UC Davis

UC Davis Previously Published Works

Title

Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases

Permalink

https://escholarship.org/uc/item/52x577fp

Journal

Nature Biotechnology, 25(7)

ISSN

1087-0156

Authors

Szczepek, Michal
Brondani, Vincent
Büchel, Janine
et al.

Publication Date

2007-07-01

Peer reviewed

Structure-based redesign of the dimerization interface
reduces the toxicity of zinc finger nucleases

Michal Szczepek [1, #], Vincent Brondani [2, #], Janine Büchel [1, #],

Luis Serrano [3], David J. Segal [2, *], and Toni Cathomen [1, *]

[1] Charité Medical School, Institute of Virology – CBF, Berlin, Germany

[2] UC Davis Genome Center and Department of Pharmacology, Davis, CA, U.S.A.

[3] Centre de Regulacio Genomica, CRG-EMBL Systems Biology Unit, Barcelona, Spain

[#] these authors contributed equally

* corresponding authors:

Toni Cathomen, Ph.D.

Charité Medical School

Institute of Virology – CBF

Hindenburgdamm 27

12203 Berlin

Germany

Tel: +49 30 8445 3820

Fax: +49 30 8445 3840

Email: toni.cathomen@charite.de

David J. Segal, Ph.D.

UC Davis Genome Center

Dept of Medical Pharmacology

4513 GBSF

451 E. Health Sciences Dr.

Davis, CA 95616

U.S.A.

Tel: +1 530 754 9134

Fax: +1 530 754 9658

Email: djsegal@ucdavis.edu

ABSTRACT

Artificial endonucleases consisting of a *Fok*I cleavage domain tethered to engineered zinc finger DNA-binding proteins have proven useful for stimulating homologous recombination in a variety of cell types. Since the catalytic domain must dimerize to become active, two subunits of these zinc finger nucleases (ZFN) are typically assembled at the cleavage site. Presumably due to cleavage at off-target sites, the use of ZFNs is often associated with significant cytotoxicity. Here, we describe a structure-based approach to reduce ZFN-induced toxicity. Rational redesign of the *Fok*I dimer interface aiming at destabilizing dimerization in combination with preventing homodimerization of the ZFN subunits was based on protein modeling and energy calculations. Cell-based recombination assays confirmed that the modified ZFNs elicit significantly reduced cytotoxicity without compromising on performance. Our results present a critical step towards the therapeutic application of the ZFN technology.

INTRODUCTION

Targeted gene modification by homologous recombination (HR) is a rare event in mammalian cells [1]. The frequency can be dramatically improved by the introduction of a targeted DNA double-strand break (DSB) near the site of the desired recombination event. Several recent studies have reported gene-targeting frequencies of 1-18% HR events per mammalian cell when the targeted DSB was introduced by natural or artificial endonucleases [2-8]. The ability to generate precise modifications of the genome – knockouts, mutations or corrections – within 1 in 100 to 1 in 5 cells presents this methodology as a potentially powerful tool for genetic studies, biotechnology and gene therapy.

The so-called zinc finger nuclease (ZFN) used in some of the above studies can be engineered to cleave at a user-determined target site. They are composed of the catalytic domain from the type IIs restriction enzyme *Fok*I tethered to an engineered zinc finger DNA-binding protein [9]. A single zinc finger motif consists of 28 amino acids and includes an α-helix that contacts three bases in the major groove of DNA. Through modular assembly of such zinc finger modules, artificial DNA-binding domains can be created to bind to a wide variety of DNA sequences [10-14]. A DNA-binding domain consisting of three zinc finger motifs has hence the potential to bind 9 bp of DNA. Many engineered 3-finger proteins have been shown to have excellent binding specificity *in vitro*, while others were less specific [10-15].

Because the active ZFN is a dimer, two subunits are typically designed to recognize the target sequence in a tail-to-tail conformation [16, 17]. The C-terminal *Fok*I domains are positioned in close proximity on the same face of the DNA helix by separating the two zinc finger binding sites by a 6 bp spacer [18]. The required combination of two ZFN subunits consisting of 3 fingers each thus provides enough specificity in any known genome by theoretically recognizing an 18 bp site. A typical application requires that two distinctive ZFN subunits bind as a heterodimer at the desired cleavage site. However, symmetry at the *Fok*I dimerization interface also permits homodimers to form, thus enabling cleavage at the respective homodimer sites. Moreover, at high ZFN concentrations dimerization could happen in the absence of correct DNA binding, resulting in activation of the nuclease and cleavage at off-target sites.

In previous work, we described two ZFNs, GZF1-N and GZF3-N, which were able to form an active heterodimer and stimulate chromosomal site-directed recombination at a frequency of 1% in 293 cells [8]. However, we and others have observed considerable cytotoxicity associated with the use of some ZFNs [7, 8], most likely due to cleavage of genomic DNA at off-target sites. Such toxicity could severely limit the potential of this nascent

technology and is, at least in part, due to the lack of allosteric regulation of the ZFN activity, as known for the natural *Fok*I enzyme, which keeps the catalytic domain in an inactive conformation in the absence of cognate DNA-binding [17]. Since the ZFNs are deficient in this structural mode of regulating the enzymatic activity, we investigated the effects of various modifications to the *Fok*I dimerization interface as a potential approach to reduce ZFN-induced cytotoxicity.

Several computational approaches have shown that novel enzyme specificities can be generated through redesign of the dimer interface [19-21]. We employed 3D protein modeling based on the published crystal structures of *Fok*I [22, 23] followed by computational analysis of dimerization to identify amino acids at the interface that adjust dimer formation. The subsequent evaluation in three cell-based HR assays confirmed that we generated variants with significantly improved ZFN performance, most likely as a result of reduced cytotoxicity. In particular, modifications that destabilized dimer formation and prevented homodimerization proved successful.

RESULTS

We hypothesized that toxicity associated with ZFNs is at least partially due to low specificity of DNA cleavage. Consequently, due to the dimeric nature of the *Fok*I nuclease domain [16, 17], altering the dimerization behavior of the two ZFN subunits could improve cleavage specificity.

Protein modeling and *in silico* mutagenesis

In order to identify and characterize the critical residues involved in dimerization of ZFNs, 3D protein models based on the crystal structure of the native *Fok*I endonuclease [22] were established. Although the published crystal structures [22, 23] do not provide a clear-cut view of the active enzyme bound to DNA, they offer useful information for structure-aided design of point mutations in order to alter the dimerization behavior of ZFNs (Fig. 1a).

Structural analysis suggested that dimerization is mainly mediated by helices α4 (residues 479-490) and α5 (residues 528-539) of the C-terminally located cleavage domain [23]. A combination of structural and functional data revealed that D483 of α4 may contact R487 in α4 of the other subunit (Fig. 1b) through bidentate hydrogen bonds [17, 23]. Close inspection of our 3D model revealed that an additional contact might be taking place between Q486 and

E490 of the opposing monomer (Fig. 1c). Furthermore, isoleucines in the interhelical sequence (I499) and in α5 (I538) may stabilize the dimer through hydrophobic interactions (Fig. 1d). Rational redesign based on this model predicts that an asymmetric dimer interface can be created by swapping a critical pair of interacting residues, like D483R or R487D (Fig. 1e). In such a case, likewise charged residues D483/D487 (DD) or R483/R487 (RR), respectively, will be situated in the same subunit (Fig. 1h), making the formation of homodimers unlikely due to electrostatic repulsion. The same rational applies to the interacting residues Q486 and E490 (Fig. 1f): variants EE (Q486E) and QK (E490K) should favor the formation of heterodimers over homodimers. Altering the hydrophobic residues I499 or I538 to small amino acids (Fig. 1g) is expected to reduce the dimerization energy without changing the preference for homo- or heterodimer formation. Due to the low protein–protein interaction energy of the resulting variants AI (I499A) and IV (I538V), dimers are expected to form only after binding of the corresponding monomers to the target sequence.

