

## SURVEY AND SUMMARY

# Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells

Sundar Durai<sup>1,3</sup>, Mala Mani<sup>1</sup>, Karthikeyan Kandavelou<sup>1,2</sup>, Joy Wu<sup>1</sup>,  
Matthew H. Porteus<sup>4,5</sup> and Srinivasan Chandrasegaran<sup>1,\*</sup>

<sup>1</sup>Department of Environmental Health Sciences, The Johns Hopkins University Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 21205-2179, USA, <sup>2</sup>Pondicherry Biotech Private Ltd., 21 Louis Pragasam Street, Pondicherry 605001, India, <sup>3</sup>Center for Bioinformatics, School of Life Sciences, Pondicherry University, Pondicherry 605014, India, <sup>4</sup>Department of Pediatrics and <sup>5</sup>Department of Biochemistry, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390, USA

Received August 2, 2005; Revised September 22, 2005; Accepted October 3, 2005

### ABSTRACT

Custom-designed zinc finger nucleases (ZFNs), proteins designed to cut at specific DNA sequences, are becoming powerful tools in gene targeting—the process of replacing a gene within a genome by homologous recombination (HR). ZFNs that combine the non-specific cleavage domain (N) of FokI endonuclease with zinc finger proteins (ZFPs) offer a general way to deliver a site-specific double-strand break (DSB) to the genome. The development of ZFN-mediated gene targeting provides molecular biologists with the ability to site-specifically and permanently modify plant and mammalian genomes including the human genome via homology-directed repair of a targeted genomic DSB. The creation of designer ZFNs that cleave DNA at a pre-determined site depends on the reliable creation of ZFPs that can specifically recognize the chosen target site within a genome. The (Cys<sub>2</sub>His<sub>2</sub>) ZFPs offer the best framework for developing custom ZFN molecules with new sequence-specificities. Here, we explore the different approaches for generating the desired custom ZFNs with high sequence-specificity and affinity. We also discuss the potential of ZFN-mediated gene targeting for 'directed mutagenesis' and targeted 'gene editing' of the plant and mammalian genome as well as the potential of ZFN-based strategies as a

form of gene therapy for human therapeutics in the future.

### INTRODUCTION

The genome of biological species encodes the DNA sequence, the digital information guiding cellular life that can now be completely deciphered. With the advent of genomics over the past decade, the genome of many microbes, plants, animals and humans have been completely sequenced and annotated. Scientists can now embark on their quest to understand and manipulate a biological system as a whole with a known core of complete genomic sequence information of a species. The genome codes for two types of information, namely, genes that are responsible for developmental and physiological function of the organism; and control elements, which modulate the levels of expression of individual genes that execute functions of life. They form gene regulatory networks with the groups of other genes to mediate physiological and developmental responses. This genomic sequence information is critical for genome engineering of plant and mammalian cells including the human cells without disturbing the program of life encoded by the genome.

Molecular biologists have long sought the ability to manipulate or modify plant and mammalian genomes including the human genome at specific-sites. How does one achieve targeted genome engineering of plant and mammalian cells? (1,2). Cells use the universal process of homologous recombination (HR) to mediate site-specific recombination to maintain their genomic integrity, especially during the repair of a

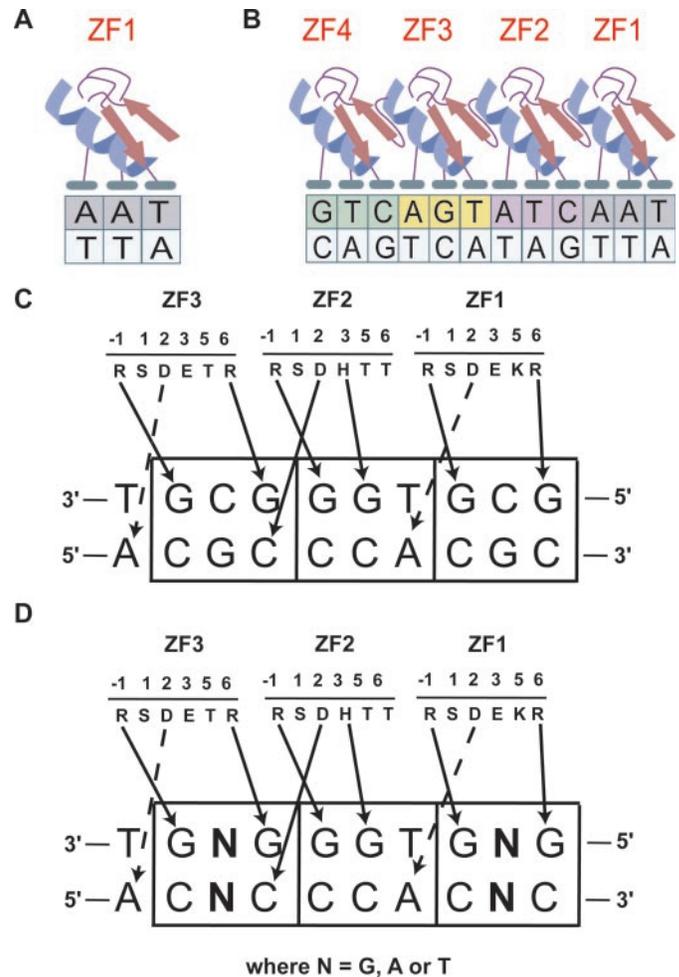
\*To whom correspondence should be addressed. Tel: 410 614 2289; Fax: 410 955 0299; Email: schandra@jhsph.edu

double-strand break (DSB). DSBs otherwise would be lethal to cells since they have the potential to scramble the digital information encoded within the genome of cells. DSB repair of a damaged chromosome by HR in a cell is the most accurate form of repair, which works via the copy and paste mechanism, usually using the homologous DNA segment from the undamaged sister-chromatid as a template. Gene targeting—the process of replacing a gene by HR—uses an extra-chromosomal fragment of donor DNA and invokes the cell's own repair machinery for gene conversion (3,4). Spontaneous gene targeting is not a very efficient process in mammalian and plant cells; about only one in a million treated cells undergo the desired gene modification. However, the introduction of a defined chromosomal break at a unique site within a genome has been shown to stimulate HR at that local site in cells to repair the DSB up to 50 000-fold (5,6).

Zinc finger nucleases (ZFNs)—proteins custom-designed to cut at specific DNA sequences—were originally developed in our lab for this purpose of delivering a targeted genomic DSB within plant and mammalian cells to enable such gene targeting experiments. Reports from several labs including ours using model systems have shown that custom-designed three-finger ZFNs find and cleave their chromosomal targets in cells; and as expected, they induce local HR at the site of cleavage (1,7–10). More recently, Urnov *et al.* (11,12) designed four-finger ZFNs that recognize an endogenous target site within the IL2R $\gamma$  gene underlying the human X-linked disease, severe combined immune deficiency (SCID) and used them for ZFN-mediated gene targeting to achieve highly efficient and permanent modification of the IL2R $\gamma$  gene in human cells—a remarkable gene modification efficiency of 18% of treated cells was obtained without selection, approximately one-third of which were altered on both X-chromosomes—attesting to the awesome power of the ZFN technology and raising the hope of applying ZFN-based strategies for human therapeutics in the future and raising the hope of applying ZFN-based strategies for human therapeutics in the future. In this review we discuss the design and selection of zinc finger proteins (ZFPs) (section 2), the development of ZFNs (section 3), the general mechanism of DSB repair (section 4), and finally, the use of ZFNs in stimulating gene targeting (section 5).

## ZFPs

The modular structure of zinc finger (ZF) motifs and modular recognition by ZF domains (Figure 1) make them the most versatile of the DNA recognition motifs for designing artificial DNA-binding proteins (13,14). Each ZF motif consists of ~30 amino acids and folds into  $\beta\beta\alpha$  structure (Figure 1A), which is stabilized by chelation of a zinc ion by the conserved Cys<sub>2</sub>His<sub>2</sub> residues. The ZF motifs bind DNA by inserting the  $\alpha$ -helix into the major groove of the DNA double helix (15). Each finger primarily binds to a triplet within the DNA substrate. Key amino acid residues at positions -1, +1, +2, +3, +4, +5 and +6 relative to the start of the  $\alpha$ -helix of each ZF motifs contribute to most of the sequence-specific interactions with the DNA site (15–17). These amino acids can be changed while maintaining the remaining amino acids as a consensus backbone to generate ZF motifs with different triplet sequence-specificities (18,19). Binding to longer DNA

