

Creation of a type IIS restriction endonuclease with a long recognition sequence

Shaun M. Lippow*, Patti M. Aha, Matthew H. Parker, William J. Blake, Brian M. Baynes and Daša Lipovšek

Codon Devices, Inc., 99 Erie Street, Cambridge, MA, 02139, USA

Received February 4, 2009; Revised February 27, 2009; Accepted March 5, 2009

ABSTRACT

Type IIS restriction endonucleases cleave DNA outside their recognition sequences, and are therefore particularly useful in the assembly of DNA from smaller fragments. A limitation of type IIS restriction endonucleases in assembly of long DNA sequences is the relative abundance of their target sites. To facilitate ligation-based assembly of extremely long pieces of DNA, we have engineered a new type IIS restriction endonuclease that combines the specificity of the homing endonuclease I-SceI with the type IIS cleavage pattern of FokI. We linked a non-cleaving mutant of I-SceI, which conveys to the chimeric enzyme its specificity for an 18-bp DNA sequence, to the catalytic domain of FokI, which cuts DNA at a defined site outside the target site. Whereas previously described chimeric endonucleases do not produce type IIS-like precise DNA overhangs suitable for ligation, our chimeric endonuclease cleaves double-stranded DNA exactly 2 and 6 nt from the target site to generate homogeneous, 5', four-base overhangs, which can be ligated with 90% fidelity. We anticipate that these enzymes will be particularly useful in manipulation of DNA fragments larger than a thousand bases, which are very likely to contain target sites for all natural type IIS restriction endonucleases.

INTRODUCTION

Restriction endonucleases, which cleave specific DNA sequences, are ubiquitous in prokaryotes (1), where their main function is defense against bacteriophage infection (2). Type II restriction endonucleases (3), which recognize specific sequences of 4–8 bp and cleave DNA into predictable, homogeneous fragments, have become invaluable tools for DNA analysis and manipulation. A subset of these enzymes, the type IIS restriction endonucleases, cleave DNA outside of their target sequence, and thus

can generate overhangs with sequence unrelated to the target sequence. This property makes type IIS restriction endonucleases particularly useful in the assembly of DNA with specified sequence from smaller fragments, where fragments with matching type-IIS-generated ends are annealed and ligated, leaving an assembled DNA product without a target-site scar at the ligation junction (Figure 1).

A limitation of type IIS restriction endonucleases in long DNA assembly is the relative abundance of their target sites. Type IIS restriction endonucleases recognize non-palindromic sequences of 5, 6 or 7 bp, which are found at an average frequency of one in 512 ($4^5/2$), 2048 ($4^6/2$), or 8192 ($4^7/2$) base pairs, respectively. While it is relatively easy to identify a type IIS restriction endonuclease that does not cut inside a typical gene-sized DNA fragment of approximately 1 kb (4), significantly longer fragments are likely to contain all known type IIS target sequences, and thus cannot be easily excised from a vector in order to be joined to another fragment. This is becoming a major limitation in the postgenomic era, when the availability of genomic information has led to the interest in the synthesis and manipulation of long DNA fragments such as operons, eukaryotic regulatory elements, chromosomes, and entire genomes (5–13).

We set out to design a novel type IIS restriction endonuclease with a significantly longer target site than that of naturally occurring enzymes of this class, to be used in ligation-based assembly of extremely long (kb–Mb) pieces of DNA. To this end we linked the DNA-cleavage domain of the type IIS restriction endonuclease FokI to a non-cleaving mutant of the homing endonuclease I-SceI (Figure 2).

FokI from *Flavobacterium okeanoikoites*, the most thoroughly studied type IIS restriction endonuclease, recognizes a 5-bp stretch of DNA and cuts the two DNA strands 9 and 13 bases (9/13) downstream from the target sequence, generating four-base, 5' overhangs (14). The structure of FokI consists of three units: the DNA-binding domain, which comprises three helix-turn-helix subdomains (two of which specifically bind DNA); a fifteen-residue linker; and a 196-residue, C-terminal,

*To whom correspondence should be addressed. Tel: +1 617 871 1174; Fax: +1 617 876 0673; Email: slippow@gmail.com

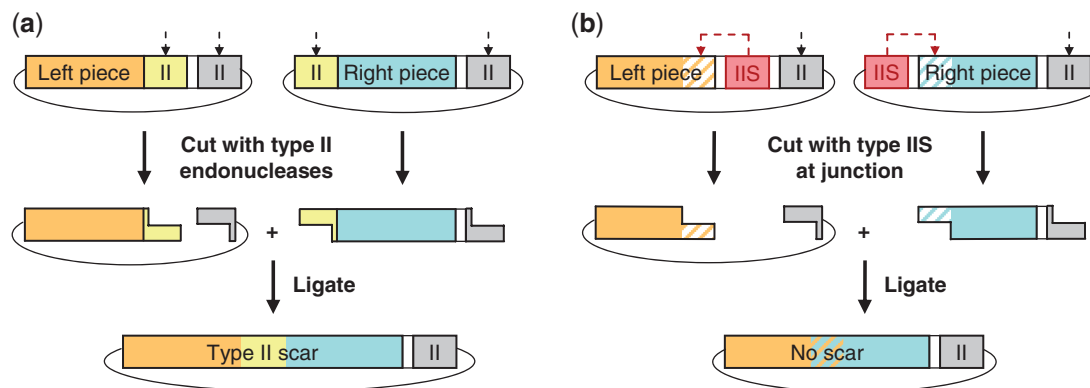


Figure 1. Ligation-based DNA assembly using restriction endonucleases. (a) DNA assembly using conventional type II endonucleases requires a type II recognition site at the junction between the two starting pieces of DNA. The junction remains in the product, precluding the construction of arbitrary DNA sequences. (b) DNA assembly using type IIS endonucleases, which allow the generation of overhangs of any arbitrary sequence, and thus sequence-independent construction.

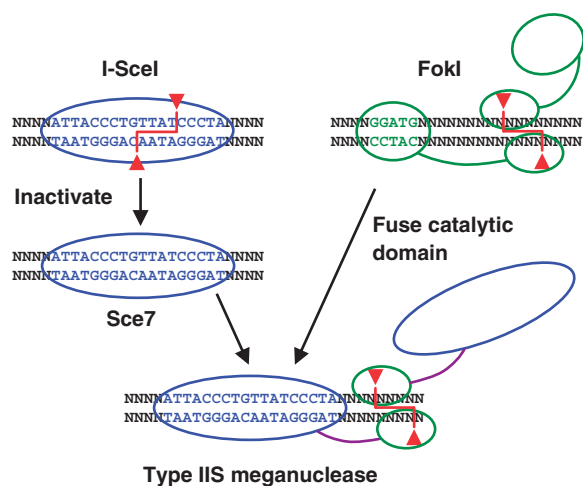


Figure 2. Overview of strategy for engineering a type IIS restriction endonuclease with a long recognition site. In the first step, the wild-type I-SceI homing endonuclease is mutated into a variant (Sce7) that binds but does not cleave DNA. Second, hybrid enzymes are constructed between Sce7 and the catalytic domain of the type IIS endonuclease FokI, using designed polypeptide linkers.

catalytic domain. In the crystal structure of the FokI–DNA complex (15), the catalytic domain is bound to the DNA-recognition domain, instead of to DNA, revealing a putative mechanism in which sequestration of the catalytic domain may contribute to the specificity of DNA cleavage (16). The structure of the FokI dimer in the absence of DNA indicates that dimerization is mediated by the catalytic domains (17), and computational modeling has been used to explain how each of the catalytic domains in the dimer are positioned to cleave one DNA strand (17). Biochemical evidence indicates that FokI cuts DNA as a dimer (18), with the DNA-recognition domain of one monomer associated with the FokI target sequence. The role of the DNA-recognition domain of the second FokI monomer is not yet clear, but it may be involved in stabilizing the dimer (19). Recent work shows that a second FokI target site can lead to FokI-mediated DNA

looping between the two sites (20–22) and enhance cleavage by at least 10-fold (23,24).

The domain-level separation of DNA-recognition and DNA-cleavage functions of FokI makes it an attractive starting point for the design of restriction endonucleases with novel specificities. The catalytic domain of FokI has previously been fused to different DNA-recognition domains such as *Drosophila* Ultrabithorax homeodomain (25), zinc finger domains (26–29), $Z\alpha$ Z-DNA binding domain (30), yeast Gal4 protein (31), hepatocyte nuclear factor-3 β protein (32), plant APETALA3 (33) and papillomavirus E2 protein (34). The resulting chimeric enzymes were found to cleave DNA outside of the sequence recognized by the DNA-recognition domain, but with relatively low activity and specificity. In particular, since the chimeric enzymes cut DNA at a number of adjacent positions, they do not generate well-defined overhangs such as those generated by natural type IIS restriction endonucleases, and thus cannot be used to cut and paste specific DNA sequences with high fidelity. Zinc-finger nucleases with higher efficiency were constructed by combining two DNA target sites, in a face-to-face orientation, by using a shorter linker between the binding and cleavage domains (35–37). Such an arrangement led to a double-stranded break between the two DNA target sites, and is now widely used for gene targeting (38–56). However, because the site of DNA cleavage is internal to the overall target site, these zinc-finger nucleases also cannot be used for ligation-based DNA assembly in place of type IIS endonucleases.

