Genome Engineering with Targetable Nucleases

Dana Carroll

Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, Utah 84112; email: dana@biochem.utah.edu

Keywords
zinc-finger nucleases, TALENs, CRISPR/Cas, gene targeting, DNA repair, nonhomologous end joining, NHEJ, homologous recombination, HR

Abstract
Current technology enables the production of highly specific genome modifications with excellent efficiency and specificity. Key to this capability are targetable DNA cleavage reagents and cellular DNA repair pathways. The break made by these reagents can produce localized sequence changes through inaccurate nonhomologous end joining (NHEJ), often leading to gene inactivation. Alternatively, user-provided DNA can be used as a template for repair by homologous recombination (HR), leading to the introduction of desired sequence changes. This review describes three classes of targetable cleavage reagents: zinc-finger nucleases (ZFNs), transcription activator–like effector nucleases (TALENs), and CRISPR/Cas RNA-guided nucleases (RGNs). As a group, these reagents have been successfully used to modify genomic sequences in a wide variety of cells and organisms, including humans. This review discusses the properties, advantages, and limitations of each system, as well as the specific considerations required for their use in different biological systems.
INTRODUCTION

Targeted Genome Manipulation

This review describes how biochemistry has assisted genetics by introducing reagents for efficient and effective genome engineering. The science of genetics relies heavily on the analysis of mutations and the phenotypes they cause. Many geneticists seek to direct mutations and precise sequence changes to particular genes of interest. Targetable nucleases provide that capability. The field is relatively new and is moving very rapidly, so it may be considered brave—more likely, foolhardy—to undertake a review at this time. I hope that this effort, at a minimum, provides a basis for understanding and evaluating advances as they accumulate.

Targeted gene replacements in a small range of organisms have been possible for decades. The groups headed by Smithies (1) and Capecchi (2) showed that DNA introduced into cultured mammalian cells can be incorporated at the homologous natural locus, albeit at very low frequency. Capecchi’s group developed methods to isolate the rare recombinants in mouse embryonic stem (ES) cells (3), which were then used to produce whole mice carrying the targeted alteration. This technology had been available even earlier for baker’s yeast and other fungi (4, 5), and Rong & Golic (6) developed a gene targeting method for Drosophila. These approaches were not effective in many other organisms due to their inherently low efficiency, the difficulty of finding the desired products in the face of much more common nonhomologous integration events, and the absence of stem cells that would allow in vitro selection prior to generation of whole animals.

The development of targetable DNA cleavage reagents greatly enhanced the efficiency of gene targeting and substantially broadened its uses.

Double-Strand Break Repair

Essentially all cells treat double-strand breaks (DSBs) in their genomic DNA as potentially lethal damage and have multiple mechanisms to repair them. Among the repair processes is homologous recombination (HR). Cells can use an unbroken sister chromatid or homologous chromosome as a template to copy appropriate information into the break site (Figure 1). It is well known that DSBs caused by ionizing radiation are often repaired by HR (7). In essentially all organisms, naturally occurring meiotic HR events—both crossing over and gene conversion—are initiated by intentional DSBs (8), as is the HR-mediated mating-type switch in fungi (9). Several groups have shown that
Figure 1
Repair of a double-strand break. A break in chromosomal DNA can be repaired by homologous recombination (HR) with a sister chromatid (a) or a homologous chromosome. If a donor DNA supplied by the experimenter is used as a template for repair, some sequences from the donor will replace those at the target (b). Alternatively, nonhomologous end joining (NHEJ) may join the broken ends inaccurately, resulting in local, small insertions and/or deletions (c).

synthetic sites for highly specific nucleases, such as HO and I-SceI, can serve as targets for HR by using a genomic template or one supplied by the experimenter (10–13).

Most cells have an alternative to HR for repairing DSBs, known as nonhomologous end joining (NHEJ) (14). In this process, broken ends are rejoined without regard for homology, and the products often contain small insertions and deletions that alter the genomic sequence (Figure 1). In many cell types, NHEJ events outnumber those that depend on HR. The conclusion from these observations is that reagents that can make unique, targeted DSBs in chromosomal DNA both generate local mutations via NHEJ and provide the means for gene editing via HR.

THE PLAYERS
Investigators have attempted to use several different types of synthetic molecules to make targeted DSBs or other DNA damage, but with limited success (15–17). The most effective reagents to date have been protein nucleases with flexible specificity. Those that are the subjects of this review are shown in Figure 2. Several research threads converged to produce the first targetable nuclease, specifically the recognition that DSBs in chromosomal DNA stimulate HR, the discovery of zinc fingers as DNA recognition modules, and the characterization of the FokI restriction endonuclease.

Zinc-finger nucleases (ZFNs) consist of DNA-binding modules derived from natural transcription factors (TFs) that are linked to the nuclease domain of the Type IIS restriction enzyme, FokI (18, 19). Because the nuclease domain must dimerize to cut DNA, two ZFN molecules are required to target a single site. Transcription activator–like effector (TALE) nucleases (TALENs) employ DNA-binding modules from bacterial TALEs linked to the same FokI cleavage domain (20). The CRISPR/Cas RNA-guided nucleases (RGNs) rely on base-pairing between a guide RNA and the DNA target for recognition and on

www.annualreviews.org • Genome Engineering with Targetable Nucleases 411
a multifunctional Cas9 protein for cleavage (21, 22). All of these types of reagents, along with their development, are described in more detail below.

Members of another class, the homing endonucleases (HEs, also known as meganucleases), have long recognition sites (15–30 bp) but do not have distinct binding and cleavage domains. Many natural HEs with a wide range of recognition sequences exist (23, 24), and investigators have expended considerable effort on engineering HEs for novel specificity, with some notable successes. I do not describe these HEs extensively here, but an excellent recent review exists (25).

**ZINC-FINGER NUCLEASES**

*FokI Cleavage Domain*

The familiar Type II restriction endonucleases—long-time workhorses of molecular biology—recognize a short DNA sequence and cut within it. Type IIS enzymes, in contrast, cut some distance from their recognition site. Chandrasegaran and colleagues (26) examined FokI, a Type IIS enzyme, and found that its recognition and cleavage activities could be isolated in separate fragments following limited proteolysis. These authors demonstrated that cleavage specificity can be altered by linking the nonspecific nuclease domain to alternative DNA-binding domains. First, they fused the nuclease domain to the homeodomain DNA-binding module from the *Drosophila* Ubx protein (27). Second, they fused it to two different sets of zinc fingers provided by the Berg lab (19). Each hybrid cleaved phage λ DNA in a unique pattern that differed from that of natural FokI. Thus, the cleavage domain can be directed to an arbitrarily chosen sequence by linkage to an appropriate DNA-binding domain.

**Zinc Fingers**

The type of zinc finger employed in ZFNs is known as Cys2His2, which refers to the four residues that coordinate a zinc atom (Figure 3). After the Klug group (28) identified this module in the *Xenopus laevis* TFIIB, Berg (29) made a remarkably accurate prediction of its fold. Pavletich & Pabo (30) determined the structure of a three-finger unit bound to its DNA target, confirming the modularity of DNA recognition (Figure 3) (31).

Each finger is relatively small (∼30 amino acids) and has one α-helix and two short β-strands (Figure 3). A finger interacts with DNA through contacts between specific protein side chains and functional groups in the major groove. The primary contacts are made by residues −1, 3, and 6 relative to the start of the α-helix (31); we usually say that each finger binds 3 bp, recognizing that this is only an approximation. At least three consecutive fingers are needed to provide adequate binding affinity, and the best three-finger combinations have an equilibrium dissociation constant (K_d) value in the low-nanomolar range. Notably, zinc fingers bind DNA backwards; in other words, the N-terminal finger contacts the 3′-terminal DNA triplet, and so on.

