Expanding the Repertoire of Target Sites for Zinc Finger Nuclease-mediated Genome Modification

Kimberly A Wilson, Abbye E McEwen, Shonda M Pruett-Miller, Jiuli Zhang, Eric J Kildebeck and Matthew H Porteus

Recent studies have shown that zinc finger nucleases (ZFNs) are powerful reagents for making site-specific genomic modifications. The generic structure of these enzymes includes a ZF DNA-binding domain and nuclease domain (Fn) that are separated by an amino acid “linker” and cut genomic DNA at sites that have a generic structure (site1)-(spacer)-(site2) where the “spacer” separates the two binding sites. In this work, we compare the activity of ZFNs with different linkers on target sites with different spacer lengths. We found those nucleases with linkers’ lengths of 2 or 4 amino acid (aa) efficiently cut at target sites with 5 or 6 base pair (bp) spacers, and that those ZFNs with a 5-aa linker length efficiently cut target sites with 6 or 7 bp spacers. In addition, we demonstrate that the Oligomerized Pool ENgineering (OPEN) platform used for making three-fingered ZF proteins (ZFPs) can be modified to incorporate modular assembly fingers (including those recognizing ANNs, CNNs, and TNNs) and we were able to generate nucleases that efficiently cut cognate target sites. The ability to use module fingers in the OPEN platform at target sites of 5–7 bp spacer lengths increases the probability of finding a ZFN target site to 1 in 4 bp. These findings significantly expand the range of sites that can be potentially targeted by these custom-engineered proteins.

Using a three-fingered ZFP platform, we define the inter-domain linker to be the stretch of amino acids from the terminal histidine pair of the third ZF to the first amino acid of the nuclease domain (Supplementary Figure S1). A growing body of literature suggests that the inter-domain linker of a ZFN can be designed to accommodate a variety of spacer lengths between the two ZFN-binding sites. Bibikova et al. showed that in Xenopus oocytes, optimal ZFN cutting was most efficient using a ZFN with a 5-aa inter-domain linker at a target site with a 6 bp spacer length, but inefficient cutting resulted at target sites with spacer lengths of 5, 7, or 6 bp using ZFNs with a wide array of inter-domain linkers. Consistent with Bibikova et al., Porteus and Baltimore found that ZFNs were more effective at stimulating gene targeting at sites with 6 bp rather than 8 bp spacer lengths. However, Urnov et al. showed that target sites with a 5 bp spacer could be efficiently targeted by ZFNs with a 4-aa inter-domain linker. Alwin et al. also reported efficient targeting at a 6 bp spacer target site, but with a different 4-aa linker. Finally, Handel et al. surveyed 11 inter-domain linker ZFN variants on targets with spacers ranging from 4 to 18 bp in mammalian cells and found that different inter-domain linkers could preferentially cleave target sites with 5, 6, 7, or 16 bp spacers. These studies suggest that the inter-domain linker may be an alterable component for target sites with variations in spacer length.

In addition to the inter-domain linker, we define the conceptually related inter-finger linker to those amino acids that occupy the positions between the individual zinc finger α-helices immediately proceeding the histidine pair, but
Expanding the Repertoire of ZFN Target Sites
Wilson et al.

Figure 1 Schematic of zinc finger nuclease (ZFN) binding, green fluorescent protein (GFP) reporter constructs, and ZFN-mediated gene targeting. (a) Two ZFN target sites on opposing strands are separated by a short-intervening sequence labeled as “spacer”. (b) This diagram highlights the use of the TQGKD 5-aa linker that joins the three-fingered ZF DNA-binding domain to the Fn domain. When a pair of ZFNs binds to their cognate-binding sites, the nuclease domains can dimerize and cut DNA. The amino acids that mediate homodimerization between two wild-type (wtFn) nuclease domains by the formation of a salt bridge are depicted. (c) The nuclease domain in the obligate heterodimer (obhetFn) has been modified to prevent homodimerization and is schematized to show that the “KK” nuclease can only dimerize with the “EL” nuclease and vice versa. (d) Schematic of the target GFP gene construct mutated by several insertions (stop codon, I-SceI recognition site, and a frameshift) and the doubled GFP-ZFN2 half-site (5′-GACGACGGC-3′, black boxes) -GACGACGGC-3′. (e) The nuclease bearing a truncated version of the GFP gene. The expressed ZFNs will bind to the GFP-ZFN2 target site in pairs to generate a DSB that stimulates homologous recombination (HR) between the reporter gene and the repair template plasmid. Repair of the DSB by homologous recombination will produce a functional GFP reporter gene. CMV, cytomegalovirus; DSB, double-strand break.

Currently, four common and publicly available methodologies exist for creating three-fingered ZFNs. First, ZFNs can be made from naturally existing ZFPs to their native target sites. Second, ZFPs can be assembled in a modular fashion by connecting individual fingers of known specificity for the target triplet subsite. Individual module fingers to target sites that included 1 bp insertions between the two-fingered array subsites, and demonstrated that ZFPs with the modified inter-finger linker could bind such sites with high affinity. In addition, a 6-aa inter-finger linker between fingers two and three is used in the four-fingered ZFNs reported by Urnov et al. to join pairs of two-fingered arrays, and these produce high frequencies of gene targeting at the cognate target site. Therefore, it appears that the inter-finger linker is also a candidate for customization to accommodate insertions within a binding site. However, it remains to be seen whether these studies can be applied to three-fingered ZFP platforms.

Figure S1. Expanding the Repertoire of ZFN Target Sites

Supplementary Figure S1. Typically, this linker length is 5 aa, but in some polydactyl ZFP domains, inter-finger linkers may have to be lengthened to allow for conformation to the periodicity of the DNA helix or to allow for extra base pairs in a ZFP recognition site. Moore et al., for example, tested 5 and 7-aa inter-finger linkers to join the ZF pairs in a six-fingered ZFP for target sites that included 1 bp insertions between the two-fingered array subsites, and demonstrated that ZFPs with the modified inter-finger linker could bind such sites with high affinity. In addition, a 6-aa inter-finger linker between fingers two and three is used in the four-fingered ZFNs reported by Urnov et al. to join pairs of two-fingered arrays, and these produce high frequencies of gene targeting at the cognate target site. Therefore, it appears that the inter-finger linker is also a candidate for customization to accommodate insertions within a binding site. However, it remains to be seen whether these studies can be applied to three-fingered ZFP platforms.

