Monomeric site-specific nucleases for genome editing

Benjamin P. Kleinstiver, Jason M. Wolfs, Tomasz Kolaczyk, Alanna K. Roberts, Sherry X. Hu, and David R. Edgell

Department of Biochemistry, Schulich School of Medicine and Dentistry, University of Western Ontario, London, ON, Canada N6A 5C1

Edited* by Marlene Belfort, University at Albany, Albany, NY, and approved April 6, 2012 (received for review November 1, 2011)

Targeted manipulation of complex genomes often requires the introduction of a double-strand break at defined locations by site-specific DNA endonucleases. Here, we describe a monomeric nuclease domain derived from GIY-YIG homing endonucleases for genome-editing applications. Fusion of the GIY-YIG nuclease domain to three-member zinc-finger DNA binding domains generated chimeric GIY-zinc-finger endonucleases (GIY-ZFEs). Significantly, the I-TevI-derived fusions (Tev-ZFES) function in vitro as monomers to introduce a double-strand break, and discriminate in vitro and in bacterial and yeast assays against substrates lacking a preferred 5′-CNNNG-3′ cleavage motif. The Tev-ZFES function to induce recombination in a yeast-based assay with activity on par with a homodimeric ZfIF268 zinc-finger nuclease. We also fused the I-TevI nuclease domain to a catalytically inactive LADGLIDAGG homing endonuclease (LHE) scaffold. The monomeric Tev-LHEs are active in vivo and similarly discriminate against substrates lacking the 5′-CNNNG-3′ motif. The monomeric Tev-ZFEs and Tev-LHEs are distinct from the FokI-derived zinc-finger nuclease and TAL effector nuclease platforms as the GIY-YIG domain alleviates the requirement to design two nuclease fusions to target a given sequence, highlighting the diversity of nuclease domains with distinctive biochemical properties suitable for genome-editing applications.

Precise genome editing often requires the introduction of a double-strand break (DSB) at defined positions (1–3), and two distinct site-specific DNA endonuclease architectures have been developed toward this goal. One of these architectures relies on reprogramming the DNA-binding specificity of naturally occurring LAGLIDADGG homing endonucleases (LHEs) to target desired sequences (4, 5). The other architecture uses the reprogrammable DNA-binding specificity of zinc-finger proteins or TAL-effector domains that are fused to the nonspecific nuclease domain of the type II restriction enzyme FokI to create chimeric zinc-finger nucleases (ZFNs) or TAL effector nucleases (TALENs) (6–8). Regardless of the architecture, the underlying biology of the component proteins imposes design challenges and the relative merits of the LHE and the ZFN/TALEN architectures are the subject of much debate in the literature (6, 9). One notable constraint imposed by the FokI nuclease domain is the requirement to function as a dimer to efficiently cleave DNA (10, 11). For any given DNA target, this necessitates the design of two distinct ZFNs (or two TALENs), such that each pair of zinc finger or TAL effector domains is oriented for FokI dimerization and DNA cleavage (12).

Expanding the repertoire of DNA nuclease domains with distinctive properties is necessary to facilitate the development of new genome-editing reagents. Indeed, a number of recent studies have explored the potential of the PuvII restriction enzyme as an alternative site-specific nuclease domain for genome-editing applications (13, 14). The PuvII chimeras, however, share similar design constraints as ZFNs and TALENs, requiring two nuclease fusions for precise targeting. In considering alternative nuclease domains for genome editing, we were intrigued by the properties of the GIY-YIG nuclease domain that is associated with a variety of proteins of diverse cellular functions (15). The small (~100 aa) globular GIY-YIG domain is characterized by a structurally conserved central three-stranded antiparallel β-sheet, with catalytic residues positioned to use a single metal ion to promote DNA hydrolysis (16–18). Intriguingly, the GIY-YIG homing endonucleases, typified by the isoschizomers I-TevI and I-Bmol (19), bind DNA as monomers (20), and generate a DSB with 2-nm, 3′ overhangs. It is unknown, however, if GIY-YIG homing endonucleases function as monomers in all steps of the reaction, as the oligomeric status during cleavage has yet to be studied. Notably, GIY-YIG homing endonucleases prefer a specific DNA sequence to generate a DSB (21, 22). For I-TevI, the bottom (1) and top (1) strand nicking sites lie within a 5′-CNN[NN](G-3′ motif (CNNNG), with the critical G-optimally positioned ~28 bp from where the helix-turn-helix (H-T-H) module of the I-TevI DNA-binding domain interacts with substrate (21, 22). From an engineering perspective, the modularity and sequence specificity of the GIY-YIG nuclease domain makes it an appealing candidate to create new endonucleases. Indeed, swapping of the I-Bmol and I-TevI catalytic and DNA-binding domains suggested that the GIY-YIG nuclease domain could be fused to unrelated DNA-targeting platforms (23).

To highlight the genome engineering potential of the GIY-YIG nuclease domain, we fused the domain to three-member zinc fingers to construct GIY-YIG zinc-finger endonucleases (GIY-ZFEs). The GIY-ZFEs are active in bacterial and yeast cells, and in vitro data show that they function catalytically as monomers and retain the cleavage specificity associated with the parental GIY-YIG nuclease domain. The GIY-YIG nuclease domain is also portable to the LHE platform, as we constructed monomeric GIY-LHEs that are active in vivo and possess ~18-bp binding specificity. We selected LHEs as a DNA-targeting domain because of the greater sequence specificity compared with three-member zinc fingers, the ability to reprogram LHE DNA-binding specificity (24–26), and recent success in generating PuvII-LHE fusions (13). Collectively, our data highlight the unique biochemical properties of the GIY-YIG nuclease domain as an alternative to the FokI nuclease domain for genome-editing applications.

