Nucleic Acids Research, 1992, Vol. 20, No. 23 6401–6407

The I-*Ceu*I endonuclease recognizes a sequence of 19 base pairs and preferentially cleaves the coding strand of the *Chlamydomonas moewusii* chloroplast large subunit rRNA gene

Philip Marshall and Claude Lemieux*
Département de Biochimie, Faculté des Sciences et de Génie, Université Laval, Québec G1K 7P4, Canada

Received May 22, 1992; Revised and Accepted November 6, 1992

ABSTRACT

The I-*Ceu*I endonuclease is a member of the growing family of homing endonucleases that catalyse mobility of group I introns by making a double-strand break at the homing site of these introns in cognate intronless alleles during genetic crosses. In a previous study, we have shown that a short DNA fragment of 26 bp, encompassing the homing site of the fifth intron in the *Chlamydomonas eugametos* chloroplast large subunit rRNA gene (Ce LSU·5), was sufficient for I-*Ceu*I recognition and cleavage. Here, we report the recognition sequence of the I-*Ceu*I endonuclease, as determined by random mutagenesis of nucleotide positions adjacent to the I-*Ceu*I cleavage site. Single-base substitutions that completely abolish endonuclease activity delimit a 15-bp sequence whereas those that reduce the cleavage rate define a 19-bp sequence that extends from position −7 to position +12 with respect to the Ce LSU·5 intron insertion site. As the other homing endonucleases that have been studied so far, the I-*Ceu*I endonuclease recognizes a non-symmetric degenerate sequence. The top strand of the recognition sequence is preferred for I-*Ceu*I cleavage and the bottom strand most likely determines the rate of double-strand breaks.

INTRODUCTION

A novel class of enzymes has been recently added to the existing repertoire of DNA endonucleases. These enzymes are called 'homing endonucleases' and are encoded by mobile introns of the group I family (1–10). In some cases, the genes encoding these endonucleases are in frame with the sequence of the preceding exon (4, 5) while in others they are contained within the intron itself (11, 12, 13). The homing endonucleases catalyse intron mobility by making a double-strand cut at the homing sites of these introns in cognate intronless alleles during genetic crosses (14). The insertion of introns in these alleles is most likely to occur via a gap repair mechanism (15).

To date, homing endonucleases have been identified in the mitochondrial large ribosomal subunit (LSU) rRNA (I-*Sce*I) (2) and *cox*1 (I-*Sce*II, I-*Sce*III and I-*Sce*IV) (4, 5, 9, 10) genes of *Saccharomyces cerevisiae*, in the nuclear LSU rRNA gene (I-*Ppo*I) of *Physarum polycephalum* (3), in the *td* (I-*Tev*I) and *sun*Y (I-*Tev*II) genes of the bacteriophage T4 (7), in the chloroplast LSU rRNA gene (I-*Ceu*I) of *Chlamydomonas eugametos* (8), in the chloroplast LSU rRNA gene (I-*Cre*I) of *Chlamydomonas reinhardtii* (16) and in the mitochondrial apocytochrome b gene (I-*Csm*I) of *Chlamydomonas smithii* (17). Many of these enzymes have been overproduced in *E.coli* and shown to generate 4-nucleotide (nt) extensions with 3′-OH overhangs (4, 5, 18, 19, 20) with the notable exception of the bacteriophage T4 endonucleases I-*Tev*I and I-*Tev*II that generate 2-nt and 3-nt extensions respectively (21, 22). Most homing endonuclease genes encode gene products that have a molecular weight of less than 30 kDa, the latter acting on their respective substrates as monomers or as dimers (23, 24). Recent studies have shown that, unlike type II restriction enzymes, the homing endonucleases I-*Sce*I and I-*Sce*II recognize non-symmetric degenerate sequences of 18 bp (18, 25). Such sequences occur infrequently in large genomes and become useful markers for their physical mapping (26, 27).

In this study, we show that the I-*Ceu*I homing endonuclease cleaves a pseudosymmetric degenerate sequence of 19 bp and generates double-strand breaks at the *C.moewusii* homing site as a result of two successive nicks that are four nucleotides apart. The top strand is preferred for I-*Ceu*I cleavage and the bottom strand is likely to determine the rate of double-strand breaks.

MATERIALS AND METHODS

Strains and enzymes

E.coli DH5α F′IQ competent cells were purchased from Bethesda Research Laboratories (Bethesda, MD.) and used as

* To whom correspondence should be addressed

recommended. Bluescript plasmid SK$^+$ was purchased from Stratagene (LaJolla, CA.). The I-*Ceu*I endonuclease was kindly provided by New England Biolabs (Beverly, MA.).

