

I-Apel: a novel intron-encoded LAGLIDADG homing endonuclease from the archaeon, *Aeropyrum pernix* K1

Norimichi Nomura*, Yayoi Morinaga, Nobuaki Shirai¹ and Yoshihiko Sako

Division of Applied Biosciences, Graduate School of Agriculture, Kyoto University, Kyoto 606-8502, Japan and ¹Industrial Research Center of Shiga, Ritto-shi, Shiga 520-3004, Japan

Received May 8, 2005; Revised and Accepted June 13, 2005

ABSTRACT

Over 50 introns have been reported in archaeal rRNA genes (rDNAs), a subset of which nests putative homing endonuclease (HEase) genes. Here, we report the identification and characterization of a novel archaeal LAGLIDADG-type HEase, I-Apel, encoded by the ApeK1.S908 intron within the 16S rDNA of *Aeropyrum pernix* K1. I-Apel consists of 222 amino acids and harbors two LAGLIDADG-like sequences. It recognizes the 20 bp non-palindromic sequence 5'-GCAAGGCTGAAAC↓TTAAAGG and cleaves target DNA to produce protruding tetranucleotide 3' ends. Either Mn²⁺ or Co²⁺ can be substituted for Mg²⁺ as a cofactor in the cleavage reaction. Of the 20 bases within the minimal recognition site, 7 are essential for cleavage and are located at positions proximal to the cleavage sites.

INTRODUCTION

Homing endonucleases (HEases) are a diverse collection of site-specific endonucleases that recognize and cleave sequences of 14–40 DNA base pairs and tolerate various degrees of degeneracy of their target sequences (1–5). *In vivo* they play a pivotal role as 'target selector proteins' in the homing of mobile introns and inteins in the genomes of bacteriophages, bacteria, archaea and unicellular eukaryotes. HEases are encoded within the mobile elements themselves, except for freestanding enzymes, such as F-SceII (HO endonuclease) and F-TevI (SegA). The hitherto identified HEases are grouped into four families, LAGLIDADG, GIY-YIG, His-Cys and HNH, on the basis of their conserved motifs. The LAGLIDADG family is the largest of the four families with at least

200 members (see REBASE at <http://rebase.neb.com/cgi-bin/azlist?homing>).

Much interest has focused on the target specificity of these HEases due to their potential as tools for genome mapping, cloning of megabase DNA fragments and gene targeting (6). Searches for naturally occurring HEases that recognize novel sequences remain important, though a few artificial rare-cutting enzymes with novel specificity have also been created (7–9). Archaea are promising sources of novel HEases, because their rRNA genes (rDNAs) are interspersed with a number of mobile introns, ~40% of which contain open reading frames (ORFs) encoding LAGLIDADG-like sequences (10–13).

Here, we report the identification and characterization of a new intron-encoded LAGLIDADG-type HEase, I-Apel, from the hyperthermophilic crenarchaeote *Aeropyrum pernix* K1. We also describe the effects on cleavage of single base pair substitutions within the I-Apel recognition sequence.

MATERIALS AND METHODS

Strains and growth conditions

A. pernix strain K1 was grown as described previously (14). *Escherichia coli* INV α F' (Invitrogen) was grown in liquid or in solid Luria-Bertani (LB) medium supplemented with ampicillin and used for plasmid manipulations; *E. coli* BL21(DE3) pLysS was used as the host for protein overexpression.

Plasmid construction

Plasmids containing wild-type and mutant target sites. The nomenclature of archaeal rDNA introns is based on a proposal by Morinaga *et al.* (12). The DNA base pairs around the insertion site (IS) of the ApeK1.S908 intron are named as shown in Figures 3A and 5. The wild-type target DNA

*To whom correspondence should be addressed. Tel: +81 75 753 6224; Fax: +81 75 753 6226; Email: nom@kais.kyoto-u.ac.jp

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

© The Author 2005. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oupjournals.org

spanning positions -12 to +12 was generated by annealing the 25 bp complementary oligonucleotides 5'-CGCAAGGCT-GAA*ACTTAAAGGAATA and 5'-ATTCCTTTAAGT*-TTCAGCCTTGCGA (asterisk indicates the IS of the ApeK1.S908 intron). The duplex was ligated into a pCR2.1 vector (Invitrogen) to yield pWT. A series of pWT derivatives with mutated target sites were prepared by cloning appropriate synthetic oligonucleotide duplexes into pCR2.1. All the cloned inserts were verified by sequencing both strands.

Expression plasmid. The I-ApeI gene was PCR amplified from *A. pernix* K1 chromosomal DNA with primers apeI-For (5'-CATATGGATGGGTATGCTTTATTTT) and apeI-Rev (5'-GGATCCTTAAGCGGCAGGCTCCACCCCT). The product was cut at its NdeI and BamHI sites (underlined) and ligated into the same sites in pET-15b (Novagen) to yield plasmid pIA1. The insert was then sequenced to eliminate PCR artifacts.

Protein expression and purification

E. coli BL21(DE3) pLysS was transformed with pIA1, and the transformants were grown at 37°C in LB broth supplemented with 50 µg/ml ampicillin to an optical density at 600 nm of 0.4. Expression of His₆I-ApeI was induced by adding isopropyl-β-D-thiogalactopyranoside to a concentration of 0.5 mM and continuing incubation for another 3 h at 37°C. The culture was then chilled on ice, and the cells were harvested by centrifugation at 4°C. The cell pellet was washed and resuspended in ice-cold buffer A (50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 0.1 mM EDTA, 10% glycerol and 0.5 mM DTT), and all subsequent steps were performed at 4°C. The cells were disrupted by sonication, and the crude lysate was centrifuged to remove intact cells and cellular debris. The resulting supernatant was heated at 60°C for 30 min, and thermally denatured proteins were removed by centrifugation. The supernatant was adjusted to a final ammonium sulfate concentration of 20% saturation and centrifuged. Ammonium sulfate was then added to the supernatant to a final concentration of 50% saturation, and the precipitate was pelleted, dissolved in buffer A and dialyzed against the same buffer. The dialysate was fractionated on a HiTrap SP column (Amersham) followed by a HiTrap heparin 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. The final fractions containing His₆I-ApeI were pooled and loaded onto a Superdex 75 pg column (Amersham) equilibrated with buffer A, and the purified protein was flash frozen and stored at -80°C.

