*Chapter 6*

# Advances in Engineering Homing Endonucleases for Gene Targeting: Ten Years After Structures

Barry L Stoddard*,†, Andrew M Scharenberg‡ and Raymond J Monnat, Jr§

Homing endonucleases (HEs) are highly site-specific endonucleases that induce homologous recombination or gene conversion *in vivo* by cleaving long (typically >18 bp) DNA target sites. Homing endonucleases are under development as tools for targeted genetic engineering applications, ranging from therapeutic gene correction to metabolic and population engineering. The first structures of homing endonucleases were reported 10 years ago. Since that time, representative structures from each of the known families of homing endonucleases have been determined, and the corresponding details of their mechanisms of DNA recognition and cleavage have been elucidated. Using this information, the LAGLIDADG homing endonuclease family has been identified as the most tractable for further modification by structure-based selection and/or engineering approaches. Most recently, successful redesign of the I-CreI endonuclease has led to the development of reagents that recognize and act on genes associated with monogenic diseases, including the human *RAG1* and *XPC* genes. These studies demonstrate the feasibility of using engineered homing endonucleases to promote efficient and target site-specific modification of chromosomal loci. Current studies are rapidly improving the throughput and efficiency of homing endonuclease design and selection, and aim to optimize the specificity and activity of the resulting endonucleases for targeted genomic applications in medicine and biotechnology.

*Keywords*: Homing endonuclease; LAGLIDADG; recombination; gene therapy; gene correction; protein engineering; genome engineering; DNA recognition.

*Corresponding author. bstoddar@fhcrc.org
†Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle WA 981089, USA
‡Department of Pediatrics, University of Washington Seattle WA 98195 and Children's Hospital and Regional Medical Center, Seattle WA 98109, USA
§Departments of Pathology and Genome Sciences, University of Washington, Seattle WA 98195, USA

# 1. Genomic Targeting: From Monogenic Diseases to Broad Genome Engineering

## 1.1. *Monogenic Disorders: A Case for Targeted Gene Repair*

Over the last decade, analyses of US and global public health issues have highlighted the increasing significance of monogenic diseases to the total health care burden, particularly for children (Baird *et al.*, 1988; Christianson *et al.*, 2006; McCandless *et al.*, 2004). The individual and societal costs imposed by medical interventions and treatment of monogenic disorders have compelled a search for curative therapeutic options for these diseases, focusing on correcting the underlying genetic defect in a patients' cells. A variety of therapies based on either of two broad approaches have been developed: (1) allogeneic stem cell transplantation, a strategy aimed at replacing a patient's defective hematopoietic stem cells with those of a normal donor; and (2) "traditional" gene therapy. In the latter case, a wild-type version of a defective gene is integrated into the genomes of a patient's cells in order to replace the function of a defective gene. In the case of diseases affecting lymphohematopoietic development and function, patient-derived hematopoietic stem cell populations can be transfected *in vitro* with the wild-type gene and then reintroduced.

The current practice of gene replacement therapy has several attendant issues. First, gene therapy involves the random insertion of foreign DNA into the genomes of stem cells, potentially resulting in the inactivation or activation of endogenous genes. Widely publicized examples of this type of problem include the development of T-cell lymphomas in four of 10 patients treated in France for X-SCID, and the development of oligoclonal proliferation of neutrophils in two of three patients treated in Germany for chronic granulomatous disease (CGD) (Abbott, 2006; Baum, 2007; Hacein-Bey-Abina *et al.*, 2003). In each of these cases, cellular proliferation appears to have been driven by retroviral insertion near a growth-related gene (LMO2 for the T-cell lymphomas in X-SCID, and MDS1/EV1 for clonal proliferations in CGD patients). In addition, liver tumors have developed following lentiviral vector-mediated factor IX hepatic gene transfer in mice (Themis *et al.*, 2005). While these results are troubling, it is still important to note that four of the eight patients who have been successfully cured in the French X-SCID trial are still doing well, and all the patients from a second X-SCID trial performed in London are well and free from adverse events (Baum, 2007; Thrasher *et al.*, 2006). In addition, new lentiviral vectors avoid the use of highly active LTR-based promoters, and thus may improve safety profiles



(Griesenbach *et al.*, 2006); however, others remain more pessimistic (Porteus, 2006).

A second issue is that it is desirable in all cases, and required in many, to use lineage-specific transcriptional control elements. However, defining such control elements is non-trivial, and may require years of experimentation. For the treatment of hemogoblinopathies such as thalassemias, native hemoglobin regulatory elements are too large to be efficiently delivered together with the relevant coding regions to hematopoietic stem cells. It has taken an extraordinary effort, by many laboratories, to define smaller transcriptional regulatory elements and to devise ways to deliver these with the current generation viral vectors. One recent successful development made use of a knowledge of $\beta$-globin transcriptional regulation together with lentiviral delivery technology to express artificial hemoglobin genes with apparently strict transcriptional control in murine hematopoietic stem cells (Malik and Arumugam, 2005; Malik *et al.*, 2005; Puthenveetil *et al.*, 2004). The extension of these approaches to large animal models remains to be demonstrated.

A final issue with gene replacement therapy is that it is best suited to treat diseases due to the complete deficiency or absence of a gene product. Gene replacement therapy is less well-suited or may not be applicable to diseases caused by the presence of an aberrantly functioning protein that may interfere with the function of the replacement normal protein. These diseases may require a more complicated approach in which simultaneous gene replacement and the suppression or knockdown of the defective gene product may be required. One example of a disease where this combined approach may be required is sickle cell disease (Sadelain, 2006; Samakoglu *et al.*, 2006). However, gene knockdown strategies involve highly conserved cellular RNAi mechanisms whose function is required for normal hematopoietic development. The long-term safety of this type of combined approach is an open question.

## 1.2. *Genome Engineering: A Targeted Approach*

A potential solution to some of the drawbacks and hurdles facing traditional gene replacement therapy is the possibility of "genome engineering," which describes an emerging discipline in which genomes of target organisms or cells are manipulated *in vivo*, using site-specific recombination to alter or add genetic information. The concept of genome engineering is not new, and dates back to experiments in the late 1970s in which ectopic DNA could be incorporated into the genome of the budding yeast *Saccharomyces*

*cerevisiae*. The success of this approach depends on endogenous homologous recombination pathways (Hinnen *et al.*, 1978; Orr-Weaver *et al.*, 1981). Depending on the exact methodology, individual yeast genes can be efficiently incorporated, deleted, mutated or corrected. However, while homologous recombination is extremely efficient in yeast, in mammalian cells it occurs at a very low frequency, often in the range of $10^{-5}$ to $10^{-7}$ per transformed cells (Doetschman *et al.*, 1987; Koller and Smithies, 1989). As we describe below, this limitation can in part be overcome by using a highly site-specific endonuclease to cleave the donor or recipient locus to stimulate targeted homologous recombination. The development of these reagents has allowed the field of genome engineering to progress dramatically over the past five years, together with the pursuit of several specific genome engineering applications (Bullard and Weaver, 2002; Glaser *et al.*, 2005; Gouble *et al.*, 2006; Tzfira and White, 2005; Vasquez *et al.*, 2001):

- *Synthetic biology.* DNA synthesis technology now allows the assembly of large artificial DNA fragments — up to several tens of thousands of base pairs — which are of sufficient size to encode entire metabolic pathways. Embedding these new pathways in a "minimal" genome or other model organism for the purpose of creating new types of synthetic organisms is the goal of the emerging science of synthetic biology (Ghatge *et al.*, 2006; Posfai *et al.*, 1999). The use of site-specific nucleases or recombinases to embed synthetic genes at specific desired target sites in model organisms represents a crucial enabling technology for synthetic biologists to create, manipulate, and control artificial genomes.
- *Pest control.* Synthetic genes encoding artificial nucleases may be used to create "selfish" genetic elements with the ability to integrate into and alter target genes while promoting their own transmission. This type of genetic drive system can strongly bias Mendelian inheritance to favor the generation of progeny that contain both the selfish genetic element and the altered target gene(s) (Burt, 2003). In modeling experiments, complete penetration of such a "hyper-dominant" allele through a population can occur in fewer than 20 generations. This strategy has been proposed as a novel means for genetic control of *Anopheles*-mediated malaria transmission by dominant transmission and inheritance of traits corresponding to resistance against *Plasmodium* infection, or by reducing the lifespan or reproductive fitness of the insect host (Chase, 2006). Practical evaluation and implementation of such a strategy are presently the subjects of a Gates Grand Challenges project focused on the control of malaria and the malaria vector *Anopheles gambiae* (Burt, 2005).
- *Gene repair.* By using standard transfection methods to introduce a site-specific nuclease and a modifying genomic template into primary lymphocytes, it has been possible to modify up to ≈5% of the target locus in the transfected cell population (Urnov *et al.*, 2005). This approach, when used to modify a mutant gene so as to restore normal function, is often termed "gene repair" or "gene correction." While gene repair has the same goal as traditional gene therapy approaches — restoration of the expression of a normally functioning protein — it has many advantages (Bertolotti, 1996). Since the endogenous gene's function is restored, the protein is expressed under the control of its natural regulatory elements, thus eliminating potential problems with inappropriate or inadequate expression of a transgene or transgene silencing. By targeting the repair with high efficiency to a single mutant locus, gene repair may also be able to dramatically reduce mutagenesis due to random insertions at other genomic locations.



## 2. Double Strand Break-Induced Gene Conversion and Gene-Specific Nucleases

Several different technologies have been developed to promote efficient targeted gene correction in mammalian cells. These include gene-targeted triplex forming oligonucleotides and hybrid RNA-DNA oligonucleotides (Kolb *et al.*, 2005) and the use of highly site-specific recombinases and transposases (Coates *et al.*, 2005). Each of these approaches has limitations related to the range of sequences that can be targeted (e.g., triplex-forming oligonucleotides), or the requirement for prior introduction of a target site (e.g., for site-specific recombinase-mediated targeting). Potentially the most versatile of all genome engineering technologies are those that make use of DNA double strand break-targeted homologous recombination for gene modification. Recent achievements (Urnov *et al.*, 2005; Arnould *et al.*, 2007) indicate that this method allows a desired genomic sequence to be altered in a precise manner, without the requirement for a selection marker or the introduction of additional exogenous DNA sequence(s).

Double strand break-targeted recombination requires the introduction or expression of a site-specific endonuclease in cells to generate a DNA double strand break at or near the desired modification site, together with the presence of a DNA repair template. Repair templates typically flank

the DNA double strand break site and include sequence modifications to be incorporated upon repair. Homologous recombination is a normal DNA repair pathway that makes use of homologous donor DNA to repair strand breaks. This pathway is highly conserved, and is used in many cells and organisms to repair DNA double strand breaks with high efficiency and fidelity or accuracy. The machinery of homologous recombination makes use of the DNA double strand break ends, together with the repair template, to incorporate DNA sequence modification(s) in the repair template into the chromosome as part of the process of DNA break repair.