In order to generate a type IIS restriction endonuclease with an extremely long target site, we fused the catalytic domain of FokI to an inactive site, we fused the catalytic domain of FokI to an inactive mutant of the I-SceI homing endonuclease, an enzyme with one of the longest known target sites (57). I-SceI is encoded on a group I mitochondrial intron of *Saccharomyces cerevisiae* (58). Like other intron-encoded homing endonucleases of the LAGLIDADG family, I-SceI catalyzes the double-stranded DNA break required for the lateral transfer of the intron that encodes it to a homologous allele that lacks that intron (59). I-SceI recognizes an 18-bp DNA sequence

and cleaves within this target sequence, generating four-base, 3' overhangs (60–63). Whereas the 18-bp target site suggests that I-SceI would recognize approximately one site in a random sequence of 4^{18} (10^{11}) base pairs, I-SceI has been shown to tolerate some nucleotide substitutions in its target site, leading to the estimate of about one target site in a random sequence of 10^8 bp (64,65).

To convert I-SceI, which resembles a type II restriction endonuclease with an extremely long target site, into the DNA-binding component of a new type IIS restriction endonuclease, we first mutated two essential active-site aspartates in I-SceI (66) to generate a series of non-catalytic mutants. The most highly expressed non-catalytic mutant, which still bound the I-SceI target sequence, was then fused to the native catalytic domain of FokI using either the native FokI linker or a series of designed linkers. We found that the chimeric enzyme with the native FokI linker (named CdnDI) and the chimeric enzyme with a designed, 20-residue linker (named CdnDII) both had type IIS endonuclease activity, cleaving double-stranded DNA at a specific distance from the I-SceI target site. Here we present the design and biochemical characterization of these enzymes, as well as the proof-of-concept for their utility in ligation-based DNA assembly.

MATERIALS AND METHODS

Modeling and design of chimeric enzymes

Structural models were built using the Schrödinger software suite (Schrödinger, LLC). The FokI dimer [2FOK (17)] was structurally aligned to the DNA-bound BamHI dimer [1BHM (67)] using residues 418–579 of FokI and residues 1–180 of BamHI, leading to the model of a dimer of FokI catalytic domains bound to DNA.

The FokI dimer was positioned on the DNA downstream of the FokI target site [1FOK (15)] by aligning the DNA backbone adjacent to the original BamHI cleavage site in the dimer complex to the known 9/13 site of FokI cleavage (1FOK: B913–B920, C922–C929; 1BHM: C1–C8, D5–D12). The wild-type FokI linker, previously defined as residues 373–387 (15), was modeled as follows: The backbone dihedral angles of residues 382–387 were modified to α -helical, and the backbone dihedral angles of residues 376–382 were adjusted to create a new backbone turn to place the 382–387 helix in line with helical residues 388–399 of the catalytic domain. Residues 375–389 were minimized and, separately, the Prime Refine Loops function was applied to generate a series of low-energy loop conformations. Alternate FokI models with the catalytic dimer at different positions downstream of the target site were generated by aligning the same 1BHM DNA base pairs to shifted base pairs of 1FOK. To model the chimeric enzymes, similar alignments were used to position the FokI dimer on DNA downstream of I-SceI [1R7M (66)].

Construction of active-site mutants of I-SceI and of chimeric endonucleases

All DNA constructs used in this study were assembled from synthetic oligonucleotides using the Codon Devices

commercial PCR- and ligation-based DNA-assembly methods and error-correction technology (68). The eight I-SceI variant genes constructed in addition to the wild-type I-SceI gene contained all possible single and double mutations of Asp 44 and Asp 145 to either alanine or asparagine (Figure 1S). Genes for the chimeric endonucleases were assembled from an upstream fragment of DNA encoding the inactive, DNA-binding I-SceI variant Sce7 (D44N, D145A); a middle fragment of DNA encoding a 10-, 15- or 20-residue linker; and a downstream fragment of DNA encoding the wild-type catalytic domain of FokI. The protein sequences of the chimeric endonucleases described in this report are detailed in the legend of Table 1. Each I-SceI variant gene and chimeric-endonuclease gene was cloned between the NcoI and XhoI sites of vector pBAD-His (A) (Invitrogen, Carlsbad, CA), transformed into Top10 OneShot cells (Invitrogen), and plated onto LB agar with 100 μ g/ml of carbenicillin (LB/carb).

Screen for high expression level of I-SceI mutant proteins in *E. coli*

To test the expression level of each I-SceI mutant, individual colonies were used to inoculate 1 ml of LB/carb, and grown overnight, with shaking, at 37°C. Ninety microliters of the overnight culture was added to 3 ml of fresh LB, and grown for 2.5–3 h, to mid-log phase, at 37°C. Arabinose was added to each culture to 0.02%, and the cultures were incubated for three more hours at 37°C. The cell pellets were recovered by centrifugation (10 min at 3220g, 4°C), stored frozen at –20°C, then thawed on ice and lysed in 350 μ l of 50 mM sodium phosphate, pH 7.0, 0.5 M NaCl, 50 mM MgCl₂, 0.5 mg/ml lysozyme, 0.05 mg/ml DNaseI, and 1 \times EDTA-free COMPLETE protease inhibitor cocktail (Roche, Indianapolis, IN). Cellular debris was removed by centrifugation for 2 min at 16 000g, 4°C. Thirty microliters of 50% slurry of TALON Superflow Metal Affinity Resin (Clontech, Mountain View, CA), pre-equilibrated with 50 mM sodium phosphate, pH 7.0, 0.5 M NaCl and 50 mM MgCl₂, were added to each, and the mixtures were rocked for an hour at 4°C. Each resin was washed three times with 400 μ l of ice-cold PBS, pH 7.4, then resuspended in 25 μ l of PBS. The resuspended resin was combined with reducing SDS–PAGE buffer, boiled at 99°C for 5 min, separated on 4–12% Bis/Tris gradient gels in MES-based buffer (Invitrogen), and detected with GelCode Blue (Pierce, Rockford, IL) (Figure 1S).

Purification of I-SceI variant Sce7 and chimeric endonucleases

A single colony harboring the sequence-verified gene of interest was grown in 100 ml LB/carb overnight, with shaking, at 37°C. The overnight culture of Sce7 was diluted 1/33 into the final volume of 21 of fresh LB/carb, grown in shaker flasks for 2.5–3 h, at 37°C, to mid-log phase, induced by adding arabinose to 0.02%, grown for three more hours at 37°C, and harvested by centrifugation for 30 min at 4785 \times g, 4°C. In contrast, chimeric endonucleases were expressed in 4–61 of LB/carb at 25°C,

for 16–18 h. The cell pellets were stored at -80°C , then thawed at 4°C in 1/50 of the original culture volume of 10 mM HEPES, pH 8.0, 1 M NaCl, 1 mM DTT, 25 mM imidazole, $1\times$ EDTA-free COMPLETE protease inhibitor cocktail and 120 $\mu\text{g}/\text{ml}$ lysozyme. Lysis buffer for purification of chimeric endonucleases also contained 1 mM PMSF. The lysate was sonicated, then clarified by centrifugation for 20 min at 6000g, 4°C , followed by filtration of the supernatant through a 0.45 μm filter. All further purification steps were performed at 4°C .

The clarified lysate was loaded onto a 1 ml HisTrap column (GE Healthcare, Piscataway, NJ) pre-equilibrated with 10 mM HEPES, pH 8.0, 1 M NaCl, 25 mM imidazole and 1 mM DTT on the AKTA Purifier chromatography system (GE Healthcare), at the flow rate of 1 ml/min. The column was washed with 20 column volumes of the equilibration buffer, then eluted with a linear gradient of 25–100 mM imidazole in equilibration buffer over 30 column volumes. The fractions containing eluted protein were pooled, concentrated using Amicon (Houston, TX) Ultra-15 centrifugal concentrators (10 kDa MWCO) to between approximately 1 ml, and filtered through a 0.2 μm filter. The partially purified protein was loaded onto a HiLoad Superdex 200 16/60 column (GE Healthcare) pre-equilibrated with 20 mM HEPES, pH 8.0, 0.5 M NaCl, 1 mM DTT, 0.1 mM EDTA, and 5% glycerol on the AKTA Purifier. The size-exclusion chromatography step was performed at 0.5 ml/min. Fractions eluted from the column at the volume corresponding to the expected molecular weight (approximately 38 kDa for the Sce7 mutant and 52 kDa for the chimeric endonucleases) were pooled, concentrated using Amicon Ultra-4 centrifugal concentrators (10 kDa MWCO) at 2000g, 4°C (to 160 $\mu\text{g}/\text{ml}$ for Sce7 and between 100 and 300 $\mu\text{g}/\text{ml}$ for the chimeric endonucleases), and stored at -20°C in 10 mM HEPES, pH 8.0, 0.25 M NaCl, 0.5 mM DTT, 0.05 mM EDTA and 50% glycerol.

DNA cleavage assay

To construct a DNA substrate for I-SceI and the chimeric endonucleases, two complementary oligonucleotides containing the I-SceI target site (5'-AATTCTGGTTCCGAA GCCTGTCCTGCACGCTAGGGATAACAGGGTAA TAATATATGAATCCAAACTAGAGCGGGGCTCTT GACGTTTGGCTCAAACGTCGTGAGACAGTTTG GTCAGTTGTAAATATCTAATATTCCAATG-3' and 5'-GATCCATTGGAATATTAGATATTTACAACCTG CCAAACGTCTCACGACGTTTTGAGCCAAACGT CAAGAGCCCCGCTCTAGTTTGGATTCATATATT ATTACCCTGTTATCCCTAGCGT GCAGGACAGGC TTCGGAACCAG-3'; the I-SceI target site is underlined) were annealed, phosphorylated, and ligated between the EcoRI and BamHI restriction sites of pUC19 (Invitrogen). The plasmid, named pSCI, was propagated in and extracted from *E. coli* OneShot Top 10 (Invitrogen), linearized by cleavage with AlwNI (New England Biolabs, Ipswich, MA) and purified using standard isopropanol/acetate precipitation.

