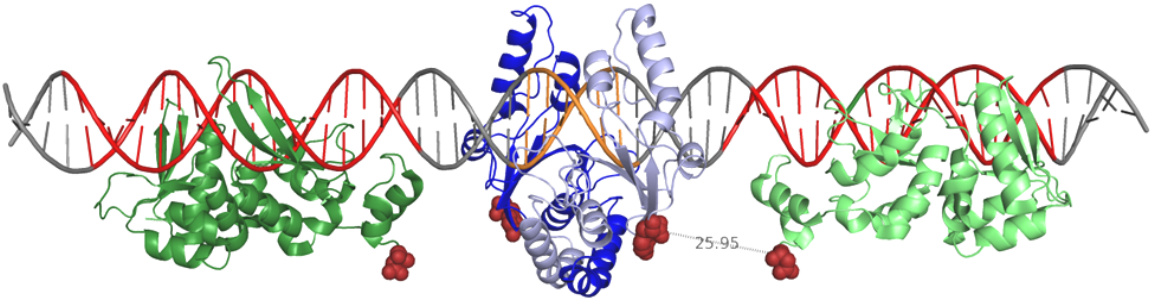
Generating meganucleases with new specificities by directed evolution and rational engineering approaches



Inauguraldissertation

zur Erlangung des Grades
Doktor der Naturwissenschaften
Dr. rer. nat.
Im Fachbereich Biologie und Chemie
der Justus-Liebig-Universität Gießen

vorgelegt von
Diplom-Biologin
Ines Fonfara

Gießen im August, 2011

Die vorliegende Arbeit wurde im Rahmen des Graduiertenkollegs *„Enzymes and Multienzyme Complexes Acting on Nucleic Acids“* (GRK 1384) und mit Finanzierung durch das DFG geförderte Schwerpunktprogramm (SPP 1170) *„Directed Evolution to Optimize and Understand Molecular Biocatalysts“* am Institut für Biochemie des Fachbereich 08 der Justus-Liebig-Universität Gießen in der Zeit von April 2008 bis Juli 2011 unter Leitung von Prof. Dr. A. Pingoud durchgeführt.

Erstgutachter: Prof. Dr. Alfred Pingoud
Institut für Biochemie
Fachbereich Biologie und Chemie
Heinrich-Buff-Ring 58
35392 Gießen

Zweitgutachter: Prof. Dr. Michael U. Martin
Professur für Immunologie
Fachbereich Biologie und Chemie
Winchesterstrasse 2
35394 Gießen

Erklärung

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

Gießen, den

(Ines Fonfara)

DANKSAGUNG

Zuerst möchte ich mich bei Prof. Dr. Alfred Pingoud für die Möglichkeit in seinem Institut Doktorarbeit machen zu dürfen bedanken. Danke auch für die wertvollen Diskussionen, Ideen und die stetige Motivation.

Herrn Prof. Dr. Michael Martin danke ich für die Übernahme des Koreferats.

Dr. Wolfgang „Wolle“ Wende danke ich für die langjährige Betreuung, das stetig offene Ohr, die Kritik, die einen anspornt doch immer nochmal ein bisschen besser zu sein, und die grenzenlose Geduld.

Prof. Dr. Peter Friedhoff gilt mein Dank für die zahlreichen Diskussionen, teilweise zwischen „Tür und Angel“, für die vielen kleinen und großen Hilfestellungen und für die Möglichkeit Mitglied des IRTG gewesen zu sein.

Meinen russischen Kollaborationspartnern Dr. Andrei Alexeevski und Alexander Grishin danke ich für die Möglichkeit bei Ihnen etwas über Bioinformatik zu lernen.

Danke an Dr. Ute Curth, für die Hilfe bei den AUZ Experimenten.

Ein besonderer Dank an Karina und Ina, die bei allen bürokratischen Schwierigkeiten stets zur Seite standen und alles geregelt haben und ein ganz großes Dankeschön an Anja, für die tolle Organisation rund um das IRTG und natürlich für das Korrekturlesen.

Ich möchte mich ausdrücklich beim gesamten Institut für Biochemie für die ausgesprochen nette Arbeitsatmosphäre bedanken. Vor allem bei den „alten“ Mitgliedern, die mich in Ihren Reihen aufgenommen und immer unterstützt haben und die zu guten Freunden geworden sind. Steffi: danke für die vielen Gespräche und Unternehmungen innerhalb und außerhalb des Labors, Katja: danke für die kleinen und großen „Pausen“, Ann-Jo: danke für den vielen Spaß im Institut und natürlich auch danach und die immer wieder interessanten Diskussionen, Roger: danke für die verrückte Zeit in Moskau und die seelisch und moralische Unterstützung und Jasmina danke für die schöne Zeit zusammen im L1 und vor allem auch danach. Danke, dass ich euch kennenlernen durfte.

Außerdem danke ich Ina D., die mich in die Geheimnisse der Laborarbeit eingeweiht hat, Sabrina für die technische Unterstützung und den Spaß zwischendurch und Caro für die super Zeit in „unserem“ Labor, ich hätte mir keine bessere Laborkollegin wünschen können.

Ein besonderes Dankeschön geht an Dr. Vera Pingoud die mir viele kleine Tricks und Kniffe beigebracht hat.

Meinen Freunden Jessica, Fenja, Judith, Janette, Stöps, Susi, Falk und David möchte ich dafür danken, dass sie immer an mich geglaubt haben, immer für mich da waren und ein offenes Ohr hatten und mir immer wieder gezeigt haben das es auch noch was anderes gibt als Arbeit. Dankeschön, ich weiß nicht, was ich ohne euch gemacht hätte.

Abschließend gilt mein Dank meinen Eltern und meinem Bruder für die Unterstützung, die Liebe und das in mich gesetzte Vertrauen, dass sie mir immer den Rücken gestärkt haben und mir damit das alles erst ermöglich haben. Dankeschön für euer Verständnis, als ich mich für diesen Weg entschieden habe.

"O verdadeiro encenador da nossa vida é o acaso – um encenador cheio de crueldade, misericórdia e encanto cativante."

Amadeu de Prado – Um ourives das palavras

„Der wirkliche Regisseur unseres Lebens ist der Zufall - ein Regisseur voll der Grausamkeit, der Barmherzigkeit und des bestrickenden Charmes."

Amadeu de Prado – Der Goldschmied der Worte

(Pascal Mercier – Nachtzug nach Lissabon)

PUBLICATIONS

Grishin A., **Fonfara I.**, Alexeevski A., Spirin S., Zanegina O., Karyagina A., Alexeyevsky D., Wende W. *Identification of conserved features of LAGLIDADG homing endonucleases.* J Bioinform Comput Biol, 2010 Jun; 8 (3):453-69

Fonfara I., Curth U., Pingoud A., Wende W., *Creating highly specific nucleases by fusion of active restriction endonucleases and catalytically inactive homing endonucleases,* Nucleic Acids Res, 2011; doi: 10.1093/nar/gkr788

PATENT APPLICATION

Application No 61/495410 *"Nuclease fusion protein and uses thereof"*

ORAL PRESENTATIONS

August 2008, SPP1170 Workshop, Bremen, Germany, *"New strategies to generate homing endonucleases of pre-defined specificity by directed evolution"*

November 2009, Minisymposium IRTG 1384, Moscow, Russia, *"A new approach for generating meganucleases"*

May 2010, Workshop IRTG 1384 "Enzymes and enzyme complexes acting on nucleic acids", Vilnius, Lithuania, *"Natural and artificial meganucleases – tools for gene therapy"*

June 2010, Off-Spring Meeting IRTG 1384, Moscow, Russia, *"Binding and bending of DNA by proteins examined via gel-based assays"*

POSTER PRESENTATIONS

June 2010, FASEB Summer Research Conference "Nucleic Acid Enzymes", Saxton River, Vermont, USA, *"Extending meganucleases via domain fusions"*

August 2010, 6th NEB Meeting on DNA Restriction and Modification, Bremen, Germany, *"Natural and artificial meganucleases – tools for gene therapy"*

TABLE OF CONTENTS

ABBREVIATIONS

Abbreviations for chemicals are listed in Table 2.1.

A	Ampère
bp	Base pair(s)
Da	Dalton
DNA	Deoxyribonucleic acid
e.g.	Exempli gratia; for example
EMSA	Electrophoretic mobility shift assay
FPLC	Fast protein liquid chromatography
g	Gram
h	Hour
HE	Homing endonuclease
i.e.	id est; that is
k	Kilo
l	Liter
LB	Luria-Bertani
m	Milli/meter
M	Molar
min	Minute
MW	Molecular Weight
n	Nano
o/n	Overnight
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
Pfu	*Pyrococcus furiosus*
rpm	Rotations per minute
sec	Second
Taq	*Thermus aquaticus*
T_M	Melting temperature
UV	Ultraviolet
V	Volt
v/v	Volume/volume
wt	Wild type
w/v	Weight/volume
x g	Gravity
Ω	Ohm
µ	Micro

1. INTRODUCTION

1.1 GENE TARGETING

After deciphering the first genome of an organism in 1995, i.e. from *Haemophilus influenzae* [1], a large number of sequences followed until finally the human genome was published in 2001 [2,3]. This opened several possibilities for gene therapy and gene targeting, e.g. the genetic information for hereditary monogenic diseases are now available and the genes can be addressed directly. This can be realized by several tools, summarized in Figure 1.1, which create specific double-strand breaks next to the mutated sequence responsible for the disease or in a certain locus where transgenes should be introduced. These double-strand breaks can increase the number of homologous recombination events [4], whose frequency is naturally quite low. By providing the correct gene in *trans* this can be used as repair matrix and exchanged for the defect gene.

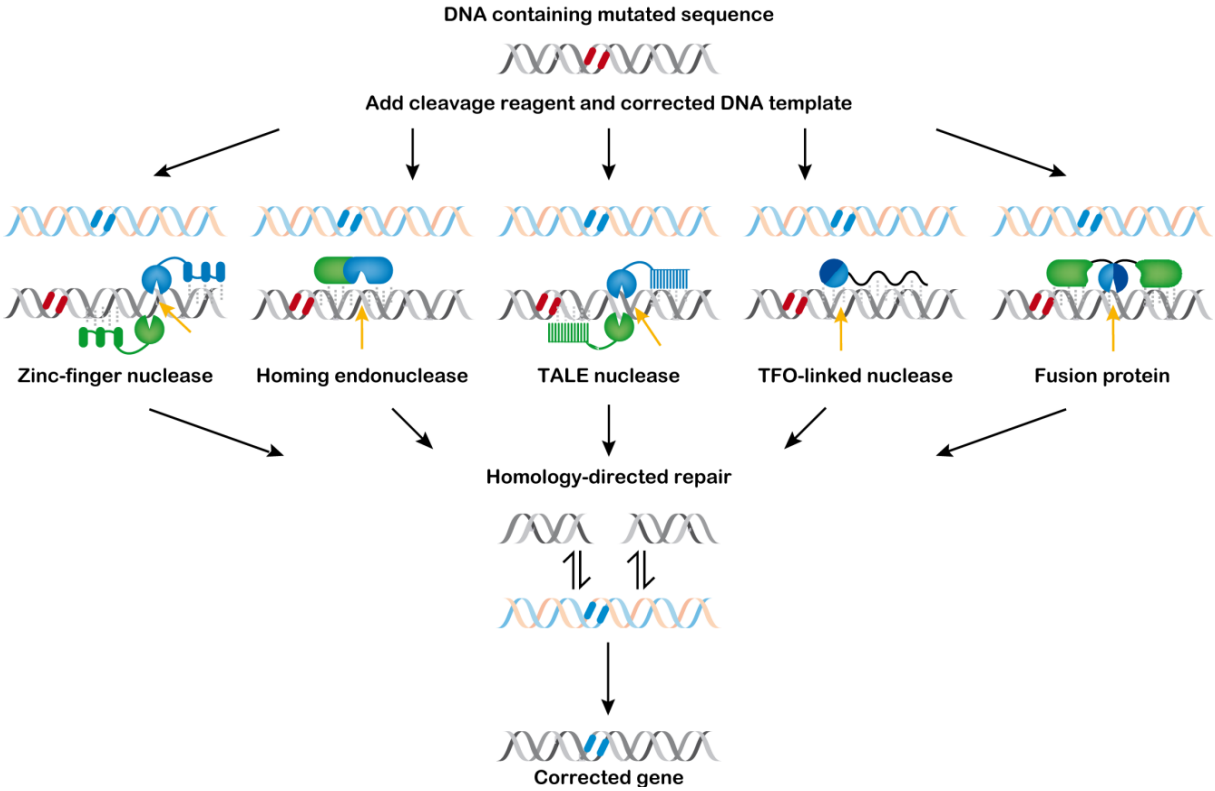


Figure 1.1: Overview of possibilities for gene targeting. *Specific double-strand breaks in proximity to a mutated sequence introduced by several tools, namely Zinc-finger nucleases, homing endonucleases, TALE-nucleases, TFO-linked nucleases and fusion proteins will be repaired via stimulating homologous recombination using the newly introduced sequence (blue, pink) as matrix. Modified after [5].*

1.1.1 HOMING ENDONUCLEASES

Comparing the highly specific nucleases presented in Figure 1.1 it is striking that homing endonucleases (HE) are the only ones already existing in nature. All the others were engineered by fusing specific DNA-binding modules to DNA-cleavage modules, i.e. restriction endonucleases, and by this prolonging their recognition site. Homing endonucleases, usually encoded by group I or group II introns or existing as inteins or freestanding open reading frames, are able to recognize sequences of up to 40 bp in length [6], which makes them perfect candidates for gene targeting tools.

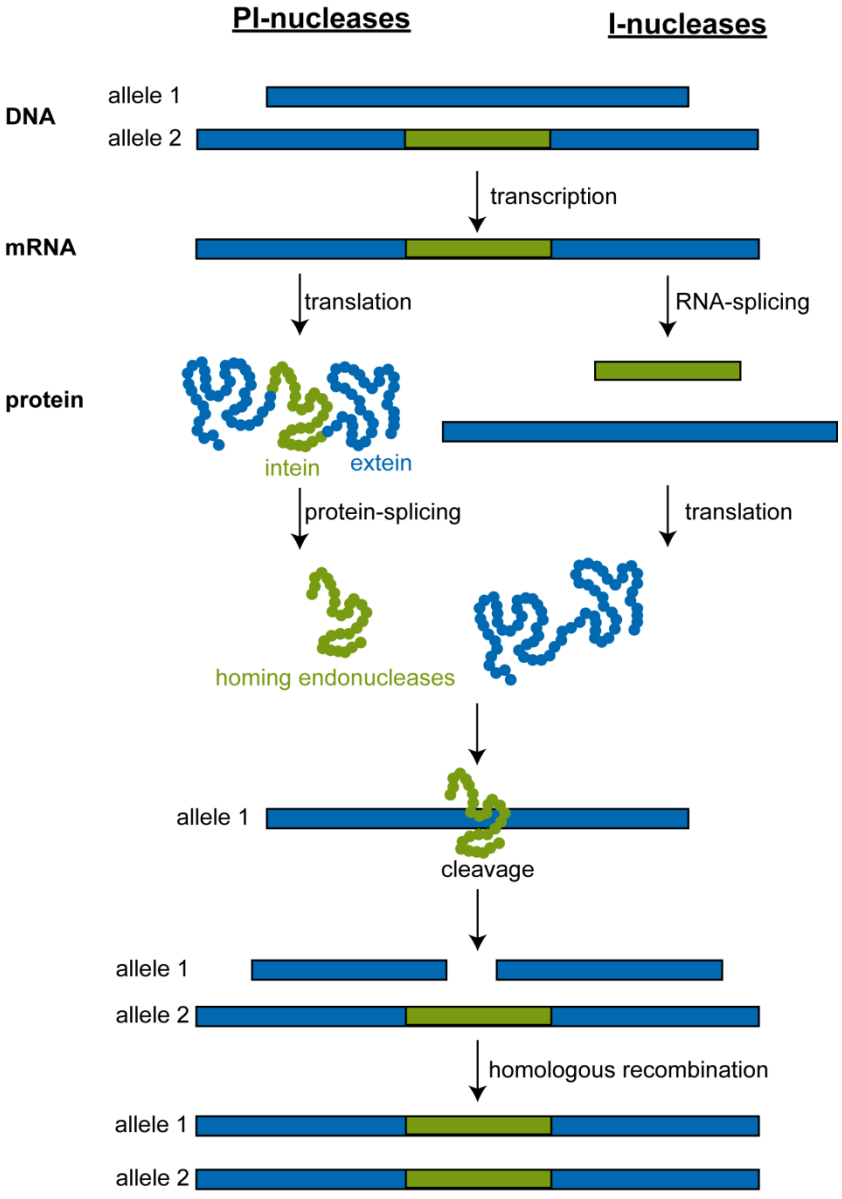


Figure 1.2: The homing process. *PI-nucleases, existing as inteins, are spliced out of a precursor-protein, while I-nucleases are encoded in introns which are translated after RNA-splicing. The homing process is the same for both, they recognize a sequence in the intronless allele, cleave there and this double-strand break is then repaired by using the intron-containing allele as repair matrix, resulting in a homozygous situation. Modified after [7]*

The main function of these enzymes is the propagation of their genes, which is realized by cleavage of a sequence on the allele lacking their own coding sequence, and this double-strand break will be repaired via homologous recombination using the other allele containing the coding sequence for the HE as matrix (Figure 1.2). By this, the maintenance of the HE-coding gene in both alleles is assured. The fact that this process, called homing is quite similar to the above described mechanism of gene targeting is another good reason for choosing HEs as specific cleavage tools.

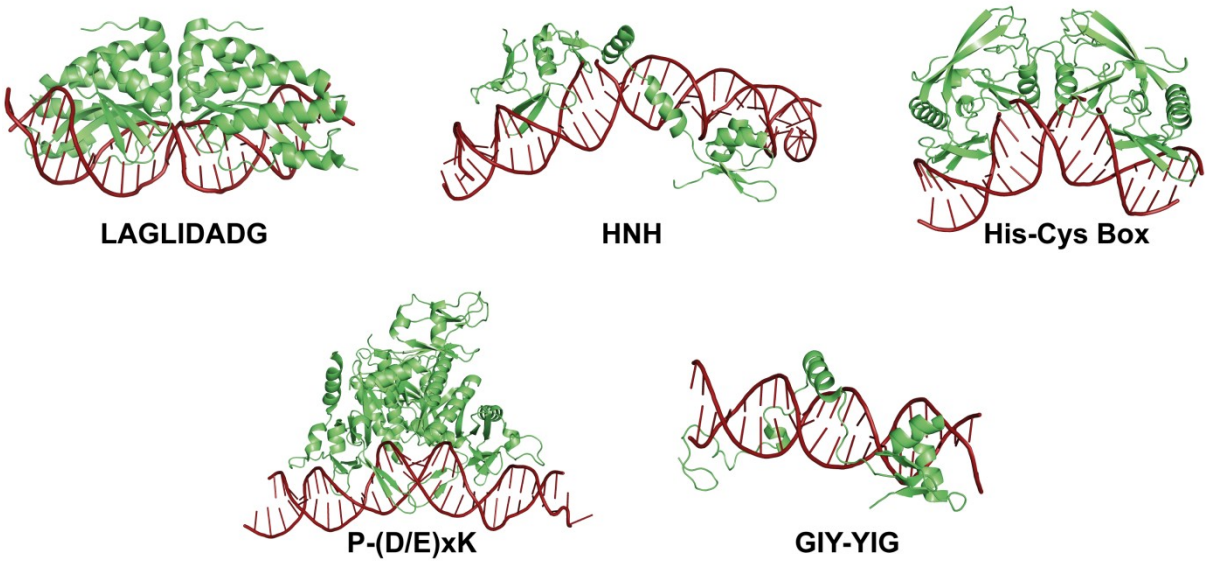


***Figure 1.3: Overview of the five homing endonuclease families with one representative for each**. I-CreI (pdb: 1G5Y) as LAGLIDADG, I-HmuI (pdb: 1U3E) as HNH, I-PpoI (pdb: 1A73) as His-Cys Box, I-Ssp6803I (pdb: 2OST) as PD-(D/E)xK and the DNA-binding domain of I-TevI (pdb: 1T2T) as GIY-YIG family member.*

According to their active center five families of homing endonucleases, namely LAGLIDADG, His-Cys Box, HNH, GIY-YIG and PD-(D/E)xK, have been described so far. For representatives of all these families structural data are available, revealing how these enzymes manage to recognize and bind their large recognition sequences (Figure 1.3). The best studied and largest of these families are the LAGLIDADG homing endonucleases. This family can be furthermore divided into two subgroups, monomeric and homodimeric proteins. The homodimeric LAGLIDADG homing endonucleases contain only one LAGLIDADG-motif per subunit (e.g. I-CreI) whereas the monomeric enzymes harbor two motifs (e.g. I-SceI). Monomeric LAGLIDADG homing endonucleases can be further divided into intron encoded proteins and those existing as inteins (e.g. PI-SceI). The last mentioned harbor an additional domain which is responsible for splicing the enzyme out of the precursor protein and also extends the DNA-binding ability. Common feature of all LAGLIDADG enzymes is their conserved αßßαßßα core topology, with the LAGLIDADG motif being at the end of the

first α-helix. The ß-sheets normally form a saddle-like structure binding specifically to the DNA [6].

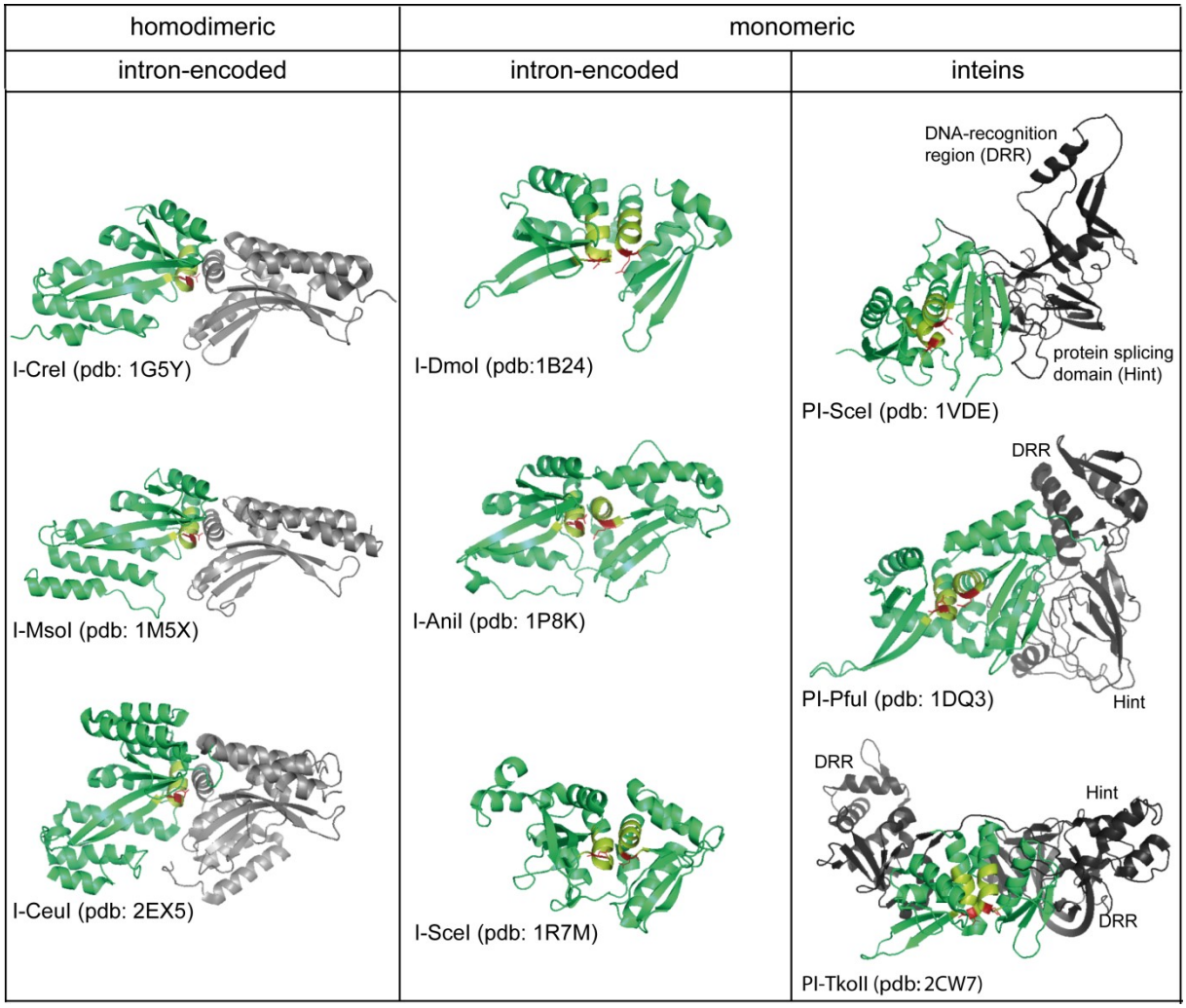


Figure 1.4: Overview of published crystal structures representing the structural subfamilies of LAGLIDADG homing endonucleases*, homodimeric, monomeric intron encoded and monomeric inteins. The LAGLIDADG motif is indicated in yellow and the catalytically active residues are red. In the homodimeric enzymes just one subunit is colored green and in the case of the monomeric inteins the additional DNA-recognition-region (DRR) and the splicing domain are dark-gray.*

Due to the relatively large recognition sequence and the tolerance of single base changes, homing endonucleases seem to have a higher plasticity in DNA binding/recognition compared to restriction enzymes. This feature might be an advantage in engineering and designing them for genomic applications (see review: [8]). Therefore, a lot of effort is currently undertaken to create homing endonucleases with an altered specificity. The four most important of these approaches are: domain/subdomain shuffling [9-11], computational redesign [12,13], screening of variants with mutations at residues predicted to be involved in DNA-binding [14-16] and *in vivo* selection assays [17-19] which are also part of the work described here and will therefore be discussed later in more detail.

Successful examples for using homing endonucleases as gene targeting tools are given by the company Cellectis, SA (Romainville, France), which developed a high-throughput screening method for variants of I-CreI [20-22]. With this assay, they screened millions of variants containing two to four amino acid changes in positions known to be responsible for protein-DNA contacts against any potential target site. This resulted in hundreds of I-CreI variants with locally changed recognition patterns, which were then used in combinatorial approaches by combining different sets of mutants in one protein, ending up with fully reengineered meganucleases (Figure 1.5).With this building-block like approach, Cellectis was able to create a library of > 30,000 I-CreI derived variants recognizing in principle almost every desired target sequence.

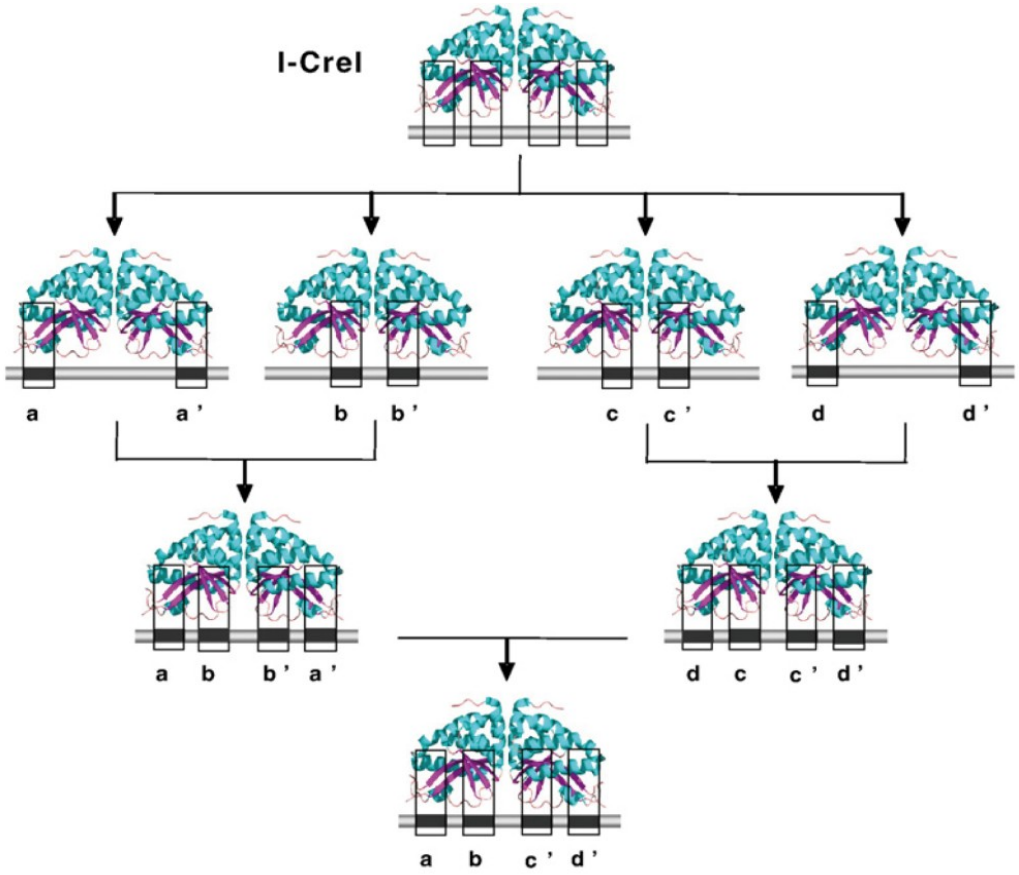


***Figure 1.5: Overview of combinatorial strategy pursued by Cellectis**. In a screening approach I-CreI variants with altered specificity were selected, whose local mutations were later on combined, creating libraries of engineered I-CreI derived meganucleases. (from [23])*

From this library of I-CreI variants two enzymes were selected so far, targeting human gene loci, namely *XPC* involved in *Xeroderma pigmentosum* [23] and *RAG1* involved in the *X-chromosome linked severe combined immunodefiency* (X-SCID) [24,25]. Furthermore, meganucleases were used for so-called virus clipping, meaning the nuclease targets viral sequences and thus is able to cleave these sequences integrated into the host genome. If no

repair matrix is available this double-strand break will be repaired via non-homologous end joining (NHEJ) leading in most cases to insertions or deletions (indels) [26], and the viral DNA will be rendered harmless. An example of this concept is reported by Grosse *et al*. [27] where recombinant Herpes-simplex viruses (HSV-1) were produced harboring an I-SceI site in their genome or meganucleases based on I-CreI were created to cleave inside the natural HSV-1 genome. In this study, they showed a 50 % reduction of viral infection when the meganuclease gene was transfected before virus infection. The repair mechanism of NHEJ is unwanted in the case of directed gene targeting, where homologous recombination should be preferred, inserting the correct gene sequence into the double-strand break.

Since Metzger *et al*. [28] reported that nicking of DNA is sufficient for stimulation of homologous recombination and provokes a lower frequency of NHEJ, mega-nickases which were studied years before [29,30], offer also an alternative as gene targeting tools.

1.1.2 ZINC-FINGER NUCLEASES

This group of chimeric nucleases is, until now, the most prominent and most widely used one. They consist of zinc-finger proteins linked to the catalytic domain of FokI. Zinc-finger proteins contain three to six zinc fingers each of about 25 amino acids with five amino acid spacers between the individual fingers. In each zinc finger a zinc ion is coordinated by two cysteine and two histidine residues, leading to the characteristic three dimensional structure (Figure 1.6).

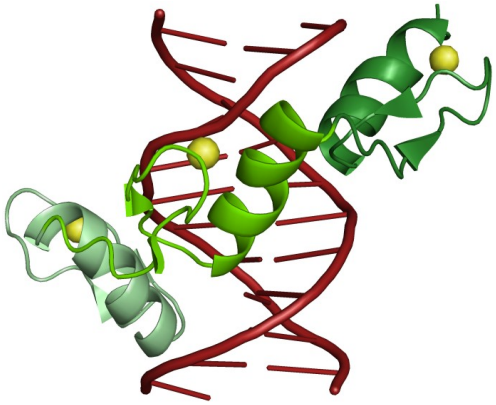


Figure 1.6: Co-crystal structure of the Zinc-finger protein ZIF268 *(pdb: 1A1L) consisting of three zinc finger motifs, each recognizing approximately three base pairs. Zinc ions are depicted in yellow.*

Residues -1, 3 and 6 in every helix form specific hydrogen bonds to a base pair triplet leading to specificity of the zinc finger and by this being the programmable part. Using phage display technology, it was possible to generate zinc-finger proteins recognizing a large

number of sequences, a method optimized and widely used by the company Sangamo Biosciences, Inc (Richmond, CA, USA) (for review see [31]). Having this powerful tool to target any possible sequence the next step was using it in combination with a cleavage module as a highly specific nuclease, so-called Zinc-finger nucleases (ZFN) (Figure 1.7) [32-34]. As cleavage module the catalytic domain of the restriction endonuclease FokI was used, which by itself is not able to recognize a specific sequence [35]. Since FokI has to dimerize prior to catalysis [36], a 16 to 36 bp sequence could be recognized. The problem in this case is that FokI can also cleave if only one subunit is properly bound to the DNA [37], which leads to cleavage of halfsites of the recognition sequence and by this to cytotoxicity. For this reason, it was aimed to change the dimerization interface of FokI and by this creating functional heterodimers [38-40].

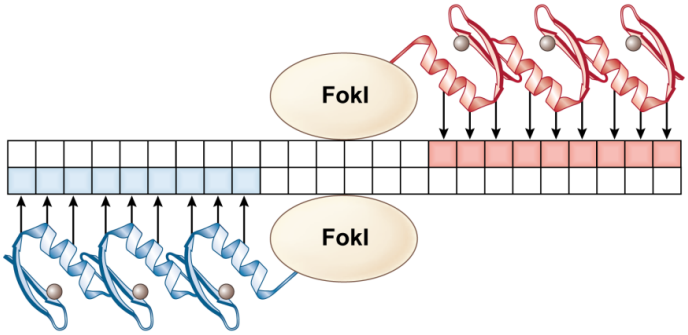


Figure 1.7: Model of a typical zinc-finger nuclease (ZFN) *consisting of the catalytic domain of FokI fused to a three zinc-finger protein recognizing 9 bp sequences. The FokI subunits dimerize for activity and by this an 18 bp tandem repeat sequence could be recognized. Modified after [31].*

Until now, ZFNs have been widely used for genome modifications in several organisms, including fruit fly, nematodes, tobacco, zebrafish and human (for review see [41]). In case of human genome modification two examples have to be highlighted. The first one is based on homologous recombination stimulated by a double-strand break at the human *IL2γR* gene locus leading to exchange of the defect gene involved in X-SCID [42,43]. The second approach, which already reached clinical trials, is based on NHEJ by creating a ZFN targeting the human CCR5, a co-receptor for HIV infection. If this receptor is not functional the patients show resistance against HIV infection [44], which implies possibilities for anti HIV therapy. Perez *et al*. [45] reported a ZFN cleaving a sequence in the CCR5 locus in primary CD34+ T-cells, which is then repaired via error-prone NHEJ leading to disruption of the CCR5 coding sequence and can confer resistance against HIV infection.

In summary, until now ZFN are the tool of choice regarding gene targeting and genome modification, but it has to be considered that they have been investigated since

15 years, which is an advantage compared to for example TALE nucleases, which have been discovered to be good candidates for gene targeting only two years ago.

1.1.3 TALE NUCLEASES

After decoding the DNA-binding and recognition mechanism of the transcription activator like effectors (TALE) from the plant pathogen *Xanthomonas* in 2009 [46], these proteins were thought to be perfect candidates for DNA-binding modules for highly specific nucleases. DNA-binding by these proteins is mediated through a tandem repeat domain containing repeat units of usually 33 - 35 amino acids, while amino acids at position 12 and 13 display the so-called repeat-variable diresidue (RVD) responsible for specific base pair contacts (Figure 1.8A) [47]. The NMR structure of 1.5 repeats of the TALE PthA revealed that one repeat consists of three α-helices forming a tetratricopeptide repeat (TPR)-like structure (Figure 1.8C) [48] with the RVDs located on the top of the helix turns. It is suggested that these repeat units encircle the DNA with the RVDs forming the base specific contacts [49].

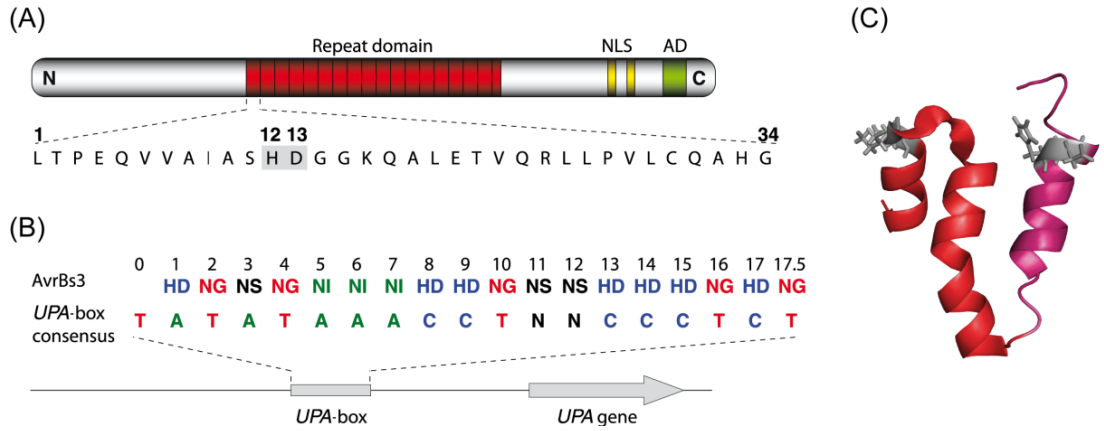


***Figure 1.8: Domain organization of TALE AvrBs3 from* Xanthomonas campestris pv. *and NMR structure of PthA from* Xanthomonas axonopodis pv. citris** *(A) the repeat domain in red containing 17.5 repeat units each with a length of 34 amino acids, with residue 12 and 13 (shaded gray) being variable. (B) Amino acid residues 12 and 13 from all 17.5 repeats of AvrBs3, with the respective nucleotide recognized by them. From [46] (C) NMR structure (pdb: 2KQ5) showing 1.5 repeats, with one repeat in red and the other half-repeat in pink, the RVDs are depicted in gray.*

A certain combination of these diresidues recognizes a distinct base pair sequence (Figure 1.8B), which makes TALEs a programmable DNA-binding module. In contrast to zinc fingers, there are well defined rules for specific recognition by the TALEs, making the use of extensive screening procedures unnecessary. This new DNA-binding module was also fused to the catalytic domain of FokI, as it was reported for ZFNs, to generate so-called TALENs (TALE-nucleases) [50-52]. Until now, these new chimeric enzymes seem to offer an

alternative to the ZFN used so far, especially because of their predictability, but further studies have to be made and particularly co-crystal structures of these enzymes should answer some open questions.

1.1.4 TFO-LINKED NUCLEASES

In this class of highly specific nucleases the DNA-binding module is a triple helix forming oligonucleotide (TFO), which can bind duplex DNA in a sequence-specific manner and by this provides a high degree of programmability [53]. This module was fused to different DNA cleavage modules, as for example restriction enzymes [54] or psoralen [55]. Disadvantage of TFO-linked nucleases is that in contrast to all other highly specific nucleases the specificity is mediated by a non-proteinaceous DNA-binding module, which makes expression of the fusion construct in cells difficult and needs direct delivery of the TFO-linked nuclease into the cells.

1.2 MEGANUCLEASES

In the beginning of the gene targeting era all highly specific nucleases were summarized by the term “meganucleases” [56], while in the meantime this name is reserved only for homing endonucleases and engineered variants thereof as they will be presented in this work.

1.2.1 DIRECTED EVOLUTION

Enzymes evolved during millions of years due to natural selection. The properties evolved, such as activity, solubility or specificity are often not suitable for applied sciences or industry. For this reason directed evolution approaches for finding optimized enzymes were developed. Usually, libraries of enzyme variants were created either by error-prone PCR [57] or DNA-shuffling [58], if for instance the structure of the enzyme is not available or by site-directed mutagenesis, if residues involved for example in DNA-binding or catalysis are known. These libraries can then be screened or selected for the desired property of the enzyme (for review see [59]). Since the work presented here deals with meganucleases or highly specific nucleases to create double-strand breaks into DNA, only systems for directed evolution of this class of enzymes will be discussed.