These model-based predictions were supported by calculations based on the computer algorithm FoldX [24], which permits a quantitative estimation of the effect of point mutations on the stability of protein–protein interactions [25, 26]. As pointed out (Fig. 1i), heterodimerization of variants DD with RR and EE with QK is energetically preferred over formation of the respective homodimers. All combinations involving variants AI and IV revealed reduced dimerization activity. The same is true for double mutant D483A/R487A (AA), which has been previously described as dimerization-defective [17]. Variant KO (D450AG) harbors a mutation in the catalytic center of the nuclease domain [17], but dimerization is not affected (data not shown).

In vitro cleavage assays

The *in vitro* activity and specificity of GZF1-N and GZF3-N homo- and heterodimers was analyzed by incubating purified ZFN proteins with a DNA fragment containing a GZF1-N homodimer site, a GZF3-N homodimer site, or a combined GZF3-N/GZF1-N heterodimer target site, referred to herein as 1-1, 3-3 and 3-1 (Fig. 2a). As predicted by FoldX, the respective homodimer target sequences were not only cleaved by wild-type GZF1-N or GZF3-N but also to some extent by variants containing the QK and IV variation (Fig. 2b, DNAs 1-1 or 3-3). Unexpectedly, also variant AI was able to cleave at the homodimeric target sequence. All other variations in the dimer interface prevented formation of homodimers and thus cleavage at homodimeric targets. Importantly, for wild-type or all variant *Fok*I domains

the heterodimer target was cleaved only by the appropriate combination of GZF1-N and GZF3-N (DNA 3-1), suggesting a high specificity of DNA-binding.

ZFN dimerization variants stimulate SSA

The potential of GZF1-N and GZF3-N to stimulate single strand annealing (SSA)-based HR in 293T cells was analyzed using a plasmid-based SSA assay (Fig. 3a). In this assay, a luciferase reporter gene was split into two inactive fragments separated by stop codons and a ZFN target site. Each fragment contained an 870 bp region of homology. Introduction of a DSB at the ZFN target site should initiate SSA recombination between the homologous regions, producing an active luciferase gene. 293T cells were transfected with the indicated SSA reporter plasmid and different ratios of empty- and ZFN-expression plasmids. ZFN expression was controlled by a PGK promoter and the assay was analyzed 24 hours post-transfection. Increasing the concentration of GZF1-N expression plasmid resulted in increased SSA activation on the 1-1 reporter (Fig. 3b), while GZF3-N failed to similarly stimulate SSA on a 3-3 homodimer reporter. SSA activation by GZF1-N/GZF3-N heterodimers was only detectable at low concentrations of GZF3-N. The dimerization-defective mutants AA–AA did not induce SSA. Thus, despite a similar *in vitro* activity, GZF1-N and GZF3-N differ dramatically in their ability to stimulate SSA recombination in a cell-based assay.

The effects of modifications to the ZFN dimer interface were evaluated next. Homodimers of variants RR, DD, QK and EE failed to stimulate SSA on their corresponding homodimer reporters (Fig. 3b), suggesting that homodimerization is significantly reduced, while variants IV and IA showed moderate to low activity. The pair-wise analysis of the individual variants on a SSA reporter containing the 3-1 heterodimer target site indicated that combinations of variants DD–RR or EE–QK displayed SSA activation (Fig. 3c), while the homologous combinations did not display any activity. In contrast, all combinations including variants AI and IV revealed some, although a rather low activity. Lastly, we probed the potential toxicity of GZF3-N using an SSA reporter containing a 1-1 homodimer site. GZF1-N-mediated SSA was rapidly attenuated upon the introduction of wild-type GZF3-N (Fig. 3d). In contrast, when using the ZFNs that contained the IV mutation, the expected gradual decrease in SSA was observed, albeit at a lower level of activity.

In conclusion, the SSA results largely confirmed the predictions made by FoldX: destabilizing dimerization (AI/IV) attenuated activity of the ZFNs while an asymmetric dimer interface (EE/QK and DD/RR) reduced homodimerization of the ZFN subunits.

ZFN dimerization variants stimulate HR-based gene repair

The efficiency of the modified ZFNs to promote gene repair was at first evaluated in a plasmid-based HR assay by co-transfecting a target plasmid and a repair plasmid [8]. To ensure that no functional EGFP is expressed, the target plasmid (Fig. 4a) contains a LacZ gene followed by stop codons and a 5'-truncated EGFP gene (∂GFP) [27]. The binding sites for I-*Sce*I as well as GZF3-N and GZF1-N were placed between the two open reading frames. The repair plasmid harbors a non-transcribed 5'-truncated LacZ-EGFP fusion gene and is designed to rescue EGFP expression through HR by generating a LacZ-EGFP fusion protein. An expression cassette for the red fluorescent protein DsRed-Express (REx) is located further downstream to mark transfected cells. ZFN expression was controlled by a PGK promoter and the percentage of green 293T cells was determined 48 hours post-transfection by flow cytometry.

In the absence of a nuclease only a few cells turned green, representing non-stimulated HR. Combined expression of the original GZF1-N and GZF3-N (WT–WT) stimulated HR to a similar extent as I-*Sce*I, with some 12% of transfected cells revealing functional correction of the GFP gene (Fig. 4b). As seen before in the SSA assay, the combination of variants with an asymmetric dimer interface (DD–RR and EE–QK) worked at least as well, if not better, than the original ZFNs. On the other hand, combinations harboring the same mutation on both subunits did not significantly activate HR. The same was true for subunits harboring a mutation in the catalytic center (KO–KO). All combinations involving the isoleucine variants AI and IV stimulated HR significantly. Attractively, upon expression of the combination IV–AI, 30% of transfected cells underwent HR, corresponding to a 2.5-fold increase as compared to the wild-type configuration.

To assess whether the modified nucleases were able to promote gene repair in a chromosomal setting, the frequency of GFP rescue was assessed in cell line 293/3-1, which contains a single integrated copy of the target locus [8]. In this assay, a CMV promoter regulated ZFN expression and the number of EGFP-positive 293/3-1 cells was determined seven days after transfection of a repair plasmid and the various ZFN combinations (Fig. 4d). In the absence of a nuclease, less than 0.1% of cells turned green. Expression of I-*Sce*I stimulated HR by 155-fold, which is in good agreement with previously published results [4, 8, 28-31], corresponding to 11% of transfected cells undergoing HR between the target locus and the repair plasmid. After transfection of the original ZFN expression vectors (WT–WT), the number of transfected cells after seven days was too low to assess the frequency of HR, probably due to high toxicity. As seen before in the episomal assays, combinations including

the asymmetric ZFN variants or the isoleucine variants stimulated HR significantly above basal HR, reaching 1.3% to 3.8%. Neither in the absence of a repair plasmid nor with the dimerization-defective (AA–AA) or the catalytic dead mutants (KO–KO) a significant increase in the HR frequency was observed. Evaluation of the expression levels by quantitative immunoblotting showed that the ZFN expression levels were roughly equivalent, with some slight decrease in the steady-state levels for the wild-type ZFNs and the QK variants (Fig. 4c). The highly significant differences regarding stimulation of HR can however not be solely reduced to changes in the ZFN concentration.

Double variants were created by combining the asymmetric ZFN mutants with the weakened dimer interface (Fig. 5a). Comparable steady-state levels of the ZFN double variants were confirmed by quantitative immunoblotting (Fig. 5b). To assess the gene repair kinetics, the amount of transfected ZFN expression vectors was increased and the extent of GFP rescue determined at day 3 and 7. For combinations DD–RR and IV–AI, a 4-fold increase in the amount of transfected DNA translated into a 2.5-fold increase in the HR frequency (Fig. 5c; 4.8% and 9.5%). The corresponding double variant RV–DA performed slightly better (9.7%), almost approaching the level of I-*Sce*I stimulated HR (19.3%). No increase was observed for EE–QK and the analogous double variant KV–EA, suggesting that higher amounts of these combinations may be toxic. As a result of the high cytotoxicity associated with the original ZFNs (WT–WT), the HR frequency was only measurable at day 3 with the lower DNA amount (0.75%). Taken together, these results establish that performance of ZFNs can be significantly improved by destabilizing dimerization of the two subunits and by promoting heterodimer formation.