A significant practical barrier to a widespread application of this accurate and efficient gene repair mechanism in genome engineering has been the requirement for endonucleases that are able to induce DNA double strand breaks at specific chromosomal target sites. Until very recently the technology to accomplish this has been lacking. However, over the past several years, two different approaches to creating enzymes capable of inducing highly site-specific DNA double strand breaks have been developed: *zinc finger nucleases* (ZFNs) and *homing endonucleases* (HEs) (Fig. 1).

ZFN site-specific nuclease technology involves the creation of artificial nucleases by appending a non-specific nuclease domain (such as the catalytic domain of the FokI restriction endonuclease) to a DNA-recognition and binding construct consisting of a tandem array of zinc fingers (Bibikova *et al.*, 2001; Porteus, 2006; Smith *et al.*, 2000). As individual zinc fingers recognize DNA triplets within the context of long cognate target sites (Beerli and Barbas, 2002; Bulyk *et al.*, 2001; Segal *et al.*, 1999), the concatenation of a series of zinc fingers of defined triplet specificity provides the possibility to create ZFNs able to bind and cleave at rare DNA targets. ZFNs have been demonstrated to induce gene correction/modification in both Drosophila and mammalian cells (Bibikova *et al.*, 2003; Porteus and Baltimore, 2003), and the highly efficient correction of disease-associated mutations in the human *IL2Rγ* gene (Urnov *et al.*, 2005).

Zinc finger nucleases have the important advantage of some capacity for modular design, and therefore ZFN technology has been the subject of intensive study over the past ten years (Porteus, 2006). While ZFN technology is clearly useful in designing nucleases able to cleave at diverse target sites, it also has several important limitations (Segal *et al.*, 2003). The main one results from the fact that DNA-recognition by individual zinc fingers is context-dependent: the identity of neighboring zinc fingers and target DNA sequences strongly influences their specificity and affinity towards



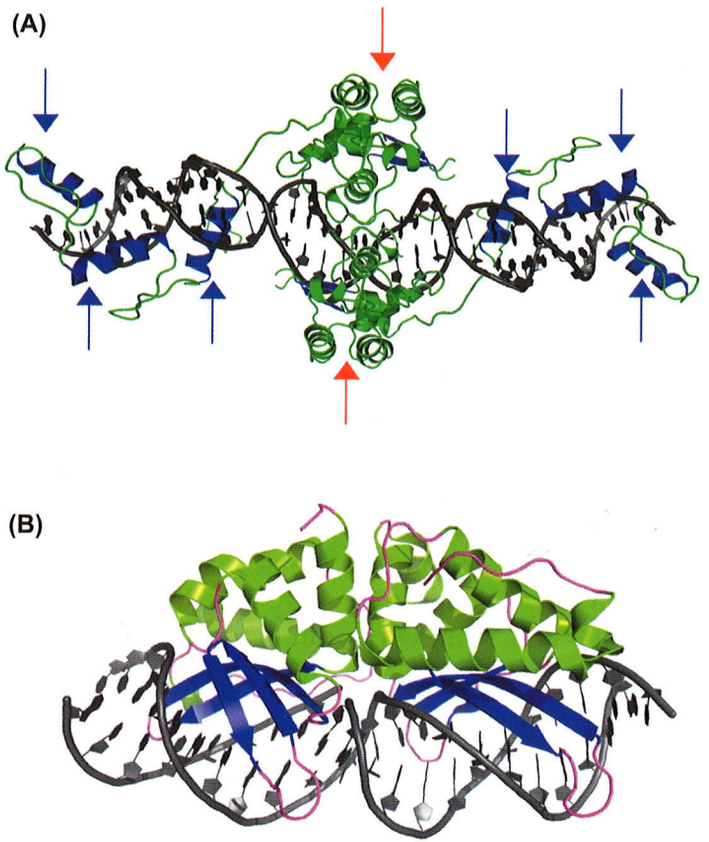


**Fig. 1.** **(A)** A hypothetical model of a zinc finger nuclease dimer bound to a DNA target site. Each subunit consists of a tandem series of three zinc fingers (blue arrows) tethered to the nuclease domain of the FokI restriction enzyme (red arrows). The model shown is based on the crystal structure of the Zif 268 zinc finger-DNA cocrystal structure (PDB code 1AAY) and the crystal structure of the FokI endonuclease bound to its DNA target (PDB code 1FOK). As structure of a zinc finger nuclease chimera has not yet been solved, the model is for illustrative purposes only. In the active complex, it is known that the nuclease domains form a closely associated transient dimer interface during double strand cleavage. The individual zinc fingers are each associated with three basepairs of the DNA target site (18 bases total). The two cognate binding sites for these modules are separated by a nonspecific sequence of DNA that harbors the sites of cleavage by the nuclease domains. **(B)** The crystal structure of a monomeric LAGLIDADG homing endonuclease (I-AniI) bound to its 19 basepair cognate target site. The active sites and cleaved phosphates are located at the center of the endonuclease-DNA complex; the target site is 19 basepairs in length.

"cognate" DNA codons. Therefore, engineering is not a simple modular process of appending zinc fingers with the appropriate triplet recognition specificity — multiple steps of selection and optimization are required. Indeed, the exquisite specificity and affinity of clinical grade reagents is

a tedious and time-consuming process and is essential for the genesis of efficient toxic-free zinc finger nucleases (e.g., Urnov *et al.*, 2005). Therefore, modular assemby of designed polydactyl zinc-finger DNA binding domains is an illusive advantage when compared to emerging customization of clinical grade homing endonucleases.

Homing endonucleases are naturally occurring, highly site-specific DNA endonucleases that are encoded as open reading frames embedded within introns or inteins (Fig. 2) (Belfort and Roberts, 1997; Stoddard, 2005). First discovered in the 1980s as part of the mobile genetic elements in yeast (Jacquier and Dujon, 1985; Kostriken *et al.*, 1983), homing endonucleases promote the mobility of introns or inteins that contain the endonuclease open reading frame by generating a site-specific DNA double strand break in a homologous gene or allele that lacks the intron or intein (Colleaux *et al.*, 1986). DNA break repair leads to lateral transfer of the element via homologous recombination to the cleaved allele, using the intron- or intein-containing allele as a repair template. Thus, homing endonuclease genes (HEGs) are selfish DNA sequences that are inherited in a dominant, non-Mendelian manner, and represent nature's application of double strand break-induced gene modification. Since their discovery, one of the first well characterized homing endonucleases (I-SceI protein from budding yeast) has been widely used to promote DNA double strand break-induced recombination in a wide variety of genomes. This work demonstrates the general utility of using homing endonucleases to induce site-specific recombination for gene repair, though with wide variations in the efficiency of repair and length of the repair template (commonly referred to as the "conversion tract") that is incorporated into the repaired allele (reviewed in Pâques and Duchateau, 2007).

Homing endonucleases are widespread and are found encoded within introns and inteins in all biological super-Kingdoms. Based on primary sequence homology, five homing enzyme families have been identified (Fig. 3), each primarily associated with a unique location in the host genomes: the LAGLIDADG endonucleases (found in archaea and fungal and algal organellar genomes), His-Cys Box family (found in protist nuclear genomes), the HNH and GIY-YIG endonucleases (found in bacteriophage) and the PD-(D/E)XK family, similar to archael Holliday junction resolvases and type II restriction endonucleases, as typified by I-SspI (found in bacterial genomes) (Stoddard, 2005). The His-Cys Box and HNH family appear to be descended from a common ancestral origin (which also gave rise to bacterial colicins and many additional DNA modifying enzymes), while the



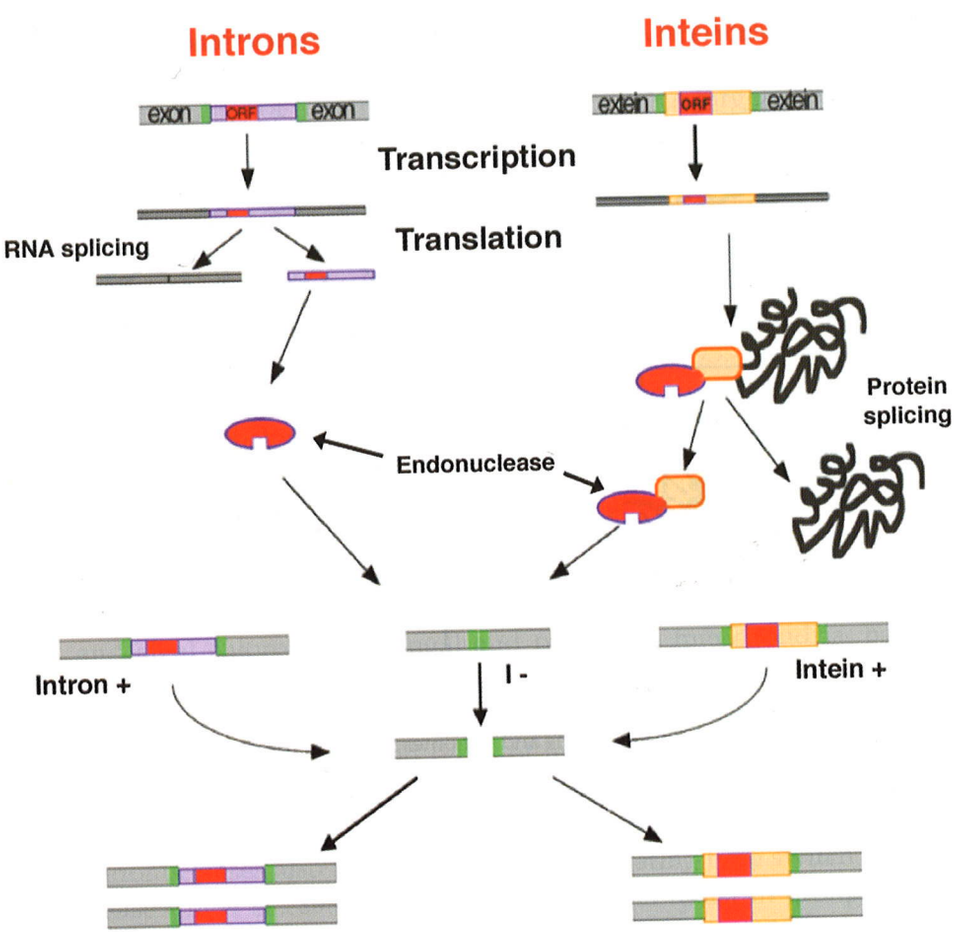


**Fig. 2.** Homing endonucleases and intron/intein mobility. Homing is the transfer of an intervening sequence (either an intron or intein) to a homologous allele that lacks the sequence, leading to gene conversion and dominant transmission and inheritance of the mobile element. Invasion of ectopic sites, including transfer across biological kingdoms and between different genomic compartments, has also been documented. Homing is initiated by an endonuclease, that is encoded within the mobile intervening sequence, that recognizes a DNA target site and generates a single or double strand break. Transfer of group I introns and inteins is initiated solely by the endonuclease activity, and is completed by cellular mechanisms that repair the strand breaks via homologous recombination, using the intron- or intein-containing allele as a template. (Based on Fig. 1 of: Stoddard BL, *Q Rev Biophys* 38: 49–95 (2005), with permission of Cambridge University Press.)

remaining three families each contain a unique catalytic scaffold and appear to have arisen independently of one another. The structure and mechanisms of representatives from each of these families have been extensively characterized over the past ten years.