Screening systems for meganucleases with altered specificity were mainly developed and used by Cellectis, as already mentioned above. These assays were either performed in yeast or CHO (*chinese hamster ovary*) cells, where recombination events due to double-strand

cleavage of the HE at the desired target sequence were detected by blue/white screening based on restoration of the *lacZ* locus [21,22]. Since screening methods are normally time-consuming, because every possible variant will be investigated individually and this could only be realized by high-through-put systems, most other systems for directed evolution of meganucleases are based on selection. In most cases, these selection systems link the activity and specificity of the meganuclease for a certain target site directly to the survival of *E.coli* cells. Figure 1.9 shows the principle strategy of such two-plasmid assays. *E.coli* cells were co-transformed with two plasmids, while one expresses the meganuclease library and the other one contains the target site to select for. The latter mentioned is called reporter plasmid and normally encodes a toxin. Expression of the meganuclease library is induced first and if there is no enzyme able to recognize and cleave the target sequence on the reporter plasmid, following induction of this toxin will lead to cell death. If there is a variant in the library which can cleave the target site, the reporter plasmid will be linearized and subsequently degraded by the *RecBCD* system of the cell [60]. Since there is no reporter plasmid available anymore, no toxin expression will be induced and the cells will survive. These surviving colonies should express homing endonuclease variants with an altered specificity. To increase the stringency of the assay, the surviving colonies are usually pooled, DNA extracted and again used to co-transform *E. coli* cells with reporter plasmid, repeating this several times.

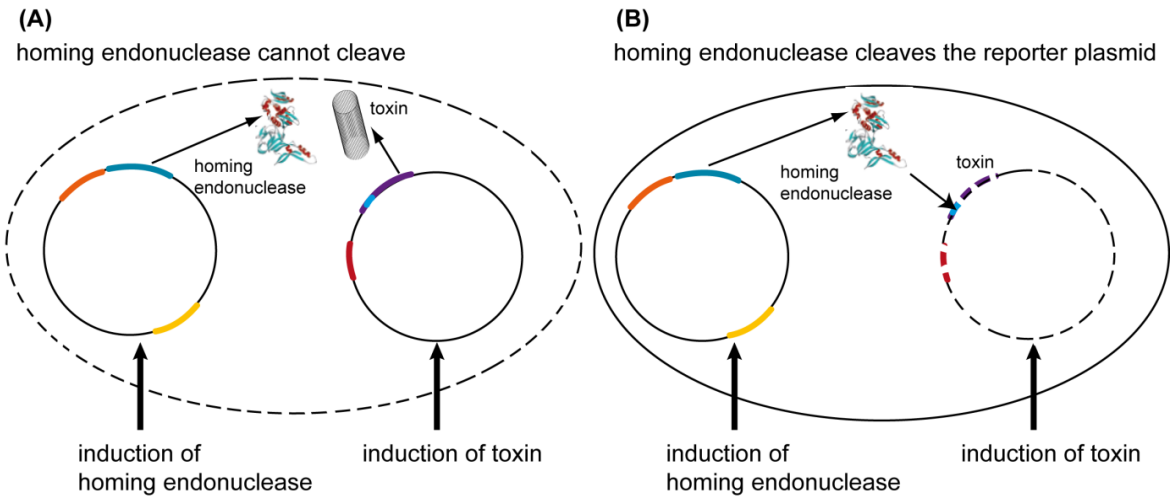


Figure 1.9: Principle of a two-plasmid selection assay. E.coli *cells were co-transformed with the plasmid expressing the homing endonuclease or variants thereof and a reporter plasmid expressing a toxin and harboring the target sites for selection. (A) The expressed HE is not able to recognize and cleave the target site on the reporter plasmid and therefore the cells will not survive toxin expression. (B) One of the HE-variants can cleave the reporter plasmid at the target site, which will be degraded by the RecBCD system of the cell and the* E.coli *cells will survive induction of toxin expression.*

The first such assay trying to find new variants of the homing endonuclease I-SceI was developed by Gruen *et al*. [17] in 2002. As toxic reporter they used Barnase (RNAse from *Bacillus amyloliquefaciens*) and the problems they were facing were a too high background of false positive cells and a too low sensitivity of the assay. These obstacles could be solved by Chen and Zhao [18] and Doyon *et al*. [19], both using CcdB (control of cell death B) as toxic reporter.

PI-SceI, a LAGLIDADG homing endonuclease existing as intein, was thought to be a suitable candidate for a gene targeting nuclease, but so far only one assay was available for selecting PI-SceI variants based on binding ability of the enzyme to different substrates [15]. For this reason, the before mentioned CcdB selection assay was used by a colleague of mine (Tobias Ullrich) to select new variants of PI-SceI. Unfortunately, the number of surviving cells in the CcdB-based selection assay was very low and it turned out that the cells only survived because of mutations in the gene encoding the toxin CcdB and not because PI-SceI cleaved its target site. A severe disadvantage of this system was that the homing endonuclease-encoding plasmid was controlled by a low-copy number promoter while the reporter plasmid was present in a high copy number in the cell. Since PI-SceI tends to stay bound to its cleavage product and by this shows a low activity rate, the number of enzyme molecules should be higher than the number of cleavage sites. In preliminary work (Diploma thesis Ines Fonfara, Giessen 2008) the existing assay was optimized to work for selection of PI-SceI variants. This optimization included the use of a high-copy plasmid for expressing the homing endonucleases and a low-copy plasmid for expressing the toxin, so that there are more homing endonuclease-expressing plasmids in the cell, than reporter plasmids. Furthermore the P_L-promoter from bacteriophage λ [61] was used to control the toxin expression, which has the advantage of tight regulation of expression by the temperature sensitive repressor cI857 [62]. This system by itself has the advantage that no additional chemical compound has to be added to the cells, only a temperature shift from 28 °C to 42 °C was needed. Another improvement was the use of a different reporter toxin, namely gvpA (gas vesicle protein A from *Anabaena flos-aquae* [63]). In contrast to the previously used toxins, this protein is only a structurally acting toxin, by building up gas vesicles inside the *E.coli* cells and the idea was that single mutations in the reporter gene would not necessarily lead to inefficacy of the toxin.

1.2.1.1 PI-SceI

This homing endonuclease from the LAGLIDADG family, existing as intein has a size of about 50 kDa. It is produced by protein splicing of the 119 kDa precursor protein VMA1

from *Saccharomyces cerevisiae*. For this reason, PI-SceI was called VDE (*VMA1 derived endonuclease*) [64] at first. After discovering an increasing number of homing endonucleases it was decided to have a common nomenclature [65], as it was done for restriction enzymes and it was renamed PI for "protein-intron" (intein) and Sce for *S.cerevisae*. PI-SceI recognizes a long asymmetric sequence of 31 bp, which is unique in the 13 million bp genome of yeast. This is one of the reasons why it was thought to be a good candidate for gene targeting. The second reason is its two-domain organization, with domain I being responsible for DNA-binding and also harboring a protein splicing activity while domain II with the two LAGLIDADG motifs is responsible for DNA-cleavage but also involved in DNA-binding (Figure 1.10) [66]. It was also shown that domain I alone is able to specifically bind DNA while domain II cannot bind DNA when expressed alone [67,68].

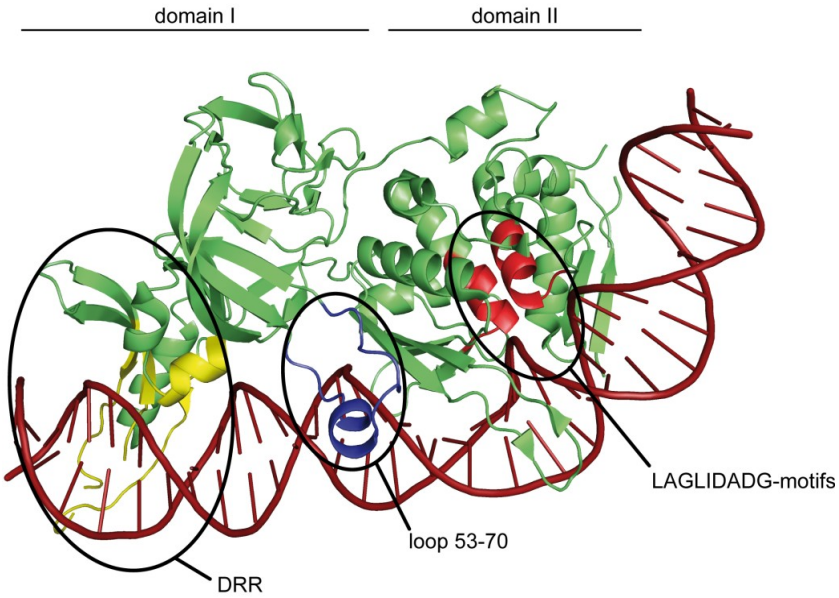


Figure 1.10: Co-crystal structure of PI-SceI *(pdb: 1LWS) showing the two domains. Domain I contains the DNA-recognition region (DRR, yellow) and the loop 53-70 (blue) which makes a flipping motion after DNA-binding [69] and domain II harbors the catalytic LAGLIDADG motifs (red).*

As observed in the co-crystal structure (Figure 1.10), PI-SceI bends the DNA by ~ 75 ° towards the major groove [70]. In this publication, it was also reported that PI-SceI tends to stay bound to one of the products after cleavage, which is the reason that this enzyme does not show a real turnover activity. The active center residues Asp 218 and Asp 326 are responsible for catalysis [71,72], both being the second Asp in the two LAGLIDADG motifs. PI-SceI cleaves its target site by leaving a 4 bp 3'overhang and requires divalent cations as cofactors. While these could be Mg^{2+} or Mn^{2+} the enzyme shows a higher activity in presence of Mn^{2+}, but also a more relaxed specificity [73]. Furthermore, the temperature optimum for cleavage activity is between 40 - 45 °C and the pH optimum is between 8.5 and 9 [64].

1.2.2 FUSION ENZYME ENGINEERING

As already reported, all generated chimeric nucleases so far, except the TFO-linked nucleases, consist of a DNA-binding module (zinc finger or TALE) and the catalytic domain of FokI as cleavage module. An alternative binding module could be also I-SceI due to its large recognition sequence and specificity, which was already reported [74], where a catalytically inactive variant of I-SceI was fused to the catalytic domain of FokI. The disadvantage of using the catalytic domain of FokI as cleavage module might be that it does not contribute to the specificity of the fusion enzymes itself, which could always result in potential off-site target cleavage. For this reason, specific nucleases, as for example type IIP restriction endonucleases, would offer an opportunity to create highly specific chimeric endonucleases, since they can contribute to some extent to the specificity of the fusion enzymes.

1.2.2.1 I-SceI

I-SceI is another member of the LAGLIDADG homing endonucleases encoded by a mitochondrial intron in *Saccharomyces cerevisiae*. This monomeric enzyme, with a size of 27.6 kDa, recognizes an asymmetric 18 bp sequence leading to 4 bp 3' overhangs, although it has to be mentioned that I-SceI shows some tolerance to mutations in the recognition site [75]. Furthermore, I-SceI tends to stay bound to one of the products after cleavage leading to a slow turnover number of this enzyme [76]. As for PI-SceI the two catalytically important aspartates 44 and 145 correspond to the second aspartate in the LAGLIDADG motifs [77]. I-SceI binds its recognition site with a K_D of ~ 9.7 nM [19], with the C-terminus binding to the upstream part and the N-terminus binding downstream where most of the protein-DNA contacts are made [77]. Since the early 90s, this enzyme is widely used for genome modification and initiation of recombination events in different cell types and organisms [78-83] and for this is still called the "gold-standard" [56].

1.3 AIM

In this work, two different approaches for generating meganucleases were pursued. The first aims at finding variants of the homing endonuclease PI-SceI with an altered specificity by directed evolution approaches. A library of variants was generated by mutating certain residues in the DRR of PI-SceI and this library was to be assayed with a previously developed two-plasmid selection system. The second approach is more rational by using the homing endonuclease I-SceI as specific DNA-binding module and fusing it to the

homodimeric restriction endonuclease PvuII. By this the recognition site of PvuII would be extended and a highly specific nuclease cleaving only a tripartite site, consisting of a PvuII site flanked by two I-SceI sites, would be generated. With the first assay, it was planned to show the possibilities of changing the specificity of an already existing meganuclease to eventually target a genomic locus. The second assay represents a proof-of-principle study to show PvuII as an alternative to the catalytic domain of FokI used as cleavage module so far. The here generated meganucleases would be used later in the so-called “safe harbor” approaches using I-SceI sites already introduced into plant genomes as “landing platform” [84].

2. MATERIAL AND METHODS

2.1 MATERIAL

2.1.1 CHEMICALS

All chemicals used listed in Table 2.1 were of high purity.

Table 2.1

Chemical	Abbreviation	Company
[α^{32}P] Deoxyadenosin-triphosphat	[α^{32}P]ATP	Hartmann Analytic
2-[4-(2-hydroxyethyl)piperazin-1-yl] ethanesulfonic acid	HEPES	AppliChem
2-Mercaptoethanol	2-Me	Merck
Acetic acid		Roth
Acrylamide-bisacrylamide solution (29:1) 40 %		AppliChem
Adenosin-triphosphat	ATP	Sigma
Agar		AppliChem
Agarose Ultra Pure ™		Invitrogen
Aluminium sulfate		AppliChem
Ammonium persulfate	APS	Merck
Ampicillin	Amp	AppliChem
Anhydrotetracyclin		IBA
Bacteriophage λ-DNA	λ-DNA	Fermentas
Bovine serum albumin	BSA	NEB
Bromphenol Blue		Merck
Chloramphenicol	Cm	AppliChem
Coomassie ® Brilliant Blue G250		AppliChem
Deoxynucleotide triphosphates	dNTPs	Fermentas
Desthiobiotine		Sigma Aldrich
Di-potassium hydrogen phosphate		Merck
Di-sodium hydrogen phosphate		Merck
Dithioerythritol	DTE	AppliChem
Dithiothreitol	DTT	AppliChem
Ethanol		Roth
Ethidium bromide		Roth
Ethylene diamine tetraacetate	EDTA	AppliChem
Glycerol		AppliChem
Glycine		Merck
Hydrochloric acid	HCl	Merck
Imidazole		Merck

Isopropyl-1-thio-ß-D-galactopyranoside	IPTG	Roth
Kanamycin sulfate	Kan	AppliChem
Magnesium acetate	Mg-acetate	Merck
Magnesium chloride	$MgCl_2$	Merck
Manganese chloride	$MnCl_2$	Merck
Ni-NTA-agarose		Quiagen
o-Phosphoric acid 87 %		Roth
Poly(deoxyinosinic-deoxycytidylic) acid sodium salt	Poly (dI-dC)	Sigma Aldrich
Polyethylene glycol	PEG	Merck
Potassium chloride	KCl	Roth
Potassium dihydrogen phosphate		Merck
Potassium glutamate		Merck
Sodium acetate	NaAc	AppliChem
Sodium chloride	NaCl	Merck
Sodium di-hydrogen phosphate		Merck
Sodium dodecyl sulfate	SDS	Roth
Sodium hydroxide	NaOH	Merck
Strep-Tactin® Sepharose®		IBA
Tetracyclin hydrochloride	Tet	AppliChem
Tetramethyl ethylene diamine	TEMED	Merck
Tris-(hydromethyl)-aminomethane	Tris	AppliChem
Tryptone		AppliChem
Xylene cyanol		Merck
Yeast extract		AppliChem

2.1.2 ENZYMES

All restriction enzymes and homing endonucleases listed in Table 2.2 were used in the recommended buffer according to the manufacturer's manual.

Table 2.2

Enzyme	**Cleavage site (5'→ 3')**	**Company**
BamHI	G|GATCC	Fermentas
BshTI	A|CCGGT	Fermentas
Bsp68I	TCG|CGA	Fermentas
DpnI	GA|TC	Fermentas
Eco105I	TAC|GTA	Fermentas
Eco47III	AGC|GCT	Fermentas
EcoRI	G|AATTC	Fermentas
EcoRV	GAT|ATC	Fermentas
EheI	GGC|GCC	Fermentas

FspI	TGC|GCA	Fermentas
HindIII	A|AGCTT	Fermentas
HpaII	C|CGG	Fermentas
I-SceI	TAGGGATAA|CAGGGTAAT	Fermentas
KpnI	GGTAC|C	Fermentas
KspAI	GTT|ACC	Fermentas
NaeI	GCC|GGC	Fermentas
NcoI	C|CATGG	Fermentas
NheI	G|CTAGC	Fermentas
NsiI	ATGCA|T	Fermentas
Pfl23II	C|GTACG	Fermentas
PmlI	CAC|GTG	Fermentas
PsiI	TTA|TAA	NEB
PvuII	CAG|CTG	Fermentas
SacI	GAGCT|C	Fermentas
SphI	GCATG|C	Fermentas
StuI	AGG|CCT	Fermentas
XhoI	C|TCGAG	Fermentas
ZraI	GAC|GTC	NEB

All other enzymes are listed in Table 2.3 and were also used according to the manufacturer's suggestion.

Table 2.3

Name	**Company**
CIAP	Fermentas
Phusion TM High-Fidelity DNA polymerase	NEB
Proteinase K	NEB
T4 DNA ligase	Fermentas
Taq-DNA polymerase	over-expressed and purified in our institute
Pfu-DNA polymerase	over-expressed and purified in our institute

2.1.3 OLIGONUCLEOTIDES

The oligonucleotides used for site-directed mutagenesis are listed in Table 2.4, with introduced restriction enzyme sites indicated in bold letters. (N= A, G, C or T; K= G, T; M= A, C)

Table 2.4

Name	Sequence (5'→ 3')	Comment
PL	GATACTGAG**GACGTC**AGCAGGACGCAC	introduce ZraI site in P_L promoter of reporter plasmid
newsubrev_B	CTGCCATTGAGCGACCTCTTTCTCCGCAC CCGACATAGA	PI-SceI newsubB
newsubrev_C	CTGCCATCGAGCGCCCTCTTTCTCCGCAC CCGACATAGA	PI-SceI newsubC
I-SceI_KpnI_3_rev	GATGCGTCCGGCCTCTATCAGC**GGTACC**A CCCTGTTATCCC	pAT_I3P, pAT_I7P
I-SceI_KpnI_8_rev	GATGCGTCCGGCCTCGATA**GGTACC**ATAT TACCCTGTTATC	pAT_I8P
I-SceI_KpnI_10_rev	CCACGATGCGTCCGGCGAGA**GGTACC**CTA TATTACCCTG	pAT_I10P
I-SceI_KpnI_12_rev	GATGCGTCCGGCCGA**GGTACC**AGCTATAT TACCCTGTTATCCC	pAT_I12P
pAT_IxP_P	CGATCACTGGA**CAGCTG**ATCGTCACGGCG	2nd PvuII site in pAT_IxP
I-SceI_D44S	CTGATCCTG**GGATCC**GCTTACATCCGTTC TCG	I-SceI_D44S
I-SceI_N152K	GTGGTAAATGGGA**TTATAA**GAAAAACTCT AC	I-SceI_N152K
I-SceI_D150C	GTAAATGGTGTTATAAGAA**GAATTC**TACC AAC	I-SceI_D150C
Oligo_078	CCATTTACC**GCCGGC**ATCCATGAACCAG	I-SceI_D145A
Oligo_079	CCTGGGAAACGCTTACATCCGTT**CACGT G**ATGAAG	I-SceI_S44N
VDE_T225A	GATTGTCTGA**CCGG**GCAGCTTTTTCGGTTG	PI-SceI_T225A
VDE_D218A	GGTTTATGGATT**GGCGCC**GGATTGTCTGA CAGGG	PI-SceI_D218A
VDE_ran_Q55H56	CAGAAAAGTNNKNNKAGAGCCCATAAAAG TGACTCAAGTCGTGAA	PI-SceI_pool Q55H56
VDE_ran_R90_L92_R94	CGTAGTGTCNNKCGTNNKTCTNNKACCAT TAAG	PI-SceI_pool R90L92R94
VDE_ran_S169H170	AGCTTTACGAACMNNMNNACCCAACAG	PI-SceI_pool S169H170
Oligo_185	TAAATTA**CTCGAG**GAATGGCCGCATATAC	PvuII_L12E
Oligo_196	GTGTGGCCACTC**CTCGAG**CAGTTTGTTC	PvuII(A)_L12E
Oligo_194	ATTATTA**GAGCTC**TGGGGTCATATACAGG	PvuII_P14G
Oligo_198	GAATGTGACCCCA**GAGCTC**CAGCAG	PvuII(A)_P14G
Oligo_186	TGGCCG**GATATC**CAGGAATATCAAGACTTAG	PvuII_H15D
Oligo_197	GATCCTGGTACTCTTG**GATATC**TGGCCAC	PvuII(A)_H15D
ww09_048	CTTCTAATTGGAGG**GTTAAC**AGTACTACCAG GACG	PvuII_T46G
Oligo_152	GGTGCTGCTGATCGG**AGGCCT**GACCGTGC TGCCGG	PvuII(A)_T46G
scPvuII_H83A_rev	GATTCATGTGG**TGCGCA**GTTGAAAAACC	PvuII_H83A
Oligo_162	GGTTCATATGA**TGCGCA**GTGCTAAAGCC	PvuII(A)_H83A
scPvuII_Y94F_rev	GGTACTTGTCTGAATT**TCGCGA**TAATTAAGG	PvuII_Y94F
Oligo_161	CCAC**GGTACC**TGGCGGAATTTCGCGATG	PvuII(A)_Y94F

Table 2.5 lists the complementary oligonucleotides containing certain target sites which were annealed and ligated into plasmids to create substrates for cleavage assays.

Table 2.5

Name	Sequence (5' → 3')	Comment
I-SceI_upper	CTATATTACCCTGTTATCCCTAGCGTAACT	I-SceI
I-SceI_lower	AGTTACGCTAGGGATAACAGGGTAATATAG	cleavage site
wt2006 upper	TCTATGTCGGGTGCGGAGAAAGAGGTAATGAAATGGCAG	PI-SceI
wt2006 lower	CTGCCATTTCATTACCTCTTTCTCCGCACCCGACATAGA	cleavage site
Sce6PvuII_up	GATCTTACGCTAGGGATAACAGGGTAATATAGGTCAGCTGGACGCA	S6P substrate
Sce6PvuII_low	GATCTGCGTCCAGCTGACCTATATTACCCTGTTATCCCTAGCGTAA	
Oligo_093	GATCTAGGGATAACAGGGTAATGGTACTCAGCTGATTCATATTACC CTGTTATCCCTA	S6P6S substrate
Oligo_094	GATCTAGGGATAACAGGGTAATATGAATCAGCTGAGTACCATTACC CTGTTATCCCTA	
Oligo_217	GATCTAGGGATAACAGGGTAATGGTACTGGTACCATTCATATTACC CTGTTATCCCTA	S6x6S substrate
Oligo_218	GATCTAGGGATAACAGGGTAATATGAATGGTACCAGTACCATTACC CTGTTATCCCTA	
Oligo_219	GATCTAGGGATAACAGGGTAATCAGGTACTCAGCTGATTCATCGAT TACCCTGTTATCCCTA	S8P8S substrate
Oligo_220	GATCTAGGGATAACAGGGTAATCGATGAATCAGCTGAGTACCTGAT TACCCTGTTATCCCTA	
Oligo_226	GATCTAGGGATAACAGGGTAATGGCTCAGCTGATTCATTACCCTGTA TATCCCTA	S4P4S substrate
Oligo_227	GATCTAGGGATAACAGGGTAATGAATCAGCTGAGCCATTACCCTGT TATCCCTA	

Table 2.6 lists the complementary oligonucleotides which were annealed and ligated between the NheI and AgeI cleavage sites located between the genes for PvuII and I-SceI to create different linker variants.

Table 2.6

Name	Sequence (5'→3')	linker name	amino acid sequence
Oligo_087	CTAGCGGTGGCTCTGGTTCTGGTT	$L_{(N)}$	ASGGSGSGSG
Oligo_088	CCGGAACCAGAACCAGAGCCACCG		
Oligo_089	CTAGCACCAAACAGCTGGTTAAGT	$L_{(+)}$	ASTKQLVKSG
Oligo_090	CCGGACTTAACCAGCTGTTTGGTG		
Oligo_091	CTAGCGGTGACTCTGGTTCTGATT	$L_{(-)}$	ASGDSGSDSG
Oligo_092	CCGGAATCAGAACCAGAGTCACCG		

Oligonucleotides for generation of PCR fragments as cleavage, competition or binding substrates are listed in Table 2.7; the corresponding template is listed as well.

Table 2.7

Name	**Sequence (5' → 3')**	**template plasmid**	**obtained fragment (length)**
CorrACYC 5'	GCCGACATGTTCTAGATCACGAGGC AGACCTCAGCGCC	gvpA_PL_I	I-SceI_star (454 bp)
corr_rev	GGAAACACGAACCCAAGCGTC		
pATlindiff 1for	CCGCCCAGTCCTGCTCGCTTC	pAT_I_10	S (292 bp)
2 Bend 3'	ATGCAAGGAGATGGCGCCCAACAGT		
pATlindiff 1for	CCGCCCAGTCCTGCTCGCTTC	pAT_I8P	S8P (298 bp)
2 Bend 3'	ATGCAAGGAGATGGCGCCCAACAGT		
HgiCI_1for	GGTAGATGACGACCATCAGGGAC	pAT_I6P6I	P (224 bp)
HgiCI_1rev	CGCCGGCTTCCATTCAGG		
lin diff kurz 5'	GCTTCGCTACTTGGAGCCACTATCGA	pAT_I6P6I	S6P6S (400 bp)
BBseqA783	ATCTCGACGCTCTCCCTTATGCG		

All other oligonucleotides used for screening (S), as second primer for generation of megaprimer (M) or amplification (A) of ligation inserts are listed in Table 2.8. As before, introduced restriction enzyme site are indicated by bold letters.

Table 2.8

Name	**Sequence (5'→ 3')**	**Comment**
pQe_for	GTATCACGAGGCCCTTTCGTCT	S
pQe_rev	CATTACTGGATCTATCAACAGGAG	S, M
VDE_SnaBI_rev	AGCTTTACGTACATGGGAACCCAACAG	S, M
VDE_rev	GGATTCTCAGTATTAAGATTATTGCG	S, M
VDE_SacI_rev	TCTTTGAAACTTCCTTGACGAGCTCAACA	M
VDE_BsiWI	TCCGCCGTTTGTCTCGTACGATTAAGG	M
VDE_R147A	TTGGTAGAATCATATGCAAAGGCTTCAAA	S
scPvuII-I-SceI delprom 5'	GATAACAATTTCACAC**CTCGAGCCATGGGCATG** **C**ACCACCCAGAT	A
scPvuII_KpnI_NsiI_rev	GCG**ATGCATGGTACC**GTGGTGATGATGATGGTGTGATC	A
T5lac 5'	CTTCAC**CTCGAG**AAATCATAAAAAATTTATTTG	A
T5lac 3'	GTTAT**GCATGC**TTAATTTCTCCTCTTTAATG	A
PvuII_S81C	TACAGGATTCATATGGTGGTGAGTGCAAAAACCTT	S
HgiCI rev2	CAGTTCTCCGCAAGAATTGATTGG	M
pACYC seq rev	GCAAGAATTGATTGGCTCCAATTCTTGG	S
PvuIISceI 5'	GGAGAAATTAA**CCATGG**GTCACCCAGATC	A

PvuIISceI 3'	GCG**AGCGCT**TTTCAGGAAAGTTTCGGAGGAG	A
PvuIISceI short 3'	GGA**AGCGCT**GTTCGGCAGTTTGTACATCATC	A
Oligo_071	C**GCTAGCCGTACGACCGGT**ATGCATAACATCAAA AAAAACCAGG	A
Oligo_072	**ACCGGTCGTACGGCTAGC**GTAAATCTTTGTCCCATGTTCC	A
omp 5'	GAGTTATTTTACCACTCCCTATCAG	S, M
omp 3'	CGCAGTAGCGGTAAACGGCAG	S
PvuII-KpnI	CAAAGATTTACGGTACCCACCATCATCATCAC	S
I-SceI_rev	GTCTGGGCGCCCCAGGTGATTACC	S
scPvuIIR129C fwd	AAATGGGAATGCAAATGGTATTCAGATGG	S
Oligo_100	TGTATTCTTTCAGCAGTTTAGAGTTCGGAC	S
Oligo_162	GGTTCATATGATGCGCAGTGCTAAAGCC	S

2.1.4 MARKERS

Markers used for DNA and protein electrophoresis are listed in Table 2.9.

Table 2.9

Name	Company
pUC 8 Mix marker	Fermentas
GeneRuler^{TM} 1kB DNA Ladder	Fermentas
PageRuler^{TM} unstained Protein Ladder	Fermentas

2.1.5 *E.COLI* STRAINS

LK111λ

($(r_K^- m_K^+)$ thi-1 thr-1 leuB6 ton A21 supE44 lacI^{q}YZ ΔM15 hfr λ^+)

This strain was used in the directed evolution approach for cloning of the reporter plasmids, since it harbors the gene for a strongly expressed cI-repressor.

TGE900

F^- su^- ilv^- bio (λcI857ΔBamΔHI)

The cI-repressor variant expressed by this strain is the temperature sensitive form cI857, which is why these cells were used for the selection assays.

JM109

F'traD36 proA$^+$B$^+$ lacI^q Δ(lacZ) M15/e14$^-$ (McrA$^-$) Δ(lac-proAB) endA1 gyrA96(Nal^r) thi-1 hsdR17 ($r_K^-m_K^+$) glnV44 relA1 recA1

For all cloning steps of the homing endonuclease expressing plasmids this strain was used.

XL10 Gold (pLGM)

endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tet^R F'[proAB lacI^{q}ZΔM15 Tn10(Tet^R Amy Cm^R)]

This strain was used for cloning and amplification of all fusion enzymes and PvuII variants, since it also contains a plasmid coding for the PvuII methyltransferase (pLGM) [85], which protects the DNA from autodigestion by PvuII. To maintain this plasmid, these cells were always grown in the presence of kanamycin. For the *in vivo* assays of the fusion enzymes the same strain, but without pLGM, was used.

DH5α

F$^-$ endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(r_K^- m_K^+), λ–

For amplification and cloning of the substrate plasmids in the fusion enzyme engineering part this strain was used.

2.1.6 BUFFERS

Water for preparing the buffers was used from MilliQ-Synthesis (Millipore) water purification system.

2.1.6.1 Protein Purification Buffer

I-SceI (Heparin column)

Lysis/Binding buffer	50 mM Tris/HCl pH 8, 0.1 mM EDTA, 200 mM NaCl
Washing buffer	10 mM Tris/HCl pH 8, 200 mM NaCl, 1 mM DTT
Elution buffer 0.5-2	10 mM Tris/HCl pH 8, 0.5, 1, 1.5 or 2 M NaCl, 1 mM DTT
Dialysis buffer	10 mM Tris/HCl pH 8, 100 mM NaCl, 1 mM DTT, 10 % (v/v) glycerol

PI-SceI (His_6-tag, Ni-NTA agarose)

Lysis buffer	10 mM HEPES/KOH pH 7, 1 M KCl, 15 mM imidazole
Washing buffer	10 mM HEPES/KOH pH 7, 200 mM KCl, 15 mM imidazole
Elution buffer	10 mM HEPES/KOH pH 7, 200 mM KCl, 200 mM imidazole
Dialysis buffer	10 mM KPi pH 7.2, 200 mM KCl, 0.1 mM EDTA, 50 % (v/v) glycerol

scP-$L_{(H)}$-$S_{(D44S)}$ (His_6-tag, Protino Ni-IDA column)

Lysis/Washing buffer	50 mM NaH_2PO_4 pH 8, 300 mM NaCl
Elution buffer	50 mM NaH_2PO_4 pH 8, 300 mM NaCl, 250 mM imidazole
FPLC gelfiltration buffer	50 mM NaPi pH 7.5, 500 mM NaCl, 0.5 mM EDTA
Dialysis buffer	50 mM NaPi pH 7.5, 100 mM NaCl, 0.5 mM EDTA, 60 % (v/v) glycerol

All further fusion enzymes and PvuII variants (Strep-tag, Strep-Tactin® Sepharose®)

Lysis/Washing buffer	100 mM Tris/HCl pH 8, 1 M NaCl, 1 mM EDTA
Washing buffer	100 mM Tris/HCl pH 8, 300 mM NaCl, 1 mM EDTA
Elution buffer	100 mM Tris/HCl pH 8, 150 mM NaCl, 1 mM EDTA, 2.5 mM desthiobiotin
FPLC gelfiltration buffer	50 mM NaPi pH 7.5, 500 mM NaCl, 0.5 mM EDTA
Dialysis buffer	50 mM NaPi pH 7.5, 100 mM NaCl, 0.5 mM EDTA, 60 % (v/v) glycerol
AUC dialysis buffer	50 mM NaPi pH 7.5, 100 or 500 mM NaCl, 4.4 % (w/v) glycerol

2.1.6.2 Cleavage buffer

Commercially purchased (Fermentas)

Buffer Tango	33 mM Tris-acetate pH 7.9, 10 mM Mg-acetate, 66 mM K-acetate, 0.1 mg/ml BSA

Buffer Green	10 mM Tris/HCl pH 7.5, 10 mM $MgCl_2$, 50 mM NaCl, 0.1 mg/ml BSA
PI-SceI	
Mg^{2+} buffer 8.5	10 mM Tris/HCl pH 8.5, 100 mM KCl, 2.5 mM $MgCl_2$, 1 mM DTT, 10 µg/ml BSA
Mg^{2+} buffer 7.5	10 mM Tris/HCl pH 7.5, 100 mM KCl, 2.5 mM $MgCl_2$, 1 mM DTT, 10 µg/ml BSA
Mn^{2+} buffer 7.5	10 mM Tris/HCl pH 7.5, 100 mM KCl, 2.5 mM $MnCl_2$, 1 mM DTT, 10 µg/ml BSA
KGB	100 mM potassium glutamate, 25 mM Tris-acetate pH 7.5, 10 mM Mg-acetate, 500 µM 2-Me, 10 µg/ml BSA
Fusion enzymes	
KGB optimized	100 mM potassium glutamate, 25 mM Tris-acetate pH 7.5, 0.8 mM Mg-acetate, 100 mM KCl, 0.5 mM 2-Me, 10 µg/ml BSA, for binding experiments Mg-acetate was left out

2.1.7 PLASMIDS

2.1.7.1 Selection assay reporter plasmids

These plasmids are based on pACYC184, in which the gene coding for the toxin gvpA (gas vesicle protein A from *Anabaena flos-aquae*) together with the phage λ promoter P_L was inserted in preliminary work. The target sites for I-SceI and PI-SceI were inserted either in the 5' coding region of *gvpA* or in the promoter region resulting in the plasmids called gvpA_I, gvpA_P, gvpA_PL_I and gvpA_PL_P respectively. The site-directed mutagenesis was done on gvpA_PL_P using the primer described in Table 2.4 to change the PI-SceI target site (gvpA_PL_PA, gvpA_PL_PB, gvpA_PL_PC). These plasmids mediate chloramphenicol resistance.

2.1.7.2 I-SceI and PI-SceI expressing plasmids

The plasmids coding for the homing endonucleases were also created in preliminary work. Both plasmids mediate ampicillin resistance; contain a pBR322 derived origin of replication and the gene for the $lacI^q$ repressor. The expression of the enzymes is controlled

by the *T5 lac* promoter. The plasmid pHisPI-SceI codes for PI-SceI with an N-terminal His_6-tag and pT5-I-SceI codes for I-SceI without any additional tag. These plasmids were the basis for creating the different variants or pools of the corresponding enzymes by site-directed mutagenesis (Table 2.4).

2.1.7.3 Fusion enzyme and PvuII expressing plasmids

The used plasmids for creating the fusion enzymes were pT5-I-SceI(D44S), pRIZ'R.PvuII-GSH, pscPvuII-HGC, pASK IBA63b-plus (IBA) and pPvuII(A)His, where PvuII(A) is a synthetic gene produced by Geneart coding for PvuII, but with a different codon usage. The cloning procedure is described in chapter 2.2.2.5.

2.1.7.4 Substrate plasmids for fusion enzymes

All substrate plasmids are based on pAT153, where the cassettes for the target sites (Table 2.5) were ligated into the BamHI cleavage site, resulting in the plasmids called pAT_S6P, pAT_S4P4S, pAT_S6P6S, pAT_S6x6S and pAT_S8P8S. The additional PvuII cleavage site was introduced via site-directed mutagenesis (Table 2.4). The substrate with only one PvuII site was pATPEB, where also the I-SceI target-site cassette was ligated in the PvuII site resulting in pAT_S. To insert PvuII sites next to this I-SceI site in distinct distances a KpnI cleavage site was inserted via site-directed mutagenesis at different distances and the PvuII target-site cassette was ligated into this site (pAT_S3P, pAT_S7P, pAT_S8P, pAT_S10P, pAT_S12P).

2.2 METHODS

2.2.1 MICROBIOLOGICAL METHODS

For all microbiological approaches LB-medium (10 g/l tryptone, 5 g/l yeast extract, 5 g/l NaCl, pH 7.5) and LB-agar plates (LB-medium with 1.5 % (w/v) agar) were used. The corresponding antibiotics were added after sterilization, before use. Table 2.10 lists the antibiotics and the used concentrations.

Table 2.10

Antibiotic	concentration (µg/ml)	
	LB-medium	agar plate
ampicillin	75	100
chloramphenicol	20	30
kanamycin	25	25
tetracycline	10	10

2.2.1.1 Preparation of electrocompetent cells

Electrocompetent *E.coli* cells for normal use were prepared using the following protocol. A 500 ml LB-medium culture (depending on the cells supplemented with antibiotics) was inoculated with 10 ml of a preculture of the corresponding cells and grown at 37 °C until OD_{600} 0.5-0.8. The cells were cooled down on ice for 30 min and then centrifuged for 15 min at 4200 rpm (Beckman, J6-HC). The cell pellet was washed with cold, sterile 10 % (v/v) glycerol 3-times (250 ml, 150 ml, 40 ml consecutively). After the last centrifugation step the pellet was resuspended in 2 ml 10 % (v/v) glycerol and 80 µl aliquots were shock-frozen with liquid nitrogen and stored at -80 °C.

For the selection system, TGE900 cells were made electrocompetent with the *gvpA* reporter plasmid. The procedure was the same as described above, except that the medium contained 60µg/ml chloramphenicol and cells were grown at room temperature (~ 23 °C).

2.2.1.2 Electroporation

Electrocompetent *E.coli* cells were thawed on ice, mixed with the plasmid DNA (usually 20-100 ng) and filled into an electroporation cuvette. The transformation was done in an electroporator (EasyjecT, EquiBio) at 1250 V, 25 mA and 25 Ω. Immediately after transformation, cells were resuspended in 800 µl LB-medium and incubated for 1 h at 37 °C (if not described differently). Afterwards usually 30-100 µl of cells were spread on an agar-plate containing the corresponding antibiotic and incubated overnight at 37 °C (if not described differently).

2.2.1.3 Selection assay “in liquid”

Electrocompetent TGE900 cells harboring the reporter plasmid were transformed with the plasmid coding for the corresponding homing endonuclease or a variant thereof. After incubation for 1 h at 28 °C, 3 ml LB-medium (+ Amp) were inoculated with 100 µl of the transformed cells, as control 100 µl of the transformation mixture were also spread on agar

plates (+ Amp/Cm) and both were incubated overnight at 28 °C. The next day, the OD_{600} of the liquid culture was measured and two times 25 ml LB-medium (+ Amp) were inoculated with a certain volume of the pre-culture according to a distinct OD_{600}. One 25 ml culture was incubated at 28 °C, the other one at 42 °C with measuring the OD_{600} every 30 min. OD_{600} values were plotted against the time and fit by non-linear regression analysis using the formula: $N = N_0 \times e^{kt}$, where N_0 is the starting OD_{600} and N the measured OD_{600} at the time t. Doubling time (Dt in h) was calculated with the formula: $Dt = \frac{ln2}{k}$.

2.2.1.4 Selection assay "on plate"

TGE900 cells harboring the respective reporter plasmid were transformed with 50 ng of the homing endonuclease expressing plasmid; the transformation mix was resuspended in 600 µl LB-medium and incubated 90 min at 28 °C. 150 µl were spread on two agar plates supplemented with Amp and 0.5 mM IPTG. One of the plates was incubated overnight at 28 °C (growth control) and the other one for 1 h at 28 °C to induce the expression of the homing endonuclease and then incubated overnight at 42 °C. After another 12 h at 28 °C the colonies were counted and the percentage of survival was calculated by dividing this number by the number of colonies on the control plate. To perform more rounds of selection the colonies from the 42 °C plate were pooled and the plasmid DNA was extracted. The possibly still existing reporter plasmid in this DNA preparation was removed by AgeI digestion and again TGE900 cells with the distinct reporter plasmid were transformed with 50 ng of this DNA. This assay was repeated up to four-times.

2.2.1.5 *In vivo* activity test of selected PI-SceI variants

E.coli TGE900 harboring gvpA_PL_PC were transformed with the same amount of plasmid DNA coding for PI-SceI_T225A, PI-SceI_B1, PI-SceI_C1 or PI-SceI_C2 respectively and spread on two agar plates supplemented with Amp and 0.5 mM IPTG. One plate was incubated at 28 °C as growth control and the other one was incubated for 1 h at 28 °C and over-night at 42 °C. The next day, the survival rate was calculated as percentage of colonies grown on the 42 °C plate with respect to the colonies grown on the control plate.