Endonuclease assays

The double-stranded DNA endonuclease activity of I-ApeI was assayed using linearized plasmids. Plasmid pWT and its derivatives were linearized with NcoI, which cleaves the vector DNA 1.6 kb from the I-ApeI target site. The resulting 4.0 kb fragments were extracted with phenol-chloroform and precipitated with ethanol. This substrate DNA (0.1 pmol) was incubated with I-ApeI (10 pmol) in a total volume of 10 µl of 10 mM Tris-HCl (pH 7.5 at 25°C), 180 mM NaCl, 10 mM MgCl₂ and 1 mM DTT at 90°C for 10 min. Unless otherwise stated, reactions were performed under these standard conditions. The assay mixtures were incubated in an iCycler thermal cycler (Bio-Rad) without oil sample overlays. The reactions

were terminated by placing the samples on ice and adding 2 µl of loading buffer containing 0.25% bromophenol blue, 1 mM EDTA and 30% glycerol. The products were separated by electrophoresis on 1% agarose gels and visualized with ethidium bromide. The gels were photographed using a GelDoc2000 digital imaging system (Bio-Rad), and the band intensities were determined using Quantity One software (PDI).

Determination of the recognition sequence and cleavage site

To identify the minimal recognition sequence, plasmids containing a series of deletions surrounding the IS of the ApeK1.S908 intron were used as substrates in the cleavage reaction. A decrease in reaction products was interpreted as an indication that the deleted base pairs lay within the recognition sequence.

For cleavage site (CS) mapping, the double-stranded DNA substrates for I-ApeI cleavage were PCR amplified from pWT using the universal primers M13(-20) and M13 reverse, both of which were biotinylated at their 5' end. Each labeled primer was used in a separate PCR together with the appropriate non-labeled partner primer. Denatured I-ApeI-digested substrates (labeled as 'X' in Figure 3D) were run on 6% polyacrylamide/8.3 M urea gels, alongside the corresponding dideoxy sequencing reactions generated from the same 5' end-labeled primers. Biotinylated DNA was detected using an Imaging High Kit (Toyobo).

The nature of the cohesive termini generated by I-ApeI digestion was also examined. I-ApeI-digested pWT was treated with T4 DNA polymerase (Takara) in the presence of dNTPs, and the resulting blunt ends of the DNA fragment were re-ligated, to generate pWTΔPE8. The sequence of pWTΔPE8 was then compared with that of pWT.

Western blotting

A polyclonal antibody against the 16-residue synthetic polypeptide SSRKTPWPTQRVKDC, which corresponds to residues 87–101 of I-ApeI, was raised in female New Zealand White rabbits. Native I-ApeI protein was partially purified from the cell-free extract of *A. pernix* K1 using HiTrap SP, HiTrap heparin and Superdex 75 pg columns as described above. Aliquots of the samples were separated on 0.1% SDS-18% PAGE and subjected to western blotting with the rabbit antiserum followed by goat anti-rabbit immunoglobulin G-alkaline phosphatase conjugate (Sigma-Aldrich) and the colorimetric substrate Western Blue (Promega).

Determination of molecular weight

To determine the molecular weight, 120 µg of the purified His₆I-ApeI protein was sedimented in a 4.6 ml 10–30% glycerol gradient in 50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 0.1 mM EDTA and 0.5 mM DTT. Size standards included 1 mg each of BSA (66 kDa), pepsin (34.7 kDa), trypsinogen (24 kDa) and lysozyme (14.3 kDa). Centrifugation was carried out in a Beckman SW50.1 rotor at 45 000 r.p.m. for 40 h at 4°C. Of the 200 µl samples that were taken from the top of the gradient, 50 µl was subjected to 0.1% SDS-18% PAGE and detected by staining with Coomassie brilliant blue or western blotting with anti-I-ApeI serum. Each fraction was diluted

1:10 in buffer A, and 1 μ l was used in the I-ApeI activity assays.

Thermostability studies

The thermostability was tested by incubating the purified His₆I-ApeI protein (10 pmol/ μ l) at 70, 80 and 90°C in buffer A. Aliquots were stored on ice at different intervals, and the remaining activities were assayed under standard conditions.

The effect of the substrate DNA on the stability of I-ApeI activity was examined by incubating a mixture containing the His₆I-ApeI protein (1 pmol/ μ l) and the NcoI-linearized pWT (0.01 pmol/ μ l) in buffer A at 90°C. The remaining activities at the intervals indicated were monitored as described above.

Structure modeling

The structural coordinates of PI-PfuI (PDB entry: 1DQ3) were used to model the LAGLIDADG interface of I-ApeI. A model was built using the 3D-JIGSAW program (15) and visualized using RasMol software (version 2.7.1.1).

