

Site- and strand-specific nicking of DNA by fusion proteins derived from Muth and I-SceI or TALE repeats

Lilia Gabsalilow, Benno Schierling, Peter Friedhoff, Alfred Pingoud and Wolfgang Wende*

Institute for Biochemistry, Justus-Liebig-University Giessen, Heinrich-Buff-Ring 58, D-35392 Giessen, Germany

Received October 10, 2012; Revised January 14, 2013; Accepted January 22, 2013

ABSTRACT

Targeted genome engineering requires nucleases that introduce a highly specific double-strand break in the genome that is either processed by homology-directed repair in the presence of a homologous repair template or by non-homologous end-joining (NHEJ) that usually results in insertions or deletions. The error-prone NHEJ can be efficiently suppressed by 'nickases' that produce a single-strand break rather than a double-strand break. Highly specific nickases have been produced by engineering of homing endonucleases and more recently by modifying zinc finger nucleases (ZFNs) composed of a zinc finger array and the catalytic domain of the restriction endonuclease FokI. These ZF-nickases work as heterodimers in which one subunit has a catalytically inactive FokI domain. We present two different approaches to engineer highly specific nickases; both rely on the sequence-specific nicking activity of the DNA mismatch repair endonuclease Muth which we fused to a DNA-binding module, either a catalytically inactive variant of the homing endonuclease I-SceI or the DNA-binding domain of the TALE protein AvrBs4. The fusion proteins nick strand specifically a bipartite recognition sequence consisting of the Muth and the I-SceI or TALE recognition sequences, respectively, with a more than 1000-fold preference over a stand-alone Muth site. TALE-Muth is a programmable nickase.

INTRODUCTION

Precise gene targeting requires custom-designed highly specific nucleases. Two basically different approaches are being pursued for this purpose, (i) protein engineering of

homing endonucleases, which results in 'meganucleases' of predefined specificity (1,2) and (ii) the fusion of a programmable DNA-binding module and a DNA-cleavage module, as exemplified by the zinc finger nucleases [ZFNs, (3–6)] and the TALE nucleases [TALENs, (7–10)]. ZFNs and TALENs use the non-specific cleavage domain of FokI as cleavage module and an array of zinc fingers or the DNA-binding domain of TAL effector proteins as DNA-binding module, respectively. Recently, also fusion proteins with a catalytically inactive I-SceI or a ZF array as DNA-binding modules and the restriction endonuclease PvuII as sequence-specific DNA-cleavage module were described (11,12), as well as fusion proteins with a catalytically inactive I-OnuI or a ZF array as DNA-binding modules and the I-TevI nuclease domain as DNA-cleavage module (13).

All nucleases mentioned above introduce DNA double-strand breaks at a specific target site that can be repaired by two different pathways: homology-directed repair (HDR) which requires a donor DNA template, or error-prone non-homologous end-joining (NHEJ), resulting in insertions, deletions or translocations. For the purpose of genome engineering by introducing sequence alterations or insertions at or near the site of the double-strand break, NHEJ is an unwanted and possibly genotoxic side reaction. It can be largely circumvented by using a DNA-nicking endonuclease [*vulgo* 'nickase', (14,15)] which had been shown to stimulate HDR (16).

Nickases occur naturally or can be obtained by engineering restriction or homing endonucleases [reviewed in (17,18)]. Naturally occurring nickases are for example the restriction endonuclease Nt.CviPII (19,20) or subunits of heterodimeric restriction endonucleases, e.g. Nt.BstD6I (21,22), Nb.BsrDI and Nb.BtsI (23). I-HmuI and I-BasI represent naturally occurring nickases in the homing endonuclease family (24–26). Restriction and homing endonucleases that usually recognize palindromic or quasi-palindromic DNA sequences and catalyze double-strand cuts making use of two catalytic centres

*To whom correspondence should be addressed. Tel: +49 641 35407; Fax: +49 641 35409; Email: wolfgang.wende@chemie.bio.uni-giessen.de

can be engineered to become nickases by inactivating one catalytic centre, as had been shown for EcoRV (27,28), FokI (29), I-SceI (30) and I-AniI (31).

ZFNs that use the non-specific DNA-cleavage domain of FokI to introduce a programmable DNA double-strand cut can be converted to a ZF-nickase by inactivating the catalytic centre of one FokI monomer in a ZFN heterodimer. This has been shown in a few studies published in 2012 (32–34), which demonstrated that the ZF-nickases allow site-specific genome modifications at the predetermined target site, while reducing unwanted mutagenesis caused by error-prone NHEJ.

It had been argued by Halford (35) that ‘the reaction mechanism of FokI excludes the possibility of targeting ZFNs to unique DNA sites’. Although this risk has been minimized by using obligate heterodimers (36,37), it was shown in two recent studies that ZFNs that form obligate heterodimers caused residual off-target cleavage (38,39). What is a problem with ZFNs might be also one with ZF-nickases, when the same FokI cleavage module is used. The question arises, whether one can substitute the catalytic domain of FokI in ZF-nickases with a different DNA-cleavage module, as we have done in three different fusion proteins, PvuII–I-SceI (12), ZF-PvuII (11) and TALE–PvuII (M. Yanik, unpublished data) that introduce double-strand cuts into DNA.

In this study, we engineered and tested two highly specific nickases, MutH–I-SceI and TALE–MutH. The DNA mismatch repair protein MutH is a naturally occurring monomeric site-specific nickase that can nick the DNA at GATC sites in un- or hemimethylated DNA (40,41). It is part of the MutHLS system from *Escherichia coli* and a few other bacteria; it serves to repair DNA mismatches, e.g. caused by rare errors of the replicative polymerases. The mismatch is recognized by MutS which recruits MutL to the site of error to form a ternary complex (42). This complex activates MutH to nick the erroneous daughter strand at the 5'-side of a hemimethylated GATC site which can be >1000 bp upstream or downstream from the mismatch (43). The nick in the unmethylated strand is the entry point for UvrD helicase to generate single-stranded DNA that is digested by exonucleases past the original mismatch. The resulting gap is filled by DNA polymerase and the DNA is ligated (44). MutH by itself hardly shows any cleavage activity on unmethylated DNA under physiological ionic strength (41,45). Only when guided by the fused DNA-binding module does MutH exhibit nicking activity at the targeted GATC site. We show here that MutH–I-SceI and TALE–MutH can be considered as site- and strand-specific nickases, and TALE–MutH, in addition, as a programmable nickase.

MATERIALS AND METHODS

Design and construction of the MutH–I-SceI and TALE–MutH fusion proteins

For the construction of the MutH–I-SceI fusion protein, a catalytically inactive variant of I-SceI (46), which had been truncated at the C-terminus (Δ C9), was fused to the C-terminal end of a cysteine-free variant of MutH

via a 10-amino-acid linker (ASENLYFQGG) harbouring a TEV protease recognition site (underlined), or for control a linker without the TEV site (TKQLVKSE). The gene for MutH (C96S), which contains the coding sequence for an N-terminal His₆-tag, and the gene for I-SceI (Δ C9 D44S D145A) were connected by the coding sequence of the linker and inserted into the expression vector pASK-IBA63b-plus (IBA, Göttingen, Germany) coding for a C-terminal Strep-tag. Thus, the expected fusion protein would be: His₆-MutH–I-SceI-Strep.

