Creating highly specific nucleases by fusion of active restriction endonucleases and catalytically inactive homing endonucleases

Ines Fonfara[1], Ute Curth[2], Alfred Pingoud[1] and Wolfgang Wende[1,*]

[1]Institut für Biochemie, Justus-Liebig-Universität Giessen, Heinrich-Buff-Ring 58, D-35392, Giessen and [2]Institut für Biophysikalische Chemie, Medizinische Hochschule Hannover, Carl-Neuberg-Str. 1, D-30625, Hannover, Germany

ABSTRACT

Zinc-finger nucleases and TALE nucleases are produced by combining a specific DNA-binding module and a non-specific DNA-cleavage module, resulting in nucleases able to cleave DNA at a unique sequence. Here a new approach for creating highly specific nucleases was pursued by fusing a catalytically inactive variant of the homing endonuclease I-SceI, as DNA binding-module, to the type IIP restriction enzyme PvuII, as cleavage module. The fusion enzymes were designed to recognize a composite site comprising the recognition site of PvuII flanked by the recognition site of I-SceI. In order to reduce activity on PvuII sites lacking the flanking I-SceI sites, the enzymes were optimized so that the binding of I-SceI to its sites positions PvuII for cleavage of the composite site. This was achieved by optimization of the linker and by introducing amino acid substitutions in PvuII which decrease its activity or disturb its dimer interface. The most specific variant showed a more than 1000-fold preference for the addressed composite site over an unaddressed PvuII site. These results indicate that using a specific restriction enzyme, such as PvuII, as cleavage module, offers an alternative to the otherwise often used catalytic domain of FokI, which by itself does not contribute to the specificity of the engineered nuclease.

INTRODUCTION

Gene targeting allows the introduction of genetic elements into chosen genomic loci. The usual approach to achieve this goal is based on homologous recombination, the natural efficiency of which is quite low but can be stimulated by double-strand breaks using engineered nucleases. For gene targeting these endonucleases should recognize unique sequences and cleave DNA highly specifically, and thereby initiate double-strand break repair by induced homologous recombination with a DNA template supplied *in trans* (1,2).

Such nucleases with very high specificity have been engineered by rational and combinatorial strategies (3): among them are (i) zinc-finger nucleases (4–9), (ii) engineered homing endonucleases (10–13) and (iii) triple helix forming oligonucleotides linked to restriction enzymes (14) or other DNA cleavage moduls (15,16).

Combining a cleavage module with a recognition module in order to generate nucleases with high specificity was first achieved with the non-specific DNA cleavage domain of the restriction endonuclease FokI and the *Ubx* homeodomain (17), which was soon followed by the fusion of zinc finger motifs to the FokI cleavage domain (18). Most chimeric nucleases produced so far use the FokI cleavage domain as cleavage module (19), and the majority of them are zinc-finger nucleases. A noteworthy exception is the fusion product consisting of a catalytically inactive homing endonuclease and the FokI cleavage domain (20), which resulted in a fusion protein recognizing the homing endonuclease recognition site and cleaving the DNA in both strands 2 and 6 nt downstream from the recognition site, as expected for the Type IIS restriction endonucleases FokI. With a similar approach the ^{5}mCpG-binding domain of a ^{5}mCpG-specific DNA glycosylase was fused to the FokI cleavage domain, in order to obtain a chimeric nuclease that cleaves DNA at ^{5}mCpG-sites (21). Similarly, the control protein C.BcII that represses the expression of the methyltransferase of the BcII R-M system was fused to the non-specific cleavage domain of BmrI; the chimeric nuclease recognizes specific sites in the vicinity of the C.BcII control sequence (22).

Recently, so called transcription activator like effector (TALE) nucleases (TALENs) were produced; they consist

*To whom correspondence should be addressed. Tel: +496419935407; Fax: +496419935409; Email: wolfgang.wende@chemie.bio.uni-giessen.de

of TALEs as recognition module (23–26) and the FokI cleavage domain as cleavage module. Similarly as zinc-fingers, TALEs can be designed to recognize different DNA sequences (23,24) and therefore promise to be useful for the production of engineered nucleases of extended specificity (27–33).

To our knowledge no attempts have been made so far to produce chimeric nucleases consisting of a homing endonuclease (as a DNA-binding module) and a restriction enzyme (as a DNA binding and cleavage module, respectively). Such chimeric nucleases ideally should have an extended specificity and should recognize a bipartite or tripartite (in case of homodimeric restriction endonucleases, see Figure 1A) recognition site, comprising that of the homing endonuclease and that of the restriction endonuclease and cleave the DNA within the recognition site of the restriction endonuclease.

Type II restriction endonucleases are highly specific enzymes; they recognize DNA sequences of 4–8 bp in length, which means that their recognition sites will occur on average every 4^4–4^8 bp. PvuII (34), a homodimeric Type IIP restriction endonuclease, for example, recognizes the palindromic double-stranded DNA sequence CAG↓CTG and cleaves it in the presence of Mg^{2+} as indicated, generating blunt ends. Homing endonucleases are even more specific, as they recognize much longer sequences than restriction endonucleases (35). I-SceI (36,37), a monomeric homing endonuclease of the LAGLIDADG family, for example, recognizes an 18 bp sequence (TAG GGATAACAGGGTAAT) and in the presence of Mg^{2+} cleaves the DNA within this sequence. Typically, homing endonucleases cleave complex genomes at a few sites only.

A fusion protein consisting of PvuII as cleavage module and a catalytically inactive I-SceI variant (= S*) as binding module, separated by a linker L_x, for example PL_xS^*, which upon dimerization (viz. S^*L_xP/PL_xS^*) could be expected to recognize a 42 bp sequence. This composite site consists of one I-SceI site, one PvuII site and one reverse complement of the I-SceI site: TAGGGA TAACAGGGTAAT — CAG↓CTG— ATTACCCTGTT ATCCCTA. Cleavage should occur in the addressed PvuII site at the position indicated. This requires that the catalytic module, i.e. the PvuII part of the fusion protein is only activated after the DNA binding modules, the two I-SceI*parts of the fusion protein, have positioned the catalytic module vis-à-vis the scissile phosphodiester bonds of the PvuII site. The PL_xS^* fusion protein should not be able to cleave an unaddressed PvuII site, i.e. a PvuII site not flanked by I-SceI sites. We demonstrate here that by an appropriate choice of the linker between PvuII and I-SceI* and by introducing amino acid substitutions into PvuII and I-SceI*, the PL_xS^* fusion protein cleaves with high preference addressed PvuII sites.
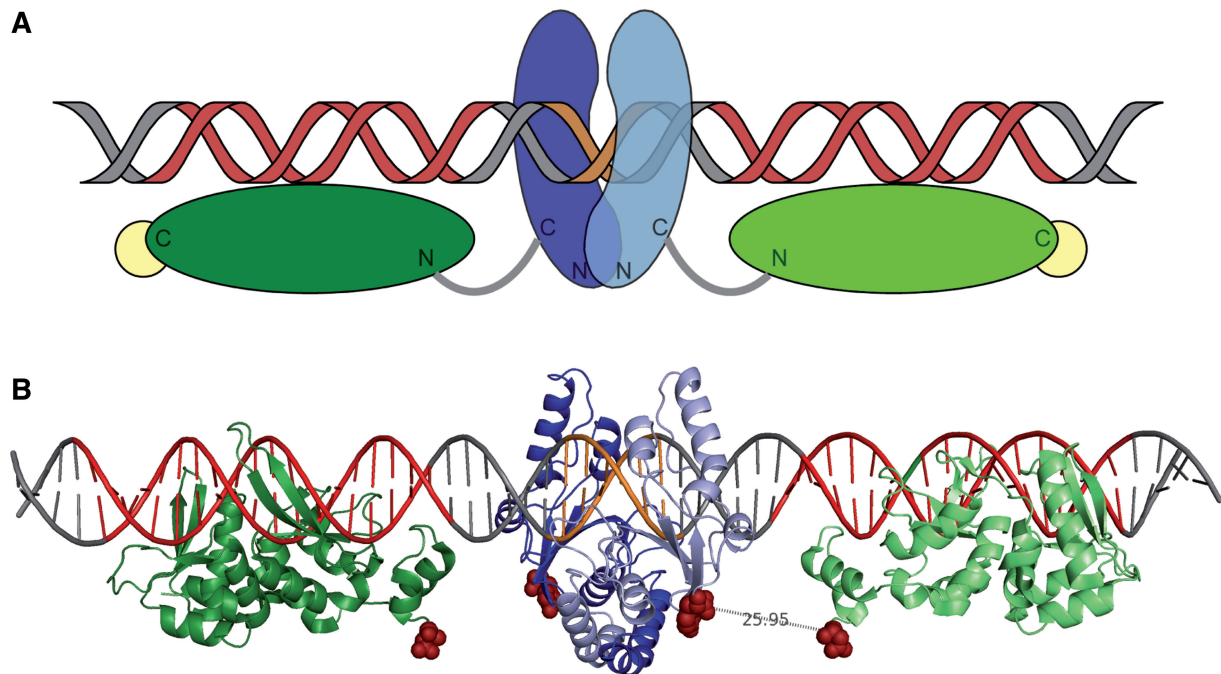

Figure 1. (**A**) Cartoon representation of the fusion enzyme. The C-terminus of one subunit of a dimeric restriction enzyme (PvuII; blue) is fused via a peptide linker (gray) to the N-terminus of a catalytically inactive homing endonuclease (I-SceI; green). This fusion enzyme binds as a homodimer to a tripartite recognition site consisting of the PvuII recognition site (orange) flanked by two I-SceI sites (red) separated by an un-specified DNA (gray). The very C-terminal Strep-tag used for affinity purification is depicted in yellow. (**B**) Model of the engineered fusion enzyme. One subunit of the fusion enzyme consists of one I-SceI [green; pdb 1r7m (40)] and one subunit of the PvuII dimer [blue, pdb 1pvi (47)]. This model was built by aligning the recognition sites from the crystal structures of the individual proteins on a DNA composed of the PvuII recognition site (orange) and two I-SceI sites (red) separated by 6 bp up- and downstream. The C-terminus of PvuII and the N-terminus of I-SceI, separated by 2.6 nm, are indicated by red spheres; this distance must be covered by a peptide linker of suitable length.

