Stony Brook University

OFFICIAL COPY

The official electronic file of this thesis or dissertation is maintained by the University Libraries on behalf of The Graduate School at Stony Brook University.

© All Rights Reserved by Author.

Zinc Finger Nucleases: Custom Enzymes for DNA Engineering

A Dissertation Presented

by

Andriy Tovkach

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Genetics

Stony Brook University

May 2009

Stony Brook University

The Graduate School

Andriy Tovkach

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree,
hereby recommend acceptance of this dissertation.

Vitaly Citovsky, Ph.D. (Advisor), Professor
Department of Biochemistry and Cell Biology

Tzvi Tzfira, Ph.D. (co-Advisor), Assistant Professor
University of Michigan

Gerald H. Thomsen, Ph.D. (Chair), Professor
Department of Biochemistry and Cell Biology

Harvard Lyman, Ph.D., Professor
Department of Biochemistry and Cell Biology

Jörg Schwender, Ph.D., Associate Scientist
Brookhaven National Laboratory

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Zinc Finger Nucleases: Custom Enzymes for DNA Engineering

by

Andriy Tovkach

Doctor of Philosophy

in

Genetics

Stony Brook University

2009

Zinc finger nucleases (ZFNs) are custom restriction enzymes that can be designed to recognize and cleave DNA in virtually any specific user-defined location. In my thesis research, I took advantage of the unique DNA-recognition properties of ZFNs to develop novel approaches for molecular cloning and plant genome engineering. First, I described expression and purification methods for ZFN10, a novel custom restriction endonuclease, and characterized its biochemical properties. I demonstrated that Ni-affinity purification was sufficient to produce a cloning-grade enzyme and that ZFN10 was tolerant to a range of target-site substitutions which could be predicted from specificities of recognition helices incorporated into the structure of its zinc finger protein (ZFP) domain. Second, I showed that ZFNs can be used for standard molecular cloning applications along with known type II restriction enzymes, and that because of their long target sites and design flexibility, they can accomplish such challenging tasks as cloning very large DNA fragments and custom-cloning native DNA molecules. Third, I developed a comprehensive method for ZFN testing, which can be used by research groups wishing to

adopt ZFN technology for plant gene targeting. The testing consists of an *in-vitro* assay involving expression of a ZFN in *E. coli* with subsequent use of the bacterial lysate to digest a plasmid carrying the ZFN target site, and of a series of plant assays based on reconstruction of a GUS reporter's expression following transient or stable delivery of a mutated GUS-encoding gene and ZFN-expressing cassettes into target plant cells. This work expands the current collection of restriction endonucleases used in molecular cloning with potentially hundreds of thousands of new enzymes and provides the foundation for widespread use of ZFNs for targeted modification of plant genomes.

Table of Contents

List of Tables

List of Figures

Acknowledgements

Getting my Ph.D. has been a long journey. When I first looked at my admission papers, I could not believe that the “normal length of study” was seven years. I ended up extending this to nine years and four departments, and at every stop on the way I met amazing people whom I would like acknowledge for their help and support.

First, I want to thank Peter Gergren, the former director of Genetics at Stony Brook, who helped me get into the program, provided his support during hard times of lab transitions and helped arrange my studies in Michigan. Without Peter, I would never have been able to push through and succeed. I would also like to thank my advisor Dr. Tzvi Tzfira for creating an intellectually stimulating environment in the lab and for understanding my passion for the applied side of research. It was thanks to Tzvika’s vision and trust that I managed to venture successfully into the emerging field of zinc finger nuclease design. I am very grateful to my committee members Vitaly Citovsky, Harvard Lyman, Wolfgang Lukowitz, Jorg Schwender and Jerry Thomsen who created a supportive, result-oriented atmosphere during the committee meetings.

Working in the lab would not have been nearly as much fun without my current and former lab mates Vardit, Avner, Mery, Kamy, Raz, Dan, Ben, Shachi, Alex and Stas. Thank you all for your plants, competent cells, tea and board games. Special thanks to Kate Bell, for her devotion and concern for students' souls, and for accompanying each paperwork hurdle with the comforting realization that there was someone out there looking out for us. I want to thank Chris Kalesis at the Stony Brook International Office

and Kelly Campbell in the MCDB department at the University of Michigan for processing my CPT forms. Big thanks to David Helfman and Adrian Krainer from Cold Spring Harbor for creating a memorable lab rotation experience and sharing their excitement about science. I am grateful to Andrei Chistoserdov, Vanessa Madrid and Josephine Aller from the Marine Science Center for welcoming me to the US and for being my family during my first months away from home.

Finally, I would like to thank my brother and my parents for always being happy to pick up my phone calls, for supporting most of my new adventures and for concluding their advice with the “of course, you know better” clause.

The text of Chapter 3 of this Dissertation in full is a reprint of the materials as it appears in Zeevi, V., Tovkach, A. and Tzfira, T. Increasing cloning possibilities using artificial zinc finger nucleases. *PNAS*, 2008, 105, 12785-12790 (Copyright 2008 National Academy of Sciences, U.S.A.). The coauthors listed in the publication participated in and supervised the research that forms the basis for this Dissertation.

The text of Chapter 4 of this Dissertation in full is a reprint of the materials as it appears in Tovkach, A., Zeevi, V. and Tzfira, T. A toolbox and procedural notes for characterizing novel zinc finger nucleases for genome editing in plant cells. *Plant J.* 2009 Feb; 57(4):747-57 (Copyright 2009 Blackwell Science). The coauthors listed in the publication participated in and supervised the research that forms the basis for this Dissertation.

Publications

1. **Tovkach, A.** and Tzfira, T. (2009) Expression, purification and characterization of an artificial zinc finger nuclease (submitted)
2. **Tovkach, A**., Zeevi, V. and Tzfira, T. (2009) Validation and expression of ZFNs and in plant cells. In press. Methods in Molecular Biology (D. Segal and McKay, eds) (in press)
3. Zeevi, V., **Tovkach, A.** and Tzfira, T. (2009) Custom DNA Cloning by Zinc Finger Nucleases. In press. Methods in Molecular Biology (D. Segal and McKay, eds) (in press)
4. Dafny-Yelin, M., **Tovkach, A.** and Tzfira, T. (2009) Integration of *Agrobacterium* T-DNA in plant cells. In: *Functional Organization of the Plant Nucleus* I. Meier (ed.), Springer, pp. 157-186
5. **Tovkach, A.**, Zeevi, V. and Tzfira, T. (2009) A toolbox and procedural notes for characterizing novel zinc finger nucleases for genome editing in plant cells. Plant J. 57:747-57.
6. Zeevi, V.*, **Tovkach, A.*** and Tzfira, T. (2008) Increasing cloning possibilities using artificial zinc finger nucleases. PNAS 105: 12785-12790.
   *Authors contributed equally to this work
7. Citovsky, V., Kozlovsky, S.V., Lacroix, B., Zaltsman, A., Dafny-Yelin, M., Vyas, S., **Tovkach, A.** and Tzfira T. (2007). Biological systems of the host cell involved in *Agrobacterium* infection. Cell. Microbiol. 9: 9-20

Chapter 1. Introduction

1.1 Reverse genetics—methods for studying functions of unknown genes

The science of genetics has traditionally focused on determining the underlying causes of mutant phenotypes by linking mutations to specific positions on chromosomes and eventually, to alterations in the DNA sequence of the studied organism (1,2). However, with the advancement of DNA sequencing projects and increasing interest in interactions among multiple genes affecting a biological process, a new task has emerged: working backwards from a predicted gene to the function or phenotype that it might confer. While the first route—linking observed phenotype with a change in DNA—holds the title of classic (also forward) genetics, the second route is referred to as reverse genetics (1,2). Plant reverse genetics is an emerging science that aims at increasing our knowledge about plant metabolic processes and at improving food security for a growing world population by introducing novel traits into domesticated crops. While in some model organisms (such as yeast and mice), reverse-genetics methods have reached a significant degree of efficiency and sophistication, adaptation of those methods or their alternatives in plants lags behind (3). Below, I discuss some of the methods of reverse genetics, specifically homologous recombination (HR), homology-directed gene silencing (HDGS, also known as RNA interference, or RNAi) indexed mutant libraries **(Figure 1.1)** and targeted genome modification with rare-cutting restriction endonucleases, with a special emphasis on their application in plant species.

1.1.1 Homologous recombination

In the context of reverse genetics (also known as gene targeting), HR refers to an exchange between a chromosomal copy of a gene and a donor DNA that shares some

sequence identity with the target gene. In addition to homologous regions, artificially introduced DNA carries a modified segment that leads to disruption of the target gene sequence, and a selection marker that helps with the identification of gene-targeting events (**Figure 1.1.A**).

The first eukaryotic model system to reap full advantage from HR-based gene-targeting methods was the budding yeast *Saccharomyces cerevisiae.* After the original demonstration in the late 1970s that yeast chromosomal segments can be efficiently replaced with altered sequences by HR (4), a variety of gene-targeting techniques were developed, such as one-step gene disruption, pop-in/pop-out gene replacement and allele recovery (described in detail by R. Rothstein (5)). The relative simplicity of these techniques combined with rapid growth, ease of transformation and a well-characterized genome, made *S. cerevisiae* a model organism of choice for studying genetic processes in eukaryotes.

Mouse is another model system that routinely utilizes HR to knock out genes of interest. Mouse embryonic stem (ES) cells, which are pluripotent cells derived from preimplantation mouse embryos, are amenable to HR-based genetic manipulation techniques, creating a range of mutations—from single nucleotide replacements to large deletions and tissue-specific deactivations of studied genes (6). Modified ES cells are injected into a developing blastocyst, which gives rise to a chimera—a mouse derived from engineered and wild-type cells. Eventually, a cross between two chimeric parents produces a line of knockout mice with targeted genome modification. As the only mammalian model amenable to gene targeting, knockout mice are an essential

experimental tool for functional analysis of mammalian genes and for studying the mechanisms of human diseases.

Although routine in yeast and mice, gene targeting by HR in higher plants presents a significant challenge, mainly because of the strong dominance of the nonhomologous end-joining (NHEJ) mechanism for foreign DNA integration into plant cells (3). Several published attempts at plant gene targeting demonstrated that during a transformation event, the transgenes, usually delivered in the form of *Agrobacterium* T-DNA, enter the NHEJ pathway and integrate in random locations throughout the genome, and thus vigorous screening procedures are required to identify the rare (10^{-3}-10^{-6}) cases of precise homology-based integration (7-10).

A typical attempt to achieve HR-based gene targeting was demonstrated in a paper published by Terada and coworkers, who managed to knock out the *Waxy* gene in rice by using strong positive-negative selection (9). The authors created a targeting vector containing the hygromycin-resistance gene flanked by the regions of homology with *Waxy*, which, in turn, were flanked by two diphtheria toxin A (DT-A) fragments, and then used *Agrobacterium* to transform their vector into rapidly dividing rice callus tissue. The DT-A genes, placed in the target construct outside of the region of homology, allowed the elimination of randomly inserted clones. For every 1500 transformants resistant to hygromycin, Terada and coworkers obtained about 100 diphtheria toxin survivors and only a single correct integration event was confirmed by PCR. Such low frequency, characteristic of this and other (8,10) reports on plant gene targeting, is most likely the reason why the reported methods could only be applied to select genes and did not evolve into universal tools for plant reverse genetics.

The efficiency and precision of gene targeting by HR makes it the technology of choice for reverse-genetics research in select model organisms. However, the limited success with this technology in higher plants underscores the need for other methods of plant genome editing.

1.1.2 Homology-directed gene silencing

HDGS is the process of downregulating gene expression at the transcriptional (transcriptional gene silencing, TGS) and post-translational (post-translational gene silencing, PTGS; also known as RNA interference, or RNAi) levels (3). It is induced by the presence of double-stranded (ds) RNA homologous to a targeted gene and leads to the formation of heterochromatin structures (TGS) or the degradation of mRNA (PTGS), thus abolishing a gene's function without permanently modifying its sequence in the genome.

Fueled by the high demand for information on gene function, and by the technique's universal nature and relative ease of application, over the last decade HDGS has become the method of choice for reverse-genetics experimentation in a wide range of model organisms. In the nematode *Caenorhabditis elegans*, which allows for HDGS experimentation by simply soaking the worms in a solution containing dsRNAs or by feeding them with *Escherichia coli* organisms that express the dsRNAs, nearly 90% of the computer-predicted genes on chromosome I have been targeted by RNAi (11). In mice, the introduction of constructs encoding small hairpin RNAs (shRNAs) in combination with the Cre-loxP system has yielded efficient tissue-specific and conditional knockdowns of endogenous genes, in a fraction of the time required to achieve similar results by HR (12).

An efficient way of inducing HDGS events in plants, which in fact forms the foundation for many popular and widely used protocols, involves expression of long intron-separated hairpin RNAs that fold back on themselves to create a double-stranded region (13). More recently, a new RNAi-based plant gene-targeting strategy was developed by Schwab and coworkers (14). It utilizes artificial micro RNAs (amiRNAs), which are much more specific than the small interfering RNAs (siRNAs) created from hairpins, and which can efficiently silence both single and multiple target genes in a predictable manner. Currently, several large-scale studies employing both intron-separated hairpin and amiRNA approaches are underway to determine plant gene function [AGRIKOLA (www.agrikola.org), ChromDB (ChromDB.org), amiRNA Central (http://2010.cshl.edu/scripts/main2.pl)], the largest of them being the AGRIKOLA project funded by the European Union and aimed at producing resources for targeting up to 25,000 *Arabidopsis* genes by RNAi (15).

The major drawbacks of HDGS as a method for studying gene function are incomplete downregulation and limited inheritance of gene-silencing effects (16). Indeed, in many cases, reducing the level of a gene's expression is not sufficient to abolish its function, which leads to ambiguities in the interpretation of experimental outcomes. At the same time, a transgene introduced into a genome for the purpose of transcription of hairpin RNAs may itself become subject to transcriptional gene silencing, abolishing its functionality and restoring its target's expression. Thus, HDGS methods, despite their contribution to genetic studies, cannot replace the need for efficient genome-modification techniques.

1.1.3 Sequence-indexed mutant libraries

The sequence-indexed mutant library is another reverse-genetics method which revealed its full potential in the model plant *Arabidopsis thaliana* through a concentrated effort aimed at sequencing flanking genomic DNA (flanking sequence tags, or FSTs) in a large collection of random T-DNA insertion lines (17). Each of these lines originates from random integration of *Agrobacterium* T-DNA in the plant genome, which can either happen in an intergenic region or disrupt the sequence of a known gene. The data on currently available insertion lines can be accessed via web interface (http://signal.salk.edu) and the seeds have been made available to the *Arabidopsis* research community through the Arabidopsis Biological Resource Center (ABRC) at Ohio State University. As of February 2009, the ABRC library contained insertion lines for 29,000 out of a total of 32,000 *Arabidopsis* genes, thus providing researchers ready access to mutants and allowing them to test their hypotheses about gene function. The drawbacks of the indexed mutant library lie in the random nature of the mutagenesis process, which in the case of *Arabidopsis* leads to a lack of hits with small gene targets and, most importantly, in the prohibitive cost of repeating such global-scale experiments in other species.

1.1.4 Targeted genome modification with rare-cutting restriction endonucleases

Yet another reverse-genetics approach, which until recently was only hypothetical, involves application of rare-cutting restriction endonucleases (also known as homing endonucleases). Homing endonucleases are a group of endonucleases with long recognition sites encoded by mobile introns of lower eukaryotes, bacteria and bacteriophages (18). By creating a double-strand DNA break (DSB), homing

endonucleases initiate a process of intron copying from an intron-containing allele of a gene to an intronless allele of the same gene, thus promoting the spread of mobile introns within a population. The key feature of homing endonucleases which makes them attractive for genome manipulation is that their recognition sites (18 bp and up) are long enough to be unique in large genomes. In the first study of the effects of rare-cutter expression on an artificial chromosomal target, conducted by Rouet and coworkers in a mouse 3T3 embryonic fibroblast cell line, induction of DSBs by I-SceI led to NHEJ-mediated small deletions in the vicinity of the breaks and to a two orders of magnitude increase in the frequency of recombination with a transgenic homologous fragment (HR-directed gene targeting) (19). From the perspective of reverse genetics, both types of events, i.e. small deletions produced by NHEJ and gene replacement by HR, would be highly desirable if created in a gene of interest.

In addition to increased frequency of HR (20) and small deletions by NHEJ (21), induction of DSBs by rare-cutters in plants leads to the capture filler sequences of various lengths (5-800 bp) originating in various regions of the genome (21). Moreover, *Agrobacterium* T-DNA, used as a vehicle for plant transgene delivery, can also integrate into rare-cutter-induced DSBs (21), in some cases with high precision (22,23), suggesting that insertion at break points by NHEJ is a normal mode of transgene entry into the plant genome. Thus, T-DNA introduction into targeted DSBs represents another possible avenue for plant gene targeting.

Unfortunately, when these first studies with rare-cutters were being performed, their potential as genome-editing tools could not be realized because their sites had to be introduced into the genome prior to the gene-targeting experiment. In addition, it was

nearly impossible to re-engineer them into native target sequences (24). However, recent advances in zinc finger nuclease (ZFN) technology, described below, present an elegant solution to the problem of targeted introduction of DSBs and subsequent genome modification by both HR and NHEJ routes.

1.2 Zinc finger nucleases

ZFNs are custom-made DNA-cutting enzymes built by adding the endonuclease domain of type II restriction endonuclease *Fok*I to a designer zinc finger protein (ZFP) (**Figure 1.2**) (25,26). The marvel of ZFN technology lies in the individual modules of the ZFPs—zinc fingers (ZFs), which are small, about 30 amino-acid-long protein domains capable of recognizing specific nucleotide triplets by creating hydrogen bonds in the major groove of the DNA. Recent work in several labs leading to the identification of ZFs with high affinity to most of the GNN, ANN, CNN, and some of the TNN triplets (27-30), created the foundation for the modular assembly of ZFs into ZFPs that can be targeted to virtually any genomic location. The second element of ZFNs, the *Fok*I endonuclease domain, is a nonspecific endonuclease performing the function of DNA cleavage after it has been delivered to a ZFN-binding site by the ZFP. Like most type II restriction nucleases, *Fok*I acts as a homodimer, and thus requires the positioning of two ZFNs on adjacent recognition sites for the cleavage to take place (**Figure 1.2**). In the most commonly used configuration, the ZFP domain of a ZFN is comprised of three or four ZFs, and thus recognizes a 9- to 12-bp target site. The *Fok*I domain's requirement for dimerization increases the size of the target site to 18 to 24 bp for a given pair of ZFNs, which makes ZFN cutting locations unique, even for genomes as large as the 3 billion bp human genome. The simplicity of their design and the ability of novel ZFNs to recognize

and cleave preselected DNA target sequences provide the prerequisite for their application in a variety of in-vitro and in-vivo DNA-engineering tasks.

1.2.1 Design of new ZFNs from modular units

Widespread adoption of a new technology depends on its simplicity. One of the attractive features of ZFNs is that their DNA-recognition domains (ZFPs) can be assembled from modular units by conventional molecular cloning techniques.

ZFN design begins with the identification of a target site and selection of recognition helices. The two competing forces determining the location of an ideal target site are the length of the DNA stretch suitable for the gene-targeting experiment (for example, the first 100 bp of an open reading frame) (**Figure 1.3**) and the availability on that stretch of consecutive triplets for which high-specificity recognition helices have been reported in the literature. In **Figure 1.3.B**, I show specificity graphs for three recognition helices, RRD-E-LNV, TSG-N-LVR and QSG-D-LRR, recognizing triplets ATG, GAT and GCA, correspondingly (27,29). While the second recognition helix (TSG-N-LVR) is highly specific for its triplet, the first one could seemingly well bind to AAG, ACG and AGG, thus reducing the specificity of the ZFN in which it is incorporated. Once recognition helices are identified, they are integrated into a ZFP backbone structure (31), first on the computer and then in vitro by either sequential replacement of ZF modules in an existing ZFP sequence (32) or de-novo assembly (33).

Currently, there are two web-based software packages that simplify the process of target-site selection and design of ZFPs—Zinc Finger Targeter, or ZiFiT (http://bindr.gdcb.iastate.edu/ZiFiT/) (34), and Zinc Finger Tools (www.zincfingertools.org) (35), both allowing a search for ZFN target sites within a

provided DNA sequence. ZiFiT has a broader selection of recognition helices, from both the public domain and Sangamo BioSciences Inc. and ToolGen Inc. collections, while Zinc Finger Tools contains bar graphs highlighting the specificity of each recognition helix (such as those shown in **Figure 1.3 B**), thus enabling the selection of target sites for which high-specificity ZFNs can be designed. In addition, Zinc Finger Tools provides more flexibility of the spacer length between two half-sites, creating more choice in the target-site search, and a convenient program for incorporating recognition helices into a ZFP sequence, which could prevent confusion arising from the aforementioned ZFN orientation on its target site.

The drawback of modular assembly is its periodical failure to create functional ZFP domains (36). Thus, prior to committing to time- and resource-consuming gene-targeting experiments, one should test the ZFN's activity in vitro; this can be accomplished, for example, by expressing it in *E. coli* and performing digestion of a plasmid with the introduced ZFN target site with a bacterial lysate or purified protein.

1.2.2 Molecular cloning potential of ZFNs

An indication that ZFNs could indeed be used as molecular cloning tools was presented in a report by Mani and coworkers (33) who, for the purpose of assaying ZFN activity, linearized plasmids carrying ZFN target sites with common type II restriction endonucleases (*Aat*II, *Ssp*I and *Xmn*I) and then cleaved them with ZFNs expressed in rabbit reticulocyte lysates. Although the topic of using ZFN-cut plasmid fragments for molecular cloning was not mentioned in the publication, one could easily imagine the wealth of new opportunities that ZFN technology would bring to recombinant DNA research.

First, application of ZFNs for molecular cloning would significantly expand the number of available unique enzyme specificities. The latest issue of Rebase (37) lists 3698 type II restriction enzymes, of which only 262 distinct specificities are available commercially. In comparison, a three-finger ZFN comes in ca. 260,000 (64^3, where 64 is the number of nucleotide triplets) different combinations and a four-finger ZFN—in 16.7 million (64^4) different combinations. In addition, most of the commercially available restriction enzymes recognize short, 4- to 6-bp target sites that have a high probability of occurring on long stretches of DNA, making them unusable for cloning large DNA fragments. The 18- to 24-bp target sites of ZFNs place them in the unique category of rare-cutters, particularly useful for the assembly of large plasmids and genome mapping. Finally, a pair of ZFNs, designed to recognize and cleave a native DNA sequence, could be used to achieve the previously unthinkable task of cloning native DNA molecules at user-defined positions.

A unique challenge in modern plant genetics and biotechnology, which ZFNs could resolve as novel restriction enzymes, is mounting multiple expression cassettes on plant multi-gene transformation vectors (38). Currently, simultaneous plant transformation with several transgenes is best achieved by assembling them on a binary *Agrobacterium* Ti (tumor-inducing) plasmid-based vector and subsequently delivering this multi-gene construct into the plant genome by *Agrobacterium* transformation (39,40). One such system consists of a pRCS binary plasmid with a multiple cloning site (MCS) containing recognition sites for several rare-cutting restriction endonucleases and a series of satellite plasmids (pSATs) that carry gene-expression cassettes surrounded by a pair of rare-cutter sites (41,42). As the pRCS plasmid grows in size in the process of assembly,

the use of rare-cutters whose recognition sites have a very low probability of occurrence within DNA sequences ensures that it will be cut in only one position (MCS) during each subsequent cassette-mounting step. As of late, the capacity of the pRCS-pSAT system has been constrained to seven transgenes by the limited commercial availability of rare-cutting endonucleases. ZFNs, which fall into the category of rare-cutters due to the large (18-24 bp) size of their target sites, could potentially be used to remove this constraint. To this end, one could expand the pRCS-pSAT system by introducing novel ZFN target sites into the MCS region of pRCS, by constructing additional pSAT plasmids with ZFN sites and by expressing and purifying enzymes for ZFN-mediated cloning. Construction of multi-gene transformation vectors with ZFNs would add an additional degree of freedom to our ability to study plant gene function and to create novel transgenic crops with artificial metabolic pathways.

1.2.3 Application of ZFNs for genome editing

As already mentioned, induction of DSBs in eukaryotic genomes by rare-cutting restriction endonucleases leads to increased frequency of HR in the vicinity of the break and/or to small deletions after break repair by NHEJ (19-23). Therefore, upon the first demonstration that ZFN injection can increase the frequency of HR in a self-recombining reporter plasmid in *Xenopus* oocytes (43), several groups initiated studies directed at revealing the potential of ZFNs as genome-editing tools in a wide range of model organisms.

For example, in *Drosophila,* a pair of designed ZFNs was used to create genomic DSBs in the *yellow* (*y*) gene involved in color development in the fly's posterior abdomen. Induction of ZFN expression by activation of a heat-shock promoter yielded

simple deletions, deletions-insertions and duplications at the *y*ZFN target site, leading to the formation of visible patches of *y* mutant tissue (somatic mosaics) in 46% of the flies (44). In the nematode *C. elegans*, a well-characterized ZFN QQR and a pair of specifically designed nucleases were used to induce mutations in the nematode somatic cells, again achieving 20% mutation frequency and observing similar deletions-insertions following NHEJ-mediated repair of DSBs created by ZFNs (45). In zebrafish, where gene targeting was previously intractable, three independent publications have emerged in the last half a year demonstrating ZFN-mediated knockouts of native genes (46-48).

The work by a research group at Sangamo Biosciences deserves special attention. They created the foundation for the application of ZFNs in human gene therapy by repairing an X-linked severe combined immune deficiency (SCID) mutation in the *IL2Rγ* gene in human K562 cells. A pair of highly specific four-ZF ZFNs was used to induce HR at the mutated *IL2Rγ* site with an extrachromosomal DNA donor, achieving a modification rate as high as 5% in the absence of any selection (49).

Thus, induction of genomic DSBs with ZFNs can indeed be used to achieve both NHEJ- and HR-based gene targeting in a variety of eukaryotic species.

1.2.4 ZFN-mediated gene targeting in plants

Successful application of ZFNs in *Xenopus* (43) and *Drosophila* (44,50) prompted attempts at ZFN-based gene targeting in plants. Two studies, by Lloyd and coworkers and by Wright and coworkers, used well-characterized ZFNs to induce gene-targeting events in artificially introduced ZFN sites by NHEJ- (51) and HR-based (52) mechanisms. In the first study, authors created an *Arabidopsis* line carrying a T-DNA insertion with a *QQR ZFN* gene and a QQR target site. Induction of QQR expression by a

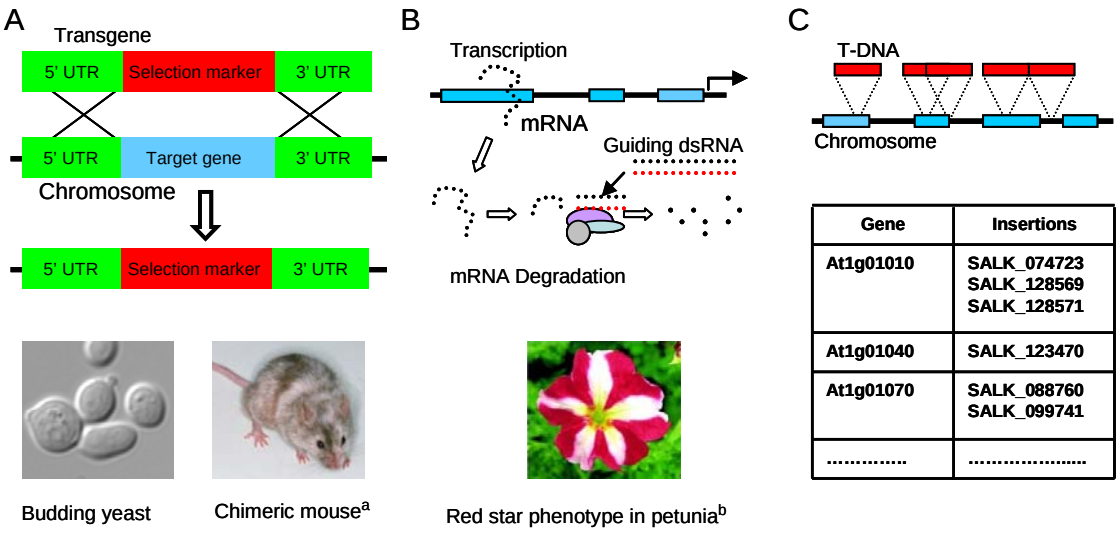
heat-shock promoter led to small deletions and insertions at the QQR target site characteristic of NHEJ-mediated DSB repair with a frequency of up to 20% (51). In the second study, authors introduced a defective GUS:NPTII reporter gene in tobacco and used electroporation technique to deliver a Zif268-ZFN-expressing plasmid and a GUS:NPTII repair template into plant protoplasts. Induction of DSBs at a Zif268 site within the defective reporter by Zif268-ZFN led to reporter repair by HR with a frequency of 10% (52). In both cases, the frequencies of gene-targeting events represented a staggering improvement over previously reported frequencies of 10^{-7} to 10^{-4}.

