GENERATION AND CHARACTERIZATION OF TRANSCRIPTION ACTIVATOR-LIKE EFFECTOR NUCLEASES

LAURA WALTL

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

Institute für Biochemie
Fachbereich 08
Justus Liebig Universität
– January 2016 –

This dissertation was prepared at the *Institue for Biochemistry* in the department of *Biology and Cemistry* (FB08) of the J*ustus Liebig University Giessen*. The DFG financed graduete colloge " Enymes and Multienzyme Complexes acting on Nucleic Acids" (IRTG 1393) sponsored the work.

Primary Referee:
Prof. Dr. Peter Friedhoff
Institut für Biochemie
Fachbereich Biologie und Chemie
Heinrich-Buff-Ring 58
35392 Giessen

Secondary Referee:
Prof. Dr. Alexander Brehm
Institut für Molekularbiologie und Tumorforschung (IMT)
Philipps-Universität Marburg
Emil-Mannkopff-Str. 2
35033 Marburg

Supervisors:
Dr. Wolfgang Wende

IRTG supervisors:

(I) Prof. Dr. Alfred Pingoud, Giessen

(II) Prof. Dr. Alexander Brehm, Marburg

(III) Prof. Dr. Virginijus Siksnys, Vilnius

Laura Waltl: *Generation and characterization of Transcription activator-like effector nucleases,* Dissertation zur Erlangung des Akademischen Grades des Doktors der Naturwissenschaften (Dr. rer. nat), © January 2016

Erkenne dich selbst!

— Thales von Milet (~ 625- ~547 v. Chr.)

Dedicated to my parents.
Gert & Angelika Waltl

ABSTRACT

Designer nucleases are developed to modify the genome in various organisms, and are promising molecular tools for the advancement of gene therapy. They should introduce only one specific double-strand or single-strand break in a complex genome in order to trigger the DNA repair mechanisms of the cell such that the DNA damage is repaired by homologous directed repair (HDR) or non-homologous end joining (NHEJ), an error-prone repair mechanism that can be applied to knock out genes. With the addition of an external DNA fragment, HDR can be exploited to correct or add a gene.

Transcription activator-like effector (TALE) nucleases belong to a platform of designer nucleases and are known to be highly specific, efficient and easy to customize for any given target. Nevertheless, there is still a need to improve the accuracy of TALE nucleases for their reliable application *in vivo*. To this end, off-target activity should be avoided without compromising cleavage efficiency. This can be achieved by optimization of the catalytic domain as well as by improving the binding domain.

The subjects of this comparative study are the homing endonuclease I-SceI, the widely used TALE-FokI fusion protein, the new RNA guided Cas9 nuclease and our recently developed TALE-PvuII variants as well as the TALE-MutH fusion protein, whose nuclease domain is a sequence- and strand-specific nickase. In the study we have analyzed the toxicity and cleavage efficiency of these specific nucleases in a yeast based single-strand annealing assay as well as *in vitro*. The present study also focus on the characterization of the *in vitro* binding mechanism of TALE. Binding characteristics of AvrBs3($\triangle$ N152-C28) were analyzed via anisotropy measurements, specific single- and double-labeled TALE variants were designed and DNA::Protein complexes were analyzed employing FRET measurements. We gained labeled variants remaining their binding characteristics and enables the investigation of the *in vitro* binding mechanism of TALE proteins.

Our results demonstrate convincingly that the monomeric TALE-MutH nickase displays less toxicity in *Saccharomyces cerevisiae* compared to the widely used dimeric TALE-FokI nuclease and is able to introduce the repair of the DNA damage via single-strand annealing as efficient as I-SceI, the GOLD STANDARD for cleaving genomic DNA. To improve the design of TALE-MutH nickases, we analyzed the cleavage behavior on various specific loci *in vitro* and *in vivo*. The results presented show that the TALE-MutH fusion protein is a promising programmable nickase for *in vivo* applications and these results were confirmed by a TALE-MutH addressing a relevant therapeutic gene.

ZUSAMMENFASSUNG

Konstruierte Nukleasen wurden entwickelt um das Genom in verschiedenen Organismen zu modifizieren und sind viel versprechende Werkzeuge zur Weiterentwicklung von effizienten Gentherapien. Spezifische Nukleasen sollten einen spezifischen Doppel- oder Einzelstrangbruch in dem komplexem Genom induzieren und hierdurch den DNA Reparaturmechanismus der Zelle auslösen. Die DNA wird hierbei durch *homologous directed repair* (HDR) oder *non-homologous end joining* (NHEJ) repariert. NHEJ ist ein zu Fehlern neigender Reparaturmechanismus und kann zur Ausschaltung eines Gens angewendet werden. Mit Hilfe eines externen DNA Fragments kann HDR zur Korrektur oder zum Hinzufügen eines Gens genutzt werden. *Transcription activator-like effector* (TALE) Nukleasen bilden eine Gruppe der konstruierten spezifischen Nukleasen und sind aufgrund ihrer hohen Spezialität, Effizienz und einfachen Anfertigung für beliebige adressierte Sequenzen bekannt. Dennoch werden weitere Verbesserungen benötigt um die Genauigkeit von TALENs für die zuverlässige *in vivo* Anwendung zu gewährleisten. Hierbei müssen *off-target* Ereignisse vermieden werden ohne dabei die Spaltaktivität zu kompromittieren. Dies kann durch die Optimierung der katalytischen Domäne, sowie durch die Verbesserung der Bindungsdomäne erreicht werden.

Die Themen dieser vergleichenden Arbeit sind die Homing Endonuklese I-SceI, das weit verwendeten TALE-FokI Fusionsprotein, die neuartige RNA gesteuerte Cas9 Nuklease und die erst kürzlich etablierten TALE-PvuII Varianten, sowie das TALE-MutH Fusionsprotein. Letzteres besitzt eine Nuklease Domäne, welche eine Sequenz- und Strand-spezifische Nickase darstellt. In der Arbeit haben wir die Toxizität und Spaltaktivität dieser spezifischen Nukleasen in Hefe durch einen *single-strand annealing assay* und *in vitro* analysiert. Die aktuelle Studie konzentriert sich auch auf die *in vitro* Charakterisierung des Bindungsmechanismus von TALE. Die Bindungscharakteristika von AvrBs3($\triangle$ N152-C28) wurden durch Anisotropiemessungen bestimmt, spezifisch einzel- oder doppelt-markierte TALE Varianten wurden hergestellt und DNA::Protein Komplexe wurden durch FRET-Messungen untersucht. Wir konstruierten fluoreszenzmarkierte Varianten die ihre Bindungscharakteristiken behielten und somit die Untersuchung der *in vitro* Bindungsmechanismen ermöglichen.

Die Ergebnisse zeigen dass die monomere TALE-MutH Nickase eine geringere Toxizität in *Saccharomyces cerevisiae* im Vergleich zu der dimer TALE-FokI Nuklease aufweist. Des weiteren ist TALE-MutH zur Reparatur des DNA Schaden durch *single-strand annealing* fähig und hierbei genauso effizient wie der Goldene Standard I-SceI. Zur

Verbesserung des TALE-MutH Designs wurde das Spaltverhalten auf verschiedene Loci *in vitro* und *in vivo* getestet. Die präsentierten Ergebnisse zeigen dass das TALE-MutH Fusionsprotein eine vielversprechende Nickase für *in vivo* Anwendungen darstellt mit welchem auch therapeutisch relevante Gene adressiert werden können.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AA	amino acid
aa	amino acids
AAV	adeno-associated virus
AD	activator domain
BFP	blue fluorescent protein
B2H	bacterial two-hyprid
bp	base pair
CC	connection cilium
CRISPR	clustered regularly interspaced short palindromic repeats
CRD	central repeat domain
crRNA	CRISPR RNA
Cy3	cyanine 3
Cy3	cyanine 5
DMD	Duchenne muscular dystrophy
DNA	desoxyribonukleinsäure
DOL	degree of labeling
DSB	double-strand break
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
F9	factor IX
FGE	formylglycine generating enzyme
FRET	Förster resonance energy transfer
gal	galactodsidase
GFP	green fluorescent protein
gRNA	guided RNA
HDR	homologous directed repair

HE homing endonuclease

HEK 293T human embryonic kidney 293 T

HEX hexachlorofluorescein

HGP Human Genome Project

HR homologous recombination

IDLV integration-deficient lentiviral vectors

IFT intraflagellar transport

INDELs insertion and deletion events

IPTG isopropyl-beta-D-thiogalactopyranosid

IS inner segments

LB Lysogeny broth

LCA leber congenital amaurosis

LPLD lipoprotein lipase deficiency

NHEJ non-homologous end-joining

NLS nuclear localisation signal

NTR N-terminal region

OD optical density

OPEN oligomerized pool engineering

OS outer segments

PAA polyacrylamind

PAGE polyacrylmide gel electrophoresis

PAM protospacer-adjacent motif

PCR polymerase chain reaction

PEI polyethylenimin

PK proteinase K

PR photoreceptor

R resistance

RE restriction endonuclease

RipTAL Ralstonia injected protein TALE

RNA ribonucleic acid

RNF RNA-guided FokI-dCas9 nuclease

RPE retinal pigment epithelium

RPTG retinitis pigmentosa GTPase regulator

RT room temperature

RVD repeat variable diresidue

S host-susceptibility

SaCas9 Staphylococcus aureus Cas9

sc single-chain

SDS sodium dodecyl sulfate

SN2 bimolecular nucleophile substitution

SOC super optimal broth with catabolite repression

spCas9 Streptococcus pyogene Cas9

SSA single-strand annealing

SSB single-strand break

TAE tris-acetic acid-EDTA

TBE tris-borate-EDTA

TPE tris-phosphate-EDTA

TALEN transcription activator-like effector nuclease

TALE transcription activator-like effector

TFO triplex-forming oligonucleotid

TLR traffic Light Reporter

tracrRNA transactivating CRISPR RNA

UPA upregulated by AvrBs3

UPT upregulated by TALE

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

Xcv Xanthomonas campestris pv. vesicatoria

XLRP X-linked form of retinitis pigmentosa

ZFN zinc-finger nuclease

Part I

INTRODUCTION

1 INTRODUCTION

In 1990, the Human Genome Project (HGP) was initiated to decipher the human genome. It took until 2003 to complete sequencing of the entire human genome, which improved our understanding of genetic disease mechanisms and resulted in the development of new therapeutic strategies. One therapeutic approach was to silence dysfunctional genes with small interfering RNAs (siRNAs), resulting in the transient down-regulation of the expression specific targeted genes. However, its use is limited by the relatively low efficiency and its unfavorably safety profile [4, 114, 50]. Another approach is a therapy, which inserts the relevant gene by packing the encoding sequences into a viral-vector. A successful application of a viral-vector based therapeutic is Glybera, it treats lipoprotein lipase deficiency (LPLD) and is marketed in Europa since 2014. Glybera is injected into the skeletal muscle and introduces a healthy lipoprotein lipase gene [84, 232]. However, editing therapies carrying the risk of an immune response to the viral-vector and the non-specific insertion of the correct gene [121]. This has been shown where gene therapy was successfully employed to cure severe combined immunodeficiency-X1 (SCID X-1) via introducing a retroviral vector containing the encoding sequence of the relevant protein. However, some patients developed leukemia-like symptoms as a result of an unspecific insertion [121, 92]. Furthermore, a lethal immunological response was reported in a therapeutic trial using adenovirus vectors for the insertion of the relevant gene [216].

Homologous recombination (HR) of an exogenous DNA template with chromosomal DNA can be used to add sequences to the genome by gene correction or gene addition. However, recombination in mammalians is a rare event and therefore very inefficient. Chromosomal HR can be increased over 1000 fold by introducing double strand breaks (DSBs) in the target, thus facilitating gene targeting [61, 130, 182, 207, 181]. Various specific nucleases have been developed to introduce specific DSBs into the genome, such as meganucleases [87, 9, 201], zinc finger nucleases (ZFNs) [220, 176, 31], transcription activator-like effector nucleases (TALENs) [155, 41, 148] and the RNA-guided Cas9 nuclease [108, 140, 39, 47, 102] (Figure 1).

Site-specific DSBs trigger the cell to repair the lesion by the error prone repair mechanism non-homology end joining (NHEJ) [131] or by homology directed repair (HDR) [30, 95, 159, 60, 187]. NHEJ introduces small insertions or deletions (INDELs), leading to a disruption of the gene [131]. Thus, it can be applied to silence a pathologic gene [189, 211].

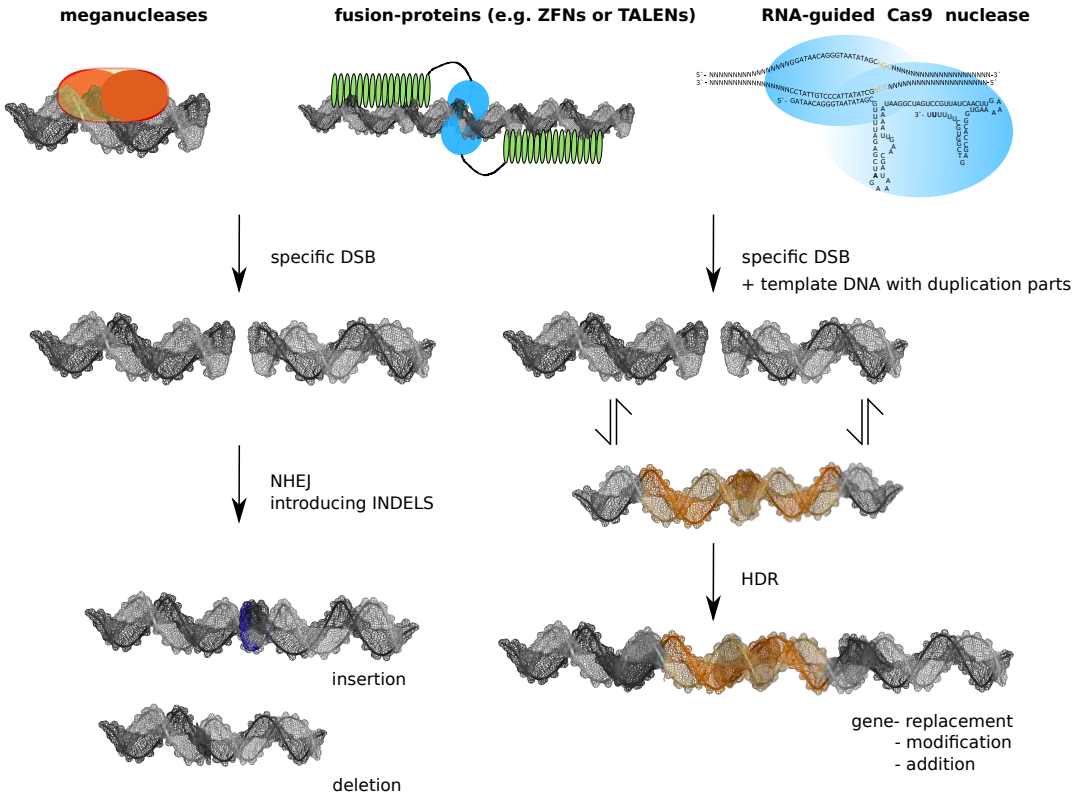

Figure 1: Gene targeting with meganucleases, fusion-proteins or RNA-guided endonucleases
Specific nucleases induce a DSB into the genome and trigger the error prone NHEJ mechanism or HDR. NHEJ results in INDELs and can disrupt the gene. HDR is based on a DNA template containing duplication fragments. HDR can bee used for gene correction, gene replacement, or gene addition.

NHEJ is the more efficient repair mechanism compared to the precise homologous direct repair and is active throughout the whole cell cycle [186, 197]. The ends of the disrupted DNA are literally sticked together via NHEJ [131]. However, if a precise gene correction or gene addition is needed, a DNA template has to be introduced, which will be assembled to the DNA via HDR [61, 130, 182, 207]. However, HDR does primarily occur during the S/G2 phase of the cell cycle [43, 35] and is therefore less active and restricted to dividing cells [50].

To increase the possibility of HDR, several modifications were made [50]. For instance, suppressing the key players of NHEJ [42], but also single-strand breaks (SSBs) were introduced instead of DSBs to the DNA locus [145, 52, 222], single-stranded oligonucleotides were used instead of double- stranded substrates [36, 146], and the similarity was increased between the DNA template and the DSB site to promote formation of intermediates during the template integration [19, 154, 158].

The improvement of gene targeting is an ongoing process. Thus, the specific nucleases have to be optimized to reduce safety risks,

such as off-target activity [228, 93, 77] and a deeper understanding of the binding mechanism could improve their specificity [142, 205, 193, 194]. Furthermore, the precise characterization of the repair mechanisms could optimize HDR efficiency [113]. Finally, improving the delivery systems to the target site could be an important step towards the effective manipulation of the human genome [121].

1.1 TOOLS FOR GENOME EDITING

The tools for genome editing are the specific nucleases, which require a number of features to be effective and safe. They have to be highly specific and highly efficient, but also easy to handle and to design. The most important specific nucleases are reviewed below.

1.1.1 *Meganucleases/Homing endonucleases*

Homing endonucleases (HEs) are one of the first engineered nucleases [61, 130, 182, 207, 181]. HEs, also known as meganucleases, target an extremely rare large recognition site (12-40 bp) and are grouped into five families based on sequence and structural motifs [206]. HEs are located and encoded within protein encoding genes. HEs are self-splicing introns (RNA) or self-splicing inteins (protein splicing). They are considered as "selfish" nucleases as they support their own proliferation. The HE induces a DSB infto the foreign DNA lacking the HE gene, triggering HDR. In turn, HDR uses the intron containing allel as the repair matrix and the HE gene becomes duplicated [141]. I-SceI belongs to the best analyzed LAGLIDADG family, characterized by the $\alpha\beta\beta\alpha\beta\beta\alpha$ motif. Two LAGLIDADG motifs form a bipartite catalytic center. Each center α-helix contains the LAGLIDADG motif and enclose one of the two acidic residues, which coordinate the divalent metal ion cofactor Mg^{2+} for cleavage of the phosphodiester bonds. The four β-sheets recognize the target by forming a saddle like structure around the DNA. HEs exist as homodimer with two copies of the LAGLIDADG motif in each monomer (e.g. I-CreI) or as monomeric enzyme containing two LAGLIDADG motifs fused by a linker (I-SceI) [201, 141, 171]. I-SceI recognizes a non-palidromic 18 bp sequence and is the established GOLDEN STANDARD due to its high specificity combined with its high nuclease activity. Meganucleases are characterized by a non-modular structure, thus engineering of their binding domains proofed to be challenging, but can be realized by rational design and/or directed evolution [218, 106].

1.1.2 *Fusion-proteins*

In contrast to meganucleases, the repetitive nature of the DNA binding region of the fusion proteins ZFNs [220, 176, 31] and TALENs

[155, 41, 148] enable them to address a high number of sequences. The programmable binding domains of ZFs and TALEs are typically fused to the non-specific cleavage domain of the Typ IIs endonuclease FokI, which functions as a dimer [117]. Improved designs have been established and are reviewed in section 1.1.4.

Another option to engineer programmable fusion proteins is to use triple-helix forming oligonucleotides (TOFs) as binding modules. TOFs build stable triple helices with the DNA through hoogsteen base pairs [63, 171]. Due to the challenging cell delivery of TOF nucleases and the slow rate of triple-helix formation, this platform of specific nucleases has no current relevance in genome editing [171].

1.1.2.1 *Zinc finger nucleases*

A set of zinc-fingers consist of three individual zinc-fingers and each Cys_2-His_2 finger is composed of 30 amino acids formed to a $\beta\beta\alpha$-motif. Two cysteine and two histidine residues coordinate one divalent zinc ion. Three defined amino acid residues located in the α-helix recognize 3 bp in the major groove of the DNA [66]. A zinc-finger binding domain comprise three to four zinc-finger sets to target 9-12 bp. Several methods have been developed to construct zinc-finger proteins [3]. For example, modular assembly, which is based on a preselected library of zinc finger modules [196] or by using the oligomerized pool engineering (OPEN) method, which selects multi-finger arrays from a library and members with binding affinity are isolated with a bacterial two-hybrid (B2H) selection method [137, 136]. Unfortunately, crosstalk between adjacent zinc fingers affects the binding specificity [103] and interfere with zinc finger assembling [178]. Additionally, ZFN revealed a high off-target activity, when compared to TALEN [156].

1.1.2.2 *Transcription activator-like effector (TALE)*

Deciphering the code of a TAL (transcription activator-like) effector of the plant pathogenic *Xanthomonas* was a crucial step to improve binding domains for highly programmable specific nucleases. TAL effectors are virulence factors, which simulate eukaryotic transcription factors [25]. Gram-negative plant-phathogenic bacteria of the genus *Xanthomonas* inject a cocktail of effector proteins by a type III secretion system into plant cells, thus causing diseases in crop plants [115]. TAL effectors are composed of a N-terminal trans-location signal, a central repeat domain (CRD) and a C-terminal region carrying a nuclear localization signal (NLS) and a transcription activator domain [24, 26] (Figure 2 (A)). TALEs enter the nucleus and bind to a UPT (upregulated by TALE) box to activate host-susceptibility (S) genes, promoting bacterial proliferation. AvrBs3 is a well studied TALE protein from the pepper (*Capsicum spec.*) and tomato (*Solanum lycopersicum*) pathogenic strain *Xanthomonas campestris* pv. *vesicato-*

ria (Xcv). AvrBs3 interacts with the S gene UPA (upregulated by AvrBs3) inducing cellular hypertrophy of mesophyll cells [152]. Due to plant-pathogen co-evolution, disease resistance (R) genes exist in some plants. For instance, the pepper plant contains a *Bs3* R gene. Here, AvrBs3 induces programmed cell death, called hypersensitive response [185].

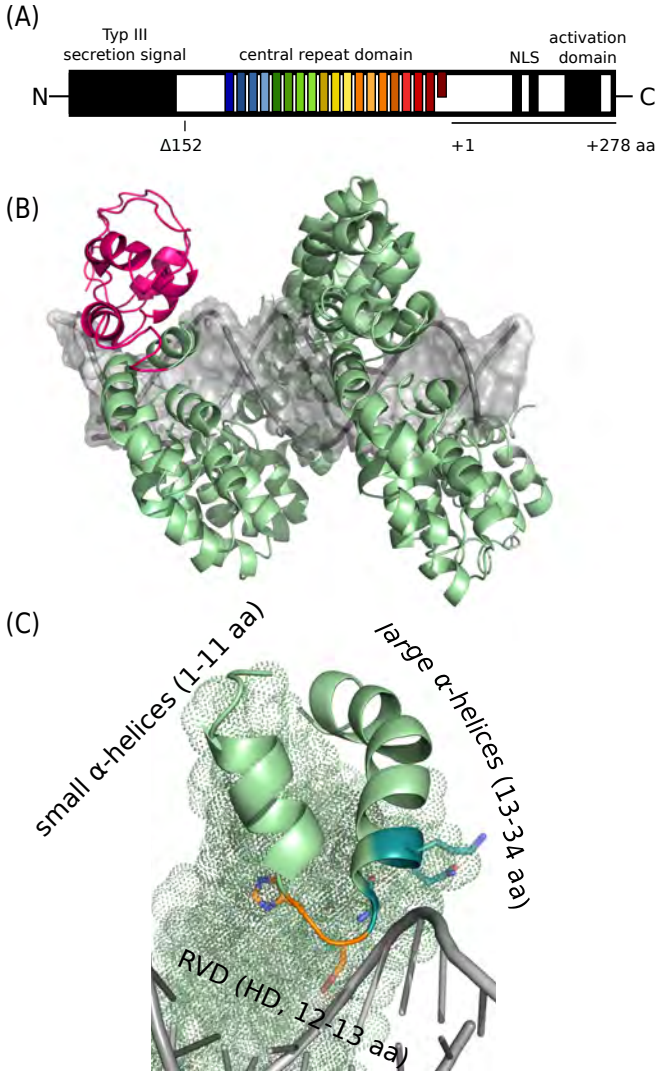

Figure 2: 3D structure of TALE (pdb: 2YPF) [205].

(A) TALE consist of the N-terminal region (NTR) carrying the type 3 secretion signal (T3SS), the central repeat domain (e.g. 17.5 RVDs), and the C-terminus containing the nuclear localization signal (NLS) and a transcription activator domain (AD). Δ152 indicates the truncated N-terminus, while the complete C-terminus carry 278 aa after its last half repeat. (B) Crystal structure of TALE bound to DNA. TALE forms a right-handed superhelix wrapped around the major groove of the DNA. The pseudo repeats at the N-terminal region (NTR) are colored in magenta, the CRD is shown in green.(B) Magnification of one repeat. The repeat variable diresidue (RVD) at position 12 and 13 is illustrated as sticks in orange and the residues G14, K16 and Q17 are also illustrated as sticks.

Another *Xcv* TAL effector is AvrBs4, which is 97% identical to AvrBs3 [28]. Its corresponding R gene *Bs4* is also found in some tomato plants, but can not be induced by AvrBs3 [152, 192]. Similar to the TALE protein from the plant pathogenic *Xanthomonas, a* TAL effector (RipTAL) from *Ralstonia solanacearum* with high sequence similarity to the TAL effector of *Xanthomonas* was identified [54], as well as a modular DNA binding protein, BurrH, from *Burkholderia rhizoxinica* [110, 204].

1.1.2.3 *Structure and binding mechanism*

The crystal structure of the binding domain of a TAL effector PthXo1 was accomplished from the rice pathogen *Xanthomonas oryzae* [139]. Later, AvrBs3 bound to DNA was crystallized by Stella *et al.* [205] (Figure 2 (B)). The central binding module is composed of highly conserved repeats, ranging from 1.5-33.5 repeats. The last repeat does always contain half of the repeat residues [24, 25]. There exist conflicting evidences regarding the necessity of the last half repeat; both a significant impact [226, 27] as well as a minor relevance [142, 248] for the binding character of TALEs have been reported. The minimum number of repeats to induce gene expression is 6.5 repeats [25] and each repeat contains ~34 conserved residues. At position 12 and 13 the so called repeat variable diresidue (RVD) is located and is responsible for the "one repeat to one base" [193] specific recognition code [25, 153]. Every repeat forms two α-helices connected by a three-residue loop carrying the RVD (Figure 2 (C)). The TAL effector bound to double stranded DNA forms a right-handed superhelix wrapped around the DNA major groove [139, 205] (Figure 2 (B)). A total of 20 unique RVD sequences have been described [48], but only a few promote specific binding. RVD "HD" prefer binding to cytosine (C), RVD "NG" binds mainly to thymine (T) and RVD "NI" binds usually to adenine (A). RVD "NN" can bind to adenine (A) and to guanine (G), whereas RVD "NS" binds to adenine (A) and to cysteine (C) [25]. RVDs "NK", "HN" or "NA" can be used to target guanine (G) [153]. The RVD "N*" missing the glycine residue at position 13, has been described to overcome the sensibility of TALEs to methylated bases [221].

The single repeats can be assembled to an artificial functional TALE by the GOLDEN GATE cloning strategy [33]. This cloning strategy enable rapid assembling of TALEs and the use of TALEs for any given target [33].

The residue at position 13 directly contacts the DNA and the residue at position 12 is responsible to stabilize the interaction between DNA and TALE [139, 57]. The residues G14, K16 and Q17 interact by non-specific polar and charge interactions with the DNA backbone [230]. Furthermore, the N-terminal region (NTR) contributes to DNA binding and consists of four continuous repeats (N_{-3}, N_{-2}, N_{-1}, N_0) forming

a right handed superhelical structure similar to the CRD, but differing in its sequence [81]. Truncated versions of TALEs have been designed to reduce the overall size of the protein. It has been shown that the N-terminus missing 152 aa and the C-terminus carrying 28 aa retains its binding ability [148]. A binding sequence beginning with a thymine (T_0) facilitates TALE binding, but other bases can also be recognized by TALE [148, 209]. Tryptophan at position 232 in repeat N_{-1} is responsible for its T_0 preference. By structure-guided library design, a N-terminal domain preferring guanine (G_0) has been constructed [124]. It has been reported that *Ralstonia solanacearum* TALE requires G_0 and carrying an arginine at position 232 in repeat N_{-1} , similar to the TALE designed by structure-guided library [124, 54].

The naturally occurring TALEs from *Xanthomonas* bind to a DNA sequence containing a 5' T_0 [25, 153]. Interestingly, there is no naturally occurring TALE protein which matches perfectly to its target promoter. For instance, AvrBs3 has two HD-A mismatches and one NG-C mismatch at its Bs3 box [142]. It has been reported that TALEs from *Xanthomonas* possess single repeats with aberrant length. These repeats can be truncated or extended, when compared to the canonical repeat carrying 34 aa. The aberrant repeats allow an additional binding conformation in which the repeat loops out and therefore introduce a local flexibility to TALE-DNA binding [183].

A detailed analyses of the RVDs revealed that the overall composition of RVDs contribute to the TALE-DNA binding affinity. The order of contribution of single RVDs has been shown to be HD $\geq$ NG $\geq$ NN >>> NI > NK. Furthermore, the TALE-DNA recognition is influenced by a polarity-effect, thus the N-terminal repeats have an higher impact on the specificity and affinity of TALE compared to the C-terminal repeats [142].

Each repeat carries a conserved cysteine (C) residue at position 30, which is located in the second α-helices in a solvent-exposed outer loop [57, 139]. The cysteine residues are required to dimerize via disulfide bridges to homodimers, which occur prior to the import into the nucleus [91, 194]. However, complex formation inhibits binding to the DNA and it has therefore been suggested that the homodimers dissociate in the nucleus. Monomeric binding of TALE to its target has been shown by EMSA (electrophoretic mobility shift assays) data [194] and by the structural data [205, 139, 57] of TALE. While substitution off all cysteine residues to serine or alanine in AvrBs3 revealed that the cysteine is required for gene activation in *planta,* the lack of cysteine residues slightly reduces the binding ability of TALEs under *in vitro* conditions. These results suggest that the substitution slightly affects the plasticity of AvrBs3 [194]. X-ray crystallography revealed that TALE undergoes a conformational change upon binding to the DNA. The superhelical pitch gets reduced from 60 A° to 35 A° [57, 230]. Furthermore, conformation elasticity of TALE was

predicted by computational studies [226, 70]. Just recently, single-molecule experiments were performed with Cy3 labeled AvrBs3. The protein was labeled by hydrazine-functionalized Cy3 organic dyes at the N-terminus carrying an unnatural aldehyde and a two-state mechanism enclosing a dynamic conformation change between search and recognition state was reported [51]. The NTR binds to non-specific DNA and facilitates 1D sliding [138, 81, 51]. An extended superhelical pitch exists, which minimizes the strong interaction between CRD and the DNA backbone. This enables sliding and hopping of the TALE protein along the DNA. The single molecule results revealed an arrested state in 1D diffusion, which indicates a transition from the search state to the recognition state. The recognition state has a more compact conformation, when compared to the search state. Furthermore, they suggested that large TALEs suffer from a "slowdown" in 1D diffusion. This results in a higher frequency of transition from search to recognition state and thus results in higher off-target rates [51].

1.1.3 *Clustered regularly interspaced short palindromic repeats (CRISPR)-associated endonuclease Cas9*

The clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas)-system is an adaptive immune system in bacteria and archaea to defend the organism against invading viruses or plasmids (Figure 3) [231].

The type II CRISPR-Cas system integrates foreign sequences (protospacer) to the CRISPR cluster and transcribes these to pre-CRISPR RNAs (pre-crRNAs), which are processed to short crRNA and hybridized with transactivating CRISPR RNA (tracrRNA) and the Cas9 protein. The complex Cas9::crRNA::tracrRNA binds to its target by RNA-mediated Watson-Crick interaction and induces a DSB. To simplify the system for genome editing, a single crRNA-tracrRNA chimera was established by fusing crRNA to a truncated tracrRNA which resulted in a guide RNA (gRNA) [108]. RNA-guided Cas9 nuclease recognizes a 20 bp target which requires the protospacer adjacent motif (PAM) sequence downstream of the target for binding and cleaving activity. The commonly used *Streptococcus pyogenes* Cas9 (spCas9) protein primarily recognizes 5' NGG PAM sequence [108, 107]. The Cas9 protein contains two nuclease motifs, RuvC and HNH. RuvC induces a SSB to the complementary strand and HNH induces a SSB into the non-complementary strand, resulting in a blunt end 3 bp upstream to the 5' PAM motif [83]. The required PAM motif represents a limitation for the application of the gRNA endonuclease Cas9. However, there have been attempts to overcome the issue of the PAM motif [107, 119, 7, 72]. Partial binding of gRNA::Cas9 was demonstrated to the PAM sequence 5' NAG [107]. The design of an engineered

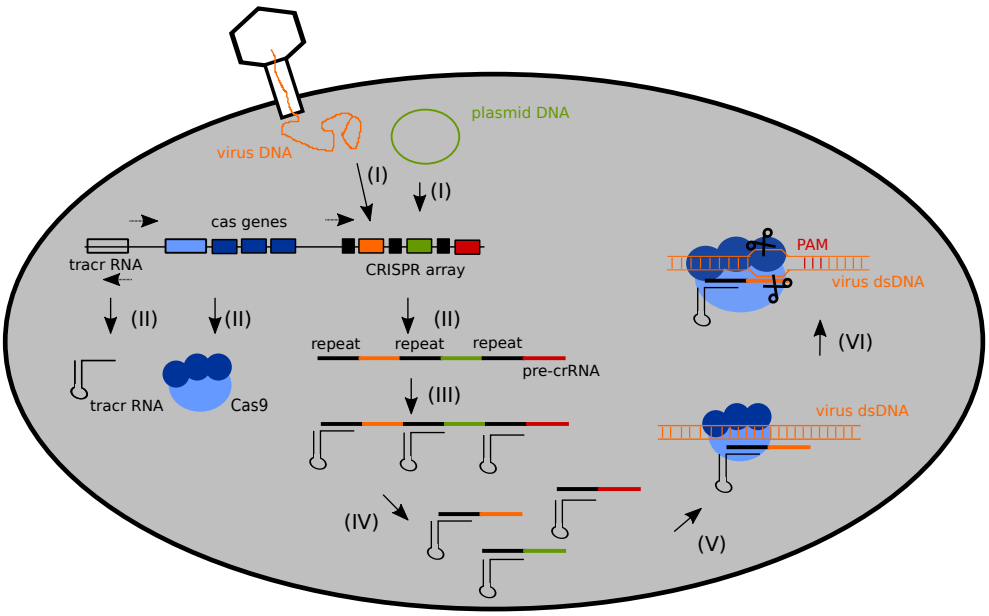

Figure 3: CRISPR-Cas system
(I) integration of a short foreign sequences (protospacer) into the CRISPR array. The protospacers are illustrated in orange, green, and red. The repeat arrays are displayed in balck. (II) Encoding tracrRNA, Cas9 protein and pre-crRNA. (III) The tracr RNA hybridizes to the repeat arrays of the pre-crRNA and (IV) are processed to short cr-RNAs. (V) Complex formation of Cas9::crRNA::tracrRNA. The complex scans the dsDNA for its target. (VI) Specific binding of the Cas9::crRNA::tracrRNA complex to the complementary protospacer sequence by Watson-Crick interaction. The PAM sequence is required to induce a DSB to the foreign DNA.

spCas9 protein with novel PAM recognition has been shown to be challenging [7], but has been recently achieved by a bacterial selection system [119]. Another attempt to improve the diversity of PAMs was the characterization of orthologues Cas9 proteins from other bacteria [119, 72]. Especially, the *Staphylococcus aureus* Cas9 (saCas9) revealed to be promising [119, 179]. It recognizes three novel PAM sites (NNGGGT, NNGAAT, NNGAGT) [119] and the protein saCas9 has the benefit to be ~1kb shorter compared to spCas9 [179]. The small size improves the packing features of the Cas9 protein into adeno-associated virus (AAV) vectors, a commonly applied delivery method. AAV vectors allow to pack 4.5 kb, which limits packing of the widely used spCas9 (~4.2 kb) and its gRNA into one AAV vector [179]. Additional a further CRISPR-Cas system was characterized containing a Cpf1 protein. In contrary to the Cas9 protein the Cpf1 protein does not require a transactivating CRISPR RNA, addresses a target site with a 5‘ T-rich PAM sequence and induces a DSB with a 4 nt or 5 nt 5‘ overhang[243].

A significant advantage of the gRNA endonuclease is the ability of re-targeting a new sequence by simply changing the sequence of the gRNA. Furthermore, the possibility to introduce multiple DSBs (mul-

tiplexing) allows new applications for the Cas9 system [47]. Nonetheless, the Cas9 nuclease does tolerate mismatches especially in the 5' upstream region [163, 101, 76]. This limits the application of Cas9 for genome editing, especially when a high level of precision is required. To improve the specificity of RNA-guided Cas9, a nickase has been established, by introducing a point mutation in either the RuvC (D10A) or HNH (H840A) domain [108, 47]. This led to the possibility to use RNA-guided Cas9 endonuclease for inducing SSBs, but also as an inactive binding domain [174, 247].

1.1.4 *Catalytic domains*

1.1.4.1 *Cleavage domains*

Different types of nucleases have been used in fusion constructs as cleavage domains [238, 78, 22, 27]. Nucleases are phosphordiesterases hydrolyzing the phosphordiester bond, thus creating most commonly 5'-phosphates and 3'-OH groups. These proteins are essential for defending the host against invading nucleic acids [21, 175], DNA replication [111] and RNA processing [2]. Phosphordiester bond cleavage occurs due to a bimolecular nucleophile substitution (S_N2), which inverts the configuration of the phosphorous. A neighboring water molecule is deprotonated to a strong nucleophile, which attacks the partially positive phosphor. Simultaneously, 3'-O^- is released. The reaction step can be catalyzed by diverse mechanisms depending on none-, one-, two-, or three-metal-ions [169, 237]. The cleavage domains analyzed in the present work belong to the two-metal-ion mechanism (FokI, PvuII and MutH) (Figure 4).

One divalent metal-ion facilitates deprotonation of the water molecule and both metal-ions stabilize the pentavalent transition intermediate. The two divalent metal-ions are coordinated by a conserved asparate, a hallmark of the two-metal-ion mechanism [236, 169]. Homing endonucleases were also used as cleavage domains in fusion constructs. The LAGLIDADG homing endonucleases I-CreI [22] and I-AniI [27] belong to the two-metal-ion-dependent nucleases and were fused to the binding domain TALE. Furthermore, one-metal-ion dependent nucleases, like the GIY-YIG homing endonuclease I-TevI [20] and the non-specific endonucleases ColE7, NucA and EndA [22] have been shown to function as cleavage domains. The cleavage domains of the analyzed specific nucleases are reviewed in detail.

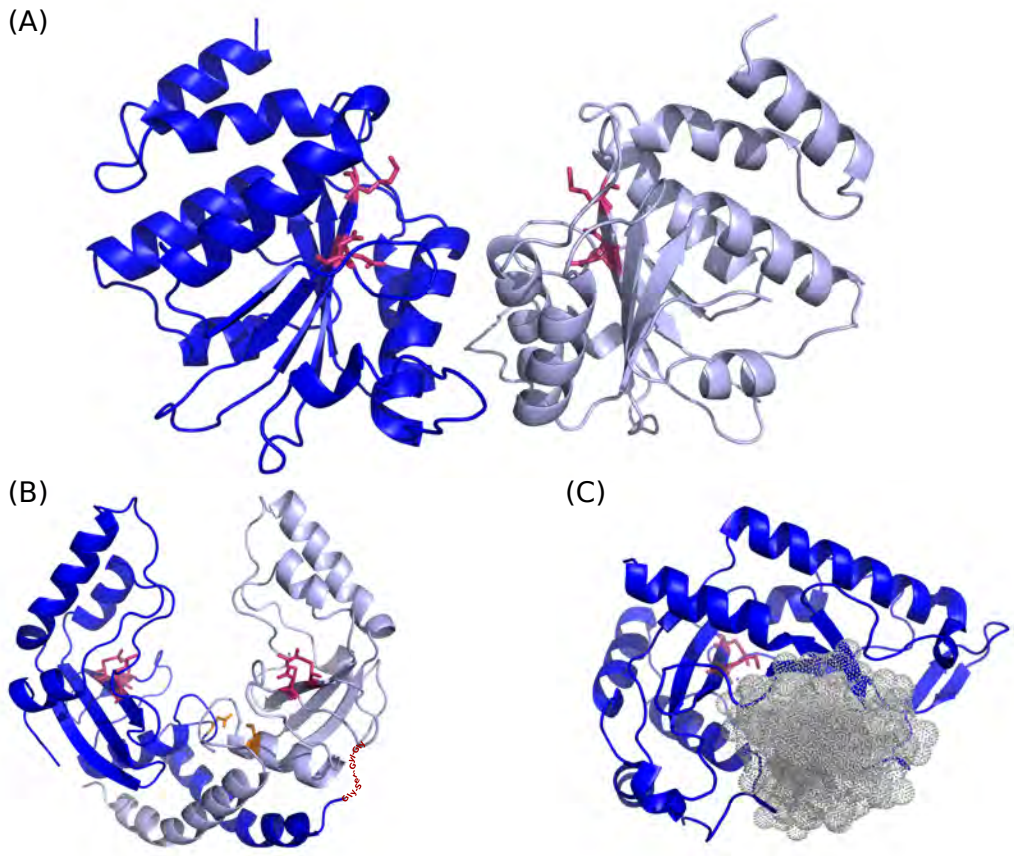

Figure 4: Cleavage domain of FokI, restriction enzyme PvuII and MutH
The essential residues of the cleavage domains are illustrated as purple sticks. One subunit is colored in blue and the second subunit in light blue. (A) Dimeric FokI (PDB: 2FOK) cleavage domains [224]. The active center is build by the residues D450, D467 and K468. (B) PvuII (PDB:1PvI) with its catalytic residues D58, E68 and K70 in each subunit [99]. Naturally, the PvuII restriction enzyme function as homodimer. To design a heterodimeric PvuII, the substitution D30K or K38D was introduced to one subunit, respectively. D30 and K38 are displayed as orange sticks. In red the linker (Gly-Ser-Gly-Gly) used to fuse the identical subunits of PvuII and to create a scPvuII variant is shown. (C) Monomeric nickase MutH interacting with its target DNA. (PDB: 2AOQ) Catalytic residues are D70, E77 and K79 [125]. Double stranded DNA is demonstrated in gray.

1.1.4.2 *FokI*

The endonuclease FokI, isolated from *Flavobacterium okeanokoites*, belongs to the type IIS restriction endonucleases.

The enzyme recognizes a 5 bp asymmetric nucleotide sequence, cleaving 9 (top-strand) and 13 (bottom-strand) nt downstream of the site (5'-GGATG(N)$_9$$^{\bigtriangledown}$NNNN$_{\bigtriangleup 13}$) [208].

Two domains shape the FokI restriction endonuclease, a N-terminal binding domain (41 kDa) and a C-terminal cleavage domain (25 kDa)

[128]. The two domains are fused by a flexible linker and the cleavage domain is kept in an compact conformation until binding to the substrate [167]. Binding to the DNA target induces a conformational change, which recruits the catalytic cleavage domain 13 nt downstream of the site to the bottom-strand. Next, a second monomeric FokI enzyme dimerizes by protein-protein interaction and is responsible to cut the the top-strand 9 nt away from the recognition site [167, 188, 23].

The second monomeric FokI endonuclease can be bound to a proximate second recognition site or is available from the free solution [93]. However, the dimerization between bound and free FokI proteins is unstable, resulting in a low cleavage rate for one-target site DNA [23]. Unless free FokI enzyme concentration is >>100 nM, since the equilibrium dissociation constant of the second FokI monomer is ≈100 nM [32]. Consequently, the catalytic domain of FokI is an unspecific cleavage domain, which requires dimerization to induce DSBs [224, 23] (Figure 4 (A)).

The catalytic domain of FokI was fused to the C-terminus of a zinc-finger domain by the Chandrasegaran group, creating the first reported ZFN [117]. Thereafter, improved ZF-FokI designs were created. One approach to increase the specificity was to design heterodimeric catalytic domains by introducing defined mutations [147, 212, 90]. Furthermore, *sharkey*, a FokI catalytic domain with increased activity was designed and is reported to be compatible with the heterodimeric mutations [90]. Another approach to optimize the construct was the design of the single-chain (sc) ZN-FokI specific nuclease. Two copies of the FokI catalytic domain were fused with a flexible $(Gly\text{-}Gly\text{-}Gly\text{-}Gly\text{-}Ser)_n$ polypetid linker to each other [149, 213]. When TALE, as a new binding domain was discovered, the cleavage domain FokI was applied to the new design [33, 129, 148]. Likewise to the scZN-FokI design, a scTALE-FokI [210] construct with a 95 aa polypeptide linker was constructed. Recently, the unspecific catalytic domain of FokI was fused to the amino-terminus of a catalytic inactive Cas9 protein to improve the off-target character of the CRISPR/-Cas9 -system [89, 219].

To enable FokI dimerization, two adjacent sites are placed 15-39 bp apart from each other with PAM sites facing outside. This new dimeric RNA-guided FokI-dCas9 nuclease (RNF) has been shown to improve the sensitivity to mismatches [8, 235].

1.1.4.3 *PvuII*

The restriction enzyme PvuII (Figure 4 (B)), isolated from *Proteus vulgaris*, belongs to the type IIP restriction enzymes. PvuII addresses a palidromic 6 bp sequence and induces a DSB through dimerization, resulting in blunt ends (5'-CAG$^{\bigtriangledown}$CTG-3'). Each PvuII subunit contains a PD...D/ExK motiv, characterized by the crucial aa of the

active center (D58, E68 and K70) [37, 170, 169]. The advantage of PvuII over the unspecific catalytic domain FokI is its additional site-specific sequence specificity [239]. Similar to the catalytic domain FokI, the PvuII enzyme was optimized for the use as cleavage domain. Thus, a scPvuII version has been designed by fusion of two identical PvuII subunits with a short aa linker (Gly-Ser-Gyl-Gly) [202, 144]. The first alternative architecture of programmable nucleases with the PvuII cleavage domain was the triple-helix-formation oligonuleotide (TFO)-scPvuII fusion construct [65]. However, this design has to be pre-incubated in the absence of Mg^{2+} metal-ions and activated with Mg^{2+} metal-ions to cleave the target site specifically. This required activation process complicates the application of TFO-scPvuII *in vivo* [239]. Furthermore, PvuII was fused to inactivated homing endonuclease I-SceI* [71], to a zinc-finger array [190] and to the TALE binding domain [239, 238]. The new site-specific architecture of programmable nucleases with PvuII proofed to be superior with the fusion constructs with FokI compared to zinc- finger-PvuII and zinc-finger-FokI [190]. PvuII connected to the truncated binding domain TALE (ΔN152-C28) was analyzed as homodimer, as monomeric single-chain construct and as heterodimer. Here, PvuII contained the T46G mutation, which is known to reduce "star-activity" and has been termed as high fidelity (hf) PvuII [250]. "Star-sites" are similar, but not identical to the recognition site. To reduce the toxicity of TALE-PvuII the cleavage domain PvuII was further optimized. Ten new PvuII variants, each with a novel single mutation, were generated by directed evolution. Five of these new variants retained the cleavage activity of the fusion construct. Among these five variants, the TALE-PvuII(T46G/G135W) variant revealed the best balance between activity and toxicity. Toxicity was analyzed in different *Escherichia coli (E. coli)* strains (JM109 and XL10Gold), as well as in an eucaryotic cell line (HEK293) [239, 238]. In addition, heterodimer PvuII subunits were designed to increase the specificity. Therefore, PvuII subunits were created containing a D30K mutation or a K38D mutation. These aa interact with each other during the dimerization of the PvuII subunits. The mutations are located centrally, but are not placed in the active center. Due to the different charges of PvuII(D30K) and PvuII(K38D), exclusively heterodimeric PvuII subunits can dimerize and induce a DSB [238].

1.1.4.4 *Programmable nickases*

To further reduce off-target activity, nucleases were modified by the utilization of nickases to induce SSBs rather than DSBs, thus favoring the occurrence of HDR over the error prone repair mechanism NHEJ [145, 52, 222].

Programmable nickases have been created by (i) inactivation of one of the binary catalytic cleavage centers (e.g. N450A FokI [116, 188,

177, 227], K223I I-SceI [157, 113] and D10A CAS9 Mali et al. [140], Ran et al. [180], Cho et al. [40]), or (ii) by fusion of a naturally occurring nickase to its binding domain, e.g. TALE-MutH [78]. Nickases have been designed to induce SSBs in the sense or anti-sense strand. This feature can be exploited to increase HDR efficiency, as it is crucial when the transcribed or the non-transcribed strand are nicked [53]. Furthermore, specific nickases have been used for double nicking (DN), thus inducing SSBs on both DNA strands, which resulted in an enhanced genome editing specificity [191, 199, 180, 40]. It has been shown by Ran *et al.* [180] that the efficiency of double nicking depends on the offset between the two target sites and on the created overhang. HDR efficiency triggered by paired Cas9 nickases has been reported to be comparable with the HDR efficiency induced by wt Cas9 endonuclease creating DSBs. Generating a 5'overhang does improve NHEJ and HDR efficiency [40, 180, 140]. While the offset can be up to 100 bp, it is optimal upto 20 bp. Finally, at least one nick should occur in the homologous part to optimize the efficiency and no improvement of HDR efficiency could be detected by generating a 3'overhang [180, 140].