MATERIALS AND METHODS

Construction of the PvuII-I-SceI fusions

The PvuII-I-SceI fusion enzyme and variants thereof were created by fusing scPvuII (38) or wt PvuII (39) via their C-terminus to the N-terminus of a catalytically inactive variant of I-SceI (I-SceI*) which had been truncated at the C-terminus (ΔC9). These last nine amino acid residues were not resolved in the co-crystal structure of I-SceI (40). For this construction, the genes coding for scPvuII or wt PvuII with the coding sequence for a C-terminal His$_6$-tag were connected to the gene coding for I-SceI(D44S) (41) and introduced into the vector pASK-IBA63b-plus (IBA) coding for a C-terminal Strep-tag. For further improvement of the fusion protein two active site residues of I-SceI were exchanged (D44N and D145A) according to (20) via PCR-based site-directed mutagenesis (42). The mutagenesis of selected residues of PvuII was performed in the same way for the fusion protein and for PvuII alone (for control reasons), whose gene was also introduced into pASK-IBA63b-plus. Between the genes for PvuII and I-SceI three restriction enzyme sites (NheI, BsiWI, AgeI) were introduced, replacing the coding sequence for the His$_6$-tag (L$_H$ $\rightarrow$ L$_6$, where L denotes the linker). By cleaving the resulting vector with NheI and AgeI, and filling in, the linkers L$_N$, L$_+$ and L$_-$ were obtained. The correctness of all genetic constructs was confirmed by sequencing over the entire coding region for PvuII-L$_x$-I-SceI, containing the different linkers L$_x$ (L$_H$, L$_6$, L$_N$, L$_+$ and L$_-$).

Purification of the enzymes

The pASK-IBA63b-plus plasmids harboring the coding sequences for the fusion enzymes or PvuII alone were introduced into *Escherichia coli* strain XL-Gold, which had been transformed before with a plasmid encoding the PvuII methyltransferase gene (pLGM). The cells were grown at 37°C to an OD$_{600}$ of 1. Protein expression was induced by adding 200 µg/l anhydrotetracycline, followed by further growth at 20°C overnight. After harvesting the cells by centrifugation the cell pellet was dissolved in 100 mM Tris–HCl, pH 8, 1 M NaCl, 1 mM EDTA, lysed by sonication and centrifuged for 30 min ($>$17 000 g) to remove cell debris. The recombinant Strep-tagged proteins were purified using Strep-Tactin Sepharose (IBA). After washing with a high-salt (100 mM Tris–HCl, pH 8, 1 M NaCl, 1 mM EDTA) and a low-salt (100 mM Tris–HCl, pH 8, 300 mM NaCl, 1 mM EDTA) washing buffer, the tagged proteins were eluted with 100 mM Tris–HCl, pH 8.0, 300 mM NaCl, 2.5 mM desthiobiotin. The purity of the eluates was analyzed by SDS-PAGE and selected fractions were pooled and dialysed overnight (50 mM Na-phosphate, pH 7.5, 100 mM NaCl, 0.5 mM EDTA, 60% glycerol). Protein concentration was determined by A$_{280}$ measurements using a molar extinction coefficient of 92 250 M^{-1} cm^{-1} at 280 nm as calculated from the amino acid composition (43). Protein concentrations are given as dimers throughout the text.

DNA cleavage assays

DNA cleavage activity assays were carried out with 50 nM of a 454 bp PCR fragment containing one I-SceI site. This substrate was incubated with either 5U I-SceI (Fermentas) in Buffer Tango (Fermentas), 100 nM PvuII or 100 nM scPL$_H$S* in Buffer Green (Fermentas) for 45 min at 37°C. Cleavage products were separated by 6% PAGE, stained with ethidium bromide and visualized by a BioDocAnalyze system (Biometra). The cleavage products obtained with scPL$_H$S* were gel purified and sequenced.

For plasmid cleavage assays, the different target sequences (Table 2) were introduced into pAT153. The plasmid substrate was used in 8 nM concentration and incubated with 8 nM of the corresponding enzyme either for 3 h or overnight at 37°C. DNA cleavage by PvuII was performed in Buffer Green (Fermentas), by the fusion enzymes in an optimized potassium glutamate buffer [KGB (44)] (100 mM K-glutamate, 25 mM Tris–acetate, pH 7.5, 100 mM KCl, 0.8 mM Mg-acetate, 10 µg/ml BSA, 0.5 mM 2-mercaptoethanol). The cleavage products were separated on 0.8% agarose gels, stained with ethidium bromide and visualized by the BioDocAnalyze system. Quantification of linear product formation was performed by densitometry.

For bacteriophage λ-DNA (which contains 15 PvuII sites) cleavage, the DNA concentration was chosen such, that 8 nM of PvuII sites (corresponding to 564 pM λ-DNA) were incubated with 8 nM fusion enzymes or PvuII in optimized KGB or Buffer Green, respectively, for 3 h at 37°C.

Analytical ultracentrifugation

Protein PL$_6$S* was purified as described above followed by size exclusion chromatography on a Superdex 200 10/300 GL column with 50 mM Na-phosphate, pH 7.5, 500 mM NaCl, 0.5 mM EDTA. Fractions containing pure enzyme were pooled and one-half was dialyzed against 50 mM Na-phosphate, pH 7.5, 500 mM NaCl, 0.47 M glycerol (high salt conditions), the other half against 50 mM Na-phosphate, pH 7.5, 100 mM NaCl, 0.47 M glycerol (low salt conditions).

Sedimentation velocity runs were performed at 5°C and 45 000 r.p.m. in a Beckman Coulter Optima XL-I using the absorbance optics of the analytical ultracentrifuge. The experiments were carried out in an An-50 Ti rotor using 12 mm standard double sector centerpieces filled with 400 µl sample. Under high-salt conditions protein concentrations between 0.2 and 3.25 µM were examined, under low salt conditions a concentration of 1.75 µM was used. Depending on protein concentration, UV-absorption was measured at 230 or 280 nm, respectively. Sedimentation profiles were analyzed using the program SEDFIT (45), providing a model of diffusion-corrected differential sedimentation coefficient distributions [c(s) distributions]. Partial specific volumes, buffer densities and viscosities were determined with the program SEDNTERP (46) and used to correct the sedimentation coefficients to s$_{20,w}$.

Competition cleavage experiments

To test the specificity of the fusion enzymes for the addressed site, radioactively labeled PCR-fragments using [α^{32}P] dATP containing either the addressed site (S6P6S, for definition see Table 2) or the unaddressed site (P) were produced. 10 nM of each of these fragments were incubated together with 10 nM of the fusion enzymes in KGB (100 mM K-glutamate, 25 mM Tris–acetate, pH 7.5, 100 mM KCl, 0.8 mM Mg-acetate, 10 µg/ml BSA, 0.5 mM 2-mercaptoethanol) at 37°C; after defined time intervals samples were taken and subjected to polyacrylamide gel electrophoresis. Cleavage products were analyzed with the Instant Imager system (Packard) and quantified with the Instant Imager software. Initial rates were calculated by linear regression analysis.

Electrophoretic mobility shift assay

The affinity of the fusion enzymes to their target sequences was determined with electrophoretic mobility shift assays (EMSAs), using radioactively labeled DNA-fragments which were produced via PCR using [α^{32}P] dATP. One fragment contained the addressed site (S6P6S) and another one only an I-SceI site (S). For K_D determination 2 nM of the radioactively labeled addressed substrate were incubated with fusion enzyme in concentrations ranging from 1–150 nM in KGB without Mg^{2+} for 30 min at room temperature. For the unaddressed substrate the enzyme concentration ranged from 20 to 180 nM. After adding 1 µl 87% glycerol the samples were loaded onto a 6% polyacrylamide Tris–acetate, pH 8.5 gel and subjected to electrophoresis for 2 h at 10 V/cm. The bands were visualized using the Instant Imager system (Packard). The intensities of bands representing free and bound DNA were quantified with the Instant Imager software and the percentage of bound DNA was calculated. The data were fit and the K_D determined using a non-linear regression analysis.

RESULTS

Design of the fusion enzyme

For creating a highly specific nuclease with extended specificity the type IIP restriction endonuclease PvuII was fused as the cleavage module to a catalytically inactive variant of the homing endonuclease I-SceI (= S*) as an additional binding module (Figure 1A). Due to the C2 symmetry of the homodimeric PvuII (= P) the fusion would result in a protein with the PvuII homodimer flanked by two I-SceI* molecules. The fusion protein, PL$_x$S*, after homodimerization, should recognize a sequence consisting of the PvuII recognition sequence flanked by two I-SceI recognition sequences, S…P…S (where S and P denote the I-SceI and PvuII recognition sequences, respectively).

In the co-crystal structure, the N-terminal part of I-SceI is an α-helix sticking out of the otherwise compact structure of this enzyme and could serve as an attachment point for PvuII, as shown in a model of the fusion protein in complex with DNA, which was generated using the pdb coordinates for I-SceI [pdb:1r7m (40)] and PvuII [pdb: 1pvi (47)] and ideal B-DNA (Figure 1B). In this model the target sites for I-SceI and PvuII are separated by 6 bp, which leads to all three proteins in the fusion protein approaching the DNA from the same side.