Randomized mutagenesis

Using an automated CYCLONE DNA synthesizer (Millipore, Bedford, MA.), mutations were introduced at random from positions 11 to 36 in the 46-mer oligonucleotide (5'-CCCCGA-ATTCTAACTATAACGGTCCTAAGGTAGCGAGGAAGCT-TCC-3') by contaminating each of the four phosphoramidites with 1.3% (v/v) of each of the three others (28). The mutagenized oligonucleotide was purified on a NENSORB20 column (NEN) as suggested by the manufacturer. Two μg of purified oligo-nucleotide contained in a total volume of 20 μl of annealing buffer (50 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 5 mM DTT) were heated at 70°C for 10 min and allowed to cool to room temperature over a 20 min period. Annealed products were incubated for 30 min at room temperature in the presence of 5 U of Klenow fragment of DNA pol I and 500 μM dNTPs. The resulting homoduplex (82 bp) was then incubated with 50 U of *Eco*RI for 2 h at 37°C and cleavage products (72 bp) were resolved on a 5% polyacrylamide gel in TBE buffer (45 mM Tris-borate, 1 mM EDTA pH 8.0). The 72-bp DNA fragment was extracted from the gel by the crush and soak method (29) and resuspended in 20 μl of deionized water. The fragment was digested with 50 U of *Hin*dIII for 2 h at 37°C, extracted twice with phenol-chloroform (1:1) and precipitated with two volumes of 95% ethanol. The DNA was resuspended in deionized water and 100 ng of this material was ligated to 100 ng of pBluescript SK$^+$ vector that had been previously linearized with *Hin*dIII and *Eco*RI. Ligations were carried out for 2 h at room temperature in 20 μl of ligation buffer (66 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT, 5 mM ATP, 1 mM EDTA, 50 μg/ml bovine serum albumin), in the presence of 1 U of T4 DNA ligase in a total volume of 20 μl. A 1 μl aliquot was used to transform *E.coli* DH5α F'IQ competent cells and transformants were plated on YT (8 g tryptone, 5 g yeast extract, 5 g NaCl, 7.5 g bacto-agar per liter) plates containing ampicillin at 100 μg/ml. Colonies were streaked onto Hybond-N membranes (Amersham). The membranes were prehybridized for 2 h at 65°C in 6×SSC (3 M NaCl, 0.3 M sodium citrate), 0.1% SDS, 10×Denhardt (0.2% Ficoll 400, 0.2% polyvinyl pyrrolidone, 0.2% bovine serum albumin), 50 μg/ml denatured salmon sperm DNA and hybridized in 6×SSC, 1%SDS, 10×Denhardt, 1mM EDTA pH 8.0, 50 μg/ml denatured salmon sperm, at room temperature for 5 h with a 26-mer (5'-TAACTATAACGGTCCTAAGGTAGCGA-3') oligonucleotide that was labelled using T4 polynucleotide kinase and 50 μCi of [γ-^{32}P] dATP (3000Ci/mmol; NEN). Hybridizations were followed by three 5 min washes in 6×SSC, 0.1%SDS at room temperature and two stringent washes of 2 min each at 65°C and at 69°C. Transformants containing plasmids with single mutations were among those which hybridized strongly to the oligonucleotide probe at 65°C but not at 69°C. These clones were verified by sequence analysis, using the T7 sequencing kit of Pharmacia. The resulting pCMW and pCMM constructs contained respectively the wild-type I-*Ceu*I recognition sequence or single-base mutations.

Site-directed mutagenesis

The oligonucleotides (5'-TAGGACCGTT(ACG)TAGTTAGA-AT-3', 5'-TTAGGACCGT(AGC)ATAGTTAGAA-3', 5'-CC-TTAGGACC(ATC)TTATAGTTAG-3' and 5'-GCTACCTTA-GCACCGTTATA-3') were used to introduce changes at nucleotide positions −7, −6, −4 and +1 of the I-*Ceu*I recognition sequence by site-directed mutagenesis. Underlined bases indicate the nucleotide changes and the letters in parentheses designate the three bases that were used to replace the nucleotide from the wild-type sequence. The procedure was carried out essentially as outlined by Kunkel et al. (30) with the following modifications. For each mutation, 10 pmole of the phosphorylated oligonucleotide were annealed to 1 pmole of single-stranded pCMW DNA in 10 μl of annealing buffer (10 mM Tris−HCl pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT). The annealing mix was heated at 70°C and allowed to cool to room temperature over a 20 min period. Oligonucleotides were elongated in the presence of 10 U of T4 DNA polymerase, 5 U of T4 DNA ligase, 50 μg of the T4 gene product 32 and of 500 μM dNTP in 50 μl of T4 DNA polymerase buffer (33 mM Tris-acetate pH 8.0, 66 mM potassium acetate, 10 mM $MgCl_2$, 5 mM DTT) (31). The reaction mix was diluted 50 fold and 1 μl was used to transform *E.coli* DH5α F'IQ competent cells.

In vitro endonuclease assays

The plasmid DNA preparations that were used for these assays were isolated using the midi-plasmid extraction kit of QIAGEN (Chatsworth, CA.). pCMM and pCMW constructs were linearized either with *Dra*I or with *Sca*I. Plasmids derived from pACYC184 were digested with *Eco*RI. These plasmid digests were extracted once with phenol-chloroform (1:1) and precipitated with 2.5 volumes of 95% ethanol. All DNAs were resuspended in deionized water and a sample was diluted 1:10 in T E buffer (10 mM Tris−HCl pH 8.0, 1 mM EDTA). Each sample was incubated in 500 μl of 1N HCl in the presence of diamino benzoic acid and quantified by fluorimetry at 520 nm. pCMM and pCMW plasmid DNA preparations were digested with 0.1 U of I-*Ceu*I for 16 h at 37°C in 20 mM Tris−HCl pH 8.8, 5 mM $MgCl_2$, 1 mM DTT. Cleavage products were analysed on a 0.8% agarose gel and stained with ethidium bromide (0.5 μg/ml). Gels were photographed using Kodak EKTAPAN film and the negatives were scanned with a Molecular Dynamics densitometer. Cleavage efficiencies were determined as described by Montheilet et al. (1990).