Natural zinc fingers are abundant in eukaryotic sequence-specific TFs, and new ones have been created by protein design (32, 33). Some of these have been adopted for use in synthetic ZFNs, but most available fingers...
and finger combinations have been derived in the laboratory by partial randomization and selection. Using phage display (34–36), Barbas and colleagues isolated new fingers with specificity for all of the 5′-GNN triplets (where N is any base) (37) and many of the 5′-ANN and 5′-CNN triplets (38, 39). Liu et al. (40) characterized a somewhat different set for the 5′-GNNs. Researchers at Gendaq, Ltd., and at Sangamo BioSciences accumulated an extensive, proprietary inventory of fingers and finger pairs. Kim et al. (41) recovered natural zinc fingers from the human genome and characterized their binding preferences. Many other fingers have been recovered from the selection schemes described below.

Using lists of zinc fingers for individual DNA triplets, one can assemble new combinations for a broad range of target sequences, and this has been done in many cases (42–46). Success is not guaranteed, however, in terms of either affinity or specificity. A finger that performs well in one context may not do as well in another (47). The limited extent of contact between consecutive fingers seems unlikely to influence DNA binding; instead, a property of the sequence-dependent disposition of the base pairs may affect interaction. Investigators have achieved the best success with modular assembly using targets composed largely of 5′-GNN triplets (43, 47).

Because of this unreliability, several groups have focused on selecting new fingers in the desired context (48, 49). Schemes based on selection for binding activity in bacteria take account of context effects but can be time consuming, and they are based largely on affinity rather than specificity. Pairs of fingers that work well together have been tabulated (50–52), and they can be used for modular assembly. More fingers can be added to increase affinity and specificity, although more is not always better (53). Because zinc-finger clusters are not in perfect registration with B-form DNA, separating two-finger units with a one-amino acid linker may be helpful (54), and longer linkers can permit base pairs to be skipped (55). For those with few targets and deep pockets, Sigma-Aldrich® sells ZFN pairs that are constructed from the Sangamo database and have been extensively tested for activity.

In assembling new zinc-finger combinations, one can choose among several different frameworks on which to append the specificity-determining residues. The most common are (a) the natural Zif268 backbone, in which all the fingers are somewhat different (43); (b) a framework in which Zif268 fingers two and three are used for each pair of new zinc fingers; and (c) one in which all fingers have the same framework, either iterations of Zif268 finger 2 or one derived from an early consensus, known as Sp1C (33, 56). Fortunately, the choice of framework seems to have only modest effects on recognition properties (46).

**Cleavage Requirements**

In characterizing the requirements for ZFN cleavage, Smith et al. (57) found that the cleavage domain must dimerize to be active, a finding that is also true for natural FokI (58) but had not been appreciated in earlier ZFN research. Because the dimer interface is very weak, two monomers are required, each of which consists of a set of zinc fingers linked to a cleavage domain (Figure 2). The requirement for dimerization turns out to be a great advantage. Because neither monomer is active, the cleavage reagent is assembled only when both partners bind their target sequences. This characteristic has also enabled independent manipulation of the cleavage domain, including the isolation of a hyperactive variant (59).

The need for dimerization highlights the importance of the size of the spacer between binding sites and the length of the linker between the zinc-finger and nuclease domains of the protein (57, 60). In cells, when the linker was very short (initially called $n = 0$, but now generally referred to as $n = 4$), the optimum spacer length was exactly 6 bp (60). A molecular model of a ZFN pair on B-form DNA supported this observation. Subsequent research has confirmed this finding and shown that 5-bp spacers can also be effective (61–63). In the
initial fusion and in almost all subsequent ones, the nuclease domain was linked to the C terminus of the zinc-finger cluster, but the reverse arrangement is probably also active.

For experiments in cells, our group injected synthetic substrates and the ZFN protein into *Xenopus* oocyte nuclei. Importantly, cleavage and break-induced recombination occurred readily in a cellular environment with a chromatin substrate (60). Because natural *FokI* is a bacterial protein, it could easily have been defeated by chromatin structure.

**TALENS**

**TALEs**

The genomes of some plant-pathogenic bacteria (*Xanthomonas* and related genera) encode proteins that are imported into the host cell nucleus and regulate genes to promote infection (64). These proteins are known as transcription activator–like effectors (TALEs), although they are actually genuine transcription activators. The proteins have a DNA-binding domain made up of tandem repeats of ~34 amino acids, and these modules are not related to any other DNA-recognition motif. The framework is highly conserved among modules, but residues in positions 12 and 13—termed repeat variable diresidues (RVDs)—vary in concert with individual base pairs in the target DNA sequence (65, 66). In natural TALEs, several different RVDs may be found for any particular base pair, but the most common ones for each base pair constitute a recognition code that seems robust (Table 1). In contrast to zinc fingers, TALE modules bind DNA forward; in other words, the N-terminal repeat binds the 5′-most base pair, and so on.

The structures of TALE repeats bound to DNA (67, 68) show a remarkable helical staircase–like form, with consecutive modules closely apposed to each other (Figure 4). The RVD of each repeat is on a loop that is directed into the major groove. Surprisingly, only residue 13 makes contact with DNA, while the residue 12 side chain folds back and makes stabilizing contacts with other residues in the module. Furthermore, the specificity of the RVDs for A:T and T:A base pairs is determined largely by steric compatibility and van der Waals contacts, not by hydrogen bonds.

Table 1 shows the accepted code for DNA recognition by TALEs, along with alternative RVDs that may be useful. The standard RVD for G is NN (asparagine in positions 12 and 13), but it recognizes A in some contexts. NK and NH have greater specificity for G, but bind more weakly (69–72). Although HD provides good specificity for C, its binding is inhibited by methylation at position 5 of the base. This inhibition may pose a problem when targeting sites in higher organisms that

---

**Table 1** TALE modules used in DNA recognition

| Base pair | Canonical module | Alternatives
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A:T</td>
<td>NI</td>
<td></td>
</tr>
<tr>
<td>C:G</td>
<td>HD</td>
<td>N&lt;sup&gt;hv&lt;/sup&gt;</td>
</tr>
<tr>
<td>G:C</td>
<td>NN</td>
<td>NK, NH&lt;sup&gt;dc&lt;/sup&gt;</td>
</tr>
<tr>
<td>T:A</td>
<td>NG</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Modules are identified by the amino acids in positions 12 and 13 of the transcription activator–like effector (TALE) repeat, using one-letter codes: D, aspartic acid; G, glycine; H, histidine; I, isoleucine; K, lysine; N, asparagine. The asterisk indicates no residue.

<sup>b</sup>Modules HD and NN, for C:G and G:C base pairs, respectively, are considered strong in terms of binding affinity; NG is called weak by some (71) and strong by others (72).

<sup>c</sup>This module alleviates inhibition by methylation on position 5 of cytosine (73) but binds more weakly than HD.

<sup>d</sup>These alternatives are more specific for G:C but have weaker affinity than NN (69, 71, 72).
contain CpG dinucleotides, given that such sequences are often heavily methylated. The RVD $N^*$ (asparagine in position 12, no residue in position 13) overcomes this inhibition (73).

Numerous schemes for assembling new combinations of TALE modules for new targets have been published (74–86), and most of the building blocks have been made publicly available. See Reference 20 for a review of methods.

**TALENs**

Once the TALE recognition code was identified and tested, the parallel to zinc-finger recognition became obvious, and it was not long before fusions to the $FokI$ cleavage domain were produced (87–89). Whereas zinc-finger clusters require no additional protein, TALEs require some protein sequence beyond the DNA-binding modules on both ends (87, 89). Note that TALE frameworks and modules for new construction have been derived from several different natural proteins, but they are treated in the literature as having essentially the same properties.

Upstream of the obvious repeats in TALE proteins are two cryptic modules (known as repeats 0 and $-1$) that resemble the repeats structurally, although not in sequence (68). These modules make contacts with T, which is the preferred base in the 0 position of the target. The last TALE repeat is only partial and is often referred to as a half-repeat. Moreover, additional C-terminal residues from the natural protein are required before linkage to the $FokI$ cleavage domain (89), perhaps for folding and stability.