2.2.1.6 *In vivo* test for fusion enzymes

When transforming *E.coli* cells with plasmids coding for restriction enzymes, they usually have to be co-transformed with plasmids expressing the companion methyltransferase to protect the *E.coli* genome against autodigestion. To test if the created fusion enzymes are specific and do not cleave unaddressed PvuII sites anymore, an *in vivo* test was performed.

For this, *E.coli* cells Xl10 Gold either harboring the plasmid for the PvuII methyltransferase or not were transformed with 50 ng of plasmid coding for the fusion enzyme or for PvuII or one of its variants as a control. After incubation of both transformation mixtures (with or without plasmid pLGM coding for M.PvuII) for 1 h at 37 °C, 30 µl were spread on agar plates supplemented with Amp (without pLGM) or with Amp/Kan (with pLGM). The plates were incubated overnight at 37 °C and analyzed the next day.

2.2.2 MOLECULAR BIOLOGY METHODS

2.2.2.1 DNA extraction

Purification of plasmid DNA was done either low-scale with the Promega PureYield™ Plasmid Miniprep System or high-scale with the Promega PureYield™ Plasmid Midiprep System according to the manufacturer's instruction. The Promega Wizard® SV Gel and PCR Clean-Up System was used according to the manual to purify PCR products for EMSA, site-directed mutagenesis or to gel-purify restriction enzyme digested plasmid fragments for ligation. The concentration of nucleic acids was determined by measuring the absorbance at 260 nm in a spectrophotometer NanoDrop® 1000 (Thermo Scientific).

2.2.2.2 Electrophoresis

Polyacrylamide gel electrophoresis

For small DNA fragments from 100 - 1200 bp polyacrylamide gel electrophoresis was used with 6 % polyacrylamide in 1 x TPE (0.9 M Tris-H_3PO_4 pH 8.2, 2 mM EDTA). The samples were mixed with 5x loading buffer (250 mM EDTA pH 8, 30 % (v/v) glycerol, 1.2 % SDS, 0.3 mg/ml bromphenol blue, 0.3 mg/ml xylencyanol) and loaded on the gel. The electrophoresis was performed in 1 x TPE for 1 h at 45 mA, stained with ethidium bromide and bands visualized using the BioDocAnalyze (Biometra) gel documentation system.

Agarose gel electrophoresis

Larger DNA fragments like plasmid DNA or λ-DNA were analyzed on 0.8 % agarose gels in 1 x TPE. The samples were also mixed with loading buffer, loaded on the gel and electrophoresis was performed in 1 x TPE for 1 h at 8 V/cm. DNA was stained with ethidium bromide and visualized with the BioDocAnalyze (Biometra) gel documentation system.

EMSA

The gels for the binding experiments contained 6 % polyacrylamide in 20 mM Tris-acetate pH 8.5. Before loading the samples supplemented with 8 % (v/v) glycerol the gel was pre-run for 30 min at 10 V/cm. The separation was in 20 mM Tris-acetate pH 8.5 for 2 - 3 h at 10 V/cm. Afterwards the gel was transferred to Whatman® paper, dried, wrapped with polythene foil and radioactivity was counted using an IstantImager system (Packard).

SDS-PAGE

Proteins were analyzed using SDS-polyacrylamide gel electrophoresis. The stacking gel contained 6 % polyacrylamide in 130 mM Tris/HCl pH 6.8, 0.1 % SDS and the separating gel contained 10 - 15 % polyacrylamide (depending on the size of the protein to be detected) in 420 mM Tris/HCl pH 8.8, 0.1 % SDS. Samples were mixed with 2x Lämmli-loading buffer (160 mM Tris/HCl pH 6.8, 2 % SDS, 5 % 2-Me, 40 % (v/v) glycerol, 0.1 % bromphenol blue) heated up to 95 °C for 5 min and loaded on the gel, which was run in 1 x SDS buffer (25 mM Tris, 0.19 M glycine, 0.1 % SDS) at 35 mA for 1 h. Afterwards, the gel was washed three times for 5 min in hot water and stained with colloidal coomassie staining solution (0.1 % Coomassie ® Brilliant Blue G250, 2 % (v/v) H_3P0_4, 5 % aluminum sulfate, 10 % ethanol). The bands were visualized using BioDocAnalyze (Biometra) gel documentation system.

2.2.2.3 Polymerase chain reaction (PCR)

PCR was used either to amplify certain genes for cloning approaches, to introduce site-directed mutations into genes or to screen for the accuracy of the two before mentioned processes. When genetic integrity of the product was important *Pfu*-polymerase or even Phusion TM High-Fidelity DNA polymerase (NEB) was used, while for mere screening purposes *Taq*-polymerase was sufficiently accurate. *Taq*- and *Pfu*-polymerase were used in the following activity buffers.

Taq 10x	100 mM Tris/HCl pH 8.8, 500 mM KCl, 1 % Triton X-100, 15 mM $MgCl_2$
Pfu 10x	200 mM Tris/HCl pH 9.0, 100 mM $(NH_4)SO_4$, 100 mM KCl, 1 % Triton X-100, 25 mM $MgSO_4$

The PCR reaction profile and the reaction mixture also differed for both enzymes:

Taq PCR profile
1. Initial step: 3 min at 95 °C
2. Denaturing: 30 sec at 95 °C
3. Primer annealing: 45 sec at 3 °C above T_M of primer with lowest T_M
4. Extension: 1 min/1000 bp at 72 °C
5. Final step: 5 min at 72 °C
Steps 2-4 repeated 29 times

Taq reaction mix
250 µM dNTPs
400 nM forward primer
400 nM reverse primer
1x *Taq* activity buffer
10-50 ng template DNA or resuspended colony
3 U *Taq*-polymerase
ad 30 µl H_2O

Pfu PCR profile
1. Initial step: 5 min at 93 °C
2. Denaturing: 30 sec at 93 °C
3. Primer annealing: 45 sec at 3 °C above T_M of primer with lowest T_M
4. Extension: 1 min/500 bp at 68 °C
5. Final step: 5 min at 68 °C
Steps 2-4 repeated 29 times

Pfu reaction mix
250 µM dNTPs
400 nM forward primer
400 nM reverse primer
1x *Pfu* activity buffer
10-50 ng template DNA
10 U *Pfu*-polymerase
ad 100 µl H_2O

The commercial enzyme Phusion TM High-Fidelity DNA polymerase (NEB) was used according to the manufacturer's manual.

2.2.2.4 Site-directed mutagenesis

To substitute certain amino acids the site-directed mutagenesis protocol reported in [86] was used. Primers were planned to carry the desired mutation and an additional restriction enzyme site (Table 2.4). *Pfu*-PCR was performed by using the plasmid coding for the enzyme to be mutated as template and a second primer which together with the mutagenic one creates a fragment of 150 - 300 bp. The resulting PCR product, so called "megaprimer", was purified (2.2.2.1) and used as primer on the same plasmid template in a second *Pfu*-PCR ("rolling circle" PCR): initial step 5 min 93 °C, 15 cycles of denaturing 1 min 93 °C, annealing 50 sec 60 °C, extension 15 min 68 °C and final step 20 min 68 °C. The obtained PCR product was digested with DpnI for 3 h at 37 °C to remove the template plasmid.

2.2.2.5 Cloning

To transfer genes from one plasmid to another, they were amplified using primers covering the 5'- and 3'-ends of the desired region and carrying specific restriction enzyme sites. The PCR was done with *Pfu* as described before (2.2.2.3). The obtained PCR product was purified (2.2.2.1) and digested with the corresponding restriction enzymes following the respective company's manual. The target vector was also digested with the same enzymes and gel purified (2.2.2.1). For ligation, usually 50 - 100 ng of vector were incubated with a 3 - 5 times molar excess of insert in a total volume of 20 µl together with 2 µl 10 x T4 ligase buffer, 2 µl PEG 4000 (for blunt ends) and 1 µl T4 DNA ligase. The ligation mix was usually incubated overnight at 16 °C.

Cloning of fusion enzyme and PvuII

For fusing scPvuII with its C-terminal His_6-tag to the N-terminus of I-SceI first the T5lac-promoter was cut out of the plasmid pT5-I-SceI(D44S) with XhoI and NsiI and the PCR fragment generated from pscPvuII_HGC with primers scPvuII-I-SceI delprom 5' and scPvuII_KpnI_NsiI_rev was ligated into the cleaved plasmid. Afterwards, the T5lac-promoter gene sequence was inserted again by amplifying it with primer pair T5 lac 5' and T5 lac 3' from pT5-I-SceI and ligating it into the XhoI and SphI cleavage sites. The same was done for the fusion of PvuII to I-SceI. The resulting plasmids coded for scP-$L_{(H)}$-$S_{(D44S)}$ and P-$L_{(H)}$-$S_{(D44S)}$. These plasmids served as template for exchanging Ser44 to Asn and Asp145 to Ala by site-directed mutagenesis on *I-SceI*. The subsequent plasmids coding for scP-$L_{(H)}$-$S_{(D44N,D145A)}$ and P-$L_{(H)}$-$S_{(D44N,D145A)}$ were the basis for further experiments. For easier purification the whole fusion constructs were PCR-amplified with PvuIISceI 5' and PvuIISceI 3'or PvuIISceI short 3' for the full-length version or a ΔC9 version of I-SceI respectively and cloned into the multiple cloning site of pASK IBA63b-plus (IBA). From now on, S* stands for I-$SceI_{(D44N, D145A)}$ and Ss* stands for the ΔC9I-$SceI_{(D44N, D145A)}$. To remove the His_6-tag between PvuII and I-SceI, which served as linker $L_{(H)}$, the whole plasmid was amplified with Oligo_071 and Oligo_072 which introduced the restriction sites for NheI, Pfl23II and BshTI. The resulting PCR fragment was cleaved with Pfl23II and religated, so that there are now the three restriction sites between the genes coding for PvuII and I-SceI: this sequence encodes $L_{(6)}$. To exchange the linker, the plasmids were cleaved with NheI and BshTI and the different linker cassettes (Table 2.6) could be ligated into the plasmid DNA. All site-directed mutagenesis steps to substitute certain amino acids in PvuII were done with the resulting plasmid. To obtain the PvuII gene alone in the same vector background, a

synthetic gene coding for PvuII, but with different codon usage (PvuII(A)), was cut out of the vector pPvuII(A)His using NcoI and Eco47III and cloned into these restriction sites of pASK IBA63b-plus (IBA), resulting in pASK_PvuII(A). This plasmid was then used as template for further site-directed mutagenesis of PvuII alone and for over-expression and purification of Strep-tagged PvuII and variants thereof.

2.2.2.6 Ethanol precipitation

To purify either DpnI-digested "rolling circle" PCR products or ligation mixtures before electroporation, usually ethanol precipitation was used. Therefore the solution (volume V) was kept on ice, 1/3 V of 3 M sodium acetate and 3 V pure ethanol were added and everything was incubated for 30 minutes on ice. Afterwards the sample was centrifuged for 30 min at 4 °C, the pellet washed once with 3 V 70 % ethanol and again centrifuged for 15 min. After removing the supernatant the pellet was dried and then dissolved in 15 µl water. Electrocompetent *E.coli* cells were transformed with 5 µl of the solution, spread on an agar plate supplemented with the corresponding antibiotic and incubated overnight at 37 °C.

2.2.2.7 Screening

To identify successful mutagenesis or cloning events, grown colonies were picked and examined with the help of analytical PCR. This PCR was a *Taq*-PCR described above (2.2.2.3) using a primer pair covering the genetic region of interest. Usually the PCR product was afterwards digested with the marker restriction enzymes, whose site was introduced in the course of cloning or mutagenesis and analyzed by polyacrylamide gel electrophoresis (2.2.2.2). Positive clones were used to inoculate LB-medium containing the corresponding antibiotics, the plasmid-DNA was purified (2.2.2.1) the next day and sent for sequencing to confirm the correctness of the changed gene.

2.2.2.8 Protein over-expression

A single colony harboring the plasmid coding for the protein to express was used to inoculate 25 ml LB-medium (with antibiotics), which was incubated overnight at 37 °C. 5 ml of this preculture were used to inoculate a 500 ml culture, which was then grown at 37 °C until an OD_{600} of ~1 was reached. In case of I-SceI and PI-SceI, protein expression was induced by adding 0.8 mM IPTG and further growth at 37 °C for 3 - 4 h. In case of the fusion proteins and PvuII in pASK vector, protein expression was induced by adding 200 µg/l anhydrotetracycline and further growth at 20 °C overnight. To control the expression in both cases, 500 µl samples before and after induction were taken, centrifuged, the pellet

was dissolved in 20 µl Lämmli-loading dye, heated up to 95 °C for 5 min and analyzed by SDS-PAGE (2.2.2.2). According to the measured OD_{600}, the same amount of cells was loaded. After the induction time, cells were harvested by centrifugation for 20 min at 4200 rpm (Beckman, J6-HC), the pellet was washed with 40 ml 1 x STE (10 mM Tris/HCl pH 8, 100 mM NaCl, 0.1 mM EDTA) and centrifuged again for further 15 min. The cell pellet was either frozen at -20 °C for later purification or used immediately for protein purification.

2.2.2.9 Protein purification

Buffers used for protein purification of the individual proteins are listed in 2.1.6.1 and therefore not described further in the following sections.

I-SceI

Since I-SceI does not carry any purification tag it was purified using a HiTrap™ Heparin HP 1ml column (GE Healthcare). The pellet was thawed on ice in 10 ml lysis/binding buffer, lysed by sonification (Branson sonifier; 6 x 1 min, with a 30 sec break, duty cycle 50 %, output control 5) and centrifuged for 30 min (20,000 rpm Beckman, JA20) to remove cell debris. After equilibrating the column with 5 ml lysis/binding buffer, the clear lysate was loaded and the column was washed with 5 ml washing buffer. Protein was eluted by adding 1ml of the elution buffer each with increasing salt concentration consecutively. Four fractions per salt concentration were collected and 15 µl of each fraction were analyzed on SDS-PAGE (2.2.2.2) for detection of I-SceI which was usually found in fraction 9 (corresponds to elution at 0.5 M NaCl). The protein containing fractions were dialyzed overnight at 4 °C against dialysis buffer.

PI-SceI

PI-SceI carries an N-terminal His_6-tag and by this could be purified using Ni-NTA agarose affinity chromatography. The cell pellet was thawed in 20 ml lysis buffer on ice, lysed by sonification and centrifuged for 30 min (20,000 rpm) to remove cell debris. 400 µl of Ni-NTA agarose suspension were equilibrated with 10 ml lysis buffer for 15 min. Before adding the clear lysate the Ni-NTA agarose had to be centrifuged 5 min at 1000 rpm (Beckman, J6-HC) to remove the buffer. The lysate and Ni-NTA agarose were incubated 1 h at 4 °C, followed by one washing step with 20 ml lysis buffer and one washing step with 20 ml washing buffer, each 10 min. The protein was eluted with 2 x 1 ml elution buffer. Fractions of each purification step were analyzed by SDS-PAGE (2.2.2.2) and protein-containing eluates were dialyzed overnight at 4 °C against dialysis buffer.

scP-L$_{(H)}$-S$_{(D44S)}$

The His_6-tag of this fusion enzyme is located between the two proteins, scPvuII and I-SceI, and for this Protino® Ni-IDA 2000 packed columns (Macherey-Nagel) were used for purification, since purification with these columns results in purer products and a higher yield than Ni-NTA agarose. The purification was performed as described in the manufacturer's manual. The SDS-PAGE following purification showed that the protein was not pure and that the eluates contained also free I-SceI and scPvuII. To remove them, subsequent preparative gel filtration using a Superdex 200 10/300 GL column (Tricorn) on an ÄKTA purifier (Amersham Biosciences) was performed using FPLC gel filtration buffer. Fractions of 500 µl were collected and purity of the fractions, showing an absorbance at 280 nM in the chromatogram, was checked with SDS-PAGE (2.2.2.2). These fractions containing the pure fusion enzyme were pooled and dialyzed overnight at 4 °C against dialysis buffer.

All other fusion enzymes and PvuII variants

Since the genes coding for these enzymes were cloned into pASK IBA63b-plus (IBA) all these proteins carry a C-terminal Strep-tag, this is why they could be purified by using Strep-Tactin® Sepharose®. The cell pellet was thawed on ice in 20 ml lysis/washing buffer, lysed by sonification and centrifuged for 30 min (20,000 rpm) to remove cell debris. 2 ml Strep-Tactin® Sepharose® were equilibrated with 5 ml lysis/washing buffer, the clear lysate was applied and afterwards the column was washed once with 3 ml lysis/washing buffer and once with 3 ml washing buffer. The protein was eluted with 500 µl elution buffer first and then with 4 x 1 ml elution buffer. The purity of the eluates was controlled by SDS-PAGE (2.2.2.2) and fractions containing the pure enzyme were dialyzed overnight at 4 °C against dialysis buffer.

To prepare the enzyme P-L$_{(6)}$-Ss* for analytical ultracentrifugation, the fractions containing the protein were further purified by preparative gel filtration using a Superdex 200 10/300 GL column (Tricorn) on an ÄKTA purifier (Amersham Biosciences) after elution from Strep-Tactin® Sepharose®. The protein-containing fractions were pooled and one half was dialyzed overnight at 4 °C against high-salt dialysis buffer (500 mM NaCl) and the other half against low-salt dialysis buffer (100 mM NaCl).

Determination of protein concentration

The concentration of the proteins was determined by measuring the absorption at 280 nm in a spectrophotometer NanoDrop® 1000 (Thermo Scientific) and using the

Lambert-Beer-Law ($A_{280} = c \times d \times \varepsilon$). The extinction coefficient ε of the proteins was calculated based on the amino acid content [87]. Protein concentrations are given as dimer concentrations throughout the text.

2.2.2.10 Analytical ultracentrifugation

To determine the oligomeric state of P-$L_{(6)}$-Ss* analytical ultracentrifugation was performed at the Hannover Medical School in the lab of PD Dr. Ute Curth in an analytical ultracentrifuge (Beckman Coulter Optima XL-A).

2.2.2.11 Activity assays

The following described activity assays for the individual proteins were partially performed in buffers described in 2.1.6.2. In some cases, the following described protocols are just basic descriptions which could differ in incubation time or concentration, which is then mentioned in the results section.

I-SceI

To compare the activity of the different I-SceI variants, 8 nM plasmid containing one I-SceI site were incubated with 1 µl of eluate 9 of each I-SceI variant in Buffer Tango at 37 °C. Since the purification of I-SceI did not result in pure enzyme and by this the concentration of I-SceI could not be measured, the same volume of eluate was always used. After 10, 30 and 60 min consecutively 10 µl of the reaction were withdrawn; the reaction was stopped by adding 5 x loading buffer and the reaction products analyzed on 0.8 % agarose gels.

PI-SceI

The activity of different PI-SceI variants was compared by incubating 3 nM plasmid DNA containing one PI-SceI site with 5 µM of each enzyme in Mg^{2+} buffer 8.5 (2.1.6.2). After 10, 30 and 60 min samples were withdrawn consecutively and the reaction was stopped by adding 5x loading buffer and analyzed on 0.8 % agarose gels. To test the activity of the newly selected variants of PI-SceI, DNA cleavage assays were performed. Usually 3.7 nM of plasmid DNA (gvpA_PL_P, gvpA_PL_Pb or gvpA_PL_Pc) were incubated with 100 nM protein for 1 h at either 37 °C or 42 °C in different buffers listed in 2.1.6.2. The reactions were stopped by adding 5x DNA loading buffer and analyzed on 0.8 % agarose gels.

scP-$L_{(H)}$-$S_{(D44S)}$ characterization and buffer determination

The first activity assay of scP-$L_{(H)}$-$S_{(D44S)}$ was performed with 50 nM of a 454 bp PCR fragment containing only an I-SceI site incubated with either 5U I-SceI (Fermentas) in 1x Buffer Tango, 100 nM PvuII or 100 nM scP-$L_{(H)}$-$S_{(D44S)}$ in Buffer Green for 45 min at 37 °C. Cleavage products were separated on a 6 % polyacrylamide gel (2.2.2.2), stained with ethidium bromide and visualized on a BioDocAnalyze system (Biometra). The cleavage products of scP-$L_{(H)}$-$S_{(D44S)}$ were gel-purified (2.2.2.1) and sent for sequencing.

To compare the activity with scPvuII and determine the best distance between the I SceI and the PvuII target site a kinetic DNA cleavage assay with the plasmids pAT_S3P, pAT_S8P, pAT_S10P and pAT_S12P was performed. For this, 8 nM plasmid DNA were incubated with 800 pM of either scPvuII or scP-$L_{(H)}$-$S_{(D44S)}$ in Buffer Green without $MgCl_2$ for 10 min at 37 °C. Then 10 mM $MgCl_2$ was added and after 1, 2, 5, 10 min samples were withdrawn consecutively from the reaction mixture and mixed with 5x DNA loading buffer. The samples were analyzed on 0.8 % agarose gels. To test if the fusion enzyme would also be active under nearly physiological conditions, KGB (2.1.6.2) with different concentrations of Mg-acetate (0.2, 0.4, 0.6, 0.8 and 1 mM consecutively) was tested as reaction buffer. 8 nM pAT_S8P were incubated with 400 pM scPvuII or scP-$L_{(H)}$-$S_{(D44S)}$ in KGB with 100 mM KCl for 10 min at 37 °C and then Mg-acetate was added, followed by further incubation for 20 min at 37 °C. The reaction was stopped by adding 5 x DNA loading buffer and analyzed on 0.8 % agarose gels.

PvuII activity in different buffers

The activity of PvuII was tested by incubating 8 nM PvuII with 8 nM plasmid DNA containing one PvuII site in either recommended Buffer Green (Fermentas), in KGB with 100 mM KCl and 10 mM magnesium acetate or KGB with 100 mM KCl and 0.8 mM magnesium acetate (optimized KGB (2.1.6.2)) for 3 h at 37 °C. The reaction was stopped by adding 5 x loading buffer and analyzed on 0.8 % agarose gels.

Plasmid cleavage assay by fusion enzymes

Activity assays were performed, if not described differently, by incubating equimolar amounts of fusion enzyme and DNA (usually 8 nM each) in optimized buffer conditions (described in 2.1.6.2) for a certain time at 37 °C. Reaction was stopped by adding 5 x loading buffer and analyzed on 0.8 % agarose gels. Kinetic experiments were done by withdrawing

samples from the reaction mixture after distinct time intervals and mixing them with 5 x loading buffer to stop the reaction.

Specificity test on plasmid DNA

Plasmid pAT_PEB which contains only one PvuII site or plasmid pAT_S6P6S_P which contains the addressed tripartite site and an additional PvuII site were linearized using HindIII according to the manufacturer's description and purified using Promega Wizard® SV Gel and PCR Clean-Up System. 8 nM of linearized plasmid were incubated either with 8 nM PvuII in Buffer Green (Fermentas) or with 8 nM of P-$L_{(H)}$-Ss*, P-$L_{(6)}$-Ss*, P-$L_{(N)}$-Ss*, P-$L_{(+)}$-Ss* or P-$L_{(-)}$-Ss* in optimized KGB overnight at 37 °C. The reaction was stopped by adding 5 x loading buffer and analyzed on 0.8 % agarose gels.

Product release test

To test if the fusion enzymes P-$L_{(6)}$-Ss* and P-$L_{(+)}$-Ss* remain bound to the cleavage product after cleavage, 2 nM of an internally [α^{32}P]dATP-labeled PCR fragment containing the tripartite recognition site S6P6S was incubated with 80 nM of fusion enzyme in KGB containing 10 mM Mg-acetate for 30 min at 37 °C. One half of the cleavage reaction was stopped by heating for 10 min to 80 °C and the other one not. Samples were loaded onto a 6 % polyacrylamide Tris-Acetate (pH 8.5) gel, which was run for 2 h at 10 V/cm; to find out whether stable enzyme-DNA product complexes are formed. The gels were analyzed using the InstantImager system (Packard).

Bacteriophage λ-DNA digestion

For bacteriophage λ-DNA (contains 15 PvuII sites) digestion, the DNA concentration was chosen in a way that 8 nM of PvuII sites (corresponds to 564 pM λ-DNA) were incubated with 8 nM fusion enzyme or PvuII in optimized KGB (2.1.6.2) or Buffer Green respectively for 3 h at 37 °C. The reaction was stopped by adding 5 x loading buffer and analyzed on 0.8 % agarose gel.

Molecular beacon assay

The beacon assay for determining kinetic parameters of PvuII variants was performed in accordance to [88]. The molecular beacon 5' (6-Fam) CGCAGCTGGGAC$(T)_{16}$GTCC CAGCTGCG (Dabcyl) 3' was purchased from Biomers. For steady-state experiments 1 nM of the respective PvuII variant was used to cleave beacon at concentrations ranging from 1 - 50 nM in Buffer Green. The fluorescence increase due to cleavage was measured with an

InfiniteF200Pro plate-reader (Tecan) for 5 min every 5 sec. Initial velocity was calculated by linear regression of the first 20 measurement points and then plotted against beacon concentration. K_M and k_{cat} were determined by fitting this graph with non-linear regression using the formula $v = \frac{v_{max} \times [S]}{K_M + [S]}$, with v being the initial velocity, *[S]* the beacon concentration and v_{max} the maximal velocity, which was divided by the protein concentration to get k_{cat}.

Competition cleavage experiments

To test the specificity of the fusion enzymes for the addressed site, radiolabeled PCR fragments using [α^{32}P]dATP containing either the addressed site (S6P6S) or the unaddressed site (P) were created by *Taq*-PCR (2.2.2.3). These fragments, each 10 nM, were both incubated with 10 nM of the fusion enzymes in optimized KGB (2.1.6.2) at 37 °C and after defined time intervals samples were withdrawn and analyzed on 6 % polyacrylamide gels (2.2.2.2). Cleavage products were analyzed with the InstantImager system (Packard) and quantified with InstantImager software. The percentage of cleavage was then plotted against time. Initial velocities were calculated by linear regression analysis.

2.2.2.12 Binding Assays

Anisotropy

The binding constant determination via anisotropy was done by annealing oligonucleotide Sce6PvuII low with oligo 5' (Hex) TTACGCTAGGGATAACAGGGTAAT ATAGGTCAGCTGGACGCA 3' to have a fluorescently labeled DNA fragment containing an I-SceI site and 6 bp away a PvuII site. The fluorescence anisotropy with an excitation at 535 nm and an emission at 556 nm was measured in a FluoroMax®4 spectrofluorimeter (HORIBA Jobin-Yvon). The excitation and emission slits were set for 10 nm band widths and for every protein concentration five measurements were taken, which were in turn the mean of data points taken in 5 seconds. A cuvette (Quartz Ultra-Micro Cell, Hellma) was filled with 100 µl solution containing 2 nM Hex-labeled DNA in 1x KGB (2.1.6.2) without magnesium-acetate to prevent cleavage. After measuring the starting anisotropy 1 µl fusion enzyme, which was diluted in 1 x reaction buffer, was added to the final concentrations of 1.0; 2.0; 3.9; 5.8; 9.6; 19.1; 56.4; 93.4 and 166.8 nM consecutively. The anisotropy values were plotted against the protein concentration and data were fitted using non-linear regression analysis using the formula: $AB = \frac{A+B+K_D}{2} - \sqrt{\left(\frac{A+B+K_D}{2}\right)^2 - A \times B}$, with AB: the amount of bound DNA, A: the protein concentration and B: the DNA concentration.

Electrophoretic mobility shift assay (EMSA)

For determination of binding constants for the fusion enzymes, EMSAs with radioactively labeled PCR fragments were performed. The shift fragments were created via *Taq*-PCR (2.2.2.3) using [α^{32}P] dATP. One fragment contained the addressed site (S6P6S) and one just an I-SceI site (I). For K_D-determination 2 nM of the radiolabeled substrate were incubated with fusion enzyme concentrations ranging from 1 - 150 nM in optimized KGB (2.1.6.2) without magnesium for 30 min at room temperature. To inhibit unspecific interaction, poly (dI-dC) was also added to the reaction at a final concentration of 10 µg/ml. For the unaddressed substrate the enzyme concentration ranged from 20 - 180 nM. After adding 1 µl 87 % glycerol the samples were loaded onto a 6 % polyacrylamide Tris-acetate (pH 8.5) gel (2.2.2.2) and run for 2 h at 10 V/cm. The bands were visualized using the InstantImager system (Packard). The band intensities for free and bound DNA were measured by the InstantImager software and the percentage of bound DNA was calculated. The data were fitted and the K_D determined using non-linear regression analysis, as described above.

3. RESULTS

In order to find highly specific nucleases which could be used in gene targeting, two different approaches have been used. The first deals with changing the specificity of a naturally occurring meganuclease, namely PI-SceI, by directed evolution experiments and the second is a more rational approach by fusing the homing endonuclease I-SceI to the type IIP restriction endonuclease PvuII and thus extending its specificity.

3.1 DIRECTED EVOLUTION OF HOMING ENDONUCLEASE PI-SCEI

Based on already published selection systems for directed evolution of homing endonucleases [17-19], a two-plasmid selection assay was developed and optimized in preliminary work (Diploma thesis Ines Fonfara, Giessen 2008).

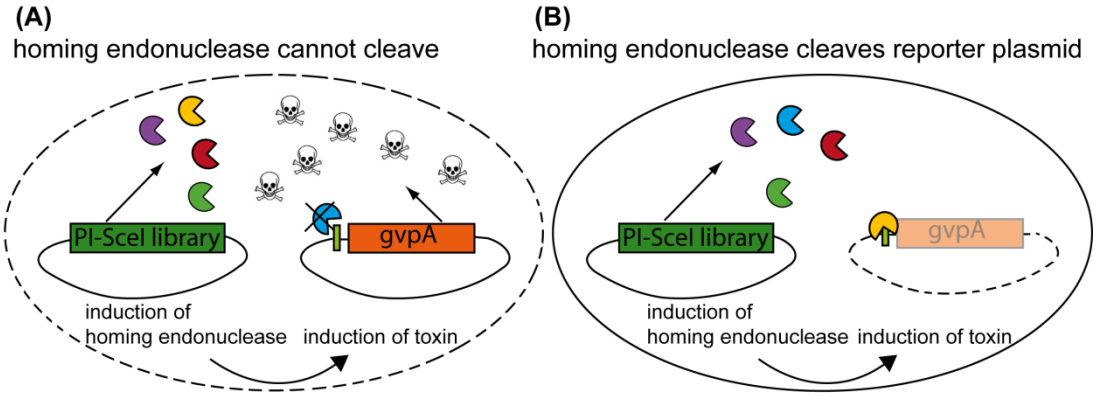


***Figure 3.1: Two-plasmid selection assay developed for selection of PI-SceI variants**. (A) None of the variants in the PI-SceI library is able to cleave the target site on the reporter plasmid and the toxin can be expressed, leading to cell death. (B) A variant of the PI-SceI library can cleave the target site and the linearized reporter plasmid will be degraded by the RecBCD system, leading to survival of the* E.coli *cell.*

This assay relies on the presence of two plasmids which were introduced into *E.coli* cells (Figure 3.1). One plasmid is the reporter harboring a gene coding for a toxin and the respective target site of the homing endonuclease. The other plasmid encodes the homing endonuclease gene or variants thereof. The toxin expression is induced by a temperature shift from 28 °C to 42 °C. If one of the encoded homing endonuclease variants is able to cleave the target site on the reporter plasmid, which will be degraded afterwards by the *RecBCD* system, the cells will survive incubation at 42 °C, meaning toxin expression. If none of the variants can cleave the target site, under inducible conditions the toxin will be expressed and the cells will die.

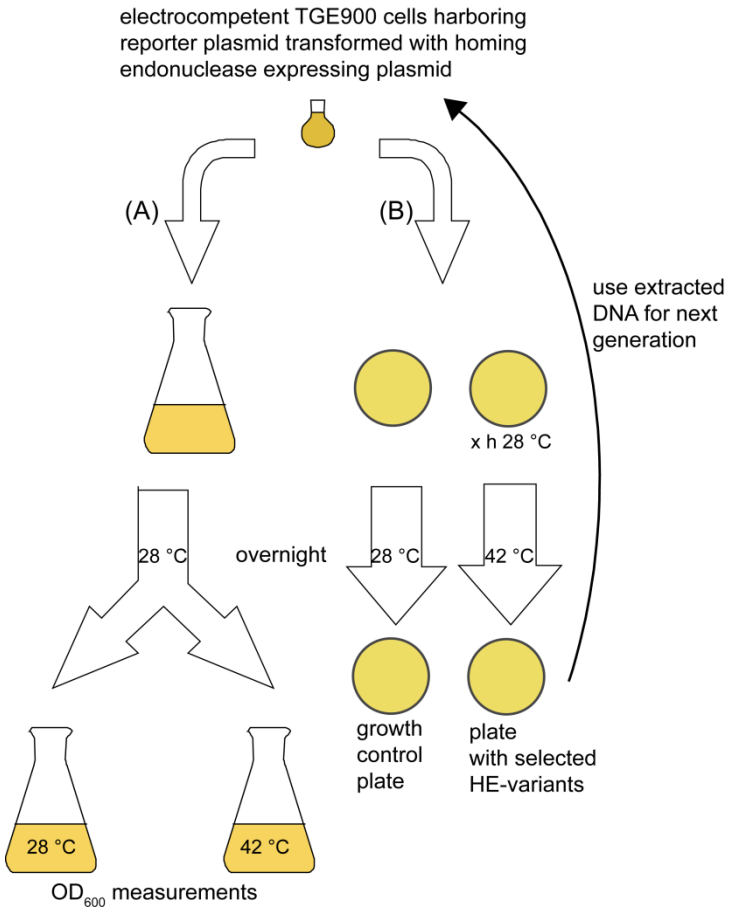


Figure 3.2: Flowchart describing the two-plasmid selection assays used in this work. *The first tests were done by the selection in liquid medium (A), where the growth of the cells under inducible conditions (42 °C) was measured directly. The final selection assay to select for variants of homing endonucleases, which cleave new sequences, was performed on agar plates (B).*

To investigate the limitations and feasibility of this two-plasmid system, the first experiments were done in liquid medium (Figure 3.2A), because the survival of the cells could be measured in real time. For this, cells harboring the reporter plasmid containing a respective target site were transformed with plasmid expressing the homing endonuclease. The transformation mixture was used for inoculating a 28 °C overnight pre-culture, which was divided the next day into two flasks, where one was incubated at non-inducible conditions (28 °C) and the other one at inducible conditions (42 °C). The OD_{600}, which is an indicator for cell growth, was measured every 30 min and doubling time was calculated (see 2.2.1.3). Since the selection assay should be used later on to find enzyme variants with new specificities, an assay "in liquid" is not suitable. For this, an assay "on plate" (2.2.1.4) was developed (Figure 3.2B). Again, cells harboring the reporter plasmid with the desired target site were transformed with plasmid DNA coding for the respective homing endonuclease or a library of homing endonuclease variants. The transformation mixture was spread on two agar-plates, where one was incubated overnight at non-inducible conditions (28 °C) to control the

growth and transformation efficiency. The other plate was incubated for a certain time at 28 °C to give the homing endonuclease some time to cleave the reporter plasmid if possible. Afterwards, the plate was incubated under inducible conditions (42 °C) to induce toxin expression. Only the cells expressing a functional homing endonuclease should be able to grow and could be used for further rounds of selection or investigation of the expressed enzyme.

3.1.1 TWO-PLASMID SELECTION ASSAY "IN LIQUID"

3.1.1.1 I-SceI

Until now, it was not possible to develop a selection system for the directed evolution of PI-SceI. All known assays are based on I-SceI, which is called the "gold-standard" [56] in terms of stimulation of homologous recombination. To determine the limitations of the developed selection assay, different I-SceI variants with stepwise decreased activity compared to the wild type were created. The substituted amino acids are depicted in Figure 3.3 [19] and mutagenesis was performed as described in 2.2.2.4. Variant I-SceI_D44S is the negative control, since the catalytically active aspartate from the first LAGLIDADG motif is mutated to serine which makes the enzyme inactive. The variant I-SceI_N152K was used, since it was reported to have ~ 80 % of the wt-activity and the double variant I-SceI_ D150C, N152K was constructed to decrease the activity of the enzyme even more.

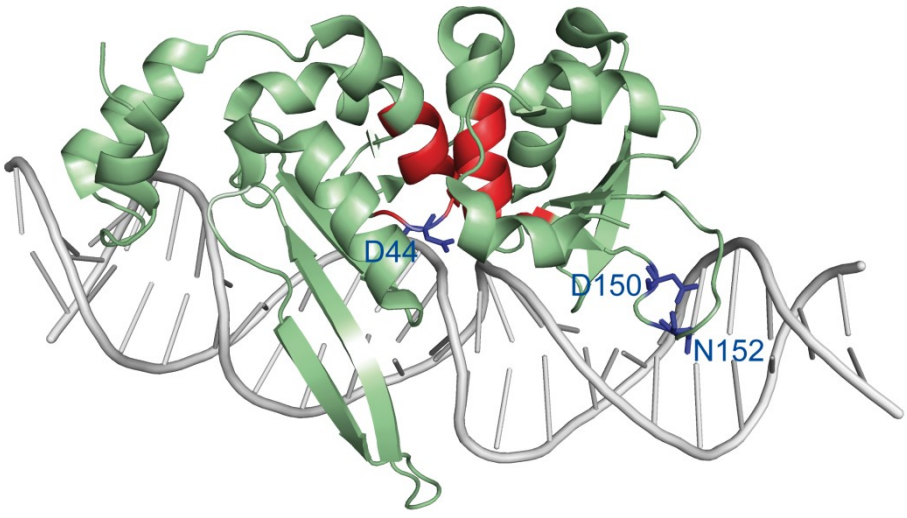


Figure 3.3: Co-crystal structure of I-SceI *(pdb: 1r7m). The LAGLIDADG helices harboring the active site with Asp 44 and Asp 145 are indicated in red. The changed amino acids are shown in blue.*

The listed variants were created, expressed and purified (2.2.2.9) and their *in vitro* activity was tested on plasmid DNA (2.2.2.11) containing an I-SceI site (Figure 3.4A). After

20 min I-SceI_wt cleaved 100 % of the substrate, while I-SceI_N152K cleaved 80 %, I-SceI_D150C,N152K ~ 40 % and I-SceI_D44S did not cleave at all after 1 h incubation. Since now enzyme variants with stepwise decreased activity were available, the selection assay could be tested for its limitations. For this, the variants of I-SceI were used in the two-plasmid selection assay "in liquid" (Figure 3.4B). The doubling times of the active variants (I-SceI_wt; I-SceI_N152K; I-SceI_D150C,N152K) were nearly the same at both temperatures, meaning that the cells grew normally at 28 °C as well as at 42 °C when toxin expression was induced. Only the inactive variant I-SceI_D44S showed a ~ 5 - fold increased doubling time, meaning the cells did not grow properly, which could be due to the expressed toxin. The doubling time under inducible conditions of I-SceI_D150C, N152K, the variant with the lowest *in vitro* activity, was 1.5 - fold lower than the activity of I-SceI_wt. An over 50 % decreased *in vitro* activity does not necessarily reflect the activity *in vivo* under selection pressure.

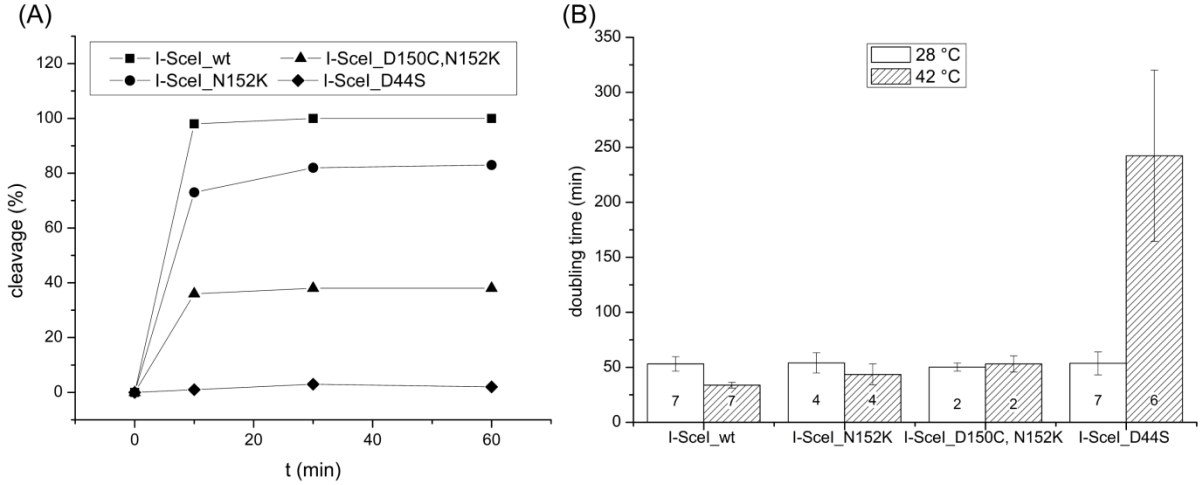


Figure 3.4 : Comparison of* in vitro *versus* in vivo *activity of I-SceI variants. *The* in vitro *activity is presented as percentage of cleavage of plasmid DNA containing an I-SceI target site plotted against time (A). The* in vivo *activity is indirectly measured as doubling time at 42 °C in the "in liquid" selection system (B).White bars represent the doubling time of cells containing the reporter plasmid gvpA_PL_I and the plasmid expressing the respective I-SceI variant at 28 °C. Shaded bars represent the doubling time at 42 °C. The respective number of experiments is indicated by the number inside the bars.*

3.1.1.2 PI-SceI

It was decided to test PI-SceI as well in this selection system. To have a control variant, the active site aspartate from the first LAGLIDADG motif was also mutated to alanine (PI-SceI_D218A, Figure 3.5). Furthermore, a variant with an improved catalytic activity compared to the wild type was created (PI-SceI_T225A, [89]).

