DNA substrate specificity and cleavage kinetics of an archaeal homing-type endonuclease from *Pyrobaculum organotrophum*

Jens Lykke-Andersen, Hoa Phan Thi-Ngoc and Roger A. Garrett*

Institute of Molecular Biology, Copenhagen University, Sølvgade 83H, DK-1307 Copenhagen K, Denmark

Received August 25, 1994; Accepted October 10, 1994

**ABSTRACT**

The protein encoded by intron 1 of the single 23S rRNA gene of the archaeal hyperthermophile *Pyrobaculum organotrophum* was isolated and shown to constitute a homing-type DNA endonuclease, I-Por. It cleaves the intron-23S rDNA of the closely related organism *Pyrobaculum islandicum* near the site of intron insertion in *Pb.organotrophum*. In contrast, no endonuclease activity was detected for the protein product of intron 2 of the same gene of *Pb.organotrophum* which, like I-Por, carries the LAGLI-DADG motif, common to group I intron-encoded homing enzymes. I-Por cleaves optimally at 80°C, with $k_{cat}$ and $K_m$ values of about 2 min$^{-1}$ and 4 nM, respectively, and generates four nucleotide 3'-overhangs and 5'-phosphates. It can cleave a 25 base pair DNA fragment encompassing the intron insertion site. A mutation-selection study established the base pair specificity of the endonuclease within a 17 bp region, from positions -6 to +11 with respect to the intron-insertion site. Four of the essential base pairs encode the sequence involved in the intron-exon interaction in the pre-rRNA that is required for recognition by the RNA splicing enzymes. Properties of the enzyme are compared and contrasted with those of eucaryotic homing endonucleases.

**INTRODUCTION**

Archaeal introns have been found in stable RNA genes of extreme halophiles and hyperthermophiles. They constitute a separate class of introns which generate a characteristic ‘bulge-helix-bulge’ motif at the exon/intron junctions of the pre-RNAs, which is recognized by a cleavage enzyme (1-6). Some pre-RNA introns contain open reading frames (ORFs) encoding putative proteins carrying the LAGLI-DADG motif (7-9). These introns circularize after excision, generating highly structured and stable RNA species in vivo (3, 6, 8).

The LAGLI-DADG motif was first characterized for group I intron-encoded proteins of lower eucaryotes (10). Many of these proteins have been identified as homing endonucleases which recognize, and cut, at the intron insertion site in intron- alleles. They recognize non-palindromic sequences, 15 to 20 base pairs in length, and produce four nucleotide 3'-overhangs with 5'-phosphates on cleavage. The cleavage can lead to intron homing, where an intron is copied from an intron- to an intron- allele (reviewed in 11, 12). More recently, LAGLI-DADG proteins have been found as protein fusion inteins, which cleave an intein- allele of the gene prior to intein homing (13-16). Other homing endonucleases, lacking the LAGLI-DADG motif, some of which carry a GIY-YIG motif, are encoded by group I introns of bacteriophages; they cleave intron- alleles, distal to the intron insertion site, creating small 3’- or 5’-overhangs (17-19). A few LAGLI-DADG proteins have been shown to exhibit maturase activity and they facilitate splicing of the introns that encode them (20-22).

Recently, two LAGLI-DADG-type proteins from archaeal hyperthermophiles were found to have DNA endonuclease activity: one is expressed as a protein fusion product together with the DNA polymerase of *Thermococcus litoralis* (16) and the other is encoded in the 23S rRNA intron of *Desulfurococcus mobilis* (7, 23). The archaeal hyperthermophile *Pyrobaculum organotrophum* has two introns in its single 23S rRNA gene which encode putative LAGLI-DADG proteins. These are both absent from the closely related organism *Pyrobaculum islandicum* (8, 24). In the present work, we show that the protein encoded by the first intron is a homing-type endonuclease. Its substrate specificity shows similarities, and significant differences, from homing endonucleases encoded by eucaryotic and bacteriophage group I introns. The protein encoded by intron 2, on the other hand, showed no endonuclease activity.

