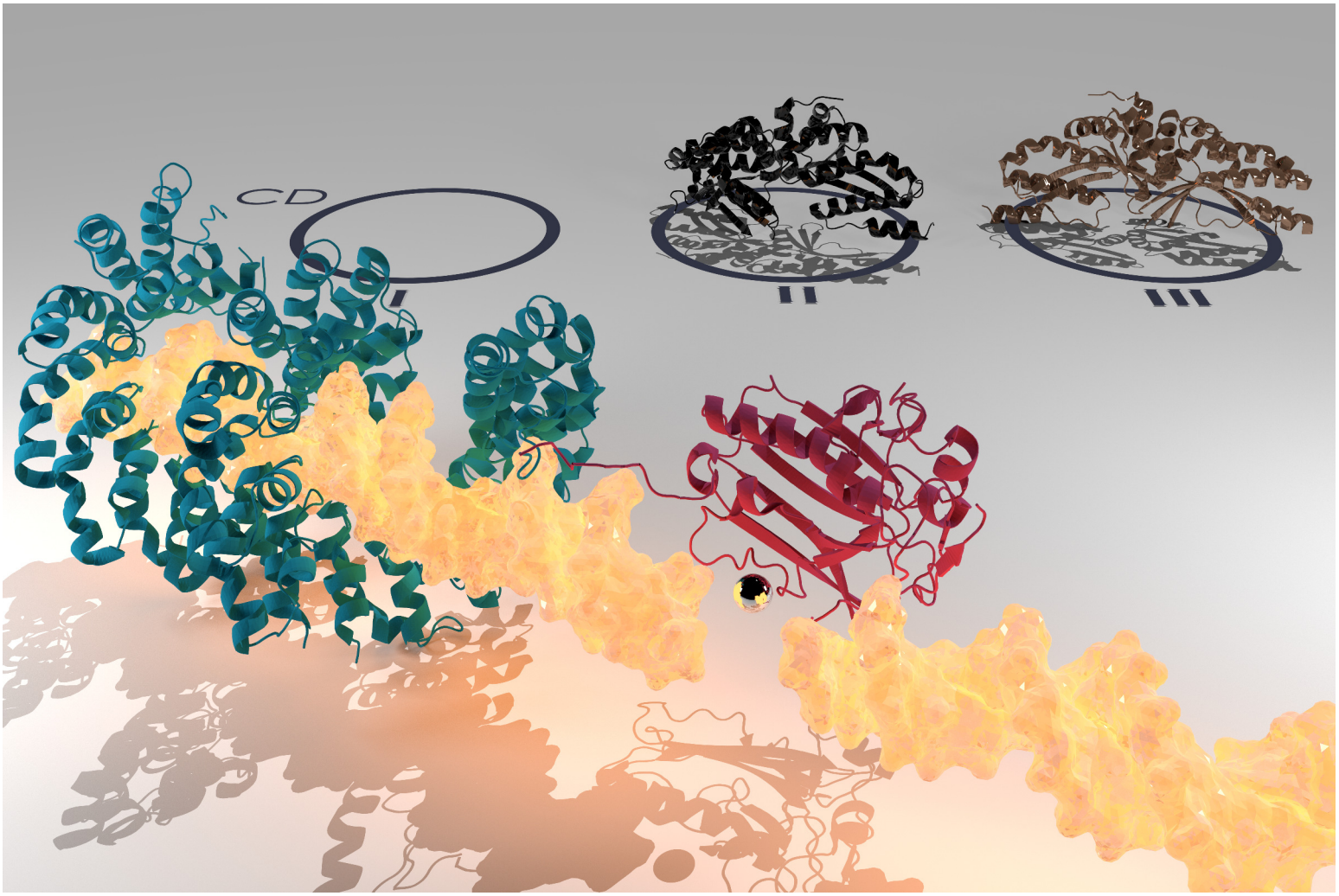
This dissertation was prepared in the research and development division *Computational Biology* at Cellectis (Paris) and at the *Institute for Biochemistry* in the department *Biology and Chemistry* (FB08) of the *Justus-Liebig-Universität Gießen* from April,2011 to July 2015. The work was supported by the DFG financed graduate college "Enzymes and Multienzyme Complexes acting on Nucleic Acids" (IRTG GRK 1384).

PRIMARY REFEREE:
Prof. Dr. Peter Friedhoff
Institut für Biochemie
Fachbereich Biologie und Chemie
Heinrich-Buff-Ring 58
35392 Gießen

SECONDARY REFEREE:
Prof. Dr. Michael U. Martin
Institut für Immunologie
Fachbereich Biologie und Chemie
Schubertstraße 81
35392 Giessen

SUPERVISORS:
Dr. George Silva and Dr. Wolfgang Wende

Fabian Bietz: *Generation and characterization of TAL effector nucleases with novel catalytic domains,* Dissertation zur Erlangung des Akademischen Grades des Doktors der Naturwissenschaften (Dr. rer. nat.), July, 2015




*I learned once again what a many-sided thing*
*is the telling of any tale*
— Gene Wolfe

*Dedicated to my parents, to my sister and to all other people*
*who I enjoyed spending time with in the last few years*

## ABSTRACT

Site-specific nucleases (SSNs) are molecular tools to introduce DNA double-strand breaks (DSBs) at definite genomic loci. DSBs can be exploited to knock-out, delete, repair or insert genes of interest. Construction of SSNs was simplified tremendously with the discovery of transcription activator-like effector (TALE) proteins. The novel TALE nucleases (TALENs) consist of a TALE-derived DNA binding domain, guiding the construct to its target site, and a nuclease domain, which is cleaving the DNA. The standard nuclease domain is the catalytic domain of type IIS restriction endonuclease FokI, which was adopted from the older zinc-finger-nuclease architecture. FokI requires dimerization for the creation of a DSB, making two TALENs necessary. Aim of this work is the replacement of this catalytic domain to generate monomeric TALENs, that simplify production and transfection. To cover nuclease domains with varying degrees of specificity, three groups were chosen: promiscuous H-N-H and DRGH nucleases (I), the more specific I-TevI catalytic domain (II) and variants of the highly-specific homing endonuclease I-CreI (III). The selected domains were adapted to the fusion scaffold via rational design strategies and tested *in vitro* and *in vivo* in several model organisms. Special properties of these domains made the generation of novel, monomeric TALENs possible. Colicin E7, Nuclease A and Endonuclease A (I) were used to create switchable TALENs and I-CreI (III) fusion yielded a highly specific construct. Investigation of the influence of the fusion terminus allowed the construction of scaffolds with multiple nuclease domains via I-TevI (II) and FokI.



## ZUSAMMENFASSUNG

Hoch-spezifische Nukleasen (SSNs) sind molekulare Werkzeuge um DNA-Doppelstrangbrüche (DSBs) an definierten Stellen im Genom zu erzeugen. Diese DSBs können für Gen-*Knockout*, -Deletion, -Reparatur oder -Insertion genutzt werden. Die Herstellung der SSNs wurde durch die Entdeckung der *Transcription Activator-Like Effector*-(TALE)-Proteine enorm vereinfacht. Die neuen TALE-Nukleasen (TALEN) bestehen aus einer TALE-basierten DNA-Bindedomäne, die das Konstrukt zu seiner Zielsequenz führt, und einer Nukleasedomäne, welche die DNA schneidet. Die gängigste Nukleasedomäne ist die katalytische Domäne der Typ IIS Restriktionsendonuklease FokI, die aus der älteren Zinkfingernukleasen-Architektur übernommen wurde. Da FokI eine

Dimerisierung zur DSB-Erzeugung erfordert, werden zwei TALENs benötigt. Ziel dieser Arbeit ist der Austausch dieser katalytischen Domäne um monomere TALEN, die Herstellung und Transfektion vereinfachen, anzufertigen. Um Nukleasedomänen mit verschiedenem Grad an Spezifität abzudecken, wurden drei Gruppen ausgewählt: unspezifische H-N-H und DRGH Nukleasen (I), die spezifische, katalytische Domäne von I-TevI (II) und Varianten der hoch-spezifischen Homingendonuklease I-CreI (III). Die gewählten Domänen wurden durch rationales Design für die Fusion adaptiert und *in vitro* sowie *in vivo* in verschiedenen Modelorganismen getestet. Die besonderen Eigenschaften der jeweiligen Domänen machte die Entwicklung neuartiger, monomerer TALEN möglich. Colicin E7, Nuklease A und Endonuklease A (I) wurden verwendet um regulierbare TALEN zu erzeugen und aus der Fusion mit I-CreI (III) entstand eine hoch-spezifischen Nuklease. Untersuchungen über den Einfluss des Fusionsterminus erlaubte die Herstellung von Gerüsten mit multiplen Nukleasedomänen mittels I-TevI (II) und FokI.

# PUBLICATIONS

Some ideas have appeared previously in the following publications:

Beurdeley, M., F. Bietz, J. Li, S. Thomas, T. Stoddard, A. Juillerat, F. Zhang, D. F. Voytas, P. Duchateau, and G. H. Silva (2013, January). Compact designer TALENs for efficient genome engineering. *Nature communications* 4, 1762

Juillerat, A., M. Beurdeley, J. Valton, S. Thomas, G. Dubois, M. Zaslavskiy, J. Mikolajczak, F. Bietz, G. H. Silva, A. Duclert, F. Daboussi, and P. Duchateau (2014, January). Exploring the transcription activator-like effectors scaffold versatility to expand the toolbox of designer nucleases. *BMC molecular biology* 15(1), 13

# CONTENTS

# LIST OF FIGURES

# LIST OF TABLES



# ACRONYMS

aa amino acids

A-NHEJ alternative non-homologous end-joining

CaMV cauliflower mosaic virus

CAR chimeric antigen receptor

CAS CRISPR-associated system

CHO chinese hamster ovary

C-HR classical homologous recombination

CIAP calf intestinal alkaline phosphatase

CMV cytomegalovirus

C-NHEJ classical non-homologous end-joining

CRISPR clustered regularly interspaced short palindromic repeats

crRNA CRISPR RNA

cTALEN compact TALEN

cv column volumes

DBD DNA-binding domain

dcTALEN dual catalytic TALEN

dHJ double Holliday junction

DNA desoxyribonucleic acid

dNTP deoxyribonucleotide triphosphate

DSB double-strand break

DSBR double-strand break repair

ds double-strand

EBE effector binding element

EDTA ethylenediaminetetraacetate

for forward primer

GMOs genetically modified organisms

gRNA guide RNA

HE homing endonuclease

HEK293 human embryonic kidney cell line 293

HR homologous recombination

INDELs insertion and deletion events

IPTG isopropyl-beta-D-thiogalactopyranosid

MCS multiple cloning site

MID multiplex identifier

NHEJ non-homologous end-joining

NLS nuclear localisation signal

O/N over night

PAGE polyacrylamide gel electrophoresis

PAM protospacer-adjacent motif

PBS phosphate-buffered saline

PCR polymerase chain reaction

RE restriction endonuclease

rev reverse primer

RGEN RNA-guided endonuclease

RipTAL *Ralstonia* injected protein TAL

RNA ribonucleic acid

RT room temperature

RVD repeat variable diresidue

SDSA synthesis-dependent strand annealing

SDS sodium dodecyl sulfate

SOE splicing by overlap extension

SSA single-strand annealing

ssDNA single-stranded DNA

SSN site-specific nuclease

ss single-strand

TALEN transcription activator-like effector nuclease

TALE transcription activator-like effector

t-DNA transfer DNA

TFO triplex-forming oligonucleotid

tracrRNA trans-activating CRISPR RNA

WT wild-type

X-Gal 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

ZFN zinc-finger nuclease

Amino acids and nucleotides are abbreviated with single-letter codes. SI units or SI-accepted units (e. g., litre, hour, °Celsius, Dalton etc.) are employed in this work.

# Part I

# INTRODUCTION

# 1 INTRODUCTION

Since the first steps in agriculture and animal husbandry, humans, albeit unintentional at first, are breeding plant and animal species towards beneficial ends. Modification on the molecular level however, started with the advent of genetics and biotechnology. Even before the discovery of the desoxyribonucleic acid (DNA) double-helix [234], people found out how to increase the natural mutation rate of organisms to yield mutants with desired traits [197, 3]. The main drawback is the random nature of mutagenesis.

Changes came when genetic backgrounds were investigated and recombinant DNA methods were developed [44]. Plant genomes could be modified with bacterial vectors or with a particle inflow gun [163, 67], while micro injection and viral vectors allowed the modification of animal genomes [100]. Even human gene therapy became a possibility with these methods [76]. One main obstacle of these approaches is the pseudo random nature of integration of the transferred DNAs. As a consequence, integration events can be cytotoxic or worse: inactivate tumor suppressor genes. Progress was made with the use of homology directed repair (see subsection 1.3.2) to insert genes. Here, a template DNA with homology regions is transfected into a cell and then integrated into the genome via the cell's own repair mechanisms (see section 1.3). Although integration is site-specific, homologous recombination is rare in somatic cells of higher organisms. Fortunately, the frequency of integration can be increased by several orders of magnitude by stimulating homologous recombination (HR) through introduction of a DNA double-strand break (DSB) in the region of interest [201, 172]. This site-specific DSB, can also be repaired via the non-homologous end-joining (NHEJ) pathway (see subsection 1.3.1), leading to small insertion and deletion events (INDELs), which can be exploited for the creation of gene knock-outs. Both however are reliant on highly specific endonucleases with programmable target site preferences (see figure 1). Approaches to provide these were the reprogramming of highly specific homing endonucleases (see section 1.2.2) and the creation of chimeric FokI based zinc-finger nuclease (ZFN)s (see section 1.2.1). These ZFNs cleave DNA as dimers, each consisting of one DNA-binding domain (DBD) and one FokI based nuclease domain. Both platforms have been applied for genome modification in research and a successful therapeutic approach for ZFNs has been reported [213].

Despite this progress adjustment of the target-site specificity and prevention of off-target cleavage remain non-trivial and time-consuming. The breakthrough for genome editing came with the deciphering of

the transcription activator-like effector (TALE) DNA recognition code (see section 2) and the application of clustered regularly interspaced short palindromic repeats (CRISPR)/ CRISPR-associated system (CAS) proteins (see section 1.2.3). Main advantages of the TALE-DBD over the zinc finger DBD are its improved specificity and ease of reprogramming. Therefore it is no surprise that they were used to replace the zinc finger DBD in the new transcription activator-like effector nuclease (TALEN) architecture. The ribonucleic acid (RNA) guided endonuclease Cas9 on the other hand proved to be even easier to retarget and was shown to be ideal for multiplex genome editing. Drawbacks are the large size of Cas9 variants [71] and significant off-site activity [232, 74, 221].

So while TALENs seem to be among the most specific approaches, they can be difficult to transfect for some applications. TALENs have been introduced into cells via plasmid DNA, mRNA or as proteins [4, 135], but this is not ideal for every cell type. Problems also occur, when viral vectors are the preferred method of transfection. Adeno-associated viruses only assemble efficiently when the expression cassette is smaller than 4.7 kb [87], which would only fit one TALEN monomer, a promoter and a small repair template. The repetitive nature of the underlying TALEN DNA seems to make delivery via integrase-defective lentiviral vectors difficult, which requires the generation of modified versions [153].

A monomeric TALEN can circumvent difficulties during transfection, reduce production cost and also broaden the range of targetable regions, since regions with only one suitable binding site can now be addressed. In this work, monomeric nuclease domains are investigated in a TALE context and optimized to fit different applications. Three different approaches were pursued: non-specific (ColE7, NucA, EndA), semi-specific (I-TevI) and highly specific nuclease domains (I-CreI). These monomeric nuclease domains made the creation of completely new switchable and multi-domain architectures possible.

## 1.1 TRANSCRIPTION ACTIVATOR-LIKE EFFECTOR NUCLEASES

TALENs possess a bipartite layout, consisting of a TALE-based DNA-binding domain and a nuclease domain, which was adopted from the older zinc-finger nuclease architecture.

### 1.1.1 *Transcription activator-like effectors*

The genus *Xanthomonas* consists of bacterial plant pathogens, which secrete TALE proteins as pathogenicity factors. TALEs are translocated into the host organisms via the bacterial type III secretion system [185]. After entering the cytoplasm, they bind to the promotor sequences of a subset of genes and increase expression [111]. One of

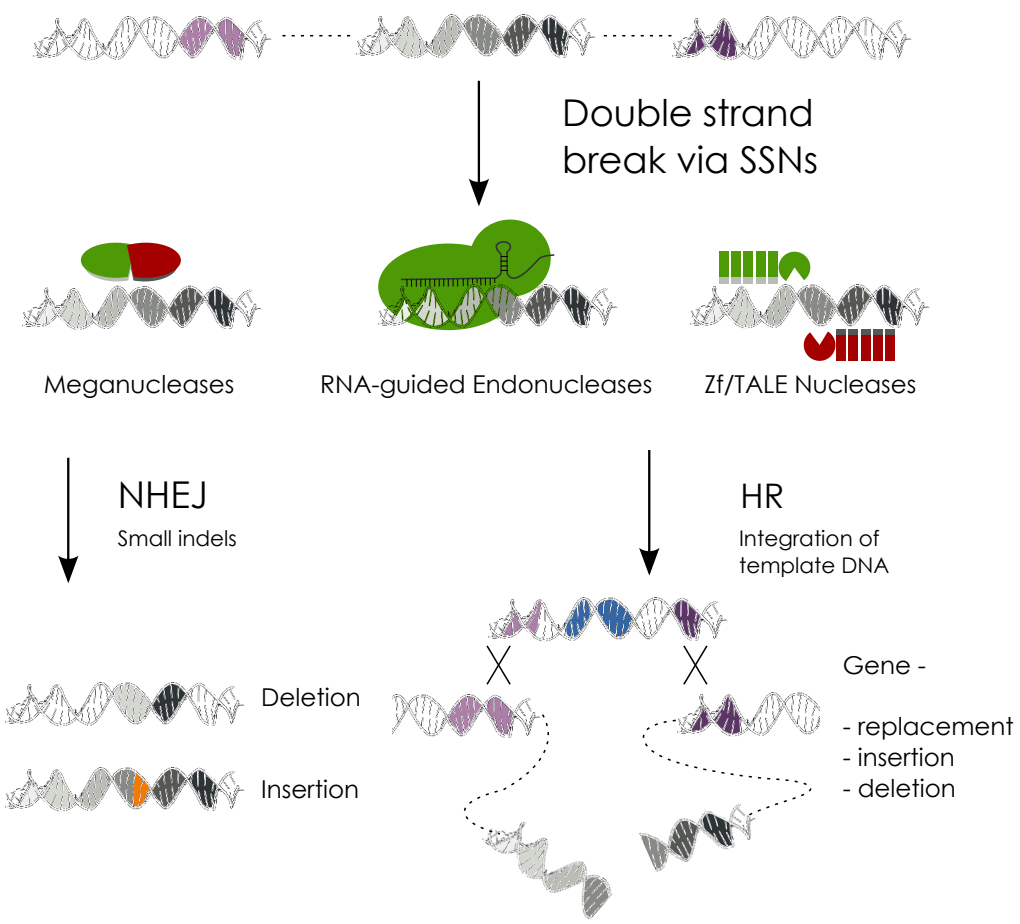



Figure 1: Site-specific double-strand breaks can be used for genome editing
A DNA double-strand break is introduced by site-specific nucleases, such as meganucleases, RNA-guided endonucleases or TALENs. The resulting lesion can be repaired (see section 1.3) by two groups of pathways: non-homologous end-joining (NHEJ) and homologous recombination (HR). Small insertion and deletion events can occur during repair via NHEJ, while the addition of a homologous repair template permits the replacement of a whole genomic region via HR.

the best characterized examples is the effector AvrBs3 of *Xanthomonas campestris* pv *vesicatoria*, which can be seen as an example for the general mechanism of TALEs.

*X. campestris* pv *vesicatoria* infects pepper plants (*Capsicum annuum*) and induces transcription of *UPA* (up-regulated by AvrBs3) genes. The transcribed gene products cause the enlargement (hypertrophy) of the plants mesophyll amongst other effects [141]. Pepper variants carrying the bacterial spot (Bs) resistance gene *Bs3* are unaffected. Interestingly, the promotor region of *Bs3* contains an effector binding element (EBE), which causes AvrBs3 itself to induce the immune response [28]. This is why this effector was first termed avirulence (Avr) protein AvrBs3. Other adaptations against TALE proteins include mutations in the host promotor regions.

It has also been shown that TAL-effector proteins are present in the beta proteobacterium *Ralstonia solanacearum,* which causes bacterial wilt in different species of plants [56]. These *Ralstonia* injected protein TAL (RipTAL) proteins share a similary code, but show some structural differences compared to *Xanthomonas* TALEs (see subsection 1.1.1.1). Cryptic TALEs, which can be engineered to bind DNA, can be found in *Burkholderia rhizoxinica* [104, 203].

#### 1.1.1.1 *Structure*

TAL effectors are comprised of a central sequence of tandem repeats; an N-terminal type III secretion sequence, a C-terminal nuclear localisation signal (NLS) and a C-terminal acidic transcriptional activation domain (see figure 2). The central sequence is responsible for DNA-binding and possesses a characteristic α-solenoid fold. Crystal structures for PthXo1 [140] and AvrBs3 [204] show that the DNA binding domain forms a super-helix following the major groove. Effectors contain between one and 34 repeats, though no function has been shown for the tiniest TALEs (<6.5 repeats) [25]. Typical repeats comprise between 33 and 36 amino acids (aa), which form a hairpin-like structure made up of two α-helices (small and large). Larger, non-canonical repeats, that contain duplications of the small or large helix, can also be found in nature. These duplications seem to reduce TALE activity slightly, but allow "skipping" of a mismatching base pair by turning single repeats outwards [182]. It is thought that this is a counter-adaptation against mutation of the EBE.

The amino acid sequence is highly conserved with the exception of position 12 and 13, located in the loop connecting the two helices. Amino acids at these positions are hyper variable and are referred to as repeat variable diresidue (RVD)s. It has been shown that these RVDs determine the DNA sequence specificity of each repeat, creating a "one repeat - one base" relationship. This relationship can be expressed with a simple code (see figure 2) [26, 156]. Commonly, the diresidue "NI" is used to target adenine (A); "HD" for cytosine (C);

"NN" for guanine (G) or adenine (A) and "NG" for thymine (T). Studies have shown that a single asparagin at position 12 (referred to as "N*" RVD) can offer a benefit when binding methylated bases [224]. Alternative RVDs like "NK" are more specific for guanine, but appear less often in nature [156] and seem to impair binding negatively [47]. The extended TALE code accommodates even more RVDs, however the more exotic ones play less of a role for biotechnological applications.

The N-terminal region prior to the first TALE-repeat is involved in DNA binding and conveys specificity to thymine or in case of RipTALs to guanine. It is therefore referred to as "To" region. This region consists of four degenerate repeats ($N_{-3}$, $N_{-2}$, $N_{-1}$ and $N_0$) which exhibit structural similarity but deviating amino acid sequence and smaller size. It has been shown for the AvrBs3 scaffold that TALE-binding domains lacking this region ($\Delta$N-terminus > 152aa) have reduced affinity to DNA or result in TALENs with lowered activity [151]. It had already been shown, that these residues were only needed for type III secretion [211]. N-termini with changed specificities have been engineered by directed evolution experiments [122]. A tryptophan at position 232, located in the loop in repeat $N_{-1}$, seems to play an important role in the specificity of the N-terminus. An alternative to using an engineered N-terminus is a substitution with a guanine-targeting RipTAL N-terminus.

The last TALE-repeat is an incomplete "half-repeat". To date, there is contradictory information about the importance of the last repeat. There seems to be little to no contribution to binding in some cases [249, 142], while other groups report a significant or overproportional impact of this last repeat [229, 27]. It has to be noted that a variety of C-terminal truncations and linkers are functional in the TALEN context [151].

#### 1.1.1.2 *Target binding and recognition*

Crystal structures suggest, that TALE DBDs undergo a conformational change when binding to DNA. The superhelix is being compressed so that the helical-pitch length is reduced from 60 Å to 35 Å [158, 59]. Several studies measured affinity to DNA, however different dissociation constants were observed [79, 142, 106]. Recent single-molecule analysis revealed that TALEs find their target sequence in a two-step process [51]. The first part seems to be a classic example of facilitated diffusion ("sliding/hopping"). The DBD binds non-specifically and then moves one dimensionally along the DNA ("1D sliding"). Interestingly, longer TALEs with more repeats move slower during this 1D sliding. This movement is combined by quick dissociation and rebinding in the near vicinity ("hopping"). It is proposed that the TALE adopts a looser conformation while searching and is then compressed upon binding of the EBE [51]. Although specificity of single repeats is conferred via the RVDs, the main energetic contributions for

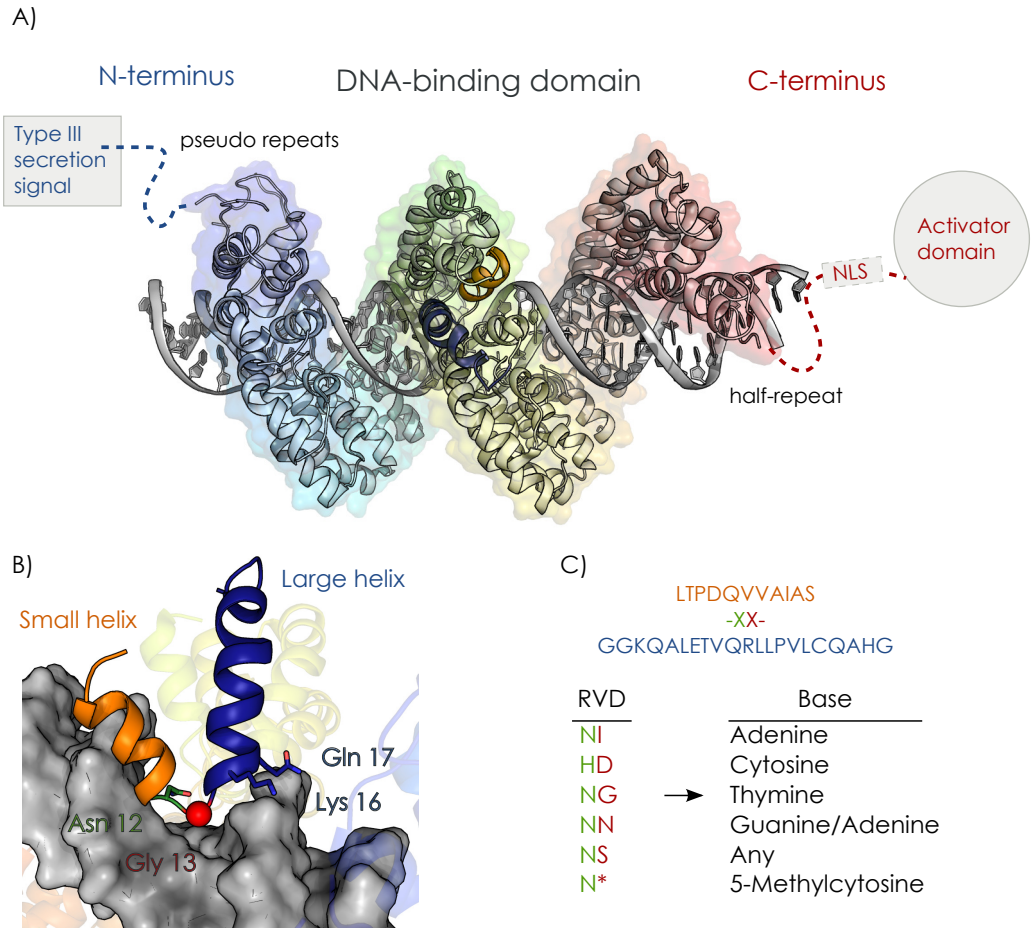



Figure 2: Structure of a TAL-effector

A transcription activator-like effector protein (A) consists of a type III secretion signal, a DNA binding domain (DBD of effector PthXo1 is given as an example, [140], PDB-ID:3UGM), an NLS and an activator domain. The DBD forms a superhelix following the major groove. It exhibits a characteristic α-solenoid topology, that consists of small repeat units. The N-terminal pseudo repeats are not well-resolved in this crystal structure. Each repeat (B) is made up of a small (orange) and a large helix (blue), linked by the repeat variable diresidue (RVD, green/red). Lysine 16 and glutamine 17 are clamping the phosphate backbone, while residue 13 (red, here gly) determines the base specificity. Shown is an example sequence for a 34 aa repeat (C). The RVD at position 12/13 determines which base can be bound. This relationship can be expressed in a simple code.

specific DNA binding seems to stem from the residues G14, K16 and Q17 [236]. Furthermore, molecular dynamic simulations suggest that the free energy contributions of amino acid 13 binding to its respective nucleotide is lower than its binding contribution to the previous nucleotide. The other RVD residue, amino acid 12, does not seem to interact with the DNA at all. One model postulates that residues G14, K16 and Q17 clamp the phosphate backbone and orientate residue 13 towards the target base [236]. Specificity is then obtained via negative discrimination of non-fitting base pairs, that means residue 13 creates steric or electrostatic clashes with wrong bases. In this model, residue 12 is thought to act as a helix breaker for the α-hairpin motif.

#### 1.1.1.3 *Beyond nuclease domains*

Since TALE proteins contain an activator domain and act as activators of expression, it is no surprise that they have been successfully applied in a variety of cell types. One example of a working activator fusion is TALE::vp16 (or vp64), with an activator domain derived from the herpex simplex virus. It has been shown that this construct can activate genes in mammalian cells [245]. The next logical step was the fusion of repressor domains (e. g., KRAB) to reduce or block gene expression of a target gene [80]. Activator and repressor TALEs have been colorfully used to create logical gates [199], bistable switches [124] and even more complex genetic circuits [133, 200]. The range of DNA-modifying enzymes also includes CpG-methyltransferases [18] and demethylases [138, 130] to examine epigenetic connections. Fusion of histon-modifying enzymes to study chromatin remodeling is also interesting in this context [146, 40]. But also inactive marker proteins have been used [152]. Marker-fusions act as fluorescent probes and help to visualize the location of certain DNA sequences in a cell. Till now, this application seems to be limited to repeating sequences [152].

Another interesting fusion partner for TALE-DBDs are recombinases and transposases [147, 164]. These proteins can allow the direct integration of a template, bypassing the cells DSB repair machinery and without the creation of potentially toxic DNA lesions (see section 1.3). Currently, off-target integration is however still a problem for these enzymes [147, 164].

### 1.1.2 *Catalytic domains*

Nucleases have many different functions in nature, such as defense against invading nucleic acids, processing of RNA, maintenance and recombination of DNA, scavenging of extracellular DNA or cytotoxicity [240]. Catalysis works through an acid-base mechanism. The general base deprotonates a nucleophile which usually attacks the scissile phosphate of the backbone in an $S_N2$ manner [168]. Suitable for TALE

fusion are endonucleases that accept dsDNA as substrate. Catalytic domains used in this work (see figure 3) can be divided into groups depending on the amount of divalent metal ions used for catalysis and structural similarities [241]. FokI and I-CreI both need two metal ions for catalysis. FokI carries a PD-(D/E)XK motif, which is typical for many type II restriction endonuclease (RE)s [169]. I-CreI is a highly-specific homing endonuclease of the LAGLIDADG family. The other nucleases in this work catalyze phosphodiester cleavage with only one metal ion. ColE7, NucA and EndA are non-specific nucleases and share a ββα-metal motif. In this family, a β-hairpin is connected to an α-helix. A histidine at the end of the first β-strand activates a water molecule which serves as the attacking nucleophile, while the metal ion is thought to stabilize the $S_N2$ transition state (see figure 4) [233]. I-TevI is a homing endonuclease of the GIY-YIG family. Although the motif is also located on a β-hairpin followed by an α-helix, the catalytic center differs from the ββα-Me family [206, 241].

#### 1.1.2.1 *FokI*

FokI is a restriction endonuclease found in *Planomicrobium okeanokoites* [91]. It is no surprise that FokI was the first catalytic domain for TALENs [42], since it has already been applied in the context of zinc-finger nucleases [114] and even meganucleases [134]. Hence, it is not only well characterized as an endonuclease, but also as a fusion domain [195, 170]. More active derivatives of FokI, like the "*Sharkey*" variant have also been engineered [88].

FokI is a member of type IIS restriction endonucleases [169], which means it cleaves its substrate at a defined distance outside of its recognition sequence ($5'$GGATG(N)$_9$▼NNNN▲$_{13}$ (▼ for top-, ▲ for bottom strand cleavage)). The reason for this is the bipartite layout of FokI [228]. There is a helix-turn-helix related DNA-recognition domain, consisting of three smaller subdomains, at the amino terminus and a DNA-cleavage domain at the carboxy terminus. The cleavage domain possesses a PD-(D/E)XK motif and uses $Mg^{2+}$ as a cofactor. It is non-specific, but contains only one catalytic center, forcing transient dimerization of two catalytic domains to effect a DNA-DSB [227, 23]. Dimerization of the catalytic domain can occur with just one DBD binding [89] and also in trans [123]. Since cleavage and recognition are separated, the DNA-recognition domain can be replaced with another specific DNA-binding domain, effectively changing the preferred substrate of this enzyme. Interestingly, two DBDs are needed to create a DSB, when the natural DBD has been replaced by a zinc-finger or a TALE-DBD. FokI is commonly fused to the C-terminus of a TALE-DBD, however N-terminal fusion is also possible, allowing cleavage by N/N dimers and N/C heterodimers [105]. In this work, TALE::FokI is used as control nuclease but also as a fusion partner for dual catalytic TALEN (dcTALEN)s.

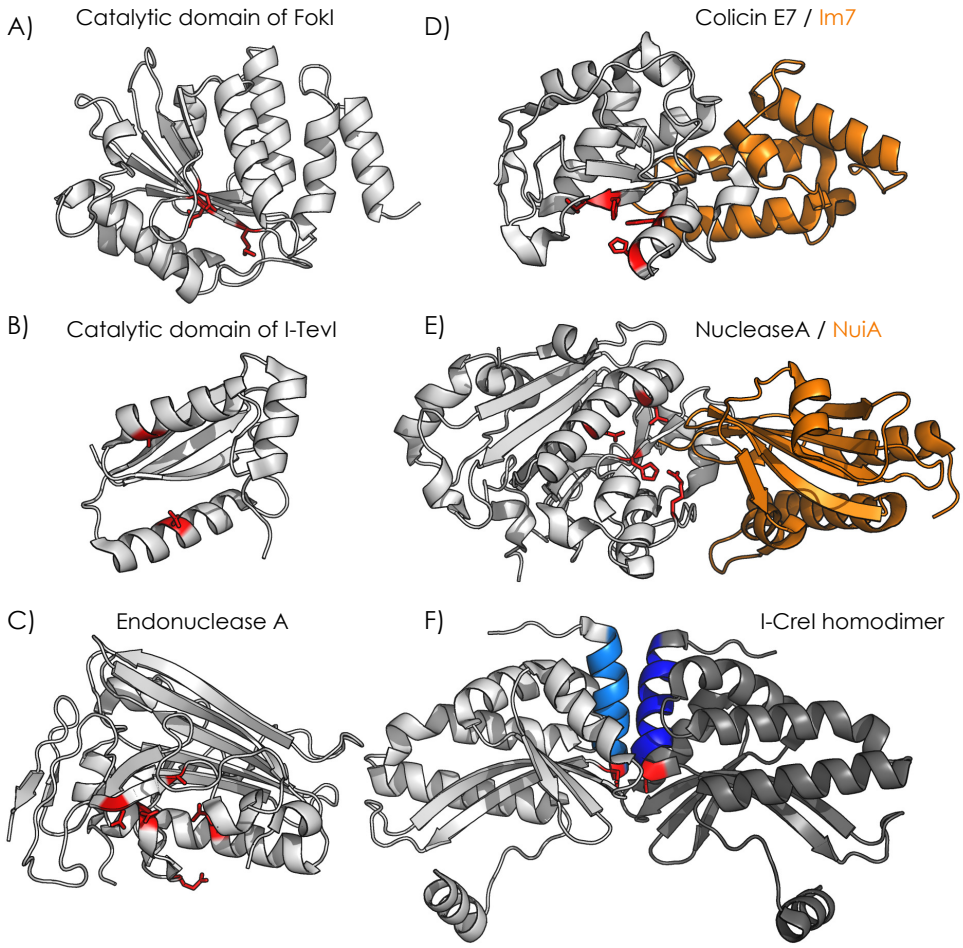



Figure 3: Overview of catalytic domains

Shown are crystal structures for all nucleases used in this work. Catalytic residues are displayed as sticks and colored in red. FokI (A, [227], PDB-ID: 2FOK) is the most used nuclease domain for ZFNs and TALENs. Catalytic residues shown are D450, D467 and K468. The catalytic domain of homing endonuclease I-TevI (B, [225], PDB-ID: 1LN0) is comparatively small. Its linker, consisting of a DNA-binding helix and a zinc-finger, are not shown in this crystal structure. Catalytic residues are E75 and R27, which was mutated to alanine for crystalization. A detailed look into the cleavage mechanism of Endonuclease A (C, [154], PDB-ID: 3OWV) can be seen in figure 4. Four histidines (H544,H545,H569 and H573) mediate cleavage by the nuclease domain of Colicin E7 (D, [207], PDB-ID: 1MZ8). Its inhibitor Im7 (orange) binds an exosite allosterically. NucA (E, ref, PDB-ID: 2O3B) on the other hand is inhibited competitively. The active site is similar to the one of EndA and consists of H124, R93, N155 and E163. It is blocked here by the inhibitor NuiA. I-CreI (E, [171], PDB-ID: 2O7M) is a dimer that is coordinated by the eponymous LAGLIDADG helices (blue). The catalytic center (D20 D20') is located at the bottom of the helices.

#### 1.1.2.2 *I-TevI*

I-TevI, which was discovered in the enterobacteria phage T4, is a member of the GIY-YIG family of homing endonuclease (HE)s [17]. These "selfish" enzymes are highly-specific endonucleases, which propagate their own intron-localised coding sequence into cognate alleles lacking this particular intron [206]. It shares its bipartite layout of discrete recognition and cleavage functions with FokI, but domain orientation differs. The catalytic domain of I-TevI is located at the amino terminus and it is connected via a linker domain to a carboxyterminal recognition domain [118]. However, the main differences to FokI lie in the catalytic domain. Contrary to FokI, the catalytic domain of I-TevI has an intrinsic 5′ CN▲NN▼G 3′ specificity and leaves a 2 bp 5′-overhang. Recent studies further suggest that the linker domain, which consists of a zinc finger and a major-groove-binding helix, conveys an additional preference for specific target sites [115]. While FokI needs to dimerize, I-TevI is thought to hydrolyze the phosphodiester bond of one strand first, rotate around the DNA and then hydrolyze the other one, similar to what has been shown for related GIY-YIG HE I-BmoI [116]. The catalytic core consists of a three-stranded antiparallel β-sheet, that is flanked by three helices [225]. Recently the catalytic domain has also been used for fusion constructs with zinc-fingers [117] and with the LAGLIDADG HE I-OnuI ("MEGATEV") [238] as DNA-binding domains . In the case of MEGATEV, two cuts are made, similar to the dualcatalytic TALENs in this work. It is therefore not surprising that I-TevI can also be used as a nuclease domain for TALENs [21, 115].

#### 1.1.2.3 *Colicin E7,*

Colicin E7 (ColE7) is an unspecific endonuclease, cleaving double-strand (ds)- and single-strand (ss)DNA and can be characterized by its histidine-asparagine-histidine (H-N-H) motive. ColE7 has no sequence specificity, but prefers to cleave by creating nicks 3′ after deoxythymidine [233].

It is part of a group of *E.coli* exotoxins known as colicines. Colicines are expressed to kill non-self cells during stress and are functionally diverse [29]. Two DNA endonucleases from this group are ColE7 and ColE9. They share structural similarities and their respective inhibitors can be adapted to block the other nuclease [127]. The nuclease domain of ColE7 is connected to a receptor-binding and a translocation domain, which transport it through the bacterial periplasm. Both domains are later separated from the catalytic domain by proteolytical cleavage between residues K446 and R447, which is also involved in binding of the phosphate backbone [194]. The active site is made up of four histidines. It is thought that H544, H569 and H573 are coordinating the metal ion, while H545 activates the nucleophile

[92]. $Zn^{2+}$as well as $Mg^{2+}$can serve as divalent metal ions. For ease of reading, "ColE7" and "Colicin E7" in this work will refer to the N-terminal catalytic domain of Colicin E7.

Because of its toxicity to the producing organism, ColE7 is blocked by the allosteric inhibitor Im7. This inhibition is one of the tightest protein-protein interactions with an affinity in the femtomolar range [112]. In contrast to I-TevI and FokI, there is no recognition domain and ColE7 is fully active as nuclease domain. Correspondingly, its activity and affinity for DNA are not preadapted for a fusion protein. Interestingly, use of the ColE7 catalytic domain for zinc-finger fusion has been theorized recently and potential constructs have been designed *in silico* [160].

#### 1.1.2.4 *NucA*

NucA is a sugar-non-specific endonuclease that is found in the cyanobacterium *Anabaena sp.* strain PCC7120 [159]. These nucleases can cut both DNA and RNA sequence independently [72]. NucA is secreted and degrades extracellular nucleic acids, probably for nutrient scavenging [143]. Intracellular toxicity is avoided via the inhibitor NuiA. In contrast to ColE7/Im7, inhibition takes place competitively at the active site[84]. There is no sequence specificity, but NucA avoids d(A) or d(T) tracts [143]. It is a member of the ββα-Me superfamily and includes a characteristic DRGH motif. In the proposed mechanism, histidine 124 acts as general base and activates the attacking water molecule [85]. Arginine 93 helps position the scissile phosphate, while glutamate 163 and asparagine 155 coordinate the stabilizing metal ion. NucA is more active with $Mn^{2+}$or $Co^{2+}$than with the more common $Mg^{2+}$ [144].

#### 1.1.2.5 *EndA*

The nuclease EndA was found in *Streptococcus pneumoniae*, where it is a virulence factor during host infection [121]. It is attached to the cell membrane and exposed to the outside of the cell [173] and it is thought to degrade the DNA scaffold of neutrophil extracellular traps [15], which are part of the mammalian innate immune system.

EndA shares a structural similarity to NucA [154]. Both contain a DRGH catalytic motive and share the characteristic ββα-Me fold. Wildtype EndA cannot be reliably expressed in *E. coli* due to its toxicity. A way to circumvent this is the use of a H160G variant, which replaces one of the catalytic histidines by glycine [148]. Although this variant is inactive by itself, activity can be restored through chemical rescue with imidazole. An overview of the catalytic center and relevant catalytic residues can be seen in figure 4. A number of mutations affecting DNA-binding and catalysis have also been identified [149] and were used in this work.

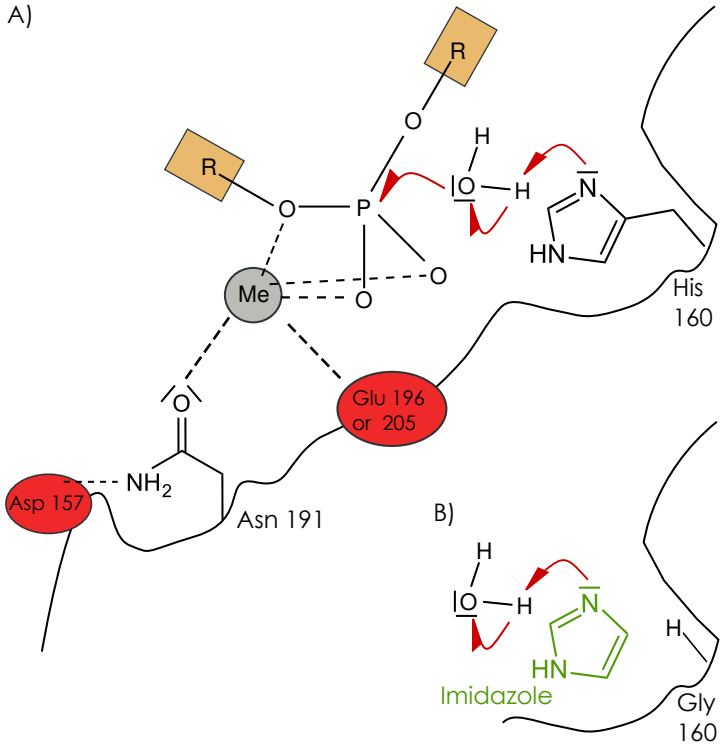



Figure 4: EndA H160G activity can be rescued chemically
Shown is the proposed catalytic mechanism of EndA (A), which is similar to that of H-N-H nucleases. Histidine 160 activates a water molecule to act as the nucleophile for the $S_N2$ in-line attack on the central phosphorous atom. The negatively charged transition state is stabilized by a metal ion, which is coordinated by asparagine 191 and a glutamate. Aspartate 157 is thought to interact with N191 to facilitate metal binding. For members of the H-N-H family, asparagine 191 is replaced by a histidine and an asparagine coordinates the two histidines.
Mutation to H160G creates a space that can be filled by imidazole (B). Imidazole can replace the histidine and activate the catalytic water.

#### 1.1.2.6 *I-CreI*

I-CreI, together with I-SceI and I-AniI, is probably the best characterized member of the LAGLIDADG family of homing endonucleases [38, 171] (see section 1.2.2). It is a dimer with a size of 18 kDa per subunit. The natural target-site is a 22 bp pseudopalindromic sequence (TCAAAACGTCGTGAGACAGTTTGG, [107]) in the 23S rRNA gene of the chloroplast genome in the green algae *Chlamydomonas rheinhardtii [218]*.

*C. rheinhardtii has been selected as "Algae of the year 2014" by the german botanical society*

Mutagenesis of I-CreI can change the prefered recognition site [192, 184]. The target sequence can be grouped into smaller blocks [202, 6] that are recognized almost discretely. Consequently, one can alter residues recognizing one block, without necessarily changing specificity for the other blocks. The cleavage site is located in the GTGA motif in the middle (-2...2). Other blocks are the 10 to 8, 7/6 and 5 to 3 regions [202, 6]. Obligate heterodimers of I-CreI need to be constructed, since most target sequences are not palindromic. This can be achieved by generating single-chain fusions [64] or by mutating residues of the LAGLIDADG helix [196]. E. g., lysine 7 on top of the helix interacts with glutamate 8′ in the other subunit. Homodimerisation is greatly reduced by changing residues 7 and 8 to the negatively charged glutamate and residues 7′ and 8′ to the positively charged arginine. A prominent example of a successful reengineering of I-CreI is a variant targeting the human *XPC* gene [8, 178], which is involved in the disease Xeroderma Pigmentosum.

#### 1.1.2.7 *PvuII, BfiI, MutH, scFokI and I-AniI*

More nuclease domains have been fused to TALE-DBDs by other groups.

- PvuII is a type IIP restriction endonuclease with a characteristic PD-(D/E)XK motif [37, 169]. It is one of the smallest REs (18 kDa) and functions as a dimer or artificial single-chain version. It cleaves a 5′ CAG▲▼CTG 3′ sequence, leaving blunt ends. TALE::PvuII fusions have to dimerize like FokI, but obligate heterodimers have also been designed [243]. The target site requirement and the heterodimerization are thought to reduce the amount of off-site activity, but the amount of targetable sites is reduced as a consequence.

- BfiI is a type IIS restriction enzyme like FokI and also requires dimerization for cleavage. DNA binding and catalytic functions are separated in different domains. BfiI is one of the few metal-ion independent DNases and uses a conserved His as nucleophile [186, 187]. The linker, which is connecting nuclease and DNA-binding domain can inhibit cleavage activity in solution. This autoinhibition could reduce off-site activity *in vivo* [242].

- MutH is a nickase, which creates single-strand nicks, from the *E. coli* DNA-mismatch repair system. It is recruited to a mispaired DNA by proteins MutS and MutL and nicks a 5′ ▼GTAC 3′ sequence [125]. MutH was fused C-terminally to TALE-DBDs and TALE::MUTH monomers were shown to work as site-specific nickases [77], similar to TALE::TEV. Drawbacks are its comparatively low activity and requirement for a GTAC sequence.

- FokI nuclease domains can also be constructed as single-chain variants. TALE::SCFOK monomers are toxic, but mutagenesis was used to reduce the activity and generate monomeric TALENs [209]. However, scFokI based constructs showed between 2- and 10-fold lower activity than their dimeric counterparts and were significantly more cytotoxic in some cases.

- I-AniI is a HE of the LAGLIDADG family. TALE fusion allowed the redirection to just one of many I-AniI target sites in the human genome. Subsequently, an engineered I-AniI variant targeting the T-cell receptor alpha-chain gene (TcrA), was successfully applied in a fusion construct *in vivo* [27].