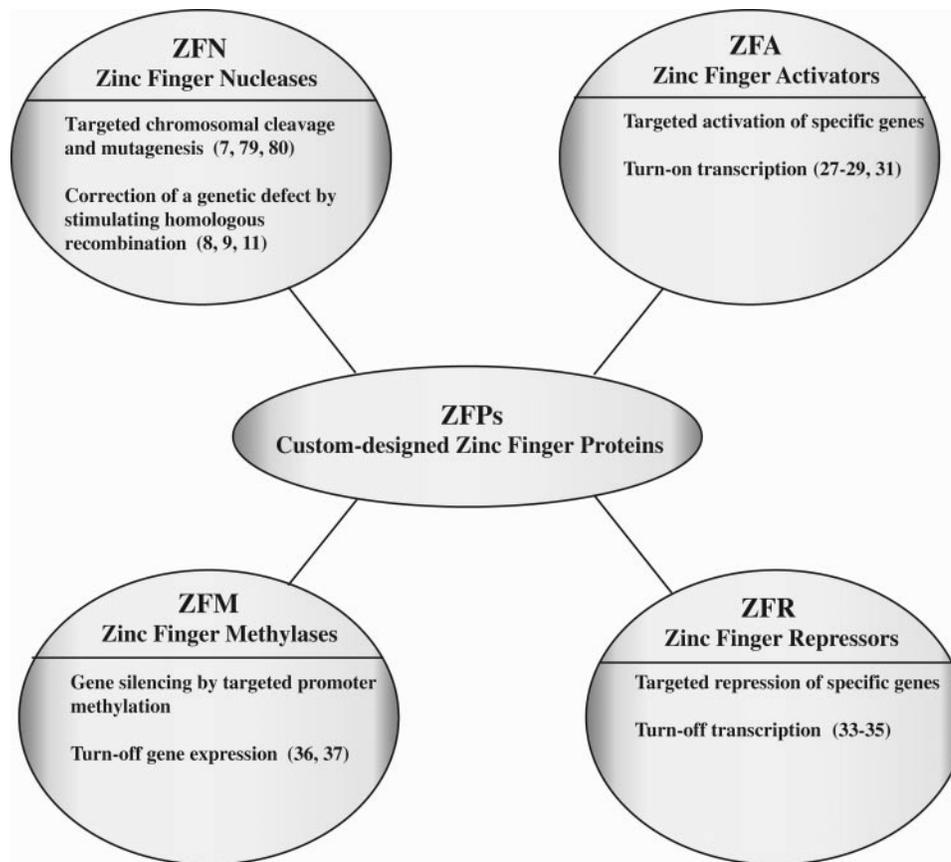


**Figure 1.** Structures of an individual ZF motif (A) and a four-finger ZFP, which is formed by linking four such individual ZF motifs (B). Zif268 bound to its cognate DNA site (C) and potential binding of Zif268 to degenerate sites (D). The key base contacts were deduced from the crystal structure of Zif268–DNA complex (44). Each finger makes contact with its target 3 bp site. In addition, Asp<sup>-</sup> at position 2 in each finger makes contact with a base outside the 3 bp site (C). Fingers 1 and 3 of Zif268 make specific contacts only with two bases of their cognate DNA triplets, while specific base contacts are seen with all the three bases of finger 2. Zif268 could potentially bind to other secondary or degenerate sites (D) as indicated, where N = G, A, T or C.

sequences is achieved by linking several of these ZF motifs in tandem to form ZFPs (Figure 1B) (20–22). The designed ZFPs provide a powerful platform technology (Figure 2) since other functionalities like non-specific FokI cleavage domain (N), transcription activator domains (A), transcription repressor domains (R) and methylases (M) can be fused to a ZFPs to form ZFNs respectively, (23–26), zinc finger transcription activators (ZFA) (27–32), zinc finger transcription repressors (ZFR) (33–35) and zinc finger methylases (ZFM) (36,37) (Figure 2).

## Design strategy for ZFPs

The creation of ZFNs that recognize and cleave any target sequence depends on the reliable creation of ZFPs that can specifically recognize that chosen target. The identification of ZF motifs that specifically recognize each of the 64



**Figure 2.** ZFP platform technology. Other functionalities like non-specific FokI cleavage domain (N), transcription activator domain (A), transcription repressor domain (R) and methylases (M) can be fused to ZFPs to form respectively ZFN, ZFA, ZFR and ZFM. The literature references for each of these chimeric ZFPs and their potential biological applications are given in brackets.

possible DNA triplets is a key step towards the construction of 'artificial' DNA-binding proteins that recognize any pre-determined target sequence within a plant or mammalian genome. Although numerous structural studies of ZFP-DNA complexes are available, the information is not yet sufficient for rational design of ZFPs that bind to any pre-selected DNA sequence (13). So far, only ZF modules that specifically recognize 5'-GNN-3' and 5'-ANN-3' triplets have been identified through design and selection strategies (38-40). More recently, ZF motif recognition preferences for the 5'-CNN-3' triplets have been published (41). A few ZF motif designs for 5'-TNN-3' triplets are also available in literature (42).

The designed ZFPs appear to have the highest affinity and sequence-specificity for their targets only when the individual ZF motif designs are chosen in the context of their neighboring fingers. The presence of Asp<sup>2</sup> at position 2 of the  $\alpha$ -helix of the preceding ZF motif promotes a cross-strand contact to a base outside the canonical triplet site, resulting in a target site overlap (15,43) (Figure 1C). While this increases the affinity of the ZFP to the target site, it also precludes the presence of a simple general recognition code for easy rational design of ZF motifs-based DNA-binding proteins. The target site overlap problem arises only when Asp is present at position 2 of the  $\alpha$ -helix of the preceding ZF motif. Thus, in many instances because of the presence Asp<sup>2</sup>, the individual ZF motifs of a ZFP rather than binding to a DNA triplet actually recognizes a

4 bp DNA target complicating the design strategy. However, many studies have shown that ZFPs with sufficient affinity and specificity suitable for many biological applications could be engineered using the known ZF motif designs by a simple assembly strategy (25).

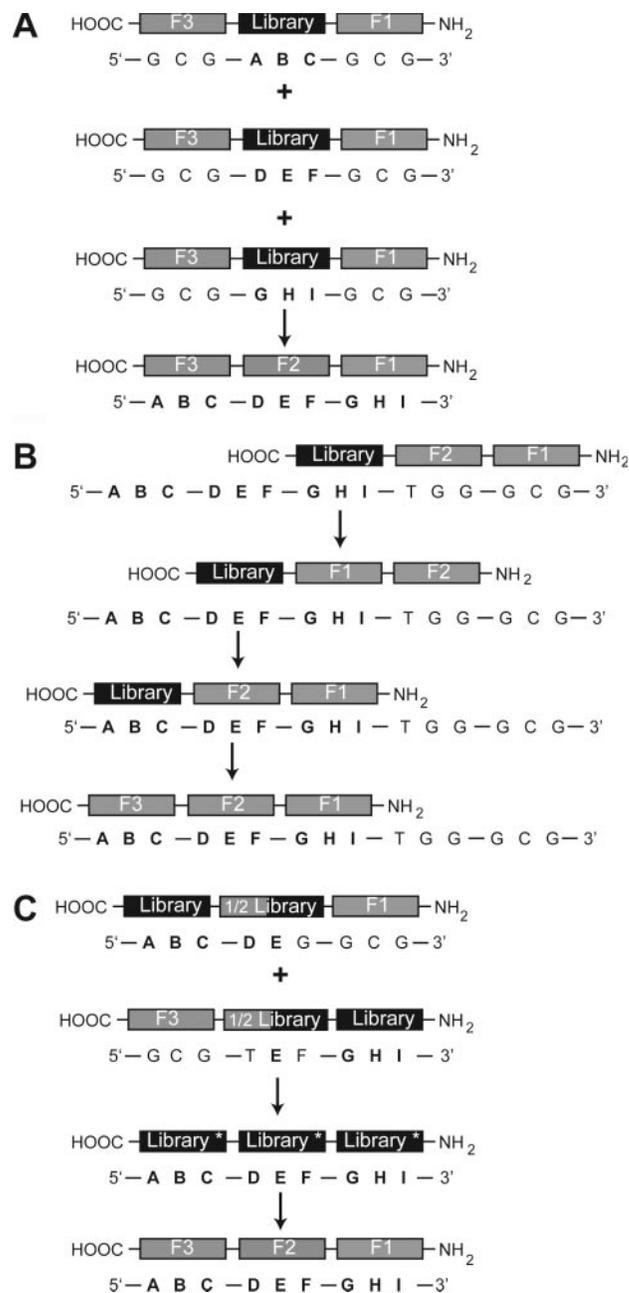
#### Selection strategies for ZFPs

The phage display selection method is commonly used to identify three-finger ZFPs that bind to a desired target DNA site. Workers in the field have utilized three-finger Zif268-derived phage libraries for selection of ZFPs that recognize specific 9 bp sequences (44,45). Here, key amino acid residues (-1 to +6 position of the  $\alpha$ -helix) that could potentially contact the DNA in each of the individual ZF motifs within a three-finger ZFP are randomized to generate a mutant library from which the desired high affinity three-finger ZFPs that bind to the chosen target site are selected. Since randomization of all the seven potential DNA contact positions of individual ZF motifs even in a three-finger ZFP results in an unmanageable number of possible mutants, only partial libraries are generally created by restricting randomizations to a few key residues within the DNA contact region of the  $\alpha$ -helix in each ZF motif of a ZFP. Furthermore, the libraries are limited by the maximal experimental transfection efficiencies (which are in the order of 10<sup>9-11</sup>) that are achievable using bacterial cells.

Several studies using Zif268-derived phage libraries for selection of three-finger ZFPs have been reported in literature (44,46,47). Although they yield the desired high affinity ZFPs, these experiments are too laborious and too cumbersome to perform and require special expertise in library construction and affinity selection using the phage display methodology. Furthermore, phage display selection of Zif268-derived libraries is constrained for two reasons: (i) The target site overlap problem posed by the presence of Asp<sup>2</sup> at position 2 of the  $\alpha$ -helix of each finger of Zif268 protein (Figure 1C); this makes selection of ZFPs using phage display libraries more difficult. (ii) The variability of the backbone residues between each of the three ZF motifs of Zif268 results in a variety of non-standard docking arrangements giving rise to different side chain-base contacts. The diversity of these contacts places obvious limits on a simple modular recognition code for the ZF motifs within the ZFPs.

As a result, research using phage display selection methods has taken two different routes. The first is based on a 'parallel' selection strategy, which is based on functional independence of individual ZF motifs in the ZFPs, i.e. each individual finger in a ZFP binds to its recognition site independent of its neighboring finger (Figure 3A). The main goal of this strategy was to find ZF motifs that bind specifically to each of the 64 possible DNA triplets, which then could be used in the design strategy for assembling the desired ZFPs by mixing and matching the various ZF motif designs as discussed above. This parallel strategy was useful in developing ZF motif designs that are specific for the 5'-GNN-3' (40,46) and 5'-ANN-3' (38) triplets. However, ZFPs assembled using these known ZF motif designs do not always have the desired sequence-specificity and affinity because not all of the available ZF designs bind to their cognate DNA triplets in a highly sequence-specific manner; they also bind degenerated sites (Figure 1). Hence, the designed ZFPs often require further optimization to obtain those with the desired sequence-specificity and high affinity.