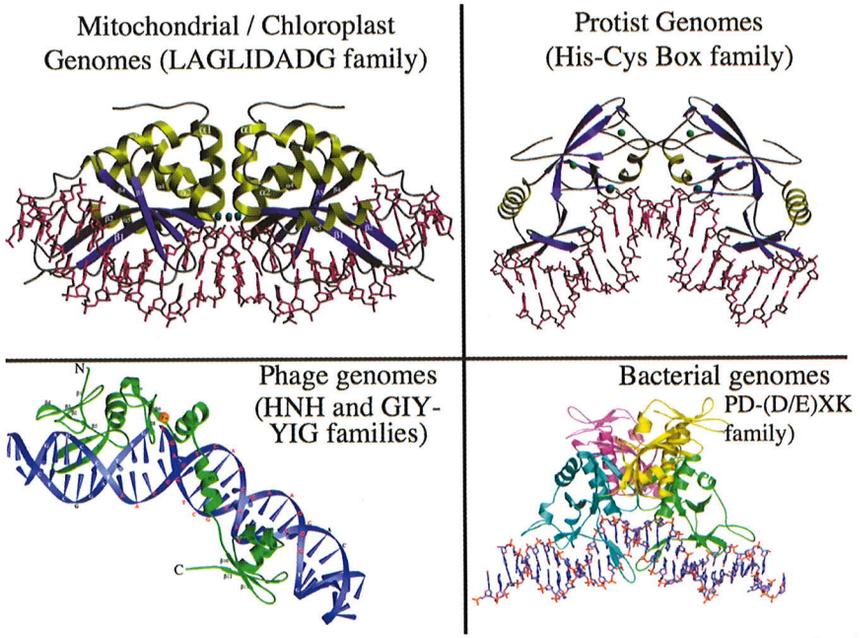


**Fig. 3.** Representative members of the known group I homing endonuclease families. Each particular family of homing endonuclease is generally constrained to mobility within their corresponding family of host organisms and genomes, presumably due to barriers imposed by the constraints of host toxicity and/or poor mobility when transferred into a foreign genome. Of these homing endonuclease families, the LAGLIDADG endonucleases have become increasingly useful for engineering.

## 3. LAGLIDADG Homing Endonuclease Proteins are Well-Suited for Protein Engineering and Genomic Targeting Applications

The LAGLIDADG homing endonucleases ("LHEs") are the largest family of such enzymes, which includes the first identified and biochemically characterized intron-encoded proteins (Dujon, 1980; Jacquier and Dujon, 1985; Kostriken *et al.*, 1983; Lazowska *et al.*, 1980). It has also been variously termed the "DOD," "dodecapeptide," "dodecamer," and "decapeptide" endonuclease family, based on the conservation of a 10-residue sequence motif (Belfort *et al.*, 1995; Belfort and Roberts, 1997; Chevalier and Stoddard, 2001; Dalgaard *et al.*, 1997; Dujon, 1989; Dujon *et al.*, 1989). LHEs are segregated into groups that possess either one or two copies of the conserved LAGLIDADG motif. Enzymes that contain a single copy of this motif act as homodimers on DNA targets sites that are palindromic or near-palindromic. In contrast,



LHEs that have two copies of the motif are monomers that possess a pair of structurally similar nuclease domains on a single peptide chain. These endonucleases are not constrained to symmetric DNA target sites. The two domains are linked in these monomeric proteins by flexible linker peptides that range from three residues to over 100 residues (Dalgaard *et al.*, 1997). LHEs that are not covalently linked to additional functional domains recognize DNA sites that range from 18 to 22 base pairs (Agaard *et al.*, 1997; Dalgaard *et al.*, 1994; Durrenberger and Rochaix, 1993; Perrin *et al.*, 1993). They cleave both DNA strands to generate cohesive four-base 3′-overhangs (Colleaux *et al.*, 1988; Durrenberger and Rochaix, 1993; Thompson *et al.*, 1992). Like most nucleases, LHEs require divalent cations for activity.

As reviewed below and in Stoddard (2005), the LAGLIDADG family appear to be homing endonucleases of choice for engineering and gene-specific applications:

(i) They represent the largest collection of known and characterized HEs, with a diverse biological host range that spans the genomes of plant and algal chloroplasts, fungal and protozoan mitochondria, bacteria and *Archaea* (Dalgaard *et al.*, 1997). One reason for the wide phylogenetic distribution of LAGLIDADG genes appears to be their remarkable ability to invade unrelated types of intervening sequences, including group I introns, archaeal introns and inteins (Belfort and Roberts, 1997; Chevalier and Stoddard, 2001). Descendents of LAGLIDADG homing endonucleases also include the yeast HO mating type switch endonuclease (Jin *et al.*, 1997), and "maturases" that assist in group I intron splicing (Delahodde *et al.*, 1989; Geese and Waring, 2001; Lazowska *et al.*, 1989; Schafer *et al.*, 1994).

(ii) They are the most specific of all known homing endonucleases, typically recognizing 19 to 22 base-pair DNA target sites. Thus, LHEs target DNA sites that are sufficiently long to ensure that they cleave one or a small number of target sites, even within complex genomes. They display low- to sub-nanomolar dissocation constants against their cognate target sites. The specificity of cognate DNA recognition and cleavage, and the structural basis for this behavior, has been extensively described both for a homodimeric LHE (I-CreI: Chevalier *et al.*, 2003) and for monomeric LHEs that are prototypes for genome engineering and correction applications (I-SceI: Gimble *et al.*, 2003; and I-AniI: Scalley-Kim *et al.*, 2007); under physiological conditions, the measured specificity of DNA recognition for that latter enzyme is at least 1 in $10^9$ random sequences.

(iii) They possess a well-defined, relatively small and highly modularized structure, and employ a DNA-binding mechanism, described in detail below, that is amenable to design and engineering. Monomeric LHEs (those containing both active sites and DNA recognition regions on a single peptide chain) are surprisingly small (ranging from ~200 to 250 amino acids total) relative to their long DNA target sites, thus simplifying the task of delivering a functional nuclease or its gene to target cells (Dalgaard *et al.*, 1997).

(iv) LHE's have been shown to be highly effective in inducing markerless modification of genes without toxicity in both prokaryotic and eukaryotic organisms. Of note, the prototypical LHE I-SceI has been used as the "gold standard" against which the activity and specificity of engineered ZFN's have been measured (Pâques and Duchateau, 2007; Porteus and Baltimore, 2003).

(v) LHE cleavage activity is tightly linked to site-specific binding of the cognate DNA target, significantly minimizing off-site cleavage activity (Chevalier *et al.*, 2004). Extensive experimentation by a wide variety of laboratories with I-SceI and other LHEs, in which transfected eukaryotic cells overexpress active endonuclease scaffolds, indicate little or no toxicity or off-site activity in a variety of experimental contexts, together with the ability to promote efficient, DNA double strand break-dependent gene conversion (Pâques and Duchateau, 2007).

(vi) LHE proteins have the potential for biochemical diversification, as they can be converted to site-specific nicking enzymes (McConnell-Smith and Stoddard, unpublished data) and can be fused to additional function protein domains at their N- and C-termini while retaining target site-specific binding affinity and activity.

### 3.1. *The Structure, Specificity and Catalytic Activity of LAGLIDADG Endonucleases*

The structures of several LHEs bound to their DNA targets have been determined [Fig. 4(A) and 4(B)]. These include several homodimers [I-CreI (Chevalier *et al.*, 2003 and 2001; Heath *et al.*, 1997; Jurica *et al.*, 1998), I-MsoI (Chevalier *et al.*, 2003) and I-CeuI (Spiegel *et al.*, 2006)], pseudosymmetric monomers [I-AniI (Bolduc *et al.*, 2003) and I-SceI (Moure *et al.*, 2003)], an artificially engineered chimeric enzyme [H-DreI (Chevalier *et al.*, 2002)] and an intein-associated endonuclease from yeast [PI-SceI (Moure *et al.*, 2002)]. Structures have also been determined of additional LHE's in the absence



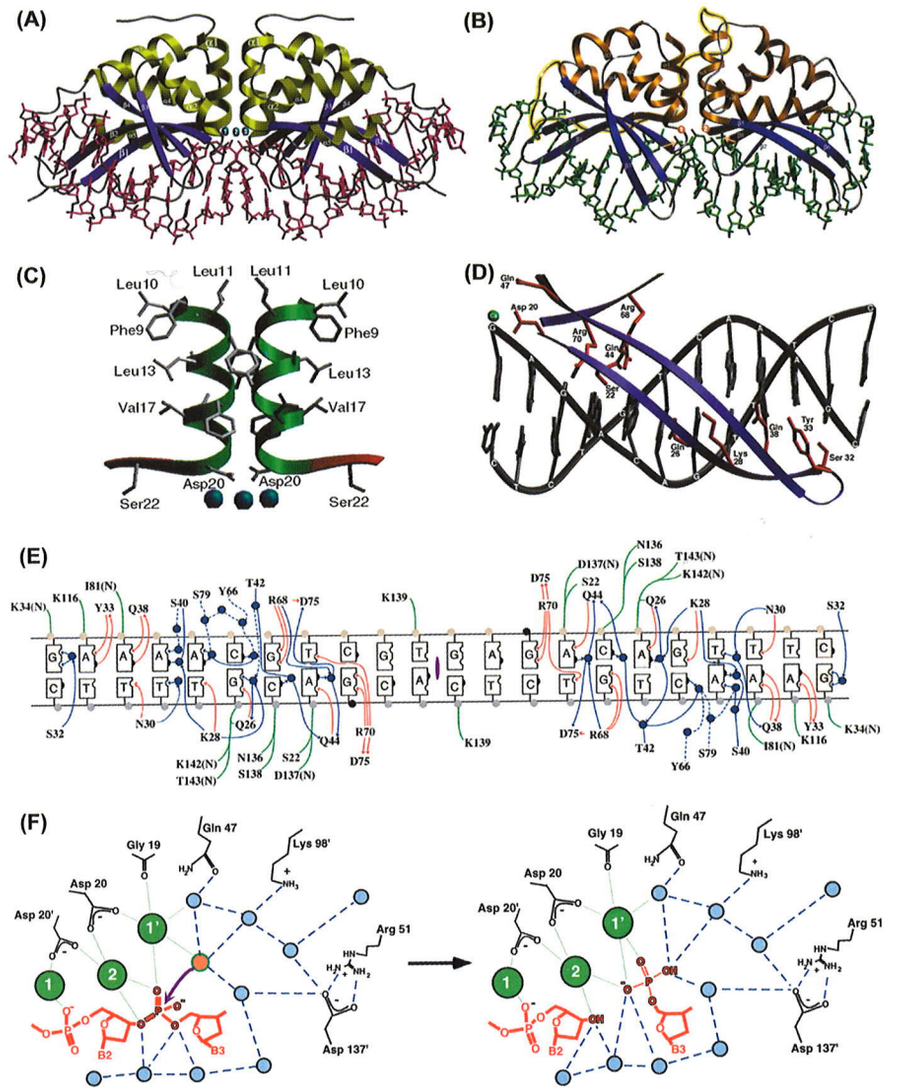


**Fig. 4.** Structural features of LAGLIDADG homing endonucleases. **(A)** DNA bound complex of the I-CreI homodimer. **(B)** DNA bound complex of the I-AniI monomer. The peptide linker connecting the two related domains is outlined in yellow. **(C)** The LAGLIDADG motifs form the helices at the domain interface of the I-CreI structure and serve a similar role in all known LAGLIDADG enzymes. **(D)** Structure and orientation of the DNA-binding antiparallel $\beta$-sheets from I-CreI. **(E)** Summary of direct and water-mediated contacts between the I-CreI enzyme and the bases of its DNA target site. Direct bonds are shown as solid arrows (red to bases, green to backbone atoms); water-mediated contacts are shown as blue dashed and solid lines. The scissile phosphates are indicated with black closed circles. **(F)** Proposed catalytic mechanism for I-CreI, as described in the text. Nucleophilic water is red; surrounding ordered water molecules are blue. (Based on Fig. 3 of: Stoddard BL, *Q Rev Biophys* 38: 49–95 (2005), with permission of Cambridge University Press.)