The above-described studies were a great demonstration of ZFN efficiency, yet they represented only the proof-of-principle stage on the way to targeting native plant genes. To this end, Maeder and coworkers reported efficient mutagenesis of the tobacco SurA locus with a pair of ZFNs constructed by a novel strategy called oligomerized pool engineering (OPEN) (53). In addition, research groups from Dow Agrosciences and Sangamo Biosciences recently reported homology-directed transgene integration in a tobacco endochitinase gene after induction of a DSB with ZFNs (54).

Illustrated advances in the targeting of both artificial and native plant genes indicate that ZFNs have great potential as gene-targeting instruments and that they could present a viable solution to the long-standing problem of plant genome engineering. At the same time, unexpectedly high failure rates of ZFNs assembled from modular units (36) and their often low gene-targeting efficiency, as evidenced from assays in human cells (53), prompted further research on efficient methods of ZFN application for plant gene targeting. In addition, although the cloning potential of ZFNs is theoretically evident

from the nature of their DNA recognition and cleaving properties, experiments need to be carried out for biochemical ZFN characterization, elucidation of their cloning properties and establishment of efficient protocols for ZFN-mediated cloning. Application of ZFNs as molecular cloning tools could generate advances in plant genome engineering and biotechnology by expanding the cloning capacity of multi-gene plant genetic transformation vectors (38).

1.3 Figures



Gene	Insertions
At1g01010	SALK_074723 SALK_128569 SALK_128571
At1g01040	SALK_123470
At1g01070	SALK_088760 SALK_099741
.....	.....

Figure 1.1. Methods of reverse genetics. A. Homologous recombination. B. RNA interference. C. Sequence-indexed insertion library. [a]Reproduced from Tarkowski et al. (55). [b]Reproduced from Koseki et al. (56).

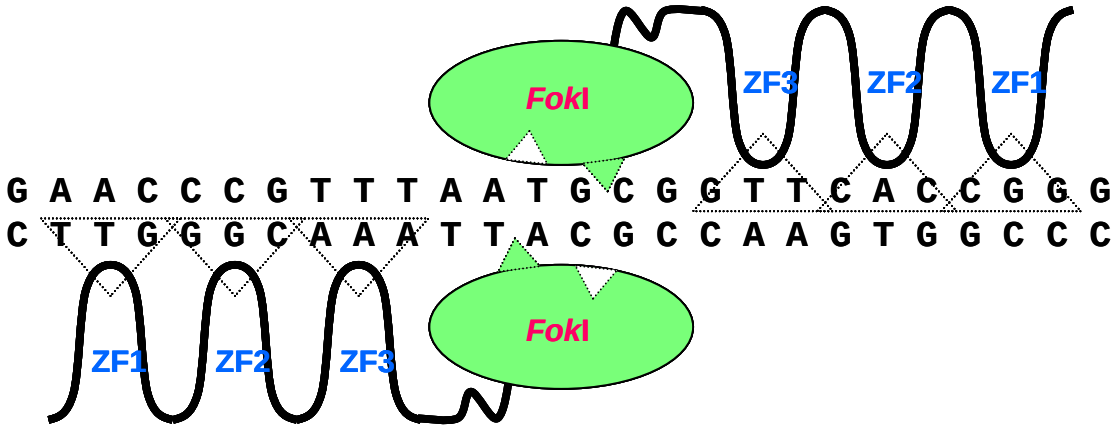


Figure 1.2. Structure and DNA-binding properties of zinc finger nucleases. Zinc finger nucleases are composed of a zinc finger domain, responsible for DNA recognition, and a *Fok*I endonuclease domain conducting DNA cleavage. Because of the *Fok*I domain's requirement for dimerization, two pairs of ZFNs have to be placed on adjacent sequences to ensure efficient DNA cutting.

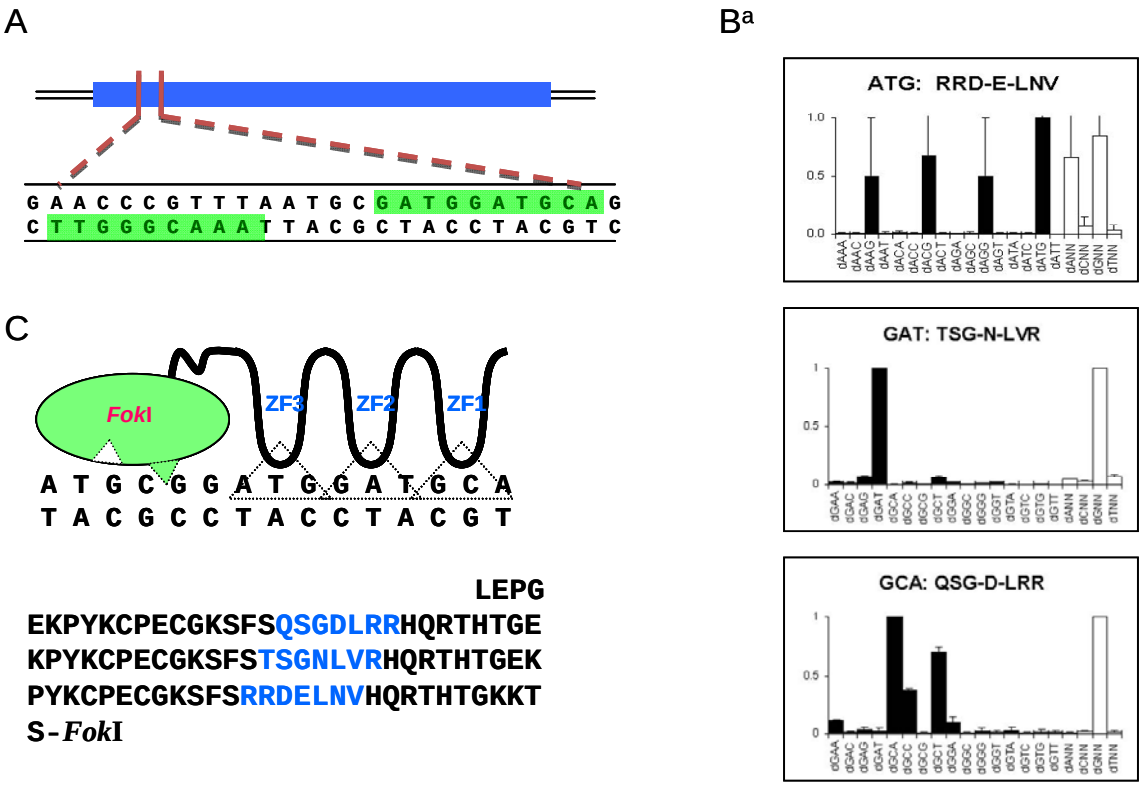


Figure 1.3. Design of ZFNs from modular units. A. Target-site selection. B. Examples of recognition helices with different degrees of specificity for their target triplets. C. Sequence of zinc finger protein domain with incorporated recognition helices.
[a]Reproduced from Dreier et al. (27,29).

Chapter 2. Expression, purification and characterization of cloning-grade zinc finger nuclease

2.1 Summary

The limited number of naturally occurring rare-cutting restriction enzymes and the slow and tedious engineering of existing restriction enzymes for novel specificities have prompted the design of new strategies for the development of restriction enzymes with specificities for long DNA sequences. One possibility is using zinc finger nucleases (ZFNs)—synthetic restriction enzymes that are custom-designed to target and cleave long DNA sequences and which have been recently shown useful for DNA cloning. Here we report on the purification and biochemical analysis of ZFN-10, a custom-made ZFN. We show that Ni-affinity and gel-filtration purification methods are sufficient to produce a cloning-grade enzyme. We show that ZFN-10 can function as an accurate and reliable ZFN using the same reagents and protocols used for naturally occurring and commercially available recombinant restriction enzymes. We also show that ZFN-10 tolerates a set of target-site substitutions which can be predicted from the specificities of recognition helices incorporated into the structure of its DNA-binding domain. The relative simplicity of ZFN-10 design, expression, purification and analysis suggests that novel ZFNs can potentially be designed and applied for various recombinant DNA applications.

2.2 Introduction

Restriction enzymes are essential and integral components of modern molecular biology research. Since their discovery in 1971, they have been invaluable for the physical mapping of DNA and have become the foundation for molecular cloning and recombinant DNA technology as a whole (57). A large number of restriction enzymes, most of which are capable of digesting short (4 to 6 bp) palindromic sequences, have been isolated from a variety of organisms. Enzymes capable of digesting non-palindromic sequences of 8 bp, 10 bp and even longer have also been identified (37,58). The restriction enzyme database REBASE (containing enzymes and genes for DNA restriction and modification) provides a list of over 3600 restriction enzymes which have been genetically and/or biochemically characterized (37). About 600 of these enzymes have been purified to the level of commercial use, targeting a relatively narrow collection of about 260 distinct 4- to 6-bp sequences and about a dozen 8- and 10-bp sequences (37). The number of enzymes capable of targeting very large (i.e. above 12 bp) sequences is extremely low: to the best of our knowledge, only six enzymes are commercially available. The latter, also known as rare-cutting restriction enzymes, are of particular interest: they are useful for the cloning of large DNA fragments, they can potentially be used for genome mapping, and they can facilitate the molecular assembly of complex DNA structures (e.g. Refs. 38,41,42,59). Rare cutters are also instrumental for genome engineering, as their expression in living cells can produce genomic double-strand breaks, leading to site-specific mutagenesis and/or increased homologous recombination in various organisms (21,60-63). Realizing the importance of rare cutters for genome engineering and as novel enzymes for molecular cloning, several groups are currently

engaged in attempts to re-engineer existing enzymes for novel specificities (e.g. Refs. 64,65,66). Such studies have led to the development of new rare cutters, but we have yet to witness the use of these recently reported variants and novel enzymes for cloning or other purposes (e.g. Refs. 64,65,66).

Zinc finger nucleases (ZFNs) are artificial restriction enzymes (26,28,33) which may provide a viable alternative to the tedious and complex re-engineering of existing rare cutters, since they can be engineered to target long DNA sequences. ZFNs are composed of a custom-designed DNA-binding domain which is fused to the *Fok*I non-specific DNA-cleavage domain. The DNA-binding domain can be composed of several fingers. Each finger can be engineered to recognize and bind a specific DNA template, and a typical ZFN monomer consists of three to four fingers, capable of recognizing and binding to a 9- to 12-bp-long target DNA sequence. When two ZFN monomers are aligned in reverse orientations, they form an active enzyme which is capable of digesting its target sequence. The length of a non-sequence-specific spacer located between the ZFN monomer binding sites is rather flexible (67,68), and while a 6-bp spacer is typical in the engineering of novel ZFNs, shorter sequences have also been shown to be digested by ZFNs (67,68). This phenomenon allows for great flexibility during the design and construction of novel ZFNs: a typical ZFN dimer which is composed of two monomers capable of binding a 9-bp DNA sequence each, can be used to target a 20- to 24-bp-long DNA sequence. ZFNs can thus be classified as rare cutters, and can potentially be used in a variety of molecular biology applications.

We recently demonstrated that ZFNs can be used for a variety of molecular cloning tasks (69). In our experiments, ZFNs accurately and specifically digested target

DNA molecules, which were subsequently ligated and recovered using common recombinant DNA protocols. We used 24-bp-long ZFN recognition sites for the assembly of a transformation vector composed of two independent expression cassettes (a task that is typically achieved using site-specific recombinases or rare cutters (38,70)) and for the cloning of a native DNA sequence. The latter was achieved using a pair of ZFNs which were specifically designed to target and digest the target DNA sequence (69). While proven useful for DNA cloning, tedious and costly in-vitro expression methods may hinder the use of ZFNs for DNA cloning. Furthermore, the presence of contaminants in *E. coli*-expressed ZFN lysates may lead to non-specific activity and low digestion efficiency, which may result in a low-quality digested product. Consequently, we ascertained that obtaining well-characterized enzyme preparations can prevent non-specific cleavage and produce more robust outcomes for ZFN-mediated cloning protocols.

The biochemical properties of hybrid restriction enzymes have been characterized by Kim et al. (71), who constructed, expressed and biochemically characterized two zinc finger protein (ZFP)-*Fok*I fusions (CP-QDR and Spl-QNR). Those authors showed that the purified enzymes were capable of digesting a λ DNA substrate into the predicted size fragments, but they also reported that both enzymes cut their DNA substrates at multiple positions. Although that study laid the foundation for the production of recombinant artificial restriction enzymes, the authors indicated that the described enzymes may still not be ready for routine laboratory use (71). Other studies have reported the construction, purification and substrate-specificity analysis of additional enzymes (e.g. Refs. 72,73,74) and many recent reports describe the construction and use of ZFNs for gene-targeting

experiments in living cells (e.g. Refs. 49,51,75-77). However, no data have been provided showing the integrity of the digested products, an important factor in the use of ZFNs for cloning purposes.

Here we describe the expression, purification and analysis of the DNA-cleavage properties of ZFN-10, a custom-made ZFN, and its use for DNA cloning. We demonstrate that Ni-affinity and gel filtration purifications are sufficient to produce a cloning-grade enzyme, as determined by ligation efficiency and sequencing analysis of the ligated products. We also show that ZFN-10 tolerates a set of target-site substitutions which can be predicted from the specificities of recognition helices incorporated into the structure of its ZFP domain.

2.3 Experimental procedures

2.3.1 DNA constructs

The ZFP domain of ZFN-10, with zinc finger moieties QSG-D-LRR (GCA) (29), TSG-N-LVR (GAT) (29) and RRD-E-LNV (ATG) (27) capable of binding the target site ATGGATGCA, was assembled from overlapping oligonucleotides as previously described (33,69,78) and cloned between *Xho*I and *Spe*I sites upstream of the *Fok*I ED in a pSAT6-*Fok*I vector (69,78) to yield pSAT6-ZFN-10. The assembled ZFN-10 was transferred as an *Nco*I-*Bam*HI fragment into the same sites of a modified pET28 expression vector, pET28-XH (78), to yield pET28XH-ZFN-10, which contains a ZFN-10-6xHis tag fusion ORF under control of a T7 bacterial expression promoter. Construction of pSAT10-MCS was as described in Zeevi et al. (69). To construct the set of pSAT6-TS-(n) DNA substrates, we first designed a collection of primer pairs (upper and lower primers) encoding the various modified ZFN-10 target sites. Primers in each

pair were capable of annealing to each other and were designed to carry *Nco*I and *Pst*I sites. The upper and lower primers' general structure was 5'CGATAGCCATGG$(n)_9$GTAAGT$(n)_9$CTGCAGTCGACG and 5'CGTCGACTGCAG$(n)_9$ACTTAC$(n)_9$CCATGGCTATCG, respectively, where $(n)_9$ was replaced by the modified ZFN-10 recognition site shown in **Figure 2.5**. (Thus, for example, 5'CGATAGCCATGGTGCATCCCTGTAAGTAGGGATGCACTGCAGTCGACG and 5' CGTCGACTGCAGTGCATCCCTACTTACAGGGATGCACCATGGCTATCG were annealed to produce pSAT6-TS-3.) Annealed primers were digested by *Nco*I and *Pst*I and were cloned into the same sites of pSAT6-MCS (41), producing the collection of pSAT6-TS-(n) target plasmids (designated pSAT6-TS-0 through pSAT6-TS-14).

2.3.2 Protein expression and purification

pET28SX-ZFN-10 was transformed into BL21 (Gold) DE3 PLysS competent cells (Stratagene). Cell culture (1 L) was grown in LB medium supplemented with kanamycin and 100 μM $ZnCl_2$ at 22°C to an OD_{600} of 0.6 and induced with 0.7 mM IPTG for 4 h. The cells were harvested by centrifugation, resuspended in 35 ml resuspension buffer (25 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% v/v glycerol and 100 μM $ZnCl_2$) and lysed twice through a French press. For His-tag protein purification, 1 ml of Ni-NTA resin (Qiagen) was added to the lysate which was then mixed on a rotary shaker for 1 h at 4°C. Resin was separated from the lysate by centrifugation for 2 min at 500*g*, washed in a wash buffer (resuspension buffer supplemented with 20 mM imidazole) by two rounds of resusupension and centrifugation (2 min at 500*g*) and loaded on a Bio-Rad Poly-Prep disposable chromatography column. The protein was eluted from the Ni-NTA resin with

3 ml of elution buffer (resuspension buffer supplemented with 500 mM imidazole). For size-exclusion chromatography, 2.5 ml of Ni-NTA-resin-purified ZFN was loaded on a HiPrep200 HS column (GE Healthcare, controlled by AKTA HPLC system) which had been pre-equilibrated with the resuspension buffer. Elution was conducted at room temperature. The column eluate was monitored for UV absorbance at 280 nm and collected in 2-ml fractions. The fractions corresponding to a UV absorbance peak, at elution volumes of 53 to 59 ml, were analyzed for ZFN activity and pooled. Total protein content for crude extract and purified proteins was determined by Bio-Rad Protein Assay.

2.3.3 Digestion and ligation analysis

Digestion of 100 ng plasmid DNA (pSAT10-MCS or pSAT6-TS) was carried out at various temperatures in NEBuffer 4 (New England Biolabs) in a total reaction volume of 25 µl. For digestion analysis, the products of the digestion reaction were separated by gel electrophoresis and their amounts were quantified by ImageJ software (79). For ligation analysis, the 2.6-kb backbone DNA fragment from pSAT10-MCS was first purified using Illustra GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare). About 0.5 ng of purified fragment was self-ligated with 2 units of T4 ligase (New England Biolabs) in a total volume of 10 µl for 2 h at room temperature and the total ligation reaction was transferred into chemically competent DH5α *E. coli* cells using standard molecular biology protocols. Recombinant colonies were counted, and randomly selected colonies were analyzed for the integrity of their ZFN-10 target site by ZFN-10 digestion and DNA sequencing.

2.4 Results and Discussion

2.4.1 Reagents for ZFN expression and analysis

We have previously shown that ZFNs can be used for DNA cloning. We demonstrated the applicability of custom-made ZFNs as a tool for the assembly of complex DNA structures and for the cloning of native DNA sequences (69). We found that while crude lysate protein extracts from ZFN-expressing *E. coli* cells are useful for *in-vitro* digestion analysis of ZFN activity, the presence of contaminating nucleases (and possibly other DNA-damaging enzymes) in the extracts leads to low quality of the digested DNA products and low ligation efficiency. We thus reasoned that high-grade purification and further biochemical characterization of custom-made ZFNs may yield reliable enzymes useful for recombinant DNA applications.

We chose the ZFN-10 monomer, which we had previously assembled and used for molecular cloning (69), as a model enzyme for our biochemical purification and characterization study. The ZFP domain of the ZFN-10 monomer was designed to recognize DNA sequence ATGGATGCA and was assembled from overlapping oligonucleotides designed to encode the recognition helices of each finger (**Figure 2.1A**), as previously described (33). The actual artificial ZFN-10 recognition site (hereafter referred to as ZFN-10-TS) is composed of two ZFN-10 monomer recognition sites placed in antiparallel orientation and separated by a 6-bp long spacer (**Figure 2.1A**). Digestion of the 6-bp spacer is expected to occur between the second and third bases, relative to the ZFN-10 binding site, on both the top and bottom strands (**Figure 2.1A**); however, other digestion patterns may occur. For expression in *E. coli*, the ZFN-10 coding sequence was subcloned in-frame to a C-terminal 6x-His tag between the T7 promoter and terminator in

pET28XH (78), a modified pET28 vector, producing pET28XH-ZFN-10 (**Figure 2.1B**). As a substrate for the analysis of ZFN enzymatic activity, we used pSAT10-MCS (69), a plasmid that contains two ZFN-10-TSs separated by a 1.2-kb fragment (**Figure 2.1C**). The 6-bp spacer sequence within the ZFN-10-TS was GTAAGT. Digestion of pSAT10-MCS using lysate of *E. coli* expressing ZFN-10 for a period of 5 min produced the expected fragments of 1.2 and 2.7 kbp (e.g. **Figure 2.1D**). A 3.9-kbp linear DNA fragment was also observed, as a result of incomplete digestion, and was attributed to the low activity of ZFN-10 in *E. coli*'s lysates. Longer incubation times and/or higher lysate concentrations typically resulted in degradation of the DNA substrate (data not shown), most likely due to the presence of contaminating endonucleases in the crude lysate preparations. The pSAT10-MCS substrate provides a simple visual means of determining the digestion activity of ZFN-10: following its digestion, pSAT10-MCS can be self-ligated and sequenced, providing viable information on the quality and accuracy of the digestion step.

2.4.2 Purification of ZFN-10

In previous reports, hybrid restriction enzymes have been purified by various means, including Ni affinity, ion exchange and gel-filtration chromatorgraphy (71,73,74). While proven useful for characterizing sequence specificity and analyzing the digestion and binding activities of these artificial enzymes, no data were presented as to the enzymes' use for recombinant DNA applications, e.g. DNA cloning. We reasoned that analyzing the digestion characteristics of an artificial ZFN in the context of DNA cloning is key to its successful integration into existing cloning protocols. We selected the commonly used *E. coli* strain BL21 (Gold) DE3 PLysS (Stratagene) for expression of ZFN-10. Pelleted

cells were lysed in a French press and the lysate was successively purified using Ni-NTA agarose beads and gel filtration. Coomassie-stained SDS-PAGE analysis of ZFN-10 from the initial lysate and each of the two purification steps is shown in **Figure 2.2**, where it can be clearly seen that capturing the His-tagged ZFN-10 by Ni affinity and further purification in a gel-filtration column reduced much of the contaminating *E. coli* proteins. A dominant band at slightly higher than 34 kDa was observed (**Figure 2.2**), in agreement with the predicted size of ZFN-10 (34.9 kDa).

Key to the applicability of ZFN monomers for the digestion of long DNA sequences is DNA-dependent dimerization at their target site. We thus investigated whether ZFN-10 proteins remain soluble and in a monomeric state following their purification by gel filtration. We calibrated the gel-filtration column with a series of molecular weight markers and estimated the molecular weight of ZFN-10, in solution, from its elution profile (**Figure 2.3**). Peak ZFN-10 eluted at 54 ml (with Kav, a phase-distribution coefficient, of 0.21), corresponding to a molecular weight of 38.2 kDA on our calibration graph. Taking into account that the calculated size of ZFN-10 is 34.9 kDa, we concluded that ZFN-10 is a globular protein and that it maintains a monomeric state in solution in the absence of DNA with its binding sites.

We calculated the yield of purified protein following the two successive purification steps (Table 1). Total protein content was quantified by Bradford assay and the amount of total active protein was experimentally estimated by digestion of pSAT10-MCS. For the purpose of this study, we defined a unit of ZFN-10 activity as the amount of protein required for cleavage of 1 μg pSAT10-MCS DNA in 1 h at 37°C in NEBuffer 4 in a total reaction volume of 25 μl. For the quantitative activity measurements, we

digested pSAT10-MCS by a series of total protein dilutions, separated the products of the cleavage reactions on a 0.7% agarose gel, measured band intensities with ImageJ software (79), calculated the percentage of cleaved product and adjusted for amount of substrate and incubation time of the reaction. The total protein content and total active protein from a 1 L *E. coli* culture are summarized in Table 1. We calculated the specific activity and total yield of active enzyme for each purification step, and found that Ni-NTA and gel filtration resulted in the recovery of correspondingly 34% and nearly 10% of the active enzyme.

2.4.3 Cloning quality of the digested DNA

We evaluated the cleavage quality of the three ZFN-10 preparations by digesting the pSAT10-MCS plasmids, self-ligating the digested products and analyzing the recombinant DNA products. We separated digested pSAT10-MCS plasmids on a 0.7% agarose gel, purified the 2.6-kb pSAT10 backbone fragments from crude lysate, Ni-affinity-purified and gel-filtration-purified ZFN-10 digestion mixtures and self-ligated them. As shown in **Figure 2.4A**, the number of independent clones obtained from ca. 0.5 ng of pSAT10-MCS purified backbone cut with Ni-affinity- or gel-filtration-purified ZFN-10 was on average 30 times higher than that obtained from cell-lysate-cut plasmid (**Figure 2.4A**). We further compared the cloning efficiency of ZFN-10-digested plasmids to that of a commercial rare-cutting restriction enzyme, PI-*Psp*I, a homing endonuclease capable of digesting 30-bp-long sequences. While the sequence degeneracy tolerated by PI-*Psp*I enzymes is still unknown, various studies have shown its usefulness for accurate DNA cloning (e.g. Refs. 41,42,80,81). As a substrate for PI-*Psp*I, we used pSAT6-MCS (41), a plasmid that is similar in size and sequence to pSAT10-MCS, except that it carries

two PI-*Psp*I recognition sites instead of the two ZFN-10-TSs. We set up a PI-*Psp*I digestion reaction, released the 2.7-kb backbone fragment of pSAT6-MCS and self-ligated it using DNA amounts, and ligation and transformation conditions similar to those used for ZFN-10-digested plasmid DNA. The efficiency of Ni-affinity- or gel-filtration-purified ZFN-10-digested plasmid DNA was comparable to that of PI-*Psp*I (**Figure 2.4A**). To further analyze the quality of the ligated products, we randomly selected 20 plasmid colonies from each ligation reaction (i.e. with cell lysate, Ni-affinity-purified and gel-filtration-purified ZFN-10) and analyzed the integrity of the ZFN-10 ligation site by digestion and DNA sequencing. To this end, we linearized the plasmids with *Sca*I and subjected the purified product to digestion with gel-filtration-purified ZFN-10. While all of the clones obtained from pSAT10-MCS digestion by Ni-affinity- or gel-filtration-purified ZFN-10 carried a functionally reconstructed ZFN-10 recognition site, most of the clones (i.e. 18 out of 20) obtained from *E. coli*-lysate-digested pSAT10-MCS had lost their ZFN-10 site, as determined by ZFN-10 digestion analysis (exemplified in **Figure 2.4B**). Sequence analysis of 10 of the defective clones obtained from the *E. coli*-lysate-digested pSAT10-MCS revealed that part of the ZFN-10 recognition site had indeed been degraded (**Figure 2.4C**), most likely by contaminating exonucleases or other contaminants in the crude DNA extract. Sequence analysis also revealed that the ZFN-10 recognition site was correctly reconstructed in 10 clones obtained from digestion with Ni-affinity- or gel-filtration-purified ZFN-10, as well as in two of the corrected clones obtained from *E. coli*-lysate-digested pSAT10-MCS (e.g. clones NiNTA1, GF1 and Crude10, **Figure 2.4C**). Our data thus indicated that Ni-affinity or gel-filtration purification is essential for obtaining high ligation yield and for accurate digestion of

target DNA substrates. It should be noted that while Ni-affinity purification did not remove all of the contamination (**Figure 2.2**), the final product was sufficiently pure for the efficient and accurate digestion of pSAT10-MCS DNA substrate.

2.4.4 Target-site specificity of ZFN-10

We demonstrated that large-scale expression and Ni-affinity purification of the expressed proteins produce sufficient amounts of recombinant ZFN for DNA cloning. However, the ZFNs need to be further characterized to facilitate their introduction as a flexible tool for recombinant DNA applications. We have previously shown that ZFNs can be very useful in the assembly of multigene transformation vectors (69), a task that can rarely be achieved using type-II restriction enzymes and that is typically performed with only a very limited number of homing endonucleases (38). It should be noted that homing endonucleases, the only type of enzyme besides ZFNs capable of recognizing and cleaving long target sites on DNA, are known to tolerate a high degree of sequence polymorphism within those sites. Argast et al. (82), for example, identified 21 different homing target sequences for each of I-*Ppo*I and I-*Cre*I endonucleases, and Bryke et al. (83) reported 60 possible homing target sites for I-*Tev*I, and demonstrated that no individual base within the enzyme's 48-bp site is absolutely required for the interaction. We thus investigated the sequence degeneracy of ZFN-10.