RESULTS AND DISCUSSION

Identification of I-ApeI

A previous study showed that the *A. pernix* 16S rRNA gene is interrupted by a 699 bp archaeal intron, ApeK1.S908 (formerly designated I α) (16). An ORF (726 bp) that uses an ATG start codon located 32 bp downstream of the 5' end of the intron has been annotated as the gene encoding a putative HEase. However, this prediction is not correct, because an assigned Shine-Dalgarno (SD) sequence for the ORF, GGGAGG, is far removed (17 bp) from the start codon. Torarinsson *et al.* (17) revealed that 86% of single genes and the first genes of operons in the *A. pernix* genome lie 7–15 bp downstream of the predominant SD sequence, GGGG. In view of these data, we reexamined the intron sequence and identified a suitable coding region for the putative HEase. The newly identified ORF (669 bp) uses an ATG start codon located 83 bp downstream of the 5' end of the intron, continues through the 3' intron-exon junction and extends 52 bp into the 3' exon. The putative SD sequence, GGGG, precedes the ORF 7 bp upstream from the start codon. This ORF was estimated to encode a protein of 222 amino acids with a molecular mass of 25.0 kDa.

Alignment with the amino acid sequences of known archaeal HEases revealed that the intronic ORF contains two copies of LAGLIDADG-like sequences (L₉VGALRDG₁₆ and V₁₁₁KGFVDAEG₁₁₉, termed P1 and P2, respectively; Figure 1A). The P1 and P2 sequences have three characteristic features: (i) a high content of hydrophobic residues, (ii) a scattering of small residues such as Gly and Ala that can form van der Waals contacts between protein backbone atoms and (iii) the presence of the acidic residues (D15, E118) capable of forming the active site. Whereas the P2 sequence has the GxxxG motif common to LAGLIDADG enzymes and transmembrane proteins (18), P1 has a one-residue gap in the C-terminal half region. Homology-based structural modeling of the ORF product using the endonuclease domain of PI-PfuI as a template suggests that its overall structure consists of two similar α/β domains ($\alpha\beta\beta\alpha\beta\beta\alpha$) that

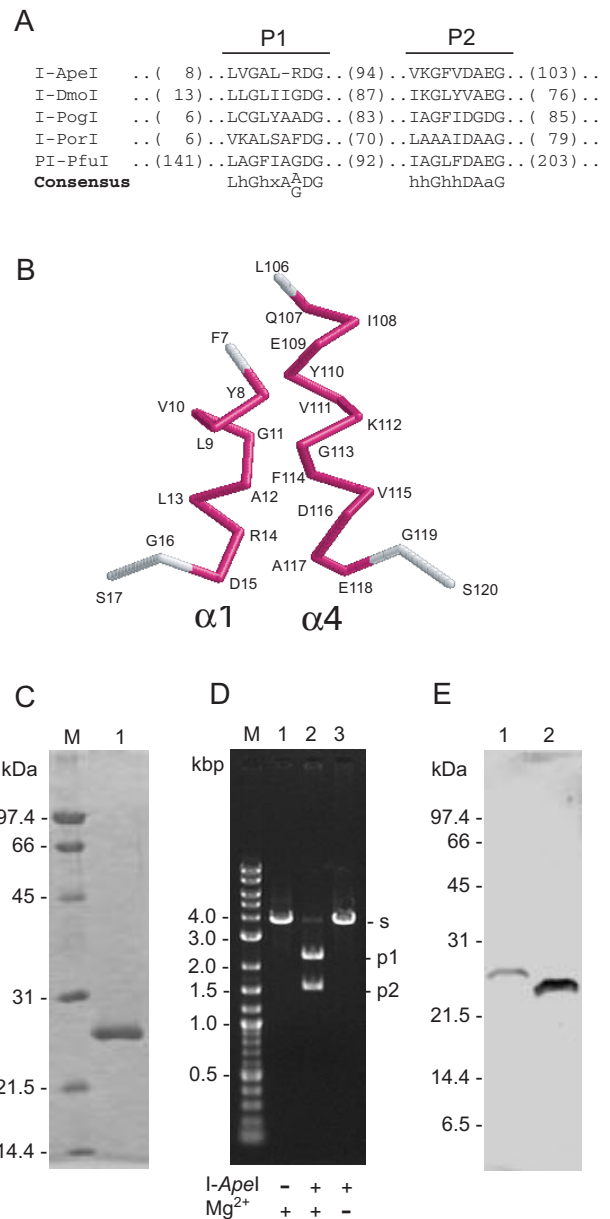


Figure 1. Identification of I-ApeI. (A) Alignment of the conserved motifs of representative archaeal intron/intein-encoded LAGLIDADG HEases, which share 13.5–17.1% overall amino acid identity. Pertinent residues in P1 and P2 are shown. Uppercase letters, conserved amino acids in standard single letter code; h, hydrophobic residue (G, A, C, F, V, L or I); a, acidic residue (D or E); x, variable residue. Numbers in parentheses give the amino acid sequence length of each region. (B) C α trace representation of a model of the LAGLIDADG interface of I-ApeI. The interfacial helix bundle consisting of $\alpha 1$ and $\alpha 4$ (P1 and P2 region) is shown in magenta color. (C) SDS-PAGE analysis of purified His₆I-ApeI (lane 1) with size standards (lane M). Purified protein (5 μ g) eluted from the Superdex 75 column was separated on a 0.1% SDS–12.5% polyacrylamide gel and stained with Coomassie brilliant blue. Positions of molecular size standards are indicated (in kDa) on the left. (D) Site-specific double-stranded DNA cleavage by I-ApeI. NcoI-cut pWT DNA (0.1 pmol) was incubated for 30 min at 90°C in the presence or absence of His₆I-ApeI (10 pmol) and Mg²⁺ (10 mM), as indicated. Positions of the substrate(s) and products (p1 and p2) are shown on the right; size markers (lane M) are shown on the left. (E) Western blot showing the production of native I-ApeI protein in the archaeal cells. An active fraction of the partially purified native enzyme (lane 2) was subjected to 0.1% SDS–18% PAGE analysis followed by immunodetection. Lane 1, purified His₆I-ApeI protein (1 μ g).