of DNA. LHE domains form an elongated protein fold that consists of a core fold with mixed $\alpha/\beta$ topology ($\alpha$-$\beta$-$\beta$-$\alpha$-$\beta$-$\beta$-$\alpha$: Heath *et al.*, 1997). The overall shape of this domain is a half-cylindrical "saddle" that averages approximately 25 Å × 25 Å × 35 Å. The surface of the saddle is formed by an anti-parallel, three- or four-stranded $\beta$-sheet that presents a large number of exposed basic and polar residues for DNA contacts and binding. The complete DNA-binding surfaces of the full-length enzymes, generated by two-fold symmetry or pseudo-symmetry, are 70 to 85 Å long and thus can accommodate DNA target sites of up to 24 base pairs.

The LAGLIDADG motifs [Fig. 4(C)] form the last two turns of the N-terminal helices in each folded domain or monomer and are packed against one another. They also contribute N-terminal, conserved acidic residues to two active sites where they help coordinate divalent cations that are essential for catalytic activity. The structure and packing of the parallel, two-helix bundle in the domain interface of the LAGLIDADG enzymes is strongly conserved among the otherwise highly diverged members of this enzyme family.

Two structurally independent, antiparallel $\beta$-sheets (one from each protein domain) are used to contact nucleotide bases within the major groove, at positions flanking the central four base pairs [Fig. 4(D)]. Despite little primary sequence homology among the LHEs outside of the motif itself, the topologies of the endonuclease domains of the enzymes visualized to date, and the shape of their DNA-bound $\beta$-sheets, are remarkably similar. A structural alignment of several endonuclease domains and subunits in their DNA-bound conformation indicates that the structure of the central core of the $\beta$-sheets is well conserved (Bolduc *et al.*, 2003). The conformations of the more distant ends of the $\beta$-strands and connecting turns are more degenerate. Base pairs $\pm 3$ to $\pm 7$ in each DNA half-site are typically recognized with higher specificity than base pairs in the less conserved, distant flanks of the DNA target. In rare cases, the core fold of LHEs can be tethered to additional functional domains involved in DNA binding (Sitbon and Pietrokovski, 2003). For example, a single copy of a canonical helix-turn-helix domain is found downstream (C-terminal) from the LAGLIDADG core of the intron-associated gene product of ORF Q0255 in yeast. This motif is similar to a conserved region of the bacterial sigma54-activator DNA-binding protein (Wintjens and Rooman, 1996).

LHEs typically make contact with 65 to 75% of possible hydrogen-bond donors and acceptors of the base pairs in the major groove. They make few or no additional contacts in the minor groove, and contact approximately one-third of the backbone phosphate groups across the homing site



sequence [Fig. 4(E)]. These contacts are split evenly between direct and water-mediated contacts. Target site DNA is progressively and gradually bent around the interface formed by antiparallel $\beta$-sheets to give an overall curvature across the entire length of the site of ~45°. The information content (specificity) of recognition and cleavage by these enzymes, at each base pair of their DNA target site, is correlated with the number and type of intermolecular contacts made by the enzyme to each base pair. In addition, specificity is increased at the individual base pairs, particularly near the center of the cleavage site, by the additional indirect contribution of sequence-specific conformational preferences of the DNA itself.

The structures of several LHEs have been determined at relatively high resolution in complex with DNA (2.4 to 1.5 Å). These high resolution structures demonstrate the presence of three bound divalent metal ions distributed across a pair of overlapping active sites, with one central metal shared between the active sites. These enzymes appear to employ a canonical two-metal mechanism for phosphodiester hydrolysis [Fig. 4(F); Chevalier *et al.*, 2001]. The active site metal ions are coordinated by conserved acidic residues from each LAGLIDADG motif, and by oxygen atoms from scissile phosphates on each DNA strand. Individual details of the structural mechanism of nucleophilic activation appear to differ between enzyme subfamilies. With the exception of the direct metal-binding residues from the LAGLIDADG motifs, the active site residues are only moderately conserved within the LAGLIDADG enzymes (Chevalier *et al.*, 2001).

### 3.2. *Engineering and Selection of LAGLIDADG Homing Endonucleases with Altered Structures and Specificities*

The past 20 years of homing endonuclease research, including recent analyses of their evolution and divergence that have been facilitated by high-throughput sequencing programs, have conclusively documented the dynamic evolution of these gene and protein families. Protein open reading frames have undergone rapid divergence, structural shuffling and recombination, continuous adaptation to and invasion of ectopic target insertion sites, rapid expansion throughout novel target lineages, and cyclical acquisition and loss. As a result, it is widely believed that the actual — and potential — site recognition repertoires of homing endonucleases are extremely broad. These data argue that if the primary mechanisms by which evolution has driven specificity changes in homing endonucleases could be duplicated in the laboratory, it should be possible to generate a wide variety

of HEs with diverse DNA target site specificities for genome engineering applications.

The structures of several hundred protein-DNA complexes have been determined and analyzed at high resolution, and there have been many corresponding attempts to survey and catalogue the identity of contacts made in those complexes (Pabo and Nekludova, 2000). These studies have repeatedly indicated that while certain DNA nucleotides display preferences for specific residue contact patterns (such as the use of arginine side chains to make direct contacts to guanine bases, or the use of the amides of glutamine and asparagine to make direct contacts to adenine bases), there is no simple one-to-one "code" describing and predicting such interactions, even for the most modular and simplest DNA recognition modules such as zinc-fingers (Wolfe *et al.*, 2000a and 2000b). This is due to the fact that side-chain contacts to DNA bases, and their individual contribution to specificity, are exquisitely sensitive to the surrounding structural and chemical context of the protein-DNA interface. This context is determined by the local backbone conformation and the structure of both the protein and the corresponding DNA target site, as well as the conformational changes that occur during complex formation and the interdependence of neighboring contact networks.

In addition to the context-dependent role of direct protein contacts, binding specificity is also driven by the sequence-dependent conformational preferences of potential target sites, which can greatly influence and increase specificity at the individual nucleotide positions (Lavery, 2005). This property is usually termed "indirect" readout of specificity, as contrasted to the specificity created through direct contacts as described above. For example, the homing endonuclease I-CreI clearly displays very strong sequence preferences in the DNA target at individual base pairs that are not directly contacted by the enzyme, due to significant DNA bending induced across the center of the target site (Chevalier *et al.*, 2003).

The implication of these studies for the engineering of novel protein-DNA specificities is clear: that structure-based redesign of direct protein-DNA contact points, together with the generation or selection of equally critical variants within the protein scaffold, will be required to produce proteins with the requisite target site specificity for genome engineering applications.



### 3.3. *Redesign of Homing Endonuclease Domain Architecture and Oligomery*

Several independent studies have demonstrated that domains or subunits from unrelated LHEs can be fused to create fully active, chimeric homing endonucleases that recognize corresponding chimeric DNA target sites (Chevalier *et al.*, 2002; Epinat *et al.*, 2003; Steuer *et al.*, 2004). This technology requires extensive repacking of the domain interface, and allows the creation of new protein scaffolds with novel specificities. These studies reinforce the idea that LAGLIDADG HEs display modularity of form and function, with individual domains and subunits being largely or entirely reponsible for the recognition and binding of individual DNA target half-sites.

Of particular note, two separate groups have generated an artificial highly specific endonuclease by fusing domains of homing endonucleases I-DmoI and I-CreI (Fig. 5); the novel proteins were termed "E-DreI" (now renamed H-DreI for **H**ybrid-**D**mo/C**reI**) and "DmoCre," respectively (Chevalier *et al.*, 2002; Epinat *et al.*, 2003). In the former study leading to the creation of H-DreI, structure-based protein engineering was accomplished by combining computational redesign and an *in vivo* protein folding and solubility screen (Chevalier *et al.*, 2002). The resulting enzyme binds a long chimeric DNA target site with nanomolar affinity, cleaving it precisely at the same phosphate groups with a rate equivalent to its natural parents. The structure of the engineered protein in complex with its DNA target demonstrated the accuracy of the protein interface redesign algorithm, and revealed how catalytic function was maintained by the creation of new, chimeric active sites. Most importantly, the mechanism of DNA recognition displayed by the chimeric endonuclease, and the identity of its residues used to make contacts to individual nucleotides, appeared to be indistinguishable from the original parental enzymes. Thus, the individual domains of LAGLIDADG endonucleases appear to be highly modular, and thus can be shuffled and recombined in order to make large numbers of different DNA-binding specificities.

