Published in final edited form as:

Methods Mol Biol. 2014 ; 1123: 1–26. doi:10.1007/978-1-62703-968-0_1.

Homing Endonucleases: From Genetic Anomalies to Programmable Genomic Clippers

Marlene Belfort and

Department of Genetics, Wadsworth Center, New York State, Department of Health Albany, NY, USA

Richard P. Bonocora

Department of Genetics, Wadsworth Center, New York State, Department of Health, Albany, NY, USA

Abstract

Homing endonucleases are strong drivers of genetic exchange and horizontal transfer of both their own genes and their local genetic environment. The mechanisms that govern the function and evolution of these genetic oddities have been well documented over the past few decades at the genetic, biochemical, and structural levels. This wealth of information has led to the manipulation and reprogramming of the endonucleases and to their exploitation in genome editing for use as therapeutic agents, for insect vector control and in agriculture. In this chapter we summarize the molecular properties of homing endonucleases and discuss their strengths and weaknesses in genome editing as compared to other site-specific nucleases such as zinc finger endonucleases, TALEN, and CRISPR-derived endonucleases.

Keywords

Homing endonucleases; Genome editing; Altered specificity endonucleases; Biotechnology; Endonuclease applications

1 History and Evolution

1.1 Discovery of the First Homing Endonuclease

In the early 1970s at the Institut Pasteur in Paris, researchers discovered a phenomenon that seemed to throw the principles of genetic inheritance described by Mendel out the window. While performing crosses with yeast strains, they discovered that a particular mitochondrial allele, known as omega (ω), was unidirectionally inherited at what was later determined to be the gene for the large ribosomal RNA subunit (LSrRNA) [1, 2]. The ω allele was shown to contain a LSrRNA gene disrupted by an intervening sequence known as a group I intron [3, 4]. At that time introns were thought to be purely “junk” DNA that were removed from a gene posttranscriptionally, leaving the mature RNA intact, fully functional, and without a discernible phenotype. How then could this junk exert such a strong influence on its own

inheritance? The mystery was finally solved when, surprisingly, the ω intron was found to contain a separate gene. That gene encodes a DNA endonuclease that recognizes and introduces a double-strand break (DSB) in LSU rRNA genes lacking the intron [5].

The endonuclease-dependent mechanism for the “super” inheritance of ω was simple and elegant and has since been shown to be a general phenomenon known as “homing” [6]. By introducing site-specific DNA breaks that then are repaired, these homing endonucleases (HEs) stimulate the cellular recombination and DNA repair processes to fix the break by simply copying the gene encoding them (the homing endonuclease gene or HEG) and flanking DNA into the broken chromosome (Fig. 1a). This gene conversion event is termed double-stranded break repair (DSBR) [7]. The end result is the proliferation of the HEG by lateral transfer.

1.2 Evolutionary Considerations

Since the discovery of ω , evidence has accumulated that the HEG is the actual mobile element and that the intron provides safe haven in the form of a phenotypically silent locus, while the intron enjoys the ride [8]. First, HEGs have been found in different genetic contexts such as different classes of intron [9], protein-splicing elements called inteins [10–12] and as free-standing genes [13]. Second, phylogenetic analysis of HEGs and the splicing elements indicated that they have different evolutionary histories [12]. Third, similar HEGs are located in different positions of similar introns [14]. Fourth, similar introns have been invaded by different HEGs [15]. Finally, disruption of the HEG abrogates homing [5, 16–18].

One might then ask whence the HEs arose. In addition to their function in self-promoting their lateral transfer, their catalytic scaffolds have been found in genes involved in general housekeeping processes such as recombination, repair, gene regulation, RNA folding, genome stability, and restriction of foreign DNA [19]. In most cases it is unclear if the chicken or the egg, namely the HEGs or the housekeeping gene, came first. Although in some cases the order is defined, as for the derivation of the HO endonuclease, which promotes yeast mating-type switching, from the intein-encoded VDE endonuclease [20, 21], most origins are ambiguous. What is clear is that these enzymes live a dynamic lifestyle that must adapt to the host gene and genome that they invade to minimize their impact on the fitness of the organism [22, 23].

HEGs within splicing elements have been proposed to undergo a lifecycle characterized by rapid invasion to fixation of a gene, slow decay of the endonuclease activity, eventual complete loss of the element and subsequent reinvasion [24]. Amendments to this model include the opportunity of the HEG to develop new functions that benefit the host organism, as for the HO endonuclease and other examples described below (Table 1), thereby preventing loss of the element, or for the HEG to transpose to a new favorable location [25, 26].

Regardless of the genetic environment in which a HEG exists, a key aspect of the homing mechanism is that the converted chromosome is now protected from cleavage by the HE. For intron-and intein-encoded HEs the associated intervening sequence lies within the

sequence recognized by the HE. Once the intron/intein is copied into the new chromosome the recognition site is split by the intervening sequence thereby preventing self-cleavage. The process for free-standing HEGs is similar, but with a specific problem: the HE recognizes a target site that can be separated by several hundred base pairs from the HEG, typically located in an adjacent gene. The mode of protection can vary here; often the recognition site contains sequence changes between the sensitive chromosome without the HEG (recipient) and the refractory chromosome with the HEG (donor), which make the chromosome with the HEG refractory to HE cleavage (known as intronless homing) [27]. The sequence change to prevent cleavage can even be an unrelated intron or other genetic element inserted within the HE recognition site (termed collaborative homing) [28, 29].

A characteristic shared by introns and HEGs may have been key to the origin of mobile introns; both prefer conserved, functionally significant sequences such as those that encode enzyme active sites [30–32]. Extant splicing elements have been retained in these locations presumably because their loss has to be precise or the coding sequence will be altered and the gene function impaired. Precise loss, on the other hand, would lead to reinvasion. HEGs benefit from recognizing a target site that is well conserved. This similarity in sites suggests that free-standing HEGs can be “preadapted” to recognize intron insertion sites.

Such a scenario is likely to have occurred at the DNA polymerase gene (gene 5) of T3-like phages. Phage T3 contains a freestanding HE F-TsII (encoded by gene 5.3), immediately downstream of gene 5. F-TsII recognizes and cleaves a site within gene 5 of related phages, within the enzyme’s catalytic center. As expected, the corresponding sequence in T3 is not cleaved. The related phage Φ I lacks a HEG downstream of gene 5, but instead has a group I intron inserted one nucleotide away from the F-TsII cleavage site. This intron contains a HEG that is similar to F-TsII and encodes a functional HE named I-TsII. Both of these HEGs cleave intronless gene 5 alleles at precisely the same location. Thus F-TsII exemplifies a HE preadapted for an intron insertion site that has since invaded an intron [17].

Since both the HEGs and splicing elements converge on the same sequences, is there an advantage to their forming a composite element? Free-standing HEGs generally cleave far from their insertion sites. As a result, transfer of the cleavage-resistant allele from the donor genome can occur without cotransfer of the HEG [17, 27]. The result is an increase of resistant alleles and therefore a concomitant reduction in homing opportunities and pressure to retain the HEG. The HEG solves this problem by coupling with the resistance element (a group I intron disrupting the HE recognition site) thereby ensuring the transfer of both. The intron also benefits as it is now intimately linked to a mobile element and will persist in the population.

1.3 HEGs from Then Till Now

In the more than 40 years since the observation of unidirectional inheritance of ω that led to the discovery of intron homing, much has been learned about the recombination process and the HEGs responsible. Although the biological role of HEGs remains elusive, the usefulness of HEGs as tools in biotechnology, medicine, agriculture, and possibly population control of disease vectors is becoming increasingly clear. In this chapter we will provide an overview

of the biochemistry and structure of HEs and how HEs can be tailored for the various applications. We further compare these enzymes to other agents of gene targeting.

2 General Properties of HEs

HEs are small proteins (< 300 amino acids) found in bacteria, archaea, and in unicellular eukaryotes (reviewed by Stoddard [33]). A distinguishing characteristic of HEs is that they recognize relatively long sequences (14–40 bp) compared to other site-specific endonucleases such as restriction enzymes (4–8 bp). These lengthy recognition sites, and the name of the first such known enzyme, ω (also known as I-SceI), have given rise to the term “meganuclease” [34]. Another feature that sets HEs apart from restriction endonucleases is their lack of absolute sequence specificity. Whereas restriction enzyme binding and/or cleavage depend on a perfect match to the recognition sequence, HEs are less discriminating, often tolerating multiple sequence changes within their recognition site [35, 36]. This is apparent at the structural level where there is a great disparity between the number of contacts made by restriction endonucleases and HEs. Restriction endonucleases exploit most of the potential hydrogen bonds between the proteins and their target sites [37] whereas HEs utilize only a fraction of the possible hydrogen bonds [38–40]. The positions that are tolerated by HEs are often those at third positions of codons, which vary naturally between organisms. Such tolerance allows homing into new sites. Despite the imperfect fidelity, the lengthy recognition sites can make HEs highly specific, often cutting large genomes only once. This attribute makes the HEs amenable to genome editing, where spurious off-site cleavages are detrimental.

HEs have been historically categorized by small conserved amino acid motifs. At least five such families have been identified: LAGLIDADG; GIY-YIG; HNH; His-Cys Box and PD-(D/E)xK, which are related to EDxHD enzymes and are considered by some as a separate family (Table 1, Fig. 2a). At a structural level, the HNH and His-Cys Box share a common fold (designated $\beta\beta\alpha$ -metal) as do the PD-(D/E)xK and EDxHD enzymes. The catalytic and DNA recognition strategies for each of the families vary and lend themselves to different degrees to engineering for a variety of applications.

