J. Biochem. 2010;148(4):429–438 doi:10.1093/jb/mvq088

JB THE JOURNAL OF BIOCHEMISTRY

Crystal structures of glycinamide ribonucleotide synthetase, PurD, from thermophilic eubacteria

Received May 15, 2010; accepted July 3, 2010; published online August 16, 2010

Gen-ichi Sampei[1,2,*], Seiki Baba[2,3,4], Mayumi Kanagawa[2,3], Hisaaki Yanai[1,2], Takeshi Ishii[1], Hiroya Kawai[3], Yoko Fukai[3], Akio Ebihara[2,†], Noriko Nakagawa[2] and Gota Kawai[2,3,‡]

[1]Department of Applied Physics and Chemistry, Faculty of Electro-Communications, The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585; [2]RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5148; [3]Department of Life and Environmental Sciences, Faculty of Engineering, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino-shi, Chiba 275-0016; and [4]Synchrotron Radiation Research Institute/SPring-8, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan

*Gen-ichi Sampei, Department of Applied Physics and Chemistry, Faculty of Electro-Communications, The University of Electro-Communications, Chofu-shi, Tokyo 182-8585, Japan. Tel: +81 42 443 5481, Fax: +81 42 443 5501, email: sampei@pc.uec.ac.jp

‡Gota Kawai, Department of Life and Environmental Sciences, Faculty of Engineering, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino-shi, Chiba 275-0016, Japan. Tel: +81 47 478 0425, Fax: +81 47 478 0425, email: gkawai@sea.it-chiba.ac.jp

†Present address: Akio Ebihara, Faculty of Applied Biological Sciences, Gifu University, 1-1, Yanagido, Gifu, 501-1193, Japan.

Glycinamide ribonucleotide synthetase (GAR-syn, PurD) catalyses the second reaction of the purine biosynthetic pathway; the conversion of phosphoribosylamine, glycine and ATP to glycinamide ribonucleotide (GAR), ADP and Pi. In the present study, crystal structures of GAR-syn's from *Thermus thermophilus*, *Geobacillus kaustophilus* and *Aquifex aeolicus* were determined in apo forms. Crystal structures in ligand-bound forms were also determined for *G. kaustophilus* and *A. aeolicus* proteins. In general, overall structures of GAR-syn's are similar to each other. However, the orientations of the B domains are varied among GAR-syn's and the MD simulation suggested the mobility of the B domain. Furthermore, it was demonstrated that the B loop in the B domain fixes the position of the β- and γ- phosphate groups of the bound ATP. The structures of GAR-syn's and the bound ligands were compared with each other in detail, and structures of GAR-syn's with full ligands, as well as the possible reaction mechanism, were proposed.

Keywords: ligand binding/molecular dynamics/purine metabolism/structure of enzyme/x-ray crystallography.

Abbreviations: AMP, adenosine monophosphate; ADP, adenosine diphosphate; ATP, adenosine triphosphate; DTT, dithiothreitol; GAR, glycinamide ribonucleotide; GAR-syn, glycinamide ribonucleotide synthetase; GMP, guanosine monophosphate; MD, molecular dynamics; 2-ME, mercaptoethanol; MME, monomethyl ether; PCR, polymerase chain reaction; PEG, polyethylene glycol; Pi, inorganic phosphate; PMSF, phenylmethanesulphonyl fluoride; PRA, phosphoribosylamine; PRPP-AT, glutamine phosphoribosyl pyrophosphate amidotransferase; Se-Met, selenomethione, selenomethionyl.

The *de novo* synthesis of purine nucleotides (AMP and GMP) proceeds by a 14-step branched pathway via IMP and all reactions in this pathway are formation reactions of the C–N bond (*1*). This pathway is common to most organisms except for a part of parasitic protozoa and, thus, genetic, biochemical and structural analyses of this pathway will lead the understanding of the organization of a biosynthetic pathway. Under this concept, we are determining 3D structures of the enzymes in this pathway of a thermophilic eubacteria, *Thermus thermophilus* and some other thermophiles as well as hyperthermophiles. *Thermus thermophilus* can grow at highest temperature (50–82°C) among organisms whose molecular biology has been studied (*2*). It is expected that valuable information on the function of proteins can be obtained by the analysis of evolutionally selected and naturally existing proteins without using the artificial mutants. Proteins of thermophiles should be useful to study the structure–function relationship because they are thermostable and can be crystallized more easily than non-thermophilic proteins.

Glycinamide ribonucleotide synthetase (GAR-syn, EC6.3.4.13), also called as PurD, catalyses the second step of the *de novo* purine nucleotide biosynthetic pathway; the conversion of phosphoribosylamine (PRA), glycine and ATP to glycinamide ribonucleotide (GAR), ADP and Pi (Fig. 1). PRA, which is generated by glutamine phosphoribosyl pyrophosphate amidotransferase (PRPP-AT) or PurF, is unstable and a kinetic study has shown that PRA directly transferred from PurF to PurD (*3*).

Wang *et al.* (*4*) has determined the crystal structure of PurD from *Escherichia coli* (*Ec*PurD, PDB ID: 1GSO) and it was shown that PurD consists of four domains labelled N, A, B and C. The N, A and C domains are clustered to form a large central core

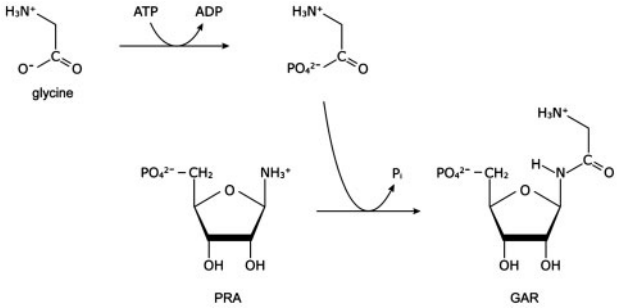

Fig. 1 The reaction catalysed by GAR-syn, also called as PurD.

structure whereas the smaller B domain is extended outward. PurDs from *Thermotoga maritima* (*Tm*PurD, 1VKZ) and human (*Hs*PurD, 2QK4) have also been determined and these three structures are similar to each other. However, the relative orientation of B domain is different; *Ec*PurD is in the open form and *Tm*PurD and *Hs*PurD seem to be in the closed form.

In the present study, we determined the crystal structures of PurDs from *T. thermophilus* (*Tt*PurD), *Geobacillus kaustophilus* (*Gk*PurD) and *Aquifex aeolicus* (*Aa*PurD). Structural characteristics, including the ligand binding and the reaction mechanism, of the three PurDs are discussed.

Materials and Methods

Expression and purification of TtPurD

Thermus thermophilus purD was amplified by PCR from *T. thermophilus* HB8 genomic DNA and was cloned into the pT7Blue (Novagen). After confirmation of the nucleotide sequence, the DNA fragment was ligated into the expression vector pET-11a (Novagen) at the *Nde*I/*Bam*HI sites.