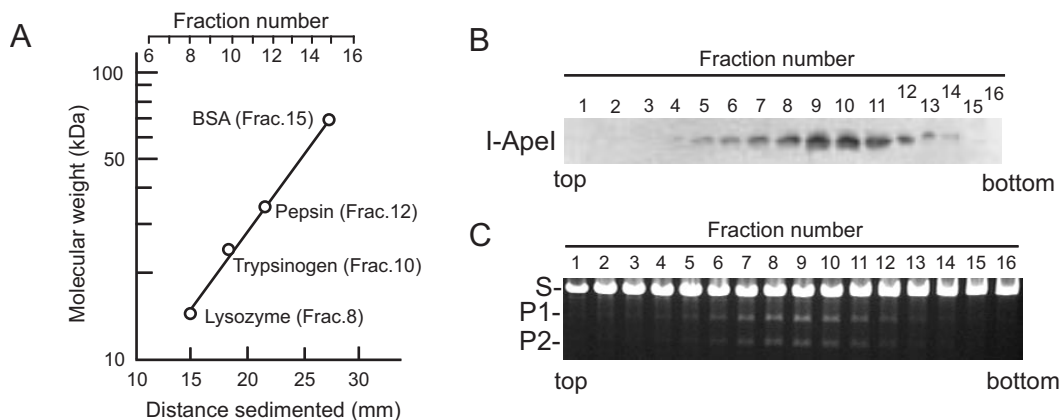


Figure 2. Sedimentation analysis of the His₆I-ApeI protein through a glycerol density gradient along with a set of molecular size standards. (A) Sedimentation of the standard proteins. The fraction in which a peak of each protein was detected is shown in parentheses. The His₆I-ApeI-containing fractions were determined by western blotting (B) and endonuclease assays (C).

assemble as a monomer with the domains related by pseudo 2-fold symmetry, and that the LAGLIDADG helices (corresponding to $\alpha 1$ and $\alpha 4$) are located at the center of this symmetry (Figure 1B). Owing to the gap in the P1 sequence, the C-terminal region of the $\alpha 1$ helix appears to have an anomalous structure. However, from these computational analyses, it remains unclear whether the aberration in the $\alpha 1$ helix has a significant effect on the formation of the LAGLIDADG two-helix bundle, which is mainly responsible for the domain interface as well as for the proper spacing of the neighboring acidic residues forming the active site.

To test the ability of the ORF product to cleave DNA in a sequence-specific manner, we prepared the N-terminal His₆-tagged recombinant protein. The ApeK1.S908 intron ORF is in-frame with the downstream exon sequence and the translation termination site is not predictable. Therefore, we expressed the ORF product from precursor rRNA using plasmid pIA1. Five purification stages were needed to obtain a homogeneous protein (His₆I-ApeI): heat treatment and ammonium sulfate fractionation, followed by three chromatographic steps (Figure 1C). A western blotting using specific anti-His₆ tag antibodies confirmed that this single band, with an apparent molecular mass (Mr) of 27 kDa, corresponds to the overexpressed protein (data not shown). Despite the presence of the His₆ tag, the recombinant protein did not bind to a Ni²⁺-affinity column under the native conditions used. Using the purified His₆I-ApeI protein, the endonuclease activity was assayed (Figure 1D). Incubation of NcoI-linearized pWT with His₆I-ApeI and Mg²⁺ at 90°C generated discrete cleavage products of 1.6 and 2.4 kb (Figure 1D, lane 2). Incubation of the same substrate with either His₆I-ApeI or Mg²⁺ on its own resulted in no cleavage (Figure 1D, lanes 1 and 3). The same activity was also demonstrated for the ORF product corresponding to translation from the excised intron RNA (data not shown). These results show that the archaeal intron ApeK1.S908 encodes a site-specific double-stranded DNA endonuclease whose activity depends on Mg²⁺.

Translation of the I-ApeI protein in the archaeal cells was confirmed by western blotting. Trace amounts of I-ApeI activity were present in the host cells. After partial purification from a cell-free extract, a fraction containing I-ApeI activity was obtained. As shown in Figure 1E, a single band with an Mr of

25 kDa was detected in an aliquot of this fraction (lane 2). The position of this band was slightly different from that of the His₆I-ApeI protein (Mr = 27 kDa; lane 1), reflecting the absence of the His₆ tag. In general, the homing mechanism of mobile introns requires the translation of intron-encoded HEases (3). This result indicates that homing might take place in natural hydrothermal environments.

In an attempt to determine its native molecular weight, the purified His₆I-ApeI was re-chromatographed on a calibrated Superdex 200 column. The V_e observed for His₆I-ApeI yielded a K_{av} of 0.81, corresponding to an invalid apparent molecular mass of ~7 kDa. This retarded elution may be due to non-specific interaction between the protein and the column. Then, we determined the size of I-ApeI by density gradient ultracentrifugation. His₆I-ApeI was detected in the resulting fractions by western blotting (Figure 2B) as well as by enzymatic assay (Figure 2C). The protein migrated at a distance of 17.1–18.9 mm (fractions 9 and 10), which corresponds to a molecular size of 21–28 kDa for a globular protein (Figure 2A). These results imply that I-ApeI is a monomeric enzyme with an Mr of 25.0 kDa.

Biochemical characterization of I-ApeI

Recognition sequence. To determine the minimal recognition sequence, a 24 bp stretch of DNA spanning the IS of the ApeK1.S908 intron was analyzed. A series of gradually truncated target sequences were prepared for the I-ApeI cleavage assays (Figure 3B and C). The initial reaction rate was reduced with substrates having deletions that included the segment from -11 to +9. I-ApeI cleaved the substrate DEL13, containing only the sequence from -11 to +9 as efficiently as the wild-type substrate, WT. These results indicate that the minimal target sequence for I-ApeI cleavage comprises 20 bp (5'-GCAAGGCTGAAACTTAAAGG). No HEase with this specificity has been reported previously. The fact that I-ApeI targets a non-palindromic DNA sequence is consistent with the conclusion that I-ApeI is a monomeric LAGLIDADG enzyme.