3 HE Families

3.1 LAGLIDADG Endonucleases

The LAGLIDADG endonucleases (LHE) make up a large, well-described family of HEs identified largely in archaea and organelles of lower eukaryotes. LHE genes exist in a variety of genomic environments, being encoded within group I introns, archaeal rRNA introns, as in frame fusions with inteins, and as freestanding genes.

LHEs have either one or two copies of the signature LAGLIDADG motif and leave 4 nt 3'-extensions. Crystallographic studies with both dual and single motif LHEs, show that the motif is present in an α -helix that is responsible for mediating protein–protein interactions [39, 41–44] (Fig. 2a). The LAGLIDADG helices act as an intramolecular interface for the dual motif endonucleases, or as the site of dimerization between identical monomers for single-motif endonucleases. These helices are also responsible for correctly positioning the

active site residues. Biochemical and structural evidence indicates that the second conserved acidic residue of the motifs (LAGLIDADG) acts in metal ion coordination critical for the catalytic activity of the endonuclease [39, 45, 46]. I-CreI catalysis involves hydrolysis of the scissile phosphates following a canonical two-metal mechanism, with a single metal in each active site [47, 48]. Target recognition is mediated by interactions between a four strand anti-parallel β -sheet from each monomer/domain [39] (Fig. 2a). LHEs with a single motif recognize largely palindromic sequences whereas the two-motif nucleases need not be symmetric.

In some cases LHEs have taken on secondary roles, as for example the aforementioned HO endonuclease [21], which initiates a gene conversion event that results in mating-type switching [20, 49]. Additionally, several LAGLIDADG proteins encoded within group I introns of yeast act as RNA maturases instead of DNA endonucleases, assisting in the folding of their cognate intron into its catalytically active conformation. Some intron-encoded proteins can function as both maturase and endonuclease [50–53] while others can be converted from a maturase to an endonuclease by mutation [54]. I-AniI, a group I intron-encoded LAGLIDADG endonuclease/maturase, has the DNA- and RNA-binding sites located at independent surfaces on the protein [50, 55, 56], suggesting that maturase activity is an adaptation independent of endonuclease function. Finally, in an extreme example of LHE adaptation, the DUF199 family of bacterial transcriptional regulators resemble dual motif LHEs lacking catalytic aspartate residues, fused at their C-terminus to a helix-turn-helix DNA-binding domain [57–59]. Although the LAGLIDADG domain alone has lost its DNA-binding capacity, it improves binding of the full length protein about fourfold over the HTH domain alone, providing an example of adaptation of both its catalytic and DNA-binding functions [19, 59].

3.2 GIY-YIG Endonucleases

GIY-YIG endonucleases are modular proteins that exist in all three domains of life. Computational studies coupled with extensive genetic, biochemical, and structural analyses of I-TevI have provided a detailed picture of a member of the GIY-YIG family of endonucleases [60–62]. I-TevI and its close relative I-BmoI are endonucleases encoded within introns interrupting the thymidylate synthase genes of bacteriophage T4 and the soil bacterium *Bacillus mojavensis*, respectively. I-TevI and I-BmoI consist of an N-terminal globular catalytic domain, that is conserved with other family members, and a modular C-terminal DNA-binding domain connected by a flexible linker (Fig. 2a) [63–65]. GIY-YIG HEs cleave their homing site leaving 2 nt 3' overhangs [14, 66, 67]. The N-terminal catalytic domain spans ~90 amino acids and contains five conserved regions [60]. The first GIY-YIG endonuclease visualized in complex with DNA is restriction enzyme R. Eco29k1, showing both tyrosines within the conserved motifs in the catalytic center [68]. For GIY-YIG restriction endonuclease Hpy1881, the GIY tyrosine was identified as a general base, with a conserved glutamate anchoring a single metal in the active site [69].

The C-terminal domains of both I-TevI and I-BmoI contribute the bulk of the DNA-binding energy. Indeed, this domain alone binds to the homing site with equal affinity to the full-length protein [14, 63]. Structural studies of the C-terminal domain of I-TevI in complex

with its homing site has revealed that the DNA-binding domain itself consists of three smaller structured subdomains linked together by elongated unstructured regions (Fig. 2a). The DNA-binding domain wraps around the minor groove of its target DNA making few base-specific contacts [40]. Consistent with this minor groove binding is the demonstration that no single base-pair in the homing site is absolutely required for cleavage [36].

In addition to GIY-YIG motifs in some restriction enzymes and associated with the reverse transcriptase of Penelope retroelements [70], the motif also occurs in enzymes involved in DNA repair and maintenance of chromosomal genome stability. These include eukaryotic flap endonucleases, eukaryotic S1 $\times$ 1–S1 $\times$ 4 resolvase, and bacterial UvrC nucleotide excision repair protein (summarized in [61]). As another example of HE adaptation, I-TevI binds to an operator site overlapping its promoter and acts as an autorepressor [71]. The operator site is not efficiently cleaved, but is bound in a similar fashion as the homing site. Apparently this repression delays translation of I-TevI, facilitating splicing of the host intron [22].

3.3 HNH Endonucleases

Members of the HNH family have been found in group I introns in plastids and phages, group II introns, and free-standing ORFs of bacteria, phages, and a cyanobacterial intein [12]. The HNH motif consists of a stretch of approximately 30 amino acids that contain three highly conserved histidine and/or asparagine residues. Structural analysis indicates a role in metal binding for the first two conserved amino acids [72–74].

Many members of the HNH family violate one or more of the canonical properties for intron-encoded HEs. Unlike HEs from other families, many HNH endonucleases nick one strand of dsDNA 5' to the intron insertion site [75, 76], although at least one HNH endonuclease, I-TevIII, has been shown to make a DSB by dimerization [77]. Another deviation from canonical HEs is demonstrated by the endonucleases I-HmuI and I-HmuII (Fig. 2a) encoded within homologous group I introns interrupting the DNA polymerase gene from the *B. subtilis* phages SPO1 and SP82, respectively. They are nicking endonucleases that cleave both intronless and intron-containing targets. These endonucleases prefer to cleave the DNA of the heterologous phage in vitro and I-HmuII has been shown to exclude the SPO1 intron and flanking regions during mixed infection [75].

In group II intron retrohoming, the HNH nuclease is part of a multidomain intron-encoded protein [12] that forms a ribonucleoprotein particle (RNP) with its cognate intron. The group II intron portion of the RNP directs the complex to the appropriate target via base-pairing [6, 78]. The intron RNA reverse splices into the sense strand of the target while the HE nicks the template strand. This nick provides a 3'-OH which is used to prime RNA-dependent DNA synthesis by a separate reverse transcriptase domain of the HE.

The HNH motif has also been identified in nonspecific colicins nucleases [79, 80], resolvases [81] and type II restriction endonucleases [82]. Interestingly, the Cas9 endonuclease involved in the clustered, regularly interspaced, short palindromic repeats (CRISPR) system of bacterial adaptive immunity from *Streptococcus thermophilus* contains two catalytic nicking domains, one of which contains an HNH motif, and the other a RuvC-

like motif. CRISPR endonucleases form RNP complexes with small bacterially-encoded RNAs that are homologous to short sequences derived from previously encountered phage or plasmids. Similar to group II intron HEs, Cas9 forms an RNP with the CRISPR RNA (crRNA), and is directed to target DNA by base pairing between the crRNA and target. A DSB results from two independent nicks in the target DNA, one by the HNH motif and the other by the RuvC-like motif [83, 84].

3.4 His-Cys Box Endonucleases

The His-Cys box endonucleases comprise a much smaller family of HEs, found in group I introns interrupting nuclear rRNA genes of lower eukaryotes [33, 85]. As the name implies, the signature of this family is a region that is rich in histidine and cysteine residues. The best studied of the His-Cys Box family is I-PpoI, a group I intron-encoded endonuclease from the slime mold *Physarum polycephalum*. The crystal structure of I-PpoI shows a dimer, where each monomer contains two histidine- and cysteine-rich sequences that coordinate separate zinc ions and function to stabilize the protein structure [86]. Like LHEs, I-PpoI binds its target DNA using antiparallel β -sheets. However, unlike the LHEs, I-PpoI induces a strong bend into the DNA target to bring the scissile phosphates into proximity with each active site (Fig. 2a).

Comparison of the structures of I-PpoI, the colicin E9 HNH endonuclease and a nonspecific nuclease from *Serratia* have identified a structural similarity at the active sites of all three. This similarity has suggested a reclassification of the two families into one known as $\beta\beta\alpha$ -Me which reflects the three secondary structural elements and the bound metal ion that define the motif [87].

3.5 PD-(D/E)xK and EDxHD Endonucleases

Restriction endonucleases share very little sequence conservation; however many contain a common catalytic fold from the PD-(D/E)xK super-family [88]. This motif appears to have been adapted by the group I intron-encoded homing endonuclease I-Ssp6803I from the cyanobacterium *Synechocystis* sp. 6803 [38, 89]. I-Ssp6803I is a small protein that functions as a tetramer to recognize a fairly large target site [38, 90, 91] (Fig. 2a). Like other HEs this recognition involves a paucity of protein–DNA contacts utilizing only one third of the possible hydrogen bonds [38]. This is in contrast to restriction endonucleases which use a high density of protein–DNA contacts to recognize small DNA target sites [38]. Presumably the under-saturation of contacts by I-Ssp6803I allows for cleavage of sequence variants as with other families of HEs. In addition to restriction endonucleases and HEs, the PD-(D/E)xK motif occurs in nucleases involved in DNA recombination, tRNA splicing, transposition, Holliday junction resolution, DNA repair, and Pol II termination [92].