Results

Construction and Validation of GIY-ZFEs. To create novel chimeric enzymes, we modeled GIY-ZFEs using existing crystal structures of the I-TevI 130C DNA binding domain and the ZfIF268 zinc finger (Fig. L4) (27, 28). One notable feature of our constructs is the polarity, as the I-TevI nuclease domain is fused to the N-terminal end of the three-finger rYa zinc-finger protein to mimic its native orientation, unlike FokI constructs that are fused to the C-terminal end of zinc-finger proteins. We modeled the ZfIF268 zinc finger in place of the H-T-H module at the C terminus of I-TevI, providing the rationale to subsequently fuse various lengths of the I-TevI N-terminal region to the rYa zinc finger that targets a sequence in the Drosophila rosy gene to create Tev-rYa zinc-finger endonucleases (Tev-ZFES) (Fig. L4) (29). The Tev-zinc


The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

1To whom correspondence should be addressed. E-mail: dedgell@uwo.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1117984109/-/DCSupplemental.
finger DNA substrates (TZ) consisted of 30–38 bps of the I-TevI td homing site joined to the 9-bp ryA target site. The TZ substrates differ in the distance of the CNNNG cleavage motif relative to the ryA-binding site (Fig. 1 B). Each TZ substrate possesses a single zinc-finger targeting sequence, rather than two head-to-head zinc-finger sites necessary for efficient ZFN cleavage. A similar set of I-BmoI-ryA fusions (Bmo-ZFEs) and substrates (BZ) were constructed (Fig. S1).

We tested the activity of the GIY-ZFEs using a well-described two-plasmid bacterial selection system, where survival is dependent on the endonuclease cleaving a target plasmid (30, 31). Eight Tev-ZFEs were tested on seven TZ substrates cloned into the reporter plasmid (Fig. 1 B and C and Table S1). In general, the survival of all Tev-ZFEs was highest against TZ substrates where the preferred CNNNG motif was positioned between 33 and 35 bp from the ryA binding site. Low survival (~4–6%) was observed for all Tev-ZFEs against the TZ1.32 substrate, but none survived on the TZ1.30 substrate. Similarly, there was no survival against the longer substrates, with the exception that the longest fusion (TevS206-ryA) exhibited ~22% survival against the TZ1.36 substrate. No survival was observed when the Tev-ZFEs were tested against the target plasmid without a target site (p11lacYwtx1). Mutation of the catalytic arginine 27 of the I-TevI nuclease domain to alanine to create TevR27A allowed cleavage by TevN201-ryA for cleavage assays and in vitro mapping. We first performed cleavage assays to determine the relationship between TevN201-ryA enzyme concentration and initial reaction velocity using a plasmid substrate with a single TZ-ryA target site. The reaction progress curves indicated an initial burst of cleavage followed by a slower rate of product accumulation (Fig. 2 A), consistent with product release being the rate-limiting step. The initial burst phase was used to estimate initial velocity, and plotting against protein concentration yielded a linear relationship (Fig. 2 A), suggesting that DNA hydrolysis catalyzed by TevN201-ryA is first order with respect to protein concentration.

The model TZ-ryA substrates were designed as a single ryA zinc-finger site fused to the I-TevI target sequence. To determine if cleavage by TevN201-ryA was influenced by additional Tev-ryA target sites, we constructed two-site plasmids that differed in whether the target sites were in the same or opposite orientations relative to each other. The single- or two-site plasmids were used in time-course cleavage assays under single-turnover conditions (~10-fold molar excess of protein to substrate) to determine reaction rates. As shown in Fig. 2 B, cleavage of the one-site plasmid yielded $k_{\text{obs(1-site)}} = 0.099 \pm 0.001 \text{ s}^{-1}$, and cleavage of the two-site plasmids with target sites in the opposite or same (Fig. 2 B) orientations generated very similar rate constants, $k_{\text{obs(2-site)}} = 0.088 \pm 0.001 \text{ s}^{-1}$ and 0.089 ± 0.001 s$^{-1}$, respectively, to the one-site plasmid. In contrast, similar experiments with FokI showed a

---

**Fig. 1.** Design and functionality of Tev-ZFEs. (A) Modeling of a Tev-zinc finger fusion with DNA substrate (light green) using structures of the I-TevI catalytic domain in green (PDB 1M8O), the I-TevI DNA-binding domain crysyal in blue (PDB 1I3D), and the Zif268 crysyal in red (PDB 1AAY). (B) The TZ-ryA substrate is colored according to the structural model. Shown is the top strand of the I-TevI td homing site substrate fused to the 5′ end of the ryA-binding site for all wild-type substrates tested. The substrate is numbered from the first base of the td homing site sequence (the numbering scheme is reverse of that used for the native td homing site). The substrates tested differ by insertion or deletion of td sequence at the junction of the td/ryA sites. (C) Percent survival of three representative Tev-ryA ZFEs in the bacterial two-plasmid selection. All Tev-ryA ZFEs were tested against plasmids containing various length substrates (T21.30–1.38), plasmids lacking a target site (p11lacY), and T21.33 plasmids with single or double mutations in the CNNNG motif (GSA and C1A/GSA) (Table S1). (D) Percent survival of TevN201-ryA and TevN201-ryB ZFEs on their cognate and reciprocal target sites. Data are plotted with SD for n ≥ 3.
significant rate enhancement for two-site plasmids relative to one-site plasmids, consistent with FokI functioning as a dimer (33). We conclude that cleavage by TevN201-ryA is non-cooperative and that efficient DNA hydrolysis does not require two sites, consistent with TevN201-ryA functioning catalytically as a monomer.