The *Tt*PurD expression plasmid was transformed into *E. coli* strain B834 (DE3) (Novagen) for overexpression. *Escherichia coli* B834 (DE3) cells carrying the pET-11a plasmid were cultured in LB medium at 37°C for a few hours. The pre-culture was inoculated at a dilution ratio of 100 : 1 (v/v) into LeMaster medium (*5*) containing Se-Met with lactose as a carbon source and cultured at 37°C for 24 h and harvested by centrifugation. The centrifuged pellet was resuspended in 20 mM Tris–HCl (pH 8.0), 50 mM NaCl and heated to 70°C for 11.5 min. All subsequent steps were performed at 4°C. After centrifugation, ammonium sulphate was added to the supernatant, and the 30–60% (w/v) fraction was applied sequentially to HiTrap-Q HP and HiTrap-Blue HP columns (Amersham Biosciences) with salt gradients in the presence of buffer containing 20 mM Tris–HCl (pH 8.0), 1 mM EDTA, 1 mM dithiothreitol (DTT) and 10 μM phenylmethanesulphonyl fluoride (PMSF). The purity of each sample was verified using SDS–PAGE, which confirmed the presence of a single band of ~45 kDa eluted from the HiTrap-Blue HP chromatography.

Expression and purification of GkPurD

Geobacillus kaustophilus purD was amplified by PCR from *G. kaustophilus* genomic DNA as the template. After confirmation of the nucleotide sequence, the DNA fragment was ligated into the expression vector pET–HisTEV carrying a tag sequence: mgsshhhhhhssgenlyfqgh (Novagen) at the *Nde*I/*Bam*HI sites. The *Gk*PurD expression plasmid was transformed into *E. coli* strain BL21 (DE3) (Novagen) for overexpression, and its recombinant strain was cultured in 2.5 l of LB medium supplemented with 50 μg/ml ampicillin at 37°C for 20 h. The centrifuged pellet was resuspended in 17 ml of 20 mM Tris–HCl (pH 8.0), 500 mM NaCl, 5 mM 2-mercaptethanol (2-ME), 1 mM PMSF, and 1 min of sonication was repeated 10 times. After removing the insoluble fraction by centrifugation, the supernatant was heated to 70°C for 13 min. All subsequent steps were performed at room temperature. After centrifugation, the supernatant was applied sequentially to HiTrap Chelating HP5, HiLoad 16/60 Superdex 200, HiTrap Chelating HP5 and RESOURCE Q. The purity of each sample was verified using SDS–PAGE, which confirmed the presence of a single band of ~37 kDa eluted from the RESOURCE Q chromatography. Between the HiLoad 16/60 Superdex 200 and HiTrap Chelating HP5 steps, Tobacco Etch Virus (TEV) protease was added to the sample but no cleavage of the His-Tag was observed.

Expression and purification of AaPurD

Aquifex aeolicus VF5 *purD* was amplified by PCR using *A. aeolicus* VF5 genomic DNA as the template. After confirmation of the nucleotide sequence, the DNA fragment was ligated into the expression vector pET-21a (Novagen) at the *Nde*I/*Bam*HI sites. The *Aa*PurD expression plasmid was transformed into *E. coli* BL21 (DE3) strain (Novagen) for overexpression, and its recombinant strain was cultured in 4.5 l of LB medium supplemented with 50 μg/ml ampicillin at 37°C for 20 h. The centrifuged pellet was resuspended in 35 ml of 20 mM Tris–HCl (pH 8.0), 500 mM NaCl, 5 mM 2-ME, 1 mM PMSF and 1 min of sonication was repeated 10 times. After removing the insoluble fraction by centrifugation, the supernatant was heated to 70°C for 13 min. All subsequent steps were performed at room temperature. After centrifugation, the supernatant was applied sequentially to HiPrep 26/10, SuperQ TOYOPERL 650 M, RESOURCE Q, CHT20-I and HiLoad 16/60 Superdex 200. The purity of each sample was verified using SDS–PAGE, which confirmed the presence of a single band of ~50 kDa eluted from the HiLoad 16/60 Superdex 200 chromatography.

Crystallization, X-ray diffraction data collection and data processing

The crystals of PurD samples were obtained by the hanging drop vapour diffusion method at 20°C. The crystallization drops were prepared by mixing, typically, 1 μl of protein solution with 1 μl of reservoir solution. Initial screening was performed by crystal screening kits including Crystal screen 1, 2, SaltRX, PEG/Ion and Index (Hampton research). Crystallization conditions were farther optimized. Typically, crystals appeared within a few days and reached maximum size over the course of an additional few weeks. Details of optimized conditions are shown in Table I. Crystals of PurD–ligand complex were obtained by using solutions containing 10 mg/ml protein and 10 mM of each ligand. We tried to obtain the PurD–ATP–Mg^{2+} and PurD–AMP–Mg^{2+} complexes by soaking apo crystals of *Gk*PurD and *Aa*PurD with ATP–Mg^{2+} or AMP–Mg^{2+}. However, X-ray diffraction of these crystals were poor and structures of the PurD–ATP–Mg^{2+} and PurD–AMP–Mg^{2+} complexes were not obtained.

X-ray intensity data used for phasing and refinements were collected at the SPring-8, beam line BL41XU, BL26B1 or BL26B2. Typically, a frozen crystal was used to measure 180° of data. Each data frame was collected with a 1° oscillation angle and 5–25 s exposure times. Camera length was 150–200 mm. Detail conditions are shown in Table II.

Collected data were processed by using HKL2000 (HKL Research, Inc.). Crystallographic parameters are also shown in Table II.

Structure determination of TtPurD

The crystal structure of *Tt*PurD was solved by the molecular replacement method using the Se-Met protein sample. A data set collected at the a wavelength of 0.9843 Å was processed with the program HKL2000 and an integrated data with the highest resolution of 2.8 Å was obtained. The phase was determined with the program AMORE (*6*) in the CCP4 package (*7*) by using a model structure based on the crystal structure of *Ec*PurD (PDB ID: 1GSO) (*4*) and refined by rigid body refinement and simulated annealing procedures with the program Crystallography & NMR System (CNS) (*8*). An atomic model was fitted into the electron density map using the graphics program O (*9*), and was refined by simulated annealing procedures with CNS. During the structure refinement, a run of simulated annealing with the non-crystallographic symmetry restraints was also applied. The refined model has R_{work} of 21.9% and R_{free}

Table I. The crystallization conditions.

