A homing endonuclease with a switch: Characterization of a twintron encoded homing endonuclease

Tuhin Kumar Guha, Georg Hausner *
Department of Microbiology, University of Manitoba, Winnipeg, MB, Canada

1. Introduction

Homing endonucleases (HEases) are encoded by homing endonuclease genes (HEGs) which can be embedded within composite mobile elements such as group I introns and group II introns, archael introns, as well as inteins (Dujon, 1989; Belcour et al., 1997; Gimble, 2000; Tocchini-Valentini et al., 2011). HEGs can be components of mobile elements that mimic introns but can in many cases with the help of host factors splice from the primary transcripts or in the case of inteins from the protein precursor (Belmany et al., 1997; Gimble, 2000; Tocchini-Valentini et al., 2011). HEGs can be factors that have not yet been invaded by this mobile composite element.

HEases promote the mobility of the HEG or the mobility of composite elements that encodes HEases by cleaving a target sequence in cognate alleles that lack HEGs or intron/intein insertions (Dujon and Belcour, 1989). HEase typically generates double-stranded breaks that are repaired by the host double-strand break repair pathway involving homologous recombination whereby the intron-containing allele is used as a template to repair the “cut” intron-less allele (Belfort et al., 2002). As HEases require long DNA recognition sites they cut infrequently within a genome and therefore are useful for DNA engineering (Stoddard, 2011; Prieto et al., 2012). These elements are also important as HEGs and mobile introns are invasive elements and therefore contribute towards the size of fungal mtDNA genomes, generate mtDNA polymorphisms, and in some instances they are associated with mtDNA instabilities in some fungi (Dujon and Belcour, 1989; Abu-Amero et al., 1995; Baidyaroy et al., 2011; Hafez et al., 2013).

Amlacher et al. (2011) determined the nuclear and mitochondrial genomes (mtDNA) for the thermophilic fungus Chaetomium thermophilum var. thermophilum La Touche (strain DSM 1495) which can tolerate temperatures up to 60°C; although optimal growth is around 45°C (Microfungus Collection and Herbarium, http://www.straininfo.net/strains/418186/browser). Examination of the mtDNA sequence for this fungus showed that the small ribosomal subunit (rns) gene contained a novel twintron at position mS1247 (Hafez et al., 2013). This fungus has been reported from many locations and has been isolated from soil and compost heaps; besides its ecological importance as a cellulytic fungus it is viewed as a source of thermostable enzymes (Wang et al., 2012; Busk and Lange, 2013).
Characterization of the rns twintron in two strains of *C. thermophilum* var. *thermophilum* ([strains DSM 1495 and UAMH 2024 (=CBS 141.64); Hafez et al. (2013)] indicated that this twintron is composed of an external group I intron that has been invaded by a group II intron (Fig. 1). The internal group II intron inserted within the external group I intron LAGLIDADG open reading frame (ORF). This posits a unique possibility whereby splicing of the internal group II intron would allow the ORF to be reconstituted (ORF). This posits a unique possibility whereby splicing of the encoded HEase. This might offer a window for engineering a HEase that based on the splicing competency of an intron might have an in vivo switch for regulating HEase activity/expression.

Recently the value of native HEases towards engineering site specific genome editing tools has been demonstrated by Takeuchi et al. (2011) on work based on HEGs inserted within the mtDNA of *thermovibrio maritimus* and *Pseudomonas aeruginosa*. In order to confirm the elimination of the DNA template, the forward primer ‘A’ (5’ TAGGGACTATG CATTGCC 3’) and the reverse primer ‘B’ (5’ TCTCTACAGCTGAGCC 3’) were applied to perform a standard PCR reaction as described above using 2 μL from the in vitro transcription reaction mixture as the template (see Fig. 2A).

To evaluate if splicing occurred cDNA was generated from the in vitro transcribed RNA by reverse transcriptase (RT) PCR. For first strand synthesis a 20-μL reaction mix was prepared containing 1 μg RNA, 0.5 μM of the reverse primer (mtr-2), 0.1 M DTT, 4 μL of 5× cDNA synthesis buffer, 1 μL of each dNTP, 40 units RNase-OUT (Invitrogen) and 15 units of Thermoscript reverse transcriptase (Invitrogen). Reverse transcription was performed at 55 °C for 1 h and stopped by heating the reaction mixture to 85 °C for 10 min. Finally, 1 μL of RNase H (2 units) was added to the reaction mixture followed by incubation at 37 °C for 20 min.