Engineering of PvuII-I-SceI fusion enzymes and finding a target with an optimum distance between the I-SceI and the PvuII recognition site

In our first construct, produced for the purpose of testing our approach, the single chain variant of PvuII, scPvuII, was fused via its C-terminal His$_6$-tag as a linker (L$_H$) to the catalytically inactive I-SceI D44S variant (41). The DNA cleavage activity of the scPvuII-L$_H$-I-SceID44S fusion protein was tested on an oligonucleotide substrate containing only an I-SceI site and no PvuII site, which we expected not to be cleaved. However, a specific cleavage pattern was obtained (Figure 2A), which after sequencing of the product (Figure 2B) turned out to be due to
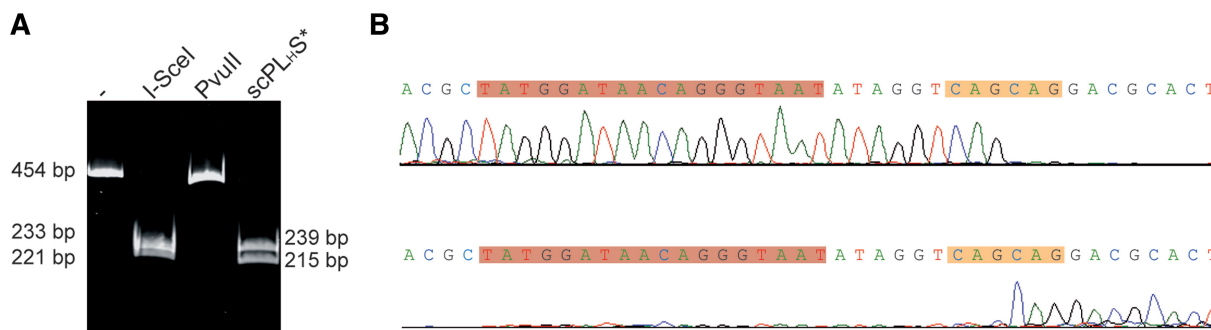

Figure 2. (**A**) Cleavage of 50 nM of a 454 bp PCR fragment containing an I-SceI recognition site. Lane 1 (-): uncleaved DNA fragment; Lane 2: cleavage products obtained by incubation of the 454 bp PCR fragment with 5 U I-SceI (Fermentas) in Buffer Tango; Lanes 3 and 4: cleavage products obtained by incubation of the 454 bp PCR fragment with 100 nM PvuII and 100 nM scPL$_H$S, respectively, in Buffer Green. (**B**) Sequencing results of the cleavage product of scPL$_H$S* from (A). The fragments were gelpurified and sequenced in forward (upper panel) and reverse (lower panel) direction. The I-SceI recognition sequence is indicated by a red bar and the cleavage site, which corresponds to a PvuII 'star' site by an orange bar. These two sites are separated by 6 bp.

cleavage at a PvuII 'star' site, (CAG↓CAG), 6 bp downstream from the I-SceI site. 'Star' sites, which differ in one base pair from the canonical recognition site, are only attacked by type II restriction enzymes at high concentrations of organic solvents such as DMSO and glycerol, at alkaline pH or in the presence of Mn^{2+}. The DNA cleavage result that we obtained, therefore, showed that (i) I-SceI activates PvuII to cleave a site that it normally would not attack and that it directs PvuII to this site, and (ii) that a distance of 6 bp between the I-SceI and the PvuII recognition site in the substrate as predicted by the structural model (Figure 1B) is acceptable for the scPvuII-I-SceI D44S fusion protein. The DNA cleavage activity of this fusion protein was tested under various buffer conditions, among them the KGB (100 mM K-glutamate, 25 mM Tris–acetate, pH 7.5, 10 mM Mg-acetate, 10 μg/ml BSA, 0.5 mM 2-mercaptoethanol) which had been developed to mimic physiological conditions (44) and to which we added 100 mM KCl and in which the concentration of Mg-acetate was decreased to 0.8 mM, in order to more closely approach the KCl and Mg^{2+} concentration prevailing *in vivo*. In this buffer the largest difference in activity between scPvuII and scPL$_H$ S* (Supplementary Figure S1) was observed, which means that the fusion construct cleaves an addressed site much better than scPvuII alone does. All further cleavage experiments of the fusion enzymes were performed in the optimized version of KGB (100 mM K-glutamate, 25 mM Tris–acetate, pH 7.5, 100 mM KCl, 0.8 mM Mg-acetate, 10 μg/ml BSA, 0.5 mM 2-mercaptoethanol).

For all other constructs the homodimer of PvuII was used instead of the single chain variant, which means that the fusion proteins contains two I-SceI* binding modules rather than one as in the scPvuII fusion proteins. The two I-SceI* modules are linked via their N-termini to the C-termini of the two PvuII subunits. For this construct, we also used another I-SceI variant, namely I-SceIΔC9,D44N,D145A; this variant has amino acid substitutions in both catalytic centers and is C-terminally truncated. For the amino acid substitutions in the catalytic center it was shown before, that this combination improves the solubility of I-SceI (20). Since fusion proteins with wtPvuII are homodimeric, their recognition site will be longer and they are likely to be more specific. wtPvuII and the I-SceIΔC9,D44N,D145A were fused using different linkers (L$_H$, L$_6$, L$_N$, L$_+$ and L$_-$) (Table 1) which all could cover the ~26 Å distance between the C-terminus of PvuII and the N-terminus of I-SceI in our structural model (Figure 1B). These fusion proteins were expressed and

purified. To test these enzymes, substrates with different distances between the I-SceI and the PvuII recognition sites (Table 2) were produced. Considering the structural model and the DNA cleavage results with scPL$_{(H)}$S* cleaving 6 bp downstream from the I-SceI site, we tested the different linker variants for specificity on the linearized plasmid substrate S6P6S_P, i.e. a substrate containing a PvuII site (P) flanked by two I-SceI sites (S) in a 6 bp distance (6P and P6, respectively) and an additional single PvuII site (_P) (Figure 3B). wtPvuII cleaves this substrate at both PvuII sites (the addressed one, which is flanked by I-SceI sites, and the other one, which is unaddressed) resulting in three major products, which shows, that PvuII by itself has no preference for one of the sites due to sequence context. The fusion enzymes in contrast cleave preferentially the addressed site (S6P6S) resulting in the fragments indicated with an arrow in Figure 3B. The unaddressed site (_P) is also cleaved but much more slowly; PL$_6$S* and PL$_+$S* cleaved the unaddressed site with an at least two orders of magnitude slower rate than the addressed site (Table 4). Still, a plasmid substrate with only an unaddressed PvuII site (Figure 3A) which is quickly cleaved by wtPvuII is also cleaved by the fusion enzymes, albeit with a lower rate. PL$_6$S* and PL$_+$S* show the slowest unaddressed cleavage rate. For this reason these variants were used for further optimization and characterization.

Oligomeric state of PL$_6$S*

PvuII is a homodimer in solution (48). In order to investigate if the fusion with I-SceI has an effect on the oligomeric state, sedimentation velocity runs in the analytical ultracentrifuge were carried out. These experiments were done at high and low salt conditions. Under high salt conditions PL$_6$S* sediments as a single species with a sedimentation coefficient of $s_{20,w} = 4.4$ S (Figure 4). Increasing the protein concentration from 0.2 to 3.25 μM had no effect on the sedimentation coefficient (data not shown), indicating that PL$_6$S* does not change its oligomerization state in this concentration range. Analysis of the diffusion broadening of the sedimenting boundary using the c(s) analysis in the program SEDFIT (45) yielded a molar mass of about 83 kg/mol. Since the molar mass of the monomer as calculated from amino acid composition is 46.7 kg/mol, the protein most probably is a homodimer. From the sedimentation analysis it can be

Table 1. Overview of the linkers used in the engineered PvuII-I-SceI fusions

Linker name	Length	Sequence	Source
L$_H$	10 AA	GSHHHHHHGT	Engineered
L$_6$	6 AA	ASRTTG	Engineered
L$_N$	10 AA	ASGGSGSGSG	Engineered
L$_+$	10 AA	ASTKQLVKSG	FokI
L$_-$	10 AA	ASGDSGSDSG	Engineered

Table 2. Overview of the substrates used

Name	Recognition sites
S	I-SceI: TAATGGGACAATAGGGAT
P	PvuII: CAGCTG
S4P4S	I-SceI-4 bp-PvuII-4 bp-I-SceI
S6x6S	I-SceI-6 bp-KpnI-6 bp-I-SceI
S6P6S	I-SceI-6 bp-PvuII-6 bp-I-SceI
S6P6S_P	I-SceI-6 bp-PvuII-6 bp-I-SceI-763 bp-PvuII
S8P8S	I-SceI-8 bp-PvuII-8 bp-I-SceI

