

# Novel protein fold discovered in the PabI family of restriction enzymes

Ken-ichi Miyazono<sup>1</sup>, Miki Watanabe<sup>2</sup>, Jan Kosinski<sup>3</sup>, Ken Ishikawa<sup>2,5</sup>,  
Masayuki Kamo<sup>1</sup>, Tatsuya Sawasaki<sup>4</sup>, Koji Nagata<sup>1</sup>, Janusz M. Bujnicki<sup>3</sup>,  
Yaeta Endo<sup>4</sup>, Masaru Tanokura<sup>1</sup> and Ichizo Kobayashi<sup>2,5,6,\*</sup>

<sup>1</sup>Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, University of Tokyo, Tokyo, 113-8657, Japan, <sup>2</sup>Department of Medical Genome Sciences, Graduate School of Frontier Science, University of Tokyo, Tokyo 108-8639, Japan, <sup>3</sup>Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, Trojdena 4, 02-109 Warsaw, Poland, <sup>4</sup>Department of Applied Chemistry, Faculty of Engineering, Ehime University, Matsuyama 790-8577, Japan, <sup>5</sup>Graduate Program in Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Tokyo 108-8639, Japan and <sup>6</sup>Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan

Received November 20, 2006; Revised February 1, 2007; Accepted February 1, 2007

## ABSTRACT

Although structures of many DNA-binding proteins have been solved, they fall into a limited number of folds. Here, we describe an approach that led to the finding of a novel DNA-binding fold. Based on the behavior of Type II restriction–modification gene complexes as mobile elements, our earlier work identified a restriction enzyme, R.PabI, and its cognate modification enzyme in *Pyrococcus abyssi* through comparison of closely related genomes. While the modification methyltransferase was easily recognized, R.PabI was predicted to have a novel 3D structure. We expressed cytotoxic R.PabI in a wheat-germ-based cell-free translation system and determined its crystal structure. R.PabI turned out to adopt a novel protein fold. Homodimeric R.PabI has a curved anti-parallel  $\beta$ -sheet that forms a ‘half pipe’. Mutational and *in silico* DNA-binding analyses have assigned it as the double-strand DNA-binding site. Unlike most restriction enzymes analyzed, R.PabI is able to cleave DNA in the absence of Mg<sup>2+</sup>. These results demonstrate the value of genome comparison and the wheat-germ-based system in finding a novel DNA-binding motif in mobile DNases and, in general, a novel protein fold in horizontally transferred genes.

## INTRODUCTION

Determination of the 3D structure of DNA-binding proteins has played a key role in the study of various cellular genetic processes involving DNA, such as transcription, replication, repair, mutagenesis and recombination. Although many structures of DNA-binding proteins have been solved, they fall into a limited number of folds. Finding a novel DNA-binding fold and, therefore, a novel mode of protein–DNA interaction from the point of view of structural biology will broaden our understanding of these processes, genomes and life. In the present work, we introduce a new approach in search of novel DNA-binding folds. This approach is based on the structural diversity of restriction enzymes and the behavior of their genes as mobile elements.

Restriction–modification systems consist of two enzymatic activities—namely, a restriction enzyme that recognizes a specific DNA sequence and introduces a double-strand break, and a cognate modification enzyme that can methylate the same sequence and thereby render it resistant to the restriction enzyme. Their genes are often tightly linked and form a restriction–modification gene complex. There are several lines of evidence that some restriction–modification gene complexes behave as selfish mobile genetic elements, similar to viral genomes and transposons (1–3). These include their attack on the host chromosome (1), the presence of restriction–modification gene complexes on a variety of mobile elements, their

\*To whom correspondence should be addressed. Tel: +81 3 5449 5326; Fax: +81 3 5449 5422; Email: ikobaya@ims.u-tokyo.ac.jp

Present address:

Masayuki Kamo, Intercyto Nano Science CO., LTD, Saito Bio-Incubator #1037-7-15, Saito-Asagi, Ibaraki, Osaka 567-0085, Japan

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

aberrant GC content and codon usage suggesting their recent transfer from distantly related organisms (2,3), and their association with genome rearrangements suggested from comparison of closely related prokaryotic genomes (4,5) and verified in the laboratory (6).

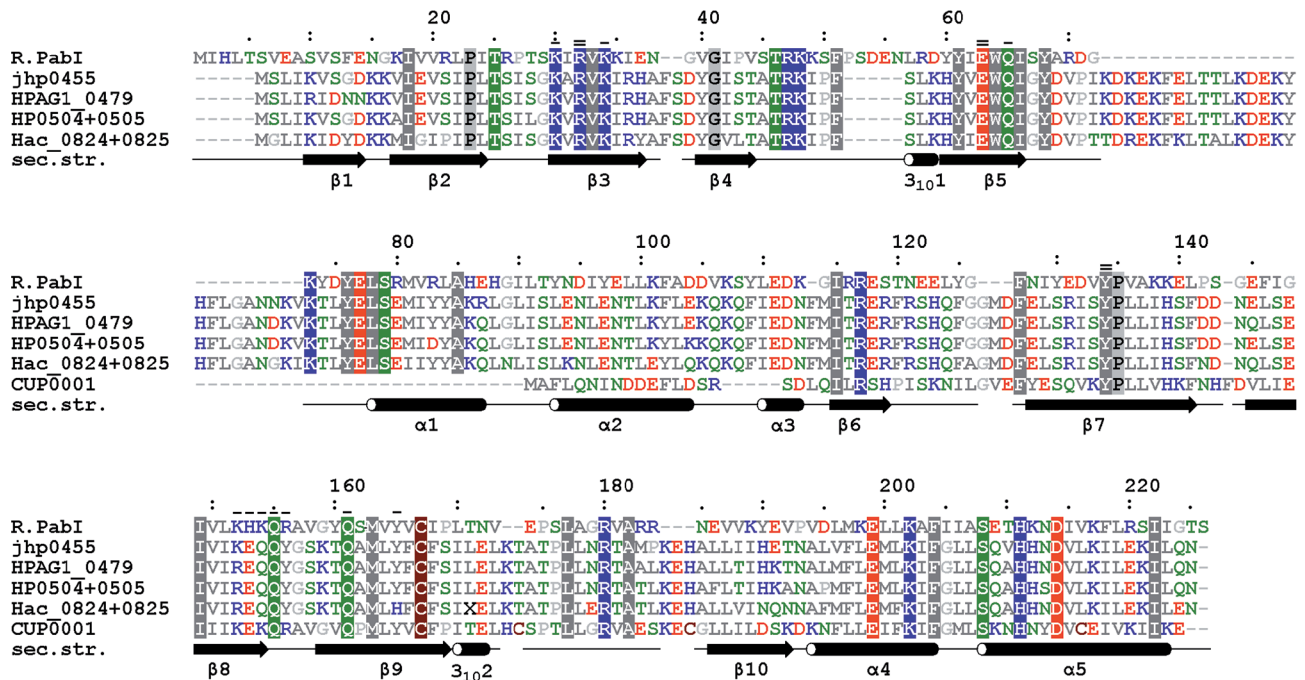
So far, restriction enzymes have been found to belong to four evolutionarily unrelated superfamilies with different folds (7). Initially, all the structurally characterized restriction enzymes belonged to the PD-(D/E)XK superfamily, which is represented by enzymes such as R.EcoRI and R.EcoRV (8). Only recently was the structure of R.BfiI solved, revealing a completely different 3D fold and membership of this enzyme in the phospholipase D (PLD) superfamily (9). A number of restriction enzymes were also predicted to belong to the HNH superfamily (10,11), and for some of these the predictions were confirmed by mutagenesis and biochemical studies, e.g. R.KpnI (12). The fourth group with a tentatively assigned different 3D fold is comprised of a few restriction enzymes from the GIY–YIG superfamily (11). However, for the majority of restriction enzyme sequences, the structure has been neither determined nor predicted, and it remains to be determined if the so far ‘unassigned’ sequences belong to one of the four ‘old’ superfamilies or to some other fold, known or unknown.