The identification of alternative binding sequences for homing endonucleases typically relies on elaborate library-screening procedures (e.g. Refs. 82,83,84). Target-site degeneracy of ZFNs, on the other hand, can potentially be predicted from available data on the specificity of individual recognition helices (27,85). This feature can simplify the prediction of possible off-target site digestion by artificial ZFNs. We analyzed the

target-site degeneracy of ZFN-10 and found that the first recognition helix may be capable of binding triplets GCA, GCT and GTC (29), the second recognition helix is rather specific and is expected to recognize only a single target triplet, GAT (29), and the third recognition helix may tolerate any substitution in the second position and an A-to-G substitution in the first position (27). This information allowed us to produce a putative list of preferred sequences for the ZFN-10 monomer (**Figure 2.5A**). We next determined whether various semi-palindromic target sequences composed of putative alternative ZFN-10 monomer recognition sites can be digested by the ZFN-10 enzyme. We constructed a series of pSAT6-TS target plasmids (designated pSAT6-TS-1 through pSAT6-TS-14) in which a modified ZFN-10-TS was placed within a distance of ca 0.9 kb from a unique *Age*I site (**Figure 2.5B, C**). A pSAT6-TS-0 plasmid, carrying the original ZFN-10-TS, was also constructed. The modified target sites were designed to probe ZFN-10's tolerance for various substitutions within its recognition site (**Figure 2.5C**). We next co-digested the pSAT6-TS plasmids with *Age*I or gel-filtration-purified ZFN-10 and *Age*I , and separated the digestion products on an agarose gel (**Figure 2.5D**). The digestion efficiency of each plasmid, quantified by ImageJ software, is given as percentage of cleaved products out of the total amount of plasmid DNA (cutting efficiency; **Figure 2.5C**).

Our data showed (**Figure 2.5C, D**) that calculated predictions can indeed be useful for determining possible target-site degeneracy of ZFNs. More specifically, we found that some of the predicted single-base-pair substitutions were indeed well tolerated by the ZFN-10 enzyme, resulting in almost complete cleavage of the corresponding target-site plasmids (e.g. A-to-G substitution of the first nucleotide (pSAT6-TS-2) or T-

to-C substitution of the second nucleotide (pSAT6-TS-4) of the ZFN-10 9-bp-long recognition site, **Figure 2.5C, D**). Combining several substitutions in the ZFN-10 24-bp-long target site completely abolished ZFN-10 digestion of some of the modified target sites (e.g. pSAT6-TS-7 and 8), but had no effect on others (e.g. pSAT6-TS-9). It is interesting to note that recognition helix 2, which has been reported to have a unique preference for the GAT triplet in a multitarget-specificity assay (29), could also bind GCT and GAC triplets when constructed in the ZFN-10 structure.

Demonstrated target-site polymorphism somewhat reduces, but does not completely abolish, the rare-cutting potential of ZFN-10 (or of other ZFNs for that matter): although some of the substitutions were well-tolerated, the tolerance did not extend to each target-site position or to each of four possible nucleotides. Most importantly, the polymorphisms followed, to a great extent, the pattern predicted from the data on recognition-helix specificity, thus allowing for selection of ZFN specificity parameters during the design stage.

2.4.5 Optimal digestion temperature and heat-deactivation of ZFN-10

Optimal digestion temperature and deactivation are critical parameters for the use of restriction enzymes in recombinant DNA applications. Although the optimal temperature for *Fok*I endonuclease digestion activity is 37^{o}C (New England Biolabs), in a biochemical digestion analysis of Sp1-QNR ZFNs, Kim et al. reported an optimal activity temperature of 22^{o}C (71). They also reported significantly reduced enzyme activity at 32^{o}C, and even complete abolishment of activity at 37^{o}C (71). Using gel-filtration-purified ZFN-10 and *Age*I-linearized pSAT6-TS-0, we determined that while the enzyme is capable of digesting its substrate at 22^{o}C, the optimal digestion temperature of ZFN-10

is 37°C (**Figure 2.6**). The discrepancy between the optimal temperatures of ZFN-10 and the Sp1-QNR ZFN can be attributed to the type of assay and substrate used to determine the activity of each enzyme. Kim et al. (71) used λ phage DNA which carried a 9-bp binding site for a single ZFN monomer. The second monomer, required for *Fok*I endonuclease domain (ED) dimerization and DNA cleavage, was likely to find its way into the cleavage complex randomly through non-specific ZFP-DNA or low-energy *Fok*I ED interactions, both of which are reduced at higher temperatures. In our experimental system, digestion of the target DNA substrate site required interaction of both monomers with the semi-palindromic targeting site. The limiting factor for the cleavage efficiency was most likely *Fok*I ED, which has an optimal temperature of 37°C. The fact that the optimal digestion activity of ZFN-11, ZFN-H2a and ZFN-H2b, three other ZFNs used for recombinant DNA applications (69), was also measured at 37°C (data not shown) further supports the notion that optimal digestion activity is imposed by the *Fok*I ED and not by the artificial DNA-binding domain.

A significant decline in ZFN activity was observed at temperatures higher than 37°C (**Figure 2.6B**). We therefore tested whether the enzyme can be permanently heat-inactivated at 65°C, the inactivation temperature of *Fok*I (New England Biolabs). We heat-treated ZFN-10 for 7 min, cooled down the heated enzyme and analyzed its activity by digestion of pSAT6-TS-0. Gel electrophoresis analysis demonstrated that short exposure to 65°C is sufficient to render ZFN-10 inactive (**Figure 2.6**).

We stored crude gel-filtration-purified ZFN-10 in 50% glycerol at -80°C and we did not notice any loss of activity after six months of storage. This was evidenced not only by their ability to digest target DNA substrates, but also by the high efficiency of the

digestion reactions and the accuracy of the digestion process, as indicated by ligation and sequence analysis of the digested products (data not shown).

2.5 Conclusions

The development of ZFNs as a novel tool for genome editing in living cells (26,35,86,87) has provided a unique opportunity to harness knowledge on the design of novel ZFNs for recombinant DNA applications. Indeed, we have previously shown that similar to meganucleases (41,42,59), ZFNs can be used to overcome the statistical limitations of many of the 6-bp cutters in assembling multigene transformation vectors (69) and can potentially assist in expanding the cloning possibilities of complex multigene transformation vectors which are currently limited by the small number of commercially available meganucleases. ZFNs can also be tailor-made to target specific sequences and used to clone native DNA sequences (69). Thus, ZFNs may provide a useful alternative to methods based on recombinases, meganucleases, transposomes and others for the construction of genomic DNA clones and the artificial cloning of large DNA sequences (e.g. gene clusters) into BAC vectors and other artificial chromosomes (88-91). It is clear, however, that the application of ZFNs as a viable tool for recombinant DNA uses depends on the quality and precision of the recombinant ZFNs.

Here we demonstrate that ZFNs can be purified to cloning grade using Ni-affinity and gel-filtration purification. We show that while gel filtration may produce a cleaner product, Ni affinity may be sufficient for recombinant DNA applications, its product's efficiency being similar to that of PI-*Psp*I, a commercial rare-cutting restriction enzyme. We also show that ZFNs can be used in co-digestion experiments, can be heat-deactivated, can digest their substrate at 37^{o}C—the optimal temperature for many other

restriction enzymes, and can be stored without loss of activity at -80^{o}C. Finally, we show that target specificity can be predicted and analyzed, a characteristic that is not common to other rare-cutting restriction enzymes. These features allow the simple introduction of ZFNs into regular cloning procedures. We have used these procedures for the expression and purification of ZFN-10 and other ZFNs for the production of large transformation cassettes (Zeevi V, Liang Z, Dafny R and Tzfira T., unpublished data). The relative simplicity of the techniques employed for the construction, expression and purification of our ZFNs make our strategy sufficiently attractive in terms of both satisfying routine lab cloning needs and potentially allowing commercial production of ZFN enzymes. In summary, we have developed a simple and efficient way of expressing and purifying ZFN-10 and our studies are expected to be useful for the production of novel ZFNs and their implementation for various applications in recombinant DNA technology.

2.6 Tables and figures

Table 2.1. Quantitative parameters of ZFN-10 purification

Purification step	Total protein content (mg)	Total activity (units)	Specific activity (units/mg)	Purification factor	Yield (%)
Crude	39.7	525,000	13,000	1	100
Ni-NTA	0.54	180,000	333,000	25.6	34
Gel filtration	0.21	48,000	229,000	17.6	9.1

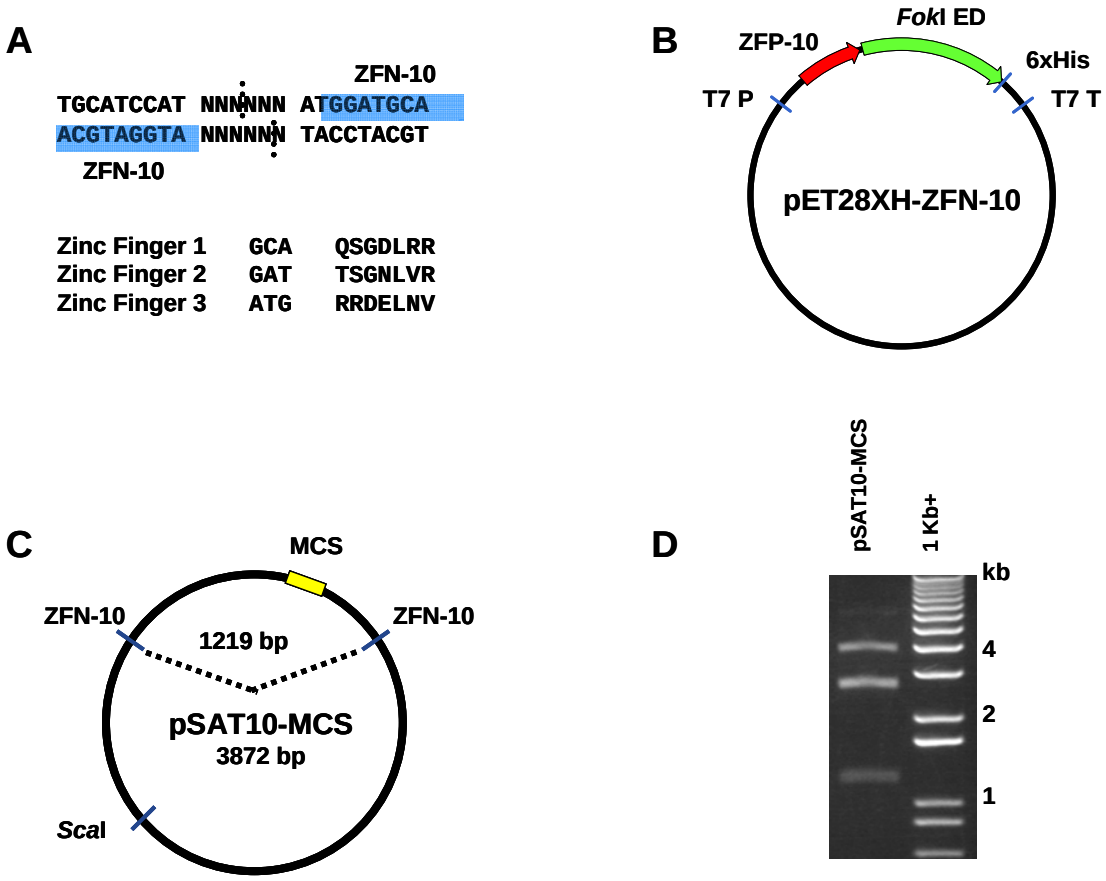


Figure 2.1. Structure of ZFN-10, expression vector and DNA substrate. (**A**) The structure of the 24-bp-long ZFN-10 recognition site and the unique amino-acid sequences in each zinc finger. DNA binding sites are in blue. The predicted cleavage sites are indicated by dotted lines. (**B**) Structure of the pET28XH-ZFN-10 expression vector. The ZFN-10-6xHis tag fusion ORF is under control of the T7 bacterial expression promoter. (**C**) Scheme of the pSAT10-MCS plasmid. Sites of the ZFN-10 palindrome-like recognition sites and the unique *Sca*I site are indicated. (**D**) Restriction analysis of pSAT10-MCS by crude ZFN-10.

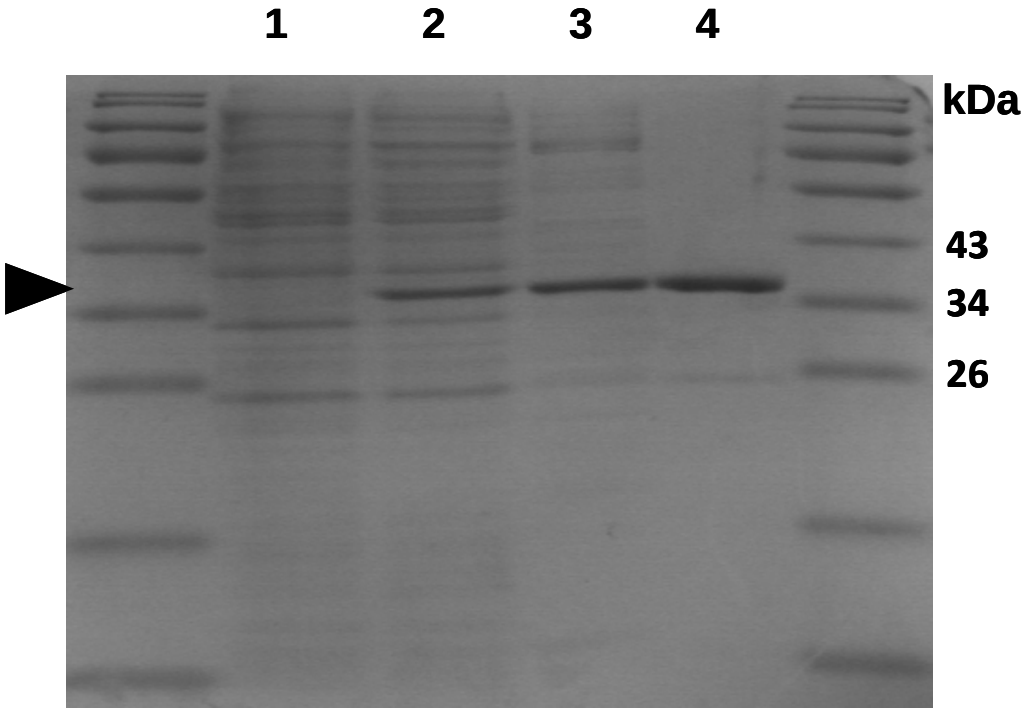


Figure 2.2. SDS-PAGE analysis of ZFN-10. Separation of total proteins from non-induced (lane 1), and IPTG-induced (lane 2) ZFN-10 protein-expressing *E. coli* cells, and of Ni-affinity-purified (lane 3) and gel-filtration-purified (lane 4) ZFN-10 protein. The location of the ca. 34-kDa band corresponding to the ZFN-10 protein is indicated by an arrowhead.

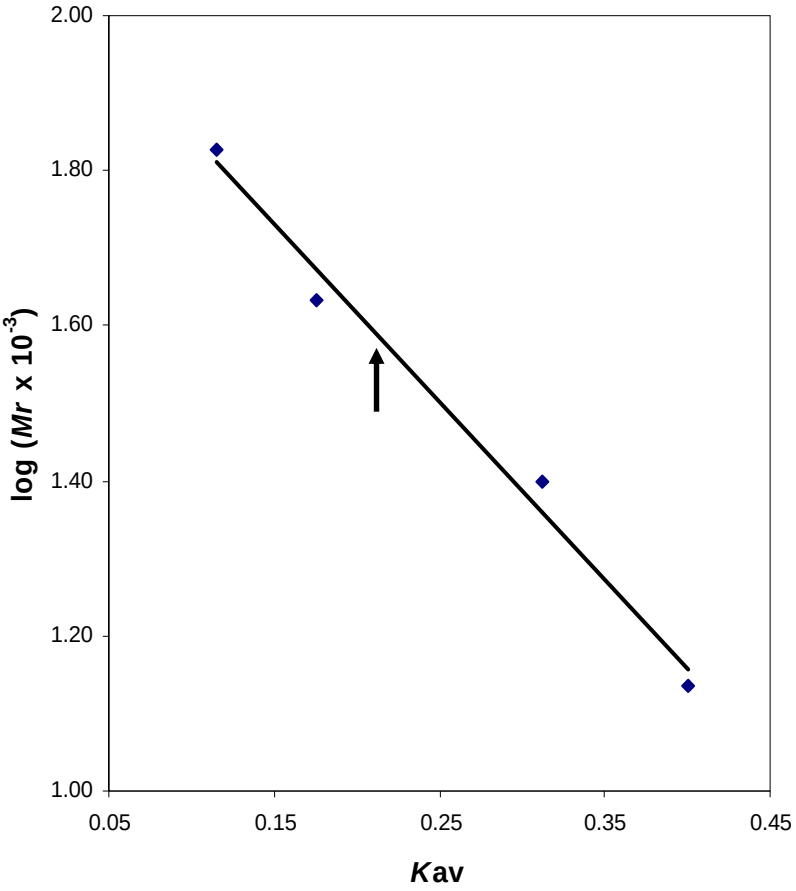
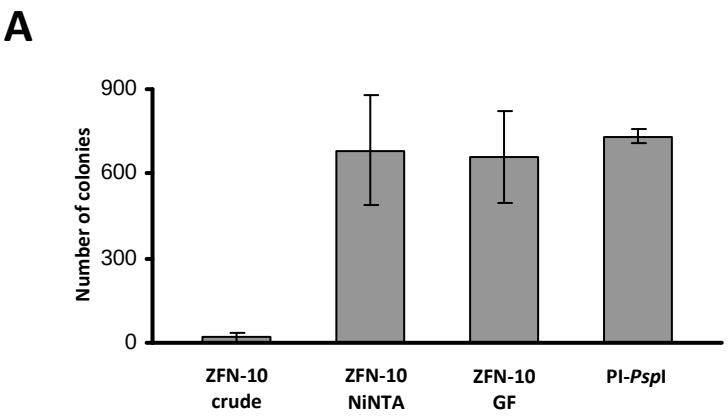
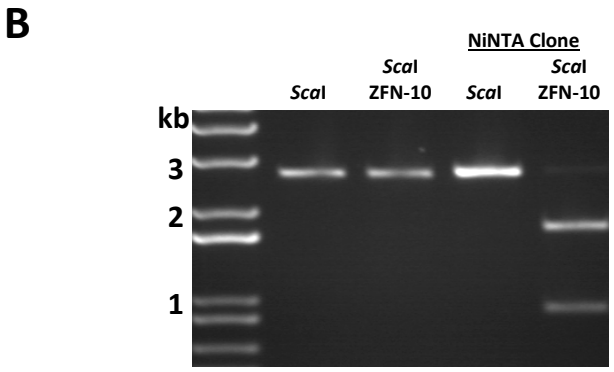


Figure 2.3. Size-exclusion chromatography of ZFN-10. Purified ZFN-10 was eluted from a HiPrep 16/60 Sephacryl S-100 HR column that had been calibrated with ribonuclease A (13.7 kDa), chymotrypsinogen A (25 kDa), ovalbumin (43 kDa), albumin (67 kDa) and blue dextran (2000 kDa) molecular mass standards. Kav values were calculated as $Kav = Ve - Vo/Vt - Vo$ (Vo = void volume, Vt = total column volume and Ve = elution volume). Arrow indicates the Kav corresponding to peak ZFN-10 activity.





C

```
                                ZFN-10          ZFN-10
pSAT10-Ø    CGACGGCCAGTGCCTGCATCCATGTAAGTATGGATGCAGTAATCATGGTCATAGCTGTTTCCTGTGTG

Crude1      CGACGGCCAGTGCCTGCATCCA----------------GTAATCATGGTCATAGCTGTTTCCTGTGTG
Crude2      CGACGGCCAGTGCCTGCATCCATGTAA----------------GGTCATAGCTGTTTCCTGTGTG
Crude3      CGACGGCCAGTGCCTGCATCCATGTAA----------------GGTCATAGCTGTTTCCTGTGTG
Crude4      CGACGGCCAGTGCCTGCATCCATGTAA----------------GGTCATAGCTGTTTCCTGTGTG
Crude5      CGACGGCCAGTGCCTGCATCCATGTAA----------------GTCATAGCTGTTTCCTGTGTG
Crude6      CGACGGCCAGTGCCTGCATCCATG----------------GTCATAGCTGTTTCCTGTGTG
Crude7      CGACGGCCAGTGCCTGCATCCATG----------------GTCATAGCTGTTTCCTGTGTG
Crude8      CGACGGCCAGTGCCTGCATCCATG----------------GTCATAGCTGTTTCCTGTGTG
Crude9      CGACGGCCAGTGCCTGCATCCATGTA----------------AGCTGTTTCCTGTGTG
Crude10     CGACGGCCAGTGCCTGC----------------------------------------TGTG
NiNTA1      CGACGGCCAGTGCCTGCATCCATGTAAGTATGGATGCATAAGTCATGGTCATAGCTGTTTCCTGTGTG
GF1         CGACGGCCAGTGCCTGCATCCATGTAAGTATGGATGCATAAGTCATGGTCATAGCTGTTTCCTGTGTG
```

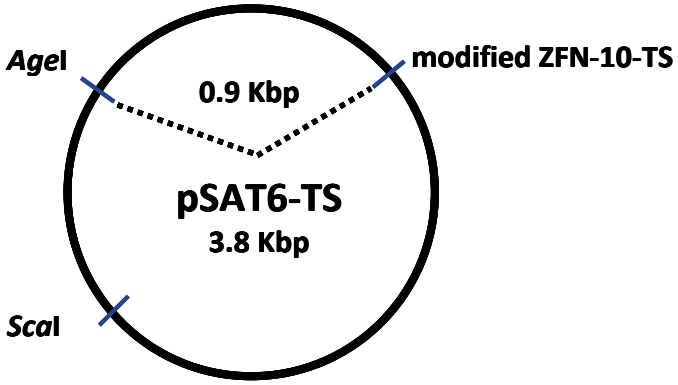
Figure 2.4. Analysis of ZFN-10-digested DNA products. (**A**) Self-ligation efficiency analysis of pSAT10-MCS and pSAT6-MCS backbones. pSAT10-MCS was digested by ZFN-10 lysate (crude), or Ni-affinity-purified (NiNTA) or gel-filtration-purified (GF) ZFN-10. pSAT6-MCS was digested by PI-*Psp*I. (**B**) Restriction analysis of representative pSAT10-ø clones derived from pSAT10-MCS digested by ZFN-10 lysate or Ni-affinity-purified ZFN-10. (**C**) Sequence analysis of representative pSAT10-ø clones derived from

pSAT10-MCS digested by ZFN-10 lysate (clones Crude1 through 10), or Ni-affinity-purified (clone NiNTA1) or gel-filtration-purified (clone GF1) ZFN-10.

A

ATGGATGCA
GG T
C C
A

B



C

pSAT6-ZFN10-TS-(n)	Target sequence	Cutting efficiency
0	TGCATCCAT-GTAAGT-ATGGATGCA	100%
1	AGCATCCAT-GTAAGT-ATGGATGCT	8%
2	GGCATCCAT-GTAAGT-ATGGATGCC	97%
3	TGCATCCCT-GTAAGT-AGGGATGCA	2%
4	TGCATCCGT-GTAAGT-ACGGATGCA	98%
5	TGCATCCTT-GTAAGT-AAGGATGCA	79%
6	TGCATCCAC-GTAAGT-GTGGATGCA	100%
7	GGCATCCCT-GTAAGT-AGGGATGCC	0%
8	TGCATCCCC-GTAAGT-GGGGATGCA	0%
9	AGCATCCTC-GTAAGT-GAGGATGCT	97%
10	TGCATCCAT-GTAAGT-ATGCATGCA	0%
11	TGCATCCAT-GTAAGT-ATGGCTGCA	94%
12	TGCATCCAT-GTAAGT-ATGGACGCA	63%
13	TGCATCCAT-GTAAGT-ATGCCTGCA	0%
14	TGCATCCAT-GTAAGT-ATGCCCGCA	0%

D

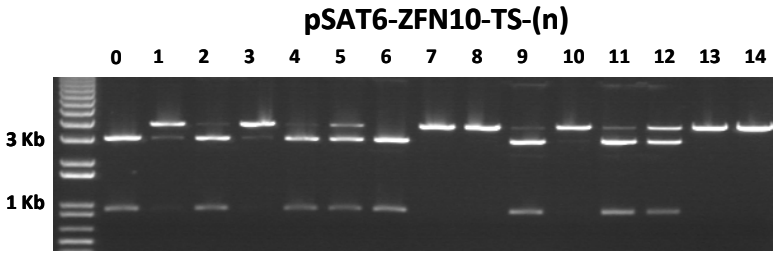
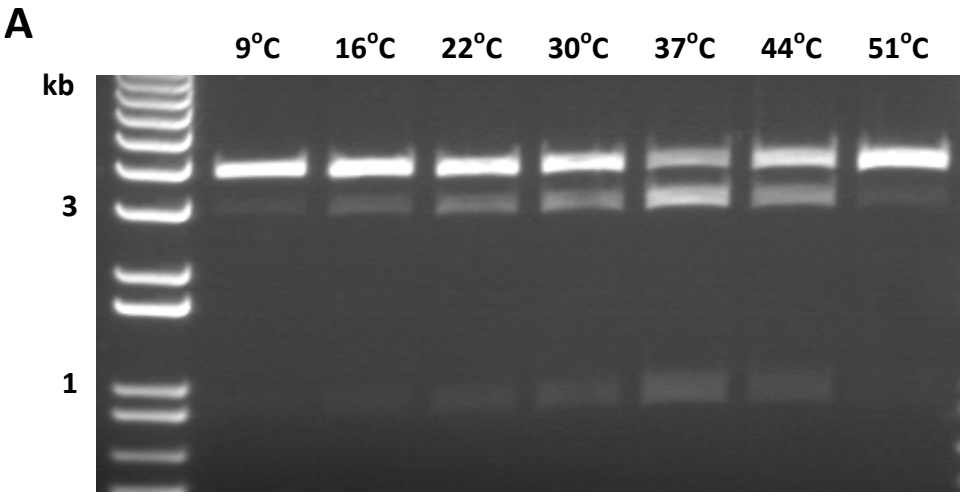
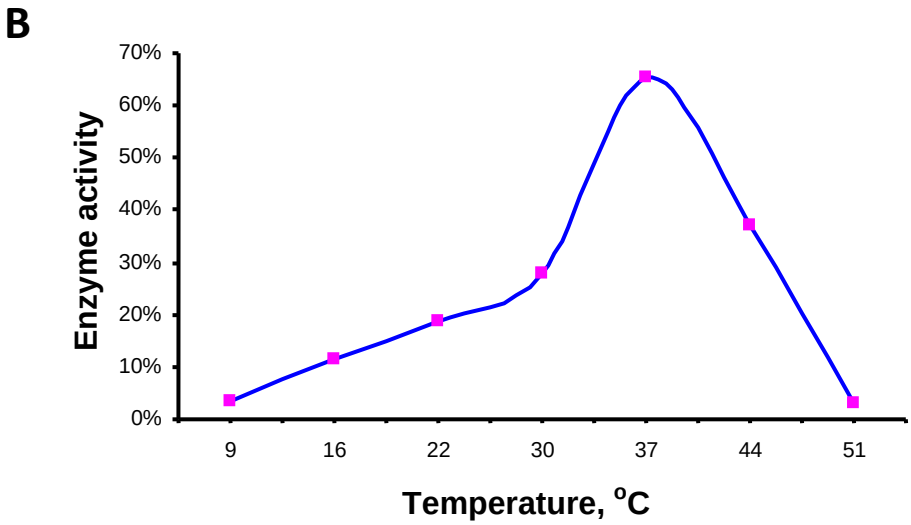


Figure 2.5. Analysis of ZFN-10 target specificity. (**A**) Prediction of ZFN-10 monomer target-site binding degeneracy. (**B**) Scheme of the pSAT6-TS substrate plasmids. Site of the ZFN-10 palindrome-like modified recognition site and the unique *Age*I site are indicated. (**C,D**) Digestion efficiency of modified semi-palindromic ZFN-10 target sites by *Age*I and ZFN-10. Digestion efficiency of different target sites (i.e. sites 1 through 14) is given as percentage of the digestion of the original target site (site 0) (**C**), as analyzed by agarose gel electrophoresis (**D**).


A
9°C
16°C
22°C
30°C
37°C
44°C
51°C
kb
3
1


B
Enzyme activity
70%
60%
50%
40%
30%
20%
10%
0%
9
16
22
30
37
44
51
Temperature, °C

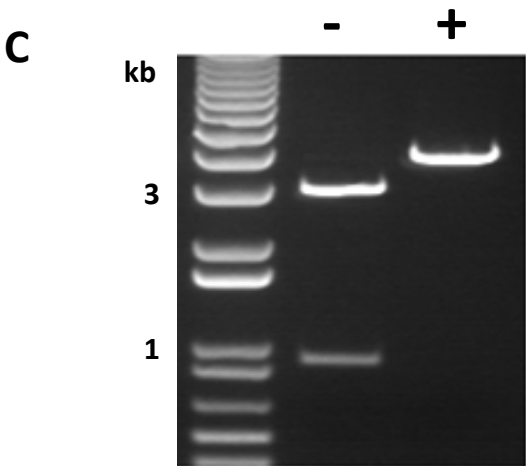

C
-
+
kb
3
1

Figure 2.6. Optimal temperature digestion and inactivation of ZFN-10. (**A,B**) Agarose gel electrophoresis analysis (**A**) and quantified temperature response curve (**B**) of pSAT6-TS-0 digestion by ZFN-10. (**C**) Agarose gel electrophoresis analysis of ZFN-10 digestion activity prior to (-) and after (+) heat inactivation.

