

**Entwicklung hochspezifischer Nickasen
für *gene targeting*-Anwendungen**

Inauguraldissertation

zur Erlangung des Grades

Doktor der Naturwissenschaften

- Dr. rer. nat. -

des Fachbereichs Biologie und Chemie

der Justus-Liebig-Universität Gießen

vorgelegt von

Lilia Gabsalilow, M.Sc.

Gießen 2014

*Первое, что находишь в жизни, последнее, к чему
протягиваешь руки, самое ценное в жизни.*

In Liebe meiner Familie gewidmet

Die vorliegende Arbeit wurde am Institut für Biochemie des Fachbereichs 08 – Biologie und Chemie - der Justus-Liebig Universität Gießen im Zeitraum von November 2010 bis Juni 2014 unter der Leitung von Prof. Dr. Alfred M. Pingoud angefertigt und durch die von der DFG finanzierte *International Research Training Group* (IRTG) „Enzymes and Multienzyme complexes acting on nucleic acids“ (GRK 1384) unterstützt.

Erstgutachter: **Prof. Dr. Alfred M. Pingoud**
Institut für Biochemie
Fachbereich 08 – Biologie und Chemie
Justus-Liebig Universität Gießen
Heinrich-Buff-Ring 58-62
35392 Gießen

Zweitgutachter: **Prof. Dr. Rainer Renkawitz**
Institut für Genetik
Fachbereich 08 – Biologie und Chemie
Justus-Liebig Universität Gießen
Heinrich-Buff-Ring 58-62
35392 Gießen

Publikationen

Gabsalilow, L., B. Schierling, P. Friedhoff, A. Pingoud & W. Wende (2013). Site- and strand-specific nicking of DNA by fusion proteins derived from MthH and I-SceI or TALE repeats. *Nucleic Acids Res.* 41, e83.

Schierling, B., N. Dannemann, L. Gabsalilow, W. Wende, T. Cathomen & A. Pingoud (2012). A novel zinc-finger nuclease platform with a sequence-specific cleavage module. *Nucleic Acids Res.* 40, 2623-38.

Publikation in Vorbereitung

Gabsalilow, L., B. Schierling, W. Wende, P. Friedhoff & A. Pingoud. Controllable nicking or double-strand cleavage with a site-specific nuclease derived from monomeric HinP1I and TALE repeats.

Eingereichte Patente

Pingoud, A., Wende, W., Yanik, M., Friedhoff, P., Gabsalilow, L., & Schierling, B. (2012). Chimeric nucleases for gene targeting. WO2013045480 A1.

Erklärung

Ich versichere, dass ich die vorgelegte Dissertation selbständig und ohne unzulässige fremde Hilfe angefertigt und verfasst habe. Jegliche Hilfsmittel wurden angegeben und alle Zitate, die ich andern Veröffentlichungen dem Wortlaut oder Sinn entsprechend entnommen habe, sind durch Quellenangaben kenntlich gemacht worden.

(Lilia Gabsalilow)

Gießen, den 21.06.2014

Abkürzungsverzeichnis

μ	mikro- (10 ⁻⁶)
% (v/v)	Volumenprozent pro Volumen
% (w/v)	Gewichtsprozent pro Volumen
°C	Grad Celsius
A	
A	Amper
Abb.	Abbildung
APS	Ammoniumpersulfat
As	Aminosäure
B	
bp	Basenpaar
BSA	Bovine Serum Albumin
bzw.	beziehungsweise
C	
D	
Da	Dalton
DNA	Desoxyribonukleinsäure
dNTP	Desoxyribonukleosidtriphosphat
DTT	1,4-Dithiothreitol
E	
EDTA	Ethylendiaminteraacetat
<i>et al.</i>	und andere
ErBr	Ethidiumbromid
EtOH	Ethanol
F	
fwd	<i>forward</i> = vorwärts
G	
x g	Schwerebeschleunigung an der Erdoberfläche
g	Gramm
H	
h	Stunde(n)
I	
IPTG	Isopropyl-β-D-Thiogalaktopyranosid
J	
K	
k	Kilo (10 ³)
L	
l	Liter
LB	Luria-Bertani
lin	linear
M	
m	milli-(10 ⁻³)
m	Meter
M	Molar
min	Minute(n)
N	
n	nano- (10 ⁻⁹)

nt	Nukleotid
O	
oc	<i>open circular</i>
OD _{600nm}	Optische Dichte bei 600 nm
o/n	über Nacht
P	
PAGE	Polyacryamid-Gelelektrophorese
PCR	Polymerase Kettenreaktion
PMSF	Phenylmethylsulfonylfluorid
Q	
R	
rev	<i>revers</i> = rückwärts
rpm	Umdrehungen pro Minute
S	
sc	<i>supercoiled</i>
SDS	Natriumdodecylsulfat
sek	Sekunde(n)
T	
Tab.	Tabelle
TAE	Tris-Acetat-EDTA
TBE	Tris-Borat-EDTA
TEMED	N,N,N',N',-Tetramethylethylendiamin
TPE	Tris-Phosphat-EDTA
Tris	Tris-(hydroxymethyl)-aminomethan
U	
U	<i>units</i> (Enzymaktivität)
UV	Ultraviolett
V	
V	Volt
Vol.	Volumen
W	
wt	Wildtyp
X	
Y	
Z	

Als Dezimaltrennzeichen wird in dieser Arbeit ein Komma verwendet. Aminosäuren werden mit den international üblichen Drei-Buchstaben- bzw. Ein-Buchstaben-Codes abgekürzt.

Inhaltsverzeichnis

1 EINLEITUNG.....	1
1.1 GENE TARGETING.....	1
1.2 HOCHSPEZIFISCHE NUKLEASEN	1
1.2.1 NATÜRLICH VORKOMMENDE HOCHSPEZIFISCHE NUKLEASEN	2
1.2.1.1 Homing Endonukleasen (HE)	2
1.2.1.2 RNA-vermittelte Nukleasen (CRISPR/Cas)	4
1.2.2 ARTIFIZIELLE HOCHSPEZIFISCHE NUKLEASEN	5
1.2.2.1 DNA-Bindungsmodul: Zinkfinger (ZF)	5
1.2.2.2 DNA-Bindungsmodul: TALE	7
1.2.2.3 DNA-Spaltungsmodule: FokI und Alternativen	9
1.3 ZYTOTOXIZITÄT UND HOCHSPEZIFISCHE NICKASEN	11
1.4 FUSIONS-KONSTRUKTE	11
1.4.1 DNA-BINDUNGSMODUL: I-SCEI	12
1.4.2 DNA-BINDUNGSMODUL: AVRBS3 BZW. AVRBS4	13
1.4.3 DNA-SPALTUNGSMODUL: MUTH	13
1.4.4 DNA-SPALTUNGSMODUL: HINP1I	16
1.5 ZIELSETZUNG.....	18
2 MATERIAL UND METHODEN	20
2.1 MATERIALIEN.....	20
2.1.1 CHEMIKALIEN	20
2.1.2 PUFFER UND MEDIEN	22
2.1.3 KITS	25
2.1.4 SÄULENMATRIX	25
2.1.5 DIALYSESCHLÄUCHE	25
2.1.6 <i>PRIMER</i> UND OLIGONUKLEOTIDE	26

2.1.7 ENZYME UND MARKER	27
2.2 METHODEN	27
2.2.1 MIKROBIOLOGISCHE METHODEN	27
2.2.1.1 Plasmide	27
2.2.1.2 <i>E.coli</i> Stämme	29
2.2.1.3 <i>E.coli</i> Viabilitätstest	30
2.2.2 MOLEKULARBIOLOGISCHE METHODEN	31
2.2.2.1 Klonierungen	31
2.2.2.2 Denaturierende Polyacrylamid-Gelelektrophorese	35
2.2.3 BIOCHEMISCHEN METHODEN	37
2.2.3.1 Proteinexpression	37
2.2.3.2 Tandem-Reinigung der Fusionskonstrukte	37
2.2.3.4 DNA-Spalttests	39
2.2.3.5 TEV Proteolyse des Peptidlinkers von Muth-I-SceI	41
3 ERGEBNISSE	42
3.1. KONSTRUKTION DER HOCHSPEZIFISCHEN NUKLEASEN	42
3.2 TANDEM-REINIGUNG DER FUSIONS KONSTRUKTE	43
3.3 SUBSTRAT-DESIGN	44
3.4 CHARAKTERISIERUNG DES FUSIONS KONSTRUKTS MUTH-I-SCEI	46
3.4.1 ANALYSE DER AKTIVITÄT UND SPEZIFITÄT VON MUTH-I-SCEI	46
3.4.2 INAKTIVIERUNG VON MUTH-I-SCEI DURCH PROTEOLYTISCHE SPALTUNG DES PEPTIDLINKERS	48
3.4.3 BESTIMMUNG DER STRANGPRÄFERENZ VON MUTH-I-SCEI	49
3.5 CHARAKTERISIERUNG DES FUSIONS KONSTRUKTS TALE-MUTH	50
3.5.1 ANALYSE DER AKTIVITÄT UND SPEZIFITÄT VON TALE-MUTH	50
3.5.2 BESTIMMUNG DER STRANGPRÄFERENZ VON TALE-MUTH MITTELS FLUOROPHOR-MARKIERTEN POLYNUKLEOTID-SUBSTRATEN	53
3.5.3 UNTERSUCHUNG VERSCHIEDENER LINKER-VARIANTEN VON TALE-MUTH	56
	IX

3.5.4 SPALTUNG EINER TRIPARTITEN ERKENNUNGSSEQUENZ DURCH TALE-MUTH	58
3.5.5 STEIGERUNG DER KATALYTISCHEN AKTIVITÄT VON TALE-MUTH	58
3.5.6 <i>E. COLI</i> VIABILITÄTSTEST	59
3.6 CHARAKTERISIERUNG DES FUSIONSKONSTRUKTS TALE-HINP1I	61
3.6.1 ANALYSE UND OPTIMIERUNG DER SPEZIFITÄT VON TALE-HINP1I	61
3.6.2 ERMITTLUNG DER AKTIVITÄT VON TALE-HINP1I EADA208 MIT VERSCHIEDENEN ADRESSIERTEN SUBSTRATEN	63
3.6.4 BESTIMMUNG DER STRANGPRÄFERENZ VON TALE-HINP1I	67
4 DISKUSSION	71
4.1 GENE TARGETING MIT HOCHSPEZIFISCHEN NUKLEASEN	71
4.2 UNTERSUCHUNG DER HOCHSPEZIFISCHEN NICKASE MUTH-I-SCEI	72
4.3 UNTERSUCHUNG DER HOCHSPEZIFISCHEN NICKASE TALE-MUTH	74
4.3.1 DOPPELSTRANG-SPALTUNG MIT TALE-MUTH	76
4.3.2 LINKER-OPTIMIERUNG VON TALE-MUTH	77
4.3.3 STEIGERUNG DER KATALYTISCHEN AKTIVITÄT DES TALE-MUTH FUSIONSKONSTRUKTS	78
4.3.2 <i>IN VIVO</i> UNTERSUCHUNG MÖGLICHER TOXISCHER EFFEKTE DURCH TALE-MUTH	79
3.4 UNTERSUCHUNG DER HOCHSPEZIFISCHEN NUKLEASE TALE-HINP1I	81
4.4.1 AKTIVITÄTS-ANALYSE DER VARIANTE TALE-HINP1I EADA208	83
4.4.2 BESTIMMUNG DER HALBWERTSZEIT VON TALE-HINP1I EADA208	84
4.4.3 STRANGPRÄFERENZ VON TALE-HINP1I EADA208	84
4.5 FAZIT UND AUSBLICK	85
5 ZUSAMMENFASSUNG	87
7 DANKSAGUNG	89
6 LITERATURVERZEICHNIS	91

1 Einleitung

1.1 gene targeting

Unter *gene targeting* versteht man allgemein die gezielte Modifikation im Erbmaterial. Komplexe Genome ortsspezifisch manipulieren zu können, hat vor allem auf dem gentherapeutischen Forschungsgebiet immer mehr an Bedeutung gewonnen. Die Möglichkeit, bestimmte Gene aus dem Genom entfernen, sie an definierten Orten im Genom integrieren oder sie gegeneinander austauschen zu können, würde einen enorm wichtigen Fortschritt in der Therapie von monogenen Erbkrankheiten und der Erforschung von Gen-Funktionen bedeuten. Um dies praktisch umzusetzen, bedient sich eine vielversprechende *gene targeting*-Methode dem Prinzip der homologen Rekombination (HR). HR ist ein zelleigener Mechanismus, welcher der genetischen Rekombination während der Meiose und der Reparatur von DNA-Doppelstrangbrüchen vor der Mitose dient. Voraussetzung dafür sind DNA-Abschnitte mit gleicher oder ähnlicher (homologer) Sequenzabfolge, die rekombiniert werden können. Diese Eigenschaft kann beim *gene targeting* genutzt werden, indem gewünschte DNA-Abschnitte als Reparaturvorlage (DNA-Reparatur-Matrix) in die Zelle eingebracht werden, um dort einen unerwünschten homologen DNA-Bereich im Genom entsprechend zu verändern. Das bloße Einbringen der DNA-Reparatur-Matrix in die Zelle reicht jedoch nicht aus, da das Ereignis der HR zu selten auftritt (~ 1 HR pro 10^6), als dass es effektiv genutzt werden könnte. Ein zusätzlicher DNA-Doppelstrangbruch im bzw. in unmittelbarer Nähe des zu ersetzenden DNA-Abschnitts erhöht die Rate der HR jedoch um ein Vielfaches (Kolb *et al.* 2005, Pingoud & Silva 2007). Die wissenschaftliche Herausforderung, die sich daraus ergibt, besteht darin, einen spezifischen DNA-Strangbruch in einen exakt definierten Bereich des Genoms einzufügen. Eine vielversprechende Aussicht, dies bewerkstelligen zu können, bieten hochspezifische Nukleasen.

1.2 Hochspezifische Nukleasen

Hochspezifische Nukleasen müssen grundsätzlich zwei Kriterien erfüllen. (i) Es muss gewährleistet sein, dass sie eine singuläre Sequenz spezifisch binden und erkennen, (ii) und dass sie innerhalb oder in der Nähe dieser Sequenz einen Einzelstrang- bzw. Doppelstrangbruch katalysieren (Pingoud & Wende 2011). Neben den natürlich vorkommenden hochspezifischen Nukleasen, gibt es auch künstlich fusionierte hochspezifischen Nukleasen, die für das *gene targeting* eingesetzt werden (Abb. 1.1).

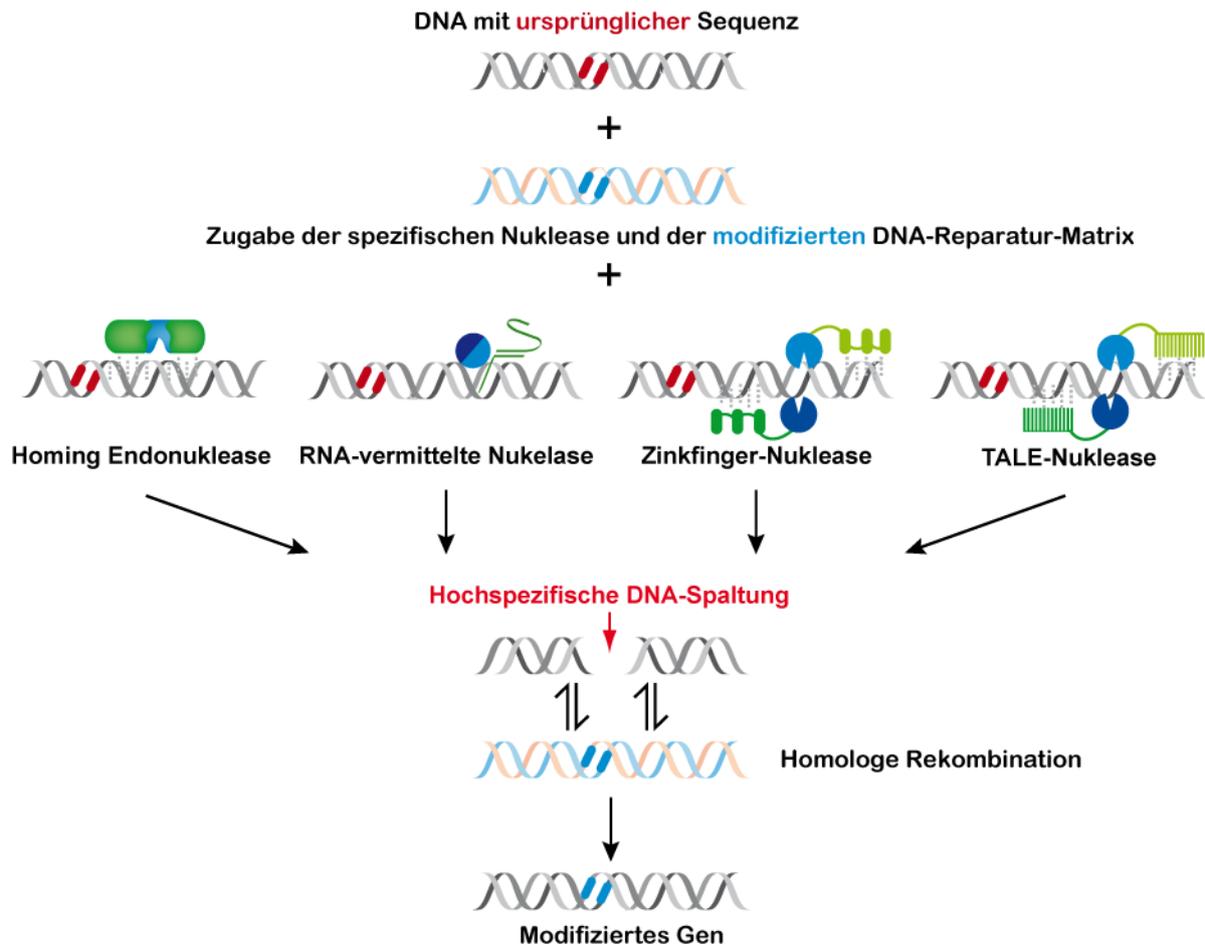


Abbildung 1.1: Schematische Darstellung der *gene targeting*-Methode mittels hochspezifischer Nukleasen. Unter Verwendung von hochspezifischen Nukleasen, wie Homing Endonukleasen, RNA-vermittelten Nukleasen, Zinkfinger-Nukleasen und TALE-Nukleasen, wird eine spezifische Sequenz (rot) erkannt und gespalten. Bei gleichzeitiger Zugabe einer homologen DNA-Reparatur-Matrix mit der gewünschten Zielsequenz (blau), wird durch homologe Rekombination die ursprüngliche Sequenz modifiziert. Verändert nach Pingoud & Wende 2011.

1.2.1 Natürlich vorkommende hochspezifische Nukleasen

Unter den unzähligen natürlich vorkommenden sequenzspezifischen Nukleasen gibt es nur einige wenige, deren Sequenzspezifität ausreicht, eine singuläre Sequenz in einem komplexen DNA-Kontext zu erkennen. Prominente Beispiele für Nukleasen, deren natürliche Funktion dahingehend modifiziert werden kann, um als hochspezifische Nukleasen zu fungieren, sind Homing Endonukleasen und RNA-vermittelte Nukleasen (Pauwels *et al.* 2014).

1.2.1.1 Homing Endonukleasen (HE)

Die genetische Kodierung von Homing Endonukleasen (HE) findet man üblicherweise in Introns und Inteinen (Stoddard 2005). *Homing* beschreibt den Transfer einer speziellen Sequenz

in ein homologes Allel, das diese Sequenz nicht aufweist (Dujon 1989). Dieser Vorgang führt somit zwangsläufig zur Weiterverbreitung dieser speziellen Sequenz, die daher auch als mobiles Element bezeichnet wird. Im einfachsten Fall ist die HE innerhalb des mobilen Elements kodiert und katalysiert nach der Translation einen DNA-Doppelstrangbruch in einem homologen Bereich (Stoddard 2005), wodurch die Reparatur durch homologe Rekombination stimuliert wird (Abb. 1.2). Die Sequenzspezifität von HE kann 12-40 Basenpaare (bp) betragen, was das statistische Vorkommen sehr rar macht (~ 1 pro 7×10^{10} bp). HE sind in fünf Familien untergliedert (LAGLIDADG, GIY-YIG, His-Cys box, H-N-H, PD-(D/E)-XK), wobei die LAGLIDADG HE die größte Familie bilden. Auf Grund der hohen Spezifität für lange Erkennungssequenzen und der gleichzeitigen katalytischen Aktivität von HE wurden diese im Zuge der Forschung auf dem Gebiet des *gene targeting* intensiv eingesetzt und weiterentwickelt (Pauwels *et al.* 2014, Stoddard 2014). Bei den offensichtlichen Vorteilen der Verwendung von HE als hochspezifische Nukleasen bleibt jedoch eine enorme Hürde bestehen, denn die Zahl der natürlich vorkommenden HE und daher auch ihre Erkennungssequenzen sind limitiert und die gezielte Veränderung auf eine gewünschte Zielsequenz nur teilweise erfolgreich und schwierig durchführbar (Grizot *et al.* 2009). HE sind nicht ohne weiteres für spezielle Erkennungssequenzen programmierbar, was für die gezielte Modifizierung bestimmter Bereiche im Genom allerdings unumgänglich ist. HE könnten aber auch verwendet werden, um sichere Bereiche im Genom, sogenannte *safe harbor*, zu adressieren (Sadelain *et al.* 2012). Diese Strategie dient der Insertion von Genen, jedoch nicht am ursprünglichen Bestimmungsort, sondern vorzugsweise in unkritischen Genombereichen. Diese Strategie hat zwar den Vorteil, dass der *safe harbor* für diverse Insertion dienen kann, allerdings muss die Erkennungssequenz für die HE erst einmal eingebracht werden. Diese Strategie wird u.a. in Kulturpflanzen angewendet (D'Halluin *et al.* 2008).

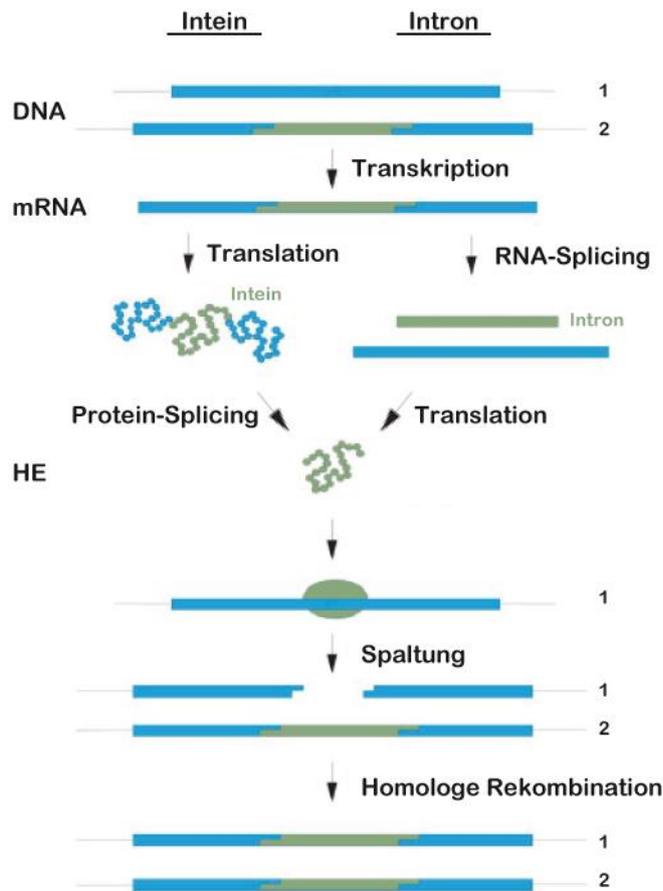


Abbildung 1.2: Schematische Darstellung des Mechanismus von Homing Endonukleasen. Das mobile Element (grün) befindet sich zunächst in Allel 2, fehlt jedoch in Allel 1. Nach der Transkription wird die Homing Endonuklease (HE) entweder direkt translatiert und liegt zunächst als Intein vor, das im Folgeschritt durch Protein-Splicing zur funktionellen Form prozessiert wird (linke Seite) oder die Prozessierung findet auf RNA-Ebene statt und das Intron kodiert direkt für die funktionale Form des Proteins (rechte Seite). Anschließend katalysiert die HE in Allel 1 einen Doppelstrangbruch (DSB) in einem Bereich, der homolog zu dem DNA-Abschnitt zu dem mobilen Element auf Allel 2 ist. Der DSB wird durch homologe Rekombination repariert, wobei eine Kopie des mobilen Elements in das Allel 1 integriert wird. Verändert nach Pingoud *et al.* 2003.

1.2.1.2 RNA-vermittelte Nukleasen (CRISPR/Cas)

Das CRISPR/Cas System (*clustered regularly interspaced short palindromic repeat/CRISPR-associated*) beschreibt ein adaptives Immunsystem in Prokaryoten, um sich vor zellfremder DNA zu schützen (Barrangou *et al.* 2007). Hierbei verfügt der potentielle Wirtsorganismus über kurze konservierte Wiederholungssequenzen (*repeats*), die wiederum durch individuelle Sequenzen (*spacer*) unterbrochen sind (CRISPR). In unmittelbarer Nähe der CRISPR befinden sich die Cas-Gene. Bei Befall des Wirtsorganismus werden Sequenzbereiche der zellfremden DNA als *spacer* in das eigene Genom integriert (Terns & Terns 2011, Wiedenheft *et al.* 2012). Bei erneutem Befall wird diese *spacer*-Sequenz zur crRNA (*cursor RNA*) prozessiert und rekrutiert ein Cas-Protein zum komplementären Bereich der eingedrungenen DNA, um diese durch Hydrolyse eines Doppelstrangbruchs unschädlich zu machen (Brouns *et al.* 2008). Jinek *et al.* 2012 konnten zeigen, dass Cas9 (eine Komponente des TypII-CRISPR/Cas-Systems) in Kombination mit einer chimären RNA eine funktionale RNA-vermittelte Nuklease bildet, die für jede beliebige Sequenz programmiert werden kann, indem die entsprechende crRNA verwendet wird (Abb. 1.3). Die chimäre RNA ist aus der crRNA und der tracrRNA (*trans-*

activating crRNA), die die Hydrolyse von Doppelstrangbrüchen im TypII-CRISPR/Cas-System unterstützt, fusioniert. Diese neuartige und zugleich leicht programmierbare hochspezifische Nuklease ist im Bereich des *gene targeting* stark in den Fokus gerückt und wurde seither in vielen Studien erfolgreich angewendet (Cong *et al.* 2013, DiCarlo *et al.* 2013, Dickinson *et al.* 2013, Hwang *et al.* 2013, Karvelis *et al.* 2013a, Karvelis *et al.* 2013b, Mali *et al.* 2013, Wang *et al.* 2013).

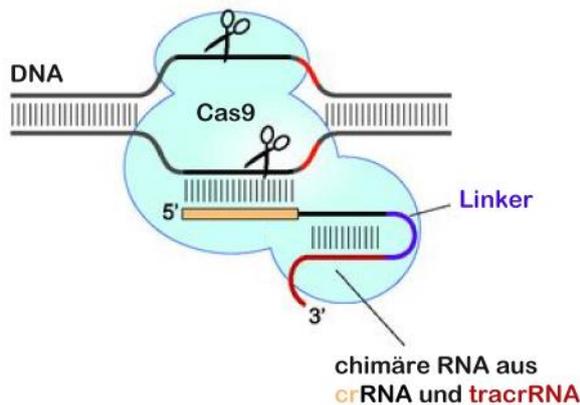


Abbildung 1.3: Sequenzspezifische Spaltung mit einer RNA-vermittelten Nuklease. Die *cursor*RNA (crRNA, beige/schwarz) ist über einen Nukleotid-Linker (blau) mit der *trans-activating* crRNA (tracrRNA, dunkelrot) verbunden und bilden die chimäre RNA. Diese rekrutiert die Endonuklease Cas9 (hellblau) zu einer DNA-Sequenz, die komplementär zur crRNA ist (beige). Cas9 hydrolysiert simultan einen Einzelstrangbruch in den gegenüberliegenden DNA-Strängen, woraus ein Doppelstrangbruch resultiert. Verändert nach Jinek *et al.* 2012.

1.2.2 Artifizielle hochspezifische Nukleasen

Hochspezifische Nukleasen können mittels rekombinanter Proteine künstlich generiert werden. Hierbei werden Proteine oder Proteindomänen, die eine erweiterte DNA-Spezifität aufweisen (DNA-Bindungsmodul) mit jenen kombiniert, die einen DNA-Strangbruch katalysieren können (DNA-Spaltungsmodul). Um eine singuläre Sequenz innerhalb eines komplexen Genoms, wie z.B. dem humanen ($3,3 \times 10^9$ Basenpaare), spezifisch erkennen zu können, muss die Erkennungssequenz der rekombinanten Nuklease mindestens 16 (4^4) Basenpaare betragen. Statistisch kommt diese Sequenz dann nur alle 4^{16} ($4,3 \times 10^9$) Basenpaare vor. Zudem ist es für die Anwendung von enormer Wichtigkeit, jede beliebige Sequenz adressieren zu können, d.h. die rekombinante Nuklease muss für eine bestimmte Zielsequenz programmierbar sein. Diesen besonderen Anforderungen entsprechen die beim *gene targeting* vielfach eingesetzten Zinkfinger-Nukleasen und TALE-Nukleasen (Carroll 2014, Pauwels *et al.* 2014).

1.2.2.1 DNA-Bindungsmodul: Zinkfinger (ZF)

Zinkfinger sind Proteine, die ein bestimmtes Motiv aufweisen. Dabei handelt es sich meist um zwei Cystein- und zwei Histidin-Reste, die in einer $\beta\beta\alpha$ -Struktur ein einzelnes Zink-Ion koordinieren (Cys2His2-Zinkfinger-Motiv) und dadurch maßgeblich die Struktur vorgeben

(Pabo *et al.* 2001) (Abb. 1.4 A). Ein einzelner Zinkfinger erkennt über bestimmte Aminosäure-Reste in der α -Helix eine spezifische Abfolge von drei Basenpaaren (Pavletich & Pabo 1991). Mehrere Zinkfinger mit unterschiedlichen Sequenzspezifitäten können tandem-artig hintereinander geschaltet vorliegen (Multi-Zinkfinger) und je nach Anzahl der Zinkfinger entsprechend lange DNA-Sequenzen erkennen (Abb. 1.4 B). Es wurden unterschiedliche Methoden entwickelt, um solche hoch-sequenzspezifischen Multi-Zinkfinger zu designen, wie z.B. *modular assembly*, *context-sensitive selection* und die „2 + 2“-Methode von Sangamo BioSciences (Cathomen & Joung 2008). Weitere Methoden zur Generierung von Multi-Zinkfingern bieten OPEN (*oligomerized pool engineering*), eine effektive und öffentlich zugängliche Plattform (Maeder *et al.* 2008) und CoDA (*context-dependent assembly*) (Sander *et al.* 2011). Die Generierung von Muti-Zinkfingern ist jedoch nicht trivial. Ein besonderes Problem stellt der sogenannte *crossstalk* dar, d.h. die Interaktion der einzelnen Zinkfinger untereinander, welche sich negativ auf die DNA-Bindung und daher auch auf die Spezifität auswirken kann. Muti-Zinkfinger müssen daher kontextabhängig selektiert und für die gewünschte Erkennungssequenz optimiert werden. Nichtsdestotrotz, wurden und werden Nuklease-gekoppelte Multi-Zinkfinger auf dem Gebiet des *gene targeting* weitverbreitet eingesetzt. Als eine der ersten artifiziellen hochspezifischen Nukleasen (Kim *et al.* 1996) wurde diese in den vergangenen Jahren in zahlreichen Studien untersucht, wobei das Potential von ZFN mehrfach bestätigt wurde u.a. in Pflanzen- (Cai *et al.* 2009) und Tiermodellen, wie *Drosophila* (Bibikova *et al.* 2002), Schmetterlingen (Merlin *et al.* 2013), Zebrafischen (Doyon *et al.* 2008), Mammalia (Remy *et al.* 2010). Aber auch bei der Modifikation des humanen Genoms sind ZFN Vorreiter. Ein erfolgreicher Ansatz war die Deletion des Gens *CCR5* in $CD4^+$ T-Zellen (Perez *et al.* 2008). Das Gen kodiert für einen Membranrezeptor, der für das HI-Virus einen der wichtigsten Eintrittspunkte in die Zelle darstellt. Durch die Gen-Deletion ist es gelungen, die Zellen weitestgehend resistent gegen HIV zu machen. Andere Ansätze adressieren mutierte Gene, die zu Krankheitsbildern führen, um diese zu korrigieren, darunter die monogene Erbkrankheit SCID (*X-linked severe combined immune deficiency*) und die Sichelzell-Krankheit (Urnov *et al.* 2005, Zou *et al.* 2011).

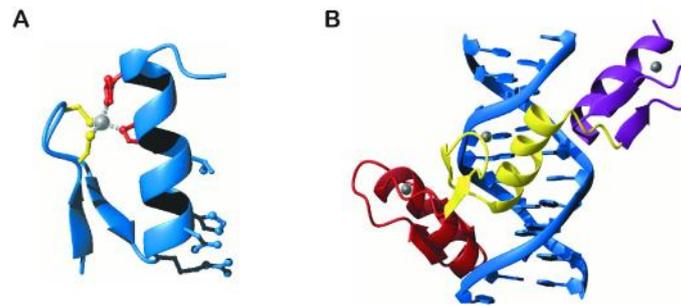


Abbildung 1.4: Zinkfinger Kristallstruktur und spezifische Interaktion eines Multi-Zinkfingers mit der DNA. (A) Cys2His2-Zinkfinger-Motiv, bestehend aus zwei β -Faltblättern und einer α -Helix ($\beta\beta\alpha$ -Struktur). Zwei Cystein- und zwei Histidin-Reste (gelb bzw. rot) koordinieren ein Zink-Ion (graue Kugel). Die für die DNA-Interaktion relevanten Aminosäure-Reste sind in der α -Helix hervorgehoben. (B) DNA-Interaktion eines Multi-Zinkfingers, bestehend aus drei Zinkfingern (rot, gelb, violett). Die DNA ist in hellblau dargestellt. Verändert nach Pabo & Grant 2001. RCSB Protein Data Bank: 1AAY.

1.2.2.2 DNA-Bindungsmodul: TALE

TALE (*transcription activator-like effectors*) bilden eine Gruppe innerhalb der Typ-III-Effektor-Proteine, die durch *Xanthomonas* Bakterien in infizierten Pflanzenzellen ausgeschieden werden. Die Gattung *Xanthomonas* ist divers und besteht aus gram-negativen Pflanzenpathogenen, die insgesamt über 200 verschiedene Pflanzen-Familien infizieren können. Von entscheidender Bedeutung für die Pathogenität ist das Hrp-TypIII-Sekretions-(T3S) System. Es bildet eine räumliche Überbrückung von der bakteriellen Membran zur pflanzlichen Plasmamembran (Buttner & Bonas 2002, Weber *et al.* 2005), um ein Gemisch aus 30-40 TypIII-Effektor-Proteine in den Wirtsorganismus einzuschleusen (Buttner & Bonas 2010). TypIII-Effektor-Proteine werden aufgrund von Sequenz- und Struktur-Homologien, sowie ihrer Funktionalität phylogenetisch unterteilt (White *et al.* 2009). TALE bilden die größte Protein-Gruppierung und binden im Wirtsorganismus an spezielle Promotorregionen, wodurch die Transkription gewisser Gene hoch- oder herunterreguliert wird (Kay & Bonas 2009). Insgesamt beeinflussen die Effektor-Proteine unterschiedliche zelluläre Abläufe im Wirtsorganismus, die die Abwehr der Pflanze schwächen und die Virulenz der Bakterien begünstigt. Typische äußere Merkmale der infizierten Pflanze sind Nassfäule, Chlorophyllmangel und nekrotisches Gewebe. Für die Funktion unumgängliche Merkmale der TALE-Proteine sind das *nuclear localization signal* (NLS) und die *acidic activation domain* (AD), die sich im C-terminalen Bereich des Proteins befinden (Boch & Bonas 2010). Das NLS dient dem Transfer der Proteine vom Zytoplasma in den Zellkern und die AD der Aktivierung der Genregulation. Entscheidend für die DNA-Erkennung ist die sogenannte *repeat*-Domäne,

die sich im mittleren Teil des Proteins befindet. Sie besteht aus sich wiederholenden DNA Sequenzen (*repeats*), die bis auf wenige Ausnahmen hochkonserviert sind. In der Natur kommen TALE-Proteine vor, die zwischen 1,5 bis 33,5 *repeats* aufweisen, wobei 15,5 bis 19,5 üblich sind (Boch & Bonas 2010). Die Anzahl der Aminosäuren (AS) eines *repeats* kann ebenfalls variieren (30-42 AS), am häufigsten sind jedoch 34 AS. Die auffälligste Abweichung innerhalb der *repeats* findet man an der AS-Position 12 und 13, den sogenannten *repeat variable diresidues* (RVD). Studien zeigen, dass die DNA-Spezifität durch die entsprechenden AS an diesen Positionen definiert werden (Boch *et al.* 2009, Moscou & Bogdanove 2009, Morbitzer *et al.* 2010). So erkennen z.B. die Aminosäure-Paare NI ein Adenin, HD ein Cytosin, NG ein Thymin, und NK ein Guanin (Abb. 1.5 B). Die Co-Kristallstrukturen von TALE-Proteinen zeigen, dass die Interaktion mit der DNA stattfindet, indem das Protein sich um die DNA windet und sich dabei in der großen Furche positioniert (Deng *et al.* 2012, Mak *et al.* 2012) (Abb. 1.5 A). Die *repeats* weisen eine beinahe identische Konformation auf. Die AS Bereiche jenseits der RVD bilden jeweils eine α -Helix, die durch den sogenannten RVD-loop (Aminosäure-Position 12, 13 und auch 14) verknüpft sind. Dabei findet lediglich eine Interaktion der RVD mit den Basen des direkten Strangs (5' \rightarrow 3') statt. Diese Interaktion wird durch die AS an der Position 13 vermittelt. Die Struktur weist darauf hin, dass die AS an der Position 12 die Funktion hat, den RVD entsprechend zu positionieren und somit nur indirekt an der DNA-Erkennung beteiligt ist. Die Erkenntnis, dass benachbarte RVD die Sequenzspezifität nicht gegenseitig beeinflussen (Moscou & Bogdanove 2009) ermöglicht es artifizielle TALE-Proteine mit gewünschten Sequenzspezifitäten relativ leicht zu konstruieren (Boch *et al.* 2009, Morbitzer *et al.* 2010). Diverse Methoden wurden vorgestellt, optimiert und weiterentwickelt und ermöglichen es inzwischen programmierte TALE-Proteine effizient und mit geringem Aufwand herzustellen (Morbitzer *et al.* 2011, Briggs *et al.* 2012, Reyon *et al.* 2012, Schmid-Burgk *et al.* 2012). TALE-Proteine sind seit der Entschlüsselung des Spezifitätsmusters vermehrt für die *gene targeting*-Methode verwendet worden. Es wurden zahlreiche Studien in Tiermodellen u. a. in *Drosophila* (Liu *et al.* 2012), Moskitos (Aryan *et al.* 2013), Seidenraupen (Ma *et al.* 2012) und Zebrafische (Zu *et al.* 2013), wie auch in humanen Zelllinien durchgeführt (Cermak *et al.* 2011, Hockemeyer *et al.* 2011, Miller *et al.* 2011).