## 1.2 TOOLS FOR GENOME EDITING

The first genetically modified organisms (GMOs) were created only shortly after the discovery of recombinant DNA technologies [98, 44]. Among the first were transformed bacteria [155, 32], but also transgenic mice [100] and plants [22]. Till now a huge variety of GMOs has been generated. Loss- and gain-of-function mutants have elucidated genetic backgrounds and therefore helped to advance medicine, agriculture and biology in general [210]. The other side are non-scientific applications. Engineered microorganisms that produce valuable metabolites like insulin, human growth hormone or chymosin [31, 86, 108] have become common place. Plants and animals have been modified to yield more or higher quality end products [70, 250, 36], to serve as bioreactors after metabolic engineering [35, 48, 13], to be more resistant to herbicides, disease or predators [223, 43, 165], to control pest populations in the wild ("gene-drive", [220, 9, 10, 14]) and interestingly also for aesthetic reasons. Blue carnations (*Moon* series, [212]), blue roses (*Suntory rose*, [109]) and fluorescent fish (*GloFish*, [157]) have been constructed, as well as bioluminescent houseplants [65]. Genetic engineering has also been applied to humans. A few hundred patients have been treated for genetic diseases like severe combined immunodeficiency, cystic fibrosis, β-thalassemia or muscular dystrophia [183, 5, 69, 145] and human T-cells are altered to be resistant to human immunodeficiency virus by deletion of the CCR5 receptor [231] or to recognize cancer cells via a chimeric antigen receptor (CAR) [61]. In 2012, a gene-therapy ("Glybera", [82]) for lipopro-

tein lipase deficiency has been approved in the European union. Recently even (non-viable) human embryos have been modified [132], which sparked controversy in the scientific community [193]. There are also integration sites of endogenous retroviruses in the human genome ("*safe harbors*") that can be used to insert transgenes [137]. Although these mutations would not be inserted into the germline, human amelioration remains of ethical concern.

Zinc-finger nucleases and homing endonucleases represent the first efforts to generate highly-specific nucleases, before the advent of TALENs and RNA-guided endonuclease (RGEN)s. Due to their ease of use, TALENs and RGENs are probably the most widely applied site-specific nuclease (SSN)s nowadays. Nevertheless, zinc-finger nucleases and meganucleases still have their specific niches. The first used methods for integration of DNA were untargeted and are also shortly mentioned.

### 1.2.1 *Zinc finger nucleases*

ZFNs are the spiritual precursors of TALENs. Both share a bipartite layout consisting of a DNA-binding domain fused to a FokI nuclease domain [114]. Both DNA-binding domains are made up of smaller units, with individual DNA specificities that form a larger array with a composite specificity. The building blocks for this array are three to six Cys2-His2 fingers. Each finger recognizes 3 bp and consists of about 30 aa in a conserved $\beta\beta\alpha$-fold [167], which explains the smaller size of ZFNs compared to TALENs. ZFNs targeting a wide range of different genes have been constructed and successfully tested *in vivo* [113, 175, 78]. However, there is crosstalk between the modules [96, 97], making their design difficult [176] and requiring approaches that take context dependence into account [139]. Another problem is increased off-target activity compared to TALENs [20, 166]. Intriguingly, it is also possible to re-engineer the zinc-fingers themselves to act as nucleases. Attachment of a lanthanide binding loop between the fingers allows cleavage in the presence of cerium ions [95].

### 1.2.2 *Homing endonucleases of the* LAGLIDADG *family*

Homing endonucleases (HEs) or meganucleases are naturally occurring endonucleases with recognition sequences between 12 and 40 bp. They can be divided into the families LAGLIDADG (I), His-Cys box and H-N-H (II) and GIY-YIG (III) [206]. There are also few examples of HEs with a PD-(D/E)XK motif [162, 247]. A trait of these nuclease is their "selfish" nature. HEs are coded in introns or inteins and a functional version of the protein is then later spliced out either on the mRNA (intron) or on the protein (intein) level. The homing endonuclease will then introduce a DNA-DSB in a cognate allel that is

lacking the coding sequence of the homing endonuclease. During homology directed repair, the intron-containing allel serves as a repair template for the damaged locus, effectively replacing its intron-less counterpart. Members of the LAGLIDADG family occur as monomers or dimers and are characterized by their αββαββα-fold. Recognition of their target is carried out by the four β-sheets, which form a saddle like structure around the DNA (see figure 3). Two central α-helices, containing the eponymous LAGLIDADG consensus, are coordinating the two subunits or subdomains, respectively. Right at the base of the helices lie the two catalytic centers, where cleavage of the phosphodiester bonds is carried out by two acidic residues. Crystal structures show three metal ions in the catalytic center. Each active site depends on two metal ions, with one metal ion being shared between the two [39, 16].

Meganucleases were among the first used for genome editing via induction of a specific DSB [41] and variants targeting a variety of genes have been engineered [7]. HEs are smaller than other SSNs and therefore easier transfected, which makes them major contenders for application in gene drive [237, 9, 220]. Their biggest drawback is the way they are adapted to a new target sequence. The whole DNA-binding region needs to be engineered via rational design and/or directed evolution [219, 99]. This prerequisite for experience, time and specialized equipment makes HEs virtually impossible to design for non-experts.

### 1.2.3 RNA-*guided endonucleases*

Clustered regularly interspaced palindromic repeats (CRISPR) and CRISPR-associated system (CAS) are parts of an adaptive immune system, which protects prokaryots against foreign nucleic acids [81]. Fragments of invading DNA are integrated into a CRISPR cluster and then later transcribed together with their spacer. The nuclease Cas9, found in the bacterial type II CRISPR systems, causes site-specific DSBs. Sequence specificity is granted by an RNA (CRISPR RNA (crRNA)) that forms watson-crick bonds with a complementary target DNA. A second RNA molecule (trans-activating CRISPR RNA (tracrRNA)) stabilizes the complex [103]. The guide RNA (gRNA), a synthetic fusion of tracr- and crRNA, is employed for biotechnical applications [93]. Due to this, Cas9 variants are also called "RNA guided endonucleases" (RGENs). Two nuclease motifs are contained in Cas9. The complementary strand is cleaved by the RuvC, the non-complimentary by the H-N-H motif [102]. It is no surprise that inactivation of one motif via mutagenesis creates a nickase [177]. Targets that can be addressed by Cas9, are limited by the necessity of a protospacer-adjacent motif (PAM). Cleavage occurs 3 nucleotids upstream of the PAM and produces blunt ends [81]. RGENs could be successfully applied in

a variety of species and even multiplexing is possible [46, 129, 235]. Retargeting the nuclease is easier than for TALENs or ZFNs, however RGENs are slightly larger, show greater off-target activity [232, 75] and are restricted by the PAM motif and limited gRNA length.

Usage of CRISPR is not limited to the creation of DSBs. Activator or repressor fusions have been reported [66, 246], inactive Cas9 can also block transcription by itself in a technique termed "CRISPR interference (CRISPRi)" [248] and selection of special PAM sequences even allows the binding of RNA [161]. Another interesting exploitation of the CRISPR system is a bioinformatic analysis (spacer oligonuclotide typing) of the CRISPR region to determine relationships of closely related species or clones [53, 226].

### 1.2.4 *TFO-linked nucleases*

It is also possible to fuse nuclease modules to a triplex-forming oligonucleotid (TFO). The nuclease is then guided to its target sequence by an oligonucleotid which forms a DNA-triplex with that sequence via hoogsten base pairs [63]. TFO-linked nucleases need to be assembled *in vitro*, given that they are DNA-protein hybrids. Another drawback is slow binding to their target sequences, allowing the possibility of off-target cleavage by the linked nuclease [63]. Therefore, TFO-linked nucleases do not play a major role for genome editing.

### 1.2.5 *Untargeted genome editing*

The probably simplest way to introduce DNA into a genome is to transport it into a cell and to look for integration. These method includes chemical or electrical transfection methods [45], micro-injection [100] but also transfection via particle inflow gun [67]. An alternative is the employment of naturally occuring vectors. The $\alpha$-proteobacterium *Agrobacterium tumefaciens* has been widely used to generate plant cultivars [163, 222]. *Agrobacterium* transfers a Ti-plasmid (tumor inducing plasmid) into plant cells via the type IV secretion system. Part of the plasmid, the so called transfer DNA (t-DNA) is then integrated into the plant genome. This t-DNA can be chosen by replacing the Ti-plasmid with a recombinant version [163].

For animals, integrative viral vectors are alternatives. Transduction by these recombinant viral vectors occurs in a "dead-end" fashion so that no new viral particles are produced after infection. This is possible by separating the essential viral genes from the packaging domain and linking this packaging domain to a therapeutic cassette [110]. The recombinant vectors are then assembled in the presence of both DNAs. Retro-, lenti-, adeno- and adeno-associated viruses have been used as vectors [76, 110, 217]. Even special properties can be exploited: herpes simplex virus can be used to specifically target

the nervous system [68]. Non-integrative viral vectors still remain important vehicles to transport SSNs and templates into a cell.

## 1.3 DNA DOUBLE-STRAND BREAK REPAIR

Double-strand breaks (DSBs) are highly toxic DNA lesions which can lead to aneuploidy or other genetic aberrations if left unrepaired. One of them being the recently described chromothripsis [205] (thripsis (greek) meaning "shattering"). During chromothripsis a singular event fragments one or more chromosomes that are then incorrectly reassembled, leading to catastrophic outcomes like cell death or even cancer. DSBs can be caused by ionizing radiation (IR) or genotoxic chemicals, but also endogenously, such as replication over a single-strand nick and subsequent replication fork collapse [54]. Not all DSBs are harmful in nature. Programmed DSBs are absolutely essential in some processes that generate genetic variability such as meiotic recombination, mating type switching, autogamy in ciliates or V(D)J- and class switch recombination in immune cells [119, 19].

DNA double-strand breaks can be repaired by two groups of pathways (see figure 5): non-homologous end-joining (NHEJ) (I) and homologous recombination (HR) (II). NHEJ and HR can be further subdivided into

- classical non-homologous end-joining (C-NHEJ) and alternative non-homologous end-joining (A-NHEJ)
- classical homologous recombination (C-HR) and single-strand annealing (SSA)

As the name implies, HR requires a homologous repair template, whereas NHEJ directly ligates DSB ends without the need of a template. All pathways compete with each other and their frequency is related to the organism, cell type and point in the cell cycle. For example, HR is much more prevalent in *S. cerevisiae* and *S. pombe* than in mammalian cells or plants [101].

Which pathway acts on the DSB is dependent on how the DNA ends are processed. Examples for mammalian functional homologs are given in brackets. In yeast, repair is initiated by the MRX (MRN) complex, formed by Mre11,Rad50 and Xrs2 that then recruits the response protein ATM or in the case of replication forks ATR (ATM related). ATM phosphorylates mediator proteins which then amplify the signal by phosphorylating other mediator proteins. This signaling cascade then "determines" the implementation of the type of repair pathway or of additional responses (e. g., cell cycle arrest or apoptosis) [54]. C-NHEJ is carried out by the heterodimer Ku and Dnl4/Lif1 (LIG4/XRCC4). It is thought that Ku physically blocks access to the DNA ends, while Dnl4 and Lif1 stabilize its binding and join the DNA. The p53 binding protein 53BP1, a protein involved in DNA damage

response, seems also to be favoring C-NHEJ. It is therefore no surprise that mutation of *Ku* or *53bp1* reduce the ratio of C-NHEJ [60].

Antagonistically to this end-protection is exonucleolytic resection, initiated by end-binding of Parp1 and carried out by endonuclease Sae2 . DNA binding proteins CtIP and BRCA1, which is known for its role in breast cancer, replace Sae2 in mammalian cells. Interestingly, lethal *Brca1* mutants can be rescued by *53bp1* mutation, presumably by removing end-protection [119]. Resection can also occur via exonuclease Exo1 (EXO1) or Sgs1 and Dna2 (BLM). Resected DNA ends can then be substrates for C-HR, SSA or A-NHEJ [60].

### 1.3.1 *Non-homologous end-joining*

#### 1.3.1.1 *Classical Non-homologous End-Joining*

As mentioned before, classical NHEJ (C-NHEJ) is reliant on the end-protection of DSBs. After Ku70/Ku80 heterodimers bind the DNA ends, they recruit DNA dependent protein kinase DNA-PKcs. Subsequently Artemis, polynucleotidkinase (PNK) and polymerase X (PolX) process the DNA until it is ligated by ligase IV and its cofactors (Xrcc44 and XLF). Although C-NHEJ is used to generate genetic variability, it is thought to be mostly error-free for blunt and cohesive DNA ends [19]. Examples are the aforementioned autogamy in ciliates, where thousands of DSBs are repaired in a short time frame, and the creation of genetic diversity in immunoglobulin genes via V(D)J recombination. In the later, recombinases RAG1 and RAG2 create DSBs and cap the ends with a hairpin. This hairpin is resolved in different ways by Artemis and variable nucleotids are attached by the terminal deoxynucleotidyl transferase (TdT). The resulting ends are then ligated via C-NHEJ, which has to be able to tolerate non-complementary ends for this [120]. Other causes that generate non-complementary DNA-ends are ionizing radiation or reactive oxygen species. Depending on the type of overhang, lesions cannot be repaired without error. It can therefore be argued that the variability or mutagenicity was introduced by previous treatment of the DNA and not by C-NHEJ itself.

#### 1.3.1.2 *Alternative Nonhomologous End-Joining*

A-NHEJ, also termed backup NHEJ (B-NHEJ) or microhomology mediated end-joining (MMEJ), is different from the aforementioned C-NHEJ in several ways. It was suggested that A-NHEJ utilizes ligase III and Xrcc1 instead of ligase 4 and Xrcc4 [19]. Moreover, it is dependent on the same processing enzymes as HR. Therefore, A-NHEJ is always mutagenic in nature since it ligates resected DNA ends. It is thought that resection of single-strands uncovers microhomologies (0-10 bp) which can be annealed and ligated, after gaps are filled and flaps are removed [126]. Nevertheless, if A-NHEJ is one or a group of pathways

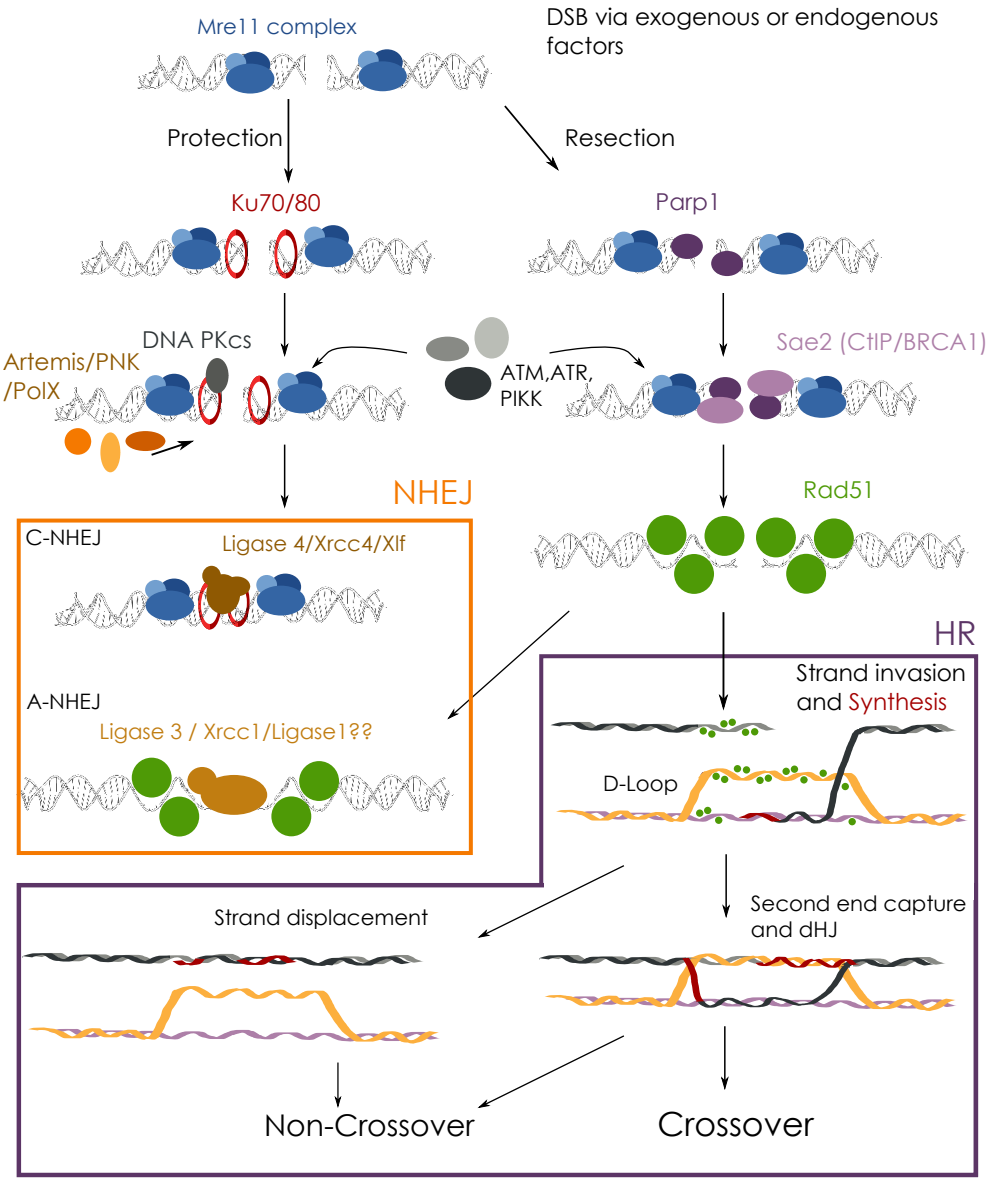



Figure 5: Repair pathways

The Mre11 complex (MRN/MRX, blues) binds first to a DNA-DSB. Which pathway is used for repair depends on the proteins that are recruited next. Ku (red) will protect the DNA, while Parp 1 (purple) will recruite exonuclease Sae2 (light purple) to resect the ends. Some end modification via Artemis/PNK and Pol X (brown) can also occur after end protection . Several mediator proteins (grey) are recruited next. Ligase 4 with its cofactors (brown) is joining the dna in the case of C-NHEJ. Rad51 (green) binds the resected ssDNA, which is then ligated via Ligase 1 or 3 (brown) in A-NHEJ or is further processed in homologous recombination. Here a strand invades the homologous strand forming the D-Loop heteroduplex. The homologous strand acts as a template for synthesis (red). Then the invading strand can be moved out during strand-displacement, which leads to non-crossover products. An alternative is the capture of the second end, creating a dHJ. This dHJ can then be resolved to yield Non-Crossover or Crossover products.

remains unclear [73]. As does its exact biological function. A-NHEJ is robust and can be used to ligate incompatible 5′ and 3′-ends, which would be an advantage over C-NHEJ [120]. Interestingly, A-NHEJ has been discovered in *E. coli* [34], which is deficient in components of C-NHEJ like Ku, suggesting that it could be an evolutionary older pathway[19].

An intriguing application for A-NHEJ and TALEs is the correction of trinucleotide repeat expansions. Here trinucleotide repeats reach abnormal lengths and cause a variety of severe disorders. A cut inside these repeats can cause their contraction via A-NHEJ [181, 180].

### 1.3.2 *Homologous Recombination*

#### 1.3.2.1 *Classical Homologous Recombination*

A defining step of HR is the invasion of a 3′-single-strand into a homologous duplex. To generate single-stranded overhangs, exonucleolytic resection of DNA ends is needed. This happens as mentioned before by Parp1 binding and resection via Sae2 (CtIP and BRCA1). Next, strand-exchange protein Rad51 (RecA in *E. coli*) binds ssDNA and stretches it. Binding of Rad51 is regulated by a complex network of regulator proteins that get sumo- or phosphorylated in the process. The Rad51-ssDNA nucleoprotein filament can then form Watson-Crick interactions with the homologous duplex DNA, creating a heteroduplex structure called the "D-loop" [101]. Rad51 now binds dsDNA and thereby stabilizes this complex. The invading 3′-end then acts as a template for DNA synthesis and thereby extends the D-loop[101]. Several polymerases ($\delta$,$\eta$ and $\varkappa$) can extend the strand, however this process is not well characterized [191]. Now repair can take two different paths that result in crossing over or non-crossing over events [119]. In the double-strand break repair (DSBR) pathway, the second end is captured by the D-loop, resulting in a double Holliday Junction (dHJ). Subsequently to DNA synthesis, these junctions can then be resolved to generate crossover or non-crossover products.

The second pathway is called synthesis-dependent strand annealing (SDSA). Here, the nascent strand is displaced and anneals to the other 3′-overlap of the original strand. Therefore, SDSA exclusively produces non-crossover products. Strand displacement can also occur after formation of the dHJ [119].

A special case of homologous recombination is meiotic recombination. Endonuclease Spo11 induces a DSB that is processed by the meiosis-specific proteins Mei5 and Sae3. Single-strand DNA is then bound by Rad51 paralogue Dmc1, which is also specific for meiosis. Although, defects in Dmc1 and Spo11 are linked to infertility, model organisms like *D. melanogaster* and *C. elegans* lack Dmc1. The subsequent repair process is biased towards DSBR and also towards crossover events [101].

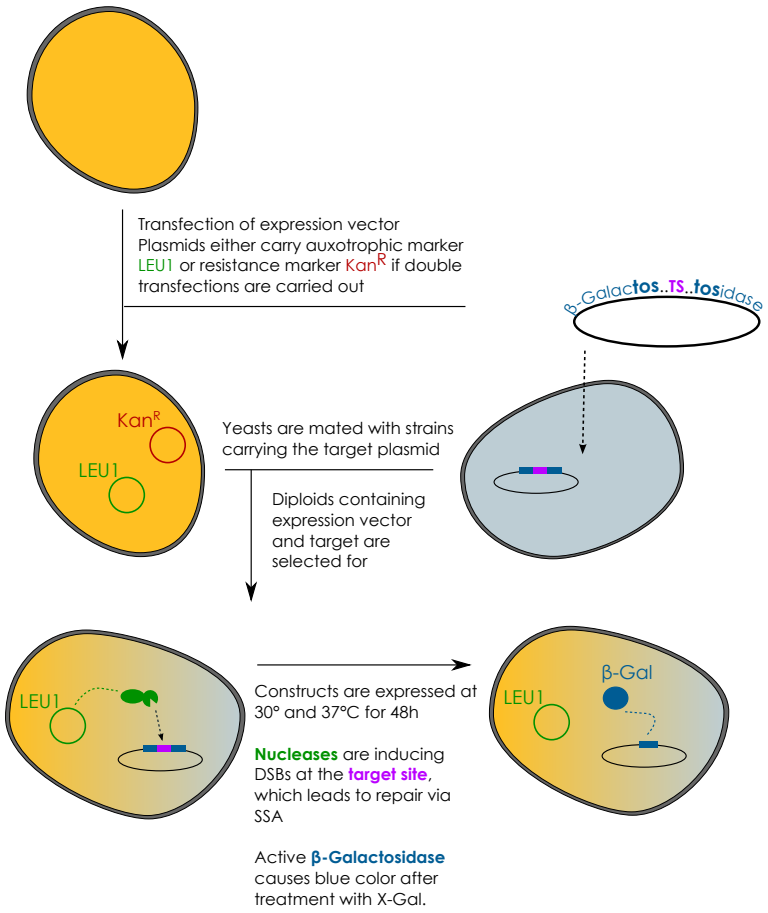



Figure 6: SSA-assays in yeast can be used to assess nuclease activity
SSA-assays are explained with the example of yeast assays used in this work. In this assay, cells harboring the nuclease plasmid were mated with cells containing the target plasmid. A disrupted β-*galactosidase* gene served as a reporter gene.

#### 1.3.2.2 *Single Strand Annealing pathway*

Single-strand annealing (SSA) is independent from the single-stranded DNA (ssDNA)-binding Rad51. Exonucleolytic end-resection uncovers homology regions that can bind each other. Flaps are then excised and the strands are ligated together [131]. Genetic information is lost during this process, similar to the end-joining by A-NHEJ.

A prominent application for this pathway is the SSA-assay [6] to profile site-specific nuclease activity *in vivo* (see figure 6). SSNs are expressed in cells harboring a target plasmid. The target plasmid contains a reporter gene, which is disrupted by a nuclease recognition site. This recognition site is flanked by homologous sequences of the reporter gene, which allows these regions to anneal after induction of a DNA-DSB and subsequent 5'-end resection. Functionality of the reporter is restored, after deletion of the disrupting recognition site and ligation of the ends. Observed amount or activity of the reporter is an indirect measurement of the nuclease activity.

## 1.4 AIM

Construction of ZFNs was focused on the DNA-binding domain, due to the inherent difficulties in designing a functional and highly-specific zinc-finger. FokI presented a suitable nuclease module to gain more specificity via its dimerization requirement. While this strategy is valid for the small ZFNs, it can limit the applications of the larger TALENs. Moreover, TALE-DBDs are easier to design and proved to cause less off-target cleavage than their predecessors [20, 166]. Therefore TALENs now present the opportunity to focus on the nuclease domain.

Aim of this work is the design of monomeric TALENs to replace or complement FokI-TALENs. Investigation of these monomeric TALENs can help understand the TALE-DBD better, since the nuclease activity is not a composite of two proteins, but originates from one singular construct. Additionally, proof-of-concept models with diverse nuclease domains can grant valuable insights into the potential and the limits of the TALE scaffold. These insights can lead to general guidelines on what type of fusions are sensible or which nuclease domains are suitable for which task. Another advantage of novel catalytic domains are special features of the domains, like control by an inhibitor protein or chemical activation.

To get a broad overview, a range of different nucleases are to be adapted via rational design. They need to be tested in plasmid based SSA-assays and also biochemically. Finally it has to be seen, if the new constructs are active on a chromosomal locus and if they have advantages over a FokI-TALEN.

# Part II

# RESULTS AND DISCUSSION

# 2 RESULTS

## 2.1 OVERVIEW

In this work, different catalytic domains were fused to a variety of TALEN scaffolds based on the AvrBs3 effector. Part of the experiments were carried out in the *Computational Biology* department of CELLECTIS, the other part was carried out at the *Institute for Biochemistry* at the JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN.

To cover different types of catalytic domains, members from three groups were selected and fused. The chosen were ranging from promiscuous H-N-H and DRGH nucleases (I) over the more restrained I-TevI nuclease domain (II) to the faithful homing endonuclease I-CreI (III) (see table 1).

Table 1: Catalytic domains
Shown are catalytic domains used in this work to create novel TALE nucleases. ColE7, nucA and EndA are unspecific H-N-H nucleases. I-TevI is a member of the GIY-YIG homing endonucleses and possesses a CNNNG specificity. LAGLIDADG HE I-CreI recognizes a 22 bp target sequence.

| NAME | COMMENTS |
|---|---|
| ColE7 | N-terminal (nuclease) domain of *E. coli* exotoxin Colicin E7 |
| NucA | Nuclease A of *Anabaena* |
| EndA | DNA-entry nuclease of *S. pneumoniae* |
| I-TevI | Nuclease domain of HE I-TevI from *enterobacteriophage T4* |
| scI-CreI | single-chain of HE I-CreI of *C. reinhardtii* |
| scI-CreI_DS | scI-CreI with degenerate scaffold mutations |
| scTcr_DS | engineered scI-CreI targeting T-cell receptor β-chain gene; degenerate scaffold |

"AvrBs3" was abbreviated "Avr" sometimes to keep construct names and target sequences short. We used a truncated TALE-DBD with an N-terminus of 136 aa (or ΔN152) and various C-termini (see figure 7). The most common length was 11 aa after the half-repeat (or ΔC220); 28 aa for EndA::TALE, 40 aa for the nptII::FokI controls. In the case of EndA::TALE::EndA, we connected a FokI-based linker to the last half-repeat. See appendix A.2.1 for the protein sequences). Due to the

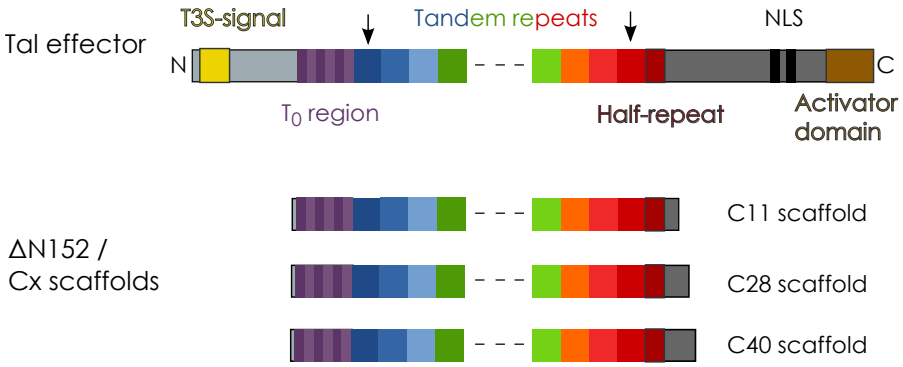



Figure 7: TAL effector and scaffolds

Layout of a *Xanthomonas*-based TAL effector. The TALE consists of an N-terminal type III secretion signal (yellow), a C-terminal acidic activator domain (brown), C-terminal nuclear localization signals (black) and a central DNA-binding region. The central region can be separated into: (I) the pseudo repeats forming the To region (purple) and (II) the 34 aa repeats (blue to red, arrows) including the incomplete, last half-repeat (darkred). Scaffolds used in this work are based on the AvrBs3 effector and were truncated, leaving 136 aa at the N-terminus and 11,28 or 40 aa at the C-terminus.

modularity of TALENs, RVD arrays can be easily exchanged to target other effector binding elements.

One important way to benchmark nuclease activity were colorimetric SSA-assays in yeast (see figure 6). The CELLECTIS platform transformed nuclease plasmids into yeast strains, which were then subsequently mated with yeast strains harboring the target plasmid. Diploids were then gridded on a plate (see figure 8) and expressed nucleases induced homology directed repair of the disrupted β-galactosidase reporter gene. Functional β-galactosidase cleaved 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-Gal), which consecutively caused blue color. The blueness of the colonies can be quantified and is a measure of specific nuclease activity. Values for blueness are ranging from 0 to 1, with 1 meaning full saturation. To be able to further grade nucleases, values at 8, 24 and 48 hours were taken and yeasts were grown at 30° and 37°C. Empty expression vector, lacking a nuclease, was used as a negative control and homing endonuclease I-SceI or a zinc-finger nuclease were used as positive controls.

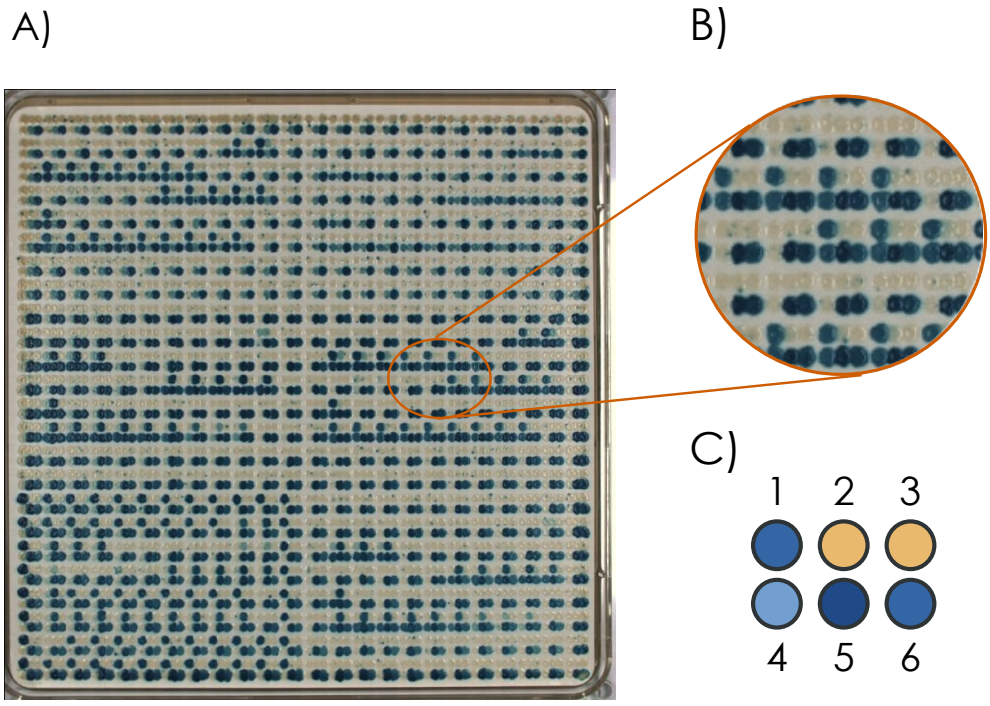



Figure 8: Nuclease activity can be measured in yeast
A standard plate (A) contains 24x24 clusters of six colonies each or 3456 colonies in total. All six colonies per cluster contain the same target plasmid, but different nuclease plasmids, which leads to different activities in different clusters (B). The amount of blue is calculated and used as a measurement for specific activity of the nuclease.
Usually, three out of six colonies per cluster are controls (C). Colonies 5 and 6 harbor the meganuclease I-SceI and a zinc-finger nuclease as positive controls. Colony 3 contains an empty expression vector and serves as a negative control. Constructs to be profiled are expressed in colonies 1,2 and 4.

## 2.2 UNSPECIFIC DOMAINS

As a starting point for monomeric TALENs we decided to look at unspecific nuclease domains. We identified Colicin E7 and Nuclease A (see table 1) as suitable nuclease domains during a preliminary screening. Both domains were then profiled in the context of a AvrBs3::CatDomain fusion in a yeast SSA-assay. Our NucA fusion construct reached relative activities between 0.6 and 0.8 , while ColE7-based TALENs stayed between 0.6 and 0.7.

Our working hypothesis was that the apparent lower activity of ColE7 was due to off-target effects and general cytotoxicity. Considering that DNA yields for the expression plasmids were also very low (~400 ng DNA / 3 ml culture $\triangleq$ 10 ng / µl eluate), we chose binding and catalytic mutants (see section 1.1.2) and benchmarked them on a group of standard targets with two effector binding elements. The provided "Avr-Spacer-Avr" targets were cleaved more or less uniformly by our mutants (see figure 9).

An exception to this was the construct AvrBs3::ColE7_H573E, which showed a surprising cleavage pattern. Relative activity reached values comparable to FokI on the targets Avr-16-Avr and Avr-17-Avr, while other targets were not cut at all, suggesting that this mutant displays a cleavage preference. The two mutants D493Q and K497A were both more active than the wild-type (WT). Though K497A displayed little off target activity, D493Q caused recombination on all off-targets. Variants R496A, H545A and H545Q were inactive; N560D and H573A showed reduced activity. Interestingly H573A cleaves targets much more uniformly than H573E. A reduction of toxicity could already be seen in *E.coli*. Inactive variants yielded the same DNA concentrations as the AvrBs3 shuttle vector (~8 µg DNA / 3ml culture), while less active variants yielded intermediate amounts of DNA.

A feature of Colicin E7 and NucA is the control of activity by their respective inhibitors Im7 and NuiA. These inhibitors allowed the construction of SSNs with an off-switch. We constructed shuttle vectors containing the sequence of the respective inhibitors upstream of our TALENs. This improved the stability of our DNA sequences and removed any toxicity observed in *E.coli*. Coexpression in yeast revealed that our TALENs can be inhibited completely (see figure 10). There was no crossinhibition and no negative effect on our AvrBs3::FokI control.

Another improvement upon our constructs was found during one of our TALE architecture experiments. For this experiment, we fused ColE7 and NucA domains between two DNA-binding domains (DBD::cat. domain::DBD). These "sandwich"-TALENs had reduced activity and were only active on a very small subset of targets (see appendix table 15). Furthermore, due to the two TALE-DBDs, which are highly repetitive by themselves, unwanted recombination of the plasmids

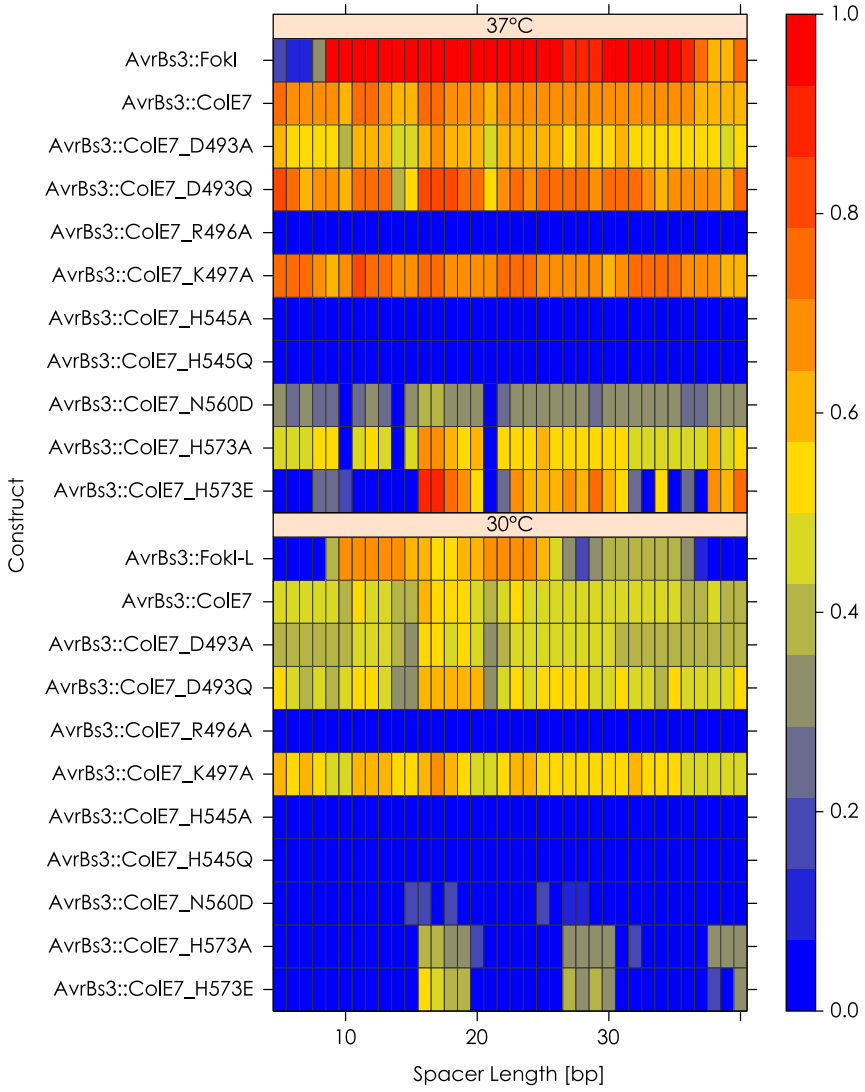



Figure 9: Mutations alter the cleavage pattern of a ColE7-based TALEN
Mutants for the ColE7 nuclease domain were assayed in yeast. Data were obtained after 48h at 30° and 37°C. Blue color stands for low, while red stands for high relative activity. Here only targets containing two AvrBs3 EBEs on opposite strands are shown. The spacer length determines the distance between the two. Targets were cut with different efficiencies, despite the TALENs working as monomers. Mutants for the catalytic residues H545 and R496 were inactive.

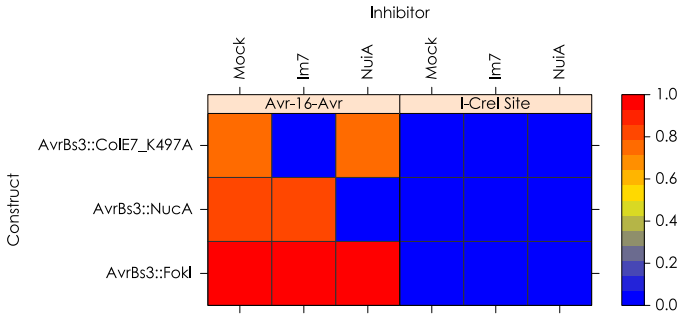



Figure 10: ColE7 and NucA constructs can be inhibited *in vivo*
AvrBs3::ColE7_K497A and AvrBs3::NucA were coexpressed with inhibitors Im7 and NuiA in yeast. Shown is the relative activity of each construct on one target plasmid Avr-16-Avr and the off-target I-CreI after 48h at 37°C. No activity was detectable, when the respective inhibitors were present.

was always a possibility. Nevertheless, we found that addition of a short linker sequence, which allowed fusions to the C-terminus of ColE7, was increasing relative activity in yeast by itself, while slightly reducing off-site activity. We decided to test this augmented catalytic domain (here termed ColE7_K497A_cfs) along with other monomeric TALEN constructs in a different RVD context. Here, we used TALE DNA-binding domains targeting the neomycin phosphotransferase II gene (see figure 11).

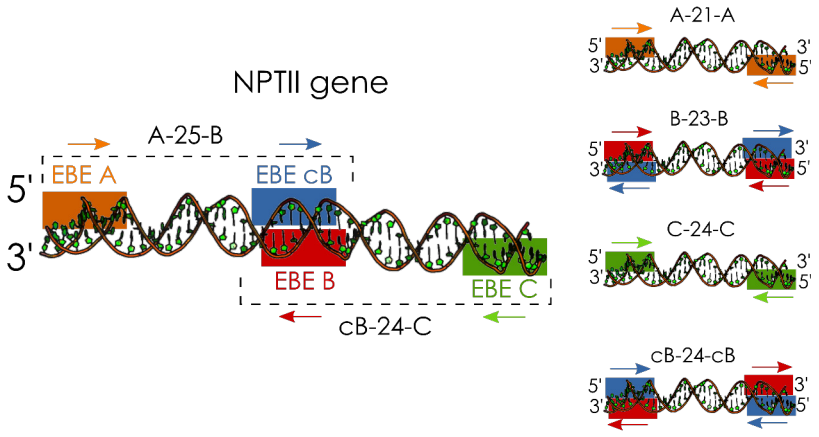



Figure 11: NPTII gene and EBEs
DBDs targeting four effector binding elements (EBE A, EBE B, EBE cB and EBE C) in the NPT II gene were chosen for our monomeric TALENs. The sequence for EBE B is partially complementary to EBE cB. Target sites had a length of either 57 or 60 bp, with two EBEs on opposite strands. Sequences of the target sites A-25-B and cB-24-C are present in the neomycin phosphotransferase II gene (*nptII*). Arrows show the directionality in 5′ to 3′ direction of each EBE.

This gene conveys resistance to a variety of aminoglycoside antibiotics (e. g., kanamycin, neomycin) and is used as a selection marker.

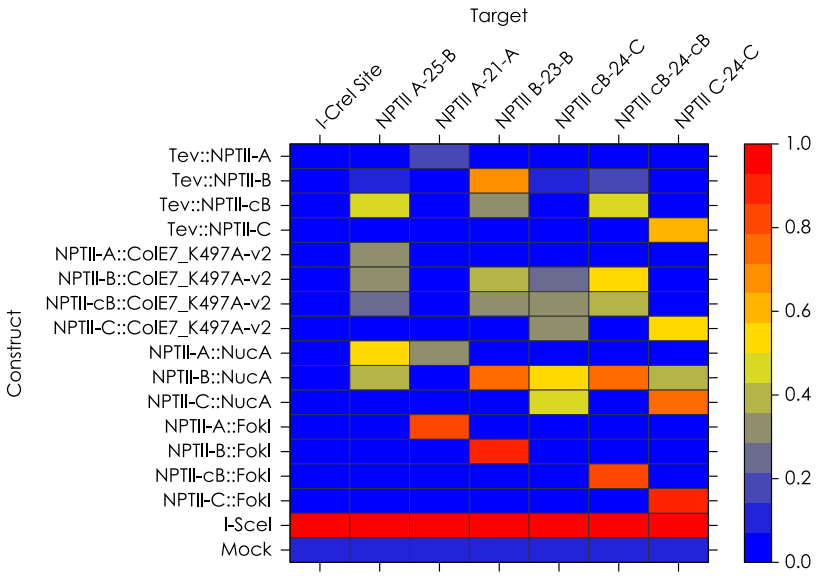



Figure 12: Monomeric TALENs induce homology directed repair on the NPTII locus
TALENs based on I-TevI, ColE7 and NucA nucleases were examined in a yeast SSA-assay. FokI constructs acted as positive controls. Target plasmids contained sequences from the neomycin phosphotransferase II gene (see figure 11). Shown are datapoints that were obtained after 48 hours at 37°C.

An *npt*II knockout can improve biosafety and reduce regulatory concerns for GMOs. We selected repeat arrays which had been made for the FokI architecture so we could compare our constructs. Profiling with a yeast SSA-assay revealed that our constructs were active (see figure 12) on their respective targets. Three monomeric TALENs, based on the nuclease domains from I-TevI, ColE7 and NucA, were tested. FokI-based constructs acted as controls.

Activity of Tev-based constructs was highly dependent on the presence of a suitable `CNNNG` sequence. For example, activity for the TALEN Tev::NPTII-cB on the target A-25-B had an average relative activity of 0.45, while there was no activity detectable on the target cB-24-C. The EBE in both cases was the same, with the only difference being a missing `CNNNG` motif for target cB-24-C.

Colicin E7 constructs did not share this limitation. All constructs displayed activity, albeit lower than for FokI. Recombination was generally higher when two EBEs were present, with the surprising exception of the construct targeting NPTII-A. This construct was virtually inactive (0.05) when a target plasmid with two A-EBEs was provided, although it displayed some activity when just one site was present.

The most active compact TALEN (cTALEN)s were based on NucA. Their relative activities reached values comparable to FokI (0.8) on targets with two EBEs. We detected one case of off-target activity. NPTII-B::NucA cut the target C-24-C, despite no suitable target being present.

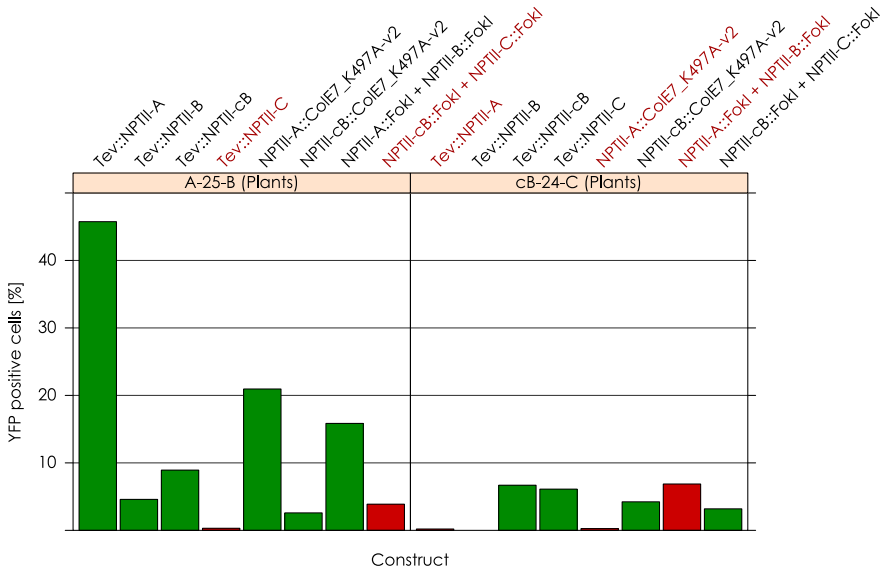



Figure 13: monomeric TALENs enable recombination in tobacco protoplasts
TALEN constructs were expressed via a 35S CaMV promoter in protoplast cells. Active constructs cut a target plasmid harboring a NPTII site which disrupted a YFP reporter gene. Homologous recombination restored the reporter gene. YFP-positive cells were then sorted and the total percentage was calculated. Transfection efficiency was controlled via a 35S CaMV YFP reporter plasmid and was at 82% ±1.7%. Constructs acting on off-targets are shown in red. See section 2.3 for more information about I-TevI TALENs.

After this initial validation we decided to see if our constructs were working in higher plants. For this, the constructs were subjected to another SSA-assay in tobacco protoplasts by Cellectis Plant Sciences [21]. Plasmids containing TALENs under the control of a 35S cauliflower mosaic virus (CaMV) promoter were cotransfected with plasmids containing either the A-B or the cB-C target site. Successful recombination restored a flanking YFP reporter gene. Cells were then sorted and the percentage of YFP-positive cells was calculated (see figure 13). Four Tev-based and two Colicin E7-based TALENs were used, whereas the NucA constructs could not be transferred into the plant expression vector.

The strongest effect had the Tev-based construct targeting the EBE A, on average reaching a rate of 45% YFP-positive cells, while the respective FokI TALEN pairs reached 16%. The FokI value was also surpassed by the ColE7 TALENs (21%) targeting this EBE. Other monomeric TALENs were less active, suggesting that TALE array or `CNNNG` motif were worse.

Surprisingly the constructs with the largest amount of off-target activity were the FokI TALENs. The FokI TALEN pair A/B even exceeded the FokI pair cB/C on its native target, despite only matching the cB EBE. Monomeric TALENs exhibited off-target activities below 0.3%.

In view of these findings, we decided to evaluate the nucleases in a mammalian system. We cloned AvrBs3::ColE7_K497A_V2 and the low activity variant AvrBs3::ColE7_H573E into a mammalian expression vector and transfected the constructs into chinese hamster ovary (CHO) cells. No activity could be detected in extrachromosomal SSA-assays.

Complete wild-type Colicin E7 consists of a translocation, a receptor-binding and a nuclease domain, which is cleaved off upon entering of the target bacterium. Lysis occurs between arginine 446 and lysine 447, which is also involved in DNA binding. To ascertain that we did not include a cryptic cleavage site in our fusion constructs, we mutated these two residues and profiled the mutants in a yeast SSA-assay (see figure 14). Tested target plasmids could be grouped into four classes: plasmids with two EBEs in the standard head to head orientation (1, a bound TALEN C-terminus faces another bound TALEN C-terminus), plasmids with two EBEs but non-standard orientation (2,N-terminus faces C-terminus), plasmids with just one EBE (3) and off-target plasmids without an AvrBs3 EBE (4).

One conservative mutation (glutamine for lysine 446 and lysine for arginine 447) and one non-conservative mutation was chosen (tyrosine for 446 and alanine for 447). Alteration of residue 446 yielded mutants with cleavage profiles comparable to AvrBs3::ColE7_K497A_V2. Interestingly, the mutant AvrBs3::ColE7_R447K displayed no detectable off-target activity. Only two off-targets were cut, however, these two targets were also cut by the AvrBs3::FokI control, making the correct identity of these targets doubtful. Activity on the standard targets remained comparable, while roughly half of the one-site targets showed no apparent recombination. The variants K446Q and R447K were then tested in a mammalian SSA-assay, since the first variant showed good activity in yeast, while the later did not show any off-site cleavage. Neither variant caused any sort of recombination.