1.1.4.5 *MutH*

MutH (Figure 4 (C)) is a 28 kDa monomeric site and strand specific latent endonuclease which requires activation by MutS and MutL associated with a mismatched base pair [125, 12]. It induces a SSB in un- and hemi-methylated strands 5' to its recognition site $^{\nabla}$GATC. The structure of MutH is surprisingly similar to the PvuII endonuclease. Furthermore, the active center of MutH contains the core aa Asp70, Glu77 and Lys79 (the DEK triad), which is also found in the PD...D/ExK motif of PvuII [12]. The monomeric endonuclease belongs to the mismatch repair system in system in *E. coli* and some other bacteria. MutS recognizes the mismatch and recruits MutL [94]. This ternary complex activates MutH, which induces a SSB at the 5´-side of a hemimethylated GATC site. The mismatch can be >1000 bp upstream or downstream of the GATC site [122]. Next, single-stranded DNA is generated and digested by an exonuclease across the original mismatch [151]. Without recruitment, the nickase MutH is barely active or inactive under physiological conditions [229, 64]. This feature was exploited by Gabsalilow *et al.* [78], as they fused the monomeric site and strand specific endonuclease to the binding domain TALE.

TALE-MutH is a strand- and site-specific programmable nickase, which has been characterized *in vitro* [78]. Dependent on the spacer length between the TALE binding site and the GATC site, the SSB is induced in the top- or bottom-strand. Targets containing a spacer of 3 bp are nicked on the bottom-strand, whereas the optimal spacer length for the top-strand cleavage is 6 bp [78]. This feature allows to

specifically target either the transcribed or the non-transcribed strand by TALE-MutH, which has been shown to be crucial in respect of HDR efficiency and therefore successful genome editing [53].

1.1.4.6 *Diversity of the catalytic domains*

Beyond the function as specific nuclease the described "tools for genome editing" were also modified for other applications. For instance, TALE proteins were fused to activator (e.g. vp16) [245, 152, 148] or repressor domains (e.g. KRAB) [82] to activate gene expression or to down-regulate or even arrest gene expression. Other designed fusion constructs were realized with DNA-methyltransferases [18], with demethylases [134], or with histon-modifying enzymes [143, 38]. To catalyze DNA integration, excision, and inversion recombinases [80, 85], or transposases [240] were fused to the binding domain TALE [79]. Additionally, TALEs were fused to fluorescent proteins to visualize target sequences [150]. Optical control of transcription was achieved by an optogenetic two-hyprid system, employing the light-inducible heterodimerizing proteins CRY2 and CIB1. This was realized by fusing CIB1 to the transactivation domain VP64 and CRY2 to TALE [120] or to the inactive Cas9 protein [172]. In the presence of blue light, a conformation change of CRY2 occurs and CIB1 forms a heterodimer with CRY2. Furthermore, inactive Cas9 protein was applied to repress transcription by blocking the RNA polymerase [174]. These examples underline the high relevance of the discussed proteins and their diverse applications.

1.2 THERAPEUTIC APPLICATIONS

A major limitation of the application of specific nucleases for genome editing is their specificity. Studies compared the specificity of the tools and revealed that the binding proteins tolerate mismatches depending on location and composition [142, 40]. For instance, TALE proteins are more sensitive to mismatches at the 5'-end of the target [142], while Cas9::gRNA tolerates mismatches at the 5'-end better than at the 3'-end [108, 47, 107]. TALENs have been shown to be the best choice when high accuracy is required (e.g. for gene correction) [155, 156, 228]. However, by choosing unique target sites [40], the possibility of double nicking with Cas9 proteins [40] and the fusion construct Cas9::FokI [219] increased the specificity and the field of application for Cas9 proteins. However, other factors affect the safe and efficient application of specific nucleases.

1.2.1 *Delivery systems*

The gene editing technologies have to be delivered into the target cell for their successful application. This can be realized by *ex vivo* or *in*

vivo editing, using viral or non-viral vector-mediated delivery or by employing physical delivery methods. The viral or non-viral vectors carry the expression cassette for the specific nucleases and in some case also the DNA template. The proteins can also be introduced to the cell as purified proteins. For *ex vivo* editing therapy, the target cell population is removed from the organ, modified and then re-engrafted into the body. This enables the usage of a wide range of delivery methods and a control of the dosage of the delivered molecules. However, *ex vivo* editing is limited to certain tissues, as the cell population needs to survive the manipulation outside the body and has to be able to return back to the tissue after transplantation. In *in vivo* therapies, the gene editing technologies are introduced directly into the diseased cells. This allows to target tissues not amenable for *ex vivo* editing therapy and the treatment of diseases affecting multiple organ systems [50]. However, major drawbacks of *in vivo* editing therapies are the possibility of an immune response to the viral-vector and the non-specific insertion of the corrected gene [121].

1.2.1.1 *viral vectors:*

Integration-deficient lentiviral vectors (IDLVs) and adeno-associated virus (AAV) are tools to deliver nucleases *ex vivo* and *in viv*o. They are charcaterized with high transduction efficacy and long-term gene expression. IDLVs provide the ability to pack the specific nuclease and the donor DNA for HDR, but showed to integrate randomly into the host genome. AAV vectors are non-pathogenic and facilitate delivery into dividing and non-dividing cells. Though, AAV vectors are limited by their carrying capacity (~4.5 kb). ZFNs are relative small (~1kb) and thus suitable for AAV vectors. However, TALENs can only be packed as monomers with a small promoter sequence into the AAV vectors. In the case of the spCas9 protein, a small orthogonal saCas9 was developed, which can be packed with gRNA to the AAV vector. Additionally, the repetitive nature of TALE-encoding genes are an interferrence, as rearrangements and truncations occur using both viral vectors [132, 121].

1.2.1.2 *non-viral vectors:*

Non-viral vectors possess a high carrying capacity, less toxicity and are easily produced. But they revealed to be less efficient and also toxic at high concentrations. Cationic liposomes and cationic polymers such as polyethylenimine (PEI) are commonly used non-viral-vectors [3]. The positive charge interact with the anionic nature of the cell membrane and triggers cellular uptake [121]. Furthermore, direct delivery of purified proteins was realized. Due to the positive charge of the zinc fingers, they are cell-penetrating proteins. Modified TALENs with cell-penetrating peptides do also enable cellular

uptake. A similar approach was used to promote cellular uptake of the Cas9 protein and its gRNA [132].

1.2.1.3 *Physical methods:*

The most commonly used physical methods for delivery of specific nucleases are electroporation and microinjection. Electroporation generates microscopic pores due to electrical pulses. By microinjection the editing technology can be introduced directly to the cell, using a needle (0.5 μm -20 μm in diameter) in combination with a specialized microscope [3].

1.2.2 *Impact of epigenetics on enzyme activation*

The accessibility of the target sequence in the chromatin structure is another aspect to consider. Relaxing the chromatin structure by the inhibition of epigenetic modulatory activity of the histone deacetylases improved the targeting frequency of specific nucleases [29]. Silencing a protein encoding gene which is required for chromatin remodeling has been shown to improve the efficiency of specific nucleases [56]. Additionally, blocking DNA-methyltransferase resulted in an increased TALENs activity [29]. Consequently, an optimized TALE RVD was designed to overcome the sensitivity to cysteine methylation [221]. Due to the epigenetic mechanisms of the cell cycle, the chromatin structure is looser in the S phase, when DNA replication occurs [13]. Thus, optimized cleavage takes place in the S phase of the cell cycle [164].

1.2.3 *Therapeutic potential*

ZFNs are the most commonly used therapeutic tool, as they are the first developed specific nucleases [117]. ZNFs addressing the CCR5 gene to treat Human immunodeficiency virus (HIV-1) resistance have already completed an early phase of human clinical trails. The CCR5 gene is disrupted by NHEJ, which prevents expression of the CCR5 receptor. This receptor is needed for viral uptake to the $CD4^+$ T cell, which serves the virus as host cell [97, 166, 214]. TALENs and Cas9 proteins were employed to treat HIV-1 in different cell lines [241, 68]. Another clinical relevant example is hemophilia B, a genetic disorder causing a deficiency of the coagulation factor IX (F9). Introducing a donor DNA with the expression sequence of factor IX by HDR increased the concentration of F9 and improved clotting time in transgenic hemophilic mice [126]. Treatment of the X-linked hereditary disease, Duchenne muscular dystrophy (DMD) was realized with ZFNs [161], TALENs [161] and the CRISPR/Cas9 system [133]. The monogenetic disease is caused by a mutation in the dystrophin gene, which is required for muscular function. Treatment was realized with ZFNs,

an *ex vivo* protocol, and manipulated undifferentiated cells were transplanted back to the limb of mice [161]. Furthermore, the newly established CRISPR/Cas9 system was used to treat DMD in mice [133].

Recently, a 11 month girl with B cell leukemia was successfully treated with universal chimeric antigen receptor (UCAR)19 T-cells. The UCAR19 T-cells were engineered by the TALEN-based editing technology of *Cellectis* to suppress the alloreactivity of the T-cells and acquire resistance to the monoclonal antibody *Alemtuzumab*, which is a drug to treat B cell leukemia [173]. These examples demonstrate the therapeutic potential of specific nucleases and the diversity of their genetic targets [121].

1.2.3.1 *Gene therapy without specific nucleases*

1.2.3.2 *Ocular gene therapy*

Treatment of photoreceptor (PR) degeneration in the eye by gene addition has been lately achieved [105, 11, 44]. The eye represent an excellent target for gene therapy [14]. Three independent clinical trails revealed the efficiency of gene therapy in patients with RPE65-associated leber congenital amaurosis (LCA) ((NCT00481546, NCT00516477, NCT00643747, clinicaltrials.gov). RPE65-associated LCA is caused by RPE65 deficiency, an isomerase in the retinal pigment epithelium (RPE) [44, 234]. RPE is a part of the blood-retinal barrier and in metabolic relationship with the adjacent PR layer of the neural retina.

The neural retina consists of the outer segment (OS) and the inner segment (IS), which are linked by the connection cilium (CC). The OS is light sensible and contains rods and cones, which are packed with the G-protein-couple receptor opsin [233, 100]. The IS is responsible for the protein production and the CC transports the proteins from the IS to the OS by intraflagellar transport (IFT) [100]. The light-sensitive chromophore 11-cis retinal is covalently bound to the G-protein-couple receptor resulting rhodopsin (RHO) (Figure 5).

Activation of rhodopsin occur by light due to isomerization of 11-cis retinal inducing a signal cascade [233]. Patients with RPE65 deficiency have degeneration of RPE65 and PR, due to the missing isomerase which results in the accumulation of all-trans retinal [233]. By introducing a recombinant AAV carrying the complete RPE65 gene, the retinal sensibility could be improved [45, 44, 11, 105].

1.2.3.3 *X-linked form of retinitis pigmentosa (XLRP)*

The successful application of gene therapy in RPE65-associated leber congenital amaurosis resulted in attempts to treat other PR degenerative diseases with gene therapy, such as the X-linked form of retinitis pigmentosa (XLRP). XLRP represents 22% of all forms of retinitis pigmentosa and 80% of all XLRP cases are caused by a mutation in the retinitis pigmentosa GTPase regulator (RPGR)-gene [100, 223, 198, 165]. The RPGR is concentrated in the CC, as well as in the OS of vertebrate PRs and plays an essential role for CC trafficking. In fact, RPGR can be found in many cells of the body in distinct splicing variants of the gene. The $RPGR^{ORF15}$ splicing variant contains a terminal exon with a repetitive glutamic acids rich domain, called open reading frame 15 (ORF15) and is found in large quantities in the PRs [100]. The ORF15 consists of an imperfect repeat motif, which carries 80% of sequence alteration, causing a dysfunctional RPGR [223, 100]. Opsin misslocation is often a hallmark for RPGR dysfunction [217, 234, 223].

Lately, AAV2/5 vectors carrying a full-length human $RPGR^{ORF15}$ cDNA were subretinally injected to four dogs which carried two distinct RPGR mutations [15]. Furthermore, the full-length $RPGR^{ORF15}$ packed into AAV 8 or 9 capsin were subretinally injected into *Rpgr*-KO mice and could rescue the retinal structure and function [234].

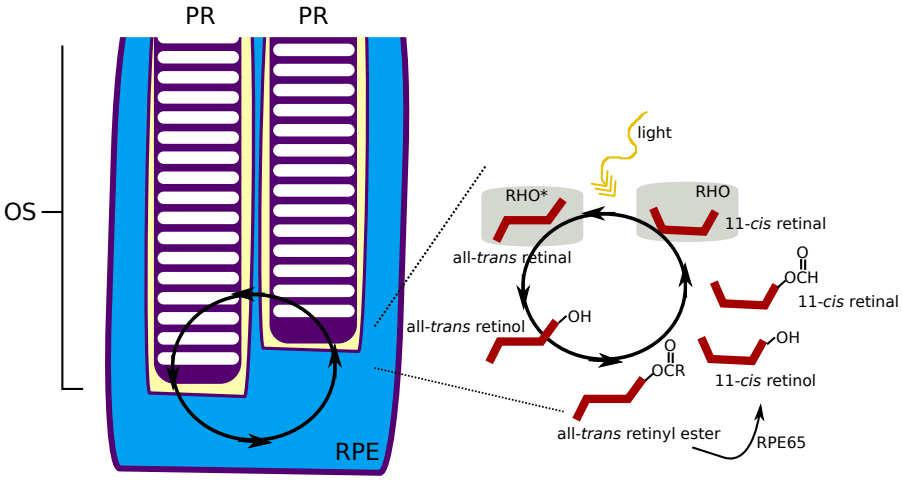

Figure 5: The outer segment of the photoreceptore cells.
The outer segments (OS) are adjacent to the retinal pigment epithelium (RPE). The light-sensitive chromophore 11-cis retinal is covalent bound to opsin in the OS. Rhodopsin is activated by light, which trigger isomerization of 11-*cis* retinal inducing a signal cascade. The resulting all-*trans* retinal gets released from the rhodopsin and then it is transported to the RPE. In the RPE it is modified to all-*trans* retinol, while it result in 11-cis retinal by two following enzymatic catalyzed steps requiring RPE65 and a dehydrogenase.

1.3 AIM

The aim of the present study was to characterize specific nucleases with a focus on the site- and strand-specific nucleases TALE-MutH as well as the binding domain AvrBs3.

We claim to analyze the mechanism of nicking and double nicking by TALE-MutH *in vitro* as well as its enzymatic activity under *in vivo* conditions using a Single-Strand-Annealing (SSA) assay in *Saccharomyces cerevisiae*. In a comprehensive study the characteristics of the monomeric nickase TALE-MutH, I-SceI and the widely used TALE-FokI will be analyzed. Additionally, the characteristics of TALE-PvuII variants and the gRNA nuclease under *in vivo* condition employing the SSA assay in *S. cerevisiae* should be defined and compared to I-SceI. The most efficient specific nucleases will be challenged by targeting a relevant therapeutic gene, such as the RPGR gene.

Furthermore, the thesis focused on the binding domain AvrBs3 and aims to define the binding mechanism of TALE. There are several studies characterizing the binding mechanism of TALE, while the exact binding mechanism still remains to be elucidated. A hindrance to investigate the binding mechanism are the lacking or restricted methods. Here, we claim to design an AvrBs3 which enable fluorescent labeling at specific positions, while retaining the binding characteristics. In 2015 AvrBs3 was labeled with Cy3 via an unnatural aldehyde and for the first time the *in vitro* binding character of AvrBs3 was described with single molecule studies [51]. In the present study, we want to confirm the findings of Cuculis *et al.* [51] with a focus on the conformation change between a non-specific TALE::DNA complex and a specific TALE::DNA complex. This should be achieved by Förster resonance energy transfer (FRET) experiments, which depend on the distances between donor fluorophore and acceptor fluorophore. Two single labeled AvrBs3 variants should be designed, which allows labeling of AvrBs3 specific and close to the N-terminus region or close to the C-terminus region. We were aspired to design an AvrBs3 labeled at two specific positions enabling to follow conformational changes within the protein.

Part II

MATERIAL AND METHODS

2 MATERIAL AND METHODS

2.1 MATERIALS

2.1.1 *Chemical*

Chemicals were applied from Merk, Roth, Sigma-Aldrich, AppliChem or Qiagen and were diluted in molecular biology grade water.

2.1.2 *Commercial DNA purification kits*

Kits used for DNA purification are listed in table 1.

Table 1: Commercial purification kits for DNA

Commercial purification kit	Provider
Wizard R SV Gel and PCR Clean – up System	Promega
PureYieldTM Plasmid Miniprep System	
PureYieldTM Plasmid Midiprep System	

2.1.3 *Columns/matrices*

By the principle of affinity chromatography the recombinant proteins were purified above Strep-, His- beads and/or Heparin columns. Labeled recombinant proteins were purified by the principle of size – exclusion chromatography with Zeba Desalting spin columns. Table 2 gives the employed column and matrices.

Table 2: Columns and matrices

Column / matrix	Provider
Ni-NTA agarose	Macherey - Nagel
Strep-Tactin R Sepharose R	IBA
HiTrap Heparin HP	GE Healthcare Life Science
Zeba Desalting spin columns (40K)	Pierce

2.1.4 *Dialyse tube*

For dialyse the dialysis tube Visking type 8/32 (Roth) was used. The semi-permeable membrane has an exclusion size of 14 kDa. The tubes were boiled in ultra–pure water and stored in 20% ethanol at 4°C. Before use the tubes were washed with ultra-pure water and dialyse buffer.

2.1.5 *Buffers and media*

In this section, all employed buffers and media are listed. Buffers and media were prepared with ultra-pure water from Mili-Q A10 ultra-pure water system (Milipore).

Buffer used for activity assays, for labeling reaction, as well as for anisotropy and FRET experiments are given in table 3.

Table 3: Buffers for assays

BUFFER	COMPOUNDS
low salt activity buffer	20 mM tris - acetate
50 mM potassium acetate	pH 7.9
	50 mM potassium acetate
	10 mM magnesium acetate
	1 mM DTT
labeling buffer	10 mM HEPES-KOH
	1 mM EDTA
	200 mM potassium chloride
	10% Glycerin
physiological activity buffer	20 mM tris - acetate
e	pH 7.5
	120 mM potassium acetate
	1 mM magnesium acetate
	0.1 M DTT
1x NEB 3.1 buffer	50 mM Tris-HCl
	pH 7.9
	100 mM NaCl
	10 mM MgCl2
	100 ug/ml BSA

To identify, quantify, and purify nucleic acid fragments, DNA electrophoresis was used. The employed buffers are recorded in table 4.

Table 4: Buffers for DNA electrophoresis

Buffer	Components
10 x Tris-phosphate-EDTA (TPE) buffer	0.9 M Tris, pH 8.2 phosphoric acid 20 mM EDTA
10 x Tris-borate-EDTA (TBE) buffer	1 M Tris pH 8.3 1 M boric acid 25 mM EDTA
50 x Tris-acetic acid-EDTA (TAE) buffer	2 M Tris pH 8.0 1 M sodium acetate 50 mM EDTA
10 x Laemmli-SDS-electrophoresis-buffer	250 mM Tris pH 8.3 1.9 M glycine SDS 1% (w/v)
5 x AAP agarose loading buffer	250 mM EDTA pH 8.0 sucrose 25% (w/v) SDS 1.2% (w/v) bromphenol blue 0.1% (w/v)
6 x SDS loading buffer	420 mM Tris-HCl pH 6.8 0.1% SDS 2-mercaptoethanol 7.2 (v/v) glycerol 36% (v/v) bromphenol blue 0.06% (w/v)
Coomassie staining solution	coomassie brilliant blue G250 0.1% (w/v) phosphoric acid 2% (v/v) aluminum sulfate 5% (w/v) ethanol 10% (v/v)

Table 5 provides buffers used for working with *Saccharomyces cerevisiae*.

Table 5: Buffers for *Saccharomyces cerevisiae*

BUFFER	COMPONENTS
transformation mix	0.1 M lithium acetate 10 mM tris-HCl 1 mM EDTA polyethylene glycol (PEG 3350) 40% (v/v)
LTE buffer	0.1 M lithium acetate 10 mM tris-HCL 1 mM EDTA
10 x BU salt	250 mM NaH_2PO_4 pH 7 195 mM Na_2HPO4 7 H_2O
10 x dropout solution missing appropriate amino acid	200 mg/L L-Adenine hemisulfate salt 200 mg/L L-Arginine HCl 200 mg/L L-Histidine HCl monohydrate 300 mg/L L-Isoleucine 1000 mg/L L-Leucine 300 mg/L L-Lysine HCl 200 mg/L L-Methionine 500 mg/L L-Phenylalanine 2000 mg/L L-Threonine 200 mg/L L-Tryptophan 300 mg/L L-Tyrosine 200 mg/L L-Uracil 1500 mg/LL-Valine

Media used for *E. coli* and *S. cerevisiae* are given in table 6.

Table 6: Media for *E. coli* and *Saccharomyces cerevisiae*

MEDIA	COMPOUNDS
Lysogeny broth (LB) media	trypton 1% (w/v)
	yeat extract 0.5% (w/v)
	NaCl 0.5% (w/v)
	pH 7.5
SOC media	bacto tryptone 2% (w/v)
	yeast extract 0.5% (w/v)
	10 mM NaCl
	0.5 mM potassium chloride
	10 mM $MgCl_2$
	10 mM $MgSO_4$
	20 mM glucose
	pH 6.5
YPAD media	yeast extract 1% (w/v)
	bacto pentone 2% (w/v)
	dextrose 2% (w/v)
	bacto agar 2% (w/v)
	adenine hemisulfate salt 7.5 x 10^{-3}% (w/v)

To grow *E. coli* colonies LB-plates containing antibiotics were used. For *S. cerevisiae* drop out plates were employed(Table 7).

Table 7: LB-plates and drop out plates

plate	compounds
LB-plate	LB-media agar 0.015 (w/v)
drop out plate (-HIS/-TRP)	yeast nitrogen base without aa 0.67% (w/v) agar 0.0 % (w/v) 1 x dropout solution (-HIS/-TRP) dextrose 2% (w/v) 0,08 g/ml X-GAL 1 x BU salt buffer
drop out plate (-His)	yeast nitrogen base without aa 0.67% (w/v) agar 0.03% (w/v) 1 x dropout solution (-HIS) dextrose 2% (w/v) 1 x BU salt buffer

2.1.6 *Fluorophores*

Several fluorophores were used to either label target DNA or different variants of AvrBs3 proteins and are listed in table 8.

Table 8: Physical properties of fluorophores used for labeling DNA and proteins. Maleimide-fluorophores were utilized for protein labeling and reacted with the cysteine residues of the AvrBs3 variants. $A_{F,max.}$ and $E_{F,\,max.}$: Absorbance and fluorescence emission maxima in nm; * correction factor of fluorophores at 280 nm and 260 nm.

Physical properties	Cy3	Cy5	HEX	Alexa594	Alexa488	Alexa647
$A_{F,max.}$[nm]	550	649	535	590	495	650
$E_{F,\,max.}$ [nm]	570	670	560	617	519	665
$\varepsilon_{max}[cm^{-1}M^{-1}]$	150,000	250,000	96,000	73,000	71,000	239,000
cf_{280}*	-	0.05	-	0.56	0.11	0.030
cf_{260}*	0.08	-	0.33	0.43	0.30	0.00

In figure 6 the chemical structure of the cyanine fluorophores Cy3 (I) and Cy5 (I) are given, as well as the chemical structure of the 6-carboxy-2 ,4,4 ,5 ,7,7 -hexachlorofluorescein succinimidyl ester HEX (II) fluorophore. Cy3 and HEX fluorophores were labeled to the DNA and Cy5 maleimide was used to label the single cysteine AvrBs3 variants.

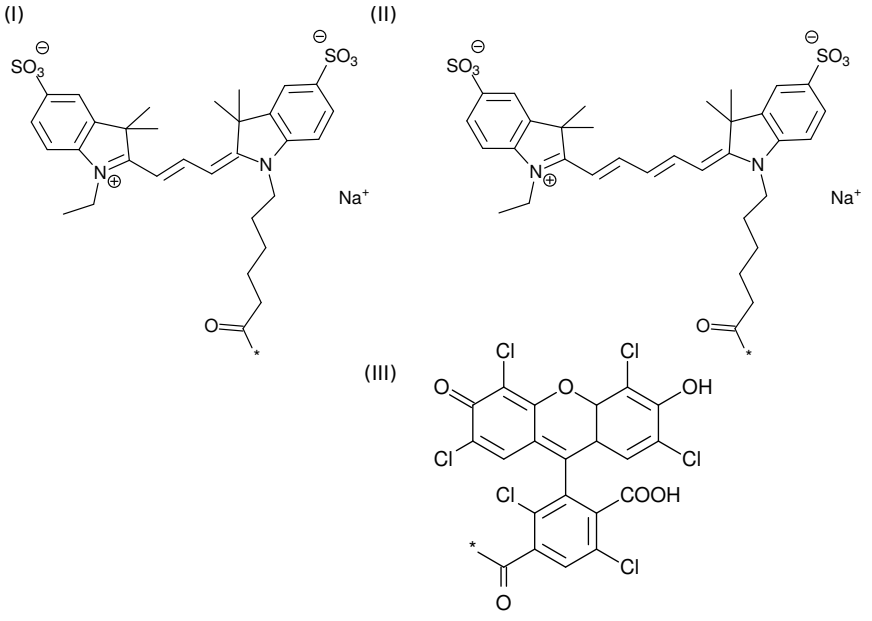

Figure 6: Chemical structures of fluorophores
Chemical structure of the fluorophores Cy3 (I), Cy5 (II) and HEX (III).

The chemical structures of the fluorophores used for labeling the single cysteine variants, as well as the double cysteine AvrBs3 variants are given in figure 7. All fluorophores used to label the protein contained a maleimide functional group.

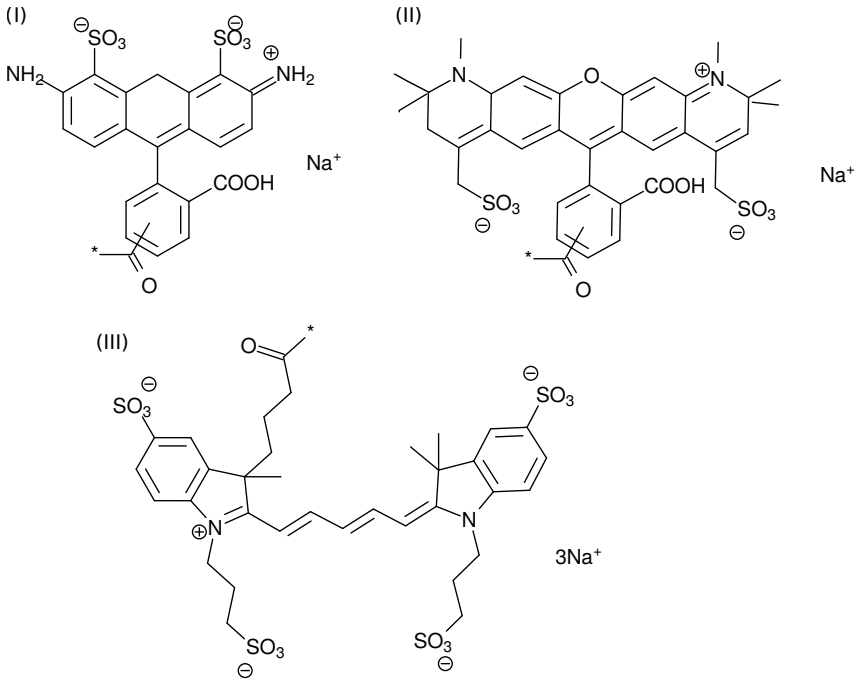

Figure 7: Chemical structures of fluorophores
Chemical structure of Alexa 488 (I), Alexa 594(II), and Alexa 647(III).

2.1.7 *Plasmids*

Plasmids were used to construct the binding domains TALE or as expression and substrate vector.

2.1.7.1 Binding domains TALE

Binding domains TALE were cloned by the Golden Gate TALEN and TAL effector kit 2.0. This plasmid Kit was a gift from Daniel Voytas and Adam Bogdanove (Addgene kit # 1000000024) [33]. It contains all required plasmid to construct a TALE binding domain.

2.1.7.2 Expression vectors

pqe – 30 vector The pQE-30 vectors can be used for protein expression in all ampicillin – sensitive *E. coli* strains containing pREP4 repressor or lacZ gene on the F-factor episome. The vector harbours a T5 – promoter, which is under control of the *lac-repressor*. The N-terminal His-tag was sub cloned to the C-terminus coding sequence and a Strep-tag was added to the N-terminal coding sequence.

PQE-30_STREP -AVRBS3(N152-C28) contains the truncated version of AvrBs3 (ΔN152-C28).

PQE-30_STREP -AVRBS4 (28)-MUTH-HIS was a gift from Gabsalilow *et al.* [78]. It expresses the fusion construct Strep-AvrBs4-MutH-His. AvrBs4 is truncated at the N-terminus and the C-terminus (ΔN152-C28). The linker between the repeat array of AvrBs4 and MutH consist of 28 amino acids behind the last half repeat of AvrBs4.

PZHY500 VECTOR The shuttle vector pZHY500 was constructed by Daniel Voytas (Addgene plasmid # 36186) as compatible yeast expression vector for TALENs [246]. It harbours a truncated TALEN architecture [246] and is a GOLDEN GATE compatible expression vector [33]. The N-terminus and C-terminus are truncated (ΔN152-C68) and the repeat array is replaced by a lacZ-gene. The homodimeric FokI nuclease is fused direct to the C-terminus. pZHY500 contains a HIS3 selectable marker, as well as ampicilline resistance. For expression in yeast a TEV-promoter is used.

CAS9 SYSTEM:

P414-TEF1P-CAS9-CYC1T For using the CRISPR-Cas system in Saccharomyces cerevisiae the Cas9 protein was expressed from the plasmid p414-TEF1p-Cas9-CYC1t, a gift of George Church [58] (Addgene plasmid # 43802). It contains the selective marker TRP1.

PRPR1_GRNA_HANDLE_RPR1T The guided RNA was constructed starting with the plasmid pRPR1_gRNA_handle_RPR1t from Lu T. [69] (Addgene plasmid # 49014).

PVAX This vector can be used for the expression of recombinant proteins in eukaryotic cells. It harbours a kanamycin resistance gene and expression is under control of a CMV promoter.

2.1.7.3 Substrate vectors

PAT153 was used as a substrate vector. It contains 19 unaddressed MutH recognition sites and an ampicillin resistance-gene.

PAT153_T-3-H/PAT153_T-6-H were made by Gabsalilow *et al. [78]* for the analysis of the *in vitro* activity and specificity of the AvrBs4_MutH protein. The vectors harbours the 18 bp AvrBs4 recognition site with different spacer length to the recognition site of MutH (Table 9).

PAT153_T3-4-P-4-T3 were made by Yanik *et al. Yanik et al. [239]* for the analysis of the *in vitro* activity and specificity of the AvrBs3-PvuII variants. In table 9 the inserted target site is listed.

Table 9: Addressed target sites
Specific sequences of the target site of AvrBs4 (T), AvrBs3(T3), MutH(H), and PvuII(P) are underlined. The number between T and H represents the spacer length in basepairs.

Target site	Abbreviation
5'-TGCAGTATAATTAATAATCCACTTGAGGATCCTCGCTACATG-3'	T-3-H
'5'-TGCAGTATAATTAATAATCCACTTGAGCCGGATCCTCGCTACATG-3'	T-6-H
'5- TCTATAAACCTAACCCTCTATCCCAGCTGTCTAAGAGGGTTAGGTTTATAGA-3'	T3-4-P-4-T3

PCP5 VECTOR from Daniel Voytas (Addgene plasmid # 35397) [33], a *E. coli*/yeast shuttle vector, was used as a reporter plasmid in yeast. It harbours a URA3 selective maker, which is flanked by a divided lacZ-gene which contains 125 bp duplication. Additionally, it encompasses a TRP1 selective marker, kanamycin resistance and chloramphenicol resistance. The pCP5 vector uses a GDP promoter. A ccdB lethal gene, which is used as a selection marker, is also integrated in between the divided lacZ gene.

2.1.8 *E. coli and* Saccharomyces cerevisiae *strains*

E. coli

XL10 – GOLD STRAIN was used for the expression of pQE_Strep-AvrBs4-MutH-His and pQE_Strep-AvrBs3, as well as for several cloning steps of pZHY500.

Genotype: endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tetR F'[proAB lacIqZΔM15 Tn10(TetR Amy CmR)]

DH5α STRAIN was mainly used to sub clone the reporter plasmid pCp5.

Genotype: F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK- mK+), λ–

JM109 STRAIN was used for cloning the TALE binding domain by employing the GOLDEN GATE strategy.

Genotype: endA1 glnV44 thi-1 relA1 gyrA96 recA1 mcrB+ Δ(lac-proAB) e14- [F' traD36 proAB+ lacIq lacZΔM15] hsdR17(rK-mK+)

JM110 STRAIN was used to multiply the substrates inserted in pAT153. JM110 lacks methylase dam and dcm and therefore produces non methylated DNA.

Genotype: rpsL thr leu thi lacY galK galT ara tonA tsx dam dcm glnV44 Δ(lac-proAB) e14- [F' traD36 proAB+ lacIq lacZΔM15] hsdR17(rK-mK+)

Saccharomyces cerevisiae

YPH500 YEAST STRAIN was used for the yeast activity assay, which is describe in 2.2.. Genotype: MATα ura3-52 lys2-801_amber ade2-101_ochre trp1-Δ63 his3-Δ200 leu2-Δ1

2.1.9 *Enzymes and markers*

Enzymes used were provided by *Thermo Scientific Fermentas* or *New England Biolabs (NEB).* Markers for gel electrophoresis are listed in table 10.

Table 10: Maker for gel electrophorese

Marker	Provider
PageRulerTM Unstained Protein Ladder	Thermo Scientific Fermentas
GeneRuler^{TM} 1kb DNA ladder	
pUC Mix Marker, 8, ready to use	

2.1.10 *Primer and oligonucleotide*

Cassettes used as inserts are summarized in table 11.

Table 11: Cassettes used as inserts
Specific sequences of the target site of TALE(T) and MutH(H) are underlined. The number between T and H represents the spacer length in basepairs.

Sequences of cassettes	Abbreviation
.....5‘-TCTCAGCGAGGATCCGGAAGTGGATTATTAATTATACTGCACATG-3‘ 3‘-CATGAGAGTCGCTCCTAGGCCTTCACCTAATAATTAATATGACGT-5‘	H-3-T
.....5‘-TGCAGTATAATTAATAATCCACTTCCGGATCGATCCGGAAGTGGATTATTAATTATACTGCACATG-3‘ 3´-CATGACGTCATATTAATTATTAGGTGAAGGCCTAGGCCTTCACCTAATAATTAATATGACGT-5‘	T-3-HH-3-T
.....5‘-TGCAGTATAATTAATAATCCACTTCCGGATCCGGAAGTGGATTATTAATTATACTGCACATG-3‘ 3´-CATGACGTCATATTAATTATTAGGTGAAGGCCTAGGCCTTCACCTAATAATTAATATGACGT-5‘	T-3-H-3-T
5‘- CTAGGCGGTACCTAATAGTAGCT - 3‘ .....3‘- CGCCATGGATTATCA - 5	stop codon
5‘- GCTGCGATTGGCAGAAGCTTGGATAACAGGGTAATATAGCGTTTTAGAGCTAGAAATAGC-3‘ 5‘- CGACGCTAACCGTCTTCGAACCTATTGTCCCATTATATCGCAAAATCTCGATCTTTATCG-3‘	insert for gRNA

Cassettes used as target DNA for anisotropy and FRET experiments are given in table 12.

Table 12: Cassettes used as target DNA
The specific DNA contains the AvrBs3 binding site (underlined) and is either labelled with Hex at the 5´ -end of the bottom strand or does not posses a fluorophore label. For fluorecence based experiments, specific DNA was labelled with Cy3 at the 5´ -end of the top or bottom strand. The scramble DNA consists of the same nucleotide composition, but in a random order.

DNA	Name	Length [bp]	Fluorophores
5‘-TTTATCTATAAACCTAACCCTCTCT-‘3 3‘-AAATAGATATTTGGATTGGGAGAGA-‘5	specific DNA	25	HEX or Cy3
5‘-ACACCTCTACCTATCTTATATCTTA-‘3 3‘-TGTGGAGATGGATAGAATATAGAAT-‘5	scramble DNA	25	-

Primer used to create magaprimer for Mutagenesis-PCR are listed in table 13.

Table 13: Primers used to create Megaprimers
The first four oligos were used to create megaprimer, which induces a AvrII cutting site to the plasmid, respectively. All the rest were used to create megaprimers to substitute C (cysteine) to M (methonine) in the given plasmids.

Sequence	Number	
5‘-GAGCTCCCTAGGCAGGCCAAGGCGACGAGG-3‘	558	558/455 (AvrII site in stuffer-MutH)
5‘-catcgcgcaatgcactgac-3‘	455	
5‘-CTGCAGGTCGACGGATCCTCCCAGACCCGGTTTAC -3‘	423	423/455 (AvrII site in pTal1)
5‘-CAGGACGTCCGCCTAGGCAGGCCAAGGCGAC-3‘	523	455/523 (AvrII site in pZHY500)
		C to M substitution
5‘-GCGGCTGTTGCCAGTACTGATGCAGGACCATGG-3‘	499	499/498 (pHD1,8,9; pNI1,6,7; NG10)
5‘-CGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAG-3‘	498	
5‘-GCTGTTGCCAGTACTGATGCAGGGAGACCCTCGAGC-3‘	501	501/498 (pNI3, pHD3)
5‘-TCCTCTAGAGGTCTCGGCAGGATCATGGCCTGACC-3‘	502	502/498 (pNG4)
5‘-CTCCTGTTGCCAGTACTGATGCAGGACCATGGCCTG-3‘	503	503/498 (pNI5, pHD5)
5‘-CCTCTAGAGGTCTCGGCAGGATCATGGCCTGACTC-3‘	695	695/498 (pHD4)
5‘-TAATACGACTCACTATAGGG-3‘	41	41/686 pTal1(AvrII)
5‘-CAGGACGTCCGCCTAGGCTGGCCAAGGCGACGA-3‘	696	

Primer used for screening and sequencing are shown in table 14.

Table 14: Primers
Oligos used for screening

Sequence	Number
5‘-GTATCACGAGGCCCTTTCGTCT-3‘	126 (for)
5‘-CATTACTGGATCTATCAACAGGAG -3‘	68 (rev)
5‘-catcgcgcaatgcactgac-3‘	455(for)
5‘-ggcgacgaggtggtcgttgg-3‘	454(rev)
5‘-ttggcgtcggcaaacagtgg -3‘	453(for)
5‘-GGAGAAAATACCGCATCAG -3‘	477(rev)
5‘-CGGTTGCGCCACCCTGTCTCCCTTGC-3‘	MutHseq4 (rev)
5‘-TCATCGCAAGACCGGCAACAGGAT -3‘	577 (rev)
5‘-CAACTTTCCACCATTGCCTTGAAAAATATCATTTA -3‘	457 (rev)
5‘-TCCCCAGAAGCTACTCATGAAGCTATTGTTG -3‘	564 (rev)
5‘-ATCCTGCACCATCGTCTGCT -3‘	476 (rev)
5‘-ATGCACTAGATCTGGCGACCACACCCGTCAG -3‘	473 (for)
5‘-CCCAGTACTAGTGAGGCCGTTGAGCACCGC -3‘	474 (rev)
5‘-TTTATCTATAAACCTAACCCTCTGGATCCCAGCTGTCTAGACATG -3‘	339 (for)
5‘-GGATCCCAGCTGTCTAGAAGAGGGTTAGGTTTATAGATAAACATG -3‘	337 (rev)
5‘-TGACCTCCCATTGATATTTAAGTTAATAAACG -3‘	487 (for)
5‘-CGTTCTCGGAGCACTGTCCGACCGCTTTGG-3‘	lab S/H (for)
5‘-GCACCGCCGCCGCAAGGAATGGTGC-3‘	t4-x-H (rev)

2.2 METHODS

Microbiological standard methods (producing competent cells, transformation of plasmids in *E. coli* cells or *Saccharomyces cerevisiae* cells etc.) and microbiological standard methods (agorse- and polyacrylamide-electrophoresis, DNA isolation, restriction digest etc.) were performed as described in *Current Protocols in Molecular Biology* and *Protein Science* [10, 46].

2.2.1 *Molecular cloning*

Molecular cloning is a technique to assemble recombinant DNA, which is then replicated in a host organism. Recombinant DNA was introduced to the host bacterium *E. coli*. In figure 8 the basic principles of molecular cloning are shown. First, an insert has to be created, which is then ligated into a plasmid. Inserts can be constructed by cassettes, PCR amplification or digestion. Cassettes used as inserts for cloning strategies are listed in table 11. These oligos were mixed in a 1:1 ratio, heated at 95°C for 3 minutes and then slowly cooled down to RT. The remaining inserts were processed either by PCR amplification of the gene of interest or by digestion of a plasmid containing the gene of interest. The amplified PCR fragments were digested by the required restriction enzymes. Receiving the insert through digestion, the DNA-fragments were separated by gel electrophoresis and the fragment were purified by the Wizard R SV Gel and PCR Clean–up System (Promega). The needed backbones were prepared by digesting the plasmid with restriction enzymes and dephosphorylation of the DNA ends by Fast AP (37°C, 15 min.). Digestion of DNA fragments or plasmids were preformed at 37° C for 1 hour. For ligation 5 nM insert, 3 nM backbone DNA, 1 x T4-ligase buffer and 1 µl T4–ligase, as suggested by the manufacture, were mixed in 20 µl. Ligation were performed at 22°C for 1 h or 16°C overnight. Z–competent cells were prepared for transformation by the Z-competent *E. coli* Transformation Kit & Buffer Set (Zymo Research) and stored at -70°C. For transformation cells were allowed to defrosted for 10 minutes on ice and 3-5 µl ligation–mixture was added to the cell aliquot. Further, the cell aliquots stayed on ice over 5–10 minutes and 400 µl SOC media was added. Cells were permitted to recover for about an hour at 37°C before distributing the cell aliquot on the essential LB–plate (37°C, o/n). Figure 8 depicts the key steps of the cloning strategies.

MUTAGENESIS-PCR [118] was used to substitute precise amino acids. First, a mega primer was amplified by PCR using primers including the desired mutation. Next, the mega primer was then used for a rolling circle PCR (Table 16), which amplified the complete vector, containing the new mutation.

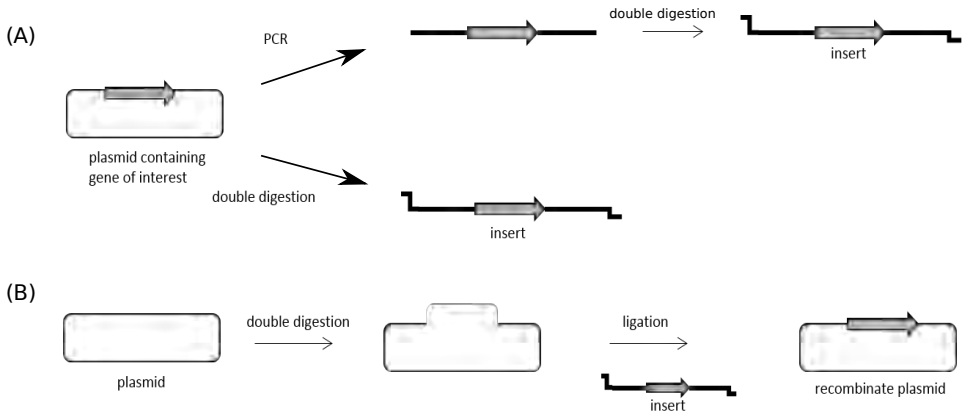

Figure 8: Basic steps of the cloning strategies
(A) Insert made by PCR amplification, followed by double digestion or by immediate double digestion of the plasmid. (B) The backbone was obtained by double digestion of the starting plasmid. The desired recombinant plasmid was gained by ligation.

Table 15: Components used for PCR are listed.

Components
100 ng DNA-matrix
400 nM 5'-primer
400 nM 3'-primer
200 µM dNTPs
1x Pfu-buffer
1 µl Pfu-polymerase

For exact amplification the pfu–polymerase of Pyrococcus furiosus (3'- 5' Exonucleaseactivity, $T_{opt.}$ = 68°C, ~ 500 nucleotide per minutes) was used. To digest the template vector the mixture was digested by DpnI, an enzyme which only cuts at methylated sites.

BLUE - WHITE SCREENING Blue -white screening was used to simplify the cloning procedure, but the same principle was also applied in the single strand annealing assay, as described later. The method is based on an enzyme, called ß-galactosidase. This enzyme is able to convert 5- bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X–Gal) to a blue dye (figure 9). ß-galactosidase is encoded by the lacZ gene of the lac operon and functions as a homotetramer. *E. coli* strains carrying the lacZ deletion mutant (lacZΔM15) express a ß-galactosidase (ω-peptide), which misses its N-terminal residues, resulting in a non functional enzyme. If the *E. coli* strain carries a plasmid, which encodes the first 59 residues (α-peptide) of ß-galactosidase the ω-peptide and the α-peptide reconstitute a functional ß-galactosidase enzyme.

Table 16: PCR-programs

PCR-programs applied for this thesis are described. The basic program is used to amplify DNA fragments. $T_{annealing}$ depend on length and composition of the primer and T_{opt}depends on the employed polymerase (e.g. Taq-polymerase $T_{opt.}$= 72°C). The rolling circle program is used for mutagenesis-PCR. Prior to that, a megaprimer is processed, which contains the relevant mutation.

basic program			rolling circle program		
temperature [C°]	time [min]		temperature [C°]	time [min]	
94	2		94	2	
92	0.75		94	0.75	
$T_{annealing}$~ 58-62	0.50		60	0.50	
$T_{opt.}$	500-1000 nt/min	go to step one and repeat 30 x	68	15	go to step one and repeat 20 x
$T_{opt.}$	5		68	20	
4	break		4	break	

The LB-plates have to be spread with 40 µl 20 mg/ml X-Gal and 40 µl of a 0.1 M Isopropyl β-D-1-thiogalactopyranoside (IPTG). IPTG is needed to inactivate the LacZ repressor in *E. coli*. The colonies growing on the plate will either be white or blue, depending on the existence of the α-peptide encoding plasmid.

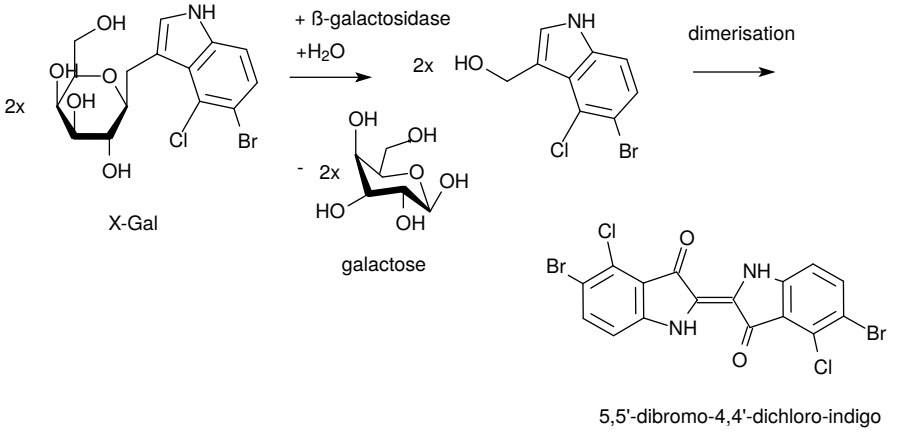

Figure 9: X-Gal converted to a blue dye

ß-galactosidase hydrolyse X-Gal to galactose and 5-bromo-4-chloro-3-hydroxyindole. Due to spontaneous dimerisation followed by oxidation, a 5,5′-dibromo-4,4′-dichloro-indigo is yielded and responsible for the blue color.

2.2.1.1 *Golden Gate*

The TALE binding domains were designed with the GOLDEN GATE TALEN and TAL Effector Kit 2.0 from Daniel Voytas and Adam Bogdanove (Addgene Kit#1000000024) [33]. The kit consists of bacterial glycerol stocks enclosing the required plasmids to construct a TALE binding domain. The kit enables to design TALE binding domains enclosing 12-31 repeats. The kit contains various plasmids. The backbone scaffold (pTAL1-4) contains the TAL-effector gene. Here the repeat array is replaced with a lacZ gene, which functions as stuffer. The lacZ gene enabled blue-white screening.

Plasmids containing the encoding sequence for one repeat (e.g. pNG1) and a tetracycline resistence and plasmids carrying the encoding sequence for a last half repeat (e.g. pLR-HD) with a spectinomycin resistence are included. Plasmids named, pFusA(1-10) or pFusB(1-10) can assemble up to 10 repeats in one plasmid and enclose a spectinomycin resistence-gene.

The principle of GOLDEN GATE cloning strategy is to use typeIIs restriction enzymes, which induce DSBs outside of their recognition sites. Due to that, sticky ends emerge which can exclusively be ligated to a compatible other sticky end. The kit enables to assemble ten repeats into one back bone in a one pot reaction. In figure 10, the principle is illustrated. The protocol of the the GOLDEN GATE TALEN and TAL Effector Kit 2.0 take five days to construct one binding domain and we followed the protocol exactly as recommended by the manufactures.

All constructed TALE binding arrays were assembled into a modified pTAL1 vector, which contained a AvrII cutting site. This enabled us to insert the TALE binding arrays into the specific expression vector. We used modified versions of pZHY500[246] for yeast expression, pVAX for expression in HEK297 cells and pQE-30 vectors for expression in *E. coli*. In table 17 the repeat composition of the designed TALE binding domains are given.

2.2.2 *vectors*

EXPRESSION VECTORS The expression vectors were modified for the insertion of the TALE binding domains. An AvrII cutting site was introduced by MUTAGENESIS-PCR to pTAL1 and pZHY500. This enabled us to remove the TALE binding domain from the pTAL1 vector with the restriction enzymes AvrII and StuI and to insert the repeat array into the modified pZHY500 expression vector.