Figure S1. Expanding the Repertoire of ZFN Target Sites

Supplementary Figure S1. Typically, this linker length is 5 aa, but in some polydactyl ZFP domains, inter-finger linkers may have to be lengthened to allow for conformation to the periodicity of the DNA helix or to allow for extra base pairs in a ZFP recognition site. Moore et al., for example, tested 5 and 7-aa inter-finger linkers to join the ZF pairs in a six-fingered ZFP for target sites that included 1 bp insertions between the two-fingered array subsites, and demonstrated that ZFPs with the modified inter-finger linker could bind such sites with high affinity. In addition, a 6-aa inter-finger linker between fingers two and three is used in the four-fingered ZFNs reported by Urnov et al. to join pairs of two-fingered arrays, and these produce high frequencies of gene targeting at the cognate target site. Therefore, it appears that the inter-finger linker is also a candidate for customization to accommodate insertions within a binding site. However, it remains to be seen whether these studies can be applied to three-fingered ZFP platforms.
an a shared Finger 2 sequence found in previously characterized ZFPs developed by OPEN or those found in natural domains. This recently described method of context-dependent assembly (CoDa) has rates of success similar to that of the OPEN protocols. While modular assembly, OPEN, and CoDa methodologies allow the versatility of creating a new ZFP for a target sequence, there remains room for improvement in the three-fingered ZFP platform to expand the range of sequence coverage.

In this study, we survey a number of modifications in ZFN architecture to target sites that carry variations in the canonical sequence of two 5'-GNNNGNN-3' half-sites separated by a 6 bp spacer (Figure 1a). To explore a range of spacer lengths from 3–7 bp, we tested the on-target and off-target cutting activities of ZFN variants with four types of inter-domain linkers (2–5 aa) in vitro and in mammalian cell-based assays. We found that target sites with 5, 6, or 7 bp spacers could be cut by ZFNs with differential efficiencies based on the inter-domain linker used. Also, we sought to determine whether a ZFN based on a three-fingered ZFP platform could continue to have efficient nuclease activity when the target half-site contains a 1 bp insertion between sub-sites 5'-GNNNGNNN-3' or 5'-GNGNNNGNN-3'. In lengthening the endogenous 5-aa inter-finger linker to 6 aa between fingers corresponding to the insertion in the target site, we found that despite being able to develop inter-finger ZFN variants that could efficiently generate DSBs, these variants were not sufficiently specific for the intended target site. Finally, to address ZFN target half-sites that include non-GNN triplets, we describe a successful hybrid methodology that combines modular assembly fingers with the OPEN platform to produce functional ZFNs. Ordinarily, the probability of finding a ZFN target composed purely of GNN triplets with a 6 bp spacer is 1 in 4,096 bp. However, based on the guidelines found in this report, the inclusion of module fingers in OPEN-based selections for target sites with 5–7 bp spacers, the probability dramatically increases to 1 in 4 bp. All together, our work not only further defines the robustness of the three-fingered ZFN platform, but also broadly expands the range of target sites that can be targeted by ZFNs.

Results

In vitro nuclease activity of the TGQKD inter-domain linker GFP-ZFN2 variant protein

Using an in vitro cutting assay, we tested the purified TGQKD 5-aa inter-domain linker variant (unmodified green fluorescent protein (GFP)-ZFN2) protein for its ability to cut DNA in vitro on a substrate. To test this inter-domain linker variant, we made a series of reporter constructs in which two GFP-ZFN2–bind- ing sites were arranged as inverted repeats separated by spacers of 3, 4, 5, 6, or 7 bp (Figure 1d). ZFN cleavage at the cognate target site results in a linear 3 kb fragment being cut into two fragments of ~1.8 and 1.2 kb. Off-target cutting leads to the disappearance of a 2.4 kb band or by the presence of other bands that are not ~1.8 or 1.2 kb. In our results, we found significant off-target cutting on all the substrates at high concentrations (4:1 ratio of protein to DNA) of ZFN protein. Since the products of this digestion had identical weights, the ZFN cutting appears specific, albeit off-target (Figure 2). The off-target cutting in vitro is consistent with the off-target DSBs ZFNs make in cells. At low concentrations of ZFN protein (0.25:1 ratio of protein to DNA), we found that the ZFNs did not cut any of the substrates efficiently in these conditions. Finally, at intermediate amounts (1:1 ratio of protein to DNA), the ZFN cut the substrates with 5, 6, or 7 bp spacers better than they cut the substrates with 3 or 4 bp spacer lengths (Figure 2, as shown by brighter specific bands in the 1:1 ratio lane for the 5, 6, and 7 bp spacer constructs). These in vitro results, however, did not accurately predict which spacer lengths the nuclease would cut efficiently when integrated into the mammalian genome (see below). For this reason, we did not pursue in vitro characterization of the other linker variants. These results suggest that ZFN in vitro activity profiles cannot necessarily predict activity in cells.

Experimental strategy for testing GFP-ZFN2 inter-domain linker variants

We have previously reported a ZFN (GFP-ZFN2) that was designed to recognize a target site 5'-GACGACGGC-3' within the GFP gene. This ZFN was previously shown to have high activity and low toxicity in prior work and binds to its target site with high affinity (~0.1 nmol/l, data not shown). We constructed a series of ZFNs from the GFP-ZFN2 with different inter-domain linkers: GS (2 aa), LRGS (4 aa), TGQKD (5 aa), and AAARA (5 aa) (Supplementary Figure S1). The LRGS, TGQKD, and AAARA linkers represent actual or modified linkers used in mammalian cells from the published literature. We made the GS inter-domain linker to explore the effect of a shorter linker. Finally, we also made modifications of the nuclease domain originally designed to prevent homodimerization and previously demonstrated in literature to reduce toxicity. We tested whether these nuclease modifications changed the activity of a ZFN on different spacer constructs.
Gene targeting by GFP-ZFN2 inter-domain linker variants on target sites with different spacer lengths

Since our in vitro studies demonstrated that an inter-domain linker variant ZFN had activity on a range of different target sites with different spacer lengths, we tested their activity on a genome-integrated reporter in mammalian cells. We created a set of stable cell lines that contained an integrated GFP reporter gene with the ZFN target sites separated by 3–7 bp, using GFP reporter constructs identical to those used as cutting substrates in the in vitro assays (Figure 1b,c). Thus, for each spacer length, we generated a different cell line (five total). The reporter in each of these cell lines contained a recognition site for I-SceI as an internal standard. By using I-SceI as an internal standard, we could compare the relative activities of the ZFN variants across different cell lines and control for variables such as local positional effects and chromatin status. Prior work has found that the activity of ZFNs can vary depending on the amount of ZFN plasmid transfected.23,30,32 We performed these experiments, therefore, at low (20 ng) and high (100 ng) amounts of transfected ZFN-expressing plasmids.