The I-TevI nuclease domain preferentially cleaves DNA within a 5′-CNNNG motif, with ↑ and ↓ representing the bottom- and top-strand nicking sites, respectively (22). I-TevI defaults to cleave at the wild-type distance on substrates in vitro when this motif is moved closer to, or distant from, the primary binding site, whereas mutants in the I-TevI-specific zinc finger cleave at the correct sequence rather than the wild-type distance on mutant substrates (34). To determine the cleavage preference of the TevN201-ryA construct, we mapped the bottom- and top-strand nicking sites using strand-specific end-labeled substrates to the CNNNG motif (Fig. 2C). Combined with data from the genetic assays showing no survival on substrates that displace the CNNNG motif (Fig. 2D), the level of cleavage seen in vitro with monomeric TevN201-ZFE is consistent with the cleavage site preference of I-TevI (Fig. S1D) (19). To extend the in vivo relevance of the Tev-ZFE fusions, we used a well-described yeast-based recombination assay to test Tev-ZFE function in a eukaryotic system (35). This assay provides a quantitative β-galactosidase readout if the nuclease cleaves its target site that is positioned between a partially duplicated lacZ gene. Furthermore, the assay allowed us to calibrate TevN201-ryA activity relative to a homodimeric FokI-Zif268 control with previously measured in vivo activity sufficient to induce recombination events for genome engineering applications (35).

In Fig. 2D, ~60-fold and ~4.7-fold more protein were required to achieve half-maximal cleavage of the double- and single-mutant substrates relative to the wild-type substrate. The greater substrate discrimination observed in the genetic assay likely reflects lower in vivo protein concentrations than those used for in vitro cleavage assays. These results show that the TevN201-ryA fusion retains the cleavage specificity of the parental I-TevI enzyme and that double nucleotide substitutions significantly reduce cleavage efficiency. To determine if Bmo-ZFEs also retained substrate specificity, the bottom- and top-strand nicking sites of the BmoN221-ryA fusion were mapped to a 5′-NNI[N]G-3′ motif, consistent with the cleavage site preference of I-BmoI (Fig. S1D) (19).

**Tev-ZFEs Function in a Yeast-Based Recombination Assay.** To extend the in vivo relevance of the Tev-ZFE fusions, we used a well-described yeast-based recombination assay to test Tev-ZFE function in a eukaryotic system (35). This assay provides a quantitative β-galactosidase readout if the nuclease cleaves its target site that is positioned between a partially duplicated lacZ gene. Furthermore, the assay allowed us to calibrate TevN201-ryA activity relative to a homodimeric FokI-Zif268 control with previously measured in vivo activity sufficient to induce recombination events for genome engineering applications (35). As shown in Fig. 3, the level of β-galactosidase activity for the TevN201-ryA fusion on its cognate TZ-ryA substrate was ~1.4-fold higher than the Zif268 ZFN control. The TevN201-ryA or Zif268 ZFN constructs displayed no activity on each other’s substrates, and activity was dependent on a functional I-TevI nuclease domain, as the TevN201R27A catalytic mutant was unable to induce recombination. Furthermore, TevN201-ryA activity was not observed on mutant substrates where one or both of the critical residues of the CNNNG motif were mutated.

In Fig. 2D, ~60-fold and ~4.7-fold more protein were required to achieve half-maximal cleavage of the double- and single-mutant substrates relative to the wild-type substrate. The greater substrate discrimination observed in the genetic assay likely reflects lower in vivo protein concentrations than those used for in vitro cleavage assays. These results show that the TevN201-ryA fusion retains the cleavage specificity of the parental I-TevI enzyme and that double nucleotide substitutions significantly reduce cleavage efficiency. To determine if Bmo-ZFEs also retained substrate specificity, the bottom- and top-strand nicking sites of the BmoN221-ryA fusion were mapped to a 5′-NNI[N]G-3′ motif, consistent with the cleavage site preference of I-BmoI (Fig. S1D) (19).

**Tev-ZFEs Function in a Yeast-Based Recombination Assay.** To extend the in vivo relevance of the Tev-ZFE fusions, we used a well-described yeast-based recombination assay to test Tev-ZFE function in a eukaryotic system (35). This assay provides a quantitative β-galactosidase readout if the nuclease cleaves its target site that is positioned between a partially duplicated lacZ gene. Furthermore, the assay allowed us to calibrate TevN201-ryA activity relative to a homodimeric FokI-Zif268 control with previously measured in vivo activity sufficient to induce recombination events for genome engineering applications (35). As shown in Fig. 3, the level of β-galactosidase activity for the TevN201-ryA fusion on its cognate TZ-ryA substrate was ~1.4-fold higher than the Zif268 ZFN control. The TevN201-ryA or Zif268 ZFN constructs displayed no activity on each other’s substrates, and activity was dependent on a functional I-TevI nuclease domain, as the TevN201R27A catalytic mutant was unable to induce recombination. Furthermore, TevN201-ryA activity was not observed on mutant substrates where one or both of the critical residues of the CNNNG motif were mutated.