Cleavage sites. The CSs were mapped by comparison of the fragments digested with I-ApeI with dideoxy sequencing ladders. From the results shown in Figure 3D, it can be inferred

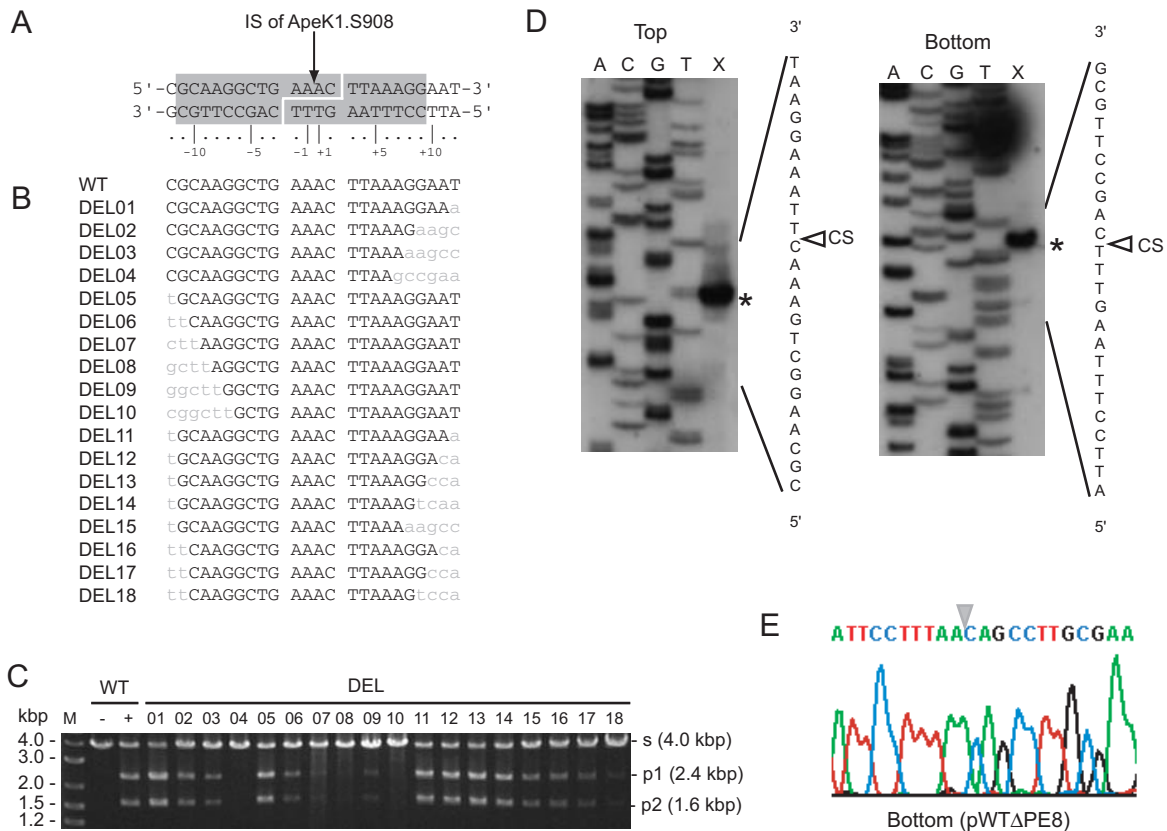


Figure 3. Recognition sequence and CS of I-ApeI. (A) Summary of recognition sequence mapping. The numbering below the sequence is described in the text. The region in which deletions and substitutions abolished cleavage is shaded in gray. The IS of the ApeK1.S908 intron is indicated by the arrow. (B) Sequences (top strand) of the recognition site mutants. WT refers to NcoI-digested pWT. The name of each deletion mutant is given on the left. Nucleotides replacing the deleted regions are represented in gray lower case. (C) I-ApeI cleavage assay for recognition site mutants. Each substrate (0.1 pmol) was incubated for 10 min at 90°C with His₆I-ApeI (10 pmol). The symbols are as described in Figure 1D. (D) CS mapping. The products of I-ApeI cleavage reactions (lane X) were subjected to electrophoresis alongside sequencing ladders. The CSs (open arrowheads) on the top strand (left panel) and bottom strand (right panel) are shown. Sequencing lanes A, C, G and T are denoted by the dideoxynucleotide species used in the reaction. The sequence of the target DNA immediately flanking the CS on each strand is shown on the right. (E) Determination of the cohesive termini generated by I-ApeI cleavage. The sequencing chromatogram for the area of interest on the bottom strand of pWTΔPE8 is shown. The gray arrowhead denotes the junction of the blunt ends produced by T4 DNA polymerase action on the I-ApeI digest.

that the I-ApeI CSs are after +2 and -2 on the top and bottom strands, respectively. Additional evidence supporting this conclusion was obtained from sequencing pWTΔPE8 (Figure 3E). The outcome indicates that cleavage by I-ApeI creates 4 bp 3'-protruding ends that are removed by the exonucleolytic activity of T4 DNA polymerase. The ends generated were a 5'-phosphate and a 3'-OH, as confirmed by ligation-recutting of the products (data not shown).

Optimum reaction conditions. The molar ratio of DNA:I-ApeI (1:100) was adjusted so that no more than 70% of the substrate was converted to the products, and the reaction rate was linear as a function of the time within 10 min. The excess amount of I-ApeI could attribute to its low thermostability as described below.

I-ApeI was active between 70 and 90°C, with a maximum at 90°C (Figure 4A). Even at temperatures ~90°C, the double-stranded DNA substrate was stabilized by Mg²⁺. However, at temperatures >95°C, I-ApeI activity could not be estimated precisely due to rapid denaturation of the DNA. I-ApeI exhibited high activity at relatively high (180–220 mM) concentrations of NaCl (Figure 4B). The optimal pH for cleavage was 7.5 (Figure 4C).