The substrates P and S6P6S are also called unaddressed and addressed substrates, respectively.
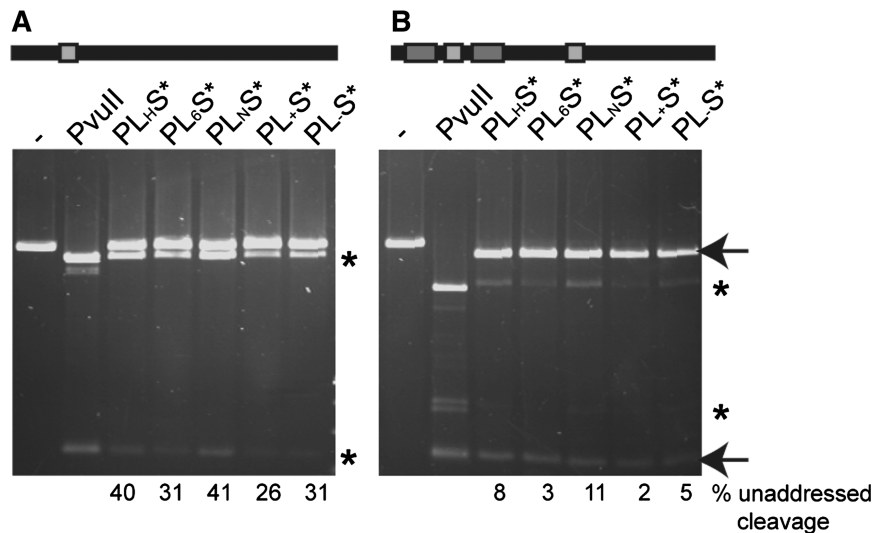
Figure 3. Cleavage activity of fusion enzymes with different linkers on linearized DNA. The DNA substrate contains either (**A**) a single PvuII site (light gray bar) or (**B**) the tripartite recognition site consisting of a PvuII site (light gray bar) flanked by two I-SceI sites (dark gray bars) each 6 bp away, and an additional unaddressed PvuII site. The experiment was performed under optimized conditions for the corresponding enzymes, namely Buffer Green for PvuII and optimized KGB for the fusion enzymes. The products of addressed cleavage at the tripartite recognition site are indicated by arrows and the products of unaddressed cleavage at PvuII sites by asterisks. The percentage of unaddressed cleavage is shown at the bottom.
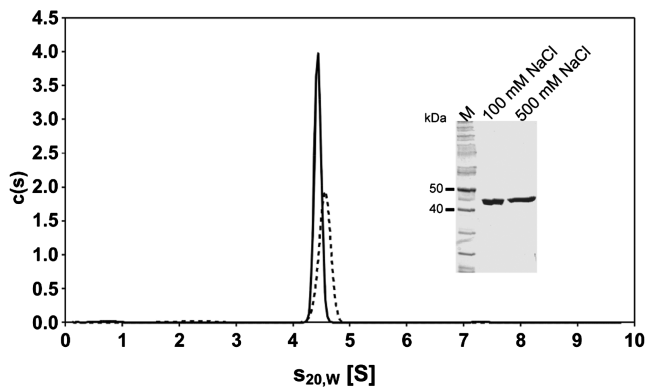

Figure 4. Sedimentation coefficient distributions in an analytical ultra-centrifugation experiment with PL$_6$S* at different salt concentrations. PL$_6$S* at concentrations of 3.3 or 3.5 μM was analyzed in sedimentation velocity experiments at 500 mM NaCl (solid line) or 100 mM NaCl (dashed line), respectively. c(s) analysis using the program SEDFIT yielded a single sedimenting species with $s_{20,W} = 4.4$ S under high salt and $s_{20,W} = 4.6$ S under low salt conditions, indicating that PL$_6$S* is a stable homodimer. The purity and stability of the protein analyzed had been determined before by SDS-PAGE of 1.7 μg protein diluted in either low or high salt buffer (see insert). The theoretical molecular weight of one subunit of PL$_6$S* is 46.7 kDa.

ruled out, therefore, that the protein exists as a monomer in solution, since it sediments faster than a sphere of the corresponding mass. For the dimer a frictional ratio of $f/f_0 = 1.58$ can be calculated from the sedimentation coefficient. As hydrated spherical proteins are expected to yield frictional ratios in the range of 1.1–1.2 (49), the shape of PL$_6$S* must deviate substantially from a sphere, suggesting that it is an elongated homodimer and/or contains flexible regions that increase the frictional coefficient of the protein. Under low salt conditions the sedimentation coefficient increased slightly to $s_{20,W} = 4.6$ S

(Figure 4), and a molar mass of about 80 kg/mol was obtained from the c(s) analysis. This indicates that the PL$_6$S* homodimer formed under low salt conditions is slightly more compact than under high salt conditions.

Identification of residues in PvuII that increase the specificity of the fusion enzyme

The fusion enzymes PL$_6$S* and PL$_+$S* are still able to cleave unaddressed PvuII-sites (Figure 3A), albeit with much lower rate than the addressed sites. There are in principle two possibilities to increase the preference of the fusion enzymes for the addressed site, which both rely on making the cleavage activity of the cleavage module PvuII dependent on the binding of the additional binding modules I-SceI* to the I-SceI site: (i) either by decreasing the activity (k_{cat}, K_m) of PvuII so that it cannot cleave DNA if not positioned by I-SceI*, or (ii) by weakening the dimer interface of PvuII to make a proper subunit–subunit interaction, which is a prerequisite for DNA cleavage, dependent on binding of the I-SceI* modules to their target sites.

For the first approach (i) the mutations T46G, known to be responsible for high fidelity (Zhenyu Zhu, New England Biolabs, personal communication) and Y94F, known to have a decreased Mg^{2+} binding ability (50) were introduced into PvuII (Figure 5, right panel) and in the fusion enzyme. Indeed the PvuII(T46G,Y94F) variant showed an ∼100-fold decrease in activity compared to wtPvuII (data not shown). To weaken the PvuII dimer interface (ii) without affecting the catalytic activity too much we decided to choose residues in the N-terminal α-helices (Figure 5, left panel) which form close contacts between the two subunits and are far away from the active center (Figure 5, lower panel). The residues L12 and H15 were substituted by Glu and Asp, respectively, and
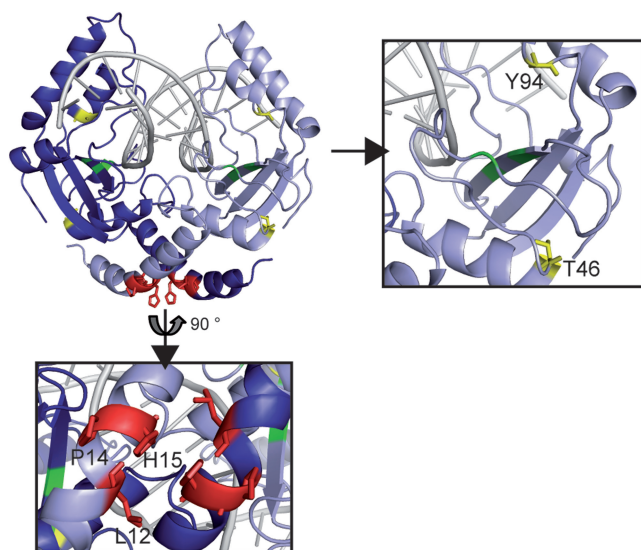
Figure 5. Structural details of PvuII (pdb 1pvi). The catalytic centre composed of the residues D58, E68 and K70 is highlighted in green. The positions of substitutions introduced into PvuII are: T46G (yellow) increases the fidelity of PvuII; Y94F (yellow) has a reduced Mg^{2+} binding ability (50); L12E, P14G and H15D (red) are likely to weaken the dimer interface.

thereby generate electrostatic repulsion between the two subunits. Furthermore the central residue of this helix P14 was substituted by Gly in order to affect the bending of the helix and by this reduce the interface contacts. These variants were introduced individually into wtPvuII and in the fusion enzyme as well. The PvuII(L12E), PvuII(P14G) and PvuII(H15D) variants were characterized with respect to their kinetic parameters, which showed, that the mutations had only a slight influence on the catalytic activity (Supplementary Table S1).

Characterization of fusion enzymes containing different amino acid substitutions in the cleavage module PvuII

We tested the effects of the amino acid substitutions T46G and Y94F in the fusion enzymes containing the L_6 or the L_+ linker by competition experiments with two internally labeled PCR fragments containing either the addressed site (S6P6S) or the unaddressed site (P) in equimolar concentration of enzyme and the two substrates ($c_{PL6/+S^*}$: c_{S6P6S}: $c_P = 1:1:1$) over 3 h. The fusion enzymes PL_6S^* (Figure 6A) and PL_+S^* (Figure 6C) containing wtPvuII as cleavage module cleave the addressed site by factors of ~300 and 100 faster than the unaddressed site, respectively (Figure 6E and Table 4). Consistent with our finding that the PvuII(T46G, Y94F) variant has a highly reduced catalytic activity compared to wtPvuII, the variants $P(T46G, Y94F)L_6S^*$ (Figure 6B) and $P(T46G, Y94F)L_+S^*$ (Figure 6D) also have a much lower activity. However, they show only a 10–20-fold faster addressed than unaddressed cleavage rate (Figure 6E and Table 4).

Since I-SceI is the part of the fusion enzymes which should be mainly responsible for specific DNA binding, we wanted to investigate, whether the affinity of the fusion enzymes to the tripartite recognition site (S6P6S) is higher

than the affinity for a 'stand-alone' I-SceI recognition site (S) and is similar to the reported K_D of 9.7 nM for I-SceI (51). For this purpose internally radiolabeled PCR fragments containing either the sites S6P6S or S were used for gel shift experiments under non-cleavage conditions. The affinity of the fusion enzymes for the S6P6S site was slightly lower than the affinity of I-SceI for its recognition site. A 'stand-alone' I-SceI site (S), however, is bound by the fusion enzyme with considerably lower affinity than the tripartite site (S6P6S) (Table 3). This indicates that PvuII is likely to contribute to the binding ability of the fusion enzyme to its specific site but disturbs the binding of I-SceI to a 'stand-alone' I-SceI site.

Optimizing the spacer length within the tripartite target sites with different distances between the PvuII and I-SceI sites

Before testing the PvuII variants with a proposed weakened dimer interface in the fusion constructs, we wanted to find out which linker is the preferred one in PvuII-I-SceI* fusion proteins for cleaving substrates with different distances between the PvuII site and the I-SceI site. For this, the linker variants PL_HS^*, PL_6S^*, PL_NS^*, PL_+S^* or PL_-S^* were assayed for cleavage activity on plasmid DNA substrates containing the target sites S4P4S, S6P6S, S6x6S or S8P8S (Table 2 and Figure 7). All variants cleave the addressed substrate with the 6 bp distance between I-SceI and PvuII site (Figure 7B) with nearly the same efficiency; in contrast, the substrate, which does not contain a PvuII target site (S6×6S) between the I-SceI sites, is not cleaved by any of the variants (Figure 7C). For the substrates with a decreased (S4P4S, Figure 7A) or increased (S8P8S, Figure 7D) distance between the I-SceI and PvuII sites differences in the cleavage efficiency were observed for the linker variants. The variants containing a charged linker (PL_+S^* or PL_-S^*) have a preference for the S4P4S substrate, PL_HS^* and PL_NS^* cleave the substrates with the following preference: S6P6S > S8P8S > S4P4S. PL_6S^* shows the preference predicted from the model (Figure 1B): S4P4S < S6P6S > S8P8S. For this reason the amino acid substitutions L12E, P14G and H15D were introduced only into PL_6S^*.