Genome comparison is a powerful approach for the identification of new restriction enzymes and accompanying modifying enzymes because of the behavior of some restriction–modification gene complexes as a mobile element (2,3,7). A new restriction enzyme, R.PabI, which

catalyzes the cleavage of 5'-GTAC generating a TA3' overhang, was found by comparing the genomes of two hyperthermophilic archaea, *Pyrococcus abyssi* and *Pyrococcus horikoshii* (13,14). R.PabI exhibited a new pattern of sequence conservation and predicted secondary structure that did not match any of the four canonical restriction-enzyme motifs (Figure 1), prompting it as a candidate for a new fold or a significant modification of one of the known folds (14).

One problem encountered in pursuing the determination of the crystal structure of R.PabI was its toxicity to the cells. The uninduced level of R.PabI killed *Escherichia coli* cells even when the cognate methyltransferase M.PabI (15) was co-expressed (M.W. and I.K., unpublished data). *In vitro* translation systems, in contrast, can synthesize almost any protein, often with high accuracy and at a speed approaching *in vivo* rates (16). Among the cell-free systems, the wheat-germ-based system is of special interest (17): the translation machine has little codon preference; translation is uncoupled from transcription, so that digestion of the template DNA by the produced restriction enzyme does not take place; and the system has little discrimination between methionine and selenomethionine (SeMet), a characteristic that is useful in crystal structure determination.

In this study, we produced R.PabI proteins, unlabeled and SeMet-labeled, in this wheat-germ-based system, obtained their crystals and solved their structure. The crystal structure, mutagenic analysis and *in silico* modeling of DNA binding of R.PabI reveals that the



**Figure 1.** Sequence alignment between R.PabI and its homologs identified in protein sequence databases. jhp0455 and HPAG1\_0479 are putative translation products of open reading frames (ORFs) from *Helicobacter pylori* strains J99 and HPAG1. HP0504+0505 and Hac\_0824+0825 are reconstructed translation products of pseudogenes disrupted by frameshift mutations in *H. pylori* 26695 and *H. acinonychis* str. Sheeba. CUP0001 is a product of an incomplete ORF from *Campylobacter upsaliensis* RM3195. Amino acid residues with similar physico-chemical properties are indicated by color (hydrophobic—gray, negative—red, positive—blue, polar—green, cysteine—brown). Identical residues are highlighted. Residues analyzed by alanine mutagenesis in this work are indicated in the top by ‘:’ (mutant is partially active) or ‘=’ (mutant’s activity is undetectable). The bottom panel shows secondary structure determined for R.PabI.

protein adopts a novel fold and a novel mode of protein–DNA interaction.

## MATERIALS AND METHODS

### Large-scale preparation of R.PabI in wheat-germ-based cell-free expression system

The new nomenclature and abbreviations for restriction enzymes and their genes (18) were used. *P. abyssi* GE5 genomic DNA was provided by Dr Yoshizumi Ishino (University of Kyushu, Japan) (14). The *pabIR* ORF was amplified from this DNA template by polymerase chain reaction (PCR) as described previously (14). It was inserted into a pEU3b vector (19), which is transcribed with SP6 polymerase. The insert was confirmed by DNA sequencing. This plasmid was named pMW40.

For crystallization, wild-type R.PabI was synthesized in a large-scale, wheat-germ-based, cell-free translation system. An extract was prepared from purified wheat embryos, and endogenous small molecules such as amino acids were removed by gel filtration (20). The mRNA was transcribed *in vitro* from the R.PabI gene that was cloned into a vector specialized for cell-free expression (see above). A large-scale cell-free protein production was carried out using the bilayer reaction method (21). Briefly, 0.5 ml of the translation mixture containing 60 A260 units of the extract, the synthesized mRNA and all of the necessary ingredients such as 20 amino acids (250  $\mu$ M each), ATP, GTP, energy-generating system and ions were placed on the bottom of a well in a 6-well plate (3.2 cm in diameter, Asahi Techno Glass Corp.), then 5.5 ml of the substrate mixture containing the same ingredients except for the extract and mRNA was carefully overlaid. After incubation at 26°C for 12 h, the protein synthesis solution containing R.PabI was heated at 90°C for 15 min and the denatured proteins and insoluble materials were removed by centrifugation (12 000 r.p.m., 4°C, 15 min). The supernatant was purified through a Heparin–Sepharose affinity column (GE Healthcare). Twenty-four wells of the above reaction gave 3 mg of R.PabI with a purity of approximately 90%. For the preparation of SeMet-labeled protein, similar reactions were carried out in the presence of SeMet (250  $\mu$ M) instead of methionine in the reaction solutions. The content of endogenous methionine in the translation mixture was measured as 2  $\mu$ M by an amino-acid analyzer (L-80800, Hitachi), so that the fraction of SeMet incorporated was calculated to be more than 99%.

### Crystallization

The crystallization conditions of R.PabI were screened by the sparse-matrix method using Crystal Screen HT, Grid Screen PEG6000 (Hampton Research) and Wizard I & II (Emerald Biostructures). Because of the low solubility of R.PabI, we used a low concentration (0.5 mg/ml) for the first screening. The crystallization experiments of R.PabI were performed by the sitting drop vapor diffusion method. Crystallization drops were made by mixing 1  $\mu$ l of the protein solution [0.5 mg/ml protein in 10 mM Tris-HCl (pH 7.5), 200 mM NaCl and 1 mM dithiothreitol (DTT)]

and an equal amount of reservoir solution. The best crystal of R.PabI appeared under the reservoir conditions containing 100 mM MES (pH 6.0) and 5% PEG6000 2 weeks later. Crystals of the SeMet variant of R.PabI were obtained under identical crystallization conditions, but were too small to collect data for enough resolution. Larger crystals of the SeMet variant were obtained when the protein solution was concentrated to 1.9 mg/ml and the buffer composition was changed to 10 mM MES (pH 6.0), 200 mM NaCl, 10 mM MgCl<sub>2</sub> and 10 mM DTT. The best crystal of the SeMet variant of R.PabI was grown under the reservoir conditions containing 50 mM MES (pH 6.8) and 1% PEG6000 using concentrated protein solution 1 day later.

### Structure determination

Our data collection and refinement statistics are summarized in Table 1. X-ray diffraction datasets of native and SeMet variant crystals of R.PabI were collected at the BL-5A and NW-12 beam-lines at the Photon Factory (Tsukuba, Japan), respectively. All the measurements were carried out under cryogenic conditions (95 K) using 20% ethylene glycol (final concentration) as the cryoprotectant. A native crystal of R.PabI diffracted X-rays to a resolution of 3.0 Å. The X-ray diffraction data were integrated and scaled with HKL2000 (22). The native R.PabI crystal belonged to the space group  $P2_1$  with the unit cell dimensions of  $a = 84.6$  Å,  $b = 114.0$  Å,  $c = 89.2$  Å and  $\beta = 116.3^\circ$ . Consideration of the values of  $V_M$  suggests that this crystal has six protein molecules per asymmetric unit ( $V_M = 2.5 \text{ \AA}^3/\text{Da}$ ; (23)). Diffraction data of the SeMet variant of R.PabI were collected and processed in the same way. A SeMet variant crystal diffracted X-rays to a resolution of 2.9 Å and belonged to the same space group

**Table 1.** Summary of data collection and refinement statistics

	R.PabI	SeMet variant of R.PabI Peak
Wave length (Å)	1.0000	0.9792
Space group	$P2_1$	$P2_1$
Unit cell dimensions		
<i>a</i> (Å)	84.6	84.6
<i>b</i> (Å)	114.0	114.2
<i>c</i> (Å)	89.2	89.4
$\beta$ (deg.)	116.3	116.3
Resolution (Å)	2.00–3.0 (3.11–3.00)	2.00–2.9 (3.00–2.90)
Completeness (%)	99.3 (96.7)	100.0 (100.0)
Unique reflection	30 687	33 650
Averaged redundancy	3.6 (3.4)	7.5 (7.6)
$R_{\text{merge}}$ (%)	6.7 (28.3)	7.7 (28.0)
$I/\sigma$	14.8 (4.5)	17.8 (5.1)
Number of non-hydrogen atoms		
Protein	10599	
Water	0	
$R_{\text{cryst}}/R_{\text{free}}$ (%)	24.9/31.8	
R.m.s. deviations from ideal values		
Bond length (Å)	0.009	
Bond angle (deg.)	1.5	
Average <i>B</i> -factor	65.4	

*Note:* Values in parentheses are for the highest-resolution shell.



with the native crystal,  $P2_1$  with the unit cell dimensions of  $a = 84.6 \text{ \AA}$ ,  $b = 114.2 \text{ \AA}$ ,  $c = 89.4 \text{ \AA}$  and  $\beta = 116.3^\circ$ .