For the TALE–MutH fusion protein, a truncated variant of the TALE protein AvrBs4, corresponding to the previously described AvrBs3 DNA-binding module (9), was fused directly to the N-terminal end of MutH. The gene for the TALE–MutH fusion contains two parts: (i) the gene of the truncated TALE variant, missing the coding sequence of the first 152 amino acids at the N-terminal end and the last 250 amino acids at the C-terminal end [28 amino acids remain after the last half-repeat (M. Yanik, unpublished data)]. (ii) The gene of MutH, which contains the coding sequence for a C-terminal His₆-tag. The two parts were connected and introduced into the expression vector pQE30 (Qiagen), coding for an N-terminal Strep-tag. Thus, the expected fusion protein would be: Strep-TALE–MutH-His₆. The sequence of both fusion constructs was confirmed by DNA sequencing of the entire coding region. We had varied the linker length between AvrBs3 and PvuII which is structurally very similar to MutH (47) and found the 28-amino-acid linker superior to a 63- and 16-amino-acid linker (M. Yanik, unpublished data). Therefore, we have chosen the 28-amino-acid linker for the TALE–MutH fusion protein.

Protein expression and purification

The expression vectors for the recombinant fusion proteins were introduced into the *E. coli* strain XL1-Blue (Stratagene). The cells were grown at 37°C to OD_{600 nm} ca. 0.7 in LB-medium containing 75 µg/ml ampicillin. Protein expression was induced by adding 200 mg/l anhydrotetracycline or 1 mM isopropyl- β -D-thiogalactopyranoside for MutH–I-SceI or TALE–MutH, respectively, followed by further growth at 20°C overnight. The cells were harvested by centrifugation and resuspended in 20 mM Tris–HCl, pH 7.9, 1 M NaCl, 20 mM imidazole, 1 mM phenylmethanesulfonyl fluoride (PMSF), lysed by sonification and centrifuged for 30 min (>17 000g) at 4°C to remove cell debris. The His₆- and Strep-tagged proteins were first purified via Ni²⁺-NTA agarose (Qiagen) by a 1 h incubation of Ni²⁺-NTA agarose with the supernatant at 4°C and washing with resuspension buffer; the protein was eluted with 20 mM Tris–HCl, pH 7.9, 1 M NaCl and 200 mM imidazole. The eluted fractions were then transferred to a column with Strep-Tactin Sepharose (IBA) for further purification. After washing with 100 mM Tris–HCl, pH 7.9, 1 M NaCl, the protein was eluted with 100 mM Tris–HCl, pH 7.9, 500 mM NaCl, 2.5 mM desthiobiotin, dialyzed overnight at 4°C against 10 mM HEPES-KOH, pH 7.9, 500 mM KCl, 1 mM

ethylenediaminetetraacetic acid (EDTA), 1 mM dithiothreitol (DTT), 50% (v/v) glycerol and stored at -20°C . The protein concentration was determined by measuring the absorbance at 280 nm using the molar extinction coefficient as determined according to Pace *et al.* (48). Protein purification was monitored by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (PAGE) analysis.

Plasmid cleavage assay

To test the activity and specificity of the fusion proteins, the substrate plasmid pAT153 was used, which was modified by introducing a DNA cassette with the appropriate target site (Table 1). Three kinds of substrates were generated: one addressed substrate and two unaddressed substrates. The addressed substrate contains the recognition site for the binding module [I-SceI (S) or AvrBs4 (T)] and a MutH (H) recognition site separated by spacers of different lengths. For MutH–I-SceI, only an addressed substrate with a 3-bp spacer was used (S-3-H) and for TALE–MutH, several spacer lengths were tested (T-x-H; $x = 1, 2, 3, 4, 5, 6, 7, 8$ and 9 bp). The two unaddressed substrates contain either the recognition site for the binding module (S or T) without a nearby MutH recognition site or with a MutH recognition site (H) and without a recognition site for a binding module. For the analysis of the cleavage reactions, 4 nM substrate plasmid and 16 nM enzyme (i.e. a 4-fold excess of enzyme over substrate) were incubated in 20 mM Tris-acetate, 120 mM K-acetate, 1 mM MgCl_2 , 50 $\mu\text{g/ml}$ bovine serum albumin (BSA), pH 7.5 at 37°C . Taking the enzyme dilution into account, the ionic strength of the reaction mixture was ca. 160 mM (which corresponds to physiological conditions). The reaction progress was measured after defined time intervals (1, 3, 10, 30, 60 and 120 min) by agarose gel electrophoresis; gels were stained with ethidium bromide, and the fluorescence of DNA bands was visualized with a BioDocAnalyze System (Biometra). The time course of the reaction (product versus time) was fitted to a monoexponential function to give the rate constants for nicking. For the determination of the strand-specificity of

MutH–I-SceI and TALE–MutH, respectively, a pAT153-derived plasmid harbouring the specific substrate cassettes S-3-H and T-3-H/T-6-H, respectively (Table 1), was used. The nicked addressed plasmid was gel-purified and sent for sequencing with primers 5'-AATAGGCGTATCACGAGGCCCTTTC-3' (binding to the bottom strand) and 5'-ACCCAGAGCGCTGCCGGCAC-3' (binding to the top strand).

TEV cleavage of the linker of MutH–I-SceI

The TEV protease cleavage of MutH–I-SceI to separate the DNA-binding and DNA-cleavage modules was performed in 50 mM Tris–HCl, pH 8.0, 0.5 mM EDTA, 1 mM DTT, 10 units/150 μl of TEV protease (Invitrogen) and 0.44 μM MutH–I-SceI in the presence of 4.4 μM of a 53-bp oligonucleotide harbouring a GATC recognition site. The reaction mixture was incubated for 90 min at 30°C . Afterwards, an aliquot was taken to test the activity of MutH–I-SceI by a plasmid cleavage assay as mentioned above, using the addressed substrate. To exclude that the loss of the catalytic activity of MutH–I-SceI after TEV proteolysis is due to the procedure, the whole experiment was performed in parallel using water instead of the TEV protease.

Cleavage assay with fluorescently labelled PCR products for the determination of the strand-specificity of TALE–MutH

To investigate whether TALE–MutH nicks a DNA substrate in the top or the bottom strand (defined by the asymmetric recognition sequence of the DNA-binding module), PCR products were generated using a forward and a reverse primer which were 5'-labelled with the fluorophores Atto 488 and Atto 647N, respectively. These experiments were carried out for the T-x-H substrates ($x = 1, 2, 3, 4, 5, 6, 7, 8$ and 9 bp). The assay was performed with 20 nM of a 211-bp polynucleotide substrate and 120 nM of TALE–MutH (i.e. a 6-fold excess of enzyme over substrate) in 20 mM Tris-acetate, 120 mM K-acetate, 1 mM MgCl_2 , 50 $\mu\text{g/ml}$ BSA, pH 7.5 at 37°C . The ionic strength of the reaction mixture was ca.

Table 1. DNA cassettes harbouring appropriate target sites that were introduced into the plasmid substrates

Cassettes for addressed substrates	
TATGCGTATAGGGATAACAGGGTAATCCGGATCCTCTGAGTGTGCTGGTAATC	S-3-H
TGCAGTATAATTAATAATCCACTTGGATCCTCGCTCCGAGACATG	T-1-H
TGCAGTATAATTAATAATCCACTTCCGGATCCTCGCTCGAGACATG	T-2-H
TGCAGTATAATTAATAATCCACTTCCGGATCCTCGCTGAGACATG	T-3-H
TGCAGTATAATTAATAATCCACTTACCGGATCCTCGCGAGACATG	T-4-H
TGCAGTATAATTAATAATCCACTTGACCGGATCCTCGCTGACATG	T-5-H
TGCAGTATAATTAATAATCCACTTGAGCCGGATCCTCGCTACATG	T-6-H
TGCAGTATAATTAATAATCCACTTGAGTCCGGATCCTCGCACATG	T-7-H
TGCAGTATAATTAATAATCCACTTGAGTACCGGATCCTCGACATG	T-8-H
TGCAGTATAATTAATAATCCACTTGAGTATCCGGATCCTCACATG	T-9-H
Cassettes for unaddressed control substrates	
TATGCGTATAGGGATAACAGGGTAATCTGTGGCTACTGAGTGTGCTGGTAATC	S
TGCAGTATAATTAATAATCCACTTCCGTGGCCTCGCTGAGACATG	T
TATGCGTAGTCAAGCCTAAGTGTAACCCGGATCCTCTGAGTGTGCTGGTAATC	H

Notes: Target sites are underlined and abbreviated with S (I-SceI), T (AvrBs4) and H (MutH). Numbers indicate spacer lengths in base pairs. Sequences are shown in the 5'–3' direction for the sense strand.