Comparative analysis of the activity of PvuII-I-SceI fusion proteins in DNA cleavage assays

The specificity of these variants was tested and as expected from the kinetic parameters for PvuII(L12E), PvuII(P14G) and PvuII(H15D) (Supplementary Figure S1), $P(L12E)L_6S^*$ (Figure 8A), $P(P14G)L_6S^*$ (Figure 8B) and $P(H15D)L_6S^*$ (Figure 8C) show a higher specific activity compared to $P(T46G,Y94F)L_6S^*$ and $P(T46G, Y94F)L_+S^*$ resulting in > 1000-fold faster addressed than unaddressed cleavage rate (Table 4). This preference is seen even at 10 times higher concentration of enzyme over the addressed and the unaddressed substrate (Supplementary Figure S2). It needs to be mentioned, however, that PL_6S^*, $P(L12E)L_6S^*$, $P(P14G)L_6S^*$ and $P(H15D)L_6S^*$ also attack substrates containing the dipartite site S6P, although much more slowly than the tripartite site S6P6S (Supplementary Figure S4). The preference for addressed
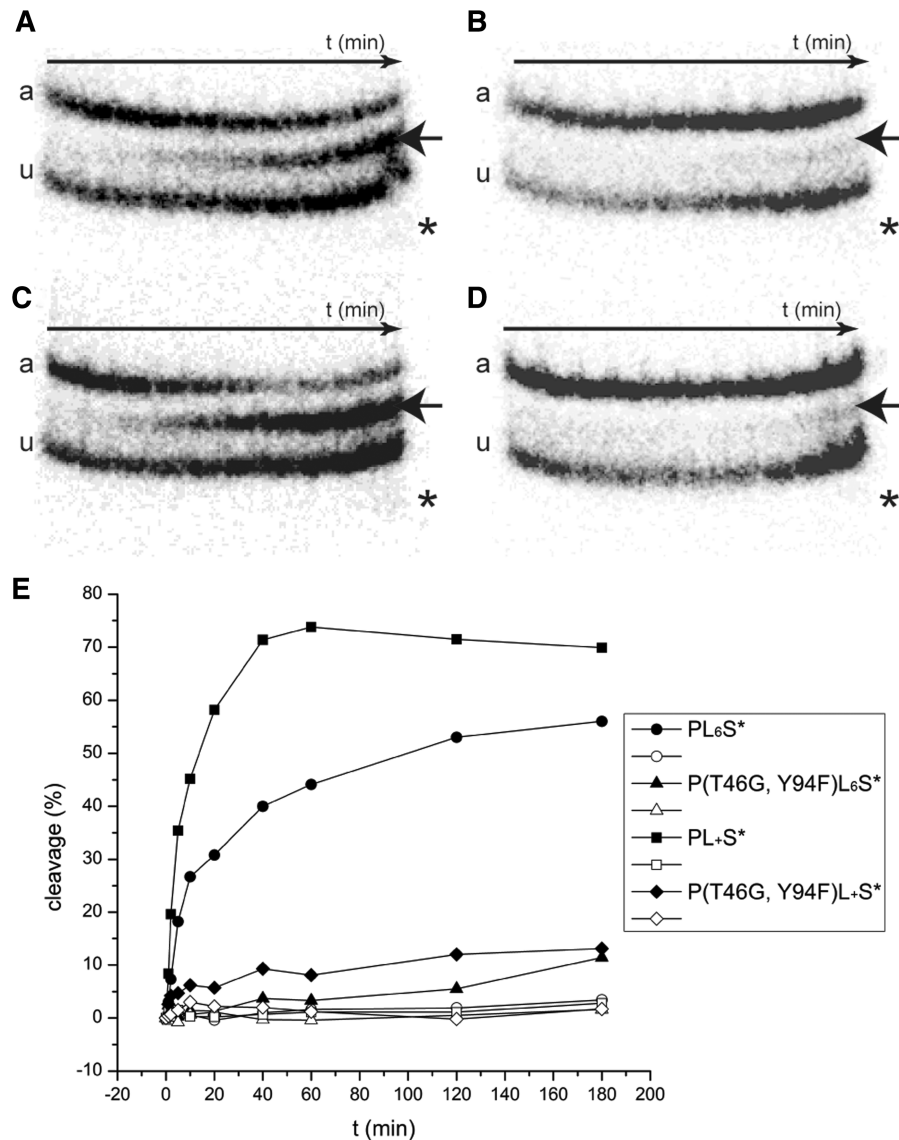
Figure 6. Competition cleavage experiments with fusion enzyme variants. The internally ^{32}P labeled DNA fragments contain either the tripartite recognition sequence S6P6S (a) or a single PvuII site (u), at equimolar concentrations of 10 nM of both substrates. The two substrates were incubated with (**A**) PL_6S^*, (**B**) $P(T46G, Y94F)L_6S^*$, (**C**) PL_+S^* and (**D**) $P(T46G,Y94F)L_+S^*$ for 3 h at 37°C; after defined time intervals samples were withdrawn from the reaction mixture and analyzed by gel electrophoresis. The cleavage product obtained by cleavage at the addressed tripartite recognition site is indicated by an arrow and the position of the expected cleavage product obtained by cleavage at the unaddressed PvuII site by an asterisk. (**E**) The autoradiograms (A–D) were quantified and the percentage of cleavage was plotted against time. Filled symbols show the addressed cleavage and open ones the unaddressed cleavage.

sites could be even higher, since for none of them unaddressed cleavage could be detected, even after overnight incubation (data not shown), which makes these fusion proteins highly specific for the addressed substrate. We determined also the binding affinity of these variants for the S6P6S substrates. The K_D-values obtained for $P(L12E)L_6S^*$, $P(P14G)L_6S^*$ and $P(H15D)L_6S^*$ were of the same magnitude as the K_D of I-SceI (Table 5). On a substrate containing only an I-SceI site no specific binding could be detected (Supplementary Figure S3).

Since off-site target cleavage is a common problem of engineered nucleases of extended specificity (ZFN, meganucleases, etc.), we wanted to test to what extent our

Table 3. Summary of the binding constants K_D determined for PL_6S^*, $P(T46G,Y94F)L_6S^*$, PL_+S^* and $P(T46G,Y94F)L_+S^*$ on DNA substrates containing the tripartite site (S6P6S) or a single I-SceI site (S)

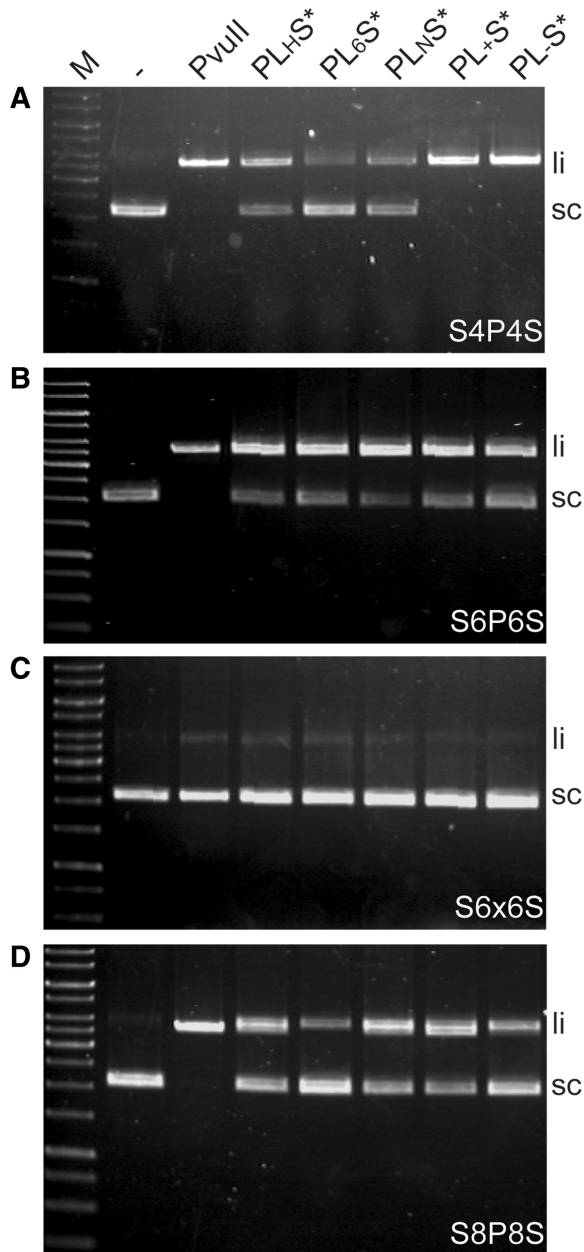| Enzyme	Substrate	K_D (nM)
PL_6S^*	S6P6S	31 ± 3
	S	144 ± 20
$P(T46G,Y94F)L_6S^*$	S6P6S	31 ± 2
	S	119 ± 16
PL_+S^*	S6P6S	18 ± 1
	S	212 ± 29
$P(T46G,Y94F)L_+S^*$	S6P6S	31 ± 3
	S	>250

Figure 7. DNA cleavage experiments with all linker variants and the different substrates at equimolar concentrations (8 nM). (**A**) S4P4S, (**B**) S6P6S, (**C**) S6x6S and (**D**) S8P8S were incubated with PvuII, PL$_H$S*, PL$_6$S*, PL$_N$S*, PL$_+$S* and PL$_-$S*, respectively, for 3 h under optimized conditions (see main text).

engineered fusion enzymes would cleave high molecular DNA at unwanted sites. For our fusion proteins PL$_x$S* the most critical unwanted cleavage would be unaddressed cleavage i.e. cleavage at PvuII sites not addressed by I-SceI sites. For this purpose we used bacteriophage λ-DNA, which contains 15 PvuII sites, in a cleavage experiment (Figure 9). All variants produced, namely the different linker variants, the T46G, Y94F variants of PvuII alone or in the fusion enzyme context and the variants L12E, P14G and H15D also of PvuII alone or in the

fusion enzyme context were used in a 1:1 ratio of enzyme over PvuII sites to see, if genomic DNA is a target for the meganuclease. The variants that we prepared at the beginning of our study (PL$_H$S*, PL$_6$S*, PL$_N$S*, PL$_+$S*, PL$_-$S*) show cleavage of λ-DNA, even though not as complete as observed with wtPvuII. The variants with a decreased catalytic activity [P(T46G,Y94F)L$_6$S*, P(T46G,Y94F)L$_+$S]* show no cleavage of λ-DNA, but since PvuII(T46G,Y94F) also shows reduced cleavage activity compared to wtPvuII, the specific effect (i.e. absence of cleavage) is not as high as for P(L12E)L$_6$S* and P(P14G)L$_6$S* which also do not cleave λ-DNA, whereas PvuII(L12E) and PvuII(P14G) do so. Surprisingly, P(H15D)L$_6$S* shows some cleavage, although it did not cleave the unaddressed substrate in the competition experiment (Figure 8C). We have verified the specificity of some of our fusion constructs also *in vivo* by demonstrating that *E. coli* cells producing fusion proteins are viable even in the absence of PvuII methyltransferase.