In order to characterize the twintron derived transcripts several primers were designed (see Fig. 2A) in order to recover potential splicing intermediates. The forward primer ‘A’ and the reverse primer ‘B’ are based on the external intron’s upstream and downstream sequences with regards to the group II intron location. Primer ‘C’ (5’ ACAGCATGCGAGAAAAGG 3’) and primer ‘D’ (5’ GTTACACATCTGCGAAGC 3’) were designed to amplify a 400 bp segment of the internal group II intron and these evaluated by submarine agarose gel electrophoresis on a 1% gel in 1× TBE [89 mM Tris base, 89 mM boric acid, 2 mM EDTA (pH 8)] buffer. The F1-T7/mtr-2 PCR product was used as the template for the in vitro transcription reaction. A 20-μL in vitro transcription reaction mixture contained the following ingredients: 0.1 μg DNA template, 0.5 mM of each NTP, 2 μL of T7 transcription buffer (Ambion/Life Technologies, Burlington, Ontario) [20 mM NaCl, 40 mM Tris–HCl (pH 7.8), 6 mM MgCl2, 2 mM spermidine, 10 mM dTT], and 20 units of T7 RNA Polymerase-Plus Enzyme Mix (Ambion). The reaction mixture was incubated for 2 h at 37 °C. For splicing the addition of NaCl and MgCl2 to a final concentration of 1.2 M NaCl and 60 mM respectively to the transcription reaction buffer was necessary in order for efficient splicing of the group II intron. Template DNA was removed by the addition of 2 units of DNasel (Fermentas, Ottawa, Canada) and incubation at 37 °C for 15 min; the reaction was stopped by adding 1 μL EDTA (50 mM) followed by a 10 min incubation at 65 °C. In order to confirm the elimination of the DNA template, the forward primer ‘A’ (5’ TAGGGACTATG CATTGCC 3’) and the reverse primer ‘B’ (5’ TCTCTACAGCTGAGCC 3’) were applied to perform a standard PCR reaction as described above using 2 μL from the in vitro transcription reaction mixture as the template (see Fig. 2A).

To demonstrate the splicing competency of the group II intron component of the S1247 twintron a splicing assay involving in vitro transcription was performed following the protocol of Salman et al. (2012). A segment of the rns gene containing the S1247 twintron (see Fig. 2A) was amplified by PCR using a forward primer containing a T7-promoter sequence (F1-T7, 5’GAATTCGACAGCTACATAGGAACTATCAAACTCCGGGG 3’ and the mtr-2 reverse primer (5’ CGAGTGTTAGTACCAATCC 3’) were designed to amplify the sequence site within the rns gene and provide some characterization of the HEase protein and demonstrate by an in vitro splicing assay that the group II intron indeed splices in a manner that allows for the interrupted HEase ORF to be reconstituted during RNA processing.

2. Material and methods

2.1. In vitro RNA splicing assay

![Fig. 1](https://example.com/fig1.png)

**Fig. 1.** A schematic representation of the mtDNA rns gene (black line) of *C. thermophilum* strains DSM 1495 and UAMH 2024 (Hafez et al., 2013) and the twintron at S1247. At the S1247 position a group IC2 intron (orange line) that encodes a double motif LAGLIDADG type HEase (green line) is interrupted by an ORF-less group II A1 intron (blue line). The arrows represent the position of primers (not to scale) within the twintron utilized in this study. A segment of the HEase ORF sequence is shown in black letters to illustrate the location of the group II intron insertion (represented by a single dot). The corresponding amino acid sequence is provided below the nucleotide sequence. The position of the group II intron is referred to as a phase 0 intron, as its positions does not disrupt a codon. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
primers were designed to provide a positive control. To amplify potential splicing intermediates 2 \mu L of cDNA (0.2 \mu g) was added to a 50-\mu L PCR reaction mixture containing 1 mM of each product-specific forward and reverse primer (see above), 0.2 mM of each dNTP, 5 \mu L of 10x OneTaq PCR buffer, and 1 unit of OneTaq DNA polymerase (New England Biolab, Pickering, Ontario). The PCR program was as follows: initial denaturation at 94 °C for 2 min, followed by 30 cycles of 94 °C for 30 s, 55 °C for 30 s, 68 °C for 2 min, followed by final extension time of 10 min at 68 °C. PCR products were resolved on a 1% agarose gel. The amplicons obtained were excised from the agarose gel and purified using the Wizard\textsuperscript{®} SV Gel and PCR Clean-Up system (Promega, Madison, Wisconsin). The gel extracted DNA fragments were sent to the DNA Technologies Unit, NCR, Saskatoon, Saskatchewan for cycle sequencing, utilizing the primers used for obtaining the amplicons.
2.2. Construction of Escherichia coli expression vector for the I-CthI HEase

Based on sequences obtained by Amlacher et al. (2011) and Hafez et al. (2013) (Genbank accessions: JN007486 and JX139037 respectively) the mS1247 twintron ORF was reconstructed by removing the group II intron sequence from the external intron’s ORF sequence. The genetic code for the HE ORF was optimized for expression in E. coli and the HEase sequence was synthesized by GenScript (New Jersey, USA). The synthesized ORF was inserted into the pET28b+ plasmid as a BamHI/Ndel fragment with an N-terminal 6x HIS-tag; this construct was named I-CthI-pET28b+. The construct was transformed into E. coli BL21 (λDE3) for protein expression, purification and biochemical studies (Fig. 3A).

Ten mL of Luria Broth (LB; peptone 10 g/L, yeast extract 5 g/L and NaCl 5 g/L; pH 7.0) media supplemented with 100 µg/mL kanamycin and 0.25% w/v glucose was inoculated with 100 µL of E. coli BL21 (λDE3) transformed with I-CthI-pET28b+ and incubated overnight (ON) with agitation at 37 °C. Five mL from the ON culture was used to inoculate 1 L of LB medium supplemented with kanamycin and 25% w/v glucose (as described above). The culture was grown at 37 °C with agitation till the OD600 reached ~0.65 and expression of the protein was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The culture was then shifted to 28 °C and incubated with agitation overnight (~16 h). Cells were harvested by centrifugation at 5000 rpm for 10 min and the resulting pellet was frozen at ~80 °C.

