FEBS 27300 FEBS Letters 544 (2003) 165–170

An archaeal homing endonuclease I-*Pog*I cleaves at the insertion site of the neighboring intron, which has no nested open reading frame[1]

Hitoshi Nakayama, Yayoi Morinaga, Norimichi Nomura*, Takuro Nunoura, Yoshihiko Sako, Aritsune Uchida

Division of Applied Biosciences, Graduate School of Agriculture, Kyoto University, Kyoto 606-8502, Japan

Received 19 March 2003; revised 25 April 2003; accepted 26 April 2003

First published online 13 May 2003

Edited by Horst Feldmann

Abstract Homing endonucleases (HEs) of the LAGLIDADG family cleave intron/inteinless cognate DNA at, or near, the insertion site (IS) of their own intron/intein. Here, we describe a notable exception to this rule. Two introns, Pog.S1205 (length 32 bp) and Pog.S1213 (664 bp), whose ISs are 8 bp apart, exist within the 16S rRNA gene of the archaeon *Pyrobaculum oguniense*. Pog.S1213 harbors a nested open reading frame (ORF) encoding a 22 kDa monomeric protein, I-*Pog*I, which contains two LAGLIDADG motifs and has optimal DNA cleavage activity at 90°C. Intriguingly, I-*Pog*I cleaves the Pog.S1205-less substrate DNA in the presence or absence of Pog.S1213. The cleavage site (CS) of I-*Pog*I does not coincide with the IS of Pog.S1213 but with that of Pog.S1205. Thus, I-*Pog*I activity both promotes the homing of its own intron, Pog.S1213, and guarantees co-conversion of the ORF-less intron Pog.S1205.
© 2003 Federation of European Biochemical Societies. Published by Elsevier Science B.V. All rights reserved.

Key words: Mobile intron; Intron homing; Homing endonuclease; LAGLIDADG family; Flanking marker co-conversion; Archaea

1. Introduction

A subset of introns undergo a process termed 'homing' [1], whereby they are efficiently integrated into intronless cognates of the intron-containing alleles. This process is initiated by site-specific DNA endonucleases (homing endonucleases; HEs) encoded within the introns themselves [2]. HEs recognize lengthy sequences (14–40 bp) that are usually centered on the intron insertion sites (ISs) of intronless alleles. Once homing has occurred, the recognition sequence is interrupted by the intron, and the resultant allele is immune to cleavage by the HE. This mode of action of HEs allows the host organism to sustain non-reciprocal homologous recombination events.

Intron-encoded HEs are phylogenetically widespread, occurring in archaea, bacteria and eukaryotes. Over 30 HEs encoded in group I, group II and archaeal introns have been characterized (see REBASE at http://rebase.neb.com/rebase/rebase.homing.html for a compilation). They are classified into three families, LAGLIDADG, GIY-YIG and ββα-Me, based on structural similarity [3]. The vast majority of known HEs belong to the LAGLIDADG family. Analysis of this family using a hidden Markov model reveals that only 8% of the amino acid residues are highly conserved and most of these occur in the catalytically essential LAGLIDADG motif [4]. Despite the low overall similarity of their primary sequences, the LAGLIDADG enzymes share several features. Firstly, they cut the cognate intronless double-stranded DNA leaving 4 nucleotide (nt) 3′ single-stranded overhangs at the cleavage site (CS). Secondly, the CS of a LAGLIDADG-type HE is located at, or close (within 6 nt) to, the IS of the intron encoding the HE itself.

Archaeal rRNA genes (rDNAs) are interspersed with introns [5,6], approximately 40% of which contain large open reading frames (ORFs) or ORF-like sequences. To date, three of the intronic ORFs have been demonstrated to correspond to typical LAGLIDADG-type HEs, I-*Dmo*I [7], I-*Por*I [8] and I-*Ape*I [9]. The distribution of archaeal rDNA introns is highly variable between species and strains, and does not agree with the phylogeny of the host organism deduced from the rDNA exon sequences. The occurrence of rDNA introns has been most frequently reported in the order *Thermoproteales* [10–13] (Fig. 1A). It is noteworthy that many strains of this taxon possess a pair of introns at contiguous positions 1205 and 1213 of the 16S rDNA (using the *Escherichia coli* numbering [14]), and that only the intron at position 1213 harbors an ORF or ORF-like sequence. In this case, it is not known which portion of the DNA sequence is recognized as a target and cleaved by the intron-encoded HE.

Pyrobaculum oguniense is a member of the order *Thermoproteales*, and its 16S rDNA contains the two introns Pog.S1205 and Pog.S1213 [13]. We have analyzed the properties of the site-specific double-stranded DNA endonuclease activity associated with I-*Pog*I, which is encoded by Pog.S1213. We show here that I-*Pog*I cleaves the Pog.S1205-less DNA in the presence and absence of Pog.S1213. Our results provide an example contrary to the dogma that the existence of an HE-encoding intron destroys the recognition sequence of the HE.

*Corresponding author. Fax: (81)-75-753 6226.
E-mail address: nom@kais.kyoto-u.ac.jp (N. Nomura).

[1] The nucleotide sequences of the two introns, Pog.S1205 and Pog.S1213, have been deposited to the GenBank/EMBL/DDBJ databases under the accession number AB087402.