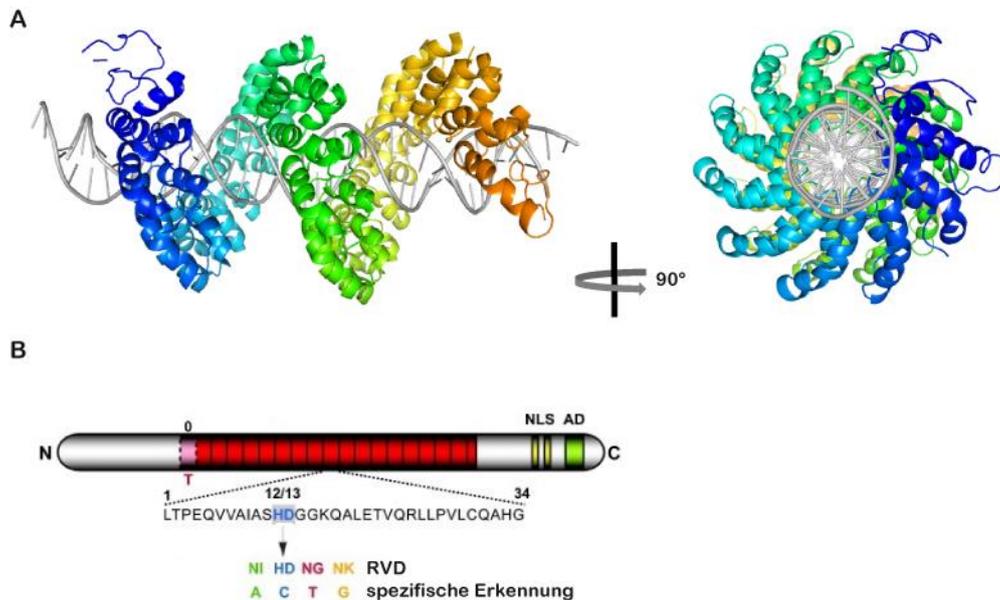


Abbildung 1.5: Spezifische TALE-DNA-Interaktion und schematische Darstellung der Proteinstruktur. (A) Co-Kristallstruktur des TALE-Proteins PthXo1. Die Teilstruktur umfasst einen Großteil der *repeat*-Domäne, die sich in der großen Furche der DNA angelagert und die DNA dabei umwickelt. RCSB Protein Data Bank: 3UGM. (B) TALE-Proteine weisen stets ein *nuclear localization signal* (NLS) sowie eine *acidic activation domain* (AD) am C-terminalen Ende auf. Die *repeat*-Domäne setzt sich aus repetitiven Aminosäure-Abfolgen zusammen, die bis auf die Positionen 12 und 13 (RVD, *repeat variable diresidues*) nur gering variieren. Abhängig vom RVD-Typ variiert die Basen-spezifische Erkennung. Verändert nach Scholze & Boch 2010.

1.2.2.3 DNA-Spaltungsmodule: FokI und Alternativen

In „klassischen“ artifiziellen hochspezifischen Nukleasen ist das DNA-Bindungsmodul üblicherweise an die Spaltungs-Domäne der Restriktionendonuklease FokI gekoppelt (ZFN, TALEN). Die FokI-Spaltungs-Domäne hat die charakteristischen Eigenschaften, dass sie sequenzunspezifisch ist und erst katalytisch aktiv wird, wenn zwei monomere Untereinheiten auf der DNA dimerisieren. Hierfür muss die Erkennungssequenz immer tripartit sein, also zwei Erkennungssequenzen des DNA-Bindungsmoduls, die einen gewissen DNA-Abschnitt flankieren, auf dem die DNA-Spaltungsmodule dimerisieren können. Für die „klassischen“ ZFN konnte neben der gezielten Modifizierung von DNA-Sequenzen, eine relativ hohe Zytotoxizität durch unspezifische Spaltung unadressierter Genom-Bereiche gezeigt werden (Perez *et al.* 2008, Pruett-Miller *et al.* 2009). Cathomen und Joung (2008) beschreiben mögliche Ursachen der unspezifischen DNA-Spaltung. Zum einen kann unspezifische DNA-Spaltung auftreten, wenn die Sequenzspezifität der Multi-Zinkfinger unzureichend ist (Abb. 1.6 A). Andererseits kann unspezifische Spaltung auftreten, wenn lediglich ein Multi-Zinkfinger an seine spezifische Sequenz gebunden hat, diese Sequenz aber nicht innerhalb der Zielsequenz

liegt (*hanging monkey*-Situation) (Abb. 1.6 B). Dies kann theoretisch im menschlichen Genom auftreten, da die Erkennungssequenz eines Multi-Zinkfingers mit drei Zinkfingern nur neun Basenpaare umfasst. Häufig müssen zwei unterschiedliche ZFN-Monomere eingesetzt werden, die in der Erkennungssequenz des Multi-Zinkfingers variieren. Um die Zielsequenz spezifisch erkennen und spalten zu können, muss ein Heterodimer gebildet werden. Das bietet jedoch die Gefahr einer potentiellen unspezifischen Spaltung, indem zwei gleiche Monomere dimerisieren (Abb. 1.6 C). Um dies zu vermeiden, wurden Monomere der FokI-Spaltungs-Domäne entwickelt, die sich in der Kontaktstelle zueinander unterscheiden (Miller *et al.* 2007, Szczypek *et al.* 2007). Mit dem Schlüssel-Schloss-Prinzip können nur komplementäre FokI-Monomere miteinander dimerisieren (Abb. 1.6 D). Auf diese Weise kann die unspezifische Spaltung durch Homodimere reduziert werden. Äquivalente Ansätze wurden auch für „klassische“ TALEN beschrieben (Wu *et al.* 2014). Ein anderer Ansatz, um die Zytotoxizität, die durch die sequenzunspezifische Natur der FokI-Spaltungs-Domäne verursacht wird, zu reduzieren, ist die Verwendung alternativer DNA-Spaltungsmodule. Dies konnte erfolgreich am Beispiel der sequenzspezifischen Restriktionendonuklease PvuII, sowohl in der Kombination mit einem Multi-Zinkfing (Schierling *et al.* 2012), also auch mit einem TALE-Protein (Yanik *et al.* 2013) gezeigt werden. Im direkten Vergleich zur der Fusion mit einem FokI-DNA-Spaltungsmodul wiesen die alternativen Konstrukte signifikant geringere Zytotoxizität auf.

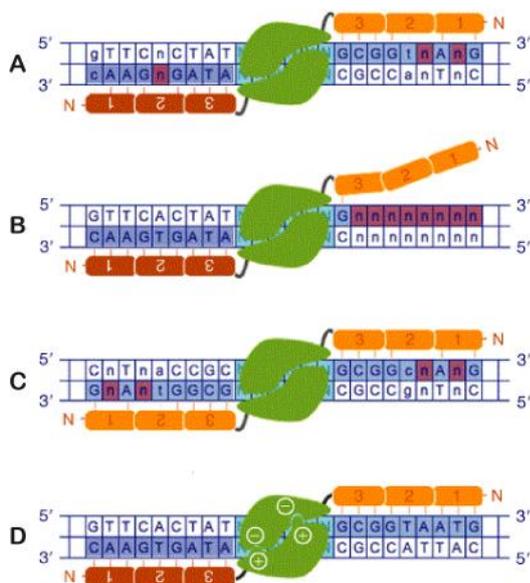


Abbildung 1.6: Mögliche Ursachen für unspezifische DNA-Spaltung durch „klassische“ ZFN und TALEN. (A) Unzureichende Sequenzspezifität der DNA-Bindungsmodule (dunkelrot bzw. orange). (B) Dimerisierung der DNA-Spaltungsmodule, obwohl lediglich ein DNA-Bindungsmodul (rot) spezifisch gebunden ist. (C) Homodimer-Bildung: zwei gleiche DNA-Bindungsmodule (orange). (D) Modifizierte Kontaktstelle zwischen FokI-Monomeren vermeidet Homodimer-Bildung. Unspezifische DNA-Erkennung sind pink hervorgehoben. Verändert nach Cathomen & Young 2008.

1.3 Zytotoxizität und hochspezifische Nickasen

Die in Abschnitt 1.2.2.3 genannten Gründe haben toxische Effekte in der Zelle, da mangels Spezifität außerhalb der Zielsequenz gespalten wird. Die klassischen hochspezifischen Nukleasen sind konstruiert, Doppelstrangbrüche zu katalysieren, um mit Hilfe einer angebotenen DNA-Reparatur-Matrix die homologe Rekombination zu stimulieren (Pingoud & Wende 2011, Sakuma & Woltjen 2014). Unspezifische Doppelstrangbrüche sind zwangsläufig toxisch für die Zelle, wenn keine homologe DNA-Reparatur-Matrix vorhanden ist. Wenn diese DNA-Reparatur-Matrix fehlt oder nicht verwendet wird, greift ein alternativer Reparatur-Mechanismus, das sogenannte *non-homologous end-joining* (NHEJ) (Lieber 2010). Hierbei wird der Doppelstrangbruch repariert, indem einzelne Nukleotide an den freien DNA-Enden deletiert oder eingefügt werden. Dies führt häufig zur Zerstörung von Genen, da das Leseraster verschoben wird. Studien konnten jedoch zeigen, dass homologe Rekombination nicht zwangsläufig durch Doppelstrangbrüche, sondern auch durch Einzelstrangbrüche (*nicks*) initiiert werden kann. Das verringert die Zytotoxizität signifikant, da *nicks* nicht durch NHEJ prozessiert werden (Davis & Maizels 2011, Metzger *et al.* 2011). Die Reduktion der Zytotoxizität durch *nick*-induzierte HR wurde auf alle bestehenden hochspezifischen Nukleasen übertragen. Im Fall von Homing Endonukleasen (McConnell Smith *et al.* 2009) und den RNA-vermittelten Nukleasen (Cong *et al.* 2013) werden Nickasen erzeugt indem eins von zwei katalytischen Zentren durch Mutationen inaktiviert wird. FokI-gekoppelte Zinkfinger und TALE-Proteine können zu Nickasen verändert werden, indem ein Heterodimer verwendet wird, bei dem ein FokI-DNA-Spaltungsmodul inaktiviert ist (Kim *et al.* 2012, Ramirez *et al.* 2012, Wang *et al.* 2012, Wu *et al.* 2014). Alle Studien zeigen, dass die Rate für NHEJ dramatisch sinkt, was die Anwendung sicherer macht. Allerdings muss man auch bedenken, dass die Rate für HR, die durch *nicks* induziert wird, geringer ist als für DSB. Relevant ist die Art der Anwendung. Bei Gen-Deletion z.B. ist NHEJ ein erwünschter Effekt, bei Gen-Insertion und -Korrektur wiederum ist NHEJ möglichst zu vermeiden.

1.4 Fusions-Konstrukte

Basierend auf den Erfahrungen vorhergehender Entwicklungen hochspezifischer Nukleasen, sollen innovative Konstrukte entstehen, die die verschiedenen Merkmale kombinieren, um Programmierbarkeit, Spezifität und Konstrukt-Größe zu verbessern Die Grundbausteine dieser Fusions-Konstrukte werden in den folgenden Kapiteln näher vorgestellt.

1.4.1 DNA-Bindungsmodul: I-SceI

I-SceI gehört zu den Homing Endonukleasen (HE) und wird der Familie LADLIDADG zugeordnet (Kapitel 1.2.1.1). Grundsätzlich können HE dieser Familie entweder als Dimer (ein LAGLIDADG-Motiv pro Untereinheit) oder als Monomer (zwei LAGLIDADG-Motive im Protein) aktiv sein (Colleaux *et al.* 1988, Dalgaard *et al.* 1997). Entscheidend für die katalytische Aktivität ist der Aspartat-Rest am Carboxyl-Terminus des Sequenz-Motivs LAGLIDADG (Gimble & Stephens 1995). Dimere HE haben immer palindrome bzw. pseudopalindrome Erkennungssequenzen. Monomere HE hingegen weisen stets eine asymmetrische Erkennungssequenz auf, wie auch I-SceI. I-SceI wird im mitochondrialen Genom des einzelligen Pilzes *Saccharomyces cerevisiae* kodiert (Colleaux *et al.* 1986), hat ein Molekulargewicht von 26 kDa und erkennt eine 18 Basenpaare lange Sequenz. Der Protein-DNA-Komplex zeigt die für LADLIDADG HE typische Sattelstruktur (Moure *et al.* 2003) (Abb. 1.7). Die katalytischen Zentren sind jeweils in der kleinen Furche der DNA positioniert, um einen Einzelstrangbruch im oberen und im unteren Strang zu katalysieren (Moure *et al.* 2008), woraus ein Doppelstrangbruch mit vier Basen 3'-Überhängen resultiert (Colleaux *et al.* 1988). Kinetische Studien weisen auf eine geringe Strangpräferenz hin, wobei der untere Strang mit einer signifikant schnelleren Rate gespalten wird (Perrin *et al.* 1993). Die Co-Kristallstruktur legt nahe, dass für die Hydrolyse vier divalente Metall-Ionen (vorzugsweise Mg^{2+}) notwendig sind (Moure *et al.* 2008). HE können neben dem Einsatz als eigenständige hochspezifische Nukleasen (Kapitel 1.2.1.1) auch der Untersuchung von neuen DNA-Spaltungsmodulen für artifizielle hochspezifische Nukleasen dienen. Dabei werden sie lediglich als DNA-Bindungsmodul in Fusionskonstrukten eingesetzt werden (Fonfara *et al.* 2012). Dafür muss die HE jedoch katalytisch inaktiviert werden, was bei LAGLIDADG HE durch die Substitution der katalytisch relevanten Aspartat-Reste relativ einfach bewerkstelligt werden kann (Gimble & Stephens 1995). Auch in diesem Fall wurde die katalytisch inaktive I-SceI-Variante D44N D145A verwendet, um die Eigenschaften eines neuen DNA-Spaltungsmoduls zu untersuchen (*proof of principle*).



Abbildung 1.7: I-SceI Co-Kristallstruktur. Sequenzspezifische Interaktion mit der 18 bp langen Erkennungssequenz. Die spezifische Sequenz wird in 5'→3' Richtung gebunden, wobei I-SceI vom C- zum N-Terminus orientiert ist. RCSB Protein Data Bank: 1R7M.

1.4.2 DNA-Bindungsmodul: AvrBs3 bzw. AvrBs4

AvrBs3 und AvrBs4 sind TALE-Proteine aus *Xanthomonas campestris* pv. *vesicatoria* und zählen zu einer der größten TALE-Familien (Kapitel 1.2.1.4). Sie weisen eine sehr starke strukturelle Homologie auf, haben jedoch unterschiedliche DNA-Bindesepezifitäten (Bonas *et al.* 1989). Ihre *repeat*-Domäne besteht aus jeweils 17,5 *repeats* und somit wird eine 18 Basen spezifische Erkennungssequenz erkannt (eine Base pro *repeat* inkl. letzten halben *repeat*). Für die Aktivität einiger TALE-Proteine ist es obligatorisch, dass ein Thymin (Boch *et al.* 2009) oder im Fall von *Ralstonia solanacearum* TALE ein Guanin (de Lange *et al.* 2013) vor der eigentlichen Erkennungssequenz vorgestellt ist (Position 0 „null“) (Abb. 1.5). Im Fall von AvrBs3 und AvrBs4 muss sich ein Thymin an dieser Position befinden, somit beträgt die gesamte Erkennungssequenz 19 Basen. Welche genaue Rolle die Position 0 spielt, ist noch nicht vollständig verstanden. Eine neuere Studie zeigt, dass die Funktion mancher TALEN davon abhängen kann, aber nicht zwangsläufig muss. Viel entscheidender dabei ist die Komposition verschiedener Faktoren, wie die RVD an der Position 1, die *repeat*-Anzahl und die Gesamtaktivität des TALE-Proteins (Schreiber & Bonas 2014). Als Beispiel für ein programmierbares DNA-Bindungsmodul wird ein AvrBs3- und entsprechend ein AvrBs4-Protein verwendet, die am N-terminalen Ende um 152 Aminosäuren (AS) und am C-terminalen Ende um 250 AS verkürzt sind (es verbleiben 28 AS nach dem letzten halben *repeat*). Diese Varianten wurden in einer Studie bereits beschrieben und sind trotz Reduktion, in der Lage ihre spezifische Erkennungssequenz zu binden (Miller *et al.* 2011).

1.4.3 DNA-Spaltungsmodul: MutH

MutH ist eine monomere Endonuklease und gehört zum DNA-*mismatch repair* (MMR) System in einigen gram-negativen Bakterien, z.B. *E. coli*. Dieses System dient der Reparatur von

Basenfehlpaarungen sowie Insertionen und Deletionen (*Indels*) von einem bis zu vier Nukleotid(en). Ein bis drei *Indel(s)* werden vergleichbar effizient repariert wie eine G-T Fehlpaarung (am besten durch das MMR-System in *E. coli* korrigiert), die Reparatur-Effizienz von vier *Indels* sinkt hingegen deutlich ab und fünf *Indels* werden nicht mehr erkannt (Learn & Grafstrom 1989, Smith & Modrich 1996). Basenfehlpaarungen und *Indels* entstehen häufig bei natürlichen zellulären Prozessen, wie DNA-Replikation oder homologe Rekombination. Das MMR System umfasst viele Komponenten, die Hauptinitiatoren sind jedoch MutS, MutL und MutH. Sobald eine Basenfehlpaarung auftritt, wird dies durch MutS detektiert (Abb. 1.8 A1). MutS bindet an die Basenfehlpaarung und aktiviert in einer ATP-abhängigen Reaktion MutL (Abb. 1.8 A2), welches wiederum ATP-abhängig die Endonuklease MutH zu einer nahegelegenen (bis zu 1000 Basenpaare) Erkennungssequenz (GATC) rekrutiert (Modrich 1991) (Abb. 1.8. A3). MutH ist nicht nur ein sequenzspezifisches sondern auch eine methylierungabhängiges Enzym (Parker & Marinus 1992, Friedhoff *et al.* 2003). Es ist daher in der Lage zu unterscheiden, ob die spezifische Erkennungssequenz methyliert ist oder nicht. Für die Reparatur des fehlerhaften Tochterstrangs ist diese Eigenschaft von entscheidender Bedeutung, denn im frühen Stadium der Replikation weist der neusynthetisierter Tochterstrand nicht die charakteristische Methylierung durch die DNA Adenin-Methyltransferase (DAM) auf (GATC) (Barras & Marinus 1989). Somit hydrolysiert MutH nach der Rekrutierung an die Erkennungssequenz lediglich den neusynthetisierten Strand durch einen DNA-Einzelstrangbruch (*nick*) vor dem ersten Guanin am 5' Ende (\downarrow GATC). In den folgenden Reaktionsschritten kommen weitere Komponenten zum Einsatz, was zum Abbau des Tochterstrangs, ausgehend vom *nick* bis über die Basenfehlpaarung hinaus, und der Resynthese eines neuen korrekten Abschnittes des Tochterstrangs führt und somit die Basenfehlpaarung repariert wird (Modrich 1991). Vollständig methylierte DNA ist resistent gegenüber der Bindung und Spaltung durch MutH (Welsh *et al.* 1987, Lee *et al.* 2005). Friedhoff *et al.* (2003) zeigten das der Aminosäure-Rest Y212 von essenzieller Bedeutung für die Diskriminierung zwischen dem modifizierten und unmodifizierten Adenin der MutH-Erkennungssequenz ist. Die Hydroxylgruppe (-OH) des Tyrosins bildet Wasserstoffbrücken mit dem unmodifizierten Adenin und eine zusätzliche Modifizierung führt aller Voraussicht nach zu sterischen Hindernissen. Hemimethylierte und vollständig unmethylierte DNA wird als Substrat angenommen, wobei eine Präferenz für Ersteres besteht. Hemimethylierte DNA wird durch MutH fester gebunden als vollständig unmethylierte DNA, was sich in einem niedrigerem K_D -Wert widerspiegelt (hemimethyliert: 0,9 μ M, unmethyliert: 3,5 μ M) und einer zehnmal

schnelleren Spaltung der hemimethylierten DNA (Lee *et al.* 2005). Hemimethylierte DNA wird lediglich im unmethylierten DNA-Strang gespalten. Unmethylierte Erkennungssequenzen werden ohne Strangpräferenz gespalten. Hierbei akkumuliert jedoch zunächst das genickte Produkt bevor es in einer weiteren langsameren Reaktion unter verlängerter Inkubation und Enzymüberschuss in einen Doppelstrangbruch konvertiert wird (Welsh *et al.* 1987, Au *et al.* 1992). Die Aktivität von MutH wird außerdem im Zuge einer Basenfehlpaarung durch MutS und MutL von 0,0017 *nicks*/min/Monomer auf 0,04-0,12 *nicks*/min/Monomer 20- bis 70-fach gesteigert (Welsh *et al.* 1987). Zusätzlich konnte gezeigt werden, dass die Aktivität von MutH abhängig vom DNA-Kontext der GATC-Sequenz ist (Lee *et al.* 2005). MutH hat ein Molekulargewicht von 28 kDa und kann grob in zwei zueinander bewegliche Subdomänen eingeteilt werden, die eine Art Klammer bilden (Ban & Yang 1998, Lee *et al.* 2005). Der N-terminale Teil der Klammer beherbergt das katalytische Zentrum, der C-terminale Teil ist hauptsächlich verantwortlich für die spezifische DNA-Bindung (Abb. 1.8 B). Das katalytische Zentrum befindet sich am inneren Ende der Klammer. Die Co-Kristallstruktur zeigt, dass die DNA von der Klammer umschlossen wird und dabei mit der großen Furche der DNA interagiert (Lee *et al.* 2005). Das katalytische Sequenz-Motiv von MutH (D(X)₆₋₃₀(E/D)XK) findet man häufig auch bei TypII-Restriktionsendonukleasen (REasen) und das aktive Zentrum zeigt große Ähnlichkeiten in der Primär-, Sekundär- und Tertiärstruktur, was auf eine divergente Entwicklung mit den Typ-II-REasen PvuII, BamHI, EcoRV, EcoRI, Sau3AI und FokI hinweist (Ban & Yang 1998). Die strukturelle Ähnlichkeit ist vor allem in der N-terminalen Subdomäne von MutH deutlich, wohingegen die C-terminale Subdomäne, die maßgeblich für die DNA-Bindung ist, große Abweichungen aufweist. Das katalytische Zentrum wird durch eine Triade der Aminosäure-Reste D70, E77 und K79 gebildet, wobei das Aspartat in allen oben genannten REasen konserviert ist und das zweite Carboxyl häufiger ein Glutamat als ein Aspartat ist. Der Lysin-Rest ist maßgeblich an der Katalyse beteiligt, indem es die Metall-Ionen-Koordination mit sequenzspezifischer Bindung und Spaltung koppelt (Lee *et al.* 2005). Für die katalytische Aktivität wird Mg²⁺ als Co-Faktor benötigt und die Reaktion verläuft nach dem Zwei-Metall-Ionen-Mechanismus. In Rahmen dieser Arbeit wurde MutH als DNA-Spaltungsmodul eingesetzt werden. Das maßgebliche Kriterium für die Auswahl dieser Endonuklease ist, dass es unter physiologischen Bedingungen selbstständig keine katalytische Aktivität aufweist, sondern die aktive Rekrutierung zu der spezifischen Erkennungssequenz benötigt. Daher ist die Idee die Rekrutierung mittels MutS und MutL durch die Kopplung an ein geeignetes DNA-Bindungsmodul zu ersetzen.

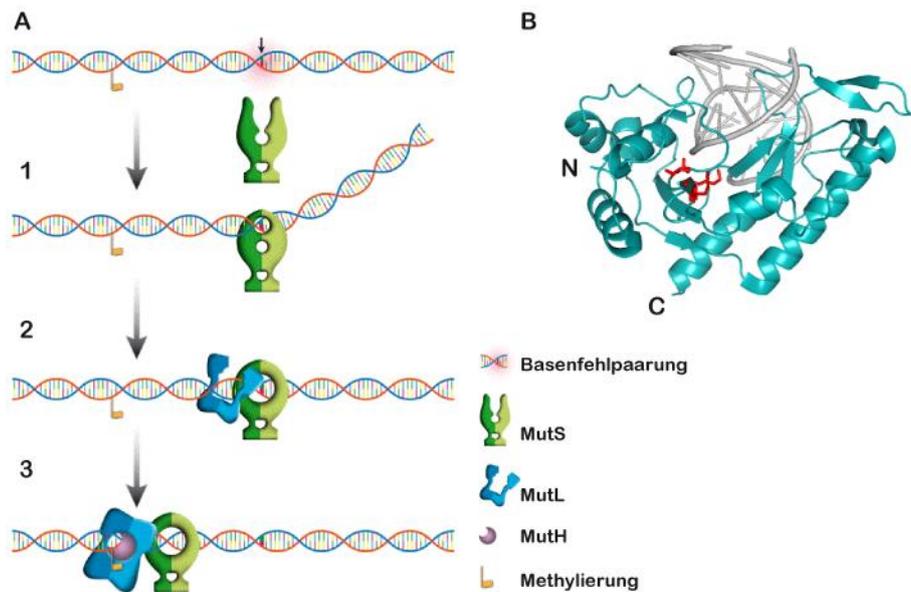


Abbildung 1.8: Schematische Darstellung des DNA-mismatch repair Systems und der MutH Co-Kristallstruktur. (A) Die fehlerhaft eingebaute Base (Pfeil) im neusynthetisierten Tochterstrang (blau) wird im 1. Schritt durch MutS (grün) detektiert. Im 2. Schritt erfolgt die Rekrutierung von MutL (blau), was wiederum die Hydrolyse einer nahegelegenen GATC-Sequenz im Tochterstrang durch MutH (violett) vermittelt (Schritt 3). Verändert nach Internetquelle mm2m.eu. (B) Interaktion von MutH mit der spezifischen Erkennungssequenz GATC. Die katalytisch relevanten Aminosäure-Reste D70, E77 und K79 sind rot hervorgehoben. RCSB Protein Data Bank: 2AOQ.

1.4.4 DNA-Spaltungsmodul: HinP1I

HinP1I stammt aus dem Bakterium *Haemophilus influenzae P1* (Shen et al. 1980) und gehört zu der weit verbreiteten Enzym-Familie der TypII-REasen. Die Datenbank REBASE umfasst mittlerweile über 3600 TypII-REasen mit mehr als 250 verschiedenen Erkennungssequenzen (Roberts et al. 2010). REasen sind Bestandteil eines Restriktions-Modifikations-Systems, welches Bakterien und Archaeen Schutz vor zellfremder DNA bietet. Die zelleigene DNA ist durch Methylierung der Erkennungssequenzen der zelleigenen REasen geschützt, wohingegen zellfremde DNA diese Modifikation nicht aufweist und daher gespalten und somit unschädlich gemacht wird. Dabei sind diese Enzyme sowohl effizient ($k_{cat}/k_{uncat}=10^{16}$) als auch spezifisch (Horton et al. 2006). TypII-REasen zeichnen sich dadurch aus, dass sie innerhalb oder in unmittelbarer Nähe ihrer Erkennungssequenz spalten, welches ein Palindrom (P) ist (Pingoud et al. 2005). Ein weiteres charakteristisches Merkmal ist, dass die enzymatische Spaltung ATP bzw. GTP unabhängig ist, jedoch divalente Metall-Ionen (idealerweise Mg^{2+}) als Co-Faktor benötigt (Bowen & Dupureur 2003). Im Gegensatz zu den meisten anderen TypII-REasen, die in ihrer aktiven Form als Homodimere oder –tetramer

vorliegen, ist HinP1I ein funktionales Monomer, das allerdings bei hohen Enzymkonzentration auch als Dimer oder Oligomer auftreten kann (Yang *et al.* 2005). Interessanterweise dimerisiert HinP1I nicht, indem die DNA-Interaktionsstellen samt katalytische Zentrum eine Klammer bilden, die die DNA umschließt - dies ist für homodimere REasen üblich - sondern die monomeren Untereinheiten mit der abgewandten Seite interagieren und somit zwei außenliegende DNA-Interaktionsstellen, die voneinander wegzeigen, vorliegen (Horton *et al.* 2006). Die Funktion des HinP1I-Homodimers ist noch nicht ganz geklärt. Aufgrund der abgewandten katalytischen Zentren ist das Homodimer nicht in der Lage einen simultanen Doppelstrangbruch einzuführen, sondern interagiert mit unabhängigen Bereichen in der DNA. *In vitro* Experimente zeigen, dass *supercoiled*-Plasmid-DNA in der *open circular*-Form akkumuliert, bevor sie vollständig linearisiert wird (Horton *et al.* 2006). Dies ist ein starkes Indiz dafür, dass die beiden Stränge nicht simultan sondern zeitlich nacheinander gespalten werden. Hierbei ist jedoch noch unklar, ob ein HinP1I-Monomer einen Einzelstrangbruch hydrolysiert und ein anderes HinP1I-Monomer einen Einzelstrangbruch auf der gegenüberliegenden Seite einführt, woraus ein Doppelstrangbruch resultiert oder wohlmöglich ein einziges HinP1I-Monomer beide Einzelstrangbrüche einführt, indem es von einem Strang zum nächsten wechselt, wie es auch für die monomere REase BcnI bekannt ist (Sasnauskas *et al.* 2011). HinP1I hat ein Molekulargewicht von 29 kDa und hydrolysiert die Sequenz G↓CGC, wobei nach dem ersten Guanin am 5' Ende gespalten wird und dabei zwei Basen 5'-Überhänge entstehen (Yang *et al.* 2005). Das katalytische Zentrum setzt sich aus den Aminosäure-Resten E18, D62 und K83 zusammen. Viele REasen weisen das katalytische Motiv PDX_n(D/E)XK auf (Pingoud *et al.* 2005). Das Motiv von HinP1I weicht davon leicht ab, zeigt aber Ähnlichkeiten zu der monomeren REase MspI: (S/T)DX17-18(Q/N)XK. Auch die strukturellen Ähnlichkeiten zu MspI sind auffällig. Da dies aber auf Sequenz-Ebene nicht der Fall ist, ist die Evolution der beiden Enzyme vermutlich konvergent (Yang *et al.* 2005). HinP1I gehört zur α/β-Protein Klasse. Bei der Interaktion mit der spezifischen Erkennungssequenz bindet die N-terminale α-Helix des Enzyms in der kleinen und das zentrale β-Faltblatt in der großen Furche der DNA (Horton *et al.* 2006). Dabei wird die DNA um ~60° gebogen und die Konformation der DNA dabei stark beeinflusst (*base flipping*), was die katalytische Funktion unterstützt. Bei der Bindung findet auch eine Konformations-Änderung des Enzyms statt, allerdings nicht im Bereich des aktiven Zentrums, welches recht rigide zu sein scheint. Die Co-Kristallstruktur von HinP1I zeigt, dass der katalytische Aminosäure-Rest K83 eine zentrale Rolle bei der Koordination des Metall-Ions sowie bei der sequenzspezifischen Erkennung spielt. Außerdem

zeigt die Struktur deutlich, dass der Mechanismus der Hydrolyse vermutlich mit zwei Metall-Ionen abläuft (Horton *et al.* 2006). Neben MutH wurde auch HinP1I im Zuge dieser Arbeit als DNA-Spaltungsmodul eingesetzt werden. Der Schwerpunkt liegt auch hier auf der monomeren Funktion der hochspezifischen Nuklease.

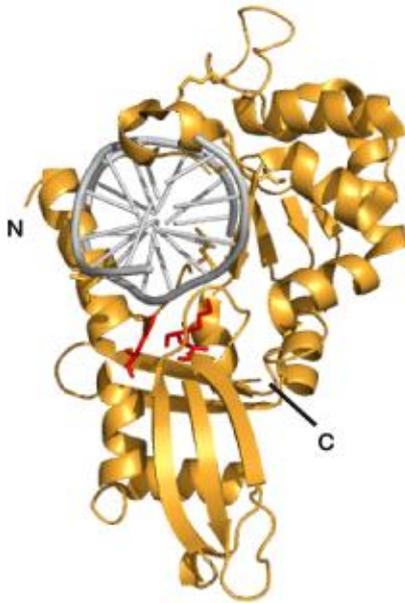


Abbildung 1.9: HinP1I Co-Kristallstruktur. Interaktion mit der spezifischen Erkennungssequenz GCGC. Die katalytisch relevanten Aminosäure-Rest E18, D62 und K83 sind rot hervorgehoben. RCSB Protein Data Bank: 2FL3.

1.5 Zielsetzung

Das Ziel dieser Arbeit war es, hochspezifische Nucleasen zu entwickeln, die eine sichere Anwendung ermöglichen und somit die Grundlage für spätere *in vivo* Studien mit möglichem therapeutischem Einsatz (Gentherapie) zu schaffen. Der besondere Fokus der innovativen Fusionskonstrukte lag auf dem Einsatz der DNA-Spaltungsmodule MutH und HinP1I. Toxische Effekte durch unspezifische Spaltung der Fusionskonstrukte sollten möglichst reduziert werden, da MutH und HinP1I sequenzspezifische Nucleasen sind, was die Spezifität des Fusionskonstrukts insgesamt begünstigt. Das bedeutet, die Erkennungssequenzen der DNA-Spaltungsmodule (MutH = GATC bzw. HinP1I = GCGC) werden durch das jeweilige DNA-Bindungsmodul (I-SceI, TALE) adressiert. Somit sollen lediglich die GATC- bzw. GCGC-Sequenz, die eine benachbarte Erkennungssequenz des DNA-Bindungsmoduls aufweist, gespalten werden, hingegen unadressierte GATC- bzw. GCGC-Sequenzen unprozessiert bleiben. Hierzu sollte vor allem die schwache Nuclease-Aktivität von MutH genutzt werden, welches eine aktive Rekrutierung zu der Erkennungssequenz benötigt, um aktiv zu sein und unadressierte GATC-Sequenz nicht selbstständig hydrolysieren kann. Die generierten hochspezifischen Nucleasen stellen funktionale Monomere dar. Dies verschafft nicht nur den

Vorteil, dass nur eine Ausführung des Konstrukts bei einer Anwendung in die Zelle eingeführt werden muss (im Gegensatz zu heterodimeren Konstrukten), sondern sollte auch für die Generierung hochspezifischer Nickasen genutzt werden. Um die bestmögliche Aktivität und Spezifität der Fusionskonstrukte zu ermitteln, sollten zudem der Peptidlinker zwischen dem DNA-Bindungsmodul und dem DNA-Spaltungsmodul variiert werden (Proteinebene), wie auch der Abstand zwischen der Erkennungssequenz des DNA-Bindungsmoduls und des DNA-Spaltungsmoduls auf DNA-Ebene. Darüber hinaus sollte die Funktion aller Fusionskonstrukte mittels kinetischen Analysen, wie Katalyse und Bindungsaffinität validiert werden.

2 Material und Methoden

2.1 Materialien

2.1.1 Chemikalien

Sämtliche verwendeten Chemikalien sind unter Angabe des Herstellers in Tabelle 2.1 aufgeführt. Alle Chemikalien haben den Reinheitsgrad *pro analysi*. Für alle Lösungen wurde Reinstwasser verwendet, das über ein Milli-Q Synthesis A10 Reinstwassersystem (Millipore) gewonnen wurde.

Tabelle 2.1: Verwendete Chemikalien mit Angabe des Herstellers.

Chemikalie	Hersteller
2-Mercaptoethanol	<i>Merck</i>
40%-Acrylamid/Bisacrylamid 29:1 und 19:1	<i>AppliChem</i>
Agar	<i>AppliChem</i>
Agarose	<i>Invitrogen</i>
Arabinose	<i>Sigma Aldrich</i>
Aluminiumsulfat	<i>AppliChem</i>
Ammoniumperoxodisulfat (APS)	<i>Merck</i>
Ammoniumsulfat	<i>AppliChem</i>
Ampicillin-Natriumsalz	<i>AppliChem</i>
Borsäure	<i>AppliChem</i>
Bromphenolblau	<i>Merck</i>
Coomassie Brilliant Blue G 250	<i>Serva</i>
dNTPs	<i>Fermentas</i>
Desthiobiotin	<i>Sigma-Aldrich</i>
Dimethylsulfoxid (DMSO)	<i>Merck</i>
Dithiothreitol (DTT)	<i>AppliChem</i>
Ethylendiamintetraacetat (EDTA)	<i>AppliChem</i>
Essigsäure	<i>Roth</i>
Ethanol	<i>Applichem</i>
Ethidiumbromid	<i>Roth</i>
Formamid	<i>Merck</i>
Glycerin	<i>AppliChem</i>

Glycin	<i>Applichem</i>
Hefeextrakt	<i>AppliChem</i>
HEPES	<i>AppliChem</i>
Imidazol	<i>AppliChem</i>
Isopropyl- β -D-thiogalactopyranosid (IPTG)	<i>AppliChem</i>
Isopropanol	<i>Roth</i>
Kaliumacetat	<i>Applichem</i>
Kaliumchlorid	<i>Roth</i>
Kaliumdihydrogenphosphat	<i>Merck</i>
Kaliumglutamat	<i>Sigma</i>
di-Kaliumhydrogenphosphat	<i>Merck</i>
Kaliumhydroxid	<i>Merck</i>
Lubrol	<i>Sigma</i>
Magnesiumchlorid	<i>Merck</i>
Magnesiumsulfat Heptahydrat	<i>Merck</i>
Mangan(II)-chlorid-terahydrat	<i>Applichem</i>
Natriumacetat	<i>AppliChem</i>
Natriumchlorid	<i>AppliChem</i>
Natriumhydroxid	<i>Merck</i>
Ni-NTA-Agarose	<i>Macherey-Nagel</i>
Phenylmethylsulfonylfluorid (PMSF)	<i>AppliChem</i>
Phosphorsäure	<i>Roth</i>
Saccharose	<i>Merck</i>
Salzsäure	<i>Merck</i>
<i>sodium dodecyl sulfate (SDS)</i>	<i>RotH</i>
Strep-Tactin Sepharose	<i>IBA</i>
Tetramethylethyldiamin (TEMED)	<i>AppliChem</i>
Tetracyclin-Hydrochlorid	<i>AppliChem</i>
Tris-(hydroxymethyl)-aminomethan	<i>AppliChem</i>
Triton X100	<i>AppliChem</i>
Trypton	<i>AppliChem</i>
Tween 20	<i>Merck</i>
Urea	<i>AppliChem</i>

2.1.2 Puffer und Medien

Sämtliche verwendeten Puffer und Medien sind unter Angabe der Zusammensetzung in Tabelle 2.2 aufgeführt. Für alle Puffer und Medien wurde Reinstwasser verwendet, das über ein Milli-Q Synthesis A10 Reinstwassersystem (Millipore) gewonnen wurde.