**MATERIALS AND METHODS**

Construction of expression vectors for the intron-encoded proteins

A Shine-Dalgarno sequence was inserted upstream from the coding sequences in order to express the putative proteins in *E.coli*. Moreover, for intron 1, the putative start codon was

*To whom correspondence should be addressed
changed from GUG to AUG. Transcription vectors for the two introns were constructed using the polymerase chain reaction (PCR). Primers 5'-AAGGAGAAATTAATGAGATAC-AGTATG and 5'-GCCCTAGGTCACAGAT were used for intron 1, and primers 5'-AAGGAGAAATTAATGAGGCTGTGATTGGT and 5'-GAACTTCACCGGTTGGA were employed for intron 2. The PCR reaction mixture contained 100 µl of 10 mM Tris—HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 0.01% gelatin, 100 µM dNTPs, 20 pmole of each primer and 50 ng of pUC19 plasmid containing a part of the 23S rRNA gene of Ph. organotrophum. The mixture was overlaid with mineral oil and after denaturing at 95°C, for 5 min, 0.5 units Taq DNA polymerase (Amersham) was added per 25 µl solution. After 30 cycles at 56°C for 0.5 min, 72°C for 1 min, and 92°C for 0.5 min, 2 units Klenow fragment were added and the mixture was incubated for 30 min at room temperature. The product was then purified on a low melting agarose gel, and cloned into the Stratagene pUT719, a derivative of pUC19 containing a T7 promoter upstream from the polylinker (25). Each construct was then sequenced in order to establish that no construction or copying errors had occurred.

Expression of the proteins

In vitro translation was performed by transcription of linearized pUT719 derivatives (25), containing the intron-ORFs, with T7 RNA polymerase, in vitro, followed by incubation with rabbit reticulocyte lysate as described earlier (23).

High expression of the proteins was obtained by transforming the T7-promoter constructs into E.coli strain BL21/DE3, containing the F-factor from E.coli XL1-blue. E.coli BL21/DE3 contains an IPTG inducible T7 RNA polymerase (26). An overnight culture of transformed BL21/DE3 cells was diluted 1:100, in LB media containing 10 µg/ml tetracyclin and 100 µg/ml ampicillin, and grown to A₆₀₀ ~ 1 at 37°C. IPTG was then added to a concentration of 0.5 mM, and the cells were grown overnight. 1.5 ml of cell culture was pelleted by centrifuging at 15 000 r.p.m. for 5 min and resuspended in 200 µl of 100 mM Tris—HCl (pH 7.5), 50 mM NaCl, 10 mM MgCl₂, 2 mM DTT, 10% glycerol, 10% DMSO, 0.1% Triton X-100, 60 µg/ml PMSF. The cells were lysed by sonication and cell debris was removed by centrifuging at 15 000 r.p.m. for 5 min. The supernatant was then added to a concentration of 0.5 mM, and the cells were grown overnight. 1.5 ml of cell culture was pelleted by centrifuging at 15 000 r.p.m. for 5 min. The supernatant was then added to a concentration of 0.5 mM, and the cells were grown overnight. 1.5 ml of cell culture was pelleted by centrifuging at 15 000 r.p.m. for 5 min. The supernatant was then added to a concentration of 0.5 mM, and the cells were grown overnight. 1.5 ml of cell culture was pelleted by centrifuging at 15 000 r.p.m. for 5 min. The supernatant was then added to a concentration of 0.5 mM, and the cells were grown overnight.
Portl (giving approximately 50% cleavage in the first cycle); 2) purifying the uncleaved substrate on an agarose gel, and 3) a standard PCR reaction on the uncleaved substrate. After the fourth cycle, the uncleaved substrate was cloned into the HincII-site of pUC19 (31). Single-mutants were identified by sequencing.

Cleavage assays
In cleavage assays with I-Portl, the substrate was present on a PvuII cleaved pUC19 derivative, containing a 205 bp PCR fragment of Pb.islandicum 23S rDNA which included the cleavage site (with or without mutation). The substrate was [32P] 5'-end-labeled by incubating with [γ-32P]ATP and T4 polynucleotide kinase (28), after 5'-dephosphorylation with calf intestinal alkaline phosphatase (Amersham). I-Portl was added to a mixture of a small amount of end-labeled DNA substrate and a specified amount of unlabeled substrate in 9 μl of the standard cleavage buffer [50 mM Hepes–KOH (pH 8.0), 100 mM KCl, 0.5 mM EDTA, 2 mM DTT, 10% glycerol, 0.1% Triton-X100]. The mixture was overlaid with 25 μl of mineral oil and heated at 80°C for 1 min. The reaction was started by adding MgCl2 (preheated) to 10 mM and maintaining at 80°C, and stopped by adding 90 μl of ice-cold 10 mM Tris–HCl (pH 8.0), 5 mM EDTA, 0.1% SDS, 0.1 mg/ml tRNA, and placing on ice before extracting once with phenol:CHCl3 (1:1) (saturated with TE buffer), and once with CHCl3. The DNA was precipitated with ethanol, washed, dried and dissolved in 10 μl loading buffer. It was then denatured at 95°C for 3 min and 8 μl was loaded on a 5% polyacrylamide/7M urea gel. After electrophoresis, the gel was dried and the band intensities were quantified in an Instant imager (Packard, Meriden, USA).