Recently, a family of HEs, EDxHD, related to the PD-(D/E)xK was discovered by a bioinformatic analysis of the Global Ocean Sampling (GOS) environmental metagenomic sequence data [93, 94]. Like the GIY-YIG and HNH HE families, the EDxHD is modular. Although the overall fold in its catalytic domain is similar to the PD-(D/E)xK fold, the active site has diverged [95]. The crystal structure of I-Bth0305I, an EDxHD endonuclease encoded in a group I intron interrupting the *recA* gene of the *Bacillus thuringiensis* 0305r8–

36 bacteriophage, supports bioinformatic evidence [96] that this family is homologous to very short patch repair (Vsr) endonucleases [19, 95]. The EDxHD HEs are also associated with split inteins encoded by a non-contiguous open reading frame, and several genes involved in DNA replication and repair [96]. Thus, involvement of enzymes related to HEs in nucleic transactions of the cell is a common feature among the different HE families.

4 Modular Semisynthetic DNA Cleavage Enzymes

In addition to HEs, modular cleavage enzymes are being assembled in the laboratory for genome engineering. Two modular types of semisynthetic site-specific cleavage enzymes are the zinc finger nucleases (ZFNs) [97] and the transcription activator-like effector (TALE) nucleases (TALENs) [98]. Both ZFNs and TALENs contain multiunit DNA-binding domains in which the specificity residues are fused to nonspecific DNA cleavage modules, typically from the type II restriction enzyme FokI (Fig. 2b). The DNA-binding domain of the ZFNs contain between three and five zinc fingers, which are fairly specific for stretches up to ~15 bp. Likewise, the modular TALE DNA-binding domain contains units of amino acid sequences each with specificity for a single nucleotide. A combination of these units forms a DNA-binding cartridge. The current most popular cleavage domain is again from FokI, which must dimerize. This requires a pair of DNA-binding cartridges each recognizing opposite strands to be fused at their C-termini to the FokI cleavage domain. The net result for both ZFNs and TALENs is a large tailored site-specific dimeric cleavage enzyme (Fig. 2b).

5 Reprogramming HEs

5.1 Birth of an Industry: Selection Systems

Almost two decades ago, when little was yet known of the basis for sequence specificity of the LHE enzymes, the ω endonuclease I-SceI was used to induce DSBR in mammalian chromosomes [99, 100]. It soon became clear that altered specificity variants of LHEs would have enormous application in both research and the biotech industry, and various selection systems were designed to isolate altered-specificity mutant enzymes. These include a bacterial blue-white colony screen based on β -galactosidase elimination [101], selection for growth based on control of a cell death protein [102], a bacterial two-hybrid selection system [103], and a yeast or mammalian white-blue screen based on repair of β -galactosidase [104]. More recently, other reporters such as URA3 and GFP have been used in the DNA repair assay, and yeast surface display has been employed as an effective way to isolate altered-specificity mutants [105]. The effectiveness of these selection systems and the perceived utility of retargeting HEs rapidly spawned several biotech companies, Collectis Bioresearch in 1999, and both Precision Biosciences and PreGenen Genome Engineering in 2006.

5.2 HEs with Altered Specificity

As structures of members of all the HE families were solved, including monomeric and dimeric LHEs (Fig. 2a), rational design coupled with randomization and screening in a high-throughput format took hold, resulting in LHEs with altered specificity [106]. Simultaneously, highly sophisticated computational reprogramming resulted in redesigned

LHE DNA binding and cleavage specificities [107]. RosettaDesign (RD) has been used to generate thousands of different mutants of the LHE I-CreI targeted towards 16 different base pair positions in the 22 bp I-CreI target site. Of these, over two-thirds had the intended new site specificity [108]. These results and those with the LHE I-MsoI demonstrate that specificity switches for multiple concerted base pair substitutions can be computationally designed, and that iteration between design and structure determination provides a route to large-scale specificity reprogramming [109]. Another approach to increasing the LHE specificity is by mining naturally occurring enzymes with different target sites that can be used as alternate platforms for reengineering [110–112]. A useful LAGLIDADG HE database and engineering server is LAHEDES <http://homingendonuclease.net/>.

5.3 Nicking Endonucleases

Another focus of attention has been enzymes that perform single-strand DNA cleavage rather than DSBs. Sequential cleavage among the GIY-YIG endonucleases [113, 114] and nicking among the HNH enzymes [115] were shown to result in gene conversion events. Later, LHEs were converted into nickases (reviewed by Chan et al. [116]). A nicking variant of I-AniI has been generated, that stimulates site-specific homologous recombination [117]. These nick-induced recombination events have two distinct and important advantages: they protect against error-prone nonhomologous end-joining reactions (Fig. 1b), and they reduce cellular toxicity [118].

5.4 Hybrid Endonucleases

Creation of hybrid endonucleases provides a different avenue to expanding HE specificity, such that the chimeric enzymes recognize corresponding hybrid target sites. This approach has been useful for both LHEs [119–122] and GIY-YIG enzymes [65] (Fig. 3). As new LHE nucleases are identified, closely related enzymes (e.g. from the *OnuI* family) can be easily combined to refine recognition specificity [123] or redesigned LHEs can undergo domain swapping (e.g. between engineered I-DmoI and I-CreI) to further expand specificities [124] (Fig. 3a). More adventuresome chimeras, true to the origin of the term in Greek mythology, have been constructed, as exemplified by a variant with a catalytically inactive LHE I-SceI fused to the restriction enzyme *PvuII* as the cleavage module [125] (Fig. 3b). Other restriction enzymes, particularly *FokI*, have been used as the cleavage domain with both ZF nucleases and TALENs (Fig. 2b). These ZFN and TALEN modular technologies are currently marketed by Addgene, Sangamo Biosciences, and Sigma-Aldrich.

The need for dimer formation by *FokI* and *PvuII* restricts cleavage to symmetrical target sequences (Figs. 2b and 3b). This limitation has been relieved by designing monomeric hybrid enzymes with the catalytic domain of GIY-YIG HE I-TevI fused to both ZF and inactive LHE scaffolds (Fig. 3c). Both the Tev-ZF and Tev-LHE monomers can induce site-specific DSBs and induce recombination in yeast [126].

Coupling designer endonucleases *in trans* with DNA end-processing enzymes is a recent strategy employed to drive productive homologous DNA repair pathways, in favor of unproductive nonhomologous end-joining [127]. This approach with TREX2, the 3' repair

exonuclease, has yielded improved targeted gene disruption in several different cell lines with ZFNs, TALENs, and an engineered I-CreI LHE.

6 Applications of HEs

Designer nucleases, including the HEs, ZFNs, and TALENs, are being used to modify the genomes of viruses, bacteria, yeast, plant, insect, and human cells, and they are revolutionizing genome engineering. Our focus here is on the HEs, which allow insertion, deletion, single-site mutation, and correction, in a highly site-specific and controlled fashion (Fig. 1b). A brief comparison of these different genome engineering tools will follow (Table 2).

6.1 HEs as Therapeutic Agents

Therapies built around HEs are still in their infancy (reviewed by Silva et al. and Marcaida et al. [128, 129]). Although the hurdles facing this technology are considerable, control over HE-engineered cells is far superior to random transgenics, as with viral vectors that can integrate indiscriminately causing off-target events. However, there is a trade-off of the relative safety of HEs; that is, efficiency on the scale of viral vectors will require not only versatile nuclease engineering for site specificity but also enhanced homologous recombination frequency. By using the aforementioned TREX2-coupling, a sevenfold increase in gene disruption of the endogenous HIV coreceptor *CCR5* was observed with an engineered I-CreI LHE over the uncoupled nuclease [127]. Targeted disruption of the *CCR5* with a ZFN is being tested in clinical trials as a therapeutic modality against HIV [130]. Additionally, LHE I-AniI is being used in a new therapeutic approach to cure cells of latent HIV infection by targeted mutagenesis of essential viral genes of the provirus [131].

Soon after the discovery of meganuclease-induced DSB repair, LHEs were used to stimulate recombination in mammalian cells [132]. This led to the notion of repairing defective genes, with potential application to patients with monogenic diseases, and particularly those that can be treated ex vivo. Among these are blood disorders, including thalasemias, porphyria, hemophilia, leukemia, skin ailments such as Xeroderma pigmentosum and melanoma, and immunodeficiencies [129]. Progress has already been made with engineered I-CreI targeted for correction of the *XPC1* gene in cells from Xeroderma pigmentosum patients [133, 134]. Similar approaches have been used to tailor I-CreI for correction of mutation of the *RAG1* gene in severe combined immunodeficiency disease (SCID) [135, 136]. The relative safety of gene correction for primary immunodeficiencies is an attractive feature of HE mediated gene therapy [137]. Similarly an I-CreI variant has been engineered to correct a dystrophin gene in Duchenne muscular dystrophy [138].