We found no evidence of targeting in cells when the spacer was 3 or 4 bp at both low and high amounts of transfected ZFN with all the different inter-domain linker variants (data not shown). Thus, we found that even though there was measurable cutting of these substrates in vitro (Figure 2), this activity could not predict the results of a cell-based reporter system (Figure 3). In the reporter with the 5 bp spacer, the 2 and 4-aa variants gave the best gene-targeting activity (Figure 3a–c). Figure 3b best demonstrates the improved...
targeting with these shorter linkers. We found that all inter-domain linker variants stimulated efficient gene targeting using the 6 bp spacer (Figure 3d,e). Finally, with the 7 bp target, we found that only the TGQKD 5-aa linker gave efficient targeting (Figure 3g–i). The cell-based assay results with the TGQKD linker in which efficient targeting was achieved best on spacers with 6 or 7 bp that is in contrast with the in vitro results where the TGQKD variant cut the constructs with spacer lengths of 5, 6, or 7 bp equally (Figure 2).

Gene targeting using obligate heterodimer GFP-ZFN2 inter-domain linker variants on target sites with different spacer lengths

Prior studies have shown that ZFN toxicity can be decreased by modifying the nuclease domain to prevent homodimerization and are called "obligate heterodimer" variants (obhetFn).28,32 In this study, we tested the modifications described in Miller et al.26 In these variants, one ZFN contains the following changes E490K, I538K (called "KK") while the other contains the changes Q486E, I499L (called "EL"), where the numbering reflects the amino acid position in the wild-type FokI nuclease domain (wtFn). We incorporated these changes into our inter-domain linker variants and tested them for targeting activity using the spacer variant reporter lines described above (Figure 1c–e). We found that just as with the wtFn, the obhetFn variants had no activity on the 3 or 4 bp spacer (data not shown), but also

| Table 1 Qualitative summary of GFP-ZFN2 inter-domain linker variants and their relative activity on target sites with various spacer lengths, expression levels, and toxicity |
|---------------------------------------------------------------|
| Relative on-target cutting at spacer lengths*                  |
| **Fn** | Linker | 3 bp | 4 bp | 5 bp | 6 bp | 7 bp | Expression | Toxicity |
|--------|--------|------|------|------|------|------|------------|----------|
| GS     | —      | ++++ | ++++ | —    | +++  | Lower |            |          |
| LRGS   | —      | +++  | ++++ | —    | +++  | Higher|            |          |
| TGQKD* | —      | —    | +++  | ++++ | +++  | Medium|            |          |
| wt     | AAARA  | —    | ++   | —    | —    | Lower |            |          |
| GS     | —      | +    | —    | —    | ++   | Lower |            |          |
| LRGS   | —      | —    | ++   | —    | —    | Lower |            |          |
| KK/EL  | TGQKD  | —    | —    | +++  | —    | ++   | Lower     |          |

GFP, green fluorescent protein; wt, wild-type; ZFN, zinc finger nuclease. †On-target ZFN-cutting activities are scaled relative to I-SceI activity set to +++; ‡Expression levels are scaled relative to GFP-ZFN2 (TGQKD-wtFn) set to +++.

Figure 4 Toxicity of inter-domain linker variant zinc finger nucleases (ZFNs). The different ZFN variants were analyzed for toxicity using two different previously described assays: a cell survival assay and a double-strand break (DSB) foci formation assay.23 In the cell survival assay, a lower percent survival relative to I-SceI is a marker of greater toxicity. In the akDSB foci formation assay, a lower percent survival relative to I-SceI is a marker of greater toxicity. In the akDSB foci formation assay, an increased number of cells with 53BP1 foci are a marker of greater toxicity. (a) Cell survival after transfection of 20 ng of each ZFN with a wild-type nuclease domain (wtFn). (b) Cell survival after transfection of 100 ng of each ZFN with a wtFn. (c) Cell survival after transfection of 100 ng of each ZFN with a modified nuclease domain to prevent homodimerization. (d) akDSB foci formation assay in which kDsBs are identified by p53BP1 foci after immunostaining. The number of foci was counted in 100 transfected cells for each condition. The cells were then grouped into three bins (0–1 foci, 2–5 foci, 6 or more foci). In prior work, we have found that a large number of cells with six or more foci correlate best with toxicity.23 As negative controls, cells were transfected with either I-SceI alone or an empty expression vector (“blank vector”). As a positive control, cells were transfected with caspase-activated DNAase (“CAD”). The 4-aa LRGS ZFN, 5-aa TGQKD ZFN, and the 5-aa AAARA ZFN all showed increases in DSB formation relative to the negative controls, whereas the 2-aa GS ZFN did not. Only in the 4-aa LRGS ZFN, however, was the increase statistically significantly different than the negative controls ($\chi^2$ analysis, $^{*}P < 0.05$).
had no activity on the 7 bp spacer (Figure 3i). With the 5 bp spacer, the 2 and 4-aa inter-domain linker obhetFn pairs gave significantly less activity than the wtFn counterparts (compare Figure 3b,c). With the TGQKD inter-domain linker, the obhetFn pair showed only 20% of the activity given by the TGQKD variant with the wtFn (Figure 3c).

Overall, the TGQKD inter-domain linker variant showed the broadest activity (spacer lengths of 5, 6, and 7 bp), which is in contrast to the AAARA inter-domain linker variant which only efficiently targets 5 and 6 bp spacers (Figure 3b,e,h). With the 6 bp spacer, the obhetFn pair gave equal activity (Figure 3c) to the wtFn with the TGQKD linker but less activity with the 2 and 4-aa inter-domain linker (compare Figure 3e,f).