Tabelle 2.2: Verwendete Puffer und Medien mit jeweiliger Zusammensetzung.

Puffer/Medium	Zusammensetzung
10x Aktivitätspuffer MutH-I-SceI	1 M Kaliumglutamat 250 mM Tris-Acetat 10 mM MgCl ₂ 5 mM 2-Mercaptoethanol 500 µg/ml BSA pH 7,6
10x Aktivitätspuffer TALE-MutH	200 mM Tris-Acetat 1,2 M Kalium-Acetat 10 mM MgCl ₂ 50 µg/ml BSA pH 7,5
10x Aktivitätspuffer TALE-HinP1I	100 mM Tris/HCL 1,5 M KCl 10 mM DTT 10 mM MgCl ₂ 1 mg/ml BSA pH 7,9
5x DNA-Auftragspuffer	25 % (w/v) Saccharose 1,2 % (w/v) SDS 250 mM EDTA 0,1 % (w/v) Bromphenolblau pH 8,0 (eingestellt mit NaOH)

2x FE-Probenpuffer	80% (v/v) Formamid 100 mM EDTA
50 x TAE	2 M Tris 1 M Natriumacetat 50 mM EDTA pH 8,0 (eingestellt mit Essigsäure)
10 x TPE	0,9 M Tris 20 mM EDTA pH 8,2 (eingestellt mit Phosphorsäure)
10x TBE	1 M Tris 1M Borsäure 25 mM EDTA pH 8,3
10 x Laemmli-SDS-Elektrophorese-Puffer	250 mM Tris 1,9 M Glycin 1 % (w/v) SDS pH 8,3
6 x Laemmli-Auftragspuffer	420 mM Tris/HCl pH 6,8 0,12 % (w/v) SDS 7,2 % (v/v) 2-Mercaptoethanol 36 % (v/v) Glycerol 0,06 % (w/v) Bromphenolblau pH 6,8
His Wasch-/Aufschlusspuffer (MutH Fusionen)	20 mM Tris/HCl 1 M NaCl 20 mM Imidazol 1mM PMSF pH 8,0

His Elutionspuffer (MutH Fusionen)	20 mM Tris/HCl 1 M NaCl 200 mM Imidazol pH 8,0
Strep Waschpuffer (MutH Fusionen)	100 mM Tris/HCl 1 M NaCl pH 8,0
Strep Elutionspuffer (MutH Fusionen)	100 mM Tris/HCl 500 mM NaCl 2,5 mM Desthiobiotin pH 8,0
Dialysepuffer (HinP1I Fusionen)	30 mM Kaliumphosphat 1 M NaCl 20 mM Imidazol 1mM PMSF 0,001% (v/v) Lubrol pH 7,4
His Wasch-/Aufschlusspuffer (HinP1I Fusionen)	30 mM Kaliumphosphat 1 M NaCl 20 mM Imidazol pH 7,4
His Elutionspuffer (HinP1I Fusionen)	30 mM Kaliumphosphat 1 M NaCl 200 mM Imidazol pH 7,4
Strep Waschpuffer (HinP1I Fusionen)	30 mM Kaliumphosphat 600 mM NaCl pH 7,4

Strep Elutionspuffer (HinP1I Fusionen)	30 mM Kaliumphosphat
	600 mM NaCl
	2,5 mM Desthiobiotin
	pH 7,4
Dialysepuffer (HinP1I Fusionen)	30 mM Kaliumphosphat
	600 mM NaCl
	0,5 mM EDTA
	0,5 mM DTT
	0,01% (v/v) Lubrol
	50% (v/v) Glycerol
	pH 7,4
Luria Bertani-Medium (LB-Medium)	1 % (w/v) Trypton
	0,5 % (w/v) Hefeextrakt
	0,5 % (w/v) NaCl
	pH 7,5

2.1.3 Kits

Im Rahmen dieser Arbeit wurden folgende Kits zur Reinigung von DNA verwendet:

- Wizard® SV Gel and PCR Clean-up System (Promega)
- PureYield™ Plasmid Miniprep System (Promega)
- PureYield™ Plasmid Midiprep System (Promega)

Die Durchführung erfolgte nach Protokoll des Herstellers.

2.1.4 Säulenmatrix

Zur Reinigung der rekombinanten Proteine durch Affinitätschromatographie über den Strep- bzw. His-tag wurde Strep-Tactin® Sepharose® (IBA) bzw. Protino® Ni-NTA Agarose (Macherey-Nagel) als Säulenmatrix verwendet.

2.1.5 Dialyseschläuche

Für die Dialyse wurde der Dialyseschlauch Visking Typ 8/32 (Roth) verwendet. Ausschlussgröße 14 kDa. Der Dialyseschlauch wurde in Reinstwasser aufgekocht und in 20%

Ethanol bei 4°C gelagert. Vor dem Gebrauch wurde der Schlauch gründlich mit Reinstwasser gespült.

2.1.6 *Primer* und Oligonukleotide

Sämtliche verwendeten *primer* und Oligonukleotide sind mit Sequenzen in Tabelle 2.3 und Tabelle 2.4 (S. 33/34) aufgeführt. Bei allen Oligonukleotiden handelt es sich um Oligodesoxynukleotide.

Tabelle 2.3: Verwendete *primer* mit Sequenz. Die Sequenzen sind in 5' - 3' Richtung dargestellt.

<i>primer/Oligonukleotid</i>	Sequenz
TALE+3'SacI	GATAACGGAGCTCCCCTCAAGTGCTGGCCGTCCTCT
MutH_fwd	AGATATACCATGGGCAGCAGC
MutH+5'SacI_fwd	TTAGAGCTCAATGTCCCAACCTCGCCCACTG
MutH+5'KpnI_fwd	TTAGGTACCATGTCCCAACCTCGCCCACTG
MutH+3'NheI_rev	AGCATTGCTAGCCTGGATCAGAAAATGACGGGCCAGTAGTG
MutH+3'His+BlpI_rev	AGCATTGCTAGCCTGGATCAGAAAATGACGGGCCAGTAGTG
HinP1I_fwd	GAGAAGTCGACGAATCTGGTAGAATTAGGATC
HinP1I+5'SacI_fwd	TAGGGAGCTCCAATCTGGTAGAATTAGGATC
HinP1I_S8A_fwd	GTAGAATTAGGAGCTAAAACCGCTGAAGAT
HinP1I_K12E_fwd	CTAAAACAGCTGAAGATGGTTTTAAAACG
HinP1I_S86A_fwd	GTAAAGCTCGTCGCGAATAAACGTGGTTTTAATC
HinP1I_H97A_fwd	CAGATAGATAAAGCCTGGCTTGC
HinP1I_R134D_rev	AAATTCTGTCATGAACATTCTGTCCTTATCTTTTGTATTTG
HinP1I_S208A_rev	CAAAGTTAATAGAGCCTCTTGGGGCAAGAGAGATATCG
HinP1I_L235D_rev	GGATCAATTTTGAATTGATCCATATTTGCGG
pQE_fwd	GTATCACGAGGCCCTTTCGTCT
pQE_rev	CATTACTGGATCTATCAACAGGAG
StPr_fwd	AATAGGCGTATCACGAGGCCCTTTC
StPr_rev	ACCCAGAGCGCTGCCGGCAC

2.1.7 Enzyme und Marker

Alle verwendeten Restriktionsenzyme werden unter Kapitel 2.2.2.1 aufgeführt und wurden von den Herstellern New England Biolabs oder Fermentas bezogen. Die Verwendung erfolgte nach Herstellerangaben. Der Marker PageRuler™ Unstained Protein Ladder (Fermentas) wurde als Größenstandard bei der gelelektrophoretischen Analyse der Fusionsproteine verwendet.

2.2 Methoden

Mikrobiologische, molekularbiologische sowie biochemische Standardmethoden werden nicht näher erläutert. Sie wurden nach Angabe des Herstellers durchgeführt oder können in *Current Protocols in Molecular Biology* und *Protein Science* (Ausubel et al. 2012, Coligan et al. 2012) nachgeschlagen werden.

2.2.1 Mikrobiologische Methoden

Alle mikrobiologischen Arbeiten wurden an einer Sterilbank unter Verwendung autoklavierter Gefäße, Medien und sonstiger Materialien durchgeführt.

2.2.1.1 Plasmide

Im Rahmen dieser Arbeit wurden Plasmide zur Proteinexpression und als Substrate verwendet. Alle verwendeten Plasmidkonstrukte wurden durch Sequenzierung entsprechender Bereiche auf ihre Richtigkeit hin überprüft.

Expressionsplasmide

pASK_His-MutH-I-SceI-Strep

Das Fusionskonstrukt MutH-I-SceI wurde generiert, indem eine katalytisch inaktive Variante von I-SceI (Fonfara *et al.* 2012), welche einen verkürzten C-terminus aufweist (Δ C9), an das C-terminale Ende einer cysteinfreien Variante von MutH (C96S) fusioniert wurde. I-SceI und MutH wurden entweder über einen 10 Aminosäure langen Peptidlinker (ASENLYFQGG) verbunden, der über eine TEV-Protease-Erkennungssequenz verfügt (unterstrichen), oder mit einem Linker ohne TEV-Erkennungssequenz (TKQLVKSE), der zur Kontrolle diente. Das Gen für MutH (C96S), welches die kodierende Sequenz für einen N-terminalen His-tag aufweist, und das Gen für I-SceI (Δ C9 D44N D145A) wurden mit der kodierenden Sequenz des jeweiligen Peptidlinkers verbunden und in das Expressionsplasmid pASK-IBA63b-plus (IBA)

kloniert, welches seinerseits für einen C-terminalen Strep-*tag* kodiert. Das erwartete Protein sollte somit wie folgt zusammengesetzt sein: His-MutH-I-SceI-Strep. Das Gen ist hinter einen Tetracyclin-sensitiven Promotor geschaltet, was eine gezielte Expression des Fusionsproteins durch die Induktion mit Anhydrotetracyclin ermöglicht. Für die selektive Kultivierung besitzt das Expressionsplasmid zudem ein Ampicillin-Resistenzgen und ist 6650 Basenpaare groß.

pQE30_Strep-AvrBs4-MutH-His und Strep-AvrBs3-HinP1I-His

Für die Generierung der Fusionskonstrukte TALE-MutH und TALE-HinP1I wurde die verkürzte Variante von AvrBs4 bzw. AvrBs3 (TALE) an das N-terminale Ende des MutH bzw. HinP1I Wildtyps (wt) fusioniert. Der kodierende Bereich besteht somit aus dem Gen für AvrBs4 bzw. AvrBs3, ohne die kodierende Sequenz für die ersten 152 Aminosäuren am N-Terminus und die letzten 250 Aminosäuren am C-Terminus (28 Aminosäuren nach dem letzten halben *repeat*) und dem Gen für MutH bzw. HinP1I, welches die kodierende Sequenz für einen His-*tag* am C-Terminus trägt. Die zwei hintereinander geschalteten Gene wurden in das Expressionsplasmid pQE30 (Qiagen) kloniert, das seinerseits für einen N-terminalen Strep-*tag* kodiert. Die erwarteten rekombinanten Proteine sollte somit wie folgt zusammengesetzt sein: Strep-AvrBs4-MutH-His bzw. Strep-AvrBs3-HinP1I-His. Das Gen ist hinter einen Lac-Promotor geschaltet. Dies ermöglicht eine gezielte Expression der Fusionsproteine durch die Induktion mit IPTG (Isopropyl- β -D-thiogalactopyranosid). Für die selektive Kultivierung besitzen die Expressionsplasmide zudem ein Ampicillin-Resistenzgen und sind 6423 (TALE-MutH) bzw. 6489 (TALE-HinP1I) Basenpaare groß. Für die verschiedenen Linker-Varianten von TALE-MutH wurde der 16 Aminosäure (AS) lange Linker L16 verwendet (SSVIPNRGVTKQLVKG). Einerseits wurde das Fusionskonstrukt ohne Linker kloniert (L0), indem das Gen von MutH unmittelbar nach dem letzten halben *repeat* des Gens von AvrBs4 folgt (TALE-L0-MutH). Andererseits ersetzt L16 die 28 AS von AvrBs4 nach dem letzten halben *repeat* (TALE-L16-MutH) oder wird hinter die 28 AS geschaltet (TALE(28)-L16-MutH). Alle hergestellten Varianten des Fusionskonstrukts TALE-HinP1I wurden ausgehend von der kodierenden Sequenz des HinP1I Wildtyps durch gezielte Mutagenese hergestellt.

Substratplasmide

Zur Analyse der Spezifität und Aktivität der Fusionskonstrukte wurde das Substratplasmid pAT153 verwendet, welches modifiziert wurde, indem eine DNA-Kassette mit der

entsprechenden Erkennungssequenz eingebracht wurde (Tabelle 2.4, S. 33/34). Es wurden zwei Arten von Substraten generiert: (i) ein adressiertes Substrat, welches die Erkennungssequenz des DNA-Bindungsmoduls trägt [I-SceI (S) oder AvrBs4 bzw. AvrBs3 (T)] und die Erkennungssequenz des DNA-Spaltungsmotifs MutH [H] oder HinP1I [P]. Die Abstände zwischen den Erkennungssequenzen variieren in der Anzahl der Basenpaare (bp). Für das Fusionskonstrukt MutH-I-SceI wurde lediglich ein adressiertes Substrat mit 3 bp Abstand untersucht (S-3-H). Für TALE-MutH und TALE-HinP1I wurden diverse Abstände getestet (T-x-H; x = 1, 2, 3, 4, 5, 6, 7, 8, 9 bzw. T-x-P; x = 0, 2, 4, 6, 8, 10, 12). (ii) Für die Fusionskonstrukte MutH-I-SceI und TALE-MutH wurden zwei Kontroll-Substrate generiert, die entweder die Erkennungssequenz für das DNA-Bindungsmodul [I-SceI (S) oder AvrBs4 (T)], aber keine benachbarte MutH-Erkennungssequenz aufweisen oder lediglich MutH-Erkennungssequenzen, aber keine Erkennungssequenz für das DNA-Bindungsmodul (H). Neben den eingebrachten DNA-Kassetten befinden sich zusätzlich 18 (Substraten für MutH-I-SceI) bzw. 19 (Substraten für TALE-MutH) nicht adressierte MutH-Erkennungssequenzen auf dem jeweiligen Plasmid. Die adressierten Substrate für TALE-HinP1I verfügen über 22 zusätzliche GATC-Sequenzen. Als unadressiertes Substrat wurde in diesem Fall das unmodifizierte pAT153 Plasmid verwendet. Es verfügt über 26 HinP1I-Erkennungssequenzen. Für die selektive Kultivierung besitzen alle Substratplasmide ein Ampicillin-Resistenzgen.

2.2.1.2 *E. coli* Stämme

Zur Proteinexpression wurde der *E. coli* Stamm XL1 Blue (Stratagene) verwendet. Für die selektive Kultivierung besitzt der Stamm ein Tetracyclin-Resistenzgen. Im Fall des Fusionskonstrukts TALE-HinP1I wt wurde zusätzlich ein Plasmid eingebracht, das das Gen für die HinP1I-Methyltransferase trägt und über eine Chloramphenicol-Resistenz verfügt.

Zur Amplifizierung der Substratplasmide für die MutH-Fusionskonstrukte wurde der *E. coli* Stamm JM110 (Stratagene) verwendet. Dieser Stamm ist defizient für die Methylasen Dam und Dcm, somit weist die DNA keine Methylierung auf. Für die selektive Kultivierung trägt der Stamm ein Streptomycin-Resistenzgen. Zur Amplifizierung der Substratplasmide für die HinP1I-Fusionen wurde der oben beschriebene *E. coli* Stamm XL1 Blue (Stratagene) genutzt.

Beide Stämme exprimieren keine der bekannten K12 Restriktionsenzyme, was gewährleistet, dass eingebrachte Fremd-DNA nicht degradiert wird. Zusätzlich sind beide Stämme

rekombinationsdefizient, sodass eingebrachte DNA nicht durch Rekombination modifiziert werden kann.

2.2.1.3 *E. coli* Viabilitätstest

Der Viabilitätstest diente der Untersuchung möglicher toxischer Effekte durch unspezifische Spaltung in der Zelle. Unspezifische Doppelstrangbrüche führen zwangsläufig zur genomischen Instabilität und somit auch zu einer verminderten Viabilität der Zelle. Der Viabilitätstest wurde in Form von diskontinuierlichen und kontinuierlichen Wachstumskurven mit folgenden Protein-Varianten durchgeführt:

- 1) TALE-MutH: zu untersuchen
- 2) TALE: Positivkontrolle
- 3) MutH: Positivkontrolle
- 4) ZFN: Negativkontrolle
- 5) TALE-HinP1I EAD: Negativkontrolle
- 6) pQE30: Expressionsplasmid leer

Alle Kontrollen bis auf MutH (3) liegen wie das zu untersuchende Fusionskonstrukt TALE-MutH (1) im Expressionsplasmid pQE30 (6) kloniert vor und sind entsprechend hinter einen Lac-Promotor geschaltet. Das Gen für MutH wt befindet sich in dem Expressionsplasmid pMQ402 (Derivat von pBAD18) und unterliegt der Kontrolle eines araBAD Promoters, der durch Arabinose induziert werden kann (Loh *et al.* 2001). Bei den Positivkontrollen TALE (2) und MutH (3) handelt es sich um das eigenständige DNA-Bindungsmodul (AvrB4) bzw. DNA-Spaltungsmodul (MutH wt) gemäß dem Fusionskonstrukt TALE-MutH. Theoretisch sollten diese keine katalytische Aktivität in der Zelle aufweisen und entsprechend keine toxischen Effekte haben, da das TALE Protein keine katalytische Aktivität hat und die Erkennungssequenz im *E. coli* Genom nicht auftritt und MutH unter physiologischen Bedingungen keine katalytische Aktivität außerhalb des *mismatch repair* Systems aufweisen sollte. ZFN (4) und TALE-HinP1I (5) wurden als Negativkontrollen eingesetzt, da sie bei vorhergehenden *in vitro* Untersuchungen Unspezifität aufweisen und somit toxische Effekte haben könnten (Schierling *et al.* 2012) (Kapitel 3.6.1). Alle Expressionsplasmide besitzen ein Ampicillin-Resistenzgen zur Selektion und wurden in dem *E. coli* Stamm JM110 untersucht (GATC-Sequenzen sind unmethyliert). Der *E. coli* Stamm wurde mit dem entsprechenden Expressionsplasmid transformiert. Die Aufzucht erfolgte in Suspensionsschüttelkulturen o/n

bei 37°C in LB-Medium, das zur Selektion mit 75 µg/ml Ampicillin und 50 µg/ml Streptomycin versetzt wurde. Unter Verwendung der genannten Antibiotika wurden am Folgetag 25 bzw. 50 ml frisches LB-Medium mit entsprechendem Volumen der o/n-Kultur auf OD_{600nm} = ~0,05 (diskontinuierlich) bzw. ~0,1 (kontinuierlich) angeimpft und der OD-Wert nach 0, 1, 2, 3, 4, 5h (diskontinuierlich) bzw. 0, 0,5, 1, 2, 3, 4, 5, 6, 7, 8h (kontinuierlich) gemessen. Bei der Messung der diskontinuierlichen Wachstumskurve wurde nach 1h mit 1 mM IPTG bzw. 0,05% (w/v) Arabinose induziert. Im Fall der kontinuierlichen Wachstumskurve erfolgte die Induktion nach 2h. Um die Bedingungen kontinuierlich zu halten, wurden die Suspensionsschüttelkulturen alle 3h im Verhältnis 1:20 mit frischem LB-Medium und entsprechender Zugabe der Antibiotika und des entsprechenden Induktors verdünnt. Für die OD-Messung der diskontinuierlichen Wachstumskurven gilt zu beachten, dass eine genaue Ermittlung des Wertes nur bis ~0,9 möglich ist (dieser Wert wird im Fall der kontinuierlichen Wachstumskurven durch die genannte Verdünnung nicht überschritten). Daher ist es notwendig, die Probe vor der Messung zu verdünnen (hier 1:5 Verdünnung mit frischem LB-Medium ab 3h-Wert).

2.2.2 Molekularbiologische Methoden

2.2.2.1 Klonierungen

Für die Konstruktion der verschieden Fusionsproteine wurden Ziel-Gene mittels PCR amplifiziert (*insert*) und dabei notwendige Schnittstellen eingefügt, um das Fragment in das gewünschte Expressionsplasmid (Vektor) einzubringen zu können. Nach der entsprechenden Restriktionsspaltung wurden 3 nM Vektor und 5 nm *insert* o/n bei 16°C mit der T4 DNA Ligase nach Herstellerangaben ligiert. Im Fall der Substratplasmide wurden Oligonukleotide so entworfen, dass sie von komplementären Schnittstellen flankiert werden, um in den Ziel-Vektor ligiert werden zu können. Anschließend wurden 55 µl des entsprechenden elektrokomponten *E. coli* Stamms ($1-2 \times 10^6$ Zellen/ml) mit 4 µl des Ligations-Ansatzes transformiert. Die Substitution gezielter Aminosäuren wurde auf Gen-Ebene mittels Mutagenese-PCR eingefügt. Hierfür wurden Oligonukleotide entworfen, die die gewünschte Mutation tragen. Diese dienten bei der PCR als *primer*, um die Mutation in das Ziel-Gen einzufügen. Das amplifizierte Fragment wurde wiederum in der *rolling circle PCR* als *primer* verwendet, wobei das vollständige Expressionsplasmid mit der gewünschten Mutation auf Basis des Ausgangs-Plasmids entsteht. Für eine akkurate Amplifizierung wurde für alle Klonierungen die Pfu-Polymerase aus *Pyrococcus furiosus* (3'-5' Exonukleaseaktivität, T_{opt} = 68°C, ~500 Nukleotide

pro Minute) verwendet. Um das Ausgangs-Plasmid zu entfernen, wurde der Ansatz anschließend mit DpnI verdaut und die DNA durch Ethanol-Fällung konzentriert (1/10 Vol. 5 M Natrium-Acetat pH 5,0 und 3 Vol. reinem Ethanol: Zentrifugation bei 20000 x g, 30', 4°C; Pellet + 3 Vol. 70% Ethanol: Zentrifugation bei 20000 x g, 20', 4°C; Pellet in 10 µl H₂O lösen). Auch hier wurden anschließend 55 µl des entsprechenden elektrokomponenten *E. coli* Stamms (1-2 x 10⁶ Zellen/ml) mit 4 µl des Ligations-Ansatzes transformiert.

Standard PCR-Ansatz (50 µl)

100 ng	DNA-Matrize
400 nM	5'- <i>primer</i>
400 nM	3'- <i>primer</i>
200 µM	dNTPs
1x	Pfu-Aktivitätspuffer
1 µl	Pfu-Polymerase

Standard PCR-Programm

94°C	2 min	} x 30
92°C	45 sek	
T _{opt}	30 sek	
68°C	500 nt/min	
68°C	5 min	
4°C	Pause	

rolling circle PCR -Programm

94°C	2 min	} x20
94°C	45 sek	
60°C	30 sek	
68°C	15 min	
68°C	20 min	
4°C	Pause	

pASK_His-MutH-I-SceI-Strep

Das Fusionskonstrukt MutH-I-SceI wurde auf Basis des Vektor-Plasmids pASK_His-PvuII-I-SceI-Strep (L₊) kloniert (Fonfara *et al.* 2012). Hierfür wurde der kodierende Bereich für His-PvuII durch Restriktionsspaltung mit den Enzymen NcoI und NheI entfernt und ein *insert* des Gens für His-MutH C96S unter Verwendung des Expressionsplasmids pMQ420_MutHC96S sowie der *primer* MutH_fwd und MutH+3'NheI_rev über PCR hergestellt. Der fwd *primer* erfasst eine NcoI Schnittstelle unmittelbar vor der Kodierung des His-tags, und der rev *primer* fügt über einen nicht komplementären Überhang am 5'-Ende eine NheI Schnittstelle hinzu (Abb.2.1).

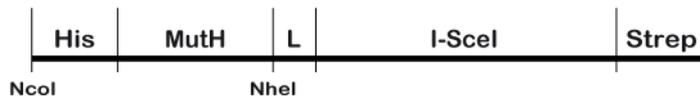
pQE30_Strep-AvrBs4(28)-MutH-His bzw. pQE30_Strep-AvrBs3(28)-HinP1I-His

Das Fusionskonstrukt TALE-MutH basiert auf dem Vektor-Plasmid pQE30_Strep-AvrBs4-PvuII-His. Das Gen für PvuII wurde durch Spaltung mit den Restriktionsenzymen SacI und BlnI entfernt. Das kodierende Gen für MutH wt wurde über die *primer* MutH+5'SacI_fwd und MutH+3'His+BlnI_rev unter Verwendung des Expressionsplasmids pMQ420_MutHwt isoliert. Dabei wird eine SacI und BlnI Schnittstelle sowie die Kodierung für einen C-terminalen His-tag in das *insert* eingeführt. Die gleiche Klonierungs-Strategie wurde auch im Fall von TALE-HinP1I angewendet. Hier dient das analoge Vektor-Plasmid pQE30_Strep-AvrBs3-PvuII-His (Yanik *et al.* 2013) als Basis. Ein *insert* des Gens für HinP1I wurde anhand des Expressionsplasmids pQE30_Strep-HinP1Iwt-His sowie der *primer* HinP1I+5'SacI_fwd und HinP1I_rev über PCR hergestellt (Abb.2.1).

Peptidlinker-Varianten

Zur Generierung der Variante TALE-L0-HinP1I (ohne Linker) wurde der kodierende Bereich von AvrBs4 mittels der *primer* pQE_fwd und TALE+3'SacI_rev anhand des Ausgangs-Plasmids pQE30_Strep-AvrBs4(28)-MutH-His amplifiziert. Dabei wurde eine SacI Schnittstelle nach dem letzten halben *repeat* von AvrBs4 eingefügt. Nach der Restriktionsspaltung des *inserts* und des Ausgangs-Plasmids (dient erneut als Vektor) mit SalI und SacI, entstand nach der Ligation der kodierende Bereich Strep-AvrBs4-MutH-His. Die Linker-Variante TALE(28)-L16-MutH wurde auf Basis des Vektor-Plasmid pQE30_Strep-AvrBs4(28)-L16-PvuII-His (analog zu pQE30_Strep-AvrBs3(28)-L16-PvuII-His; Yanik *et al.* 2013) hergestellt. Der kodierende Bereich für PvuII-His wurde über Spaltung mit KpnI und BlnI eliminiert und durch den kodierenden Bereich für MutH wt-His ersetzt. Das *insert* wurde mit den *primern* MutH+5'KpnI und MutH+3'His+BlnI_rev unter Verwendung des Expressionsplasmids pMQ420_MutHwt isoliert. Dabei entstehen die flankierenden Schnittstellen für KpnI und BlnI. Nach analogem Verdau und Ligation liegt die kodierende Sequenz für pQE30_Strep-AvrBs4(28)-L16-MutH-His vor. Im Fall der Variante TALE-L16-MutH wurde das AvrBs4-Gen durch Restriktion mit SalI und SacI aus dem Expressionsplasmid pQE30_Strep-AvrBs4(28)-L16-MutH-His entfernt. Wiederum wurde dieses Plasmid verwendet, um den kodierenden Bereich für AvrBs4 mittels der *primer* pQE_fwd und TALE+3'SacI_rev (SacI Schnittstelle nach dem letzten halben *repeat* von AvrBs4) zu amplifizieren. Nach entsprechender Restriktion und Ligation entsteht der kodierende Bereich für das Konstrukt Strep-TALE-L2-MutH-His.

MutH-I-SceI



TALE-MutH/HinP1I



Abbildung 2.1: Für die Klonierung relevante Restriktionsschnittstellen der Expressionsplasmide. Dargestellt ist der kodierende Bereich für die Konstrukte MutH-I-SceI und TALE-MutH bzw. TALE-HinP1I. L symbolisiert den in einigen Konstrukten verwendeten Peptidlinker.

Varianten von pQE30_Strep-AvrBs3(28)-HinP1I-His

Die TALE-HinP1I-Variante EAD wurde mittels zielgerichteter Mutagenese des HinP1I wt DNA-Spaltungsmoduls generiert. Die Aminosäure-Substitutionen K12E, S86A bzw. R134D wurden mit den *primer*-Paaren HinP1I_K12E_fwd und pQE_rev, HinP1_S86A_fwd und pQE_rev bzw. HinP1_S86A_fwd und HinP1I_R134D_rev eingeführt. Die zusätzliche Aminosäure-Substitution S8A (HinP1I_S8A_fwd und HinP1I_R134D_rev), H97A (HinP1I_H97A_fwd und HinP1I_R134D_rev), S208A (HinP1_S86A_fwd und HinP1I_S208A_rev) bzw. L235D (HinP1_S86A_fwd und HinP1I_L235D_rev) wurden analog durchgeführt.

Plasmidsubstrate

Die DNA-Kassetten für die Plasmidsubstrate von MutH-I-SceI wurden über die EcoRV Schnittstelle in den Vektor pAT153 eingeführt. Für die DNA-Kassetten der Plasmidsubstrate von TALE-MutH und TALE-HinP1I wurden die Schnittstellen KpnI und SphI genutzt, wobei beide Schnittstellen nach der Ligation eliminiert werden. In Tabelle 2.4 ist jeweils ein Strang der doppelsträngigen DNA-Kassette gezeigt.

Tabelle 2.4: Eingebrachte DNA-Kassetten mit entsprechenden Erkennungssequenzen. Erkennungssequenzen sind unterstrichen und abgekürzt mit S (I-SceI), T (AvrBs4) und H (MutH) bzw. T (AvrBs3) und P (HinP1I). Die Zahlen geben den Abstand zwischen den Erkennungssequenzen in Basenpaaren an. Sequenz in 5' - 3' Richtung dargestellt.

DNA-Kassetten für adressierte Plasmidsubstrate

TATGCGTATAGGGATAACAGGGTAATCCGGATCCTCTGAGTGTGCTGGTAATC	S-3-H
TGCAGTATAATTAATAATCCACTTGGATCCTCGCTCCGAGACATG	T-1-H
TGCAGTATAATTAATAATCCACTTCCGGATCCTCGCTCGAGACATG	T-2-H
TGCAGTATAATTAATAATCCACTTCCGGATCCTCGCTGAGACATG	T-3-H
TGCAGTATAATTAATAATCCACTTACCGGATCCTCGCGAGACATG	T-4-H
TGCAGTATAATTAATAATCCACTTGACCGGATCCTCGCTGACATG	T-5-H
TGCAGTATAATTAATAATCCACTTGAGCCGGATCCTCGCTACATG	T-6-H
TGCAGTATAATTAATAATCCACTTGAGTCCGGATCCTCGCACATG	T-7-H
TGCAGTATAATTAATAATCCACTTGAGTACCGGATCCTCGACATG	T-8-H
TGCAGTATAATTAATAATCCACTTGAGTATCCGGATCCTCACATG	T-9-H
TTTATCTATAAACCTAACCCCTCTGCGCCACATAGCAGGATCCCATG	T-0-P
TTTATCTATAAACCTAACCCCTCTCCGCGCCACATAGCAGGATCCCATG	T-2-P
TTTATCTATAAACCTAACCCCTCTACC GCGCCACATAGCAGGATCCCATG	T-4-P
TTTATCTATAAACCTAACCCCTCTAATACCGCGCCACATAGCAGGATCCCATG	T-6-P
TTTATCTATAAACCTAACCCCTCTATAATACCGCGCCACATAGCAGGATCCCATG	T-8-P
TTTATCTATAAACCTAACCCCTCTGGATAATACCGCGCCACATAGCAGGATCCCATG	T-10-P
TTTATCTATAAACCTAACCCCTCTCGGGATAATACCGCGCCACATAGCAGGATCCCATG	T-12-P

DNA-Kassetten für unadressierte Plasmidsubstrate

TATGCGTATAGGGATAACAGGGTAATCTGTGGCTACTGAGTGTGCTGGTAATC	S
TGCAGTATAATTAATAATCCACTTCCGTGGCCTCGCTGAGACATG	T
TATGCGTAGTCAAGCCTAAGTGTAAACCGGATCCTCTGAGTGTGCTGGTAATC	H

2.2.2.2 Denaturierende Polyacrylamid-Gelelektrophorese

Die denaturierende Polyacrylamid-Gelelektrophorese dient dazu, DNA Moleküle unter denaturierenden Bedingungen, d.h. im einzelsträngigen Zustand zu untersuchen. Diese Methode wird vor allem genutzt, um Einzelstrangbrüche in linearen dsDNA Molekülen detektieren zu können, denn unter nativen Bedingungen sind Einzelstrangbrüche nicht von dem intakten Ausgangsprodukt zu unterscheiden. In dieser Arbeit wurde die denaturierende Polyacrylamid-Gelelektrophorese genutzt, um die Strangpräferenz der Fusionskonstrukte

TALE-MutH und TALE-HinP1I zu analysieren. Der Ansatz für ein denaturierendes Polyacrylamid-Gel sah wie folgt aus:

1 x TBE

10% (v/v) Acrylamid; Acrylamid-Bisacrylamid-Vernetzungsgrad 19:1

7 M Harnstoff

30% (v/v) Formamid

Die Polymerisationsreaktion wurde mit TEMED (0,2 % (v/v)) und APS (0,008 % (w/v)) gestartet. Die zu untersuchenden Reaktionen wurde mit 0,5 Vol. 2 x FE-Probenpuffer versetzt, zur Denaturierung bei 95°C für 3 min erhitzt und anschließend auf Eis gekühlt. Als Laufpuffer diente 1 x TBE-Puffer. Die Elektrophorese erfolgte bei 30 V/cm für 45 min, wobei das Gel ohne Proben einen Vorlauf für mindestens 20 min bei gleichen Bedingungen hatte. Zur Visualisierung der DNA sowie möglicher Spaltprodukte wurden Fluorophor-markierte *primer* verwendet, mit denen die Polynukleotid-Substrate hergestellt wurden. Hierfür wurde ein fwd *primer* mit dem Fluorophor Atto 488 und ein rev *primer* mit dem Fluorophor Att647N am jeweiligen 5'-Ende in einer PCR eingesetzt. Unter Einsatz der adressierten Plasmidsubstrate als *template* entstehen 211 bp (T-x-H) bzw. 216 bp (T-4-P), 218 bp (T-6-P) und 220 bp (T-8-P) lange Polynukleotid-Substrate (T-x-H; x = 1, 2, 3, 4, 5, 6, 7, 8, 9). Die möglichen Spaltprodukte im oberen als auch im unteren Strang (definiert anhand der asymmetrischen Erkennungssequenz von AvrBs4 bzw. AvrBs3) lassen sich sowohl anhand der Größe als auch durch die Markierung des Fluorophores unterscheiden. Zum Beispiel: mit dem adressierten Polynukleotid-Substrats T-3-H können potentiell ein 169 Nukleotid (nt) langes Spaltprodukt entstehen, das ein grünes Fluoreszenz-Signal zeigt, wenn das Substrat im oberen Strang gespalten wird, und ein 38 nt langes Spaltprodukt, das ein rotes Fluoreszenz-Signal zeigt, wenn das Substrat im unteren Strang gespalten wird. Im Fall, dass das Substrat ungespalten vorliegt, werden beide Fluorophor-Signale überlagert und es wird ein gelbes Signal detektiert (Abb.2.1).

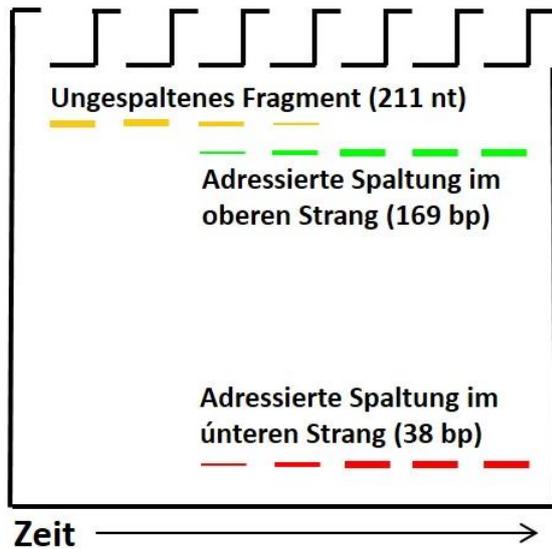


Abbildung 2.2: Schematische Darstellung der denaturierenden Polyacrylamid-Gelelektrophorese zur Analyse der Strangpräferenz des Fusionskonstrukts TALE-MutH anhand des adressierten Fluorophormarkierten PCR Substrats T-3-H. Der obere Strang ist mit dem Fluorophor Atto 488, der untere hingegen mit Atto 647N markiert. Bei Spaltung des oberen (unteren) Strangs entstehen grüne (rote) Fragmente, die 169 (38) nt groß sind. Die ungespaltene DNA erscheint als 211 nt großes gelbes Fragment, durch die Überlagerung der beiden Fluorophore.

2.2.3 Biochemischen Methoden

2.2.3.1 Proteinexpression

Für die Proteingewinnung wurde der *E. coli* Stamm (Kapitel 2.2.1.2) mit dem entsprechenden Expressionsplasmid (Kapitel 2.2.1.1) transformiert. Die Aufzucht von *E. coli* erfolgte in Suspensionsschüttelkulturen bei 37°C in LB-Medium, das zur Selektion mit 75 µg/ml Ampicillin und 10 µg/ml Tetracyclin versetzt wurde. Bei einem OD_{600nm}-Wert von ~0,7 erfolgte die Induktion mit 200 mg/l Anhydrotetracyclin bzw. 1 mM IPTG für die Expression von MutH-I-SceI bzw. TALE-MutH und TALE-HinPII, gefolgt von einer weiteren Inkubation bei 20°C o/n.

2.2.3.2 Tandem-Reinigung der Fusionskonstrukte

Alle in dieser Arbeit untersuchten rekombinanten Fusionsproteine verfügen über einen His- und Strep-tag zur Protein-Isolierung aus dem Zellextrakt. Die Reinigung erfolgte stets über Affinitätschromatographie mit beiden tags im Tandemverfahren (erst via Strep-tag und anschließend über His-tag) für eine möglichst reine Isolierung des Zielproteins.