The frozen pellet was thawed on ice and resuspended in lysis buffer [50 mM Tris–HCl (pH 8.0), 100 mM NaCl, 10% (w/v) glycerol, 6 mM β-mercaptoethanol] at a ratio of 5 mL of buffer to 1 g of cells (wet weight). Cells were homogenized twice using the French press and the resulting lysate was centrifuged at 12,000 rpm for 20 min at 4 °C to pellet cell debris. The clear cell lysate (about 8 mL) was added to 3 mL of Ni-NTA resin (Qiagen, Toronto, Ontario) and incubated at 4 °C with agitation for 20 min. The sample along with the Ni-NTA slurry was loaded onto a Ni-NTA super flow column (Qiagen) and the washing steps were as listed: wash 1: 30 mL of washing buffer (WB) [50 mM Tris–HCl (pH 8.0), 100 mM NaCl, 10% (w/v) glycerol, 6 mM β-mercaptoethanol]; wash 2: 30 mL of WB buffer with 25 mM of imidazole; and wash 3: 30 mL of WB buffer with 50 mM of imidazole. The protein was eluted in Elution buffer (EB) [WB supplemented with 250 mM imidazole] and collected in 10 fractions (1 mL each). After confirming the presence of pure I-CthI protein among these 10 fractions by denaturing SDS polyacrylamide gel (12.5%) electrophoresis, the fractions showing the desired protein were pooled. Excess imidazole was removed by dialysis using the dialysis buffer (50 mM Tris–HCl (pH 8.0), 100 mM NaCl, 6 mM β-mercaptoethanol) and the slide-a-lyzer dialysis cassette with a 10 kDa molecular weight cut-off (MWCO) (Millipore, Billerica, USA) according to the manufacturer’s protocol. The dialysed sample was concentrated by the Amicon concentrator (model 8050) using a Millipore YM-10 membrane to a final volume of 1 mL. Ultimately, the pure protein was mixed with 200 µL of protein storage buffer [50 mM Tris–HCl (pH 8.0), 400 mM NaCl, 0.5 mM DTT, 10% (w/v) glycerol] and kept at ~80 °C.

2.3. Endonuclease assay

The putative HEase was evaluated for activity by performing endonuclease assays. A substrate plasmid containing the HEase recognition site was designed as follows. A segment of the C. thermophilum rns sequence (JN007486) was constructed that is composed of sequences flanking the mS1247 twintron (321 nucleotide upstream and 148 nucleotide downstream of the twintron) based on predictions of Hafez et al. (2013). This 469 bp sequence was synthesized (GenScript, New Jersey, USA) and inserted into the EcoRV site within the pUC57 vector (2.7 kb). The substrate plasmid was named Cth-rns.pUC57 and its size is 3.1 kb (see Fig. 4A).

To serve as the negative control, a C. thermophilum rns segment that contains the twintron plus the flanking exon sequences was PCR amplified (as outlined previously) using the rns-F1 forward (5’ CGTGCAGCGACTGCCGG 3’) and mtsr-2 reverse primers (Hafez...
The substrate and control plasmids were transformed into E. coli DH5α and the plasmids were purified with the Wizard® Plus Miniprep DNA purification kit (Promega) according to the manufacturer’s protocol. The endonuclease assay reaction mixtures contained: 25 μg/mL substrate or the negative control plasmid or the substrate PCR product, 5 μL In Vitrogen Buffer React #3 (100 mM NaCl, 50 mM Tris–HCl, pH 7.9 and 10 mM MgCl2) supplemented with 1 mM DTT, 5 μL I-CthI (9 μg) and H2O to achieve a final volume of 50 μL. Cleavage reactions were incubated at 37 °C and 10 μL aliquots were withdrawn at 0, 30, 60, 90 and 120 min. These aliquots were treated with the addition of 2 μL of 200 mM EDTA (pH 8.0) and 1 μL of proteinase K (1 mg/mL) in order to stop the endonuclease reactions; the aliquots were incubated for 30 min at 37 °C. The products of these assays were resolved on a 1% agarose gel.

2.4. I-CthI cleavage site mapping

The Cth-rns.pUC57 construct was incubated with I-CthI and the linearized substrate plasmid was subsequently treated with T4 DNA polymerase (Invitrogen), which due to its ability to hydrolyze 3’-overhangs generates blunt DNA ends (Bae et al., 2009). Forty μL of linearized plasmid (25 μg/mL) was treated with 10 units of T4 DNA polymerase (Invitrogen; 5 u/μL) in a reaction volume of 100 μL that included 20 μL of 5x T4 DNA polymerase buffer, 20 μL dNTP mixture (0.5 mM) and H2O to achieve the final volume of 100 μL. The reaction was incubated at room temperature (~24 °C) for 20 min and then placed on ice for 5 min and finally terminated by heating the reaction mixture at 70 °C for 10 min. The DNA was recovered from the reaction mixture with the Wizard® SV Gel and PCR Clean-Up system (Promega). The linearized and now blunt-ended plasmid DNA was religated. The 50 μL ligation mixture contained the following ingredients, 20 μL (0.25 μg) of T4 DNA polymerase treated cleaved plasmid, 2 μL of T4 DNA Ligate (Invitrogen) (1 u/μL), 10 μL of 5x Ligase buffer, and H2O to achieve the final volume 50 μL. The reaction was incubated at room temperature for 2 h and thereafter the ligation mixture was diluted 5-fold and 10 μL of this dilution was used to transform E. coli (DH5α). Potential transformants were plated on LB (plus ampicillin) agar plates and single colonies were used to inoculate 5 mL LB (plus ampicillin) broth cultures. Plasmids were recovered from the overnight cultures with the Wizard® Plus Miniprep DNA purification kit (Promega) and the plasmid DNAs were sent to the DNA Technologies Unit, NRC, Saskatoon, Saskatchewan for cycle sequencing (using the M13F forward and M13R reverse primers for both treated and untreated substrate). The chromatograms from these sequencing reactions were aligned manually and compared with the GeneDoc program (version 2.7; Nicholas et al., 1997); in particular the HEase and T4 DNA polymerase treated substrate plasmid sequence with the untreated substrate plasmid sequence. Nucleotides missing in the sequence of the substrate treated with I-CthI when compared to the original untreated substrate plasmid derived sequence allowed for the determination of the HEase cleavage site.