The I-TevI nuclease domain preferentially cleaves DNA within a 5′-CNNNG motif, with ↑ and ↓ representing the bottom- and top-strand nicking sites, respectively (22). I-TevI defaults to cleave at the wild-type distance on substrates in vitro when this motif is moved closer to, or distant from, the primary binding site, whereas mutants in the I-TevI-specific zinc finger cleave at the correct sequence rather than the wild-type distance on mutant substrates (34). To determine the cleavage preference of the TevN201-ryA construct, we mapped the bottom- and top-strand nicking sites using strand-specific end-labeled substrates to the CNNNG motif (Fig. 2C). Combined with data from the genetic assays showing no survival on substrates that displace the CNNNG motif from an optimal position, our data suggest that in the context of a ryA fusion, the TevN201 domain acts as a molecular ruler with a distance preference.

To further demonstrate TevN201-ryA cleavage preference, we introduced mutations in the CNNNG motif that were previously shown to drastically reduce I-TevI cleavage efficiency (Fig. 2D) (21, 22). Significantly, we observed no survival under selective conditions in the two-plasmid assay on plasmids carrying either the single G5A (CNNNA) or double C1A/G5A (ANNNA) substitutions (Fig. 1C), equivalent to positions C-27 and G-23 of the I-TevI nd substrate, respectively. We also performed in vitro cleavage assays on wild-type and mutant substrates with increasing concentrations of TevN201-ZFE to determine the amount of protein required for half-maximal cleavage (EC_{0.5max}). As shown in Fig. 2D, ~60-fold and ~4.7-fold more protein were required to achieve half-maximal cleavage of the double- and single-mutant substrates relative to the wild-type substrate. The greater substrate discrimination observed in the genetic assay likely reflects lower in vivo protein concentrations than those used for in vitro cleavage assays. These results show that the TevN201-ryA fusion retains the cleavage specificity of the parental I-TevI enzyme and that double nucleotide substitutions significantly reduce cleavage efficiency. To determine if Bmo-ZFEs also retained substrate specificity, the bottom- and top-strand nicking sites of the BmoN221-ryA fusion were mapped to a 5′-NNI[N]G-3′ motif, consistent with the cleavage site preference of I-BmoI (Fig. S1D) (19).
in the TZ1.33 substrate. Collectively, these assays show that the I-TevI nuclease domain functions in a eukaryotic system with activity on par to a characterized ZFN.

I-TevI Nuclease Domain Is Portable to the LAGLIDADG Architecture.

To demonstrate that the I-TevI nuclease domain functions in the context of DNA-targeting platforms with greater specificity than three-member zinc fingers, we constructed fusions of the domain to a catalytically inactive monomeric single-chain LAGLIDADG homing endonuclease (Tev-LHE). As with the Tev-ZFE constructs, we modeled a Tev-LHE chimera using the cocrystal of I-OnuI with its DNA substrate such that the I-TevI nuclease and linker domains were fused to the N terminus of I-OnuI, which is partially disordered in the structure (Fig. 4A) (25). Based on this model, we fused TevN201G4 and TevK203 fragments to a catalytically dead I-OnuI E1 E22Q mutant. A series of model DNA substrates were constructed by fusing the td target site to the I-OnuI E1 binding site in the human MAO-B gene, differing in the position of the CNNNG cleavage motif relative to the I-OnuI E1 site (TO1.12 to TO1.30) (Fig. 4B).

In the bacterial two-plasmid selection, we found that the TevN201G4-Onu and TevK203-Onu fusions were active against a range of DNA substrates. Notably, the fusions displayed maximal survival on longer targets (TO1.26, TO1.28, and TO1.30), and lower survival against shorter targets (TO1.18 and TO1.20 targets). The two groups of substrate differ by approximately one helical turn of DNA, meaning that the preferred CNNNG motif would be presented on the same face of the substrate, even though the motif is closer to the I-OnuI E1 binding site on the shorter targets. Similar periodic cleavage patterns have been observed in vitro with I-TevI on substrates with a displaced CNNNG motif (36). This result also implies that the N terminus of I-OnuI possesses inherent flexibility to allow the I-TevI nuclease domain to search out the CNNNG motif, in contrast to the ruler-like behavior of the Tev-ZFE constructs, likely because the zinc-finger N terminus is inflexible. Importantly, the Tev-Onu fusions were not active against the TZ-ryA zinc-finger substrates (Table S2), showing that the LHE, and not the I-TevI linker, directs DNA targeting. Survival was also dependent on an active I-TevI nuclease domain, as TevR27A fusions in the context of the I-OnuI E22Q mutant did not survive (Fig. 4D). Conversely, the targeting and activity of wild-type I-OnuI E1 was not affected by fusion of the I-TevI domain, because the TevR27A-OnuWT fusions survived against TO substrates (Fig. 4D).

The apparent flexibility of the N terminus and the greater specificity of I-OnuI prompted us to test fusions containing shorter fragments of the I-TevI nuclease domain (Fig. 4A). Based on structural and genetic data, we constructed TevS114-Onu, TevD127-Onu, TevN140-Onu, TevN169-Onu, and TevD184G2-Onu fusions, progressively removing amino acid residues of I-TevI that make specific base pair contacts to the td substrate (28) (Fig. 4A). Notably, the TevS114, TevD127, TevN140, and...