160 mM (which corresponds to physiological conditions). The reaction progress was analysed after defined time intervals (3, 10, 30, 60, 120 and 180 min) by denaturing PAGE; the fluorescence of DNA bands was visualized with the VersaDoc Imaging System (Bio Rad).

RESULTS

Design of the fusion constructs

Two principally different types of approaches had been recently used by us to produce highly specific nucleases for the purpose of genome engineering: (i) fusion proteins with a specificity defined by a catalytically inactive homing endonuclease [I-SceI, (12)] and (ii) fusion proteins with a programmable specificity defined by a ZF array (11) or a TALE protein (M. Yanik, unpublished data). In both cases, the Type II restriction endonuclease PvuII was the DNA-cleavage module. We have now used these two different approaches in generating highly specific nickases, MutH–I-SceI and TALE–MutH. In analogy to our PvuII–I-SceI construct (12), we have produced a MutH–I-SceI fusion protein with an N-terminal His₆-tag and a C-terminal Strep-tag. Likewise, in analogy to our ZF- (11) and TALE-PvuII constructs (M. Yanik, unpublished data), a TALE–MutH fusion protein with an N-terminal Strep-tag and a C-terminal His₆-tag was produced (Figure 1).

Determination of the cleavage preference of MutH–I-SceI for the addressed substrate

The rates of DNA nicking were determined with unmethylated plasmid substrates with three different recognition sites: (i) the addressed bipartite recognition site composed of an I-SceI recognition site next to a MutH recognition site (GATC), (ii) a stand-alone I-SceI

recognition site and (iii) a stand-alone MutH recognition site. All three plasmid substrates have 19 additional GATC sites. As shown in Figure 2, only the plasmid substrate with the addressed bipartite site is nicked. Even at a 4-fold excess of enzyme over unaddressed substrate and at prolonged incubation time (2 h), no non-specific nicking or cleavage is observed. Given the sensitivity of the assay, these results show that the plasmid substrate with the addressed site is nicked by a factor of 1000 faster than the other plasmid substrates. Actually, the preference for the addressed site over an unaddressed site might exceed the factor of 1000, because (i) the plasmids used in the assay contain 19 unaddressed sites, i.e. stand-alone GATC sites and only one addressed site and (ii) in the determination of the rate of nicking of the undressed substrate, the lower limit of detection had been reached.

Inactivation of the specific nicking activity of MutH–I-SceI by proteolytic separation of the binding and the cleavage module

The activity of the MutH–I-SceI fusion construct can be suppressed by separating the DNA-binding and DNA-cleavage modules by pre-incubating MutH–I-SceI with the TEV protease which cleaves the linker peptide that connects the two modules. Figure 3 shows that the covalent linkage of I-SceI and MutH is necessary for nicking of the addressed site. There is no nicking or cleavage observed when I-SceI and MutH are present in the reaction mixture but not covalently linked to each other. If the MutH–I-SceI fusion construct, in which the linker does not contain a TEV protease cleavage site, is pre-incubated with TEV protease, the addressed substrate is nicked, demonstrating that TEV protease does not cleave MutH or I-SceI at a cryptic TEV protease recognition site.

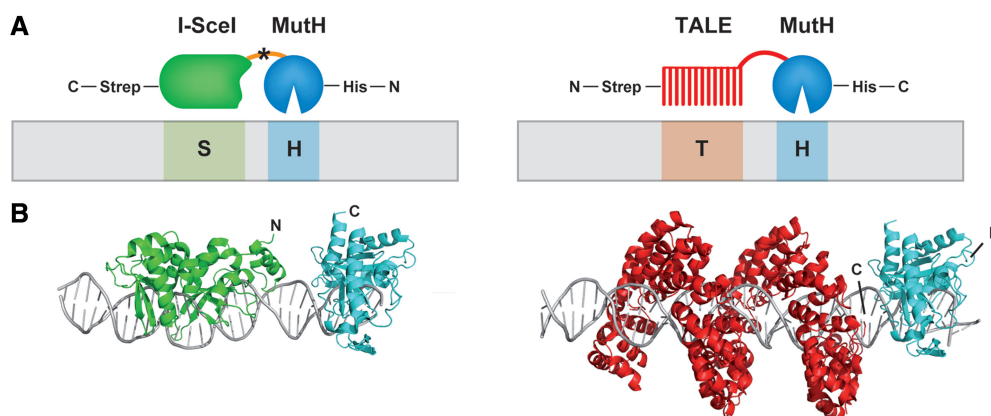


Figure 1. (A) Scheme of the fusion constructs MutH–I-SceI and TALE–MutH binding to their respective recognition sites (5′→3′). The highly specific interaction is mediated mainly by the binding module, I-SceI or the DNA-binding domain of the TALE protein AvrBs4, respectively. The monomeric cleavage module MutH is recruited by the binding module to nick the GATC site in close proximity to the I-SceI or TALE site. MutH is fused to I-SceI via a 10-amino-acid linker (ASENLYFQGG, shown as an orange line; the star indicates a TEV protease recognition site). In the case of the TALE–MutH construct, no additional linker was introduced; the fusion was done 28 amino acids (shown as red line) after the last half-repeat. (B) Model of the designed fusion constructs bound to specific DNA, but without the oligopeptide linking the DNA-binding and DNA-cleavage module. The colour code can be deduced from the scheme, I-SceI in green, the TALE protein in red (only the 23 repeats of the 3UGM structure are shown) and MutH in blue. Modelling was done using the crystal structures 1R7M (I-SceI), 3UGM (TALE) and 2A0Q (MutH). The model was generated with PyMOL and 3D-DART (56).