ZFN dimerization variants elicit lower toxicity

Several of the preceding experiments suggested that improved ZFN performance is closely linked to reduced toxicity. One of the earliest responses after the creation of a DSB is phosphorylation of histone H2AX (γ-H2AX) and formation of repair foci near the DSB [32, 33]. For this reason, ZFN-associated genotoxicity was determined by quantitatively assessing the level of γ-H2AX in response to expression of the ZFNs in transfected HT1080 cells. The assay was validated by incubating the cells with etoposide, a known inducer of DSBs [34]. As determined by flow cytometry and immunofluorescence, increasing the concentration of etoposide led to elevated levels of γ-H2AX in the cells (Fig. 6a) as well as the formation of γ-H2AX containing repair foci (Fig. 6d). As compared to the catalytic dead mutant (KO–KO),

expression of the original ZFNs (WT–WT) significantly increased the overall level of γ-H2AX (Fig. 6b), the number of highly γ-H2AX-positive cells (Fig. 6c), and the number of repair foci (Fig. 6d). On the other hand, combinations involving the asymmetric DD–RR or the weak IV–AI dimer interface revealed a significantly reduced genotoxicity. In good agreement with the chromosomal gene repair data, QK–EE was more toxic than the other variants and the lowest toxicity was measured for the double variant combining DD–RR with IV–AI (RV–DA).

In concert with the previous experiments, these assays confirm that targeted modifications to the *Fok*I dimerization interface significantly reduce ZFN-induced toxicity and, in doing so, enhance performance of these artificial nucleases.

DISCUSSION

The ability to change deliberately the genetic program of a living cell not only allows the study of gene function but also the treatment of patients suffering from inherited disorders. This report substantiates that custom nucleases are a valuable tool to stimulate targeted manipulations of the genome and demonstrates for the first time that specific alterations within the dimer interface of these nucleases attenuate ZFN-associated toxicity, while promoting robust HR.

ZFN-induced toxicity most likely results from cleavage at off-target sites, which is determined by multiple factors, including the specificity of DNA-binding and DNA-cleavage. Here we hypothesized that targeted changes to the dimer interface of ZFNs will reduce toxicity associated with ectopic expression of ZFNs. Using 3D protein modeling and computer-based energy calculations, we identified potential residues in the *Fok*I dimer interface that are involved in dimerization of ZFNs. The potential of computer-assisted protein design is emphasized by the fact that 10 different modifications to the dimer interface with a total of 94 combinations thereof were scrutinized *in silico* in relatively short time. Only promising candidates were subsequently generated and evaluated in our cell-based assays.

Two general strategies were pursued to reduce ZFN-associated toxicity. The exchange of interacting charged residues in the protein–protein interface created asymmetric dimerization variants that contained likewise charged residues in the same subunit. In the chromosomal HR assay, variant DD–RR performed better than EE–QK, probably owing to the fact that at high ZFN concentrations also EE–EE homodimers could form, which might contribute to the observed toxicity. This is in agreement with the computer analysis,

predicting that DD–DD and RR–RR homodimers are much less favored than EE–EE and QK–QK homodimers. Also the isoleucine variants IV–AI, which are expected to dimerize only after binding to the respective target site, performed well. The combination of both strategies proved especially promising. At high concentrations, the variant combining DD/RR with IV/AI stimulated HR almost as well as the gold standard I-*Sce*I. This is probably owned to the fact that genotoxicity for this double variant was hardly detectable.

A scan of the human genome revealed 28104 binding sites for GZF1, 12647 GZF3 sites, a single heterodimeric GZF3-GZF1 site (the inserted target locus), and no binding sites for the homodimers GZF1-GZF1 and GZF3-GZF3 in the human genome (data not shown). This suggests that the effective ZFN concentration rather than cleavage at homodimeric sites seems to determine toxicity. The actual mechanism by which the variants overcome toxicity is speculative and complicated by a lack of knowledge of many key details. The natural *Fok*I restriction enzyme exists in solution and binds DNA as a monomer. To cleave DNA, two monomers associate through their catalytic domain to form a dimer with two active sites. *In vitro* studies showed that an active *Fok*I dimer can also form with just one subunit being bound to its recognition site [35]. We speculate that analogous to *Fok*I, a ZFN subunit binds to its respective target site as a monomer and becomes activated once it forms a dimer with a second subunit. At high concentrations, dimerization-on-the-fly, that is a dimeric ZFN complex with one monomer not attached to DNA, could form and contribute to toxicity. By destabilizing the dimer interface, as in the case of the isoleucine variants, these off-target effects would be lost, because the interaction energy for these complexes to form is too low. Variants that create an asymmetric dimerization interface may work in a related way. Preventing the formation of homodimers lowers the number of possible interaction partners and hence the number of active ZFNs.

Two studies apparently identified non-toxic ZFNs that contained a wild-type *Fok*I interface [2, 36]. These ZFNs were possibly composed of zinc finger domains that had better DNA-binding specificity than those used in this study. Depending on the specificity of the zinc finger DNA-binding domain, the dimeric ZFN complex may recognize a more or a less closely related family of 18-bp target sequences, which suggests that an active complex can also form if one or both monomers bind at appropriately spaced pseudo-sites at reduced affinity. Higher specificity of DNA-binding consequently means that less ZFN monomers are bound to pseudo-sites, which translates into less cleavage events. Characterizing the *in vitro* and cell-based affinity and specificity of all these reported ZFN would help to address this issue.

In summary, our results present both forced heterodimerization and destabilized interaction between two monomers as successful methods that can be combined to significantly reduce ZFN-associated toxicity without compromising on the activity. These findings present a critical step towards the therapeutic application of the ZFN technology.

5

METHODS

FoldX-based mutagenesis and optimization of structure

Mutagenesis modeling was done with FoldX version 2.65 (http://foldx.embl.de) [24, 37],

10 which tests different rotamers and allows neighbor side chains to move. The program first changes all residues to alanine, then it mutates to the selected residue (while moving the neighbor residues). For positions, where the amino acid residues of the model structure are moved, the program is able to store the original rotamer, and only if the interaction energy is better, the new rotamer will be considered. After reconstruction, all models have been passed

15 through an optimization by using the repair function of FoldX. During this procedure, FoldX identifies those residues that have e.g. bad torsion angles or van der Waals clashes. The way it operates is as follows: first it mutates the selected position to alanine and annotates the side-chain energies of the neighbor residues. Then it mutates the alanine to the selected amino acid and re-calculates the side-chain energies of the same neighbor residues. Those that exhibit an

20 energy difference are mutated to themselves to see if another rotamer will be more favorable. This procedure contains an additional function, where all side-chains are slightly moved in order to eliminate small steric clashes. The advantage of this is that it quickly eliminates small local clashes and gains time by decreasing the number of real rotamer searches. In this FoldX version we ensured that the same neighbors, which are moved in the mutant, are also moved

25 in the wild-type protein. Thus, for each mutant a specific reference wild-type protein is generated, which results in a significant improvement in the prediction of mutation effects (Stricher, F., Shcymkowitz, J., Borg, J., Serrano, L., manuscript in preparation).

Energy calculations using FoldX

30 Energy calculations of the *Fok*I dimers have been done using FoldX. The energy function includes terms that have been found to be important for protein stability, where the energy of unfolding (ΔG) of a target protein is calculated using the equation:

$$\Delta G = \Delta G_{vdw} + \cdot \Delta G_{solvH} + \cdot \Delta G_{solvP} + \Delta G_{wb} + \Delta G_{hbond} + \Delta G_{el} + \Delta Gkon + T \cdot \Delta S_{mc} + T \cdot \Delta S_{sc} + T \cdot \Delta S_{tr}$$

ΔG_{vdw} is the sum of the Van der Waals contributions of all atoms with respect to the same interactions with the solvent. ΔG_{solvH} and ΔG_{solvP} is the difference in solvation energy for apolar and polar groups, respectively, when going from the unfolded to the folded state. ΔG_{hbond} is the free energy difference between the formation of an intra-molecular hydrogen-bond compared to inter-molecular hydrogen-bond formation (with solvent). ΔG_{wb} is the extra stabilizing free energy provided by a water molecule making more than one hydrogen-bond to the protein (water bridges) that cannot be taken into account with non-explicit solvent approximations. ΔG_{el} is the electrostatic contribution of charged groups, including the helix dipole. ΔG_{kon} reflects the effect of electrostatic interactions on the k_{on} [38]. ΔS_{mc} is the entropy cost for fixing the backbone in the folded state. This term is dependent on the intrinsic tendency of a particular amino acid to adopt certain dihedral angles. ΔS_{sc} is the entropic cost of fixing a side chain in a particular conformation. ΔS_{sc} is the loss of translational and rotational entropy upon making the complex. The energy values of ΔG_{vdw}, ΔG_{solvH}, ΔG_{solvP} and ΔG_{hbond} attributed to each atom type have been derived from a set of experimental data, and ΔS_{mc} and ΔS_{mc} have been taken from theoretical estimates. The Van der Waals' contributions are derived from vapor to water energy transfer, while in the protein we are going from solvent to protein [24]. Details are described in previous publications [24, 37, 38].