2.5. Temperature profile and thermal stability of the I-CthI protein

To test the effect of temperature on I-CthI cleavage activity, the endonuclease assay was performed as described in Section 2.3, but the reactions were incubated at a temperature range from 25°C to 85°C at 10 degree intervals for 1 h. The reactions were terminated as described previously and the products of these assays were resolved on a 1% agarose gel.

Temperature stability was also examined by circular dichroism spectropolarimetry. The CD spectra were acquired on a JASCO J-810 spectropolarimeter-fluorimeter calibrated with (+)-10-

---

**Fig. 4.** (A) Schematic overview of the in vitro endonuclease assay. An intron-less section (400 bp) of the rns gene was constructed and cloned in pUC57 (2.7 kb) to serve as the substrate plasmid (3.1 kb). A second construct with the target site interrupted by the intron (at mS1247) was used as the negative control (see Section 2). (B) In vitro endonuclease cleavage assay with the C. thermophilum twintron encoded HEase; a 1% agarose gel showing the results of the endonuclease assay. Lane C = uncut control substrate (in pUC57); lane L = linearized substrate plasmid (cleaved with BamHI); numbers on top of each lane represent incubation times in minutes at 37 °C. In all instances 1 μg plasmid was treated with the 5 μL aliquot of HEase (9 μg) and an arrow shows the linearized substrates. Similarly, the 650 bp PCR substrate was cleaved into two fragments (sear arrows) when incubated with the same concentration of the HEase. (C) In vitro endonuclease cleavage assay with the C. thermophilum twintron encoded HEase when challenged with the negative control plasmid. In lane Cn (=control/non-substrate plasmid) no HEase was added and in the lane ′+HE′ the non-substrate plasmid was incubated with the same concentration of HEase as above for one hour at 37 °C. No cutting was observed and only high molecular weight supercoiled and concatenated plasmid DNAs was observed. For gels depicted in (B) and (C) the lane denoted as ‘M’ shows the 1 kb plus DNA ladder (Invitrogen).

et al., 2013). The resulting PCR product (2.7 kb) was resolved on a 1% agarose gel and extracted from the gel using the Wizard® SV Gel and PCR Clean-Up system (Promega). The purified 2.7 kb DNA fragment was cloned into the pCR4 TOPO vector using the TOPO TA Cloning Kit ingredients and protocols (Invitrogen).

Finally the HEase was challenged with a PCR product that contained the putative I-CthI target site. This template was generated to rule out the possibility that the putative HEase cuts in the pUC57 vector sequence, the DNA sequence containing only the substrate region was PCR amplified from the Cth-rns.pUC57 (i.e. substrate plasmid) using M13F forward (5′ GTTAAGACGACGCCCAG 3′) and M13R reverse (5′ CAGGAAAACAGTATGAC 3′) primers using the following conditions: initial denaturation 94 °C for 1 min, followed by 30 cycles of 94 °C for 30 s, 52 °C for 30 s, 68 °C for 30 s, and a final elongation time of 5 min at 68 °C.

---

**Reference:**

Bae, T.K. Guha, G. Hausner / Fungal Genetics and Biology 65 (2014) 57–68
camphorsulfonic acid and purged with N₂ at 20 L/min. CD spectra of the protein samples were measured in the far UV region (180–250 nm) using 0.05–0.10 cm path length quartz cuvettes at the initial setting at 35 °C, a scan rate of 10 nm/min, and a response time of 0.8 s. CD spectra were corrected by baseline subtraction and were converted to mean residue ellipticity (MRE) according to the following formula: \[ \theta_{200} = \frac{M \cdot \theta}{c \cdot l} \] where \( \theta_{200} \) are 10° deg cm²/dmol, \( M \) is the molecular mass of the His₆-HE (34.8 kDa or 1.2 mg/mL), \( \theta \) is the measured ellipticity in millidegrees, \( l \) is the path length of the cuvette in cm (0.1 cm), \( c \) is the protein concentration in g/L, and \( n \) is the number of amino acid residues in the protein (305). Temperature was controlled during thermal denaturation experiments of the His₆-HEase sample using the Peltier device connected to the spectropolarimeter. Spectra were recorded for the following temperature range: from 25 °C to 85 °C at 10 degree intervals for 30 min at each tested temperature.

The structure of the twintron encoded HEase was predicted with the online Protein Homology/analogY Recognition Engine V2.0 (PHYRE2) program [http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index; Kelley and Sternberg, 2009].

2.6. Phylogenetic analysis of the twintron encoded ORF and related LAGLIDADG HEs

The online resource BLASTp (Altschul et al., 1997) was used to retrieve sequences that were related to the mS1247 twintron ORF. The LAGLIDADG ORF amino acid sequences were aligned with the online PRALINE multiple sequence alignment program (Simossis and Heringa, 2005) and the alignment was further refined with the GeneDoc program.