Protein PDB ID		*Tt*PurD 2IP4	*Gk*PurD 2YS7	*Gk*PurD 2YRW	*Gk*PurD 2YRX	*Gk*PurD 2YS6	*Aa*PurD 2YYA	*Aa*PurD 2YW2
Protein sample	Buffer	20 mM Tris–HCl, pH 8.0, 150 mM NaCl, 1 mM DTT	20 mM Tris–HCl pH 8.0, 200 mM NaCl, 1 mM DTT	20 mM Tris–HCl pH 8.0, 200 mM NaCl, 1 mM DTT	20 mM Tris–HCl pH 8.0, 200 mM NaCl, 1 mM DTT	20 mM Tris–HCl pH 8.0, 200 mM NaCl, 1 mM DTT	20 mM Tris–HCl pH 8.0, 200 mM NaCl, 1 mM DTT	20 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM DTT
	Conc.	24.7 mg/ml	10 mg/ml	10 mg/ml	10 mg/ml	10 mg/ml	10 mg/ml	10 mg/ml
	Se-Met	Yes	No	No	No	No	No	No
Co-crystal					10 mM ATP	10 mM ATP, 10 mM Gly		10 mM ATP
Reservoir solution			CS1 4	Salt RX 52	CS2 13	PEG/ION Screen 39	INDEX 65	INDEX 65
		30% PEG MME 5000, 0.2 M ammonium sulphate, 0.1 M MES pH 6.5	0.1 M Tris–HCl pH 8.5, 2 M ammonium sulphate	2.4 M di-ammonium hydrogen phosphate, 0.1 M Tris–HCl pH 8.5	0.2 M ammonium sulphate, 0.1 M sodium acetate trihydrate pH 4.6, 30% w/v PEG MME2000	0.2 M sodium dihydrogen phosphate monohydrate, 20% w/v PEG 3350	0.10 M succinic acid pH 7.0, 15% w/v PEG 3350	0.10 M ammonium acetate, 0.1 M bis-Tris pH 5.5, 17% w/v PEG 10000
Crystallization		1 µl : 1 µl, vapour diffusion 20°C	1 µl : 1 µl, vapour diffusion 20°C	1 µl : 1 µl, vapour diffusion 20°C	1 µl : 1 µl, vapour diffusion 20°C	0.5 µl : 0.5 µl, vapour diffusion 20°C	1 µl : 1 µl, vapour diffusion 20°C	0.5 µl : 0.5 µl, vapour diffusion 20°C
Cryo protectant		10% glycerol	None	20% glycerol	20% glycerol, 10 mM ATP	20% glycerol, 10 mM ATP, 10 mM Gly	20% glycerol	none

of 23.8% at 2.8 Å resolution. The atomic coordinates of Se-Met *Tt*PurD have been deposited (RCSB ID: rcsb039841, PDB ID: 2IP4).

Structure determination of GkPurD and AaPurD

Data sets of *Gk*PurD and *Aa*PurD collected at the wavelength of 1.0 Å was processed and integrated with the program HKL2000. The crystal structures of *Gk*PurD and *Aa*PurD were solved by the molecular replacement method using the program MOLREP (*10*) in the CCP4 suite (*7*) by using a model structure based on the crystal structure of *Tt*PurD (PDB ID: 2IP4) and refined with the program CNS (*8*). During the structure refinement, a run of simulated annealing refinement, rigid-body refinement, pick water molecules in the electron density map and individual B-factor refinement were also applied. An atomic model was fitted into the electron density map using the graphics program XtalView/Xfit (*11*). The atomic coordinates of *Gk*PurD in four forms, apo, complexes with PO_4^{3-}, AMP/PO_4^{3-} and AMP/glycine, have been deposited (RCSB ID: rcsb027067, rcsb027056, rcsb027057, rcsb027066 and PDB ID: 2YS7, 2YRW, 2YRX, 2YS6, respectively). The atomic coordinates of *Aa*PurD in two forms, apo and ATP bound forms, have been deposited (RCSB ID: rcsb027286, rcsb027206 and PDB ID: 2YYA, 2YW2, respectively).

The Ramachandran plots were prepared by the programme MolProbity (*12*). Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization and Informatics at the University of California, San Francisco (supported by National Institutes of Health P41 RR-01081) (*13*). Residues around the bound ligands were examined by Chimera as well as the program LIGPLOT (*14*).

Molecular dynamics simulation

Molecular dynamics (MD) simulations were performed with the program AMBER 7 (*15*). Initial models were prepared by inserting missing residues using the modelling software ICM-pro (Molsoft). For *Tt*PurD, Se-Met residues were replaced by Met and, for *Gk*PurD and *Hs*PurD, tag-sequences in the N-terminal were removed. For each protein, after the removal of the crystallographic water molecules, the charge was neutralized by adding sodium ions and the protein was surrounded by TIP3 water molecules in a box with a buffer distance between the wall of the box and the closest atom in the solute of 8.0 Å. Summary of the model treatment can be found in the Supplementary data. An energy minimization in 1000 iterations was applied to the system in the presence of the positional restraint for the protein molecule with a force constant of 10.0. Temperature of the system was increased to 300 K in 10 ps (10,000 steps) and, then, a constant pressure simulation in 50 ps was performed to equilibrate the density of the system with a force constant of 5.0 for the positional restraints. The force constant for the positions restraintal was decreased to 0.5 and a constant volume simulation in 50 ps was performed. Finally, a productive simulation in constant volume without positional restraints was performed for 1 ns (1,000,000 steps). The rectangular periodic boundary condition and the particle-mesh Ewald procedure for the long-range electrostatic interactions were used and the non-bonded cutoff was 9 Å. The trajectory of the productive simulation was processed by the program ptraj in the AMBER suite (*15*).

Results and Discussion

Structures of TtPurD, GkPurD and AaPurD

Seven structures were solved and the crystal data and refinement statistics are summarized in Table II. Figure 2 shows the determined structures as well as the previously determined structures of GAR-syn's. In general, all six structures are similar to each other. However, position of the B domain, indicated by red, varied among structures; *Tt*PurD, *Gk*PurD and *Ec*PurD are in the open form, and *Aa*PurD, *Tm*PurD and *Hs*PurD are in the closed form.

Structure of TtPurD. Each asymmetric unit contains two monomeric molecules with 49% solvent content.

Table II. The crystal data and refinement statistics.

