© 1999 Oxford University Press
Nucleic Acids Research, 1999, Vol. 27, No. 21 **4167–4174**

PI-*Pfu*I and PI-*Pfu*II, intein-coded homing endonucleases from *Pyrococcus furiosus*. I. Purification and identification of the homing-type endonuclease activities

Kayoko Komori[1], Naoki Fujita[1,3], Kenji Ichiyanagi[1,2], Hideo Shinagawa[3], Kosuke Morikawa[2] and Yoshizumi Ishino[1,*]

[1]Department of Molecular Biology and [2]Department of Structural Biology, Biomolecular Engineering Research Institute (BERI), 6-2-3 Furuedai, Suita, Osaka 565-0874, Japan and [3]Department of Molecular Microbiology, Research Institute for Microbial Diseases, Osaka University, 3-1 Yamadaoka, Suita, Osaka 565-0871, Japan

Received July 26, 1999; Revised and Accepted September 15, 1999

ABSTRACT

We screened for proteins with specific binding activity to Holliday junction DNA from the hyperthermophilic archaeon *Pyrococcus furiosus* and found a protein that has specific affinity for DNA with a branched structure, like a three-way or four-way junction. The protein was identified as one of the two inteins encoded in the gene for ribonucleotide reductase (RNR) by gene cloning. These two inteins were spliced out from the precursor protein as polypeptides with molecular weights of 53.078 and 43.976 kDa, respectively. The amino acid sequences of these inteins have two copies of the LAGLIDADG motif, which is found in the site-specific DNA endonucleases. The purified proteins actually cleaved double-stranded DNA with the sequence of the intein$^-$ allele, and, therefore, they were designated PI-*Pfu*I and PI-*Pfu*II. They generate a 4 bp 3′-OH overhang with a 5′-phosphate, like other known homing endonucleases originating from inteins. The optimal conditions of the DNA cleavage reaction, including temperature, pH, and concentrations of KCl and $MgCl_2$, have been determined. The high affinity for junction DNA of PI-*Pfu*I was confirmed using the purified protein.

INTRODUCTION

Inteins are proteins that are spliced out from a precursor at the protein level (for reviews see 1–6). The splicing is known to be processed autocatalytically, which means that the splicing ability is entirely encoded by the inteins themselves. Inteins have been shown to be present in all three domains of life, Bacteria, Eukarya and Archaea. As a result of archaeal genome sequencing projects (7–10), many inteins have been predicted in their genomes, and about 100 inteins have been registered in the database InBase to date (11; http://www.neb. com/neb/inteins.html). Among the registered sequences, 25 proteins have been experimentally determined to be functional inteins. Since an intein is a type of movable genetic element that promotes DNA rearrangement, it may be an important factor in the evolution of genomes. The molecular mechanism of its invasion into intein-less sites of DNA is also interesting for molecular biologists, especially in the field of DNA recombination. Moreover, the autocatalytic reaction of protein splicing would be useful for protein engineering. These properties of the intein have been catching many researchers' attention, and publications related to inteins have been increasing dramatically.

Most inteins have the LAGLIDADG motif in their sequence, which is characteristic of a family of homing endonucleases (reviewed in 12,13). The endonucleases in this category make a double-stranded break in the DNA site, at each 12–40 base long specific recognition sequence, and eight inteins have been experimentally demonstrated to have this activity (11). Even though several regions, including the LAGLIDADG motif described above, are conserved among the inteins (9), the entire amino acid sequences of the inteins are not conserved, which is related to the fact that each intein can recognize a specific nucleotide sequence on the DNA, like the restriction endonucleases. To elucidate the structural basis for the ability of the homing endonucleases in the LAGLIDADG motif family to specifically recognize and cleave a DNA homing site, the crystal structures have been solved for PI-*Sce*I (14), I-*Cre*I (15) and I-*Dmo*I (16). Furthermore, the crystal structure of I-*Cre*I bound to its homing site DNA has been determined (17).

During the course of our screening for Holliday junction binding proteins from a hyperthermophilic archaeon, *Pyrococcus furiosus*, we identified a protein that can bind specifically to the junction DNA. Further analyses showed that the protein was one of the two inteins proposed from the sequence in the ribonucleotide reductase (RNR) precursor protein (18). In this report, we confirm that the two intein-like sequences are spliced out from the precursor protein produced in *Escherichia coli*. The biochemical properties of the purified inteins produced in *E.coli* have been characterized. Each intein has a heat stable endonuclease activity that cleaves double-stranded DNA at the intein-less allele. These inteins are now registered in InBase as PI-*Pfu*I and PI-*Pfu*II, respectively.

*To whom correspondence should be addressed. Tel: +81 6 6872 8208; Fax: +81 6 6872 8219; Email: ishino@beri.co.jp