Plasmids

Target plasmid pCMV.LacZs31∂GFP, repair plasmids pUC.Zgfp/Rex and pUC.Zgfp, and the nuclease expression vectors pRK5.LHA-Sce1, pRK5.GZF1-N, pRK5.GZF3-N, pPGK.GZF1-N and pPGK.GZF3-N have been described before [8]. Point mutations were introduced into the ZFN expression plasmids using site-directed mutagenesis. SSA reporter plasmids were generated in a pGL3 vector (Promega) background. PCR was used to amplify a 5'- and a 3'-fragment of the luciferase gene, each containing an 870 bp region of homologous overlap. The 5' fragment was inactive due to the truncation at the 3'-end and additionally contained a terminal stop codon followed by an introduced *EcoR*I site. The 3'-fragment was inactive due to the truncation at the 5'-end, which also contained an *EcoR*I site. These fragments were then cloned into pGL3 between the *Xho*I and *BamH*I sites using the *EcoR*I site to join the fragments. The ZFN target sites are contained in the junction of the fragments (Fig. 3a). To generate bacterial expression vectors, GZF1-N and GZF3-N were PCR amplified and subcloned into the *Xba*I and *Sal*I sites of pMal2-c2X (New England Biolabs), which appends an N-terminal maltose binding protein (MBP) domain for purification. Detailed maps and sequence files of the plasmids can be obtained upon request.

MBP-ZFN purification and *in vitro* assay

MBP-ZFN fusion proteins were expressed in *E. coli* strain ER2508 (New England Biolabs). Expression of ZFN protein was induced for 2 hours with IPTG (0.3 mM) in the presence of $ZnCl_2$ (0.1 mM). The cells were subsequently lysed and sonicated in Lysis Buffer (50 mM Tris-HCl (pH 8.0), 20 mM KCl, 2 mM $MgCl_2$, 1 mM DTT, 0.2% Triton X-100, and Complete protease inhibitor without EDTA (Roche)). Protein purification using amylose resin (New England Biolabs) and elution in Lysis Buffer containing 10 mM maltose was performed as recommended by the manufacturer. Proteins were stored in 30% glycerol at –20°C until use. ZFN protein activities were analyzed in Cleavage Buffer (10 mM Tris-HCl (pH 8.0), 90 mM KCl, 0.1 mM $ZnCl_2$, 5 mM DTT) using 0.5 µg (100 fmol) of a linear DNA substrate corresponding to the SSA reporter plasmids (cleaved at *Xho*I) and 50 ng (500 fmol) of each purified protein. The reactions were incubated for 15 minutes at 25°C before adding $MgCl_2$ to 10 mM, followed by an additional incubation of 30 minutes at 37°C and analysis on a 1% agarose gel.

Cell lines and cell-based recombination assays

Cell lines HT1080, 293T and 293/3-1 [8] were grown in DMEM supplemented with 10% fetal calf serum.

For the SSA assay, 293T cells were transfected in a 96 well plate format using lipofectamine 2000 transfection reagent (Invitrogen). 200 ng of pPGK expression vectors and 100 ng of reporter plasmids were co-transfected into 50,000 cells/well. The ZFN-expressing vectors were mixed at different ratios (200/0, 160/40, 100/100, 40/160, 0/200) to analyze activity. After 24 hours, the cells were lysed in 1X lysis buffer (Promega) and luciferase activity was determined using BrightGlo (Promega) in a plate luminometer (Molecular Devices).

For the episomal HR assay, 293T cells in 24-well plates were transfected by calcium phosphate precipitation with 20 ng of a target plasmid, 1 µg of the repair plasmid (pUC.Zgfp/REx) or a repair control (pUC.REx), and 0.5 µg of a PGK-driven endonuclease expression vector encoding the ZFNs, I-*Sce*I (pRK5.LHA-Sce1) or a control vector (pCMV.Luc). Two days after transfection, 50,000 cells were analyzed by flow cytometry (FACSCalibur, BD Bioscience) to determine the percentage of EGFP- and DsRed-Express(REx)-positive cells.

For the chromosomal HR assay, 293/3-1 cells in 12-well plates were transfected by calcium phosphate precipitation with 2 μg of the repair plasmid (pUC.Zgfp) or a repair control (pUC118), 100 or 400 ng of a CMV-driven nuclease expression vector or control vector (pCMV.Luc), and 10 ng of pDsRed-Express-N1 (Clontech). Three and 7 days after transfection, 50,000 cells were analyzed by flow cytometry to determine the percentage of EGFP- and REx-positive cells. The number of REx-positive cells at day 3 was used to normalize for transfection efficiency. If the percentage of transfected cells at day 3 dropped below 1%, results were not evaluated.

Statistical significance was determined using the student's t-test.

Quantitative immunoblotting

293T cells in 35 mm wells were transfected by calcium phosphate precipitation with 800 ng of a CMV-controlled ZFN expression plasmid, 200 ng of pEGFP (Clontech) and pUC118 to 4 μg. After 30 h cells were harvested and resuspended in lysis buffer [8]. Equal amounts of proteins (50 μg) were separated by SDS polyacrylamide gel electrophoresis and transferred to Immobilon-P membrane (Millipore). The blots were concomitantly incubated with a rabbit anti-HA antibody (Novus Biologicals #NB600-363, 1:5000) and with mouse anti-GFP antibody MAB3580 (Chemicon, 1:5000) for 2 h at RT. After incubation with secondary antibodies conjugated with IRDye-680 or IRDye-800CW (Li-Cor Bioscience, both 1:20,000), protein expression was visualized and quantitated using an Odyssey scanner (Li-Cor Bioscience) according the manufacturer's instructions.

Indirect immunofluorescence

Immunofluorescence was basically performed as described [39, 40]. Briefly, HT1080 cells on gelatin-coated 13 mm glass cover slips were transfected with 200 ng each of a CMV-controlled ZFN expression plasmid and pUC118 to 800 ng using lipofectamine 2000 (Invitrogen). Alternatively, cells were exposed to etoposide at the indicated concentrations for 60 min 2 h prior to harvesting. After 30 h, cells were fixed with 4% paraformaldehyde in PBS and permeabilized in 0.5% Triton X-100. After blocking, cells were incubated for 1 h with mouse anti-γH2AX (Upstate #05-636; 1:1000) and rabbit anti-HA (Novus Biologicals #NB600-363; 1: 2000) in 3% BSA in PBS, followed by staining for 1 h with Cy2-conjugated goat anti-mouse, Cy3-conjugated goat anti-rabbit (both Jackson ImmunoResearch; 1:500) and 1 μg/ml 4,6-diamidino-2-phenyindole (DAPI; Sigma) in 1% BSA in PBS. Cover slips were mounted with Fluoromount-G (Southern Biotechnology) and analyzed by fluorescence

microscopy.

Quantitative genotoxicity assay

HT1080 cells in 12-well plates were transfected with 400 ng each of a CMV-controlled nuclease expression vector, 400 ng of pEGFP-tubulin [8], and pUC118 to 1.6 μg using Lipofectamine 2000 (Invitrogen). Alternatively, cells were exposed to etoposide at the indicated concentrations for 60 min 2 h prior to harvesting. Cells were collected after 30 h by trypsinization and fixed in 75% EtOH. After blocking in 3% BSA in PBS, cells were stained for 2 h at 4°C with mouse anti-γH2AX (Upstate #05-636; 1:1000) in 3% BSA, followed by incubation for 1 h at 4°C with AlexaFluor594-conjugated goat anti-mouse (Molecular Probes; 1:200) in 1% BSA in PBS. 25,000 cells were analyzed by flow cytometry (FACSCalibur, BD Bioscience) to determine the level of γH2AX in EGFP-positive cells. Statistical significance was assessed using the student's t-test.