Protein	*Tt*PurD	*Gk*PurD	*Gk*PurD	*Gk*PurD	*Gk*PurD	*Aa*PurD	*Aa*PurD
Se-Met	Yes	No	No	No	No	No	No
Ligands	SO_4^{2-}	–	PO_4^{3-}	AMP, PO_4^{3-}	AMP, Gly	–	ATP
	2IP4	2YS7	2YRW	2YRX	2YS6	2YYA	2YW2
Beamline	BL41XU	BL26B1	BL26B1	BL26B2	BL26B2	BL26B2	BL26B2
Detector	MAR165	Jupiter210	Jupiter210	Jupiter210	Jupiter210	Jupiter210	Jupiter210
Wavelength	0.979	1	1	1	1	1	1
Oscillation angle	1	1	1	1	1	1	1
Oscillation range	180	180	180	270	180	270	270
Camera distance (mm)	150	150	180	160	200	200	150
Exposure time (s)	8	5	8	15	25	8	8
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_1$	$P1$
Cell dimensions	$a=72.33$,	$a=51.23$,	$a=51.24$,	$a=50.79$,	$a=50.79$,	$a=97.26$,	$a=51.23$,
	$b=98.10$,	$b=83.55$,	$b=83.62$,	$b=83.36$,	$b=83.43$,	$b=45.15$,	$b=83.55$,
	$c=121.20$	$c=94.73$	$c=95.47$	$c=96.45$	$c=96.54$	$c=103.48$,	$c=94.73$,
						$\beta=103.87$	$\alpha=97.65$,
							$\beta=102.98$,
							$\gamma=106.65$
Solvent content (%)	49	41	42	46	42	47	50
*V*m (Å^3/Da)	2.39	2.10	2.12	2.27	2.12	2.32	2.44
Nmol/asym	2	1	1	1	1	2	2
No. of water molecules	24	160	151	186	82	89	311
Resolution range of	50.00–2.80	41.78–2.21	41.81–2.20	48.23–1.90	38.29–2.21	49.55–2.40	35.42–1.80
reflections used (Å)	(2.90–2.80)	(2.34–2.21)	(2.34–2.20)	(2.02–1.90)	(2.34–2.21)	(2.55–2.40)	(1.91–1.80)
Completeness (%)	99.3 (97.1)	97.7 (84.3)	98.3 (95.8)	95.5 (88.6)	95.1 (80.4)	92.9 (82.5)	92.9 (82.1)
R-merge I (%)	10.0 (29.0)	5.5 (15.1)	7.0 (22.2)	10.2 (65.5)	8.9 (42.4)	6.9 (27.7)	5.5 (28.0)
Avg I/Avg σ(I)	19.8 (4.6)	56.4 (16.5)	38.2 (8.6)	24.4 (2.7)	18.5 (3.2)	19.9 (3.3)	18.9 (2.3)
Multiplicity	6.7 (5.7)	6.9 (6.7)	7.0 (7.2)	10.3 (9.4)	6.8 (6.3)	4.8 (4.2)	2.8 (2.8)
number of reflections	21939	20854	21504	32811	21104	33971	79536
R-factor (working) (%)	21.8	20.8	21.4	21.1	20.4	22.5	20.3
free *R*-factor (%)	23.8	25.1	23.2	24.4	23.0	24.9	22.7
Stereochemical ideality							
Bond	0.006	0.006	0.006	0.005	0.026	0.037	0.006
Angle	1.3	1.3	1.3	1.3	2.2	3.4	1.5
Dihedral	23.6	23.5	23.2	23.3	25.8	26.9	23.4
Improper	0.91	0.78	0.86	0.88	2.18	2.99	1.09
Ramachandran plot (%)[a]							
Favoured regions	91.0	97.8	97.1	97.8	97.8	91.3	97.7
Allowed regions	98.4	99.8	100.0	100.0	99.8	98.6	99.9
Outliers	1.6	0.2	0.0	0.0	0.2	1.4	0.1

[a]Analysd by the program MolProbity (*12*).

The structures of two molecules in the asymmetric unit are similar to each other with the backbone r.m.s.d. of 0.67 Å. The B domain shows structural difference between the two molecules (Fig. 3A); backbone r.m.s.d was 0.83 Å for the B domain whereas 0.35, 0.45, 0.50 Å for the N, A and C domains, respectively. For *Ec*PurD in apo form, the electron density is not observed for the B loop of the B domain. In contrast, the electron density of the B loop was clearly observed for *Tt*PurD although the B-factor of the region is rather high (Fig. 3B).

In the *Tt*PurD structure, two sulphate ions are located in the active site cleft, as described below, where probably phosphate groups of ligands bind to.

Structure of GkPurD. We determined the crystal structures of *Gk*PurD in four forms; apo, complexes with PO_4^{3-}, AMP/ PO_4^{3-} and AMP/glycine. Despite the differences in bound ligands, overall structures of the four forms are almost identical to each other (Fig. 4A).

Locations of bound AMP for the two forms are almost same and agree well to the predicted ADP binding site by Wang *et al.* (*4*) (Fig. 4B). The adenine base is recognized by Ala118, Glu185, Glu186, Tyr187, Leu188, Ile282, Glu283 and Met273; the ribose is recognized by Glu192, Lys214, Lys143 and Asn224; and the phosphate by Lys103. The positions of AMP in the two structures, AMP/ PO_4^{3-} and AMP/glycine, are slightly different as shown in Fig. 4B. This is probably because the protein is in the open form and the bound AMP is not strongly fixed. It should be noted that ATP but not AMP was used as the ligand for crystallization. Thus, there is a possibility that bound compound is ATP and the β- and γ-phosphate groups are disordered in the crystal.

Figure 4C shows the glycine bound to *Gk*PurD. The glycine is recognized by Asn285, Arg287, Gly289, Asp290, Pro291, Glu292, Tyr269 and Gly226. Especially, the side chain of Arg287 and backbone of Asp290 directly interact with the bound glycine. Lys214 and Asp212 are located between the glycine and adenine nucleotide.

Structure of AaPurD. We determined the crystal structures of *Aa*PurD in two forms; apo and ATP bound forms. For each form, two protein molecules

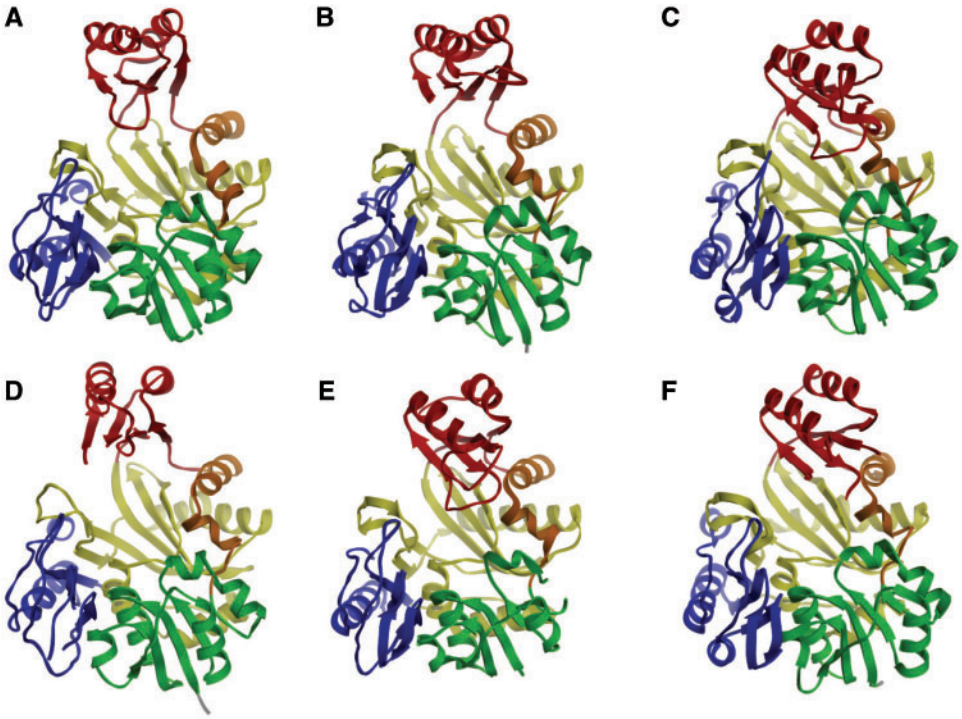

Fig. 2 Crystal structures of PurDs. (A) *Tt*PurD, (B) *Gk*PurD, (C) *Aa*PurD, (D) *Ec*PurD (1GSO), (E) *Tm*PurD (1VKZ) and (F) *Hs*PurD (2QK4). For all, molecule A was shown. The N, B′, B, A and C domains are indicated by green, orange, red, yellow and blue, respectively.

