Three new active members of the I-OnuI family of homing endonucleases
Iman M. Bilto, Tuhin K. Guha, Alvan Wai, and Georg Hausner

Abstract: In vitro characterization of 3 LAGLIDADG-type homing endonucleases (HEs) (I-CcaI, I-CcaII, and I-AstI) that belong to the I-OnuI family showed that they are functional HEs that cleave their respective cognate target sites. These endonucleases are encoded within group ID introns and appear to be orthologues that have inserted into 3 different mitochondrial genes: rns, rnl, and cox3. The endonuclease activity of I-CcaI was tested using various substrates, and its minimum DNA recognition sequence was estimated to be 26 nt. This set of HEs may provide some insight into how these types of mobile elements can migrate into new locations. This study provides additional endonucleases that can be added to the catalog of currently available HEs that may have various biotechnology applications.

Key words: mtDNA, mobile introns, protein purification, endonuclease assays.

Résumé : La caractérisation in vitro de 3 endonucléases de homing de la classe LAGLIDADG (I-CcaI, I-CcaII et I-AstI) appartenant à la famille I-OnuI a montré qu’elles constituent des endonucléases de homing fonctionnelles qui clivent leurs sites cibles cognats respectifs. Ces endonucléases sont codées dans des introns du groupe ID et semblent être des orthologues qui se sont insérés à l’intérieur de 3 gènes mitochondriaux différents: rns, rnl et cox3. L’activité endonucléase de I-CcaI a été testée à l’aide de différents substrats et sa séquence d’ADN de reconnaissance minimale était estimée à 26 nt. Cet ensemble d’endonucléases de homing peut fournir un aperçu de la manière par laquelle ces types d’éléments mobiles peuvent migrer vers d’autres sites nouveaux. Cette étude ajoute de nouvelles endonucléases au catalogue d’endonucléases de homing actuellement disponibles qui peuvent avoir différentes applications biotechnologiques. [Traduit par la Rédaction]

Mots-clés : ADNmt, introns mobiles, purification des protéines, dosage de l’activité endonucléase.

Introduction

Homing endonucleases (HEs) require long DNA target sites (12 to 40 bp) and, therefore, are highly specific DNA cleavage enzymes (Belfort and Roberts 1997; Chevalier and Stoddard 2001; Lambert et al. 2016). Homing endonuclease genes (HEGs) are embedded within elements such as group I and sometimes group II and archaeal introns (Dujon 1989; Belcour et al. 1997; Toor and Zimmerly 2002; Mullineux et al. 2010; Tocchini-Valentini et al. 2011). HEGs can also be freestanding and they can form the DNA-cutting component of inteins (Gimble 2000; Hafez and Hausner 2012). HEGs are mobile genetic elements that can generate double-stranded DNA breaks in a cognate allele, inducing the cell’s DNA double-strand break repair mechanism to repair the break by using the HEG-containing DNA as a template (Dujon 1989; Belfort et al. 2002; Burt and Trivers 2006). HE activity not only results in repair that involves the nonreciprocal transfer of the HEG into new sites, but it also can be associated with co-conversion of markers flanking the HEG (or intron) insertion site (Mueller et al. 1996a, 1996b; Parker et al. 1999; Muñoz et al. 2012).

HEs can be assigned to 6 categories on the basis of presence of conserved amino acid motifs. Concerning inteins and group I and group II introns, 4 types are the most relevant: LAGLIDADG (LHE), GIY-YIG, H-N-H, and His-Cys box (Stoddard 2005; Skowronek and Bujnicki 2007; Hafez and Hausner 2012). The LAGLIDADG and GIY-YIG families are most frequently encountered in fun-
gal mitochondrial group I introns (Haugen et al. 2005; Hausern 2012). LHEs generate staggered cuts with 4 nt 3′ overhangs at their cleavage site (Belfort and Roberts 1997; Chevalier and Stoddard 2001). LHEs in their active configuration can act as homodimers when the peptide contains 1 LAGLIDADG motif or they can act as monomers when the peptide contains 2 LAGLIDADG motifs (Dalgaard et al. 1997; Lucas et al. 2001; Haugen and Bhattacharya 2004). Some HEs have been shown to tolerate some sequence degeneracy within their recognition sites, which gives HEGs (and their host element) the flexibility to invade new sites and thus provides the possibility of HEGs to avoid elimination by genetic drift (Scalley-Kim et al. 2007; Barzel et al. 2011; Edgell et al. 2011). Specificity with some allowance for variations among alleles makes HEs potential tools in genome editing such as gene replacements, targeted mutagenesis, gene drive mechanisms for pest control, and in designing vectors systems that require unique endonuclease target sites (Hafez and Hauser 2012). Fungal mitochondrial genomes have shown to be a resource for isolating new HEs that have new target specificities (Stoddard 2011; Baxter et al. 2012; Jacoby et al. 2012; Chan et al. 2013; Hafez et al. 2014a). Overall, few HES have been biochemically characterized so far (Marcaida et al. 2010; Prieto et al. 2012), and this limits the application of HES in targeting a wide variety of sites.