For phylogenetic analyses, only those segments of the alignment where all sequences could be aligned unambiguously were retained. Phylogenetic estimates were generated by the programs contained within the Molecular Evolutionary Genetic Analysis program package (MEGA 5.2; Tamura et al., 2011) and the MrBayes program v3.1 (Ronquist and Huelsenbeck, 2003). With MEGA phylogenetic trees were generated with the Maximum parsimony (MP), Neighbor joining (NJ), and Maximum likelihood (ML) methods. MEGA 5.2 was also used for determining the best fit substitution model for ML analysis; thus for ML analysis the WAG + G + F model was applied and for all programs the bootstrap option was selected (1000 replicates) in order to obtain estimates for the confidence levels for the major nodes present within the phylogenetic trees (Felsenstein, 1985).

The MrBayes program was used for Bayesian analysis and the parameters for amino acid alignments were the mixed model setting. The Bayesian inference of phylogenies was initiated from a random starting tree and four chains were run simultaneously for 2,000,000 generations; trees were sampled every 100th generation. The first 40% of trees generated were discarded (“burn-in”) and the remaining trees were used to compute the posterior probability values and majority rule consensus tree. Phylogenetic trees were drawn with the TreeView program (Page, 1996) using the MrBayes tree files, and the phylogenetic tree was annotated with Corel Draw™ (Corel Corporation Ltd., Ottawa, Canada).

3. Results

3.1. In vitro splicing of the internal group II intron reconstitutes the LAGLIDADG ORF

Initial attempts using RT PCR with whole cell RNA as a template to recover precursor rns transcripts failed and we could only recover mature rns transcripts (as previously characterized in Hafez et al., 2013; also see Fig. 1). In order to demonstrate that the internal group II intron could splice we set up in vitro transcription assays based on a segment of the rns gene that contained the twintron region. These assays allowed for the generation of splicing competent group II intron RNAs, however only under high salt concentrations (1.2 M NaCl) could we recover splicing intermediates along with processed RNAs (Fig. 2B, C, D). The splicing reaction products were analyzed by RT-PCR utilizing primers A and B (based on external intron sequences that flank the internal group II intron). Among the observed cDNAs was a dominant band at 1.1 kb (Fig. 2D), the expected size for cDNAs from transcripts were the group II was spliced out. Along with the 1.1 kb fragment we observed what appears to be cDNAs generated from unspliced versions (1.9 kb) and various shorter fragments. Those were not further investigated. As a positive control cDNAs were generated based on the internal group II with primers C and D generating a 400 bp fragment (see Fig. 2C and D); serving as a negative control was a sample without the RT step to ensure all DNA was removed from the in vitro transcription assay (Fig. 2B). The 1.1 kb cDNA PCR product was excised from the agarose gel and submitted for DNA sequence analysis and the resulting data were compared with the genomic version of the mS1247 twintron. Comparative sequence analysis within the GeneDoc program showed that the 1.1 kb cDNA was the result of the group II intron being spliced out and the joining of the flanking external intron sequences. Also as predicted previously by in silico analysis the “internal intron”/external intron splice junction (Hafez et al., 2013) corresponds to a phase 0 position with regards to the coding region of the group I intron ORF and the splicing of the group II intron allows the group I ORF located within the P 9.1 loop to be reconstituted into a continuous reading frame potentially encoding a functional HEase.

3.2. Overexpression and purification of the twintron encoded homing endonuclease

The codon optimized HEase ORF was overexpressed in E. coli, however optimization with regards to IPTG concentration and temperature during induction was required. The transformed E. coli culture was grown overnight and the recombinant HEase protein expressed at all temperatures tested (28 °C and 37 °C) including RT (~22 °C) except at 16 °C. Overall induction with 0.5 mM IPTG followed by incubation at 28 °C was determined to be the best expression conditions. Upon harvesting and lysing the cells the sized protein migrated on an SDS–PAGE at around 29 kDa close to the predicted size of the twintron encoded HEase 32 kDa (Fig. 3B). The purification of the protein was achieved by affinity chromatography involving Ni-NTA Superflow resin. A step up gradient with buffers containing 25 mM and 50 mM imidazole was used to remove the background proteins while a buffer containing 250 mM imidazole was used in protein elution; the purification of the protein was monitored by SDS–PAGE (Fig. 3C). The desired fractions were pooled, dialyzed and the protein concentrated to a final concentration of 2.2 mg/mL. The HEase protein was purified in sufficient concentrations to pursue endonuclease and cleavage assays plus to conduct some investigations with regards to the thermal stability of the twintron encoded protein.

3.3. The mS1247 twintron encoded I-CthI is an active endonuclease

In vitro endonuclease assays were performed by incubating the purified I-CthI with circular and linearized versions of the substrate plasmid that contains the putative rns target site (Fig. 4A). The endonuclease activity of the enzyme was tested at five different time points (O, 30, 60, 90 and 120 min) and at 60 min I-CthI completely linearized the circular substrate plasmid (3.1 kb). Cleavage activity was already observed at time “0” time
point, however time “0” means that the HEase was added and thereafter the stop buffer plus protease K was added. Therefore there was a short time period that allowed the enzyme to digest the substrate. Moreover, when a PCR derived DNA fragment (700 bp) containing the rns target site was used as the substrate, the enzyme cleave the PCR product yielding two fragments (400 and 300 bp, Fig. 4B). So both circular (supercoiled) and linear substrate molecules are cleaved by I-CthI. However, I-CthI did not cleave the non-substrate plasmid even after one hour of incubation at 37 °C (Fig. 4C).