The second approach is based on a 'sequential' selection strategy, where individual fingers are selected in the context of its neighbors, and thereby, circumvent the constraints placed by the target site overlap problem (Figure 3B) (44). However, this requires construction of multiple ZFP libraries and multiple selections for each and every ZFP that is desired, which is a tedious and time-consuming effort. A variation of the 'sequential' selection theme is the 'bipartite' strategy, which utilizes two pre-generated ZFP libraries in each of which one-and-a half fingers of a three-finger ZFP are partially randomized at the key residues that make contact with DNA (Figure 3C) (47,48). The N-terminal part of the ZFP is randomized in one library while the C-terminal half is randomized in the other. Selection is done in parallel using the 5' and 3' halves, respectively of the target sequence. Selections from the individual libraries are then recombined and selected again using the full target site to obtain the ZFP with the desired specificity and affinity. Although they yield the desired high affinity ZFPs, all these selection approaches are very labor intensive and too cumbersome to perform routinely and require special expertise. Furthermore, selection approaches become untenable with increasing number of ZF motifs within the ZFPs, since this will result in an exponential increase in the



**Figure 3.** Strategies for the production of ZFPs with desired DNA-binding specificity [adapted from reference (14)]. (A) Parallel selection; (B) Sequential selection; and (C) Bipartite selection. Parallel selection (A) is based on functional independence of individual ZF motifs in the ZFPs, i.e. each individual finger in a ZFP binds to its recognition site independent of its neighboring finger. In 'sequential' selection strategy (B), individual fingers are selected in the context of its neighbors, and thereby, circumvent the constraints placed by the target site overlap problem (Figure 1). However, this requires construction of multiple ZFP libraries and multiple selections for each and every ZFP that is desired, which is a tedious and time-consuming effort. The 'bipartite' strategy (C) utilizes two pre-generated ZFP libraries in each of which one-and-a half fingers of a three-finger ZFP are partially randomized at the key residues that make contact with DNA. The N-terminal part of the ZFP is randomized in one library while the C-terminal half is randomized in the other. Selection is done in parallel using the 5' and 3' halves, respectively of the target sequence. Selections from the individual libraries are then recombined and selected again using the full target site to obtain the ZFP with the desired specificity and affinity. Although they yield the desired high affinity ZFPs, all these selection approaches are very labor intensive and too cumbersome to perform routinely and require special expertise. Asterisks indicate pre-selected libraries.

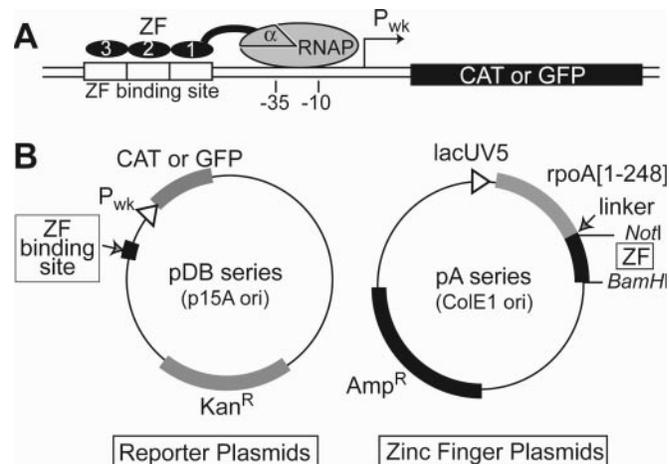
ZFP library members, all of whom cannot be sampled in a single phage display selection experiment.

### ZFP selection strategies in an *in vivo* context

The phage display method discussed above requires several rounds of ZFP selection and these selections do not occur in an *in vivo* setting. Alternate systems with a number of advantages exist including a bacterial two-hybrid system developed by Joung *et al.* (49) in which a desired ZFP–DNA interaction was linked to transcriptional activation of the yeast *HIS3* gene. These systems have the advantage that ZFP with the desired affinity can be identified in a single round (instead of multiple rounds for phage display) and that the affinity is selected for in an *in vivo* setting. As in all genetic reporter systems, the correlation of protein activity (i.e. binding affinity of a ZFP) of a particular clone to the transcription of a reporter gene must be done under the assumption that the protein concentration in the cell does not vary greatly from clone to clone.

The bacterial two-hybrid system developed by Joung *et al.* (49,50) links ZFP-binding to expression of the yeast *HIS3* gene that can complement a defect in growth in *Escherichia coli* cells bearing a deletion of the *hisB* gene. In this system, the ZFPs are fused to the yeast Gal11P protein. Gal11P has affinity for the yeast Gal4 protein. Gal4 is fused to a fragment of the  $\alpha$ -subunit of *E. coli* RNA polymerase (RpoA[1–248]). The ZFP–Gal11P and Gal4–RpoA[1–248] fusions are encoded on separate, compatible plasmids. The desired ZFP-binding site is positioned on the episome at position –63 in a promoter upstream from the *HIS3* gene. Thus, binding of the ZFP–Gal11P fusion to this binding site recruits the Gal4–RpoA[1–248] fusion which in turn recruits the rest of the RNA polymerase. The stimulation of *HIS3* transcription that results allows for growth of a *hisB* strain on selective minimal media plates. The stringency of the system can be increased by the addition of the *HIS3* competitive inhibitor 3-aminotriazole to the plates to select for ZFPs with very high affinity. This system has many advantages over phage display including: (i) single-step isolation of candidates in an *in vivo* context from libraries  $>10^8$  in size and (ii) the bypass of complications associated with the export of proteins to the cell membranes. This bacterial two-hybrid system, combined with a domain shuffling strategy, has been successfully used to select ZFPs with sub 100 pM dissociation constants for their desired target (50).

More recently, we have developed bacterial one-hybrid systems for selecting the desired ZFP and testing its sequence-specificity *in vivo* (51). This system was derived from a bacterial one-hybrid system developed by Hochschild *et al.* (52,53) that used *lacZ* as a reporter gene in a genetic screen (Figure 4A). In our system, we have two different reporter genes chloramphenicol acetyltransferase ( $\text{Cm}^R$ ) and green fluorescent protein (GFP) (Figure 4B). Our system utilizes two plasmids. On the first, a desired ZFP [in our case a ZFP named  $\Delta\text{QNK}$ ; for more details see (54)] is fused to a fragment of the  $\alpha$ -subunit of RNA polymerase (RpoA– $\Delta\text{QNK}$ ) via a 22 amino acid linker whose length has been optimized using incremental truncation. On the second plasmid (reporter plasmid), the binding site for the ZFP is centered in a promoter at position –55 relative to the start of transcription (the position was also optimized using incremental truncation) upstream either from the genes for  $\text{Cm}^R$  or GFP. Binding



**Figure 4.** One-hybrid system for detection of ZFP–DNA interactions. (A) Schematic of genetic selection system for interrogating ZFP–DNA interactions. (B) Plasmids for one-hybrid genetic selection system. The reporter gene, either chloramphenicol acetyltransferase (CAT), or GFP is located downstream from a weak *lac* derivative promoter ( $P_{wk}$ ) on pDB series plasmids. A 9 bp target site for binding by the ZFP is located at a specific distance from the start of transcription. On the pA series of plasmids, the gene for the ZFP is fused to a fragment of the  $\alpha$ -subunit of RNA polymerase (*rpoA*[1–248]) via a sequence coding for an amino acid linker. Binding of the RpoA[1–248]–ZFP fusion to the 9 bp site in the reporter plasmid recruits the other RNA polymerase subunits to stimulate transcription of the reporter gene.

of the ZFP domain of RpoA– $\Delta\text{QNK}$  to this binding site results in a 10-fold increase in chloramphenicol resistance or a 8-fold increase in green fluorescence (51). Single base changes in the  $\Delta\text{QNK}$ -binding site (sites to which the ZFP does not bind) eliminate the transcriptional stimulation. Both systems have been used to select for ZFPs that selectively bind the motif 5′-GGGGCAGAA-3′ from a library of 160 000 variants (on the protein level) in which the amino acids at positions –1, 2, 3 and 6 of the middle finger were randomized. The results obtained were consistent with those obtained from the phage display methodology. Since transcription levels may or may not reflect affinities of the ZFPs, we converted four of the selected ZFPs namely NDDR, QETR, QSNK and QSTK (where the capital letters represent the amino acid at positions –1, 2, 3, and 6 of the  $\alpha$ -helix, respectively) into ZFNs and tested them for sequence-specific cleavage of the  $\Delta\text{QNK}$  substrate. NDDR-FN and QETR-FN cut the plasmid substrate with the inverted  $\Delta\text{QNK}$  sites as well as  $\Delta\text{QNK}$ -FN, while QSNK-FN and QSTK-FN, as expected, were less active (S. Durai, M. Mani and S. Chandrasegaran, unpublished data). These results are consistent with their observed activity in our one-hybrid green fluorescence reporter system. This suggests that the one-hybrid single-reporter systems could be used as an easier to perform system for interrogating ZFP–DNA interactions.

## ZFNS

### Domain structure of FokI endonuclease

FokI restriction enzyme, a bacterial type IIS restriction endonuclease recognizes the non-palindromic pentadeoxyribonucleotide, 5′-GGATG-3′: 5′-CATCC-3′, in duplex DNA and cleaves 9/13 nt downstream of the recognition site (55,56).

It does not recognize any specific-sequence at the site of cleavage. This implies the presence of two separate protein domains within FokI: one for sequence-specific recognition of DNA and the other for the endonuclease activity. Once the DNA-binding domain is anchored at the recognition site, a signal is transmitted to the endonuclease domain, probably through allosteric interactions, and the cleavage occurs. We reasoned that one may be able to swap the FokI recognition domain with other naturally occurring DNA-binding proteins that recognize longer DNA sequences or other designed DNA-binding motifs to create chimeric nucleases.