The crystal structure of R.PabI was determined by the single wavelength anomalous diffraction (SAD) phasing method using the diffraction data set of the SeMet variant of R.PabI. The selenium substructure was solved using the SnB program (24). A total of 19 selenium sites were determined in the asymmetric unit. The initial phase was calculated with the program SHARP (25) using the coordinates of selenium sites solved by SnB program. Phase calculation resulted in an overall figure of merit (FOM) of 0.31 for the resolution range of 20–2.9 Å. After that, density modification and initial model building was performed with the program RESOLVE (26). Molecular models of 759 residues (56% of the total) were automatically built with this calculation. The initial model of the SeMet variant of R.PabI was refined and manually rebuilt with the programs CNS (27) and XtalView (28), using 10% (randomly chosen) of the reflections to calculate the  $R_{\text{free}}$ . The partially refined model was transformed into five other subunits using the program MOLREP (29) in CCP4 (30), and refined with the program REFMAC5 (31) with non-crystallographic symmetry (NCS) restraints. The crystal structure of native R.PabI was determined by a molecular replacement method using the coordinates of the SeMet variant structure of R.PabI with the program MOLREP. The final structure of R.PabI was refined and built using a native crystal diffraction data set (20–3.0 Å) with the program CNS (without NCS restraints) and XtalView. We did not use NCS refinement in the final step of refinement because  $R$ -factor and  $R_{\text{free}}$  values became worse when refined with NCS restraint. We could not build coordinates of water molecules in this structure. Though we see some electron density peaks around R.PabI molecules, we could not clearly determine whether they were from water molecules or simply from noises.

### Construction, small-scale expression and assay of R.PabI mutants

Site-directed mutagenesis of pKI2 (=pEU-NII::pabIR, made from pEU-NII, a vector in PROTEIOS™ kit below) with specific primers was carried out using QuikChange Site-directed Mutagenesis™ kit (Stratagene), following the manufacturer's instruction. The synthetic oligonucleotides used (synthesized by Operon Biotechnologies) are listed in Supplementary Table 1. All the mutant plasmids were checked by DNA sequencing (Takara Bio). Wild-type R.PabI protein and the R.PabI mutant proteins were synthesized using PROTEIOS™ ver.2 (TOYOBO), a kit for small-scale, wheat-germ-based, cell-free protein synthesis, by the bilayer method as described previously (14,19). After synthesis, we heated the resulting solution at 90°C for 5 min to remove denatured proteins and insoluble materials by centrifugation. Expression of the R.PabI mutants in a soluble form in the supernatant was confirmed by SDS-PAGE.

To assess the restriction activity of the mutants, a 2559 bp linear substrate DNA with a single recognition sequence (5'-GTAC) was treated as described

previously (14). Its cleavage by R.PabI produces two DNA fragments of 563 bp and 1996 bp, respectively (14). The reaction mixture contained 50 mM MES (pH 6.5 at 25°C), 100 mM NaCl, 0.2 µg of this DNA and appropriate concentrations of each of the above R.PabI enzyme preparations. The mixture was incubated at 75°C for 1 h. To remove small molecules of mRNA present in the wheat-germ extract, the reaction mixture was treated with RNase A (14). The DNAs were separated by electrophoresis through a 1% agarose gel in TAE buffer (= 50 mM Tris-HCl pH8.0, 20 mM CH<sub>3</sub>COONa, 1 mM EDTA) and visualized with ultraviolet light after ethidium bromide staining.

### DNA binding

The heated supernatant of R.PabI (wild-type and R32A mutant) were incubated with 5 nM of <sup>33</sup>P-labeled, 40-mer double-stranded oligonucleotides containing one (#1: 5' GGACGCTTCACCGGATGTACAGGCATGCGACG ACCCCTAG 3' and its complement) or no R.PabI site (#2: 5' GGACGCTTCACCGGATGCTAAGGCATGC GACGACCCCTAG 3' and its complement) in a buffer (50 mM MES, pH 6.0 and 100 mM NaCl) at 25°C for 10 min. The free DNA and enzyme-bound DNA were separated on 8% native polyacrylamide gel in 1 × TBE buffer (50 mM Tris-HCl, pH 8.0, 50 mM boric acid, 1 mM EDTA) and autoradiographed.

### R.PabI DNA cleavage in the absence of added divalent cation

R.PabI was prepared as described previously (14). In brief, the protein was synthesized using wheat-germ-based cell-free protein synthesis system PROTEIOS™ (TOYOBO) and purified by heat treatment and chromatography through Heparin-Sepharose, followed by dialysis and concentration. To assess the restriction activity, the same substrate described in the section of mutant analysis was used. The reaction mixture contained buffer A (= 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM DTT), 0.2 µg of this DNA, the above purified R.PabI, and either 1 mM EDTA or 10 mM MgCl<sub>2</sub>. The mixture was incubated at 85°C for 1 h. The DNAs were separated by electrophoresis through 1% agarose gel in TAE buffer (= 50 mM Tris-HCl pH 8.0, 20 mM CH<sub>3</sub>COONa, 1 mM EDTA) and visualized with ultraviolet light after ethidium bromide staining.

### In silico analysis

Structure was analyzed with CCP4 (30) and APBS (32) and visualized with PyMol. The DNA-binding region was predicted using the PreDs program (33). PreDs classifies each residue on the surface of a protein to the DNA-binding and non-DNA binding regions using a statistical evaluation function (34). The evaluation function was developed based on an analysis of the shape and electrostatic properties of DNA-binding regions in structures of 63 protein-DNA complexes. GRAMM 1.03 (35) was used in the low-resolution docking mode to generate 2000 alternative docking orientations between the idealized symmetrical structure of R.PabI dimer

(two identical copies of chain A) and idealized B-DNA 24-mer with the PabI site. The construction of an idealized dimer was necessary for docking because chain B exhibited missing density in the predicted DNA-binding interface. All 2000 orientations were filtered to retain only those matching the DNA-binding site that was predicted using PreDs (33) and clustered. The variant with the best shape complementarity and minimal number of clashes was refined manually to obtain symmetrical positioning of B-DNA to each monomer of R.PabI.

## RESULTS

### Preparation in wheat-germ-based cell-free protein synthesis system

Our attempts to establish recombinant plasmids for expression of R.PabI in *E. coli* strains were not successful, apparently because even the slightest expression of this 4 bp cutter is toxic to the cells. We were, however, able to prepare R.PabI in the wheat-germ-derived, cell-free translation system (19), in a native form and in a form with SeMet substitution, in amounts that were sufficient for crystal structure determination (Materials and methods section). After synthesis *in vitro*, the solution containing R.PabI, which is hyperthermostable (14), was heated at 90°C for activation (14) and for eliminating most of the endogenous proteins by centrifugation. R.PabI was purified through a Heparin–Sepharose affinity column. The above reaction gave 3 mg of R.PabI of approximately 90% purity. Selenomethionyl R.PabI was prepared using similar reactions, except for the inclusion of SeMet instead of methionine. The content of SeMet incorporated was calculated to be more than 99% of the total methionine residues from the ratio of methionine and SeMet in the reaction solution (see Materials and methods section).