Chapter 3. Increasing cloning possibilities using artificial zinc finger nucleases

3.1 Summary

The ability to accurately digest and ligate DNA molecules of different origins is fundamental to modern recombinant DNA research. Only a handful of enzymes are capable of recognizing and cleaving novel and long DNA sequences, however. The slow evolution and engineering of new restriction enzymes calls for alternative strategies to design novel and unique restriction enzymes capable of binding and digesting specific long DNA sequences. Here we report on the use of zinc finger nucleases (ZFNs)—hybrid synthetic restriction enzymes that can be specifically designed to bind and cleave long DNA sequences—for the purpose of DNA recombination. We show that novel ZFNs can be designed for the digestion of specific sequences and can be expressed and used for cloning purposes. We also demonstrate the power of ZFNs in DNA cloning by custom-cloning a target DNA sequence and assembling dual-expression cassettes on a single target plasmid, a task that rarely can be achieved using type-II restriction enzymes. We demonstrate the flexibility of ZFN design and the ability to shuffle monomers of different ZFNs for the digestion of compatible recognition sites through ligation of compatible ends and their cleavage by heterodimer ZFNs. Of no less importance, we show that ZFNs can be designed to recognize and cleave existing DNA sequences for the custom-cloning of native target DNA molecules.

3.2 Introduction

DNA cloning is fundamental for nearly every niche of molecular biology research, and site-specific endonucleases are the cornerstones of modern recombinant DNA technology (57,92). The burden of site-specific DNA digestion falls on a wide variety of restriction enzymes isolated from different organisms. A recent update of the REBASE (enzymes and genes for DNA restriction and modification) database lists >3600 type II restriction enzymes that have been biochemically or genetically characterized (37). But although >600 of these restriction enzymes are currently commercially available, only 262 distinct sequences are targeted (37), most of them 4- to 6-bp palindromic sequences. Thus, not only are many specificities still unavailable, but the number of enzymes capable of recognizing 8- and 10-bp-long sequences is rather small. The situation is even more critical with respect to meganucleases—restriction enzymes capable of targeting very large (>12 bp) sequences. To the best of our knowledge, only six different meganucleases are currently available commercially.

Although many of the 6-bp cutters are widely used for various recombinant DNA applications, their application can be somewhat limited by their high statistical occurrence within various DNA sequences (93). Furthermore, their use for cloning large DNA fragments or for the assembly of complex DNA structures is nearly impossible. Thus, for example, the construction of large artificial DNA molecules (which may contain several reparative elements and/or 6- and 4-bp-long recognition sites) could be facilitated by the introduction of restriction enzymes with long and novel target sequences (e.g., 8- and 10-bp cutters and meganucleases).

Meganucleases (also known as rare cutters) have been particularly useful for *in vivo* applications. Their high specificity makes them extremely useful for genome engineering in mammalian, animal, and plant cells (60,61,63) and for novel cloning purposes (64-66). Various groups are engaged in searching for, modifying, and reengineering existing meganucleases for the targeting of novel sequences (64-66). But the extreme specificity of meganucleases (and of other restriction enzymes, for that matter) poses a great challenge for their engineering (24), and we have yet to witness the commercialization and widespread use of some of the recently reported variant and novel meganucleases (64-66).

Recently, zinc finger nucleases (ZFNs), a new type of restriction enzymes, have been developed as a novel tool for genome engineering in living cells (46,49,51). ZFNs provide an excellent alternative to the tedious and potentially impractical reengineering of existing meganucleases, because they can be engineered to digest virtually any long stretch of DNA sequence (87). ZFNs are modular proteins composed of a specifically engineered zinc finger DNA-binding domain fused to the *Fok*I non-sequence-specific DNA-cleavage domain (shown in **Figure 3.1A**). The zinc finger DNA-binding domain is composed of several fingers, each engineered to bind a specific DNA triplet. Monomers of two ZFNs must be properly aligned and function together to digest a double-stranded target DNA sequence (26,94). The length of the spacer between the two monomers, where *Fok*I actually digests the target DNA, is rather flexible; with a 6-bp spacer, the nuclease has been proposed to typically digest between the second and fourth nucleotides, leaving a 2-base-long 5′ overhang. Thus, a ZFN dimer, composed of two monomers, each engineered to bind a 9-bp DNA sequence, can be used to target a 20- to

24-bp DNA sequence. It also should be noted that custom-designed ZFNs may bind not only to their target sites, but also to similar 9-bp sequences. But due to the very low likelihood of finding two similar 9-bp-long sequences arranged in reverse orientation in a semipalindromic fashion, it is likely that custom-designed ZFNs will exhibit high specificity toward their target sequence. Furthermore, ZFNs with four or even five zinc fingers possibly can be designed if higher specificity is required. Thus, ZFNs can be classified as supermeganucleases, which have potential uses in various cloning tasks.

Recently, several artificially designed ZFNs have been successfully used as powerful tools for gene targeting in various species (46,49,51,87). The rules and protocols for the design, assembly, and testing of new ZFNs have been laid out in several recent publications (32,94,95), but to the best of our knowledge, the application of ZFNs and an evaluation of their potential for *in vitro* DNA cloning purposes have not yet been reported. In this article, we describe the construction, expression, and purification of several ZFNs and their successful use in recombinant DNA procedures. We demonstrate the power of ZFNs in performing tasks that are nearly impossible using the current repertoire of 6-bp restriction enzymes. We also demonstrate the uniqueness of ZFNs, which lies in the ability to shuffle and combine monomers of different ZFNs for cloning purposes, as well as the potential for digesting and cloning native target DNA molecules. Given that >200,000 different combinations are possible with a 9-bp sequence, the ability to engineer ZFNs to bind to such sequences represents a significant increase in the number of restriction enzymes that possibly can be developed and may be useful for DNA cloning.

3.3 Experimental procedures

3.3.1 Molecular assembly of ZFNs

The ZFN DNA-binding regions (i.e., zinc finger protein domains) for ZFN10, ZFN11, ZFN-H2a, and ZFN-H2b were designed based on a zinc-finger framework consensus sequence developed by Desjarlais and Berg (96). The detailed recognition helices are given in Table 3.1. Each zinc finger protein was assembled from sets of overlapping backbone oligos (BBOs) and sequence-dependent oligos (SDOs) as described previously (33), using a high-fidelity *Pfu* DNA polymerase (Stratagene). A comprehensive list of the BBO and SDO primers used for the assembly of ZFN10, ZFN11, ZFN-H2a, and ZFN-H2b is given in Table 3.2 (Table 3.2). Each PCR was composed of 5 pM BBO and SDO primers and 200 pM 5′CCGCTCGAGCTGAAAAACCTTACAAGTGTCC3′ and 5′GGACTAGTCCTCCAGTATGAGTACGTTGATG3′ primers and was carried out for 35 cycles. PCR products were cloned into the XhoI and SpeI sites of a modified pET28 vector, resulting in fusion of the zinc finger protein with the *Fok*I endonuclease domain from pHS::QQR-QEQ/2300 (51) and a 6xHis-tag at the C terminus of the protein driven by the T7 promoter.

3.3.2 Assembly of modified pSAT and pRCS plasmids

To construct pSAT10-MCS, pSAT11-MCS, and pSAT12.1-MCS, the entire plasmid backbone of pSAT6-MCS (41) was PCR-amplified using the appropriate set of primers (Table 3.3). The PCR product was digested by *Age*I and *Not*I, and ligated with the *Age*I-*NotI* 1.2-kb fragment from pSAT6-MCS (41). The YFP-CHS, DsRed2-P, and CHRD-RFP expression cassettes were cloned as *Age*I-*Not*I fragments from pSAT6-EYFP-C1-

CHS, pSAT4-DsRed2-P, and pSAT6(A)-CHRD-mRFP-N1, respectively, producing pSAT10-YFP-CHS, pSAT11-DsRed2-P, and pSAT12.1-CHRD-RFP. pRCS11 was constructed by modifying the MCS of pRCS2 (41); this was done by subsequent cloning of self-annealed pairs of primers (Table 3.3) encoding the ZFN10 and ZFN11 sites into the *Kpn*I and *Sma*I sites, respectively, of pRCS2.

3.3.3 Expression of ZFN protein

For *E. coli* expression, ZFN expression plasmids (i.e., pET28-ZFN10, pET28-ZFN11, pET28-ZFN-H2a, and pET28-ZFN-H2b) were transformed into BL21 GOLD (DE3) PlyS cells (Stratagene). The cells were cultured in 100 ml of Luria broth medium supplemented with 50 μg/μl of kanamycin and 100 μM $ZnCl_2$ and grown at 22°C. At an OD_{600} of 0.6, ZFN expression was induced by 0.7 mM IPTG for 3 h at 22°C. Cells were harvested by centrifugation; resuspended in 35 ml of 25 mM Tris·HCl [pH 7.5], 300 mM NaCl, 5% (vol/vol) glycerol, and 100 μM $ZnCl_2$; and lysed twice using a French press. Proteins were loaded on 0.5 ml of Ni-NTA agarose beads (Qiagen) and eluted with 1 ml of buffer containing 500 mM imidazole. Eluted proteins were stored at −20°C in 50% glycerol. Alternatively, the Expressway *in vitro* protein synthesis system (Invitrogen) was used for *in vitro* expression of ZFN proteins. Crude *in vitro*-produced proteins, with or without further purification through Ni-NTA agarose beads, were used for *in vitro* digestion experiments.

3.3.4 Digestion, ligation, and transformation into *E. coli*

Digestion of ≈200 ng of plasmid DNA was carried out in 10 mM Tris [pH 8.8], 50 mM NaCl, 1 mM DTT, 100 μM $ZnCl_2$, 50 μg/ml of BSA, and 100 μg/ml of tRNA in a total reaction volume of 20–30 μl. Typically, 0.05–1 μl of purified enzyme was added to the

reaction, gently mixed, and preincubated for 30 min at room temperature, to allow the enzyme to bind to its target sequences. $MgCl_2$ was added next, to a final concentration of 5 mM, and then the reaction was incubated for 2–40 min at room temperature. Cleaved fragments were separated by gel electrophoresis and purified using a GFX Gel Band Purification Kit (Amersham). Compatible ends in the target plasmids were dephosphorylated using shrimp alkaline phosphatase (Fermentas), and fragments were ligated with T4 ligase (NEB) and then transferred to chemically competent DH5α *E. coli* cells using standard molecular biology protocols.

3.3.5 Microbombardment and confocal microscopy

For biolistic delivery, 50 µg of DNA was adsorbed onto 10 mg of 1-µm gold particles (Bio-Rad) according to the manufacturer's instructions, and then microbombarded onto the leaf epidermis of greenhouse-grown *Nicotiana tabacum* cv. Turk plants. For cobombardment of two constructs, the plasmids were mixed at a 1:1 molar ratio before the adsorption onto gold beads. The microbombardment was performed at a pressure of 150 psi using a portable Helios gene gun system (PDS-1000/He; Bio-Rad). After incubation for 16–24 h at 25°C to allow expression of the transfected DNA, the plant tissues were viewed directly under a confocal laser-scanning microscope (TCS SP5; Leica). Enhanced YFP was excited by an argon laser at 514 nm, and fluorescence was monitored between 520 and 540 nm. DsRed2 was excited by a helium-neon laser at 543 nm, and fluorescence was monitored between 570 and 630 nm. Chlorophyll fluorescence was monitored >660 nm. All experiments were repeated at least three times, with 10–20 cells exhibiting fluorescent signal examined in each independent experiment.

3.4 Results

3.4.1 ZFN-digested DNA ends can be ligated to each other

One of the basic requirements for newly isolated or recombinant restriction enzymes is their ability to accurately and specifically digest donor and recipient DNA molecules. We thus tested whether an artificial ZFN could specifically and accurately digest target DNA molecules, and also whether the products of such digestion could be purified and ligated using common molecular cloning protocols. Toward this end, we first engineered the plasmid pSAT10-MCS (multiple cloning site) to carry two identical 24-bp sequences, each composed of a pair of 9-bp sequences, acting as a palindrome, separated by a 6-bp nonpalindromic sequence (**Figure 3.1B**). We next engineered ZFN10, a ZFN monomer with three zinc fingers capable of binding in a specific orientation to the ATGGATGCA sequences on pSAT10-MCS, as illustrated in **Figure 3.1A**. Because an active ZFN10 site is actually composed of two 9-bp sequences making up a "palindrome-like" recognition site, two ZFN10 proteins were expected to function as a homodimer and cleave within the 6-bp spacer, as reported previously for other ZFNs (72). It is important to note that we designed the spacer sequence such that if digestion occurs between the second and fourth nucleotides, located on the upper and lower strands, respectively (**Figure 3.1A**), then the overhanging parts of the digested molecules can easily ligate to each other.

We expressed ZFN10 in *E. coli* cells, but the crude protein extract proved useless for accurate digestion of our target pSAT10-MCS plasmid, leading to nearly complete degradation of the sample DNA (data not shown). Thus, we further purified the enzyme by binding it to Ni-NTA agarose beads (**Figure 3.1C**). Indeed, Ni-NTA–purified ZFN10 was able to release a specific ZFN10-flanked 1.2-kb fragment from pSAT10-MCS

(**Figure 3.1D**). Similarly, *in vitro*-expressed ZFN10 also released a well defined ZFN10-flanked fragment from pSAT10-MCS (data not shown). Thus, both *E. coli*– and *in vitro*-expressed ZFNs were useful for the efficient and specific digestion of target DNA molecules, and we used them alternately throughout this study.

We next purified the ZFN10-digested plasmid backbone from the gel and subjected it to a simple *in vitro* ligation procedure. We used commercial T4 ligase in an overnight ligation reaction and then transferred the entire reaction into chemically competent DH5α *E. coli* cells. We recovered a large number of recombinant colonies and analyzed their structure by both restriction pattern analysis and DNA sequencing. **Figure 3.1D** shows that whereas the digestion of the parental pSAT10-MCS by *Age*I and *Not*I (which are located internal to the ZFN10; **Figure 3.1B**), or solely by ZFN10, released a 1.2-kb fragment, digestion of the self-ligated plasmid's backbone (pSAT- ) was limited to ZFN10. The plasmid no longer could be digested by *Age*I or *Not*I (compare the *Age*I/*Not*I and ZFN10 digestions of pSAT10-MCS and pSAT-  **Figure 3.1D**). Indeed, pSAT-  was missing the entire MCS fragment and was 1.2 kb shorter than its parental plasmid pSAT10-MCS (compare the *Nde*I digestions of pSAT10-MCS and pSAT-  **Figure 3.1D**). This digestion pattern suggests that after removal of the ZFN10-digested 1.2-kb fragment, ligation of the two ZFN10 sites resulted in the accurate reconstruction of a new ZFN10 site at the ligation site. Indeed, DNA sequencing of several recombined clones clearly showed the accurate assembly of a new ZFN10 site (**Figure 3.1E**). Thus, ZFNs not only can be expressed using both *E. coli* and *in vitro* expression systems, but also can be used for accurate digestion, and the ZFN-digested DNA molecules can be accurately ligated and recovered using common recombinant DNA protocols.

3.4.2 ZFNs can be used to assemble large and complex DNA constructs

Once we found that ZFN10 can be used to digest and self-ligate single DNA molecules, we then set out to determine whether ZFNs can be designed and used to recombine DNA fragments from different plasmids. We chose the challenge of assembling two expression cassettes onto a single plasmid. This task cannot typically be achieved using type II restriction enzymes, because of the high occurrence of their recognition sites in the two expression cassettes and the cloned genes (38). One approach to assembling multiple expression cassettes onto a single plasmid is to flank each cassette with novel sequences and use rare-cutters for their digestion and cloning into an acceptor plasmid (38,41). Basing our strategy on this approach, we decided to construct two plasmids in which each expression cassette was flanked by a different ZFN. We began by constructing pSAT10-YFP-CHS and pSAT11-DsRed2-P, a pair of satellite plasmids (41) carrying plant expression cassettes for the yellow fluorescent protein (YFP)-tagged endoplasmic reticulum protein chalcone synthase (CHS) (97) and the DsRed2-tagged protein P of Sonchus Yellow Net Virus (SYNV) (98), respectively. The YFP-CHS and DsRed2-P expression cassettes were flanked by ZFN10 and ZFN11 24-bp-long recognition sites, respectively (**Figure 3.2A**). We also engineered ZFN11, a ZFN monomer capable of binding to the 9-bp sequences composing the ZFN11 recognition site (data not shown). We constructed pRCS11, a binary plasmid, to carry the ZFN10 and ZFN11 recognition sites on its MCS (**Figure 3.2B**). We successively cloned the YFP-CHS and DsRed2-P expression cassettes from pSAT10-YFP-CHS and pSAT11-DsRed2-P using ZFN10 and ZFN11, respectively, into their corresponding sites on pRCS11. Bombardment of the resultant plasmid, pRCS11[YFP-CHS][DsRed2-P], into tobacco mesophyll cells

demonstrated the successful mounting of both expression cassettes onto the acceptor plasmid.

Consistent with previous observations (97,98), DsRed2-P was found in both the cell nucleus and the cytoplasm, whereas YFP-CHS decorated the rough endoplasmic reticulum and the nucleus (**Figure 3.2C**). Furthermore, when we cobombarded pRCS11[YFP-CHS][DsRed2-P] with pSAT3-N (41), a plasmid expressing free SYNV N protein, the DsRed2-P accumulated predominantly in the nucleus (compare the expression pattern of DsRed2-P in the presence and absence of N; **Figure 3.2C**), in accordance with the N protein's function: translocating the P protein into the host cell nucleus (98).

We next analyzed the ligation junction site sequences, and found that ZFN-digested cassettes were accurately ligated into their corresponding cloning sites on pRCS11 (**Figure 3.2D**). Collectively, our data suggest that ZFNs can be used successfully for the accurate digestion and recombination of DNA fragments from different sources and can be useful for unique tasks, such as the assembly of complex and multigene transformation vectors.

3.4.3 ZFNs can be used to clone native DNA molecules

The potential use of ZFNs for cloning purposes goes far beyond the cloning of artificially assembled DNA sequences. An important feature of ZFNs is their ability to recognize and cleave existing native sequences. We chose the *Arabidopsis* ATP-dependent chloroplast-protease-encoding FtsH2 gene as a model sequence to demonstrate how ZFNs can be designed and used to assist in the cloning of existing DNA molecules. To further demonstrate the flexibility of ZFN design (i.e., targeting various lengths of

sequences), we selected a 22-bp-long target sequence and constructed a pair of new ZFN monomers designated ZFN-H2a and ZFN-H2b, capable of binding two target sequences on our model genomic sequence separated by a 4-bp-long spacer (**Figure 3.3A**). We next used ZFN-H2a and ZFN-H2b as a heterodimer ZFN to digest their target and linearize a plasmid carrying a genomic fragment containing the FtsH2 gene. The linearized plasmid was gel-purified and further digested with *Kpn*I, which released the expected 1.5-kb fragment. Because digestion with ZFNs leaves a 5′ overhang, we used Klenow fragment to fill in the recessed 3′ ends of the digested site, and cloned the semiblunted fragment into the *Sma*I and *Kpn*I sites of our target plasmid, pSL301. Sequence analyses of the ligation junction between the FtsH2 fragment and the *SmaI* site on pSL301 (**Figure 3.3B**) in recombinant molecules revealed the different cleavage point made by the ZFN-H2a/ZFN-H2ba pair digested at the 4-bp spacer between the 9-bp DNA-binding recognition sites (**Figure 3.3B**). These observations are in line with a previous report (72) showing that ZFN digestion patterns may depend on the length of the spacer between the ZFN monomers' binding sites. Thus, our data demonstrate that ZFNs can be designed to facilitate *in vitro* cloning of not only artificial, but also native DNA sequences.

3.4.4 ZFN monomers can be shuffled for digestion compatibility

The unique structure of a typical ZFN recognition site allows for flexible design in producing ZFNs with compatible restriction patterns. With the proper spacer sequence, pairs of ZFNs can be designed to function as compatible enzymes during the ligation of two DNA molecules (similar to the compatibility existing between, e.g., *Bam*HI and *Bgl*II). To demonstrate this feature for artificially designed ZFNs, we linearized pSAT12.1-CHRD-RFP, a plasmid carrying a ZFN-H2a recognition sequence (**Figure**

3.4A) and the RFP-tagged chloroplast ChrD protein expression cassette, and cloned it into the ZFN10 recognition site on pRCS11 (**Figure 3.4B**). Bombardment of the resultant plasmid, pRCS11[10/12.1-CHRD-RFP], into tobacco mesophyll cells demonstrated the functionality of the CHRD-RFP expression cassette, as evidenced by a distinct red signal in the chloroplasts (data not shown). More importantly, sequence analyses of the ligation-junction sites revealed accurate ligation between the ZFN-H2a- and ZFN10-digested DNA fragments (**Figure 3.4C**).

Naturally, because the palindrome-like ZFN10 and ZFN-H2a recognition sequences at the ligation sites were lost after the ligation (**Figure 3.4C**), we predicted that ZFN10 and ZFN-H2a would not be able to digest any of these sites on the recombined plasmid. To examine this hypothesis, we isolated a 1-kb *Apa*LI-*Hin*dIII fragment containing a ZFN-H2a/ZFN10 recombination site from pRCS11[10/12.1-CHRD-RFP] and redigested it with either ZFN10 or ZFN-H2a. Indeed, as shown in **Figure 3.4D**, although the fragment could be digested by *Kpn*I, an enzyme with recognition sites located near the ZFN-H2a/ZFN10 junction recognition site, neither ZFN10 nor ZFN-H2a alone was capable of digesting the recombined site. More importantly, because both ZFN10 and ZFN-H2a could each still bind to their corresponding 9-bp sites and because they share a similar structure typical of all ZFNs, they potentially could function as heterodimers and digest at the junction site. Indeed, digestion of the 1-kb *Apa*LI-*Hin*dIII fragment with a combination of ZFN10 and ZFN-H2a resulted in a digestion pattern similar to that after digestion with *KpnI* alone (**Figure 3.4D**), suggesting that the ZFN10-ZFN-H2a heterodimer indeed is capable of binding and digesting the ZFN10/ZFN-H2a

junction recognition site. Thus, our data demonstrate that ZFN monomers can be shuffled and matched for digestion, cloning, and analysis of recombined molecules.

3.5 Discussion

Novel restriction enzymes with unique specificities, particularly enzymes that are capable of recognizing and cleaving long sequences, are extremely useful for genome analysis and manipulation, assembly of complex DNA structures, and genome editing (99-101). Thus, although other molecular techniques, such as overlap extension polymerase chain reaction (PCR), *Sfi*I-based ligation, multiGATEWAY recombination, and uracil-DNA glycosylase, can be used for the construction of complex DNA structures (38,102), the search continues for novel and rare-cutting restriction enzymes that can facilitate the cloning of DNA molecules. Various approaches have been taken to creating artificial rare cutters, most involving engineering existing enzymes for unique specificities. For example, a rational approach combined with a cell-based high-throughput assay was recently used to identify functional variants of the naturally occurring rare cutter I-*Cre*I (65). In another approach, a unique positive and negative selection system was developed to direct the *in vivo* evolution of a I-*Sce*I mutant with altered DNA-cleavage specificities (103). Although such approaches potentially can be used to develop additional variants and novel restriction enzymes, they have proven quite challenging. In fact, most of these methods require special expertise (24,104), and they rarely if ever help in the prediction or assembly of new restriction enzymes with predetermined novel target sequences. Thus, for example, although Lanio *et al.* (105) used a structure-guided approach to alter the recognition sequence of *Eco*RV and produced several variants with altered selectivity, none had the desired (and predicted) selectivity. Because the structure of many restriction

enzymes is not modular by nature (in contrast to, e.g., zinc finger transcription factors), reengineering such enzymes to bind novel sequences while maintaining their specific enzymatic activity is a challenge (104).

The development of ZFNs as a tool for genome engineering (87) laid the foundation for the design and assembly of novel restriction enzymes. Important advantages of ZFNs as rare-cutting restriction enzymes include the clear separation between their DNA-binding and catalytic domains, and the ability to design them from the ground up. The rules and protocols for the design and assembly of novel ZFNs have been laid out in several recent publications (32,33,94,95,106), and a collection of ZFNs already has been assembled and successfully used in gene-targeting experiments in various species (46,49,51,87). During the course of this study, we successfully used publicly available knowledge and information for the design and construction of four novel ZFNs, all of which were then successfully used for cloning purposes. We used basic protein expression and purification methods for the production of our proteins. These protocols are similar to those used for the development and analysis of other ZFN proteins (33,46,94,95); thus, we suggest that similar strategies can be used to produce novel ZFNs suitable for *in vitro* DNA digestion in cloning tasks. We also used routine molecular biology procedures for digestion, purification, and ligation of our target and vector DNA molecules, further supporting the notion that self-produced ZFNs can be easily implemented into existing recombinant DNA protocols.

The biochemical properties and digestion characteristics of ZFNs have been the subject of several reports (33,72). When properly aligned on their target DNA, two ZFN monomers will act together to digest spacers of different lengths and will leave various

lengths of 5′ overhang (33,72,94). This phenomenon allows for great flexibility in the design and use of ZFN monomers for the cloning of native sequences, and also in the construction of artificial complex DNA constructs. We demonstrated this flexibility during the cloning of the FtsH2 genomic clone (**Figure 3.3**) and the assembly of the CHRD-RFP expression cassette onto a plant-transformation vector (**Figure 3.4**). We showed that although ZFN-H2a was originally designed to function together with ZFN-H2b in digesting a unique 22-bp-long target site in the FtsH2 genomic sequence, it also could be used to target a 24-bp palindrome-like recognition site carrying a 6-bp spacer. Furthermore, we showed that different ZFN monomers potentially can be shuffled to act together, because the product of ZFN-H2a was ligated into a ZFN10-digested vector, and product could then be digested by a mixture of ZFN-H2a and ZFN10 (**Figure 3.4D**).

ZFN-H2a and ZFN-H2b were designed to target a native genomic sequence and were intentionally located only a short (4-bp) distance from each other (**Figure 3.3A**). This design allowed us to demonstrate that, in line with the notion that the ZFN digestion pattern depends on the length of the target spacer (72), digestion of the 4-bp-long sequence indeed occurred at different positions on the spacer (**Figure 3.3B**). Although we cannot rule out the possibility that the digestion of a 6-bp-long spacer also produces various 5′ overhang products, a sequence analysis of the ZFN-H2a/ZFN10 ligation junctions indicates that at least in the cases that we examined, digestion of both the target and vector sequences occurred between the third and fourth nucleotides of the spacer (**Figure 3.4C**). Rare cutters' ability to achieve site-specific digestion of such long spacers allows for another important use, in directional cloning (**Figure 3.2**). Directional cloning is particularly important during the assembly of complex multigene transformation

vectors, where the direction of several expression cassettes in relation to others may affect the expression patterns of the cloned genes. Whereas the assembly of multigene expression cassettes possibly can be acheived by various means, methods that are based on the use of rare-cutting restriction enzymes, are expected to be more relevant to the assembly of multigene transformation vectors (38) and other complex DNA structures. The latter approaches certainly will benefit from the introduction of a much larger collection of novel rare cutters. Although in this article we have described the use of only four new rare cutters for cloning purposes, the previous descriptions of many others (46,49) along with the relative ease of design and assembly of such enzymes, support the notion that various novel ZFNs can be constructed and introduced for recombinant DNA applications. Furthermore, because ZFN monomers with four or more fingers potentially can be designed to target sequences of 12 bp or longer, they can be used to produce ZFNs with extremely high specificity.

To conclude, here we have shown for the first time that ZFNs can be used for accurate *in vitro* cloning of DNA molecules of both artificial and native DNA sequences. Furthermore, we have shown that ZFNs can be used in conjunction with commercially available restriction enzymes, and that pairs of ZFNs can be designed for compatibility. Because ZFNs can be specifically tailored to suit individual needs (32,33,94), they represent a powerful and novel addition to the molecular biologist's toolbox.