Affinitätschromatographie über Ni-NTA Agarose

Der His-tag kann spezifisch über Ni-NTA Agarose gebunden werden. Diese Interaktion wird genutzt, um die Proteine aus dem Zellextrakt zu isolieren. Zunächst wurden die Zellen nach der Proteinexpression durch Zentrifugation (3600 x g, 4°C, 15 min) aus der Lösung isoliert und das

Zellpellet in 40 ml His Wasch-/Aufschlusspuffer resuspendiert. Um die Zellen aufzuschließen, wurden diese 5-mal für 40 sek mit Ultraschall behandelt (Duty Cycle 50 %, Output Control 5). Zwischen den Behandlungen wurde für 30 sek pausiert, um eine mögliche Proteindenaturierung durch Wärme zu vermeiden. Um den Zellextrakt von den Zelltrümmern zu trennen, wurde die Probe im nächsten Schritt für 30 min bei 31000 x g und 4°C zentrifugiert. Der Überstand wurde mit 1 ml equilibrierter Ni-NTA Agarose (Equilibrierung in 30 ml His Wasch-/Aufschlusspuffer unter Schütteln bei 4°C für 30 min) 1h lang bei 4°C inkubiert, anschließend wurde die Ni-NTA Agarose 2-mal mit 40 ml His Wasch-/Aufschlusspuffer im *batch*-Verfahren gewaschen und auf eine Säule (Poly-Prep, 0,0 x 4 cm, *BioRad*) übertragen. Die Gewinnung der Proteine erfolgte 3-mal mit 1 ml His Elutionspuffer.

Affinitätschromatographie über Strep-Tactin Sepharose

Für die anschließende Reinigung der Fusionsproteine über den Strep-tag wurde 1 ml Strep-Tactin Sepharose auf eine Säule aufgetragen (Poly-Prep, 0,0 x 4 cm, *BioRad*) und mit Strep Waschpuffer equilibriert (4 ml). Anschließend wurde das gesamte Eluat der vorherigen His-Reinigung auf die Säule aufgetragen und mit 4 ml Strep Waschpuffer gewaschen. Die Elution der Proteine von der Strep-Tactin Sepharose erfolgte mit 5-mal 500 µl Strep Elutionspuffer. Die Eluate wurden abschließend bei 4°C über Nacht gegen Dialysepuffer dialysiert und bei -20°C gelagert. Die Proteinkonzentration wurde unter Verwendung des Extinktionskoeffizienten und der Messung der Absorption bei 280 nm ermittelt (*Pace et al.* 1995). Alle Schritte der Strep- und His-Aufreinigung wurden mittels SDS-PAGE analysiert. Im Rahmen dieser Arbeit wurden ausschließlich 12 %ige Trenngele und 5 %ige Sammelgele verwendet. Das Trenngel besteht aus 12 % (v/v) Acrylamid/Bisacrylamid (29:1), 0,42 M Tris/HCl pH 8,8 und 0,1 % (w/v) SDS. Das Sammelgel besteht aus 5 % (v/v) Acrylamid/Bisacrylamid (29:1), 0,13 M Tris/HCl pH 6,8 und 0,1 % (w/v) SDS. Die Polymerisationsreaktion wurde mit 0,2 % (v/v) TEMED und 0,008 % (w/v) APS gestartet. Vor dem Beladen des Gels wurden die Proteinproben mit 0,2 Vol. 5 x Laemmli-Auftragspuffer vermischt und für 5 min bei 95°C denaturiert. Die Elektrophorese erfolgte bei 35 mA für 1h. Als Laufpuffer wurde 1 x Laemmli-SDS-Elektrophorese-Puffer verwendet. Kolloidales Coomassie diente der Markierung der Proteine. Die Gele wurden über Nacht gefärbt und anschließend mittels des BioDocAnalyze Systems (*Biometra*) digitalisiert.

2.2.3.4 DNA-Spalttests

Im Rahmen dieser Arbeit wurden DNA-Spalttests durchgeführt, um die Aktivität sowie die sequenzspezifische Spaltung der generierten Fusionskonstrukte zu untersuchen. Hierfür wurde die Spaltaktivität auf den adressierten Substraten ermittelt und die Spaltpräferenz gegenüber den nicht adressierten Kontroll-Substraten verglichen. Weiterhin wurden DNA-Spalttests genutzt, um eine eventuelle Strangpräferenz der Fusionskonstrukte zu untersuchen. Die Analyse der Spaltprodukte erfolgte im Fall der Plasmidsubstrate über ein 0,8%iges Agarose-Gel (Lauf bei 10 V/cm für 1,5 h in 1x TAE). Die Gele wurden mit EtBr behandelt und die Fluoreszenz der DNA-Banden mit dem BioDocAnalyze System (Biometra) detektiert. Im Fall der Fluorophor-markierten Substrate wurden die Spaltprodukte durch denaturierende Polyacrylamid-Gelelektrophorese analysiert (Kapitel 2.2.2.2). Die Visualisierung der DNA-Banden erfolgte mittels VersaDoc Imaging System (Bio Rad). Die Spaltraten für MutH-I-SceI und TALE-MutH wurden durch Anpassung der Werte (rel. Anteil gespaltener DNA pro Zeiteinheit) durch die monoexponentielle Funktion $y=A*(1-\exp(-1*k*x))$ ermittelt. Im Fall von TALE-HinP1I wurde die Software KinTek Explorer zur Auswertung genutzt.

Plasmid-Spalttest

Für die Analyse der Plasmid-Spaltaktivität von MutH-I-SceI und TALE-MutH wurden 4 nM Plasmidsubstrat und 16 nM Enzym in 1 x Aktivitätspuffer bei 37 °C inkubiert (entspricht einem vierfachen Enzymüberschuss gegenüber dem Substrat). Unter Berücksichtigung der Enzymverdünnung beträgt die Ionenstärke des Reaktionsansatzes ca. 160 mM, was physiologischen Bedingungen entspricht. Die Plasmidspaltung wurde nach definierten Zeitintervallen (1, 3, 10, 30, 60, 120 min) durch Agarose-Gelelektrophorese analysiert. Die Untersuchung der tripartiten Substrate erfolgte bei 4 nM Plasmidsubstrat und 32 nM Enzym (entspricht einem achtfachen Enzymüberschuss gegenüber dem Substrat). Für die Steigerung der Aktivität von TALE-MutH wurde statt 1 mM MgCl₂ 0,5 mM MnCl₂ (und 10% DMSO) eingesetzt. Im Fall der TALE-HinP1I-Varianten wurden 10 nM Plasmidsubstrat und 80 nM Enzym eingesetzt (entspricht einem 8-fachen Enzymüberschuss gegenüber dem Substrat) mit Ausnahme des TALE-HinP1I wt. Hierbei wurden 0,1 nM verwendet (entspricht einem 100-fachen Substrat-Überschuss gegenüber dem Enzym). Die Spaltung der adressierten Plasmidsubstrate wurde unter oben genannten Bedingungen durchgeführt. Die Analyse der

Spaltung des unadressierten Plasmidsubstrats erfolgte nach den Zeitintervallen (0,5, 1, 2, 3, 5, 24 h).

Bestimmung der Strangpräferenz von TALE-MutH und TALE-HinP1I mit Fluorophor-markierten Polynukleotid-Substraten

Für die Analyse einer mögliche Strangpräferenz von TALE-MutH und TALE-HinP1I wurden adressierte PCR-Substrate mittels 5'-Fluorophor-markierten *primern* hergestellt. Um unterscheiden zu können, ob die adressierte GATC bzw. GCGC Erkennungssequenz präferenziell im oberen oder im unteren Strang gespalten wird (definiert durch die asymmetrische Erkennungssequenz von AvrBs4 bzw. AvrBs3), wurden fwd bzw. rev *primer* eingesetzt, die mit dem Fluorophoren Atto 488 bzw. Atto 647N markiert sind. Die Untersuchung wurde für alle adressierten Substrate von TALE-MutH (T-x-H; x=1, 2, 3, 4, 5, 6, 7, 8, 9) und die Basenpaarabstände 4, 6, und 8 für TALE-HinP1I (T-x-P; x=4, 6, 8) durchgeführt. Die Spaltaktivität wurde unter 1 x Aktivitätspuffer, 20 nM des Polynukleotid-Substrats und 120 nM TALE-MutH bzw. 160 nM TALE-HinP1I (entspricht einem sechs- bzw. achtfachem Enzymüberschuss gegenüber dem Substrat) bei 37°C ermittelt. Unter Berücksichtigung der Enzymverdünnung entspricht die Ionenstärke des Reaktionsansatzes ca. 160 mM, was physiologischen Bedingungen entspricht. Die Plasmidspaltung wurde nach definierten Zeit-Intervallen (1, 3, (5), 10, 30, 60, 120, 180 min) durch denaturierende Polyacrylamid- Gelelektrophorese analysiert.

Bestimmung der Strangpräferenz von MutH-I-SceI und TALE-MutH mittels Sequenzierung

Eine weitere Methode, um die Strangpräferenz bis auf die genaue Nukleotid-Position zu analysieren, ist die Sequenzierung des Spaltprodukts. Hierfür wurden die adressierten Plasmidsubstrate mit den eingebrachten DNA-Kassetten S-3-H bzw. T-3-H/T-6-H (Tab. 2.4) verwendet. Nach vollständiger Einzelstrang-Spaltung der Substrate, gemäß den Ergebnissen des Plasmid-Spalttests (nach 2h, bei 37°C), wurde die DNA gelgereinigt und mit den *primern* StPr_fwd (komplementär zum unteren Strang) und StPr_rev (komplementär zum oberen Strang) sequenziert.

2.2.3.5 TEV Proteolyse des Peptidlinkers von MutH-I-SceI

Die proteolytische Spaltung des Peptidlinkers dient zur physikalischen Trennung von DNA-Bindungs- und DNA-Spaltungsmodul. Die Proteolyse erfolgte in 1 x TEV Aktivitätspuffer, 10 *units*/150 ml TEV Protease, 0,44 mM MutH-I-SceI und 4,4 mM eines 53 bp langen Oligonukleotids mit einer GATC-Sequenz für 90 min bei 30°C. Nach der Inkubation wurde ein Aliquot des Reaktionsansatzes für einen Plasmid-Spalttest entnommen, wobei das adressierte Substrat verwendet wurde. Um auszuschließen, dass der Verlust der katalytische Aktivität von MutH-I-SceI durch die Prozedur bzw. eine intrinsische TEV Spaltstelle zustande kommt, wurde zur Kontrolle eine Ansatz mit H₂O statt der TEV Protease bzw. eine Variante ohne TEV-Erkennungssequenz im Peptidlinker (TKQLVKSE) eingesetzt.

3 Ergebnisse

Maßgeblich für die Funktion von Nukleasen für das *gene targeting* ist (i) die Fähigkeit, eine singuläre Sequenz in einem komplexen DNA-Kontext hochspezifisch zu erkennen und (ii) innerhalb oder in der Nähe dieser spezifischen Sequenz einen DNA-Strangbruch (Einzel- oder Doppelstrangbruch) zu katalysieren. Um zu gewährleisten, dass eine Sequenz innerhalb eines komplexen Genoms nur einmal auftritt, muss diese z.B. für das humane Genom ($3,3 \times 10^9$ Basenpaare) mindestens 16 (4^4) Basenpaare lang sein. Statistisch kommt diese Sequenz dann nur alle 4^{16} ($4,3 \times 10^9$) Basenpaare vor. Um diese hohe Spezifität zu erreichen, werden meist Fusionskonstrukte eingesetzt, die aus einem hochspezifischen DNA-Bindungsmodul und einem DNA-Spaltungsmodul zusammengesetzt sind (Kapitel 1.2).

3.1. Konstruktion der hochspezifischen Nukleasen

Bei den Fusionskonstrukten MutH-I-SceI dient die katalytisch inaktive Homing Endonuklease I-SceI D44N D145A als DNA-Bindungsmodul und bei TALE-MutH bzw. TALE-HinP1I die *repeat*-Domäne der TALE-Proteine AvrBs4 bzw. AvrBs3 (Abb. 3.1). I-SceI, sowie AvrBs3 und AvrB4 haben eine 18 Basenpaare (bp) lange Erkennungssequenz (wobei ein Thymin an der Position 0 für diese TALE-Proteine obligatorisch ist und somit die Erkennungssequenz um 1 bp erweitert wird (\rightarrow 19 bp)). Alle drei DNA-Bindungsmodule erfüllen das Kriterium eine singuläre Erkennungssequenz in einem komplexen DNA-Kontext erkennen zu können. Als DNA-Spaltungsmodule wurden die Endonukleasen MutH und die Restriktion Endonuklease HinP1I verwendet. Beide Enzyme sind monomer und haben eine vier Basenpaar lange Erkennungssequenz: MutH (\downarrow GATC), HinP1I (G \downarrow CGC). Zur affinitätschromatographischen Reinigung der rekombinanten Fusionskonstrukte verfügen diese über einen Strep-*tag* und einen His-*tag*. Im Zuge der Optimierung des Designs wurden verschiedene Peptidlinker zwischen dem DNA-Bindungsmodul und dem DNA-Spaltungsmodul verwendet. Für das Fusionskonstrukt MutH-I-SceI wurde ein zehn Aminosäure langer Peptidlinker (ASENLYFQGG) verwendet, der über eine TEV-Protease-Erkennungssequenz verfügt (unterstrichen) und ein acht Aminosäure langer Linker ohne TEV-Erkennungssequenz (TKQLVKSE), der zur Kontrolle diente. Der Einfluss unterschiedlicher Peptidlinker auf die Aktivität und Spezifität sind im Folgenden lediglich für TALE-MutH ermittelt. Für das Fusionskonstrukt TALE-HinP1I wurde die Charakterisierung anhand der besten Linker-Variante TALE(28)-HinP1I durchgeführt.

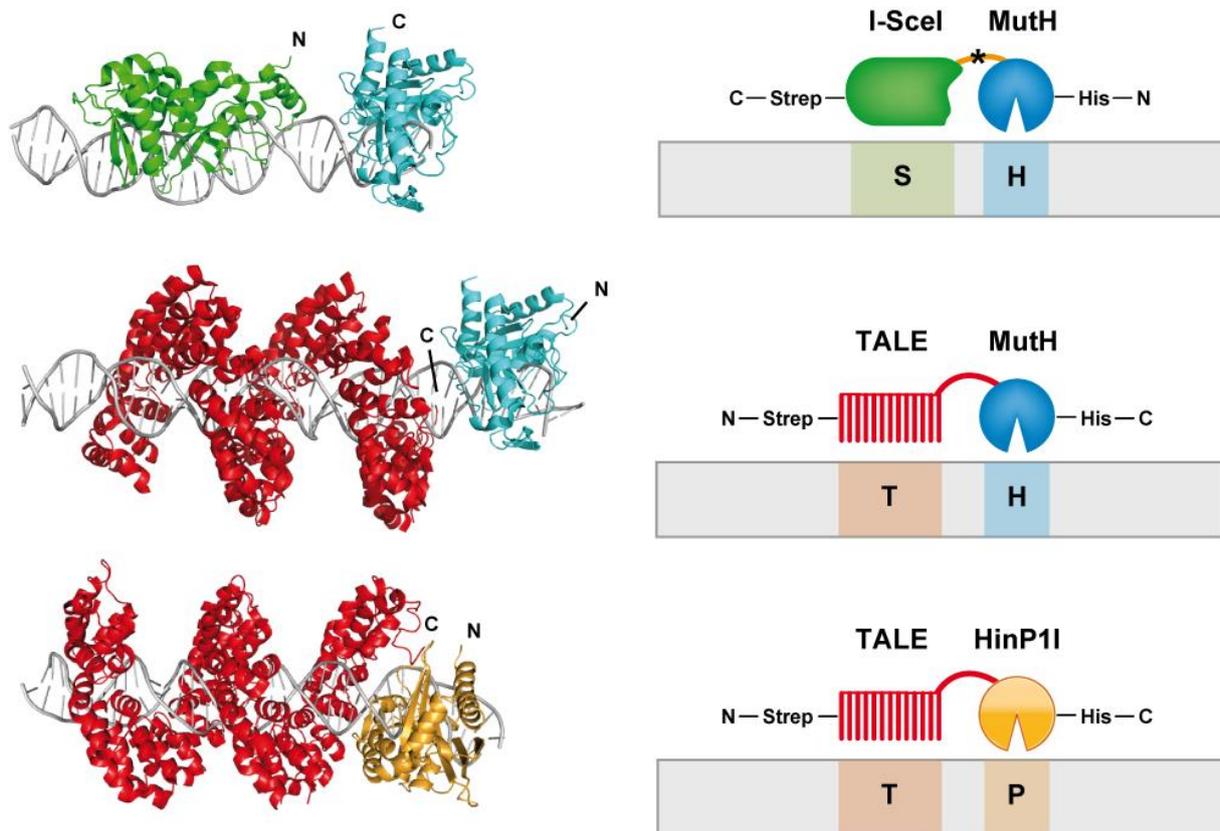


Abbildung 3.1: Modell und schematische Darstellung der Fusionskonstrukte in Interaktion mit der spezifischen Erkennungssequenz. Die spezifische Interaktion wird durch das DNA-Bindungsmodul I-SceI (grün) bzw. der *repeat*-Domäne des TALE-Proteins (rot) vermittelt und rekrutiert dadurch das monomere DNA-Spaltungsmodul MutH (blau) bzw. HinP1I (orange) zu seiner spezifischen Erkennungssequenz GATC (H) bzw. GCGC (P). MutH ist über einen 10 Aminosäure langen Linker (ASENLYFQGG) an I-SceI fusioniert (orange). Der Stern impliziert eine TEV-Protease-Erkennungssequenz. Im Fall von TALE-MutH und TALE-HinP1I wurde kein zusätzlicher Linker verwendet. MutH bzw. HinP1I ist direkt an die letzten 28 Aminosäuren von AvrBs4 bzw. AvrBs3 gekoppelt. Die Modelle wurden anhand der Kristallstrukturen erstellt, RCSB Protein Data Bank: 1R7M (I-SceI), 3UGM (TALE), 2AOQ (MutH), 2FL3 (HinP1I). Im Modell ist die Linker-Region nicht dargestellt. Für das TALE-Protein sind 23 *repeats* von 3UGM gezeigt. Die schematische DNA ist in 5'→3' orientiert.

3.2 Tandem-Reinigung der Fusionskonstrukte

Die Reinigung aller Fusionskonstrukte erfolgte über den terminalen His- und Strep-*tag* durch Affinitätschromatographie. Um das jeweilige Fusionsprotein möglichst ohne andere unspezifische Proteine zu isolieren, wurde es erst über Ni-NTA Agarose und anschließend nochmals über Strep-Tactin Sepharose gereinigt. Die einzelnen Schritte der Reinigung wurden über SDS-PAGE analysiert und werden hier am Beispiel des Fusionskonstrukts TALE-MutH dargestellt (Abb. 3.2). Die analysierten Proben bestehen aus den pelletierten Zelltrümmern (P) und dem Überstand (Ü) nach Sonifizierung und Zentrifugation, sowie dem Durchlauf (DF), den Wasch (W)- und Elutions-Fraktionen (E) nach Bindung an das Säulenmaterial (B, *beads*) und

dem dialysierten Protein nach der Reinigung (D). Die Analyse zeigt, dass die Fusionskonstrukte spezifisch über die Tandem-Reinigung von allen anderen Proteinen getrennt werden können. Nach der ersten Affinitätschromatographie über Ni-NTA Agarose sind neben der gewünschten Proteinbande, deren elektrophoretischen Mobilität bei etwa 107 kDa erwartet wird, weitere Proteinbanden zu erkennen. Diese können allerdings während der anschließenden Affinitätschromatographie über Strep-Tactin Sepharose zum großen Teil abgetrennt werden, und das Dialysat stellt sich in der gelelektrophoretischen Analyse als homogene Bande dar, die der erwarteten Größe des rekombinanten Proteins entspricht.

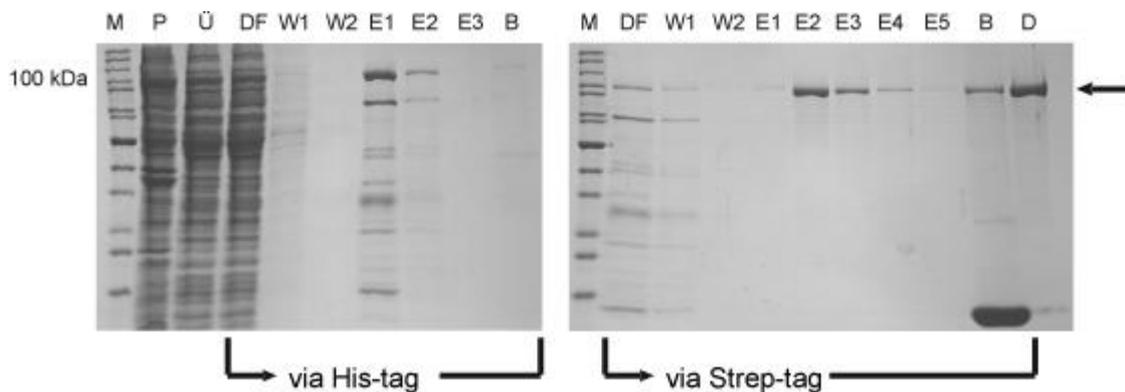


Abbildung 3.2: Tandem-Reinigung von TALE-MutH. Pelletierte Zelltrümmer (P) und Überstand (Ü) nach Sonifizierung und Zentrifugation. Durchfluss (DF), Wasch- (W), Elutions-Fractionen (E) und Dialysat (D) nach Bindung an Strep Tactin Sepharose bzw. Ni-NTA Agarose (B). Größe von TALE-MutH: 107 kDa (Pfeil). Analyse über SDS-PAGE. Marker: Page Ruler Protein Ladder.

3.3 Substrat-Design

Um die Spezifität und Aktivität der konstruierten Nukleasen zu testen, wurden zwei Typen von Substraten generiert (Abb. 3.3). (i) Zum einen handelt es sich um ein adressiertes Substrat, welches in der unmittelbaren Nähe der Erkennungssequenz für das DNA-Bindungsmodul (BM: I-SceI (S), AvrBs3 oder AvrBs4 (T)) die Erkennungssequenz für das DNA-Spaltungsmodul (SM: MutH (H), HinP1I (P)) aufweist (BM-x-SM; x = Abstand in Basenpaaren). Die Variation des Abstands zwischen dem DNA-Bindungs- und Spaltungsmodul dient vor allem der Optimierung der Spalteffizienz. So wird durch die Kombination des Peptidlinkers auf Proteinebene und des Basenpaarabstands auf DNA-Ebene die bestmögliche Bedingung ermittelt, um die spezifische Bindung des DNA-Bindungsmoduls und Spaltung des DNA-Spaltungsmoduls zu gewährleisten. Das adressierte Substrat dient in erster Linie der Untersuchung der Aktivität der Fusionskonstrukte. (ii) Im Gegensatz dazu dienen die

unadressierten Substrate zur Untersuchung der Spezifität der Fusionskonstrukte. Das unadressierte Substrat weist entweder die Erkennungssequenz des DNA-Bindungsmoduls (BM: I-SceI (S), AvBrs3 (T)) auf, jedoch keine nahegelegene Erkennungssequenz für das DNA-Spaltungsmodul (unadressiertes Substrat BM), oder es ist lediglich die Erkennungssequenz für das DNA-Spaltungsmodul (SM: MutH (H), Hinp1I (P)), aber keine Erkennungssequenz für das DNA-Bindungsmodul vorhanden (unadressiertes Substrat SM). Bei absoluter Spezifität des Fusionskonstrukts sollte keine Spaltung der unadressierten Substrate stattfinden. Für das Fusionskonstrukt TALE-Hinp1I wurde im Fall des unadressierten Substrats nur ein Plasmid mit der Erkennungssequenz für das DNA-Spaltungsmodul, aber ohne DNA-Bindungsmodul verwendet. Alle Substrate wurden hergestellt, indem DNA-Kassetten mit den entsprechenden Erkennungssequenzen in Plasmide eingebracht wurden. An dieser Stelle sei bemerkt, dass auf Grund des natürlichen Vorkommens der Erkennungssequenz von MutH (GATC) und Hinp1I (GCGC), diese neben den gezielt eingeführten DNA-Kassetten zahlreich auf den Plasmidsubstraten vorkommen (18 bis 21 zusätzliche Erkennungssequenzen; abhängig von der Klonierungsstrategie). Neben dem Einsatz der Plasmide als Substrate wurde für die Fusionskonstrukte TALE-MutH und TALE-Hinp1I PCR-Substrate anhand der Plasmide hergestellt.

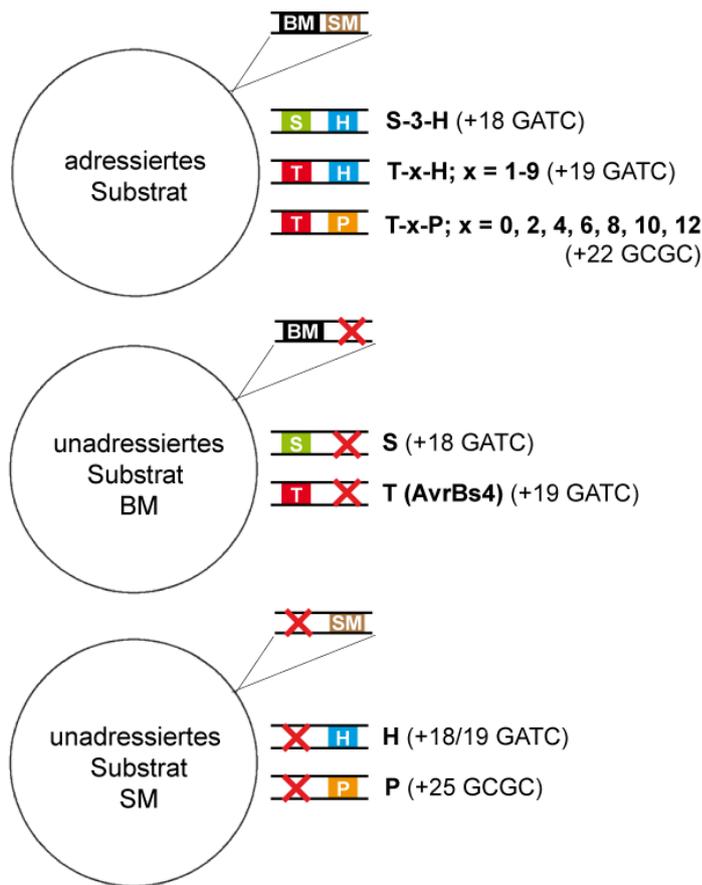


Abbildung 3.3: Schematische Darstellung der Substrat-Konstrukte. Zur Charakterisierung der Spezifität und Aktivität der Fusionskonstrukte wurden zum einen adressierte Substrate generiert, die die Erkennungssequenz für das DNA-Bindungsmodul (BM; I-SceI = S bzw. TALE = T) und eine nahegelegene Erkennungssequenz für das DNA-Spaltungsmodul (SM; MutH = H bzw. HinPII = P) aufweisen, wobei der Abstand variieren kann. Zum anderen wurden bis zu zwei unadressierte Substrate zur Kontrolle generiert. Diese weisen entweder die Erkennungssequenz für das DNA-Bindungsmodul auf, jedoch keine nahegelegene Erkennungssequenz für das DNA-Spaltungsmodul (unadressiertes Substrat BM) oder es ist keine Erkennungssequenz für das DNA-Bindungsmodul aber Erkennungssequenzen für das DNA-Spaltungsmodul vorhanden (unadressiertes Substrat SM). Die Anzahl der GATC (MutH) und GCGC (HinPII) in den Klammern verweist auf die zusätzlichen Sequenzen, die neben der Modifizierung im Plasmid vorkommen.

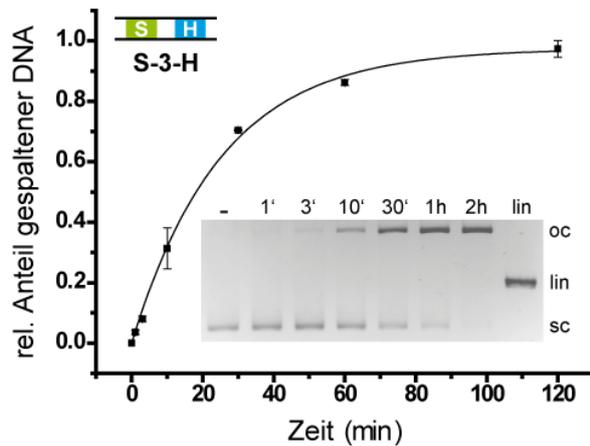
3.4 Charakterisierung des Fusionskonstrukts MutH-I-SceI

3.4.1 Analyse der Aktivität und Spezifität von MutH-I-SceI

Das Fusionskonstrukt MutH-I-SceI wurde mittels dreier Substrate auf Aktivität und Spezifität getestet. Das adressierte Plasmidsubstrat weist die spezifische I-SceI-Erkennungssequenz (S) sowie eine drei Basenpaare entfernte MutH-Erkennungssequenz (H; GATC) auf (S-3-H). Hingegen verfügen die unadressierten Plasmidsubstrate entweder über eine I-SceI-Erkennungssequenz ohne eine nahegelegene MutH-Erkennungssequenz (S) oder es existiert keine I-SceI-Erkennungssequenz auf dem Plasmid (H). Die zeitabhängige Spaltung zeigt deutlich, dass ausschließlich das adressierte Plasmidsubstrat von MutH-I-SceI gespalten wird. Das *supercoiled*-Plasmid wird innerhalb von zwei Stunden vollständig umgesetzt, hingegen findet mit beiden unadressierten Plasmidsubstraten während der gesamten Reaktion keine detektierbare Hydrolyse statt, obwohl das Enzym in vierfachem Überschuss vorliegt (Abb. 3.4). Besonders hervorzuheben ist, dass das adressierte Plasmidsubstrat lediglich in die *open circular*-Form konvertiert wird, also die adressierte MutH-Erkennungssequenz nur in einem einzelnen DNA-Strang gespalten wird. Die ermittelten Spaltraten zeigen eine 1000-fache

Spaltpräferenz für das adressierte Plasmidsubstrat im Vergleich zu den Kontroll-Plasmiden. Die tatsächliche Spaltpräferenz könnte allerdings weit höher liegen, zum einen weil sich auf dem adressierten Plasmidsubstrat neben der adressierten MutH-Erkennungssequenz 19 weitere GATC-Sequenzen befinden und zum anderen weil bei diesem Test die erreichten Detektionsuntergrenzen keine anderen Aussagen erlauben.

Adressiertes Substrat



Unadressierte Substrate

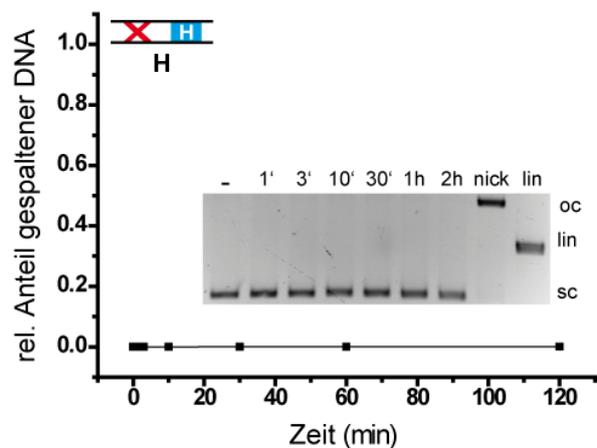
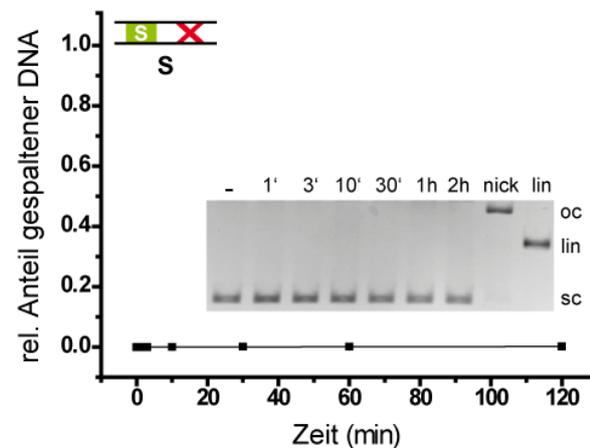


Abbildung 3.4: Zeitabhängige Spaltung des adressierten und unadressierten Substrats mit MutH-I-SceI. In dem Plasmid-Spalttest wurde ein vierfacher Enzymüberschuss gegenüber dem Substrat eingesetzt (16 nM Protein, 4 nM DNA). Das adressierte Plasmidsubstrat weist die spezifische I-SceI-Erkennungssequenz (S) sowie eine drei bp entfernte MutH-Erkennungssequenz (H; GATC) auf (S-3-H). Die unadressierten Plasmidsubstrate verfügen entweder über eine I-SceI-Erkennungssequenz ohne eine nahegelegene MutH-Erkennungssequenz (S) oder es existiert keine I-SceI-Erkennungssequenz (H). Jedes Plasmid hat 19 weitere GATC-Sequenzen. Die möglichen Plasmidformen und -positionen sind mit *supercoiled* (sc), *open circular* (oc) und linear (lin) bezeichnet. Als Kontrolle dient der Ansatz ohne MgCl₂ (-), sowie die oc- (*nick*) und die lin-Form des Plasmids. Das adressierte Plasmidsubstrat wird von MutH-I-SceI durch einen Einzelstrangbruch in die oc-Form konvertiert, wohingegen die unadressierten Plasmidsubstrate ungespalten bleiben. Die gemessene Spaltrate für die adressierte Spaltung beträgt $4.0 \times 10^{-2} \text{ min}^{-1}$. Die Spaltraten für die unadressierten Plasmidsubstrate liegen unterhalb des Detektionslimits von $4 \times 10^{-5} \text{ min}^{-1}$. Analyse über Agarose-Gelelektrophorese.

3.4.2 Inaktivierung von MutH-I-SceI durch proteolytische Spaltung des Peptidlinkers

Um die Abhängigkeit des DNA-Spaltungsmoduls vom DNA-Bindungsmodul zu untersuchen, wurde der verknüpfende Peptidlinker proteolytisch gespalten und die Module somit entkoppelt. Nach Inkubation des Fusionsproteins mit der TEV Protease zeigt der anschließende Plasmid-Spalttest, dass das adressierte Substrat nicht gespalten wird (Abb. 3.5 B). Zur Kontrolle erfolgten zwei parallele Ansätze. Der erste Kontrollansatz soll demonstrieren, dass der katalytische Verlust von MutH-I-SceI nicht auf die Durchführung des Experiments zurückzuführen ist, da das Fusionskonstrukt nach gleicher Vorgehensweise, aber ohne den Einsatz der TEV Protease, die erwartete Spaltaktivität zeigt (Abb. 3.5 A). Für den zweiten Kontrollansatz wurde eine Variante des Fusionskonstrukts verwendet, die keine TEV-Erkennungssequenz im Peptidlinker aufweist, aber sonst vergleichbare Aktivität und Spezifität zeigt (Ergebnisse hier nicht gezeigt). Auf diese Weise kann ebenfalls ausgeschlossen werden, dass das Fusionsprotein eine TEV-Erkennungssequenz jenseits des Peptidlinkers besitzt. Auch in diesem Fall bleibt die katalytische Aktivität erhalten (Abb. 3.5 C). Unter physiologischen Pufferbedingungen ist MutH somit nur katalytisch aktiv ist, wenn es in räumliche Nähe der Erkennungssequenz durch das DNA-Bindungsmodul I-SceI rekrutiert wird.

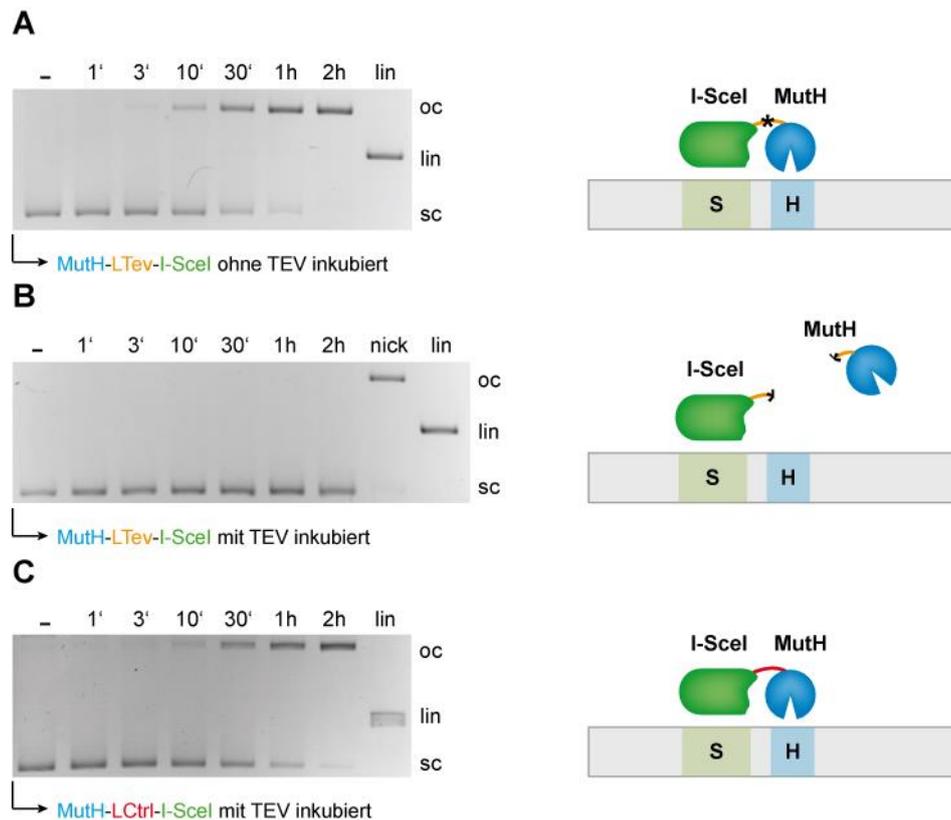


Abbildung 3.5: Vergleich der katalytischen Aktivität von MutH-I-SceI in Fusion und nach räumlicher Trennung des DNA-Bindungs- und -Spaltungsmoduls. Zeitabhängige Spaltung des adressierten Substrats mit MutH-I-SceI. In dem Plasmid-Spalttest wurde ein vierfacher Enzymüberschuss gegenüber dem Substrat eingesetzt (16 nM Protein, 4 nM DNA). Das adressierte Plasmidsubstrat weist die spezifische I-SceI-Erkennungssequenz (S) und eine drei bp entfernte MutH-Erkennungssequenz (H; GATC) auf (S-3-H), sowie 19 weitere GATC-Sequenzen auf dem Plasmid. (A) Plasmid-Spalttest mit vorheriger Inkubation von MutH-I-SceI ohne TEV. (B) Plasmid-Spalttest mit vorheriger Inkubation von MutH-I-SceI mit TEV. (C) Plasmid-Spalttest mit MutH-I-SceI-Variante ohne TEV-Erkennungssequenz im Peptidlinker, vorheriger Inkubation mit TEV. Die möglichen Plasmidformen und -positionen sind mit *supercoiled* (sc), *open circular* (oc) und linear (lin) bezeichnet. Als Kontrolle dient der Ansatz ohne MgCl₂ (-), sowie die oc- (*nick*) und die lin-Form des Plasmids. Analyse über Agarose-Gelelektrophorese.