As a first step, we probed the domain structure of FokI restriction endonuclease by limited proteolysis using trypsin. Our studies revealed an N-terminal DNA-binding domain and a C-terminal domain with non-specific DNA cleavage activity (57,58). Waugh and Sauer (59,60) later showed that single amino acid substitutions decouple the DNA-binding and strand scission activities of FokI endonuclease. The mode of DNA-binding by FokI has been analyzed by DNA footprinting methods (61,62). These studies show a lack of protection at the cleavage site of DNA by FokI. Wah *et al.* (63,64) reported the crystal structures of FokI and FokI-bound to its cognate site, which confirmed the modular nature of FokI endonuclease. The endonuclease domain appears to be sequestered in a 'piggyback' fashion by the recognition domain by protein-protein interactions. This is consistent with the DNA footprinting analysis. Thus, the crystal structure of FokI endonuclease is in complete agreement with the model derived from rigorous biochemical studies. Later reports have shown that the release of the endonuclease domain from the recognition domain enables it to swing over the DNA cut site and that the cleavage occurs only when FokI is bound to its cognate site and only in the presence of magnesium ions. Further studies on the mechanism of cleavage by FokI have shown that the dimerization of the endonuclease domain is required for FokI to produce a DSB (65).

### Engineering of ZFNs

The modular nature of FokI endonuclease suggested that it might be feasible to engineer chimeric nucleases by fusing other DNA-binding proteins to the cleavage domain of FokI. This turned out to be true. We created several novel chimeric endonucleases by fusing the FokI cleavage domain to other DNA-binding proteins. The latter include the three common eukaryotic DNA-binding motifs namely the helix-turn-helix motif, the ZF motif and the basic helix-loop-helix protein containing a leucine zipper motif (bzip). The first chimeric endonuclease was created by linking the *Drosophila Ubx* homeodomain to the cleavage domain of FokI endonuclease (66). This was followed by the fusion of classical Cys<sub>2</sub>His<sub>2</sub> ZFPs and the N-terminal 147 amino acids of the yeast Gal4 protein (contains the bzip motif), respectively, to the cleavage domain of FokI (23,67). Such fusions were shown to make specific cuts *in vitro* very close to their expected recognition sequences. Since our first report, other labs have reported making FokI nuclease domain fusions with other DNA-binding proteins (68–71). FokI fusions have also been used to identify high and low affinity binding sites of transcription factors *in vitro* (72), to study recruitment of various factors to

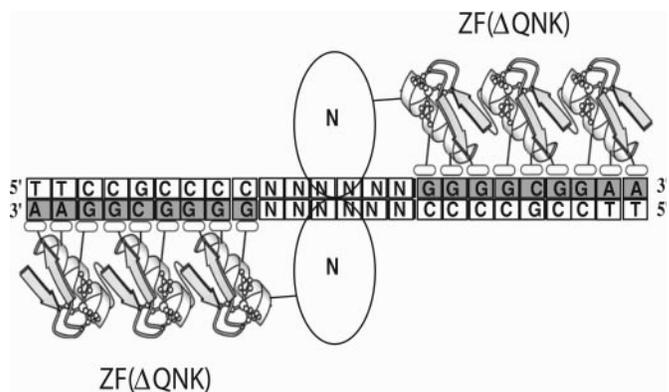
the promoter sites *in vivo* using the method called protein position identification with a nuclease tail (PIN\*POINT) assay (69,73–76) and to study Z-DNA conformation specific proteins (68,77).

The most important group of chimeric nucleases is the ZFNs. ZF-DNA-binding motifs, because of their modular structure and modular binding, as discussed above, offer an attractive framework for designing ZFNs with tailor-made sequence-specificities (1,2,78). Several three-ZFPs, each recognizing a 9 bp sequence, have been fused to the non-specific endonuclease domain of FokI to form ZFNs. The binding specificity of ZFPs has been shown to correlate directly with the cleavage specificity of the corresponding engineered ZFNs (25,54).

### Mechanism of double-strand DNA cleavage by ZFNs

Studies on the mechanism of double-strand cleavage by ZFN has shown that it requires dimerization of the nuclease domain (Figure 5) in order to cut the DNA substrate (24,79). ZFN dimerization, and hence double-strand cleavage, seems to be facilitated by two closely oriented inverted binding sites (that is the binding sites are positioned in an head to tail orientation on the top and bottom strands of the DNA substrate; Figure 5) (24). These results are consistent with the literature reports on the mechanism of DNA cleavage by FokI restriction enzyme. The first clue to FokI dimerization through its cleavage domain came from the elegant work of Neil Aggarwal's lab on the crystal structure of native FokI restriction endonuclease. Subsequent kinetic studies also indicated that dimerization of FokI restriction enzyme through its nuclease domain is essential in order to cut the double-stranded DNA substrate (63,65).

Since a three-finger ZFN requires two copies of the 9 bp recognition sites in a tail-to-tail orientation in order to dimerize and produce a DSB, it effectively has an 18 bp recognition site, which is long enough to specify a unique genomic address in plants and mammals. In principle, the binding sites need not be identical provided the ZFNs that bind both sites are present. We have shown that two ZFNs with different



**Figure 5.** A three-finger ZFN (called  $\Delta$ QNK-FN) bound to its target sequence. Since the three-finger ZFN requires two copies of the 9 bp recognition sites in an inverted orientation in order to dimerize and produce a DSB, it effectively has an 18 bp recognition site. The three-finger ZFP ( $\Delta$ QNK) was fused to the FokI cleavage domain (N or FN) to create the ZFN,  $\Delta$ QNK-FN (54).

sequence-specificities collaborate as a heterodimer to produce a DSB when their binding sites are appropriately placed and oriented with respect to each other (8,24). Since the recognition specificity of ZFPs can be manipulated experimentally, ZFN could be engineered to target a unique site within a plant or a mammalian genome.

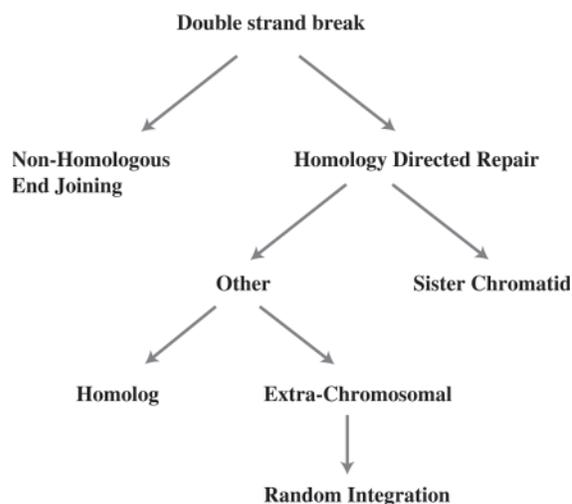
A comparison of the ZFN cleavage of a substrate containing two inverted ZFN sites and a substrate containing a single ZFN site has shown that the two inverted ZFN sites appear to facilitate DSB cleavage to a great extent (79). Furthermore, the two inverted ZFN sites are cleaved more efficiently when they are positioned closer to each other than further apart but a minimal spacer of 4–6 bp between the sites must be maintained for efficient cleavage (79).

Incidentally, this DNA cleavage mechanism also explains the observed exquisite sequence-specificity of ZFNs and their successful application to *in vivo* gene targeting experiments in frog oocytes (8), in *Drosophila* (7,80), plants (81) and human cells (9) that are discussed in detail in Section 5. Recombinant cells from these experiments would not have proven to be viable since statistically one could expect a large number of single sites to be available within the plant and mammalian genome for cleavage for each of the three-finger ZFNs used in these studies. We have speculated that the cytotoxicity observed as a result of continued over expression of a pair of three-finger ZFNs (each recognizing a 9 bp DNA sequence) in cells is probably due to ZFN cleavage at secondary degenerate sites positioned in an inverted orientation. We expect that improvements in the sequence-specificity and selectivity of the ZFNs will make them less toxic to cells. In the recent study by Urnov *et al.* (11), where use of four-finger ZFNs (expected to recognize a 24 bp DNA sequence) have been shown to promote highly sequence-specific cleavage in human cells while exhibiting decreased cytotoxicity. This suggests that one possible way to increase the sequence-specificity of ZFNs is to increase the number of ZF motifs within the ZFNs.

### Sequence-specificity and affinity of ZFNs

The sequence-specificity and affinity of the ZFNs are only as good as the ZFPs that are used to engineer them. This is because in many instances the individual ZF motifs within ZFPs usually make sequence-specific contacts with only two of the bases within the cognate triplet (15,44) (Figure 1C). The additional base specific cross-strand contact from the presence of Asp<sup>2</sup> at position +2 of the  $\alpha$ -helix of the neighboring finger that precedes the ZF motif increases the affinity and specificity of the ZF motif for its triplet subsites. If this is absent, then only two bases are generally recognized within the cognate DNA triplet, which more often than not, could result in ZF motifs recognizing other degenerate sites (Figure 1D). Because of this, even though a set of three-finger ZFP are expected to recognize an 18 bp target in theory, the actual recognition site is anywhere between 12 and 18 bp depending on the specificity of the chosen individual ZF designs for their cognate triplets. As discussed above, it appears that ZFNs could be engineered to be highly sequence-specific by adding more fingers to the three-finger ZFPs, thereby, making them recognize a larger target DNA sequence.