---

**Fig. 3.** Tev-ZFES can induce recombination in a eukaryotic system. Shown are normalized β-galactosidase units from a yeast-based recombination assay for the indicated nuclease/substrate combinations. Activity was normalized to a homodimeric FokI-Zif268 ZFN-positive control. Data are plotted with SD for n = 4.

**Fig. 4.** Design and functionality of Tev-LHEs. (A) Modeling of a Tev-Onu E1 fusion with DNA substrate (light green) using structures of the I-TevI catalytic domain in green (PDB 1MK0), the I-TevI DNA-binding domain cocrystal in blue (PDB 1I3J), and the I-OnuI cocrystal in red (PDB 3QQY). Shown are fusion points at which the I-TevI fragment has been shortened. (B) The Tev-Onu E1 (TO) substrate is colored according to the structural model. Shown is the top strand of the I-TevI td homing site substrate fused to the 5’ end of the Onu E1-binding site. The substrates are numbered from the first base of the td homing site sequence and differ by the deletion of td nucleotides at the junction of the td/Onu E1 sites. (C) Percent survival of Tev-LHEs in the bacterial two-plasmid selection with various length target sites (TO1.12–1.30). All Tev-LHEs tested were in the I-OnuI E1 E22Q background. (D) Percent survival of TevR27A(N201G4)-OnuE1 and TevR27A(N201G4)-OnuE1(E22Q) on TO1.30, TO1.30G5A, and T21.33. Data are plotted with SD for n = 3.
TevN169 removed the α-helix that binds in the minor groove, as well as residues shown by structural data to make base-specific contacts (28). The TevS114 fusion point lies at the boundary of the deletion tolerant region of the I-TevI linker, and represents a functionally minimal GIY-YIG nuclease domain (36, 37). We found that the shorter fusions were not active against the longer TO1.28 and TO1.30 substrates, yet displayed the same periodic activity on the shorter substrates (Fig. 4C and Table S2). A single exception was the TevD184G2 fusion that showed low survival against the TO1.22 substrates, against which no other fusion survived. No survival was observed on mutant substrates that contained single (CNNNA) or double (ANNNA) mutations in the CNNNG motif, recapitulating the necessity for an appropriately positioned CNNNG as seen with the Tev-ZFE fusions.

5′-CNNNG-3′ Cleavage Motif Is Not Limiting for Targeting. An important consideration in the design of GIY-ZFEs or GIY-LHEs for genome-editing applications is the targeting requirements, notably the need for the CNNNG di-nucleotide cleavage motif (Fig. 5A). In a complex genome of \(~3 \times 10^9\) bp, the statistically predicted occurrence of the CNNNG motif is once every 15 bp, assuming a 50% GC content. To determine if the frequency of the CNNNG motif would be limiting for targeting applications, we examined 35 bp flanking 8,829 computationally predicted ZFN sites on zebrafish chromosome 1 for the occurrence of the CNNNG motif (38). As shown in Fig. 5B, the motif is highly represented at all positions within a 35-bp window relative to the ZFN sites. Of the 8,829 sites examined, 88% (7,845) of ZFN sites possessed at least one motif within 35 bp of the predicted binding site (Fig. 5C). These requirements contrast sharply with those of the recently described PuvII-LHEs and PuvII-ZFNs that require the 6-bp 5′-CAGCTG-3′ PuvII site in addition to the LHE or ZF binding site (13, 14). Of the 8,829 ZFN sites, 97% lacked a PvuII site within the 35-bp window (Fig. S3). Thus, the requirement for a di-nucleotide cleavage motif in the context of a GIY-ZFE or GIY-LHE will not severely limit potential targeting sites.

Discussion

Here, we provide evidence that the GIY-YIG nuclease domain is a potential alternative to the currently used FokI nuclease domain for genome-editing applications. We show that the I-TevI GIY-YIG nuclease domain is portable to two reprogrammable DNA-binding scaffolds, the three-member zinc fingers, and LAGLIDADG homing endonucleases. The Tev-ZFE and Tev-LHE fusions are active in vitro and in vivo, with the activity of the Tev-ZFE in a yeast-based recombination assay on par with that of a characterized ZFN. We foresee the monoclonal nature of the Tev-LHEs and Tev-ZFEs as a key advantage over existing ZFNs and TALENs, because a single fusion protein need be designed to target a given sequence, rather than two ZFNs or TALENs required to promote dimerization of the FokI nuclease domain (12). Moreover, the fact that the I-TevI nuclease domain possesses a preferred cleavage motif adds another layer of specificity to targeting requirements, potentially limiting DSBs at off-target sites that do not possess the cleavage motif.