6.2 HEs in Insect Vector Control

An exciting potential application of HEGs is to effect insect vector control by promoting the spread of engineered insects through populations, and thereby curtailing spread of such diseases as malaria. Attempts at spreading deleterious mutations through vector populations using HE technology have therefore been initiated. The feasibility of the method has been demonstrated in both a *Drosophila melanogaster* model and in *Anopheles gambia*

mosquitoes. In the fly model, using the LHE I-SceI it was shown that high rates of homing can be achieved within spermatogonia and in the female germline for successful deployment of a HEG-based gene drive strategy [139]. Likewise, it was demonstrated with the same HEG that a synthetic genetic element, consisting of mosquito regulatory regions, can increase transmission to progeny in transgenic *Anopheles* cage populations. Again, expression of I-SceI in the male germline induces high rates of site-specific gene conversion, resulting in the I-SceI gene acquisition that accounts for the observed genetic drive [140]. Similarly the I-PpoI gene in male mosquitoes was capable of introducing high levels of infertility in target populations in cage trials [141]. Through models of mosquito population genetics and malaria epidemiology combined with currently available HEG transmission data, it was concluded that HEG-based approaches could have a transformational effect on malaria control [142].

6.3 HEs in Agriculture

Recently, LHE I-CreI has been modified for agricultural applications in maize. The endonuclease gene was delivered to immature embryos to generate transgenic plants, with deletions and insertions detected at the HE cut site [143]. To improve prospects of nuclease-mediated improvement of plants, multigene plant transformation vectors have recently been constructed, with a cloning system based on ZFNs and HEs [144]. Viral vectors are also available for endonuclease delivery as a novel approach to plant engineering [145]. Thus HEs are being tailored for lofty applications in medicine, public health, and agronomy.

7 Comparison of Protein Targeting Technologies for Genome Engineering

TALENs (Fig. 2b) have been touted as “Genomic Cruise Missiles” as one of Science magazines breakthroughs of 2012. The striking structures of the TALENs explains their modular specificity [146, 147]. Nevertheless, TALENs like ZFNs and meganucleases must be engineered for each new DNA target, all with both targeting and cleavage constraints. The benefits and drawbacks of ZFNs, TALENs, and HEs for treatment of chronic viral infections such as HIV, hepatitis B, and Herpes Simplex viruses have been considered [148]. The difference among these three genome engineering tools is basically as follows (Table 2): First, the ZFNs and TALENs that exist as fusions to FokI generate DSBs with 5' extensions, whereas HEs generate nicks or 3' extensions. Second, since TALEN modules recognize single bases, they access broader sequence space than ZF modules which target nucleotide triplets, and as a result not all specific nucleotides are recognized. On the other hand, HEs are limited by the need to engineer naturally occurring enzymes, but these have an increasing cleavage repertoire as more enzymes are discovered. Third, these foregoing considerations result in higher cleavage specificity for the HEs and TALENs than for ZFNs. Fourth, context effects of ZFNs and HEs are more common than with TALENs [149]. Fifth, because of these specificity differences, off-target cleavage can be problematic with ZFNs and to a lesser extent with TALENs, whereas HEs tend to be most specific. Off-target effects not only can result in undesirable end-points, but also can induce toxicity in the cell. Sixth, targeting the modular ZFNs and TALENs is technically more straightforward than redirecting the specificity of HEs, where the DNA-binding and cleavage determinants reside in the same molecule. Finally, this coexistence of specificity and cleavage functions of the

HEs accounts for their compactness relative to the modular ZFNs and TALENs, a great advantage for application, particularly vectorization. For the above reasons, the pros and cons of the three types of protein endonucleases need to be considered for any particular application, and one needs to keep abreast of emerging subtleties, such as a recent report that associates different mutation profiles with ZFNs and TALENs [150].

8 RNA as a Player in Targeted Genome Editing

The challenge to protein endonuclease technology may come from readily programmable RNAs. Group II introns were the first RNAs to be used for gene targeting (reviewed by Lambowitz and Zimmerly and Cui and Davis [78, 151]). These introns themselves recognize their native DNA targets in an aforementioned retromobility reaction by an exon binding sequence (EBS) base-pairing with the DNA over ~14 bp. By reprogramming the intron EBS, highly site-specific retargeting occurs. Custom services for specific targetrons are available for use in several bacterial systems from both Sigma-Aldrich and Targetronics, although barriers remain in eukaryotes [78]. The targeting RNA must be in the form of a ribonucleoprotein, and entry into the nucleus remains the major obstacle. Nevertheless group II introns have the potential to not only integrate into DNA but also to make site-specific DSBs.

A new RNA-based approach for precise and efficient gene targeting that utilizes CRISPR/Cas9 derived from a bacterial immunity system [152, 153] is rapidly bursting on the scene [154–158]. The uniqueness of these RNA-guided endonucleases (RGENs) is based on designing guide RNAs that make the editing process highly flexible and facile, since the protein nuclease does not require engineering for retargeting. The bacterial CRISPR/Cas9 system has been successfully modified to accommodate mammalian and zebrafish transcription and translation requirements and kits are already being offered (by Addgene). Potential constraints of the CRISPR/Cas9 approach are the requirement for an adjacent proto-spacer sequence (an NGG triplet), genomic DNA accessibility due to chromatin and methylation states, and RNA secondary structure. However, a useful feature of the CRISPR/Cas9 engineering tool and one that distinguishes it from the other genome targeting technologies at this point is its utility for multiplexed editing. This versatility offers simultaneous change of several sites within a single genome.

Ironically, in both these cases of RNA-based targeting, the protein clippers are members of the HNH HE family. The technology is coming full circle, and developing at an extraordinary pace: >40 years since the genetic effects of ω were observed, 10 years since ZFNs were developed, 2 years since TALENs were offered to the world, and months since RGENs were discovered and spread through the literature, research, and corporate laboratories alike.

9 Summary

The HE field has indeed come a long way over 40 years, from the discovery of a robust recombination event that accounted for non-Mendelian inheritance in yeast mitochondria, through a detailed understanding of the structure and function of the endonuclease family responsible for the event, followed by clever manipulation of DNA cleavage enzymes used

to edit genomes in highly targeted fashion. Tailoring endonuclease specificity has found broad application that started with research in genome engineering of bacterial, fungal, and mammalian cells and is now being used in the fields of agriculture and human health. Endonucleases are not only being groomed as anti-viral agents, but also for gene therapy of monogenic diseases, while the prospect of controlling malaria mosquitoes in sub-Saharan Africa, using the very enzyme that led to the discovery of HEs, becomes ever more promising.

Acknowledgments

We thank Matt Stanger for preparing the figures, Barry Stoddard for providing the images for Fig. 1a, and Rebecca McCarthy for preparing the manuscript. Research in the Belfort lab is supported by NIH grants GM39422 and GM44844.

References

1. Netter P, Petrochilo E, Slonimski PP, Bolotin-Fukuhara M, Coen D, Deutsch J, Dujon B. Mitochondrial genetics. VII. Allelism and mapping studies of ribosomal mutants resistant to chloramphenicol, erythromycin and spiramycin in *S. cerevisiae*. *Genetics*. 1974; 78:1063–1100. [PubMed: 4281750]
2. Bolotin M, Coen D, Deutsch J, Dujon B, Netter P, Petrochilo E, Slonimski PP. La recombinaison des mitochondries chez *Saccharomyces cerevisiae*. *Bull Inst Pasteur*. 1971; 69:215–239.
3. Bos JL, Heyting C, Borst P, Arnberg AC, van Bruggen EFJ. An insert in the single gene for the large ribosomal RNA in yeast mitochondrial DNA. *Nature*. 1978; 275:336–338. [PubMed: 357991]
4. Faye G, Dennebouy N, Kujawa C, Jacq C. Inserted sequence in the mitochondrial 23S ribosomal RNA gene of the yeast *Saccharomyces cerevisiae*. *Mol Gen Genet*. 1979; 168:101–109. [PubMed: 372737]
5. Jacquier A, Dujon B. An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell*. 1985; 41:383–394. [PubMed: 3886163]
6. Lambowitz AM, Belfort M. Introns as mobile genetic elements. *Annu Rev Biochem*. 1993; 62:587–622. [PubMed: 8352597]
7. Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. The double-strand-break repair model for recombination. *Cell*. 1983; 33:25–35. [PubMed: 6380756]
8. Belfort M. Phage T4 introns: self-splicing and mobility. *Annu Rev Genet*. 1990; 24:363–385. [PubMed: 2088173]
9. Dalgaard JZ, Garrett RA, Belfort M. A site-specific endonuclease encoded by a typical archaeal intron. *Proc Natl Acad Sci USA*. 1993; 90:5414–5417. [PubMed: 8390663]
10. Kane PM, Yamashiro CT, Wolczyk DF, Neff N, Goebel M, Stevens TH. Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science*. 1990; 250:651–657. [PubMed: 2146742]
11. Hirata R, Oshumi Y, Nakano A, Kawasaki H, Suzuki K, Anraku Y. Molecular structure of a gene, VMA1, encoding the catalytic subunit of H⁺ -translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. *J Biol Chem*. 1990; 265:6726–6733. [PubMed: 2139027]
12. Dalgaard JZ, Klar A, Moser MJ, Holley WR, Chatterjee A, Mian IS. Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the H-N-H family. *Nucleic Acids Res*. 1997; 25:4626–4638. [PubMed: 9358175]
13. Sharma M, Ellis RL, Hinton DM. Identification of a family of bacteriophage T4 genes encoding proteins similar to those present in group I introns of fungi and phage. *Proc Natl Acad Sci USA*. 1992; 89:6658–6662. [PubMed: 1631169]