### Framework for DSB repair in mammalian cells



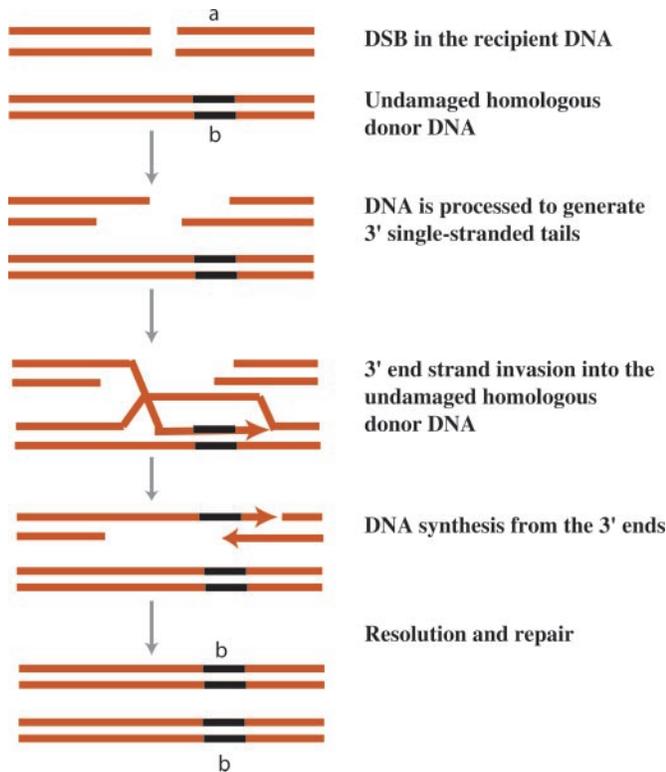
**Scheme 1.** Framework for DSB repair in mammalian cells.

### MECHANISM OF DOUBLE-STRAND BREAK REPAIR IN MAMMALIAN CELLS

The stimulatory nature of DSBs for gene targeting is related to how cells naturally repair DSBs. DSBs are naturally occurring events that have the potential to cause chromosomal rearrangements or cell death. But cells have redundant mechanisms to repair DSBs. The two primary mechanisms are by non-homologous end-joining (NHEJ) and HR (Scheme 1).

In NHEJ, the broken ends are simply ligated back together. If the DSB is ‘clean’ and the broken ends are compatible, repair of a DSB by NHEJ will usually be non-mutagenic. But occasionally, repair of a clean DSB by NHEJ will be mutagenic. If the DSB is complex—creating ends that are not compatible, then repair by NHEJ will necessarily be mutagenic. That is, the repaired DNA will have sequence differences, either the insertion or deletion of nucleotides, which were not present in the original segment.

The second major mechanism that cells repair a DSB is by HR. Conceptually repair of a DSB by HR is by a ‘copy and paste’ mechanism. Mechanistically, repair of a DSB proceeds in the following fashion and has been reviewed elsewhere (82). Here we briefly describe the synthesis-dependent strand-annealing (SDSA) model of DSB repair by HR (Figure 6). First, the ends of the DSB are resected to generate 3′ single-stranded tails. The generation of these tails requires the Mre11/Rad50/Nbs1 multi-protein complex but the exact nuclease that generates the 3′ tails remains to be identified. These 3′ tails are protected from nucleolytic degradation by coating with RPA. The next step in HR is strand invasion. In this step, Rad51 displaces RPA from the single-stranded DNA (in conjunction with Rad52 and BRCA2) and the Rad51/ssDNA nucleofilament invades an undamaged homologous segment of DNA. Then primed DNA synthesis occurs as the undamaged DNA serves as a template for DNA replication with the invading strand serving as the primer. After DNA synthesis the complex becomes unraveled and the DNA strands pair with their original partners. The newly synthesized DNA then serves as a



**Figure 6.** ZFN-induced homology-directed DSB repair in mammalian cells. Initially, gene 'a' is undamaged. The repair of the gene by HR occurs after the induction of a DSB. The DSB may arise spontaneously such as by damage from reactive oxygen species during normal metabolism, induced randomly by the exposure to ionizing radiation, or induced specifically by ZFNs. The DSB is then processed to form free 3' single-strand tails, a process that requires the Mre11/Rad50/Nbs1 complex. The HR machinery, through the actions of the strand invasion protein, Rad51, then uses the free 3' ends to invade a homologous repair template/donor. How the machinery identifies a homologous repair donor remains unclear but it is likely that simple physical proximity plays an important role. In the normal repair of a DSB, the repair donor is the sister-chromatid and thus the template is identical to the damaged allele. In gene targeting, the repair donor would be an extra-chromosomal piece of DNA that could have sequence differences. In this figure the letter 'b' and dark bars denotes those sequence differences. After strand invasion, primed DNA synthesis occurs to generate new undamaged DNA using the undamaged DNA as a template. The process is completed by the annealing of the new strand of DNA with its original partner and subsequent use of that new DNA to template DNA synthesis. Through this process, original allele ('a') is converted into the repair allele ('b') while the undamaged allele is unchanged. This donation of information from the undamaged allele to the damaged allele is called gene conversion.

template for DNA replication for its partner strand with the result that a complete undamaged double-strand piece of DNA is formed. The DNA used as the template is unperturbed and this is why the mechanism can be conceptualized as 'copy and paste' rather than 'cut and paste.'

The fidelity of the repair by HR depends on its choice of homologous DNA to use as a repair template. Usually, the cell uses the sister-chromatid as a template. Since the sister-chromatid is identical to the damaged chromosomal DNA using it as a template will result in a perfect form of repair, no matter how complex the original DNA break was. Rarely, however, the cell will use an alternative piece of homologous DNA as the repair template—either the homolog chromosome or some other random piece of homologous DNA. If there

are sequence differences between the homologous template and the damaged DNA, then those differences will be transferred to the damaged allele. For example, loss of heterozygosity can occur if the chromosomal homolog is used as a repair template. Finally, if the cell uses an introduced extra-chromosomal segment of DNA as the repair template, then gene targeting will result. Thus, gene targeting is the result of the cell using its own endogenous repair mechanism to repair a DSB.

### Random integration versus gene targeting

In gene targeting, an exogenous piece of DNA is introduced into the cell. An important aspect to understanding gene targeting is to consider what happens to that piece of DNA. There are four general outcomes: (i) It can be degraded in which case the introduction will be of no consequence in the long-term. (ii) It can be maintained as an episomal fragment. Unless it contains a mammalian origin of replication, such as that occurs on some Epstein-Barr virus (EBV)-based vectors, however, the piece of DNA will eventually be lost as the cell divides. On the other hand, if the cell is not dividing, then the fragment of DNA can be maintained episomally for long periods of time. This seems to be a frequent occurrence, e.g. when hepatocytes are infected with adeno-associated virus. (iii) It can be used as a template for the repair of a DSB by HR and subsequently lost. If this occurs, then gene targeting occurs as sequence information from the introduced DNA fragment is transferred to a homologous segment of the host genome. (iv) It can integrate randomly into the genome. In random integration, it is not just information that is transferred to the host genome, but the actual DNA fragment is integrated into the host genome. But because it is integrating in a non-homologous fashion, the integration of the DNA is fundamentally a mutagenic event. Surprisingly little is known about the mechanism of random integration in mammalian cells. Interestingly, random integration can occur at sites of DSBs—it appears that the cell uses the piece of DNA to 'patch' a break. When this occurs, there is no surrounding sequence homology to suggest that the integration was through an HR mechanism. Certain types of random integration also appear to be dependent on NHEJ. One of the appeals of gene targeting is that it is a precise way to manipulate the genome. In thinking about how to utilize gene targeting most effectively, one must not only think about how to increase the frequency that the extra-chromosomal DNA is used as a repair template for the repair of a DSB by HR, but also about how to decrease the frequency of the extra-chromosomal DNA integrating randomly and thereby causing undesired mutations.

### ZFN-MEDIATED GENE TARGETING

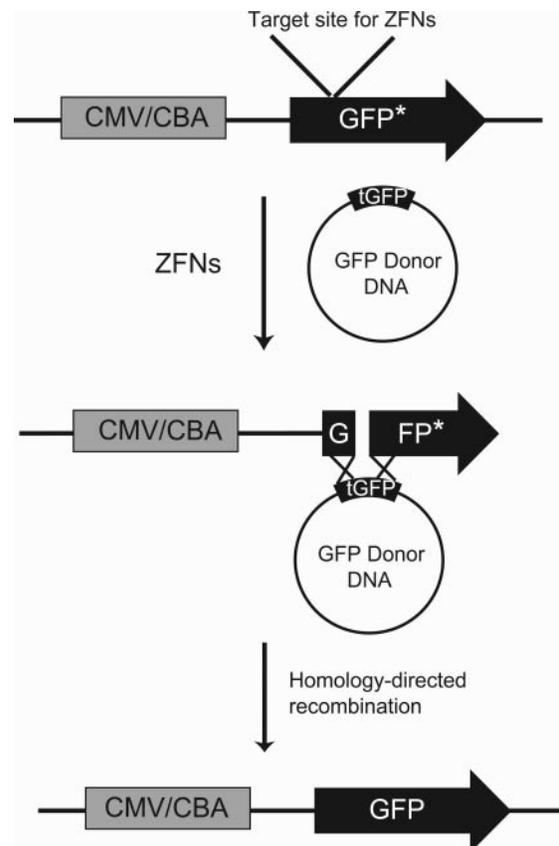
#### Homology-directed repair by HR versus mutagenic repair by NHEJ

We showed stimulation of HR through targeted cleavage using ZFNs in frog oocytes via homology-directed repair. In a proof-of-principle experiment, we have shown in collaboration with Dana Carroll's lab that the three-finger ZFN ( $\Delta$ QQR-FN) finds and cleaves its target *in vivo* (8). This was tested by microinjection of DNA substrates and the enzyme into frog eggs.

Double-strand cleavage required an inverted repeat of the 9 bp target. When the appropriate sites were placed in the recombination substrate, this DNA was cleaved in frog oocytes by the injected enzyme and HR ensued. These microinjection experiments have shown that >90% of the substrate was cleaved *in vivo* and almost all cleaved substrates underwent a specific-type of HR by a mechanism called 'single-strand annealing'. Although, in this proof-of-principle experiment, an extra-chromosomal target was used to test the idea, two important conclusions can be drawn from this study. First, ZFNs find and cleave their targets within cells. Second, the targeted DSB stimulates HR at the site of cleavage for homology-directed repair of DSB.