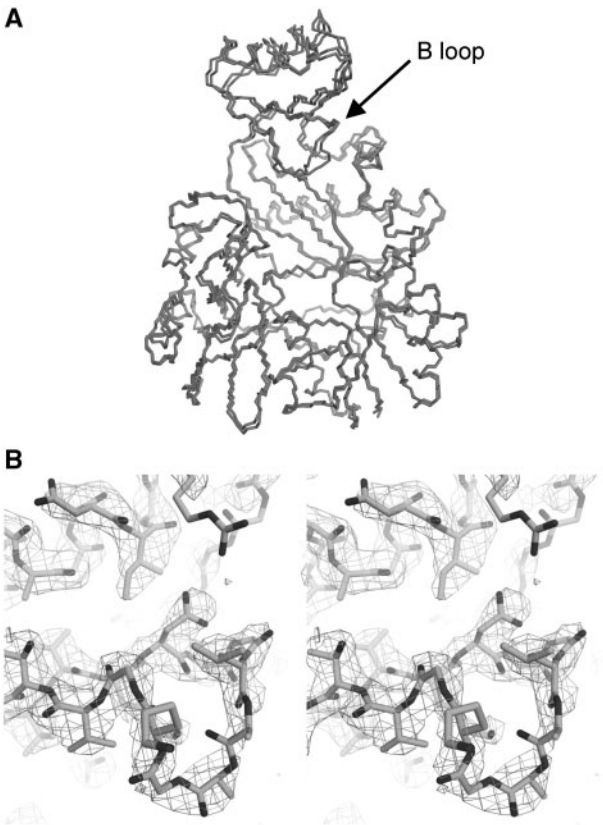

Fig. 3 Crystal structure of *Tt*PurD. (A) Superposition of two molecules in the asymmetric unit. (B) A stereo view of the electron density map around the B loop of *Tt*PurD. The $2F_{obs}$–F_{calc} map calculated by CNS was shown at the level of 1.0 σ.

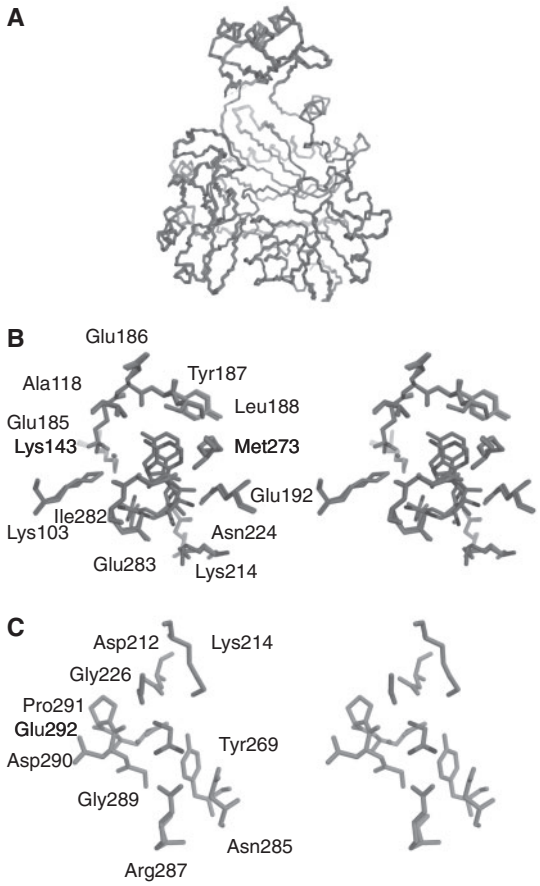

Fig. 4 Crystal structure of *Gk*PurD. (A) Superposition of the four structures of *Gk*PurD without ligands. (B) A stereo view of a superposition of the AMP binding sites and the bound AMP molecules of 2YRX and 2YS6. (C) A stereo view of the glycine binding site with the bound glycine molecule (2YS6).

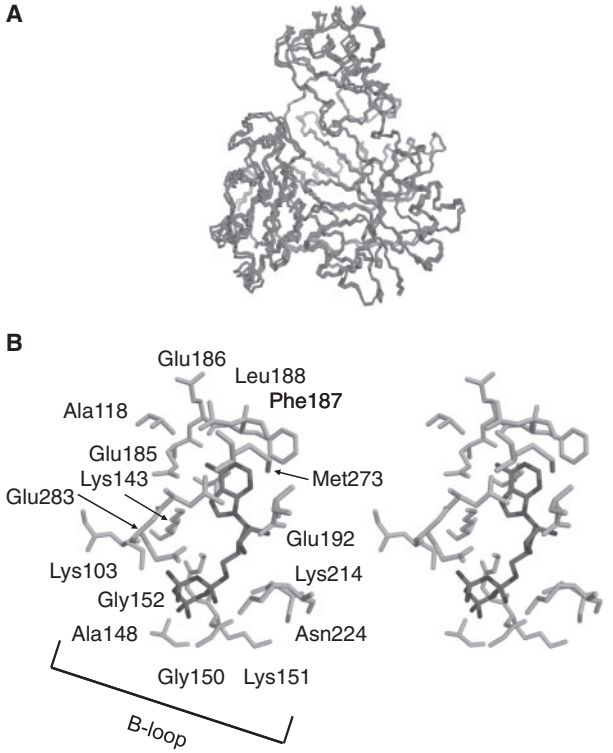

Fig. 5 Crystal structure of *Aa*PurD. (A) Superposition of the four structures of *Aa*PurD. (B) A stereo view of the ATP binding site with the bound ATP molecule (2YW2). Labels for Val141 and Leu282 holding the adenine base and Ala153 bound to the ribose are not shown.

are found in the asymmetric unit and the four structures are in general similar to each other (Fig. 5A).

Figure 5B shows the ATP binding site of *Aa*PurD. The adenine base is recognized by Glu185, Glu186, Phe187, Leu188, Leu282, Glu283 and Met273, and these residues are corresponding to those of *Gk*PurD. In addition to these seven residues, Val141 and Ala118 also located close to the adenine base. The ribose is recognized by Glu192, Lys214, Lys143, and these residues are also corresponding to those of *Gk*PurD. The α-phosphate group is recognized by Asn224, which is corresponding to Asn224 of *Gk*PurD. The β- and γ-phosphate groups of ATP are recognized by the B loop including Ala148, Gly150, Lys151, Glu152, Ala153 and Lys103 of the B′ domain (for domain definition, see below). Clearly, it was indicated that the B loop fixes the positions of the β- and γ-phosphate groups of ATP in the closed form.

Structural comparison among GAR-syn's (overall structure)

Sequence identities among *Tt*PurD, *Gk*PurD, *Aa*PurD and *Ec*PurD are 49–54%. Sequence identities of *Hs*PurD and *Tm*PurD to the other four PurDs are 44–50% and 32–40%, respectively (see Supplementary data). Despite the high sequence identities, the orientation of the B domain is varied among PurDs as mentioned earlier. The orientation of the B domain is not affected by the binding of adenine nucleotides. For example, *Gk*PurD is in the open form regardless of the binding of AMP whereas *Aa*PurD is in the closed form regardless the binding of ATP. As described below, MD simulation showed the hinge-like motion for all six PurDs suggesting that each PurD can take both the open and closed forms. Although the binding of glycine did not affect the B domain orientation of *Gk*PurD, binding of PRA may mediate the conformational change. Crystal packing can also affect the conformation and further analyses including crystallization with PRA and longer MD simulations are required.