Another interesting find was the difference between the targets with two EBEs. ColE7 TALENs cut targets with the standard EBE orientation better than non-standard targets. Furthermore, these non-standard targets were cut as well or worse than targets with just one EBE, suggesting that the location of the DSB in relation to the homology regions is more important than the amount of DSBs. There was also the additional variability that could be seen when just examining the standard FokI targets (see figure 9).

To evaluate these findings, we compared a ColE7 variant fused to the TALE N-terminus to the original AvrBs3::ColE7 (see figure 15). Both constructs cut the standard targets best. ColE7::AvrBs3 cut the atypical two EBE plasmids slightly better than the one target site plasmids. A striking difference between the two constructs was their behavior on the off-targets. While AvrBs3::ColE7 cut all plasmids with a relative activity of at least 0.09 with an average of 0.12,

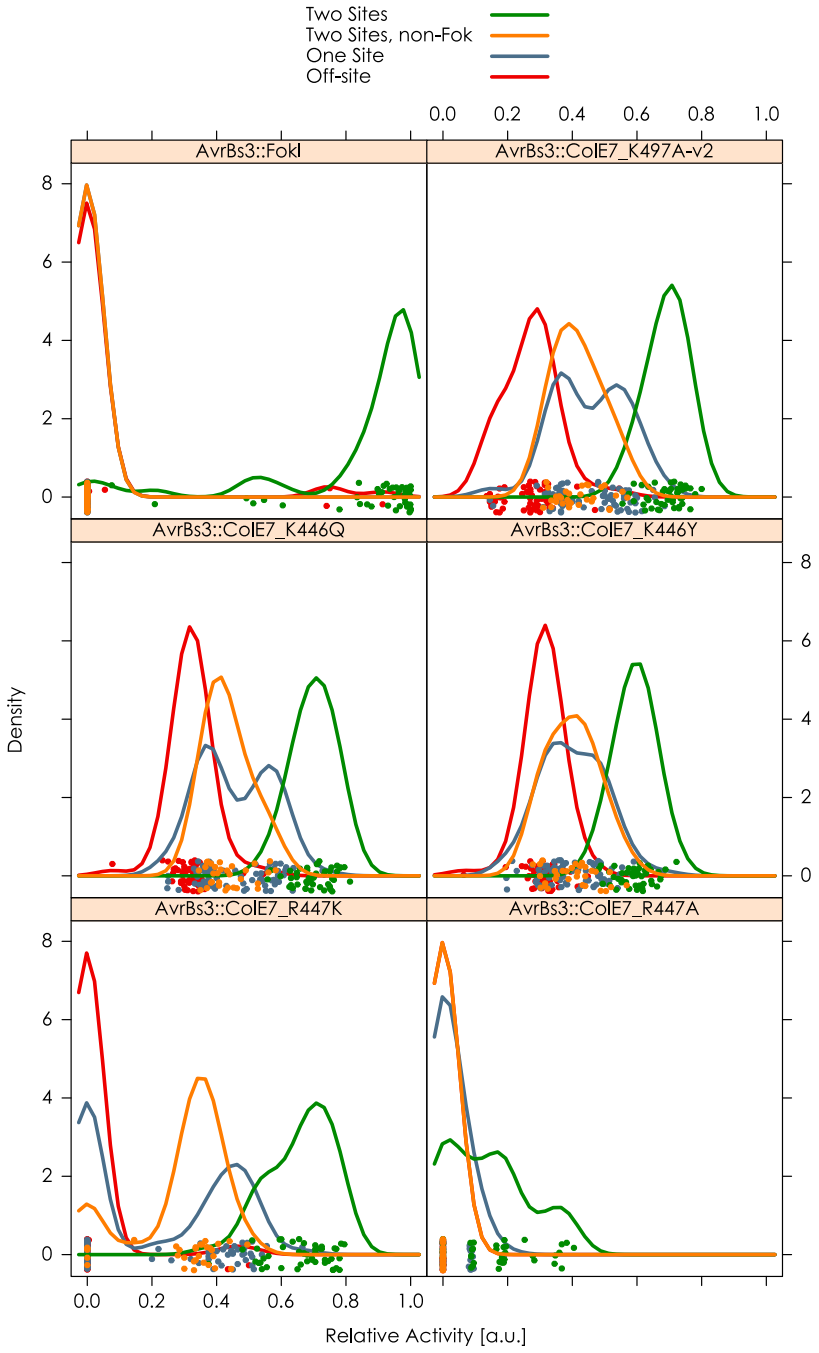



Figure 14: Residues 446 and 447 have a strong effect on activity
Mutants were profiled in a yeast SSA-assay. Each point represents at least two measurements at 37°C after 48h on the same target. Green points represent targets that contain two AvrBs3 sites in "head-to-head (C-C)" orientation, which is suitable for FokI. Orange targets also contain two target sites, however in the "tail-head (N-C)" orientation. Blue targets contain one target site, while red represents off-targets wihout any AvrBs3-EBE. Relative activity is given on a scale of 0 to 1. As a reference, the cleavage profiles of AvrBs3::FokI (top left) and AvrBs3::ColE7_R497K_v2 (top right) are given. Both 446 mutants are slightly more active than the WT (see figure 15) and exhibit a profile similar to the one of R497K_v2. Conservative mutation of the residue R447 (bottom left) removes detectable off-site activity, while reducing overall activity slightly.

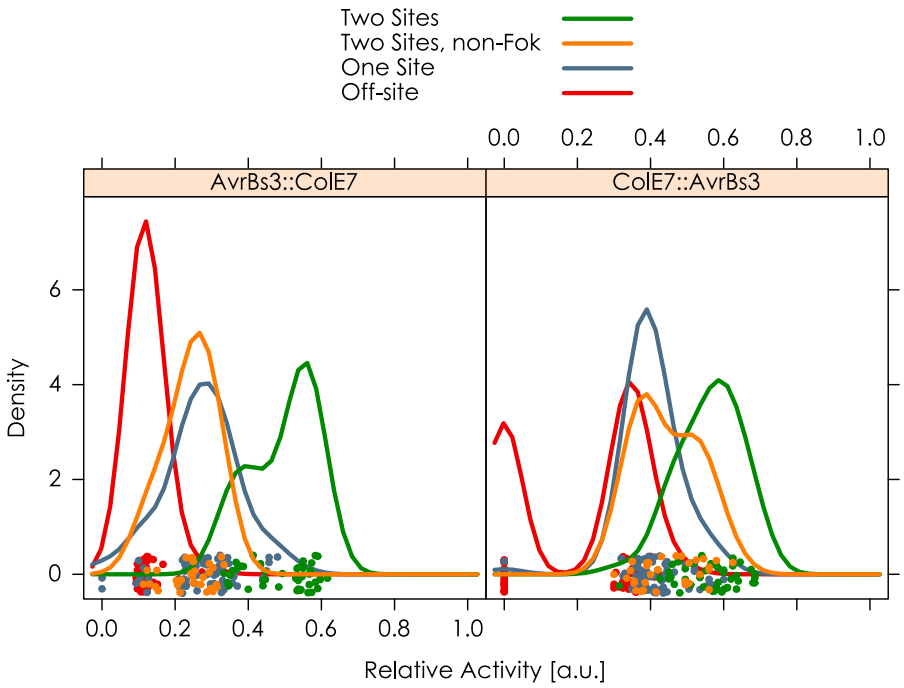



Figure 15: The fusion terminus has an impact on ColE7's cleavage pattern
Two constructs based on the same TALE-DBD and catalytic domain, but with different domain arrangement, were tested in a yeast SSA-assay. The orientation of the catalytic domain has an impact on the cleavage profile. The N-terminal fusion ColE7::AvrBs3 is more active than its C-terminal counterpart AvrBs3::ColE7. There are two populations of off-targets for the N-terminal fusion. One population is not cut at all, while the other is cut with a relative activity of around 0.3. This is not the case for the c-terminal fusion. Here, all off-targets are cut with equal activity.

ColE7::AvrBs3 cut half the off-target plasmids with an activity comparable to the single EBE plasmids, while no recombination could be measured for the other half. This lead us to believe that off-site activity is either star-site like or eluded detection in this type of assay.

We tested some of the variants with a smaller TALE-DBD (13.5 repeats) targeting part of the recombination activating gene *RAG2* (see figure 16). *RAG2* is a target of interest for gene therapy since a defective version can cause severe combined immunodeficiency (SCID). Yeast SSA-assays showed that for the shorter construct relative activity, but also off-target activity, decreased compared to the longer AvrBs3 DBD (17.5 repeats). It has to be noted that both arrays cannot be compared adequately, since the recognition sequences are different.

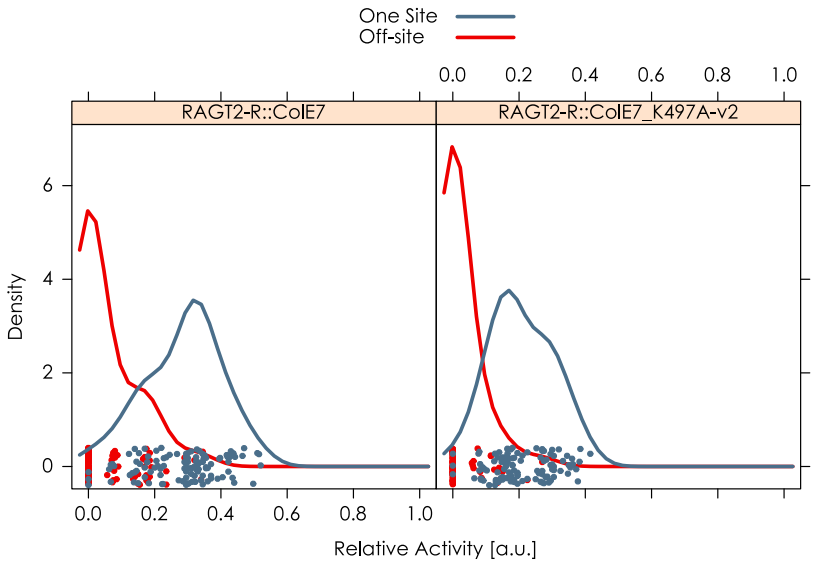



Figure 16: Activity of TALE::ColE7 is reduced for the RAG TALE-DBD
Two ColE7-based constructs were profiled in a yeast SSA-assay. Activity for the target sequences reached 0.4, while it reached 0.6 for an AvrBs3-based TALEN (see figure 14). The majority of off-target plasmids were not cut.

To study unspecific nuclease domains further, we decided to investigate cleavage of an unspecific nuclease domain *in vitro*. It had been shown that inactive EndA mutants could be chemically rescued *in vitro* [148]. Our nuclease variants contained the mutations H160A or H160G, which replaced a catalytic histidine of the H-N-H motif with a glycine or alanine. Addition of imidazole restored the activity. This system allowed us to safely express constructs and then assess specificity *in vitro*. For this, we fused the DNA-entry nuclease (EndA) of *Streptococcus pneumoniae* to the N-terminus of the AvrBs3 DNA-binding domain. The chimeric nuclease was then purified with two affinity purifications. First via binding of its C-terminal His-tag to a Ni-NTA column, then via binding to a heparin column. Heparin "mimics" the structure of DNA and can be used to purify DNA-binding enzymes, but also kations in general. The elution profile for the heparin column can be seen in figure 17.

Purity of the eluates was then determined via sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) (for an example purification, see figure 18). All TALENs had higher eletrophoretic mobilities than what their molecular mass suggested. Resequencing of the expression vectors and use in cleavage assays confirmed however, that the constructs were not truncated.

Cleavage assays proved that the construct EndA_H160G::AvrBs3 completely degraded our target plasmid, given enough time (see figure 19). To reduce this off-site activity, we decided to lower the activ-

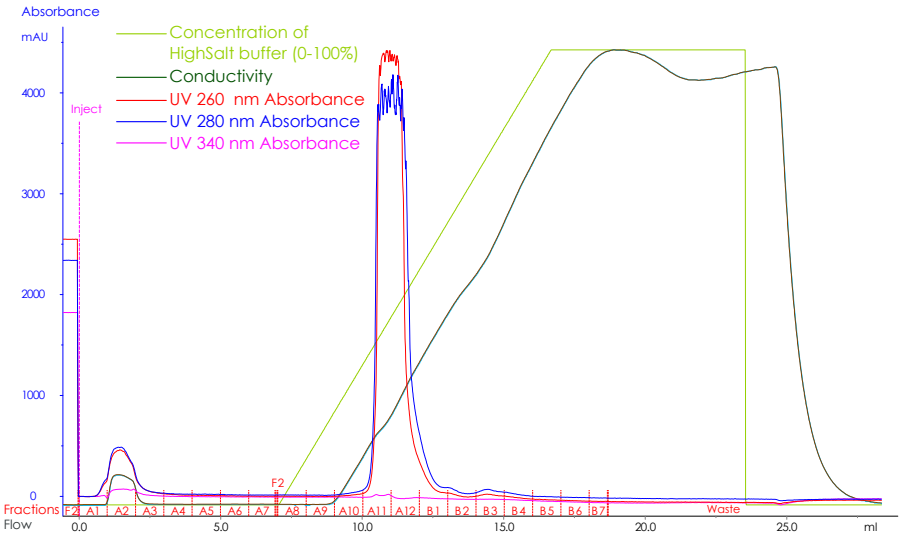



Figure 17: EndA_H160G::AvrBs3 binds to a heparin column and elutes at 400 mM of NaCl
Shown is UV absorbance at 260, 280 and 340 nm, conductivity and the concentration of the HighSalt-buffer. A preequilibrated heparin column was injected with His-tag-purified eluates and then washed. Samples were eluted by increase of the salt concentration up to a total of 1 M NaCl at 100% HighSalt buffer. Here, the protein samples eluted in fractions A11, A12 and B1, at a sodium chloride concentration of about 400 mM NaCl.

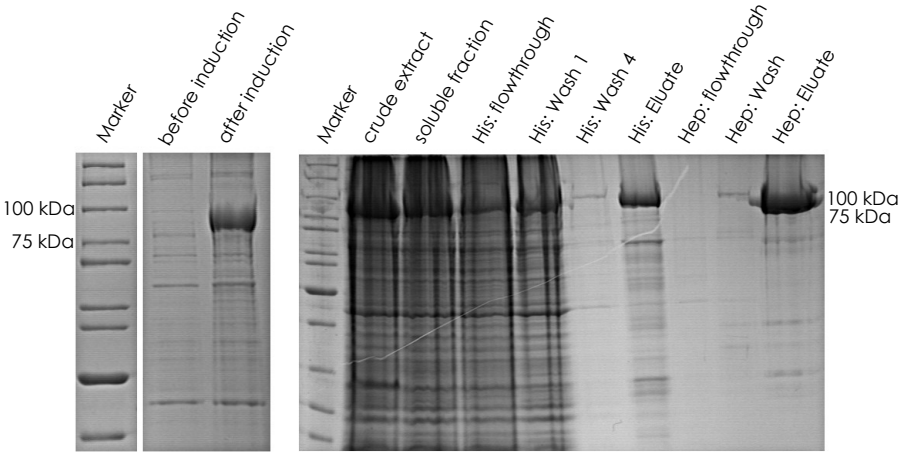



Figure 18: Purification of EndA::AvrBs3::snap
Shown on the left are samples from an XL10 GOLD expression culture for EndA_H160G_Q186A_N202A::AvrBs3::snap before and after induction with IPTG. On the right are samples: after cell lysis via sonication (crude fraction), soluble fractions after centrifugation and flowthrough, wash, eluate for the His-tag and heparin affinity chromatographies. The protein of interest is EndA_H160G_Q186A_N202A::AvrBs3::snap with a molecular mass of 126 kDa. Its electrophoretic mobility is comparable to a much smaller protein between 100 and 75 kDa. The acrylamid percentage of the gels was 12.5%.

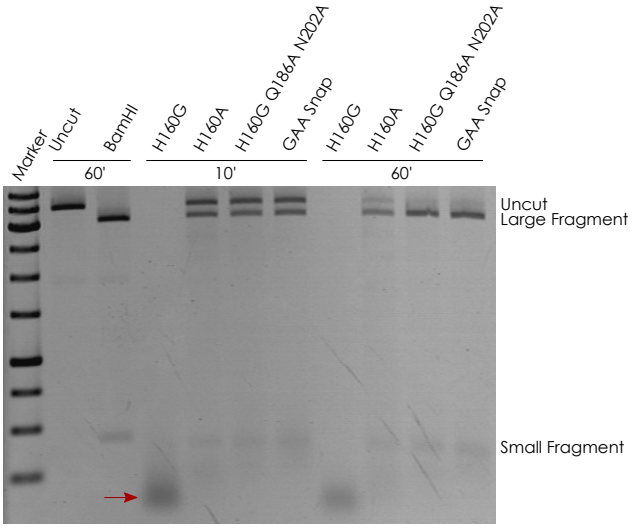



Figure 19: Mutation of the EndA domain diminishes enzymatic activity
Shown are different EndA::AvrBs3 variants cutting a plasmid target harboring an AvrBs3 EBE. 50 nM protein were used to digest 10 nM of plasmid DNA, in 100 mM NaCl, 50 mM Tris-HCl pH 7.9, 10 mM $MgCl_2$, 100 µg/ml BSA. The H160G mutant completely degrades (red arrow) the DNA. The mutations H160A and Q186A / N202A reduced the activity. Addition of the SNAP-tag seemed to have little impact.

ity of the nuclease domain. This was achieved by replacing residue 160 with an alanine. It is thought that the imidazole can fit better into the active site when His 160 was replaced by Glycine. This approach reduced the activity (see figure19), however its application *in vivo* is difficult in some types of cells, due to the chemical activation. Since the goal was to engineer the nuclease to be specific and then revert the H160 mutation, we chose two mutations that have been shown to reduce EndA activity [149]. For this reason we introduced the mutations Q186A N202A, which had the intended effect. While the original H160G mutant completely degrades the DNA in already 10 minutes at 37°C, the same amount of enzyme for the mutants produced clean cleavage products after an hour, although some degradation products can be seen. Another variant additionally contained a C-terminal SNAP-tag. This tag is a modified version of the O6-alkylguanine DNA alkyltransferase [83] and can be used to covalently label proteins via a benzyl group. It has been demonstrated that TALE::SNAP and SNAP::TALE fusion proteins still bind to DNA [90]. Addition of this tag seemed to slightly improve specificity, which can be seen in figure 20.

Activity of the constructs depended on several factors. High salt concentration lowered, while low salt concentration promoted on- and off-site cleavage (see figure 21). Varying degrees of specificity could be obtained, depending on the buffer composition (see appendix 49).

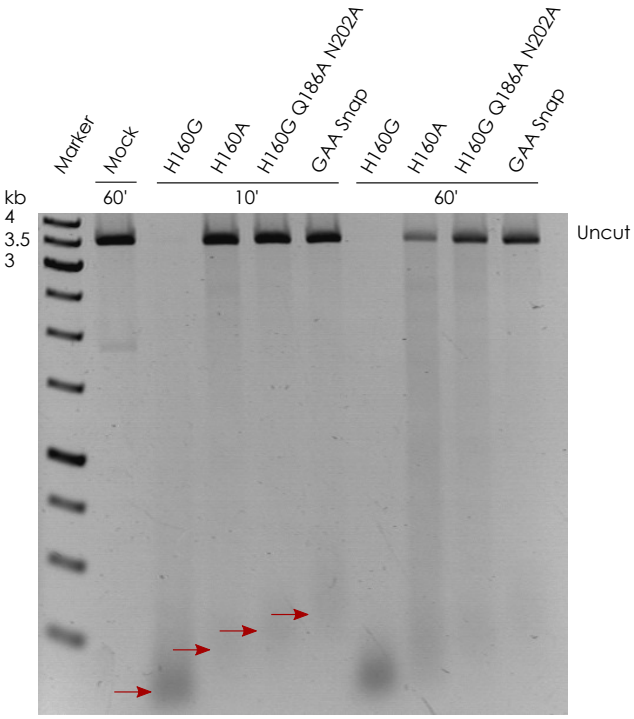



Figure 20: EndA mutants exhibit reduced off-target cleavage
10 nM of a linearized plasmid, lacking the AvrBs3 EBE, were incubated with 50 nM of different EndA mutants in 100 mM NaCl, 50 mM Tris-HCl pH 7.9, 10 mM $MgCl_2$, 100 µg/ml BSA. The mutant closest to the wild-type (H160G) completely degraded the plasmid DNA in less than 10 minutes. Mutation of the catalytic site (H160A) reduced the amount of off target activity. Alteration of the DNA binding residues (H160G/Q186A/N202A and H160G/Q186A/N202A SNAP) decreased it even further. Still some degradation products (red arrows) and a smear of unspecifically cut DNA can be seen.

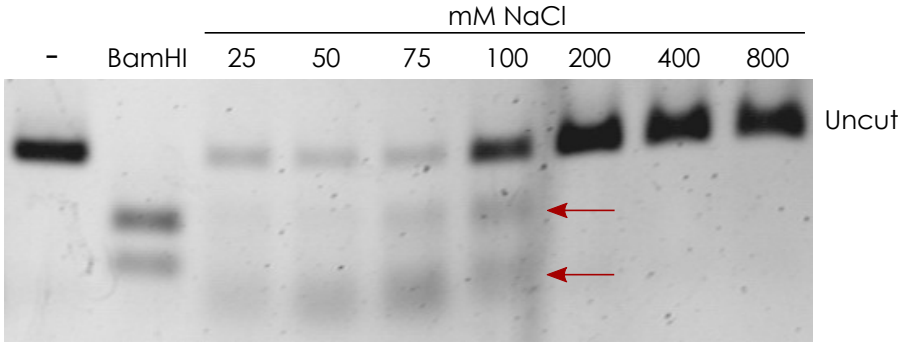



Figure 21: High salt concentration inhibits cleavage
A PCR fragment (594 bp, 30 nM) harboring an AvrBs3 EBE was cleaved by EndA_H160G::AvrBs3 (50 nM, 15 min at 20°C, 50 mM Tris pH 8, 50 mM imidazole, 10 mM $MgCl_2$). The nuclease proved to be active at low sodium chloride concentrations, however the amount of specifically cut fragments (red arrows) was highest around a NaCl concentration of 100 mM. While lower salt concentrations reduced the amount of uncut fragment most, they also showed the highest amounts of degradation.

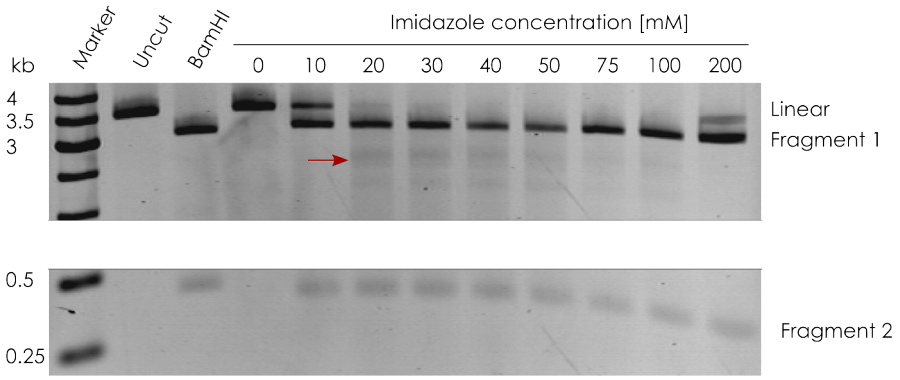



Figure 22: EndA_H160G::AvrBs3 is active over a wide range of imidazole concentrations
Shown is a cleavage assay (50 mM EndA_H160G::AvrBs3, 10 mM addressed plasmid, 15 min at 20°C, 50 mM Tris pH 8, 100 mM NaCl, 10 mM $MgCl_2$) to determine imidazole requirements for DNA cleavage by EndA_H160G::AvrBs3. The construct is active for the most concentrations under these reaction conditions. Unspecific cleavage products appear (red arrow).

The constructs proved to be functional over a wide range of imidazole concentrations (see figure 22).

To determine potential strand specificities, we examined cleavage fragments of a fluorophor labeled PCR product during a timecourse assay with denaturing urea native PAGE. Double strand breaks will only show up in the native gel, while nicks appear in the denaturing gel (see figure 23).

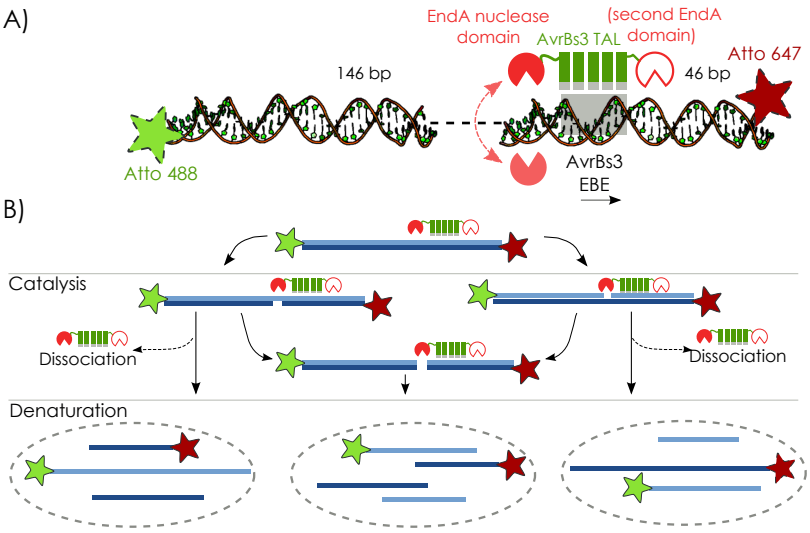



Figure 23: Overview of EndA::TALE cleavage assay
Substrate (A) for the cleavage assay is a 211 bp PCR fragment labeled with Atto488 (green, top strand) and Atto647 (red, bottom strand). An AvrBs3-EBE is located in the later half of the fragment. During catalysis, a single strand is cut first and then converted to a double strand break (B). Nicked intermediates can be discriminated via denaturing PAGE.

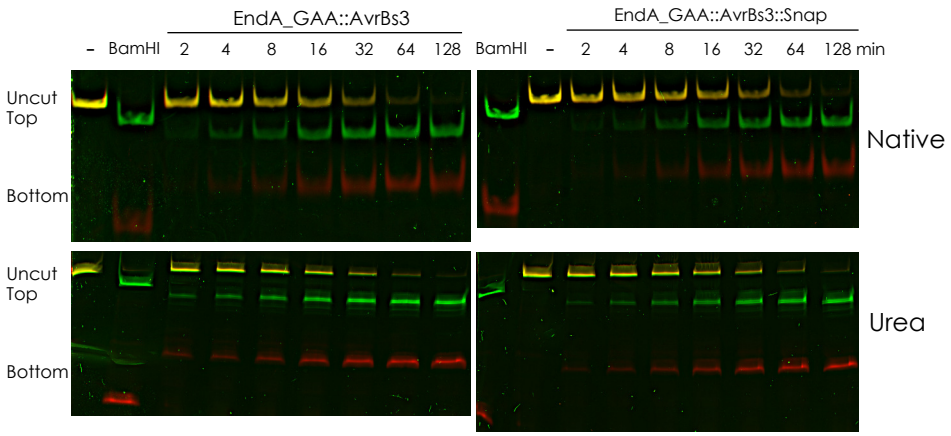



Figure 24: EndA_GAA constructs do not show a strand preference
Shown are composite images of Atto488 (green) and Atto647n (red) fluorescence for a native and a denaturing (urea) gel. A PCR product was cleaved and then half of the sample was analysed via native-, the other half with urea PAGE. Green fluorescence reprensents the top-, red fluorescence the bottom strand. TALE and DNA concentration were both at 20 nM. Paa concentration was 8% for the native and 10% for the denaturing gel. Assays were carried out at 37°C in 100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 µg/ml BSA and pH 7.9. BamHI was used as a positive control. The amount of nicked- and cut DNA increases over time. For the analysis see figures 25 and 26.

It is also possible to discriminate top- and bottom strand nicks via different fluorescence. By subtracting DSBs (native) from total nicks (urea), one can conclude the amount of single nicked products. Example gels for EndA_H160G_Q186A_N202A::AvrBs3 and its SNAP-tagged version can be seen in figure 24.

The analysis of the gels (see figure 25) showed that there was no strand preference for the constructs. Interestingly the SNAP-tagged variant produced fewer nicked DNA strands, but overall the same amount of double strand breaks (see figure 26). This is suggesting that the initial nicks are slower for EndA_H160G_Q186S_N202A::AvrBs3::SNAP, but the nicked intermediate states are converted faster to DSBs. Under the assay conditions, no cleavage could be observed for PCR fragments lacking the AvrBs3-target site.

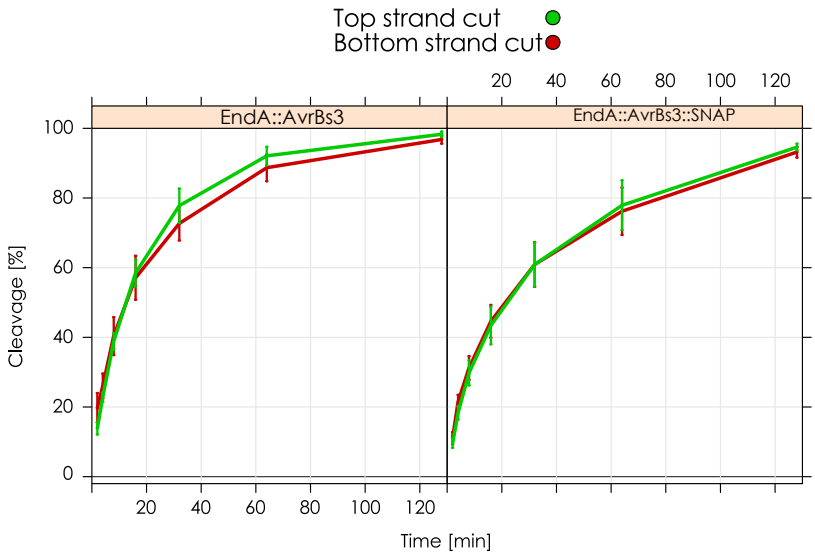



Figure 25: There is no detectable strand-preference for N-terminally fused EndA

Shown are percentages of top- (green) and bottom strand (red) cut PCR-product from time course cleavage assays (see figure 24 (n=5)). The two constructs EndA_H160G_Q186A_N202A::AvrBs3 (here abbreviated: EndA::AvrBs3) and EndA_H160G_Q186A_N202A::AvrBs3::SNAP (here abbreviated: EndA::AvrBs3::SNAP) show no preferences for bottom- or top strand cleavage (EndA::AvrBs3 p=0.47;EndA::AvrBs3::Snap p=0.30).

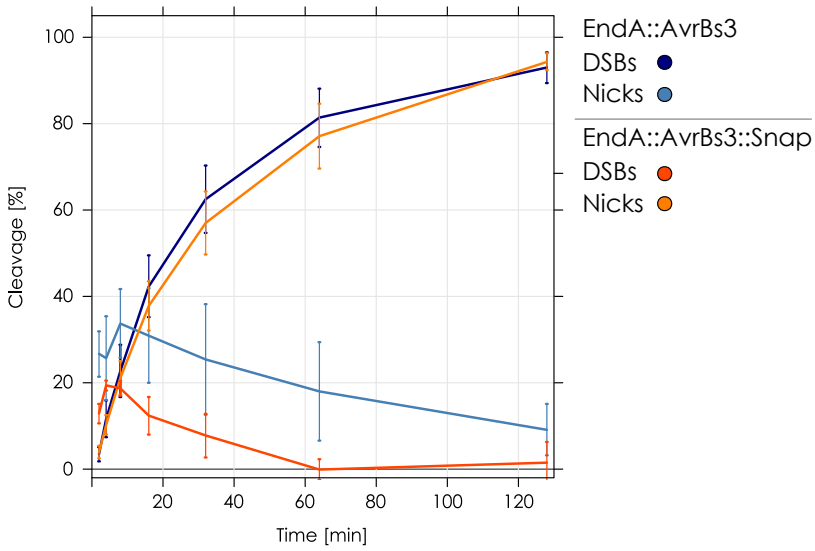



Figure 26: Addition of a SNAP-tag changes the cleavage pattern of an EndA TALEN

Shown is the amount of double-strand breaks in a PCR-product from time course cleavage assays (see figure 24 (n=5)). The two constructs EndA_H160G_Q186A_N202A::AvrBs3 (blues, here abbreviated: EndA::AvrBs3) and EndA_H160G_Q186A_N202A::AvrBs3::SNAP (oranges, here abbreviated: EndA::AvrBs3::SNAP) produce the same amount of DNA-DSBs (p=0.32), but the amount of single-nicked substrate differs significantly ($p=3*10^{-4}$ for just top-nicked, $p=6*10^{-4}$ for just bottom-nicked substrates). No preferences for bottom- or top strand cleavage were found.

Since the EndA-scaffold worked as a monomer, we decided to fuse a second EndA nuclease domain to the C-terminus to create a tripartite enzyme (for more dual catalytic TALENs see section 2.5). After confirming that both domains were active (e. g., appendix 49), the previous assays were redone with the new EndA_H160G_Q186A_N202A::AvrBs3::EndA_H160G_Q186A_N202A. Interpretation of the results is more difficult, since there are two DNA single strands (top and bottom), which can be cut at two places (up- and downstream of the AvrBs3-EBE). This means that one DNA molecule can be converted into 16 different forms. It has also to be noted that a single strand with two nicks cannot be distinguished from a single strand with a nick close to the fluorophore. For examples a downstream (or C-terminal) - top strand nick will disappear in this assay, as soon as the top strand is also nicked upstream (by the N-terminal nuclease domain). This works vice versa for the bottom strand.

Analysis via PAGE showed (see figure 27) that the C-terminal nuclease domain was less active overall, but that this lack of activity stemmed from a lack of bottom strand cleavage (see figures 28 and 29). These findings are consistent with observations made with N-terminal ColE7 and N-terminal I-TevI TALENs (see section 2.3). Together they suggest that domains fused directly to the N-terminus are more flexible than domains fused to the C-terminus via commonly used linkers.

To test our EndA constructs *in vivo*, we reverted the H160G mutation in the Q186A / N202A context, so that the EndA domain is constitutively active. While the wild-type H160 mutant could not be cloned, mutants H160Q and H160N were generated successfully. The idea was that the amino groups in glutamine and asparagine could take over the function of the nitrogen in the histidine sidechain and activate the catalytic water molecule (see figure 4 for an overview of the cleavage mechanism of H-N-H nucleases). Plasmids for both mutants could be propagated in *E. coli*, however only H160N could be expressed, despite correct sequences for both. *In vitro* assays after purification showed no detectable activity for EndA_H160N_Q186A_N202A::AvrBs3, suggesting that a different replacement is needed for histidine 160.

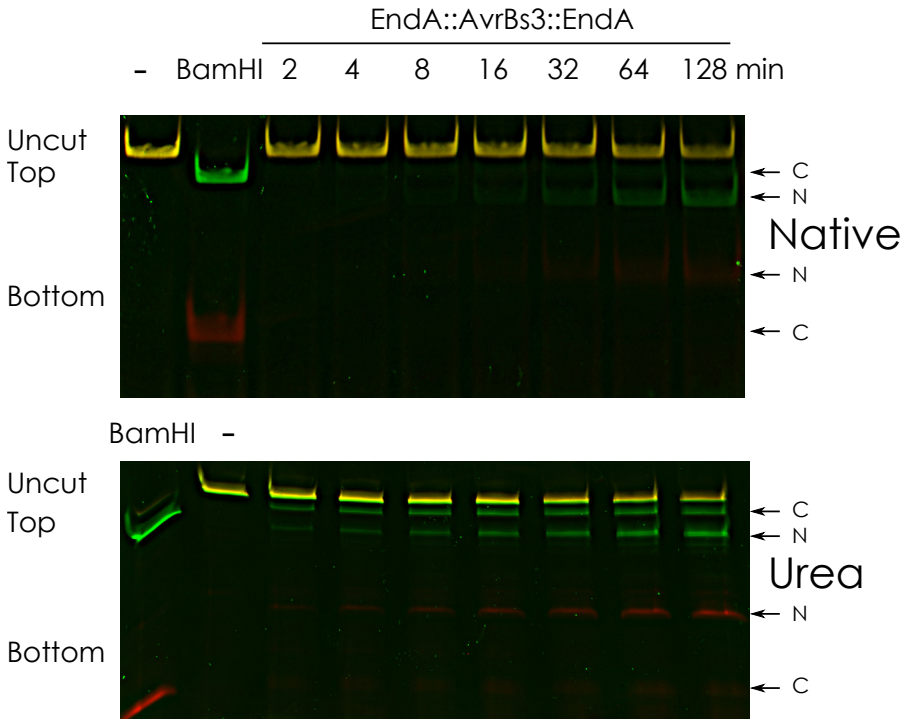



Figure 27: A C-terminal EndA domain is less efficient than an N-terminal one

Shown are composite images of ATTO488 (green) and ATTO647N (red) fluorescence for a native and a denaturing (urea) gel. A PCR product was cleaved and then half of the sample was analyzed via native-, the other half with urea PAGE. Green fluorescence represents the top-, red fluorescence the bottom strand. TALE concentration was 80 nM and DNA concentration was at 20 nM. PAA concentration was 8% for the native and 10% for the denaturing gel. Assays were carried out at 37°C in 100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 µg/ml BSA and pH 7.9. BamHI was used as a positive control. ENDA_H160G_Q186A_N202A(GAA)::AVRBS3::ENDA(GAA) is overall less active than its single domain counterpart. Although the C-terminal domain produces less DSBs than the N-terminal one, it generates a more similar amount of top strand nicks. For a detailed analysis see figures 28 and 29.

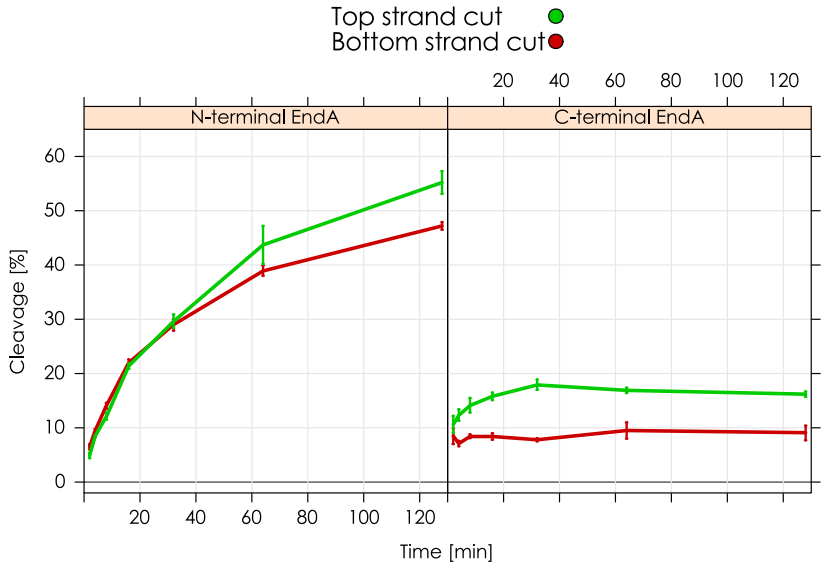



Figure 28: The C-terminal EndA domain in a tripartite construct has a preference for top strand cleavage
Shown are top- (green) and bottom (red) strand cleavage for the N- and C-terminal EndA domain of a tripartite TALEN (see figure 27 (n=3)). The C-terminal EndA domain has a clear preference ($p=4*10^{-12}$) for top strand cleavage. It has to be noted that top strand cleavage for the C-terminal domain and bottom strand cleavage for the N-terminal domain are underestimated in this assay, due to conversion of the long fragments to shorter fragments by the opposite domain (see text).

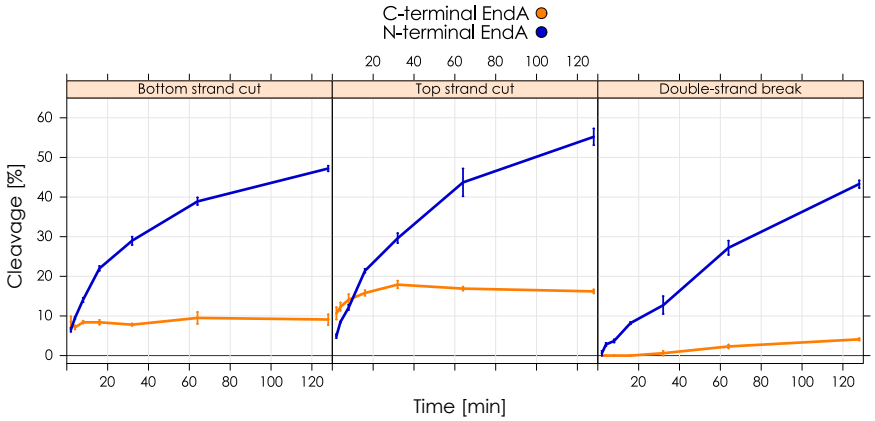



Figure 29: The N-terminal EndA domain in a tripartite TALEN construct is more active than the C-terminal one
Shown are bottom-, top- and double-strand cleavage for the N- (blue) and C-terminal (orange) EndA domains of the tripartite TALEN EndA_H160G_Q186A_N202A::AvrBs3::EndA_H160G_Q186A_N202A (see figure 27). The N-terminal domain generated more strand breaks in total, however the C-terminal domain produced more top strand breaks in the first few minutes. The real amount of top strand cleavage by the C-terminal domain is underestimated (see text).

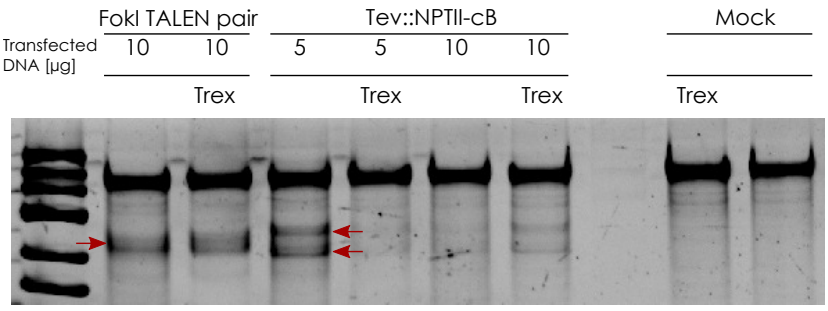



Figure 30: Genome modification by an I-TevI-based TALEN can be verified via T7 assay
Shown are PCR fragments generated from genomic DNA (obtained from CHO-K1 cells treated with nucleases), which were cut by T7 endonuclease 1. 5 or 10 µg of plasmids harboring a nuclease targeting the NPTII locus were transfected. Cleavage products (red arrows) of the T7 endonuclease 1 mirror the amount of gene modification *in vivo*. The I-TevI-based TALEN had a greater impact at a transfection amount of 5 µg than at 10 µg. Cotransfection of a plasmid harboring human exonuclease Trex2 instead of a mock plasmid reduced the amount of cleavage when 5 µg of the TEV::TALE plasmid where transfected and slightly increased the amount, when 10 µg were transfected.

## 2.3 I-TEVI FUSION CONSTRUCTS

As an alternative to our unspecific catalytic domains we chose the nuclease domain of homing endonuclease I-TevI. A trait of this domain is its "CNNNG" specificity that should reduce the number of potential target sites and off-targets, which it is thought to do in nature. I-TevI binds to its own promotor for autorepression, while the missing CNNNG motif prevents the catalytic domain from cleaving [62].

We fused the catalytic domain to the N-terminus of TALE-DNA binding domains, mimicking the natural I-TevI structure (N- to C-terminus: nuclease domain - linker - DNA-binding domain). TEV::TALE fusion constructs targeting the neomycin phosphotransferase II gene worked successfully as monomers in yeast (see figure 12) and in tobacco protoplasts (see figure 13). We also tested the constructs on an endogenous target in CHO-K1 cells. In some cases we cotransfected a plasmid coding for the human 3' to 5' exonuclease Trex2, which can act as an enhancer reagent for mutagenic NHEJ [58].

A quick overview can be achieved by generating PCR fragments of the modified genomic region and then using them in a T7 endonuclease 1 assay. Here, PCR fragments are denatured, reannealed and mismatching duplexes are subsequently cut by the T7 endonuclease 1. The amount of cleavage products is an indicator for the amount of genome modification, since they stem from annealing wild-type DNA copies with mutated forms. The assay proved that the TEV::TALE was active (see figure 30), which was then confirmed by deep sequencing (on average ~4.3% modification for 5 µg samples, see [21]).

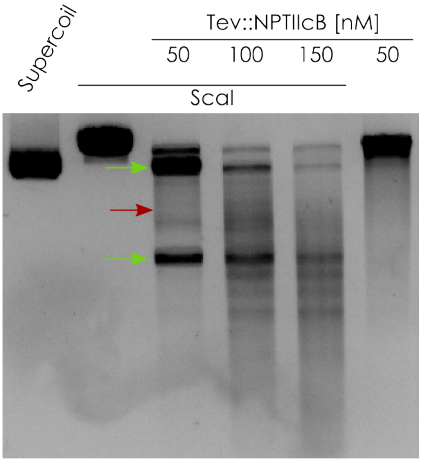



Figure 31: TEV::NPTII-CB IS ACTIVE BUT SHOWS SIGNS OF OFF-SITE CLEAVAGE
A supercoiled plasmid (10 nM, 10 kb) harboring a neomycin phosphotransferase II gene was incubated (1h at 37°C in 50 mM Tris/HCl pH 7.9, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT) with varying concentrations of the TEV::NPTII-CB TALEN. ScaI was used to linearize part of the samples. A five fold excess of protein (50 nM) over DNA generated specific cleavage products (green arrows), but also produced unspecifically cut plasmid (red arrow). The amount of degraded plasmid increased with protein concentration with the highest amount of off-target activity at a 15-fold excess. Linearization could also be achieved by the TALEN (right lane), however a smear of unspecific cleavage products appeared.

TEV::NPTII-CB was also purified and assayed *in vitro* to assess fidelity. Our nuclease generated specific double strand breaks, but also a smear of unspecific degradation products (see figure 31). The amount of degradation products is proportional to the nuclease concentration. Despite these findings, we did not observe any toxicity for Tev-based constructs in viability assays (see [21]). Interestingly, even a FokI-based nuclease generates off-site products *in vitro* (see appendix 50), suggesting that off-site activity *in vivo* is either lower or less relevant.

We also tried an "unnatural" arrangement by attaching the I-TevI nuclease domain to the C-terminus of a TALE-DBD, generating a TALE::TEV fusion. Surprisingly, this changed the cleavage mechanism of the nuclease. While the I-TevI nuclease domain causes DSBs when fused to the N-terminus, it only creates nicks when fused to the C-terminus. Two EBEs were needed to achieve homologous recombination in a yeast SSA-assay (see figure 32). An additional requirement was the presence of the "CNNNG" motif between the two binding sites. This nickase activity was sufficient to introduce genes in human embryonic kidney cells (see [21]).

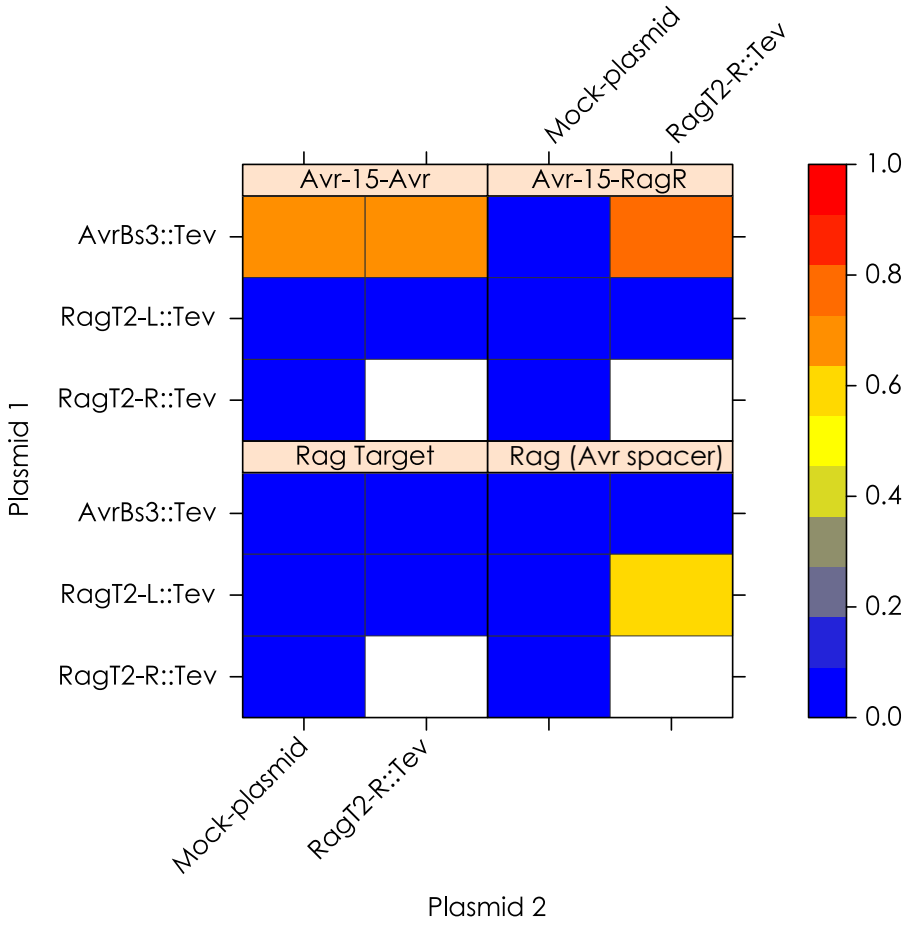



Figure 32: Two TALENs are needed to stimulate SSA when the I-TevI nuclease domain is connected C-terminally

A TALE::TEV fusion targeting the AvrBs3 site and a TALE::TEV pair targeting the *RAG* locus ("L" and "R") were profiled in a yeast-SSA assay. Shown are relative activities after incubation for 48h at 37°C. Nucleases were tested on four targets with effector binding elements for Avr, RagT2-L and RagT2-R. The "Avr-15-Avr" target could be cut by the AvrBs3::Tev nuclease, since there are two Avr EBEs. Target "Avr-15-RagR" required the Avr and the RagT2-R nuclease. The complete "Rag target" was only cut by the Rag-TALEN pair when the spacer was replaced by the spacer of the Avr-target, because the later contained the needed "`CNNNG`" motif.