Abbreviations: CS, cleavage site; IS, insertion site; rDNA, rRNA gene; HE, homing endonuclease

0014-5793/03/$22.00 © 2003 Federation of European Biochemical Societies. Published by Elsevier Science B.V. All rights reserved.
doi:10.1016/S0014-5793(03)00497-6

2. Materials and methods

2.1. Strains and growth conditions

P. oguniense TE7 (JCM10595) was grown at 90°C under aerobic conditions as described previously [4]. *E. coli* INVαF′ (Invitrogen) was used for plasmid manipulations and grown in liquid or on solid Luria–Bertani (LB) media supplemented with appropriate antibiotics. *E. coli* BL21(DE3) pLysS was used as the host strain for protein overexpression.

2.2. Plasmid construction

2.2.1. Intronless target plasmids. Introns are named according to the newly proposed nomenclature of archaeal rDNA introns [6]. The Pog.S1205(+)/Pog.S1213(−) target DNA was generated by annealing the complementary oligonucleotides 5′-cgctgag*GCCCCGA-A^AGCCCCGGGGCTGCACA and 5′-GTGCAGCCCCGGGGC-T^TTCGGGGC*ctcagcgA (* and ^ indicate the ISs of Pog.S1205 and Pog.S1213, respectively; the internal sequence of Pog.S1205 is represented by lowercase letters). The Pog.S1205(−)/Pog.S1213(−) target DNA was generated by annealing the complementary oligonucleotides 5′-TCAGTAT*GCCCCGAA^ACCCCGGGGCTGCACA and 5′-GTGCAGCCCCGGGGT^TTCGGGGC*ATACTGAA. The Pog.S1205(−)/Pog.S1213(+) target DNA was generated by annealing the complementary oligonucleotides 5′-TCAGTAT*GCCCCGAA-^**agccgtggccccagg**A and 5′-**cctggggccacggct**^TTCGGGGC*ATACT-GAA (the internal sequence of Pog.S1213 is represented by bold lowercase letters). The three synthetic duplexes were independently cloned into the pCR2.1 vector (Invitrogen) to yield pTD(+/−), pTD(−/−) and pTD(−/+), respectively.

*2.2.2. Plasmids used to determine the I-*Pog*I recognition site.* Plasmids containing variants of the Pog.S1205(−)/Pog.S1213(−) DNA sequence differing by gradual truncation from either end of the putative recognition site were constructed. These plasmids contain altered target sites with deletions between positions 1199 and 1228 (*E. coli* numbering) and were created by ligating a series of synthetic oligonucleotide duplexes into pCR2.1. Cloned inserts were confirmed by sequencing.

2.2.3. Expression plasmid. The full-length I-*Pog*I gene (*pogI*) was polymerase chain reaction (PCR) amplified from *P. oguniense* chromosomal DNA using primers I-PogIF (5′-CATATGGTGAAGACC-TGGGACTACCTC) and I-PogIR (5′-GGATCCCTAGAGTCTGT-CTGTCGTCCCTG). The two primers were designed to place an initiation codon (ATG) and an *Nde*I site at the 5′ end, and a *Bam*HI site at the 3′ end of the I-*Pog*I coding region. The His_6I-*Pog*I expression plasmid pET-pogI was constructed by ligating the PCR product into *Nde*I–*Bam*HI-digested pET-15b (Novagen). The insert was sequenced on both strands to confirm the correct sequence.

2.3. Protein expression and purification

E. coli BL21(DE3)/pLysS carrying pET-pogI was grown in LB supplemented with ampicillin (50 μg ml^{-1}) at 37°C to an optical density at 600 nm of 0.6, at which time expression of His_6I-*Pog*I was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. After an additional 6 h of incubation at 37°C, the cells were harvested and disrupted by sonication in buffer A (50 mM Tris–HCl (pH 8.0), 200 mM NaCl, 0.1 mM ethylenediamine tetraacetic acid (EDTA), 0.5 mM dithiothreitol, 10% glycerol). The supernatant was incubated at 80°C for 30 min to denature and precipitate most of the *E. coli* proteins. The heat-treated supernatant was dialyzed overnight against buffer NiS (50 mM sodium phosphate buffer (pH 8.0), 300 mM NaCl, 10 mM imidazole) and run on a chelating column (HiTrap, Amersham), charged with 50 mM $NiSO_4$, and equilibrated in buffer NiS. Bound proteins were eluted over a linear gradient of 10–500 mM imidazole. Fractions containing His_6I-*Pog*I were pooled and dialyzed against buffer A. The dialysate was fractionated on a HiTrap heparin column (Amersham) followed by a HiTrap SP column (Amersham). Each column was pre-equilibrated with buffer A, and bound proteins were eluted with a linear gradient of 0.2–1 M NaCl. This was followed by chromatography on a Superdex 200 column (Amersham) equilibrated in buffer A. The final fraction was judged to be near homogeneity, since it exhibited a single band by Coomassie blue staining after 0.1% sodium dodecyl sulfate (SDS)–14% polyacrylamide gel electrophoresis (PAGE). This preparation was used throughout this study.

