## Tetrameric structure of the restriction DNA glycosylase R.Pabl in complex with nonspecific double-stranded DNA

Delong Wang, Ken-ichi Miyazono, and Masaru Tanokura\*

Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan

\*To whom correspondence should be addressed. Tel: +81-3-5841-5165; Fax: +81-3-5841-8023; Email: amtanok@mail.ecc.u-tokyo.ac.jp

PabI	1	β1 >TT	β2	$\beta 3$	η1 β4	η2 22222-
PabI S.hellenicus_DSM_12710 C.australicus_RC3 H.pylori_HPAGI H.bilis_ATCC_43879 H.cinaedi_PAGU611 C.coli_H56	MIHLTSVE .MINMVE .MAKSIIN .MAKSIIN .MKI .MKI	ASVSFENGK CEVRWEDGR PTITIEEGK IRIDNNKKV QKCENQQ QQCENKQ MNFK	IVIRLPITRPT IVIGLLPITLPT IHINFFLKNNA IEVSIPLTSIS IFVEIPLTTQS IFVEIPLTTQS IDYELPLTSVA	SKIRVKKI SKVRVKRN ASKIRLKER GKVRVKIR GKTRVKTR GKTRVKTR GKIRIKOR	ENGVGIPVS G.E. PIP RNYMGERFL HAFSDYGIS NSFYEYGLP NSFYEYGLP STFNDYGLP O	TRKKSFPSDENLRDY ARQTKLREDD RPDEALKVDSA TATRKIPFSLKH TATRQIPFSQKH TATRQIPFSQKH VAPTKININVKH OOOOO
PabI PabI S.hellenicus_DSM_12710 C.australicus_RC3 H.pylori_HPAGI H.bilis_ATCC_43879 H.cinaedi_PAGU611 C.coli_H56	β5 Y IEWQISY LLEWQISY YVEWQIGY YVEWQIGY YIEWQIGY YIEWQIGY YUEWQIGY	TT 79 ARDGK KKDEE DTKVKKDGD DVPIKDKEK DVDKNDKEK DVDKNDKEK DVDKNDKEK MVAG	YDY VLI YKVQHLEGDNY FELTIKDEKY LALSTLGQ LALSTLRQ 	LYIRERNG HFLGANG. EFIGANG. SNFIGANG.	E QRVKKYPAE KVKLYE KTKALYE KTKALYE KTKALYK KDKKLYE	αl 80 90 LSRMVRLAHEHGILT AGKMLEIAYNRGIIT LMVIIKYALDLGFLR LSEMIYYAKQLGLIS LSEYIYYFTQWGIIT LSEYLYYFTQWGIIT LSDIIFQFFKHNIIL
PabI S.hellenicus_DSM_12710 C.australicus_RC3 H.pylori_HPAGI H.bilis_ATCC_43879 H.cinaedi_PAGU611 C.coli_H56	COORDER CONTRACTOR CON	XFADDV DYAVKV NLTEEYES KYLEKQKQ. AFLQNIKD. AFLQNIKD. NFLENNEE.	α3 222.2 110 KSYLED.KC PQTFDKQFF KNYLDNFF DEFLDSRSDLC DEFLDSRSDLC DEFLDSRSDLC DEFLDSRSDLC DEFLDSRSDLC	β6 120 LKESTNR LLGEKTDN LLSHPVS LLSHPVS LLSHPVS LLSHFVS	TT	β7 140 EDVYPVAKKEL.P FRHIPIIHKDL.D LSRISYPLLIHSFDD ESQVKYPLLVHKFSH ESQVKYPLLVHKFSH ESVVYPLLVKFNN
PabI PabI S.hellenicus_DSM_12710 C.australicus_RC3 H.pylori_HPAGI H.bilis_ATCC_43879 H.cinaedi_PAGU611 C.coli_H56	β8       T     150       SGEFIGIV     NGCFVEAE       KDVFIEME     N.QLSEIV       FDVLVEII     FDVLVEII       FDVLVEII     NEFLSEII	16 LKHKORAVG LKHKORAVG IKEQOYGSK IKEKORAVG IKEKORAVG IKEKORAJG	β9 17 20 20 20 20 20 20 20 20 20 20	η3 TNVE RNVVTNDA KILKK LELKT TELHC TELHC HLLKNING	TT 180 . PSLAGRVA VGSLVGRTA  ATPLLNRTA SPVLLGRVA NPVLLGRVA ERNFLNRCI	β10   TT 190   RRNE VVKYE   VVKYE VPV   RPKE VVKWFPTK   KQENIYAWIADENNI ALKEHALLTIHKTNA   ESKE CGLLILDSKDK   ESKE CGLLILDSKDK   ESKE KQYLEISRNNI   OO OO
PabI S.hellenicus DSM 12710 C.australicus RC3 H.pylori_HPAGI H.bilis ATCC 43879 H.cinaedi_PAGU611 C.coli_H56	α4 QQQQQQQ 200 DLMKELLK NDIIEIMK NIFEKVLI LMFLEMLK CFLLEIFK NFLLEIFK NIFLEMLK	20000 200 210 AFIIASETH TFAVLSQKH SFALASEQH IFGLLSQAH IFGMLSKSH IFGMLSKNH IFGLLSNNH	α5 220 220 KNDTVKFIRS REDVIEIIER NKDIRELLEI HNDVLKILEK NYDVCEIVKII NYDVCEIVKII RYDVLQILEFI	IGTS ISKF LKNI KVNIKSLT KVNIKSLT LNSK		