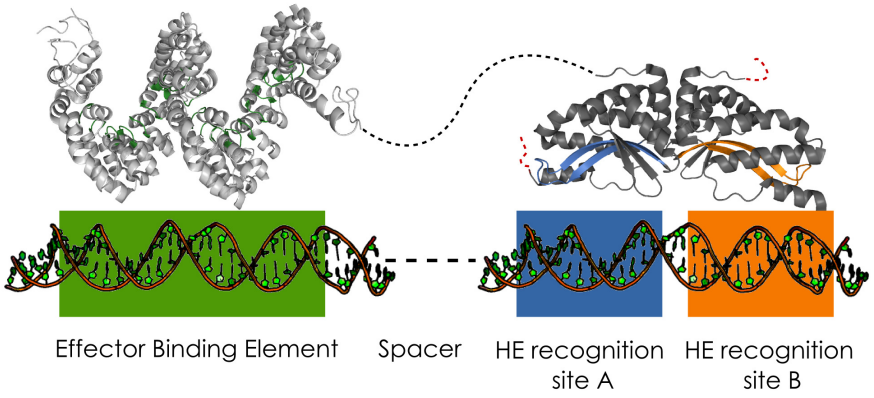



Figure 33: Layout of TAL::MEGAS and their target sequences
A TALE-DBD (light grey) is connected via a linker to the N-terminus of scI-CreI (dark grey). RVDs are highlighted in green consistend with their effector binding element which is also shown in green. EBE and homing endonuclease recognition site are separated by a variable spacer. The HE recognition site can be separated into two parts (blue and orange) and each part is bound by one domain due the bipartite nature of LAGLIDADG HEs. Cleavage occurs between the recognition sites. The consensus cleavage sequence is GTAC.

## 2.4 TALE::MEGANUCLEASE FUSIONS

I-CreI was chosen as a model for highly specific catalytic domains. Since I-CreI is a member of the dimeric group of LAGLIDADG homing endonucleases, we first generated a single-chain variant termed scI-CreI by fusing the C- and N-terminus via a linker. The subsequent monomeric variant was then attached to the C-terminus of a TALE-DBD scaffold [see figure 33] with another linker. This second linker is an α-helix based on a natural linker from GIY-YIG homing endonuclease I-TevI, which in that context connects the nuclease domain to its binding domain. We also introduced several mutations into scI-CreI to create the degenerate scaffold scI-CreI_DS. Parts of these mutations had been identified in broad-specificity variants of I-CreI within a mutant library [Claudia Bertonati, personal communication]. The reasoning behind this was to broaden the specificity of I-CreI while reducing its affinity; effectively increasing the number of possible targets for the catalytic domain without increasing overall toxicity.

The first TALE-DBD we tested was the well characterized AvrBs3 in our cT11-scaffold [see figure 7]. As an additional control we also fused an empty scaffold just containing a TALE N- and C-terminus without any repeats. We then profiled our constructs in yeast SSA-assays to assess activity.

ScI-CreI cleaved all provided target plasmids, when an I-CreI recognition site is present, while the derivative scI-CreI_DS exhibited only moderate activity (see figure 34) and no activity at 30°C growth condi-

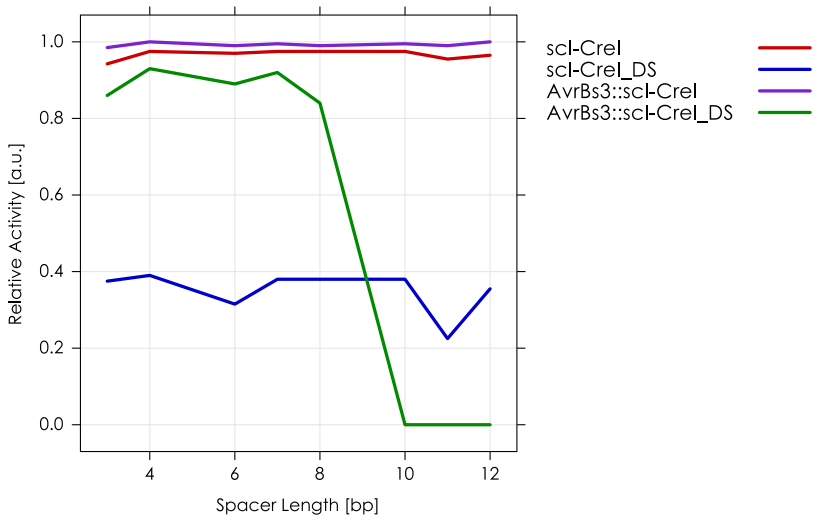



Figure 34: AvrBs3-DBD fusion modifies the activity of scI-CreI
Relative β-galactosidase activity as a consequence of DSB induced recombination in a yeast SSA-assay. Values shown are mean values obtained 48h after transfection with growth at 37°C. Both scI-CreI constructs (red and purple) caused a saturated signal under these conditions. The degenerate scaffold mutant (blue) is less active, but activity can be restored by fusion of a TALE-DBD (green). This rescued activity is dependent on the distance between effector binding element and HE recognition site.

tions. Another difference between the two variants is the influence of the TALE-DBD domain. The Avr fusion boosted the activity of scI-CreI on all targets and improvements could be seen at early timepoints or low temperatures (e. g., 37°C & 8h Avr-05-Cre: relative activity of 0.35 [scCre] to 0.65 [Avr::scCre]). This increase was seen independently of proximity or presence of an Avr-EBE. Interestingly, even the empty scaffold control increased activity. While the TALE-DBD uniformly augmented cleavage for scI-CreI, it interfered with cleavage when fused to scI-CreI_DS. If the distance between Avr-effector binding element and HE recognition site was greater than 10 bp or the EBE is missing, no signal could be detected. The same is true for the control lacking repeats.

Additionally we created the variants scI-CreI_DS-N and scI-CreI_DS-C. Both were hetero domain variants with the first one containing the DS mutations in the N-terminal domain and the second one in the C-terminal domain. Profiling revealed some differences. Both were slightly less active than scI-CreI, but unlike scI-CreI_DS both had saturating activity (see appendix 16). They had increased activity in the TALE context, but like scI-CreI did not seem to depend on the spacer length, suggesting that both subdomains need to be altered for full restriction of activity.

As a proof of concept, we then decided to test our constructs for their ability to promote homology directed repair in a mammalian model. We used a CHO-cell line harboring a chromosomal β-galactosidase

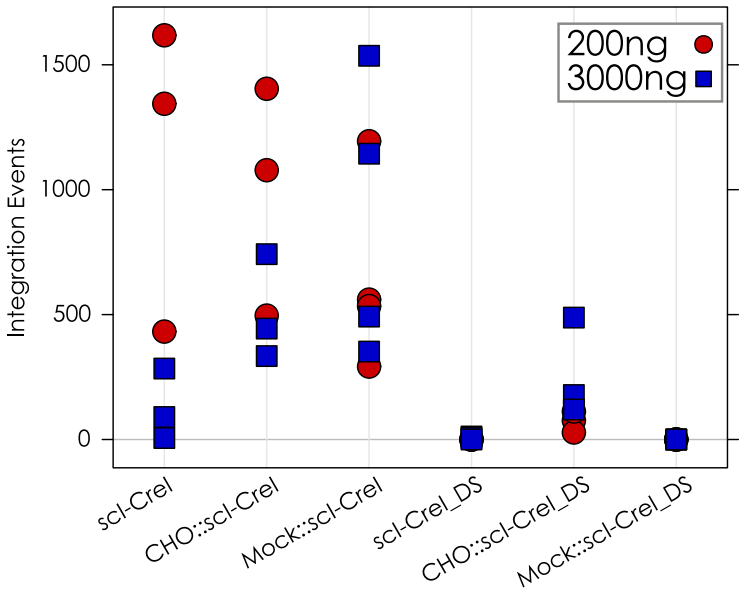



Figure 35: TAL::SCI-CREI exhibits no toxicity at large transfection concentrations

CHO cells, containing an artificial, chromosomal I-CreI site, were transfected with 200 (red) or 3000 ng (blue) of construct plasmid and 2 µg of LacZ repair matrix. Cleavage of the I-CreI site and subsequent homology directed repair yielded LacZ expressing CHO cells. Cells were stained with X-gal after three days of growth. Blue cell clusters were considered as integration events.

SCI-CREI efficiency decreases dependent on the transfection dose (see table 2). There are almost no blue cells on the SCI-CREI_DS plate, however fusion of a TALE domain targeting the outside region of the chromosomal target (CHO::SCI-CREI_DS) rescues activity. Interestingly, there is no discernible difference for the CHO-TALE and the Mock-TALE domain, when fused to SCI-CREI.

gene, which is disrupted by an I-CreI target site. Two different doses of our expression vectors were then cotransfected with a repair matrix. After three days, we fixated the cells and stained them with X-Gal. Plates were then incubated for another day and groups of blue cells were counted.

For SCI-CREI, there were significantly less integration events for the high transfection dose (see figure 35 and table 2). However, this dose dependent effect disappears for both TAL::SCI-CREI constructs. Interestingly, there seems to be little difference in integration efficiency between the TALE array recognizing the chromosomal target region (CHO::) and the TALE-DBD recognizing a mock sequence (MOCK::). This is suggesting that the apparent boost in recombination stems from an overall reduction of activity of SCI-CREI and thereby reduction of toxicity. Contrary to this, the correct TALE-DBD is very important for the degenerate scaffold. While SCI-CREI_DS caused almost no integration events by itself, activity could be restored when fusing the matching array. On the other hand, addition of the wrong TALE-DBD (MOCK::)

Table 2: Mean amount of integration events in transfected CHO cells
Shown are mean amount of events for each construct and dose of transfected DNA. No events were seen when just an empty expression vector and the repair matrix were transfected. Transfection of a LacZ expression vector yielded a large amount of blue cells after staining.
The amount of integration events is significantly ($p = 2.6*10^{-2}$) lower for the 3 µg than for the 0.2 µg transfection of scI-CreI (see figure 36 for effects of toxicity). For the scI-CreI_DS transformations, integration events only occured at higher DNA concentration.

| CONSTRUCT | DNA [µg] | MEAN EVENTS [] | SEM |
|---|---|---|---|
| scI-CreI | 0.2 | 1131.3 | ±358.5 |
| | 3 | 126.7 | ±82.3 |
| CHO::scI-CreI | 0.2 | 992.7 | ±265.6 |
| | 3 | 506.7 | ±121.9 |
| Mock::scI-CreI | 0.2 | 645 | ±192.7 |
| | 3 | 881 | ±278.3 |
| scI-CreI_DS | 0.2 | 0 | ±0 |
| | 3 | 6 | ±3.5 |
| CHO::scI-CreI_DS | 0.2 | 72 | ±24.3 |
| | 3 | 262 | ±114.3 |
| Mock::scI-CreI_DS | 0.2 | 0 | ±0 |
| | 3 | 0 | ±0 |
| Mock | 3 | 0 | ±0 |
| LacZ expression vector | 2 | +++ | |

shut down the construct.

The difference between the single-chain variant and the TAL::MEGA was clearest when directly looking at the petridishes (see figure 36). Not only was the amount of integration events lower but also the overall amount of cells was reduced for the 3 µg scI-CreI samples. There were almost empty regions with only single cells. The round morphology suggested that these cells were either dead or dying. On the other hand, cells transfected with CHO::scI-CreI formed a lawn at 3 µg of transfected DNA. Even at high concentrations, no toxicity could be observed for any of the DS variants. These findings are consistent with viability assays done in CHO cells (see appendix 51).

The next step was to estimate the specificity of our constructs, since the DS mutations were introduced to allow the nuclease domain to address more targets. Therefore, we profiled the two meganucle-

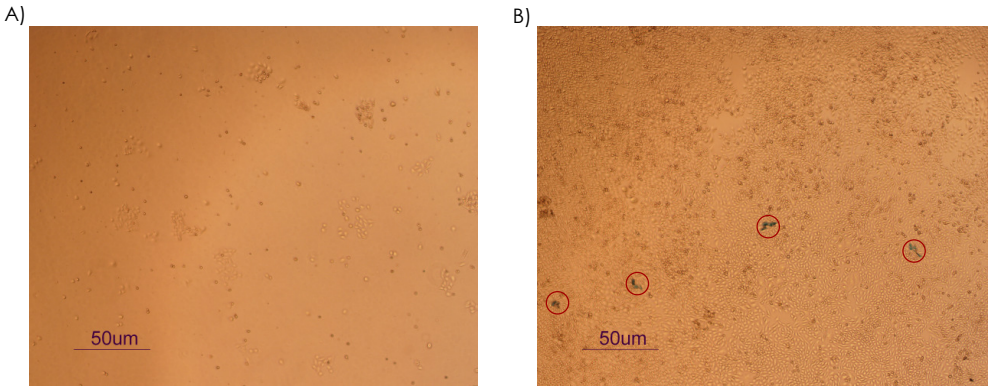



Figure 36: The number of viable cells is low at high scI-CreI doses
Shown are sections of plates containing CHO cells which were transfected with scI-CreI derivates. There were empty patches and overall fewer cells after transfection of 3 µg scI-CreI vector (A), while plates are covered with cells after transfection of 3 µg CHO::scI-CreI vector (B). Cells groups expressing β-galactosidase turned blue (red circle). Each group was considered as one integration event.

ases and their TALE fusions on a library of targets in a yeast SSA-assay (see figure 37). The library consisted of a total of 234 targets, which contained variations of a palindromic I-CreI target site (+12 `TCNNNNCGTCGT-ACGACGXXXXGA` -12). It had been shown previously that the palindromic target is cut at least as good as the pseudo-palindromic WT target [6].

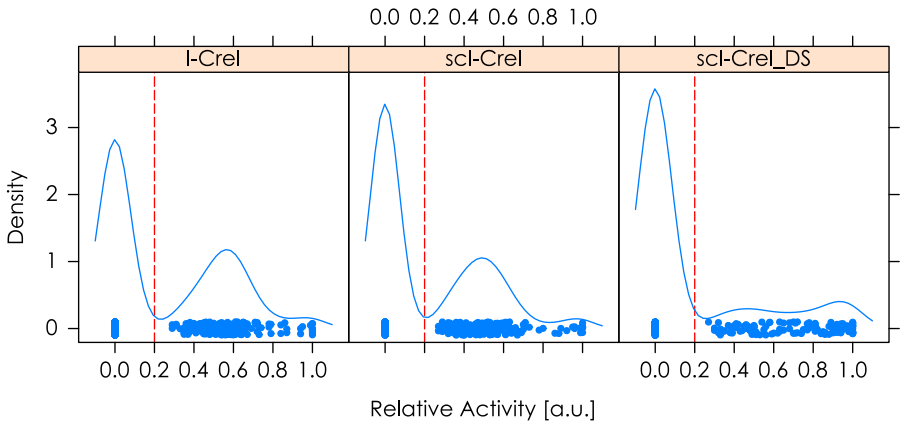



Figure 37: Degenerate scaffold mutations alter the range of addressed targets
I-CreI and two scI-CreI variants are profiled on a library of I-CreI target site derivatives in a yeast SSA-assay. In this library, nucleotides at 10 to 7 bp positions of the palindromic recognition sequences (+12 `TCNNNNCGTCGT-ACGACGXXXXGA` -12) vary. Each dot represents one measurement after 48 hours of incubation at 37°C. Relative activities ≤0.2 were normalized to 0. The y-axis shows the estimated density of the data. I-CreI and scI-CreI show similar profiles. scI-CreI_DS is inactive on more targets, but also cleaves more targets with a high relative activity compared to scI-CreI.

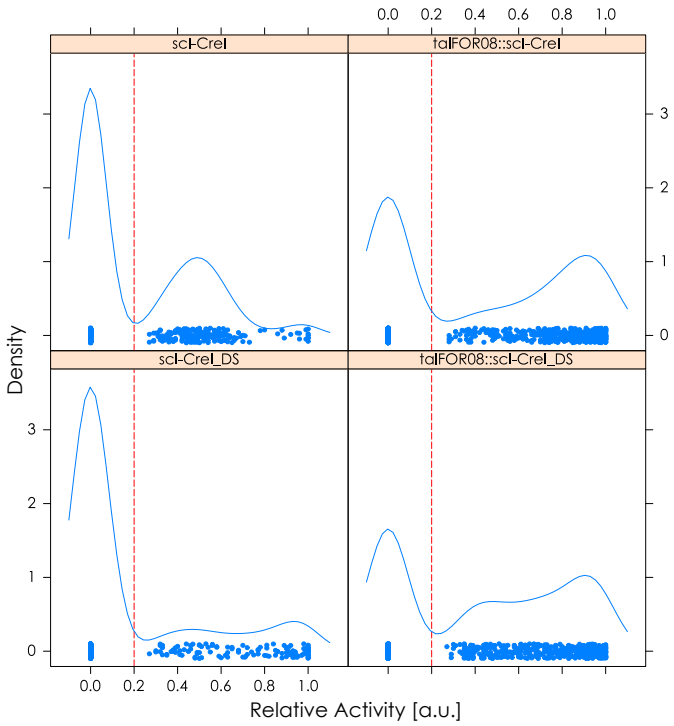



Figure 38: Addition of a TALE-DBD improves a meganucleases ability to tolerate mismatches
The same library as in figure 37 was used to profile TAL::SCI-CREI fusion constructs. The TALE-DBDs recognize a sequence 8 bp upstream (TALFOR08) of the I-CreI sequence. DBD fusion reduces the amount of uncut (<0.2) targets while increasing overall activity.

There seems to be very little difference between the single-chain and the dimer forming I-CreI. Both variants cleaved a small subset of targets well, a larger amount with moderate efficiency (0.5 ±0.2) and the majority of targets were not cleaved at all. The WT was slightly more active than our single-chain construct. This profile differs for the degenerate scaffold. It cuts more targets with a high efficiency, but less targets overall.

TALE-DBD fusion has a positive effect on both single-chains (see figure 38). and increased the mean activity significantly (***, both $p = 2.2 * 10^{-16}$). Detected signals were generally higher and less targets were not cut. TALE-DBD fusion had a more prominent effect on the degenerate scaffold than on the regular single-chain. By itself, SCI-CREI caused DSBs on more targets than SCI-CREI_DS. However after TALE fusion this ratio switches: TALFOR08::SCI-CREI_DS acts on more targets than its predecessor TALFOR08::SCI-CREI.

To have a look at the base preference, the targets can be plotted in a heat map with the bases at position 9 and 10 on the x- and bases at position 7 and 8 on the y-axis (see figure 39). Mean relative activity is then illustrated as a color spectrum.

As previously seen, fusion of the TALFOR08 array increased relative activity for all previously cleaved targets. Activity on some of the un-

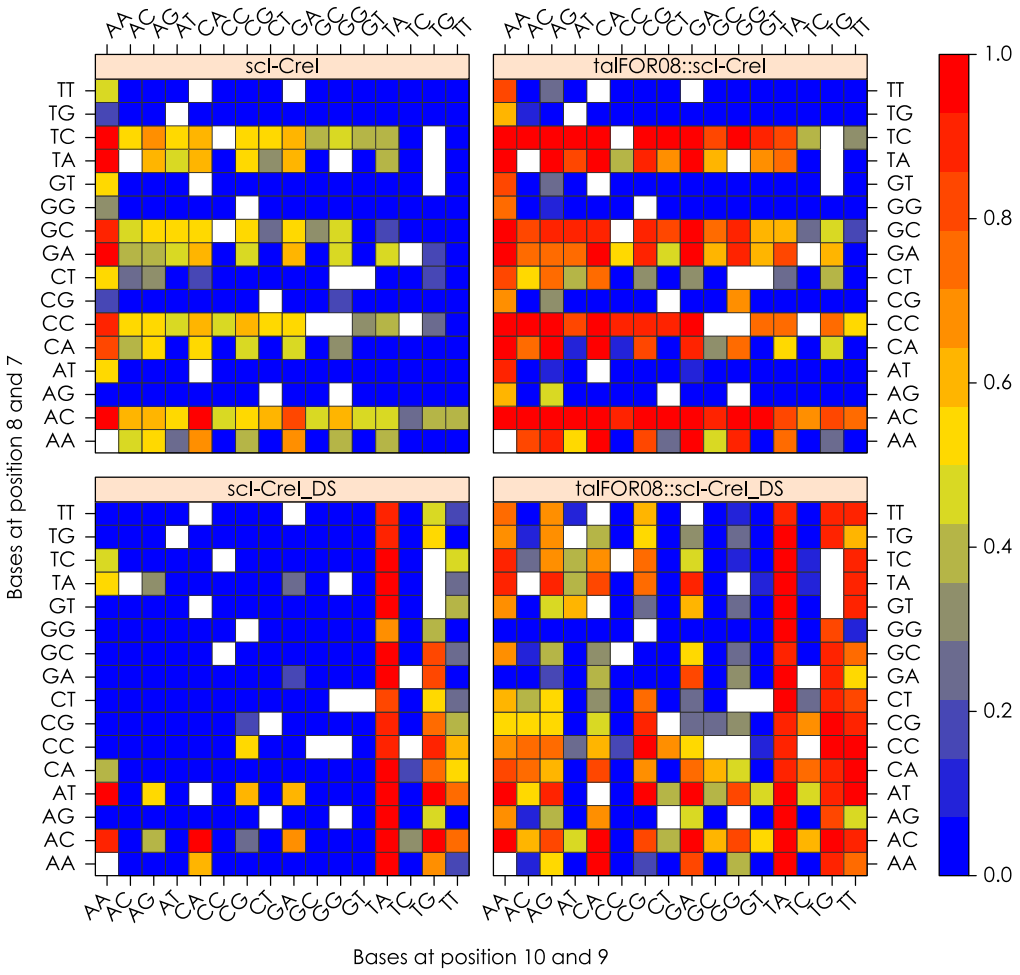



Figure 39: Cleavage profile of scI-CreI and TAL::scI-CreI
All 234 profiled sequences are ordered in a matrix with bases at position 10 and 9 on the x- and bases at position 8 and 7 on the y-axis. Mean values for the relative activity of clones measured after incubation for 48 hours at 37°C are plotted. Red stands for high activity; blue for low or no activity. The cleavage profile of scI-CreI and the degenerate scaffold differs (see figure 40).

cleaved targets could be restored by TALFOR08, however there are targets that remain without detectable activity. More interesting are the differences in cleavage patterns, when comparing the two cleavage domains. Since I-CreI's pseudopalindromic target sequence in *Chlamydomonas rheinhardtii* is "AAAA" and "AAAC" at position 10 to 7 (or -10 to -7 respectively), it is no surprise that scI-CreI had a "AANM" base preference. Adenine is the clear preference at position 10 and thymine is tolerated poorly. A purine base is preferred at position 9, although targets with pyrimidine bases are also cleaved well, if a cytosine is present at position 7. Furthermore, position 7 has the biggest influence on activity and cytosine or adenine are required in most cases. The majority of targets with an "NNNK" (T or G) sequence are not cut, even in the TALE context.

scI-CreI_DS exhibited an almost completely different cleavage pattern. ScI-CreI's rejection of thymine at position 10 changed to a preference in the degenerate scaffold. The importance of an adenine or cytosine (M) at base 7 has been completely removed for the degenerate scaffold. In fact, without the TALE-DBD only few targets that do not posses a thymine at position 10, were addressed at all. The before mentioned preference for a purine base at position 9 could be seen in the TALE fusion context. Interestingly, the pyrimidine base thymine is accepted nicely at position 9, if there is also a thymine at position 10. "TTNN" targets were generally cleaved with a high relative efficiency. A consensus (see figure 40) of well cut targets (relative activity $\geq$0.85) was generated for a simpler overview.

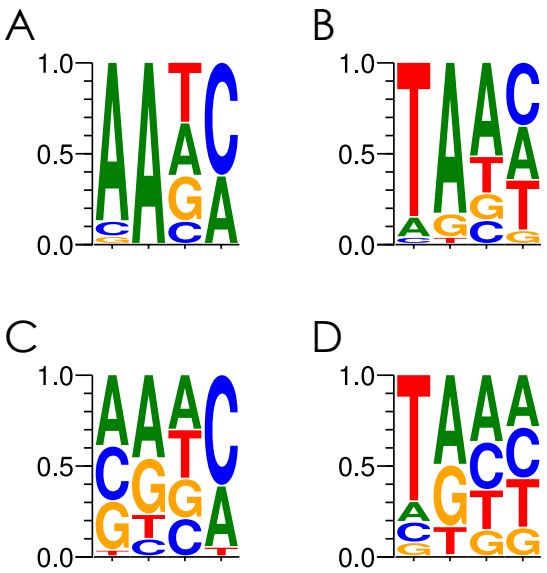



Figure 40: TALE-DBD fusion broadens the 10-7 bp consesus sequence
Shown are probability consensus sequences of the 10 to 7 bp positions in a palindromic I-CreI wt target for scI-CreI (A), scI-CreI_DS (B), talFOR8::scI-CreI (C) and talFOR8::scI-CreI_DS (D). Targets that were cut with a relative activity of $\geq$0.85 were included in the consensuses.

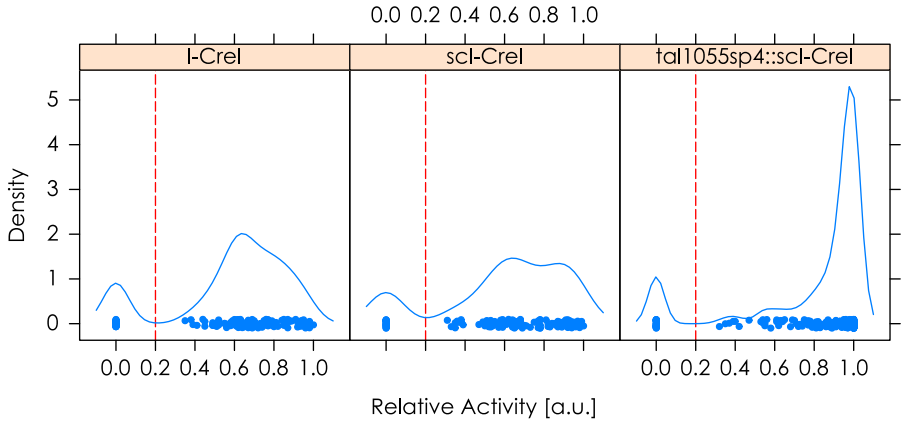



Figure 41: Relative activity is increased after TALE fusion
Constructs I-CreI, scI-CreI and tal1055sp4::scI-CreI were profiled on a library for the cleavage sequence at position -2 to 2. Since the "-12 to -3" and "3 to 12" sequences are palindromic, only 128 target plasmids are needed. Data points obtained after 48h at 37°C were plotted. While there is only a small difference between I-CreI and its single-chain derivative, a larger amount of targets is cleaved at high activity when scI-CreI was fused to a TALE-DBD.

Similar boosting effects could be seen when fusing the TALE array "talREV02::", which bound 2 bp distant from the target site on the opposite DNA strand. Differences between talREV02::scI-CreI_DS and talFOR08::scI-CreI_DS were subtle, but the later had an overall higher activity. One set of measurements for talREV02::scI-CreI did not yield any detectable signals, thereby reducing the mean. Data for this array were therefore not included.

Since TALE-DBD fusion allowed the addressing of a much wider range of targets, we decided to profile our constructs on the -2 to 2 region, which is the area of cleavage. It is thought that, apart from interactions by the catalytic center to the phosphate backbone, there are few protein/DNA interactions in this region [128]. Therefore it is difficult to engineer specificity in this region without affecting cleavage.

*In the "-2 to 2 bp" library, a target plasmid containing the sequence "...TC-AAAA-GA..." contains "...TC-TTTT-GA..." on the complementary strand*

We found that the TALE-DBD had a very prominent effect on activity of scI-CreI (see figure 41). Although the amount of uncut targets did not change much, the number of targets that were cut with a relative efficiency of $\geq$0.85 almost tippled. This change is reflected nicely in the consensus sequence (see figure 42).

For all constructs, the palindromic "GTAC" was the most highly ranked and "GTGA", the "-2 to 2" sequence in I-CreI's natural target, was also visible. Overall, the consensus of well cleaved target sequences was much wider for the tal::scI-CreI fusion. We also profiled scI-CreI_DS, although the 10 to 7 region (here: "AAAA") of our library was not optimal. Only two targets were cleaved at all (GTAC and GTGC), nonetheless the DBD rescued activity for 26 targets with

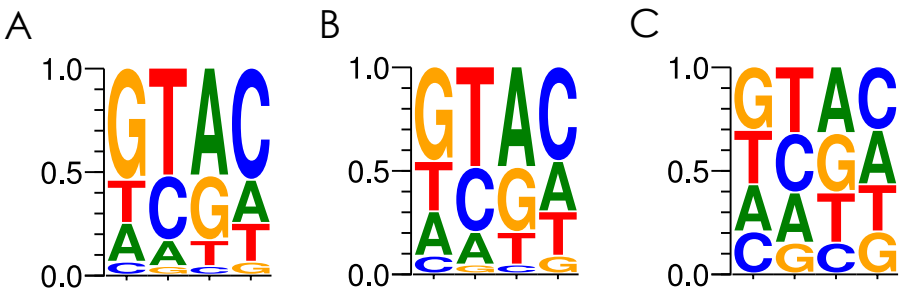



Figure 42: The cleavage consensus sequence is expanded for TALE constructs
A cleavage consensus for I-CreI (A), scI-CreI (B) and TAL::scI-CreI (C) was built by gathering all sequences in a -2 to 2 library that were cut with a relative activity of ≥0.85. The palindromic consensus GATC is clearly visible in A and B; more bases are tolerated in C.

four of them reaching activities above 0.85.

Due to the positive effects that TALE fusion and the degenerate scaffold mutations had, we chose to test both modifications into an already established meganuclease. The engineered I-CreI variant scTcr is targeting the gene for the human t-cell receptor β-chain, but displayed toxicity in mammalian cells. So we introduced the degenerate scaffold mutations and profiled our new meganuclease in a yeast SSA-assay (see figure 43), along with two TALE::MEGANUCLEASE versions of it. Since scTcr is targeting a non-palindromic site, we included palindromic versions of the first (A) and the second half (B) of the target site termed TCR_AA and TCR_BB.

Our scTcr_DS was active on all targets containing a TCR target site, but also on the off target TCR_AA. The TAL::MEGA on the other hand needed two correct DNA motifs to induce a double-strand break: effector binding element and meganuclease recognition site. No activity was detectable, when EBE (TCR- / TCR_AA- Target) or HE recognition site (EBE_TCR-05-Cre / EBE_TCR-null) were missing. There was also a noticeable effect of spacer length between EBE and meganuclease site. TalTCR::scTcr_DS was more active on targets with a small spacer and less active on targets with a longer spacer (>10bp). This effect can also be seen at earlier timepoints (8 h and 24 h). Unlike its relative scI-CreI_DS, it is working at 30°C albeit with much lower efficiency and only on an optimized target (relative activity of ~0.4). Moreover activity was not completely shut down for spacers longer than 8 bp. Fusion of an Avr TALE array inactivated the construct when there was no AvrBs3 site present. Nonetheless, both tested TALE arrays (talTCR:: and Avr::) did not abolish activity on an optimized TCR target with the "-2 to 2" sequence of "GTAC" instead of "GTAA". Here we saw homology directed repair with relative efficiencies between 0.5 and 0.6, even when no EBE was present.

In view of our positive yeast-SSA results and because there was no dose/response related impairment of viability in CHO cells (see

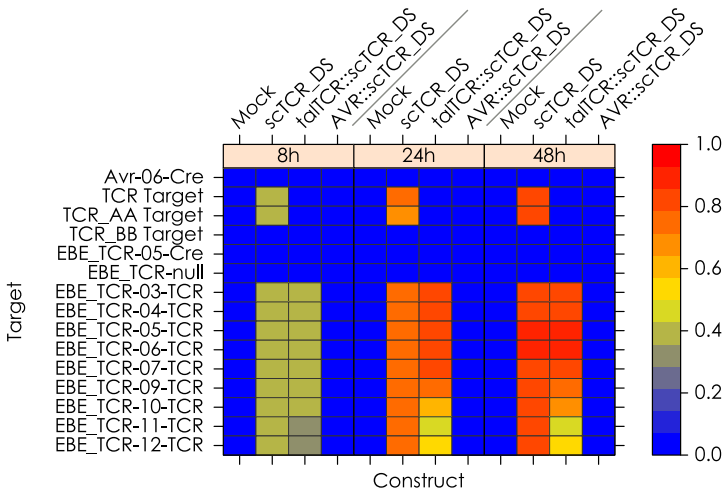



Figure 43: Meganuclease scTcr_DS is more specific and more active after TALE-DBD fusion
Relative β-galactosidase activity as a consequence of DSB induced recombination in a yeast SSA-assay. Blue color signifies no or low activity; red signifies saturation. All values were measured at 37°C. The TAL::MEGANUCLEASE talTCR::scTcr_DS is active on targets containing effector binding element and meganuclease recognition site (EBE_TCR-spacer-TCR); there is no detectable activity on the TCR-target or the palindromic off-target TCR-AA.

appendix 52), we tested our constructs in human embryonic kidney cell line 293 (HEK293) cells. The cell line were transfected with 3 µg of expression plasmids for our nucleases. To enhance the amount of INDELs, 2 µg of a plasmid expressing human 3' -> 5' exonuclease Trex2 were cotransfected in some of the samples. In the case of the TAL::MEGA without Trex2, 9 µg DNA were transfected to determine toxicity. Two samples were taken: one during the first passage (Early) of the cells and another one after seven days (Late). We extracted the genomic DNA and determined the total percentage of insertion and deletion events via next generation sequencing.

The nuclease scTcr caused more NHEJ events than our version scTcr_DS (see figure 44). However in the scTcr samples, the amount of INDELs that are present after seven days (Late) was much lower than at the earlier time point. This is suggesting that cells with an active scTcr meganuclease either died between the two time points or at least divided less. The variant talTCR::scTcr_DS – a TCR-targeting TALE-DBD fused to the degenerate meganuclease scTcr_DS – induced a similar amount of INDELs as scTcr without any of the reduction after seven days. Consistently even with high transfection concentrations of 9 µg (see table 3), no decrease could be seen for our TAL::MEGA. In fact, the amount of INDELs even increased slightly over time.

Without Trex2, our constructs showed little activity, especially scTcr_DS was virtually inoperative when compared to the mock control. Only

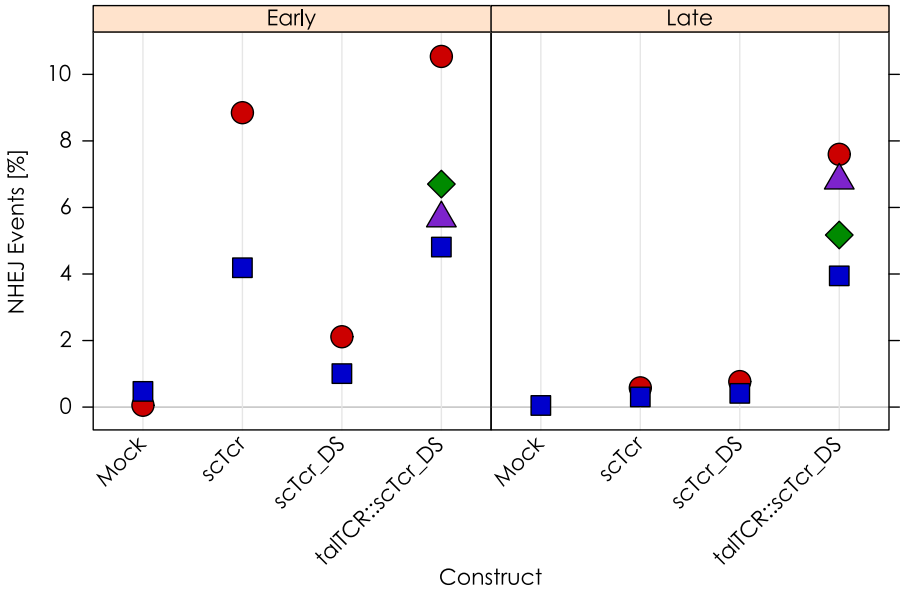



Figure 44: TCR meganucleases induce INDELs *in vivo*

Two plasmids, one harboring the construct and one harboring human exonuclease Trex2, were transfected into HEK293 cells. A sample (Early) was taken between 48 and 60 hours, a second sample (Late) was taken after seven days. Genomic DNA was then extracted and the TcrB genomic region was amplified via PCR. INDELs were then quantified by deep sequencing.

The amount of insertions and deletions is a result of nuclease mediated NHEJ. TALTCR::SCTCR_DS is as active as SCTCR during the first passage of cells (Early); however, the induced mutations are significantly higher ($p = 5.8*10^{-3}$) after seven days (Late).

Table 3: Mean amount of NHEJ events in the TcrB region of transfected HEK293 cells
While the amount of INDELs decreases for the scTcr variant, no significant decrease can be seen for talTCR::scTcr_DS. Cotransfection of Trex2 promotes mutagenic NHEJ by increasing the amount of deletions.

| CONSTRUCT | TREX | SAMPLE | MEAN NHEJ [%] | SEM |
|---|---|---|---|---|
| scTcr | + | Early | 6.52 | ±2.33 |
| | + | Late | 0.44 | ±0.14 |
| | - | Early | 0.64 | ±0.13 |
| | - | Late | 0.31 | ±0.05 |
| scTcr_DS | + | Early | 1.56 | ±0.55 |
| | + | Late | 0.59 | ±0.18 |
| | - | Early | 0.12 | ±0.06 |
| | - | Late | 0.09 | n/a |
| talTCR::scTcr_DS | + | Early | 6.93 | ±1.26 |
| | + | Late | 5.88 | ±0.82 |
| talTCR::scTcr_DS 9 µg transfection | - | Early | 0.67 | ±0.16 |
| | - | Late | 1.10 | ±0.22 |
| Mock | + | Early | 0.26 | ±0.21 |
| | + | Late | 0.05 | n/a |
| | - | Early | 0.27 | ±0.22 |
| | - | Late | 0.04 | ±0.01 |

when examining the sequences directly could one see that scTcr_DS was active without Trex2, since almost all of the positive sequences in two of the Mock controls stemmed from two insertion artifacts.

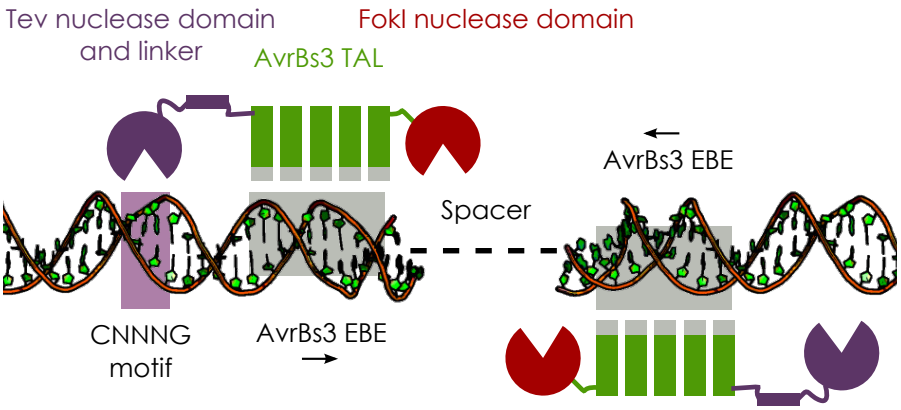



Figure 45: Layout of a Tev::tale::FokI construct binding its target site
A FokI nuclease domain (red) is fused to the C-terminus of an AvrBs3 TALE-DBD (green); an I-TevI nuclease- and part of its linker domain (purple) are fused to the N-terminus. Two EBEs (grey) and an adjacent `CNNNG` motif (light purple) have to be present on the DNA for full activity.

## 2.5 DUAL CATALYTIC TALENS

Seeing that the TALE-DBD proved to be compatible to a large variety of domains and fusion orientations, we chose to design nucleases with two catalytic domains (dcTALEN). Figure 45 shows the dcTALEN Tev::AvrBs3::FokI. Its tripartite architecture consists of the I-TevI nuclease domain and its linker, the TALE-DBD and the FokI nuclease domain. The construct can act as a monomeric TALEN due to the I-TevI domain. However two EBEs are needed to allow the FokI domain to dimerize.

We tested Tev::AvrBs3::FokI in extrachromosomal SSA-assays in CHO-K1 cells (see figure 46). This assays works similar to the yeast version, with one of the differences being that ONPG (o-Nitrophenyl-β-D-galactopyranosid) instead of X-Gal is used to ascertain β-galactosidase activity. Three target plasmids with two AvrBs3 EBEs and differing spacers were chosen: Avr-12-Avr, Avr-15-Avr and Avr-18-Avr. One `CNNNG` site for the I-TevI nuclease domain is located upstream of one AvrBs3 site and downstream of the first LacZ homology region in the Avr-12-Avr and Avr-18-Avr plasmids. The Avr-15-Avr plasmid is based on a different backbone and is lacking this `CNNNG` site. A spacer length of 18 bp is unsuitable for our FokI constructs and causes reduced activity. As a result of this Avr-12-Avr is ideal for both catalytic domains, Avr-15-Avr is suitable for FokI- and Avr-18-Avr is suitable for Tev-constructs.

Tev::AvrBs3 was cleaving the Avr-12-Avr and Avr-18-Avr plasmids, while it was practically inactive on Avr-15-Avr (see table 5). AvrBs3::FokI showed a different pattern: it induced recombination on Avr-12-Avr and Avr-15-Avr, while exhibiting reduced activity on Avr-18-Avr. Tev::AvrBs3 was overall less effective than AvrBs3::FokI. Our dcTALEN Tev::AvrBs3::FokI surpassed both predecessors on all targets. The biggest effect can be

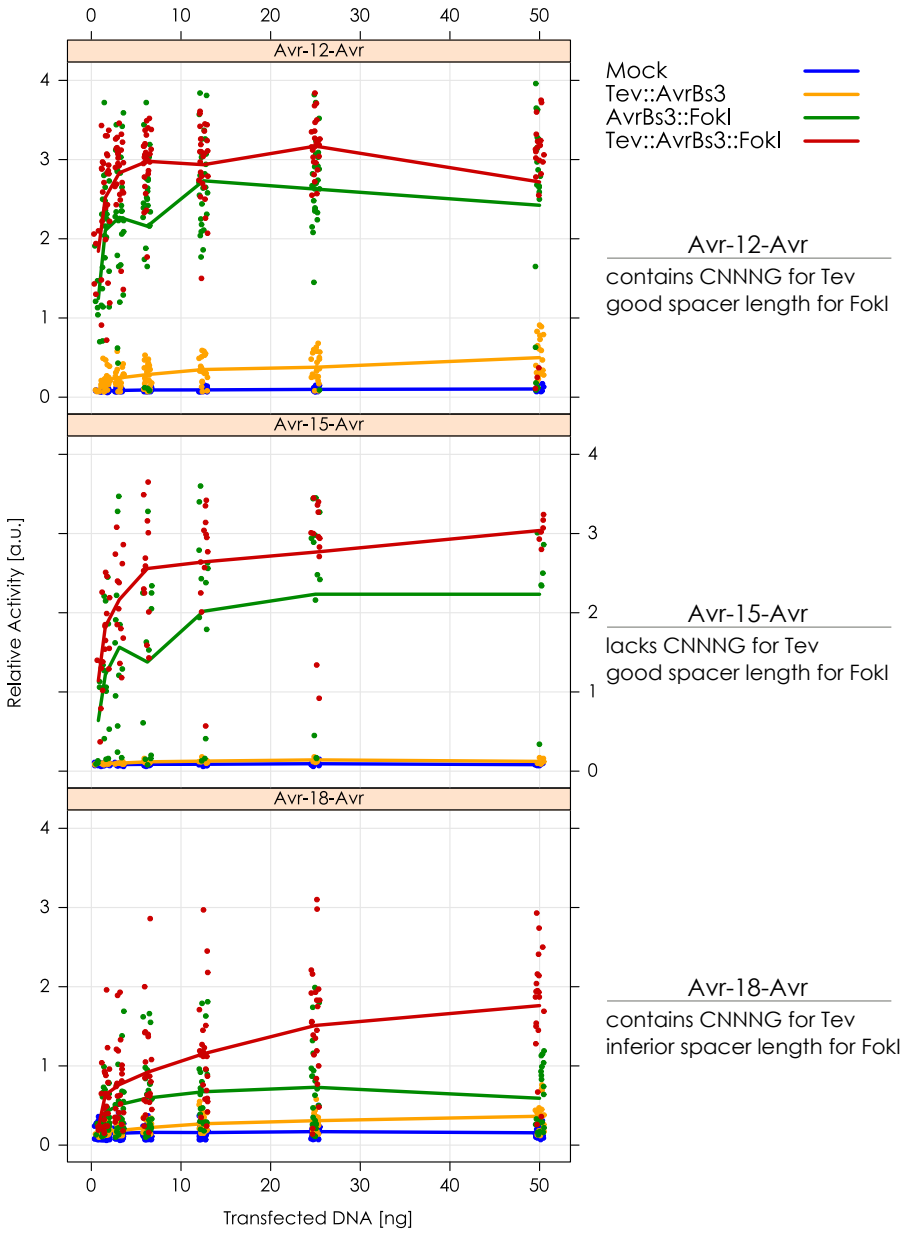



Figure 46: dcTALEN TEV::AVRBS3::FOKI is highly active in CHO cells
Various talen constructs were tested in a CHO-based SSA-assay. A plasmid coded LacZ gene is restored via homology directed repair and relative activity is quantified after ONPG staining. Curves represent mean activity. Target "Avr-12-Avr" is suitable for both catalytic domains, "Avr-15-Avr" lacks the necessary "CNNNG" motiv for Tev and "Avr-18-Avr" has a poor spacer distance for our FokI constructs. Assayed were the constructs TEV::AVRBS3 (orange), AVRBS3::FOKI (green), TEV::AVRBS3::FOKI (red) and an empty vector control ("Mock", blue). A jitter of 0.5 was introduced on the X-axis to reduce overlapping of points; this has no impact on the average curve. TEV::AVRBS3::FOKI induces on average more recombination than AVRBS3::FOKI, even when no suitable CNNNG motif is near the EBE (Avr-15-Avr).

seen on Avr-18-Avr, where the dcTALEN is twice as active as the standard Fok-TALEN. A positive effect of the Tev domain can be seen on the Avr-15-Avr target, even though it is inactive by itself.

Maximum activity and half maximal effective concentration can be determined after regression analysis (see figure 47 and table 4). TEV::AVRBS3::FOKI reached higher $E_{max}$ values than AVRBS3::FOKI, in agreement with the average values. The effect of the dcTALEN was greater than the combined effect of its single components, suggesting some kind of synergistic effect of the two nuclease domains.

We chose to test the inactive mutant Tev_R27A fused to an AvrBs3 DBD and to a AVRBS3::FOKI TALEN, since a boosting effect of Tev also occured on the Avr-15-Avr target, which lacks a nearby `CNNNG` motif. Tev_R27A fused to an AVRBS3 scaffold did not show any activity and was indistinguishable from the empty vector control. Interestingly, there was no difference (Avr-12-Avr: p=0.529, Avr-15-Avr: p=0.121) between TEV::AVRBS3::FOKI and TEV_R27A::AVRBS3::FOKI on the spacer 12 and 15 targets; both exhibited increased activity over AVRBS3::FOKI (see also table 5). On the other hand, TEV::AVRBS3::FOKI was significantly more active than TEV_R27A::AVRBS3::FOKI on the Avr-18-Avr target ($p < 2*10^{-16}$). In fact, the inactive dcTALEN had half of the relative activity of AVRBS3::FOKI. This is suggesting that the boosting effect of the N-terminal Tev domain can be divided into two parts: a DNA-binding (see Avr-15-Avr) and a cleavage effect (see Avr-18-Avr).

In addition we tested a fusion construct containing a single-chain version of human 3' to 5' exonuclease Trex2, which can increase mutagenic NHEJ [58]. This SCTREX::AVRBS3::FOKI construct induced recombination on all three targets despite the addition of the comparatively large scTrex. However, the construct displayed reduced cleavage on the spacer 12 and 15 targets compared to the FokI control.