YEAST EXPRESSION VECTOR PZHY500_AVRBS4-FOKI: The sequence encoding AvrBs4 was inserted to create the coding sequence for the homodimeric AvrBs4-FokI.

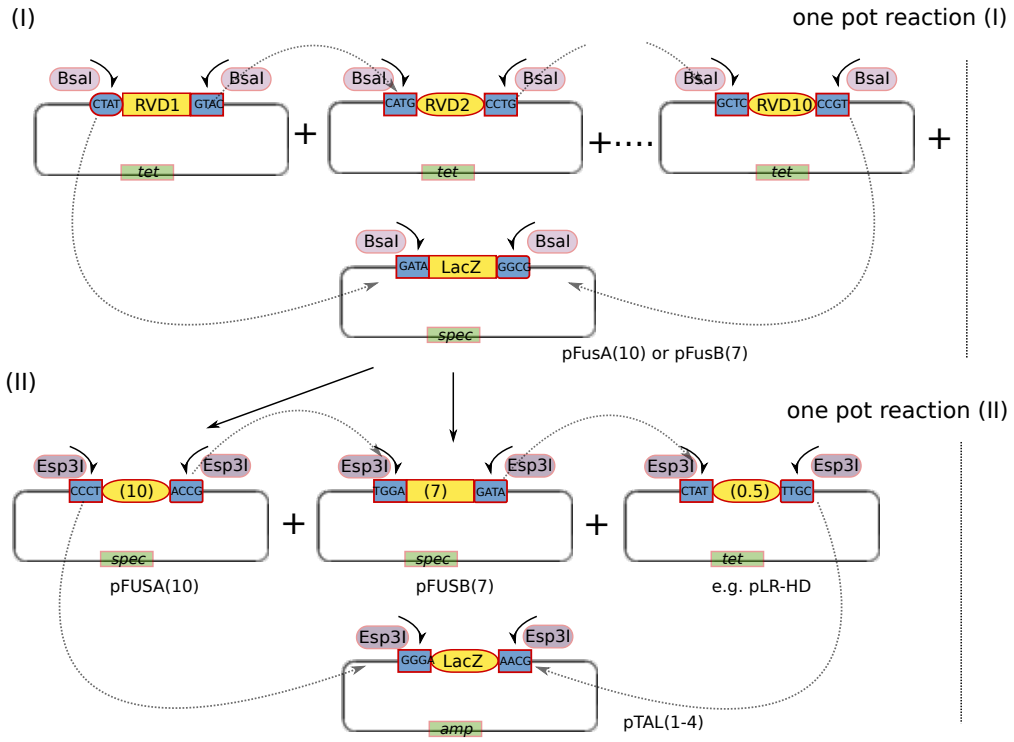

Figure 10: Golden Gate cloning strategy
(I) The first one pot reaction can assemble up to ten repeats into the pFusA(1-10) or pFusB(1-10) backbone. Each repeat is cut out by the restriction enzyme BasI, which induces a DSB outside of the recognition site. Due to the resulting sticky ends, the fragments are ligated to each other in a specific order. (II) The second one pot reaction assembles pFusA(10), pFusB(7), pLR-HD, and pTAL1 by the same principle, but using the restriction enzyme Esp3I.

Table 17: Repeat composition assembled with *the* Golden Gate cloning strategy
Two naturally occurring TALE binding domains were designed (AvrBs4 and AvrBs3). Additionally, five binding domains addressing a GATC site (I) and (II) in the RPGR-gene were constructed.

Repeat composition	Name
NI1 NG2 Ni3 Ni4 NG5 NG6 NI7 Ni8 NG9 NI10 NI1 NG2 HD3 HD4 NI5 HD6 NG7 NG	AvrBs4
HD1 NG2 NS3 NG4 NI5 NI6 NI7 HD8 HD9 NG10 NS1 NS2 HD3 HD4 HD5 NG6 HD7 NG	AvrBs3
	TALE(RPGR)
HD1 NI2 NH3 NH4 NH5 NH6 NH7 NH8 NH9 NI10 NH1 NI2 NH3 NI4 Nh5 NI6 NH7 NI	1(I)
NG1 NG2 NH3 NI4 NG5 HD6 HD7 HD8 HD9 NH10 NG1 HD2 NI3 HD4 HD5 HD	2(I)
HD1 NI2 HD3 NI4 NH5 Ni6 NG7 NG8 NG9 NH10 NI1 NG2 HD3 HD4 HD5 NH6 NG7 HD8 NO	3(I)
HD1 NG2 NI3 NN4 NI5 HD6 NG7 HD8 NI9 HD10 HD1 NG2 NG3 HD4 NG5 HD6 NI7 HD8 NI9 NN10 NI	1(II)
NH1 NG2 NG3 NG4 HD5 NG6 NI7 NI8 NI9 HD10 NG1 NH2 NI3 HD4 HD5 NG6 NG7 HD8 NG9 HD10 NI1 HD	2(II)

PZHY500_AVRBS4: A negative control vector (AvrBs4-stop) was designed by replacing the homodimeric FokI cleavage domain by the stop codon cassette (5'- CTAGGCGGTACCTAATAGTAGCT – 3', 3'-GATCCGCCATGGATTATCATCGA-5') inserted in the AvrII and HindIII sites.

PZHY500_I-SCEI: The plasmid coding for I-SceI was created by amplifying the I-SceI gene using pASK_His-MutH-I-SceI-Strep as a template and inserting the PCR fragment into the vector pZHY500 following he cutting with XbaI and SacI.

PZHY500_AVRBS4-MUTH: To construct an analogue yeast expression vector for AvrBs4-MutH the lacZ-gene was amplified by PCR and inserted by StuI and SacI as a stuffer replacing TALE in the *E. coli* expression vector pQE30_Strep-TALE-MutH-His6 [78]. Next, stuffer-MutH was amplified and integrated into pZHY500 using StuI and Ecl136, resulting in pZHY500_stuffer-MutH. An AvrII site was introduced by mutagenesis PCR and the stuffer (lacZ-gene) was replaced by the encoding sequence of AvrBs4.

PZHY500_AVRBS3-PVUII (VARIANTS): To construct analogue yeast expression vectors for AvrBs3-PvuII variants, the MutH-encoding gene in pZHY500_lacZ-MutH was replaced by the respective PvuII (hfPvuII, scPvuII, D30KPvuII, and K38DPvuII) encoding gene using AvrII and Ecl136 sites. The stuffer (lacZ-gene) was replaced by the encoding sequence of AvrBs3 using the AvrII and the StuI site.

HEK297 EXPRESSION VECTOR PVAX_TALE(RPGR): All TALE arrays addressing the RPGR-gene were inserted into pVAX_stuffer-MutH. The encoding sequence stuffer-MutH was gained as described above and inserted into the pVAX vector using StuI and Ecl136 sites. The requiered pVAX-TALE(RPGR) vectors were constructed by using the AvrII and StuI cutting sites.

Cysteine to methionine substitution in AvrBs3

Additionally, different variants of the TAL effector AvrBs3 were designed using the plasmid kit for the generation of TALENs as described by Daniel Voytas and Adam Bogdanove (Addgene kit # 1000000024)[33]. Cysteine residues at position 30 in all single 17 repeats were substituted by MUTAGENESIS PCR to methionine. Furthermore, the cysteine in the backbone pTAL1 was removed by mutagenesis PCR. An AvrII cutting site was introduced into the backbone pTAL1, to relocate AvrBs3 variants into an *E. coli* expression vector pQE30_Strep-AvrBs3.

Table 18: AvrBs3 variants
TALEs made with Golden Gate strategy. TALEs were cloned into pTAL1 + AvrII backbone. RVDs which are underlined include cysteine to methionine substitution. The last half repeat contains no cysteine. (I) represent AvrBs3 without any cysteine; (A) represent AvrBs3 (A) carry one cysteine residue in repeat domain one; (B) AvrBs3 (B) enclose one cysteine residue in repeat domain sixteen; (AB) AvrBs3 (AB) contain one cysteine residue in repeat domain one and one cysteine domain in repeat number sixteen.

repeat composition	name
HD1 NG2 NS3 NG4 NI5 NI6 NI7 HD8 HD9 NG10 NS1 NS2 HD3 HD4 HD5 NG6 HD7 NG	(I)
HD1 NG2 NS3 NG4 NI5 NI6 NI7 HD8 HD9 NG10 NS1 NS2 HD3 HD4 HD5 NG6 HD7 NG	(A)
HD1 NG2 NS3 NG4 NI5 NI6 NI7 HD8 HD9 NG10 NS1 NS2 HD3 HD4 HD5 NG6 HD7 NG	(B)
HD1 NG2 NS3 NG4 NI5 NI6 NI7 HD8 HD9 NG10 NS1 NS2 HD3 HD4 HD5 NG6 HD7 NG	(AB)

CRISPR-CAS SYSTEM: The guided RNA was constructed starting with the plasmid pRPR1_gRNA_handle_RPR1t from Lu T. [69] (Addgene plasmid # 49014). Using a one-step Gibson assembly reaction a 20 bp reporter (5'-GGATAACAGGGTAATATAGC-3') was introduced into the gRNA- expression plasmid. The target site addressed overlap with the I-SceI target site.

GIBSON ASSEMBLY The Gibson Assembly® Cloning Kit from NEB was used to design the gRNA. The HindIII site in pRPR1_gRNA_handle_RPR1t was employed to insert the specific sequence via an one-step Gibson assembly. The cassette used for insertion carried the 20 bp homologous sequence upstream and downstream of the HindIII site. In between the 20 bp duplications sequence, the 20 bp sequence addressing the Cas9 target site is located. HindIII digestion was followed by 5'-end resection, the homologous parts anneal with each other and the DNA polymerase extends the 3'-ends. Finally, the DNA ligase seals the nicks. The procedure was performed as recommended by the manufactures.

SUBSTRATES The substrate plasmids pat153_T-3-H-3-T, pat153_T-3-HH-3-T and pat153_H-3-T were constructed as has been previously described by Gabsalilow *et al.* [78]. The appropriate DNA-cassette enclosing the target site was inserted to pat153 by using the KpnI and SphI cutting sites.

To design and construct the *E. coli*/Yeast shuttle vectors the target sites of the different pat153 plasmids were amplified and inserted via the BglII and SpeI site into pCP5, a kind gift from Daniel Voytas [33] (Addgene plasmid # 35397).

The target site (pCP5_T3-4-P-4-T3) for the AvrBs3-PvuII variants was constructed similarly by Dr. Mert Yanik.

2.2.3 *Protein expression and purification*

AvrBs3-MutH

Strep-AvrBs4(28) –MutH-His (Figure 13), containing a Strep-tag and a His-tag, was purified by the tandem purification, recently described by Gabsalilow *et al.* [78]. *E. coli* strain XL-10 Gold was transformed with the expression plasmid. 25 ml LB-media containing ampicillin (75 µg/ml) was inoculated and incubated over night at 37°C. Next, 500 ml LB-media, containing ampicillin, was inoculated with the overnight culture and shaken at 37°C until an optical density (OD_{600}) of 0.7 -0.9 was reached. Expression was induced with 1 mM isopropyl-β-d-thiogalactopyranoside (IPTG). The LB-media was incubated over night at 20°C. The cell suspension was centrifuged (3600 x g, 4°C; 15 min). The supernatant was removed and cells were washed with 40 ml His-binding-/ washing buffer (20mM tris-HCl, 1 M NaCl, 20 mM imidazole, 1mM phenylmethanesulfonyl fluoride (PMSF), pH 7.9). Cell lysis was achieved by sonification (Duty Cycle 50 %, output Control 5; 12 x 15 sec) and cell extracts were centrifuged (31000 x g, 30 min at 4°C). Meanwhile Ni-NTA agarose (Qiagen) was equilibrated with 30 ml His - binding -/ washing buffer for 30 min at 4°C. The equilibrated Ni–NTA was incubated with the clean cell extract for 1 h at 4°C and washed twice afterwards with 40 ml His-binding-/ washing buffer. The Ni-NTA beads were transferred to a Poly–Prep chromatography column (Poly-Prep, BioRad)) and the protein was eluted with 3 x 1 ml elution buffer (20mM tris-HCl, 1 M NaCl, 200 mM imidazole, pH 7.9). The eluates of the His–purification were loaded on a column with Strep–Tactin sepharose (IBA). The Strep-Tactin spharose was washed with 4 ml Strep-binding-/ washing buffer (100 mM tris-HCl, 1 M NaCl, pH 7.9) and the protein was eluted with 5 x 500 µl Strep–elution buffer (100 mM tris-HCl, 500 mM NaCl, 2.5 mM, desthiobiotin, pH 7.9). The eluates were combined and concentrated at 4°C by dialysis into 10 mM HEPES-KOH, pH 7.9, 500 mM KCl, 1 mM EDTA, 1 mM dithiothreitol (DTT) and 50% (v/v) glycerol. The protein was stored at -20°C and concentrations were determined by absorbance (280 nm) according to Pace *et al.* [162].

AvrBs3 variants

Strep-AvrBs3 variants (Figure 33), containing a Strep-tag were purified using a Strep-Tactin sepharose column as describe in chapter 2.2.3. Except that the supernatant after cell lysis was filtered with a syringe filter (0.45 µm Celluloseacetat Membran, VWR) and loaded on the Strep –Tactin sepharose column. AvrBs3 variants were labeled in the next step or purified in a second step over heparin sepharose columns. The protein eluate of the Strep-tag purification was diluted to a salt concentration of 160 mM. The ion exchange chromatography was performed with Äktapurifier 10 Chromatographie System (GE

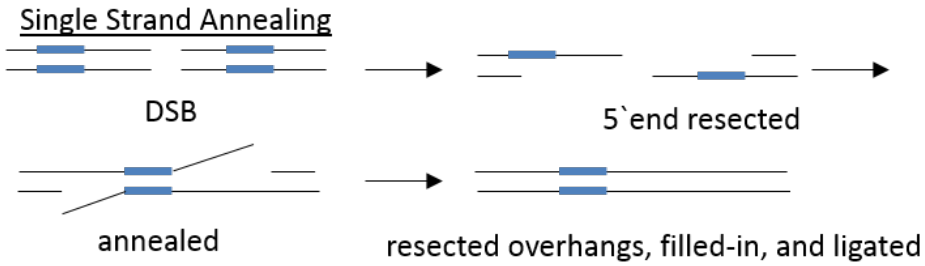

Figure 11: Single strand annealing
Specific nuclease induces a DSB and triggers SSA. The blue arrows indicate the partially duplicated fragments of the lacZ gene. DSB is followed by 5'-end resection and the duplicated parts anneal to each other. This results in overhangs, which are resected and the gabs are filled with nucleotides. Finally, the junction is ligated and resulting in a functionally lacZ gene.

Healthcare). The gradient was shaped by a high (1000 mM NaCl) and low salt (150 mM NaCl) buffer and AvrBs3 variants eluated at a NaCl concentration of 285 nM. The eluate were combined and concentrated at 4°C by dialysis into 10 mM HEPES-KOH, pH 7.9, 500 mM KCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM dithiothreitol (DTT), 50% (v/v) glycerol.

2.2.4 *Eukaryote cell systems*

Single strand annealing assay

To determine the specific *in vivo* activity and efficiency of the analyzed nucleases, a yeast based recombinant assay was established according to the method previously described by Cermak et al. [33]. Briefly, successful cleavage of the target positioned between two partially duplicated fragments of the lacZ gene reconstitutes a functional lacZ gene due to subsequent single strand annealing (figure 11), thus enabling a quantitative analysis of the nuclease activity. Yeast cells YPH500 were co-transformed with a 1:1 ratio of the reporter plasmid (pCp5_addressed site) and the expression plasmid (pZHY500_specific nuclease). Three different amounts of respective inserted DNA were tested (1000 ng, 2000 ng, and 3000 ng). 200 µl of the transformation-mixture were evenly dispersed on a drop out media plate, containing 0.08 mg/ml 5- bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X–GAL). After 5 days of incubation at 30°C, outgrowth of blue and white colonies was counted. All designed nucleases were compared to I-SceI on target T-3-HH-3-T. Levels of significance were determined by unpaired T-tests ($P < 0.05$ *, $P < 0.01$ **, $P < 0.001$ ***).

2.2.4.1 *HEK293T cytotoxicity assay*

The TALE-MutH constructs were analyzed in terms of their cytoxicity and activity in HEK293T cells together with Dr. Yanik in the retinal gene therapy lab of Prof. Dr. Dr. Knut Stieger.

CYTOTOXICITY ASSAY [49]: To determine the nuclease-assocciated toxicity, HEK293T cells were seeded in a 24-well plate at a density of 150,000 cells per well in Dulbecco's modified Eagle media supplemented with 10% fetal bovine serum and penicillin/streptomycin (Invitrogen). The cells were transfected using polyethylenimin (PEI)-mediated transfection [203] containing the TALE-MutH pair TALE-RPGR-1(I)-MutH (600 ng)/ TALE-RPGR-2 (I))-MutH (600ng) expression vectors and the eGFP expression vector (100ng, eGFP-N1, Clontech). For comparison, the toxic ZFNs GZF1N/GZF3N [6] (kindly provided by T. Cathomen, University Medical Center Freiburg, Germany), the GOLDEN STANDARD nuclease I-SceI (I-SceI was cloned into the pvax plasmid, Invitrogen) and an empty vector pcDNA3.1 (Invitrogen) were analyzed similar. The number of eGFP-postitve cells were determined by flow cytometry (FACSCalibur; BD Biosciences) at day 2 and 4 after the transfection. The survival rate is analyzed by counting the decrease of eGFP-positive cells in this interval and is normalized to I-SceI.

TRAFFIC LIGHT REPORTER (TLR) SYSTEM: The traffic light reporter system has been described by Certo *et al. [34]* and was employed to determine the activity of the TALE-MutH pair (TALE-RPGR-1(I)-MutH/ TALE_RPGR-2 (I))-MutH). The TLR system enclosing a I-SceI recognition site was modified by replacing the incorrect mCherry encoding sequence to the incorrect blue (BFP) encoding sequence using the InFusion cloning (Clontech) system. NHEJ, triggered by I-SceI, causes a frame shift which result in the correct BFP reading frame. BFP positive cells were counted by flow cytometry (FACSCalibur; BD Biosciences, kindly provided by H. Hackstein, JLU Gießen). The recognition site for the TALE-MutH pair was introduced adjacent to the I-SceI target site using a AvrII and BsiWI restriction site.

HEK293T cells were seeded similarly as for the cytoxicity assay. After 24 h, cells were transfected with the PEI transfection mix (PEI (0.1 g/l, pH: 5.5, 150 mM NaCl) containing 600 ng of the TLR plasmid enclosing the specific target site and 600 ng of the specific nuclease expression vector. Cells were harvested at day 3 after the transfection and the number of BFP positive cells was determined by flow cytometry (FACSCalibur; BD Biosciences).

2.2.5 *Cleavage assays*

To determine the cleavage specificity and activity *in vitro* two cleavage assays, recently described by Gabsalilow *et al.* [78] were used.

Cleavage assay on plasmid

The *in vitro* activity of Strep-AvrBs4(28)-MutH-His_6 on H-3-T and T-3-H were analyzed on a plasmid containing the addressed site and 18 unaddressed GATC sites. 8 nM substrate were incubated with 32 nM Strep-AvrBs4(28)–MutH-His6 at 37°C (20 mM Tris-acetat, 120 mM K-acetat, 1mM MgCl2, 50 µg/ml bovine serum albumin (BSA), pH 7,5). Aliquots of the reaction were stopped with 50 mM EDTA at defined time intervals (1', 3', 10', 30', 60', 120', 180') and analyzed by an 1.2% agarose gel electrophoresis. The gels were stained by ethidium bromide and analyzed using LabImage 1D.

Cleavage assay on fluorescently labelled PCR fragments

To determine the strand-specificity of the substrates fluorescently labelled PCR products were used. To amplify the PCR products (T-3-H, H-3-T, T-6-H, T-3-HH-3-T and T-3-H-3-T), 5'-end labeled (Atto 488 or Atto 647N) forward and reverse primers were used. 20 nM labeled PCR-substrate were incubated with six-fold excess of enzyme at 37°C (20 mM Tris-acetat, 120 mM K-acetat, 1 mM $MgCl_2$, 50 µg/ml (BSA), pH 7.5). Reaction were stopped by 50 mM EDTA at defined time points (3', 10', 30', 60', 120', 180') and analyzed by denaturingpolyacrylamide gel electrophoresis (10% polyacrylamide 19:1 (PAA), 7M urea, formamide 30%, 1 x TBE).

Similarly, we incubated the targets with AvrBs4-MutH for 180 min, followed by digestion with 1 µl Proteinase K (10 min at 37°C), followed by protein heat deactivation for 10 min at 95°C. New 120 nM AvrBs4-MutH was added and allowed to digest the substrate for another 60 min at 37°C. The defined reaction steps were stopped with 50 mM EDTA and analyzed by denaturing PAGE (heat: 400 V, 40 mA, 30 min; run: 300 V, 30 mA, 45 min) and native 10% (PAA) gels (run: 300 V, 30 mA, 45 min). The fluorescent images were visualized with TyphoonTM FLA 9500. Alexa 488 was excited at 473 nm using the BPG1 (560-580 nm) filter and Atto647N was excited at 635 nm using the LPR (> 665 nm) filter.

2.2.6 *Fluorescence based methods*

Labeling of AvrBs3 variants

The proteins were all labeled by a thiol Michael-type 1,4-addition reaction, creating a covalent connection between the thiol group of the cysteine residue and the maleimide fluorophores. The reaction mechanism is shown in figure 12. The thiol group is deprotonated in presence of a base. Next, the the thiol group is added to the double bond of the maleimide creating an intermediate (I). Hydrolysis of the intermediate result in the succinimide-thiolether (II).

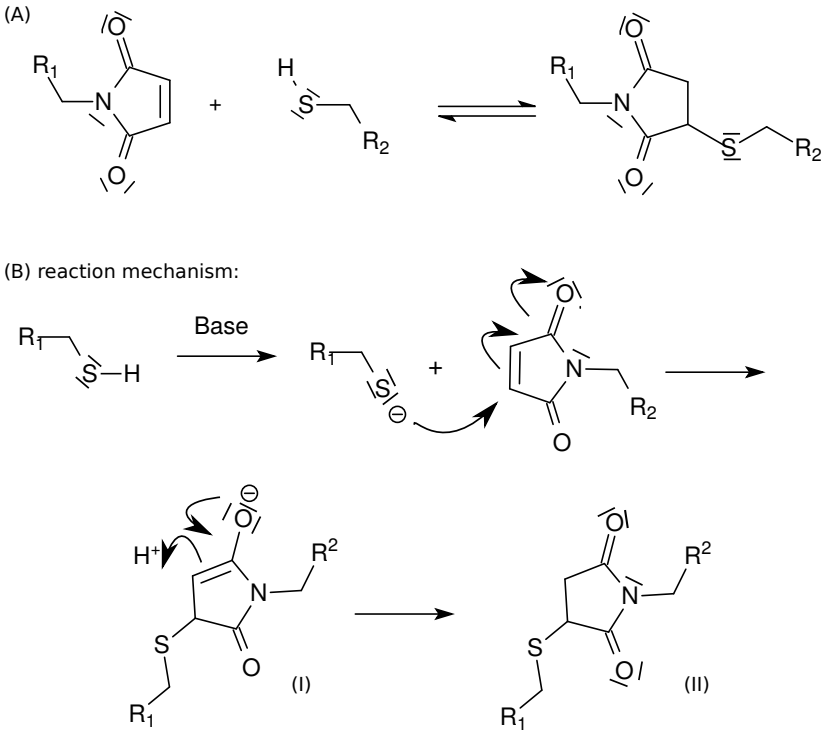

Figure 12: Thiol Michael-type addition reaction
(A) Equilibrium reaction of the thiol group of the cysteine with the maleimide fluorophore. (B) Reaction mechanism of the thiol Michael-type 1,4-addition. The thiol group is deprotonated in presence of a base and is added to the maleimide by creating an intermediate (I). Hydrolysis results in the succinimide-thiolether (II).

LABELING CONDITIONS: 5 mM redox reagent Dithiothreitol (DTT) was added to 80 µM of the protein and incubated at 0°C for 1 h. Dialyse puffer was exchanged to an adjusted puffer for labeling (10 mM 2-(4-(2-hydroxyethyl)-1-piperazinyl)-ethansulfonsäure (HEPES) /KOH pH 8,0, 1 mM EDTA, 200 mM KCl, 10% glycerol) by a zeba spin desalting column 40 K. The fourfold excess of the maleimide dye was supplemented to the single cysteine AvrBs3 variants and the twofold excess of maleimide dye was added to the AvrBs3 (AB) variant. The

protein was incubated at 0°C under exclusion of light for 1-2 hours. AvrBs3 variants were purified by the zeba spin desalting column 40 K and eluted twice with the high salt labeling puffer (10 mM 2-(4-(2-hydroxyethyl)-1-piperazinyl)-ethansulfonsäure (HEPES)/KOH pH 8,0, 1 mM ethylenediaminetetraacetic acid (EDTA), 500 mM KCl, 10% glycerol). The aliquots were frozen with liquid nitrogen and stored at -80°C.

2.2.6.1 *Degree of labeling (DOL)*

The degree of labeling was determined after each labeling reaction of a protein with a fluorophore. DOL describes the ratio between the dye and the protein concentration (Formula (1)).

$$DOL = \frac{c(dye)}{c(protein)} \tag{1}$$

The concentrations of the dye and the protein were measured with the spectrophotometer NanoDrop® 1000 (Thermo Scientific).

$$c(dye) = \frac{A_{max}}{\varepsilon_{max} * d} \tag{2}$$

$$c(protein) = \frac{A_{prot.}}{\varepsilon_{prot} * d} \tag{3}$$

$$A_{prot.} = A_{280nm} - A_{max} * cf_{280} \tag{4}$$

$$cf_{280} = \frac{\varepsilon_{max}}{\varepsilon_{280nm}} \tag{5}$$

Similar to the determination of the protein concentration the DNA concentration was analyzed and could be used to define the DOLs of the labeled DNAs. The absorption maximum of protein exist at 280 nm, whereas the absorption maximum of DNA is at 260 nm. Thus, the absorption at 260 nm ($A_{260\ nm}$) was measured and the extinction coefficient ε_{260nm} was required to calculate the DNA concentration.

2.2.6.2 *Electrophoretic Mobility Shift Assay (EMSA)*

EMSA can be used to analzye DNA-protein or RNA-protein interaction. We applied an EMSA based on a fluorescently labeled DNA. A 25 bp long DNA containing the AvrBs3 recognition site labeled with Cy3 at the 5'-end of the lower strand was used. The unbound DNA contains a high mobility due to its small size. Binding of protein to

the DNA results in a reduction in the mobility and can be analyzed on a gel. 5'-end Cy3 labeled DNA (20 nM) was princubated under high salt conditions (150 mM) in 1x NEB 3.1 buffer with protein (0-600 nM) for 5 minutes on ice. The reaction was stopped with 50 mM EDTA and the whole reaction was loaded to a 4% polyacrylamid gel (35 mA, 100 V, 20 min, 4°C). The fluorophore Cy3 (filter: BPG1(560-580 nm)) and Alexa 594 (filter: LPG (>575 nm)) was excited the the green SHG 532 nm, the fluorophore Alexa488 (filter: BPB1 (520-540 nm)) was excited with the blue LD laser (473 nm), Alexa 647 (filter: LPR (>665 nm)) was excited with the red LD laser (635 nm) and the gel was visualized with the TyphoonTM FLA 9500.

Fluorescence anisotropy

Fluorescence anisotropy is a tool to analyze binding constants and kinetics. It depends on the rotation time of the fluorophore part or the whole molecule. An advantage of fluorescence anisotropy is that only one molecule has to be labeled. For instance, using fluorescence anisotropy to investigate DNA::protein binding, the DNA can carry the fluorophore or the protein. Therefore, the protein does not has to be modified, thus excluding the risk of changing the binding character of the analyzed protein.

The fluorophore is excited with linear polarized light. It absorbs a photon, which results in a higher energy level of the valence electrons. The fluorescence lifetime τ describe the timephase the higher energy level exist. The energy is mostly released in parts by oscillation and the remaining energy is emitted by a photon. The fluorophore molecule has to be oriented in the angle of the polarized light to absorb the photon. If the molecule tumbling is "slow" the emitted photon will be polarized similar to the applied polarized light. whereas, a small fluorophore molecule will change the orientation before emitting the photon, thus resulting in a different orientation. Taken together, fluorescence anisotropy is dependent on the rotation lifetime ϕ and the fluorescence lifetime τ of the fluorophore. To detect this effect, the fluorescence intensity is measured parallel and perpendicular to the polarized light by aligning the polariser and the analyzer to each other [123].

Binding characteristics of the AvrBs3 variants to the DNA were analyzed by fluorescence anisotropy. The assays were performed under low salt conditions (60 mM) in assay puffer (20 mM tris-acetat, 50 mM kkalium-acetat, 10 mM magnesium-acetat, 1 mM DTT, pH 7.9) and fluorescence anisotropy was measured by using the Fluoromax-4 Spectroflurometer to determine the K_D value. The 25 bp long DNA contained the AvrBs3 binding site and was labeled with Hex at the 5´ prime end of the bottom strand. To determine the K_D-value 0.5 nM of the DNA was set constant and the protein was titrated to the solution.

To define the active rate of protein, the titration was repeated with 10 nM DNA. The kinetic experiments were performed under the same buffer conditions, but using the Tecan infinitie F200 pro plate reader. 10 nM labelled DNA with 50 nM AvrBs3 was used to achieve complete binding. For competition experiments unlabeled specific DNA (180 nM) and scramble DNA (180 nM) were used.

Förster-Resonanzenenergietarnsfer (FRET)

FRET is a physical process, which enable to measure the acceptor emission resulted by donor excitation. Thus, FRET requires a donor fluorophore and an acceptor fluorophore, which can be labeled to on protein (intramolecular) or to diverse biomolecules (intermolecular). The radiation-free energy transfer occur, when defined requirements are given. The donor emission spectra have to overlap with the acceptor excitation spectra (I); the orientation of the transient dipole moments have to bee favourable (II); and the proximity between donor and acceptor fluorophore has to bee given (1-10 nm) [74, 75, 73, 244].

2.2.6.3 *FRET experiment between Cy3 and Cy5*

The FRET experiments were performed under low salt conditions (80 mM) with assay puffer (20 mM tris-acetat, 50 mM kalium-acetate, 10 mM magnesium-acetat, 1 mM DTT, pH 7.9). The specific DNA (Tabel 12) was labeled with Cy3 at the 5'-end of the bottom (specific DNA (I)) or the top (specific DNA (II)) strand. AvrBs3 (A)/AvrBs3 (B) labeled with Cy5 was titrated to 1 nM of the labelled specific DNA, respectively.

To calculate the FRET index ($nF/F_D{}^{ex_D em_D}[\%]$) the three cube method described by Youvan *et al. [242]* was used (Formula 6). This method describes no FRET efficiency, but a values, which change with FRET [244]. Three measurements were required, when donor and acceptor fluorophores were presence. First, the donor was excited and emission in acceptor ($F^{ex_D em_A}$) and donor ($F^{ex_D em_D}$) channel was detected. Next, the acceptor was excited and emission was measured in the acceptor channel, resulting in the $F^{ex_A em_A}$ signal.

The donor fluorophore Cy3 was excited at 520 ± 5 nm and the acceptor fluorophore Cy5 was excited at 650 ± 5 nm using the Fluoromax-4 Spectrofluorometer. To consider crosstalk of Cy3 and Cy5 both Cy3 labelled DNA and Cy5 labelled AvrBs3 were excited individually, and the resulting emissions in the donor channel and in the acceptor channel were measured (Formula 7 and 8) . FRET efficiency was calculated with the following formulas.

$$nF = F^{ex_D em_A} - F^{ex_A em_A} * \alpha) - (F^{ex_D em_D} * \beta) \tag{6}$$

$$\alpha = \frac{F_A^{e_D em_A}}{F_A^{e_A em_A}} \tag{7}$$

$$\beta = \frac{F_D^{e_D em_A}}{F_D^{e_D em_D}} \tag{8}$$

nF: indicates the corrected fluorescence signal
F: indicates the fluorescence signal in presence of the donor and the acceptor
F_A*: indicates the fluorescence signal in presence of the acceptor*
F_D*: indicates the fluorescence signal in presence of the donor*
ex_D*: donor excitation*
ex_A*: acceptor excitation*
em_D*: donor emission*
em_A*: acceptor emission*

FRET efficiency can be described by the Förster equation (9). FÖRSTER RADIUS (R_o) describes the distance between donor and acceptor fluorophores, when 50% energy transfer takes place and it depends on the fluorophhores used. The actual distance between donor and acceptor is given with r_o. Formula (9) can be converted to calculate r_o, if FRET efficiency is given (Formula 10).

$$E = \frac{R_0^6}{R_0^6 + r_0^6} \tag{9}$$

$$r_0 = \sqrt[6]{\frac{R_0^6}{E - R_0^6}} \tag{10}$$

A stoichiometric method which results in a FRET efficiency was described by Hoppe *et al. [98].* They applied the correction factor γ, which describes the relative excitability of the donor and acceptor at the donor excitation wavelength [244].

$$Ef_A* = \gamma * \frac{nF}{\alpha * F_A^{ex_A em_A}} \tag{11}$$

$$\gamma = \frac{{\varepsilon_A}^{exD}}{{\varepsilon_D}^{exD}} \tag{12}$$

$\varepsilon_A{}^{exD}$: *extinction coefficient of the acceptor fluorophore at the donor excited wavelength*
$\varepsilon_B{}^{exD}$: *extinction coefficient of the donor fluorophore at the donor excited wavelength*

In the experiment we excited Cy3 at 520 nm to minimize cross talk. Thus, we needed the $\varepsilon_A{}^{520nm}$ *and* $\varepsilon_B{}^{520\ nm}$ to calculate the correction factor γ.

Additional the the donor quench (E_q) was expressed in percentage [%].

$$E_q = \frac{F_D^{ex_D em_D} - F^{ex_D em_D}}{F_D^{ex_D em_D}} \tag{13}$$

Part III

RESULTS AND DISCUSSION

3 RESULTS

3.1 CHARACTERIZATION OF HIGHLY SPECIFIC NUCLEASES *in vitro* AND *in vivo*

This chapter characterizes the monomeric strand and site specific nickase AvrBs4-MutH under *in vitro* and *in vivo* conditions with a focus on the optimization of the target design. Additionally, AvrBs4-FokI and AvrBs3-PvuII variants were studied in *Saccharomyces cerevisiae* and compared to the GOLDEN STANDARD meganuclease I-SceI. Furthermore, the RNA guided Cas9 protein was evaluated in yeast. Finally, the most promising TALEN with the optimized target design was applied to a therapeutic relevant gene causing X-linked retinitis pigmentosa.

3.1.1 *Characterization of TALENs*

3.1.1.1 AvrBs4-MutH

AvrBs4-MutH designed by Gabsalilow *et al.* [78], is a truncated AvrBs4 protein fused to the nickase MutH. The coding sequence of the N-terminus of AvrBs4 is shortened by 152 aa resulting in 138 aa and the C-terminus contains 28 aa. The nickase MutH, which addresses the 4 bp sequence (GATC), is fused to the C-terminus of AvrBs4 via two serine residues. The TALEN protein holds a Strep-tag at the N-terminus and a His-tag at the C-terminus, which enable a tandem purification (Figure 13).

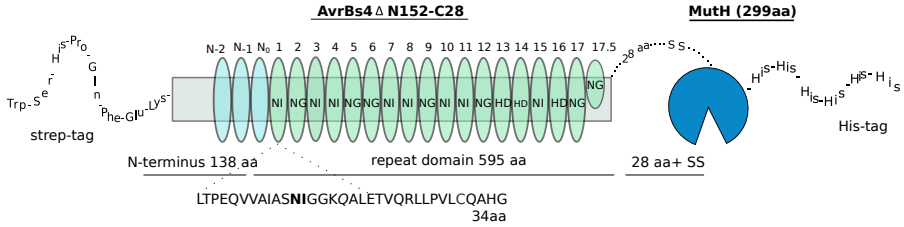

Figure 13: Architecture of AvrBs4-MutH
AvrBs4 contains a N-terminus with 138 aa, a repeat domain with 595 aa, and a C-terminus with 28 aa. The Strep-tag is fused to the N-terminus, which consist of the non-canonical repeats N_0, N_{-1} and N_{-2}. The binding domain encloses 17.5 repeats. Every repeat holds 34 highly conserved aa. In position 12 and 13 the aa residues are crucial for DNA recognition (**bold**). Two serines (S-S) connect the truncated AvrBs4 to the nickase MutH. The nickase carries a His-tag at the C-terminus and addresses a GATC site.

Previous *in vitro* experiments revealed that AvrBs4-MutH is capable to induce strand and site specific SSBs. Dependent on the spacer length between the AvrBs4 addressed site and the GATC sequence, the target is nicked on the bottom strand (3 bp spacer) or on the top-strand (6 bp spacer) (Figure 14). These characteristics of the nickase AvrBs4-MutH have been described by Gabsalilow *et al.* [78] in detail. For the present study, AvrBs4-MutH was analyzed on specifically designed target sequences. The construction of the new recognition sites were based on recent findings in the filed of genome targeting [53, 40, 199, 191]. In respect of HDR efficiency, the importance of nicking either the transcribed or the non-transcribed strand has been emphasized [53]. Accordingly, the target T-3-H, T-6-H and H-3-T were used to characterize AvrBs3-MutH under *in vitro* and *in vivo* conditions. Double nicking by specific nickases has been shown to reduce off-target activity [40, 199, 191]. This encouraged us to analyze the mechanism of double nicking by AvrBs4-MutH on the double targets T-3-HH-3-T and T-3-H-3-T. Three single targets (T-3-H; H-3-T and T-6-H) were designed to be nicked by AvrBs4-MutH and two double targets were created to detect double nicking. All targets comprise an I-SceI target site and are visualized in figure 14.

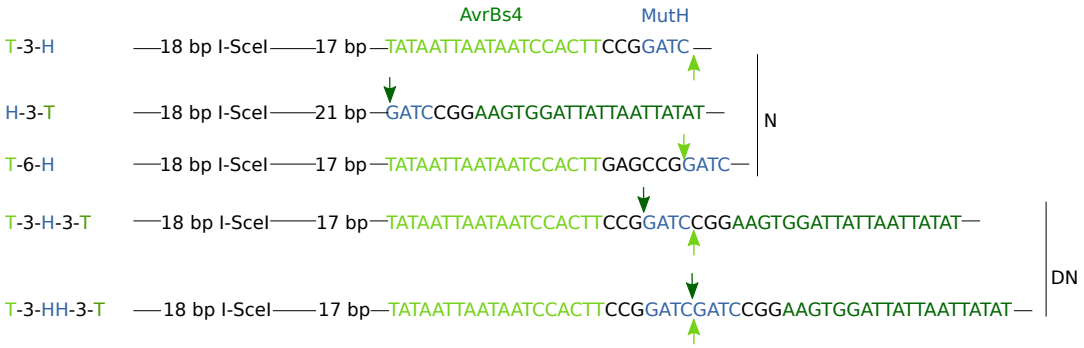

Figure 14: Design of specific single and double targets
I-SceI target site encloses 18 bp and can be found in all targets 17-21 bp in front of the binding site of AvrBs4-MutH. TALE binding site (T) is marked in green, whereas the recognition site of MutH (H) is shown in blue. The spacer length between the two binding sites is given in bp (x = 3, 6). Arrows demonstrate the nicking positions of MutH. T-3-H, H-3-T, and T-6-H are nicked (N) at a single position and are termed single targets. T-3-HH-3-T and T-3-H-T carry two positions to induce SSB and are termed double targets.

3.1.1.2 *in vitro Analysis*

The cutting mechanism and activity of AvrBs4-MutH were analyzed by employing a plasmid cleavage assay and a cleavage assay using different fluorescently 5'-end labeled PCR products. First, the inverted single target sites (T-3-H and H-3-T) were analyzed and compared with a plasmid cleavage assay (Figure 15). AvrBs4-MutH induced almost two times faster ($k = 3.\ 8 * 10^{-2} \pm 9.\ 6 \times 10^{-4}\ min^{-1}$) SSBs to target H-3-T, than to target T-3-H ($k = 2.\ 2 \times 10^{-2} \pm 2.\ 9 \times 10^{-3}\ min^{-1}$).

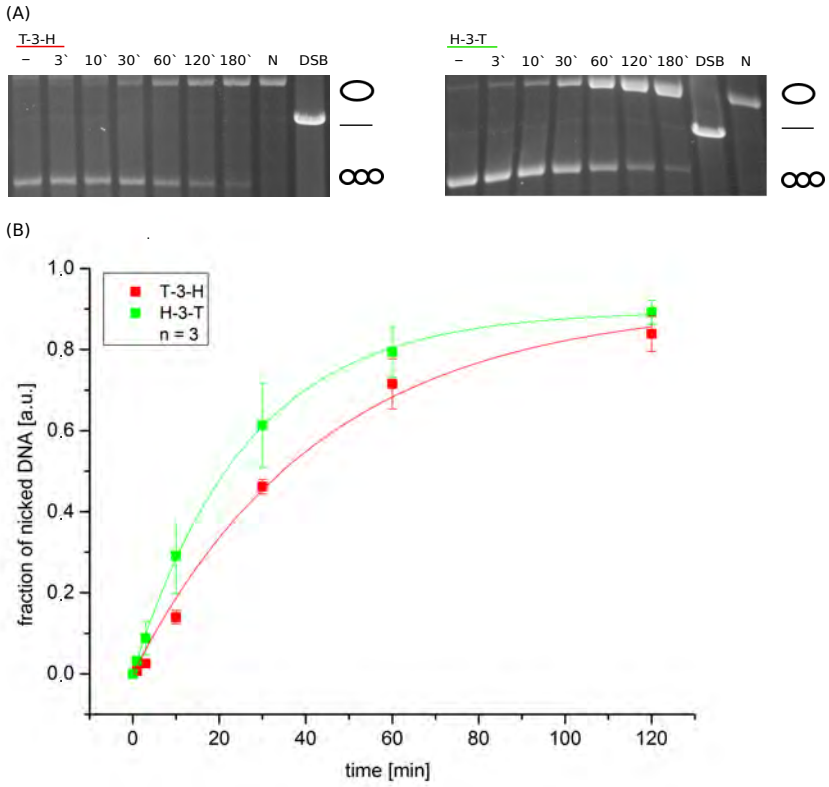

Figure 15: Kinetic cleavage assay on T-3-H and H-3-T

Fraction of nicked DNA of T-3-H and H-3-T, analyzed by a plasmid cleavage assay. 8 nM substrate was incubated with a four fold excess of AvrBs4-MutH and time points were taken at 3', 10', 30', 60', 120', 180'. (A) Agarose gels (0.8%) stained with ethidium bromide. The gel demonstrates the super-coiled DNA, linear DNA, and open circle DNA. Substrates were digested by the nicking enzyme Nt.BspQ1 (N) and BamHI (DSB), respectively. (B) Quantitative evaluation of the cleavage assay. Each measuring point represents the mean ± standard error indicated by error bars for each point. (n) number of independent experiments. AvrBs4-MutH had a cleavage rate constant for H-3-T of $k = 3.\ 8 \times 10^{-2} \pm 9.\ 6 \times 10^{-4}\ min^{-1}$ and for T-3-H $k = 2.\ 2 \times 10^{-2} \pm 2.9 \times 10^{-3}\ min^{-1}$.

Next, we identified which DNA strand was nicked by AvrBs4-MutH using the cleavage assay on 5'-end labeled PCR fragments. The principle of the cleavage assay with fluorescently 5'-end labeled PCR products is depicted in figure 16 by using the example of the double target T-3-H-3-T. Relative to the strand which is cut on the fluorescently 5'-end labeled PCR product by AvrBs3-MutH, either a long fragment (eg. 169 nt) with a red fluorescent signal or a short fragment (eg. 59 nt) with a green fluorescent signal result.

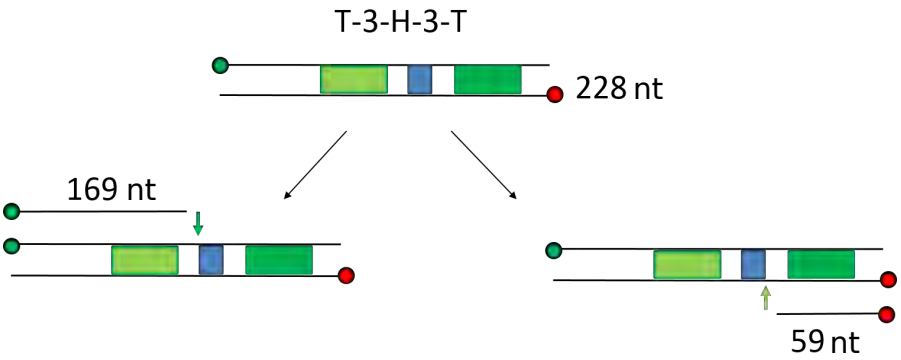

Figure 16: Cleavage assay with labeled PCR fragments
Cleavage assay with fluorescently 5'-end labeled PCR fragments exemplified by the double target T-3-H-3-T. The introduction of a SSB into the top strand results in a long fragment (e.g. 169 nt) containing a green fluorescent signal (Atto 647N). While the SSB introduced into the bottom strand will result in a short fragment (e.g. 59 nt) with a red fluorescent signal (Atto 488). In case of the target T-3-H-3-H, the total size of the fragment is 228 bp.

First, the results of the single targets were analyzed on denaturing PAGE and native PAA gels. While the denaturing PAGE gel revealed which DNA strand is nicked, the occurrence of DSBs was detected by the analysis of native PAA gels. T-3-H was cleaved on the bottom strand, whereas H-3-T and T-6-H were nicked on the top strand (Figure 17 (A)). No DSBs were induced (Figure 17 (B)). The analysis of the fluorescently 5'-end labeled PCR double targets on a denaturing PAGE revealed that top and bottom strands were nicked in a 1:1 ratio (Table 19 and Figure 18 (A) and 19). No double strand breaks were detectable after 180 min of digestion. However, removing AvrBs4-MutH by Proteinase K (PK) and adding new AvrBs4-MutH after deactivation of PK led to double nicks in particular on T-3-H-3-T (Figure 18 (B)). The analysis of enzyme kinetics on double targets revealed that AvrBs3-MutH nicked T-3-H-3-T three to four times faster than T-3-HH-3-T (Figure 19) and (Table 19).

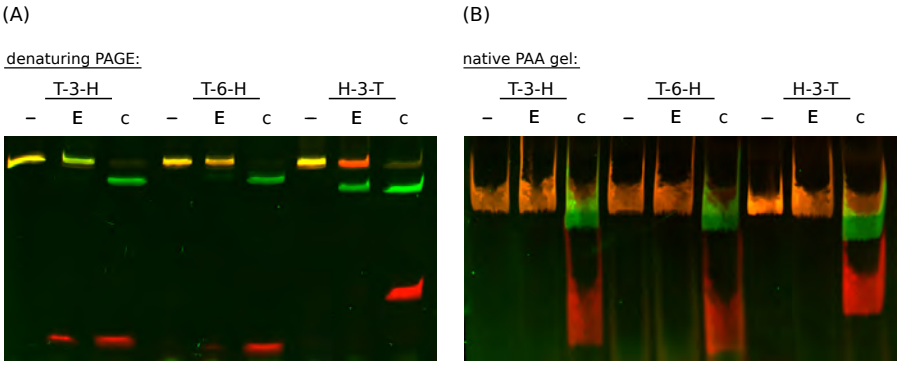

Figure 17: *In vitro* characterization of AvrBs4-MutH on single targets (T-3-H, T-6-H and H-3-T)
Cleavage assay with fluorescently 5'-end labeled PCR products analyzed on a denaturing PAGE (A) and a native PAA gel (B). (-) uncut labeled PCR target (20 nM). (E) Target preincubated for 180 min with six fold excess of AvrBs4-MutH. (c) Target digested by BamHI. The uncut fragment is colored in orange. Fragments resulting from top strand nicking were labeled with a green fluorophore (Atto 488), while fragments resulting from bottom strand nicks were labeled with a red fluorophore (Atto 647N). T-3-H was nicked on the bottom strand, whereas T-6-H and H-3-T were nicked on the top strand. AvrBs4-MutH induced no DSBs on the single targets.

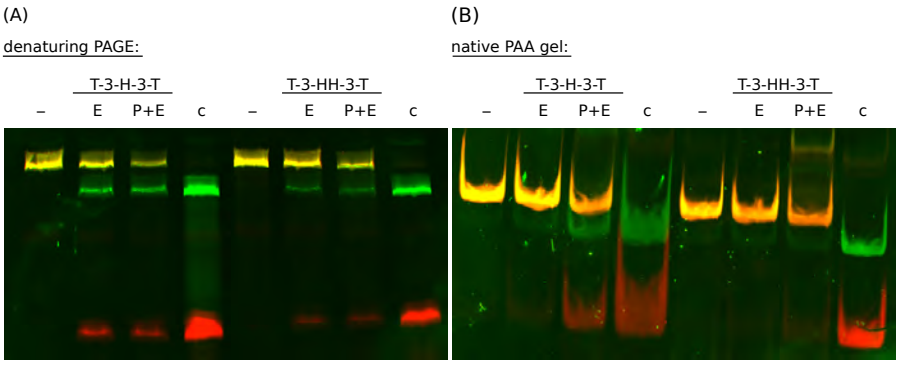

Figure 18: *In vitro* characterization of AvrBs4-MutH on double targets (T-3-H-3-T and T-3-HH-3-T).
Cleavage assay with 5'-end labelled PCR double targets (T-3-H-3-T and T-3-HH-3-T) analyzed on a denaturing PAGE (A) and a native PAA gel (B). (-) uncut labeled PCR target (20 nM); (E) target incubated with six fold excess of AvrBs4-MutH for 180 min; (P+E) AvrBs4-MutH preincubated on target site (E) was removed with PK (10 min, 37°C) and additional protein (120 nM) was incubated for 60 min after deactivation of PK; (c) control, target digested by BamHI (T-3-H-3-T) or ClaI (T-3-HH-3-T). The uncut fragment is shown in orange. Fragments resulting from top strand nicking were labeled with a green fluorophore (Atto 488), while fragments nicked at the bottom strand were labeled with a red fluorophore (Atto 647N). After treatment with PK and adding new AvrBs4-MutH after deactivation of PK DSBs were detectable in particular on T-3-H-3-T.