One targeting consideration for chimeric GIY-YIG endonucleases is the DNA sequence requirement of the I-TevI linker. The I-TevI linker is a complex structure, consisting of defined structural elements with distinct roles in I-TevI function (28, 34, 36). The primary role of the linker is to position the nuclease domain on substrate for cleavage at the CNNNG motif, which is found at a defined distance from the binding site on naturally occurring I-TevI substrates. However, the linker can direct the nuclease domain in vitro to search out displaced CNNNG motifs on both native and nonnative substrates with insertions or deletions, albeit with reduced cleavage efficiency. Our Tev-LHE fusions recapitulate this distance versus sequence behavior in vivo, as the fusions can cleave displaced CNNNG motifs with a periodicity that parallels the helical nature of DNA. We partially attribute this ability of the Tev-Onu fusions to the flexible N terminus of I-Onu. The substrate flexibility of different length Tev-Onu fusions is an important consideration for targeting, as CNNNG motifs at various positions relative to the LHE binding site would be accessible by the choice of the appropriate Tev-LHE fusion. In contrast, the apparently inflexible N terminus of the three-member zinc fingers constrains cleavage to a distance of 33–36 bp from the ryA-binding site, mimicking the spacing of the CNNNG motif on native ryA substrate. Our longest Tev-ZFE and Tev-LHE fusions encompass all of the known elements of the I-TevI linker that make multiple base-specific and nonspecific contacts to DNA (28). However, biochemical studies revealed that I-TevI retains significant cleavage activity on substrates with multiple substitutions in the central region of its cognate DNA substrate that is contacted by the linker, equivalent to positions 6–33 of our longest chimeric substrates (39). The shortest Tev-LHE fusions do not contain any linker elements that are known to make base-specific DNA contacts, and cleave only at the preferred CNNNG motif. This observation implies that the I-TevI linker may contact substrate nucleotides adjacent to the CNNNG motif. Potential contacts may play a role in the positioning of the nuclease domain, rather than being necessary for cleavage, and any preference may be related to regulating the position of the nuclease domain on substrate or the maintenance of DNA-structure (36).

Future work on Tev-ZFEs and Tev-LHEs will require a detailed dissection of binding affinity and specificity, and characterization of cellular toxicity that results from cleavage at off-target sites. In their current form, the targeting specificity of the Tev-ZFEs is a function of the three-zinc finger domain, which could be further enhanced by addition of zinc fingers to generate a four-, five-, or six-zinc finger
fusion to increase specificity, as has been done with a variety of ZFNs (40). In contrast, the ~18-bp specificity of LHEs is sufficient to direct targeting and cleavage at endogenous loci in human cells. LHEs, however, are tolerant of nucleotide substitutions within their recognition sequence, and I-OnuI E1 cleaves off-target sites that differ by one or two nucleotide substitutions (25). In the context of Tev-LHEs, decoupling of DNA-cleavage and DNA-binding activity by using a catalytically dead LHE scaffold, combined with the requirement for a preferred I-TevI CNNNG cleavage motif, would significantly reduce cleavage at off-target sites (Fig. S4). Another advantage of the decoupled activities of Tev-LHEs is that they would not require reoptimization of catalytic activity that is often necessary in LHEs that have been reprogrammed to bind nonnative target sites (25, 41). Similar to the exploration of alternative DNA-binding platforms (2), it is imperative to incorporate nuclease domains with distinct biochemical properties into the genome-engineering pipeline to create highly precise tools. With further optimization, the I-TevI nuclease domain may become an alternative to the FokI-derived ZFNs and TALENs.

Materials and Methods
See detailed SI Materials and Methods for further discussion, Fig. S5 for amino acid sequences of GIY-ZFES and Tev-LHEs, Table S3 for strains and plasmids used in this study, and Table S4 for oligonucleotides used in this study. Tev-ZFE and Tev-LHE fusion vectors and hybrid target sites were modeled in PyMOL using the I-TevI 130C (PDB 1I3J), Zif268 (PDB 1AAY), and I-OnuI (PDB 3QQY) cocrystal structures (25, 27, 28). The in vivo activity of fusions was determined using a two-plasmid bacterial selection (31), and a yeast-based reporter assay was used to calibrate activity of the Tev-ZFE against a characterized ZFN (35). TevN201-αNy was purified using nickel affinity chromatography to determine the in vitro biochemical properties of Tev-ZFES. Cleavage assays were performed as previously described (42). A custom Perl script (Dataset S1) was created to determine CNNNG site occurrences relative to 8,829 predicted ZFN sites on zebrafish chromosome 1 (38).

Acknowledgments. We thank Dr. Adam Bogdanove for providing Saccharomyces cerevisiae strains and reporter assay plasmids; Dr. Chris Brandl for yeast culture and assay reagents; and Dr. Greg Gloer, Dr. Megan Davey, Clarice Schmidt, Julie Genereaux, and Russell Dickson for invaluable advice and assistance. This work was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (to D.R.E.), and an Alexander Graham Bell Canada Graduate Scholarship from the Natural Sciences and Engineering Research Council of Canada (to B.P.K.).