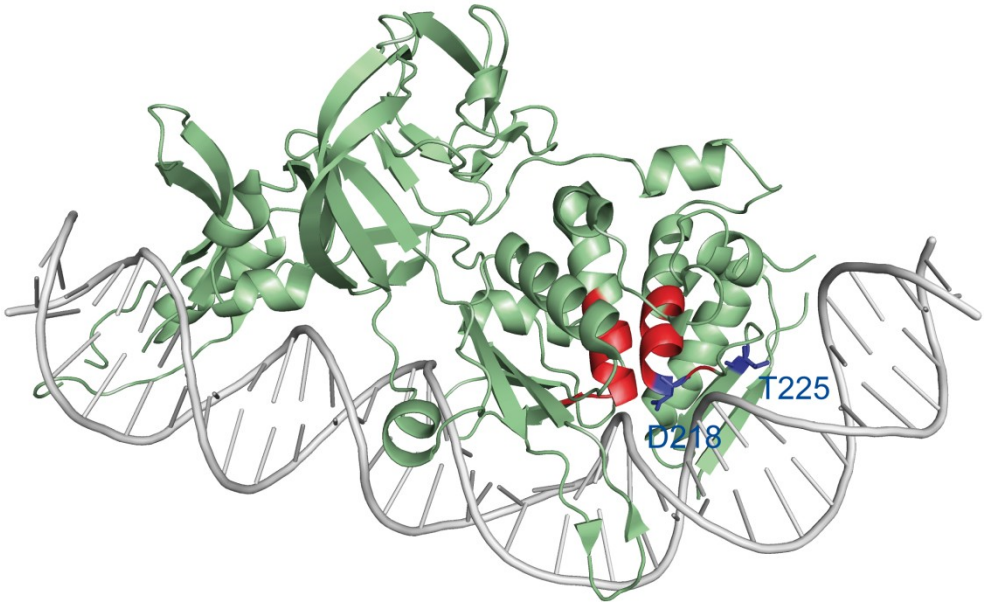


Figure 3.5: Co-crystal structure of PI-SceI *(pdb:1lws) The LAGLIDADG helices harboring the active site Asp 218 and Asp 316 are colored red. The mutated residues are indicated in blue.*

The three variants PI-SceI_wt, PI-SceI_D218A and PI-SceI_T225A were expressed and purified (2.2.2.9) and the *in vitro* activity of PI-SceI_wt and PI-SceI_T225A was tested on plasmid DNA (2.2.2.11) containing one PI-SceI site (Figure 3.6A). PI-SceI_T225A cleaved 100 % of the substrate after 1 h incubation, while PI-SceI_wt cleaved ~ 60 % in the same time, which is consistent with the 3-fold increased activity of PI-SceI_T225A reported before [89]. PI-SceI_D218A did not show any cleavage activity in comparable assays (data not shown).

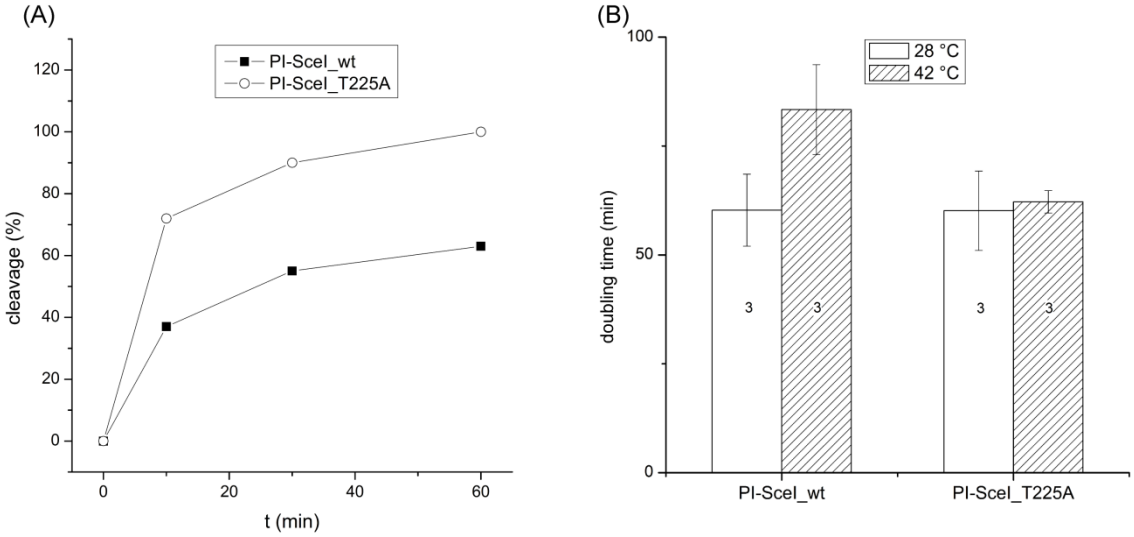


Figure 3.6: Comparison of* in vitro *and* in vivo *activity of PI-SceI_wt and PI-SceI_T225A. *The percentage of cleavage of plasmid DNA containing one PI-SceI site representing the* in vitro *activity is shown in (A) and the* in vivo *activity represented by the doubling time during the "in liquid" selection assay using the reporter gvpA_PL_P is shown in (B). Labeling is according to Figure 3.4.*

To test the *in vivo* activity of both PI-SceI variants and their suitability for the developed selection system, the same selection assay "in liquid", as described above, for I-SceI, was performed. This time cells harboring the reporter plasmid with the PI-SceI site in the promoter region were transformed with plasmids coding for the PI-SceI variants. The obtained doubling times shown in Figure 3.6B confirmed the *in vitro* data. At 28 °C the doubling times were comparable to these from I-SceI, but at inducible conditions the doubling time of cells expressing PI-SceI_wt was increased, whereas the doubling time of the more active variant PI-SceI_T225A was comparable to that of I-SceI_D150C, N152K.

3.1.2 TWO-PLASMID SELECTION ASSAY "ON PLATE"

3.1.2.1 Determining the parameters

In the assay developed in preliminary work, the target site for the homing endonuclease was inserted in the 5' coding region of the toxic reporter gene (gvpA_P). To see if there is an increased effect, this target site was also inserted in the promoter region of the reporter plasmid (gvpA_PL_P). Both reporter plasmids were used in a selection assay "on plate" to determine which has the best efficiency. Furthermore in this first assay it has been tested how long the incubation time of the plate at 28 °C should be, before the toxin expression at 42 °C is induced (Figure 3.7A). Since PI-SceI_T225A was the better variant in the previous test, cells harboring one of the two reporter plasmids were transformed with DNA coding for this enzyme variant. The transformation mixture was spread as described above on one control plate which was incubated overnight at 28 °C and on four other plates containing also 0.5 mM IPTG to induce homing endonuclease expression. These four plates were incubated for 1 - 4 h at 28 °C and than overnight at 42 °C. The position of the target site has a large effect on the survival of cells under inducible conditions: when the target site is located in the promoter region, the cells were able to grow on the plate incubated at 42 °C, while when the target site is located in the coding region of the toxin, only a small number of cells grew on these plates. Furthermore, the incubation time at 28 °C before inducing the toxin expression seems to be an important parameter. The longer this incubation, the more cells can grow. This may be due to the fact that expressed PI-SceI has more time in the cell to cleave and by this destroys the reporter plasmid.

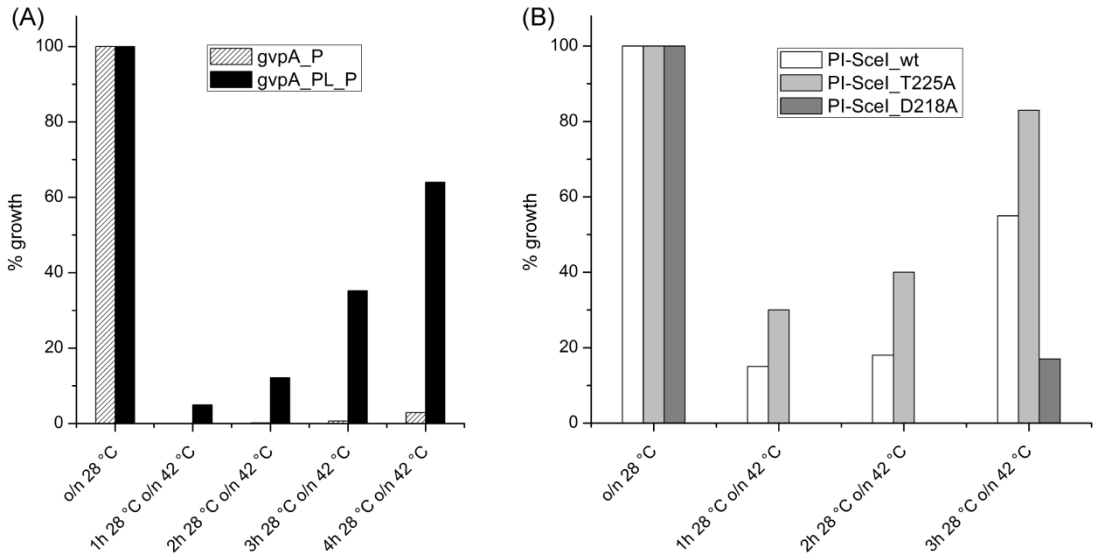


Figure 3.7: Comparison of different reporter plasmids and PI-SceI variants for the selection assay "on plate". *(A) TGE900 cells containing one of two reporter plasmids harboring the PI-SceI recognition site either in the 5' coding region of gvpA (gvpA_P, shaded bars) or in the promoter region (gvpA_PL_P, black bars) were transformed with plasmid coding for PI-SceI_T225A. The number of colonies on the control plate grown at 28 °C was set to 100 % growth and the percentage of colonies on the other plates was calculated accordingly. (B) Cells containing the reporter plasmid gvpA_PL_P were transformed with plasmid coding for the different PI-SceI variants PI-SceI_wt (white bars), PI-SceI_T225A (light gray bars) and PI-SceI_D218A (gray bars), the percentage of growth was calculated as for (A).*

To confirm that PI-SceI_T225A would be the best starting point for further selection experiments, all three variants of PI-SceI, namely PI-SceI_wt, the more active variant PI-SceI_T225A and the inactive variant PI-SceI_D218A were also tested in the selection assay "on plate" (Figure 3.7B). This assay was performed with the reporter plasmid gvpA_PL_P which showed the best results in the previous experiment. Again, the cells transformed with plasmid coding for PI-SceI_T225A showed a higher survival under inducing conditions (42 °C) compared to wild type which corresponds to the lower doubling time and the higher *in vitro* cleavage activity of this enzyme (Figure 3.6). When the plates were incubated for 3 h at 28 °C before inducing the toxin expression, the cells transformed with plasmid coding for the inactive variant PI-SceI_D218A also showed ~ 20 % surviving colonies, which could be due to a loss of the reporter plasmid. Summing the so far obtained results up, further experiments will be done with reporter plasmid containing the target site in the promoter region, variant PI-SceI_T225A as starting point for mutagenesis and 1 h incubation of the plate at 28 °C before incubating at 42 °C overnight, to start the toxin expression.

3.1.2.2 Validation of the assay

The goal of this assay would be to select single variants of a library which are able to cleave a certain target site. To imitate this situation, the selection assay "on plate" was performed with a mixture of plasmids coding for the active PI-SceI_T225A and the inactive PI-SceI_D218A, while the ratio of active to inactive was one to hundred. This mixture was used to transform *E.coli* harboring gvpA_PL_P and the assay was performed as described above. The next day the colonies from the 42 °C plate were pooled, plasmid DNA was extracted (2.2.2.1) and a sample digested with EheI, which specifically cleaves the plasmid pHisPI-SceI_T225A only once and pHisPI-SceI_D218A twice. By this procedure, the plasmids having the same size could be separated on an agarose gel (Figure 3.8A, red and green bars). The extracted DNA was then used for another round of selection assay (see also Figure 3.2). Again the next day the colonies grown on the 42 °C plate were pooled, DNA extracted and the same procedure as described above was performed four times. The agarose gel in Figure 3.8A shows the EheI-digested samples of the DNA extraction after every round of generation, and it is obvious that the amount of pHisPI-SceI_T225A (red bar) increased in every generation, while the amount of pHisPI-SceI_D218A (green bar) decreased.

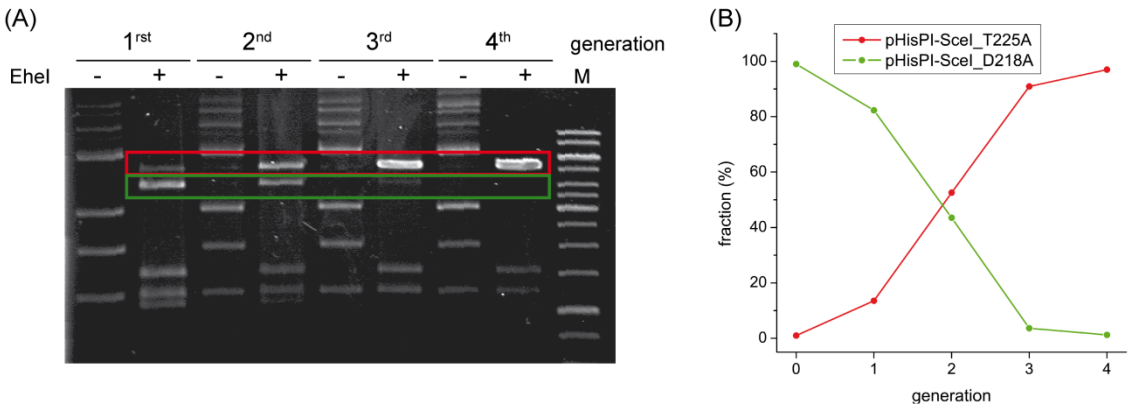


Figure 3.8: Enrichment of pHisPI-SceI_T225A trough four generations of selection assay "on plate". *(A) Cells harboring the reporter plasmid gvpA_PL_P were transformed with a mixture of pHisPI-SceI_T225A and pHisPI-SceI_D218A in a molar ratio of 1:100. After every round of selection assay, the colonies grown at 42 °C were pooled, DNA extracted, a sample digested with EheI (cleaves pHisPI-SceI_T225A once, pHisPI-SceI_D218A twice) and analyzed on an agarose gel. Then the next round of selection was started by using the previously extracted DNA for transformation of TGE900 cells (gvpA_PL_P). pHisPI-SceI_T225A which is linearized by EheI is indicated by a red bar, pHisPI-SceI_D218A which is cleaved twice by EheI is indicated by a green bar. M: Gene Ruler ™ 1kb DNA ladder (B) The fractions of pHisPI-SceI_T225A and pHisPI-SceI_D218A in percent plotted over four generations, while generation 0 represents DNA used for the first transformation containing 1 % pHisPI-SceI_T225A and 99 % pHisPI-SceI_D218A.*

To make this effect clearer the percentage of both plasmids on the total amount of extracted DNA were calculated for every generation (Figure 3.8B), while generation 0 means

the input, i.e. 1% pHis_PI-SceI_T225A and 99 % pHisPI-SceI_D218A. It can be seen, that already after two generations the relation between both plasmids was inverted and the plasmid coding for the active form of PI-SceI was enriched after 3 generations.

3.1.2.3 Selection of PI-SceI variants with new specificity

Since the two-plasmid selection assay was now tested for feasibility to select for new PI-SceI variants, the next step would be the creation of enzyme variant libraries and creation of new target sites to select for. One approach to create mutant libraries of enzymes is by error-prone PCR [57,90], where the whole gene coding for the enzyme is mutated randomly. Since this would result in a large library of variants, where the catalytic center of the enzyme could also be affected and by this the number of inactive variants would be too high, it was decided to do a semi-rational approach. Only a few residues, known from the co-crystal structure to be involved in base pair contacts to the target site, were mutated using degenerated primers [91]. The main principle of this technique is, that the primer (Table 2.4) used to amplify the area of interest contains any possible nucleotide (NNN) at the position where a certain amino acid should be mutated randomly. Thus at this position all combinations of the 4 nucleotides are possible. This could result, in theory, in every possible amino acid at this position. In this approach, the triplet "NNK" was chosen in the degenerated primers VDE_ran_Q55H56 and VDE_ran_R90_L92_R94, where "N" stands for A, G, T or C and "K" for G or T and "MNN" in the primer VDE_ran_S169H170 where "M" stands for A or C. By using these triplets the probability of creating stop codons (TGA, TAA, TAG) is decreased. Three groups of 2 - 3 residues have been chosen according to published binding studies (summarized in [92]). These groups are called pool QH (residues Q55, H56; orange), pool RLR (residues R90, L92, R94; yellow) and pool SH (residues S169, H170; magenta) and their position in the PI-SceI enzyme and respective contacts to the DNA are indicated in Figure 3.9. Pool QH was generated for later approaches, where the changes in the recognition site were planned to be extended. All these residues were also chosen because they are situated in another domain than the active site residues D218 and D326, and thus the catalytic activity of PI-SceI is not affected. According to the marked binding positions of the designated amino acid residues, these base pairs of the target site were changed to create the new targets B and C as described in 2.2.2.4 using the primers listed in Table 2.4 respectively (Figure 3.9B), which were later on called gvpA_PL_PB and gvpA_PL_PC. The base pair changes are based on studies of [93], where they reported the lowest cleavage efficiency of PI-SceI_wt on these changed nucleotides.

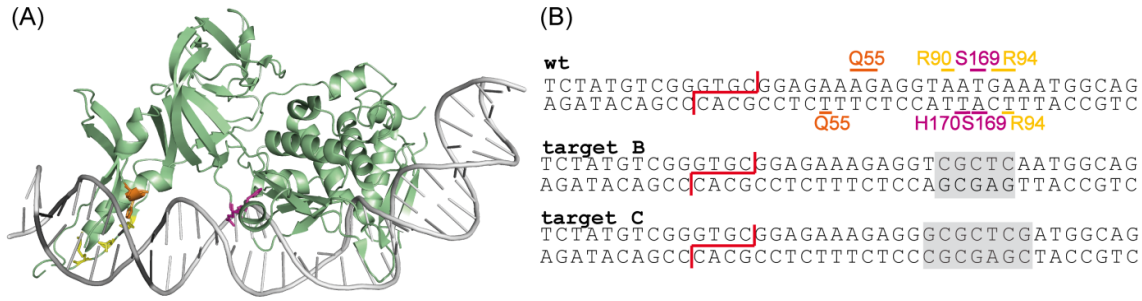


Figure 3.9: Overview of amino acids used for creating pools of PI-SceI and the new target sites. *(A) Co-crystal structure of PI-SceI with annotated amino acids to create pool QH (residues Q55, H56; orange), pool RLR (residues R90, L92, R94; yellow) and pool SH (S169, H170; magenta), respectively. (B) PI-SceI wt-target and the two new targets B and C with the cleavage site indicated by a red line. The known contacts formed by the amino acids to be mutated are indicated in the upper panel (color coding according to (A)). The base pairs changed in the new targets B and C are shaded gray.*

The mutagenesis was done as described in 2.2.2.4 by using pHisPI-SceI_T225A as template and the primers reported in Table 2.4. The created libraries were sent for sequencing to validate the success of the mutagenesis and to see if the nucleotides are equally distributed or if there was a bias. Figure 3.10 shows the summary of these results and except for some nucleotide positions the distribution was mostly equal. It has to be mentioned again, that for pool QH and pool RLR the last position of every triplet should be only guanine or thymine and for pool SH the first position of every triplet should be only adenine or cytosine.

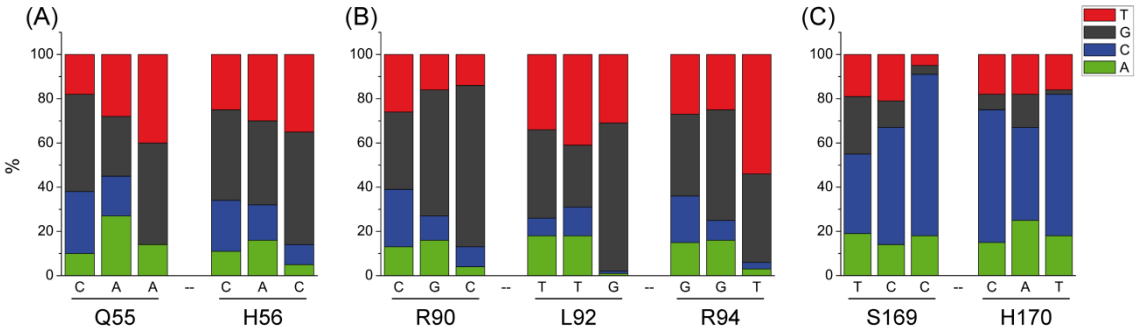


Figure 3.10: Summary of sequencing results for the created PI-SceI libraries*. The percentage of every nucleotide appearing in the sequencing results of pool QH (A), pool RLR (B) and pool SH (C) is represented as fraction of 100 %. On the x-coordinate the nucleotides of the wild type sequence are indicated, with the corresponding amino acid shown below. T: thymine, G: guanine, C: cytosine, A: adenine*

Since pool RLR, with three residues to be mutated, has the highest variability and is the most distant from the active center, this library was used in the first approach to transform *E.coli* harboring the reporter plasmids gvpA_PL_PB or gvpA_PL_PC respectively. The selection assay "on plate" was performed, as described above, over two generations. After the last round two colonies were chosen from the 42 °C plate selected against gvpA_PL_PB and three colonies were chosen from the 42 °C plate selected against gvpA_PL_PC, DNA was

extracted (2.2.2.1) and sent for sequencing to see which variants appeared. Table 3.1 lists the variants selected against gvpA_PL_PB (PI-SceI_B1 and PI-SceI_B2) and against gvpA_PL_PC (PI-SceI_C1, PI-SceI_C2, PI-SceI_C3).

Table 3.1: Overview of selected amino acids at positions 90, 92 and 94 after two rounds of selection against the new targets B and C respectively. The triplets coding for these amino acids are given in brackets. The variants were named after the corresponding target site, they were selected for. Variants B1, C1 and C2 were over-expressed and purified.

name	**position 90**	**position 92**	**position 94**
PI-SceI_wt	Arg (CGC)	Leu (TTG)	Arg (GGT)
PI-SceI_B1	Gly (GGT)	Gly (GGG)	Arg (AGG)
PI-SceI_B2[a]	Arg (CGG)	Leu (TTG)	Arg (GGT)
PI-SceI_C1	Arg (AGG)	Gly (GGG)	Val (GTT)
PI-SceI_C2	Gly (GGG)	Ser (TCG)	Arg (AGG)
PI-SceI_C3[b]	Arg (AGG)	Gly (GGG)	Val (GTT)

[a] same as wild type [b] same as c1

For all selected variants only one or two of the three amino acids had been changed. In case of PI-SceI_B2 the amino acids of PI-SceI_wt had been recovered, while against the new target C the same variation of amino acids had been selected twice. It is also to mention that all amino acids were substituted by small uncharged amino acid residues, such as glycine, valine or serine. Overall, the observed changes in nucleotide sequence could be regarded as expected from the sequencing results (Figure 3.10).

Variants PI-SceI_B1, PI-SceI_C1 and PI-SceI_C2 were over-expressed and purified as described for PI-SceI_wt (2.2.2.9). To see if they cleave the target sites they were selected for, specifically over the wild type target (Figure 3.11), their *in vitro* activity was tested as described in 2.2.2.11. The activity was always compared to PI-SceI_wt under different buffer conditions listed in 2.1.6.2 and using the wt target or the new target B or C as substrate. The first test was done under conditions optimal for PI-SceI_wt (Mg^{2+} buffer pH 8.5, 37 °C, [70]), but at these conditions the new variants did not cleave the new target sites more preferably, but the wt target with comparable efficiency as PI-SceI_wt, with PI-SceI_C1 having the lowest cleavage activity of 15 % after 1 h.

Since the selection assay was mainly performed at 42 °C and it also has been reported that PI-SceI shows a better activity at this temperature [64] all activity tests were performed at 42 °C as well (right panel Figure 3.11). Under these conditions (Mg^{2+} buffer pH 8.5, 42 °C) the new variants did not cleave their respective target sites, but only the wt-target. The only

difference was that PI-SceI_wt cleaved its target site to ~ 80 %, while PI-SceI_C1 did not cleave this site at all and the other two variants only to 30 %.

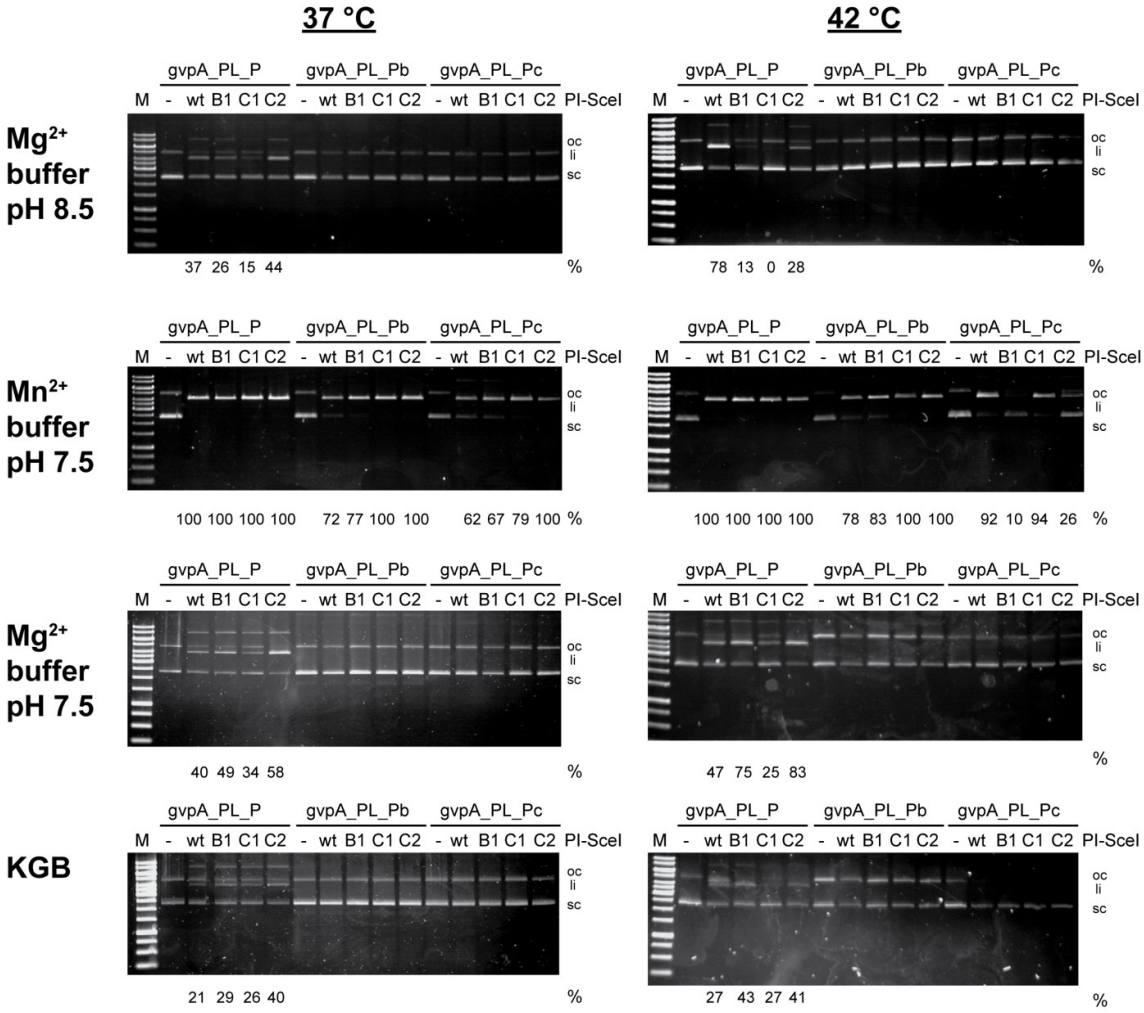


Figure 3.11: Summary of* in vitro *activity assays of the selected variants PI-SceI_B1, PI-SceI_C1 and PI-SceI_C2 under different buffer and temperature conditions. *The selected variants and PI-SceI_wt were tested on plasmid DNA containing either the wt-target or the new targets B or C. The tested buffers are indicated on the left side and were tested either at 37 °C (left panel of gel pictures) or at 42 °C (right panel of gel pictures). The percentage of linearized DNA is indicated below each gel picture; if there is nothing written, no cleavage occurred. M: Gene Ruler ^{TM} 1 kb DNA Ladder, oc: open circle, li: linear, sc: super coiled plasmid DNA*

It is known that PI-SceI shows a higher activity but also a more relaxed specificity in presence of Mn^{2+} instead of Mg^{2+} [73]. For this, the activity of the variants was also tested in Mn^{2+} buffer at 37 °C and 42 °C (second row Figure 3.11). Under these conditions the wt target site was cleaved by all enzymes completely after 1 h incubation. The new target B was completely cleaved only by the variants selected against target C. PI-SceI_wt and PI-SceI_B1 cleaved this substrate only to 70 - 80 %. The target C was only cleaved to completion by PI-SceI_C2, while the other variants cleaved only to 60 - 70 % at 37 °C. This

is in contrast to incubation at 42 °C, where only PI-SceI_wt and PI-SceI_C1 cleaved this substrate nearly completely and PI-SceI_C2 cleaved just 26 % and PI-SceI_B1 just 10 %.

To test if the pH has an influence on the specificity of the selected enzymes, the activity assay was also performed in Mg^{2+} buffer pH 7.5. Still, the new variants did not cleave their respective target sites, but showed increased cleavage activity on wt target. Finally, the activity assay was performed in KGB, which was reported to be a more physiological buffer [94], at 37 °C and 42 °C (last row Figure 3.11). Still, even under these "quasi"-physiological conditions the new variants did not cleave target B or target C, although they must have been active in *E.coli* cells.

In summary, to our surprise, the new variants cleaved the target sites they were selected for only in the presence of Mn^{2+} which is not physiological. Even under these conditions there was only a slight preference of PI-SceI_B1 for the target B at 42 °C detectable, while the other variants PI-SceI_C1 and PI-SceI_C2 showed no preference for the target C under the conditions tested. The lack of specificity is also represented by cleavage of the wt target by all variants under all conditions. For this reason, the *in vivo* specificity of these variants was tested as described in 2.2.1.5. The obtained survival rates are listed in Table 3.2.

Table 3.2: In vivo *activity of selected PI-SceI variants, represented by survival rate in a two-plasmid assay on plate using gvpA_PL_PC as reporter plasmid. The survival rate is the percentage of colonies grown on the 42 °C plate in relation to the number of colonies on the control plate.*

enzyme	**survival rate**
PI-SceI_T225A	0.02 %
PI-SceI_B1	0.8 %
PI-SceI_C1	1.0 %
PI-SceI_C2	1.7 %

All selected variants showed a 40 - 85 times higher survival rate than PI-SceI_wt, meaning that in the cell these variants were obviously able to cleave the new target C, which is in contrast to the obtained *in vitro* data.

3.2 ENGINEERING MEGANUCLEASES BY A FUSION APPROACH

In the second approach for generating highly specific nucleases a more rational strategy was pursued, which is based on already published approaches, where a DNA-binding module is fused to a DNA cleavage module, as for example Zinc-finger nucleases [32] or TFO-linked restriction enzymes [54].

In this work, I-SceI was used as a DNA-binding module and the type IIP restriction endonucleases scPvuII or PvuII served as DNA cleavage module. To investigate what the fusion enzyme would look like, a model (Figure 3.12) was built. The enzymes were arranged in a way that they all approach the DNA from the same direction, resulting in a distance between the C-terminus of PvuII or scPvuII and the N-terminus of I-SceI of 26 Å. This distance has to be covered by a suitable peptide linker, which is not shown in the model. Furthermore in this model, the corresponding target sites of the enzymes, which are colored red and orange, are separated by 6 bp.

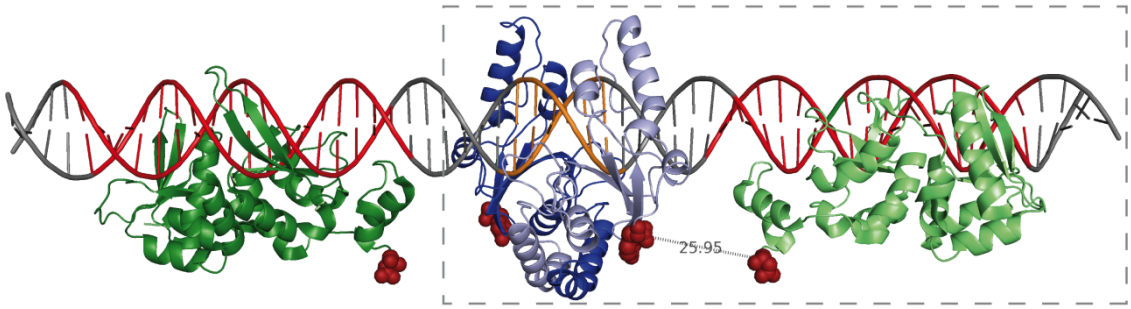


***Figure 3.12: Model of the engineered fusion enzyme**. The whole picture would correspond to the fusion of PvuII (blue, pdb:1pvi) and I-SceI (green, pdb: 1r7m) resulting in a dimeric protein. The structure framed by the dashed box would correspond to the fusion of scPvuII to I-SceI resulting in a monomeric enzyme. This model was built using Pymol by aligning the recognition sites from the crystal structures of the individual proteins on a DNA composed of the PvuII recognition site (orange) and two I-SceI sites (red) 6 bp up- and downstream of it. The C-terminus of PvuII and the N-terminus of I-SceI, separated by 2.6 nm, are indicated by red spheres; this distance must be covered by a peptide linker of suitable length.*

3.2.1 GENERATION OF SCPVUII-I-SCEI FUSION ENZYME

3.2.1.1 Purification of scP-$L_{(H)}$-$S_{(D44S)}$

Cloning of the genes coding for scPvuII and the inactive form of I-SceI (D44S) was performed as described in section 2.2.2.5. The fusion protein was over-expressed in *E.coli* and purified (2.2.2.9) via the His_6-tag which served in this case as linker ($L_{(H)}$) between the C-terminus of scPvuII and the N-terminus of I-SceI_D44S. The SDS-PAGE following purification in Figure 3.13A shows that the eluate was not pure and it seemed possible that

there was still scPvuII in the fractions, which would later affect the activity assays of scP-$L_{(H)}$-$S_{(D44S)}$. For this reason, a subsequent gel filtration was performed. The chromatogram (Figure 3.13B) shows a single peak eluting between 13 - 16 ml, which corresponds to pure scP-$L_{(H)}$-$S_{(D44S)}$ (insert Figure 3.13B). The fractions containing the pure protein were dialyzed for long-term storage. This protein preparation with a concentration of 10 µM was used for the following experiments.

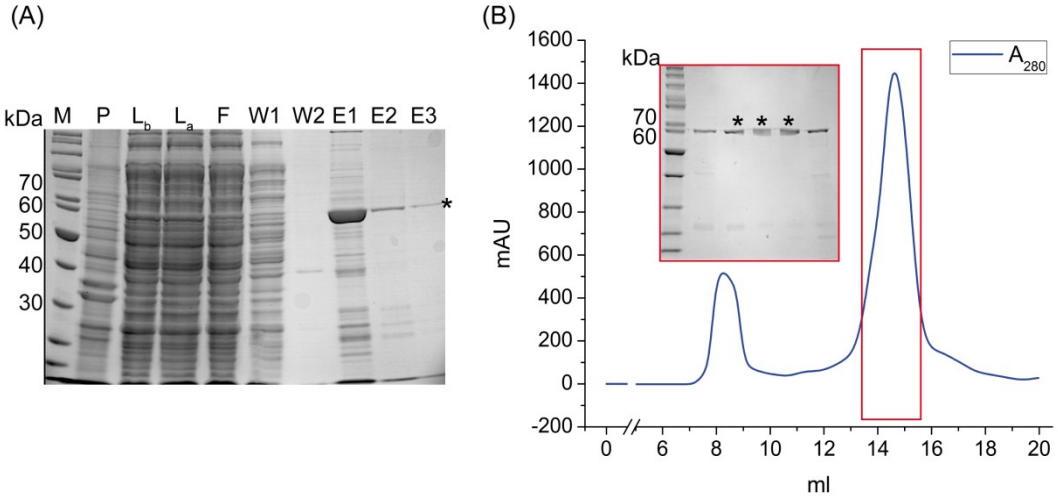


Figure 3.13: Purification of scP-$L_{(H)}$-$S_{(D44S)}$. *The first purification step was with affinity purification over Protino® Ni-IDA columns (A), while scP-$L_{(H)}$-$S_{(D44S)}$ elutes mainly in fraction E1 with an expected molecular weight of 65.6 kDa and is marked with an asterisk.(B) Chromatogram of fraction E1 from (A) after the second purification over Superdex 200 10/300 GL column. The absorbance at 280 nm is plotted against the elution volume. Fractions eluting from 13.25 - 15.75 ml (red frame) were analyzed on SDS-PAGE (insert B) and those containing pure scP-$L_{(H)}$-$S_{(D44S)}$ (*) were dialyzed overnight. M: PageRuler™ unstained Protein Ladder; P:pellet; $L_{b/a}$: lysate before and after filtration through a 0.45 µm filter; F:flowthrough; W1/2: wash; E1/2/3: eluates*

3.2.1.2 Activity assay of scP-$L_{(H)}$-$S_{(D44S)}$

To test if scP-$L_{(H)}$-$S_{(D44S)}$ has any unspecific cleavage activity, a PCR fragment containing only an I-SceI site (and no PvuII site) was used for a first activity assay (2.2.2.11). This test showed a specific cleavage product for scP-$L_{(H)}$-$S_{(D44S)}$, which was comparable to the cleavage product of I-SceI alone (Figure 3.14A). To spot the cleavage site, the products were gel-purified (2.2.2.1) and sent for sequencing. The results shown in Figure 3.14B revealed that the DNA was cleaved at the sequence CAG|CAG, which is in fact a PvuII "star" site (i.e. a site differing in one bp from the canonical PvuII recognition site) 6bp downstream of the I-SceI site. With this experiment, the proposed model (Figure 3.12) where the target sites for PvuII and I-SceI are separated by 6 bp could be confirmed. Furthermore, this seems to be consistent because only in this arrangement all enzymes bind the DNA from the same direction.

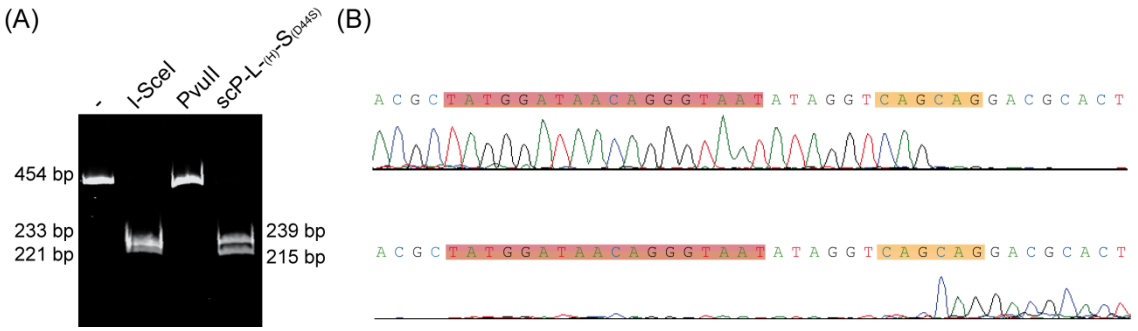


Figure 3.14: Activity test of scP-L$_{(H)}$-S$_{(D44S)}$ and sequencing results of the cleavage fragments. *(A) Cleavage of PCR fragment containing only an I-SceI recognition site. Lane 1 (-): uncleaved DNA fragment; Lane 2: cleavage products obtained by incubation of the PCR fragment with I-SceI (Fermentas); Lane 3 and 4: cleavage products obtained by incubation of the PCR fragment with PvuII and scP-L$_{(H)}$-S$_{(D44S)}$, respectively. (B) Sequencing results of the cleavage product obtained with scP-L$_{(H)}$-S$_{(D44S)}$ from (A). The fragments were gel-purified and sequenced in forward (upper panel) and reverse (lower panel) direction. The I-SceI recognition sequence is indicated by a red bar and the cleavage site which corresponds to a PvuII "star" site by an orange bar.*

To determine if there was any distance preference, scP-L$_{(H)}$-S$_{(D44S)}$ was tested for activity on plasmids pAT_S3P, pAT_S8P, pAT_S10P and pAT_S12P (2.2.2.11), where now a PvuII site (P) was in certain distance (indicated by the number in the plasmid name) to the I-SceI site (S).