Synthetic DNA substrates

The synthetic DNA substrate was prepared by mixing equimolar amounts (100 ng) of complementary 50-mer oligonucleotides (5'-GCAGGAATTCAACTATAACGGTCCTAAGGTAGCGA-AATTCAAGCTTATCG-3' and 5'-CGATAAGCTTGAATT-TCGCTACCTTAGGACCGTTATAGTTGAATTCCTGC-3'). One or the other oligonucleotide was 5'-labelled with [γ-^{32}P] dATP (3000 Ci/mmol; NEN) in a total volume of 10 μl of kinase buffer (66 mM Tris−HCl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT) by using T4 polynucleotide kinase. The oligonucleotides were heated at 70°C for 5 min and cooled to room temperature over a 20 min period in a total volume of 20 μl of annealing buffer (50 mM Tris−HCl pH 8.0, 10 mM $MgCl_2$, 50 mM NaCl, 5 mM DTT). The annealed products were resolved on a 15% polyacrylamide gel in the presence of a 10-fold excess of unlabelled duplexes and extracted using the 'crush and soak' method (29). Each duplex (15 nM) was incubated with 0.1 U of I-*Ceu*I endonuclease, in a total volume of 20 μl, for 60 min at 37°C. Aliquots were taken at different time points and cleavage products were analysed on both native 15% polyacrylamide and

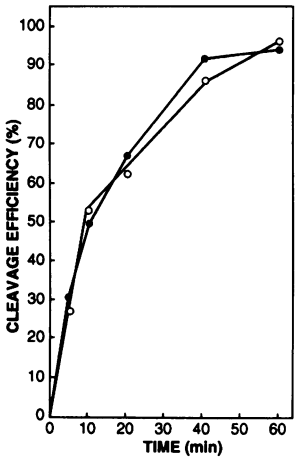

Figure 1. The natural substrate of the I-*Ceu*I endonuclease is contained within a 26-bp sequence. Plasmids (7.5 nM) pCMW (open circles) and pBHS (closed circles) were linearized at a unique *Sca*I site and incubated with 0.1 U of I-*Ceu*I endonuclease for 60 min at 37°C. Aliquots were removed at 0, 5, 10, 20, 40, 60 min time points and cleavage products were analysed on a 0.8% agarose gel. Cleavage efficiencies were determined by densitometry scanning and plotted against the time of incubation.

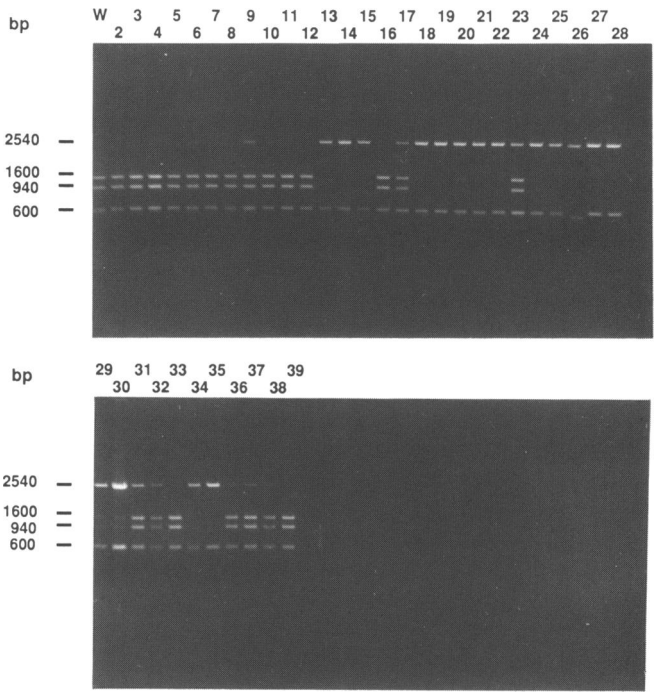

Figure 2. In vitro endonuclease assays with pCMM single-base mutants. *Dra*I digests (300 ng) of pCMW (W) and of pCMM mutants (numbered at the top of the gel) were incubated with 0.1 U of I-*Ceu*I endonuclease for 16 h at 37°C. Cleavage products (open triangles) were analysed on a 0.8% agarose gel and stained with ethidium bromide (0.5 μg/ml).

denaturing 7% polyacrylamide urea-gels. After fixing, the gels were dried, autoradiographed using preflashed KODAK X-Omat AR films and scanned with a Molecular Dynamics densitometer.