References
Supporting Information

Kleinstiver et al. 10.1073/pnas.1117984109

SI Materials and Methods

Bacterial Strains and Plasmid Construction. Escherichia coli strains DH5α and ER2566 (New England Biolabs) were used for plasmid manipulations and protein expression, respectively. E. coli strains BW25141(ΔDE3) was used for genetic selection assays (1). A complete description of all plasmids used in this study are listed in Table S3, and oligonucleotides are listed in Table S4. The ryA and ryB zinc-finger genes were synthesized by Integrated DNA Technologies with 5′- BamHI and 3′-XhoI sites and a C-terminal 6-histidine tag, and cloned into pACYC/duet-1 to generate pACYCryAZf+H and pACYCryBZF+H, respectively. A stop codon was introduced at the 3′ end of the ryAZ/ry gene using Quikchange (Stratagene) to generate pACYCryAZf. To create GIY-zinc finger endonucleases (GIY-ZFEs), the I-Tevl and 1-BmoI GIY-YIG domains were PCR-amplified from bacteriophage T4 gDNA and pACYC/Bindol, respectively, and cloned into the NcoI/BamHI sites of pACYCryAZf+H, pACYCryAzi, and pACYCryBzF+H. TevN201-ryA and TevN201r27A were subcloned into the XbAl and EcoRV sites of pTAL3 to generate the expression plasmids for the yeast reporter assay (pYTZN201 and pYTZN201r27A). To generate I-Tevl LADGLIDADG homing endonucleases (Tev-LHes), the I-Onul E1 gene was amplified with BamHI and SalI ends to clone into the BamHI and XhoI sites of pACYC/duet-1 (PciI) to create pACYCOnuE1(+H). This vector was subsequently Quikchanged to introduce an E22Q mutation in I-Onul E1 to create pACYCOnuE1E22Q (+H). I-Tevl catalytic domains were amplified as above and cloned into PciI/BamHI of pACYC/OnuE1+E22Q (+H). The R27A mutants of I-Tevl-zinc finger endonucleases (Tev-ZFEs) and Tev-LHes were generated using Quikchange mutagenesis. Hybrid GIY-ZFE and Tev-LHIE target constructs (Figs. 1B and 4B, and Tables S1 and S2) were cloned into the toxic plasmid p11-lac-ywtx1 to generate reporter plasmids for the bacterial selection. Tev-ryA and Bmo-ryA target sites were cloned into pSP72 for in vitro cleavage assays. The Tev-ryA site hybrid homing site was also cloned into LITMUS28i using BamHI and XhoI to generate pTZHS1.35. The two-site Tev-ZF plasmids were created by subcloning the PvuII/HpaI fragment from pTZHS1.35 into the SwaI site of pTZHS1.35 to generate pTZHS2.35 and pTZHS3.35 (with the second TZHS in either orientation). The G5A or C1A/G5A mutations were introduced into pToxTZ and pTZHS plasmids by Quikchange mutagenesis. To generate the target plasmids for the yeast reporter assay, the TZ-ryA target sites from toxic plasmids containing TZ1.33, TZ1.33G5A, and TZ1.33C1A/G5A were amplified and cloned into the BglII and SpeI sites of pCP5.1. All constructs were verified by sequencing, and the amino acid sequences of all GIY-ZFEs and Tev-LHes constructed are provided in Fig. S5.

Two-Plasmid Genetic Selection. The two plasmid genetic selection was performed as previously described (1), with toxic (reporter) plasmids containing hybrid TZ-ryA, TZ-ryB, BZ-ryA, TO target sites (Tables S1 and S2), mutant target sites (with G5A or C1A/ G5A substitutions), or plasmids lacking a target site (p11-lac-ywtx1). Survival percentage was calculated by dividing the number of colonies observed on selective plates by those observed on nonselective plates.

Yeast Reporter Assay. Transformants of Saccharomyces cerevisiae YPH500(a) with Tev-ZFE constructs and YPH499(a) with target constructs were grown overnight (~230 rpm) at 30 °C in synthetic complete medium lacking histidine (Tev-ZFEs) or lacking tryptophan and uracil (targets). Tev-ZFEs and targets were mated by adding equal densities (~400 μL) of overnight culture to 1 mL YPD and left stationary for 5-6 h at 30 °C. Cells harvested by centrifugation were washed in 1 mL and resuspended in 4 mL of synthetic medium lacking histidine and tryptophan before shaking overnight at 30 °C. Cells were harvested by centrifugation, washed in 1 mL Z buffer (60 mM Na2PO4, 40 mM NaH2PO4, 10 mM KCl, 1 mM MgSO4, pH 7.0), and suspended in 250 μL Z buffer. The suspension was diluted 20-fold into 1 mL Z buffer containing 0.27% β-mercaptoethanol, and 75 μL CHCl3 and 45 μL 0.1% SDS were added before vortexing. Lysates were preincubated at 30 °C before the addition of 100 μL 4 mg/mL ortho nitrophenol (ONPG). Reactions proceeded until a yellow color developed whereby progress was stopped by the addition of 300 μL 1 M Na2CO3. Stopped reactions were pelleted and the absorbance of the supernatant was analyzed at 420 nm and 550 nm.

Protein Purification. Cultures overexpressing either TevN201-ZFE or BmoN221-ZFE were grown at 37 °C to an OD600~0.5 and expression induced by 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG; Bio Basic) overnight at 15 °C. Cells were harvested by centrifugation at 8,983 × g for 12 min, resuspended in binding buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mMimidazole, 5% (vol/vol) glycerol, and 1 mM DTT), and lysed by homogenization at 15,000 psi. The cell lysate was clarified by centrifugation at 20,400 × g, followed by sonication for 30 s, and centrifugation at 20,400 × g for 15 min. The clarified lysate was loaded onto a 1 mL HiTrap-HP column (GE Healthcare), washed with 15 mL binding buffer and then 10 mL wash buffer [20 mM Tris- HCl (pH 8.0), 500 mM NaCl, 50 mM imidazole, 5% (vol/vol) glycerol, and 1 mM DDT]. Bound proteins were eluted in 1.5-mL fractions in four 5-mL step elutions with increasing concentrations of imidazole. Fractions containing GIY-ZFEs were dialyzed twice against 1 L dialysis buffer [20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 5% (vol/ vol) glycerol, and 1 mM DDT] before storage at −80 °C.