When a new TALEN target is chosen, the nuclease domain must still be encouraged to dimerize, so two binding sites are necessary. Each site should start with a T in the DNA and have at least an additional 10–12 bp, although 15–21 bp are more common. The binding sites are in opposite orientation and are separated by 12–20 bp due to the additional protein sequence between the TALE modules and the nuclease domain. A site of the form $TN_{18}$ is commonly referred to as an 18-bp site, and the protein that binds it carries 17.5 repeats, ignoring the initial T and acknowledging the final half-repeat.

**CRISPR/Cas RNA-GUIDED NUCLEASES**

**CRISPR Systems in Bacteria**

The most recent entry to the targetable nuclease family derives from an adaptive immune system that is widespread among bacteria and archaea (90). In simple terms, these microbes capture short DNA segments (20–50 bp) from invading viruses and plasmids and integrate them into their own genomes between copies of a repeat sequence (also 20–50 bp). The resulting arrays have been dubbed CRISPRs (short for clustered regularly interspaced short palindromic repeats) (91). When a member of the same virus or plasmid family invades again, the corresponding spacer sequences direct
cleavage of the incoming genome, thereby combating the intrusion.

In the Type II CRISPR system, only three components are required for cleavage: a processed RNA transcript of the CRISPR array containing one spacer and part of a repeat (crRNA); a second RNA with partial complementarity to the repeat, known as tracrRNA; and the CRISPR-associated Cas9 protein. Cas9 has two nuclease active sites, one to cleave each strand of the target DNA. It is guided by the crRNA/tracrRNA complex, which forms an RNA/DNA hybrid between the spacer sequence and the homologous sequence in the target (known as the protospacer because it gave rise to the spacer in a prior infection). Cas9 also requires a very short sequence in the target immediately downstream of the hybrid region, namely a protospacer adjacent motif (PAM). Jinek et al. (92) fused the crRNA and tracrRNA of *Streptococcus pyogenes* into a single guide RNA (sgRNA) that induced efficient Cas9 cleavage in vitro.

**CRISPR/Cas Nucleases**

The genome-engineering possibilities presented by the CRISPR/Cas system were immediately obvious to several groups. The Jinek et al. (92) paper appeared online at the end of June 2012. By January 2013, four groups had applied these reagents to genomic targets in human cells (93–96); one described its use in zebrafish (97); and a flood of additional reports followed. A consensus among these papers is that the sgRNA should have a 20-nt guide sequence corresponding to the desired target followed by 80 nt of hybrid crRNA/tracrRNA (95, 98). This sgRNA structure is often referred to as +85 in the literature.

The basic approach is quite simple (Figure 2). A target site is selected on the basis of the experimenter’s preferences. In addition to the 20 bp to be bound by the sgRNA, there must be an appropriately located PAM. For the *S. pyogenes* Cas9, this sequence is simply 5′ NGG. RNA/DNA hybrids slightly shorter than 20 bp have also been used successfully (99, 100). sgRNAs are typically produced by transcription in vitro or in vivo, and the promoters used will specify the base(s) at the 5′ end (one G for an RNA polymerase III promoter, two Gs for the commonly used SP6, T3, and T7 promoters). Initially, researchers were careful to have this base match the target, but doing so appears not to be strictly necessary (93, 101).

The first crystal structures of the Cas9 protein, alone (102) and in complex with sgRNA and a single strand of target DNA (103), and cryo-EM reconstructions (102) show how the protein domains are arranged and how recognition and cleavage are accomplished. Because of this protein’s multiple functions, little of the protein sequence is dispensable. There are extensive contacts with the RNA/DNA hybrid and with the constant portion of the sgRNA, as well as separate domains for cleavage of each of the target strands. These structures provide a basis for engineering versions of Cas9 that may have enhanced specificity and altered PAM recognition, among other valuable features.

The significant advantages of the CRISPR/Cas system versus ZFNs and TALENs are as follows.

1. A single protein is required, and it is always the same—no protein engineering is needed.
2. Targeting depends on base pairing, so sgRNA design requires only knowledge of the Watson–Crick rules.
3. New sgRNAs are very easily produced.
4. Because of advantages 1–3, it is feasible to attack multiple targets simultaneously with mixed sgRNAs (see the sections titled Rodents and Human Applications, below).

**Other Constructs**

Some researchers have made other types of cleavage reagents by using the flexible recognition modules described above. Specifically, they have created hybrids between HEs and zinc fingers or TALE domains to make recognition more flexible (104, 105). A restriction enzyme (106) and a nicking enzyme (107) have
been fused to TALEs to increase their specificity. Moreover, engineered zinc fingers and TALEs have been linked to other types of domains, including transcriptional regulatory sequences (108). A nuclease-dead version of the Cas9 protein, when linked to such a domain, becomes an RNA-guided TF (109, 110). Fusions to recombinases have also been produced (111, 112).

**GENOME ENGINEERING**

**A Little History**

After learning the basic requirements for ZFN cleavage—finger number, dimerization, linker and spacer lengths—it was time to attack a genuine genomic locus. With help from the Golic group, we chose the yellow gene of *Drosophila melanogaster* as a target. This gene has an easily visible mutant phenotype; gene activity is cell autonomous, so somatic mosaics can be identified; Rong & Golic (6) had successfully targeted yellow with their approach, which did not involve target cleavage; and we found a promising ZFN target in the coding sequence that consisted entirely of GNN triplets for which fingers had been characterized. We produced the novel zinc-finger combinations needed to address this target. Using heat shock to induce ZFN expression from integrated transgenes, we found both mutant patches in the injected generation and germ-line transmission of the mutant phenotype resulting from NHEJ repair of the induced DSB (113).

Using FLP and I-SceI to excise and linearize a donor DNA (6, 114), we also demonstrated homologous gene replacement (115). Effective ZFN cleavage and recombination were also demonstrated in cultured human cells, both with new zinc-finger combinations for genomic targets (63) and at synthetic targets for ZFNs of known specificity (116). These processes were carried out in other cells and organisms, and the list of species in which genomic targets have been modified is now quite long (Table 2).

**Delivery**

All of the above nucleases have been used in various biological systems to create targeted mutations and sequence replacements. In each case, success depended on the ability to deliver all the reagents efficiently and functionally to the cells or organisms under study. For cultured cells, the options for expressing the nuclease proteins include plasmid DNA transfection, viral vector delivery, and transfection with synthetic mRNA. When DNA is introduced, the nuclease coding sequences must be flanked by appropriate expression modules, including promoter, enhancer, and polyA addition signals. The proteins themselves should carry a nuclear localization signal. ZFN proteins inherently have an excess of positive charge that allows them to be taken up by cells directly from the culture medium (117), but this is not true for TALENs or Cas9. Long donor DNAs can also be supplied on plasmid or viral vectors, and oligonucleotide donors can simply be added to the medium.

When whole organisms are being addressed with targetable nucleases, other methods of delivery are required, and they must be adapted to each particular situation. In many cases, direct injection of nuclease mRNAs or DNA expression constructs into embryos is very effective. In zebrafish, for example, mRNAs are injected into the embryo cytoplasm shortly after fertilization (118, 119). Early expression of the proteins leads to both somatic and germ-line mutations via NHEJ. The same approach has been effectively used with many other organisms, including mammals (120–125). *Drosophila* embryos are multinucleate at the stage when injections are performed, but directing the injection to the posterior site of the germ-line precursor cells enhances the recovery of transmissible mutations (126). For all organisms, donor DNAs can simply be added to the mRNA injection mix.