Previously, Monteiro-Vitorello et al. (2009) reported an intron within the mtDNA rns gene inserted at position ms915 (designation of rns intron positions and naming is according to the nomenclature proposed by Johansen and Haugen 2001: S, small subunit ribosomal gene; m, mitochondria) in the chestnut blight fungus Cryptonectria parasitica. Hafez et al. (2013) reannotated this intron and noted that it is actually inserted at position m917; this study showed that this intron is a nested (or twintron) group ID intron where the internal group ID intron encodes a double motif LAGLIDADG-type open reading frame (ORF) and it is inserted into an ORF-less external group ID intron. Further, this study noted that orthologues of this HEG are present in other fungal species encoded within group ID introns inserted in rnl, cox3, nad5, and nad6 genes. Exploration of these HES may provide some clues on how HEGs can move into different sites thereby promoting their spread within a genome, within a population, or between different species. The objectives for this study were to identify and characterize active members of the ms917 HES family (i.e., orthologues of the ms917 LHE). The study focused on 3 members that are inserted within 3 different genes: (i) the ms917 LHE version that is encoded within an rns intron at position S917 in Ceratocystis cacaofunesta (causative agent of wilt disease in cacao), (ii) a LHE encoded within intron 1 in the rnl gene in C. cacaofunesta (Ambrosio et al. 2013), and (iii) a LHE that is encoded within the third intron of cox3 in Annulohypoxylon stygium (an endophyte).

Materials and methods

Phylogenetic analysis of HE sequences

The LHE data set was based on sequences extracted from the NCBI database with the ms917 intron-encoded LHE amino acid (aa) sequence of C. cacaofunesta (GenBank accession AFO38132.1) used as a query in blastp. Forty sequences were extracted from GenBank and aligned with MAFFT (Katoh and Standley 2013) and PRALINE (Simossis and Heringa 2005). If necessary, manual adjustments were made to the alignment with GeneDoc (Nicholas et al. 1997). The aa alignment was analyzed with neighbour joining (Saitou and Nei 1987) as implemented in the MEGA (version 7) program (Kumar et al. 2016). Distances were calculated with JTT (Jones et al. 1992) and its default settings were selected along with the complete deletion of gap option. The bootstrap option was implemented (1000 replicates) to assess the level of support for the tree topology (Felsenstein 1985).

HEs naming nomenclature

HEs are named according to the nomenclature proposed by Belfort and Roberts (1997). HES encoded by group I introns have the prefix “i-” followed by abbreviations for the genus and species name. The final Roman numeral distinguishes multiple enzymes that have been characterized for that organism. The HES examined in this study are designated as follows: i-Ccal for the intron-encoded protein (IEP) encoded by the rns1 of C. cacaofunesta, I-Ccall for the IEP encoded by the rnl1 of C. cacaofunesta, and I-Astl for the IEP encoded by the cox3i3 of A. stygium.

Construction of expression vectors and substrate plasmids

The HE sequences were retrieved from GenBank: I-Ccal (AFO38132.1), I-Ccall (AFO38136.1), and I-Astl (AHB33504.1). To allow for efficient expression in Escherichia coli, the genetic code for these HE ORFs were codon optimized for E. coli. The pET-28b(+) vector was used for assembling the HE expression constructs (GenScript, New Jersey, USA). The LHE ORFs were inserted at the BamHI or NdeI restriction enzyme site and the vector provided an N-terminal 6× Histidine (His)-tag and the T7 promoter. The 3 expression vectors were named pl-Ccal, pl-Ccall, and pl-Astl.