Divalent metal ions are essential cofactors for all phosphoryl-transfer reactions catalyzed by nucleases (19). I-ApeI cleavage required Mg²⁺ (Figures 1D and 4D); excess EDTA completely inhibited the reaction. Interestingly, I-ApeI preferred Mn²⁺ as a cofactor at low concentrations (1–5 mM) to Mg²⁺, without altering its cleavage site on the wild-type substrate DNA. I-ApeI had no activity at 10 mM Mn²⁺ because of aggregation of the substrate DNA. Co²⁺ (1 mM) was able to substitute for Mg²⁺ as a cofactor, but resulted in less efficient cleavage (83%). We did not examine whether star activity was created in the presence of Mg²⁺ or Co²⁺. No activity was observed when Mg²⁺ was replaced by Ca²⁺, Ba²⁺, Zn²⁺, Ni²⁺ or Sr²⁺, although the presence of these divalent cations did not cause aggregation of the substrate DNA.

In order to evaluate the thermostability of I-ApeI, the time course of its heat inactivation was monitored at temperatures of 70, 80 and 90°C. There was no apparent activity loss after the enzyme was incubated alone for 10 min at 70°C, whereas the I-ApeI apoenzyme was extremely labile at 90°C and retained no activity after 2 min of incubation (Figure 4E). Nevertheless, significant protection against thermal inactivation was exerted by the substrate DNA (Figure 4F). After 2, 4

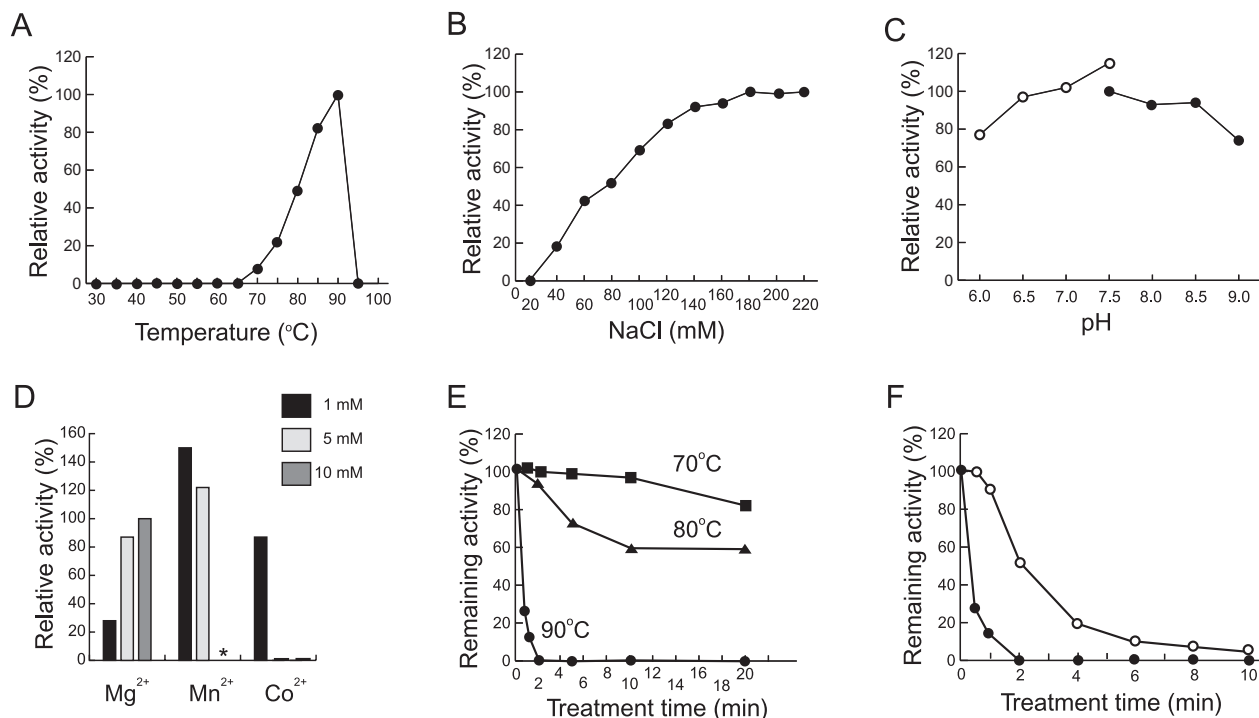


Figure 4. Optimization of the I-ApeI cleavage reaction. Effect of temperature (A), NaCl concentration (B), pH (C) and divalent cations (D) on I-ApeI activity. The NcoI-cut pWT DNA substrate was incubated for 10 min with His₆I-ApeI under the conditions indicated. The efficiency of cleavage in the standard reaction conditions described in the text was set at 100%. In (C), open and closed circles indicate activities in the assays using 10 mM PIPES-NaOH and 10 mM Tris-HCl buffers, respectively. In (D), an asterisk indicates inability to determine activity due to aggregation of the substrate DNA. I-ApeI activity was completely inhibited in the presence of 5 mM or 10 mM Co²⁺, although no obvious aggregation of the substrate DNA occurred. (E) Thermostability of I-ApeI. The enzyme was incubated at 70°C (squares), 80°C (triangles) and 90°C (circles) for the time periods indicated. The remaining activities after the heat treatments are shown. (F) Effect of the substrate DNA on the stability of I-ApeI at 90°C. The enzyme was incubated in the presence (open circles) or absence (closed circles) of NcoI-digested pWT. The remaining activities at the intervals indicated are plotted as fractions of the initial activity.