RESULTS
Intron 1 encodes a DNA endonuclease, intron 2 does not
In order to establish whether the introns in the 23S rRNA gene of Pb.organotrophum (Fig. 1) encode homing endonucleases, cell extracts of Pb.organotrophum were tested for DNA endonuclease activity. Fragments of the intron - 23S rRNA gene of Pb.islandicum were cloned and incubated with cell extracts of Pb.islandicum and Pb.organotrophum (Fig. 2a, lanes 1 and 2, respectively). Only one cut was produced by the cell extract of Pb.organotrophum confirming preliminary evidence reported earlier (32). The fragment sizes correspond approximately to cleavage at the insertion site of intron 1. This suggested that only one of the protein products was an endonuclease and, therefore, DNA fragments, containing the two ORFs, were inserted after a T7 RNA polymerase promoter in plasmid pUT719 and the proteins were expressed in rabbit reticulocyte lysate (see Materials and Methods) (Fig 2b). The proteins were then assayed for endonuclease activity (Fig. 2a, lanes 3 to 5). The protein encoded by intron 1 cleaved intron - 23S rDNA, while that encoded by intron 2 showed no activity over the temperature range 50 to 95°C (data not shown). The DNA endonuclease is named I-Portl, according to the convention for homing endonucleases (33). The protein from intron 2, which is of unknown function, is labeled pPO2.

In order to obtain higher expression of I-Portl, we transformed the pUT719-derivative, containing the ORF of intron 1, into E.coli strain BL21/DE3 (26). Partially purified I-Portl, expressed from this strain, was used in the following experiments.

Figure 2. Cleavage assays on cell extracts and isolated intron-encoded proteins. (a) Ethidium bromide stained agarose gel showing cleavage of 23S rRNA of Pb.organotrophum after incubation with 1: cell extract of Pb.islandicum, 2: cell extract of Pb.organotrophum, 3: negative control, 4: I-Portl and 5: pPO2. Samples 3 to 5 were present in rabbit reticulocyte lysates. The insert (substrate) was a ~3.5 kbp BamHI fragment containing most of the single 23S rRNA gene of Pb.islandicum (24), cloned into phage vector M13mp18 and linearized with PvuII. Cleavage conditions are described in Materials and Methods. M: phage λ digested with BstEII. Crude cell extracts of Pb.organotrophum and Pb.islandicum were prepared by a procedure described earlier (3). (b) Autoradiogram showing expression of I-Portl and pPO2 in rabbit reticulocyte lysate, and a negative control of reticulocyte lysate incubated with [35S]-methionine. Proteins were expressed in the presence of [35S]-methionine (see Materials and Methods), separated on a 12% polyacrylamide/SDS gel and autoradiographed.
The temperature optimum of \( l\text{-Porl} \) was established by incubating the enzyme with \( [\text{32P}]\)-5'-end-labeled DNA substrate over the range 20°C to 95°C (Fig. 3a). Activity was observed between 56°C and 95°C, with optimal cutting at 80°C (Fig. 3b). This correlates with the high growth temperature of \( \text{Pb} \text{.organotrophum} \) (78°C to 102°C) (34). No activity was observed at 37°C, even at very high enzyme concentrations (data not shown).

**Temperature dependence of \( l\text{-Porl} \)**

The exact cleavage site of \( l\text{-Porl} \) was established using a primer extension approach (27). A double stranded DNA substrate was prepared by primer extension, using oligodeoxynucleotide primers annealed to sites 50–100 nucleotides upstream, and downstream, from the putative cleavage site, on the 23S rRNA gene of \( \text{Pb} \text{.islandicum} \) which was cloned into the phage vectors M13mp18 and M13mp19. The products of the primer extension reactions were incubated with \( l\text{-Porl} \), and subsequently run on 6% polyacrylamide/urea sequencing gels, alongside dideoxy-sequencing reactions with the same primers (35) (Fig. 4). Cuts were detected 5 bp and 1 bp downstream from the intron insertion site, on the coding and noncoding strand, respectively, which generated 3'-overhangs of four nucleotides. Furthermore, ligation of the cleaved substrate, by T4 DNA ligase, established that 5'-phosphates were created in the cleavage reaction (data not shown). Since the cleavage products were separated on a denaturing gel (see Materials and Methods), the degree of cleavage on each strand could be examined, and was shown to be equal (data not shown). Cleavage also occurred when dITP was substituted for dGTP in the primer extension reaction (data not shown).