Supplementary Figure 1. Amino acid sequence alignment of R.Pabl and its homologues. Invariant residues are highlighted with red boxes, and conserved residues are shown in red text. The secondary structure of R.Pabl in the R.Pabl-nonspecific dsDNA complex is indicated by helices ( $\alpha$  and  $\eta$  (3<sub>10</sub>)-helices), arrows ( $\beta$ -strand), and TT ( $\beta$ -turn). Residues that were mutated to alanine for the structure determination are indicated by black boxes. Residues that recognize the DNA bases and the sugar-phosphate backbones are marked with black and open circles, respectively. Residues analysed by mutagenesis are marked with black triangles. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 5' - G C A C T A G T T C G A A C T A G T G C - 3' 3' - C G T G A T C A A G C T T G A T C A C G - 5' 20'19'18'17'16'15'14'13'12'11'10' 9' 8' 7' 6' 5' 4' 3' 2' 1'

а



Supplementary Figure 2. Structure determination of the R.Pabl-nonspecific dsDNA complex. (**a**) Sequence of a dsDNA used for the structure determination of the R.Pabl-nonspecific dsDNA complex. (**b**) Wall-eyed stereo image of the R.Pabl-nonspecific dsDNA complex in one asymmetric unit. (**c**) Wall-eyed stereo image of the bound ssDNA and its composite omit map (1.7  $\sigma$ , magenta). The structure of DNA is shown as a stick model. (**d**) Composite omit map (1.7  $\sigma$ , magenta) of Arg70 and Asp71 of protomer A. (**e**) Composite omit map (1.7  $\sigma$ , magenta) of Arg70 and Asp71 of protomer B.



Supplementary Figure 3. Structure comparison of R.Pabl in the DNA-free state, the nonspecific dsDNA binding state, and the specific dsDNA binding state. (**a**) Wall-eyed stereo image of the superposition of the DNA-free state (magenta) and the nonspecific DNA binding state (green and cyan) of the R.Pabl dimers. (**b**) Wall-eyed stereo image of the superposition of the nonspecific DNA binding state (green and cyan) and the specific DNA binding state (yellow) of the R.Pabl dimers.



Supplementary Figure 4. The entire gel images of Figure 4.



Supplementary Figure 5. Gel filtration analysis of the R.Pabl mutants (cyan) and their nonspecific dsDNA complexes (red). The R.Pabl Y68F D71R and Y68F R70D D71R mutants were eluted with buffer containing 10 mM MES (pH 6.0) and 600 mM NaCl to avoid nonspecific binding to the column. The other R.Pabl mutants, the R.Pabl-nonspecific dsDNA complexes, and the nonspecific dsDNA were eluted with buffer containing 10 mM MES (pH 6.0) and 300 mM NaCl. The peak positions of the marker proteins are indicated by the black triangles at the top of the chromatogram. The molecular weight of DNA free R.Pabl Y68F, R32A E63A, Y68F R70D, Y68F D71R, and Y68F R70D D71R mutants are estimated to be 38,000, 38,000, 42,000, 38,000, and 40,000, respectively, indicating that R.Pabl mutants form homodimers in solution (the molecular weight of R.Pabl Y68F D71R, and Y68F R70D D71R mutants in complex with nonspecific dsDNA are estimated to be 86,000, 75,000, 78,000, 65,000, and 80,000, respectively. The R.Pabl Y68F R70D X68F R70D X6

X R.Pabl recognition site (5'-GTAC-3')

b

а

5'- fluorescein - G G A T G C A T G A G T A C G A G G A C C A T C - 3' 3'- C C T A C G T A C T C A T G C T C C T G G T A G -5'