RESULTS

Recognition sequence of the I-*Ceu*I endonuclease

a) The natural substrate of the I-CeuI endonuclease is contained within a short sequence of 26 bp. In a previous study, we showed that a short sequence of 26 bp (5'-TAACTATAACGGTCCTA-AGGTAGCGA-3') extending from position −13 to position +13 with respect to the Ce LSU·5 intron insertion site can be cleaved by the I-*Ceu*I endonuclease (20). To use this sequence as an efficient substrate for the study of its recognition site, we compared the cleavage rate of the 26-bp sequence to that of a 575-bp cpDNA fragment that contains the insertion site of the Ce LSU·5 intron and encompasses the last 275 bp of the fourth intron and the first 300 bp of the fifth exon of the LSU rRNA gene of *C.moewusii* (8). The 26-bp sequence and the cpDNA fragment were both cloned into the plasmid vector pBluescript SK^+ to generate the recombinant plasmids pCMW (this study) and pBHS respectively (20). The two plasmids were linearized at a unique *Sca*I site and incubated in the presence of 0.1 U of I-*Ceu*I endonuclease for 60 min at 37°C. Cleavage products from different time points were analysed on a 0.8% agarose gel and cleavage efficiencies were determined by densitometry scanning. As shown in Figure 1, the 26-bp sequence is cleaved as rapidly as the 575-bp cpDNA fragment indicating that the short sequence can be used to study the interactions that occur between I-*Ceu*I and its natural substrate from the chloroplast genome of *C.moewusii*.

b) Random mutagenesis of the I-CeuI recognition sequence. In order to determine the recognition sequence of the I-*Ceu*I endonuclease, single-base substitutions were introduced by random mutagenesis in the 26-bp region encompassing the Ce LSU·5 intron insertion site in the chloroplast LSU rRNA gene of *C.moewusii*. The mutant oligonucleotides that were generated by using this strategy were cloned between the *Hin*dIII and *Eco*RI sites of the pBluescript SK^+ vector. Preliminary screening of the transformants was performed by differential hybridization and clones that contained potential point mutations were analysed by sequence analysis. Approximately 40% of the 300 mutants that were sequenced contained single mutations. The remaining clones harbored wild-type sequences and a few contained multiple mutations. Overall, 34 single-base mutations were obtained for all but four positions. Mutations for these positions were later obtained by site-specific mutagenesis. All 38 single-base mutations, presented in Table 1, were tested for I-*Ceu*I cleavage at pH 8.8 in the presence of 5 mM $MgCl_2$ for 16 h at 37°C. As the I-*Ceu*I endonuclease is highly sensitive to salt concentration, contaminations of inhibitory substances was eliminated as a possible cause for the absence of I-*Ceu*I cleavage by showing that contructs that harbor the wild-type I-*Ceu*I site are cleaved in the presence of pCMM mutants that are not cleaved (data not shown). The commercial preparation of I-*Ceu*I endonuclease used in this study is almost homogeneous and free of any detectable nuclease activity (Davis, T.B., Jack, W. and Schildkraudt, I. unpublished observations). Incubation periods were extended to 16 h to achieve complete cleavage of plasmid DNA containing the wild-type I-*Ceu*I recognition sequence. Following successful cleavage by I-*Ceu*I, *Dra*I digests of pCMM mutants generate three fragments of 600, 1600 and 940 bp, in contrast to the two fragments of 600 and 2540 bp that are produced by *Dra*I cleavage alone (Fig. 2). Incomplete cleavage by the I-*Ceu*I endonuclease yields four fragments of 600, 940, 1600 and 2540 bp.

Table 1. Positions of single-base mutations and their effect on I-*Ceu*I cleavage

Mutant [a]	Position	Mutation	Cleavage efficiency [b]
2	-13	T→A	100
3	-12	A→G	100
4	-12	A→C	100
5	-11	A→C	100
6	-9	T→A	98
7	-8	A→T	98
8	-8	A→C	100
9	-7	T→G	84
10	-7	T→A	98
11	-6	A→T	99
12	-5	A→G	98
13	-5	A→T	0
14	-4	C→A	0
15	-3	G→A	14
16	-3	G→T	100
17	-3	G→C	70
18	-2	G→A	0
19	-2	G→T	0
20	-1	T→A	15
21	+1	C→G	0
22	+2	C→A	0
23	+3	T→A	63
24	+4	A→T	10
25	+4	A→C	0
26	+5	A→C	0
27	+6	G→A	0
28	+6	G→T	5
29	+7	G→A	13
30	+7	G→T	8
31	+8	T→C	73
32	+9	A→T	76
33	+9	A→C	97
34	+10	G→A	0
35	+10	G→C	0
36	+11	C→A	88
37	+12	G→T	87
38	+12	G→C	89
39	+14	G→C	100

[a]All mutants tested are numbered from 2 to 39. Strain number 1 corresponds to the wild-type sequence in our collection.
[b]The efficiency of cleavage at the mutant sites is expressed as the percentage of cleaved products relative to the substrate present in the reaction.