Suitable substrates for testing the above LHEs were designed as follows. The corresponding host gene sequences were examined and 100 nt of exon sequences flanking the intron insertion sites (~200 nt) were combined, synthesized, and inserted into the pUC57 (2.7 kb) vector (GenScript, New Jersey, USA). These substrates were named as follows: prms-SUB for testing the C. cacaofunesta rns1 IEP (I-Ccal), prnl-SUB for testing the C. cacaofunesta rnl1 IEP (I-Ccall), and pcox3-SUB for testing the A. stygium cox3i3 IEP (I-Astl).

All plasmids were maintained in E. coli DH5α (Thermo Fisher Scientific, Burlington, Ontario) and the plasmids
were purified with the Presto™ Mini Plasmid kit (FroggaBio, Toronto, Ontario). Transformed cells were kept at −80 °C in 80% glycerol stocks (Sambrook et al. 1989) and purified plasmids were stored at −20 °C.

Minimum recognition sequence and substrate plasmids for the I-CcaI
In addition to the prns-SUB, a set of 8 substrate plasmids (prns-SUB2 to -SUB9) were designed to estimate the minimum recognition sequence required for the LHE activity for the mS917 IEP (I-CcaI) of C. cacaofungea. Essentially, the original substrate sequence was modified by shortening the potential recognition sequence 2 nt at a time on either side of the LHE cleavage site by inserting nucleotide triplets that disrupt the potential LHE recognition sequence (Bae et al. 2009; see supplementary Fig. 1). Expression and purification of recombinant proteins (HEs)
Conditions for the overexpression and purification of the LHE proteins have been described previously in Hafez et al. (2014a, 2014b). Briefly the LHE expression constructs were transferred into chemically competent E. coli BL21 (λDE3) cells (New England Biolabs, Whitby, Ontario) for protein expression, which was induced with isopropyl β-D-thiogalactopyranoside (IPTG) when cells reached an absorbance of 0.50 at OD 600. The purification of the His-tagged LHE proteins was performed on Ni-NTA columns. All the washing and elution fractions were collected and analyzed on a SDS polyacrylamide gel (12.5%). Electrophoresis was used to evaluate HE protein purity. Samples were dialysed in di-alysis buffer (50 mmol/L Tris–HCl (pH 8.0), 100 mmol/L NaCl, 6 mmol/L β-mercaptoethanol) using the Slide-A-Lyzer™ dialysis cassette (Millipore, Billerica, Massachusetts, USA) with a 10 kDa molecular mass cutoff to remove the imidazole. Dialysed samples were concentrated to 1 mL with the Amicon concentrator (Model 8050) using a Millipore YM-10 membrane. Usually, the purified, concentrated protein (LHE) was immediately assayed for its endonuclease activity before storing. However, for long-term storage, 200 μL of protein storage buffer (50 mmol/L Tris–HCl (pH 8.0), 400 mmol/L NaCl, 0.5 mmol/L DTT, 10% (m/v) glycerol) was added to the purified protein and the samples were stored at −80 °C.

Endonuclease assays
The purified HEs were challenged with the appropriate DNA substrate plasmids to assess their endonuclease potential. Methodology is based on those previously pre-sented in Hafez et al. (2014a, 2014b). Briefly, the endonuclease reaction mixtures contained the following: 25 μL of substrate plasmid (25 ng/μL), 5 μL of Invitrogen® Buffer React #3 (100 mmol/L NaCl, 50 mmol/L Tris–HCl (pH 8.0), and 10 mmol/L MgCl2) supplemented with 1 mmol/L DTT, 5 μL of LHE protein (9 μg), and H2O to reach a final volume of 50 μL. The endonuclease reactions were incubated at 37 °C, and 10 μL aliquots were withdrawn at the following time points: 0, 30, 60, 90, and 120 min. The endonuclease reactions were stopped by adding 2 μL of 200 mmol/L EDTA (pH 8.0) and 1 μL of proteinase K (1 mg/ml) to the reaction aliquots. Finally, the reaction products were resolved on a 1% agarose gel ran at 80 volts and gels were stained with ethidium bromide (0.5 μg/mL).

