Meganucleases (MNs) are highly specific enzymes that can induce homologous recombination in different types of cells, including mammalian cells. Consequently, these enzymes are used as scaffolds for the development of custom gene-targeting tools for gene therapy or cell-line development. Over the past 15 years, the high resolution X-ray structures of several MNs from the LAGLIDADG family have improved our understanding of their protein-DNA interaction and mechanism of DNA cleavage. By developing and utilizing high-throughput screening methods to test a large number of variant–target combinations, we have been able to re-engineer scores of I-CreI derivatives into custom enzymes that target a specific DNA sequence of interest. Such customized MNs, along with wild-type ones, have allowed for exploring a large range of biotechnological applications, including protein-expression cell-line development, genetically modified plants and animals and therapeutic applications such as targeted gene therapy as well as a novel class of antivirals.

Keywords: cell-line development/gene therapy/I-CreI/homing endonuclease/protein–DNA interaction engineering

Precise genome engineering requires the ability to direct biochemical activities involving protein-DNA (or ligand-DNA) interactions (e.g. DNA strand cleavage, homologous recombination and integration) to target, with a high degree of efficacy, sequence-specific sites. Meganucleases (MNs), also called homing endonucleases, are a class of highly sequence-specific and efficient enzymes that were discovered in yeast. The impact of a mitochondrial protein (an MN) on homologous recombination, first described in the 1980s (Jacquier and Dujon, 1985), was more fully studied and characterized, resulting in gene correction and insertion experiments in mammalian cells in the early 1990s (Rouet et al., 1994; Choulika et al., 1994, 1995; Smith et al., 1995; Donoho et al., 1998; Pierce et al., 2001). MNs can induce site-specific double-strand breaks (DSBs) and thereby stimulate homologous recombination by more than a 1000-fold in cultured cells (Donoho et al., 1998; Szczepek et al., 2007). Homologous recombination induced by MNs has been used in a variety of cell types and organisms, including mammalian cells, mice, plants, Drosophila, Escherichia coli and trypanosomes (Paques and Duchateau, 2007). MN-induced DSBs can also be repaired by non-homologous end-joining (NHEJ), an error-prone process that frequently results in micro-insertions or -deletions at the cleavage site (Liang et al., 1998).

MNs of the LAGLIDADG family: structure and function

MNs are generally encoded within introns or inteins although freestanding members also exist (Chevalier and Stoddard, 2001). The nomenclature of these proteins is similar to that of restriction endonucleases, with additional prefixes indicating different enzyme classes: I, for intron-encoded (e.g. I-CreI); PI, for intein-encoded (e.g. PI-SceI); and F, for freestanding (e.g. F-SceI) (Roberts et al., 2003).

MNs can be divided into five families based on sequence and structural motifs: LAGLIDADG, GIY-YIG, HNH, His-Cys box and PD- (D/E)XK (Orlowski et al., 2007; Zhao et al., 2007). The DNA targets of MNs consist of long (14–40 bp) sequences that are recognized and cleaved with high specificity in vitro and in vivo. Given the extended recognition sequence many MNs can tolerate target-site polymorphisms, with little or no loss in binding or cleavage activity. Nevertheless, such long target sites occur rarely in a whole genome; only one I-SceI target site is found in the 13 000 kbp yeast genome.

The most well studied family is that of the LAGLIDADG (Grishin et al., 2010) proteins, which can be found in all kingdoms of life. They catalyze the lateral transfer of their coding intron/intein by cleaving homologous alleles that lack the intron/intein sequence to initiate a recombination-dependent repair event that uses the intron/intein-containing allele as a repair template. LAGLIDADG proteins can exhibit up to two primary activities: (i) an RNA maturase activity that facilitates splicing of their own intron and/or (ii) a highly specific endonuclease activity resulting in the cleavage of the exon–exon junction sequence wherein their intron resides.

LAGLIDADG meganucleases (LMNs) can be homodimeric (e.g. I-CreI), targeting a palindromic or pseudopalindromic DNA sequence, or monomeric enzymes with two subdomains (e.g. I-SceI), targeting non-palindromic DNA sites (Fig. 1). Each monomer or subdomain contains a single...
copy of the LAGLIDAGG motif that forms a helix at the dimer interface, with the conserved penultimate acidic residues forming part of the active site. Several LAGLIDAGG protein structures have been solved alone or in complex with their DNA target, providing insight into their mechanisms of DNA recognition, binding and catalysis (Heath et al., 1997; Jurica et al., 1998; Silva et al., 1999; Chevalier et al., 2002, 2003, 2004; Moure et al., 2002, 2003; Bolduc et al., 2003; Spiegel et al., 2006; Marcaida et al., 2008; Redondo et al., 2008).