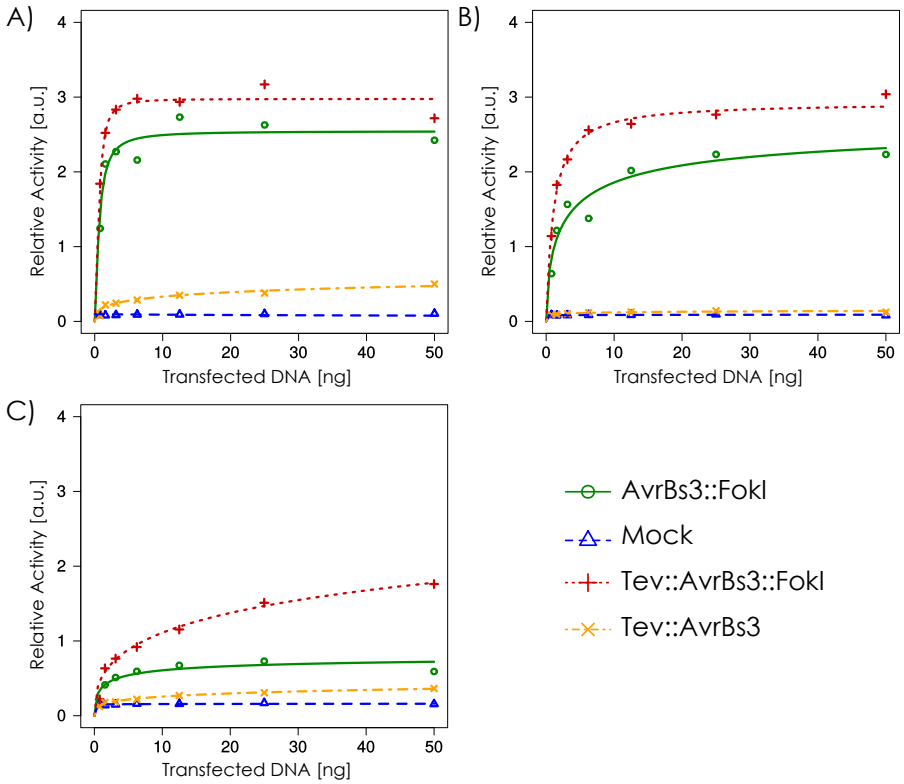



Figure 47: Potency can be estimated via regression analysis
Three parameter logistic models were used to fit the data displayed in figure 46. Resulting regression curves for target plasmids Avr-12-Avr (A), Avr-15-Avr (B) and Avr-18-Avr (C) are shown. Only average points are displayed for clarity reasons. No reliable maximum values could be estimated for Tev::AvrBs3 (orange) since it never reached saturation. Tev::AvrBs3::FokI induced more recombination than AvrBs3::FokI on all target plasmids. Calculated values for $E_{max}$ and $EC_{50}$ can be found in table 4.

Table 4: Regression analysis (see figure 47) yields values for maximal effect ($E_{max}$) and half maximal effective concentration ($EC_{50}$)
The effect of Tev::AvrBs3::FokI is greater than the one of AvrBs3::FokI on the 12 and 15 target, however the $EC_{50}$ values are in a similar range. Since Tev::AvrBs3::FokI did not reach saturation when assayed on Avr-18-Avr, $E_{max}$ and $EC_{50}$were greatly overestimated here. AvrBs3::FokI was much less active when the Avr-18-Avr target was provided.

| Construct | Target | $E_{max}$ [a.u.] | $EC_{50}$ [ng] |
|---|---|---|---|
| AvrBs3::FokI | Avr-12-Avr | 2.54 | 0.63 |
| Tev::AvrBs3::FokI | | 2.97 | 0.61 |
| AvrBs3::FokI | Avr-15-Avr | 2.71 | 2.84 |
| Tev::AvrBs3::FokI | | 2.92 | 1.06 |
| AvrBs3::FokI | Avr-18-Avr | 0.84 | 1.60 |
| Tev::AvrBs3::FokI | | 5.02 | 261 |

Table 5: The impact of the Tev catalytic domain goes beyond its catalytic activity
Shown are mean values at a transfection dose of 25 ng. Tev_r27a::AvrBs3::FokI possessed an inactive Tev domain, but yielded similar activities like Tev::AvrBs3::FokI on the Avr-12 and -15 targets. This effect was converse on the Avr-18 target; here addition of the inactive Tev domain impaired cleavage, while an active Tev domain greatly improved cleavage over AvrBs3::FokI. FokI was still active when the scTrex exonuclease was fused N-terminally.

| Construct | Target | Mean 25 | sem | 95% ci | n |
|---|---|---|---|---|---|
| Tev::AvrBs3 | Avr-12-Avr | 0.38 | ±0.04 | ±0.07 | 23 |
| AvrBs3::FokI | | 2.6 | ±0.14 | ±0.29 | 26 |
| Tev::AvrBs3::FokI | (Tev: "+" | 3.17 | ±0.06 | ±0.12 | 26 |
| TevR27A::AvrBs3::FokI | Fok: "+") | 3.1 | ±0.16 | ±0.35 | 12 |
| scTrex::AvrBs3::FokI | | 1.8 | ±0.41 | ±1.04 | 6 |
| Mock | | 0.10 | ±0.005 | ±0.01 | 20 |
| Tev::AvrBs3 | Avr-15-Avr | 0.14 | ±0.01 | ±0.02 | 9 |
| AvrBs3::FokI | | 2.2 | ±0.36 | ±0.79 | 12 |
| Tev::AvrBs3::FokI | (Tev: "-" | 2.8 | ±0.23 | ±0.51 | 12 |
| TevR27A::AvrBs3::FokI | Fok: "+") | 3.3 | ±0.29 | ±1.25 | 3 |
| scTrex::AvrBs3::FokI | | 1.4 | ±0.22 | ±0.55 | 6 |
| Mock | | 0.09 | ±0.007 | ±0.01 | 13 |
| Tev::AvrBs3 | Avr-18-Avr | 0.31 | ±0.03 | ±0.05 | 23 |
| AvrBs3::FokI | | 0.7 | ±0.12 | ±0.24 | 23 |
| Tev::AvrBs3::FokI | (Tev: "+" | 1.5 | ±0.16 | ±0.32 | 23 |
| TevR27A::AvrBs3::FokI | Fok: "±") | 0.33 | ±0.05 | ±0.12 | 12 |
| scTrex::AvrBs3::FokI | | 0.7 | ±0.19 | ±0.49 | 6 |
| Mock | | 0.17 | ±0.01 | ±0.03 | 40 |

# 3

# DISCUSSION

The main limitation of the nuclease domain FokI lies in its need to dimerize for DNA cleavage. Two effector binding elements are needed and both should respect the guidelines described in [30]: thymine at position 0, no thymine at position 1, no adenine at position 2, T at last position is advantageous and the base composition should be within two standard deviations of the average. Additionally, both EBEs have to be within a defined distance to get optimal cleavage. This is further complicated by the accessibility of a genomic locus [52] and the need to avoid potential off-sites. For a monomeric TALEN, only one effector binding element has to fulfill these prerequisites. We could show in our experiments on the neomycin phosphotransferase II locus, (see section 3.1) that a dimeric TALEN pair can be impaired by its weakest DBD and that a monomeric TALEN containing the better DBD can circumvent this. Monomeric nucleases can broaden the range of suitable loci and can also act as tools to determine the quality of single TALE-DBDs. However, the even greater advantage of these TALENs over regular TALENs is their size. Some viral vectors are difficult to use for regular TALENs. Adeno-associated viruses can only carry a limited payload of DNA [87], which is too little for two TALENs (>2.8 kb each), a promoter and a repair template DNA. The larger non-integrative lentiviruses would be able to accommodate a TALEN pair, but are unable to package the highly repetitive nucleases. A small size remains a big advantage even when other methods of delivery are used. Transfection quality mRNA in sufficient quantities and viral vectors in general are expensive to produce. So monomeric TALENs halve the cost of production and make some types of transfections feasible.

A CATALYTIC DOMAIN should fulfill several prerequisites, to be suitable for a fusion construct. Information about

- structure
- biochemical properties
- biological function and interactions

should be available. A crystal structure is needed to choose suitable fusion orientations and linkers. Biochemical information is necessary to select residues for mutagenesis and to predict activity (E. g., "will the construct be active at a certain temperature?","how fast is the turnover?"). And lastly, the biological context is required, to determine possible interaction partners. A different strategy would be to

pick endogenous nuclease domains. The advantage of this approach is that the nuclease domain is already adapted to the host organism and it is obviously possible to control the cleavage reaction somehow. Additionally, the nuclease domain could recruit factors that are beneficial for the processing of the DSB. However one would need to show, if these advantages even apply in the TALEN context and how many species are covered by a specific construct. Another problem is the aforementioned recruitment and regulation. This recruitment could also work in reverse, i. e., the TALEN could be recruited via its catalytic domain. It is not clear if the chimeric nuclease could then still assume the function of its natural predecessor. Moreover, the nuclease domain could be down-regulated or even be inactivated by the cell. Another question is, if the function and interactions of the target catalytic domain are really understood well enough.

For these reasons, we chose heterologous catalytic domains that preferably do not interact with other proteins in the target cell.

### 3.1 COLE7, NUCA AND ENDA

On paper, unspecific nuclease domains are the ideal fusion partners for DNA-binding domains, since they do not reduce the amount of addressable sequences. Ideally, the nuclease domain is guided to its target by the TALE-DBD, creates the double-strand break at a well-defined distance to the EBE, dissociates and is then quietly degraded by the proteasome. Unfortunately, off-target activity exists. Part of the off-target activity is owed to the specificity of the TALE-DBD [106]. Since some RVDs can bind multiple bases and since repeat arrays can tolerate single mismatches, off-sites resembling the target site are bound with lower activity. These types of off-sites can be minimized with careful selection of the target-site and knowledge of the target genome.

Another type of off-site activity has not been well studied, since it is prevented by the obligatory dimerization of FokI in standard-TALENs. TALE-DBDs need to bind the DNA to find their target via facilitated (1D-) diffusion [51]. During this process, the DBD slides 1-dimensionally along the DNA, "pauses", dissociates and rebinds to the DNA ("hopping"). While the TALEN is "searching", it is bound to DNA, but not to its recognition sequence. Therefore, nearly any cleavage occurring while searching will be at an off-target. Off-target activity has been described for other monomeric TALE-fusions such as methyltransferases or demethylases [138, 130].

More off-site cleavage arises as a result of the linker connecting DNA-binding and catalytic domain. A flexible linker can allow cleavage *in trans*, as has been already suggested for FokI based SSNs [123, 89]. A rigid linker on the other hand can impair cleavage, since the nuclease module has to be at the right angle and orientation to bind

its DNA. For these reasons, a nuclease domain should be designed to only cleave when it has been in proximity of a certain DNA for a "longer" period, which is the case when the DBD is bound to the EBE. This can be achieved by reducing its affinity or by slowing down catalysis.

The unspecific catalytic domains used in this work are well characterized members of the ββα-Me family of nucleases. Hence, there is structural and biochemical data available (see section 1.1.2), which was resorted to for design of the constructs and mutagenesis. Modification of the relevant catalytic or binding residues allowed us to adjust the activity of our fusion constructs.

COLICIN E7 AND NUCLEASE A -based TALE-fusion plasmids proved to be difficult to clone and amplify in *E. coli* DH5-α. At first glance, it is surprising that the constructs were apparently toxic, even though expression had not been induced. In fact, the nearest promoter in front of the fusion proteins is >1 kb upstream, meaning that leaky expression should be extremely low. Like typical cloning strains, DH5-α is a recA mutant that is therefore deficient for homology directed repair. This prevents it from recombining highly repetitive sequences (e. g., TALENs), but slows growth and makes it more susceptible to DNA damage (e. g., by heterologous nucleases). More evidence for this is that mutagenesis of catalytic residues increased the DNA yield and inactive variants had the same yields as the shuttle vector. New shuttle vectors, that coexpressed the respective inhibitors, solved the problem and allowed easier manipulation of the plasmid (see section 2.2). Yeast on the other hand seemed to tolerate the nucleases well and could also be used to clone the inhibitor-free yeast expression vectors. Nonetheless, mutation of ColE7 residues suggests that there is also a cytotoxic effect. Mutation of K497A, D493Q and H573E should reduce the catalytic activity of the nuclease *in vitro* [208], but showed increased recombination efficiency *in vivo*. These three mutants are less active and as a consequence also less toxic. Interestingly, AvrBs3::ColE7_H573E did not cleave all targets and was inactive on some, while AvrBs3::ColE7_H573A cleaved the targets uniformly. An explanation can be the size of the replaced residue. While the large glutamate sterically clashes with the bases, the smaller alanine does not. On the other hand, both nucleases show a very similar cleavage pattern at 30° C, suggesting that the targets they cut share some characteristics.

More interesting, however, are the interactions of the site-specific nucleases with their respective inhibitors. The ability to regulate the activity of a nuclease is an advantageous trait. This way a time window in which SSNs are active can be assigned, so that off-site cleavage is kept to a minimum. One way to achieve this is the generation of photo-switchable restriction enzymes [188], whose underlying design

principles could also be carried over to SSNs [244]. However, this class of photo-sensitive proteins has been chemically modified and cannot be simply expressed *in vivo*. A different approach has been recently described for RGENs. Here, a cell-permeable molecule binds to an added intein, thereby converting the Cas9/intein chimera from an on- to an off-state [55]. Our approach for TALENs simply required expression of the natural inhibitor. No activity could be detected in our SSA-assays when the inhibitors were coexpressed, which is not surprising given the tight binding of the inhibitors. These are the first examples of TALENs that can be shut off.

The next logical step was testing their activities as monomers on a relevant target. Targeting the neomycin phosphotransferase II (*nptII-*) gene was a sensible choice, since it is one of the most widely used resistance markers in plant biotechnology. Removal of this marker can improve biosafety and reduce regulatory concerns. Constructs we tested were active in yeast and worked as cTALENs. I-TevI-based TALENs will be covered in the section 3.2. WT NucA was slightly more active than the mutated ColE7_K497A catalytic domain we used, nonetheless both were active on all targets that contained their respective EBEs. One NucA TALEN cut a plasmid without the fitting EBE, which was the only case of off-target cleavage in this experiment. More interestingly, we saw that SSA-experiments in yeast did not correlate well with SSA-experiments in tobacco protoplasts. The two tested ColE7-based nucleases were highly active in protoplasts and exhibited off-site activity <1%. In fact, our FokI-controls showed clear off-site activity, although just one EBE was present for the TALEN pair. This can be seen as a prime example of one of the previously described off-site cleavage types of FokI. Plant expression vectors for NucA could not be generated at a high enough concentration – again highlighting the toxicity of this protein.

There are different possibilities to explain the disparity in activity between yeast and plants. Tobacco protoplasts simply provide different reaction conditions than *S. cerevisiae.* E. g., the protoplasts were grown at temperatures below 30°C and contain different concentrations of divalent metals and ions in general. Altering the temperature of a cell for a shorter period could be a promising way to regulate nuclease activity. This is probably more feasible for plants and microorganisms than for homoiothermic animal cells.

Although fusions of the ColE7 variant K497A_v2 (see figure 13) worked in a higher eukaryote, they did not show any activity in mammalian cells. Our working hypothesis included two scenarios. ColE7_K497A_v2 TALENs still possessed a rest of off-site activity, which could be detrimental for the cells. Another possibility was the proteolytic cleavage of the linker connecting the TALE-DBD to ColE7. Full Colicin E7 consists of an C-terminal catalytic domain – here referred to as ColE7 – and N-terminal translocation and receptor-binding do-

mains, which transport the catalytic domain through the bacterial periplasm to its target. The catalytic domain is then cleaved off between lysine 446 and arginine 447. Both residues were still present in our construct. Still, the rest of the protease recognition sequence and the periplasmic protease itself are missing, making this scenario unlikely. The perceived toxicity in our cloning strains can also not be fully explained by this. A ColE7 nuclease domain could hardly enter an *E.coli* cell, even if it was cleaved off a TALEN in the periplasm, since the receptor-binding domain is not present in our construct.

Despite this scenario being unlikely for mammalian cells, we mutate the two residues. Arginine 447 has also been described to be very important for DNA-binding [194], which fit our idea of reducing affinity to limit off-site activity. As expected, mutation of the residue 446 had little impact. This residue is cleaved off in nature and therefore plays no role for catalysis or binding. In the TALEN context, mutation of this residue is basically equivalent to changing residues in the linker. Residue 446 can alter binding only indirectly by influencing residue 447 or by changing the orientation of ColE7 to the binding domain.

Residue 447 on the other hand had a bigger impact. The constructs lost most of their activity after non-conservative mutation of arginine 447 to alanine, which is in agreement with the findings of Németh *et al.* [160], who used the non-conservative R447G mutation on the free catalytic domain. However, our conservative mutation to lysine removed all detectable off-site activity, while keeping an on-target activity comparable to the most active constructs. The only drawback being a loss of activity on some targets. Which type of on-targets are affected remains unclear and needs to be further elucidated in benchmarks.

The nuclease AvrBs3::ColE7_R447K showed no activity in mammalian cells, despite its promising properties in yeast. The intracellular concentrations of $Zn^{2+}$, the co-factor of ColE7, lie between 5 pM and 1 nM in mammalian cells and in the femtomolar range for *E. coli* [49]. Zinc concentrations for plants vary over a large range [198]. ColE7 is also able to accept the more ubiquitous $Mg^{2+}$ ion, anyway. While the differences in metal ion concentrations cannot satisfactorily explain the discrepancy between our model organisms, there are still the aforementioned variations in the cytosolic composition and reaction temperature.

The importance of the fusion terminus was another point of interest. While there have been fusions to the TALE N-terminus in the dimeric FokI context [105], there is little known for monomeric nucleases. For this reason, we fused WT ColE7 to the AvrBs3 N-terminus and compared it to the C-terminal fusion. The profiling in yeast SSA-assays revealed increased activity compared to the C-terminal fusion. Relative activities for the two EBE targets shifted from 0.3-0.6 for

AvrBs3::ColE7 to 0.4-0.7 for ColE7::AvrBs3. Similar increases were seen for the single EBE targets. What is surprising are the activities on unaddressed target plasmids. Two populations could be seen for ColE7::AvrBs3, with one population not being cut at all and another population being cut with the same efficiency as monomeric targets. On the other hand, for Avrbs3::ColE7 there was only one off-target population, whose members were cut with low efficiency. This could be an artifact, which can occur for values near the cutoff for activity. Quantification of the relative activity is not possible for values below 0.2. While yeast SSA-assays are useful to determine the fidelity of a construct, they are ill-suited to quantify toxicity. No dose/response dependent toxicity is shown in this setup and dead yeast cells are not looked for. We could not test our ColE7 constructs in a standardized CHO-based toxicity assay, due to them not being active in mammalian cells. *In vitro* assays were needed to further investigate a cTALEN with an unspecific nuclease domain. For this, we chose Endonuclease A from *S. pneumoniae.*

ENDONUCLEASE A variants, which could be chemically rescued had been described [148]. Here, mutation of the catalytic histidine to glycine creates a cavity, which can be occupied by imidazole to replace the missing histidine. This way one can express the nuclease, without damaging the cell. Activity is then later restored via addition of imidazole. Chemical rescue of NucA had also been demonstrated, albeit with much lower activity than for EndA [148]. This concept can not only be used to facilitate expression and investigation of the nuclease but also allows the aforementioned time-dependent control of the nuclease.

We fused EndA H160X to an AvrBs3 N-terminus, due to the perceived higher activity of ColE7 in that context. As expected, protein could be expressed in large quantities. The purified cTALEN could then be characterized biochemically. However, the same problems that were seen for ColE7 TALENs also occurred for EndA TALENs after activation. In the assays, DNA fragments stemming from specific cleavage events appeared early, but were then fully degraded. It would be an interesting experiment to sequence the unspecific band, to see if it is an AvrBs3 EBE, which was protected by the DBD.

Mutation of binding and catalytic residues reduced the problem of off-site activity, showing that the untargeted activity emanated from the TALEN and not from a nuclease contamination. We saw that EndA_H160A::AvrBs3 was less active than EndA_H160G::AvrBs3, which is consistent with the observations made by Midon *et al.* for the nuclease itself. The mutations Q186A and N202A greatly reduced off-target activity. The constructs were also not impaired after fusion of a C-terminal SNAP-tag, which could allow monitoring of the nuclease *in vivo* via fluorescence labeling.

Activity of EndA TALENs could also be modulated by external factors. Lower temperature means reduced molecular motion and therefore allowed us to investigate the H160G mutant without it degrading the target plasmids immediately. Besides one can adjust the ion concentration. A high ion concentration can "mask" the DNA, thereby reducing binding of DNA-binding proteins. A more interesting option to adjust the activity is the imidazole concentration, since other factors can or should normally not be changed for *in vivo* applications. EndA TALENs were active over a long millimolar range. The cavity in the catalytic center is occupied less often, if the imidazole concentration is too low. If it is too high, then the imidazole will interfere with DNA binding.

Another point of interest is the cleavage mechanism of monomeric TALENs. For this reason, a double-fluorescently-labeled DNA was cleaved during a time course assay. Half of each sample was analyzed via standard PAGE, the other half was analyzed by denaturing PAGE. This way one could determine and compare the amount of nicks and DSBs. Different fluorophores allowed the discrimination of top- and bottom strand cleavage. No strand preference could be seen for the two constructs EndA_H160G_Q186A_N202A(GAA)::AvrBs3 and EndA_GAA::AvrBs3::SNAP. That means both strands are either cut with equal probability or cleavage of the second strand occurs shortly after cleavage of the first cut strand. To see if this is the case one can compare the amount of uncut DNA in the PAGE and the denaturing PAGE. Nicks and DSBs will be seen in the denaturing PAGE, while standard PAGE only shows DSBs. The amount of pure nicks can then be obtained by subtraction. While there was no difference in specific activity for the two tested constructs, there was a difference in nicks. EndA_GAA::AvrBs3 caused significantly more nicked intermediates than the SNAP-tagged version. This could mean that EndA_GAA::AvrBs3::SNAP either dissociates less often after making the first cut or that the second cut is made faster after the first one. The SNAP-tag is a synthetic version of the human $O^6$-alkylguanine DNA alkyltransferase. It is 19.4 kDa in size and was engineered not to bind DNA [83]. Therefore, interactions of the SNAP tag to the TALE-bound DNA should be unspecific in nature, although the possibility of some residual affinity for DNA in this context cannot be excluded.

To be able to directly compare the influence of the two fusion termini, we fused two EndA modules onto the TALE-DBD. The construct EndA_H160G_Q186A_N202A::AvrBs3::EndA_H160G_Q186A_N202A was then used in the previously explained fluorescence assay. The great advantage of this setup is that the two domains are present in equal purity and equimolar concentrations and can therefore be directly compared. However, one has to keep in mind that the results are only valid for the specific polypeptid linkers that were used and that there could be a context dependence between the two catalytic do-

mains (i. e., one domain could influence binding of the domain on the other side). The specific activity of this dcTALEN was lower than for the N-terminal EndA TALENs, which does not necessarily mean that the actual activity is also lower. Nevertheless, clear differences between the two EndA domains could be observed. The N-terminal EndA had no strand preference and generated more DSBs than the C-terminal one. The reason for this seems to lie in the strand preference of C-terminal EndA. It cleaves the top strand faster than N-terminal EndA, but it barely cleaves the bottom strand. Bottom strand cleavage is around five fold lower than for N-terminal EndA, despite the fact that in this assay bottom strand cuts for N-terminal EndA are underestimated. An explanation for this could be a difference in flexibility of the two catalytic domains. E. g., the C-terminal linker could prohibit EndA from reaching the bottom strand. The fusion terminus is discussed further in section 3.2.

Since the ameliorated constructs showed very little off-target cleavage *in vitro*, we reverted the H160G mutation to remove the imidazole activation. While imidazole is not overly toxic ($LD_{50}$, mouse, oral: 880 mg/kg), it induces programmed cell death in some tissues [94] and is considered harmful in general. Therefore, it would not be suitable for application in mammalian cells.

EndA_H160_Q186A_N202A::AvrBs3 could not be cloned, suggesting that leaky expression is already too toxic for the cells, similar to what was observed for ColE7 and NucA TALENs. Since the reversal of the mutation was not possible, we used conservative mutation of histidine 160 instead. Asparagine and glutamine both contain an amino group that could possibly activate the catalytic water for the cleavage reaction, but not as well as the histidine. Indeed, the EndA_Q186A_N202A::AvrBs3 variants H160Q and H160N could both be cloned, showing that they are less toxic. The glutamine variant could not be expressed, suggesting that it is active, but still too toxic for the expression strain. Glutamine has a comparable size to histidine and the amino group should be positioned at a similar position as the imino group of the histidine. The asparagine variant EndA_H160N_Q186A_N202A::AvrBs3 could be expressed and purified, but was not active *in vitro*. A possible explanation is that the amino group of asparagine is not arranged properly for the activation of the nucleophile. Histidine is hard to substitute due to its special properties, but cysteine can be considered as a more conservative replacement and could be tried.

## 3.2 I-TevI

Though the use of the I-TevI nuclease domain seems to be only a variation of the previous approaches, there are some ameliorations over the unspecific domains we tested. I-TevI can be divided into a

DNA-binding domain, a linker and a nuclease domain. The nuclease domain therefore already possesses natural adaptations for a fusion. One of them being its CNNNG specificity and another one being the linker, who has DNA-binding affinity and positions the catalytic domain at the right distance to the binding domain. Both these features may also reduce off-site activity in chimeric constructs.

Fusion of the I-TevI nuclease domain and the linker to a TALE-DBD N-terminus proved to be active in yeast, plants and mammals. These findings are consistent with what has been described for TEV::zinc-finger fusions [117] and TEV::Meganucleases [238]. The only drawback of the I-TevI architecture lies in the requirement for a CNNNG motif, which has not been fully characterized yet. I. e., there seem to be some NNNs that are cut well, while others are cut with very low efficiency [21, 115]. There also seems to be a sequence preference of the linker domain, causing a small subset of targets not to be cleaved, which contrasts with previously described findings [57]. Despite benchmarking of the Tev domain requiring more work, compact TEV::TALENs could be used for genome editing with good activity in multiple organisms. TEV::TALEs could even be purified and tested *in vitro*. Wild-type I-TevI is highly toxic and cannot be expressed in the usual way [239]. This was not the case for our constructs, suggesting that the TALE is a "safer" DBD than the natural I-TevI DBD. Interestingly, no toxicity could be observed *in vivo* [21], while unspecific degradation was seen *in vitro*. This could be a nuclease contamination, since the TEV::TALENs were only purified via NI-NTA. On the other hand, it could also be a consequence of the reaction conditions. We also saw similar *in vitro* degradation of DNA for FokI TALENs (see appendix 50), which are considered safe in general. *In vitro* experiments are useful to investigate cleavage patterns and to elucidate cleavage mechanisms, but they do not take into account the DNA-repair systems, the proteasome or the chromatin structure of a living cell. Additionally, enzyme concentrations have mostly been in excess to substrate. Therefore, toxicity of a nuclease can be predicted *in vitro*, but has to be determined *in vivo*. Several suitable techniques to profile off-site activity of SSNs have been described recently [232].

THE DIFFERENCE BETWEEN THE TWO FUSION TERMINI could be seen especially well with I-TevI fusions. These fusions worked as cTALENs, when the catalytic domain was fused in the natural order (i. e., N- catalytic domain - linker - DBD -C). On the other hand, fusion to the C-terminus yielded a construct that was active as a nickase, so that two nucleases were needed to get a double-strand break. Contrary to TALE::FOKI, a TALE::TEV monomer could induce recombination *in vivo* (see [21]), showing that activity stemmed from single-strand nicks and not from a dimerization of two Tev domains. These observations further support what was found for the EndA fu-

sions. Flexibility of the nuclease domain seems to be greater at the N-terminus than at the C-terminus of the TALE-DBD. A cleavage mechanism has been proposed for homing endonuclease I-BmoI, which is also a member of the GIY-YIG family. It is thought that the nuclease domain of I-BmoI first cleaves one strand and then rotates to cleave the other strand [116]. Maybe a similar rotation is impaired for the Tev catalytic domain by the C-terminus.

## 3.3 MEGANUCLEASE FUSIONS

At first glance, meganucleases seem to be an unconventional choice of nuclease domain, since they are already highly specific by themselves. In fact, the recognition sequence of I-CreI (22 bp) is longer than the one of AvrBs3 (17.5 bp). It is therefore no surprise that meganucleases of the LAGLIDADG family have been some of the first enzymes to be applied for genome engineering (see section 1.2.2). The dimeric nature of the standard TALE::FokI architecture reduces the amount of off-target cleavage by requiring two DNA-binding modules. By using a meganuclease as catalytic domain, we integrated this second DNA binding module into the nuclease domain, creating a monomeric version. The price of slimmer constructs are three main drawbacks of meganucleases: reprogramming to the desired target site is not as straightforward as for TALENs or RGENs, since the whole DNA/protein interface has to be engineered (I); the ability to tolerate single base pair mismatches can allow constructs to cleave off-sites (II) and the range of addressable sequences is intrinsically limited by the meganuclease scaffold (III). Therefore, creating an application-related meganuclease remains a time- and work-consuming task, although there have been multiple approaches via rational design or selection [219, 99].

The concept of a DBD::MEGANUCLEASE fusion already occurred at least once in nature: the homothallic switching endonuclease in *Saccharomyces cerevisiae*, involved in mating type switching, is a meganuclease of the LAGLIDADG family with a short DNA-binding zinc-finger domain [12]. Presence of the zinc finger is essential for activity. TALE fusions to the LAGLIDADG HE I-AniI have been described by Boissel *et al.* [27].

Our constructs can mitigate the shortcomings mentioned above. The difficulties in tailoring a meganuclease to the right target-site can be circumvented by using a low-affinity variant with a broader specificity. Precision is ensured by the TALE-DBD, which is guiding the meganuclease. Thereby, the nuclease domain can tolerate different targets, but will only cut those that are adjacent to a fitting EBE. Proof of this can be seen in figure 34. Here, our broad specificity variant scI-CreI_DS displayed reduced activity, which could be rescued to WT activity by the AvrBs3 TALE-DBD. Moreover, the TALE-DBD pre-

vented scI-CreI_DS from cleaving target sequences that were lacking an adjacent EBE. No activity could be detected when the EBE was not present or further away than 8 bp. While the spacers with a length between 0 and 8 bp did not show any difference in our assay, Boissel *et al.* [27] found that only spacers between 5 and 10 bp worked satisfactorily; with a 7 bp spacer working best. This can be explained by a difference in architectures: NΔ154/C+63 (TAL), Glycine/Serine-based linkers and I-AniI (MEGA) vs NΔ153/C+11 (TAL), I-TevI-derived linker and scI-CreI (MEGA). Since the distance between the last TALE-repeat and the nuclease domain is larger, it is no surprise that the Boissel construct cannot tolerate small spacer lengths. Interestingly, we did not see any spacer- or even EBE-dependence for WT TAL::scI-CreI, even though DNA affinities for I-CreI ($K_d$: 7 ± 1 nM [192]) and I-AniI ($K_d$: 8.4 ± 3 nM [33]) are thought to be in the same range. The different C-termini and linker seem to have a larger impact on I-AniI than on scI-CreI.

It is not surprising that the TALE-DBD could not prevent the original scI-CreI from cutting, given that a detergent is needed to remove I-CreI from its target site *in vitro* [230]. It could however reduce the amount of off-target activity, and thereby cytotoxicity in a mammalian assay (see figure 2 and 36, see appendix 51).

Interestingly, the reduction of cytotoxicity could also be achieved by fusing a binding domain that did not target a genomic region. Affinity to off-sites is by definition lower than for a main target site. Interference of the binding domain is not enough to have an impact on cleavage of I-CreI's recognition sequence, however, it seems to be able to impede cleavage of the lower affinity off-sites. Not necessarily only by binding its own EBE, but maybe also by unspecific DNA binding during target search or by other types of steric interference.

When interpreting the degenerate scaffold results one has to take into account that the recognition sequence changed with introducing our mutations (see figure 39). Meaning that the WT recognition sequence represents an off-target for scI-CreI_DS which accordingly is cut less often. The activity on this off-target can now be increased by fusion of the TALE-DBD. Here, the TALE-DBD will bind its EBE, which is located adjacent to the homing endonuclease recognition site. This increases the local concentration of the nuclease domain, which consequently increases the probability of cleavage.

In continuative experiments we evaluated the differences between the specificities of scI-CreI and scI-CreI_DS as well as the target range of our fusion constructs. For this reason, we first focused on a library of target plasmids that had varied the 10 to 7 (-10 to -7) region of the idealized palindromic I-CreI target. Our monomeric I-CreI was only marginaly less active than its dimeric predecessor, which was also the case in [64]. The degenerate scaffold mutations, however, had a more prominent effect. The new variant could cleave more

target plasmids with high efficiency, but it also cleaved more target plasmids not at all. After examining the well-cut targets (see figure 37), we found that the requirement for an adenine at position 10 had been changed to a thymine and the requirements at position 7 had been relaxed (see figures 39 and 40). TALE fusion further widened the amount of acceptable targets (see figure 38). So while the meganuclease is able to tolerate more mismatches and has a reduced accuracy after fusion, there still is a net gain in specificity for the total construct. Similar observations were made for the cleavage region 2 to -2 (see figures 41 and 42). Many targets that were cut with medium efficiency before, saturated the assay after TALE-DBD fusion. However, the amount of uncut targets did not change, suggesting that there are some sterical constraints in the catalytic center which cannot be overcome by increasing the time the nuclease is bound to its substrate.

Profiling in yeast showed that the insights we obtained from scI-CreI could be carried over to scTcr. This synthetic meganuclease targets the T-Cell Receptor α-chain and can be applied in generation of CARs, which proved to be difficult due to scTcr's cytotoxicity. TALE fusion abolished cleavage of targets without fitting EBE (see figure 43) and reduced activity on targets with more distant EBEs (spacer length >10 bp), again suggesting that the unbound TALE domain is interfering with meganucleases binding. We also saw the aforementioned boosting effect on target plasmids with a meganuclease recognition site adjacent (3-10 bp) to an EBE. In light of this, it is no surprise that our degenerate scaffold mutations also reduced toxicity in CHO cells (see appendix 52) and during our test in HEK293 cells, where we compared the amount of NHEJ INDELs between day 2 (Early) and day 7 (Late) samples. It is expected that if an active nuclease is cytotoxic, there will be a difference between the early and late samples. Cells expressing a sufficient amount of nucleases will induce detectable INDELs in the target region. However, a toxic nuclease will cause DNA damage to accumulate over time, causing a decrease in detected NHEJ since the respective cells are dying. As expected, we see this decrease for our original scTcr Meganuclease. Introduction of our degenerate scaffold mutations reduced the activity, probably by slightly altering the recognition sequence. Fusion of the TALE-DBD talTCR reverses this drop and rescues our nuclease to at least the values of the original scTcr. The most interesting observation, however, is that the NHEJ INDELs introduced by talTCR::scTcr are stable over time and even slightly increased, indicating that toxicity is greatly reduced here, too.

Addition of a TALE-DBD proved to augment activity and specificity of our I-CreI-based meganucleases.

## 3.4 TALENS WITH TWO CATALYTIC DOMAINS

Our novel, tripartite TALEN design (see figure 45) containing two nuclease domains proved to be an augmentation over the older architectures. Tev and FokI complemented each other. More than one FokI domain per DBD rarely makes sense, since two catalytic domains oriented in the right fashion are needed for cleavage. On the other hand, two I-TevI domains fused to one DBD can work, if there are `CNNNG` sites at the right distance up- and downstream of the EBE, thereby limiting the range of targets again. Anyway, the C-terminal Tev domain only produces single-strand nicks. The Tev::TALE::FokI architecture surpassed the already high activity of FokI TALENs. Our experiments showed that the fusion construct caused higher levels of recombination than their predecessors (see figure 46). This was expected, since two nuclease domains should cause more DSBs than just one. Interestingly, this boost was not restricted to target sites that were optimal for both constructs. We could see an increase in activity even on targets that were lacking a suitable `CNNNG` motif for Tev. In agreement with these findings, a construct with the inactive Tev variant R27A also caused this increase. This is suggesting that Tev's affinity for DNA is more impactful on this target than its ability as an endonuclease. By fusing the Tev domain, the overall affinity for DNA is increasing, which in turn is enhancing FokI's ability to effect a DSB.

The reverse situation is the case, when we looked at a target that is less suitable for FokI. On our Avr-18-Avr targets, AvrBs3::FokI TALENs had poorer activities compared to targets with shorter or longer spacers. Fusion of an inactive Tev nuclease domain reduced this activity even further. The TALE-DBDs are positioning the two FokI domains in an unfavorable orientation, which is further exacerbated by the added affinity of the Tev_R27A domain. Contrary, we see a large increase in activity on the same target plasmid, when WT-Tev is fused. Here, the Tev domain can cleave the plasmid upstream of the EBE. This DSB relaxes the DNA, which could have a beneficial effect on the positioning of the two FokI domains. Here, the activity of the dcTALEN is even greater than the absolute activities of the single TALENs.

An interesting side note is the activity of the scTrex::tale::FokI fusions. The single-chain version of the human exonuclease TREX2 increases the amount of deletion events, when cotransfected with a site-specific nuclease. Here, it adds another 491 aa (53 kDa) to the TALEN. The resulting construct, with a total size of 155 kDa, could still induce a DSB, albeit with lower efficiency. If the exonuclease still improves non-conservative DNA-repair mechanisms when it is fused to a TALEN, has to be further determined.

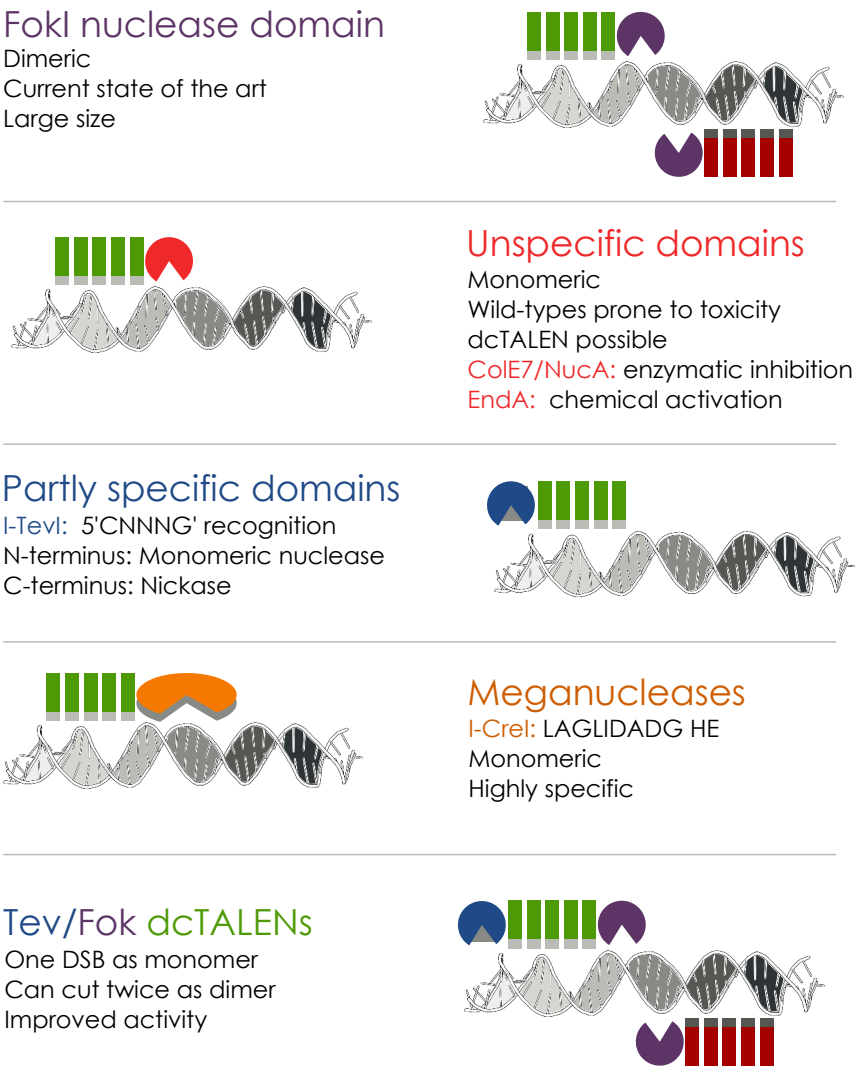



Figure 48: Summary of catalytic domains
Shown are the types of catalytic domains used in this work and their special features.

## 3.5 SUMMARY

The FokI nuclease domain has been the state of the art for ZFNs for over a decade. A new DBD that is more specific and easier to engineer came with the discovery of TALEs. Now one could approach the engineering of the nuclease domain, since two main problems of the ZFN architecture had been suddenly solved. A comparison of the nuclease domains used in this work can be seen in figure 48. The unspecific domains that were used in this work were active as monomers in a variety of organisms. Flaws they all shared at first, were the susceptibility to off-site activity and consequential toxicity. Both could be reduced by lowering the overall activity of the nuclease domain via mutagenesis. The goal of this approach is that the nuclease can only cleave when it is held in proximity of the DNA for a long enough time. Cleavage *in trans*, while the TALE-DBD is bound, is then ideally prevented by a short enough linker. Other types of toxicity are a consequence of the method and cannot be fully excluded.

E. g., cycles of "cleavage ↷ repair ↷ recleavage" will cause the loss of genetic information eventually. Ideally, the recognition site disappears during homology directed repair or alternative NHEJ. Two different ideas that could avoid these problems have been presented in this work. The first is a time-dependent control of the nuclease. SSNs are only active in a time window, to minimize the collateral damage they could cause. Transfection of mRNA or protein can achieve this, thereby also circumventing the unwanted integration of vector DNA into the target genome. Switchable nucleases could allow an even tighter control. Remaining nucleases could be shut off by addition of their inhibitor. Or activity could be restrained by addition of only a small amount of imidazole. Imidazole can permeate the cellular membrane, so intracellular imidazole concentration could be diluted by changing the growth medium.

A different approach is the excision of the recognition site. TALENs with two catalytic domains can remove the EBE, by creating double-strand breaks up- and downstream of it. EndA::TAL::EndA are able to do this *in vitro*, but unfortunately have not been tested *in vivo* yet. It should nevertheless be remembered that the nuclease domains create different types of strand-breaks depending on their fusion terminus, which was nicely seen for I-TevI and EndA. Both constructs used here had different linkers (N-terminus: GlySer (EndA) / partial WT- I-TevI linker (Tev); C-Terminus: FokI linker (EndA) / 11 residues (Tev), see appendix A.2.1), suggesting that the difference originates mostly in the structure of the TALE-DBD and not so much from the linkers. It would be interesting to see, if other TALEN architectures behave similarly. E. g., can the nickase TAL::MutH [77] work as a cleavase, when MutH is fused to the N-terminus or is scPvuII::AvrBs3 more active than AvrBs3::scPvuII [243]?

The dcTALEN TEV::TAL::FOKI described in this work, however, displayed a great activity *in vivo*, which surpassed a regular FokI TALEN. A similar construct with two catalytic domains is the meganuclease fusion to the I-TevI nuclease domain, which has been described recently by [238]. Alternatively one can also use multiple standard TALENs or gRNAs, with the downside of an even larger transfection load.

ANOTHER TYPE OF INHERENT TOXICITY stems from the natural properties of the DNA-binding domain. TALE proteins have to be able to tolerate mismatches. Otherwise a small mutation in the effector binding element could prevent the TALE from binding, which would be beneficial for the host organism. Therefore, TALEs have to be undiscriminating enough to counteradapt mutations of the EBE, but be specific enough to not bind too many off-sites. TALEs share this dichotomy with the selfish homing endonucleases: here counteradaption and ability to propagate to new alleles conflicts with off-site cleavage and toxicity. This is also the reason, why I-CreI is slightly toxic in mammalian cells. Understandably, inaccuracy is neither a desired trait for TALE-DBDs, nor for synthetic meganucleases. Both shortcomings can be fixed by creating a TALE::MEGANUCLEASE fusion. The highly-specific nuclease domain compensates for weaknesses in the TALE-DBD, while the TALE-DBD prevents the HE from binding to off-sites. Integration of part of the recognition into the catalytic domain combined with the mere length (e. g., ~40 bp for AVRBS3::SCCRE) of the target site, make TALE::MEGAS extremely specific SSNs. This safety and their monomeric layout makes them ideal for genetherapy.

Certain types of DNA-damage appears however, no matter how precise the nuclease is. Wrong ends can be joined during NHEJ or the target chromosome can form a crossing-over product with the repair template during HR. Recombinase and transposase fusions [147, 164] can integrate a template without the necessity of the DSB-repair system. Although the main drawback for this system is off-site integration at the moment. TALE::recombinases might replace SSNs in the future, when the recombinases have been engineered sufficiently for a fusion context.

THE FOUNDATION for successful engineering of a fusion domain is sufficient information about it. Structural knowledge, but also classic biochemical insights into mechanisms and relevant residues are invaluable. *In vitro* experiments may not fully imitate the conditions inside a cell, but they nevertheless provide practical evidence and will still help us understand biochemical systems better in the future.

## 3.6 CONCLUDING REMARKS

Site-specific genome engineering represented a paradigm shift in the modification of genomes. Earlier, genomes were mutated as a whole or therapeutic factors and beneficial genes were integrated randomly. The next revolution came with the second generation of site-specific nucleases. The design of meganucleases or ZFNs requires time and experience, while TALENs and RGENs make the addressing of a specific sequence almost trivial. This is reflected in the magnitude of how fast researchers have adapted this new technique. New SSNs have been applied successfully in virtually every model organism and greatly augmented the ability to study them. It is therefore no surprise that genome editing has been named "Method of the Year 2011" by NATURE METHODS [1] and runner-up for "Breakthrough of the Year 2012" by SCIENCE [2]. Beyond nucleases, synthetic biologists have started to create complex artificial genetic circuits that artfully apply TALE fusions.

Novel or improved DNA modifying domains will allow new therapies and new applications in the future. The author of this work hopes that some of the findings described here may prove helpful then.

# Part III

# MATERIALS AND METHODS

# 4 MATERIALS AND METHODS

Materials and methods are described in this chapter. Sequences for proteins (A.2.1) and primers (A.2.2), as well as the compositions of standard buffers, can be found in the appendix.

## 4.1 MOLECULAR CLONING

Coding sequences were subcloned into their recipient vectors (table 11) with standard molecular biological methods. First, sequences were obtained by PCR or by restriction digest from a plasmid. Fragments were then cleaned up with a PCR-clean up kit, if clipped fragments were small enough to be removed (<60bp). Otherwise samples were gel purified. Standard cloning strains DH5α or XL10 GOLD were transformed with the ligated plasmids and grown in 3-4 ml cultures over night. Correct integration of plasmids was verified by screening and sequencing. An overview of the subcloned fragments can be found in table 6, while the recipient scaffolds can be found in table 7.

### 4.1.1 *Restriction digests*

Between two and four µg of DNA were digested (see table 8) for 10 minutes per µg. Restriction endonucleases were obtained from NEB and THERMO SCIENTIFIC. Buffers and incubation temperatures were adjusted accordingly.

Table 8: Layout for restriction digests

| REAGENT | FINAL CONCENTRATION | VOLUME |
|---|---|---|
| Plasmid DNA | 2-4 µg | var |
| Buffer | 1x | 5 µl |
| Enzyme 1 | var | 0.25 - 1 µl |
| (Enzyme 2) | var | 0.25 - 1 µl |
| Water | | to final volume |
| Total | | 50 µl |

### 4.1.2 *Gel purification*

Restriction digests were supplemented with standard DNA loading dye (e.g. *Orange* DNA *Loading Dye 6x* (LIFE TECHNOLOGIES), *Gel Load-*

Table 6: Cloning strategies
Shown are DNA fragments for catalytic domains, fusion terminus in the final construct, sequence size as well as the 5′ and 3′ restriction sites that were used for subcloning. Protein sequences can be found in appendix A.2.1.

| NAME | FUSION TERMINUS | 5′ SITE | 3′ SITE | SIZE [bp] |
|---|---|---|---|---|
| Col E7 | N | NcoI | BamHI | 414 |
| | C | Kpn2I | EagI | 426 |
| NucA | C | BamHI | EagI | 768 |
| EndA | N | NsiI | BamHI | 747 |
| | C | AvrII | BlpI | 759 |
| Tev | N | NcoI | Kpn2I | 567 |
| | C | BamHI | EagI | 573 |
| Fok | C | Kpn2I | EagI | 639 |
| scCre | C | BamHI | EagI | 1104 |
| scCreDS | C | BamHI | EagI | 1104 |
| scTcr | C | BamHI | EagI | 1104 |
| scTrex | N | NcoI | BamHI | 729 |
| SNAP tag | C | SacI | HindIII | 588 |

Table 7: Recipient scaffolds
Shown are tale scaffolds in vectors pCLS7865 and pQE30, which can accept catalytic domains. EndA dual catalytic talens used the N-ter_AvrBs3(28) scaffold, while Tev::tale::Fok nucleases were generated by first constructing TEV::TALE and TALE::FOK and subsequent digest with NsiI in the N-terminal pseudo repeat region of the tale. Note that NsiI (A▲TGCA▼T), instead of NcoI (C▼CATG▲G), is used to cut at the start codon in pQE30.

| NAME | MULTIPLE CLONING SITE | MCS LOCATION |
|---|---|---|
| Tal11_NFS1<br>(N-terminal Fusion Scaffold) | NcoI / AgeI /<br>Kpn2I / BamHI | 5′ / N |
| Tal11_CFS1<br>(C-terminal Fusion Scaffold) | BamHI / Kpn2I /<br>BglII / EagI | 3′ / C |
| Nter_AvrBs3_(28)_His | (5′) NsiI/BamHI/SalI/Bsu36I<br>(3′) AvrII/SacI/HindIII/BlpI | both |

Table 9: Layout for ligations

| REAGENT | FINAL CONCENTRATION | VOLUME |
|---|---|---|
| Recipient vector | 5 ng / µl | var |
| Insert fragment | 2-6x molar excess | var |
| T4 DNA ligase buffer (10x) | 1x | 1 µl |
| T4 DNA ligase | 1-3u | 1 µl |
| Water | | to final volume |
| Total | | 10 µl |

*ing Dye* (NEB) or *AAP* buffer 5x (ethylenediaminetetraacetate (EDTA) 250 mM, all (w/v): 25 % Sucrose, 1,2 % SDS , 0,01 % Bromophenolblue , 0,01 % Xylencyanol, NaOH to pH 8.0)) and completely loaded onto a 1% agarose gel buffered in 1x TAE. Gels were run at 5V / cm until satisfying separation was achieved. Fragments were visualized by ethidium bromide staining and then cut out of the gel under UV light. Time under the UV was limited to a minimum to avoid nicking of the DNA and accompanying low ligation efficiencies. Isolated fragments were then purified using the *PCR Clean Up and Gel Extraction*-Kit (MACHEREY NAGEL) or the *Wizard SV gel and PCR clean up system* (PROMEGA).