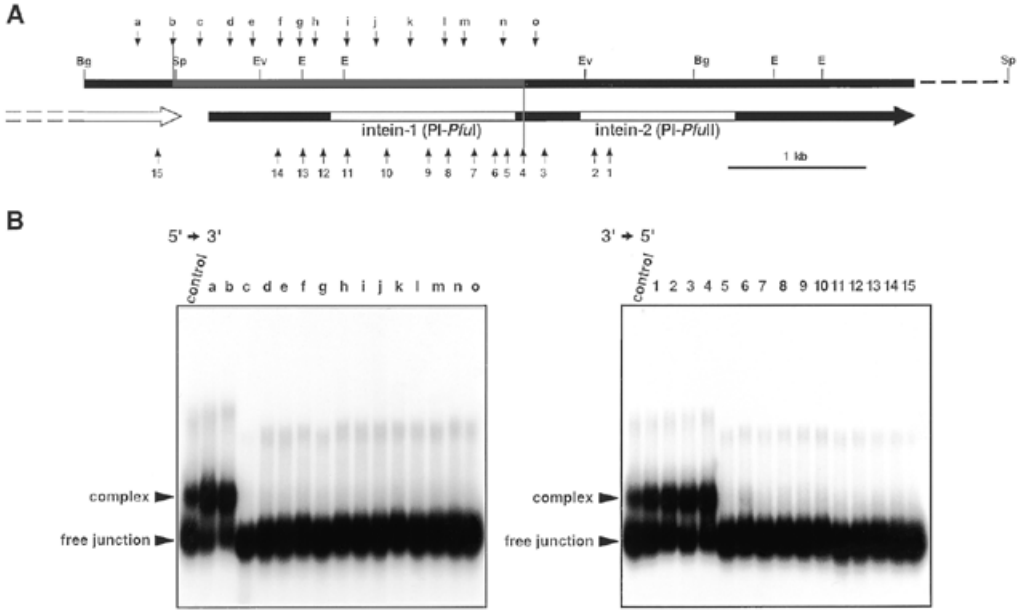

Figure 1. Determination of the region responsible for production of the Holliday junction binding protein. The 5′- and 3′-ends of the nested deletion mutants from each side are indicated by a–o and 1–15, respectively, in (**A**). Heat-treated cell extracts from each deletion mutant were subjected to the gel retardation assay using a ^{32}P-labeled synthetic four-way junction as probe. An autoradiograph of a 5% PAGE is shown in (**B**). The letters above the lanes correspond to the mutants indicated in (A). A heated extract of the clone having pFU45 was used as a positive control (lane control). Abbreviations in the restriction map: Bg, *Bgl*II; E, *Eco*RI; Ev, *Eco*RV; Sp, *Sph*I.

MATERIALS AND METHODS

Materials

Pyrococcus furiosus strain Vc1 (19) was obtained from Deutsche Sammlung von Mikroorganismen (Braunschweig, Germany). Cultivation of the strain, preparation of its DNA and construction of the cosmid-based library were as described earlier (20,21), except that the sonicated extracts were heated at 85°C instead of 100°C (22). *Escherichia coli* JM109, plasmids pUC18 and pUC118, restriction endonucleases, modification enzymes, PCR-related products and oligonucleotides were purchased from Takara Shuzo (Kyoto, Japan). *Escherichia coli* BL21(DE3) and pET21d were from Novagen (Madison, WI). [γ-^{32}P]ATP was obtained from NEN Life Science Products (Boston, MA).

Gel retardation assay for screening

Synthetic Holliday junctions, HJ (non-homologous sequence in the junction center) and HJm (homologous sequence in the core), and other forms of DNA including normal duplex, three-way junctions (3J and 3Jm), and looped-out DNA were constructed as described earlier using the same oligonucleotides (22). The reaction mixtures for the gel retardation assay (20 µl) contained 5 µl of heat-treated supernatant and ^{32}P-labeled DNA (0.1 pmol) in binding buffer (20 mM Tris–acetate, pH 8.0, 0.5 mM Mg acetate, 1 mM DTT, 100 µg/ml BSA, 5% glycerol) and were incubated at 55°C for 10 min, after which 5 µl of a gel loading buffer, containing 20 mM Tris–HCl, pH 8.0, 10% glycerol, and 0.1% bromophenol blue, was added. Fifteen microliters of each mixture were subjected to 5% polyacrylamide gel electrophoresis using TAM buffer (20 mM Tris–acetate, 0.5 mM Mg acetate, pH 7.8). The electrophoretic profiles were visualized by autoradiography.

DNA sequencing

Two plasmids, pFU45 and pFU60, in which the 4.5 kb *Bgl*II fragment and 6.0 kb *Sph*I fragment (Fig. 1) were inserted, respectively, into pUC118, were used for preparing nested deletion clones in both directions using the kilosequence deletion kit (Takara Shuzo), and the nucleotide sequences were analyzed with a fluorescent DNA sequencer, ABI 377A (PE Applied Biosystems, Foster City, CA).

Subcloning and expression of the intein genes

The regions encoding the inteins were amplified by PCR individually from pFU60. PCR primers corresponding to the translational initiation (F1 and F2) and termination regions (R1 and R2) were designed: F1 (5′-CGCACGAGCCATGGCCA-TAGACGGAAAGGCCAAG-3′) and R1 (5′-ACGCTGGATCCT-ATTAGTTGTGGACGAAAATCATT-3′) for intein-1 (PI-*Pfu*I); F2 (5′-CAGCCGTACCATGGCTGTCGTTGGAGACACT-AGAAT-3′) and R2 (5′-CGCATGGATCCTATTAGTTGTG-GCTCATGAAGCCG-3′) for intein-2 (PI-*Pfu*II). The F1 and F2 primers contain *Nco*I recognition sequences at the initiation codons, and R1 and R2 contain *Bam*HI sequences just after the termination codons of intein-1 and intein-2, respectively (underlined). PCR was performed in the standard reaction mixture optimized for *Pfu* DNA polymerase (Stratagene, La Jolla, CA), and the products were inserted into the pET21d expression vector after digestion with *Nco*I and *Bam*HI. The entire nucleotide sequences of the inserted DNAs were confirmed using the resultant plasmids, pFINT1 and pFINT2.

Purification of inteins

Escherichia coli BL21(DE3) carrying pFINT1 or pFINT2 was grown at 37°C with shaking in 500 ml of L broth containing 50 mg of ampicillin. When the A_{600} of the culture reached 0.6, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM, and the culture was incubated for a further 6 h. The cells were harvested and disrupted by sonication in buffer A, containing 50 mM Tris–HCl, pH 8.0, 2 mM 2-mercaptoethanol and 10% glycerol. The supernatant was incubated at 80°C for 15 min to denature and precipitate most of the *E.coli* proteins. The heat-treated supernatant was subjected to anion exchange chromatography (Resource Q; Amersham Pharmacia, Uppsala, Sweden) followed by affinity chromatography (Hitrap Heparin; Amersham Pharmacia). The purity of the target protein in each fraction was verified by SDS–PAGE, and the final fractions were quantified using a protein assay kit (Micro BCA Protein Assay Reagent; Pierce, Rockford, IL).

Protein splicing in *E.coli*

A part of the RNR gene, either an *Eco*RV fragment containing intein-1 and its vicinity or an *Eco*RI fragment containing intein-2 and its vicinity, was inserted into the pET21d vector with adjustment of the translational frame to produce the precursors. The resultant plasmids, pINT1EV and pINT2E, were introduced into *E.coli* BL21(DE3) and the target regions of the precursor protein were produced. Products were separated by 10% SDS–PAGE followed by electroblotting onto a polyvinylidene difluoride membrane. The blots were processed with an enhanced chemiluminescence system (Amersham Pharmacia) as described earlier (22).