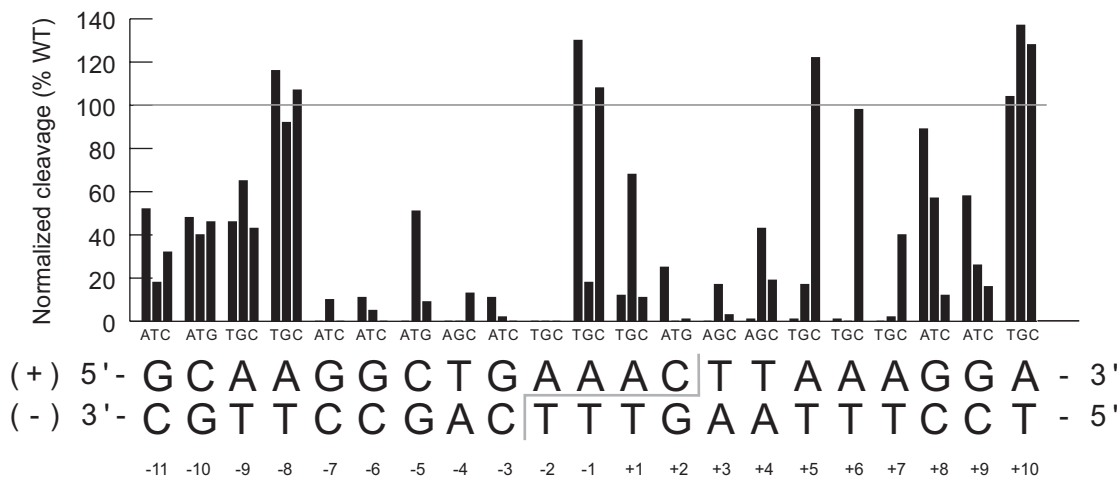


Figure 5. Target-sequence preference of I-ApeI. Top: diagram showing the positions of single base pair substitutions in I-ApeI mutant substrates and the associated cleavage activities. Below each column are shown the base substitutions on the upper strand. Cleavage activity obtained with each mutant substrate is shown as a percentage of the WT. Values are means of triplicate assays. Bottom: the wild-type recognition sequence into which single substitutions were introduced. The staggered line indicates the I-ApeI cleavage site.

or 6 min of incubation with the substrate DNA at 90°C, the enzyme still retained 51, 18 and 7%, respectively, of its catalytic activity. These results led us to hypothesize that some molecular interaction between I-ApeI and genomic DNA could be responsible for the prevention of the enzyme decay in the hyperthermophilic archaeal cells.

Single base pair substitutions that influence I-ApeI cleavage efficiency. To identify the DNA bases essential for I-ApeI cleavage, we performed systematic assays with a set of variant substrates with single base pair substituted one by one with the three alternative bases. A total of 63 single base pair variants at 21 positions from -11 to +10 were prepared. Each of the

variant substrates was incubated with His₆I-ApeI for 10 min, and the extents of cleavage observed are shown in Figure 5.

Mutations at positions -7, -6, -4, -3, -2, +2 and +3 significantly reduced the efficiency of cleavage. A >80% loss of I-ApeI function was observed with any of the alternative bases at these positions, indicating that these seven bases are strongly involved in the DNA-protein interaction. The effect of base substitutions at positions -3 and +3 are of particular note, since similar results have been obtained in mutational experiments on the substrates of other LAGLID-ADG enzymes, such as I-CreI (20), I-CeuI (21), I-SceI (22), I-SceII (23), I-PorI (24) and PI-SceI (25). This leads to the general rule that the bases located close outside the CSs are essential for recognition by LAGLIDADG enzymes. In agreement with this inference, structural data on DNA co-crystals of I-CreI, I-MsoI, I-AniI, I-SceI, E-DreI and PI-SceI have demonstrated that the bases at positions -3 and +3 make direct contact via hydrogen bonds with residues of the proteins (8,26–30). Conversely, substitution of -8 had little or no effect on the extent of cleavage, suggesting that the base at this position is of minor importance in the cleavage reaction. Biased base preferences were observed at the positions of -5, -1, +1, +4, +5, +6, +7, +8 and +9.

In conclusion, the preferred sequence for I-ApeI cleavage is 5'-RCANGGCTG'ADRC↓TTMMADR [N = A or C or G or T, D = A or G or T, R = A or G, M = A or C; ^ and ↓ indicate the cleavage site on the bottom and top strand, respectively]. The effects of multiple substitutions within the recognition sequence on I-ApeI cleavage reaction are yet to be determined.

ACKNOWLEDGEMENTS

This work was supported by a Grant-in-Aid for Scientific Research on Priority Areas 'Genome Biology', and by the National Project on Protein Structural and Functional Analyses, from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. Y.M. was the recipient of a fellowship from the Japan Society for the Promotion of Science (JSPS). Funding to pay the Open Access publication charges for this article was provided by MEXT.

Conflict of interest statement. None declared.