To observe cleavage of the linearized, purified pSCI substrate by the chimeric endonucleases, 400 ng (11 nM)

of the plasmid DNA were incubated for 4 h under different reaction conditions. For the chimeric endonuclease CdnDI, the optimal cleavage conditions were 50 nM endonuclease in 20 mM Tris-HCl, pH 9.0, 25 mM NaCl, 10 mM MgCl_2 , 1 mM dithiothreitol, 0.1 mg/ml bovine serum albumin, and 2% glycerol, at 37°C . For the chimeric endonuclease CdnDII, the optimal cleavage conditions were 100 nM endonuclease in 20 mM Tris-HCl, pH 9.0, 25 mM NaCl, 10 mM MgCl_2 , 1 mM dithiothreitol, 0.1 mg/ml bovine serum albumin, and 2% glycerol, at 42°C . After the incubation, the products of the reaction were separated on a 0.8% agarose E-gel (Invitrogen). The DNA bands were visualized with ultraviolet light and quantified by densitometry using an AlphaImager HP imager (AlphaInnotech, San Leandro, CA) and ImageJ software (<http://rsb.info.nih.gov/ij/>).

To test the ability of chimeric endonucleases to cleave supercoiled DNA, 400 ng of circular pSCI were digested with CdnDI or CdnDII under the optimized cleavage conditions listed above, and the products were separated on a 0.8% agarose E-gel (Invitrogen).

Determination of dissociation constants

The DNA reagent for determining the dissociation constants between I-SceI variant Sce7 or chimeric endonucleases and DNA was prepared by annealing two 80-base oligonucleotides, 5'-GAATTCTGGTTCCGAA GCCTGTCCTGCACGCTAGGGATAACAGGGTAA ATAATATATGAATCCAAACTAGAGCGGGGCTC T-3' and 5'-AGAGCCCCGCTCTAGTTTGGATTCATA TATTATTACCCTGTTATCCCTAGCGTGCAGGAC AGGCTTTCGGAACCAGAATTC-3' (the I-SceI target site is underlined).

To measure the dissociation constants, the binding protein of interest, at a range of concentrations between 2 and 400 nM, was incubated with 10 nM DNA substrate in 20 μl of 20 mM Tris-HCl, pH 9.0, 25 mM NaCl, 10 mM CaCl_2 , 1 mM DTT, 2% glycerol and 0.1 mg/ml BSA, for 20 min at 37°C (Sce7 and CdnDI) or 42°C (CdnDII). Two microliters of $10\times$ loading buffer were added to each sample, and 10 μl of each mixture were loaded onto 6% polyacrylamide DNA retardation gels (Invitrogen), and run for 30 min at 175 V. The gels were stained with SYBR Gold (Invitrogen) and visualized with ultraviolet light. DNA-containing bands were quantified by densitometry as described above. Concentrations of bound and free protein were calculated from the input amount of DNA and from ratios of band intensities, and the data were fit using Origin software (Originlab, Northampton, MA) to the following equation:

$$[\text{Bound protein}] = \frac{[\text{DNA}]_{\text{total}} \times [\text{Free protein}]}{(K_d + [\text{Free protein}])}$$

Determination of cleavage sites of chimeric endonucleases

Linearized pSCI was cleaved by CdnDI or CdnDII chimeric endonucleases under optimal cleavage conditions listed above. The resulting 1900-bp and 900-bp

DNA fragments were separated on a 1.2% agarose gel, then extracted from the gel using a Gel Extraction Kit (Qiagen, Valencia, CA). The 1900-bp fragment was sequenced using oligonucleotide primer 5'-ATTGCGCCATTCAGGCTGCGC-3', and the 900-bp fragment was sequenced using oligonucleotide primer 5'-CACTTTATGCTTCCGGCTCG-3'. The end of each fragment was deduced from the point where sequencing data terminated.

Specificity of chimeric endonucleases

Eighteen double-stranded, 80-bp DNA reagents for evaluating endonuclease specificities were constructed as described above for determination of dissociation constants, except that each reagent contained a single base-pair substitution in the I-SceI target site. The following mutations were used in the different positions in the target site: 1: T→C, 2: A→C, 3: G→T, 4: G→T, 5: G→T, 6: A→C, 7: T→C, 8: A→C, 9: A→C, 10: C→T, 11: A→C, 12: G→T, 13: G→T, 14: G→T, 15: T→C, 16: A→C, 17: A→C, 18: T→C. The effect of each substitution on binding was assessed by following the procedure described above for determination of dissociation constants at 37°C, with the following modifications: For each of the 18 mutated DNA reagents and for the original DNA reagent containing wild-type I-SceI target site, each protein was evaluated at a single concentration, which corresponded to approximately three times its K_d for the wild-type I-SceI target sequence (185 nM for I-SceI, 40 nM for Sce7, 50 nM for CdnDI). The percentage of each DNA substrate that bound to each protein was determined by densitometry. Each measurement was conducted three to four times. For each single-substitution substrate, the average percentage bound was normalized to the average percentage bound to the wild-type I-SceI target sequence.

Assembly of DNA fragments generated by chimeric endonucleases

To test the integrity of 5' overhangs generated by the chimeric endonucleases, we carried out ligation-based DNA assemblies between DNA fragments with CdnD- or BsaI-generated overhangs. The DNA fragments to be ligated, Q-Sce, T-Sce, Q-2.567 and T-2.567, were generated by hybridizing and phosphorylating complementary oligonucleotides, yielding double-stranded DNA with the following sequence:

Q-Sce:

TTCATGAGACGATCTCCTTCCTCTTGATGGCTGT
AATAATAGCTCTAGGGCGATGTAAAGACAACG
GATTCATATATTATTACCCTGTTATCCCTAGCGT
GCAGGACAGGCTTCGGAACCTGAG

T-Sce:

TTCAATATATTATTACCCTGTTATCCCTAGCGTG
CAGGACAGGCTTCGGAACCGGAGACGTTGACA
ACATGAAGTAAACAGCGTAAGATGTACCACAT
GAAATTGCGATGAGGAAATCTATGAG

Q-2.567:

TTCATGAGACGATCTCCTTCCTCTTGATGGCTGT
AATAATAGCTCTAGGGCGATGTAAAGACAACCTT

CAGAGGCCATGTGAAGGACGCAACATGTGCAG
GTCTCTGTGCGCTCTTCTGTGTGAG

T-2.567:

TTCAGTGTGAAGAGCGTGCTGAGACCGTGCAT
GACTGCGTCATACCAACAGTAGGAGACGTTGA
CAACATGAAGTAAACAGCGTAAGATGTACCAC
ATGAAATTGCGATGAGGAAATCTATGAG

The fragments were cloned into pUC-based vector pTS (Supplementary Data, Section 3) which had been linearized with BsmBI and BsaI. Vectors based on pTS that contained inserts Q-Sce, T-Sce, Q-2.567 and T-2.567 were named pTS-Q-Sce, pTS-T-Sce, pTS-Q-2.567 and pTS-T-2.567, respectively.

To test ligation, each left-hand insert was generated by digesting donor vector pTS-Q-Sce or pTS-Q-2.567 with BsmBI and either CdnDI, CdnDII, or BsaI. Each right-hand insert was generated by digesting donor vector pTS-T-Sce or pTS-T-2.567 with BtgZI and either CdnDI, CdnDII, or BsaI. Each enzyme was used under its optimal reaction conditions. The unpurified digestion products were combined with the acceptor vector, pCK (Supplementary Data, Section 3), which had been linearized with BsmBI and BsaI. The mixture was ligated using T4 DNA ligase (New England Biolabs) and transformed into One Shot Top 10 chemically competent cells (Invitrogen). The transformation mixture was plated onto LB agar plates with 12.5 µg/ml chloramphenicol and 25 µg/ml kanamycin. Fidelity of ligation junctions was confirmed by DNA sequencing of the entire pCK insert and insert-vector junctions for 48 colonies per ligation.

RESULTS

Identification of the I-SceI-based DNA-binding module

After mutating the catalytic aspartate residues 44 and 145 of the homing endonuclease I-SceI (66) to either asparagine or alanine, we found that all eight single and double mutants were expressed in *E. coli* at a higher level than was wild-type I-SceI, with mutant Sce7 (D44N, D145A) yielding the highest amount of soluble protein (Figure 1S). Purified Sce7 does not cut linear, double-stranded DNA containing the native I-SceI target site (Figure 3a), but it does bind to that DNA sequence (Figure 3b and c).