Endonuclease assay

The oligonucleotide duplexes (40 bp in length) home 1 and home 2, with nucleotide sequences corresponding to the insertion sites of the genes for intein-1 and intein-2, respectively (20 bp from each end of the N- and C-terminal exteins), were used as the substrates for an endonuclease reaction after being 5′-labeled with ^{32}P. The 30 bp duplexes home 1′, with 15 bp each of the N- and C-terminal exteins, and home 2′, with 20 bp of the N- and 10 bp of the C-terminal exteins, were also used. Alternatively, these oligonucleotides were inserted into pUC118 at the *Hinc*II site and the resultant plasmids, p140 and p140C (both directions for PI-*Pfu*I), and p240 and p240C (those for PI-*Pfu*II), were directly used for digestion reactions. The 50 µl cleavage reaction mixture contained 10 mM Tris–HCl, pH 8.8, 10 mM $MgCl_2$, 1 mM DTT, 150 mM KCl and 1 pmol of the oligonucleotide substrate (20 nM) or 0.5 µg of the plasmid substrate (76 nM), and was incubated at either 55°C (for oligonucleotide substrates) for 30 min or 80°C (for plasmid substrates) with endonucleases (0.4 µM) for 5 min. The mixture was analyzed by 6% PAGE followed by autoradiography or by 1% agarose gel electrophoresis with ethidium bromide staining. Phenol/chloroform extraction was necessary before loading the reaction mixture onto the gels, because the DNA band was shifted when the reaction products were applied directly.

Determination of the cleavage site

To determine the exact cleavage positions of PI-*Pfu*I and PI-*Pfu*II in the recognition sequence, a primer extension method was used as described earlier (23). Single-stranded DNAs prepared from p140, p140C, p240 and p240C were annealed with a ^{32}P-labeled universal sequencing primer (M4; Takara Shuzo) that can bind to the position upstream of the multicloning site of pUC118, and were divided into two tubes. Primer extension reactions were performed using *Bca*BEST DNA polymerase (Takara Shuzo). The reaction products were then digested with PI-*Pfu*I or PI-*Pfu*II, and were loaded on a 6% polyacrylamide gel with 8 M urea together with the primer extension products with dideoxynucleotide triphosphates from the same primer-templates. For delineation of the minimal cleavage sequences, another primer extension method was done as described earlier (24).

RESULTS

Cloning of the genes for PI-*Pfu*I and PI-*Pfu*II

In order to identify the proteins in *P.furiosus* cells that can bind specifically to Holliday junction DNA, we employed a strategy that involved screening for heat stable binding activities to a synthetic Holliday junction from the *P.furiosus* protein library. Five hundred independent *E.coli* clones, transformed by a cosmid-based library of *P.furiosus* genomic DNA, were cultivated, and cell-free extracts were incubated at 85°C for 10 min to inactivate most of the *E.coli* proteins. The resultant heated supernatants were subjected to gel retardation assay using a synthetic Holliday junction (HJ) as probe. In this way, 10 clones with binding activity were obtained. To characterize the *P.furiosus* genomic DNA regions inserted into the cosmids, a Southern blot analysis was done, which showed that all of the clones contained the same restriction fragment (data not shown). Then, common *Bam*HI fragments were excised from the cosmid vector and digested with *Bgl*II. Each fragment was cloned into the *Bam*HI site of the pUC118 vector. Gel retardation assays using these subclones showed that a 4.5 kb *Bgl*II fragment (pFU45) was responsible for the specific binding activity (data not shown). Nested deletion clones in both directions were constructed, and the entire nucleotide sequence of the fragment was determined. From the nucleotide sequence analysis, the region we cloned was matched to that including a part of the RNR gene published earlier (18). Two inteins were predicted to exist from the deduced amino acid sequence of the gene, and the gel retardation assay using the deletion clones indicated that the region encoding the upstream intein (intein-1) was responsible for the production of a specific Holliday junction binding protein (Fig. 1).

Overproduction of PI-*Pfu*I and PI-*Pfu*II

In order to analyze the biochemical properties of the intein-like proteins, we made overexpression systems for the regions for intein-1 and intein-2. The expected proteins were highly produced in *E.coli* BL21(DE3) cells with IPTG induction. The two recombinant proteins were heat stable, and heating the sonicated cell extracts at 80°C for 15 min was an effective step for their purification. Two sequential column chromatographies were sufficient to purify the two proteins to near homogeneity (Fig. 2). The amounts of purified intein-1 and intein-2 from 500 ml cultures of each recombinant *E.coli* were 4.0 and 4.5 mg, respectively.

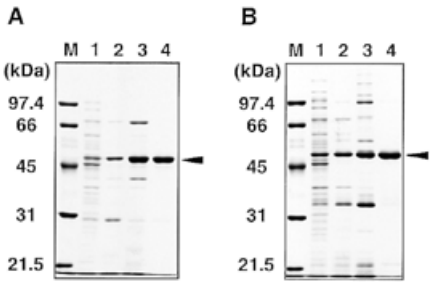

Figure 2. Purification of intein-1 (PI-*Pfu*I) (**A**) and intein-2 (PI-*Pfu*II) (**B**). The two inteins produced in *E.coli* were purified, and fractions from the different stages (5 µg each) were analyzed by 10% SDS–PAGE. Lane 1, supernatant after sonication; lane 2, supernatant after heat treatment (80°C for 15 min); lane 3, Resource Q column chromatography; lane 4, HiTrap Heparin column chromatography. Lane M contains size marker proteins (Bio-Rad). The gels were stained with Coomassie brilliant blue. The two inteins are indicated by arrowheads.