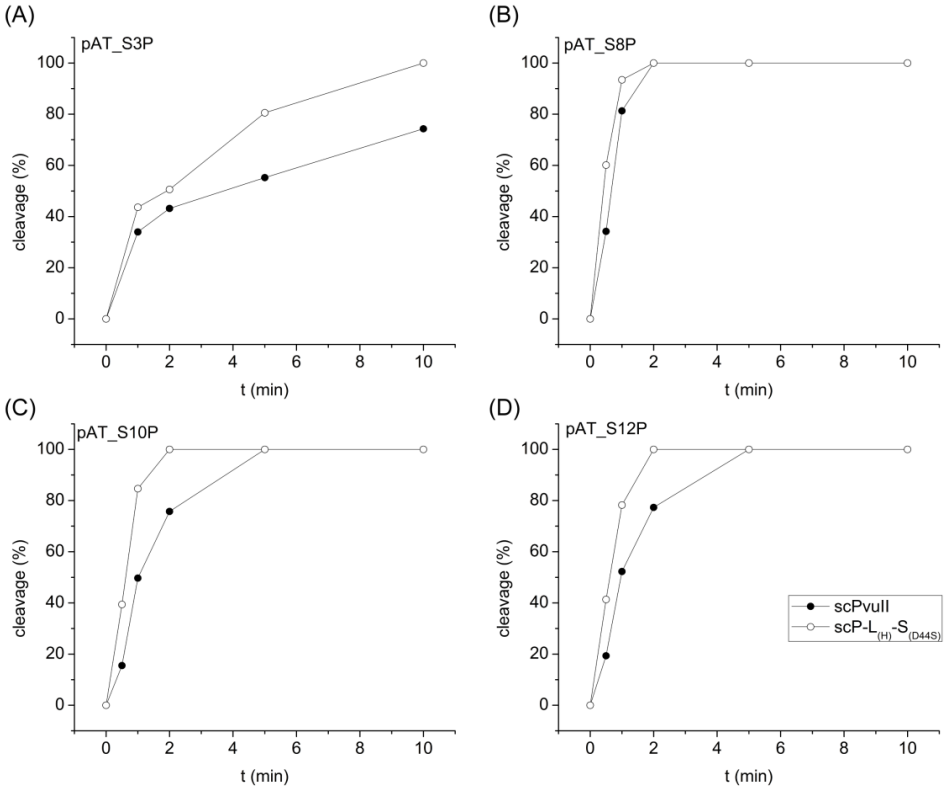


Figure 3.15: Comparison of the cleavage activities of scPvuII and scP-L$_{(H)}$-S$_{(D44S)}$. *The percentage of cleavage of pAT_S3P (A), pAT_S8P (B), pAT_S10P (C) and pAT_S12P (D) by scPvuII (●) or scP-L$_{(H)}$-S$_{(D44S)}$ (○) is plotted against time.*

Figure 3.15 shows that there was no significant preference for any of the substrates. pAT_S3P (Figure 3.15A) was cleaved more slowly by scP-$L_{(H)}$-$S_{(D44S)}$ but for some reason it is also cleaved more slowly by scPvuII. Overall, one has to point out that scP-$L_{(H)}$-$S_{(D44S)}$ cleaves all the substrates slightly faster than scPvuII.

3.2.1.3 Optimizing the buffer conditions

For future applications, like gene targeting, meganucleases should be active in a cellular environment. To optimize the fusion enzymes in this direction, the activity buffer condition used should mimic physiological conditions. The chosen starting point was KGB [94], reported to be a suitable buffer for a number of restriction endonucleases and closer to physiological conditions then most buffers recommended by companies. The optimization of the buffer included adding 100 mM KCl and decreasing the amount of magnesium acetate, since the recommended 10 mM would be too high for cellular conditions (see 2.2.2.11). To test the optimal concentration of Mg^{2+}, plasmid pAT_S8P was cleaved either by scPvuII or scP-$L_{(H)}$-$S_{(D44S)}$ in the presence of different Mg-acetate concentrations.

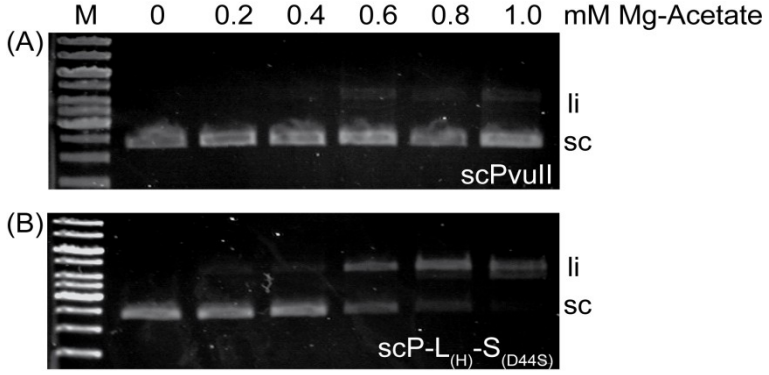


Figure 3.16: Titration of Mg-acetate in KGB. *DNA cleavage experiment of plasmid pAT_S8P comparing scPvuII (A) and scP-$L_{(H)}$-$S_{(D44S)}$ (B) in KGB [94] (100 mM potassium glutamate, 25 mM Tris-acetate pH 7.5, 100 mM KCl, 10 µg/ml BSA, 0.5 mM 2-Me) with different concentrations of magnesium-acetate. M: GeneRuler™ 1kB DNA Ladder, sc: supercoiled, li: linear*

Figure 3.16 shows that at a magnesium acetate concentration of about 0.8 mM the highest difference in activity between scPvuII and scP-$L_{(H)}$-$S_{(D44S)}$ was observed, which means that the fusion construct cleaved an addressed site better than PvuII alone did. All further cleavage experiments with fusion enzymes were performed in an optimized version of KGB (100 mM potassium glutamate, 25 mM Tris-acetate pH 7.5, 100 mM KCl, 0.8 mM magnesium acetate, 10 µg/ml BSA, 0.5 mM 2-Me). $PvuII_{wt}$ was also tested for activity in different buffers (Figure 3.17) as described in 2.2.2.11, but as already reported [95] the activity of PvuII is highly dependent on the Mg^{2+} concentration and there was no cleavage at

magnesium acetate concentrations of 0.8 mM observable, but only at 10 mM in KGB and as expected in the recommended buffer. Therefore, all further controls with PvuII or a variant thereof were performed in Buffer Green.

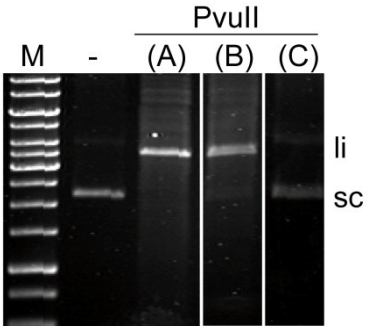


Figure 3.17: PvuII activity in different buffer conditions*. PvuII was incubated with equimolar concentration of 8 nM plasmid DNA containing one PvuII site either (A) in recommended Buffer Green, (B) in KGB with 10 mM Mg-acetate or (C) in KGB with 0.8 mM Mg-acetate. M: GeneRuler™ 1kB DNA Ladder sc: supercoiled, li: linear*

3.2.2 GENERATION OF PVUII-I-SCEI FUSION ENZYMES WITH DIFFERENT LINKER VARIANTS

To extend the recognition sequence of the fusion enzyme, $PvuII_{wt}$ was fused to I-SceI as described in 2.2.2.5, since this enzyme has to dimerize and it has to recognize a tripartite sequence consisting of a PvuII site flanked by two I-SceI sites (see model Figure 3.12).

3.2.2.1 Changing the purification strategy

Furthermore, it was planned to remove the His_6-tag between PvuII and I-SceI. For this, the whole fusion construct was cloned into pASK IBA63b-plus (IBA), containing a C-terminal Strep-tag for protein purification. Additionally, two different variants of I-SceI were cloned, one full-length and the other one a shortened variant, where the last 9 amino acids were removed from the C-terminus, which is according to the sequence resolved in the co-crystal structure [77]. The nomenclature will be S for full-length and Ss for the shortened variant of I-SceI. The affinity purification (example P-$L_{(H)}$-Ss Figure 3.18) using this Strep-tag resulted in cleaner purification products than the previous one, using the His_6-tag (see Figure 3.13) and no additional purification step via gel filtration was needed. All further fusion enzymes carry a C-terminal Strep-tag.

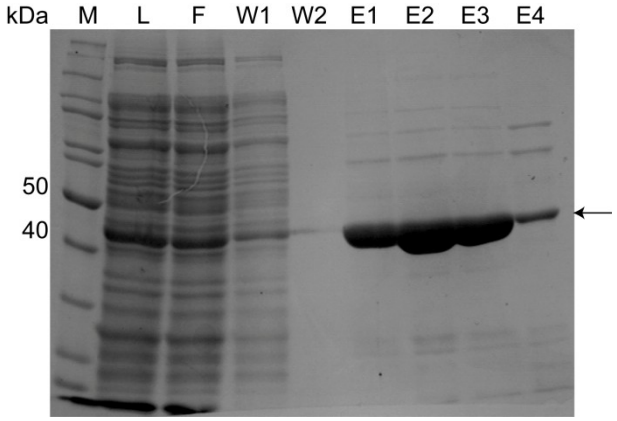


Figure 3.18: Example of an SDS-PAGE after purification of Strep-tagged fusion enzyme*. The estimated size of P-L$_{(H)}$-Ss was 47.2 kDa and the purified enzyme is indicated by an arrow. M: PageRuler^{TM} unstained Protein Ladder; L: lysate; F: flowthrough; W1/2: wash; E1/2/3/4: eluates*

3.2.2.2 Active site mutations of I-SceI

To optimize the fusion enzymes even further, different catalytically inactive variants of I-SceI inspired by [74] were produced. Site-directed mutagenesis with oligonucleotides listed in Table 2.4 was done according to 2.2.2.4. The selected combinations were additional to the already existing I-SceI_D44S, D145, the variants I-SceI_D44S, D145A and I-SceI_D44N, D145A. These variants were chosen because it was reported that they are more soluble, and moreover with the double variants two negative charges next to the DNA were removed and by this possibly the binding affinity of I-SceI to DNA increased. The mutations were introduced into P-L$_{(H)}$-Ss, over-expressed and purified as described in 2.2.2.9. The DNA-binding behavior of these variants was tested by anisotropy measurements (2.2.2.12) in KGB without and with 50 or 100 mM KCl, respectively. Table 3.3 lists the determined K_D-values for DNA-binding of P-L$_{(H)}$-Ss$_{(D44S)}$, P-L$_{(H)}$-Ss$_{(D44S, D145A)}$ and P-L$_{(H)}$-Ss$_{(D44N, D145A)}$ under various conditions.

Table 3.3: DNA-binding constants (K_D) of PvuII-I-SceI fusion enzymes with different active site variants of I-SceI

name	KGB	KGB + 50 mM KCl	KGB + 100 mM KCl
P-L$_{(H)}$-Ss$_{(D44S)}$	13 ± 2 nM	31 ± 7 nM	39 ± 17 nM
P-L$_{(H)}$-Ss$_{(D44S, D145A)}$	12 ± 2 nM	16 ± 4 nM	13 ± 5 nM
P-L$_{(H)}$-Ss$_{(D44N, D145A)}$	8 ± 2 nM	5 ± 1 nM	9 ± 4 nM

It was shown that the K_D increases for all variants with increasing salt concentration, but overall the variant P-L$_{(H)}$-S$_{(D44N, D145A)}$ has the highest binding affinity for DNA. For this reason, the active site mutations of I-SceI D44N and D145A were used in all further fusion

enzymes and were abbreviated as S* and Ss* respectively. The nomenclature for the I-SceI variants used is summarized in Table 3.4.

Table 3.4: Summary of I-SceI nomenclature and corresponding meaning

abbreviation	meaning
S*	full length I-SceI_D44N, D145A
Ss*	ΔC9 I-SceI_D44N, D145A

3.2.2.3 Engineering of the linker

After removing the His_6-tag which served so far as linker $L_{(H)}$, different linker variants were cloned in, using a building block strategy. The oligonucleotides to create linkers $L_{(+)}$, $L_{(N)}$ and $L_{(-)}$ are listed in Table 2.6 and all linkers eventually used are listed in Table 3.5. These linkers were cloned into the PvuII-I-SceIΔC9 fusion enzyme, resulting in P-$L_{(6)}$-Ss*, P-$L_{(N)}$-Ss*, P-$L_{(+)}$-Ss* and P-$L_{(-)}$-Ss*.

Table 3.5: Overview of linkers used in the engineered PvuII-I-SceI fusions.

Linker name	length	sequence	source
$L_{(H)}$	10 AA	GSHHHHHHGT	engineered
$L_{(6)}$	6 AA	ASRTTG	engineered
$L_{(N)}$	10 AA	ASGGSGSGSG	engineered
$L_{(+)}$	10 AA	ASTKQLVKSG	FokI
$L_{(-)}$	10 AA	ASGDSGSDSG	engineered

P-$L_{(6)}$-Ss and P-$L_{(+)}$-Ss* are the most specific of the fusion enzymes with different linkers*

The enzymes were over-expressed and purified as described in 2.2.2.9 and tested for specificity using linearized pAT_PEB (contains one PvuII site) or linearized pAT_S6P6S_P (contains tripartite site and additional PvuII site) as described in 2.2.2.11. Figure 3.19B shows that all the variants cleaved the plasmid pAT_S6P6S_P first at the addressed tripartite site resulting in products indicated by an arrow. Still, they also cleaved the unaddressed site which results in the products marked with an asterisk. This unaddressed cleavage reached a maximum of 11 % after overnight digestion. P-$L_{(6)}$-Ss* and P-$L_{(+)}$-Ss* displayed the lowest unaddressed cleavage activity. This was confirmed by cleavage of plasmid pAT_PEB which contains only a PvuII site (Figure 3.19A), where again the previously mentioned variants cleaved less efficiently than the others. As control both plasmids were also cleaved by PvuII,

which showed cleavage at all PvuII sites independent of neighboring I-SceI. To some extent "star activity" by PvuII was also observable.

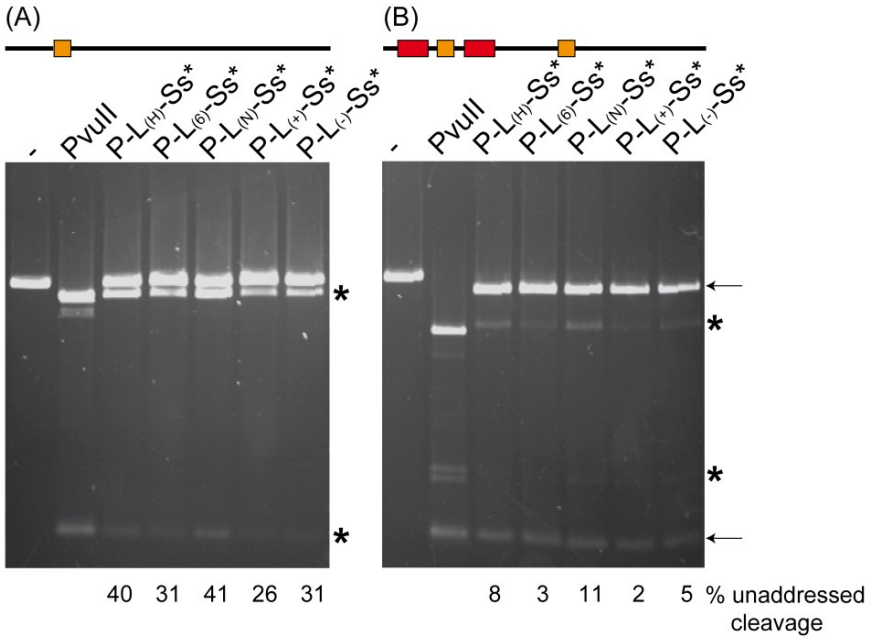


Figure 3.19: Cleavage activity of fusion enzymes with different linkers on linearized DNA *containing either (A) a single PvuII site (pAT_PEB; orange bar) or (B) the tripartite recognition site consisting of a PvuII site (orange bar) flanked by two I-SceI sites (red bars) each 6 bp away, and an additional unaddressed PvuII site (pAT_S6P6S_P). The products of addressed cleavage at the tripartite recognition site are indicated by arrows and the products of unaddressed cleavage at PvuII sites by asterisks. The percentage of unaddressed cleavage is shown at the bottom.*

P-L$_{(6)}$-Ss had the highest preference for 6 bp distance between PvuII and I-SceI sites*

Since the engineered linkers differ in length and amino acid composition it was also tested if there is any preference for a certain distance between the PvuII site and the flanking I-SceI sites. Furthermore, it was tested if there has to be a PvuII site at all between the I-SceI sites, or if it would be enough that the two binding modules bind to their respective sites and that PvuII cleaves the DNA only because of its close proximity to it. For this the activity of the fusion enzymes with different linker variants was tested on plasmids pAT_S4P4S, pAT_S6P6S and pAT_S8P8S, where the distance between I-SceI and PvuII sites is 4, 6 and 8 bp respectively (see 2.2.2.11). Additionally, plasmid pAT_S6x6S, which harbors a random palindromic 6 bp sequence between the two I-SceI-sites separated by 6 bp on each site, was used for the cleavage experiment. Figure 3.20C shows first of all that none of the fusion enzymes cleaved substrate pAT_S6x6S, which mean that there has to be a PvuII site between the two I-SceI sites for specific cleavage.

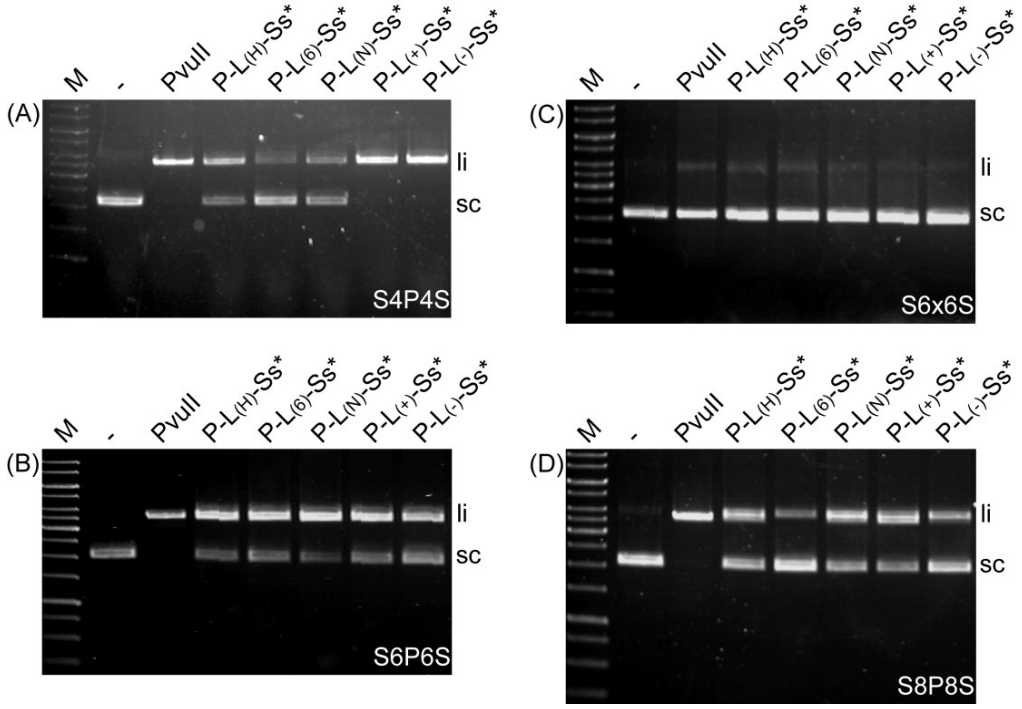


Figure 3.20: DNA cleavage experiments with all linker variants and the different substrates *at equimolar concentrations (8 nM): (A) S4P4S, (B) S6P6S, (C) S6x6S and (D) S8P8S were incubated with PvuII, P-L$_{(H)}$-Ss*, P-L$_{(6)}$-Ss*, P-L$_{(N)}$-Ss*, P-L$_{(+)}$-Ss* and P-L$_{(-)}$-Ss*, respectively, for 3 h at optimized conditions. M: GeneRuler^{TM} 1kB DNA Ladder sc: supercoiled, li: linear*

The substrates containing a PvuII site between the two I-SceI sites were all cleaved by the fusion enzymes, but with different efficiencies. It is obvious that P-L$_{(6)}$-Ss* cleaved pAT_S6P6S to a higher extent than pAT_S4P4S and pAT_S8P8S, suggesting that it is more specific for the 6 bp distance between the target sites. P-L$_{(H)}$-Ss* and P-L$_{(N)}$-Ss* cleaved pAT_S6P6S also to a higher extent than the other two substrates, but not as prominent as P-L$_{(6)}$-Ss* did. Surprisingly, P-L$_{(+)}$-Ss* and P-L$_{(-)}$-Ss*, the two variants with the charged linker, showed a high preference for the substrate with the smallest distance between the target sites (Figure 3.20A), although their linker is longer than the one of P-L$_{(6)}$-Ss*. Taken together all the results for the different linker variants it was decided to do further optimization on the variants P-L$_{(6)}$-Ss* and P-L$_{(+)}$-Ss*.

3.2.2.4 What is the oligomeric state of the PvuII-I-SceI fusion enzymes?

Since it is known that PvuII occurs as dimer in solution [96], it had to be investigated if the fusion of I-SceI to PvuII may have changed the oligomeric state. As example for the PvuII-I-SceI fusion enzymes, the oligomerization was investigated by sedimentation velocity runs of P-L$_{(6)}$-Ss* in an analytical ultracentrifugation. The experiments and analysis were performed in the lab of PD Dr. Ute Curth at the Medical School Hannover. The sedimentation experiments were done under high-salt (500 mM NaCl) and low-salt (100 mM NaCl)

conditions to see if there was a salt-dependent sedimentation behavior of the enzyme. Figure 3.21 shows that under high-salt conditions P-$L_{(6)}$-Ss* sediments as a single species with a $s_{20,W}$ of 4.4 S. This value slightly increases under low-salt conditions resulting in a $s_{20,W}$ of 4.6 S. The analysis of the diffusion broadening of the sedimentation boundary using the program SEDFIT [97] yielded a molar mass of 83 kg/mol and 80 kg/mol, respectively. The molar mass for P-$L_{(6)}$-Ss* calculated from the amino acid composition is about 46.7 kg/mol, suggesting that the enzyme is most probably a dimer in solution, which seems to be salt-independent. Analysis of the frictional ratios, which would be 1.1 to 1.2 for a hydrated, spherical protein [98], revealed that, with a value of 1.58 for the dimeric P-$L_{(6)}$-Ss*, it differs from a spherical shape and probably has a more elongated form, consistent with the proposed model (Figure 3.12).

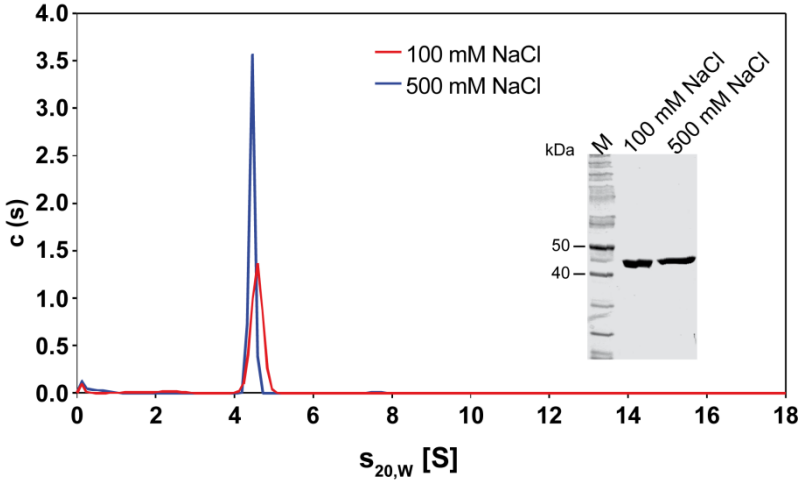


Figure 3.21: Sedimentation profile of the analytical ultracentrifugation run *of P-$L_{(6)}$-Ss* at low (red line) and high salt (blue line) conditions. The protein sediments under low salt conditions with a $s_{20,W}$ of 4.6 S and under high salt conditions with $s_{20,W}$ of 4.4 S. The purity and stability of the used protein was determined by SDS-PAGE (inset) of 1.7 µg of protein diluted in the respective buffer. The theoretical molecular mass of one subunit of P-$L_{(6)}$-Ss* is 46.7 kDa.*

3.2.3 Optimizing the PvuII-I-SceI fusion enzymes

The so far created fusion enzymes showed a certain specificity for the addressed site, but there was still residual activity on the unaddressed sites observable. To optimize the specificity further it was planned to weaken the cleavage module PvuII, so that in the ideal case it cleaves DNA only when it is in close proximity due to binding of the I-SceI modules. Therefore, two different approaches, namely a more functional and a more structural one, have been performed. On one side, PvuII was to be weakened by decreasing its catalytic activity and/or its DNA-binding affinity and on the other side the PvuII dimerization interface was to be disturbed. Both approaches involved introducing certain amino acid changes. The amino acids to be changed are indicated in Figure 3.22, with the right panel showing the

residues used to reduce the "star activity" (T46), decrease the binding affinity (H83) and weaken the catalytic activity (Y94). The residues mutated to disrupt the dimerization interface (L12, P14, H15) are shown in the lower panel. Mutagenesis was performed as described in 2.2.2.4 using oligonucleotides listed in Table 2.4 .

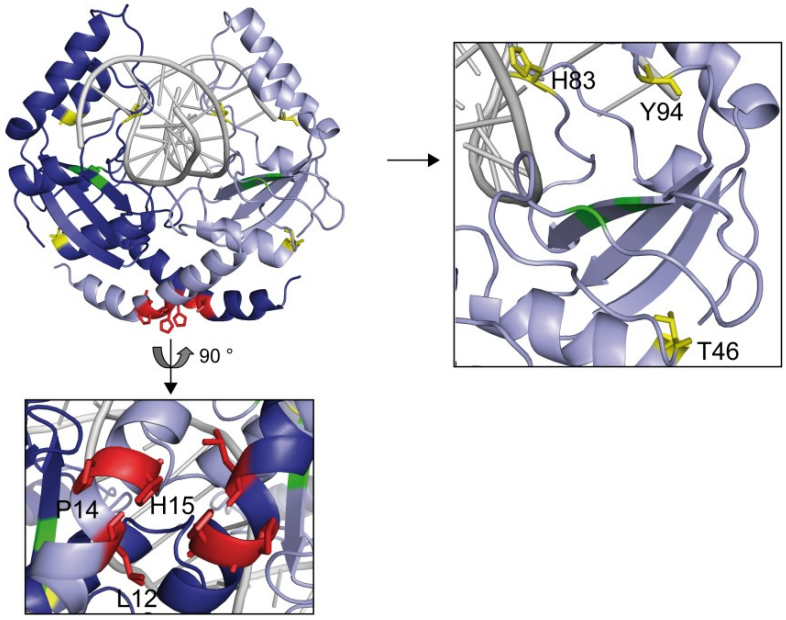


Figure 3.22: Overview of amino acids changed in PvuII (pdb 1pvi). *Positions T46, Y94 and H83 changed to reduce "star activity", weaken the catalytic activity and reduce the DNA-binding are depicted in yellow. To disrupt the dimer interface residues L12, P14 and H15 shown in red were mutated. The catalytic center composed of the residues D58, E68 and K70 is highlighted green.*

3.2.3.1 Weakening the DNA-binding and catalytic properties of PvuII

The first optimization step included mutation of residue Thr46 to Gly which is known to be responsible for reduced "star activity" of PvuII (Zheng, NEB; personal communication). This variant was combined either with the H83A or with the Y94F mutation or both. $PvuII_{(H83A)}$ [99] is known to have a reduced binding ability to DNA, and $PvuII_{(Y94F)}$ [100] is reported to have a decreased affinity for binding the second metal ion and by this a reduced catalytic activity. These mutations were introduced into P-$L_{(6)}$-Ss*, P-$L_{(+)}$-Ss*, and PvuII, either as double mutants in the combination T46G, H83A or T46G, Y94F or as triple mutants T46G, H83A, Y94F. The resulting proteins were over-expressed, purified and characterized for cleavage activity, specificity and DNA binding.

Variants containing the Y94F mutation were specific for the tripartite recognition site

Proteins $P_{(T46G,\ H83A)}$-$L_{(6)}$-Ss*, $P_{(T46G,\ H83A)}$-$L_{(+)}$-Ss*, $P_{(T46G,\ Y94F)}$-$L_{(6)}$-Ss* and $P_{(T46G,\ Y94F)}$-$L_{(+)}$-Ss* were tested for their cleavage activity (2.2.2.11) on linearized plasmid pAT_S6P6S_P (tripartite recognition site and single PvuII site) or pAT_S7P_P (dipartite

recognition site and single PvuII site). Figure 3.23 shows that all enzymes cleaved plasmid pAT_S6P6S_P after 1 h incubation only at the addressed site, while plasmid pAT_S7P_P was cleaved at the addressed site only by the variants containing the H83A mutation. This indicated that the variants containing the Y94F mutation have certain specificity for the tripartite recognition site over just the dipartite one. Overall, none of the tested enzymes cleave the unaddressed PvuII site which was also on the plasmids, indicated by the cleavage experiment with PvuII. Since the final nuclease with extended specificity should cleave as specific as possible, only the variants containing the mutations T46G, Y94F were further characterized.

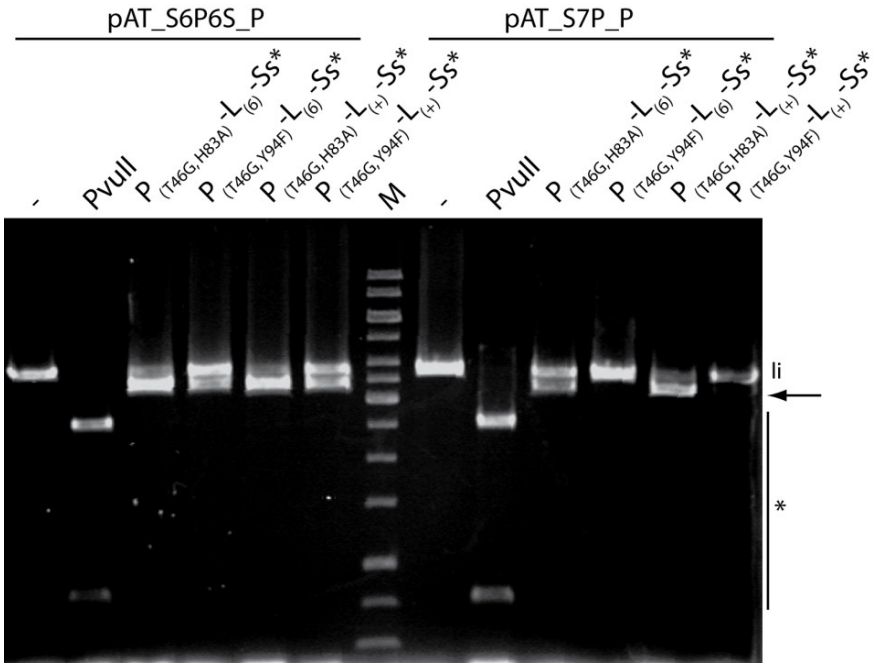


Figure 3.23: Comparison of cleavage activity of $P_{(T46G,H83A)}$-$L_{(6)}$-Ss*, $P_{(T46G, H83A)}$-$L_{(+)}$-Ss*, $P_{(T46G, Y94F)}$-$L_{(6)}$-Ss* and $P_{(T46G, Y94F)}$-$L_{(+)}$-Ss* *of linearized plasmid containing the tripartite (pAT_S6P6S_P) or the dipartite (pAT_S7P_P) recognition sequence. Cleavage by PvuII serves as control, the anticipated unaddressed cleavage product is indicated by an asterisk. Addressed cleavage is shown by an arrow. M: GeneRuler™ 1kB DNA Ladder, li: uncleaved linear plasmid DNA*

Characterization of $P_{(T46G, Y94F)}$-$L_{(6)}$-Ss and $P_{(T46G, Y94F)}$-$L_{(+)}$-Ss**

In the previous experiment no unaddressed cleavage by the fusion enzymes was observable, but since in this assay the incubation time was only 1 h, it was tested if longer incubation of equimolar amounts of enzyme to plasmid DNA leads to unspecific products. The cleavage reaction was incubated up to 21 h. Even after this long time, no appearance of unaddressed cleavage products was visible (Figure 3.24). Both enzymes cleaved only the addressed site after 21 h under optimized conditions.

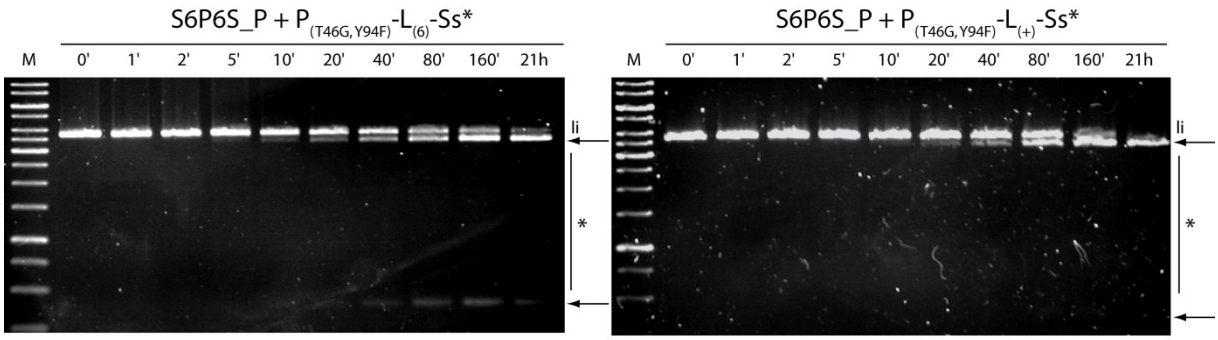


Figure 3.24: Overnight kinetics of $P_{(T46G, Y94F)}$-$L_{(6)}$-Ss* and $P_{(T46G, Y94F)}$-$L_{(+)}$-Ss* on linearized pAT_S6P6S_P. *Equimolar amounts of enzyme and DNA were incubated at 37 °C for distinct time intervals. The product of addressed cleavage is indicated by an arrow, the expected range of unaddressed product appearance is indicated by an asterisk. M: GeneRuler™ 1kB DNA Ladder, li: uncleaved linear plasmid DNA*

Since at equimolar amounts of enzyme and DNA, no unaddressed cleavage could be detected after 21 h incubation, it was to be tested if increasing enzyme concentrations lead to unspecific cleavage. For this, either linearized pAT_S6P6S_P, as used before, or pAT_PEB, which contains only one PvuII site, were incubated with increasing amounts of $P_{(T46G, Y94F)}$-$L_{(6)}$-Ss* or $P_{(T46G, Y94F)}$-$L_{(+)}$-Ss* (4 - 256 nM).

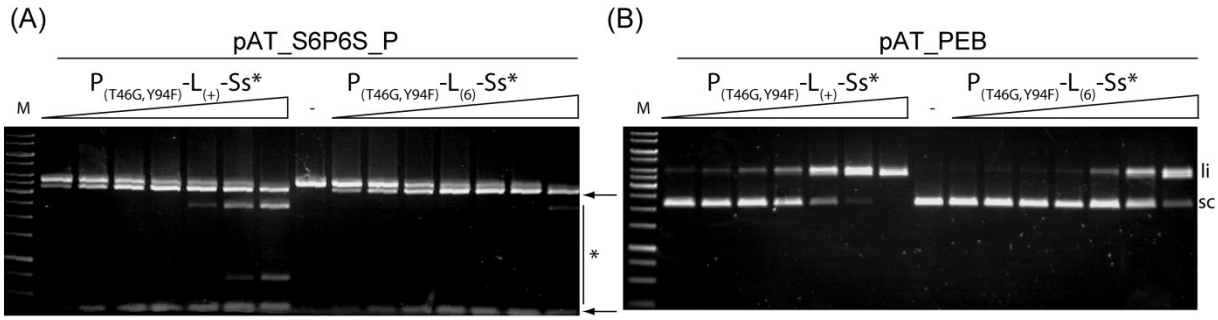


Figure 3.25: Overnight cleavage of plasmid by increasing amounts of $P_{(T46G, Y94F)}$-$L_{(6)}$-Ss* or $P_{(T46G, Y94F)}$-$L_{(+)}$-Ss*. *8 nM linearized pAT_S6P6S_P (A) or pAT_PEB (B) were incubated with enzyme concentrations of 4, 8, 16, 32, 64, 128 and 256 nM consecutively. The addressed cleavage product in (A) is indicated by arrows, the unaddressed cleavage product by an asterisk. M: GeneRuler™ 1kB DNA Ladder sc: supercoiled, li: linear*

Plasmid pAT_S6P6S_P (Figure 3.25A), which contains the addressed site along with the PvuII site, was cleaved unspecifically only at an > 8 - fold molar excess of protein over DNA for $P_{(T46G, Y94F)}$-$L_{(+)}$-Ss* and at an > 16 - fold molar excess of protein over DNA for $P_{(T46G, Y94F)}$-$L_{(6)}$-Ss*. Plasmid pAT_PEB (Figure 3.25B), containing only a PvuII site, was cleaved by $P_{(T46G, Y94F)}$-$L_{(+)}$-Ss* already at equimolar concentrations and by $P_{(T46G, Y94F)}$-$L_{(6)}$-Ss* at > 8 - fold molar excess of protein over DNA. These results indicated that in the presence of an addressed site, this is preferred over an unaddressed site. All previous tests for specificity were done with plasmid DNA which was then analyzed on agarose gels stained with ethidium bromide. To have a more sensitive and accurate

quantification method, competition experiments using radiolabeled PCR-fragments were performed (2.2.2.11). For this, equimolar concentrations of internally [α^{32}P]dATP-labeled DNA fragments containing either the addressed site S6P6S or the unaddressed single PvuII site were incubated with the same molar concentration of fusion enzymes containing either the $L_{(6)}$ or the $L_{(+)}$ linker and either $PvuII_{wt}$ or $PvuII_{(T46G, Y94F)}$ as cleavage modules. Cleavage was analyzed after certain time points for up to 3 h.