We were able to show that the variants P(L12E)L$_6$S* and P(P14G)L$_6$S* do not cleave PvuII sites on high molecular weight DNA and wanted to see finally, if these fusion enzymes have a preference for a certain distance between the I-SceI and PvuII sites. For this purpose, plasmid DNA containing either the S4P4S, S6P6S or S8P8S target was tested for cleavage by P(L12E)L$_6$S*, P(P14G)L$_6$S* and P(H15D)L$_6$S*. The amount of linear DNA produced was determined by densitometry and the mean of three experiments was plotted (Figure 10). All enzymes cleave S6P6S preferentially, but to different degrees. Whereas P(L12E)L$_6$S* has an ∼2-fold preference for S6P6S over S4P4s and S8P8S, and P(H15D)L$_6$S* a 1.5- and 2-fold preference over S4P4S and S8P8S, respectively, P(P14G)L$_6$S* shows a 8–10-fold preference for S6P6S over the other substrates. Taking all the results obtained together, P(P14G)L$_6$S* is the fusion enzyme of choice, which cleaves highly preferentially the addressed substrate in the presence of unaddressed substrate, does not cleave unaddressed PvuII sites in λ-DNA and has a specificity for substrate with two I-SceI sites and a PvuII site in the middle separated by 6 bp on each site. These properties make this enzyme a suitable candidate to be tested for gene targeting in further *in vivo* experiments.

DISCUSSION

Nucleases that are specific enough to recognize and cleave only unique DNA sequences in complex genomes are needed for the purpose of introducing new genetic material into complex genomes by homologous recombination. In gene targeting, new DNA sequences (e.g. complete genes) are introduced by making a specific double-strand cut at the locus of interest, which is repaired by recombination with the new DNA supplied in *trans*. There are two options to achieve this: (i) to use a nuclease that recognizes a sequence that had been specifically introduced into the genome at the locus of interest in a previous genome engineering experiment, or (ii) to reprogram a specific nuclease, or to program a non-specific nuclease to recognize a
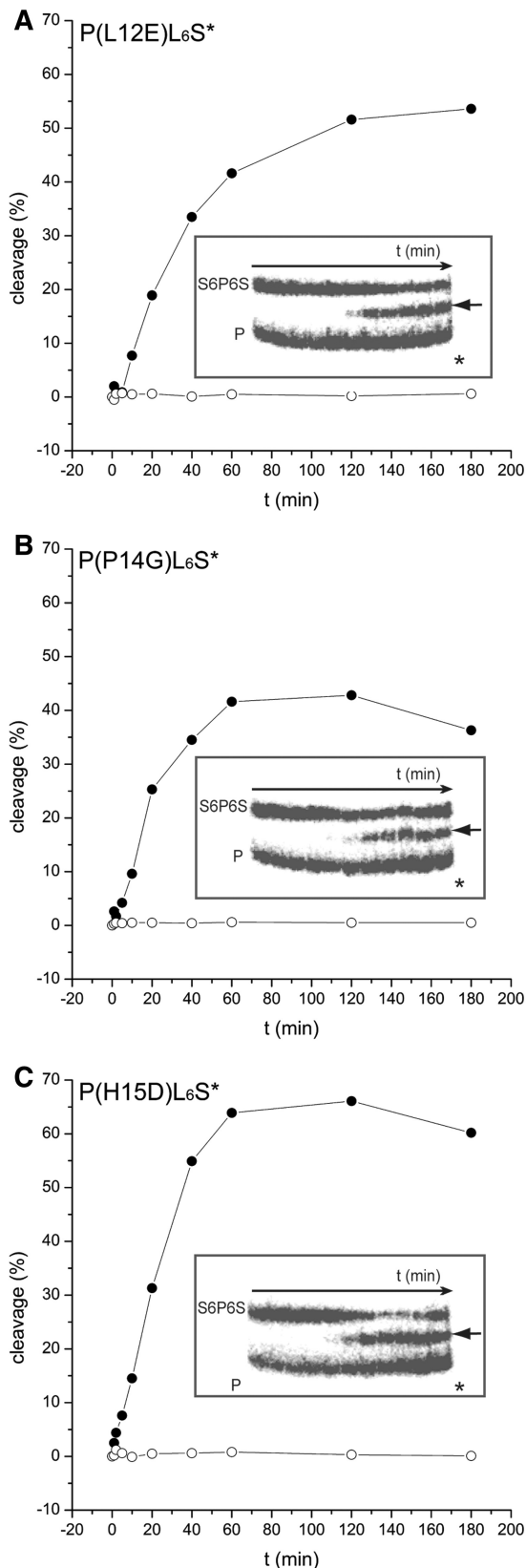
Figure 8. Competition cleavage experiments with the fusion enzyme variants containing amino acid substitutions in the dimer interface and internally ^{32}P labeled DNA fragments containing the addressed tripartite recognition site (S6P6S) or the unaddressed PvuII site alone (P). The experiment was performed as described in Figure 5.

certain sequence at the locus of interest. We have pursued here the first option, namely to produce a nuclease with a unique specificity that could be used in particular for introducing new genes into genomes, in which recognition sequences for highly specific nucleases are already present or had been introduced for future targeted genome optimization as it has been done for example in agronomic crops (52,53).

Nucleases with extended specificity so far have been produced mainly by engineering homing endonucleases ('meganucleases') or by fusing a non-specific nuclease, typically the DNA cleavage domain of the FokI restriction endonuclease and three to four zinc fingers ($\rightarrow$ZFN). In the new approach described here, we utilize the same architecture as utilized for ZFNs, triple helix forming oligonucleotide-linked nucleases, or TALENs, in which a DNA-binding module is fused to a DNA-cleavage module to generate an enzyme, whose distinct properties can be exploited to target a unique sequence in complex genomes. The binding module of choice for our purpose is derived from an inactive (denoted by *) variant of the monomeric homing endonuclease I-SceI (I-Sce*) that specifically recognizes an 18-bp sequence, but because of amino acid substitutions in the active site (Asp44$\rightarrow$Ser or Asn; Asp145$\rightarrow$Ala) is only competent in specific DNA binding but not in cleavage [cf. (20)]. In addition, I-SceI* lacks the nine C-terminal amino acid residues, which are not resolved in the co-crystal structure (40). For the DNA-cleavage module, we employ variants of the homodimeric type II restriction endonuclease PvuII. Different from the non-specific cleavage domain of FokI in the ZFNs, PvuII in our fusion proteins not only supplies the catalytic function, but also, in principle, could increase the overall specificity. Our PvuII-I-SceI* fusion proteins are homodimers (as shown by analytical ultracentrifugation), consisting of the central PvuII homodimer, whose subunits are fused at their C-termini to the N-termini of I-SceI*, which carries a C-terminal Strep-tag (Figure 1A). They bind the tripartite recognition site, consisting of a central PvuII recognition site and two flanking I-SceI recognition sites with higher affinity than a 'stand-alone' I-SceI recognition site.

A critical issue for the usefulness of engineered nucleases with extended specificity, in general, is that cleavage at so-called 'off-target sites' must be prevented (54). For PvuII-I-SceI* fusion proteins it is mandatory that they do not attack an unaddressed PvuII site (i.e. a PvuII site not flanked by I-SceI sites), and should prefer PvuII sites flanked on both sides with an I-SceI site over PvuII sites flanked on only one side with an I-SceI site, although we do not regard cleavage at PvuII sites flanked on only one side with an I-SceI site as critical, given the length of the

Figure 8. Continued
The autoradiogram is shown in the insert, the result of the quantification is shown as a reaction progress curve (percent cleavage *vs.* time): (**A**) P(L12E)L$_6$S*, (**B**) P(P14G)L$_6$S* and (**C**) P(H15D)L$_6$S*. The arrow indicates the cleavage product obtained by addressed cleavage (closed circle) and the asterisk the position of the expected cleavage product of unaddressed cleavage (open circle).

Table 4. Summary of activities on addressed and unaddressed substrates and resulting specificities determined by competition cleavage experiments[a] for all tested fusion enzyme variants

Enzyme	Activity (nM min^{-1})		Specificity
	Addressed	Unaddressed	(Addressed/ unaddressed)
PL_6S^*	0.88 ± 0.07	0.0026 ± 0.0004	338
PL_+S^*	1.00 ± 0.08	0.014 ± 0.004	72
$P(T46G, Y94F)L_6S^*$	0.018 ± 0.005	0.0010 ± 0.0004	18
$P(T46G, Y94F)L_+S^*$	0.025 ± 0.009	0.0023 ± 0.0008	11
$P(L12E)L_6S^*$	0.39 ± 0.05	$<0.0001^b$	>1000
$P(P14G)L_6S^*$	0.47 ± 0.02	$<0.0001^b$	>1000
$P(H15D)L_6S^*$	0.44 ± 0.06	$<0.0001^b$	>1000

[a]carried out in triplicate.
[b]below detection limit.