REFERENCES

- Belfort, M. and Roberts, R.J. (1997) Homing endonucleases: keeping the house in order. *Nucleic Acids Res.*, **25**, 3379–3388.
- Chevalier, B.S. and Stoddard, B.L. (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res.*, **29**, 3757–3774.
- Belfort, M., Derbyshire, V., Parker, M.M., Cousineau, B. and Lambowitz, A.M. (2002) Mobile introns: pathways and proteins. In Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds), *Mobile DNA II*. ASM Press, Washington DC, pp. 761–783.
- Galburt, E.A. and Stoddard, B.L. (2002) Catalytic mechanisms of restriction and homing endonucleases. *Biochemistry*, **41**, 13851–13860.
- Guhan, N. and Muniyappa, K. (2003) Structural and functional characteristics of homing endonucleases. *Crit. Rev. Biochem. Mol. Biol.*, **38**, 199–248.
- Rong, Y.S. and Golic, K.G. (2000) Gene targeting by homologous recombination in *Drosophila*. *Science*, **288**, 2013–2018.
- Chevalier, B.S., Kortemme, T., Chadsey, M.S., Baker, D., Monnat, R.J. and Stoddard, B.L. (2002) Design, activity, and structure of a high specific artificial endonuclease. *Mol. Cell*, **10**, 895–905.
- Epinat, J.C., Arnould, S., Chames, P., Rochaix, P., Desfontaines, D., Puzin, C., Patin, A., Zanghellini, A., Paques, F. and Lacroix, E. (2003) A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res.*, **31**, 2952–2962.
- Sussman, D., Chadsey, M., Fauce, S., Engel, A., Bruett, A., Monnat, R.J., Stoddard, B.L. and Seligman, L.M. (2004) Isolation and characterization of new homing endonuclease specificities at individual target site positions. *J. Mol. Biol.*, **342**, 31–41.
- Nomura, N., Morinaga, Y., Kogishi, T., Kim, E.J., Sako, Y. and Uchida, A. (2002) Heterogeneous yet similar introns reside in identical positions of the rRNA genes in natural isolates of the archaeon *Aeropyrum pernix*. *Gene*, **295**, 43–50.
- Redder, P., Faguy, D., Brugger, K., She, Q. and Garrett, R.A. (2002) Archaeal mobile DNA. In Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds), *Mobile DNA II*. ASM Press, Washington, DC, pp. 1060–1073.
- Morinaga, Y., Nomura, N. and Sako, Y. (2002) Population dynamics of archaeal mobile introns in natural environments: a shrewd invasion strategy of the latent parasitic DNA. *Microbes Environ.*, **17**, 153–163.
- Itoh, T., Nomura, N. and Sako, Y. (2003) Distribution of 16S rRNA introns among the family *Thermoproteaceae* and their evolutionary implications. *Extremophiles*, **7**, 229–233.
- Sako, Y., Nomura, N., Uchida, A., Ishida, Y., Morii, H., Koga, Y., Hoaki, T. and Maruyama, T. (1996) *Aeropyrum pernix* gen. nov., sp. nov., a novel aerobic hyperthermophilic archaeon growing at temperatures up to 100°C. *Int. J. Syst. Bacteriol.*, **46**, 1070–1077.
- Bates, P.A., Kelley, L.A., MacCallum, R.M. and Sternberg, M.J.E. (2001) Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins*, **5**(Suppl.), 39–46.
- Nomura, N., Sako, Y. and Uchida, A. (1998) Molecular characterization and postsplicing fate of three introns within the single rRNA operon of the hyperthermophilic archaeon *Aeropyrum pernix* K1. *J. Bacteriol.*, **180**, 3635–3643.
- Torarinsson, E., Klenk, H.P. and Garrett, R.A. (2005) Divergent transcriptional and translational signals in *Archaea*. *Environ. Microbiol.*, **7**, 47–54.
- Silva, G.H. and Belfort, M. (2004) Analysis of the LAGLIDADG interface of the monomeric homing endonuclease I-DmoI. *Nucleic Acids Res.*, **32**, 3156–3168.
- Suck, D. (1992) Nuclease structure and catalytic function. *Curr. Opin. Struct. Biol.*, **2**, 84–92.
- Argast, G.M., Stephens, K.M., Emond, M.J. and Monnat, R.J. (1998) I-PpoI and I-CreI homing site sequence degeneracy determined by random mutagenesis and sequential *in vitro* enrichment. *J. Mol. Biol.*, **280**, 345–353.
- Marshall, P. and Lemieux, C. (1992) The I-CeuI endonuclease recognizes a sequence of 19 base pairs and preferentially cleaves the coding strand of the *Chlamydomonas moewusii* chloroplast large subunit rRNA gene. *Nucleic Acids Res.*, **20**, 6401–6407.
- Perrin, A., Buckle, M. and Dujon, B. (1993) Asymmetrical recognition and activity of the I-SceI endonuclease on its site and on intron-exon junctions. *EMBO J.*, **12**, 2939–2947.
- Wernette, C., Saldanha, R., Smith, D., Ming, D., Perlman, P.S. and Butow, R.A. (1992) Complex recognition site for the group I intron-encoded endonuclease I-SceII. *Mol. Cell Biol.*, **12**, 716–723.
- Lykke-Andersen, J., Thi-Ngoc, H.P. and Garrett, R.A. (1994) DNA substrate specificity and cleavage kinetics of an archaeal homing-type endonuclease from *Pyrobaculum organotrophum*. *Nucleic Acids Res.*, **22**, 4583–4590.
- Gimble, F.S. and Wang, J. (1996) Substrate recognition and induced DNA distortion by the PI-SceI endonuclease, an enzyme generated by protein splicing. *J. Mol. Biol.*, **263**, 163–180.
- Jurica, M.S., Monnat, R.J. and Stoddard, B.L. (1998) DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-CreI. *Mol. Cell*, **2**, 469–476.
- Moure, C.M., Gimble, F.S. and Quijoch, F.A. (2002) Crystal structure of the intein homing endonuclease PI-SceI bound to its recognition sequence. *Nature Struct. Biol.*, **9**, 764–770.

28. Chevalier,B., Turmel,M., Lemieux,C., Monnat,R.J. and Stoddard,B.L. (2003) Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-CreI and I-MsoI. *J. Mol. Biol.*, **329**, 253–269.
29. Bolduc,J.M., Spiegel,P.C., Chatterjee,P., Brady,K.L., Downing,M.E., Caprara,M.G., Waring,R.B. and Stoddard,B.L. (2003) Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing factor. *Genes Dev.*, **17**, 2875–2888.
30. Moure,C.M., Gimble,F.S. and Quioco,F.A. (2003) The crystal structure of the gene targeting homing endonuclease I-SceI reveals the origins of its target site specificity. *J. Mol. Biol.*, **334**, 685–695.