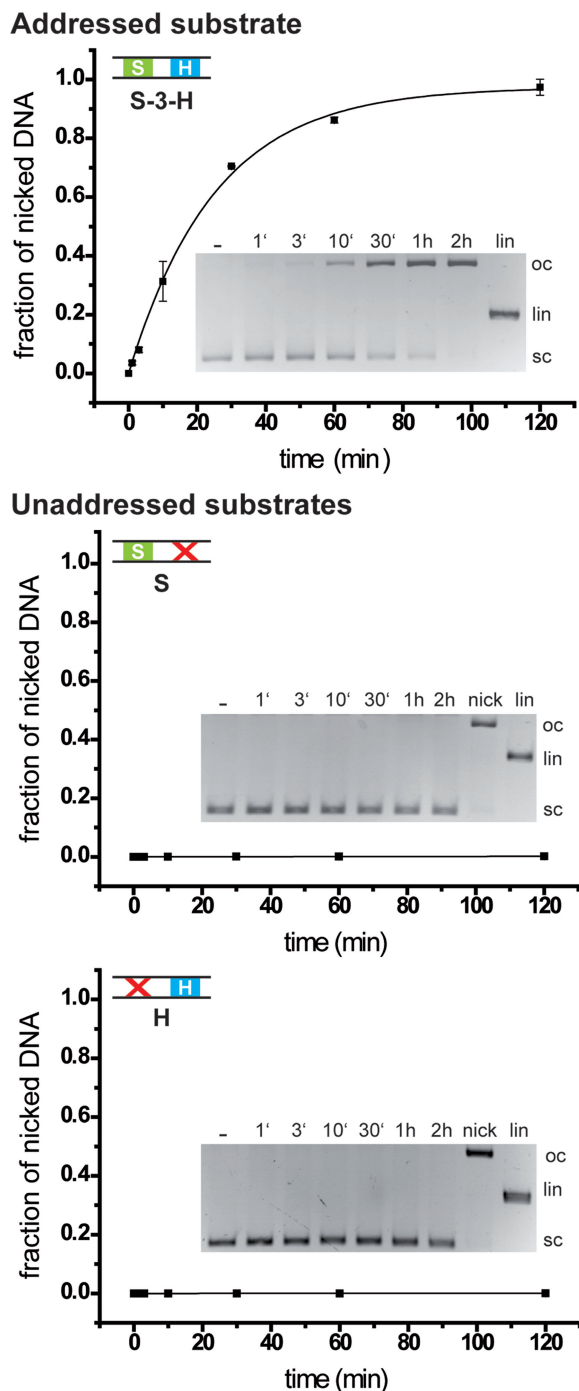


Figure 2. Kinetics of DNA nicking by the MutH-I-SceI fusion construct of addressed and unaddressed substrates analysed with plasmid cleavage assays using a 4-fold excess of enzyme over substrate (16 nM enzyme, 4 nM substrate). The addressed substrate plasmid contains an I-SceI and a MutH recognition site separated by 3 bp (S-3-H). The unaddressed substrate plasmids (controls) contain either an I-SceI recognition site without a nearby MutH recognition site (S) or no I-SceI recognition site (H). Each of the substrates has 19 additional GATC sites distributed over the plasmid. The position of the bands representing supercoiled (sc), open circular (oc) and linear (lin) forms of the plasmid is indicated. The first lane shows the reaction without $MgCl_2$ (-). 'Nick' and 'lin' represent the controls for nicking and cleavage activity, respectively, and correspond to the open circular and linear forms of the plasmid, respectively. The electrophoretic analysis shows that the addressed substrate plasmid is nicked by the fusion construct MutH-I-SceI, whereas the unaddressed substrate

Determination of the strand specificity of MutH-I-SceI

Since in the fusion protein MutH-I-SceI, the DNA-binding and the cleavage module were fused with a flexible linker, the question arose, whether both strands can be nicked by MutH or whether one strand is preferred. Guided by the molecular model shown in Figure 1, we expected the bottom strand to be cleaved preferentially if not exclusively. The strand specificity of MutH-I-SceI was determined by sequencing the nicked product. As shown in Figure 4, the top strand remains intact and the bottom strand is cleaved 5' of the GATC site.

Determination of the cleavage preference of TALE-MutH for the addressed substrate

Similar experiments as carried out to determine the preference of the MutH-I-SceI fusion protein for addressed over unaddressed substrate cleavage were carried out for the TALE-MutH fusion proteins. As shown in Figure 5, the TALE-MutH fusion constructs are specific for the addressed bipartite recognition site, consisting of the TALE target site (here we used the AvrBs4 target site) in a defined distance next to a MutH recognition site. We tested spacer lengths of 1, 2, 3, 4, 5, 6, 7, 8 and 9 bp between the two sites on the addressed substrate and found that a distance of 3 bp (T-3-H) is optimal (Figure 5C) for nicking. No double-strand breaks were detectable.

Investigation of the strand specificity of TALE-MutH

To determine the strand specificity of TALE-MutH, we have used a 211-bp PCR product that carried different fluorophores on the 5'-ends of the top and bottom strand, respectively. Depending on which strand is preferentially nicked, the electrophoretic analysis of the nicking reaction would yield a different characteristic fluorescence image. As shown in Figure 6, TALE-MutH nicks the bottom strand of the substrate T-3-H with more than two orders of magnitude higher preference over the top strand. Intriguingly, the substrate T-6-H is nicked preferentially in the top strand, albeit by a factor of ca. 3 more slowly than the bottom strand of T-3-H. Similar experiments were performed also for the other substrates: T-1-H to T-4-H are nicked preferentially in the bottom strand, whereas T-5-H to T-9-H are nicked preferentially in the top strand. This change in preference for bottom strand nicking to top strand nicking could be correlated with the finding shown in Figure 5 that there is a decrease in the rate of nicking between T-3-H and T-6-H. Similarly, as for MutH-I-SceI (Figure 4), determination of strand specificity was also determined for the substrates T-3-H and T-6-H by sequencing. The results show that the bottom strand (T-3-H) and the top strand (T-6-H), respectively, are nicked only at the \downarrow GATC site.

Figure 2. Continued

plasmids are not nicked or cleaved. The calculated apparent cleavage rate constant for the addressed plasmid substrate is $4.0 \times 10^{-2} \text{ min}^{-1}$, whereas the apparent cleavage rate constants for both unaddressed substrates are below the detection limit of approximately $4 \times 10^{-5} \text{ min}^{-1}$.

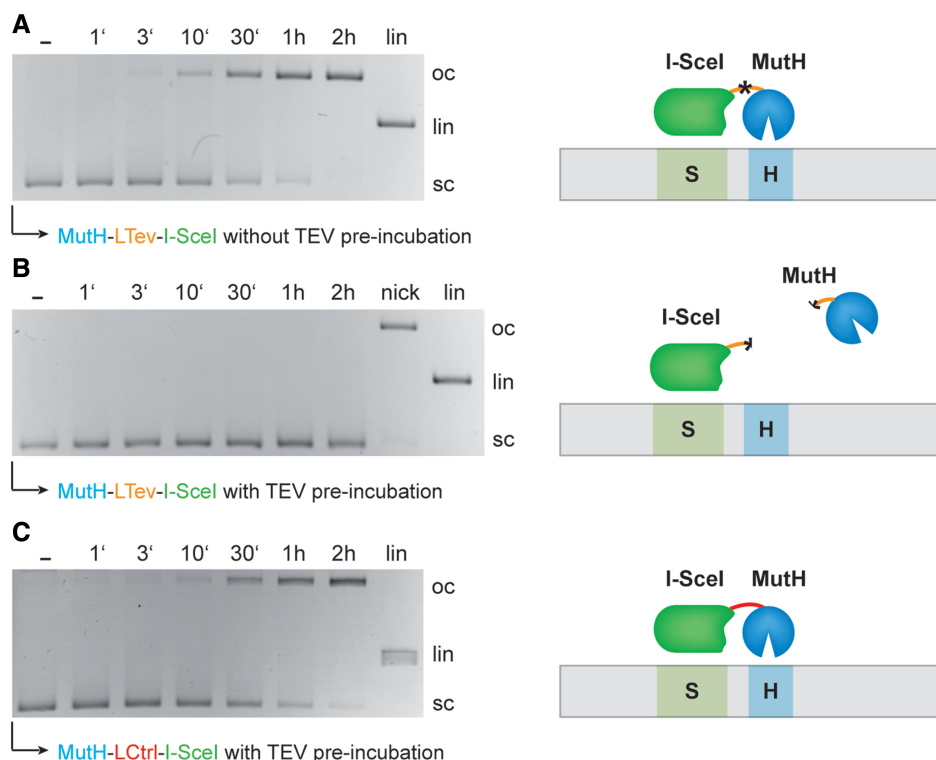


Figure 3. Effect of separating the I-SceI and the MutH module on addressed substrate nicking. The kinetics of DNA-nicking by the MutH–I-SceI fusion construct of the addressed substrate was analysed with a plasmid cleavage assay (16 nM enzyme, 4 nM substrate). The addressed substrate plasmid contains an I-SceI and a MutH recognition site separated by 3 bp (S-3-H). The substrate has 19 additional GATC sites distributed over the plasmid. (A) Assay performed with MutH–I-SceI that had not been pre-incubated with TEV protease. (B) Assay performed with MutH–I-SceI that had been pre-incubated with TEV protease. (C) For control, we have also incubated the fusion protein, in which the linker did not contain a TEV protease cleavage site, with the TEV protease. The position of the bands representing supercoiled (sc), open circular (oc) and linear (lin) forms of the plasmid is indicated. The first lane shows the reaction without $MgCl_2$ (–). ‘Nick’ and ‘lin’ are the controls for cleavage and nicking activity and correspond to the open circular and linear forms of the plasmid, respectively. Whereas in (A) and (C), nicking is observed, in (B), it is not.