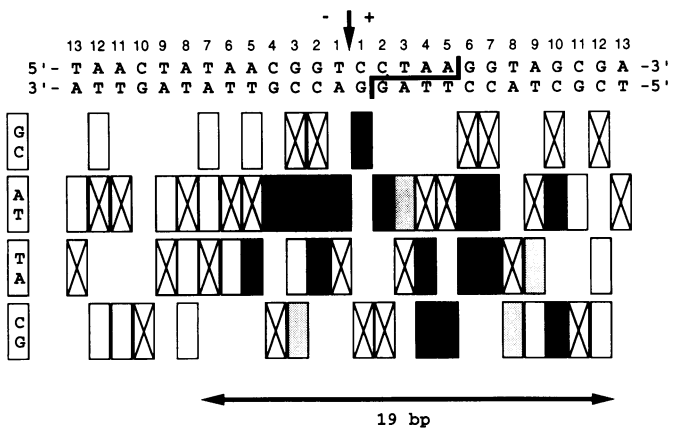

Figure 3. Recognition sequence of the I-*Ceu*I endonuclease. The top sequence corresponds to the nucleotide sequence of the oligonucleotide in which single-base mutations were introduced; + and − indicate positions that are downstream or upstream of the Ce LSU·5 homing site respectively. The arrow denotes the position of the Ce LSU·5 intron insertion site site and the staggered line the I-*Ceu*I cleavage site. The single-base mutants tested are divided into the four categories represented by rectangles with different shading intensities that are inversely proportional to the cleavage efficiency (Table 1) at the mutation sites: black boxes, no cleavage; dark boxes, less than 25% cleavage; light boxes, 25–75% cleavage; and open boxes, greater than 75% cleavage. The cross-hatched rectangles represent the wild-type sequence.

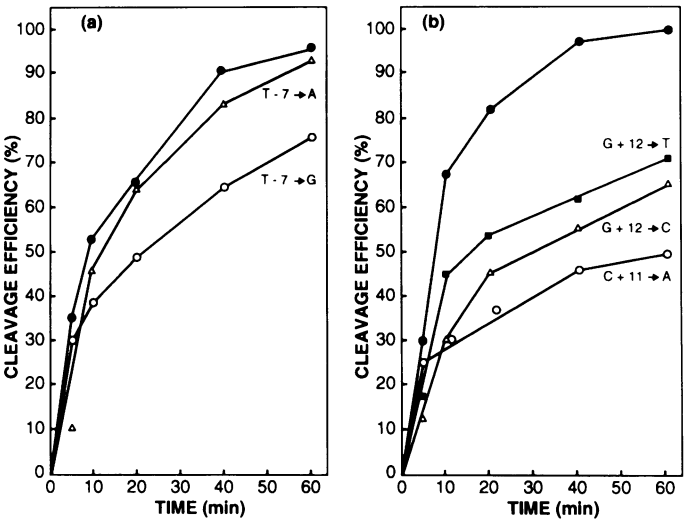

Figure 4. Positions −7, +11 and +12 are recognized by the I-*Ceu*I endonuclease. The cleavage rate obtained with pCMM plasmids (7.5 nM) containing point mutations at positions −7, +11, and +12 were compared to the rate obtained with the pCMW plasmid containing the wild-type sequence. In panel **A)**, plasmid pCMW (closed circles) and mutants numbered 9 (open circles) and 10 (open triangles) (Table 1) were linearized at the unique *Sca*I site and incubated with 0.1 U of I-*Ceu*I endonuclease for 60 min at 37°C. Aliquots were removed at 0, 5, 10, 20, 40 and 60 min and cleavage products were analysed on a 0.8% agarose gel. In panel **B)** plasmid pCMW (closed circle) and mutants numbered 36 (open circles), 37 (dark squares) and 38 (open triangles) (see Table1) were treated as described in **A**.

Mutations that completely abolish endonuclease cleavage delimit a sequence of 15 bp extending from position −5 to position +10 with respect to the Ce LSU·5 intron insertion site. Although many point mutations contained within this sequence inhibit I-*Ceu*I cleavage, several allowed either partial or complete cleavage (Fig. 3). Since we have not examined the effect of all possible mutations on I-*Ceu*I cleavage, the present study does not allow us to determine the relative importance of all nucleotide positions in the I-*Ceu*I recognition site. We consider, however, that nucleotide positions −2, +4, +6, +7 and + 10 are probably the most important for I-*Ceu*I recognition as two mutations at each of these positions strongly inhibit cleavage. This is in sharp contrast to nucleotide positions −3, +3, +8, +9 that minimally affect cleavage. The same results were observed for the nucleotides bordering the 15-bp region; point mutations at these positions allowed cleavage efficiencies that are greater than 80%.

To determine whether the nucleotides at positions −7, +11 and +12 affect or not I-*Ceu*I cleavage, kinetics of product accumulation were determined using linearized pCMM substrates containing mutations at positions −7, +11 and +12 and compared to those of pCMW. If single-base changes at the above-mentioned positions affect recognition and/or cleavage, the rate of cleavage for these mutant substrates is expected to be reduced with respect to the wild-type sequence. As shown in fig. 4, a T→G substitution at position −7 reduces the cleavage rate only slightly whereas point mutations at positions +11 and +12 reduce the cleavage rate by approximately 40 to 50%. Note that mutations extending outside the −7 to +12 region do not affect I-*Ceu*I cleavage.