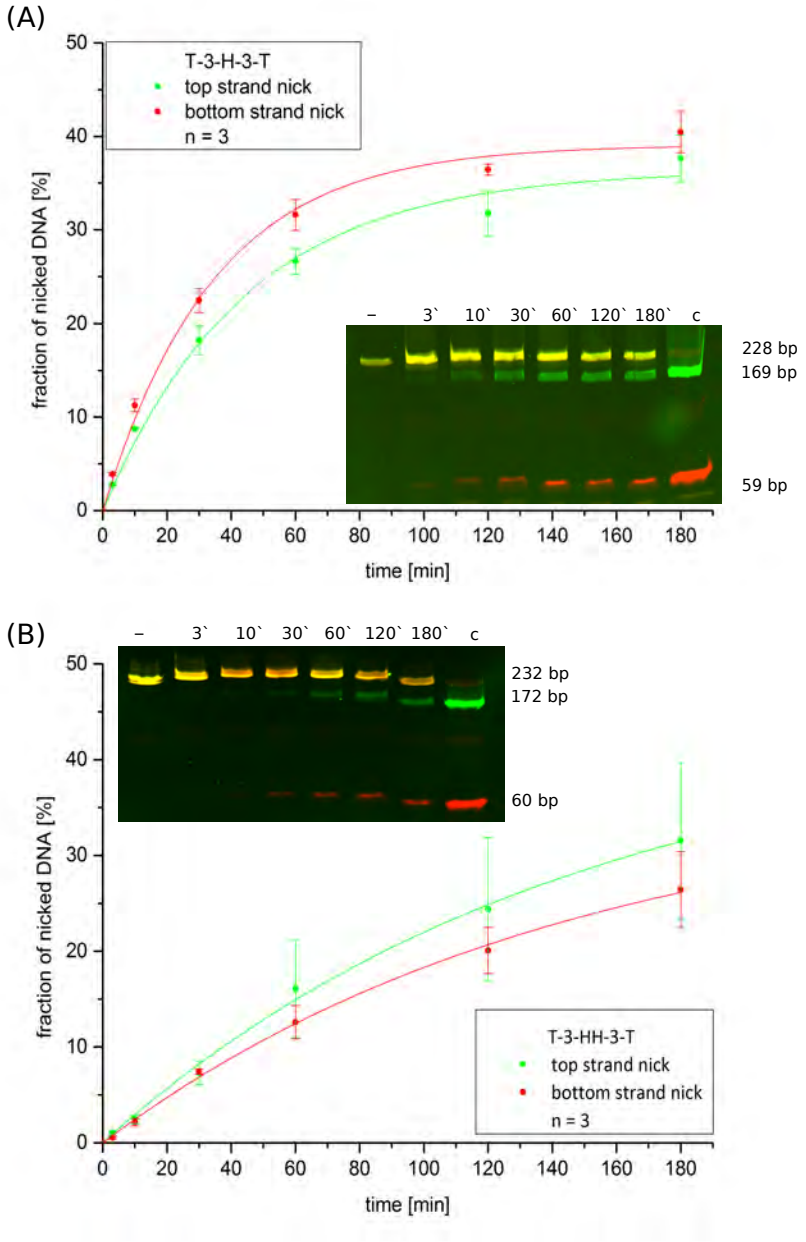

Figure 19: Kinetics of DNA nicking of AvrBs4-MutH on the double targets (T-3-H-3-T, T-3-HH-3-T)
Fluorescently 5'-end labeled PCR target (20 nM) was incubated with a six fold excess of AvrBs4-MutH and time points were taken at 3', 10', 30', 60', 120', 180'. Each measuring point represents the mean ± standard error indicated by error bars for each point. (n) number of independent experiments. AvrBs4-MutH nicked T-3-H-3-T (A) three to four times faster than T-3-HH-3-T (B). Cleavage rate constants of AvrBs4-MutH (n= 3): T-3-H-3-T: 0.023 ± 2.5 x 10^{-2} min^{-1} (top strand), 0.029 ± 2.3 x 10^{-3} min^{-1} (bottom strand); T-3-HH-3-T : 6.8 x10^{-3} ± 9.4 x 10^{-4} min^{-1} (top strand), 6.9 x10^{-3} ± 6.7 x 10^{-4} min^{-1} (bottom strand).

Table 19: Enzyme activity of AvrBs4-MutH for two independent double targets
Top and bottom strand-specific cleavage rate constant K [min^{-1}] of AvrBs3-MutH is shown for the double targets T-3-HH-T and T-3-H-3-T. The binding site of AvrBs4 is marked with T, the GATC binding site of MutH is labeled with H, and the spacer length between the binding sites is given in bp.

		K [min^{-1}]
T-3-HH-3-T	top strand	$6.8 \times 10^{-3} \pm 9.4 \times 10^{-4}$
	bottom strand	$6.9 \times 10^{-3} \pm 6.7 \times 10^{-4}$
T-3-H-3-T	top strand	$2.3 \times 10^{-2} \pm 2.5 \times 10^{-3}$
	bottom strand	$2.9 \times 10^{-2} \pm 2.3 \times 10^{-3}$

3.1.1.3 *in vivo Analysis*

To challenge the established specific nucleases (I-SceI or TALE-FokI), TALE-MutH has to provide excellent *in vivo* efficiency. To investigate the efficiency of TALE-MutH under *in vivo* conditions and to validate the *in vitro* results, a single strand *in vivo* assay in yeast (Figure 20), as has been previously reported by Cermak *et al*. [33], was employed. *S. cerevisiae* is an eukaryotic microorganisms, providing a fast and cost efficient model for *in vivo* conditions by using the single strand annealing assay. The successful cleavage of the target positioned between two partially duplicated fragments of the *lacZ* gene reconstituted a functional *lacZ* gene is reconstituted by subsequent single strand annealing (SSA) (Figure 11). This experimental approach enables a quantitative analysis of the nuclease activity (Figure 20). The functional *lacZ* gene encodes a *lacZ* fragment which dimerize with the *lacZ* deletion mutant resulting in an active β-gal.

The ß-gal activity of TALE-MutH was compared to I-SceI and AvrBs4-FokI. I-SceI induces DSBs within its non-palidromic 18 bp target sequence and is the established GOLDEN STANDARD due to its high specificity combined with high nuclease activity. The homodimeric AvrBs4-FokI induces DSBs within a 36 bp (2 x 18 bp) substrate. The repeat array of AvrBs4 was introduced to a truncated variant of the TALE protein via the GOLDEN GATE cloning strategy (Methods 2.2.1.1). The AvrBs4-MutH design was similar to the *in vitro* design, but lacking the His- and Strep-tag. Similar to AvrBs4-MutH, the N-terminus of AvrBs4-FokI lacks the coding sequence of 152 aa residues, but contains a 68 aa as linker to the unspecific cleavage domain FokI. A yeast expression plasmid with AvrBs4 without catalytic domain served as negative control.

Additionally, several variants of AvrBs3-PvuII were analyzed by the SSA assay, as PvuII proofed to be a promising alternative to the unspecific cleavage domain FokI. The specific nuclease PvuII increases the number of bp in the target site (42 bp) [239].

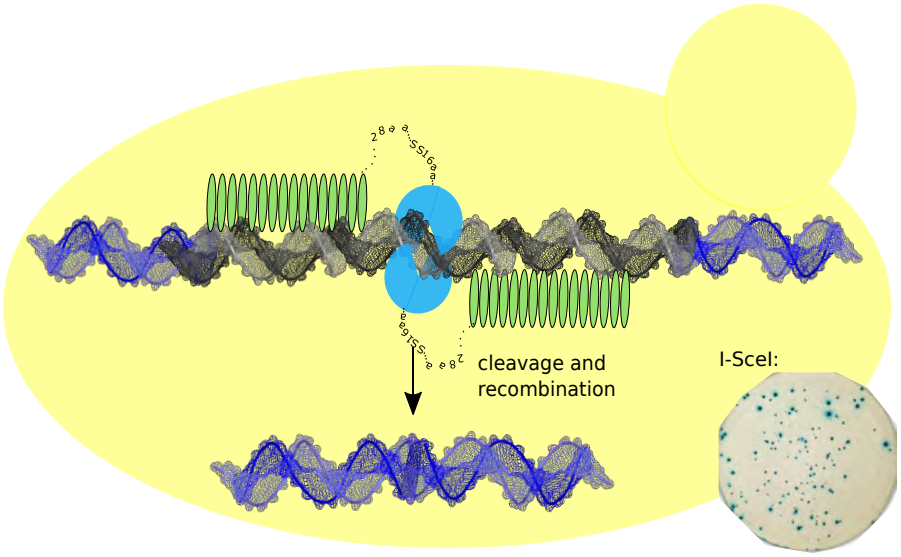

Figure 20: Single strand annealing assay
A yeast based recombinant assay was used to analyze the efficiency of the specific nucleases. The specific nucleases bind to the recognition sequence, which is flanked by two partially duplicated fragments of the *lacZ* gene (blue), SSA is triggered and a functional *lacZ* gene is reconstructed. Right corner: Yeast colonies (blue) with I-SceI expression plasmid and target plasmid growing on a matrix containing X-Gal.

3.1.1.4 *AvrBs4-MutH*

Two versions of the viability assay were used. First, yeast cells were transformed with expression plasmids, but no target plasmid was present. The total amount of yeast colonies of AvrBs4 without nuclease were compared to I-SceI, AvrBs4-FokI and AvrBs4-MutH and no significant difference in the total amount of colonies could be detected (Figure 21 (A)). Second, yeast cells were transformed with expression plasmids in the presence of the plasmid with the defined target for the specific nucleases. This resulted in a significant reduction of the total amount of yeast colonies compared to the total amount of yeast colonies transfected with AvrBs4 without nuclease activity, except in the case of AvrBs4-MutH on the target T-3-H and H-3-T (Figure 21 (B)). AvrBs4-FokI revealed the lowest number of colonies, followed by AvrBs4-MutH on the double target T-3-HH-3-T. However, when compared to I-SceI, AvrBs4-MutH with the targets T-3-H-3-T, T-3-H, and H-3-T led to higher colony counts, while colony counts with the target T-6-H were comparable to I-SceI.

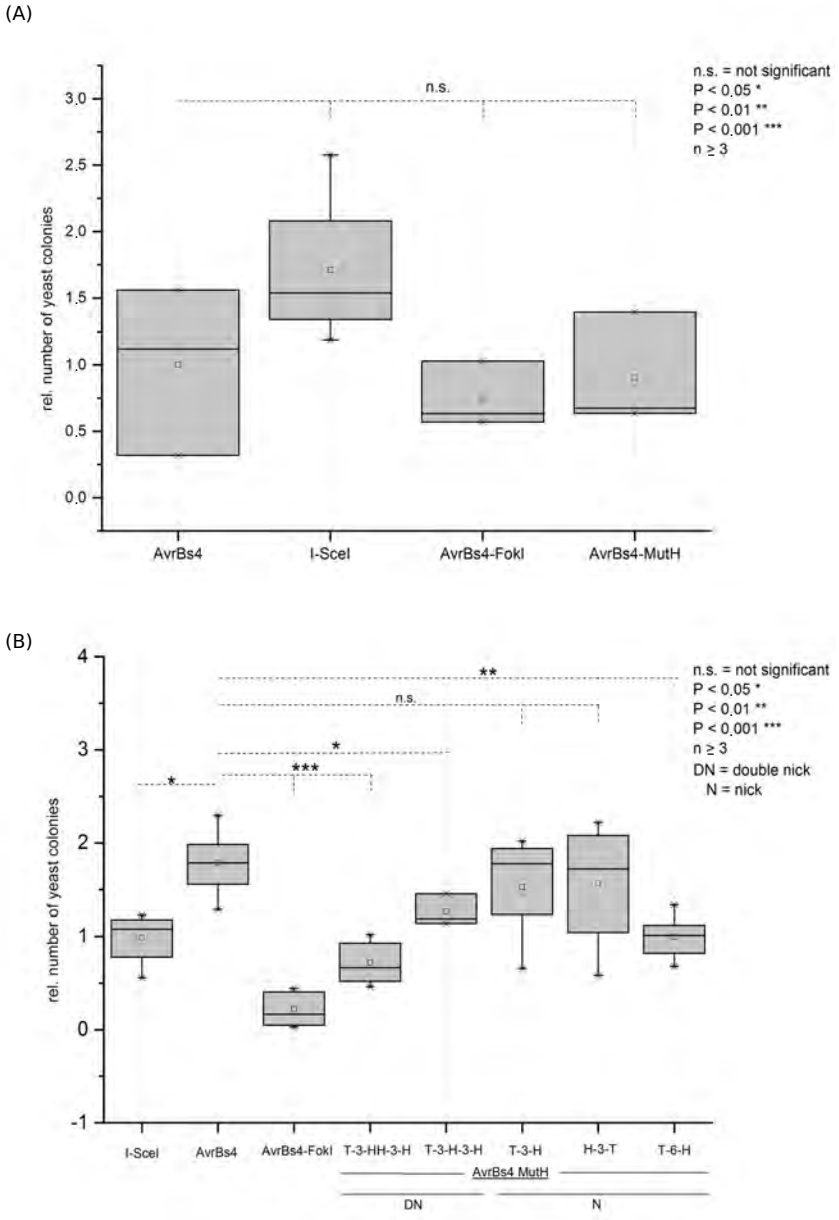

Figure 21: Relative survival of yeast colonies

In vivo characterization of AvrBs4-MutH as nicking and double nicking enzyme compared to established specific nucleases (I-SceI, AvrBs4-FokI). The specific nucleases are compared to AvrBs4 without nuclease activity and significant differences were determined by the unpaired T-test ($n \geq 3$, n.s. = not significant, $P < 0.05$ *, $P < 0.01$ **, $P < 0.001$ ***). Targets underlined with DN (double nick) carry two positions to induce a SSB, while N (nick) indicate a single position to induce SSB. (A) Total amount of yeast colonies transfected with expression plasmid for the specific nuclease without the target plasmid. (B) Quantification of yeast colonies transfected with the specific nuclease and the specific target plasmid.

To analyze the specific activity of the nucleases on different targets, the ratio between blue and unstained colonies was calculated and compared to AvrBs4 without nuclease activity (Figure 22 (A and B)).

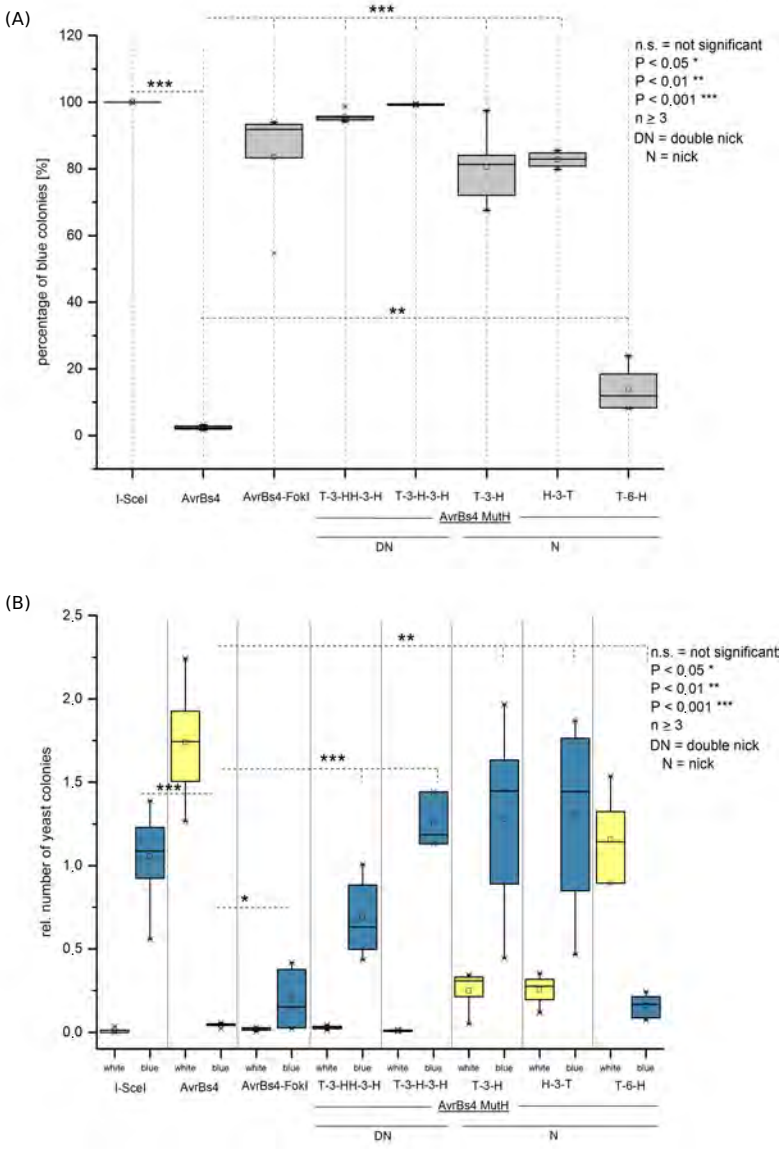

Figure 22: β-gal activity of specific nucleases (I-SceI, AvrBs4-FokI, and AvrBs4-MutH) in *S. cerevisiae*
In vivo characterization of AvrBs4-MutH as nicking and double nicking enzyme compared to established specific nucleases (I-SceI, AvrBs4-FokI). The specific nucleases are compared to AvrBs4 without nuclease activity and significant differences were determined by the unpaired T-test ($n \geq 3$, n.s. = not significant, $P < 0.05$ *, $P < 0.01$ **, $P < 0.001$ ***). Targets underlined with DN (double nick) carry two positions to induce a SSB, while N (nick) indicate a single position to induce SSB. (A) Fraction of blue colonies expressed as percentage [%] of the different specific nucleases. (B) Relative number of white and blue colonies of the nucleases on their specific targets.

When compared to AvrBs4 all constructs on all target combinations showed a significantly higher β-gal activity. Even the single targets showed a significantly higher β-gal activity. This finding indicates that nicking can trigger SSA. The GOLDEN STANDARD I-SceI and AvrBs4-MutH addressing the double target T-3-H-3-T resulted the highest activity, which approximated 100% of stained colonies (I-SceI: 100 ± 1%; T-3-H-3-T: 99 ± 1%), followed by AvrBs4-MutH on the double target T-3-HH-3-T (T-3-HH-3-T: 95 ± 2%) and AvrBs4-FokI (AvrBs4-FokI: 83 ± 17%). The single target T-6-H (8 ± 3%); revealed a three to four fold lower activity when compared to T-3-H (84 ± 9%) and H-3-T (84 ± 2%) (Figure 22 (A)).

Next, the number of white and blue colonies relative to the results of AvrBs4 were analyzed (Figure 22 (B)). Here, AvrBs4-FokI and AvrBs4-MutH targeting T-6-H showed relatively low activities, while the highest enzyme activities were detected for the combinations of AvrBs4-MutH with T-3-H-3-T, T-3-H, and H-3-T. This indicates that AvrBs4-MutH has a greater *in vivo* efficiency compared to the GOLDEN STANDARD I-SceI. While incubation over two additional days led to a further increase in the intensity of the blue-staining of the T-3-H and H-3-T colonies, all other constructs did not change the intensity of the blue color (Figure 23).

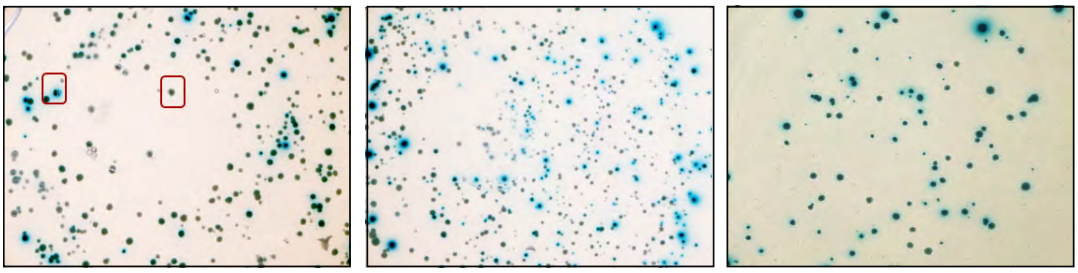

Figure 23: Yeast colonies growing on X-gal containing matrix
Clipping from the image of X-GAL containing plates. Yeast cells (YPH500) were co-transformed with a 1:1 ratio of expression plasmid and reporter plasmid. Cells were allowed to grow for 5 days (30°C) on media plates, containing 0.08 mg/ml X-GAL. AvrBs4-MutH on T-3-H (*left*) and AvrBs4-MutH on H-3-T (*center*) resulted in white, light blue, and dark blue colonies. As example, a light blue colony and a dark blue colony are marked in red (*left*). I-SceI on T-3-HH-3-T containing the I-SecI recognition sequence (*right*) let to dark blue colonies.

3.1.1.5 *AvrBs3-PvuII variants*

It has been recently shown that the catalytic unspecific domain of FokI can also be replaced by the DNA specific cleavage domain PvuII [239]. To investigate the efficiency of the "tool box" of TALENs, AvrBs3-PvuII variants (Figure 24) were analyzed using the yeast based recombinant single strand annealing assay.

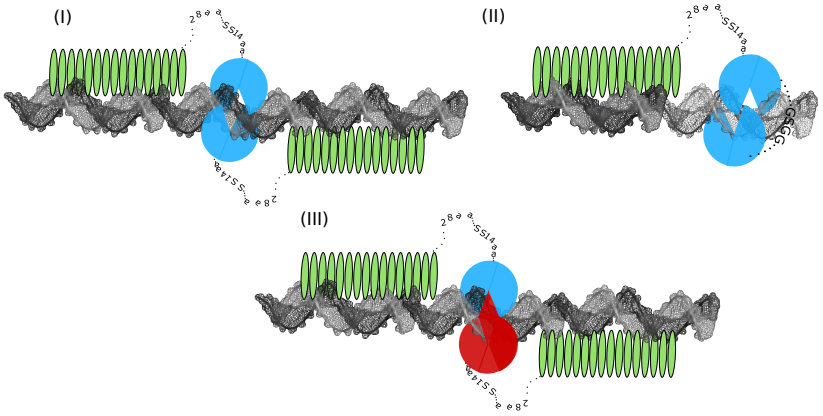

Figure 24: AvrBs3-PvuII variants
The binding domain AvrBs3 is fused to the specific cleavage domain PvuII via a linker containing the remaining 28 aa of the C-terminus from AvrBs3, a 14 aa linker from the FokI protein and two additional serines. The three TALE-PvuII variants analyzed by the SSA assay in *Saccharomyces cerevisiae* are shown. (I) AvrBs3-hf-PvuII, a homodimeric TALEN, containing the high fidelity (hf) variant of PvuII(T46G) and addressing a tripartial substrate of 42 bp. (II) AvrBs3-sc-PvuII holds homodimeric PvuII domains, which are fused to each other by a 4 aa linker (GSGG). The single chain variant target a bipartial sequence with 24 bp and contains the T46G mutation. (III) AvrBs3-hd-PvuII, a heterodimeric TALEN, includes the T46G mutation and either a PvuII domain with the substitution D30K or a converse substitution K38D. Together, the two PvuII variants yield an active PvuII nuclease.

AvrBs3-PvuII coding sequences were imported into the yeast expression plasmid. Similar to the previous TALEN constructs, the AvrBs3 repeat array (Figure 33 (A)) was established by employing the Golden Gate cloning strategy (Methods 2.2.1.1) and introduced into a truncated TALE protein (NΔ152-C28). Additional to the C28 linker from the TALE protein, a 14 aa linker based on the FokI-linker (VIPNRGVTKQLVKG) [225] and two additional serines link the binding module to the cleavage domain. This design was previously used for ZFNs [190] and adopted by Yanik *et al.* [239] for TALE-PvuII variants. Three distinct AvrBs3-PvuII variants were characterized by the single strand annealing assay and are described in Figure 24. (I) AvrBs3-hf-PvuII [239] is a homodimeric TALEN containing the high fidelity (hf) variant of PvuII(T46G) to prevent star site activity and it addresses a tripartial substrate. Based on this architecture, the single chain vari-

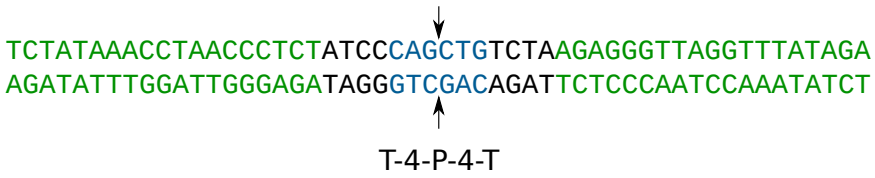

Figure 25: Recognition sequence of AvrB3-PvuII variants
All AvrBs3-PvuII variants were analyzed on the T-4-P-4-T target. Including the 6 bp target site of PvuII (P) flanked by the the target sequence of AvrBs3 (T). The spacer length between the two target sequences is given in bp and the arrow refer to the cutting position of PvuII.

ant (II) AvrBs3-sc-PvuII [239] was designed and includes a T46G substitution. The homodimeric PvuII domains of AvrBs3-sc-PvuII (II) are linked to each other by a 4 aa linker (GSGG) [202] and the monomeric TALEN recognizes a bipartial sequence. (III) AvrBs3-hd-PvuII [238] is a heterodimeric (hd) TALEN carrying a PvuII domain, including a D30K substitution or a PvuII domain, which contains a K38D mutation. Together, the two PvuII domains built an active PvuII cleavage domain. All AvrBs3-variants induce DSBs and were analyzed on the tripartial target T-4-P-4-T (Figure 25), which has been reported to be the optimal substrate design [239].

Beginning with the analysis of viability, equal amounts of expression vector were introduced to the yeast cells (Figure 26 (A)). The I-SceI results were set at 1.0 and all further results were correlated to I-SceI. The total count of yeast colonies of the specific AvrBs3-PvuII variants were compared to the count of yeast colonies of AvrBs4 without nuclease activity. AvrBs3-hf-PvuII (I) and AvrBs3-sc-PvuII (II) resulted in a significant reduction ($P < 0.001$ ***) of the number of yeast colonies, when compared to the colony numbers of AvrBs4. The count of yeast colonies of AvrBs3-hd-PvuII (III) compared to AvrBs4 showed no significant differences. Furthermore, no significant different number of yeast colonies were detected, when single substitutions of the heterodimeric TALEN (e.g. AvrBs3-K38D-PvuII) were analyzed individually and compared to AvrBs4.

Next, the expression vector, as well as the target plasmid were introduced into the yeast cells and the amount of yeast colonies were analyzed (Figure 26 (B)). Again AvrBs3-hf-PvuII (I) and AvrBs3-sc-PvuII (II) yielded a significant reduction ($P < 0.05$ *) of the number of yeast colonies compared to AvrBs4 without nuclease activity. The total amount of yeast colonies of AvrBs3-hd-PvuII and AvrBs3-K38D-PvuII showed a significant decrease ($P < 0.05$ *) when compared to the values of AvrBs4. AvrBs3-D30K-PvuII had no significant different number of yeast colonies, when compared to AvrBs4.

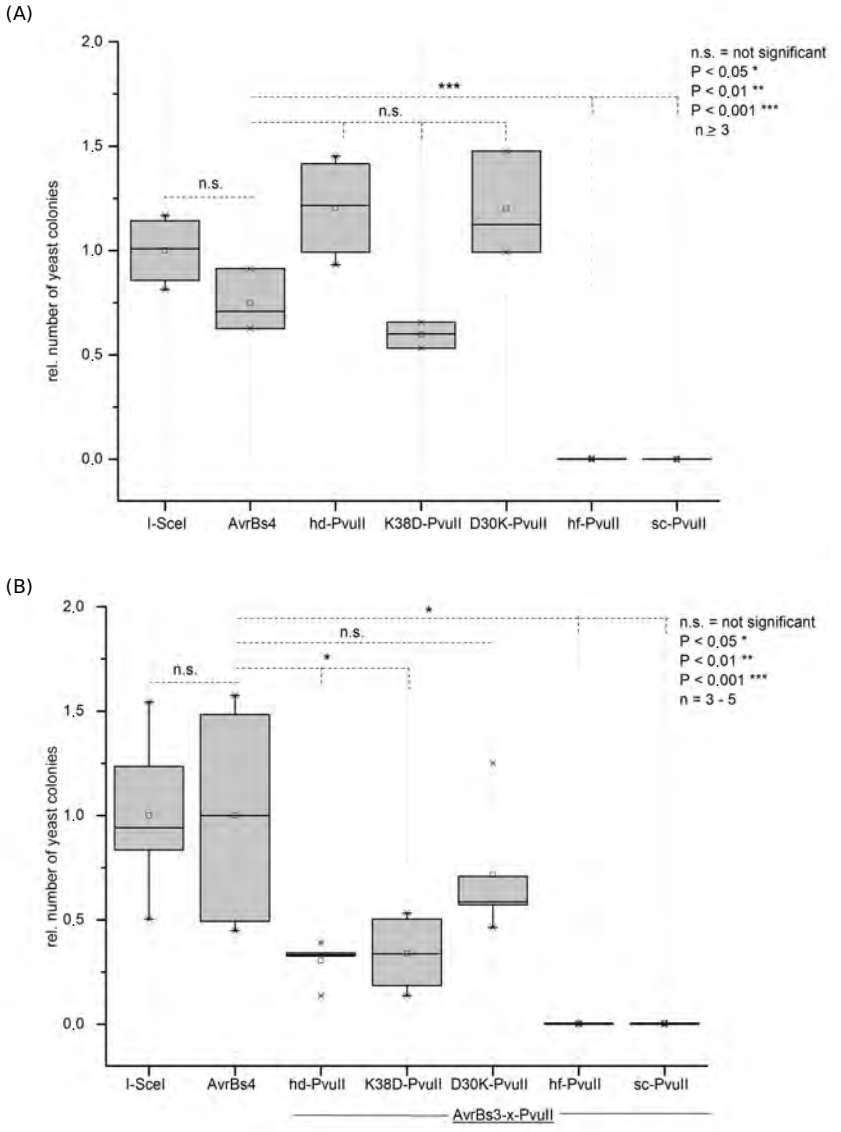

Figure 26: Relative survival of yeast colonies
The number of yeast colonies relative to the results of I-SceI is given. The specific nucleases are compared to AvrBs4 without nuclease activity and significant differences were determined by the unpaired T-test ($n \geq 3$, n.s. = not significant, $P < 0.05$ *, $P < 0.01$ **, $P < 0.001$ ***). (A) Relative amount of yeast colonies transformed with expression plasmid for the specific nuclease without any target. (B) Quantification of yeast colonies transfected with the specific nuclease and the specific target plasmid.

Next, the β-gal activity of the specific nucleases was determined. First, the ratio between white (no functional ß-gal) and blue (functional ß-gal) colonies was calculated and compared to AvrBs4 without nuclease activity (Figure 27). AvrBs3-hf-PvuII (I) and AvrBs3-sc-PvuII (II) revealed no significant different β-gal activity, when compared to AvrBs4 (percentage of blue colonies: 4 ± 2%) , while AvrBs3-

hd-PvuII (III) had a significantly higher β-gal activity ($P < 0.001^{***}$; percentage of blue colonies: 63 ± 10%). Still, AvrBs3-K38D-PvuII reached a higher ß-gal activity (95 ± 3%) when compared to the heterodimeric construct. AvrBs3-D30K-PvuII uncovered no significantly higher β-gal activity (2 ± 1%) compared to the AvrBs4 results. The relative number of white and blue colonies is shown in figure 28. AvrBs3-hf-PvuII (I) and AvrBs3-sc-PvuII (II) revealed no significant differences in enzyme activity when compared to AvrBs4 without nuclease (relative number of blue colonies: 0.03 ± 0.01). AvrBs3-hd-PvuII (III) (relative number of blue colonies: 0.2 ± 0.1) and the individual analysis of AvrBs3-K38D-PvuII (relative number of blue colonies: 0.3 ± 0.2) disclosed a significantly higher number of blue colonies ($P < 0.01^{**}$) compared to the relative number of blue colonies of AvrBs4. The individual analysis of AvrBs3-D30K-PvuII (relative number of blue colonies: 0.01 ± 0.01) showed no significantly increased activity compared to AvrBs4 without nuclease activity.

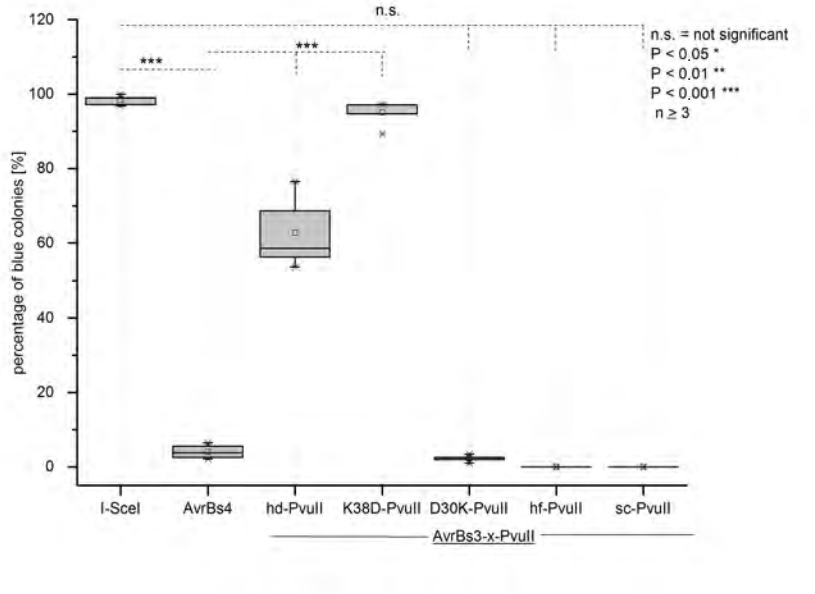

Figure 27: Relative β-gal activity of specific nucleases (I-SceI and AvrBs3-PvuII variants) in *S. cerevisiae* as analyzed by the ratio of blue colonies expressed in percentage [%]
Single Strand Annealing assay in yeast. All results were normalized to I-SceI. Levels of significance were determined by unpaired T-tests ($n \geq 3$, n.s. = not significant, $P < 0.05$ *, $P < 0.01$ **, $P < 0.001$ ***). Percentage of blue colonies [%] of the different specific nucleases. Relative number of white and blue colonies of the nucleases on their specific targets.

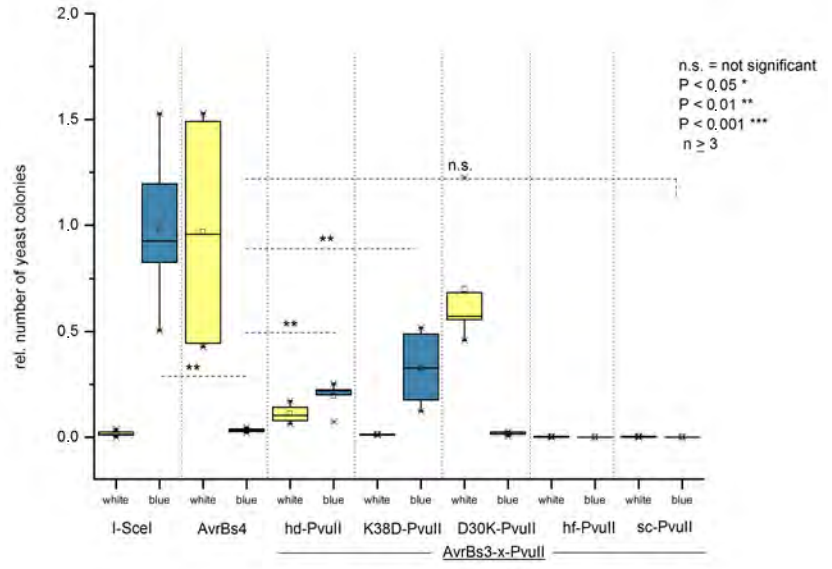

Figure 28: β-gal activity of specific nucleases (I-SceI and AvrBs3-PvuII variants) in *S. cerevisiae*
The specific nucleases are compared to AvrBs4 without nuclease activity and significant differences were determined by the unpaired T-test ($n \geq 3$, n.s. = not significant, $P < 0.05$ *, $P < 0.01$ **, $P < 0.001$ ***). Relative number of white and blue colonies of the nucleases on their specific targets.

3.1.2 *Characterization of RNA guided Cas9 protein in vivo*

Due to simple handling and the ability of multiplexing RNA guided Cas9 nuclease appears to be a promising new specific nuclease [127, 191, 249]. The β-gal activity of a type II bacterial Cas9 protein on a target overlapping the I-SceI recognition sequence (Figure 29), was analyzed by employing the SSA assay in *S. cerevisiae* (Figure 30).

The spCas9 kindly supplied by George Church [58] addresses a 20 bp sequence overlapping the I-SceI target in the T-3-H-3-T substrate. At the end of the 20 bp target sequence, the required 5'-NGG protospacer-adjacent motif (PAM) is placed (Figure 30). The architecture of the gRNA is based on the empty *S. cerevisiae* expression vector from Timothy Lu [69]. The 20 nucleotides (nt) complementary to the DNA target were introduced to the empty expression vector by using the Gibson assembly cloning strategy. Analyzing the Cas9::gRNA system in yeast revealed a significantly higher β-gal activity ($P < 0.001^{***}$; β-gal activity: 23 ± 0.1%) compared to AvrBs4 without nuclease activity (Figure 30).

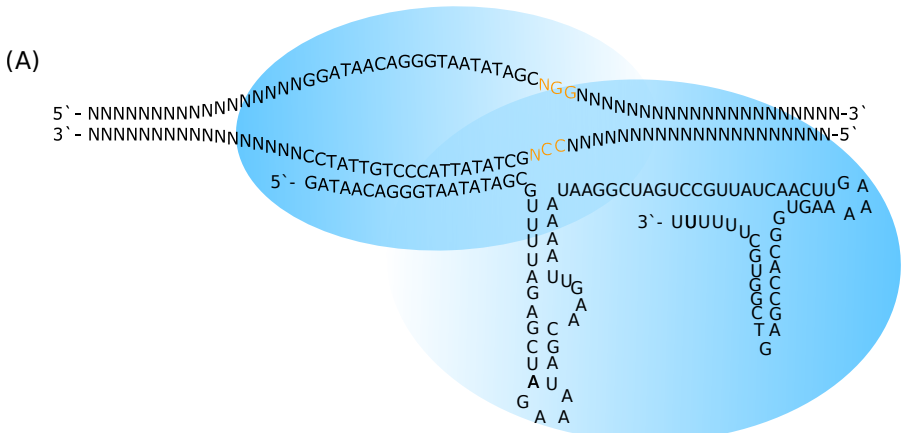

Figure 29: Architecture of Cas9::gRNA and its recognition sequence
(A) Cas9::gRNA addresses a 20 bp target which overlap with the I-SceI target sequence in T-3-H-3-T. At the end of the 20 bp target the 5'-NGG protospacer-adjacent motif (PAM) is placed. gRNA, carrying 20 nt complementary to the DNA target sequence recognizes its target via Watson-crick base pairing. (B) Target of I-SceI is illustrated in bold letters. The addressed sequence of Cas9::gRNA is underlined in black and the PAM sequence is displayed in orange. The addressed site is positioned 9 bp in front of the target site (T-3-H-3-T) of AvrBs4-MutH.

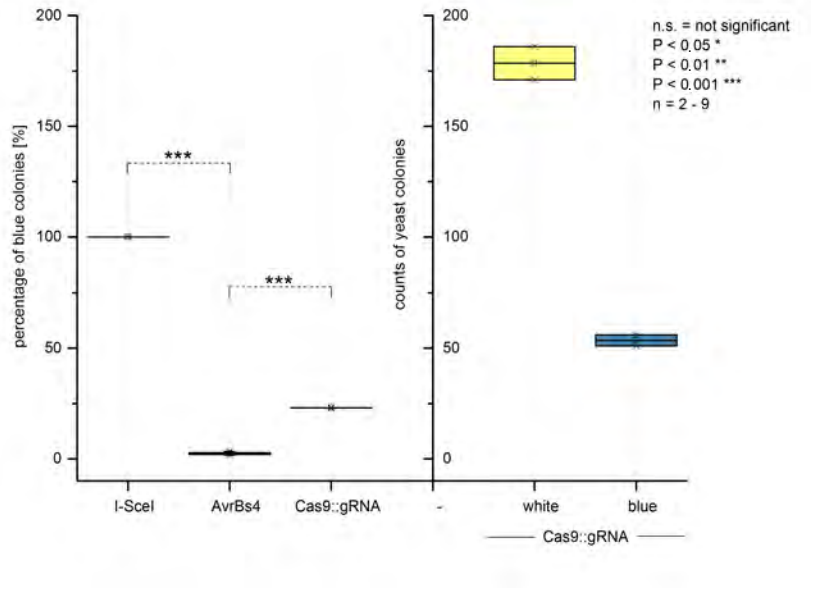

Figure 30: ß-gal activity of Cas9::gRNA in *S. cerevisiae*
On the left site the ß-gal activity of Cas9::gRNA is compared to AvrBs4 without nuclease activity and significant differences were determined by the unpaired T-test ($n \geq 3$, n.s. = not significant, $P < 0.05$ *, $P < 0.01$ **, $P < 0.001$ ***). On the right site the absolute counts of white and blue colonies are shown.

3.1.3 *Alignment search of the addressed sites in the genome of Saccharomyces cerevisiae*

One key limitation of the specific nucleases is off-target activity. We therefore searched the yeast genome for recognition sites analyzed in the SSA assay. Alignments in the genome of *S. cerevisiae* to the addressed sequences were search by using the Basic Local Alignment Search Tool (BLAST) [5]. The S288c strain is the most widely applied yeast strain and its genome has been completely sequenced. YPH500 is driven from the YNN216 strain, which is congenic to the S288c strain and differs only in one locus and a linked segment of chromosome [200].

First, no hits for any of the recognition sequences were found within the genome of *S. cerevisiae* holding an identity of 100%. It has been described that specific nucleases can address a target site including mismatches. How many mismatches are tolerated depends on the family of the specific nuclease and the location of the mismatch within the target sequence. Beginning with the I-SceI recognition site, which consist of 18 bp, the best hit was found in the chromosome XII (National Center for Biothechnology Information (NCBI) reference sequence: NC_001144.5) with an identity of 77% (14 nt /18 nt), but missing the 3'-end. It is known, that I-SceI has a two step recognition mechanisms. First, a strong binding take place to the 3'-end of the target (seed region) followed by binding of I-SceI to the 5'-end region [168]. Consequently, mismatches at the 3'-end have a high impact on the specificity of I-SceI. Thus, alignments were searched which included at least the 9 bp seed region at the 3'-end. The two best alignments, including the 3'-end, were identified in the chromosome XVI (NCBI reference sequence: NC_001148.4) (identity: 66% (12 nt / 18 nt); Table 20). Furthermore, single hits were found with an identity to the I-SceI target sequence of 61% (11 nt / 18 nt) in chromosomes XII (NCBI reference sequence: NC_001144.5) and XIII (NCBI reference sequence: NC_001145.3).

Next, the target sequence of AvrBs4-MutH (T-3-H) was analyzed. In contrast to I-SceI, TALE proteins bind more efficiently to the 5'-end of the target [142]. There were several hits to the AvrBs4 target with an identity of 68% (13 nt / 19 nt), including the 10 bp 5'-end of the AvrBs4 recognition site (Table 20). But there was no AvrBs4 recognition site next to a GATC site with a distance of 3 bp. Furthermore, there was no sequence containing the GATC site and the 5'-end of the TALE recognition site.

Table 20: Identity of the best hits enclosing the seed region of the specific target site
The size of the specific target site of each investigated specific nuclease are given in bp. The seed region can be located at the 5'end or at the 3'end of the recognition site. TALE binding site in combination with the additional specific sequence of the nuclease do not exist (n.e.: non-existence).

		identity [%] of the best hits enclosing seed region		
specific nuclease	target size [bp]	5'end seed region	3'end seed region	+ nuclease target
I-SceI	18 bp	-	66	-
AvrBs4-MutH	18 bp	68	-	n.e.
AvrBs3-PvuII	18 bp	68	-	n.e.
Cas9::gRNA	20 bp + 5' NGG	-	52	-
	20 bp + 5'NAG	-	52	-

Furthermore, a fraction of the target site of AvrBs3-PvuII (T-4-P) was analyzed. One alignment was detected in chromosome XII, including the 10 bp 5'-end containing an identity for the AvrBs3 sequence of 68% (13 nt / 19 nt), but not containing the PvuII-site (Table 20). The best hit to the AvrBs3 target site was found in chromosome VII with an identity of 79% (15 nt / 19 nt), yet missing the 5'-end sequence and the PvuII recognition site.

Finally, the sequence addressed by Cas9::gRNA was analyzed. To target a sequence with the Cas9::gRNA construct a 5'-NGG a 5'-NAG PAM is required [107]. Additionally, Cas9::gRNA is more sensitive to mismatches at the 3'-end (15 bp to the PAM) [108, 47, 107]. The yeast genome was searched for hits including one of the PAMs and parts of the 3'-end target sequence. Three alignments containing the 5'-NGG PAM motif were detected, one in chromosome VIII (NCBI reference sequence: NC_001140.6) with an identity of 52% (12 nt / 23 nt) (Table 20) and two with an identity of 48% (11 nt / 23 nt) in distinct chromosomes. Similar results were found with a 5'-NAG PAM (Table 20). One hit was detected in the chromosome IX (NCBI reference sequence: NC_001141.2) with an identity of 52% (12 nt / 23 nt) and two alignments with an identity of 48% (11 nt / 23 nt) were found in chromosome II (NCBI reference sequence: NC_001143.8) and in chromosome XI (NCBI reference sequence: NC_001143.9).

3.1.4 *Addressing X-linked Retinitis Pigmentosa with specific nucleases*

3.1.4.1 *Design of TALE-MutH proteins addressing the retinitis pigmentosa GTPase regulator (RPGR) gene*

Due to the promising results of the SSA assay, TALE-MutH was redesigned to target the retinitis pigmentosa GTPase regulator (RPGR) gene, which is involved in the inherited eye disease X-linked retinitis pigmentosa (RP) (Table 17). In figure 31 the constructed TALENs are illustrated.

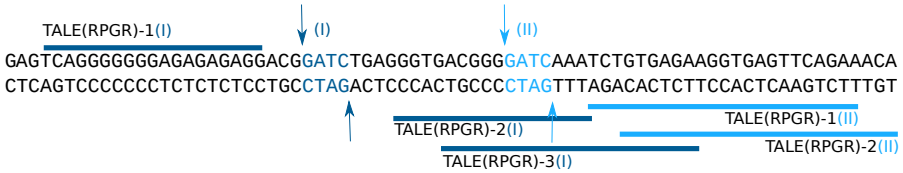

Figure 31: Sequence within the RPGR gene targeted by the designed TALE-MutH variants
The illustrated sequence is a part of ORF15 within the RPGR gene. The GATC site (I) (blue) is addressed by three of the five constructed TALE-MutH variants (TALE(RPGR)-1 to 3 (I)). Two of them (TALE(RPGR)-1(I) and TALE(RPGR)-2(I) address a target site which contain a 3 bp spacer between the TALE recognition site and the GATC site (I). TALE(RPGR)-3(I) recognizes a sequence including a 6 bp spacer between the two recognition sites. For the second GATC site (II) (light blue) two additional TALE-MutH variants were created. Recognition site of TALE(RPGR)-1(II) includes a 3 bp spacer and is therefore nicked on the sense strand. Due to its 6 bp spacer TALE(RPGR)-2(II) nicks the anti-sense strand on the target sequence.