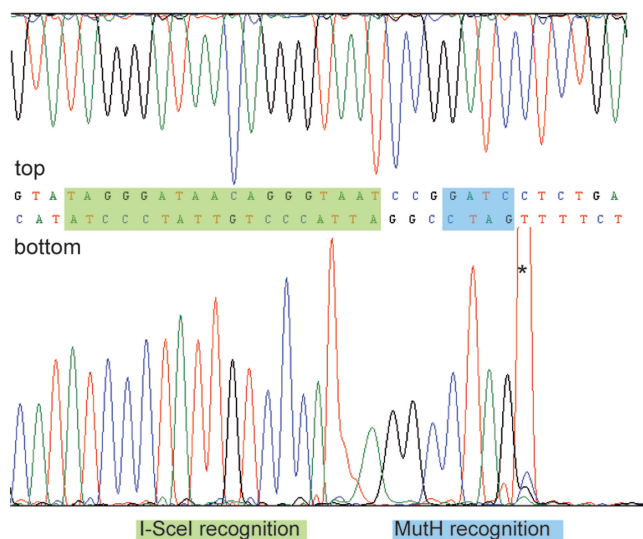


Figure 4. Determination of the strand specificity of MutH–I-SceI by sequence analysis of the nicked addressed plasmid after cleavage with the fusion construct MutH–I-SceI. The cleavage product was gel-purified and sequenced (top strand and bottom strand) over the specific site (S-3-H). For the purpose of illustration, the sequence is shown in the reverse and complementary orientation. The recognition sites of I-SceI and MutH are highlighted by green and blue bars, respectively. The sequencing results indicate that the bottom strand is cleaved 5' of the addressed GATC site. *Taq polymerase artefact.

DISCUSSION

The efficiency of homologous recombination can be increased by several orders of magnitude by a specific double-strand break at the locus of interest (49). The targeted insertion of DNA into a pre-defined locus by homologous recombination requires highly specific nucleases (50,51). Such nucleases became available with the fusion of zinc fingers to the FokI cleavage domain (52). It was demonstrated, however, that engineered double-strand-specific nucleases could introduce mutations when the double-strand break is repaired by the error-prone NHEJ pathway (53) rather than by HDR and that nicking enzymes suppress NHEJ (14,15), which means that DNA nicks can initiate efficient gene correction, with less genomic instability than a targeted DNA double-strand break. Ramirez *et al.* (33), Wang *et al.* (34) and Kim *et al.* (32) introduced Zinc finger nickases (ZFNs) for targeted gene insertion and showed that they induce HDR with reduced mutagenic effects. As there are other architectures for highly specific double-strand-specific nucleases [e.g. (11,12), M. Yanik, unpublished data], it should be possible to generate highly site- and strand-specific nickases other than the ZF-based nickases. In this article, we have shown that specificity-determining DNA-binding modules (catalytically inactive I-SceI and the DNA-binding