The obvious next step was to show site-specific cleavage of a chromosomal target within cells by ZFNs. Just this was done by Dana Carroll's lab, which reported targeted chromosomal cleavage and mutagenesis in *Drosophila melanogaster* (80). A pair of ZFPs that targeted the yellow (*Y*) gene on the X-chromosome of *Drosophila* were designed and then converted into ZFNs. The nucleases were cloned under the control of a heat shock promoter and then introduced into *Drosophila* genome via *P* element-mediated transformation. Only when both ZFNs were expressed in the developing larvae did somatic mutations occur specifically in the *Y* gene. The yellow mosaics were observed in 46% of males expressing both ZFNs. No yellow mosaics were observed in control larvae that expressed a single ZFN or without heat-shock. Germ line *Y* mutations were recovered from 5.7% of males but none from the females tested. This was as expected in female fruit flies because the targeted DSB would be repaired by recombination with the uncut *Y* homolog. In males, only simple ligation via NHEJ is available to repair the chromosomal damage. DNA sequence analysis of the mutants revealed that all mutations were small deletions and/or insertions localized to the targeted site, which was typical of repair by NHEJ (80–83). Many of these mutations led to frameshifts or deletion of essential codons of the *Y* gene. As a result, the male larvae emerging from the heat shock experiment showed yellow patches in the otherwise dark posterior abdomen. In subsequent experiments, Carroll's lab has recovered *y* mutations in female fruit flies, but always at a much lower frequency than that in male flies. This suggests that if the cleavage efficiency of ZFN was high enough to cut both homologs of the *Y* gene simultaneously, one can produce female germline mutations. Since accurate repair from the homolog appears not to be fully efficient, it might be possible to induce germline mutations in autosomal recessive genes using ZFNs (80). Furthermore, since DSB stimulates mutagenic repair that essentially operates in all organisms, targeted cleavage by ZFN could facilitate generation of directed mutations in a variety of cells and organisms. Bibikova *et al.* (7) then went on to show that gene targeting is enhanced using designed ZFNs in *Drosophila* by ~10-fold via homology-directed repair over those observed without the targeted cleavage (7). These logical experiments, which extended the proof-of-principle frog oocyte studies, demonstrated that the ZFNs could find and cleave a target site embedded within a chromosome and that ZFN-mediated genomic DSB induces HR using an extra-chromosomal DNA fragment for repair.

Simultaneously, Porteus and Baltimore (9) described a system based on the correction of a mutated GFP gene to study



**Figure 7.** Diagram of the GFP gene targeting reporter system. ZFN-mediated gene targeting results in the conversion of GFP negative cells into GFP positive cells, which is detected using flow cytometry. GFP\*, mutated GFP gene; tGFP, truncated GFP gene.

gene targeting in human somatic cells (Figure 7). In this assay, targeting is measured by the conversion of GFP negative cells to GFP positive cells. Since this conversion can be measured by flow cytometry, millions of cells can be analyzed in a short period of time. Using this system, they have shown that gene targeting using three-finger ZFNs is stimulated over 2000-fold by targeted chromosomal cleavage. An important finding from their work is that continued over expression of the three-finger ZFNs in human cells was cytotoxic; as much as 75% of the targeted cells were lost due to cytotoxicity.

More recently, Urnov *et al.* (11) have shown that ZFNs could be made less cytotoxic by engineering highly sequence-specific ZFNs. They assembled two four-finger ZFNs, which together recognize and cut a 24 bp site in the IL2R $\gamma$  gene. ZFNs were further optimized for their sequence-specific cleavage by tinkering with individual ZF motifs within the ZFPs and testing them in cells for ZFN-mediated correction of a mutated GFP gene as discussed above (9). These were then used for ZFN-mediated gene targeting in human cells; they achieved highly efficient and permanent modification of the IL2R $\gamma$  gene. A remarkable gene modification efficiency of 18% of treated cells was obtained without selection, approximately one-third of which were altered on both X-chromosomes. By using a set of four-finger ZFNs the target recognition was enlarged from 18 to 24 bp. This along with further optimization at the level of individual ZF motifs

of the ZFPs as expected yielded ZFNs with high sequence-specificity and selectivity that were less toxic to cells. Thus, increasing of the sequence-specificity of ZFNs appears to greatly reduce their cytotoxicity. Thus, these studies have established that ZFN-mediated gene targeting as a powerful research tool that could be used to generate 'targeted mutation' by NHEJ, and 'gene correction' by HR in cells and organisms.

In summary, a number of conclusions can be made about the applicability of ZFNs to gene targeting from these studies. A simple assembly approach is most likely to result in useful ZFNs if the target sequence fits the 5'-GNNGNNGNN-3' consensus. A recent report by Alwin *et al.* (84) shows that it is possible to create the assembly approach by an active ZFN to a target sequence that contained a mixture of ANN and GNN triplets; but it remains to be seen how easy it is to target such sequences in general. With the recent publication by Dreier *et al.* (41) of the ZF motif preferences for the CNN triplets, the repertoire of potential binding sequences may increase still further. But it still remains to be determined whether active ZFNs can be made by mixing CNN fingers with GNN and/or ANN fingers since the ZF designs for ANN and CNN triplets are not as well tested and established experimentally as those of the GNN triplets.

ZFNs made by simple assembly strategy are likely to be useful for experimental research purposes, so if the investigator desires to use them for such purposes, the assembly approach may be good enough. A ZFP design web site will soon be available at Carlos Barbas laboratory (<http://www.scripps.edu/mb/barbas/zfdesign/zfdesignhome.php>). This program will identify potential ZFP-binding sites within any target sequence using the collection of ZFPs that the Barbas lab has generated over the years. But it appears that the assembled ZFPs in general have less specificity and affinity than those made by selection strategies. Therefore, the ZFNs for human therapeutics would likely have to be at least in part derived from a selection strategy. Currently, selection strategies for ZFPs require a certain degree of expertise but we hope that other strategies will be developed to perform these experiments with ease and thereby broaden the application of ZFN-mediated gene targeting to various biological and biomedical applications.

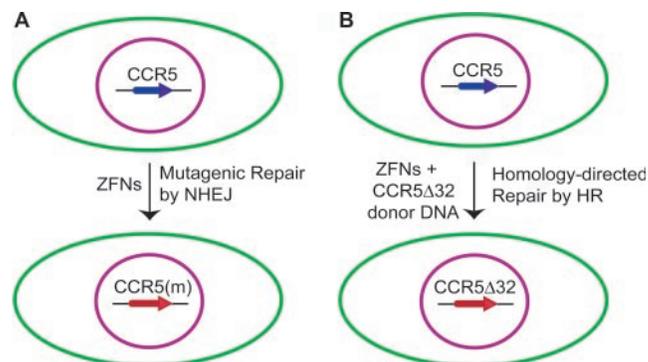
The published data appears to suggest that four-finger ZFNs may well be better than three-finger ZFNs. However, such a comparison is made difficult by the fact that the four-finger ZFNs were made using selection strategy followed by assembly while a simple assembly approach was used to make the three-finger ZFNs. Furthermore, at this juncture the experimental comparison of ZFN-mediated gene targeting between 3, 4, 5 and even 6-finger ZFNs remains quite limited by the fact that there is no published data regarding the use of either 5- or 6-finger ZFNs for gene targeting. We hope that this will be rectified in the near future and direct comparisons of such ZFNs for both their specificity of targeting and their lack of cytotoxic effects will determine which number of fingers within a ZFN is optimal. Just as determining which strategy to use for the design of ZFPs (assembly versus selection), it may also be that the optimal number of fingers may be determined by the ultimate use of the ZFN. It may be that one needs to invest the additional energy into making specific 4, 5 or even 6-finger ZFNs only if the goal is to use them for treatment of human diseases.

Given the current status of ZFN technology, it would be reasonable to expect that an investigator could use ZFNs to facilitate targeting their gene of interest if the gene contained the following consensus sequence 5'-NNCNNC(N)<sub>4-6</sub>GNNGNNGNN-3'. Such a sequence would be expected to occur approximately once every 1000–1500 bp. Efficient gene targeting occurs best if the target sequence is <100 bp from the site of the DSB but they can also occur at much greater distances (85–87). This ability to target at a distance from the site of the ZFNs further increases the probability that this technology can be used to target a specific gene within a plant or a mammalian genome including the human genome. So far, the technology has remained in the hands of a few selected labs because of the difficulty associated with generating highly sequence-specific ZFNs (87). Because of the promise of ZFN technology, scientists are working hard to refine and improve the design and selection strategies for creating highly sequence-specific ZFNs that are needed for various biomedical applications.

## FUTURE OUTLOOK

ZFNs are becoming powerful research tools for highly efficient and permanent site-specific modification of various types of cells, organisms, plants and animals. It must be emphasized that the sequence-specificity and affinity of the engineered ZFNs are only as good as the ZFPs that are used to construct them. While designed three-finger ZFNs may be sufficient to achieve targeted genome engineering of plants and animal cells in most cases, optimized four-finger ZFNs with higher sequence-specificity and affinity would likely yield even better results. Optimized four-finger proteins appear to have both higher rates of targeting and less cytotoxicity in mammalian cells. Coupling ZFN-mediated gene targeting to a positive-negative selection protocol (3) might lead to further improvements in gene modification efficiency.

There is also excitement in the scientific community that ZFN-based strategies will make targeted correction of a genetic defect feasible in the future, especially in treating monogenic diseases. Several areas of research appear to be converging and coalescing to make gene therapy a reality.



**Figure 8.** ZFN-mediated directed mutagenesis in human cells. (A) Targeted disruption of the CCR5 gene by NHEJ (mutagenic repair) using engineered ZFNs. Cells are transfected with ZFNs alone. CCR5 (m) depicts mutant CCR5 gene. (B) Targeted disruption of the CCR5 gene by ZFN-induced homology-directed repair. In this experiment, cells are transfected with both ZFN and CCR5Δ32 (or mutant CCR5 DNA).