3.6 Tables and figures

Table 3.1. Recognition helices for ZFN10, ZFN11, ZFN-H2a, and ZFN-H2b

	Target (5'→ 3')	Recognition Helix I	Recognition Helix II	Recognition Helix III
ZFN10	ATGGATGCA	QSGDLRR[a]	TSGNLVR[a]	RRDELNV[b]
ZFN11	TTGTGGGAA	QSSNLQK[c]	RSDHLTT[d]	RSDSLTK[e]
ZFN-H2a	GTTGGTGCT	QSSDLTR[f]	TSGHLVR[a]	TSGSLVR[a]
ZFN-H2b	GCTGCTGTC	DRSALAR[f]	QSSDLTR[f]	QSSDLTR[f]
[a](29), [b](27), [c](107), [d](30),[e](33), [f](106)				

Table 3.2. Primers used for the assembly of ZFN10, ZFN11, ZFN-H2a, and ZFN-H2b. Underlined letters indicate complementary regions between BBO and the corresponding SDO primers

Name	Use	Sequence (5'→ 3')
BBO1	Backbone primer 1	GAAAAACCTTACAAGTGTCCTGAATGTGGAAAGTCTTTTTCT
BBO2	Backbone primer 2	CAGCGAACACACACAGGTGAGAAGCCATATAAATGCCCAGAATGTGGTAAATCATTCAG
BBO3	Backbone primer 3	CAACGGACCCACAfCCGGGGAGAAGCCATTTAAATGCCCTGAGTGCGGGAAGAGTTTTT
ZFP10-SDO1	Helix primer I in ZFP10	ACCTGTGTGTGTTCGCTGGTGACGACGCAAATCTCCAGACTGAGAAAAAGACTTTCCACA
ZFP10-SDO2	Helix primer II in ZFP10	CCCGGTGTGGGTCCGTTGGTGACGAACCAAATTTCCAGAAGTACTGAATGATTTACCACA
ZFP10-SDO3	Helix primer III in ZFP10	TCCAGTATGAGTACGTTGATGAACATTCAATTCATCACGACGTGAAAAACTCTTCCCGCAC
ZFP11-SDO1	Helix primer I in ZFP11	ACCTGTGTGTGTTCGCTGGTGCTTCTGAAGGTTGCTAGACTGAGAAAAAGACTTTCCACA
ZFP11-SDO2	Helix primer II in ZFP11	CCCGGTGTGGGTCCGTTGGTGAGTAGTAAGATGATCGGAACGACTGAATGATTTACCACA
ZFP11-SDO3	Helix primer III in ZFP11	TCCAGTATGAGTACGTTGATGCTTAGTTAGTGAATCGGAACGTGAAAAACTCTTCCCGCAC
ZFN-H2a-SDO1	Helix primer I in ZFN-H2a	ACCTGTGTGTGTTCGCTGGTGACGAGTAAGATCAGAAGACTGAGAAAAAGACTTTCCACA
ZFN-H2a-SDO2	Helix primer II in ZFN-H2a	CCCGGTGTGGGTCCGTTGGTGACGAACAAGATGTCCAGAAGTACTGAATGATTTACCACA
ZFN-H2a-SDO3	Helix primer III in ZFN-H2a	TCCAGTATGAGTACGTTGATGACGAACAAGAGATCCAGAAGTTGAAAAACTCTTCCCGCAC
ZFN-H2b-SDO1	Helix primer I in ZFN-H2b	ACCTGTGTGTGTTCGCTGGTGACGAGCAAGAGCAGAACGATCAGAAAAAGACTTTCCACA
ZFN-H2b-SDO2	Helix primer II in ZFN-H2b	CCCGGTGTGGGTCCGTTGGTGACGAGTAAGATCAGAAGACTGACTGAATGATTTACCACA
ZFN-H2b-SDO3	Helix primer III in ZFN-H2b	TCCAGTATGAGTACGTTGATGACGAGTAAGATCAGAAGACTGTGAAAAACTCTTCCCGCAC

Table 3.3. Primers used for the construction of pSAT10, pSAT11, pSAT12.1, and pRCS11. Underlined letters indicate ZFN target sites.

Name	Used in assembly of	Sequence (5'→ 3')
SAT10-F	pSAT10	ATAAGAATGCGGCCGCTGCATCCATGTAAGTATGGATGCAGTAATCATGGTCATAGCTGTTTCC
SAT10-R-V2	pSAT10	GACGCACCGGTTGCATCCATACTTACATGGATGCAGGCACTGGCCGTCGTTTTACAACG
SAT11-F	pSAT11	ATAAGAATGCGGCCGCTTCCCACAAGTAAGTTTGTGGGAAGTAAGTATGGATGCAGTAATCATGGTCATAGCTGTTTCC
SAT11-R	pSAT11	GACGCACCGGTTTCCCACAAACTTACTTGTGGGAAGGCACTGGCCGTCGTTTTACAACG
SAT12-F	pSAT12.1	ATAAGAATGCGGCCGCAGCACCAACGTAAGTGTTGGTGCTGTAAGTATGGATGCAGTAATCATGGTCATAGCTGTTTCC
SATI2-R	pSAT12.1	GACGCACCGGTAGCACCAACACTTACGTTGGTGCTGGCACTGGCCGTCGTTTTACAACG
RCS11-UP-1	pRCS11	TCCCCCGGGTTCCCACAAACTTACTTGTGGGAAAGCACCAACACTTACGTTGGTGCTCCCGGGGGA
RCS11-DW-1	pRCS11	TCCCCCGGGAGCACCAACGTAAGTGTTGGTGCTTTCCCACAAGTAAGTTTGTGGGAACCCGGGGGA
RCS11-UP-2	pRCS11	GGGGTACCTGCATCCATACTTACATGGATGCAGGTACCCC
RCS11-DW-2	pRCS11	GGGGTACCTGCATCCATACTTACATGGATGCAGGTACCCC

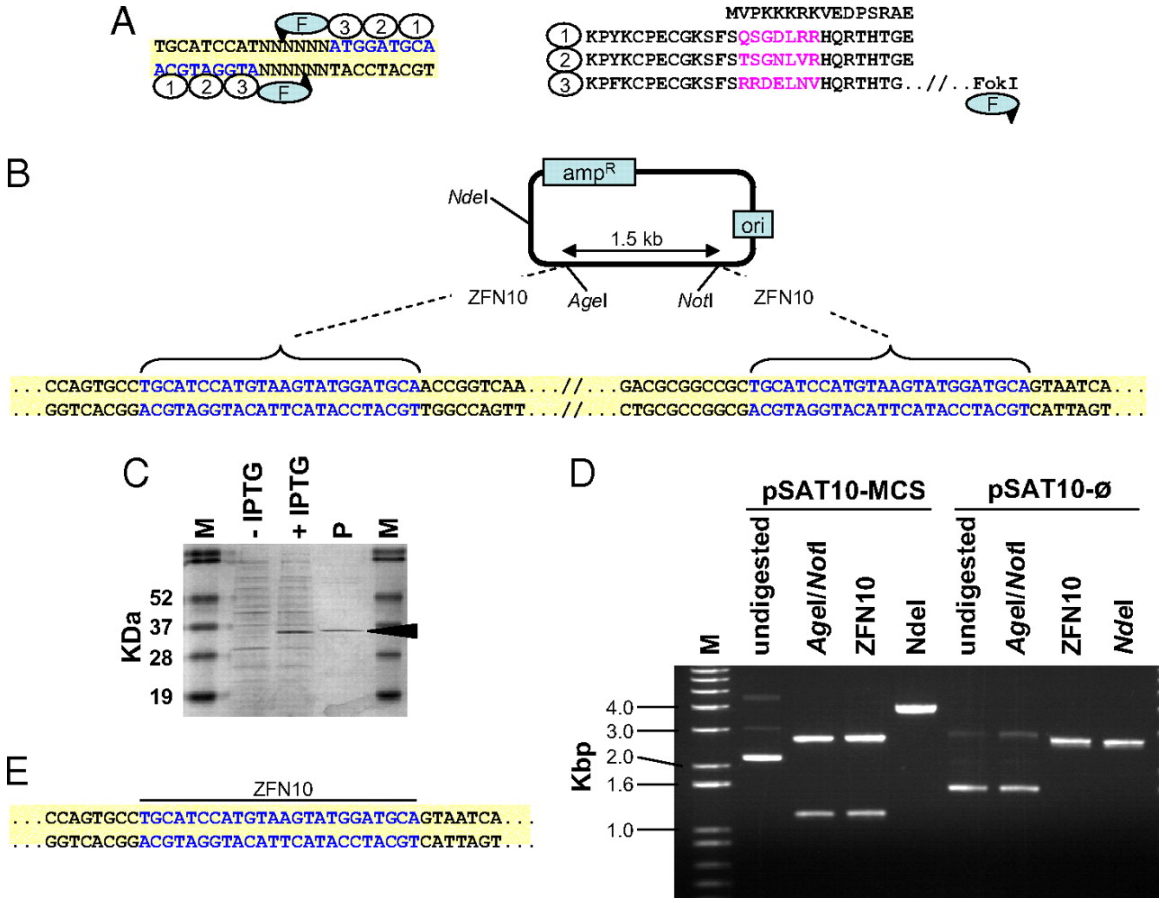


Figure 3.1. Structure and expression of ZFNs and their use for DNA cloning. A. The structure of a typical 24-bp-long ZFN recognition site and its corresponding zinc finger protein, exemplified by ZFN10. DNA triplets (blue) and their binding zinc fingers (ovals) are numbered correspondingly. The unique amino acid sequences in each zinc finger are in purple. The *Fok*I cleavage domain (F) is linked to the zinc finger protein's carboxyl terminus, and the predicted cleavage sites are indicated by arrowheads. B. Scheme of the pSAT10-MCS plasmid. Sequences of the ZFN10 palindrome-like recognition sites are in blue. C. Separation of total crude extract from induced (+IPTG) and noninduced (-IPTG) ZFN10 protein-expressing *E. coli* cells and of purified ZFN10 protein (P). The 34-kDa band corresponding to the ZFN10 protein is indicated. D. Restriction analysis of the parental pSAT10-MCS and its progeny plasmid pSAT10-Ø. E. Sequence analysis of the ZFN10 ligation site in one of the pSAT10-Ø plasmids. Sequences of the reconstructed ZFN10 palindrome-like recognition sites are in blue.

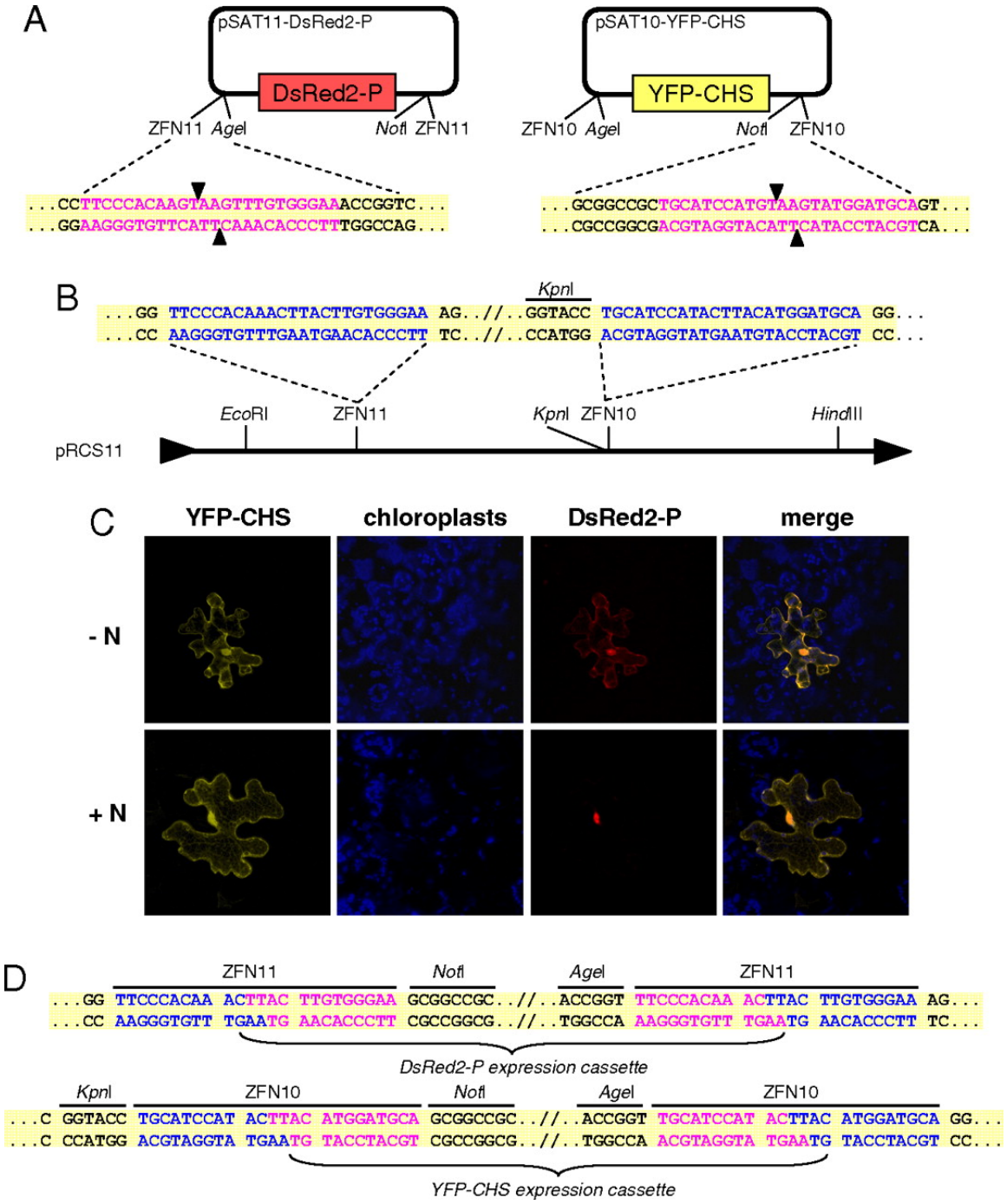


Figure 3.2. Assembly of dual-expression cassettes by ZFNs. A. Schemes of pSAT10-YFP-CHS and pSAT11-DsRed2-P. Sequences of ZFN10 and ZFN11 recognition sites are in purple, and the predicted cleavage sites in each sequence are indicated by arrowheads. B. Scheme of the pRCS11 acceptor plasmid. Sequences of ZFN10 and ZFN11 recognition sites in the plasmid's MCS are in blue. C. Detection of YFP-CHS and DsRed2-P expression from pRCS11[YFP-CHS][DsRed2-P]-bombarded plant cells in the presence (+N) or absence (-N) of the N protein of SYNV. Expression of YFP-CHS, DsRed-P, and chloroplast autofluorescence is shown in yellow, blue, and red, respectively. The images are projections of several confocal sections. D. Sequence

analysis of the ZFN10 and ZFN11 ligation sites in pRCS11[YFP-CHS][DsRed2-P]. ZFN10 and ZFN11 sequences derived from the acceptor plasmid pRCS11 are in blue, and those derived from the inserted expression cassettes are in purple.

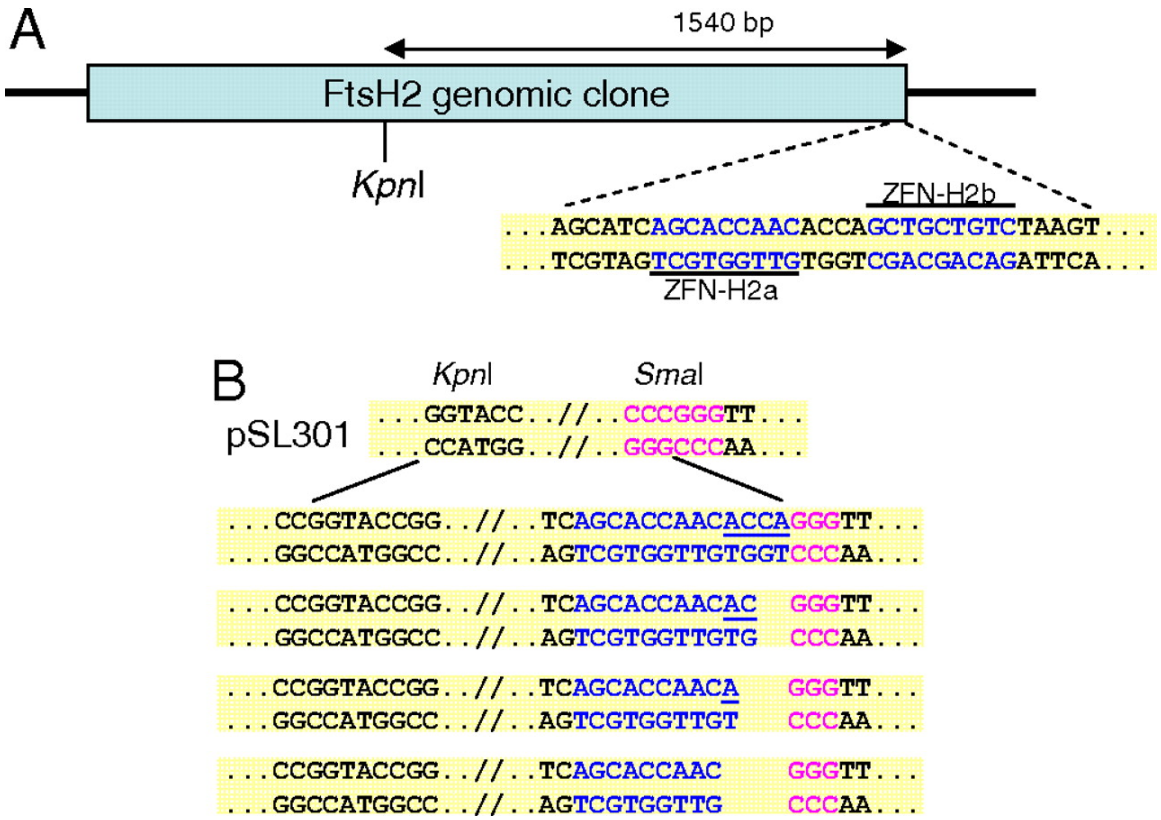


Figure 3.3. Custom-cloning of a target DNA sequence. A. Outline of the *Arabidopsis* FtsH2 genomic clone. Sequences of the ZFN-H2a and ZFN-H2b binding sites are shown in blue. B. Sequence analysis of ZFN-H2a-ZFN-H2b/*Sma*I and *Kpn*I/*Kpn*I junctions in various pSL301-AtFtsH2 clones. Sequences derived from the SmaI site on acceptor plasmid pSL301 are in purple, and those derived from the inserted DNA are in blue. Filled-in nucleotides are underlined.

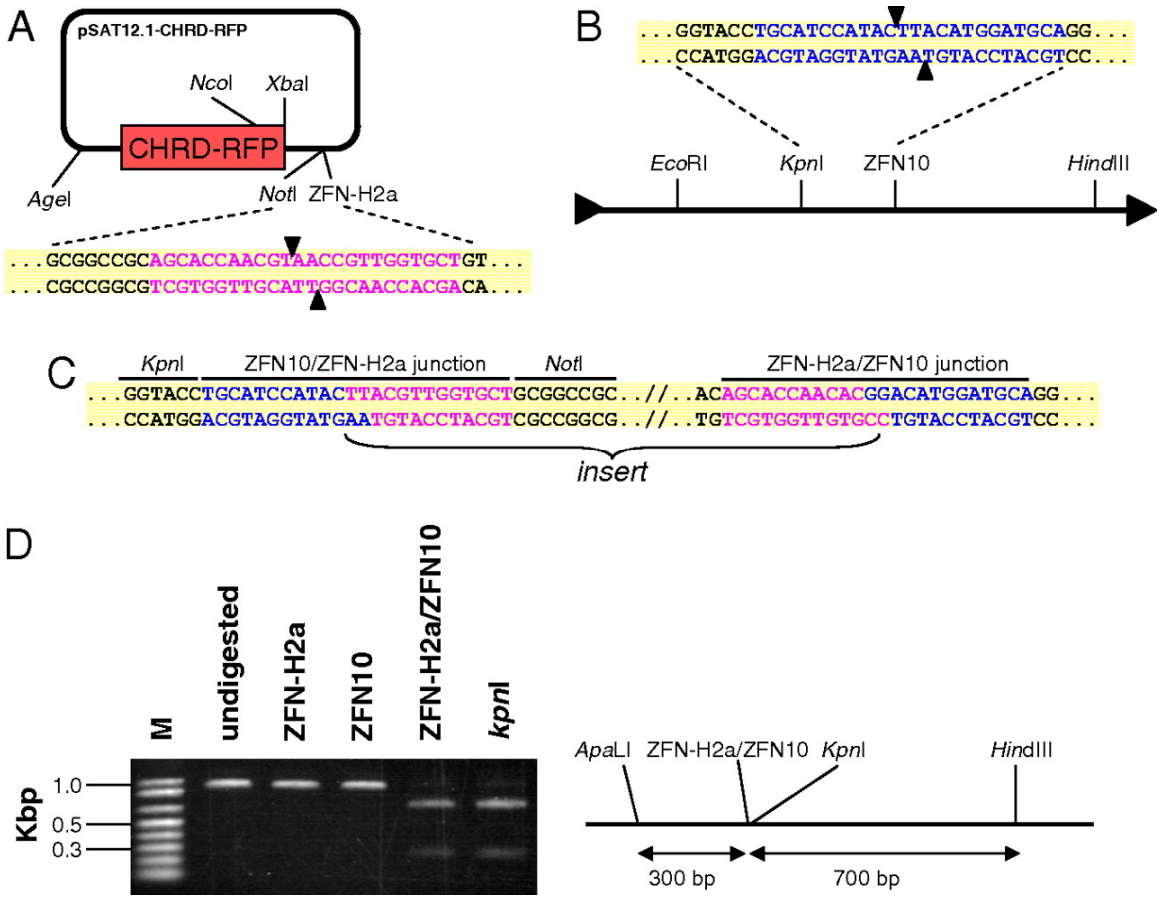


Figure 3.4. DNA cloning using compatible ZFNs. A. Scheme of pSAT12.1-CHRD-RFP. The ZFN-H2a recognition-site sequence is in purple, and the predicted cleavage sites are indicated by arrowheads. B. Scheme of the pRCS11 acceptor plasmid. The ZFN10 recognition site in the plasmid's MCS is in blue. C. Sequence analysis of the ZFN10/ZFN-H2a junctions at the ligation sites in pRCS11[10/12.1-CHRD-RFP]. The ZFN10 recognition sequences derived from the acceptor plasmid pRCS11 are in blue, and the ZFN-H2a sequences derived from pSAT12.1-CHRD-RFP are in purple. D. Restriction analysis of an ≈1-kb-long *Apa*LI-*Hin*dIII fragment from pRCS11[10/12.1-CHRD-RFP]. The fragment's restriction map, which includes a ZFN-H2a/ZFN10 ligation junction, is on the right.

Chapter 4. A toolbox and procedural notes for characterizing novel zinc finger nucleases for genome editing in plant cells

4.1 Summary

Induction of double-strand breaks (DSBs) in plant genomes can lead to increased homologous recombination or site-specific mutagenesis at the repair site. This phenomenon can potentially be used for gene targeting applications in plant cells upon the induction of site-specific genomic DSBs using zinc finger nucleases (ZFNs). ZFNs are artificial restriction enzyme, custom-designed to cleave a specific DNA sequence. Tools and methods for ZFN assembly and validation could potentially boost their application for plant gene targeting. Here we report on the design of biochemical and *in-planta* methods for analyses of newly designed ZFNs. Cloning begins with *de-novo* assembly of the DNA-binding regions of new ZFNs from overlapping oligonucleotides containing modified helices responsible for DNA-triplet recognition, and the fusion of the DNA-binding domain with a *Fok*I endonuclease domain in a dedicated plant expression cassette. Following the transfer of fully assembled ZFNs into *Escherichia coli* expression vectors, bacterial lysates were found to be most suitable for *in-vitro* digestion analysis of palindromic target sequences. A set of three *in-planta* activity assays were also developed to confirm the nucleic activity of ZFNs in plant cells. The assays are based on the reconstruction of GUS expression following transient or stable delivery of a mutated *uidA* and ZFN-expressing cassettes into target plants cells. Our tools and assays offer cloning flexibility and simple assembly of tested ZFNs and their corresponding target sites into *Agrobacterium* binary plasmids, allowing efficient implementation of ZFN-validation assays *in planta*.

4.2 Introduction

The targeting and modification of native genomic sequences (gene targeting, GT) in yeast and animal cells is often made possible by homologous recombination (HR) between the donor and chromosomal DNA (108). Several attempts have been made, with only little reported success, to introduce efficient HR-based GT methods for use in plant species. These include, for example, using a strong positive- and negative-selection scheme (9), or overexpressing a HR-related yeast protein in *Arabidopsis* plants (10). It seems that the dominance of non-homologous end joining (NHEJ) over HR in plant DNA repair, and by implication the integration of foreign DNA into the plant genome, does not permit efficient HR between donor and chromosomal DNA (3,109,110). Methods are therefore needed to increase the rate of HR or interfere with the random foreign DNA integration process in plants.

Enhanced HR can potentially be achieved in plant cells by creating genomic DSBs at the target site (111). Thus, for example, expression of I-*Sce*I, a rare-cutting restriction enzyme, has been shown to lead to a significant increase in HR-mediated GT in tobacco plants (20,60), and expression of HO endonuclease has been shown to enhance intrachromosomal recombination in *Arabidopsis* plants (112). Expression of rare-cutters can also lead to site-specific integration of foreign DNA molecules and to induction of site-specific mutagenesis. For example, expression of I-*Sce*I and I-*Ceu*I led to deletions, insertions and targeted integration of foreign DNA molecules in the break sites of tobacco plants (21-23). The potential of using rare-cutters to increase HR rate or as site-specific mutagens is thus very clear (101,113). Nevertheless, their use is limited to rarely occurring natural recognition sites or to artificial target loci. Furthermore, the ability to

re-engineer rare-cutters for novel DNA-target specificities is both difficult and tedious (101) which limits their use for GT experiments. One possible route to overcoming these technological challenges is the use of zinc finger nucleases (ZFNs) - synthetic restriction enzymes which can be specifically designed to cleave virtually any long stretch of double-stranded DNA sequence (for recent reviews see Refs. 26,87,94). Indeed, expression of ZFNs in living cells has been shown to produce genomic DSBs, leading to enhanced HR, site-specific mutagenesis and targeted gene addition in various species (e.g. Refs. 43,49,75,114-116). Two independent reports have demonstrated the use of ZFNs for genome modifications in plant cells. In one, ZFNs were used to increase the frequency of HR in tobacco plants (52) and in the other, ZFNs were used to induce site-specific mutagenesis in *Arabidopsis* plants (51).

The rules and protocols for the assembly of new ZFNs have been the subject of several excellent recent reports (e.g. Refs. 32,33,94), and various tools and vectors for the design and assembly of novel ZFNs are available from Carlos Barbas laboratory (www.zincfingertools.org) and the Zinc Finger (ZF) Consortium (http://www.zincfingers.org). It should be noted, however, that while we can potentially design ZFNs to target virtually any genomic sequence, caution needs to be exercised in their construction based solely on computational methods and their subsequent use for targeting experiments in plant cells (and any other species for that matter). Ramirez et al. (36), for example, recently reported an unexpected failure rate for the modular assembly of new engineered zinc finger domains and pointed to some of the experimental obstacles that can hinder the assembly and evaluation of novel ZFNs. Thus, the application of ZFN technology for GT experiments requires not only the design and assembly of novel ZFNs,

but also proper validation of their activity in living cells. More specifically, it is important not only to verify that newly designed ZFNs can not only bind and digest *in vitro* target sequences, but can also digest target sequences *in vivo* (Cathomen et al., 2008; Porteus, 2008).

Realizing the importance of analyzing and testing the binding and digestion activities of new ZFNs, several groups have developed assays for their analysis. These include, for example, a bacterial cell-based reporter assay that analyzes the DNA-binding activities of zinc finger domains (32), a cell culture-based system suitable for analyzing the digestive activity of newly designed ZFNs in mammalian cells (117) and a digestion assay for *in-vitro* analysis of ZFN digestion activity (33). While these assays can facilitate the assembly and analysis of ZFNs for *in-vitro* use and for GT in mammalian cells and other model organisms (e.g. Refs. 69,117,118), no specialized vectors, tools or protocols have been described to assist with the development and validation of ZFNs in plant species.

Here we describe a set of assays and vectors suitable for the assembly, expression and functional analysis of novel ZFNs in plant cells. In addition to an *in-vitro* digestion assay, we designed three *in-planta* digestion assays which can be used for functional analysis of newly assembled ZFNs on transiently delivered and chromosomally integrated target DNA molecules. The flexibility of our vector design, which allows easy assembly and transfer of ZFN coding sequences between bacterial and plant expression vectors, the ability to clone ZFN expression cassettes onto *Agrobacterium* binary vectors and the simplicity of our *in-planta* digestion assays which are based on reconstruction of

the GUS reporter gene in living cells, provide the user with a useful collection of plasmids and assays for the development of ZFNs for plant research and biotechnology.