### 4.1.3 *Ligations*

Stocks of the recipient vector were dephosphorylated with calf intestinal alkaline phosphatase (CIAP) to increase efficiency. Between 1 and 4 µg of vector were treated with 1 µl of CIAP (PROMEGA or NEB) in 50 µl of total volume for 30 minutes at 37°C. Vectors were purified with the *Wizard SV gel and PCR clean up system* (PROMEGA) or the *PCR Clean up and Gel Extraction Kit* (MACHEREY NAGEL). Reagents for the ligation were mixed as shown in table 9. Inserts were added in two to six fold molar excess over the recipient. Solutions were incubated at room temperature (RT) between 20 minutes and one hour. T4 ligase was obtained from LIFETECHNOLOGIES, PROMEGA or NEB.

### 4.1.4 *Escherichia coli Strains*

Several *E. coli* strains were used, depending on the application. Plasmid DNA was amplified in DH5α, XL1BLUE and XL10GOLD. Proteins were expressed in XL10GOLD (for T5 promoters) or BL21(DE3) (for T7 promoters). Genotypes are:

- BL21(DE3): $F^-$ ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ araB::T7RNAP-tetA

- DH5α: $F^-$ endA1 glnV44 $thi^-$1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17($r_K^-$ $m_K^+$), λ–
- XL1BLUE: endA1 gyrA96($nal^R$) thi-1 recA1 relA1 lac glnV44 F'[ ::Tn10 $proAB^+$ $lacI^q$ Δ(lacZ)M15] hsdR17($r_K^-$ $m_K^+$)
- XL10GOLD: endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 $tet^R$ F'[proAB $lacI^qZ$ΔM15 Tn10($Tet^R$ Amy $Cm^R$)]

### 4.1.5 *Transformations*

Bacterial aliquots of 50µl were gently thawed on ice for ten minutes and then incubated for another five minutes with five µl of a ligation solution or between one to five ng of supercoiled plasmid DNA.

- Chemically competent DH5α or BL21 were then heatshocked for 30s at 42°C, placed on ice for 2 minutes and 400µl of SOC medium were added. Cells were then grown at 37°C for 30-45 minutes in SOC-medium and between 10 and 100 µls were plated out on LB agar (LB medium supplemented with 1.5% agar (w/v)). The 30 minute growth step was skipped if ampicillin resistance was the selection marker. Agar plates were then placed at 37°C over night.
- Z. competent XL10GOLD did not require a heat shock and could be plated directly onto the agar plates, if ampicillin was used. Otherwise a 30 to 45 minute growth step with SOC medium had to be added.

### 4.1.6 *Screening*

A colony PCR (see 4.1.10) was used to screen for successfully inserted constructs. Amplicon length is then determined by standard agarose gel electrophoresis or PAGE. Examples for agarose gel electrophoresis and PAGE can be found in [11].

### 4.1.7 *Plasmid purification*

Plasmids were extracted from 3 ml cultures with the Miniprep-kits *NucleoSpin* (MACHEREYNAGEL) or *Wizard Plus* (PROMEGA). When endotoxin-free or more DNA was required, larger cultures were cleaned up with the *NucleoBond Xtra Midi EF* kit (MACHEREYNAGEL) or the *Wizard Plus Midi* kit (PROMEGA).

### 4.1.8 *DNA quantification*

The *NanoDrop 1000* system by THERMOSCIENTIFIC was used to quantify DNA and fluorophores.

### 4.1.9 *Sequencing*

Sequencing results were obtained from MWG-BIOTECH (EUROFINS) or LGC-GENOMICS.

### 4.1.10 *Polymerase Chain Reactions*

Depending on the requirements, different polymerases were selected. Annealing temperatures were adjusted depending on primer length and composition.

- The *Thermus aquaticus* (*Taq*) (PROMEGA or preparation by Heike Büngen) or the *REDTaq* polymerase (SIGMA-ALDRICH) was used for colony PCR to screen for successful ligations. Here a bacterial colony is transferred into 10 µl of water. Then 10 µl of master mix (0.5 µM Primer forwards, 0.5 µM Primer reverse, *Taq* buffer [Final: 10 mM Tris/HCl pH 9.0, 50 mM potassium chloride, 0.1% Triton X 100 (v/v), 1.5 mM magnesium chloride] or *REDTaq* buffer, 0.2 µl *Taq* polymerase) were added and a PCR was carried out.

- PCRs with *Pyrococcus furiosus* (*Pfu*) (Heike Büngen) or Phusion (*Phu*) High-Fidelity (NEB) DNA polymerases were chosen for standard cloning procedures. A standard layout would be: 5 to 15 ng of plasmid DNA, 0.5 µM Primer forward, 0.5 µM Primer reverse, 200 µM deoxyribonucleotide triphosphate (dNTP)s, buffer (1x Pfu: 20 mM Tris/HCl pH 9.0, 10 mM ammonium sulfate, 10 mM potassium chloride, 0.1% Triton X 100 (v/v), 2.5 mM magnesium hyposulfite; 1x Phu: see NEB), 1 µl polymerase (*Pfu*, Phu), nuclease-free water to 50 µl.

- *QuikChange Lightning* or *Lightning Multi Site-Directed Mutagenesis Kits* (AGILENT) were used according to protocol. A megaprimer PCR, utilizing *Pfu* polymerase, was applied in a similar manner to modify single nucleotides or codons. Here a fragment is amplified from a plasmid template by mismatch PCR. This PCR introduces point mutations via mismatching oligos. The resulting fragment acts as a primer for a second round of PCR, which replicates the whole plasmid. Afterwards, the methylated template plasmids were removed by DpnI digest. Linear plasmid copies were ligated and then transformed into a cloning strain.

Table 10: Cycle times and temperatures of various PCRs
Annealing temperatures were adjusted depending on primer length and composition.

| STEP | STANDARD PCR | MEGAPRIMER 2 | HERCULASE |
|---|---|---|---|
| Initial denat. | 60″, 94°C | 5′, 93°C | 2′, 95°C |
| Denaturation | 10-20″, 94°C | 1′,93°C | 10-20″, 95°C |
| Annealing | 10-45″, 50-60°C | 50″, 60°C | 20″, 50-60°C |
| Extension | 1′ / kb (*Taq*)+15″, 68°C | 15″, 68°C | 30″, 72°C |
| | 0.5′ / kb (*Pfu*)+15″, 72°C | | |
| | 0.5′ / kb (Phu)+15″, 72°C | | |
| Cycles | 30-33 | 18 | 30 |
| Final ext. | 2x Ext., same °C | 20″, 68°C | 2′, 72°C |

- splicing by overlap extension (SOE) PCRs were introducing restriction sites and point mutations. Here two DNA fragments of a region of interest were generated in two separate PCR reactions. This way mutations were introduced via mismatch PCR. Both fragments overlapped so they could hybridize and act as a template in a second PCR. The resulting fragments were recombinant fusions of both original PCR products and were subcloned into plasmids of choice via restriction and ligation. They used the standard *Pfu* setup, but with an annealing time of 1 minute for the second PCR.
- To amplify genomic regions of DNA for the surveyor assay and for deep sequencing, two rounds of PCR with the high-fidelity polymerase *Herculase II Fusion* (AGILENT) were applied according to the manufacturer's instructions. The first round amplified the target region, while the second round was a nested PCR to add multiplex identifier (MID) tags and adapter sequences for 454 sequencing. Here the reverse primer was the same for all samples of one genomic locus, while the forward primers contained different MIDs to distinguish the constructs later.

## 4.2 DNA SYNTHESIS

### 4.2.1 *DNA oligonucleotides*

Primers were aquired from EUROFINS or BIOMERS.

### 4.2.2 *TALE scaffold generation*

TALE constructs used in this work are based on the AvrBs3 scaffold (see figure 7). For the N-terminus, a length of 136 amino acids was chosen, resulting in a truncation of Δ152 aa. The C-terminus lengths were either 11, 28 or 40 residues after the last TALE half-repeat. Two BsmBI restriction-sites are located in the central region between the TALEN N- and C-terminus. The type IIS restriction endonuclease BsmBI was then used to excise the RVD-less central region leaving compatible overhangs for the ligation of the RVD array.

### 4.2.3 *TALE repeat synthesis*

The AvrBs3 binding domain was used for most experiments to be able to compare constructs. Other RVD-arrays were synthesized by the CELLECTIS PLATFORM. During synthesis, small DNA fragments ("blocks") coding for combinations of two repeats were fixated to a biotin matrix, ligated to a new set of blocks and then removed from the matrix by restriction digest. This process was iterated until the sequence of direpeat building blocks reached the desired length. The final DNA fragment was cloned into a shuttle vector and provided for further cloning.

#### 4.2.3.1 *Insertion of RVDs into TALE backbones*

The provided shuttle vectors were first predigested with the restriction endonuclease SfiI and subsequently gelpurified. The gelpurified fragments were further digested with the REs BbvI and SfaNI. Fragments were then cleaned up with the *PCR Clean Up and Gel Extraction* Kit (MACHEREY NAGEL). The resulting fragment was ligated into TALE scaffolds predigested with BsmBI.

### 4.2.4 *Catalytic domains*

Synthesized DNAs and Plasmid vectors containing catalytic domains of I-TevI/ColE7/NucA or subunits of I-CreI were bought from GENECUST. Plasmids containing EndA mutants were provided by Marika Midon and Heike Büngen. The SNAP-tag was obtained from NEB.

### 4.2.5 *Vectors*

Several different vector plasmids were used in this work dependent on the intended purpose. All vectors contained a pUC origin of replication. Constructs were assembled in the cloning vectors pCLS7865, pCLS12081 (ColE7 and NucA) or pQE30 (EndA).

Table 11: List of plasmids

NcoI and EagI could be used to subclone sequences into pCLS7865, pCLS12081, pCLS0368, pCLS0542 and pCLS7763. Constructs had to be transferred from the shuttle vector into pcls1853 via digest with XhoI and BssHII. NsiI and HindIII (or BlpI) could be used to clone fragments into our pQE30 derivate.

| NAME | PURPOSE | COMMENTS |
|---|---|---|
| pCLS7865 | cloning-/shuttle vector (CSV) | Amp resistance |
| pCLS12081 | CSV with inhibitor (Im7 or NuiA) | based on 7865 |
| pCLS0368 | expression (*E.coli*) | T7 promoter |
| pCLS0542 | expression (yeast) | Gal 1/10 promoter<br>Leu1 auxotrophic<br>marker (yeast) |
| pCLS7763 | expression (yeast) | Gal 1/10 promoter<br>Kan resistance (yeast) |
| pCLS1853 | expression (mammals) | CMV promoter |
| pCLS15603 | expression (plants) | CaMV 35S promoter |
| pCLS0002 | mock controls | empty |
| pCLS0003 | mock controls | empty |
| pQE30 | cloning / expression (*E. coli*) | T5 promoter |
| pAT PEB | *in vitro* target plasmids | |

### 4.2.6 *Target plasmids and PCR fragments*

Yeast and mammalian target plasmids were provided by the CELLECTIS PLATFORM. The pAT PEB (negative control) and pAT PEB 339 (AvrBs3 EBE) plasmids were provided by Mert Yanik. Small DNA fragments and fluorescently labeled targets were prepared via *Pfu*-PCR with a pAT PEB 339 template. For primers see appendix (A.2.2).

## 4.3 PROTEINPURIFICATION

Constructs were either expressed in XL10 GOLD or BL21 cells, according to the protocols below.

### 4.3.1 *In* XL10GOLD

Starter cultures of XL10 GOLD cells harboring EndA expression plasmids were grown over night (O/N) at 37°C. Expression cultures were then inoculated with the starter culture in a 1% ratio. Cells were grown at 37°C to an $OD_{600}$ of 0.5. Cultures were cooled down to 20°C, while "before induction" samples were taken. Induction was started with a final concentration of 1 mM IPTG. Growth continued over night at 20°C for 18h.

Cells were harvested and resuspended in 20 ml lysis/equilibration/wash buffer (1 M NaCl) per 500 ml of culture. The mixture was then sonicated on ice for 5x 30 seconds with 30 second pause in between (Output control: 5, duty cycle 50%). The cell lysate was centrifuged (BECKMANN COULTER JA-20 rotor, 19 000 rpm or ~43 000 rcf) and filtered through a 0.45 µM filter. A *Protino-Ni-IDA 2000* column (MACHEREYNAGEL) was prequilibrated, before the filtrate was poured onto it. The column was then washed in succession with 4x4 ml buffer with decreasing salt concentrations (1, 0.8, 0.6 and 0.3 M NaCl). Proteins were eluted with 1.5 ml elution buffer (MACHEREYNAGEL).

The eluates were diluted with 50 mM sodium phosphate to a sodium chloride concentration of 150 mM (this step was skipped if the 300 mM NaCl Low-Salt buffer was used), before being loaded on a prequilibrated 1 ml *HiTrap Heparin* column in the *Äkta* HPLC system (both GE HEALTHCARE). The column was washed with 5 column volumes (CV) of Low-Salt buffer. Elution was accomplished with a linear sodium chloride gradient over 10 CV to a final concentration of 1 M (100% High-Salt buffer). UV/Vis Absorbance at 280 nm was measured and peak fractions were collected. Buffers are shown in table 12.

#### 4.3.1.1 *Dialysis*

Nucleases were dialyzed (see table 12) O/N and the buffer was exchanged in the morning to allow four more hours of dialysis. Dialysis membranes with a molecular weight cut-off of 10 kDA were used

Table 12: Buffers for *Protino* and *HiTrap* purification

| BUFFER | REAGENT | CONCENTRATION |
|---|---|---|
| *Lysis/Equilibration/Wash* buffer | NaCl | 1000-300 mM |
| pH 8.0 | $NaH_2PO_4$ | 50 mM |
| *Elution* buffer (MACHEREYNAGEL) | NaCl | 300 mM |
| pH 8.0 | $NaH_2PO_4$ | 50 mM |
| | Imidazole | 250 mM |
| Low-Salt buffer | NaCl | 300 (or 150) mM |
| | Tris/HCl pH 8.0 | 50 mM |
| | EDTA | 1 mM |
| | DTT | 1 mM |
| | Glycerol | 5 % (v/v) |
| High-Salt buffer | NaCl | 1 M |
| | Tris/HCl pH 8.0 | 50 mM |
| | EDTA | 1 mM |
| | DTT | 1 mM |
| | Glycerol | 5 % (v/v) |
| Dialysis buffer | NaCl | 500 M |
| | Tris/HCl pH 8.0 | 50 mM |
| | EDTA | 1 mM |
| | DTT | 1 mM |
| | Glycerol | 50 % (v/v) |

(e. g., *SnakeSkin* Dialysis Tubing, LIFETECHNOLOGIES). Afterwards TALENs were separated into 50 µl aliquots, flash frozen in liquid nitrogen and stored at -80°C.

#### 4.3.1.2 *SDS-PAGE*

Protein samples were analyzed via standard SDS-PAGE. Gels were run in the *MiniProtean Tetra System* (BIORAD) for 45 min at 35 mA. Gels were either self-cast (Separating gel: 10, 12 or 15% acryl amide (v/v), 0.1 % SDS (w/v), 840 mM TrisHCl pH 8.8; Stacking gel: 6% acryl amide (v/v), 0.1% SDS (w/v), 125 mM TrisHCl pH 6.8) or precast *MiniProtean TGX* (BIORAD) gels were used. Gels were then stained with coomassie brilliant blue (e. g., 0.1% coomassie brilliant blue G250 (w/v), 2% phosphoric acid (v/v), 5% aluminum sulfate (w/v), 10% ethanol (v/v)).

#### 4.3.1.3 *Protein quantification*

The absorbance at 280 nm was measured and concentrations were determined with a calculated extinction coefficient (see 4.6).

### 4.3.2 *In* BL21

Expression cultures were grown to an $OD_{600}$ of 0.5 and then induced with IPTG. Growth continued O/N at 20°C for 18h. Cells were harvested (resuspenion in *Native IMAC Lysis* buffer), sonicated and the constructs TEV::NPTII-CB TALENs and AVRBS3::FOKI were then purified with the *Profinia Protein Purification Instrument* from BIORAD. For this the *Mini-Profinity Profinia IMAC Cartridge* from the *Native IMAC Purification*-kit (BIORAD) was used according to the manufacturer. Buffers can be found in table 13.

### 4.3.3 *Cleavage assays in vitro*

#### 4.3.3.1 *EndA cleavage assays*

All concentrations were as described in the results section. Water, buffer and DNA were mixed first and kept on ice. Protein was added from a fresh dilution of an ENDA::TALE aliquot. Imidazole was added to the desired concentration and the samples were placed in a preheated thermocycler after mixing. The reaction was stopped with *AAP* buffer 5X (EDTA 250 mM, all (w/v): 25 % Sucrose, 1,2 % SDS , 0,01 % bromophenol blue , 0,01 % xylencyanol, NaOH to pH 8.0). Samples were then analyzed with standard agarose gel electrophoresis or PAGE (e. g., see [11]).

Table 13: Buffers for Profinia IMAC

| BUFFER | REAGENT | CONCENTRATION |
|---|---|---|
| *Native IMAC Lysis* buffer | KCl | 300 mM |
| pH 8.0 | $KH_2PO_4$ | 50 mM |
| | Imidazole | 5 mM |
| *Native IMAC Wash buffer 1* | KCl | 300 mM |
| pH 8.0 | $KH_2PO_4$ | 50 mM |
| | Imidazole | 5 mM |
| *Native IMAC Wash buffer 2* | KCl | 300 mM |
| pH 8.0 | $KH_2PO_4$ | 50 mM |
| | Imidazole | 10 mM |
| *Native IMAC Elution buffer* | KCl | 300 mM |
| pH 8.0 | $KH_2PO_4$ | 50 mM |
| | Imidazole | 250 mM |
| *Desalting* buffer | NaCl | 137 mM |
| pH 7.4 | KCl | 2.7 mM |
| | $Na_2HPO_4$ | 4.3 mM |
| | $KH_2PO_4$ | 8.1 mM |

#### 4.3.3.2 *Fluorescence assays and denaturing* PAGE

For the timecourse assays to determine strand specificity, a master mix was used. Two samples per construct were taken at 2,4,8,16,32 and 64 minutes and pippetted into stop buffers. Samples for the native PAGE were stopped with *AAP* buffer without bromophenol blue and xylencyanol. The 2X stop buffer for the denaturing PAGE contained 80% formamide, 0.5 % SDS and 0.5 mM EDTA. Both samples were incubated for 2 minutes at 80°C to ensure denaturation of the protein.

Half of the samples were separated in the dark by standard 8% PAGE. The other samples were analyzed via denaturing gels ( 10% Acrylamide, 0.1 M Tris, 0.1 M borate, 2.5 mM EDTA, 40% formamide (v/v), 42% urea (w/v)). Replacement of borate with taurin and addition of glycerol to the stop buffer yielded equal qualtiy gels and did not improve the method. The *MiniProtean* (BIORAD) system was used to run the gels. Two denaturing gels were prerun at 400 V and 40 mA for 25 minutes, to heat up the gel. The actual run was carried out in the dark at 300 V and 30 mA for 1 h with a 100 W power supply. Gels were scanned with the *Typhoon FLA9500 biomolecular imager* (GE HEALTHCARE). Band intensity was calculated with IMAGEJ (see section 4.6).

#### 4.3.3.3 *Other cleavage assays*

TEV::TALE and TALE::FOKI *in vitro* cleavage assays were carried out similar to the standard EndA assays. Assays were started by addition of the protein. Concentrations and conditions were as described for the specific experiment.

## 4.4 YEAST SINGLE STRAND ANNEALING ASSAYS

Profiling in yeast was carried out by the CELLECTIS yeast platform. At least two clones were assayed per nuclease construct and experiment. Experiments were repeated on average three times, but at least twice.

Yeast expression plasmids (542 or 7763) containing SSNs were transfected into yeast strains. These nuclease strains were gridded with a colony gridder (QPIXII from GENETIX LIMITED) on a nylon filter. Compatible yeast strains harboring a target plasmid were added on top in a second gridding process. The filter membranes were placed on agar containing YPGlycerol-rich medium and were allowed to mate O/N at 30°C. Filters were then transferred to synthetic medium, without leucine and tryptophan, with glucose (2% (w/v)) as carbon source. If needed for coexpression experiments, G418 (Geneticin) was added to select for Kanamycin resistance. Cells were grown for 5 days at 30°C and diploids were selected for. The filters were then transferred onto YPGalactose-rich medium to express the TALENs for 48 h at

30°C and 37°C. Subsequently, filters were placed on agarose medium (1% agarose (w/v), 0.02% X-Gal (w/v) in 0.5M sodium phosphate buffer, pH 7.0, 0.1% SDS (w/v), 6% dimethyl formamide (v/v), 7 mM β-mercaptoethanol) and incubated at 30°C and 37°C for 48 h. The filters were scanned after 8 h, 24 h and 48 h, to measure β-galactosidase activity. Median values for the color of the colonies were used to quantify blueness. White colonies were assigned the activity value 0, while dark blue, saturated colonies were assigned the activity value 1. β-galactosidase activity is then used as a measurement for homologous recombination efficiency. Plasmid DNA for the expression vectors could later be extracted and validated by sequencing, to ascertain the identity of a nuclease.

#### 4.4.0.4 *In vivo cloning*

Sequences coding for SSNs could also be transferred into the expression vectors via *in vivo* cloning in yeast. For this a linearized (BssHII) fragment, coding for the nuclease, was cotransfected with a linearized recipient vector (NcoI / EagI). A functional expression vector is then generated in yeast via homologous recombination, since both fragments contain two homology regions of ~40 bp.

## 4.5 CELL CULTURES

### 4.5.1 *General*

#### 4.5.1.1 *Strains*

Three cell lines were used in this work:

- CHO cells, strain K1
- CHO $\pi$-10 cells, harboring an optimized, palindromic I-CreI target site on a chromosomal locus
- HEK293 cells

#### 4.5.1.2 *Cell growth and passages*

Cells were grown at 37°C and 5% $CO_2$. For passage, cells were washed in phosphate-buffered saline (PBS), trypsinized and resuspended in F-12 K (CHO) or DMEM (HEK293) complete medium (GIBKO). 2 mM L-glutamine, penicillin (100 IU / ml), streptomycin (100 mg / ml), amphotericin B (*Fungizone*, 0.25 mg / ml) and 10% fetal bovine serum were supplemented.

#### 4.5.1.3 *Cell numbers*

The *Countess Automated Cell Counter* from LIFETECHNOLOGIES in combination with Trypan Blue staining was used to determine the cell number.

#### 4.5.1.4 *Expression*

Ectopic expression was driven by a cytomegalovirus (CMV) promoter in plasmid pCLS1853.

#### 4.5.1.5 *CHO-K1*

CHO-K1 cells (1 x $10^6$ cells) were electroporated with the *Nucleofector Kit T* for CHO-K1 cells (LONZA) according to the provided protocol. Between 200 ng and 10 µg of expression plasmid were transfected. Expression vectors for scTREX, repair matrices or "empty" mock-plasmids were sometimes cotransfected, so that total transfection amounts were at 10 or 15 µg. The cells were then plated in a 10 cm petridish in complete medium (F-12K medium (GIBKO) supplemented with 2mM L-glutamine, penicillin (100 IU / ml), streptomycin (100 mg / ml), amphotericin B (*Fungizone*, 0.25 mg / ml) and 10% fetal bovine serum).

#### 4.5.1.6 *Hek293*

Aliquots of 1.2 * $10^6$ HEK293 cells were plated onto a 10-cm dish and grown O/N. Five µg of DNA were transfected the next day with the *Lipofectamin 2000* reagent (LIFE TECHNOLOGIES), as specified by the manufacturers instructions.

#### 4.5.1.7 *Transfection efficiency*

Transfection efficiency was measured by blue- or cyan fluorescent protein with the *MACSQuant* Flow Cytometer (MILTENYI).

#### 4.5.1.8 *Genomic DNA extraction and analysis*

Two or three days and seven days after transfection, genomic DNA was extracted with the *DNeasy Blood and Tissue* Kit from QIAGEN, according to the manufacturer's instructions. Extracted DNA was then amplified with two rounds of *Herculase* PCR with specific primers flanking the endogenous target site (see 4.1.10). Amplicons were purified via magnetic beads using the *Agencourt AMPure XP* system (BECKMAN COULTER), according to the manufacturer's instructions. Samples were analyzed via T7 endonuclease 1 assay was carried out or with a 454 sequencing system (454 LIFE SCIENCES). NHEJ events were considered if insertions or deletions were detected within the expected cleavage site.

Table 14: List of X-gal staining buffers

| Buffer | Reagent | Concentration |
|---|---|---|
| Fixation buffer | sodium phosphate pH 7.3 | 100 mM |
| | $MgCl_2$ | 1 mM |
| | glutaraldehyde | 0.5% (v/v) |
| Wash buffer | sodium phosphate pH 7.3 | 100 mM |
| | $MgCl_2$ | 1 mM |
| | NP-40 | 0.02% (v/v) |
| Coloration buffer | sodium phosphate pH 7.3 | 100 mM |
| | $MgCl_2$ | 1 mM |
| | potassium ferricyanide | 5 mM |
| | potassium ferrocyanide | 5 mM |
| | X-Gal | 0.1% (w/v) |

#### 4.5.1.9 *T7 endonuclease 1 assays*

The purified PCR products were also analysed via a T7 endonuclease 1 assay. 200 ng of DNA were denatured and reannealed in NEB buffer 2 and a volume of 19 µl. One µl of T7 endonuclease 1 (~10 U, NEB) was added and the samples were incubated at 37°C for 15 minutes. The reaction was stopped by addition of 2 µl 0.5 M EDTA and samples were analyzed via PAGE.

#### 4.5.1.10 *X-Gal staining of CHO cells*

A CHO strain containing a chromosomal I-CreI target site, flanked by a defective β-galactosidase reporter, was grown and transfected as described earlier. Two micrograms of a LacZ repair matrix were cotransfected with each nuclease. After three days, the medium was removed and cells were first washed with 10 ml PBS, then with 5 ml fixation buffer and 5 ml wash buffer. After removal of the wash buffer, 5 ml of X-Gal containing coloration buffer were added and the plates were incubated O/N at 37°C. Plates were then examined via confocal microscopy.

#### 4.5.1.11 *Dose-response assays in CHO-K1 on an extrachromosomal target*

Single-strand annealing assays were also carried out on extrachromosomal targets in CHO-K1 cells. Analogous to the yeast assays, a disrupted β-galactosidase gene with homology regions was used as a reporter. However, ONPG (o-Nitrophenyl-β-D-galactopyranosid) replaced the X-Gal from the yeast SSA-assays. The readout was carried out by measuring the optical density at 420 nm. For each experiment, several 96-well plates, containing decreasing amounts of mammalian expression plasmids, and a 96-well plate, containing SSA-target plas-

mids, were needed. These were provided to the CELLECTIS PLATFORM for transfection using an automated *Velocity 11 BioCel* system (BIOCEL). Transfection and readout were carried out as described in [21]. Relative nuclease activities for each data point were then returned.

#### 4.5.1.12 *Toxicity assay*

Cell survival assays shown in the appendix, were performed by the CELLECTIS PLATFORM as described in [21].

### 4.5.2 *Plant* SSA*-assays*

ColE7, NucA and I-TevI based constructs were cloned into 35S CaMV based expression vector pCLS17693. Single-strand annealing assays in tobacco protoplasts were carried out by CELLECTIS PLANT SCIENCES as described in [21].

## 4.6 SOFTWARE

A number of specialized programs was used in this work.

- Graphical representations of biomolecules in this work were generated with PYMOL [190].
- Gel images were analyzed and quantified with IMAGEJ [189].
- The development environment RSTUDIO was used to sort, group and to analyze data as well as to generate figures with the language "R" [174].
- Figures were edited with INKSCAPE [215] and GIMP [214].
- Sequence and sequencing data was managed with VECTORNTI [136].
- The JAVA frontend of EMBOSS (European Molecular Biology Open Software Suite, [179]) was used for standard bioinformatic applications such as translations, reverse complementations, or calculations of molecular masses and extinction coefficients.
- Cleavage probability consensus sequences were generated with WEBLOGO [50].
- The picture on the dedication page was rendered with the BLENDER package using the CYCLES engine [24].
- This document was typeset with the LaTeX editor LyX [216] utilizing the *Classicthesis* template by André Miede [150].

# Part IV

# APPENDIX

# APPENDIX

## A.1 SUPPLEMENTARY EXPERIMENTS

Experiments that further complement the data are presented here.

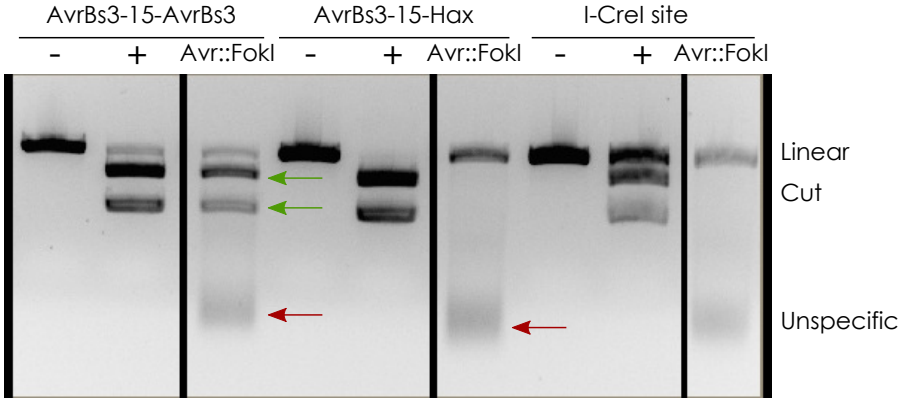



Figure 50: AvrBs3::FokI causes unspecific cleavage *in vitro*
100 nM of AvrBs3::FokI were used to cleave 10 nM of 3 kb plasmids harboring two, one or zero AvrBs3 EBEs. Samples were incubated for one hour at 37°C in 50 mM Tris/HCl pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$. Kpn2I was used as a positive control for the Avr-targets, WT I-CreI was used as a control for the I-CreI target plasmid. The target, with two AvrBs3 sites, is cut specifically (green arrows) by Avr::FokI. Unspecific cleavage products (red arrows) appear for all three targets, though.

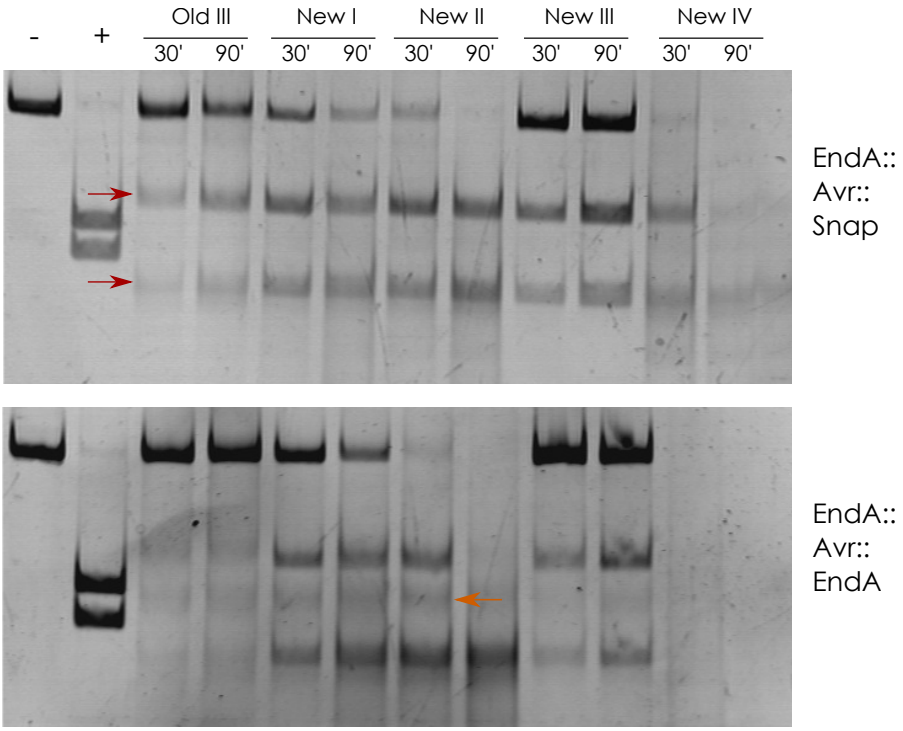



Figure 49: Buffer composition changes the cleavage pattern of EndA
A 440 bp PCR target (30 nM), harboring an AvrBs3 EBE, was incubated with 200 nM of EndA_H160G_Q186A_N202A::AvrBs3::SNAP or EndA_H160G_Q186A_N202A::AvrBs3::EndA(GAA) for 30 or 90 minutes in different buffers. The target fragments were cleaved by the N-terminal EndA domain (red arrows) and also by the C-terminal domain (orange arrow), albeit with lower activity. Buffers with low salt concentration caused DNA degradation, which can be seen as a smear. Buffers:
Old III: 100 mM NaCl, 25 mM Tris/HCl, 1 mM MgCl2, pH8
NewI: 10mM Bis-Tris-Propane-HCl, 10mM $MgCl_2$, 1mM DTT, pH7.0
NewII: 50mM NaCl, 10mM Tris/HCl, 10mM $MgCl_2$, 1mM DTT, pH7.9
NewIII: 50 mM Tris/HCl pH7.9, 100 mM NaCl, 10 mM MgCl2, 100µg/ml BSA
NewIV: 50mM Potassium Acetate, 20mM Tris-acetate, 10mM Magnesium Acetate, 1mM DTT, pH7.9

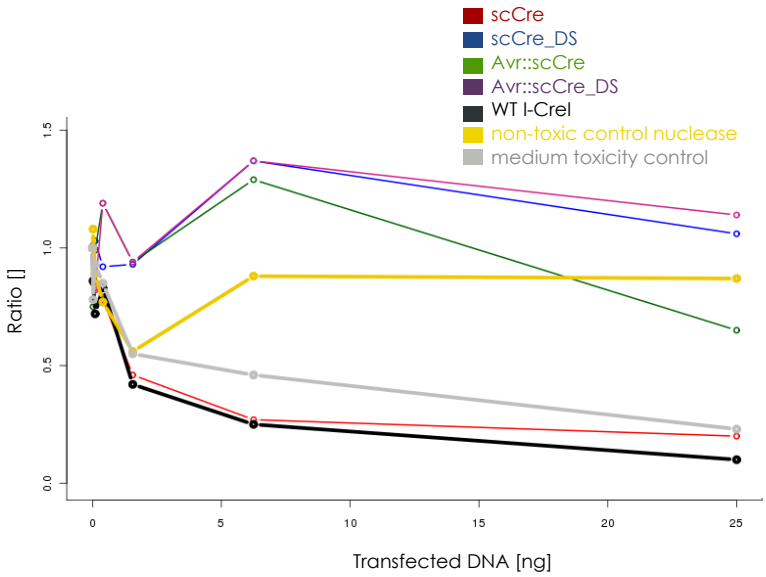



Figure 51: TALE fusion reduces toxicity of I-CreI variants
Shown is a dose-response assay in CHO cells to determine nuclease toxicity *in vivo*. The single-chain I-CreI variant (red) is as toxic as the wild-type (black), while the degenerate scaffold scCre_DS (blue) is non-toxic. Fusion of the AvrBs3 DBD (green and purple) reduces toxicity.

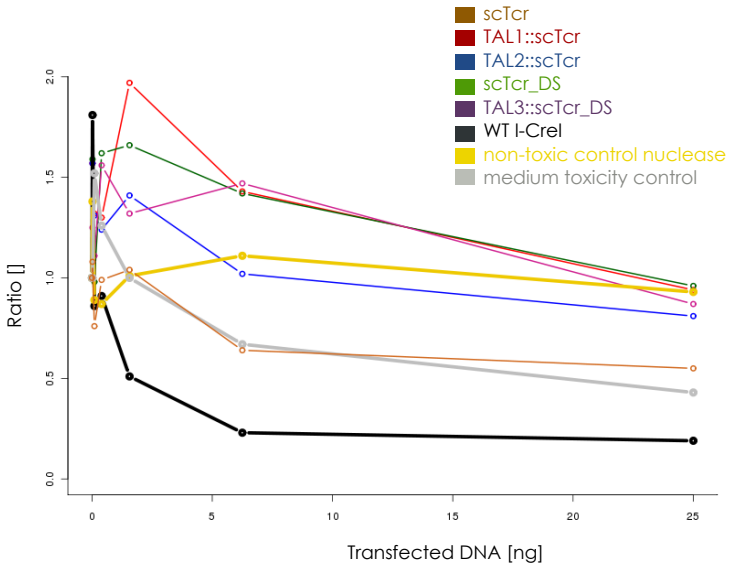



Figure 52: TALE fusion reduces toxicity of scTcr variants
Shown is a dose-response assay in CHO cells to determine nuclease toxicity *in vivo*. Engineered I-CreI variant scTcr (brown) is as toxic as a medium toxicity control (grey). Introduction of the degenerate scaffold mutations (green) or TALE-DBD fusion (red, blue, purple) removes measurable toxicity.

Table 15: Sandwich TALEN architecture
Shown are relative activities in a yeast SSA-assay at 37°C after 48h for the nuclease RagR::ColE7::Avr on exemplary targets. The EBEs on Rag-spacer-Avr targets are present on the same (!) strand and should allow binding of the complete construct in the right orientation. The nuclease was not active on the majority of assayed targets. Including the suitable Rag-Avr targets.

| CONSTRUCT | TARGET | ACTIVITY |
|---|---|---|
| Avr-ColE7-RagR | RagR-21-Avr | 0 |
| | Rag-Target | 0 |
| | Avr-21-Avr | 0,21 |

Table 16: Avr::scCre constructs
Shown are relative activities in a yeast SSA-assay at 37°C after 48h for Avr::scCre constructs on exemplary targets. N and C denominate to which subdomain of scCre the degenerate scaffold mutations have been applied. CFS is the RVD-less TALE fusion-scaffold. The mixed N/C variants are slightly more active than the full degenerate scaffold. Like WT scCre, they can cleave their target, even when no EBE is in proximity. Constructs are still fully active with the empty TALE-scaffold.

| CONSTRUCT VS TARGET | AVR-07-CRE | AVR-10-CRE | CRE | TCR |
|---|---|---|---|---|
| scCre | 1 | 1 | 1 | 0 |
| scCre_DS | 0,72 | 0,66 | 0,72 | 0 |
| scCre_DS-N | 0,84 | 0,89 | 0,96 | 0 |
| scCreDS-C | 1 | 1 | 1 | 0 |
| CFS::scCre | 1 | 1 | 1 | 0 |
| CFS::scCre_DS | 0 | 0 | 0 | 0 |
| CFS::scCre_DS-N | 1 | 1 | 1 | 0 |
| CFS::scCreDS-C | 0,97 | 0,97 | 1 | 0 |
| Avr::scCre | 1 | 1 | 1 | 0 |
| Avr::scCre_DS | 0,93 | 0 | 0 | 0 |
| Avr::scCre_DS-N | 0,99 | 0,93 | 1 | 0 |
| Avr::scCre_DS-C | 0,98 | 0,95 | 1 | 0 |

## A.2 SEQUENCES

### A.2.1 *Proteinsequences*

Given are protein sequences for catalytic domains and the AvrBs3 binding domains used in this work.

CATALYTIC DOMAIN OF COLICIN E7, N-TERMINAL FUSION
Blue: Colicin E7, residues 445-576
Red: Glycine/Serine linker

MASGSKRNKPGKATGKGKPVNNKWLNNAGKDLGSPVPDRIANKLRD
KEFKSFDDFRKKFWEEVSKDPELSKQFSRNNNDRMKVGKAPKTRTQ
DVSGKRTSFELHHEKPISQNGGVYDMDNISVVTPKRHIDIHRGKGS

CATALYTIC DOMAIN OF COLICIN E7, C-TERMINAL FUSION
Red: Glycine/Serine linker
Blue: Colicin E7, residues 445-576

GSSGSKRNKPGKATGKGKPVNNKWLNNAGKDLGSPVPDRIANKLRD
KEFKSFDDFRKKFWEEVSKDPELSKQFSRNNNDRMKVGKAPKTRTQ
DVSGKRTSFELHHEKPISQNGGVYDMDNISVVTPKRHIDIHRGKGS
SAD*

NUCLEASE A, C-TERMINAL FUSION
Red: Glycine/Serine linker
Blue: Nuclease A residues, 24-274

GSQVPPLTELSPSISVHLLLGNPSGATPTKLTPDNYLMVKNQYALS
YNNSKGTANWVAWQLNSSWLGNAERQDNFRPDKTLPAGWVRVTPSM
YSGSGYDRGHIAPSADRTKTTEDNAATFLMTNMMPQTPDNNRNTWG
NLEDYCRELVSQGKELYIVAGPNGSLGKPLKGKVTVPKSTWKIVVV
LDSPGSGLEGITANTRVIAVNIPNDPELNNDWRAYKVSVDELESLT
GYDFLSNVSPNIQTSIESKVDNAAD*

Endonuclease A, N-terminal fusion
Blue: DNA-entry nuclease A, residues 30-274
Red: Glycine/Serine linker

MHSAPNSPKTNLSQKKQASEAPSQALAESVLTDAVKSQIKGSLEWN
GSGAFIVNGNKTNLDAKVSSKPYADNKTKTVGKETVPTVANALLSK
ATRQYKNRKETGNGSTSWTPPGWHQVKNLKGSYTHAVDRGGLLGYA
LIGGLDGFDASTSNPKNIAVATAWANQAQAEYSTGQAYYESKVRKA
LDQNKRVRYRVTLYYASNEDLVPSASQIEAKSSDGELEFNVLVPNV
QKGLQLDYRTGEVTVTQGS

Endonuclease A, C-terminal fusion
Blue: DNA-entry nuclease A residues, 31-274
Orange: Polyhistidine-Tag

GTAPNSPKTNLSQKKQASEAPSQALAESVLTDAVKSQIKGSLEWNG
SGAFIVNGNKTNLDAKVSSKPYADNKTKTVGKETVPTVANALLSKA
TRQYKNRKETGNGSTSWTPPGWHQVKNLKGSYTHAVDRGGLLGYAL
IGGLDGFDASTSNPKNIAVATAWANQAQAEYSTGQAYYESKVRKAL
DQNKRVRYRVTLYYASNEDLVPSASQIEAKSSDGELEFNVLVPNVQ
KGLQLDYRTGEVTVTQHHHHHH*

Catalytic domain and linker of I-TevI, N-Terminal fusion
Blue: I-TevI, residues 2-183
Red: Serine/Glycine linker

MAKSGIYQIKNTLNNKVYVGSAKDFEKRWKRHFKDLEKGCHSSIKL
QRSFNKHGNVFECSILEEIPYEKDLIIERENFWIKELNSKINGYNI
ADATFGDTCSTHPLKEEIIKKRSETVKAKMLKLGPDGRKALYSKPG
SKNGRWNPETHKFCKCGVRIQTSAYTCSKCRNRSGENNSFFNHKHS
QGPSG

CATALYTIC DOMAIN AND LINKER OF I-TEVI, C-TERMINAL FUSION
Blue: I-TevI, residues 2-183
Red: Glycine/Serine Linker

GSKSGIYQIKNTLNNKVYVGSAKDFEKRWKRHFKDLEKGCHSSIKL
QRSFNKHGNVFECSILEEIPYEKDLIIERENFWIKELNSKINGYNI
ADATFGDTCSTHPLKEEIIKKRSETVKAKMLKLGPDGRKALYSKPG
SKNGRWNPETHKFCKCGVRIQTSAYTCSKCRNRSGENNSFFNHKHS
QGPSAD*

CATALYTIC DOMAIN OF FOKI, C-TERMINAL FUSION
Red: Glycine/Serine linker
Blue: Catalytic domain of variant FokI-L, residues 381-583

GSSGPNRGVTKQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNS
TQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGV
IVDTKAYSGGYNLPIGQADEMQRYVEENQTRNKHINPNEWWKVYPS
SVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGE
MIKAGTLTLEEVRRKFNNGEINFGSSAD*

SINGLE-CHAIN I-CREI, C-TERMINAL FUSION
Red: Glycine/Serine linker
Purple:DNA binding helix (I-TevI WT residues 184-199)
Orange: Single-chain linker
Blue: I-CreI variant (residues 2-163 and 6-163)
Green: LAGLIDADG motif

GSDITKSKISEKMKGQGPSGNTKYNKEFLLYLAGFVDGDGSIIAQI
KPNQSYKFKHQLSLTFQVTQKTQRRWFLDKLVDEIGVGYVRDRGSV
SDYILSEIKPLHNFLTQLQPFLKLKQKQANLVLKIIEQLPSAKESP
DKFLEVCTWVDQIAALNDSKTRKTTSETVRAVLDSLSEKKKSSPAA
GDSSVSNSEHIAPLSLPSSPPSVGSNKEFLLYLAGFVDGDGSIIAQ
IKPNQSYKFKHQLSLTFQVTQKTQRRWFLDKLVDEIGVGYVRDRGS
VSDYILSEIKPLHNFLTQLQPFLKLKQKQANLVLKIIEQLPSAKES
PDKFLEVCTWVDQIAALNDSKTRKTTSETVRAVLDSLSEKKKSSP*

SINGLE-CHAIN I-CREI, DEGENERATE-SCAFFOLD, C-TERMINAL FUSION
Red: Glycine/Serine linker
Purple:DNA binding helix (I-TevI WT residues 184-199)
Orange: Single-chain linker
Blue: I-CreI variant (residues 2-163 and 6-163)
Green: LAGLIDADG motif

GSDITKSKISEKMKGQGPSGNTKYNKEFLLYLAGFVDGDGSIIACI
RPNQTCKFKHQLSLTFQVTQKTQRRWFLDKLVDEIGVGYVRDRGSV
SDYILSEIAPLHNFLTQLQPFLKLKQKQANLVLKIIEQLPSAKESP
DKFLEVCTWVDQIAALNDSKTRKTTSETVRAVLDSLSEKKKSSPAA
GDSSVSNSEHIAPLSLPSSPPSVGSNKEFLLYLAGFVDGDGSIIAC
IRPNQTCKFKHQLSLTFQVTQKTQRRWFLDKLVDEIGVGYVRDRGS
VSDYILSEIKPLHNFLTQLQPFLKLKQKQANLVLKIIEQLPSAKES
PDKFLEVCTWVDQIAALNDSKTRKTTSETVRAVLDSLSEKKKSSP*

ENGINEERED MEGANUCLEASE SCTCR, C-TERMINAL FUSION
Red: Glycine/Serine linker
Purple:DNA binding helix (I-TevI WT residues 184-199)
Orange: Single-chain linker
Blue: I-CreI variant (residues 2-163 and 6-163)
Green: LAGLIDADG motif

GSDITKSKISEKMKGQGPSGNTKYNEEFLLYLAGFVDGDGSIVAQI
KPHQSCKFKHQLRLTFNVTQKTQRRWFLDKLVDEIGVGHVYDSGSV
SYYQLSEIKPLHNFLTQLQPFLELKQKQANLVLKIIEQLPSAKESP
DKFLEVCTWVDQVAALNDSKTRKTTSETVRAVLDSLSEKKKSSPAA
GDSSVSNSEHIAPLSLPSSPPSVGSNKKFLLYLAGFVDSDGSIIAQ
IKPNQSYKFKHYLSLTFCVTQKTQRRWFLDKLVDRIGVGYVQDRGS
VSDYRLSEIKPLRNFLTQLQPFLKLKQKQANLVLKIIEQLPSAKES
PDKFLEVCTWVDQVAALNDSKTRKTTSETVRAVLDSLSEKKKSSP*

SNAP-TAG, N-TERMINAL FUSION
Red: Glycine/Serine linker
Blue: SNAP-tag, corresponding to residues 4-181 in O6-alkylguanine-DNA alkyltransferase (human)
Orange: Polyhistidine-tag

GSSDKDCEMKRTTLDSPLGKLELSGCEQGLHEIKLLGKGTSAADAV
EVPAPAAVLGGPEPLMQATAWLNAYFHQPEAIEEFPVPALHHPVFQ
QESFTRQVLWKLLKVVKFGEVISYQQLAALAGNPAATAAVKTALSG
NPVPILIPCHRVVSSSGAVGGYEGGLAVKEWLLAHEGHRLGKPGLG
PAGGSHHHHHH*

SCTREX, N-TERMINAL FUSION
Red: Glycine/Serine linker
Blue:Three prime repair exonuclease 2 (human), residues 45-279 (x2)
Green: internal linker

MGSEAPRAETFVFLDLEATGLPSVEPEIAELSLFAVHRSSLENPEH
DESGALVLPRVLDKLTLCMCPERPFTAKASEITGLSSEGLARCRKA
GFDGAVVRTLQAFLSRQAGPICLVAHNGFDYDFPLLCAELRRLGAR
LPRDTVCLDTLPALRGLDRAHSHGTRARGRQGYSLGSLFHRYFRAE
PSAAHSAEGDVHTLLLIFLHRAAELLAWADEQARGWAHIEPMYLPP
DDPSLEATPPQTGLDVPYSEAPRAETFVFLDLEATGLPSVEPEIAE
LSLFAVHRSSLENPEHDESGALVLPRVLDKLTLCMCPERPFTAKAS
EITGLSSEGLARCRKAGFDGAVVRTLQAFLSRQAGPICLVAHNGFD
YDFPLLCAELRRLGARLPRDTVCLDTLPALRGLDRAHSHGTRARGR
QGYSLGSLFHRYFRAEPSAAHSAEGDVHTLLLIFLHRAAELLAWAD
EQARGWAHIEPMYLPPDDPSLEAGGGGSGGGGS

AvrBs3(11) N-terminal fusion scaffold
"NS" repeats have been replaced with the more stringent "NI" repeats in this version
Red: Glycine/Serine linker
Light blue: N-terminal pseudo repeats, AvrBs3 residues 153-288
Blue: TALE-repeats, AvrBs3 residues 289-866
Orange: RVDs
Green: half-repeat, AvrBs3 residues 867-886
Purple: AvrBs3 C-terminus, residues 887-897

MANTGGSSGVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFT
HAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGA
RALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNA
LTGAPLNLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPQQ
VVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQ
ALETVQALLPVLCQAHGLTPQQVVAIASNGGGKQALETVQRLLPVL
CQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVV
AIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQAL
ETVQALLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQ
AHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPQQVVAI
ASNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALET
VQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQAH
GLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIAS
HDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQ
RLLPVLCQAHGLTPQQVVAIASNGGKQALETVQRLLPVLCQAHGL
TPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNG
GGRPALESIVAQLSRPDPSAD*

AvrBs3(11) C-terminal fusion scaffold
"NS" repeats have been replaced with the more stringent "NI" repeats in this version
Light blue: N-terminal pseudo repeats, AvrBs3 residues 153-288
Blue: TALE-repeats, AvrBs3 residues 289-866
Orange: RVDs
Green: half-repeat, AvrBs3 residues 867-886
Purple: AvrBs3 C-terminus, residues 887-897
Red: Glycine/Serine linker

MVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALS
QHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLT
VAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNL
TPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNG
GGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQAL
LPVLCQAHGLTPQQVVAIASNGGGKQALETVQRLLPVLCQAHGLTP
EQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGG
KQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLP
VLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQ
VVAIASHDGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNGGGKQ
ALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVL
CQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVV
AIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQAL
ETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQ
AHGLTPQQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAI
ASHDGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNGGGRPALES
IVAQLSRPDPGSSGRSSAD*

NTER_AVRBS3(28) SCAFFOLD
Red: Glycine/Serine linker
Light blue: N-terminal pseudo repeats, AvrBs3 residues 153-288
Blue: TALE-repeats, AvrBs3 residues 289-866
Orange: RVDs
Green: half-repeat, AvrBs3 residues 867-886
Purple: AvrBs3 C-terminus, residues 887-914
Underlined: Polyhistidine-tag

MHSGSVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHI
VALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALE
ALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGA
PLNLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPQQVVAI
ASNGGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNSGGKQALET
VQRLLPVLCQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAH
GLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIAS
NIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQ
ALLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGL
TPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNG
GGKQALETVQRLLPVLCQAHGLTPEQVVAIASNSGGKQALETVQAL
LPVLCQAHGLTPEQVVAIASNSGGKQALETVQRLLPVLCQAHGLTP
EQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGG
KQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLP
VLCQAHGLTPQQVVAIASNGGGRPALETVQRLLPVLCQAHGLTPEQ
VVAIASHDGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNGGGRP
ALESIVAQLSRPDPALAALTNDHLVALACLGSSHHHHHH*

AvrBs3-DBD IN EndA::TALE::EndA
Red: Glycine/Serine linkers
Light blue: N-terminal pseudo repeats, AvrBs3 residues 153-288
Blue: TALE-repeats, AvrBs3 residues 289-866
Orange: RVDs
Green: half-repeat, AvrBs3 residues 867-886
Purple: WT FokI-linker, residues 379-391

GSVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVAL
SQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALL
TVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLN
LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPQQVVAIASN
GGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNSGGKQALETVQR
LLPVLCQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLT
PEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIG
GKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALL
PVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPE
QVVAIASHDGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNGGGK
QALETVQRLLPVLCQAHGLTPEQVVAIASNSGGKQALETVQALLPV
LCQAHGLTPEQVVAIASNSGGKQALETVQRLLPVLCQAHGLTPEQV
VAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQA
LETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLC
QAHGLTPQQVVAIASNGGGRPALETVQRLLPVLCQAHGLTPEQVVA
IASHDGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNGGGRPALE
GSSVIPNRGVTKQLVKGT

### A.2.2 *Oligonucleotides*

The following tables contain oligonucleotide sequences (forward primer (for) and reverse primer (rev)) used in this work for mutagenesis, screening or insertion of multiple cloning site (MCS)s. Sequences for common primers (e. g., T7 forward, M13 reverse, etc.) are not shown here.