Cell-based assays to measure toxicity of GFP-ZFN2 inter-domain linker variants

To determine whether the inter-domain linker GFP-ZFN2 variants had different off-target effects in cells leading to extraneous DSBs that may result in cell death, we used two assays: (i) a cell-based survival assay and (ii) a DSB foci formation assay. In the cell-based survival assay, we demonstrated that at low amounts of ZFN transfected, the GFP-ZFN2 inter-domain linker variants did not show appreciable toxicity (Figure 4a). At the higher amount of ZFN transfected, we found that the TGQKD variant showed some increased toxicity and the LRGS variant showed the most (compare Figure 4a,b). Results from the foci formation assay show similar results; only the LRGS inter-domain linker variant with the wtFn had significantly ($P < 0.05$) more cells with six or more foci than the non-toxic controls (Figure 4d).

Since toxicity increases as the amount of ZFN transfected increases, we determined the relative expression levels of the inter-domain linker ZFN variants (Supplementary Figure S1a–c). We found that the LRGS-wtFn variant had higher levels of expression relative to all other variants and correlates to the higher levels of toxicity seen in Figure 4a,b,d. These expression studies suggest that the toxicity of the LRGS inter-domain linker variant with the wtFn may be more the result of its expression level rather than any intrinsic property of the ZFN architecture itself (Figure 4a–c and Supplementary Figure S2a,c). These experiments also highlight that small amino acid changes in the ZFN, as few as two, can have a dramatic impact on expression in mammalian cells. In contrast, we found that the LRGS-obhetFn variants did not express as well as the ZFNs with the wtFn (Supplementary Figure S2a,b). This lower expression may partially account for the lower frequencies of gene targeting (Figure 3c,f,i).

Table 1 summarizes our findings for inter-domain linker ZFN variant activity.

| Target site | Full site sequence | ZFN Finger 3 | Finger 2 | Finger 1 |
|-------------|--------------------|--------------|----------|----------|
| GFP1/2      | 5′-ACC ATC TTC-gaattc-GAC GAC GGC-3′ | pGFP1 QHPNLTR VAHNLTR TROKLGV |  |  |
| Left:      | GNN GNN GNN |  |  |  |
| Right:     | GNN GNN GNN |  |  |  |
| F2-ACG     | 5′-TAC CGT GTC-caagac-GGA GAC GAG-3′ | pGFP2 EGGNLMR DRSNLTRa APSKLD |  |  |
| Left:      | GNN ACG GNN |  |  |  |
| Right:     | GNN GNN GNN |  |  |  |
| F1-CAG     | 5′-CTG CTC AAC-atcgcc-GTG GCT GAC-3′ | pJZ90A DOGNLIR RTDTRLRA AAVLVR |  |  |
| Left:      | GNN GNN GNN |  |  |  |
| Right:     | GNN GNN GNN |  |  |  |
| F2-AAC     | 5′-TCC CAC AGC-tcctg-GGC AAC GTG-3′ | pJZ108 TNNVLNT DRSNLTR KHNLTR |  |  |
| Left:      | GNN ACC GNN |  |  |  |
| Right:     | GNN GNN GNN |  |  |  |
| F2-AAG, F2-TGG | 5′-CTC CTT GCC-tagtc-GGA TGG GCA-3′ | pEK3-L4 KNASLGH RKDNLKNa RMSNLDR |  |  |
| Left:      | GNN AAG GNN |  |  |  |
| Right:     | GNN TGG GNN |  |  |  |

ZFN, zinc finger nuclease; ZFP, zinc finger protein.

*Modular zinc fingers, all amino acid identities are given in the order of −1 to +6 position within the α-helix of the individual zinc finger.
Expanding the Repertoire of ZFN Target Sites

Wilson et al.

between target subsites (5′-GACNGACGCGC-3′ and 5′-GACGACNGGC-3′) in a similar strategy described in Moore et al.15 We chose to use single guanine, adenine, or thymine insertions to detect any potential sequence preferences for these unbound regions of the target site. In the GFP-ZFN2 ZFP, the pre-existing inter-finger linker TGEKP was modified to TGSEKP or TGSEQD in the F1-F2 or F2-F3 positions without additional changes to the original recognition helices of the individual fingers (Supplementary Figures S1 and S3). We then explored this concept further by using the OPEN method to incorporate the 6-aa inter-finger linker variations and make new three-fingered libraries to the normal 9 bp GFP2 target half-site sequence and four 10 bp versions. The resulting ZFPs were converted to ZFNs and further details can be found in Supplementary Table S1.

To test the nuclease activity of these new GFP-ZFN2 inter-finger linker variants, we constructed a GFP-based extra-chromosomal single-strand annealing (SSA) assay where the normal and mutated GFP2 target half-sites were paired with the GFP1 half-site and inserted between repeated sections of the GFP gene (five total). These reporter plasmids were cotransfected with plasmids expressing GFP-ZFN1 and one of the GFP-ZFN2 inter-finger linker variants in a combinatorial fashion (Supplementary Table S1). Briefly, our results suggest that within the three-fingered ZFP platform, extension of the 5-aa inter-finger linker by the addition of a serine to a 6-aa linker to accommodate a 10 bp target half-site was associated with measurable off-target nuclease activity and a sequence preference for thymine insertions. Our efforts to adapt the OPEN method for this purpose through the incorporation of TGSEKP linkers into the three-fingered libraries were met with high failure rates in producing functional ZFPs. Based on this data, we were unable to find an architectural modification that expand the range of ZFN target half-sites from 9 to 10 bp and do not believe that this strategy can help broaden the repertoire of ZFN target sites from 9 to 10 bp without decreases in specificity. We are also unable to conclude that specifically matching 6-aa inter-finger linkers to the positions the target site insertions in this platform can be systematically accomplished for three reasons: (i) the high failure rate of inter-finger linker ZFN variants made by either modifying an existing three-fingered ZFPs or those generated by OPEN protocols; (ii) discrepancies between sites with high nuclease activity in the SSA assays when compared with the site selections of ZFP origin; and (iii) significant levels of

Table 3 Target site criteria and estimated probabilities

| Target site conditions | Probability |
|-----------------------|-------------|
| “GNN” triplets with a 6 bp spacer | 1 in 4,096 bp |
| “GNN” triplets with a 5, 6, or 7 bp spacer | 1 in 1,382 bp |
| OPEN B2H triplets with a 6 bp spacer | 1 in 109 bp |
| OPEN B2H triplets with a 5, 6, or 7 bp spacer | 1 in 42 bp |
| OPEN B2H triplets with one module for F1 or F2 on ONE side with 5, 6, or 7 bp spacer | 1 in 7 bp |
| OPEN B2H triplets with one module for F1 or F2 on BOTH sides with 5, 6, or 7 bp spacer | 1 in 4 bp |

B2H, bacterial-2-hybrid; ZFN, zinc finger nuclease. *The estimated probability of finding ZFN target sites were calculated using Monte Carlo simulations assuming 50% GC sequence content.