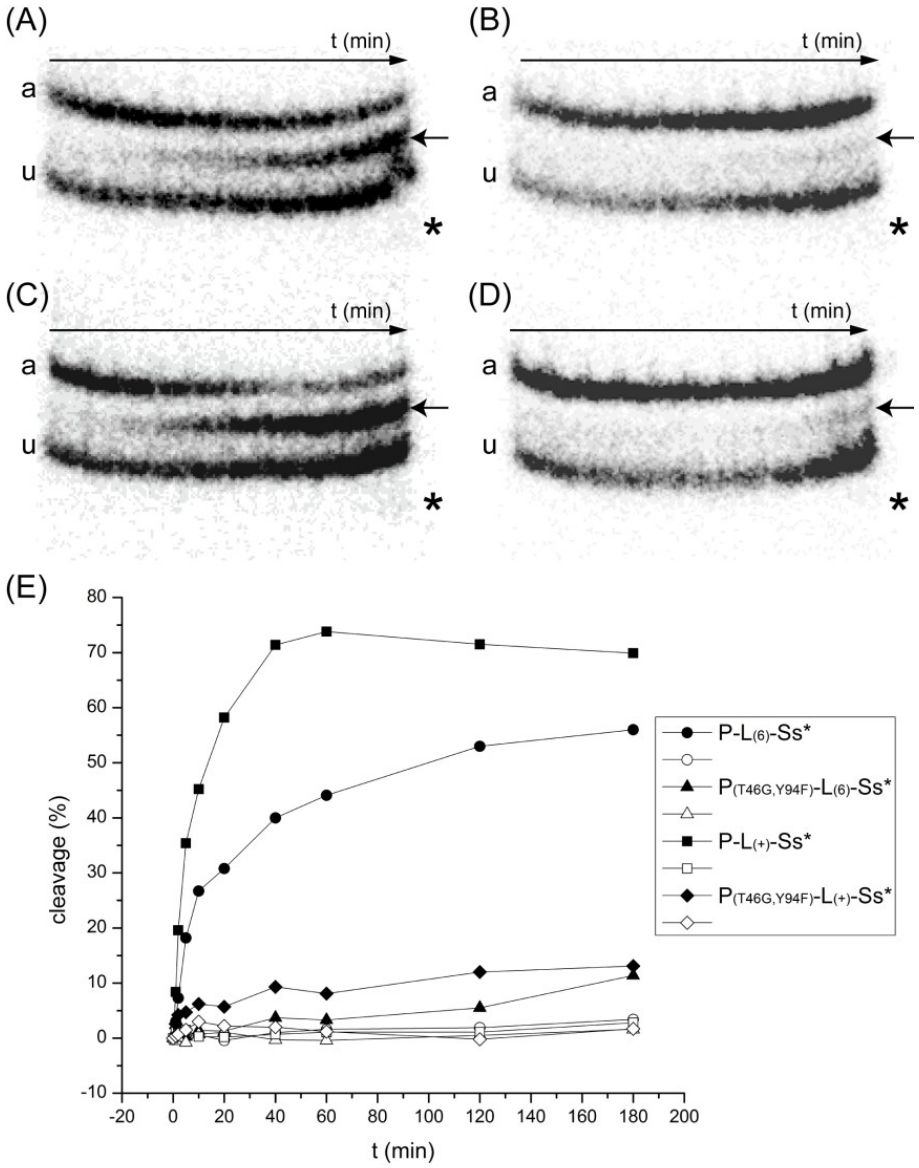


Figure 3.26: Competition cleavage experiments *with fusion enzyme variants of internally [α^{32}P]dATP-labeled PCR-fragments containing the tripartite recognition site S6P6S (a) or a single PvuII site (u) at equimolar concentrations of both substrates. The two substrates were incubated with (A) P-$L_{(6)}$-Ss*, (B) $P_{(T46G, Y94F)}$-$L_{(6)}$-Ss*, (C) P-$L_{(+)}$-Ss* and (D) $P_{(T46G, Y94F)}$-$L_{(+)}$-Ss* for up to 3 h at 37 °C. The cleavage product obtained by cleavage at the addressed site is indicated by an arrow, the position of the expected cleavage product obtained after unaddressed cleavage by an asterisk. (E) The autoradiograms (A-D) were quantified and the percentage of cleavage plotted against time. Filled symbols show addressed, open ones unaddressed cleavage.*

Fusion enzymes P-$L_{(6)}$-Ss* and P-$L_{(+)}$-Ss* (Figure 3.26A and C) containing $PvuII_{wt}$ as cleavage module cleaved the addressed site highly preferentially, leading to a specificity factor of 338 and 72, respectively (see also Table 3.8). In contrast, the variants $P_{(T46G, Y94F)}$-$L_{(6)}$-Ss* and $P_{(T46G, Y94F)}$-$L_{(+)}$-Ss* (Figure 3.26B and D) showed only 11 and 13 % addressed cleavage product after this time and ~ 2 % unaddressed cleavage (Figure 3.26E), which results in a specificity factor of 18 and 11 (Table 3.8). With this, overall reduced activity of around 18-fold, the variants containing $PvuII_{(T46G, Y94F)}$ as cleavage module do not show an obvious preference for the addressed site. It was remarkable that even fusion enzymes containing $PvuII_{wt}$ as cleavage module reached some kind of a plateau after certain time, but this was not at full cleavage. To investigate the reason for this, radiolabeled DNA-fragments containing the tripartite recognition site (S6P6S) were cleaved by P-$L_{(6)}$-Ss* and P-$L_{(+)}$-Ss* as described in 2.2.2.11. One cleavage reaction was stopped by heating the sample for 10 min at 80 °C and the other one not. If the cleavage reaction was not heated up, the protein sticked to the cleavage products, resulting in an enzyme-product complex, indicated by a shift in Figure 3.27. After heating, the complex was destroyed and free cleavage product was observable. This might be indicating that the fusion enzymes do not have a turn over like classical enzymes or only a really slow one.

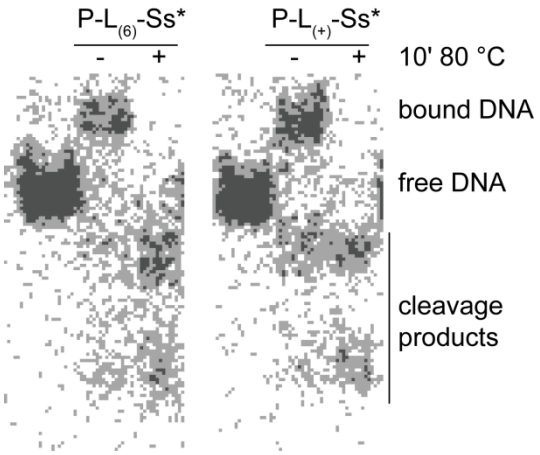


Figure 3.27: Product release experiment of 80 nM P-$L_{(6)}$-Ss* and P-$L_{(+)}$-Ss *with 2 nM internally [α^{32}P]dATP-labeled PCR fragment containing the tripartite recognition site (S6P6S) incubated 30 min at 37 °C. One reaction mixture per protein was heated to 80 °C for 10 min, the other one not. The free DNA, without addition of enzyme, the bound DNA and the cleavage products are indicated on the right side.*

Since I-SceI is the part of the fusion enzymes responsible for DNA-binding, and it was shown that these enzymes bind even their product, it was to be investigated whether the binding of the fusion enzymes is corresponding to the reported K_D for I-SceI of 9.7 nM [19]. For this purpose, internally radiolabeled PCR-fragments, containing either the site S6P6S or just an I-SceI site, were used for gel shift experiments (2.2.2.12) under non-cleavage

conditions (Table 3.6). Overall, the binding constants on the addressed substrate for the fusion enzymes P-$L_{(6)}$-Ss*, $P_{(T46G,\ Y94F)}$-$L_{(6)}$-Ss*, P-$L_{(+)}$-Ss* and $P_{(T46G,\ Y94F)}$-$L_{(+)}$-Ss* were around 2 - 3 times higher than for I-SceI. A single I-SceI site was bound at least 12 times less effectively by the fusion enzymes. Obviously the variant of PvuII fused to I-SceI has no dramatic effect on the binding behavior of the fusion enzyme, but on the contrary the fusion of PvuII might have a disruptive effect on binding of I-SceI to DNA.

Table 3.6: Summary of determined binding constants for P-$L_{(6)}$-Ss, $P_{(T46G,\ Y94F)}$-$L_{(6)}$-Ss*, P-$L_{(+)}$-Ss* and $P_{(T46G,\ Y94F)}$-$L_{(+)}$-Ss* on DNA substrates containing the tripartite site (S6P6S) or a single I-SceI site (S)*

enzyme	substrate	K_D [nM]
P-$L_{(6)}$-Ss*	S6P6S	31 ± 3
	S	144 ± 20
$P_{(T46G,Y94F)}$-$L_{(6)}$-Ss*	S6P6S	31 ± 2
	S	119 ± 16
P-$L_{(+)}$-Ss*	S6P6S	18 ± 1
	S	212 ± 29
$P_{(T46G,Y94F)}$-$L_{(+)}$-Ss*	S6P6S	31 ± 3
	S	> 250

Comparing the cleavage activity of $PvuII_{wt}$, $PvuII_{(T46G,\ Y94F)}$ and $PvuII_{(T46G,\ H83A,\ Y94F)}$

The competition cleavage experiment shown in Figure 3.26 revealed that the fusion enzyme variants containing $PvuII_{(T46G,\ Y94F)}$ showed a low catalytic activity compared to the fusion enzymes containing $PvuII_{wt}$ as cleavage module. To characterize this further, the single cleavage modules $PvuII_{wt}$, $PvuII_{(T46G,\ Y94F)}$ and $PvuII_{(T46G,\ H83A,\ Y94F)}$ were tested in a plasmid cleavage assay. Figure 3.28 shows that introducing the substitutions T46G and Y94F decreases the catalytic activity of PvuII about 100-fold and the triple variant with substitutions T46G, H83A, Y94F has an even 200-fold decreased activity compared to $PvuII_{wt}$.

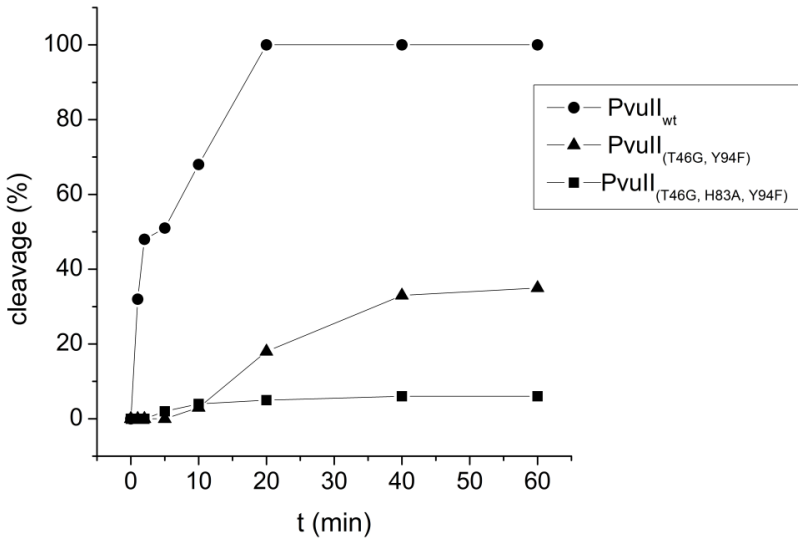


Figure 3.28: Comparison of cleavage percentage of plasmid DNA over time of PvuII$_{wt}$ (●), PvuII $_{(T46G, Y94F)}$ (▲) and PvuII$_{(T46G, H83A, Y94F)}$ (■). *8 nM of plasmid DNA containing one PvuII site and 400 pM enzyme were incubated in Buffer Green at 37 °C. After certain time intervals samples were withdrawn, reaction was stopped and analyzed on 0.8 % agarose gels.*

In vivo test of fusion enzymes

Since the engineered meganucleases were to be used for gene targeting experiments in cells, the off-site target cleavage was measured by an *in vivo* test in *E.coli* as described in section 2.2.1.6. Therefore, *E.coli* cells either expressing the PvuII methyltransferase or not were transformed with plasmids coding for P-L$_{(6)}$-Ss*, P-L$_{(+)}$-Ss* (Figure 3.29A), P$_{(T46G, Y94F)}$-L$_{(6)}$-Ss*, P$_{(T46G, Y94F)}$-L$_{(+)}$-Ss* (Figure 3.29B), P$_{(T46G, H83A, Y94F)}$-L$_{(6)}$-Ss* and P$_{(T46G, H83A, Y94F)}$-L$_{(+)}$-Ss* (Figure 3.29C). As control the same assay was also performed with plasmids coding for PvuII$_{wt}$ (Figure 3.29A), PvuII$_{(T46G,Y94F)}$ (Figure 3.29B) and PvuII$_{(T46G, H83A,Y94F)}$ (Figure 3.29C).

The control plates with cells expressing the PvuII methyltransferase showed for all transformed plasmids independent of the expressed protein, full colony growth (Figure 3.29, left panel). In absence of methyltransferase, cells transformed with plasmids coding for PvuII$_{wt}$ and PvuII$_{(T46G, Y94F)}$ showed no surviving colonies, indicating that these variants cleaved the genomic PvuII site. The cells transformed with plasmids coding for the fusion enzymes showed, independent of the linker variant, surviving colonies in absence of methyltransferase. The amount of surviving cells depended on the variant of PvuII used as cleavage module, with cells expressing the fusion enzyme containing PvuII$_{(T46G, H83A, Y94F)}$ showing the most survivors. Unfortunately, cells expressing this variant of PvuII alone also showed growth in the absence of methyltransferase, indicating that this variant was not able to

cleave the genomic PvuII site. Obviously the triple mutant of PvuII has a too low cleavage activity, as was already shown in Figure 3.28.

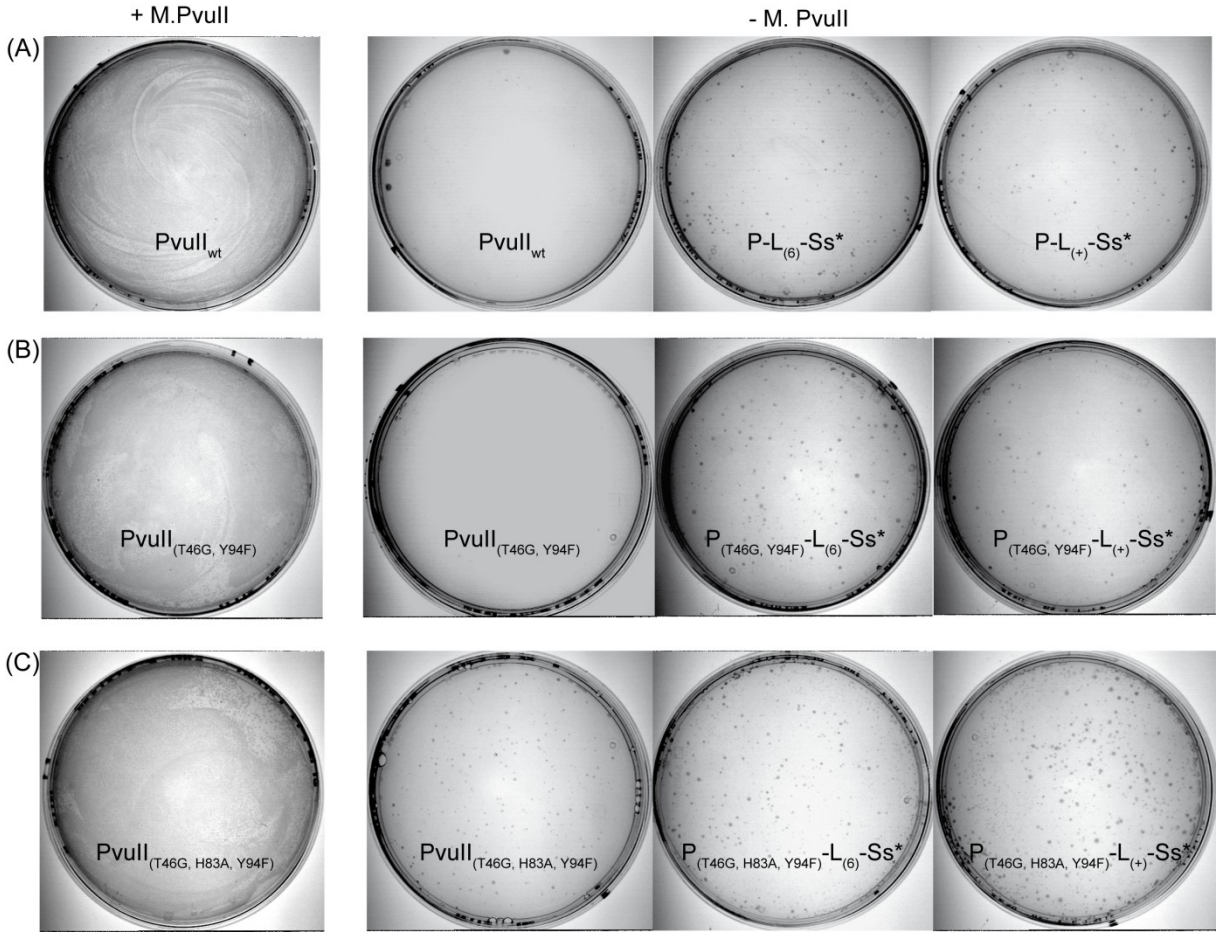


Figure 3.29: Overview of* in vivo *tests with either PvuII alone or with fusion enzymes containing PvuII$_{wt}$ (A), PvuII$_{(T46G, Y94F)}$ (B) or PvuII$_{(T46G, H83A, Y94F)}$ (C). *Control plates representing the amount of grown colonies in presence of M.PvuII (left panel) show exemplarily the results from transformation of plasmids expressing PvuII alone or one of its variants.*

3.2.3.2 Disrupting the dimer interface of PvuII

In section 3.2.2.4 it was demonstrated that the fusion enzymes occur as dimers in solution. The next approach was trying to destabilize the dimerization interface of PvuII, so that dimerization, and by this activity, is dependent on binding of both I-SceI modules to their respective target site in a defined manner. Since the activity of PvuII should not be affected, the residues to be chosen had to lie as far as possible away from the active center, but should be still involved in forming the PvuII dimer. As depicted in Figure 3.22 (lower panel), residues from the first α-helix were used. In the crystal structure of the PvuII dimer, these symmetry related helices interact with each other and are involved in stabilizing the dimer formation. Residue His15 was chosen, since this residue lies in both subunits next to each other and by mutating it to Asp, a negative patch will be introduced so these residues could

repel each other and affect the dimerization interface. The same strategy was followed by mutating Leu12 to Glu. The substitution of Pro14 in the middle of the helix for Gly was suggested by X. Cheng (personal communication). This was supposed to change the bending behavior of the helix and by this reduce the interface contacts. All these mutations were introduced individually in PvuII or in P-$L_{(6)}$-Ss*, which showed the highest specificity for the substrate, in which the I-SceI and PvuII sites are separated by 6 bp (Figure 3.20) and which showed over-all the lowest unspecific cleavage (Figure 3.25). The resulting enzymes were over-expressed and purified.

Characterization of $PvuII_{(L12E)}$, $PvuII_{(P14G)}$, $PvuII_{(H15D)}$

The influence of the introduced substitutions was first tested in PvuII, to exclude that they have such a negative effect on the catalytic activity of PvuII as the mutations reported above (T46G, H83A, and Y94F). For this, steady state beacon cleavage assays were performed as described in 2.2.2.11. As shown in Table 3.7, the kinetic parameters for the newly created variants are over-all lower than those reported for PvuII [85], namely 7 - fold for $PvuII_{(L12E)}$, 1.5 - fold for $PvuII_{(P14G)}$ and 4 - fold for $PvuII_{(H15D)}$.

Table 3.7: Kinetic parameter represented by k_{cat}/K_M values for PvuII variants determined by steady state beacon cleavage assay.

Enzyme	k_{cat}/K_M [min^{-1} M^{-1}]
$PvuII_{wt}$	3.5×10^9 *
$PvuII_{(L12E)}$	4.8×10^8
$PvuII_{(P14G)}$	2.4×10^9
$PvuII_{(H15D)}$	9.0×10^8

* taken from [85]

Since the activity was decreased by these substitutions only to a maximum factor of 7, these results were promising and so all three fusion enzymes $P_{(L12E)}$-$L_{(6)}$-Ss*, $P_{(P14G)}$-$L_{(6)}$-Ss* and $P_{(H15D)}$-$L_{(6)}$-Ss* were further characterized.

$P_{(L12E)}$-$L_{(6)}$-Ss, $P_{(P14G)}$-$L_{(6)}$-Ss* and $P_{(H15D)}$-$L_{(6)}$-Ss* have comparable activity as P-$L_{(6)}$-Ss**

The results for the PvuII alone variants with the substitutions L12E, P14G or H15D showed that their activity is comparable to that of $PvuII_{wt}$. Therefore, it was tested if this would be influenced by the fusion of I-SceI to them. The first assay performed with the variants $P_{(L12E)}$-$L_{(6)}$-Ss*, $P_{(P14G)}$-$L_{(6)}$-Ss* and $P_{(H15D)}$-$L_{(6)}$-Ss* was a DNA cleavage experiment (2.2.2.11) with linearized plasmid DNA containing either just the PvuII site (pAT_PEB) or

the tripartite site and an additional single PvuII site (pAT_S6P6S_P). Figure 3.30A shows that none of the fusion enzymes cleaved the plasmid containing only a PvuII site, whereas pAT_S6P6S_P was cleaved by them only at the addressed site (Figure 3.30B).

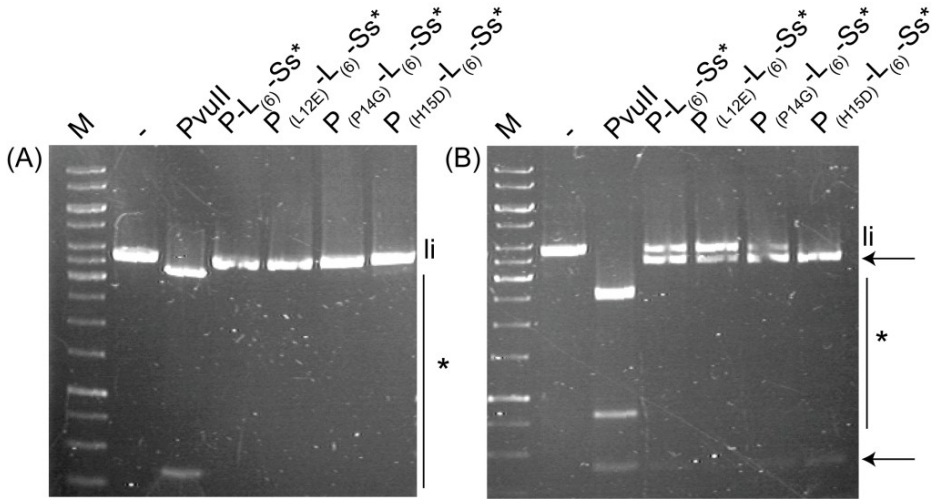


Figure 3.30: DNA cleavage experiment of linearized plasmid pAT_PEB (A) and pAT_S6P6S_P (B). *Equimolar amounts of DNA and enzymes PvuII, P-$L_{(6)}$-Ss*, $P_{(L12E)}$-$L_{(6)}$-Ss*, $P_{(P14G)}$-$L_{(6)}$-Ss* or $P_{(H15D)}$-$L_{(6)}$-Ss* were incubated for 3 h under optimal conditions. The expected unaddressed cleavage product is indicated by an asterisk, the addressed cleavage product by an arrow. M: GeneRuler™ 1kB DNA Ladder, li: uncut linearized plasmid DNA*

$P_{(L12E)}$-$L_{(6)}$-Ss, $P_{(P14G)}$-$L_{(6)}$-Ss* and $P_{(H15D)}$-$L_{(6)}$-Ss* are highly specific for the addressed site*

To investigate the specificity for the addressed site of the variants $P_{(L12E)}$-$L_{(6)}$-Ss*, $P_{(P14G)}$-$L_{(6)}$-Ss* and $P_{(H15D)}$-$L_{(6)}$-Ss* further and with a more sensitive method, competition cleavage experiments over time with internally [α^{32}P]dATP-labeled PCR fragments were performed (2.2.2.11). Figure 3.31 shows that the previously obtained results could be confirmed. The specific activity of the variants with amino acid substitutions in the dimer interface, more precisely $P_{(L12E)}$-$L_{(6)}$-Ss*, $P_{(P14G)}$-$L_{(6)}$-Ss* and $P_{(H15D)}$-$L_{(6)}$-Ss*, is ~ 2 - fold lower than that for P-$L_{(6)}$-Ss* (Table 3.8).

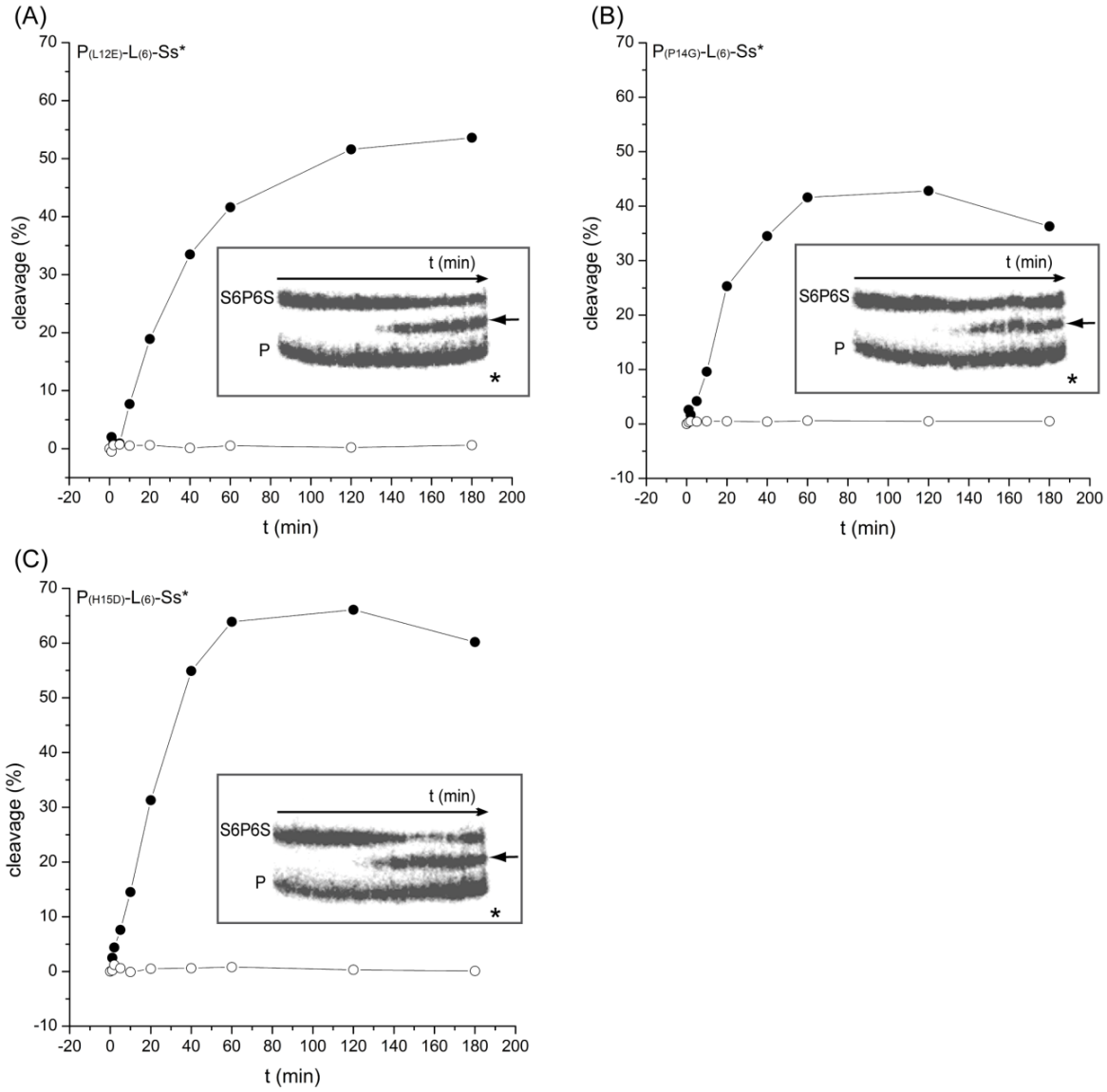


Figure 3.31: Competition cleavage experiments with fusion enzymes $P_{(L12E)}$-$L_{(6)}$-Ss* (A), $P_{(P14G)}$-$L_{(6)}$-Ss* (B) and $P_{(H15D)}$-$L_{(6)}$-Ss* (C) *and internally [α^{32}P]dATP-labeled DNA fragments containing the addressed tripartite recognition site (S6P6S) or the unaddressed PvuII site (P). The autoradiograms are shown as inserts; the result of the quantification is shown as a reaction progress curve (percent cleavage vs. time). The products obtained by addressed cleavage (●) are indicated by an arrow, the products expected by unaddressed cleavage (○) are indicated by an asterisk.*

In contrast, the unaddressed cleavage could not be detected for any of them even after 3 h incubation, leading to a degree of specificity of over 1000 for all the variants (Table 3.8), compared to the variant containing $PvuII_{wt}$ as cleavage module, which showed a specificity of ~ 300. Again, after certain time the enzymes reached a kind of plateau in activity, which could be explained as before with the enzymes sticking on the cleavage products and by this not having a turnover.

Table 3.8: Summary of activities and specificities derived from competition cleavage experiments[§] for all tested fusion enzyme variants

enzyme	activity [nM min^{-1}]		specificity
	addressed	unaddressed	[addressed/ unaddressed]
P-$L_{(6)}$-Ss*	0.88 ± 0.07	0.0025 ± 0.0004	338
P-$L_{(+)}$-Ss*	1.00 ± 0.08	0.014 ± 0.004	72
$P_{(T46G, Y94F)}$-$L_{(6)}$-Ss*	0.018 ± 0.005	0.0010 ± 0.0004	18
$P_{(T46G, Y94F)}$-$L_{(+)}$Ss*	0.025 ± 0.009	0.0023 ± 0.0004	11
$P_{(L12E)}$-$L_{(6)}$-Ss*	0.39 ± 0.05	< 0.0001[a]	> 1000
$P_{(P14G)}$-$L_{(6)}$-Ss*	0.47 ± 0.02	< 0.0001[a]	> 1000
$P_{(H15D)}$-$L_{(6)}$-Ss*	0.44 ± 0.06	< 0.0001[a]	> 1000

[a] below detection limit
[§] carried out in triplicate

Determination of the binding constant

As for all the other variants described before, the binding constant for the enzymes $P_{(L12E)}$-$L_{(6)}$-Ss*, $P_{(P14G)}$-$L_{(6)}$-Ss* and $P_{(H15D)}$-$L_{(6)}$-Ss* was determined by EMSA (2.2.2.12), to investigate if the introduced substitutions have an influence on the specific binding behavior of the fusion enzyme. Surprisingly, the determined binding constants, now on substrate S6P6S (Table 3.9) were similar to the reported K_D of I-SceI. For $P_{(L12E)}$-$L_{(6)}$-Ss* and $P_{(P14G)}$-$L_{(6)}$-Ss* the K_D was about 9 nM and for $P_{(H15D)}$-$L_{(6)}$-Ss* with 2 nM even better than for I-SceI. It was not possible to determine any binding parameters for these enzymes on substrate containing only an I-SceI site, since the gel shift experiments always resulted in undefined bands, which could be a sign for unspecific interaction of the enzyme with this site.

Table 3.9: Summary of the determined binding constants for $P_{(L12E)}$-$L_{(6)}$-Ss, $P_{(P14G)}$-$L_{(6)}$-Ss* and $P_{(H15D)}$-$L_{(6)}$-Ss* on DNA substrate containing the tripartite site (S6P6S).*

enzyme	substrate	K_D [nM]
$P_{(L12E)}$-$L_{(6)}$-Ss*	S6P6S	9 ± 1
$P_{(P14G)}$-$L_{(6)}$-Ss*	S6P6S	9 ± 1
$P_{(H15D)}$-$L_{(6)}$-Ss*	S6P6S	2.3 ± 0.4

$P_{(P14G)}$-$L_{(6)}$-Ss has a high preference for the target site S6P6S*

As already reported, the variant P-$L_{(6)}$-Ss* shows preferred cleavage for the substrate where I-SceI and PvuII sites are separated by 6 bp (Figure 3.20). To see if the variants with substitutions in the dimer interface, which are based on this enzyme, show the same or even a higher substrate preference, plasmid DNA containing either the target S4P4S, S6P6S or S8P8S was cleaved by these enzymes (2.2.2.11). The DNA cleavage experiments were performed three times and the mean percentage of cleaved DNA is plotted for each enzyme-substrate combination in Figure 3.32.

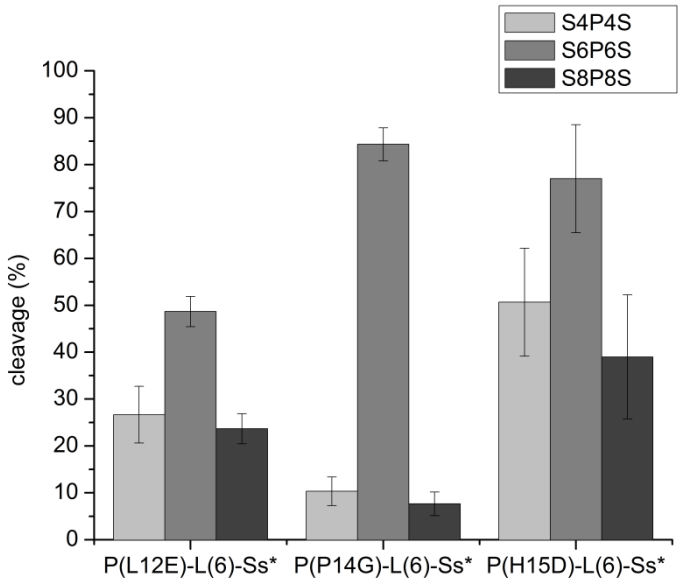


Figure 3.32: Evaluation of DNA cleavage assays with three different plasmid substrates containing the sites S4P4S (light gray), S6P6S (gray) and S8P8S (dark gray). *The variants $P_{(L12E)}$-$L_{(6)}$-Ss*, $P_{(P14G)}$-$L_{(6)}$-Ss* and $P_{(H15D)}$-$L_{(6)}$-Ss* were tested for cleavage of these substrates in equimolar concentration of DNA:protein for 3 h at 37 °C under optimized conditions. The amount of linear DNA was estimated by densitometry of the ethidium bromide stained agarose gels and the mean and standard deviation were calculated and plotted (n=3).*

All tested enzymes, $P_{(L12E)}$-$L_{(6)}$-Ss*, $P_{(P14G)}$-$L_{(6)}$-Ss* and $P_{(H15D)}$-$L_{(6)}$-Ss*, showed preferential cleavage of target site S6P6S, but to different extents. While $P_{(L12E)}$-$L_{(6)}$-Ss* has a ~ 2 - fold preference for S6P6S over S4P4S and S8P8S, and $P_{(H15D)}$-$L_{(6)}$-Ss* a 1.5 - and 2 - fold preference over S4P4S and S8P8S respectively, the variant $P_{(P14G)}$-$L_{(6)}$-Ss* shows with 8 – 10 - fold the highest preference for S6P6S over the other substrates.

3.2.4 OFF-SITE TARGET CLEAVAGE TEST

Off-site target cleavage is a common problem for all engineered nucleases with extended specificity and had to be tested for the here created variants as well. The previously mentioned *in vivo* test is one method to test this and worked good for the variants P-$L_{(6)}$-Ss*,

$P_{(T46G, Y94F)}$-$L_{(6)}$-Ss*, P-$L_{(+)}$-Ss* and $P_{(T46G, Y94F)}$-$L_{(+)}$-Ss* (Figure 3.29), but not for the variants with substitutions in the dimer interface (data not shown). Since there is no obvious explanation for this, the potential of these variants for off-site target cleavage was tested by incubating the enzymes with bacteriophage λ DNA (2.2.2.11), which contains 15 PvuII sites. The concentrations were chosen in a way that the ratio between PvuII sites to enzyme was 1:1 (Figure 3.33).

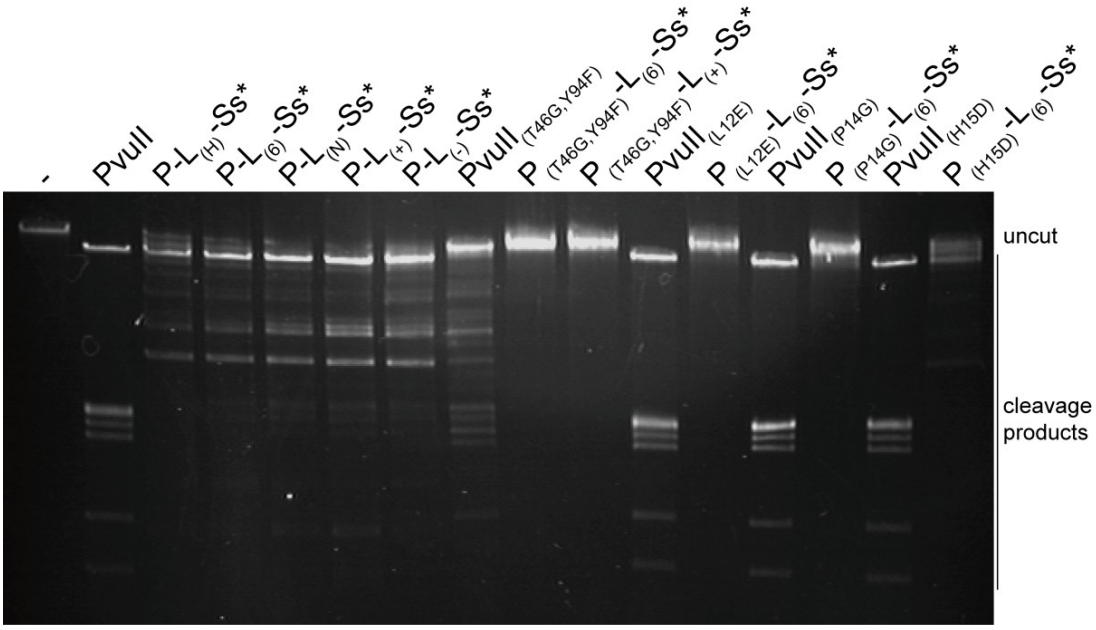


Figure 3.33: DNA cleavage experiment of bacteriophage λ-DNA *with all created fusion enzymes and all PvuII variants. DNA and enzyme concentration were chosen such that the molar ratio between enzyme and PvuII sites was 1:1.*

The first variants (P-$L_{(H)}$-Ss*, P-$L_{(6)}$-Ss*, P-$L_{(+)}$-Ss*, P-$L_{(N)}$-Ss* and P-$L_{(-)}$-Ss*), having $PvuII_{wt}$ as cleavage module, showed cleavage of λ-DNA, even though not as complete as PvuII alone. As expected, the catalytically weakened variants $P_{(T46G, Y94F)}$-$L_{(6)}$-Ss* and $P_{(T46G, Y94F)}$-$L_{(+)}$-Ss*, having $PvuII_{(T46G, Y94F)}$ as cleavage module, showed no cleavage of λ-DNA, but since $PvuII_{(T46G, Y94F)}$ alone showed also just reduced cleavage activity compared to $PvuII_{wt}$, the specific effect is not so prominent (comparable to Figure 3.26). Variants $P_{(L12E)}$-$L_{(6)}$-Ss* and $P_{(P14G)}$-$L_{(6)}$-Ss* showed no cleavage of λ-DNA although $PvuII_{(L12E)}$ and $PvuII_{(P14G)}$ alone cleaved the DNA as complete as $PvuII_{wt}$. This result is also consistent with the finding in Figure 3.31, where these fusion enzymes did not cleave unaddressed substrate at all. Surprisingly, $P_{(H15D)}$-$L_{(6)}$-Ss* which did not show unaddressed cleavage in the before mentioned experiment, seemed to cleave λ-DNA at least once, resulting in two faint cleavage product bands (Figure 3.33).

4. DISCUSSION

Homing endonucleases with their large recognition sequences (14 - 40 bp) and their high specificity for DNA binding and cleavage are suitable tools for gene targeting and genome engineering approaches. In this work, two different strategies were pursued to create enzymes utilizable for exactly these applications. The first method used was directed evolution to change the specificity of a well-studied homing endonuclease, namely PI-SceI. This enzyme, consisting of two domains [101], a DNA binding domain (I) and a DNA cleavage domain (II), was thought to be a good candidate for this study since the DNA-binding specificity can be changed without necessarily affecting the catalytic center of the enzyme. The second approach aimed in the same direction, by rational engineering of a nuclease with an extended specificity composed of a DNA-binding and a DNA-cleavage module. As already mentioned above, homing endonucleases have advantageous features making them good DNA-binders. For this purpose, as DNA-binding module in the engineering approach a catalytically inactive but DNA binding competent variant of the homing endonuclease I-SceI was fused to the restriction enzyme PvuII. Since restriction enzymes are well studied enzymes and used for several molecular biology methods to cleave DNA, PvuII was considered to be a suitable choice as cleavage module. Therefore, I-SceI and PvuII were fused with an exchangeable linker and optimized in direction of application as nuclease with extended specificity for gene targeting approaches.

4.1 DIRECTED EVOLUTION OF PI-SCEI

In 1997 Seligman *et al.* [102] described the first *in vivo* assay linking DNA-cleavage activity of an enzyme to the survival of *E.coli* cells. Since this time, a few groups tried to use comparable assays to find enzymes, i.e. homing endonucleases, with new cleavage specificities. All these approaches are described as directed evolution, meaning selection of suitable enzyme variants out of a library of mutant enzymes. These assays were performed as phage-display [20], as screening assays in yeast [21] or mammalian cells [103] or as selection assays in *E.coli* cells [17-19]. The selection assays in *E.coli* especially are based on two-plasmid systems, where one plasmid is the reporter carrying the target site of choice and expressing a toxin and the second plasmid encoding the homing endonuclease library. The assay developed in this work relies on the same principle. In the past, several efforts have been made to find a suitable selection assay for the directed evolution of PI-SceI. The problem was that this enzyme cleaves DNA relatively slowly and the toxin-expression was not repressed tightly enough, leading to background expression of toxic reporter before

cleavage of the respective target site on the reporter plasmid by the homing endonuclease. This caused selection pressure on the cells, which in turn could cause mutations in the gene coding for the toxin, producing a too high number of false positive cells surviving, even though PI-SceI did not cleave. To circumvent these problems, the selection assay was optimized in previous work (Diploma thesis Ines Fonfara, Giessen 2008). The optimization consisted of using the tightly regulated P_L-promoter [61] for the reporter gene and gvpA as toxic reporter, which is a structural rather than a functional toxin. Mutations caused by selection pressure would perhaps not have such drastic effects, which mean that there are no functional residues in the protein which could be affected. This protein only forms gas vesicles, which are likely to also be formed, if there are some residues mutated. Further optimization on the assays used before, were inserting the target site in the coding region of the reporter gene and using low-copy number plasmids for the reporter and high-copy number plasmids for the expression of the homing endonuclease. By this, the efficiency of the assay should be increased.