Modeling and design of chimeric enzymes

In order to gain insight into the design of chimeric endonucleases comprising the FokI cleavage domain, we first investigated the mechanism of wild-type FokI following the approach previously described by Wah *et al.* (17). We were particularly interested in addressing how dimerized FokI cleavage domains make a double-stranded break in DNA precisely 9 and 13 bases downstream of the DNA target site. To compensate for the lack of direct structural information for FokI bound to and cleaving DNA, we merged information from the structure of monomeric FokI bound to DNA in an inactive, sequestered conformation (15); the structure of dimeric FokI without DNA (17); and a structure of BamHI, a structural homolog

to the FokI cleavage domain, complexed with DNA (67). We assumed that BamHI is a useful model for the orientation of the FokI cleavage domains relative to substrate DNA due to the structural similarity between the two

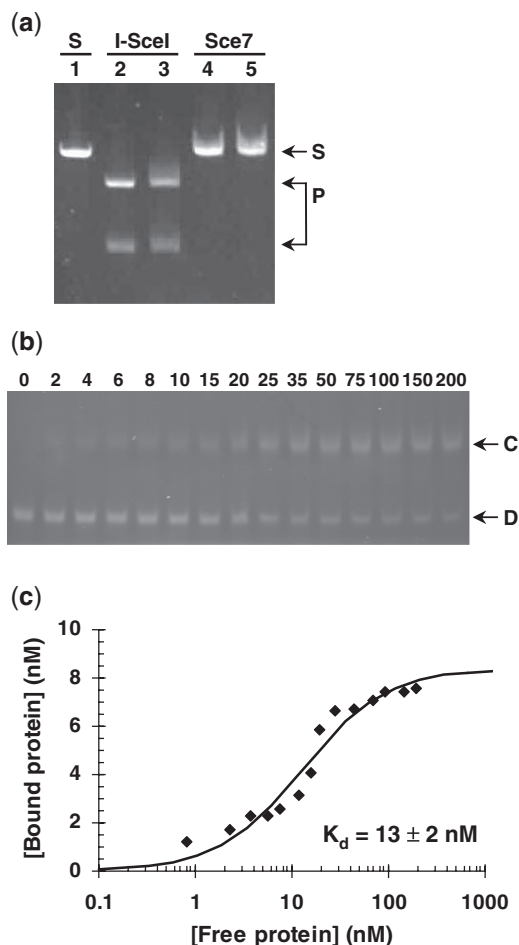


Figure 3. Characterization of Sce7. (a) DNA-cleavage assay using linearized plasmid pSCI (6.8 nM) which contains a single site recognized by I-SceI. Lane 1 (S): linearized plasmid; lanes 2 and 3: I-SceI at 275 nM and 1.1 μ M, respectively. Lanes 4 and 5: Sce7 at 140 nM and 280 nM, respectively. P: DNA fragments resulting from cleavage of pSCI at the I-SceI recognition site. (b) Sce7 DNA-binding assay, showing DNA band retardation with increasing concentrations of Sce7 (shown in nM). D: double-stranded, linear DNA fragment of 80 bp with a single I-SceI recognition site. C: complex between the DNA fragment and Sce7. (c) Sce7 binding-affinity curve fit derived from the gel in (b) using densitometry.

endonucleases and the fact that they both produce four-base, 5' overhangs.

We found that modeling of FokI was able to distinguish the native 9/13 cleavage pattern from incorrect cleavage positions. Only with the dimer of FokI catalytic domains at the 9/13 position can full-length FokI be modeled to both avoid steric clashes between the DNA-recognition and catalytic domains, and to allow the two domains to stay within reach of the 15-residue linker, while preserving key intramolecular contacts between the beginning of the linker and the recognition domain. Because of the intrinsic twist in double-stranded DNA, repositioning the catalytic domains by even 1 bp resulted in a dramatic change in the distances between the termini of the recognition and catalytic domains. With the cleavage domains positioned 10/14, the 15-residue linker would be overly stretched, but a longer linker could be adequate. This analysis is consistent with the results of previous studies in which four or seven residues were inserted in the native linker, leading to cleavage at position 10/14 in addition to 9/13 (69).

Several criteria were considered in designing the linkers to connect the I-SceI variant Sce7 to the FokI catalytic domain. Because our goal was a new type IIS endonuclease that would generate well-defined, ligation-quality overhangs, it was critical that our chimeric enzyme cleave at a distinct location on DNA. In other words, the linker should allow the cleavage domain to reach and cut at only one position downstream of the target site.

Starting with models of dimerized FokI catalytic domains bound to DNA at different positions downstream of the I-SceI/Sce7 target site, we eliminated those models in which Sce7 and either catalytic FokI domain clashed sterically. From the remaining models, we focused on those where the C-terminus of Sce7 was within 35 Å of at least one of the N-termini of the catalytic FokI dimer. Theory for the end-to-end distance of flexible polypeptides as a function of the number of residues (70,71) indicated that our new linkers should be in the range of 10–20 residues. We chose mostly polar and charged residues to maximize the solubility of the linker. To maintain conformational flexibility, we avoided amino acids with a preference for α -helical or β -strand secondary structure (72,73), and included many glycines. In addition to the four novel, designed sequences, which comprised two different lengths and two different amino-acid compositions, we included the native, fifteen-residue linker from wild-type FokI (Table 1).

Table 1. The linkers used in the chimeric I-SceI/FokI enzymes and a summary of experimentally determined enzyme cleavage properties^a

Linker sequence	Source	# AAs	Linker name	Enzyme name	Cleavage at I-SceI target site	Non-specific cleavage	Overhang
QFVIPNRGVTKQLVK	FokI	15	FokL	CdnDI	Yes	None	2/6
GGSGDRDDSDPSDKNDGSGG	Design	20	20D	CdnDII	Yes	Minor	2/6
GGSGGSDGSGNGGSGSGGG	Design	20	20S		Yes	Minor	2/6
GGDSRDSGG	Design	10	10D		Yes	Moderate	ND
GGSGGDGSGG	Design	10	10S		Yes	Extensive	1/5

^aThe sequence of each full-length chimeric protein was: [MGHHHHHHENLYFQSGS] [Residues 3–225 of I-SceI as in pdb 1R7M (66)] [Linker as defined above] [Residues 388–579 of FokI as in pdb 1FOK (15) and 2FOK (17)]. ND: not determined.

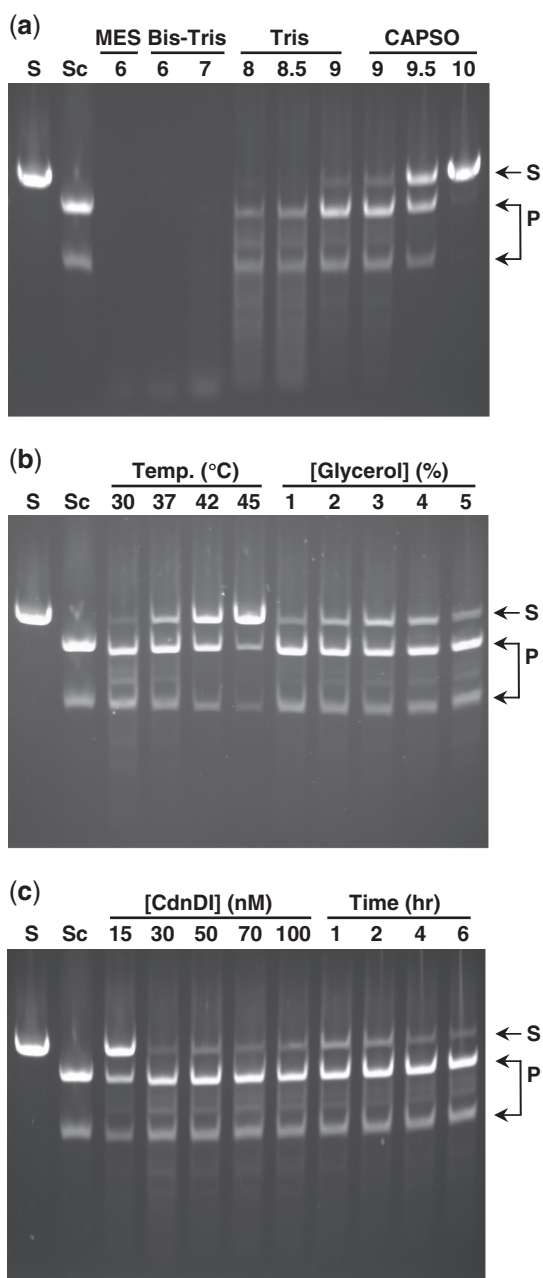


Figure 4. DNA cleavage with the hybrid endonuclease CdnDI near its optimal conditions, showing the effects of single-parameter perturbations. S: linearized plasmid containing a single I-SceI site; Sc: wild-type I-SceI; P: expected products from single cleavage near the I-SceI site. All samples contained 11 nM linearized plasmid. Samples with CdnDI were incubated at its optimal conditions (4h, 50 nM CdnDI, 20 mM Tris-HCl, pH 9.0, 25 mM NaCl, 10 mM MgCl₂, 1 mM dithiothreitol, 0.1 mg/ml bovine serum albumin, 2% glycerol, 37°C) except for the single parameter variation as indicated. (a) Effect of pH. (b) Effect of temperature and % glycerol. (c) Effect of CdnDI concentration and time.

Further details of the modeling and design procedures used and the results obtained are described in Supplementary Data, Section 2, Table 1S, and Figures 2S and 3S.