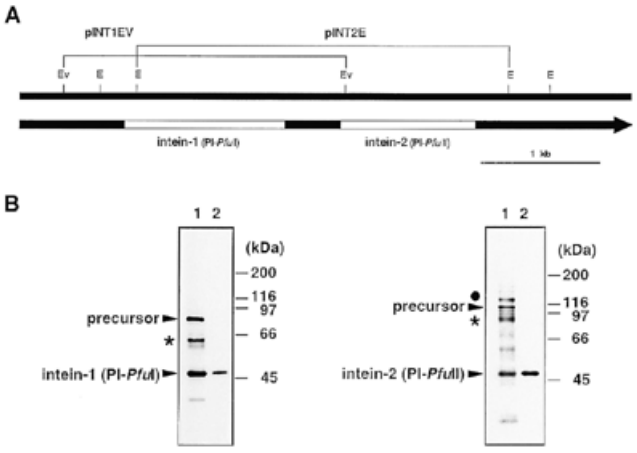

Figure 3. Identification of protein splicing in *E.coli*. DNA fragments containing the intein-1 (PI-*Pfu*I) gene and its flanking regions or the intein-2 (PI-*Pfu*II) gene and its flanking regions were cloned into the expression vector (**A**), and the resultant plasmids, pINT1EV and pINT2E, were introduced into *E.coli*. The cell extracts of each transformant were subjected to 10% SDS–PAGE (lane 1). Purified inteins were loaded in lane 2. The proteins were detected by western blotting (**B**). The closed circle and asterisk show the branched intermediate and the product from cleavage at only one splice junction, respectively, as predicted from their mobilities.

Protein splicing in *E.coli*

To determine whether the two intein-like sequences in the RNR precursor have protein splicing ability, the plasmids pINT1EV and pINT2E, expressing a portion of the RNR precursor, were constructed. In each of the plasmids, a gene fragment encoding one of the inteins and its vicinity was inserted (Fig. 3A). These plasmids were introduced into *E.coli*, and the produced proteins were analyzed by western blotting using antisera against the purified inteins. As shown in Figure 3B, bands corresponding to intein-1 and intein-2 and their precursors were observed in both cases. The precursor proteins were more strongly detected when the *E.coli* strains were cultured at 20 instead of 37°C (data not shown). The N-terminal amino acid sequences (nine residues each) of the bands were completely matched to the deduced sequences of intein-1 and intein-2 (the N-terminal cysteine was not detected). This result shows that both proteins are actually inteins that have autocatalytic splicing activity. These two proteins were also detected in the crude extract of *P.furiosus* cells by western blotting, which means that both intein-1 and intein-2 are produced in *P.furiosus* cells, possibly by protein splicing. Bands of precursors and splicing intermediates were not detected (data not shown).

Identification of the site-specific endonuclease activity

The purified inteins were subjected to an endonuclease assay. In general, intein-coded endonucleases recognize an intein-less (junction of the two extein genes) sequence (12,13). The target sizes for the reported homing endonucleases in this category vary, ranging from 12 to 40 bp; therefore, we prepared 40 bp duplex DNAs, home 1 and home 2, including the junction of the two extein genes (described in Materials and Methods) as substrates for the endonuclease assay. These substrates were cleaved by the inteins at one site, as shown in Figure 4A, although they were not completely cleaved. Then, home 1 and home 2 were inserted into the pUC118 plasmid. The resultant plasmids, p140 and p240, were cleaved almost completely by each enzyme, as shown in Figure 4B. Neither of the two inteins cleaved the plasmid containing the DNA fragment that includes the intein gene-inserted junctions (data not shown). These two inteins were designated PI-*Pfu*I and PI-*Pfu*II, respectively, in accordance with the proposed rules for nomenclature of intein-coded homing endonucleases (3). The dependency of cleavage efficiency on the length of the substrate DNA shows that the structural and torsional environments of the recognition sites are important for PI-*Pfu*I and PI-*Pfu*II, as reported for PI-*Sce*I (25).

As shown in Figure 5, a divalent cation was essential for their cleavage activities. Less than 10 mM $MgCl_2$ was optimal for both endonucleases, as shown below (Fig. 7). Manganese, but neither calcium nor zinc, could substitute for magnesium. As reported for PI-*Sce*I (26), an extra site in pUC118 was cleaved inefficiently in the presence of Mn^{2+} (Fig. 5). This relaxation of substrate recognition, which is the so-called star activity of type II restriction endonucleases (24), was also observed in the PI-*Pfu*I reaction in the absence of KCl (described below).

Determination of the cleavage site and delineation of the recognition sequence

To determine the exact position cleaved by PI-*Pfu*I and PI-*Pfu*II in their recognition sequences, we performed primer extension experiments, as described in Materials and Methods. Both enzymes cleaved the DNA to produce a 3′-overhang of four bases, as shown in Figure 6. The cleavage position of PI-*Pfu*I includes the junction of the two extein genes. In contrast, that of PI-*Pfu*II is within the gene for the C-terminal extein. The generated ends were a 5′-phosphate and a 3′-OH, as confirmed by ligation-recutting of the products (data not shown). To delimit the recognition sequences, we tried a primer extension method, as described (24), and expected that both enzymes needed at least 30 nt for cleavage (data not shown). A footprinting experiment of PI-*Pfu*I also showed that 30 bases around the cleavage site were protected from DNase I digestion in both strands (data not shown). Therefore, we made oligonucleotides with 30 bp containing the extein junctions to confirm the

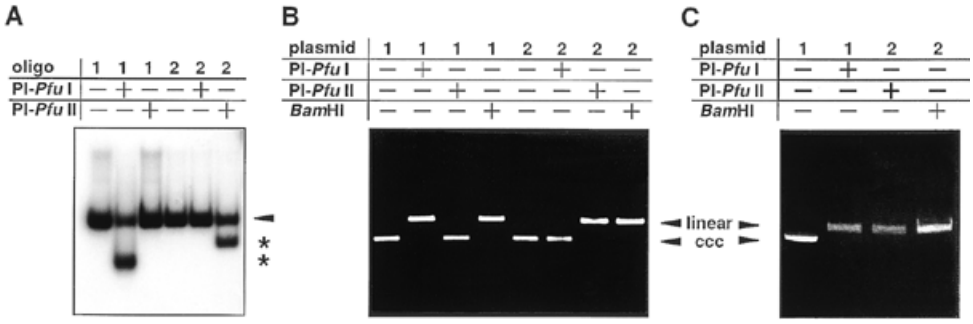

Figure 4. Endonuclease activity of intein-1 (PI-*Pfu*I) and intein-2 (PI-*Pfu*II). The 40 bp oligonucleotide duplexes home 1 and home 2 were directly used as substrates after labeling with ^{32}P (**A**), or were used after insertion into plasmid pUC118 (**B**). The 30 bp oligonucleotides with their recognition sequences (home 1′ and home 2′) were used after insertion into the same plasmid (**C**). Substrate 1 indicates home 1 (A), p140 (B) and home 1′ (C), and substrate 2 indicates home 2 (A), p240 (B) and home 2′ (C), respectively.