The complete nucleotide sequence of the human genome is now available. Also, rapid progress is being made in stem cell research. The first applications of ZFN-based strategies as a form of gene therapy to treat human diseases will likely occur in *ex-vivo* gene therapy using stem cells. Here, the desired cells can be identified through selection, expanded in culture and replenished into patients. The initial gene targets in human stem cells will likely include IL2R $\gamma$  gene and  $\beta$ -globin gene for ZFN-mediated gene correction and CCR5 gene for ZFN-mediated directed mutagenesis (Figure 8) (1,2,11). It must be emphasized that ZFN-based strategies as a form of gene therapy is still at its infancy. Several issues, like efficient gene delivery into the targeted cells and immune response to ZFNs, etc. still remain that need to be addressed very systematically and carefully before ZFNs can be considered for human therapeutics. Since the ZFNs will only be expressed transiently, it decreases the probability that they will invoke an undesired immune response but this remains to be tested. These are discussed in detail elsewhere (2). We expect that custom-designed ZFNs, the new type of molecular scissors that are engineered to target a unique site within the human genome, will contribute and greatly aid the feasibility of targeted and site-specific engineering of the human genome in the future. Ethical issues aside, we anticipate that over the next decade or so the technical problems associated with gene delivery will be overcome and that gene therapy will be routinely used in a clinical setting. This will finally signify a paradigm shift in the treatment of human diseases.

## ACKNOWLEDGEMENTS

This work was funded by a grant from National Institutes of Health (GM 53923). We also thank Environmental Health Sciences Center Core Facility (supported by grant ES 03819) for synthesis of oligonucleotides. K.K. was supported for a year by the Johns Hopkins University Center for AIDS Research (CFAR; Grant # P30 AI42855) grant. Funding to pay the Open Access publication charges for this article was provided by JHU.

*Conflict of interest statement.* None declared.

## REFERENCES

- Kandavelou, K., Mani, M., Durai, S. and Chandrasegaran, S. (2004) In Pingoud, A.M. (ed.), *Nucleic Acids and Molecular Biology*. Springer Verlag, Berlin Vol. 14, pp. 413–434.
- Kandavelou, K., Mani, M., Durai, S. and Chandrasegaran, S. (2005) 'Magic' scissors for genome surgery. *Nat. Biotechnol.*, **23**, 686–687.
- Capecchi, M.R. (1989) Altering the genome by homologous recombination. *Science*, **244**, 1288–1292.
- Capecchi, M.R. (1989) The new mouse genetics: altering the genome by gene targeting. *Trends Genet.*, **5**, 70–76.
- Jasin, M. (1996) Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet.*, **12**, 224–228.
- Vasquez, K.M., Marburger, K., Intody, Z. and Wilson, J.H. (2001) Manipulating the mammalian genome by homologous recombination. *Proc. Natl Acad. Sci. USA*, **98**, 8403–8410.
- Bibikova, M., Beumer, K., Trautman, J.K. and Carroll, D. (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science*, **300**, 764.
- Bibikova, M., Carroll, D., Segal, D.J., Trautman, J.K., Smith, J., Kim, Y.G. and Chandrasegaran, S. (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol. Cell. Biol.*, **21**, 289–297.
- Porteus, M.H. and Baltimore, D. (2003) Chimeric nucleases stimulate gene targeting in human cells. *Science*, **300**, 763.
- Wilson, J.H. (2003) Pointing fingers at the limiting step in gene targeting. *Nat. Biotechnol.*, **21**, 759–760.
- Urnov, F.D., Miller, J.C., Lee, Y.L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D. and Holmes, M.C. (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, **435**, 646–651.
- High, K.A. (2005) Gene therapy: the moving finger. *Nature*, **435**, 577–579.
- Pabo, C.O., Peisach, E. and Grant, R.A. (2001) Design and selection of novel Cys2His2 zinc finger proteins. *Annu. Rev. Biochem.*, **70**, 313–340.
- Beerli, R.R. and Barbas, C.F., III (2002) Engineering polydactyl zinc-finger transcription factors. *Nat. Biotechnol.*, **20**, 135–141.
- Pavletich, N.P. and Pabo, C.O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268–DNA complex at 2.1 Å. *Science*, **252**, 809–817.
- Shi, Y. and Berg, J.M. (1995) A direct comparison of the properties of natural and designed zinc-finger proteins. *Chem. Biol.*, **2**, 83–89.
- Elrod-Erickson, M. and Pabo, C.O. (1999) Binding studies with mutants of Zif268. Contribution of individual side chains to binding affinity and specificity in the Zif268 zinc finger–DNA complex. *J. Biol. Chem.*, **274**, 19281–19285.
- Desjarlais, J.R. and Berg, J.M. (1993) Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA-binding proteins. *Proc. Natl Acad. Sci. USA*, **90**, 2256–2260.
- Wolfe, S.A., Greisman, H.A., Ramm, E.I. and Pabo, C.O. (1999) Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code. *J. Mol. Biol.*, **285**, 1917–1934.
- Liu, Q., Segal, D.J., Ghiara, J.B. and Barbas, C.F., III (1997) Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc. Natl Acad. Sci. USA*, **94**, 5525–5530.
- Beerli, R.R., Segal, D.J., Dreier, B. and Barbas, C.F., III (1998) Toward controlling gene expression at will: specific regulation of the *erbB-2/HER-2* promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc. Natl Acad. Sci. USA*, **95**, 14628–14633.
- Kim, J.S. and Pabo, C.O. (1998) Getting a handhold on DNA: design of poly-zinc finger proteins with femtomolar dissociation constants. *Proc. Natl Acad. Sci. USA*, **95**, 2812–2817.
- Kim, Y.G., Cha, J. and Chandrasegaran, S. (1996) Hybrid restriction enzymes: zinc finger fusions to FokI cleavage domain. *Proc. Natl Acad. Sci. USA*, **93**, 1156–1160.
- Smith, J., Bibikova, M., Whitby, F.G., Reddy, A.R., Chandrasegaran, S. and Carroll, D. (2000) Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res.*, **28**, 3361–3369.
- Mani, M., Kandavelou, K., Fei, J.D., Durai, S. and Chandrasegaran, S. (2005) Design, engineering and characterization of zinc finger nucleases. *Biochem. Biophys. Res. Commun.*, **335**, 447–457.
- Alwin, S., Gere, M.B., Guhl, E., Effertz, K., Barbas, C.F., III, Segal, D.J., Weitzman, M.D. and Cathomen, T. (2005) Custom zinc-finger nucleases for use in human cells. *Mol. Ther.*, **12**, 610–617.
- Liu, P.Q., Rebar, E.J., Zhang, L., Liu, Q., Jamieson, A.C., Liang, Y., Qi, H., Li, P.X., Chen, B., Mendel, M.C. *et al.* (2001) Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. *J. Biol. Chem.*, **276**, 11323–11334.
- Rebar, E.J., Huang, Y., Hickey, R., Nath, A.K., Meoli, D., Nath, S., Chen, B., Xu, L., Liang, Y., Jamieson, A.C. *et al.* (2002) Induction of angiogenesis in a mouse model using engineered transcription factors. *Nature Med.*, **8**, 1427–1432.
- Zhang, L., Spratt, S.K., Liu, Q., Johnstone, B., Qi, H., Raschke, E.E., Jamieson, A.C., Rebar, E.J., Wolffe, A.P. and Case, C.C. (2000) Synthetic zinc finger transcription factor action at an endogenous chromosomal site. Activation of the human erythropoietin gene. *J. Biol. Chem.*, **275**, 33850–33860.
- Bartsevich, V.V., Miller, J.C., Case, C.C. and Pabo, C.O. (2003) Engineered zinc finger proteins for controlling stem cell fate. *Stem Cells*, **21**, 632–637.