Additional engineering experiments have reinforced this concept. Domains isolated from homing endonucleases and from inteins (self-splicing protein ligases, that are often associated with HEs) can be fused and shuffled into artificial, bifunctional inteins with novel DNA-binding specificities and/or activities. One such construct was assembled by inserting a gene that expresses one of the two I-CreI subunits into the *Mycobacterium xenopi* GyrA mini-intein (Fitzsimons-Hall *et al.*, 2002). This engineered intein displayed appropriate protein splicing, and produced a homodimeric

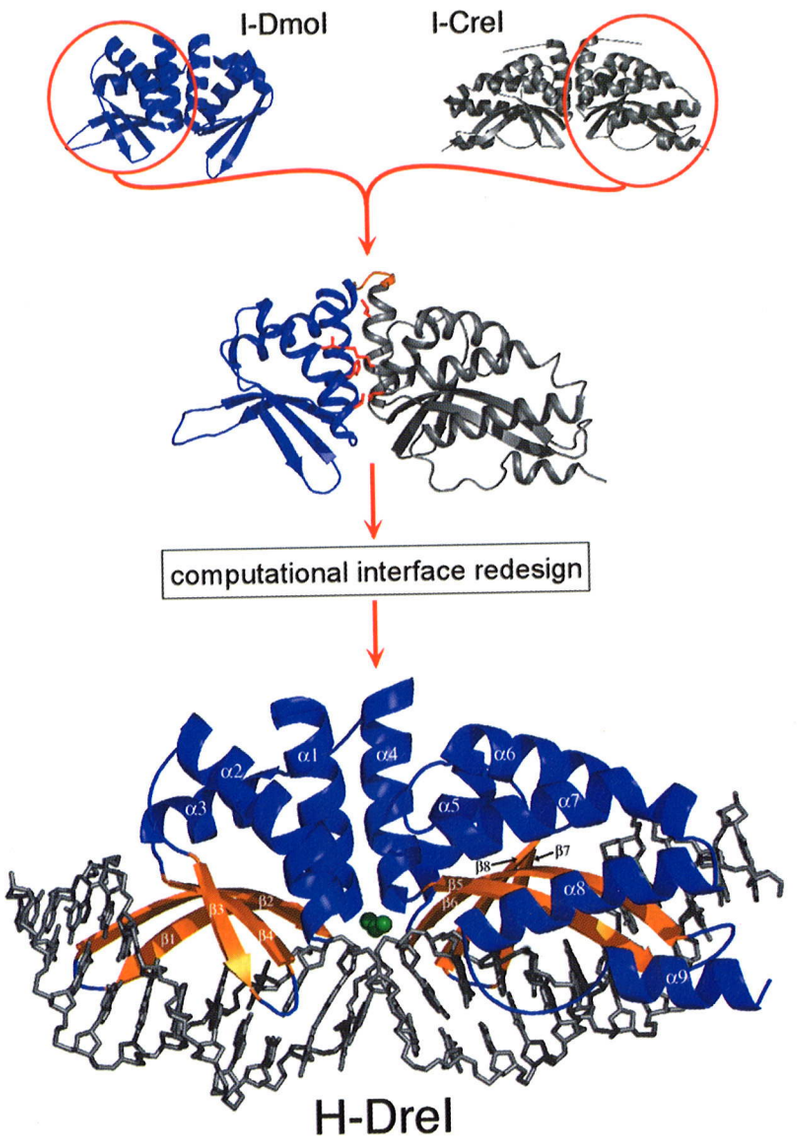


**Fig. 5.** Generation of an artificial, chimeric homing endonuclease. Individual domains of the I-DmoI and I-CreI parental enzymes (a monomeric LHE from an archaea host and a homodimeric LHE from a green algae, respectively) were fused into a single peptide chain, and computational protein engineering was employed to redesign and stabilize the novel domain interface. The resulting protein retains the overall thermal stability, specificity and cleavage rate of the parental enzymes, while acting at a chimeric DNA target site. (Based on Fig. 12 of: Stoddard BL, *Q Rev Biophys* **38**: 49–95 (2005), with permission of Cambridge University Press.)



site-specific endonuclease activity identical to naturally occurring I-CreI. Separate experiments have also demonstrated that the PI-SceI protein splicing domain can be used as a site-specific DNA binding module in chimeric protein constructs: domain swapping between the PI-SceI and a homologue from *Candida tropicalis* (PI-CtrIP) was conducted to design altered specificity proteins (Steuer *et al.*, 2004).

A related experiment has demonstrated that a single chain, monomeric endonuclease can be generated from a homodimer predecessor, I-CreI (Epinat *et al.*, 2003). This construct was shown to initiate homologous recombination in both yeast and mammalian cells. Finally, the role and mutability of LAGLIDADG interface residues has been examined by grafting side chains from the homodimeric I-CreI into the corresponding positions in the monomeric I-DmoI enzyme, resulting in enzymes with novel nicking activities and oligomeric properties (Silva and Belfort, 2004). Subsequent experiments with this same enzyme have demonstrated that individual domains from the monomeric wild-type protein can be engineered to form stable and functional homodimers, again illustrating the modularity of LHE scaffolds (Silva *et al.*, 2006).

### 3.4. *Redesign of Homing Endonuclease DNA Contacts and Specificity*

Several methods have been used to alter the DNA target site specificity of LHE's (Figs. 6 and 7). These have focused in large part on the mutation of individual DNA-protein interface side chains that contact specific DNA target site base pairs. These strategies can be broadly divided into:

(i) Protocols that rely on a selection for high affinity DNA binding activity.
(ii) Protocols that rely on a selection for efficient cleavage activity. Examples have been described (summarized below) of methods that rely on the elimination of a gene (usually by destruction of a plasmid and a resident protein coding sequence) or on complementation of a gene (by cleavage-induced homologous recombination of two dysfunctional enzyme alleles into a wild-type, functional reading frame). This latter method is the basis of the most successful "high-throughput" endonuclease redesigns to date, which have progressed to targeting of physiologically relevant gene sequences. Those experiments are summarized at the end of this section.
(iii) Protocols that rely on structure-based computational redesign of DNA-protein interface contact surfaces and residues. It is important to note

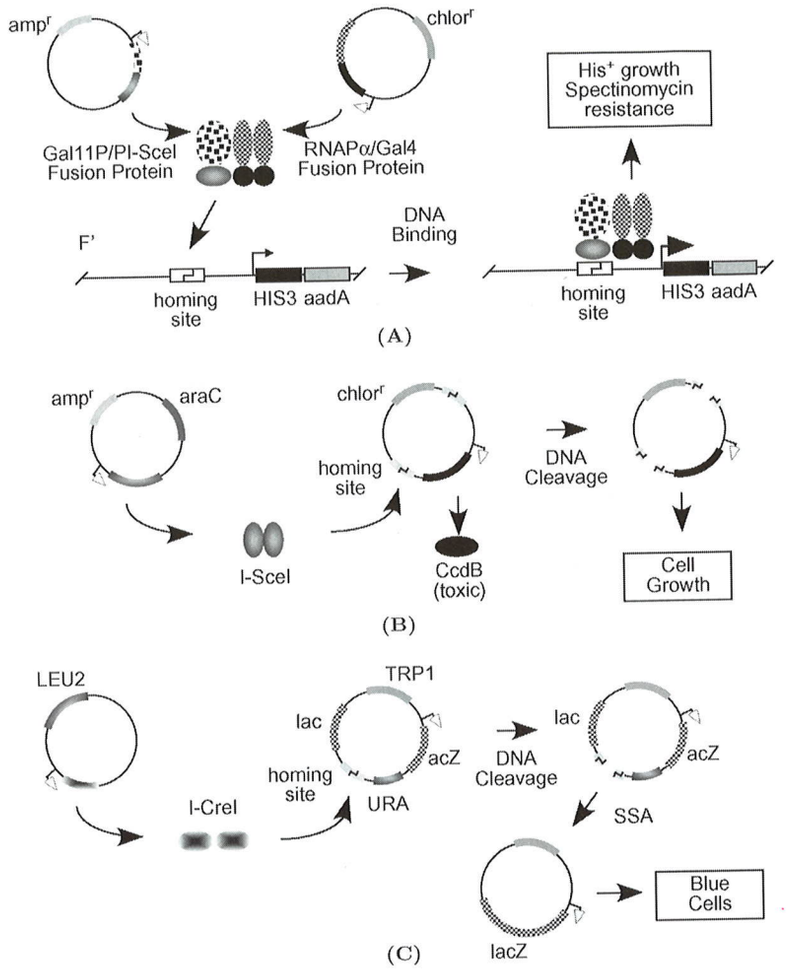


**Fig. 6.** Selections and screens for altered specificity homing endonucleases. **(A)** Selection for DNA binding activity using a two-hybrid system (Gimble *et al.*, 2003). Gal11P/PI-SceI variants expressed from a plasmid library that bind to a homing site increase the expression of *HIS3* and *aadA* by recruiting the RNAPα/Gal4 fusion protein to the weak $P_{lac}$ promoter on the F′. Thus, cells that express a PI-SceI protein capable of binding a chosen DNA target sequence grow on histidine-selective and spectinomycin-selective media. **(B)** Selection for DNA cleavage activity (Chen and Zhao, 2005; Doyon and Liu, 2006). Co-existence of two plasmids kills bacterial cells because one expresses the toxic CcdB gene product. Selection can be made for I-SceI enzyme variants that cleave I-SceI target sites on the CcdB plasmid because they eliminate it and allow the cells to survive. **(C)** Screen for homologous recombination activity (Arnould *et al.*, 2006). Expression of I-CreI from a plasmid library cleaves a target site within an interrupted *lacZ* gene located on a reporter plasmid. This DSB stimulates single strand annealing (SSA) of two direct repeats that flank the site, leading to restoration of a functional *lacZ* gene and the appearance of a blue colony. (Figure and caption reproduced from: Gimble FS, *Gene Ther Regul* **3**: 33–50 (2007), with permission of the author and of World Scientific Publishing.)



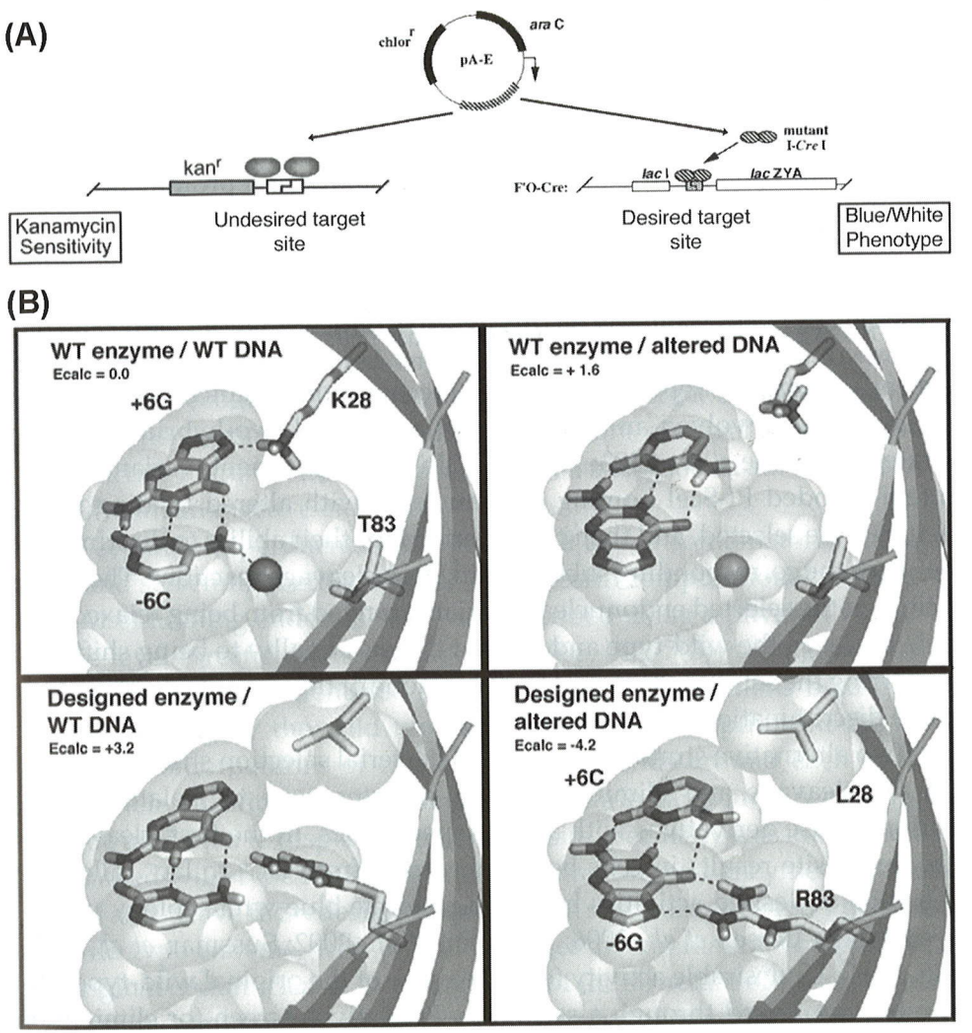


**Fig. 7.** From *in vivo* selection to *in silico* screening. **(A)** Screen for homing endonuclease based on cleavage and gene elimination (Seligman *et al.*, 2002 and 2004). Plasmid-encoded I-CreI derivatives that bind and cleave an I-CreI homing site located on an F′ lead to its elimination. Concomitant loss of an adjacent kanamycin antibiotic marker yields kanamycin-sensitive cells. **(B)** Alteration of specificity of the I-MsoI homing endonuclease by computational structure-based redesign (Ashworth *et al.*, 2006). The four panels provide a comparison of the experimentally determined interactions in wild-type and redesigned cognate complexes (*upper left* and *lower right*) and predicted interactions in non-cognate complexes, as well as predicted relative energies of the complexes.