Because no point mutation were recovered at position +13, the 3′ boundary of the I-*Ceu*I recognition site could lie at position +12 or +13. To resolve this ambiguity, a 19-bp sequence

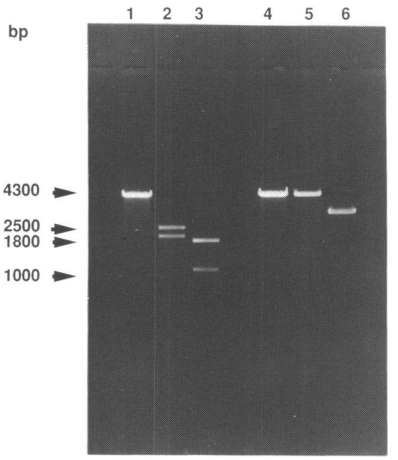

Figure 5. A 19-bp sequence from the chloroplast genome of *C.moewusii* mediates I-*Ceu*I cleavage. Linearized pACYC184 (lanes 1and 4), pRSC (lanes 2 and 5) and pCMW (lanes 3 and 6) plasmids were incubated in the presence (lanes 1, 2, 3) and absence (lanes 4, 5, 6) of 0.1 U of I-*Ceu*I endonuclease for 2 h at 37°C. Cleavage products were analysed on a 0.8% agarose gel and stained with ethidium bromide (0.5 μg/ml) and their respective molecular weights are as indicated.

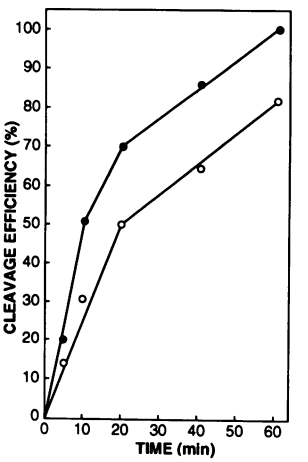

Figure 6. The sequence of 19 bp is not cleaved as efficiently as the 26-bp sequence. The plasmids (7.5 nM) pRSC (open circles) and pAHS (closed circles) were incubated with 0.1 U of I-*Ceu*I endonuclease for 60 min at 37°C. Aliquots were taken at time points of 0, 5, 10, 20, 40, 60 min and cleavage products were analysed on a 0.8% agarose gel. Cleavage efficiencies were plotted against the time of incubation.

(5′-TAACGGTCCTAAGGTAGCG-3′) extending from position −7 to position +12 was cloned into the *Bam*HI site of the pACYC184 vector and tested for I-*Ceu*I cleavage. The resulting pRSC construct was linearized at a unique *Eco*RI site and shown to be completely cleaved following an incubation of 2 h at 37°C (Fig. 5). As the initial rate of cleavage observed with pRSC is slightly reduced compared to that of the wild-type sequence (Fig. 6), we conclude that position +13 alone or combined with other nucleotides outside the proposed recognition sequence of 19 bp is required for full recognition of the natural I-*Ceu*I cleavage site.

I-*Ceu*I shows preferential cleavage for the coding strand

As the I-*Ceu*I endonuclease recognizes a different sequence on each DNA strand, we hypothesized that cleavage of one strand would be preferred to the cleavage of the other strand (32). To test this hypothesis, we synthesized a double-stranded substrate of 50 bp that contained a single copy of the I-*Ceu*I recognition sequence. In the experiments described below, only one of the two oligonucleotides constituting this substrate carried a label at its 5′ end, allowing us to monitor I-*Ceu*I cleavage of each strand separately.

The synthetic DNA substrate was incubated in the presence of I-*Ceu*I for 60 min at 37°C. Aliquots were removed at different time points and cleavage products were separated on both native and denaturing polyacrylamide urea-gels. Single-strand nicks are revealed in denaturing conditions whereas double-strand breaks are revealed on native gels. As shown in Fig. 7, the results indicate that the nicks made in the top (coding) strand accumulate much faster than nicks in the bottom strand; the latter progressing at the same rate as that of double-strand breaks. We suggest that cleavage of the bottom strand is 'rate-limiting' with respect to the production of double-strand breaks at the Ce LSU·5 homing site. As non-synthetic substrates are also cleaved in this manner (data not shown), we do not believe that cleavage preference depends on the nature of the DNA substrate used. Since the present experiment only monitors the formation the cleavage products, it is unclear whether this preference occurs at the recognition or at the cleavage steps.

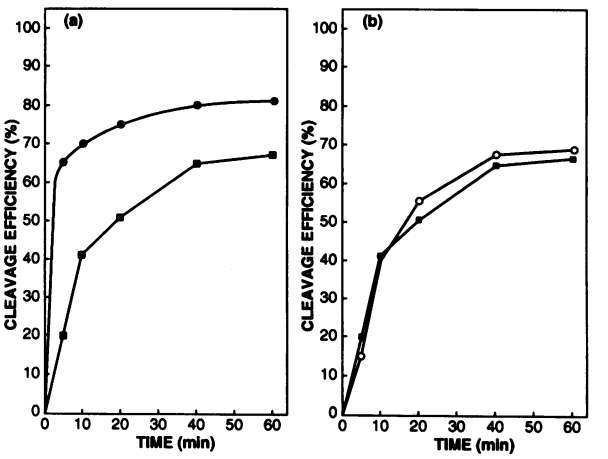

Figure 7. Asymmetrical cleavage at the I-*Ceu*I homing site. This panel compares the kinetics of I-*Ceu*I cleavage at the top and bottom strands to the kinetics of double-strand break accumulation. End-labelled DNA substrates (15 nM) were incubated with 0.1 U of I-*Ceu*I endonuclease for 60 min at 37°C. Aliquots were removed at 0, 5, 10, 20, 40, 60 min and analysed on both non-denaturing 12% polyacrylamide gels and denaturing 7% polyacrylamide urea-gels. The values in panel **A)** represent the fraction of substrate molecules that underwent nicking (closed circle) at the top strand or double-strand cleavage (dark squares). In panel **B)**, the values represent the fraction of substrate molecules that underwent nicking in the bottom strand (open circle) or cleavage (dark squares).