In some situations, the biology of the organism interferes with a straightforward approach to delivery. Despite early success with NHEJ mutagenesis in zebrafish by use of ZFNs,
Table 2  Organisms with genomes modified by targetable nucleases*  

<table>
<thead>
<tr>
<th>Organism</th>
<th>ZFNs</th>
<th>TALENs</th>
<th>CRISPR/Cas</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Invertebrates</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drosophila</td>
<td>113, 115, 126</td>
<td>145, 181</td>
<td>182–184</td>
</tr>
<tr>
<td>Silkworm</td>
<td>213</td>
<td>211, 212</td>
<td>214</td>
</tr>
<tr>
<td>Mosquito</td>
<td>216</td>
<td>217, 218</td>
<td>—</td>
</tr>
<tr>
<td>Cricket</td>
<td>242</td>
<td>242</td>
<td>—</td>
</tr>
<tr>
<td>Butterfly</td>
<td>203</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Caenorhabditis elegans</td>
<td>130, 134</td>
<td>133, 134</td>
<td>132, 133, 135–141</td>
</tr>
<tr>
<td>Other nematodes</td>
<td>—</td>
<td>133</td>
<td>133</td>
</tr>
<tr>
<td>Sea urchin</td>
<td>243</td>
<td>244</td>
<td>—</td>
</tr>
<tr>
<td>Ciona</td>
<td>245</td>
<td>246</td>
<td>—</td>
</tr>
<tr>
<td><strong>Plasmodium</strong></td>
<td>247</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Vertebrates</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zebrafish</td>
<td>118, 119</td>
<td>76, 127, 185</td>
<td>97, 128, 189, 190</td>
</tr>
<tr>
<td>Medaka</td>
<td>248</td>
<td>249</td>
<td>—</td>
</tr>
<tr>
<td>Catfish</td>
<td>250</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Rainbow trout</td>
<td>251</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Xenopus</td>
<td>252</td>
<td>253, 254</td>
<td>255–257</td>
</tr>
<tr>
<td>Rat</td>
<td>123</td>
<td>198, 259</td>
<td>258, 260</td>
</tr>
<tr>
<td>Mouse</td>
<td>120, 125, 261, 262</td>
<td>199, 201</td>
<td>200, 202, 260</td>
</tr>
<tr>
<td>Rabbit</td>
<td>263</td>
<td>264</td>
<td>—</td>
</tr>
<tr>
<td>Goat</td>
<td>265</td>
<td>266</td>
<td>—</td>
</tr>
<tr>
<td>Pig</td>
<td>124</td>
<td>121, 266</td>
<td>266</td>
</tr>
<tr>
<td>Cow</td>
<td>210</td>
<td>121, 266</td>
<td>—</td>
</tr>
<tr>
<td>Chicken</td>
<td>267</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Monkey</td>
<td>268</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Human</td>
<td>63, 116, 233</td>
<td>75, 80, 89, 172</td>
<td>93–96, 220, 221, 269</td>
</tr>
<tr>
<td><strong>Plants</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Arabidopsis</em></td>
<td>194, 195</td>
<td>75</td>
<td>196, 278</td>
</tr>
<tr>
<td>Tobacco</td>
<td>205</td>
<td>206</td>
<td>196</td>
</tr>
<tr>
<td>Maize</td>
<td>207</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Soybean</td>
<td>270</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Rice</td>
<td>—</td>
<td>271, 272</td>
<td>196, 200, 273</td>
</tr>
<tr>
<td>Bunchgrass</td>
<td>—</td>
<td>272</td>
<td>—</td>
</tr>
<tr>
<td>Cabbage</td>
<td>—</td>
<td>274</td>
<td>—</td>
</tr>
<tr>
<td>Barley</td>
<td>—</td>
<td>275</td>
<td>—</td>
</tr>
<tr>
<td>Wheat</td>
<td>—</td>
<td>—</td>
<td>273, 276</td>
</tr>
<tr>
<td>Sorghum</td>
<td>—</td>
<td>—</td>
<td>196</td>
</tr>
<tr>
<td><em>Nicotiana benthamiana</em></td>
<td>—</td>
<td>—</td>
<td>277, 278</td>
</tr>
<tr>
<td>Petunia</td>
<td>208</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Chlamydomonas</td>
<td>279</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

(Continued)
investigators could not achieve repair from a donor template. When incorporation of an oligonucleotide donor was observed following TALEN or Cas9 cleavage, many of the products only partially matched expectation; they looked homologous on one end but nonhomologous on the other (127, 128). More recently, fully homologous products were obtained with a long donor (129), albeit at levels far below that of NHEJ mutants. Zebrafish embryos go through very rapid cell divisions following fertilization. During this phase, prompt repair o fD S B sb yN H E Js e e m st ob ep r e f e r a b l et o more deliberate, template-directed HR; this preference is also imposed on experimental manipulations.

An equally challenging situation is presented by the popular nematode *Caenorhabditis elegans*. Somatic mutagenesis was readily achieved with ZFNs, at both genomic and extrachromosomal targets (130). These worms are notorious for their ability to suppress transgene expression in the germ line via RNA interference (RNAi) (131), which probably frustrated early attempts to produce transmissible mutations. Recently, researchers successfully used mRNA injection or DNA injection to achieve germ-line mutagenesis, including HR, with all three nuclease platforms (132–140). Cho et al. (141) employed a novel scheme based on injection of Cas9 protein-sgRNA complexes, so no mRNA translation was required and RNAi was evaded. Surprisingly, in at least some of these cases (141), mutagenesis appears to have occurred in the germ line of the F1 generation, indicating that the nuclease persists for several days. Although the frequencies are typically not high, successfully modified worms are readily isolated by screening.

### Nuclease-Induced Mutations

When a nuclease successfully cleaves its desired target, what are the consequences? As discussed above, small-scale sequence changes are often introduced at the break by NHEJ. These changes are typically deletions and/or insertions of a few or some tens of base pairs. By chance, two-thirds of them will create a frameshift in a protein-coding sequence, often completely inactivating the gene product. If the alteration is in a critical region of the protein, even a multiple of 3 bp can produce a null mutation. The level of sequence alteration is typically assayed by polymerase chain reaction (PCR) amplification of the region, followed by DNA sequencing, by a gel electrophoresis assay based on the mismatch-specific Surveyor (142, 143) or T7EI endonuclease (41), or by high-resolution melt analysis (144).

The spectrum of mutations following TALEN cleavage is more biased toward deletions, and somewhat longer ones, than that observed with ZFNs (145, 146), presumably reflecting events that occur in the longer spacer between binding sites for TALENs. My assessment of published data indicates that Cas9 cleavage also leads to longer deletions. What limits them (if anything) is unclear, given that

---

### Table 2 (Continued)

<table>
<thead>
<tr>
<th>Organism</th>
<th>ZFNs</th>
<th>TALENs</th>
<th>CRISPR/Cas</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Miscellaneous</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese hamster ovary cells</td>
<td>219, 280</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Yeast</td>
<td>—</td>
<td>87</td>
<td>204</td>
</tr>
<tr>
<td>Bacteria</td>
<td>—</td>
<td>—</td>
<td>281</td>
</tr>
<tr>
<td>Hepatitis B virus</td>
<td>282</td>
<td>283</td>
<td>—</td>
</tr>
<tr>
<td>HIV-1</td>
<td>284</td>
<td>—</td>
<td>285</td>
</tr>
</tbody>
</table>

References describe the use of zinc-finger nucleases (ZFNs), transcription activator-like–effector nucleases (TALENs), and CRISPR/Cas nucleases in various organisms. Not all existing papers are cited, but the earliest ones and significant advances are provided. A similar table appears in another recent review (286).
neither the RNA/DNA hybrid region nor the PAM is spared from deletion. Occasionally, longer insertions and deletions of up to several hundred base pairs are produced. Attempts to bias repair toward long deletions by coexpressing exonucleases have been only moderately successful (147, 148). With all of the nucleases, occasional repair products have base substitutions at the repair site, suggesting that an error-prone DNA polymerase may be involved in the process.