Structure of the B and A domains

The B and A domains, which form the ATP binding site, are classified into the ATP grasp family (*16*, *17*). The ATP grasp family includes, for example, D-alanine:D-alanine ligase, biotin carboxylase and glutathione synthetase, despite the low sequence similarity. These enzymes all utilize similar ATP-dependent catalytic mechanisms even though they catalyse different chemical reactions.

Two α-helices connecting the N and B domains can be included in the N domains based on the structural inspection as is the case in ref (*4*). However, some residues of the helices are included in sequence of the ATP grasp family and the conserved lysine residue, Lys103 of *Aa*PurD, located in one of the helices interacts to the phosphorus group of bound ATP. Thus, the two helices structurally associates with the B domain rather than the N domain and may be called as the B′ domain. The structure of the B′, B and A domains are well conserved among the three enzymes; PurD, PurK and PurT (Supplementary data). The MD analysis also suggests that the B′ domain is structurally independent of the N domain as described below.

Structure of the N and C domains

The structures of the N domains or ribonucleotide binding domains including the P loop with the sequence of GGR are well-conserved among PurD, PurK, PurT and PurN. The structures and location of the C domain are also conserved among PurD, PurK and PurT. Probably because PurN does not contain the ATP binding domain, the relative location of the C domain is different from those of other three enzymes; the N and C domains of PurN are linked by a short linker polypeptide (Supplementary data). It should be noted that the structure of the C domain of PurN shows some similarity to those of other three enzymes, indicating the possibility that the PurD, PurK and PurT evolved from an ancestral enzyme which was produced by insertion of the ATP binding domain to the ancestral enzyme of PurN.

MD simulation of PurD

In order to examine the flexibility of the linker between the B domain and the N/A domains, we performed the MD simulation. Figure 6 shows the conformational fluctuation of each residues of *Tt*PurD as expressed by the scale of the B-factor. When overall structures

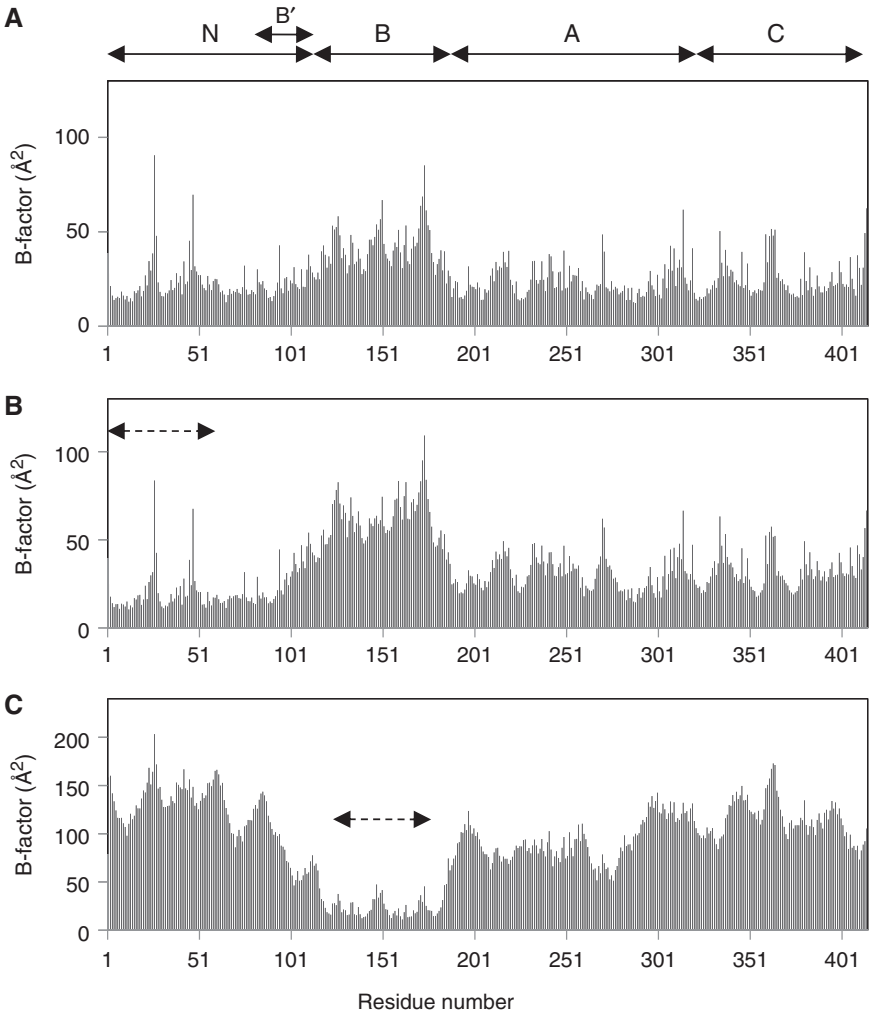

Fig. 6 Conformational fluctuations of *Tt*PurD deduced from the trajectory of the MD simulation. Atomic fluctuations are represented for each residue as the scale of the B-factor. (A) Each structure was superimposed to the first structure by all residues. (B) Each structure was superimposed to the first structure by a part of N-domain (1–60), as indicated by the arrow of the short dashed line. (C) Superimposition by a part of B-domain (123–176). Position of each domain is indicated by arrows on the top.

are superimposed, the B domain shows only slightly higher fluctuation than other domains as shown in Fig. 6A. However, when the first 60 residues of the N domain was superimposed, only the B domain shows high fluctuation (Fig. 6B), indicating that the A and C domains are fixed to the N domain whereas the B domain moves relative to other domains. In contrast, when the B domain is superimposed, other domains showed high fluctuation (Fig. 6C). The same tendencies were observed for the remaining five PurDs and the fluctuation of the B domain for the open form is larger than that for the closed form (Supplementary data). Although the movement of the B domain is not large enough to change from the open form to the closed form or vice versa in the current MD simulation, it was shown that the linkage between B domain and other domains has a tendency to exhibit the hinge-like motion. It should be noticed that the amount of fluctuation of the B′ domain is intermediate between those of the B domain and other domains, suggesting that this region is fixed neither of the B domain nor other domains. Thus, the movement of the B domain is due to the hinge-like motion at the B′ domain and the position 184 (*Tt*PurD) at the border of the B and A domains. The movement of the B domain probably contributes to the binding of the β- and γ-phosphate groups of ATP as well as the promotion of the enzymatic reaction, and the B′ domain may contribute to the movement of the B domain.