LAGLIDADG HEases tend to generate cohesive termini by generating staggered cuts with 3’ 4 nucleotide single stranded overhangs. T4 DNA polymerase can hydrolyze 3’ overhangs and thus blunt the cleaved DNA fragment and therefore one can indirectly characterize 3’ overhangs by comparative sequence analysis with uncut substrate DNA molecules (Bae et al., 2009). The T4 DNA polymerase treated and religated I-CthI-cleaved substrate plasmid sequence when compared with the sequence of the uncut substrate plasmid showed that a 5’ AAGA 3’ segment was removed from the sense strand (Fig. 5A). So the cleavage site mapping experiment showed that I-CthI cleaves 8 bp downstream of the twintron insertion site (sense strand) or 4 bp downstream of position S1247 at the antisense strand (Fig. 5B).

3.4. The effect of temperature on I-CthI endonuclease activity and stability

I-CthI was challenged with the substrate plasmid at a temperature range starting from 25 °C in 10 °C increments until 85 °C (Fig. 6A). The enzyme showed activity at 25, 35 and 45 °C; albeit

---

**Fig. 5.** (A) Cleavage site mapping for the *C. thermophilum* twintron encoded HEase. The cleavage site was mapped by comparing uncut with I-CthI treated substrate DNAs. Cleavage by I-CthI generates a staggered cut with 4 nucleotide 3’ overhang in the substrate plasmid at the enzymes target site. The cleaved ends were blunted using T4 DNA polymerase. The religated plasmid was sequenced and compared to the sequence of the untreated substrate plasmid in order to map the cleavage site by scanning for a 4 bp deletion in the T4 DNA polymerase treated cleaved substrate plasmid. (B) Schematic representation of the I-CthI cleavage site near the twintron insertion sequence. Proposed cleavage sites are indicated by open triangles; and a vertical line represents the twintron insertion site. The HEase cleavage site is 8 nt downstream of the twintron insertion site with regards to the sense strand or 4 nt downstream with regards to the antisense strand.
at 25 °C and 45 °C there appeared to be evidence for the presence of considerable amount of uncut substrate. No visible cleavage activity was noted at 55 °C and above (Fig. 6A).

The stability of the I-CthI protein was evaluated by circular dichroism (CD) spectropolarimetry. Circular dichroism allows for the prediction of protein secondary structural features, by splitting plane polarized light into its left and right components and by monitoring the differences in absorbance between the two components (Kelly et al., 2005). When the sample is subjected to different temperatures during the CD analysis one can evaluate the thermal stability of the protein by monitoring changes in its secondary structure. Far-UV CD spectra’s of the I-CthI protein over a temperature range from 25 to 85 °C (Fig. 6B) showed a shift of the bands at 222 nm and 208 nm towards less negative values, reflecting the reduced fraction of α-helical segments (Greenfield, 2006). In addition we noted that there was a shift of bands at 210 nm towards the negative direction possible showing an increase in disorder within the protein (Kelly et al., 2005; Greenfield, 2006). These changes were particular noticeable at 55 °C and above. This loss of α-helical segments correlates with the data from the endonuclease assays that showed a reduction of endonuclease activity at 45 °C and a lack of detectable cutting activity at or above 55 °C.

The online program PHYRE2 was used to examine the secondary and tertiary structure of I-CthI (Fig. 6C). The program showed that the protein is comprised of ten alpha helices (39%) and nine beta strands (27%), the confidence key of these regions was found to be 100% when compared with the crystal structure of another homing endonuclease I-Onul12 (PDB accession number: 3pqyA). There were also a few disordered regions in the proteins tertiary structure starting from amino acid 134 to 164 and the overall disorderness was estimated to be 22%.

3.5. Phylogenetic position of the I-CthI HEase

The programs utilized to infer phylogenetic relationships for the aligned LAGLIDADG data set yielded similar tree topologies and in all cases the trees received moderate node support values although deeper notes received poor support values (Fig. 7). However, strong support was noted for the node that unites the I-CthI twintron ORF with other ms1247 group I intron encoded ORFs suggesting that most likely these introns share a common ancestor and the group II intron was probably inserted more recently within the C. thermophilum ancestral ms1247 intron. Also Bayesian, ML and NJ analysis provided significant support for suggesting that the ms1247 intron ORFs share a common ancestry to group I intron ORFs located within protein coding genes such as nad4L and cox1. It is also worthwhile to note that within this data set only C. thermophilum can be classified as a thermophile.

4. Discussion

4.1. The twintron encoded split ORF encodes an active homing endonuclease

Characterization of the ms1247 twintron encoded reconstituted ORF showed that it encodes a functional HEase that cuts downstream of the twintron insertion site. The results also showed that the twintron encoded ORF, which is interrupted by a group II intron, is functional when the internal group II intron is removed. The twintron encoded HEase was able to cleave both plasmid and linear (PCR product) substrates that contained the rns target