14. Edgell DR, Shub DA. Related homing endonucleases I-BmoI and I-TevI use different strategies to cleave homologous recognition sites. *Proc Natl Acad Sci USA*. 2001; 98:7898–7903. [PubMed: 11416170]
15. Mota EM, Collins RA. Independent evolution of structural and coding regions in a *Neurospora* mitochondrial intron. *Nature*. 1988; 332:654–656. [PubMed: 2965793]
16. Quirk SM, Bell-Pedersen D, Belfort M. Intron mobility in the T-Even phages: high frequency inheritance of group I introns promoted by intron open reading frames. *Cell*. 1989; 56:455–465. [PubMed: 2644046]
17. Bonocora RP, Shub DA. A likely pathway for formation of mobile group I introns. *Curr Biol*. 2009; 19:223–228. [PubMed: 19200727]
18. Macreadie IG, Scott RM, Zinn AR, Butow RA. Transposition of an intron in yeast mitochondria requires a protein encoded by that intron. *Cell*. 1985; 41:395–402. [PubMed: 3886164]
19. Taylor GK, Stoddard BL. Structural, functional and evolutionary relationships between homing endonucleases and proteins from their host organisms. *Nucleic Acids Res*. 2012; 40:5189–5200. [PubMed: 22406833]
20. Haber, JE.; Wolfe, KH. Function and evolution of HO and VDE endonucleases in fungi. In: Belfort, M.; Derbyshire, V.; Stoddard, BL.; Wood, DW., editors. *Homing endonucleases and inteins*. Springer; Berlin: 2005. p. 161-175.
21. Pietrokovski S. Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. *Protein Sci*. 1994; 3:2340–2350. [PubMed: 7756989]
22. Gibb EA, Edgell DR. Better late than never: delayed translation of intron-encoded endonuclease I-TevI is required for efficient splicing of its host group I intron. *Mol Microbiol*. 2010; 78:35–46. [PubMed: 20497330]
23. Stoddard BL, Belfort M. Social networking between mobile introns and their host genes. *Mol Microbiol*. 2010; 78:1–4. [PubMed: 20545861]
24. Goddard MR, Burt A. Recurrent invasion and extinction of a selfish gene. *Proc Natl Acad Sci USA*. 1999; 96:13880–13885. [PubMed: 10570167]
25. Gimble FS. Invasion of a multitude of genetic niches by mobile endonuclease genes. *FEMS Microbiol Lett*. 2000; 185:99–107. [PubMed: 10754232]
26. Gogarten JP, Hilario E. Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. *BMC Evol Biol*. 2006; 6:94. [PubMed: 17101053]
27. Liu Q, Belle A, Shub DA, Belfort M, Edgell DR. SegG endonuclease promotes marker exclusion and mediates co-conversion from a distant cleavage site. *J Mol Biol*. 2003; 334:13–23. [PubMed: 14596796]
28. Zeng Q, Bonocora RP, Shub DA. A free-standing homing endonuclease targets an intron insertion site in the psbA gene of cyanophages. *Curr Biol*. 2009; 19:218–222. [PubMed: 19200728]
29. Bonocora RP, Zeng Q, Abel EV, Shub DA. A homing endonuclease and the 50-nt ribosomal bypass sequence of phage T4 constitute a mobile DNA cassette. *Proc Natl Acad Sci USA*. 2011; 108:16351–16356. [PubMed: 21930924]
30. Loizos N, Tillier ERM, Belfort M. Evolution of mobile group I introns: recognition of intron sequences by an intron-encoded endonuclease. *Proc Natl Acad Sci USA*. 1994; 91:11983–11987. [PubMed: 7991569]
31. Edgell DR, Stanger MJ, Belfort M. Coincidence of cleavage sites of intron endonuclease I-TevI and critical sequences of the host thymidylate synthase gene. *J Mol Biol*. 2004; 343:1231–1241. [PubMed: 15491609]
32. Koufopanou V, Goddard MR, Burt A. Adaptation for horizontal transfer in a homing endonuclease. *Mol Biol Evol*. 2002; 19:239–246. [PubMed: 11861883]
33. Stoddard BL. Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. *Structure*. 2011; 19:7–15. [PubMed: 21220111]
34. Paques F, Duchateau P. Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Curr Gene Ther*. 2007; 7:49–66. [PubMed: 17305528]
35. Colleaux L, D'Auriol L, Galibert F, Dujon B. Recognition and cleavage site of the intron-encoded *omega* transposase. *Proc Natl Acad Sci USA*. 1988; 85:6022–6026. [PubMed: 2842757]

36. Bryk M, Quirk SM, Mueller JE, Loizos N, Lawrence C, Belfort M. The *td* intron endonuclease makes extensive sequence tolerant contacts across the minor groove of its DNA target. *EMBO J*. 1993; 12:2141–2149. [PubMed: 8491202]
37. Pingoud A, Jeltsch A. Structure and function of type II restriction endonucleases. *Nucleic Acids Res*. 2001; 29:3705–3727. [PubMed: 11557805]
38. Zhao L, Bonocora RP, Shub DA, Stoddard BL. The restriction fold turns to the dark side: a bacterial homing endonuclease with a PD-(D/E)-XK motif. *EMBO J*. 2007; 26:2432–2442. [PubMed: 17410205]
39. Jurica MS, Monnat RJ Jr, Stoddard BL. DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-*CreI*. *Mol Cell*. 1998; 2:469–476. [PubMed: 9809068]
40. Van Roey P, Waddling CA, Fox KM, Belfort M, Derbyshire V. Intertwined structure of the DNA-binding domain of intron endonuclease I-TevI with its substrate. *EMBO J*. 2001; 20:3631–3637. [PubMed: 11447104]
41. Silva G, Dalgaard JZ, Belfort M, Van Roey P. Crystal structure of the thermostable archaeal intron-encoded endonuclease I-*DmoI*. *J Mol Biol*. 1999; 286:1123–1136. [PubMed: 10047486]
42. Duan X, Gimble FS, Quioco FA. Crystal structure of PI-*SceI*, a homing endonuclease with protein splicing activity. *Cell*. 1997; 89:555–564. [PubMed: 9160747]
43. Ichihyanagi K, Ishino Y, Ariyoshi M, Komori K, Morikawa K. Crystal structure of an archaeal intron-encoded homing endonuclease PI-PfuI. *J Mol Biol*. 2000; 300:889–901. [PubMed: 10891276]
44. Heath PJ, Stephens KM, Monnat RJ Jr, Stoddard BL. The structure of I-*CreI*, a group I intron-encoded homing endonuclease. *Nat Struct Biol*. 1997; 4:468–476. [PubMed: 9187655]
45. Gimble FS, Stephens BW. Substitutions in conserved dodecapeptide motifs that uncouple the DNA binding and DNA cleavage activities of PI-*SceI* endonuclease. *J Biol Chem*. 1995; 270:5849–5856. [PubMed: 7890714]
46. Schottler S, Wende W, Pingoud V, Pingoud A. Identification of Asp218 and Asp326 as the principal Mg²⁺ binding ligands of the homing endonuclease PI-*SceI*. *Biochemistry*. 2000; 39:15895–15900. [PubMed: 11123916]
47. Chevalier BS, Monnat RJ Jr, Stoddard BL. The homing endonuclease I-*CreI* uses three metals, one of which is shared between the two active sites. *Nat Struct Biol*. 2001; 8:312–316. [PubMed: 11276249]
48. Chevalier B, Sussman D, Otis C, Noel AJ, Turmel M, Lemieux C, Stephens K, Monnat RJ Jr, Stoddard BL. Metal-dependent DNA cleavage mechanism of the I-*CreI* LAGLIDADG homing endonuclease. *Biochemistry*. 2004; 43:14015–14026. [PubMed: 15518550]
49. Klar AJ. The yeast mating-type switching mechanism: a memoir. *Genetics*. 2010; 186:443–449. [PubMed: 20940334]
50. Bolduc JM, Spiegel PC, Chatterjee P, Brady KL, Downing ME, Caprara MG, Waring RB, Stoddard BL. Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing factor. *Genes Dev*. 2003; 17:2875–2888. [PubMed: 14633971]
51. Ho Y, Kim SJ, Waring RB. A protein encoded by a group I intron in *Aspergillus nidulans* directly assists RNA splicing and is a DNA endonuclease. *Proc Natl Acad Sci USA*. 1997; 94:8994–8999. [PubMed: 9256423]
52. Wenzlau JM, Saldanha RJ, Butow RA, Perlman PS. A latent intron-encoded maturase is also an endonuclease needed for intron mobility. *Cell*. 1989; 56:421–430. [PubMed: 2536592]
53. Belfort M. Two for the price of one: a bifunctional intron-encoded DNA endonuclease-RNA maturase. *Genes Dev*. 2003; 17:2860–2863. [PubMed: 14665667]
54. Szczepanek T, Lazowska J. Replacement of two non-adjacent amino acids in the *S. cerevisiae* bi2 intron-encoded RNA maturase is sufficient to gain a homing-endonuclease activity. *EMBO J*. 1996; 15:3758–3767. [PubMed: 8670880]
55. Chatterjee P, Brady KL, Solem A, Ho Y, Caprara MG. Functionally distinct nucleic acid binding sites for a group I intron encoded RNA maturase/DNA homing endonuclease. *J Mol Biol*. 2003; 329:239–251. [PubMed: 12758073]