2.4. I-Pog*I cleavage reactions*

Plasmids pTD(+/−), pTD(−/−) and pTD(−/+) were linearized with *Nco*I, which cuts the vector DNA 1.6 kb from the I-*Pog*I target site. For use as substrates, the resulting 4.0 kb fragments were extracted with phenol–chloroform and precipitated with ethanol. Substrate DNA (0.5 μg) was incubated with I-*Pog*I (0.5 μg) in a total volume of 10 μl of 33 mM Tris–acetate (pH 7.9 at 25°C), 10 mM Mg–acetate, 0.5 mM dithiothreitol and 66 mM K–acetate at 90°C for 10 min. Unless stated otherwise, pTD(−/−) linearized with *Nco*I was used as a standard substrate for endonuclease assays. During incubation, the assay mixture was overlaid with 25 μl mineral oil; reactions were stopped on ice. The reaction products were separated by electrophoresis on 1.5% agarose gels and visualized by ethidium bromide staining. The gels were photographed using a GelDoc2000 digital imaging system (Bio-Rad) and the extent of the reaction was determined using NIH image software (ver. 1.62).

2.5. Determination of the I-Pog*I CS*

Target DNA was amplified from pTD(−/−) using universal primers M13(−20) (5′-GTAAAACGACGGCCAGT) and M13 reverse (5′-CAGGAAACAGCTATGAC), both of which were biotinylated at the 5′ end. Each labeled primer was used in separate PCRs with the non-labeled partner primer. The reactions were cycled 40 times at 96°C for 30 s, 60°C for 30 s and 72°C for 10 s. Cleavage reactions were performed as above except that the PCR-generated substrates were used. Reaction products were heat-denatured and separated by electrophoresis on 6% polyacrylamide, 8.3 M urea gels, together with sequencing reactions of pTD(−/−) DNA generated from the same 5′ end-labeled primers. Biotinylated DNA was detected using an Imaging High kit (Toyobo).

2.6. Mapping of the I-Pog*I recognition site*

Plasmids containing a series of deletions surrounding the I-*Pog*I CS were linearized with *Nco*I, which cuts the vector DNA 1.6 kb from the I-*Pog*I site. The resulting 4.0 kb DNA fragments were used as substrates in the cleavage reactions as described above.

3. Results

3.1. The archaeal intron Pog.S1213 encodes a site-specific double-stranded DNA endonuclease, I-Pog*I, a new member of the LAGLIDADG family*

The *P. oguniense* 16S rDNA is interrupted by two introns, Pog.S1205 and Pog.S1213, which are 32 and 664 bp in length and inserted after positions 1205 and 1213 (*E. coli* numbering), respectively. At the intron–exon junctions, both of the introns are predicted to fold into an RNA secondary structure consisting of a bulge-helix-bulge motif, which is typical of archaeal introns [15]. Pog.S1213 has a nested free-standing ORF (579 bp in length) that uses a GTG start codon located 50 bp downstream of the 5′ end of Pog.S1213. The deduced ORF product is a 21.9 kDa, 192 amino acid polypeptide, which contains two copies of the LAGLIDADG motif (Fig. 1B). A potential ribosome-binding site, GAG, precedes the ORF 13 bp from the start codon.

The N-terminal His_6-tagged protein (His_6I-*Pog*I) was overexpressed in *E. coli* BL21(DE3) pLysS cells after induction with IPTG. His_6I-*Pog*I was recovered from soluble crude extracts, and heat treatment followed by five sequential column chromatographies were sufficient to purify His_6I-*Pog*I to near homogeneity (Fig. 2A). An immunoblot analysis using specific anti-His_6 tag antibodies confirmed that this single band, with an apparent molecular mass (M_r) of 29.6 kDa, corresponds to the overexpressed protein (data not shown). Unexpectedly high apparent molecular mass on SDS–PAGE can be attributed to the loose denaturation of this thermostable enzyme. The yield of purified His_6I-*Pog*I from 1 l culture of recombinant *E. coli* was 1.0 mg.