Table 5. Summary of the binding constants determined for $P(L12E)L_6S^*$, $P(P14G)L_6S^*$ and $P(H15D)L_6S^*$ on DNA substrates containing the tripartite site (S6P6S)

Enzyme	Substrate	K_D (nM)
$P(L12E)L_6S^*$	S6P6S	9 ± 1
$P(P14G)L_6S^*$	S6P6S	9 ± 1
$P(H15D)L_6S^*$	S6P6S	2.3 ± 0.4

Figure 9. DNA cleavage experiments with all fusion enzymes and all PvuII variants and bacteriophage λ-DNA, which harbors 15 PvuII sites. DNA and enzyme concentrations were chosen such, that the molar ratio between enzyme and PvuII sites is 1:1.

Figure 10. Evaluation of DNA cleavage assays with three different plasmid substrates containing the sites S4P4S (light gray), S6P6S (gray) and S8P8S (dark gray). The variants $P(L12E)L_6S^*$, $P(P14G)L_6S^*$ and $P(H15D)L_6S^*$ were tested for cleavage of these substrates in equimolar concentrations of DNA and protein for 3 h at 37°C under optimized conditions. The amount of linear DNA was estimated by densitometry of the ethidium bromide stained agarose gel and the mean and standard deviation were calculated and plotted ($n = 3$).

combined recognition sequences of I-SceI and PvuII. We have employed a 2-fold strategy to achieve the required specificity of our fusion protein using either different linkers or introducing different amino acid substitutions into PvuII.

In a preliminary experiment, we had linked a catalytic-ally inactive version of I-SceI via its N-terminus to the C-terminus of the single chain version of His$_6$-tagged PvuII (scPvuII), in order to confirm that such a construct would not attack a DNA with one I-SceI site and without a PvuII site. To our surprise it turned out that the DNA was cleaved. Sequencing of the product identified the site of cleavage as a PvuII 'star' site (CAGCAG), which differs from a canonical PvuII site (CAGCTG) by a single base pair. Furthermore, at low Mg^{2+} ('physiological') concentrations the PvuII-I-SceI* fusion protein cleaves a plasmid DNA containing a composite site consisting of an I-SceI site and a PvuII site 8 bp downstream much better than scPvuII. These results showed that the scPvuII-I-SceI* fusion protein is functional and that I-SceI* can direct PvuII to a site that is related to a PvuII site. This result prompted us to pursue our experiments with fusion proteins consisting of wtPvuII and I-SceI*.

We had generated a structural model for PvuII-I-SceI* (Figure 1B). It suggested that the fusion protein designed to cleave a quasi-palindromic site consisting of a PvuII site flanked at a distance of ~6 bp by two I-SceI sites oriented in opposite directions could be engineered with two linkers between the N-terminus of I-SceI* and the C-terminus of PvuII, ~6–10 amino acid residues in length. This fusion protein would approach the DNA from one side. Of course, the distance of the PvuII and I-SceI sites and the
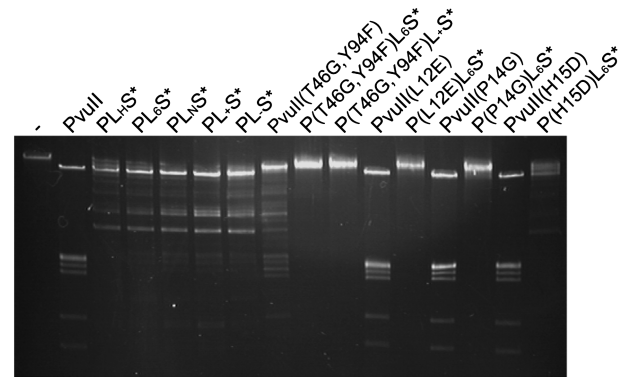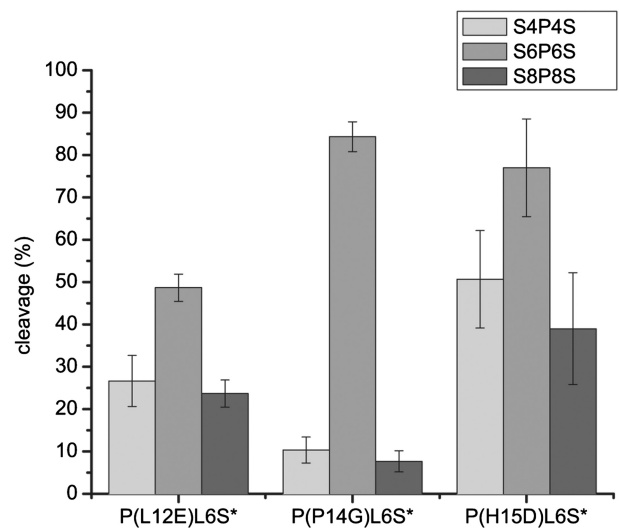
linker length and its amino acid composition needed to be optimized. It turned out that a 6 bp distance and a linker of 6 amino acids (ASRTTG) proved to be optimal. A flexible 10 amino acid linker (ASTKQLVKSG) was also acceptable.

Since PvuII-I-SceI* fusion proteins even with optimized linkers were not absolutely specific for the addressed site, we introduced amino acid substitutions into PvuII that are known to affect K_m and k_{cat}. The intention was to increase the contribution of I-SceI* to specificity relative to PvuII, both in the ground (K_m) and in the transition state (k_{cat}). It is interesting to note that slowing down catalysis by introducing the T46G and Y94F amino acid substitutions into PL_6S^* and PL_+S^* results in a decrease of specificity [compare for example $PL_6S^* \equiv 338$ with

$P(T46G,Y94F)L_6S^* \equiv$ 18]. This can be explained by an 'allosteric coupling' between the binding and the cleavage module in our fusion enzymes. A consequence of this coupling could be that the effect of reduced cleavage activity by the substitutions T46G and Y94F is only seen on the addressed site where PvuII is positioned correctly by the I-SceI mojety. On the unaddressed site this effect is not seen because the fusion enzyme is not able to bind in a catalytically productive manner. This is illustrated for example by our finding that PL_6S^* does not bind to an isolated PvuII site.

Since proper juxtaposition of the two subunits of PvuII is essential for double-strand cleavage we figured that interference with the subunit/subunit interface might affect PvuII activity but not that of the fusion protein, which is kept in a catalytically proficient dimer state by I-SceI* binding to its recognition sites. Although amino acid substitutions in the subunit/subunit interface did not abolish the DNA cleavage activity of PvuII, indicating that PvuII remains dimeric, they increased the specificity of the PvuII-I-SceI* fusion protein so much that no un-addressed DNA cleavage could be observed anymore. Presumably, the amino acid substitutions in the PvuII dimer interface impair DNA access for the catalytic center of PvuII in the fusion enzyme if the I-SceI/DNA interaction does not take place and thereby cannot support productive dimer formation of PvuII on the DNA.

In summary, given the detection limit of our DNA cleavage experiments, the preference of our best PvuII-I-SceI* fusion proteins is better than 1000 for an addressed site over an unaddressed site.

CONCLUSIONS

We have produced a fusion protein consisting of a LAGLIDADG homing endonuclease and a type IIP restriction enzyme for highly specific DNA cleavage as it is required for gene targeting. Nucleases with extended specificity that contain a DNA binding module fused to a DNA cleavage module invariably make use of the non-specific DNA cleavage domain of the type IIS restriction endonuclease FokI, as exemplified by the zinc finger nucleases. They have the advantage of being programmable by choosing appropriate zinc fingers to recognize series of trinucleotide sequences. The ZFN concept has been very successful, in particular after amino acid substitutions were introduced into the subunit-subunit interface in order to prevent homodimerization (55,56). Still, the problem of off-site cleavage, although very much reduced, still exists, as was systematically analyzed in two very recent studies (57,58).

We have wondered, whether other DNA cleavage modules could be used instead of the non-specific DNA cleavage domain of FokI to widen the repertoire of modules to be combined in engineering chimeric nucleases. In principle, also a specific DNA-cleavage module, for example a type IIP restriction endonuclease, could be used, as long as one can make DNA cleavage absolutely dependent on the binding of the DNA binding module.

In this case one could take advantage of the specificity mediated by the cleavage module in addition to the DNA recognition by the DNA binding module. Our study shows that this is possible, as demonstrated for the tripartite site comprising the recognition sequences of the I-SceI homing endonuclease and the PvuII restriction endonuclease. To achieve the needed extraordinary specificity it is necessary to optimize the DNA binding module, the linker and the DNA cleavage module. Presumably, every fusion construct must be optimized individually. Our fusion enzyme is not freely programmable. However, for gene targeting purposes, one may introduce extended recognition sequences sites in the genome of interest at a so called 'safe harbor' locus, which can then be used reliably for genome engineering.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online: Supplementary Material and Methods, Supplementary Table 1, Supplementary Figures 1–4 and Supplementary References [38 and 59].

REFERENCES

1. Smih,F., Rouet,P., Romanienko,P.J. and Jasin,M. (1995) Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Res.*, **23**, 5012–5019.
2. Choulika,A., Perrin,A., Dujon,B. and Nicolas,J.F. (1995) Induction of homologous Recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **15**, 1968–1973.
3. Pingoud,A. and Wende,W. (2011) Generation of novel nucleases with extended specificity by rational and combinatorial strategies. *ChemBioChem*, **12**, 1495–1500.
4. Wu,J., Kandavelou,K. and Chandrasegaran,S. (2007) Custom-designed zinc finger nucleases: what is next? *Cell. Mol. Life Sci.*, **64**, 2933–2944.
5. Carroll,D. (2008) Progress and prospects: zinc-finger nucleases as gene therapy agents. *Gene Ther.*, **15**, 1463–1468.
6. Cathomen,T. and Joung,J.K. (2008) Zinc-finger nucleases: the next generation emerges. *Mol. Ther.*, **16**, 1200–1207.
7. Urnov,F.D., Rebar,E.J., Holmes,M.C., Zhang,H.S. and Gregory,P.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.*, **11**, 636–646.
8. Klug,A. (2010) The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Q. Rev. Biophys.*, **43**, 1–21.
9. Handel,E.M. and Cathomen,T. (2011) Zinc-finger nuclease based genome surgery: it's all about specificity. *Curr. Gene Ther.*, **11**, 28–37.