Fig. 6. Effect of temperature on I-CthI endonuclease activity. (A) 1% agarose gel showing the effect of temperature on the in vitro endonuclease activity when the reactions were incubated at a temperate range from 25 °C to 85 °C in 10 degree intervals for 1 h. The lane marked M contains the 1 Kb plus DNA ladder (Invitrogen) and the lane marked C contains a control where no HEase was added to the substrate plasmid. The enzyme appears to cut most efficiently at 35 °C while the cleavage activity diminishes as the temperature rises (linearized substrate shown by arrow). (B) Temperature stability for I-ChI examined by circular dichroism spectropolarimetry. Spectra were recorded for the following temperate range: from 25 °C to 85 °C in 10 degree intervals for 30 min and the corresponding CD spectra were plotted. The values recorded between 208 nm and 222 nm slowly move towards the 0 base line as the temperature increases. This region of the spectra is indicative for alpha helices and a shift towards less negative values corresponds to the loss of the structural integrity in those alpha helices. Also note that the decrease in the spectras at 195 nm over the tested temperature range is indicative of an increase in structural disorder, i.e. accumulation of random coils in the proteins secondary structure. (C) An in silico model for the I-ChI protein generated by the PHYRE2 program. The program identified the double motif LAGLIDADG I-Onul (PDB: 3pqyA) HEase protein as a template for folding I-ChI. The model shows the symmetrical nature of this protein and the alpha helices and beta sheets along with amino terminal (N) and carboxyl terminal (C) have been marked. The beta sheets arrange in a configuration that forms the DNA binding surface of the HEase and the LAGLIDADG motifs contribute toward the active site of these enzymes (Stoddard, 2006).
site. So this HEase has the potential of providing this composite element with the ability to insert into cognate alleles that lack an insertion.

We also showed that the internal group II intron has the potential to self-splice under laboratory conditions in the presence of high salt concentration. This is commonly observed in group II in-
tron in vitro splicing assays (Toor et al., 2006; Mullineux et al., 2010; Fedorova, 2012) and one typically assumes that under cellular conditions either intron encoded or host factors or both facilitate the formation of splicing competent group II intron RNA ribozyme configurations (Bonen and Vogel, 2001; Lambowitz and Zimmerly, 2011; Hausner, 2012).

Nested introns or twintrons, combinations of self-splicing introns embedded within another potentially self-splicing intron have been noted in protists rDNA, fungal mtDNAs and algal chloroplast genomes (Drager and Hallick, 1993; Einvik et al., 1998a; Hafez et al., 2013). Analogous spliceosomal version of nested introns/ twintrons also exits in Fungi and Metazoan nuclear genomes (Filippi et al., 2013; Janice et al., 2013). These nested elements may offer insights on the evolution of complex nuclear introns (Janice et al., 2013; Suzuki et al., 2013) or in the case of group I and group II introns the evolution of composite organellar mobile elements and novel ribozymes (Einvik et al., 1998b; Khan and Archibald, 2008; Moreira et al., 2012; Pombert et al., 2012). A similar situation to the mS1247 twintron has been noted in the rnl-rps3 (ribosomal protein 3) locus in Grommannia piceiperda, here a twintron (mL2449) has an internal group I intron interrupting the external group I intron’s rps3 open reading frame, but upon splicing of the internal intron the rps3 ORF is reconstituted thus allowing for the potential expression of the RPS3 protein (Rudski and Hausner, 2012). Obviously the evolution of nested elements requires compatibility and co-evolution among its various constituents in order to allow for efficient splicing and expression of protein products; raising questions with regards to their mode of evolution as a simple neutral evolution model (Goddard and Burt, 1999; Cogarten and Hilario, 2006) may not explain their persistence within a population. Such complex multicomponent associations would be eliminated almost immediately if mutations could accumulate within a population. Such complex multicomponent associations would be eliminated almost immediately if mutations could accumulate within a population. Therefore the production of the HEase is important but one would expect that excess production would be a drain on the host system.

With regards to the mS1247 internal group II intron this element might benefit from this association by gaining a neutral location thus minimizing impact to the host genome plus the ORF less group II intron still has the ability to be mobilized as part of the twintron unit by the DNA based mobility mechanism that drives group I intron homing. Although this needs experimental investigations one can speculate that the splicing of the group II introns may be a regulatory step that can control the amount of HEase protein that can be translated from the processed group I intron RNA. The survival of these elements depends on their ability to spread into cognate (or new) sites that lack introns and by minimizing their impact on the host genome (Gillham, 1994; Goddard and Burt, 1999; Edgell et al., 2011). Therefore the production of the HEase is important but one would expect that excess production would be a drain on the host system.

4.2. Origin of the twintron

The origin of the group II intron at this stage cannot be determined; blastn searches did not reveal the presence of similar introns within the NCBI data base. Various scenarios can be envisioned as to the origin of the internal intron of the mS1247 twintron. It has been demonstrated that group II introns that lack ORFs may not have lost their ability for homing or ectopic integration into new sites. For example Moran et al. (1995) noted that in yeast a mtDNA group II intron lacking reverse transcriptase activity could move based on a DNA-level recombination mechanism. It is also possible that the mS1247 internal intron was mobilized by retrotransposition facilitated by a trans-acting factor provided by a related reverse transcriptase-encoding group II intron. With regards to invading ectopic sites two mechanisms have been proposed, one involving reverse splicing of the intron RNA into an ectopic site within another RNA thereby reverse splicing into an ectopic site within tRNA (Zimmerly et al., 1995a). Reverse splicing within RNA would require two additional steps, reverse transcription into cDNA and the integration of the cDNA into the host genome by recombination (Mueller et al., 1993; Zimmerly et al., 1995a,b; Bonen and Vogel, 2001; Cousineau et al., 2000). So far strong experimental evidence for reverse splicing into RNA is still lacking but evidence for ectopic integration of group II intron RNA into DNA sites that resemble the intron’s native homing sites (IBS1 and IBS2) has been demonstrated (Yang et al., 1998, 1998; Dickson et al., 2001).