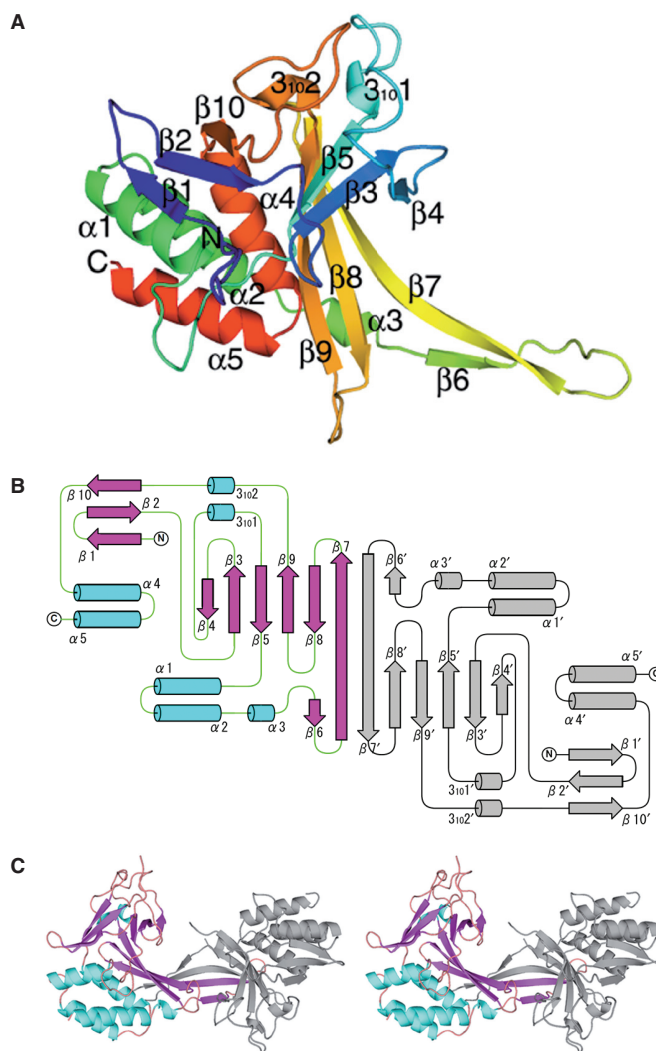
### Structure determination

The structure of R.PabI was solved by the single wavelength anomalous diffraction (SAD) method. The current model has been refined to an *R*-factor of 24.9% (*R*<sub>free</sub> 31.8%) as the diffraction data 20–3.0 Å resolution covers 96% of the total number of atoms (AB, CD and EF dimers of R.PabI). Due to the poor electron densities and low resolution, we could not build or sufficiently refine the structure of several residues in the N-termini of subunit B, D and F and in some loop regions. Though magnesium ion, frequently used in the catalytic reaction of Type II restriction endonucleases, was included in the SeMet R.PabI crystal, we could not detect binding of magnesium ion to SeMet R.PabI molecule. PROCHECK (36) analysis indicates that 96.3% of the residues were in the most favored or additional allowed regions in the Ramachandran plot (37).

### Overall structure: a novel fold

The overall structure and topology of the R.PabI protomer are illustrated in Figure 2A and B. The R.PabI protomer is composed of 10 β strands, 5 α helices

and 2 <sub>310</sub> helices. The R.PabI monomer folds into a α/β structure with a topology of ββββ<sub>310</sub>βαααββββ<sub>310</sub>βαα. Six β strands—β4–β3–β5–β9–β8–β7—form an extended antiparallel β-sheet that is curved to form an extended groove, which is the unique architecture of R.PabI. Approximately half of the convex surface of this β-sheet forms a sandwich with a nearly perpendicular additional antiparallel β-sheet that is formed by the N-terminal β-hairpin (strands β1 and β2) and strand β10 (see also Supplementary Figure S2). Finally, two pairs of α-helices form an interlocked bundle, which is partially inserted into one side of the sandwich and covers the other half of the convex surface of the main β-sheet.



**Figure 2.** Structure of R.PabI. (A) The protomer structure of R.PabI (subunit A). Color coding runs from blue at the N-terminal region to red at the C-terminal region. Secondary structure assignments are labeled on the ribbon model. (B) Topology diagram of R.PabI dimer. Helices and strands are shown as cylinders and arrows, respectively. Antiparallel β-sheets consisting of β4–β3–β5–β9–β8–β7–β7′–β6′ and β6–β7–β7′–β8′–β9′–β5′–β3′–β4′ form two half-pipe structures. (C) Stereo diagram of the dimer structure of R.PabI. One protomer is colored in light blue (α and <sub>310</sub> helices) and pink (β strands), and the other is colored in gray.



Searches of the Protein Data Bank with a number of programs (DALI, VAST, TOPSCAN, FATCAT and SSM) indicate that there are no proteins with a globally similar 3D structure to the overall fold of R.PabI. We found only partial matches: the fragment 134–207 of R.PabI shows limited similarity to the gelsolin fold (Supplementary Figure S1), whereas the helical bundle (aa 78–104 and 194–222) has an analogous substructure in a globally dissimilar fold of a carotenoid-binding protein (1m98 in the PDB, data not shown). Thus, the R.PabI monomer exhibits a novel fold. This fold is partially similar to that of the  $\beta$  barrel motif in lipocalin present in human tears (38), although R.PabI possesses only half of a side of the barrel and exhibits a different  $\beta$ -strand topology. Thus, we refer to this fold as a ‘half pipe’.

Homodimerization of R.PabI is mediated by the very long  $\beta 7$  strand (Figure 2A) of both protomers and a highly curved anti-parallel  $\beta$ -sheet is formed by strands of both protomers (residue range of 128–141). By this intermolecular interaction, the anti-parallel  $\beta$ -sheet forming the half pipe structure is extended by 2  $\beta$ strands ( $\beta 4$ – $\beta 3$ – $\beta 5$ – $\beta 9$ – $\beta 8$ – $\beta 7$ – $\beta 6'$ ; Figure 2B and C). The buried surface area of the R.PabI dimer is 3198 Å<sup>2</sup>, which is equivalent to 12.8% of the total surface area.