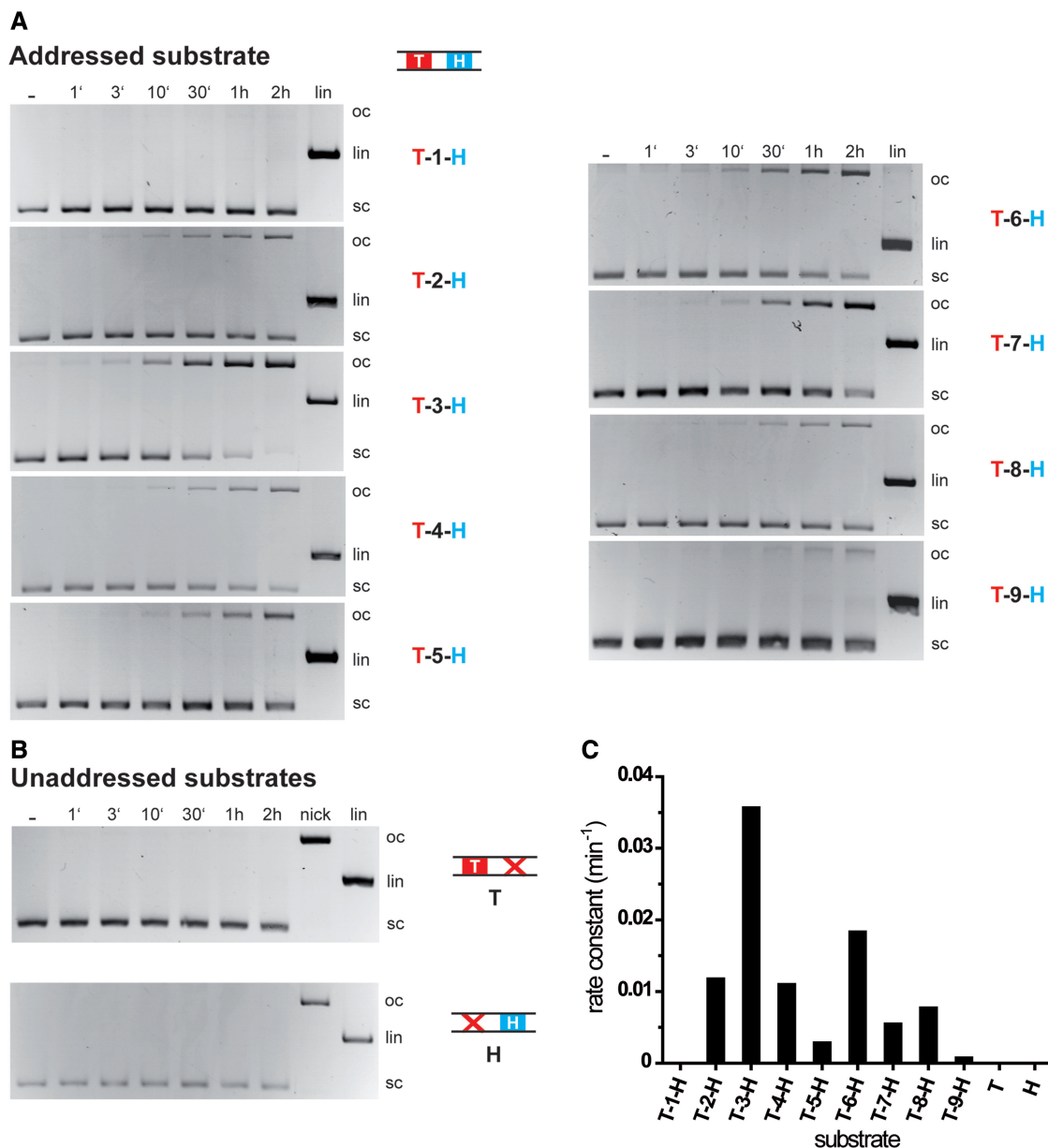


Figure 5. Kinetics of DNA-cleavage by TALE–MutH fusion constructs of addressed and unaddressed substrates analysed with plasmid cleavage assays using a 4-fold excess of enzyme over substrate (16 nM enzyme, 4 nM substrate). (A) Catalytic activity of TALE–MutH with a plasmid substrate containing an AvrBs4 and a MutH recognition site separated by x base pairs (T- x -H; $x = 1-9$ bp). (B) Catalytic activity on the unaddressed substrate plasmids (controls) containing either an AvrBs4 recognition site without a nearby MutH recognition site (T) or no AvrBs4 recognition site (H). Each of the substrates has 18 additional GATC sites distributed over the plasmid. The position of the bands representing supercoiled (sc), open circular (oc) and linear (lin) forms of the plasmid is indicated. The first lane shows the reaction without $MgCl_2$ ('-'). 'Nick' and 'lin' represent controls for nicking and cleavage activity, respectively, and correspond to the open circular and linear forms of the plasmid, respectively. (C) Calculated rate constants for TALE–MutH nicking the above-mentioned substrates. The electrophoretic analysis shows that the addressed substrate plasmid is nicked by the fusion construct TALE–MutH, whereas the unaddressed substrate plasmids remain uncleaved. The graph shows the quantitative evaluation of the experiments shown in (A) and (B). Note that the enzyme exhibits the best activity on the addressed substrate with the 3 bp spacer and no activity with the stand-alone TALE (T) or MutH (H) recognition sites. The quantitative analyses reveal the following apparent cleavage rate constants: $1.3 \times 10^{-4} \text{ min}^{-1}$ (T-1-H), $1.2 \times 10^{-2} \text{ min}^{-1}$ (T-2-H), $3.6 \times 10^{-2} \text{ min}^{-1}$ (T-3-H), $1.1 \times 10^{-2} \text{ min}^{-1}$ (T-4-H), $0.3 \times 10^{-2} \text{ min}^{-1}$ (T-5-H), $1.8 \times 10^{-2} \text{ min}^{-1}$ (T-6-H), $0.6 \times 10^{-2} \text{ min}^{-1}$ (T-7-H), $0.8 \times 10^{-2} \text{ min}^{-1}$ (T-8-H) and $8.9 \times 10^{-4} \text{ min}^{-1}$ (T-9-H); the nicking rate for T and H was below the detection limit of approximately $4 \times 10^{-5} \text{ min}^{-1}$.

domain of a TALE protein, respectively) can be fused to a specific nicking enzyme (MutH) to produce a highly sequence- and strand-specific nickase. The fusion proteins that we obtained, MutH–I–SceI and TALE–MutH, recognize their respective bipartite recognition sequence,

consisting of the recognition site of the DNA-binding module and the MutH recognition site (GATC). They only nick one strand 5' of the GATC site and do not nick stand-alone GATC sites or any other sites, making them potentially useful tools for site-specific nicking of

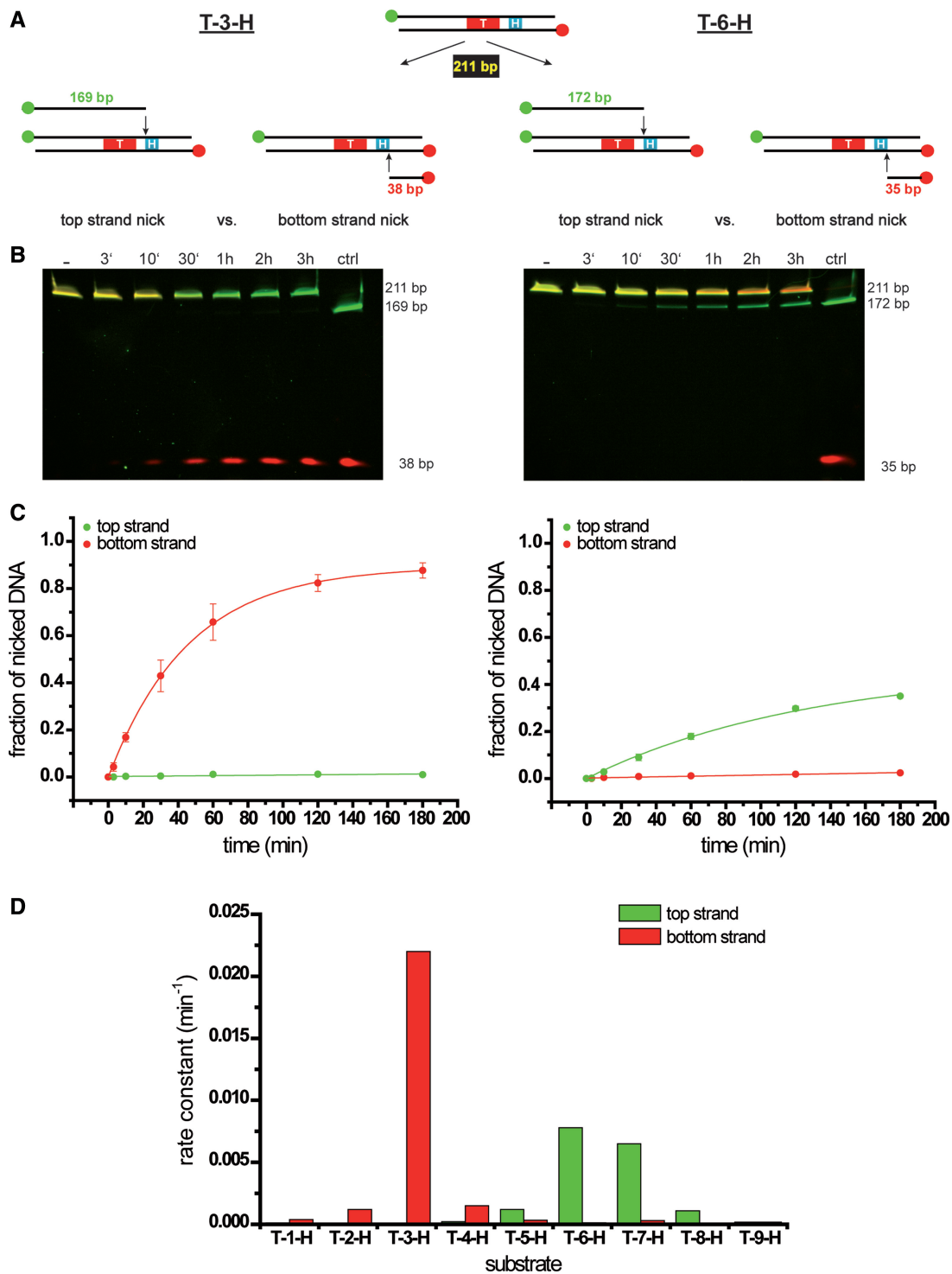


Figure 6. Determination of the strand specificity of TALE-MutH. (A) Scheme of the reaction: the 211 bp polynucleotide substrate harbours the AvrBs4 recognition site (T) and the MutH (H) recognition site separated by 3 (T-3-H; left) or 6 (T-6-H; right) bp. The substrates are 5'-labelled at the top strand with Atto 488 (green) and with Atto 467 N (red) at the bottom strand. The expected nicking products have a length of 169 (172) and 38 (35) bp for top and bottom strand nicking of the T-3-H substrate (T-6-H substrate). Which strand is attacked preferentially can be determined in the electrophoretic analysis by the characteristic fluorescence of the products. (B) DNA nicking kinetics of the addressed polynucleotide substrates T-3-H and T-6-H (right) by TALE-MutH. The nicking products for top and bottom strand of the T-3-H (T-6-H) substrate are indicated with 169 (172) and 38 (35) bp, respectively. The first lane shows the control reaction without MgCl_2 ('-'). The lane designated as 'ctrl' shows the result of the cleavage of the PCR substrate by BamHI which overlaps the MutH site and is shown here to indicate the size of the nicking products. (C) The graphs show the quantitative evaluation of the experiments shown in (B). The calculated cleavage rate constants for T-3-H (T-6-H) top strand nicking is $6.0 \times 10^{-5} \text{min}^{-1}$ ($7.8 \times 10^{-3} \text{min}^{-1}$) and for bottom strand nicking $2.2 \times 10^{-2} \text{min}^{-1}$ ($1.3 \times 10^{-4} \text{min}^{-1}$). The analysis of the strand specificity shows that the addressed PCR substrate with the 3 bp spacer is preferentially nicked in the bottom strand by the fusion construct TALE-MutH, whereas the substrate with the 6 bp spacer is preferentially nicked in the top strand. In (D), the rate constants for nicking are shown for all substrates investigated: T-1-H to T-9-H.