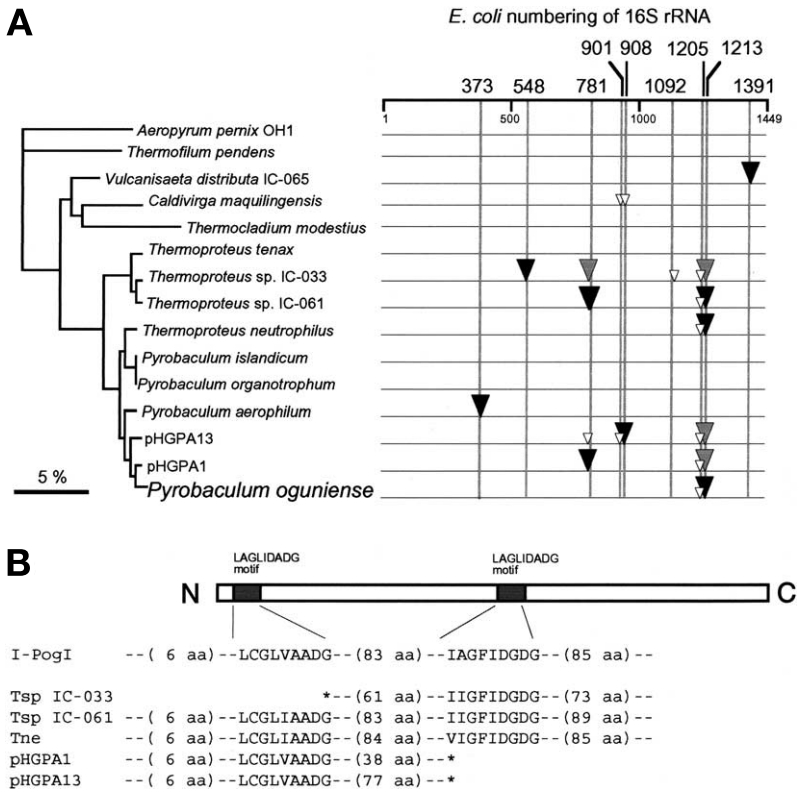

Fig. 1. A: Intron distribution in the 16S rDNAs of members of the order *Thermoproteales*. Phylogenetic relationships between the 12 strains and two environmental clones based on the 16S rDNA exon sequences are shown in the left-hand panel. The tree was constructed using the neighbor-joining method as described previously [5]; the sequence of *Aeropyrum pernix* was used as an outgroup. The scale bar indicates 0.05 substitutions per site. Triangles in the schematic diagrams on the right-hand panel indicate the ISs of introns. ORF-containing introns are represented by black triangles, ORF-less short introns by small open triangles and long introns (650–770 bp) whose nested ORFs might incur frameshift mutations are denoted by gray triangles. Numbers refer to the positions of the residues in the *E. coli* 16S rRNA sequence, which correspond to the residue immediately preceding the intron. GenBank accession numbers are as follows: *A. pernix* OH1, AB078015; *Thermofilum pendens*, X14935; *V. distributa* IC-065, AB063639; *C. maquilingensis*, AB013926; *Thermocladium modestius*, AB005296; *Thermoproteus tenax*, M35966; *Thermoproteus* sp. IC-033, AB009616; *Thermoproteus* sp. IC-061, AB00969617; *Thermoproteus neutrophilus*, AB009618; *Pyrobaculum islandicum*, L07511; *P. organotrophum*, AB063647; *P. aerophilum*, L07510; *P. oguniense*, AB087402; pHGPA1, AB027539; pHGPA13, AB027540. B: Alignment of the LAGLIDADG motifs of the amino acid sequences encoded within the position 1213 introns of *Pyrobaculum* and *Thermoproteus* 16S rDNAs. One LAGLIDADG motif has been lost in the sequences marked with *. Tsp IC-033, *Thermoproteus* sp. IC-033; Tsp IC-061, *Thermoproteus* sp. IC-061; Tne, *T. neutrophilus*.

To determine its native molecular mass, the purified His_6I-*Pog*I was rechromatographed on a calibrated Superdex 200 column. The V_e observed for His_6I-*Pog*I yielded a K_{av} of 0.60, which corresponded to an apparent molecular mass of 22.7 kDa (Fig. 2B). Taking into account the mass of the His_6 tag, this indicates that I-*Pog*I is a 21.9 kDa monomeric enzyme in solution and is likely to possess a globular shape.

The endonuclease activity of the purified His_6I-*Pog*I protein was assayed. Plasmids carrying target DNA lacking one or both introns were constructed (see Section 2). Incubation of *Nco*I-linearized pTD(−/−) and pTD(−/+) with His_6I-*Pog*I at 90°C resulted in the appearance of two discrete cleavage products of 1.6 and 2.4 kb (Fig. 3A, lanes 2 and 6). After incubation of the same substrates with His_6I-*Pog*I at 37°C, no cleavage was observed (data not shown). On the other hand, *Nco*I-linearized pTD(+/−) DNA was resistant to cleavage by His_6I-*Pog*I even at 90°C (Fig. 3A, lane 4), suggesting that the IS of Pog.S1205 is located within the recognition sequence.

3.2. Catalytic properties and optimal reaction conditions of *I*-Pog*I*

The endonuclease activity of purified His_6I-*Pog*I was assayed under different conditions, with varying temperature,

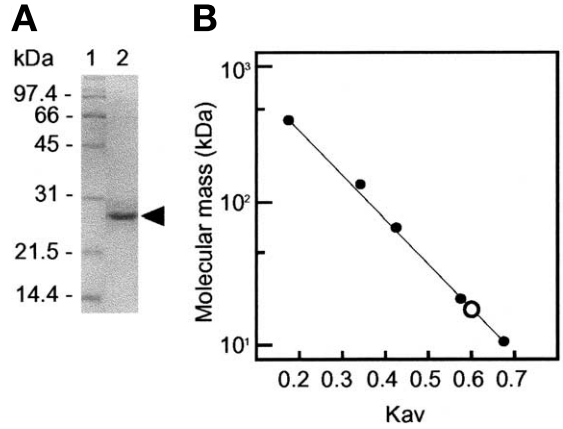

Fig. 2. A: SDS–PAGE analysis of the purified His_6I-*Pog*I. Lane 1, molecular mass markers; lane 2, purified protein (5 μg) eluted from the Superdex 200 column. B: Molecular mass determination of His_6I-*Pog*I by gel filtration under non-denaturing conditions. Molecular mass standards and the corresponding observed K_{av} values are: ferritin, 440 kDa, 0.18; aldolase, 158 kDa, 0.34; bovine serum albumin, 66 kDa, 0.42; chymotrypsinogen A, 25 kDa, 0.59; cytochrome *c*, 12.4 kDa, 0.67. The open circle indicates the elution position of His_6I-*Pog*I.

pH and salt concentration. I-*Pog*I activity displayed a fairly sharp temperature dependence, with a maximum at 90°C (Fig. 3B). At temperatures above 95°C, activity could not be estimated precisely due to the rapid denaturation of the substrate DNA. The optimal pH for cleavage activity was between 8.0 and 9.0.