3.4.3 Bestimmung der Strangpräferenz von MutH-I-SceI

Aus den Plasmid-Spalttests ergibt sich, dass das adressierte Plasmidsubstrat lediglich durch einen Einzelstrangbruch attackiert wird. Um zu bestimmen, ob der Einzelstrang präferenziell im oberen oder im unteren Strang hydrolysiert wird (definiert durch die asymmetrische Erkennungssequenz des TALE-Proteins), wurde das adressierte Plasmidsubstrat nach vollständiger Spaltung durch MutH-I-SceI mittels Sequenzierung analysiert. Hierzu wurden entsprechende *primer* zur Sequenzierung eingesetzt, die vor und hinter der MutH-I-SceI-Erkennungssequenz binden. Das Ergebnis zeigt, dass der obere Strang vollständig sequenziert werden konnte, wohingegen die Sequenzierung des unteren Strangs am 5'GATC abbricht

(Abb. 3.6). Das entspricht dem Spaltmusters von MutH und deutet darauf hin, dass nach Spaltung des adressierten Plasmidsubstrats mit dem Fusionskonstrukt MutH-I-SceI der obere Strang intakt bleibt, wohingegen der untere Strang gespalten wird.

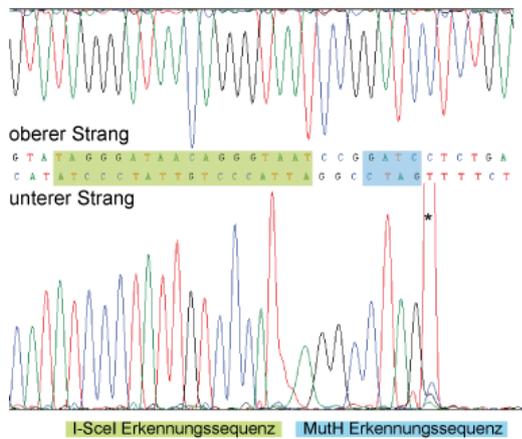


Abbildung 3.6: Analyse der Strangpräferenz von MutH-I-SceI durch Sequenzierung des adressierten Plasmidsubstrats nach vollständiger Hydrolyse. Das Spaltprodukt wurde gelgereinigt und die Erkennungssequenz (S-3-H) mittels spezifischer *primer* sequenziert. Zur Illustration sind die Sequenzierungsergebnisse revers und komplementär dargestellt. Die Erkennungssequenz von I-SceI bzw. MutH sind grün bzw. blau unterlegt. Die Sequenzierung des untern Strangs bricht am 5'GATC ab. Der Stern (*) verweist auf einen Artefakt der Taq-Polymerase.

3.5 Charakterisierung des Fusionskonstrukts TALE-MutH

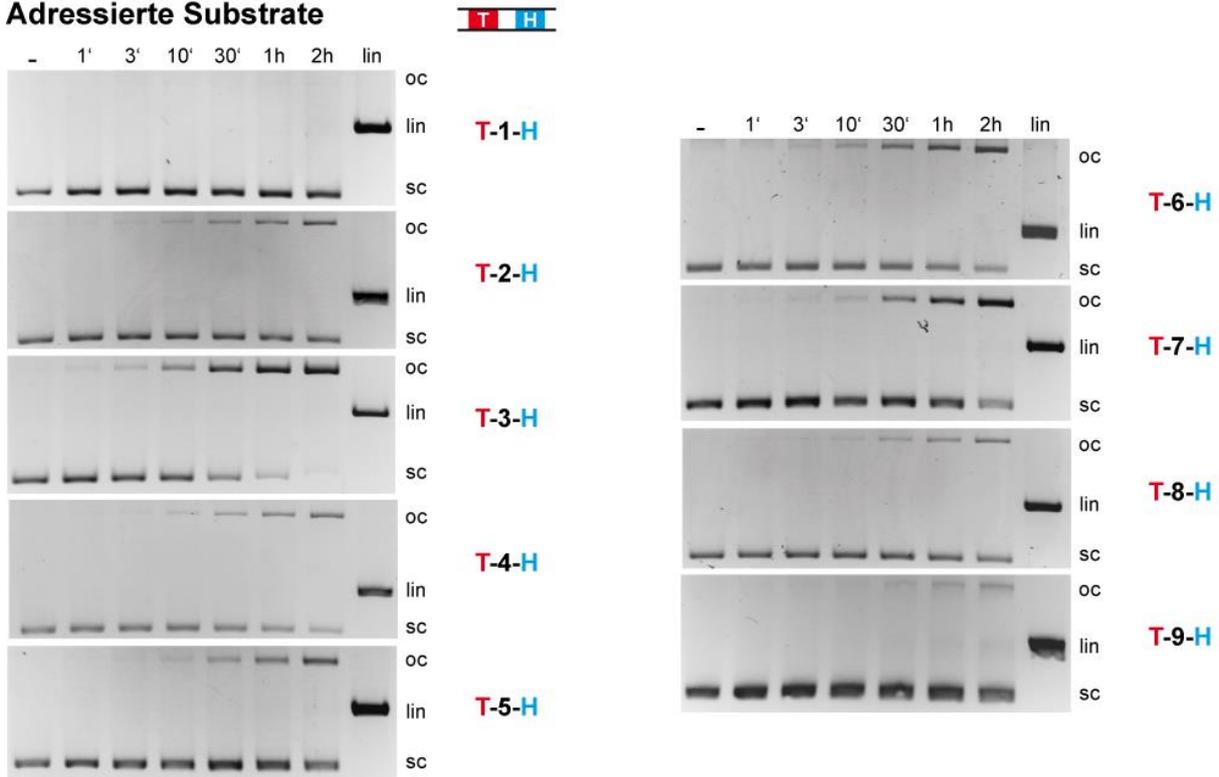
3.5.1 Analyse der Aktivität und Spezifität von TALE-MutH

Analog zum Plasmid-Spalttest mit MutH-I-SceI wurde auch die Aktivität und Spezifität des Fusionskonstrukts TALE-MutH mittels dreier Substrate getestet. Das adressierte Plasmidsubstrat weist die spezifische AvrBs4-Erkennungssequenz (T) sowie eine nahegelegene MutH-Erkennungssequenz (H; GATC) auf. In diesem Fall wurden neun unterschiedliche adressierte Plasmidsubstrate getestet, die im Basenpaarabstand zwischen der AvrBs4- und MutH-Erkennungssequenz variieren (T-x-H: x = 1, 2, 3, 4, 5, 6, 7, 8, 9). Die unadressierten Plasmidsubstrate weisen entweder eine AvrBs4-Erkennungssequenz ohne eine nahegelegene MutH-Erkennungssequenz (T) auf oder es existiert keine AvrBs4-Erkennungssequenz auf dem Plasmid (H). TALE-MutH ist lediglich auf den adressierten Plasmidsubstraten katalytisch aktiv und weist keine detektierbare Aktivität auf den Kontroll-Substraten auf (Abb. 3.7 A und B). Die Spaltaktivität mit den adressierten Substraten variiert stark mit dem Abstand zwischen den Erkennungssequenzen von AvrBs4 und MutH. Die ermittelten Spaltraten zeigen, dass das adressierte Plasmidsubstrat mit drei Basenpaaren Abstand zwischen AvrBs4- und MutH-Erkennungssequenz am schnellsten umgesetzt wird, gefolgt vom Substrat mit sechs Basenpaaren Abstand (Abb. 3.7 C). Für die Plasmidsubstrate jenseits dieser beiden Maxima sinkt die Spaltrate deutlich ab. Vergleichbar mit dem Plasmid-Spalttest mit MutH-I-SceI findet auch hier die katalytische Aktivität lediglich in Form von Einzelstrangbrüchen statt. Die

ermittelten Spaltraten zeigen wie im Fall des Fusionskonstrukts MutH-I-SceI eine 1000-fache Spaltpräferenz für das adressierte Plasmidsubstrats T-3-H gegenüber den Kontroll-Plasmiden. Die tatsächliche Spaltpräferenz könnte auch hier weit höher liegen, zum einen weil sich auf dem adressierten Plasmidsubstrat neben der adressierten MutH-Erkennungssequenz 19 weitere GATC-Sequenzen befinden und zum anderen weil bei diesem Test die erreichten Detektionsuntergrenzen keine andere Aussagen erlauben.

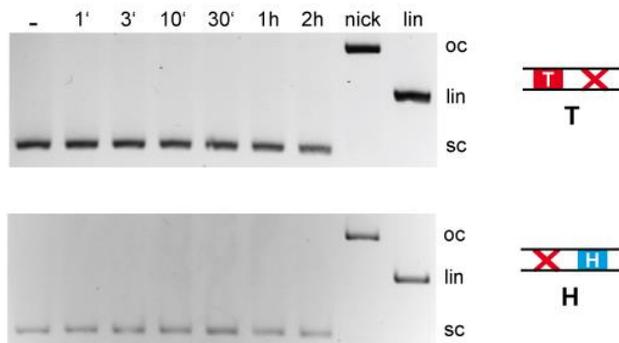
A

Adressierte Substrate



B

Unadressierte Substrate



C

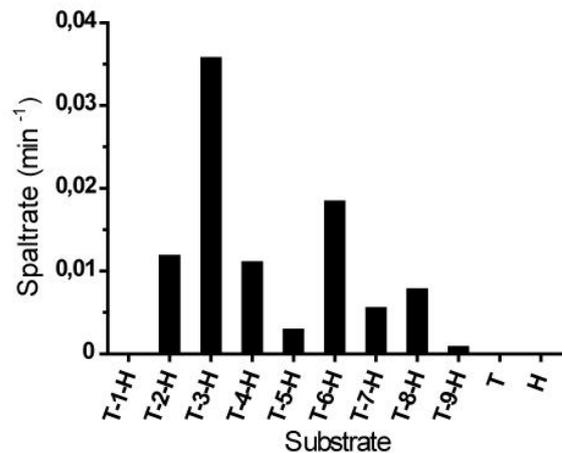


Abbildung 3.7: Zeitabhängige Spaltung des adressierten und unadressierten Substrats mit TALE-MutH. In dem Plasmid-Spaltest wurde ein vierfacher Enzymüberschuss gegenüber dem Substrat eingesetzt (16 nM Protein, 4 nM DNA). **(A)** Die adressierten Plasmidsubstrate weisen die spezifische AvrBs4-Erkennungssequenz (T) sowie eine nahegelegene MutH-Erkennungssequenz (H; GATC) auf (T-x-H: x = 1, 2, 3, 4, 5, 6, 7, 8, 9). Die Plasmidsubstrate werden abhängig vom Basenpaarabstand zwischen den Erkennungssequenzen unterschiedlich schnell von TALE-MutH durch einen Einzelstrangbruch in die oc-Form konvertiert. **(B)** Die unadressierten Plasmidsubstrate verfügen entweder über eine AvrBs4-Erkennungssequenz ohne eine nahegelegene MutH-Erkennungssequenz (T) oder es existiert keine AvrBs4-Erkennungssequenz (H). Mit den Kontrollsubstrate fand keine detektierbare katalytische Aktivität statt. Jedes Plasmid hat 18 weitere GATC-Sequenzen. Die möglichen Plasmidformen und -positionen sind mit *supercoiled* (sc), *open circular* (oc) und linear (lin) bezeichnet. Als Kontrolle dient der Ansatz ohne MgCl₂ (-), sowie die oc-(*nick*) und die lin-Form des Plasmids. **(C)** Graphische Darstellung der berechneten Spaltraten für die oben genannten Plasmid-Spaltests. Die beste Spaltung ergab sich für das adressierte Substrat mit 3 bp Abstand zwischen den Erkennungssequenzen. Die ermittelten Spaltraten

betragen: $1,3 \times 10^{-4} \text{ min}^{-1}$ (T-1-H); $1,2 \times 10^{-2} \text{ min}^{-1}$ (T-2-H); $3,6 \times 10^{-2} \text{ min}^{-1}$ (T-3-H); $1,1 \times 10^{-2} \text{ min}^{-1}$ (T-4-H); $0,3 \times 10^{-2} \text{ min}^{-1}$ (T-5-H); $1,8 \times 10^{-2} \text{ min}^{-1}$ (T-6-H); $0,6 \times 10^{-2} \text{ min}^{-1}$ (T-7-H); $0,8 \times 10^{-2} \text{ min}^{-1}$ (T-8-H) und $8,9 \times 10^{-4} \text{ min}^{-1}$ (T-9-H). Die Spaltraten für die unadressierten Plasmidsubstrate liegen unterhalb des Detektionslimits von $4 \times 10^{-5} \text{ min}^{-1}$. Analyse über Agarose-Gelelektrophorese.

3.5.2 Bestimmung der Strangpräferenz von TALE-MutH mittels Fluorophor-markierten Polynukleotid-Substraten

Zur Analyse und Illustration der Strangpräferenz von TALE-MutH wurden Fluorophor-markierte Polynukleotid-Substrate verwendet. Diese wurden hergestellt indem zwei Oligonukleotide, die jeweils am 5'-Ende mit den Fluorophor Atto 488 bzw. Atto 647N gekoppelt sind, als fwd bzw. rev *primer* für die PCR eingesetzt wurden und die spezifische Zielsequenz des Fusionsproteins T-x-H ($x = 1, 2, 3, 4, 5, 6, 7, 8, 9$) auf dem Plasmidsubstrat (*template*) amplifiziert wurde. Das Polynukleotid-Substrat ist 211 Basenpaare lang, wobei die Spaltprodukte bei Einzelstrangbrüchen im oberen Strang 169-172 Nukleotide lang sind und ein grünes Signal detektiert werden kann bzw. bei Einzelstrangbrüche im unteren Strang 38-35 Nukleotide lang sind und rot fluoreszieren (Abb. 3.8 A). Der Plasmid-Spalttest zeigte, dass es zwei deutliche Maxima der Aktivität mit den verschiedenen adressierten Substraten gab: T-3-H und T-6-H. Bei jeweils kleineren und größeren Abständen zwischen den Erkennungssequenzen sank die Aktivität wieder ab (Abb. 3.7 C). Die strangspezifische Untersuchung ergab, dass das Fusionskonstrukt TALE-MutH verschiedene Strangpräferenzen für das adressierte Substrat mit drei und sechs Basenpaaren Abstand hat. Die Spaltung des Substrats T-3-H erfolgt präferenziell im unteren Strang, wohingegen das Substrat T-6-H ausschließlich im oberen Strang gespalten wird (Abb. 3.8 B). Wie der Plasmid-Spalttest bereits zeigt, wird das Substrat T-6-H langsamer umgesetzt als T-3-H. Unter den gleichen Voraussetzungen wird T-3-H vollständig, T-6-H jedoch nur zu etwa 40 % gespalten (Abb. 3.8 C). Die Spaltrate des unteren Strangs von T-3-H ist 2800-fach höher im Vergleich zu der Hydrolyse des oberen Strangs. Die Analyse der restlichen adressierten Substrate hat ergeben, dass der Wechsel der Strangpräferenz zwischen 4 und 5 Basenpaaren Abstand erfolgt (Abb. 3.8 D). Bei Substraten mit einem bis vier Basenpaaren Abstand zwischen der AvrBs4- und MutH-Erkennungssequenz wird der Einzelstrangbruch im unteren Strang hydrolysiert. Die Einzelstrangspaltung der Substrate mit 5 bis 9 Basenpaaren Abstand erfolgt hingegen im oberen Strang. Analog zu der Ermittlung der Strangpräferenz von MutH-I-SceI (Kapitel 3.4.3) fand auch in diesem Fall die Analyse auf Nukleotid-Ebene für die spezifisch gespaltenen Plasmidsubstrate T-3-H und T-6-H statt, mit

dem Ergebnis, dass die Hydrolyse spezifisch am 5'GATC stattfindet (Ergebnisse hier nicht gezeigt).

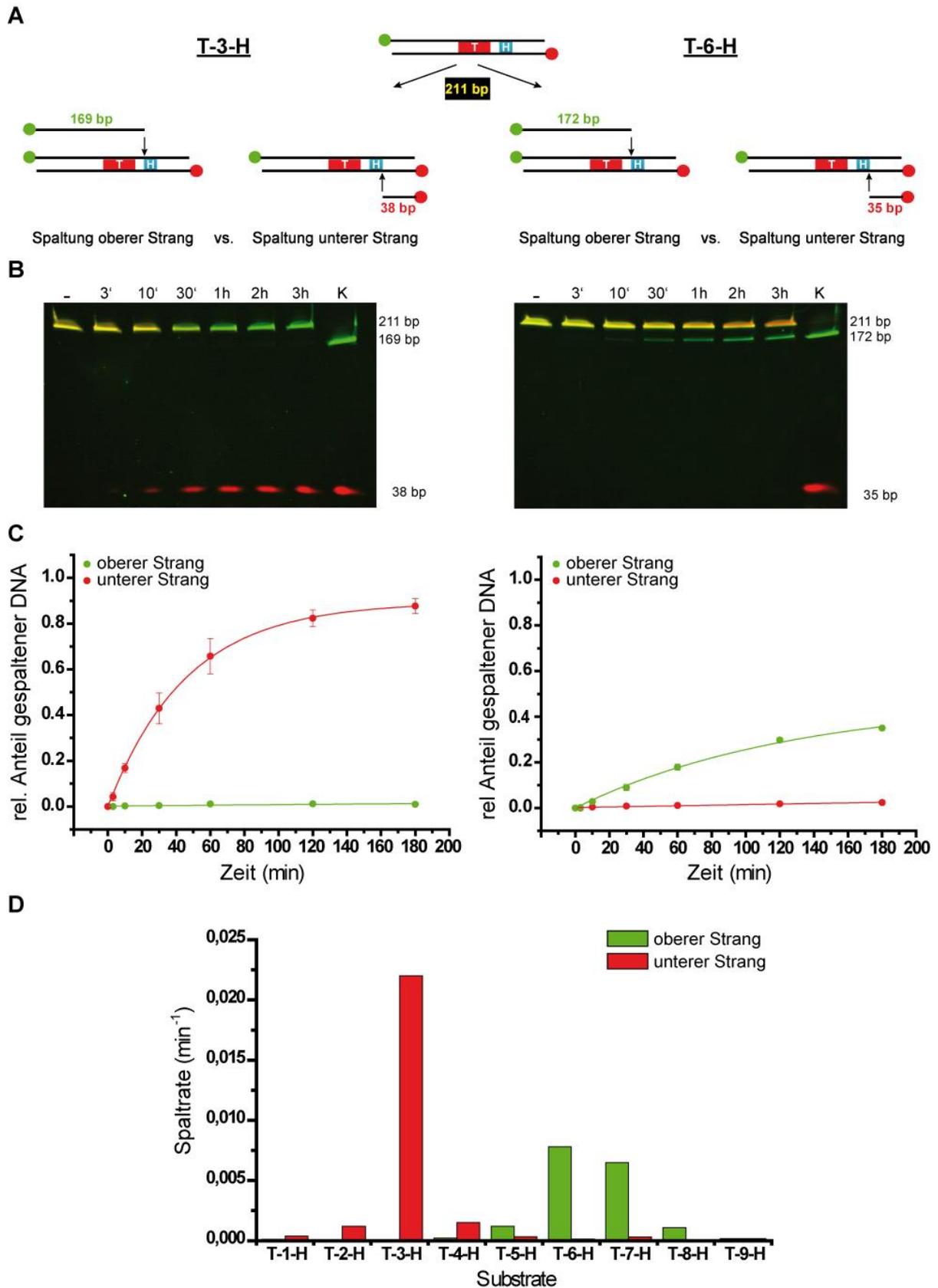


Abbildung 3.8: Analyse der Strangpräferenz von TALE-MutH mittels zeitabhängiger Spaltung der adressierten Fluorophor-markierten Polynukleotid-Substrate. In dem Aktivitätstest wurde ein sechsfacher

Enzymüberschuss gegenüber dem Substrat eingesetzt (120 nM Protein, 20 nM DNA). Die adressierten Substrate weisen die spezifische AvrBs4-Erkennungssequenz (T) sowie eine nahegelegene MutH-Erkennungssequenz (H; GATC) auf (T-x-H: x: 1, 2, 3, 4, 5, 6, 7, 8, 9). **(A)** Schema der Fluorophor-markierten Polynukleotid-Substrates T-3-H bzw. T-6-H und der möglichen Spaltprodukte. Die Länge des ungestalteten Substrats beträgt 211 bp. Erfolgt der Einzelstrangbruch im oberen (unteren) Strang entstehen 169 (38) nt lange Fragmente für T-3-H und 172 (35) nt lange Fragmente für T-6-H. Grüner: Atto 488. Roter: Atto 647N. **(B)** Detektion des Fluoreszenzsignals nach zeitabhängiger Spaltung und Analyse über denaturierende Polyacrylamid-Gelelektrophorese. Gelbes Fluoreszenzsignal: Überlagerung von Atto 488 und Atto 647N (ungespaltene Substrate). Grünes Fluoreszenzsignal: Atto 488 (Einzelstrangbruch im oberen Strang). Rotes Fluoreszenzsignal: Atto 647N (Einzelstrangbruch im unteren Strang). **(C)** Das Substrat T-3-H wird durch TALE-MutH präferenziell im unteren Strang gespalten, das Substrat T-6-H hingegen im oberen Strang. Ermittelte Spaltrate für Hydrolyse des oberen (unteren) Strangs $6,0 \times 10^{-5} \text{ min}^{-1}$ ($2,2 \times 10^{-2} \text{ min}^{-1}$) für T-3-H und $7,8 \times 10^{-3} \text{ min}^{-1}$ ($1,3 \times 10^{-4} \text{ min}^{-1}$) für T-6-H. **(D)** Graphische Darstellung der berechneten Spaltraten für alle adressierten Substrate (T-x-H, x = 1, 2, 3, 4, 5, 6, 7, 8, 9).

3.5.3 Untersuchung verschiedener Linker-Varianten von TALE-MutH

Nach ausgiebiger Untersuchung des ersten TALE-MutH Konstrukts wurden weitere Varianten hergestellt, indem der Peptidlinker zwischen AvrBs4 und MutH variiert wurde. Im ursprünglichen Konstrukt ist MutH nach 28 Aminosäuren (AS) mit AvrBs4 nach dem letzten halben *repeat* fusioniert (im Folgenden TALE(28)-MutH). Darauf aufbauend wurde in einer weiteren Variante ein zusätzlicher 16 AS langer Linker zwischen die Fusionspartner eingebracht (TALE(28)-L16-MutH). Außerdem wurden zwei zusätzliche Alternativen getestet: die Fusion zwischen AvrBs4 und MutH ohne die 28 AS nach dem letzten halben *repeat*, aber mit dem oben genannten 16 AS langem Linker (TALE-L16-MutH) und die direkte Fusion von MutH nach dem letzten halben *repeat* von AvrBs4, also ohne Linker (TALE-L0-MutH). Analog zu den vorherigen Plasmid-Spalttests wurde die Aktivität der Linker-Varianten mittels des adressierten Plasmidsubstrats ermittelt. Auch hier wurden alle neun adressierten Plasmidsubstrate verwendet, die im Basenpaarabstand zwischen der AvrBs4- und MutH-Erkennungssequenz variieren (T-x-H = 1, 2, 3, 4, 5, 6, 7, 8, 9). Abweichend zu den vorherigen Experimenten, wurde in diesem Fall aus Gründen der Vergleichbarkeit die Spaltrate ermittelt, indem angenommen wurde, dass jedes Substrat theoretisch zu 100% umgesetzt werden kann, insofern Aktivität detektiert wird (dies führt teilweise zu kleinen Abweichungen der Spaltraten für TALE-MutH gegenüber den oben genannten Ergebnissen). Die Plasmid-Spalttests haben ergeben, dass die Variation des Peptidlinkers im Fusionskonstrukt keine Verbesserung der Aktivität zur Folge hat, die effizienteste Spaltung zeigt sich für TALE(28)-MutH auf den Plasmidsubstraten T-3-H und T-6-H (Abb. 3.9 A). Ein ähnliches Muster der Spaltraten erkennt man auch für die Variante ohne Linker (TALE-L0-MutH), wenn auch mit niedrigeren

Aktivitäten. Hier ist die Präferenz für die Substrate allerdings um ein Basenpaar nach vorne verschoben und liegt bei T-2-H und T-5-H (Abb. 3.9 B). Trotz theoretisch längerem Linker tritt dieser Effekt für TALE-L16-MutH nicht auf. Diese Variante bevorzugt lediglich kürzere Abstände (2 und 3 Basenpaare) zwischen den Erkennungssequenzen von AvrBs4 und MutH (Abb. 3.9 C). Die geringste Aktivität wurde für TALE(28)-L16-MutH gemessen. Die Spaltraten sind für alle getesteten Substrate sehr niedrig, so dass sich keine deutliche Präferenz bestimmen lässt (Abb. 3.9 D).

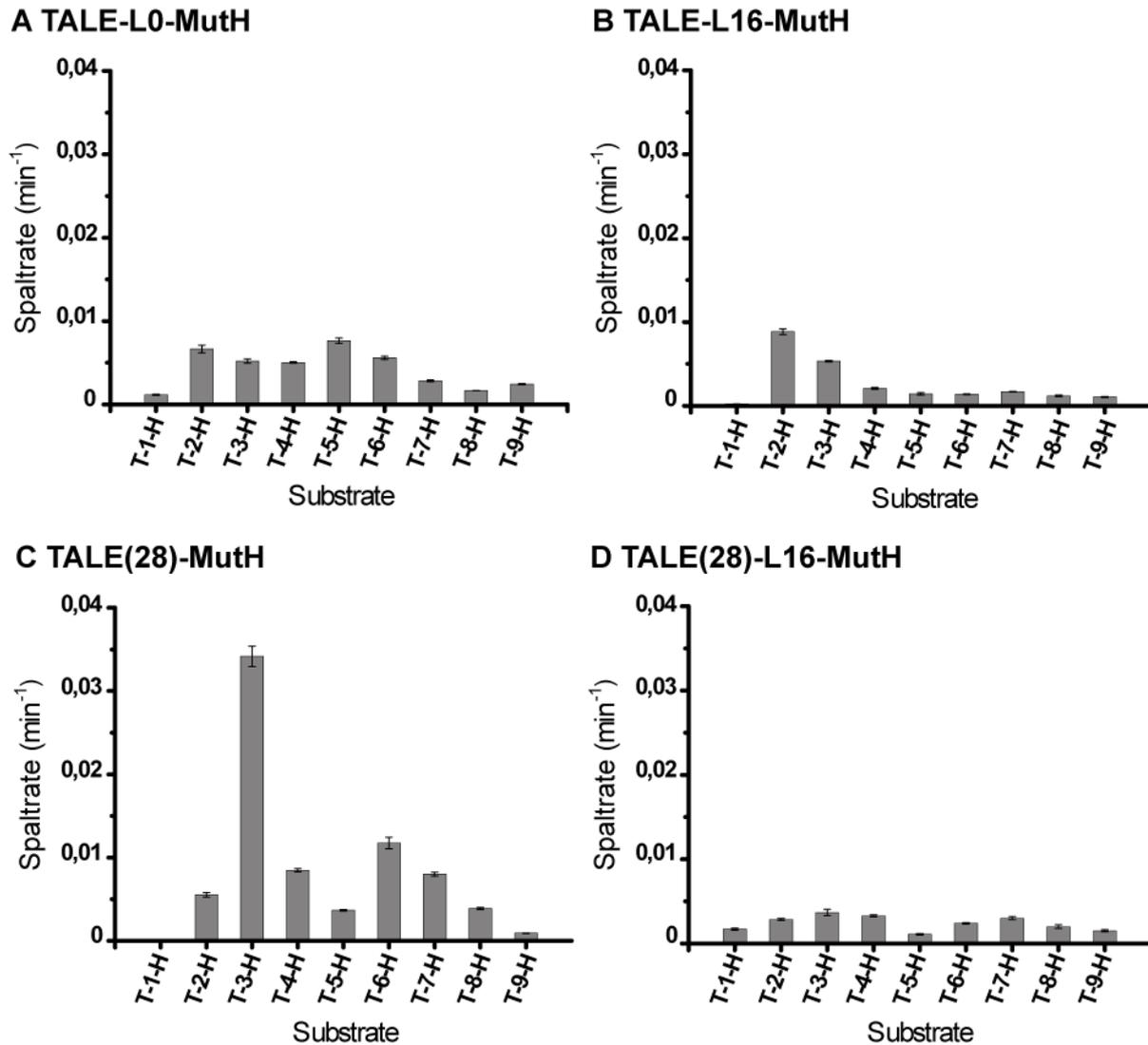


Abbildung 3.9: Berechnete Spaltraten der verschiedenen TALE-MutH Linker-Varianten nach zeitabhängiger Spaltung der adressierten Plasmidsubstrate. In dem Plasmid-Spalttest wurde ein vierfacher Enzymüberschuss gegenüber dem Substrat eingesetzt (16 nM Protein, 4 nM DNA). Die adressierten Plasmidsubstrate weisen die spezifische AvrBs4-Erkennungssequenz (T) sowie eine nahegelegene MutH-Erkennungssequenz (GATC) auf (T-x-H: x: 1, 2, 3, 4, 5, 6, 7, 8, 9). Jedes Plasmid hat 18 weitere GATC-Sequenzen. **(A)** TALE-L0-MutH spaltet bevorzugt die Substrate T-2-H und T-5-H. **(B)** TALE-L16-MutH weist Substratpräferenzen für 2 und 3 Basenpaare Abstand zwischen den Erkennungssequenzen auf.

(C) TALE(28)-MutH zeigt insgesamt die höchsten Spaltraten mit präferenzierter Spaltung der Substrate T-3-H und T-6-H. (D) TALE(28)-L16-MutH hat die niedrigste Spaltaktivität mit keiner deutlichen Substratpräferenz.

3.5.4 Spaltung einer tripartiten Erkennungssequenz durch TALE-MutH

Basierend auf den vorhergehenden Resultaten, dass das Fusionskonstrukt TALE-MutH eine obligatorische Nickase ist, wurde untersucht, ob durch entsprechende Substrat-Architektur auch Doppelstrangbrüche durch das Fusionskonstrukt katalysiert werden können. Hierfür wurden zwei Plasmidsubstrate generiert, die zwei AvrBs4-Erkennungssequenzen (T) aufweisen, die wiederum mit jeweils drei Basenpaaren Abstand durch eine bzw. zwei aufeinanderfolgende MutH-Erkennungssequenz(en) (H) voneinander getrennt sind (T-3-H-3-T bzw. T-3-HH-3-T). Entgegen der Erwartung zeigt der Plasmid-Spalttest keinerlei Aktivität mit dem Substrat T-3-H-3-T (Abb. 3.10 A) und im Vergleich zu dem bipartiten Substrat nur schwache Aktivität mit dem Substrat T-3-HH-3-T, das in einer schwachen *open circular*-Bande resultiert (Abb. 3.10 B).

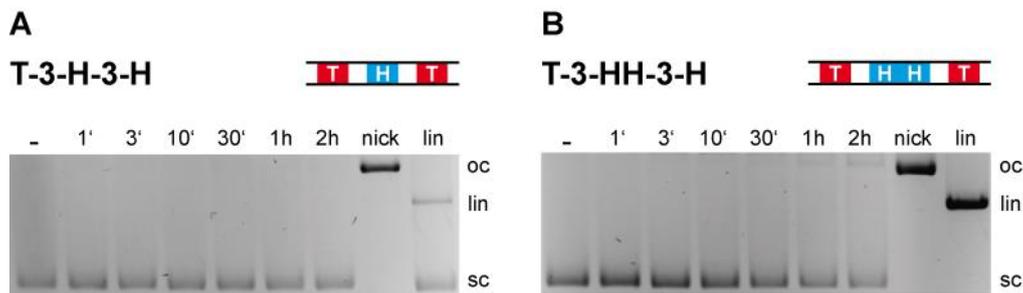


Abbildung 3.10: Zeitabhängige Spaltung des tripartiten Substrats mit TALE-MutH. In dem Plasmid-Spalttest wurde ein achtfacher Enzymüberschuss gegenüber dem Substrat eingesetzt (32 nM Protein, 4 nM DNA). (A) Das Plasmidsubstrate weist zwei spezifische AvrBs4-Erkennungssequenzen (T) auf, die mit jeweils drei Basenpaare Abstand durch eine MutH-Erkennungssequenz (H; GATC) getrennt werden (T-3-H-3-T). (B) Das Plasmidsubstrate weist zwei spezifische AvrBs4-Erkennungssequenzen (T) auf, die mit jeweils drei Basenpaare Abstand durch zwei aufeinanderfolgende MutH-Erkennungssequenzen (H; GATC) getrennt werden (T-3-HH-3-T). Für beide tripartite Substrate konnte kaum katalytische Aktivität mit TALE-MutH detektiert werden. Beide Plasmide weisen 19 weitere GATC-Sequenzen auf. Die möglichen Plasmidformen und -positionen sind mit *supercoiled* (sc), *open circular* (oc) und linear (lin) bezeichnet. Als Kontrolle dient der Ansatz ohne $MgCl_2$ (-), sowie die oc-(*nick*) und die lin-Form des Plasmids. Analyse über Agarose-Gelelektrophorese.

3.5.5 Steigerung der katalytischen Aktivität von TALE-MutH

Für einige Restriktionsenzyme ist bekannt, dass die Verwendung des divalenten Metall-Ions Mn^{2+} anstelle von Mg^{2+} die katalytische Aktivität steigern kann (Hsu & Berg 1978, Woodhead *et al.* 1981). Dieser Effekt kann auch durch Anwesenheit des organischen Lösungsmittels

DMSO (Dimethylsulfoxid) erreicht werden (Malyguine *et al.* 1980, Wei *et al.* 2008). Um zu untersuchen, ob die Katalyse des Fusionskonstrukts TALE-MutH ebenfalls unter diesen Bedingungen beeinflusst wird, wurde ein Plasmid-Spalttest mit dem adressierten Substrat T-3-H analog zu Kapitel 3.5.1 unter Verwendung von 0,5 mM MnCl_2 und 10% DMSO anstelle von 1 mM MgCl_2 durchgeführt. Tatsächlich wurde das Substrat mit MnCl_2 12-mal schneller gespalten. Das Plasmid liegt bereits nach 10 Minuten fast vollständig in der *open circular*-Form vor und wird nach 30 Minuten geringfügig auch linearisiert (Abb. 3.11 A). Die katalytische Aktivität wird mit DMSO zusätzlich um das 2-fache gesteigert. In diesem Fall wird das Plasmid innerhalb von 3 Minuten komplett durch Einzelstrangbrüche umgesetzt und es findet eine stärkere Linearisierung statt (Abb. 3.11 B).

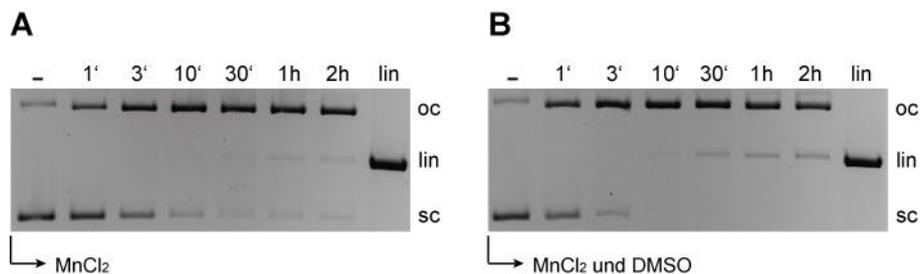


Abbildung 3.11: MnCl_2 - und DMSO-abhängige Steigerung der katalytischen Aktivität von TALE-MutH. In dem Plasmid-Spalttest wurde ein vierfacher Enzymüberschuss gegenüber dem Substrat eingesetzt (16 nM Protein, 4 nM DNA). Als Substrat diente das adressierte Plasmidsubstrat T-3-H (AvrBs4-Erkennungssequenz (T) getrennt durch 3 Basenpaare von der MutH-Erkennungssequenz (H)). **(A)** Verwendung von 0,5 mM MnCl_2 . Ermittelte Spaltrate: $4,1 \times 10^{-1} \text{ min}^{-1}$. **(B)** Verwendung von 0,5 mM MnCl_2 und 10% DMSO. Ermittelte Spaltrate: $8,1 \times 10^{-1} \text{ min}^{-1}$. Unter Verwendung des divalenten Metall-Ions Mn^{2+} wird das Substrat innerhalb von 10 Minuten vollständig in die oc-Form konvertiert. Durch Zusatz von DMSO wird die Aktivität nochmals gesteigert. Das Plasmid hat 18 weitere GATC-Sequenzen. Die möglichen Plasmidformen und -positionen sind mit *supercoiled* (sc), *open circular* (oc) und linear (lin) bezeichnet. Als Kontrolle dient der Ansatz ohne MgCl_2 (-), sowie die lin-Form des Plasmids. Analyse über Agarose-Gelelektrophorese.

3.5.6 *E. coli* Viabilitätstest

Der *E. coli* Viabilitätstest wurde in Form von diskontinuierlichen und kontinuierlichen Wachstumskurven untersucht. Hierfür wurden die Proteine TALE-MutH, TALE, MutH, ZFN und TALE-HinPII in einem *dcm/dam* negativen *E. coli* Stamm exprimiert. Durch die fehlende Dam Methyltransferase sind die GATC-Sequenzen der bakteriellen genomischen DNA nicht geschützt und können prinzipiell von MutH gespalten werden. Verursachen die obengenannten Proteine bei Expression Doppelstrangbrüche durch unspezifische Spaltung, führt das zur genomischen Instabilität der Zellen, die in einem reduzierten Wachstum der Kultur erkennbar wird. Die eigenständigen Fusionseinheiten TALE und MutH, sowie das „leere“

Expressionsplasmid pQE30 dienen als Kontrolle und sollten theoretisch keinen negativen Effekt auf das Zellwachstums zeigen. Das Fusionsprotein TALE-HinP1I EAD und die ZFN lassen hingegen auf Grund von *in vitro* Ergebnissen toxische Effekte *in vivo* vermuten (Schierling *et al.* 2012) (Kapitel 3.6.1). Die diskontinuierlichen Wachstumskurven wurden stündlich in einem Zeitraum von 5 Stunden gemessen. Das kontinuierliche Zellwachstum wurde über acht Stunden erfasst und die Zellen durch regelmäßige Verdünnung mit frischem LB-Medium in der exponentiellen Wachstumsphase gehalten. Durch das kontinuierliche Wachstum können toxische Effekte erhöht und somit besser detektiert werden als mit der klassischen, diskontinuierlichen *batch*-Kultur. Das Zellwachstum wurde bei annähernd gleicher OD_{600nm} gestartet (diskontinuierlich: $\sim 0,05$; kontinuierlich: $\sim 0,1$) und die Proteinexpression nach einer (diskontinuierlichen) bzw. zwei Stunden (kontinuierlich) entsprechend induziert. Bis zu drei Stunden zeigt das kontinuierliche Zellwachstum der verschiedenen Varianten keine nennenswerten Unterschiede, ab der vierten Stunde steigt die OD_{600nm} für die Kultur, die das Fusionskonstrukt TALE-HinP1I EAD exprimiert, nicht weiter an (Abb. 3.12 B), wobei ein geringer Anstieg ab der sechsten Stunde zu verzeichnen ist. Auch die ZFN-Variante zeigt ab der vierten Stunde ein stetig schwindendes Zellwachstum im Vergleich zu den anderen Varianten. Hervorzuheben ist aber, dass das zu untersuchende Fusionskonstrukt TALE-MutH das Zellwachstum von *E. coli* in der diskontinuierlichen als auch kontinuierlichen Kultur nicht negativ beeinflusst, da diese Zellen im Vergleich zu den nicht toxischen Kontrollen (TALE, MutH und pQE30) gleich gut wachsen. Ein ähnliches Ergebnis zeigt sich auch bereits bei den diskontinuierlichen Wachstumskurven ab. Nach der Induktion setzt sich das Wachstum der TALE-HinP1I EAD-Variante stark gegenüber den anderen Varianten ab (Abb. 3.12 A). Auch die ZFN-Variante zeigt ein etwas vermindertes Wachstum. Interessant ist ebenfalls, dass bei beiden Messverfahren die Expression von MutH das Zellwachstum geringfügig begünstigt wird.