DISCUSSION

We have determined the recognition sequence of the I-*Ceu*I endonuclease by introducing point mutations in a short 26-bp region that harbors the Ce LSU·5 homing site in the LSU rRNA gene of *C.moewusii*. Point mutations that completely abolish I-*Ceu*I cleavage delimit a 15-bp sequence whereas those that reduce

the cleavage rate define a 19-bp sequence (5'-TAACGGTCCT-AAGGTAGCG-3') extending from position −7 to position +12 with respect to the Ce LSU·5 intron insertion site. As observed for the homing endonucleases I-*Sce*I and I-*Sce*II, the I-*Ceu*I endonuclease recognizes a degenerate sequence at certain positions throughout the entire recognition site. For example, nucleotides at positions −7, −6, −3, +3, +8, +9, +11, and +12 can undergo several mutations and still allow I-*Ceu*I cleavage while others at positions −2, +4, +6, +7, +10 severely affect cleavage. The importance of positions −5 and −3 is not clear because point mutations at each of the two positions generate cleavage efficiencies ranging from 0% to 100% (Table 1). Observations of this nature have also been reported for the endonucleases I-*Sce*I and I-*Sce*II of yeast mitochondria. Compared to the I-*Sce*II endonuclease, the I-*Ceu*I and I-*Sce*I endonucleases appear to be more specific, the more relaxed specifity of I-*Sce*II being accentuated when the endonuclease is synthesized in *E.coli* (33).

Although the 19-bp sequence is not cleaved as efficiently as the wild-type sequence, it does mediate complete I-*Ceu*I cleavage in a foreign context. The importance of nucleotides, outside this sequence, is rather difficult to address as they affect I-*Ceu*I cleavage only slightly. Similar conclusions were drawn from an *in vivo* study on site recognition by the yeast HO nuclease which showed that point mutations that reduce HO recognition and/or cleavage delimit a recognition core of 8 non-contiguous bp while multiple mutations outside this core delimit a sequence as large as 30 bp (34). If recognition by the I-*Ceu*I endonuclease follows a similar rule, a combination of point mutations outside the 19-bp region is likely to affect cleavage.

The I-*Ceu*I recognition sequence has no axis of symmetry *per se*, but contains an imperfect palindrome (5'-CCTAAGG-3') extending from positions +1 to +7 that may hold an important role in the recognition and/or cleavage step(s) because most changes within this sequence abolish I-*Ceu*I cleavage. As a result, the recognition sequence of the I-*Ceu*I endonuclease is said to be pseudosymmetric rather than asymmetric. Because of this latter feature, we proposed that cleavage reactions are quite different for each DNA strand. Although we have shown that I-*Ceu*I cleavage occurs preferentially at the top strand and that cleavage of the bottom strand probably determines the rate of double-strand breaks, it is not clear whether this preference occurs at the recognition or at the cleavage level. Asymmetric cleavage was also reported in studies involving the λ integrase (35). In this particular case, orderly cleavage ensures sequential strand transfer reactions that happen at the *att* sites shared by the lambda and *E.coli* genomes. Because the mechanism by which λ integrates into the chromosome of *E.coli* is fundamentally distinct from the gap repair model that was proposed for intron homing, the significance of asymmetric cleavage by I-*Ceu*I in the transmission of group I introns during genetic crosses remains to be elucidated.

Since point mutations, belonging to the pCMM series, have not been tested in the chloroplast of *Chlamydomonas*, it remains to be seen if the nucleotide sequence, designated as the I-*Ceu*I recognition sequence, promotes the homing of the Ce LSU·5 intron during the genetic crosses. On a statistical basis, the recognition of a 19-bp sequence is sufficient to ensure homing specificity in the chloroplast genome of *C.moewusii* (292 kbp). From the set of experiments described in this study, we have shown that the I-*Ceu*I homing endonuclease could be useful for the analysis of very large genomes. This endonuclease would be particularly useful to analyse bacterial genomes because it recognizes a sequence that is highly conserved among LSU rRNA genes. The number of I-*Ceu*I homing sites encountered in bacterial genomes is expected to depend upon the copy number of ribosomal DNA operons. Consistent with the prediction is the finding that the I-*Ceu*I endonuclease cleaves the *E.coli* genome at nine distinct sites (Davis, T.B. and Schildkraudt, I., unpublished observations). Although mapping of these sites has not been carried out, it is most likely that cleavage occured at each of the seven copies of the ribosomal operon in the *E.coli* genome. This would explain the toxic effect caused by the expression of the I-*Ceu*I endonuclease gene in *E.coli* (8).