REFERENCES

1. Vasquez, K.M., Marburger, K., Intody, Z. & Wilson, J.H. Manipulating the mammalian genome by homologous recombination. *Proc Natl Acad Sci U S A* **98**, 8403-8410 (2001).

2. Urnov, F.D. et al. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646-651 (2005).

3. Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol* **14**, 8096-8106 (1994).

4. Choulika, A., Perrin, A., Dujon, B. & Nicolas, J.F. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. *Mol Cell Biol* **15**, 1968-1973 (1995).

5. Epinat, J.C. et al. A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res* **31**, 2952-2962 (2003).

6. Porteus, M.H. Mammalian gene targeting with designed zinc finger nucleases. *Mol Ther* **13**, 438-446 (2006).

7. Porteus, M.H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**, 763 (2003).

8. Alwin, S. et al. Custom zinc-finger nucleases for use in human cells. *Mol Ther* **12**, 610-617 (2005).

9. Kim, Y.G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A* **93**, 1156-1160 (1996).

10. Liu, Q., Xia, Z., Zhong, X. & Case, C.C. Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *J Biol Chem* **277**, 3850-3856 (2002).

11. Segal, D.J., Dreier, B., Beerli, R.R. & Barbas, C.F., 3rd Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc Natl Acad Sci U S A* **96**, 2758-2763. (1999).

12. Dreier, B., Beerli, R.R., Segal, D.J., Flippin, J.D. & Barbas, C.F., 3rd Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem* **276**, 29466-29478. (2001).

13. Dreier, B. et al. Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem* **280**, 35588-35597 (2005).

14. Blancafort, P., Magnenat, L. & Barbas, C.F., 3rd Scanning the human genome with combinatorial transcription factor libraries. *Nat Biotechnol* **21**, 269-274 (2003).

15. Hurt, J.A., Thibodeau, S.A., Hirsh, A.S., Pabo, C.O. & Joung, J.K. Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc Natl Acad Sci U S A* **100**, 12271-12276 (2003).

16. Smith, J. et al. Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res* **28**, 3361-3369 (2000).

17. Bitinaite, J., Wah, D.A., Aggarwal, A.K. & Schildkraut, I. FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci U S A* **95**, 10570-10575 (1998).

18. Bibikova, M. et al. Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol* **21**, 289-297 (2001).

19. Silva, G.H., Belfort, M., Wende, W. & Pingoud, A. From monomeric to homodimeric endonucleases and back: engineering novel specificity of LAGLIDADG enzymes. *J Mol Biol* **361**, 744-754 (2006).

20. Sims, P.A., Menefee, A.L., Larsen, T.M., Mansoorabadi, S.O. & Reed, G.H. Structure and catalytic properties of an engineered heterodimer of enolase composed of one active and one inactive subunit. *J Mol Biol* **355**, 422-431 (2006).

21. Bolon, D.N., Grant, R.A., Baker, T.A. & Sauer, R.T. Specificity versus stability in computational protein design. *Proc Natl Acad Sci U S A* **102**, 12724-12729 (2005).

22. Wah, D.A., Hirsch, J.A., Dorner, L.F., Schildkraut, I. & Aggarwal, A.K. Structure of the multimodular endonuclease FokI bound to DNA. *Nature* **388**, 97-100 (1997).

23. Wah, D.A., Bitinaite, J., Schildkraut, I. & Aggarwal, A.K. Structure of FokI has implications for DNA cleavage. *Proc Natl Acad Sci U S A* **95**, 10564-10569 (1998).

24. Schymkowitz, J.W. et al. Prediction of water and metal binding sites and their affinities by using the Fold-X force field. *Proc Natl Acad Sci U S A* **102**, 10147-10152 (2005).

25. van der Sloot, A.M. et al. Designed tumor necrosis factor-related apoptosis-inducing ligand variants initiating apoptosis exclusively via the DR5 receptor. *Proc Natl Acad Sci U S A* **103**, 8634-8639 (2006).

26. Kölsch, V., Seher, T., Fernandez-Ballester, G.J., Serrano, L. & Leptin, M. Control of Drosophila gastrulation by apical localization of adherens junctions and RhoGEF2. *Science* **315**, 384-386 (2007).

27. Li, X. et al. Deletions of the Aequorea victoria green fluorescent protein define the minimal domain required for fluorescence. *J Biol Chem* **272**, 28545-28549 (1997).

28. Miller, D.G., Petek, L.M. & Russell, D.W. Human gene targeting by adeno-associated virus vectors is enhanced by DNA double-strand breaks. *Mol Cell Biol* **23**, 3550-3557 (2003).

29. Sargent, R.G., Brenneman, M.A. & Wilson, J.H. Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination. *Mol Cell Biol* **17**, 267-277 (1997).

30. Porteus, M.H., Cathomen, T., Weitzman, M.D. & Baltimore, D. Efficient gene targeting mediated by adeno-associated virus and DNA double-strand breaks. *Mol Cell Biol* **23**, 3558-3565 (2003).

31. Smih, F., Rouet, P., Romanienko, P.J. & Jasin, M. Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Res* **23**, 5012-5019 (1995).

32. Rogakou, E.P., Boon, C., Redon, C. & Bonner, W.M. Megabase chromatin domains involved in DNA double-strand breaks in vivo. *J Cell Biol* **146**, 905-916 (1999).

33. Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S. & Bonner, W.M. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* **273**, 5858-5868 (1998).

34. Banath, J.P. & Olive, P.L. Expression of phosphorylated histone H2AX as a surrogate of cell killing by drugs that create DNA double-strand breaks. *Cancer Res* **63**, 4347-4350 (2003).

35. Catto, L.E., Ganguly, S., Milsom, S.E., Welsh, A.J. & Halford, S.E. Protein assembly and DNA looping by the FokI restriction endonuclease. *Nucleic Acids Res* **34**, 1711-1720 (2006).

36. Beumer, K., Bhattacharyya, G., Bibikova, M., Trautman, J.K. & Carroll, D. Efficient gene targeting in Drosophila with zinc-finger nucleases. *Genetics* **172**, 2391-2403 (2006).

37. Guerois, R., Nielsen, J.E. & Serrano, L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol* **320**, 369-387 (2002).

38. Vijayakumar, M. et al. Electrostatic enhancement of diffusion-controlled protein-protein association: comparison of theory and experiment on barnase and barstar. *J Mol Biol* **278**, 1015-1024 (1998).

39. Cathomen, T. & Weitzman, M.D. A functional complex of adenovirus proteins E1B-55kDa and E4orf6 is necessary to modulate the expression level of p53 but not its transcriptional activity. *J Virol* **74**, 11407-11412. (2000).

40. Han, J., Hendzel, M.J. & Allalunis-Turner, J. Quantitative analysis reveals asynchronous and more than DSB-associated histone H2AX phosphorylation after exposure to ionizing radiation. *Radiat Res* **165**, 283-292 (2006).

Figure 1 Structure-based redesign of the *Fok*I dimer interface. (**a**) Schematic of ZFN. All ZFNs contain an N-terminal HA tag followed by three zinc-finger modules, a short AAARA linker and the nuclease domain of *Fok*I. The sequence of helix α4 (positions 480-490) along with positions 499 and 538 is shown for wild-type (WT) and all variants. The two-letter code defining the variants and the corresponding point mutations are indicated on the left. (**b-d**) Model of wild-type *Fok*I dimer interface and variants (**e-g**). The models are based on the crystal structure of *Fok*I and depict amino acids 479 through 539. Critical residues implicated in dimerization of the wild-type nuclease are highlighted: D483–R487 in (b), Q486–E490 in (c), and I499/I538 in (d). Dimerization of variant DD with RR is shown in (e), EE with QK in (f), and AI with IV in (g). Acidic residues are depicted in green, basic and polar residues in red, hydrophobic residues in pink, and small amino acids in orange. (**h**) Schematic of strategy. The crucial residues implicated in dimerization are depicted for the wild-type configuration (top) and for the variants (bottom). (**i**) *In silico* computation of dimerization energy. The interaction energy of two variant ZFN subunits was calculated using FoldX version 2.65. The graph indicates the interaction energy with respect to the combination of two wild-type subunits (ΔΔG). For the heterodimers, the graph shows the average of the two calculated values in case slightly different interaction energies were obtained. The respective combinations are indicated on the bottom.