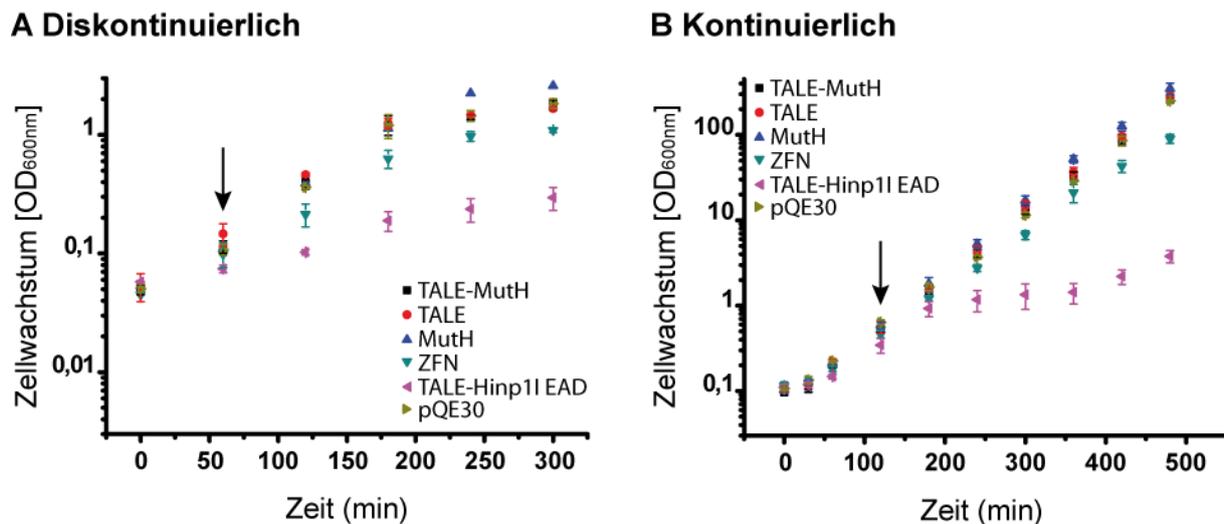


Abbildung 3.12: *E. coli* Viabilitätstest zur Untersuchung toxischer Effekte des Fusionskonstrukts TALE-MutH *in vivo*. Zeitabhängige logarithmische Darstellung der Optischen Dichte bei 600 nm (OD_{600nm}). Zur Kontrolle dient das eigenständige DNA-Bindungsmodul (TALE) und DNA-Spaltungsmodul (MutH), sowie eine rekombinante Zinkfinger Nukleasen (ZFN), das Fusionskonstrukt TALE-HinP1I EAD und das unmodifizierte Expressionsplasmid (pQE30). (A) Die Suspensionsschüttelkulturen wurden bei 37°C inkubiert und die OD nach definierten Zeitintervallen bestimmt (statisch). (B) Die Suspensionsschüttelkulturen wurden bei 37°C inkubiert und die OD nach definierten Zeitintervallen bestimmt, wobei das Medium regelmäßig erneuert wurde, um das Wachstum in der exponentiellen Phase beizubehalten (kontinuierlich). Der Pfeil impliziert den Induktionszeitpunkt der Expression.

3.6 Charakterisierung des Fusionskonstrukts TALE-HinP1I

Anmerkung: Der folgende Abschnitt basiert auf einem ursprünglichen Postdoc-Projekt des Instituts. Die Konstruktion aller Plasmidsubstrate und der Varianten TALE-HinP1I wt, TALE-HinP1I EAD sowie die entsprechende Charakterisierung wurde von Dr. Benno Schierling geplant und ausgeführt. Das gesamte Projekt wurde durch die technische Expertise von Bedriska Reitz unterstützt.

3.6.1 Analyse und Optimierung der Spezifität von TALE-HinP1I

Zu Beginn wurde die Aktivität der TALE-HinP1I-Varianten auf dem unadressierten Plasmidsubstrat untersucht, um zu ermitteln, ob die Spezifität hinreichend ist. Zunächst wurde die Aktivität des TALE-HinP1I wt getestet mit dem Ergebnis, dass das Kontrollsubstrat bei einem Enzym-Substrat Verhältnis von 1:100 innerhalb von zwei Stunden fast vollständig degradiert wird (Abb. 3.13 A). Um die Spezifität des Fusionskonstrukts zu erhöhen, wurden gezielte Aminosäure-Substitutionen im DNA-Bindungsmodul HinP1I eingeführt. Hierbei konzentrierte man sich auf Aminosäuren, von denen man aus der Co-Kristallstruktur schließen

kann, dass sie an der Protein-DNA Interaktion beteiligt sind. Dieser Ansatz zielt darauf hin, die Interaktion zwischen HinP1I und der DNA abzuschwächen (die katalytische Aktivität jedoch beizubehalten) und somit das DNA-Bindungsmodul maßgeblich für die Bindung verantwortlich zu machen. Die Spezifität wird erhöht, indem HinP1I die eigene Erkennungssequenz nicht mehr selbständig binden kann und eine katalytische Aktivität nur ermöglicht wird, indem die HinP1I-Erkennungssequenz sich in unmittelbarer Nähe einer TALE-Erkennungssequenz befindet und somit HinP1I durch die Fusion aktiv durch das TALE Protein rekrutiert wird. Eine erfolgreiche Kombination aus Aminosäure-Substitutionen stellt die TALE-HinP1I EAD-Variante dar (Abb. 3.13 B), bei der die Spezifität enorm gesteigert werden konnte. Im Vergleich zu TALE-HinP1I wt wird das unadressierte Substrat bei einem Enzym-Substrat Verhältnis von 8:1 selbst nach Langzeit-Inkubation (24 Stunden) nicht vollständig degradiert (Abb. 3.13 A). Allerdings ist eine gewisse Unspezifität trotzdem zu detektieren. Eine Umsetzung des Plasmids von der *supercoiled-* zur *open circular*-Form beginnt bereits nach 30 Minuten und resultiert letztendlich in der potentiell toxischen Linearisierung des Plasmids nach drei Stunden. Da die Variante TALE-HinP1I EAD zwar sehr vielversprechend aber trotzdem unspezifisch ist, wurden vier einzelne Aminosäure-Substitutionen zusätzlich eingebracht (Abb. 3.13 C). Im Fall der Varianten TALE-HinP1I EADA8 und TALE-HinP1I EADA97 konnte die Spezifität im Vergleich zu der TALE-HinP1I EAD-Variante nicht maßgeblich gesteigert werden. Auch hier wird das unadressierte Substrat nach 24 Stunden zum Teil linearisiert (Abb. 3.13 A). Vielversprechender scheinen hingegen die Varianten TALE-HinP1I EADD235 und TALE-HinP1I EADA208 zu sein. Erstere zeigt eine stärkere unspezifische Umsetzung des unadressierten Substrats von der sc- zur oc-Form, aber beide Varianten zeigen selbst nach Langzeit-Inkubationen keine unspezifischen Doppelstrangbrüche und weisen somit eine zureichende Spezifität auf (Abb. 3.13 A).

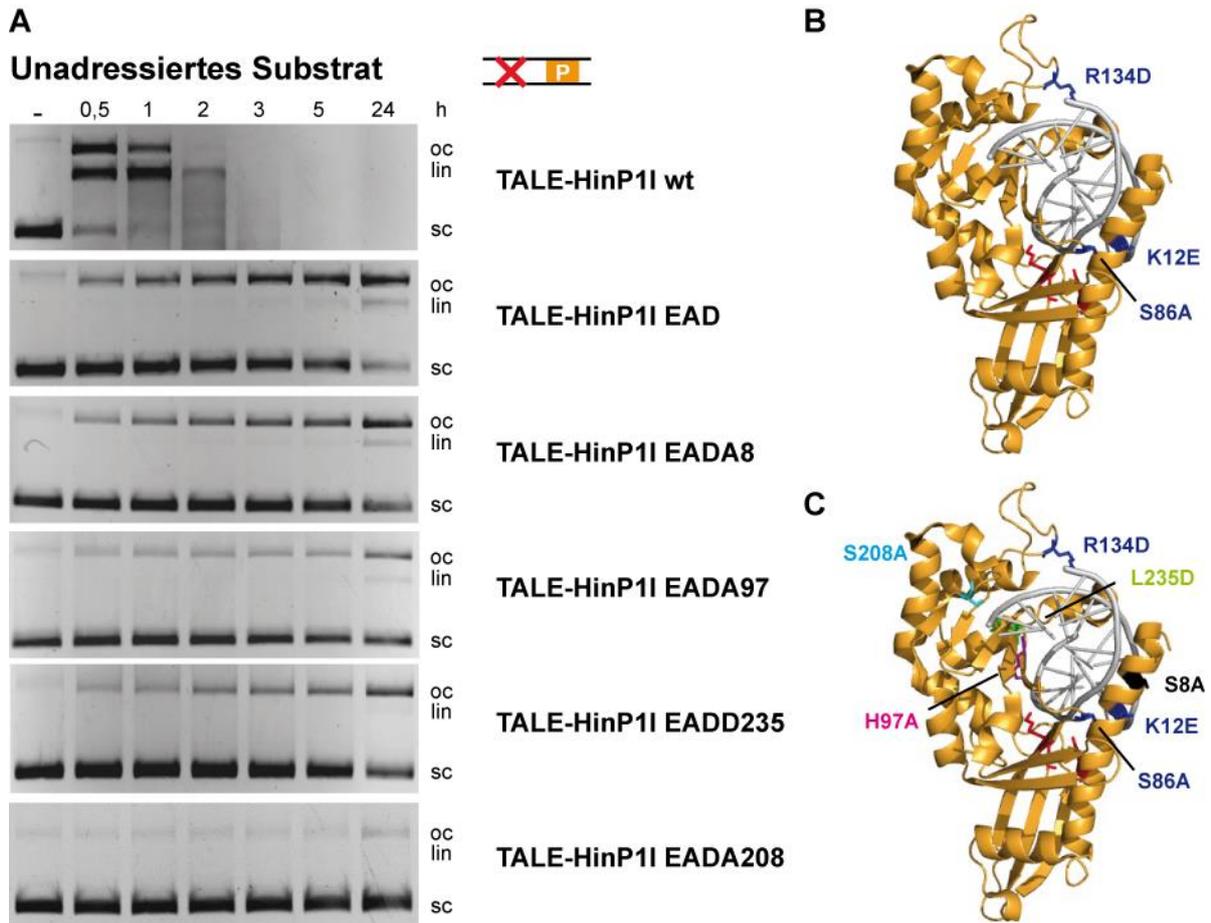


Abbildung 3.13: Zeitabhängige Spaltung des unadressierten Substrats mit TALE-HinP1I und die Optimierung des Konstrukts. (A) Mit Ausnahme von TALE-MutH wt wurde in dem Plasmid-Spaltttest ein achtfacher gegenüber dem Substrat eingesetzt (80 nM Protein, 10 nM DNA; TALE-HinP1I wt = 0,1 nM). Das unadressierte Plasmidsubstrat weist keine AvrBs3-Erkennungssequenz (X) sondern lediglich 26 HinP1I-Erkennungssequenz (P; GCGC) auf. Durch gezielte Aminosäure-Substitutionen kann die Aktivität auf dem unadressierten Substrat durch die Varianten TALE-HinP1I EAD, TALE-HinP1I EADA8 und TALE-HinP1I EADA97 gegenüber dem Fusionskonstrukt TALE-MutH wt reduziert werden. Kaum unspezifische Spaltung konnte mit den Varianten TALE-HinP1I EADD235 und TALE-HinP1I EADA208 detektiert werden. Die möglichen Plasmidformen und –positionen sind mit *supercoiled* (sc), *open circular* (oc) und linear (lin) bezeichnet. Als Kontrolle dient der Ansatz ohne MgCl₂ (-). Analyse über Agarose-Gelelektrophorese. (B) und (C) Die gezielten Aminosäure-Substitutionen sind in der Co-Kristallstruktur des DNA-Spaltungsmoduls HinP1I angedeutet. Das katalytische Zentrum ist rot dargestellt. RCSB Protein Data Bank: 2FL3.

3.6.2 Ermittlung der Aktivität von TALE-HinP1I EADA208 mit verschiedenen adressierten Substraten

Analog zum Plasmid-Spaltttest mit MutH-I-SceI und TALE-MutH wurde die bestmögliche Aktivität der Variante TALE-HinP1I EADA208 mittels adressierter Plasmidsubstrate ermittelt, die im Basenpaarabstand zwischen den Erkennungssequenz von AvrBs3 (T) und HinP1I (P) variieren (T-x-P; x = 0, 2, 4, 6, 8, 10, 12). Die Analyse hat gezeigt, dass Substrate mit keinen

bzw. kleinen Basenpaarabständen (T-0-P und T-2-P) oder großen Abständen (T-10-P und T-12-P) sich nicht eignen, da nur sehr geringe katalytische Aktivität detektiert werden konnte (Abb. 3.14 A). Die Substrate T-4-P, T-6-P und T-8-P werden hingegen unter gleichen Voraussetzung z.T. vollständig prozessiert. Es fällt jedoch sofort auf, dass sich das Spaltmuster zwischen diesen drei Plasmidsubstrate stark unterscheidet (Abb. 3.14 A und B). Das Plasmidsubstrat mit vier Basenpaaren Abstand zwischen den Erkennungssequenzen wird innerhalb der Reaktionszeit zum größten Teil aus der *supercoiled*-, über die *open circular*-, in die linearisierte Form überführt. Im Fall von T-6-P ist die sc-Form des Plasmids bereits nach 10 Minuten nicht mehr vorhanden, allerdings wird die oc-Form während der restlichen Reaktionszeit nicht komplett linearisiert, sondern ist am Ende immer noch zu ~60% vorhanden (Abb. 3.14 B). Ist der Abstand auf acht Basenpaare erweitert, findet nahezu keine Linearisierung des Plasmids statt. Das Substrat wird in die oc-Form überführt und nicht mehr weiter prozessiert (Abb. 3.14 B). Die Substrate T-6-P und T-8-P werden überwiegend durch Einzelstrangbrüche umgesetzt, wohingegen mit dem Substrat T-4-P überwiegend Doppelstrangspaltung stattfindet (Abb. 3.14 C). Besonders hervorzuheben ist, dass in allen Fällen die oc-Form akkumuliert und auf die monomere Aktivität von HinP1I zurück schließen lässt. Das adressierte Plasmidsubstrat T-6-P, das den besten Umsatz des sc-Plasmids aufweist, wird 3,5-mal schneller gespalten im Vergleich zum unadressierten Kontroll-Plasmid (Kapitel 3.6.1). Doch auch in diesem Fall könnte die tatsächliche Spezifität höher sein, wenn man bedenkt, dass 22 unadressierte HinP1I-Erkennungssequenzen auf dem Plasmid vorkommen. Allerdings muss auch angemerkt werden, dass die z.T. unspezifische Spaltung dieser Sequenzen bei der Detektion der adressierten Spaltung ebenfalls erfasst wird.

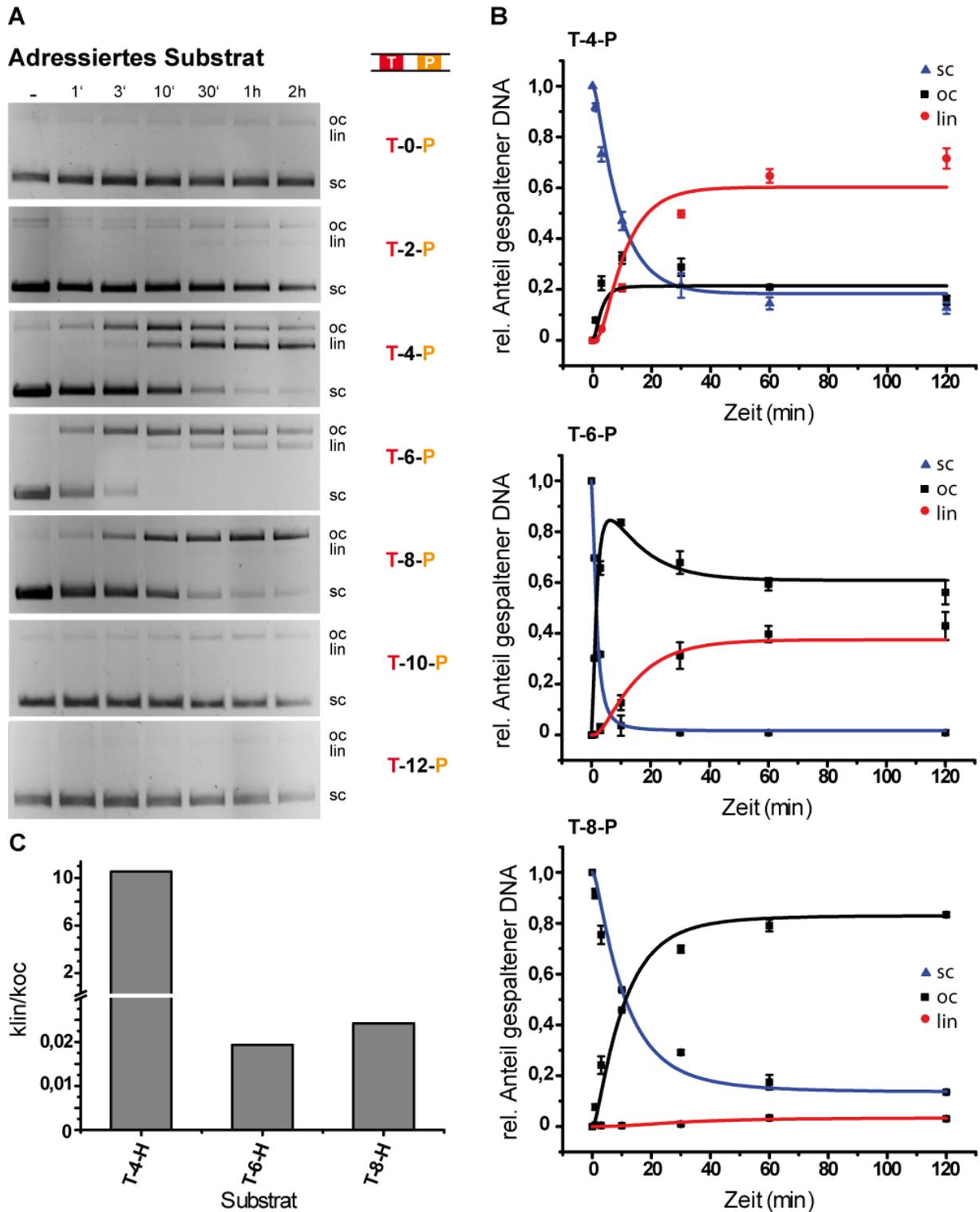


Abbildung 3.14: Zeitabhängige Spaltung des adressierten Substrats mit TALE-HinP1I EADA208. In dem Plasmid-Splitttest wurde ein achtfacher Enzymüberschuss gegenüber dem Substrat eingesetzt (80 nM Protein, 10 nM DNA). (A) Die adressierten Plasmidsubstrate weisen die spezifische AvrBs3-Erkennungssequenz (T) sowie eine nahegelegene HinP1I-Erkennungssequenz (P; GCGC) auf (T-x-P: x: 0, 2, 4, 6, 8, 10, 12). Die Plasmidsubstrate T-4-P, T-6-P und T-8-P werden abhängig vom Basenpaarabstand zwischen den Erkennungssequenzen unterschiedlich schnell von TALE-HinP1I EADA208 gespalten, wobei das Spaltmuster zwischen oc- und lin-Produkt variiert. Mit den Substraten T-0-P, T-2-P sowie T-10-P und T-12-P wurde keine bis

schwache Aktivität gemessen. Jedes Plasmid hat 22 weitere GCGC-Sequenzen. Die möglichen Plasmidformen und -positionen sind mit *supercoiled* (sc), *open circular* (oc) und linear (lin) bezeichnet. Als Kontrolle dient der Ansatz ohne $MgCl_2$ (-). Analyse über Agarose-Gelelektrophorese. **(B)** Graphische Darstellung und Anpassung der Spaltkinetik für die oben genannten Plasmid-Spalttests mit den Substraten T-4-P, T-6-P und T-8-P. Die insgesamt beste Spaltung wurde für das adressierte Substrat mit 6 bp Abstand zwischen den Erkennungssequenzen gemessen. Die ermittelten Spaltraten für oc (lin) betragen: $4,0 \times 10^{-3} \text{ min}^{-1}$ (0 min^{-1}) für T-0-P, $6,6 \times 10^{-3} \text{ min}^{-1}$ ($1,9 \times 10^{-1} \text{ min}^{-1}$) für T-2-P, $1,1 \times 10^{-1} \text{ min}^{-1}$ ($1,2 \text{ min}^{-1}$) für T-4-P, $1,9 \text{ min}^{-1}$ ($3,6 \times 10^{-2} \text{ min}^{-1}$) für T-6-P, $9,0 \times 10^{-2} \text{ min}^{-2}$ ($2,2 \times 10^{-3} \text{ min}^{-1}$) für T-8-P, $4,8 \times 10^{-3} \text{ min}^{-1}$ ($5,4 \times 10^{-3} \text{ min}^{-1}$) für T-10-P, $9,6 \times 10^{-3} \text{ min}^{-1}$ ($3,2 \times 10^{-1} \text{ min}^{-1}$) für T-12-P. **(C)** Graphische Darstellung des Quotienten aus den berechneten Spaltraten (k) für lin und oc.

Die Ergebnisse der Plasmid-Spalttests von TALE-HinP1I EADA208 mit unterschiedlichen adressierten Substraten sowie die Anpassung der Datenpunkte weisen darauf hin, dass sich die Reaktion über die Zeit verlangsamt und nach etwa einer Stunde fast vollkommen zum Erliegen kommt, obwohl das Substrat nicht komplett umgesetzt ist (Abb. 3. 14 B). Um zu untersuchen, ob das Enzym an Aktivität verliert, wurde die Halbwertszeit unter Reaktionsbedingungen ermittelt. Um gegebenenfalls auszuschließen, dass der Aktivitätsverlust lediglich auf die Variante TALE-HinP1I EADA208 zutrifft, wurde die Halbwertszeit auch für TALE-HinP1I EADD235 gemessen. Letztere ist etwas aktiver, zeigt aber im Spezifitätstest ebenfalls keine Linearisierung des unadressierten Substrats (Abb. 3.13 A). Beide Enzyme wurden unter Ausschluss von $MgCl_2$ unter Reaktionsbedingungen (Aktivitätspuffer, Substrat, Temperatur) für 0, 15, 30 bzw. 60 Minuten inkubiert und anschließend ein Aktivitätstest mit dem adressierten Substrat T-6-P durchgeführt. Für beide Varianten konnte gezeigt werden, dass die Aktivität mit zunehmender vorheriger Inkubation sinkt (Abb. 3.15). TALE-HinP1I EAD208 setzt nach 15-minütiger Inkubation das komplette Substrat um, jedoch bereits mit einer verminderten Spaltrate. Nach 30 bzw. 60 Minuten Inkubation werden nur noch ~90 bzw. ~70% des Substrats umgesetzt, wobei die Spaltrate stetig sinkt. Da die Gesamtaktivität von TALE-HinP1I EADD235 etwa doppelt so hoch ist, wird das Substrat auch noch nach 30-minütiger Inkubation vollständig umgesetzt und liegt nach 60 Minuten Inkubation immer noch bei ~90%. Aber auch hier sinkt die Spaltrate vergleichbar zu TALE-HinP1 EADA208, so dass für Beide etwa die gleiche Halbwertszeit von 20 Minuten (TALE-HinP1I EADA208) und 16 Minuten (TALE-HinP1I EADD235) ermittelt wurde.

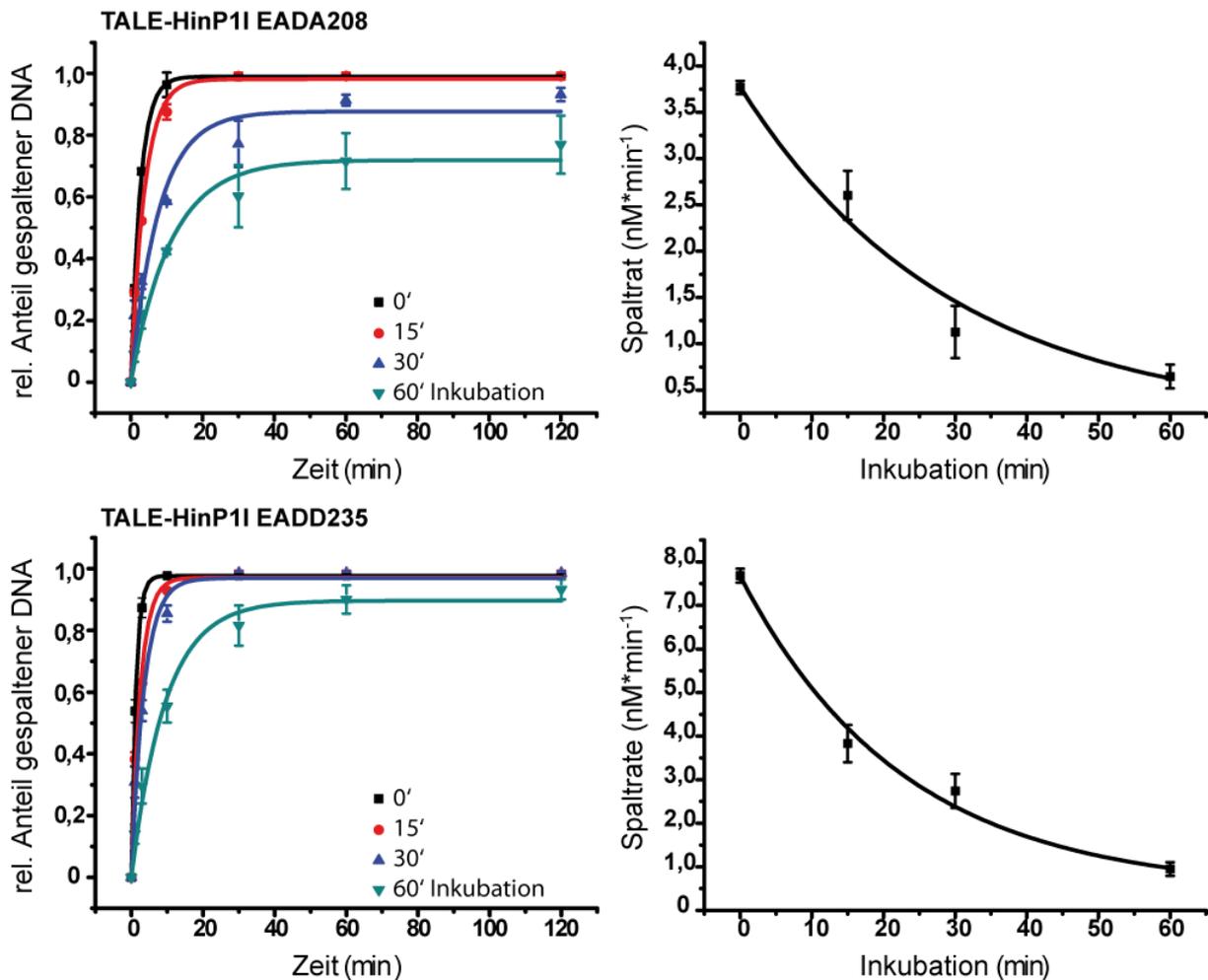


Abbildung 3.15: Bestimmung der Halbwertszeit von TALE-HinP1I EADA208 und TALE-HinP1I EADD235. Beide Fusionskonstrukte wurden unter Reaktionsbedingungen (Aktivitätspuffer, Substrat, Temperatur wie in den Plasmid-Spaltests), aber Ausschluss von $MgCl_2$ inkubiert und anschließend die Aktivität bestimmt. In dem Plasmid-Spaltest wurde ein achtfacher Enzymüberschuss gegenüber dem Substrat eingesetzt (80 nM Protein, 10 nM DNA). Es wurde das adressierte Plasmidsubstrat T-6-P verwendet. Mit zunehmender Vorinkubationszeit sinkt die Spaltrate. Die ermittelten Halbwertszeiten (λ) betragen 20 Minuten für TALE-HinP1I EADA208 und 16 Minuten für TALE-HinP1I EADD235. Anpassung Spaltrate gegen Vorinkubationszeit mit $y = A1 \cdot \exp(-x/t1) + y0 \rightarrow \lambda = t1 \cdot \ln(2)$. Die ermittelten Spaltraten betragen: $3,8 \times 10^{-1} \text{ min}^{-1}$ (0°), $2,7 \times 10^{-1} \text{ min}^{-1}$ (15°), $1,3 \times 10^{-2} \text{ min}^{-1}$ (30°) und $9,0 \times 10^{-2} \text{ min}^{-1}$ (60°) für TALE-HinP1I EADA208 bzw. $7,9 \times 10^{-1} \text{ min}^{-1}$ (0°), $3,9 \times 10^{-1} \text{ min}^{-1}$ (15°), $2,8 \times 10^{-1} \text{ min}^{-1}$ (30°) und $1,1 \times 10^{-1} \text{ min}^{-1}$ (60°) für TALE-HinP1I EADD235.

3.6.4 Bestimmung der Strangpräferenz von TALE-HinP1I

Entsprechend der Ermittlung der Strangpräferenz von TALE-MutH (Kapitel 3.5.2) wurden auch in diesem Fall Fluorophor-markierte Polynukleotid-Substrate hergestellt indem zwei Oligonukleotide (jeweils am 5' Ende mit den Fluorophor Atto 488 bzw. Atto 647N verbunden) als *primer* eingesetzt wurden, um die spezifische Zielsequenz des Fusionsproteins T-x-P ($x = 4, 6, 8$) zu amplifizieren. Einzelstrangbrüche im oberen Strang resultieren als Spaltprodukt mit

grünen Signal, Einzelstrangbrüche im unteren Strang werden als rotes Signal detektiert (Abb. 3.16 A). Das Substrat T-4-P wird innerhalb von drei Stunde nahezu komplett gespalten, wobei sowohl der obere als auch der untere Strang prozessiert werden. Letzterer wird allerdings etwa dreimal schneller gespalten (Abb. 3.16 B und C). Die präferenzielle Spaltung des unteren Strangs ist für das Substrat T-6-P deutlicher zu erkennen. Der unter Strang wird ca. viermal besser gespalten als der obere Strang. In diesem Substrat wird der untere Strang im Vergleich zu den anderen beiden Substraten generell am besten umgesetzt (die Spaltrate ist zweimal höher als für T-4-P und 13-mal höher als für T-8-P). Der obere Strang wird während der gesamten Reaktion hingegen nur zu ~30% gespalten (Abb. 3.16 C). Im Fall des Substrats T-8-P wird ausschließlich der untere Strang gespalten. Die Spalt-Präferenz gegenüber dem oberen Strang beträgt etwa 400. Die Spaltrate ist allerdings deutlich geringer als für die Substrate mit vier und sechs Basenpaaren Abstand zwischen der TALE und HinP1I-Erkennungssequenz. Insgesamt wird der untere Strang während der Reaktion zu ~60% prozessiert.

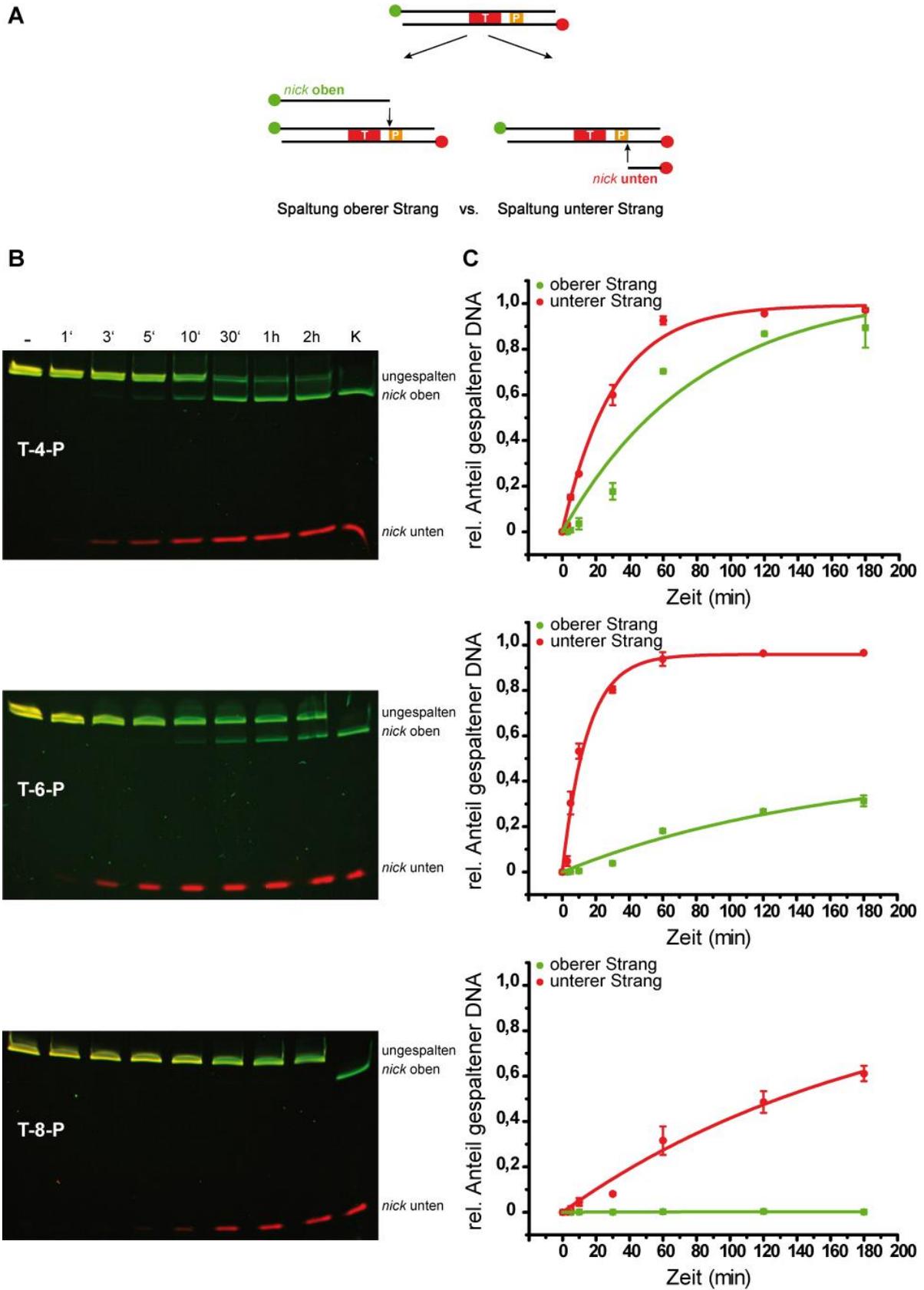


Abbildung 3.16: Analyse der Strangpräferenz von TALE-HinP1I mittels zeitabhängiger Spaltung der adressierten Fluorophor-markierten Polynukleotid-Substrate. In dem Aktivitätstest wurde ein achtfacher

Enzymüberschuss gegenüber dem Substrat eingesetzt (160 nM Protein, 20 nM DNA). Die adressierten Substrate weisen die spezifische AvrBs3-Erkennungssequenz (T) sowie eine nahegelegene HinP1I-Erkennungssequenz (P, GCGC) auf (T-x-P: x = 2, 4, 6). **(A)** Schema der Fluorophor-markierten Polynukleotid-Substrate und der möglichen Spaltprodukte. Erfolgt der Einzelstrangbruch im oberen (unteren) Strang entsteht ein grün (rot) fluoreszierendes Signal. Grün: Atto 488. Rot: Atto 647N. **(B)** Detektion des Fluoreszenzsignals nach zeitabhängiger Spaltung und Analyse über denaturierende Polyacrylamid-Gelelektrophorese. Gelbes Fluoreszenzsignal: Überlagerung von Atto 488 und Atto 647N (ungespaltenes Substrat). Grünes Fluoreszenzsignal: Atto 488 (Einzelstrangbruch im oberen Strang). Rotes Fluoreszenzsignal: Atto 647N (Einzelstrangbruch im unteren Strang). **(C)** Substrate werden durch TALE-HinP1I präferenziell im unteren Strang gespalten. Ermittelte Spaltrate für die Hydrolyse des oberen (unteren) Strangs $1,2 \times 10^{-2} \text{ min}^{-1}$ ($3,3 \times 10^{-2} \text{ min}^{-1}$) für T-4-P, $6,3 \times 10^{-3} \text{ min}^{-1}$ ($6,7 \times 10^{-2} \text{ min}^{-1}$) für T-6-P und $1,2 \times 10^{-5} \text{ min}^{-1}$ ($5,2 \times 10^{-3} \text{ min}^{-1}$) für T-8-P. *nick* = Spaltung.