Similar to the naturally occurring TALE binding domains (AvrBs4 and AvrBs3), the new binding domains were reconstructed with the GOLDEN GATE (methods 2.2.1.1) cloning strategy and introduced into an eukaryotic expression vector containing the coding sequence of MutH. Mutations causing X-linked retinitis pigmentosa are located inside the ORF15 of the RPGR gene. The TALENs were designed to address a sequence within the ORF15 nearby a GATC site. A correct DNA matrix will be introduced to the gene for gene editing. This can occur by inducing DSBs, SSBs, or by double nicking. Three TALE-MutH were designed to address the GATC site (I). TALE(RPGR)-1(I)-MutH and TALE(RPGR)-2(I)-MutH contain a 3 bp spacer on their addressed sites in between the recognition site of TALE and the GATC site. The design is based on the target site construction T-3-H-3-T. TALE(RPGR)-3(I)-MutH also addresses the GATC site (I), but recognizes the target site with a 6 bp spacer. This result in two heterodimeric TALE(RPGR)-MutH (TALE(RPGR)-1(I)-MutH / TALE(RPGR)-2(I)-MutH and TALE (RPGR)-2(I)-MutH / TALE(RPGR)-3(I)-MutH), which are able to induce double nicking within the ORF15. The second GATC site (II) can be addressed by

TALE(RPGR)-1(II)-MutH and TALE(RPGR)-2(II)-MutH. TALE(RPGR)-1(II)-MutH addresses the GATC site (II) with a 3 bp spacer in between the TALE target site and GATC site, whereas TALE(RPGR)-2(II)-MutH contains a spacer distance of 6 bp in between the TALE target site and GATC site. The heterodimerc TALE-MutH introduces a nick into the bottom strand and into the upper strand. Additionally, both GATC sites (I and II) can be addressed (e.g. TALE(RPGR)-1(I)-MutH / TALE(RPGR)-1(II)-MutH) resulting in a double nick, 13 bp apart from each other. Finally, all TALE-MutH constructs can also be applied individually to induce one SSB into the ORF15. Two more GATC sites are located in the ORF15, which represent further possible target for TALE-MutH constructs.

3.1.4.2 *Characterization of TALE-MutH specific nucleases for the RP GTPase regulator gene*

TALE-MutH constructs addressing RP GTPase regulator gene were analyzed together with Dr. Mert Yanik in the retinal gene therapy lab of Prof. Dr. Dr. Knut Stieger.

The heterodimer TALE(RPGR)-1 (I) -MutH/ TALE(RPGR)-2 (I) -MutH described in section 3.1.4.1 were analyzed under *in vivo* conditions in cooperation with the group Retinal gene therapy of Prof. Dr. Dr. Knut Stieger at the University of Giessen (FB 11). The specific nucleases were analyzed together with Dr. Mert Yanik episomal in HEK293T cells. Due to a DNA break NHEJ is triggered and a frame shift occurs, which results in a functional gene expressing blue fluorescent protein (BFP). The BFP was detected by flow cytometry and the results were compared to the GOLDEN STANDARD I-SceI (Figure 32 (A)). Analysis of TALE(RPGR)-1 (I) and TALE(RPGR)-2 (I) addressing the GATC (I) site revealed a similar NHEJ activity, when compared to I-SceI.

Additionally, cell toxicity studies were performed, by co-transfection of mCherry and the specific nucleases expression plasmids (Figure 32 (B)). A pair of ZF-FokI [6] and the TALE-MutH pair (TALE(RPGR)-1 (I)/TALE(RPGR)-2 (I)) were analyzed using the cell toxicity assay.

The empty control vector (pcDNA3) served as negative control. The ZF-FokI (survival rate: 42 ± 18%) pair led to a significantly reduced survival rate when compared to I-SceI (survival rate: 100 ± 8%). Whereas the heterodimeric TALE(RPGR)-MutH (survival rate: 89 ± 15%) resulted in a comparable survival rate to I-SceI. The empty vector pcDNA3 (survival rate: 100 ± 8%) showed a similar survival rate as I-SceI.

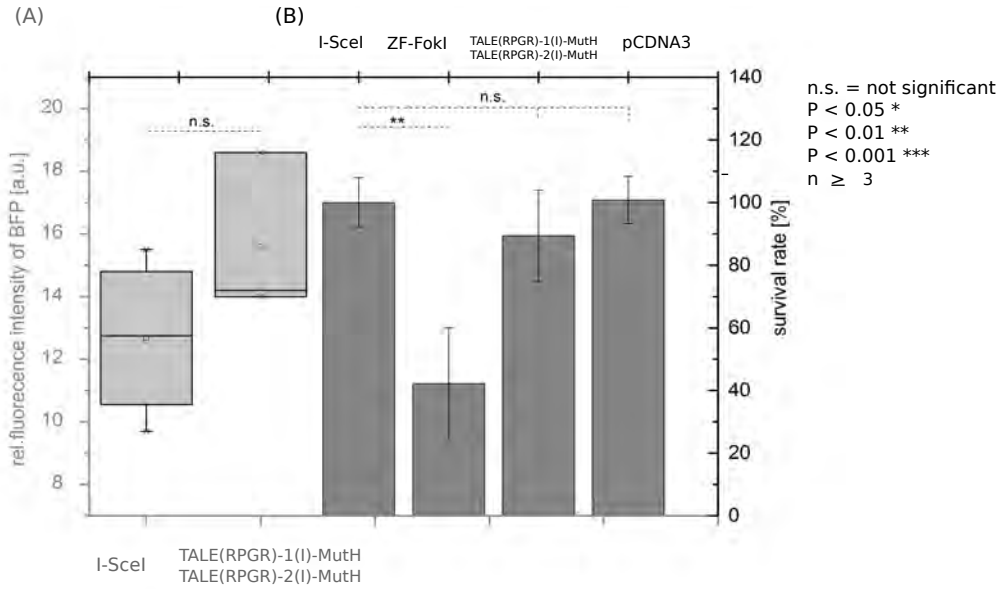

Figure 32: Analysis of TALE-MutH addressing GATC (I) site within the ORF15 of the RPGR gene

Levels of significance were determined by unpaired T-tests ($n \geq 3$, n.s. = not significant, $P < 0.05$ *, $P < 0.01$ **, $P < 0.001$ ***). (A) NHEJ activity analysis of TALE(RPGR)-1 (I) and TALE(RPGR)- 2 (I) addressing the GATC (I) site, compared to NHEJ activity of I-SceI. NHEJ activity result in a functional gene, expressing BFP. (B) Cell toxicity was analyzed by the co-transfection of mCherry and specific nucleases expression plasmids. Survival rate was defined as the decrease of mCherry-positive cells from day 2 to day 5. Number of mCherry-positive cells were detected by flow cytometry and data were normalized to the I-SceI results. The empty vector (pcDNA3) served as negative control.

3.2 CHARACTERIZATION OF BINDING CHARACTERISTICS OF AVRBS3

This chapter provides a detailed analysis of the TALE binding protein AvrBs3(△ N152-C28). The AvrBs3(△ N152-C28) protein was redesigned to enable labeling at specific positions in the protein, thus resulting in several AvrBs3 variants. These new designed TALEs provide promising possibilities to analyze the DNA binding properties of TALEs and facilitate the analysis of the off-target problematic.

3.2.1 *Variants of AvrBs3*

The modified protein AvrBs3(△ N152-C28) is a truncated version of the naturally occurring AvrBs3. The N-terminus lacks 152 aa and the C-terminus holds 28 aa. All AvrBs3(△ N152-C28) variants carry a strep-tag at the N-terminus and were generated by the GOLDEN GATE cloning strategy (methods 2.2.1.1). Based on the mechanism of a thiol Michael-type addition reaction cysteine residues can react with the double bond of maleimide to succinimide-thiolethers. This mechanism can be used to label cysteine residues with maleimide-fluorophores. Thus, various AvrBs3 mutants were created and labeled at distinct positions (Figure 33).

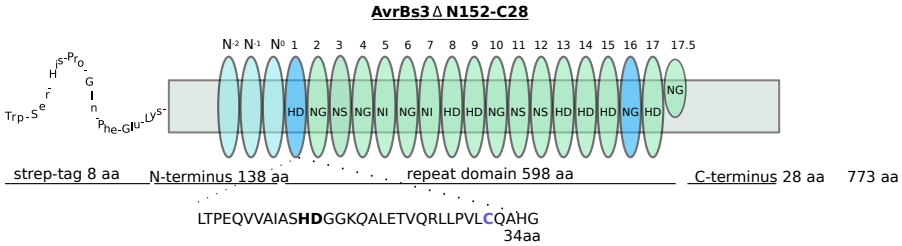

Figure 33: Architecture of AvrBs3 variants
All AvrBs3 variants contain a N-terminus with 138 aa, a repeat domain with 598 aa and a C-terminus with 28 aa. The Strep-tag is fused to the N-terminus, which comprises N_0, N_{-1} and N_{-2}. The binding domain encloses 17.5 repeats. Each repeat holds 34 highly conserved aa. In position 12 and 13 the aa are crucial for DNA recognition (**bold**). Each repeat contains one cysteine residue located at position 30. In the newly constructed variants, cysteine residues are mutated to methionine in each repeat except for repeat number 1 and/or 16 (blue) depending on the specific AvrBs3 variants (AvrBs3 (*A*), AvrBs3 (*B*), and AvrBs3 (*AB*)).

The AvrBs3 binding domain contains 17.5 repeats and each repeat consists of 34 highly conserved aa, which enclose a cysteine residue in position 30 (Figure 34 (A)). To permit labeling at defined positions, the cysteine residues in the specific repeats were substituted to methionine residues via site-directed mutagenesis PCR.

The web server PoPMuSiC [55] was used to estimate the stability change of these single-site mutations of cysteine. It revealed that methionine has the smallest shift in its thermodynamic stability (ΔΔ G = 0.3 [kcal/mol]) when compared to cysteine (Figure 34 (B)). While methionine belongs to the non-polar aa and cysteine to the polar aa, they both include an alkyl chain and a sulfur atom.

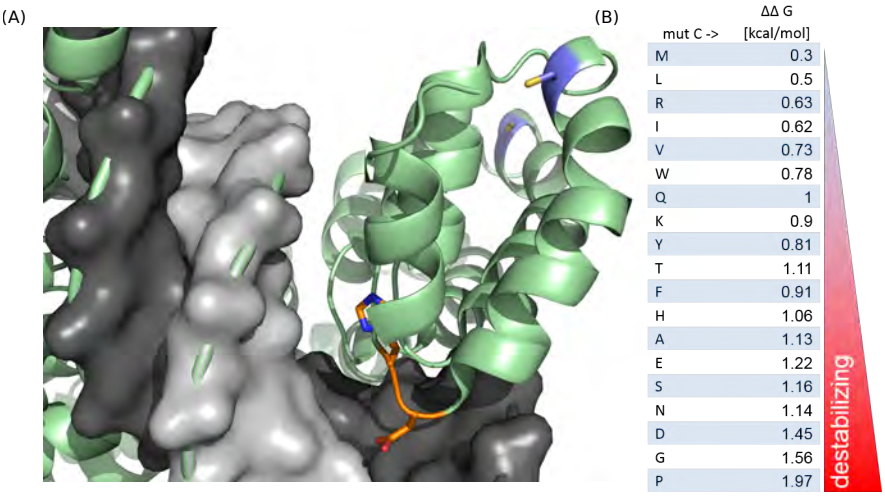

Figure 34: Single-site mutation of the cysteine residue
(A) Crystal structure of AvrBs3 bound to DNA with a focus to one repeat is shown (pdb: 2YPF) [205]. The alpha-turn-alpha structure of this repeats is exhibited and aa 12 (H) and 13 (D) (orange) recognize the nucleotide cytosine. The cysteine residue at position 30 in each repeat is marked in blue and its thiol group in orange. (B) Prediction of AvrBs3 mutant stability changes ΔΔ G [kcal/mol]. Methionine has the lowest stability change with ΔΔ G = 0.3 [kcal/mol].

Having the choice between plasmids encoding each repeat either with or without cysteine residues allowed to generate AvrBs3 variants with different numbers and positions of cysteine residues in its binding domain. Beside the variant AvrBs3(△ N152-C28), which contains a cysteine residue in each repeat, three additional AvrBs3(△ N152-C28) variants were created (Figure 33). Two of theses AvrBs3(△ N152-C28) variants ((*A*) and (*B*)) contain a single cysteine residue in the entire protein and all remaining cysteine residues were mutated to methionine. AvrBs3 (*A*) encloses a single cysteine residue in its first repeat and AvrBs3 (*B*) has its single cysteine residue in repeat sixteen. These two variants enable to label AvrBs3(△ N152-C28) with a single fluorophore at the beginning of the binding domain (AvrBs3 (*A*)) or at the end of the binding domain (AvrBs3 (*B*)). Furthermore, an AvrBs3(△ N152-C28) variant was generated which allows labeling simultaneously at two positions in the binding domain (AvrBs3 (*AB*)). This variant carries two cysteine residues, one cysteine is located in the first repeat and one cysteine in repeat sixteen.

To enable labeling, the cysteine residue has to be accessible for the used dye in a way that it does not influence the protein functionality. Regarding the crystal structure of AvrBs3 the cysteine residue at position 30 in each repeat is in an exposed position and should be available for labeling. Figure 34 (A) shows the crystal structure of one repeat unit and the location of the cysteine residue is marked.

Purification of the AvrBs3(△ N152-C28) variants was examined by affinity purification using a strep-tag fused to their N-terminus. In figure 35 (A), an example of a strep-tag purification of AvrBs3 (***B***) analyzed with a 12% SDS-polyacryamide gel electrophoresis (SDS-PAGE) is given. AvrBs3 protein has a molecular weight of 80 kDa and was analyzed on the 12% SDS-polyacrylamid gel above 85 kDa. TALEs used for characterization of unlabeled AvrBs3 variants (e.g. determination of K_D-values) were purified in an additional step, using heparin sepharose affinity chromatography. The AvrBs3(△ N152-C28) variants eluded at a NaCl concentration of 285 nM (Figure 35 (B)).

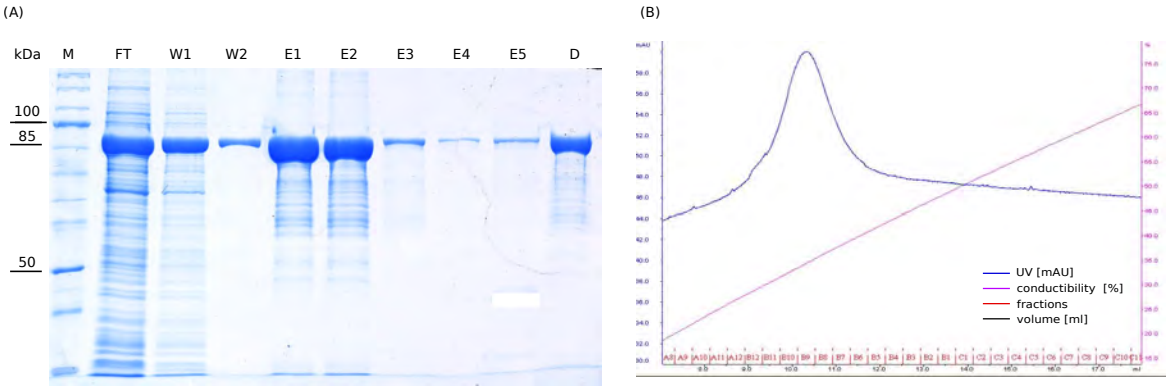

Figure 35: Purification of AvrBs3 (***B***)
(A) Strep-tag purification, taking the example of AvrBs3. This 12% SDS-polyacrylamide gel shows the marker (PageRuler^{TM}Unstained Protein Ladder) and each working step of the strep-tag purification: Marker (M), flow through (FT), washing step 1 and 2 (W1 and W2), elution steps (E1-5), and dialysis (D). AvrBs3 band appears above the 85 kDa band. (B) Heparin sepharose affinity chromatography: In a second step AvrBs3 was purified over a heparin sepharose column. A conductibility gradient with high salt buffer B (0-100%) was created and AvrBs3 eluted at a NaCl concentration of 285 nM. Blue = absorption [mAU] at 280 nm; magenta = conductivity [%], red = fractions.

3.2.2 *Binding character of the AvrBs3 variants*

The K_D-values of AvrBs3(△N152-C28) and AvrBs3 (***B***) were determined by fluorescence anisotropy (Figure 36 and Figure 37). Protein was added in increasing concentrations to 5'-end Hex labeled specific DNA (0.5 nM). The K_D-values of both AvrBs3 proteins were about 1 nM (AvrBs3(△N152-C28): 1.9 ± 0.8 nM , AvrBs3 (***B***): 0.9 ± 0.3 nM (Table 21; Figure 36).

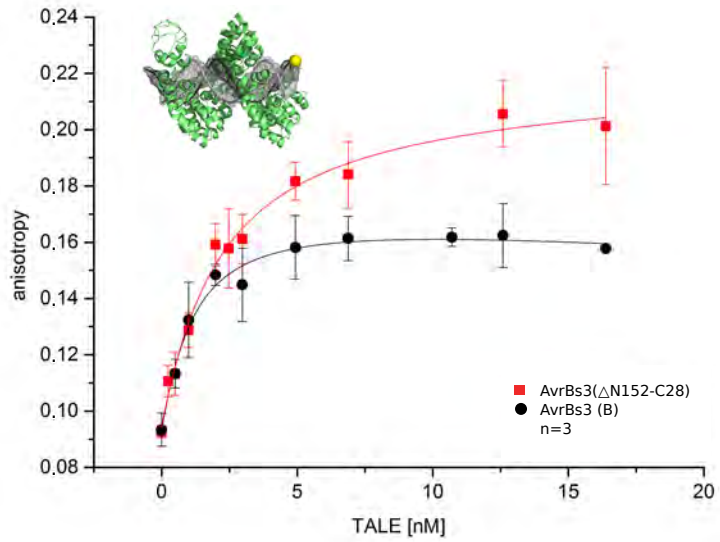

Figure 36: K_D-values of AvrBs3(△N152-C28) and AvrBs3 (***B***)
K_D-values were analyzed by fluorescence anisotropy. A 25 bp long specific DNA (0.5 nM) labeled with HEX at the 5'-end of the bottom strand was used. Labeling position of HEX is marked in yellow in the crystal structure of AvrBs3 bound to specific DNA (pdb: 2YPF) [205]. TALE variants were titrated with increasing concentration to specific DNA. Each measuring point represents the mean ± standard error indicated by error bars for each point. n : number of independent experiments. Measuring points were fitted with a non-linear regression. K_D-values determined: 1.9 ± 0.8 nM (AvrBs3(△N152-C28)) and 0.9 ± 0.3 nM (AvrBs3 (***B***)).

The active fraction of protein was determined by titrating the protein with a total DNA concentration (10 nM) much higher than the K_D-values (Figure 37). Under these conditions the K_D-value cannot be determined, but the number of binding positions can be analyzed. The experiment revealed, that 13% of AvrBs3(△N152-C28) and 15% of AvrBs3 (***B***) of the purified stock solutions were active. The calculated K_D-values of 100% active protein were 0.23 ± 0.07 nM for AvrBs3(△N152-C28) and 0.14 ± 0.05 nM for AvrBs3 (***B***) (Table 21).

Next, the rate constants for dissociation were analyzed for the AvrBs3 variants. Here, the 5'-end HEX labeled specific DNA competed with

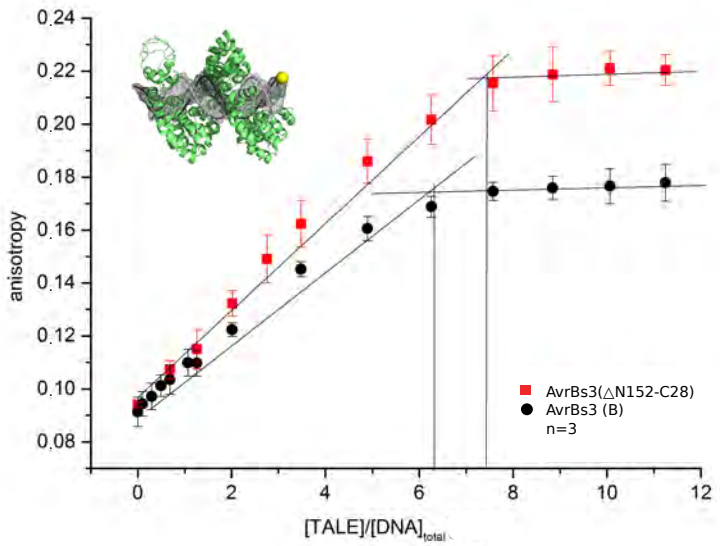

Figure 37: Determination of active fractions of AvrBs3(△N152-C28) and AvrBs3 (*B*)
To determine the active fraction of TALE variants a 5'-end Hex labeled specific DNA with a concentration (10 nM) much higher then the K_D-values was applied. Labeling position of HEX is marked in yellow in the crystal structure of AvrBs3 bound to specific DNA (pdb: 2YPF) [205]. TALE variants were titrated with increasing concentration to the HEX labeled specific DNA. Each measuring point represents the mean ± standard error indicated by error bars for each point. n : number of independent experiments. Active fractions: 13% (AvrBs3(△N152-C28)) and 15% (AvrBs3(B)).

unlabeled DNA (Figure 38 (A)). HEX labeled specific DNA (10 nM) was incubated for 5 minutes with an excess of TALE (50 nM). Next, unlabeled-specific DNA or -scramble DNA was added to compete with HEX labeled specific-DNA (Figure 38 (B)). As a negative control, buffer was added instead of unlabeled specific/scramble DNA. Rate constant of AvrBs3 (*B*) for scramble DNA was k_{-1}= 9.6 x 10^{-3} ± 1.3 x10^{-3} $s^{-1}$with t = 103 ± 11 s and rate constant for specific DNA was k_{-1}= 1.3 x 10^{-3} ± 4.2 x 10^{-5} $s^{-1}$with t = 795 ± 27 s (Table 21). Competition with unlabeled scramble DNA was fitted by a one-phase exponential decay function, whereas unlabeled specific DNA was fitted by two-phase exponential decay function.

Almost identical rate constants were determined by adding unlabeled scramble DNA and specific DNA consecutive to labeled DNA with an excess of AvrBs3 (*B*). Rate constant of AvrBs3 (*B*) for scramble DNA was k_{-1}= 9.7 x 10^{-3} ± 1.1 x10^{-3} $s^{-1}$with t = 102 ± 11 s and rate constant for specific DNA was k_{-1}= 1.5 x 10^{-3} ± 1.1 x 10^{-4} $s^{-1}$with t = 650 ± 45 s.

Consecutively to the competition experiment, AvrBs3 (***B***) was digested by Proteinase K (PK) and baseline values of DNA with unbound protein were reached (Figure 39 (A); Table 21).

AvrBs3(△N152-C28) (k_{-1}= 2.1 x 10^{-3} ± 3.8 x10^{-3} s^{-1}, t = 465 ± 83 s) competed with unlabeled unspecific DNA 4.5 time slower, compared to AvrBs3 (***B***) (k_{-1}= 9.7 x 10^{-3} ± 1.1 x10^{-3} s-1, t = 102 ± 27 s) (Figure 39 (B), Table 21). Competition to unlabeled specific DNA could not be fitted.

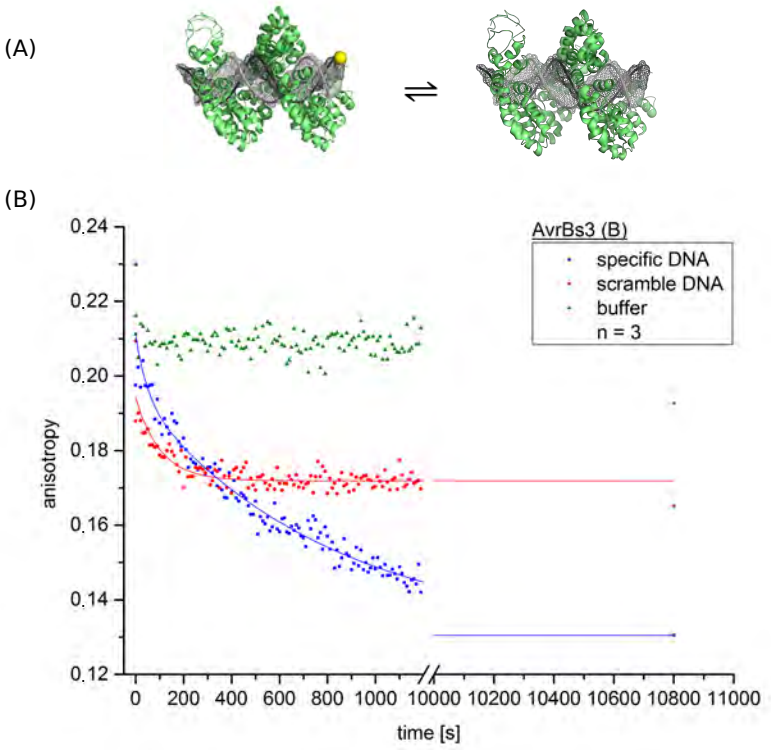

Figure 38: Individual competition of AvrBs3 (***B***).
(A) Crystal structure of AvrBs3 (pdb: 2YPF) [205] binding to 23 bp specific labeled DNA and unlabeled DNA. Labeling position of HEX is shown in yellow. (B) Individual competition: specific HEX labeled DNA was incubated with AvrBs3 (***B***) (50 nM). Then unlabeled specific DNA (180 nM; blue), unlabeled scramble DNA (180 nM; red), or buffer (green) was added to compete with specific DNA. Competition with unlabeled scramble DNA is fitted with a one-phase exponential decay function and competition with unlabeled specific DNA is fitted with a two-phase exponential decay function. n : number of independent experiments. Rate constant of AvrBs3 (***B***) with scramble DNA: k_{-1}= 9.6 x 10^{-3} ± 1.3 x10^{-3} $s^{-1}$with t = 103 ± 11 s. Rate constant of AvrBs3 (***B***) with specific DNA: k_{-1}= 1.3 x 10^{-3} ± 4.2 x 10^{-5} $s^{-1}$with t = 795 ± 27 s.

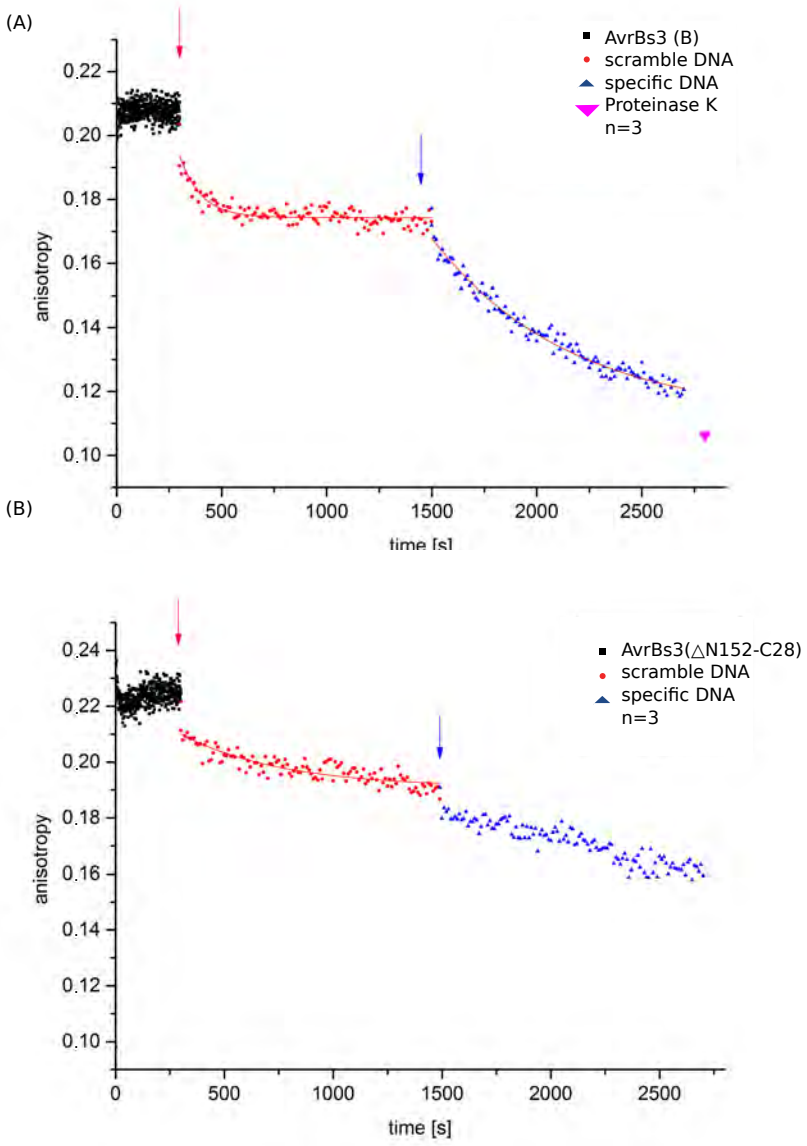

Figure 39: Consecutive competition of AvrBs3 (***B***)
(A) Specific HEX labeled DNA was incubated with AvrBs3 (***B***) (50 nM) for 5 min (black), then unlabeled scramble DNA (180 nM) was added and fluorescence anisotropy changes were measured for 20 min (red). Thereafter, unlabeled specific DNA (180 nM) was added and fluorescence anisotropy was monitored over further 20 min (blue). At the end Proteinase K was added to digest the AvrBs3 (***B***) and the starting value was analyzed (magenta). Data points were fitted with a one-phase exponential decay function. Rate constant (B) Specific HEX labeled DNA was preincubated with AvrBs3 (***B***) (50 nM) for 5 min (black), then unlabeled scramble DNA (180 nM) was added and fluorescence anisotropy changes were measured for 20 min (red). Thereafter, unlabeled specific DNA (180 nM) was added and fluorescence anisotropy was detected over further 20 min (blue). Data points were fitted with an one-phase exponential decay function. Competition to unlabeled specific DNA could not be fitted.

Table 21: The K_D-values of AvrBs3(△N152-C28) and AvrBs3 (***B***)
The K_D-values of the TALE variants are listed for direct comparison. Considering the determined active fraction of 13% (AvrBs3(△N152-C28)) and 15% AvrBs3 (***B***) the K_D-values improve to 0.24 ± 0.09 nM (AvrBs3(△N152-C28)) and 0.14 ± 0.05 nM (AvrBs3 (***B***)) for 10% active TALE. Rate constants of dissociation for AvrBs3(△N152-C28) and AvrBs3 (***B***) for individual competition assay (1) and consecutive competition experiment (2) are also given.

		AvrBs3(△N152-C28)	AvrBs3 (***B***)
K_D [nM]		1.9 ± 0.8	0.9 ± 0.3
K_D [nM] 100% active protein		0.24 ± 0.09	0.14 ± 0.05
t [s];k^{-1} [s^{-1}] (1)	scramble DNA	-	103 ± 11 9.6 x 10^{-3} ± 1.3 x 10^{-3}
	specific DNA	-	795 ± 27 1.3 x 10^{-3} ± 4.2 x 10^{-5}
t [s]; k^{-1} [s^{-1}] (2)	scramble DNA	465 ± 83 2.1 x 10^{-3} ± 3.8 x10^{-4}	102 ± 11 9.7 x 10^{-3} ± 1.1 x 10^{-3}
	specific DNA	-	650 ± 45 1.5 x 10^{-3} ± 1.1 x 10^{-4}

3.2.2.1 *Electrophoretic mobility shift assay of AvrBs3 (AB)*

Binding of AvrBs3 (***AB***) on a 5´-end Cy3 labeled 25 bp specific fragment was analyzed with an electrophoretic mobility shift assay (EMSA) on a 4% PAA gel. AvrBs3 was added with increasing concentration ranging from 0-600 nM to 5'-end Cy3 labeled specific DNA (20 nM) and mixtures were incubated for 5 min (0°C) (Figure 40). DNA total concentration was much higher then the K_D-value of the protein ($[DNA]_{total} >> K_D$). Thus, the K_D-value cannot be determined by this experimental set up. We expected to gain a curve, which increases linearly to its plateau. A plateau indicating the concentration, where all binding positions on the DNA are occupied. As expected, increasing protein concentration resulted in a linearly increasing curve (Figure 40 (B)).

Two DNA shifts have been detected by EMSA. The shift (I) increases linearly to a plateau of 0.46 ± 0.06 rel. fluorescence intensity, followed by a slight decrease due to the second shift. The shift (II) begins at a protein concentration between 250 nM and 450 nM and does not reach a plateau at 600 nM AvrBs3. In figure 40 (B), shift (I) and (II) are plotted and the rel. fluorescence intensity of shift (I) and (II) are added.

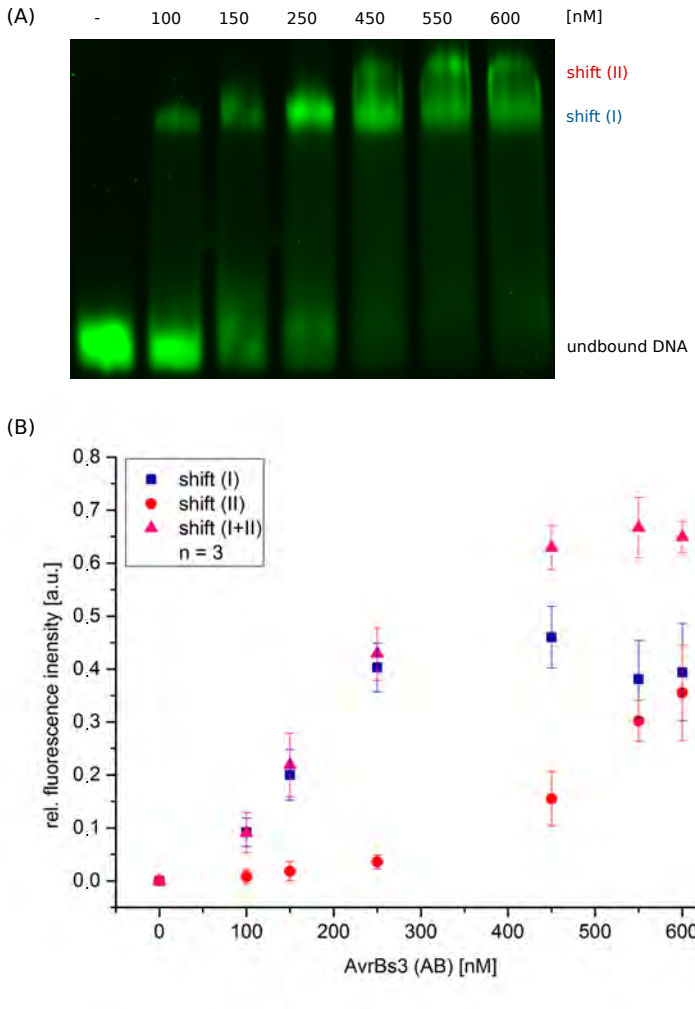

Figure 40: EMSA of AvrBs3 (**AB**)
AvrBs3 (**AB**) in increasing concentrations (0-600 nM) were preincubated for 5 min (0°C) with 20 nM 5´-end Cy3 labeled specific DNA. (A) EMSA was analyzed on a 4% PAA gel and the fluorophore Cy3 was excited and visualized with Typhoon^{TM} FLA 9500 (532 nm). The 5´-end Cy3 labeled specific DNA is shown in green. Three distinct settings were detected, unbound DNA, shift (I) (blue), and shift (II) (red). (B) Relative fluorescence intensity [a.u.] of shift (I) (blue), shift (II) (red), and shift (I+II) (magenta) are plotted.

3.2.3 *Labeled AvrBs3 variants*

Since the mutated AvrBs3(△N152-C28) variants retained their binding characteristic, we analyzed if this is also the case for the labeled AvrBs3(△N152-C28) variants. The single cysteine variants were labeled with the fluorophore Cy5 as well as Alexa 488. Binding efficiencies were detected with anisotropy and FRET distance measurements. The double cysteine variant AvrBs3 (***AB***) was labeled (Alexa 488, Alexa 594, and Alexa 647) and binding characteristics were analyzed with EMSA.

3.2.3.1 *Labeling of single cysteine variants with Alexa 488 fluorophore*

First, the applicability of the fluorophore Alexa 488 for labeling AvrBs3 (***B***) was analyzed. The labeling reaction was performed with an excess of dye (1:4) for 1.30 h and resulted in a degree of labeling (DOL) of 158%. The binding characteristic was analyzed via anisotropy measurements on the 5'-end Hex labeled specific DNA (Figure 41). The K_D-value obtained for AvrBs3 (***B***)-Alexa 488 (K_D = 1.3 ± 0.3 nM) was similar to the K_D-values of the unlabeled AvrBs3 variants.

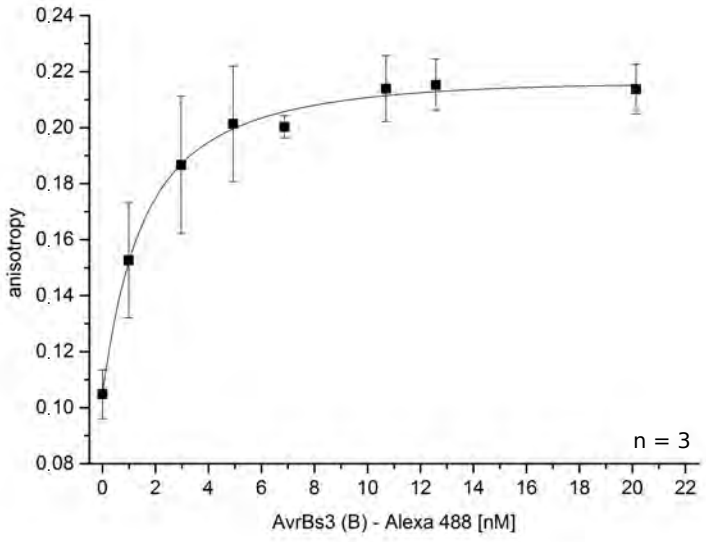

Figure 41: Determination of K_D-value of AvrBs3 (***B***)-Alexa 488
Anisotropy measurement of AvrBs3 (***B***) labeled with the fluorophore Alexa 488 at the 5'-end Hex labeled DNA (0.5 nM). Protein was added with increasing concentrations to the 5'-end Hex labeled specific DNA. Each measuring point represents the mean ± standard error indicated by error bars for each point. n : number of independent experiments. Data points were fitted with a non-linear regression. K_D-value: 1.3 ± 0.3 nM.

3.2.3.2 *Labeling of single cysteine variants with Cy5 fluorophore*

Next, the single cysteine variants were labeled with the maleimide fluorophore Cy5. Labeling reactions were performed with an excess of dye (1:4) for 90 min (Figure 42 (A)). Labeled protein was separated from the free fluorophore using a Zeba Desalting column (40K) (Figure 42 (B)).

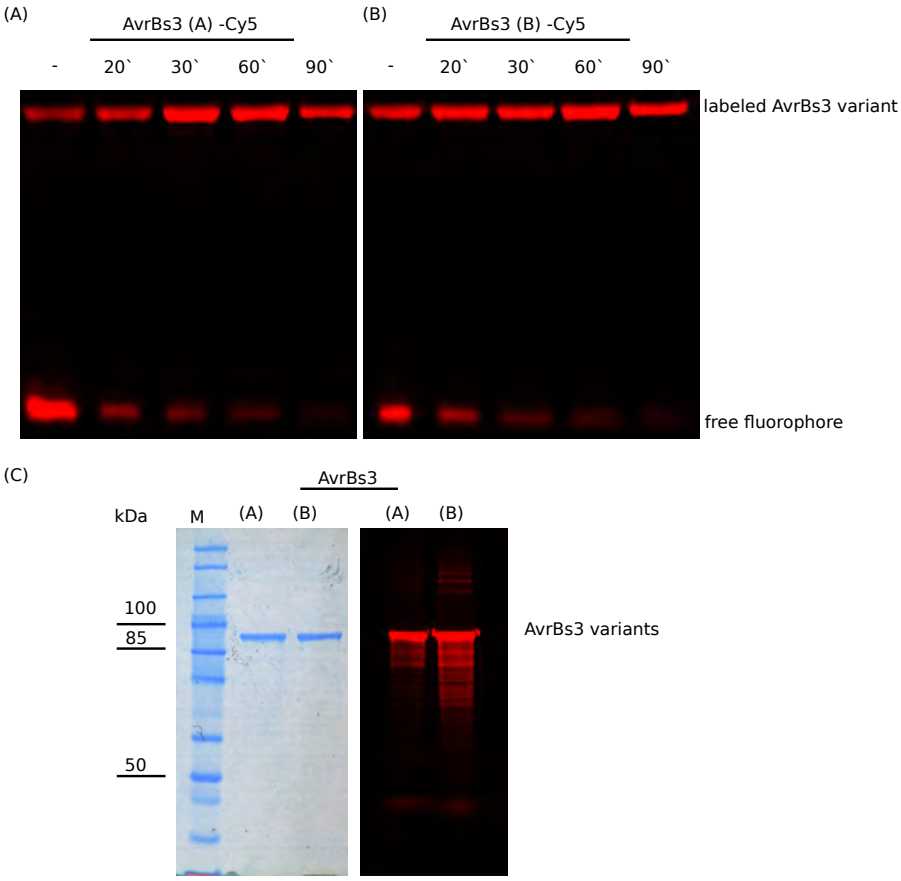

Figure 42: SDS-PAGE of single labeled AvrBs3 variants with flourophore Cy5

Single cysteine variants of AvrBs3 were labeled with an excess of Cy5 maleimide fluorophore (1:4) and analyzed by 12% SDS-polyacrylamide gel electrophoresis. (A) Time resolved kinetics of the labeling reaction of AvrBs3 (***A***) (left) and AvrBs3 (***B***) (right). 2 µg protein was loaded in each lane. (-) represents the time point, when the fluorophore was added to the protein. The respective time points were taken (20', 30', 60', 90') and Cy5 fluorophore was excited at 635 nm with the Typhoon^{TM} FLA 9500. (B) Images of the single cysteine AvrBs3 variants (4 µg) after purification with Zeba Desalting spin column (40K). *Left*: Coomassie Brilliant Blue stained 12% SDS-gel with Marker (PageRuler^{TM} Unstained Protein Ladder; lane 1), AvrBs3 (***A***) (lane 2), and AvrBs3 (***B***) (lane 3). *Right*: Fluorescence image of the 12% SDS-gel visualized with Typhoon^{TM} FLA 9500. The SDS-gel showed that AvrBs3 variants was analyzed above 85 kDa.

The Coomassie Brilliant Blue stained SDS-gel revealed that AvrBs3 variants run above 85 kDa (Figure 42 (B, left)). The SDS-gel can be visualized with fluorescence excitation (635 nm) (Figure 42 (B, right)). The degree of labeling of the single cysteine variants was analyzed. AvrBs3 (*A*)-Cy5 revealed a degree of labeling (DOL) of 115% and AvrBs3 (*B*)-Cy5 a DOL of 82%.

3.2.3.3 *Förster resonance energy transfer (FRET) between Cy5 labeled AvrBs3 single cysteine variants and Cy3 labeled DNA*

FRET enables a detailed analysis of the binding character of AvrBs3. The fluorophore Cy5 results in an effective FRET pair with fluorophore Cy3. Here, Cy3 served as a donor fluorophore and Cy5 as the acceptor fluorophore. The FÖRSTER RADIUS of Cy3 to Cy5 is given with 60 Å [96]. Binding characteristics of the two Cy5 labeled single cysteine AvrBs3 variants were analyzed on two distinctively labeled 5'-end Cy3 specific DNAs (1 nM). The proteins were added to each DNA with increasing concentrations. Binding of the AvrBs3 variants to the DNA resulted in a donor quench and an increased indirectly acceptor signal (Figures 43, 54 and 55).

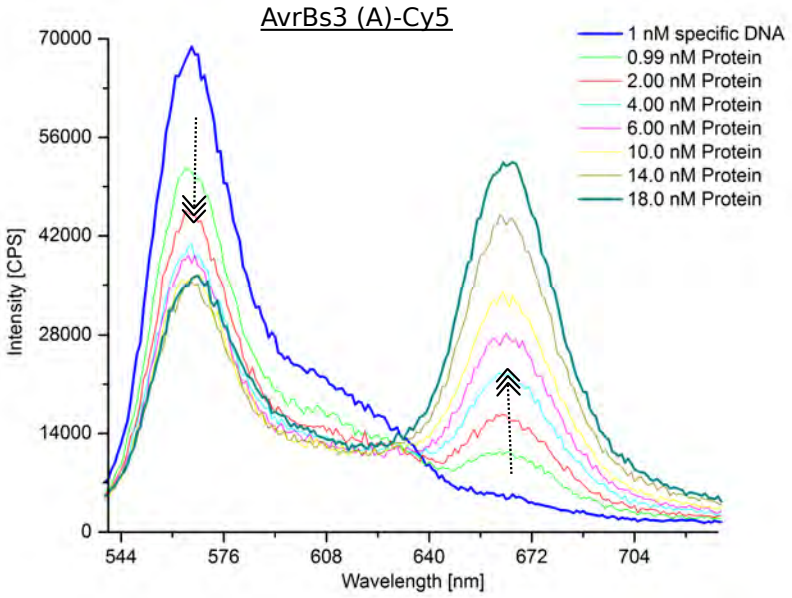

Figure 43: Emission spectra of titrating Cy5 labeled AvrBs3 to 5'-end Cy3 labeled specific DNA (I)
Representative example of emission spectra. AvrBs3 (*A*)-Cy5 analyzed on DNA (I) (1nM). Protein was titrated with increasing concentrations (0, 0.99, 2, 4, 6, 10, 14, 18 [nM]) to 1 nM DNA (I).

This FRET indexes (nF) were retrieved by measuring the acceptor emission resulted by the donor excitation ($F^{ex_D em_A}$) and calculated by the three-cube method described by Youvan *et al.* [242]. This does not represent the FRET efficiency, but provides an index which change with FRET [17, 244]. The FRET signal had to be corrected via the crosstalk between the donor and the acceptor fluorophores. To receive the correction factors α and β the fluorophores had to be excited both individual in the FRET channel ($F^{ex_D em_A}$), the acceptor individual in the acceptor channel ($F_A{}^{ex_A em_A}$), and the donor individual in the donor channel ($F_D{}^{ex_D em_D}$). The donor was excited at 520 nm to reduce cross talk. The resulted spectra are given in figure 56 (appendix) for the donor fluorophore and figure 44 for the acceptor fluorophore. Furthermore, the cross talk considering the fluorescence signal $F_A{}^{ex_D em_D}$ and $F_D{}^{ex_A em_A}$ have to be analyzed. However, the FRET pair Cy3 and Cy5 resulted in a fluorescence signal $F_A{}^{ex_D em_D}$ and $F_D{}^{ex_A em_A}$ close to zero, thus this correction can be neglected in this experiment.

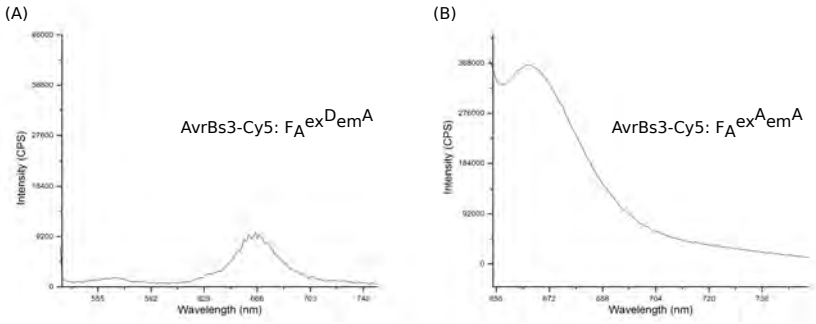

Figure 44: Spectra of AvrBs3-Cy5 to determined the correction factor α
The correction factor α had to be defined to correct the crosstalk between Cy3 and Cy5. Acceptor fluorophore was excited at 520 nm (ex^D) and emission was detected in the acceptor channel as well as the acceptor was excited at 650 nm (ex^A) and emission was detected in the acceptor channel. The factor α is characterized by the fluorescence signal $F_A{}^{ex_D em_A}$(A) divided with the fluorescence signal $F_A{}^{ex_A em_A}$(B). The $F_A{}^{ex_D em_A}$(A) indicates a partly direct excitation of the acceptor at 520 nm.

Both DNAs enclose 25 bp, but DNA (I) carries the Cy3 fluorophore at the 5'-end on the top strand (Figure 45 (A)) and DNA (II) is labeled at the 5'-end on the lower strand (Figure 46 (A)). The theoretical distances between Cy5 and Cy3 were measured on the crystal structure of AvrBs3 (pdp. 2YPF) [205]. The crystal structure contains a 23 bp long DNA, however in the experiment a 25 bp DNA was used. We added per missing bp 3.4 Å to the distance between Cy3 and Cy5. The measured distances r_0* are given in table 22, in figure 45 (A), and in figure 46 (A).

FRET measurements of AvrBs3 (*A*)-Cy5 with DNA (I) revealed a K_D-value of 0.6 ± 0.3 [nM]. The FRET measurements of AvrBs3 (*B*)-Cy5 on DNA (I) also resulted in a comparable K_D-value of 0.6 ± 0.3 [nM] (Figure 45 (B) and Table 22).