### Prediction of active sites

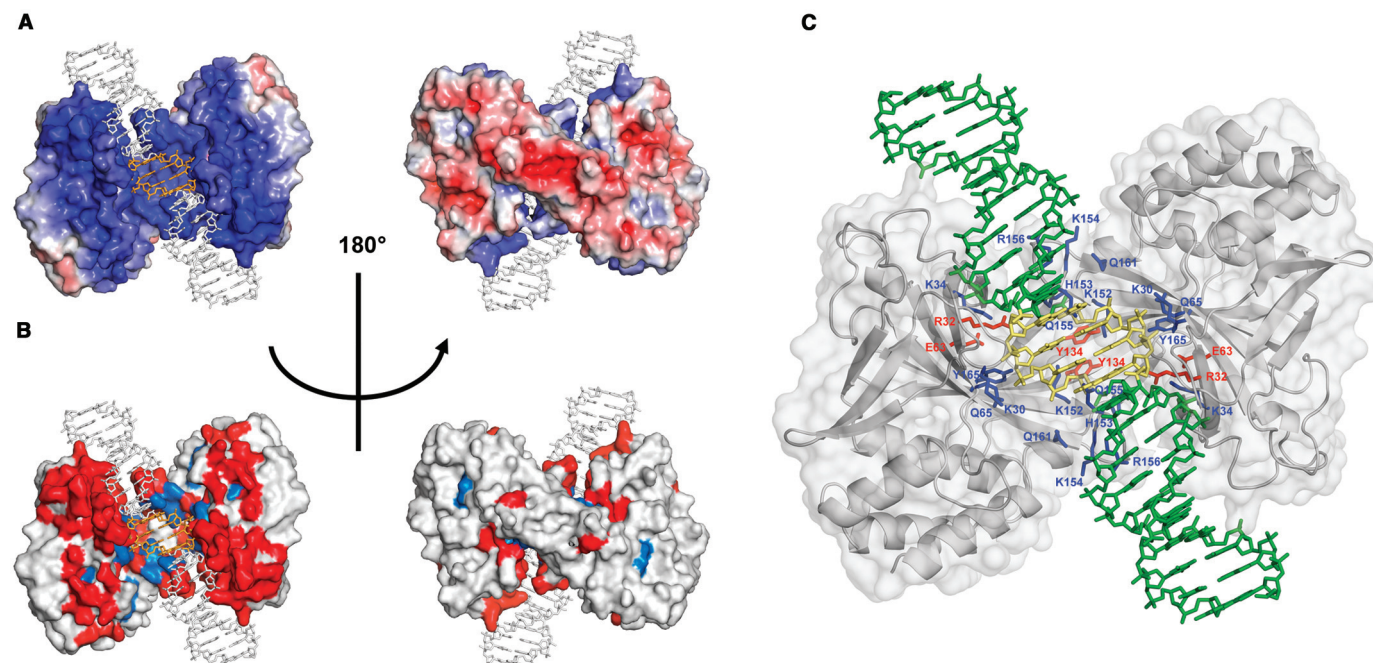
Mapping of the electrostatic potential on the molecular surface of dimeric R.PabI reveals that a large patch with

positive charges encompasses the extended groove that is formed between two protomers (Figure 3A), which suggests that this extended groove can interact with the negatively charged target DNA. In fact, this groove was also predicted to be the DNA-binding region by the PreDs program (Figure 3B), which takes into account the shape and electrostatic properties of the molecular surface (33). Two pairs of  $\beta$ -hairpins ( $\beta 3/\beta 4$  and  $\beta 8/\beta 9$ ) protrude from each monomer into the concave side of the sheet, forming a very rugged groove (Figure 2C).

Structural analysis of R.PabI also enables prediction of residues involved in sequence recognition and catalysis of phosphodiester bond cleavage. Charged or neutral residues located in the groove, such as Lys30, Arg32, Lys34, Glu63, Gln65, Tyr134, Lys152, His153, Lys154, Gln155, Arg156 and Gln161, could be involved in sequence recognition and cleavage. Interestingly, some of these residues were predicted to be involved in catalysis and/or DNA binding by amino acid sequence alignment (14).

### Mutant analysis

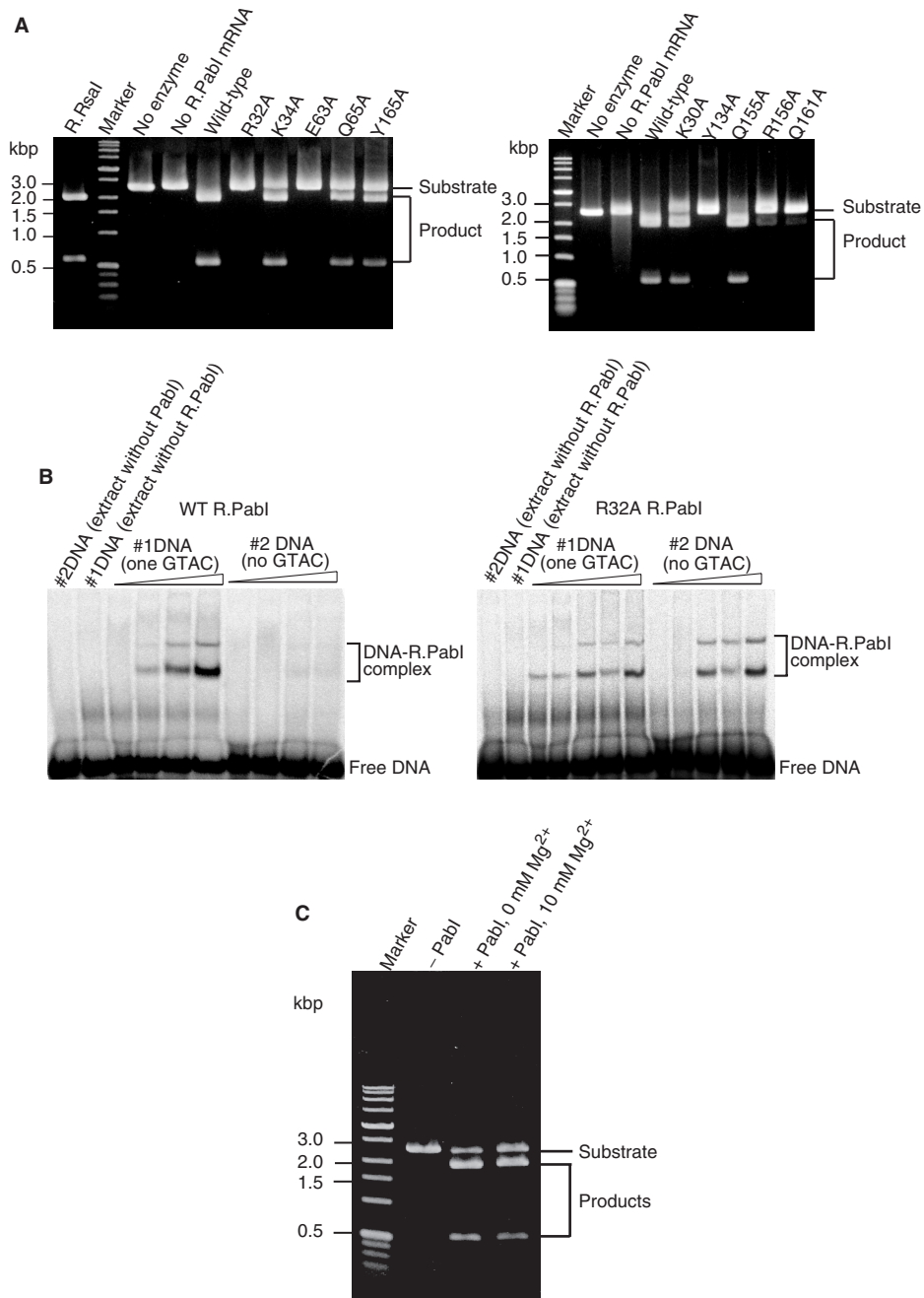
To evaluate functional importance, we constructed alanine-substitution mutants for 13 residues—Lys30, Arg32, Lys34, Glu63, Gln65, Tyr134, Lys152, His153, Lys154, Gln155, Arg156, Gln161 and Tyr165—and expressed them in the wheat-germ-based *in vitro* translation system. Specific DNA cleavage by approximately the same amount (as estimated by SDS-PAGE) of the soluble form



**Figure 3.** A model of R.PabI complexed with DNA. (A) Mapping of the electrostatic potential (as calculated by APBS tools) onto the surface of the R.PabI structure in a low-resolution model of a complex with DNA. The blue color indicates a positive charge and the red color indicates a negative charge. The DNA is shown in gray, with specifically recognized nucleotides in light orange. (B) DNA-binding residues (predicted by PreDs) mapped onto the surface of R.PabI structure in a low-resolution model of a complex with DNA. The residues that are predicted to bind DNA are shown in red, whereas the residues that are predicted not to be involved in DNA binding are shown in light gray. The DNA is shown in green, with specifically recognized nucleotides in light orange. (C) Functionally important residues of R.PabI. The protein is shown as gray ribbons together with the transparent surface of a space-filling view. The DNA is shown in green, with specifically recognized nucleotides in yellow. Residues in which a mutation abolished cleavage activity of R.PabI are shown as red sticks. Residues in which a mutation decreased cleavage activity of R.PabI are represented as blue sticks.

of each mutant enzyme was examined with a linear substrate DNA with a single recognition site. As a result, Arg32Ala, Glu63Ala and Tyr134Ala mutants showed no detectable cleavage activity, whereas Lys30Ala, Lys34Ala,

Gln65Ala, Lys152Ala, His153Ala, Lys154Ala, Gln155Ala, Arg156Ala, Gln161Ala and Tyr165Ala showed a lower activity than the wild-type R.PabI enzyme (Figure 4A and data not shown).