Figure 2 *In vitro* cleavage specificity. (**a**) Schematic of *in vitro* cleavage assay. A linear DNA substrate containing two binding sites for GZF1-N (1-1), two sites for GZF3-N (3-3), or one site each for GZF3-N and GZF1-N (3-1) is cleaved in different size products. (**b**) Analysis of homo- and heterodimer cleavage reactions. Purified ZFNs were incubated with the linear DNA substrate and the extent of cleavage analyzed by agarose gel electrophoresis. Reactions analyzing a single ZFN contained an equal concentration of purified maltose binding protein (•). Control reactions were substrate alone (−) and cleavage with *EcoR*I (+).

Figure 3 Single strand annealing assay. (**a**) Schematic of SSA assay. The reporter plasmids encode two truncated inactive parts of the luciferase gene with overlapping repeated sequence (LR and RR). Following a DSB inserted by the ZFNs, a functional luciferase gene is generated by a SSA reaction. (**b-d**) Cellular activity of wild-type and variant ZFNs. 293T cells were transfected with a reporter plasmid and ZFN expression vectors mixed at different

ratios (200/0, 160/40, 100/100, 40/160, 0/200) as indicated on top. The extent of SSA was determined in a plate luminometer after 24 h and the graphs illustrate fold induction over background.
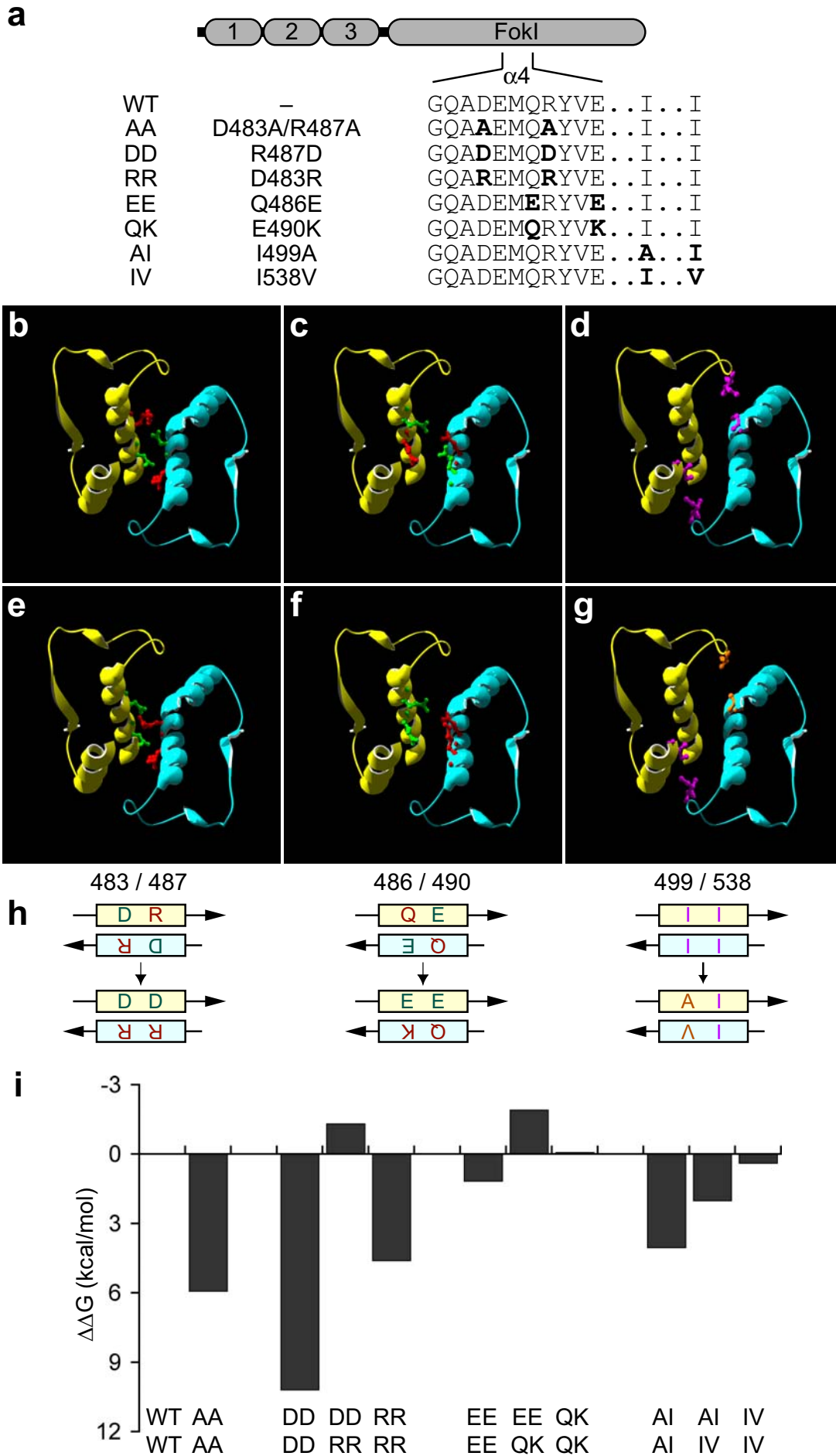
Figure 4 Stimulation of HR with episomal and chromosomal target locus. (**a**) Experimental set up. Repair and target plasmid are described in the text. The stop codons (bold), the recognition site for I-*Sce*I (boxed), and the binding sites for zinc fingers GZF1 and GZF3 (grey boxes) are highlighted. (**b**) Episomal gene repair. 293T cells transfected with repair plasmid, target plasmid and the nuclease expression vectors as indicated were analyzed by flow cytometry after 2 days. The graph displays the fraction of EGFP-positive cells in relation to transfected cells. "control" indicates transfection with a control vector while * points out statistically significant