Ligand binding

To understand the catalytic mechanism of PurD, models of ligands bound to *Tt*PurD, in the open form, and *Aa*PurD, in the closed form, were prepared based on analyses described earlier. First, the PRA binding site was estimated by using the structure of *Ec*PurT with GAR (*18*). Figure 7A shows that, if the N domains of *Tt*PurD, *Gk*PurD and *Ec*PurT were superposed, the phosphate group of GAR bound to

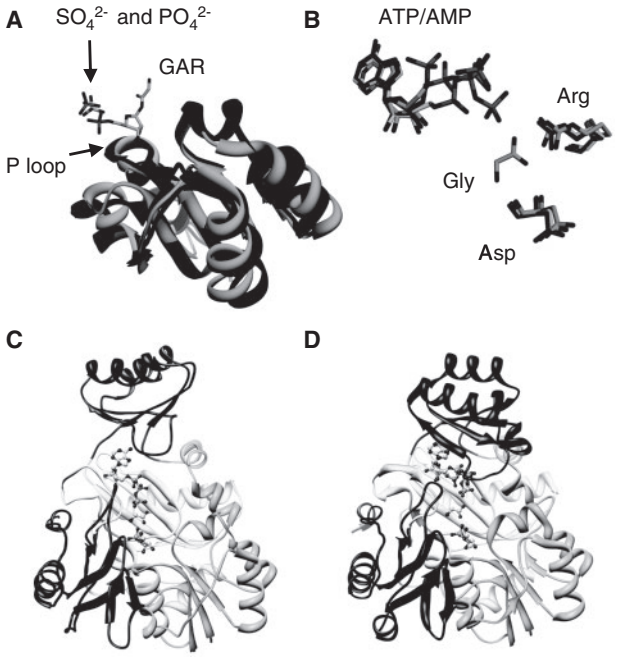

Fig. 7 Modelling of PurD with full ligands. (A) Superposition of the N domains of *Tt*PurD (2IP4), *Gk*PurD (2YRX) and *Ec*PurT (1KJ8). The SO_4^{2-} bound to *Tt*PurD, PO_4^{2-} bound to *Gk*PurD and GAR bound to *Ec*PurT are also shown. (B) The bound adenine nucleotides and glycine. The N, A and C domains are superimposed. The adenine bases on *Aq*PurD (2YW2), *Ec*PurT (1KJ8) and *Gk*PurD (2YS6) are overlapped well. The glycine recognizing Arg and Asp residues of *Gk*PurD as well as the corresponding Arg and Asp residues of *Tt*PurD and *Aq*PurD are also shown. (C) A model of *Tt*PurD with full ligands in the open form. (D) A model of *Aa*PurD with full ligands in the closed form.

*Ec*PurT is well overlapped to the SO_4^{2-} bound to *Tt*PurD and the PO_4^{2-} bound to *Gk*PurD. Thus, we put the GAR molecule on the *Tt*PurD as well as *Aa*PurD, and modified it to PRA by removing the glycine moiety. In the *Tt*PurD, PRA is recognized by Arg11 (backbone) and Glu12 in the N domain and Arg371 in the C domain and Asp284 in the A domain. The corresponding residues in *Aa*PurD are Arg11, Glu12, Arg378 and Arg290. These four residues are conserved among the six PurDs, and this recognition of PRA is likely common among PurDs. As already mentioned earlier, the structures of N, A and C domains can be well overlapped among PurDs analysed so far. Thus the glycine molecule can be located on the *Tt*PurD and *Aa*PurD just similar to that bound to *Gk*PurD. In *Gk*PurD, the glycine was recognized mainly by Arg287 and Asp290. In *Tt*PurD, Arg281 and Asp284 exist exactly on the same positions as the Arg287 and Asp290, respectively, of *Gk*PurD as shown in Fig. 7B. The Arg287 and Asp 290 of *Aa*PurD are also located exactly at the same positions as those of *Gk*PurD. An adenine nucleotide can be unambiguously located on *Tt*PurD because adenine nucleotides bind to the same position in *Aa*PurD, *Gk*PurD as well as *Ec*PurT (Fig. 7B). The orientation of γ-phosphate group of ATP is different between *Aa*PurD and *Ec*PurT, and the conformation on *Ec*PurT was chosen based on the vicinity to the bound glycine because the γ-phosphate group should be transferred to glycine molecule during the reaction. The presence of Mg^{2+} may cause the conformational difference of the phosphate groups between *Aa*PurD, without Mg^{2+}, and *Ec*PurT with Mg^{2+}.

It was reported that the kinetic mechanism of PurD is the sequential, ordered mechanism; PRA, ATP and glycine bind to PurD in this order and, after the reaction, PO_4^{2-}, ADP and GAR are released in this order (*19*). Figure 7C and D show the overall images of *Tt*PurD and *Aa*PurD with full ligands, respectively. Clearly, in the closed form (Fig. 7D), the bound ATP is covered by the B domain and may not be released from the protein. Thus, after the binding of ATP to the protein in the open form with PRA, the conformation of the protein changes to the closed form. Glycine must bind to the protein after PRA and ATP because the bound glycine seems to close the binding pathways of PRA and ATP. After the reaction, the conformation changes to the open form to release ADP and PRA. Accordingly, this model on the ligand binding is consistent with the reported kinetic mechanism.

Possible mechanism of reaction

Based on the structures discussed earlier, possible mechanism of reaction was proposed as shown in Fig. 8 with the residue numbers for *Aa*PurD. Starting with the binding of three ligands, ATP, glycine and PRA, as described earlier, the oxygen of the glycine may attack the γ-phosphorus (Fig. 8A). In the full ligand model, the oxygen of the glycine, the γ-phosphorus and the oxygen connecting the γ and β-phosphorus atoms are located in line, suggesting the bimolecular nucleophilic substitution reaction to form the glycylphosphate as an intermediate. This is consistent with the result by Shen *et al.* (*19*) showing that the oxygen of glycine was transferred to phosphate during the reaction. Following reaction should occur with the attack of the nitrogen of PRA to the carbonyl carbon of glycylphosphate as shown in Fig. 8B. However, in the current model, the distance between the reacting two atoms are too far to react, suggesting that further conformational change should occur during the reaction. In the crystal structures of *Tt*PurD (2IP4) and GkPurD (2YRX), SO_4^{2-} and PO_4^{2-} bind to the side chain of Arg11, respectively. The Arg11, located in the P loop of N domain, is conserved among the six PurDs and also are spatially close to the Arg287, which recognizes the glycine. Possible mechanism of the release of the phosphate is that Arg11 traps the produced phosphate to release as shown in Fig. 8C. In the second step of the reaction, which is supposed to be the nucleophilic acyl substitution, the ammonium group of PRA should be deprotonated to be an amino group and the carboxylate of the glycylphospahte should be protonated. Thus, another possibility of the role of Arg11 is to separate the