In this work, the previously developed selection assay, as described in Figure 3.2A, was tested for its limitations in terms of activity of the homing endonuclease used. As indicator for a working system, the doubling time of cells containing the reporter plasmid, carrying the respective target site, and transformed with plasmid expressing the corresponding homing endonuclease was calculated. At 28 °C, the cells grew normally independent of the expressed enzyme, since toxin-expression is not induced under these conditions. In contrast, at 42 °C, when toxin expression is induced, only those cells will grow where an active homing endonuclease is expressed and able to recognize and cleave the target site on the reporter plasmid. The assay was applied in liquid culture, since the success of the selection can be measured in real time, while for the selection of clones this would not be appropriate. To select a single active and specifically cleaving enzyme variant out of a library of variants, it has to be feasible to select single colonies. This was achieved by the selection assay “on plate” (Figure 3.2B). To evaluate the limitations of the first assay “in liquid”, different variants of I-SceI (overview in Figure 3.3), the so-called “gold-standard” [56] in terms of inducing homologous recombination, were created. These enzymes showed *in vitro* step-wise decreased activity compared to the wild type enzyme (see Figure 3.4A) and gave an idea of how active an enzyme has to be to be suitable for the selection system. Although the variant I-SceI_D150C, N152K showed the lowest activity compared to wild type enzyme, the doubling time determined *in vivo* under inducible conditions was just slightly higher than for I-SceI_wt. For this reason, it was decided to test the assay also with PI-SceI which has a lower

activity compared to I-SceI. In order to have a more promising starting position PI-SceI_T225A, reported to have a 3-fold higher activity than PI-SceI_wt [89], was also examined in the selection assay. The increased activity of this variant compared to the wild type enzyme was confirmed by *in vitro* cleavage experiments (Figure 3.6A) and had also an effect on the doubling time under inducible conditions, which was ~ 1.5 - fold lower for cells expressing PI-SceI_T225A compared to cells expressing PI-SceI_wt. These first experiments, with the selection assay "in liquid", showed that there is a link between the *in vitro* activity of an enzyme and the survival of cells, harboring the reporter plasmid with the respective target site, under inducible conditions. The less active an enzyme was, the higher was the doubling time at 42 °C, since less cells could survive the inducible conditions at 42 °C, due to the reporter plasmid not being cleaved by the homing endonuclease.

As already described above, the selection assay "in liquid" is not suitable for selection of variants out of a library and for this, the selection assay "on plate" was tested for its feasibility. It turned out that the position of the target site on the reporter plasmid and the duration of the incubation at 28 °C before incubating the plate at 42 °C have an influence on the efficacy of the assay. If the target site is positioned in the coding region of gvpA (gvpA_P) less than 5 % of the transformed cells were able to survive 42 °C, whereas those cells harboring reporter plasmid with the target site in the region of the P_L promoter (gvpA_PL_P) showed up to ~ 70 % survivors (Figure 3.7A). These plates also contained IPTG to induce the expression of the homing endonuclease, and the longer the pre-incubation at 28 °C, the more enzyme molecules have been expressed and have cleaved the target site on the reporter plasmid. This results in a direct correlation between pre-incubation at 28 °C and the number of surviving colonies. On the other hand, the pre-incubation should not be too long, as seen in Figure 3.7B, where on the plate which was incubated for 3 h at 28 °C colonies grew, even though the cells were transformed with an inactive variant of PI-SceI. It is possible that even at 28 °C there is a slight background expression of the toxin, which sets the cells under selection pressure, leading to loss of the reporter plasmid, because there was no chloramphenicol on the plate to maintain the plasmid. For this, a real fine-tuned system between giving the enzyme enough time to cleave the target site and not incubating the cells too long to get false positives, is needed. The difference in activity between PI-SceI_wt and PI-SceI_T225A is also observable in Figure 3.7B, where again PI-SceI_T225A showed always a ~ 1.5 – 2 - fold increased number of surviving cells. The following selection assays "on plate" were therefore performed using gvpA_PL_P as reporter plasmid, PI-SceI_T225A

as starting point for creating libraries and 1 h pre-incubation at 28 °C before incubating the plates at 42 °C.

The practicability of the developed assay to select specific active variants out of a library was tested with a mixture of plasmids expressing the active variant and the inactive one in relation one to hundred. As shown in Figure 3.8, after four generations of selection the active variant PI-SceI_T225A could be enriched completely over the inactive PI-SceI_D218A. This experiment was also performed with a relation of one active in a thousand, but no enrichment of the active variant was observable after four generations, meaning that the library of variants should not be too large. For this reason, it was decided not to do random mutagenesis over the whole PI-SceI gene, but only to select a few residues to randomize using degenerated primers. These residues were chosen according to previously published studies on DNA-binding of PI-SceI, since the goal was to find an enzyme variant recognizing a new target sequence.

To have predominantly catalytically active variants in the library, and by this increase the chance of selecting an active enzyme, residues in domain I were chosen. Cluster of two to three residues were used to create enzyme libraries (Figure 3.9A). Residues R90 and R94 form a positive cluster, which was examined by alanine substitutions [89] and turned out to be crucial for DNA-binding, this cluster contained the residues, R90, L92 and R94. The other two clusters, namely Q55/H56 and S169/H170 were chosen after analysis of the co-crystal structure of PI-SceI [66] and identification of potential hydrogen bonds and hydrophobic interactions between these amino acid residues and the PI-SceI target sequence (summarized in [92]). The successful generation of the enzyme-pools was confirmed by sequencing (Figure 3.10), which showed that the distribution of the nucleotides was not equal at all positions, especially the first positions of the triplets coding for S169 and H170, and the last position of every triplet coding for Q55, H56, R90, L92 and R94, which was strived for, to decrease the chance of generating stop-codons. Overall, at all other chosen positions every nucleotide was present to at least 10 %, so it was decided to test these pools in the selection assay “on plate”. Since R90 and R94 in the pool RLR (R90, L92, R94) were already experimentally confirmed to be involved in DNA-binding and this pool has the highest variability with three residues to change, it was decided to examine this one first in the selection assay. Gimble *et al.* [93] substituted all positions in the PI-SceI recognition sequence to other nucleotides individually, and tested the influence of these positions as PI-SceI cleavage activity on the substrate. These changes were then introduced at the positions of the target site, known from the co-crystal

structure to be involved in contacts to the before mentioned amino acids used to generate the enzyme variant pools (Figure 3.9B). After two rounds of selection of pool RLR against the targets B and C, colonies surviving on the 42 °C plates were examined for the obtained enzyme variant. As listed in Table 3.1 five colonies were obtained, two against target B and three against target C, while PI-SceI_B2 is the same as PI-SceI_C1 and PI-SceI_C3 has the same amino acid composition as PI-SceI_wt. Overall, it was obvious that only one or two of the three amino acids were substituted by others and at least one of them was the same as in the wild type enzyme.

To investigate if the selected variants really show a specific cleavage of the new target sequences, as expected, considering that they had to cleave the respective target site on the reporter plasmid to survive, *in vitro* plasmid cleavage assays were performed. Figure 3.11 summarizes the results of this test, where plasmids containing either the wt-target, target B or target C were cleaved by PI-SceI_wt, PI-SceI_B1, PI-SceI_C1 and PI-SceI_C2 under different buffer and temperature conditions. Surprisingly, under none of the buffer conditions the newly selected variants cleaved the sequences they were selected for, except for the buffer containing Mn^{2+} as divalent metal ion. But PI-SceI_wt also cleaved the targets B and C under these conditions, and it is already known that PI-SceI shows relaxed specificity in presence of Mn^{2+} [73]. It is also noteworthy that, even under Mn^{2+} conditions, only PI-SceI_B1 showed some preference for its target B, while the other variants selected against target C did not show a preferred cleavage of exactly this sequence *in vitro*. Furthermore, the newly selected variants still cleaved the wild type target, under Mn^{2+} conditions completely and under the other buffer conditions to different extents, with PI-SceI_C1 cleaving less than the others. It was thought that possibly the temperature has an influence on the activity, since the selection assay was mainly performed at 42 °C. Unfortunately, this did not make the new variants cleave their target *in vitro* nor did it affect the cleavage of the wild type target. Taken all this into account, it was not possible to confirm the results obtained by the selection assay. Because of this, it was attempted to reproduce it *in vivo* by transforming cells harboring the reporter plasmid gvpA_PL_C with a plasmid coding for PI-SceI_T225A or one of the selected variants. The obtained survival rates reflected the specificity of the new enzyme variants for the target C. Above all, a difference between the wild type enzyme and the selected variants was now observable. The survival rate of PI-SceI_T225A was at least 40 times less than for the selected variants.

Obviously it is not as easy as expected to compare results obtained *in vivo* and *in vitro*. Reasons for this could be the different conditions *in vivo*, which are not fully reproducible *in vitro* and the selection pressure present in the cell due to the expression of toxic reporter gene, if the homing endonuclease is not able to cleave the target sequence. Apparently, all selected amino acid substitutions were from large amino acids as arginine and leucine (Figure 4.1B) to smaller amino acids like glycine, serine and valine. Figure 4.1C shows the most frequently obtained amino acids at the corresponding positions, namely glycine at position 90 and 92 and valine at position 94. The amino acids of pool RLR are all located in a loop, which is possibly flexible, because it is not completely resolved in the crystal structure, but probably ordered in the wild type enzyme due to the two positive charges (R90 and R94) interacting with the DNA in the co-crystal. It is also striking, that at least one of these two arginines is recovered in all the selected variants, indicating that these two residues may be important for the integrity of the DNA-binding by the loop-region.

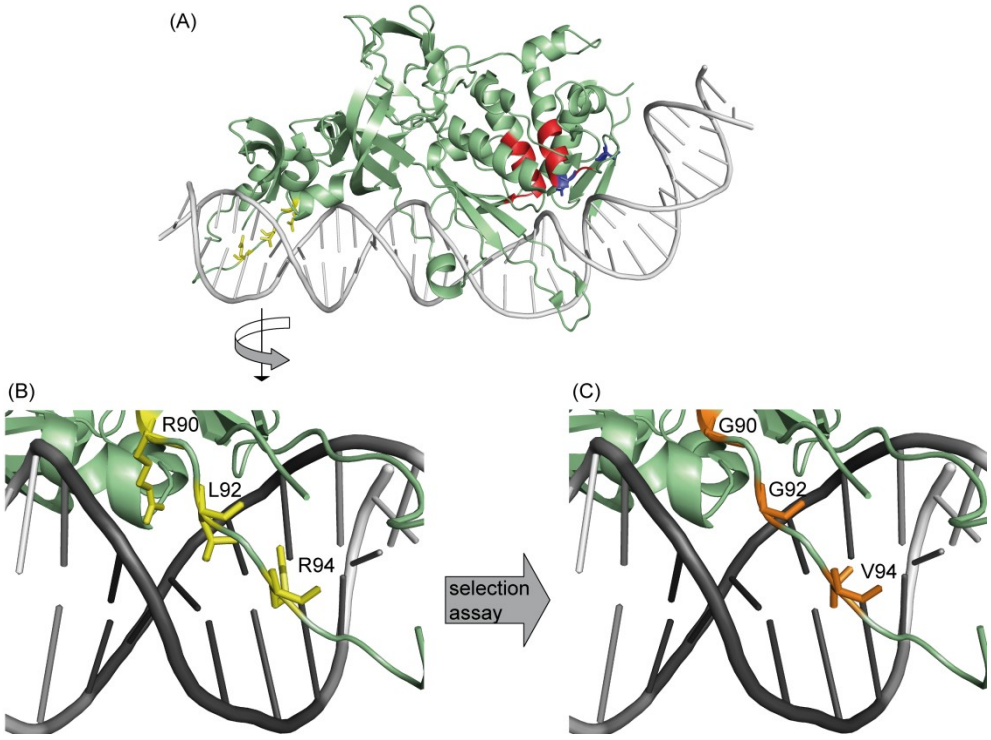


Figure 4.1: (A) PI-SceI co-crystal structure *with LAGLIDADG helices shown in red and the catalytic active aspartates in blue. (B) Cut-out showing pool RLR consisting of R90, L92 and R94 colored yellow. (C) Overview of mainly obtained amino acids after selection against target B or target C shown in orange; this figure is a model made by* in silico *mutagenesis using Pymol and not reflecting the real structure of the enzyme. The changed base pairs of the DNA in target B and target C are colored dark gray and gray respectively.*

If now one of these amino acids is changed to an uncharged one, this loop may become more flexible again and will not have such a strong influence on DNA-recognition and binding. This might be a reason for selecting exactly these amino acids, because then the

DNA recognition and specificity is more relaxed. If this was the case, the selected variants should also cleave the new targets *in vitro*. Since they do not cleave the target sites they were selected for, but only the wild type sequence, there has to be a more complex network of amino acids communicating with each other, and this relatively rational approach with selecting only a few amino acids to change may be not suitable. For this reason, it was decided not to pursue the selection assays with pools QH and SH, since these libraries were made each only of two amino acid residues and the comprising amino acids were not investigated experimentally for DNA-binding before.

This work showed that it is not straight-forward to predict the residues of such large enzymes, like homing endonucleases, possibly involved in DNA-recognition and just changing them. Maybe it would be more promising to do mutagenesis of not only three residues but the whole binding domain of PI-SceI and apply this library to the selection system. But for this, the assay described here has to be improved to find an active and specific variant in the large amount of variants. For this, possibly the first generations of selection could be performed in liquid media to enrich the active and specific variants and then only the last step would be performed on plate, to select single enzyme variants.

A further improvement of the assay would be introducing the wild type recognition sequence for PI-SceI on the homing endonuclease expressing library plasmid as a kind of feedback control, preventing selection of enzyme variants with a broadened rather than a more specific cleavage activity, which was perhaps the case for the variants selected here. Summing all this up, an assay was developed with which it is theoretically possible to select enzyme variants with new specificities, although this new cleavage pattern was only observed *in vivo*. The assay is the first one developed which couples cleavage activity of PI-SceI to survival of *E.coli* cells. All previously mentioned and published selection assays for directed evolution of homing endonucleases were based on cleavage activity of I-SceI or I-CreI and for PI-SceI only one assay based on DNA-binding was reported [15] so far.

Directed evolution of homing endonucleases as it was described above, is neither a rational nor an easy to control approach. Stringent selection systems are needed to find homing endonucleases with a new specificity and not just a broadened one. Furthermore, only a few base pairs in the recognition sequence should be altered at once, meaning a lot of selection cycles are required and moreover, assays based on negative selection always carry the problem of false positive results, due to the loss of the reporter plasmid. This could be

another reason for the variants selected in this work, since the obtained *in vitro* activity does not correspond to the *in vivo* activity of the enzymes.

4.2 ENGINEERING PVUII-I-SCEI FUSION ENZYMES

A more rational approach to create endonucleases with new specificities was aimed for by fusion of a DNA-binding module to a DNA-cleavage module, a method widely applied currently by several groups as well as by biotech and pharmaceutical companies. The most popular DNA-cleavage module in most cases is the unspecific catalytic domain of the restriction endonuclease FokI, which was fused to several DNA-binding modules, including Zinc fingers e.g. [104,105], TALEs (transcription activator like effectors) e.g. [50-52] and an inactive form of the homing endonuclease I-SceI [74]. Using this DNA-cleavage module could be problematic, since it does not specifically recognize and bind DNA by itself, which could lead to unspecific cleavage without binding of the DNA-binding modules to their respective target sites, the so-called off-site target cleavage. For this reason, it was decided to use a specific type IIP restriction enzyme as DNA-cleavage module, namely PvuII. As DNA-binding module, an inactive form of the LAGLIDADG homing endonuclease I-SceI was used and the aim was to combine the recognition sequences of the restriction enzyme and the homing endonuclease resulting in a long sequence, appearing ideally only once in a genome.

As a first attempt a model of the PvuII I-SceI fusion enzyme was constructed (Figure **3.12**), which indicates a distance of 6 bp between the two recognition sequences for the respective enzymes as the best arrangement of the proteins, entering the DNA all from the same direction. It is to mention that the fusion was planned by connecting the C-terminus of PvuII with the N-terminus of I-SceI, since this N-terminus, consisting of two α-helices, seems to "hang out" of the otherwise compact structure of I-SceI and by this could maybe serve as a kind of linker between the proteins. The distance to be covered by a linker was estimated from the model to be ~ 26 Å, assuming that the N-terminal part of I-SceI is located as in the co-crystal structure. The first fusion protein consisted of scPvuII fused to I-SceI_D44S, a catalytically inactive variant of I-SceI, also used in the directed evolution project as negative control. This fusion protein carried a His_6-tag as linker $L_{(H)}$ between the genes coding for scPvuII and I-SceI, used for affinity purification of the enzyme. As shown in Figure 3.13A the proteins resulting from this purification were not completely pure, which could have been due to either breaking of the fusion enzyme or incomplete translation. In either case, there was the danger that scPvuII alone was purified, since it harbors the C-terminal His_6-tag and free scPvuII in the protein preparation would falsify DNA-cleavage experiments. To prevent this, a subsequent gel filtration was performed (Figure 3.13B), resulting in pure fusion protein which could now be tested for activity and particularly specificity, since the main goal of the whole fusion enzymes approach was finding an enzyme recognizing and cleaving a sequence

consisting of the target sites for PvuII and I-SceI together in a certain arrangement, which has to be identified.

In the first test, controlling if there is any unspecific activity of scP-$L_{(H)}$-$S_{(D44S)}$, performed with substrate containing only an I-SceI site but no PvuII site, cleavage of exactly this substrate was observed. Surprisingly, these cleavage products seemed to run at the same height as the products obtained by cleavage of the substrate with I-SceI (Figure 3.14A), indicating that the cleavage occurred at the I-SceI site, which would be unlikely since the fusion enzyme contained the inactive variant of I-SceI. To pinpoint the exact cleavage position, the products were gel-purified and sequenced, revealing that the substrate was cleaved by scP-$L_{(H)}$-$S_{(D44S)}$ at a position 9 bp downstream the I-SceI site. It turned out that this cleavage position lies in the middle of the PvuII "star" site CAG|CAG which is coincidentally located 6 bp downstream of the I-SceI site, exactly like it was proposed in the model. "Star" sites are sequences which differ in one base pair position from the cognate recognition site of the enzyme. PvuII by itself has a certain potential to attack "star" sites, if the buffer conditions are not ideally chosen, if there is too much enzyme in relation to DNA or if the incubation is too long, but in this activity assay PvuII alone did not show any unspecific cleavage, although there are in total four potential "star" sites on the substrate used, differing in one base pair from the PvuII target sequence CAG|CTG. The fact that the fusion protein cleaves also just the one site, addressed by the 6 bp upstream I-SceI site, shows that, due to binding of I-SceI to its respective site, PvuII is directed to the DNA and cleaves a site which it would normally not attack. This already shows a certain specificity of the fusion protein and it also shows that the fusion is done correctly, since I-SceI is obviously able to bind DNA and PvuII is able to cleave DNA at a certain distance away from the I-SceI binding site, the two prerequisites necessary for further improvements.

The model and the first activity test suggest a distance of 6 bp between I-SceI and PvuII target sites, which could not be confirmed completely by the plasmid cleavage assay summarized in Figure 3.15, where target sites with different distances of 3, 8, 10 and 12 bp between the I-SceI and the PvuII site were assayed for cleavage by scP-$L_{(H)}$-$S_{(D44S)}$ and scPvuII. It turned out, that there is no obvious preference for any of the substrates, although the cleavage rate of scP-$L_{(H)}$-$S_{(D44S)}$ of plasmid pAT_S3P is slower than for the other substrates, but scPvuII also shows decreased activity on this substrate. Evidently scP-$L_{(H)}$-$S_{(D44S)}$ shows, for all the substrates, a faster cleavage than scPvuII does, which could

be due to binding of I-SceI to its site and by this directing PvuII to the DNA and its recognition site, while scPvuII has to search for it and then cleave.

The first activity assays of scP-$L_{(H)}$-$S_{(D44S)}$ were performed in Buffer Green recommended as ideal cleavage buffer for PvuII, which is otherwise not suitable for further improvements. Nucleases with extended specificity as they were created here are supposed to be used *in vivo* later, and because of this the buffer conditions should already imitate physiological conditions. In this work, KGB was the initial point for optimizing the buffer conditions, including addition of 100 mM KCl, to end up with an ionic strength close to that of the cellular milieu. Furthermore, the concentration of Mg-acetate had to be decreased, compared to the 10 mM originally used in the published KGB [94]. The activity of scP-$L_{(H)}$-$S_{(D44S)}$ compared to scPvuII was tested with plasmid containing the addressed site S8P, with Mg-acetate concentrations ranging from 0 - 1 mM (Figure 3.16). At a concentration of 0.8 mM Mg-acetate, which is closer to cellular conditions than the reported 10 mM, the difference in cleavage activity between scP-$L_{(H)}$-$S_{(D44S)}$ and scPvuII was the highest, which means that scP-$L_{(H)}$-$S_{(D44S)}$ cleaved the plasmid completely while scPvuII showed nearly no cleavage under these conditions. For this reason, all further experiments were performed in the so-called optimized KGB supplemented with 100 mM KCl and 0.8 mM Mg-acetate. Since PvuII is not active under these conditions and its activity is dependent on the concentration of Mg^{2+} [95] in the buffer (Figure 3.17), all control experiments with PvuII or a variant thereof were still performed in the recommended buffer. This observation was already a first hint that a specific nuclease with extended specificity was created by fusion of a catalytically inactive homing endonuclease to the restriction enzyme PvuII. Obviously, I-SceI is the dominating binding part of the fusion enzyme, bringing PvuII in close proximity to the DNA and its respective target site, and by this promoting the cleavage activity of PvuII, which would otherwise, under the buffer conditions mentioned (optimized KGB), not bind and cleave its recognition sequence on its own.

To increase the recognition sequence of the meganuclease and by this create an even more specific enzyme, $PvuII_{wt}$ was fused to I-SceI instead of the single chain variant, which needs the dimerization of PvuII prior to catalysis and by this the fusion enzyme has to recognize a tripartite sequence consisting of the PvuII target site, flanked by two I-SceI sites. Further improvement of the fusion enzyme included cloning of the construct into a different vector system coding for a C-terminal Strep-tag (pASK IBA-63b plus) and by this being independent of the His_6-tag in between the genes coding for PvuII and I-SceI. With this, it

was possible to remove this linker variant ($L_{(H)}$) and exchange it for others. A further advantage of using the C-terminal Strep-tag for affinity purification was that now no additional purification by gel filtration was needed, since only completely translated fusion enzymes can be purified (Figure 3.18).

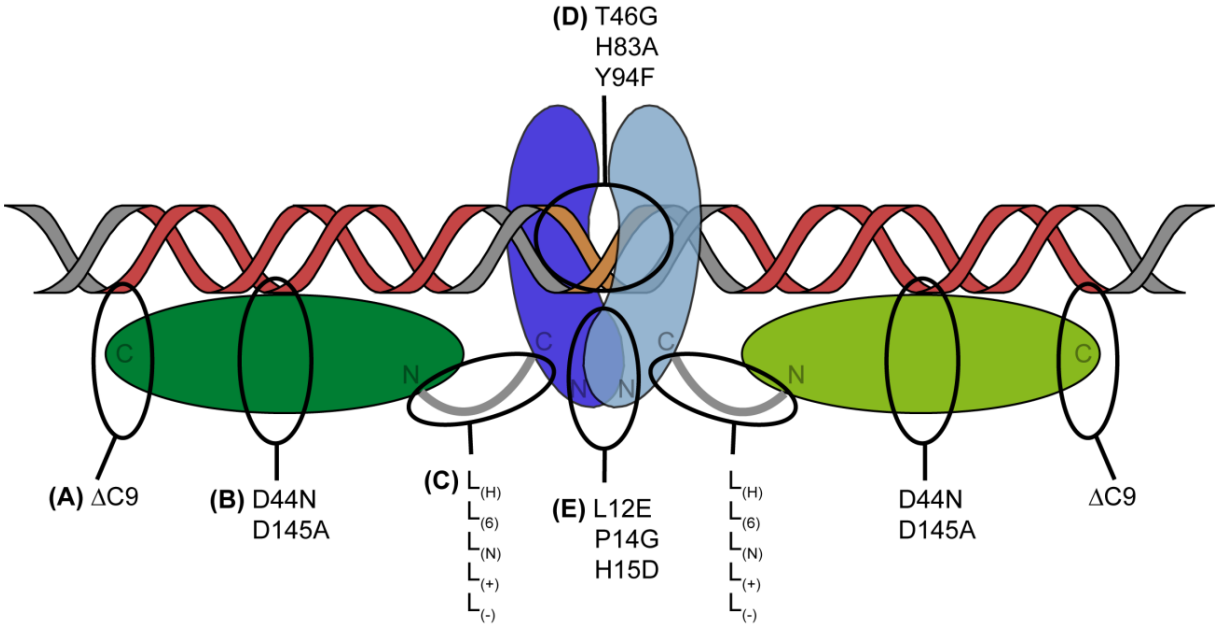


Figure 4.2: Summary of all optimization steps applied to the PvuII-I-SceI fusion enzyme *shown here as cartoon, with I-SceI in green and PvuII in blue. (A) The C-terminal deletion of 9 amino acids, (B) substitutions in the LAGLIDADG motif of I-SceI making it a catalytically inactive enzyme, (C) different linker variants used, (D) substitutions in PvuII decreasing its catalytic activity and (E) substitutions in PvuII disrupting the dimerization interface.*

The DNA-binding module was also subjected to improvements by using a C-terminally shortened variant lacking nine amino acids (ΔC9) and substituting Asp 145 from the second LAGLIDADG motif, in addition to Asp 44, the catalytically active residue in the first LAGLIDADG motif, to ensure that there is absolutely no residual cleavage activity by I-SceI. Furthermore, by this the negatively charged residues lying close to the DNA were removed which might increase the binding affinity of I-SceI. The combinations of active site variants created are listed in Table 3.3, with their respective binding constants determined with substrate containing a PvuII site with an I-SceI site 6 bp away. The variant P-$L_{(H)}$-$Ss_{(D44N, D145A)}$ turned out to have the lowest K_D and by this the best binding behavior, so ΔC9 I-$SceI_{(D44N, D145A)}$ was chosen to be the binding module for all further fusion enzymes (Figure 4.2B). Interestingly, exactly this variant of I-SceI, except for the deletions at the C-terminus (Figure 4.2A), was reported in [74] to be the most soluble variant of I-SceI in their experiments, a fact that could also be profitable for the work described here. As already mentioned in Table 3.4, I-$SceI_{(D44N, D145A)}$ΔC9 will be called Ss* throughout the text.

The different linker variants chosen (Table 3.5, Figure 4.2C) were, with one exception, all ten amino acids long, since this length turned out to be suitable, based on the experiments performed with the fusion enzyme containing the His_6-tag ($L_{(H)}$), also composed of ten amino acids. The stated exception $L_{(6)}$, consisting of 6 amino acids, was actually not planned to be a linker, but it arose by introducing three restriction enzyme sites between the genes for PvuII and Ss*. However, the resulting fusion enzyme was also tested. Other linker variants used were $L_{(N)}$ composed of Gly and Ser, $L_{(+)}$ as a part of the FokI linker [106] and, since there was a neutral and a positively charged linker, also $L_{(-)}$ composed of Gly and Ser and two additional Asp as negatively charged linker was used. A first specificity test of the fusion enzymes carrying the different linker, using plasmid DNA containing an addressed site (I-SceIsite-6bp-PvuIIsite-6bp-I-SceI; S6P6S) and a single PvuII site (Figure 3.19B), revealed that P-$L_{(6)}$-Ss* and P-$L_{(+)}$-Ss* showed the least unaddressed cleavage at the single PvuII site. Overall, it has to be mentioned that all the fusion enzymes cleaved the addressed site (S6P6S) first. Since $PvuII_{wt}$ cleaves both sites (addressed and unaddressed) with the same efficiency, resulting in three major cleavage products, it can be ruled out that there is a common preference for one of the sites due to sequence context, which means the fusion enzymes show a specific cleavage of the addressed target site. The activity test with plasmid DNA containing only a PvuII (Figure 3.19A) site showed that there is still a need to optimize the fusion enzymes, since all of them still attack this site, resulting in 30 - 40 % of cleavage product after overnight incubation. To determine if there are any distance constraints, meaning that the 6 bp distance between the target sites as proposed in the model, is a prerequisite for cleavage, different substrates with 4 and 8 bp between I-SceI and PvuII sites were also generated. As further control, a substrate with no PvuII site, but only a random palindromic 6 bp sequence between the two I-SceI sites in a 6 bp distance, was constructed. Figure 3.20 summarizes the results of cleavage assays on different targets, and it is obvious that all of the fusion enzymes cleave substrate S6P6S with nearly the same efficiency, but surprisingly the variants containing a charged linker (P-$L_{(+)}$-Ss* and P-$L_{(-)}$-Ss*) showed a high preference for the substrate with the shortest distance between I-SceI and PvuII sites. These enzymes were also able to cleave the substrates with the 8 bp distance. The other fusion enzymes, containing either the His_6-tag $L_{(H)}$ or the neutral linker $L_{(N)}$, cleaved the substrates with following preference: S4P4S < S8P8S < S6P6S and only the variant with the shortest linker P-$L_{(6)}$-Ss* showed the preference predicted from the model (Figure 3.12), with substrate S6P6S being the best cleaved substrate. This could be explained by the length of this linker which is maybe too short, since although the target sites are closer together in S4P4S, it has to be considered that

in this arrangement, the target sites moved along the torsion of the DNA and the enzymes would not enter the DNA from the same direction anymore, which would again necessitate a longer linker. The same would of course be true for the 8 bp distance between the target sites. From the first cleavage experiment, where a PvuII "star" site 6 bp away from an I-SceI site was cleaved, it could be expected that, although there is no PvuII site between the binding module sites, this substrate would be cleaved. But as shown in the cleavage experiment with S6x6S as substrate, there has to be a PvuII site between the two I-SceI sites for cleavage (Figure 3.20C), which means that PvuII is not just a cleavage module but that it also contributes to the specificity of the fusion enzyme. This finding is different from the fusion enzymes where the catalytic domain of FokI was used as cleavage module (e.g. Zn-finger nucleases [107]), which by itself does not recognize and bind DNA specifically.

As shown in the model (Figure 3.12), the fusion enzymes having $PvuII_{wt}$ as cleavage module have to dimerize to form an active enzyme and the question arising from this was if the enzyme is already present as dimer in solution, as it is known for PvuII alone [96], or if the fusion to I-SceI affects the dimerization potential. This would mean that the PvuII subunits of the fusion enzyme dimerize on the DNA due to binding of the I-SceI binding modules to their target sites in a distinct manner. This property would be advantageous for the purpose of using the fusion enzyme as nuclease with extended specificity in gene targeting and would avoid a so-called "hanging monkey" situation, where the cleavage module by itself binds its recognition sequence without being specifically directed by the binding modules. Sedimentation velocity runs of P-$L_{(6)}$-Ss* in an analytical ultracentrifugation performed under high- and low-salt conditions revealed that the fusion enzyme occurs also as dimer in solution, independent of protein concentration and salt content of the buffer, indicating that the I-SceI-fusion had no effect on the oligomeric state of PvuII. This experiment also confirmed the structure proposed in the model of the fusion enzyme, having an elongated shape, rather than forming a spherical protein-complex, which is obvious, since the fusion enzyme is composed of formerly stand-alone proteins probably having no tendency to interact with each other.

Summarizing all obtained results so far, a homodimeric fusion protein was constructed, showing a pronounced specificity for an addressed tripartite recognition site consisting of a PvuII site flanked by two I-SceI sites ideally 6 bp apart. Different linker variants show different extents of specificity, with P-$L_{(6)}$-Ss* and P-$L_{(+)}$-Ss* being the favorites so far. For this reason, these two protein variants were the starting point for further optimization. All further improvements were planned to happen on the cleavage module, to

increase the specificity and make the cleavage activity of PvuII more dependent on binding of I-SceI to the DNA. As described in Figure 3.22, two different approaches were followed. One was based on the catalytic properties of PvuII, which should be decreased by certain amino acid substitutions and the other one relied on the dimeric structure of PvuII, which should be disturbed by mutagenesis in the dimer interface of the protein.

In the first more functional approach, substitutions T46G, Y94F or H83A (Figure 3.22, right panel; Figure 4.2D) were introduced into PvuII in different combinations, with T46G always being present. This mutation is responsible for abolishing the "star activity" of PvuII (Zheng, NEB, personal communication), a feature suitable for the goal pursued here. Additional to this, Y94F was reported to have a reduced binding ability to the second metal ion [100], resulting in a reduced catalytic activity while H83A was reported to be involved in DNA-binding [99]. These mutations were introduced as double or triple mutants into PvuII, either in the fusion enzyme context or alone. That these substitutions already have a big impact on the catalytic activity of PvuII is shown in Figure 3.28. $PvuII_{(T46G,\ Y94F)}$ has a 100 - fold reduced cleavage activity and $PvuII_{(T46G,\ H83A,\ Y94F)}$ an even 200 - fold reduced activity compared to $PvuII_{wt}$. The variant $PvuII_{(T46G,\ H83A)}$ was not tested, since an experiment, accomplished before, showed that fusion enzyme variants containing this cleavage module did not have a specificity only for the tripartite recognition site but also for the dipartite, consisting of a PvuII site with just one I-SceI site next to it. However, this experiment (Figure **3.23**) showed that all tested variants ($P_{(T46G,H83A)}$-$L_{(6)}$-Ss*, $P_{(T46G,Y94F)}$-$L_{(6)}$-Ss*, $P_{(T46G,H83A)}$-$L_{(+)}$-Ss* and $P_{(T46G,Y94F)}$-$L_{(+)}$-Ss*) did not cleave the unaddressed site also present on the substrates. The variants containing $PvuII_{(T46G,\ Y94F)}$ as cleavage module in addition showed a specificity for the addressed tripartite site (S6P6S) consisting of a PvuII site flanked by two I-SceI sites 6 bp apart and did not cleave the dipartite site (S7P). Since this observation already means a gain in specificity, only variants containing $PvuII_{(T46G,\ Y94F)}$ were used for further characterization, including detailed enzyme kinetics. Both enzymes tested, $P_{(T46G,\ Y94F)}$-$L_{(6)}$-Ss* and $P_{(T46G,\ Y94F)}$-$L_{(+)}$-Ss*, showed no cleavage of the unaddressed site on a plasmid containing the addressed (S6P6S) and the unaddressed (P) PvuII site even after overnight incubation (Figure 3.24), although they did cleave the addressed site completely after this time. Since in this experiment enzyme and DNA were incubated in equimolar amounts, but the plasmid has two sites for the enzyme to attack, the ratio of enzyme to cleavable sites is 1:2. To test if the enzyme will still be specific if this ratio changes, enzyme titration experiments were performed with enzyme to DNA ratios of 0.5:1 – 32:1, with plasmid DNA containing the addressed and the unaddressed site (S6P6S_P) (Figure 3.25A) or

plasmid DNA containing only a PvuII site (Figure 3.25B). Actually, $P_{(T46G, Y94F)}$-$L_{(6)}$-Ss* cleaves the PvuII site in presence of the addressed site only at a 16 - fold excess of protein over DNA and in absence of an addressed site at 8 - fold excess. The other fusion protein with the linker $L_{(+)}$ is not as specific, cleaving PvuII sites in presence of S6P6S at 8 - fold excess and in absence of the addressed site already at equimolar concentrations of protein and DNA.

So far, the evaluation of the cleavage activity was performed by ethidium bromide staining, which is not sensitive enough to also detect small traces of cleavage product, why competition cleavage experiments using internally radiolabeled PCR-fragments harboring either the addressed site (S6P6S) or a single PvuII site (P) were done. These experiments, summarized in Figure 3.26, revealed that there is at least some unaddressed cleavage (~ 3 % after 3 h) in the presence of equimolar amounts of addressed site by the fusion proteins, independent of the linker ($L_{(6)}$ or $L_{(+)}$) or the PvuII variant (wt or T46G, Y94F) used. In terms of addressed cleavage, the variants with $PvuII_{wt}$ as cleavage module show remarkably specific cleavage activity while the variants $P_{(T46G, Y94F)}$-$L_{(6)}$-Ss* and $P_{(T46G, Y94F)}$-$L_{(+)}$-Ss* showed 34 and 16 times less activity, which is indeed not as dramatic as for the variants with PvuII alone, showing a 100 - fold drop in activity (compare Figure 3.28), but still too little to see any effect on the specificity of these variants, particularly because these enzyme variants show nearly the same extent of unaddressed cleavage as P-$L_{(6)}$-Ss* and P-$L_{(+)}$-Ss*. This results in a specificity factor of 18 for $P_{(T46G, Y94F)}$-$L_{(6)}$-Ss* and 11 for $P_{(T46G, Y94F)}$-$L_{(+)}$-Ss* (Table 3.8), which indicates a loss in specificity compared to the same variants having $PvuII_{wt}$ as cleavage module.

During all cleavage experiments by the fusion enzyme, it was observable that after a certain time the cleavage activity of the enzymes did not increase anymore, although not all substrate was turned over, the activity reached a plateau. One explanation for this could be, that not all enzyme molecules, which of course absorb at A_{280} and by this contribute to the concentration, were still active. Another explanation was derived from the knowledge that homing endonucleases have the tendency to stick to their cleavage products after cleavage [70,108]. This was tested by simply analyzing the cleavage products of P-$L_{(6)}$-Ss* and P-$L_{(+)}$-Ss* on a native gel either after heating the sample or not (Figure 3.27). When not heated up, the DNA appeared as shifted in the gel, while after heating, and by this denaturing the enzyme sticking on the DNA, the cleavage products were observed. This result indicated that the fusion enzymes are *per se* no “real” enzymes which must have a turn-over, a fact that could be useful for application *in vivo*, since in the ideal case the fusion enzyme has to cleave

only once at the desired position. Sticking to the products should not be an obstacle for initiating homologous recombination, as homing endonucleases also tend to do this and still their only function is stimulation of homologous recombination to preserve the propagation of genes coding for them [65]. These results obtained were already a sign for I-SceI being the major part of the fusion enzymes involved in DNA-binding, as it was aimed to. The binding constants derived by EMSA on specific substrate (S6P6S) confirmed this, although these K_Ds were 2 - 3 times higher than the one reported for I-SceI alone [19]. A single I-SceI site is bound at least 12 times less efficient by the fusion enzyme, which indicates that PvuII also contributes to the binding ability of the fusion enzyme to its specific site but disrupts the binding of I-SceI to its site alone. The obtained data were in principle independent of the linker and the PvuII variant used.

Although the so far created and optimized fusion enzyme variants $P_{(T46G,\ Y94F)}$-$L_{(6)}$-Ss* and $P_{(T46G,\ Y94F)}$-$L_{(+)}$-Ss* did not show any increase in specificity compared to the enzymes containing $PvuII_{wt}$ as cleavage module, an *in vivo* test in *E.coli* was performed, investigating the potential off-site target cleavage by these variants. Figure 3.28 summarizes the results obtained thereby, indicating that the plasmids expressing $PvuII_{wt}$ or $PvuII_{(T46G,\ Y94F)}$ need the presence of their companion methyltransferase to protect the *E.coli* genome from autodigestion, while *E.coli* transformed with plasmids encoding the different fusion enzymes (P-$L_{(6)}$-Ss*, P-$L_{(+)}$-Ss*, $P_{(T46G,\ Y94F)}$-$L_{(6)}$-Ss*, $P_{(T46G,\ Y94F)}$-$L_{(+)}$-Ss*, $P_{(T46G,\ H83A,\ Y94F)}$-$L_{(6)}$-Ss* and $P_{(T46G,\ H83A,\ Y94F)}$-$L_{(+)}$-Ss*) in the absence of methyltransferase showed surviving colonies, indicating that these enzymes did not attack genomic PvuII sites (whereof there are ~ 1800 in the *E.coli* DH10B genome) to the same extent as PvuII does. It is to mention that *E.coli* transformed with plasmid coding for $PvuII_{(T46G,\ H83A,\ Y94F)}$ also showed surviving colonies in the absence of M.PvuII confirming the too low activity of this enzyme variant (compare Figure 3.28). Overall, the amount of surviving cells in presence of fusion enzyme and absence of methyltransferase is not as high as would be expected from the control plates, showing the number of colonies growing in presence of M.PvuII, an indication that there is still off-site target cleavage by the fusion enzymes. Additionally, there is again a discrepancy between *in vitro* and *in vivo* data observable, because the variants having $PvuII_{(T46G,\ Y94F)}$ as cleavage module show by trend a higher amount of surviving cells although the factor for specificity of these variants obtained *in vitro* was at least one order of magnitude lower than for the fusion proteins containing $PvuII_{wt}$ (Table 3.8). The higher survival of cells for these variants could possibly be explained by their overall decreased activity.