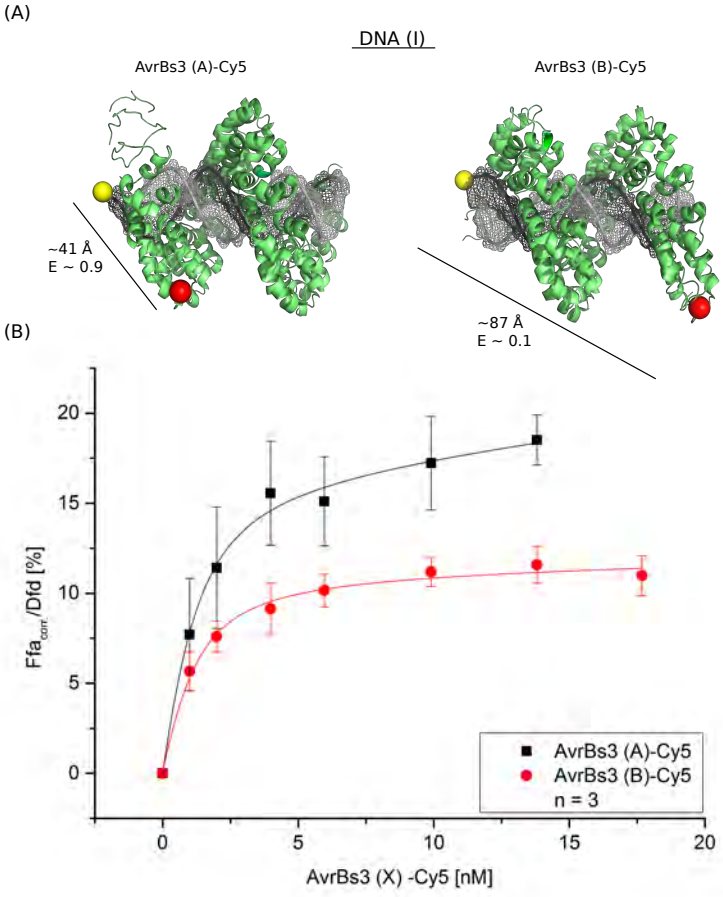

Figure 45: FRET distance measurement of AvrBs3 (*A*)-Cy5 and AvrBs3 (*B*)-Cy5 on DNA (I)
Single cysteine AvrBs3-Cy5 variants were analyzed via FRET distance measurements on 5′-end Cy3 labeled specific DNA (I) (1nM). (A) Crystal structure of AvrBs3 bound to 23 bp specific DNA [205]. 5′-end Cy3 labeled DNA (I) carried the fluorophore Cy3 (yellow) at the top strand 4 bp ahead of the target sequence. Both AvrBs3-Cy5 variants ((A) and (B)) were analyzed on DNA (I) resulting in a close distance (AvrBs3 (*A*)-Cy5: 41 Å) or a far distance (AvrBs3 (*B*)-Cy5: 87 Å) between Cy3 (yellow) and Cy5 (red). Theoretically calculated FRET efficiencies are given with E. (B) FRET index was calculated by formula (6) and then the corrected FRET signal (nF) was divided with the donor signal ($F^{ex_D em_D}$). Each measuring point represents the mean of FRET efficiency ± standard error indicated by error bars for each point. n : number of independent experiments. The data points were fitted with a non-linear regression curve. K_D-values: 0.63 ± 0.3 [nM] (AvrBS3 (*A*)-Cy5); 0.62 ± 0.3 [nM] (AvrBs3 (*B*)-Cy5).

Performing the FRET measurements with DNA (II) resulted in a K_D-value for AvrBs3 (*A*)-Cy5 of 1 ± 0.6 [nM], while for AvrBs3 (*B*)-Cy5 a K_D-value of 6 ± 3 [nM] was determined (Figure 46 (B) and Table 22).

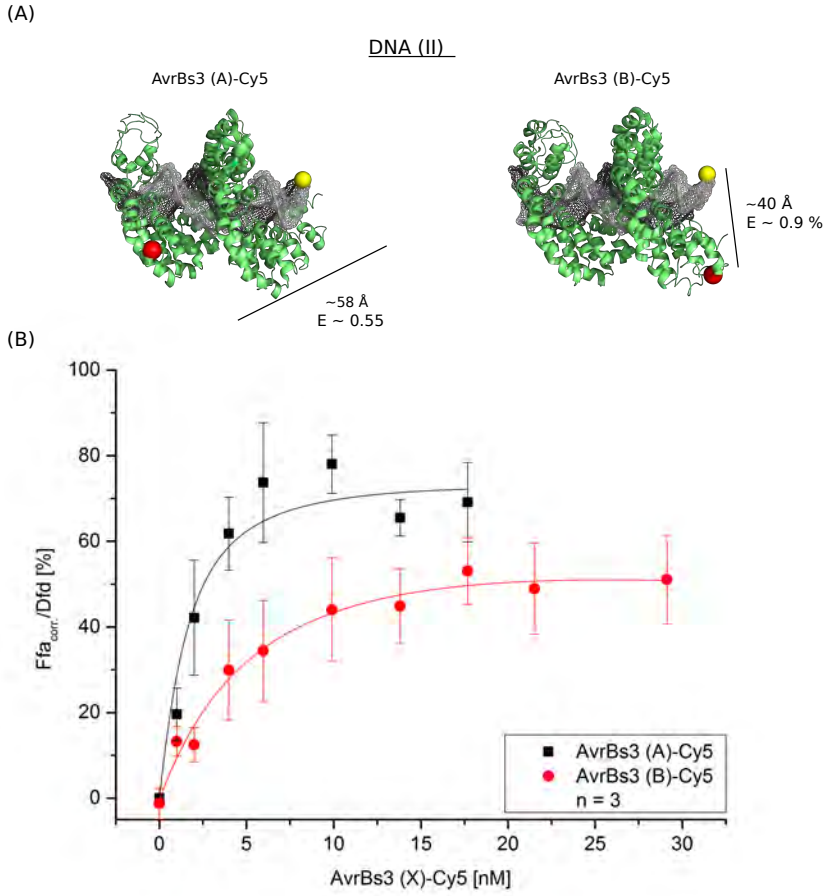

Figure 46: FRET distance measurement of AvrBs3 (*A*)-Cy5 and AvrBs3 (*B*)-Cy5 on DNA (II)
Single cysteine AvrBs3-Cy5 variants were analyzed via FRET distance measurements on 5'-end Cy3 labeled specific DNA (II) (1nM). (A) Crystal structure of AvrBs3 bound to 23 bp specific DNA (pdb: 2YPF) [205]. 5'-end Cy3 labeled DNA (II) carried the fluorophore Cy3 (yellow) at the bottom strand 2 bp behind the target sequence. Both AvrBs3-Cy5 variants ((A) and (B)) were analyzed on DNA (II), resulting in a far distance (AvrBs3 (*A*)-Cy5: ~58 Å) or a close distance (AvrBs3 (*B*)-Cy5: ~40 Å) between Cy3 (yellow) and Cy5 (red). Theoretically calculated FRET efficiencies are given with E. (B) FRET index was calculated by formula (6) and then the corrected FRET signal (nF) was divided with the donor signal ($F^{ex_D em_D}$). Each measuring point represents the mean of FRET efficiency ± standard error indicated by error bars for each point. n : number of independent experiments. The data points were fitted with a non-linear regression curve. K_D-values: 1.1 ± 0.6 [nM] (AvrBS3 (*A*)-Cy5); 6 ± 3 [nM] (AvrBs3 (*B*)-Cy5).

The theoretical FRET efficiencies were calculated using the measured distance r_0* with the mathematic formula (9) and are listed in table 22. While it was expected to result a high FRET efficiency for the "close-state", such as AvrBs3 (*A*) on DNA(I) (r_0*= 41 Å) and AvrBs3 (*B*) on DNA(II) (r_0*= 40 Å), a low FRET efficiency was expected for the "far-state", such as AvrBs3 (*B*) on DNA(I) (r_0*= 87 Å) and AvrBs3 (*A*) on DNA(II) (r_0*= 58 Å).

Table 22: FRET distance measurements on DNA (I) and DNA (II)
Distance measurements between Cy3 and Cy5 were preformed in PyMOL [195] and result in r_0 [Å]*. The experimentally determined K_D-values, the theoretically calculated FRET efficiencies (E) as well as the measured FRET efficiencies (Ef_A*) (mean values ± standard error for three measuring points with the highest protein concentration) are listed. The FRET efficiencies (Ef_A*) were calculated with formula (11). The FRET efficiencies (Ef_A*) were used to define r_0 [Å]** with formula (10).

	r_0* [Å]	E	K_D-value [nM]	Ef_A*	r_0** [Å]	Eq
DNA (I)						
AvrBs3 (*A*)	41	0.9	0.6 ± 0.3	0.42 ± 0.03	63	0.46 ± 0.02
AvrBs3 (*B*)	87	0.1	0.6 ± 0.3	0.52 ± 0.06	59	0.4 ± 0.1
DNA (II)						
AvrBs3 (*A*)	58	0.55	1 ± 0.6	0.69± 0.12	52	0.44 ± 0.06
AvrBs3 (*B*)	40	0.9	6 ± 3	0.74± 0.06	50	0.64 ± 0.02

Hoppe *et al.* [98] described a stoichiometric method based on the three cube-quantification that allows to measure FRET efficiency. We calculated the FRET efficiency (Ef_A*) for the plateaus with this formula (11). The resulted FRET efficiencies (Ef_A*) were used to recalculate the distance between Cy3 and Cy5 (r_0**) via formula (10). Formula (10) is the result of converting formula (9) towards r_0. The distances r_0** are given in table 22. The calculations of FRET efficiencies (Ef_A*) on both DNAs revealed not the expected ratio of FRET efficiencies between "close-state" and "far-state" when compared to the theoretically calculated values (E).

The distance r_0** of AvrBs3 (*A*) on DNA(I) was ~34% greater as the distance r_0* measured within the crystal structure, while AvrBs3 (*A*) on DNA(II) resulted in ~10% reduced r_0** compared to the measured r_0*. Similarly AvrBs3 (*B*) resulted in ~32% reduced distance r_0** on DNA(I), while gaining ~20% increase distance r_0** on DNA(II) when compared to the measured r_0* distance.

Furthermore, we calculated with formula (13) the donor quench for the plateau. It resulted in a similar donor quench for all constructs, except AvrBs3 (*B*)-Cy5 on DNA (II) which revealed an increased quenching effect.

3.2.3.4 *Can AvrBs3 (AB) be labeled without affecting binding characteristics?*

It has been discussed, that TALE binding proteins have a two-state search mechanism. On the one hand, the crystal structure revealed a looser conformation of unbound TALE protein, when compared to the bound state [57, 230]. On the other hand, a two-state mechanism was described recently in a single molecule study, distinguishing between a search state and a recognition state [51]. Additionally, the conformation elasticity of TALE was predicted with computational studies [226, 70]. Based on these observations, we aimed to reveal the conformational changes of TALE proteins with FRET distance measurements. To obtain a AvrBs3 variant which enable FRET distance measurements within the TALE protein, an active AvrBs3 variant carrying two distinct fluorophores had to be designed.

3.2.3.5 *Fluorophores for labeling AvrBs3 (AB)*

An AvrBs3 (**AB**) variant enclosing two cysteine residues, was labeled with single fluorophores to analyze the compatibility of the fluorophores and the binding character of TALE. AvrBs3 (**AB**) was labeled with the fluorophores Alexa 594, Alexa 488, or Alexa 647. Labeling reactions were performed for 2 h for each variant (Figure 47).

Labeling efficiency of AvrBs3 (**AB**)-Alexa 594 (Figure 47 (A), 48 ± 2%) and AvrBs3 (**AB**)-Alexa 488 (Figure 47 (B), 82 ± 2%) reached their plateau at approximately 60 min. AvrBs3 (**AB**) labeling reaction with Alexa 647 Figure 47 (C)) did not reach its plateau phase during the 2 h time period and achieved a labeling efficiency of 69%.

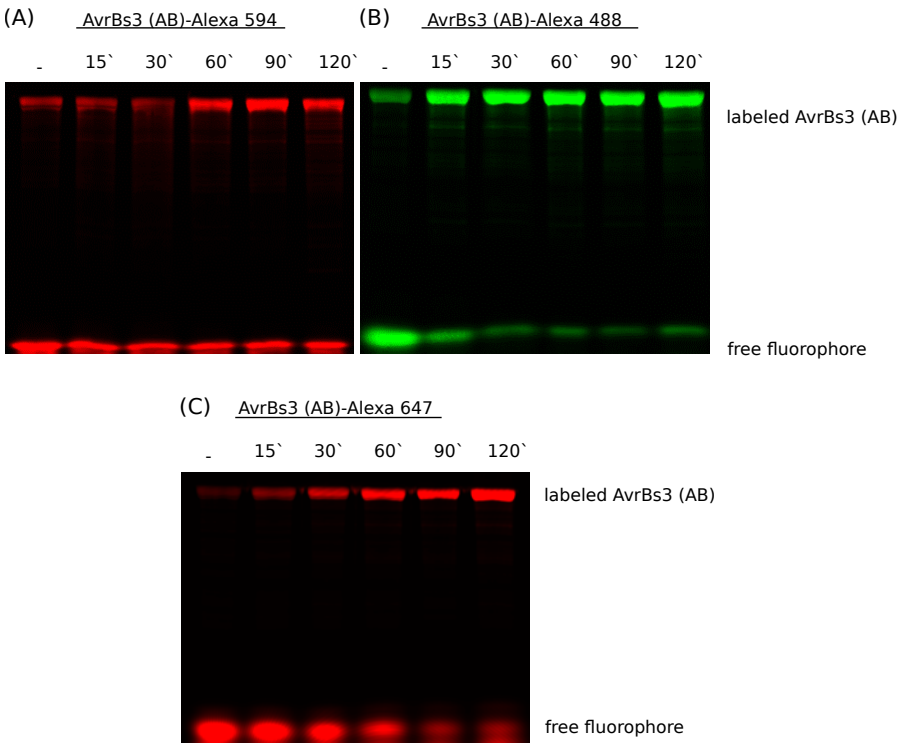

Figure 47: SDS-PAGE of AvrBs3 (**AB**) labeled with Alexa 594, Alexa 488, or Alexa 647

AvrBs3 (**AB**) was labeled with three distinct fluorophores: (A) Alexa 594, (B) Alexa 488, and (C) Alexa 647. The ratio of protein to dye corresponds to a 1:4 ratio. (-) represents starting point of the labeling reaction. Respective time points were take during the labeling reaction (-, 15', 30', 60', 90', 120'). Labeling reactions were analyzed with 12% SDS-PAGE and fluorescent images were visualized with Typhoon^{TM} FLA 9500. Each lane contained 2 µg protein.

Following the labeling reactions proteins were purified with a Zeba Desalting column (40 K) and DOLs were analyzed (AvrBs3 (**AB**)-Alexa 594: 91%; AvrBs3 (**AB**)-Alexa 488: 77%; AvrBs3 (**AB**)-Alexa 647: 144%). The binding characteristic of the single labeled AvrBs3 variants were detected with EMSA on a 5'-end Cy3 labeled specific DNA and are illustrated in figure 48 and figure 49. To ensure that the protein was located on the DNA with increasing concentrations, the fluorophore labeled to AvrBs3 (**AB**) was also exited on the corresponding 4% PAA gel. The two resulting pictures were merged and are displayed in figure 48. The 5'-end Cy3 labeled specific DNA is colored in green and the corresponding fluorophore is displayed in red. The analysis of DNA and protein results in a yellow spot. The merged pictures proofed that the labeled proteins did bind to the specific DNA (Figure 48 (B), (C) and Figure 49), except AvrBs3 (**AB**) labeled with Alexa 594 (Figure 48 (A) and Figure 49). The relative fluorescence intensities of the shifted DNA was plotted in figure 49 (B).

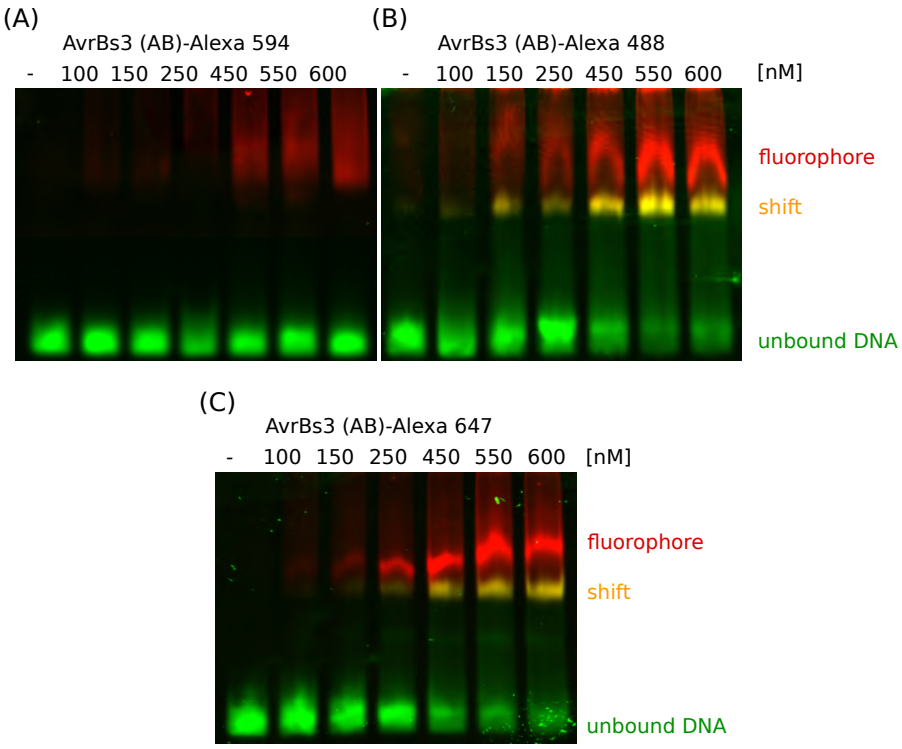

Figure 48: Merge fluorescent images of AvrBs3 (**AB**) labeled with single fluorophores (Alexa 594, Alexa 488, or Alexa 647)
Merged fluorescence images of the 4% PAA gels were visualized with Typhoon^{TM} FLA 9500. 5´-end Cy3 labeled DNA is excited at 532 nm and colored in green. Each fluorophore was exited with a specific wavelength and is shown in red. Region in which protein overlaps with DNA is displayed in yellow. (A) Merged fluorescence images of AvrBs3 (**AB**)-Alexa 594 using the green SHG laser (532 nm). Image was processed afterwards. No DNA shift is detected. (B) Merged fluorescence images of AvrBs3 (**AB**)-Alexa 488 were excited using the blue LD laser (473 nm). DNA is shifted, due to the binding characteristic of AvrBs (AB)-Alexa 488. (C) AvrBs3 (**AB**)-Alexa 647 was excited with the red LD laser (635 nm). Protein binding resulted in a DNA shift.

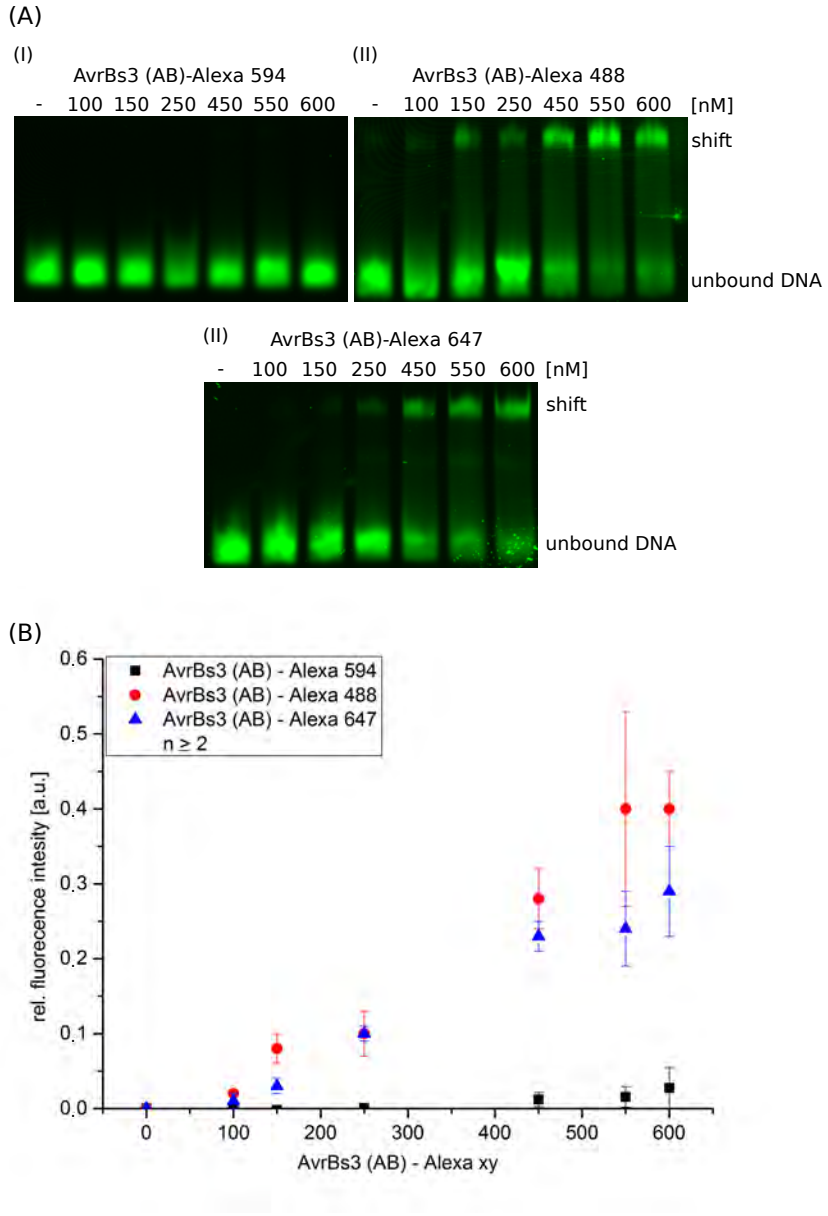

Figure 49: EMSA with AvrBs3 (**AB**) labeled with distinct fluorophores (Alexa 594, Alexa 488, and Alexa 647)
The protein binding was analyzed on a 25 bp long 5'-end Cy3 labeled specific DNA (20 nM, green) via EMSA. TALE protein was added in increasing concentrations to the 5'-end Cy3 labeled specific DNA (-, 100, 150, 250, 450, 550, 600 [nM]). (A) The 4% PAA gels of the EMSA from AvrBs3 (**AB**)-Alexa 594 (I), AvrBs3 (**AB**)-Alexa 488 (II), and AvrBs3 (**AB**)-Alexa 647 (II). Fluorescence images were visualized with Typhoon™ FLA 9500 using the green SHG laser (532 nm). (B) Relative fluorescence intensity [a.u.] of the shifted DNA bands were plotted from each labeled AvrBs3 (**AB**) variant. AvrBs3 (**AB**)-Alexa 488 (red) and AvrBs3 (**AB**)-Alexa 647 (blue) bound to the specific DNA, but AvrBs3 (**AB**)-Alexa 594 (black) lost its binding characteristic.

3.2.3.6 *Consecutive labeling of AvrBs3 (AB) with Alexa 488 and Alexa 647*

Based on these results, AvrBs3 (**AB**) was labeled consecutively with Alexa 488 and Alexa 647 (Figure 50).

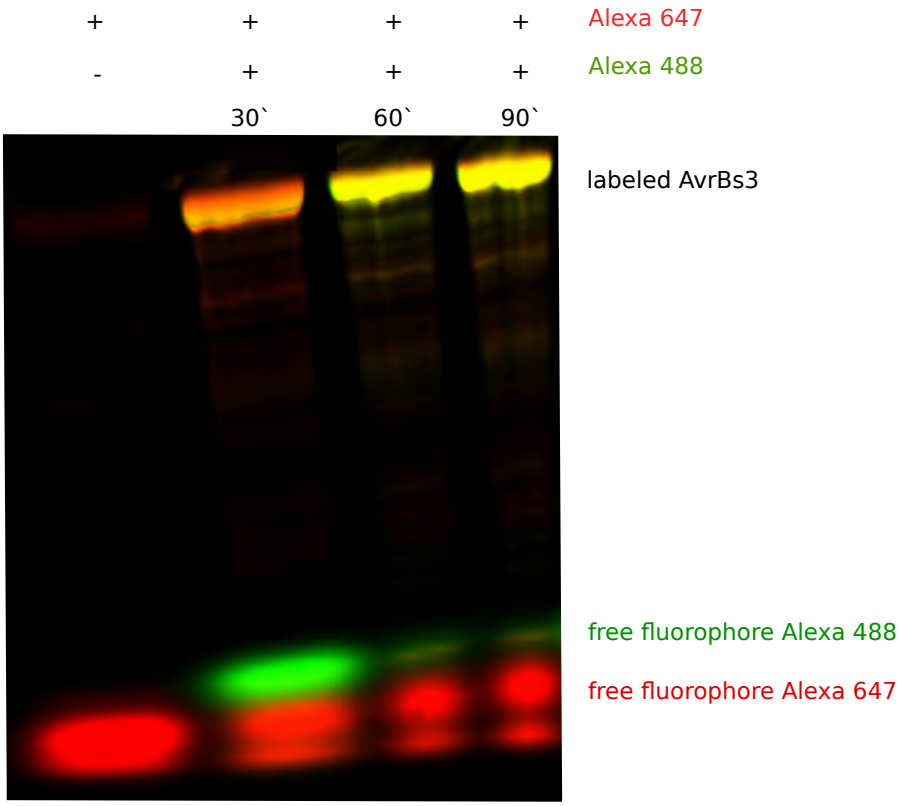

Figure 50: Double labeling of AvrBs3 (**AB**) with Alexa 647 and Alexa 488
Labeling reaction of AvrBs3 (**AB**) with Alexa 647 and Alexa 488 analyzed with a 12% SDS-PAGE. In each lane 2 µg protein was loaded. Merged fluorescence image was visualized with Typhoon™ FLA 9500, using the red LD laser (635 nm) and the blue LD laser (473 nm). Fluorescence signals: Alexa 647 (red) and Alexa 488 (green). Overlap of both fluorophores resulted in yellow color.

The labeling reaction was started with Alexa 647 for 30 min, then Alexa 488 was added to the reaction. The fluorophores were given to the reaction in a 1:2 ratio to the protein. After purification of the protein with Zeba Desalting columns (40K), the protein had a degree of labeling of 123% (Alexa 488) and 57% (Alexa 647). Next, the emission spectra of AvrBs3 (**AB**)-Alexa 488-Alexa 647 was detected, which indicated a FRET signal (Figure 51), thus a location proximity was proofed. The binding characteristic of AvrBs3 (**AB**)-Alexa 488-Alexa 647 was analyzed with EMSA, as described above (Figure 52). No DNA shift was detected.

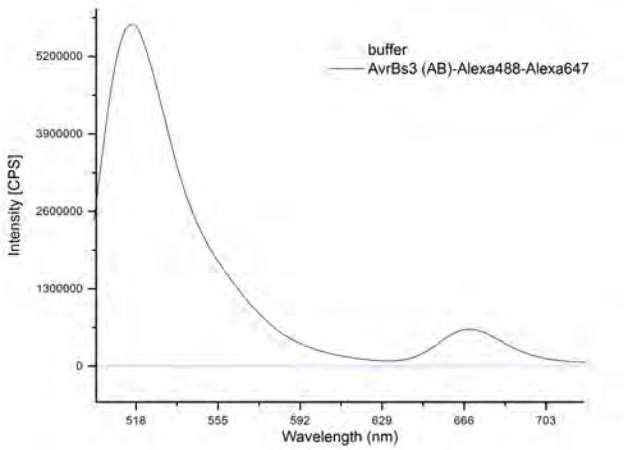

Figure 51: Emission spectra of AvrBs3 (**AB**)-Alexa488-Alexa647
Alexa 488 was excited at 488 ± 5 nm using the Fluoromax-4 Spectroflurometer. Emission spectra was detected between 500-720 nm. The donor fluorophore Alexa488 resulted in a maximum emission spectra at ~ 518 nm and the acceptor fluorophore Alexa 647 showed a peak at ~ 670 nm.

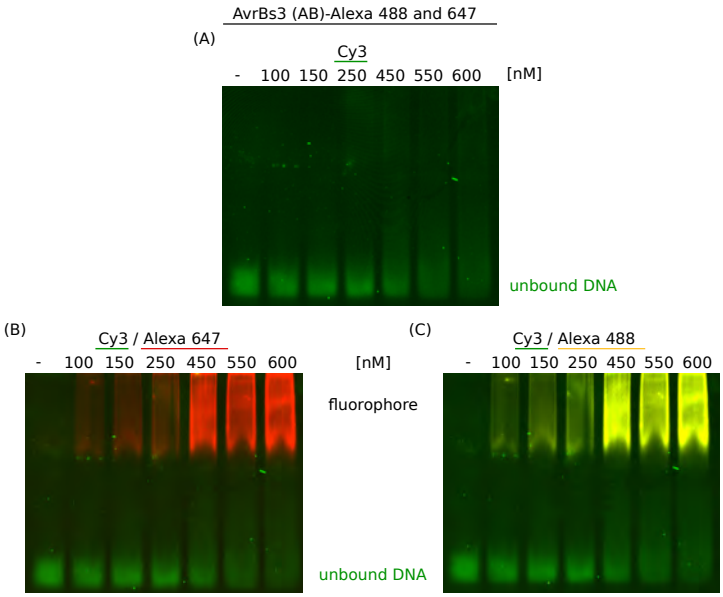

Figure 52: Binding characteristics of AvrBs3 (**AB**) - Alexa 484-Alexa 647
Binding was analyzed on a 4% PAA gel via EMSA. Protein was added with increasing concentrations (-, 100, 150, 250, 450, 550, 600 [nM]) to the 20 bp long 5´-end Cy3 labeled specific DNA (20 nM). The fluorescence images were visualized with Typhoon^{TM} FLA 9500 using the green SHG laser (532 nm, Cy3), blue LD laser (473 nm, Alexa 488), and red LD laser (635 nm, Alexa 647). (A) 5'-end labeled Cy3 specific DNA. (B); (C): merged fluorescence images of excited DNA and each excited fluorophore. No DNA shift was detectable.

4 DISCUSSION

4.1 CHARACTERIZATION OF HIGHLY SPECIFIC NUCLEASES *in vitro* AND *in vivo*

Gene editing has been nominated to be the method of the year in 2011 by the journal Nature Methods [1]. It is a promising tool for curing monogenetic human diseases [50]. This can be realized using specific nucleases to disrupt a dysfunctional gene with NHEJ [131] or through gene replacement using an exogenous DNA template by the precise HDR [30, 95, 159, 60, 187]. In many monogenetic diseases gene replacement is required for efficient treatment [50]. However, HDR is less active than NHEJ and occurs only in the S/G2 phase of the cell cycle [43, 50]. In contrast, NHEJ occurs during the whole cell cycle [43, 50], but causes INDELs and chromosomal rearrangements [112, 16]. Furthermore, off-target effects are a major limitation for the successful application of gene editing. Off-target activity can induce unwanted chromosomal rearrangements, resulting in deletions, inversions, and translocations, thus causes cellular toxicity and long term consequences [104, 112]. Furthermore, there are concerns that off-target activity could activate oncogenes or inactivate tumor suppressors. Several types of engineered nucleases have been developed to optimize the efficiency and specificity of the cutting tools, such as monomeric TALE-FokI [210], TALE-PvuII [238, 239] or Cas9-FokI [89]. While even the most specific nucleases which induce DSBs have been reported to cause mutations at off-target sites [228, 93, 77], the newly designed Cas9-FokI seems to induce solely specific DSBs [89].

Specific nucleases inducing SSBs have been continuously optimized and represent promising tools for gene therapy [116, 157, 140, 78]. While the induction of SSBs increases HDR efficiency and reduces off-target effects, the overall efficiency of gene editing is reduced [145, 52, 222]. It has been recently reported that suppressing the key players of NHEJ increased efficiency of HDR by 4-5 fold [42]. Furthermore, it has been demonstrated that a nick in the transcribed strand triggers HDR eightfold more efficiently compared to a nick in the non-transcribed strand [53]. These findings encouraged us to analyze the site- and strand-specific nickase TALE-MutH under *in vitro* and *in vivo* conditions. Furthermore, studies with paired Cas9 nickases reported low off-target effects and genome editing efficiencies comparable to DSBs Schiml et al. [191], Shen et al. [199], Ran et al. [180], Cho et al. [40]. Thus, we aimed to adapt this double nicking approach to TALE-MutH and to optimize its specificity and efficacy.

TALE-MutH is a specific nuclease, which has been designed and analyzed under *in vitro* conditions by Gabsalilow *et al.* [78]. These studies revealed that TALE-MutH cuts either the lower strand or the upper strand depending on the target design. To optimize the therapeutic application of TALE-MutH we tested various target designs. Our data indicate towards the high therapeutic potential of TALE-MutH to enhance gene editing even when compared to the established specific nucleases I-SceI and TALE-FokI. To proof the therapeutical applicability of TALE-MutH, we exchanged the naturally occurring TALE binding protein AvrBs4 with engineered TALE binding domains, addressing the relevant gene target for X-linked retinitis pigmentosa which resulted in a robust activity.

4.1.0.7 *AvrBs4-MutH*

SINGLE TARGETS: The characteristics of TALE-MutH on the inverted single targets (T-3-H and H-3-T) were analyzed by employing an *in vitro* plasmid cleavage assay. TALE-MutH induced almost two times faster nicks to H-3-T, when compared to the single target T-3-H. However, analyzing the two single targets with the *in vivo* SSA assay led to no differences in the β-galactosidase (β-gal) activity. β-gal activity occurs due to the induced DNA breaks, which trigger SSA repair mechanisms, restoring the gene encoding the missing β-gal domain. Thus, the SSA assay did not confirm the finding by Davis *et al.* [53], that a nick in the transcribed strand triggers HDR more effectively, than a nick in the non-transcribed strand. However, the β-gal activity of the two single targets (T-3-H: 84 ± 9% ;H-3-T : 84 ± 2%) were close to the I-SceI β-gal activity (I-SceI : 100 ± 1%). Following the analysis of the relative number of blue yeast colonies (Figure 22 (B)), the single targets T-3-H and H-3-T revealed a better activity than the β-gal activity of I-SceI. Additionally, we analyzed TALE-MutH on the single target T-6-H, which had been described by Gabsalilow *et al.* [78] to cut the top strand 3 times slower then the bottom strand on T-3-H under *in vitro* conditions. We were able to confirm a reduced activity of TALE-MutH on T-6-H by employing the *in vivo* SSA assay.

Interestingly, SSAs were triggered by inducing SSBs. While it is well recognized that HDR can be induced by SSBs with a low efficiency [145, 52, 222], the finding that SSAs can also be stimulated by SSBs has just recently been reported by Katz *et al.* [113].

These results indicated that TALE-MutH on single targets seem to be a promising tool for the *in vivo* application. Furthermore, when comparing the survival rate of TALE-MutH to the negative control (AvrBs4 without nuclease activity), no significant differences could be detected.

Analyzing the survival rate with the specific single targets only T-6-H revealed a significant reduction of the survival rate, when compared to the negative control. The survival rate of T-6-H proofed to be

comparable to that of the GOLDEN STANDARD I-SceI and the inverted single targets revealed even an increased survival rate when compared to I-SceI. The latter finding is of high interest considering that TALE-MutH revealed a surprising high β-gal activity on the inverted single target.

DOUBLE TARGETS: TALE-MutH was analyzed on two double targets to investigate the double nicking approach of TALE-MutH. The first double target enclosed two adjacent GATC sites (T-3-HH-3-T). While adjacent GATC sites do not occur naturally, the experiment demonstrated the enzyme activity on a target without a distance between the two GATC sites. The second double target contained one GATC site flanked by the recognition sites of AvrBs4 (T-3-H-3-T). A pair of TALE-MutH introduced a site-specific DSB, thus creating a blunt end on T-3-HH-3-T and a 4 bp 5'overhang on T-3-H-3-T.

By using paired Cas9 nickases, it has been shown that the created 5' overhang improves NHEJ and HDR efficiency to a comparable or even higher level when compared to induced blunt ends, while generating a 3' overhang resulted in no improvement of the NHEJ and HDR efficiency [180, 140, 40]. Ran *et al.* [180] reported that no HDR was detectable with paired Cas9 nickases creating a 3' overhang. Additionally, double nicking reduced the off-target effect by 50–1000 fold [180, 140, 40]. This encouraged us to analyze the characteristics of double nicking with TALE-MutH. The double targets were analyzed *in vitro* under denaturing conditions on a 5' labeled PCR fragment. Here, TALE-MutH nicked T-3-H-3-T three to four times faster than T-3-HH-3-T. This finding indicated towards steric hindrance on the double target with two adjusted sites. Additionally, the *in vitro* kinetics revealed that TALE-MutH induces SSBs in a 1:1 ratio to top- and bottom-strand on both target sites. The question arise if TALE-MutH induced a DSB or only nicked the PCR fragments. Analyzing the 5' labeled PCR fragments on native gels revealed that under *in vitro* conditions, no DSBs were induced unless TALE-MutH were digested with PK and new TALE-MutH were added. In this case, DSBs were detected particularly on the target T-3-H-3-T. When one TALE-MutH binds to its specific site, the target site is blocked under *in vitro* conditions.

The *in vivo* SSA assay revealed that β-gal activity on T-3-H-3-T (T-3-H-3-T: 99 ± 1%) is similar to the β-gal activity of I-SceI and the double target T-3-HH-3-T (T-3-HH-3-T: 95 ± 2%) is close to the β-gal activity of I-SceI. Both double targets showed a slightly higher β-gal activity, when compared to the inverted single targets. However, the relative number of blue colonies revealed a low activity on T-3-HH-3-T, whereas comparable levels to the inverted single target sites were found for T-3-H-3-T. Furthermore, the survival rates indicated that T-3-H-3-T is the preferred double target. Both targets showed a

significant reduction in the relative number of yeast colonies, when compared to AvrBs4 without nuclease activity, indicating a toxic effect on cell growth. Whereas T-3-HH-3-T had a lower survival rate compared to I-SceI, the survival rate of T-3-H-3-T was between I-SceI and the inverted single targets. Taken together, T-3-H-3-T proofed to be an excellent target design combining high efficiency and low toxicity. However, it remains unclear if an optimized off-set between the two GATC sites in T-3-HH-3-T could improve SSA efficiency, as it has been reported for the paired Cas9 nickases [180]. The TALE-MutH design is limited by the availability of GATC sites in the genome and GATC sites close to each other are rare. Finally, the double target T-3-H-3-T revealed to be a suitable target design to induce a strand- and site-specific DSBs with a efficiency comparable to I-SceI.

4.1.1 *Beyond AvrBs4-MutH*

4.1.1.1 *AvrBs4-FokI*

AvrBs4-FokI was the first TALEN construct and is based on the design of ZF-FokI [117]. The cleavage module consists of the non-specific cleavage domain of FokI. The cleavage domains have to dimerize to induce DSBs [167, 188, 23] and due to its non-specific nature high off-target effects have been reported [93]. We characterized TALE-FokI in our yeast assays, analyzing survival rates and β-gal activity. Comparing the survival rate without inducing the target plasmid, revealed no significant differences when compared to the negative control. However, analyzing the survival rates with target plasmids led to the highest reduction of the survival rate of all analyzed constructs, indicating towards toxic effects on cell growth.

The β-gal activity of AvrBs4-FokI (AvrBs4-FokI: 83 ± 17%) revealed a robust β-gal activity. Yet, the relative number of blue colonies showed the toxic effects of AvrBs4-FokI. Both, I-SceI and TALE-MutH on all tested targets displayed a higher specificity and efficiency, when compared to TALE-FokI. In table 23 the *in vivo* characteristics of I-SceI, TALE-FokI and TALE-MutH are summarized.

4.1.1.2 *AvrBs3-PvuII*

An alternative architecture of the specific TALENs is the combination of the site-specific cleavage domain PvuII with the binding module TALE [238, 239]. Our data showed that the TALE-MutH variant has superior characteristics for gene editing, when compared to the TALE-PvuII variants. PvuII was initially used as an alternative cleavage domain to FokI in combination with the binding module TOF [65]. Similar to FokI, PvuII has to dimerize to induce DSBs. The main advantage of PvuII is its site-specific character, which is based on the recognition of a palindromic 6 bp sequence [37, 170, 169]. Due

Table 23: Overview of the I-SceI, TALE-FokI and TALE-MutH results
Significant difference between analyzed AvrBs3-PvuII variants and AvrBs4 without nuclease activity were determined by the unpaired T-test: n $\geq$ 3, n.s. = not significant, P < 0.05 *, P < 0.01 **, P < 0.001 ***

	in vivo			
	Survival rate		β-gal activity	
	with target site	without target site	blue colonies [%]	rel. number of yeast colonies
I-SceI	*	n.s.	***	***
AvrBs3-FokI	***	n.s.	***	*
TALE-MutH	-	n.s.	-	-
TALE-MutH				
T-3-H	n.s.	-	***	**
H-3-T	n.s.	-	***	**
T-6-H	**	-	**	**
T-3-HH-3-T	*	-	***	***
T-3-H-3-T	*	-	***	***

to the limitation of TOF-sc-PvuII for *in vivo* applications the specific nuclease were optimized using binding modules such as ZFs or TALEs [239, 190]. Previous, investigations reported that TALE-PvuII had superior features, when compared to ZN-PvuII [190]. Today, three TALE-PvuII variants are available (TALE-hf-PvuII, TALE-sc-PvuII and TALE-hd-PvuII). The variants were designed and optimized by Yanik *et al.* [239, 238]. Using the two distinct yeast viability assays, we were able to confirm the high toxicity of TALE-hf-PvuII and TALE-sc-PvuII, which has been previously reported by Yanik *et al.* [239, 238] in *E.Coli* strains. To optimize the TALE-PvuII design, a heterodimeric variant was designed by introducing a mutation in each subdomain of PvuII [238]. Under *in vitro* conditions, the variants AvrBs3-K38D-PvuII and AvrBs3-D30K-PvuII have been shown to be inactive on their own, but when introduced together they induce specific DSBs. Similar results were achieved when the variants were analyzed in HEK293 cells, however with a reduced activity [238]. We could not reproduce this finding in yeast cells. Here, isolated AvrBs3-D30K-PvuII showed no activity, while AvrBs3-K38D-PvuII (β-gal activity: 95 ± 3%) revealed a similar β-gal activity as I-SceI. The relative number of blue colonies of AvrBs3-K38D-PvuII was significantly higher, when compared to the negative control. However, it revealed a reduced activity, when compared to I-SceI. Analyzing the

heterodimeric AvrBs3-hd-PvuII combination revealed a reduced activity, when compared to AvrBs3-K38D-PvuII. Thus, it can be suggested that AvrBs3-K38D-PvuII is active in yeast.

The viability assay without a specific target revealed no significant difference in the relative amount of cell survival between the negative control and the analyzed TALE-PvuII variants. However, the viability assay with specific targets showed a reduced cellular growth for AvrBs3-hd-PvuII and AvrBs3-K38D-PvuII.

Taken together, the different AvrBs3-PvuII variants proofed to possess major limitations in the SSA assay and the viability assays. We were able to confirm the high toxicity of TALE-hf-PvuII and TALE-sc-PvuII with our SSA assay, but the reported function of the heterodimeric AvrBs3-hd-PvuII could not be reproduced. Contrasting to the reported characterization of AvrBs3-K38D-PvuII [238], isolated AvrBs3-K38D-PvuII revealed to be able to induce DSBs in yeast. An overview of the characteristics of AvrBs3-variants is given in table 24.

Table 24: Overview of the nuclease activity of AvrBs3-PvuII variants
Significant difference between analyzed AvrBs3-PvuII variants and AvrBs4 without nuclease activity were determined by the unpaired T-test: n $\geq$ 3, n.s. = not significant, P < 0.05 *, P < 0.01 **, P < 0.001 ***

	in vivo			
	Survival rate		β-gal activity	
	with target site	without target site	blue colonies [%]	rel. number of yeast colonies
I-SceI	n.s.	n.s.	***	**
AvrBs3-hf-PvuII	*	n.s.	***	**
AvrBs3-K38D-PvuII	*	n.s.	***	**
AvrBs3-D30K-PvuII	n.s.	n.s.	n.s.	n.s.
AvrBs3-sc-PvuII	*	***	n.s.	n.s.
AvrBs3-hd-PvuII	*	***	n.s.	n.s.

4.1.1.3 *Cas9 endonuclease*

The expression plasmid for Cas9 endonucleases in yeast has been established in 2013 [58] and we aimed to investigate the β-gal activity of Cas9 endonucleases with the SSA system. We were able to confirm

a significant induction of β-gal activity in *S. cerevisiae* by the Cas9 endonuclease, when compared to the negative control. Still, compared to I-SceI the Cas9 system showed low β-gal activity. Three plasmids had to be introduced; the expression plasmid of Cas9, the gRNA plasmid, and the target plasmid. The expression plasmid for Cas9 nuclease and the target plasmid carry the same selective marker (TRP1). Thus, our results could be significantly influenced by transformation efficiency.

4.1.1.4 *Alignment search of the addressed sites in the genome of Saccharomyces cerevisiae*

We screened the yeast genome of the S288c strain for similar off-target sites of the employed specific nucleases. No 100% similar off-target site for the analyzed on-target sites were found. Next, the seed regions of the target site for each specific nuclease was analyzed. AvrBs4 and AvrBs3 revealed the highest off-target similarity with 68% containing the 5' seed region. This finding is in agreement with the low cell growth of AvrBs4-FokI. Yet, no similar off-target was found including the 10 bp of the 5' seed region and the addressed site of MutH (5'-$^{\triangledown}$GATC-3') or PvuII (5'-CAG$^{\triangledown}$CTG-3'). For the on-target of I-SceI, the best two hits had a similarity of 66%, enclosing the 3' seed region of I-SceI. The on-target for Cas9::gRNA revealed off-target site hits with a similarity of 52% containing the PAM site and the 3' seed region. Cas9::gRNA has been shown to possess a lower specificity, compared to the TALE binding domains. Furthermore, the Cas9 system has been reported to tolerate up to 5 mismatches [101, 163, 76]. It is therefore, that the difference of the off-target sites and on-target sites are adequate to limit off-target effect.

THERAPEUTIC TARGET: The next step was to analyze if the naturally occurring binding module AvrBs4 could be exchanged by an engineered TALE binding domain without losing the specificity and efficiency of the designed nuclease TALE-MutH. We targeted the relevant gene in X-linked retinitis pigmentosa, which is caused by mutations in the RPGR-gene. 80% of the mutations leading to RPGR dysfunction are located in the ORF15, which consists primarily of an imperfect repeat motif [100, 223, 198, 165]. Treatment of X-linked retinitis pigmentosa could be realized by insertion of a AAV vector enclosing the encoding sequence for the RPGR as has been demonstrated in dogs [15] and mice [234].

Distinct GATC sites are located in the ORF15 and TALE binding modules were created by employing the GOLDEN GATE cloning strategy. The newly designed TALE-MutH constructs (TALE(RPGR)-1 (I) -MutH/ TALE(RPGR)-2 (I) -MutH) were analyzed in terms of their activity and toxicity in HEK293T cells. They addressed a heterodimeric double target site containing one GATC site flanked with the two

TALE binding sites. The design is based on the double target T-3-H-3-T and the heterodimeric TALE-MutH induces a strand- and site-specific DSB resulting in a 4 bp 5'overhang. The activity was detected with a frame shift assay, which is triggered by NHEJ. The high frequencies of NHEJ and the independence of a DNA template simplify the analysis of nuclease activity, when compared with the detection of HDR efficiency. Thus, we analyzed NHEJ activity to characterize nuclease activity on the therapeutic target.

It revealed similar NHEJ activity for the heterodimeric TALE-MutH design (TALE(RPGR)-1 (I) -MutH/ TALE(RPGR)-2 (I) -MutH) , when compared with the GOLDEN STANDARD I-SceI. Additionally, no significant difference between the survival rates of I-SceI and the heterodimeric TALE-MutH constructs were found. Furthermore, an empty plasmid (pcDNA) was analyzed as a negative control. No significant difference between the survival rates of I-SceI and empty vector were detected, while the ZF-FokI [6] construct revealed a reduced survival rate. The episomal *in vivo* analysis of TALE-MutH on the therapeutic target underlined its high potential as a specific and efficient tool for gene targeting.

Our results indicate that TALE-MutH can challenge the established specific nucleases (I-SceI and TALE-FokI), when addressing the relevant gene causing 80% of X-linked retinitis pigmentosa [100, 223, 198, 165]. The target site T-3-H-3-T, proofed to be a powerful target design to induce double nicks. Further important steps are the analysis of HDR efficiency of TALE-MutH on the T-3-H-3-T targets, charactering the exogenous DNA template, such as its overall length, the length of the homologous parts, or its double stranded or single stranded DNA character. Additionally, the effect of TALE-MutH addressing its target site on a genome have to be investigated. Here, the chosen genome should be screened for similar off-target sites and the accessibility of the target sequence in the chromatin structure remains to be analyzed.

Nevertheless, TALE-MutH requires a GATC site to induce a SSBs and TALE binding modules have to be designed for each new target site respecting several conditions, such as the required thymine at the beginning of the target sequence [193] and the higher sensitivity to mismatches at the 5'end of the addressed site [142].

The GOLDEN GATE cloning strategy enables to design TALE binding modules for any given target within a 5 days protocol [67]. TALE binding modules can be very specific [215, 160, 59], but depending on the addressed target site, the TALE domain can also looses its specific features [88]. Computer programs have been developed to predict the best composition of a TALE binding domain [184, 86, 62]. Yet, an improved understanding of the TALE binding mechanism remains to be investigated.

4.2 CHARACTERIZATION OF BINDING CHARACTERISTICS OF AVRBS3

The second part of this thesis concentrates on the characterization of the naturally occurring TALE protein AvrBs3. Four distinct AvrBs3(△N152-C28) variants (two single cysteine variants, a double cysteine variant, and a variant containing all cysteine residues) were expressed and characterized. The K_D-values of the AvrBs3 variants were analyzed and functional labeled AvrBs3 variants were designed.

TALE proteins have been used as binding module for specific nucleases and were successfully employed in genome and epigenome editing technologies [109, 48, 135]. Deciphering of the recognition code of TALE [25] and the introduction of the one pot cloning strategy GOLDEN GATE represent important steps toward a therapeutic use of the TALE proteins [33]. Theoretically, any given target can be addressed by an engineered TALE protein. Despite these developments, off-target effects of TALENs were described and limit their application [88]. Improving the knowledge on the binding mechanism of TALE has been a central aim of many researchers [142, 205, 193, 194]. For instance, a polarity effect has been recognized, indicating that the 5' -end is more sensitive to mismatches [142]. Furthermore, the impact of adjacent repeat domains were investigated and computer programs were developed to identify the best fitting TALE construct for a specific target [184]. Still, a better understanding of the binding mechanism of TALE proteins is required, which could result in an improved prediction of potential off-target effects.