The size of the recognition sequence was established using a primer extension approach (27). This is based on the principle that cleavage only occurs when a complete substrate is synthesized before insertion of a dideoxynucleotide, with resultant elimination of a band from the sequencing gel. After performing the dideoxy-sequencing reactions from primers used in the cleavage site determination, \( l\text{-Porl} \) was added and the mixture was incubated at 80°C. The sequencing gels, illustrated in Fig. 4, indicate a minimal substrate size of about 20 bp. Therefore, a 25 bp DNA fragment which covered this region was synthesized and incubated with \( l\text{-Porl} \). The results, shown in Fig. 5, demonstrate that it was cut, although one strand was affected more than the other; this probably reflects the partial dissociation of the 9-mer fragment on the right hand strand (Fig. 4) after cleavage at 80°C.

**Steady-state kinetics**

Steady-state kinetics measurements were performed on \( l\text{-Porl} \) in order to gain more insight into the mechanisms of protein-DNA recognition and DNA cleavage. The \( l\text{-Porl} \) substrate was present in a 205 bp DNA fragment of \( \text{Pb} \text{.islandicum} \) contained in a linearized pUC19 derivative. A time course of cleavage is shown in Fig. 6a and a half reciprocal plot of initial cleavage rate (\( v_{0} \)) versus the initial substrate concentration (\( S \)) in Fig. 6b, where 1 unit \( l\text{-Porl} \) is defined as the amount of enzyme which cleaved 1 \( \mu \text{g} \) of substrate (2891 bp) in one hour, at 80°C, in standard cleavage buffer (see Materials and Methods). The specific activity of \( l\text{-Porl} \) was estimated at about 4 fmole/unit enzyme by SDS-polyacrylamide gel electrophoresis. From these data, the \( k_{\text{cat}} \) value was estimated at 2 min\(^{-1} \) (at 80°C) and the \( K_{m} \) value was calculated to be 4 \( \text{nM} \) (Fig. 6a+b). This indicates that the binding constant, \( K_{b} \), for \( l\text{-Porl} \) on the wild-type cleavage site, is \( > 250 \mu \text{M}^{-1} \) (since \( K_{b} > 1/K_{m} \) and \( K_{m} = 4 \text{nM} \)).

**Mutation-selection analysis of the cleavage site**

Mutations were generated in the cleavage site by mutagenic PCR. Single-site mutants with low cleavage rates were selected (see Materials and Methods). Those mutations which impaired or abolished \( l\text{-Porl} \) cleavage were concentrated in a 17 bp region, between positions -6 and +11 with respect to the intron insertion site (Fig. 4). It was found also, as for the wild-type substrate, that each of the mutants that exhibited impaired cleavage was cut equally on both strands (data not shown).
Figure 4. Determination of the cleavage-site and the minimal substrate size. Autoradiograms showing polymerization extensions treated with (+) or without (−) I-Porl; dideoxy-sequencing reactions alone (A, G, C, T), and dideoxy-sequencing reactions with I-Porl included (+ I-Porl). (T/+) indicates a mixture of the (+) sample and the sequencing (T) track. All extensions were from primers annealed downstream (left) or upstream (right) from the putative insertion site of intron 1 (see Fig. 1). The sequence of the DNA 25-mer (see Fig. 5) is shown and the minimal substrate is boxed. The cleavage sites are indicated by horizontal arrows and the intron insertion site is marked by asterisks. Vertical arrows indicate the direction of electrophoresis. The experimental procedures are described in Materials and Methods.

Double reciprocal plots are illustrated for I-Porl and some of the mutant substrates in Fig. 6c which illustrate the range of curves obtained, many of which overlapped and are therefore not shown. $k_{cat}$ values derived for all of the mutants are listed in Table 1. They varied from 2 to 80% of the value for the wild-type substrate. The $K_m$ values, in contrast, which are also listed in Table 1, were almost constant, between 4 and 5.5 nM, for all the substrates. The data were summarized in Fig. 7 where the mutants which showed impaired cleavage are indicated above the DNA substrate sequence. $k_{cat}$ values are also expressed, in histogram form, as percentage of the value for the wild-type substrate. Mutations located at positions −6, −5, −4, −3 and +8 produced uncleavable substrates (see also Table 1).