The presence of 20–300 mM KCl significantly stimulated the rate of substrate cleavage. This stimulation by K^+ appeared to be somewhat specific to this monovalent cation and not merely an effect of ionic strength, since significantly lower (50–80%) activity was observed when NaCl was substituted for KCl. I-*Pog*I cleavage activity also exhibited dependence on Mg^{2+}, with an optimum at concentrations of 1–10 mM. An excess of EDTA completely inhibited the reaction. Mn^{2+} could be substituted for Mg^{2+} without altering the CS specificity of I-*Pog*I. Other divalent cations, namely Co^{2+}, Zn^{2+} and Cu^{2+}, allowed partial activity at concentrations below 0.5 mM. I-*Pog*I activity as a percentage of the activity in the presence of 0.1 mM Mg^{2+} was 80% (Co^{2+}), 40% (Zn^{2+}) and 20% (Cu^{2+}) (0.1 mM of each cation was used). In contrast, no activity was observed when Ca^{2+} or Ba^{2+} was used at the concentrations tested (0.01–20 mM).

3.3. *I-*Pog*I cleaves at the IS of the neighboring ORF-less intron Pog.S1205*

To map precisely the CS of I-*Pog*I, the intronless DNA substrates, which were 5′ end-labeled on either the top (coding) or bottom (template) strand, were digested with His_6I-*Pog*I. The cleavage products were separated on a denaturing polyacrylamide gel alongside sequencing ladders generated from the same 5′ end-labeled primers. As shown in Fig. 4A, the CSs are after positions 1205 and 1201 (*E. coli* numbering of 16S rDNA) on the top and bottom strands, respectively. I-*Pog*I cleaves to create cohesive ends with four bases protruding at the 3′ termini, as do most known LAGLIDADG-type endonucleases. Fig. 4B summarizes these results and shows the location of the double-strand break (DSB) on the cDNA sequence of *P. oguniense* 16S rRNA.

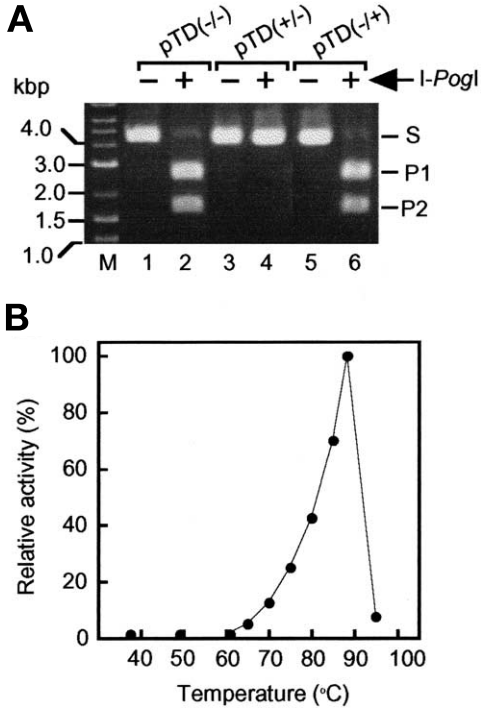

Fig. 3. I-*Pog*I exhibits endonuclease activity when the substrate DNA lacks the neighboring Pog.S1205 intron. A: The *Nco*I-cut pTD(−/−) DNA (lanes 1 and 2), pTD(+/−) DNA (lanes 3 and 4) and pTD(−/+) DNA (lanes 5 and 6) were incubated for 30 min at 90°C with and without His_6I-*Pog*I. Lanes 2, 4 and 6 contain His_6I-*Pog*I; lanes 1, 3 and 5 contain substrates incubated under the same conditions without protein. Substrates are indicated by S, and products by P1 and P2, on the right. Lane M, molecular mass markers. B: Effect of temperature on I-*Pog*I activity. The *Nco*I-cut pTD(−/−) DNA substrate was incubated for 5 min with His_6I-*Pog*I at the temperatures indicated. The relative cleavage efficiencies were estimated from band intensities on ethidium bromide-stained gels.