GFP-ZFN2 inter-finger linker variants

In order to accommodate helical periodicity or extra nucleotides between target subsites, modifications to lengthen inter-finger linkers can be found in some previously published four- and six-fingered ZFPs.6,14,15 Therefore, we hypothesized that if the three-fingered ZFP platform could be adapted to bind to longer sequences, then the 9 bp target half-site repertoire may be increased by the inclusion of 10 bp sites. To study this possibility, we mutated the GFP2 target half-site (5′-GACGACGCGC-3′) with the insertion of single nucleotides

Figure 6 Nuclease activity of hybrid zinc finger nucleases (ZFNs) as measured by extrachromosomal repair of a green fluorescent protein (GFP)-based reporter plasmid. (a) The target sites listed in Table 2 are inserted between two repeated regions of the GFP gene to create a GFP-based reporter plasmid. When cotransfected, the expressed ZFNs cut the target site and the resulting double-strand break (DSB) is repaired by single-strand annealing repair mechanisms to produce a functional GFP gene. (b) Each reporter plasmid (20 ng) was cotransfected into HEK293 cells with 100 ng of each ZFN-expressing plasmid in appropriate pairs. Extrachromosomal repair of the resulting DSB produces a fluorescent protein (GFP)-based reporter plasmid. (Supplementary Table S1). Briefly, our results suggest that within the three-fingered ZFP platform, extension of the 5-aa inter-finger linker by the addition of a serine to a 6-aa linker to accommodate a 10 bp target half-site was associated with measurable off-target nuclease activity and a sequence preference for thymine insertions. Our efforts to adapt the OPEN method for this purpose through the incorporation of TGSEKP linkers into the three-fingered libraries were met with high failure rates in producing functional ZFPs. Based on this data, we were unable to find an architectural modification that expand the range of ZFN target half-sites from 9 to 10 bp and do not believe that this strategy can help broaden the repertoire of ZFN target sites from 9 to 10 bp without decreases in specificity. We are also unable to conclude that specifically matching 6-aa inter-finger linkers to the positions the target site insertions in this platform can be systematically accomplished for three reasons: (i) the high failure rate of inter-finger linker ZFN variants made by either modifying an existing three-fingered ZFPs or those generated by OPEN protocols; (ii) discrepancies between sites with high nuclease activity in the SSA assays when compared with the site selections of ZFP origin; and (iii) significant levels of
off-target cutting (Supplementary Table S1). It is possible that the 6-aa inter-finger linkers disrupt context-dependent binding between individual fingers, or that the TGSEKP and TGSQKD are not optimal linkers for accommodating single-nucleotide insertions in the three-fingered ZFP platform.

A hybrid modular assembly and OPEN strategy to engineer active ZFNs

In our prior work, we integrated a single modular finger into the OPEN platform to generate an active ZFN.23 For this case, the target site followed the canonical 5′-GNN1NGNN-3′ sequence. Encouraged, we hypothesized that we could expand this hybrid form of ZFP engineering (the incorporation of module fingers in OPEN-based selections) to make active ZFNs to target sites that contain non-GNN triplets. For this study, we identified four different full ZFN target sites where there is at least one ANN, TNN, or CNN (Table 2). We then adapted the module fingers that recognize the non-GNN subsites by including them in OPEN strategy at the PCR-based step for randomly recombining the single finger archives to make the three-fingered cassettes used in full site selection (schematized in Figure 5). In doing so, the module finger is included in the screening for context-dependent binding, which is a crucial feature of the OPEN and CoDA platforms. The resulting ZFNs from this hybrid method of ZFP development are also listed in Table 2.

Repair of an extrachromosomal GFP reporter by SSA using ZFNs developed from hybrid methodologies

We then tested the ZFNs made by the hybrid method for nuclease activity with SSA strategies (Figure 6a). Briefly, the four target sites were inserted into repeated sections of the GFP gene to create reporter plasmids that were cotransfected with the corresponding pair of ZFNs listed in Table 2. The activity of these ZFNs were normalized relative to the GFP1/2-positive control and reported as a percentage of that activity (Figure 6b). We found that when the JZ90A/JZ100 ZFN pair (which targets an ACG triplet with an F2 module for one ZFN) is approximately one-third as active as the GFP1/2 control pair. In comparison, the JZ154/JZ144 ZFN pair, which also contains an ANN module for F2, exhibits nuclease activity for the F2-AAC site that approaches the GFP1/2 standard (~85%). We also installed a module finger recognizing the F1-CAG for the JZ90C2 ZFN. When used as a pair with the JZ108 ZFN, the mean nuclease activity for the F1-CAG site exceeds the positive control (~125%), although this is not statistically significant (P = 0.19, Student's one-tailed t-test). In addition to studying ZFN pairs with only one modular finger, we used the F2-AAG, F2-TGG site as an opportunity to investigate the feasibility of using two modular fingers. Both the Ek3-L4 and Ek3-R3 ZFNs have F2 modules and as a pair show half as much nuclease activity for the F2-AAG, F2-TGG site relative to the activity of the GFP1/2 ZFNs. In prior work, we have shown that we could substitute a single finger at F2 that recognizes a GNN triplet into the OPEN selection strategy, and here we expand that successful protocol by demonstrating that we can introduce modules that recognize ANN, TNN, and CNN triplets into fingers F1 or F2 in combination with the OPEN methodology.23 In light of these results, we sought to estimate the expansion of the ZFN repertoire. Assuming 50% GC content in randomized sequences, our Monte Carlo-based simulations predict that the probability of finding a ZFN target site is now 1 in 4 bp (see Discussion, Table 3).