that high resolution crystal structures of the wild-type protein-DNA complexes have been used in various contexts, either to enable the targeted redesign of homing endonuclease-DNA contacts at individual residues (thus bypassing selection approaches altogether), or to facilitate more efficient mutational screening of enzyme libraries (by identifying, and thus greatly reducing, the number of protein residues to be randomized). The use of such crystal structures is a prominent feature of the most successful LHE redesigns described to date.

### 3.4.1. *Redesign of specificity at individual DNA base pairs*

An early attempt to reprogram the DNA recognition specificity of a homing endonuclease involved an adaptation of a bacterial two-hybrid screening strategy [Gimble *et al.*, 2003; Fig. 6(A)]. In this experiment, variants of the intein-encoded PI-SceI homing endonuclease with altered binding specificities were selected, and then characterized for their ability to discriminate between the corresponding substrates in DNA cleavage reactions. The specificities of the selected endonuclease variants ranged from being relaxed (i.e., able to cleave the wild-type and mutant targets equally) to being shifted to preferring the selection targets. However, none of the variants displayed the same discrimination as wild-type PI-SceI.

Two alternative strategies have used bacterial selection strategies based on the cleavage and elimination of a reporter gene to isolate homing endonuclease derivatives with altered specificities. In the first, cleavage of the target site results in cells being converted from $lac^+$ to $lac^-$, allowing selection of desired activities based on a simple blue-white colony phenotype screen [Rosen *et al.*, 2006; Seligman *et al.*, 2002; Sussman *et al.*, 2004; Fig. 7(A)]. Undesirable activity (e.g., cleavage of the original wild-type site) can be suppressed through a secondary "negative" screen for elimination of an essential reporter (such as an antibiotic resistance marker). Using this method, endonuclease mutants with single or double amino-acid substitutions, at positions predicted to make base-specific DNA contacts, were assayed against appropriate DNA target site mutants. Enzyme variants with shifted specificities, but with reduced ability to discriminate between cognate and miscognate sites, were typically obtained. Crystallographic analyses of several of these altered mutants in complex with their new cognate targets demonstrated that the "modularity" of protein-DNA recognition, previously described for entire protein domains, extends to the level of individual amino-acid side chains (Rosen *et al.*, 2006; Sussman *et al.*, 2004).



A similar bacterial selection strategy has been described in which a toxic gene product results in cell death, unless a homing endonuclease variant that cleaves a homing site within the toxin expression vector is also present [Fig. 6(B)]. Two versions of this screen have been described: the first utilizing the nonspecific ribonuclease barnase (Gruen *et al.*, 2002), and the other describing a more easily controlled system based on the "control of cell death B" (CcdB) protein (Doyon *et al.*, 2006). In the latter screening strategy, CcdB expression led to very low rates of background survival, without requiring additional gene expression control elements such as those required to use the more toxic barnase protein. As with the bacterial selection described in the preceding paragraph, the CcdB system can be tailored as a positive selection for cleavage of a desired target, or as a negative selection, to strongly disfavor the recognition of a non-cognate DNA target site sequence. This system, when optimized, can lead to nearly 100% survival of cells expressing an active homing endonuclease, against a background survival of $<1$ in $5 \times 10^4$ with an inactive enzyme.

In separate experiments that avoided all use of combinatorial mutagenesis and screening methodologies, a successful computational redesign of the I-*Mso*I homing endonuclease was described using a physically realistic atomic level forcefield [Ashworth *et al.*, 2006; Fig. 7(B)]. Using an *in silico* screen, investigators identified single base pair substitutions predicted to disrupt binding by the wild type enzyme, and then optimized the identities and conformations of clusters of amino acids around each of these unfavorable substitutions by using Monte Carlo sampling. A redesigned enzyme predicted to display altered target site specificity was identified that maintained wild-type binding affinity. The redesigned enzyme was found to bind and cleave the redesigned recognition site more effectively than does the wild type enzyme, with a level of target discrimination comparable to the original endonuclease. Determination of the structure of the redesigned nuclease-recognition site complex by X-ray crystallography confirmed the accuracy of the computationally-predicted interface.

### 3.4.2. *High-throughput screening and targeting of genomic loci*

Recently, a commercial research and development group in France (Cellectis, Inc.) has developed a powerful eukaryotic assay system that reports on the generation of double-strand break-induced homologous recombination, rather than only DNA cleavage [Arnould *et al.*, 2006; Chames *et al.*, 2005; Fig. 6(C)]. In this assay, the function of a gene required for growth or for another easily scoreable phenotype is restored through the action

of a homing endonuclease. Prior to expression and action of the HE, the gene sequence is interrupted with an insert containing a desired HE cut site flanked by two direct repeats. In the most recently described version of this screen, the endonuclease expression construct and the "reporter" construct are located in separate yeast strains, allowing the investigator to introduce the HE (or a library of HE variants) to a target site by mating. The mating of the two required yeast strains can be automated, and thus done as a high-throughput assay. Thus, the same library of endonuclease variants can be efficiently screened against multiple DNA target site variants, in parallel or rapid sequential experiments.

Using this method, the DNA recognition specificity of I-CreI LHE has been thoroughly analyzed. Small endonuclease mutant libraries resulting from the randomization of two to four amino acids (corresponding to the "nearest protein neighbors" of individual DNA basepairs) were individually screened against all potential variant cognate sequences. This analysis identified individual mutations in the I-CreI protein scaffold that were associated with shifts in specificity at individual target site base pair positions. From these studies, hundreds of endonuclease mutants with altered specificies were identified, catalogued and archived (Arnould *et al.*, 2006). Many of these variants displayed cleavage activities and levels of site discrimination that were at least equivalent to the wild-type endonuclease, thus providing a starting point for more ambitious endonuclease redesign to physiological targets in eukaryotic, mammalian and even human genomes.

Using the approach summarized above, derivatives of the I-CreI LHE have been generated that display sequence-specific cleavage and recombination activity against the human *RAG1* gene (Smith *et al.*, 2006) at the site of mutations, giving rise to a rare subset of severe combined immunodeficiency disease (or SCID) phenotypes. I-CreI variants directed at the human *XPC* (xeroderma pigmentosum complementation group C) gene (Arnould *et al.*, 2007) were also generated. XPC, when a mutant, confers an extreme UV or sun-sensitivity phenotype together with a predisposition to sunlight-induced skin cancer. Both of these human disease genes are candidates for corrective genetic therapies. Furthermore, in the latter experiments, the modified I-CreI derivatives cleaving sequences from the XPC gene were found to induce a high level of gene targeting, similar to levels observed with the wild-type I-CreI or I-SceI scaffolds. This is the first time an engineered homing endonuclease has been used in mammalian cells to target and modify a chromosomal target locus.



### 3.4.3. *Surface display of homing endonucleases: avoiding bottlenecks in mutagenesis and screening?*

While the methodologies summarized above have shown great promise in delivering tailor-made gene-specific targeting reagents, they are limited by the requirement to first generate and then screen libraries of endonuclease mutants. Moreover, the selections and screens discussed above require the generation of unique intracellular reporter constructs that must be redesigned and constructed for each target sequence of interest. In order to address these issues, investigators have recently described a novel system where LHE proteins might be rapidly mutated *in vivo* and then screened to identify and isolate endonuclease variants with a new DNA target specificity (Fig. 8; Volna *et al.*, 2007). This work has built on a recent study that describes the evolution and selection of fluorescent protein spectral properties using the somatic hypermutation machinery of lymphocytes (Wang *et al.*, 2004).

Investigators demonstrated that LHEs can be expressed on the plasma membrane of a lymphocyte cell line by targeting the expression of an LHE-CD80 transmembrane fusion protein to the secretory pathway. Surface-expressed LHEs faithfully recapitulate the properties of the native enzymes in solution, as assessed by flow cytometric analysis of both the binding and the cleavage of fluorescently conjugated double-stranded DNA target site oligonucleotides. Identification of endonuclease expression clones with the desired DNA recognition properties was highly specific, allowing discrimination of endonucleases with binding preferences differing by only a single base pair. Furthermore, target sequence-specific LHE interactions with these DNA target site probes under conditions that limit substrate cleavage allow both the identification and sorting or enrichment of clones expressing LHE variants with the highest site specificity and affinity for further characterization. This coupled rapid analysis of LHE-DNA interactions on the cell surface, together with concurrent sorting or enrichment, should substantially accelerate the generation and isolation of novel endonuclease variants with unique DNA target specificities.