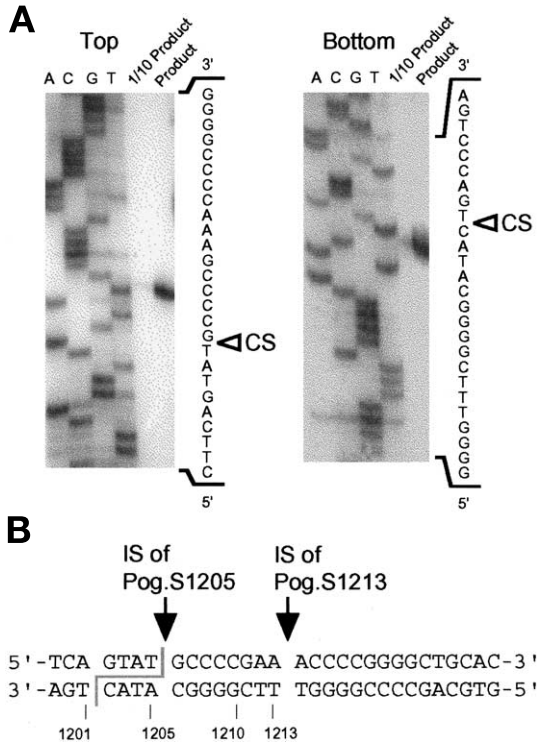

Fig. 4. Mapping of the I-*Pog*I CS. A: The products of I-*Pog*I cleavage reactions were diluted (1/10) and subjected to electrophoresis alongside sequencing ladders. Top strand cleavage is shown on the left panel, and bottom strand cleavage on the right panel. Sequencing lanes are denoted by the dideoxynucleotide species used in the reaction. The sequence of the target DNA immediately flanking the CS (represented by an arrowhead labeled CS) on each strand is shown on the right. B: The sequence surrounding the CS in the cDNA of *P. oguniense* 16S rRNA. The staggered line indicates the CS for I-*Pog*I. The numbering below the sequence is relative to the *E. coli* 16S rRNA sequence. The ISs of the two introns, Pog.S1205 and Pog.S1213, are shown by arrows.

3.4. *I-*Pog*I can cleave the Pog.S1205-less substrate DNA regardless of whether Pog.S1213 is present or not*

In general, LAGLIDADG-type HEs recognize long DNA sequences (14–30 bp) surrounding the ISs of their own introns/inteins [3]. To delimit the minimal recognition sequence of I-*Pog*I, a 30 bp stretch of DNA spanning the IS of Pog.S1213 was analyzed. A series of gradually truncated target sequences were prepared for cleavage assays using His_6I-*Pog*I (Fig. 5C). No reaction product was interpreted as an indication that the deleted base pairs lie within the recognition sequence. Fig. 5B and C show that His_6I-*Pog*I does not cleave the substrates with deletions that include the segment from

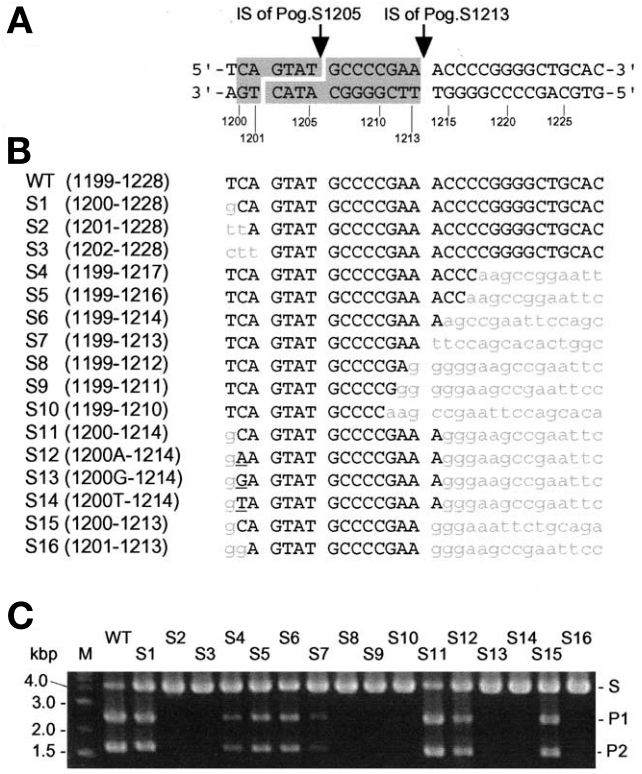

Fig. 5. Delimitation of the minimal recognition sequence of I-*Pog*I. A: Summary of recognition sequence mapping; symbols and numbering are as in Fig. 4B. The region in which deletions or substitutions abolished cleavage is shaded in gray. B: Sequences (sense strand) of the recognition site mutants. WT refers to *Nco*I-digested pTD(−/−). The name of each deletion mutant is given on the left; numbers in parentheses correspond to the positions of wild-type nucleotides (uppercase letters) remaining after deletion (relative to the *E. coli* 16S rRNA sequence). In substrates S12, S13 and S14, the C at position 1200 was substituted by the nucleotide shown in underlined uppercase letters. Nucleotides replacing the deleted regions are represented in gray lowercase. C: I-*Pog*I cleavage assay for recognition site mutants. Each substrate was incubated for 10 min at 90°C with His_6I-*Pog*I. Symbols are as in Fig. 3A.

positions 1200 to 1213 (*E. coli* numbering of 16S rDNA). Assays using the substrates with base substitutions at position 1200 (S12, S13, S14) confirm that this position is included in the recognition sequence. In addition, the substrate S15, which only contains the sequence from positions 1200 to 1213, was efficiently cleaved. These results indicate that the sequence necessary for cleavage by I-*Pog*I comprises at least 14 bp (Fig. 5A), and that the existence of Pog.S1213 does not eliminate the I-*Pog*I cleavage reaction. The latter supports the finding that I-*Pog*I cleaves the *Nco*I-linearized pTD(−/+) DNA (Fig. 3A, lane 6).