Having characterized hundreds of NHEJ products in *Drosophila*, we found several insertions for which the source can be identified. Some are direct or inverted copies of short sequences very close to the break. A plausible model for generation of such insertions has been proposed (149), but it accounts for only a minority of all insertions. Although there is no direct evidence to suggest it, short insertions that have no obvious source might be produced by untemplated DNA additions to the ends at the DSB. Among the longest insertions are some that correspond to genomic sequences very far from the break. In one case, the insertion is a direct or indirect product of reverse transcription, given that it matches an exon–exon junction with an intron cleanly removed (150). Apparently, what we lump into the term NHEJ is actually a range of processes, all of which are designed to fix a break before more serious consequences occur.

One can make other genomic changes by using the targetable nucleases. Large deletions, thousands of base pairs long, have been produced by inducing cleavage at two widely separated sites in human cells (151), pigs (121), and zebrafish (152). Although each break is usually repaired independently, some events join one end from each break and eliminate the intervening DNA. Inversions and duplications can also be detected at low levels (153). Making a break on each of two different chromosomes has allowed for the construction of translocations that mimic those found in human cancers (154, 155). Researchers have also made large deletions on a single chromatid by making a single cut and providing a donor DNA with one homology at the break and another to a sequence some distance away (156).

**Homologous Recombination Versus Nonhomologous End Joining**

Often an experimenter wishes to recover products of HR with a donor template, in preference to NHEJ mutants. Doing so can be challenging because NHEJ dominates DSB repair in many situations, although this preference also varies among cell types. When frequencies are high enough, the desired HR products can be identified with molecular analyses.

As part of a study of the genetic requirements for genome engineering with ZFNs in *Drosophila*, we found that knocking out a key component of the major NHEJ pathway, DNA ligase IV, greatly improved the ratio of HR to NHEJ repair without significantly reducing the overall recovery of sequence alterations (126, 150). A *lig4* knockout also reduced the level of NHEJ mutagenesis in *C. elegans* (130). These organisms tolerate the loss of ligase IV, but many others do not. The identification of a small-molecule inhibitor that is specific for that ligase and could be used for transient depletion of its activity more broadly would be beneficial.

Another way to influence the outcome of nuclese cleavage is to change the type of break produced. Several groups have made derivatives of HEs, ZFNs, and Cas9 that cut only one strand at the target (94, 96, 157–160). Nicks are not substrates for NHEJ, but they can stimulate HR. A low level of NHEJ mutagenesis still takes place, perhaps because of progression of a replication fork into the nick, but it is substantially reduced. HR is also significantly lower with nicks than with DSBs, but not as dramatically reduced as NHEJ is. Nickases have been proposed to be safer alternatives to full nucleases, given that off-target single-strand breaks should not lead to mutations. The trade-off in efficiency can be considerable, however. A useful reporter, known as Traffic Light, was developed to make rapid distinctions between HR and NHEJ products (161).
When engineering HR, several practical concerns arise. The donor DNA must have sufficient homology to the target. With long, double-stranded donors, most investigators use homologies in the range between 1,000 and 4,000 bp, which are approximately equally distributed on the two sides of the break. The fact that longer homologies and perfect matches to the target are not required distinguishes break-induced targeting from traditional methods used in mammalian cells. Oligonucleotide donors as short as 40 bases have been successfully used (156); of course, they must span the break.

Another issue is how much sequence from the donor is incorporated at the target. This parameter is equivalent to what geneticists call conversion tracts, and it is important for two reasons: (a) Once a nuclease is known to be effective, it can be used to introduce desired sequence changes throughout a region, if conversion tracts are long, and (b) some applications (e.g., translational fusions) require exact insertion or substitution of sequences from the donor, so making a cut very close to the insertion site is necessary when conversion tracts are short. In *Drosophila*, conversion tracts after ZFN cleavage are gratifyingly long (many kilobases) (162), but they tend to be rather short in mammalian cells (100–200 bp) (163). Knowing what limits tract length and how that might be manipulated would be very useful.

**Off-Target Cleavage**

The above sections focus on ways to address an intended target. None of the nucleases has perfect DNA recognition specificity, so we should also be concerned about possible cleavage and mutagenesis elsewhere in the genome. In fact, the first genomically targeted ZFNs were toxic due to excessive cleavage and were effective only at moderate expression levels (113, 164). Cleavage and mutagenesis at off-target sites complicate the intended use in experimental organisms, but this problem can usually be ameliorated by repeated outcrossing or by studying combinations of independently isolated mutant alleles. For use in human gene therapy and in human food sources, such secondary mutations must be thoroughly characterized for safety or else scrupulously avoided.

How can we assess the extent and location of off-target effects? General assays, such as staining repair foci, provide only a broad assessment of the problem (142, 165). A popular approach is to make educated guesses about the identity of potential secondary targets and to monitor them individually. For example, experimental measures of nuclease sequence preferences can be turned into predictions of at-risk genomic sites. These preferences have been obtained through selection for preferred binding (166–168) or by in vitro cleavage of partially randomized substrates (169). Simple searches for genomic sequences with the fewest mismatches to the desired target have also been performed.

Candidate secondary targets have been monitored by PCR amplification and either a gel assay or deep sequencing. Often the level of mutation at such sites is undetectable, but in some instances it is distressingly high. Off-target cleavage is often a property of only one of a pair of nucleases that apparently has poor discrimination against related targets and/or excess affinity (164). A major step forward was the introduction of obligate heterodimer modifications of the FokI cleavage domain (142, 165). Interestingly, Smith et al. (57) anticipated these modifications by making changes at the dimer interface that prevented cleavage by a homodimeric ZFN. The first generation of obligate heterodimers also reduced cleavage activity in some contexts, but second-generation adjustments (170) corrected this flaw and are now routinely used for ZFNs and TALENs.

An unbiased method to identify secondary targets and to assay cleavage and mutagenesis at those sites is required. Gabriel et al. (171) developed such a method by capturing ends produced in cells by ZFNs for a target in the human CCR5 gene using ligation to the DNA of an integration-defective lentivirus. Whether some bias exists in this experiment is difficult to assess, but if so, it is expected to be minor. This approach and two others guided by in vitro
data (167, 169) all agree that the closely related CCR2 gene is the most commonly affected secondary target. For the less commonly mutated sites, there is much less agreement among the studies.

TALENs are generally less toxic and apparently more specific than ZFNs (144, 172–175). Remarkably, single mismatches between modules and base pairs can significantly decrease binding, even when there are still 15 or more matches. Because of the tight apposition of consecutive modules (67, 68), a single mismatch may disrupt adjacent contacts, propagating the disruption beyond a single base pair. Mismatches near the 5′ end of the target are more disruptive than ones near the 3′ end (72).

As mentioned above, alternative modules for some base pairs exist that can increase specificity (69, 71, 72).

As of early 2014, it is too early to pass judgment on CRISPR/Cas specificity, but several very recent studies suggest that this may be a significant issue (98, 99, 176–178). Some secondary targets—even ones with multiple mismatches to the sgRNA—are mutated at rates similar to that of the desired target. In principle, 20 bp are more than adequate to specify a unique site in the human genome, but not all positions in the RNA/DNA hybrid are equally specified. Positions closest to the PAM are most stringent, whereas those nearer the 5′ end of the RNA are less so, and multiple mismatches can apparently be tolerated. The Cas9 protein presumably stabilizes the duplex between the sgRNA and the target and also determines the base-matching requirements.

The relaxed specificity in this system may be adaptive (178a). Precise specification of 22 bp (20 in the sgRNA, 2 in the PAM) seems excessive if the goal is simply to find a match in a viral genome of less than 10^6 bp and avoid cutting in the host genome of less than 10^7 bp. Viral genomes are subject to constant variation and selection, so the next infecting agent will certainly differ slightly from the one that established a CRISPR insert. By accommodating mismatches, the host defense allows for this variation. It may be possible to engineer the Cas9 protein to require longer or more perfect matches, or Cas9 from other species may have more inherent specificity.