DNA in general and precision genome engineering in particular.

MutH is a site-specific DNA nicking enzyme which in its natural function requires complex formation with activated MutL to be directed to its target site, a hemimethylated GATC site that is nicked in the unmethylated strand. By itself and at physiological ionic strength and Mg^{2+} concentration (which we had deliberately chosen to be prepared for applications *in vivo*), MutH does not attack unmethylated GATC sites, but is strictly dependent on a covalent coupling to a DNA-binding module, as was shown here for MutH–I-SceI. Proteolytic separation of the DNA-binding and DNA-cleavage module prevents DNA cleavage. This finding suggests that MutH cannot bind in a productive manner to the bipartite recognition site, unless it is positioned properly by the I-SceI (or TALE) module as part of the fusion protein.

We had shown before (12) that catalytically inactive I-SceI can serve as a specific DNA-binding module for the restriction endonuclease PvuII. Our results with MutH–I-SceI demonstrate that this is also possible with other nucleases, here a nicking enzyme that is structurally related to PvuII (47). MutH–I-SceI recognizes a unique sequence. If such a sequence is introduced into a complex genome, this sequence could be used as a target site for genome engineering, as it was done with I-SceI sites for *in planta* gene targeting (54).

Of particular interest is the TALE–MutH fusion protein, which in contrast to MutH–I-SceI is programmable to recognize almost any DNA sequence. Similar to the MutH–I-SceI, the TALE–MutH fusion protein requires a GATC site next to the recognition sequence of its DNA-binding module, the TALE recognition sequence. This requirement should not limit the usefulness of this programmable nickase, as GATC sites occur on average every 256 bp and are unmethylated in eukaryotic genomes. In contrast to the ZF-nickases based on FokI (32–34), TALE–MutH only needs one TALE protein for targeting to a specific site, which reduces the size of a MutH-based TALE-nickase by 50% compared with a FokI-based TALE-nickase, because MutH functions as a monomer whereas FokI is a functional dimer. As the DNA-binding module of TALE proteins can be used to program the restriction endonuclease PvuII to cleave a bipartite recognition sequence consisting of the TALE recognition sequence and the PvuII recognition sequence (M. Yanik, unpublished data), we believe that other single-strand-specific nucleases can function as highly specific nickases in fusion proteins consisting of a DNA-binding module and a DNA-nicking module. Examples are restriction enzymes such as Nt.CviPII (19,20) or subunits of heterodimeric restriction endonucleases, e.g. Nt.BstD6I (21,22), Nb.BsrDI and Nt.BtsI (23) (Nt or Nb indicate the specificity for top or bottom strand nicking). Different from these naturally occurring nicking enzymes, the TALE–MutH fusion protein can nick the upper or lower strand, depending on the distance between the TALE recognition site and the GATC sequence. For example, with a spacer length of 2–4 bp (optimally T-3-H), the bottom strand is nicked with

several hundred-fold preference over the other strand. In contrast, with a spacer length of 5–8 bp (optimally T-6-H), the rate of nicking decreases but now the top strand is preferred. Substrates with a spacer length of 1 and 9 bp are hardly attacked at all. This suggests that there is some flexibility in the junction of the DNA-binding module and the DNA-cleavage module which allows MutH to reach the scissile phosphodiester bond either in the bottom strand or in the top strand, depending on the distance between the MutH recognition site relative to the TALE recognition site. We believe that decreasing or increasing the length of the linker between TALE and MutH could have a similar effect, i.e. changing the preference for bottom or top strand nicking. As we have not determined the optimal linker between AvrBs4 and MutH experimentally, but rather extrapolated the linker length (28 aa) in our TALE–MutH fusion protein from a previously constructed TALE–PvuII fusion construct, it could be that the specificity could be further increased by slight modifications in the linker length. Likewise, we have not optimized the DNA-binding module of TALE protein but rather used the natural AvrBs4 protein sequence; it could well be that exchanging some of the RVDs by ‘strong’ RVDs (55) could even further increase specificity.

ACKNOWLEDGEMENTS

The authors thank Mert Yanik, Marika Midon, Ines Fonfara, Andreas Marx and Roger Heinze for fruitful discussions and plasmids; Laura Walzl for assistance and Anja Drescher for critical reading of the article. L.G. is a member of the International Research Training Group ‘Enzymes and Multienzyme Complexes acting on Nucleic Acids’ funded by the Deutsche Forschungsgemeinschaft (DFG): B.S. has been a recipient of a grant (Just’us) by the Justus-Liebig-University Giessen.

FUNDING

Deutsche Forschungsgemeinschaft [International Research Training Group GRK 1384, and Excellence Cluster ECCPS] and the Justus-Liebig-University Giessen [Just’us]. Funding for open access charge: DFG and the Justus-Liebig-University Giessen.

Conflict of interest statement. None declared.

REFERENCES

1. 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.
2. Silva, G., Poirot, L., Galetto, R., Smith, J., Montoya, G., Duchateau, P. and Paques, F. (2011) Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Curr. Gene Ther.*, **11**, 11–27.
3. 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.
4. Ramalingam, S., Kandavelou, K., Rajenderan, R. and Chandrasegaran, S. (2011) Creating designed zinc-finger nucleases with minimal cytotoxicity. *J. Mol. Biol.*, **405**, 630–641.