4. Discussion

In this report, we demonstrated a number of unusual features in the mode of I-*Pog*I action. Firstly, the recognition sequence of I-*Pog*I does not stretch on both sides of the IS of the Pog.S1213 intron, although the gene encoding I-*Pog*I is borne on Pog.S1213. Secondly, the recognition sequence of I-*Pog*I is not destroyed by the presence of its own intron, Pog.S1213, but by the presence of the neighboring ORF-less intron, Pog.S1205. Thirdly, the CS for I-*Pog*I lies in the IS of the Pog.S1205 intron. Finally, the center of the CS for I-*Pog*I is located in a position that divides its recognition sequence into remarkably uneven lengths (4:10 in the number of nucleotide residues). These properties have not previously been observed in known LAGLIDADG-type HEs. To date, it has been shown that the two intron-encoded HEs of the ββα-Me family, I-*Hmu*I and I-*Hmu*II, cleave far from the intron ISs and are not inhibited by the presence of the introns, although the biological implications of the nicking endonucleases are distinct from the present finding [16].

The intron homing mechanism is based on the killer/anti-killer principle. The HE ('killer' element) is deleterious for genomes with a vacant (intron-free) homing site, while the existence of the intron ('anti-killer' element), carrying the HE gene, antagonizes this activity. In general, it is critical for the proliferative nature of the intron that its IS is located within the recognition sequence of the HE. The situation we described here is, however, slightly different. The Pog.S1213 intron seems unlikely to be involved in the anti-killer effect, although its IS is located within the recognition sequence of I-*Pog*I. Instead, the neighboring Pog.S1205 intron appears to play a role as an anti-killer element for I-*Pog*I activity. Our findings raise the possibility of the co-conversion of the two introns, Pog.S1205 and Pog.S1213. Following a DSB created by I-*Pog*I, a recombination-dependent repair event would be induced from its CS. In this process, the Pog.S1205 intron is guaranteed to integrate into the intronless allele with very high frequency, because the CS of I-*Pog*I is identical to the IS of Pog.S1205. The Pog.S1213 intron is transferred as a flanking marker. The failure of Pog.S1205 to integrate into the recipient would result in cleavage of the chromosomal

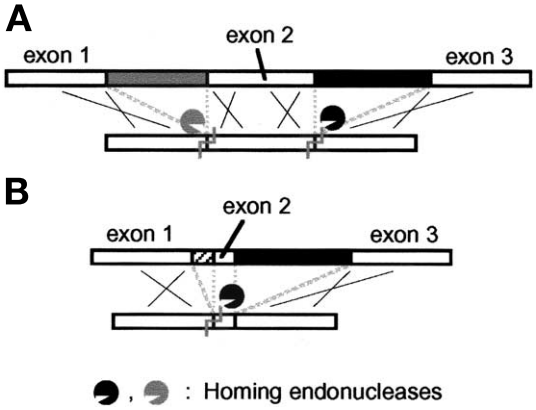

Fig. 6. Homing versus co-homing. A: Multiple independent homing of mobile introns. Black and gray boxes represent introns with a nested HE gene. An endonuclease encoded within the intron cleaves the intronless allele at, or close to, the IS of its own intron. After DSB repair, insertion of the intron destroys the recognition sequence of the HE. Each intron produces the HE and confers immunity on the intron-containing allele against the attack of the HE, acting as both killer and anti-killer in one to propagate its own copies throughout the population. B: A proposed mechanism for co-homing by I-*Pog*I. Hatched and filled boxes represent the introns Pog.S1205 and Pog.S1213, respectively. I-*Pog*I cleaves at the IS of the neighboring Pog.S1205 intron. Although Pog.S1213 produces I-*Pog*I, the existence of Pog.S1213 does not abolish cleavage by I-*Pog*I; instead Pog.S1205 serves as an anti-killer element against the killer I-*Pog*I and the roles of killer and anti-killer are divided between the two introns. DSB repair initiates recombination between the homologous regions within exons 1 and 3, resulting in the entire segment containing the two introns being transferred to the intronless alleles.

DNA by I-*Pog*I, leading to cell death. We propose that this model should be referred to as 'co-homing' (Fig. 6). Consistent with the co-homing model is the intron insertion pattern observed within the 16S rDNA of the order *Thermoproteales*. Several members of this taxon possess the two introns at positions 1205 and 1213 of the 16S rDNA, but the remainder have neither of them. To date, no example has been found where an intron occurs only at either position 1205 or 1213 (Fig. 1A).