Cleavage Assays. Single time-point cleavage assays to determine the EC0.5max of TevN201-ryA were performed in buffer containing 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM MgCl2, 5% (vol/vol) glycerol, 1 mM DTT and 10 mM pTZHS1.33. Reactions were incubated for 3 min at 37 °C, stopped with 5 μL stop solution (100 mM EDTA, 40% (vol/vol) glycerol, and bromophenol blue), and electrophoresed on a 1% agarose gel before staining with ethidium bromide and analysis on an AlphaImager3400 (Alpha Innotech). The EC0.5max was determined by fitting the data to the equation

\[
E_c(\text{[endo]}) = \frac{f_{\text{max}}[\text{[endo]}]^H}{E_{0.5\text{max}} + [\text{[endo]}]^H},
\]

where \(f_{\text{[endo]}}\) is the fraction of substrate cleaved at concentration of TevN201-ryA [endo], \(f_{\text{max}}\) is the maximal fraction cleavage, with 1 being the highest value, and \(H\) is the Hill constant that was set to 1. The initial reaction velocity was determined using supercoiled plasmid substrate with varying concentrations of TevN201-ryA (0.7 nM to 47 nM) and buffer as above. Aliquots were removed at various times, stopped and analyzed as above. The data for product appearance was fitted to the equation

\[
P = A(1 - e^{-k_tt}) + k_f,
\]

where \(P\) is product (in nanomolars), \(A\) is the magnitude of the initial burst, \(k_t\) is the rate constant (s⁻¹) of the initial burst phase,
and \( k_2 \) is the steady state rate constant (s\(^{-1}\)). The two-site plasmid cleavage assays were conducted as above, using 10 nM pTZHS2.33 or pTZHS3.33 as substrates, and \(~90 \) nM purified TevN201-ryA. The \( k_{\text{obs}} \) rate constants were calculated from the decay of supercoiled substrate by fitting to the equation

\[
[C] = [C_0]\exp(-k_{\text{obs}}t),
\]

where \([C] \) (2) is the concentration (nM) of supercoiled plasmid at time \( t \), \([C_0] \) is the initial concentration of supercoiled substrate (nM), and \( k_1 \) is the first order rate constant (in s\(^{-1}\)).

At least three independent trials were conducted for each dataset.

**Cleavage Mapping.** Mapping of cleavage sites was performed as previously described (3). Briefly, primers were individually end-labeled with \( \gamma \)-\( ^{32} \)P ATP, and used in PCR reactions with pTox or pSP72 plasmids carrying TZ-ryA or BZ-ryA target sites to generate strand-specific substrates. The substrates were incubated with purified protein as above, and electrophoresed in 8% denaturing gels alongside sequencing ladders generated by cycle sequencing (US Biologicals) with the same end-labeled primers.

**Bioinformatics.** The distribution of the CNNNG cleavage motif was examined using a custom Perl script (Dataset S1) in a 35-bp window of ZFN sites predicted for *Dania rerio* chromosome 1 (Ensembl release 51) (4). Briefly, 40 bp flanking the downstream region of each predicted ZFN was extracted from the corresponding zebrafish chromosome 1 cDNA, and searched for position \( i = C \) and position \( i + 4 = G \), with the occurrences of each CNNNG reported at position \( i \) (the C of the motif). The number of CNNNG motifs for unique ZFN sites was fit to a binomial distribution, and plotted in R.

Fig. S2. TevN201-ryA purification for in vitro experiments. (A) Purification of TevN201-ryA. Shown is a representative SDS/PAGE gel; C, crude lysate; E, elution; FT, flow-through from metal-affinity column; IND, induced culture; M, marker with molecular weights in kilodaltons indicated on the left; UN, uninduced culture; W, wash. (B) Graphic representation of cleavage assays with 90 nM TevN201-ZFE and 10 nM two-site pTZ1.33 plasmid with target sites in the same orientation. Data are plotted as averages of three independent replicates with SDs; FLL, full-length linear; L1+L2, linear products; OC, open-circle (nicked); SC, supercoiled.

Fig. S3. PvuII site analysis. (A) Shown is the distribution of the 5′-CAGCTG-3′ motif in a 35-bp window flanking 8,829 predicted ZFN sites on zebrafish chromosome 1. The number of occurrences of the ‘C’ of the motif at each distance is indicated. (B) Unique ZFN sites were grouped according to the number of occurrences of the 5′-CAGCTG-3′ motif in the 35-bp window.
Fig. S4. Occurrence of the 5′-CNNNG-3′ motif upstream of I-OnuI E1 off-target sites. Shown is 37-nt of upstream sequence adjacent to the 22-nt I-OnuI E1 MAO-B target site, along with 19 predicted off-target sites (1). CNNNG motifs are highlighted in red, with only 3 of 19 predicted I-OnuI E1 off-target sites containing a CNNNG motif at a targetable distance by Tev-LHE fusions. Nucleotide differences of the off-target sites to the I-OnuI E1 site are indicated in magenta lowercase font.

Fig. S5. (Continued)
Fig. S5. Amino acid sequences of GIY-ZFEs and Tev-LHEs.

<table>
<thead>
<tr>
<th>Table S1. Tev-ZFE selection data</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Table S2. Tev-LHE selection data</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Table S3. Strains and plasmids used in this study</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Table S4. Oligonucleotides used in this study</th>
</tr>
</thead>
</table>
Other Supporting Information Files

Dataset S1 (DOC)