Mapping of the I-CcaI, I-CcaII, and I-AstI cleavage sites
The in vitro cleavage site mapping strategy was based on Bae et al. (2009) and has been previously described (Guha and Hausner 2014; Hafez et al. 2014a, 2014b). The substrate plasmids were digested with the corresponding HE described above, and the linearized substrates were cut from the gel and purified with the Wizard® SV Gel and PCR Clean-Up system (Promega; Thermo Fisher Scientific). Forty microliters of linearized substrate was treated with T4 DNA polymerase (T4 DNA pol) to generate blunt ends by removing the characteristic 4 nt 3'-OH overhangs generated by LAGLIDADG-type HEs (Bae et al. 2009). The T4 DNA pol treatment reaction contained the following components: 40 μL of linearized substrate plasmid (25 ng/μL), 2 μL of T4 DNA pol (5 units/μL), 20 μL of 5× T4 DNA polymerase buffer, 20 μL of dNTP mixture (0.5 mmol/L), and H2O to achieve the final volume of 100 μL. The reaction was incubated at room temperature (24 °C) for 20 min, and thereafter the tubes were placed on ice for 5 min and finally the reactions were terminated by transferring to 70 °C for 10 min. The DNA was purified with the Wizard-SV Gel and PCR Clean-Up system (Promega). The blunted substrate DNA was religated in the following reaction mixture: 20 μL (0.25 μg) of T4 DNA pol-treated DNA, 2 μL of T4 DNA ligase (1 U/μL; Invitrogen), 10 μL of 5× ligase buffer, and the addition of H2O to achieve a final volume of 50 μL. Ligation reactions were incubated at room temperature for 2 h. Thereafter, the ligation mixtures were diluted 5-fold, and a 10 μL aliquot of the dilution was transformed into chemically competent E. coli DH5α cells. The transformed cells were plated onto LB agar plates supplemented with ampicillin (100 μg/mL). Single colonies were picked from these plates and these were used to inoculate 5 mL of LB broth (100 μg/mL ampicillin) tubes that were incubated for ~18 h at 37 °C. The plasmids were purified by the Wizard® Plus Minipreps DNA Purification system (Promega) and they were sent to the DNA Technologies Unit (NRC, Saskatoon, Saskatchewan) for cycle sequencing using the M13 forward and reverse primers. Untreated substrate plasmids were sent along
ORFs were double-motif LAGLIDADG HEGs and their group ID does not contain an ORF (Hafez et al. 2013). Internal group ID intron encodes the LHE and external are complex and appear to be nested introns where the mS917 introns in inserted into at least 5 different genes, including rRNA. Logues appears (bootstrap support of node = 98%) to have HE protein overexpression in E. coli and subsequent characterization were noted to be difficult for several members of the ms917 group. Although all sequences were codon optimized for expression in E. coli, I-Cpa-917 (ms917 nested version; GenBank accession No. AAB84210.1) and I-CcalIII (nad5; GenBank accession No. AFO38108.1) could be expressed but showed no activity, and I-CcalV (nad6; GenBank accession No. AFO38135.1) failed to express.

**Temperature requirement and the minimal DNA recognition sequence for I-Ccal**

The I-Ccal HE (mS917 IEP) was evaluated with regards to its activity at different temperatures. Endonucleases assays were set up as described previously except the reactions were incubated for 1 h at 20, 30, 37, 40, and 50 °C. In addition, the I-Ccal LHE was further characterized concerning its minimum recognition sequence requirement. Here, the LHE was challenged with various substrate plasmids containing sequences modified to provide various lengths of the putative recognition sequence. The preparation of the substrate plasmids (prns-SUB2 to -SUB9 have been described above) and the endonuclease assays were performed at 37 °C for 2 h as described above.

**Results**

**Phylogeny of ms917 HEs and related LAGLIDADG-type ORFs**

Forty LHE sequences were extracted from NCBI by using the ms917 IEP (I-Ccal) sequence from C. cacaofunesta as a query. Phylogenetic analysis with neighbour joining (Fig. 1) showed that 10 sequences, all encoded within group ID introns, appear to be derived from a common ancestor. This grouping includes 4 ms917 IEPs along with 3 IEPs encoded within an intron located within the nad5 gene, 1 IEP encoded within the mls gene, 1 IEP from the cox3 gene, and 1 IEP from the nad6 gene. This set of orthologues appears (bootstrap support of node = 98%) to have inserted into at least 5 different genes, including rRNA and protein-coding genes. It is also worth noting that the ms917 introns in Ophiocordyceps tricentri and C. parasitica are complex and appear to be nested introns where the internal group ID intron encodes the LHE and external group ID does not contain an ORF (Hafez et al. 2013). Originally, the goal was to characterize a member from each host gene (including nad5 and nad6) and from the nested ms917 intron arrangements; however, we could only overexpress IEPs and show activity for IEPs from C. cacaofunesta (rns), C. cacaofunesta (mls), and A. stygium (cox3).