It has been well documented in a number of organisms that multiple mobile introns exist within a gene encoding rRNA (see the Comparative RNA Web Site at http://www.rna.icmb.utexas.edu/). In these cases, each intron harbors a HE gene, and it is assumed that the introns are transferred into the cognate intronless allele by multiple independent homing events, thus each intron plays a dual role of killer and anti-killer. In the co-homing model however, a single DSB would result in simultaneous transfer of two neighboring introns, one of which has no nested HE gene. The roles of killer and anti-killer are strictly divided between the two introns, and both introns are indispensable for their propagation. Therefore, the two models, conventional homing and co-homing, are clearly distinguishable (Fig. 6). The selective constraints that could have generated the co-homing phenomenon have not yet been determined.

The LAGLIDADG-type HEs described to date possess either one or two LAGLIDADG motifs, and the two-motif enzymes fold into monomers [3]. This is also the case for I-*Pog*I. As revealed by several crystal structures, two-motif enzymes have two topologically similar domains with approximate dyad symmetry that are connected by a highly variable peptide linker [17–19]. Each domain contains a conserved LAGLIDADG motif, of which the first seven residues are part of an α-helix forming the axis of the domain interface. The DNA-binding surface of these enzymes is saddle-shaped and made up of four antiparallel β-strands from each domain. Binding of the DNA target sequence is dictated by independent sets of interactions made between individual domains and individual halves of the recognition sequence separated by the CS [20–22]. It is possible that the approximately equal size of the two domains correlates to the location of the CSs of LAGLIDADG-type enzymes at approximately the midpoint of the recognition sequence. In contrast, the exceptional location of the I-*Pog*I CS within the recognition sequence leads us to predict that the structure of the DNA-binding surface of I-*Pog*I is distinct from those of known LAGLIDADG enzymes. Crystallographic studies to determine the structure of I-*Pog*I would be of great interest to evaluate this hypothesis.

Acknowledgements: We would like to thank Prof. Ichizo Kobayashi (University of Tokyo) for helpful discussion. This work was supported by a Grant-in-Aid for Scientific Research on Priority Areas (C) 'Genome Biology', and by the National Project on Protein Structural and Functional Analyses, from the MEXT of Japan. Y.M. acknowledges support from a JSPS fellowship.

References

[1] Dujon, B., Belfort, M., Butow, R.A., Jacq, C., Lemieux, C., Perlman, P.S. and Vogt, V.M. (1989) Gene 82, 115–118.
[2] Belfort, M., Derbyshire, V., Parker, M.M., Cousineau, B. and Lambowitz, A.M. (2002) in: Mobile DNA II (Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M., Eds.), pp. 761–783, ASM Press, Washington, DC.
[3] Chevalier, B.S. and Stoddard, B.L. (2001) Nucleic Acids Res. 29, 3757–3774.
[4] Dalgaard, J.Z., Klar, A.J., Moser, M.J., Holly, W.R., Chatterjee, A. and Mian, I.S. (1997) Nucleic Acids Res. 25, 4626–4638.
[5] Nomura, N., Morinaga, Y., Kogishi, T., Kim, E., Sako, Y. and Uchida, A. (2002) Gene 295, 43–50.
[6] Morinaga, Y., Nomura, N. and Sako, Y. (2002) Microbes Environ. 17, 153–163.
[7] Dalgaard, J.Z., Garrett, R.A. and Belfort, M. (1993) Proc. Natl. Acad. Sci. USA 90, 5414–5417.
[8] Lykke-Andersen, J., Thi-Ngoc, H.P. and Garrett, R.A. (1994) Nucleic Acids Res. 22, 4583–4590.
[9] Nomura, N. (1998) Ph.D. thesis, Kyoto University, Kyoto.
[10] Itoh, T., Suzuki, K. and Nakase, T. (1998) Arch. Microbiol. 170, 155–161.
[11] Takai, K. and Horikoshi, K. (1999) Appl. Environ. Microbiol. 65, 5586–5589.
[12] Itoh, T., Suzuki, K., Sanchez, P.C. and Nakase, T. (1999) Int. J. Syst. Bacteriol. 49, 1157–1163.
[13] Sako, Y., Nunoura, T. and Uchida, A. (2001) Int. J. Syst. Evol. Microbiol. 51, 303–309.
[14] Noller, H.F. (1984) Annu. Rev. Biochem. 53, 119–162.
[15] Lykke-Andersen, J., Aagaard, C., Semionenkov, M. and Garrett, R.A. (1997) Trends Biochem. Sci. 22, 326–331.
[16] Goodrich-Blair, H. and Shub, D.A. (1996) Cell 84, 211–221.
[17] Duan, X., Gimble, F.S. and Quiocho, F.A. (1997) Cell 89, 555–564.
[18] Silva, G., Dalgaard, J.Z., Belfort, M. and Van Roey, P. (1999) J. Mol. Biol. 286, 1123–1136.
[19] Ichiyanagi, K., Ishino, Y., Ariyoshi, M., Komori, K. and Morikawa, K. (2000) J. Mol. Biol. 300, 889–901.
[20] Jurica, M.S., Monnat, R.J. and Stoddard, B.L. (1998) Mol. Cell 2, 469–476.
[21] Chevalier, B.S., Kortemme, T., Chadsey, M.S., Baker, D., Monnat, R.J. and Stoddard, B.L. (2002) Mol. Cell 10, 895–905.
[22] Moure, C.M., Gimble, F.S. and Quiocho, F.A. (2002) Nat. Struct. Biol. 9, 764–770.