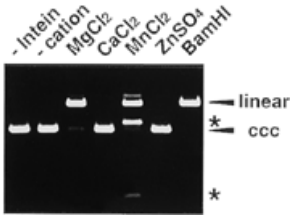

Figure 5. Effect of divalent cations on the endonuclease activities of PI-*Pfu*I. Purified PI-*Pfu*I was dialyzed against buffer without divalent cations and was subjected to the endonuclease assay with various cations using plasmid p140 as the substrate. The reactions containing the DNA substrate (76 nM) and the enzyme (0.2 μM) were analyzed by 1.0% agarose gel electrophoresis. Arrows and stars indicate the linear plasmid cleaved at the recognition site and the products of the star activity, respectively.

minimum size for recognition/cleavage. These d30mers were digested by PI-*Pfu*I and PI-*Pfu*II only when they were inserted into the plasmid, as shown in Figure 4C, but were not digested by themselves. The flanking sequences of the plasmid at the insertion site of the d30mer are completely different from those in home 1, but one base matches that in home 2. These results show that at least 30 and 31 nt sequences including the extein–extein junctions can be recognized for cleavage by PI-*Pfu*I and PI-*Pfu*II, respectively.

Optimal reaction conditions for PI-*Pfu*I and PI-*Pfu*II

Optimal conditions for the endonuclease activity of PI-*Pfu*I and PI-*Pfu*II were determined using the plasmids p140 and p240, respectively, as DNA substrates (Fig. 7). Both endonucleases worked over a wide range of temperature, pH and concentrations of $MgCl_2$ and KCl, but PI-*Pfu*II required more stringent conditions than PI-*Pfu*I. PI-*Pfu*I was more heat resistant than PI-*Pfu*II. Maximum activities for PI-*Pfu*I and PI-*Pfu*II were obtained at 80–100 and 80–90°C, respectively. PI-*Pfu*I was much more tolerant of changes in pH and salt concentration. A neutral pH, 1–10 mM $MgCl_2$ and <200 mM KCl were clearly suitable for PI-*Pfu*II. Under acidic conditions, open circular DNAs nicked at one strand were detected, in contrast to the highly concerted cleavage of the two strands under the optimal conditions for both PI-*Pfu*I and PI-*Pfu*II. During these experiments we found that PI-*Pfu*I cleaved one additional site in p140 at a low concentration of KCl. The sizes of the cleaved products were the same as those from the reaction with $MnCl_2$, as mentioned above. Therefore, we determined the cleavage site in pUC118 by PI-*Pfu*I by the same primer extension method, after estimating the position by the digestion patterns of several restriction enzymes. The determined recognition sequence was substantially similar to the original recognition sequence of PI-*Pfu*I, as shown below (Fig. 10).

Specific binding of PI-*Pfu*I to Holliday-structured DNA

We cloned the gene for PI-*Pfu*I by specific binding to Holliday-structured DNA, as described above. As shown in Figure 8, PI-*Pfu*I, but not PI-*Pfu*II, bound to the two different Holliday-structured DNAs (HJ and HJm). PI-*Pfu*I also bound to the three-way junction. However, it did not bind to the looped-out DNA molecule containing 10 unpairable bases. Double-stranded DNAs containing 1 base mismatch or 1 base deletion were not substrates of PI-*Pfu*I for specific binding (data not shown). To investigate whether PI-*Pfu*I recognizes the nucleotide sequence of one of the four arms in the Holliday junction, duplex DNAs with the sequences of each of the four oligonucleotides constituting the Holliday junction were used for a gel retardation experiment; however, no specific binding was observed (data not shown). To confirm the specificity of the PI-*Pfu*I binding ability, competition experiments were done. As shown in Figure 9, binding of PI-*Pfu*I to the Holliday junction or to the double-stranded DNA home 1 was inhibited only by these two DNAs and not by a duplex homing site, home 2, of PI-*Pfu*II. The binding of PI-*Pfu*I to its homing site was inhibited by an excess amount of the Holliday junction, but not by a non-specific duplex DNA. These results clearly show that PI-*Pfu*I has a specific affinity for Holliday junction DNA. The reason why complex formation did not decrease with an equimolar probe/competitor is probably that the protein concentrations were higher enough than probe concentrations in these experiments to not be affected by the competitors. The apparent dissociation constants (K_d) of PI-*Pfu*I obtained from the gel retardation assay were 180 and 160 nM for HJ and HJm, respectively. These values indicate that the affinities of PI-*Pfu*I for Holliday junctions are one order of magnitude lower than that for the duplex homing site (home 1), which