Expression and activity of chimeric enzymes

We found that the activity and specificity of the chimeric enzymes on linearized plasmid DNA containing a single

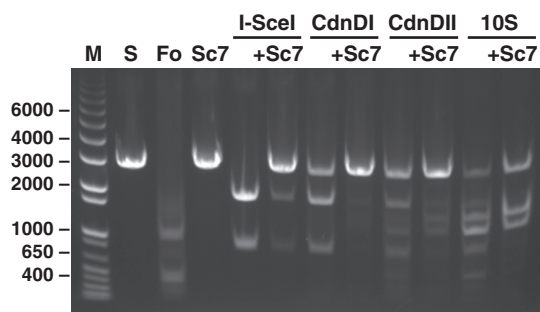


Figure 5. DNA cleavage by hybrid endonucleases and protection of substrate DNA from cleavage by excess Sce7. The addition of 110 nM Sce7 to any of the three hybrid endonucleases blocks cleavage at the I-SceI recognition site, but does not block non-specific cleavage by CdnDII or the hybrid endonuclease containing the 10S linker. M: DNA size standards (in base pairs); S: linearized plasmid containing a single I-SceI site and several FokI sites, without enzyme (11 nM in all assays); Fo: wild-type FokI (2 units); Sc7: Sce7 protein (110 nM); I-SceI: 185 nM; CdnDI: 50 nM, its optimal conditions; CdnDII: 100 nM, its optimal conditions.

copy of the I-SceI target sequence were affected by both reaction conditions and linker sequence (Figure 4). Under optimized reaction conditions (4h incubation with 50 nM CdnDI or 100 nM CdnDII in 20 mM Tris-HCl, pH 9.0, 25 mM NaCl, 10 mM MgCl₂, 1 mM dithiothreitol, 0.1 mg/ml bovine serum albumin and 2% glycerol; at 37°C for CdnDI or 42°C for CdnDII), the chimeric enzymes CdnDI and CdnDII each cut the linearized substrate at a single site to produce two DNA fragments of the expected length (Figure 5). In addition to this pair of major products, a minor trace of additional DNA fragments indicative of non-specific cleavage was observed for CdnDII. In contrast, the remaining chimeric enzymes produced additional DNA fragments under the conditions tested, indicative of cleavage at one or more alternative sites (Table 1). The addition of excess Sce7 protein blocked cleavage at the I-SceI target site, but did not block cleavage at other sites (Figure 5). Under their optimal cleavage conditions, both CdnDI and CdnDII cleaved supercoiled plasmid DNA that contained a single copy of the I-SceI target site with efficiency comparable to that on linearized DNA.

Cleavage sites of chimeric enzymes

For wild-type I-SceI and chimeric enzymes CdnDI and CdnDII, the two products of DNA cleavage were extracted, purified, and sequenced in the direction toward the I-SceI target site to identify the exact location of the cleavage site for each enzyme (Figure 6). For other chimeric enzymes, only the two products of DNA cleavage of length indistinguishable from those found for I-SceI were extracted, purified and sequenced.

Sequencing traces for each enzyme tested dropped off sharply at a specific site, which corresponded to the 5' end of that DNA cleavage product, indicating that DNA cleavage occurred predominantly at a single position (Figure 6). For wild-type I-SceI, the sequences of the DNA products were consistent with the published cleavage site within the recognition sequence (62). For all chimeric enzymes, the 5' end sequences of the two DNA

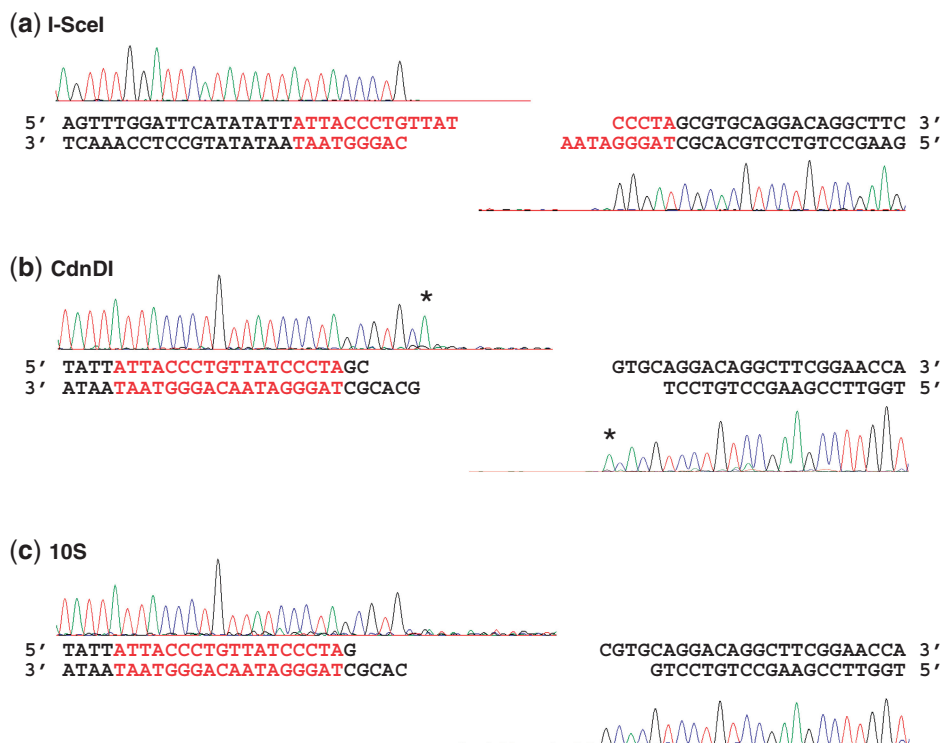


Figure 6. Run-off sequencing of purified cleavage products of I-SceI, CdnDI and the hybrid endonuclease with the 10S linker. Each cleavage product was sequenced toward the cleavage site. The sharp drop-off after a particular base indicates the end of the fragment, and thus the site of DNA cleavage. (a) I-SceI. The fragment ends generated by I-SceI are in agreement with its previously reported internal cleavage site (62). (b) CdnDI. The overlapping fragments generated by CdnDI indicate 5', four-base overhangs resulting from cuts at positions 2/6 relative to the I-SceI recognition site. The asterisks denote non-templated additions of 'A', as discussed in 'Results' section. CdnDII and the hybrid endonuclease with the 20S linker produced run-off sequencing results indistinguishable from those for CdnDI. (c) The hybrid enzyme with the 10S linker. The overlapping fragments generated by this enzyme indicate 5', four-base overhangs resulting from cuts at positions 1/5 relative to the I-SceI recognition site.

products overlapped, indicating that these enzymes produce 5' overhangs. Often, an additional 'A' was observed after the final base in a fragment (e.g. Figure 6b, both traces); we attribute this phenomenon to the previously described, non-templated addition of 'A' (dATP) by the Taq polymerase used in the sequencing reactions (74). Whereas the final 'A' signal in the right-hand trace of Figure 6b does not match the sequence of the DNA template, and is clearly a product of non-templated addition, the final 'A' signal in the left-hand trace of Figure 6b is less straightforward to interpret due to the 'A' present at that position in the template DNA. This final 'A' in this left-hand trace could be either the result of a non-templated polymerase addition, or else a part of the true sequencing trace. If it is the result of a non-templated polymerase addition, the trace indicates that the CdnDI enzyme produced a four-base, 2/6 overhang; alternatively, if it corresponds to an actual 'A' in the fragment sequence, the trace indicates that the enzyme produced a five-base, 2/7 overhang. We distinguished between these possibilities by testing the left-hand product of CdnDI for its ability to ligate to four-base overhangs generated by BsaI, which match the predicted CdnDI 2/6 overhang. The high-fidelity ligation of these test overhangs with matching four-base overhangs generated by BsaI, as described below in the section 'Assembly

of DNA fragments generated by chimeric enzymes', suggested that the 'A' signal was indeed due to a non-templated polymerase addition, and that CdnDI produced four-base, 2/6 overhangs.

The sequences of DNA fragments produced by CdnDI, CdnDII and the chimeric enzyme-containing linker 20S all drop off at the exact same sites, which correspond to the ends of four-base, 5' overhangs, two and six bases downstream of the target site (2/6). When tested on a substrate with a different DNA sequence downstream of the I-SceI target site, CdnDI still produced the same four-base, 2/6 overhang. The chimeric enzyme containing linker 10S generates a four-base, 5', 1/5 overhang (Figure 6c), but the cleavage pattern of the chimeric enzyme containing linker 10D was not homogenous and reproducible enough for the cleavage site to be determined (Table 1).

Affinity and specificity of CdnD enzymes for the I-SceI target site

Binding affinities of purified I-SceI, Sce7, CdnDI and CdnDII for the I-SceI target site were estimated using a gel-retardation assay under conditions that inhibit DNA cleavage by the active enzymes. We estimated the K_d for I-SceI to be $62 (\pm 16)$ nM, which is similar to its previously measured K_m value of 34 nM at pH 9.5 (63). The inactive homing endonuclease mutant, Sce7, and chimeric enzymes

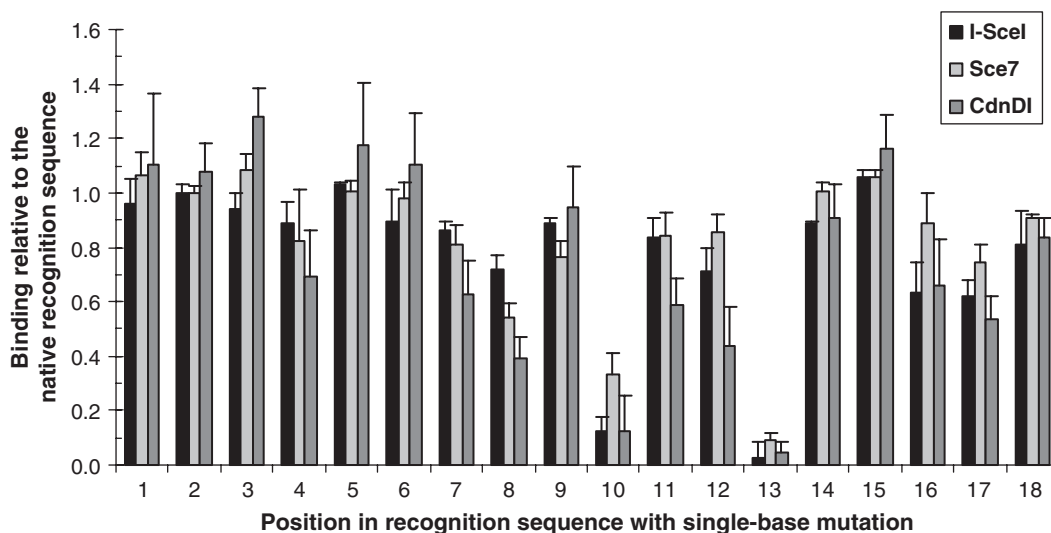


Figure 7. Binding specificity of I-SceI, Sce7 and CdnDI to near-native I-SceI target sequences. Each of the eighteen positions in the native target sequence were changed sequentially to a cytosine or thymine. Wild-type I-SceI, the inactive variant Sce7, and the hybrid enzyme CdnDI, each at the concentration three times of its apparent equilibrium dissociation constant to the native target sequence, were incubated with each variant target sequence. Protein-bound DNA was separated from free DNA in a gel-retardation assay and quantified by densitometry.