Three groups (99, 101, 179) describe an alternative that makes the CRISPR system more specific. They produced a version of Cas9 with only one nuclease active site and used two sgRNAs to direct it simultaneously to a pair of sequences near each other on opposite strands of the target. When both sites were nicked, a DSB was produced and led to both NHEJ and HR repair. This paired nicking approach both increases specificity by requiring dual recognition (analogous to ZFNs and TALENs) and reduces off-target effects by limiting single events to nicks.

Fu et al. (100) found that, in several cases, sgRNAs that are slightly truncated compared with the standard 20-nt guide sequence had improved specificity with little loss of on-target efficacy. These 17- and 18-nt guides have enhanced specificity presumably because shorter duplexes are more sensitive to mismatches in secondary targets.

A practical consideration is that high nuclease concentrations, although favoring efficient cleavage of the desired target, endanger off-target sites. Specifically, as the nuclease concentration nears saturation, the ratio of off- to on-target cleavage rises. Reducing the nuclease concentration improves this ratio, but at the cost of less-efficient on-target events.

APPLICATIONS

Model Organisms

Because of space constraints, I cannot describe all of the contexts in which ZFNs, TALENs, and CRISPR/Cas have been successfully applied to genomic targets in various organisms and cell types. Instead, I cite some examples and provide additional references. Table 2 lists these applications, but it will soon be out of date.

**Drosophila.** The first organism successfully targeted with ZFNs was the fruit fly (113). By
now, numerous genes have been knocked out in this organism with ZFNs (126, 164), TALENs (145, 180, 181), and CRISPR/Cas (182–184); these include genes in which null mutations had not previously been available. With these reagents, initial experiments were aimed at targets that provide an easily scored mutant phenotype, but mutation frequencies are often high enough that the desired alleles can be recovered using straightforward molecular analysis (126, 145, 184). Both NHEJ mutagenesis and HR have been achieved with an embryo injection procedure (126) by use of either long double-stranded DNAs or single-stranded oligonucleotides as donors (162). The ease of studying mutant phenotypes made Drosophila a good organism in which to evaluate several parameters that affect targeting efficiency. Because of the availability of other powerful genetic tools for Drosophila, ZFNs were not broadly adopted for flies, but with the ease of construction and high success rate observed for TALENs and CRISPR/Cas, this situation is likely to change.

**Zebrafish.** Despite the popularity of zebrafish in genetic and developmental studies, effective gene manipulation procedures were not available for them until the advent of targetable nucleases. Many genes were successfully mutated with ZFNs (118, 119), but the frequencies were often rather low. TALENs have provided much higher rates (76, 144, 185–188), enabling identification of new mutants in the injected generation as well as the isolation of germ-line variants. The CRISPR reagents also work well in fish (97, 189, 190), and several genes have been mutated simultaneously (192). As mentioned above, the biology of the zebrafish embryo frustrated initial attempts to achieve HR, but this problem has been overcome with the higher cleavage efficiency of TALENs and CRISPR/Cas (127–129).

**Arabidopsis.** Initial experiments with Arabidopsis targeted transgenes that had been inserted in the genome (193), but later efforts manipulated genuine genomic loci (75, 194–196). A continuing issue with this and with other plants is the delivery of the nucleases and donor DNAs. Arabidopsis is readily transfected using the T-DNA plasmid from Agrobacterium tumefaciens, but it has limited carrying capacity for exogenous sequences.

**Rodents.** Rat researchers point out that their organism has greater physiological similarity to humans in many respects than does the mouse and that it is more widely used for pharmacological testing. The absence of robust rat ES cells meant that the powerful gene-targeting approach used for mice (197) is not available for this organism. The demonstration that ZFNs can produce targeted mutations by direct embryo injection generated considerable interest (123). TALENs have been effectively used in rats as well (198). Direct embryo injection also works well in mice (120, 122, 125, 199–201), and this approach could replace currently used cell-culture steps and eliminate at least one generation on the path to recovering useful mutants.

Wang et al. (202) demonstrated the unique power of the CRISPR system by expressing the Cas9 protein in mouse ES cells along with up to five different sgRNAs. This experiment led to the recovery of cells with simultaneous disruptions in all the alleles of these genes. In mouse embryos, coinjection of Cas9 mRNA with two sgRNAs enabled biallelic mutation of both targets with high efficiency. Finally, coinjection of oligonucleotide donor DNAs produced targeted sequence changes at both sites. In principle, a very large number of sgRNAs, with or without donors, can be coinjected and the emerging newborns screened for any combination of induced sequence changes.

**Others.** The situation with C. elegans is described above. Targetable nucleases have been used successfully in Xenopus, sea urchin, Chlamydomonas, and many other organisms (Table 2). In the monarch butterfly, investigators mutated genes that affect circadian rhythms, with the long-term goal of dissecting migration behavior (203). The genome of the yeast Saccharomyces cerevisiae has been targeted.
with nucleases (87, 204), but given the ease of
gene replacement without target cleavage in
this organism, nuclease-mediated approaches
may have little impact.

Economically Important Organisms
The targetable nucleases offer an attractive al-
ternative to standard trait identification and
breeding in agricultural plants and animals.
Known mutations can be transferred from one
breed or cultivar to another without disrupt-
ing a favorable genetic background, and novel
mutations can be easily produced.

Plants. Tobacco protoplasts are readily trans-
formed and subsequently regenerated into
whole plants, making tobacco a good candidate
for early experiments with ZFNs and TALENs
(205, 206). Maize can also be handled experi-
mentally in embryonic callus culture and regen-
erated. Shukla et al. (207) showed that a single
allele governing phosphate production can be
manipulated in this plant. Nuclease coding se-
quences and donor DNAs were delivered on
plasmids in both the above cases. Plant viruses
have also been used for this purpose (208), but
like T-DNA, they have limited capacity for car-
rying exogenous sequences. Other plants that
have been targeted include soybean, rice, cab-
bage, barley, and bunchgrass (Table 2).

Livestock. Imagine increasing skeletal muscle
mass (i.e., meat) genetically in cows and pigs
(121), making pig organs more suitable for hu-
mans, transplantation (124) or curing econom-
ically important genetic diseases of livestock
(209). The tools for accomplishing this task
have been developed using ZFNs, TALENs,
and CRISPR/Cas. Direct embryo injection of
mRNAs and donor DNA is an effective deliv-
ery method, as is somatic cell-nuclear transfer
to enucleated eggs following in vitro manip-
ulation of the genome (121, 124, 210). Other
food animals that have been successfully tar-
geted with ZFNs and TALENs include rabbit,
catfish, and trout (Table 2). Large animals can
also serve as models for human genetic diseases,
sometimes with greater relevance than rodents.

Others. The silkworm has been targeted with
ZFNs, TALENs, and CRISPR/Cas (211–214).
Although the dramatic example of making fluo-
rescent silk (215) employed a transposon vec-
tor, not nuclease targeting, the latter approach
opens the door to a wide variety of other manip-
ulations. Economically important in a different
sense, mosquitos are vectors for transmission
developing tropical diseases. Initial success
in targeting the genome of Aedes aegypti with
ZFNs (216) and TALENs (217), and in Anophe-
les gambiae with TALENs (218), has been re-
ported. When large-scale production of mam-
nalian proteins is needed, Chinese hamster
ovary cells are often used; these cells have been
subjected to nuclease targeting, including three
successive rounds of gene knockouts (219).

Human Applications
Much nuclease-based genome engineering re-
search has focused on cultured human cells be-
cause of potential uses in human health. Tar-
getting with all of the nuclease classes is simple
to perform in established lines that are readily
transformed. Achieving efficient delivery to pri-
mary cells and to other cell types is more chal-
lenging, so appropriate vectors must be cho-
sen. For example, Reyon et al. (80) attacked 96
genomic targets with TALENs and recovered
mutations in 84 of them at frequencies rang-
ing from 2.5% to 56%, on the basis of the
T7EI assay. With greater sensitivity, more tar-
gets might have been deemed successful. Very
recently two groups demonstrated the multi-
plexing capabilities of the CRISPR/Cas nucle-
ases by simultaneously targeting tens of thou-
sands of human genes in populations of cultured
cells (220, 221).