Supplementary Figure 6. DNA glycosylase activities of the R.Pabl mutants. (a) Schematic diagram of three R.Pabl substrates (24 bp, 500 bp, and 3000 bp) used in the DNA glycosylase assays. The recognition sequence (5'-GTAC-3') is located at the centre of each substrate. After R.Pabl treatment and the cleavage of AP sites by NaOH, these substrates are separated into two products with equal lengths. (b) Sequence of dsDNA (24 bp) used for the DNA glycosylase assay. The R.Pabl recognition sequence is indicated by a red box. (c) DNA glycosylase activity assay using 24 bp dsDNA as a substrate. First,  $0.2 \mu$ M of the 24 bp substrate dsDNA and  $0.4 \mu$ M of the R.Pabl dimer were mixed and incubated at 45°C. Each experiment was repeated three times. (d) DNA glycosylase activity assay using 500 bp dsDNA as a substrate. First, 5.9 nM of the R.Pabl dimer were mixed and incubated at 45°C. Each experiment was repeated three times. (e) DNA glycosylase activity assay using 3000 bp dsDNA as a substrate. First, 5.9 nM of the substrate DNA and 80 nM of the R.Pabl dimer were mixed and 80 nM of the substrate DNA and 80 nM of the R.Pabl dimer were mixed and incubated at  $45^{\circ}$ C. Each experiment was a substrate. First, 5.9 nM of the substrate DNA and 80 nM of the R.Pabl dimer were mixed and incubated at  $45^{\circ}$ C. Each experiment was a substrate. First, 5.9 nM of the substrate DNA and 80 nM of the R.Pabl dimer were mixed and incubated at  $45^{\circ}$ C. Each experiment was repeated three times. (e) DNA glycosylase activity assay using 3000 bp dsDNA as a substrate. First, 5.9 nM of the substrate DNA and 80 nM of the R.Pabl dimer were mixed and incubated at  $45^{\circ}$ C. Each experiment was repeated three times.

Protein				H-bonds			DNA		
Chain	Residue	Main/	Atom	Distance	Chain	Base	Base/	Atom	
ID		Side	Name	(Å)	ID		Backbone	Name	
А	Thr46	Side	OG1	2.44	C'	Cyt4'	Backbone	OP2	
А	Arg47	Main	Ν	2.96	C'	Cyt4'	Backbone	OP2	
А	Arg47	Main	Ν	3.17	C'	Cyt4'	Backbone	O5'	
А	Lys48	Main	Ν	3.41	C'	Cyt4'	Backbone	OP2	
А	Lys48	Side	NZ	2.87	C'	Ade3'	Backbone	OP1	
А	Lys49	Main	Ν	2.68	C'	Ade3'	Backbone	OP2	
А	Ser50	Main	Ν	3.73	C'	Ade3'	Backbone	OP1	
А	Ser50	Main	Ν	3.38	C'	Ade3'	Backbone	OP2	
А	Ser29	Main	Ν	2.98	С	Ade12	Backbone	OP2	
А	Lys30	Main	Ν	3.83	С	Ade12	Backbone	OP1	
В	Thr25	Side	OG1	3.62	C'	Ade16'	Backbone	O3'	
В	Thr25	Side	OG1	2.51	C'	Gua17'	Backbone	OP1	
В	Ser45	Side	OG	3.70	C'	Ade16'	Backbone	OP1	
В	Arg47	Side	NE	2.82	C'	Gua17'	Backbone	OP2	
В	Arg156	Side	NH1	3.68	C'	Thy8'	Base	O4	
В	Arg156	Side	NH1	2.70	C'	Gua7'	Base	O6	
В	Val158	Main	Ν	3.38	C'	Thy5'	Backbone	OP1	
В	Val158	Main	Ν	3.29	C'	Thy5'	Backbone	OP2	
В	Arg184	Side	NE	3.24	C'	Thy18'	Backbone	OP2	
В	Asn185	Side	Ν	2.76	C'	Gua17'	Backbone	OP1	
В	Asn185	Side	ND2	3.17	C'	Gua17'	Backbone	O3'	
В	Asn185	Side	ND2	3.33	C'	Thy18'	Backbone	OP1	
В	Ser29	Main	Ν	2.76	С	Thy9	Backbone	OP1	
В	Ser29	Side	OG	2.76	С	Thy9	Backbone	OP1	

Supplementary Table 1. Hydrogen bonds between R.Pabl and DNA in the R.Pabl-nonspecific dsDNA complex.