CdnDI and CdnDII bind substrate DNA more tightly than does wild-type I-SceI, with the estimated K_d of $12 (\pm 2)$, $19 (\pm 6)$ and $5.0 (\pm 1.2)$, respectively.

The binding specificities of I-SceI, Sce7 and CdnDI for the native 18-bp I-SceI target sequence were investigated by comparing their binding to a series of near-native target sequences (Figure 7) to their binding to the native target sequence. For the majority of single-base substitutions, all three proteins retain at least 50% of binding observed for the native binding sequence, as measured by the proportion of DNA bound to each protein. Mutations at positions 10 and 13, however, have significant, deleterious effects on binding, with a similar effect on all three proteins. Overall, the specificity profile across all 18 positions is essentially unchanged from the wild-type I-SceI to Sce7 and CdnDI.

Assembly of DNA fragments generated by chimeric enzymes

We tested the utility of the chimeric endonucleases CdnDI and CdnDII in ligation-based assembly of DNA fragments. To assemble a pair of inserts, each was excised from its donor vector by cutting with a standard type IIS endonuclease to generate an overhang compatible with the acceptor vector, and with either CdnDI, CdnDII or BsaI to generate an overhang for pairing with the second insert. As described in 'Materials and Methods' section, the donor vectors were designed so that, for each enzyme, both the fragment containing the recognition sequence and the fragment separated from the recognition sequence were tested. The pair of inserts with putative complementary overhangs were then ligated to each other to test the junction of interest, ligated into the acceptor vector and sequenced.

BsaI, a standard type IIS endonuclease that generates 5', four-base overhangs, was used to generate test

Table 2. Ligation fidelity for DNA fragments generated by CdnDI and CdnDII

Enzyme used to generate each side of junction		Fraction of colonies with correct sequence
Left side	Right side	
CdnDI	CdnDI	43/48 = 90%
CdnDI	BsaI	45/47 = 96%
BsaI	CdnDI	44/47 = 91%
CdnDII	CdnDII	39/48 = 81%
CdnDII	BsaI	48/50 = 96%
BsaI	CdnDII	42/48 = 87%

overhangs with perfect complementarity to the 5' overhang that would be generated by CdnDI or CdnDII if these enzymes cut in the predicted 2/6 pattern. As shown in Table 2, ligation of BsaI-generated overhangs to CdnDI-generated overhangs yielded more than 90% of clones with the correctly assembled insert. Similar high fidelity was achieved when CdnDI-generated overhangs were ligated to complementary CdnDI-generated overhangs. Together, these results add further evidence that CdnDI generates homogeneous, four-base, 5' overhangs, two and six bases downstream of the target site (2/6). Ligations that involved one or two CdnDII-generated fragments had only a slightly lower fidelity (81–96%) than the ligations of CdnDI-generated fragments, consistent with the same, 2/6 cleavage pattern for the two endonucleases.

DISCUSSION

This study adds to the large number of designed chimeric endonucleases engineered by linking the catalytic domain of FokI (25–37) or the catalytic domain of BmrI (75–77),

another type IIS endonuclease, to unrelated DNA-binding domains. What distinguishes our most successful engineered chimeric endonucleases, CdnDI and CdnDII, from their predecessors is that they each make a single, double-stranded break in DNA at a precisely defined site outside their recognition region, in a manner similar to that of type IIS restriction endonucleases. Previously described chimeric endonucleases either cleave DNA outside their recognition site, but generate a heterogeneous mixture of products (25–28,30–32,34,75–77); or else cleave at a precisely defined site, which must be located between two face-to-face recognition sites, as is the case of some highly efficient zinc-finger nucleases (35–37,48). As a consequence, the DNA products of CdnDI and CdnDII, but not the products of earlier engineered chimeric endonucleases, have no sequence restrictions and can be ligated to complementary sticky ends to generate defined-sequence products.

Biochemical characterization of five engineered chimeric endonucleases constructed using different linkers revealed that, whereas all five enzymes cleaved linearized plasmid containing a single I-SceI target site and produced the two fragments of the expected size, only CdnDI produced no trace of additional cleavage products. The other four chimeric endonucleases produced two or more additional DNA fragments, indicative of cleavage at a small number of additional, defined sites. The two highest-molecular-weight undesired cleavage products, which were produced at the highest level by the chimeric enzyme with the 10S linker (Figure 5), are consistent in size with cleavage of the linearized plasmid at a site that matches 12 of the 18 bases of the I-SceI target site. One possible explanation for cleavage at this partially matching site is that Sce7 binds to secondary, lower-affinity sites and still directs the catalytic domain to cleave DNA due to decreased catalytic-domain sequestration. In wild-type FokI, the DNA-recognition domain has been shown to interact with the catalytic domain, likely sequestering the catalytic domain from DNA cleavage until the specific, high-affinity target site on DNA is found (16). In the chimeric enzymes, however, the Sce7 DNA-binding module would not be expected to sequester the FokI catalytic domain. Nevertheless, it is possible that, in the case of the most specific chimeric enzyme, CdnDI, its relatively hydrophobic linker promotes nonspecific interactions between Sce7 and the catalytic domain, mimicking the sequestration mechanism of wild-type FokI and thus protecting this site from cleavage by CdnDI. The more hydrophilic linkers of the other four chimeric endonucleases tested are less likely to provide such sequestration-like interactions. An alternative possible explanation is that the linkers used in the chimeric enzymes other than CdnDI provide some additional contacts with DNA that lead to cleavage at the site with only 12 matching base pairs.

In addition to binding-site specificity, linkers of different length and composition also affect the exact length and position of overhangs on the cleavage products. Overall, the results agree with the prediction that a chimeric enzyme with a shorter, 10-residue linker would cleave closer to the target site than a chimeric enzyme with a

longer, 20-residue linker. We find that the chimeric enzyme with the native FokI, fifteen-residue linker cleaves DNA at the same sites (2/6) as the two chimeric enzymes with 20-residue linkers, producing 5', four-base overhangs, while the chimeric enzyme with the ten-residue linker, 10S, cleaves closer to the target site, at 1/5.

Enzyme characterization under different reaction conditions revealed that the parameter with the greatest effect on activity and specificity of the chimeric enzymes was buffer pH. For example, CdnDI was completely inactive at pH 10; active with a significant proportion of specific, single-site cleavage between pH 8–9.5; and active leading to complete, non-specific degradation of substrate DNA at pH 6–7 (Figure 4a). This effect can be only partially explained by what is known about the pH dependence of the two parent endonucleases, FokI and I-SceI. Our findings are consistent with the fact that FokI is active in the pH range of 5–10, with 80% maximum activity or higher at pH 7–9, and with a sharp drop in activity from pH 9 to 10 (78). It appears that activity of the FokI catalytic domain at pH 10 is too low for appreciable cleavage by CdnDI on the 2–6 h scale, even when the FokI catalytic domain is positioned on DNA substrate by the Sce7-binding domain. On the other hand, we cannot easily explain the loss of cleavage specificity by CdnDI at pH 6–7. It does not appear likely that the loss of specificity is due to Sce7-independent DNA binding and cleavage by the CdnDI FokI catalytic domain, since it has been shown that FokI with a binding-deficient recognition domain does not cleave DNA (18). Also, whereas activity of wild-type I-SceI activity is optimal at pH 9–10 and drops sharply as pH decreases to 6, this loss of activity has been shown to be due primarily to a decrease in V_{max} —not to an increase in K_m or a loss of enzyme stability (63). One remaining possible explanation for the loss of CdnD specificity at low pH is that, at low pH, Sce7 gains affinity for other DNA sequences without a change in affinity for the native target site. The resulting Sce7-mediated binding to DNA sequences other than the standard I-SceI target sequence may be the reason for the observed promiscuous DNA cleavage by the associated FokI catalytic domain.

We demonstrated that the homogeneous, 5', four-base overhangs generated by cleavage of our chimeric, type IIS endonucleases CdnDI and CdnDII can be ligated to complementary sticky ends with a high fidelity and efficiency. Since the long target sequence of the I-SceI-derived DNA-recognition domain imparts on our chimeric endonucleases the specificity for a much less common DNA sequence than that recognized by natural type IIS endonucleases, we expect that CdnDI and CdnDII will be particularly useful in ligation-based assembly of DNA fragments in the kilobase to megabase range.