increase in HR (p<0.001) above non-stimulated HR (control). (**c**) ZFN expression levels. Transfected 293T cells were harvested after 30 h and lysates probed with an antibody against the HA tag. Proteins were visualized using an Odyssey scanner. "cto" indicates transfection with a luciferase expression vector. (**d**) Comparison of dimerization variants. 293/3-1 cells, which contain a single copy of the target locus, were transfected with 100 ng of the nuclease expression vectors and repair plasmid (RP) or a control RP (cto). The columns designate the fraction of EGFP-positive cells 7 days post-transfection normalized for transfection efficiency. * and ** indicate statistically significant increase in HR (p<0.01 and p<0.001) above non-stimulated HR (cto, column 3); "bd" stands for "below detection" in the case the number of REx-positive cells was too low due to cell death (<1% at day 3 or <0.1% at day 7).
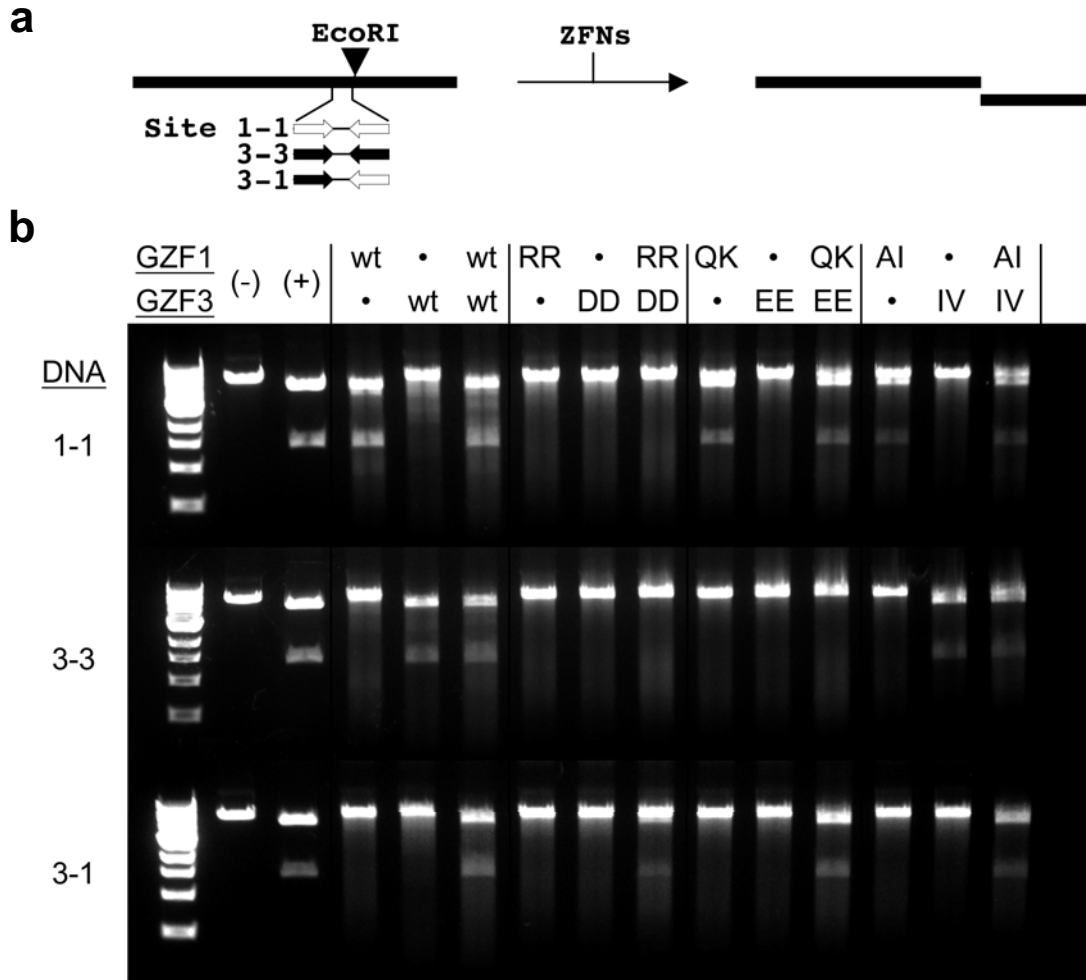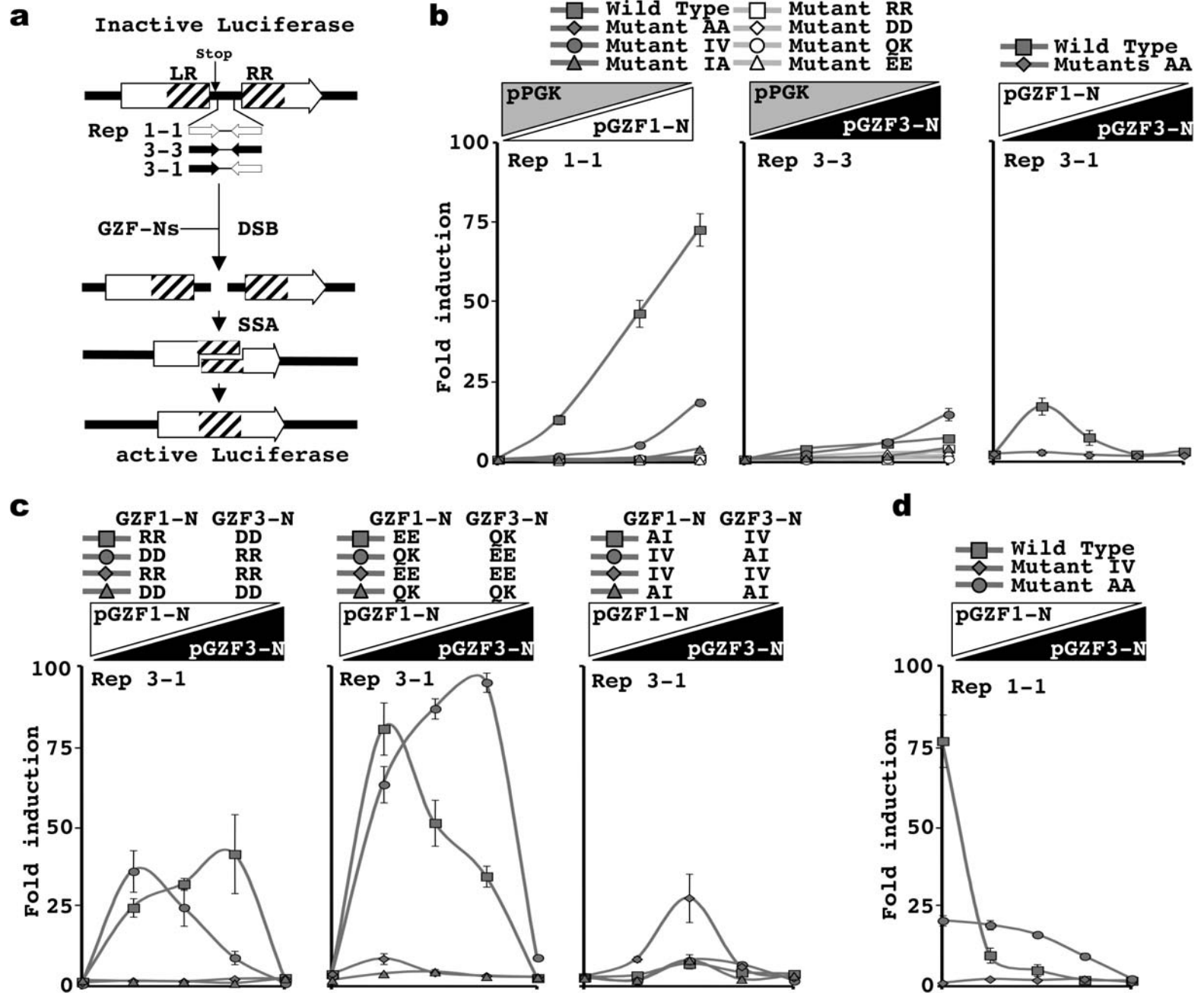
Figure 5 Stimulation of chromosomal HR by single and double variant ZFNs. (**a**) Schematic of double variant ZFNs. The sequence of helix α4 (positions 480-490) along with positions 499 and 538 is shown for wild-type (WT) and the double variants. The two-letter code and the corresponding point mutations are indicated on the left. (**b**) Expression levels of ZFN double variants. Transfected 293T cells were harvested after 30 h and lysates probed with an antibody against the HA tag. Proteins were visualized using an Odyssey scanner. "cto" indicates transfection with a luciferase expression vector. (**c**) Titration of dimerization variants and kinetics of gene repair. 293/3-1 cells were transfected with repair plasmid and nuclease expression vector as indicated (100 or 400 ng). The columns designate the fraction of EGFP-positive cells 3 and 7 days post-transfection. * and ** indicate statistically significant increase in HR (p<0.01 and p<0.001) above non-stimulated HR (cto). "bd" stands