4.3 Experimental procedures

4.3.1 DNA constructs

Standard DNA amplification and cloning methods were used during the construction of all our vectors, unless noted otherwise. To produce a plant ZFP cloning vector that carries a NLS and the *Fok*I endonuclease domain, we first cloned the NLS coding sequence by annealing the

5’CATGCCATGGTGCCAAAAAAGAAGAGAAAGGTAGAAGACCCCTCTCGAGC GG and

5’CCGCTCGAGAGGGGTCTTCTACCTTTCTCTTCTTTTTTGGCACCATGGCATG primers and cloning the DNA product as an *Nco*I-*Xho*I fragment into pSAT6-MCS (41), producing pSAT6.35SP.NLS-MCS. We next PCR-amplified the *Fok*I endonuclease domain from pHS::QQR-QEQ/2300 (51) using the

5’ACGCGTCGACGGACTAGTCAAAAGTGAACTGG and

5’GAAGATCTTTAGGATCCAAAGTTTATCTCGCCGTTATTAAA primers and cloned it as a *Sal*I-*Bgl*II fragment into the *Sal*I and *Bam*HI sites of pSAT6.35SP.NLS-MCS, producing pSAT6.35SP.NLS-*Fok*I. Finally, we transferred the NLS-*Fok*I expression cassette as an *Age*I-*Not*I fragment from pSAT6.35SP.NLS-*Fok*I into pSAT4-MCS (41), producing pSAT4.35SP.NLS-*Fok*I. To produce pET28.XH we first annealed the 5’AACTTTGGATCCCTCGAGTAGGACTGCAGGCTTGCGGCCGCACTCGA and 5’TCGAGTGCGGCCGCAAGCCTGCAGTCCTACTCGAGGGATCCAAAGTT primers and cloned the DNA product as a *Bam*HI-*Not*I fragment into the pET28c(+)

bacterial expression vector (Invitrogen) producing the intermediate plasmid pET28.XX. We next removed the spacer between the *BamH*I site and the 6xHis coding sequence by cleaving pET28.XX at two *Xho*I sites and self-ligating it. The resultant pET28.XH contained unique *Nco*I and *Bam*HI sites allowing the assembly of a ZFN::6xHis fusion in a single cloning step.

To produce a constitutive plant expression cassette for the ZFN QQR, we PCR-amplified the QQR coding sequence from pHS::QQR-QEQ/2300 and cloned it as an *Nco*I-*Bam*HI fragment into pSAT4.35SP.NLS-*Fok*I, producing pSAT4.35SP.QQR. To produce a heat-shock-inducible plant expression cassette for QQR, we PCR-amplified the *HSP18.2* promoter sequence from pHS::QQR-QEQ/2300 and cloned it into the *Age*I-*Nco*I sites of pSAT4.35SP.QQR, replacing the tandem CaMV 35S promoter with an *HSP18.2* promoter, and producing pSAT4.hspP.QQR. To produce a constitutive plant expression cassette for ZFN3, the ZFP was first assembled from a set of overlapping backbone oligonucleotides (BBO primers:
5'GAAAAACCTTACAAGTGTCCTGAATGTGGAAAGTCTTTTTCT,
5'CAGCGAACACACACAGGTGAGAAGCCATATAAATGCCCAGAATGTGGTAA
ATCATTCAG and
5'CAACGGACCCACACCGGGGAGAAGCCATTTAAATGCCCTGAGTGCGGGAA
GAGTTTTT) and sequence-dependent oligonucleotides (SDO primers:
5'ACCTGTGTGTGTTCGCTGGTGACGAGTAAGATCAGAAGACTGAGAAAAAGA
CTTTCCACA,
5'CCCGGTGTGGGTCCGTTGGTGACGAACAAGATGTCCAGAAGTACTGAATGA
TTTACCACA and

5'TCCAGTATGAGTACGTTGATGACGAACAAGAGATCCAGAAGTTGAAAAACT CTTCCCGCAC) as previously described (33) using high-fidelity *Pfu* DNA polymerase (Stratagene). The PCR mixture, consisting of 5 pM each of BBO and SDO primers and 200 pM each of 5'CCGCTCGAGCTGAAAAACCTTACAAGTGTCC3' and 5'GGACTAGTCCTCCAGTATGAGTACGTTGATG3' primers, was carried out for 35 cycles. The PCR product was cloned into the *Xho*I and *Spe*I sites of pSAT4.NLS-*Fok*I, producing pSAT4.35SP.ZFN3.

For expression of ZFN in *E. coli* cells, the QQR and ZFN3 coding sequences were transferred from pSAT4.35SP.QQR and pSAT4.35SP.ZFN3 as *Nco*I-*Bam*HI fragments and cloned into pET28.XH, producing pET28.XH-QQR and pET28.XH-ZFN3, respectively. The ZFN3 target plasmid pSAT6.ZFN3-TS was constructed by annealing the 5'CGATAGCCATGGAGCACCAACACCAGTTGGTGCTCTGCAGTCGACG and 5'CGTCGACTGCAGAGCACCAACTGGTGTTGGTGCTCCATGGCTATCG primers and cloning the DNA product as an *Nco*I-*Pst*I fragment into pSAT6-MCS. To produce mutated GUS-expression vectors carrying the QQR zinc finger recognition sites, the *uidA* coding sequence was PCR-amplified from pRTL2-GUS using the forward 5'GGGGTACCATGTTCTTCCCCTCCTGAGGGGAAGAATTACGTCCTGTAGAAA CCCC and reverse 5'CGGGGTACCATGTTACGTCCTGTAGAAACCCC primers and cloning the PCR product into the *Kpn*I-*Bam*HI sites of pSAT6A-MCS (119), producing pSAT6A.QQR-TS*::GUS. A similar strategy was used to construct pSAT6A.ZFN3-TS*::GUS, using the forward 5' GGGGTACCATGAGCACCAACTCCTGAGTTGGTGCTTTACGTCCTGTAGAAAC CCC and the above-mentioned reverse primer.

The binary vectors pRCS2.[ZFN3-TS*::GUS] and pRCS2.[QQR-TS*::GUS] were constructed by transferring the ZFN3-TS*::GUS and QQR-TS*::GUS expression cassettes from pSAT6A.ZFN3-TS*::GUS and pSAT6A.QQR-TS*::GUS, respectively, as PI-*Psp*I fragments into pRCS2 (41). The binary vectors pRCS2.[ZFN3][ZFN3-TS*::GUS] and pRCS2.[QQR][QQR-TS*::GUS] were constructed by transferring the ZFN3 and QQR constitutive expression cassettes from pSAT4.35SP.ZFN3 and pSAT4.35SP.QQR as I-*Sce*I fragments into pRCS2.[ZFN3-TS*::GUS] and pRCS2.[QQR-TS*::GUS], respectively. The binary vectors pRCS2.[KAN][ZFN3-TS*::GUS] and pRCS2.[KAN][QQR-TS*::GUS] were constructed by transferring the ZFN3-TS*::GUS and QQR-TS*::GUS expression cassettes from pSAT6A.ZFN3-TS*::GUS and pSAT6A.QQR-TS*::GUS, respectively, as PI-*Psp*I fragments into pRCS2-nptII (119). The binary vectors pRCS2.[ZFN3] and pRCS2.[QQR] were produced by removing the ZFN3-TS*::GUS and QQR-TS*::GUS expression cassettes from pRCS2.[ZFN3][ZFN3-TS*::GUS] and pRCS2.[QQR][QQR-TS*::GUS], respectively. The pRCS2.[KAN][hspP.QQR][QQR-TS*::GUS] binary vector was constructed by cloning the heat-shock-induced QQR expression cassette from pSAT4.hspP.QQR into pRCS2.[QQR-TS*::GUS].

4.3.2 ZFN protein expression and in-vitro digestion assay

For *E. coli* expression, ZFN expression plasmids (pET28.XH-ZFN3 and pET28.XH-QQR) were transformed into BL21 GOLD (DE3) PlyS cells (Stratagene). The cells were cultured in 100 ml of LB medium supplemented with 50 μg/ml kanamycin and 100 μM $ZnCl_2$ and grown at 22^{o}C. At an OD_{600} of 0.6, ZFN expression was induced by 0.7 mM IPTG for 3 h at 22^{o}C. Cells were harvested by centrifugation, resuspended in 35 ml of 25

mM Tris-HCl pH 7.5, 300 mM NaCl, 5% (v/v) glycerol and 100 μM $ZnCl_2$ and lysed twice using a French Press. The cell lysates were used directly for target-site plasmid cutting in the biochemical assay. Digestion of the target-site plasmids or DNA fragments was carried out in NEBuffer 4 (New England Biolabs) for 30 min at 37°C using 0.25-2.5 μl of cell lysates. Double-digestion with ZFN3 and *Age*I was performed by combining both enzymes in the same reaction mix.

4.3.3 Plant transformation

Binary vectors were transferred into *Agrobacterium tumefaciens* strain EHA105. For the T-DNA repair assay, *Agrobacterium* cells were grown overnight at 28°C in LB medium supplemented with spectinomycin and streptomycin (200 μg/ml each), collected by centrifugation and resuspended in an induction medium (10.5 g/L K_2HPO_4, 4.5 g/L KH_2PO_4, 1 g/L $(NH_4)_2PO_4$, 0.5 g/L Na-citrate, 1 g/L glucose, 1 g/L fructose, 4 g/L glycerol, 120 mg/L $MgSO_4$, 2 g/L MES, pH 5.6) supplemented with 100 μM acetosyringone and antibiotics, grown to an OD_{600} of 0.6, again collected by centrifugation and resuspended in an infiltration medium (10 mM $MgSO_4$, 10 mM MES, pH 5.6) supplemented with 200 μM acetosyringone. The *Agrobacterium* infiltration method (120) and detached tobacco leaves were used for the delivery of T-DNA molecules for the T-DNA repair assay. The leaf-disk transformation method (121) with only a callus-induction medium instead of a shoot-regeneration medium was used for the production of transgenic calli for the *in-planta* T-DNA repair assay using the pRCS2.[KAN][ZFN3-TS*::GUS] or pRCS2.[KAN][QQR-TS*::GUS] vectors. For activation of the *uidA* gene in the *in-planta* T-DNA repair assay, transgenic calli were re-infected with *Agrobacterium* cells carrying the appropriate binary vectors (i.e.

pRCS2.[ZFN3] or pRCS2.[QQR]) for 48 h at 25°C on a hormone-free MS medium. For the whole-plant DNA repair assay, *Arabidopsis* plants were transformed using the flower-dip transformation method (122) and transgenic plants, tissues and/or seedlings were heat-shocked at 42°C for 30 min. Infected leaves, transgenic calli, and heat-shocked leaves and seedlings were stained with chromatogenic substrate X-Gluc as described previously (123).

4.3.4 Molecular analysis of mutagenized transgenic Arabidopsis plants

Total plant DNA was extracted from leaves of heat-shocked transgenic *Arabidopsis* plants using the phenol-chloroform method. Total DNA was digested with *Dde*I for 1 h and the region surrounding the ZFN target site was PCR-amplified using primers 5'CATGCCATGGCTATCCTTCGCAAGACCCTTCC and 5'GACTAGTGAGCATTACGCTGCGATG. The resulting PCR fragment was re-digested with *Dde*I, and its undigested fraction was re-amplified and cloned as an *Nco*I-*Spe*I fragment into pSAT5-MCS (41). Randomly selected colonies were then selected and sequenced.

4.4 Results

4.4.1 Outline of ZFN functional assays and vector systems

To facilitate the analysis of newly assembled ZFNs, we designed a battery of assays which allow functional analysis of ZFN monomers' activity *in vitro* and *in planta*. **Figure 4.1** outlines the set of assays for ZFN activity and illustrates the main vector components required for each assay. Each assay is composed of a ZFN expression cassette and a target sequence on which the ZFN activity is to be tested. Since each assay was designed

for the analysis of ZFN activity under a different setting, we decided to develop the different ZFN expression plasmids with maximal flexibility, allowing the developer of novel ZFNs easy and simple transfer of ZFN components between the various vectors. **Figure 4.2** illustrates the structures and generic names of the ZFN expression vectors, their derivatives and other components of our ZFN assembly and functional assays. In line with other reports which describe the development of ZFN for animal and human cell applications (e.g., Cathomen et al., 2008; Porteus, 2008), we propose that custom-designed ZFNs be assayed not just for their *in-vitro* digestion activity but also for their ability to target stably integrated genomic sequences, before considering them for experiments involving the targeting of native sequences.

4.4.2 ZFN plant expression vectors and assembly of ZFNs

The rules and protocols for the assembly of custom-designed ZFNs have been described in several papers (e.g. Refs. 32,33) and we adapted and slightly modified the protocol described by Mani et al. (33) for the *de-novo* synthesis of zinc finger proteins (ZFPs). We constructed the pSAT4.35SP.NLS-*Fok*I vector to facilitate the assembly of a new ZFN monomer into a plant expression vector using a single cloning step. The pSAT4.35SP.NLS-*Fok*I vector's key features are the presence of the *Fok*I endonuclease domain, a nuclear localization signal (NLS) and the *Xho*I and *Spe*I recognition sites in the multi cloning site (MCS) located between the NLS and *Fok*I (**Figure 4.2**, ZFP cloning). The zinc finger DNA-binding domain can be cloned into the *Spe*I and *Xho*I sites following its assembly using a combination of overlapping oligonucleotides and Klenow and PCRs, or by PCR amplification and transfer of a ZFP domain from existing ZFNs . This cloning strategy results in the construction of the pSAT4.35SP.ZFN plant expression

vector in which the expression of a newly assembled ZFN is controlled by the CaMV 35S tandem promoter (see generic plasmid pSAT4.35SP.NLS-*Fok*I and its descendent pSAT4.35SP.ZFN vector in **Figure 4.2**). The 35S promoter can also be easily replaced using virtually any other promoter (e.g. the heat-shock promoter, hspP) in a single cloning step using *Age*I and *Nco*I (**Figure 4.2**, plant ZFN expression) and the entire plant expression cassette can be mounted onto pRCS-based *Agrobacterium* binary vectors using rare cutters (**Figure 4.2**, binary plant transformation) for transformation experiments in plant cells.

We used pSAT4.35SP.NLS-*Fok*I for the construction of pSAT4.35SP.QQR and pSAT4.35SP.ZFN3, two plant expression cassettes carrying the QQR (43) and ZFN3 (**Figure 4.3**) zinc finger nucleases. The QQR is a well-defined ZFN which has been used in various studies (e.g. Refs. 43,51) and the ZFN3 is a custom-made enzyme which we designed to bind the 9-bp-long sequence 5'AGCACCAAC (**Figure 4.3A**). We used these enzymes during the development of our *in-vitro* and *in-planta* digestion assays.

4.4.3 In-vitro ZFN digestion assay

Previous studies have emphasized the importance of analyzing the *in-vitro* cleavage activity of custom-designed ZFNs prior to their use for *in-vivo* applications (26,32,33,117). Mani et al. (33), for example, described a protocol for the rapid characterization of ZFNs for their sequence-specific cleavage activity and constructed a set of vectors to allow the conversion of assembled ZFPs to ZFNs and for the construction of target sequences in pUC18. While these protocols and plasmids can be very useful for *in-vitro* analysis of any given ZFN, we decided to produce our own bacterial expression vector which will be compatible with our plant vectors. We also

propose a simple strategy for cloning ZFN recognition sequences into pSAT-based vectors to produce target plasmids (e.g. pSAT6.ZFN-TS, **Figure 4.1**), useful for simple *in-vitro* digestion analysis. We constructed the bacterial expression vector pET28.XH-ZFN to carry unique *Nco*I and *Bam*HI sites between the T7 promoter and a 6xHis tag, to allow transfer of the entire ZFN from pSAT4.35SP.ZFN plasmids in a single cloning step (**Figure 4.2**). To demonstrate our *in-vitro* assay, we first transferred the QQR and ZFN3 from pSAT4.35SP.QQR and pSAT4.35SP.ZFN3, producing pET28.XH-QQR and pET28.XH-ZFN3, respectively. We next transferred these vectors into BL21 *E. coli* cells for overexpression of their encoded enzymes and we used crude bacterial lysate (**Figure 4.3B**) to assay their ability to digest their corresponding recognition sites on target plasmids. The digestion activity of QQR, expressed from our pET28.XH-QQR plasmid, was demonstrated by its ability to digest its semi-palindromic target site which was engineered on the HS::QQR-QEQ/2300 plasmid (51). We released a 3.4-kb QEQ containing a *Not*I fragment from pHS::QQR-QEQ/2300, and its digestion with QQR released the expected 2.1- and 1.3-kb DNA fragments (**Figure 4.3C**). The 3.4-kb *Not*I fragment also carried a unique *Kpn*I recognition site next to the QQR recognition site and its digestion with *Kpn*I produced a restriction pattern similar to that of QQR (**Figure 4.3C**), thus indicating that QQR indeed cleaved its corresponding site within the target fragment. It should be noted that while the digestion of the 3.4-kb *Not*I fragment by QQR was only partial, this does not provide conclusive evidence as to the specific activity of this enzyme since we did not purify it to the level of a commercial enzyme such as *Kpn*I. We further demonstrated the application of our *in-vitro* digestion assay by analyzing the digestion activity of ZFN3. To this end, we first constructed a target plasmid, designated

pSAT6.ZFN3-TS, by annealing two complementary primers coding for the ZFN3 semi-palindromic target site and cloning them between the adjacent *Nco*I and *Pst*I sites in pSAT6-MCS. The resultant pSAT6-MCS-based vector structure (**Figure 4.1**) allows for single-step digestion analysis of ZFN activity by comparing the *Age*I/ZFN and *Age*I/*Nco*I double-digestion patterns. Indeed, the digestion pattern of pSAT6.ZFN3-TS by *Age*I and ZFN3 was similar to that of *Age*I and *Nco*I (**Figure 4.3C**), indicating that the ZFN3 can cleave its corresponding site within its target plasmid. A similar cloning strategy of annealed primers can be used to assemble other ZFN target sites into pSAT6-MCS, allowing for simple *in-vitro* analysis of custom-designed ZFNs.

4.4.4 In-planta T-DNA repair assay

Confirming the *in-vitro* digestion activity of novel ZFNs is only the first step, albeit a crucial one, of their functional analysis *in vivo*. We designed a T-DNA repair assay in which the ZFN digestion activity can be analyzed by GUS staining in plant cells. The assay's first component is the plant ZFN expression cassette pSAT4.35SP.ZFN. The assay's second component is a pSAT5.ZFN-TS*::GUS plant expression cassette carrying a mutated *uidA* gene which is engineered to carry the TGA (stop) codon within the 6-bp spacer of the ZFN target site (**Figure 4.2**, plant GUS reporter repair), leading to premature termination of *uidA* translation in plant cells. Digestion of the ZFN recognition site and its successive repair can lead to deletion and/or mutation of the stop codon and consequent activation of the *uidA* reporter gene. Assembly of the T-DNA repair assay requires the construction of a binary plasmid carrying the constitutive expression cassettes from pSAT4.35SP.ZFN and pSAT5.ZFN-TS*::GUS (i.e. the generic plasmid pRCS2.[ZFN][ZFN-TS*::GUS], **Figure 4.1**). The pSAT4.35SP.ZFN vector is based on

the pSAT4 backbone plasmid that allows transferring the entire ZFN expression cassette onto pRCS-based binary vectors (e.g., pRCS2, **Figure 4.2** and Ref. 41) using I-*Sce*I. The pSAT5.ZFN-TS*::GUS vector that carries the mutated *uidA* gene (ZFN-TS*::GUS) under the control of the tandem 35S promoter is based on the pSAT5 backbone plasmid (41), allowing transfer of the ZFN-TS*::GUS expression cassette onto pRCS-based binary vectors using I-*Ceu*I. Mounting both expression plasmids onto a single binary vector ensures efficient delivery of the tested ZFN and its target sequence into the target plant cells and allows monitoring the functional activity of the ZFN by GUS staining.

We demonstrated the functionality of our *in-planta* T-DNA repair assay by analyzing the activity of ZFN3 and QQR ZFNs in tobacco cells. We constructed pSAT5.ZFN3-TS*::GUS and pSAT5.QQR-TS*::GUS and mounted them and their corresponding ZFN expression cassettes (*i.e.* pSAT4.35SP.ZFN3 and pSAT4.35SP.QQR, respectively) onto pRCS2 binary plasmids, producing pRCS2.[ZFN3][ZFN3-TS*::GUS] and pRCS2.[QQR][QQR-TS*::GUS], respectively. The general structure of pRCS2.[ZFN3][ZFN3-TS*::GUS] and the mutated *uidA* region which was engineered to contain the ZFN3 target site and the TGA codon is shown in **Figure 4.4A**. Also shown in **Figure 4.4A** is a putative outcome of ZFN3 digestion and misrepair of the DSBs by the plant's NHEJ machinery, which can potentially lead to elimination of the stop codon and reconstruction of the *uidA* reporter gene. Indeed, agroinfiltration of the pRCS2.[ZFN3][ZFN3-TS*::GUS] into *Nicotiana benthamiana* leaves resulted in positive GUS staining (**Figure 4.4B**), indicating the functional activity of the ZFN3 protein in plant cells. Agroinfiltration of pRCS2.[QQR][QQR-TS*::GUS] also resulted in positive GUS staining, although with higher intensity (**Figure 4.4C**), which may be attributed to

QQR's higher DNA-binding affinity relative to that of ZFN3. GUS expression was not observed following agroinfiltration of binary plasmids carrying the ZFN3-TS*::GUS (**Figure 4.4D**), or QQR-TS*::GUS (**Figure 4.4E**) expression cassettes without their corresponding ZFN expression cassettes, indicating that ZFN expression is required for digestion and activation of the *uidA* reporter gene from the *Agrobacterium*'s T-DNA, *in planta*.

4.4.5 In-planta transgene repair assay

The ability of ZFNs to digest T-DNA molecules in plant cells provides another level of confidence for their potential use in living cells. Nevertheless, we recommend testing the ZFN's ability to digest genomically embedded DNA sequences prior to their deployment for GT experiments, as suggested by other studies aimed at developing ZFNs for non-plant species (e.g., Cathomen et al., 2008; Porteus, 2008). To this end, we designed the transgene repair assay in which the ZFN's activity is tested on a mutated *uidA* expression cassette which has first been integrated into the genome of transgenic tobacco calli. The assay's first component is a binary transformation vector that carries the ZFN-TS*::GUS reporter expression cassette and a selectable marker (e.g. *nptII*) (**Figure 4.1**, generic plasmid pRCS2.[KAN][ZFN-TS*::GUS]). This vector can be used to produce transgenic calli which lack functional *uidA* translation. The second component is a binary vector, carrying the ZFN expression cassette, driven by the tandem 35S promoter (**Figure 4.1**, generic plasmid, pRCS2.[ZFN]). This vector is used for infection of the transgenic calli and for digestion of the ZFN-TS*::GUS expression cassette in these tissues.

We demonstrated the functionality of our transgene repair assay by testing the activity of ZFN3 and QQR ZFNs in tobacco cells. We constructed pRCS2.[KAN][ZFN3-

TS*::GUS] and pRCS2.[KAN][QQR-TS*::GUS] and used them for the production of kanamycin-resistant transgenic tobacco calli. As expected, these transgenic calli did not express functional *uidA*, as indicated by their negative GUS staining (**Figure 4.5A** and **B**). We next constructed pRCS2.[ZFN3] and pRCS2.[QQR] and used them in wounding-infection experiments with the ZFN3-TS*::GUS- and QQR-TS*::GUS-transgenic calli, respectively, which resulted in positive GUS staining (**Figure 4.5C** and **D**), indicating the functionality of these ZFNs in digesting integrated DNA molecules

4.4.6 A whole-plant DNA repair assay

The *in-planta* transgene repair assay provides clear experimental evidence of the analyzed ZFN's digestion activity on genomically integrated DNA sequences. We next developed a whole-plant DNA repair assay as a tool for analyzing, at the molecular level, the outcome of their digestion and the repair of the digested DNA in transgenic plants. The assay calls for the assembly of a multi-gene plant transformation vector, which carries a plant selectable marker, the mutated *uidA* reporter gene cloned under the control of a constitutive promoter and the ZFN protein cloned under the control of an inducible promoter (**Figure 4.1**, pRCS2.[KAN][hspP.ZFN][ZFN-TS*::GUS] generic plasmid). We used the *Arabidopsis* heat-shock-inducible promoter *HSP18.2* (124), but other types of inducible promoters may be considered. The vector can be used to produce transgenic plants and to monitor ZFN activity in various tissues at different developmental stages. We used this assay to monitor the activity of QQR ZFN in different tissues of *Arabidopsis* transgenic plants and to molecularly characterize the outcome of various *uidA* digestion and repair events in these plants. To this end, we constructed pRCS2.[KAN][hspP.QQR][QQR-TS*::GUS], produced several kanamycin-resistant

Arabidopsis plants and used heat shock to induce the expression of the QQR enzyme in mature plants and in their offspring. Induction of the ZFN in fully developed mature transgenic leaves resulted in activation of the *uidA* reporter gene, as indicated by the appearance of GUS-expressing spots on the induced leaves (**Figure 4.6A**). Induction of ZFNs in transgenic seeds resulted in the appearance of GUS-expressing sectors (**Figure 4.6B**) and the production of chimeric tissues (e.g. **Figure 4.6C**). We allowed few of the heat-shocked transgenic seedlings to develop, stained them and found that they exhibited extensive GUS staining (e.g. **Figure 4.6D**). No GUS expression was observed in uninduced transgenic plants.

We further analyzed several samples from ZFN-induced transgenic seedlings by PCR amplification of the mutated QQR-TS*::GUS region. While repair of induced DSBs by restriction enzymes has been shown to create various site-specific mutations, including large deletions and/or insertions at the break site (21,51,125), we focused our analysis on identifying small, single-nucleotide deletions and/or substitutions in the mutated ORF which could lead to reconstruction of an active *uidA* gene. To this end, we first digested the induced plants' total DNA extract with *Dde*I in order to reduce the contamination of non-mutated CTGAG-sequence DNA. We next used PCR to amplify the undigested DNA, cloned the PCR products and analyzed them by DNA sequencing. **Figure 4.7** shows the sequencing results of some of these clones, which revealed the presence of nucleotide replacement and small deletions and/or additions, some of which can explain the reconstruction of a functional *uidA* gene. This analysis can thus be used to predict at least part of the possible outcome of ZFN-mediated GT in plant cells.

4.5 Discussion

ZFNs have been proven instrumental for genome editing in various species. Expression of model and tailor-made ZFNs in model organisms such as *Drosophila*, *Caenorhabditis*, *Xenopus*, *Danio* and human cell lines has been shown to lead to various outcomes of genome editing. These include site-specific mutagenesis, gene replacement by HR, and targeted gene addition at specific locations (e.g. Refs. 43,45-47,49,75,115,116,118). ZFNs have also been shown to enhance HR and to lead to site-specific mutagenesis in *Arabidopsis* and tobacco model plants (51,52) and more recently, Dow AgroSciences and Sangamo BioSciences announced the modification of maize and canola genomes by ZFNs (http://files.shareholder.com/downloads/SGMO/0x0x113031/d58602ac-aed0-4118-8955-e98e4004aa9d/SGMO_News_2007_6_19_General_Releases.pdf). Thus, ZFNs represent a novel and powerful technology which may eventually lead to the development of much needed GT methods for plant species (126).

The application of ZFN technology for the GT of native sequences requires overcoming several technological barriers. Proper selection of the binding site for a pair of ZFN monomers on the target sequence is the first step in the process of using this technology for genome editing in living cells. The rules for selecting target sites have been described in several excellent papers (34,35,117) and the developer of novel ZFNs can also use a collection of data sets describing the DNA-binding domains for 49 out of 64 possible triplets (35), as well as publicly available web-based programs (www.zincfingers.org and www.zincfinger-tools.org) to assist with the design of novel ZFNs. Once designed, the developer is faced not only with the challenge of assembling the ZFNs but also with validating their activity *in vitro* and *in vivo*, prior to their

deployment for genome editing experiments in their target organism. Various cloning methods can potentially be used for the assembly of custom ZFNs. The restriction-digest-based modular assembly, for example (32), is based on successive cloning of individual fingers from a collection of finger modules and the production of a multi-finger construct which can then be used for the assembly of custom-made ZFNs. In another strategy (117), several fingers are first independently amplified using a combination of general and finger-specific primers and then assembled together *via* a second overlapping PCR step. We adopted the strategy of *de-novo* assembly (33) of custom-designed ZFPs from a collection of overlapping backbone and sequence-dependent oligonucleotides (**Figure 4.2**, ZFP assembly). This strategy, which is based on annealing and ligating the assembled oligonucleotides followed by PCR amplification of the resultant fragment, allows the construction of novel ZFP domains in a single reaction. To facilitate the use of this strategy for the assembly of novel ZFNs for plant research, we constructed pSAT4.NLS-*Fok*I (**Figure 4.2**, ZFP cloning), which allows for single-step fusion of a ZFP with a NLS and the *Fok*I endonuclease domain, producing pSAT4.35SP.ZFN (**Figure 4.2**, plant ZFN expression). No less important was the construction of the pET28.XH bacterial expression vector (**Figure 4.2**, bacterial ZFP expression) which allows the simple transfer of ZFN coding sequences from pSAT4.NLS-*Fok*I. The pSAT4.35SP.ZFN and pET28.XH-ZFN expression vectors are the key plasmids for the analysis of newly assembled ZFNs.