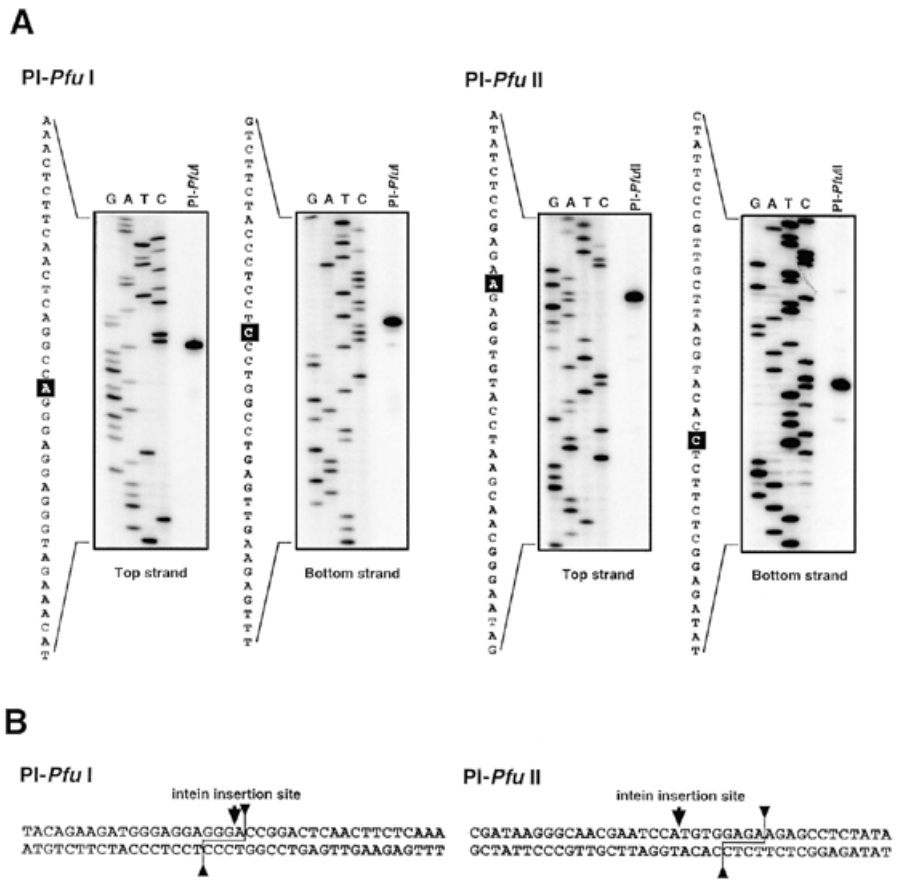

Figure 6. Determination of the cleavage sites of PI-*Pfu*I and PI-*Pfu*II in their recognition sequences. Gel electrophoresis of the primer extension reaction, followed by PI-*Pfu*I or PI-*Pfu*II digestion is shown in (**A**). The cleavage positions were determined by comparing the product bands with sequence ladders obtained in parallel. The recognition sequences and the cleavage positions (indicated by arrowheads) of PI-*Pfu*I and PI-*Pfu*II are shown in (**B**).

gave a K_d value of 11 nM, as shown in the accompanying paper (27). A weaker competition of HJ or HJm, but not of home 1, with home 2 for PI-*Pfu*II binding was observed in the competition experiment shown in Figure 9, which indicates that PI-*Pfu*II may also prefer to bind to a Holliday junction rather than to normal duplex DNA, even though direct binding of PI-*Pfu*II to HJ or HJm was not observed in the gel retardation assay.

DISCUSSION

We have determined that the two inteins found in *P.furiosus* have two functions: splicing in *E.coli* and acting as homing-type endonucleases. Both inteins, which were named PI-*Pfu*I and PI-*Pfu*II based on their endonuclease activities, recognize about 30 bases in their intein-less (homing) sequences and cleave double-stranded DNA. Due to their rare cutting properties, PI-*Pfu*I and PI-*Pfu*II are useful enzymes for genome analyses and gene manipulation including genome mapping, construction of cloning vectors, and gene targeting, like other homing endonucleases (reviewed in 13).

Both endonucleases are highly heat stable, and they are suitable for studies of structure–function relationships and also for practical applications. Recently, a useful technique for protein engineering using *trans*-splicing was proposed. Southworth *et al.* (28) showed that protein splicing occurs from splitting the *Psp* Pol-1 intein precursor proteins, which were produced in separate *E.coli* hosts. We used PI-*Pfu*I for *trans*-splicing independently, and succeeded in segmental isotope labeling of the C-terminal domain of the *E.coli* RNA polymerase α subunit with ^{15}N (29). This technique is especially advantageous for protein structure analysis by NMR spectroscopy.

This is the first report of an intein with binding specificity for Holliday junction DNA. This finding is interesting in relation to a model mechanism for the transfer of an intein gene or intron by the double-strand break–repair pathway through a Holliday junction intermediate (reviewed in 12). However, it is not known whether all (or most) of the homing endonucleases are endowed with this specific binding ability. No binding of PI-*Pfu*II to the Holliday junction was detected in the gel retardation assay under the same conditions used for PI-*Pfu*I. The HMG box proteins are known to show structure-selective binding of four-way DNA junctions (30,31). The K_d value of PI-*Pfu*I binding to four-way junctions is comparable to those of the HMG box proteins published elsewhere (32). However, Pohler *et al.* (32) suggested that recognition of a DNA junction by HMG proteins may be due to an accidental similarity in the DNA structure to the normal binding site on the chromosomal DNA, because the HMG proteins prefer to bind to the open square form of the junction. The binding of PI-*Pfu*I to four-way junction DNA decreased with an increasing amount of Mg^{2+},

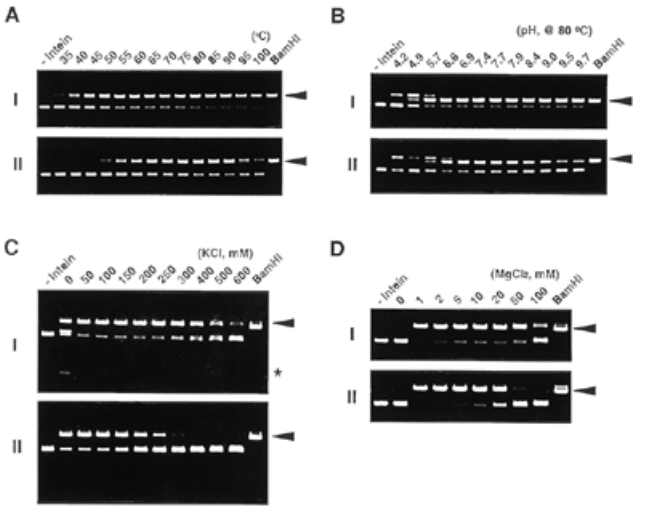

Figure 7. Effect of temperature, pH, ionic strength and divalent cation concentration on the endonuclease activity of PI-*Pfu*I and PI-*Pfu*II. Except for the variables indicated, the cleavage reactions were done using the p140 and p240 plasmids as substrates under the standard assay conditions. I and II indicate PI-*Pfu*I and PI-*Pfu*II, respectively, in (**A**) (temperature), (**B**) (pH), (**C**) (KCl concentration) and (**D**) ($MgCl_2$ concentration). The concentrations of the substrate and the enzyme in the reaction mixture were the same as those in Figure 5. *Bam*HI-digested p140 or p240 was loaded as a control for the digested linear DNA product in all panels. The products derived from the star activity of PI-*Pfu*I are indicated by an asterisk.