31. Dai,Q., Huang,J., Klitzman,B., Dong,C., Goldschmidt-Clermont,P.J., March,K.L., Rokovich,J., Johnstone,B., Rebar,E.J., Spratt,S.K. *et al.* (2004) Engineered zinc finger-activating vascular endothelial growth factor transcription factor plasmid DNA induces therapeutic angiogenesis in rabbits with hindlimb ischemia. *Circulation*, **110**, 2467–2475.
32. Rebar,E.J. (2004) Development of pro-angiogenic engineered transcription factors for the treatment of cardiovascular disease. *Expert Opin. Investig. Drugs*, **13**, 829–839.
33. Snowden,A.W., Zhang,L., Urnov,F., Dent,C., Jouvenot,Y., Zhong,X., Rebar,E.J., Jamieson,A.C., Zhang,H.S., Tan,S. *et al.* (2003) Repression of vascular endothelial growth factor A in glioblastoma cells using engineered zinc finger transcription factors. *Cancer Res.*, **63**, 8968–8976.
34. Bartsevich,V.V. and Juliano,R.L. (2000) Regulation of the MDR1 gene by transcriptional repressors selected using peptide combinatorial libraries. *Mol. Pharmacol.*, **58**, 1–10.
35. Ren,D., Collingwood,T.N., Rebar,E.J., Wolffe,A.P. and Camp,H.S. (2002) PPARgamma knockdown by engineered transcription factors: exogenous PPARgamma2 but not PPARgamma1 reactivates adipogenesis. *Genes Dev.*, **16**, 27–32.
36. Xu,G.L. and Bestor,T.H. (1997) Cytosine methylation targeted to pre-determined sequences. *Nature Genet.*, **17**, 376–378.
37. McNamara,A.R., Hurd,P.J., Smith,A.E. and Ford,K.G. (2002) Characterisation of site-biased DNA methyltransferases: specificity, affinity and subsite relationships. *Nucleic Acids Res.*, **30**, 3818–3830.
38. Dreier,B., Beerli,R.R., Segal,D.J., Flippin,J.D. and Barbas,C.F.,III (2001) 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.
39. Dreier,B., Segal,D.J. and Barbas,C.F.,III (2000) Insights into the molecular recognition of the 5'-GNN-3' family of DNA sequences by zinc finger domains. *J. Mol. Biol.*, **303**, 489–502.
40. Liu,Q., Xia,Z., Zhong,X. and Case,C.C. (2002) Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *J. Biol. Chem.*, **277**, 3850–3856.
41. Drier,B., Fuller,R.P., Segal,D.J., Lund,C., Blancafort,P., Huber,A., Kokschi,B. and Barbas,C.F.,III (2005) 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.* (in press, doi:10.1074/jbc.M506654200).
42. Jamieson,A.C., Miller,J.C. and Pabo,C.O. (2003) Drug discovery with engineered zinc-finger proteins. *Nature Rev. Drug Discov.*, **2**, 361–368.
43. Pavletich,N.P. and Pabo,C.O. (1993) Crystal structure of a five-finger GLI–DNA complex: new perspectives on zinc fingers. *Science*, **261**, 1701–1707.
44. Greisman,H.A. and Pabo,C.O. (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science*, **275**, 657–661.
45. Isalan,M., Klug,A. and Choo,Y. (1998) Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers. *Biochemistry*, **37**, 12026–12033.
46. Segal,D.J., Dreier,B., Beerli,R.R. and Barbas,C.F.,III (1999) 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. USA*, **96**, 2758–2763.
47. Isalan,M. and Choo,Y. (2001) Engineering nucleic acid-binding proteins by phage display. *Methods Mol. Biol.*, **148**, 417–429.
48. Isalan,M., Klug,A. and Choo,Y. (2001) A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nat. Biotechnol.*, **19**, 656–660.
49. Joung,J.K., Ramm,E.I. and Pabo,C.O. (2000) A bacterial two-hybrid selection system for studying protein–DNA and protein–protein interactions. *Proc. Natl Acad. Sci. USA*, **97**, 7382–7387.
50. Hurt,J.A., Thibodeau,S.A., Hirsh,A.S., Pabo,C.O. and Joung,J.K. (2003) Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc. Natl Acad. Sci. USA*, **100**, 12271–12276.
51. Durai,S., Bosley,A.D., Abulencia,A.B., Chandrasegaran,S. and Ostermeier,M. (2005) A bacterial one-hybrid selection system for interrogating zinc finger–DNA interactions. *Comb Chem High Throughput Screen.* (accepted).
52. Dove,S.L., Joung,J.K. and Hochschild,A. (1997) Activation of prokaryotic transcription through arbitrary protein–protein contacts. *Nature*, **386**, 627–630.
53. Hu,J.C., Kornacker,M.G. and Hochschild,A. (2000) *Escherichia coli* one- and two-hybrid systems for the analysis and identification of protein–protein interactions. *Methods*, **20**, 80–94.
54. Smith,J., Berg,J.M. and Chandrasegaran,S. (1999) A detailed study of the substrate specificity of a chimeric restriction enzyme. *Nucleic Acids Res.*, **27**, 674–681.
55. Sugisaki,H. and Kanazawa,S. (1981) New restriction endonucleases from *Flavobacterium okeanokoites* (FokI) and *Micrococcus luteus* (MluI). *Gene*, **16**, 73–78.
56. Szybalski,W., Kim,S.C., Hasan,N. and Podhajska,A.J. (1991) Class-IIS restriction enzymes—a review. *Gene*, **100**, 13–26.
57. Li,L. and Chandrasegaran,S. (1993) Alteration of the cleavage distance of FokI restriction endonuclease by insertion mutagenesis. *Proc. Natl Acad. Sci. USA*, **90**, 2764–2768.
58. Li,L., Wu,L.P. and Chandrasegaran,S. (1992) Functional domains in FokI restriction endonuclease. *Proc. Natl Acad. Sci. USA*, **89**, 4275–4279.
59. Waugh,D.S. and Sauer,R.T. (1993) Single amino acid substitutions uncouple the DNA-binding and strand scission activities of FokI endonuclease. *Proc. Natl Acad. Sci. USA*, **90**, 9596–9600.
60. Waugh,D.S. and Sauer,R.T. (1994) A novel class of FokI restriction endonuclease mutants that cleave hemi-methylated substrates. *J. Biol. Chem.*, **269**, 12298–12303.
61. Li,L., Wu,L.P., Clarke,R. and Chandrasegaran,S. (1993) C-terminal deletion mutants of the FokI restriction endonuclease. *Gene*, **133**, 79–84.
62. Yonezawa,A. and Sugiura,Y. (1994) DNA-binding mode of class-IIS restriction endonuclease FokI revealed by DNA footprinting analysis. *Biochim. Biophys. Acta.*, **1219**, 369–379.
63. Wah,D.A., Bitinaite,J., Schildkraut,I. and Aggarwal,A.K. (1998) Structure of FokI has implications for DNA cleavage. *Proc. Natl Acad. Sci. USA*, **95**, 10564–10569.
64. Wah,D.A., Hirsch,J.A., Dorner,L.F., Schildkraut,I. and Aggarwal,A.K. (1997) Structure of the multimodular endonuclease FokI bound to DNA. *Nature*, **388**, 97–100.
65. Bitinaite,J., Wah,D.A., Aggarwal,A.K. and Schildkraut,I. (1998) FokI dimerization is required for DNA cleavage. *Proc. Natl Acad. Sci. USA*, **95**, 10570–10575.
66. Kim,Y.G. and Chandrasegaran,S. (1994) Chimeric restriction endonuclease. *Proc. Natl Acad. Sci. USA*, **91**, 883–887.
67. Kim,Y.G., Smith,J., Durgasha,M. and Chandrasegaran,S. (1998) Chimeric restriction enzyme: Gal4 fusion to FokI cleavage domain. *Biol. Chem.*, **379**, 489–495.
68. Kim,Y.G., Kim,P.S., Herbert,A. and Rich,A. (1997) Construction of a Z-DNA-specific restriction endonuclease. *Proc. Natl Acad. Sci. USA*, **94**, 12875–12879.
69. Kim,Y.G., Shi,Y., Berg,J.M. and Chandrasegaran,S. (1997) Site-specific cleavage of DNA–RNA hybrids by zinc finger/FokI cleavage domain fusions. *Gene*, **203**, 43–49.
70. Huang,B., Schaeffer,C.J., Li,Q. and Tsai,M.D. (1996) Splase: a new class IIS zinc-finger restriction endonuclease with specificity for Sp1 binding sites. *J. Protein Chem.*, **15**, 481–489.
71. Lee,J.S., Lee,C.H. and Chung,J.H. (1998) Studying the recruitment of Sp1 to the beta-globin promoter with an *in vivo* method: protein position identification with nuclease tail (PIN\*POINT). *Proc. Natl Acad. Sci. USA*, **95**, 969–974.
72. Ruminy,P., Derambure,C., Chandrasegaran,S. and Salier,J.P. (2001) Long-range identification of hepatocyte nuclear factor-3 (FoxA) high and low-affinity binding sites with a chimeric nuclease. *J. Mol. Biol.*, **310**, 523–535.
73. Kim,M.K., Lee,J.S. and Chung,J.H. (1999) *In vivo* transcription factor recruitment during thyroid hormone receptor-mediated activation. *Proc. Natl Acad. Sci. USA*, **96**, 10092–10097.
74. Lee,J.S., Lee,C.H. and Chung,J.H. (1999) The beta-globin promoter is important for recruitment of erythroid Kruppel-like factor to the locus control region in erythroid cells. *Proc. Natl Acad. Sci. USA*, **96**, 10051–10055.
75. Lee,J.S., Ngo,H., Kim,D. and Chung,J.H. (2000) Erythroid Kruppel-like factor is recruited to the CACCC box in the beta-globin promoter but not to the CACCC box in the gamma-globin promoter: the role of the

- neighboring promoter elements. *Proc. Natl Acad. Sci. USA*, **97**, 2468–2473.
76. Lee, C.H., Murphy, M.R., Lee, J.S. and Chung, J.H. (1999) Targeting a SWI/SNF-related chromatin remodeling complex to the beta-globin promoter in erythroid cells. *Proc. Natl Acad. Sci. USA*, **96**, 12311–12315.
77. Kim, Y.G., Lowenhaupt, K., Schwartz, T. and Rich, A. (1999) The interaction between Z-DNA and the Zab domain of double-stranded RNA adenosine deaminase characterized using fusion nucleases. *J. Biol. Chem.*, **274**, 19081–19086.
78. Chandrasegaran, S. and Smith, J. (1999) Chimeric restriction enzymes: what is next? *Biol. Chem.*, **380**, 841–848.
79. Mani, M., Smith, J., Kandavelou, K., Berg, J.M. and Chandrasegaran, S. (2005) Binding of two zinc finger nuclease monomers to two specific sites is required for effective double-strand DNA cleavage. *Biochem. Biophys. Res. Commun.*, **334**, 1191–1197.
80. Bibikova, M., Golic, M., Golic, K.G. and Carroll, D. (2002) Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics*, **161**, 1169–1175.
81. Lloyd, A., Plaisier, C.L., Carroll, D. and Drews, G.N. (2005) Targeted mutagenesis using zinc-finger nucleases in *Arabidopsis*. *Proc. Natl Acad. Sci. USA*, **102**, 2232–2237.
82. West, S.C. (2003) Molecular views of recombination proteins and their control. *Nature Rev. Mol. Cell. Biol.*, **4**, 435–445.
83. Gloor, G.B., Moretti, J., Mouyal, J. and Keeler, K.J. (2000) Distinct *P*-element excision products in somatic and germline cells of *Drosophila melanogaster*. *Genetics*, **155**, 1821–1830.
84. Stark, J.M., Richardson, C., Ommundsen, M. and Jasin, M. (2004) Rad51 overexpression promotes alternative double-strand break repair pathways and genomic instability. *Oncogene*, **23**, 546–553.
85. Elliot, B., Richardson, C., Winderbaum, J., Nickoloff, J.A. and Jasin, M. (1998) Gene conversion tracts from double-strand repair in mammalian cells. *Mol. Cell. Biol.*, **18**, 93–101.
86. Porteus, M.H. (2005) Mammalian gene targeting with designed zinc finger nucleases. *Mol. Ther.* (in press, doi:10.1016/j.ymthe.2005.08.003).
87. Scott, C.T. (2005) The zinc finger nuclease monopoly. *Nat. Biotechnol.*, **23**, 915–918.