Dedicated bacterial expression of ZFN vectors for *in-vitro* analysis of custom-designed ZFNs has been described in several reports. The pET15b:N vector, for example (33), was constructed for *in-vitro* transcription-translation of ZFNs and their validation in

conjunction with pUC18-based target sequences, and the pMAL-c2X-based bacterial expression vector was designed for MBP (maltose-binding protein) tagging of ZFNs (95). The latter enables protein purification on amylase resin and the use of purified ZFNs in *in-vitro* activity analysis on target DNA sequences. We found that cell lysates (**Figure 4.3B**) from ZFN-expressing *E. coli* cells using our modified pET28.XH vector were useful for *in-vitro* digestion analysis of target DNA molecules (**Figure 4.3C**). Thus, while the use of an *in-vitro* expression method and/or epitope-based extraction may produce higher protein concentrations and/or a more highly purified product, simple crude protein extraction may also be sufficient for analysis of newly assembled ZFNs. Furthermore, while *in-vitro* digestion analysis is typically performed by sequential digestion of target DNA molecules using ZFNs and commercial enzymes (e.g. Refs. 33,95), we found that double digestion can be performed in a single reaction, which greatly simplifies the *in-vitro* validation process. The qualitative nature of our in vitro digestion assay does allow quantifying ZFNs specific activity, yet it is sufficient for determining their functionally.

In-vitro digestion analysis is just the first step in validating ZFN activity for *in-vivo* applications. Two basic approaches have been developed to assay novel ZFNs in living cells. The first approach is based on analyzing the binding activity of the ZFNs' DNA-binding domain and it has been adapted, for example, by Wright et al. (32), who developed a two-hybrid reporter system for analysis of ZFP binding activity in bacterial cells. In their system, expression of the ZFP-Gal11P and alpha-Gal4 hybrid proteins led to activation of the *lacZ* reporter gene in bacterial cells upon binding of the ZFP with its target site, which was engineered upstream of the *lacZ* promoter. In another report, Meng

et al. (47) developed a one-hybrid activation system in which fusion of ZFP to RNA polymerase resulted in activation of the *HIS* and *URA* genes in bacterial cells. Both systems offer a robust, rapid and easily performed assay for analyzing ZFP-binding activity, but they do not allow for direct evaluation of the ZFN digestion activity in target cells. The second approach overcomes this limitation since it is based on analyzing ZFN digestibility, and not only binding, in living cells. One such system was recently described by Doyon et al. (46) who developed a yeast-based chromosomal reporter system in which the activity of novel ZFNs could be measured following the repair of a ZFN-mediated DSB in the yeast genome. Building on the notion that the digestion of tester sequences may provide more reliable information on the potential activity of newly assembled ZFNs in digesting native sequences in living cells, we adapted the second approach during the construction of our reporter ZFN-validation assays for plant cells.

ZFN-mediated DSBs of episomal or chromosomally embedded or encoded target sequences can potentially be repaired by HR or lead to enhanced HR at the break site. Nevertheless, genomic DSBs in many organisms, and particularly in plant species, are typically repaired by the cell's NHEJ machinery. Misrepair of broken edges can lead to deletions, insertions and/or substitutions and even capture of foreign DNA molecules at the repaired site. This wide array of possible outcomes could potentially hinder the efficiency of our reporter gene repair assay which can tolerate only short, in-frame deletions and/or alterations of a few nucleotides at the mutated *uidA* internal stop codon. Nevertheless, we have clearly shown that our *in-planta* repair assays are sensitive enough to detect proper activation of the mutated *uidA* gene. This was further supported by molecular analysis of several repair events produced by a whole-plant DNA repair assay

which revealed the reconstruction of a functional *uidA* gene (**Figure 4.7**).

Several key features of our plant-specific ZFN-validation assays make them useful for simple and robust analysis of custom-made ZFNs. First, our vector system is based on the structure of our previously described pSAT plant expression vector system which was specifically designed to facilitate the assembly of multi-gene expression cassettes (41). This feature is exemplified during the construction of the T-DNA repair assay where both the ZFN and the mutated GUS expression cassettes were mounted onto a single binary plasmid using a pair of rare cutters which was then used in a single transient transformation experiment (**Figure 4.4**). Second, our gene-activation assays are based on the *uidA* gene encoding the sensitive β-glucuronidase enzyme. This allows monitoring of ZFN activity using a simple biochemical staining procedure, which has been reported as useful in a wide range of plant species, tissues and cell types (123). We should also note that while the expression of ZFNs is driven by the tandem 35S promoter in our pSAT-based vectors, the promoter and terminator regions can easily be replaced by other regulatory elements (38,119) which may be more suitable for the ZFN developer's species of choice. This feature is exemplified during the conversion of pSAT4.35SP.QQR to pSAT4.hspP.QQR and the use of the latter plasmid for heat-shock induction of the QQR ZFN in transgenic *Arabidopsis* plants (**Figure 4.6**). Third, our transgene repair and whole-plant DNA repair assays, which call for the assembly of *Agrobacterium* stable transformation vectors and the recovery of transgenic tissues and/or whole plants (**Figure**s **5** and **6**), can support the use of different selection genes (**Figure 4.2** and Refs. 38,119), providing the user with greater flexibility in the choice of selectable marker. Fourth, our set of vectors provides the user with a variety of options for analyzing the

digestive activity of novel ZFNs. Thus, while we recommend a systematic analysis of novel ZFNs before their deployment for targeting experiments, the users of our system can choose to analyze their ZFNs at different levels of confidence, depending on their choice of assay.

In summary, we describe the construction and experimental application of a new set of assays and supporting plasmids for the validation of ZFN activity in plant cells. These complement a long line of assays and vectors which have been specifically designed to facilitate the assembly, analysis and use of ZFNs in non plant species. Given the simplicity of our assay system, its cloning flexibility and its compatibility with a much larger array of vectors carrying a large selection of promoters, terminators, selectable markers and reporter genes (38,41,119), our hope is that this system will facilitate experimentation with ZFN technology in plant species.

4.6 Figures

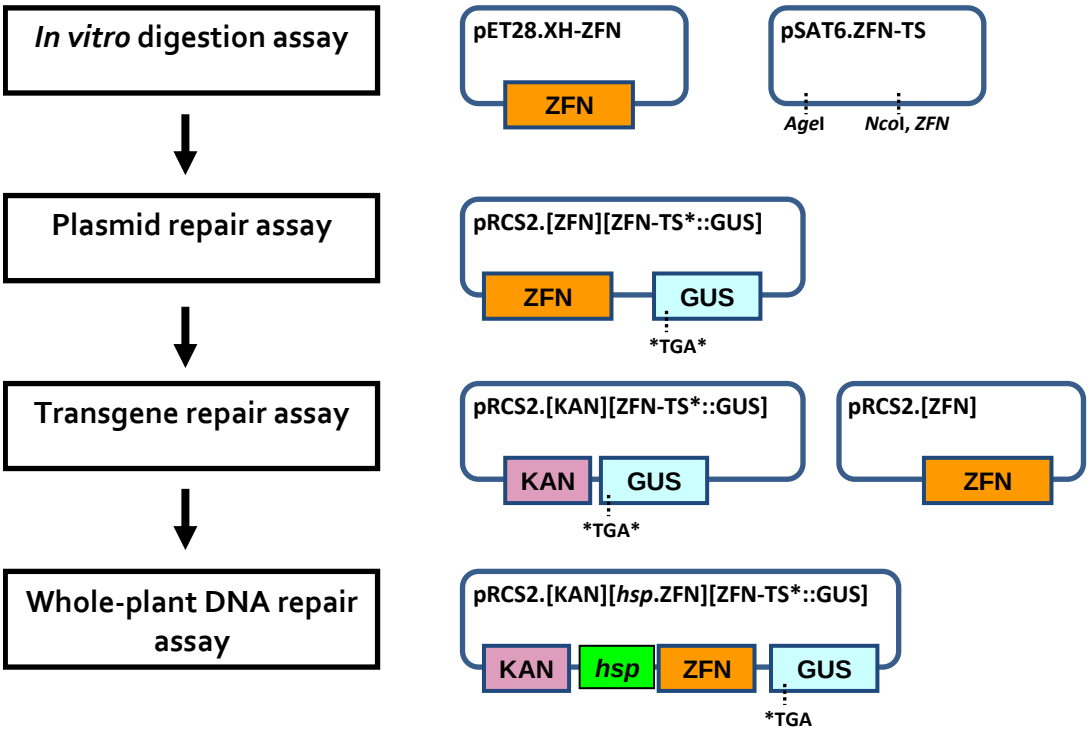


Figure 4.1. Outline of ZFN functional assays and their vector systems. The step-by-step comprehensive analysis of novel ZFNs is composed of four distinct assays (left panel), each based on monitoring ZFN activity using a defined set of vectors (right panel). The *in-vitro* digestion assay tests the digestion activity of an *E. coli*-expressed ZFN on its recognition site cloned on a target vector. The T-DNA repair assay requires the assembly of a dual-expression cassette on an *Agrobacterium* binary vector and tests the ability of a constitutively expressed ZFN to digest and repair a mutated GUS reporter gene cloned on the T-DNA region of that vector. The transgene repair assay calls for separating the ZFN expression cassette from its target site based on the activation of a mutated GUS reporter gene in transgenic calli, while the whole-plant repair assay activates the mutated GUS reporter gene in seedling or mature plant tissues upon specific activation of the ZFN. TS, ZFNs target site.

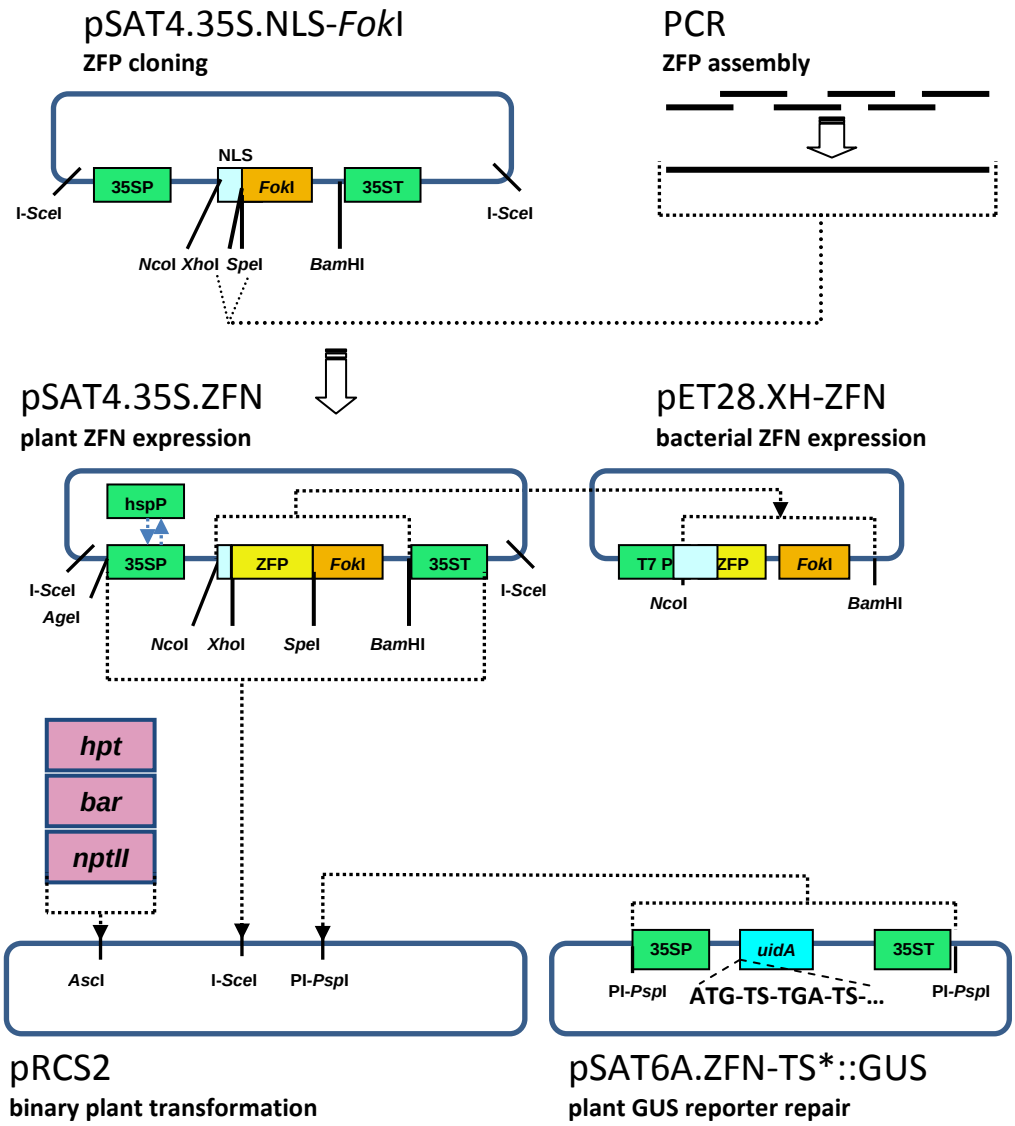
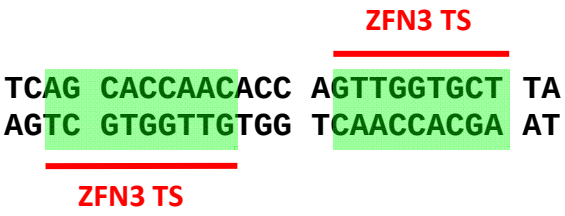


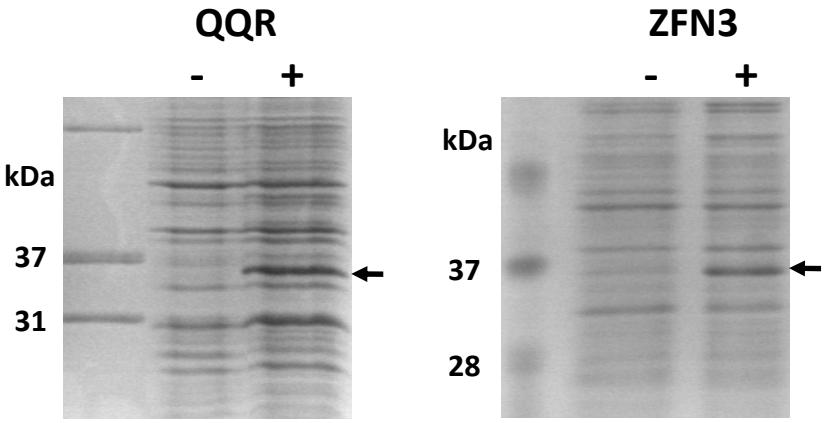
Figure 4.2. Structural features of the ZFN assembly and expression vector systems. A ZFP coding sequence can be assembled by Klenow/PCR using a combination of overlapping backbone and sequence-dependent oligonucleotides fused to the *Fok*I endonuclease domain in pSAT4.35SP.NLS-*Fok*I, producing the plant expression vector pSAT4.35SP.ZFN. The entire ZFN coding sequence can be transferred onto a pET28.XH-based vector producing a pET28.XH-ZFN vector, suitable for ZFN expression in bacterial cells and for *in-vivo* digestion assays. The pSAT4.35SP.ZFN can be modified by replacing the 35S constitutive promoter with a heat-shock-inducible promoter, producing a plasmid that is useful for the whole-plant DNA repair assay. A plant selectable marker, a ZFN and a mutated GUS reporter expression cassette can be mounted onto a pRCS2-based binary plant transformation vector using a combination of rare-cutting restriction enzymes and can then be used for various *in-planta* assays.

A



ZFP3 MVPKKKRKVEDPSRAE
KPYKCPECGKSFSQSSDLTRHQRTHTGE
KPYKCPECGKSFSTSGHLVRHQRTHTGE
KPFKCPECGKSFSTSGSLVRHQRTHTG…

B



C

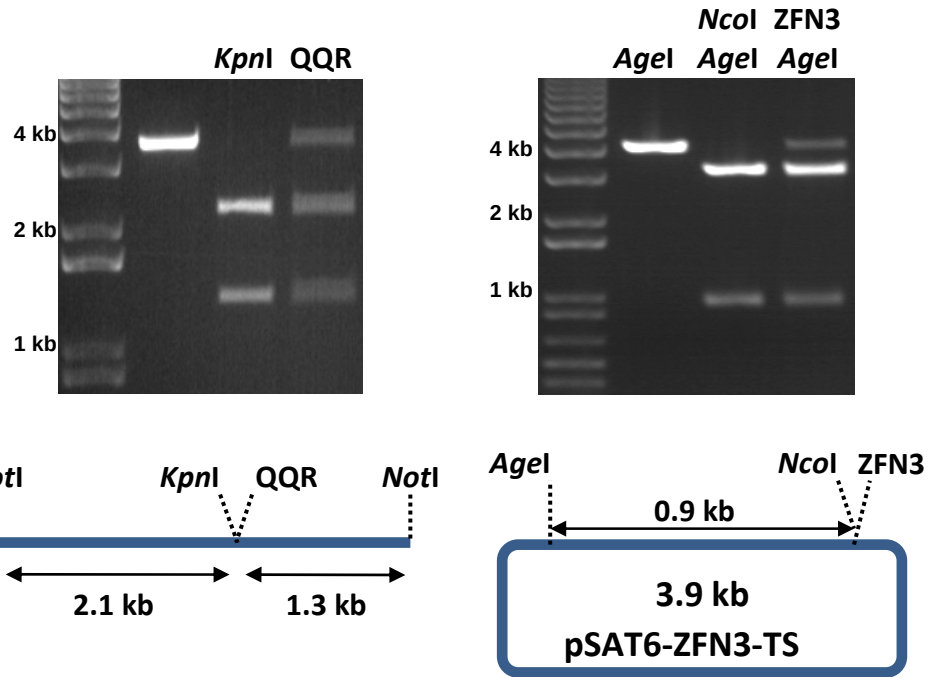


Figure 4.3. *In-vitro* digestion analysis of ZFNs. A. The structure of the ZFN3 24-bp palindrome-like recognition sequence and its corresponding ZFP. ZFN3 DNA-binding sites are highlighted in green and the unique amino-acid sequences for each zinc finger are given in blue. B. Separation of total crude extract from IPTG-induced (+) and non-induced (-) QQR and ZFN3 protein-expressing *E. coli* cells. C. Restriction analysis of a 3.4-bp *Not*I DNA fragment carrying the QQR palindrome-like recognition sequence (left) and the pSAT6.ZFN3-TS plasmid carrying the ZFN3 palindrome-like recognition sequence (right) by their corresponding ZFNs. Also shown are the schematic maps for each target DNA plasmid.

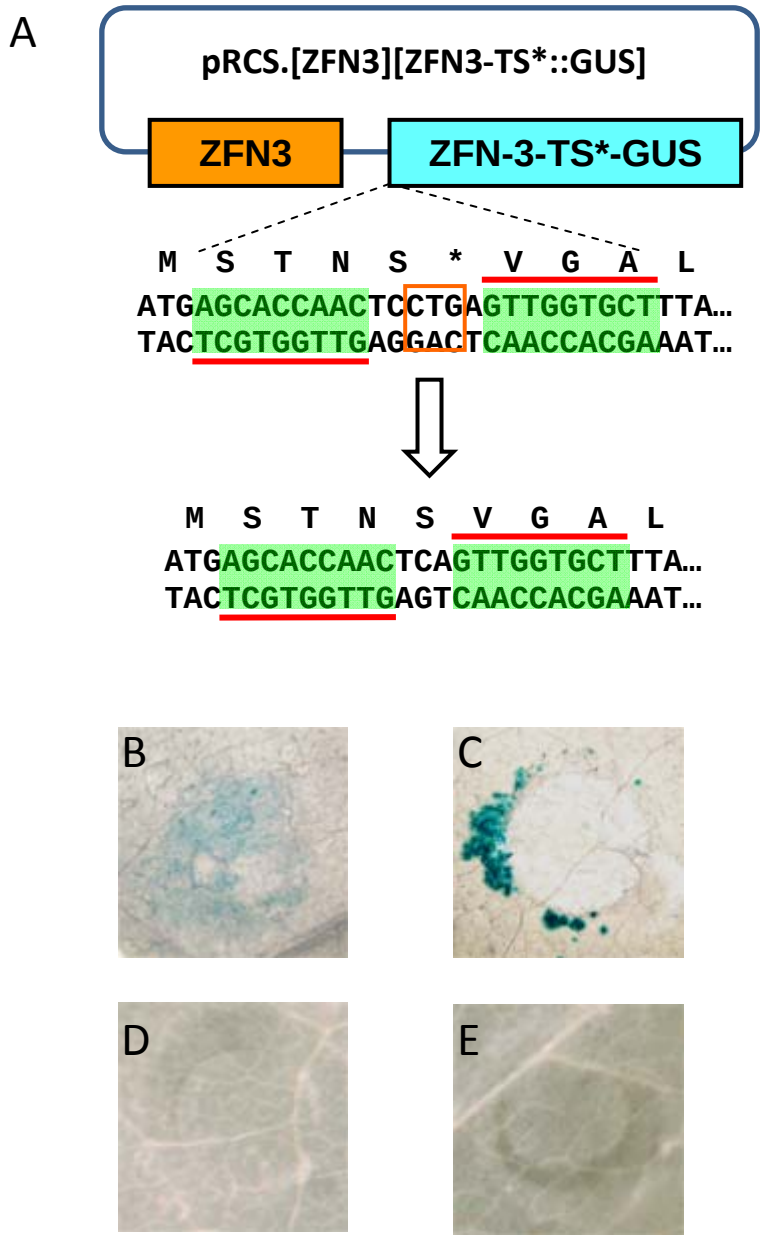


Figure 4.4. T-DNA repair assay in tobacco leaves. A. The structure of a T-DNA repair assay plasmid is exemplified by a binary plasmid constructed to test ZFN3 activity in plant cells. Reconstruction of a mutated GUS-encoding gene is exemplified by putative deletion of a CTG sequence (in brown box) which may occur following the digestion and misrepair of the ZFN3 recognition site *in planta*. B-E. Transformed tobacco leaves infected with QQR (B and D) or ZFN3 (C and E) mutated GUS expression cassettes, with (B and C) or without (D and E) their corresponding digesting ZFNs on the T-DNA molecule.

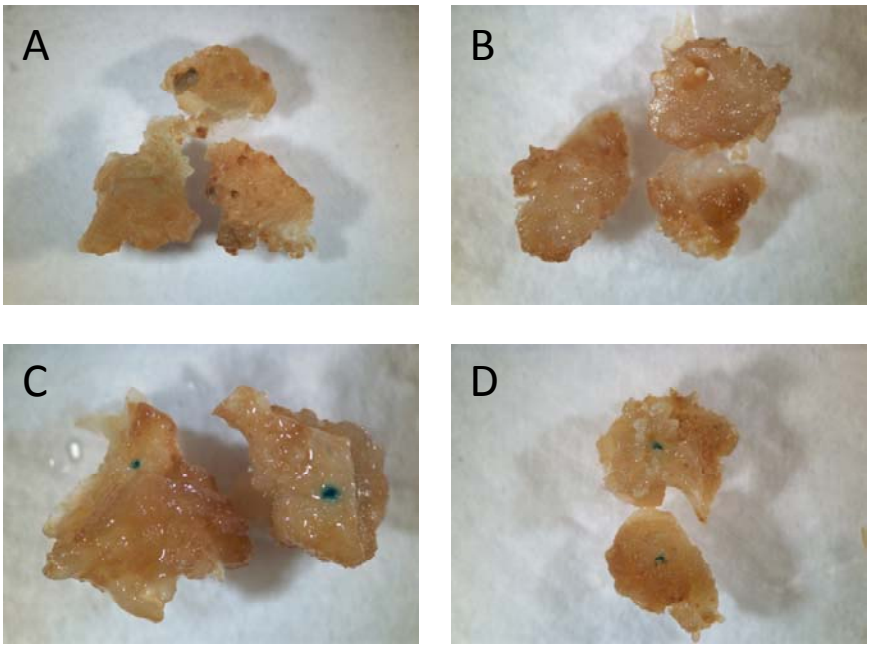


Figure 4.5. Transgene repair assay in tobacco calli. A,B. X-Gluc-stained kanamycin-resistant calli carrying mutated GUS-encoding sequences engineered with a QQR (A) or ZFN3 (B) target site. C,D. Detection of mutated GUS-encoding sequence activation by X-Gluc staining of transgenic calli engineered with a QQR (C) or ZFN3 (D) target site and inoculated with their corresponding ZFNs.

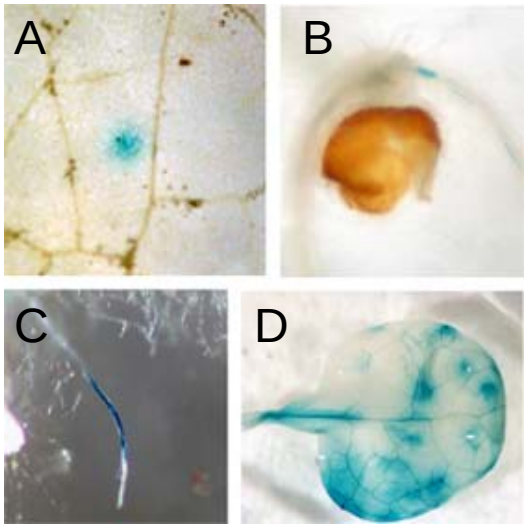


Figure 4.6. Whole-plant DNA repair assay in *Arabidopsis*. Detection of ZFN activity in mature leaves (A), germinating seedlings (B), elongating root (C) and developing leaves (D) of transgenic *Arabidopsis* plants engineered to carry a heat-shock-inducible QQR and its corresponding mutated GUS-encoding sequence.

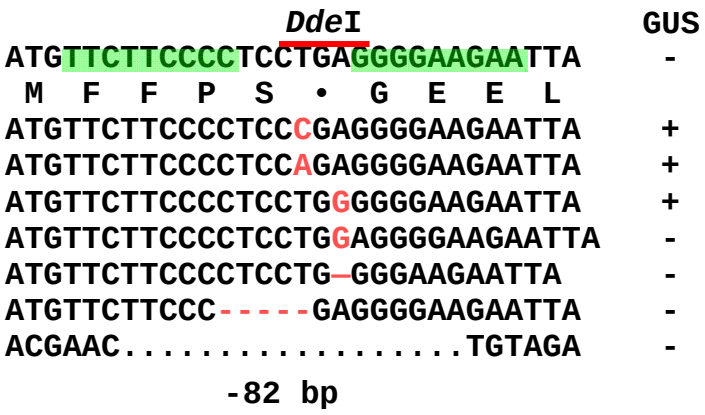


Figure 4.7. Molecular analysis of ZFN-induced mutations in transgenic *Arabidopsis* plants. The ZFN binding sites on the mutated GUS-encoding sequence's upper strands are highlighted in green. The location of the *Dde*I recognition site is indicated by a red bar. The outcome of positive (+) or negative (-) GUS expression in the different types of ZFN-mediated deletions or substitutions is indicated on the left.

Chapter 5. Concluding remarks

5.1 Summary

This dissertation project provides new insights into the biochemical properties of ZFNs, reveals their remarkable molecular cloning potential and lays the foundation for the widespread application of ZFNs in plant genome engineering. To achieve this project's goals, I applied a modular assembly approach to construct several novel ZFNs and subsequently used those ZFNs for a variety of in-vitro and in-vivo DNA engineering tasks.

I started by designing, assembling and purifying a ZFN, termed ZFN10, and characterizing its biochemical properties (Chapter 2). Ni-affinity and gel-filtration-column purification steps proved to be sufficient to obtain a homogeneous preparation of ZFN10 as evidenced by coomassie staining of an SDS-polyacrylamide gel, while attempts at digestion and subsequent self-ligation of pSAT10-MCS plasmid revealed that Ni purification was both necessary and sufficient to obtain a cloning-grade enzyme. One of the most interesting findings was the target-site flexibility of ZFN10: it was capable of tolerating a wide range of target-site substitutions, most of which were predicted from published specificity data on ZFN10 recognition helices.

Building upon my ZFN-purification protocols, I assembled and purified several additional enzymes that I used to demonstrate the molecular cloning properties of ZFNs (Chapter 3). I showed that ZFNs conveniently integrate into standard molecular cloning protocols and that they efficiently function in double-digestion reactions together with commercially available type II restriction endonucleases. Because of the low probability of occurrence of 18-bp target sites on long stretches of DNA, ZFNs can be also applied to

cloning large DNA fragments, such as gene-expression cassettes, which were successfully mounted on an *Agrobacterium* binary plasmid in my experiments. I demonstrated that a pair of specially designed ZFNs can be used for the truly unique task of custom-cloning a native DNA molecule.