## 4. Concluding Remarks

The first X-ray structures of homing endonuclease (I-CreI and PI-SceI) were reported within one month of each other in 1997 (Duan *et al.*, 1997; Heath *et al.*, 1997). These were followed a year later by the first DNA-bound cocrystal structure (again I-CreI) (Jurica *et al.*, 1998). The subsequent 10 years have

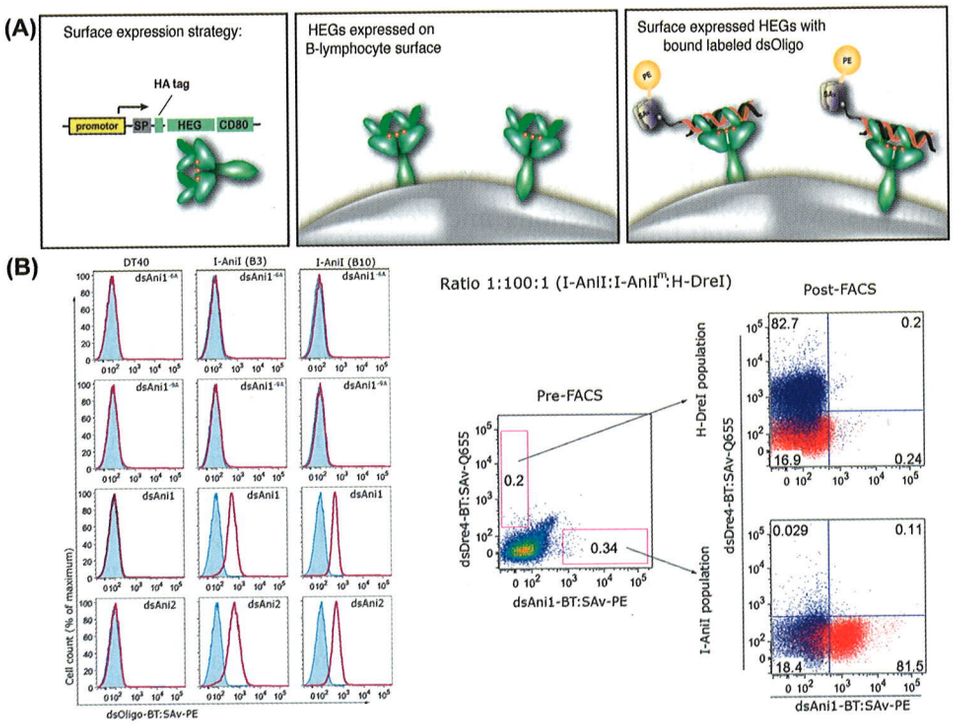


**Fig. 8.** DNA binding by surface displayed homing endonucleases. **(A)** An HA-tagged LAGLIDADG homing endonuclease is expressed on the surface of vertebrate cells through fusion with the transmembrane domain and tail of murine CD80. The expressed homing endonucleases I-AniI and H-DreI retain their DNA recognition, binding and cleavage activity, and binding can be quantitatively analyzed through the use of fluorescently labeled, double-stranded (ds) DNA oligonucletides and flow cytometry. "SP" = signal peptide. **(B)** *Left*: The fluorescence staining intensity using dsDNA probes is dependent on the surface expression and display of a homing endonuclease, and is capable of discriminating between the binding of a wild-type target site vs. miscognate sites harboring single base pair substitutions. DT40: parental cell line without surface HE expression. I-Ani (B3) and (B10): two individual clones expressing wild-type I-AniI homing endonuclease on the surface. dsAni1 and dsAni2: distinct dsDNA probes encoding different I-AniI target sites that are efficiently recognized, bound, and cleaved by recombinant soluble I-AniI. dsAniI-6A and dsAni1-9A: two dsDNA probes harboring I-AniI target sites that contain single base pair substitutions which eliminate cleavage activity of the recombinant soluble I-AniI. *Right*: enrichment through flow sorting of rare cells expressing a desired binding specificity from an excess (approximately 100–500 fold) of undesired HE binding activities.

produced a remarkable increase in our understanding of the biology, structure, function and mechanism of homing endonucleases. Representative structures from each of the major homing endonuclease families have been determined together with their catalytic mechanisms. At least six different methods for the modification and redesign of homing endonuclease DNA



target specificity have also been developed. During this period, a small number of homing endonuclease proteins have been assayed for activity and toxicity *in vivo*, and have been used to investigate mechanistic aspects of DNA double strand break repair and recombination. The first engineered homing endonuclease variants have also been described and characterized, and the most recent of these have been shown to induce DNA cleavage and recombination at chromosomal target loci *in vivo*. This collective experience has established the feasibility of using homing endonuclease proteins for genome engineering, together with the data needed to further develop homing endonuclease proteins for basic science and engineering applications.

The development of engineered homing endonucleases as reagents for basic biology (to facilitate the introduction or analysis of double- or single-strand breaks, covalent DNA modifications or trans-acting protein factors at specific DNA target sites in living cells) has matured rapidly over the past decade, and their further use and development in this area is virtually guaranteed. In contrast, the use of homing endonucleases as therapeutic reagents for disease treatment or prevention is still in its infancy and will require many years of dedicated work. Among the most important, unanswered questions are: the specificity profile of such reagents in biologically relevant contexts (e.g., stem cells), how efficient they are in generating target site-specific DNA double strand breaks *in vivo*, the range and efficiency with which individual DNA double-strand breaks lead to desired — and undesired — molecular outcomes in populations of cells, and the long-term toxicity, mutagenicity and immunogenicity of homing endonucleases when used as genome engineering, therapeutic or disease prevention reagents. These questions are pertinent now that the feasibility of using homing endonuclease proteins for genome engineering has been demonstrated, and when, as noted by Pâques and Duchateau, "the time when therapeutic applications were pure fantasy is now gone" (Pâques and Duchateau, 2007).

## References

Abbott A. (2006) Questions linger about unexplained gene-therapy trial death. *Nat Med* **12**: 597.

Agaard C, Awayez MJ, Garrett RA. (1997) Profile of the DNA recognition site of the archaeal homing endonuclease I-DmoI. *Nucl Acids Res* **25**:1523–30.

Arnould S, Chames P, Perez C, *et al.* (2006) Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J Mol Biol* **355**: 443–58.

Arnould S, Perez C, Cabaniols J-P, *et al.* (2007) Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *J Mol Biol* **371**: 49–65.

Ashworth J, Havranek JJ, Duarte CM, *et al.* (2006) Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature* **441**: 656–59.

Baird PA, Anderson TW, Newcombe HB, Lowry RB. (1988) Genetic disorders in children and young adults: a population study. *Am J Hum Genet* **42**: 677–93.

Baum C. (2007) Fourth case of leukaemia in the first SCID-X1 gene therapy trial, and the diversity of gene therapy. *http://www.esgct.org/upload/4th_CaseofLeukemial.pdf*

Beerli RR, Barbas CF. (2002) Engineering polydactyl zinc-finger transcription factors. *Nat Biotechnol* **20**: 130–41.

Belfort M, Reaban ME, Coetzee T, Dalgaard JZ. (1995) Prokaryotic introns and inteins: a panoply of form and function. *J Bacter* **177**: 3897–903.

Belfort M, Roberts RJ. (1997) Homing endonucleases — keeping the house in order. *Nucl Acids Res* **25**: 3379–88.

Bertolotti R. (1996) Recombinase-mediated gene therapy: strategies based on Lesch-Nyhan mutants for gene repair/inactivation using human RAD51 nucleoprotein filaments. *Biogenic Amines* **12**: 487–98.

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

Bibikova M, Carroll D, Segal DJ, *et al.* (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol* **21**: 289–97.

Bolduc JM, Spiegel PC, Chatterjee P, *et al.* (2003) Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing cofator. *Genes Dev* **17**: 2875–88.

Bullard DC, Weaver CT. (2002) Cutting-edge technology: IV. Genomic engineering for studies of the gastrointestinal tract in mice. *Am J Physiol Gastrointest Liver Physiol* **283**: G1232–37.

Bulyk ML, Huang XH, Choo Y, Church GM. (2001) Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. *Proc Natl Acad Sci USA* **98**: 7158–63.

Burt A. (2003) Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proc R Soc Lond B Biol Sci* **270**: 921–28.

Burt A. (2005) Gates Grand Challenge Project. In: *Homing Endonuclease Genes: New Tools for Mosquito Population Engineering and Control. www.gcgh.org.*

Chames P, Epinat JC, Guillier S, *et al.* (2005) *In vivo* selection of engineered homing endonucleases using double-strand break induced homologous recombination. *Nucl Acids Res* **33**: e178.

Chase CD. (2006) Genetically engineered cytoplasmic male sterility. *Trends Plant Sci* **11**: 7–9.

Chevalier B, Sussman D, Otis C, *et al.* (2004) Analysis of the symmetric metal-dependent I-CreI homing endonuclease mechanism. *Biochemistry* **43**: 4015–26.



Chevalier B, Turmel M, Lemieux C, *et al.* (2003) Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-CreI and I-MsoI. *J Mol Biol* **329**: 253–69.

Chevalier BS, Kortemme T, Chadsey MS, *et al.* (2002) Design, activity and structure of a highly specific artificial endonuclease. *Mol Cell* **10**: 895–905.

Chevalier BS, Monnat RJ Jr, Stoddard BL. (2001) The homing endonuclease I-CreI uses three metals, one of which is shared between the two active sites. *Nat Struct Biol* **8**: 312–16.

Chevalier BS, Stoddard BL. (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucl Acids Res* **29**: 3757–74.

Christianson A, Howson CP, Modell B. (2006) Global Report on Birth Defects: The Hidden Toll of Dying and Disabled Children. March of Dimes, White Plains, New York.

Coates CJ, Kaminski JM, Summers JB, *et al.* (2005) Site-directed genome modification: derivatives of DNA-modifying enzymes as targeting tools. *Trends Biotechnol* **23**: 407–19.

Colleaux L, D'Auriol L, Betermier M, *et al.* (1986) Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into *E. coli* as a specific double strand endonuclease. *Cell* **44**: 521–33.

Colleaux L, D'Auriol L, Galibert F, Dujon B. (1988) Recognition and cleavage site of the intron-encoded omega transposase. *Proc Natl Acad Sci USA* **85**: 6022–26.

Dalgaard JZ, Garrett RA, Belfort M. (1994) Purification and characterization of two forms of I-*Dmo*I, a thermophilic site-specific endonuclease encoded by an archaeal intron. *J Biol Chem* **269**: 28885–92.

Dalgaard JZ, Klar AJ, Moser MJ, *et al.* (1997) 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 HNH family. *Nucl Acids Res* **25**: 4626–38.

Delahodde A, Goguel V, Becam AM, *et al.* (1989) Site-specific DNA endonuclease and RNA maturase activities of two homologous intron-encoded proteins from yeast mitochondria. *Cell* **56**: 431–41.

Doetschman T, Gregg RG, Maeda N, *et al.* (1987) Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature* **330**: 576–78.

Doyon JB, Pattanayak V, Meyer CB, Liu DR. (2006) Directed evolution and substrate specificity profile of homing endonuclease I-SceI. *J Am Chem Soc* **128**: 2477–84.

Duan X, Gimble FS, Quiocho FA. (1997) Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. *Cell* **89**: 555–64.

Dujon B. (1980) Sequence of the intron and flanking exons of the mitochondrial 21S rRNA gene of yeast strains having different alleles at the omega and rib-1 loci. *Cell* **20**: 185–97.

Dujon B. (1989) Group I introns as mobile genetic elements: facts and mechanistic speculations — a review. *Gene* **82**: 91–114.

Dujon B, Belfort M, Butow RA, *et al.* (1989) Mobile introns: definition of terms and recommended nomenclature. *Gene* **82**: 115–18.

Durrenberger F, Rochaix J-D. (1993) Characterization of the cleavage site and the recognition sequence of the I-*Cre*I DNA endonuclease encoded by the chloroplast ribosomal intron of *Chlamydomonas reinhardtii*. *Mol Gen Genet* **236**: 409–14.

Epinat JC, Arnould S, Chames P, *et al.* (2003) A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucl Acids Res* **31**: 2952–62.