LMN subdomains adopt a similar αβαβαβα fold, with the LAGLIDAGG motif constituting the terminal region of the first helix. Two such αβ domains assemble to form the functional protein, with the resulting β-sheets creating a saddle-shaped DNA binding interface. An assortment of LMN structures in complex with their DNA target allowed for pinpointing the main determinants of the protein-DNA interaction (Stoddard, 2005). Protein structures were also useful for understanding the main determinants of the protein-DNA interaction (Stoddard, 2005). Protein-DNA interaction energy calculations have been used to modify the specificity of LMNs (Ashworth et al., 2006, 2010).

The coexpression of two I-CreI variants in the same cell yields a heterodimeric species allowing for the cleavage of a target of interest (Arnould et al., 2006, 2007; Smith et al., 2006). To abolish homodimer formation that results from the protein coexpression, an obligate heterodimer (Fajardo-Sanchez et al., 2008) and/or single-chain molecule (Epinat et al., 2003; Grizot et al., 2009a,b) strategy is applied. In vitro and in vivo studies have confirmed that many of the active variants obtained by this method preserve the essential properties of the wild-type I-CreI scaffold, including structure, stability, cleavage efficiency and narrow sequence specificity.

In parallel, several novel monomeric MN chimeras have also been generated from homodimeric proteins (I-CreI and I-MsoI) or from domain swapping between subdomains of I-Dmol and I-CreI. These active chimeric MNs have hybrid specificity derived from that of each initial half-parents target (Chevalier et al., 2002; Epinat et al., 2003; Silva et al., 2006; Grizot et al., 2009a,b; Li et al., 2009). Monomeric MNs that require the vectorization of a single polypeptide represent a potential improvement for gene-targeting applications.

**MN applications: in cellulo research tools**

The first experiments of the 1990s enabled modest gene-targeting frequencies of 10⁻³–10⁻⁴ (per cell) (Rouet et al., 1994; Choulia et al., 1994, 1995; Smith et al., 1995). Various levels of recombination have been observed since (reviewed in Paques and Duchateau, 2007), with up to 10% recombination achieved using I-SceI in HEK-293 cells (Szczepke et al., 2007).

As MNs are highly sequence specific and facilitate targeted homologous recombination in different cell types from various species (Smith et al., 2006; Arnould et al., 2007; Grizot et al., 2009a,b), these enzymes are useful templates for the development and commercialization of research tools dedicated to in cellulo genome customization, including transgene insertion (knock-in), gene disruption (knock-out) and modulation of gene expression.

For very sequence specific knock-in purposes, Cellectis Bioresearch has already developed dedicated products and protocols amendable for transgene integration into several types of secondary cell lines, e.g. CHO-K1, CHO-S, NIH-3T3 or HEK-293. Two lines of knock-in products have been successfully created: (i) cGPS(R) (cellular Genome Positioning System), in which the cell line has been engineered to incorporate a single copy of a DNA fragment that contains not only the target site of a natural MN (I-CreI or I-SceI) but also genetic elements required for selection and (ii) cGPS Custom(R), which dispenses the pre-engineered
cell line in favor of a custom-made MN engineered to specifically recognize an endogenous target site in a dedicated genome (Cabaniols and Paques, 2008).

In association with Servier laboratories, a case study was undertaken to compare the cGPS CHO-K1 system versus classical transfection methods for the development of cell-based assays dedicated to high-throughput screening (Fig. 2) (Cabaniols et al., 2009). From the results gained, cGPS presents a clear advantage as it is shown to be more rapid (2-week protocol), very efficient (>95% of isogenic targeted integration) and allows for homogenous and stable expression of different transgene elements over time (at least 23 weeks of culture).

**MN applications: gene targeting**

Most current gene therapy strategies for inherited diseases are based on a complementation approach: a virus-borne functional copy of the variant gene is randomly inserted into the genome, resulting in a phenotypic correction of the genetic defect (Hacein-Bey-Abina et al., 2008; Howe et al., 2008; Stein et al., 2010). However, this approach has several limitations, including the risk of transgene silencing, potential disruption of endogenous genes or transcriptional activation of neighboring genes such as proto-oncogenes. An alternative approach for monogenic inherited diseases involves the use of MNs to precisely engineer a chosen locus.