56. Geese WJ, Kwon YK, Wen X, Waring RB. *In vitro* analysis of the relationship between endonuclease and maturase activities in the bi-functional group I intron-encoded protein, I-AniI. *Eur J Biochem.* 2003; 270:1543–1554. [PubMed: 12654010]
57. Knizewski L, Ginalski K. Bacterial DUF199/COG1481 proteins including sporulation regulator WhiA are distant homologs of LAGLIDADG homing endonucleases that retained only DNA binding. *Cell Cycle.* 2007; 6:1666–1670. [PubMed: 17603302]
58. Kaiser BK, Clifton MC, Shen BW, Stoddard BL. The structure of a bacterial DUF199/WhiA protein: domestication of an invasive endonuclease. *Structure.* 2009; 17:1368–1376. [PubMed: 19836336]
59. Kaiser BK, Stoddard BL. DNA recognition and transcriptional regulation by the WhiA sporulation factor. *Sci Rep.* 2011; 1:156. [PubMed: 22355671]
60. Kowalski JC, Belfort M, Stapleton MA, Holpert M, Dansereau JT, Pietrokovski S, Baxter SM, Derbyshire V. Configuration of the catalytic GIY-YIG domain of intron endonuclease I-TevI: coincidence of computational and molecular findings. *Nucleic Acids Res.* 1999; 27:2115–2125. [PubMed: 10219084]
61. Dunin-Horkawicz S, Feder M, Bujnicki JM. Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics.* 2006; 7:98. [PubMed: 16646971]
62. Van Roey P, Meehan L, Kowalski J, Belfort M, Derbyshire V. Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI. *Nat Struct Biol.* 2002; 9:806–811. [PubMed: 12379841]
63. Derbyshire V, Kowalski JC, Dansereau JT, Hauer CR, Belfort M. Two-domain structure of the *td* intron-encoded endonuclease I-TevI correlates with the two-domain configuration of the homing site. *J Mol Biol.* 1997; 265:494–506. [PubMed: 9048944]
64. Liu QQ, Dansereau JT, Puttamadappa SS, Shekhtman A, Derbyshire V, Belfort M. Role of the interdomain linker in distance determination for remote cleavage by homing endonuclease I-TevI. *J Mol Biol.* 2008; 379:1094–1106. [PubMed: 18499124]
65. Liu Q-Q, Derbyshire V, Belfort M, Edgell DR. Distance determination by GIY-YIG intron endonucleases: discrimination between repression and cleavage functions. *Nucleic Acids Res.* 2006; 34:1755–1764. [PubMed: 16582101]
66. Bell-Pedersen D, Quirk SM, Bryk M, Belfort M. I-TevI, the endonuclease encoded by the mobile *td* intron, recognizes binding and cleavage domains on its DNA target. *Proc Natl Acad Sci USA.* 1991; 88:7719–7723. [PubMed: 1881913]
67. Belle A, Landthaler M, Shub DA. Intronless homing: site-specific endonuclease SegF of bacteriophage T4 mediates localized marker exclusion analogous to homing endonucleases of group I introns. *Genes Dev.* 2002; 16:351–362. [PubMed: 11825876]
68. Mak AN, Lambert AR, Stoddard BL. Folding, DNA recognition, and function of GIY-YIG endonucleases: crystal structures of R.Eco29kI. *Structure.* 2010; 18:1321–1331. [PubMed: 20800503]
69. Sokolowska M, Czapinska H, Bochtler M. Hpy188I-DNA pre- and post- cleavage complexes—snapshots of the GIY-YIG nuclease mediated catalysis. *Nucleic Acids Res.* 2011; 39:1554–1564. [PubMed: 20935048]
70. Pyatkov KI, Arkhipova IR, Malkova NV, Finnegan DJ, Evgen'ev MB. Reverse transcriptase and endonuclease activities encoded by Penelope-like retroelements. *Proc Natl Acad Sci USA.* 2004; 101:14719–14724. [PubMed: 15465912]
71. Edgell DR, Derbyshire V, Van Roey P, LaBonne S, Stanger MJ, Li Z, Boyd TM, Shub DA, Belfort M. Intron-encoded homing endonuclease I-TevI also functions as a transcriptional autorepressor. *Nat Struct Mol Biol.* 2004; 11:936–944. [PubMed: 15361856]
72. Ko T-P, Liao C-C, Ku W-Y, Chak K-F, Yuan HS. The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein. *Structure.* 1999; 7:91–102. [PubMed: 10368275]
73. Kleanthous C, Kuhlmann UC, Pommer AJ, Ferguson N, Radford SE, Moore GR, James R, Hemmings AM. Structural and mechanistic basis of immunity toward endonuclease colicins. *Nat Struct Biol.* 1999; 6:243–252. [PubMed: 10074943]

74. Friedhoff P, Franke I, Meiss G, Wende W, Krause KL, Pingoud A. A similar active site for non-specific and specific endonucleases. *Nat Struct Biol.* 1999; 6:112–113. [PubMed: 10048918]
75. Goodrich-Blair H, Shub DA. Beyond homing: competition between intron endonucleases confers a selective advantage on flanking genetic markers. *Cell.* 1996; 84:211–221. [PubMed: 8565067]
76. Landthaler M, Begley U, Lau NC, Shub DA. Two self-splicing group I intron in the ribonucleotide reductase large subunit of *Staphylococcus aureus* phage Twort. *Nucleic Acids Res.* 2002; 30:1935–1943. [PubMed: 11972330]
77. Robbins JB, Stapleton M, Smith D, Dansereau JT, Derbyshire V, Belfort M. Homing endonuclease I-TevIII: dimerization as a means to a double strand break. *Nucleic Acids Res.* 2007; 35:1589–1600. [PubMed: 17289754]
78. Lambowitz AM, Zimmerly S. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol.* 2011; 3:a003616. [PubMed: 20463000]
79. Gorbalenya AE. Self-splicing group I and group II introns encode homologous (putative) DNA endonucleases of a new family. *Protein Sci.* 1994; 3:1117–1120. [PubMed: 7920259]
80. Shub DA, Goodrich-Blair H, Eddy SR. Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. *Trends Biochem Sci.* 1994; 19:402–404. [PubMed: 7817395]
81. Raaijmakers H, Toro I, Birkenbihl R, Kemper B, Suck D. Conformational flexibility in T4 endonuclease VII revealed by crystallography: implications for substrate binding and cleavage. *J Mol Biol.* 2001; 308:311–323. [PubMed: 11327769]
82. Bujnicki JM, Radlinska M, Rychlewski L. Polyphyletic evolution of type II restriction enzymes revisited: two independent sources of second-hand folds revealed. *Trends Biochem Sci.* 2001; 26:9–11. [PubMed: 11165501]
83. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. *Nature.* 2012; 482:331–338. [PubMed: 22337052]
84. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA.* 2012; 109:E2579–E2586. [PubMed: 22949671]
85. Johansen S, Embley TM, Willassen NP. A family of nuclear homing endonucleases. *Nucleic Acids Res.* 1993; 21:4405. [PubMed: 8415009]
86. Flick KE, Jurica MS, Monnat RJ Jr, Stoddard BL. DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-*PpoI*. *Nature.* 1998; 394:96–101. [PubMed: 9665136]
87. Kuhlmann UC, Moore GR, James R, Kleanthous C, Hemmings AM. Structural parsimony in endonuclease active sites: should the number of homing endonuclease families be redefined? *FEBS Lett.* 1999; 463:1–2. [PubMed: 10601625]
88. Bujnicki JM. Crystallographic and bioinformatic studies on restriction endonucleases: inference of evolutionary relationships in the “midnight zone” of homology. *Curr Protein Pept Sci.* 2003; 4:327–337. [PubMed: 14529527]
89. Orlowski J, Boniecki M, Bujnicki JM. I-Ssp6803I: the first homing endonuclease from the PD-(D/E)XK superfamily exhibits an unusual mode of DNA recognition. *Bioinformatics.* 2007; 23:527–530. [PubMed: 17242028]
90. Bonocora RP, Shub DA. A novel group I intron-encoded endonuclease specific for the anticodon region of tRNA(fMet) genes. *Mol Microbiol.* 2001; 39:1299–1306. [PubMed: 11251845]
91. Zhao L, Pellenz S, Stoddard BL. Activity and specificity of the bacterial PD-(D/E)XK homing endonuclease I-Ssp6803I. *J Mol Biol.* 2009; 385:1498–1510. [PubMed: 19038269]
92. Steczkiewicz K, Muszewska A, Knizewski L, Rychlewski L, Ginalski K. Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. *Nucleic Acids Res.* 2012; 40:7016–7045. [PubMed: 22638584]
93. Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen JA, Hoffman JM, Remington K, et al. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* 2007; 5:e77. [PubMed: 17355176]
94. Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K, Eisen JA, Heidelberg KB, Manning G, Li W, et al. The Sorcerer II Global Ocean sampling expedition: expanding the universe of protein families. *PLoS Biol.* 2007; 5:e16. [PubMed: 17355171]