4.2.0.5 *Determination of K_D-values*

The K_D-values of AvrBs3 with different lengths of its N- and C-terminus were analyzed under various conditions using several methods [205, 142, 194]. Schreiber *et al.* [194] described the K_D-value of AvrBs3(△ N152-C16) with similar N- terminus length to AvrBs3(△ N152-C28) using fluorescence anisotropy on a 36 bp long DNA under physiological conditions (c(NaCl) = 150 mM, pH: 7.5). Their studies revealed a higher K_D-value of the AvrBs3(△ N152-C16) (K_D= 282 ± 18 nM), when compared to the K_D-value of the AvrBs3(△ N152-C28) (K_D= 1.9 ± 0.8 nM) as we have determined on a 25 bp long DNA under low salt conditions (60 mM). These divergent results could be explained by the extended DNA and the high salt concentration employed by Schreiber *et al.* [194]. Additionally, they measured the K_D-values of wtAvrBs3 (K_D-value = 23.2 ± 1.6 nM) and AvrBs3(△ N152-C278) carrying the entire C-terminus (K_D= 99.2 ± 7.5 nM), suggesting that a truncation of the C-terminus to 16 aa has a high impact on the binding characteristics of AvrBs3. Stella *et al.* [205] measured the K_D-value of wtAvrBs3 (K_D= 33.6 ± 4,9 nM) with fluorescence anisotropy on a 20 bp long DNA under physiological conditions (c(NaCl) = 150 mM, pH: 7.5). The K_D-values of wtAvrBs3 in both studies were compa-

rable. However, using the same conditions, but isothermal titration calorimetry for detection resulted in a K_D-value of 88 ± 32 nM [205].

A third group reported K_D-values of 3.2 ± 1.6 nM (AvrBs3(△ N143-C54)) and 3.9 ± 0.1 nM (AvrBs3(△ N0-C192)) using EMSA under low salt conditions (c (KCl)= 60 mM) [142]. These findings were similar to the K_D-value of AvrBs3(△ N152-C28) (K_D= 1.9 ± 0.8 nM) measured in our study. Furthermore, the ratio of active and inactive AvrBs3 influences the K_D-value. Here, protein expression, purification, and storage have an impact on the ratio between active and inactive protein and could be important factors for comparing the analyzed K_D-values. We determined the ratio of active AvrBs3(△ N152-C28) to estimate the K_D-value for 100% active protein. About 13% of AvrBs3(△ N152-C28) was active, resulting in a K_D-value of 0.24 ± 0.09 nM. The high inactive portion of AvrBs3(△ N152-C28) indicates towards a relevant degeneration under *in vitro* conditions. An activity assay with the fusion construct AvrBs3-PvuII revealed a 50% reduction of the cutting activity under physiological conditions in 4 h [238].

4.2.0.6 *Cysteine variants*

The characterization of the crystal structure of the apoenzyme and the bound TALE protein was a mayor step in identifying the binding character of the TALE::DNA interaction. Here, a superhelical conformation of TALE was shown to interact with the major groove of the DNA [139, 205] and a positive charge along the inner surface contributes to the non-specific binding energy of the TALE::DNA interaction [57, 81]. The positive charged amino acids G14, K16 and Q17 interact with the negatively charged phosphate backbone of the DNA, thus supplying the main binding energy. Interestingly, the so called di-residues at position 12 and 13 do not significantly contribute to the binding energy, despite the fact that residue 13 is the only known AA performing direct interactions with the DNA. This suggests, that binding to the specific DNA occurs by negative discrimination due to steric and electrostatic clashes between DNA and the di-residues [230, 57, 139]. Furthermore, the crystal structure and computer based analysis of TALE binding revealed a conformation elasticity of the binding module, resulting in a loose conformation as apoenzyme and a more tight conformation when bound to its target site [57, 230, 226, 70]. It has been reported, that the superhelical turn of TALE is reduced from 60 A° to 35 A° in its bound state [57]. Recently, the AvrBs3 binding mechanism has been described on the single-molecule level [51]. To enable this, AvrBs3 was labeled with Cy3 at the N-terminus via an unnatural aldehyde. The TALE protein was tagged with a six AA motif and the formylglycine generating enzyme (FGE) converted the cysteine enclosed in the motif on protein level. The single-molecule experiment revealed a conformation dynamic between search and recognitions state, as well as the interactions between NTR and non-specific DNA

that facilitate 1D diffusion [51]. The main advantage of the labeling strategy used by Cuculis *et al.* [51], is the lack of mutations which could impact the binding character of AvrBs3. However, AvrBs3 hold a tag at the N-terminus, which could influences the binding characteristics due to steric clashes. Moreover, the fluorophore is fused to the flexible linker and therefore precise distance measurements via FRET are not possible. Another strategy to label proteins is to create single cysteine variants. While the fluorophore is located closer to the protein, the cysteine residue has to be accessible for the labeling reaction and the removed cysteine residues are not allowed to have a mayor impact to the binding characteristics of the protein. AvrBs3 has a cysteine residue at position 30 in each repeat domain [57, 139]. In AvrBs3(ΔN152-C278) Schreiber *et al.* [194] substituted all cysteine residue to serine or alanine residues ($AvrBs3(C30S)_{Rep}$ / $AvrBs3(C30A)_{Rep}$). This study proofed that the cysteine residues are required for gene activation in *planta*. Interestingly, the specific interaction between $AvrBs3(C30S)_{Rep}$::DNA is only minimally reduced under *in vitro* conditions, when compared to AvrBs3(ΔN152-C278) containing all cysteine residues. The cysteine residues are required for dimerization via disulfide bridges to homodimers, which facilitate the nuclear transport [91, 194]. $AvrBs3(C30S)_{Rep}$ revealed reduced complex formation to AvrBs3(ΔN152-C278) due to the missing cysteine residues. Furthermore, Schreiber *et al.* [194] postulated that cysteine substitution resulted in a plasticity change of AvrBs3. The missing cysteine residues led to a small conformation variance in each repeat. This effect might increase with increasing repeat numbers and result in the incapability of $AvrBs3(C30S)_{Rep}$ to activate genes in *planta*.

In the present thesis, the cysteine residue was substituted to methionine, as methionine has the least shift to cysteine in its thermodynamic stability ($\Delta\Delta$ G = 0.3 [kcal/mol]) when compared to all other aa. Serine and alanine revealed a shift in the thermodynamic stability above 1 [kcal/mol]. We designed three distinct cysteine variants. AvrBs3 (A) carrying one cysteine residue in repeat one, AvrBs3 (B) encloses a cysteine residue in repeat 16, and AvrBs3 (AB) contains two cysteine residues. First, we analyzed the K_D-value of AvrBs3 (B) and were able to show that the K_D-value of AvrBs3 (B) was similar to the K_D-value of AvrBs3(△N152-C28) containing all cysteine residues. Consequently, the cysteine residues did not affect the binding characteristics of AvrBs3 under *in vitro* conditions regarding the determined K_D-values. While the substitution of cysteine to methionine revealed similar binding characteristics compared to AvrBs3(△N152-C28), the substitution described by Schreiber *et al.* [194] resulted in a slightly decreased affinity of $AvrBs3(C30S)_{Rep}$ compared to AvrBs3(ΔN152).

Next, we performed competition assays in which AvrBs3 (**B**) bound to its specific target site (AvrBs3 (**B**)::DNA) competed with an ex-

cess of non-specific or specific free DNA. The experiment revealed a non-specifc binding state and a specific binding state. This finding was supported by an individual competition assay and a consecutive competition assay with AvrBs3 (**B**). Our results suggest, that several AvrBs3 (**B**) are clustered to the specific AvrBs3 (**B**)::DNA complex. Adding an excess of unspecific DNA to the specific AvrBs3 (**B**)::DNA complex resulted in a fast displacement of the non-specifc bound proteins, followed by a slow displacement of the specific interaction (AvrBs3 (**B**)::DNA) when an excess of specific DNA was added. Removing all specific interactions was achieved by digesting AvrBs3 (**B**) with Proteinase K.

On the contrary, AvrBs3(△N152-C28) carrying all cysteine residues did result in an even slower decrease of the specific interaction (AvrBs3 (△N152-C28)::DNA) by adding an excess of specific DNA to the specific complex. These findings indicate towards a stronger AvrBs3(△N152-C28)::DNA interaction, when compared to the cysteine substitution variant (AvrBs3 (**B**)::DNA).

4.2.0.7 *Labeling single cysteine AvrBs3 variants*

Next, we labeled the AvrBs3 cysteine variants and analyzed their binding characteristics under *in vitro* conditions. AvrBs3 (**B**) labeled with Alexa 488 resulted in a DOL of 158% and a K_D value of 1.3 ± 0.3 nM. This proofed that the fluorophore Alexa 488 is compatible with the binding characteristics of AvrBs3. The K_D-value is comparable with AvrBs3(ΔN152-C28).

AvrBs3 (**A**) and AvrBs3 (**B**) were also labeled with the Cy5 fluorophore and DOLs of 115% (AvrBs3 (**A**)-Cy5) and 82% (AvrBs3 (**B**)-Cy5) were achieved. The binding characteristics of the variants were analyzed on two distinct 25 bp long 5′ -end Cy3 labeled specific DNAs. DNA(I) carried the Cy3 fluorophore at the 5′ -end on the top strand and DNA(II) held the Cy3 fluorophore at the 5′ -end on the lower strand. Cy5 and Cy3 are a known FRET pair with a FÖRSTER RADIUS OF 60 Å [96]. The FRET measurement depends on the overlapping absorption spectra of donor and acceptor fluorophores, the distance between the chosen fluorophores and the orientation between the donor and acceptor fluorophores to each other.

THE K_D -VALUES of the different variants were analyzed on the two DNA designs with FRET measurements. The measured FRET signals were corrected by the cross-talk between donor and acceptor fluorophores. The FRET indexes (nF) were calculated by the three-cube method described by Youvan *et al.* [242]. This method does not represent the FRET efficiency, but provides an value, which changes with FRET [17, 244], thus it enables determination of K_D -values.

On DNA(I) both Cy5 labeled AvrBs3 variants revealed comparable K_D -values to the unlabeled single cysteine variant. On DNA (II)

AvrBs3 (**A**)-Cy5 also showed a similar K_D -value when compared to the unlabeled single cysteine variant, while AvrBs3 (**B**) revealed an increased K_D -value of 6 ± 3 nM, when compared to the anisotropy results of the unlabeled cysteine variants.

THE BINDING CHARACTER of AvrBs3 was studied by using the specifically labeled AvrBs3 variants on DNA(I) and DNA(II). AvrBs3 (**A**)-Cy5 on DNA(I) (r_0*= 41 Å) and AvrBs3 (**B**) on DNA II) (r_0*= 40 Å) was expected to have a "close-state" to the Cy3 label on the DNAs leading to an increased FRET efficiency, when compared to AvrBs3 (**B**)-Cy5 on DNA(I) (r_0*= 58 Å) and AvrBs3 (**A**)-Cy5 on DNA(II) (r_0*= 87 Å), representing the "far-state".

We measured the theoretical distance between donor and acceptor fluorophores within the crystal structure of AvrBs3 on a 23 bp long DNA (pdb: 2YPF) [205]. The DNAs used in this thesis were 25 bp long and therefore 3.4 Å were added for each missing nucleotide. The theoretical FRET efficiencies were calculated with the measured distance between donor and acceptor fluorophores using formula (9). These data revealed that the difference between "close-state" and "far-state" is more pronounced on DNA(I) (E(AvrBs3 (**A**)= 0.9; E(AvrBs3 (**B**))= 0.1), compared to DNA(II) (E(AvrBs3 (**A**)= 0.55; E(AvrBs3 (**B**))= 0.9). Additionally, the difference between the theoretically calculated FRET efficiencies of AvrBs3 (**B**)-Cy5 on the two DNAs is larger as the observed FRET efficiencies of AvrBs3 (**A**)-Cy5 on the two different DNAs.

Furthermore, we calculated the FRET efficiency (Ef_A*) for the three measuring points with the highest protein concentration with a stoichiometric method that is based on the three cube-quantification described by Hoppe *et al.* [98]. The resulted FRET efficiencies (Ef_A*) were used to recalculate the distance between Cy3 and Cy5 (r_0**) as decribed by formula (11).

The measured FRET efficiencies (Ef_A*) on both DNAs revealed not the expected ratio of FRET efficiencies between "close-state" and "far-state" when compared to the theoretically calculated values (E). The distance r_0** of AvrBs3 (**A**) on DNA(I) was ~34% greater as measured within the crystal structure, while AvrBs3 (**A**) on DNA(II) resulted in ~10% reduced r_0** compared to the measured r_0*. Similarly AvrBs3 (**B**) resulted in ~32% reduced distance r_0** on DNA(I), while gaining ~20% increase distance r_0** on DNA(II) when compared to the measured r_0* distance (Figure 53). These finding indicates towards diverged orientation of the Cy3 and Cy5 fluorophore as expected. The fluorophores were covalent bound to the cysteine residue allowing a certain elasticity of the fluorophores. Due to the plastic elasticity of AvrBs3 and the fluorophore Cy5 the location of Cy5 fluorophore was hardly described. A further step would be the calculation of an expected flexibility of Cy5 according to its structural size.

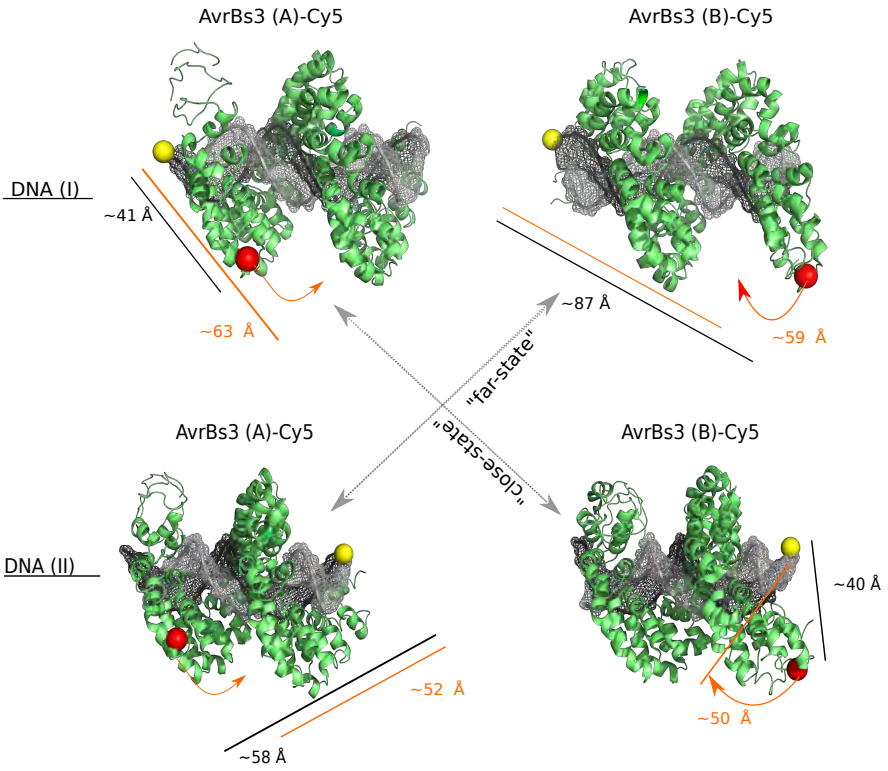

Figure 53: "close-state" and "far-state"
The distances on DNA(I) and DNA(II) of AvrBs3 (**A**)-Cy5 and AvrBs3 (**B**)-Cy5 are illustrated on the crystal structure AvrBs3 (pdb: 2YPF) [205]. The cysteine residues used for the labeling reaction with Cy5 is colored in red, the position of Cy3 is displayed in yellow. The distances r_0* were measured within the crystal structure AvrBs3 (pdb: 2YPF) [205] and are displayed in black, while r_0** yield by the measured FRET efficiencies (Ef_A*) are colored in orange. The arrows indicate the displacement of Cy5.

The resulted distances r_0** indicated that the fluorophore Cy5 labeled to the N-terminus of AvrBs3 (**A**) is shifted to the C-terminus, as well as the Cy5 fluorophore labeled to the C-terminus (AvrBs3 (B)) is shifted to the N-terminus. These would explain the greater distances for the "close-state" and the reduced distances for the "far-states" (Figure 53).

The Ef_A* were calculated by the formula (11) described by Hoppe *et al.* [98]. However, Hoppe *et al.* [98] also described an extended formula which corrects the FRET signal for the donor quench, thus the FRET efficiencies could be defined more accurate. Additionally, FRET signal could also be influenced by different factors which complicate the comparison between the single experiments. For instance, the different DOLs of the labeled DNA and AvrBs3-Cy5, further quenching effects, and preferred orientation of Cy3 and Cy5 within one of the DNA::AvrBs3 complexes when compared with another DNA::AvrBs3 complex.

Quenching in term of FRET has also become a tool to investigate ligand binding, thus we calculated the donor quench for the plateau. It resulted in a similar donor quench for all constructs, except AvrBs3 (**B**)-Cy5 on DNA (II) which revealed an increased quenching effect. This finding indicates towards similar binding characteristics of the AvrBs3::DNA complexes.

4.2.0.8 *Labeling of the double cysteine variant AvrBs3 (AB):*

It has been described, that TALE has a conformational elasticity. Binding to DNA results in a tighter superhelical conformation [57, 230, 226, 70], but conformational changes between non-specific and specific binding has also been described with single-molecule experiments [51]. We aimed to show this effect with FRET distance measurements. We designed a double cysteine AvrBs3 (**AB**) variant, carrying a cysteine residue in repeat 1 and in repeat 16. We expected, that upon binding to the DNA the distance between donor and acceptor fluorophore would decrease, resulting in an increased FRET signal.

The first challenge was to design a functional double labeled AvrBs3 variant. The binding characteristics of unlabeled AvrBs3 (**AB**) was analyzed by EMSA, which revealed two shifts of the labeled DNA. This suggested that the first shift contributes to the specific binding and the second shift is a result of non-specific binding clustered to the specific AvrBs3 (**AB**)::DNA complex. The next step was to find donor and acceptor fluorophores compatible with the binding characteristics of AvrBs3 (**AB**). AvrBs3 (**AB**) was labeled either with Alexa 488, Alexa 594, or Alexa 647; EMSA was performed to analyze the binding capacity. AvrBs3 (**AB**) labeled with Alexa 488 and Alexa 647 kept its binding characteristics, but AvrBs3 (**AB**)-Alexa 647 showed decreased binding to the specific DNA. AvrBs3 (**AB**)-Alexa 594 completely lost its binding characteristics. Alexa 488 and Alexa 647 have overlapping absorption spectra and function as a FRET pair. Therefore, AvrBs3 (A**B**) was labeled consecutively with Alexa 488 and Alex 647. Based on the isolated labeling kinetics with Alexa 488 and Alexa 647, we knew that Alexa 647 reacts slowly with the cysteine residue of AvrBs3 (**AB**). Therefore, AvrBs3 (**AB**) was incubated for 30 min with Alexa 647 and then Alexa 488 was added. Despite this, a DOL for Alexa 488 of 123% and for Alexa 647 of 57% were measured. Thus, Alexa 647 showed a reduced labeling capacity for AvrBs3 (**AB**). Alexa 488 (M = 721 [g/mol]) has a small molecular weight, when compared to Alexa 647 ($\approx$ 1250 [g/mol]). This could result in steric hindrance, which causes the reduced labeling capacity of Alexa 647 to AvrBs3 (**AB**).

Next, we analyzed if AvrBs3 (**AB**)-Alexa 488- Alexa 647 kept its binding characteristics by employing EMSA. However, it revealed a complete loss of the protein::DNA interaction. Thus the FRET pair

Alexa 488 and Alexa 647 could not be used for this experimental approach. We could proof that AvrBs3 (**AB**)-Alexa 488- Alexa 647 displayed a FRET signal. Consequently, the donor and acceptor fluorophores were close to each other, which indicates towards an aggregation of the protein due to the covalently bound labels.

Further investigations identifying compatible FRET pairs with AvrBs3 (**AB**) binding capacity are required. For instance, the FRET pair Cy3 and Cy5 could be analyzed, but also Cy5 in combination with Alexa 488 could be a promising attempt, as Alexa 488 has shown an advantageous labeling capacity.

In conclusion, we analyzed the *in vitro* binding characteristics of AvrBs3(△ N152-C28) under low salt conditions and revealed a similar K_D-value to the reported value of Meckler *et al.* [142]. The other groups determined the binding characteristics under physiological conditions and measured higher K_D-values [194, 205]. We were able to design two single cysteine variants and one double cysteine variant, which kept their *in vitro* binding characteristics. The competition assays revealed a stronger protein::DNA interaction of AvrBs3(△ N152-C28), when compared to the single cysteine variant. It remains unclear, if the substitution of cysteine to methionine improves gene activation under *in vivo* condition, when compared to the $AvrBs3(C30S)_{Rep}$ designed by Schreiber *et al.* [194].

Labeled single cysteine variants with Alexa 488 and Cy5 kept their binding characteristics and could be used to further investigate the binding mechanisms of the AvrBs3 *in vitro*.

The FRET efficiencies measurements (Ef_A*) with AvrBs3 (**A**)-Cy5 and AvrBs3 (**B**)-Cy5 on DNA()) and DNA(II) were not in concordance withe the expected "close-state" and "far-state". The results indicated towards a displacement of the Cy5 fluorophore when compared to the expected location within the crystal structure.

Finally, a double cysteine AvrBs3 labeled with a donor and an acceptor fluorophore was designed. Isolated labeled AvrBs3 (**AB**) with Alexa 488 or Alexa 647 kept its binding capacity, while consecutively labeling resulted in a loss of function.

Part IV

APPENDIX

A

APPENDIX

A.1 SUPPLEMENTARY EXPERIMENTS

Experimental data additional to the results section.

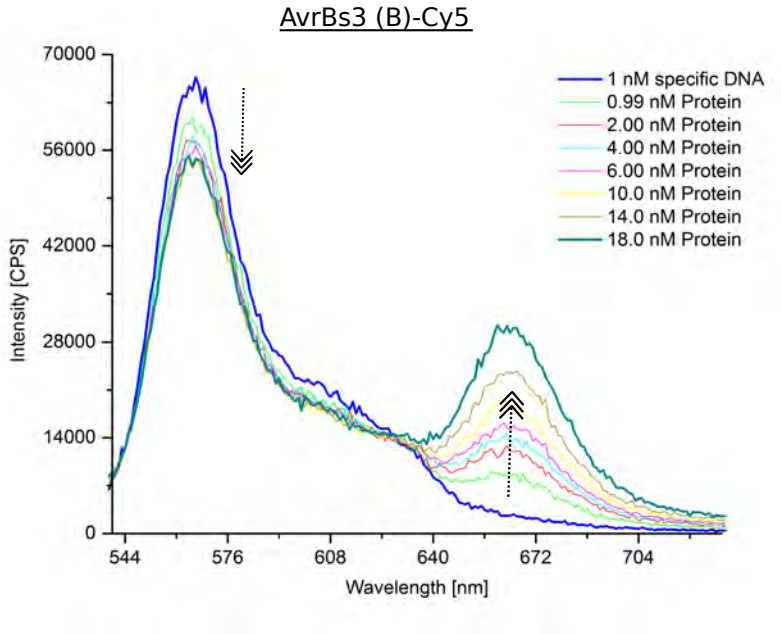

Figure 54: Emission spectra of titrating Cy5 labeled AvrBs3 (B) to 5'-end Cy3 labeled specific DNA (I)
Example of emission spectra. Proteins were titrated in increasing concentrations (0, 0.99, 2, 4, 6, 10, 14, 18 [nM]) to 1 nM DNA (I). AvrBs3 (B)-Cy5 analyzed on DNA (I) (1nM).

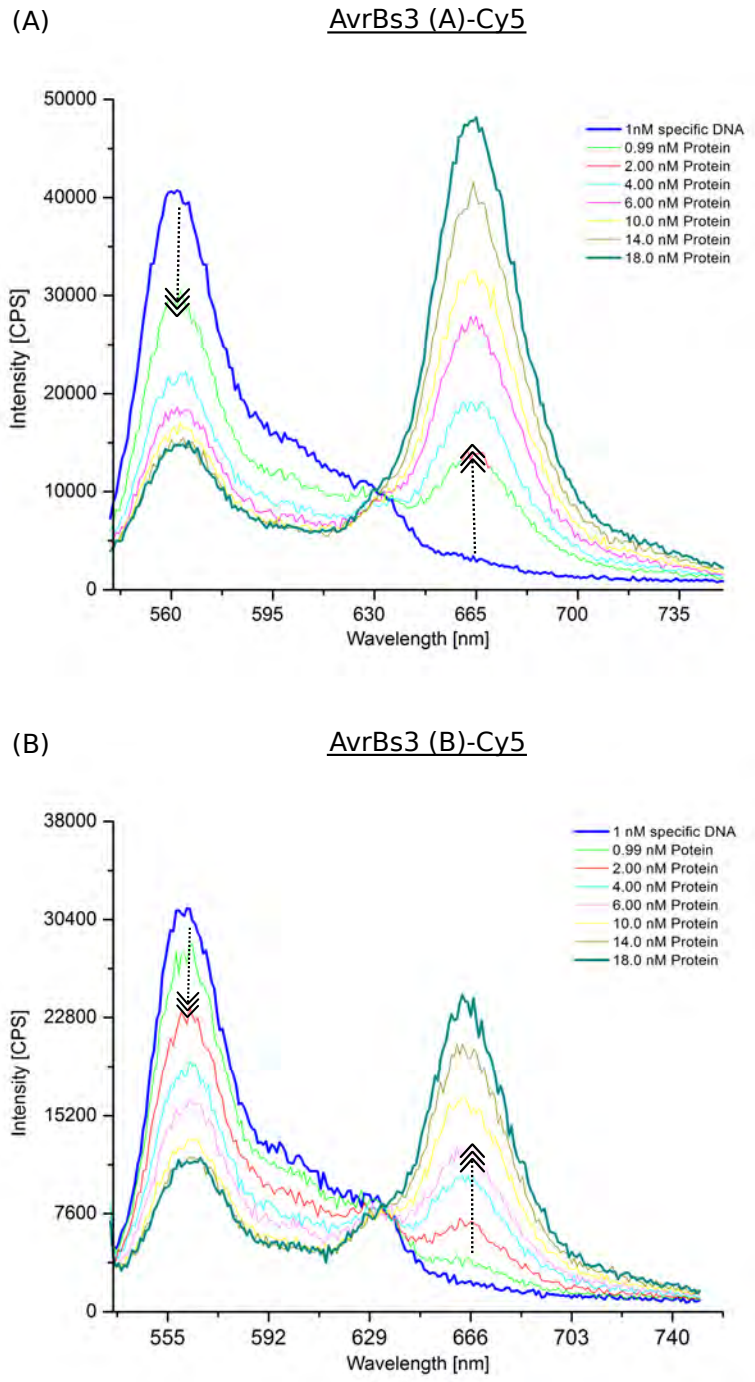

Figure 55: Emission spectra of titrating Cy5 labeled AvrBs3 to 5'-end Cy3 labeled specific DNA (II)
Example of emission spectra. Proteins were titrated in increasing concentrations (0, 0.99, 2, 4, 6, 10, 14, 18 [nM]) to 1 nM DNA (I). (A) AvrBs3 (A)-Cy5 analyzed on DNA (II) (1nM). (B) AvrBs3 (B)-Cy5 analyzed on DNA (II) (1nM)

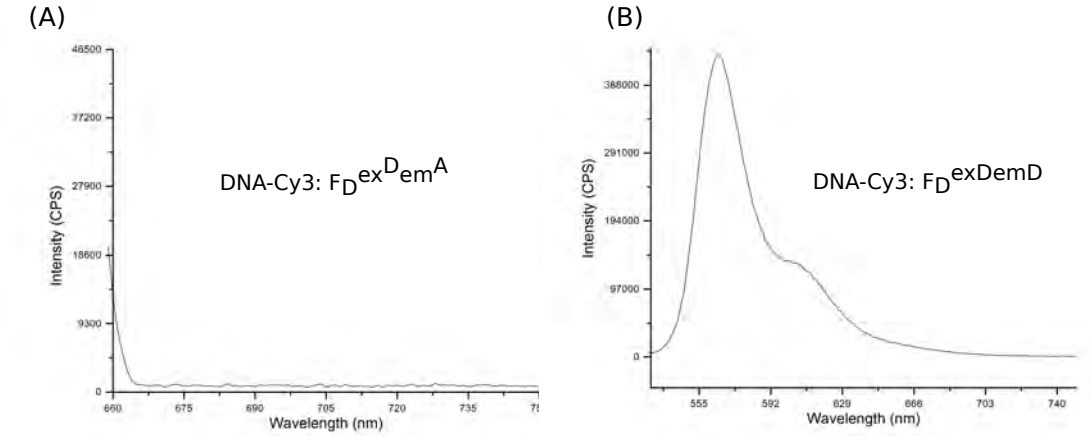

Figure 56: Spectra of DNA-Cy3 to determined the correction factor β
The correction factor β had to be defined to correct the crosstalk between Cy3 and Cy5. Donor fluorophore was excited at 520 nm (ex^D) and emission was detected in the donor channel and in the acceptor channel. The factor β is characterized by the fluorescence signal $F_D{}^{ex_D em_A}$(A) divided with the fluorescence signal $F_D{}^{ex_D em_D}$(B).

BIBLIOGRAPHY

[1] (2012). Method of the year. *Nature Methods*, 9(1). (Cited on page 101.)

[2] Abelson, J., Trotta, C. R., & Li, H. (1998). tRNA splicing. *Journal of Biological Chemistry*, 273(21):12685–12688. (Cited on page 12.)

[3] Ain, Q. U., Chung, J. Y., & Kim, Y.-H. (2015). Current and future delivery systems for engineered nucleases: ZFN,TALEN and RGEN. *Journal of Controlled Release*, 205:120–127. (Cited on pages 6, 18, and 19.)

[4] Akshay, K. V., Jared, G., Christina, G.-V., Renta, H., Dinah, S., Rachel, M., Tony, d. F., & John, M. (2010). A status report on RNAi therapeutics. *Silence*, 1(14). (Cited on page 3.)

[5] Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and psi-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17):3389–3402. (Cited on page 74.)

[6] Alwin, S., Gere, M. B., Guhl, E., Effertz, K., Barbas, C. F., Segal, D. J., Weitzman, M. D., & Cathomen, T. (2005). Custom zinc-finger nucleases for use in human cells. *Molecular Therapy*, 12(4):610–617. (Cited on pages 47, 77, and 108.)

[7] Anders, C., Niewoehner, O., Duerst, A., & Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, 513(7519):569–573. (Cited on pages 10 and 11.)

[8] Aouida, M., Eid, A., Ali, Z., Cradick, T., Lee, C., Deshmukh, H., Atef, A., AbuSamra, D., Gadhoum, S. Z., Merzaban, J., Bao, G., & Mahfouz, M. (2015). Efficient fdCas9 Synthetic endonuclease with improved specificity for precise genome engineering. *PLoS ONE*, 10(7):e0133373. (Cited on page 14.)

[9] Arnould, S., Delenda, C., Grizot, S., Desseaux, C., Pâques, F., Silva, G., & Smith, J. (2011). The I-CreI meganuclease and its engineered derivatives: applications from cell modification to gene therapy. *Protein Engineering Design and Selection*, 24(1-2):27–31. (Cited on page 3.)

[10] Ausubel, F., Brent, R., Kingston, R., Moore, D., Seidmann, J., Smith, J., & Struhl., K. (2012). Current protocols in molecular biology. *John Wiley & Sons*. (Cited on page 38.)

[11] Bainbridge, J. W., Mehat, M. S., Sundaram, V., Robbie, S. J., Barker, S. E., Ripamonti, C., Georgiadis, A., Mowat, F. M., Beattie, S. G., Gardner, P. J., Feathers, K. L., Luong, V. A., Yzer, S., Balaggan, K., Viswanathan, A., de Ravel, T. J., Casteels, I., Holder, G. E., Tyler, N., Fitzke, F. W., Weleber, R. G., Nardini, M., Moore, A. T., Thompson, D. A., Petersen-Jones, S. M., Michaelides, M., van den Born, L. I., Stockman, A., Smith, A. J., Rubin, G., & Ali, R. R. (2015). Long-term effect of gene therapy on leber's congenital amaurosis. *The New England Journal of Medicine*, 372(20):1887–1897. (Cited on page 20.)

[12] Ban, C. & Yang, W. (1998). Structural basis for MutH activation in E.coli mismatch repair and relationship of muth to restriction endonucleases. *The EMBO Journal*, 17(5):1526–1534. (Cited on page 16.)

[13] Bell, O., Tiwari, V. K., Thomä, N. H., & Schübeler, D. (2011). Determinants and dynamics of genome accessibility. *Nat Rev Genet*, 12(8):554–564. (Cited on page 19.)

[14] Bellingrath, J.-S. & Fischer, M. (2015). Gentherapie als behandlungskonzept für erbliche netzhauterkrankungen. *Der Ophthalmologe*, 112(9):720–727–. (Cited on page 20.)

[15] Beltran, W. A., Cideciyan, A. V., Lewin, A. S., Iwabe, S., Khanna, H., Sumaroka, A., Chiodo, V. A., Fajardo, D. S., Román, A. J., Deng, W.-T., Swider, M., Alemán, T. S., Boye, S. L., Genini, S., Swaroop, A., Hauswirth, W. W., Jacobson, S. G., & Aguirre, G. D. (2012). Gene therapy rescues photoreceptor blindness in dogs and paves the way for treating human X-linked retinitis pigmentosa. *Proceedings of the National Academy of Sciences of the United States of America*, 109(6):2132–2137. (Cited on pages 21 and 107.)

[16] Bennett, C. B., Lewis, A. L., Baldwin, K. K., & Resnick, M. A. (1993). Lethality induced by a single site-specific double-strand break in a dispensable yeast plasmid. *Proceedings of the National Academy of Sciences of the United States of America*, 90(12):5613–5617. (Cited on page 101.)

[17] Berney, C. & Danuser, G. (2003). FRET or no FRET: A quantitative comparison. *Biophysical Journal*, 84(6):3992–4010. (Cited on pages 91 and 112.)

[18] Bernstein, D. L., Le Lay, J. E., Ruano, E. G., & Kaestner, K. H. (2015). TALE-mediated epigenetic suppression of CDkn2A increases replication in human fibroblasts. *The Journal of Clinical Investigation*, 125(5):1998–2006. (Cited on page 17.)

[19] Beumer, K. J., Trautman, J. K., Mukherjee, K., & Carroll, D. (2013). Donor DNA utilization during gene targeting with zinc-

finger nucleases. *G3: Genes | Genomes | Genetics*, 3(4):657–664. (Cited on page 4.)

[20] Beurdeley, M., Bietz, F., Li, J., Thomas, S., Stoddard, T., Juillerat, A., Zhang, F., Voytas, D. F., Duchateau, P., & Silva, G. H. (2013). Compact designer TALENs for efficient genome engineering. *Nature Communications*, 4:1762. (Cited on page 12.)

[21] Bickle, T. A. & Krüger, D. H. (1993). Biology of DNA restriction. *Microbiological Reviews*, 57(2):434–450. (Cited on page 12.)

[22] Bietz, F. (2015). *Generation and characterization of TAL effectore nucleases with novel catalytic domains*. PhD thesis, Justus Liebig Universität Giessen. (Cited on page 12.)

[23] Bitinaite, J., Wah, D. A., Aggarwal, A. K., & Schildkraut, I. (1998). FokI dimerization is required for DNA cleavage. *Proceedings of the National Academy of Sciences of the United States of America*, 95(18):10570–10575. (Cited on pages 14 and 104.)

[24] Boch, J. & Bonas, U. (2010). Xanthomonas AvrBs3 family-type III effectors: Discovery and function. *Annu. Rev. Phytopathol*, 48(1):419–436. (Cited on pages 6 and 8.)

[25] Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., & Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, 326(5959):1509–1512. (Cited on pages 6, 8, 9, and 109.)

[26] Bogdanove, A. J., Schornack, S., & Lahaye, T. (2010). TAL effectors: finding plant genes for disease and defense. *Current Opinion in Plant Biology*, 13(4):394–401. (Cited on page 6.)

[27] Boissel, S., Jarjour, J., Astrakhan, A., Adey, A., Gouble, A., Duchateau, P., Shendure, J., Stoddard, B. L., Certo, M. T., Baker, D., & Scharenberg, A. M. (2013). megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering. *Nucleic Acids Research*, 42(4):2591–2601. (Cited on pages 8 and 12.)

[28] Bonas, U., Conrads-Strauch, J., & Balbo, I. (1993). Resistance in tomato to xanthomonas campestris pv vesicatoria is determined by alleles of the pepper-specific avirulence gene AvrBs3. *Molecular and General Genetics*, 238(1-2):261–269. (Cited on page 8.)

[29] Bultmann, S., Morbitzer, R., Schmidt, C. S., Thanisch, K., Spada, F., Elsaesser, J., Lahaye, T., & Leonhardt, H. (2012). Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Research*, 40(12):5368–5377. (Cited on page 19.)

[30] Capecchi, M. (1989). Altering the genome by homologous recombination. *Science*, 244(4910):1288–1292. (Cited on pages 3 and 101.)

[31] Carroll, D. (2011). Genome engineering with zinc-finger nucleases. *Genetics*, 188(4):773–782. (Cited on pages 3 and 5.)

[32] Catto, L. E., Ganguly, S., Milsom, S. E., Welsh, A. J., & Halford, S. E. (2006). Protein assembly and DNA looping by the FokI restriction endonuclease. *Nucleic Acids Research*, 34(6):1711–1720. (Cited on page 14.)

[33] Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J. A., Somia, N. V., Bogdanove, A. J., & Voytas, D. F. (2011). 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 pages 8, 14, 32, 33, 34, 41, 43, 44, 46, 63, and 109.)

[34] Certo, M. T., Ryu, B. Y., Annis, J. E., Garibov, M., Jarjour, J. V., Rawlings, D. J., & Scharenberg, A. M. (2011). Tracking genome engineering outcome at individual DNA breakpoints. *Nature methods*, 8(8):671–676. (Cited on page 47.)

[35] Chapman, J., Taylor, M., & Boulton, S. Playing the end game: DNA double-strand break repair pathway choice. *Molecular Cell*, 47(4):497–510. (Cited on page 4.)

[36] Chen, F., Pruett-Miller, S. M., Huang, Y., Gjoka, M., Duda, K., Taunton, J., Collingwood, T. N., Frodin, M., & Davis, G. D. (2011). High-frequency genome editing using ssdna oligonucleotides with zinc-finger nucleases. *Nature methods*, 8(9):753–755. (Cited on page 4.)

[37] Cheng, X., Balendiran, K., Schildkraut, I., & Anderson, J. E. (1994). Structure of pvuii endonuclease with cognate DNA. *The EMBO Journal*, 13(17):3927–3935. (Cited on pages 15 and 104.)

[38] Cho, H.-S., Gu Kang, J., Lee, J.-H., Lee, J.-J., Jeon, S. K., Ko, J.-H., Kim, D.-S., Park, K.-H., Kim, Y.-S., & Kim, N.-S. (2015). Direct regulation of e-cadherin by targeted histone methylation of taleset fusion protein in cancer cells. *Oncotarget*, 6(27):–. (Cited on page 17.)

[39] Cho, S. W., Kim, S., Kim, J. M., & Kim, J.-S. (2013). Targeted genome engineering in human cells with the cas9 RNA-guided endonuclease. *Nature Biotechnology*, 31(3):230–232. (Cited on page 3.)

[40] Cho, S. W., Kim, S., Kim, Y., Kweon, J., Kim, H. S., Bae, S., & Kim, J. S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Ressearch*, 24(1):132–141. (Cited on pages 16, 17, 58, 101, and 103.)

[41] Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A. J., & Voytas, D. F. (2010). Targeting DNA double-strand breaks with tal effector nucleases. *Genetics*, 186(2):757–761. (Cited on pages 3 and 6.)

[42] Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., & Kuhn, R. (2015). Increasing the efficiency of homology-directed repair for crispr-cas9-induced precise gene editing in mammalian cells. *Nature Biotechnology*, 33(5):543–548. (Cited on pages 4 and 101.)

[43] Ciccia, A. & Elledge, S. J. (2010). The DNA damage response: Making it safe to play with knives. *Molecular Cell*, 40(2):179–204. (Cited on pages 4 and 101.)

[44] Cideciyan, A. V. (2010). Leber congenital amaurosis due to rpe65 mutations and its treatment with gene therapy. *Progress in retinal and eye research*, 29(5):398–427. (Cited on page 20.)

[45] Cideciyan, A. V., Jacobson, S. G., Beltran, W. A., Sumaroka, A., Swider, M., Iwabe, S., Roman, A. J., Olivares, M. B., Schwartz, S. B., Komáromy, A. M., Hauswirth, W. W., & Aguirre, G. D. (2013). Human retinal gene therapy for leber congenital amaurosis shows advancing retinal degeneration despite enduring visual improvement. *Proceedings of the National Academy of Sciences of the United States of America*, 110(6):E517–E525. (Cited on page 20.)

[46] Coligan, J., Dunn, B., Speicher, D., & Wingfield, P. (2012). Current protocols in protein science. *John Wiley & Sons*. (Cited on page 38.)

[47] Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex genome engineering using crispr/cas systems. *Science (New York, N.Y.)*, 339(6121):819–823. (Cited on pages 3, 12, 17, and 75.)

[48] Cong, L., Zhou, R., Kuo, Y.-c., Cunniff, M., & Zhang, F. (2012). Comprehensive interrogation of natural tale DNA binding modules and transcriptional repressor domains. *Nature communications*, 3:968–968. (Cited on pages 8 and 109.)

[49] Cornu, T. I., Thibodeau-Beganny, S., Guhl, E., Alwin, S., Eichtinger, M., Joung, J., & Cathomen, T. (2007). DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Molecular Therapy*, 16(2):352–358. (Cited on page 47.)

[50] Cox, D. B. T., Platt, R. J., & Zhang, F. (2015). Therapeutic genome editing: prospects and challenges. *Nature Medecine*, 21(2):121–131. (Cited on pages 3, 4, 18, and 101.)

[51] Cuculis, L., Abil, Z., Zhao, H., & Schroeder, C. M. (2015). Direct observation of TALE protein dynamics reveals a two-state search mechanism. *Nature Communications*, 6:–. (Cited on pages 10, 22, 95, 110, 111, and 115.)

[52] Davis, L. & Maizels, N. (2011). DNA nicks promote efficient and safe targeted gene correction. *PLoS ONE*, 6(9):e23981–. (Cited on pages 4, 15, 101, and 102.)

[53] Davis, L. & Maizels, N. (2014). Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proceedings of the National Academy of Sciences of the United States of America*, 111(10):E924–E932. (Cited on pages 16, 17, 58, 101, and 102.)

[54] de Lange, O., Schreiber, T., Schandry, N., Radeck, J., Braun, K. H., Koszinowski, J., Heuer, H., Strauß, A., & Lahaye, T. (2013). 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 pages 8 and 9.)

[55] Dehouck, Y., Kwasigroch, J. M., Gilis, D., & Rooman, M. (2011). Popmusic 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinformatics*, 12:151–151. (Cited on page 80.)

[56] Delacôte, F., Perez, C., Guyot, V., Mikonio, C., Potrel, P., Cabaniols, J.-P., Delenda, C., Pâques, F., & Duchateau, P. (2011). Identification of genes regulating gene targeting by a high-throughput screening approach. *Journal of Nucleic Acids*, 2011:947212–. (Cited on page 19.)

[57] Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.-K., Shi, Y., & Yan, N. (2012). Structural basis for sequence-specific recognition of DNA by tal effectors. *Science*, 335(6069):720–723. (Cited on pages 8, 9, 95, 110, 111, and 115.)

[58] DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J., & Church, G. M. (2013). Genome engineering in saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Research*, 41(7):4336–43. (Cited on pages 33, 72, and 106.)

[59] Ding, Q., Lee, Y.-K., Schaefer, E. A. K., Peters, D. T., Veres, A., Kim, K., Kuperwasser, N., Motola, D. L., Meissner, T. B., Hendriks, W. T., Trevisan, M., Gupta, R. M., Moisan, A., Banks, E., Friesen, M., Schinzel, R. T., Xia, F., Tang, A., Xia, Y., Figueroa, E., Wann, A., Ahfeldt, T., Daheron, L., Zhang, F., Rubin, L. L., Peng, L. F., Chung, R. T., Musunuru, K., & Cowan, C. A. (2012). A TALEN genome

editing system to generate human stem cell-based disease models. *Cell stem cell*, 12(2):238–251. (Cited on page 108.)

[60] Doetschman, T., Gregg, R. G., Maeda, N., Hooper, M. L., Melton, D. W., Thompson, S., & Smithies, O. (1987). Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature*, 330(6148):576–578. (Cited on pages 3 and 101.)

[61] Donoho, G., Jasin, M., & Berg, P. (1998). Analysis of gene targeting and intrachromosomal homologous recombination stimulated by genomic double-strand breaks in mouse embryonic stem cells. *Molecular and Cellular Biology*, 18(7):4070–4078. (Cited on pages 3, 4, and 5.)

[62] Doyle, E. L., Booher, N. J., Standage, D. S., Voytas, D. F., Brendel, V. P., VanDyk, J. K., & Bogdanove, A. J. (2012). Tal effector-nucleotide targeter (TALE-NT) 2.0: tools for tal effector design and target prediction. *Nucleic Acids Research*, 40:W117–W122. (Cited on page 108.)

[63] Duca, M., Vekhoff, P., Oussedik, K., Halby, L., & Arimondo, P. B. (2008). The triple helix: 50 years later, the outcome. *Nucleic Acids Research*, 36(16):5123–5138. (Cited on page 6.)

[64] E. Thomas, A. Pingoud, P. F. (2002). An efficient method for the preparation of long heteroduplex DNA as substrate for mismatch repair by the escherichia coli MutHLS system. *Biological Chemistry*. (Cited on page 16.)

[65] Eisenschmidt, K., Lanio, T., Simoncsits, A., Jeltsch, A., Pingoud, V., Wende, W., & Pingoud, A. (2005). Developing a programmed restriction endonuclease for highly specific DNA cleavage. *Nucleic Acids Research*, 33(22):7039–7047. (Cited on pages 15 and 104.)

[66] Elrod-Erickson, M., Benson, T. E., & Pabo, C. O. (1998). High-resolution structures of variant zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure*, 6(4):451–464. (Cited on page 6.)

[67] Engler, C., Gruetzner, R., Kandzia, R., & Marillonnet, S. (2009). Golden gate shuffling: A one-pot DNA shuffling method based on type iis restriction enzymes. *PLoS ONE*, 4(5):e5553. (Cited on page 108.)

[68] Fadel, H. J., Morrison, J. H., Saenz, D. T., Fuchs, J. R., Kvaratskhelia, M., Ekker, S. C., & Poeschla, E. M. (2014). TALEN knockout of the psip1 gene in human cells: Analyses of hiv-1 replication and allosteric integrase inhibitor mechanism. *Journal of Virology*, 88(17):9704–9717. (Cited on page 19.)

[69] Farzadfard, F., Perli, S. D., & Lu, T. K. (2013). Tunable and multifunctional eukaryotic transcription factors based on crispr/-cas. *ACS Synthetic Biology*, 2(10):604–613. (Cited on pages 33, 44, and 72.)

[70] Flechsig, H. (2014). Tales from a spring - superelasticity of tal effector protein structures. *PLoS ONE*, 9(10):e109919–. (Cited on pages 10, 95, 110, and 115.)

[71] Fonfara, I., Curth, U., Pingoud, A., & Wende, W. (2011). Creating highly specific nucleases by fusion of active restriction endonucleases and catalytically inactive homing endonucleases. *Nucleic Acids Research*, 40(2):847–860. (Cited on page 15.)

[72] Fonfara, I., Le Rhun, A., Chylinski, K., Makarova, K. S., Lecrivain, A. L., Bzdrenga, J., Koonin, E. V., & Charpentier, E. (2014). Phylogeny of cas9 determines functional exchangeability of dual-rna and cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Research*, 42(4):2577–90. (Cited on pages 10 and 11.)

[73] Förster, T. (1946). Energiewanderung und fluoreszenz. *Naturwissenschaften*, 33(6):166–175. (Cited on page 52.)

[74] Förster, T. (1948). Zwischenmolekulare energiewanderung und fluoreszenz. *Annalen der Physik*, 437(1-2):55–75. (Cited on page 52.)

[75] Förster, T. (2012). Energy migration and fluorescence. *Journal of Biomedical Optics*, 17(1):0110021–01100210. (Cited on page 52.)

[76] Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., & Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*, 31(9). (Cited on pages 12 and 107.)

[77] Gabriel, R., Lombardo, A., Arens, A., Miller, J. C., Genovese, P., Kaeppel, C., Nowrouzi, A., Bartholomae, C. C., Wang, J., Friedman, G., Holmes, M. C., Gregory, P. D., Glimm, H., Schmidt, M., Naldini, L., & von Kalle, C. (2011). An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nature Biotechnology*, 29(9):816–823. (Cited on pages 5 and 101.)

[78] Gabsalilow, L., Schierling, B., Friedhoff, P., Pingoud, A., & Wende, W. (2013). 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 12, 16, 33, 43, 44, 45, 48, 57, 58, 101, and 102.)

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

[80] Gaj, T., Mercer, A. C., Sirk, S. J., Smith, H. L., & Barbas, C. F. (2013b). A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. *Nucleic Acids Research*, 41(6):3937–3946. (Cited on page 17.)