Beyond the obvious knockins and knock-
outs, additional inventive applications have
been developed. Lombardo et al. (222) targeted
a sequence in a relatively early exon of the
IL-2Rγ gene with ZFNs and used HR with
a donor DNA to insert all downstream exons
at the site. This maneuver, also used in a mouse model of hemophilia (223), provides a means to correct disease-causing mutations throughout a large portion of a gene using a single nuclease target. Gutschner et al. (224) used ZFNs and HR to insert a strong polyA addition signal near the 5′ end of the gene for a long noncoding RNA. This process allows one to disrupt production of the RNA without needing to know what portions of its sequence are critical to its function. Hu et al. (225) demonstrated the versatility of TALEN targeting by knocking out the core sequence of a microRNA—a very small target indeed.

An exciting prospect is genome manipulation in human pluripotent cells—both ES cells and induced pluripotent stem (iPS) cells (172, 223, 226–232). Yusa et al. (233) clearly demonstrated how targetable nucleases could be used for therapeutic applications. These authors generated iPS cells from a patient carrying a mutation in the α1-antitrypsin gene and used a complex strategy involving ZFNs to correct the mutation without leaving a mark. This strategy led to the isolation of cells targeted at one allele and a proportion with biallelic correction. The cells were differentiated in culture and injected into immune-deficient mice to demonstrate the efficacy of the gene correction. In the future, these cells could be introduced back into the original patient. Only a single off-target exon mutation was introduced upon ZFN treatment, although 25 mutations accumulated during the initial derivation of the iPS cells (233).

A clinical trial using ZFNs has been under way for several years, and others have been approved. The initial target is the gene for CCR5 (167, 234), the most common coreceptor for HIV-1. Natural human CCR5 mutants are healthy and are resistant to the development of AIDS after infection. This trial uses ZFNs delivered by an adenovirus vector to introduce NHEJ mutations into the CCR5 gene in T cells derived from HIV-positive patients (168). The cells—a substantial proportion of which now carry null mutations in CCR5—are rein infused into the person who originally donated them. Trial results have not yet been published, but reports made at meetings are very promising. An obvious extension would be to perform the targeting in hematopoietic stem cells so that the therapeutic benefits will be long-lasting (235).

An alternative to correcting a mutated sequence is to deliver an intact version of the missing gene. Doing so would be indicated, for example, when the disease mutation is a large deletion. Random integration of a therapeutic gene has inherent dangers, as demonstrated by X-SCID trials in which a retroviral vector activated an oncogene by integration in its vicinity (236). Several groups have initiated the use of nucleases to insert genes into so-called safe-harbor sites in the human genome—loci in which an integrated gene would be expressed, but would have little danger of causing unwanted effects. One such site is AAVS1 (237), the preferred integration site for adenoassociated virus. Placement of a therapeutic gene in a safe harbor requires that it be provided with all necessary regulatory sequences, which is usually less desirable than manipulating the natural locus and relying on endogenous controls but would permit placing the transgene under inducible control, if that were desired. Other potential clinical applications that are being explored include (a) correcting hemophilia mutations in the Factor IX gene (223); (b) restoring the reading frame in some cases of Duchenne muscular dystrophy (175); (c) knocking out the other coreceptor for HIV-1, CXCR4 (238); (d) correcting mutations in hemoglobinopathies, including sickle cell anemia (230); and (e) creating universal donor cells by eliminating human leukocyte class I antigen expression (239).

**CHOOSING A NUCLEASE**

Regarding the choice of platform, it seems safe to say that the TALEN and CRISPR/Cas systems are replacing ZFNs for routine research laboratory use on the basis of their design advantages, efficacy, and reliability. Specificity remains an issue in the CRISPR system, but its simplicity and the fact that it is readily multiplexed are unique advantages.
It is too early to sound the death knell for ZFNs, however. Some of the current ZFN pairs are very highly engineered, very effective, and very specific. The ZFNs for CCR5, for example, will probably persist both as clinical reagents and as a standard to be met by future designs. ZFNs also have the advantage of small size. Coding sequences for two four-finger proteins occupy just over 2 kb and are readily carried by plasmid or viral vectors. TALEN coding sequences are approximately 3 kb each, and the Cas9 coding sequence is more than 4 kb.

LOOKING AHEAD

Targetable nucleases are valuable tools for genetic manipulation and analysis. To date, many published studies have focused on optimizing the construction and demonstrating the activity of the nucleases. Practical applications are now being explored in the areas of medicine and food production, as described above. In model organisms, we can expect increasing analyses of specific gene functions and development of human disease models that utilize nuclease-induced genome modifications.

Why the range of cleavage efficiency varies so greatly (and often fails outright) among targets within each class of nuclease remains a puzzle. There is no obvious correlation with affinity, although this lack of correlation has not been extensively evaluated. People often wonder whether chromatin structure plays a role. No systematic study has explored this relationship, but both active and inactive genes are targeted with equal efficiencies. The issue may be moot for most researchers, given that attempting a few designs for any one target gene has been sufficient to guarantee success. As demands for finer control and higher cleavage activity increase, however, the question may be worth addressing.

At present, it is difficult to imagine a type of nuclease that would have greater simplicity than base-pairing for target selection, but we may yet be surprised. In the CRISPR arena, components from organisms other than S. pyogenes may prove useful (240). For example, different PAM requirements, both looser and stricter, will broaden the range of accessible targets, and some Cas9 proteins might be less tolerant of mismatches between the target DNA and sgRNA. TALE modules from genera other than Xanthomonas could provide additional materials for TALEN designs (241). Considering the rate at which new studies are appearing and the speed with which interest in genome engineering is increasing, we can safely predict only that this area will continue to witness exciting developments.

SUMMARY POINTS

1. Three nuclease platforms—ZFNs, TALENs, and CRISPR/Cas—are very effective in generating targeted sequence changes in the genomes of a wide variety of organisms.

2. CRISPR/Cas nucleases are particularly easy to design and use but have shown a lack of specificity.

3. Local mutations are generated at the site of cleavage by NHEJ, and specific sequence changes can be introduced from a donor DNA template.

4. How the nucleases are delivered, and the expected outcomes, depends on the biology of the organism or cells being studied.

5. Applications of targetable nucleases include genome editing in model organisms, genetic improvement of economically important plants and animals, and human gene therapy.
FUTURE ISSUES

1. Specificity of nuclease cleavage is a significant issue; additional methods to reduce and detect off-target effects will be welcome.

2. With all three platforms, some new designs work well, whereas others function only marginally or not at all. It would be useful to know whether nuclease design, target accessibility, or other factors limit efficacy.

3. Applications to crop plants and food animals will require proof of efficacy and safety to gain regulatory approval and public acceptance.

4. The use of targetable nucleases in conjunction with human stem cells promises to be an exciting area of future therapeutic applications.

DISCLOSURE STATEMENT

I am coholder of a patent on chromosomal mutagenesis using ZFNs, and I receive license royalties from Sangamo BioSciences, Inc.

ACKNOWLEDGMENTS

I am grateful to the people who have worked in my lab on targetable nucleases and to the many collaborators and colleagues who have introduced me to a variety of fascinating experimental organisms. I apologize for any oversights I have made in citing relevant work. Research in my lab has been supported by grants from the National Institutes of Health, currently R01 GM078571, and in part by a University of Utah Cancer Center support grant.