**Figure 4.** Functional analysis. (A) DNA cleavage. A linear DNA substrate was cleaved with approximately the same amount of each mutant R.PabI enzyme at 75°C for 1 h. Marker: Perfect DNA<sup>TM</sup> Markers, 0.1–12 kb (Novagen). R.RsaI: cut with R.RsaI, a neoschizomer of R.PabI. No enzyme: only the substrate DNA was incubated. No R.PabI mRNA: the substrate DNA was incubated with wheat germ extract that had not been treated with R.PabI mRNA. After electrophoresis through 1% agarose gel, DNA was visualized under ultraviolet light after ethidium bromide staining. (B) Binding to DNA with and without a recognition sequence. The DNA binding reactions were carried out with 5 nM of a duplex oligonucleotide either with one (#1) or no (#2) R.PabI site (Materials and methods section) with increasing amounts of each enzyme preparation (1.3, 2.6, 4.0, 5.3, 6.6 μl). The ‘no enzyme’ controls contain 5.3 μl of wheat-germ extract processed without mRNA. The shift of electrophoretic mobility was examined with 8% native polyacrylamide gel. (C) R.PabI DNA cleavage in the absence of divalent cation. The reaction mixture was incubated for 1 h at 85°C. Marker: Perfect DNA<sup>TM</sup> Markers, 0.1–12 kb (Novagen). Substrate: a linear DNA substrate with a single recognition site for R.PabI. +R.PabI, 0 mM Mg<sup>2+</sup>: the substrate DNA was cleaved with a suboptimal concentration of R.PabI in buffer A containing 1 mM EDTA. +R.PabI, 10 mM Mg<sup>2+</sup>: the substrate DNA was cleaved with the same concentration of R.PabI in buffer A containing 10 mM Mg<sup>2+</sup> (the same as for Figure 4A).

### A model for DNA binding

To predict the DNA-binding mode of R.PabI, we performed computational docking of a 24-mer ideal B-DNA to the idealized symmetrical structure of the R.PabI dimer (constructed from two identical copies of chain A). The largest clusters of docking solutions were consistent with PreDs prediction, by which DNA should bind in the positively charged cleft. Figure 3 shows an idealized averaged docking solution (corrected manually to obtain symmetrical positioning of B-DNA to each monomer of R.PabI). The binding surfaces show a high degree of complementarity—loops comprising of aa 154–159 and aa 43–49 make contacts in the major groove of DNA, and aa 25–29 make contacts in the minor groove. However, the fit is not perfect, and a number of protein–DNA clashes exist. Apparently, the structure of the R.PabI dimer, especially the loop regions that interact with DNA, and/or the DNA itself undergo conformational changes upon complex formation (e.g. the DNA might be bent or distorted in some other way).

Despite the imperfect fit of B-DNA in the groove of the R.PabI structure, the docking model is consistent with the results of site-directed mutagenesis with the mutants. Indeed, according to the docking model, side chains of all residues except Glu63 and Tyr165 (Arg32, Lys34, Gln65, Lys152, His153, Lys154) can form close contacts with DNA (Figure 3C).

### Functional analyses

We examined DNA binding of R.PabI in an electrophoretic mobility shift assay (Figure 4B left). Its binding to an oligoduplex with one recognition sequence (see Materials and methods for sequence) was stronger than that to an oligoduplex without one. We then analyzed Arg32 residue, which is located in the large groove (Figure 3C) and was shown to be essential for the cleavage activity (see above). Binding of its alanine-substitution (R32A) mutant protein turned out to be comparable for the cognate DNA and non-cognate DNA (Figure 4B right), a result indicating that Arg32 is responsible for specific binding to the recognition sequence. This finding explains why the R32A mutant is defective in cleavage activity and provides further support for a model in which the groove serves as the site of DNA binding.

Requirement of  $Mg^{2+}$  or other divalent cations for DNA cleavage is a general feature of Type II restriction enzymes with the exception of R.BfiI, a member of the PLD/Nuc family (8,39). Interestingly, R.PabI was able to cleave DNA in the absence of added  $Mg^{2+}$  (Figure 4C, lane 3) as well as in the presence of 10 mM  $Mg^{2+}$  (lane 4). DNA cleavage was observed even in the presence of 10 mM EDTA (data not shown). These results indicate the unique nature of R.PabI and its interaction with the DNA.

### DISCUSSION

The crystal structure of R.PabI restriction enzyme reported in this work reveals a new 3D fold. This result provides a proof of principle for our strategy in the search for

restriction enzymes of novel folds through the following steps: (i) to compare closely related genome sequences to find evidence of recent horizontal gene transfer of a putative restriction–modification gene complex, (ii) to identify a methyltransferase gene through its conserved motifs, (iii) to identify the restriction enzyme candidate as an ‘ORFan’ with no detectable sequence similarity to any protein family and (iv) to predict whether it is compatible with known folds or is likely to assume a new fold. One candidate protein obtained by this approach indeed showed restriction enzyme activity (14), and here we demonstrate that it assumes a novel fold. This result demonstrates that bioinformatics can be useful not only for the identification of homologies to previously known structures, but also for the prioritization of candidates for experimental validation of potential new folds.

Toxicity to cells represents a serious problem in expressing proteins for structure determination. The wheat-germ-based, cell-free translation system was able to bypass this problem by providing the R.PabI protein in a sufficient amount and quality for crystal structure determination. SeMet labeling *in vivo* often results in low incorporation and low productivity. The low incorporation causes heterogeneity in the protein sample, which is not good for crystallization. In contrast, the wheat-germ cell-free system has the advantage of efficient SeMet labeling. To our knowledge, this report represents the first scientific publication of crystal structure determination using protein that has been entirely prepared by the wheat-germ-based cell-free translation system.

In this report, we determined the crystal structure of R.PabI at 3.0 Å resolution. Homodimer R.PabI forms a highly curved anti-parallel  $\beta$  sheet with a positively charged, extended groove on one side and a negatively charged surface on the other. Comparison with the known protein structures suggests that R.PabI exhibits a novel fold. Mutational and *in silico* analyses of DNA-binding indicate that R.PabI binds double-stranded DNA in the groove (Figure 3). R.PabI also provides a novel model of DNA binding. Although TATA-box binding protein (TBP) recognizes TATA sequence by its curved anti-parallel  $\beta$  sheet as R.PabI (40), its mode of DNA binding is different from that modeled for R.PabI with DNA (Figure 3C) with respect to the extent of DNA bending and the mutual orientation of the protein and the DNA. The exact mode of protein–DNA interactions remains to be characterized through high-resolution analysis of the co-crystal structure. We recently obtained R.PabI–DNA co-crystals, which, however, have not yet been solved completely (Miyazono, K. Watanabe, M. Nagata, K. Kobayashi, I. and Tanokura, M. unpublished data).

Mutational analysis has revealed that there are three residues, Arg32, Glu63, and Tyr134 (present in  $\beta$ 3,  $\beta$ 5 and  $\beta$ 7, respectively), which are indispensable for the catalytic activity of R.PabI. This evidence suggests that these residues would be the catalytic residues. This prediction is consistent with the results of docking simulation. The distance between the two putative catalytic sites of the R.PabI dimer is approximately 18 Å (the distance between two midpoints of putative catalytic residues).