Discussion

Expanding the range of sites that can be targeted by ZFNs is an important part of the development of ZFN-mediated genome modification. Therefore, we investigated whether modification of different architectural elements of the ZFN could expand the design density. In this study, we explored variations in the target site including variations in the spacer lengths between ZFN-binding sites, single base pair insertions between triplet recognition sites in the target sites, and the inclusion of non-GNN triplets as part of the target site. We also explored modifications in the ZFN architecture including changes in the ZFN inter-domain linkers, inter-finger linkers, and protocols for ZFP creation.

Previous work has demonstrated that full ZFN target sites with 5, 6, 7, and 16 bp spacers could be targeted efficiently at a chromosomal locus.1,6,8,12,13,29 To target sites with these different spacer lengths, however, requires ZFNs with slightly different architectures. Our work not only validates the efficacy of targeting such sites, but also provides additional inter-domain linker solutions to differences in target site spacer length (Table 1). More specifically, we have also found that sites with 3 or 4 bp spacers could not be targeted efficiently. In our work, target sites with 5 bp spacer lengths showed the most efficient ZFN-mediated gene targeting when the linker between the ZF DNA-binding domain and the nuclease domain is 2 or 4 aa (Figure 3g,h). In contrast, our work demonstrates that ZFN variants with a 5-aa TGQKD inter-domain linker can efficiently target sites with a 7 bp spacer.

The evaluation of new ZFNs by an in vitro cutting assay could be a convenient way of assessing on-target and off-target effects (Figure 2). Unfortunately, we find that the ability of a ZFN to cut a target site in vitro does not always predict its ability to cut its target when embedded in the genome (Figure 3). Thus in evaluating a ZFN, in vitro assays alone are not sufficient.

Minimizing ZFN toxicity is also an important aspect of ZFN development.12,23,26,30 In this work, we have confirmed that engineering modifications in the nuclease domain to produce obligate heterodimerization can reduce toxicity (Figure 4c,d). The data also shows that these modifications, however, can result in reduced nuclease activity. How these modifications reduce toxicity requires further investigation to elucidate the full mechanism. The elimination of homodimer formation should only result in a twofold reduction in toxicity (preventing homodimer (ZFN1 with ZFN1 or ZFN2 with ZFN2), but published studies have demonstrated a significantly greater decrease.23,28,32 One hypothesis is that these modifications reduce toxicity not just by eliminating homodimerization but also by decreasing the affinity of the nuclease domains for each other, thereby requiring a more stable complex between the ZF DNA-binding domain and its target binding site to form before cutting can occur. In addition, we did not find that the
length or content of the linker affected toxicity. Instead, we found that the length and content of the linker and modifications of the nuclease domain could affect expression levels, and that the change in expression correlated best with toxicity (Figure 4 and Supplementary Figure S1a–c). This expression data also suggests the possibility of improving toxicity if ZFN expression could be controlled.\textsuperscript{34}

In addition to variations in target site spacer lengths, we investigated the possibility of developing ZFNs to 9 bp target half-sites containing 1 bp insertions between subsites (10 bp total). We hypothesized that lengthening the inter-finger linker to from 5 aa to 6 aa at the corresponding position of the insertion may allow a ZFN to conform to the new spatial requirements in the DNA helix for efficient nuclease activity (Supplementary Figure S3). Unfortunately, our efforts resulted in very few functional ZFNs. Any inter-finger linker variant ZFN with efficient nuclease activity demonstrated a preference for thymine insertions and significant amounts of off-target cutting (Supplementary Table S1). One possibility is that the TGSEKP or TGSKQD inter-finger linkers may not provide an optimal solution to this problem and whether solutions exist in modifying either the length or amino acid content of this linker remains a possibility. It is more likely, however, that active three-fingered ZFPs are highly dependent on each inter-finger interaction and may not tolerate disruption of these interactions by the extension of the inter-finger linker in the way that four-fingered or six-fingered ZFPs can.\textsuperscript{6,14,15}

There are three common and publicly available methods for engineering novel DNA-binding domains for ZFNs: modular assembly, OPEN, and CoDA. While individual module fingers that target most of the possible 64 triplet-binding sites have been published, there is also a report of a low success rate in using modular assembly in making active ZFNs, particularly for target sites that contain a non-GNN subsite.\textsuperscript{21,22} In contrast, the OPEN and CoDA methods have higher rates of success, but the target subsite coverage is significantly lower than is possible with modular assembly.\textsuperscript{24,25,27} In this work, we show that we can use a hybrid modular assembly/OPEN method (Figure 5) to generate active ZFNs that target sites which contain non-GNN subsites, including sites that contain ANN, TNN, and CNN triplets. We tested this strategy on four different targets that contained five different non-GNN subsites and were successful at all four targets (Table 2 and Figure 6b). These results, therefore, provide a strong rationale to using this hybrid strategy to target sites that are not amenable to targeting by the OPEN or CoDA method due to limited sequence coverage.

In summary, this work provides a guide to identify possible full ZFN target sites that vary from the canonical structure of two 5’-GNNNGNGNN-3’ target half-sites separated by a 6 bp spacer and methods to adapt ZFN architecture to cut those sites. The probability of finding such canonical sites is 1 in 4,096 bp and may not be in sufficient proximity to a desired locus for achieving high frequencies of gene modification.\textsuperscript{34}

However, when target site criteria is expanded to include 5, 6, or 7 bp spacer and the possibility of using module fingers and OPEN protocols together, the theoretical probability of finding a candidate target site can increase by three orders of magnitude to 1 in 4 bp (Table 3). Thus, these results provide guidelines that should be immediately useful to researchers attempting to develop ZFNs to perform efficient, site-specific genome modifications to a much wider range of target sequences.

Materials and methods

**GFP-ZFN2 inter-domain linker variants with the wtFn.** The three-fingered ZFP of the GFP-ZFN2 was developed through the B2H protocols of the OPEN methodology described previously.\textsuperscript{23} The DNA-binding domain recognizes the 9 bp target half-site 5’-GACGACGACG-3’ and the recognition helices for the three fingers are as follows: Finger 1: APSKLD; Finger 2: DRSNLTR; Finger 3: EGGNLMR. Using standard molecular biology procedures, three variants of the original nuclease were made: one with a 2-aa linker (GS), a 4-aa linker (LRGS), and another 5-aa linker (AAARA).\textsuperscript{6,12} All ZFN variants were cloned using the previously characterized GFP-ZFN2-B2H vector plasmid as a template, which already has the TGQKD inter-domain linker and wtFn and referred to as GFP-ZFN2 in this study. All of the ZFNs in this paper were cloned into expression vectors with a cytomegalovirus promoter.