LITERATURE CITED


# Contents

Journeys in Science: Glycobiology and Other Paths  
*Raymond A. Dwek* ................................................................. 1

Lipids and Extracellular Materials  
*William Dowhan* ..................................................................... 45

Topological Regulation of Lipid Balance in Cells  
*Guillaume Drin* ........................................................................ 51

Lipidomics: Analysis of the Lipid Composition of Cells and Subcellular Organelles by Electrospray Ionization Mass Spectrometry  
*Britta Brügger* ......................................................................... 79

Biosynthesis and Export of Bacterial Lipopolysaccharides  
*Chris Whitfield and M. Stephen Trent* ........................................ 99

Demystifying Heparan Sulfate–Protein Interactions  
*Ding Xu and Jeffrey D. Esko* ..................................................... 129

Dynamics and Timekeeping in Biological Systems  
*Christopher M. Dobson* .......................................................... 159

Metabolic and Nontranscriptional Circadian Clocks: Eukaryotes  
*Akhilesh B. Reddy and Guillaume Rey* ........................................ 165

Interactive Features of Proteins Composing Eukaryotic Circadian Clocks  
*Brian R. Crane and Michael W. Young* ....................................... 191

Metabolic Compensation and Circadian Resilience in Prokaryotic Cyanobacteria  
*Carl Hirschie Johnson and Martin Egli* ........................................ 221

Activity-Based Profiling of Proteases  
*Laura E. Sanman and Matthew Bogyo* ........................................ 249

Asymmetry of Single Cells and Where That Leads  
*Mark S. Bretscher* .................................................................... 275

Bringing Dynamic Molecular Machines into Focus by Methyl-TROSY NMR  
*Rina Rosenzweig and Lewis E. Kay* ............................................ 291
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chlorophyll Modifications and Their Spectral Extension</td>
<td>Min Chen</td>
<td>317</td>
</tr>
<tr>
<td>Enzyme Inhibitor Discovery by Activity-Based Protein Profiling</td>
<td>Micah J. Niphakis and Benjamin F. Cravatt</td>
<td>341</td>
</tr>
<tr>
<td>Expanding and Reprogramming the Genetic Code of Cells and Animals</td>
<td>Jason W. Chin</td>
<td>379</td>
</tr>
<tr>
<td>Genome Engineering with Targetable Nucleases</td>
<td>Dana Carroll</td>
<td>409</td>
</tr>
<tr>
<td>Hierarchy of RNA Functional Dynamics</td>
<td>Anthony M. Mustoe, Charles L. Brooks, and Hasib M. Al-Hashimi</td>
<td>441</td>
</tr>
<tr>
<td>High-Resolution Structure of the Eukaryotic 80S Ribosome</td>
<td>Gulnara Yusupova and Marat Yusupov</td>
<td>467</td>
</tr>
<tr>
<td>Histone Chaperones: Assisting Histone Traffic and Nucleosome Dynamics</td>
<td>Zachary A. Gurard-Levin, Jean-Pierre Quivy, and Geneviève Ahmouzni</td>
<td>487</td>
</tr>
<tr>
<td>Human RecQ Helicases in DNA Repair, Recombination, and Replication</td>
<td>Deborah L. Croteau, Venkateswarlu Popuri, Patricia L. Opresko, and Vilhelm A. Bohr</td>
<td>519</td>
</tr>
<tr>
<td>Intrinsically Disordered Proteins and Intrinsically Disordered protein Regions</td>
<td>Christopher J. Oldfield and A. Keith Dunker</td>
<td>553</td>
</tr>
<tr>
<td>Mechanism and Function of Oxidative Reversal of DNA and RNA Methylation</td>
<td>Li Shen, Chun-Xiao Song, Chuan He, and Yi Zhang</td>
<td>585</td>
</tr>
<tr>
<td>Progress Toward Synthetic Cells</td>
<td>J. Craig Blain and Jack W. Szostak</td>
<td>615</td>
</tr>
<tr>
<td>PTEN</td>
<td>Carolyn A. Worby and Jack E. Dixon</td>
<td>641</td>
</tr>
<tr>
<td>Regulating the Chromatin Landscape: Structural and Mechanistic Perspectives</td>
<td>Blaine Bartbolomew</td>
<td>671</td>
</tr>
<tr>
<td>RNA Helicase Proteins as Chaperones and Remodelers</td>
<td>Inga Jarmoskaite and Rick Russell</td>
<td>697</td>
</tr>
</tbody>
</table>
Selection-Based Discovery of Druglike Macrocyclic Peptides
   Toby Passioura, Takayuki Katoh, Yuki Goto, and Hiroaki Suga .................. 727

Small Proteins Can No Longer Be Ignored
   Gisela Storz, Yuri I. Wolf, and Kumaran S. Ramamurti .......................... 753

The Scanning Mechanism of Eukaryotic Translation Initiation
   Alan G. Hinnebusch ................................................................. 779

Understanding Nucleic Acid–Ion Interactions
   Jan Lipfert, Sebastian Doniach, Rhiju Das, and Daniel Herschlag .............. 813

Indexes

   Cumulative Index of Contributing Authors, Volumes 79–83 ..................... 843

   Cumulative Index of Article Titles, Volumes 79–83 ............................... 847

Errata

An online log of corrections to Annual Review of Biochemistry articles may be found at
http://www.annualreviews.org/errata/biochem
New From Annual Reviews:

Annual Review of Statistics and Its Application

Volume 1 • Online January 2014 • http://statistics.annualreviews.org

Editor: Stephen E. Fienberg, Carnegie Mellon University
Associate Editors: Nancy Reid, University of Toronto
Stephen M. Stigler, University of Chicago

The Annual Review of Statistics and Its Application aims to inform statisticians and quantitative methodologists, as well as all scientists and users of statistics about major methodological advances and the computational tools that allow for their implementation. It will include developments in the field of statistics, including theoretical statistical underpinnings of new methodology, as well as developments in specific application domains such as biostatistics and bioinformatics, economics, machine learning, psychology, sociology, and aspects of the physical sciences.

Complimentary online access to the first volume will be available until January 2015.

TABLE OF CONTENTS:

- A Systematic Statistical Approach to Evaluating Evidence from Observational Studies, David Madigan, Paul E. Stang, Jesse A. Berlin, Martijn Schuemie, J. Marc Overhage, Marc A. Suchard, Bill Dumouchel, Abraham G. Hartzema, Patrick B. Ryan
- The Role of Statistics in the Discovery of a Higgs Boson, David A. van Dyk
- Brain Imaging Analysis, F. DuBois Bowman
- Statistics and Climate, Peter Guttorp
- Climate Simulators and Climate Projections, Jonathan Rougier, Michael Goldstein
- Probabilistic Forecasting, Tilmann Gneiting, Matthias Katzfuss
- Bayesian Computational Tools, Christian P. Robert
- Bayesian Computation Via Markov Chain Monte Carlo, Radu V. Craiu, Jeffrey S. Rosenthal
- Build, Compute, Critique, Repeat: Data Analysis with Latent Variable Models, David M. Blei
- Structured Regularizers for High-Dimensional Problems: Statistical and Computational Issues, Martin J. Wainwright
- High-Dimensional Statistics with a View Toward Applications in Biology, Peter Bühlmann, Markus Kalisch, Lukas Meier
- Next-Generation Statistical Genetics: Modeling, Penalization, and Optimization in High-Dimensional Data, Kenneth Lange, Jeannette C. Papp, Janet S. Sinsheimer, Eric M. Sobel
- Breaking Bad: Two Decades of Life-Course Data Analysis in Criminology, Developmental Psychology, and Beyond, Elena A. Erosheva, Ross L. Matsueda, Donatello Telesca
- Event History Analysis, Niels Keiding
- Statistical Evaluation of Forensic DNA Profile Evidence, Christopher D. Steele, David J. Balding
- Using League Table Rankings in Public Policy Formation: Statistical Issues, Harvey Goldstein
- Statistical Ecology, Ruth King
- Estimating the Number of Species in Microbial Diversity Studies, John Bunge, Amy Willis, Fiona Walsh
- Dynamic Treatment Regimes, Bibhas Chakraborty, Susan A. Murphy
- Statistics and Related Topics in Single-Molecule Biophysics, Hong Qian, S.C. Kou
- Statistics and Quantitative Risk Management for Banking and Insurance, Paul Embrechts, Marius Hofert

Access this and all other Annual Reviews journals via your institution at www.annualreviews.org.