5. Rahman,S.H., Maeder,M.L., Joung,J.K. and Cathomen,T. (2011) Zinc-finger nucleases for somatic gene therapy: the next frontier. *Hum. Gene. Ther.*, **22**, 925–933.
6. Carroll,D. (2011) Genome engineering with zinc-finger nucleases. *Genetics*, **188**, 773–782.
7. Christian,M., Cermak,T., Doyle,E.L., Schmidt,C., Zhang,F., Hummel,A., Bogdanove,A.J. and Voytas,D.F. (2010) TAL effector nucleases create targeted DNA double-strand breaks. *Genetics*, **186**, 757–761.
8. Li,T., Huang,S., Jiang,W.Z., Wright,D., Spalding,M.H., Weeks,D.P. and Yang,B. (2010) TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.*, **39**, 359–372.
9. 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.
10. 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.*, **39**, 9283–9293.
11. Schierling,B., Dannemann,N., Gabsalilow,L., Wende,W., Cathomen,T. and Pingoud,A. (2012) A novel zinc-finger nuclease platform with a sequence-specific cleavage module. *Nucleic Acids Res.*, **40**, 2623–2638.
12. Fonfara,I., Curth,U., Pingoud,A. and Wende,W. (2012) Creating highly specific nucleases by fusion of active restriction endonucleases and catalytically inactive homing endonucleases. *Nucleic Acids Res.*, **40**, 847–860.
13. Kleinstiver,B.P., Wolfs,J.M., Kolaczyk,T., Roberts,A.K., Hu,S.X. and Edgell,D.R. (2012) Monomeric site-specific nucleases for genome editing. *Proc. Natl Acad. Sci. USA*, **109**, 8061–8066.
14. Metzger,M.J., McConnell-Smith,A., Stoddard,B.L. and Miller,A.D. (2011) Single-strand nicks induce homologous recombination with less toxicity than double-strand breaks using an AAV vector template. *Nucleic Acids Res.*, **39**, 926–935.
15. Davis,L. and Maizels,N. (2011) DNA nicks promote efficient and safe targeted gene correction. *PLoS One*, **6**, e23981.
16. van Nierop,G.P., de Vries,A.A., Holkers,M., Vrijzen,K.R. and Goncalves,M.A. (2009) Stimulation of homology-directed gene targeting at an endogenous human locus by a nicking endonuclease. *Nucleic Acids Res.*, **37**, 5725–5736.
17. Zheleznaya,L.A., Kachalova,G.S., Artyukh,R.I., Yunusova,A.K., Perevyazova,T.A. and Matvienko,N.I. (2009) Nicking endonucleases. *Biochemistry*, **74**, 1457–1466.
18. Chan,S.H., Stoddard,B.L. and Xu,S.Y. (2011) Natural and engineered nicking endonucleases—from cleavage mechanism to engineering of strand-specificity. *Nucleic Acids Res.*, **39**, 1–18.
19. Xia,Y.N., Morgan,R., Schildkraut,I. and Van Etten,J.L. (1988) A site-specific single strand endonuclease activity induced by NYs-1 virus infection of a Chlorella-like green alga. *Nucleic Acids Res.*, **16**, 9477–9487.
20. Chan,S.H., Zhu,Z., Van Etten,J.L. and Xu,S.Y. (2004) Cloning of CviPII nicking and modification system from chlorella virus NYs-1 and application of Nt.CviPII in random DNA amplification. *Nucleic Acids Res.*, **32**, 6187–6199.
21. Zheleznaya,L.A., Perevyazova,T.A., Alzhanova,D.V. and Matvienko,N.I. (2001) Site-specific nickase from bacillus species strain d6. *Biochemistry*, **66**, 989–993.
22. Kachalova,G.S., Rogulin,E.A., Yunusova,A.K., Artyukh,R.I., Perevyazova,T.A., Matvienko,N.I., Zheleznaya,L.A. and Bartunik,H.D. (2008) Structural analysis of the heterodimeric type IIS restriction endonuclease R.BspD6I acting as a complex between a monomeric site-specific nickase and a catalytic subunit. *J. Mol. Biol.*, **384**, 489–502.
23. Xu,S.Y., Zhu,Z., Zhang,P., Chan,S.H., Samuelson,J.C., Xiao,J., Ingalls,D. and Wilson,G.G. (2007) Discovery of natural nicking endonucleases Nb.BsrDI and Nb.BtsI and engineering of top-strand nicking variants from BsrDI and BtsI. *Nucleic Acids Res.*, **35**, 4608–4618.
24. Shen,B.W., Landthaler,M., Shub,D.A. and Stoddard,B.L. (2004) DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J. Mol. Biol.*, **342**, 43–56.
25. Landthaler,M., Shen,B.W., Stoddard,B.L. and Shub,D.A. (2006) I-BaI and I-HmuI: two phage intron-encoded endonucleases with homologous DNA recognition sequences but distinct DNA specificities. *J. Mol. Biol.*, **358**, 1137–1151.
26. Landthaler,M. and Shub,D.A. (2003) The nicking homing endonuclease I-BaI is encoded by a group I intron in the DNA polymerase gene of the *Bacillus thuringiensis* phage Bastille. *Nucleic Acids Res.*, **31**, 3071–3077.
27. Wende,W., Stahl,F. and Pingoud,A. (1996) The production and characterization of artificial heterodimers of the restriction endonuclease EcoRV. *Biol. Chem.*, **377**, 625–632.
28. Stahl,F., Wende,W., Jeltsch,A. and Pingoud,A. (1996) Introduction of asymmetry in the naturally symmetric restriction endonuclease EcoRV to investigate intersubunit communication in the homodimeric protein. *Proc. Natl Acad. Sci. USA*, **93**, 6175–6180.
29. Sanders,K.L., Catto,L.E., Bellamy,S.R. and Halford,S.E. (2009) Targeting individual subunits of the FokI restriction endonuclease to specific DNA strands. *Nucleic Acids Res.*, **37**, 2105–2115.
30. Niu,Y., Tenney,K., Li,H. and Gimble,F.S. (2008) Engineering variants of the I-SceI homing endonuclease with strand-specific and site-specific DNA-nicking activity. *J. Mol. Biol.*, **382**, 188–202.
31. McConnell Smith,A., Takeuchi,R., Pellenz,S., Davis,L., Maizels,N., Monnat,R.J. Jr and Stoddard,B.L. (2009) Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-Anil LAGLIDADG homing endonuclease. *Proc. Natl Acad. Sci. USA*, **106**, 5099–5104.
32. Kim,E., Kim,S., Kim,D.H., Choi,B.S., Choi,I.Y. and Kim,J.S. (2012) Precision genome engineering with programmable DNA-nicking enzymes. *Genome Res.*, **22**, 1327–1333.
33. Ramirez,C.L., Certo,M.T., Mussolino,C., Goodwin,M.J., Cradick,T.J., McCaffrey,A.P., Cathomen,T., Scharenberg,A.M. and Joung,J.K. (2012) Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Res.*, **40**, 5560–5568.
34. Wang,J., Friedman,G., Doyon,Y., Wang,N.S., Li,C.J., Miller,J.C., Hua,K.L., Yan,J.J., Babiarz,J.E., Gregory,P.D. *et al.* (2012) Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. *Genome Res.*, **22**, 1316–1326.
35. Halford,S.E., Catto,L.E., Pernstich,C., Rusling,D.A. and Sanders,K.L. (2011) The reaction mechanism of FokI excludes the possibility of targeting zinc finger nucleases to unique DNA sites. *Biochem. Soc. Trans.*, **39**, 584–588.
36. 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.
37. Szczeppek,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.
38. Gabriel,R., Lombardo,A., Arens,A., Miller,J.C., Genovese,P., Kaepfel,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.
39. 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.
40. Lee,J.Y., Chang,J., Joseph,N., Ghirlando,R., Rao,D.N. and Yang,W. (2005) MutH complexed with hemi- and unmethylated DNAs: coupling base recognition and DNA cleavage. *Mol. Cell*, **20**, 155–166.
41. Welsh,K.M., Lu,A.L., Clark,S. and Modrich,P. (1987) Isolation and characterization of the *Escherichia coli* mutH gene product. *J. Biol. Chem.*, **262**, 15624–15629.
42. Hall,M.C. and Matson,S.W. (1999) The *Escherichia coli* MutL protein physically interacts with MutH and stimulates the MutH-associated endonuclease activity. *J. Biol. Chem.*, **274**, 1306–1312.
43. Lahue,R.S., Su,S.S. and Modrich,P. (1987) Requirement for d(GATC) sequences in *Escherichia coli* mutHLS mismatch correction. *Proc. Natl Acad. Sci. USA*, **84**, 1482–1486.

44. Modrich,P. and Lahue,R. (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu. Rev. Biochem.*, **65**, 101–133.
45. Thomas,E., Pingoud,A. and Friedhoff,P. (2002) An efficient method for the preparation of long heteroduplex DNA as substrate for mismatch repair by the *Escherichia coli* MutHLS system. *Biol. Chem.*, **383**, 1459–1462.
46. 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.
47. Ban,C. and Yang,W. (1998) Structural basis for MutH activation in *E. coli* mismatch repair and relationship of MutH to restriction endonucleases. *EMBO J.*, **17**, 1526–1534.
48. 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.
49. Jasin,M. (1996) Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet.*, **12**, 224–228.
50. Porteus,M.H. and Baltimore,D. (2003) Chimeric nucleases stimulate gene targeting in human cells. *Science*, **300**, 763.
51. 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.
52. 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.
53. 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.
54. Fauser,F., Roth,N., Pacher,M., Ilg,G., Sanchez-Fernandez,R., Biesgen,C. and Puchta,H. (2012) In planta gene targeting. *Proc. Natl Acad. Sci. USA*, **109**, 7535–7540.
55. Streubel,J., Blucher,C., Landgraf,A. and Boch,J. (2012) TAL effector RVD specificities and efficiencies. *Nat. Biotechnol.*, **30**, 593–595.
56. van Dijk,M. and Bonvin,A.M. (2009) 3D-DART: a DNA structure modelling server. *Nucleic Acids Res.*, **37**, W235–W239.