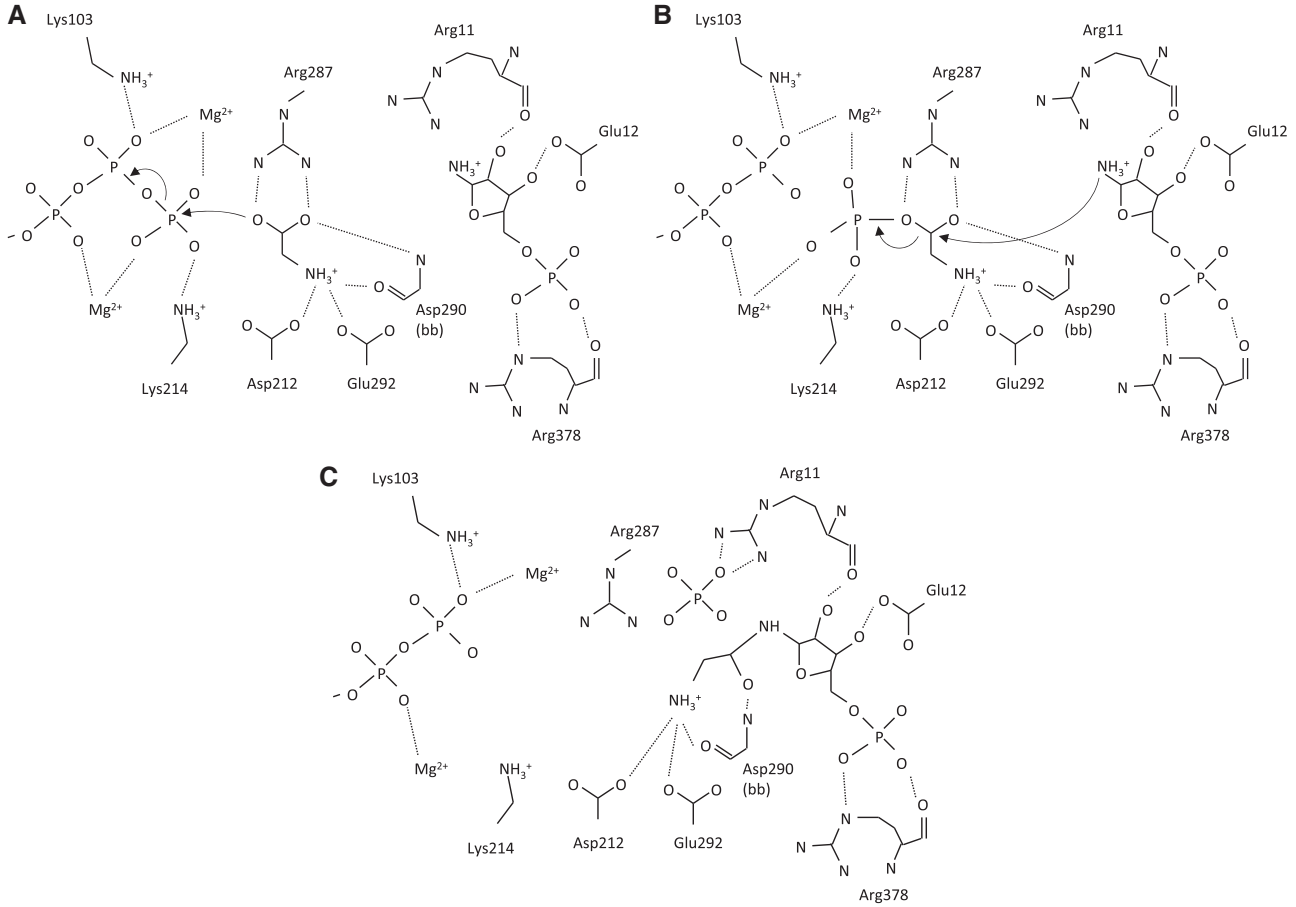

Fig. 8 Possible reaction mechanism catalyzed by PurD. (A) The reaction starts by the nucleophilic attack of the oxygen of the glycine to the γ-phosphorus. (B) The second stage of the reaction is the attack of the nitrogen of PRA to the carbonyl carbon of the intermediate, glycylphosphate. Before the reaction, the ammonium group of PRA should be deprotonated and the carboxylate of the glycylphospahte should be protonated. (C) At the last stage of the reaction, Arg11 may trap the produced phosphate to release.

carboxylate of the glycylphosphate from Arg287 to initiate the second step of the reaction.

Supplementary Data

Supplementary Data are available at *JB* Online.

Acknowledgements

We thank Drs A. Takénaka and O. Nureki for their useful comments and assistances on the protein crystallography. Some of synchrotron radiation experiments were performed on beamlines BL41XU and BL38B1 at SPring-8 with the approval of the Japan Synchrotron Radiation Research Institute (JASRI; proposal No. 2004A0721-NL1-np-P3k and 2008A1120, respectively).

Conflict of interest

None declared.

References

1. Zalkin, H. and Dixon, J.E. (1992) *De novo* purine nucleotide biosynthesis. *Prog. Nucleic Acid Res. Mol. Biol.* **42**, 259–287
2. Yokoyama, S., Hirota, H., Kigawa, T., Yabuki, T., Shirouzu, M., Terada, T., Ito, Y., Matsuo, Y., Kuroda, Y., Nishimura, Y., Kyogoku, Y., Miki, K., Masui, R., and Kuramitsu, S. (2000) Structural genomics projects in Japan. *Nature Struct. Biol.* **7**, 943–945
3. Rudolph, J. and Stubbe, J. (1995) Investigation of the mechanism of phosphoribosylamine transfer from glutamine phosphoribosylpyrophosphate amidotransferase to glycinamide ribonucleotide synthetase. *Biochemistry* **34**, 2241–2250
4. Wang, W., Kappock, T.J., Stubbe, J., and Ealick, S.E. (1998) X–ray crystal structure of glycineamide ribonucleotide synthetase from *Escherichia coli*. *Biochemistry* **37**, 15647–15662
5. LeMaster, D.M. and Richards, F.M. (1985) ^{1}H-^{15}N heteronuclear NMR studies of *Escherichia coli* thioredoxin in samples isotopically labeled by residue type. *Biochemistry* **24**, 7263–7268
6. Navaza, J. (1994) AMoRe: an automated package for molecular replacement. *Acta Cryst.* **A50**, 157–163
7. Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) A graphical user interface to the CCP4 program suite. *Acta Cryst.* **D59**, 1131–1137
8. Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse–Kunstleve, R.W., Jiang, J.S.,

Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., and Warren, G.L. (1998) Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Cryst.* **D54**, 905–921
9. Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991) Improved methods for the building of protein models in electron density maps and the location of errors in these models. *Acta Cryst.* **A47**, 110–119
10. Vagin, A. and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022–1025
11. McRee, D.E. (1999) XtalView/Xfit—A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165
12. Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, III W.B., Snoeyink, J., Richardson, J.S., and Richardson, D.C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383
13. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612
14. Wallace, A.C., Laskowski, R.A., and Thornton, J.M. (1995) LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng.* **8**, 127–134
15. Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham, III T.E., Wang, J., Ross, W.S., Simmerling, C.L., Darden, T.A., Merz, K.M., Stanton, R.V., Cheng, A.L., Vincent, J.J., Crowley, M., Tsui, V., Gohlke, H., Radmer, R.J., Duan, Y., Pitera, J., Massova, I., Seibel, G.L., Singh, U.C., Weiner, P.K., and Kollman, P.A. (2002) AMBER 7, University of California, San Francisco
16. Galperin, E. and Koonin, E.M. (1997) A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity. *Protein Sci.* **6**, 2639–2643
17. Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E.L., Eddy, S.R., and Bateman, A. (2009) The Pfam protein families database. *Nucleic Acids Res.* **38**, D211–D222
18. Thoden, J.B., Firestine, S.M., Benkovic, S.J., and Holden, H.M. (2002) PurT-encoded glycineamide ribonucleotide trasnformylase. Accomodation of adenosine nucleotide analogs within the active site. *J. Biol. Chem.* **277**, 23898–23908
19. Shen, Y., Rudolph, J., Stern, M., Stubbe, J., Flannigan, K.A., and Smith, J.M. (1990) Glycinamide ribonucleotide synthetase from *Escherichia coli*: cloning, overproduction, sequencing, isolation, and characterization. *Biochemistry* **29**, 218–227