### A.2.3 *Targets*

Recognition sites for TALEs and meganucleases are given in table 25.

Table 17: General cloning oligonucleotides

| NAME | SEQUENCE | COMMENTS |
|---|---|---|
| CMP_001 | ACACGCAAACACAAATACACAGCGGCCTTGCCACCATGG | pCLS7865 (for) (and 542) |
| CMP_003 | AGAAGTCCAAAGCTTCAGCTGCTGCAGGCTCGAGGAGCTC | pCLS7865 (rev) (and 542) |
| CMP_124 | GCTGGAGAGCATTGTTGCCCAGTTATCTCGCCC | TALE C-ter (for) |
| CMP_125 | CCTGTTGCTGCTGGCTGTAGCCGAGCGTGCGTAG | TALE N-ter (rev) |
| CMP_169 | GGAGAGGACACGCACGAGATCTG | pCLS15603 (for) |
| CMP_170 | GGCAACAGGATTCAATCTTAAG | pCLS15603 (rev) |
| CMP_145 | TAAGGAGGAAACATTCATGGGAGACGGACACAGCATCCTCGTGGCGTCTCAGGCCAACTGATTTAAT | Insert (sense) to create pCLS12081 |
| CMP_146 | TAAATCAGTTGGCCTGAGACGCCACGAGGATGCTGTGTCCGTCTCCCATGAATGTTTCCTCCTTAAT | Insert (anti-sense) to create pCLS12081 |
| CMP_171 | AGCTGTTCTAGAGGCGCGCCACACGCAAACACAAATAC | primer (for) to insert 7865 MCS into 15603 |
| CMP_172 | AAAGCTAGCGCTCGCCGGCGCTCGAGGAGCTCTTATCAGTGG | primer (rev) to insert 7865 MCS into 15603 |
| Oligo_571 | TTCCGGATCCGTCGACCTTAGAACATTGGGATATTCACAACAACAGCAGGAGAAGATTAAGCC | TALE Nter EndA-MCS insert (sense) |
| Oligo_572 | TTAGGCTTAATCTTCTCCTGCTGTTGTTGTGAATATCCCAATGTTCTAAGGTCGACGGATCCGGAATGCA | TALE Nter EndA-MCS insert (anti-sense) |
| Oligo_68 | CATTACTGGATCTATCAACAGGAG | pQE30 (rev) |
| Oligo_126 | GTATCACGAGGCCCTTTCGTCT | pQE30 (for) |
| Oligo_241 | ATTGAGCAGACCCGATCCTGCTTTGG | TALE C-ter (for) |
| Oligo_283 | CCTGCTGTTGTTGTGAATATCCCAAT | TALE N-ter (rev) |
| Tal_N-Ter_Rev | CTGTAGATCTAACCTTAGGCTTAATCTTCTCC | TALE N-ter (rev) |

Table 18: ColE7 SOEing primers

| NAME | SEQUENCE | COMMENTS |
|---|---|---|
| CMP_082 | GAATTTCTTGCGAAAGTCTGCAAAACTCTTGAATTC | D493A (rev) |
| CMP_083 | GAATTCAAGAGTTTTGCAGACTTTCGCAAGAAATTC | D493A (for) |
| CMP_084 | CTTCCCAGAATTTCTTGGCAAAGTCATCAAAACTC | R496A (rev) |
| CMP_085 | GAGTTTTGATGACTTTGCCAAGAAATTCTGGGAAG | R496A (for) |
| CMP_86 | CACTTCTTCCCAGAATTTTGCGCGAAAGTCATCAAAAC | K497A (rev) |
| CMP_87 | GTTTTGATGACTTTCGCGCAAAATTCTGGGAAGAAGTG | K497A (for) |
| CMP_88 | GAGAGATTGGTTTTTCCTGATGCAATTCAAAAC | H545Q (rev) |
| CMP_89 | GTTTTGAATTGCATCAGGAAAAACCAATCTCTC | H545Q (for) |
| CMP_90 | GTAACCACAGAGATGTCGTCCATGTCGTAAACG | N560D (rev) |
| CMP_91 | CGTTTACGACATGGACGACATCTCTGTGGTTAC | N560D (for) |
| CMP_92 | GAGGATCCTTTACCGCGAGCAATATCGATATGTCTTTTC | H573A (rev) |
| CMP_93 | GAAAAGACATATCGATATTGCTCGCGGTAAAGGATCCTC | H573A (for) |
| CMP_94 | GAGGATCCTTTACCGCGTTCAATATCGATATGTCTTTTC | H573E (rev) |
| CMP_95 | GAAAAGACATATCGATATTGAACGCGGTAAAGGATCCTC | H573E (for) |
| CMP_153 | GAGAGATTGGTTTTTCAGCATGCAATTCAAAAC | H545A (rev) |
| CMP_154 | GTTTTGAATTGCATGCTGAAAAACCAATCTCTC | H545A (for) |
| CMP_155 | CTTACCTGGTTTGTTGCGCTGAGATCCGGAGGATCC | K446Q (rev) |
| CMP_156 | GGATCCTCCGGATCTCAGCGCAACAAACCAGGTAAG | K446Q (for) |
| CMP_157 | CTTACCTGGTTTGTTGCGGTAAGATCCGGAGGATCC | K446Y (rev) |
| CMP_158 | GGATCCTCCGGATCTTACCGCAACAAACCAGGTAAG | K446Y (for) |
| CMP_159 | GCCTTACCTGGTTTGTTGGCCTTAGATCCGGAGGATC | R447A (rev) |
| CMP_160 | GATCCTCCGGATCTAAGGCCAACAAACCAGGTAAGGC | R447A (for) |
| CMP_161 | GCCTTACCTGGTTTGTTCTTCTTAGATCCGGAGGATC | R447K (rev) |
| CMP_162 | GATCCTCCGGATCTAAGAAGAACAAACCAGGTAAGGC | R447K (for) |
| CMP_163 | GCCTTACCTGGTTTGTTCTTCTGAGATCCGGAGGATCC | K446Q, R447K (rev) |
| CMP_164 | GGATCCTCCGGATCTCAGAAGAACAAACCAGGTAAGGC | K446Q, R447K (for) |
| CMP_165 | GCCTTACCTGGTTTGTTGGCGTAAGATCCGGAGGATCC | K446Y,R447A (rev) |
| CMP_166 | GGATCCTCCGGATCTTACGCCAACAAACCAGGTAAGGC | K446Y,R447A (for) |
| CMP_167 | GAATTTCTTGCGAAAGTCCTGAAAACTCTTGAATTC | D493Q (rev) |
| CMP_168 | GAATTCAAGAGTTTTCAGGACTTTCGCAAGAAATTC | D493Q (for) |

Table 19: Inhibitor related primers

| NAME | SEQUENCE | COMMENTS |
|---|---|---|
| CMP_147 | CGATTGCCATGGAACTGAAGAACTCCATCAGCG | Im7 (for), rem. BamHI |
| CMP_148 | CCACCACCATGGCCACCAAGACCAACAGCGAGATTC | NuiA (for), rem. BamHI |

Table 20: EndA primers

| NAME | SEQUENCE | COMMENTS |
|---|---|---|
| Oligo_573 | TAATATGCATTCCGCACCTAATAGTCCCAAAACC | EndA (for), +NsiI |
| Oligo_574 | CTGGGATCCCTGAGTTACAGTTACTTCTCC | EndA (rev), +BamHI |
| 160H_Forward | GATCGAGGTCACTTGCTCGGGTATGCCTTAATCG | SOE (for), G160H |
| 160H_Reverse | CCCGAGCAAGTGACCTCGATCGACTGCATGCG | SOE(rev), G160H |
| X160_Reverse | ACCTCGATCGACTGCATGCGTATAAG | SOE (rev), X160 |
| H160_Forward2 | ATGCAGTCGATCGAGGTCATTTGCTCGGGTATGCC | SOE (for), G160H |
| N160_Forward | ATGCAGTCGATCGAGGTAACTTGCTCGGGTATGCC | SOE (for), G160N |
| Q160_Forward | ATGCAGTCGATCGAGGTCAATTGCTCGGGTATGCC | SOE (for), G160Q |

Table 21: *In vitro* substrate oligos

| NAME | SEQUENCE | COMMENTS |
|---|---|---|
| Oligo_209 | GCTATATGCGTTGATGCAATTTCTATGCGCACCC | pAT (for) |
| Oligo_396 | ACCCAGAGCGCTGCCGGCAC | pAT (rev) |
| l/b S/H_For | CGTTCTCGGAGCACTGTCCGACCGCTTTGG | Atto488 top (for) |
| lab T4-x-H_rev | GCACCGCCGCCGCAAGGAATGGTGC | Atto647 bottom (rev) |

Table 22: SNAP-tag primers

| NAME | SEQUENCE | COMMENTS |
|---|---|---|
| Oligo_597 | AACATATGCATGGGAGCTCAGACAAAGATTGC-GAAATG | SNAP (for), for N&C, +NsiI/SacI |
| Oligo_598 | AATTAAGCTTTTAGTGGTGATGATGATGGTGG-GATCCGCCTGCAGGTCCC | SNAP (rev), for N&C, +BamHI/6xHis/HindIII |

Table 23: I-CreI primers

| NAME | SEQUENCE | COMMENTS |
|---|---|---|
| CMP_191 | AATACCGGATCCAACAAAGAGTTCCTGCTGTA | Cre (for), +BamHI |
| CMP_192 | AATACCAGATCTAACAAAGAGTTCCTGCTGTA | Cre (for), +BglII |
| CMP_193 | TGCAGGCTCGAGTTACTACGGGGAGGATTTCT-TCTTCTCGCTC | Cre (rev), first mono. |
| CMP_194 | TTTGTTGGATCCAACAGATGGAGGAGAGGAAGGC | Cre (rev), second mono. |
| CMP_195 | TTTGTTGGATCCAACAGATGGAGGAGAGGAAGGC | Cre linker (rev) |

Table 24: I-TevI primers

| NAME | SEQUENCE | COMMENTS |
|---|---|---|
| CMP_069 | GCCACCGGATCCAAGTCTGGCATCTACCAGATTAAG | I-TevI, (for) |
| CMP_118 | GAAGTGCCTCTTCCATGCTTTCTCAAAGTCCTTG | SOE (rev), R27A |
| CMP_071 | TGGTATCGGCCGATGGACCTTGAGAATGCTTGTGG | I-TevI, (rev) |
| CMP_119 | CAAGGACTTTGAGAAAGCATGGAAGAGGCACTTC | SOE (for), R27A |

Table 25: Recognition sites

| NAME | SEQUENCE | COMMENTS |
|---|---|---|
| AvrBs3 | TCTATAAACCTAACCCTCT | |
| RagT2-L | TATATTTAAGCACTTTAT | |
| RagT2-R | TGTTTATGGTTACTTAT | |
| NPTII-A | TCCTTGCGCAGCTGTG | |
| NPTII-B | TAGCAGCCAGTCCCTTC | |
| NPTII-cB | TGAAGCGGGAAGGGACT | |
| NPTII-C | TGACAGGAGATCCTGCC | |
| I-CreI | TCAAAACGTCGTACGACGTTTTGA | optimized |
| TCR (EBE) | TGCTGGTCAGCGCCC | TALE |
| TCR | GATGGCCATGGTAAGCAGGAGGGC | Meganuclease |

Table 26: Common media

| BUFFER | COMPONENT | CONCENTRATION |
|---|---|---|
| LB-medium (lysogeny broth) | Yeast extract | 5 g/l |
| | Tryptone | 10 g/l |
| | NaCl | 5 g/l |
| SOB/SOC medium | Bacto-tryptone | 2 g / 100 ml |
| pH 6-7 with NaOH | Yeast extract | 0.5 g / 100 ml |
| | NaCl | 10 mM |
| | KCl | 0,5 mM |
| | $MgCl_2$ | 10 mM |
| | $MgSO_4$ | 10 mM |
| For SOC medium | after autoclaving | |
| | add glucose | 20 mM |
| Phosphate-buffered saline (PBS) | NaCl | 137 mM |
| | KCl | 2.7 mM |
| | $Na_2HPO_4$ | 10 mM |
| | $KH_2PO_4$ | 1.8 mM |

## A.3 BUFFERS, MEDIA AND CHEMICALS

Common media, electrophoresis buffers and antibiotics can be found in this section. Chemicals used in this work were molecular biology grade and were obtained from AppliChem, Merk, Roth, Qiagen or Sigma-Aldrich.

Table 27: Electrophoresis buffers

| BUFFER | COMPONENT | CONCENTRATION |
|---|---|---|
| TAE buffer (1X) | Tris | 40 mM |
| pH 8.0 with acetic acid | Sodium acetate | 20 mM |
| | EDTA | 1 mM |
| TBE buffer (1X) | Tris | 100 mM |
| pH 8.3 with boric acid | Borate | 100 mM |
| | EDTA | 2.5 mM |
| TPE buffer (1X) | Tris | 90 mM |
| pH 8.2 with phosphoric acid | Phosphate | 90 mM |
| | EDTA | 2 mM |
| TTE buffer 1X | Tris | 100 mM |
| pH 9.0 | Taurine | 28.8 mM |
| | EDTA | 1 mM |
| Laemmli running buffer | Tris | 25 mM |
| | Glycine | 190 mM |
| | SDS | 1% (w/v) |

Table 28: Antibiotics

| NAME | CONCENTRATION (PLATE) | CONCENTRATION (MEDIUM) |
|---|---|---|
| Ampicillin | 100 (200) µg / ml | 75 µg / ml |
| Chloramphenicol | 30 µg / ml | 20 µg / ml |
| Kanamycin | 25 µg / ml | 25 µg / ml |
| Tetracylcin | 10 µg / ml | 10 µg / ml |

# BIBLIOGRAPHY


[1] (2012, January). Method of the Year 2011. *Nat Meth 9*(1), 1–1. (Cited on page 87.)

[2] (2012). The Runners-Up. *Science 338*(6114), 1525–1532. (Cited on page 87.)

[3] Ahloowalia, B., M. Maluszynski, and K. Nichterlein (2004). Global impact of mutation-derived varieties. *135*(2), 187–204–. (Cited on page 3.)

[4] Ain, Q. U., J. Y. Chung, and Y.-H. Kim (2014). Current and future delivery systems for engineered nucleases: ZFN, TALEN and RGEN. *Journal of Controlled Release 205*, 120–127. (Cited on page 4.)

[5] Alton, E., M. Stern, R. Farley, A. Jaffe, S. Chadwick, J. Phillips, J. Davies, S. Smith, J. Browning, M. Davies, M. Hodson, S. Durham, D. Li, P. Jeffery, M. Scallan, R. Balfour, S. Eastman, S. Cheng, A. Smith, D. Meeker, and D. Geddes (1999, March). Cationic lipid-mediated CFTR gene transfer to the lungs and nose of patients with cystic fibrosis: a double-blind placebo-controlled trial. *The Lancet 353*(9157), 947–954. (Cited on page 16.)

[6] Arnould, S., P. Chames, C. Perez, E. Lacroix, A. Duclert, J.-C. Epinat, F. Stricher, A.-S. Petit, A. Patin, S. Guillier, S. Rolland, J. Prieto, F. J. Blanco, J. Bravo, G. Montoya, L. Serrano, P. Duchateau, and F. Pâques (2006, January). Engineering of Large Numbers of Highly Specific Homing Endonucleases that Induce Recombination on Novel DNA Targets. *Journal of Molecular Biology 355*(3), 443–458. (Cited on pages 15, 24, and 57.)

[7] Arnould, S., C. Delenda, S. Grizot, C. Desseaux, F. Pâques, G. H. Silva, and J. Smith (2011, January). The I-CreI meganuclease and its engineered derivatives: applications from cell modification to gene therapy. *Protein engineering, design & selection : PEDS 24*(1-2), 27–31. (Cited on page 18.)

[8] Arnould, S., C. Perez, J.-P. Cabaniols, J. Smith, A. Gouble, S. Grizot, J.-C. Epinat, A. Duclert, P. Duchateau, and F. Pâques (2007, August). Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *Journal of Molecular Biology 371*(1), 49–65. (Cited on page 15.)

[9] Aryan, A., M. A. E. Anderson, K. M. Myles, and Z. N. Adelman (2013, March). Germline excision of transgenes in Aedes aegypti

by homing endonucleases. *Scientific Reports 3*, 1603–. (Cited on pages 16 and 18.)

[10] Aryan, A., K. M. Myles, and Z. N. Adelman (2014, February). Targeted genome editing in aedes aegypti using TALENs. *Methods*. (Cited on page 16.)

[11] Ausubel FM, Brent R, K. R. M. D. S. J. S. J. S. K. (1995). *Current Protocols in Molecular Biology*. Wiley. (Cited on pages 94 and 101.)

[12] Bakhrat, A., M. S. Jurica, B. L. Stoddard, and D. Raveh (2004, March). Homology modeling and mutational analysis of Ho endonuclease of yeast. *Genetics 166*(2), 721–8. (Cited on page 80.)

[13] Bansal, S. and T. P. Durrett (2015). Camelina sativa: An ideal platform for the metabolic engineering and field production of industrial lipids. *Biochimie* (0), –. (Cited on page 16.)

[14] Basu, S., A. Aryan, J. M. Overcash, G. H. Samuel, M. a. E. Anderson, T. J. Dahlem, K. M. Myles, and Z. N. Adelman (2015). Silencing of end-joining repair for efficient site-specific gene insertion after TALEN/CRISPR mutagenesis in aedes aegypti. *PNAS 112*(13), 201502370. (Cited on page 16.)

[15] Beiter, K., F. Wartha, B. Albiger, S. Normark, A. Zychlinsky, and B. Henriques-Normark (2006, February). An Endonuclease Allows Streptococcus pneumoniae to Escape from Neutrophil Extracellular Traps. *Current Biology 16*(4), 401–407. (Cited on page 13.)

[16] Belfort, M. and R. P. Bonocora (2014). Homing Endonucleases: From Genetic Anomalies to Programmable Genomic Clippers. *Methods in molecular biology (Clifton, N.J.) 1123*, 1–26. (Cited on page 18.)

[17] Bell-Pedersen, D., S. M. Quirk, M. Bryk, and M. Belfort (1991, September). I-TevI, the endonuclease encoded by the mobile td intron, recognizes binding and cleavage domains on its DNA target. *PNAS 88*(17), 7719–7723. (Cited on page 12.)

[18] Bernstein, D. L., J. E. Le Lay, E. G. Ruano, and K. H. Kaestner (2015, May). TALE-mediated epigenetic suppression of CDKN2A increases replication in human fibroblasts. *J Clin Invest 125*(5), 1998–2006. (Cited on page 9.)

[19] Bétermier, M., P. Bertrand, and B. S. Lopez (2014). Is Non-Homologous End-Joining Really an Inherently Error-Prone Process? *PLoS Genetics 10*(1). (Cited on pages 20, 21, and 23.)

[20] Beumer, K. J., J. K. Trautman, M. Christian, T. J. Dahlem, C. M. Lake, R. S. Hawley, D. J. Grunwald, D. F. Voytas, and D. Carroll (2013, August). Comparing Zinc Finger Nucleases and Transcription Activator-Like Effector Nucleases for Gene Targeting in

Drosophila. *G3: Genes | Genomes | Genetics* 3(10), 1717–1725. (Cited on pages 17 and 25.)

[21] Beurdeley, M., F. Bietz, J. Li, S. Thomas, T. Stoddard, A. Juillerat, F. Zhang, D. F. Voytas, P. Duchateau, and G. H. Silva (2013, January). Compact designer TALENs for efficient genome engineering. *Nature Communications 4*, 1762. (Cited on pages 12, 36, 50, 51, 79, and 107.)

[22] Bevan, M. W., R. B. Flavell, and M.-D. Chilton (1983, July). A chimaeric antibiotic resistance gene as a selectable marker for plant cell transformation. *Nature 304*(5922), 184–187. (Cited on page 16.)

[23] Bitinaite, J., D. A. Wah, A. K. Aggarwal, and I. Schildkraut (1998, April). FokI dimerization is required for DNA cleavage. *PNAS 95*(18), 10570–10575. (Cited on page 10.)

[24] Blender Online Community (2015). *Blender - a 3D modelling and rendering package*. Blender Institute, Amsterdam: Blender Foundation. www.lyx.org. (Cited on page 107.)

[25] Boch, J. and U. Bonas (2010, July). Xanthomonas AvrBs3 Family-Type III Effectors: Discovery and Function. *Annu. Rev. Phytopathol. 48*(1), 419–436. (Cited on page 6.)

[26] Boch, J., H. Scholze, S. Schornack, A. Landgraf, S. Hahn, S. Kay, T. Lahaye, A. Nickstadt, and U. Bonas (2009, December). Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science 326*(5959), 1509–1512. (Cited on page 6.)

[27] Boissel, S., J. Jarjour, A. Astrakhan, A. Adey, A. Gouble, P. Duchateau, J. Shendure, B. L. Stoddard, M. T. Certo, D. Baker, and A. M. Scharenberg (2014). MegaTALs: A rare-cleaving nuclease architecture for therapeutic genome engineering. *Nucleic Acids Research 42*, 2591–2601. (Cited on pages 7, 16, 80, and 81.)

[28] Bonas, U., R. Stall, and B. Staskawicz (1989). Genetic and structural characterization of the avirulence gene avrBs3 from Xanthomonas campestris pv. vesicatoria. *218*(1), 127–136–. (Cited on page 6.)

[29] Cascales, E., S. K. Buchanan, D. Duché, C. Kleanthous, R. Lloubès, K. Postle, M. Riley, S. Slatin, and D. Cavard (2007, March). Colicin Biology. *Microbiology and Molecular Biology Reviews 71*(1), 158–229. (Cited on page 12.)

[30] Cermak, T., E. L. Doyle, M. Christian, L. Wang, Y. Zhang, C. Schmidt, J. A. Baller, N. V. Somia, A. J. Bogdanove, and D. F. Voytas (2011, March). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research 39*(12), e82–e82. (Cited on page 71.)

[31] Chan, S. J., J. Weiss, M. Konrad, T. White, C. Bahl, S. D. Yu, D. Marks, and D. F. Steiner (1981, September). Biosynthesis and periplasmic segregation of human proinsulin in Escherichia coli. *PNAS 78*(9), 5401–5405. (Cited on page 16.)

[32] Chang, A. C. Y. and S. N. Cohen (1974, April). Genome Construction Between Bacterial Species In Vitro: Replication and Expression of Staphylococcus Plasmid Genes in Escherichia coli. *PNAS 71*(4), 1030–1034. (Cited on page 16.)

[33] Chatterjee, P., K. L. Brady, A. Solem, Y. Ho, and M. G. Caprara (2003). Functionally distinct nucleic acid binding sites for a group I intron encoded RNA maturase/DNA homing endonuclease. *Journal of Molecular Biology 329*(03), 239–251. (Cited on page 81.)

[34] Chayot, R., B. Montagne, D. Mazel, and M. Ricchetti (2010, February). An end-joining repair mechanism in Escherichia coli. *PNAS 107*(5), 2141–2146. (Cited on page 23.)

[35] Chen, Q. and H. Lai (2014, November). Gene Delivery into Plant Cells for Recombinant Protein Production. *BioMed Research International 2015*, 932161–. (Cited on page 16.)

[36] Chen, R., G. Xue, P. Chen, B. Yao, W. Yang, Q. Ma, Y. Fan, Z. Zhao, M. Tarczynski, and J. Shi (2008). Transgenic maize plants expressing a fungal phytase gene. *17*(4), 633–643–. (Cited on page 16.)

[37] Cheng, X., K. Balendiran, I. Schildkraut, and J. E. Anderson (1994, September). Structure of PvuII endonuclease with cognate DNA. *The EMBO Journal 13*(17), 3927–3935. (Cited on page 15.)

[38] Chevalier, B. S., R. J. Monnat, and B. L. Stoddard (2001, April). The homing endonuclease I-CreI uses three metals, one of which is shared between the two active sites. *Nat Struct Mol Biol 8*(4), 312–316. (Cited on page 15.)

[39] Chevalier, B. S. and B. Stoddard (2001, April). Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Research 29*(18), 3757–3774. (Cited on page 18.)

[40] Cho, H.-S., J. Gu Kang, J.-H. Lee, J.-J. Lee, S. K. Jeon, J.-H. Ko, D.-S. Kim, K.-H. Park, Y.-S. Kim, and N.-S. Kim (2015). Direct regulation of E-cadherin by targeted histone methylation of TALE-SET fusion protein in cancer cells. *Oncotarget; Advance Online Publications: Page 3*, –. (Cited on page 9.)

[41] Choulika, A., A. Perrin, B. Dujon, and J. F. Nicolas (1995, April). Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of saccharomyces cerevisiae. *Molecular and Cellular Biology 15*(4), 1968–1973. (Cited on page 18.)

[42] Christian, M., T. Cermak, E. L. Doyle, C. Schmidt, F. Zhang, A. Hummel, A. J. Bogdanove, and D. F. Voytas (2010, July). Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics 186*(2), 757–761. (Cited on page 10.)

[43] Cletus, J., V. Balasubramanian, D. Vashisht, and N. Sakthivel (2013). Transgenic expression of plant chitinases to enhance disease resistance. *35*(11), 1719–1732–. (Cited on page 16.)

[44] Cohen, S. N., A. C. Y. Chang, H. W. Boyer, and R. B. Helling (1973, November). Construction of Biologically Functional Bacterial Plasmids In Vitro. *PNAS 70*(11), 3240–3244. (Cited on pages 3 and 16.)

[45] Cohen, S. N., A. C. Y. Chang, and L. Hsu (1972, August). Nonchromosomal Antibiotic Resistance in Bacteria: Genetic Transformation of Escherichia coli by R-Factor DNA. *PNAS 69*(8), 2110–2114. (Cited on page 19.)

[46] Cong, L., F. Ran, D. Cox, S. Lin, and R. Barretto (2013, July). Multiplex genome engineering using CRISPR/Cas systems. *Science 25*(7), 778–785. (Cited on page 19.)

[47] Cong, L., R. Zhou, Y.-c. Kuo, M. Cunniff, and F. Zhang (2012, July). Comprehensive Interrogation of Natural TALE DNA Binding Modules and Transcriptional Repressor Domains. *Nature Communications 3*, 968–968. (Cited on page 7.)

[48] Cooper, C., E. Maga, and J. Murray (2015). Production of human lactoferrin and lysozyme in the milk of transgenic dairy animals: past, present, and future. pp. 1–10–. (Cited on page 16.)

[49] Cousins, R. J., J. P. Liuzzi, and L. a. Lichten (2006). Mammalian zinc transport, trafficking, and signals. *Journal of Biological Chemistry 281*(34), 24085–24089. (Cited on page 75.)

[50] Crooks, G. E., G. Hon, J.-M. Chandonia, and S. E. Brenner (2004, January). Weblogo: A sequence logo generator. *Genome Research 14*(6), 1188–1190. (Cited on page 107.)

[51] Cuculis, L., Z. Abil, H. Zhao, and C. M. Schroeder (2015). Direct observation of TALE protein dynamics reveals a two-state search mechanism. *Nature Communications 6*, 7277. (Cited on pages 7 and 72.)

[52] Daboussi, F., M. Zaslavskiy, L. Poirot, M. Loperfido, A. Gouble, V. Guyot, S. Leduc, R. Galetto, S. Grizot, D. Oficjalska, C. Perez, F. Delacôte, A. Dupuy, I. Chion-Sotinel, D. Le Clerre, C. Lebuhotel, O. Danos, F. Lemaire, K. Oussedik, F. Cédrone, J.-C. Epinat, J. Smith, R. J. Yáñez-Muñoz, G. Dickson, L. Popplewell, T. Koo,

T. VandenDriessche, M. K. Chuah, A. Duclert, P. Duchateau, and F. Pâques (2012, March). Chromosomal context and epigenetic mechanisms control the efficacy of genome editing by rare-cutting designer endonucleases. *Nucleic Acids Research 40*(13), 6367–6379. (Cited on page 71.)

[53] Dale, J., D. Brittain, A. Cataldi, D. Cousins, J. Crawford, J. Driscoll, H. Heersma, T. Lillebaek, T. Quitugua, N. Rastogi, R. Skuce, C. Sola, D. van Soolingen, and V. Vincent (2001). Spacer oligonucleotide typing of bacteria of the Mycobacterium tuberculosis complex: recommendations for standardised nomenclature [The Language of Our Science]. *The International Journal of Tuberculosis and Lung Disease 5*(3), 216–219. (Cited on page 19.)

[54] Daley, J. M., H. Niu, A. S. Miller, and P. Sung (2015). Biochemical mechanism of DSB end resection and its regulation. *DNA Repair* (0), –. (Cited on page 20.)

[55] Davis, K. M., V. Pattanayak, D. B. Thompson, J. A. Zuris, and D. R. Liu (2015, May). Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat Chem Biol 11*(5), 316–318. (Cited on page 74.)

[56] de Lange, O., T. Schreiber, N. Schandry, J. Radeck, K. H. Braun, J. Koszinowski, H. Heuer, A. Strauß, and T. Lahaye (2013, August). Breaking the DNA-binding code of Ralstonia solanacearum TAL effectors provides new possibilities to generate plant resistance genes against bacterial wilt disease. *New Phytologist 199*(3), 773–786. (Cited on page 6.)

[57] Dean, A. B., M. J. Stanger, J. T. Dansereau, P. Van Roey, V. Derbyshire, and M. Belfort (2002, April). Zinc finger as distance determinant in the flexible linker of intron endonuclease I-TevI. *PNAS 99*(13), 8554–8561. (Cited on page 79.)

[58] Delacôte, F., C. Perez, V. Guyot, M. Duhamel, C. Rochon, N. Ollivier, R. Macmaster, G. H. Silva, F. Pâques, F. Daboussi, and P. Duchateau (2013, January). High Frequency Targeted Mutagenesis Using Engineered Endonucleases and DNA-End Processing Enzymes. *PloS one 8*(1), e53217–. (Cited on pages 50 and 68.)

[59] Deng, D., C. Yan, X. Pan, M. Mahfouz, J. Wang, J.-K. Zhu, Y. Shi, and N. Yan (2012). Structural Basis for Sequence-Specific Recognition of DNA by TAL Effectors. *Science 335*(6069), 720–723. (Cited on page 7.)

[60] Deriano, L. and D. B. Roth (2013). Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. *Annual review of genetics 47*, 433–55. (Cited on page 21.)

[61] Dotti, G., S. Gottschalk, B. Savoldo, and M. K. Brenner (2014, January). Design and Development of Therapies using Chimeric Antigen Receptor-Expressing T cells. *Immunological reviews 257*(1), 10.1111/imr.12131–. (Cited on page 16.)

[62] Edgell, D. R., V. Derbyshire, P. V. Roey, S. LaBonne, M. J. Stanger, Z. Li, T. M. Boyd, D. A. Shub, and M. Belfort (2004, October). Intron-encoded homing endonuclease I-TevI also functions as a transcriptional autorepressor. *Nat Struct Mol Biol 11*(10), 936–944. (Cited on page 50.)

[63] Eisenschmidt, K., T. Lanio, A. Simoncsits, A. Jeltsch, V. Pingoud, W. Wende, and A. Pingoud (2005, November). Developing a programmed restriction endonuclease for highly specific DNA cleavage. *Nucleic Acids Research 33*(22), 7039–7047. (Cited on page 19.)

[64] Epinat, J.-C., S. Arnould, P. Chames, P. Rochaix, D. Desfontaines, C. Puzin, A. Patin, A. Zanghellini, F. Pâques, and E. Lacroix (2003, March). A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Research 31*(11), 2952–2962. (Cited on pages 15 and 81.)

[65] Evans, A. (2014). Glowing plants: Natural lighting with no electricity. *www.glowingplant.com*. (Cited on page 16.)

[66] Farzadfard, F., S. D. Perli, and T. K. Lu (2013, July). Tunable and Multifunctional Eukaryotic Transcription Factors Based on CRISPR/Cas. *ACS Synthetic Biology 2*(10), 604–613. (Cited on page 19.)

[67] Finer, J., P. Vain, M. Jones, and M. McMullen (1992). Development of the particle inflow gun for DNA delivery to plant cells. *11*(7), 323–328–. (Cited on pages 3 and 19.)

[68] Fink, D. J., N. A. DeLuca, W. F. Goins, and J. C. Glorioso (1996, March). Gene transfer to neurons using herpes simplex virus-based vectors. *Annu. Rev. Neurosci. 19*(1), 265–287. (Cited on page 20.)

[69] Finotti, A., L. Breda, C. W. Lederer, N. Bianchi, C. Zuccato, M. Kleanthous, S. Rivella, and R. Gambari (2015, February). Recent trends in the gene therapy of $\beta$-thalassemia. *Journal of Blood Medicine 6*, 69–85. (Cited on page 16.)

[70] Fitzpatrick, J. L., H. Akbarashandiz, D. Sakhrani, C. A. Biagi, T. E. Pitcher, and R. H. Devlin (2011, February). Cultured growth hormone transgenic salmon are reproductively out-competed by wild-reared salmon in semi-natural mating arenas. *Aquaculture 312*(1-4), 185–191. (Cited on page 16.)

[71] Fonfara, I., A. Le Rhun, K. Chylinski, K. S. Makarova, A.-L. Lécrivain, J. Bzdrenga, E. V. Koonin, and E. Charpentier (2013, October). Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Research 42*(4), 2577–2590. (Cited on page 4.)

[72] Friedhoff, P., B. Kolmes, O. Gimadutdinow, W. Wende, K. L. Krause, and A. Pingoud (1996, July). Analysis of the mechanism of the Serratia nuclease using site-directed mutagenesis. *Nucleic Acids Research 24*(14), 2632–2639. (Cited on page 13.)

[73] Frit, P., N. Barboule, Y. Yuan, D. Gomez, and P. Calsou (2014, May). Alternative end-joining pathway(s): Bricolage at DNA breaks. *DNA Repair 17*(0), 81–97. (Cited on page 23.)

[74] Frock, R. L., J. Hu, R. M. Meyers, Y.-J. Ho, E. Kii, and F. W. Alt (2015, February). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nature Biotechnology 33*(2), 179–186. (Cited on page 4.)

[75] Fu, Y., J. A. Foden, C. Khayter, M. L. Maeder, D. Reyon, J. K. Joung, and J. D. Sander (2013, June). High frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology 31*(9), 822–826. (Cited on page 19.)

[76] Gabriel, R., R. Eckenberg, A. Paruzynski, C. C. Bartholomae, A. Nowrouzi, A. Arens, S. J. Howe, A. Recchia, C. Cattoglio, W. Wang, K. Faber, K. Schwarzwaelder, R. Kirsten, A. Deichmann, C. R. Ball, K. S. Balaggan, R. J. Yáñez Muñoz, R. R. Ali, H. B. Gaspar, L. Biasco, A. Aiuti, D. Cesana, E. Montini, L. Naldini, O. Cohen-Haguenauer, F. Mavilio, A. J. Thrasher, H. Glimm, C. von Kalle, W. Saurin, and M. Schmidt (2009). Comprehensive genomic access to vector integration in clinical gene therapy. *Nature Medicine 15*(12), 1431–1436. (Cited on pages 3 and 19.)

[77] Gabsalilow, L., B. Schierling, P. Friedhoff, A. Pingoud, and W. Wende (2013, January). Site- and strand-specific nicking of DNA by fusion proteins derived from MutH and I-SceI or TALE repeats. *Nucleic Acids Research 41*(7), e83–e83. (Cited on pages 16 and 85.)

[78] Gaj, T., C. a. Gersbach, and C. F. Barbas (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology 31*(7), 397–405. (Cited on page 17.)

[79] Gao, H., X. Wu, J. Chai, and Z. Han (2012). Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell research 22*(12), 1716–20. (Cited on page 7.)

[80] Garg, A. and J. Lohmueller (2012, August). Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Research 40*(15), 7584–7595. (Cited on page 9.)

[81] Gasiunas, G., T. Sinkunas, and V. Siksnys (2014). Molecular mechanisms of CRISPR-mediated microbial immunity. *Cellular and Molecular Life Sciences 71*(3), 449–465. (Cited on page 18.)

[82] Gaudet, D., J. Methot, S. Dery, D. Brisson, C. Essiembre, G. Tremblay, K. Tremblay, J. de Wal, J. Twisk, N. van den Bulk, V. Sier-Ferreira, and S. van Deventer (2013, April). Efficacy and long-term safety of alipogene tiparvovec (AAV1-LPLS447X) gene therapy for lipoprotein lipase deficiency: an open-label trial. *Gene Ther 20*(4), 361–369. (Cited on page 16.)

[83] Gautier, A., A. Juillerat, C. Heinis, I. R. Corrêa Jr., M. Kindermann, F. Beaufils, and K. Johnsson (2008, February). An Engineered Protein Tag for Multiprotein Labeling in Living Cells. *Chemistry & Biology 15*(2), 128–136. (Cited on pages 42 and 77.)

[84] Ghosh, M., G. Meiss, and A. Pingoud (2007). The nuclease a-inhibitor complex is characterized by a novel metal ion bridge. *Journal of Biological Chemistry 282*(8), 5682–5690. (Cited on page 13.)

[85] Ghosh, M., G. Meiss, A. Pingoud, R. E. London, and L. C. Pedersen (2005). Structural Insights into the Mechanism of Nuclease A, a $\beta\beta\alpha$ Metal Nuclease from Anabaena. *Journal of Biological Chemistry 280*(30), 27990–27997. (Cited on page 13.)

[86] Gray, G. L., J. S. Baldridge, K. S. McKeown, H. L. Heyneker, and C. N. Chang (1985). Periplasmic production of correctly processed human growth hormone in Escherichia coli: natural and bacterial signal sequences are interchangeable. *Gene 39*(2-3), 247–254. (Cited on page 16.)

[87] Grieger, J. C. and R. J. Samulski (2005, April). Packaging Capacity of Adeno-Associated Virus Serotypes: Impact of Larger Genomes on Infectivity and Postentry Steps. *Journal of Virology 79*(15), 9933–9944. (Cited on pages 4 and 71.)

[88] Guo, J., T. Gaj, and C. F. Barbas (2010, May). Directed evolution of an enhanced and highly efficient FokI cleavage domain for Zinc Finger Nucleases. *Journal of Molecular Biology 400*(1), 96–107. (Cited on page 10.)

[89] Halford, S., L. Catto, C. Pernstich, D. Rusling, and K. Sanders (2011). The reaction mechanism of FokI excludes the possibility of targeting zinc finger nucleases to unique DNA sites. *Biochemical Society Transactions 39*(2), –. (Cited on pages 10 and 72.)

[90] Hetzert, B. (2014, September). Charakterisierung der DNA-Bindung von AvrBs3 und AvrBs3-SNAP-tag Fusionsproteinen mittels Fluoreszenz-basierter Methoden. Master's thesis, Universität Giessen. (Cited on page 42.)

[91] Hiroyuki, S. and K. Susumu (1981, December). New restriction endonucleases from Flavobacterium okeanokoites (FokI) and Micrococcus luteus (MluI). *Gene 16*(1-3), 73–78. (Cited on page 10.)

[92] Hsia, K.-C., K.-F. Chak, P.-H. Liang, Y.-S. Cheng, W.-Y. Ku, and H. S. Yuan (2004, February). DNA Binding and Degradation by the HNH Protein ColE7. *Structure 12*(2), 205–214. (Cited on page 13.)

[93] Hsu, P. D., E. S. Lander, and F. Zhang (2014, June). Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell 157*(6), 1262–1278. (Cited on page 18.)

[94] Iguchi, K., S. Usui, R. Ishida, and K. Hirano (2002). Imidazole-induced cell death, associated with intracellular acidification, caspase-3 activation, DFF-45 cleavage, but not oligonucleosomal DNA fragmentation. *7*(6), 519–525–. (Cited on page 78.)

[95] Imanishi, M., S. Negi, and Y. Sugiura (2010). Non-FokI-Based Zinc Finger Nucleases. In J. P. Mackay and D. J. Segal (Eds.), *Methods in Molecular Biology*, Volume 649, pp. 337–349–. Humana Press. (Cited on page 17.)

[96] Isalan, M., Y. Choo, and A. Klug (1997, March). Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *PNAS 94*(11), 5617–5621. (Cited on page 17.)

[97] Isalan, M., A. Klug, and Y. Choo (1998, September). Comprehensive DNA Recognition through Concerted Interactions from Adjacent Zinc Fingers. *Biochemistry 37*(35), 12026–12033. (Cited on page 17.)

[98] Jackson, D. A., R. H. Symons, and P. Berg (1972, October). Biochemical Method for Inserting New Genetic Information into DNA of Simian Virus 40: Circular SV40 DNA Molecules Containing Lambda Phage Genes and the Galactose Operon of Escherichia coli. *PNAS 69*(10), 2904–2909. (Cited on page 16.)

[99] Jacoby, K. and A. Scharenberg (2014). Homing Endonuclease Target Determination Using SELEX Adapted for Yeast Surface Display. In D. R. Edgell (Ed.), *Methods in Molecular Biology*, Volume 1123, pp. 165–190–. Humana Press. (Cited on pages 18 and 80.)

[100] Jaenisch, R. and B. Mintz (1974, April). Simian Virus 40 DNA Sequences in DNA of Healthy Adult Mice Derived from Preimplantation Blastocysts Injected with Viral DNA. *PNAS 71*(4), 1250–1254. (Cited on pages 3, 16, and 19.)

[101] Jasin, M. and R. Rothstein (2013). Repair of strand breaks by homologous recombination. *Cold Spring Harbor perspectives in biology 5*(11). (Cited on pages 20 and 23.)

[102] Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier (2012, August). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science 337*(6096), 816–821. (Cited on page 18.)

[103] Jinek, M., F. Jiang, D. W. Taylor, S. H. Sternberg, E. Kaya, E. Ma, C. Anders, M. Hauer, K. Zhou, S. Lin, M. Kaplan, A. T. Iavarone, E. Charpentier, E. Nogales, and J. a. Doudna (2014, March). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science 343*(6176), 1247997. (Cited on page 18.)

[104] Juillerat, A., C. Bertonati, G. Dubois, V. Guyot, S. Thomas, J. Valton, M. Beurdeley, G. H. Silva, F. Daboussi, and P. Duchateau (2014, January). BurrH: a new modular DNA binding protein for genome engineering. *Scientific Reports 4*, 3831. (Cited on page 6.)

[105] Juillerat, A., M. Beurdeley, J. Valton, S. Thomas, G. Dubois, M. Zaslavskiy, J. Mikolajczak, F. Bietz, G. H. Silva, A. Duclert, F. Daboussi, and P. Duchateau (2014, January). Exploring the transcription activator-like effectors scaffold versatility to expand the toolbox of designer nucleases. *BMC molecular biology 15*(1), 13. (Cited on pages 10 and 75.)

[106] Juillerat, A., G. Dubois, J. Valton, S. Thomas, S. Stella, A. Maréchal, S. Langevin, N. Benomari, C. Bertonati, G. H. Silva, F. Daboussi, J.-C. Epinat, G. Montoya, A. Duclert, and P. Duchateau (2014, February). Comprehensive analysis of the specificity of transcription activator-like effector nucleases. *Nucleic Acids Research 42*(8), 5390–5402. (Cited on pages 7 and 72.)

[107] Jurica, M. S., R. J. Monnat Jr., and B. L. Stoddard (1998, October). DNA Recognition and Cleavage by the LAGLIDADG Homing Endonuclease I-Cre I. *Molecular Cell 2*(4), 469–476. (Cited on page 15.)

[108] Katsuhiko, N., K. Yoshiyuki, H. Makoto, U. Takeshi, and B. Teruhiko (1982, October). Expression of cloned calf prochymosin gene sequence in Escherichia coli. *Gene 19*(3), 337–344. (Cited on page 16.)

[109] Katsumoto, Y., M. Fukuchi-Mizutani, Y. Fukui, F. Brugliera, T. A. Holton, M. Karan, N. Nakamura, K. Yonekura-Sakakibara, J. Togami, A. Pigeaire, G.-Q. Tao, N. S. Nehra, C.-Y. Lu, B. K. Dyson, S. Tsuda, T. Ashikari, T. Kusumi, J. G. Mason, and Y. Tanaka (2007, November). Engineering of the Rose Flavonoid Biosynthetic Pathway Successfully Generated Blue-Hued Flowers Accumulating Delphinidin. *Plant and Cell Physiology 48*(11), 1589–1600. (Cited on page 16.)