I also developed a set of plasmids and methods for the construction and validation of ZFNs used in plant gene targeting research (Chapter 4). ZFP regions of ZFNs are assembled from a set of overlapping oligonucleotides and cloned into pSAT4.35S.NLS-*Fok*I plasmid, from where they can be conveniently transferred into pET28.XH-ZFN for in-vitro expression and testing, as well as into a binary pRCS2 plasmid for in-vivo activity-validation assays. The assays, which include a transient plasmid-repair assay, a transgene-repair assay and a whole-plant DNA-repair assay, also employ a modified GUS reporter, which can be repaired only in the presence of ZFN activity. My system of assays and accompanying plasmids creates a comprehensive toolbox for testing ZFNs prior to their application in plant gene targeting.

5.2 Open questions

5.2.1 Balance between target-site specificity and flexibility of ZFNs

ZFNs have gained much popularity in recent years as genome-engineering tools due to their modular structure, allowing for design flexibility, and the length of their target sites, which confers high specificity to ZFN-mediated DNA cleavage. Yet, as I show in Chapter 2, ZFN specificity is not absolute: the tested ZFN10 tolerated a wide range of substitutions in its target site. This fact should alert researchers planning their ZFN gene-targeting experiments and direct their attention towards selection of recognition helices with high triplet-binding specificity and towards more vigorous testing of novel ZFNs.

Fortunately, phage-display data on binding specificity are available for most of the known recognition helices (27-30) and, as can be seen from Chapter 2 and Figure 2.5, ZFN specificity for the most part correlates with individual recognition helix predictions. At the same time, incorporation of recognition helixes into a ZFN structure could potentially confer additional binding flexibility, as in the case of the recognition helix TSG-N-LVR (recognizing triplet GAT), which was highly specific in the phage display (29) (Figure 1.3) but tolerated position two and three substitutions in my experiments. I believe that additional studies encompassing a broader range of nucleotide positions within ZFN target sites and utilizing ZFNs with highly specific and flexible recognition helices could shed more light on the balance between target-site specificity and flexibility of ZFNs. Ultimately, I suggest using the more specific four-ZF ZFNs and an application-driven level of design and testing of novel ZFNs: from a routine check of nonspecific cutting with a large plasmid for ZFNs used in molecular cloning, to phage-display-based module selection methods (127) for ZFNs devised for human therapeutic applications.

5.2.2 Mutagenic effect of ZFNs on the plant genome

In light of the flexibility of certain ZFNs to their target sites, the question of ZFNs' effect on plant genomes is becoming more relevant. Reliable methods of assessing the mutagenic effect of ZFNs would facilitate public acceptance of transgenic crops created with ZFN technology, and allow for a more unequivocal interpretation of ZFN gene-targeting experiments in basic research. One such method, which employs antibody-mediated detection of proteins associated with sites of DNA damage to access the mutagenic effects of ZFNs, was used recently in mammalian cell culture (128) and could potentially be adopted for plant models.

Off-target DNA cleavage by ZFNs can be also looked into with another method, which applies to plants with well-characterized genomes, such as *Arabidopsis thaliana*, and involves bioinformatics genome screening for sites similar to a ZFN's binding site. In the case of the well-characterized model ZFN QQR, I identified 168 sites in the *Arabidopsis* genome that differ by three base pairs from the original QQR binding site. In-vitro analysis of some of them with purified QQR enzyme revealed five genomic locations that could potentially be cleaved by QQR (Tovkach and Tzfira, unpublished data). Similarly, ZFN3 (described in Chapter 4) has 104 sites with 3-bp differences, and at least six of them can be cleaved by purified ZFN3 enzyme (Tovkach and Tzfira, unpublished data).

On the other hand, one could capitalize on nonspecific ZFN cleavage of DNA by designing a ZFN with low target-site specificity and using it as a tool for genome mutagenesis. Such a mutagenesis tool would offer levels of predictability (due to the remaining ZFN specificity) and control (due to the controlled ZFN expression from inducible promoters) that would position it above the currently used radiation, chemical, or transposon-based techniques.

5.2.3 Targeting native genes in plants

Despite the four years that have passed since the publication describing mutagenesis of an artificial target site with QQR (51), ZFN-mediated gene targeting is far from being a routine procedure in plant species. Two documented instances of native gene modifications in plants (53,54) come from specialized research groups and serve as evidence of hard effort rather than the much anticipated revolution in plant genetics. With the wide availability of methods for ZFN design and construction and the recent

publication of a toolbox of methods for the characterization of novel ZFNs in plants (Chapter 4) (78), the last frontier most likely lies in providing ZFNs with access through chromatin structures en route to their genomic target sites.

Directed chromatin remodeling in *Arabidopsis* has been successfully performed previously by overexpression of the yeast *RAD54* gene—a member of the *SWI2*/*SNF2* chromatin-remodeling gene family (10). In that experiment, authors achieved a two orders of magnitude increase in the frequency of rare homologous recombination events, suggesting that chromatin remodeling by *RAD54*-like activity may be a useful tool in targeted plant-genome modification. Thus, coexpression of ZFNs with *yRAD54* and similar factors (*yRAD51*, *yRAD52*, *yRAD55*) may prove useful for ZFN-mediated targeting of native genes.

An alternative approach involves the design of ZFNs with very high affinity, which can potentially outperform chromatin elements in the competition for DNA binding. Recent data by Maeder and coworkers suggest that ZFNs with high affinities and specificities have a much better chance of modifying genomic targets (53). Although the design of such highly specific ZFNs may involve laborious selection techniques (as compared to the relatively simple modular assembly), the successful modification of a target gene may be worth the effort.

References

1. Griffiths, A. J. F. (2002) *Modern genetic analysis: integrating genes and genomes.* . New York: W.H. Freeman and Co.
2. Alonso, J. M. and Ecker, J. R. (2006) Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in Arabidopsis. *Nat Rev Genet* **7**, 524-536.
3. Britt, A. B. and May, G. D. (2003) Re-engineering plant gene targeting. *Trends Plant Sci* **8**, 90-95.
4. Scherer, S. and Davis, R. W. (1979) Replacement of chromosome segments with altered DNA sequences constructed in vitro. *Proc Natl Acad Sci U S A* **76**, 4951-4955.
5. Rothstein, R. (1991) Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. *Methods Enzymol* **194**, 281-301.
6. Evans, M. J., Carlton, M. B. and Russ, A. P. (1997) Gene trapping and functional genomics. *Trends Genet* **13**, 370-374.
7. Kempin, S. A., Liljegren, S. J., Block, L. M., Rounsley, S. D., Yanofsky, M. F. and Lam, E. (1997) Targeted disruption in Arabidopsis. *Nature* **389**, 802-803.
8. Hanin, M., Volrath, S., Bogucki, A., Briker, M., Ward, E. and Paszkowski, J. (2001) Gene targeting in Arabidopsis. *Plant J* **28**, 671-677.
9. Terada, R., Urawa, H., Inagaki, Y., Tsugane, K. and Iida, S. (2002) Efficient gene targeting by homologous recombination in rice. *Nat Biotechnol* **20**, 1030-1034.

10. Shaked, H., Melamed-Bessudo, C. and Levy, A. A. (2005) High frequency gene targeting in *Arabidopsis* plants expressing the yeast RAD54 gene. *Proc Natl Acad Sci USA* **102**, 12265-12269.

11. Fraser, A. G., Kamath, R. S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M. and Ahringer, J. (2000) Functional genomic analysis of C. elegans chromosome I by systematic RNA interference. *Nature* **408**, 325-330.

12. Coumoul, X. and Deng, C. X. (2006) RNAi in mice: a promising approach to decipher gene functions in vivo. *Biochimie* **88**, 637-643.

13. Smith, N. A., Singh, S. P., Wang, M. B., Stoutjesdijk, P. A., Green, A. G. and Waterhouse, P. M. (2000) Total silencing by intron-spliced hairpin RNAs. *Nature* **407**, 319-320.

14. Schwab, R., Ossowski, S., Riester, M., Warthmann, N. and Weigel, D. (2006) Highly specific gene silencing by artificial microRNAs in Arabidopsis. *Plant Cell* **18**, 1121-1133.

15. Hilson, P., Allemeersch, J., Altmann, T. & other authors (2004) Versatile gene-specific sequence tags for Arabidopsis functional genomics: transcript profiling and reverse genetics applications. *Genome Res* **14**, 2176-2189.

16. Small, I. (2007) RNAi for revealing and engineering plant gene functions. *Curr Opin Biotechnol* **18**, 148-153.

17. Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., Stevenson, D. K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C. C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C.,

Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D. E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W. L., Berry, C. C. and Ecker, J. R. (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**, 653-657.

18. Belfort, M. and Roberts, R. J. (1997) Homing endonucleases: keeping the house in order. *Nucleic Acids Res* **25**, 3379-3388.

19. Rouet, P., Smih, F. and Jasin, M. (1994) Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol* **14**, 8096-8106.

20. 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.

21. Salomon, S. and Puchta, H. (1998) Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. *EMBO J.* **17**, 6086-6095.

22. Chilton, M.-D. M. and Que, Q. (2003) Targeted integration of T-DNA into the tobacco genome at double-strand breaks: new insights on the mechanism of T-DNA integration. *Plant Physiol.* **133**, 956-965.

23. Tzfira, T., Frankmen, L., Vaidya, M. and Citovsky, V. (2003) Site-specific integration of *Agrobacterium tumefaciens* T-DNA *via* double-stranded intermediates. *Plant Physiol.* **133**, 1011-1023.

24. Jeltsch, A., Wenz, C., Wende, W., Selent, U. and Pingoud, A. (1996) Engineering novel restriction endonucleases: principles and applications. *Trends Biotechnol* **14**, 235-238.

25. Wu, J., Kandavelou, K. and Chandrasegaran, S. (2007) Custom-designed zinc finger nucleases: what is next? *Cell Mol Life Sci* **64**, 2933-2944.

26. 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.

27. Dreier, B., Beerli, R. R., Segal, D. J., Flippin, J. D. and Barbas, C. F., 3rd (2001) Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem* **276**, 29466-29478.

28. Dreier, B., Fuller, R. P., Segal, D. J., Lund, C. V., Blancafort, P., Huber, A., Koksch, B. and Barbas, C. F., 3rd (2005) Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem* **280**, 35588-35597.

29. Dreier, B., Segal, D. J. and Barbas, C. F., 3rd (2000) Insights into the molecular recognition of the 5'-GNN-3' family of DNA sequences by zinc finger domains. *J Mol Biol* **303**, 489-502.

30. Segal, D. J., Dreier, B., Beerli, R. R. and Barbas, C. F., 3rd (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc Natl Acad Sci U S A* **96**, 2758-2763.

31. Krizek, B. A., Amann, B. T., Kilfoil, V. J., Merkle, D. L. and Berg, J. M. (1991) A consensus zinc finger peptide: design, high-affinity metal binding, a pH-

dependent structure, and a His to Cys sequence variant. *Journal of the American Chemical Society* **113**, 4518-4523.

32. Wright, D. A., Thibodeau-Beganny, S., Sander, J. D., Winfrey, R. J., Hirsh, A. S., Eichtinger, M., Fu, F., Porteus, M. H., Dobbs, D., Voytas, D. F. and Joung, J. K. (2006) Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat Protoc* **1**, 1637-1652.

33. Mani, M., Kandavelou, K., Dy, F. J., Durai, S. and Chandrasegaran, S. (2005) Design, engineering, and characterization of zinc finger nucleases. *Biochem Biophys Res Commun* **335**, 447-457.

34. Sander, J. D., Zaback, P., Joung, J. K., Voytas, D. F. and Dobbs, D. (2007) Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucleic Acids Res* **35**, W599-605.

35. Mandell, J. G. and Barbas, C. F., 3rd (2006) Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res* **34**, W516-523.

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

37. Roberts, R. J., Vincze, T., Posfai, J. and Macelis, D. (2007) REBASE--enzymes and genes for DNA restriction and modification. *Nucleic Acids Res* **35**, D269-270.

38. Dafny-Yelin, M. and Tzfira, T. (2007) Delivery of multiple transgenes to plant cells. *Plant Physiol* **145**, 1118-1128.

39. Hellens, R., Mullineaux, P. and Klee, H. (2000) Technical Focus:a guide to Agrobacterium binary Ti vectors. *Trends Plant Sci* **5**, 446-451.

40. Tzfira, T., Kozlovsky, S. V. and Citovsky, V. (2007) Advanced expression vector systems: new weapons for plant research and biotechnology. *Plant Physiol* **145**, 1087-1089.

41. Tzfira, T., Tian, G.-W., Lacroix, B., Vyas, S., Li, J., Leitner-Dagan, Y., Krichevsky, A., Taylor, T., Vainstein, A. and Citovsky, V. (2005) pSAT vectors: a modular series of plasmids for autofluorescent protein tagging and expression of multiple genes in plants. *Plant Mol Biol* **57**, 503-516.

42. Goderis, I. J., De Bolle, M. F., Francois, I. E., Wouters, P. F., Broekaert, W. F. and Cammue, B. P. (2002) A set of modular plant transformation vectors allowing flexible insertion of up to six expression units. *Plant Mol Biol* **50**, 17-27.

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

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

45. Morton, J., Davis, M. W., Jorgensen, E. M. and Carroll, D. (2006) Induction and repair of zinc-finger nuclease-targeted double-strand breaks in Caenorhabditis elegans somatic cells. *Proc Natl Acad Sci U S A* **103**, 16370-16375.

46. Doyon, Y., McCammon, J. M., Miller, J. C., Faraji, F., Ngo, C., Katibah, G. E., Amora, R., Hocking, T. D., Zhang, L., Rebar, E. J., Gregory, P. D., Urnov, F. D.

and Amacher, S. L. (2008) Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat Biotechnol* **26**, 702-708.

47. Meng, X., Noyes, M. B., Zhu, L. J., Lawson, N. D. and Wolfe, S. A. (2008) Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol* **26**, 695-701.

48. Foley, J. E., Yeh, J. R., Maeder, M. L., Reyon, D., Sander, J. D., Peterson, R. T. and Joung, J. K. (2009) Rapid mutation of endogenous zebrafish genes using zinc finger nucleases made by Oligomerized Pool ENgineering (OPEN). *PLoS ONE* **4**, e4348.

49. 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.

50. Bibikova, M., Beumer, K., Trautman, J. K. and Carroll, D. (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science* **300**, 764.

51. Lloyd, A., Plaisier, C. L., Carroll, D. and Drews, G. N. (2005) Targeted mutagenesis using zinc-finger nucleases in *Arabidopsis*. *Proc Natl Acad Sci USA* **102**, 2232-2237.

52. Wright, D. A., Townsend, J. A., Winfrey, R. J., Jr., Irwin, P. A., Rajagopal, J., Lonosky, P. M., Hall, B. D., Jondle, M. D. and Voytas, D. F. (2005) High-frequency homologous recombination in plants mediated by zinc-finger nucleases. *Plant J* **44**, 693-705.

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

54. Cai, C. Q., Doyon, Y., Ainley, W. M., Miller, J. C., Dekelver, R. C., Moehle, E. A., Rock, J. M., Lee, Y. L., Garrison, R., Schulenberg, L., Blue, R., Worden, A., Baker, L., Faraji, F., Zhang, L., Holmes, M. C., Rebar, E. J., Collingwood, T. N., Rubin-Wilson, B., Gregory, P. D., Urnov, F. D. and Petolino, J. F. (2008) Targeted transgene integration in plant cells using designed zinc finger nucleases. *Plant Mol Biol*.

55. Tarkowski, A. K., Jagiello, K., Czolowska, R. and Ozdzenski, W. (2005) Mouse chimaeras developed from electrofused blastocysts: new evidence for developmental plasticity of the inner cell mass. *Int J Dev Biol* **49**, 909-914.

56. Koseki, M., Goto, K., Masuta, C. and Kanazawa, A. (2005) The star-type color pattern in Petunia hybrida 'red Star' flowers is induced by sequence-specific degradation of chalcone synthase RNA. *Plant Cell Physiol* **46**, 1879-1883.

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

58. Roberts, R. J., Vincze, T., Posfai, J. and Macelis, D. (2005) REBASE--restriction enzymes and DNA methyltransferases. *Nucleic Acids Res* **33**, D230-232.

59. Lin, L., Liu, Y. G., Xu, X. and Li, B. (2003) Efficient linking and transfer of multiple genes by a multigene assembly and transformation vector system. *Proc Natl Acad Sci U S A* **100**, 5962-5967.

60. Puchta, H., Dujon, B. and Hohn, B. (1996) Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. *Proc Natl Acad Sci USA* **93**, 5055-5060.

61. Thermes, V., Grabher, C., Ristoratore, F., Bourrat, F., Choulika, A., Wittbrodt, J. and Joly, J. S. (2002) I-SceI meganuclease mediates highly efficient transgenesis in fish. *Mech Dev* **118**, 91-98.

62. 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.

63. Windbichler, N., Papathanos, P. A., Catteruccia, F., Ranson, H., Burt, A. and Crisanti, A. (2007) Homing endonuclease mediated gene targeting in Anopheles gambiae cells and embryos. *Nucleic Acids Res* **35**, 5922-5933.

64. Fajardo-Sanchez, E., Stricher, F., Paques, F., Isalan, M. and Serrano, L. (2008) Computer design of obligate heterodimer meganucleases allows efficient cutting of custom DNA sequences. *Nucleic Acids Res*.

65. Arnould, S., Chames, P., Perez, C., Lacroix, E., Duclert, A., Epinat, J. C., Stricher, F., Petit, A. S., Patin, A., Guillier, S., Rolland, S., Prieto, J., Blanco, F. J., Bravo, J., Montoya, G., Serrano, L., Duchateau, P. and Paques, F. (2006) Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J Mol Biol* **355**, 443-458.

66. Smith, J., Grizot, S., Arnould, S., Duclert, A., Epinat, J. C., Chames, P., Prieto, J., Redondo, P., Blanco, F. J., Bravo, J., Montoya, G., Paques, F. and Duchateau, P. (2006) A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res* **34**, e149.

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

68. Shimizu, Y., Bhakta, M. S. and Segal, D. J. (2009) Restricted spacer tolerance of a zinc finger nuclease with a six amino acid linker. *Bioorg Med Chem Lett*.

69. Zeevi, V., Tovkach, A. and Tzfira, T. (2008) Increasing cloning possibilities using artificial zinc finger nucleases. *Proc Natl Acad Sci U S A* **105**, 12785-12790.

70. Karimi, M., De Meyer, B. and Hilson, P. (2005) Modular cloning in plant cells. *Trends Plant Sci* **10**, 103-105.

71. 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.

72. 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.

73. Huang, B., Schaeffer, C. J., Li, Q. and Tsai, M. D. (1996) Splase: a new class IIS zinc-finger restriction endonuclease with specificity for Sp1 binding sites. *J Protein Chem* **15**, 481-489.

74. 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.

75. Beumer, K., Bhattacharyya, G., Bibikova, M., Trautman, J. K. and Carroll, D. (2006) Efficient gene targeting in *Drosophila* with zinc-finger nucleases. *Genetics* **172**, 2391-2403.

76. Porteus, M. H. (2006) Mammalian gene targeting with designed zinc finger nucleases. *Mol Ther* **13**, 438-446.

77. Ekker, S. C. (2008) Zinc finger-based knockout punches for zebrafish genes. *Zebrafish* **5**, 121-123.

78. Tovkach, A., Zeevi, V. and Tzfira, T. (2009) A toolbox and procedural notes for characterizing novel zinc finger nucleases for genome editing in plant cells. *Plant J* **57**, 747-757.

79. Collins, T. J. (2007) ImageJ for microscopy. *Biotechniques* **43**, 25-30.

80. Lee, L. Y., Fang, M. J., Kuang, L. Y. and Gelvin, S. B. (2008) Vectors for multi-color bimolecular fluorescence complementation to investigate protein-protein interactions in living plant cells. *Plant Methods* **4**, 24.

81. Asselbergs, F. A. and Rival, S. (1996) Creation of a novel, versatile multiple cloning site cut by four rare-cutting homing endonucleases. *Biotechniques* **20**, 558-562.

82. Argast, G. M., Stephens, K. M., Emond, M. J. and Monnat, R. J., Jr. (1998) I-PpoI and I-CreI homing site sequence degeneracy determined by random mutagenesis and sequential in vitro enrichment. *J Mol Biol* **280**, 345-353.

83. Bryk, M., Quirk, S. M., Mueller, J. E., Loizos, N., Lawrence, C. and Belfort, M. (1993) The td intron endonuclease I-TevI makes extensive sequence-tolerant contacts across the minor groove of its DNA target. *EMBO J* **12**, 4040-4041.

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

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

86. Camenisch, T. D., Brilliant, M. H. and Segal, D. J. (2008) Critical parameters for genome editing using zinc finger nucleases. *Mini Rev Med Chem* **8**, 669-676.

87. Porteus, M. H. and Carroll, D. (2005) Gene targeting using zinc finger nucleases. *Nat Biotechnol* **23**, 967-973.

88. Zhou, F., Li, Q. and Gao, S. J. (2009) A sequence-independent in vitro transposon-based strategy for efficient cloning of genomes of large DNA viruses as bacterial artificial chromosomes. *Nucleic Acids Res* **37**, e2.

89. Liu, X., Peng, D., Luo, Y., Ruan, L., Yu, Z. and Sun, M. (2009) Construction of an Escherichia coli to Bacillus thuringiensis shuttle vector for large DNA fragments. *Appl Microbiol Biotechnol* **82**, 765-772.

90. Viret, J. F. (1993) Meganuclease I-SceI as a tool for the easy subcloning of large DNA fragments devoid of selection marker. *Biotechniques* **14**, 325-326.

91. Rozwadowski, K., Yang, W. and Kagale, S. (2008) Homologous recombination-mediated cloning and manipulation of genomic DNA regions using Gateway and recombineering systems. *BMC Biotechnol* **8**, 88.

92. Brownlee, C. (2005) Danna and Nathans: Restriction enzymes and the boon to modern molecular biology. *Proc Natl Acad Sci U S A* **102**, 5909.

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

94. Carroll, D., Morton, J. J., Beumer, K. J. and Segal, D. J. (2006) Design, construction and in vitro testing of zinc finger nucleases. *Nature Protocols* **1**, 1329-1341.

95. Cathomen, T., Segal, D. J., Brondani, V. and Muller-Lerch, F. (2008) Generation and functional analysis of zinc finger nucleases. *Methods Mol Biol* **434**, 277-290.

96. Desjarlais, J. R. and Berg, J. M. (1993) Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins. *Proc Natl Acad Sci U S A* **90**, 2256-2260.

97. Saslowsky, D. and Winkel-Shirley, B. (2001) Localization of flavonoid enzymes in Arabidopsis roots. *Plant J* **27**, 37-48.

98. Goodin, M. M., Dietzgen, R. G., Schichnes, D., Ruzin, S. and Jackson, A. O. (2002) pGD vectors: versatile tools for the expression of green and red fluorescent protein fusions in agroinfiltrated plant leaves. *Plant J.* **31**, 375-383.

99. Caccio, S., Camilli, R., La Rosa, G. and Pozio, E. (1998) Establishing the Cryptosporidium parvum karyotype by NotI and SfiI restriction analysis and Southern hybridization. *Gene* **219**, 73-79.

100. Shen, W., Huang, Y., Tang, Y., Liu, D. P. and Liang, C. C. (2005) A general method to modify BACs to generate large recombinant DNA fragments. *Mol Biotechnol* **31**, 181-186.

101. Paques, F. and Duchateau, P. (2007) Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Curr Gene Ther* **7**, 49-66.

102. Lu, C., Mansoorabadi, K. and Jeffries, T. (2007) Comparison of multiple gene assembly methods for metabolic engineering. *Appl Biochem Biotechnol* **137-140**, 703-710.

103. 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.

104. Collins, C. H., Yokobayashi, Y., Umeno, D. and Arnold, F. H. (2003) Engineering proteins that bind, move, make and break DNA. *Curr Opin Biotechnol* **14**, 665.

105. Lanio, T., Jeltsch, A. and Pingoud, A. (2000) On the possibilities and limitations of rational protein design to expand the specificity of restriction enzymes: a case study employing EcoRV as the target. *Protein Eng* **13**, 275-281.

106. Liu, Q., Xia, Z., Zhong, X. and Case, C. C. (2002) Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *J Biol Chem* **277**, 3850-3856.

107. Kim, Y. G., Shi, Y., Berg, J. M. and Chandrasegaran, S. (1997) Site-specific cleavage of DNA-RNA hybrids by zinc finger/FokI cleavage domain fusions. *Gene* **203**, 43-49.

108. Jasin, M., Moynahan, M. E. and Richardson, C. (1996) Targeted transgenesis. *Proc Natl Acad Sci USA* **93**, 8804-8808.

109. Siebert, R. and Puchta, H. (2002) Efficient repair of genomic double-strand breaks by homologous recombination between directly repeated sequences in the plant genome. *Plant Cell* **14**, 1121-1131.

110. Tzfira, T., Li, J., Lacroix, B. and Citovsky, V. (2004) *Agrobacterium* T-DNA integration: molecules and models. *Trends Genet.* **20**, 375-383.

111. Puchta, H. (2002) Gene replacement by homologous recombination in plants. *Plant Mol Biol* **48**, 173-182.

112. Chiurazzi, M., Ray, A., Viret, J. F., Perera, R., Wang, X. H., Lloyd, A. M. and Signer, E. R. (1996) Enhancement of somatic intrachromosomal homologous recombination in Arabidopsis by the HO endonuclease. *Plant Cell* **8**, 2057-2066.

113. Carroll, D. (2004) Using nucleases to stimulate homologous recombination. *Methods Mol Biol* **262**, 195-207.

114. Porteus, M. H. and Baltimore, D. (2003) Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**, 763.

115. 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 USA* **104**, 3055-3060.

116. Santiago, Y., Chan, E., Liu, P. Q., Orlando, S., Zhang, L., Urnov, F. D., Holmes, M. C., Guschin, D., Waite, A., Miller, J. C., Rebar, E. J., Gregory, P. D., Klug, A. and Collingwood, T. N. (2008) Targeted gene knockout in mammalian cells by

using engineered zinc-finger nucleases. *Proc Natl Acad Sci U S A* **105**, 5809-5814.

117. Porteus, M. (2008) Design and testing of zinc finger nucleases for use in mammalian cells. *Methods Mol Biol* **435**, 47-61.

118. Carroll, D., Beumer, K. J., Morton, J. J., Bozas, A. and Trautman, J. K. (2008) Gene targeting in Drosophila and Caenorhabditis elegans with zinc-finger nucleases. *Methods Mol Biol* **435**, 63-77.

119. Chung, S. M., Frankman, E. L. and Tzfira, T. (2005) A versatile vector system for multiple gene expression in plants. *Trends Plant Sci* **10**, 357-361.

120. Van der Hoorn, R. A., Laurent, F., Roth, R. and De Wit, P. J. (2000) Agroinfiltration is a versatile tool that facilitates comparative analyses of Avr/cf-9-induced and Avr4/cf-4-induced recrosis. *Mol Plant Microbe Interact* **13**, 439-446.

121. Gallois, P. and Marinho, P. (1995) Leaf disk transformation using Agrobacterium tumefaciens-expression of heterologous genes in tobacco. *Methods Mol Biol* **49**, 39-48.

122. Zhang, X., Henriques, R., Lin, S. S., Niu, Q. W. and Chua, N. H. (2006) Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. *Nat Protoc* **1**, 641-646.

123. Jefferson, R. A., Kavanagh, T. A. and Bevan, M. W. (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6**, 3901-3907.

124. Takahashi, T. and Komeda, Y. (1989) Characterization of two genes encoding small heat-shock proteins in *Arabidopsis thaliana*. *Mol Gen Genet* **219**, 365-372.

125. Kirik, A., Salomon, S. and Puchta, H. (2000) Species-specific double-strand break repair and genome evolution in plants. *EMBO J* **19**, 5562-5566.

126. Tzfira, T. and White, C. (2005) Towards targeted mutagenesis and gene replacement in plants. *Trends Biotechnol* **23**, 567-569.

127. Isalan, M. and Choo, Y. (2001) Rapid, high-throughput engineering of sequence-specific zinc finger DNA-binding proteins. *Methods Enzymol* **340**, 593-609.

128. Miller, J. C., Holmes, M. C., Wang, J., Guschin, D. Y., Lee, Y. L., Rupniewski, I., Beausejour, C. M., Waite, A. J., Wang, N. S., Kim, K. A., Gregory, P. D., Pabo, C. O. and Rebar, E. J. (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol* **25**, 778-785.