**GFP-ZFN2 inter-domain linker variants with the obhetFn domain.** The obhetFn domains were made by PCR mutagenesis using the GFP-ZFN2-B2H vector as a template. The “KK” variant contains the E490K and I538K mutations and the “EL” variant contains the Q486E and I499L mutations where the numbering is with respect to the wt FokI enzyme.\textsuperscript{28} Six ZFNs with obhetFns (three KK/EL pairs) were made for the GS, LRGS, and TGQKD inter-domain linkers.

**Generation of reporter cell lines with target sites of different spacer lengths.** The ZFN linker variants were tested for targeting at sites with different spacer lengths using the previously reported GFP gene-targeting assay.\textsuperscript{1,2} In this reporter, an inverted repeat of the GFP2-binding site was inserted into the middle of a mutated GFP gene and adjacent to the recognition site for I-SceI using standard molecular biology techniques (Figure 1d,e). Separate reporter plasmids in which the GFP-ZFN2–binding sites were separated by either 3, 4, 5, 6, or 7 bp were made, and each reporter was used to generate a monoclonal HEK293 cell lines with a single copy of the reporter as previously described.\textsuperscript{1,8}

**In vitro cutting assay.** The in vitro cutting assay was performed using a protocol to be described elsewhere (A.E.M. and M.H.P., unpublished data). Briefly, His-tagged GFP-ZFN2 (TGQKD inter-domain linker and wtFn) was purified using a Ni\textsuperscript{2+} metal affinity column. We then linearized the five different target construct plasmids with the 3–7 bp spacer lengths using a Spel digest and combined purified ZFN protein at a concentration of 0.3 nmol/l into a gel. Cutting of the target constructs was evaluated by agarose gel electrophoresis.

**GFP-based gene-targeting assays.** Gene-targeting assays were done as described.\textsuperscript{1} ZFNs were transfected at 20 and 100 ng with 300 ng of repair template (donor) plasmid into HEK293 cells. All GFP-ZFN2 linker variants with KK or EL nuclease domains were transfected at 100 ng each with 300 ng of repair template plasmid. As an internal control for each experiment, separate wells were transfected with the repair...
template and GFP-ZFN2 (TGQKD inter-domain linker and wtFn) at 200 ng. Transfection efficiency was determined 48 hours post-transfection and gene-targeting rates were measured at 72 hours post-transfection. Repair of the mutated GFP gene to restore function was measured by flow cytometry using a FACS Calibur (Becton-Dickerson, San Jose, CA). Each of these experiments was performed between 3 and 11 times. All gene-targeting events were normalized to the transfection efficiency and reported as a relative percentage of I-SceI activity (set to 100%). By normalizing to the internal I-SceI standard, we could compare the efficiency of ZFN variants across different cell lines. Significance was measured using Student’s one-tailed t-test (P < 0.05) and error bars indicate (SEM).

**Flow cytometry assay for cell survival.** Cell survival assays were performed as previously described. Briefly, HEK293 cells were transfected with 300 ng of a GFP expression plasmid along with 20 or 100 ng of GFP-ZFN2 variants and analyzed by flow cytometry for GFP expression at days 2 and 6 after transfection. Percent survival is calculated as ratios of percent GFP+ populations at day 2 and day 6: ((ZFN day 6/ ZFN day 2)/(I-SceI day 6/I-SceI day 2)) × 100%.

**Foci formation assay and immunofluorescence staining.** Primary cultures of human foreskin fibroblasts were maintained and transfected with nucleofection techniques using program U-23 (Lonza, Basel, Switzerland). Briefly, 1 million fibroblasts were co-nucleofected using 2 μg of GFP expression plasmids and 2 μg of each ZFN variant-expressing plasmid. DSBs were visualized 48 hours post-nucleofection though incubation with rabbit α-p53BP1 primary antibody (Cell Signaling, Danvers, MA) and goat α-rabbit Rhodamine Red-X secondary antibody (Invitrogen, Carlsbad, CA). Cells were mounted onto slides using Vectashield with 4′,6-diamidino-2-phenylindole (DAPI) (Vector Laboratories, Burlingame, CA). Foci within the GFP+ fibroblasts were manually counted in a blinded manner and binned according to the number of foci/nucleus. Bins were set at 0–1 focus as background or minimal toxicity, 2–5 foci as moderate toxicity, and 6+ foci as severe toxicity. Differences between ZFN variants were analyzed statistically by a χ² analysis with a significance threshold of P < 0.05.

**Western blot analysis.** Expression of ZFNs was determined by western blotting using an anti-FLAG antibody (Sigma, St Louis, MO) to the FLAG tag at the N-terminus of each ZFN using standard procedures. Briefly, each ZFN was transfected into HEK293 cells and cell lysates were made 48 hours post-transfection. Densitometric analysis was performed using Image J software.

**Generation of GFP2 inter-finger linker ZFNs variants using the OPEN method.** First, using the GFP-ZFN2-B2H vector plasmid as a template, the canonical 5-aa TGEKP inter-finger linkers at either the Finger1-Finger2 (F1-F2) or Finger2-Finger3 (F2-F3) junctions were mutated to TGSEKP or TGSKQKD, but the original α-helices and TGQKD inter-domain linker to create four new ZFN variants were preserved using standard molecular biology techniques. New ZFPs were generated with B2H selections in the OPEN methodology using the same ZF pools used to generate the original ZFP of the GFP-ZFN2. The mutant TGSEKP inter-finger linker was incorporated into the three-fingered cassettes to generate two new libraries, F1-TGSEKP-F2 and F2-TGSEKP-F3, for stage two B2H selections. These libraries were then interrogated in five bacterial reporter strains bearing either the normal GFP2 target site or one of four GFP2 target site variants with an 1 bp insertion between subsets: 5′-GACGGACGGG-3′, 5′-GACGTGACGCC-3′, 5′-GACGACTG GCC-3′, and 5′-GACGACAGGCC-3′. New ZFNs with TGQKD inter-domain linkers were cloned from the rescued ZFPs found in the surviving colonies.