Because the residues included in the  $\beta$ -strands forming the curved  $\beta$ -sheet possess lower B values than the residues in the other parts, the structure of the DNA-binding groove and, especially, the distance of the two catalytic sites of an R.PabI dimer, will not change significantly by DNA binding. On the other hand, the structure of loops connecting the  $\beta$  strands ( $\beta 3$ – $\beta 4$  and  $\beta 8$ – $\beta 9$ ), which show higher B values, will change by DNA binding. Meanwhile, DNA structure bound to R.PabI will be significantly similar to the ideal B-form DNA in this model. If B-DNA is cleaved with a 3'-TA overhang, the distance between two scissile phosphates is about 13 Å. The contacts between the proposed catalytic residues and the scissile phosphates in the DNA cannot be modeled with confidence. However, as shown in Figure 3C, the position of both scissile phosphates accurately matches the position of essential residues; hence, we expect that only a minor adjustment of the R.PabI active site occurs at DNA binding. Our preliminary DNA binding analysis of the extract containing the R32A mutant (Figure 4B) suggested that Arg32 is also involved in specific recognition of the recognition sequence. The definitive conclusion would require the use of purified R.PabI and the R32A mutant, preferably in connection with the analysis of the R.PabI-DNA co-crystal structure that remains to be solved.

Tyr134 is present in  $\beta 7$ , with which two R.PabI protomers interact to form a dimer (Figure 2B and C). We do not yet know whether the lack of detectable restriction enzyme activity in the Y134A mutant (Figure 4A) is due to its defect in dimer formation. The dimerization will be discussed again below.

The crystal structure of R.PabI has revealed a novel fold with a new putative active site, which was previously never observed for any restriction endonucleases, thus raising the number of experimentally solved 'restriction endonuclease folds' to three [PD-(D/E)XK, PLD and half pipe]. The spatial localization of R.PabI secondary structure elements is similar to that in PD-(D/E)XK enzymes, but their directionality and connectivity are different and significantly more complex (Supplementary Figure S2 A, B), arguing for its independent evolutionary origin. R.PabI is also topologically different from other two-layer proteins that use  $\beta$ -sheet for DNA binding, e.g. nucleases from the LAGLIDADG superfamily or the TBP (Supplementary Figure S2).

The great majority of restriction enzymes, in particular members of the well-studied PD-(D/E)XK superfamily, require  $Mg^{2+}$  (8). The requirement of divalent cations was observed also for Type II restriction enzymes from two other superfamilies, e.g. MboII, from the HNH superfamily (11,41,42), and MraI from the GIY-YIG superfamily (11,43). Thus far, R.BfiI, a member of the PLD superfamily, was the only restriction enzyme of experimentally determined structure that cleaves DNA in the absence of metal ions (39). Here, we demonstrated that R.PabI, despite having no structural similarity to R.BfiI, also cleaves DNA in the absence of the  $Mg^{2+}$  ion (Figure 4C). This is in accordance with the absence of any metal ion peaks in X-ray analysis of a SeMet R.PabI crystal prepared in the presence of  $Mg^{2+}$  and provides

another line of evidence for the unique nature of the cleavage reaction by R.PabI and, by extrapolation, its homologs. PD-(D/E)XK family requires  $Mg^{2+}$  (8), but PLD/Nuc family does not (39). Requirement of divalent cations was observed with MboII, a Type II restriction enzyme similar to HNH family of homing endonucleases (11,41,42), and MraI, a Type II restriction enzyme similar to GIY-YIG family of homing endonucleases (11,43).

R.PabI protomers form a dimer structure, as is the case with other Type II restriction endonucleases that recognize a palindrome. Because the dimerization mode is an important factor for the determination of the cleavage pattern of DNA, there is a moderate correlation between quaternary structures of restriction enzymes and their DNA cleavage patterns (44). Restriction endonucleases that cleave DNA with a 5' overhang, such as R.EcoRI and R.BamHI, dimerize through contact of helices of the core domain (a 'forehead' of the catalytic domain, where the active site corresponds to a 'mouth'), whereas enzymes that produce blunt DNA ends, such as R.EcoRV and R.PvuII, dimerize using mainly the N-terminal extension (a 'chin' of the catalytic domain). Restriction endonucleases that generate a 3' overhang, exemplified by R.BglII and R.PabI, dimerize using one side of the subunit (a 'cheek' of the catalytic domain). R.PabI dimerization involves interaction between  $\beta$ -strands that protrude from the protein core of the monomers, which leads to mutual extension of both  $\beta$ -sheets. This dimerization mode is similar to that of R.BglII, despite the absence of tertiary structural similarity on the level of the monomer.

The R.PabI homologs in Epsilon proteobacteria share overall organization in secondary structure with R.PabI (Figure 1). However, they are more similar to each other than to R.PabI; for example, in the regions of  $\beta 5$  to  $\alpha 1$ ,  $\beta 6$  to  $\beta 7$  and around  $3_{10}2$ . The region between  $\beta 5$  to  $\alpha 1$  of R.PabI homologs is rich in acidic and basic residues and poor in hydrophobic residues and, therefore, may form a flexible loop. We do not yet know whether these differences are related to the biology of these archaeal and eubacterial groups or to the hyperthermophilicity of R.PabI.

PabI sites (5' GTAC) are not rare in *Pyrococcus* genomes but are extremely rare in *Helicobacter* genomes (REBASE), probably through the selection by past attacks of the R.PabI homolog. It is possible that PabI family were originally present in Epsilon-proteobacteria and invaded *Pyrococcus* more recently so that the number of sites has not yet dropped. It is also possible that *Pyrococcus* genomes with archaeal chromatin are more resistant to these enzymes than eubacterial *Helicobacter* genomes (45).

## ACCESSION CODES

Protein Data Bank: Coordinates and structure factors have been submitted to the RCSB Protein Data Bank with Accession Codes 2DVY.

## SUPPLEMENTARY DATA

Supplementary data are available on the NAR online.

## ACKNOWLEDGEMENTS

The synchrotron-radiation experiments were performed at BL-5A and NW-12 in the Photon Factory (Tsukuba, Japan) (Proposal No. 2003S2-002). This work was supported in part by the grants to M.T., I.K. and Y.E. from the National Project on Protein Structural and Functional Analyses (Protein 3000) of the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan. I.K. and M.W. were supported in part by the 21<sup>st</sup> century COE project of 'Elucidation of Language Structure and Semantic behind Genome and Life System' and the 'Grants-in-Aid for Scientific Research' from Japan Society for the Promotion of Science (JSPS) to I.K. (13141201, 15370099 and 17310113) and M.W. (1654071). M.W. was a JSPS Research Fellow (DC-2). J.M.B. and J.K. were supported by NIH (Fogarty International Center grant R03 TW007163-01). J.K. is the recipient of a scholarship from the Postgraduate School of Molecular Medicine at the Medical University of Warsaw. Y.E. and T.S. were partially supported by Special Coordination Funds for Promoting Science and Technology by MEXT. Funding to pay the Open Access publication charge was provided by MEXT.

*Conflict of interest statement.* None declared.