ACKNOWLEDGEMENTS

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada and 'Le Fonds pour la Formation de Chercheurs et l'Aide à la Recherche'. C.L. is Scholar in the Evolutionary Biology Program of the Canadian Institute for Advanced Research and P.M. is supported by a postgraduate scholarship from 'Le Fonds pour la Formation de Chercheurs et l'Aide à la Recherche'.

REFERENCES

1. Perlman, P.S. and Butow, R.A. (1989) Science **246**, 1106−1109.
2. Colleaux, L., d'Auriol, L., Betermier, M., Cottarel, G., Jacquier, A., Galibert, F. and Dujon, B. (1986) Cell **44**, 521−533.
3. Muscarella, D.E. and Vogt, V.M. (1989) Cell **56**, 443−454.
4. Wenzlau, J.M., Saldanha, R.J., Butow, R.A. and Perlman, P.S. (1989) Cell **56**, 421−430.
5. Delahodde, A., Goguel, V., Becam, A.M., Creusot, F., Perea, J., Banroques, J. and Jacq, C. (1989) Cell **56**, 431−441.
6. West, D.K., Changchien, L.-M., Maley, G.F. and Maley, F. (1989) J. Biol. Chem. **264**, 10343−10346.
7. Quirk, S.M., Bell-Pedersen, D. and Belfort, M. (1989) Cell **56**, 455−465.
8. Gauthier, A., Turmel, M. and Lemieux, C. (1991) Curr. Genet. **19**, 43−47.
9. Séraphin, B., Faye, G., Hatat, D. and Jacq, C. (1992) Gene (Amst.) **113**, 1−8.
10. Sargueil, B., Delahodde, A., Hatat, D., Tian, G.L., Lazowska, J. and Jacq, C. (1991) Mol. Gen. Genet. **225**, 340−341.
11. Dujon, B. (1980) Cell **20**, 185−197.
12. Lemieux, C. and Lee, R.W. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 4166−4170.
13. Gott, J.M., Zeeh, A., Bell-Pedersen, D., Ehrenman, K., Belfort, M. and Shub, D.A. (1988) Genes Dev. **2**, 1791−1799.
14. Dujon, B. (1989) Gene (Amst.) **82**, 91−113.
15. Szostak, J.W., Orr-Weaver, T.L. and Rothstein, R.J. (1983) Cell **33**, 25−35.
16. Durrenberger, F. and Rochaix, J.-D. (1991) EMBO J. **10**, 3495−3501.
17. Colleaux, L., Michel-Wolwertz, M.-R., Matagne, R.F. and Dujon, B. (1990) Mol. Gen. Genet. **223**, 288−296.
18. Colleaux, L., d'Auriol, L., Galibert, F. and Dujon, B. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 6022−6026.
19. Muscarella, D.E., Ellison, E.L., Ruoff, B.E. and Vogt, V.M. (1990) Mol. Cell. Biol. **10**, 3386−3396.
20. Marshall, P. and Lemieux, C. (1991) Gene (Amst.) **104**, 241−245.
21. Chu, F.K., Maley, G., Pedersen-Lane, J., Wang, A-M. and Maley, F. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 3574−3578.
22. Bell-Pedersen, D., Quirk, S.M., Clyman, J. and Belfort, M. (1990) Nucleic Acids Res. **18**, 3763−3770.
23. Monteilhet, C., Perrin, A., Thierry, A., Colleaux, L. and Dujon, B. (1990) Nucleic Acids Res. **18**, 1407−1413.
24. Wernette, C., Saldahna, R., Perlman, P.S. and Butow, R.A. (1990) J. Biol. Chem. **265**, 18976−18982.
25. Wernette, C., Saldanha, R., Smith, D., Ming, D., Perlman, P.S. and Butow, R.A. (1992) Mol. Cell. Biol. **12**, 716−723.
26. Vollrath, D., Davis, R., Connelly, C. and Hieter, P. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 6027−6031.
27. Thierry, A., Perrin, A., Boyer, J., Fairhead, C., Dujon, B., Frey, B. and Schmitz, G. (1991) Nucleic Acids Res. **19**, 189−190.

28. Hill, D.E. (1989). In Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K., (eds) Current protocols in molecular biology, John Wiley and Sons, New York, Vol. 1, unit 8.2.A.
29. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 560–564.
30. Kunkel, T.A., Roberts, J.D. and Zakour, R.A. (1987) Methods Enzymol. **154**, 367–382.
31. Craik, C.S., Largman, C., Fletcher, T., Roczniak, S., Barr, P.J., Fletterick, R. and Ruttler, W. (1985) Science **228**, 291–297.
32. Lesser, D.R., Kurpiewski, M. and Jen-Jacobson, L. (1990) Science **250**, 776–785 .
33. Sargueil, B., Hatat, D., Delahodde, A. and Jacq, C. (1990) Nucleic Acids Res. **18**, 5659–5665.
34. Nickoloff, J., Singer, J. and Heffron, F. (1990) Mol. Cell. Biol. **10**, 1174–1179.
35. Nunes-Duby, S.E., Matsumoto, L. and Landy, A. (1989) Cell **59**, 197–206.