95. Taylor GK, Heiter DF, Pietrokovski S, Stoddard BL. Activity, specificity and structure of I-Bth0305I: a representative of a new homing endonuclease family. *Nucleic Acids Res.* 2011; 39:9705–9719. [PubMed: 21890897]
96. Dassa B, London N, Stoddard BL, Schueler-Furman O, Pietrokovski S. Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family. *Nucleic Acids Res.* 2009; 37:2560–2573. [PubMed: 19264795]
97. Porteus MH, Carroll D. Gene targeting using zinc finger nucleases. *Nat Biotechnol.* 2005; 23:967–973. [PubMed: 16082368]
98. Mussolino C, Morbitzer R, Lutge F, Dannemann N, Lahaye T, Cathomen T. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res.* 2011; 39:9283–9293. [PubMed: 21813459]
99. Chouluka A, Perrin A, Dujon B, Nicolas JF. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of *Saccharomyces cerevisiae*. *Mol Cell Biol.* 1995; 15:1968–1973. [PubMed: 7891691]
100. Rouet P, Smih F, Jasin M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol.* 1994; 14:8096–8106. [PubMed: 7969147]
101. Seligman LM, Chisholm KM, Chevalier BS, Chadsey MS, Edwards ST, Savage JH, Veillet AL. Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic Acids Res.* 2002; 30:3870–3879. [PubMed: 12202772]
102. Gruen M, Chang K, Serbanescu I, Liu DR. An *in vivo* selection system for homing endonuclease activity. *Nucleic Acids Res.* 2002; 30:e29. [PubMed: 11917035]
103. Gimble FS, Moure CM, Posey KL. Assessing the plasticity of DNA target site recognition of the PI-SceI homing endonuclease using a bacterial two-hybrid selection system. *J Mol Biol.* 2003; 334:993–1008. [PubMed: 14643662]
104. Chames P, Epinat JC, Guillier S, Patin A, Lacroix E, Paques F. In vivo selection of engineered homing endonucleases using double-strand break induced homologous recombination. *Nucleic Acids Res.* 2005; 33:e178. [PubMed: 16306233]
105. Jarjour J, West-Foyle H, Certo MT, Hubert CG, Doyle L, Getz MM, Stoddard BL, Scharenberg AM. High-resolution profiling of homing endonuclease binding and catalytic specificity using yeast surface display. *Nucleic Acids Res.* 2009; 37:6871–6880. [PubMed: 19740766]
106. Arnould S, Chames P, Perez C, Lacroix E, Duclert A, Epinat JC, Stricher F, Petit AS, Patin A, Guillier S, et al. Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J Mol Biol.* 2006; 355:443–458. [PubMed: 16310802]
107. Ashworth J, Havranek JJ, Duarte CM, Sussman D, Monnat R Jr, Stoddard BL, Baker D. Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature.* 2006; 441:656–659. [PubMed: 16738662]
108. Ulge UY, Baker DA, Monnat RJ Jr. Comprehensive computational design of mCreI homing endonuclease cleavage specificity for genome engineering. *Nucleic Acids Res.* 2011; 39:4330–4339. [PubMed: 21288879]
109. Ashworth J, Taylor GK, Havranek JJ, Quadri SA, Stoddard BL, Baker D. Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. *Nucleic Acids Res.* 2010; 38:5601–5608. [PubMed: 20435674]
110. Takeuchi R, Lambert AR, Mak AN, Jacoby K, Dickson RJ, Gloor GB, Scharenberg AM, Edgell DR, Stoddard BL. Tapping natural reservoirs of homing endonucleases for targeted gene modification. *Proc Natl Acad Sci USA.* 2011; 108:13077–13082. [PubMed: 21784983]
111. Jacoby K, Metzger M, Shen BW, Certo MT, Jarjour J, Stoddard BL, Scharenberg AM. Expanding LAGLIDADG endonuclease scaffold diversity by rapidly surveying evolutionary sequence space. *Nucleic Acids Res.* 2012; 40:4954–4964. [PubMed: 22334611]
112. Li H, Ulge UY, Hovde BT, Doyle LA, Monnat RJ Jr. Comprehensive homing endonuclease target site specificity profiling reveals evolutionary constraints and enables genome engineering applications. *Nucleic Acids Res.* 2012; 40:2587–2598. [PubMed: 22121229]

113. Mueller JE, Smith D, Bryk M, Belfort M. Intron-encoded endonuclease I-TevI binds as a monomer to effect sequential cleavage via conformational changes in the *td* homing site. *EMBO J*. 1995; 14:5724–5735. [PubMed: 8521829]
114. Carter JM, Friedrich NC, Kleinstiver B, Edgell DR. Strand-specific contacts and divalent metal ion regulate double-strand break formation by the GIY-YIG homing endonuclease I-BmoI. *J Mol Biol*. 2007; 374:306–321. [PubMed: 17936302]
115. Landthaler M, Shub DA. The nicking homing endonuclease I-BasI is encoded by a group I intron in the DNA polymerase gene of the *Bacillus thuringiensis* phage Bastille. *Nucleic Acids Res*. 2003; 31:3071–3077. [PubMed: 12799434]
116. Chan SH, Stoddard BL, Xu SY. Natural and engineered nicking endonucleases—from cleavage mechanism to engineering of strand-specificity. *Nucleic Acids Res*. 2011; 39:1–18. [PubMed: 20805246]
117. McConnell Smith A, Takeuchi R, Pellenz S, Davis L, Maizels N, Monnat RJ Jr, Stoddard BL. Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. *Proc Natl Acad Sci USA*. 2009; 106:5099–5104. [PubMed: 19276110]
118. Metzger MJ, McConnell-Smith A, Stoddard BL, Miller AD. Single-strand nicks induce homologous recombination with less toxicity than double-strand breaks using an AAV vector template. *Nucleic Acids Res*. 2011; 39:926–935. [PubMed: 20876694]
119. Chevalier BS, Kortemme T, Chadsey MS, Baker D, Monnat RJ, Stoddard BL. Design, activity, and structure of a highly specific artificial endonuclease. *Mol Cell*. 2002; 10:895–905. [PubMed: 12419232]
120. Epinat JC, Arnould S, Chames P, Rochaix P, Desfontaines D, Puzin C, Patin A, Zanghellini A, Paques F, Lacroix E. A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res*. 2003; 31:2952–2962. [PubMed: 12771221]
121. Silva GH, Belfort M. Analysis of the LAGLIDADG interface of the monomeric homing endonuclease I-DmoI. *Nucleic Acids Res*. 2004; 32:3156–3168. [PubMed: 15190132]
122. Silva GH, Belfort M, Wende W, Pingoud A. From monomeric to homodimeric endonucleases and back: engineering novel specificity of LAGLIDADG enzymes. *J Mol Biol*. 2006; 361:744–754. [PubMed: 16872628]
123. Baxter S, Lambert AR, Kuhar R, Jarjour J, Kulshina N, Parmeggiani F, Danaher P, Gano J, Baker D, Stoddard BL, et al. Engineering domain fusion chimeras from I-OnuI family LAGLIDADG homing endonucleases. *Nucleic Acids Res*. 2012; 40:7985–8000. [PubMed: 22684507]
124. Grizot S, Epinat JC, Thomas S, Duclert A, Rolland S, Paques F, Duchateau P. Generation of redesigned homing endonucleases comprising DNA-binding domains derived from two different scaffolds. *Nucleic Acids Res*. 2010; 38:2006–2018. [PubMed: 20026587]
125. Fonfara I, Curth U, Pingoud A, Wende W. Creating highly specific nucleases by fusion of active restriction endonucleases and catalytically inactive homing endonucleases. *Nucleic Acids Res*. 2012; 40:847–860. [PubMed: 21965534]
126. Kleinstiver BP, Wolfs JM, Kolaczyk T, Roberts AK, Hu SX, Edgell DR. Monomeric site-specific nucleases for genome editing. *Proc Natl Acad Sci USA*. 2012; 109:8061–8066. [PubMed: 22566637]
127. Certo MT, Gwiazda KS, Kuhar R, Sather B, Curinga G, Mandt T, Brault M, Lambert AR, Baxter SK, Jacoby K, et al. Coupling endonucleases with DNA end-processing enzymes to drive gene disruption. *Nat Methods*. 2012; 9:973–975. [PubMed: 22941364]
128. Silva G, Poirot L, Galetto R, Smith J, Montoya G, Duchateau P, Paques F. Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Curr Gene Ther*. 2011; 11:11–27. [PubMed: 21182466]
129. Marcaida MJ, Munoz IG, Blanco FJ, Prieto J, Montoya G. Homing endonucleases: from basics to therapeutic applications. *Cell Mol Life Sci*. 2010; 67:727–748. [PubMed: 19915993]
130. Cannon P, June C. Chemokine receptor 5 knockout strategies. *Curr Opin HIV AIDS*. 2011; 6:74–79. [PubMed: 21242897]