## REFERENCES

- Naito, T., Kusano, K. and Kobayashi, I. (1995) Selfish behavior of restriction-modification systems. *Science*, **267**, 897–899.
- Kobayashi, I. (2001) Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res.*, **29**, 3742–3756.
- Kobayashi, I. (2004) Restriction-modification systems as minimal forms of life. In Pingoud, A. (ed), *Restriction Endonucleases* Springer, Berlin, pp. 19–62.
- Nobusato, A., Uchiyama, I., Ohashi, S. and Kobayashi, I. (2000) Insertion with long target duplication: a mechanism for gene mobility suggested from comparison of two related bacterial genomes. *Gene*, **259**, 99–108.
- Alm, R.A., Ling, L.S., Moir, D.T., King, B.L., Brown, E.D., Doig, P.C., Smith, D.R., Noonan, B., Guild, B.C. *et al.* (1999) Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature*, **397**, 176–180.
- Handa, N., Nakayama, Y., Sadykov, M. and Kobayashi, I. (2001) Experimental genome evolution: large-scale genome rearrangements associated with resistance to replacement of a chromosomal restriction-modification gene complex. *Mol. Microbiol.*, **40**, 932–940.
- Bujnicki, J.M. (2003) Crystallographic and bioinformatic studies on restriction endonucleases: inference of evolutionary relationships in the "midnight zone" of homology. *Curr. Protein Pept. Sci.*, **4**, 327–337.
- Pingoud, A., Fuxreiter, M., Pingoud, V. and Wende, W. (2005) Type II restriction endonucleases: structure and mechanism. *Cell Mol. Life Sci.*, **62**, 685–707.
- Grazulis, S., Manakova, E., Roessle, M., Bochtler, M., Tamulaitiene, G., Huber, R. and Siksnys, V. (2005) Structure of the metal-independent restriction enzyme BfiI reveals fusion of a specific DNA-binding domain with a nonspecific nuclease. *Proc. Natl. Acad. Sci. USA*, **102**, 15797–15802.
- Aravind, L., Makarova, K.S. and Koonin, E.V. (2000) Holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. *Nucleic Acids Res.*, **28**, 3417–3432.
- Bujnicki, J.M., Radlinska, M. and Rychlewski, L. (2001) Polyphyletic evolution of type II restriction enzymes revisited: two independent sources of second-hand folds revealed. *Trends Biochem. Sci.*, **26**, 9–11.
- Saravanan, M., Bujnicki, J.M., Cymerman, I.A., Rao, D.N. and Nagaraja, V. (2004) Type II restriction endonuclease R.KpnI is a member of the HNH nuclease superfamily. *Nucleic Acids Res.*, **32**, 6129–6135.
- Chinen, A., Uchiyama, I. and Kobayashi, I. (2000) Comparison between *Pyrococcus horikoshii* and *Pyrococcus abyssi* genome sequences reveals linkage of restriction-modification genes with large genome polymorphisms. *Gene*, **259**, 109–121.
- Ishikawa, K., Watanabe, M., Kuroita, T., Uchiyama, I., Bujnicki, J.M., Kawakami, B., Tanokura, M. and Kobayashi, I. (2005) Discovery of a novel restriction endonuclease by genome comparison and application of a wheat-germ-based cell-free translation assay: PabI (5'GTA/C) from the hyperthermophilic archaeon *Pyrococcus abyssi*. *Nucleic Acids Res.*, **33**, e112.
- Watanabe, M., Yuzawa, H., Handa, N. and Kobayashi, I. (2006) Hyperthermophilic DNA methyltransferase M.PabI from the archaeon *Pyrococcus abyssi*. *Appl. Environ. Mol.*, **72**, 5367–5375.
- Pavlov, M.Y. and Ehrenberg, M. (1996) Rate of translation of natural mRNAs in an optimized in vitro system. *Arch. Biochem. Biophys.*, **328**, 9–16.
- Vinarov, D.A., Lytle, B.L., Peterson, F.C., Tyler, E.M., Volkman, B.F. and Markley, J.L. (2004) Cell-free protein production and labeling protocol for NMR-based structural proteomics. *Nat. Methods*, **1**, 149–153.
- Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., Blumenthal, R.M., Degtyarev, S., Dryden, D.T. *et al.* (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.*, **31**, 1805–1812.
- Sawasaki, T., Hasegawa, Y., Tsuchimochi, M., Kamura, N., Ogasawara, T., Kuroita, T. and Endo, Y. (2002) A bilayer cell-free protein synthesis system for high-throughput screening of gene products. *FEBS Lett.*, **514**, 102–105.
- Madin, K., Sawasaki, T., Ogasawara, T. and Endo, Y. (2000) A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. *Proc. Natl. Acad. Sci. USA*, **97**, 559–564.
- Sawasaki, T., Ogasawara, T., Morishita, R. and Endo, Y. (2002) A cell-free protein synthesis system for high-throughput proteomics. *Proc. Natl. Acad. Sci. USA*, **99**, 14652–14657.
- Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
- Matthews, B.W. (1968) Solvent content of protein crystals. *J. Mol. Biol.*, **33**, 491–497.
- Weeks, C.M. and Miller, R. (1999) The design and implementation of SnB version 2.0. *J. Appl. Cryst.*, **32**, 120–124.
- De La Fortelle, E. and Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.*, **276**, 472–494.
- Terwilliger, T.C. (2000) Maximum-likelihood density modification. *Acta Crystallogr. Sect. D. Biol. Crystallogr.*, **56**, 965–972.
- Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M. *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. Sect. D. Biol. Crystallogr.*, **54**, 905–921.
- McRee, D.E. (1999) XtalView/Xfit—A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.*, **125**, 156–165.
- Tepljakov, A. and Vagin, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.*, **30**, 1022–1025.
- Collaborative Computational Project, N. (1994) The CCP4 suite: programs for protein crystallography. *Acta Cryst.*, **D50**, 760–763.
- Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D. Biol. Crystallogr.*, **53**, 240–255.
- Baker, N.A., Sept, D., Joseph, S., Holst, M.J. and McCammon, J.A. (2001) Electrostatics of nanosystems: application to

- microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA*, **98**, 10037–10041.
33. Tsuchiya, Y., Kinoshita, K. and Nakamura, H. (2005) PreDs: a server for predicting dsDNA-binding site on protein molecular surfaces. *Bioinformatics*, **21**, 1721–1723.
34. Tsuchiya, Y., Kinoshita, K. and Nakamura, H. (2004) Structure-based prediction of DNA-binding sites on proteins using the empirical preference of electrostatic potential and the shape of molecular surfaces. *Proteins*, **55**, 885–894.
35. Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A.A., Aflalo, C. and Vakser, I.A. (1992) Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. USA*, **89**, 2195–2199.
36. Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, **26**, 283–291.
37. Ramachandran, G.N. and Sasisekharan, V. (1968) Conformation of polypeptides and proteins. *Adv. Protein Chem.*, **23**, 283–438.
38. Breustedt, D.A., Korndorfer, I.P., Redl, B. and Skerra, A. (2005) The 1.8-Å crystal structure of human tear lipocalin reveals an extended branched cavity with capacity for multiple ligands. *J. Biol. Chem.*, **280**, 484–493.
39. Sapranaukas, R., Sasnauskas, G., Lagunavicius, A., Vilkaitis, G., Lubys, A. and Siksnys, V. (2000) Novel subtype of type IIs restriction enzymes. BfiI endonuclease exhibits similarities to the EDTA-resistant nuclease Nuc of *Salmonella typhimurium*. *J. Biol. Chem.*, **275**, 30878–30885.
40. Kim, Y., Geiger, J.H., Hahn, S. and Sigler, P.B. (1993) Crystal structure of a yeast TBP/TATA-box complex. *Nature*, **365**, 512–520.
41. Bujnicki, J.M. (2004) Molecular phylogenetics of restriction endonucleases. In Pingoud, A. (ed), *Restriction Endonucleases* Springer, Berlin, pp. 63–93.
42. Bocklage, H., Heeger, K. and Muller-Hill, B. (1991) Cloning and characterization of the MboII restriction–modification system. *Nucleic Acids Res.*, **19**, 1007–1013.
43. Wani, A.A., Stephens, R.E., D'Ambrosio, S.M. and Hart, R.W. (1982) A sequence specific endonuclease from *Micrococcus radiodurans*. *Biochim. Biophys. Acta*, **697**, 178–84.
44. Deibert, M., Grazulis, S., Janulaitis, A., Siksnys, V. and Huber, R. (1999) Crystal structure of MunI restriction endonuclease in complex with cognate DNA at 1.7 Å resolution. *EMBO J.*, **18**, 5805–5816.
45. Kobayashi, I., Ishikawa, K. and Watanabe, M. (2007) Chromatin as anti-restriction adaptation: a hypothesis based on restriction enzymes of a novel structure. *Proceedings of International Symposium on Extremophiles and their Applications*, **205**, 167–174.