for "below detection" in the case the number of REx-positive cells was too low due to cell death (<1% at day 3 or <0.1% at day 7).
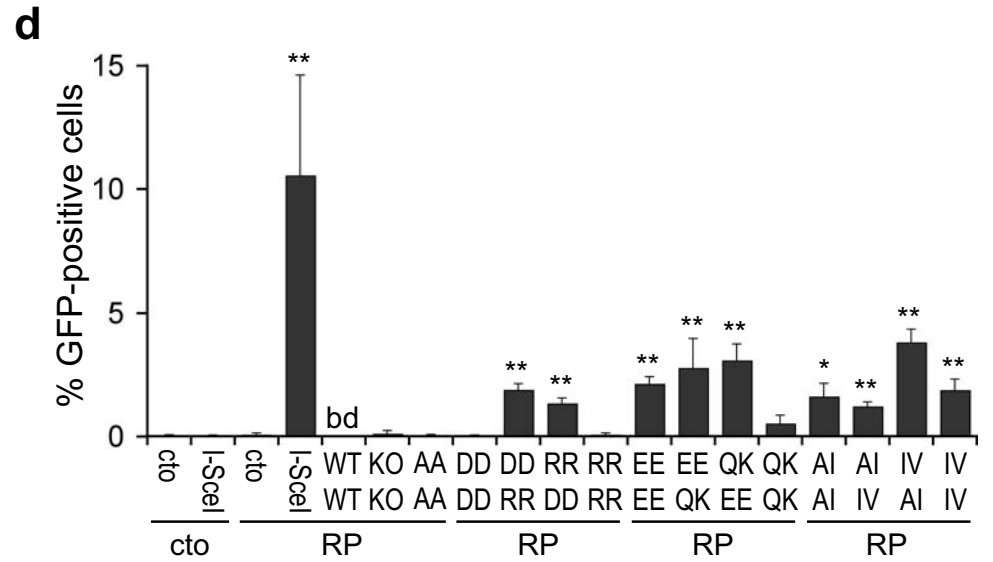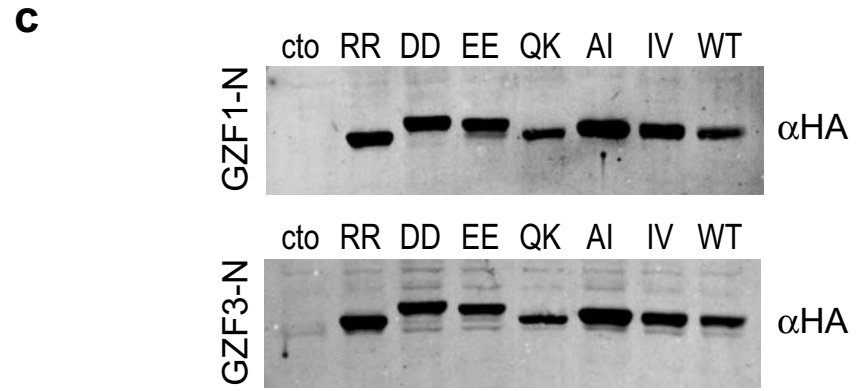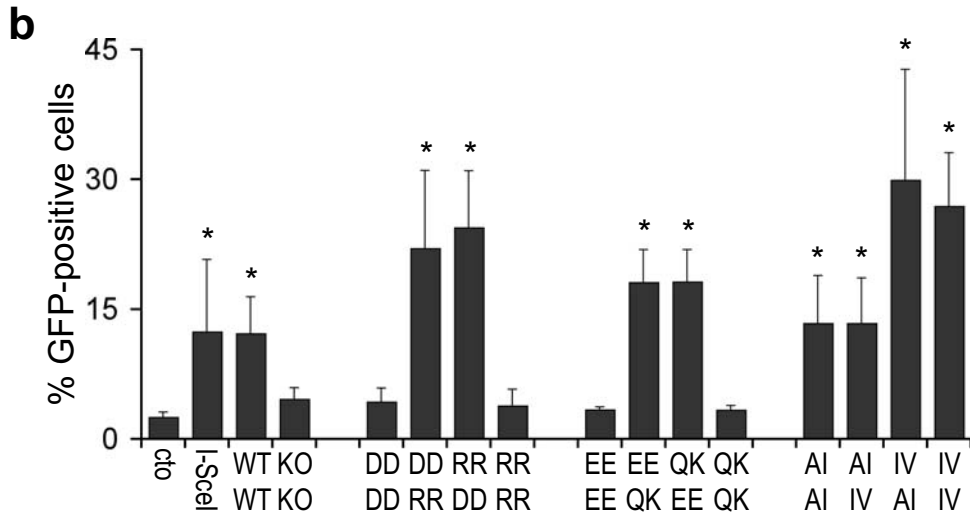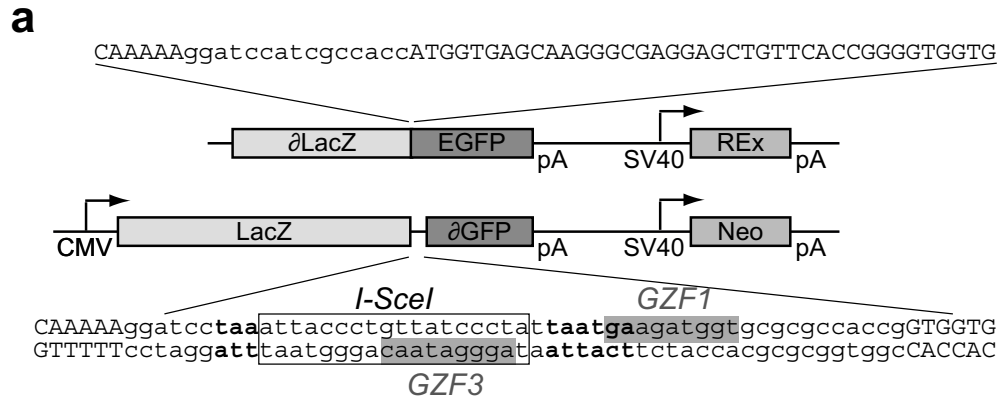
Figure 6 ZFN-associated genotoxicity. (**a, b**) Flow cytometric analysis of cellular γ-H2AX expression. HT1080 cells were exposed to etoposide or transfected with pEGFP-tubulin and nuclease expression vectors as indicated. Cellular γ-H2AX expression levels were determined after 30 h by flow cytometry and are representatively shown for etoposide treated cells (a) and cells transfected with the catalytic dead ZFN (KO–KO), a double variant (RV–DA) and the original ZFN (WT–WT). The average mean fluorescence intensity (MFI) is indicated. (**c**) Percentage of highly γ-H2AX-positive cells. After flow cytometric analysis of cellular γ-H2AX expression levels, a gate was drawn (dotted line) to encompass about 1% of untreated cells or cells transfected with a control nuclease (KO–KO). The columns denote the number of γ-H2AX-positive cells as fraction of transfected cells and represent the average of 3 experiments. # indicates statistically significant reduction of toxicity as compared to wild-type ZFN ($p<0.02$). (**d**) Induction of nuclear γ-H2AX foci. HT1080 cells were exposed to etoposide (top panels) or transfected with ZFN expression vectors as indicated on the left. Cells were fixed after 30 h and stained with antibodies against γ-H2AX (top and right hand panels), HA (to detect ZFN) and DAPI, as indicated. Due to the lack of a nuclear localization signal, the ZFNs are localized throughout the cell.
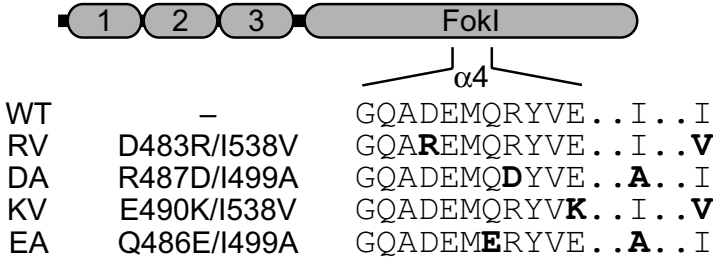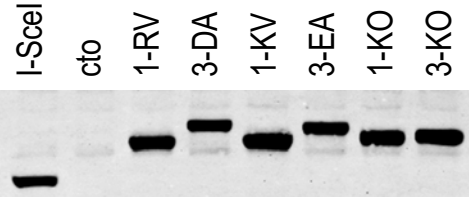
a

WT	–	GQADEMQRYVE..I..I
AA	D483A/R487A	GQA**A**EMQ**A**YVE..I..I
DD	R487D	GQA**D**EMQ**D**YVE..I..I
RR	D483R	GQA**R**EMQ**R**YVE..I..I
EE	Q486E	GQADEM**E**RYV**E**..I..I
QK	E490K	GQADEM**Q**RYV**K**..I..I
AI	I499A	GQADEMQRYVE..**A**..**I**
IV	I538V	GQADEMQRYVE..**I**..**V**

i

a

CAAAAAggatccatcgccaccATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTG

b

c

d

a

b

c