Assuming that the internal group II intron is a recent addition to the mS1247 intron ORF locus sequence comparisons between the three currently available examples may hint at the possibility of ectopic integration (Fig. 8). The sequence upstream of the site where the group II intron has inserted within the C. thermophilum twintron shows some conservation. So this sequence among the mS1247 intron ORFs may indeed have had sufficient resemblance in C. thermophilum mS1247 intron to the native homing site of the internal group II intron. This would have allowed for the group II intron EBS1 and EBS2 sequences to interact with ORF based IBS1 and IBS2 sequences and facilitate the insertion of the intron by either reverse splicing at the RNA level or, by means of a trans-acting reverse transcriptase, reverse splicing at the DNA level (Fig. 8).

4.3. A homing endonuclease with a possible “on” switch

Currently several gene targeting endonucleases are being developed based on scaffolds that include group II intron based targetrons, zinc-fingers, TALENs, the CRISPR/cas9 system and HEases (Karberg et al., 2001; Stoddard, 2011; Hafez and Hausner, 2012; Gaj et al., 2013; Ran et al. 2013; Marton et al., 2013). The advantage of native HEases is their high degree of specificity and that they can be uncovered by exploring mobile introns, i.e. they do not need to be synthesized de novo. In general HEases require long DNA recognition sites and therefore cut infrequently within a genome; this makes them useful for DNA engineering (i.e. genome editing) (Gimble, 2005, 2007). HEases are currently employed to induce mutations and for gene replacement strategies (Storici et al., 2003; Gimble, 2005; Marcaida et al., 2010; Siegl et al., 2010). With regards to gene replacements the strategy employed is to generate a double stranded break in the targeted gene and...
by co-transforming simultaneously the cells with a segment of DNA that shares homology with the target sequence and therefore homologous recombination would allow for gene replacement (Stoddard, 2011).

Usually HEase assays originating from non-thermophiles are performed at 30 to 37°C (Kowalski and Derbyshire, 2002; Sethuraman et al., 2009) and the twintron encoded HE demonstratedendonuclease activity over a wide temperature range from 25°C up to 45°C but it appears to lose activity above 45°C; the latter based on the CD spectropolarimetry temperature assays is probably due to loss of structural integrity. This would suggest that I-CthI has not yet adapted to higher temperatures and maybe this HEG has invaded the C. thermophylum var. thermophilum mtDNA recently by horizontal transmission. Overall the temperature range for I-CthI endonuclease activity is similar to that employed for many restriction enzymes so I-CthI could be readily utilized as a rare cutting endonuclease under standard laboratory conditions. In biotechnological applications it is sometimes desirable to control HEase activity. The twintron ORF studied herein offers a system where the internal group II intron could be the key to engineer an endonuclease with an "on switch". Splicing of the group II intron could be a regulatory step that allows for the maturation of the HEase transcript and thus translation of the ORF. This may have applications in bacterial systems which are more amenable to group II intron splicing (Toor et al., 2006; Yao and Lambowitz, 2007; Yao et al., 2013) unlike eukaryotic systems where so called debranching enzymes can degrade transcripts that have complex folds such as those generated by group II intron RNAs (Mastroianni et al., 2008). A potential trigger for the splicing of the internal group II intron might involve controlling the availability of magnesium as has been demonstrated by Truong et al. (2013) for E. coli. In general controlling the cleavage activity of HEases is considered a desirable feature in order to optimize these elements as tools for gene replacements as it would allow for controlling HEase activity based on cellular conditions and thus reduce or delay potential toxic effects on the cell when HEase activity is not desirable (Posey and Gimble, 2011).

Double motif LAGLIDADG HE are compact monomeric endonucleases and the genes that encode them can be recovered from many microbial and organellar genomes (Hafez and Hausner, 2012; Hafez et al., 2014) thus bioprospecting for HEGs might be an avenue of acquiring protein scaffolds that can be used directly or modified in order to develop tools for genome editing by targeting specific sequences (Baxter et al., 2012). The I-CthI is a promising addition to the currently characterized fungal derived HEases for future applications in biotechnology (Taylor et al., 2012). The mSI247 twintron also demonstrates how composite mobile elements can evolve by different categories of mobile introns inserting into one another.

Acknowledgments

We also would like to thank Dr. Mohamed Hafez for discussions and sharing ideas about the mSI247 intron and Dr. Joe O’Neil (Department of Chemistry, University of Manitoba) for generously allowing access to his laboratory and assistance the circular dichroism spectropolarimetry work. We would like to thank Dr. P.C. Loewen (Department of Microbiology, University of Calgary) for mentorship and support. Finally we would like to thank Dr. Steve Zimmerly (Biological Sciences, University of Calgary) for his suggestions on group II intron in vitro splicing assays. This research is supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada to G.H. We also gratefully acknowledge support for TKG from a University of Manitoba Faculty of Science Graduate Award.

References


Buck, K.W., Lange, T., 2013. Cellulolytic potential of thermophilic species from four fungal orders. Amb Express 3, 47.


Moreira, S., Breton, S., Burger, G., 2012. Unscrambling genetic information at the
Marton, I., Honig, A., Omid, A., De Costa, N., Marhevka, E., Cohen, B., Zuker, A.,
Lambowitz, A.M., Zimmerly, S., 2011. Group II introns: mobile ribozymes that