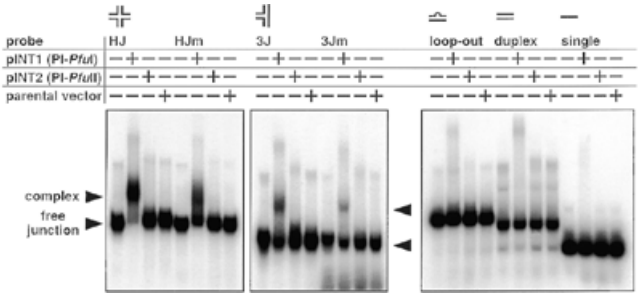

Figure 8. DNA binding specificities of PI-*Pfu*I and PI-*Pfu*II. The DNA binding abilities of PI-*Pfu*I and PI-*Pfu*II were investigated using various forms of synthetic DNAs. The conditions for the binding reaction and the gel electrophoresis are described in Materials and Methods.

which changes the four-way junction from the open square to the stacked X structure (data not shown). The binding of PI-*Pfu*I to the junction DNA may also be accidental, because of the similarity of the bent structure of the substrate to the homing endonuclease, as analyzed in detail for PI-*Sce*I (25,33). We performed a circular permutation analysis and showed that the substrate DNA bends when it binds to PI-*Pfu*I (27). To determine whether specific binding of the homing endonuclease to the Holliday junction has biological significance, the binding mode and its preferred conditions must be analyzed in more detail.

The phylogeny of the inteins is also interesting, as described earlier (5,34). Recently, the complete genome sequence of

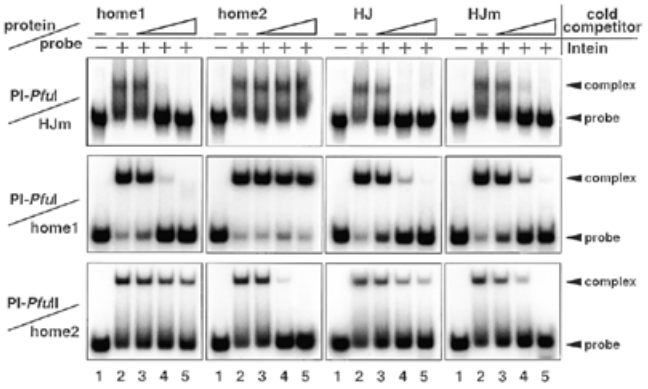

Figure 9. Competition analysis of DNA binding abilities. PI-*Pfu*I or PI-*Pfu*II (40 nM) and ^{32}P-labeled DNA probes (5 nM) were incubated with various unlabeled DNAs, and the bound complexes were detected by gel retardation assay. In each panel: lane 1, ^{32}P-labeled probe; lane 2, ^{32}P-labeled probe and protein indicated at the left side; lanes 3–5, unlabeled DNAs indicated at the top were added as competitors at 5, 50 and 500 nM, respectively.

Pyrococcus horikoshii OT3 became available in a database (10). The amino acid sequence of RNR and the nucleotide sequence of its genes in *P.horikoshii* are 93 and 82% identical, respectively, to those in *P.furiosus.* The intein corresponding to PI-*Pfu*II was found in the RNR of *P.horikoshii* at exactly the same position as that of *P.furiosus*, and their sequences have 73% identity with each other. The sequences of the intein insertion sites are almost the same (27 of 30 bases are identical) between the two species. There is no intein in *P.horikoshii* corresponding to PI-*Pfu*I. This fact indicates that *P.furiosus* acquired PI-*Pfu*I through horizontal transfer from another source, after divergence of the two pyrococcal species. It is also possible that *P.horikoshii* lost one of the inteins in the RNR gene after divergence. A comparison of the recognition sequence of PI-*Pfu*I with the corresponding intein-less sequence of *P.horikoshii* OT3 shows that 19 out of 30 bases are identical (Fig. 10). PI-*Pfu*I may digest the *P.horikoshii* genome at this site, and horizontal transfer of this intein gene may occur between the two species, because the pUC18 plasmid is digested by PI-*Pfu*I at a sequence with 14 out of 30 bases identical to its recognition sequence under some conditions, as described above. Therefore, these two strains may be useful for analyzing the molecular mechanism of the intein homing event. The sequences corresponding to the PI-*Pfu*I insertion site in the RNR genes from *Archaeoglobus fulgidus* and *Thermoplasma acidophilum* are also very similar. In their sequences, 23 and 20 out of 30 bases are identical, respectively (Fig. 10). The genes may also be cleaved by PI-*Pfu*I. In addition, the sequences of the RNR genes from *A.fulgidus* and *T.acidophilum* corresponding to the region of the PI-*Pfu*II insertion site are also very similar, and 19 and 21 out of 30 bases are identical, respectively (Fig. 10). PI-*Pfu*II may cleave these sites in the genomes of these organisms. Homing of the yeast VMA intein into an empty *VMA1* gene (*VMA1*Δ*vde*) is the only example that has been proven *in vivo* thus far (35).

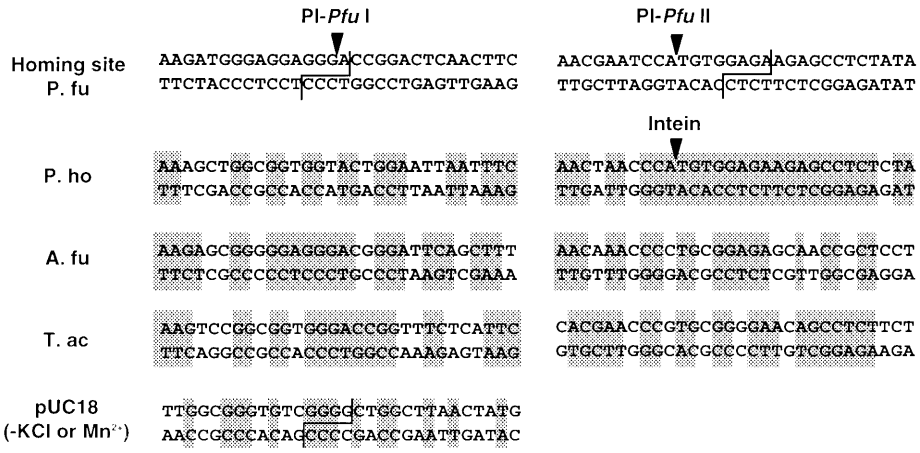

Figure 10. Comparison of the nucleotide sequences of the regions corresponding to insertion of the intein in the ribonucleotide reductase genes found in Archaea. P. fu, *P.furiosus*; P. ho, *Pyrococcus horikoshii*; A. fu, *Archaeoglobus fulgidus*; T. ac, *Thermoplasma acidophilum*. The sequence in pUC118 that can be recognized and cleaved by the star activity of PI-*Pfu*I (with Mn^{2+} or in the absence of KCl) is also shown at the bottom. The sequences identical to the authentic recognition sequence are shadowed.