After trying to optimize the fusion proteins by decreasing the catalytic properties of PvuII, which turned out to be not as successful as expected, the second approach by interrupting the dimerization interface of PvuII was undertaken (Figure 4.2). This time, first of all the PvuII alone variants harboring the substitutions L12E, P14G or H15D were examined for their catalytic properties (Table 3.7), which turned out to be at maximum 7 - fold lower ($PvuII_{(L12E)}$) than those for $PvuII_{wt}$. Since all the substitutions were located in the first N-terminal α-helix (Figure 3.22, lower panel), distant from the active center of PvuII, these results were expected. The aim of introducing the mutations into this helix was disturbing the dimer interface by creating on one hand a negative patch (L12E and H15D) and on the other hand changing the bending behavior of the helix (P14G) and by this reducing the interface contacts. The resulting fusion enzymes containing the before mentioned mutations individually in the PvuII cleavage module showed cleavage activity of plasmid DNA containing the addressed (S6P6S) and an additional unaddressed PvuII site (P) only at the addressed site and no cleavage of plasmid DNA containing only a PvuII site (Figure 3.30). Since the addressed cleavage obtained in this experiment was nearly complete and the PvuII alone variants containing the mutations L12E, P14G or H15D also showed activity comparable to $PvuII_{wt}$ it was expected that these variants also show higher cleavage activity in the competition assay than the variants containing $PvuII_{(T46G,\ Y94F)}$ as cleavage module. This was exactly the case as shown in Figure 3.31, where all three variants $P_{(L12E)}$-$L_{(6)}$-Ss*, $P_{(P14G)}$-$L_{(6)}$-Ss* and $P_{(H15D)}$-$L_{(6)}$-Ss* showed remarkable cleavage of the addressed site, while no unaddressed cleavage could be observed, resulting in a factor of specificity of these enzymes for the addressed site of over 1000 (Table 3.8). This factor might be even higher, since the amount of unaddressed cleavage is below the detection limit. The activity of these variants at the addressed substrate is only slightly lower than the addressed activity of P-$L_{(6)}$-Ss* (Table 3.8), which again confirms the results obtained before that show that the substitutions in the dimer interface do not have such a strong impact on the overall activity of the fusion enzyme but a large effect on their specificity. Surprisingly, also the binding affinity on the addressed site of the variants $P_{(L12E)}$-$L_{(6)}$-Ss*, $P_{(P14G)}$-$L_{(6)}$-Ss* and $P_{(H15D)}$-$L_{(6)}$-Ss* (Table 3.9) is higher than for the variants containing $PvuII_{wt}$ or $PvuII_{(T46G,\ Y94F)}$ as cleavage module (Table 3.3), while no specific binding of a stand-alone I-SceI site could be observed (not shown). A possible explanation for this could be that due to the mutations in the interface, PvuII is more flexible and by this maybe interferes with I-SceI and disturbs the binding of I-SceI to its site alone, while in the presence of a PvuII site between two I-SceI sites, PvuII can bind and is not affecting I-SceI anymore.

In the very first modeling approach of the fusion enzyme (Figure 3.12), it was suggested that a 6 bp distance between I-SceI and PvuII target sites would be the best, since then all enzymes enter the DNA from the same side, and also in the DNA cleavage assay of the different linker variants, pAT_S6P6S was the preferred substrate for P-$L_{(6)}$-Ss* (Figure 3.20), which is the variant used in the optimized enzymes last. If the fusion enzymes had not only a preference for an addressed site, meaning PvuII site flanked by two I-SceI sites, but furthermore had a preference for a certain distance between these sites, this enzyme would be even more specific. Plasmid cleavage assays with substrates containing the sites S4P4S, S6P6S and S8P8S (Figure 3.32), respectively, comparing the activity of $P_{(L12E)}$-$L_{(6)}$-Ss*, $P_{(P14G)}$-$L_{(6)}$-Ss* and $P_{(H15D)}$-$L_{(6)}$-Ss*, revealed that indeed $P_{(P14G)}$-$L_{(6)}$-Ss* showed a 8 - 10 - fold preference for S6P6S over S4P4S and S8P8S, making this enzyme the most specific of all. The other two variants also showed a preference for S6P6S over the other substrates, but with at maximum two fold. Finally, all fusion enzyme variants obtained so far, containing the different linker and PvuII variants, were tested for cleavage at unaddressed PvuII sites by λ-DNA digestion (Figure 3.33). When using equimolar concentration of enzyme and PvuII sites, the "early" variants, containing $PvuII_{wt}$ as cleavage module and the different linker, showed cleavage of λ-DNA, albeit not as extensive as PvuII alone does. The variants containing $PvuII_{(T46G, Y94F)}$ as cleavage module did not show any cleavage of λ-DNA, which is not surprising, since $PvuII_{(T46G, Y94F)}$ by itself does not cleave as much as $PvuII_{wt}$ and was also shown to have a 100 - fold less efficient cleavage than $PvuII_{wt}$ (Figure 3.28). More impressive are the results obtained for $P_{(L12E)}$-$L_{(6)}$-Ss* and $P_{(P14G)}$-$L_{(6)}$-Ss* which do not show any cleavage of λ-DNA, although $PvuII_{(L12E)}$ and $PvuII_{(P14G)}$ cleaved the DNA completely, comparable to $PvuII_{wt}$. This result confirmed the calculated specificity of over 1000 (Table 3.8) for addressed over unaddressed sites for these variants. For some reason, $P_{(H15D)}$-$L_{(6)}$-Ss* showed some cleavage of λ-DNA, even though this enzyme also has a specificity of > 1000 for addressed over unaddressed cleavage. It could be that this enzyme attacks addressed sites first in the presence of unaddressed ones, and then sticks to the cleavage product (see Figure 3.27), which makes it unable to attack unaddressed sites, but if there are just PvuII sites, it will attack these, which shows that the calculated specificity does not necessarily refer to the "real" specificity of the enzyme.

Summing all the results up, it was possible to engineer a highly specific nuclease by fusing a catalytically inactive homing endonuclease to the restriction enzyme PvuII. The final fusion enzyme $P_{(P14G)}$-$L_{(6)}$-Ss* cleaves with > 1000 - fold preference addressed sites in the presence of unaddressed sites, but also does not attack PvuII sites alone. Furthermore, it has

specificity for the 6 bp distance between I-SceI and PvuII target sites. Of course the last step has to be *in vivo* testing of this enzyme, which should ideally be in eukaryotic cells, e.g. plant cells, where already endeavors have been made to introduce I-SceI target sequences as so-called “safe harbors” into crop plant cells for gene targeting [84] [109].

4.3 CONCLUSION

In this work, two different approaches for generating meganucleases have been described, with both having their advantages and disadvantages. The first assay based on directed evolution aims for changing the specificity of an already existing meganuclease, while the second approach based on rational engineering enlarges the specificity of a restriction enzyme by fusion to a catalytically inactive homing endonuclease. Advantage of the directed evolution approach is that one can select for enzymes recognizing a specific sequence existing in a genome and afterwards target this sequence directly by the meganuclease for gene targeting as it is done in high throughput by the company Cellectis SA. They created a large library of I-CreI subunits recognizing different sequences and are able to combine these subunits to engineer enzymes recognizing sequences at specific loci i.e. *XPC* [23], *RAG1* [24] or HSV1 [27].

For PI-SceI, the object of investigation in this work, until now, no selection assays have been presented, although it was thought that this enzyme with its large recognition sequence of > 31 bp would be a suitable candidate. Development of a selection assay for PI-SceI revealed the disadvantages of directed evolution systems, i.e. the selection pressure in negative selection leading to false positive clones, the high expenditure of time, since only a few base pairs in the recognition sequence should be changed at once and particularly the lack of rationality in these approaches. The last fact is the striking one, since homing endonucleases like PI-SceI are complex enzymes with a large network of residues responsible for DNA-recognition and it turned out not to be straight forward to pick out certain residues that are maybe involved in DNA-binding and just changing them, as it was done here. For a successful changing of the recognition pattern, larger parts of the enzyme have to be assayed in a selection system, which could be problematic, since the larger the region to mutate, the higher the probability to affect also the activity of the enzyme. A further obstacle in the selection assay and also in the later described fusion enzyme approach is always the discrepancy between *in vivo* and *in vitro* results for the meganucleases. In the directed evolution part, *in vivo* variants of PI-SceI with an obviously new specificity have been obtained, which could not be confirmed by *in vitro* studies.

The advantage of the second approach is the rationality by combining a DNA-binding and a DNA-cleavage module to enlarge the recognition sequence of the last mentioned, a principle used at the moment intensively to generate highly specific nucleases. In this context Zinc-finger nucleases, e.g. [110], and TALE-nucleases, e.g. [50], have to be noted, both relying on fusion of the catalytic domain of FokI as cleavage module to programmable DNA-binding modules. In this work, a catalytically inactive form of the homing endonuclease I-SceI served as binding module which is *per se* not programmable, unless by directed evolution approaches as described before to find variants with new specificities. Still, this does not have to be a disadvantage, since there have already been a lot of efforts made on introducing I-SceI target sites for example in plant cells. For this reason, the here developed fusion enzyme could be a candidate for gene targeting in crop species, a field which is right now under intensive investigation. In contrast to the above mentioned approaches, where they used an unspecific cleavage module, which by itself does not contribute to the specificity of the nuclease, in this work the specific type IIP restriction endonuclease PvuII was used. This might have the advantage that the number of off-site target cleavage sites would be reduced, since PvuII recognizes the 6 bp sequence CAGCTG, which still occurs 10^4 - 10^8 times in complex genomes. However, by optimizing the linker between I-SceI and PvuII and by introducing certain substitutions into PvuII, affecting the dimer interface, it could be shown that the final fusion enzyme variant, namely $P_{(P14G)}$-$L_{(6)}$-Ss*, cleaves addressed sites with > 1000-fold preference over unaddressed sites, does not attack stand-alone PvuII site in λ-DNA and has also a preference for a specific distance of 6 bp between the I-SceI and PvuII target sites. The results obtained so far *in vitro* for the fusion enzyme are very promising and the next step would be *in vivo* testing of the enzyme. Since there have also been efforts made in controlling the activity of PvuII by either caging [111] or light inducible switches [112], this could also be applied to the fusion enzyme, making its cleavage activity controllable. Nevertheless it could be shown, that a specific restriction endonuclease, rather than only the catalytic domain of FokI, is a suitable alternative cleavage module to be used in creating highly specific nucleases.

5. SUMMARY

In the post-genomic era, highly specific nucleases became important tools for directed gene targeting and gene therapy. They are used to specifically cleave genomic sequences, involved in for instance diseases and by creating this double-strand break, homologous recombination, an otherwise rarely occurring event in the cell, will be stimulated. Most commonly used are zinc-finger nucleases, consisting of specific zinc-finger modules responsible for recognizing DNA, fused to the catalytic domain of the restriction enzyme FokI. Some of these zinc-finger nucleases are already in clinical trial phases. As alternative in the past three years so-called TALE-nuclease came into the focus of research. Their advantage might be the high predictability of specificity and their possibly lower cytotoxic effects due to off-site target-cleavage, which have to be confirmed in the future.

In this work, two other approaches for generating highly specific nucleases were presented. The first one is based on directed evolution, where the specificity of a naturally occurring meganuclease was to be changed. The investigated enzyme belonging to the LAGLIDADG family of homing endonucleases was PI-SceI. Mutant libraries of this enzyme were assayed in a two-plasmid selection system, selecting for variants cleaving a target site differing from the natural PI-SceI target in up to seven bp. Since structural and biochemical data for PI-SceI are available, only residues known to be involved in DNA-binding were used for mutagenesis. The obtained variants, showing *in vivo* an up to 100 - fold increase in specificity did not show the same result *in vitro*, under all conditions tested. This showed that results obtained *in vivo* and *in vitro* do not necessarily correspond to each other and that the developed selection assay needs further optimization.

As second approach, a more rational strategy similar to zinc-finger or TALE nucleases was pursued, by fusion of a catalytically inactive variant of the LAGLIDADG homing endonuclease I-SceI to the type IIP restriction enzyme PvuII. With this, an alternative to the often used catalytic domain of FokI as cleavage module was presented. After optimizing the linker between the two proteins and several mutations in I-SceI to abolish any catalytic activity (D44N, D145A) and mutations in PvuII to either decrease its activity (T46G, H83A or Y94F) or disturb the dimerization interface (L12E, P14G or H15D) a final variant, namely $P_{(P14G)}$-$L_{(6)}$-Ss*, was obtained. This enzyme showed *in vitro* an over 1000 - fold preference for the addressed (PvuII site flanked by two I-SceI sites) over an unaddressed site (PvuII site alone). Furthermore, this protein had an up to 10 - fold preference for the distance of 6 bp between the respective I-SceI and PvuII sites and showed no cleavage of genomic DNA

(bacteriophage λ-DNA). The next step would be *in vivo* testing of this enzyme variant, ideally *in planta*, since there are already I-SceI target sites in several crop species available that could serve as “landing platforms” for gene targeting.

6. ZUSAMMENFASSUNG

In der post-genomischen Ära haben sich hoch spezifische Nukleasen als wichtige Werkzeuge für die Gentherapie entwickelt. Sie sollen genutzt werden, um genomische DNS-Sequenzen, die zum Beispiel in monogenetisch vererbbaren Krankheiten involviert sind, spezifisch zu spalten. Der erzeugte Doppelstrangbruch stimuliert den zelleigenen Reparatur-mechanismus der homologen Rekombination, ein Ereignis, das ansonsten selten in der Zelle vorkommt. Die momentan am Häufigsten genutzten hochspezifischen Nukleasen sind sogenannte Zink-Finger Nukleasen. Diese bestehen aus Zink-Finger Modulen, die ver-antwortlich für die spezifische DNS-Bindung sind, fusioniert an die katalytische Domäne des Restriktions-Enzyms FokI. Einige dieser Zink-Finger Nukleasen befinden sich schon in klinischen Versuchs-Phasen. In den letzten drei Jahren sind, als Alternative zu den Zink-Finger Nukleasen, sogenannte TALE-Nukleasen in den Fokus der Forschung gerückt. Ihr Vorteil ist die Programmierbarkeit ihrer Spezifität und die damit verbundene wahrscheinlich geringere Zytotoxizität. All das wird sich allerdings in der Zukunft noch bestätigen müssen.

In der hier vorgestellten Arbeit werden zwei andere Ansätze zur Generierung hoch spezifischer Nukleasen präsentiert. Der erste basiert auf gerichteter Evolution, wobei die Spezifität einer natürlich vorkommenden Meganuklease geändert werden soll. Das dabei untersuchte Enzym, PI-SceI, gehört der Familie der LAGLIDADG-Homingendonukleasen an. Bibliotheken von Enzym-Varianten wurden in einem zuvor entwickelten Zwei-Plasmid System gegen Erkennungssequenzen, die sich von der natürlichen PI-SceI Erkennungs-sequenz in bis zu sieben Basenpaaren unterscheiden, selektioniert. Aufgrund bekannter struktureller und biochemischer Daten bezüglich PI-SceI, wurden nur Aminosäuren der Mutagenese unterworfen, welche in der DNS-Bindung involviert sind. Die in dieser Arbeit selektionierten Varianten von PI-SceI zeigten *in vivo* eine bis zu 100 - fach gesteigerte Spezifität gegenüber der neuen Erkennungssequenz, ein Ergebnis, dass *in vitro* unter keiner der getesteten Bedingungen bestätigt werden konnte. Dies ließ darauf schließen, dass *in vivo* und *in vitro* beobachtete Ergebnisse nicht notwendigerweise übereinstimmen und dass der hier entwickelte Assay weiter optimiert werden muss.

Der zweite Ansatz verfolgt eine rationalere Strategie, ähnlich derer von Zink-Finger und TALE-Nukleasen. Hierbei wurde eine katalytisch inaktive Variante der LAGLIDADG Homingendonuklease I-SceI an das spezifische Typ IIP Restriktionsenzym PvuII fusioniert. Letzteres soll eine Alternative zur ansonsten so häufig genutzten katalytischen Domäne von FokI darstellen. Das entstandene Fusions-Enzym wurde weiterhin in der *Linker*-Region,

welche I-SceI und PvuII miteinander verbindet, optimiert. Außerdem wurden Mutationen in I-SceI eingeführt, um sämtliche katalytische Aktivität zu unterbinden (D44N und D145A). Auch PvuII wurde optimiert, indem entweder durch Aminosäure-Substitutionen die katalytische Aktivität reduziert wurde (T46G, H83A oder Y94F) oder indem das Dimerisierungspotential verringert wurde (L12E, P14G oder H15D). Das Resultat all dieser Optimierungen stellt die finale Variante $P_{(P14G)}$-$L_{(6)}$-Ss* dar. Dieses Enzym zeigt *in vitro* eine über 1000 - fache Präferenz für eine adressierte Erkennungssequenz (PvuII Erkennungssequenz flankiert von zwei I-SceI Erkennungssequenzen) gegenüber einer unadressierten Sequenz (PvuII Erkennungssequenz allein). Außerdem hat es eine bis zu 10 - fache Präferenz für einen Abstand von sechs Basenpaaren zwischen den I-SceI und PvuII Erkennungssequenzen und zeigt keinerlei Spaltung von genomischer DNS (Bakteriophagen λ DNS). Als nächsten Schritt sollte $P_{(P14G)}$-$L_{(6)}$-Ss* *in vivo* getestet werden. Dies könnte idealerweise in Pflanzen stattfinden, da es mittlerweile in einer Auswahl von bestimmten Kulturpflanzen I-SceI Erkennungssequenzen als Insertionsstellen für genetische Modifikationen gibt.

7. REFERENCES

1. Smith, H.O., Tomb, J.F., Dougherty, B.A., Fleischmann, R.D. and Venter, J.C. (1995) Frequency and distribution of DNA uptake signal sequences in the Haemophilus influenzae Rd genome. *Science*, **269**, 538-540.
2. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921.
3. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A. *et al.* (2001) The sequence of the human genome. *Science*, **291**, 1304-1351.
4. Smih, F., Rouet, P., Romanienko, P.J. and Jasin, M. (1995) Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Res.*, **23**, 5012-5019.
5. Pingoud, A. and Wende, W. (2011) Generation of novel nucleases with extended specificity by rational and combinatorial strategies. *ChemBioChem*, **12**, 1495-1500.
6. Stoddard, B.L. (2005) Homing endonuclease structure and function. *Q. Rev. Biophys.*, **38**, 49-95.
7. Pingoud, A., Pingoud, V. and Wende, W. (2003) Homingendonukleasen - Vielversprechende Werkzeuge für das Genomengineering. *BIOspektrum*, **5**, 592-595.
8. Gimble, F.S. (2005) In Belfort, M., Derbyshire, V., Stoddard, B. L. and Wood, D. W. (eds.), *Homing endonucleases and inteins*. Springer-Verlag, Berlin Heidelberg, pp. 175-190.
9. Chevalier, B.S., Kortemme, T., Chadsey, M.S., Baker, D., Monnat, R.J. and Stoddard, B.L. (2002) Design, activity, and structure of a highly specific artificial endonuclease. *Mol. Cell*, **10**, 895-905.
10. Steuer, S., Pingoud, V., Pingoud, A. and Wende, W. (2004) Chimeras of the homing endonuclease PI-SceI and the homologous Candida tropicalis intein: a study to explore the possibility of exchanging DNA-binding modules to obtain highly specific endonucleases with altered specificity. *ChemBioChem*, **5**, 206-213.
11. Silva, G.H., Belfort, M., Wende, W. and Pingoud, A. (2006) From monomeric to homodimeric endonucleases and back: engineering novel specificity of LAGLIDADG enzymes. *J. Mol. Biol.*, **361**, 744-754.
12. Ashworth, J., Havranek, J.J., Duarte, C.M., Sussman, D., Monnat, R.J., Jr., Stoddard, B.L. and Baker, D. (2006) Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature*, **441**, 656-659.
13. Ashworth, J., Taylor, G.K., Havranek, J.J., Quadri, S.A., Stoddard, B.L. and Baker, D. (2010) Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. *Nucleic Acids Res.*, **38**, 5601-5608.
14. Seligman, L.M., Chisholm, K.M., Chevalier, B.S., Chadsey, M.S., Edwards, S.T., Savage, J.H. and Veillet, A.L. (2002) Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic Acids Res.*, **30**, 3870-3879.
15. Gimble, F.S., Moure, C.M. and Posey, K.L. (2003) Assessing the plasticity of DNA target site recognition of the PI-SceI homing endonuclease using a bacterial two-hybrid selection system. *J. Mol. Biol.*, **334**, 993-1008.
16. Sussman, D., Chadsey, M., Fauce, S., Engel, A., Bruett, A., Monnat, R., Jr., Stoddard, B.L. and Seligman, L.M. (2004) Isolation and characterization of new homing endonuclease specificities at individual target site positions. *J. Mol. Biol.*, **342**, 31-41.
17. Gruen, M., Chang, K., Serbanescu, I. and Liu, D.R. (2002) An in vivo selection system for homing endonuclease activity. *Nucleic Acids Res.*, **30**, e29.

18. Chen, Z. and Zhao, H. (2005) A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic Acids Res.*, **33**, e154.
19. Doyon, J.B., Pattanayak, V., Meyer, C.B. and Liu, D.R. (2006) Directed evolution and substrate specificity profile of homing endonuclease I-SceI. *J. Am. Chem. Soc.*, **128**, 2477-2484.
20. Chames, P., Epinat, J.C., Guillier, S., Patin, A., Lacroix, E. and Paques, F. (2005) In vivo selection of engineered homing endonucleases using double-strand break induced homologous recombination. *Nucleic Acids Res.*, **33**, e178.
21. Arnould, S., Chames, P., Perez, C., Lacroix, E., Duclert, A., Epinat, J.C., Stricher, F., Petit, A.S., Patin, A., Guillier, S. *et al.* (2006) Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J. Mol. Biol.*, **355**, 443-458.
22. Smith, J., Grizot, S., Arnould, S., Duclert, A., Epinat, J.C., Chames, P., Prieto, J., Redondo, P., Blanco, F.J., Bravo, J. *et al.* (2006) A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res.*, **34**, e149.
23. Arnould, S., Perez, C., Cabaniols, J.P., Smith, J., Gouble, A., Grizot, S., Epinat, J.C., Duclert, A., Duchateau, P. and Paques, F. (2007) Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *J. Mol. Biol.*, **371**, 49-65.
24. Grizot, S., Smith, J., Daboussi, F., Prieto, J., Redondo, P., Merino, N., Villate, M., Thomas, S., Lemaire, L., Montoya, G. *et al.* (2009) Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. *Nucleic Acids Res.*, **37**, 5405-5419.
25. Munoz, I.G., Prieto, J., Subramanian, S., Coloma, J., Redondo, P., Villate, M., Merino, N., Marenchino, M., D'Abramo, M., Gervasio, F.L. *et al.* (2011) Molecular basis of engineered meganuclease targeting of the endogenous human RAG1 locus. *Nucleic Acids Res.*, **39**, 729-743.
26. 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.
27. Grosse, S., Huot, N., Mahiet, C., Arnould, S., Barradeau, S., Clerre, D.L., Chion-Sotinel, I., Jacqmarcq, C., Chapellier, B., Ergani, A. *et al.* (2011) Meganuclease-mediated Inhibition of HSV1 Infection in Cultured Cells. *Mol. Ther.*, **19**, 694-702.
28. Metzger, M.J., McConnell-Smith, A., Stoddard, B.L. and 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.
29. Niu, Y., Tenney, K., Li, H. and Gimble, F.S. (2008) Engineering variants of the I-SceI homing endonuclease with strand-specific and site-specific DNA-nicking activity. *J. Mol. Biol.*, **382**, 188-202.
30. McConnell Smith, A., Takeuchi, R., Pellenz, S., Davis, L., Maizels, N., Monnat, R.J., Jr. and Stoddard, B.L. (2009) Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 5099-5104.
31. Klug, A. (2010) The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu. Rev. Biochem.*, **79**, 213-231.
32. Kim, Y.G., Cha, J. and Chandrasegaran, S. (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. U. S. A.*, **93**, 1156-1160.
33. Smith, J., Berg, J.M. and Chandrasegaran, S. (1999) A detailed study of the substrate specificity of a chimeric restriction enzyme. *Nucleic Acids Res.*, **27**, 674-681.

34. Smith, J., Bibikova, M., Whitby, F.G., Reddy, A.R., Chandrasegaran, S. and Carroll, D. (2000) Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res.*, **28**, 3361-3369.
35. Li, L., Wu, L.P. and Chandrasegaran, S. (1992) Functional domains in Fok I restriction endonuclease. *Proc. Natl. Acad. Sci. U. S. A.*, **89**, 4275-4279.
36. Wah, D.A., Bitinaite, J., Schildkraut, I. and Aggarwal, A.K. (1998) Structure of FokI has implications for DNA cleavage. *Proc. Natl. Acad. Sci. U. S. A.*, **95**, 10564-10569.
37. Catto, L.E., Ganguly, S., Milsom, S.E., Welsh, A.J. and Halford, S.E. (2006) Protein assembly and DNA looping by the FokI restriction endonuclease. *Nucleic Acids Res.*, **34**, 1711-1720.
38. Szczepek, M., Brondani, V., Buchel, J., Serrano, L., Segal, D.J. and Cathomen, T. (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat. Biotechnol.*, **25**, 786-793.
39. 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. *et al.* (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.*, **25**, 778-785.
40. Guo, J., Gaj, T. and Barbas, C.F., 3rd. (2010) Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. *J. Mol. Biol.*, **400**, 96-107.
41. Carroll, D. (2008) Progress and prospects: zinc-finger nucleases as gene therapy agents. *Gene Ther.*, **15**, 1463-1468.
42. 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. and Holmes, M.C. (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, **435**, 646-651.
43. Moehle, E.A., Rock, J.M., Lee, Y.L., Jouvenot, Y., DeKelver, R.C., Gregory, P.D., Urnov, F.D. and Holmes, M.C. (2007) Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc. Natl. Acad. Sci. U. S. A.*, **104**, 3055-3060.
44. Samson, M., Libert, F., Doranz, B.J., Rucker, J., Liesnard, C., Farber, C.M., Saragosti, S., Lapoumeroulie, C., Cognaux, J., Forceille, C. *et al.* (1996) Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature*, **382**, 722-725.
45. 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. *et al.* (2008) Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat. Biotechnol.*, **26**, 808-816.
46. Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509-1512.
47. Moscou, M.J. and Bogdanove, A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 1501.
48. Murakami, M.T., Sforca, M.L., Neves, J.L., Paiva, J.H., Domingues, M.N., Pereira, A.L., Zeri, A.C. and Benedetti, C.E. (2010) The repeat domain of the type III effector protein PthA shows a TPR-like structure and undergoes conformational changes upon DNA interaction. *Proteins*, **78**, 3386-3395.
49. Scholze, H. and Boch, J. (2011) TAL effectors are remote controls for gene activation. *Curr. Opin. Microbiol.*, **14**, 47-53.

50. Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J. and Voytas, D.F. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, **186**, 757-761.
51. Li, T., Huang, S., Jiang, W.Z., Wright, D., Spalding, M.H., Weeks, D.P. and Yang, B. (2011) TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.*, **39**, 359-372.
52. Mahfouz, M.M., Li, L., Shamimuzzaman, M., Wibowo, A., Fang, X. and Zhu, J.K. (2011) De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 2623-2628.
53. Schleifman, E.B., Chin, J.Y. and Glazer, P.M. (2008) Triplex-mediated gene modification. *Methods Mol. Biol.*, **435**, 175-190.
54. Eisenschmidt, K., Lanio, T., Simoncsits, A., Jeltsch, A., Pingoud, V., Wende, W. and Pingoud, A. (2005) Developing a programmed restriction endonuclease for highly specific DNA cleavage. *Nucleic Acids Res.*, **33**, 7039-7047.
55. Majumdar, A., Muniandy, P.A., Liu, J., Liu, J.L., Liu, S.T., Cuenoud, B. and Seidman, M.M. (2008) Targeted gene knock in and sequence modulation mediated by a psoralen-linked triplex-forming oligonucleotide. *J. Biol. Chem.*, **283**, 11244-11252.
56. Paques, F. and Duchateau, P. (2007) Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Curr. Gene. Ther.*, **7**, 49-66.
57. Cirino, P.C., Mayer, K.M. and Umeno, D. (2003) Generating mutant libraries using error-prone PCR. *Methods Mol. Biol.*, **231**, 3-9.
58. Zhao, H., Chockalingam, K. and Chen, Z. (2002) Directed evolution of enzymes and pathways for industrial biocatalysis. *Curr. Opin. Biotechnol.*, **13**, 104-110.
59. Tracewell, C.A. and Arnold, F.H. (2009) Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Curr. Opin. Chem. Biol.*, **13**, 3-9.
60. Kuzminov, A. and Stahl, F.W. (1997) Stability of linear DNA in recA mutant Escherichia coli cells reflects ongoing chromosomal DNA degradation. *J. Bacteriol.*, **179**, 880-888.
61. Remaut, E., Stanssens, P. and Fiers, W. (1981) Plasmid vectors for high-efficiency expression controlled by the PL promoter of coliphage lambda. *Gene*, **15**, 81-93.
62. Lieb, M. (1966) Studies of heat-inducible lambda bacteriophage. I. Order of genetic sites and properties of mutant prophages. *J. Mol. Biol.*, **16**, 149-163.
63. Hayes, P.K., Walsby, A.E. and Walker, J.E. (1986) Complete amino acid sequence of cyanobacterial gas-vesicle protein indicates a 70-residue molecule that corresponds in size to the crystallographic unit cell. *Biochem. J.*, **236**, 31-36.
64. Gimble, F.S. and Thorner, J. (1993) Purification and characterization of VDE, a site-specific endonuclease from the yeast Saccharomyces cerevisiae. *J. Biol. Chem.*, **268**, 21844-21853.
65. Belfort, M. and Roberts, R.J. (1997) Homing endonucleases: keeping the house in order. *Nucleic Acids Res.*, **25**, 3379-3388.
66. Moure, C.M., Gimble, F.S. and Quiocho, F.A. (2002) Crystal structure of the intein homing endonuclease PI-SceI bound to its recognition sequence. *Nat. Struct. Biol.*, **9**, 764-770.
67. Pingoud, V., Grindl, W., Wende, W., Thole, H. and Pingoud, A. (1998) Structural and functional analysis of the homing endonuclease PI-sceI by limited proteolytic cleavage and molecular cloning of partial digestion products. *Biochemistry*, **37**, 8233-8243.
68. Grindl, W., Wende, W., Pingoud, V. and Pingoud, A. (1998) The protein splicing domain of the homing endonuclease PI-sceI is responsible for specific DNA binding. *Nucleic Acids Res.*, **26**, 1857-1862.

69. Noel, A.J., Wende, W. and Pingoud, A. (2004) DNA recognition by the homing endonuclease PI-SceI involves a divalent metal ion cofactor-induced conformational change. *J. Biol. Chem.*, **279**, 6794-6804.
70. Wende, W., Grindl, W., Christ, F., Pingoud, A. and Pingoud, V. (1996) Binding, bending and cleavage of DNA substrates by the homing endonuclease Pl-SceI. *Nucleic Acids Res.*, **24**, 4123-4132.
71. Christ, F., Schoettler, S., Wende, W., Steuer, S., Pingoud, A. and Pingoud, V. (1999) The monomeric homing endonuclease PI-SceI has two catalytic centres for cleavage of the two strands of its DNA substrate. *EMBO J.*, **18**, 6908-6916.
72. Schottler, S., Wende, W., Pingoud, V. and Pingoud, A. (2000) Identification of Asp218 and Asp326 as the principal Mg2+ binding ligands of the homing endonuclease PI-SceI. *Biochemistry*, **39**, 15895-15900.
73. Gimble, F.S. and Thorner, J. (1992) Homing of a DNA endonuclease gene by meiotic gene conversion in Saccharomyces cerevisiae. *Nature*, **357**, 301-306.
74. Lippow, S.M., Aha, P.M., Parker, M.H., Blake, W.J., Baynes, B.M. and Lipovsek, D. (2009) Creation of a type IIS restriction endonuclease with a long recognition sequence. *Nucleic Acids Res.*, **37**, 3061-3073.
75. Monteilhet, C., Perrin, A., Thierry, A., Colleaux, L. and Dujon, B. (1990) Purification and characterization of the in vitro activity of I-Sce I, a novel and highly specific endonuclease encoded by a group I intron. *Nucleic Acids Res.*, **18**, 1407-1413.
76. Perrin, A., Buckle, M. and 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.
77. Moure, C.M., Gimble, F.S. and Quiocho, 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.
78. Thierry, A., Perrin, A., Boyer, J., Fairhead, C., Dujon, B., Frey, B. and Schmitz, G. (1991) Cleavage of Yeast and Bacteriophage-T7 Genomes at a Single Site Using the Rare Cutter Endonuclease I-Sce-I. *Nucleic Acids Res.*, **19**, 189-190.
79. Plessis, A., Perrin, A., Haber, J.E. and Dujon, B. (1992) Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. *Genetics*, **130**, 451-460.
80. Puchta, H., Dujon, B. and Hohn, B. (1993) Homologous Recombination in Plant-Cells Is Enhanced by in-Vivo Induction of Double-Strand Breaks into DNA by a Site-Specific Endonuclease. *Nucleic Acids Res.*, **21**, 5034-5040.
81. Lukacsovich, T., Yang, D. and Waldman, A.S. (1994) Repair of a specific double-strand break generated within a mammalian chromosome by yeast endonuclease I-SceI. *Nucleic Acids Res.*, **22**, 5649-5657.
82. Machida, C., Onouchi, H., Koizumi, J., Hamada, S., Semiarti, E., Torikai, S. and Machida, Y. (1997) Characterization of the transposition pattern of the Ac element in Arabidopsis thaliana using endonuclease I-SceI. *Proc. Natl. Acad. Sci. U. S. A.*, **94**, 8675-8680.
83. Cohen-Tannoudji, M., Robine, S., Choulika, A., Pinto, D., El Marjou, F., Babinet, C., Louvard, D. and Jaisser, F. (1998) I-SceI-induced gene replacement at a natural locus in embryonic stem cells. *Mol. Cell. Biol.*, **18**, 1444-1448.
84. D'Halluin, K., Vanderstraeten, C., Stals, E., Cornelissen, M. and Ruiter, R. (2008) Homologous recombination: a basis for targeted genome optimization in crop species such as maize. *Plant Biotechnol. J.*, **6**, 93-102.
85. Simoncsits, A., Tjornhammar, M.L., Rasko, T., Kiss, A. and Pongor, S. (2001) Covalent joining of the subunits of a homodimeric type II restriction endonuclease: single-chain PvuII endonuclease. *J. Mol. Biol.*, **309**, 89-97.

86. Joly, E. and Kirsch, R.D. (1998) An improved PCR-mutagenesis strategy for two-site mutagenesis or sequence swapping between related genes. *Nucleic Acids Res.*, **26**, 1848-1850.
87. Gill, S.C. and von Hippel, P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.*, **182**, 319-326.
88. Huang, Q. and Quinones, E. (2008) A spectroscopic method to determine the activity of the restriction endonuclease EcoRV that involves a single reaction. *Anal. Biochem.*, doi:10.1016/j.ab.2008.1004.1038.
89. He, Z.N., Crist, M., Yen, H.C., Duan, X.Q., Quiocho, F.A. and Gimble, F.S. (1998) Amino acid residues in both the protein splicing and endonuclease domains of the PI-SceI intein mediate DNA binding. *J. Biol. Chem.*, **273**, 4607-4615.
90. Wilson, D.S. and Keefe, A.D. (2001) Random mutagenesis by PCR. *Curr Protoc Mol Biol*, **Chapter 8**, Unit8 3.
91. Chusacultanachai, S. and Yuthavong, Y. (2004) Random mutagenesis strategies for construction of large and diverse clone libraries of mutated DNA fragments. *Methods Mol. Biol.*, **270**, 319-334.
92. Grishin, A., Fonfara, I., Alexeevski, A., Spirin, S., Zanegina, O., Karyagina, A., Alexeyevsky, D. and Wende, W. (2010) Identification of conserved features of LAGLIDADG homing endonucleases. *J. Bioinform. Comput. Biol.*, **8**, 453-469.
93. Gimble, F.S. and Wang, J. (1996) Substrate recognition and induced DNA distortion by the PI-*Sce*I endonuclease, an enzyme generated by protein splicing. *J. Mol. Biol.*, **263**, 163-180.
94. McClelland, M., Hanish, J., Nelson, M. and Patel, Y. (1988) KGB: a single buffer for all restriction endonucleases. *Nucleic Acids Res.*, **16**, 364.
95. Xie, F. and Dupureur, C.M. (2009) Kinetic analysis of product release and metal ions in a metallonuclease. *Arch. Biochem. Biophys.*, **483**, 1-9.
96. Athanasiadis, A., Vlassi, M., Kotsifaki, D., Tucker, P.A., Wilson, K.S. and Kokkinidis, M. (1994) Crystal structure of PvuII endonuclease reveals extensive structural homologies to EcoRV. *Nat. Struct. Biol.*, **1**, 469-475.
97. Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.*, **78**, 1606-1619.
98. Lebowitz, J., Lewis, M.S. and Schuck, P. (2002) Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci.*, **11**, 2067-2079.
99. Nastri, H.G., Evans, P.D., Walker, I.H. and Riggs, P.D. (1997) Catalytic and DNA binding properties of PvuII restriction endonuclease mutants. *J. Biol. Chem.*, **272**, 25761-25767.
100. Spyridaki, A., Matzen, C., Lanio, T., Jeltsch, A., Simoncsits, A., Athanasiadis, A., Scheuring-Vanamee, E., Kokkinidis, M. and Pingoud, A. (2003) Structural and biochemical characterization of a new Mg(2+) binding site near Tyr94 in the restriction endonuclease PvuII. *J. Mol. Biol.*, **331**, 395-406.
101. Duan, X., Gimble, F.S. and Quiocho, F.A. (1997) Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. *Cell*, **89**, 555-564.
102. Seligman, L.M., Stephens, K.M., Savage, J.H. and Monnat, R.J., Jr. (1997) Genetic analysis of the Chlamydomonas reinhardtii I-CreI mobile intron homing system in Escherichia coli. *Genetics*, **147**, 1653-1664.
103. Epinat, J.C., Arnould, S., Chames, P., Rochaix, P., Desfontaines, D., Puzin, C., Patin, A., Zanghellini, A., Paques, F. and Lacroix, E. (2003) A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res.*, **31**, 2952-2962.
104. Wu, J., Kandavelou, K. and Chandrasegaran, S. (2007) Custom-designed zinc finger nucleases: what is next? *Cell. Mol. Life Sci.*, **64**, 2933-2944.

105. Handel, E.M. and Cathomen, T. (2011) Zinc-finger nuclease based genome surgery: it's all about specificity. *Curr. Gene. Ther.*, **11**, 28-37.
106. Wah, D.A., Hirsch, J.A., Dorner, L.F., Schildkraut, I. and Aggarwal, A.K. (1997) Structure of the multimodular endonuclease FokI bound to DNA. *Nature*, **388**, 97-100.
107. Carroll, D., Morton, J.J., Beumer, K.J. and Segal, D.J. (2006) Design, construction and in vitro testing of zinc finger nucleases. *Nat. Protoc.*, **1**, 1329-1341.
108. Mueller, J.E., Smith, D. and Belfort, M. (1996) Exon coconversion biases accompanying intron homing: battle of the nucleases. *Genes Dev.*, **10**, 2158-2166.
109. Yang, M., Djukanovic, V., Stagg, J., Lenderts, B., Bidney, D., Falco, S.C. and Lyznik, L.A. (2009) Targeted mutagenesis in the progeny of maize transgenic plants. *Plant Mol. Biol.*, **70**, 669-679.
110. Durai, S., Mani, M., Kandavelou, K., Wu, J., Porteus, M.H. and Chandrasegaran, S. (2005) Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Res.*, **33**, 5978-5990.
111. Silanskas, A., Foss, M., Wende, W., Urbanke, C., Lagunavicius, A., Pingoud, A. and Siksnys, V. (2011) Photocaged variants of the MunI and PvuII restriction enzymes. *Biochemistry*, **50**, 2800-2807.
112. Schierling, B., Noel, A.J., Wende, W., Hien le, T., Volkov, E., Kubareva, E., Oretskaya, T., Kokkinidis, M., Rompp, A., Spengler, B. *et al.* (2010) Controlling the enzymatic activity of a restriction enzyme by light. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 1361-1366.