4 Diskussion

4.1 *Gene targeting* mit hochspezifischen Nukleasen

Die gezielte Manipulation des Genoms durch Verwendung hochspezifischer Nukleasen, die in einem DNA-Bereich kontrolliert spalten und sich damit den zelleigenen Reparaturmechanismus zunutze machen, um mit Hilfe einer DNA-Reparatur-Matrix eben diesen Bereich zu verändern, eröffnet enorme Möglichkeiten für verschiedenste Anwendungen. Nicht zuletzt deswegen wurde dieser Ansatz von der Zeitschrift Nature zur Methode des Jahres 2011 „gewählt“ (2012). Die Anwendungen erstrecken sich von der Untersuchung von Modell-Organismen, die der Analyse der Funktion bestimmter Gene dienen, über die „Manipulation“ von ökonomisch wichtigen Kulturpflanzen, bis hin zu gentherapeutische Ansätzen in menschlichen Zelllinien (Carroll 2014). Eine absolut sichere Anwendung kann jedoch nach wie vor nicht garantiert werden und stellt immer noch eine wissenschaftliche Herausforderung dar. Um eine singuläre Sequenz in einem komplexen Genom zu erkennen und hydrolytisch zu spalten, reicht es nicht aus, die Erkennungssequenz mit einer hochspezifischen Nuklease zu adressieren, vielmehr muss sichergestellt werden, dass es keine Spaltung von nicht adressierten DNA-Sequenzen gibt (*off-site cleavage*). Verschiedenste Studien haben gezeigt, dass hochspezifische Nukleasen, die durchaus für eine singuläre Spaltung in einem komplexen Genom eine ausreichend lange Erkennungssequenz haben, trotzdem toxische Effekte durch unspezifische Spaltung aufweisen (Gabriel *et al.* 2011, Pattanayak *et al.* 2011, Cradick *et al.* 2013, Fu *et al.* 2013). Ursachen könnten hierfür die Funktion des DNA-Bindungsmoduls sein. Dabei kann es eine Rolle spielen, dass das Protein zwangsläufig auch unspezifische oder sequenzähnliche DNA-Bereiche binden muss, um die eigentliche Erkennungssequenz zu finden. Das ermöglicht jedoch auch eine unspezifische Katalyse durch das mit dem DNA-Bindungsmodul gekoppelte DNA-Spaltungsmodul. Hinzu kommt die Problematik der heterodimeren hochspezifischen Nukleasen, die erst durch Kombination zweier spezifischer DNA-Bindungsmodule, die mit einem unspezifischen DNA-Spaltungsmodul verbunden sind, eine DNA spezifisch spalten. Abgesehen von der gewünschten Heterodimerisierung, treten aufgrund der unabhängigen DNA-Bindungsmodule Spezifitäts-Probleme auf, wie sie bildhaft durch die *hanging monkey*-Situation und Homo-Dimerisierung wiedergegeben wird (Abb. 1.6). Diese Art von Unspezifität kann durch monomere hochspezifische Nukleasen vermieden werden. Zudem bieten monomere hochspezifische Nukleasen den Vorteil, dass weniger biologisches Material für eine *in vivo* Anwendung in die Zelle eingebracht werden

muss. Eine weitere Ursache für unspezifische Effekte stellen die systemimmanenten sequenzunspezifische DNA-Spaltungsmodule dar (z.B. die unspezifische DNA-Spaltungsdomäne von FokI), wobei die Spezifität durch Nutzung sequenzspezifischer DNA-Spaltungsmodule verbessert werden kann, da die Aktivität auf eine bestimmte Sequenz beschränkt ist (Schierling *et al.* 2012, Yanik *et al.* 2013). Die Bindungsaffinität des DNA-Spaltungsmoduls kann zudem durch gezielte Aminosäure-Substitutionen soweit abgeschwächt werden, dass die Nuklease die eigene Erkennungssequenz möglichst nur noch bindet und spaltet, wenn es durch das gekoppelte DNA-Bindungsmodul in räumliche Nähe zur Spaltstelle gebracht wird. Dies ist im Fall von MutH natürlicherweise durch die Kopplung im MutS-MutL-System gegeben und bedarf keiner artifiziellen Modifikation. MutH ist unter physiologischen Bedingungen nur aktiv, wenn es im Zuge einer Basenfehlpaarung durch MutL zu seiner Erkennungssequenz rekrutiert wird (Modrich 1991). Das DNA-Bindungsmodul soll im Fusionskonstrukt somit die natürliche Funktion von MutL ersetzen. Eine weitere Erkenntnis, die die Anwendung des *gene targeting* mit hochspezifischen Nukleasen sicherer macht, ist dass homologe Rekombination (HR), die durch einen Einzelstrangbruch stimuliert wird, weniger zytotoxische Effekte zur Folge hat als die, die auf einen Doppelstrangbruch folgen (Davis & Maizels 2011, Metzger *et al.* 2011). Das beruht auf der Tatsache, dass unspezifische Einzelstrangbrüche entweder durch HR prozessiert oder fehlerfrei durch Ligation repariert werden. Im Gegenzug werden Doppelstrangbrüche, die nicht durch HR repariert werden, durch den alternativen Reparatur-Mechanismus NHEJ (*non-homologous end-joining*) prozessiert, der zu unvorhergesehenen Nukleotid-Insertionen und Deletionen (*indels*) an der DNA-Bruchstelle führt. Diese *indels* führen jedoch meist zur Gen-Zerstörung und sind daher toxisch für die Zelle. Durch die Verwendung hochspezifischer Nickasen, die die DNA nur in einem Strang spalten, kann das NHEJ umgangen und die Sicherheit der Anwendung erhöht werden.

4.2 Untersuchung der hochspezifischen Nickase MutH-I-SceI

Die Fusion der katalytisch inaktiven Variante von I-SceI mit der monomeren Endonuklease MutH ermöglicht die zielgerichtete Spaltung definierter GATC-Sequenzen. Es konnte gezeigt werden, dass MutH-I-SceI ausschließlich GATC-Sequenzen hydrolysiert, die in einem gewissen Abstand zu der Erkennungssequenz von I-SceI auftreten. Fehlt die Erkennungssequenz von I-SceI oder ist der Abstand zwischen I-SceI-Erkennungssequenz und GATC-Sequenz zu groß, findet keine katalytische Aktivität statt. MutH ist ein wesentlicher Bestandteil des DNA-*mismatch repair* Systems und benötigt im Zuge einer Basenfehlpaarung

die aktive Rekrutierung durch MutS und MutL zu der eigenen Erkennungssequenz (GATC), um unter physiologischen Bedingungen katalytisch aktiv zu sein (Modrich 1991). Hier konnte gezeigt werden, dass diese aktive Rekrutierung von MutH auch durch die Fusion an ein DNA-Bindungsmodul gewährleistet werden kann. Die eigens schwache katalytische Aktivität von MutH bietet einen enormen Vorteil bei dem Einsatz als DNA-Spaltungsmodul in hochspezifischen Nukleasen. Dadurch wird das Risiko unspezifischer Spaltung reduziert, weil die Prozessierung der DNA hauptsächlich durch das DNA-Bindungsmodul vermittelt wird. Im Fall der Fusion mit MutH ist die katalytische Spaltung unter physiologischen Bedingungen streng mit der Aktivität des DNA-Bindungsmoduls gekoppelt. Werden beide Fusionskomponenten räumlich voneinander getrennt, weist MutH keinerlei eigene katalytische Aktivität mehr auf. Vielmehr noch bietet die Verwendung von MutH als DNA-Spaltungsmodul eine sichere Anwendung, da das Fusionskonstrukt seine Zielsequenz durch Einzelstrangbrüche prozessiert (siehe z.B. Spaltung des unteren Strangs im Substrat T-3-H). In den vergangenen Jahren wurde durch unterschiedliche Studien mehrfach bestätigt, dass die homologe Rekombination, die bei der Insertion und Korrektur von Genomsequenzen genutzt wird, ebenfalls durch Einzelstrangbrüche stimuliert werden kann und diese Strategie gegenüber Doppelstrangbrüchen Vorteile bringt, da der alternative und weniger akkurate Reparaturweg NHEJ (*non-homologous end-joining*) nicht greifen kann (McConnell Smith *et al.* 2009, Davis & Maizels 2011, Metzger *et al.* 2011, Kim *et al.* 2012, Ramirez *et al.* 2012, Wang *et al.* 2012, Cong *et al.* 2013, Mali *et al.* 2013, Wu *et al.* 2014). Somit könnte mit der hochspezifischen Nickase MutH-I-SceI potentielle Gen-Mutationen und -Inaktivierungen bei der Anwendung vermieden werden. Die Verwendung der katalytisch inaktiven I-SceI-Variante als DNA-Bindungsmodul konnte bereits erfolgreich gezeigt werden (Fonfara *et al.* 2012). Basierend auf dieser Studie diene das Fusionskonstrukt MutH-I-SceI im Rahmen dieser Arbeit als *proof of principle*, um die Eigenschaften des DNA-Spaltungsmoduls MutH zu analysieren. Trotz der ausgeprägten Vorteile des DNA-Bindungsmoduls I-SceI, wie erweiterter Erkennungssequenz (18 bp) sowie effektive und spezifische DNA-Bindungs-Eigenschaften, bleibt die unüberwindbare Problematik der Programmierbarkeit auf eine beliebige Erkennungssequenz bestehen (Grizot *et al.* 2009). Auch wenn die beliebige Adressierung spezifischer Sequenzen im Genom für einige Anwendungen unumgänglich ist, kann MutH-I-SceI trotzdem verwendet werden, um sogenannte *safe harbor*-Stellen im Genom für die Insertion bestimmter DNA-Sequenzen zu nutzen (D'Halluin *et al.* 2008). Einen *safe harbor* kann die eingebrachte

Erkennungssequenz in einem unkritischen Genombereich darstellen, der zu (multiplen) Gen-Insertion genutzt werden kann.

4.3 Untersuchung der hochspezifischen Nickase TALE-MutH

Nach anfänglicher Charakterisierung des DNA-Spaltungsmoduls MutH im Fusionskonstrukt MutH-I-SceI wurde im Folgeschritt das DNA-Bindungsmodul I-SceI gegen das TALE-Protein AvrBs4 ersetzt. TALE-Proteine sind in den vergangenen fünf Jahren im Feld der hochspezifischen Nukleasen stark in den Fokus gerückt. Das entscheidende Merkmal dieser Proteingruppe ist, dass sie sehr simpel und effektiv auf eine beliebige Zielsequenz hin programmiert werden können (Boch *et al.* 2009, Morbitzer *et al.* 2010), was für *gene targeting*-Anwendungen essenziell ist. Für das Fusionskonstrukt TALE-MutH konnten vergleichbare Ergebnisse zu MutH-I-SceI erzielt werden (Kapitel 4.2). Zum einen zeigt TALE-MutH eine exklusive Spaltung der adressierten Substrate, d.h. dass die MutH-Erkennungssequenz durch die daneben liegende TALE-Erkennungssequenz adressiert wird und MutH lediglich katalytische Aktivität aufweist, wenn es aktiv durch das TALE-Protein an seinen Erkennungssequenz rekrutiert wird. Zum anderen findet die Prozessierung der Zielsequenz in Form von Einzelstrangbrüchen statt. Um die Aktivität des Fusionskonstrukts ausgiebiger zu untersuchen, wurden für TALE-MutH verschieden adressierte Substrate hergestellt, die im Basenpaarabstand zwischen TALE-Erkennungssequenz und MutH-Erkennungssequenz variieren. Insgesamt wurden alle Abstände zwischen einem und neun Basenpaar(en) untersucht. Die Optimierung der Substrate wird häufig bei der Charakterisierung artifizieller hochspezifischer Nukleasen durchgeführt, da bekannt ist, dass es einen enormen Einfluss auf die Aktivität haben kann (Bibikova *et al.* 2001, Eisenschmidt *et al.* 2005, Yanik *et al.* 2013). Dieser Effekt konnte auch in diesem Fall beobachtet werden. Das Ergebnis zeigt, dass im Fall von TALE-MutH das Substrat mit drei Basenpaaren Abstand zwischen den Erkennungssequenzen am schnellsten umgesetzt wurde, gefolgt vom Substrat mit sechs Basenpaaren Abstand. Interessanterweise sinkt die Spaltrate jenseits dieser beiden Abstände deutlich ab, so dass sich die Spaltrate für T-3-H und T-6-H als deutliche Maxima absetzten (Abb. 3.7). Die Bestimmung möglicher Strangpräferenzen für diese beiden Substrate hat ergeben, dass sich das Fusionskonstrukt im Spaltmechanismus für die Substrate T-3-H und T-6-H unterscheiden. Das adressierte Substrat mit drei Basenpaaren Abstand zwischen den Erkennungssequenzen wird im unteren Strang gespalten, wohingegen das adressierte Substrat mit sechs Basenpaaren Abstand im oberen Strang hydrolysiert wird. Bei der Ermittlung der

Strangpräferenzen für die restlichen adressierten Substrate wurde gezeigt, dass die Strangpräferenz zwischen vier und fünf Basenpaaren Abstand wechselt. Der Effekt des Strangpräferenz-Wechsels könnte auf die Konformation des Fusionskonstrukts zurückzuführen sein. Die Architektur des Substrates setzt voraus, dass das DNA-Bindungsmodul seine spezifische Sequenz in 5'-3' Richtung bindet, und zwar in der Orientierung vom N- zum C-Terminus des TALE-Proteins (Abb. 4.1). Die Erkennungssequenz von MutH ist ein Palindrom und die Definition des unteren und oberen Strangs wird durch die asymmetrische Erkennungssequenz des TALE-Proteins definiert. Die hydrolytische Spaltung der GATC-Sequenz erfolgt vor dem Guanin, wobei MutH sich vom N- zum C-Terminus zur 5'-3' Richtung der DNA orientiert (Abb. 4.1). Da der Plasmid-Spalttest ergeben hat, dass das adressierte Substrat im unteren Strang gespalten wird, wenn die Erkennungssequenz mit drei Basenpaaren Abstand hinter die TALE-Erkennungssequenz geschaltet ist, muss das DNA-Spaltungsmodul bzw. die Linker-Region (Bereich zwischen dem letztem halben *repeat* von AvrBe4 und MutH) in der Konformation des Fusionsproteins umgelagert sein (Abb. 4.1 A). Im Gegenteil dazu wird das adressierte Substrat mit sechs Basenpaaren Abstand zwischen der Erkennungssequenz von TALE und MutH im oberen Strang gespalten. Da der Abstand in diesem Fall größer ist, ist die GATC-Sequenz im unteren Strang durch eine Umlagerung nicht mehr zugänglich. Allerdings ermöglicht diese Substrat-Architektur den Zugang an die obere GATC-Sequenz. Dabei ändert sich die Linker-Region wahrscheinlich in eine eher ausgestreckte Konformation (Abb. 4.1 B). Jedoch bedingt diese Konformation eine Verringerung der Spaltaktivität und ist daher eine Alternative, hat jedoch nicht die besten Voraussetzungen für eine effektive Spaltung

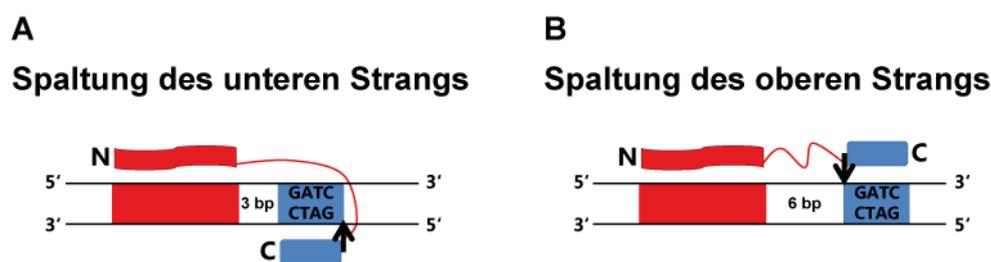


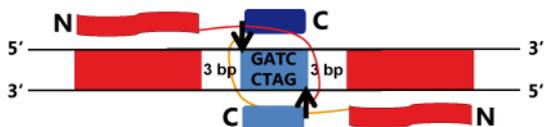
Abbildung 4.1: Hypothetisches Modell zur Strangspezifität von TALE-MutH. TALE (rot) bindet die Erkennungssequenz in 5'-3'-Richtung in der Orientierung vom N- zum C-Terminus. MutH (blau) ist bei der Bindung der 5'-GATC-3'-Sequenz ebenfalls vom N-zum C-Terminus orientiert. (A) Spaltung des unteren Strangs im adressierten Substrat T-3-H. Die Linker-Region lagert sich für die Zugänglichkeit des unteren Strangs um. (B) Spaltung des oberen Strangs im adressierten Substrat T-6-H. Die Linker-Region weist eine gestreckte Konformation auf. Die hydrolytische Spaltung am 5' zum Guanin ist durch einen Pfeil angedeutet.

4.3.1 Doppelstrang-Spaltung mit TALE-MutH

Die Architektur einer tripartiten Erkennungssequenz sollte dazu dienen, um zu untersuchen, ob eine Doppelstrang-Spaltung mit TALE-MutH möglich ist. Basierend auf den vorherigen DNA-Spalttests müsste die tripartite Sequenz durch zwei unabhängige Einzelstrangbrüche in einem Doppelstrangbruch resultieren. Entgegen der Erwartung konnte jedoch keinerlei Spaltung auf dem Substrat T-3-H-3-T beobachtet werden. Ist eine weitere GATC-Sequenz in das tripartite Substrat geschaltet (T-3-HH-3-T), findet ein geringfügiger Umsatz statt. Ein möglicher Grund für dieses Ergebnis ist, dass sich die Fusionsproteine gegenseitig bei der Funktion behindern. Wie bereits oben spekuliert wurde, lagert sich die Linker-Region von TALE-MutH bei der Spaltung des unteren Strangs des Substrats T-3-H um. Diese Umlagerung von beiden Seiten könnte zu sterischen Hindernissen führen, die die katalytische Aktivität inhibiert (Abb. 4.2 A). Die Vergrößerung des Abstands zwischen den Fusionsproteinen scheint die sterische Blockierung etwas aufzulockern (Abb. 4.2 B), was die Aktivität begünstigt. Allerdings findet trotzdem keine Doppelstrangspaltung statt. Aber auch im Fall einer sterischen Inhibition der Aktivität, stellt sich die Frage, wieso die Einzelstrangbrüche nicht zeitlich unabhängig stattfinden können. Hierbei ist zu beachten, dass im Plasmid-Spalttest ein achtfacher Enzymüberschuss gegenüber dem Substrat eingesetzt wird. Das könnte dazu führen, dass die tripartite Erkennungssequenz dauerhaft belegt ist und daher der Enzym-Substrat Komplex kaum zerfällt. Um diesen Effekt zu überprüfen, müsste die Untersuchung unter Bedingungen wiederholt werden, die sich einem stöchiometrischen Verhältnis von Enzym und Substrat annähern. Aber auch wenn der Test bei Enzymüberschuss durchgeführt wurde, deutet das Ergebnis nichtsdestotrotz auf eine starke Protein-DNA-Interaktion hin, die aufgrund der relativ schwachen Aktivität des Fusionskonstrukts nicht unbedingt zu vermuten ist. Eine weitere Möglichkeit zu untersuchen, ob es möglich ist Doppelstrangbrüche mit TALE-MutH zu katalysieren, wäre es, das tripartite Substrat mit sechs Basenpaar Abstand zwischen den TALE-Erkennungssequenzen und der GATC-Sequenz herzustellen. Basierend auf den vorhergehenden Ergebnissen stellen sterische Hindernisse mit diesem Substrat vielleicht weniger ein Problem dar. Das Substrat mit sechs Basenpaaren Abstand wird zwar nicht so effektiv gespalten wie das mit drei Basenpaaren Abstand, jedoch erfolgt bei der Prozessierung möglicherweise keine Umlagerung der Linker-Region, was die Funktion von beiden Seiten ermöglicht und in einem Doppelstrangbruch resultiert (Abb. 4.2 C).

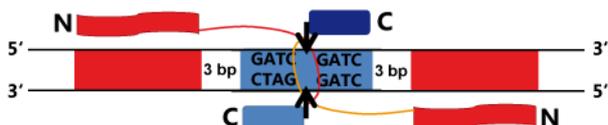
A

T-3-H-3-T



B

T-3-HH-3-T



C

T-6-H-6-H

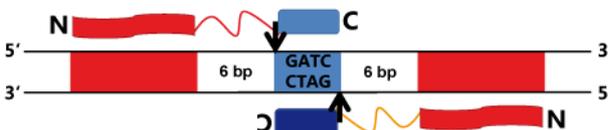


Abbildung 4.2: Hypothetisches Modell zur Spaltung einer tripartiten Erkennungssequenz. TALE (rot) bindet die Erkennungssequenz in 5'-3' Richtung in der Orientierung vom N- zum C-Terminus. MutH (blau) ist bei der Bindung der 5'-GATC-3'-Sequenz ebenfalls vom N-zum C-Terminus orientiert. (A) Spaltung des Substrats T-3-H-3-T. Die Umlagerung der Linker-Region führt zu sterischen Hindernissen (B) Spaltung des Substrats T-3-HH-3-T. Die räumliche Erweiterung der tripartiten Erkennungssequenz durch eine zusätzliche GATC-Sequenz vermindert das sterische Problem geringfügig. (C) Spaltung des Substrats T-6-H-6-T. Die Linker-Region weist eine gestreckte Konformation auf, was eventuell keine sterische Hindernisse zur Folge hat. Die hydrolytische Spaltung am 5' zum Guanin ist durch einen Pfeil angedeutet.

4.3.2 Linker-Optimierung von TALE-MutH

Die Optimierung der Funktion von TALE-MutH wurde neben der Ermittlung des optimalen Substrats auf DNA-Ebene ebenfalls auf Protein-Ebene durch Variation des Peptidlinkers zwischen TALE und MutH durchgeführt. Die vollständige Eliminierung der Linker-Region zwischen dem letzten halben *repeat* des TALE-Proteins und MutH hat zur Folge, dass sich die Maxima der Spaltrate für die adressierten Substrate gegenüber der Linker-Variante mit 28 Aminosäure nach dem letzten halben *repeat* (TALE(28)-MutH) nach vorne verschieben. TALE-L0-MutH bevorzugt die Spaltung der Substrate mit zwei und fünf Basenpaaren Abstand zwischen den Erkennungssequenzen, TALE(28)-MutH präferiert dagegen drei und sechs Basenpaare Abstand. Es ist denkbar, dass die Verkürzung des Abstands zwischen TALE und MutH im Fusionskonstrukt die generelle Verschiebung der Spaltpräferenzen zu kürzeren Basenpaarabständen bedingt. Allerdings setzen sich die Maxima der Spaltraten gegenüber den anderen adressierten Substraten nicht mehr so deutlich ab wie bei TALE(28)-MutH. Auch ist die Aktivität insgesamt verringert. Zwar wurde eine mögliche Strangpräferenz für die Linker-Varianten nicht bestimmt, es wäre jedoch nicht unwahrscheinlich, dass die beiden Maxima auch hier aufgrund unterschiedlicher Strangpräferenzen auftreten. Interessanterweise konnte dieses Ergebnis nicht im Fall der Linker-Variante TALE-L16-MutH beobachtet werden. Diese

Variante hat ebenfalls einen verkürzten Peptidlinker gegenüber TALE(28)-MutH, aber die Spaltpräferenz liegt hier deutlich im Bereich der kürzeren Abstände zwischen der Erkennungssequenz von TALE und MutH. Abgesehen von T-2-H und T-3-H werden die restlichen Substrate nur geringfügig umgesetzt. Wie für TALE-L0-MutH ist die Spalteffizienz im Vergleich zu TALE(28)-MutH niedriger. Es stellt sich die Frage, ob auch diese Variante einen Wechsel der Strangpräferenz aufweist, da nur ein Maximum der Spaltrate beobachtet wurde. Diese Variante verdeutlicht, dass die Optimierung des Peptidlinkers nicht so trivial ist, wie das Design unterschiedlicher adressierter Substrate. Allgemein verringert sich die Flexibilität des Linkers mit abnehmender Länge, allerdings spielt auch die Zusammensetzung der Aminosäuren eine entscheidende jedoch kaum vorhersehbare Rolle für die Konformation des Proteins. Die längste Linker-Variante TALE(28)-L16-MutH zeigt die schlechteste Spaltaktivität unter allen untersuchten Varianten. Hier fällt auf, dass es keine eindeutige Präferenz für eines der untersuchten Substrate gibt. Möglicherweise erlaubt der lange Peptide-Linker eine erhöhte Flexibilität des Fusionskonstrukts, was wiederum alle Substrate zugänglich macht. Auch hier wäre ein Wechsel der Strängpräferenz denkbar. Zusammenfassend konnte festgestellt werden, dass die ursprüngliche Variante TALE(28)-MutH bereits die beste Variante unter allen untersuchten Fusionskonstrukten ist und der Peptidlinker, wie auch anderen Studien zeigen (Handel *et al.* 2009, Fonfara *et al.* 2012, Yanik *et al.* 2013), einen maßgeblichen Einfluss auf die Aktivität des Fusionskonstrukts haben kann.

4.3.3 Steigerung der katalytischen Aktivität des TALE-MutH Fusionskonstrukts

Neben dem Einfluss der Substrat- und Protein-Konstruktion spielen andere Faktoren bei der Ermittlung der Aktivität eine kritische Rolle. Bei allen durchgeführten DNA-Spalttests fällt auf, dass die Aktivität der MutH-Fusion relativ schwach ist. Alle Analysen erfolgten mit Enzymüberschuss, um einen effektiven Umsatz des Substrats zu erhalten. Deutlich wird auch der direkte Vergleich von TALE-MutH und der unmodifizierten Variante TALE-HinPII wt. Dieses Ergebnis kann viele Ursachen haben, wie z.B. die Qualität der Enzym-Präparation. Möglicherweise muss ein Überschuss eingesetzt werden, weil nicht alle Enzyme nach der Reinigung aktiv sind. Es wäre auch denkbar, dass das Protein seine Aktivität unter den Reaktionsbedingungen verliert. Ein anderer Grund könnte die Funktion des DNA-Bindungs- und Spaltungsmoduls im Einzelnen und in der Fusion sein. Dabei spielen kinetische Kriterien wie Bindungs- und Spaltkonstanten eine Rolle. Für das Fusionskonstrukt TALE-MutH stellt

sich die Frage, ob die relativ schwache Aktivität wohlmöglich an einer ineffizienten Bindung des TALE-Proteins oder einer schwachen katalytischen Aktivität durch MutH bedingt wird. Kinetische Untersuchungen des DNA-*mismatch repair* Systems zeigen, dass MutH im Kontext seiner natürlichen Funktion höhere Aktivitäten aufweisen kann (Lee *et al.* 2005). Für einige Restriktionsenzyme konnte die Aktivität durch Verwendung des divalenten Metall-Ions Mn^{2+} statt Mg^{2+} und durch den Zusatz von DMSO gesteigert werden (Hsu & Berg 1978, Malyguine *et al.* 1980, Woodhead *et al.* 1981, Wei *et al.* 2008). Die Anwendung dieser Bedingungen ergab, dass auch die Aktivität von TALE-MutH enorm erhöht werden konnte. In erster Linie zeigt dieses Ergebnis, dass die Aktivität durch die Katalyse und nicht durch die Bindung des TALE-Proteins limitiert wird. Das stimmt ebenfalls mit den Ergebnissen des Konstrukts MutH-I-SceI überein. Die Aktivität beider Konstrukte ist vergleichbar und für I-SceI ist bekannt, dass es die Erkennungssequenz sehr effizient bindet (Perrin *et al.* 1993). Eine ebenfalls sehr effiziente spezifische Bindung des TALE-Proteins würde ebenfalls die Hypothese untermauern, dass dadurch die Spaltung des tripartiten Substrats inhibiert wird. Es ist denkbar, dass die volle katalytische Aktivität von MutH nicht nur durch die bloße Rekrutierung an die spezifische Erkennungssequenz ausgebildet wird, sondern zusätzlich durch andere Faktoren, wie die Interaktion mit MutL, stimuliert wird. Die ermittelte Spaltrate für TALE-MutH entspricht in etwa der Spalteffizienz eines nicht durch das DNA-*mismatch repair* Systems aktivierten MutHs (Lee *et al.* 2005). In Anwesenheit einer Basenfehlpaarungsvermittelten Rekrutierung durch MutS und MutL ist die katalytische Aktivität von MutH gesteigert. Allerdings wurde die Aktivierung von MutH nicht systematisch untersucht und daher können keine Aussagen diesbezüglich getroffen werden. Es wäre möglich die katalytische Aktivität des Fusionskonstrukts TALE-MutH in Anwesenheit von MutL zu ermitteln, um zu untersuchen, ob die Aktivität begünstigt werden kann. Allerdings sei an dieser Stelle hervorgehoben, dass eine schwache hochspezifische Nuklease nicht zwangsläufig von Nachteil ist. Im Gegenteil zeigen stark aktive hochspezifische Nukleasen toxische Effekte (Vergleiche TALE-HinP1I EAD). Die Aktivität sollte jedoch ausreichen, um die Anwendung durchzuführen.

4.3.2 *In vivo* Untersuchung möglicher toxischer Effekte durch TALE-MutH

Neben der Charakterisierung von TALE-MutH *in vitro* wurden mögliche toxische Effekte auch *in vivo* untersucht. Die Analyse erfolgte in Form von Viabilitätstests in *E.coli*. Anhand von

diskontinuierlichen und kontinuierlichen Wachstumskurven wurde ein möglicher Effekt auf das Wachstum unter Induktion des Fusionskonstrukts relativ zu den verwendeten Kontrollen ermittelt. Erfreulicherweise zeigte das bakterielle Wachstum mit TALE-MutH keinen Unterschied zu den Wachstumskurven der Kontrollen mit TALE, MutH und pQE30. Das eigenständige DNA-Bindungs- und Spaltungsmodul sollte theoretisch keinen negativen Einfluss auf das Wachstum haben. Das TALE-Protein besitzt keine Erkennungssequenz im *E.coli* Genom, außerdem weist es keine katalytische Aktivität auf, was potentielle toxische Doppelstrangbrüche zur Folge haben könnte. MutH ist zwar katalytisch aktiv, sollte jedoch unter physiologischen Bedingungen keine Aktivität außerhalb des DNA-*mismatch repair* Systems aufweisen und daher auch keine toxischen Effekte haben. Es wäre denkbar, dass unter extremen Bedingungen, wie Überexpression und kontinuierlichem Wachstum ein negativer Effekt auf die Viabilität auftreten würde, aber im Gegenteil scheint das Wachstum unter MutH-Expression geringfügig begünstigt zu sein. Das leere Expressionsplasmid pQE30 diente als Null-Kontrolle und weist ebenfalls keinen Wachstumsunterschied zu TALE-MutH auf, so dass man aus diesem Experiment schließen kann, dass das Fusionskonstrukt keine toxischen Effekte durch mangelnde Spezifität verursacht. Die verwendeten Negativ-Kontrollen zeigen ein vermindertes Wachstum. Der Effekt für die ZFN ist im Fall der diskontinuierlichen Wachstumskurve kaum festzustellen. Erst der erhöhte Druck durch das dauerhafte exponentielle Wachstum unter kontinuierlichen Bedingungen zeigt einen stetigen Abfall der Viabilität. Wie bereits erwähnt können mit ZFN Spezifitäts-Probleme auftreten, die durch das sequenzspezifischen DNA-Spaltungsmodul FokI bedingt werden. Speziell für dieses Konstrukt konnte in einer Studie bereits festgestellt werden, dass die Spezifität unter Verwendung der katalytischen Domäne von FokI als DNA-Spaltungsmodul gegenüber einem DNA-Spaltungsmodul, das sequenzspezifisch ist, geringer ist (Schierling *et al.* 2012). Folglich kann das reduzierte Zellwachstum auf einen toxischen Effekt durch unzureichende Spezifität der ZFN zurückgeführt werden. Deutlichere toxische Effekte konnten mit der TALE-HinP1I EAD-Variante gezeigt werden. Ein reduziertes Wachstum konnte bereits unter diskontinuierlichen Bedingungen registriert werden. Hier ist das Wachstum bereits zum Zeitpunkt der Induktion gemindert und setzt sich im weiteren Verlauf gegenüber den restlichen Varianten deutlich ab. Der Kurvenverlauf ist eher flach und erreicht während der gesamten Messung keine exponentielle Phase, bevor die Wachstumskurve in die stationäre Phase übergeht. Unter kontinuierlichen Bedingungen befindet sich das Wachstum der TALE-HinP1I EAD-Variante bis zu einer Stunde nach der Induktion auf dem Level der restlichen Kulturen.

Anschließend kommt es nahezu zu einem Stillstand des Wachstums. Erst nach drei Stunden (also ab Stunde sieben) steigt die Kurve wieder ein wenig an. Dieser Effekt lässt sich vielleicht dadurch erklären, dass der durch die HinP1I EAD-Variante verursachte Selektionsdruck nach einer gewissen Zeit eine Subpopulation entstehen lässt, die durch spontane Mutation kein funktionsfähiges TALE-HinP1I EAD produzieren und dadurch weiter wachsen können.

3.4 Untersuchung der hochspezifischen Nuklease TALE-HinP1I

Vor der eigentlichen Charakterisierung der finalen Variante TALE-HinP1I EADA208 war erst einmal die Optimierung der Spezifität notwendig. Der Plasmid-Spalttest ergab deutlich, dass die reine Fusion von HinP1I wt an das TALE-Protein unzureichend war, um die Nuklease an eine definierte Erkennungssequenz zu adressieren. Die gezielte Substitution geeigneter Aminosäuren erhöhte die Spezifität enorm. Dabei lag der Fokus auf Aminosäuren, von denen bekannt ist, dass sie an der Interaktion mit der DNA beteiligt sind (Abb. 4.3). Oft handelt es sich dabei um positiv geladene Aminosäure-Reste, die mit negativ geladenen Phosphatgruppen der DNA in Wechselwirkung treten. Im Fall der ersten Substitution K12E wird die positiv geladene Aminosäure Lysin (K) gegen die negativ geladenen Glutaminsäure (E) ausgetauscht. Dadurch wird die Protein-DNA-Interaktion nicht weiterhin begünstigt, sondern im Gegenteil abgeschwächt, weil es zur Abstoßung der negativen Ladungen kommt. Die gleiche Strategie wurde mit der Substitution R134D verfolgt (positiv geladenes Arginin (R) gegen negativ geladene Glutaminsäure (E)). Bei der Substitution S86A ersetzt man das ungeladene polare Serin (S; Hydroxylgruppe) mit dem ungeladenen aliphatische Alanin (A). Hierbei ist die Wirkung nicht entgegengesetzt, aber begünstigt trotzdem die Schwächung der Protein-DNA-Interaktion, denn das Serin könnte potentiell eine Wasserstoffbrücke zu einem nahegelegenen Nukleotid ausbilden. Auch wenn die TALE-HinP1I EAD-Variante durch die Modifikation um ein Vielfaches spezifischer ist, tritt auf dem Kontroll-Substrat weiterhin unspezifische Spaltung auf, die für eine Anwendung nicht akzeptabel ist. Der toxische Effekt spiegelt sich ebenfalls im *E. coli*-Viabilitätstest durch ein vermindertes Wachstum gegenüber den Kontrollansätzen wieder. Vier weitere Substitutionen wurden jeweils im Kontext der TALE-HinP1I EAD-Variante untersucht. Das Ziel aller Substitutionen ist es, die Bindung von HinP1I soweit abzuschwächen, dass es erst katalytisch aktiv wird, wenn es durch das TALE-Protein in die Nähe der Erkennungssequenz rekrutiert wird und nicht in der Lage ist, unadressierte GCGC-Sequenzen zu binden und zu spalten. Dabei muss allerdings trotzdem eine gewisse Bindungsaffinität bestehen bleiben, damit die Katalyse erfolgen kann. Die nachfolgende

Substitution sollte die Bindungsaffinität daher nicht mehr drastisch abschwächen, wie im Fall der Substitutionen S8A und S208A (analog zu S68A). Die Aminosäuren H97 und L235 bilden keinen direkten Kontakt zu einer Phosphatgruppe aus, befinden sich jedoch in unmittelbarer Nähe des DNA-Rückgrats. Hier wurde das polare Histidin ebenfalls gegen ein Alanin (H97A) substituiert und das unpolare aliphatische Leucin gegen die negative Asparaginsäure (L235D). Die Varianten mit der zusätzlichen Substitution TALE-HinP1I EADA8 und TALE-HinP1I EADA97 zeigten keine deutliche Verbesserung gegenüber der Tripel-Variante TALE-HinP1I EAD. Dies ist ein Beispiel dafür, dass nicht zwangsläufig jede Substitution einen Effekt haben muss, auch wenn die Aminosäuren mit Bedacht ausgewählt wurden. Hingegen reicht die zusätzliche Substitution der Varianten TALE-HinP1I EADA208 und TALE-HinP1I EADD235, um die Spezifität soweit zu erhöhen, dass keinerlei Doppelstrangbrüche selbst bei Enzymüberschuss und nach Langzeitinkubation zu detektieren sind. Allerdings konnten unspezifische Einzelstrangbrüche verzeichnet werden. Primär stellen unspezifische Einzelstrangbrüche jedoch kein Problem für die Genom-Stabilität dar. Einzelstrangbrüche sind die häufigste Form der DNA-Schäden (~10000/Tag/Zelle) und werden z.B. durch oxidativen Stress oder ionische Strahlung verursacht und sind sogar Intermediate im zelleigenen Metabolismus (Davis & Maizels 2014). Die Zelle ist in der Lage, Einzelstrangbrüche effizient, fehlerfrei und daher ohne negative Folgen zu prozessieren.

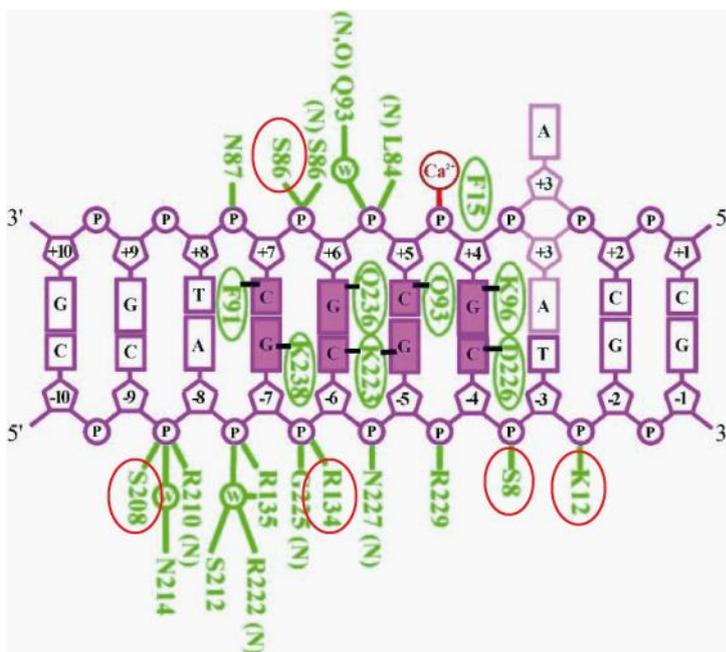


Abbildung 4.3: Darstellung der HinP1I-DNA-Interaktion. Gezeigt sind die Aminosäuren, die direkt oder indirekt an der DNA-Interaktion beteiligt sind (grün). Die DNA ist in violett dargestellt, wobei die Erkennungssequenz von HinP1I (GCGC) hervorgehoben ist. Fünf der sieben Aminosäuren, die in den TALE-HinP1I-Varianten substituiert wurden, sind rot umrandet. Geändert nach Horten *et al.* 2006.