Compared to most natural restriction enzymes, our engineered, chimeric endonucleases have relatively low activity. The enzymes must be present in excess of DNA and incubated for several hours to cleave most of the substrate. Despite the fact that one of the parent enzymes for this study, wild-type FokI, cleaves at a similarly slow rate when only one DNA recognition site is present on the template (23,24), we cannot rule out the possibility that

CdnDI and CdnDII may not turn over, i.e. may not be true enzymes. Nevertheless, the fact that an easily purified amount of CdnDI or CdnDII can be used to cleave microgram amounts of DNA and generate ligation-competent, homogeneous sticky ends demonstrates that these engineered endonucleases are sufficiently active for use in cloning, and will be particularly welcome in situations that require selectivity beyond that of standard type IIS enzymes.

Our most successful engineered endonucleases to date, CdnDI and CdnDII, both recognize the same DNA sequence and cut at the same distance from the target site. In the future, the repertoire of type IIS endonucleases with long target sequences could be expanded by attaching the FokI catalytic domain to other deactivated homing endonucleases with different target sites or to other DNA-binding proteins that recognize long DNA sequences. Alternatively, the *Sce7* DNA-recognition domain described in this study could be engineered to bind specifically to different sequences, as described previously for active homing endonucleases (64,79–94).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Kathy Galle, Serge Gottschalk, Brad Chapman, Sara Haserlat and Eric Devroe of the Codon Devices Production team for the synthesis of the DNA clones used in this project; and Anna Huang, Caitlin Vestal, Gina Prophete, Edward Kelliher, Nerline Grand-Pierre and Mei Lin for sequencing support.

FUNDING

Codon Devices, Inc., research budget. Funding for open access charge: Codon Devices, Inc., research budget.

Conflict of interest statement. None declared.

REFERENCES

- Raleigh, E.A. and Brooks, J.E. (1998) In de Bruijn, F.J., Lupski, J.R. and Weinstock, G.M. (eds), *Bacterial Genomes*, Chapman & Hall, New York, N.Y., pp. 78–92.
- Arber, W. (1979) Promotion and limitation of genetic exchange. *Science*, **205**, 361–365.
- Pingoud, A., Fuxreiter, M., Pingoud, V. and Wende, W. (2005) Type II restriction endonucleases: structure and mechanism. *Cell Mol. Life Sci.*, **62**, 685–707.
- Xu, L., Chen, H., Hu, X., Zhang, R., Zhang, Z. and Luo, Z.W. (2006) Average gene length is highly conserved in prokaryotes and eukaryotes and diverges only between the two kingdoms. *Mol. Biol. Evol.*, **23**, 1107–1108.
- Yount, B., Curtis, K.M. and Baric, R.S. (2000) Strategy for systematic assembly of large RNA and DNA genomes: transmissible gastroenteritis virus model. *J. Virol.*, **74**, 10600–10611.
- Cello, J., Paul, A.V. and Wimmer, E. (2002) Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science*, **297**, 1016–1018.
- Smith, H.O., Hutchison, C.A. 3rd, Pfannkoch, C. and Venter, J.C. (2003) Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proc. Natl Acad. Sci. USA*, **100**, 15440–15445.
- Shevchuk, N.A., Bryksin, A.V., Nusinovich, Y.A., Cabello, F.C., Sutherland, M. and Ladisch, S. (2004) Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously. *Nucleic Acids Res.*, **32**, e19.
- Kodumal, S.J., Patel, K.G., Reid, R., Menzella, H.G., Welch, M. and Santi, D.V. (2004) Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc. Natl Acad. Sci. USA*, **101**, 15573–15578.
- Forster, A.C. and Church, G.M. (2006) Towards synthesis of a minimal cell. *Mol. Syst. Biol.*, **2**, Article Number 45.
- Holt, R.A., Warren, R., Flibotte, S., Missirlis, P.I. and Smailus, D.E. (2007) Rebuilding microbial genomes. *Bioessays*, **29**, 580–590.
- Itaya, M., Fujita, K., Kuroki, A. and Tsuge, K. (2008) Bottom-up genome assembly using the *Bacillus subtilis* genome vector. *Nat. Methods*, **5**, 41–43.
- Gibson, D.G., Benders, G.A., Andrews-Pfannkoch, C., Denisova, E.A., Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., Algire, M.A. *et al.* (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science*, **319**, 1215–1220.
- Sugisaki, H. and Kanazawa, S. (1981) New restriction endonucleases from *Flavobacterium okeanokoites* (FokI) and *Micrococcus luteus* (MluI). *Gene*, **16**, 73–78.
- Wah, D.A., Hirsch, J.A., Dorner, L.F., Schildkraut, I. and Aggarwal, A.K. (1997) Structure of the multimodular endonuclease FokI bound to DNA. *Nature*, **388**, 97–100.
- Waugh, D.S. and Sauer, R.T. (1994) A novel class of FokI restriction endonuclease mutants that cleave hemi-methylated substrates. *J. Biol. Chem.*, **269**, 12298–12303.
- Wah, D.A., Bitinaite, J., Schildkraut, I. and Aggarwal, A.K. (1998) Structure of FokI has implications for DNA cleavage. *Proc. Natl Acad. Sci. USA*, **95**, 10564–10569.
- Bitinaite, J., Wah, D.A., Aggarwal, A.K. and Schildkraut, I. (1998) FokI dimerization is required for DNA cleavage. *Proc. Natl Acad. Sci. USA*, **95**, 10570–10575.
- Vanamee, E.S., Berriman, J. and Aggarwal, A.K. (2007) An EM view of the FokI synaptic complex by single particle analysis. *J. Mol. Biol.*, **370**, 207–212.
- Catto, L.E., Ganguly, S., Milsom, S.E., Welsh, A.J. and Halford, S.E. (2006) Protein assembly and DNA looping by the FokI restriction endonuclease. *Nucleic Acids Res.*, **34**, 1711–1720.
- Gemmen, G.J., Millin, R. and Smith, D.E. (2006) Tension-dependent DNA cleavage by restriction endonucleases: two-site enzymes are ‘switched off’ at low force. *Proc. Natl Acad. Sci. USA*, **103**, 11555–11560.
- Gemmen, G.J., Millin, R. and Smith, D.E. (2006) DNA looping by two-site restriction endonucleases: heterogeneous probability distributions for loop size and unbinding force. *Nucleic Acids Res.*, **34**, 2864–2877.
- Vanamee, E.S., Santagata, S. and Aggarwal, A.K. (2001) FokI requires two specific DNA sites for cleavage. *J. Mol. Biol.*, **309**, 69–78.
- Bath, A.J., Milsom, S.E., Gormley, N.A. and Halford, S.E. (2002) Many type IIs restriction endonucleases interact with two recognition sites before cleaving DNA. *J. Biol. Chem.*, **277**, 4024–4033.
- Kim, Y.G. and Chandrasegaran, S. (1994) Chimeric restriction endonuclease. *Proc. Natl Acad. Sci. USA*, **91**, 883–887.
- 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.
- 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.
- 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.
- 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.
- Kim, Y.G., Kim, P.S., Herbert, A. and Rich, A. (1997) Construction of a Z-DNA-specific restriction endonuclease. *Proc. Natl Acad. Sci. USA*, **94**, 12875–12879.