**HE protein overexpression in E. coli and purification**

In total, 10 ms917 LHEs and related LAGLIDADG HEs were identified and protein overexpression was attempted for 3 LHEs ORFs embedded within introns located in the rns, mls, and cox3 genes. All of the studied ORFs were double-motif LAGLIDADG HEGs and their lengths ranged from 433 to 551 aa. Three enzymes were expressed successfully and purified (supplementary Fig. 2). I-AstI (50 kDa) and I-CcalII (47 kDa) were partially purified, whereas I-Ccal HE (49.5 kDa) was consistently overexpressed in E. coli and purified from Ni-NTA columns. Therefore, this particular LHE was further analyzed concerning temperature preference and for estimating its target sequence length requirement for DNA recognition and cleavage activity.

The overexpression of LHE ORFs in E. coli and subsequent characterization were noted to be difficult for several members of the ms917 group. Although all sequences were codon optimized for expression in E. coli, I-Cpa-917 (ms917 nested version; GenBank accession No. AAB84210.1) and I-CcalIII (nad5; GenBank accession No. AFO38108.1) could be expressed but showed no activity, and I-CcalV (nad6; GenBank accession No. AFO38135.1) failed to express.

**Endonuclease activity of the ms917 IEP orthologues**

Endonuclease assays were performed by incubating the purified LHEs (I-Ccal, I-AstI, I-CcalII) with the appropriate plasmid substrates (prns-SUB, pcox3-SUB, and prns-SUB). The HE activity was tested at 37 °C at different time periods (0, 30, 60, 90, and 120 min), and the best completely linearized substrates (2.8 kb) were already observed at 30 min (Fig. 2) for I-Ccal, I-AstI, and I-CcalII. The control assays, i.e., untreated (no HE) substrates (prns-SUB, pcox3-SUB, prns-SUB), showed no cleavage.

The in vitro cleavage mapping assay for I-Ccal, I-AstI, and I-CcalII confirmed that these LHEs are active and generate staggered ends with a 4 nt overhang at the 3’ end (Fig. 3). The HE-cleaved and T4 DNA pol-treated substrates (prns-SUB, pcox3-SUB, prns-SUB) showed that the following 4 nt were removed: 5’-TAAT-3’, 5’-ATAC-3’, and 5’-ATGC-3’ by I-Ccal, I-AstI, and I-CcalII, respectively.

I-Ccal activity was tested at different temperatures (20, 30, 37, 40, and 50 °C) (Fig. 4). The enzyme showed activity at 30, 37, and 40 °C, as the expected linearized product can be observed on the agarose gel. However, assays performed at 20 and 50 °C yielded multiple products indicating partial or no activity, as the substrate plasmid appears in its various supercoiled forms on the agarose gel (Fig. 4).

The I-Ccal enzyme was further characterized to estimate the minimal recognition sequence. The I-Ccal endonuclease was challenged with 8 different synthesized substrates containing variations of the known target sequence. The results showed that the enzyme lost its activity when the target sequence was less than 14 nt upstream of the actual cleavage site and less than 12 nt downstream from the cleavage site, i.e., the enzyme required a DNA recognition site that is estimated to be about 26 nt in length (Fig. 5).
Fig. 1. Phylogenetic tree inferred using the neighbour-joining method. The optimal tree with the sum of branch length = 26.1071834 is shown. The percentages at the nodes are based on bootstrap analysis (1000 replicates). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the JTT matrix-based method (Jones et al. 1992) and are in the units of the number of amino acid substitutions per site. The analysis involved 40 amino acid sequences and ambiguous alignment positions were removed for each sequence pair. There were a total of 781 positions in the final data set. The node marked with black pentagon supports the monophyly of the mS917 clade of homing endonucleases. The black circles indicate the position of those members that were tested in this study for activity. The node marked by the black square supports the monophyly of the I-Onul family of homing endonucleases.
Fig. 2. Gel images for the in vitro endonuclease assays. (a) I-CcaI was incubated with the circular prns-SUB plasmid at 37 °C for 30 min. (b) I-CcaII was incubated with the circular prnl-SUB plasmid for 30 min. (c) I-AstI was incubated with the circular pcox3-SUB plasmid at room temperature overnight. Lanes denoted “L” contain the 1 kb plus DNA ladder (Thermo Fisher Scientific); lanes denoted “C” contain the untreated plasmid substrate.