Fitzsimons-Hall M, Noren CJ, Perler FB, Schildkraut I. (2002) Creation of an artificial bifunctional intein by grafting a homing endonuclease into a mini-intein. *J Mol Biol* **323**: 173–79.

Geese WJ, Waring RB. (2001) A comprehensive characterization of a group IB intron and its encoded maturase reveals that protein-assisted splicing requires an almost intact intron RNA. *J Mol Biol* **308**: 609–22.

Ghatge M, Palaniappan N, Das Choudhuri S, Reynolds K. (2006) Genetic manipulation of the biosynthetic process leading to phoslactomycins, potent protein phosphatase 2A inhibitors. *J Ind Microbiol Biotechnol* **33**: 589–99.

Gimble FS, Moure CM, Posey KL. (2003) Assessing the plasticity of DNA target site recognition of the PI-SceI homing endonuclease using a bacterial two-hybrid selection system. *J Mol Biol* **334**: 993–1008.

Glaser S, Anastassiadis K, Stewart AF. (2005) Current issues in mouse genome engineering. *Nat Genet* **37**: 1187–93.

Gouble A, Smith J, Bruneau S, *et al.* (2006) Efficient in toto targeted recombination in mouse liver by meganuclease-induced double-strand break. *J Gene Med* **8**: 616–22.

Griesenbach U, Geddes DM, Alton EW. (2006) Gene therapy progress and prospects: cystic fibrosis. *Gene Ther* **13**: 1061–67.

Gruen M, Chang K, Serbanescu I, Liu DR. (2002) An *in vivo* selection system for homing endonuclease activity. *Nucl Acids Res* **30**: 29–34.

Hacein-Bey-Abina S, Von Kalle C, Schmidt M, *et al.* (2003) LMO2–associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**: 415–19.

Heath PJ, Stephens KM, Monnat RJ, Stoddard BL. (1997) The structure of I-*Cre*I, a group I intron-encoded homing endonuclease. *Nat Struct Biol* **4**: 468–76.

Hinnen A, Hicks JB, Fink GR. (1978) Transformation of yeast. *Proc Natl Acad Sci USA* **75**: 1929–33.

Jacquier A, Dujon B. (1985) An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell* **41**: 383–94.

Jin Y, Binkowski G, Simon LD, Norris D. (1997) Ho endonuclease cleaves MAT DNA *in vitro* by an inefficient stoichiometric reaction mechanism. *J Biol Chem* **272**: 7352–59.



Jurica MS, Monnat RJ Jr, Stoddard BL. (1998) DNA recognition and cleavage by the LAGLIDADG homing endonuclease I- CreI. *Mol Cell* **2**: 469–76.

Kolb AF, Coates CJ, Kaminski JM, *et al.* (2005) Site-directed genome modification: nucleic acid and protein modules for targeted integration and gene correction. *Trends Biotechnol* **23**: 399–406.

Koller BH, Smithies O. (1989) Inactivating the beta 2-microglobulin locus in mouse embryonic stem cells by homologous recombination. *Proc Natl Acad Sci USA* **86**: 8932–35.

Kostriken R, Strathern JN, Klar AJ, *et al.* (1983) A site-specific endonuclease essential for mating-type switching in *Saccharomyces cerevisiae*. *Cell* **35**: 167–74.

Lavery R. (2005) Recognizing DNA. *Q Rev Biophys* **38**: 339–44.

Lazowska J, Claisse M, Gargouri A, *et al.* (1989) Protein encoded by the third intron of cytochrome *b* gene in *Saccharomyces cerevisiae* is an mRNA maturase. Analysis of mitochondrial mutants, RNA transcripts proteins and evolutionary relationships. *J Mol Biol* **205**: 275–89.

Lazowska J, Jacq C, Slonimski PP. (1980) Sequence of introns and flanking exons in wild-type and box3 mutants of cytochrome *b* reveals an interlaced splicing protein coded by an intron. *Cell* **22**: 333–48.

Malik P, Arumugam PI. (2005) Gene therapy for {beta}-thalassemia. *Hematol Am Soc Hematol Educ Program*: 45–50.

Malik P, Arumugam PI, Yee JK, Puthenveetil G. (2005) Successful correction of the human Cooley's anemia beta-thalassemia major phenotype using a lentiviral vector flanked by the chicken hypersensitive site 4 chromatin insulator. *Ann N Y Acad Sci* **1054**: 238–49.

McCandless SE, Brunger JW, Cassidy SB. (2004) The burden of genetic disease on inpatient care in a children's hospital. *Am J Hum Genet* **74**: 121–27.

Moure C, Gimble F, Quiocho F. (2002) Crystal structure of the intein homing endonuclease PI-SceI bound to its recognition sequence. *Nat Struct Biol* **9**: 764–70.

Moure CM, Gimble FS, Quiocho FA. (2003) The crystal structure of the gene targeting homing endonuclease I-SceI reveals the origins of its target site specificity. *J Mol Biol* **334**: 685–96.

Orr-Weaver TL, Szostak JW, Rothstein RJ. (1981) Yeast transformation: a model system for the study of recombination. *Proc Natl Acad Sci USA* **78**: 6354–58.

Pabo CO, Nekludova L. (2000) Geometric analysis and comparison of protein-DNA interfaces: Why is there no simple code for recognition? *J Mol Biol* **301**: 597–624.

Pâques F, Duchateau P. (2007) Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Curr Gene Ther* **7**: 49–66.

Perrin A, Buckle M, Dujon B. (1993) Asymmetrical recognition and activity of the I-SceI endonuclease on its site and on intron-exon junctions. *EMBO J* **12**: 2939–47.

Porteus MH. (2006) Mammalian gene targeting with designed zinc finger nucleases. *Mol Ther* **13**: 438–46.

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

Posfai G, Kolisnychenko V, Bereczki Z, Blattner FR. (1999) Markerless gene replacement in *Escherichia coli* stimulated by a double-strand break in the chromosome. *Nucl Acids Res* **27**: 4409–15.

Puthenveetil G, Scholes J, Carbonell D, *et al.* (2004) Successful correction of the human beta-thalassemia major phenotype using a lentiviral vector. *Blood* **104:** 3445–53.

Rosen LE, Morrison HA, Masri S, *et al.* (2006) Homing endonuclease I-CreI derivatives with novel DNA target specificities. *Nucl Acids Res* **34**: 4791–800.

Sadelain M. (2006) Recent advances in globin gene transfer for the treatment of beta-thalassemia and sickle cell anemia. *Curr Opin Hematol* **13**: 142–48.

Samakoglu S, Lisowski L, Budak-Alpdogan T, *et al.* (2006) A genetic strategy to treat sickle cell anemia by coregulating globin transgene expression and RNA interference. *Nat Biotechnol* **24**: 89–94.

Scalley-Kim M, McConnell-Smith A, Stoddard BL. (2007) Coevolution of a homing endonuclease and its host target sequence. *J Mol Biol*, in press.

Schafer B, Wilde B, Massardo DR, *et al.* (1994) A mitochondrial group-I intron in fission yeast encodes a maturase and is mobile in crosses. *Curr Genet* **25**: 336–41.

Segal DJ, Beerli RR, Blancafort P, *et al.* (2003) Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry* **42**: 2137–48.

Segal DJ, Dreier B, Beerli RR, Barbas CF. (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5′-GNN-3′ DNA target sequences. *Proc Natl Acad Sci USA* **96**: 2758–63.

Seligman L, Chisholm KM, Chevalier BS, *et al.* (2002) Mutations altering the cleavage specificity of a homing endonuclease. *Nucl Acids Res* **30**: 3870–79.

Silva GH, Belfort M. (2004) Analysis of the LAGLIDADG interface of the monomeric homing endonuclease I-DmoI. *Nucl Acids Res* **32**: 3156–68.

Silva GH, Belfort M, Wende W, Pingoud A. (2006) From monomeric to homodimeric endonucleases and back: engineering novel specificity of LAGLIDADG enzymes. *J Mol Biol* **361**: 744–54.

Sitbon E, Pietrokovski S. (2003) New types of conserved sequence domains in DNA-binding regions of homing endonucleases. *Trends Biochem Sci* **28**: 473–77.

Smith J, Bibikova M, Whitby FG, *et al.* (2000) Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucl Acids Res* **28**: 3361–69.

Smith J, Grizot S, Arnould S, *et al.* (2006) A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucl Acids Res* **34**: e149.

Spiegel PC, Chevalier B, Sussman D, *et al.* (2006) The structure of I-CeuI homing endonuclease: evolving asymmetric DNA recognition from a symmetric protein scaffold. *Structure* **14**: 869–80.



Steuer S, Pingoud V, Pingoud A, Wende W. (2004) Chimeras of the homing endonuclease PI-SceI and the homologous *Candida tropicalis* intein: a study to explore the possibility of exchanging DNA-binding modules to obtain highly speciric endonucleases with altered specificity. *Chembiochem* **5**: 206–13.

Stoddard BL. (2005) Homing endonuclease structure and function. *Q Rev Biophys* **38**: 49–95.

Sussman DJ, Chadsey M, Fauce S, *et al.* (2004) Isolation and characterization of new homing endonuclease specificities at individual target site positions. *J Mol Biol* **342**: 31–41.

Themis M, Waddington SN, Schmidt M, *et al.* (2005) Oncogenesis following delivery of a nonprimate lentiviral gene therapy vector to fetal and neonatal mice. *Mol Ther* **12**: 763–71.

Thompson AJ, Yuan X, Kudlicki W, Herrin DL. (1992) Cleavage and recognition pattern of a double-strand-specific endonuclease (I-*Cre*I) encoded by the chloroplast 23S rRNA intron of *Chlamydomonas reinhardtii*. *Gene* **119**: 247–51.

Thrasher AJ, Gaspar HB, Baum C, *et al.* (2006) Gene therapy: X-SCID transgene leukaemogenicity. *Nature* **443**: E5–E6; discussion E6–E7.

Tzfira T, White C. (2005) Towards targeted mutagenesis and gene replacement in plants. *Trends Biotechnol* **23**: 567–69.

Urnov FD, Miller JC, Lee YL, *et al.* (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**: 646–51.

Vasquez KM, Marburger K, Intody Z, Wilson JH. (2001) Manipulating the mammalian genome by homologous recombination. *Proc Natl Acad Sci USA* **98**: 8403–10.

Volna P, Jarjour J, Baxter S, *et al.* (2007) Flow cytometric analysis of DNA binding and cleavage by cell surface-displayed homing endonucleases. *Nucl Acids Res* **35**: 2748–58.

Wang L, Jackson WC, Steinbach PA, Tsien RY. (2004) Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proc Natl Acad Sci USA* **101**: 16745–49.

Wintjens R, Rooman M. (1996) Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. *J Mol Biol* **262**: 294–313.

Wolfe SA, Nekludova L, Pabo CO. (2000a) DNA recognition by Cys2His2 zinc finger proteins. *Ann Rev Biophys Biomol Struct* **29**:183–212.

Wolfe SA, Ramm EI, Pabo CO. (2000b) Combining structure-based design with phage display to create new Cys2His2 zinc finger dimers. *Structure* **8**: 739–50.