**Hybrid method for creating ZFPs using modular fingers and the OPEN protocols.** We identified four new full ZFN target sites in which one or both of the half-sites have a non-GNN triplet. The names and sequences are as follows: F2-AAC: 5′-TCGCCACAGGG-3′; F1-CAG: 5′-CTGCTC AAC-ctgcc-GTGGCTGAC-3′; F2-AAC: 5′-TCCACACG-t cctg-GGCAACGTG-3′; and F2-AAG, F2-TGG: 5′-CTCCCTTG CC-tagct-GGATGGGCA-3′. Modular assembly fingers (Addgene; http://www.addgene.org) were used in the finger positions in the non-GNN target half-site. The modules were incorporated into the recombinant PCR reaction to generate three-fingered cassettes for making five hybrid libraries in addition to the three other conventional OPEN-based libraries at the stage two B2H selections. These three-fingered libraries were then interrogated in corresponding bacterial reporter strains bearing the appropriate target half-site. New ZFNs with TGQKD inter-domain linkers were cloned from the rescued DNA-binding domains found in the surviving colonies, with the exception of the ZFNs developed to target the F2-AAC site for which the GS inter-domain linker was used.

**Generating GFP-based SSA reporter plasmids.** Using standard molecular biology techniques, nine SSA reporter plasmids were made by inserting one full ZFN target site into repeated sequences within the GFP gene. These include the normal GFP1/2 ZFN target site (5′-ACCATCTTC-aaattc-GGCC-3′) as a positive control, four GFP2 insertion target half-site variants which were paired with the GFP1 half-site (5′-GAAGATGGT-3′) with a 6 bp spacer to create a new full target site, and the four ZFN sites identified to have non-GNN triplets listed above.

**SSA assays for ZFN variants.** Extrachromosomal SSA assays were performed in HEK293 cells by transfection similar to those described previously. Briefly, 20 ng of a SSA reporter plasmid (described above) with 100 ng of each of the plasmids expressing a single ZFN. Repair of the mutated GFP gene by the endogenous SSA repair mechanism to restore GFP function was measured by flow cytometry at 48 hours post-transfection. The GFP-ZFN2 inter-finger linker variants were cotransfected with the GFP-ZFN1 which recognizes the GFP1 half-site and the recognition helices for the three fingers are as follows: Finger 1: TROKVLGV; Finger 2: VAHNLTR; Finger 3: QHPNLTR (previously described as GFP1.4-B2H). The hybrid ZFP ZFNs were cotransfected as pairs corresponding to the chosen target site. As a positive control, GFP-ZFN2 and GFP-ZFN1 were transfected with an SSA reporter plasmid with the normal GFP1/2 ZFN target site. Activities of all ZFNs were normalized to this positive control and reported as a percentage of that activity. Each of these experiments was performed three to four times and error is reported as SEM.
Determining the probability of finding ZFN target sites. Based on the target site guidelines informed by this study, we calculated the probability of finding such sites using Monte Carlo simulations. For each sample in our Monte Carlo simulation of size 10^5, 32 bp of random sequence were generated and tested against the patterns described in Table 3. We assumed 50% GC content and as inputs, used the publicly available modular assembly fingers (Addgene) and B2H-based OPEN single finger pools. Estimated probabilities summarized here have been rounded, but are based on our results for the 99.7% confidence interval. By allowing 5, 6, or 7 bp spacer lengths in the target site and using hybridized methods to generate ZFPs, the probability of finding a full ZFN site in close proximity to a locus of interest increases dramatically relative to the probability of only considering canonical sites.

Supplementary material

Figure S1. Amino acid sequence of GFP-ZFN2 DNA-binding domain.

Figure S2. Expression analysis of GFP-ZFN2 linker variants.

Figure S3. Inter-finger linker strategy to expand target half-sites from 9 to 10 bp.

Table S1. Nuclease activities of GFP-ZFN2 inter-finger linker variants.

Acknowledgments. We thank Sandeep Burma (University of Texas Southwestern Medical Center at Dallas) for his assistance and advice in developing the foci assay and Toni Cathomen (Hannover Medical School) for discussions with regards to the AAARA inter-domain linker. This work was supported by National Institutes of Health (R01 HL079295), The March of Dimes, a career development award from the AAARA, and a fellowship from the Texas Southwestern Medical Center by the State of Texas. The authors declared no conflict of interest.}

Supplementary Information accompanies this paper on the Molecular Therapy–Nucleic Acids website (http://www.nature.com/mtna).

1. Porteus, MH and Baltimore, D (2003). Chimeric nucleases stimulate gene targeting in human cells. Science 300: 763.

2. Dursi, S, Mani, M, Kandavelou, K, Wu, J, Porteus, MH and Chandrasegaran, S (2005). Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. Nucleic Acids Res 33: 5978–5980.

3. Bibboka, M, Beumer, K, Trautman, JK and Carroll, D (2003). Enhancing gene targeting with designed zinc finger nucleases. Science 300: 764.

4. Santiago, Y, Chan, E, Liu, PQ, Orlando, S, Zhang, L, Umov, FD et al. (2008). Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. Proc Natl Acad Sci USA 105: 5809–5814.

5. Perez, EE, Wang, J, Miller, JC, Jouvenot, Y, Kim, KA, Liu, O et al. (2008). Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat Biotechnol 26: 806–816.

6. Umov, FD, Miller, JC, Lee, YL, Beausejour, CM, Rock, JM, Augustus, S et al. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature 435: 646–651.

7. Lombardo, A, Genovese, P, Beausejour, CM, Colleoni, S, Lee, YL, Kim, KA et al. (2007). Gene editing in human stem cells using zinc-finger nucleases and integrase-defective lentiviral vector delivery. Nat Biotechnol 25: 1298–1306.

8. Porteus, MH (2006). Mammalian gene targeting with designed zinc finger nucleases. Mol Ther 13: 438–446.

9. Moehle, EA, Moehle, EA, Rock, JM, Rock, JM, Lee, YL, Lee, YL et al. (2007). Targeted gene addition into a specified location in the human genome using designed zinc-finger nucleases. Proc Natl Acad Sci USA 104: 3055–3060.

10. Smith, J, Bibboka, M, Whitby, FD, Reddy, AR, Chandrasegaran, S and Carroll, D (2000). Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. Nucleic Acids Res 28: 3361–3369.