Competitive binding experiments were performed to gain more insight into I-Porl specificity. The wild-type substrate on the linearized pUC19 was treated with I-Porl in the presence of the following putative competitors: pUC19 DNA (a) lacking an insert, (b) with inserts carrying the single intron/exon junctions of intron 1, and (c) with inserts containing the uncleavable single mutants. The results are presented as double reciprocal plots (Fig. 6d) and demonstrate that (a) pUC19 DNA lacking the insert competes weakly in the cleavage reaction, while (b) insertion of intron/exon junctions, or (c) uncleavable substrates (data not shown), did not enhance the competitive capacity of pUC19 DNA. It was inferred, therefore, that I-Porl does not bind specifically to the intron/exon junctions or to the uncleavable single-mutants. The inhibition constant of pUC19 DNA without an insert (2686 bp) was approximately 14 nM. Assuming that there is no specific binding to pUC19, and no sequence similarity with the substrate site was detected, then the unspecific DNA binding constant of I-Porl is about 0.025 μM$^{-1}$ base pair (less than 1/10$^4$ of the specific binding, see above).

A similar estimate was obtained when pUC19 DNA was replaced by phage λ DNA in the cleavage reaction mixture.

**DISCUSSION**

I-Porl is a homing-type endonuclease

I-Porl, encoded by intron 1 of the single 23S rRNA gene of *Pb.organotrophum*, constitutes a site-specific DNA endonuclease.
The endonuclease, which is expressed in vivo, cleaves the intron-23S rRNA gene of the closely related organism *Pb.islandicum*, near the site of intron insertion in *Pb.organotrophum*. The cleavage reaction generates four nucleotide 3’-extensions with 5’-phosphates as found for another archaeal intron-encoded endonuclease I-Dmol (23). The minimal substrate is approximately 20 bp. These cleavage properties are similar to those of the homing endonucleases encoded by group I introns of eucarya (reviewed in 12). In contrast, pPO2, encoded by intron 2, exhibited no DNA endonuclease activity either in the isolated protein or in cell extracts of *Pb.organotrophum*, and identification of its function must await further research.

**Substrate specificity of I-Porl**

Random mutation of the DNA substrate, and selection for impaired or abolished endonuclease cleavage, produced seventeen single-site mutants, concentrated within a 17 bp region. Mutations at five positions yielded uncleavable substrates and mutations at nine other positions produced impaired cleavage. No mutations were detected at positions T(−2), A(−1) and A(+10) (see Fig. 7) and we infer, given the selection procedure, that mutations at these positions do not strongly impair cleavage by I-Porl. One can also calculate from the number of clones of uncleavable single-site mutants (Table 1), and the AT—GC/GC—AT ratio generated by the mutational procedure (see Materials and Methods), that the probability that we failed to detect transitions at these positions that would abolish cleavage, is about 25%. By a similar argument, the probability that transitional mutations in the G-C base pairs flanking the 17 base pair region (Fig. 7) could abolish cleavage, and remain undetected in our experiments, is about 1%.

Several single mutants of the substrate showed reduced...
cleavage, on both strands, relative to the wild-type substrate and they yielded $k_{cat}$ values between 2 and 80% of that of the wild-type substrate (Fig. 7). Furthermore, since the $K_m$ values are fairly constant for these substrates (4 to 5.5 nM) (Table 1), one can infer that I-PorI binds more strongly to substrates with higher $K_m$ values, unless the off-rate of the enzyme is much higher than the catalytic rate. This implies, in turn, that the mutations affect both binding to the DNA substrate and the cleavage.

The results indicate some important differences from the substrate recognition mechanism of eucaryotic homing endonucleases. First, the general level of substrate specificity of I-PorI (Table 1; Fig. 7), appears to lie between the high degree of specificity observed for I-Scel and I-Cell (36, 37), and the lower degree of specificity found for I-Scell (38, 39). Furthermore, I-Scel, from yeast, binds specifically to its downstream exon-intron junction, and to the downstream half of the substrate generated by the cleavage reaction, which may facilitate strand transfer during the intron homing process (40). In contrast, no specific binding of I-PorI was observed to the exon-intron junctions (Fig. 6d). Moreover, we infer that I-PorI did not bind to either of the exon halves generated in the cleavage reaction because no cleavage inhibition was observed when wild-type substrate was added in excess of I-PorI (Fig. 6a).

**Sequence constraints and intron mobility**

Although the mobility of intron 1 of *Pb.organotrophum* has not yet been demonstrated, it has been shown that the archaeal homing endonuclease (I-Dmol)-encoding intron of *Dc.mobilis* (23) cannot undergo homing within the *Dc.mobilis* genome, but can spread through intron - 23S rRNA genes in a population of *Sulfobolus acidocaldarius*, where the intron insertion sequence is conserved (Aagaard, C. and Garrett, R.A., unpublished). One question that this phenomenon raises is: what are the sequence constraints on such mobility? An important prerequisite for an archaeal intron homing event is that the sequence bordering the insertion site on the DNA is compatible with the formation of the conserved intron cleavage site in the archaeal pre-23S rRNA.

**REFERENCES**