4.4.1 Aktivitäts-Analyse der Variante TALE-HinP1I EADA208

Da die Variante TALE-HinP1I EADA208 die beste Spezifität gegenüber dem unadressierten Substrat zeigte, wurde sie im Weiteren ausführlicher charakterisiert. Die Plasmid-Spalttests mit den adressierten Substraten haben ergeben, dass kurze und lange Basenpaarabstände zwischen den Erkennungssequenzen von TALE und HinP1I sich nicht eignen, da sie nahezu nicht gespalten werden. Abstände von vier, sechs und acht Basenpaaren werden dagegen unter den Reaktionsbedingungen effektiv umgesetzt. Interessanterweise variiert das Spaltpattern stark mit dem verwendeten Substrat. Je größer der Abstand umso stärker ist die Tendenz, dass das Substrat durch einen Einzelstrangbruch prozessiert wird. Ausgehend von T-4-P, wo das Produkt überwiegend linear vorliegt, wird mit T-8-P lediglich Einzelstrangspaltung detektiert. Untersuchungen zu HinP1I haben gezeigt, dass das Enzym als Monomer in der Lösung vorliegt und auch das Spaltpattern dem eines funktionellen Monomers entspricht (Yang *et al.* 2005, Horton *et al.* 2006). Auch in diesem Fall zeigt sich das Spaltpattern eines funktionellen Monomers. Selbst das adressierte Plasmidsubstrat mit vier Basenpaaren Abstand, was zum größten Teil linearisiert wird, akkumuliert zuvor deutlich die *open circular*-Form. Das bedeutet, dass das Plasmid erst in einem Strang und zeitlich versetzt im gegenüberliegenden Strang durch einen Einzelstrangbruch hydrolysiert wird. Es wurde über die Möglichkeit diskutiert, dass HinP1I trotz diesen Spaltpatterns ein funktionelles Dimer sein könnte, wobei die Katalyse in einem Strang schneller erfolgt als in dem Gegenüberliegenden und dadurch das Plasmidsubstrat ebenfalls in der *oc*-Form akkumulieren würde. Diese Hypothese kann mit diesem Ergebnis ausgeschlossen werden. Vielmehr liegt die Vermutung nahe, dass ein Doppelstrangbruch durch ein und dasselbe Monomer und nicht durch zwei unabhängige Monomere hydrolysiert wird. Um Letzteres jedoch endgültig auszuschließen, müsste man die Spaltung unter äquimolaren Bedingungen untersuchen oder die Bindungskonstante des Fusionskonstrukts bestimmen und diese mit den Spaltkonstanten vergleichen. Sollte sich das Spaltpattern bei äquimolaren Enzym-Substrat-Verhältnissen nicht ändern bzw. die Prozessierung des Doppelstrangbruchs schneller verlaufen als das Enzym von der DNA dissoziiert, würde das ein Indiz dafür sein, dass der Doppelstrangbruch durch ein und dasselbe HinP1I-Monomer katalysiert wird. Der Plasmid-Spalttest sowie die Anpassung der Umsatzkurven haben ebenfalls ergeben, dass die Aktivität sich während der Reaktion stetig verlangsamt und nahezu vollständig zum Erliegen kommt.

4.4.2 Bestimmung der Halbwertszeit von TALE-HinP1I EADA208

Um die Halbwertszeit zu bestimmen, d.h. die Zeit in der das Enzym unter Bedingungen der Spaltung aktiv bleibt, wurde der gesamte Ansatz unter Reaktionsbedingungen aber unter Ausschluss von Mg^{2+} für verschieden Zeitintervalle inkubiert und anschließend die Aktivität getestet. Die Spaltrate des Enzyms sinkt mit steigender Inkubationszeit. Die Untersuchung der Variante TALE-HinP1I EADD235 zeigt, dass dieser Effekt nicht ausschließlich auf dem untersuchten Konstrukt beruht. Die ermittelte Halbwertszeit für TALE-HinP1I EADA208 beträgt etwa 20 Minuten. Das widerspricht jedoch der Anpassung der Umsatzraten, bei dem ein Wert von etwa 5 Minuten angenommen werden musste. Das könnte zur Ursache haben, dass die Dissoziationskonstante für das *supercoiled*- und *open circular*-Substrat von ~ 10 nM für die Anpassung falsch angenommen wurde (dieser Wert wurde anhand von K_D -Werten der TALE-Protein-DNA-Interaktion geschätzt; Ergebnisse hier nicht gezeigt) oder die Halbwertszeit wird durch weitere Faktoren negativ beeinflusst, insbesondere auch durch die Anwesenheit von Mg^{2+} . Das könnte überprüft werden, indem der Reaktionsansatz in Abwesenheit von Substrat aber mit Mg^{2+} inkubiert wird. Zudem könnte der Enzym-Produkt-Komplex die Aktivität inhibieren. Um das zu überprüfen könnte nach Ende der Reaktion frisches Enzym zur Reaktion gegeben werden, um zu testen, ob das Substrat zu Ende prozessiert wird.

4.4.3 Strangpräferenz von TALE-HinP1I EADA208

Durch die Wahl des Basenpaarabstandes kann bestimmt werden, ob die DNA durch einen Einzelstrangbruch oder einen Doppelstrangbruch prozessiert wird. TALE-HinP1I EADA208 spaltet alle adressierten Substrate präferenziell im unteren Strang, wobei die Tendenz immer weiter steigt je größer der Abstand wird und letztendlich bei einem Abstand von acht Basenpaaren lediglich im unteren Strang gespalten wird. Die Spaltrate nimmt jedoch mit zunehmendem Basenpaarabstand insgesamt ab. Der Effekt der Strangpräferenz lässt sich höchstwahrscheinlich auf die Konformation des Fusionskonstrukts zurückführen, das die Prozessierung des unteren Strangs durch Positionierung von HinP1I begünstigt, die Spaltung des oberen Strangs jedoch nur z.T. zulässt, und je weiter die HinP1I-Sequenz wegrückt, umso schwieriger bis unmöglich wird es, den oberen Strang zu kontaktieren. Bemerkenswert ist, dass dieser Effekt im Fall von TALE-MutH nicht zu beobachten ist. Der Wechsel der Strangpräferenz von TALE-MutH erfolgt sehr abrupt und die Prozessierung des nicht präferierten Strangs ist nur mit zwei Fluorophor-markierten Substraten zu detektieren (T-5-H und T-7-H). Eine mögliche Erklärung wäre, dass MutH die bereits genickte GATC-Sequenz

wesentlich langsamer spaltet als eine ungespaltene Sequenz (Welsh *et al.* 1987, Au *et al.* 1992). Eine andere Erklärung wäre aber auch, dass HinP1I den Strang bei der Hydrolyse aktiv wechselt, wie es auch für das Restriktionsenzym BcnI bekannt ist (Sasnauskas *et al.* 2011). Das würde die Hypothese zum Spaltmechanismus untermauern, dass ein einziges HinP1I-Monomer den Doppelstrangbruch katalysiert.

4.5 Fazit und Ausblick

Mit allen drei Fusions-Ansätzen konnten erfolgreich hochspezifische Nukleasen generiert werden. Dabei hat sich gezeigt, dass die natürliche Funktion MutH während des DNA-*mismatch repair* Mechanismus durch MutS und MutL an eine GATC-Sequenz zu rekrutieren, damit es einen Einzelstrangbruch hydrolysiert, durch die Fusion an ein DNA-Bindungsmodul artifiziell ersetzt werden kann. Es scheint, dass man MutH mit verschiedenen DNA-Bindungsmodulen fusionieren kann, zumindest zeigt es sowohl mit I-SceI als auch dem TALE-Protein keine wesentlichen Unterschiede in der Funktion. Es werden ausschließlich GATC-Sequenzen gespalten, die sich in unmittelbarer Nähe der TALE-Erkennungssequenz befinden. MutH ist dem entsprechend nur aktiv, wenn es durch das TALE-Protein rekrutiert wird. Nicht adressierte GATC-Sequenzen werden hingegen nicht umgesetzt. Durch die schwache Eigenaktivität von MutH unter physiologischen Bedingungen ist es sehr stark von der Funktion des DNA-Bindungsmoduls abhängig, was für eine konkrete Anwendung von enormen Vorteil ist, da unspezifische Spaltung durch die natürliche Funktion von MutH, nämlich die stringente Rekrutierung durch MutS über MutL, vermieden wird. Hinzu kommt, dass das DNA-Spaltungsmodul MutH lediglich Einzelstrangbrüche katalysiert und somit der inakkurate Reparatur-Mechanismus NHEJ in der Zelle vermieden wird. Das würde die Sicherheit der Anwendung zusätzlich verbessern. Eine spezielle Eigenschaft bietet außerdem die Strangpräferenz, die abhängig vom Basenpaarabstand zwischen der Erkennungssequenz für das DNA-Bindungsmodul und der GATC-Sequenz variiert werden kann. Das könnte Untersuchungen dienen, für die eine Strang-Diskriminierung relevant ist. Die schwache Nukleaseaktivität wie im Fall von MutH konnte im alternativen DNA-Spaltungsmodul HinP1I durch gezielte Aminosäure-Substitutionen artifiziell erzeugt werden. Das Beispiel demonstriert die Notwendigkeit, die katalytische Funktion abhängig von der Zielsequenz zu machen. Anderenfalls kann das DNA-Spaltungsmodul z.T. selbständig agieren, und das führt zu unspezifischer Spaltung. Das Fusionskonstrukt TALE-HinP1I EADA208 konnte dahingehend modifiziert werden, dass es eine adressierte GCGC-Sequenz mit hinreichender Spezifität

prozessiert. Die monomeren Fusionskonstrukte TALE-MutH und TALE-HinP1I EADA208 unterschieden sich jedoch in der Art der DNA-Hydrolyse. TALE-MutH ist eine obligatorische Nickase. TALE-HinP1I EADA208 ist allerdings auch in der Lage Doppelstrangbrüche zu katalysieren. Jedoch hat auch dieses Konstrukt eine Strangpräferenz. Diese liegt im Gegensatz zu TALE-MutH lediglich auf dem unteren Strang und kann durch einen entsprechenden Basenpaarabstand der Erkennungssequenzen zu einem Einzelstrangbruch verändert werden. Die ausgiebige *in vitro*-Charakterisierung der Fusions-Konstrukte würde eine weitere *in vivo*-Untersuchung befürworten. Die verwendeten TALE-Proteine lassen sich einfach und schnell austauschen und würden die robuste Funktion der DNA-Spaltungsmodul wahrscheinlich nicht gravierend beeinflussen. Je nach Art der Anwendung haben beide DNA-Spaltungsmodule vorteilhafte Eigenschaften. Für eine möglichst sicheres und akkurates *gene targeting* durch homologe Rekombination, wie Gen-Insertion und –Korrektur, eignet sich die hochspezifische Nickase TALE-MutH. Zielt der Ansatz jedoch auf eine sichere und effiziente Gen-Inaktivierung, würde sich eher TALE-HinP1I empfehlen. Bei diesem Ansatz wird das NHEJ bewusst ausgenutzt, indem ein Doppelstrangbruch in das Ziel-Gen eingefügt wird, aber keine DNA-Reparatur-Matrix angeboten wird. Die fehlerhafte Reparatur führt somit zu Verschiebung des Leserasters und dadurch zwangsläufig zu Inaktivierung des Gens.

5 Zusammenfassung

Die gezielte Veränderung des genetischen Erbmaterials eröffnet u.a. die Möglichkeit Funktionen bestimmter Gene zu erforschen, ökonomisch wichtige Kulturpflanzen ertragreicher zu machen und ist vor allem auf dem gentherapeutischen Forschungsgebiet in den Fokus gerückt, da es einen enorm wichtigen Fortschritt in der Therapie gewisser Erbkrankheiten bedeuten könnte. Eine methodische Genom-Modifikation kann mit hochspezifischen Nukleasen erreicht werden, die eine singuläre Zielsequenz im Genom spalten. Je nach eingesetzter Nuklease können Doppel- oder Einzelstrangbrüche eingeführt werden. Die Reparatur des DNA-Schadens kann über homologe Rekombination (HR) in Anwesenheit einer homologen DNA-Reparatur-Matrix erfolgen oder alternativ, wenn eine DNA-Reparatur-Matrix fehlt oder nicht genutzt wird, kann ein Doppelstrangbruch den Reparaturmechanismus des *non-homologous end-joining* (NHEJ) stimulieren. NHEJ führt jedoch gewöhnlich zu Insertion und Deletion an der DNA-Bruchstelle, weshalb die Reparatur fehlerbehaftet ist. Daher wird dieser Ansatz für die gezielte Inaktivierung von Genen genutzt. Für Anwendung wie z.B. Gen-Korrektur oder Gen-Insertion, können hingegen Nickasen herangezogen werden, die die HR stimulieren, NHEJ aber effektiv vermeiden und daher das Risiko für unerwünschte Mutationen senken. Im Rahmen dieser Arbeit werden drei innovative hochspezifische Nukleasen generiert und charakterisiert, die auf den DNA-Spaltungsmodulen MutH und HinP1I basieren. MutH und HinP1I sind monomere Endonukleasen, infolgedessen stellen die Fusionskonstrukte funktionale Monomere dar. Im Gegensatz zu dem häufig genutzten DNA-Spaltungsmodul FokI, sind MutH und HinP1I sequenzspezifisch für eine vier Basenpaare lange Erkennungssequenz, somit können zytotoxische Effekte durch unspezifisches Spalten der DNA weiter reduziert werden. Für eine sichere Anwendung ist zusätzlich die Funktion der DNA-Spaltungsmodule streng mit der Funktion des DNA-Bindungsmoduls gekoppelt, da die Nukleaseaktivität von Natur aus schwach ist, wie im Fall von MutH, bzw. bei HinP1I durch gezielte Mutationen vermindert wurde. Es konnte gezeigt werden, dass durch die Fusion der katalytisch inaktiven Variante I-SceI bzw. der DNA-Bindungsdomäne des TALE-Proteins AvrBs4 an MutH eine hochspezifische Nickase generiert werden konnte, die eine adressierte Zielsequenz 1000-mal schneller spaltete als eine unspezifische Sequenz und keinerlei toxischen Effekte in *E. coli* hat. Darüber hinaus weist das Fusionskonstrukt TALE-MutH eine Strangpräferenz auf, die abhängig von dem Basenpaarabstand zwischen TALE- und MutH-Erkennungssequenz ist und daher steuerbar ist. Mit der Fusion der DNA-Bindungsdomäne des TALE-Proteins AvrBs3 und einer abgeschwächten Variante von HinP1I entsteht eine

monomere hochspezifische Nuklease, die sowohl einen Doppelstrang-, als auch einen Einzelstrangbruch prozessieren kann. Hierbei ist ebenfalls der Abstand zwischen den Erkennungssequenz ausschlaggebend für den Spaltmechanismus. Dabei wird die adressierte Zielsequenz im Vergleich zu unspezifischen Sequenzen deutlich präferiert. Folglich wurden in dieser Arbeit zwei hochspezifische Nukleasen entwickelt, die für jede Art der Anwendung genutzt werden können, um komplexe Genome gezielt zu modifizieren.

7 Danksagung

In erster Linie gilt mein Dank meinem Doktorvater Prof. Dr. Alfred Pingoud, sowie meinen wissenschaftlichen „Ziehv Vätern“ Prof. Dr. Peter Friedhoff und Dr. Wolfgang Wende. Ich danke Euch nicht nur für die Bereitstellung dieses tollen Projekts, sondern auch für die hervorragende Betreuung und Eure Hilfe, wenn sie gefordert war. Ping, Sie waren mir immer ein guter Chef gewesen, den ich auch menschlich sehr zu schätzen weiß. Ihr Enthusiasmus ist ansteckend, auch wenn Sie gelegentlich mal über das Ziel hinausschießen ;-). Peter, Deine wissenschaftliche Kreativität kennt keine Grenzen. Ich werde unsere Diskussion, die sich gerne auch mal im Detail verlieren, sehr vermissen. Wolle, du hast wirklich für jedes Problem eine Lösung und Dir wird vollkommen zurecht ein goldenes Händchen nachgesagt! Ich danke euch für eure Unterstützung jeglicher Art. Ihr wart hervorragende Lehrer!!!

In besonderer Weise möchte ich mich bei meinen Eltern und meinem Bruderherz bedanken. Es ist nur schwer in Worte zu fassen, wie wichtig Ihr mir seid! Eure Unterstützung, ob in guten oder in schlechten Phasen, war und ist enorm wichtig für mich. Ich danke Euch für den bedingungslosen Rückhalt und ich bin mehr als froh Euch an meiner Seite zu haben. Ich liebe Euch!

Ein ganz spezieller Dank gilt außerdem meinem ehemaligen Betreuer Dr. Benno Schierling. Ich bin über glücklich, dass du mich damals unter deine Fittiche genommen hast :-). Ich habe meine Ausbildung bei dir sehr genossen. Du bist ein außergewöhnlicher Mensch und ich verdanke dir so viel!!!

Ganz besonders möchte ich auch meiner lieben ehemaligen Kollegin Bedriska Reitz danken! Es hat so viel Spaß gemacht mit dir zusammen zu arbeiten. Ich vermisse die Zeit wirklich sehr. Ohne Dich und Benno wäre all das nicht möglich gewesen. Ich hab Dich lieb!

Liebe Laura Waltl und Sabrina Stiehler, Euch danke ich für einen wirklich amüsanten Arbeitsalltag. Die gemeinsamen lustigen und oft verrückten Mittagessen werde ich sehr vermissen! Auch Euch hab ich sehr lieb!

Außerdem möchte ich mich herzlich bei Ines Fonfara, Roger Heinze, Andreas Marx, Heike Büngen, Ines Winkler, Daniel Kutscher, Mert Yanik und Carolin Menges für Rat und Tat besonders zu meiner Anfangszeit im Labor bedanken! Es ist schön Kollegen wie euch zu haben.

Ich bedanke mich zudem bei Anja Drescher, Ina Stingel und Karina Urbach. Auf Euch ist einfach Verlass, wenn man Euch braucht!

Ein ganz lieber Dank gilt auch Petra Baumgart. Die gute Seele des Instituts, wie der Ping so schön sagt. Ohne Dich würde der Laden einfach nicht laufen!

Ich danke auch meinen lieben Kollegen Ping Ping Li und Fabian Bietz. Ping Ping, du bist einfach eine wundervolle Person. Fabian, du machst den besten Wodka ;-)

Der Arbeitsgruppe von Prof. Dr. Elena Kubareva und Prof. Dr. Tatjana Oretskaya, insbesondere Mayya Monakhova und Liudmila Abrosimova danke ich für einen aufregenden und lehrreichen Laboraustausch in Moskau!

Ich bedanke mich bei Prof. Dr. Renkawitz für die freundliche Übernahme des Zeitgutachtens.

Zum Schluss möchte ich mich ganz herzlichst bei meinen fleißigen Korrektur-Lesern bedanken: mein lieber Kollege Andreas Marx, mein lieber Freund Fabian „The Snake“ Weidenhausen und meinem Bruderherz Wadim Gabsalilow! Ein Lieber Dank geht auch an meinen Übersetzer und besten Freund Yuriy Lis. Natürlich möchte ich auch meine restlichen guten Freunde mit einem Dank würdigen für aufmunternde Wort, motivierende Ansprachen, regelmäßige Erkundigungen über das Wohlbefinden und kraftschöpfende Biere im Proberaum (KomA 13 4-ever!). Ich bedanke mich ebenfalls bei meinen Kollegen des IRTGs. Wir hatten eine geile Zeit! Moskau + St. Petersburg 2013 - never forget! :-D Es hat Spaß gemacht und ich werde Euch mit großer Sicherheit sehr vermissen!

6 Literaturverzeichnis

(2012). Method of the Year 2011. *Nat Methods* 9, 1.

Aryan, A., Anderson, M. A., Myles, K. M. & Adelman, Z. N. (2013). TALEN-Based Gene Disruption in the Dengue Vector *Aedes aegypti*. *PLoS One* 8, e60082.

Au, K. G., Welsh, K. & Modrich, P. (1992). Initiation of methyl-directed mismatch repair. *J. Biol. Chem.* 267, 12142-12148.

Ausubel, F., Brent, R., Kingston, R., Moore, D., Seidmann, J., Smith, J. & Struhl, K. (2012). Current protocols in molecular biology. John Wiley & Sons.

Ban, C. & Yang, W. (1998). Structural basis for MutH activation in E.coli mismatch repair and relationship of MutH to restriction endonucleases. *EMBO J.* 17, 1526-1534.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A. & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709-1712.

Barras, F. & Marinus, M. G. (1989). The great GATC: DNA methylation in E. coli. *Trends Genet.* 5, 139-143.

Bibikova, M., Carroll, D., Segal, D. J., Trautman, J. K., Smith, J., Kim, Y. G. & Chandrasegaran, S. (2001). Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol. Cell. Biol.* 21, 289-297.

Bibikova, M., Golic, M., Golic, K. G. & Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics* 161, 1169-1175.

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, 1509-1512.

Boch, J. & Bonas, U. (2010). Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu. Rev. Phytopathol.* 48, 419-436.

Bonas, U., Stall, R. E. & Staskawicz, B. (1989). Genetic and structural characterization of the avirulence gene *avrBs3* from *Xanthomonas campestris* pv. *vesicatoria*. *Mol. Gen. Genet.* 218, 127-136.

Bowen, L. M. & Dupureur, C. M. (2003). Investigation of restriction enzyme cofactor requirements: a relationship between metal ion properties and sequence specificity. *Biochemistry* 42, 12643-12653.

Briggs, A. W., Rios, X., Chari, R., Yang, L., Zhang, F., Mali, P. & Church, G. M. (2012). Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Res.* 40, e117.

Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., Dickman, M. J., Makarova, K. S., Koonin, E. V. & van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321, 960-964.

Buttner, D. & Bonas, U. (2002). Port of entry--the type III secretion translocon. *Trends Microbiol.* 10, 186-192.

Buttner, D. & Bonas, U. (2010). Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiol. Rev.* 34, 107-133.

Cai, C. Q., Doyon, Y., Ainley, W. M., Miller, J. C., Dekelver, R. C., Moehle, E. A., Rock, J. M., Lee, Y. L., Garrison, R., Schulenberg, L., Blue, R., Worden, A., Baker, L., Faraji, F., Zhang, L., Holmes, M. C., Rebar, E. J., Collingwood, T. N., Rubin-Wilson, B., Gregory, P. D., Urnov, F. D. & Petolino, J. F. (2009). Targeted transgene integration in plant cells using designed zinc finger nucleases. *Plant Mol. Biol.* 69, 699-709.

Carroll, D. (2014). Genome engineering with targetable nucleases. *Annu. Rev. Biochem.* 83, 409-439.

Cathomen, T. & Joung, J. K. (2008). Zinc-finger nucleases: the next generation emerges. *Mol. Ther.* 16, 1200-1207.

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 Res.* 39, e82.

Coligan, J., Dunn, B., Speicher, D. & Wingfield, P. (2012). *Current protocols in protein science*. John Wiley & Sons.

Colleaux, L., d'Auriol, L., Betermier, M., Cottarel, G., Jacquier, A., Galibert, F. & Dujon, B. (1986). Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into *E. coli* as a specific double strand endonuclease. *Cell* 44, 521-533.

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

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*.

Cradick, T. J., Fine, E. J., Antico, C. J. & Bao, G. (2013). CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res.* 41, 9584-9592.

D'Halluin, K., Vanderstraeten, C., Stals, E., Cornelissen, M. & Ruiters, R. (2008). Homologous recombination: a basis for targeted genome optimization in crop species such as maize. *Plant Biotechnol J* 6, 93-102.

Dalgaard, J. Z., Moser, M. J., Hughey, R. & Mian, I. S. (1997). Statistical modeling, phylogenetic analysis and structure prediction of a protein splicing domain common to inteins and hedgehog proteins. *J. Comput. Biol.* 4, 193-214.

Davis, L. & Maizels, N. (2011). DNA nicks promote efficient and safe targeted gene correction. *PLoS One* 6, e23981.

Davis, L. & Maizels, N. (2014). Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proc. Natl. Acad. Sci. U. S. A.* 111, E924-932.

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 Phytol.* 199, 773-786.

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, 720-723.

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 Res.*

Dickinson, D. J., Ward, J. D., Reiner, D. J. & Goldstein, B. (2013). Engineering the *Caenorhabditis elegans* genome using Cas9-triggered homologous recombination. *Nat Methods* 10, 1028-1034.

Doyon, Y., McCammon, J. M., Miller, J. C., Faraji, F., Ngo, C., Katibah, G. E., Amora, R., Hocking, T. D., Zhang, L., Rebar, E. J., Gregory, P. D., Urnov, F. D. & Amacher, S. L. (2008). Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat. Biotechnol.* 26, 702-708.

Dujon, B. (1989). Group I introns as mobile genetic elements: Facts and mechanistic speculations — a review. *Gene* 82, 91-114.

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 Res.* 33, 7039-7047.

Fonfara, I., Curth, U., Pingoud, A. & Wende, W. (2012). Creating highly specific nucleases by fusion of active restriction endonucleases and catalytically inactive homing endonucleases. *Nucleic Acids Res.* 40, 847-860.

Friedhoff, P., Thomas, E. & Pingoud, A. (2003). Tyr212: a key residue involved in strand discrimination by the DNA mismatch repair endonuclease MthH. *J. Mol. Biol.* 325, 285-297.

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. *Nat Biotech* advance online publication.

Gabriel, R., Lombardo, A., Arens, A., Miller, J. C., Genovese, P., Kaeppl, 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. *Nat. Biotechnol.* 29, 816-823.

Gimble, F. S. & Stephens, B. W. (1995). Substitutions in conserved dodecapeptide motifs that uncouple the DNA binding and DNA cleavage activities of *Pi-SceI* endonuclease. *J. Biol. Chem.* 270, 5849-5856.

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

Handel, E. M., Alwin, S. & Cathomen, T. (2009). Expanding or restricting the target site repertoire of zinc-finger nucleases: the inter-domain linker as a major determinant of target site selectivity. *Mol. Ther.* 17, 104-111.

Hockemeyer, D., Wang, H., Kiani, S., Lai, C. S., Gao, Q., Cassady, J. P., Cost, G. J., Zhang, L., Santiago, Y., Miller, J. C., Zeitler, B., Cherone, J. M., Meng, X., Hinkley, S. J., Rebar, E. J., Gregory, P. D., Urnov, F. D. & Jaenisch, R. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. *Nat. Biotechnol.* 29, 731-734.

Horton, J. R., Zhang, X., Maunus, R., Yang, Z., Wilson, G. G., Roberts, R. J. & Cheng, X. (2006). DNA nicking by *HinP1I* endonuclease: bending, base flipping and minor groove expansion. *Nucleic Acids Res.* 34, 939-948.

Hsu, M. & Berg, P. (1978). Altering the specificity of restriction endonuclease: effect of replacing Mg^{2+} with Mn^{2+} . *Biochemistry* 17, 131-138.

Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., Peterson, R. T., Yeh, J. R. & Joung, J. K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.*

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

Karvelis, T., Gasiunas, G., Miksys, A., Barrangou, R., Horvath, P. & Siksnys, V. (2013a). crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*. *RNA Biol* 10.

Karvelis, T., Gasiunas, G. & Siksnys, V. (2013b). Programmable DNA cleavage in vitro by Cas9. *Biochem. Soc. Trans.* 41, 1401-1406.

Kay, S. & Bonas, U. (2009). How *Xanthomonas* type III effectors manipulate the host plant. *Curr. Opin. Microbiol.* 12, 37-43.

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 Res.* 22, 1327-1333.

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

Kolb, A. F., Coates, C. J., Kaminski, J. M., Summers, J. B., Miller, A. D. & Segal, D. J. (2005). Site-directed genome modification: nucleic acid and protein modules for targeted integration and gene correction. *Trends Biotechnol.* 23, 399-406.

Learn, B. A. & Grafstrom, R. H. (1989). Methyl-directed repair of frameshift heteroduplexes in cell extracts from *Escherichia coli*. *J. Bacteriol.* 171, 6473-6481.

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. *Mol. Cell* 20, 155-166.

Lieber, M. R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* 79, 181-211.

Liu, J., Li, C., Yu, Z., Huang, P., Wu, H., Wei, C., Zhu, N., Shen, Y., Chen, Y., Zhang, B., Deng, W. M. & Jiao, R. (2012). Efficient and Specific Modifications of the *Drosophila* Genome by Means of an Easy TALEN Strategy. *J Genet Genomics* 39, 209-215.

Loh, T., Murphy, K. C. & Marinus, M. G. (2001). Mutational analysis of the MutH protein from *Escherichia coli*. *J. Biol. Chem.* 276, 12113-12119.

Ma, S., Zhang, S., Wang, F., Liu, Y., Liu, Y., Xu, H., Liu, C., Lin, Y., Zhao, P. & Xia, Q. (2012). Highly Efficient and Specific Genome Editing in Silkworm Using Custom TALENs. *PLoS ONE* 7, e45035.

Maeder, M. L., Thibodeau-Beganny, S., Osiaik, A., Wright, D. A., Anthony, R. M., Eichinger, M., Jiang, T., Foley, J. E., Winfrey, R. J., Townsend, J. A., Unger-Wallace, E., Sander, J. D., Muller-Lerch, F., Fu, F., Pearlberg, J., Gobel, C., Dassie, J. P., Pruett-Miller, S. M., Porteus, M. H., Sgroi, D. C., Iafrate, A. J., Dobbs, D., McCray, P. B., Jr., Cathomen, T., Voytas, D. F. & Joung, J. K. (2008). Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell* 31, 294-301.

Mak, A. N., Bradley, P., Cernadas, R. A., Bogdanove, A. J. & Stoddard, B. L. (2012). The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* 335, 716-719.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., Dicarlo, J. E., Norville, J. E. & Church, G. M. (2013). RNA-Guided Human Genome Engineering via Cas9. *Science*.

Malyguine, E., Vannier, P. & Yot, P. (1980). Alteration of the specificity of restriction endonucleases in the presence of organic solvents. *Gene* 8, 163-177.

McConnell Smith, A., Takeuchi, R., Pellenz, S., Davis, L., Maizels, N., Monnat, R. J., Jr. & Stoddard, B. L. (2009). Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-Anil LAGLIDADG homing endonuclease. *Proc. Natl. Acad. Sci. U. S. A.* 106, 5099-5104.

Merlin, C., Beaver, L. E., Taylor, O. R., Wolfe, S. A. & Reppert, S. M. (2013). Efficient targeted mutagenesis in the monarch butterfly using zinc-finger nucleases. *Genome Res.* 23, 159-168.

Metzger, M. J., McConnell-Smith, A., Stoddard, B. L. & Miller, A. D. (2011). Single-strand nicks induce homologous recombination with less toxicity than double-strand breaks using an AAV vector template. *Nucleic Acids Res.* 39, 926-935.

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. *Nat. Biotechnol.* 25, 778-785.

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. *Nat. Biotechnol.* 29, 143-148.

Modrich, P. (1991). Mechanisms and biological effects of mismatch repair. *Annu. Rev. Genet.* 25, 229-253.

Morbitzer, R., Romer, P., Boch, J. & Lahaye, T. (2010). Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* 107, 21617-21622.

Morbitzer, R., Elsaesser, J., Hausner, J. & Lahaye, T. (2011). Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Res.* 39, 5790-5799.

Moscou, M. J. & Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science* 326, 1501.

Moure, C. M., Gimble, F. S. & Quioco, F. A. (2003). The crystal structure of the gene targeting homing endonuclease I-SceI reveals the origins of its target site specificity. *J. Mol. Biol.* 334, 685-695.

Moure, C. M., Gimble, F. S. & Quioco, F. A. (2008). Crystal structures of I-SceI complexed to nicked DNA substrates: snapshots of intermediates along the DNA cleavage reaction pathway. *Nucl. Acids Res.*, in press.

Pabo, C. O., Peisach, E. & Grant, R. A. (2001). Design and selection of novel Cys2His2 zinc finger proteins. *Annu. Rev. Biochem.* 70, 313-340.

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 Sci.* 4, 2411-2423.

Parker, B. O. & Marinus, M. G. (1992). Repair of DNA heteroduplexes containing small heterologous sequences in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 89, 1730-1734.

Pattanayak, V., Ramirez, C. L., Joung, J. K. & Liu, D. R. (2011). Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Methods* 8, 765-770.

Pauwels, K., Podevin, N., Breyer, D., Carroll, D. & Herman, P. (2014). Engineering nucleases for gene targeting: safety and regulatory considerations. *New Biotechnology*.

Pavletich, N. P. & Pabo, C. O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 252, 809-817.

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., Orange, J. S., 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. *Nat. Biotechnol.* 26, 808-816.

Perrin, A., Buckle, M. & Dujon, B. (1993). Asymmetrical recognition and activity of the I-SceI endonuclease on its site and on intron-exon junctions. *EMBO J.* 12, 2939-2947.

Pingoud, A., Pingoud, V. & Wende, W. (2003). Homingendonukleasen - Vielversprechende Werkzeuge für das Genomengineering. *BIOspektrum* 5, 592-595.

Pingoud, A., Fuxreiter, M., Pingoud, V. & Wende, W. (2005). Type II restriction endonucleases: structure and mechanism. *Cell. Mol. Life Sci.* 62, 685-707.

Pingoud, A. & Silva, G. H. (2007). Precision genome surgery. *Nat. Biotechnol.* 25, 743-744.

Pingoud, A. & Wende, W. (2011). Generation of novel nucleases with extended specificity by rational and combinatorial strategies. *Chembiochem* 12, 1495-1500.

Pruett-Miller, S. M., Reading, D. W., Porter, S. N. & Porteus, M. H. (2009). Attenuation of zinc finger nuclease toxicity by small-molecule regulation of protein levels. *PLoS Genet.* 5, e1000376.

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 Res.*

Remy, S., Tesson, L., Menoret, S., Usal, C., Scharenberg, A. M. & Anegon, I. (2010). Zinc-finger nucleases: a powerful tool for genetic engineering of animals. *Transgenic Res.* 19, 363-371.

Reyon, D., Tsai, S. Q., Khayter, C., Foden, J. A., Sander, J. D. & Joung, J. K. (2012). FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotechnol.* 30, 460-465.

Roberts, R. J., Vincze, T., Posfai, J. & Macelis, D. (2010). REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* 38, D234-236.

Sadelain, M., Papapetrou, E. P. & Bushman, F. D. (2012). Safe harbours for the integration of new DNA in the human genome. *Nat. Rev. Cancer* 12, 51-58.

Sakuma, T. & Woltjen, K. (2014). Nuclease-mediated genome editing: At the front-line of functional genomics technology. *Dev. Growth Differ.* 56, 2-13.

Sander, J. D., Dahlborg, E. J., Goodwin, M. J., Cade, L., Zhang, F., Cifuentes, D., Curtin, S. J., Blackburn, J. S., Thibodeau-Beganny, S., Qi, Y., Pierick, C. J., Hoffman, E., Maeder, M. L., Khayter, C., Reyon, D., Dobbs, D., Langenau, D. M., Stupar, R. M., Giraldez, A. J., Voytas, D. F., Peterson, R. T., Yeh, J. R. & Joung, J. K. (2011). Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat. Methods* 8, 67-69.

Sasnauskas, G., Kostiuik, G., Tamulaitis, G. & Siksnys, V. (2011). Target site cleavage by the monomeric restriction enzyme BcnI requires translocation to a random DNA sequence and a switch in enzyme orientation. *Nucleic Acids Res.* 39, 8844-8856.

Schierling, B., Dannemann, N., Gabsalilow, L., Wende, W., Cathomen, T. & Pingoud, A. (2012). A novel zinc-finger nuclease platform with a sequence-specific cleavage module. *Nucleic Acids Res.* 40, 2623-2638.

Schmid-Burgk, J. L., Schmidt, T., Kaiser, V., Honing, K. & Hornung, V. (2012). A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat. Biotechnol.*

Scholze, H. & Boch, J. (2010). TAL effector-DNA specificity. *Virulence* 1, 428-432.

Schreiber, T. & Bonas, U. (2014). Repeat 1 of TAL effectors affects target specificity for the base at position zero. *Nucleic Acids Res.*

Shen, S. X., Li, Q. L., Yan, P. F., Zhou, B., Ye, S. Y., Lu, Y. Y. & Wang, D. B. (1980). Restriction endonucleases from three strains of *Haemophilus influenzae*. *Sci Sin.* 23(11):1435-42.

Smith, J. & Modrich, P. (1996). Mutation detection with MthH, MutL, and MutS mismatch repair proteins. *Proc. Natl. Acad. Sci. U. S. A.* 93, 4374-4379.

Stoddard, B. L. (2005). Homing endonuclease structure and function. *Q. Rev. Biophys.* 38, 49-95.

Stoddard, B. L. (2014). Homing endonucleases from mobile group I introns: discovery to genome engineering. *Mob DNA* 5, 7.

Szcepek, 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. *Nat. Biotechnol.* 25, 786-793.

Terns, M. P. & Terns, R. M. (2011). CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.* 14, 321-327.

Urnov, F. D., Miller, J. C., Lee, Y. L., Beausejour, C. M., Rock, J. M., Augustus, S., Jamieson, A. C., Porteus, M. H., Gregory, P. D. & Holmes, M. C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 435, 646-651.

Wang, H., Yang, H., Shivalila, Chikdu S., Dawlaty, Meelad M., Cheng, Albert W., Zhang, F. & Jaenisch, R. (2013). One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell*.

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 Res.* 22, 1316-1326.

Weber, E., Ojanen-Reuhs, T., Huguet, E., Hause, G., Romantschuk, M., Korhonen, T. K., Bonas, U. & Koebnik, R. (2005). The type III-dependent Hrp pilus is required for productive interaction of *Xanthomonas campestris* pv. *vesicatoria* with pepper host plants. *J. Bacteriol.* 187, 2458-2468.

Wei, H., Therrien, C., Blanchard, A., Guan, S. & Zhu, Z. (2008). The Fidelity Index provides a systematic quantitation of star activity of DNA restriction endonucleases. *Nucleic Acids Res.* 36, e50.

Welsh, K. M., Lu, A. L., Clark, S. & Modrich, P. (1987). Isolation and characterization of the *Escherichia coli* mutH gene product. *J. Biol. Chem.* 262, 15624-15629.

White, F. F., Potnis, N., Jones, J. B. & Koebnik, R. (2009). The type III effectors of *Xanthomonas*. *Mol. Plant Pathol.* 10, 749-766.

Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482, 331-338.

Woodhead, J. L., Bhawe, N. & Malcolm, A. D. (1981). Cation dependence of restriction endonuclease EcoRI activity. *Eur. J. Biochem.* 115, 293-296.

Wu, Y., Gao, T., Wang, X., Hu, Y., Hu, X., Hu, Z., Pang, J., Li, Z., Xue, J., Feng, M., Wu, L. & Liang, D. (2014). TALE nickase mediates high efficient targeted transgene integration at the human multi-copy ribosomal DNA locus. *Biochem. Biophys. Res. Commun.* 446, 261-266.

Yang, Z., Horton, J. R., Maunus, R., Wilson, G. G., Roberts, R. J. & Cheng, X. (2005). Structure of HinP1I endonuclease reveals a striking similarity to the monomeric restriction enzyme MspI. *Nucleic Acids Res.* 33, 1892-1901.

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, e82539.

Zou, J., Mali, P., Huang, X., Dowey, S. N. & Cheng, L. (2011). Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease. *Blood* 118, 4599-4608.

Zu, Y., Tong, X., Wang, Z., Liu, D., Pan, R., Li, Z., Hu, Y., Luo, Z., Huang, P., Wu, Q., Zhu, Z., Zhang, B. & Lin, S. (2013). TALEN-mediated precise genome modification by homologous recombination in zebrafish. *Nat Methods*.