REFERENCES

1. Shub,D.A. and Goodrich-Blair,H. (1992) *Cell*, **71**, 183–186.
2. Cooper,A.A. and Stevens,T.H. (1993) *Bioessays*, **15**, 667–674.
3. Perler,F.B., Davis,E.O., Dean,G.E., Gimble,F.S., Jack,W.E., Neff,N., Noren,C.J., Thomer,J. and Berfort,M. (1994) *Nucleic Acids Res.*, **22**, 1125–1127.
4. Anraku,Y. (1997) *Genes Cells*, **2**, 359–367.
5. Perler,F.B., Olsen,G.J. and Adam,E. (1997) *Nucleic Acids Res.*, **25**, 1087–1093.
6. Perler,F.B. (1998) *Cell*, **92**, 1–4.
7. Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D., Sutton,G.G., Blake,J.A., FitzGerald,L.M., Clayton,R.A., Gocayne,J.D. *et al*. (1996) *Science*, **273**, 1058–1073.
8. Smith,D.R., Doucette-Stamm,L.A., Deloughery,C., Lee,H.-M. Dubois,J., Aldredge,T., Bashirzadeh,R., Blakely,D., Cook,R., Gilbert,K. *et al*. (1997) *J. Bacteriol*., **179**, 7135–7155.
9. Klenk,H.P., Clayton,R.A., Tomb,J., White,O., Nelson,K.E., Ketchum,K.A., Dodson,R.J., Gwinn,M., Hickey,E.K., Peterson,E.J. *et al*. (1997) *Nature*, **390**, 364–370.
10. Kawarabayasi,Y., Sawada,M., Horikawa,H., Haikawa,Y., Hino,Y., Yamamoto,S., Sekine,M., Baba,S., Kosugi,H., Hosoyama,A. *et al*. (1998) *DNA Res.*, **5**, 55–76.
11. Perler,F.B. (1999) *Nucleic Acids Res.*, **27**, 346–347.
12. Mueller,J.E., Bryk,M., Loizos,N. and Belfort,M. (1993) In Linn,S.M., Lloyd,R.S. and Roberts,R.J. (eds), *Nucleases*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 111–143.
13. Belfort,M. and Roberts,R.J. (1997) *Nucleic Acids Res.*, **25**, 3379–3388.
14. Duan,X., Gimble,F.S. and Quiocho,F.A. (1997) *Cell*, **89**, 555–564.
15. Heath,P.J., Stephens,K.M., Monnat,R.J.,Jr and Stoddard,B.L. (1997) *Nature Struct. Biol.*, **4**, 468–476.
16. Silva,G.H., Dalgaard,J.Z., Belfort,M. and Roey,P.V. (1998) *J. Mol. Biol.*, **286**, 1123–1136.
17. Jurica,M.S., Monnat,R.J.,Jr and Stoddard,B.L. (1998) *Mol. Cell*, **2**, 469–476.
18. Riera,J., Robb,F.T., Weiss,R. and Fontecave,M. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 475–478.
19. Fiala,G. and Stetter,K.O. (1986) *Arch. Microbiol.*, **145**, 56–61.
20. Uemori,T., Ishino,Y., Toh,H., Asada,K. and Kato,I. (1993) *Nucleic Acids Res.*, **21**, 259–265.
21. Uemori,T., Sato,Y., Kato,I., Doi,H. and Ishino,Y. (1997) *Genes Cells*, **2**, 499–512.
22. Komori,K., Sakae,S., Shinagawa,H., Morikawa,K. and Ishino,Y. (1999) *Proc. Natl Acad. Sci. USA*, **96**, 8873–8878.
23. Brown,N.L. and Smith,M. (1980) *Methods Enzymol.*, **65**, 391–404.
24. Wenzlau,J.M., Saldahna,R.J., Butow,R.A. and Perlman,P.S. (1989) *Cell*, **56**, 421–430.
25. Wende,W., Grindle,W., Christ,F., Pingoud,A. and Pingoud,V. (1996) *Nucleic Acids Res.*, **24**, 4123–4132.
26. Gimble,F.S. and Thorner,J. (1993) *J. Biol. Chem.*, **268**, 21844–21853.
27. Komori,K., Ichiyanagi,K., Morikawa,K. and Ishino,Y. (1999) *Nucleic Acids Res.*, **27**, 4175–4182.
28. Southworth,M.W., Adam,E., Panne,D., Byer,R., Kautz,R. and Perler,F.B. (1998) *EMBO J.*, **17**, 918–926.
29. Yamasaki,T., Otomo,T., Oda,N., Kyogoku,Y., Uegaki,K., Ito,N., Ishino,Y. and Nakamura,H. (1998) *J. Am. Chem. Soc.*, **120**, 5591–5592.
30. Bianchi,M.E. (1995) In Lilley,D.M.J. (ed.), *DNA–Protein: Structural Interactions*. IRL Press, Oxford, UK, pp. 177–200.
31. Read,C.M., Cary,P.D., Crane-Robinson,C., Driscoll,P.C., Carrillo,M.O.M. and Norman,D.G. (1995) In Eckstein,F. and Lilley,D.M.J. (eds), *Nucleic Acids and Molecular Biology*. Springer-Verlag, Berlin, Germany, Vol. 9, pp. 222–250.
32. Pohler,J.R.G., Norman,D.G., Bramham,J., Bianchi,M.E. and Lilley,D.M.J. (1998) *EMBO J.*, **17**, 817–826.
33. Gimble,F.S. and Wang,J. (1995) *J. Mol. Biol.*, **263**, 163–180.
34. Chute,I.C., Hu,Z. and Liu,X.-Q. (1998) *Gene*, **210**, 85–92.
35. Gimble,F.S. and Thorner,J. (1992) *Nature*, **357**, 301–306.