Discussion

To avoid elimination, HEGs have to either gain new functions or insert into new sites (Goddard and Burt 1999). The mS917 clade of HEs appears to have inserted into at least 5 different target sites. This observation is also relevant to the application of LHEs as rare cutting DNA enzymes in biotechnology. Considerable efforts have been made to redesign HEs to bind and cut at different target sites. This study presents examples where closely related LHEs have naturally adapted to different target sites. This warrants further investigation in the future to establish what features and/or amino acid changes might be involved in allowing LHEs to recognize new target sites.

The phylogenetic tree (Fig. 1) shows that the mS917 LHE orthologues belong to the I-Onul family of HEs (node supported at 100%). The I-Onul HE was originally described and characterized by Gibb and Hausner (2005) and Sethuraman et al. (2009). This endonuclease was studied in more detail by Takeuchi et al. (2011), who showed that I-Onul can be engineered to recognize and cut genes involved in monogenic human diseases. For example, the MAO-B gene (involved in neurodegenerative disorders including Parkinson’s disease), which contains a DNA sequence that differs from the native I-Onul target site by only 5 bp (Takeuchi et al. 2011). Recently, Lambert et al. (2016) further demonstrated the potential utility of this family of LHEs in biotechnology applications. This information might be useful in developing strategies for engineering LHEs to target sites located within genes of economic importance, such as alleles associated with human diseases.

The LAGLIDADG family of HEs have been explored as potential genome editing tools (Gimble 2000; Stoddard 2005, 2011, 2014). Therapeutic applications demand high precision in gene modification activity and HEs are considered compact target-specific “molecular scissors” with little known issues with regards to off-target activities (Stoddard 2014; Takeuchi et al. 2014; Cox et al. 2015; Lambert et al. 2016). However, one potential drawback for this class of enzymes is the nonmodular configuration; the DNA recognition and cleavage functions are combined within the same protein domain, thus engineering of LHEs is challenging (Hafez and Hausner 2012).

One of the limitations of the I-Onul family, with regards to retargeting them for genome editing applications, is the conserved nature of the 4 bp sequence comprising the central recognition motif (CRM, i.e., the 3’ cleavage site overhangs). The 3 members of the mS917 clade characterized in this study displayed some variability within the CRM (5’-TAAT-3’ for I-CcaI, 5’-ATAC-3’ for I-AstI, and 5’-ATGC-3’ for I-CcaII). This feature could make the mS917 clade an attractive LHE protein scaffold that could be modified for targeting a wide variety of potential target sites (Lambert et al. 2016).
Fig. 3. Cleavage site mapping results. Schematic for the rns (a), rnl (b), and cox3 (c) substrates and the cleavage sites for I-CcaI, I-CcaII, and I-AstI, respectively. The staggered cuts and the 4 central recognition motifs for the homing endonucleases are indicated. The 4 nt 3’ overhangs generated by each of the enzyme and removed by T4 DNA polymerase are shown in red. The intron insertion sites were designated based on reference sequences; for the rns and rnl genes the insertion sites (IS) are based on Escherichia coli rDNA (AB035922.1), as proposed by Johansen and Haugen (2001). The cox3 intron IS is based on comparison with the Saccharomyces cerevisiae cox3 sequence (KP263414.1). [Colour online.]