[110] Kay, M. A., J. C. Glorioso, and L. Naldini (2001, January). Viral vectors for gene therapy: the art of turning infectious agents into

vehicles of therapeutics. *Nature Medicine 7*(1), 33–40. (Cited on page 19.)

[111] Kay, S., S. Hahn, E. Marois, G. Hause, and U. Bonas (2007, October). A Bacterial Effector Acts as a Plant Transcription Factor and Induces a Cell Size Regulator. *Science 318*(5850), 648–651. (Cited on page 4.)

[112] Keeble, A. H., N. Kirkpatrick, S. Shimizu, and C. Kleanthous (2006, March). Calorimetric Dissection of Colicin DNase−Immunity Protein Complex Specificity. *Biochemistry 45*(10), 3243–3254. (Cited on page 13.)

[113] Kim, S. and J.-S. Kim (2010, December). Targeted genome engineering via zinc finger nucleases. *Plant Biotechnology Reports 5*(1), 9–17. (Cited on page 17.)

[114] Kim, Y. G., J. Cha, and S. Chandrasegaran (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *PNAS 93*(3), 1156–1160. (Cited on pages 10 and 17.)

[115] Kleinstiver, B. P., L. Wang, J. M. Wolfs, T. Kolaczyk, B. McDowell, X. Wang, C. Schild-Poulter, A. J. Bogdanove, and D. R. Edgell (2014, June). The I-TevI nuclease and linker domains contribute to the specificity of monomeric TALENs. *G3 (Bethesda, Md.) 4*(6), 1155–65. (Cited on pages 12 and 79.)

[116] Kleinstiver, B. P., J. M. Wolfs, and D. R. Edgell (2013, February). The monomeric GIY-YIG homing endonuclease I-BmoI uses a molecular anchor and a flexible tether to sequentially nick DNA. *Nucleic Acids Research 41*(10), 5413–5427. (Cited on pages 12 and 80.)

[117] Kleinstiver, B. P., J. M. Wolfs, T. Kolaczyk, A. K. Roberts, S. X. Hu, and D. R. Edgell (2012, May). Monomeric site-specific nucleases for genome editing. *PNAS 109*(21), 8061–8066. (Cited on pages 12 and 79.)

[118] Kowalski, J. C., M. Belfort, M. A. Stapleton, M. Holpert, J. T. Dansereau, S. Pietrokovski, S. M. Baxter, and V. Derbyshire (1999). Configuration of the catalytic GIY-YIG domain of intron endonuclease I-TevI: coincidence of computational and molecular findings. *Nucleic Acids Research 27*(10), 2115–2125. (Cited on page 12.)

[119] Krejci, L., V. Altmannova, M. Spirek, and X. Zhao (2012). Homologous recombination and its regulation. *Nucleic Acids Research 40*(13), 5795–5818. (Cited on pages 20, 21, and 23.)

[120] Kumar, V., F. W. Alt, and V. Oksenych (2014, February). Functional Overlaps Between XLF and The ATM-dependent DNA Double Strand Break Response. *DNA Repair 16*, 11–22. (Cited on pages 21 and 23.)

[121] Lacks, S., B. Greenberg, and M. Neuberger (1975, July). Identification of a deoxyribonuclease implicated in genetic transformation of Diplococcus pneumoniae. *Journal of Bacteriology 123*(1), 222–232. (Cited on page 13.)

[122] Lamb, B. M., A. C. Mercer, and C. F. Barbas (2013, July). Directed evolution of the TALE N-terminal domain for recognition of all 5′ bases. *Nucleic Acids Research 41*(21), 9779–9785. (Cited on page 7.)

[123] Laurens, N., D. A. Rusling, C. Pernstich, I. Brouwer, S. E. Halford, and G. J. L. Wuite (2012, February). DNA looping by foki: the impact of twisting and bending rigidity on protein-induced looping dynamics. *Nucleic Acids Research*, –. (Cited on pages 10 and 72.)

[124] Lebar, T., U. Bezeljak, A. Golob, M. Jerala, L. Kadunc, B. Pirš, M. Stražar, D. Vučko, U. Zupančič, M. Benčina, V. Forstnerič, R. Gaber, J. Lonzarić, A. Majerle, A. Oblak, A. Smole, and R. Jerala (2014, September). A bistable genetic switch based on designable DNA-binding domains. *Nature Communications 5*, 5007. (Cited on page 9.)

[125] Lee, J. Y., J. Chang, N. Joseph, R. Ghirlando, D. N. Rao, and W. Yang (2005, October). MutH Complexed with Hemi- and Unmethylated DNAs: Coupling Base Recognition and DNA Cleavage. *Molecular Cell 20*(1), 155–166. (Cited on page 16.)

[126] Lee, K. and S. E. Lee (2007, August). Saccharomyces cerevisiae Sae2- and Tel1-Dependent Single-Strand DNA Formation at DNA Break Promotes Microhomology-Mediated End Joining. *Genetics 176*(4), 2003–2014. (Cited on page 21.)

[127] Levin, K. B., O. Dym, S. Albeck, S. Magdassi, A. H. Keeble, C. Kleanthous, and D. S. Tawfik (2009, October). Following evolutionary paths to protein-protein interactions with high affinity and selectivity. *Nat Struct Mol Biol 16*(10), 1049–1055. (Cited on page 12.)

[128] Li, H., U. Y. Ulge, B. T. Hovde, L. a. Doyle, and R. J. Monnat (2012, March). Comprehensive homing endonuclease target site specificity profiling reveals evolutionary constraints and enables genome engineering applications. *Nucleic Acids Research 40*(6), 2587–98. (Cited on page 61.)

[129] Li, J.-F., J. Aach, J. E. Norville, M. McCormack, D. Zhang, J. Bush, G. M. Church, and J. Sheen (2013, August). Multiplex and homologous recombination-mediated plant genome editing via guide RNA/Cas9. *Nature Biotechnology 31*(8), 688–691. (Cited on page 19.)

[130] Li, K., J. Pang, H. Cheng, W.-P. Liu, J.-M. Di, H.-J. Xiao, Y. Luo, H. Zhang, W.-T. Huang, M.-K. Chen, L.-Y. Li, C.-K. Shao, Y.-H. Feng, and X. Gao (2015). Manipulation of prostate cancer metastasis by locus-specific modification of the CRMP4 promoter region using chimeric TALE DNA methyltransferase and demethylase. *Oncotarget 6*(12). (Cited on pages 9 and 72.)

[131] Li, X. and W.-D. Heyer (2008). Homologous recombination in DNA repair and DNA damage tolerance. *Cell research 18*(1), 99–113. (Cited on page 24.)

[132] Liang, P., Y. Xu, X. Zhang, C. Ding, R. Huang, Z. Zhang, J. Lv, X. Xie, Y. Chen, Y. Li, Y. Sun, Y. Bai, Z. Songyang, W. Ma, C. Zhou, and J. Huang (2015). CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *6*(5), 363–372–. (Cited on page 17.)

[133] Lienert, F., J. P. Torella, J.-H. Chen, M. Norsworthy, R. R. Richardson, and P. A. Silver (2013, July). Two- and three-input TALE-based AND logic computation in embryonic stem cells. *Nucleic Acids Research 41*(21), 9967–9975. (Cited on page 9.)

[134] Lippow, S. M., P. M. Aha, M. H. Parker, W. J. Blake, B. M. Baynes, and D. Lipovšek (2009, March). Creation of a type IIS restriction endonuclease with a long recognition sequence. *Nucleic Acids Research 37*(9), 3061–3073. (Cited on page 10.)

[135] Liu, J., T. Gaj, J. T. Patterson, S. J. Sirk, and C. F. Barbas (2014, January). Cell-penetrating peptide-mediated delivery of TALEN proteins via bioconjugation for genome engineering. *PloS one 9*(1), e85755. (Cited on page 4.)

[136] Lu, G. and E. N. Moriyama (2004, December). Vector NTI, a balanced all-in-one sequence analysis suite. *Briefings in Bioinformatics 5*(4), 378–388. (Cited on page 107.)

[137] Luo, Y., M. Rao, and J. Zou (2014, May). Generation of GFP reporter human induced pluripotent stem cells using AAVS1 safe harbor Transcription Activator-Like Effector Nuclease. *Current protocols in stem cell biology 29*, 5A.7.1–5A.7.18. (Cited on page 17.)

[138] Maeder, M. L., J. F. Angstman, M. E. Richardson, S. J. Linder, V. M. Cascio, S. Q. Tsai, Q. H. Ho, J. D. Sander, D. Reyon, B. E. Bernstein, J. F. Costello, M. F. Wilkinson, and J. K. Joung (2013, October). Targeted DNA Demethylation and Endogenous Gene Activation Using Programmable TALE-TET1 Fusions. *Nature Biotechnology 31*(12), 1137–1142. (Cited on pages 9 and 72.)

[139] Maeder, M. L., S. Thibodeau-Beganny, J. D. Sander, D. F. Voytas, and J. K. Joung (2009, September). Oligomerized Pool ENgineering (OPEN): An "Open-Source" Protocol for Making Customized Zinc Finger Arrays. *Nature protocols 4*(10), 1471–1501. (Cited on page 17.)

[140] Mak, A. N.-S., P. Bradley, R. A. Cernadas, A. J. Bogdanove, and B. L. Stoddard (2012, January). The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science 335*(6069), 716–719. (Cited on pages 6 and 8.)

[141] Marois, E., G. Van den Ackerveken, and U. Bonas (2002, July). The Xanthomonas Type III Effector Protein AvrBs3 Modulates Plant Gene Expression and Induces Cell Hypertrophy in the Susceptible Host. *MPMI 15*(7), 637–646. (Cited on page 6.)

[142] Meckler, J. F., M. S. Bhakta, M.-S. Kim, R. Ovadia, C. H. Habrian, A. Zykovich, A. Yu, S. H. Lockwood, R. Morbitzer, J. Elsäesser, T. Lahaye, D. J. Segal, and E. P. Baldwin (2013, April). Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Research 41*(7), 4118–28. (Cited on page 7.)

[143] Meiss, G., I. Franke, O. Gimadutdinow, C. Urbanke, and a. Pingoud (1998). Biochemical characterization of Anabaena sp. strain PCC 7120 non-specific nuclease NucA and its inhibitor NuiA. *European journal of biochemistry / FEBS 251*(3), 924–934. (Cited on page 13.)

[144] Meiss, G., O. Gimadutdinow, B. Haberland, and a. Pingoud (2000). Mechanism of DNA cleavage by the DNA/RNA-nonspecific Anabaena sp. PCC 7120 endonuclease NucA and its inhibition by NuiA. *Journal of Molecular Biology 297*(2), 521–534. (Cited on page 13.)

[145] Mendell, J. R., L. Rodino-Klapac, Z. Sahenk, V. Malik, B. K. Kaspar, C. M. Walker, and K. R. Clark (2012, May). Gene Therapy for Muscular Dystrophy: Lessons Learned and Path Forward. *Neuroscience letters 527*(2), 90–99. (Cited on page 16.)

[146] Mendenhall, E. M., K. E. Williamson, D. Reyon, J. Y. Zou, O. Ram, J. K. Joung, and B. E. Bernstein (2013, September). Locus-specific editing of histone modifications at endogenous enhancers using programmable TALE-LSD1 fusions. *Nature Biotechnology 31*(12), 1133–1136. (Cited on page 9.)

[147] Mercer, A. C., T. Gaj, R. P. Fuller, and C. F. Barbas (2012, August). Chimeric TALE recombinases with programmable DNA sequence specificity. *Nucleic Acids Research 40*(21), 11163–11172. (Cited on pages 9 and 86.)

[148] Midon, M., O. Gimadutdinow, G. Meiss, P. Friedhoff, and A. Pingoud (2012). Chemical Rescue of Active Site Mutants of S. pneumoniae Surface Endonuclease EndA and Other Nucleases of the HNH Family by Imidazole. *ChemBioChem 13*(5), 713–721. (Cited on pages 13, 40, and 76.)

[149] Midon, M., P. Schäfer, A. Pingoud, M. Ghosh, A. F. Moon, M. J. Cuneo, R. E. London, and G. Meiss (2011). Mutational and biochemical analysis of the DNA-entry nuclease EndA from Streptococcus pneumoniae. *Nucleic Acids Research 39*(2), 623–634. (Cited on pages 13 and 42.)

[150] Miede, A. (2015). *A Classic Thesis Style – An Homage to The Elements of Typographic Style*. https://bitbucket.org/amiede/classicthesis/. (Cited on page 107.)

[151] Miller, J. C., S. Tan, G. Qiao, K. A. Barlow, J. Wang, D. F. Xia, X. Meng, D. E. Paschon, E. Leung, S. J. Hinkley, G. P. Dulay, K. L. Hua, I. Ankoudinova, G. J. Cost, F. D. Urnov, H. S. Zhang, M. C. Holmes, L. Zhang, P. D. Gregory, and E. J. Rebar (2011, February). A TALE nuclease architecture for efficient genome editing. *Nature Biotechnology 29*(2), 143–148. (Cited on page 7.)

[152] Miyanari, Y., C. Ziegler-Birling, and M.-E. Torres-Padilla (2013, November). Live visualization of chromatin dynamics with fluorescent TALEs. *Nat Struct Mol Biol 20*(11), 1321–1324. (Cited on page 9.)

[153] Mock, U., K. Riecken, B. Berdien, W. Qasim, E. Chan, T. Cathomen, and B. Fehse (2014, September). Novel lentiviral vectors with mutated reverse transcriptase for mRNA delivery of TALE nucleases. *Scientific Reports 4*, 6409. (Cited on page 4.)

[154] Moon, A. F., M. Midon, G. Meiss, A. Pingoud, R. E. London, and L. C. Pedersen (2010, October). Structural insights into catalytic and substrate binding mechanisms of the strategic EndA nuclease from Streptococcus pneumoniae. *Nucleic Acids Research 39*(7), 2943–2953. (Cited on pages 11 and 13.)

[155] Morrow, J. F., S. N. Cohen, A. C. Y. Chang, H. W. Boyer, H. M. Goodman, and R. B. Helling (1974, May). Replication and Transcription of Eukaryotic DNA in Esherichia coli. *PNAS 71*(5), 1743–1747. (Cited on page 16.)

[156] Moscou, M. J. and A. J. Bogdanove (2009, December). A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science 326*(5959), 1501–1501. (Cited on pages 6 and 7.)

[157] Muir, W. M. (2004, July). The threats and benefits of GM fish. *EMBO Reports 5*(7), 654–659. (Cited on page 16.)

[158] Murakami, M. T., M. L. Sforça, J. L. Neves, J. H. Paiva, M. N. Domingues, A. L. A. Pereira, A. C. de Mattos Zeri, and C. E. Benedetti (2010, December). The repeat domain of the type III effector protein PthA shows a TPR-like structure and undergoes conformational changes upon DNA interaction. *Proteins 78*(16), 3386–3395. (Cited on page 7.)

[159] Muro-Pastor, A. M., E. Flores, A. Herrero, and C. P. Wolk (1992, October). Identification, genetic analysis and characterization of a sugar-non-specific nuclease from the cyanobacterium Anabaena sp. PCC 7120. *Molecular Microbiology 6*(20), 3021–3030. (Cited on page 13.)

[160] Németh, E., G. K. Schilli, G. Nagy, C. Hasenhindl, B. Gyurcsik, and C. Oostenbrink (2014, June). Design of a colicin E7 based chimeric zinc-finger nuclease. *Journal of Computer-Aided Molecular Design 28*(8), 841–850. (Cited on pages 13 and 75.)

[161] OConnell, M. R., B. L. Oakes, S. H. Sternberg, A. East-Seletsky, M. Kaplan, and J. A. Doudna (2014, September). Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature 516*(7530), 263–266. (Cited on page 19.)

[162] Orlowski, J., M. Boniecki, and J. M. Bujnicki (2007). I-Ssp6803I: the first homing endonuclease from the PD-(D/E)XK superfamily exhibits an unusual mode of DNA recognition. *Bioinformatics 23*(5), 527–530. (Cited on page 17.)

[163] Otten, L., H. De Greye, J. Hernalsteens, M. Van Montagu, O. Schieder, J. Straub, and J. Schell (1981). Mendelian transcmission of genes introduced into plants by the Ti plasmids of Agrobacterium tumefaciens. *183*(2), 209–213–. (Cited on pages 3 and 19.)

[164] Owens, J. B., D. Mauro, I. Stoytchev, M. S. Bhakta, M.-S. Kim, D. J. Segal, and S. Moisyadi (2013, July). Transcription activator like effector (TALE)-directed piggyBac transposition in human cells. *Nucleic Acids Research 41*(19), 9197–9207. (Cited on pages 9 and 86.)

[165] Padgette, S. R., K. H. Kolacz, X. Delannay, D. B. Re, B. J. LaVallee, C. N. Tinius, W. K. Rhodes, Y. I. Otero, G. F. Barry, D. A. Eichholtz, V. M. Peschke, D. L. Nida, N. B. Taylor, and G. M. Kishore (1995). Development, Identification, and Characterization of a Glyphosate-Tolerant Soybean Line. *35*, 1451–1461. (Cited on page 16.)

[166] Pattanayak, V., J. P. Guilinger, and D. R. Liu (2014). Determining the specificities of TALENs, Cas9, and other genome editing enzymes. *Methods in enzymology 546*, 47–78. (Cited on pages 17 and 25.)

[167] Pavletich, N. and C. Pabo (1991, May). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science 252*(5007), 809–817. (Cited on page 17.)

[168] Pingoud, A., M. Fuxreiter, V. Pingoud, and W. Wende (2005). Type II restriction endonucleases: structure and mechanism. *62*(6), 685–707–. (Cited on page 9.)

[169] Pingoud, A. and A. Jeltsch (2001). Structure and function of type II restriction endonucleases. *Nucleic Acids Research 29*(12), 2437–2447. (Cited on pages 10 and 15.)

[170] Pingoud, A. and G. H. Silva (2007, July). Precision genome surgery. *Nature Biotechnology 25*(7), 743–744. (Cited on page 10.)

[171] Prieto, J., P. Redondo, D. Padró, S. Arnould, J.-C. Epinat, F. Pâques, F. J. Blanco, and G. Montoya (2007, March). The C-terminal loop of the homing endonuclease I-CreI is essential for site recognition, DNA binding and cleavage. *Nucleic Acids Research 35*(10), 3262–3271. (Cited on pages 11 and 15.)

[172] Puchta, H., B. Dujon, and B. Hohn (1996, May). Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. *PNAS 93*(10), 5055–5060. (Cited on page 3.)

[173] Puyet, A., B. Greenberg, and S. A. Lacks (1990, June). Genetic and structural characterization of endA: A membrane-bound nuclease required for transformation of Streptococcus pneumoniae. *Journal of Molecular Biology 213*(4), 727–738. (Cited on page 13.)

[174] R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0. (Cited on page 107.)

[175] Rahman, S. H., M. L. Maeder, J. K. Joung, and T. Cathomen (2011, June). Zinc-Finger Nucleases for Somatic Gene Therapy: The Next Frontier. *Human Gene Therapy 22*(8), 925–933. (Cited on page 17.)

[176] Ramirez, C. L., J. E. Foley, D. A. Wright, F. Muller-Lerch, S. H. Rahman, T. I. Cornu, R. J. Winfrey, J. D. Sander, F. Fu, J. A. Townsend, T. Cathomen, D. F. Voytas, and J. K. Joung (2008, May). Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat Meth 5*(5), 374–375. (Cited on page 17.)

[177] Ran, F. A., P. D. Hsu, C.-Y. Lin, J. S. Gootenberg, S. Konermann, A. Trevino, D. A. Scott, A. Inoue, S. Matoba, Y. Zhang, and F. Zhang (2013, August). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell 154*(6), 1380–1389. (Cited on page 18.)

[178] Redondo, P., J. Prieto, I. G. Munoz, A. Alibés, F. Stricher, L. Serrano, J.-P. Cabaniols, F. Daboussi, S. Arnould, C. Perez, P. Duchateau, F. Pâques, F. J. Blanco, and G. Montoya (2008, November). Molecular basis of xeroderma pigmentosum group C DNA recognition by engineered meganucleases. *Nature 456*(7218), 107–11. (Cited on page 15.)

[179] Rice, P., I. Longden, and A. Bleasby (2000, June). EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics 16*(6), 276–277. (Cited on page 107.)

[180] Richard, G.-F. (2015). Shortening trinucleotide repeats using highly specific endonucleases: a possible approach to gene therapy? *Trends in Genetics*, 1–10. (Cited on page 23.)

[181] Richard, G.-F., D. Viterbo, V. Khanna, V. Mosbach, L. Castelain, and B. Dujon (2014, January). Highly specific contractions of a single CAG/CTG trinucleotide repeat by TALEN in yeast. *PloS one 9*(4), e95611. (Cited on page 23.)

[182] Richter, A., J. Streubel, C. Blücher, B. Szurek, M. Reschke, J. Grau, and J. Boch (2014, January). A TAL effector repeat architecture for frameshift binding. *Nature Communications 5*, 3447. (Cited on page 6.)

[183] Rivat, C., G. Santilli, H. B. Gaspar, and A. J. Thrasher (2012, June). Gene Therapy for Primary Immunodeficiencies. *Human Gene Therapy 23*(7), 668–675. (Cited on page 16.)

[184] Rosen, L. E., H. A. Morrison, S. Masri, M. J. Brown, B. Springstubb, D. Sussman, B. L. Stoddard, and L. M. Seligman (2006, August). Homing endonuclease I-CreI derivatives with novel DNA target specificities. *Nucleic Acids Research 34*(17), 4791–4800. (Cited on page 15.)

[185] Ryan, R. P., F.-J. Vorhölter, N. Potnis, J. B. Jones, M.-A. Van Sluys, A. J. Bogdanove, and J. M. Dow (2011, May). Pathogenomics of Xanthomonas: understanding bacterium-plant interactions. *Nat Rev Micro 9*(5), 344–355. (Cited on page 4.)

[186] Sapranauskas, R., G. Sasnauskas, A. Lagunavicius, G. Vilkaitis, A. Lubys, and V. Siksnys (2000, October). Novel Subtype of Type IIs Restriction Enzymes: BfiI enodnuclease exhibits similarities to the EDTA-resistant nuclease Nuc of Salmonella typhimurium. *Journal of Biological Chemistry 275*(40), 30878–30885. (Cited on page 15.)

[187] Sasnauskas, G., L. Zakrys, M. Zaremba, R. Cosstick, J. W. Gaynor, S. E. Halford, and V. Siksnys (2010, January). A novel mechanism for the scission of double-stranded DNA: BfiI cuts both 3′-5′ and 5′-3′ strands by rotating a single active site. *Nucleic Acids Research*, –. (Cited on page 15.)

[188] Schierling, B., A.-J. Noël, W. Wende, L. T. Hien, E. Volkov, E. Kubareva, T. Oretskaya, M. Kokkinidis, A. Römpp, B. Spengler, and A. Pingoud (2009, December). Controlling the enzymatic activity of a restriction enzyme by light. *PNAS 107*(4), 1361–1366. (Cited on page 73.)

[189] Schneider, C. A., W. S. Rasband, and K. W. Eliceiri (2012, July). NIH Image to ImageJ: 25 years of image analysis. *Nat Meth 9*(7), 671–675. (Cited on page 107.)

[190] Schrödinger, LLC (2015, August). The PyMOL Molecular Graphics System, Version 1.7.0. (Cited on page 107.)

[191] Sebesta, M., P. Burkovics, S. Juhasz, S. Zhang, J. E. Szabo, M. Y. Lee, L. Haracska, and L. Krejci (2013, September). Role of PCNA and TLS polymerases in D-loop extension during homologous recombination in humans. *DNA Repair 12*(9), 691–698. (Cited on page 23.)

[192] Seligman, L. M., K. M. Chisholm, B. S. Chevalier, M. S. Chadsey, S. T. Edwards, J. H. Savage, and A. L. Veillet (2002). Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic Acids Research 30*(17), 3870–3879. (Cited on pages 15 and 81.)

[193] Sharma, A. and C. T. Scott (2015, June). The ethics of publishing human germline research. *Nature Biotechnology 33*(6), 590–592. (Cited on page 17.)

[194] Shi, Z., K. F. Chak, and H. S. Yuan (2005). Identification of an essential cleavage site in ColE7 required for import and killing of cells. *Journal of Biological Chemistry 280*(26), 24663–24668. (Cited on pages 12 and 75.)

[195] Silva, G., L. Poirot, R. Galetto, J. Smith, G. Montoya, P. Duchateau, and F. Pâques (2010, December). Meganucleases and Other Tools for Targeted Genome Engineering: Perspectives and Challenges for Gene Therapy. *Current Gene Therapy 11*(1), 11–27. (Cited on page 10.)

[196] Silva, G. H., M. Belfort, W. Wende, and A. Pingoud (2006, August). From Monomeric to Homodimeric Endonucleases and Back: Engineering Novel Specificity of LAGLIDADG Enzymes. *Journal of Molecular Biology 361*(4), 744–754. (Cited on page 15.)

[197] Simpson, I. N. and C. E. Caten (1979). Recurrent Mutation and Selection for Increased Penicillin Titre in Aspergillus nidulans. *Microbiology 113*(2), 209–217. (Cited on page 3.)

[198] Sinclair, S. A. and U. Krämer (2012). The zinc homeostasis network of land plants. *Biochimica et Biophysica Acta - Molecular Cell Research 1823*(9), 1553–1567. (Cited on page 75.)

[199] Siuti, P., J. Yazbek, and T. K. Lu (2013, May). Synthetic circuits integrating logic and memory in living cells. *Nature Biotechnology 31*(5), 448–452. (Cited on page 9.)

[200] Siuti, P., J. Yazbek, and T. K. Lu (2014, June). Engineering genetic circuits that compute and remember. *Nat. Protocols 9*(6), 1292–1300. (Cited on page 9.)

[201] Smih, F., P. Rouet, P. J. Romanienko, and M. Jasin (1995, December). Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Research 23*(24), 5012–5019. (Cited on page 3.)

[202] Smith, J., S. Grizot, S. Arnould, A. Duclert, J.-C. Epinat, P. Chames, J. Prieto, P. Redondo, F. J. Blanco, J. Bravo, G. Montoya, F. Pâques, and P. Duchateau (2006). A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Research 34*(22), e149. (Cited on page 15.)

[203] Stella, S., R. Molina, B. López-Méndez, A. Juillerat, C. Bertonati, F. Daboussi, R. Campos-Olivas, P. Duchateau, and G. Montoya (2014, July). BuD, a helix-loop-helix DNA-binding domain for genome modification. *Acta crystallographica. Section D, Biological crystallography 70*(Pt 7), 2042–52. (Cited on page 6.)

[204] Stella, S., R. Molina, I. Yefimenko, J. Prieto, G. Silva, C. Bertonati, A. Juillerat, P. Duchateau, and G. Montoya (2013, June). Structure of the AvrBs3-DNA complex provides new insights into the initial thymine-recognition mechanism. *Acta Crystallographica Section D: Biological Crystallography 69*(Pt 9), 1707–1716. (Cited on page 6.)

[205] Stephens, P. J., C. D. Greenman, B. Fu, F. Yang, G. R. Bignell, L. J. Mudie, E. D. Pleasance, K. W. Lau, D. Beare, L. A. Stebbings, S. McLaren, M.-L. Lin, D. J. McBride, I. Varela, S. Nik-Zainal, C. Leroy, M. Jia, A. Menzies, A. P. Butler, J. W. Teague, M. A. Quail, J. Burton, H. Swerdlow, N. P. Carter, L. A. Morsberger, C. Iacobuzio-Donahue, G. A. Follows, A. R. Green, A. M. Flanagan, M. R. Stratton, P. A. Futreal, and P. J. Campbell (2010, November). Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. *Cell 144*(1), 27–40. (Cited on page 20.)

[206] Stoddard, B. L. (2005). Homing endonuclease structure and function. *Quarterly Reviews of Biophysics 38*(1), 49–95. (Cited on pages 10, 12, and 17.)

[207] Sui, M., L. Tsai, K. Hsia, and L. Doudeva (2002, December). phosphate binding in the H-N-H motif: crystal structures of the nuclease domain of ColE7/Im7 in complex with a phosphate ion and different divalent metal. *Protein Science 11*(12), 2947–2957. (Cited on page 11.)

[208] Sui, M.-J., L.-C. Tsai, K.-C. Hsia, L. G. Doudeva, W.-Y. Ku, G. W. Han, and H. S. Yuan (2002, September). Metal ions and phosphate

binding in the H-N-H motif: Crystal structures of the nuclease domain of ColE7/Im7 in complex with a phosphate ion and different divalent metal ions. *Protein Science 11*(12), 2947–2957. (Cited on page 73.)

[209] Sun, N. and H. Zhao (2014). A single-chain TALEN architecture for genome engineering. *Molecular BioSystems 10*(3), 446–53. (Cited on page 16.)

[210] Sussman, M. R., R. M. Amasino, J. C. Young, P. J. Krysan, and S. Austin-Phillips (2000, December). The Arabidopsis Knockout Facility at the University of Wisconsin-Madison. *Plant Physiology 124*(4), 1465–1467. (Cited on page 16.)

[211] Szurek, B., O. Rossier, G. Hause, and U. Bonas (2002, October). Type III-dependent translocation of the Xanthomonas AvrBs3 protein into the plant cell. *Molecular Microbiology 46*(1), 13–23. (Cited on page 7.)

[212] Tanaka, Y., F. Brugliera, and S. Chandler (2009, December). Recent Progress of Flower Colour Modification by Biotechnology. *International Journal of Molecular Sciences 10*(12), 5350–5369. (Cited on page 16.)

[213] Tebas, P., D. Stein, W. W. Tang, I. Frank, S. Q. Wang, G. Lee, S. K. Spratt, R. T. Surosky, M. A. Giedlin, G. Nichol, M. C. Holmes, P. D. Gregory, D. G. Ando, M. Kalos, R. G. Collman, G. Binder-Scholl, G. Plesa, W.-T. Hwang, B. L. Levine, and C. H. June (2014, March). Gene Editing of CCR5 in Autologous CD4 T Cells of Persons Infected with HIV. *The New England journal of medicine 370*(10), 901–910. (Cited on page 3.)

[214] The GIMP Team (2015). *GIMP - the GNU Image Manipulation Program*. www.gimp.org. (Cited on page 107.)

[215] The Inkscape Team (2015). *Inkscape - A powerful, free design tool*. www.inkscape.org. (Cited on page 107.)

[216] The LyX Team (2015). *LyX - The Document Processor*. www.lyx.org. (Cited on page 107.)

[217] Thomas, C. E., A. Ehrhardt, and M. A. Kay (2003, May). Progress and problems with the use of viral vectors for gene therapy. *Nat Rev Genet 4*(5), 346–358. (Cited on page 19.)

[218] Thompson, A. J., X. Yuan, W. Kudlicki, and D. L. Herrin (1992, October). Cleavage and recognition pattern of a double-strand-specific endonuclease (I-CreI) encoded by the chloroplast 23S rRNA intron of Chlamydomonas reinhardtii. *Gene 119*(2), 247–251. (Cited on page 15.)

[219] Thyme, S. B., Y. Song, T. J. Brunette, M. D. Szeto, L. Kusak, P. Bradley, and D. Baker (2014, April). Massively parallel determination and modeling of endonuclease substrate specificity. *Nucleic Acids Research 42*(22), 13839–13852. (Cited on pages 18 and 80.)

[220] Traver, B. E., M. A. E. Anderson, and Z. N. Adelman (2009, October). Homing endonucleases catalyze double-stranded DNA breaks and somatic transgene excision in Aedes aegypti. *Insect molecular biology 18*(5), 623–633. (Cited on pages 16 and 18.)

[221] Tsai, S. Q., Z. Zheng, N. T. Nguyen, M. Liebers, V. V. Topkar, V. Thapar, N. Wyvekens, C. Khayter, A. J. Iafrate, L. P. Le, M. J. Aryee, and J. K. Joung (2015, February). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature Biotechnology 33*(2), 187–197. (Cited on page 4.)

[222] Tzfira, T. and V. Citovsky (2006, April). Agrobacterium-mediated genetic transformation of plants: biology and biotechnology. *Current Opinion in Biotechnology 17*(2), 147–154. (Cited on page 19.)

[223] Vaeck, M., A. Reynaerts, H. Hofte, S. Jansens, M. De Beuckeleer, C. Dean, M. Zabeau, M. V. Montagu, and J. Leemans (1987, July). Transgenic plants protected from insect attack. *Nature 328*(6125), 33–37. (Cited on page 16.)

[224] Valton, J., A. Dupuy, F. Daboussi, S. Thomas, A. Maréchal, R. Macmaster, K. Melliand, A. Juillerat, and P. Duchateau (2012, September). Overcoming Transcription Activator-like Effector (TALE) DNA Binding Domain Sensitivity to Cytosine Methylation. *The Journal of Biological Chemistry 287*(46), 38427–38432. (Cited on page 7.)

[225] Van Roey, P., L. Meehan, J. C. Kowalski, M. Belfort, and V. Derbyshire (2002, November). Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI. *Nat Struct Mol Biol 9*(11), 806–811. (Cited on pages 11 and 12.)

[226] Vergnaud, G., D. Zhou, M. Platonov, C. Pourcel, R. Yang, A. Anisimov, H. Neubauer, S. Balakhonov, A. Rakin, S. Dentovskaya, I. Grissa, Y. Song, Y. Cui, O. Gorgé, and Y. Li (2007). Analysis of the Three Yersinia pestis CRISPR Loci Provides New Tools for Phylogenetic Studies and Possibly for the Investigation of Ancient DNA. In R. Perry and J. Fetherston (Eds.), *Advances In Experimental Medicine And Biology*, Volume 603, pp. 327–338–. Springer New York. (Cited on page 19.)

[227] Wah, D. A., J. Bitinaite, I. Schildkraut, and A. K. Aggarwal (1998, April). Structure of FokI has implications for DNA cleavage. *PNAS 95*(18), 10564–10569. (Cited on pages 10 and 11.)

[228] Wah, D. A., J. A. Hirsch, L. F. Dorner, I. Schildkraut, and A. K. Aggarwal (1997, July). Structure of the multimodular endonuclease FokI bound to DNA. *Nature 388*(6637), 97–100. (Cited on page 10.)

[229] Wan, H., J.-p. Hu, K.-s. Li, X.-h. Tian, and S. Chang (2013, January). Molecular dynamics simulations of DNA-free and DNA-bound TAL effectors. *PloS one 8*(10), e76045. (Cited on page 7.)

[230] Wang, J., H. H. Kim, X. Yuan, and D. L. Herrin (1997, October). Purification, biochemical characterization and protein-DNA interactions of the I-CreI endonuclease produced in Escherichia coli. *Nucleic Acids Research 25*(19), 3767–3776. (Cited on page 81.)

[231] Wang, W., C. Ye, J. Liu, D. Zhang, J. T. Kimata, and P. Zhou (2014, November). CCR5 Gene Disruption via Lentiviral Vectors Expressing Cas9 and Single Guided RNA Renders Cells Resistant to HIV-1 Infection. *PloS one 9*(12), e115987–. (Cited on page 16.)

[232] Wang, X., Y. Wang, X. Wu, J. Wang, Y. Wang, Z. Qiu, T. Chang, H. Huang, R.-j. Lin, and J.-k. Yee (2015). Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *33*(2). (Cited on pages 4, 19, and 79.)

[233] Wang, Y., W. Yang, and C. Li (2007, January). Structural basis for sequence-dependent DNA cleavage by nonspecific endonucleases. *Nucleic Acids Research 35*(2), 584–594. (Cited on pages 10 and 12.)

[234] Watson, J. D. and F. H. C. Crick (1953, April). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature 171*(4356), 737–738. (Cited on page 3.)

[235] Wei, W., H. Xin, B. Roy, J. Dai, Y. Miao, and G. Gao (2014, June). Heritable Genome Editing with CRISPR/Cas9 in the Silkworm, Bombyx mori. *PloS one 9*(7), e101210–. (Cited on page 19.)

[236] Wicky, B. I. M., M. Stenta, and M. Dal Peraro (2013, January). TAL effectors specificity stems from negative discrimination. *PloS one 8*(11), e80261. (Cited on page 9.)

[237] Windbichler, N., P. A. Papathanos, F. Catteruccia, H. Ranson, A. Burt, and A. Crisanti (2007, July). Homing endonuclease mediated gene targeting in Anopheles gambiae cells and embryos. *Nucleic Acids Research 35*(17), 5922–5933. (Cited on page 18.)

[238] Wolfs, J. M., M. DaSilva, S. E. Meister, X. Wang, C. Schild-Poulter, and D. R. Edgell (2014, February). MegaTevs: single-chain dual nucleases for efficient gene disruption. *Nucleic Acids Research 42*(13), 8816–8829. (Cited on pages 12, 79, and 86.)

[239] Wu, W., D. W. Wood, G. Belfort, V. Derbyshire, and M. Belfort (2002, September). Intein-mediated purification of cytotoxic endonuclease I-TevI by insertional inactivation and pH-controllable splicing. *Nucleic Acids Research 30*(22), 4864–4871. (Cited on page 79.)

[240] Yang, W. (2011a). *Nucleases: diversity of structure, function and mechanism.*, Volume 44. (Cited on page 9.)

[241] Yang, W. (2011b). Nucleases: diversity of structure, function and mechanism. *Quarterly Reviews of Biophysics 44*(01), 1–93. (Cited on page 10.)

[242] Yanik, M. (2013). *Generierung hochspezifischer Nukleasen für die zielgerichtete Spaltung genomischer DNA-Sequenzen*. Ph. D. thesis, Justus Liebig Universität Giessen. (Cited on page 15.)

[243] Yanik, M., J. Alzubi, T. Lahaye, T. Cathomen, A. Pingoud, and W. Wende (2013, November). TALE-PvuII Fusion Proteins - Novel Tools for Gene Targeting. *PloS one 8*(12), e82539–. (Cited on pages 15 and 85.)

[244] Zaremba, M. and V. Siksnys (2010, January). Molecular scissors under light control. *PNAS 107*(4), 1259–1260. (Cited on page 74.)

[245] Zhang, F., L. Cong, S. Lodato, S. Kosuri, G. Church, and P. Arlotta (2011, January). Programmable Sequence-Specific Transcriptional Regulation of Mammalian Genome Using Designer TAL Effectors. *Nature Biotechnology 29*(2), 149–153. (Cited on page 9.)

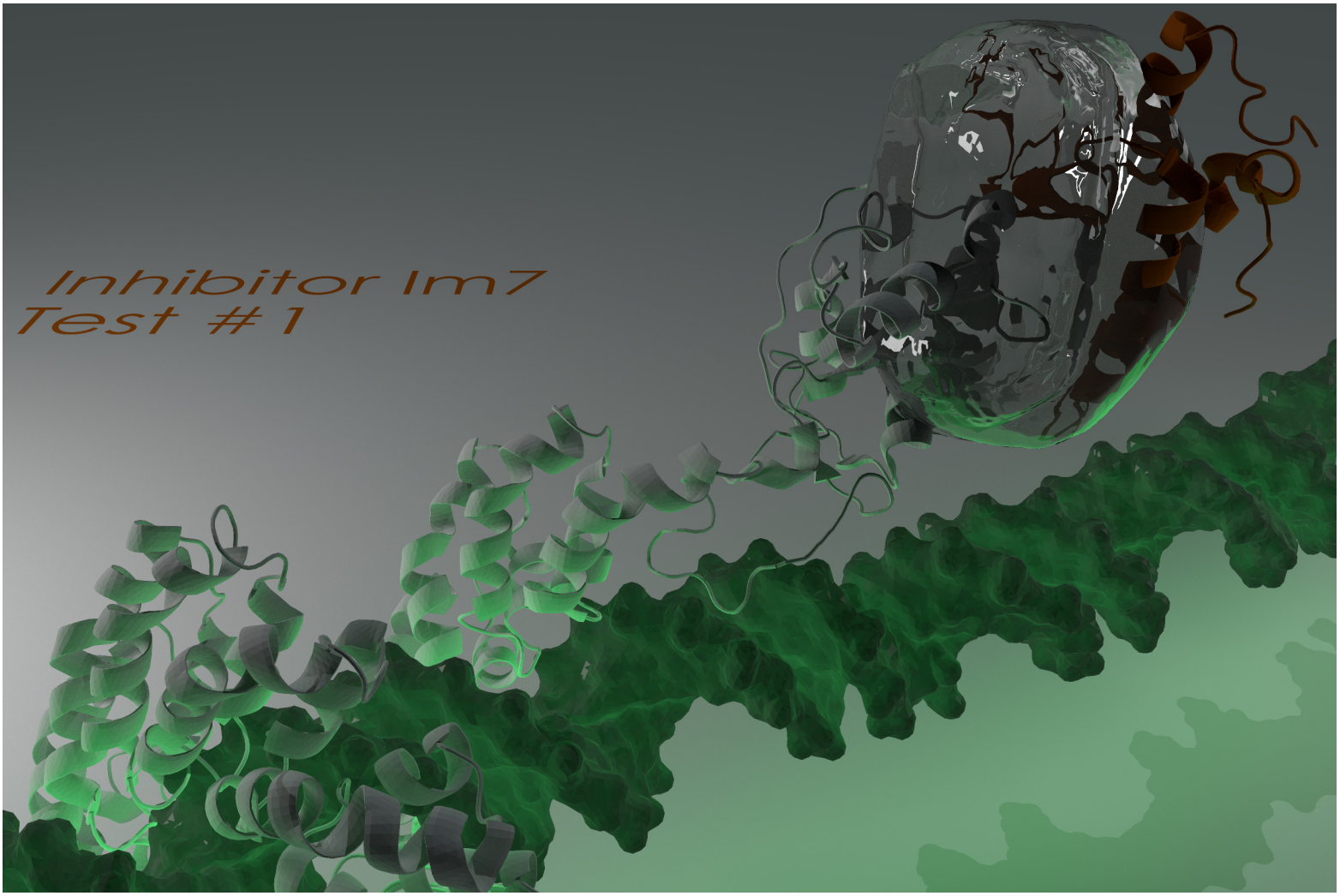
[246] Zhang, Z., E. Wu, Z. Qian, and W.-S. Wu (2014, November). A multicolor panel of TALE-KRAB based transcriptional repressor vectors enabling knockdown of multiple gene targets. *Scientific Reports 4*, 7338–. (Cited on page 19.)

[247] Zhao, L., R. P. Bonocora, D. A. Shub, and B. L. Stoddard (2007, March). The restriction fold turns to the dark side: a bacterial homing endonuclease with a PD-(D/E)-XK motif. *The EMBO Journal 26*(9), 2432–2442. (Cited on page 17.)

[248] Zhao, Y., Z. Dai, Y. Liang, M. Yin, K. Ma, M. He, H. Ouyang, and C.-B. Teng (2014, January). Sequence-specific inhibition of microRNA via CRISPR/CRISPRi system. *Scientific Reports 4*, 3943–. (Cited on page 19.)

[249] Zheng, C.-K., C.-L. Wang, X.-P. Zhang, F.-J. Wang, T.-F. Qin, and K.-J. Zhao (2014, February). The last half-repeat of transcription activator-like effector (TALE) is dispensable and thereby TALE-based technology can be simplified. *Molecular plant pathology*, 1–29. (Cited on page 7.)

[250] Zhu, X., H. Gong, Q. He, Z. Zeng, J. S. Busse, W. Jin, P. C. Bethke, and J. Jiang (2015, June). Silencing of vacuolar invertase and asparagine synthetase genes and its impact on acrylamide formation of fried potato products. *Plant Biotechnology Journal*, n/a–n/a. (Cited on page 16.)




*"Myself," said the drone sniffily, "I have never been able to see what virtue there could be in something that was eighty percent water."*

— Iain M. Banks

## ACKNOWLEDGMENTS

I would like to thank Professor Dr. Peter Friedhoff for becoming my first referees and for his insights. I really appreciated the remarks and comments on my work, especially the ones I did not like to hear at first.

I would also like to thank Professor Dr. Michael Martin for becoming my second referee.

I would like to thank Professor Dr. Alfred Pingoud for accepting me in the IRTG and for his thoughts on my topic. Learning the value of scientific discussion and of asking the right questions was absolutely invaluable. His view that every talk is an opportunity to share knowledge and to learn instead of a chore, benefited all the scientists he trained.

My thanks also go to Dr. Elena Evguenieva-Hackenberg and to Professor Dr. Virginijus Siksnys for agreeing to become my local and international IRTG advisor.

More thanks go to Dr. George Silva, for his ideas and his supervision. I learned a lot and I miss the endless discussions about specificity, accuracy, activity, efficiency but also about music, various TV shows and life in general.

Thanks also goes to Dr. Wolfgang Wende for his support, his supervision, his patience and his abilities regarding technical equipment (which are bordering on witchcraft). His comments made my talks and documents cleaner and easier to understand.

A "merci beaucoup" goes to Dr. Marine Beurdeley, who was always very helpful and never showed any signs of annoyance. Her friendliness and positive attitude made working with her really enjoyable.

Another thank you goes to Dr. Marika Midon for providing plasmids, for her experience with EndA and for her encouraging comments.

I would like to thank Professor Dr. Katja Sträßer for her help. I hope she will enjoy her time at the institute as much as I have enjoyed mine.

My thanks goes to all the other people at the *Institute for Biochemistry* and at CELLECTIS. I really had a great time with you all:

- to Heike for her saintlike patience and her omniscience regarding the location of every object in the institute.
- to Sabrina and Julia for spending lunches with me and for discussing which companion animals are objectively the best: dogs (Sabrina), cats (Julia) or plants (Fabian).
- to Laura for sharing an office with me and for her general cheerfulness. It is a good thing that there are still people who (literally and figuratively) don't know the meaning of the word "Äbsch".
- to Lili for inventing syncytium muffins. This cutting-edge technology, combining bionics, inventors spirit and completely overfilling the muffin cups, will surely cause a paradigm shift in the field of cupcakes.
- to Pingping for being a really great colleague. Thanks for teaching me how to make enough dumplings for several hundred people.
- a special thanks goes to Andi, Dennis, Laki and Mert. I miss spending nights at the institute, playing Starcraft2 and eating pizza.

- to Anja for organizing the IRTG (Saint Petersburg was great!). Thankfully, her expertise also includes site-specific integration of commas, which is far superior to my pseudo-random method. At least one sane person in the institute, who knows of the dangers of extreme temperatures (>25°C) and sunlight (see chapter 1.3).
- to Ina and Karina for helping me with every possible document that crossed my path.
- to Petra who stops the whole institute from drifting into chaos.
- to all the students I supervised during their theses or internships: to Yvonne, Kerstin, Bianca, Chad and Kevin. Thanks for your work on the project. Supervising you helped me better understand the background of my work.
- to all the new ones in the lab (Max, Rashmi, Birte, Christoph) – Don't burn down the lab after I am gone! – and to all the others that spent some time at the institute (Maya, Sascha, Houda, Franziska, Friedrich, Mona, Kimia, Dani, Alex, Julian, Manuel, Famou, Lea, Oxana, Miguel and Matthias)
- to Alex, Claudia, Gwen, Julien, Séverine, Kevin, Alan, Coline and Rachel. You were fantastic colleagues and the evenings spent at the *Frog* were truly fun. I hope to visit you soon.
- another special thanks goes to the people from the platform for doing the yeast transformations, especially to Anne-Sophie for accepting every late *fiche de transfert*.
- to the people of the IRTG, even to those whose plan to survive the end of civilization is complete rubbish. Stealing an honest person's potatoes is *disreputable* and *honorless*, Kristina! Furthermore it will not work, because I will have TONS of barbwire by then.

I would also like to thank:

- Franzi, Freddy, Norbert, Richard and Tobi for welcoming me back in Germany as if I had never been gone. (In the nice way, not the "Oh-we-did-not-realize-you-were gone"-way). It is great knowing you.
- all the people from my chess club SJ Herborn 1998 and all the other players, who I know for most of my life. I would not want to trade the time and discussions with you: "So, imagine you have a cube with n dimensions. Now imagine there is a person standing on every corner of the cube. Each person is wearing a differently colored T-shirt. Then you calculate...".

- the people in my DSA group. For similar reasons as above: "Do you think we could transport an unconscious person, that is stable (i. e., 10 < LE < 15), over 3 meters of height with a 10 meter rope, without causing severe (i. e., SP < $\frac{Konstitution}{2}$) injury?"

- the people from TEAM HAGE MULTIGAMING (Est. 2006, lessons 50€/h). I will always consider you guys ~~good frien...~~ ~~41/0/0 HuS...~~ ...people who somehow found out how to start up their computer by trial and error. To use an inspiring and deep quote: "Also, Leute. Nicht im Feuer stehen; Haut raus was geht; und vor allem: ..."

- my flatmates from Paris: Stefan and Max and especially my neighbor Julie, without whose help I would have spent my time in Paris in a room without electricity or water.

Last but not least I would like to thank my family for their support: to my sister Daniela and her husband Martin; to my aunts and uncles and cousin; to my grandparents, especially to my two grandfathers, who sadly could not witness the conclusion of my thesis. And finally, this dissertation would not have been possible without my parents, who always supported me with practical help and advice, without ever talking me into things. Without you, I would be a completely different person than I am today.

# DECLARATION

Ich erkläre: Ich habe die vorgelegte Dissertation selbständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten.“

*Giessen, 5. August 2015*

Fabian Bietz