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

11. Galetto,R., Duchateau,P. and Paques,F. (2009) Targeted approaches for gene therapy and the emergence of engineered meganucleases. *Expert Opin. Biol. Ther.*, **9**, 1289–1303.

12. Marcaida,M.J., Munoz,I.G., Blanco,F.J., Prieto,J. and Montoya,G. (2010) Homing endonucleases: from basics to therapeutic applications. *Cell. Mol. Life Sci.*, **67**, 727–748.

13. Stoddard,B.L. (2011) Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. *Structure*, **19**, 7–15.

14. Eisenschmidt,K., Lanio,T., Simoncsits,A., Jeltsch,A., Pingoud,V., Wende,W. and Pingoud,A. (2005) Developing a programmed restriction endonuclease for highly specific DNA cleavage. *Nucleic Acids Res.*, **33**, 7039–7047.

15. Simon,P., Cannata,F., Concordet,J.P. and Giovannangeli,C. (2008) Targeting DNA with triplex-forming oligonucleotides to modify gene sequence. *Biochimie*, **90**, 1109–1116.

16. Tsai,T.L., Shieh,D.B., Yeh,C.S., Tzeng,Y., Htet,K., Chuang,K.S., Hwu,J.R. and Su,W.C. (2010) The down regulation of target genes by photo activated DNA nanoscissors. *Biomaterials*, **31**, 6545–6554.

17. Kim,Y.G. and Chandrasegaran,S. (1994) Chimeric restriction endonuclease. *Proc. Natl Acad. Sci. USA*, **91**, 883–887.

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

19. Kandavelou,K., Mani,M., Durai,S. and Chandrasegaran,S. (2004) Engineering and applications of chimeric nucleases. In Pingoud,A. (ed.), *Restriction Endonucleases*. Springer, Berlin, pp. 413–434.

20. Lippow,S.M., Aha,P.M., Parker,M.H., Blake,W.J., Baynes,B.M. and Lipovsek,D. (2009) Creation of a type IIS restriction endonuclease with a long recognition sequence. *Nucleic Acids Res.*, **37**, 3061–3073.

21. Fomenkov,A., Too,P.H., Chan,S.H., Vaisvila,R., Cantin,B.A., Mazzola,L., Tam,V. and Xu,S.Y. (2008) Targeting DNA 5mCpG sites with chimeric endonucleases. *Anal. Biochem.*, **381**, 135–141.

22. Chan,S.H., Bao,Y., Ciszak,E., Laget,S. and Xu,S.Y. (2007) Catalytic domain of restriction endonuclease BmrI as a cleavage module for engineering endonucleases with novel substrate specificities. *Nucleic Acids Res.*, **35**, 6238–6248.

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

24. Boch,J., Scholze,H., Schornack,S., Landgraf,A., Hahn,S., Kay,S., Lahaye,T., Nickstadt,A. and Bonas,U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509–1512.

25. Scholze,H. and Boch,J. (2011) TAL effectors are remote controls for gene activation. *Curr. Opin. Microbiol.*, **14**, 47–53.

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

27. Christian,M., Cermak,T., Doyle,E.L., Schmidt,C., Zhang,F., Hummel,A., Bogdanove,A.J. and Voytas,D.F. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, **186**, 757–761.

28. Li,T., Huang,S., Jiang,W.Z., Wright,D., Spalding,M.H., Weeks,D.P. and Yang,B. (2011) TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.*, **39**, 359–372.

29. Miller,J.C., Tan,S., Qiao,G., Barlow,K.A., Wang,J., Xia,D.F., Meng,X., Paschon,D.E., Leung,E., Hinkley,S.J. *et al.* (2011) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.*, **29**, 143–148.

30. Zhang,F., Cong,L., Lodato,S., Kosuri,S., Church,G.M. and Arlotta,P. (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat. Biotechnol.*, **29**, 149–153.

31. Wood,A.J., Lo,T.W., Zeitler,B., Pickle,C.S., Ralston,E.J., Lee,A.H., Amora,R., Miller,J.C., Leung,E., Meng,X. *et al.* (2011) Targeted genome editing across species using ZFNs and TALENs. *Science*, **333**, 307.

32. Mussolino,C., Morbitzer,R., Lutge,F., Dannemann,N., Lahaye,T. and Cathomen,T. (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res.*, August 5 (doi:10.1093/nar/gkr1597; epub ahead of print).

33. Cermak,T., Doyle,E.L., Christian,M., Wang,L., Zhang,Y., Schmidt,C., Baller,J.A., Somia,N.V., Bogdanove,A.J. and Voytas,D.F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.*, **39**, e82.

34. Gingeras,T.R., Greenough,L., Schildkraut,I. and Roberts,R.J. (1981) Two new restriction endonucleases from Proteus vulgaris. *Nucleic Acids Res.*, **9**, 4525–4536.

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

36. Jacquier,A. and Dujon,B. (1985) An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell*, **41**, 383–394.

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

38. Simoncsits,A., Tjornhammar,M.L., Rasko,T., Kiss,A. and Pongor,S. (2001) Covalent joining of the subunits of a homodimeric type II restriction endonuclease: single-chain PvuII endonuclease. *J. Mol. Biol.*, **309**, 89–97.

39. Tao,T. and and Blumenthal,R.M. (1992) Sequence and characterization of pvuIIR, the PvuII endonuclease gene, and of pvuIIC, its regulatory gene. *J. Bacteriol.*, **174**, 3395–3398.

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

41. Gruen,M., Chang,K., Serbanescu,I. and Liu,D.R. (2002) An in vivo selection system for homing endonuclease activity. *Nucleic Acids Res.*, **30**, e29.

42. Joly,E. and Kirsch,R.D. (1998) An improved PCR-mutagenesis strategy for two-site mutagenesis or sequence swapping between related genes. *Nucleic Acids Res.*, **26**, 1848–1850.

43. Pace,C.N., Vajdos,F., Fee,L., Grimsley,G. and Gray,T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.*, **4**, 2411–2423.

44. McClelland,M., Hanish,J., Nelson,M. and Patel,Y. (1988) KGB: a single buffer for all restriction endonucleases. *Nucleic Acids Res.*, **16**, 364.

45. Schuck,P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.*, **78**, 1606–1619.

46. Laue,T.M., Shah,B.D., Ridgeway,T.M. and Pelletier,S.L. (1992) Computer-aided interpretation of analytical sedimentation data for proteins. In Harding,S.E., Rowe,A.J. and Horton,J. (eds), *In Analytical Ultracentrifugation in Biochemistry and Polymer Science*. Royal Society of Chemistry Cambridge, England, pp. 90–125.

47. Cheng,X., Balendiran,K., Schildkraut,I. and Anderson,J.E. (1994) Structure of PvuII endonuclease with cognate DNA. *EMBO J.*, **13**, 3927–3935.

48. Athanasiadis,A., Vlassi,M., Kotsifaki,D., Tucker,P.A., Wilson,K.S. and Kokkinidis,M. (1994) Crystal structure of PvuII endonuclease reveals extensive structural homologies to EcoRV. *Nat. Struct. Biol.*, **1**, 469–475.

49. Lebowitz,J., Lewis,M.S. and Schuck,P. (2002) Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci.*, **11**, 2067–2079.

50. Spyridaki,A., Matzen,C., Lanio,T., Jeltsch,A., Simoncsits,A., Athanasiadis,A., Scheuring-Vanamee,E., Kokkinidis,M. and Pingoud,A. (2003) Structural and biochemical characterization of a new Mg(2+) binding site near Tyr94 in the restriction endonuclease PvuII. *J. Mol. Biol.*, **331**, 395–406.

51. 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.

52. D'Halluin,K., Vanderstraeten,C., Stals,E., Cornelissen,M. and Ruiter,R. (2008) Homologous recombination: a basis for targeted

genome optimization in crop species such as maize. *Plant Biotechnol. J.*, **6**, 93–102.

53. Yang,M., Djukanovic,V., Stagg,J., Lenderts,B., Bidney,D., Falco,S.C. and Lyznik,L.A. (2009) Targeted mutagenesis in the progeny of maize transgenic plants. *Plant Mol. Biol.*, **70**, 669–679.

54. Petek,L.M., Russell,D.W. and Miller,D.G. (2010) Frequent endonuclease cleavage at off-target locations in vivo. *Mol. Ther.*, **18**, 983–986.

55. Miller,J.C., Holmes,M.C., Wang,J., Guschin,D.Y., Lee,Y.L., Rupniewski,I., Beausejour,C.M., Waite,A.J., Wang,N.S., Kim,K.A. *et al.* (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.*, **25**, 778–785.

56. Szczepek,M., Brondani,V., Buchel,J., Serrano,L., Segal,D.J. and Cathomen,T. (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat. Biotechnol.*, **25**, 786–793.

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

58. Gabriel,R., Lombardo,A., Arens,A., Miller,J.C., Genovese,P., Kaeppel,C., Nowrouzi,A., Bartholomae,C.C., Wang,J., Friedman,G. *et al.* (2011) An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat. Biotechnol.*, **29**, 816–823.

59. Huang,Q. and Quinones,E. (2008) A spectroscopic method to determine the activity of the restriction endonuclease EcoRV that involves a single reaction. *Anal. Biochem.*, April 27 (doi:10.1016/ j.ab.2008.04.038; epub ahead of print).