[81] Gao, H., Wu, X., Chai, J., & Han, Z. (2012). Crystal structure of a TALE protein reveals an extended n-terminal DNA binding region. *Cell Research*, 22(12):1716–1720. (Cited on pages 9, 10, and 110.)

[82] Garg, A., Lohmueller, J. J., Silver, P. A., & Armel, T. Z. (2012). Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Research*, 40(15):7584–7595. (Cited on page 17.)

[83] Gasiunas, G., Sinkunas, T., & Siksnys, V. (2013). Molecular mechanisms of CRISPR-mediated microbial immunity. *Cellular and Molecular Life Sciences*, 71(3):449–465. (Cited on page 10.)

[84] Gaudet, D., Méthot, J., & Kastelein, J. (2012). Gene therapy for lipoprotein lipase deficiency. *Current Opinion in Lipidology*, 23(4):–. (Cited on page 3.)

[85] Gersbach, C. A., Gaj, T., Gordley, R. M., Mercer, A. C., & Barbas, C. F. (2011). Targeted plasmid integration into the human genome by an engineered zinc-finger recombinase. *Nucleic Acids Research*, 39(17):7868–7878. (Cited on page 17.)

[86] Grau, J., Wolf, A., Reschke, M., Bonas, U., Posch, S., & Boch, J. (2013). Computational predictions provide insights into the biology of TAL effector target sites. *PLoS Computational Biology*, 9(3):e1002962. (Cited on page 108.)

[87] Grosse, S., Huot, N., Mahiet, C., Arnould, S., Barradeau, S., Clerre, D. L., Chion-Sotinel, I., Jacqmarcq, C., Chapellier, B., Ergani, A., Desseaux, C., Cédrone, F., Conseiller, E., Pâques, F., Labetoulle, M., & Smith, J. (2010). Meganuclease-mediated inhibition of HSV1 infection in cultured cells. *Molecular Therapy*, 19(4):694–702. (Cited on page 3.)

[88] Guilinger, J. P., Pattanayak, V., Reyon, D., Tsai, S. Q., Sander, J. D., Joung, J. K., & Liu, D. R. (2014a). Broad specificity profiling of TALENs results in engineered nucleases with improved DNA cleavage specificity. *Nature methods*, 11(4):429–435. (Cited on pages 108 and 109.)

[89] Guilinger, J. P., Thompson, D. B., & Liu, D. R. (2014b). Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nature Biotechnology*, 32(6):577–582. (Cited on pages 14 and 101.)

[90] Guo, J., Gaj, T., & Barbas, C. F. (2010). 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 14.)

[91] Gürlebeck, D., Szurek, B., & Bonas, U. (2005). Dimerization of the bacterial effector protein avrbs3 in the plant cell cytoplasm prior to nuclear import. *The Plant Journal*, 42(2):175–187. (Cited on pages 9 and 111.)

[92] Hacein-Bey-Abina, S., von Kalle, C., Schmidt, M., Le Deist, F., Wulffraat, N., McIntyre, E., Radford, I., Villeval, J.-L., Fraser, C. C., Cavazzana-Calvo, M., & Fischer, A. (2003). A serious adverse event after successful gene therapy for x-linked severe combined immunodeficiency. *The New England Journal of Medicine*, 348(3):255–256. (Cited on page 3.)

[93] Halford, S., Catto, L. E., Pernstich, C., Rusling, D. A., & Sanders, K. L. (2011). The reaction mechanism of foki excludes the possibility of targeting zinc finger nucleases to unique dna sites. *Biochemical Society Transactions*. (Cited on pages 5, 14, 101, and 104.)

[94] Hall, M. C. & Matson, S. W. (1999). The escherichia coli mutl protein physically interacts with MutH and stimulates the muth-associated endonuclease activity. *Journal of Biological Chemistry*, 274(3):1306–1312. (Cited on page 16.)

[95] Hinnen, A., Hicks, J. B., & Fink, G. R. (1978). Transformation of yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 75(4):1929–1933. (Cited on pages 3 and 101.)

[96] Hohng, S., Joo, C., & Ha, T. (2004). Single-molecule three-color FRET. *Biophysical Journal*, 87(2):1328–1337. (Cited on pages 90 and 112.)

[97] Holt, N., Wang, J., Kim, K., Friedman, G., Wang, X., Taupin, V., Crooks, G. M., Kohn, D. B., Gregory, P. D., Holmes, M. C., & Cannon, P. M. (2010). Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. *Nature Biotechnology*, 28(8):839–847. (Cited on page 19.)

[98] Hoppe, A., Christensen, K., & Swanson, J. A. (2002). Fluorescence resonance energy transfer-based stoichiometry in living cells. *Biophysical Journal*, 83(6):3652–3664. (Cited on pages 53, 94, 113, and 114.)

[99] Horton, J., Bonventre, J., & Cheng, X. (1998). How is modification of the dna substrate recognized by the PvuII restriction endonuclease. *Biological Chemistry*, 319(4-5):451–8. (Cited on page 13.)

[100] Hosch, J., Lorenz, B., & Stieger, K. (2010). RPGR: Role in the photoreceptor cilium, human retinal disease, and gene therapy. *Ophthalmic Genetics*, 32(1):1–11. (Cited on pages 20, 21, 107, and 108.)

[101] Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G., & Zhang, F. (2013). DNA targeting specificity of RNA-guided cas9 nucleases. *Nat Biotech*, 31(9):827–832. (Cited on pages 12 and 107.)

[102] Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., Peterson, R. T., Yeh, J.-R. J., & Joung, J. K. (2013). Efficient in vivo genome editing using RNA-guided nucleases. *Nature Biotechnology*, 31(3):227–229. (Cited on page 3.)

[103] Isalan, M., Choo, Y., & Klug, A. (1997). Synergy between adjacent zinc fingers in sequence-specific DNArecognition. *Proceedings of the National Academy of Sciences of the United States of America*, 94(11):5617–5621. (Cited on page 6.)

[104] Jackson, S. P. & Bartek, J. (2009). The DNA-damage response in human biology and disease. *Nature*, 461(7267):1071–1078. (Cited on page 101.)

[105] Jacobson, S. G., Cideciyan, A. V., Roman, A. J., Sumaroka, A., Schwartz, S. B., Heon, E., & Hauswirth, W. W. (2015). Improvement and decline in vision with gene therapy in childhood blindness. *The New England Journal of Medicine*, 372(20):1920–1926. (Cited on page 20.)

[106] Jacoby, K. & Scharenberg, A. (2014). Homing endonuclease target determination using selex adapted for yeast surface display. In: *Methods in Molecular Biology*, D. R. Edgell, ed., volume 1123, pages 165–190–. Humana Press. (Cited on page 5.)

[107] Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). CRISPR-assisted editing of bacterial genomes. *Nature Biotechnology*, 31(3):233–239. (Cited on pages 10, 17, and 75.)

[108] Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096):816–821. (Cited on pages 3, 10, 12, 17, and 75.)

[109] Joung, J. K. & Sander, J. D. (2012). TALENs: a widely applicable technology for targeted genome editing. *Nature reviews. Molecular cell biology*, 14(1):49–55. (Cited on page 109.)

[110] Juillerat, A., Bertonati, C., Dubois, G., Guyot, V., Thomas, S., Valton, J., Beurdeley, M., Silva, G. H., Daboussi, F., & Duchateau, P. (2014). BurrH: a new modular DNA binding protein for genome engineering. *Scientific Reports*, 4:3831–. (Cited on page 8.)

[111] Kao, H.-I. & Bambara, R. A. (2003). The protein components and mechanism of eukaryotic okazaki fragment maturation. *Critical Reviews in Biochemistry and Molecular Biology*, 38(5):433–452. (Cited on page 12.)

[112] Karran, P. (2000). DNA double strand break repair in mammalian cells. *Current Opinion in Genetics & Development*, 10(2):144–150. (Cited on page 101.)

[113] Katz, S. S., Gimble, F. S., & Storici, F. (2014). To nick or not to nick: Comparison of i-scei single- and double-strand break-induced recombination in yeast and human cells. *PLoS ONE*, 9(2):e88840–. (Cited on pages 5, 16, and 102.)

[114] Kay, M. A. (2011). State-of-the-art gene-based therapies: the road ahead. *Nature Reviews Genetics*, 12(5):316–328. (Cited on page 3.)

[115] Kay, S., Hahn, S., Marois, E., Hause, G., & Bonas, U. (2007). A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science*, 318(5850):648–651. (Cited on page 6.)

[116] Kim, E., Kim, S., Kim, D. H., Choi, B.-S., Choi, I.-Y., & Kim, J.-S. (2012). Precision genome engineering with programmable DNA-nicking enzymes. *Genome Research*, 22(7):1327–1333. (Cited on pages 15 and 101.)

[117] Kim, Y. G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America*, 93(3):1156–1160. (Cited on pages 6, 14, 19, and 104.)

[118] Kirsch, R. D. & Joly, E. (1998). An improved PCR-mutagenesis strategy for two-site mutagenesis or sequence swapping between related genes. *Nucleic Acids Research*, 26(7):1848–1850. (Cited on page 38.)

[119] Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., Gonzales, A. P. W., Li, Z., Peterson, R. T., Yeh, J.-R. J., Aryee, M. J., & Joung, J. K. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, 523(7561):481–485. (Cited on pages 10 and 11.)

[120] Konermann, S., Brigham, M. D., Trevino, A., Hsu, P. D., Heidenreich, M., Cong, L., Platt, R. J., Scott, D. A., Church, G. M., & Zhang, F. (2013). Optical control of mammalian endogenous transcription and epigenetic states. *Nature*. (Cited on page 17.)

[121] LaFountaine, J. S., Fathe, K., & Smyth, H. D. (2015). Delivery and therapeutic applications of gene editing technologies ZFNs,

TALENs, and CRISPR/Cas9. *International Journal of Pharmaceutics*, 494(1):180–194. (Cited on pages 3, 5, 18, and 20.)

[122] Lahue, R. S., Su, S. S., & Modrich, P. (1987). Requirement for d(GATC) sequences in escherichia coli mutHLS mismatch correction. *Proceedings of the National Academy of Sciences of the United States of America*, 84(6):1482–1486. (Cited on page 16.)

[123] Lakowicz, J. R. (2010). *Principles of Fluorescence Spectroscopy*. Springer. (Cited on page 51.)

[124] Lamb, B. M., Mercer, A. C., & Barbas, C. F. (2013). Directed evolution of the TALE N-terminal domain for recognition of all 5′ bases. *Nucleic Acids Research*, 41(21):9779–9785. (Cited on page 9.)

[125] Lee, J. Y., Chang, J., Joseph, N., Ghirlando, R., Rao, D. N., & Yang, W. (2005). MutH complexed with hemi- and unmethylated DNAs: Coupling base recognition and DNA cleavage. *Molecular Cell*, 20(1):155–166. (Cited on pages 13 and 16.)

[126] Li, H., Haurigot, V., Doyon, Y., Li, T., Wong, S. Y., Bhagwat, A. S., Malani, N., Anguela, X. M., Sharma, R., Ivanciu, L., Murphy, S. L., Finn, J. D., Khazi, F. R., Zhou, S., Paschon, D. E., Rebar, E. J., Bushman, F. D., Gregory, P. D., Holmes, M. C., & High, K. A. (2011). In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature*, 475(7355):217–221. (Cited on page 19.)

[127] Li, J.-F., Aach, J., Norville, J. E., McCormack, M., Zhang, D., Bush, J., Church, G. M., & Sheen, J. (2013). Multiplex and homologous recombination-mediated plant genome editing via guide RNA/Cas9. *Nature Biotechnology*, 31(8):688–691. (Cited on page 72.)

[128] Li, L., Wu, L. P., & Chandrasegaran, S. (1992). Functional domains in Fok i restriction endonuclease. *Proceedings of the National Academy of Sciences of the United States of America*, 89(10):4275–4279. (Cited on page 14.)

[129] Li, T., Huang, S., Jiang, W. Z., Wright, D., Spalding, M. H., Weeks, D. P., & Yang, B. (2010). Tal nucleases (TALNs): hybrid proteins composed of TAL effectors and Foki DNA-cleavage domain. *Nucleic Acids Research*, 39(1):359–372. (Cited on page 14.)

[130] Liang, F., Han, M., Romanienko, P. J., & Jasin, M. (1998). Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proceedings of the National Academy of Sciences*, 95(9):5172–5177. (Cited on pages 3, 4, and 5.)

[131] Lieber, M. R. (2008). The mechanism of human nonhomologous DNA end joining. *Journal of Biological Chemistry*, 283(1):1–5. (Cited on pages 3, 4, and 101.)

[132] Liu, J., Gaj, T., Yang, Y., Wang, N., Shui, S., Kim, S., Kanchiswamy, C. N., Kim, J.-S., & Barbas III, C. F. (2015). Efficient delivery of nuclease proteins for genome editing in human stem cells and primary cells. *Nat. Protocols*, 10(11):1842–1859. (Cited on pages 18 and 19.)

[133] Long, C., McAnally, J. R., Shelton, J. M., Mireault, A. A., Bassel-Duby, R., & Olson, E. N. (2014). Prevention of muscular dystrophy in mice by crispr/cas9-mediated editing of germline DNA. *Science*, 345(6201):1184–1188. (Cited on pages 19 and 20.)

[134] Maeder, M. L., Angstman, J. F., Richardson, M. E., Linder, S. J., Cascio, V. M., Tsai, S. Q., Ho, Q. H., Sander, J. D., Reyon, D., Bernstein, B. E., Costello, J. F., Wilkinson, M. F., & Joung, J. K. (2013a). Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nature Biotechnology*, 31(12):1137–1142. (Cited on page 17.)

[135] Maeder, M. L., Linder, S. J., Reyon, D., Angstman, J. F., Fu, Y., Sander, J. D., & Joung, J. K. (2013b). Robust, synergistic regulation of human gene expression using TALE activators. *Nature methods*, 10(3):243–245. (Cited on page 109.)

[136] Maeder, M. L., Thibodeau-Beganny, S., Osiak, A., Wright, D. A., Anthony, R. M., Eichtinger, M., Jiang, T., Foley, J. E., Winfrey, R. J., Townsend, J. A., Unger-Wallace, E., Sander, J. D., Felix Müller-Lerch, F., Fu, F., Pearlberg, J., Göbel, C., Dassie, J. P., Pruett-Miller, S. M., Porteus, M. H., Sgroi, D. C., Iafrate, A. J., Dobbs, D., McCray, P. B., Cathomen, T., Voytas, D. F., & Joung, J. K. (2008). Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Molecular cell*, 31(2):294–301. (Cited on page 6.)

[137] Maeder, M. L., Thibodeau-Beganny, S., Sander, J. D., Voytas, D. F., & Joung, J. K. (2009). Oligomerized pool enegineering (OPEN): An "open-source" protocol for making customized zinc finger arrays. *Nature protocols*, 4(10):1471–1501. (Cited on page 6.)

[138] Mak, A. N.-S., Bradley, P., Bogdanove, A. J., & Stoddard, B. L. (2012a). TAL effectors: function, structure, engineering and applications. *Current opinion in structural biology*, 23(1):93–99. (Cited on page 10.)

[139] Mak, A. N.-S., Bradley, P., Cernadas, R. A., Bogdanove, A. J., & Stoddard, B. L. (2012b). The crystal structure of TAl effectorPthXo1 bound to its DNA target. *Science*, 335(6069):716–719. (Cited on pages 8, 9, 110, and 111.)

[140] Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., Yang, L., & Church, G. M. (2013). CAS9 transcriptional

activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*, 31(9). (Cited on pages 3, 16, 101, and 103.)

[141] Marcaida, M., Muñoz, I., Blanco, F., Prieto, J., & Montoya, G. (2010). Homing endonucleases: from basics to therapeutic applications. 67(5):727–748. (Cited on page 5.)

[142] Meckler, J. F., Bhakta, M. S., Kim, M.-S., Ovadia, R., Habrian, C. H., Zykovich, A., Yu, A., Lockwood, S. H., Morbitzer, R., Elsäesser, J., Lahaye, T., Segal, D. J., & Baldwin, E. P. (2013). Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Research*, 41(7):4118–4128. (Cited on pages 5, 8, 9, 17, 74, 108, 109, 110, and 116.)

[143] Mendenhall, E. M., Williamson, K. E., Reyon, D., Zou, J. Y., Ram, O., Joung, J. K., & Bernstein, B. E. (2013). Locus-specific editing of histone modifications at endogenous enhancers. *Nature Biotechology*, 31(12):1133–1136. (Cited on page 17.)

[144] Meramveliotaki, C., Kotsifaki, D., Androulaki, M., Hountas, A., Eliopoulos, E., & Kokkinidis, M. (2007). Purification, crystallization, X-ray diffraction analysis and phasing of an engineered single-chain pvuii restriction endonuclease. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 63(Pt 10):836–838. (Cited on page 15.)

[145] Metzger, M. J., McConnell-Smith, A., Stoddard, B. L., & Miller, A. D. (2010). Single-strand nicks induce homologous recombination with less toxicity than double-strand breaks using an AAV vector template. *Nucleic Acids Research*, 39(3):926–935. (Cited on pages 4, 15, 101, and 102.)

[146] Miller, D. G., Wang, P.-R., Petek, L. M., Hirata, R. K., Sands, M. S., & Russell, D. W. (2006). Gene targeting in vivo by adeno-associated virus vectors. *Nature Biotechology*, 24(8):1022–1026. (Cited on page 4.)

[147] Miller, J. C., Holmes, M. C., Wang, J., Guschin, D. Y., Lee, Y.-L., Rupniewski, I., Beausejour, C. M., Waite, A. J., Wang, N. S., Kim, K. A., Gregory, P. D., Pabo, C. O., & Rebar, E. J. (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. *Nature Biotechology*, 25(7):778–785. (Cited on page 14.)

[148] Miller, J. C., Tan, S., Qiao, G., Barlow, K. A., Wang, J., Xia, D. F., Meng, X., Paschon, D. E., Leung, E., Hinkley, S. J., Dulay, G. P., Hua, K. L., Ankoudinova, I., Cost, G. J., Urnov, F. D., Zhang, H. S., Holmes, M. C., Zhang, L., Gregory, P. D., & Rebar, E. J. (2011). A TALE nuclease architecture for efficient genome editing. *Nature Biotechology*, 29(2):143–148. (Cited on pages 3, 6, 9, 14, and 17.)

[149] Minczuk, M., Papworth, M. A., Miller, J. C., Murphy, M. P., & Klug, A. (2008). Development of a single-chain, quasi-dimeric zinc-finger nuclease for the selective degradation of mutated human mitochondrial DNA. *Nucleic Acids Research*, 36(12):3926–3938. (Cited on page 14.)

[150] Miyanari, Y., Ziegler-Birling, C., & Torres-Padilla, M.-E. (2013). Live visualization of chromatin dynamics with fluorescent TALEs. *Nat Struct Mol Biol*, 20(11):1321–1324. (Cited on page 17.)

[151] Modrich, P. & Lahue, R. (1996). Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu. Rev. Biochem.*, 65(1):101–133. (Cited on page 16.)

[152] Morbitzer, R., Römer, P., Boch, J., & Lahaye, T. (2010). Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, 107(50):21617–21622. (Cited on pages 7, 8, and 17.)

[153] Moscou, M. J. & Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science*, 326(5959):1501–1501. (Cited on pages 8 and 9.)

[154] Moynahan, M. E. & Jasin, M. (2010). Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat Rev Mol Cell Biol*, 11(3):196–207. (Cited on page 4.)

[155] Mussolino, C., Alzubi, J., Fine, E. J., Morbitzer, R., Cradick, T. J., Lahaye, T., Bao, G., & Cathomen, T. (2014). TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity. *Nucleic Acids Research*, 42(10):6762–6773. (Cited on pages 3, 6, and 17.)

[156] Mussolino, C., Morbitzer, R., Lütge, F., Dannemann, N., Lahaye, T., & Cathomen, T. (2011). A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Research*, 39(21):9283–9293. (Cited on pages 6 and 17.)

[157] Niu, Y., Tenney, K., Li, H., & Gimble, F. S. (2008). Engineering variants of the I-SceI homing endonuclease with strand-specific and site-specific DNA nicking activity. *Journal of molecular biology*, 382(1):188–202. (Cited on pages 16 and 101.)

[158] Orlando, S. J., Santiago, Y., DeKelver, R. C., Freyvert, Y., Boydston, E. A., Moehle, E. A., Choi, V. M., Gopalan, S. M., Lou, J. F., Li, J., Miller, J. C., Holmes, M. C., Gregory, P. D., Urnov, F. D., & Cost, G. J. (2010). Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. *Nucleic Acids Research*, 38(15):e152–e152. (Cited on page 4.)

[159] Orr-Weaver, T. L., Szostak, J. W., & Rothstein, R. J. (1981). Yeast transformation: a model system for the study of recombination. *Proceedings of the National Academy of Sciences of the United States of America*, 78(10):6354–6358. (Cited on pages 3 and 101.)

[160] Osborn, M. J., Starker, C. G., McElroy, A. N., Webber, B. R., Riddle, M. J., Xia, L., DeFeo, A. P., Gabriel, R., Schmidt, M., Von Kalle, C., Carlson, D. F., Maeder, M. L., Joung, J. K., Wagner, J. E., Voytas, D. F., Blazar, B. R., & Tolar, J. (2013). TALEN-based gene correction for epidermolysis bullosa. *Molecular Therapy*, 21(6):1151–1159. (Cited on page 108.)

[161] Ousterout, D. G., Perez-Pinera, P., Thakore, P. I., Kabadi, A. M., Brown, M. T., Qin, X., Fedrigo, O., Mouly, V., Tremblay, J. P., & Gersbach, C. A. (2013). Reading frame correction by targeted genome editing restores dystrophin expression in cells from duchenne muscular dystrophy patients. *Molecular Therapy*, 21(9):1718–1726. (Cited on pages 19 and 20.)

[162] Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., & Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. *Protein Science : A Publication of the Protein Society*, 4(11):2411–2423. (Cited on page 45.)

[163] Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E., Doudna, J. A., & Liu, D. R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnology*, 31(9):839–843. (Cited on pages 12 and 107.)

[164] Pauwels, K., Podevin, N., Breyer, D., Carroll, D., & Herman, P. (2014). Engineering nucleases for gene targeting: safety and regulatory considerations. *New Biotechnology*, 31(1):18–27. (Cited on page 19.)

[165] Pelletier, V., Jambou, M., Delphin, N., Zinovieva, E., Stum, M., Gigarel, N., Dollfus, H., Hamel, C., Toutain, A., Dufier, J.-L., Roche, O., Munnich, A., Bonnefont, J.-P., Kaplan, J., & Rozet, J.-M. (2007). Comprehensive survey of mutations in RP2 and RPGR in patients affected with distinct retinal dystrophies: genotype-phenotype correlations and impact on genetic counseling. *Human Mutation*, 28(1):81–91. (Cited on pages 21, 107, and 108.)

[166] Perez, E. E., Wang, J., Miller, J. C., Jouvenot, Y., Kim, K. A., Liu, O., Wang, N., Lee, G., Bartsevich, V. V., Lee, Y.-L., Guschin, D. Y., Rupniewski, I., Waite, A. J., Carpenito, C., Carroll, R. G., S Orange, J., Urnov, F. D., Rebar, E. J., Ando, D., Gregory, P. D., Riley, J. L., Holmes, M. C., & June, C. H. (2008). Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nature Biotechology*, 26(7):808–816. (Cited on page 19.)

[167] Pernstich, C. & Halford, S. E. (2011). Illuminating the reaction pathway of the Foki restriction endonuclease by fluorescence resonance energy transfer. *Nucleic Acids Research*, 40(3):1203–1213. (Cited on pages 14 and 104.)

[168] Perrin, A., Buckle, M., & Dujon, B. (1993). Asymmetrical recognition and activity of the I-SceI endonuclease on its site and on intron-exon junctions. *The EMBO Journal*, 12(7):2939–2947. (Cited on page 74.)

[169] Pingoud, A., Fuxreiter, M., Pingoud, V., & Wende, W. (2005). Type II restriction endonucleases: structure and mechanism. 62(6):685–707. (Cited on pages 12, 15, and 104.)

[170] Pingoud, A. & Jeltsch, A. (2001). Structure and function of type II restriction endonucleases. *Nucleic Acids Research*, 29(18):3705–3727. (Cited on pages 15 and 104.)

[171] Pingoud, A. & Wende, W. (2011). Generation of novel nucleases with extended specificity by rational and combinatorial strategies. *ChemBioChem*, 12(10):1495–1500. (Cited on pages 5 and 6.)

[172] Polstein, L. R. & Gersbach, C. A. (2015). A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nature Chemical Biology*, 11(3):198–200. (Cited on page 17.)

[173] Qasim, W., Amrolia, P. J., Samarasinghe, S., Ghorashian, S., Zhan, H., Stafford, S., Butler, K., Ahsan, G., Gilmour, K., Adams, S., Pinner, D., Chiesa, R., Chatters, S., Swift, S., Goulden, N., Pegg, K., Thrasher, A. J., Veys, P., & Pule, M. First clinical application of talen engineered universal CAR19 T cells in B-ALL (oral and poster abstracts; session 801: Gene therapy and transfer: Poster i). *Presented at the 57th Annual Meeting and Exposition, December 5-8, 2015, Orlando*. (Cited on page 20.)

[174] Qi, L., Larson, M., Gilbert, L., Doudna, J., Weissman, J., Arkin, A., & Lim, W. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152(5):1173–1183. (Cited on pages 12 and 17.)

[175] Raleigh, E. & Brooks, J. (1998). Restriction modification systems: Where they are and what they do. In: *Bacterial Genomes*, F. de Bruijn, J. Lupski, & G. Weinstock, ed., pages 78–92–. Springer US. (Cited on page 12.)

[176] Ramalingam, S., Kandavelou, K., Rajenderan, R., & Chandrasegaran, S. (2010). Creating designed zinc finger nucleases with minimal cytotoxicity. *Journal of molecular biology*, 405(3):630–641. (Cited on pages 3 and 5.)

[177] Ramirez, C. L., Certo, M. T., Mussolino, C., Goodwin, M. J., Cradick, T. J., McCaffrey, A. P., Cathomen, T., Scharenberg, A. M., & Joung, J. K. (2012). Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Research*, 40(12):5560–5568. (Cited on page 16.)

[178] Ramirez, C. L., Foley, J. E., Wright, D. A., Muller-Lerch, F., Rahman, S. H., Cornu, T. I., Winfrey, R. J., Sander, J. D., Fu, F., Townsend, J. A., Cathomen, T., Voytas, D. F., & Joung, J. K. (2008). Unexpected failure rates for modular assembly of engineered zinc fingers. *Nature Methods*, 5(5):374–375. (Cited on page 6.)

[179] Ran, F. A., Cong, L., Yan, W. X., Scott, D. A., Gootenberg, J. S., Kriz, A. J., Zetsche, B., Shalem, O., Wu, X., Makarova, K. S., Koonin, E. V., Sharp, P. A., & Zhang, F. (2015). In vivo genome editing using staphylococcus aureus Cas9. *Nature*, 520(7546):186–191. (Cited on page 11.)

[180] Ran, F. A., Hsu, P. D., Lin, C.-Y., Gootenberg, J. S., Konermann, S., Trevino, A., Scott, D. A., Inoue, A., Matoba, S., Zhang, Y., & Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, 154(6):1380–1389. (Cited on pages 16, 101, 103, and 104.)

[181] Redondo, P., Prieto, J., Munoz, I. G., Alibes, A., Stricher, F., Serrano, L., Cabaniols, J.-P., Daboussi, F., Arnould, S., Perez, C., Duchateau, P., Paques, F., Blanco, F. J., & Montoya, G. (2008). Molecular basis of xeroderma pigmentosum group c DNA recognition by engineered meganucleases. *Nature*, 456(7218):107–111. (Cited on pages 3 and 5.)

[182] Richardson, C., Moynahan, M. E., & Jasin, M. (1998). Double-strand break repair by interchromosomal recombination: suppression of chromosomal translocations. *Genes & Development*, 12(24):3831–3842. (Cited on pages 3, 4, and 5.)

[183] Richter, A., Streubel, J., Blücher, C., Szurek, B., Reschke, M., Grau, J., & Boch, J. (2014). A TAL effector repeat architecture for frameshift binding. *Nature Communications*, 5. (Cited on page 9.)

[184] Rogers, J. M., Barrera, L. A., Reyon, D., Sander, J. D., Kellis, M., Keith Joung, J., & Bulyk, M. L. (2015). Context influences on TALE-DNA binding revealed by quantitative profiling. *Nat Commun*, 6:–. (Cited on pages 108 and 109.)

[185] Römer, P., Hahn, S., Jordan, T., Strauß, T., Bonas, U., & Lahaye, T. (2007). Plant pathogen recognition mediated by promoter activation of the pepper bs3 resistance gene. *Science*, 318(5850):645–648. (Cited on page 7.)

[186] Rothkamm, K., Krüger, I., Thompson, L. H., & Löbrich, M. (2003). Pathways of DNA double-strand break repair during the mammalian cell cycle. *Molecular and Cellular Biology*, 23(16):5706–5715. (Cited on page 4.)

[187] San Filippo, J., Sung, P., & Klein, H. (2008). Mechanism of eukaryotic homologous recombination. *Annu. Rev. Biochem.*, 77(1):229–257. (Cited on pages 3 and 101.)

[188] Sanders, K. L., Catto, L. E., Bellamy, S. R. W., & Halford, S. E. (2009). Targeting individual subunits of the foki restriction endonuclease to specific DNA strands. *Nucleic Acids Research*, 37(7):2105–2115. (Cited on pages 14, 15, and 104.)

[189] Santiago, Y., Chan, E., Liu, P.-Q., Orlando, S., Zhang, L., Urnov, F. D., Holmes, M. C., Guschin, D., Waite, A., Miller, J. C., Rebar, E. J., Gregory, P. D., Klug, A., & Collingwood, T. N. (2008). Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proceedings of the National Academy of Sciences*, 105(15):5809–5814. (Cited on page 3.)

[190] Schierling, B., Dannemann, N., Gabsalilow, L., Wende, W., Cathomen, T., & Pingoud, A. (2011). A novel zinc-finger nuclease platform with a sequence-specific cleavage module. *Nucleic Acids Research*, 40(6):2623–2638. (Cited on pages 15, 68, and 105.)

[191] Schiml, S., Fauser, F., & Puchta, H. (2014). The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in arabidopsis resulting in heritable progeny. *Plant J*, 80(6):1139–1150. (Cited on pages 16, 58, 72, and 101.)

[192] Schornack, S., Peter, K., Bonas, U., & Lahaya, T. (2005). Expression levels of avrBs3 -like genes affect recognition specificity in tomato Bs4 - but not in pepper Bs3 -mediated perception. *Molecular Plant-Microbe Interactions*, 18(11):1215–25. (Cited on page 8.)

[193] Schreiber, T. & Bonas, U. (2014). Repeat 1 of TAL effectors affects target specificity for the base at position zero. *Nucleic Acids Research*. (Cited on pages 5, 8, 108, and 109.)

[194] Schreiber, T., Sorgatz, A., List, F., Blüher, D., Thieme, S., Wilmanns, M., & Bonas, U. (2015). Refined requirements for protein regions important for activity of the TALE AvrBs3. *PLoS ONE*, 10(3). (Cited on pages 5, 9, 109, 111, and 116.)

[195] Schröder, C. M.deringer, L. (2008). The pymol molecular graphics system. Version 1.1r.1. (Cited on page 94.)

[196] Segal, P., B., P., B., Dreier, Effertz, Huber, Koksch, L., M., & andSe, V. (2003). Evaluation of a modular strategy for the construction of novel polydactyl zinc finger dna-binding proteins. *Biochemistry*, 42(7):2137–2148. (Cited on page 6.)

[197] Sharma, S. (2007). Age-related nonhomologous end joining activity in rat neurons. *Brain Research Bulletin*, 73(1-3):48–54. (Cited on page 4.)

[198] Sharon, D., Sandberg, M. A., Rabe, V. W., Stillberger, M., Dryja, T. P., & Berson, E. L. RP2 and RPGR mutations and clinical correlations in patients with X-linked retinitis pigmentosa. *The American Journal of Human Genetics*, 73(5):1131–1146. (Cited on pages 21, 107, and 108.)

[199] Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., Wang, L., Hodgkins, A., Iyer, V., Huang, X., & Skarnes, W. C. (2014). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nature Methods*, 11(4):399–402. (Cited on pages 16, 58, and 101.)

[200] Sikorski, R. S. & Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in saccharomyces cerevisiae. *Genetics*, 122(1):19–27. (Cited on page 74.)

[201] Silva, G., Poirot, L., Galetto, R., Smith, J., Montoya, G., Duchateau, P., & Pâques, F. (2010). Meganucleases and other tools for targeted genome engineering: Perspectives and challenges for gene therapy. *Current Gene Therapy*, 11(1):11–27. (Cited on pages 3 and 5.)

[202] Simoncsits, A., Tjörnhammar, M.-L., Raskó, T., Kiss, A., & Pongor, S. (2001). Covalent joining of the subunits of a homodimeric type ii restriction endonuclease: single-chain PvuII endonuclease1. *Journal of Molecular Biology*, 309(1):89–97. (Cited on pages 15 and 69.)

[203] Söllü, C., Pars, K., Cornu, T. I., Thibodeau-Beganny, S., Maeder, M. L., Joung, J. K., Heilbronn, R., & Cathomen, T. (2010). Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. *Nucleic Acids Research*, 38(22):8269–8276. (Cited on page 47.)

[204] Stella, S., Molina, R., López-Méndez, B., Juillerat, A., Bertonati, C., Daboussi, F., Campos-Olivas, R., Duchateau, P., & Montoya, G. (2014). Bud, a helix-loop-helix DNA-binding domain for genome modification. *Acta Crystallographica D*, 70(7):2042–2052. (Cited on page 8.)

[205] Stella, S., Molina, R., Yefimenko, I., Prieto, J., Silva, G., Bertonati, C., Juillerat, A., Duchateau, P., & Montoya, G. (2013). 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 pages 5, 7, 8, 9, 80, 82, 83, 84, 91, 92, 93, 109, 110, 113, 114, and 116.)

[206] Stoddard, B. L. (2005). Homing endonuclease structure and function. *Quarterly Reviews of Biophysics*, 38(01):49–95. (Cited on page 5.)

[207] Storici, F., Durham, C. L., Gordenin, D. A., & Resnick, M. A. (2003). Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proceedings of the National Academy of Sciences*, 100(25):14994–14999. (Cited on pages 3, 4, and 5.)

[208] Sugisaki Hiroyuki, K. S. (1981). New restriction endonucleases from flavobacterium okeanokoites (FokI) and micrococcus luteus (mlui). *Gene*, 16(8):73–78. (Cited on page 13.)

[209] Sun, N., Liang, J., Abil, Z., & Zhao, H. (2012). Optimized tal effector nucleases (TALENs) for use in treatment of sickle cell disease. *Molecular BioSystems*, 8(4):1255–1263. (Cited on page 9.)

[210] Sun, N. & Zhao, H. (2014). A single-chain TALEN architecture for genome engineering. *Molecular BioSystems*, 10(3):446–453. (Cited on pages 14 and 101.)

[211] Sung, Y. H., Baek, I.-J., Kim, D. H., Jeon, J., Lee, J., Lee, K., Jeong, D., Kim, J.-S., & Lee, H.-W. (2013). Knockout mice created by TALEN-mediated gene targeting. *Nature Biotechnology*, 31(1):23–24. (Cited on page 3.)

[212] Szczepek, M., Brondani, V., Buchel, J., Serrano, L., Segal, D. J., & Cathomen, T. (2007). Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nature Biotechology*, 25(7):786–793. (Cited on page 14.)

[213] Takashi Mino, Yasuhiro Aoyama, T. S. (2009). Efficient double-stranded DNA cleavage by artificial zinc-finger nucleases composed of one zinc-finger protein and a single-chain FokI dimer. *Journal of Biotechnology*. (Cited on page 14.)

[214] Tebas, P., Stein, D., Tang, W. W., Frank, I., Wang, S. Q., Lee, G., Spratt, S. K., Surosky, R. T., Giedlin, M. A., Nichol, G., Holmes, M. C., Gregory, P. D., Ando, D. G., Kalos, M., Collman, R. G., Binder-Scholl, G., Plesa, G., Hwang, W.-T., Levine, B. L., & June, C. H. (2014). 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 19.)

[215] Tesson, L., Usal, C., Menoret, S., Leung, E., Niles, B. J., Remy, S., Santiago, Y., Vincent, A. I., Meng, X., Zhang, L., Gregory, P. D., Anegon, I., & Cost, G. J. (2011). Knockout rats generated by embryo microinjection of TALENs. *Nature Biotechnology*, 29(8):695–696. (Cited on page 108.)

[216] Thomas, C. E., Ehrhardt, A., & Kay, M. A. (2003). Progress and problems with the use of viral vectors for gene therapy. *Nat Rev Genet*, 4(5):346–358. (Cited on page 3.)

[217] Thompson, D. A., Khan, N. W., Othman, M. I., Chang, B., Jia, L., Grahek, G., Wu, Z., Hiriyanna, S., Nellissery, J., Li, T., Khanna, H., Colosi, P., Swaroop, A., & Heckenlively, J. R. (2012). Rd9 is a naturally occurring mouse model of a common form of retinitis pigmentosa caused by mutations in RPGR-ORf15. *PLoS ONE*, 7(5):e35865. (Cited on page 21.)

[218] Thyme, S. B., Song, Y., Brunette, T. J., Szeto, M. D., Kusak, L., Bradley, P., & Baker, D. (2014). Massively parallel determination and modeling of endonuclease substrate specificity. *Nucleic Acids Research*, 42(22):13839–13852. (Cited on page 5.)

[219] Tsai, S. Q., Wyvekens, N., Khayter, C., Foden, J. A., Thapar, V., Reyon, D., Goodwin, M. J., Aryee, M. J., & Joung, J. K. (2014). Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature Biotechnology*, 32(6):569–576. (Cited on pages 14 and 17.)

[220] Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., & Gregory, P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nature Reviews Genetics*, 11(9):636–646. (Cited on pages 3 and 5.)

[221] Valton, J., Dupuy, A., Daboussi, F., Thomas, S., Maréchal, A., Macmaster, R., Melliand, K., Juillerat, A., & Duchateau, P. (2012). Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. *The Journal of Biological Chemistry*, 287(46):38427–38432. (Cited on pages 8 and 19.)

[222] van Nierop, G. P., de Vries, A. A. F., Holkers, M., Vrijsen, K. R., & Goncalves, M. A. F. V. (2009). Stimulation of homology-directed gene targeting at an endogenous human locus by a nicking endonuclease. *Nucleic Acids Research*, 37(17):5725–5736. (Cited on pages 4, 15, 101, and 102.)

[223] Vervoort, R., Lennon, A., Bird, A. C., Tulloch, B., Axton, R., Miano, M. G., Meindl, A., Meitinger, T., Ciccodicola, A., & Wright, A. F. (2000). Mutational hot spot within a new RPGR exon in X-linked retinitis pigmentosa. *Nature Genetics*, 25(4):462–466. (Cited on pages 21, 107, and 108.)

[224] Wah, D. A., Bitinaite, J., Schildkraut, I., & Aggarwal, A. K. (1998). Structure of FokI has implications for DNA cleavage. *Proceedings of the National Academy of Sciences of the United States of America*, 95(18):10564–10569. (Cited on pages 13 and 14.)

[225] Wah, D. A., Hirsch, J. A., Dorner, L. F., Schildkraut, I., & Aggarwal, A. K. (1997). Structure of the multimodular endonuclease FokI bound to DNA. *Nature*, 388(6637):97–100. (Cited on page 68.)

[226] Wan, H., Hu, J.-p., Li, K.-s., Tian, X.-h., & Chang, S. (2013). Molecular dynamics simulations of DNA-free and DNA-bound tal effectors. *PLoS ONE*, 8(10):e76045. (Cited on pages 8, 10, 95, 110, and 115.)

[227] Wang, J., Friedman, G., Doyon, Y., Wang, N. S., Li, C. J., Miller, J. C., Hua, K. L., Yan, J. J., Babiarz, J. E., Gregory, P. D., & Holmes, M. C. (2012). Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. *Genome Research*, 22(7):1316–1326. (Cited on page 16.)

[228] Wang, X., Wang, Y., Wu, X., Wang, J., Wang, Y., Qiu, Z., Chang, T., Huang, H., Lin, R.-J., & Yee, J.-K. (2015). Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nature Biotechnology*, 33(2):175–178. (Cited on pages 5, 17, and 101.)

[229] Welsh, K. M., Lu, A. L., Clark, S., & Modrich, P. (1987). Isolation and characterization of the escherichia coli muth gene product. *Journal of Biological Chemistry*, 262(32):15624–15629. (Cited on page 16.)

[230] Wicky, B. I. M., Stenta, M., & Dal Peraro, M. (2013). TAL effectors specificity stems from negative discrimination. *PLoS ONE*, 8(11):e80261–. (Cited on pages 8, 9, 95, 110, and 115.)

[231] Wiedenheft, B., Sternberg, S. H., & Doudna, J. A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature*, 482(7385):331–338. (Cited on page 10.)

[232] Wirth, T., Parker, N., & Ylä-Herttuala, S. (2013). History of gene therapy. *Gene*, 525(2):162–169. (Cited on page 3.)

[233] Wright, A. F., Chakarova, C. F., Abd El-Aziz, M. M., & Bhattacharya, S. S. (2010). Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait. *Nature Reviews Genetics*, 11(4):273–284. (Cited on page 20.)

[234] Wu, Z., Hiriyanna, S., Qian, H., Mookherjee, S., Campos, M. M., Gao, C., Fariss, R., Sieving, P. A., Li, T., Colosi, P., & Swaroop, A. (2015). A long-term efficacy study of gene replacement therapy for

RPGR-associated retinal degeneration. *Human Molecular Genetics*. (Cited on pages 20, 21, and 107.)

[235] Wyvekens, N., Topkar, V. V., Khayter, C., Joung, J. K., & Tsai, S. Q. (2015). Dimeric CRISPR RNA-Guided FokI-dCas9 Nucleases directed by truncated gRNAs for highly specific genome editing. *Human Gene Therapy*, 26(7):425–431. (Cited on page 14.)

[236] Yang, W. (2008). An equivalent metal ion in one- and two-metal-ion catalysis. *Nature Structural & Molecular Biology*, 15(11):1228–1231. (Cited on page 12.)

[237] Yang, W. (2011). Nucleases: diversity of structure, function and mechanism. *Quarterly Reviews of Biophysics*, 44(01):1–93. (Cited on page 12.)

[238] Yanik, M. (2013). *Generierung hochspezifischer Nukleasen für die zielgerichtete Spatlung genomischer DNA-sequenzen*. PhD thesis, Justus Liebig Universität Giessen. (Cited on pages 12, 15, 69, 101, 104, 105, 106, and 110.)

[239] Yanik, M., Alzubi, J., Lahaye, T., Cathomen, T., Pingoud, A., & Wende, W. (2013). Tale-PvuII fusion proteins - novel tools for gene targeting. *PLoS ONE*, 8(12):e82539. (Cited on pages 15, 33, 63, 68, 69, 101, 104, and 105.)

[240] Yant, S. R., Huang, Y., Akache, B., & Kay, M. A. (2007). Site-directed transposon integration in human cells. *Nucleic Acids Research*, 35(7):e50–e50. (Cited on page 17.)

[241] Ye, L., Wang, J., Beyer, A. I., Teque, F., Cradick, T. J., Qi, Z., Chang, J. C., Bao, G., Muench, M. O., Yu, J., Levy, J. A., & Kan, Y. W. (2014). Seamless modification of wild-type induced pluripotent stem cells to the natural CCR5Δ32 mutation confers resistance to HIV infection. *Proceedings of the National Academy of Sciences*, 111(26):9591–9596. (Cited on page 19.)

[242] Youvan, D., Silva, C., & Yang, M. (1997). Calibration of fluorescence resonance energy transfer in microscopy using genetically engineered GFP derivatives on nickel chelating beads. *Biotechnology*, 3:1–18. (Cited on pages 52, 91, and 112.)

[243] Zetsche, B., Gootenberg, J., Abudayyeh, O., Slaymaker, I., Makarova, K., Essletzbichler, P., Volz, S., Joung, J., van der Oost, J., Regev, A., Koonin, E., & Zhang, F. Cpf1 is a single rna-guided endonuclease of a class 2 Crispr-cas system. *Cell*, 163(3):759–771. (Cited on page 11.)

[244] Zeug, A., Woehler, A., Neher, E., & Ponimaskin, E. Quantitative intensity-based fret approaches–a comparative snapshot. *Biophysical Journal*, 103(9):1821–1827. (Cited on pages 52, 53, 91, and 112.)

[245] Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G., & Arlotta, P. (2011). Programmable sequence-specific transcriptional regulation of mammalian genome using designer tal effectors. *Nature Biotechnology*, 29(2):149–153. (Cited on page 17.)

[246] Zhang, Y., Zhang, F., Li, X., Baller, J. A., Qi, Y., Starker, C. G., Bogdanove, A. J., & Voytas, D. F. (2012). Transcription activator-like effector nucleases enable efficient plant genome engineering. *Plant Physiology*, 161(1):20–27. (Cited on pages 33 and 41.)

[247] Zhao, Y., Dai, Z., Liang, Y., Yin, M., Ma, K., He, M., Ouyang, H., & Teng, C.-B. (2014). Sequence-specific inhibition of microRNA via CRISPR/CRISPRi system. *Scientific Reports*, 4:3943–. (Cited on page 12.)

[248] Zheng, C.-K., Wang, C.-L., Zhang, X.-P., Wang, F.-J., Qin, T.-F., & Zhao, K.-J. (2014). The last half-repeat of transcription activator-like effector (TALE) is dispensable and thereby tale-based technology can be simplified. *Molecular Plant Pathology*, 15(7):690–697. (Cited on page 8.)

[249] Zhou, H., Liu, B., Weeks, D. P., Spalding, M. H., & Yang, B. (2014). Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. *Nucleic Acids Research*, 42(17):10903–10914. (Cited on page 72.)

[250] Zhu, Z., Quimby, A., Guan, S., Sun, D., Huang, Y., Lai, X., Chan, S.-h., Li, X., Xu, S.-y., & Zhang, C. (2009). High fidelity restriction endonucleases. (Cited on page 15.)

"Es gibt Reichtümer, an denen man zugrunde geht, wenn man sie nicht mit anderen teilen kann."

— Momo oder Die seltsame Geschichte von den Zeit-Dieben und von dem Kind, das den Menschen die gestohlene Zeit zurückbrachte
von Michael Ende

ACKNOWLEDGMENTS

- I would like to thank Prof. Dr. Peter Friedhoff for becoming my first referee. Especially, I like to thank him for his enthusiasm about science, which steadily encouraged me to improve.

- I also like to thank Prof. Dr. Alexander Brehm for becoming my second referee and for his support as mine IRTG supervisor.

- In remembrance of Prof. Dr. Alfred Pingoud I like to thank him for accepting me in the IRTG and giving me the opportunity to be a part of the biochemistry institute at the JLU Gießen. In the short time I got to know him, he taught me a lot about science and about life.

- My thank also go to Dr. Wolfgang Wende for his kind support.

- I like to thank Prof. Dr. Dr. Knut Stieger for being a outstanding cooperation partner.

- I like to thank Dr. Mert Yanik for his practical advises at the beginning of my thesis, but also for being an excellent cooperation partner. I will keep in mind his big smile and hope all the best for his future.

- I also like to thank Professor Dr. Virginijus Siksnys for accepting to be my international IRTG advisor.

- My thank also go to Prof. Dr. Katja Sträßer for the smoothly association of the groups.

My thank go to all my coworker and friends who accompany me during the last three years:

- I like to thank Dr. Lilia Gabsalilow for teaching me the basics in biochemistry technicals and for all the fun hours together in our lab. She was the best partner to play tricks on Andi, Dennis and Laki and a very good room mate on our journeys.

- Also I like to thank Sabrina Stiehler for helping me in all imaginable situations...I will miss our everyday talks!

- I would like to thank Dr. Fabian Bietz for all his support in the lab and his patient to explain to me the LyX system. Particularly I enjoined to discuss with him about the coloration of our figures and about the most important question "what will be cooked tonight".

- My thank also go to Julia Gotthard for the lovely trip to Hannover, for all the discussions during lunch time and for her support in the lab.

- I like to thank Dr. Maximilian Reuter for always providing support in the lab and I hope we will still meet up for that bottle of wine.

- I like to thank Dr. Anja Drescher for organization of the unforgettable trips with the IRTG (Moscow, St. Petersburg, and Rauischholzhausen). This unique time I will always keep in mind. I also like to thank Anja for her cheerful attitude and wish her all the best in Munich.

- I like to thank Ina and Karina for all support I got to fill in all possible documents.

- I like to thank Petra for being the best cleaning woman a lab can imagine.

- My thank go to Heike, Rashmi, Birte, Christoph and Pingping for completing our lovely group. I wish Pingping and her growing family all the best for the future.

- My thanks go to Dr. Bernhard Remes for his kind support and for the outstanding discussions during our coffee breaks.

- My thank go to Kristina Kovac'for the extraordinary discussion about life, while emptying a bottle of wine (or maybe two...).

- I like to thank my sister for her support, particular at end end of the thesis.

- I like to thank Philipp for being my backup in all imaginable life situation. I like to thank him for his patient and for his support. I enjoy every second with you and hope we will have plenty of them in our common future.

- At the end I like to thank my parents for always supporting me and believing in me. Thanks for all possibilities in life I got from you and thanks that I can always trust you with my small and big problems which cross my way.

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, January 2016

Laura Waltl