31. Kim, Y.G., Smith, J., Durgesh, M. and Chandrasegaran, S. (1998) Chimeric restriction enzyme: Gal4 fusion to FokI cleavage domain. *Biol. Chem.*, **379**, 489–495.
32. Ruminy, P., Derambure, C., Chandrasegaran, S. and Salier, J.P. (2001) Long-range identification of hepatocyte nuclear factor-3 (FoxA) high and low-affinity binding sites with a chimeric nuclease. *J. Mol. Biol.*, **310**, 523–535.
33. Lariguet, P., Dunand, C., Herzog, M. and Vachon, G. (1999) APETALA3-nuclease hybrid protein: a potential tool for APETALA3 target gene mutagenesis. *Plant Sci.*, **148**, 19–30.
34. Horner, S.M. and DiMaio, D. (2007) The DNA binding domain of a papillomavirus E2 protein programs a chimeric nuclease to cleave integrated human papillomavirus DNA in HeLa cervical carcinoma cells. *J. Virol.*, **81**, 6254–6264.
35. 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.
36. 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.
37. Mani, M., Smith, J., Kandavelou, K., Berg, J.M. and Chandrasegaran, S. (2005) Binding of two zinc finger nuclease monomers to two specific sites is required for effective double-strand DNA cleavage. *Biochem. Biophys. Res. Commun.*, **334**, 1191–1197.
38. 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.
39. Bibikova, M., Beumer, K., Trautman, J.K. and Carroll, D. (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science*, **300**, 764.
40. Porteus, M.H. and Baltimore, D. (2003) Chimeric nucleases stimulate gene targeting in human cells. *Science*, **300**, 763.
41. 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.
42. Wright, D.A., Townsend, J.A., Winfrey, R.J., 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.
43. 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.
44. Mandell, J.G., Barbas, C.F. and 3rd., (2006) Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res.*, **34**, W516–523.
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. USA*, **103**, 16370–16375.
46. Carroll, D., Morton, J.J., Beumer, K.J. and Segal, D.J. (2006) Design, construction and in vitro testing of zinc finger nucleases. *Nat. Protocols*, **1**, 1329–1341.
47. Moehle, E.A., Rock, J.M., Lee, Y.L., Jouvenot, Y., DeKaveler, 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.
48. 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.
49. 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.
50. Lombardo, A., Genovese, P., Beausejour, C.M., Colleon, S., Lee, Y.L., Kim, K.A., Ando, D., Urnov, F.D., Galli, C., Gregory, P.D. et al. (2007) Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat. Biotechnol.*, **25**, 1298–1306.
51. Cornu, T.I., Thibodeau-Beganny, S., Guhl, E., Alwin, S., Eichinger, M., Joung, J. and Cathomen, T. (2008) DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger Nucleases. *Mol. Ther.*, **16**, 352–358.
52. Maeder, M.L., Thibodeau-Beganny, S., Osia, A., Wright, D.A., Anthony, R.M., Eichinger, M., Jiang, T., Foley, J.E., Winfrey, R.J., Townsend, J.A. et al. (2008) Rapid “Open-Source” engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell*, **31**, 294–301.
53. Porteus, M.H. and Carroll, D. (2005) Gene targeting using zinc finger nucleases. *Nat. Biotechnol.*, **23**, 967–973.
54. Porteus, M.H. (2006) Mammalian gene targeting with designed zinc finger nucleases. *Mol. Ther.*, **13**, 438–446.
55. 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.
56. Wu, J., Kandavelou, K. and Chandrasegaran, S. (2007) Custom-designed zinc finger nucleases: What is next? *Cell Mol. Life Sci.*, **64**, 2933–2944.
57. Thierry, A., Perrin, A., Boyer, J., Fairhead, C., Dujon, B., Frey, B. and Schmitz, G. (1991) Cleavage of yeast and bacteriophage T7 genomes at a single site using the rare cutter endonuclease I-Sce I. *Nucleic Acids Res.*, **19**, 189–190.
58. Dujon, B. (1980) Sequence of the intron and flanking exons of the mitochondrial 21S rRNA gene of yeast strains having different alleles at the omega and rib-1 loci. *Cell*, **20**, 185–197.
59. Dujon, B. (1989) Group I introns as mobile genetic elements: facts and mechanistic speculations—a review. *Gene*, **82**, 91–114.
60. Colleaux, L., Dauriol, 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.
61. Dujon, B., Colleaux, L., Jacquier, A., Michel, F. and Monteilhet, C. (1986) In Wickner, R.B., Hinnebusch, A., Lambowitz, A.M., Gunsalus, I.C. and Hollaender, A. (eds), *Extrachromosomal Elements in Lower Eucaryotes*. Plenum, New York, pp. 5–27.
62. 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. USA*, **85**, 6022–6026.
63. Monteilhet, C., Perrin, A., Thierry, A., Colleaux, L. and Dujon, B. (1990) Purification and characterization of the in vitro activity of I-Sce I, a novel and highly specific endonuclease encoded by a group I intron. *Nucleic Acids Res.*, **18**, 1407–1413.
64. 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.
65. Scalley-Kim, M., McConnell-Smith, A. and Stoddard, B.L. (2007) Coevolution of a homing endonuclease and its host target sequence. *J. Mol. Biol.*, **372**, 1305–1319.
66. 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.
67. Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I. and Aggarwal, A.K. (1995) Structure of Bam HI endonuclease bound to DNA: partial folding and unfolding on DNA binding. *Science*, **269**, 656–663.
68. Carr, P.A., Park, J.S., Lee, Y.J., Yu, T., Zhang, S. and Jacobson, J.M. (2004) Protein-mediated error correction for de novo DNA synthesis. *Nucleic Acids Res.*, **32**, e162.
69. Li, L. and Chandrasegaran, S. (1993) Alteration of the cleavage distance of Fok I restriction endonuclease by insertion mutagenesis. *Proc. Natl Acad. Sci. USA*, **90**, 2764–2768.
70. Zhou, H.X. (2003) Quantitative account of the enhanced affinity of two linked scFvs specific for different epitopes on the same antigen. *J. Mol. Biol.*, **329**, 1–8.
71. Zhou, H.X. (2004) Polymer models of protein stability, folding, and interactions. *Biochem.*, **43**, 2141–2154.
72. Williams, R.W., Chang, A., Juretic, D. and Loughran, S. (1987) Secondary structure predictions and medium range interactions. *Biochim. Biophys. Acta*, **916**, 200–204.

73. Wilmot, C.M. and Thornton, J.M. (1988) Analysis and prediction of the different types of beta-turn in proteins. *J. Mol. Biol.*, **203**, 221–232.
74. Clark, J.M. (1988) Novel non-templated nucleotide addition reactions catalyzed by prokaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res.*, **16**, 9677–9686.
75. 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.
76. Zhang, P., Bao, Y., Higgins, L. and Xu, S.Y. (2007) Rational design of a chimeric endonuclease targeted to NotI recognition site. *Protein Eng. Des. Sel.*, **20**, 497–504.
77. Fomenkov, A., Too, P.H.-M., Chan, S.-H., Vaisvila, R., Cantin, B.A., Mazzola, L., Tam, V. and Xu, S.-y. (2008) Targeting DNA 5mCpG sites with chimeric endonucleases. *Analytical Biochemistry*, **381**, 135–141.
78. Kaczorowski, T., Skowron, P. and Podhajska, A.J. (1989) Purification and characterization of the FokI restriction endonuclease. *Gene*, **80**, 209–216.
79. Seligman, L.M., Chisholm, K.M., Chevalier, B.S., Chadsey, M.S., Edwards, S.T., Savage, J.H. and Veillet, A.L. (2002) Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic Acids Res.*, **30**, 3870–3879.
80. Chevalier, B.S., Kortemme, T., Chadsey, M.S., Baker, D., Monnat, R.J. and Stoddard, B.L. (2002) Design, activity, and structure of a highly specific artificial endonuclease. *Mol. Cell*, **10**, 895–905.
81. Steuer, S., Pingoud, V., Pingoud, A. and Wende, W. (2004) Chimeras of the homing endonuclease PI-SceI and the homologous *Candida tropicalis* intein: a study to explore the possibility of exchanging DNA-binding modules to obtain highly specific endonucleases with altered specificity. *ChemBioChem*, **5**, 206–213.
82. Silva, G.H., Belfort, M., Wende, W. and Pingoud, A. (2006) From monomeric to homodimeric endonucleases and back: engineering novel specificity of LAGLIDADG enzymes. *J. Mol. Biol.*, **361**, 744–754.
83. Gimble, F.S., Moure, C.M. and Posey, K.L. (2003) Assessing the plasticity of DNA target site recognition of the PI-SceI homing endonuclease using a bacterial two-hybrid selection system. *J. Mol. Biol.*, **334**, 993–1008.
84. Chen, Z.L. and Zhao, H.M. (2005) A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic Acids Res.*, **33**, Article Number E154.
85. Volna, P., Jarjour, J., Baxter, S., Roffler, S.R., Monnat, R.J., Stoddard, B.L. and Scharenberg, A.M. (2007) Flow cytometric analysis of DNA binding and cleavage by cell surface-displayed homing endonucleases. *Nucleic Acids Res.*, **35**, 2748–2758.
86. Sussman, D., Chadsey, M., Fauce, S., Engel, A., Bruett, A., Monnat, R. Jr, Stoddard, B.L. and Seligman, L.M. (2004) Isolation and characterization of new homing endonuclease specificities at individual target site positions. *J. Mol. Biol.*, **342**, 31–41.
87. Rosen, L.E., Morrison, H.A., Masri, S., Brown, M.J., Springstubb, B., Sussman, D., Stoddard, B.L. and Seligman, L.M. (2006) Homing endonuclease I-CreI derivatives with novel DNA target specificities. *Nucleic Acids Res.*, **34**, 4791–4800.
88. Ashworth, J., Havranek, J.J., Duarte, C.M., Sussman, D., Monnat, R.J. Jr, Stoddard, B.L. and Baker, D. (2006) Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature*, **441**, 656–659.
89. Epinat, J.C., Arnould, S., Chames, P., Rochaix, P., Desfontaines, D., Puzin, C., Patin, A., Zanghellini, A., Paques, F. and Lacroix, E. (2003) A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res.*, **31**, 2952–2962.
90. Arnould, S., Chames, P., Perez, C., Lacroix, E., Duclert, A., Epinat, J.C., Stricher, F., Petit, A.S., Patin, A., Guillier, S. *et al.* (2006) Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J. Mol. Biol.*, **355**, 443–458.
91. Smith, J., Grizot, S., Arnould, S., Duclert, A., Epinat, J.C., Chames, P., Prieto, J., Redondo, P., Blanco, F.J., Bravo, J. *et al.* (2006) A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res.*, **34**, e149.
92. Arnould, S., Perez, C., Cabaniols, J.P., Smith, J., Gouble, A., Grizot, S., Epinat, J.C., Duclert, A., Duchateau, P. and Paques, F. (2007) Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *J. Mol. Biol.*, **371**, 49–65.
93. Jurenaite-Urbanaviciene, S., Serksnaite, J., Kriukiene, E., Giedriene, J., Venclovas, C. and Lubys, A. (2007) Generation of DNA cleavage specificities of type II restriction endonucleases by reassortment of target recognition domains. *Proc. Natl Acad. Sci. USA*, **104**, 10358–10363.
94. Chen, Z., Wen, F., Sun, N. and Zhao, H. (2009) Directed evolution of homing endonuclease I-SceI with altered sequence specificity. *Protein Eng. Des. Select.* doi:10.1093/protein/gzp001.