Fig. 4. Effect of temperature on I-CcaI endonuclease activity. The plasmid substrate was incubated with I-CcaI at different temperatures (ranging from 20 to 50 °C). The lanes denoted “C” (control) contain the untreated substrate plasmid, while the lanes denoted “L” contain the 1 kb plus DNA ladder (Thermo Fisher Scientific). The arrow indicates the position of the linearized plasmids.
Fig. 5. Gel images for the I-CcaI cleavage assays with various substrates (see supplementary Fig. 11). I-CcaI was incubated for 30 min with 9 different plasmids (control prn5-SUB and prn5-SUB2 to prn5-SUB9) containing various length versions of the target region for I-CcaI. Lanes denoted “L” contain the 1 kb plus DNA ladder (Invitrogen); lanes denoted “U” contain the untreated (no homing endonuclease) corresponding plasmid substrate (prn5-SUB to prn5-SUB9); Lanes denoted “EcoRI” contain positive controls where the substrates were incubated with the EcoRI restriction enzyme to show the migration of the cleaved substrate plasmids. Panel a shows the results for substrates that failed to be cut by I-CcaI; panel b shows the results for those substrates that were cleaved by I-CcaI. Substrate plasmids that are untreated (U) or have been treated with EcoRI are the negative and positive controls, respectively. Note: c, position of cut plasmids; u.r., the position of uncut relaxed plasmids; u.s., the position of uncut supercoiled plasmids. Panel c indicates how many nucleotides upstream and downstream (upstream/downstream) of the cleavage site (indicated by arrowhead) are present within the corresponding substrate plasmid. Uncut plasmids appear to behave as supercoiled and relaxed plasmids, thus their migration patterns on the gel are quite variable, migrating faster or slower in comparison with linearized versions.
Recent studies showed that the central 4 nt within the CRM are not in direct contact with the LHE, but they are still very important for controlling enzyme activity, as this region undergoes bending during cleavage (Curuksu et al. 2009; Lambert et al. 2016). This is referred to as “indirect readout”, where protein–DNA interactions are due to sequence-dependent interactions (involving conformation and flexibility of the DNA) between base pairs and amino acid sequences and not due to direct interactions (H-bonding) between amino acids and nucleotides; this has been observed with regards to restriction enzymes and HEs (Molina et al. 2012; Yamasaki et al. 2012).

The study showed that I-CcaI, I-CaII, and I-AstI can be overexpressed in E. coli and purified utilizing the N-terminal His-tag in sufficient amounts to demonstrate their activity as endonucleases and to map their cleavage sites. The most consistent LHE in our analyses with regards to overexpression and purification was I-CcaI. It was noted that the optimum activity was at 37 °C, and activity was highly reduced or lost at 20 and 50 °C, respectively. In previous studies, it was shown that endonuclease activity for LHEs tends to be around 37 °C (see I-SceI group and I-CreI group; Dürrenberger and Rochaix 1993; Fonfara et al. 2012). In contrast, the I-DmoI LHE endonuclease, recovered from the thermophile Desulfuroccocus mobilis, shows optimal activity at 65 °C (Dalgaard et al. 1993; Silva and Belfort 2004).

Challenging the I-CcaI with 8 different synthesized substrates allowed for gaining better insights into the length of the DNA sequence required by the LHE, which appears to be around 26 nt. This is in the expected range for LHEs belonging to the I-ONuI family that tend to be targets sites around 22 nt (Lambert et al. 2016). In conclusion, this study presents 3 new members of the I-ONuI family of LHEs, and although recently alternative genome editing reagents have become very popular, the low off-target activities make LHEs an attractive set of enzymes that warrant further exploration (Cox et al. 2015; Lambert et al. 2016). Engineering modular meganucleases by combining the LHE DNA cutting domains with more programmable DNA binding domains might be a promising direction for utilizing LHEs that belong to the I-ONuI family (Hafez and Hausner 2012; Roisell et al. 2014; Wolf et al. 2014, 2016; Romano Ibarra et al. 2016; Guha et al. 2017).

Acknowledgements

This work is supported by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) to G.H., I.M.B., and T.K.G. Funding support from the Faculty of Science Graduate Award program (University of Manitoba) and the University of Manitoba Faculty of Graduate Studies GETS program. I.M.B. acknowledges the Islamic Development Bank Merit Scholarship Programme that supported part of her graduate program. The authors thank the efforts of an anonymous reviewer.

References


Mullineux, S.T., Costa, M., Bassi, G.S., Michel, F., and Hausner, G. 2010. A group II intron encodes a functional LAGLIDADG homeing endonuclease and self-splices under...