The ‘ideal’ approach to gene therapy is to use MNs to directly correct the mutated gene. However, this approach has practical limitations since the correction efficiency decreases rapidly as the distance from the initial DNA DSB increases. Thus, this approach can be more easily envisioned as a treatment for monogenic diseases in which a prevalent mutation is responsible for the majority of cases. A more general strategy is to use MNs to integrate a complete or partial cDNA of the affected gene upstream of a deleterious mutation. Alternatively, MNs could be used to promote site-specific integration at a ‘safe harbor’, a locus that has been chosen to minimize the probability of insertional mutagenesis as well as to maintain long-term and high-level transgene expression.

Several studies have found the frequency of successful induced gene-targeting events to range from $10^{-2}$ to $10^{-1}$ (per cell) when using a chromosomal reporter system and I-SceI in embryonic stem cells (Smith et al., 1995; Alwin et al., 2005). Recently, I-CreI derived MNs were shown to successfully induce efficient targeted recombination in endogenous genes at two different human loci, XPC and RAG1 (Smith et al., 2006; Arnould et al., 2007; Grizot et al., 2009a,b). The RAG1 MN could be used to induce 6% recombination in 293 cells with a relative specificity comparable to I-SceI (Grizot et al., 2009a,b). These frequencies should be sufficient for gene repair strategies in which the corrected cells have a selective advantage, such as with severe combined immunodeficiency (SCID). However, for other therapeutic purposes a selection scheme may be necessary to enrich for targeted cells.

**MN applications: virus clipping**

MNs targeting viral sequences represent a novel class of antiviral agents. Many chronic viral infections are due to double-stranded DNA viruses or viruses that involve a double-stranded DNA intermediate during their replicative...
Thus, MNs could be used to specifically cleave and either partially excise or eliminate viral DNA from infected cells and thus render them virus free (Fig. 3). The advantage of this approach, compared with standard anti-viral treatments that block various steps of viral replication, is the possibility of targeting the latent form of the virus.

Experiments with recombinant Herpes simplex virus 1 (HSV-1) containing an I-SceI recognition site demonstrated that viral replication could be dramatically reduced in cells that have been transfected with an I-SceI expression plasmid prior to infection (Galetto et al., 2009). Furthermore, I-CreI derived MNs designed to specifically cleave the HSV-1 genome were shown to have a significant inhibitory effect on HSV1 infection at low and moderate multiplicities of infection (i.e. 10^-3 – 1), reducing the viral load by >50% compared with control cells (Grosse et al., submitted for publication). For viruses in which the viral genome is integrated into the host cell, initial experiments suggest that cleavage in both LTRs may result in viral genome depletion, by ligation of the two generated breaks or tandem repeat recombination (Choulika et al., 1994; Liang et al., 1998; Perez et al., 2005). Hence, MNs may be developed into broadly useful antiviral compounds possibly providing efficient therapies where no cure is currently available.

**Conclusion**

Previous and ongoing studies have validated MNs as bona fide bioresearch tools with practical and therapeutic applications. Work done with I-SceI-mediated recombination in diverse organisms such as bacteria (Flanagan et al., 2008; Yu et al., 2008), mosquito (Windbichler et al., 2007), fly (Rong and Golic, 2000; Muggert et al., 2008) plant (Puchta, 2002; Yang et al., 2009) and transgenic mice (Gouble et al., 2006) has proved the potential of the technology, and is being paralleled by work with I-CreI or I-CreI derivatives in Drosophila (Rong et al., 2002), plants (Gao et al., 2010) and mammalian cells (Cabaniols and Paques, 2008; Grizot et al., 2009a,b). At the same time, Zinc-finger Nucleases, another promising type of rare-cutting endonuclease, have been used for a large number of similar applications (for review, see Porteus and Carroll 2005), including targeted gene correction or insertion (Urnov et al., 2005), gene knock-out (Perez et al., 2008) and virus clipping (Cradick et al., 2010). For all rare-cutting endonucleases, one of the major goals is to achieve the highest level of specificity, especially for therapeutic applications (Carroll, 2008). For LAGLIDADG MNs, new approaches include the use of targeted ‘mega-nickases’ (Niu et al., 2008; McConnell Smith et al., 2009) to provide high levels of induced homologous recombination with a minimization in the frequency of NHEJ. Nevertheless, in any application, MN genotoxicity will need to be assessed as it represents an inherent potential problem of enzymes that act on nucleic acids, even when the apparent DNA specificity indicates otherwise.

**Funding**

This work has been funded by Cellectis.

**References**