131. Aubert M, Ryu BY, Banks L, Rawlings DJ, Scharenberg AM, Jerome KR. Successful targeting and disruption of an integrated reporter lentivirus using the engineered homing endonuclease Y2 I-AniI. *PLoS One*. 2011; 6:e16825. [PubMed: 21399673]
132. Jasin M. Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet*. 1996; 12:224–228. [PubMed: 8928227]
133. Arnould S, Perez C, Cabaniols JP, Smith J, Gouble A, Grizot S, Epinat JC, Duclert A, Duchateau P, Paques F. Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *J Mol Biol*. 2007; 371:49–65. [PubMed: 17561112]
134. Redondo P, Prieto J, Munoz IG, Alibes A, Stricher F, Serrano L, Cabaniols JP, Daboussi F, Arnould S, Perez C, et al. Molecular basis of xeroderma pigmentosum group C DNA recognition by engineered meganucleases. *Nature*. 2008; 456:107–111. [PubMed: 18987743]
135. Grizot S, Smith J, Daboussi F, Prieto J, Redondo P, Merino N, Villate M, Thomas S, Lemaire L, Montoya G, et al. Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. *Nucleic Acids Res*. 2009; 37:5405–5419. [PubMed: 19584299]
136. Munoz IG, Prieto J, Subramanian S, Coloma J, Redondo P, Villate M, Merino N, Marenchino M, D'Abramo M, Gervasio FL, et al. Molecular basis of engineered meganuclease targeting of the endogenous human RAG1 locus. *Nucleic Acids Res*. 2011; 39:729–743. [PubMed: 20846960]
137. Pessach IM, Notarangelo LD. Gene therapy for primary immunodeficiencies: looking ahead, toward gene correction. *J Allergy Clin Immunol*. 2011; 127:1344–1350. [PubMed: 21440291]
138. Chapdelaine P, Pichavant C, Rousseau J, Paques F, Tremblay JP. Meganucleases can restore the reading frame of a mutated dystrophin. *Gene Ther*. 2010; 17:846–858. [PubMed: 20393509]
139. Chan YS, Naujoks DA, Huen DS, Russell S. Insect population control by homing endonuclease-based gene drive: an evaluation in *Drosophila melanogaster*. *Genetics*. 2011; 188:33–44. [PubMed: 21368273]
140. Windbichler N, Menichelli M, Papathanos PA, Thyme SB, Li H, Ulge UY, Hovde BT, Baker D, Monnat RJ Jr, Burt A, et al. A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature*. 2011; 473:212–215. [PubMed: 21508956]
141. Klein TA, Windbichler N, Deredec A, Burt A, Benedict MQ. Infertility resulting from transgenic I-PpoI male *Anopheles gambiae* in large cage trials. *Pathog Glob Health*. 2012; 106:20–31. [PubMed: 22595271]
142. Deredec A, Godfray HC, Burt A. Requirements for effective malaria control with homing endonuclease genes. *Proc Natl Acad Sci USA*. 2011; 108:E874–E880. [PubMed: 21976487]
143. Gao H, Smith J, Yang M, Jones S, Djukanovic V, Nicholson MG, West A, Bidney D, Falco SC, Jantz D, et al. Heritable targeted mutagenesis in maize using a designed endonuclease. *Plant J*. 2010; 61:176–187. [PubMed: 19811621]
144. Zeevi V, Liang Z, Arieli U, Tzfira T. Zinc finger nuclease and homing endonuclease-mediated assembly of multigene plant transformation vectors. *Plant Physiol*. 2012; 158:132–144. [PubMed: 22082504]
145. Vainstein A, Marton I, Zuker A, Danziger M, Tzfira T. Permanent genome modifications in plant cells by transient viral vectors. *Trends Biotechnol*. 2011; 29:363–369. [PubMed: 21536337]
146. Mak AN, Bradley P, Cernadas RA, Bogdanove AJ, Stoddard BL. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science*. 2012; 335:716–719. [PubMed: 22223736]
147. Deng D, Yan C, Pan X, Mahfouz M, Wang J, Zhu JK, Shi Y, Yan N. Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science*. 2012; 335:720–723. [PubMed: 22223738]
148. Schiffer JT, Aubert M, Weber ND, Mintzer E, Stone D, Jerome KR. Targeted DNA mutagenesis for the cure of chronic viral infections. *J Virol*. 2012; 86:8920–8936. [PubMed: 22718830]
149. Meckler JF, Bhakta MS, Kim MS, Ovadia R, Habrian CH, Zykovich A, Yu A, Lockwood SH, Morbitzer R, Elsaesser J, et al. Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Res*. 2013; 41:4118–4128. [PubMed: 23408851]
150. Kim Y, Kweon J, Kim JS. TALENs and ZFNs are associated with different mutation signatures. *Nat Methods*. 2013; 10:185. [PubMed: 23396284]

151. Cui X, Davis G. Mobile group II intron targeting: applications in prokaryotes and perspectives in eukaryotes. *Front Biosci.* 2007; 12:4972–4985. [PubMed: 17569624]
152. Burgess DJ. Technology: a CRISPR genome-editing tool. *Nat Rev Genet.* 2013; 14:80. [PubMed: 23322222]
153. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* 2012; 337:816–821. [PubMed: 22745249]
154. Mali P, Yang L, Esvelt KM, Aach J, Guell M, Dicarlo JE, Norville JE, Church GM. RNA-guided human genome engineering via Cas9. *Science.* 2013; 339:823–826. [PubMed: 23287722]
155. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science.* 2013; 339:819–823. [PubMed: 23287718]
156. Cho SW, Kim S, Kim JM, Kim JS. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol.* 2013; 31:230–232. [PubMed: 23360966]
157. Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol.* 2013; 31:227–229. [PubMed: 23360964]
158. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol.* 2013; 31:223–239.

Fig. 1.

Endonuclease-mediated DNA repair. **(a)** HE-induced DSB mediates homing. The HEG (*blue pacman symbol*), often encoded by an intron or intein (*blue bar*), cleaves an uninterrupted homing site. The DNA ends of the cleaved recipient (*red*) engage in double-strand-break repair (DSBR), such that the cut allele is repaired by gene conversion with homologous intact DNA. **(b)** Gene targeting by an exogenous endonuclease. An engineered endonuclease (*blue and red*) cleaves a desired target, which can be repaired with an allele that inserts or corrects DNA using a homologous template (*left*), or is repaired by nonhomologous end joining (NHEJ), an error-prone process, that inactivates the cleaved gene

Fig. 2. Endonuclease–DNA interactions. **(a)** Five families of HEs are shown with examples indicated in parenthesis: LAGLIDADG (I-CreI), GIY-YIG (I-TevI), HNH (I-HmuI), His-Cys Box (I-PpoI) and PD(D/E)xK (I-Ssp6803I). I-CreI binds DNA as a homodimer, while other examples of the family are monomeric (*cartoons below*). Arrowheads point to the LAGLIDADG helixes. The I-TevI binding domain is shown on DNA, whereas the catalytic domain is freestanding. A model of full-length GIY-YIG HE binding is shown in *cartoon form below the structure*. **(b)** Synthetic endonucleases, ZFNs and TALENs, are fusions

between the respective repeated binding units and the FokI catalytic domain, which dimerizes on the DNA target

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 3. Homing endonuclease fusions. **(a)** LAGLIDADG Fusion. An I-Dmol/I-CreI fusion is depicted cleaving a hybrid target site. **(b)** I-SceI-PvuII fusion is shown. **(c)** GIY-YIG fusions. The catalytic domain of I-TevI is fused with I-OnuI or zinc-finger units as DNA binding modules

Table 1

Homing endonucleases described in this chapter

Family	Member	Characteristics	Origin	Notable features	Related host functions
LAGLIDADG	⊙ I-SceI	Monomer	<i>Saccharomyces cerevisiae</i> mitochondria	First HE to be discovered	HO mating-type endonuclease
	HO	Monomer	<i>Cerevisiae</i> nucleus	Mating-type switch	Intron-encoded maturases
	I-CreI	Homodimer	<i>Chlamydomonas reinhardtii</i> chloroplast	First LHE structure solved	Bacterial transcriptional regulators
				Scaffold for retargeting	
GIY-YIG	I-MsoI	Homodimer	<i>Monomastix</i> chloroplast	Scaffold for retargeting	
	I-Anil	Monomer	<i>Aspergillus nidularis</i> mitochondria	Nicking variant isolated	
	I-OnuI	Homodimer	<i>Ophiostoma novo-ulmi</i>	Family of closely related enzymes	
	I-TevI	Modular DNA binding and catalytic domains	Coliphage T4	First GIY-YIG structure solved	Restriction enzymes
	I-BmI	Modular	<i>Bacillus mojavensis</i>	Catalytic domain used in fusions	Penelope retroelements
				Similar catalytic domain to I-TevI	Eukaryotic flap endonuclease
					Eukaryotic Slx1-Slx4 resolvase
					Bacterial UvrC excision repair
PD(D/E)xK	I-Ssp6803I	Tetramer	<i>Synechocystis</i>	Motif also in recombinases, resolvases and DNA repair enzymes	Restriction enzyme
					Enzymes of recombination, tRNA splicing, DNA resolution
His-Cys Box	I-PpoI	Bends DNA	<i>Physarum polycephalum</i> nucleus	Colicin-like	
HNH	I-HnuI	Nicking enzyme	<i>B. subtilis</i> phage SP01	Recognizes intron and intronless targets	Group II intron endonuclease
	I-HnuII	Nicking enzyme	<i>B. subtilis</i> phage SP82	Excludes SPO1 DNA polymerase locus in mixed infection	Cas9 CRISPR nuclease
	I-TevIII	Makes DSB	Coliphage RB3	Makes DSB by dimerization	Bacterial colicins
	F-TsII	Nicking enzyme	Coliphage T3	Freestanding, downstream of gene 5	Restriction enzymes
	I-TsII	Nicking enzyme	Coliphage ϕ I	In intron within gene 5	Resolvase

Table 2

Comparison of protein-targeted genome engineering

	HE	ZFN	TALEN
DNA ends at DSB	3' extension	5' extension	5' extension
Sequence availability	Limited by existing enzymes	Limited by triplet recognition of ZFs	Limited by first nucleotide (T)
Context effects	Common	Common	Uncommon
Specificity	High	Intermediate	High
Off-target effects	Low	Intermediate	Low
Enzyme engineering	Complex	Intermediate	Straightforward
Size compactness (modularity)	Small (monomeric)	Large (monomeric and dimeric)	Very large (dimeric)